COCHRAN, STEVEN NICHOLAS, M.S. Genetic Switching during Protein Synthesis and the Role of Quadruplexes in HIV-1's *nef* Gene. (2018)
Directed by Dr. E. Will Taylor. 28 pp.

HIV-1 is still a prominent global public health issue responsible for the deaths of over 1 million people annually. It is a very complex virus that has been extensively studied over the past decade. The virus has adapted several mechanisms to maintain a compact genome size while also producing required essential gene products in order to successfully replicate itself in an efficient manner. One of the ways it has been able to achieve this goal is to not obey the "rules" of the central dogma of molecular biology. The flow of genetic information from DNA to RNA to protein, and with the outdated idea of one gene producing one protein, is oversimplified with our current knowledge of the virus. One way to maintain its strict, compact genome size requirement is to have inherent multiple reading frames in the same copy of its RNA. By utilizing a ribosomal frameshift, HIV-1 is able to produce multiple proteins without the addition of more genetic material. This has been extensively studied in the past, namely in the Gag-Pol polyprotein, which is only produced as a result of a ribosomal frameshift in the -1 direction. We are presenting a new mechanism by which HIV-1 induces a -1 ribosomal frameshift.

Nef is a protein produced by the *nef* gene of HIV-1 and functions to improve viral replication and pathogenicity. It is known to downregulate CD4 and MHC Class I receptors in order to evade the host's immune cell response. Deletion of the *nef* gene in the virus has been shown to prevent long-term patients who have been diagnosed with

HIV-1 from progressing to AIDS, indicating the importance of the Nef protein and its role of pathogenesis.

There are several requirements for a ribosomal frameshift to occur. These requirements, containing a heptameric sequence with the format of 'X XXY YYZ' and a pseudoknot or other RNA secondary structure separated by a length of 6-12 oligonucleotides, are all found in the *nef* gene. We hypothesize a -1 ribosomal frameshift is induced by the formation of a quadruplex (QPX). By utilizing an *in vitro* frameshift assay with a dual reporter fluorescence protein vector, where we have cloned two copies of *nef* gene fragments (one wild type and another with 3 specific bases modified to remove potential G QPX formation) in between a cerulean fluorescent protein and an orange fluorescent protein, we have found that the QPX in this region of *nef* is essential for inducing a -1 ribosomal frameshift. When three specific guanines in the *nef* region are modified into adenine bases, the ribosomal frameshifting event that occurs in the wild type is significantly reduced ($p<0.0001$), highlighting the importance between the formation of a QPX in the *nef* gene of HIV-1 and its role in inducing a frameshift in the -1 direction during protein synthesis.

GENETIC SWITCHING DURING PROTEIN SYNTHESIS AND THE ROLE OF QUADRUPLEXES IN HIV-1'S *NEF* GENE

by

Steven Nicholas Cochran

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2018

Approved by

_____
Committee Chair

APPROVAL PAGE

This thesis written by STEVEN NICHOLAS COCHRAN has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

1.1 Human Immunodeficiency Virus Type 1 Origin and Impact

Human Immunodeficiency Virus (HIV) is believed to have originated from the Democratic Republic of Congo dating back to 1920. It is structurally similar to Simian Immunodeficiency Virus (SIV), an immune deficiency disease found in monkeys and apes. The link between SIV and HIV stems from one particular SIV strain, called SIVcpz, that infects chimpanzees and is almost identical to HIV-1 in structure (1). SIVcpz was then transmitted to humans when chimpanzees were hunted for their meat and eaten (6,34). The virus is transmitted via transmission of bodily fluids, notably blood (34). The virus was able to spread globally by 1980, and is considered a pandemic. According to the Joint United Nations Programme on HIV/AIDS, it is estimated that 36,700,000 people worldwide are currently living with HIV. Every year, there are 1,800,000 new infections and about 1,000,000 deaths due to complications from AIDS-related illnesses across the world. Since the start of the epidemic, there have been more than 35,000,000 deaths (1,2). While there have been overall recent reductions in new infections in the world population due to access of healthcare and increased education, the importance of HIV research and potential therapeutics cannot be overstated.

1.2 HIV Virology and Structure

HIV is a part of the lentivirus subfamily of retroviruses, that, once allowed to replicate, causes chronic infection to the organism and degradation of its immune system (3). HIV-1 targets all cells with a CD4 receptor, including CD4 T-cells, macrophages, dendritic cells, and monocytes, but mainly targets activated T-lymphocytes. It is spherical in shape and is around 120 nm in diameter, which is markedly smaller (about 60 times) than the cells it infects (4). It contains a double-stranded RNA genome about 9.2 kb in size and only 9 genes: *gag, pol, env, vif, vpr, vpu, tat, rev,* and *nef,* which are held inside a viral protein matrix, p24, also known as a capsid. Surrounding the capsid is a lipid envelope which contains several glycoproteins embedded in the membrane. Of note, gp120 and gp41 are named after the sizes in kDa and are ultimately responsible for binding, fusion, and infection of host cells (3,5). Figure 1 shows the schematic diagram of HIV-1.

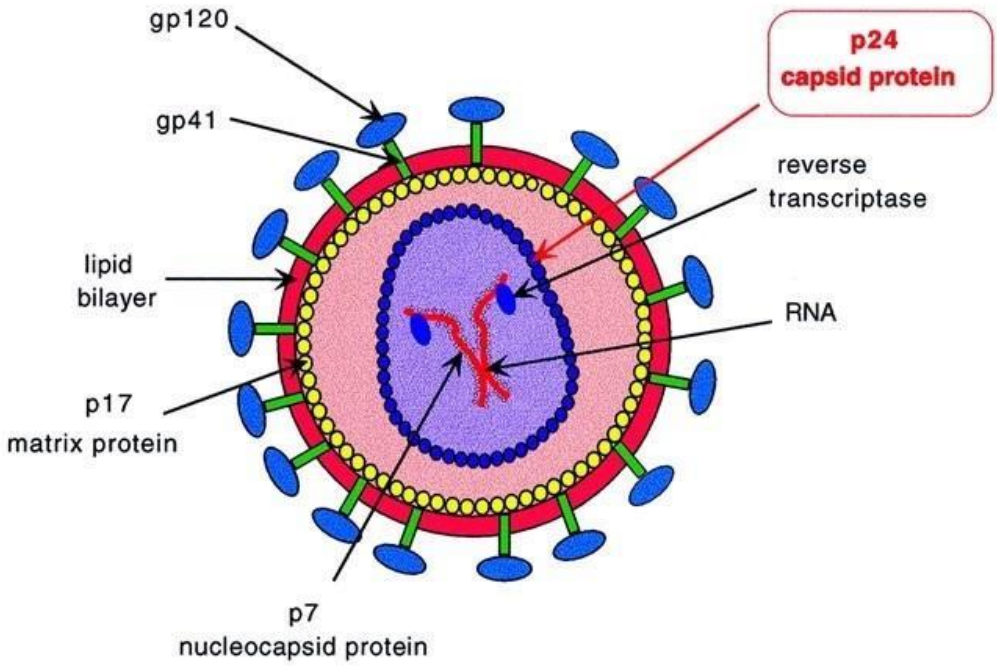

Figure 1. Schematic Diagram of HIV-1 with Labeled Proteins and Enzymes. (Berthet-Colominas, C et al. "Head-to-Tail Dimers and Interdomain Flexibility Revealed by the Crystal Structure of HIV-1 Capsid Protein (p24) Complexed with a Monoclonal Antibody Fab." (*The EMBO Journal* 18.5 (1999): 1124–1136.)

