Rev-Independent Export of Intron containing HIV-1 Env mRNAs

Rev-unabhängiger Export von Intron-haltigen HIV-1 Env mRNAs

Inaugural-Dissertation

zur Einlagen des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Anastasia Nicole Ritchie
aus Spokane, USA

Düsseldorf, October 2024

aus dem Institut für Virologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Heiner Schaal

2. Prof. Dr. Michael Feldbrügge

Tag der mündlichen Prüfung:

“The difference between screwing around and science is writing it down.”
Adam Savage

Abstract

Sequences in the HIV-1 late phase genes that lead to the inhibition of gene expression through nuclear retention are labeled as INS (instability) or CRS (*cis*-acting repressive sequences) elements. Despite extensive investigation on the HIV-1 life cycle, the exact mechanism underlying the nuclear retention of these late phase HIV-1 mRNAs in the absence of Rev is still unknown. We hypothesized that these RNA elements are bound by RNA-binding proteins, which are responsible for retaining these INS/CRS containing mRNAs in the nucleus. Previous research has implicated known RNA-binding proteins both the hnRNP and SR families in nuclear retention and export respectively. The HEXplorer algorithm, originally developed to analyze splicing regulatory sequence elements, can generally distinguish between sequences likely to bind these two protein families without specifying the exact protein. However, given how distinctive the binding regions of these two RNA binding protein families are, we can show that the described CRS/INS elements have a significantly higher probability of binding hnRNP than SR proteins.

In this work, mutated HIV-1 subgenomic expression vectors were created, guided by the HEXplorer algorithm to remove potential nuclear retention regulatory elements and introduce potential nuclear export regulatory elements without altering the underlying amino acid sequence. The results of this work show that the HEXplorer optimized HIV-1 *env* mutant mRNA were exported from the nucleus independent of Rev expression, albeit producing significantly less protein than the wild type *env*. This work also confirms that these Rev-independent *env* transcripts are still exported from the nucleus during CRM1 inhibition, hinting towards a switch in the export pathway. Concatenation and shortening of the mutated CRS region showed that while the length of the mutated region contributed to the nuclear export pattern, there were specific regions that exerted more influence on the nuclear export pattern of the transcript than others. Additionally, reduced production of HIV-1 Env protein led to decreased entry of HIV-1 pseudoviral particles despite similar levels of corresponding RNA, and a potential role for Rev in inhibiting NXF1-mediated nuclear export.

In addition to highlighting the role of splicing regulatory elements in other aspects of gene expression, the reduced production of HIV-1 Env protein and decreased viral entry suggest that the HIV-1's complex nuclear export system evolved to provide sufficient Env protein for productive viral entry. Overall, this work provides insight into the interconnected regulation of several gene expression steps and an algorithm that can predict and replicate elements controlling mRNA maturation and expression, in addition to predicting splice site usage.

Zusammenfassung

Sequenzen in den HIV-1-Spätphasengenen, die zur Hemmung der Genexpression durch nukleare Retention führen, werden als INS- (Instabilität) oder CRS-Elemente (cis-acting repressive sequences) bezeichnet. Trotz umfangreicher Untersuchungen des HIV-1-Lebenszyklus ist der genaue Mechanismus, der der nukleären Retention dieser Spätphasen-HIV-1-mRNAs in Abwesenheit von Rev zugrunde liegt, noch unbekannt. Wir stellten die Hypothese auf, dass diese RNA-Elemente durch RNA-bindende Proteine gebunden werden, die dafür verantwortlich sind, dass diese INS/CRS-haltigen mRNAs im Zellkern verbleiben. In früheren Untersuchungen wurden bekannte RNA-bindende Proteine sowohl aus der hnRNP- als auch aus der SR-Familie in die Kernretention bzw. den Kernexport einbezogen. Der HEXplorer-Algorithmus, der ursprünglich für die Analyse von Sequenzelementen zur Spleißregulierung entwickelt wurde, kann im Allgemeinen zwischen Sequenzen unterscheiden, die wahrscheinlich diese beiden Proteinfamilien binden, ohne das genaue Protein zu spezifizieren. Da die Bindungsregionen dieser beiden RNA-bindenden Proteinfamilien jedoch sehr unterschiedlich sind, können wir zeigen, dass die beschriebenen CRS/INS-Elemente eine deutlich höhere Wahrscheinlichkeit haben, hnRNP- als SR-Proteine zu binden.

In dieser Arbeit wurden mutierte subgenomische HIV-1-Expressionsvektoren erstellt, die durch nach der HEXplorer- Vorhersage generiert wurden, um potenzielle regulatorische Elemente für die Kernretention zu entfernen und potenzielle regulatorische Elemente für den Kernexport einzuführen, ohne die zugrunde liegende Aminosäuresequenz zu verändern. Die Ergebnisse dieser Arbeit zeigen, dass die mit HEXplorer optimierte HIV-1 env-Mutanten mRNA unabhängig von der Rev-Expression aus dem Zellkern exportiert wird, wenngleich sie deutlich weniger Protein produziert als der Wildtyp env. Diese Arbeit bestätigt auch, dass diese Rev-unabhängigen env-Transkripte auch während der CRM1-Hemmung aus dem Zellkern exportiert werden, was auf eine Änderung des Exportweges hindeutet. Die Verkettung und Verkürzung der mutierten CRS-Region zeigte, dass die Länge der mutierten Region zwar zum Kernexportmuster beiträgt, es aber bestimmte Regionen gibt, die einen größeren Einfluss auf das Kernexportmuster des Transkripts haben als andere. Darüber hinaus führte eine verringerte Produktion des HIV-1 Env-Proteins zu einem verringerten Eindringen von pseudoviralen HIV-1-Partikeln trotz ähnlicher Mengen an entsprechender RNA, was auf eine mögliche Rolle von Rev bei der Hemmung des NXF1-vermittelten Kernexports hinweist.

Die verringerte Produktion des HIV-1 Env-Proteins und der verringerte Eintritt des Virus deuten darauf hin, dass sich das komplexe Kernexportsystem des HIV-1 entwickelt hat, um genügend Env-Protein für einen produktiven viralen Eintritt bereitzustellen. Insgesamt bietet diese Arbeit einen Einblick in die miteinander verbundene Regulierung mehrerer Schritte der Genexpression und einen Algorithmus, der neben der Vorhersage der Nutzung von Spleißstellen auch Elemente vorhersagen und replizieren kann, die die

mRNA-Reifung und -Expression kontrollieren.

Table of Contents

1 Introduction

This work investigates the role of instability (INS) and cis-acting repressive sequences (CRS) in nuclear export pathway selection of HIV-1 late phase transcripts. As HIV-1 is dependent on host cell pathways for gene expression and maturation, this chapter will introduce the HIV-1 replication cycle, eukaryotic gene expression, and nuclear export pathways.

1.1 The Human Immunodeficiency Virus 1 (HIV-1) Replication Cycle

In 1981, several cases of advanced and unexplainable immunodeficiency began to appear in previously healthy individuals (Gottlieb, Schroff et al. 1981). Two years later, a novel T-lymphotropic retrovirus was isolated from a patient developing the acquired immunodeficiency syndrome (AIDS). Later called the human immunodeficiency virus (HIV), this virus was discovered to be the source of AIDS (Barré-Sinoussi, Chermann et al. 1983, Gallo, Sarin et al. 1983) by depleting the immune system through preferential infection of CD4+ cells of the immune compartment (Klatzmann, Champagne et al. 1984). In 1998, HIV-1 was found in several samples pre-dating the AIDS epidemic, with the earliest known isolate being from 1959 (Zhu, Korber et al. 1998). There is still no known cure for HIV-1, and it remains a major health issue worldwide. In 2022, there were approximately 39 million people living with HIV-1 and 1.3 million new cases (WHO 2023). While anti-retroviral treatment (ART) can suppress viral loads to undetectable levels, preventing transmission and the development of AIDS, the therapy is a lifelong commitment (Trono, Van Lint et al. 2010, Chun, Moir et al. 2015). Even though ART has dramatically increased life expectancy of those living with HIV-1, there are several issues with long term use, including treatment adherence, severe side effects, and drug resistance mutations (Clutter, Jordan et al. 2016, Hazen, Halbur et al. 2021). Drug resistant mutations are a major contributor to therapy failure, as alternative treatment is limited (Pennings 2013).

HIV-1 is a complex retrovirus that stably integrates with the host genome after reverse transcription of its RNA genome after which the virus is dependent on the host cell for gene expression and maturation. The virions are typically 100 nm in diameter and have a lipid membrane that is derived from the host. This membrane is studded with viral envelope glycoprotein trimers composed of HIV-1 proteins gp120 and gp41. Within the membrane, the two copies of the viral RNA genome and retroviral enzymes are contained within a conical capsid (Ramdas, Sahu et al. 2020). The HIV- replication cycle can be separated into several steps: attachment, membrane fusion, reverse transcription, integration, transcription, splicing, translation, assembly, budding, and maturation, described in further detail below.

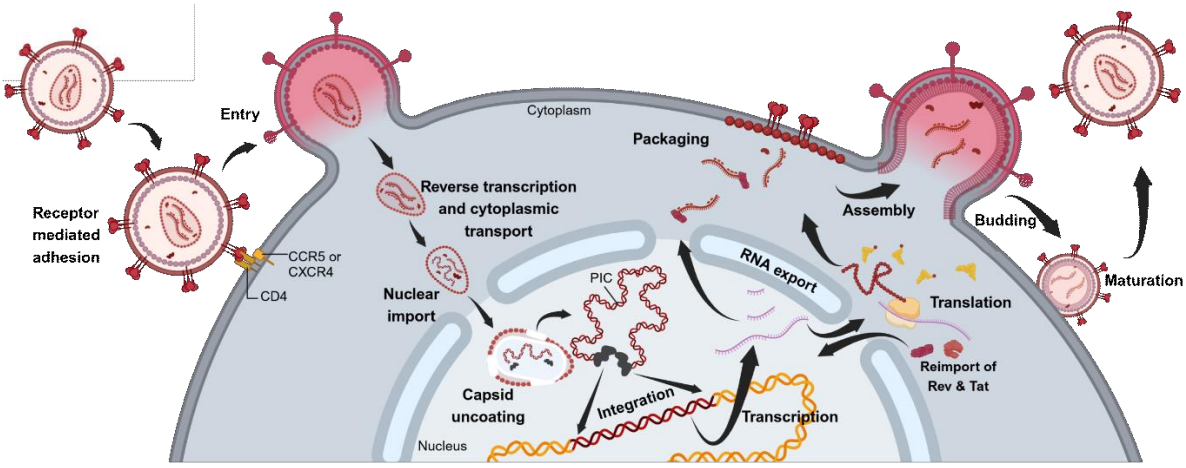


Figure 1.1.1. The HIV-1 replication cycle. The viral envelope protein (gp120) recognizes the cellular CD4 receptor and co-receptors (CXCR4 or CCR5), mediating attachment and viral fusion. The viral capsid is released into the host cell cytoplasm. As the capsid is transported to the nucleus, the viral genome is reverse transcribed into double stranded DNA (dsDNA). The capsid enters the nucleus and is uncoated to form the pre-integration complex (PIC). The viral dsDNA genome is inserted into the host cell genome by the viral integrase (IN), resulting in provirus formation. The HIV-1 provirus uses host cell mechanisms to generate three distinct mRNA classes: the fully spliced 2 kb class, the intron containing 4 kb class, and the unspliced 9 kb class. Once exported to the cytoplasm, the 2 kb class is translated into regulatory proteins, and the 4 kb and 9 kb classes are translated into structural and enzymatic proteins. The 9 kb mRNA also functions as genomic RNA for newly assembling virions. The newly assembled virions are released from the host cell by budding and mature into new infectious HIV-1 particles. This figure was created in Biorender and adapted from (Ramdas et. Al 2020).

1.1.1 Attachment and Membrane Fusion

First, the virus must attach to the host cell and fuse with the target cell plasma membrane. Attachment and recognition of the host cell is mediated by the viral envelope glycoprotein gp120 (Env), which recognizes the host cell cluster of differentiation receptor 4 (CD4), which is expressed on immune cells such as T-helper cells (CD4+ T-cells), macrophages (MQ), and dendritic cells (DCs) (Pope, Betjes et al. 1994). The binding of gp120 to CD4 induces conformational changes that allow gp120 to also bind to cellular co-receptors CXCR4 or CCR5 (Wilen, Tilton et al. 2012). Another conformational change allows the insertion of the hydrophobic Env subunit gp41 fusion peptide into the cell membrane. This ultimately results in fusion of the viral membrane with the host cell membrane (Doms and Moore 2000).

1.1.2 Uncoating, Reverse Transcription, and Nuclear Import

After successful fusion, the viral capsid is released into the host cell cytoplasm. The capsid contains two copies of the positive sense HIV-1 RNA genome and the virally encoded enzymes, reverse transcriptase (RT), integrase (IN), and protease (Pro) and other cellular proteins (Miller, Farnet et al. 1997). The intact capsid then commandeers the host microtubule network and migrates along the cytoskeleton to the host nuclear pore

(McDonald, Vodicka et al. 2002). The intact, or nearly intact capsid is imported into the nucleus by interacting with host protein cleavage and polyadenylation specificity factor 6 (CPSF6). The capsid uncoats very close to the genomic integration site (Burdick, Li et al. 2020). Figure 1.1.2 shows the process of reverse transcription, which occurs before the capsid uncoating in the nucleus (Selyutina, Persaud et al. 2020).The reverse transcription complex (RTC) contains the viral RT enzyme, and simultaneously transcribes the viral (+) RNA genome into a double stranded DNA copy and degrades the RNA genome (Fassati and Goff 2001).

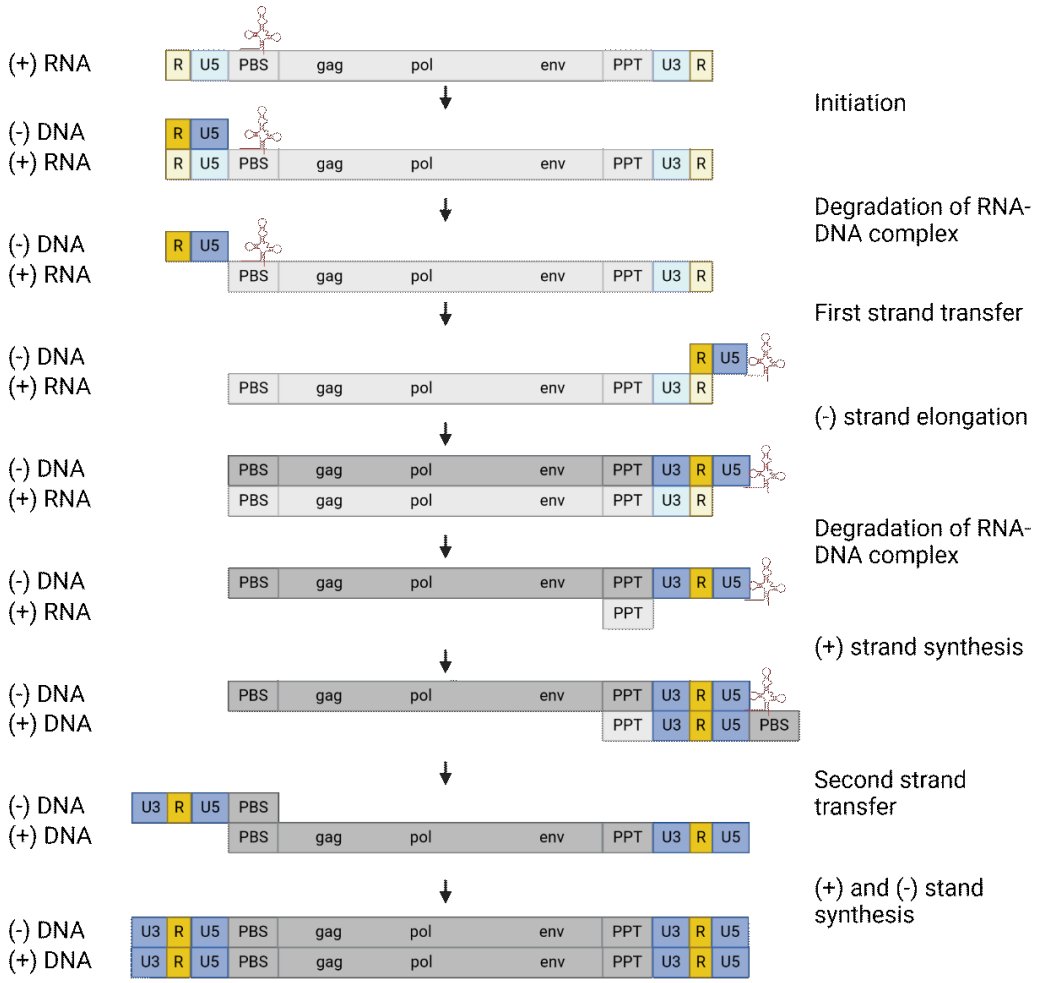


Figure 1.1.2. Schematic of HIV-1 reverse transcription (RT). The viral (+) strand RNA is bound by the tRNA Lys3 at the primer binding site (PBS), which primes reverse transcription initiation. The (-) DNA strand is synthesized from 5' to 3' until the 5' vRNA end. RNaseH activity of the RT degrades the RNA-DNA duplex. After the first strand transfer, the complementary repeated region (R) of the new (-) DNA strand and the (+) RNA, pair at the 3' end of the (+) RNA to continue the (-) DNA strand synthesis. The RNA-DNA duplex is degraded, except for the polypurine tract (PPT), which serves as a primer for the (+) DNA strand synthesis. The second strand transfer aligns the complementary PBS sequences and results in the elongation of both the (-) and (+) DNA strands, resulting in a double stranded DNA molecule enclosed by long terminal repeats (LTRs). Figure created with Biorender and adapted from (Hu and Hughes 2012).

In addition to the viral RT enzyme, the RTC also contains viral RNA, host-derived transfer-ribonucleic acid Lysine 3 ($tRNA^{Lys3}$) which primes reverse transcription, the eukaryotic translational elongation factor (eEF1A), newly synthesized DNA, and several other viral and host factors (Isel, Ehresmann et al. 1995, Fassati and Goff 2001). The $tRNA^{Lys3}$ binds the primer binding site (PBS) at the 5' end of the viral genomic RNA (vRNA) initiating the reverse transcription by RT with the help of several cellular and viral factors (Isel, Ehresmann et al. 2010, Sundquist and Kräusslich 2012). The minus strand complementary cDNA is synthesized from the PBS to the 5' end of the vRNA and the RT RNaseH activity cleaves the RNA directly after cDNA synthesis. The newly synthesized cDNA fragment transfers to the 3' end of the vRNA, using the complementarity of the repeated region (R) found in HIV-1's long terminal repeats (LTR). After this transfer, the (-) strand cDNA synthesis continues. After the RT RNaseH activity degrades the vRNA, except for the polypurine tract (PPT), the (+) strand cDNA synthesis occurs using the PPT as a primer, resulting in double-stranded DNA (dsDNA). The reverse transcription of the new (+) DNA strand ends at the 5' PBS of the template. At this point, the second-strand transfer occurs using the complementary PBS sequences and the RT continues on both DNA strands, generating the full-length dsDNA (Charneau, Alizon et al. 1992, Arhel 2010, Hu and Hughes 2012). Capsid disassembly releases the now dsDNA viral genome associated with host and viral factors in the pre-integration complex (Khiytani and Dimmock 2002).

1.1.3 Integration

Once the pre-integration complex is generated, the viral integrase (IN) enzyme inserts the viral dsDNA into transcriptionally active AT-rich euchromatin regions within the host genome in a process that is dependent on host proteins and nuclear architecture (Campbell and Hope 2015, Marini, Kertesz-Farkas et al. 2015, Ciuffi 2016). First, the integrase binds the newly synthesized dsDNA and orients the two LTR-containing ends into the IN's catalytic domain. The integrase then cleaves two nucleotides from the LTR ends resulting in a free 3'OH group on each end. Once the viral genome is ready for insertion, the integrase scans the host genome for a target insertion site, preferring short palindromic cytidine-adenine sequences (Wu, Li et al. 2005). Once the target site is found, the viral dsDNA is integrated via direct esterification initiated by the free electrons from the 3'OH ends on the LTRs. This reaction happens sequentially with a stable intermediate complex that has only one DNA end joined. The leftover 5' overhangs on the viral DNA are repaired by host ligases (Li, Mizuuchi et al. 2006). The integrated viral DNA is now called a provirus and relies on host machinery for gene expression.

HIV-1 genes are transcribed, spliced and exported to the cytoplasm for translation by host cell processes, in a tightly regulated process which is discussed in more detail in Chapter 1.2. This tightly regulated gene expression process occurs in two phases, the early and late phase, which is regulated by complex alternative splicing patterns. While regulatory elements are expressed during the early phase of HIV-1 gene expression, structural proteins are produced during the late phase. Once the structural viral proteins are synthesized, HIV-1 virions begin to assemble at the host cell membrane.

1.1.4 Assembly, Budding, and Maturation

The newly synthesized Gag and Gag-Pol proteins translocate to the host plasma membrane and form a hexameric network, which assembles with the membrane-bound gp120/gp41 Env trimers to package the viral genomic RNA dimers. The NC domain of Gag interacts with the Ψ-packaging signal in the 5'UTR of the plus-stranded viral RNA genome to promote packaging (Lever, Gottlinger et al. 1989, Rein 2019). The 5'UTR of the full-length HIV-1 mRNA can have multiple secondary structures, which either promote translation by exposing the Gag start codon and sequestering the dimerization initiation site (DIS), or promote packaging by exposing the DIS and occluding the Gag start codon (Abbink and Berkhout 2003, Sundquist and Kräusslich 2012). Viral genome packaging is regulated by recognition of the dimerized viral genome, allowing HIV-1 to strictly control the number of copies that are packaged into each mature virion. By packaging two copies of the viral genome, HIV-1 can productively infect if one of the genomes is damaged, as well as allowing recombination during reverse transcription to promote genetic drift (Nikolaitchik, Dilley et al. 2013).

When all the necessary components are gathered at the plasma membrane, the p6 domain recruits the endosomal sorting complex required for transport (ESCRT) to catalyze membrane fission. ESCRT brings the membrane sites into close contact so that membrane fission can occur, completing the budding process (Henne, Buchkovich et al. 2011, Sundquist and Kräusslich 2012). Budded viral particles are not fully matured and must undergo several steps to become infectious virions. Each released immature virion contains eight to ten Env trimers and a Gag layer that has a gap where the viral particles have been pinched from the membrane (Carlson, Briggs et al. 2008, Briggs and Kräusslich 2011). Maturation begins when the viral protease self cleaves and proceeds to cleave the Gag-Pol poly protein into active reverse transcriptase and integrase, and the 55 kDa Gag precursor into mature Gag proteins MA, CA, NC, and p6. Cleaving the Gag-Pol polyprotein allows the immature Gag lattice to reassemble into the mature conical capsid, which houses two copies of the single-stranded RNA genome, integrase, and reverse transcriptase (Pornillos, Ganser-Pornillos et al. 2011, Sundquist and Kräusslich 2012, Freed 2015).

1.2 Eukaryotic Gene Expression and HIV-1 Expression Strategy

The majority of a cell's genetic information is encoded in DNA within the nucleus. This code is expressed through a series of steps to produce proteins necessary for cell survival. Tight regulation of gene expression in eukaryotic cells is essential and disruption of this delicate process often leads to disease in the given organism (Levine and Tjian 2003, Herz, Hu et al. 2014).

1.2.1 Transcription

Gene expression typically begins with transcription of genomic DNA, which is restricted to the nucleus, into a complementary RNA strand by RNA polymerase II (Pol II). Transcription begins with initiation, and once the transcription machinery is recruited, new nucleotides are added to the RNA strand in a process called elongation. When the full gene has been transcribed, Pol II will pause and release both the DNA template and nascent RNA, after which Pol II can be recruited to the initiation site for another gene. Multiple polymerases can simultaneously transcribe the same gene.

In eukaryotes, there are three separate RNA polymerases; I, II, and III (Roeder and Rutter 1969). RNA Polymerase I transcribes genes encoding 18S and 28S ribosomal RNAs (rRNAs) and RNA Polymerase III transcribes genes encoding the 5S rRNA and tRNAs. Here I focus on RNA Polymerase II (Pol II), which is responsible for generating messenger RNAs (mRNAs) that encode protein-coding sequences (Roeder and Rutter 1970). Transcription is initiated at the transcriptional start site (TSS), a defined position imbedded in a core promoter region at the 5' end of a given gene. This core promoter region serves as a binding platform for the transcriptional machinery consisting of Pol II and its general transcription factors (GTFs). While the TSS alone can only initiate a basal level of transcription, it can be further enhanced by the presence of enhancers, regulatory elements which bind transcription factors and co-factors(Banerji, Rusconi et al. 1981, Yáñez-Cuna, Dinh et al. 2012). There are three main types of Pol II core promoters identified by mapping endogenous transcription initiation sites (Smale and Kadonaga 2003, Kadonaga 2012).

First, the best described core promoter elements are the TATA-box and initiator (Inr) motifs. These motifs have sharp initiation patterns and are primarily active in terminally differentiated adult tissue cells. They have a strong association with genes that show tissue-specific expression (Yamashita, Suzuki et al. 2005, Lenhard, Sandelin et al. 2012). The highly conserved TATA-box motif (TATAWAAR, W = A or T, R = A or G) is usually found 30 nucleotides upstream of a single dominant TSS (Grosschedl and Birnstiel 1980, Ponjavic, Lenhard et al. 2006) and is recognized by a GTF called the TATA-box-binding-protein (TBP), a subunit of TFIID (Transcription Factor, RNA Polymerase II, D). Shown in Figure 1.2.1., TBP binding recruits Pol II and mediates the assembly of the pre-initiation

complex (PIC), DNA duplex melting and pre-messenger RNA (pre-mRNA) synthesis (Lenhard, Sandelin et al. 2012). The Inr is a very common core promoter element found at the +1 transcription start site and typically has a consensus sequence of $YYA_{(+1)}NWYY$ (W = A or T, Y = C or T, N = A, T, C, or G) (Javahery, Khachi et al. 1994). In the absence of a TATA-box, the Inr can initiate basal transcription by recruiting the TFIID to begin PIC assembly. The Inr can also cooperate with other core promoter motifs, however this cooperation is sensitive to their distance from the A+1 position (Javahery, Khachi et al. 1994, Kaufmann and Smale 1994).

Second, the core promoters of ubiquitously expressed genes, such as housekeepers, show a more dispersed transcription initiation pattern than the TATA-box and Inr motifs. They tend to lack TATA-box motifs and have a short CpG island that overlaps with the TSS (Yamashita, Suzuki et al. 2005). These unmethylated CpG islands are likely to contain transcription factor Sp1 binding motifs (GGGCGG), which then lead to TFIID recruitment through intermediary cofactors (Pugh and Tjian 1990, Walker, Faik et al. 1990). This core promoter type is also used by SV40 coupled with a TATA-box, however the Sp1 binding sites are sufficient to initiate transcription when the TATA-box is removed (Dynan and Tjian 1983, Everett, Baty et al. 1983).

And third, the developmentally regulated core promoters bind key transcription factors involved in embryo development and morphogenesis. These bivalent promoters are characterized by large CpG islands that often extend into the gene itself and markers for repression by Polycomb group proteins (PcG) (Schwartz, Kahn et al. 2010). They tend to have only an Inr element, or an Inr supported by a downstream promoter element. This enrichment of downstream promoter elements compared to other promoter types is compatible with the increase of backtracking, where the polymerase moves backward on the DNA template and displaces the RNA 3’ end from the active site, observed at these promoter sites (Engström, Ho Sui et al. 2007, Nechaev, Fargo et al. 2010). Additionally, these promoters show low levels of transcription until differentiation, when they become transcriptionally active or repressed according to pathway specific gene expression requirements (Bernstein, Mikkelsen et al. 2006, Lenhard, Sandelin et al. 2012).

Once a stable PIC is formed, PIC subunit TFIIH will phosphorylate Ser5 of the C terminal domain (CTD) of Pol II to begin elongation. After transcribing a short stretch of RNA, Pol II will become trapped in a promoter proximal paused position found in most genes. This pause in transcription is regulated by unphosphorylated Negative Elongation Factor (NELF) and DRB Sensitivity Inducing Factor (DSIF). This pause in transcription is alleviated when CDK9, a subunit of the positive transcription elongation factor B (p-TEFb), phosphorylates Ser2 of the CTD and Spt5 subunit of DSIF, after which the polymerase will continue RNA synthesis (Yamaguchi, Takagi et al. 1999, Shandilya and Roberts 2012).

During elongation, the phosphorylated CTD of Pol II will recruit the capping enzyme, a set of proteins consisting of triphosphatases, guanyl transferases, and methyl transferases, which deposits a m7G cap on the 5' end of the transcript when the transcript is 20-40 nucleotides long. This cap is vital for protection from exonuclease attack, and for downstream functions such as nuclear export and translation (Cho, Takagi et al. 1997, Komarnitsky, Cho et al. 2000). Eventually, Pol II will come across a polyadenylation signal (AAUAAA) followed by a GU-rich sequence which allows the recruitment of the cleavage/polyadenylation machinery. As shown in Figure 1.2.1., the cleavage and polyadenylation factor (CSPF) and cleavage stimulatory factor (CstF) which are associated with Pol II during elongation, will then bind to the polyadenylation signal and GU-rich region respectively leading to cleavage and release of Pol II (Kuehner, Pearson et al. 2011, Shandilya and Roberts 2012). After cleavage and release of Pol II, the polyadenylation polymerase (PAP) will polyadenylate the 3' end of the transcript and the portion of the transcript left with the Pol II is quickly degraded by the exonuclease RatI/Xrn2 (Rodríguez-Molina and Turtola 2023). After the release of Pol II, the pre-mRNA will be further processed and matured before being exported to the cytoplasm for translation and Pol II will be returned to its hypophosphorylated state so that it can initiate another round of transcription. Several GTFs can remain bound to the promoter after Pol II has escaped the PIC to allow for reinitiation. It has also been shown that the promoter and termination signal can interact through gene looping where TFIIB interacts with CSPF and CstF to facilitate rapid reinitiation (El Kaderi, Medler et al. 2009, Shandilya and Roberts 2012).

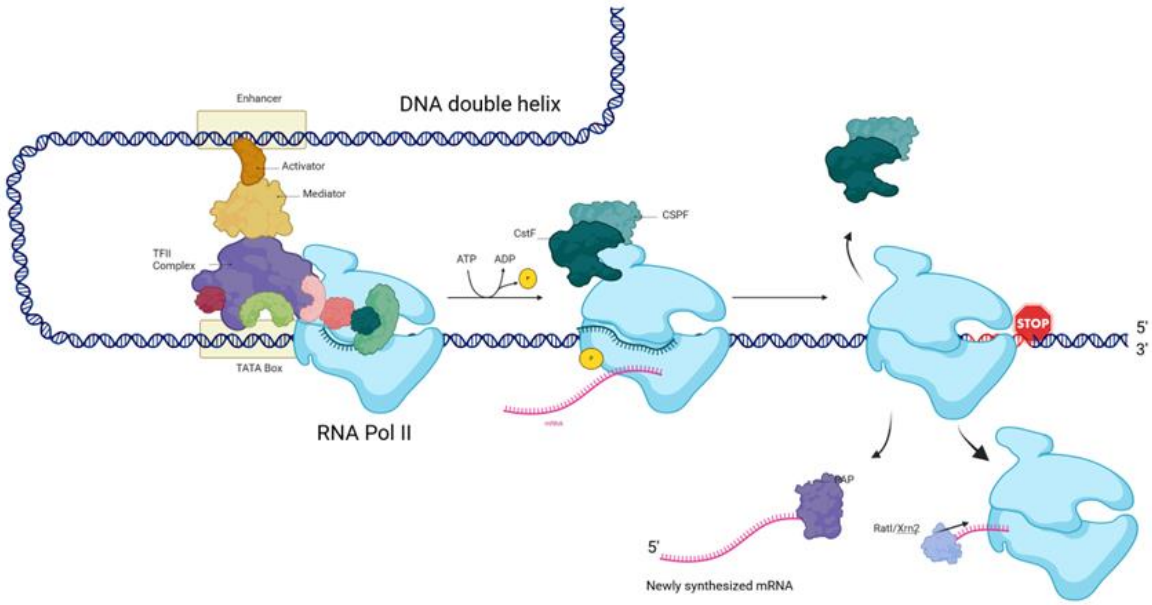


Figure 1.2.1. A model of eukaryotic transcription. Step 1 begins with initiation through recruitment to the promoter using the TFII complex and an enhancer. The cleavage and polyadenylation factor (CSPF) and cleavage stimulatory factor (CstF), which are associated with RNA Pol II during elongation initiate termination, and dissociate from the polymerase. The fully synthesized mRNA is released, and the polyadenylation polymerase poly adenylates the 3' end of the newly synthesized mRNA. The Ran fragment remaining with RNA Pol II is degraded by RatI/Xrm2. Imaged created with Biorender. Adapted from (El Kaderi, Medler et al. 2009).

1.2.1.1 Transcription in HIV-1

Previously discussed in 1.1.1, RNA Polymerase II (Pol II) requires a promoter to begin transcription. In the context of HIV-1, this promoter sequence is imbedded in the long terminal repeat (LTR), which is found on both ends of the integrated genome. Although both LTRs can function as promoters, the 5' LTR promoter exhibits significantly higher transcriptional activity (Klaver and Berkhout 1994). The LTR promoter initiates transcription efficiently, but complete transcription of the provirus is impaired due to poor elongation efficiency of the polymerase. HIV-1 encodes its own potent trans-activator protein, Tat. The poor basal transcription of the LTR produces sufficient Tat to stimulate transcription elongation for productive transcription. Tat interacts with the transactivation response region (TAR) of HIV-1 and recruits transcription elongation factor p-TEFb to the nascent RNA polymerases (Wei, Garber et al. 1998, Karn and Stoltzfus 2012). This produces a positive feedback loop and increases transcription by 200-300-fold (Sodroski, Rosen et al. 1985, Kao, Calman et al. 1987, Karn 1999). Transcription of the proviral DNA results in a full-length pre-mRNA, which contains open reading frames (ORFs) that encode at least eighteen proteins and polyprotein isoforms (Jäger, Cimermancic et al. 2011).

1.2.2 Splicing

Two decades ago, the human genome project found that our genome encodes approximately 20,000 protein-coding genes with a median length of 1.3 kb and an average of 7.5 introns (International Human Genome Sequencing 2004, Hong, Scofield et al. 2006). These introns are removed from the pre-mRNA by essential splicing machinery to form a mature protein-coding mRNA. 95% of these transcripts can produce multiple protein isoforms from a single pre-mRNA through a process called alternative splicing (Pan, Shai et al. 2008, Wang, Sandberg et al. 2008). Higher organisms, such as humans, use alternative splicing as a tool to support complexity by expanding the proteome that can be produced from a limited number of genes. However, with increased complexity comes increased opportunity for disease (Cartegni, Chew et al. 2002, Ule and Blencowe 2019). Splicing initiation typically depends on the recognition of three different elements, the 5' splice site (splice donor; SD), the 3' splice site (splice acceptor; SA), and the branch point sequence (BPS) shown in Figure 1.2.5.

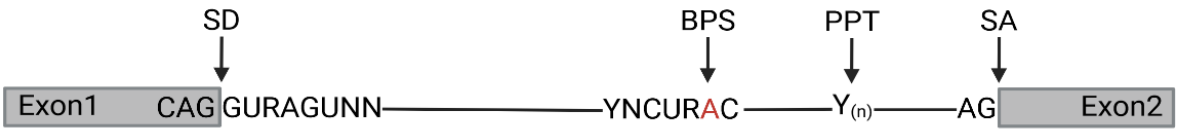


Figure 1.2.2. Architecture of a human intron. The conserved sequences at the splice donor (SD), branch point sequence (BPS), polypyrimidine tract (PPT), and splice acceptor (SA) are shown. The grey boxes indicate the flanking exons, and the black line represents the intron. R = A or G, Y = U or C, and N = A, U, G, or C. adapted from (Will and Lührmann 2011).

The splice donor is located at the 5’ end of the intron and is defined by an 11-nucleotide sequence containing a highly conserved GU at positions +1 and +2 (CAG/**GU**RAGUNN (R = A or G, N = A, G, U, or C, “/” indicates the exon-intron border)). The splice acceptor is located at the 3’ end of the intron to be excised and contains a highly conserved AU at positions -2 and -1(Smith, Chu et al. 1993). The BPS is typically found 100 - 180 nucleotides upstream of the splice acceptor and is recognized by the sequence YNCUR**A**C (Y = U or C, R = A or G, N = A, G, U, or C). In higher eukaryotes, a polypyrimidine tract (PPT) can also be found between the BPS and splice acceptor (Senapathy, Shapiro et al. 1990, Lopez and Séraphin 1999, Clark and Thanaraj 2002, Sheth, Roca et al. 2006).

The splicing reaction is based on two transesterifications that are catalyzed by the spliceosome and which are modeled in Figure 1.2.6. The spliceosome is composed of five uridine-rich small nuclear RNAs (U snRNAs) that are associated with a specific set of proteins to form small nuclear ribonucleoproteins (U snRNPs: U1, U2, U4, U5, U6) (Lerner and Steitz 1979). The uridine-rich sequence found in U snRNAs, known as the Sm site, is located at the 3’ end of U1, U2, U4, and U5 snRNAs. Sm and like-Sm proteins will form a ring around these sequences (Raker, Hartmuth et al. 1999, Urlaub, Raker et al. 2001).

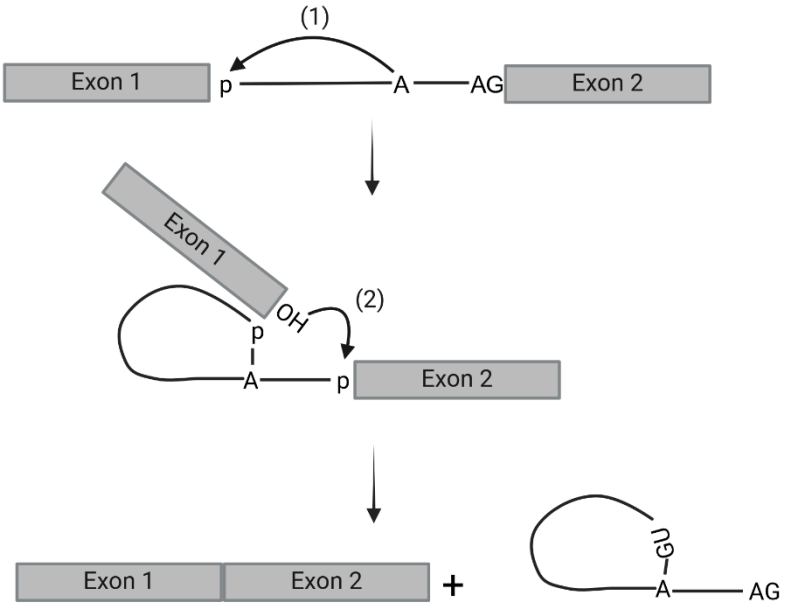


Figure 1.2.3. The two-step transesterification reaction in pre-mRNA splicing. In the first transesterification reaction, the OH group from the branchpoint adenosine attacks the phosphate group at the splice donor. This forms the intron lariat and brings the exposed free OH group from the splice donor into proximity with the splice acceptor. The free splice donor OH group attacks the phosphate group at the splice acceptor resulting in exon-exon fusion and release of the intron lariat. The grey boxes indicate exons, and the black line represents the intron. Adapted from (Will and Lührmann, 2011).

The early spliceosomal complex (E-complex) is the first complex formed during spliceosome assembly and depends on the correct recognition of splice sites. First, the U1 snRNP will recognize and bind to the splice donor through base pair complementarity

between the 5’ U1 snRNA and the 11-nucleotide splice donor site. While a strong splice donor will have high complementarity to the 5’ U1 snRNA (Zhuang and Weiner 1986), this interaction is rather weak and typically stabilized by SR proteins or the cap binding structure (Staknis and Reed 1994, Pabis, Neufeld et al. 2013). There is also evidence that the CTD of Pol II is involved in recruiting the U1 snRNP to the splice donor, as splicing occurs cotranscriptionally (Zhang, Aibara et al. 2021). In addition to U1 binding the splice donor, splicing factor 1 (SF1) will bind to the BPS and two U2 snRNP auxiliary factor (U2AF) subunits, U2AF65 and U2AF35 bind to the PPT and AG of the splice acceptor, respectively (Matera and Wang 2014). Following U1-splice donor binding, DEAD-box Helicases Prp5 and Sub2 displace SF1, and recruit the U2 snRNP to the BPS, forming the branch point helix (Perriman and Ares 2010, Liang and Cheng 2015). This forms the pre-spliceosome or A-complex.

Once the A-complex is formed, the U1 and U2 snRNP associated components interact with each other, bringing the splice donor and splice acceptor into closer proximity and forming the exon recognition complex (Parker, Siliciano et al. 1987, Zhuang, Goldstein et al. 1989). At the branch helix formed by the U2 snRNP and BPS, the branch point adenosine is exposed without the 2’OH group being accessible for transesterification (Plaschka, Lin et al. 2017). At this point, the inactive spliceosome, also called the pre-B-complex, can be formed by recruiting the tri-snRNP U4/U6.U5. The pre-B-complex will remain catalytically inactive until the U6 snRNA is no longer chaperoned by the U4 snRNA and the splice donor is released by the U1 snRNP (Bringmann, Appel et al. 1984, Hashimoto and Steitz 1984, Nguyen, Galej et al. 2015, Agafonov, van Santen et al. 2016). The DEAD-box helicase, Prp28 will initiate spliceosome activation by transferring the splice donor from the U1 snRNP to the flexible loop withing the U6 snRNA (Staley and Guthrie 1999, Charenton, Wilkinson et al. 2019). Additionally, the U5 snRNA loop 1 will stabilize the 5’ exon during branching and later for alignment of the 5’ and 3’ exons during exon ligation (Newman and Norman 1992, Sontheimer and Steitz 1993, Zhan, Yan et al. 2018).

Once the U6/splice donor duplex is formed, the spliceosome will undergo major conformational changes and release the U1 snRNP to form the B-complex. The active site will be formed when the RNA helicase Brr2 separates U4 and U6 (Raghunathan and Guthrie 1998) folds and associates the U6 snRNA with part of the U2 snRNA (Zhan, Yan et al. 2018, Charenton, Wilkinson et al. 2019). Although the catalytic core is fully formed, the distance between the BPS adenosine’s 2’OH group is still too far from the phosphate group of the splice donor’s guanine to mount the nucleophilic attack necessary for the first transesterification reaction (Haselbach, Komarov et al. 2018). To activate the spliceosome and enable branching, DEAH-box ATPase Prp2 remodels the spliceosome so that the BP adenosine moves into the active site and forms the catalytically active B*-complex (Will and Lührmann 2011).

The 2'OH group from the BPS adenosine attacks the phosphodiester group at the splice donor guanine, resulting in a cleaved 5'exon and the formation of the lariat intron/3'exon intermediate. This will leave the 5'exon guanine linked to the BPS adenosine and form the C-complex (Galej, Toor et al. 2018). The C-complex transforms into the C*-complex when the active site remodels and the splice acceptor docks (Fica, Oubridge et al. 2017). The next transesterification reaction is referred to exon ligation and fuses the two exons bookending the intron being excised. The newly exposed 3'OH of the 5' exon attacks the phosphodiester group at the splice acceptor, resulting in the ligation of the two exons, and the release of the lariat intron (Wahl, Will et al. 2009). The spliceosome then forms the post-catalytic complex, or P-complex, where the release of newly ligated exons is catalyzed by the DEAH-box ATPase Prp22 (Company, Arenas et al. 1991, Schwer 2008). During the splicing process, an exon-exon junction complex (EJC) is deposited approximately 20-24 nucleotides upstream of the newly ligated exons by spliceosome component CWC22 (Steckelberg, Boehm et al. 2012). These EJCs play vital roles in nuclear export, translation and degradation of these mRNAs (Boehm and Gehring 2016).

In higher eukaryotes, variability in splicing pattern is a major source of protein diversity and key to genetic regulation (Black 2003). In addition to constitutive splicing, 95% of human genes are alternatively spliced (Pan, Shai et al. 2008). There are several different forms of non-canonical splicing, including exon skipping, intron retention among others which are depicted in Figure 1.2.7.

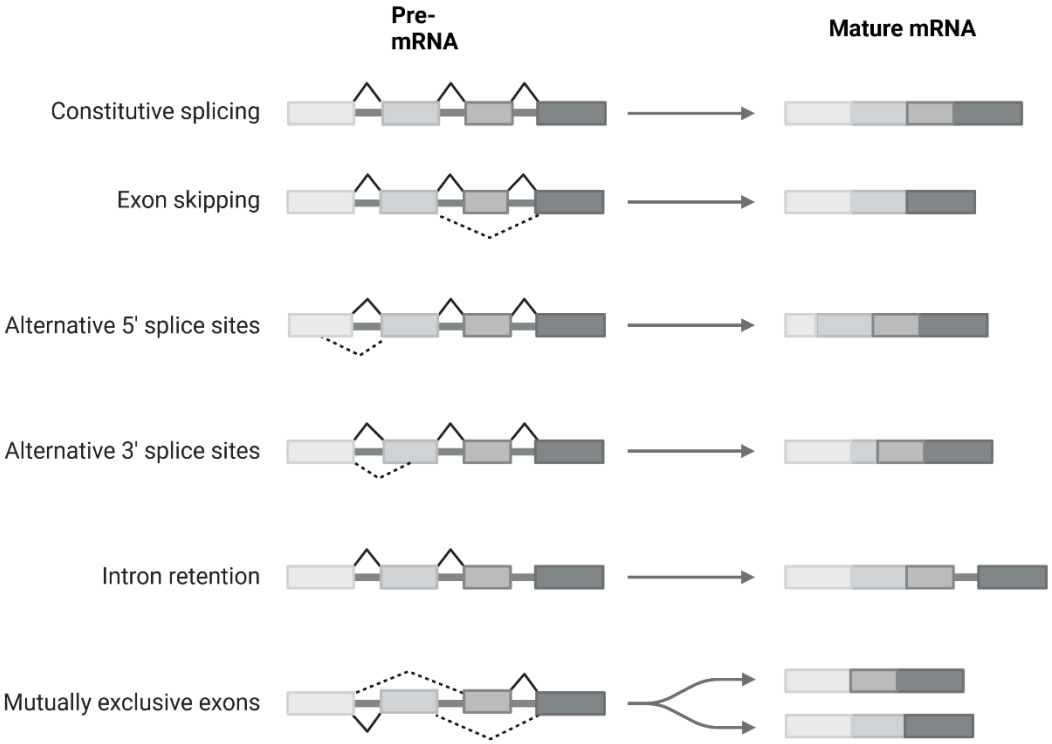


Figure 1.2.4. Schematic of alternative splicing events. Constitutive splicing results in a mature mRNA that contains all designated exons, alternative splicing allows for modification of the final mature mRNA transcript. The methods by which alternative splicing can alter the mature transcript are shown above. Boxes indicate exons and the grey lines represent introns.

These alternative splicing events are typically not due to spliceosome inaccuracy, but rather pressure from *trans*-acting regulatory proteins on splice site selection (Fox-Walsh and Hertel 2009). While SR proteins enhance the usage of downstream splice donors and upstream splice acceptors, they repress the usage of upstream splice donors and downstream splice acceptors. The opposite is true for the hnRNP family (Erkelenz, Mueller et al. 2013, Ptok, Müller et al. 2019). By selecting different combinations of exons and splice sites, genes can produce numerous mRNA isoforms (Wang, Sandberg et al. 2008).

1.2.2.1 Bioinformatic tools

As mentioned previously, initiation of the splicing reaction is dependent on splice site recognition, which is dependent on the intrinsic strength of the splice site as well as the presence of cellular splicing factors. These *trans*-acting splicing regulatory proteins (SRPs) recognize and bind *cis*-acting splicing regulatory elements (SREs) found on the transcript. Together, these elements are referred to as the splicing code (Wang and Burge 2008, Barash, Calarco et al. 2010). There are two major families of *trans*-acting SRPs, the serine and arginine-rich splicing factor (SR) proteins and the heterogeneous nuclear ribonuclear protein (hnRNP) family (Dreyfuss, Matunis et al. 1993, Manley and Krainer 2010). Proteins from these families can act as either splicing enhancers or silencers depending on their relative position towards the specific splice site, depicted in Figure 1.2.5.

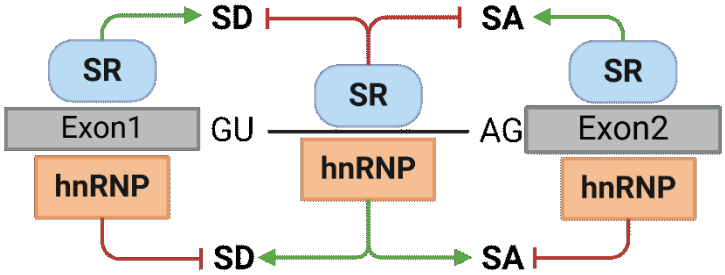


Figure 1.2.5. The position dependent effect of SR and hnRNPs on splice site usage. SR proteins promote downstream splice donor usage and upstream splice acceptor usage while inhibiting upstream splice donor usage and downstream splice acceptor usage. The opposite is true for hnRNP proteins. Grey boxes indicate exons, and the black line represents the intron. SR proteins are shown in blue and the hnRNP protein is shown in orange.

There are several different bioinformatic tools that were developed to calculate the strength of splice sites as well as identify potential SREs in the vicinity of the splice sites. Three of these tools are described below. As previously discussed, the intrinsic strength of the splice donor is dependent on its complementarity to the single stranded 5' end of the U1 snRNA. The HBond score (HBS) calculates the intrinsic splice donor strength based on the predicted hydrogen bond formation between the U1 snRNA and the 11-nucleotide long splice donor (CAG/GURAGUNN (R = A or G, N = A, U, C, or G, "/" = exon-intron border)). The website for HBond score calculation can be accessed through the following

link: https://rna.hhu.de/HBond/ (Kammler, Leurs et al. 2001, Freund, Asang et al. 2003).

The intrinsic strength of the splice acceptor is more complex. It is dependent on the pyrimidine content of the PPT, the distance between the BPS and the splice acceptor, and the complementarity between the BPS and the U2 snRNA (Wahl, Will et al. 2009). The 3' maximum entropy score (MaxEntScore) calculates the intrinsic splice acceptor strength from 23-nucleotide long sequences of the 3' intron exon border starting at the -20 intronic position to the invariant AG splice acceptor nucleotides (up to the +3 downstream exonic position). The 3'MaxEntScore website can be accessed through the following link: http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html (Yeo and Burge 2004).

The HEXplorer algorithm is a valuable tool for analyzing SRE motifs in a sequence of interest, as mutations within SREs can cause aberrant splicing, nuclear export, frameshift mutations, and pre-mature translation termination. The HEXplorer score is based on hexamer weights calculated using a relative Enhancer and Silencer Classification by Unanimous Enrichment (RESCUE) type approach (Fairbrother, Yeh et al. 2002) from a dataset of 43,464 constitutively spliced canonical annotated human exons. Hexamer distributions near splice sites were calculated using 100 nucleotides up and downstream of splice donors (10,407 "strong" splice donors, HBS ≥ 17.0, and 10,359 "weak" splice donors, HBS ≤ 13.5) excluding the 11-nucleotide splice donor. Differential hexamer frequencies were calculated for intronic and exonic sequences (Z_{EI}). SREs enriched around weak splice sites were thought to play a larger role in splice donor recognition than those surrounding stronger splice donors. Therefore, hexamers surrounding these weak splice sites contain more splicing enhancers than the hexamers surrounding strong splice sites. Using the position dependent Z_{EI} score, an average score was calculated for each nucleotide from the overlapping hexamers (HZ_{EI}) (Erkelenz, Theiss et al. 2014). The plot created from the HZ_{EI} scores illustrates potential splicing enhancing and splicing silencing sequences, where a positive HEXplorer score predicts a likelihood of SRSF protein binding, and a negatively scored sequence predicts a likelihood of hnRNP protein binding. Additionally, the HZ_{EI} score can be used to monitor the impact of mutations on SREs by calculating the difference between a wild-type sequence and the mutant sequence (ΔHZ_{EI}). The HEXplorer algorithm can be accessed through the following link: https://rna.hhu.de/HEXplorer/ (Erkelenz, Theiss et al. 2014).

1.2.2.2 Splicing in HIV-1

The 43S ribosomal subunit will start scanning from the 5' cap on HIV-1 mRNAs, and initiate translation at the first efficient start codon. To produce multiple proteins from a single transcript, the HIV-1 mRNA must be extensively spliced to bring start codons of downstream open reading frames in close proximity to the 5'CAP structure (Schwartz, Felber et al. 1990, Purcell and Martin 1993). Shown in Figure 1.2.2., HIV-1 uses five splice

donors (SD1, SD2, SD2b, SD3, SD4) and eight splice acceptors (SA1, SA2, SA3, SA4c, SA4a, SA4b, SA5, SA7) to produce over 50 viral mRNA isoforms, which results in the balanced expression of viral proteins (Stoltzfus 2009, Ocwieja, Sherrill-Mix et al. 2012, Sertznig, Hillebrand et al. 2018).

As previously discussed, splice site selection depends on both the intrinsic strength of a splice site and the surrounding splicing regulatory elements. Most HIV-1 splice sites are intrinsically weak, resulting in inefficient splicing that is primarily determined by surrounding SREs. After splicing, HIV-1 transcripts can be grouped into 3 distinct classes based on their size: 2kb, 4kb, and 9kb. The 2kb class is intronless and encodes the regulatory proteins Rev and Tat, as well as the accessory protein Nef. The intron-containing 4kb class mRNAs encode the Env glycoprotein, Vpu, Vif, and Vpr. The unspliced 9kb class functions as genomic RNA and encodes the Gag and Gag-Pol polyproteins (Purcell and Martin 1993). Expression of these three classes is temporally regulated, with the 2kb class being expressed during the early phase of HIV-1 infection, the 4kb class is the most abundant during the intermediate phase, and the 9kb class expressed primarily during the late phase (Kim, Byrn et al. 1989).

As previously stated, transcripts from the 2kb class typically encode early phase genes: Tat, Rev, and Nef. These transcripts are intronless and characterized by the lack of a Gag-Pol reading frame as well as the loss of a large portion of the Env open reading frame. To achieve this, 2 kb transcripts are generally spliced from SD1 to the SA cluster before SD4 to remove Gag-Pol, as well as from SD4 to SA7 to remove Env (Purcell and Martin 1993, Chang, Sova et al. 2011). Removal of the Env open reading frame has been shown to be dependent on the initial removal of Gag-Pol (Bohne, Wodrich et al. 2005).

Full length *tat* transcripts contain two exons and are generated by splicing SD1 to SA3 and SD4 to SA7. Alternative splicing events also result in Tat-expressing transcripts that vary in the presence of leader exons 2 and 3, as well as a *tat* isoform that only includes the first tat exon. This isoform is found in the 4kb class of HIV-1 mRNAs and produces a truncated Tat protein (Purcell and Martin 1993). Rev, the other regulatory protein produced in the 2 kb class, is encoded by at least 12 alternatively spliced variants. These variants are generated using SA4a, SA4b, and SA4b, located upstream of the Rev start codon. In addition to the two regulatory proteins, the accessory protein Nef is produced in the 2kb class. The *nef* ORF is encoded at the 3'end of the HIV-1 genome and is generated by the inclusion of exon 5. SA5 is located in the SA cluster upstream of SD4. Approximately 80% of all HIV-1 splice acceptor events occur at SA5. Like Tat, Rev and Nef may also have varying leader exons (Schwartz, Felber et al. 1990, Stoltzfus 2009).

Like the 2kb class of HIV-1 mRNAs, the 4kb class are spliced from SD1 to the SA cluster upstream of SD4 and thus lack the Gag-Pol ORF. However, they are characterized by the retention of intron 4 (SD4 to SA7), giving them the moniker "intron-containing" mRNAs.

With intron 4, these transcripts also retain the Rev Responsive Element (RRE) and are dependent on the transactivating protein Rev for export from the nucleus. These mRNAs encode the envelope protein (Env), as well as several accessory proteins; Vif, Vpr, and Vpu (Malim, Hauber et al. 1989, Purcell and Martin 1993, Bohne, Wodrich et al. 2005).

The Env and Vpu ORFs are encoded together on a bicistronic transcript that has 16 different spliced isoforms. These isoforms vary with leader exons 2 and 3, exon 4cab and exon 5 (Purcell and Martin 1993). Env expression has been found to be dependent on the presence of the GAR sequence and the integrity of the 5' ss. The GAR sequence is a short purine-rich bidirectional splicing regulatory element found upstream of SD4 that binds SRSF1 and is necessary for the activation of the SA cluster upstream of SD4 as well as *env* and *vpu* expression. Mutations in this region effectively reduce *env* expression by limiting SRSF1 interaction (Kammler, Leurs et al. 2001, Caputi, Freund et al. 2004, Asang, Hauber et al. 2008).

As one of the least common HIV-1 transcripts at 1% of the total 4kb class, Vif expression is tightly controlled because Vif inhibits viral protein processing and replications when highly expressed. The *vif* mRNA is singly spliced from SD1 to SA1, which produces a 23 kDa protein. For this to occur, SA1 must be recognized while SD2 is repressed, a process regulated by several SREs (Purcell and Martin 1993, Akari, Fujita et al. 2004, Mandal, Exline et al. 2009). Lastly, *vpr* transcripts are spliced from SD1 to SA2, with only one spliced isoform that varies with exon 2. Just like *vpu*/*env* and *vif*, the *vpr* start codon is located in a downstream intron, and SD3 must be silenced for the *vpu* mRNA to be generated (Kammler, Otte et al. 2006, Exline, Feng et al. 2008, Brillen, Walotka et al. 2017). One SRE, a G run in exon 3, controls exon selection between Vpu and Vif leader exons and disruption of it leads to an overexpression of Vpr and a subsequent reduction of Vif (Widera, Erkelenz et al. 2013, Widera, Hillebrand et al. 2014).

The majority of HIV-1 pre-mRNA transcripts remain unspliced, resulting in the 9 kb class of HIV-1, which functions as the viral genome. Additionally, this transcript encodes the open reading frame for the *gag/pro/pol* genes which produce several polyproteins that generate viral enzymes and structural proteins. Like the 4kb class, these transcripts retain the RRE in intron 4 and are dependent on Rev for nuclear export (Malim, Hauber et al. 1989, Emery and Swanstrom 2021). There are several potential methods by which HIV-1 regulates intron retention to produce the unspliced 9kb genome. One theory suggests that there are multiple configurations of the highly structured regions in the 5'UTR, some which favor SD1 usage, and others that block it. Once splicing at SD1 has been blocked, all downstream splicing is eliminated, resulting in genomic RNA (Mueller, van Bel et al. 2014, Brigham, Kitzrow et al. 2019, Esquiaqui, Kharytonchyk et al. 2020, Emery and Swanstrom 2021). Both spliced and intron-containing HIV-1 transcripts make their way to the cytoplasm using distinct host cell nuclear export pathways.

1.2.3 Nuclear Export

Most of the genetic information in eukaryotic cells is stored as DNA in the nucleus, surrounded by the nuclear membrane and movement across this membrane is tightly controlled. The newly synthesized and spliced mRNAs must make their way out of the nucleus to be accessible to the cytoplasmic ribosomes, which will translate them into protein. Nuclear export is mediated by several different pathways; however, we will focus primarily on the NXF1 and CRM1 nuclear export pathways. The mRNAs are packaged into mRNPs in the nucleus, the composition of which varies depending on the export pathway. They make their way through the nuclear pore complexes which pepper the nuclear membrane and act as gateways between the nucleus and cytoplasm. Once in the cytoplasm, the mRNA transcripts can recruit ribosomal subunits and begin the translation process.

1.2.3.1 The Nuclear Pore Complex

The nuclear pore complex (NPC) is the gatekeeper of the nucleus, and anything that crosses the nuclear membrane must meet its very selective requirements. Composed of approximately 1000 protein subunits, the NPC is around 110 MDa, making it one of the largest macromolecule assemblies found in cells. Each NPC guards a central channel that is approximately 425 Å in diameter (Watson 1959), and can be divided into three main sections: the nuclear basket, the central pore, and the cytoplasmic filaments (Lin and Hoelz 2019).

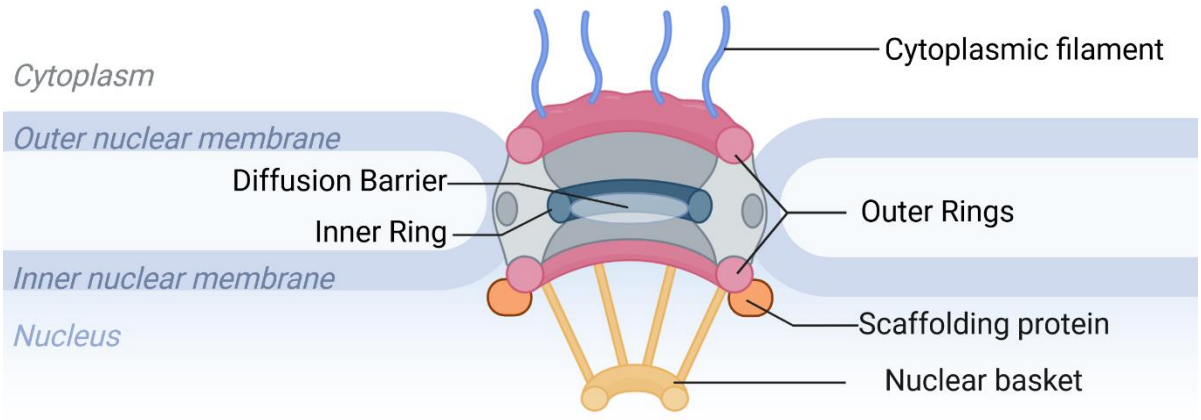


Figure 1.2.6. Schematic of the nuclear pore complex (NPC). A cutaway view depicting half of an NPC is shown, depicting the structure and orientation of the nuclear basket, inner pore, and cytoplasmic filaments. Image created with Biorender. Adapted from (Lin and Hoelz, 2019).

The proteins that make up the NPC are called nucleoporins, which are highly conserved in eukaryotes (Neumann, Lundin et al. 2010). The size of these nucleoporins ranges from 30-358 kDa, and form large, stable subcomplexes that contribute to the eight-fold

modular architecture of the NPC. While there is no formalized naming system, most nucleoporins are referred to by their molecular weight, e.g. Nup88 (Lutzmann, Kunze et al. 2002, Lin and Hoelz 2019).

The coat nuclear complex (CNC), also referred to as the Y-complex, is the best structurally characterized subcomplex in the NPC. Found on the outer ring of the central pore on both sides of the membrane, it acts as a structural scaffold for other nucleoporins. In humans, the CNC is composed of 10 nucleoporins that form an elongated Y shape, thus the term Y-complex (Lutzmann, Kunze et al. 2002, Stuwe, Correia et al. 2015, Lin and Hoelz 2019). Compared to the CNC, the inner ring subcomplexes are more heterogeneous, with each nucleoporin binding several other nucleoporins to create an intricate network (Kosinski, Mosalaganti et al. 2016). The heart of the central pore is a passive diffusion barrier primarily composed of intrinsically disordered sequences enriched in phenylalanine and glycine repeats (Denning, Patel et al. 2003). While macromolecules less than 40 kDa can generally cross this barrier passively, larger macromolecules tend to require assistance from specialized nuclear transport factors (Timney, Raveh et al. 2016). This passive barrier is not necessarily a size exclusion barrier, as a higher concentration of hydrophobic surface residues will allow proteins larger than 40 kDa to also passively diffuse across the barrier (Naim, Zbaida et al. 2009).

Molecules attempting to exit the nucleus will first encounter a basket-like structure on the nuclear side of the NPC (Jarnik and Aebi 1991). Composed primarily of three unique nucleoporins, they attach to the central pore by interacting with components of the CNC (Kim, Fernandez-Martinez et al. 2018). The nuclear basket serves as a platform for mRNP remodeling before export to the cytoplasm (Saroufim, Bensidoun et al. 2015), as well as interacts with nucleocytoplasmic transport machinery and even transcriptional regulation machinery (Köhler and Hurt 2010, Kim, Fernandez-Martinez et al. 2018).

The cytoplasmic filaments interact with the CNC to attach to the nuclear pore in a manner similar to the nuclear basket subcomplexes. Composed of seven nucleoporins, these long flexible subcomplexes extend into the cytoplasm and have a complex biochemical structure with structured interaction domains connected by long flexible linkers (Jarnik and Aebi 1991). These filaments contain several binding sites for nucleocytoplasmic transport factors as well as regulating the activity of the DEAD-box helicase DDX19 which terminates the nuclear export process (Napetschnig, Kassube et al. 2009, Kim, Fernandez-Martinez et al. 2018).

1.2.3.2 The NXF1 Pathway

A group of highly conserved nuclear export factors (NXFs) are responsible for a majority of nuclear export. In humans, these factors are labeled NXF1-6, with NXF1 being the best characterized. NXF1 (also called TAP) is involved in the export of approximately 75%

mRNA transcripts in a genome wide analysis, including both spliced and intronless transcripts (Herold, Teixeira et al. 2003). In the N-terminal domain, NXF1 contains an RNA binding domain (RBD), a leucine-rich repeat, and a nuclear localization signal. The leucine-rich region is particularly important for export of mRNAs containing a viral constitutive transport element (CTE) and the nuclear localization signal interacts with import receptor Transportin to facilitate NXF1's entry into the nucleus (Grüter, Tabernero et al. 1998). The C-terminal domain contains a nuclear transport factor 2 (NTF2) -like domain and a ubiquitin-associated (UBA) domain, both of which participate in nucleoporin binding. Additionally, the NTF2-like domain facilitates NXF1 heterodimerization with export cofactor p15 (Suyama, Doerks et al. 2000, Erkmann and Kutay 2004).

As previously discussed, many viruses export their mRNA transcripts from the nucleus by recruiting NXF1 directly with a secondary structure called the CTE. The double loop structure interacts with the leucine-rich domain and RBD of NXF1, leading to export to the cytoplasm (Grüter, Tabernero et al. 1998). Other viruses unable to directly recruit NXF1 will often do so through protein-protein interactions. This is also the main way that processed mRNAs will recruit NXF1 for export to the cytoplasm. There are several NXF1/RNA interaction mediators, such as members of the REF family and other subcomponents of the EJC (previously discussed in 1.1.3.1) (Stutz, Bachi et al. 2000, Le Hir, Gatfield et al. 2001). Among these NXF1/RNA interaction mediators are several members of the SR protein family, which interact with the N-terminal region of NXF1 in a manner similar to the REF family (Huang, Gattoni et al. 2003). In addition to interacting with mRNA through chaperones, NXF1 can bind mRNA directly with the N-Terminal arginine-rich RBD. This RBD is typically in a closed conformation but will become exposed when NXF1 binds to the TREX complex (Viphakone, Hautbergue et al. 2012).

NXF1-containing mRNPs are directed to the nuclear pore by both chaperones and NXF1 itself. One of the best studied chaperones is the TRanscription and EXport complex-2 (TREX2), a multi-subunit complex that links transcription to nuclear export. TREX-2 subunit GANP binds to the C-terminus of NXF1 and directs it to the nuclear basket (Wickramasinghe, McMurtrie et al. 2010). However, depletion of GANP only decreased export of 50% of known NXF1 export targets, implying that only specific "fast-track" mRNA classes are GANP dependent (Wickramasinghe, Andrews et al. 2014). Other studies have shown that NXF1 is capable of binding to several different nucleoporins using a nucleoporin binding domain at the C-terminus to mediate non- "fast-track" mRNA export (Bachi, Braun et al. 2000). However, docking at the NPC does not guarantee export to the cytoplasm, with some studies showing only a 25-36% success rate (Siebrasse, Kaminski et al. 2012, Ma, Liu et al. 2013, Kelich and Yang 2014).

NXF1 requires the cofactor NXT1 to translocate to the cytoplasm. The NXF1/NXT1 heterodimer moves its cargo through the NPC by direct interaction with the phenylalanine-glycine repeats at the heart of the central pore (Matzat, Berberoglu et al. 2008). Once the mRNP exits the central pore, it associates with the cytoplasmic filaments. The mRNA cargo is released into the cytoplasm, and the export factors are recycled back to the nucleus. Release of the mRNA is mediated by DEAD box helicase DDX19 in an ATP-dependent manner in response to binding of signaling molecule inositol hexakisphosphate (Alcázar-Román, Tran et al. 2006). The cytoplasmic filament component RanBP2 binds several export factors, including NXF1, to facilitate nuclear reimportation (Forler, Rabut et al. 2004). Translocation through the nuclear pore via NXF1 is rapid, commonly occurring in less than 200 ms, including docking and release (Grünwald and Singer 2010, Delaleau and Borden 2015).

1.2.3.3 The CRM1 Pathway

Originally discovered due to its association with a chromosome structure distortion, the chromosome maintenance protein 1 (CRM1, also known as exportin 1) was later found to be the most productive member of the Karyopherin family β of nuclear transporters (Adachi and Yanagida 1989, Fornerod, Ohno et al. 1997). This family of proteins has a low sequence identity, but shares several physical properties, including their dependence on Ran-GTP for interaction with their cargo (Xu, Farmer et al. 2010). As one of the larger Karyopherin exportins at 120 kDa, CRM1 contains 21 anti-parallel α-helix repeats (HEAT repeats) that form a ring shape. The concave inner surface of the ring hosts Ran-GTP while a hydrophobic groove is formed by HEAT repeats 11 and 12 on the outer surface of the ring (Dong, Biswas et al. 2009). CRM1 primarily exports proteins, but also exports a small subset of mRNAs (Fung and Chook 2014). CRM1 recognizes leucine-rich nuclear export signals (NES) found in its protein cargo. These NES are typically 8-13 amino acids and contain hydrophobic residue patterns, which interact with the hydrophobic groove on CRM1 (Xu, Farmer et al. 2012). Export of mRNAs by CRM1 involves an RBP intermediate that contains a NES signal. One of the first NES signals was discovered in the HIV-1 protein Rev, which mediates export of intron-containing viral mRNAs by binding a *cis*-acting mRNA secondary structure and interacting with CRM1 (Fischer, Huber et al. 1995, Fung and Chook 2014). In host genes, CRM1 can be recruited to the previously described ARE sequences (1.1.2.1.1) by the *trans*-acting protein HuR. HuR binds ARE sequences and interacts with accessory proteins that promote CRM1-mediated export through leucine-rich NES (Brennan, Gallouzi et al. 2000, Wu, Tong et al. 2019). The cap-binding protein eIF4E is also able to engage in CRM1 export through the leucine-rich NES of accessory protein LRPPRC (Topisirovic, Siddiqui et al. 2009, Scott, Aguilar et al. 2019). However, CRM1 is unable to bind mRNA and can only export RNA in the presence of an RBP adapter.

Like other Karyopherins, CRM1 activity is Ran-GTP dependent, and CRM1-mediated nuclear export is controlled by the Ran-GTP/Ran-GDP gradient across the nuclear membrane (Askjaer, Jensen et al. 1998). Ran is typically in a GTP state in the nucleus and a GDP state in the cytoplasm. This gradient across the nuclear membrane is maintained by compartmentalized Ran regulators. In the nucleus, chromatin tethered RCC1 maintains a GTP state through efficient guanidine nucleotide exchange. In the cytoplasm, RanGAP1 catalyzes the hydrolysis of GTP to GDP at the cytoplasmic filaments of the NPC (Izaurralde, Kutay et al. 1997). Once the RanGTP-CRM1-Cargo complex has formed, it is able to interact with the nuclear basket and move through the central pore. Once through the central pore, the complex interacts with the cytoplasmic filaments, particularly Nup358. Also known as Ran Binding Protein 2 (RanBP2), Nup358 hydrolyzes RanGTP into RanGDP with the assistance of RanBP1 or RanGAP. The CRM1-RanGDP complex has a much lower NES affinity than CRM1-RanGTP, leading to the release of the NES-containing cargo into the cytoplasm (Koyama and Matsuura 2010). Once the cargo is released, RanGDP is ferried back into the nucleus by nuclear transport factor 2 (NTF2) and is regenerated back into RanGTP (Fung and Chook 2014, Scott, Aguilar et al. 2019, Borden 2020).

1.2.3.4 Nuclear Export of HIV-1 transcripts

Enrichment of the individual classes on HIV-1 mRNAs is temporally regulated, and the nuclear export pathways differ between early and late phase transcripts (Kim, Byrn et al. 1989). Nuclear export plays an important role in maintaining the delineation between late and early phase gene expression as well as contributing to HIV-1 mRNA stability. Intron-containing HIV-1 transcripts are degraded in the nucleus, thus when transcripts are exported to the cytoplasm, they are more stable (Felber, Hadzopoulou-Cladaras et al. 1989). As previously discussed in 1.2.4.2, most eukaryotic mRNAs are exported via the NXF1 Pathway. In the context of HIV-1, however, only the multiply spliced 2 kb class transcripts are exported via the NXF1 pathway. These transcripts no longer contain introns and recruit NXF1 using Exon Junction Complexes and other mediators. Among the proteins encoded in the 2kb class is the transactivator protein Rev, which is vital for the export of intron containing late phase genes (Köhler and Hurt 2007, Emery and Swanstrom 2021).

Late phase genes include the 4 kb and 9 kb class of mRNAs, characterized by their retention of intron 4, which contains the Rev Responsive Element. The RRE facilitates nuclear export of these intron-retaining transcripts when they would otherwise be retained in the nucleus. After translation, Rev is imported into the nucleus where it binds the RRE at stem loop 2 (SL2) (Karn, Dingwall et al. 1991). This results in the multimerization of up to six Rev at the RRE and the recruitment of PIMT, a hypermethylation enzyme (Malim, Hauber et al. 1989, Daugherty, D'Orso et al. 2008). PIMT modifies the 7-methylguanosine (m7G) cap to a trimethylguanosine (TMG) cap to

help recruit the nuclear export factor CRM1 to the transcript (Yedavalli and Jeang 2010). As discussed in 1.1.4.3, CRM1 is a member of the karyopherin family that recognizes leucine-rich nuclear export signals (NES). Once recruited by the TMG cap, CRM1 binds to the leucine-rich NES at the C-terminal end of Rev. CRM1 uses this interaction to transport the intron containing HIV-1 transcripts to the cytoplasm, where they are translated or packaged into budding virions (Fischer, Huber et al. 1995, Fischer, Pollard et al. 1999).

1.2.3.5 Inhibition

In addition to being dysregulated in cancer, nuclear export is also altered and often hijacked during viral infections (Turner, Dawson et al. 2012, Guo, Zhu et al. 2023). Discussed below are several small molecules and proteins that have been found to inhibit nuclear export in a pathway dependent manner.

Small molecules are often used to regulate cellular processes in many different diseases including cancer and immune disorders. Due to this feature, drug discovery has become an important field to combat drug resistances in infections and cancer as well as discover new treatments for well-known diseases and disorders.

Originally discovered as an antifungal antibiotic isolated from a strain of Streptomyces, Leptomycin B (LB) inhibits nuclear export of Rev and Rev-dependent mRNA, but not nuclear import (Hamamoto, Gunji et al. 1983, Wolff, Sanglier et al. 1997). Further investigation found that LB does not specifically inhibit Rev-mediated export, but rather its cellular effector, CRM1 (Fornerod, Ohno et al. 1997). Leptomycin B binds covalently to cystine-528 in the NES-binding groove of CRM1 and prevents CRM1 from interacting with its cargo (Kudo, Matsumori et al. 1999). CRM1 and LB binding becomes irreversible after lactone hydrolysis, leading to LB's high potency as an inhibitor of CRM1 (Sun, Carrasco et al. 2013).

Although LB is a potent inhibitor of CRM1 mediated nuclear export and had great potential as a cancer therapy, it showed significant toxicity during clinical trials (Newlands, Rustin et al. 1996). To harness the cancer-fighting benefits of CRM1, several selective inhibitors of nuclear export (SINEs) were developed. Among those, KPT-330 showed an inhibition of CRM1 binding as well as reduced tumor proliferation in mice. KPT-330 also interacts with cystine-528 to prevent CRM1 interaction with its cargo, but in a non-covalent manner (Azmi, Aboukameel et al. 2013). KPT-330 has been approved for treatment of cancer under the name Selinexor (Landes, Moore et al. 2023).

In addition to chemical inhibition, DNA and some RNA viruses that replicate in the nucleus must use the host cell nuclear export machinery to export their transcripts to the cytoplasm. In addition to hijacking the host's nuclear export pathways, many viruses prevent immune responses by inhibiting host mRNA nuclear export (Zhang, Xie et al. 2019,

Yang, Chang et al. 2022, Guo, Zhu et al. 2023). In addition to facilitating CRM1-mediated nuclear export, HIV-1's Rev has been shown to inhibit NXF1 recruitment to RRE-containing mRNAs by inhibiting the recruitment of NXF1/TREX to the cap binding structure. Taniguchi et al shows that Rev likely competes with TREX components Aly/REF to interact with CBP80, a component of the cap-binding structure. However, the mechanism of Rev's inhibition of NXF1 export of RRE containing mRNAs still needs elucidation (Taniguchi, Mabuchi et al. 2014).

Influenza A, a negative stranded RNA virus that replicates in the nucleus, also inhibits host mRNA nuclear export by specifically inhibiting the NXF1 pathway (Satterly, Tsai et al. 2007). Influenza A protein non-structural protein 1 (NS1) interacts with the nucleoporin binding domain (NTF2L), to inhibit NXF1/NXT1's ability to escort mRNA through the nuclear pore. By preventing NXF1-mediated nuclear export with NS1, influenza inhibits the expression of many interferons and other immune-regulated genes. Virus mutants with an NS1 that is unable to inhibit NXF1 are significantly attenuated, emphasizing the importance of this viral replication strategy for influenza (Zhang, Xie et al. 2019).

1.2.4 Post-transcriptional Regulation

Although a great deal of gene expression regulation occurs at transcription, post-transcriptional control is vital to cell and organism health. Post-transcriptional control is mediated by two major components: *cis*-acting regulatory elements (CRS) and the *trans*-acting factors that interact with them. *Cis*-acting regulatory elements are sequences found throughout a transcript, including the untranslated regions and coding sequences. *Trans*-acting factors are RNA-binding proteins (RBPs) and microRNAs that interact with these *cis*-acting sequences to determine the fate of the mRNAs they bind. The interplay between *trans*-acting factors and a given mRNA allows for transcript specific regulation necessary for vital cell decisions (Vlasova-St.Louis and Sagarsky 2018).

1.2.4.1 Cis-acting Regulatory Elements

Conserved sequences within a transcript regulate most aspects of its metabolism, from conformation to the post-transcriptional fate of the protein it produces. Since the success of the human genome project (and even before), the list of conserved *cis*-acting sequences continues to grow, though we still lack a detailed understanding of the mechanisms by which many of these CRS recruit and arrange *trans*-acting factors. There are several common and relevant elements detailed below.

Two elements were identified in the 3'UTR of mRNA transcripts that correlated with increased stability and mRNA turnover: AU-rich elements (AREs) and GU-rich elements (GREs). AREs are composed of a repeating pentamer, typically AUUUA, however there are several AREs with similar functions but slightly different sequences (Peng, Chen et al.

1996). These elements have since been identified in both the untranslated regions and protein coding regions of cytokine, proto-oncogene, and transcription factor mRNAs (Chen and Shyu 1995, Chen, Xu et al. 1995). The second sequence, GREs, occur in approximately 8% of human transcripts and are considered essential regulators of splicing, stability, and translation (Halees, Hitti et al. 2011). These sequences have since been found to interact with a wide range of *trans*-acting RNA binding proteins, which are then able to control the fate of the ARE and GRE harboring transcripts, from splicing to cytoplasmic decay. Additionally, the fate of these transcripts changes during times of cellular stress due to the wide range of RBPs they can bind (Vlasova-St.Louis and Sagarsky 2018).

Several studies have found that short position-specific motifs (4-6 nucleotides in length) were able to predict RNA isoform abundance by mediating pre-mRNA/RBP interactions (Hui, Hung et al. 2005, Cereda, Pozzoli et al. 2014). These sequences serve as either splicing enhancers or silencers and are conventionally classified as exonic splicing enhancers (ESE), exonic splicing silencers (ESS), intronic splicing enhancers (ISE), or intronic splicing silencers (ISS). However, recent research has revealed that the function of the same sequence can change according to the SRE position relative to the splice site, thus these terms can be misleading. These elements are typically bound by Serine and Arginine-rich family proteins (SR) and heterogenous nuclear ribonucleoproteins (hnRNPs), both of which exert both enhancing and silencing effects depending on their relationship to the splice site in question (Erkelenz, Mueller et al. 2013). Additionally, these elements function additively and an increase in SREs leads to an increased influence on splice site selection (Huh and Hynes 1994, Wang and Burge 2008).

In addition to binding specific sequences, RBPs can also bind mRNA secondary structures. They can recruit RBPs through interaction with specific shapes created by these secondary structures which then influence gene expression (Olsen, Nelbock et al. 1990). Viruses make particularly good use of these elements to initiate translation and nuclear export. The human immunodeficiency virus 1 (HIV-1) exports its late phase expression transcripts using the Rev responsive element (RRE) to recruit nuclear export factor CRM1 through an intermediary viral RBP (Malim, Hauber et al. 1989, Fornerod, Ohno et al. 1997). On the other hand, the Mason-Pfizer Monkey Virus (MPMV) recruits the nuclear export factor NXF1 using a secondary structure called the constitutive transport element (CTE) (Kang, Bogerd et al. 2000). In addition to facilitating nuclear export, there are several secondary structures that can initiate translation independent of the typical ribosomal recruitment strategies. These elements are called internal ribosomal entry sites (IRES) and are found in viruses such as picornaviruses and enteroviruses as well as several cellular genes (Jang, Kräusslich et al. 1988, Pelletier and Sonenberg 1988, Lopez-Lastra, Rivas et al. 2005).

In the context of HIV-1, subgenomic regions throughout the HIV-1 viral genome have been found to negatively impact nuclear export of late phase transcripts independent of unused splice sites. These regions were labeled instability (INS) elements or cis-acting repressive sequences (CRS) and are highlighted in grey in Figure 1.2.2. Most of these regions are found in intronic sequences, which are retained in sequences that are trapped in the nucleus in the absence of Rev. Although many of these sequences were identified in the HIV-1 genome, no commonly recognizable sequence has been identified and the mechanism is still unclear (Rosen, Terwilliger et al. 1988, Cochrane, Jones et al. 1991, Maldarelli, Martin et al. 1991, Brighty and Rosenberg 1994, Ostermann, Ritchie et al. 2021).

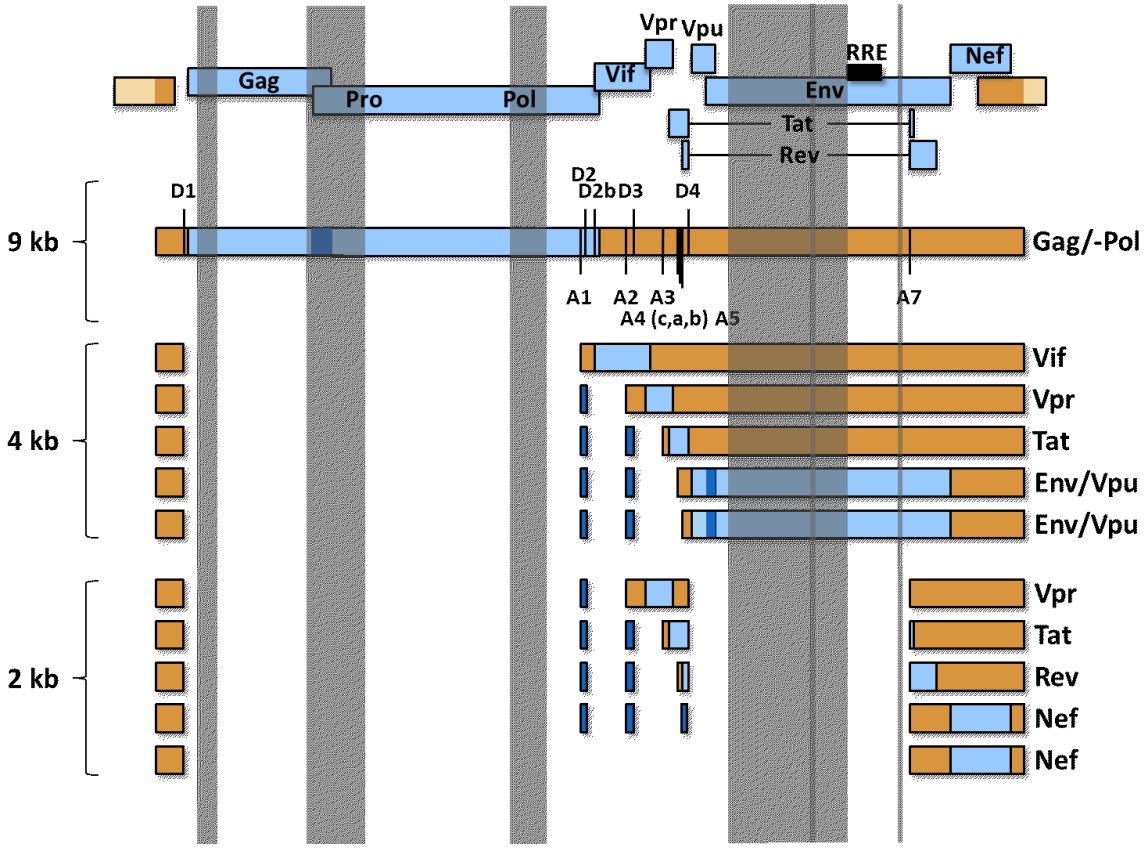


Figure 1.2.6. Schematic of HIV-1 mature transcripts and defined CRS. The HIV-1 provirus with viral ORFs is shown in blue. Untranslated regions are shown in orange and overlapping open reading frames are in dark blue. Gray shaded bands highlight positions of HIV-1 CRS that inhibit cytoplasmic RNA accumulation within the full genome, as well as spliced isoforms. D1 to D4, 5' splice (donor) sites; A1 to A7, 3' splice (acceptor) sites. Adapted from (Ostermann, Ritchie et al. 2021).

In several cases, these CRS were found to interact with host cell RNA-binding proteins including several members of the hnRNP family. Two of these proteins, hnRNPC and hnRNPA1, have previously been shown to play a role in retaining viral transcripts in the nucleus (Black, Luo et al. 1995, Sokolowski and Schwartz 2001). hnRNPC is retained in the nucleus by a nuclear retention signal (NRS) that is capable of overriding nuclear export signals (NES) to retain associated transcripts in the nucleus (Nakielny and Dreyfuss 1996).

While hnRNPC is almost exclusively nuclear, hnRNPA1 shuttles between the nucleus and cytoplasm and has been shown to facilitate nuclear export of cellular transcripts (Monette, Ajamian et al. 2009, Roy, Durie et al. 2014). However, several studies show that hnRNPA1 is also involved in the nuclear retention of transcripts, though the mechanism by which it determines the localization of transcripts is still unknown (Lund, Milev et al. 2012). Since these CRS appear exclusively in structural intron-containing genes in the context of HIV-1, it is possible that they play a role inhibiting their expression until the optimal time point during infection (Ostermann, Ritchie et al. 2021).

1.2.4.2 Trans-acting Factors

C*is*-acting sequences control mRNA fate by binding *trans*-acting factors such as RNA-binding proteins (RBPs) and miRNAs, which affect the downstream gene expression process. All RBPs contain at least one RNA-binding domain, though they often have several. These RNA-binding domains, which tend to be quite small, use a wide variety of methods to interact with RNA. Some of the most common RNA binding domains are helicase domains and RNA recognition motifs (RRM) (Corley, Burns et al. 2020). The helicase domains, such as those found in DEAD- and DEAH-box proteins, use ATP to remodel or bind RNA. They are generally polymers that tend to form interactions with phosphates and sugars, but have also been shown to bind to bases on occasion (Jankowsky 2011). Estimated to occur in almost 1% of all human proteins, RRMs bind between 2-8 single-stranded RNA nucleotides using hydrogen bonds and sequential stacking interactions, and each RRM has its own sequence preference (Cléry, Blatter et al. 2008). There are two major families of RBPs that have been found to regulate several steps of the gene expression process: serine and arginine-rich proteins (SR) and the heterogeneous ribonucleoproteins (hnRNP).

Originally discovered as splicing regulators, SR proteins have since been found to regulate everything from transcription to translation. As the name suggests, they are characterized by the C-terminal serine and arginine-rich (RS) domain, which participates in both protein-protein and protein-RNA interactions. In addition to the C-terminal domain, all SR proteins harbor an N-terminal RRM. Most of the SR proteins are restricted to the nucleus, however several shuttle between the nucleus and the cytoplasm, which is depicted in figure 1.2.3. SR proteins are not functionally equivalent and have different expression and regulatory patterns (Zahler, Neugebauer et al. 1993, Long and Caceres 2009, Manley and Krainer 2010).

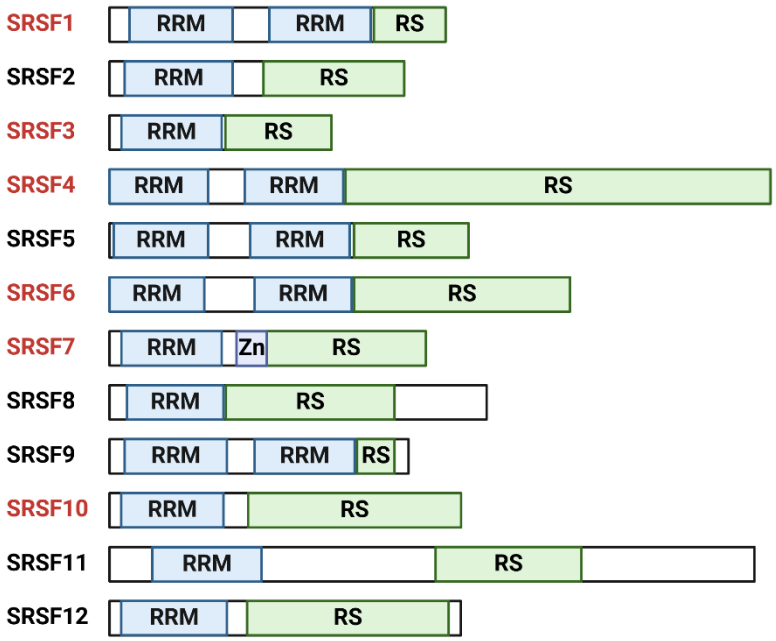


Figure 1.2.7. Schematic of known SR protein family members. The SR family of proteins is currently composed of 12 proteins, all of which contain at least an RNA recognition motif and one Serine-Arginine rich domain. These proteins are typically referred to as SRSFs. The names of the SR proteins that shuttle between the nucleus and cytoplasm are notated in red, while those that do not show shuttling activity are in black. The RNA Recognition Motif (RRM), Serine Arginine Rich Domain (RS) and Zinc Finger Motif (Zn) are notated in the image.

SR protein functions are typically regulated by phosphorylation via SR specific kinases. SR protein activity and the location of the shuttling members of the family is regulated by phosphorylation. In fact, de-phosphorylation and subsequent re-phosphorylation are very important for the function of cytoplasmic SR proteins (Zhou and Fu 2013). Additionally, aberrant expression of SR proteins can lead to a wide variety of diseases including cancer (Jeong 2017, Latorre and Harries 2017, Zheng, Peng et al. 2020, Ayyildiz, Bergonzoni et al. 2023).

Like SR proteins, heterogeneous nuclear ribonucleoproteins (hnRNPs) are a large family of RNA binding proteins that regulate several gene expression processes, and altered expression of these proteins has been linked to several diseases including cancer (Bekenstein and Soreq 2013, Gallardo, Hornbaker et al. 2016). However, as we see in Figure 1.2.4, hnRNP characteristics are not as homogeneous and structured as SR proteins. All hnRNPs have at least one of four RNA binding domains: RRM, quasi-RRM, RGG box, or a KH domain. Additionally, they often have proline-, glycine-, or acid-rich auxiliary domains (Dreyfuss, Matunis et al. 1993, Geuens, Bouhy et al. 2016).

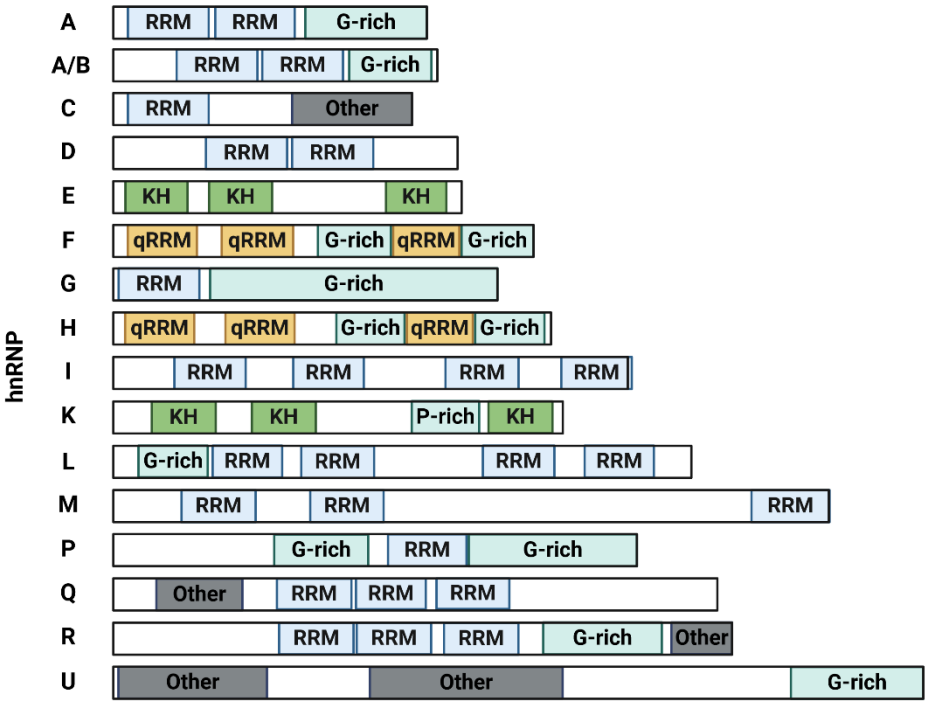


Figure 1.2.4. A schematic of hnRNP family members. The hnRNP are named alphabetically from A1 to U and have a wide range of molecular weights ranging from 34 to 120 kDa. The proteins of the hnRNP family contain at least one of four different RNA-binding domains: RNA recognition motif (RRM), quasi-RNA recognition motif (qRRM), K-homology domain (KH), and an RNA-binding domain consisting of Arg-Gly-Gly repeats (G-rich). This figure was made using Biorender and adapted from (Geuens et al. 2016).

The variety of binding domains and other elements found in the hnRNP family lead to a wide variety of functions and cellular location. Most hnRNP proteins contain a nuclear location signal (NLS) but this can be overcome by post-translational modifications or interactions with other hnRNPs (Kim, Hahm et al. 2000, Chaudhury, Chander et al. 2010). These proteins are commonly modified by phosphorylation, methylation, ubiquitination, and sumoylation, which regulates cellular localization and biological activity (Xu, Wu et al. 2019, Barrera, Ramos et al. 2020).

MicroRNAs are short regulatory sequences that interact with *cis*-acting regulatory sequences and RBPs to influence the gene expression (Ciafrè and Galardi 2013). Some miRNA have been found to occlude RBP recognition sites and regulate RBP-mRNA interaction through either direct competition or remodeling secondary structures to limit CRS exposure (Jens and Rajewsky 2015). Deregulation of miRNAs has recently been associated with many human diseases and disorders (Perbellini, Greco et al. 2011, Vlasova-St.Louis and Sagarsky 2018).

1.2.5 Translation

Once the mature mRNA has been transported from the nucleus to the cytoplasm, translation can occur. Translational regulation plays a critical role in many fundamental biological processes, and modulation of this step has a major impact on gene expression (Hershey, Sonenberg et al. 2019). The eukaryotic translation process takes the message encoded in protein-coding mRNA and matches amino acids to the 3-nucleotide codons using tRNAs. These amino acids are then strung together to form a full protein in a process consisting of four steps: initiation, elongation, termination, and ribosome recycling (Blanchet and Ranjan 2022).

Translation in eukaryotes is initiated on either a cap-dependent or cap-independent basis. In cap-dependent initiation, the initiation factors are recruited to the 5'UTR through interaction with the m7G that was deposited on the 5' end of the mRNA during transcription. Cap-independent translation does not require this cap to recruit translation initiation factors and is commonly used by genes involved in apoptosis and stress induced responses, as well as by viruses (Lopez-Lastra, Rivas et al. 2005).

First, the small 40S ribosomal subunit is pre-loaded with the initiator methionyl tRNA (Met-tRNA$_i$) with the assistance of the GTP-bound form of eukaryotic initiation factor 2 (eIF2), though eIF1, eIF1A, eIF3, and eIF5 are also thought to be involved in this process. This will form the 43S pre-initiation complex (Sokabe, Fraser et al. 2012). Mediated by the m7G cap and eIF4F, this 43S pre-initiation complex will find the 5' end of the mRNA and begin to scan through the 5'UTR until it encounters a suitable translational start sequence, characterized by the presence of the AUG start codon, which is complementary to the Met-tRNA$_i$ anticodon (Kozak and Shatkin 1979, Sherman, McKnight et al. 1980, Kolitz, Takacs et al. 2009). Not every AUG codon is recognized equally, as the surrounding nucleotides influence the intrinsic strength of the start codon, particularly positions -3 and +1. The optimal translational initiation sequence, often called the Kozak sequence in mammals, is 5' GCCR**AUG**G 3' (R = A or G) (Kozak 1984, Kozak 1986). The 43S pre-initiation complex arrests scanning once a suitable AUG start codon is encountered and disassociates several eIFs. After conformational rearrangements, eIF5B will catalyze the binding of the large 60S ribosomal subunit to form the 80S translation initiation complex (Pestova, Lomakin et al. 2000), with the Met-tRNA$_i$ anticodon base paired to the AUG-codon found in the Kozak sequence in the P-site of the ribosome, forming an elongation competent complex (Jackson, Hellen et al. 2010).

Ribosomal scanning of the 5'UTR is not always so straight forward. AUG codons within a bad context can be bypassed, and the 40S subunit will continue to an AUG with a better context further downstream in an event called leaky scanning (Kozak 1986). Leaky scanning is frequently used by viruses, such as HIV-1. HIV-1 envelope (Env) start codon is preceded by a start codon for Vpu, and the weak context for the Vpu AUG allows leaky

scanning to produce Env protein as well as Vpu from the same mRNA (Schwartz, Felber et al. 1992). Another way for a scanning 40S ribosomal subunit to skip an AUG is through ribosomal shunting, when it encounters a large secondary structure and rather than unwinding it, the 40S subunit skips over it and any AUGs that it may contain (Fütterer, Kiss-László et al. 1993, Yueh and Schneider 1996). Lastly, rather than skipping the upstream AUG, the ribosome translates both using the reinitiation mechanism. When the mRNA contains a short upstream open reading frame (uORF), the 40S subunit will not disengage allowing it to find downstream ORFs and continue translating (Morris and Geballe 2000, Lopez-Lastra, Rivas et al. 2005).

An RNA structure that allows the assembly of the translational machinery near or at the start codon was discovered during the investigation of picornaviruses (Jang, Kräusslich et al. 1988, Pelletier and Sonenberg 1988). Later called the Internal Ribosomal Entry Site (IRES, previously discussed in 1.1.2.1.3), they still require specific cellular factors called IRES *trans*-acting factors (ITAFs) to function despite being entirely independent from the 5’ and 3’ ends of the RNA (Chen and Sarnow 1995, Belsham and Sonenberg 2000). IRES-mediated translation initiation is strictly dependent on the tertiary structure integrity that is supported by RNA-protein interactions as well as RNA-RNA interactions (Lafuente, Ramos et al. 2002, Fernández-Miragall and Martínez-Salas 2003, Martínez-Salas and Fernández-Miragall 2004). In addition to viral genes, several cellular genes contain 5’UTR IRES elements, including oncogenes, growth factors, and stress response genes. These IRES mediated genes were found to be actively transcribed even in conditions where cap-dependent translation was shut off, likely to help the cell cope with stress (Akiri, Nahari et al. 1998, Stoneley, Chappell et al. 2000, Fernandez, Yaman et al. 2001, Lopez-Lastra, Rivas et al. 2005).

The 80S ribosomal complex moves over the mRNA from the 5’ end to the 3’ end, one codon at a time (one codon is comprised of three nucleotides), extending the growing protein chain. This process is assisted by multiple eukaryotic elongation factors (eEFs) and aminoacyl-tRNAs (aa-tRNAs). Figure 1.2.10 shows that within the 80S ribosomal complex, there are three codon-reading sites: the aminoacyl site (A-site), the peptidyl site (P-site), and the exit site (E-site). The incoming aa-tRNAs bind to the A-site of the ribosome through aa-tRNA anticodon complementarity with the mRNA codon in that site. This allows a peptide bond to be formed between the amino acids at the A-site and P-site. The used tRNAs are deacylated and the ribosomal machinery translocates, disassociating the previous tRNA from the E-site, moving the unloaded tRNA to the E-site, leaving the A-site empty for the next aa-tRNA. During the elongation process, the ribosome removes proteins bound to the transcript, such as EJC (Dever and Green 2012).

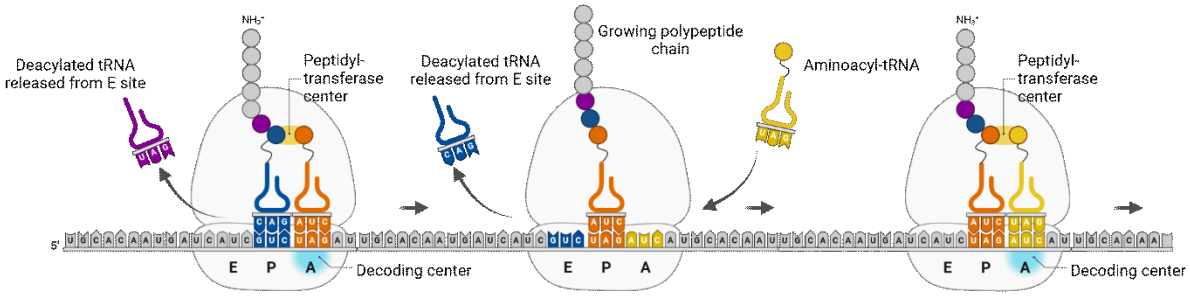


Figure 1.2.10. Schematic of the competent ribosome and elongation cycle. The large ribosomal subunit is drawn transparent to visualize tRNAs and mRNA binding to the decoding center at the interface between the large and small ribosomal subunits the peptidyl transferase center in the large subunit. An aminoacyl-tRNA with a complementary anti-codon is accommodated into the A site, and a peptide bond is formed. Peptide bond formation is accompanied by transition of the A- and P-site tRNAs into the P and E sites, respectively. The deacylated tRNA is released from the E-site and continues with the next cycle of elongation. Image created with Biorender. Adapted from (Dever and Green 2012).

The elongation process advances along the transcript until the ribosome encounters one of three stop-codons (UAA, UAG, UGA), which indicate the end of a designated ORF. These stop codons lack a complementary tRNA anticodon and are instead recognized by the eukaryotic release factor 1 (eRF1). When bound by eRF3 and GTP, it forms a ternary complex in the ribosomal A-site. The GTP is hydrolyzed by eRF3 inducing a conformational change in eRF1 (Frolova, Le Goff et al. 1994). This results in the hydrolysis of the nascent polypeptide from the polypeptidyl-tRNA occupying the P-site (Dever and Green 2012, Schuller and Green 2018). Translation termination is enhanced by cytoplasmic poly-A binding protein 1 (PABPC1) binding to the amino-terminal region of eFR3 (Hoshino, Imai et al. 1999, Ivanov, Mikhailova et al. 2016). Finally, the polypeptide, eFR3, and GDP are released and eRF1 is recruited to the ATP-binding cassette subfamily E member 1 (ABCE1). ABCE1 recycles the 80S ribosome machinery by separating it into the 40S and 60S subunits (Pisarev, Skabkin et al. 2010).

1.2.5.1 Codon Optimization

There are 64 possible codons in the genetic code, each one composed of three nucleotides. However, because there are only 20 amino acids, most amino acids are represented by multiple codons. This phenomenon is called codon redundancy and the figure below can be used to visualize this (Sonneborn 1965). Redundancy is believed to minimize the harmful effects of point mutations as well as maintain the underlying RBP code that regulates processes such as splicing without hindering the amino acid sequence (Cartegni, Chew et al. 2002, Kimchi-Sarfaty, Oh et al. 2007). Despite the synonymous nature of these redundant codons, they are not used uniformly and there is an organism-dependent bias to some codons over others. The degree of codon usage bias (CUB) correlates directly with gene expression (Grantham, Gautier et al. 1980, Sabi and Tuller 2014).

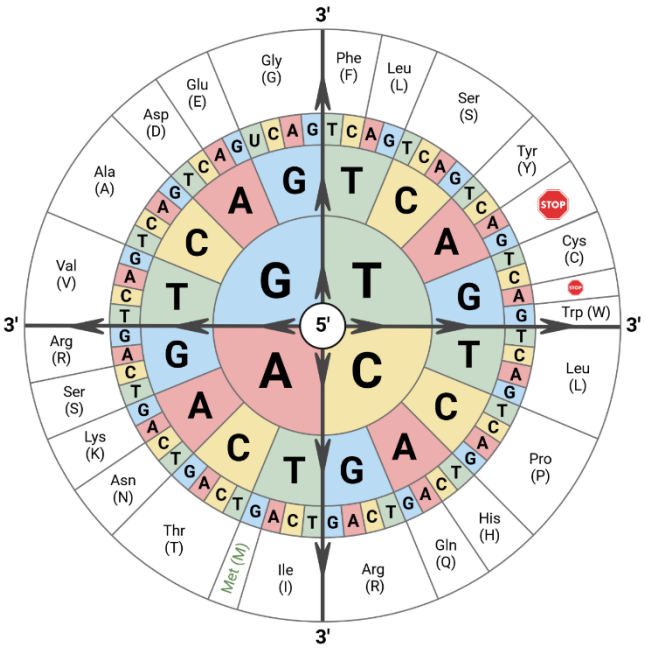


Figure 1.2.11. The codon redundancy wheel. The codon sequence is read from the center of the wheel outwards and used to translate codons (three nucleotide long codes) to the amino acid they code for. Most amino acids are encoded by several different codons. Image created in Biorender.

The abundance of tRNAs has similarly evolved to complement codon bias, with more frequently used codons and complementary iso-tRNA (tRNAs that carry the same amino acid with different anticodons) enriched in highly expressed genes (Ikemura 1985). With the rise of gene therapy technologies, many companies and researchers optimize the synthetic gene sequence to contain codons complementary to the most abundant iso-tRNAs to increase protein production (Ill and Chiou 2005, Davies and Flower 2007). However, synonymous codon exchanges can have surprising and unanticipated effects, such as altered protein conformation, stability and even function that can result in disease (Tsai, Sauna et al. 2008, Sauna and Kimchi-Sarfaty 2011, Agashe, Martinez-Gomez et al. 2013, Mauro and Chappell 2014).

1.2.5.2 Translation and Post-translational Modification in HIV-1

Transcripts from the HIV-1 2kb class only have one open reading frame (ORF) and are translated similarly to host cell genes. Alternatively, HIV-1 also produces a bicistronic message, as well as polyproteins. Translation of the bicistronic *env/vpu* transcript occurs at the rough endoplasmic reticulum (RER) and uses a leaky scanning mechanism to produce both proteins from a single transcript. The *vpu* start codon found upstream of the *env* start codon is poorly recognized due to a poor Kozak sequence. This allows the ribosomal pre-initiation complex to continue scanning and initiate translation at the downstream *env* start codon (Krummheuer, Johnson et al. 2007). The newly synthesized Env precursor gp160 partially translocates to the RER lumen using an N-terminal signal-sequence (Sundquist and Kräusslich 2012). Full translocation to the RER lumen is

prevented by a hydrophobic sequence in the transmembrane domain (TMD), leaving the cytoplasmic tail of Env in the cytoplasm, and the extracellular domain in the RER lumen (Checkley, Luttge et al. 2011). During translation, the N-terminal signal sequence is removed while oligosaccharide chains are attached to the extracellular domain in the RER lumen using N-glycosidic bonds. The individual gp160 precursor proteins will then trimerize and translocate to the Golgi complex. Once at the Golgi complex, cellular furin or furin-like proteases cleave the gp160 precursor into the surface glycoprotein (gp120) and transmembrane protein (gp41) (Hallenberger, Bosch et al. 1992, Checkley, Luttge et al. 2011).

In addition to acting as the genome, the 9 kb mRNA class produces two precursor proteins: the 55 kDa Gag precursor, and the 160 kDa Gag-Pol polypeptide. The 55 kDa Gag precursor is produced by traditional translation initiation and elongation and ends at a stop codon found in intron 1. This precursor protein contains the matrix (MA), capsid (CA), nucleocapsid (NC), p6 domains, and two spacer peptides, SP1 and SP2. These proteins are highly expressed and serve as structural building blocks for mature HIV-1 virions (Freed 2015).

To synthesize the 160 kDa polypeptide from the same 9 kB transcript, the 55 kDa stop codon needs to be ignored. HIV-1 achieves this using a “slippery sequence”. The slippery sequence consists of six uridines and an adenine (UUUUUUA) followed by a hairpin structure that is approximately 200 nucleotides upstream of the Gag stop codon that produces the 55 kDa precursor protein. These features cause the translocating ribosome to pause and induces a -1 ribosomal frameshift in 5-10% of all translational events (Dulude, Baril et al. 2002, Girnary, King et al. 2007). If the slippery sequence is read correctly the downstream Gag precursor stop codon remains in frame. When the ribosome slips in the slippery sequence, all six uridines are recognized, but the adenine is detected twice, “slipping” the ribosome to a different open reading frame downstream of the poly-U sequence that encodes the *pro* and *pol* genes. This produces the 160 kDa Gag-Pro-Pol polypeptide, which contains the viral protease, reverse transcriptase, and integrase. This system produces ten to twenty more Gag precursor molecules than Gag-Pro-Pol polypeptide, maintaining a Gag to Gag-Pol ratio that supports proper structural organization of progeny virions (Jacks, Power et al. 1988, Freed 2015).

2 Materials and Methods

2.1 Cloning of Recombinant DNA Vectors

Within this work, eukaryotic cell cultures were transfected with recombinant DNA expression vectors to investigate the gene expression intensity and the modulation of nuclear export patterns. To this end, existing recombinant DNA expression vectors were modulated using several different cloning strategies.

2.1.1 Cloning PCR

Polymerase Chain Reaction (PCR) is a common method used to modify short sequences (<100 nt) in recombinant DNA vectors when cloning. Up to two primers were designed containing the desired altered nucleotide sequence in addition to sufficient complementarity to the original sequence to allow for primer annealing and a compatible restriction enzyme recognition site to the plasmid back bone of choice. Cloning PCR reactions were carried out in 0.2 µL PCR tubes (Starlab) with a final volume of 50 µL. The reagents shown below were mixed together, and centrifuged for 10 seconds (Centrifuge 5417C, Eppendorf).

Table 2.1.1 Cloning PCR Reaction

Reagent	**Volume**	**Distributor**
Template Plasmid	2 µL (1:1000 dilution, approximately 1ng/mL)	
ddH_2O	39.5 µL	
10x Expand™ High Fidelity Buffer (with 15 mM $MgCl_2$)	5 µL	Roche
DNTP mix (10mM dATP, dCTP, dTTP, dGTP)	1 µL	Roche
Forward Primer (100 pmol)	1 µL	Metabion
Reverse Primer (100 pmol)	1 µL	Metabion
Expand™ High Fidelity DNA Polymerase (3.5 U/µL)	0.5 µL	Roche

The tubes were placed in the thermocycler (Professional Biometra TRIO Thermocycler, Analytik Jena), the lid was tightly closed, and the following program was started.

Table 2.1.2 Cloning PCR Thermocycler Program

	Temperature	**Time [mm:ss]**	
Initial denaturation, activation of the "hot-start" polymerase	94 ℃	03:00	
Denaturation	94℃	00:30	34X
Primer Binding	51-62℃	01:00	
Elongation	72℃	01:00	
Final Elongation	72℃	10:00	

The optimal primer annealing temperature was calculated using the NEB Tm calculator (www.neb.com/Tmcalculator).

2.1.1.1 Separation of DNA Fragments on a 1% Agarose Gel

The PCR amplifications were analyzed by gel electrophoresis on a 1% agarose gel (0.5g LE Agarose, Biozym; 50 mL 1x TBE (89 mM Tris-borate (pH 8), 2 mM EDTA, Sigma-Aldrich)). The gel was microwaved until fully dissolved and two drops of 0.025% ethidium bromide solution (250 µg/mL in a dropper bottle, Roth) were added. The gel was poured into a prepared electrophoresis chamber (Febikon) and allowed to solidify. The chamber was filled with 1x TBE buffer until both the gel and electrodes were submerged. For the cloning PCR, 5 µL of the reaction was diluted with 2 µL DNA loading dye. The samples were run alongside a DNA ladder (GeneRuler 1kb DNA-Ladder, Thermo Fisher). The gel electrophoresis was carried out at 75 mA for approximately 20-45 minutes, after which the DNA was visualized and imaged with UV light (312 nm, INTASUV-Systems). When the PCR product had the expected size, the remaining sample was eluted using the Monarch PCR & DNA Cleanup Kit (5µg) (NEB).

2.1.1.2 PCR cleanup

PCR products were purified using the "Monarch PCR & DNA Cleanup Kit" from NEB. The PCR product was diluted with 225 µL DNA Cleanup binding buffer according to the ratio recommended by the manufacturer, and the kit was used as recommended. The product was then eluted from the column with 17 µL ddH_2O, and directly used for restriction enzyme digestion.

2.1.2 Restriction digestion

The restriction digestion of plasmid backbone and insert were carried out in individual 1.5 mL tubes (SafeSeal, Starstedt), on a heating block (Thermostat plus, eppendorf) for 30-120 minutes. The restriction enzymes were purchased from NEB (20U/µL) and the optimal buffers and temperatures were chosen according to NEB recommendations. The digestion reactions were prepared as follows:

Table 2.1.3 Cloning Backbone Restriction Digestion

Digestion Plasmid Backbone	**Volume**
DNA Plasmid	1 µg
NEB Buffer	2 µL
NEB Restriction Enzyme I	1 µL (20U/µL)
NEB Restriction Enzyme II	1 µL (20U/µL)
ddH_2O	ad 20 µL

Table 2.1.4 Cloning Insert Restriction Digestion

Digestion PCR Product	**Volume**
PCR product	16 µL
NEB Buffer	2 µL
NEB Restriction Enzyme I	1 µL (20U/µL)
NEB Restriction Enzyme II	1 µL (20U/µL)
	ad 20 µL

2.1.2.1 Separation and extraction of digested DNA on a 1% agarose gel

After digestion, 3 µL of DNA loading dye was added to the restriction digestion and they were separated by electrophoresis on a 1% Agarose gel prepared according to section 2.1.1.1. The entire 23 µL of the digestion and dye was loaded on the agarose gel and run at 75 mA for 30-120 minutes. The gel was visualized on a UV table (312nm, INTAS®UV-Systems) The desired DNA fragment size was excised from the gel and extracted using the QIAquick® Gel Extraction Kit from Qiagen. The excised band was dissolved in 300 µL Buffer QG at 50°C, and the kit was used according to manufacturer's specifications. The DNA fragment was eluted with 30 µL ddH_2O. The DNA fragment could then be used for ligation or stored at -20°C.

2.1.2.2 Separation of digested DNA on a 0.8% low-melt agarose gel

The backbone and excised DNA fragment were separated by gel electrophoresis on a 0.8% Low melt Agarose gel (0.4 g Low-Melt Agarose (Sieve GP Agarose, Biozym); 50 mL 1x

TB1/10E (89 mM Tris-borate (pH 8), 0.2 mM EDTA, Sigma-Aldrich)). The gel was stirred and microwaved until fully dissolved and two drops of 0.025% ethidium bromide solution (250 µg/mL in a dropper bottle, Roth) were added. The gel was poured into a prepared UV-permissive electrophoresis chamber (Febikon) and allowed to solidify. The chamber was filled with 1x TB1/10E buffer until both the gel and electrodes were submerged. The backbone digestion was mixed with 3 µL of DNA loading dye, loaded on to the low melt agarose gel and run at 35 mA for 30-120 minutes. When the sample was sufficiently separated, the gel was visualized on a UV table (312nm, INTAS®UV-Systems). The desired band was excised using a clean scalpel (Feather® Safeshield Scalpel) and stored in a 1.5 mL tube (SafeSeal, Starstedt). The backbone was then melted at 65℃ for 10 minutes and used directly for ligation. These fragments were stored at -20℃.

2.1.3 Ligation

The digested backbone and DNA fragment were ligated together using ATP-dependent T4 DNA Ligase (5 U/µL, Thermo Fisher) to catalyze the reaction. The ligation reagents were mixed according to the following table.