Since HIV is a retrovirus and cannot replicate on its own, it must utilize the host's cellular machinery in order to produce viral proteins and proliferate more virions. The growth cycle is broken down into four stages:

1. Infection
2. DNA formation and integration
3. Gene production
4. Virus assembly and maturation

Figure 2 shows the major biochemical steps that occur in the growth cycle of HIV-1.

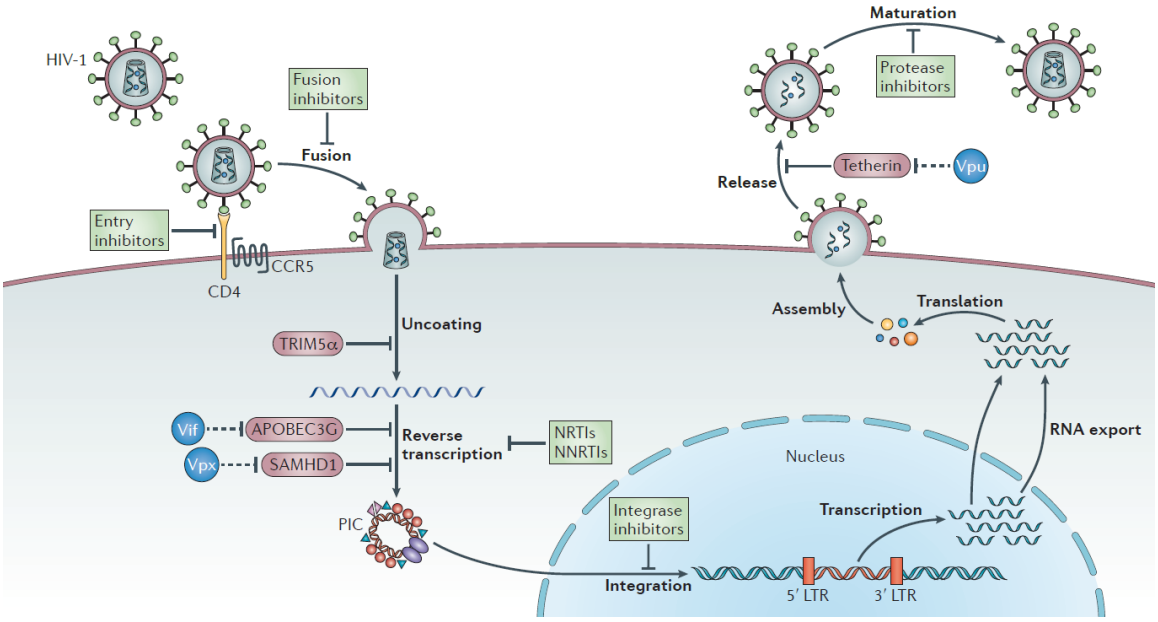

Figure 2. The Major Biochemical Steps for HIV-1's Growth Cycle. The red boxes indicate HIV restriction factors (TRIM5α, APOBEC3G, and SAMHD1). The green boxes represent antiretroviral drug treatments and what stage they target. The green boxes represent antagonist protein targets (Vpu, Vif, Vpx). (Barré-Sinoussi, Françoise, et al. "Past, Present and Future: 30 Years of HIV Research." *Nature Reviews Microbiology*, vol. 11, no. 12, Dec. 2013, pp. 877-883.)

A typical life cycle for HIV begins with binding of gp120 to the cell's CD4 and coreceptors, either CCR5 or CXCR4, which fuses the virus to the cell and allows the entry of the viral genetic material once the capsid is degraded. Once inside, reverse transcriptase (RT) converts viral RNA into its complementary DNA, while integrase incorporates this new viral DNA once inside the nucleus into the host's genome. The DNA is then transcribed normally and produces the mRNA copy, which is then translated into viral polyproteins. Once these proteins are assembled, the virus is able to bud out of the cell. During maturation, proteases cleave the polyproteins and fully functional, infectious viruses are able to continue the cycle (10). Once infected, cells contain

400,000-2,500,000 virions (3). The HIV replication cycle happens relatively fast, in about 2.6 days on average, and is characterized by a high mutation rate, because the viral polymerase (reverse transcriptase) is not subject to cellular checkpoints and DNA repair mechanisms. These errors are usually present during the reverse transcription and RNA polymerase II stages of the cycle (14,15). Most CD4 T-cells that are infected die eventually due to severe cytotoxicity; however, resting T-cells can act as a reservoir for dormant HIV, resulting in latent infection that allows the virus to remain undetected for about 44 months, even if an infected person is undergoing anti-retroviral therapy (12,13). Taken together, both the high rates of mutation and virus latency help the virus avoid the host's immune system and develop resistance to anti-retroviral drugs.

1.3 HIV Genome and Gene Products

HIV-1 contains a guanosine triphosphate cap on its 5' end, a sequence of repeating adenosine bases known as a poly-A tail on its 3' end, and nine open reading frames (ORF) that vary in size and sometimes overlap (7). These function as a way to assist in binding of the sequence to ribosomes in a cell, stabilize the nucleic acid and prevent it from enzymatic degradation once inside a cell, and allow a smaller, more compact genome size with improved efficiency, respectively. HIV produces 15 different protein products from its nine genes. The three genes *gag*, *pol*, and *env*, are found in all retroviruses and are responsible for essential polyproteins that are required for virion formation and propagation. *Gag*, (group-specific antigen) codes for proteins that make up the viral matrix (MA or p17), capsid (CA, or p24), nucleocapsid (NC or p9), and p6, which is required to separate the virion from the cell membrane (15,16). *Pol*

(polymerase) codes for the three major enzymes required to produce viral DNA and subsequent proteins, protease, reverse transcriptase, and integrase. *Env* synthesizes gp160 and, upon glycosylation and cleavage, produces two subunits: gp120 and gp41 which were previously described (1,7,8,9). Two remaining genes, *tat* and *rev* produce essential regulatory proteins that aid in viral propagation. *Tat* (trans-activator of transcription) binds to an RNA stem loop structure which recruits other transcription factors in order to enhance viral transcription (17). *Rev* (regulator of virion proteins) localizes near the nucleus of the cell in order to export the incompletely spliced *gag*, *pol*, and *env* gene products out of the nucleus and allow them to be translated. The final four remaining genes, *vpr* (viral protein R), *vpu (viral protein U)*, *vif* (virion infectivity factor), and *nef* (negative factor) are involved in the infectivity and pathogenicity of the virus.

Interestingly, in HIV-1, *gag* and *pol* are in separate reading frames (*pol* is in a -1 reading frame compared to *gag*), with the 3' end of *gag* overlapping with the 5' end of *pol* by around 200 bases as seen in Figure 3. Furthermore, Pol proteins are only produced as a fusion protein Gag-Pol, despite *pol* having no start codon in its RNA sequence. The *gag-pol* frameshifting sequence has been studied extensively by Varnam et al. in 1988 (18). They determined the Gag-Pol fusion proteins are synthesized when the ribosome has a frameshift in the -1 sequence relative to *gag*, and only occurs around 5-10% of the time. This process is under translational control. The N-terminus of the *pol* gene encodes the protease that cleaves itself from the Gag-Pol protein, the reverse transcriptase enzyme, and the integrase as previously described. Figure 3 depicts HIV-1's genome and the relative positions of its genes, as well as the various frameshifts and overlaps in genes

(19). The *tat* and *rev* genes also have two coding exons each that are separated on either side of and overlapping the *env* gene.

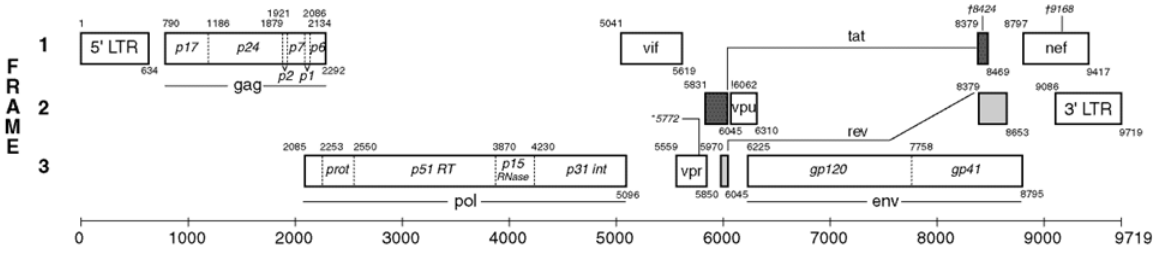

Figure 3. The Genome of HIV-1. The open reading frames (ORF) are shown in rectangles with the location of the genes' starting positions shown by the number in the top left corner. Korber et al., Numbering Positions in HIV Relative to HXB2CG, in the database compendium, *Human Retroviruses and AIDS*, 1998.

1.4 Genetic Code Expanded Through Reading Frames

The classic theory in molecular biology, which describes the flow of genetic information from DNA to RNA to protein, has since been outdated. While this is generally understood as a universal theme among living organisms, that DNA is transcribed into RNA, which is then read by ribosomes in codons of 3 bases long, it fails to describe the flow of genetic information of retroviruses. The previous description "one gene, one protein", with the advancement of molecular biology and genetics, is no longer true and is an oversimplification. As previously described, there are multiple contradictions between the central dogma and HIV-1. First, HIV-1's genome codes for reverse transcriptase, which goes the opposite direction of the classic theory; it takes viral RNA and generates complementary double stranded DNA, which the host cell will then transcribe back into RNA, and later makes proteins. This adds a layer of complexity to how the flow of information exists in modern biology. Secondly, as shown in Figure 3,

several of the coding exons in HIV-1's genome overlap and do not function normally as a eukaryotic organism's genome would. While the central dogma may not apply completely to all cellular processes, retroviruses have been able to take advantage of the "rules" of biology and become successful parasites.

The 9.2 kb size of HIV-1's genome (7), in comparison to a human's 3 billion kb, is drastically smaller. As a result, retroviruses have evolved mechanisms in order to maximize the efficiency of their genome while maintaining its small size. One way to achieve a feature of alternative RNA splicing that is akin to the eukaryotic genome is by utilizing ribosomal frameshifting and in some cases, termination suppression. In addition to the Gag-Pol fusion protein formed from the *gag-pol* genes, ribosomal frameshifting in the -1 reading frame has also been seen in the *env* gene via a "hungry codon" phenomenon (20, 29). The readthrough of a UGA stop codon has been shown in the *nef* gene via "anti-sense" tethering between viral mRNA and host selenoproteins (21). These phenomena can allow the production of multiple isoforms of protein from the same sequence of RNA. In nature, these processes occur at a low frequency which suggests the proteins that are formed as a result of frameshifting or readthrough of a stop codon are not critical to survival of the virion, but instead modulate pathogenicity, virion infectivity, and signal transduction pathways.

Codons are read in triplet from an mRNA strand in order to produce amino acid polymers and form proteins. Protein synthesis starts with the AUG codon and ends with one of three stop codons, UGA, UAA, and UAG. Codons are degenerate, meaning

multiple codons, when read by the ribosome, will produce the same amino acid. For example, the amino acid alanine is produced by the mRNA sequences GCU, GCC, GCA, and GCG. The third position in the codon, also known as the wobble position, is degenerate and less important to determining which amino acid will be added to the chain (23). Regardless of what the third nucleic acid in the codon is, if the first two are both read as "GC", alanine will be added to the growing polymer by the tRNA molecule during protein synthesis. There are three reading frames for a 5' to 3' sequence of mRNA: zero, which is the standard reading frame, +1, and -1. Ribosomal frameshifting occurs when the ribosome is displaced one nucleic acid in the +1 or -1 direction. In nature, the frequency of a frameshift occurring is low but not uncommon and is not limited to only viruses (24).