Table 2.1.5 Ligation Reaction

Reagent	Volume	Distributor
ddH_2O	14 µL	
Plasmid Backbone	1 µL	
Insert Fragment	2 µL	
10X T4 Ligase Buffer (100mM $MgCl_2$, 100mM DTT, 10mM ATP, 500mM Tris-HCl pH 7.5)	2 µL	Thermo Fisher
T4 Ligase	1 µL (5U/µL)	Thermo Fisher

The ligation set up was mixed by gently pipetting up and down, and incubated at room temperature for 45 minutes, or at 16℃ overnight. This ligation mixture was then used for transformation of chemically competent *E. Coli* cells or stored at -20℃.

2.1.4 Alternatives to classical cloning procedures

2.1.4.1 Two-step Cloning PCR

Substitutions, deletions, or insertions can be introduced to a DNA plasmid within primers used for PCR amplification. This method can be particularly useful when the region that needs to be altered lacks nearby unique restriction enzyme recognition sites.

This method required 3 separate PCR reactions. The first two PCRs were done concurrently with PCR amplifying the 5' region of the gene, and PCR 2 amplifying the 3' region of the sequence. The primers were designed so that the reverse primer of PCR 1 and forward primer of PCR 2 overlapped and contained the desired alteration. The forward primer of PCR 1 and reverse primer of PCR 2 contained the desired restriction enzyme recognition sites. Using Expand™ High Fidelity polymerase (Roche), the first two PCR reactions were carried out according to 2.1.1. The optimal primer annealing temperature was calculated using the NEB Tm calculator (www.neb.com/Tmcalculator), with a final primer concentration of 300 nM. When the PCR was finished, 5 µL of the PCR product was mixed with 2 µL DNA loading dye and run on a 1% Agarose gel according to 2.1.1.1. When PCRs 1 and 2 were both successful, they were cleaned up using the "Monarch® PCR & DNA Cleanup Kit" from NEB and eluted in 20 µL ddH_2O. These PCR products were then used as the template for PCR 3 or stored at -20°C. PCR 3 was mixed in 0.2 mL reaction tubes (Starlab) according to the table below and quickly centrifuged.

Table 2.1.6 Two-step Cloning PCR

Reagent	**Volume**	**Distributor**
Template	1 µL each of PCRs 1 and 2	
ddH_2O	39.5 µL	
10x Expand™ High Fidelity Buffer (with 15 mM $MgCl_2$)	5 µL	Roche
DNTP mix (10mM dATP, dCTP, dTTP, dGTP)	1 µL	Roche
Forward Primer of PCR 1 (100 pmol)	1 µL	Metabion
Reverse Primer of PCR 2 (100 pmol)	1 µL	Metabion
Expand™ High Fidelity DNA Polymerase (3.5 U/µL)	0.5 µL	Roche

The tubes were placed in a thermocycler, the lid was tightly closed, and the following program was run for 34 cycles.

Table 2.1.7 Two-step cloning PCR Thermocycler Program

	Temperature	**Time [mm:ss]**	
Initial denaturation, activation of the "hot-start" polymerase	94 ℃	03:00	
Denaturation	94℃	00:30	35 x
Primer Binding	51-62℃	01:00	
Elongation	72℃	01:00	
Final Elongation	72℃	10:00	

The primer annealing temperature was chosen according to the melting temperature of the primers used. After the program finished, 5 µL of PCR 3 was mixed with 2 µL DNA loading dye and separated on a 1% agarose gel according to 2.1.1.1. If the DNA insert was successfully produced, the fragment was digested and ligated into the chosen backbone.

2.1.4.2 Oligo Annealing

When the desired nucleotide modification was located in a short fragment encompassed by unique restriction enzyme recognition sites, complementary oligos were designed containing the desired alteration and annealed to form the insert. The oligos were mixed together in equal quantities (10 µL each of 100 µM stocks; Metabion) in a 1.5 mL reaction tube and incubated for 3 minutes at 95℃. Then the heating block was switched off, and the oligo mixture was allowed to cool down slowly in the heating block for 45 minutes. This annealed oligo mixture was digested and ligated with the linearized backbone of choice at a ratio of 3:1.

2.1.4.3 In-Fusion® Seamless Cloning

The In-Fusion® Snap Assembly Kit from Takara Bio can directionally insert DNA fragments into a linearized vector without the need for complementary restriction enzyme recognition sites. This method was used when the insertion site lacked unique or sticky restriction enzyme recognition sites.

Using the Q5 polymerase (NEB), the whole plasmid backbone was amplified in a single PCR reaction, with 15 bp overlapping ends to the desired DNA insert. The DNA insert was also amplified using Q5 polymerase, and the primers for both the backbone and insert were designed using the Takara Primer design tool (https://www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools). The Q5 PCR reaction was mixed in 0.2 mL PCR tubes (Starlab) according to the table below and quickly centrifuged.

Table 2.1.8 In-Fusion® Seamless Cloning PCR

Reagent	**Volume**	**Distributor**
Template	2 µL (1:1000 dilution, approximately 1ng/mL)	
ddH_2O	39.5 µL	
5x Q5 Reaction Buffer	10 µL	NEB
dNTP mix (10mM dATP, dCTP, dTTP, dGTP)	1 µL	Roche
Forward Primer of PCR 1 (10 µM)	1.5 µL	Metabion
Reverse Primer of PCR 2 (10 µM)	1.5 µL	Metabion
Q5 High Fidelity DNA Polymerase (2 U/µL)	0.5 µL	NEB

The tubes were placed in a thermocycler, the lid was tightly closed, and the following program was started.

Table 2.1.9 In-Fusion® Seamless Cloning PCR Thermocycler Program

	Temperature	**Time [mm:ss]**	
Initial denaturation, activation of the "hot-start" polymerase	98°C	01:00	
Denaturation	98°C	00:30	35 x
Primer Binding	50-72°C	00:30	
Elongation	72°C	00:30/kb	
Final Elongation	72°C	20:00	

Following the PCR, 5 µL of the PCR product was mixed with 2 µL DNA loading dye and separated according to 2.1.1.1. When the desired PCR product was generated, it was purified by "Monarch® PCR & DNA Cleanup Kit" from NEB according to 2.1.1.2. The purified PCR products were mixed together for assembly according to the table below.

Table 2.1.10 In-Fusion® Seamless Cloning Ligation

Reagent	Volume	Distributor
Purified Insert	10-200ng	
Linearized Vector	50-200 ng	
5x In-Fusion Snap Assembly Master Mix	2 µL	Takara Bio
ddH_2O	Up to 10 µL	

The quantity of vector and insert were determined using the Molar ratio calculator for In-Fusion cloning provided by Takara Bio (https://www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools). The reaction was incubated at 50℃ for 15 minutes and then placed on ice. The reactions were directly transformed and stored at -20℃.

2.1.5 Plasmid Transformation into chemically competent *E. Coli* cells

The ligated DNA plasmids were transformed into chemically competent *E. Coli* cells, which amplified the plasmid and expressed the encoded ampicillin resistance gene. Colonies amplifying this plasmid were selected via growth in the presence of ampicillin. The DH10β strain from NEB was used. The bacteria were stored in 50 µL aliquots at -80℃. For each transformation, the bacteria were thawed and further aliquoted into 10 µL in 1.5 mL tubes (SafeSeal, Starstedt). Then, 6 µL of ligation mixture was added, mixed by gentle flicking of the tube, and incubated for 20 minutes on ice. The bacteria were "heat-shocked" for 45 seconds at 42℃ (Thermostat plus, Eppendorf) to increase the DNA uptake. After incubating the heat-shocked bacteria on ice for 5 minutes, 800 µL of Luria Broth Base Medium (LB-Medium, Invitrogen) without antibiotics was added, and the cells were incubated on a spinning wheel (40 cycles/minute) for 1-2 hours at 37℃. Then, 400 µL of the transformations were spread on an ampicillin-containing agar plate (100 µg/mL, Roche) and incubated overnight at 37℃. Only bacteria that expressed the ampicillin resistance gene found on the transformed plasmid were able to replicate and form colonies. These colonies were picked for DNA-Mini-Preps.

2.1.6 Plasmid DNA minipreparation

Colonies were picked using a clean pipette tip and transferred into a test tube containing 5 mL of ampicillin-containing (100 µg/mL, Roche) LB-Medium (Invitrogen). These test tubes were incubated on a spinning wheel (40 cycles/minute) overnight at 37℃.

The next day, 2 mL of the bacterial culture were transferred to 2 mL reaction tubes for DNA isolation and centrifuged for 1 minute at 14,000 rpm (Centrifuge 5417C, Eppendorf). The remaining 3 mL of bacterial culture were stored at 4℃. The supernatant from the pelleted bacteria was discarded, and the pellets resuspended in 300 µL chilled Buffer 1 (50 nM Tris-HCl: pH 7.5, 10mM EDTA, 400 µg/mL RNase A) by scratching the tube on a metal grid. The resuspended bacteria were lysed with 300 µL Buffer 2 (0.2 M NaOH, 1%SDS) for 5 minutes at room temperature. The alkaline lysis was stopped with 300 µL chilled Buffer 3 (3 M KAc: pH 5.5) mixed by inversion. The cellular bacterial components were pelleted by centrifugation for 15 minutes at 14,000 rpm at 4℃. The supernatant was transferred to a fresh 1.5 mL reaction tube containing 600 µL of Isopropanol and mixed by inversion. The plasmid DNA was precipitated by centrifugation for 30 minutes at 14,000 rpm at 4℃. The supernatant was aspirated, and the pellet washed twice with 150 µL 70% ethanol for 10 minutes at 14,000 rpm at 4℃. The DNA pellet was air dried for 5-10 minutes and then resuspended in 20 µL ddH_2O. From here the mini-preps were analyzed by restriction enzyme digestion and stored at -20℃.

The obtained DNA plasmids were digested with specific restriction enzymes that showed a visible distinction between the original backbone and desired new plasmid. The enzymes were chosen by reference to the *in-silico* build model of the plasmids using Geneious Software (Geneious Version R10, (https://www.geneious.com)). The software also provided the expected band size, which were compared to the observed bands.

For the digestion analysis, the DNA-mini-prep were digested with up to 3 different restriction enzymes (20 U/µL, NEB) in one reaction. The digestion was carried out in 1.5 mL reaction tubes with reagents added according to the following table.

Table 2.1.11 Cloning Check Digestion

Reagent	**Volume**	**Distributor**
DNA-mini-prep	5 µL	
10x Digestion Buffer (decided based on restriction enzyme requirements)	2 µL	NEB
Restriction Enzyme	0.3 µL/each enzyme	NEB
ddH_2O	Up to 20 µL	

The digestion was carried out in a heating block (Thermostat plus, Eppendorf) for 20-60 minutes at the optimal temperature for the chosen restriction enzymes. Then, 3 µL of DNA loading dye was added to the digestion, quickly centrifuged, and 10 µL of each digestion were separated on a 1% agarose gel according to 2.1.1.1 for 20-60 minutes depending on the size of the expected bands. To identify positive clones, the digestion

reactions were run alongside a DNA ladder (GeneRuler 1kb DNA-Ladder, Thermo Fisher). If the isolated DNA restriction digest pattern matched the expected digestion pattern, the samples were sent for Sanger sequencing by Eurofins and the remaining 3 mL of bacterial culture were used to inoculate 100 mL of ampicillin-containing (100 µg/mL, Roche) LB-Medium (Invitrogen).

2.1.7 Plasmid DNA midipreparation

To obtain a larger quantity of plasmid DNA, the remaining 3 mL from the positive DNA-mini-prep was used to inoculate 100 mL of ampicillin-containing (100 µg/mL, Roche) LB-Medium (Invitrogen) and incubated overnight on a shaker (GCL) at 37℃. Glycerin stocks for long term storage were made by diluting 700 µL of the bacterial culture with 300 µL of 100% glycerin (Roth), vortexing, and storing at -80℃. DNA plasmids were purified from these 100 mL of bacterial culture using the DNA-Mid-Kit from Qiagen.

The bacterial culture was split into two 50 mL tubes (Greiner Bio-One™) and centrifuged (3K30 Sartorius, Sigma Laboratory Centrifuges) at 10,000 rpm for 10 minutes at 4℃. The supernatant was discarded, and the bacteria of both pellets were reunited by resuspension in 4 mL of chilled Buffer 1 (50 mM Tris-HCl; pH 7.5, 10 mM EDTA; pH 8, 400 µg/mL RNase A, Qiagen Midi Kit). The bacteria were lysed by adding 4 mL of Buffer 2 (0.2 M NaOH, 1% SDS, Qiagen Midi Kit), inverting the tubes, and incubation at room temperature for 5 minutes. The alkaline lysis was stopped by adding 4 mL of chilled Buffer 3 (3M KAc; pH 5.5, Qiagen Midi Kit) to the mixture, and inverting. The cell components were pelleted by centrifuging the mixture at 12,000 rpm for 15 minutes at 4℃. The supernatant was then loaded on to a Qiagen-tip 100 (Qiagen Midi Kit) equilibrated with 4 mL Buffer QBT (750 mM NaCl, 50 mM MOPS; pH7, 15% Isopropanol (v/v), Qiagen Midi Kit) avoiding the pelleted cell components. After the supernatant had passed through the filter, the filter was washed twice with 10 mL Buffer QC (1 M NaCl, 50 mM MOPS: pH 7, 15% Isopropanol (v/v), Qiagen Midi Kit), and the flow through was discarded. The bound plasmid DNA was eluted using 5 mL of Buffer QF (1.25 M NaCl, 50 mM Tris-HCl; pH 8.5, 15% Isopropanol (v/v), Qiagen Midi Kit) into a plastic test tube (Nalgene) containing 3.5 mL Isopropanol. The eluted and precipitated plasmid DNA was pelleted by centrifugation (Heraeus Megafuge 8R, Thermo Scientific) for 30 minutes at 14,000 rpm and 4℃. The supernatant was discarded, and the pellet was washed twice with 2 mL 70% Ethanol at 14,000 rpm for 10 minutes at 4℃ each. The pellet was air dried at room temperature, dissolved in 100 µL ddH_2O, and transferred to a 1.5 mL reaction tube (SafeSeal, Starstedt).

The concentration of the isolated plasmid was determined using the Spectrophotometer (DS-11+ Spectrophotometer, DeNovix) at 260 and 280 nm. A diluted aliquot with a concentration of 1 µg/µL was made by dilution with ddH_2O. The plasmids were stored at -20℃.

2.1.8 Plasmids and cloning strategies

The original pcEnv plasmid was cloned by Dr. Frank Hillebrand, as well as the CMV Gag RRE and CMV Gag 4XCTE plasmids. The pcTat and pcRev vectors were cloned by Dr. Marek Widera and kindly provided for use in this work. The original SVtat-rev-envRL was also cloned by Dr. Marek Widera, and kindly provided for use in this work. Additionally, Dr. Alan Cochrane from the University of Toronto kindly provided the pNL4-3 ΔE GFP plasmid for use in this work.

The plasmids produced for this work and their respective cloning strategies are provided in Table 2.1.12 below. Figure 2.1.1 shows the LTR Env SD4- plasmid that was mutated to produce the HEXplorer optimized mutants analyzed in this work.

Table 2.1.12. Cloning strategies for newly generated plasmids.

Plasmid Name:	Insert Template:	Backbone:	Cloning path:	Confirmed on:
pc Rev IR Tat	pWPI-BSD[p126]	pcTat	Inserted IRES from pWPI-BSD[p126] with primers 6850/6851 and Rev from pcRev with primers 6852/6853	05.07.2023
pc IR Tat	pWPI-BSD[p126]	pcTat	Inserted IRES from pWPI-BSD[p126] with primers 6850/6851	05.07.2023
SVtat-rev-envRL RRE0	pcENV RRE0	SVtat-rev-envRL RRE	Inserted RRE0 from pcENV RRE0 with Primers 7019/6843	16.07.2023
SVtat-rev-envRL RREpos	RREpos gene strand from Biocat	SVtat-rev-envRL RRE	Inserted RREpos from RREpos genestrand with Primers 7017/6843	16.07.2023

SVtat-rev-envRL RREneg	RRE neg gene strand from Biocat	SVtat-rev-envRL RRE	Inserted RREneg from RRE neg genestrand with Primers 7018/6843	16.07.2023
LTR SD4-ENV CRSwt RRE0	pcENV RRE0	LTR SD4-ENV CRSwt RRE0	Insert digested from pcENV RRE0 (EcoNI/Ale-V2)	18.07.2023
LTR SD4-ENV CRSpos RREpos*BB	LTR SD4- Env CRSpos RRE pos	LTR SD4-ENV CRSwt RREwt	Insert digested from LTR SD4- Env CRSpos RRE pos (EcoRI-HF/XbaI)	18.07.2023
LTR SD4-ENV CRSwt RREpos	LTR SD4- Env CRSwt RRE pos	LTR SD4-ENV CRSwt RREwt	Insert digested from LTR SD4- Env CRSwt RRE pos (Bsu36I/XbaI)	24.07.2023
LTR SD4-ENV CRSpww RREwt	LTR SD4- Env CRSpos RREwt	LTR SD4-ENV CRSwt RREwt	Insert digested from LTR SD4- Env CRSpos RREwt (KpnI-HF/SacII)	02.08.2023
LTR SD4-ENV CRSppw RREwt	LTR SD4- Env CRSpos RREwt	LTR SD4-ENV CRSwt RREwt	Insert digested from LTR SD4- Env CRSpos RREwt (KpnI-HF/Bsu36I)	02.08.2023
LTR Env SD4-wpwRw	LTR Env SD4-pppRw	LTR Env SD4-wwwRw	DNA Fragment from LTR Env SD4-pppRw (SacII/Bsu36I)	12.02.2024
LTR Env SD4-wwpRw	LTR Env SD4-pppRw	LTR Env SD4-wwwRw	DNA Fragment from LTR Env SD4-pppRw (Bsu36IEcoNI)	12.02.2024

LTR Env SD4-pwpRw	LTR Env SD4-wwwRw	LTR Env SD4-pppRw	DNA Fragment from LTR Env SD4-wwwRw (SacII/Bsu36I)	15.02.2024
LTR Env SD4-wppRw	LTR Env SD4-wwwRw	LTR Env SD4-pppRw	DNA Fragment from LTR Env SD4-wwwRw (KpnI/SacII)	12.02.2024
LTR Env SD4-1w2p2pRw	LTR Env SD4-wwwRw	LTR Env SD4-pppRw	PCR product from wwwRw (7126/7127) (Bsu36I/EcoNI)	29.04.2024
LTR Env SD4-2p2p3pRw	LTR Env SD4-pppRw	LTR Env SD4-pppRw	DNA Fragment from PCR product from pppRw (7126/7127) (KpnI/SacII)	29.04.2024
LTR Env SD4-2p2p2pRw	LTR Env SD4-pppRw	LTR Env SD4-2p2p3pRw	PCR product from pppRw (7128/7129) (Bsu36I/EcoNI)	10.05.2024
LTR Env SD4-3p2p2pRw	LTR Env SD4-pppRw	LTR Env SD4-1w2p2pRw	PCR product from pppRw (7128/7129) (KpnI/SacII)	10.05.2024
LTR Env SD4-nppRw	Genestrand from Biocat CRRneg	LTR Env SD4-pppRw	Genestrand from Biocat CRRneg (KpnI/SacII)	10.05.2024
LTR Env SD4-pnpRw	Genestrand from Biocat CRRneg	LTR Env SD4-pppRw	Genestrand from Biocat CRRneg (SacII/Bsu36I)	10.05.2024

CMV Vpu-RL 4XCTE	CMV SD1 GAG 4XCTE	CMV Vpu-RL RRE	PCR fragment from CMV SD1 GAG 4XCTE (7119/7120)	20.02.2024
SV40 Vpu-RL 4XCTE	CMV SD1 GAG 4XCTE	SV40 Vpu-RL RRE	PCR fragment from CMV SD1 GAG 4XCTE (7121/7120)	20.02.2024
CMV SD1 GAG RRE-4XCTE	CMV SD1 GAG RRE	CMV SD1 GAG 4XCTE	PCR fragment from CMV SD1 GAG RRE (7117/7118)	20.02.2024
CMV SD1 GAG 4XCTE-RRE	CMV SD1 GAG RRE	CMV SD1 GAG 4XCTE	PCR fragment from CMV SD1 GAG RRE (7130/7118)	20.02.2024

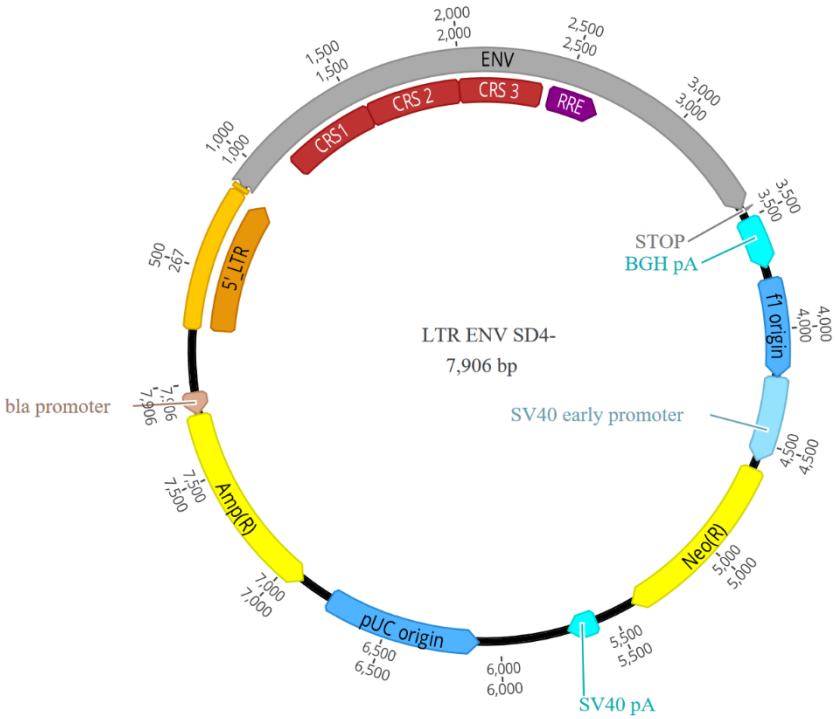


Figure 2.1.1. Schematic of the LTR Env SD4- plasmid. The LTR promoter is shown in orange, the *env* open reading frame is shown in gray, the CRS are in red. The antibiotic resistances are shown in yellow, and also origins of replication in dark blue, poly A sites in turquoise, eukaryotic promotors in light blue, and prokaryotic promoters in brown. The name and size of the plasmid are shown in the center.

To investigate the effect of Rev on nuclear export of mutants while accommodating the need for Tat to activate the LTR promoter, two plasmids were produced. Shown below, pc IR Tat, which is referred to as the Mock plasmid, and pc Rev IR Tat, which was cotransfected in the Rev containing samples.

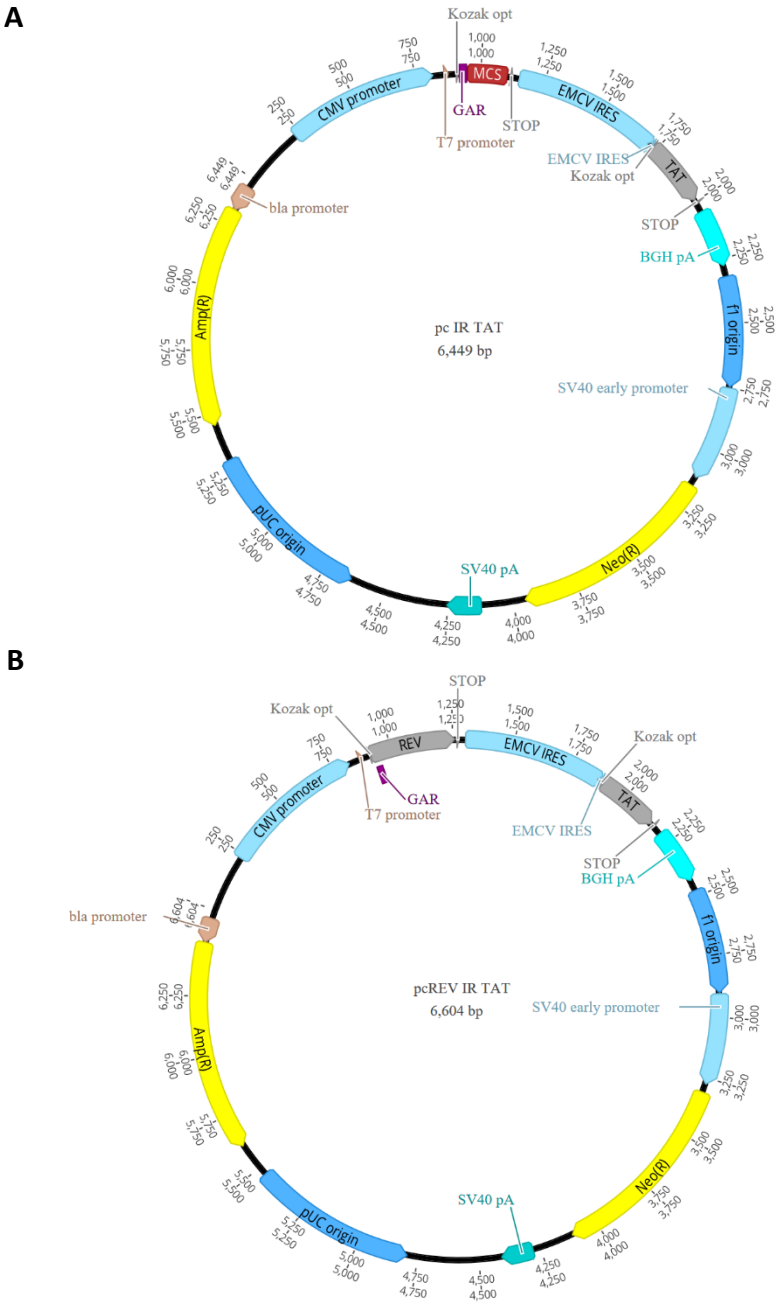


Figure 2.1.2. Schematic of pc IR Tat and pcRev IR Tat. The *rev* and *tat* open reading frames are shown in gray, the IRES is labeled in light blue, and the GAR regulatory sequence is in purple, and the multiple cloning site (MCS) is in red. The antibiotic resistances are shown in yellow, and also origins of replication in dark blue, poly A sites in turquoise, eukaryotic promotors in light blue, and prokaryotic promoters in brown. The name and size of the plasmid are shown in the center. A) Schematic of pc IR Tat. B) Schematic of pcRev IR Tat.

2.1.8.1 Gene Strands

The used gene strands were ordered from and synthesized by Biocat GmbH. They were delivered as dry circularized dsDNA plasmids that were ready for cloning upon resuspension in nuclease-free ddH_2O. The sequences for the ordered gene strands are shown below.

CAGACCTGGAGGAGGCGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAA
AAATTGAACCATTAGGAGTAGCACCCACCAAGGCTAAACGGAGAGTTGTTCAGAGGGAAAAAC
GAGCTGTTGGGATAGGGGCGCTATTTCTTGGCTTTCTTGGGGCGGCGGGGAGCACTATGGGAG
CTGCCTCAATGACGTTAACTGTTCAGGCTAGACAGTTGCTCTCTGATATTGTTCAACAGCAAAATA
ACTTATTACGAGCTATTGAGGCACAGCAGCACCTACTCCAGTTGACTGTTTGGGGCATAAAGCAA
CTGCAAGCGAGAATACTGGCTGTTGAGAGATATTTGAAAGACCAGCAGTTATTAGGCATTTGGG
GCTGCTCGGGCAAGCTAATCTGTACCACTGCTGTGCCTTG

Table 2.1.13. Sequences of the gene strands ordered from Biocat. Restriction enzyme recognition sequences shown in red.

Name	**Sequence (5‘ -> 3‘)**
Env CRS (pos)	TTATGGGGTACCTGTGTGGAAGGAAGCAACCACCACTCTATTTTGTGCATCAGA GCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGT ACCAACAGATCCGAATCCGCAAGAAGTCGTGCTCGTCAACGTCACCGAGAACT TCAACATGTGGAAGAACGACATGGTGGAGCAGATGCACGAGGACATCATCAG CCTATGGGACCAATCGCTGAAGCCGTGCGTGAAGCTGACGCCGCTATGCGTAT CGCTGAAGTGCACCGACCTGAAGAACGACACCAACACCAACAGCAGCAGCGG AAGGATGATCATGGAGAAAGGAGAAATCAAGAACTGCAGCTTCAACATCAGC ACCAGCATCCGAGACAAGGTGCAGAAGGAGTACGCCTTCTTCTACAAGCTGGA CATCGTGCCAATCGACAACACCAGCTACCGGCTGATCAGCTGCAACACCAGCG TCATCACGCAAGCGTGCCCGAAAGTCTCCTTCGAGCCGATCCCGATCCACTACT GCGCACCCGCGGGATTCGCCATCCTGAAGTGCAACAACAAGACCTTCAACGGA ACTGGACCGTGCACCAACGTCAGCACCGTCCAATGCACGCACGGCATCCGACC AGTCGTCAGCACGCAGCTGCTGCTGAACGGATCGCTGGCAGAAGAAGATGTC GTCATCAGAAGCGCCAACTTCACCGACAACGCCAAGACCATCATCGTCCAACTC AACACCAGCGTGGAGATCAACTGCACAAGACCGAACAACAACACAAGAAAGA GCATCCGGATCCAGCGAGGACCTGGAAGAGCGTTCGTCACCATCGGCAAGATC GGCAACATGCGACAAGCGCACTGCAACATCAGCAGAGCGAAGTGGAACGCCA CGCTGAAGCAGATCGCCAGCAAGCTGCGAGAGCAGTTCGGCAACAACAAGAC CATCATCTTCAAGCAGTCCTCAGGAGGAGATCCAGAAATCGTGACGCACAGCT TCAACTGCGGAGGAGAATTTTTCTACTGCAACTCAACGCAGCTCTTCAACAGCA

	CCTGGTTCAACAGCACCTGGAGCACAGAAGGAAGCAACAACACCGAAGGAAG CGACACCATCACGCTGCCGTGCCGCATCAAGCAGTTCATCAACATGTGGCAAG AAGTCGGAAAAGCGATGTACGCACCACCAATCAGCGGCCAGATCCGCTGCAGC AGCAACATCACCGGACTGCTGCTGACAAGAGATGGAGGAAACAACAACAACG GAAGCGAGATCTTCCGACCTGGAGGAGGAGAC
Env CRS (neg)	TTATGGGGTACCTGTGTGGAAGGAAGCAACCACCACTCTATTTTGTGCATCAGA TGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGT ACCCACTGACCCTAATCCCCAGGAAGTTGTTTTAGTTAATGTAACTGAAAATTTT AATATGTGGAAAAATGATATGGTAGAGCAAATGCATGAGGATATTATTTCTTT GTGGGATCAATCTTTAAAACCTTGTGTTAAATTAACTCCCTTGTGTGTTTCTTTA AAATGTACTGATTTAAAAAATGATACTAATACTAATAGTAGTAGTGGGCGTATG ATTATGGAAAAGGGGGAAATTAAAAATTGTTCTTTTAATATTTCTACTAGTATT AGGGATAAAGTTCAAAAGGAATATGCTTTTTTTTATAAACTTGATATAGTTCCC ATAGATAATACTAGTTATAGGTTAATTTCTTGTAATACTTCTGTTATTACTCAGG CTTGCCCTAAAGTTTCTTTTGAGCCTATTCCCATTCATTATTGTGCCCCCGCGGG TTTTGCTATTTTAAAATGTAACAACAAGACTTTTAATGGGACAGGGCCTTGTAC TAATGTTTCTACTGTTCAATGCACTCATGGGATTAGGCCTGTTGTTTCCACCCAA CTTTTATTAAATGGCTCTCTTGCTGAGGAGGATGTTGTTATTAGGTCTGCTAATT TTACTGATAATGCTAAAACTATAATTGTTCAATTAAATACTTCTGTTGAAATTAA TTGTACTAGGCCTAATAATAATACTAGGAAAAGTATCCGGATCCAGAGGGGGC CAGGGAGGGCTTTTGTTACAATTGGGAAAATAGGGAATATGAGGCAGGCTCA TTGTAATATTTCTAGGGCTAAATGGAATGCTACTTTAAAACAAATAGCTAGTAA ACTTAGGGAGCAATTTGGTAACAACAAGACTATTATTTTTAAACAATCCTCAGG AGGGGACCCTGAGATTGTTACTCATTCTTTTAATTGTGGGGGGGAATTTTTTTA TTGCAATAGCACTCAATTATTTAATAGCACTTGGTTTAATAGTACTTGGTCTACT GAGGGGTCTAATAATACTGAGGGGTCTGATACTATTACTCTCCCTTGCCGGATC AAACAATTTATTAATATGTGGCAGGAAGTGGGGAAGGCTATGTATGCCCCTCC CATTTCTGGGCAAATACGATGCAGTAGTAATATAACAGGGTTGCTTTTAACTAG GGATGGGGGTAATAATAATAATGGGTCTGAAATTTTTAGGCCTGGAGGAGGA GAC
RRE (neut)	CAGACCTGGAGGAGGCGATATGAGGGACAATTGGAGAAGTGAATTATATAAA TATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCTAAACG GAGAGTTGTTCAGAGGGAAAAACGAGCTGTTGGGATAGGGGCGCTATTTCTT GGCTTTCTTGGGGCGGCGGGGAGCACTATGGGAGCTGCCTCAATGACGTTAAC TGTTCAGGCTAGACAGTTGCTCTCTGATATTGTTCAACAGCAAAATAACTTATTA CGAGCTATTGAGGCACAGCAGCACCTACTCCAGTTGACTGTTTGGGGCATAAA GCAACTGCAAGCGAGAATACTGGCTGTTGAGAGATATTTGAAAGACCAGCAGT TATTAGGCATTTGGGGCTGCTCGGGCAAGCTAATCTGTACCACTGCTGTGCCTT

RRE (pos)	CAGACCTGGAGGAGGAGACATGCGAGACAACTGGAGAAGCGAGCTCTACAAG TACAAGGTCGTGAAGATCGAGCCGCTCGGCGTCGCACCAACGAAGGCGAAGA GAAGAGTCGTCCAACGAGAGAAGCGAGCAGTCGGCATCGGAGCGCTGTTCCT CGGATTCCTCGGAGCTGCTGGATCGACAATGGGAGCAGCATCGATGACGCTGA CCGTCCAAGCAAGACAGCTGCTGAGCGACATCGTCCAGCAGCAGAACAACCTG CTGCGAGCCATCGAAGCGCAGCAGCACCTGCTGCAGCTGACCGTCTGGGGCAT CAAGCAGCTGCAAGCGAGGATCCTGGCCGTCGAGCGCTACCTGAAGGACCAG CAGCTGCTCGGAATATGGGGATGCAGCGGCAAGCTGATCTGCACCACTGCTGT GCCTTG
RRE (neg)	CAGACCTGGAGGAGGCGATATGAGGGACAATTGGAGAAGTGAATTATATAAA TATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCTAAGAG GAGAGTTGTTCAGAGGGAAAAAAGGGCTGTTGGGATAGGGGCTCTTTTTTTGG GCTTTTTGGGGGCGGCAGGCTCTACTATGGGGGCTGCTAGCATGACTTTAACT GTTCAGGCCCGCCAACTTCTTTCTGATATTGTTCAGCAACAAAATAATCTTTTAC GGGCTATAGAGGCTCAGCAACATTTATTACAATTAACTGTTTGGGGGATTAAAC AATTACAGGCTAGAATTTTAGCTGTTGAGAGATATTTAAAAGATCAGCAATTAT TAGGCATTTGGGGCTGCTCAGGGAAATTAATTTGTACCACTGCTGTGCCTTG

2.2 Eukaryotic Cell Culture

In this thesis, three adherent eukaryotic cell cultures were used for analysis: HeLa, TZMbL, and HEK293T. They were grown in T75 cell culture flasks in 10 mL of Dulbecco's Modified Essential Medium (DMEM, Gibco) that was supplemented with 10% Fetal Bovine Serum (FBS Supreme, PAN Biotech) and 1% Pen/Strep (Gibco, 10,000U/ml Penicillin, 10,000µg/ml Streptomycin). All cell lines in this thesis were incubated at 37°C and 5% CO_2 (Heraeus BBD6220, Thermo Scientific).

2.2.1 Maintenance

The cells were split twice a week and reseeded with fresh medium. When splitting the cells, the medium was removed, and the cells were gently washed with phosphate buffered saline (PBS, Gibco) twice. The cells were detached from the bottom of the flask using 1.5 mL 0.05% Trypsin-EDTA (Gibco). The trypsin was spread evenly over the cell layer and then removed. After a 1–5-minute incubation at 37°C, the loosened cells were collected with 10 mL of fresh DMEM (10% FBS, 1% Pen/Strep), which additionally stops the trypsin digestion. The cell number was determined using a C-Chip Disposable Hemocytometer (NanoEntek) and 0.5-2 mL of cells were replated in a new T75 flask with 10 mL of fresh DMEM (10% FBS, 1% Pen/Strep), spread evenly over the bottom of the flask and returned to the incubator.

2.2.1.1 HeLa

The HeLa cell line was the first immortalized cell line developed and is still commonly used in scientific research. The cell line is derived from a cervical epithelial carcinoma cell sample taken from a patient, Henrietta Lacks, without her permission in 1951. HeLa cells were used in the development of the polio vaccine in 1953 (Scherer, Syverton et al. 1953), and were finally sequenced 2013 (Landry, Pyl et al. 2013).

2.2.1.2 TZMbL

The TZMbL cell line is derived from the HeLa cell line with a parental cell line of JC.53 and highly expresses CD4 CCR5, and CXCR4. This cell line has been altered to indicate HIV-1 infection by expressing integrated copies of Firefly Luciferase and β-galactosidase genes under the control of HIV-1 LTR promoters. They are extremely sensitive to infection by diverse HIV-1 strains including both macrophage and T-cell tropic strains. TZMbL cells are used as tools for infection control and titer determination (Platt, Wehrly et al. 1998).

2.2.1.3 HEK293T

Human embryonic kidney (HEK) 293T cells were derived from the kidney cells of an aborted healthy female fetus. Frank Graham transfected them with adenovirus 5 DNA, which resulted in a stable cell line (Graham, Smiley et al. 1977). Additionally, they express a stably transfected plasmid encoding a temperature sensitive mutant of the SV40 large T-antigen, which allows the replication of plasmids that contain a SV40 origin of replication (ORI) (DuBridge, Tang et al. 1987). Here, HEK293T cells are used to generate pseudoviral particles and HIV-1 virus stocks.

2.2.2 Cell concentration determination and seeding

The concentration of a cell suspension was determined to standardize the number of cells seeded across assays. To this end, 10 µL of the cell suspension was mixed with 10 µL of 0.4% trypan blue stain (Gibco) and half of this mixture was loaded into the counting chamber (C-Chip, Neubauer Improved, NanoEnTek). The cells from two of the four counting chambers were counted using an inverted microscope (Nikon Eclipse TS100) at a 40x magnification with a click counter. Trypan-stained dead cells were excluded from the count and the number of living cells were calculated using the following equation:

$$\frac{Total\ Cell\ Count}{Number\ of\ Counted\ Squares} \times Dilution\ Factor \times 10{,}000 = Cells\ per\ mL$$

And the desired dilution of cells for the assay was calculated by the following equation:

$$\frac{Cells\ per\ mL}{Desired\ Cells\ per\ mL} - 1 = Volume\ of\ Media\ Needed\ per\ mL\ Cell\ Suspension\ (mL)$$

Diluted adherent cells were then seeded into 6-well, 12-well, or 24-well plates for transfection experiments. The cells were diluted to a concentration of 100,000 cells per mL (DMEM, 10% FBS, 1% Pen/Strep) for fluorescent in situ hybridization studies and 2 mL of the diluted cells were added to 6-well plates (TPP®, Merck) containing 18mm glass coverslips (Marienfeld). For all other studies, the cells were diluted to a concentration of 150,000 cells per mL (DMEM, 10% FBS, 1% Pen/Strep) and plated at a volume of 0.5 mL per well for 24-well plates, 1 mL per well for 12-well plates, and 2 mL for 6-well plates. The plates were gently rocked to spread the diluted cells throughout each well. The seeded plates were then incubated overnight at 37°C and 5% CO_2.

2.2.3 Transfection of eukaryotic cell lines

Within this work, human cell lines were transfected with different DNA plasmids. DNA plasmids were transfected using TransIT®-LT1 Transfection Reagent (Mirus bio). The transfection reagent was used with a ratio of 2 µL TransIT®-LT1 per µg of DNA plasmid. For gene expression assays, 1 µg of each DNA plasmid per 1 mL of cell culture to be transfected was aliquoted into a fresh 1.5 mL reaction tube (SafeSeal, Starstedt). A second 1.5 mL reaction tube was prepared for each transfection with 100 µL of DMEM per mL of cell culture and 2 µL TransIT®-LT1 per 1 µg of plasmid. This mixture was incubated at room temperature for five minutes on a sterile bench and transferred to the corresponding plasmid-containing tube. This mixture was incubated at for a further 15 minutes at room temperature before being gently pipetted dropwise onto the seeded cells at a ratio of 100 µL per mL of cell culture media (DMEM, 10% FBS, 1% Pen/Strep), and distributed via gently rocking of the plate. The transfected cells were incubated at 37°C and 5% CO_2 until harvest for downstream assays.

2.2.4 Inhibitors

2.2.4.1 KPT330

KPT330 was used as a noncovalent small molecule inhibitor of the nuclear export inhibitor of the CRM1 nuclear export factor. It blocks CRM1 mediated nuclear export by interacting with the hydrophobic groove that CRM1 uses to bind the leucine-rich sequences of its export targets. The used 10mM KPT330 stock was kindly provided by Dr. Alan Cochrane from the University of Toronto. Once thawed, the KPT330 stock was aliquoted in 250 µM stocks of 50 µL each to avoid multiple freeze thaw steps. To inhibit CRM1 mediated nuclear export, the stock solution was added to the cell culture medium for a final concentration of 250 nM and incubated at 37°C and 5% CO_2 for 18 hours.

2.3 Production of Pseudoviral particles

On the first day, $3x10^6$ HEK293T were seeded into each T75 cell culture flask, in 10ml DMEM medium supplied with 10% FCS and 1% Pen/Strep. After overnight incubation at 37°C and 5% CO_2, the cells were transfected with the mixture of the needed plasmids (6µg of pNL4-3 ΔE GFP, 0.5 µg Env mutants) using the transfection reagent TransIT®-LT1 (Mirus bio). The following plasmids were used:

- pNL4-3 ΔE GFP, encoding the viral RNAs *gag, pol, tat,* and *rev*
- LTR Env SD4-, encoding for a mutated HIV-1 glycoprotein
    - LTR Env SD4- wwwR^w
    - LTR Env SD4- pwwR^w
    - LTR Env SD4- ppwR^w
    - LTR Env SD4- pppR^w
    - LTR Env SD4- pppR^p
    - pcVSVg, encoding the glycoprotein of VSV (vesicular stomatitis virus)

For the transfection, 13 µL of TransIT®-LT1 (Mirus bio) was diluted in 250 µL of serum-free DMEM. After a 5-minute incubation, the *Trans*IT-LT1 -containing serum was added to 6µg of pNL4-3 ΔE GFP and 0.5 µg of the glycoprotein mutants and incubated for 15 minutes. When the incubation time was over, the *Trans*IT-LT1 /plasmid mixture was gently pipetted into the corresponding plate and spread evenly across the cell lawn by gentle movement of the flask. The cells were incubated overnight at 37°C and 5% CO_2. The next day, the DMEM medium was removed from the cells and replaced with 4.5 mL IMDM, supplied with 10% FBS and 1% Pen/Strep, and incubated for a further 48 hours at 37°C and 5% CO_2. After 48 hours, the supernatant of each flask was harvested and filtered using a 0.45 µM syringe filter and aliquoted into 1 mL stocks and stored at -80°C.

2.4 RNA Analysis

2.4.1 Total RNA isolation

To analyze the abundance and splicing pattern of transcripts expressed by the transfected plasmids, total RNA was isolated from eukaryotic cells. The cells were washed twice with PBS before being incubated in 2 mM EDTA in PBS for 2-5 minutes at room temperature to loosen adherent cells. The cells were pipetted into 1.5 mL tubes (SafeSeal, Starstedt) and shortly centrifuged. The supernatant was discarded, and the cell pellet was lysed in 500

µL solution D (SolD; 4M guanidinium thiocyanate, 25mM sodium citrate, 0.5% sarcosyl, 0.1M β-mercaptoethanol). The RNA isolation was either interrupted at this point by freezing the samples at -20°C or proceeding directly to the phenol/chloroform extraction. The following steps were carried out under a hood, the tubes and reagents were kept on ice and centrifugation steps were carried out in a pre-cooled centrifuge at 4°C. Each tube was supplied with 7.2 µL β-mercaptoethanol (Sigma Aldrich), 50 µL 2M sodium acetate (pH 4), and 500 µL phenol (Roti®-Aqua-Phenol, Roth), prepared as a master mix. Afterward, 103 µL chloroform/IAA (24:1) was added to each sample and vortexed for 15 seconds, during which the samples became cloudy white. The samples were incubated on ice for ten minutes to facilitate phase separation and then centrifuged at 10,000 rpm and 4°C for 20 minutes (Heraeus Megafuge 8R, Thermo Fisher). The water phase (2 x 200 µL) was transferred to a fresh 1.5 mL tube without disturbing the phase separation, leaving the denser phenol phase in the original tube. The separated water phase was mixed with an equal volume of isopropanol (400 µL), vortexed and incubated at -20°C for 1 hour to overnight to precipitate the RNA. The RNA was then centrifuged in a precooled centrifuge at 4°C and 10,000 rpm for 30 minutes. The supernatant was removed, and the RNA pellet was washed twice with 150 µL 70% Ethanol, followed by a 10-minute centrifugation at 4°C and 10,000 rpm. After the final 70% Ethanol wash was removed, the pellet was allowed to dry on ice for 10 minutes with an open lid. The RNA was resuspended in 10 µL of Millipore water and the concentration of the isolated RNA was measured via photometric analysis using the Nanodrop 100 spectral photometer (ND-1000 Version-3.7.0). Isolated RNA and stored at -80°C.

2.4.2 Northern Blot analysis

With Northern blot analysis, isolated total RNA is separated by size, transferred, and crosslinked to a membrane where specific transcripts can be detected by specially designed DIG-labeled probes that are complementary to the RNA of interest. This technique allows the comparison of RNA abundancy as well as size without further amplification.

For analysis via Northern blot, 4-6 µg of total RNA of a given sample is separated on a denaturing 1% agarose gel ((1g agarose powder (Biozym LE Agarose)), 85ml ddH2O, 10ml 10x MEM (200mM MOPS, 50mM Sodium acetate, 10mM EDTA, pH 7), 5.5ml formaldehyde (37%, Rotipuran®, Roth)) under a fume hood. The total RNA samples are prepared by mixing them with 1 µL recombinant DNase I (10 U/µL, Roche) and bringing them to a volume of 7 µL. Samples were incubated with the DNase for 20 minutes at 37°C (Thermomixer comfort, Eppendorf), followed by 10 minutes at room temperature. One volume (7 µL) of 2X RNA loading dye (Thermo Fischer) was added and incubated at 70°C for 10 minutes. After this incubation, the samples were put on ice until the gel solidified. Then, 13.5 µL of the sample was loaded into a pocket individually and separated at 60V

for 30-180 minutes.

Following the separation, the gel was removed from the running chamber and the ribosomal bands were visualized on a UV table (312nm, INTAS®UV-Systems), and blotted overnight. The blot was stacked in a plastic tray in the following order, starting from the bottom:

- Approx. 20 paper towels, halved and stacked on top of each other
- 3 dry Whatman™ papers (3 MM CHR, GE Healthcare)
- 1 Whatman™ paper (3 MM CHR, GE Healthcare) preincubated in 20X SSC (3M NaCl, 300mM tri-sodium-citrate)
- 1 positively charged nylon membrane (Roche), 1st preincubation in ddH_2O, second preincubation in 20X SSC (3M NaCl, 300mM tri-sodium-citrate)
- 1 pre-run denaturing agarose gel, 1st preincubation in ddH_2O, second preincubation in 20X SSC (3M NaCl, 300mM tri-sodium-citrate)
- 3 Whatman™ papers (3 MM CHR, GE Healthcare) preincubated in 20X SSC (3M NaCl, 300mM tri-sodium-citrate)

Two plastic bowls filled with 20X SSC were put on either side of the blot setup and a connected by a strip of Whatman™ paper, where the middle of the strip sat on top of the blot set up. A plastic tray was set on top of the blot setup and weight was stacked on top of said plate. This capillary blotting was left overnight.

The next morning, the blot setup was dismantled, and the membrane reduced to the size of the gel, and the corner of the blot was removed for orientation. The RNA was then UV-crosslinked to the membrane (CL-1000 Ultraviolet Crosslinker, UVP, Energy: 1200x100µJ/cm2) and the 28S (5.0 kb) and 18S (1.9 kb) ribosomal bands were marked with pencil to serve as a size standard. The crosslinked membrane was transferred to a hybridization bottle (Thermo Fischer) and washed twice with ddH_2O. The membrane was pre-hybridized with 10 mL 1X DIG Easy Hyb hybridization solution (Roche) for 2 hours at 55°C in a hybridization oven (Biometra OV 5, 5 rpm). The prehybridization solution was removed, and replaced with specific digoxigenin (DIG)-labeled PCR probes (DIG-11-dUTP alkali-labile; Roche) in 10 mL 1X DIG Easy Hyb hybridization solution and incubated overnight at 55°C.

The next day, the DIG-labeled probe was removed, and the membrane was first washed twice with ddH_2O, and second, twice with stringent wash buffer I (2x SSC, 0.1% SDS) at room temperature. This was followed by two 20-minute washes in stringent wash buffer

II (0.2x SSC, 0.1% SDS) at 68°C in the hybridization oven. After this the membrane was washed twice more in ddH_2O at room temperature and transferred to a plastic box containing maleic acid buffer (0.1M maleic acid, 150mM NaCl, pH 7.5). The blot was then blocked with 1X northern blot blocking reagent (Roche) in maleic acid buffer for 45 minutes on a shaker. After blocking, the membrane was incubated with anti-digoxigenin-AP, Fab fragments (sheep, Roche) diluted 1:20,000 in 1x northern blot blocking solution for 1 hour at room temperature on the shaker. After blocking, the membrane was washed 3 times in maleic acid buffer for 10 minutes each at room temperature, and the bands were visualized with CDP star (Roche) for chemiluminescent reactions (1:100 in AP buffer [0.1M Tris HCl, 0.1M NaCl, pH 9.5]; Roche) using the Lumi-Imager™ F1 (INTAS).

2.4.2.1 DIG-Labeled PCR probes

Specific digoxigenin (DIG)-labeled probes designed for the detection of complementary RNAs in a northern blot are created through two sequential PCR reactions. The first PCR is done with unlabeled dNTPs (Qiagen), using plasmid DNA with the desired sequence as a template. This is followed by a second PCR with DIG-labeled dNTPs (DIG-11-dUTPalkali-labile; Roche) using the purified first PCR as a template.

The first PCR reaction is carried out in 0.2 µL PCR tubes (Starlab) with a final volume of 50 µL according to the protocol in 2.1.1. To verify the success and product size of the first PCR, 2 µL of the first PCR reaction were separated on a 1% agarose gel. The rest of the PCR was cleaned up (Monarch® PCR & DNA Cleanup Kit (5µg), NEB), eluted in 20 µL ddH_2O to be used as the template for the second PCR. A 100 µL DIG-Labeled dNTP stock solution was created using DIG-labeled dUTPs and unlabeled dNTPs for use in the second PCR reaction and stored at -20°C in 10 µL aliquots. 35 µL of the DIG-labeled dUTPs (1 mM DIG-11-dUTP alkali-labile; Roche) were mixed with 1 µL dATP (10 mM), 1 µL dGTP (10mM), 1 µL dCTP (10 mM), 0.65 µL dTTP (10 mM), and 61.35 µL DEPC-ddH2O. The second DIG-labeled PCR was carried out in 0.2 µL PCR tubes (Starlab) with a final volume of 40 µL and using the same primer pair as the first PCR. 4 µL of the prepared DIG-labeled dNTPs were mixed with 2 µL of the prepared, purified PCR probe, 4 µL 10x Expand™ High Fidelity Buffer (with 15 mM $MgCl_2$, Roche), 2 µL forward primer (1:10, 20 pmol, Metabion), 2 µL reverse primer (1:10, 20 pmol, Metabion), 25.5 µL ddH2O, and 0.5 µL Expand™ High Fidelity DNA polymerase (3.5 U/µL, Roche). The PCR program from 2.1.1 was used. The PCR product from the second DIG-labeled PCR (1 µL) was compared to the first PCR (1 µL) on a 1% agarose gel. The DIG-labeled PCR product travels more slowly than the unlabeled PCR product. The DIG-labeled probe was stored at -20°C. The sequence of the DIG-Labeled probes used for northern blot analysis in this thesis are a 153 bp fragment complementary to the HIV-1 Exon 7, produced using LTR Env SD4- $wwwR^w$ as template DNA, produced with Primers #3387 (5'-TTGCTCAATGCCACAGCCAT-3') and #3388 (5'-TTTGACCACTTGCCACCCAT-3').

2.4.3 Analysis of isolated RNAs by PCR amplification

2.4.3.1 cDNA synthesis

To eliminate DNA contamination in the isolated RNA samples before cDNA synthesis, 1 µg of RNA was brought to a volume of 9 µL and incubated with 1 µL DNase I (10 U/µL, Roche) in a 1.5 mL reaction tube. The mixture was incubated at 37°C for 20 minutes, followed by 10 minutes at room temperature. The DNase was heat inactivated at 70°C for 5 minutes, and then the samples were put on ice. 1 µL of ddH_2O, 1 µL of dNTP mix (10mM dATP, dCTP, dGTP, and dTTP, Qiagen) and 1 µL oligo(dT) primers (diluted 1:20, Roche) were added to each RNA sample in a master mix and incubated at 65°C for 5 minutes, before being placed on ice again. 4 µL of 5x first strand buffer (FSB, Invitrogen), 1 µL of DTT (0.1 M, Invitrogen), 1 µL of RNasin® Ribonuclease Inhibitor (40 U/µL, Promega), and 1 µL of SuperScript™ III Reverse Transcriptase (200 U/µL, Invitrogen) were added to the RNA samples in a mastermix and placed in a thermocycler (Professional Biometra TRIO Thermocycler, Analytik Jena). The cDNA program runs at 50°C for 60 minutes, followed by 72°C for 15 minutes. The resulting cDNA can either be directly used for PCR analysis or stored at -20°C.

2.4.3.2 Quantitative RT-q-PCR

Also referred to as qPCR, Quantitative RT-q-PCR monitors the amplification of PCR products between each cycle by measuring the intensity of fluorescence produced by either specific probes that bind to the DNA of interest, or with the addition of SYBR-green dye. SYBR-green is a fluorescent dye that intercalates with double stranded DNA, therefore as the amount of double-stranded DNA doubles with every cycle, so does the fluorescence produced by SYBR-green. The number of cycles needed to pass a predetermined threshold (C_T) can be back calculated to determine the initial amount of target cDNA in the sample when normalized to the amplification of a housekeeping gene. This results in the ΔC_T and is used to compare different samples in the same experiment.

The qPCR reactions were performed in Optical Fast-Reaction Tubes ((8-strips), 0,1mL, Thermo) with Optical Caps ((8 Caps/Strip), Thermo) with a final volume of 20 µL. 2 µL of prepared cDNA was added to the tube, and then 18 µL of target specific master mix (10 µL of Forget-Me-Not™ EvaGreen® qPCR Master Mix (Low ROX, Biotium), 1 µL forward primer (1:10, 10 pmol, Metabion), 1 µL reverse primer (1:10, 10 pmol, Metabion), and 6 µL ddH2O). As a negative control, ddH_2O was used to replace the cDNA for each primer pair. The tubes were closed, and the qPCR reaction mix was briefly centrifuged (Eppendorf Centrifuge 5430) before being loaded into the 7500 Real Time PCR system (Applied Biosystems) and the following program was run:

Table 2.4.1 qPCR Program

	Temperature	**Time [mm:ss]**	
Denaturation	95 °C	02:00	
Amplification	95°C 60°C	00:10 01:00	40X
Melting curve	95°C 63°C 95°C	00:00 00:30 00:00	
Cooling	40°C	00:30	

To calculated relative gene expression from the qPCR from the C_T values, the ΔC_T between the GOI and the reference gene was calculated first by subtracting the C_T(GOI) from the C_T(Ref). The expression of each GOI was then determined by calculating the $e^{\Delta CT}$, and the ratio between the treated and untreated samples was determined by division of these values, ($e^{\Delta CT}$ (treated)/ $e^{\Delta CT}$ (untreated)).

The primer pairs used for qPCR amplification are shown in Table 2.4.2, with the name of the amplified gene and the primers used for amplification with the corresponding primer numbers and corresponding sequences.

Table 2.4.2 qPCR Primers

Amplified GOI	**Primer Pair**	**Primer Sequences (5‘ to 3‘)**
GADPH	#3502/#3503	CCACTCCTCCACCTTTGAC / ACCCTGTTGCTGTAGCCA
HIV-1 Env	#3387/#3388	TTGCTCAATGCCACAGCCAT / TTTGACCACTTGCCACCCAT

2.4.4 Fluorescent in-situ hybridization (FISH)

Fluorescent in-situ hybridization uses up to 48 fluorescently labeled DNA probes that are complementary to the target sequence to visualize the location of the target transcripts in fixed and permeabilized cells. These probes (custom design, Stellaris®) can be designed using the “Stellaris Probe Designer” tool (https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer). The probes designed and used to image *env* transcripts in vivo are shown below in Figure 2.4.1 and Table 2.4.3.

Table 2.4.3 Sequences of Cy-3 labeled probes used to detect Env transcripts in FISH assays.

Probe #	Sequence (5'-3')
1	GATACTTCTCCTTCACTCTC
2	TATCCCAAGGAGCATGGTG
3	GTAGCACTACAGATCATCA
4	GACTGTGACCCACAATTTT
5	CACACAGGTACCCCATAAT
6	ATAGAGTGGTGGTTGCTTC
7	GCTTTAGCATCTGATGCAC
8	TTATGTACCTCTGTATCAT
9	TTATGTACCTCTGTATCAT
10	ATGAGTTTTCCAGAGCAACC
11	AACTAGCATTCCAAGGCACA
12	TGTTCCAGAGATTTATTACT
13	CCAGGTCATGTTATTCCAAA
14	TTGTTAATTTCTCTGTCCCA
15	GGAGTGTATTAAGCTTGTGT
16	CTGGTTTTGCGATTCTTCAA
17	ATTCTTGTTCATTCTTTTCT
18	GCCCATTTATCTAATTCCAA
19	CCACAGCCAATTTGTTATGT
20	ATTCTTAAACCTACCAAGCC
21	CTGCCTAACTCTATTCACTA
22	CTCGGGATTGGGAGGTGGGT
23	GGATCCGTTCACTAATCGA
24	GATCGTCCCAGATAAGTGC
25	TGAAGAGGCACAGGCTCCG
26	AGTCTCTCAAGCGGTGGTA
27	CCTCGTTACAATCAAGAGT
28	CGTCCCAGAAGTTCCACAA
29	CACCAATATTTGAGGGCTTC
30	TCCTGACTCCAATACTGTAG
31	TATGGCTGTGGCATTGAGCA
32	CTTCTATAACCCTATCTGTC
33	AGCTCTATAAGCTGCTTGTA
34	CTTCTAGGTATGTGGCGAA
35	TTTCCAAGCCCTGTCTTAT
36	ACGCGGCCGCTAGCAAAAT
37	AACGGGCCCTCTAGATTCT

38	GAGGCTGATCAGCGGGTTT
39	CTGGCAACTAGAAGGCACAG
40	CAAGGAAGGCACGGGGGAG
41	ACCTACTCAGACAATGCGAT
42	CTGCTATTGTCTTCCCAATC
43	CATAGAGCCCACCGCATC

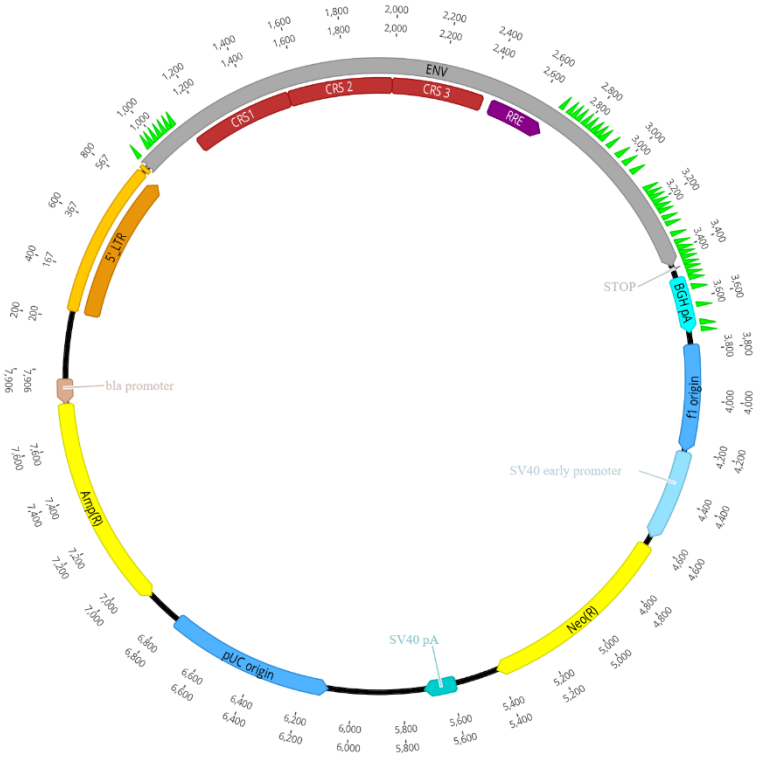


Figure 2.4.1. Schematic showing the position of Stellaris Fluorescent In Situ Hybridization probes on the LTR Env SD4- plasmid. The LTR promoter is shown in orange, the *env* open reading frame is shown in gray, the CRS are in red, and the FISH probes are shown in green. The antibiotic resistances are shown in yellow, and also origins of replication in dark blue, poly A sites in turquoise, eukaryotic promotors in light blue, and prokaryotic promoters in brown.

For FISH analysis to be effective, cells must be at 50-60% confluency at the time of harvest. For harvest, coverslips (Marienfeld) with cells grown on them were in a 12 well plate (TPP) and washed twice with 1 mL PBS to remove any media and debris. The cells were fixed in 1 mL 3.7% Formaldehyde in PBS (Gibco) for 15 minutes at room temperature before being washed twice in 1 mL PBS again. The cells were then permeabilized in 70% Ethanol between 1 hour and 1 week at 4℃. The plate that cells were being permeabilized in was sealed with parafilm to prevent evaporation of the Ethanol. After the permeabilization, cells were washed with FISH Wash Buffer (2X SSPE (0.3 M Sodium Chloride, 0.02 M Sodium Hydrogen Phosphate, 2 mM EDTA, pH 7.4) 10% Formamide (Roth), ddH_2O), and

then incubated in fresh FISH wash buffer for 30 minutes at 37°C. The hybridization buffer (10% Formamide, 12.5 nM Env probe-Cy3, Stellaris® Hybridization buffer) was aliquoted into 35 µL droplets on a labeled parafilm sheet in a humidity chamber. After incubation, the coverslips were dabbed dry on a Kimwipe and deposited cell side down on top of the hybridization mixture droplet and tapped to remove bubbles. After every coverslip was placed, the humidity chamber was sealed and incubated at 37°C overnight in a dark box. The next day, the coverslips were gently removed from the hybridization chamber and placed cell-side up into a fresh 12 well plate. The coverslips were washed with FISH Wash buffer and then incubated in FISH Wash buffer for 30 minutes at 37°C in a dark box. The FISH Wash buffer was removed and replaced with DAPI Wash Buffer (5 ng/mL DAPI,2X SSPE (0.3 M Sodium Chloride, 0.02 M Sodium Hydrogen Phosphate, 2 mM EDTA, pH 7.4) and incubated for 30 minutes at 37°C in a dark box to stain the nuclei. After the nuclei staining, the coverslips were quickly rinsed in FISH Wash buffer to remove excess DAPI stain and then incubated in FISH Wash buffer for 30 minutes at 37°C in a dark box. After the last incubation, the coverslips were mounted on a microscope slide (Marienfeld) using 8 µL VECTASHIELD Antifade Mounting Medium (Vector Labs) and sealed with clear quick-dry nail polish (DM). Once the sealant is dry, the samples are ready to image.

2.4.4.3 Image Acquisition and Analysis

The slides were imaged via fluorescence microscopy and images from separate channels (DAPI-blue), (Cy3-red) were saved as separate files. Ten cells form each coverslip were imaged for quantification and analyzed using the ImageJ Macro "Intensity_Ratio_Nuclei_Cytoplasm" (Intensity Ratio Nuclei Cytoplasm Tool, RRID:SCR_018573) to determine the average percent of cytoplasmic fluorescence of transfected cells in each condition. Images were also processed for visual analysis by removing the background and merging the channels. The individual channels were displayed in monochrome to show the fluorescent signal with higher contrast, and the composite images were displayed in color to show the localization of the fluorescent signal.

2.5 Protein Analysis

2.5.1 Western Blot

2.5.1.1 Protein isolation from cells

To analyze the abundance and size of proteins expressed by the transfected plasmids, protein was isolated from whole eukaryotic cells. To remove media residue, the cells were washed twice with PBS before being incubated in 2 mM EDTA in PBS for 2-5 minutes at room temperature to loosen adherent cells. The cells were pipetted into 1.5 mL tubes (SafeSeal, Starstedt) and shortly centrifuged. The supernatant was discarded, and the cell

pellet was lysed with 20-60 µL RIPA buffer (25mM Tris-HCl (pH 7.6), 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, protease inhibitor cocktail) (EMD Millipore). The lysed cells were frozen at -80°C for at least 10 minutes and then centrifuged at 14,000 rpm for 10 minutes at room temperature. The supernatant was transferred to fresh 1.5 mL tubes and mixed with an equal volume of 2x sample buffer (60mM Tris-HCl (pH 6.8), 24% glycerol, 2% SDS, 14.4mM β-Mercaptoethanol, 1% bromophenol blue). After boiling at 95°C for 10 minutes, the samples were either stored at -20°C or loaded on an SDS-page for western blot analysis.

2.5.1.2 Protein isolation from viral supernatant

To isolate protein from pseudoviral stocks, which contains pseudovirions produced with the Env mutants, 500 µL of viral stock was carefully underlayered with 300 µL 20% Sucrose Buffer (1 mL 1 M Tris(pH 7.5), 3.3 mL 3 M NaCl, 200 µL 0.5 M EDTA, 20 g sucrose, ad. 100 mL ddH2O) and centrifuged for 60 minutes at 50,000 g and 4°C (Sigma laboratory centrifuge 3K30) to pellet the viruses in the media. The supernatant was removed, and the pellet was resuspended in 10 µL RIPA buffer (25 mM Tris-HCl (pH 7.6), 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, protease inhibitor cocktail) (EMD Millipore). The lysed sample was mixed with 10 µL 2X sample buffer (60 mM Tris-HCl (pH 6.8), 24% glycerol, 2% SDS, 14.4 mM β-Mercaptoethanol, 1% bromophenol blue), vortexed and boiled at 95°C for 10 minutes.

2.5.1.3 Sample Analysis

To separate proteins by their molecular weight, the samples were incubated in a sodium dodecyl sulfate (SDS) containing sample buffer and loaded onto polyacrylamide gels for gel electrophoresis. The “Mini-PROTEAN® Tetra Vertical Electrophoresis Cell” from Bio-Rad was used to pour and run gels made using the “TGX Stain-Free FastCast Acrylamide Kit, 12%” from Bio-Rad according to instructions. The running Gel was poured into the chamber first, and a 2 cm gap from the top of the glass plate was left to accommodate the stacking gel. This gap was filled with 100% Isopropanol (VWR Chemicals) to seal the gel from oxygen while it polymerized. Once the Running Gel polymerized, the isopropanol was removed, and the Stacking Gel was layered on top of the polymerized Running gel. The comb was added, and the gel was left to polymerize. When the Stacking Gel had solidified, the comb was carefully removed, and the pockets washed with 1X Laemmli buffer (25 mM Tris (pH 8.8), 192 mM glycine, 0.1% SDS) using a syringe and needle. The gel was set into the “Mini-PROTEAN® Tetra Electrode Assembly” (Bio-Rad) and tank, which was then filled with 1X Laemmli buffer. Before loading, the samples were boiled at 95°C for 10 minutes. The peqGOLD Protein Marker IV (VWR Life Science) was used as a molecular weight ruler. The gel was run for 30- 90 minutes with a constant current of 17.5 mA per gel (Power Pac 3000, Bio-Rad).

In order to analyze the size-separated proteins with specific antibodies, they were transferred to a Nitrocellulose membrane (Amersham™ Protran™ 0.45µm NC, Nitrocellulose Blotting Membrane, GE Healthcare). The ""Mini Trans-Blot® Cell" from Bio-Rad was used for this step. The gel and membrane were assembled and held together by the "Mini Gel Holder Cassette" from Bio-Rad. All components of the blotting assembly were pre-incubated in Tank-Blot buffer (1X Laemmli buffer + 20% methanol) and stacked in the following order.

1. Black Plastic Grid
2. One foam pad
3. Two Whatman™ papers (3 MM CH, GE Healthcare)
4. Protein Separation Gel
5. Nitrocellulose membrane
6. Two Whatman™ papers (3 MM CH, GE Healthcare)
7. One foam pad
8. White Plastic Grid

Before adding the second foam pad and White Plastic Grid, the blotting assembly was pressed and rolled with a glass pipette to remove air bubbles. The assembled blotting cassette was loaded into the "Mini Trans-Blot Central Core" in the electrophoresis tank, taking care to orient the assembly according to the color-coded electrodes. The tank was filled with Tank-Blot buffer and run at a current of constant of 150mA, 100V for 15 minutes, followed by constant 300mA, 100V for 20 minutes was applied (Power Pac 3000, Bio-Rad).

Once removed from the blotting set up, the membrane was washed in 1X TBST (100 mM Tris-HCl (pH 8), 1.5 M NaCl, 0.5% Tween-20) and the success of the blotting step was analyzed by the transfer of the protein molecular weight marker. After washing, the membrane was blocked in 5% milk powder in TBST for 60 minutes on a shaker at room temperature. The primary antibody was then added (in 5% milk powder in 1X TBST) and incubated at 4°C on a shaker overnight. The next day, the primary antibody was removed, and the membrane was washed in 1X TBST 3 times for 10 minutes each at room temperature on a shaker. This was followed by a 60-minute incubation with the secondary antibody (in 5% milk powder in 1X TBST) at room temperature on a shaker. The blot was once again washed in 1X TBST 3 times for 10 minutes each at room temperature on a shaker and the molecular weight markers were marked with the WesternSure® Pen (Li-Cor). The membrane was transferred to a new plastic tray and developed by a 3-minute incubation with Pierce™ ECL Western Blotting Solution (Thermo Scientific). The blots were visualized using the Lumi-Imager™ F1 (INTAS). The primary and secondary

antibodies are detailed in Table 2.5.1.3 below.

Table 2.5.1. Antibodies used in this work.

Antibody	Type	Dilution	Distributor	Cat. No.
gt anti-gp120	Primary	1:5000	abcam	ab21179
rb anti-ERK2	Primary	1:1000	abcam	ab32081
ms anti-GAPDH	Primary	1:5000	abcam	ab8245
rb anti-p24	Primary	1:6000	abcam	ab32352
ms anti-SR proteins (1H4)	Primary	1:1000	Invitrogen	339400
ms anti-SR proteins (16H3)	Primary	1:1000	Life Technologies	339300
dn anti-goat	Secondary	1:2000	abcam	ab97110
gt anti-rabbit	Secondary	1:2000	Invitrogen	A27036
rb anti-mouse	Secondary	1:2000	Invitrogen	A27025

2.5.2 Luciferase activity

Cells that have been transfected with plasmids that contain a Renilla luciferase (*Renilla Reniformis*) or a Firefly luciferase (*Photinus Pyralis*) ORF were analyzed for luciferase activity in the presence of the respective substrates. To remove media residue, the cells were washed twice with PBS before being incubated in 2 mM EDTA in PBS for 2-5 minutes at room temperature to loosen adherent cells. The cells were pipetted into 1.5 mL tubes (SafeSeal, Starstedt) and shortly centrifuged. The supernatant was discarded, and the cell pellet was lysed in 200 µL 1X Lysis-Juice (PJK). The samples were vortexed and frozen at -80°C. The samples could either be stored at -80°C for later analysis or thawed after 20 minutes for immediate analysis.

To measure Luciferase activity, substrates were prepared. For the Renilla luciferase assay, Renilla substrate (coelenterazine in reconstitution buffer, PJK, stored at -80°C) was mixed with Renilla-Juice reaction buffer PJK, stored at 4°C) at a ratio of 1:50. This mixture must be made fresh. For the Firefly luciferase assay the firefly substrate (D-Luciferin and ATP, PJK) were dissolved in the Beetle-Juice reaction buffer (PJK, stored at 4°C) in the volume necessary. This mixture can be stored at -80°C and thawed when necessary. Both substrate mixtures must be protected from light.

The luciferase activity was measured in white flat-bottomed 96 well plates (Nunc™, Thermo Fisher). 20 µL of the sample was pipetted into the wells and each sample was analyzed in quadruplicate. The measurement was performed by the Tecan Infinite® 200 machine and the i-control 1.12 software, where the injection volume was 100 µL and the integration time was 10000 ms. The light emitted by the luciferase activity was measured in relative light units (RLU).

2.5.3 Syncytia assay

Env, the viral glycoprotein of HIV-1, is capable of inducing membrane fusion when expressed in cells that express the membrane protein CD4, resulting in syncytia formation. This mechanism relies on Env to be correctly folded and incorporated into the cell membrane. CD4-expressing adherent cells that had been transfected with env expression plasmids were analyzed for syncytia formation, in conjunction with a β-galactosidase assay, which results in the production of an insoluble blue color to aid in imaging the cells. β-galactosidase is stably transfected in the TZM-bL cell line under an LTR promoter, making the production of β-galactosidase dependent on the presence of the HIV-1 encoded protein Tat. Tat is co-transfected in all conditions either in an pMCS IRES Tat plasmid for the Rev-free conditions or in a pRev IRES Tat plasmid for the Rev containing conditions. Due to the nature of transient double transfection, the delivery of Tat to individual cells is not even and results in varying levels of β-galactosidase induction, which causes inconsistent intensity of the blue dye. The cells were washed twice with 1 mL PBS and fixed by a 10-minute incubation with 500 µL pre-chilled Fixing Solution (0.25% glutaraldehyde, 0.8% formaldehyde in PBS) at 4°C. The fixing solution was removed, and the cells were again washed twice with 1 mL PBS. Then, the cells were incubated with 500 µL of staining solution (0.4 mg/mL X-gal, 4 mM K3[Fe(CN)6], 4 mM K4[Fe(CN)6], 2 mM MgCl2 in PBS) at 37°C for a minimum of 4 hours. The stained cells were analyzed via microscope, and ten random fields per condition were imaged. The number of nuclei per syncytium in each image was counted and averaged with other images from the same condition, in order to quantify the size and number of syncytia produced.

2.5.4 Enzyme-Linked Immunosorbent Assay (ELISA)

The ELISAs can be used to measure proteins in a sample as a quantitative alternative to the Western blot. The typical protein used to detect HIV-1 infection is the capsid protein, p24, which is produced by the Gag ORF. This capsid protein is present in both virions as well as infected cells. It can also be produced by cells transfected with a Gag expression vector. In order to prepare viral stock for analysis via ELISA, 100 µL of viral stock was mixed with 300 µL of 0.1% Triton-X100 in PBS for the p24 ELISA, and 400 µL of 0.1% Triton-X100 in PBS for the Architect analysis (see 2.5.4.2) and vortexed. To harvest cells expressing p24, the cells were washed twice with PBS before being incubated in 2 mM EDTA in PBS for 2-5 minutes at room temperature to loosen adherent cells. The cells were pipetted into 1.5 mL tubes (SafeSeal, Starstedt) and shortly centrifuged. The supernatant was discarded, and the cell pellet was lysed in 400 µL of 0.1% Triton-X100 in PBS for the p24 ELISA, and 500 µL of 0.1% Triton-X100 in PBS for the Architect analysis and vortexed.

2.5.4.1 p24 ELISA

The p24 plate ELISA was carried out using the HIV-1 Gag p24 DuoSet ELISA kit from Bio-techne according to the instruction included with the kit. Additionally, the Ancillary kit from Bio-techne was used to provide plates, p24 Wash Buffer, and Reagent Diluent.

2.5.4.2 Architect p24 Analysis

The Abbot Architect i2000SR is a routine diagnostics analyzer developed for medical diagnostics. The HIV-1 Ag/Ab assays detect the anti-HIV antibody response against the transmembrane protein and presence of p24 antigen in the sample. The samples were pipetted into Architect analysis tubes, and any bubbles were removed. The samples were scanned, labelled, and HIV-1 Ag/Ab analysis was performed. The resulting data was expressed as signal before cutoff (S/CO) and used for comparative analysis between individual samples.

3 Results

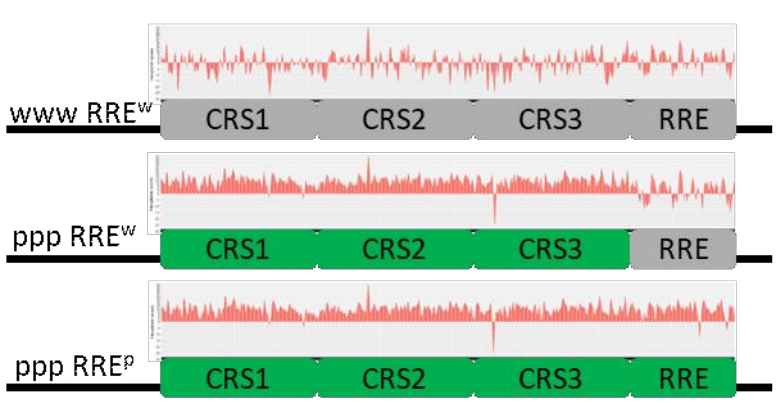
3.1 HEXplorer-guided mutation results in Rev-independent nuclear export of Env transcripts

This work investigates sequences that regulate nuclear export patterns in late phase HIV-1 mRNAs and how the bioinformatically guided alteration of these sequences via HEXplorer-guided mutation can result in the alteration of nuclear export pathway as well as remove dependence on the virally encoded nuclear export factor, Rev. This work was based on the HEXplorer algorithms' ability to predict the presence of splicing regulatory elements (SREs) that interact with RNA binding proteins, such as SR proteins (Erkelenz, Theiss et al. 2014). These SR proteins, in addition to regulating splice site selection, have also been linked to NXF1-mediated nuclear export of mRNAs (Müller-McNicoll, Botti et al. 2016). SR proteins play extensive roles in the HIV-1 life cycle and have been shown to regulate transcription, splicing, translation, and potentially nuclear export of viral transcripts, among other mechanisms (Mahiet and Swanson 2016). Despite extensive investigation of the HIV-1 life cycle, the molecular mechanism underlying the nuclear retention of unspliced gag, gag/pol, and intron-containing HIV-1 *env* mRNAs in the absence of Rev is still unknown. Although the presence of recognized but unused splice sites retains mRNAs in the nucleus under certain conditions (Mikaélian, Krieg et al. 1996), they do not fully account for the retention of HIV-1 late expression transcripts. Other sequences in the HIV-1 gag, pol, and env genes that lead to the inhibition of gene expression are labeled as INS (instability) or CRS (cis-acting repressor sequences) elements (Mikaélian, Krieg et al. 1996). Analysis of such elements in the HIV-1 genome was done by applying the HEXplorer algorithm originally developed to predict the location of potential splicing regulatory elements that offer binding sites for splicing regulatory proteins. Since HEXplorer predicts the location of regulatory sequences that offer binding sites for splicing regulatory proteins involved in the regulation of both splicing and nuclear export, we hypothesized that HEXplorer may be used to identify novel motifs and further characterize previously described INS/CRS elements by predicting sequences that offer binding sites for certain regulatory proteins associated with nuclear retention.

First, to test whether HEXplorer-guided mutations could alter nuclear export patterns of CRS containing HIV-1 mRNAs, previously described CRS found in the HIV-1 open reading frame were analyzed via HEXplorer and found to have an average HEXplorer score of -0.46 per nucleotide, compared to the -0.1 to 1.0 of the regions not described as CRS. The previously described CRSs begin 226 nt downstream of the *env* start codon and end 1,494 nt downstream of the *env* start codon, directly upstream of the RRE located between nucleotides 1,494 and 1,806 nt. As the RRE participates in the nuclear export of 4 kb and 9 kb class transcripts, it was included in the region selected for mutational analysis of

putative CRS. Subgenomic Env expression vectors were created by placing the *env* ORF into a vector containing an LTR promoter and BGH poly A site without the HIV-1 splice donor (SD4) found upstream of the *env* start codon to prevent splicing. The CRS region from 226 to 1,494 nt was separated into three separate sub-regions, which are referred to as CRS1, CRS2, and CRS3. These regions were mutated to obtain the most positive HEXplorer score possible, optimizing the sequences potential to bind SR proteins. The previously published algorithm ModCon (Ptok, Müller et al. 2021) was used to generate sequences maintaining the amino acid sequence, but resulting in generating the highest HEXplorer score possible, referred to the HEXplorer-optimized sequences. Figure 3.1.1.A shows the schematic of the Env expression vectors as well as the newly generated mutants, indicating the wildtype sequence in grey and the HEXplorer-optimized sequences in green.

A



B

		Per Nucleotide		Total	
Sequence	Length (in Nucleotides)	Wildtype	Positive	Wildtype	Positive
CRS 1	434	-1.56	9.49	-678.83	4129.01
CRS 2	444	-0.32	10.38	-139.84	4606.91
CRS 3	390	-1.72	9.30	-672.67	3627.16
RRE	312	2.30	9.44	718.64	2936.11

Figure 3.1.1. Cis-acting repressive sequences found in HIV-1 env are mutated according to HEXplorer algorithm predictions. These CRS were mutated to be as HEXplorer positive as possible without altering the underlying amino acid sequence. A) The Env mutants along with their HEXplorer plot, where green indicates HEXplorer optimized, and grey indicates the wildtype sequence. Mutants are referred to the name on the left which indicates positive fragments with a (p) and wildtype with a (w) and a graph of the HEXplorer score is shown above the mutated region. B) The table shows the length of each region and the HEXplorer scores of both the wildtype and mutated sequences.

In Figure 3.1.1.B the wildtype and optimized HEXplorer score of each of the CRS sub-regions are displayed as both a total score, and a per nucleotide score (per nt), along with the length of each of these regions in nucleotides. All three CRS sub-regions are between 390 to 444 nucleotides long. However, the longest region, CRS2, at 444 nucleotides long,

has the least HEXplorer negative wildtype sequence per nt at -0.32, but the most positive HEXplorer optimized sequence at 10.38 per nt. Additionally, the length of the CRS subregion does not always correlate to HEXplorer score. Figure 3.1.1.B shows the length and HEXplorer score of each CRS sub-region. Both mutants in Figure 3.1.1 contain the HEXplorer optimized CRS1, CRS2, and CRS3 sub-regions, however one mutant retains the wildtype Rev responsive element (RRE) while the other contains a HEXplorer-optimized RRE, referred to as RREpos. The mutation of the RRE to create RREpos is predicted to alter the secondary structure and is consequently unlikely to interact with Rev. RREwt already has a positive HEXplorer score at 2.30 per nt, though RREpos is more positive at 9.44 per nt.

The wildtype HIV-1 RRE recruits the virally encoded Rev protein based on its stem loop IIb, i.e. the Rev binding site (Rausch and Le Grice 2015), which mediates the recruitment of nuclear export factor CRM1. Disruption of stem loop IIb inhibits Rev interaction with the RRE, and thus Rev-mediated nuclear export. Therefore, to analyze the ability of RREpos to recruit the nuclear export factor Rev, RREpos was compared to the wildtype RRE using UNAFold (Zuker 2003, Markham and Zuker 2008). Figure 3.1.2.A shows secondary structure folding patterns predicted by UNAFold. The colors of the nucleotides in the predicted secondary structures indicate the probability of correct base pairing, with red indicating >99.9% probability of correct base pairing. In the RREwt secondary structure prediction, the black arrow is pointing to the expected, well-studied RRE Stem loop IIb secondary structure. In UNAfold, these nucleotides are indicated mostly in red, indicating a high probability that this is the correct base pairing. In addition to lacking the typical wildtype RRE secondary structure, the RREpos has fewer red nucleotides which predicts a lower possibility of base pairing in the predicted secondary structure compared to RREwt.

As the prediction from UNAfold indicated that the RRE secondary structure is severely altered in RREpos, RREpos's inability to facilitate nuclear export in the presence of Rev was confirmed using Renilla luciferase expression vectors that are dependent on an RRE for expression. These expression vectors were created with RREwt, RREpos, and RRE0, a mutant of the RREwt that contains point mutations between position 1,587 and 1,596 nt within stem loop IIb, which inhibits Rev binding, and thus Rev-mediated nuclear export without altering the underlying amino acid sequence as used in (Schaal, Klein et al. 1995). HeLa cells were transfected with the RRE-dependent luciferase expression vectors at a 1:1 ratio with either a mock plasmid or a Rev expression vector and then cellular protein samples analyzed for luciferase production to determine export efficiency of each RRE version.

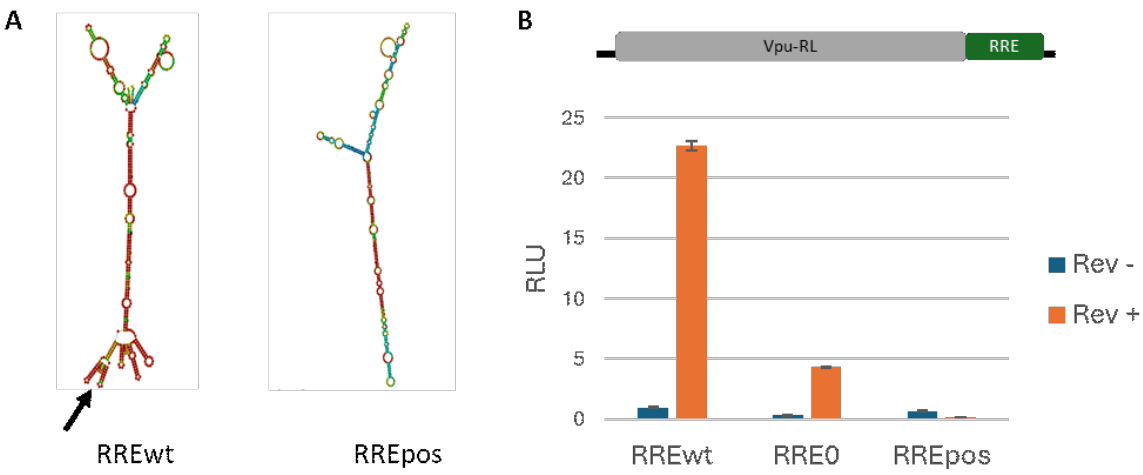


Figure 3.1.2. Mutated RRE sequences do not facilitate Rev-mediated export. HeLa cells were seeded 24 hours prior to transfection at 1.5 x 10^5 cells per mL. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The cells were harvested using 200 µL Lysis juice per mL cell culture 24 hours after transfection and 20 µL of sample were analyzed using a Tecan plate reader in quadruplicate. A) HEXplorer guided mutation of the Rev Responsive Element (RRE) alters secondary structure and thus the ability to support export via Rev-binding. The secondary structure of the wildtype and HEXplorer positive RRE were analyzed by UNAfold. The following colors correspond to probabilities of correct base pairing: black: <0.01, magenta: 0.01–0.10, blue: 0.10–0.35, cyan: 0.35–0.65, green: 0.65–0.90, yellow: 0.90–0.99, orange: 0.99–0.999, and red: > 0.999. The black arrow points to the stem loop (II), the initial interaction site of Rev. B) Luciferase expression vectors that require an export element for luciferase production were created with various export elements derived from HIV-1. HeLa cells were transfected with one of three expression plasmids containing a wildtype RRE (RREwt), an RRE that contains point mutations in the stem loop (II) that minimizes Rev binding (RRE0), and a HEXplorer optimized RRE (RREpos). These expression plasmids were transfected with either a mock expression vector or a Rev expression vector and analyzed for luciferase activity 24 hours after transfection.

In figure 3.1.2.B, all the luciferase expression vectors produce little to no luciferase in the absence of Rev. The RREwt construct showed a significant increase in luciferase production in the presence of Rev, while the RRE0 construct has a much smaller and more modest increase in luciferase production in the presence of Rev. This observation clearly shows that the point mutations in stem loop IIb of the RRE secondary structure limits Rev binding, and thus Rev-mediated export. The RREpos on the other hand produces less luciferase in the presence of Rev than without. Taken together with the disruption of the predicted secondary structure, the lack of luciferase in the presence of Rev from the RREpos construct indicates that RREpos is unable to bind Rev to facilitate nuclear exports of RRE-dependent mRNAs.

After confirming that the RREpos sequence and thus mutants containing RREpos do not facilitate nuclear export through Rev binding, the Env expression vectors detailed in 3.1.1.A were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via FISH to determine the subcellular localization of the transcripts in the presence and absence of Rev. The cells were fixed in 3.7% Formaldehyde and

incubated overnight at 37°C with Cy3 fluorescently labeled RNA probes complementary to the unmutated regions of the Env gene. The nuclei of the cells were stained with DAPI and mounted on glass microscope slides using Vectashield. Twenty cells from each condition from two separate assays were imaged and analyzed to determine the percentage of fluorescent Env mRNA found in the cytoplasm compared to the nucleus.

In figure 3.1.3.A, the wildtype *env* transcripts co-localize with the DAPI nuclear stain in the absence of Rev, showing that they are restricted to the nucleus. In contrast, the HEXplorer-optimized Env mutants show fluorescent *env* signal in the cytoplasm regardless of whether Rev is present or not, showing that nuclear export of these mutants is no longer Rev-dependent. Figure 3.1.3.B shows that in the presence of Rev, the percentage of cytoplasmic Env transcript increases by 3.38-fold. In contrast to the wildtype, the HEXplorer positive *env* mutants are found in the cytoplasm regardless of whether Rev is present or not, and the percentage of cytoplasmic Env transcript increases by 1.28-fold for the $pppR^{w}$ mutant and 1.01-fold for the $pppR^{p}$ mutant. The minimal change in cytoplasmic fluorescence in the HEXplorer-optimized mutants accompanied by the images in 3.1.3.A showing signal for *env* mRNA in the cytoplasm both with and without Rev indicates that the HEXplorer-optimized Env mutants can be exported without the assistance of Rev. Additionally, the minimal change in cytoplasmic fluorescence of $pppR^{w}$ and $pppR^{p}$ indicate that mutating RREwt to RREpos in the presence of the HEXplorer-optimized CRS has very little effect on the nuclear export pattern, and that the change in nuclear export pattern is primarily due to the change in the HEXplorer score of the CRS.

Despite not altering the amino acid sequence, changes in the RNA binding profile can have a major effect on the fate of the mRNA. To determine whether HEXplorer optimization of the CRS within the open reading frame of HIV-1 *env* is able to produce protein, HeLa cells were transfected with the Env mutants and analyzed via western blot. The samples were then analyzed for the presence of gp120, a protein produced by the *env* transcript, and the housekeeper gene ERK2 via western blot. The signal from the western blot was quantified and normalized to $wwwR^{w}$ + Rev. The results from three separate biological replicates were averaged and used to produce figure 3.1.4.B.

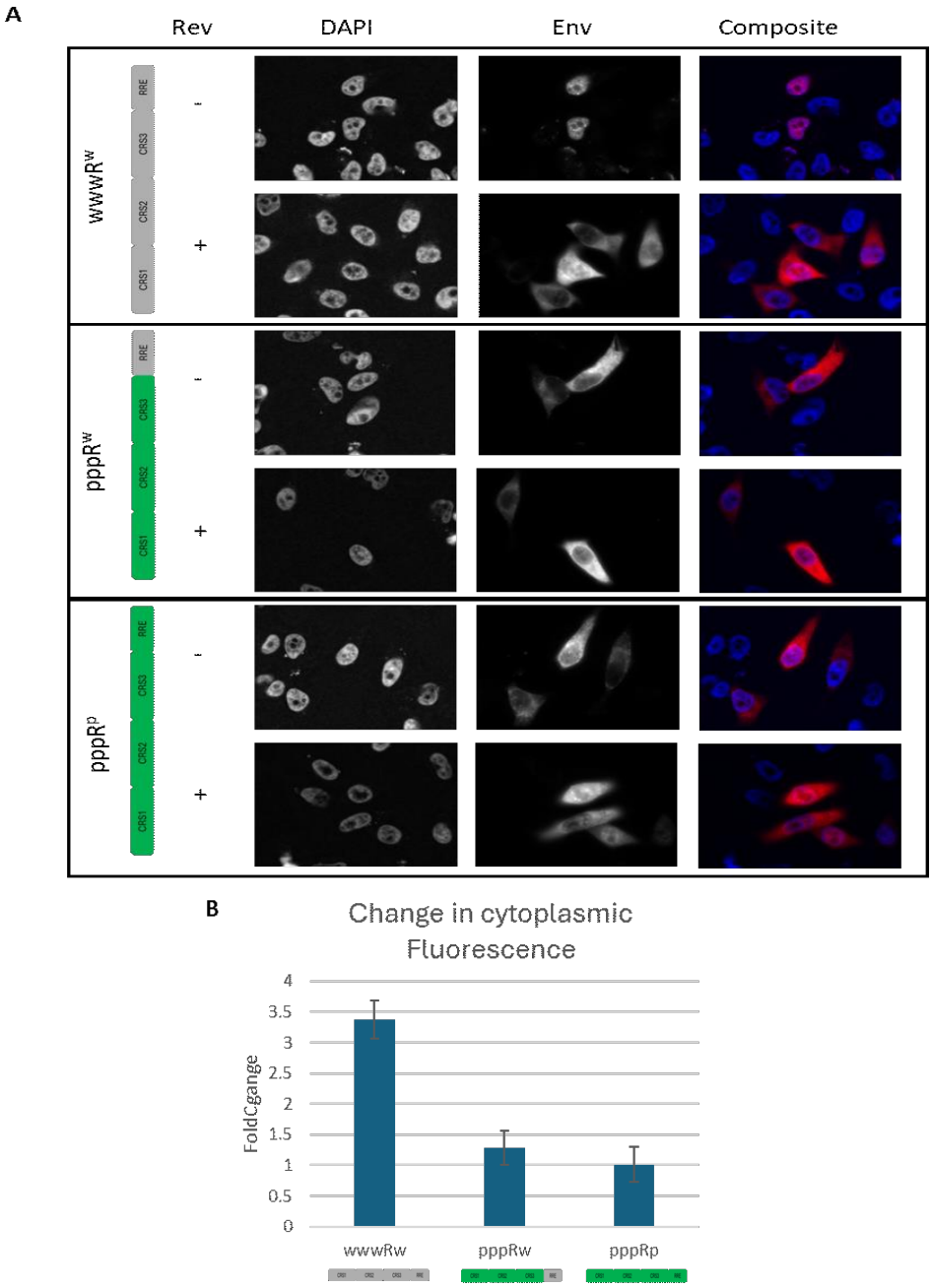


Figure 3.1.3. The HEXplorer positive mutants are exported from the nucleus in the absence of Rev. HeLa cells were seeded 24 hours prior to transfection at 1.0 x 10^5 cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were co-transfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested 24 hours after transfection. After incubating overnight, the cells were fixed and incubated with custom Env specific fluorescent RNA probes from Stellaris and imaged. A) The Rev-dependent wildtype Env is retained in the nucleus in the absence of Rev, showing a dependency on Rev for nuclear export. The HEXplorer positive Env expression transcripts are found in the cytoplasm with and without Rev, indicating that nuclear export is not Rev dependent. B) The percentage of cytoplasmic fluorescence was calculated for a total of 20 cells per condition from two separate assays, and normalized according to the Rev-free condition to show the change in cytoplasmic fluorescence upon the addition of Rev.

In figure 3.1.4, the HEXplorer-optimized Env mutants produced protein in the absence of Rev, however they produce less than the wildtype in the presence of Rev. Figure 3.1.4.A clearly shows that wwwR^{w} is dependent on Rev for protein production as transcripts retained in the nucleus in the absence of Rev cannot be translated. In contrast, the Rev-independent mutants (pppR^{w} and pppR^{p}) produce gp120 both in the presence and absence of Rev, indicating that they are not dependent on Rev for export to the cytoplasm. Figure 3.1.4.B shows that in the absence of Rev, pppR^{w} produces 47.9% of protein that wwwR^{w} does in the presence of Rev, and the quantity decreases even further to 34.9% in the presence of Rev. This decrease in protein production occurs to a greater extent in the pppR^{p} mutant, producing only 19.9% signal intensity in the absence of Rev, and 14.8% signal intensity in the presence of Rev relative to wwwR^{w} + Rev. Additionally, unlike the wwwR^{w} expression vector, the Rev-independent mutants (pppR^{w} and pppR^{p}) show slightly decreased gp120 production in the presence of Rev compared to the mock conditions, with pppR^{w} producing 13.0% less protein, and pppR^{p} producing 5.1% less protein.

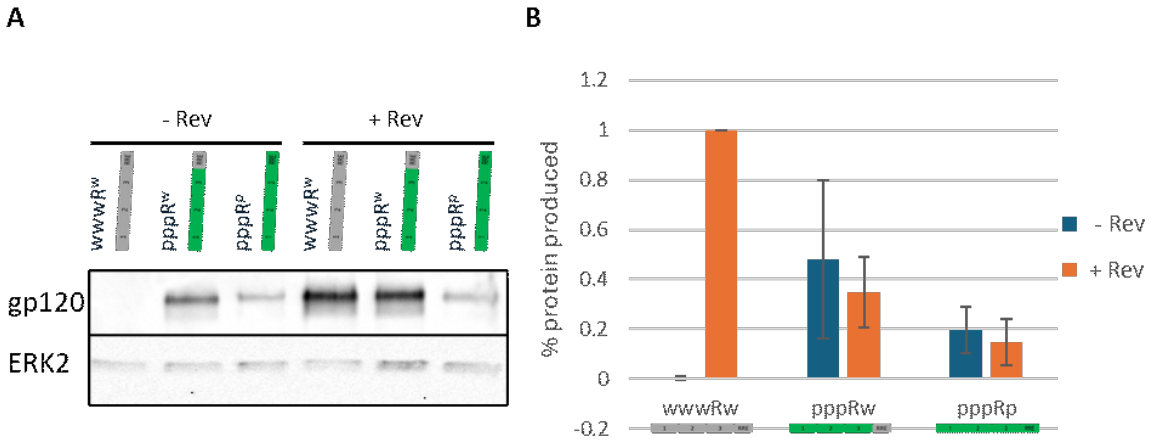


Figure 3.1.4. Rev-independent Env mutants produce protein in the absence of Rev. HeLa cells were seeded 24 hours prior to transfection at 1.5 x 10^{5} cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were co-transfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested 24 hours after transfection. The samples were lysed in RIPA buffer and analyzed via western blot. The signal from the western blot was quantified and normalized to wwwRw + Rev. The results shown in this figure were averaged from three separate biological replicates. A) western blot of Env Mutants. B) Change in gp120 levels from the Mock transfection. While all the Rev-independent Env mutants produce some protein in the absence of Rev, the wildtype Env increases greatly in the presence of Rev, showing a dependency on Rev for nuclear export. The Rev-independent mutants (pppRw and pppRp) produce similar quantities of gp120 regardless of the presence of Rev, consistent with their independence of Rev for nuclear export.

Env protein typically partially translocates to the rough endoplasmic reticulum lumen after translation and is transported to the cell membrane after maturation in the Golgi complex. Although these Env mutants still produce protein, albeit significantly less than the wildtype (wwwR^{w}) in the presence of Rev, this does not mean that the protein is successfully translocated to and incorporated in the cell membrane. To investigate if the

Env protein produced by these mutants can induce membrane fusion in cells expressing CD4, the TZMbL cell line, which expresses CD4, was transfected with the Env mutants and either a Mock plasmid or a Rev expression vector and analyzed for syncytium formation. After incubating overnight, the cells were fixed and incubated with β-galactosidase substrate and imaged. The β-galactosidase activity is independent of Env expression; however, the blue dye, which is dependent on the presence of Tat, aids in visualizing syncytia. Tat is encoded in both the Mock plasmid and Rev expression vector under an IRES. The intensity of the β-galactosidase dye is independent of the Env activity and may vary between syncytia.

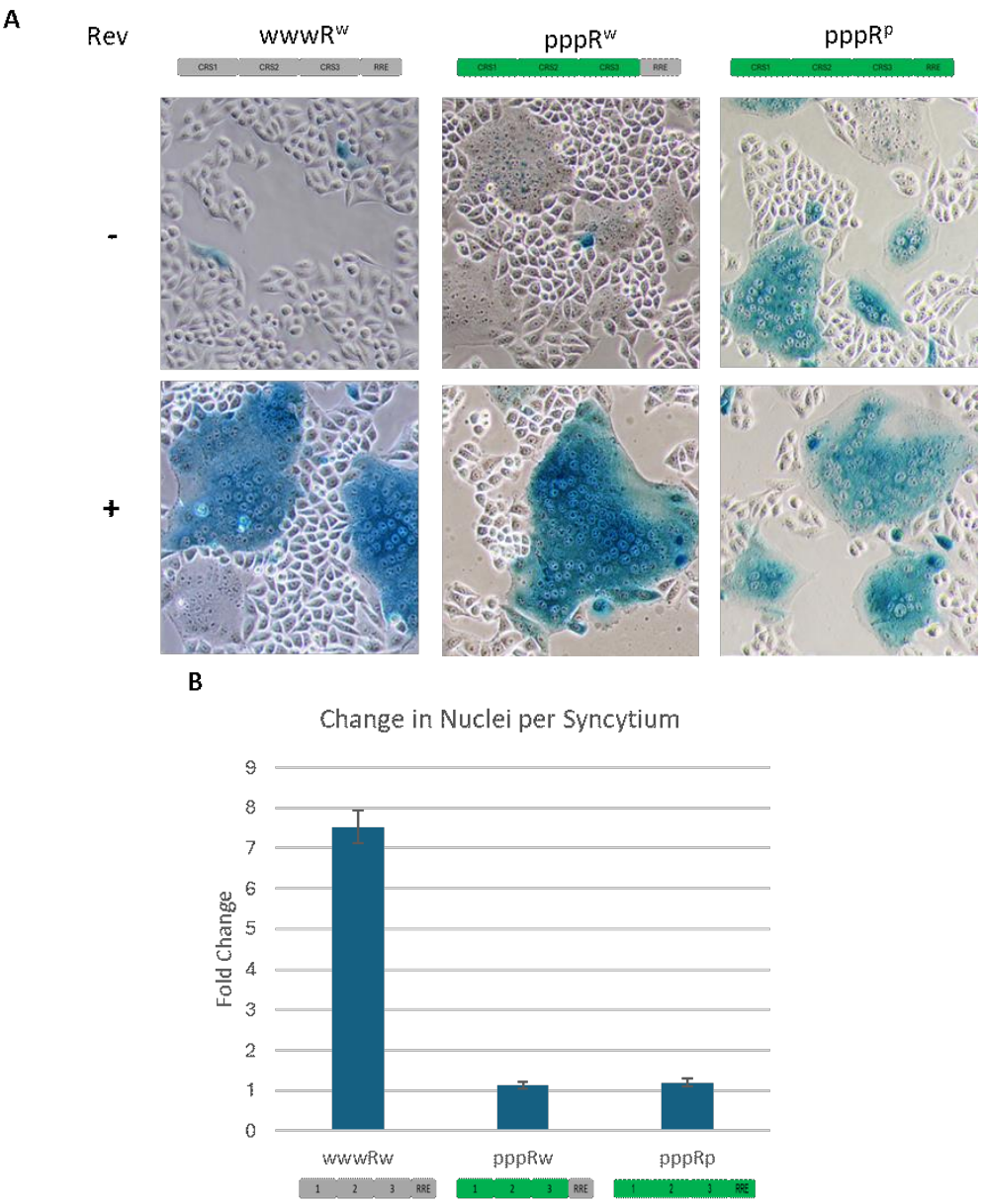


Figure 3.1.5. The Rev-independent Env mutants induce membrane fusion in CD4+ cells in the absence of Rev. TZMbL cells were seeded 24 hours prior to transfection at 2.0 x 10^5 cells per mL. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were co-transfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and fixed with 3.7% formaldehyde and incubated with β-galactosidase substrate 24 hours after transfection. Ten fields per assay from two biologically separate assays were imaged from each condition and the nuclei per syncytium counted and quantified. A) The wildtype Env shows a large increase in syncytium formation with the addition of Rev, showing a dependency on the Rev-RRE system for nuclear export. The Rev-independent mutants (pppRw and pppRp) produce large syncytia in the absence of Rev, indicating that they do not rely on Rev for nuclear export. B) The number of nuclei per syncytium was counted for ten random fields in two separate assays. It was then divided by the values seen in the absence of Rev to show the change in nuclei per syncytium for each condition.

Figure 3.1.5.A shows that the wwwR^{w} Env expression vector is only able to produce syncytia in the presence of Rev, while the Rev-independent mutants (pppR^{w} and pppR^{p}) can produce large syncytia in both the presence and absence of Rev. Interestingly, the conditions lacking Rev show less β-galactosidase activity compared to the conditions with Rev. In Figure 3.1.5.B, the wwwR^{w} shows a 7.9-fold increase in syncytium formation with the addition of Rev, showing a dependency on the Rev-RRE system for nuclear export. The Rev-independent mutants (pppR^{w} and pppR^{p}) have little to no change in nuclei per syncytium when Rev is present with a fold change of 1.13 and 1.19 respectively. Taken together with figure 3.1.5.A, this data indicates that HEXplorer optimization of the large CRS found in HIV-1 Env does not affect the maturation and function of the Env protein produced by the pppR^{w} and pppR^{p} expression vectors.

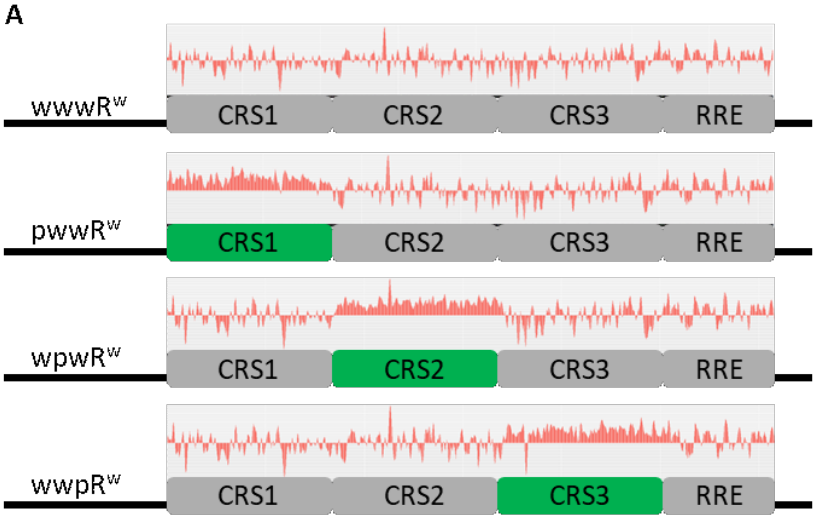
Overall, this chapter has shown that mRNAs produced by the HEXplorer optimized Env expression vectors are exported to the cytoplasm and produce functional Env protein even in the absence of Rev. Additionally, the presence of Rev seemed to result in a modest decrease of protein quantity in the pppR^{w} and pppR^{p} mutants, though the presence of Rev had little to no effect on the change in cytoplasmic fluorescence or the ability of these mutants to produce syncytia. The HEXplorer optimized mutants pppR^{w} and pppR^{p} are exported to the cytoplasm independent of the Rev-RRE export system, however, this was achieved through the mutation of 1,268 nt and 1,580 nt respectively. It is possible that the alteration in the nuclear export pattern is due to a specific region and not reliant on the entirety of the CRS region.

3.2 Rev independence is dependent on quantity of HEXplorer optimized region

In chapter 3.1, Env expression vectors in which the entire CRS region within the intron of the Env open reading frame were HEXplorer optimized displayed export to the cytoplasm in the absence of Rev. This CRS region is very large at 1,268 nucleotides long. Therefore, it is possible that rather than one large repressive sequence, there is a shorter sequence retaining the Env transcript in the nucleus contained within this CRS. To identify and further define this sequence, the HEXplorer optimized CRS1, CRS2, and CRS3 subregions were individually cloned into the wildtype Env expression vector to produce mutants with only a single HEXplorer optimized sub-region. Figure 3.2.1.A shows a schematic of the single positive Env mutants along with the HEXplorer score profile. These mutants were designed with the wildtype RRE to minimize the extent of the mutated region. In Figure 3.2.1.B, the HEXplorer score of the overall CRS region and RRE per nt, and total are shown. Each of the single HEXplorer optimized sub-regions has a similar HEXplorer score. The HEXplorer score of the analyzed region increases from -0.46 to 2.70 per nt when only CRS1 is HEXplorer optimized, which is the largest increase in HEXplorer score among the single positive mutants, even though CRS1 does not have the largest mutated region. The

overall HEXplorer score of the analyzed region increases from -0.46 to 2.62 per nt when only CRS2 is optimized, and up to 2.55 per nt when only CRS3 is optimized. This indicates that while the quantity of the HEXplorer region plays a role in nuclear export, specific sequences or even individual SR proteins may play a larger role in nuclear export than others.

A



B

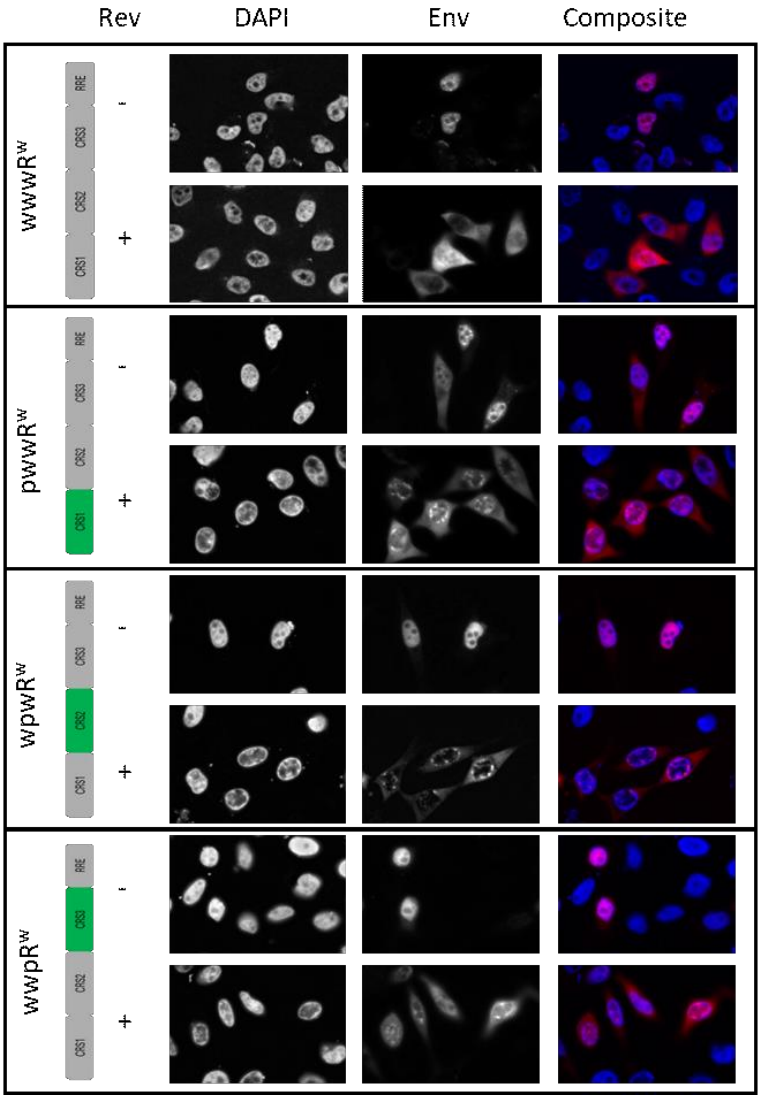
Mutant	Per/nt	Total
wwwR^{w}	-0.46	-719.07
pwwR^{w}	2.70	4188.88
wpwR^{w}	2.62	4064.82
wwpR^{w}	2.55	3952.82

Figure 3.2.1. Single cis-acting repressive sequence subregions found in HIV-1 env are mutated according to HEXplorer algorithm predictions. The individual CRS elements were mutated to be as HEXplorer positive as possible without altering the underlying amino acid sequence and interchanged to produce Env mutants with varying HEXplorer plots and scores without altering the underlying amino acid sequence. A) Single HEXplorer optimized sub-region Env mutants along with their HEXplorer plot, green indicating HEXplorer optimized sequence, and grey indicates the wildtype sequence. Mutants are referred to the name on the left which indicates positive fragments with a (p) and wildtype with a (w). B) The table shows total and per nucleotide HEXplorer score of the overall CRS regions.

The Env expression vectors shown in figure 3.2.1.A were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via RNA FISH to determine the subcellular localization of the transcripts in the presence and absence of Rev. HeLa cells were transfected with the Env expression vectors and either a mock vector or a Rev expression vector. Samples were analyzed via RNA FISH for subcellular localization in the presence and absence of Rev. After incubating overnight, the cells were fixed and incubated with custom *env* specific fluorescent RNA probes from Stellaris, imaged, and analyzed to determine the percentage of fluorescent Env mRNA found in the cytoplasm compared to the nucleus using the Intensity Ratio Nuclei Cytoplasmic Tool in Image J (Intensity Ratio Nuclei Cytoplasm Tool, RRID:SCR_018573).

In 3.2.2.A, the wildtype Env transcripts are retained in the nucleus in the absence of Rev, and a strong signal appears in the cytoplasm in the presence of Rev. However, the pwwR^{w} mutant Env expression vector appears to have a weak cytoplasmic mRNA signal, despite most of the signal being nuclear. The cytoplasmic signal appears stronger than in the absence of Rev. Additionally, the pwwR^{w} transcript still shows a strong nuclear mRNA signal in the presence of Rev. This observation indicates that the pwwR^{w} mutant is exported at low levels in the absence of Rev, though the presence of Rev increases the amount of cytoplasmic RNA. The cytoplasmic fluorescence of the pwwR^{w} in both the presence and absence of Rev is less intense than that of the wildtype. In contrast, the other two single HEXplorer optimized sub-region *env* expression vectors appear to be primarily nuclear in the absence of Rev, with a definite increase in cytoplasmic fluorescence in the presence of Rev. This indicates that the wpwR^{w} and wwpR^{w} *env* mutants are likely exported in a Rev-dependent manner, as well as suggesting that it is not HEXplorer score alone contributing to Rev-independent nuclear export. Figure 3.2.2.B shows that the cytoplasmic fluorescence of the pwwR^{w} mutant increases by 1.21-fold. This low increase in cytoplasmic fluorescence is likely due to the low level of cytoplasmic *env* mRNA signal in the absence of Rev. The other single HEXplorer-optimized *env* expression vectors wpwR^{w} and wwpR^{w} show an increase in cytoplasmic fluorescence by 1.81 and 1.50-fold respectively, which is similar to the wildtype increase of 1.74-fold. Indicating that cytoplasmic fluorescence increases in the presence of Rev. Taken together with the images in 3.2.2.A, the data indicates that *env* mutants wpwR^{w} and wwpR^{w} are still dependent on Rev for nuclear export. Interestingly, *env* mutant pwwR^{w} appears to be exported at low levels in the absence of Rev. However, the increase in cytoplasmic fluorescence in the presence of Rev suggests that Rev contributes to nuclear export of this transcript.