1.5 HIV-1's *nef* Gene Exhibits Ribosomal Frameshifting

Nef is a small, 27 kDa, myristoylated protein produced in the early stage of the HIV-1 life cycle and is known to be fundamental to viral pathogenicity and replication. Once it is produced, it is localized to the cell membrane. It downregulates both CD4 receptors and MHC class 1 molecules that are found on the surface of cells that signal an immune response. Nef internalizes CD4 receptors and redirects them into the cell's lysosomes, suggesting that Nef interacts with intracellular trafficking pathways (22). It has been shown that deletion of the *nef* gene in humans with long-term nonprogressive HIV-1 halts the progression of AIDS, which indicates its role in disease progression (23). The *nef* gene exhibits a potential overlapping region with an "ideal" XXXYYZ pattern -1 frameshifting signal upstream from a G-quadruplex (QPX), which we suspect induces

and regulates frameshifting. QPXs have been studied extensively in eukaryotes, but only to a lesser extent have they been studied in viruses. Several quadruplex-stabilizing drugs have been found and documented, notably TMPYP4, berberine, and Reservatrol. Monovalent cations, in particular, Na+ and K+ ions, have been shown to help stabilize quadruplexes as well. The scope of this project is to better understand what happens during a ribosomal frameshift, as well as the function the G-quadruplex has in *nef*. If we are able to understand the function of a QPX and its role on *nef*, we could potentially use them as a target for pharmaceutical agents, possibly affecting its pathogenicity or reduce its virulence.

CHAPTER II

G-QUADRUPLEX AND PSEUDOKNOT FUNCTION IN THE HIV-1 NEF CODING REGION AS A FUNCTION OF CO-TRANSLATIONAL GENETIC SWITCHING DURING PROTEIN SYNTHESIS

2.1 Abstract

Quadruplexes (QPX) are formed as a result of repeating guanine bases in nucleic acids that form tetrads. These G-tetrads are stabilized by Hoogsteen hydrogen bonding, a unique variant of Watson-Crick base pairing. QPX also feature a monovalent cation positioned in the middle of the tetrad for additional electrostatic stability. In the coding region of the *nef* gene of HIV-1, there is an ideal heptameric "slippery" sequence, UUUAAAA, that is 7 base pairs upstream from a QPX. This region has been shown to feature a -1 ribosomal frameshift. Because this region is in the highly stable polypurine tract, a highly conserved region in all variants of HIV-1, we suspect this region is responsible for formation of a QPX and its role in inducing a ribosomal frameshift in the -1 direction. Using an *in vitro* frameshift assay and a dual-reporter fluorescent protein vector, in which cloned presumed *nef* gene fragments have been added in between two fluorescent proteins, we have shown that the -1 ribosomal frameshift is dependent upon a QPX being formed.

2.2 Introduction

G-quadruplexes (QPX) are nucleic acid structures that form in guanine-rich sequences of both DNA and RNA. Repeating G bases fold over and stack on top of one another, forming inter- or intramolecular four-stranded structures, known as a G-tetrad. They are stabilized by Hoogsteen hydrogen bonding, a unique variant of standard Watson-Crick base pairing, with a monovalent cation (usually Na+ or K+) located at the center sandwiched between the pairs of tetrads for additional stability through electrostatic interactions (27). QPXs can fold intramolecularly from a single strand of repeating G bases, or intermolecularly through dimerization or tetramerization of separate sequences. Depending on the orientation of the sequences, they can be designated as parallel, where the 5’ and 3’ ends line up together, antiparallel, where one strand is 5’ to 3’ and another strand runs opposite in the 3’ and 5’ direction, or mixed, a combination of the two (27). The tetrads that result from these secondary interactions are planar and can stack on top of one another, forming a QPX, as shown in Figure 4.

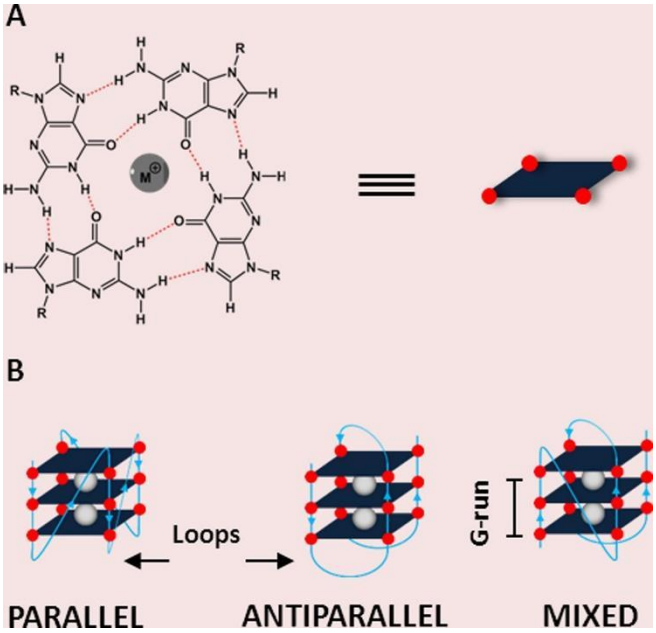

Figure 4. Chemical Structures and Schematic Planar Representations of QPXs. A) *left*: a chemical structure of the tetrad shown of 4 guanine bases surrounding a monovalent cation, shown as M+. *Right*: the planar representation of the QPX with the red dots each signifying a guanine. The red dashed lines indicate Hoogsteen base pairing. B) the folding orientations parallel, antiparallel, or mixed. (Ruggiero, E. et al.; G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy, *Nucleic Acids Research*, Volume 46, Issue 7, 20 April 2018, Pages 3270–3283)

QPXs have been identified in numerous genomes including mammalian, protozoa, yeasts, bacteria, and of course, viruses, and are a focus of nucleic acid research. QPXs are not randomly distributed but are generalized in areas of genomic importance, namely telomeres, oncogenes, gene promoters, and origins of DNA replication (25,27). They have also been found in other open reading frames and untranslated regions of the eukaryotic genome, suggesting they are involved in regulating biological pathways such as transcription, translation, replication, and stability of the genome. In the example of telomeres, QPXs have been shown to inhibit telomerase, the enzyme responsible for

adding bases to the end of chromosomes, which prevents extension of the sequence of DNA. If QPXs are able to stabilize and prevent telomerase activity in tumor cells, a new treatment could be proposed in treating cancer and limiting tumor growth (25). If we can achieve a better understanding of the mechanisms behind the formation, structure and stability of QPXs, we could potentially use them as drug targets for clinical applications.

QPX formation in the *nef* gene of HIV-1 has been previously studied and documented using computational and experimental approaches (28). Richter et al. determined three potential QPX formation sites in the *nef* gene, Nef 8528, Nef 8547, and Nef 8624 (Figure 5). Ribosomal frameshifting in the *nef* gene has already been shown to occur (21). With this information and using the potential QPX locations as described in the Richter study above, we are proposing that one of the three QPX that they identified, Nef 8624, is essential in order for the -1 ribosomal frameshift to occur in the *nef* gene. If confirmed, we can demonstrate the co-translational genetic switching that occurs between multiple protein isoforms in HIV-1. Additionally, if we are able to inhibit the QPX and the resulting ribosomal frameshift in *nef,* we could prevent the formation of these protein isoforms, which could serve as a potential target for antiviral drugs.