A


Rev
DAPI
Env
Composite
wwwR^w
pwwR^w
wpwR^w
wwpR^w
RRE
CRS3
CRS2
CRS1
-
+

B

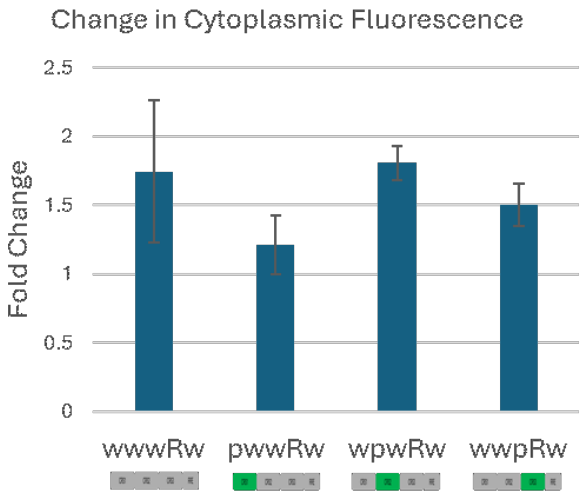
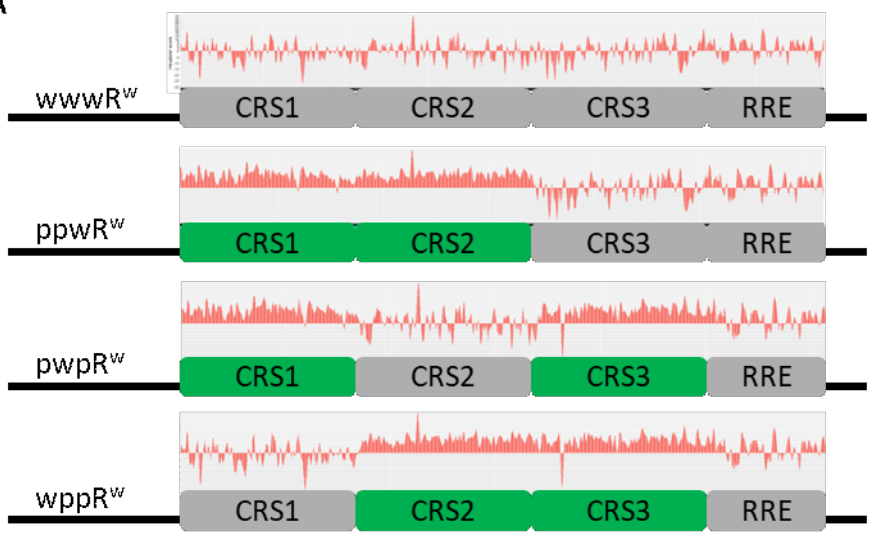

Change in Cytoplasmic Fluorescence
Fold Change
2.5
2
1.5
1
0.5
0
wwwRw
pwwRw
wpwRw
wwpRw

Figure 3.2.2. The single HEXplorer optimized mutants are not exported from the nucleus in the absence of Rev. HeLa cells were seeded 24 hours prior to transfection at 1.0 x 10^5 cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested 24 hours after transfection. The cells were fixed in 3.7% formaldehyde and incubated overnight at 37°C with Cy-3 fluorescently labeled RNA probes complementary to the unmutated regions of the Env gene. The nuclei of the cells were stained with DAPI and mounted to glass microscope slides using Vectashield mounting medium. Twenty cells from each condition from two separate assays were imaged and analyzed for cytoplasmic fluorescence. A) The Rev-dependent wildtype env is retained in the nucleus in the absence of Rev, showing a dependency on Rev for nuclear export. The single HEXplorer optimized env expression transcripts are primarily found in the nucleus in the absence of Rev, indicating that they are still exported in a Rev-dependent manner. B) The percentage of cytoplasmic fluorescence was calculated for a total of 20 cells per condition from two separate assays, compared to the Rev-free condition to show the change in cytoplasmic fluorescence upon the addition of Rev.

In Figure 3.2.2, two of the single HEXplorer optimized sub-region Env expression vectors appear to be dependent on Rev for nuclear export, wpwR^w and wwpR^w. The third single HEXplorer optimized Env expression vector pwwR^w showed a low level of nuclear export in the absence of Rev, however the low intensity of the cytoplasmic fluorescence shows that the Rev-independent export capacity is weak. Therefore, it is possible that increasing the quantity of HEXplorer optimized sequences in the transcript could lead to a stronger export signal in the absence of Rev. To this end, *env* expression vectors with two HEXplorer optimized subregions were created. Schematics of these expression vectors are shown in figure 3.2.3.A, along with the HEXplorer score profiles of these *env* expression vectors. For comparison to the single HEXplorer optimized subregion *env* expression vectors, the double HEXplorer optimized sub-region Env expression vectors contain the wildtype RRE sequence. In figure 3.2.3.B, HEXplorer score of the overall CRS region and RRE per nt, and total are shown for each double HEXplorer optimized expression vector. Each of the double HEXplorer optimized subregion vectors has a similar HEXplorer score. The HEXplorer score of the analyzed region increases from -0.46 to 5.73 per nt when CRS1 and CRS2 are HEXplorer optimized, which is the largest increase in HEXplorer score among the double HEXplorer optimized mutants, which correlates that it is also the longest mutated sequence among the double HEXplorer optimized mutants. The overall HEXplorer score of the analyzed region increases from -0.46 to 5.64 per nt when CRS1 and CRS3 are optimized, and up to 5.57 per nt when CRS2 and CRS3 are optimized.

A



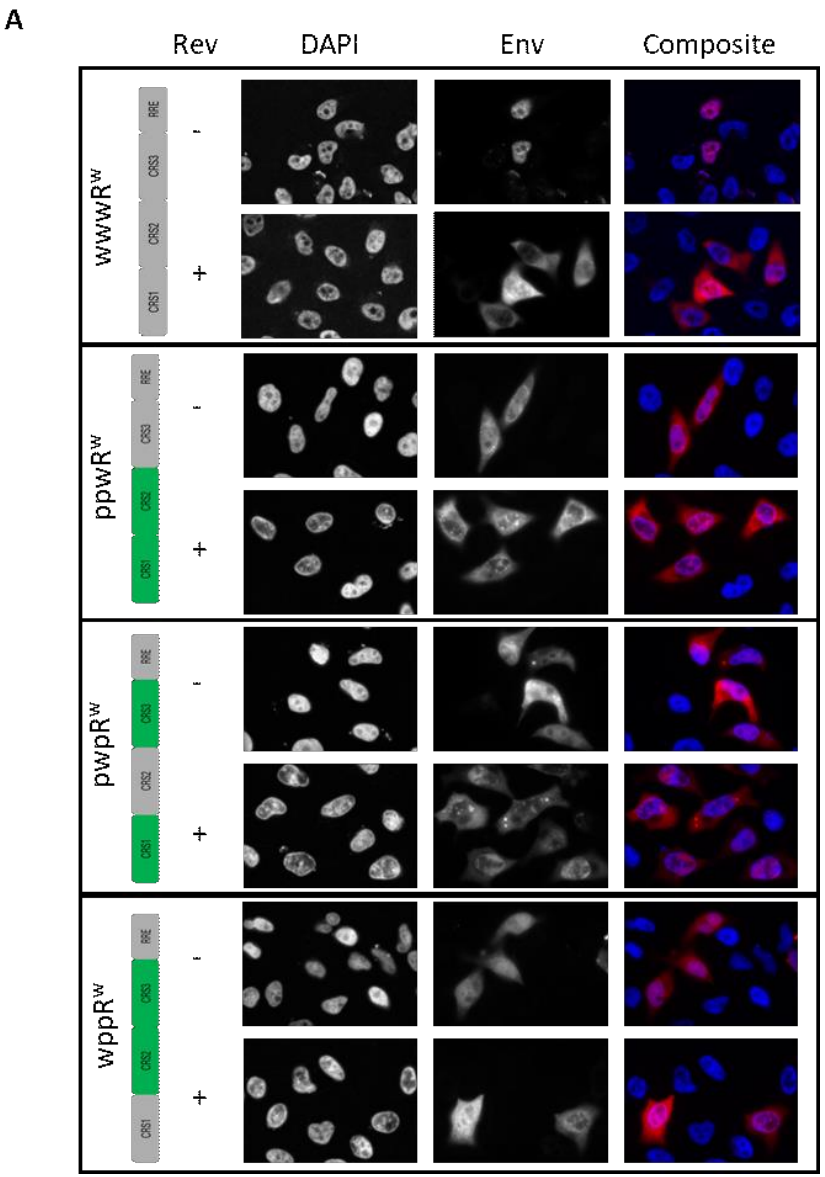
B

Mutant	Per/nt	Total
wwwR^{w}	-0.46	-719.07
ppwR^{w}	5.73	8859.88
pwpR^{w}	5.64	8752.45
wppR^{w}	5.57	8637.53

Figure 3.2.3. Two cis-acting repressive sequence sub-regions found in HIV-1 env are mutated according to HEXplorer algorithm predictions. These CRS were mutated to be as HEXplorer positive as possible without altering the underlying amino acid sequence and interchanged to produce env mutants with varying HEXplorer profiles and scores without altering the underlying amino acid sequence. A) The double HEXplorer optimized subregion env mutants along with their HEXplorer profile, where green indicates HEXplorer optimized, and grey indicates the wildtype sequence. Mutants are referred to by the name on the left which indicates positive fragments with a (p) and wildtype with a (w). B) The table shows total and per nucleotide HEXplorer score of the overall CRS regions.

The Env expression vectors shown in 3.2.3.A were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via FISH to determine the subcellular localization of the transcripts in the presence and absence of Rev. HeLa cells were transfected with the Env expression vectors and either a mock vector or a Rev expression vector. They were then analyzed via RNA FISH for nuclear export patterns in the presence and absence of Rev. After incubating overnight, the cells were fixed and incubated with custom Env specific fluorescent RNA probes from Stellaris, imaged, and analyzed to determine the percentage of fluorescent *env* mRNA found in the cytoplasm compared to the nucleus using the Intensity Ratio Nuclei Cytoplasmic Tool in Image J (Intensity Ratio Nuclei Cytoplasm Tool, RRID:SCR_018573).

In figure 3.2.4.A, the wildtype Env transcripts are retained in the nucleus in the absence of Rev, and a strong signal appears in the cytoplasm in the presence of Rev. The ppwR^{w} mutant shows a strong cytoplasmic mRNA signal in the absence of Rev, and this signal does not appear to be affected by the presence of Rev, continuing to show strong cytoplasmic mRNA signal. Similar effects are observed with the pwpR^{w} and wppR^{w} mutants, where cytoplasmic mRNA signal is clearly visible both in the absence and presence of Rev. The fluorescence of the pwpR^{w} mutant in both the presence and absence of Rev appears to be slightly weaker than the ppwR^{w} and wppR^{w} mutants. The presence of strong mRNA fluorescent signal in the cytoplasm of all three of the double HEXplorer optimized mutants both in the presence and absence of Rev indicates that these mutants are exported to the cytoplasm independent of the Rev-RRE export pathway. Figure 3.2.4.B shows that the wildtype *env* expression vector, in which the cytoplasmic fluorescence increases 1.74-fold upon the addition of Rev. The ppwR^{w} mutant has a fold change of 1.10, showing that there is very little change in cytoplasmic fluorescence. Likewise, the pwpR^{w} mutant has a fold change of 0.97, indicating that there is no change in cytoplasmic fluorescence. The wppR^{w} mutant however does show a 1.28-fold increase in cytoplasmic fluorescence in the presence of Rev, which could indicate that the RBPs that bind CRS1 contribute more to nuclear export regulation than those that bind the CRS2 and CRS3 regions. Taken together with the results shown in 3.2.4.A, the data indicates that the presence of Rev has little to no effect on the cytoplasmic fluorescence of the ppwR^{w} and pwpR^{w} double HEXplorer Env expression vectors, showing that these mutants are not reliant on Rev for export to the cytoplasm. The wppR^{w} shows clear cytoplasmic signal in the absence and presence of Rev, however the cytoplasmic signal increases by 1.28-fold in the presence of Rev. It is possible that this mutant does not reach its maximum export capability in the absence of Rev, thus the cytoplasmic signal is boosted in the presence of Rev through the wildtype RRE. However, the wppR^{w} mutant is exported to the cytoplasm in the absence of Rev, showing that it does not rely on Rev to be exported to the cytoplasm.



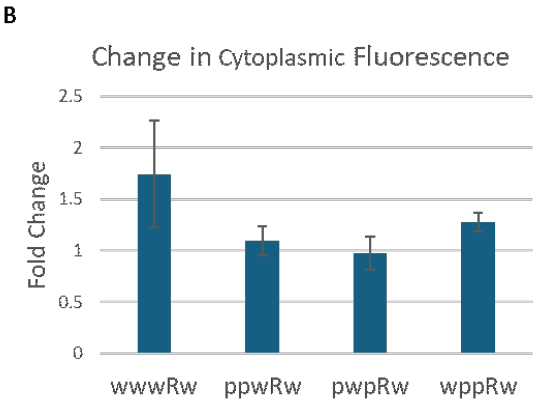


Figure 3.2.4. The double HEXplorer optimized mutants are exported from the nucleus in the absence of Rev. HeLa cells were seeded 24 hours prior to transfection at 1.0 x 10^5 cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested 24 hours after transfection. The cells were fixed in 3.7% Formaldehyde and incubated overnight at 37°C with Cy-3 fluorescently labeled RNA probes complementary to the unmutated regions of the Env gene. The nuclei of the cells were stained with DAPI and mounted to glass microscope slides using Vectashield mounting medium. Twenty cells from each condition from two separate assays were imaged. A) The Rev-dependent wildtype Env is retained in the nucleus in the absence of Rev, showing a dependency on Rev for nuclear export. The double HEXplorer optimized Env expression transcripts appear to be exported in the absence and presence of Rev, indicating that they are still exported in a Rev-independent manner. B) The percentage of cytoplasmic fluorescence was calculated for a total of 20 cells per condition from two separate assays, compared to the Rev-free condition to show the change in cytoplasmic fluorescence upon the addition of Rev.

As previously seen in figure 3.1.4, HEXplorer optimized transcripts such as $pppR^{w}$ and $pppR^{p}$ do not produce as much protein as the wildtype env expression vector in the presence of Rev. Therefore, to investigate the quantity of protein produced by the single and double HEXplorer optimized subregion Env expression vectors, HeLa cells were transfected with both single and double positive mutants with and without Rev and analyzed for Env protein production via western blot in Figure 3.2.5. The Env expression vectors shown in 3.2.2.A and 3.2.4.A were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via western blot to determine the relative quantity of protein produced in the presence and absence of Rev. The signal from the western blot was quantified via optical density measurement and normalized to wwwRw + Rev. The results from three separate biological replicates were averaged and are displayed in figure 3.2.5.B.

Figure 3.2.5.A shows that the wildtype Env expression vector produces no gp120 protein in the absence of Rev, but a large amount of protein in the presence of Rev. The single and double HEXplorer optimized subregion *env* expression vectors all produce gp120 protein in the absence of Rev, however several produce more protein in the presence of Rev. The single positive mutants ($pwwR^{w}$, $wpwR^{w}$, $wwpR^{w}$) and $pwpR^{w}$ mutant all show a large increase in gp120 production when Rev is introduced. The double positive $ppwR^{w}$ and $wppR^{w}$ mutants produce more protein than the single positive $pwwR^{w}$, $wpwR^{w}$, $wwpR^{w}$, and the double positive $pwpR^{w}$ mutants in the absence of Rev, but it appears to be less protein than the wildtype Env expression vector in the presence of Rev. Additionally, the $ppwR^{w}$, $wppR^{w}$ mutants appear to produce similar amounts of protein both in the absence and presence of Rev. Of the single HEXplorer optimized Env expression vectors, pwwRw produces the most gp120 in the absence of Rev, however $wwpR^{w}$ produces the most protein in the presence of Rev. In the double HEXplorer optimized subregion Env expression vectors, $pwpR^{w}$ in the presence of Rev produces the most protein, whereas the $ppwR^{w}$ mutant produces the most protein in the absence of Rev. Figure 3.2.5.B shows the quantification of three separate blots normalized to a percentage of the average protein quantity produced by the wildtype Env expression vector in the presence of Rev.

In figure 3.2.5.B, the quantity of protein produced was compared to the quantity of protein produced by the wildtype Env expression vector, where the wildtype produces 0.00% protein in the absence of Rev and 100% protein in the presence of Rev. The $pwwR^{w}$ produces 7% protein in the absence or Rev, but that increases by 44% to produce 51% protein compared to the wildtype Env expression vector with Rev. The other two single HEXplorer optimized Env expression vectors also produce more protein in the presence of Rev, with the $wpwR^{w}$ mutant protein production increasing from 4% to 29%, and the $wwpR^{w}$ mutant protein production increasing from 2% to 69% when Rev is present. The $wwpR^{w}$ mutant has the largest increase in protein production in the presence of Rev of

the single HEXplorer optimized Env expression vectors. The double HEXplorer optimized Env expression vectors have a more varied phenotype. Despite having *env* mRNA signal in the cytoplasm in figure 3.2.4.A, the amount of protein produced by the pwpR^w mutant increases 55% from 7% to 62%. The other two double HEXplorer optimized mutants, however, produce more similar amounts of protein in the presence and absence of Rev. The wppR^w mutant produces 12% of gp120 protein in the absence or Rev, and only increases by 9% to 21%. The ppwR^w mutant has a potentially small decrease in protein, with the average amount of protein dropping by 10% from 47% in the absence of Rev to 37% in the presence of Rev. Taken together, this data shows that the pwwR^w, wpwR^w, wwpR^w, and pwpR^w HEXplorer optimized Env expression vectors appear to produce significantly more protein in the presence of Rev, and thus are not efficiently exported in the absence of Rev. However, the ppwR^w and wppR^w HEXplorer optimized Env expression vectors showed little to no change in the presence of Rev compared to in the absence of Rev. In line with previous data shown in Figure 3.2.2, the single positive pwwR^w Env expression vector produces the most protein in the absence of Rev compared to the other single positive Env expression vectors, supporting the hypothesis that the RNA binding proteins binding the HEXplorer optimized CRS1 have a stronger nuclear export signal.

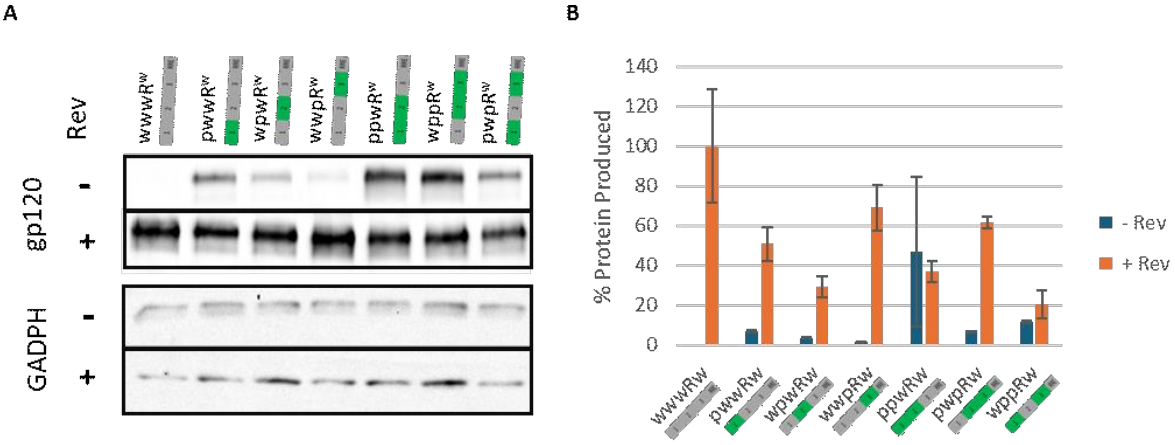


Figure 3.2.5. Rev-independent Env mutants produce protein in the absence of Rev. HeLa cells were seeded 24 hours prior to transfection at 1.5 x 10^5 cells per mL into wells. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested 24 hours after transfection. The samples were lysed in RIPA buffer and analyzed for the presence of gp120, a protein produced by the env transcript, and the housekeeper gene ERK2 via western blot. A) western blot of Env Mutants. B) The graph shows protein quantity from the Mock transfection. While all the Env mutants produce some protein in the absence of Rev, the wildtype and single HEXplorer optimized Env increases greatly in the presence of Rev, showing a dependency on Rev for nuclear export. Most of the double positive mutants produce similar quantities of protein, regardless of the presence of Rev, except for pwpRw.

To determine if the Env protein produced by these single and double HEXplorer optimized Env expression vectors are capable of inducing membrane fusion in permissive cells, TZMbL cells were transfected with the Env mutants with either a Mock or a Rev expression vector and analyzed for syncytium formation. After incubating overnight, the cells were fixed and incubated with β-galactosidase substrate and imaged. The β-galactosidase activity is independent of Env expression and instead dependent on Tat expression; however, the blue dye aids in visualizing the syncytium formation.

In Figure 3.2.6.A, the wildtype Env expression vector does not produce large syncytia in the absence of Rev, however in the presence of Rev, large syncytia form. The pwwR^{w} mutant appears to produce small syncytia in the absence of Rev, though there are large and extensive syncytium formation in the presence of Rev. The other two single HEXplorer optimized subregion Env expression vectors, wpwR^{w} and wwpR^{w} show no syncytium formation in the absence of Rev, and like the wildtype and pwwR^{w} mutant, large syncytia in the presence of Rev. This indicates that the single HEXplorer optimized env expression vectors rely on the presence of Rev to induce large syncytia. The double positive HEXplorer optimized subregion Env expression vectors, ppwR^{w}, pwpR^{w} and wppR^{w} all show large syncytia in the absence and presence of Rev. In the absence of Rev, the β-galactosidase activity is surprisingly consistently low, despite its expression not being dependent on Env expression. The ability of the double HEXplorer optimized Env mutants to induce syncytium formation regardless of the presence of Rev indicates that they do not rely on the Rev-RRE export pathway for export to the cytoplasm and subsequent expression. Figure 3.2.6.B shows the change in nuclei per syncytium in the presence of Rev. The wildtype Env expression vector produced the largest change in nuclei per syncytium at 12.60-fold more nuclei per syncytium in the presence of Rev. The single HEXplorer optimized Env expression vectors all show an increase in nuclei per syncytium over 2-fold, with pwwR^{w}, wpwR^{w}, and wwpR^{w} having a change in nuclei per syncytium of 2.79-fold, 4.49-fold, and 6.03-fold respectively. The double HEXplorer optimized Env expression vectors however show similar sized syncytia in both the absence and presence of Rev, expressed by a fold change close to 1. The nuclei per syncytium of the ppwR^{w} mutant increases by 1.16-fold, the pwpR^{w} mutant by 1.13-fold, and the wppR^{w} mutant by 1.23-fold. Taken together with figure 3.2.6.A, this data indicates that the wildtype and single HEXplorer optimized Env expression vectors are dependent on Rev for export and subsequent translation, while the double HEXplorer optimized Env expression vectors can produce sufficient functional protein to induce syncytium formation in the absence of Rev.

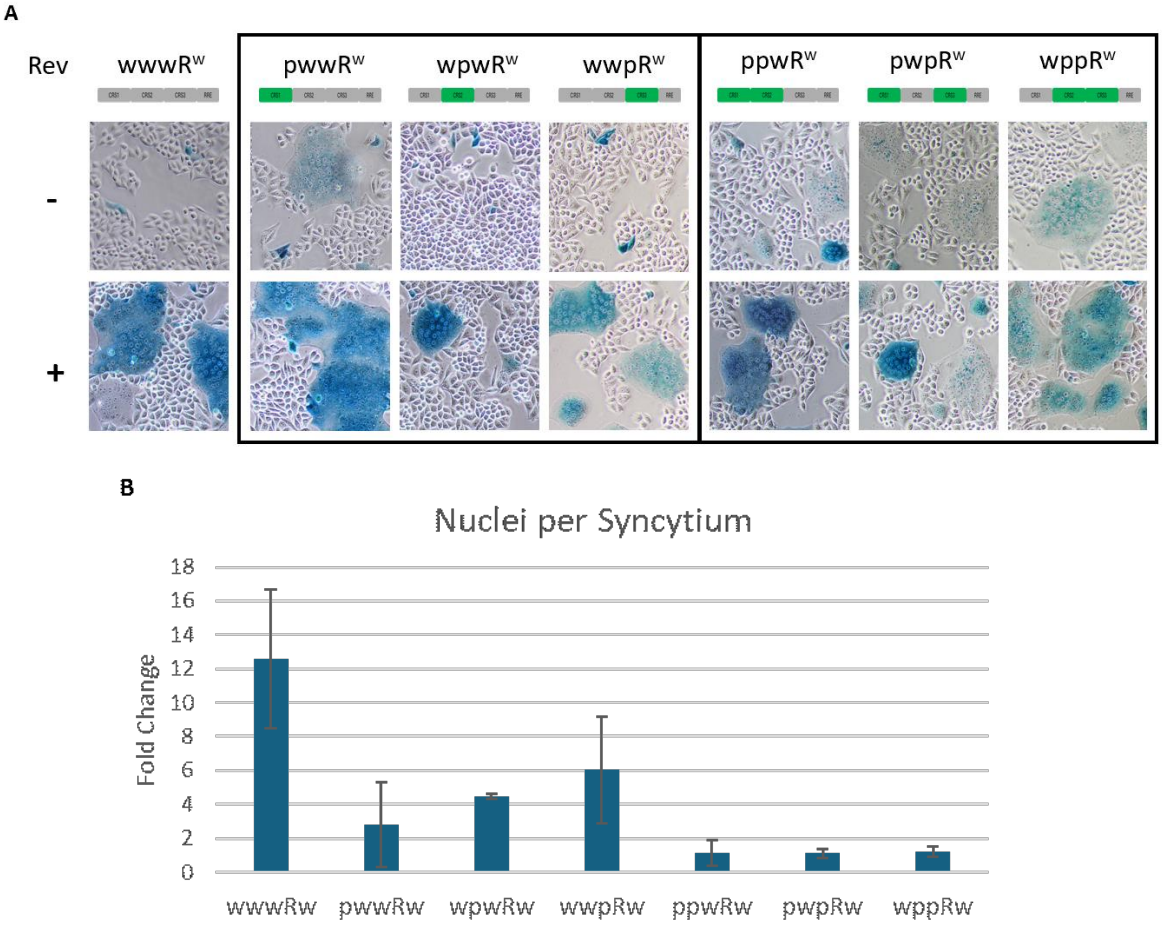


Figure 3.2.6. The Rev-independent Env mutants induce membrane fusion in CD4+ cells in the absence of Rev. TZMbL cells were seeded 24 hours prior to transfection at 2.0 x 10^5 cells per mL. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and incubated with β-galactosidase substrate 24 hours after transfection. Ten fields per assay from two biologically separate assays were imaged from each condition and the nuclei per syncytium were counted and quantified. A) The wildtype and single HEXplorer optimized Env show little to no syncytium formation in the absence of Rev, and a large increase in syncytia with the addition of Rev, showing a dependency on the Rev-RRE system for nuclear export. The double positive mutants produce large syncytia in the absence of Rev, indicating that they do not rely on Rev for nuclear export. B) The number of nuclei per syncytium was counted for twenty random fields from two separate assays. It was then normalized to absence of Rev to show the change in nuclei per syncytium for each condition.

While all the mutants produce functional Env protein that can produce syncytia in the presence of Rev, the introduction of regions rich in splicing regulatory elements has the potential to activate previously unused splice sites. To establish that the differences in expression and localization of transcripts produced by the HEXplorer optimized Env expression vectors is not due to alternative splicing of the transcripts, a northern blot and qPCR were performed. HeLa cells were transfected with all the HEXplorer optimized Env expression vectors with and without Rev and analyzed for RNA presence in figure 3.2.7.A and RNA quantity in figure 3.2.7.B. The Env expression vectors shown in 3.1.1A, 3.2.1.A, and 3.2.3.A were transfected into HeLa cells with either a mock plasmid or a Rev

expression vector to be analyzed via northern blot and RT-qPCR to determine the relative size and quantity of mRNA produced in the presence and absence of Rev. HeLa cells were transfected with the Env mutants in the presence of Rev. After 24 hours, RNA was harvested and analyzed via northern blot and RT-qPCR for plasmid-derived *env* transcript presence and quantity.

In figure 3.2.7.A, the mRNA signal produced by the subgenomic *env* expression vectors all appeared directly below the 28S ribosomal signal, which is typically 4-5 kb long. The position of the Env mutants relative to the 28S signal is consistent with the expected size of the HIV-1 4kb class transcripts. The HEXplorer positive mutants all appear to be the same length as the wildtype, as well as appear to be the expected length of *env* expression transcripts, eliminating the likelihood that the introduction of splicing regulatory element binding sites activated a previously unused splice site. However, the signal produced by the single and double positive mutants varied in intensity, showing that different quantities of mRNA appear to be produced. The wpwR^{w}, wwpR^{w}, pwpR^{w}, and wppR^{w} Env expression vector mutants appear to produce similar if not more *env* transcript than the wildtype, while the pwwR^{w}, ppwR^{w}, and pppR^{w} mutants appear to produce much less. As there appears to be a large difference in quantity of *env* mRNA despite normalizing the total RNA loaded, a qPCR was done to further investigate the quantity of RNA produced by the *env* mutants and how it is altered compared to the wildtype Env expression vector.

In Figure 3.2.7.B, HIV-1 transcripts are degraded quickly when retained in the nucleus and are more stable in the cytoplasm. Therefore, the quantity of transcripts found in samples that are primarily nuclear should contain less env transcript than the samples where the transcript is found in the cytoplasm. As expected, there is a 21.52-fold increase of mRNA produced by the wildtype Env expression vector in the presence of Rev as measured by qPCR. The single HEXplorer optimized Env expression vector wpwR^{w} shows a similar increase in mRNA quantity with a 24.57-fold change. The other two single HEXplorer positive expression vectors only show a modest increase in mRNA quantity with a 3.57-fold increase for the pwwR^{w} mutant and 4.46-fold increase for the wwpR^{w} mutant. Most of the double positive mutants show a small increase in mRNA in the presence of Rev. The ppwR^{w} and pwpR^{w} mutants show a small increase in mRNA of 1.65-fold and 2.18-fold in the presence of Rev. The last double HEXplorer optimized mutant wppR^{w}, however shows an increase in mRNA of 10.35-fold. Lastly, the mutant with the full CRS region HEXplorer optimized, pppR^{w}, shows no change in mRNA quantity, with a fold change of 1.02. Based on this data, all the Env expression vectors produce transcripts that are the same length, ruling out the activation of alternative splice sites. When compared to figure 3.2.5, we see that despite producing the most gp120 in the absence of Rev of the single mutants, the pwwR^{w} mutant has the lowest accumulation of RNA. This may be due to increased translation efficiency of this transcript. Overall, this figure confirms that Rev-independent export of the mutated *env* transcripts was not due to activation of

alternative splicing, but rather reversal of the intron 4 nuclear retention signal first described in (Mikaélian, Krieg et al. 1996).

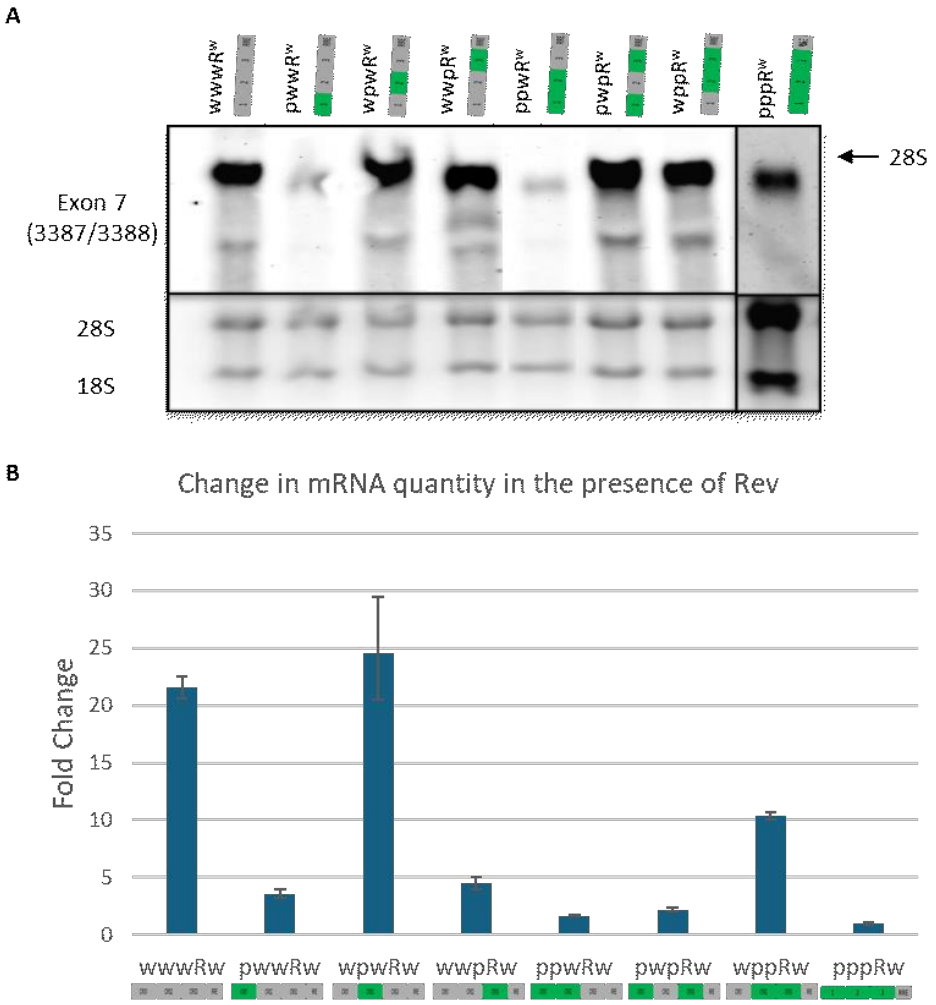


Figure 3.2.7. Env mutants were analyzed via northern blot using a probe specific for Exon 7 of HIV-1. HeLa cells were seeded 24 hours prior to transfection at 1.5 x 10^5 cells per mL into wells. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested for RNA 24 hours after transfection. The RNA was then isolated and six µg of the Rev cotransfected samples were loaded onto an agarose gel for analysis via northern blot using a probe specific for Exon 7 of HIV-1 that is present in the transcript. cDNA was made for samples transfected in the presence and absence of Rev and analyzed via RT-qPCR. The relative RNA quantity from the qPCR was analyzed to show the fold change in mRNA quantity in the presence of Rev. A) The northern blot shows that the mutant env transcripts are all the same length and appear to be slightly smaller than the 28S ribosomal RNA. B) RT-qPCR shows the relative change in mRNA quantities of Rev-independent mutants in the presence of Rev, compared to without.

Overall, this chapter has shown that the ability of the HEXplorer optimized CRS region is not completely due to a specific sequence to induce robust nuclear export in the absence of Rev. In fact, the quantity of HEXplorer optimized CRS seems to play a larger role in predicting Rev-independent nuclear export. Additionally, there is no change in the size of the transcript due to the introduction of splicing regulatory element binding sites activating a previously unused splice site, indicating that the changes in gene expression and localization are likely due to alterations in export patterns rather than alternative splicing. This alteration in export patterns may be due to an alteration in the nuclear export pathway. Rev facilitates nuclear export by recruiting the nuclear export factor CRM1, which is not the most common nuclear export pathway for mRNAs. On the other hand, several SR proteins, whose binding is likely increased in HEXplorer optimized sequences, have been shown to recruit the canonical mRNA nuclear export factor, NXF1. While HEXplorer identifies regions predicted to bind SR proteins, it does not identify individual SR protein binding sites. The difference in individual CRS nuclear export pattern may be due to differences in nuclear export properties between individual SR proteins that are not differentiated by the HEXplorer algorithm.

3.3 Rev-independent Env Mutants are exported independently of CRM1.

The canonical export pathway for HIV-1 late phase gene expression requires recruitment of the karyopherin CRM1 to RRE-containing transcripts by the virally encoded protein Rev. While several of the HEXplorer-guided *env* mutants show export to the cytoplasm in the absence of Rev, this does not necessarily mean that they are exported independent of CRM1. To investigate the export pathway these Rev-independent mutants use, cells expressing the *env* mutants were incubated in the presence of the CRM1 inhibitor KPT330. The *env* expression vectors shown in 3.1.2.A, 3.2.2.A, and 3.2.4.A were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via RNA FISH to determine the subcellular localization of the transcripts in the presence and absence of Rev. HeLa cells were transfected with the Env expression vectors and either a mock vector or a Rev expression vector. After incubating for 6 hours, 250 nM KPT330 or an equivalent volume of DMSO was added to select wells and incubated for 18 hours prior to fixation. The samples were analyzed via FISH for nuclear export patterns in the presence and absence of Rev and KPT330 using custom Env specific fluorescent RNA probes from Stellaris, imaged, and analyzed to determine the percentage of fluorescent Env mRNA found in the cytoplasm compared to the nucleus.

In figure 3.3.1.A, the transcripts produced by the wildtype Env expression vector appear to be retained in the nucleus in the absence of Rev. When Rev is present, the transcript can localize to the cytoplasm. When the CRM1 inhibitor KPT330 is introduced, export of the wildtype transcript is inhibited, and is primarily found in the nucleus. In figure 3.3.1.B, the cytoplasmic fluorescence of the wildtype Env expression vector increases by 3.38-fold

in the presence of Rev. When KPT330 is introduced, there is a decrease in cytoplasmic fluorescence by 2.6-fold compared to the wildtype in the absence of Rev. The modest decrease in fluorescence in the presence of Rev is likely due to the incomplete inhibition of CRM1 by KPT330. This observation confirms that the wildtype Env expression vector relies on the interaction between Rev and CRM1 for nuclear export. The HEXplorer optimized Env expression vectors, on the other hand, show strong cytoplasmic signal in the absence of Rev or presence of KPT330. This finding indicates that in addition to being exported in a Rev independent manner, they are also exported via an alternative nuclear export pathway from CRM1. The HEXplorer optimized Env expression vectors, $pppR^{w}$ and $pppR^{p}$, are less affected by the introduction of KPT330. In the presence of Rev, the cytoplasmic fluorescence increases by 1.28-fold, which decreases to 1.12-fold when KPT330 is introduced. In the presence of Rev, the change in cytoplasmic fluorescence is 1.01-fold, which decreases to 0.84-fold when KPT330 is introduced. Taken together, this data shows that KPT330 reduces cytoplasmic export of the Rev-RRE dependent wildtype Env expression vector, while having little effect on the Rev-independent Env expression vectors, $pppR^{w}$ and $pppR^{p}$. This implies that the HEXplorer optimized Env expression vectors do not use the RRE-Rev-CRM1 interaction to facilitate nuclear export and instead rely on an alternative nuclear export pathway.

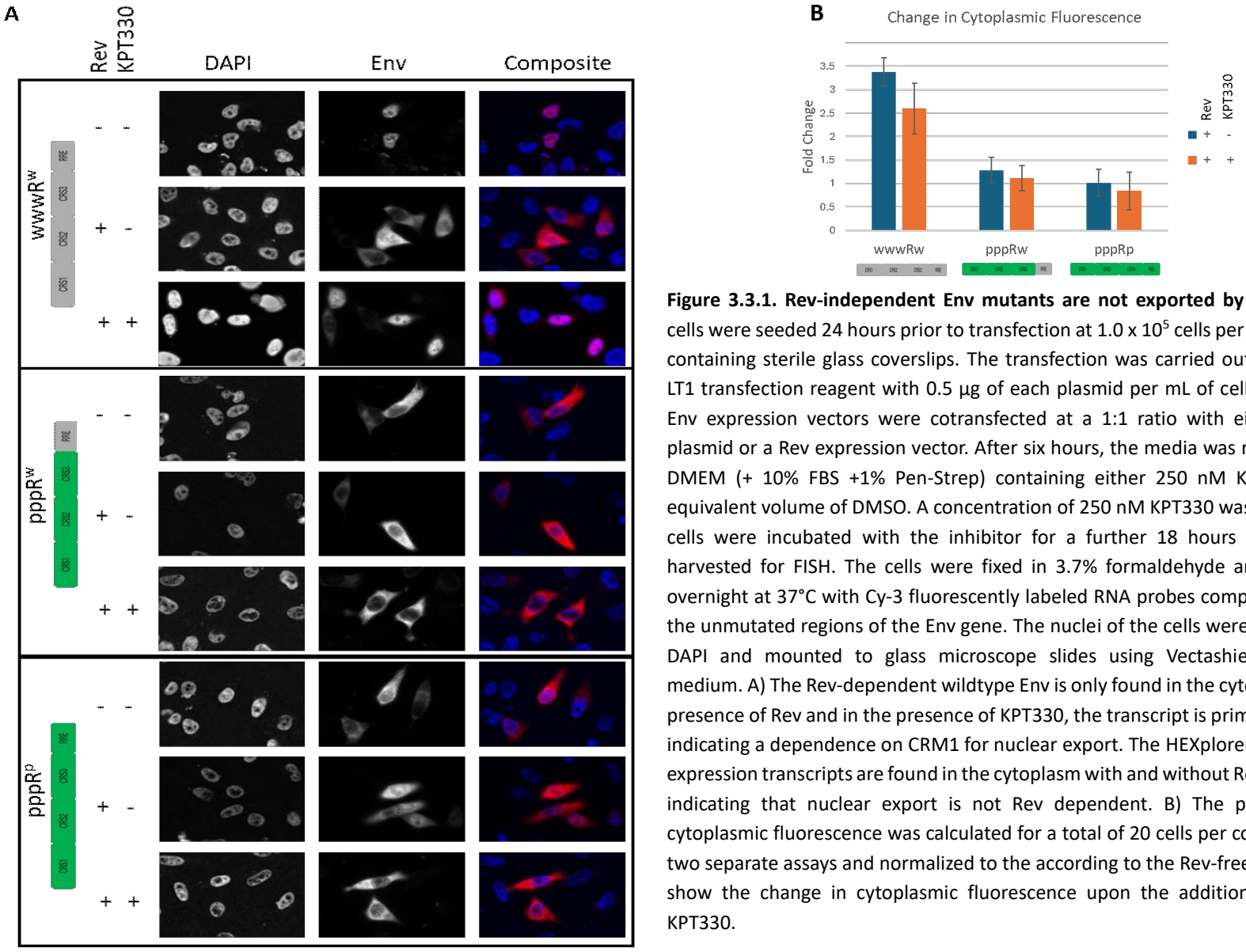


Figure 3.3.1. Rev-independent Env mutants are not exported by CRM1. HeLa cells were seeded 24 hours prior to transfection at 1.0 x 10^5 cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector. After six hours, the media was replaced with DMEM (+ 10% FBS +1% Pen-Strep) containing either 250 nM KPT330 or an equivalent volume of DMSO. A concentration of 250 nM KPT330 was chosen. The cells were incubated with the inhibitor for a further 18 hours before being harvested for FISH. The cells were fixed in 3.7% formaldehyde and incubated overnight at 37°C with Cy-3 fluorescently labeled RNA probes complementary to the unmutated regions of the Env gene. The nuclei of the cells were stained with DAPI and mounted to glass microscope slides using Vectashield mounting medium. A) The Rev-dependent wildtype Env is only found in the cytoplasm in the presence of Rev and in the presence of KPT330, the transcript is primarily nuclear, indicating a dependence on CRM1 for nuclear export. The HEXplorer positive Env expression transcripts are found in the cytoplasm with and without Rev or KPT330, indicating that nuclear export is not Rev dependent. B) The percentage of cytoplasmic fluorescence was calculated for a total of 20 cells per condition from two separate assays and normalized to the according to the Rev-free condition to show the change in cytoplasmic fluorescence upon the addition of Rev and KPT330.

As previously shown, HEXplorer optimized transcripts such as $pppR^{w}$ and $pppR^{p}$ do not produce as much protein as the wildtype env expression vector in the presence of Rev. Therefore, to investigate the quantity of protein produced by the HEXplorer optimized Env expression vectors in the presence of KPT330, HeLa cells were transfected with the Env expression vectors shown in figure 3.1.1.A with and without Rev and analyzed for Env protein production via western blot. HeLa cells were transfected with the Env mutants and either a Mock plasmid or a Rev expression vector. After incubating for 6 hours, 250 nM KPT330 or an equivalent volume of DMSO was added to select wells. After incubating overnight, the cells were harvested for a western blot. The signal from the western blot was quantified and normalized to $wwwR^{w}$ in the presence of Rev. The results from three separate biological replicates were averaged and shown in figure 3.3.2.B.

In figure 3.3.2.A, we see that the wildtype Env expression vector produces no gp120 in the absence of Rev, but shows a strong signal in the presence of Rev. When the CRM1 inhibitor, KPT330, is introduced, there is a moderate decrease in signal, however gp120 is still present. In the HEXplorer optimized $pppR^{w}$ and $pppR^{p}$ mutants produce gp120 signal independent of Rev or presence KPT330. Consistent with figure 3.1.4, the $pppR^{p}$ mutant produces less gp120 signal than the $pppR^{w}$ mutant in all conditions. In figure 3.3.2.B, protein quantities are compared to the wildtype Env expression vector in the presence of Rev, which is normalized to 100%. When Rev is along with the wildtype Env expression vector, the gp120 produced increases significantly from 0.5% to 100%. In the presence of KPT330, the quantity of protein produced by the wildtype Env expression vector reduces to 48% of control gp120 production. The HEXplorer optimized Env expression vector $pppR^{w}$ in the absence of Rev only produces 48% of the gp120 produced by the wildtype vector in the presence of Rev, and this value decreases to 35% in the presence of Rev. A further decrease is observed in the presence of KPT330 to 23%. The $pppR^{p}$ mutant behaves in a similar manner, and in the absence of Rev produced 20% of the protein produced by the wildtype Env expression vector in the presence of Rev. This number decreases to 15% protein production in the presence of Rev and even further to 9% protein production in the presence of KPT330. This indicates that the wildtype Env expression vector's ability to produce protein is diminished when CRM1 is inhibited showing that it is dependent on CRM1 for nuclear export. The decrease in protein production in the presence of KPT330 for the HEXplorer optimized Env expression vectors, $pppR^{w}$ and $pppR^{p}$, is less drastic, implying that these mutants have little to no dependency on CRM1 for nuclear export. Additionally, both HEXplorer optimized mutant Env expression vectors show a small decrease in protein production when Rev is present, which may indicate potential for Rev to inhibit expression of transcripts not exported via CRM1. This data indicates that, while the wildtype transcript utilizes the RRE-Rev-CRM1 interactions for gene expression the HEXplorer optimized Env expression vectors show little dependence on CRM1 for expression. Furthermore, the $pppR^{p}$ mutant is exported independently of the RRE-Rev-CRM1 interaction as it lacks a functional RRE secondary

structure (chapter 3.1). Taken together, this data indicates that the Rev-independent mutants are also exported independent of the CRM1 nuclear export pathway.

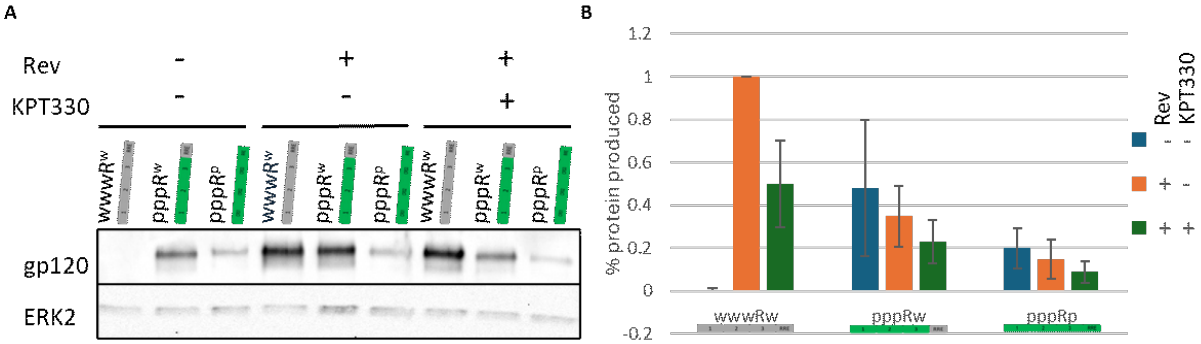


Figure 3.3.2. Rev-independent Env mutants produce protein in the absence of Rev. HeLa cells were seeded 24 hours prior to transfection at 1.5 x 10^5 cells per mL into wells. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector. After six hours, the media was replaced with DMEM (+ 10% FBS +1% Pen-Strep) containing either 250 nM KPT330 or an equivalent volume of DMSO. A concentration of 250 nM KPT330 was chosen as it showed a significant effect on wildtype Env transcript export without inducing major cell death. The cells were incubated with the inhibitor for a further 18 hours before being harvested for FISH. The samples were lysed in RIPA buffer and analyzed for the presence of gp120, a protein produced by the env transcript, and the housekeeper gene ERK2 via western blot. A) Western blot of Env Mutants. B) Change in protein quantity from the Mock transfection. While all the Rev-independent Env mutants produce some protein in the absence of Rev, the wildtype Env increases greatly in the presence of Rev, showing a dependency on Rev for nuclear export. Additionally, the quantity of protein decreases when KPT330 is introduced, showing that the wildtype transcript relies on CRM1 for export. The Rev-independent mutants (pppRw and pppRp) produce similar quantities of protein regardless of the presence of Rev or KPT330, indicating that they do not rely on the Rev-RRE system, and thus CRM1, for nuclear export.

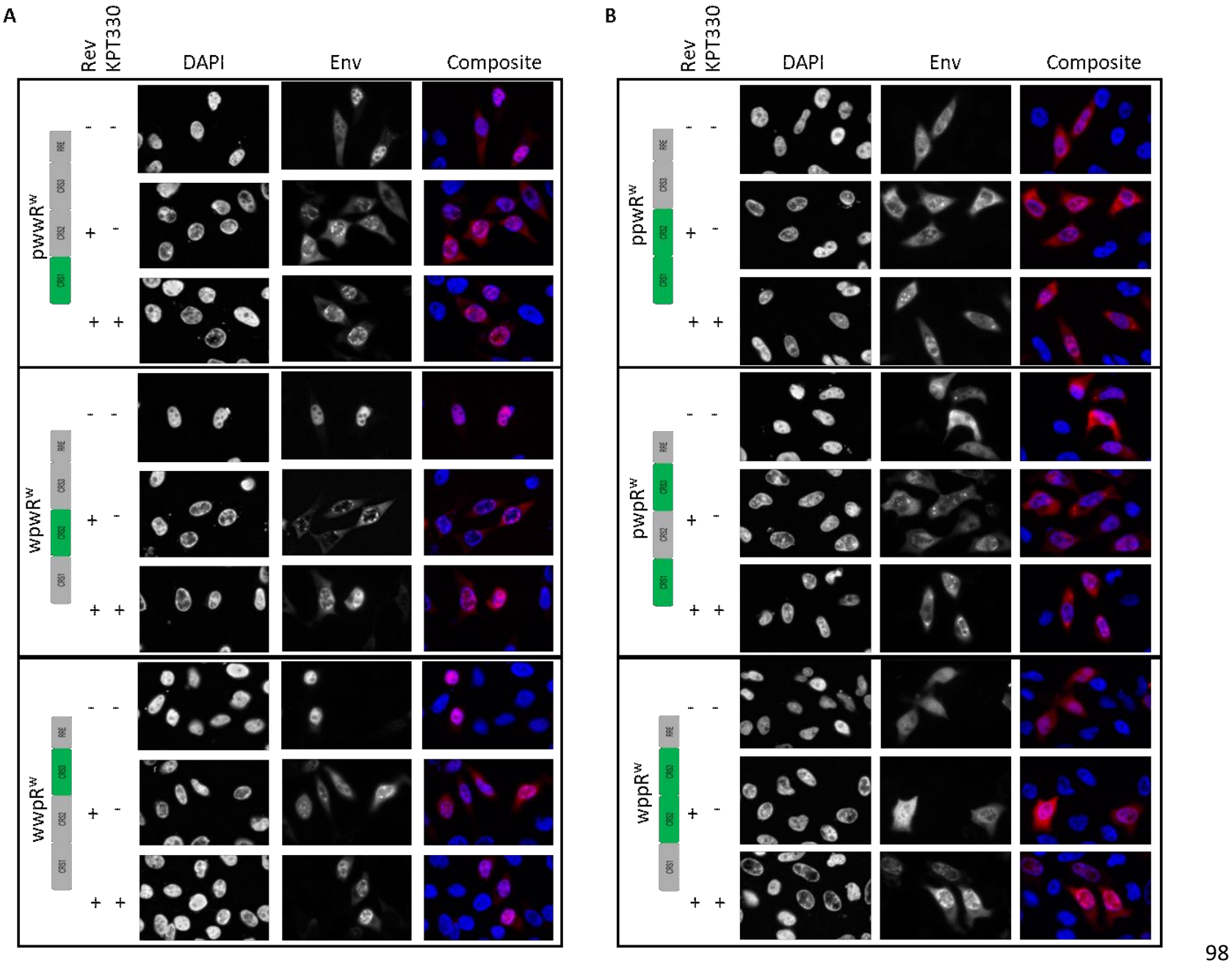
As the HEXplorer optimized Env expression vectors, $pppR^w$ and $pppR^p$, appear to be not only exported in a Rev-independent manner, but also exported by a different pathway than the canonical Rev-RRE mediated CRM1 pathway, the export pathway of the single and double positive HEXplorer optimized subregion Env expression vectors were also investigated. To investigate the export pathway requirements of these single and double positive Env mutants, cells expressing the Env mutants were grown in the presence of the CRM1 inhibitor KPT330. The Env expression vectors shown in 3.2.2.A and 3.2.4.A were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via FISH to determine the subcellular localization of the transcripts in the presence and absence of Rev. A concentration of 250 nM KPT330 was chosen as it showed a significant effect on wildtype Env transcript export without inducing major cell death. HeLa cells were transfected with the Env expression vectors and either a mock vector or a Rev expression vector. After incubating for 6 hours, 250 nM KPT330 was added to select wells, and an equivalent volume of DMSO is added to the control wells. They were then analyzed via FISH for nuclear export patterns in the presence and absence of Rev and KPT330. After incubating overnight, the cells were fixed and incubated with custom Env

specific fluorescent RNA probes from Stellaris and imaged. Twenty cells from each condition from two separate assays were imaged and analyzed to determine the percentage of fluorescent Env mRNA found in the cytoplasm compared to the nucleus.

In figure 3.3.3.A, the single HEXplorer optimized subregion vectors have been cotransfected with Rev and treated with the CRM1 inhibitor KPT330. With the pwwR^{w} mutant, the transcript appears to be primarily nuclear in the absence of Rev, albeit with a weak cytoplasmic signal. The strength of the cytoplasmic signal increases in the presence of Rev, however there is still a strong nuclear signal. When KPT330 is introduced, the cytoplasmic signal again decreases. With the other single HEXplorer optimized mutants, wpwR^{w} and wwpR^{w} mutants, the Env transcript signal is clearly retained in the nucleus, with cytoplasmic transcript appearing in the presence of Rev. When KPT330 is present, there is a decrease in cytoplasmic transcript, however, the cytoplasmic signal is not completely reversed, potentially because KPT330's inhibition of CRM1 is non-covalent and thus incomplete. In figure 3.3.3.B, the double HEXplorer optimized subregion vectors, ppwR^{w}, pwpR^{w}, and wppR^{w}, have been cotransfected with Rev and treated with the CRM1 inhibitor KPT330. All of the double HEXplorer optimized mutants show cytoplasmic *env* transcript signal in the absence of Rev, with little change in the presence of Rev. In the presence of KPT330 a strong cytoplasmic signal for all three double HEXplorer optimized mutants was observed, indicating that KPT330 had little to no effect on cellular localization of these mutants.

In figure 3.3.3.C, the change in cytoplasmic fluorescence from the condition lacking Rev for each mutant is shown. The wildtype cytoplasmic fluorescence increases by 1.74-fold in the presence of Rev. In accordance with its dependence on CRM1, the cytoplasmic fluorescence produced by the wildtype Env expression vector only increases by 1.29-fold in the presence of KPT. This indicates that while the inhibition of CRM1 by KPT330 is not complete, it still decreases CRM1-dependent export. The pwwR^{w} mutant however shows a low increase in cytoplasmic fluorescence in the presence of Rev of 1.21-fold, likely due to the faint cytoplasmic signal. However, when KPT330 is present, the cytoplasmic fluorescence decreases to 0.88-fold of pwwR^{w} in the absence of Rev. This decrease in cytoplasmic fluorescence beyond the condition in the absence of Rev indicates that pwwR^{w} is exported using RRE-Rev-CRM1 interactions. The other two single HEXplorer optimized transcripts show a much larger increase in cytoplasmic fluorescence in the presence of Rev, where the wpwR^{w} mutant increases cytoplasmic fluorescence by 1.81-fold, and the wwpR^{w} increases it by 1.50-fold. When KPT330 is introduced, the amount of cytoplasmic fluorescence decreases to 1.30 for the wpwR^{w} mutant and 1.04-fold for the wwpR^{w} mutant. In the case of the wpwR^{w} and wwpR^{w} mutants the decrease in cytoplasmic fluorescence in the presence of KPT330 is of a similar scale to that of the wildtype, confirming that they are also exported via the CRM1 pathway. For the double HEXplorer optimized Env expression vectors, ppwR^{w} and pwpR^{w} there is little to no

increase in cytoplasmic fluorescence in the presence of Rev, with each showing a change of 1.10-fold and 0.97-fold respectively. The wppR^{w} mutant does however show a moderate 1.28-fold increase in cytoplasmic fluorescence in the presence of Rev. When in the presence of both Rev and KPT330, the quantity of cytoplasmic fluorescence shows very little change from the condition in the absence of Rev, with ppwR^{w} decreasing to 0.92-fold, pwpR^{w} decreasing to 0.81-fold, and wppR^{w} producing a change of 1.03-fold in cytoplasmic fluorescence. Overall, the change in cytoplasmic fluorescence between the conditions containing Rev and those containing both Rev and KPT330, KPT330 has a larger impact on the cellular localization of the single HEXplorer optimized Env transcripts than on the double HEXplorer optimized transcripts. Taken together with the data from 3.3.3.A and 3.3.3.B, the single HEXplorer optimized transcripts are more reliant on CRM1 for nuclear export than the double HEXplorer optimized transcripts.