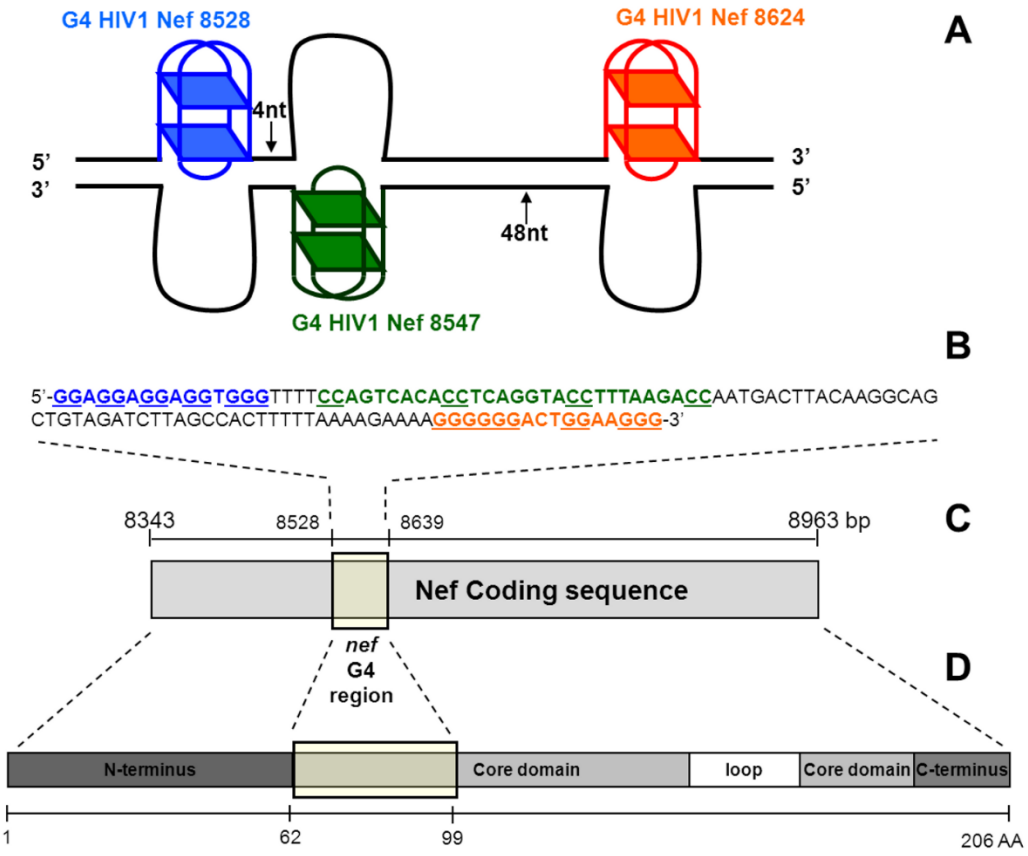

Figure 5. The Three Putative QPX Formation Sites in the *Nef* Gene and Corresponding Sequence and Location in the Gene. In blue, Nef 8528, green, Nef 8547, and orange, Nef 8624. (Perrone, Rosalba et al. "Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 *nef* Coding Region: Implications for Antiviral Activity." Ed. Jianming Qiu. *PLoS ONE* 8.8 (2013): e73121. *PMC*. Web. 20 June 2018.)

As mentioned before, ribosomal frameshifting is a result of the ribosome in a cell that, instead of reading in the zero frame, moves forward or backward one single nucleotide. HIV-1 has several ORF's that spans its entire genome as seen previously in Figure 3. The formation of the Gag-Pol and fusion protein exemplifies ribosomal frameshifting in the -1 direction, enabling it to regulate the expression of its genes while keeping its

genome relatively compact (29). There are three requirements for the induction of a -1 ribosomal frameshift:

1. A heptameric slippery sequence where the frameshifting is set to occur. For optimal efficiency, the sequence 'X XXY YYZ' is considered to be ideal due to the degeneracy of codons. The wobble position is the least important when determining the amino acid that will be produced. Because of this, when the tRNA molecule is ready to add an amino acid to the growing peptide and it encounters the code 'XXY', slippage could occur in the -1 reading frame and the ribosome will read that codon as 'XXX'. The degeneracy of the code and the wobble position allows this frameshift and translation of the new sequence to occur.
2. A spacer sequence about 6-12 nucleotides long.
3. A downstream pseudoknot or other large RNA structure such as a hairpin or stem-loop.

The exact mechanism of the frameshift and how it is induced by encountering a pseudoknot or QPX is not entirely clear. It has been shown that when the ribosome is translating a sequence of mRNA and encounters one these structures, it pauses over the heptameric slippery sequence. During the pause, the ribosome is able to slip into another reading frame, causing the frameshift. The downstream pseudoknot or QPX is essential for this process to occur in the -1 direction. However, since the sequence that makes up the pseudoknot or QPX also needs to be read by the ribosome and translated into protein, the stability and structure of the pseudoknot or QPX is also important to its function. If the structure is too stable, the ribosome may not be able to translate the remainder of the

mRNA. Quadruplexes in DNA have been shown to stall replication, transcription, and inhibit DNA binding protein function in cells, possibly linking its role in cell division and development (30). The geometry and surface charge of a QPX might also have an effect on the efficiency of the frameshifting rate.

The putative Nef 8624 site described by Richter et al. happens to be located just a few bases downstream of an ideal heptameric slippery sequence required for a -1 frameshift, UUUAAAA. Downstream to this sequence, there is an RNA pseudoknot located about 7 nucleotide bases away, which overlaps exactly with the region containing the potential QPX Nef 8624. This sequence is located in a highly conserved region of HIV-1's RNA called the polypurine tract (PPT), an area of repeating purines. PPTs are relatively resistant to degradation by the ribonuclease H activity of reverse transcriptase and is also the site of plus-strand DNA synthesis (31). When the UUUAAAA sequence is mutated to UUUAAGA, (changing the codon AAA that produces lysine into AAG, coding for arginine), and the cell is arginine deficient, arginine acts as a 'hungry codon'. A hungry codon is described when a particular amino acid, in this case arginine, is required to for peptide formation, but the aminoacyl tRNA is in short supply. Arginine is known to play a role in preventing oxidative stress, so if the cell is Arg deficient, and when the ribosome encounters the codon for Arg, AAA, it pauses, and a -1 frameshifting event is more likely to occur (20). Similarly, we are hypothesizing that when the ribosome encounters a QPX or pseudoknot, it pauses, resulting in a -1 frameshift. The QPX of Nef 8624 is located in the PPT and has a potential role in enhancing the -1 ribosomal frameshifting frequency by causing the ribosome to pause during translation. Because Nef 8624 was identified as

having the highest stability of the three possible QPX formation sites (28), we hypothesize that when the ribosome encounters this stable structure, it briefly pauses due to a physical obstruction, enhancing the -1 ribosomal frameshift.