A
B
Rev
KPT330
DAPI
Env
Composite
pwwR^w
wpwR^w
wwpR^w
ppwR^w
pwpR^w
wppR^w
CRS1
CRS2
CRS3
RRE
+
-

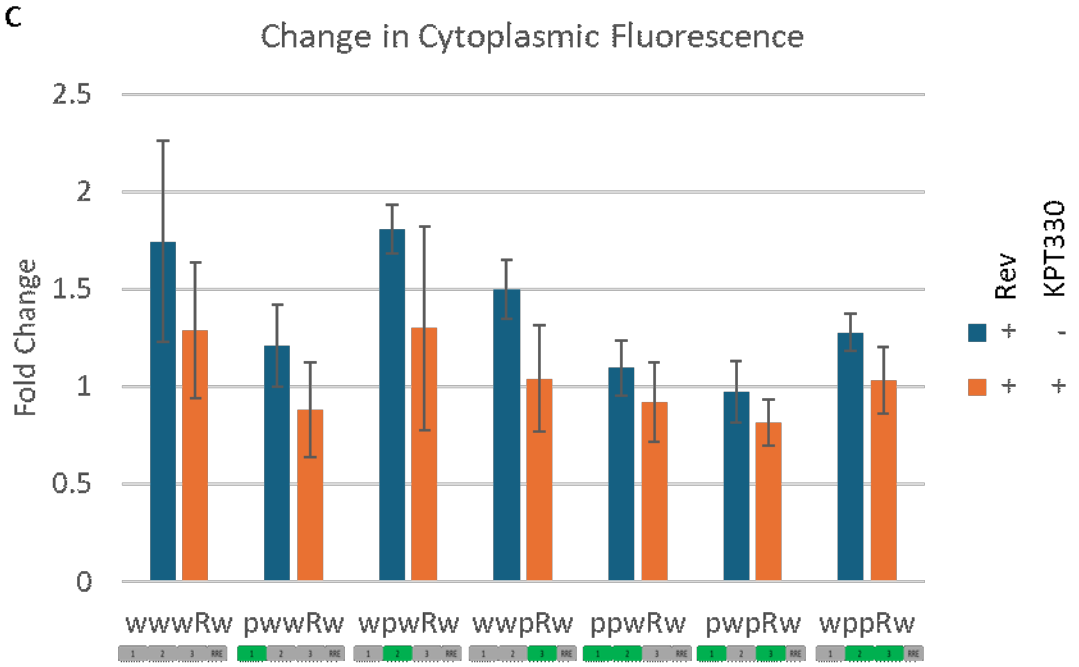


Figure 3.3.3. Rev-independent Env mutants are not exported by CRM1. HeLa cells were seeded 24 hours prior to transfection at 1.0 x 10^5 cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector. After six hours, the media was replaced with DMEM (+ 10% FBS +1% Pen-Strep) containing either 250 nM KPT330 or an equivalent volume of DMSO. The cells were incubated with the inhibitor for a further 18 hours before being harvested for FISH. The cells were fixed in 3.7% Formaldehyde and incubated overnight at 37°C with Cy3 fluorescently labeled RNA probes complementary to the unmutated regions of the Env gene. The nuclei of the cells were stained with DAPI and then mounted to glass microscope slides using Vectashield mounting medium. Twenty cells from each condition from two separate assays were imaged and analyzed to determine the percentage of fluorescent env mRNA found in the cytoplasm compared to the nucleus. A) The Rev-dependent wildtype Env is only found in the cytoplasm in the presence of Rev. However, in the presence of KPT330, the transcript is once again primarily nuclear, indicating a dependence on CRM1 for nuclear export. The single HEXplorer optimized Env expression vectors follow this trend, showing that these mutants use CRM1 for export. B) While the wildtype and single HEXplorer optimized Env expression vectors are affected by CRM1 inhibition via KPT330, the double HEXplorer optimized Env expression transcripts are found in the cytoplasm with and without Rev or KPT330, indicating that nuclear export is not Rev dependent. C) The percentage of cytoplasmic fluorescence was calculated for a total of 20 cells per condition from two separate assays and normalized to the according to the Rev-free condition to show the change in cytoplasmic fluorescence upon the addition of Rev and KPT330.

As previously shown, HEXplorer optimized transcripts such as $pppR^w$ and $pppR^p$ do not produce as much protein as the wildtype env expression vector in the presence of Rev. Therefore, to investigate the quantity of protein produced by the HEXplorer optimized subregion Env expression vectors in the presence of KPT330, HeLa cells were transfected with the Env expression vectors shown in figure 3.2.2.A and 3.2.4.A with and without Rev and analyzed for Env protein production via western blot. HeLa cells were transfected

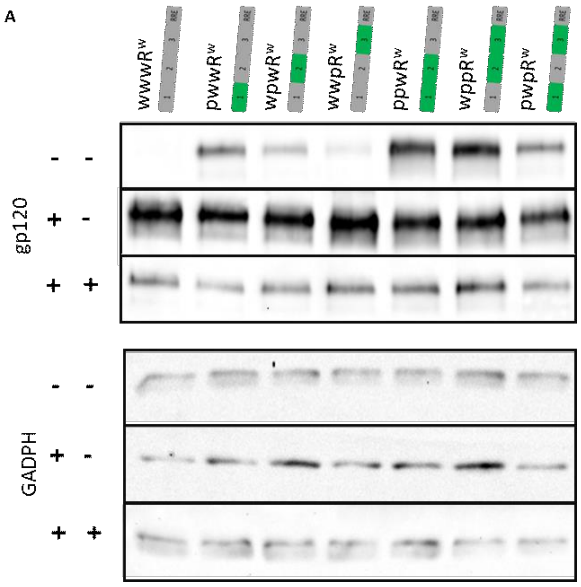
with the Env mutants and either a Mock plasmid or a Rev expression vector. After incubating for 6 hours, 250 nM KPT330 was added to select wells, and an equivalent volume of DMSO is added to the control wells. After incubating overnight, the cells were harvested for a western blot. A concentration of 250 nM KPT330 was chosen as it showed a significant effect on wildtype Env transcript export without inducing major cell death. The signal from the western blot was quantified and normalized to wwwR^{w} in the presence of Rev. The results from three separate biological replicates were averaged and results are shown in figure 3.3.4.B.

In figure 3.3.4.A, the wildtype Env expression vector produces no gp120 in the absence of Rev and increases significantly when Rev is present. When KPT330 is present in the Rev-containing condition we see a significant decrease in gp120 signal. The single HEXplorer optimized subregion Env expression vectors, pwwR^{w}, wpwR^{w}, and wwpR^{w}, follow a similar trend with a faint gp120 signal in the absence of Rev, which then increases significantly in the presence of Rev as expected from previous results. Similar to the wildtype vector, when both Rev and KPT330 are present, the level of gp120 decreases. This observation indicates that KPT330 prevents gp120 expression from these mutants by inhibiting CRM1 mediated nuclear export. The pwpR^{w} also seems to be dependent on CRM1 for gp120 protein production. It has a faint gp120 signal in the absence of Rev that increases in the presence of Rev, but when KPT330 is introduced, the signal weakens. The ppwR^{w} mutant has a less clear phenotype. There is a strong gp120 signal in the absence of Rev, indicating that gp120 production is not dependent on Rev, yet the gp120 signal weakens when KPT330 is introduced. On the other hand, the wppR^{w} mutant shows very little change in gp120 protein signal between the conditions, with the gp120 signal being slightly stronger in the presence of Rev than in the absence of Rev or presence of KPT330 and Rev combined.

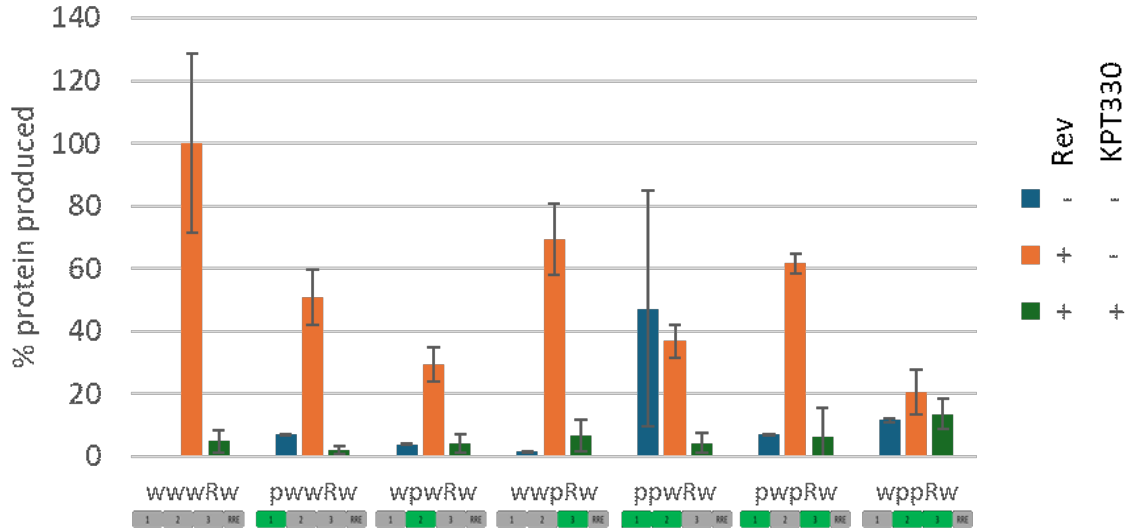
In figure 3.3.4.B and 3.3.4.C, the quantity of protein is normalized to the wildtype Env expression vector in the presence of Rev, and the quantity of protein produced is shown as a percentage of protein produced by the wildtype in the presence of Rev. The wildtype shows an increase in gp120 protein production in the presence of Rev, which decreases to 5% in the presence of KPT330. In the presence of Rev, the protein produced by the pwwR^{w} mutant increased by 44%, while the wpwR^{w} mutant only increased by 26%. The wwpR^{w} mutant shows the greatest increase of protein production in the presence of Rev, increasing by 68%. Similar to the wildtype Env expression vector, the single HEXplorer positive mutants show a significant decrease in protein production when KPT330 is present. When KPT330 is introduced, the pwwR^{w} produces 5% less than in the absence of Rev and KPT330, while the wpwR^{w} mutant produces nearly the same amount of protein as in the presence of Rev, and the wwpR^{w} mutant produces 5% more protein than in the absence of Rev. The protein produced by the double HEXplorer optimized subregion Env expression vector ppwR^{w} actually decreases by 10% when Rev is introduced, while the

pwpR^{w} mutant increases by 55%, and the wppR^{w} mutant increases by a more modest 9% of the protein produced by the wildtype Env in the presence of Rev. When KPT330 is introduced, the protein produced by the ppwR^{w} mutant decreases by 43% compared to in the absence of Rev and KPT330. Alternatively, pwpR^{w} produces a similar amount of protein in the presence of KPT330 and Rev as in the absence of Rev. The protein produced by the wppR^{w} mutant in the presence of Rev decreases by 7% when KPT330 is introduced. This data shows that the single positive mutants are clearly dependent on the Rev-CRM1 nuclear export pathway, whereas the double positive mutants do not show a clear pattern. The ppwR^{w} mutant produces protein independent of Rev, yet protein production decreases when CRM1 is inhibited. The pwpR^{w} mutant however appears to produce Env in a Rev-dependent manner, and the double positive mutant, wppR^{w} produces Env in a mostly Rev-independent manner. This inconclusive pattern may be due to each CRS having independent effects on nuclear export through unique RNA binding protein environments.

A



B



C

Rev	KPT330	wwwRw	pwwRw	wpwRw	wwpRw	ppwRw	pwpRw	wppRw
-	-	0.00	6.95	3.72	1.55	47.14	6.88	11.55
+	-	100.00	50.81	29.36	69.25	36.87	61.66	20.61
+	+	4.79	1.93	3.93	6.60	4.32	6.15	13.49

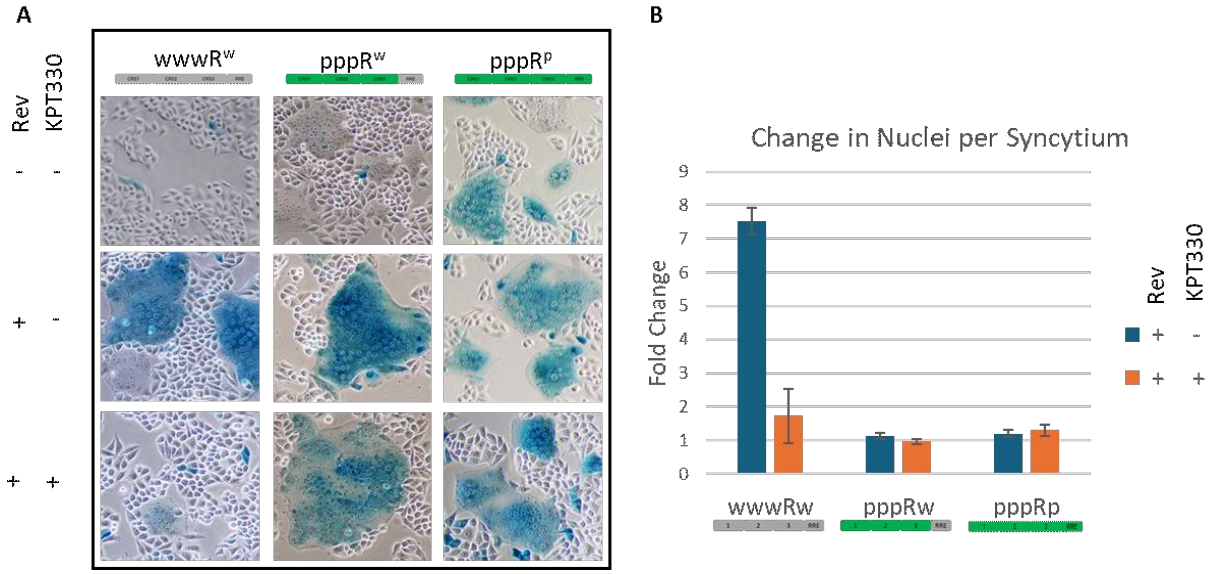
Figure 3.3.4. Rev-independent Env mutants produce protein in the absence of Rev. HeLa cells were seeded 24 hours prior to transfection at 1.5 x 10^5 cells per mL into wells. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector. After six hours, the media was replaced with DMEM (+ 10% FBS +1% Pen-Strep) containing either 250 nM KPT330 or an equivalent volume of DMSO. The samples were lysed in RIPA buffer and analyzed for the presence of gp120, a protein produced by the env transcript, and the housekeeper gene GAPDH via western blot. A) Western blot of Env Mutants. B) Change in protein quantity from the Mock transfection. While all the Rev-independent Env mutants produce some protein in the absence of Rev, the wildtype Env increases greatly in the presence of Rev, showing a dependency on Rev for nuclear export. Additionally, the quantity of protein decreases when KPT330 is introduced, showing that the wildtype transcript relies on CRM1 for export. The Rev-independent mutants produce similar quantities of protein regardless of the presence of Rev or KPT330, indicating that they do not rely on the Rev-RRE system, and thus CRM1, for nuclear export. C) Table showing the percentage of protein produced by the single and double positive mutants HEXplorer in the presence and absence of Rev and KPT330 compared to the wildtype Env expression vector in the presence of Rev.

In addition to analyzing the quantity of protein produced by the Env mutants, the function of the protein produced was analyzed via syncytia assay. To investigate the HEXplorer optimized Env expression vector's ability to induce membrane fusion in the presence of KPT330, the Env expression vectors shown in 3.1.1.A were transfected into TZMbL cells. TZM-bL cells were transfected with the Env Mutants and either a Mock plasmid or a Rev expression vector. After incubating for 6 hours, 250 nM KPT330 or an equivalent volume of DMSO was added to select wells. After incubating overnight, the cells were fixed and incubated with β-galactosidase substrate and imaged. The β-galactosidase activity is independent of Env expression; however, the blue dye induced by Tat expression aids in visualizing syncytium formation.

Figure 3.3.5.A shows that the wwwR^w Env expression vector is only able to produce syncytia in the presence of Rev, while the Rev-independent mutants (pppR^w and pppR^p) produce large syncytia in both the presence and absence of Rev. Interestingly, the conditions lacking Rev show less β-galactosidase activity compared to the conditions with Rev. Upon the addition of KPT330 to the wildtype Env expression vector, the size of the syncytia decreases, though they do not completely disappear. This indicates that the inhibition of CRM1 leads to a decrease in membrane fusion in permissive cells transfected with the wildtype Env. However, KPT330 does not induce a change in the size of syncytia produced by the HEXplorer optimized mutants, pppR^w and pppR^p. This finding indicates that sufficient Env protein is produced and localized to the cell membrane independent of CRM1. In Figure 3.1.5.B, the wwwR^w shows a 7.9-fold increase in syncytium formation with the addition of Rev, showing a dependency on the Rev-RRE system for nuclear export. In the presence of KPT330, there is only an increase of nuclei per syncytium of 2.04-fold, indicating that inhibiting CRM1 leads to 4-fold inhibition of Env expression from the wildtype Env expression vector. The pppR^w and pppR^p mutants however have little to no change in nuclei per syncytium when Rev is present with a fold change of 1.13 and 1.19 respectively. In the presence of KPT330, these HEXplorer positive Env expression vectors

show no difference in the size of the syncytia, with $pppR^{w}$ decrease with 0.96-fold, and size of nuclei produced by $pppR^{p}$ only increasing by 1.31-fold. Taken together, this data indicates that the wildtype Env expression vector is reliant on CRM1 to produce functional protein, but the extensively HEXplorer optimized vectors are exported when CRM1 is inhibited, indicating that they are exported by an alternative nuclear export pathway.

Figure 3.3.5. The Rev-independent Env mutants induce membrane fusion in CD4+ cells in the absence of



Rev. The TZMbL cells were seeded 24 hours prior to transfection at 2.0 x 10^5 cells per mL. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector. After six hours, the media was replaced with DMEM (+ 10% FBS +1% Pen-Strep) containing either 250 nM KPT330 or an equivalent volume of DMSO. After incubating at 37°C for a further 18 hours, the cells were fixed and incubated with β-galactosidase substrate. Ten fields per assay from two biologically separate assays were imaged from each condition and the nuclei per syncytium counted and quantified. A) The wildtype Env shows a large increase in syncytium formation with the addition of Rev, showing a dependency on the Rev-RRE system for nuclear export. The Rev-independent mutants (pppRw and pppRp) produce large syncytia in the absence of Rev, indicating that they do not rely on Rev for nuclear export. B) The number of nuclei per syncytium was counted for ten random fields in two separate assays. It was then normalized to absence of Rev to show the change in nuclei per syncytium for each condition.

In addition to analyzing the quantity of protein produced by the Env mutants, the function and surface expression of the protein produced was analyzed using a syncytia assay. Therefore, to investigate the HEXplorer optimized subregion Env expression vector's ability to induce membrane fusion in the presence of KPT330, the Env expression vectors shown in 3.2.1.A and 3.2.3.A were transfected into TZM-bL cells. TZMbL cells were transfected with the Env Mutants and either a Mock plasmid or a Rev expression vector. After incubating for 6 hours, 250 nM KPT330 was added to select wells, and an equivalent volume of DMSO is added to the control wells. After incubating overnight, the cells were fixed and incubated with β-galactosidase substrate and imaged.

In figure 3.3.6.A, the wildtype Env expression vector does not produce any large syncytia in the absence of Rev, however in the presence of Rev, large syncytia form. The introduction of KPT330 greatly decreases the formation of syncytium in the wildtype Env expression vector. The pwwR^w mutant appears to produce small syncytia in the absence of Rev, though there are large and extensive syncytia in the presence of Rev. The other two single HEXplorer optimized subregion Env expression vectors, wpwR^w and wwpR^w show no syncytia in the absence of Rev, and the wwpR^w mutant produces large syncytia in the presence of Rev like the wildtype and pwwR^w. The wpwR^w mutant, however, appears to produce smaller syncytia than the wildtype, pwwR^w, and wwpR^w. In the presence of the CRM1 inhibitor KPT330, the size of syncytia formed by the pwwR^w and wwpR^w mutants decreases. The wpwR^w mutant, however, still produces small syncytia in the presence of both Rev and KPT330. This indicates that the single HEXplorer optimized env expression vectors pwwR^w, and wwpR^w relies on the interaction between Rev and CRM1 for nuclear export, whereas the wpwR^w mutant does not appear to be affected by CRM1 inhibition. The double positive HEXplorer optimized subregion Env expression vectors in figure 3.3.6.A, ppwR^w, pwpR^w and wppR^w, all show large syncytium formation in the absence and presence of Rev as well as in the presence of KPT330. The ability of the double HEXplorer optimized Env mutants to induce syncytium formation regardless of the presence of Rev or KPT330 indicates that they do not rely on the RRE-Rev-CRM1 interaction for export to the cytoplasm and subsequent expression.

Figure 3.2.6.B shows the change in nuclei per syncytium in the presence of Rev. The wildtype Env expression vector produced the largest change in nuclei per syncytium at 12.60-fold more nuclei per syncytium in the presence of Rev, which decreases to 4.15-fold when KPT330 is introduced. The single HEXplorer optimized Env expression vectors all show an increase in nuclei per syncytium over 2-fold in the presence of Rev, with pwwR^w, wpwR^w, and wwpR^w having a change in nuclei per syncytium of 2.79-fold, 4.49-fold, and 6.03-fold respectively. In the presence of KPT330, this increase in syncytia size decreases to 1.72-fold for the pwwR^w mutant and 2.37-fold for the wwpR^w mutant. The size of syncytia formed by the wpwR^w mutant does not change, remaining at approximately 4.5-fold when KPT330 is introduced. The double HEXplorer optimized Env expression vectors however show similar sized syncytia in both the absence and presence of Rev as well as the presence of KPT330, expressed by a fold change close to 1. The nuclei per syncytium of the ppwR^w mutant increases by 1.16-fold, the pwpR^w mutant by 1.13-fold, and the wppR^w mutant by 1.23-fold. In the presence of KPT330, the size of the syncytia of the ppwR^w mutant increases by 1.13-fold, the pwpR^w mutant by 0.93-fold, and the wppR^w mutant by 1.04-fold. Taken together with the data presented in figure 3.3.6.A, this indicates that the wildtype and single HEXplorer optimized Env expression vectors are dependent on Rev and its interaction with CRM1 for export and subsequent translation, while the double HEXplorer optimized Env expression vectors produce sufficient functional protein to induce syncytium formation in the presence of KPT330.

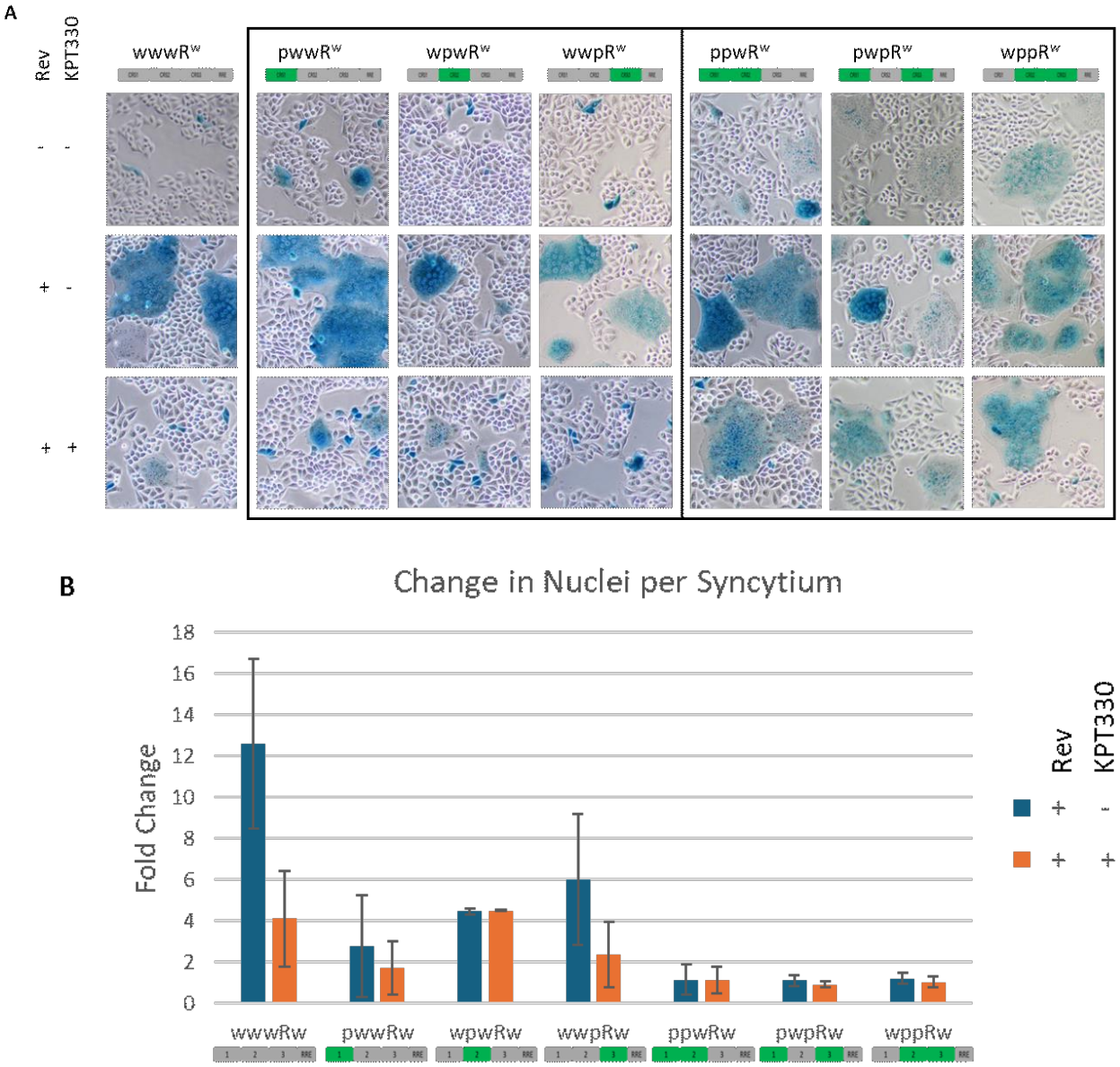


Figure 3.3.6. The Rev-independent Env mutants induce membrane fusion in CD4+ cells in the absence of Rev. The TZM-bL cells were seeded 24 hours prior to transfection at 2.0 x 10^5 cells per mL. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector. After six hours, the media was replaced with DMEM (+ 10% FBS +1% Pen-Strep) containing either 250 nM KPT330 or an equivalent volume of DMSO. After incubating at 37°C for a further 18 hours, the cells were fixed and incubated with β-galactosidase substrate. Ten fields per assay from two biologically separate assays were imaged from each condition and the nuclei per syncytium counted and quantified. A) The wildtype and Rev-dependent Env expression vectors show large increases in syncytium formation with the addition of Rev and decrease of syncytium formation in the presence of KPT330, showing a dependency on the Rev-RRE system for nuclear export. The Rev-independent mutants produce large syncytia in the absence of Rev, indicating that they do not rely on Rev or CRM1 for nuclear export. B) The number of nuclei per syncytium was counted for ten random fields in two separate assays. It was then normalized to absence of Rev to show the change in nuclei per syncytium for each condition.

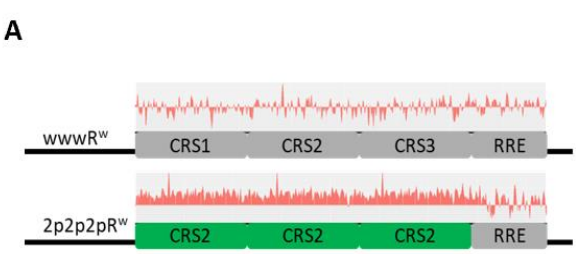
Overall, this chapter has shown that the Env expression vectors that are not dependent on the Rev-RRE interaction for nuclear export are also exported independent of CRM1 through inhibition of CRM1 via KPT330. This further confirms that the double HEXplorer optimized, and fully optimized Env expression vectors do not rely on the RRE-Rev-CRM1 interaction for nuclear export. It is important to note that there were several inconsistencies, particularly with the pwwR^{w} and pwpR^{w} mutants in protein production, which do not correlate to the RNA FISH data. This indicates that while the size of the region that is HEXplorer optimized plays a role in nuclear export patterns, there may be yet unrecognized sequence specific mechanisms due to the binding of different SR proteins involved in regulating nuclear export.

3.4 Specific protein RNA binding sites still need to be considered when predicting nuclear export patterns

While chapter 3.2 shows that the quantity of HEXplorer optimized sequence plays a larger role in the selection of nuclear export pathway, it is still possible that individual CRS sequences still exert influence over nuclear export pathway selection. To further investigate this possibility, the most positive HEXplorer optimized CRS subregion, CRS2, was concatenated by three, and cloned in place of the wildtype CRS region. As the concatenated vectors do not maintain the *env* ORF amino acid sequence, these vectors are unable to produce protein and were only analyzed via FISH. The concatenated Env vector and wildtype env vectors shown in 3.4.1.A were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via RNA FISH to determine the subcellular localization of the transcripts in the presence and absence of Rev. Ten cells from each condition were imaged and analyzed to determine the percentage of fluorescent Env mRNA found in the cytoplasm compared to the nucleus.

Figure 3.4.1.A shows a schematic of the wildtype Env expression vector, and the CRS2 concatenated Env vector, 2p2p2pR^{w}. As shown in figure 3.1.1.B, CRS2 is not only the longest CRS subregion by 10 nucleotides, but also has the highest HEXplorer score per nucleotide. Concatenation of the CRS2 produced the longest and most HEXplorer optimized mutant CRS region possible. The concatenated 2p2p2pR^{w} vector no longer contains an intact Env ORF, and is unable to produce protein, however the region complementary to the FISH probes remains intact. This figure also shows the HEXplorer profile of the mutated region that is analyzed in Figure 3.4.1.B. In figure 3.4.1.B, the HEXplorer score of the CRS and RRE region of the wildtype Env expression vector is compared to the HEXplorer score of the mutated region and RRE of the CRS2 concatenated Env vector. While the wildtype CRS and RRE region have a HEXplorer score of -0.46 per nt, the 2p2p2pR^{w} vector has a HEXplorer score of 9.13 per nt. In total, the wildtype Env expression vector has a total HEXplorer score of -719.07, compared to the total HEXplorer score of 15,469.81 for the 2p2p2pR^{w} vector. Figure 3.4.1.C shows that the

transcript produced by the wildtype Env expression vector is retained in the nucleus in the absence of Rev, however in the presence of Rev there is clear, strong cytoplasmic Env transcript signal. The transcripts produced by the 2p2p2pR^{w} vector, however, show weak cytoplasmic signal both in the absence and presence of Rev. In the presence of Rev, the strength of the nuclear signal increases. The cytoplasmic fluorescence in the absence of Rev indicates that this transcript is exported independently of Rev-RRE interaction, however the strong nuclear signal in the presence of Rev indicates that the transcript is sequestered in the nucleus. The increase of nuclear fluorescence in the presence of Rev may even indicate that Rev is inhibiting the export of the concatenated 2p2p2pR^{w} transcripts.



B

Mutant	Per/nt	Total
wwwR^{w}	-0.46	-719.07
2p2p2pR^{w}	9.13	15469.81

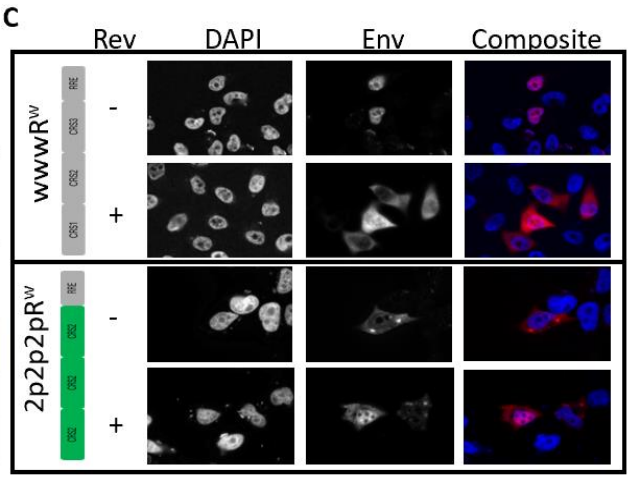


Figure 3.4.1. CRS2 does not contribute to nuclear export. The CRS2 cis-acting repressive sequence subregion found in HIV-1 env was mutated according to HEXplorer algorithm predictions and concatenated to produce an Env mutant containing three HEXplorer optimized CRS2 subregions in place of the wildtype CRS, altering the underlying amino acid sequence. HeLa cells were seeded 24 hours prior to transfection at 1.0 x 10^{5} cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested 24 hours after transfection. The cells were fixed in 3.7% Formaldehyde and incubated overnight at 37°C with Cy-3 fluorescently labeled RNA probes complementary to the unmutated regions of the Env gene. The nuclei of the cells were stained with DAPI and mounted to glass microscope slides using Vectashield mounting medium. A) Schematics of the wildtype and 2p2p2pRw Env vectors along with their HEXplorer plot, where green indicates HEXplorer optimized, and grey indicates the wildtype sequence. Mutants are referred to the name on the left which indicates positive fragments with a (p) and wildtype with a (w). B) The table shows total and per nucleotide HEXplorer score of the overall CRS regions in A. C) The Env vectors in A were transfected in HeLa cells and analyzed via FISH for cellular localization of the Env transcript. Ten cells from each condition were imaged and analyzed to determine the percentage of fluorescent Env mRNA found in the cytoplasm compared to the nucleus.

Figure 3.4.1 shows that the concatenated Env vector is not dependent on Rev for export to the cytoplasm, and it is likely that this effect remains when only two CRS2 are present. To investigate this possibility, the first CRS2 in the concatenated sequence was replaced with the wildtype CRS1 to produce the 1w2p2pR^{w} Env vector. The wildtype CRS1

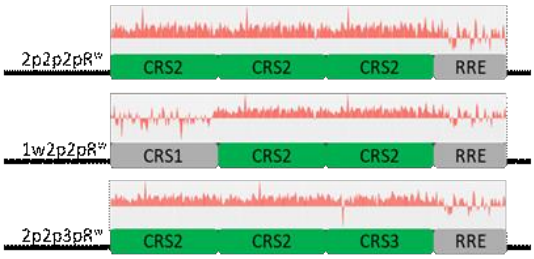
sequence was chosen because in the single HEXplorer optimized subregion Env expression vectors analyzed in figure 3.2.3, the mutants that still contained a wildtype CRS1 subregion, wpwR^{w} and wwpR^{w}, appeared more Rev dependent than the pwwR^{w} mutant. This hints that the wildtype CRS1 is a stronger CRS than CRS2 or CRS3. Additionally, the third CRS2 in the concatenated sequence was replaced by the HEXplorer optimized CRS3 sequence. The HEXplorer optimized CRS3 sequence was chosen because it is the shortest CRS subregion and has the least positive HEXplorer score, and thus will reduce the total HEXplorer score the most, while maintaining a HEXplorer optimized CRS region. To investigate the nuclear export patterns of the partially concatenated mutants, the 1w2p2pR^{w} and 2p2p3pR^{w} Env vectors were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via RNA FISH to determine the subcellular localization of the transcripts in the presence and absence of Rev.

In figure 3.4.1, the concatenated 2p2p2pR^{w} has more nuclear fluorescence in the presence of Rev, decreasing the percentage of cytoplasmic fluorescence. To investigate the change in cytoplasmic fluorescence of the concatenated 2p2p2pR^{w} vector and partially concatenated mutants, the 1w2p2pR^{w} and 2p2p3pR^{w} Env vectors, they were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via RNA FISH to determine the subcellular localization of the transcripts in the presence and absence of Rev. Ten cells from each condition were imaged and analyzed to determine the percentage of fluorescent Env mRNA found in the cytoplasm compared to the nucleus.

Figure 3.4.2.A shows a schematic of the wildtype Env expression vector, and the partial CRS2 concatenated Env vectors, 1w2p2pR^{w} and 2p2p3pR^{w}. The amino acid sequence of these partially concatenated Env vectors was altered, thus this vector is unable to produce Env protein. This figure also shows the HEXplorer profile of the mutated region that is analyzed in Figure 3.4.2.B. In figure 3.4.2.B, the HEXplorer score of the CRS and RRE region of the wildtype Env expression vector is compared to the HEXplorer score of the mutated region and RRE of the CRS2 concatenated Env vector. While the wildtype CRS and RRE region has a HEXplorer score of -0.46 per nt, the 1w2p2pR^{w} vector has a HEXplorer score of 6.12 per nt and the 2p2p3pR^{w} vector has a HEXplorer score of 8.89 per nt. In total, the wildtype Env expression vector has a total HEXplorer score of -719.07, compared to the total HEXplorer score of 10,232.61 for the 1w2p2pR^{w} vector and 13,709.70 for the 2p2p3pR^{w} vector. Figure 3.4.1.C shows that the transcript produced by the wildtype Env expression vector is retained in the nucleus in the absence of Rev. However, in the presence of Rev there is a clear, strong cytoplasmic Env transcript signal. The transcripts produced by the 1w2p2pR^{w} vector show a similar pattern with primarily nuclear signal in the absence of Rev, and a strong cytoplasmic env transcript signal in the presence of Rev. The transcripts produced by the 2p2p3pR^{w} vector, however, have cytoplasmic signal both in the absence and presence of Rev. In the presence of Rev, the nuclear signal decreases.

The presence of cytoplasmic fluorescence in the absence of Rev indicates that the while 2p2p3pR^{w} vector transcript can be exported independently of Rev-RRE mediation, Rev still has a significant effect on the nucleo-cytoplasmic distribution of the Env transcript. The 1w2p2pR^{w} vector has a similar nuclear export pattern to the wwwR^{w} construct, showing that it is Rev dependent. When compared to the nuclear export pattern of the 2p2p2pR^{w} vector, which is weakly exported from the nucleus in the presence and absence of Rev, we see that CRS1 wildtype its sufficient to restore Rev dependence, while the CRS3 positive sequence results in a clear cytoplasmic signal both in the presence and absence of Rev. As previously discussed in chapter 3.2, this nuclear export pattern could be due to varying nuclear export functions of the proteins binding these CRS that are not defined by the HEXplorer algorithm.

A



B

Mutant	Per/nt	Total
2p2p2pR^{w}	9.13	15469.81
1w2p2pR^{w}	6.12	10232.61
2p2p3pR^{w}	8.89	13709.70

C

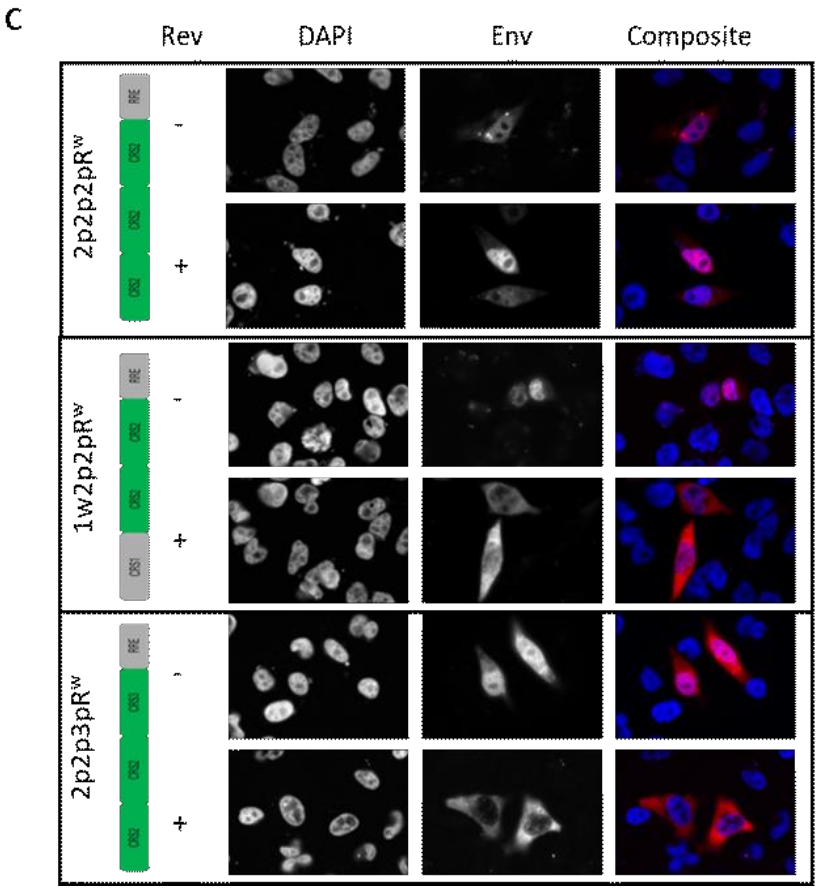


Figure 3.4.2. Single CRS reverses the nuclear export pattern of the CRS2 concatenation vector. The CRS2 cis-acting repressive sequence subregion found in HIV-1 env was mutated according to HEXplorer algorithm and concatenated to produce an Env mutant containing two HEXplorer optimized CRS2 subregions and either the HEXplorer optimized CRS3 or wildtype CRS1 in place of the wildtype CRS, altering the underlying amino acid sequence. HeLa cells were seeded 24 hours prior to transfection at 1.0 x 10^5 cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested 24 hours after transfection. The cells were fixed in 3.7% Formaldehyde and incubated overnight at 37°C with Cy-3 fluorescently labeled RNA probes complementary to the unmutated regions of the Env gene. The nuclei of the cells were stained with DAPI and mounted to glass microscope slides using Vectashield mounting medium. A) Schematics of the wildtype and mutant Env vectors along with their HEXplorer plot, where green indicates HEXplorer optimized, and grey indicates the wildtype sequence. Mutants are referred to the name on the left which indicates positive fragments with a (p) and wildtype with a (w). B) The table shows total and per nucleotide HEXplorer score of the overall CRS regions in A. C) The Env vectors in A were transfected in HeLa cells and analyzed via FISH for cellular localization of the Env transcript.

Figure 3.4.3 shows the changes in cytoplasmic fluorescence between the condition without Rev and in the presence of Rev. The cytoplasmic fluorescence produced by the wildtype *env* transcript increases by 1.74-fold in the presence of Rev, indicating that the presence of Rev increases nuclear export, and thus the percentage of cytoplasmic fluorescence. The percentage of cytoplasmic fluorescence produced by the 2p2p2pR^w Env vector decreases by 0.62-fold in the presence of Rev, implying that Rev inhibits the export of the transcripts produced by the 2p2p2pR^w Env vector. However, when the wildtype CRS1 is reintroduced in the 1w2p2pR^w mutant, the cytoplasmic fluorescence increases by 2.66-fold. Taken together with the data from figure 3.4.2.C, this indicates that the wildtype CRS1 can override the double concatenation, restoring the transcript's reliance on Rev for nuclear export. In contrast, the 2p2p3pR^w Env vector has moderate 1.29-fold increase in cytoplasmic fluorescence when Rev is added, showing that it is Rev-independent. This indicates that by replacing one of the HEXplorer optimized CRS2 sequences with a HEXplorer optimized CRS3 subregion, nuclear export is no longer inhibited. As shown with CRS1 in Chapter 3.2, this data suggests that the specific RNA binding proteins predicted by HEXplorer score may have varying effects on nuclear export that cannot be fully defined by the HEXplorer algorithm. In this case, CRS2 may be primarily bound by non-shuttling proteins.

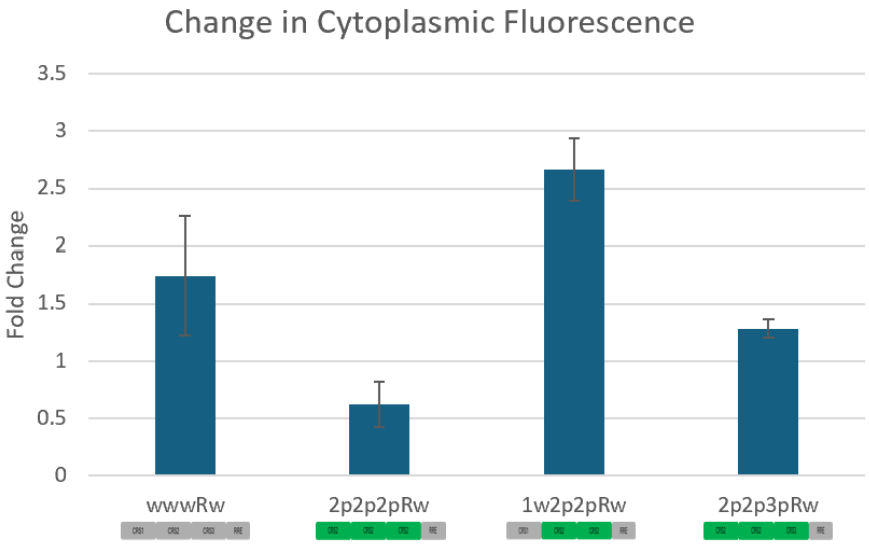


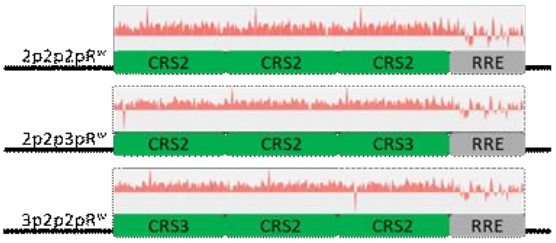
Figure 3.4.3. The change in cytoplasmic fluorescence of concatenated HEXplorer optimized CRS Env mutants. The CRS2 cis-acting repressive sequence subregion found in HIV-1 env was mutated according to HEXplorer algorithm and concatenated to produce an Env mutant containing two HEXplorer optimized CRS2 subregions and either the HEXplorer optimized CRS3 or wildtype CRS1 in place of the wildtype CRS, altering the underlying amino acid sequence. HeLa cells were seeded 24 hours prior to transfection at 1.0 x 10^5 cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested 24 hours after transfection. The cells were fixed in 3.7% Formaldehyde and incubated overnight at 37°C with Cy-3 fluorescently labeled RNA probes complementary to the unmutated regions of the Env gene. The nuclei of the cells were stained with DAPI and mounted to glass microscope slides using Vectashield mounting medium. The cytoplasmic fluorescence of 20 cells from each condition from two biological replicates was calculated and displayed as a ratio of the change in fluorescence between two conditions of the same construct.

In figure 3.4.2, the re-introduction of the CRS3 positive vector in the 2p2p3pR^w mutant results in significantly more cytoplasmic fluorescence. To investigate if this effect is due to the proximity of CRS3 with the RRE, the 3p2p2pR^w mutant was created. The 3p2p2pR^w and 2p2p3pR^w Env vectors were transfected into HeLa cells with either a mock plasmid or a Rev expression vector to be analyzed via RNA FISH to determine the subcellular localization of the transcripts in the presence and absence of Rev and imaged via fluorescence microscopy. Ten cells from each condition were imaged and analyzed to determine the percentage of fluorescent Env mRNA found in the cytoplasm compared to the nucleus.

Figure 3.4.4.A shows a schematic of the 2p2p2pR^w Env vector, and the partial CRS2 concatenated Env vectors, 2p2p3pR^w and 3p2p2pR^w. The amino acid sequence of these partially concatenated Env vectors was altered, thus this vector cannot produce Env protein. This figure also shows the HEXplorer score graph of the mutated region that is analyzed in Figure 3.4.2.B. In figure 3.4.2.B, the HEXplorer score of the CRS and RRE region of the 2p2p2pR^w Env vector is compared to the HEXplorer score of the mutated region

and RRE of the CRS2 concatenated Env vector. While the 2p2p2pR^{w} CRS and RRE region has a HEXplorer score of 9.13 per nt, the 2p2p3pR^{w} vector has a slightly decreased HEXplorer score of 8.89 per nt and the 3p2p2pR^{w} vector has a similar HEXplorer score of 8.85 per nt. In total, the 2p2p2pR^{w} Env vector has a total HEXplorer score of 15,469.81, compared to the total HEXplorer score of 13,709.70 for the 2p2p3pR^{w} vector and 13,783.57 for the 3p2p2pR^{w} vector. Figure 3.4.1.C shows that the transcripts produced by the 2p2p2pR^{w} inefficiently exported in the absence and presence of Rev, indicating that this inefficient export is not mediated by Rev. In the presence of Rev, however, the strength of the nuclear signal increased. The transcripts produced by the 2p2p3pR^{w} vector, however, have cytoplasmic signal both in the absence and presence of Rev, though export is more efficient in the presence of Rev. RNA produced by the 3p2p2pR^{w} vector shows extensive accumulation in the cytoplasm in the absence and presence of Rev. The lack of change in cytoplasmic fluorescence of both the 2p2p3pR^{w} and 3p2p2pR^{w} vectors indicates that the transcript is exported independently of Rev-RRE mediation. Figure 3.4.4.D shows the changes in cytoplasmic fluorescence between the condition without Rev and in the presence of Rev. The cytoplasmic fluorescence of the wildtype Env vector increases by 1.74-fold in the presence of Rev, indicating that the presence of Rev increases nuclear export, and thus the percentage of cytoplasmic fluorescence. However, the percentage of cytoplasmic fluorescence produced by 2p2p2pR^{w} Env vector decreases by 0.62-fold in the presence of Rev, implying that Rev may inhibit the export of the transcripts produced by the 2p2p2pR^{w} Env vector. The 2p2p3pR^{w} Env vector has moderate 1.29-fold increase in cytoplasmic fluorescence when Rev is added. When the HEXplorer optimized CRS3 fragment is placed upstream of the concatenated double HEXplorer optimized CRS2 fragment, the cytoplasmic fluorescence increases by 1.13-fold. This indicates that by replacing one of the HEXplorer optimized CRS2 sequences with a HEXplorer optimized CRS3 subregion, nuclear export is no longer inhibited, and that this effect is not dependent on proximity to the RRE.

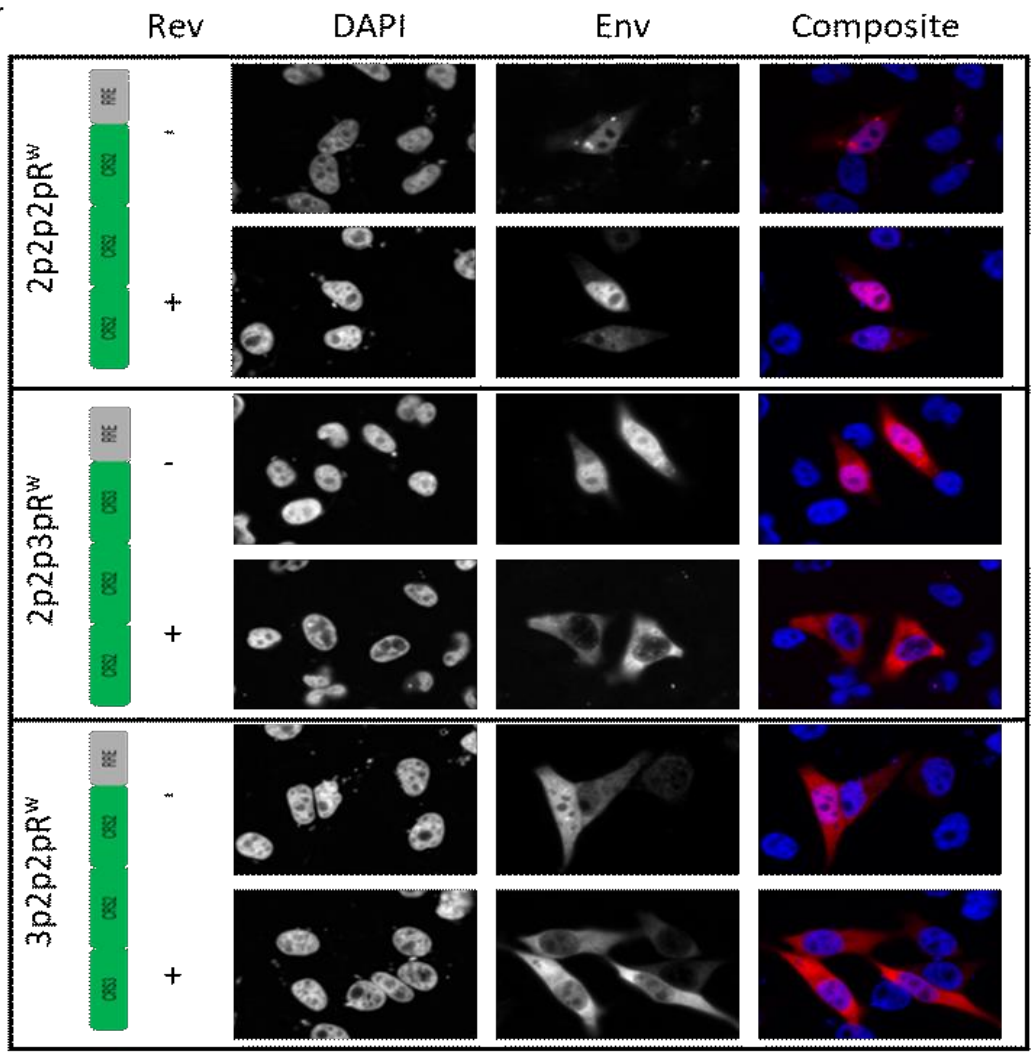
A



B

Mutant	Per/nt	Total
2p2p2pR^{w}	9.13	15469.81
2p2p3pR^{w}	8.89	13709.70
3p2p2pR^{w}	8.85	13783.57

C



D

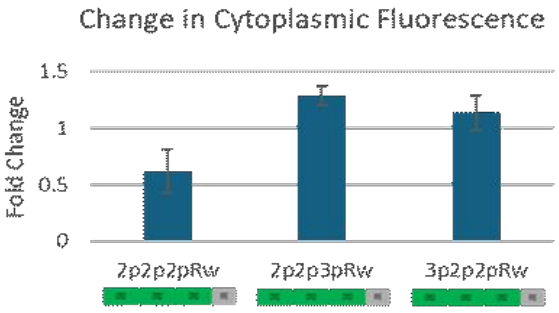


Figure 3.4.4. The nuclear export signal of CRS3 positive is position independent. The CRS2 cis-acting repressive sequence subregion found in HIV-1 env was mutated according to HEXplorer algorithm predictions without altering the underlying amino acid sequence and concatenated to produce a vector containing two HEXplorer optimized CRS2 subregions and the HEXplorer optimized CRS3 in place of the wildtype CRS1. HeLa cells were seeded 24 hours prior to transfection at 1.0 x 10^5 cells per mL into wells containing sterile glass coverslips. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Env vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector and harvested 24 hours after transfection. The cells were fixed in 3.7% Formaldehyde and incubated overnight at 37°C with Cy-3 fluorescently labeled RNA probes complementary to the unmutated regions of the Env gene. The nuclei of the cells were stained with DAPI and mounted to glass microscope slides using Vectashield mounting medium. A) Schematics of the wildtype and mutant Env vectors along with their HEXplorer plot, where green indicates HEXplorer optimized, and grey indicates the wildtype sequence. Mutants are referred to the name on the left which indicates positive fragments with a (p) and wildtype with a (w). B) The table shows total and per nucleotide HEXplorer score of the overall CRS regions in A. C) The Env vectors in A were transfected in HeLa cells and analyzed via FISH for cellular localization of the Env transcript. D) The cytoplasmic fluorescence of 20 cells from each condition in two biological replicates was calculated and displayed as a ratio of the change in fluorescence between two conditions of the same construct.

Overall, this data shows that while the quantity of HEXplorer optimized sequence influences nuclear export patterns and pathways, shorter sequences with specialized nuclear export signals also impact nuclear export patterns. Nuclear export regulation is complex, and pathway selection likely relies on several factors, including both RNA binding protein environment and specific sequence encoded signals. While previous chapters show that altering the RNA-binding protein environment via HEXplorer-guided mutation can induce a change in nuclear export pattern, this data shows that specialized sequences still need to be considered when predicting nuclear export patterns.

3.5 Env Mutants limit viral entry

The subgenomic Env expression vector encodes the envelope protein of HIV-1. The HIV-1 envelope protein is responsible for mediating viral entry into the host cell by interacting with host cell receptors such as CD4 to fuse the viral and host cell membrane, releasing the viral capsid into the host cytoplasm. Env is expressed in the late phase of HIV-1 gene expression, thus the timing and abundance of protein produced is strictly regulated. Altering the underlying HEXplorer score potentially disrupts the strict control maintained over Env expression, which could lead to changes in the ability of virions containing Env mutants to mediate membrane fusion. To investigate the effect of HEXplorer optimized Env expression vectors have on the ability of these virions to mediate viral entry and thus infectivity of the virion, pseudotyped virions were produced using the HEXplorer optimized expression vectors and used to transduce TZM-bL cells. Luciferase activity was analyzed 24 hours after transduction.

Figure 3.5.1.A shows a schematic showing the production of pseudoviral particles generated with the HEXplorer optimized Env expression vectors. In figure 3.5.1.A, HEK293T cells are cotransfected with the HEXplorer optimized Env expression vectors pNL4-3 ΔE GFP to produce pseudoviral particles generated with the Env produced by the mutant Env expression plasmids. TZM-bL cells, which are stably transfected with luciferase under a Tat-responsive promoter, were then transduced with equivalent quantities of pseudoviral particles, determined by p24 ELISA. The samples were subsequently analyzed for firefly luciferase 24 hours after transduction. Figure 3.5.1.B shows the relative luciferase activity induced by the pseudoviral mutants in relative light units (RLU), normalized to the luciferase activity induced by virions produced with the wildtype Env. The pseudoviral particles generated by cotransfecting pNL4-3 ΔE GFP with a mock plasmid were used as a negative control. All luciferase values were normalized to the wildtype pseudoviral particles when diluted by 2. The mock pseudovirus induces similar luciferase activity regardless of dilution. In contrast the pseudoviral particles produced with the wildtype Env expression vector induce significantly more luciferase activity. The luciferase activity produced by the wildtype pseudoviral particles was set to 100% when diluted two-fold. Accordingly, the luciferase activity nearly halves (55.72%) when the pseudoviral particles are diluted by 4 and decreases even further to 32.06% when diluted by 8. This pattern is expected, due to the decrease in viral particles per well in the dilution series. The HEXplorer optimized Env expression vector, $pppR^{w}$ produces significantly less luciferase activity compared to the wildtype Env pseudoviral particles. When diluted by 2, the $pppR^{w}$ pseudoviral particles produce 31.09% luciferase activity compared to the wildtype at the same dilution. This is comparable to the luciferase activity produced by the wildtype when diluted by 8. The luciferase activity produced by the $pppR^{w}$ pseudoviral particles continues to decrease when further diluted, to 18.04% when diluted by 4 and 14.37% when diluted by 8. The pseudoviral particles produced by the HEXplorer optimized Env expression vector, $pppR^{p}$, also induces significantly less luciferase activity compared to the wildtype pseudoviral particles. When diluted by 2, these pseudovirions produce 10.07% of the luciferase activity produced by the wildtype at the same dilution. This decreases to 8.43% when diluted by 4 and 8.81% when diluted by 8. The pseudoviral particles generated with $pppR^{p}$ seem to produce little to no luciferase activity when compared to the mock pseudoviral particles, while the luciferase activity induced by the $pppR^{w}$ pseudoviral particles is low, but still measurable. The decrease in ability to induce luciferase activity correlates with the decrease of Env protein production shown in figure 3.1.4, indicating that quantity of Env protein included in the pseudoviral particles correlates to the ability of the virion to induce membrane fusion.