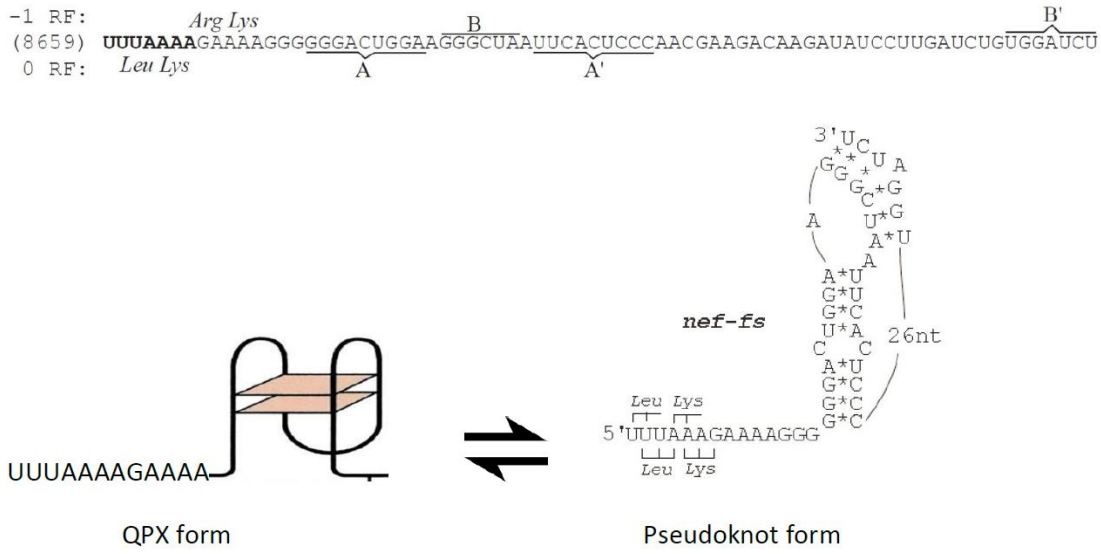

Figure 6. *Top*: The mRNA Sequence in HIV-1's *Nef* Gene that can Assume the Quadruplex. *Bottom left*: Quadruplex form. *Bottom right*: Psuedoknot form. These structures show the alternate molecular conformations of the viral RNA. In this study, we hope to directly link the QPX that is believed to form in Nef 8624 and its role in inducing a -1 ribosomal frameshift. If we alter the *nef* gene and eliminate the suspected QPX formation, frameshifting will not occur, implicating its role in the process.

2.3 Materials and Methods

The first step was to create two expression vectors containing the putative HIV-1 nef frameshift site previously predicted by Taylor and coworkers (21), in one of which is cloned the wild-type putative QPX-forming *nef* gene fragment; the second vector contains a mutated, QPX-inhibited *nef* gene fragment. In both constructs, the wild-type or mutated nef frameshift sequences are inserted between two fluorescent protein reporter genes, in order to assess the frameshift efficiency *in vitro*. The QPX-inhibited *nef* gene fragment was created to mutate specific guanine bases into adenine bases, effectively disabling the possibility of forming a QPX. Three specific guanines were mutated into adenines in order to remove the potential for QPX formation. This maintains the slippery sequence that is critical to enable the -1 frameshift, as well as the predicted RNA hairpin loop, and only changes one amino acid (glycine to glutamine) in the transframe fusion protein sequence, which should not affect protein yield significantly. These three guanine bases were selected because their absence (via mutation to adenines) would have the highest probability of disabling formation of the QPX, considering the work on the putative QPX formation study done by Richter et al. (28). The construct was designed to have an upstream crulean fluorescent protein (CFP) reporter gene and a downstream orange fluorescent protein (OFP) reporter gene that is in the -1 reading frame. When allowed to grow, all the transfected cells will produce CFP. However, because the OFP is in the -1 reading frame, orange fluorescence will only occur when a frameshift happens. Fluorescence was measured using an EVOS fluorescence microscope and quantified using ImageJ NIH software. If our hypothesis is correct, the frameshift frequency of the

mutated, QPX-inhibited *nef* gene fragment is expected to be significantly less than the wild type *nef* gene fragment.

2.3.1 Plasmid Construct

To construct our plasmids, we used the mCerulean C1 (4731 bp) expression vector which contains the enhanced cerulean fluorescent protein (ECFP) gene (722 bp) upstream of the multi cloning site (added by AddGene.org). The *nef* gene has an inherent BglII (A'GATCT) restriction site upstream of the frameshift site, which allows the ligation of the nef gene inserts via another BglII restriction site in mCerulean C1, which was digested using the BglII restriction enzyme at the multicloning site. ECFP is the reporter gene upstream from the inserted nef frameshift sequence in the constructs. We are using pmOrange N1 mammalian vector (obtained from Clontech lab. Inc.) as a source for the orange fluorescent protein gene with multicloning sites added, which is located between a BamHI (G'GATCC) restriction site and a downstream multicloning site. ECFP was first ligated with the *nef* inserts via the BglII site, which produced ECFP attached with the wild type (WT) *nef* and mutated (mut) *nef* gene fragments. A second ligation occurred between the attached ECFP-*nef* into the mOrange vector, using a BamHI restriction site. The closed, completed construct was achieved using NheI sites to ligate both mCerulean and pmOrange.

2.3.2 Plasmid *nef* Inserts

Nef-fs Wild type (*nef*-fs WT):

ctg tag atc tta gcc act ttt taa aag aaa agg ggg gac tgg aag ggc taa ttc act ccc aaag

Nef-fs WT mutated to create 3' BamH1 site:

ctg tag atc tta gcc act ttt taa aag aaa agg ggg gac tgg aag ggc taa ttc gga tcc aaag

Nef-fs mutant to disrupt quadruplex formation (*nef*-fs mut):

ctg tag atc tta gcc act ttt taa aag aaa aga agg gac tgg aag agc taa ttc gga tcc aaag

The bases in blue indicate the BglII site, while the bases in red indicate the BamHI site that was added to the end of the *nef* gene. The bases in teal indicate the specific guanine bases that were mutated into adenine. The sequences in the middle indicate the -1 frameshifting site and location of the proposed QPX formation. The nef-fs WT and nef-fs mut were designed and ordered from IDT Inc. The inserts were mutated to create a BamHI site at the 3' end in order to ligate with pmOrange. *Nef* originally has a BglII restriction site upstream from the QPX-forming region.

2.3.3 *In Vitro* Frameshifting Assay

The construct was transformed into JM109 competent cells and grown on lysogeny broth (LB) plates containing kanamycin. The clones were harvested, amplified, and purified using a DNA purification kit. The plasmids were sent to Eurofins Genomics to be sequenced to confirm the formation of the construct. HEK-293 mammalian cells were incubated at 37° C and 5% CO_2 atmosphere on a black clear-bottom 96 well plate seeded with 0.5x10^5 cells per well. After growing to confluency, the cells were transfected with the plasmid constructs, Lipofectamine 2000 (Invitrogen), OptiMEM I,

and 10% FBS-enriched DMEM for 12 hours, after which the media was replaced with 100 μl of the enriched DMEM and allowed to grow for 4 days. The expressed fluorescence protein levels were then observed using an EVOS fluorescence microscope. The parent plasmids mCerulean and pmOrange were used as positive controls, while untransfected cells were used as a negative control. Fluorescence images were taken at 10 different well locations for each variable treatment. The intensity in relative light intensity units was measured using NIH ImageJ software (35). Statistical analysis was done using Graphpad software.

2.4 Results and Discussion

CFP and OFP intensities of both nef-fs WT and nef-fs mut were measured and compared to the controls as listed above. The upstream CFP was expressed in both nef-fs WT (3780±498) (SEM) and nef-fs mut (5675±962) (SEM) constructs, with only a minor decrease in nef-fs WT as compared to nef-fs mut. This relative frameshifting difference is not statistically significant and this is within experimental error ($p=0.0998$). However, there is a highly significant ($p<0.0001$) decrease in OFP expression shown in nef-fs mut (211 ± 200) (SEM) as compared to nef-fs WT (1983 ± 255) (SEM). This indicates the nef-fs WT construct undergoes frameshifting in the -1 direction and expresses OFP, while nef-fs mut, with its mutated DNA sequence that inhibits the putative quadruplex from forming, does not. As expected, the untransfected cells produced no fluorescence and, compared to nef-fs mut OFP production, there is no significant difference ($p=0.31$). In the nef-fs WT construct, there is a highly significant ($p<0.0001$) amount of frameshifting occurring (OFP 1983 ± 255) (SEM) as compared to the untransfected cells

(OFP 0 ± 0). These results are consistent with our hypothesis that the previously identified quadruplex in the *nef* gene of HIV-1, that by serving as the RNA structural element in the bipartite frameshift signal (along with the required slippery sequence), plays an essential role in inducing a -1 ribosomal frameshift in this important viral gene. This is a functional role of a QPX that has not been reported previously.

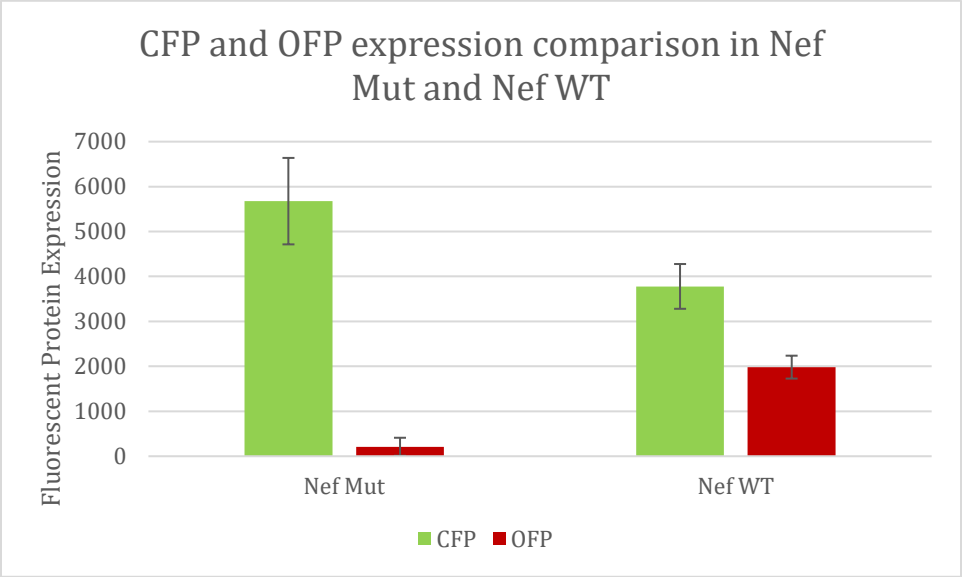

Figure 7. CFP and OFP Expression Levels in *Nef*-fs mut and *Nef*-fs WT Constructs.

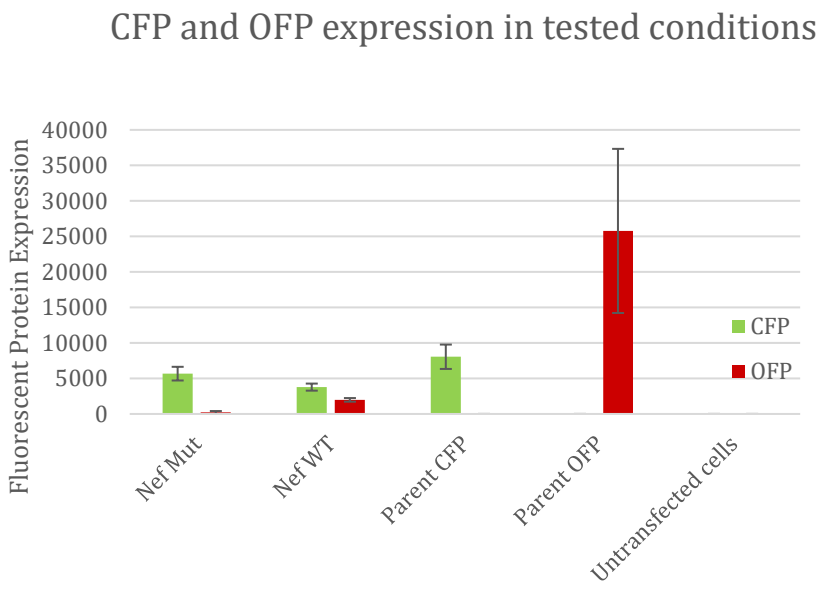

Figure 8. CFP and OFP Expression Levels in the Tested Variables. From left: nef-fs mut, nef-fs WT, parent CFP, parent OFP, and untransfected cells.