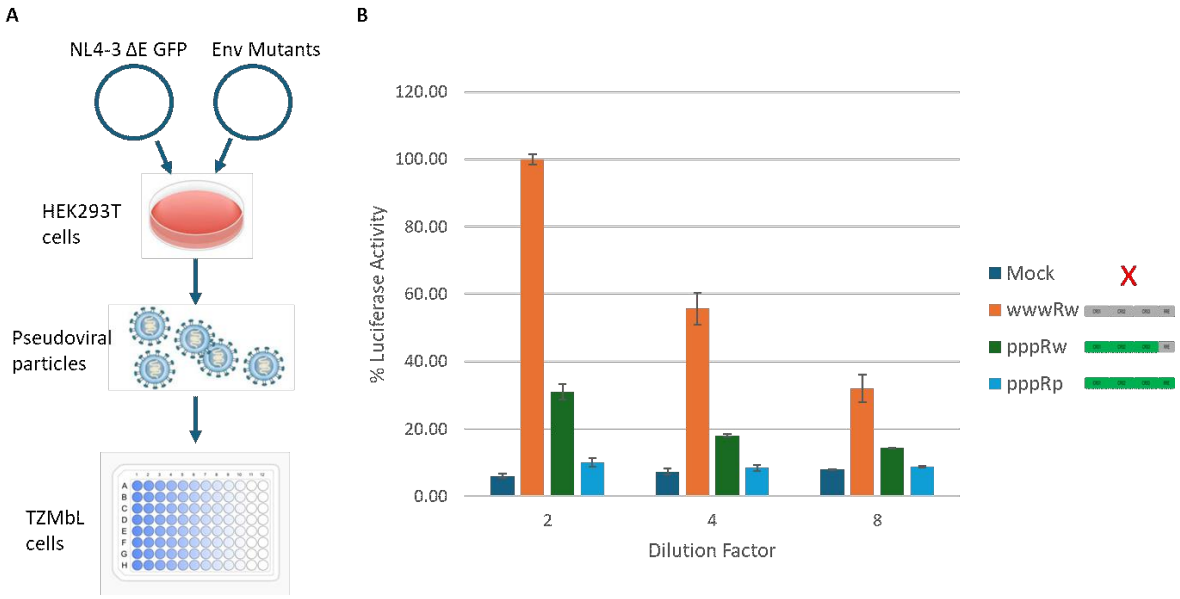


Figure 3.5.1 Virions produced with the HEXplorer optimized Env expression vectors are limited in their ability to infect permissive cells. Pseudoviral particles were generated using the HEXplorer optimized Env expression vector and used to transduce TZMbL cells. TZM-bL cells are stably transfected with a HIV-1 Tat driven Firefly Luciferase expression vector that uses the HIV-1 LTR sensitive to HIV-1 Tat expression. HEK293T cells were plated at a concentration of 3.0 x 10^5 24 hours before transfection. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of the Env expression plasmids per 10 mL of cell culture. An HIV-1 producing plasmid where the Env open reading frame is replaced with the GFP ORF, pNL4-3 ΔE GFP, was cotransfected with the wildtype and HEXplorer optimized Env vectors at a ratio of 12:1. The transfected HEK293T cells were incubated for 48 hours at 37°C before the supernatant was harvested and filtered via a 0.45 µm syringe filter and aliquoted for storage at 80°C. These pseudoviral aliquots were thawed, analyzed for p24 capsid production via ELISA, and normalized according to the p24 concentration for subsequent transduction. TZM-bL cells were seeded 24 hours prior to transduction at 2.0 x 10^5 cells per mL in DMEM (10% FBS, 1% PenStrep). The media was removed from the TZM-bL cells and replaced with fresh media containing a serial dilution of the pseudoviral particles. The cells were incubated for 24 hours at 37°C, and then harvested using Lysis Juice and analyzed for firefly luciferase activity using a Tecan plate reader. A) A schematic showing the experimental plan and production of pseudoviral vectors. B) TZMbL cells were transduced via serial dilution with pseudovirons produced with the mutant Env expression vectors and analyzed for Luciferase production after 24 hours of incubation.

To investigate the effect that the HEXplorer optimized Env expression vector mutations have on the composition of these virions, pseudoviral virions were produced using the HEXplorer optimized expression vectors and normalized to the quantity of p24. These particles were concentrated via sucrose cushion and analyzed via western blot. Figure 3.5.2.A shows a western blot where the samples were analyzed for the presence of gp120, a protein produced by the Env transcript, and p24, a protein produced by pNL4-3 ΔE GFP. There is no gp120 produced by the Mock condition of positive control (VSVg) as expected, and they both produce a strong p24 signal. Interestingly, the p24 signal becomes weak in the presence of the wildtype and HEXplorer optimized Env expression vector. Contrarily, the quantity of gp120 increases, with the wildtype producing the least amount of gp120, and the $pppR^p$ mutant producing the most. Figure 3.5.2.B shows the quantification from

two separate assays. As seen in the western blot, the mock transfection and the positive control, VSVg, produce no gp120 protein, but they do produce more p24 than the wildtype and HEXplorer optimized Env pseudoviral particles at 23830 U for the Mock condition and 14242 U for the positive control. The wildtype pseudoviral particle appears to contain the least gp120 at 6484 U, and the most p24 of the pseudoviral particles produced with the mutant Env expression vectors at 4272 U. The HEXplorer optimized Env expression vectors $pppR^{w}$ and $pppR^{p}$ produce similar quantities of gp120 at 16355 U and 17155 U respectively. These mutants also appear to produce similar quantities of p24, with the $pppR^{w}$ pseudoviral particles producing 2255 U and $pppR^{p}$ pseudoviral particles producing 1546 U. Taken together, there appears to be a switch between p24 and gp120 presence when Env expression vectors are introduced. In the context of lentiviral production, this may be due to dysregulation of Env expression through the alteration of nuclear export pathway, leading to excess Env inclusion in budding lentiviral particles.

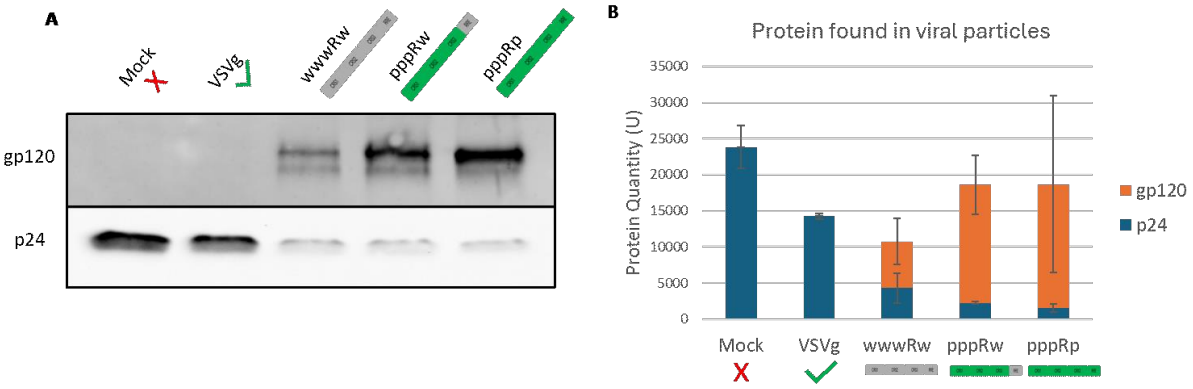


Figure 3.5.2 Quantity of protein present in pseudoviral particles generated with HEXplorer optimized Env expression vectors. HEK293T cells were plated at a concentration of 3.0 x 10^5 24 hours before transfection. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of the Env expression plasmids per 10 mL of cell culture. An HIV-1 producing plasmid where the Env open reading frame is replaced with the GFP ORF, pNL4-3 ΔE GFP, was cotransfected with the wildtype and HEXplorer optimized Env vectors at a ratio of 12:1. The transfected HEK293T cells were incubated for 48 hours at 37°C before the supernatant was harvested and filtered via a 0.45 µm syringe filter and aliquoted for storage at 80°C. These pseudoviral aliquots were thawed, analyzed for p24 capsid production via ELISA, and normalized according to the p24 concentration. The normalized pseudoviral stocks were concentrated by centrifuging with a 20% sucrose cushion at 50,000 rpm. The resulting pellet was lysed in RIPA buffer and analyzed via western blot. Two western blots were quantified for this figure. A) The pseudoviral particles produced by transfecting HEK293T cells with pNL4-3 ΔE GFP and the HEXplorer optimized Env expression vectors were concentrated via centrifugation with a 20% sucrose cushion and analyzed via western blot. B) The western blot was quantified to show comparative intensities of p24 (blue) and gp120 protein (orange).

The single and double HEXplorer optimized regions display various levels of Rev dependence in chapter 3.2, and produce varying levels of protein in the presence and absence of Rev. To investigate the effect that the HEXplorer optimized subregion Env expression vector mutations have on the ability of these virions to mediate viral entry and thus infectivity of the virion, pseudoviral virions were produced using the HEXplorer

optimized expression vectors, pwwR^w, ppwR^w, and ppR^w, and used to transduce TZM-bL cells which were analyzed for luciferase activity 24 hours after transduction. TZM-bL cells are stably transfected with a Firefly Luciferase expression vector that uses an inducible promoter sensitive to HIV-1 infection.

Figure 3.5.3 shows the relative luciferase activity induced by the pseudoviral mutants in relative light units (RLU). The pseudoviral particles generated by cotransfecting pNL4-3 ΔE GFP with a mock plasmid were used as a negative control. All luciferase values were normalized to the wildtype pseudoviral particles when diluted by 2. The mock pseudovirus induces similar luciferase activity regardless of dilution, producing 6.08% of the luciferase activity compared to the wildtype pseudovirus diluted by 2. When diluted by 4, it produces 7.30% and when diluted by 8, it produces 8.00% of the luciferase activity compared to the wildtype pseudovirus diluted by 2. In contrast, the pseudoviral particles produced with the wildtype Env expression vector induce significantly more luciferase activity. The wildtype pseudoviral particles produce 100% luciferase activity when diluted by 2, however this decreases to 55.72% when diluted by 4, and decreases even further to 32.06% when diluted by 8. This pattern is expected, due to the decrease in viral particles per well in the dilution series. Surprisingly, when diluted by 2, the pseudoviral particles generated with pwwR^w induce 157.16% compared to the wildtype pseudoviral particles at the same dilution. This then decreases to 89.95% when diluted by 4 and 54.56% when diluted by 8. It appears that the single positive pseudoviral particle induces more membrane fusion than wildtype pseudoviral particles. The pseudoviral particles produced with the ppwR^w Env expression vector on the other hand, only produce 26.56% compared to the wildtype pseudoviral particles at the same dilution of 2. This decreases to 16.38% when diluted by 4 and 12.43% when diluted by 8. When diluted by 2, the pppR^w pseudoviral particles produced with the HEXplorer optimized Env expression vector, pppR^w, induce 31.09% luciferase activity compared to the wildtype at the same dilution. This is comparable to the luciferase activity produced by the wildtype when diluted by 8. There is a clear trend between the Rev-dependent and Rev-independent Env expression vectors. It appears that the pseudoviral particles produced with the Rev-dependent wildtype and pwwR^w Env expression vectors are able to induce substantial luciferase activity, whereas the pseudoviral particles produced with the Rev-independent ppwR^w and pppR^w Env expression vectors produce very little.

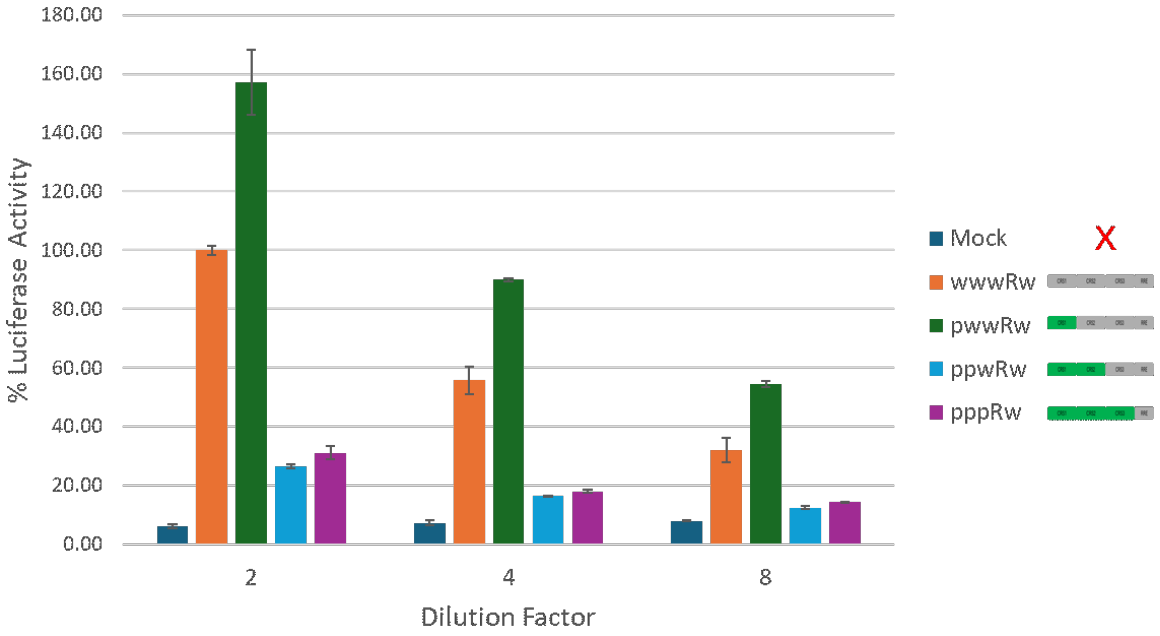


Figure 3.5.3 Virions produced with the HEXplorer optimized Env expression vectors are limited in their ability to infect permissive cells. HEK293T cells were plated at a concentration of 3.0 x 10^5 24 hours before transfection. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of the Env expression plasmids per 10 mL of cell culture. An HIV-1 producing plasmid where the Env open reading frame is replaced with the GFP ORF, pNL4-3 ΔE GFP, was cotransfected with the wildtype and HEXplorer optimized Env vectors at a ratio of 12:1. The transfected HEK293T cells were incubated for 48 hours at 37°C before the supernatant was harvested and filtered via a 0.45 µm syringe filter and aliquoted for storage at 80°C. These pseudoviral aliquots were thawed, analyzed for p24 capsid production via ELISA, and normalized according to the p24 concentration for subsequent transduction. TZM-bL cells were seeded 24 hours prior to transduction at 2.0 x 10^5 cells per mL in DMEM (10% FBS, 1% PenStrep). The media from the TZMbL cells was removed and replaced with fresh media containing a serial dilution of the normalized pseudoviral particles. The cells were incubated for 24 hours at 37°C and harvested using Lysis Juice and analyzed for firefly luciferase activity using a Tecan plate reader. B) TZM-bL cells were transduced via serial dilution with pseudovirons produced with the mutant Env expression vectors and analyzed for Luciferase production after 24 hours of incubation.

In order to investigate the effect that the pwwR^w and ppwR^w HEXplorer optimized Env expression vector mutations have on the composition of these virions, pseudoviral virions were produced using the HEXplorer optimized expression vectors and analyzed via western blot. Figure 3.5.4.A shows a western blot where the samples were analyzed for the presence of gp120, a protein produced by the Env transcript, and p24, a protein produced by pNL4-3 ΔE GFP. There is no gp120 produced by the Mock condition or positive control (VSVg) as expected, and they both produce a strong p24 signal. Interestingly, the p24 signal becomes weak in the presence of the wildtype and varying HEXplorer optimized Env expression vectors. Contrarily, the quantity of gp120 increases, with the wildtype producing the least amount of gp120, and the ppwR^w mutant producing the most. Figure 3.5.2.B shows the quantification from two separate assays. As seen in the in the blot, the mock transfection and the positive control, VSVg, produce no gp120 protein, but they do appear to produce more p24 than the wildtype and HEXplorer

optimized Env pseudoviral particles at 23830 U for the Mock condition and 14242 U for the positive control. The wildtype pseudoviral particle appears to contain the least gp120 at 6485 U, and the most p24 of the pseudoviral particles produced with the mutant Env expression vectors at 4272 U. The single HEXplorer optimized Env expression vector pwwR^{w} contains 9718 U of gp120 and 2406 U of p24. Pseudoviral particles produced with double HEXplorer optimized Env expression vector, ppwR^{w}, appears to contain both more gp120, and simultaneously less p24, with 13743 U gp120 and 632 U p24. These mutants also appear to contain less gp120 than the pppR^{w} pseudoviral particles, which contains 16355 U gp120, and 2255 U p24. Taken together, there appears to be a switch between p24 and gp120 presence when Env expression vectors are introduced.

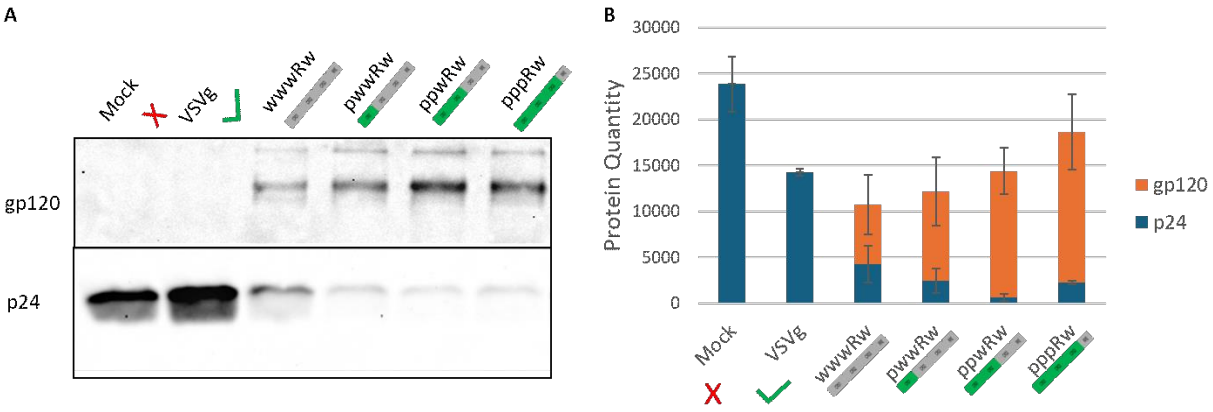


Figure 3.5.4 Quantity of protein present in pseudoviral particles generated with wildtype HEXplorer optimized Env expression vectors. HEK293T cells were plated at a concentration of 3.0 x 10^5 24 hours before transfection. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of the Env expression plasmids per 10 mL of cell culture. An HIV-1 producing plasmid where the Env open reading frame is replaced with the GFP ORF, pNL4-3 ΔE GFP, was cotransfected with the wildtype and HEXplorer optimized Env vectors at a ratio of 12:1. The transfected HEK293T cells were incubated for 48 hours at 37°C before the supernatant was harvested and filtered via a 0.45 µm syringe filter and aliquoted for storage at 80°C. These pseudoviral aliquots were thawed, analyzed for p24 capsid production via ELISA, and normalized according to the p24 concentration. The normalized pseudoviral stocks were concentrated by centrifuging with a 20% sucrose cushion at 50,000 rpm. The resulting pellet was lysed in RIPA buffer and analyzed via western blot. Two western blots were quantified for this figure. A) The pseudoviral particles produced by transfecting HEK293T cells with pNL4-3 ΔE GFP and the HEXplorer optimized Env expression vectors were concentrated via centrifugation with a 20% sucrose cushion and analyzed via western blot. B) The western blot was quantified to show comparative intensities of p24 (blue) and gp120 protein (orange).

Overall, this data shows that extensive mutation of the CRS found in intron 4 affects the infectability of virions produced with these HEXplorer optimized Env expression vectors. The decrease in luciferase activity seems to be associated with the Rev dependence of the Env expression vector used. Additionally, this trend in ability to induce luciferase activity does not appear to be correlated to the quantity of protein harvested from the viral stocks.

3.6 Rev may inhibit host gene expression

In figure 3.2.4, the addition of Rev appears to decrease the quantity of protein produced by both HEXplorer optimized Env expression vectors, $pppR^{w}$ and $pppR^{p}$. Additionally, in figure 3.4.3, the percentage of cytoplasmic fluorescence decreases in the presence of Rev. Taken together, this data implies that Rev may impede the expression of transcripts not reliant on CRM1 for export from the nucleus such as host cell transcripts. To investigate Rev's potential ability to inhibit host cell gene expression, HIV-1 sub genomic expression vectors encoding the *Gag-Pol* ORF, which are dependent on export elements for nuclear export, were developed and analyzed for Gag production in the absence and presence of Rev. These expression vectors were transfected in HeLa cells and analyzed for protein production in the absence and presence of Rev.

Figure 3.6.1.A shows a schematic of the Gag-Pol expression vector used to analyze Rev's effect on gene expression relative to the export pathway used. The promoter is shown in blue while the Gag-Pol ORF is shown in grey. At the 3' end of the Gag-Pol open reading frame the Export element is shown in orange. This export element encodes either the RRE export element found in HIV-1, or four concatenated constitutive transport elements (CTE) from the Mason Pfizer Monkey virus (Aibara, Katahira et al. 2015). The RRE mediates nuclear export by recruiting the CRM1 export factor through the virally encoded Rev protein, while the CTE element directly interacts with the NXF1 export factor to facilitate nuclear export of CTE containing transcripts. In addition to single pathway export elements, the RRE and 4XCTE were inserted into the same expression vector producing the Gag-Pol RRE-4XCTE and Gag-Pol 4XCTE-RRE expression vectors. Figure 3.6.1.B shows that the level of Gag protein produced by the RRE dependent vector increases by 2.68-fold in the presence of Rev, compared to the more modest increase of 1.33-fold for the 4XCTE-containing Gag-Pol expression vector in the presence of Rev. However, when the export elements are added together, the expression of Gag decreases in the presence of Rev. In the presence of Rev, the Gag-Pol expression of the vector containing the RRE upstream of the 4XCTE element decreases to 0.78-fold compared to in the presence of Rev, and the Gag-Pol expression of the vector containing the 4XCTE upstream of the RRE element decreases to 0.81-fold. This indicates that while the presence of Rev has little to no effect on the expression of a viral gene that uses an NXF1 recruitment element alone, transcripts that can recruit both CRM1 and NXF1 show a decrease in gene expression.

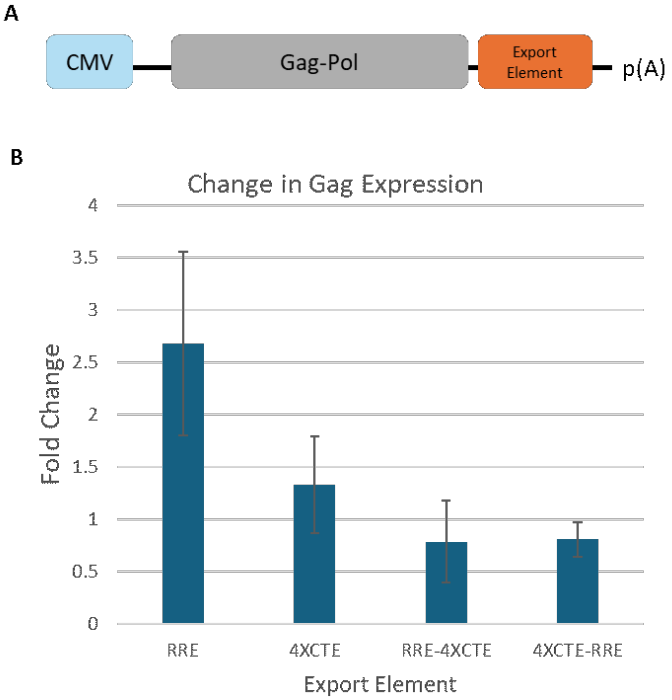


Figure 3.6.1 Rev may decrease gene expression of transcripts exported via NXF1. HeLa cells were seeded 24 hours prior to transfection at 1.5 x 10^5 cells per mL. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The Gag-Pol expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector. After 24 hours, the cells were harvested using 0.1% Triton-X100 in PBS for analysis via ELISA. A) A schematic of the Gag-Pol expression vector, where the Promoter is shown in blue, the Gag-Pol open reading frame is shown in grey, and the export element is shown in orange. B) The change in Gag expression upon the addition of Rev of Gag-Pol expression vectors with varying export elements.

Although the presence of Rev showed negligible effect on the Gag-Pol expression vectors, it is possible that the inhibition is specific to non-viral genes. To investigate this, Renilla luciferase expression vectors dependent on export elements were developed and analyzed for luciferase production in the absence and presence of Rev. Luciferase expression vectors were transfected in HeLa cells and analyzed for luciferase activity in the absence and presence of Rev. Figure 3.6.2.A shows a schematic of the Renilla luciferase vector used to analyze Rev's effect on gene expression relative to the export pathway used. The promoter is shown in blue while the Renilla luciferase ORF is shown in grey. At the 3' end of the Luciferase open reading frame the Export element is shown in orange. The export element used was either the RRE export element found in HIV-1, or four concatenated constitutive transport elements (CTE) from the Mason Pfizer Monkey virus. The RRE mediates nuclear export by recruiting the CRM1 export factor through the virally encoded Rev protein, while the CTE element directly interacts with the NXF1 export factor to facilitate nuclear export of CTE containing transcripts. Figure 3.6.2.B shows the change in luciferase activity produced by the luciferase expression vector shown above using a CMV promoter and the RRE and CTE export elements in the presence and absence of Rev. When Rev is present, the luciferase activity produced by the RRE containing CMV-luciferase vector increases by 12.83-fold. The Luciferase expression vector that relies on

the 4XCTE export element shows little to no change in the presence of Rev, with a change in luciferase activity of 0.99-fold. This lack of change in the Luciferase-4XCTE expression vector may be because the CMV promoter is highly transcribed, which can obscure changes in gene expression due to post transcriptional regulation. To investigate whether the lack of visible effect is due to overexpression from the CMV reporter, Luciferase-export element expression vectors containing a weaker promoter, SV40, were made. Figure 3.6.2.C shows that when under an SV40 promoter, the luciferase activity produced by the RRE-containing Luciferase expression vector increases by 20.55-fold. The increase in change in luciferase activity upon the addition of Rev is 1.60-fold higher when using the SV40 promoter as compared to the CMV promoter, showing that post-transcriptional regulation has more impact on genes expressed using the SV40 promoter. The Luciferase activity produced by the 4XCTE-reliant Luciferase expression vector, however, decreases to 0.53-fold when Rev is present. This indicates that Rev is inhibiting the expression the SV40 Luciferase 4XCTE expression vector.

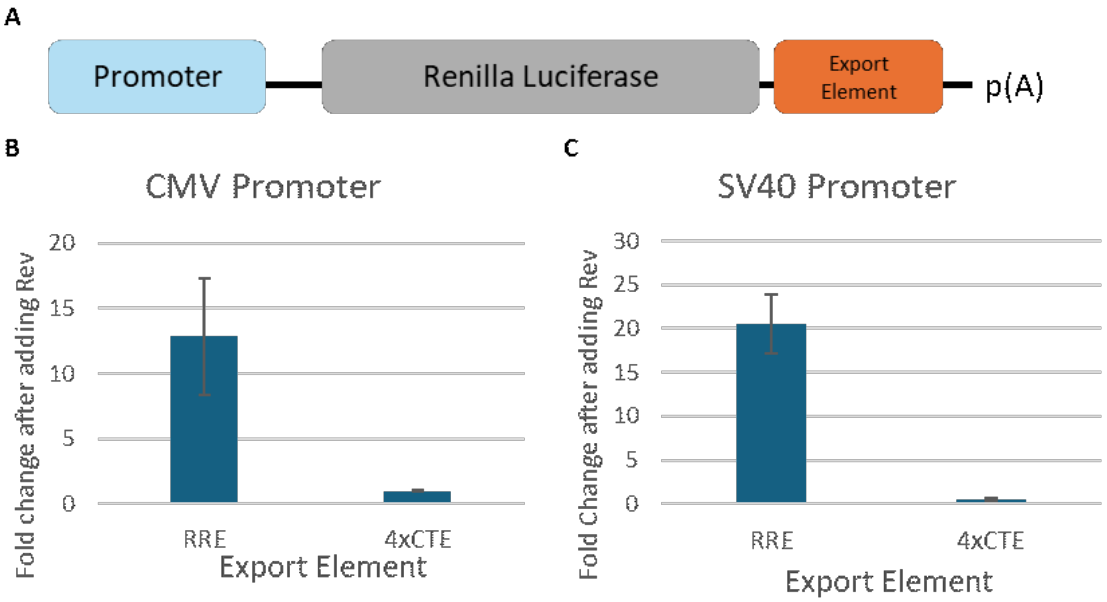


Figure 3.6.2 Rev decreases luciferase activity in transcripts exported via NXF1. HeLa cells were seeded 24 hours prior to transfection at 1.5 x 10^5 cells per mL. The transfection was carried out using Mirus LT1 transfection reagent with 0.5 µg of each plasmid per mL of cell culture. The luciferase expression vectors were cotransfected at a 1:1 ratio with either a mock plasmid or a Rev expression vector. The cells were harvested using 200 µL Lysis juice per mL cell culture 24 hours after transfection and 20 µL of each sample were analyzed using a Tecan plate reader in quadruplicate. A) A schematic of the Luciferase expression vector, where the Promoter is shown in blue, the luciferase open reading frame is shown in grey, and the export element is shown in orange. B) Fold change of Luciferase activity when Rev is present, and the Luciferase expression vectors use a CMV promoter. C) Fold change of Luciferase activity when Rev is present, and the Luciferase expression vectors use an SV40 promoter.

Overall, this data shows that Rev is capable of inhibiting gene expression of transcripts that use the NXF1 export pathway, although this effect seems to be restricted to genes transcribed under a weak promoter. Additionally, transcripts that can recruit both CRM1 and NXF1 export factors show a decrease in gene expression, potentially due to competition between the export pathways. This data indicates that while Rev is responsible for the export of late phase transcripts of HIV-1, it may also have a secondary role in inhibiting expression of host cell genes.

4 Discussion

This work aimed to identify and alter the function of sequences that regulate nuclear export patterns of HIV-1 late phase genes through bioinformatically guided adjustment of SRE predictions. The results of this work show that while mutating repressive sequences according to HEXplorer predictions can alter the nuclear export pathway, there are still specific sequence elements that exert further influence on nuclear export patterns. This work also illustrates the complexity of post-transcriptional gene regulation and the multi-faceted role that the underlying genetic code plays in RNA maturation and expression.

4.1 Mutating Cis-acting Repressive sequences according to HEXplorer results in a change in nuclear export pathway

After integration into the host cell genome, HIV-1 gene expression is tightly regulated and occurs in two phases, the early and late phase genes. Late phase genes are retained in the nucleus by *cis*-acting repressive sequences (CRS) until the early phase HIV-1 encoded protein, Rev, translocates to the nucleus. Analysis of previously identified HIV-1 sequences from literature described as inhibiting nuclear export for splicing regulatory element (SRE) potential using the HEXplorer algorithm showed that several of the previously mapped CRS had negative HEXplorer scores (Ostermann, Ritchie et al. 2021). Negative HEXplorer scores indicate that these sequences have a high potential of containing SREs typically found in intronic sequences, and thus RNA binding protein families that typically suppress downstream splice donor usage, such as members of the hnRNP protein family. In addition to regulating splice site usage, hnRNPs have been found to regulate nuclear export, typically by repressing it. Over the course of the 1990's, several CRS were found in the ORFs of late phase HIV-1 genes, such as *gag-pol*, and *env* (Rosen, Terwilliger et al. 1988, Cochrane, Jones et al. 1991, Schwartz, Felber et al. 1992). Later, potential *trans*-acting factors, such as hnRNP C, were also identified as binding to the CRS-containing RNAs (Suh, Seguin et al. 2003). The correlation between CRS binding of hnRNP proteins and HEXplorer prediction of SREs that interact with hnRNP proteins indicates that HEXplorer can analyze elements regulating nuclear export.

Overall, this thesis has shown that HEXplorer optimized mRNAs encoding Env are exported to the cytoplasm and produce functional Env protein in the absence of Rev. The quantity of HEXplorer optimized sequence seems to play a role in predicting Rev-independent nuclear export, which occurs independent of the canonical CRM1 nuclear export pathway. While the size of the region that is HEXplorer optimized plays a role in nuclear export patterns, there are also unrecognized sequence specific mechanisms involved in regulating nuclear export. In virions produced with these mutants, the

extensive mutation of the intron 4 CRS region affects the ability of virions produced with these HEXplorer optimized Env expression vectors to enter the host cells. Additionally, this thesis indicates that Rev is capable of inhibiting expression of transcripts that use the NXF1 export pathway, although this effect seems to be restricted to genes transcribed under a weak promoter. This indicates that while Rev is responsible for the export of late phase transcripts of HIV-1, it may also have a yet undescribed secondary role in inhibiting expression of host cell genes.

In initial experiments, transcripts produced by the LTR Env SD4- mutant $pppR^{w}$, containing the HEXplorer optimized sequence shown in figure 3.1.1, are found in the cytoplasm in the absence of Rev (figure 3.1.3). Additional mutation of the RRE to eliminate the secondary structure (figure 3.1.2) and HEXplorer optimize the RRE sequence produced the LTR Env SD4- $pppR^{p}$ mutant. Like the LTR Env SD4- $pppR^{w}$, this mutant is exported to the cytoplasm in the absence of Rev. However, unlike the $pppR^{w}$ mutant, this mutant is entirely unable to bind Rev to facilitate Rev-mediated export, as shown in figure 3.1.2.B. The presence of both mutant transcripts in the nucleus shows that bioinformatically optimizing sequences for the presence of exonic SREs can induce export of transcripts typically restrained to the nucleus. The major difference between the $pppR^{w}$ mutant and the $pppR^{p}$ mutant is the ability to interact with Rev. In figure 3.1.3.B, the cytoplasmic fluorescence produced by the $pppR^{w}$ mutant transcripts increases slightly in the presence of Rev. In contrast, the $pppR^{p}$ mutant, which is unable to interact with Rev, shows no change in nuclear export upon the addition of Rev. Since this mutant does not contain a functional RRE, the only export signal it can use is found in the HEXplorer optimized CRS region. This implies that while the HEXplorer optimized region can induce nuclear export, either by abolishing the binding of retention factors or by introducing binding sites for supporting factors, this export mechanism is not as efficient or optimal as the Rev-RRE system.

Later experiments showed that the wildtype is dependent on the Rev-CRM1 interaction for nuclear export, and export of these transcripts is accordingly inhibited by CRM1 inhibition. Nuclear export of the Rev-independent transcripts produced by the $pppR^{w}$ and $pppR^{p}$ mutants, however, is not hindered by CRM1 inhibition (figure 3.3.1). This finding indicates that by mutating the CRS, the *env* transcripts produced by these mutants are exported by a different pathway than the one they were originally destined for. This data provides insight into the underlying RBP code influencing nuclear export pathway selection, as well as mechanisms for retaining transcripts in the nucleus until certain conditions have been met, ex. the presence of Rev. While these Rev-independent Env expression vectors are not reliant on CRM1 for nuclear export, the exact nuclear export pathway is unclear.

Apart from the regulation of viral gene expression, eukaryotic cells can use intron retention to rapidly alter the concentration of specific mRNAs available for translation in the cytoplasm in response to stressors. The cell can use this mechanism in two ways. First, by stockpiling immediate early response transcripts in the nucleus through retention of an intron that contains a CRS and excising the retained intron in response to stress to release the transcript to the cytoplasm for subsequent translation. Or second, by retaining introns to keep a transcript in the nucleus in response to stress, so that they can be spliced and released during stress recovery (Wegener and Müller-McNicoll 2018). Consequently, the ability to predict or select nuclear export pathways as well as identify nuclear retention signals in mRNA sequences could lead to a better understanding of temporal gene expression regulation, both in response to stress and other factors.

4.2 Quantity vs Intensity of nuclear export signals

Mutating smaller sections of the CRS region to identify specific sequences that exerted greater influence over nuclear export pathways showed that while CRS1 appeared to have a slightly more significant role in supporting nuclear retention than the other two CRS fragments (figure 3.2.2), none of the individual CRS were entirely responsible for the switch from Rev-dependent expression to Rev-independent expression. When the larger region of the intron 4 CRS (over half) was mutated according to HEXplorer predictions, the switch from Rev-dependent to Rev-independent nuclear export was readily apparent in all variations of this mutant (figure 3.2.4). Based on this information, it appears that while some elements express a more significant nuclear retention signal, quantity of nuclear retention signal appears to have a greater impact on inducing Rev-dependent nuclear export.

Further investigation of the effect of nuclear export sequence intensity compared to the quantity of nuclear export sequence led to the creation of CRS subregion concatenation vectors. By concatenating the most positive HEXplorer optimized subregion, CRS2, the Rev-independence is not as strong as expected (figure 3.4.1), showing weak cytoplasmic fluorescence that appears to decrease in the presence of Rev. Not only does this suggest that expression of transcripts containing certain sequences are downregulated by Rev, but it also indicates that despite having a large HEXplorer optimized region, this sequence is not efficiently exported and may rely on the presence or interaction with other specific sequences. Supplementing the concatenated CRS2 regions with a HEXplorer optimized CRS3 sequence downstream of the double CRS2 sequence (2p2p3pR^{w}) resulted in an increase in cytoplasmic fluorescence, and a phenotype similar to that of the pppR^{w} mutant in figure 3.1.3 (figure 3.4.2 and 3.4.3). When this additional HEXplorer optimized sequence was moved upstream of the double concatenated CRS sequence, the phenotype remained the same (figure 3.4.4). This suggests that the effect of adding another HEXplorer optimized sequence is position independent. In contrast, adding the

wildtype CRS1 upstream of a double concatenated CRS2 resulted in a Rev-dependent phenotype, agreeing with the data from figure 3.2.2 that CRS1 exerts the strongest effect on nuclear export patterns, compared to the other *env* CRS subregions. Further investigation on the interplay between export signal intensity and quantity could include rearranging the pattern of the CRS1 containing CRS2 concatenated mutant to investigate the importance of relative position of this CRS to the strength of the signal. Additionally, dividing CRS1 into smaller sections within the 2p2p2pR^w mutant would further define the intense nuclear retention signal into a sequence that is able to be analyzed via RNA-pulldown and subsequent mass spectrometry analysis to identify the *trans*-acting factors involved.

When analyzing viral nuclear export regulatory sequences in HIV-1, whether they inhibit or support nuclear export, the sequences tend to be quite long, ranging from 48 nt to 692 nt, with the average being 285 nt (Ostermann, Ritchie et al. 2021). Taking into account the typical length of these regulatory sequences as well as the length of introns in nuclear stockpiled transcripts previously discussed, it is likely a combination of quantity and specific sequences that interact to exert targeted influence over mRNA fate (Wegener and Müller-McNicoll 2018). A further understanding of the interplay between specific retention sequences and the quantity of retention signal needed could shed light on diseases caused by dysregulated nuclear export, and potentially allow for treatment by promoting the use of another export pathway.

4.3 HIV-1 strictly regulates gene expression

HIV-1 separates gene expression into early and late phases by regulating nuclear export through splicing, where the CRS are spliced out in early phase HIV-1 genes and retained in late phase genes. By altering the CRS in the ORF of the *env* gene, the expression of this gene is no longer restricted to late phase expression, as it is exported to the cytoplasm in the absence of Rev. In figure 3.5.1, pseudoviral particles produced with Rev-independent Env expression vectors pppR^w and pppR^p are less capable of supporting viral entry than those produced with the wildtype Env expression vector. This pattern continues with pseudoviral particles produced with the pwwR^w and ppwR^w Env expression vectors, which are mostly Rev-dependent and Rev-independent respectively. The pseudoviral particles induce significantly less viral entry than those produced with the Rev-dependent vectors (figure 3.5.3). While this effect may also be attributed to the decrease in protein production, the increase in Env protein in pseudoviral particles produced with these mutants contradicts this (figure 3.5.4).

Jordan-Paiz and colleagues investigated the effect of codon optimization and deoptimization on Env protein production and found that while codon deoptimization has typically been used to generate attenuated viruses, they saw little change in Env protein

production. Although they thought that the synonymous mutations may have altered splicing through the mutation of SREs, they saw no alterations in splicing patterns. It is far more likely, considering the data shown in this work, that the optimization and deoptimization resulted in a disruption of the CRS regulating nuclear export patterns (Jordan-Paiz, Franco et al. 2021). In addition to being exported during late phase HIV-1 gene expression, the produced Env glycoprotein is cleaved and folded very slowly (Li, Luo et al. 2000). The retardation of *env* mRNA nuclear export and protein maturation indicates that the temporal regulation of the HIV-1 glycoprotein is important for HIV-1 fitness. So, while in addition to producing less protein, the disruption of the temporal nuclear export regulation could induce the decrease in viral entry shown in figures 3.5.1 and 3.5.3. While the exact reason for the strict temporal control HIV-1 retains over viral gene expression is unknown, it is not inconsequential.

The large CRS found in the *env* ORF is also found in the 9 kb class of HIV-1 mRNAs, in addition to multiple other CRS. While this work does not investigate the effect on HEXplorer guided mutation of CRS in the context of full proviral gene expression, the massive alteration of the SRE environment in intron 4 could possibly lead to changes in splicing pattern among other possible effects. Including the alteration of Env expression pattern and escape of temporal controls that HIV-1 uses to ensure gene expression fitness, the disruption of carefully balanced expression patterns could lead to an attenuation of viral production and fitness. The slight decrease in Env produced by the Rev-independent mutants in the presence of Rev (figure 3.1.4) and decrease in cytoplasmic fluorescence from the triple concatenated CRS 2 mutant in the presence of Rev (figure 3.4.3) indicates that the viral protein Rev may also have a secondary role in inhibiting host cell gene expression. In support of this, Taniguchi and colleagues found that Rev can suppress TAP/NXF1 binding of transcripts containing an RRE (Taniguchi, Mabuchi et al. 2014). Rev's inhibition of NXF1-dependent gene expression was further investigated, showing that, in specific conditions, Rev does inhibit expression of transcripts containing an NXF1 specific export element (figures 3.6.1 and 3.6.2). The exact conditions and mechanisms of this function are poorly understood and should be further investigated.

4.4 Altering the nuclear export pathway effects protein production

Initial experiments investigating the $pppR^{w}$ mutant showed that there is a slight increase in cytoplasmic fluorescence in the presence of Rev, indicating that Rev-independent export results in less cytoplasmic transcript. In figure 3.1.4, both the $pppR^{w}$ and $pppR^{p}$ mutants produce less protein both in the presence and absence of Rev than the wildtype in the presence of Rev. Despite producing less protein than the wildtype, the HEXplorer optimized mutants can produce the same size syncytia in the presence and absence of Rev, potentially indicating that this change in protein production doesn't appear to affect protein function. However, this may be due a saturation effect where Env expression over

a certain level has no effect on syncytium formation. A similar pattern emerges in the single and double HEXplorer optimized CRS subregion mutants, all the mutants produce less protein than the wildtype Env expression vector in the presence of Rev. The function of the protein also seems to fall along the lines Rev-dependent and Rev-independent mutants, though most of the Rev-independent mutants appear to produce smaller syncytia than the wildtype in the presence of Rev (figure 3.4.7).

The steps of gene expression are interconnected and many proteins that regulate gene expression mechanisms play roles in several processes. By altering the underlying RBP environment to change nuclear export patterns, the RBPs that regulate splicing, transcription, localization, and even translation have been altered (Dreyfuss, Matunis et al. 1993, Long and Caceres 2009). In addition to the suboptimal export of the mutant *env* transcripts, it is possible that the decrease in protein produced could also be due to dysregulation of other gene expression processes and needs to be further investigated. The consequences of less protein expressed by these Env expression vectors results in decreased ability to induce membrane fusion.

HIV-1 has very few envelope proteins on the surface of the virion, on average 8-10 Env trimers, which has been hypothesized to aid in evasion of humoral immunity (Klein and Bjorkman 2010, Stano, Leaman et al. 2017). As the Env glycoprotein trimers are responsible for mediating membrane fusion between the virion and host cell, or between two cells in the case of the syncytia assay, it is unexpected that the triple and double positive Env mutants can induce a similar syncytium formation in the presence and absence of Rev, and that the relative quantity of protein production doesn't appear to coincide with relative size of syncytia produced (figures 3.2.5 and 3.2.6). These mutants do, however, affect the entry capability of pseudoviral particles produced with them despite appearing to contain more Env protein (figures 3.5.1 and 3.5.2). This is incongruent with previous studies showing that pseudovirions with nearly 10-fold more Env trimers than the wildtype virus displayed both more infectivity and more activation of Env-specific B-cells (Stano, Leaman et al. 2017). The composition of these pseudoviral particles and their function needs to be further elucidated, particularly their potential ability to induce humoral immune responses. In addition to the composition of the pseudo viral particles, the conformational states and ability of the protein produced by these mutants to secrete Env glycoprotein should also be considered (Kalyanaraman, Pal et al. 1988, Nguyen, Wang et al. 2023).

4.5 Not-so-silent mutations and the underlying RNA-binding protein environment

As previously discussed, HIV-1 gene regulation relies on both splicing and nuclear export to separate and maintain temporal regulation of the late and early phases of gene

expression. In host gene expression as well as viral gene expression, these two steps are intertwined and RNA binding proteins that influence splicing also recruit nuclear export proteins (Liu and Mertz 1995, Chi, Wang et al. 2014, Khan, Hou et al. 2021). While originally discovered as splicing regulators, both the SR protein family and hnRNP protein family are now recognized as master regulators of gene expression. Though they play roles in several phases of gene expression, these protein families may interact to authorize transcripts for export or retain them in the nucleus. Mammalian cells have coupled gene expression steps into a pipeline and use RNA binding protein families to chaperone and sort transcripts through said pipeline (Cullen 2000, Valencia, Dias et al. 2008). The HEXplorer algorithm's ability to predict and analyze the code governing the underlying RNA protein binding environment to identify sequences that regulate mRNA maturation and expression is an important step in decoding the human genome at an epigenetic level and provides a better understanding of the effects caused by synonymous mutations.

Beyond the HIV-1 specific consequences of altered gene expression by these mutants, this data sheds light on synthetic gene optimization. Current gene optimization to produce recombinant proteins involves codon optimization, which involves replacing rare codons with more common codons to try and increase translation efficiency. While this does increase protein production in some cases, synonymous codons are not equivalent and can alter protein conformation and function through stalling or slowing the ribosome, which can alter folding pathways (Tsai, Sauna et al. 2008, Mauro 2018). As our understanding of gene expression progresses, synthetic genes and guided mutations are becoming more viable options to treat diseases and conditions that are caused by deleterious mutations and dysregulation of vital processes (Kabaria, Bae et al. 2024, Shi, Hamann et al. 2024). In addition to treating disease, the production of synthetic antibodies rely on our understanding of the RNA code underlying gene expression, not just codon optimization (Fridy, Li et al. 2014). Better understanding of the underlying RNA binding code and how it effects protein production could lead to more efficient and higher production of recombinant proteins used in a multitude of ways (de Marco 2015, Hunter, Yuan et al. 2019). While HEXplorer can optimize the underlying RBP environment to produce protein, it does not consider rare codon usage that could inhibit gene expression at the translation level like codon optimization. Further investigation of HEXplorer optimization that considers codon optimization and rare codon usage could be a step towards more optimal recombinant protein production for a variety of research and even potentially therapeutic uses.

4.6 Conclusion

In conclusion, mutating the CRS found in the HIV-1 ORF according to HEXplorer to bioinformatically optimize the presence of SREs lead to Rev-independent export of these

mutant *env* transcripts, as well as altering the nuclear export pathway of these transcripts. While further analysis showed that quantity of SRE optimized region was the most reliable factor for predicting a switch in nuclear export patterns of *env* mutants, there were also specific sequences involved in nuclear export regulation. Unsurprisingly, the change in nuclear export pathway also resulted in a decrease in protein production, though this may be due to rare codon usage and the viral lineage of the Env expression vector. This decrease in protein production had downstream consequences as virions produced with these mutants had significantly reduced membrane fusion capabilities. Overall, the data shown in this work provides initial insight into analysis and potential replication of elements that regulate gene expression. The work presented in this thesis provides insight into the regulation of several interconnected steps of gene expression and an algorithm that can predict and replicate elements controlling mRNA maturation and expression, along with its original purpose of predicting splice site usage.

5 Bibliography

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

Adachi, Y. and M. Yanagida (1989). "Higher order chromosome structure is affected by cold-sensitive mutations in a Schizosaccharomyces pombe gene crm1+ which encodes a 115-kD protein preferentially localized in the nucleus and its periphery." J Cell Biol **108**(4): 1195-1207.

Agafonov, D. E., M. van Santen, B. Kastner, P. Dube, C. L. Will, H. Urlaub and R. Lührmann (2016). "ATPγS stalls splicing after B complex formation but prior to spliceosome activation." Rna **22**(9): 1329-1337.

Agashe, D., N. C. Martinez-Gomez, D. A. Drummond and C. J. Marx (2013). "Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme." Mol Biol Evol **30**(3): 549-560.

Aibara, S., J. Katahira, E. Valkov and M. Stewart (2015). "The principal mRNA nuclear export factor NXF1:NXT1 forms a symmetric binding platform that facilitates export of retroviral CTE-RNA." Nucleic Acids Res **43**(3): 1883-1893.

Akari, H., M. Fujita, S. Kao, M. A. Khan, M. Shehu-Xhilaga, A. Adachi and K. Strebel (2004). "High level expression of human immunodeficiency virus type-1 Vif inhibits viral infectivity by modulating proteolytic processing of the Gag precursor at the p2/nucleocapsid processing site." J Biol Chem **279**(13): 12355-12362.

Akiri, G., D. Nahari, Y. Finkelstein, S. Y. Le, O. Elroy-Stein and B. Z. Levi (1998). "Regulation of vascular endothelial growth factor (VEGF) expression is mediated by internal initiation of translation and alternative initiation of transcription." Oncogene **17**(2): 227-236.

Alcázar-Román, A. R., E. J. Tran, S. Guo and S. R. Wente (2006). "Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export." Nat Cell Biol **8**(7): 711-716.

Arhel, N. (2010). "Revisiting HIV-1 uncoating." Retrovirology **7**: 96.

Asang, C., I. Hauber and H. Schaal (2008). "Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer." Nucleic Acids Res **36**(5): 1450-1463.

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**(50): 33414-33422.

Ayyildiz, D., G. Bergonzoni, A. Monziani, T. Tripathi, J. Döring, E. Kerschbamer, F. Di Leva, E. Pennati, L. Donini, M. Kovalenko, J. Zasso, L. Conti, V. C. Wheeler, C. Dieterich, S. Piazza, E. Dassi and M. Biagioli (2023). "CAG repeat expansion in the Huntington's disease gene shapes linear and circular RNAs biogenesis." PLoS Genet **19**(10): e1010988.

Azmi, A. S., A. Aboukameel, B. Bao, F. H. Sarkar, P. A. Philip, M. Kauffman, S. Shacham and R. M. Mohammad (2013). "Selective inhibitors of nuclear export block pancreatic cancer cell proliferation and reduce tumor growth in mice." Gastroenterology

144(2): 447-456.

Bachi, A., I. C. Braun, J. P. Rodrigues, N. Panté, K. Ribbeck, C. von Kobbe, U. Kutay, M. Wilm, D. Görlich, M. Carmo-Fonseca and E. Izaurralde (2000). "The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates." Rna **6**(1): 136-158.

Banerji, J., S. Rusconi and W. Schaffner (1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." Cell **27**(2 Pt 1): 299-308.

Barash, Y., J. A. Calarco, W. Gao, Q. Pan, X. Wang, O. Shai, B. J. Blencowe and B. J. Frey (2010). "Deciphering the splicing code." Nature **465**(7294): 53-59.

Barré-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vézinet-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**(4599): 868-871.

Barrera, A., H. Ramos, J. Vera-Otarola, L. Fernández-García, J. Angulo, V. Olguín, K. Pino, A. J. Mouland and M. López-Lastra (2020). "Post-translational modifications of hnRNP A1 differentially modulate retroviral IRES-mediated translation initiation." Nucleic Acids Res **48**(18): 10479-10499.

Bekenstein, U. and H. Soreq (2013). "Heterogeneous nuclear ribonucleoprotein A1 in health and neurodegenerative disease: from structural insights to post-transcriptional regulatory roles." Mol Cell Neurosci **56**: 436-446.

Belsham, G. J. and N. Sonenberg (2000). "Picornavirus RNA translation: roles for cellular proteins." Trends Microbiol **8**(7): 330-335.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber and E. S. Lander (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-326.

Black, A. C., J. Luo, C. Watanabe, S. Chun, A. Bakker, J. K. Fraser, J. P. Morgan and J. D. Rosenblatt (1995). "Polypyrimidine tract-binding protein and heterogeneous nuclear ribonucleoprotein A1 bind to human T-cell leukemia virus type 2 RNA regulatory elements." J Virol **69**(11): 6852-6858.

Black, D. L. (2003). "Mechanisms of alternative pre-messenger RNA splicing." Annu Rev Biochem **72**: 291-336.

Blanchet, S. and N. Ranjan (2022). Translation Phases in Eukaryotes. Ribosome Biogenesis: Methods and Protocols. E. KD. New York, Humana**:** 217-228.

Boehm, V. and N. H. Gehring (2016). "Exon Junction Complexes: Supervising the Gene Expression Assembly Line." Trends Genet **32**(11): 724-735.

Bohne, J., H. Wodrich and H. G. Kräusslich (2005). "Splicing of human immunodeficiency virus RNA is position-dependent suggesting sequential removal of introns from the 5' end." Nucleic Acids Res **33**(3): 825-837.

Borden, K. L. B. (2020). "The Nuclear Pore Complex and mRNA Export in Cancer." Cancers **13**(42).

Brennan, C. M., I. E. Gallouzi and J. A. Steitz (2000). "Protein ligands to HuR modulate its interaction with target mRNAs in vivo." J Cell Biol **151**(1): 1-14.

Briggs, J. A. and H. G. Kräusslich (2011). "The molecular architecture of HIV." J Mol Biol **410**(4): 491-500.

Brigham, B. S., J. P. Kitzrow, J. C. Reyes, K. Musier-Forsyth and J. B. Munro (2019). "Intrinsic

conformational dynamics of the HIV-1 genomic RNA 5'UTR." Proc Natl Acad Sci U S A **116**(21): 10372-10381.

Brighty, D. W. and M. Rosenberg (1994). "A cis-acting repressive sequence that overlaps the Rev-responsive element of human immunodeficiency virus type 1 regulates nuclear retention of env mRNAs independently of known splice signals." Proc Natl Acad Sci U S A **91**(18): 8314-8318.

Brillen, A. L., L. Walotka, F. Hillebrand, L. Müller, M. Widera, S. Theiss and H. Schaal (2017). "Analysis of Competing HIV-1 Splice Donor Sites Uncovers a Tight Cluster of Splicing Regulatory Elements within Exon 2/2b." J Virol **91**(14).

Bringmann, P., B. Appel, J. Rinke, R. Reuter, H. Theissen and R. Lührmann (1984). "Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing." Embo j **3**(6): 1357-1363.

Burdick, R. C., C. Li, M. Munshi, J. M. O. Rawson, K. Nagashima, W. S. Hu and V. K. Pathak (2020). "HIV-1 uncoats in the nucleus near sites of integration." Proc Natl Acad Sci U S A **117**(10): 5486-5493.

Campbell, E. M. and T. J. Hope (2015). "HIV-1 capsid: the multifaceted key player in HIV-1 infection." Nat Rev Microbiol **13**(8): 471-483.

Caputi, M., M. Freund, S. Kammler, C. Asang and H. Schaal (2004). "A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression." J Virol **78**(12): 6517-6526.

Carlson, L. A., J. A. Briggs, B. Glass, J. D. Riches, M. N. Simon, M. C. Johnson, B. Müller, K. Grünewald and H. G. Kräusslich (2008). "Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis." Cell Host Microbe **4**(6): 592-599.

Cartegni, L., S. L. Chew and A. R. Krainer (2002). "Listening to silence and understanding nonsense: exonic mutations that affect splicing." Nat Rev Genet **3**(4): 285-298.

Cereda, M., U. Pozzoli, G. Rot, P. Juvan, A. Schweitzer, T. Clark and J. Ule (2014). "RNAmotifs: prediction of multivalent RNA motifs that control alternative splicing." Genome Biol **15**(1): R20.

Chang, S. T., P. Sova, X. Peng, J. Weiss, G. L. Law, R. E. Palermo and M. G. Katze (2011). "Next-generation sequencing reveals HIV-1-mediated suppression of T cell activation and RNA processing and regulation of noncoding RNA expression in a CD4+ T cell line." mBio **2**(5).

Charenton, C., M. E. Wilkinson and K. Nagai (2019). "Mechanism of 5' splice site transfer for human spliceosome activation." Science **364**(6438): 362-367.

Charneau, P., M. Alizon and F. Clavel (1992). "A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication." J Virol **66**(5): 2814-2820.

Chaudhury, A., P. Chander and P. H. Howe (2010). "Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles." Rna **16**(8): 1449-1462.

Checkley, M. A., B. G. Luttge and E. O. Freed (2011). "HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation." J Mol Biol **410**(4): 582-608.

Chen, C. Y. and P. Sarnow (1995). "Initiation of protein synthesis by the eukaryotic

translational apparatus on circular RNAs." Science **268**(5209): 415-417.
Chen, C. Y. and A. B. Shyu (1995). "AU-rich elements: characterization and importance in mRNA degradation." Trends Biochem Sci **20**(11): 465-470.
Chen, C. Y., N. Xu and A. B. Shyu (1995). "mRNA decay mediated by two distinct AU-rich elements from c-fos and granulocyte-macrophage colony-stimulating factor transcripts: different deadenylation kinetics and uncoupling from translation." Mol Cell Biol **15**(10): 5777-5788.
Chi, B., K. Wang, Y. Du, B. Gui, X. Chang, L. Wang, J. Fan, S. Chen, X. Wu, G. Li and H. Cheng (2014). "A Sub-Element in PRE enhances nuclear export of intronless mRNAs by recruiting the TREX complex via ZC3H18." Nucleic Acids Res **42**(11): 7305-7318.
Cho, E. J., T. Takagi, C. R. Moore and S. Buratowski (1997). "mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain." Genes Dev **11**(24): 3319-3326.
Chun, T. W., S. Moir and A. S. Fauci (2015). "HIV reservoirs as obstacles and opportunities for an HIV cure." Nat Immunol **16**(6): 584-589.
Ciafrè, S. A. and S. Galardi (2013). "microRNAs and RNA-binding proteins: a complex network of interactions and reciprocal regulations in cancer." RNA Biol **10**(6): 935-942.
Ciuffi, A. (2016). "The benefits of integration." Clin Microbiol Infect **22**(4): 324-332.
Clark, F. and T. A. Thanaraj (2002). "Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human." Hum Mol Genet **11**(4): 451-464.
Cléry, A., M. Blatter and F. H. Allain (2008). "RNA recognition motifs: boring? Not quite." Curr Opin Struct Biol **18**(3): 290-298.
Clutter, D. S., M. R. Jordan, S. Bertagnolio and R. W. Shafer (2016). "HIV-1 drug resistance and resistance testing." Infect Genet Evol **46**: 292-307.
Cochrane, A. W., K. S. Jones, S. Beidas, P. J. Dillon, A. M. Skalka and C. A. Rosen (1991). "Identification and characterization of intragenic sequences which repress human immunodeficiency virus structural gene expression." J Virol **65**(10): 5305-5313.
Company, M., J. Arenas and J. Abelson (1991). "Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes." Nature **349**(6309): 487-493.
Corley, M., M. C. Burns and G. W. Yeo (2020). "How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms." Mol Cell **78**(1): 9-29.
Cullen, B. R. (2000). "Connections between the processing and nuclear export of mRNA: evidence for an export license?" Proc Natl Acad Sci U S A **97**(1): 4-6.
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**(6): 824-834.
Davies, M. N. and D. R. Flower (2007). "Harnessing bioinformatics to discover new vaccines." Drug Discov Today **12**(9-10): 389-395.
de Marco, A. (2015). "Recombinant antibody production evolves into multiple options aimed at yielding reagents suitable for application-specific needs." Microb Cell Fact **14**: 125.
Delaleau, M. and K. L. Borden (2015). "Multiple Export Mechanisms for mRNAs." Cells **4**(3): 452-473.