2.5 Future Work and Research

A 0% frameshift efficiency with modified stop codons could be added as an additional negative control and the difference between the nef-fs mut, QPX-inhibited construct could be compared. QPX-stabilizing drugs such as TMPYP4 could be added to both *nef*-fs WT and *nef*-fs mut in order to further showcase the decreased efficiency of frameshifting that occurs when a QPX is unable to form, as TMPYP4 is expected to have no effect on the *nef*-fs mutant construct. Further future work could be done using a different neuroblastoma cell line. Previously, when neuroblastoma cells were introduced to aconitine, a sodium channel opening drug, a decrease in stability of the QPX found in the ChAT gene was observed (32). An assay involving various monovalent cations in

different concentrations using this cell line could help determine the role of Na+, K+, and Li+ in quadruplex stability.

REFERENCES

1. UN Joint Programme on HIV/AIDS (UNAIDS), *Global AIDS Update - 2016*, June 2016,
2. Fettig, Jade et al. “Global Epidemiology of HIV.” *Infectious disease clinics of North America* 28.3 (2014): 323–337. *PMC*. Web. 12 June 2018.
3. Hutchinson, J. F. The Biology and Evolution of Hiv. *Annual Review of Anthropology* 2001, *30* (1).
4. Dubey, Anubha et al. “Machine learning model for HIV1 and HIV2 enzyme secondary structure classification.” (2011).
5. Various (2008). *HIV Sequence Compendium 2008 Introduction* (PDF). Retrieved June 11, 2018.
6. Gao, F. et al (1999) 'Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes' Nature 397(6718):436-441
7. Watts, Joseph M. et al. “Architecture and Secondary Structure of an Entire HIV-1 RNA Genome.” *Nature* 460.7256 (2009): 711–716. *PMC*. Web. 13 June 2018.
8. HIV i-BASE. “HIV Genome: Genetic Structure and Function of HIV Explained.” *i-Base*, i-base.info/qa/factsheets/hiv-genome-explained.
9. Beilhartz, Greg L., and Matthias Götte. “HIV-1 Ribonuclease H: Structure, Catalytic Mechanism and Inhibitors.” *Viruses* 2.4 (2010): 900–926. *PMC*. Web. 13 June 2018.
10. Barré-Sinoussi, Françoise, et al. "Past, Present and Future: 30 Years of HIV Research." *Nature Reviews Microbiology*, vol. 11, no. 12, Dec. 2013, pp. 877-883. EBSCO*host*, doi:10.1038/nrmicro3132.
11. Berthet-Colominas, C et al. “Head-to-Tail Dimers and Interdomain Flexibility Revealed by the Crystal Structure of HIV-1 Capsid Protein (p24) Complexed with a Monoclonal Antibody Fab.” *The EMBO Journal* 18.5 (1999): 1124–1136. *PMC*. Web. 19 June 2018.
12. Siliciano, J. D., Kajdas, J., Finzi, D., Quinn, T. C., Chadwick, K., Margolick, J. B., Kovacs, C., Gange, S. J., and Siliciano, R. F. (2003). Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nature medicine, 9(6):727--728.
13. Finzi, D. *et al. Nat. Med.* 5, 512–517 (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy.
14. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M., and Ho, D. D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science, 271(5255):1582--1586

15. Mohammadi, P., (2014). “Studies in computational biology of HIV-1” (Doctoral dissertation). Retrieved from https://doi.org/10.3929/ethz-a-010403338
16. Müller, B. *et al.* (2002). Structure and Assembly: The Late-Domain-Containing Protein p6 is the Predominant Phosphoprotein of Human Immunodeficiency Virus Type 1 Particles. J. Viriol. February 2002 76:3 1015-1024.
17. Vaishnav YN, Wong-Staal F (1991). "The biochemistry of AIDS". *Annu. Rev. Biochem*. 60: 577–630.
18. T. Jacks, et al. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol gene expression. Nature 331 (1988) 280-83.
19. Korber et al., Numbering Positions in HIV Relative to HXB2CG, in the database compendium, *Human Retroviruses and AIDS*, 1998
20. E.W. Taylor, et al. (2005) A −1 frameshift in the HIV-1 *env* gene is enhanced by arginine deficiency via a hungry codon mechanism. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 579 (1-2) 125-132.
21. Premadasa, L. (2016). “Co-Translational Genetic Switching during Protein Synthesis: The HIV-1 Nef Gene as a Paradigm.” (Doctoral dissertation).
22. Piguet V, and Trono D. “The Nef Protein of Primate Lentiviruses.” *Reviews in Medical Virology*, vol. 9, no. 2, 1999, pp. 111–20.
23. Kirchhoff, F., PhD., Greenough, T. C., M.D., Brettler, D. B., M.D., Sullivan, J. L., M.D., & Desrosiers, R. C., PhD. (1995). Brief report: Absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. *The New England Journal of Medicine*, *332*(4), 228-232.
24. Dinman, Jonathan D. “Programmed Ribosomal Frameshifting Goes Beyond Viruses: Organisms from All Three Kingdoms Use Frameshifting to Regulate Gene Expression, Perhaps Signaling a Paradigm Shift.” *Microbe (Washington, D.C.)* 1.11 (2006): 521–527. Print.
25. Moye A, Porter, K, Cohen S, Phan T, Zyner K, Sasaki N, Lovrecz C, Beck J, Bryan T. (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. Nature Communications 6, 7643
26. Zhou Z, et al. (2016). Codon usage is an important determinant of gene expression levels largely through its effects on transcription. PNAS October 11, 2016. 113 (41) E6117-E6125
27. Emanuela Ruggiero, Sara N Richter; G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy, *Nucleic Acids Research*, Volume 46, Issue 7, 20 April 2018, Pages 3270–3283
28. Perrone, Rosalba et al. “Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 *nef* Coding Region: Implications for Antiviral Activity.” Ed. Jianming Qiu. *PLoS ONE* 8.8 (2013): e73121. *PMC*. Web. 20 June 2018.
29. Langer S and Sauter D (2017) Unusual Fusion Proteins of HIV-1. *Front. Microbiol*. 7:2152.
30. Maizels, Nancy. “G4-Associated Human Diseases.” *EMBO Reports* 16.8 (2015): 910–922. *PMC*. Web. 21 June 2018.

31. Fitzgerald, Megan E., and Alexander C. Drohat. “Studies of RNA/DNA Polypurine Tracts.” *Chemistry & biology* 15.3 (2008): 203–204. *PMC*. Web. 21 June 2018.
32. Baghaee, S. (2016). “Ionic Modulation of QPX Stability as a Nano - Switch Regulating Gene Expression in Neurons” (Doctoral dissertation).
33. CDC. “HIV/AIDS.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 16 Mar. 2018, www.cdc.gov/hiv/basics/transmission.html.
34. Avert. “Origin of HIV & AIDS.” *AVERT*, Avert, 8 May 2018, www.avert.org/professionals/history-hiv-aids/origin.
35. Abramoff MD, Magalhaes PJ, Ram SJ: Image Processing with ImageJ. Biophotonics International 2004, 11(7):36-42.