Denning, D. P., S. S. Patel, V. Uversky, A. L. Fink and M. Rexach (2003). "Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded." Proc Natl Acad Sci U S A **100**(5): 2450-2455.

Dever, T. E. and R. Green (2012). "The elongation, termination, and recycling phases of translation in eukaryotes." Cold Spring Harb Perspect Biol **4**(7): a013706.

Doms, R. W. and J. P. Moore (2000). "HIV-1 membrane fusion: targets of opportunity." J Cell Biol **151**(2): F9-14.

Dong, X., A. Biswas, K. E. Süel, L. K. Jackson, R. Martinez, H. Gu and Y. M. Chook (2009). "Structural basis for leucine-rich nuclear export signal recognition by CRM1." Nature **458**(7242): 1136-1141.

Dreyfuss, G., M. J. Matunis, S. Piñol-Roma and C. G. Burd (1993). "hnRNP proteins and the biogenesis of mRNA." Annu Rev Biochem **62**: 289-321.

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**(1): 379-387.

Dulude, D., M. Baril and L. Brakier-Gingras (2002). "Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1." Nucleic Acids Res **30**(23): 5094-5102.

Dynan, W. S. and R. Tjian (1983). "Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II." Cell **32**(3): 669-680.

El Kaderi, B., S. Medler, S. Raghunayakula and A. Ansari (2009). "Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries." J Biol Chem **284**(37): 25015-25025.

Emery, A. and R. Swanstrom (2021). "HIV-1: To Splice or Not to Splice, That Is the Question." Viruses **13**(2).

Engström, P. G., S. J. Ho Sui, O. Drivenes, T. S. Becker and B. Lenhard (2007). "Genomic regulatory blocks underlie extensive microsynteny conservation in insects." Genome Res **17**(12): 1898-1908.

Erkelenz, S., W. F. Mueller, M. S. Evans, A. Busch, K. Schöneweis, K. J. Hertel and H. Schaal (2013). "Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms." Rna **19**(1): 96-102.

Erkelenz, S., S. Theiss, M. Otte, M. Widera, J. O. Peter and H. Schaal (2014). "Genomic HEXploring allows landscaping of novel potential splicing regulatory elements." Nucleic Acids Res **42**(16): 10681-10697.

Erkmann, J. A. and U. Kutay (2004). "Nuclear export of mRNA: from the site of transcription to the cytoplasm." Exp Cell Res **296**(1): 12-20.

Esquiaqui, J. M., S. Kharytonchyk, D. Drucker and A. Telesnitsky (2020). "HIV-1 spliced RNAs display transcription start site bias." Rna **26**(6): 708-714.

Everett, R. D., D. Baty and P. Chambon (1983). "The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter." Nucleic Acids Res **11**(8): 2447-2464.

Exline, C. M., Z. Feng and C. M. Stoltzfus (2008). "Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression." J Virol **82**(8): 3921-3931.

Fairbrother, W. G., R. F. Yeh, P. A. Sharp and C. B. Burge (2002). "Predictive identification

of exonic splicing enhancers in human genes." Science **297**(5583): 1007-1013.
Fassati, A. and S. P. Goff (2001). "Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1." J Virol **75**(8): 3626-3635.
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**(5): 1495-1499.
Fernández-Miragall, O. and E. Martínez-Salas (2003). "Structural organization of a viral IRES depends on the integrity of the GNRA motif." Rna **9**(11): 1333-1344.
Fernandez, J., I. Yaman, R. Mishra, W. C. Merrick, M. D. Snider, W. H. Lamers and M. Hatzoglou (2001). "Internal ribosome entry site-mediated translation of a mammalian mRNA is regulated by amino acid availability." J Biol Chem **276**(15): 12285-12291.
Fica, S. M., C. Oubridge, W. P. Galej, M. E. Wilkinson, X. C. Bai, A. J. Newman and K. Nagai (2017). "Structure of a spliceosome remodelled for exon ligation." Nature **542**(7641): 377-380.
Fischer, U., J. Huber, W. C. Boelens, I. W. Mattaj and R. Lührmann (1995). "The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs." Cell **82**(3): 475-483.
Fischer, U., V. W. Pollard, R. Lührmann, M. Teufel, M. W. Michael, G. Dreyfuss and M. H. Malim (1999). "Rev-mediated nuclear export of RNA is dominant over nuclear retention and is coupled to the Ran-GTPase cycle." Nucleic Acids Res **27**(21): 4128-4134.
Forler, D., G. Rabut, F. D. Ciccarelli, A. Herold, T. Köcher, R. Niggeweg, P. Bork, J. Ellenberg and E. Izaurralde (2004). "RanBP2/Nup358 provides a major binding site for NXF1-p15 dimers at the nuclear pore complex and functions in nuclear mRNA export." Mol Cell Biol **24**(3): 1155-1167.
Fornerod, M., M. Ohno, M. Yoshida and I. W. Mattaj (1997). "CRM1 is an export receptor for leucine-rich nuclear export signals." Cell **90**(6): 1051-1060.
Fox-Walsh, K. L. and K. J. Hertel (2009). "Splice-site pairing is an intrinsically high fidelity process." Proc Natl Acad Sci U S A **106**(6): 1766-1771.
Freed, E. O. (2015). "HIV-1 assembly, release and maturation." Nat Rev Microbiol **13**(8): 484-496.
Freund, M., C. Asang, S. Kammler, C. Konermann, J. Krummheuer, M. Hipp, I. Meyer, W. Gierling, S. Theiss, T. Preuss, D. Schindler, J. Kjems and H. Schaal (2003). "A novel approach to describe a U1 snRNA binding site." Nucleic Acids Res **31**(23): 6963-6975.
Fridy, P. C., Y. Li, S. Keegan, M. K. Thompson, I. Nudelman, J. F. Scheid, M. Oeffinger, M. C. Nussenzweig, D. Fenyö, B. T. Chait and M. P. Rout (2014). "A robust pipeline for rapid production of versatile nanobody repertoires." Nat Methods **11**(12): 1253-1260.
Frolova, L., X. Le Goff, H. H. Rasmussen, S. Cheperegin, G. Drugeon, M. Kress, I. Arman, A. L. Haenni, J. E. Celis, M. Philippe and et al. (1994). "A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor." Nature **372**(6507): 701-703.
Fung, H. Y. and Y. M. Chook (2014). "Atomic basis of CRM1-cargo recognition, release and inhibition." Semin Cancer Biol **27**: 52-61.

Fütterer, J., Z. Kiss-László and T. Hohn (1993). "Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA." Cell **73**(4): 789-802.

Galej, W. P., N. Toor, A. J. Newman and K. Nagai (2018). "Molecular Mechanism and Evolution of Nuclear Pre-mRNA and Group II Intron Splicing: Insights from Cryo-Electron Microscopy Structures." Chem Rev **118**(8): 4156-4176.

Gallardo, M., M. J. Hornbaker, X. Zhang, P. Hu, C. Bueso-Ramos and S. M. Post (2016). "Aberrant hnRNP K expression: All roads lead to cancer." Cell Cycle **15**(12): 1552-1557.

Gallo, R. C., P. S. Sarin, E. P. Gelmann, M. Robert-Guroff, E. Richardson, V. S. Kalyanaraman, D. Mann, G. D. Sidhu, R. E. Stahl, S. Zolla-Pazner, J. Leibowitch and M. Popovic (1983). "Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS)." Science **220**(4599): 865-867.

Geuens, T., D. Bouhy and V. Timmerman (2016). "The hnRNP family: insights into their role in health and disease." Hum Genet **135**(8): 851-867.

Girnary, R., L. King, L. Robinson, R. Elston and I. Brierley (2007). "Structure-function analysis of the ribosomal frameshifting signal of two human immunodeficiency virus type 1 isolates with increased resistance to viral protease inhibitors." J Gen Virol **88**(Pt 1): 226-235.

Gottlieb, M. S., R. Schroff, H. M. Schanker, J. D. Weisman, P. T. Fan, R. A. Wolf and A. Saxon (1981). "Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency." N Engl J Med **305**(24): 1425-1431.

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**(1): 59-74.

Grantham, R., C. Gautier, M. Gouy, R. Mercier and A. Pavé (1980). "Codon catalog usage and the genome hypothesis." Nucleic Acids Res **8**(1): r49-r62.

Grosschedl, R. and M. L. Birnstiel (1980). "Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo." Proc Natl Acad Sci U S A **77**(3): 1432-1436.

Grünwald, D. and R. H. Singer (2010). "In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport." Nature **467**(7315): 604-607.

Grüter, P., C. Tabernero, C. von Kobbe, C. Schmitt, C. Saavedra, A. Bachi, M. Wilm, B. K. Felber and E. Izaurralde (1998). "TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus." Mol Cell **1**(5): 649-659.

Guo, J., Y. Zhu, X. Ma, G. Shang, B. Liu and K. Zhang (2023). "Virus Infection and mRNA Nuclear Export." Int J Mol Sci **24**(16).

Halees, A. S., E. Hitti, M. Al-Saif, L. Mahmoud, I. A. Vlasova-St Louis, D. J. Beisang, P. R. Bohjanen and K. Khabar (2011). "Global assessment of GU-rich regulatory content and function in the human transcriptome." RNA Biol **8**(4): 681-691.

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**(6402): 358-361.

Hamamoto, T., S. Gunji, H. Tsuji and T. Beppu (1983). "Leptomycins A and B, new antifungal antibiotics. I. Taxonomy of the producing strain and their fermentation, purification and characterization." J Antibiot (Tokyo) **36**(6): 639-645.

Haselbach, D., I. Komarov, D. E. Agafonov, K. Hartmuth, B. Graf, O. Dybkov, H. Urlaub, B. Kastner, R. Lührmann and H. Stark (2018). "Structure and Conformational Dynamics of the Human Spliceosomal B(act) Complex." Cell **172**(3): 454-464.e411.

Hashimoto, C. and J. A. Steitz (1984). "U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle." Nucleic Acids Res **12**(7): 3283-3293.

Hazen, R. J., D. Halbur, B. Mills, H. S. Kirkham and J. Hou (2021). "Evaluation of Medication Therapy Issues, Resolutions, and Adherence Among Persons With HIV in the Pharmacist-Led Patient-Centered HIV Care Model." J Acquir Immune Defic Syndr **88**(1): 96-102.

Henne, W. M., N. J. Buchkovich and S. D. Emr (2011). "The ESCRT pathway." Dev Cell **21**(1): 77-91.

Herold, A., L. Teixeira and E. Izaurralde (2003). "Genome-wide analysis of nuclear mRNA export pathways in Drosophila." Embo j **22**(10): 2472-2483.

Hershey, J. W. B., N. Sonenberg and M. B. Mathews (2019). "Principles of Translational Control." Cold Spring Harb Perspect Biol **11**(9).

Herz, H.-M., D. Hu and A. Shilatifard (2014). "Enhancer malfunction in cancer." Molecular Cell **53**(6): 859-866.

Hong, X., D. G. Scofield and M. Lynch (2006). "Intron size, abundance, and distribution within untranslated regions of genes." Mol Biol Evol **23**(12): 2392-2404.

Hoshino, S., M. Imai, T. Kobayashi, N. Uchida and T. Katada (1999). "The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-Poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein." J Biol Chem **274**(24): 16677-16680.

Hu, W. S. and S. H. Hughes (2012). "HIV-1 reverse transcription." Cold Spring Harb Perspect Med **2**(10).

Huang, Y., R. Gattoni, J. Stévenin and J. A. Steitz (2003). "SR splicing factors serve as adapter proteins for TAP-dependent mRNA export." Mol Cell **11**(3): 837-843.

Huh, G. S. and R. O. Hynes (1994). "Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element." Genes Dev **8**(13): 1561-1574.

Hui, J., L. H. Hung, M. Heiner, S. Schreiner, N. Neumüller, G. Reither, S. A. Haas and A. Bindereif (2005). "Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing." Embo j **24**(11): 1988-1998.

Hunter, M., P. Yuan, D. Vavilala and M. Fox (2019). "Optimization of Protein Expression in Mammalian Cells." Curr Protoc Protein Sci **95**(1): e77.

Ikemura, T. (1985). "Codon usage and tRNA content in unicellular and multicellular organisms." Mol Biol Evol **2**(1): 13-34.

Ill, C. R. and H. C. Chiou (2005). "Gene therapy progress and prospects: recent progress in transgene and RNAi expression cassettes." Gene Ther **12**(10): 795-802.

International Human Genome Sequencing, C. (2004). "Finishing the euchromatic sequence of the human genome." Nature **431**: 218-223.

Isel, C., C. Ehresmann, G. Keith, B. Ehresmann and R. Marquet (1995). "Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA(3Lys) (template/primer)." J Mol Biol **247**(2): 236-250.

Isel, C., C. Ehresmann and R. Marquet (2010). "Initiation of HIV Reverse Transcription." Viruses **2**(1): 213-243.

Ivanov, A., T. Mikhailova, B. Eliseev, L. Yeramala, E. Sokolova, D. Susorov, A. Shuvalov, C.

Schaffitzel and E. Alkalaeva (2016). "PABP enhances release factor recruitment and stop codon recognition during translation termination." Nucleic Acids Res **44**(16): 7766-7776.

Izaurralde, E., U. Kutay, C. von Kobbe, I. W. Mattaj and D. Görlich (1997). "The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus." Embo j **16**(21): 6535-6547.

Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr and H. E. Varmus (1988). "Characterization of ribosomal frameshifting in HIV-1 gag-pol expression." Nature **331**(6153): 280-283.

Jackson, R. J., C. U. Hellen and T. V. Pestova (2010). "The mechanism of eukaryotic translation initiation and principles of its regulation." Nat Rev Mol Cell Biol **11**(2): 113-127.

Jäger, S., P. Cimermancic, N. Gulbahce, J. R. Johnson, K. E. McGovern, S. C. Clarke, M. Shales, G. Mercenne, L. Pache, K. Li, H. Hernandez, G. M. Jang, S. L. Roth, E. Akiva, J. Marlett, M. Stephens, I. D'Orso, J. Fernandes, M. Fahey, C. Mahon, A. J. O'Donoghue, A. Todorovic, J. H. Morris, D. A. Maltby, T. Alber, G. Cagney, F. D. Bushman, J. A. Young, S. K. Chanda, W. I. Sundquist, T. Kortemme, R. D. Hernandez, C. S. Craik, A. Burlingame, A. Sali, A. D. Frankel and N. J. Krogan (2011). "Global landscape of HIV-human protein complexes." Nature **481**(7381): 365-370.

Jang, S. K., H. G. Kräusslich, M. J. Nicklin, G. M. Duke, A. C. Palmenberg and E. Wimmer (1988). "A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation." J Virol **62**(8): 2636-2643.

Jankowsky, E. (2011). "RNA helicases at work: binding and rearranging." Trends Biochem Sci **36**(1): 19-29.

Jarnik, M. and U. Aebi (1991). "Toward a more complete 3-D structure of the nuclear pore complex." J Struct Biol **107**(3): 291-308.

Javahery, R., A. Khachi, K. Lo, B. Zenzie-Gregory and S. T. Smale (1994). "DNA sequence requirements for transcriptional initiator activity in mammalian cells."

Jens, M. and N. Rajewsky (2015). "Competition between target sites of regulators shapes post-transcriptional gene regulation." Nat Rev Genet **16**(2): 113-126.

Jeong, S. (2017). "SR Proteins: Binders, Regulators, and Connectors of RNA." Mol Cells **40**(1): 1-9.

Jordan-Paiz, A., S. Franco and M. A. Martinez (2021). "Synonymous Codon Pair Recoding of the HIV-1 env Gene Affects Virus Replication Capacity." Cells **10**(7).

Kabaria, S. R., Y. Bae, M. E. Ehmann, A. M. Beitz, B. A. Dorn, E. L. Peterman, K. S. Love, D. S. Ploessl and K. E. Galloway (2024). "Programmable promoter editing for precise control of transgene expression." bioRxiv.

Kadonaga, J. T. (2012). "Perspectives on the RNA polymerase II core promoter." Wiley Interdiscip Rev Dev Biol **1**(1): 40-51.

Kalyanaraman, V. S., R. Pal, R. C. Gallo and M. G. Sarngadharan (1988). "A unique human immunodeficiency virus culture secreting soluble gp160." AIDS Res Hum Retroviruses **4**(5): 319-329.

Kammler, S., C. Leurs, M. Freund, J. Krummheuer, K. Seidel, T. O. Tange, M. K. Lund, J. Kjems, A. Scheid and H. Schaal (2001). "The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an

unstable env pre-mRNA." Rna **7**(3): 421-434.
Kammler, S., M. Otte, I. Hauber, J. Kjems, J. Hauber and H. Schaal (2006). "The strength of the HIV-1 3' splice sites affects Rev function." Retrovirology **3**: 89.
Kang, Y., H. P. Bogerd and B. R. Cullen (2000). "Analysis of cellular factors that mediate nuclear export of RNAs bearing the Mason-Pfizer monkey virus constitutive transport element." J Virol **74**(13): 5863-5871.
Kao, S. Y., A. F. Calman, P. A. Luciw and B. M. Peterlin (1987). "Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product." Nature **330**(6147): 489-493.
Karn, J. (1999). "Tackling Tat." J Mol Biol **293**(2): 235-254.
Karn, J., C. Dingwall, J. T. Finch, S. Heaphy and M. J. Gait (1991). "RNA binding by the tat and rev proteins of HIV-1." Biochimie **73**(1): 9-16.
Karn, J. and C. M. Stoltzfus (2012). "Transcriptional and posttranscriptional regulation of HIV-1 gene expression." Cold Spring Harb Perspect Med **2**(2): a006916.
Kaufmann, J. and S. T. Smale (1994). "Direct recognition of initiator elements by a component of the transcription factor IID complex." Genes Dev **8**(7): 821-829.
Kelich, J. M. and W. Yang (2014). "High-resolution imaging reveals new features of nuclear export of mRNA through the nuclear pore complexes." Int J Mol Sci **15**(8): 14492-14504.
Khan, M., S. Hou, S. Azam and H. Lei (2021). "Sequence-dependent recruitment of SRSF1 and SRSF7 to intronless lncRNA NKILA promotes nuclear export via the TREX/TAP pathway." Nucleic Acids Res **49**(11): 6420-6436.
Khiytani, D. K. and N. J. Dimmock (2002). "Characterization of a human immunodeficiency virus type 1 pre-integration complex in which the majority of the cDNA is resistant to DNase I digestion." J Gen Virol **83**(Pt 10): 2523-2532.
Kim, J. H., B. Hahm, Y. K. Kim, M. Choi and S. K. Jang (2000). "Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm." J Mol Biol **298**(3): 395-405.
Kim, S. J., J. Fernandez-Martinez, I. Nudelman, Y. Shi, W. Zhang, B. Raveh, T. Herricks, B. D. Slaughter, J. A. Hogan, P. Upla, I. E. Chemmama, R. Pellarin, I. Echeverria, M. Shivaraju, A. S. Chaudhury, J. Wang, R. Williams, J. R. Unruh, C. H. Greenberg, E. Y. Jacobs, Z. Yu, M. J. de la Cruz, R. Mironska, D. L. Stokes, J. D. Aitchison, M. F. Jarrold, J. L. Gerton, S. J. Ludtke, C. W. Akey, B. T. Chait, A. Sali and M. P. Rout (2018). "Integrative structure and functional anatomy of a nuclear pore complex." Nature **555**(7697): 475-482.
Kim, S. Y., R. Byrn, J. Groopman and D. Baltimore (1989). "Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression." J Virol **63**(9): 3708-3713.
Kimchi-Sarfaty, C., J. M. Oh, I. W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar and M. M. Gottesman (2007). "A "silent" polymorphism in the MDR1 gene changes substrate specificity." Science **315**(5811): 525-528.
Klatzmann, D., E. Champagne, S. Chamaret, J. Gruest, D. Guetard, T. Hercend, J. C. Gluckman and L. Montagnier (1984). "T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV." Nature **312**(5996): 767-768.
Klaver, B. and B. Berkhout (1994). "Comparison of 5' and 3' long terminal repeat promoter function in human immunodeficiency virus." J Virol **68**(6): 3830-3840.

Klein, J. S. and P. J. Bjorkman (2010). "Few and far between: how HIV may be evading antibody avidity." PLoS Pathog **6**(5): e1000908.

Köhler, A. and E. Hurt (2007). "Exporting RNA from the nucleus to the cytoplasm." Nat Rev Mol Cell Biol **8**(10): 761-773.

Köhler, A. and E. Hurt (2010). "Gene regulation by nucleoporins and links to cancer." Mol Cell **38**(1): 6-15.

Kolitz, S. E., J. E. Takacs and J. R. Lorsch (2009). "Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation." Rna **15**(1): 138-152.

Komarnitsky, P., E. J. Cho and S. Buratowski (2000). "Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription." Genes Dev **14**(19): 2452-2460.

Kosinski, J., S. Mosalaganti, A. von Appen, R. Teimer, A. L. DiGuilio, W. Wan, K. H. Bui, W. J. Hagen, J. A. Briggs, J. S. Glavy, E. Hurt and M. Beck (2016). "Molecular architecture of the inner ring scaffold of the human nuclear pore complex." Science **352**(6283): 363-365.

Koyama, M. and Y. Matsuura (2010). "An allosteric mechanism to displace nuclear export cargo from CRM1 and RanGTP by RanBP1." Embo j **29**(12): 2002-2013.

Kozak, M. (1984). "Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo." Nature **308**(5956): 241-246.

Kozak, M. (1986). "Bifunctional messenger RNAs in eukaryotes." Cell **47**(4): 481-483.

Kozak, M. (1986). "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes." Cell **44**(2): 283-292.

Kozak, M. and A. J. Shatkin (1979). "Characterization of translational initiation regions from eukaryotic messenger RNAs." Methods Enzymol **60**: 360-375.

Krummheuer, J., A. T. Johnson, I. Hauber, S. Kammler, J. L. Anderson, J. Hauber, D. F. Purcell and H. Schaal (2007). "A minimal uORF within the HIV-1 vpu leader allows efficient translation initiation at the downstream env AUG." Virology **363**(2): 261-271.

Kudo, N., N. Matsumori, H. Taoka, D. Fujiwara, E. P. Schreiner, B. Wolff, M. Yoshida and S. Horinouchi (1999). "Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region." Proc Natl Acad Sci U S A **96**(16): 9112-9117.

Kuehner, J. N., E. L. Pearson and C. Moore (2011). "Unravelling the means to an end: RNA polymerase II transcription termination." Nat Rev Mol Cell Biol **12**(5): 283-294.

Lafuente, E., R. Ramos and E. Martínez-Salas (2002). "Long-range RNA-RNA interactions between distant regions of the hepatitis C virus internal ribosome entry site element." J Gen Virol **83**(Pt 5): 1113-1121.

Landes, J. R., S. A. Moore, B. R. Bartley, H. Q. Doan, P. L. Rady and S. K. Tyring (2023). "The efficacy of selinexor (KPT-330), an XPO1 inhibitor, on non-hematologic cancers: a comprehensive review." J Cancer Res Clin Oncol **149**(5): 2139-2155.

Landry, J. J., P. T. Pyl, T. Rausch, T. Zichner, M. M. Tekkedil, A. M. Stütz, A. Jauch, R. S. Aiyar, G. Pau, N. Delhomme, J. Gagneur, J. O. Korbel, W. Huber and L. M. Steinmetz (2013). "The genomic and transcriptomic landscape of a HeLa cell line." G3 (Bethesda) **3**(8): 1213-1224.

Latorre, E. and L. W. Harries (2017). "Splicing regulatory factors, ageing and age-related disease." Ageing Res Rev **36**: 165-170.

Le Hir, H., D. Gatfield, E. Izaurralde and M. J. Moore (2001). "The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay." Embo j **20**(17): 4987-4997.

Lenhard, B., A. Sandelin and P. Carninci (2012). "Metazoan promoters: emerging characteristics and insights into transcriptional regulation."

Lerner, M. R. and J. A. Steitz (1979). "Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus." Proc Natl Acad Sci U S A **76**(11): 5495-5499.

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

Levine, M. and R. Tjian (2003). "Transcription regulation and animal diversity." Nature **424**(6945): 147-151.

Li, M., M. Mizuuchi, T. R. Burke, Jr. and R. Craigie (2006). "Retroviral DNA integration: reaction pathway and critical intermediates." Embo j **25**(6): 1295-1304.

Li, Y., L. Luo, D. Y. Thomas and C. Y. Kang (2000). "The HIV-1 Env protein signal sequence retards its cleavage and down-regulates the glycoprotein folding." Virology **272**(2): 417-428.

Liang, W. W. and S. C. Cheng (2015). "A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence." Genes Dev **29**(1): 81-93.

Lin, D. H. and A. Hoelz (2019). "The Structure of the Nuclear Pore Complex (An Update)." Annu Rev Biochem **88**: 725-783.

Liu, X. and J. E. Mertz (1995). "HnRNP L binds a cis-acting RNA sequence element that enables intron-dependent gene expression." Genes Dev **9**(14): 1766-1780.

Long, J. C. and J. F. Caceres (2009). "The SR protein family of splicing factors: master regulators of gene expression." Biochem J **417**(1): 15-27.

Lopez-Lastra, M., A. Rivas and M. I. Barria (2005). "Protein synthesis in eukaryotes: The growing biological relevance of cap-independent translation initiation." Biological Research **38**: 121-146.

Lopez, P. J. and B. Séraphin (1999). "Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition." Rna **5**(9): 1135-1137.

Lund, N., M. P. Milev, R. Wong, T. Sanmuganantham, K. Woolaway, B. Chabot, S. Abou Elela, A. J. Mouland and A. Cochrane (2012). "Differential effects of hnRNP D/AUF1 isoforms on HIV-1 gene expression." Nucleic Acids Res **40**(8): 3663-3675.

Lutzmann, M., R. Kunze, A. Buerer, U. Aebi and E. Hurt (2002). "Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins." Embo j **21**(3): 387-397.

Ma, J., Z. Liu, N. Michelotti, S. Pitchiaya, R. Veerapaneni, J. R. Androsavich, N. G. Walter and W. Yang (2013). "High-resolution three-dimensional mapping of mRNA export through the nuclear pore." Nat Commun **4**: 2414.

Mahiet, C. and C. M. Swanson (2016). "Control of HIV-1 gene expression by SR proteins." Biochem Soc Trans **44**(5): 1417-1425.

Maldarelli, F., M. A. Martin and K. Strebel (1991). "Identification of posttranscriptionally active inhibitory sequences in human immunodeficiency virus type 1 RNA: novel level of gene regulation." J Virol **65**(11): 5732-5743.

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**(6212): 254-257.

Mandal, D., C. M. Exline, Z. Feng and C. M. Stoltzfus (2009). "Regulation of Vif mRNA splicing by human immunodeficiency virus type 1 requires 5' splice site D2 and an exonic splicing enhancer to counteract cellular restriction factor APOBEC3G." J Virol **83**(12): 6067-6078.

Manley, J. L. and A. R. Krainer (2010). "A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins)." Genes Dev **24**(11): 1073-1074.

Marini, B., A. Kertesz-Farkas, H. Ali, B. Lucic, K. Lisek, L. Manganaro, S. Pongor, R. Luzzati, A. Recchia, F. Mavilio, M. Giacca and M. Lusic (2015). "Nuclear architecture dictates HIV-1 integration site selection." Nature **521**(7551): 227-231.

Markham, N. R. and M. Zuker (2008). "UNAFold: software for nucleic acid folding and hybridization." Methods Mol Biol **453**: 3-31.

Martínez-Salas, E. and O. Fernández-Miragall (2004). "Picornavirus IRES: structure function relationship." Curr Pharm Des **10**(30): 3757-3767.

Matera, A. G. and Z. Wang (2014). "A day in the life of the spliceosome." Nat Rev Mol Cell Biol **15**(2): 108-121.

Matzat, L. H., S. Berberoglu and L. Lévesque (2008). "Formation of a Tap/NXF1 homotypic complex is mediated through the amino-terminal domain of Tap and enhances interaction with nucleoporins." Mol Biol Cell **19**(1): 327-338.

Mauro, V. P. (2018). "Codon Optimization in the Production of Recombinant Biotherapeutics: Potential Risks and Considerations." BioDrugs **32**(1): 69-81.

Mauro, V. P. and S. A. Chappell (2014). "A critical analysis of codon optimization in human therapeutics." Trends Mol Med **20**(11): 604-613.

McDonald, D., M. A. Vodicka, G. Lucero, T. M. Svitkina, G. G. Borisy, M. Emerman and T. J. Hope (2002). "Visualization of the intracellular behavior of HIV in living cells." J Cell Biol **159**(3): 441-452.

Mikaélian, I., M. Krieg, M. J. Gait and J. Karn (1996). "Interactions of INS (CRS) elements and the splicing machinery regulate the production of Rev-responsive mRNAs." J Mol Biol **257**(2): 246-264.

Miller, M. D., C. M. Farnet and F. D. Bushman (1997). "Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition." J Virol **71**(7): 5382-5390.

Monette, A., L. Ajamian, M. López-Lastra and A. J. Mouland (2009). "Human immunodeficiency virus type 1 (HIV-1) induces the cytoplasmic retention of heterogeneous nuclear ribonucleoprotein A1 by disrupting nuclear import: implications for HIV-1 gene expression." J Biol Chem **284**(45): 31350-31362.

Morris, D. R. and A. P. Geballe (2000). "Upstream open reading frames as regulators of mRNA translation." Mol Cell Biol **20**(23): 8635-8642.

Mueller, N., N. van Bel, B. Berkhout and A. T. Das (2014). "HIV-1 splicing at the major splice donor site is restricted by RNA structure." Virology **468-470**: 609-620.

Müller-McNicoll, M., V. Botti, A. M. de Jesus Domingues, H. Brandl, O. D. Schwich, M. C. Steiner, T. Curk, I. Poser, K. Zarnack and K. M. Neugebauer (2016). "SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export." Genes Dev **30**(5): 553-566.

Naim, B., D. Zbaida, S. Dagan, R. Kapon and Z. Reich (2009). "Cargo surface hydrophobicity is sufficient to overcome the nuclear pore complex selectivity barrier." Embo j **28**(18): 2697-2705.

Nakielny, S. and G. Dreyfuss (1996). "The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals." J Cell Biol **134**(6): 1365-1373.

Napetschnig, J., S. A. Kassube, E. W. Debler, R. W. Wong, G. Blobel and A. Hoelz (2009). "Structural and functional analysis of the interaction between the nucleoporin Nup214 and the DEAD-box helicase Ddx19." Proc Natl Acad Sci U S A **106**(9): 3089-3094.

Nechaev, S., D. C. Fargo, G. dos Santos, L. Liu, Y. Gao and K. Adelman (2010). "Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila." Science **327**(5963): 335-338.

Neumann, N., D. Lundin and A. M. Poole (2010). "Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor." PLoS One **5**(10): e13241.

Newlands, E. S., G. J. Rustin and M. H. Brampton (1996). "Phase I trial of elactocin." Br J Cancer **74**(4): 648-649.

Newman, A. J. and C. Norman (1992). "U5 snRNA interacts with exon sequences at 5' and 3' splice sites." Cell **68**(4): 743-754.

Nguyen, H. T., Q. Wang, S. Anang and J. G. Sodroski (2023). "Characterization of the Human Immunodeficiency Virus (HIV-1) Envelope Glycoprotein Conformational States on Infectious Virus Particles." J Virol **97**(3): e0185722.

Nguyen, T. H., W. P. Galej, X. C. Bai, C. G. Savva, A. J. Newman, S. H. Scheres and K. Nagai (2015). "The architecture of the spliceosomal U4/U6.U5 tri-snRNP." Nature **523**(7558): 47-52.

Nikolaitchik, O. A., K. A. Dilley, W. Fu, R. J. Gorelick, S. H. Tai, F. Soheilian, R. G. Ptak, K. Nagashima, V. K. Pathak and W. S. Hu (2013). "Dimeric RNA recognition regulates HIV-1 genome packaging." PLoS Pathog **9**(3): e1003249.

Ocwieja, K. E., S. Sherrill-Mix, R. Mukherjee, R. Custers-Allen, P. David, M. Brown, S. Wang, D. R. Link, J. Olson, K. Travers, E. Schadt and F. D. Bushman (2012). "Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing." Nucleic Acids Res **40**(20): 10345-10355.

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**(4944): 845-848.

Ostermann, P. N., A. Ritchie, J. Ptok and H. Schaal (2021). "Let It Go: HIV-1 cis-Acting Repressive Sequences." J Virol **95**(15): e0034221.

Pabis, M., N. Neufeld, M. C. Steiner, T. Bojic, Y. Shav-Tal and K. M. Neugebauer (2013). "The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells." Rna **19**(8): 1054-1063.

Pan, Q., O. Shai, L. J. Lee, B. J. Frey and B. J. Blencowe (2008). "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing." Nat Genet **40**(12): 1413-1415.

Parker, R., P. G. Siliciano and C. Guthrie (1987). "Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA." Cell **49**(2): 229-239.

Pelletier, J. and N. Sonenberg (1988). "Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA." Nature **334**(6180): 320-325.

Peng, S. S., C. Y. Chen and A. B. Shyu (1996). "Functional characterization of a non-AUUUA AU-rich element from the c-jun proto-oncogene mRNA: evidence for a novel class of AU-rich elements." Mol Cell Biol **16**(4): 1490-1499.

Pennings, P. S. (2013). "HIV Drug Resistance: Problems and Perspectives." Infect Dis Rep **5**(Suppl 1): e5.

Perbellini, R., S. Greco, G. Sarra-Ferraris, R. Cardani, M. C. Capogrossi, G. Meola and F. Martelli (2011). "Dysregulation and cellular mislocalization of specific miRNAs in myotonic dystrophy type 1." Neuromuscul Disord **21**(2): 81-88.

Perriman, R. and M. Ares, Jr. (2010). "Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing." Mol Cell **38**(3): 416-427.

Pestova, T. V., I. B. Lomakin, J. H. Lee, S. K. Choi, T. E. Dever and C. U. Hellen (2000). "The joining of ribosomal subunits in eukaryotes requires eIF5B." Nature **403**(6767): 332-335.

Pisarev, A. V., M. A. Skabkin, V. P. Pisareva, O. V. Skabkina, A. M. Rakotondrafara, M. W. Hentze, C. U. Hellen and T. V. Pestova (2010). "The role of ABCE1 in eukaryotic posttermination ribosomal recycling." Mol Cell **37**(2): 196-210.

Plaschka, C., P. C. Lin and K. Nagai (2017). "Structure of a pre-catalytic spliceosome." Nature **546**(7660): 617-621.

Platt, E. J., K. Wehrly, S. E. Kuhmann, B. Chesebro and D. Kabat (1998). "Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1." J Virol **72**(4): 2855-2864.

Ponjavic, J., B. Lenhard, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki and A. Sandelin (2006). "Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters." Genome Biol **7**(8): R78.

Pope, M., M. G. Betjes, N. Romani, H. Hirmand, P. U. Cameron, L. Hoffman, S. Gezelter, G. Schuler and R. M. Steinman (1994). "Conjugates of dendritic cells and memory T lymphocytes from skin facilitate productive infection with HIV-1." Cell **78**(3): 389-398.

Pornillos, O., B. K. Ganser-Pornillos and M. Yeager (2011). "Atomic-level modelling of the HIV capsid." Nature **469**(7330): 424-427.

Ptok, J., L. Müller, P. N. Ostermann, A. Ritchie, A. T. Dilthey, S. Theiss and H. Schaal (2021). "Modifying splice site usage with ModCon: Maintaining the genetic code while changing the underlying mRNP code." Comput Struct Biotechnol J **19**: 3069-3076.

Ptok, J., L. Müller, S. Theiss and H. Schaal (2019). "Context matters: Regulation of splice donor usage." Biochim Biophys Acta Gene Regul Mech **1862**(11-12): 194391.

Pugh, B. F. and R. Tjian (1990). "Mechanism of transcriptional activation by Sp1: evidence for coactivators." Cell **61**(7): 1187-1197.

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**(11): 6365-6378.

Raghunathan, P. L. and C. Guthrie (1998). "RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2." Curr Biol **8**(15): 847-855.

Raker, V. A., K. Hartmuth, B. Kastner and R. Lührmann (1999). "Spliceosomal U snRNP core

assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner." Mol Cell Biol **19**(10): 6554-6565.

Ramdas, P., A. K. Sahu, T. Mishra, V. Bhardwaj and A. Chande (2020). "From Entry to Egress: Strategic Exploitation of the Cellular Processes by HIV-1." Front Microbiol **11**: 559792.

Rausch, J. W. and S. F. Le Grice (2015). "HIV Rev Assembly on the Rev Response Element (RRE): A Structural Perspective." Viruses **7**(6): 3053-3075.

Rein, A. (2019). "RNA Packaging in HIV." Trends Microbiol **27**(8): 715-723.

Rodríguez-Molina, J. B. and M. Turtola (2023). "Birth of a poly(A) tail: mechanisms and control of mRNA polyadenylation." FEBS Open Bio **13**(7): 1140-1153.

Roeder, R. G. and W. J. Rutter (1969). "Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms."

Roeder, R. G. and W. J. Rutter (1970). "Specific Nucleolar and Nucleoplasmic RNA Polymerases."

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**(7): 2071-2075.

Roy, R., D. Durie, H. Li, B. Q. Liu, J. M. Skehel, F. Mauri, L. V. Cuorvo, M. Barbareschi, L. Guo, M. Holcik, M. J. Seckl and O. E. Pardo (2014). "hnRNPA1 couples nuclear export and translation of specific mRNAs downstream of FGF-2/S6K2 signalling." Nucleic Acids Res **42**(20): 12483-12497.

Sabi, R. and T. Tuller (2014). "Modelling the efficiency of codon-tRNA interactions based on codon usage bias." DNA Res **21**(5): 511-526.

Saroufim, M. A., P. Bensidoun, P. Raymond, S. Rahman, M. R. Krause, M. Oeffinger and D. Zenklusen (2015). "The nuclear basket mediates perinuclear mRNA scanning in budding yeast." J Cell Biol **211**(6): 1131-1140.

Satterly, N., P. L. Tsai, J. van Deursen, D. R. Nussenzveig, Y. Wang, P. A. Faria, A. Levay, D. E. Levy and B. M. Fontoura (2007). "Influenza virus targets the mRNA export machinery and the nuclear pore complex." Proc Natl Acad Sci U S A **104**(6): 1853-1858.

Sauna, Z. E. and C. Kimchi-Sarfaty (2011). "Understanding the contribution of synonymous mutations to human disease." Nat Rev Genet **12**(10): 683-691.

Schaal, H., M. Klein, P. Gehrmann, O. Adams and A. Scheid (1995). "Requirement of N-terminal amino acid residues of gp41 for human immunodeficiency virus type 1-mediated cell fusion." J Virol **69**(6): 3308-3314.

Scherer, W. F., J. T. Syverton and G. O. Gey (1953). "Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix." J Exp Med **97**(5): 695-710.

Schuller, A. P. and R. Green (2018). "Roadblocks and resolutions in eukaryotic translation." Nat Rev Mol Cell Biol **19**(8): 526-541.

Schwartz, S., B. K. Felber, D. M. Benko, E. M. Fenyö and G. N. Pavlakis (1990). "Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1." J Virol **64**(6): 2519-2529.

Schwartz, S., B. K. Felber and G. N. Pavlakis (1992). "Distinct RNA sequences in the gag region of human immunodeficiency virus type 1 decrease RNA stability and inhibit

expression in the absence of Rev protein." J Virol **66**(1): 150-159.
Schwartz, S., B. K. Felber and G. N. Pavlakis (1992). "Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs." Mol Cell Biol **12**(1): 207-219.
Schwartz, Y. B., T. G. Kahn, P. Stenberg, K. Ohno, R. Bourgon and V. Pirrotta (2010). "Alternative epigenetic chromatin states of polycomb target genes." PLoS Genet **6**(1): e1000805.
Schwer, B. (2008). "A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release." Mol Cell **30**(6): 743-754.
Scott, D. D., L. C. Aguilar, M. Kramar and M. Oeffinger (2019). The Biology of mRNA: Structure and Function., Advances in Experimental Medicine and Biology.
Selyutina, A., M. Persaud, K. Lee, V. KewalRamani and F. Diaz-Griffero (2020). "Nuclear Import of the HIV-1 Core Precedes Reverse Transcription and Uncoating." Cell Rep **32**(13): 108201.
Senapathy, P., M. B. Shapiro and N. L. Harris (1990). "Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project." Methods Enzymol **183**: 252-278.
Sertznig, H., F. Hillebrand, S. Erkelenz, H. Schaal and M. Widera (2018). "Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet." Virology **516**: 176-188.
Shandilya, J. and S. G. Roberts (2012). "The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling." Biochim Biophys Acta **1819**(5): 391-400.
Sherman, F., G. McKnight and J. W. Stewart (1980). "AUG is the only initiation codon in eukaryotes." Biochim Biophys Acta **609**(2): 343-346.
Sheth, N., X. Roca, M. L. Hastings, T. Roeder, A. R. Krainer and R. Sachidanandam (2006). "Comprehensive splice-site analysis using comparative genomics." Nucleic Acids Res **34**(14): 3955-3967.
Shi, S., C. A. Hamann, J. C. Lee and J. M. Brunger (2024). "Use of CRISPRoff and synthetic Notch to modulate and relay endogenous gene expression programs in engineered cells." Front Bioeng Biotechnol **12**: 1346810.
Siebrasse, J. P., T. Kaminski and U. Kubitscheck (2012). "Nuclear export of single native mRNA molecules observed by light sheet fluorescence microscopy." Proc Natl Acad Sci U S A **109**(24): 9426-9431.
Smale, S. T. and J. T. Kadonaga (2003). "The RNA Polymerase II Core Promoter." Annual Rev. Biochem **72**: 449-479.
Smith, C. W., T. T. Chu and B. Nadal-Ginard (1993). "Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns." Mol Cell Biol **13**(8): 4939-4952.
Sodroski, J., C. Rosen, F. Wong-Staal, S. Z. Salahuddin, M. Popovic, S. Arya, R. C. Gallo and W. A. Haseltine (1985). "Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat." Science **227**(4683): 171-173.
Sokabe, M., C. S. Fraser and J. W. Hershey (2012). "The human translation initiation multi-factor complex promotes methionyl-tRNAi binding to the 40S ribosomal subunit." Nucleic Acids Res **40**(2): 905-913.
Sokolowski, M. and S. Schwartz (2001). "Heterogeneous nuclear ribonucleoprotein C

binds exclusively to the functionally important UUUUU-motifs in the human papillomavirus type-1 AU-rich inhibitory element." Virus Res **73**(2): 163-175.

Sonneborn, T. M. (1965). Degeneracy of the Genetic Code: Extent, Nature, and Genetic Implications. Evolving Genes and Proteins. V. Bryson and H. J. Vogel, Academic Press**:** 377-397.

Sontheimer, E. J. and J. A. Steitz (1993). "The U5 and U6 small nuclear RNAs as active site components of the spliceosome." Science **262**(5142): 1989-1996.

Staknis, D. and R. Reed (1994). "SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex." Mol Cell Biol **14**(11): 7670-7682.

Staley, J. P. and C. Guthrie (1999). "An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p." Mol Cell **3**(1): 55-64.

Stano, A., D. P. Leaman, A. S. Kim, L. Zhang, L. Autin, J. Ingale, S. K. Gift, J. Truong, R. T. Wyatt, A. J. Olson and M. B. Zwick (2017). "Dense Array of Spikes on HIV-1 Virion Particles." J Virol **91**(14).

Steckelberg, A. L., V. Boehm, A. M. Gromadzka and N. H. Gehring (2012). "CWC22 connects pre-mRNA splicing and exon junction complex assembly." Cell Rep **2**(3): 454-461.

Stoltzfus, C. M. (2009). "Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication." Adv Virus Res **74**: 1-40.

Stoneley, M., S. A. Chappell, C. L. Jopling, M. Dickens, M. MacFarlane and A. E. Willis (2000). "c-Myc protein synthesis is initiated from the internal ribosome entry segment during apoptosis." Mol Cell Biol **20**(4): 1162-1169.

Stutz, F., A. Bachi, T. Doerks, I. C. Braun, B. Séraphin, M. Wilm, P. Bork and E. Izaurralde (2000). "REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export." Rna **6**(4): 638-650.

Stuwe, T., A. R. Correia, D. H. Lin, M. Paduch, V. T. Lu, A. A. Kossiakoff and A. Hoelz (2015). "Nuclear pores. Architecture of the nuclear pore complex coat." Science **347**(6226): 1148-1152.

Suh, D., B. Seguin, S. Atkinson, B. Ozdamar, A. Staffa, A. Emili, A. Mouland and A. Cochrane (2003). "Mapping of determinants required for the function of the HIV-1 env nuclear retention sequence." Virology **310**(1): 85-99.

Sun, Q., Y. P. Carrasco, Y. Hu, X. Guo, H. Mirzaei, J. Macmillan and Y. M. Chook (2013). "Nuclear export inhibition through covalent conjugation and hydrolysis of Leptomycin B by CRM1." Proc Natl Acad Sci U S A **110**(4): 1303-1308.

Sundquist, W. I. and H. G. Kräusslich (2012). "HIV-1 assembly, budding, and maturation." Cold Spring Harb Perspect Med **2**(7): a006924.

Suyama, M., T. Doerks, I. C. Braun, M. Sattler, E. Izaurralde and P. Bork (2000). "Prediction of structural domains of TAP reveals details of its interaction with p15 and nucleoporins." EMBO Rep **1**(1): 53-58.

Taniguchi, I., N. Mabuchi and M. Ohno (2014). "HIV-1 Rev protein specifies the viral RNA export pathway by suppressing TAP/NXF1 recruitment." Nucleic Acids Res **42**(10): 6645-6658.

Timney, B. L., B. Raveh, R. Mironska, J. M. Trivedi, S. J. Kim, D. Russel, S. R. Wente, A. Sali and M. P. Rout (2016). "Simple rules for passive diffusion through the nuclear pore complex." J Cell Biol **215**(1): 57-76.

Topisirovic, I., N. Siddiqui and K. L. Borden (2009). "The eukaryotic translation initiation factor 4E (eIF4E) and HuR RNA operons collaboratively regulate the expression of survival and proliferative genes." Cell Cycle **8**(7): 960-961.

Trono, D., C. Van Lint, C. Rouzioux, E. Verdin, F. Barré-Sinoussi, T. W. Chun and N. Chomont (2010). "HIV persistence and the prospect of long-term drug-free remissions for HIV-infected individuals." Science **329**(5988): 174-180.

Tsai, C. J., Z. E. Sauna, C. Kimchi-Sarfaty, S. V. Ambudkar, M. M. Gottesman and R. Nussinov (2008). "Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima." J Mol Biol **383**(2): 281-291.

Turner, J. G., J. Dawson and D. M. Sullivan (2012). "Nuclear export of proteins and drug resistance in cancer." Biochem Pharmacol **83**(8): 1021-1032.

Ule, J. and B. J. Blencowe (2019). "Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution." Mol Cell **76**(2): 329-345.

Urlaub, H., V. A. Raker, S. Kostka and R. Lührmann (2001). "Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure." Embo j **20**(1-2): 187-196.

Valencia, P., A. P. Dias and R. Reed (2008). "Splicing promotes rapid and efficient mRNA export in mammalian cells." Proc Natl Acad Sci U S A **105**(9): 3386-3391.

Viphakone, N., G. M. Hautbergue, M. Walsh, C. T. Chang, A. Holland, E. G. Folco, R. Reed and S. A. Wilson (2012). "TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export." Nat Commun **3**: 1006.

Vlasova-St.Louis, I. A. and C. Sagarsky (2018). "Mammalian Cis-Acting RNA Sequence Elements." InTech.

Wahl, M. C., C. L. Will and R. Lührmann (2009). "The spliceosome: design principles of a dynamic RNP machine." Cell **136**(4): 701-718.

Walker, J. I., P. Faik and M. J. Morgan (1990). "Characterization of the 5' end of the gene for human glucose phosphate isomerase (GPI)." Genomics **7**(4): 638-643.

Wang, E. T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth and C. B. Burge (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-476.

Wang, Z. and C. B. Burge (2008). "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." Rna **14**(5): 802-813.

Watson, M. L. (1959). "Further observations on the nuclear envelope of the animal cell." J Biophys Biochem Cytol **6**(2): 147-156.

Wegener, M. and M. Müller-McNicoll (2018). "Nuclear retention of mRNAs - quality control, gene regulation and human disease." Semin Cell Dev Biol **79**: 131-142.

Wei, P., M. E. Garber, S. M. Fang, W. H. Fischer and K. A. Jones (1998). "A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA." Cell **92**(4): 451-462.

WHO. (2023, July 13 2023). "HIV and AIDS." World Health Organization, from https://www.who.int/news-room/fact-sheets/detail/hiv-aids.

Wickramasinghe, V. O., R. Andrews, P. Ellis, C. Langford, J. B. Gurdon, M. Stewart, A. R. Venkitaraman and R. A. Laskey (2014). "Selective nuclear export of specific classes of mRNA from mammalian nuclei is promoted by GANP." Nucleic Acids Res **42**(8): 5059-5071.

Wickramasinghe, V. O., P. I. McMurtrie, A. D. Mills, Y. Takei, S. Penrhyn-Lowe, Y. Amagase,

S. Main, J. Marr, M. Stewart and R. A. Laskey (2010). "mRNA export from mammalian cell nuclei is dependent on GANP." Curr Biol **20**(1): 25-31.

Widera, M., S. Erkelenz, F. Hillebrand, A. Krikoni, D. Widera, W. Kaisers, R. Deenen, M. Gombert, R. Dellen, T. Pfeiffer, B. Kaltschmidt, C. Münk, V. Bosch, K. Köhrer and H. Schaal (2013). "An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA." J Virol **87**(5): 2707-2720.

Widera, M., F. Hillebrand, S. Erkelenz, A. A. Vasudevan, C. Münk and H. Schaal (2014). "A functional conserved intronic G run in HIV-1 intron 3 is critical to counteract APOBEC3G-mediated host restriction." Retrovirology **11**: 72.

Wilen, C. B., J. C. Tilton and R. W. Doms (2012). "HIV: cell binding and entry." Cold Spring Harb Perspect Med **2**(8).

Will, C. L. and R. Lührmann (2011). "Spliceosome structure and function." Cold Spring Harb Perspect Biol **3**(7).

Wolff, B., J. J. Sanglier and Y. Wang (1997). "Leptomycin B is an inhibitor of nuclear export: inhibition of nucleo-cytoplasmic translocation of the human immunodeficiency virus type 1 (HIV-1) Rev protein and Rev-dependent mRNA." Chem Biol **4**(2): 139-147.

Wu, M., C. W. S. Tong, W. Yan, K. K. W. To and W. C. S. Cho (2019). "The RNA Binding Protein HuR: A Promising Drug Target for Anticancer Therapy." Curr Cancer Drug Targets **19**(5): 382-399.

Wu, X., Y. Li, B. Crise, S. M. Burgess and D. J. Munroe (2005). "Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses." J Virol **79**(8): 5211-5214.

Xu, D., A. Farmer and Y. M. Chook (2010). "Recognition of nuclear targeting signals by Karyopherin-β proteins." Curr Opin Struct Biol **20**(6): 782-790.

Xu, D., A. Farmer, G. Collett, N. V. Grishin and Y. M. Chook (2012). "Sequence and structural analyses of nuclear export signals in the NESdb database." Mol Biol Cell **23**(18): 3677-3693.

Xu, Y., W. Wu, Q. Han, Y. Wang, C. Li, P. Zhang and H. Xu (2019). "Post-translational modification control of RNA-binding protein hnRNPK function." Open Biol **9**(3): 180239.

Yamaguchi, Y., T. Takagi, T. Wada, K. Yano, A. Furuya, S. Sugimoto, J. Hasegawa and H. Handa (1999). "NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation." Cell **97**(1): 41-51.

Yamashita, R., Y. Suzuki, S. Sugano and K. Nakai (2005). "Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity." Gene **350**(2): 129-136.

Yáñez-Cuna, J. O., H. Q. Dinh, E. Z. Kvon, D. Shlyueva and A. Stark (2012). "Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding." Genome Res **22**(10): 2018-2030.

Yang, C. C., C. H. Chang, H. L. Chen, M. C. Chou, C. J. Yang, R. S. Jhou, E. Y. Huang, H. C. Li, C. S. Suen, M. J. Hwang and C. Shih (2022). "CRM1-spike-mediated nuclear export of hepatitis B virus encapsidated viral RNA." Cell Rep **38**(10): 110472.

Yedavalli, V. S. and K. T. Jeang (2010). "Trimethylguanosine capping selectively promotes expression of Rev-dependent HIV-1 RNAs." Proc Natl Acad Sci U S A **107**(33): 14787-14792.

Yeo, G. and C. B. Burge (2004). "Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals." J Comput Biol **11**(2-3): 377-394.

Yueh, A. and R. J. Schneider (1996). "Selective translation initiation by ribosome jumping in adenovirus-infected and heat-shocked cells." Genes Dev **10**(12): 1557-1567.

Zahler, A. M., K. M. Neugebauer, W. S. Lane and M. B. Roth (1993). "Distinct functions of SR proteins in alternative pre-mRNA splicing." Science **260**(5105): 219-222.

Zhan, X., C. Yan, X. Zhang, J. Lei and Y. Shi (2018). "Structures of the human pre-catalytic spliceosome and its precursor spliceosome." Cell Res **28**(12): 1129-1140.

Zhang, K., Y. Xie, R. Muñoz-Moreno, J. Wang, L. Zhang, M. Esparza, A. García-Sastre, B. M. A. Fontoura and Y. Ren (2019). "Structural basis for influenza virus NS1 protein block of mRNA nuclear export." Nat Microbiol **4**(10): 1671-1679.

Zhang, S., S. Aibara, S. M. Vos, D. E. Agafonov, R. Lührmann and P. Cramer (2021). "Structure of a transcribing RNA polymerase II-U1 snRNP complex." Science **371**(6526): 305-309.

Zheng, X., Q. Peng, L. Wang, X. Zhang, L. Huang, J. Wang and Z. Qin (2020). "Serine/arginine-rich splicing factors: the bridge linking alternative splicing and cancer." Int J Biol Sci **16**(13): 2442-2453.

Zhou, Z. and X. D. Fu (2013). "Regulation of splicing by SR proteins and SR protein-specific kinases." Chromosoma **122**(3): 191-207.

Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. M. Sharp and D. D. Ho (1998). "An African HIV-1 sequence from 1959 and implications for the origin of the epidemic." Nature **391**(6667): 594-597.

Zhuang, Y. and A. M. Weiner (1986). "A compensatory base change in U1 snRNA suppresses a 5' splice site mutation." Cell **46**(6): 827-835.

Zhuang, Y. A., A. M. Goldstein and A. M. Weiner (1989). "UACUAAC is the preferred branch site for mammalian mRNA splicing." Proc Natl Acad Sci U S A **86**(8): 2752-2756.

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

6 Publications

Müller L, Andrée M, Moskorz W, Drexler I, Hauka S, Ptok J, Walotka L, Grothmann R, Hillebrandt J, **Ritchie A**, Peter L, Walker A, Timm J, Adams O, Schaal H. Adjusted COVID-19 booster schedules balance age-dependent differences in antibody titers benefitting risk populations. Front Aging. 2022 Oct 12;3:1027885.

Müller L, Ptok J, Nisar A, Antemann J, Grothmann R, Hillebrand F, Brillen AL, **Ritchie A**, Theiss S, Schaal H. Modeling splicing outcome by combining 5'ss strength and splicing regulatory elements. Nucleic Acids Res. 2022 Aug 26;50(15):8834-8851.

Müller L, Andrée M, Moskorz W, Drexler I, Walotka L, Grothmann R, Ptok J, Hillebrandt J, **Ritchie A**, Rabl D, Ostermann PN, Robitzsch R, Hauka S, Walker A, Menne C, Grutza R, Timm J, Adams O, Schaal H. Age-dependent Immune Response to the Biontech/Pfizer BNT162b2 Coronavirus Disease 2019 Vaccination. Clin Infect Dis. 2021 Dec 6;73(11):2065-2072.

Ostermann PN*, **Ritchie A***, Ptok J, Schaal H. Let It Go: HIV-1 *cis*-Acting Repressive Sequences. J Virol. 2021 Jul 12;95(15):e0034221.

Ptok J, Müller L, Ostermann PN, **Ritchie A**, Dilthey AT, Theiss S, Schaal H. Modifying splice site usage with *ModCon*: Maintaining the *genetic code* while changing the underlying *mRNP code*. Comput Struct Biotechnol J. 2021 May 21;19:3069-3076.

7 Erklärung

I declare under oath that I have compiled my dissertation independently and without any undue assistance by third parties under consideration of the “Principles for Safeguarding Good Scientific Practice at Heinrich Heine University Düsseldorf”. This thesis has not been submitted to any other faculty. Furthermore, I have not made any unsuccessful or successful doctoral attempts to date.

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Diese Arbeit wurde bei keiner anderen Fakultät vorgelegt. Weiterhin habe ich bisher keine erfolglosen oder erfolgreichen Promotionsversuche unternommen.

Anastasia Nicole Ritchie

Düsseldorf, den

8 Acknowledgements

I would like to take this opportunity to express my gratitude to everyone who made this work possible and whose support ensured the completion of my doctoral thesis. First and foremost, I want to thank Prof. Dr. Heiner Schaal for inviting me to pursue my doctoral research in his group. His wealth of ideas and meticulous attention to detail provided the foundation for this work, and his boundless enthusiasm for exploring the unknown truly inspired me. I am also grateful to the Molecules of Infection (MOI) Graduate School IV for funding my research and allowing me to engage with a diverse range of researchers and topics. Additionally, I would like to thank Prof. Dr. Michael Feldbrügge and Dr. Lisa Müller for their mentorship throughout my doctoral journey. I also wish to express my gratitude to our collaborator, Prof. Dr. Alan Cochrane, for his advice and support, as well as welcoming me to his lab during my exchange.

I'd like to extend my thanks to my colleagues and friends; my time in the lab and at the institute would not have been as enjoyable without their support and camaraderie. Additionally, I'm deeply grateful to my family and friends for their unwavering encouragement and for always listening, no matter how confusing. A special thanks goes to Guru for his understanding and unwavering support during weekends and evenings filled with thesis writing, sharing both the frustrations and the joys, along with countless cups of chai.