PREMADASA, LAKMINI, Ph.D. Co-Translational Genetic Switching during Protein Synthesis: The HIV-1 Nef Gene as a Paradigm. (2016)
Directed by Dr. Ethan W. Taylor. 79 pp.

The old maxim of "one gene, one mRNA, one protein" no longer holds, especially with viral genes. It is possible for one mRNA to encode several proteins of unrelated functions in overlapping reading frames of a single oligonucleotide, or for an additional protein domain to be added on to a protein at the C-terminal by the readthrough of a stop codon. The question of how and when stop does not always mean stop, how slippage from one reading frame into another is controlled, and the factors that trigger those genetic switches, are the subjects of this research.

The focus of the project is the HIV-1 nef gene, which has examples of both of these types of co-translational switching events (translational frameshifting and stop codon readthrough). Nef is a myristoylated protein expressed in the early stage of the HIV-1 life cycle, which functions as a fundamental factor for efficient viral replication and pathogenesis.

One of the notable features of nef is the highly conserved 3'-UGA stop codon, and the potential for the protein to be extended by about 30 amino acids if readthrough of that stop codon can occur. We hypothesize that antisense tethering interactions (ATIs) between viral mRNA and host selenoprotein mRNA enables capture of the host selenocysteine insertion sequence (SECIS) element to enable the expression of virally encoded selenoprotein modules via translation of in frame UGA stop codons as selenocysteine (SeC).

This mRNA hijack mechanism was predicted theoretically using computational analysis and was experimentally supported at the DNA level by gel shift assay. Readthrough of UGA was proved at the mRNA level by fluorescence microscopy image analysis and flow cytometry of transfected HEK 293 cells with engineered reporter gene plasmid vector constructs, in which the downstream reporter gene can only be expressed if the UGA is translated.

siRNA knockdown of thioredoxin reductase 1 (TR1) mRNA in transfected cells resulted in decreased GFP expression, consistent with the hypothesis that host-virus mRNA tethering may enable selenocysteine incorporation for the stop codon readthrough. Furthermore quantitative analysis of TR1 mRNA knockdown demonstrated using RT-PCR confirmed that the siRNA treatment results in approximately 20% knockdown of TR1.

The HIV-1 nef coding region features a potential -1 frameshift site with a potential overlapping gene region near the middle of the coding sequence. A sequence matching the pattern (XXXYYYZ) of a known -1 frameshifting “slippery sequence” signal is present in the nef sequence at this point, immediately upstream of a G-quadruplex (QPX) sequence that serves to regulate frameshifting. An in vitro frameshift assay using a dual reporter vector was constructed, in which the putative HIV-1 nef-fs sequence with QPX was cloned between two fluorescent reporter genes. Cells transfected with this construct showed orange fluorescence, which is only possible if the -1 frameshifting occurs. Treating the transfected cells with QPX stabilizing synthetic drug

TMPYP4 increased the frameshifting efficiency by 27%, specifically confirming the role of the QPX as an enhancer of -1 frameshifting efficiency.

CO-TRANSLATIONAL GENETIC SWITCHING DURING PROTEIN SYNTHESIS: THE HIV-1 NEF GENE AS A PARADIGM

by

Lakmini Premadasa

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Greensboro
2016

Approved by

_____
Committee Chair

APPROVAL PAGE

This dissertation written by Lakmini Premadasa has been approved by the following committee of the Faculty of the Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____
Ethan Will Taylor, Ph.D.

Committee Members _____
Dennis R. LaJeunesse, Ph.D.

_____
Daniel Joseph Christian Herr, Ph.D.

_____
Jan Ruzicka, Ph.D.

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGMENTS

It was a great pleasure for me to conduct my research in the field of Nanoscience. While proceeding with my research I realized that success could only be gained through gradual experience and knowledge. This research could be of great use to those who work directly with these materials.

Firstly, I use this opportunity to extend my sincere gratitude to my supervisor, Dr. Ethan Will Taylor for the invaluable guidance, encouragement and unfailing support throughout my research project that has led to this scientific accomplishment. He was always there for me throughout this research, and never hesitated to allocate his valuable time on my behalf.

Next, I would like to express my gratitude towards my committee members, Dr. Dennis LaJeunesse, Dr. Daniel Herr and Dr. Jan Ruzicka for being on my committee. I am indebted to the academic and technical staff of the Department of Nanoscience of the University of North Carolina at Greensboro, for their continual help that has allowed me to conduct research work.

Finally, I would like to extend my heartfelt gratitude to my loving parents and family members. The encouragement they gave me and the patience they had with me throughout this period is highly appreciated.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

1.1 Basic Virology of Human Immunodeficiency Virus

More than 30 million people are currently infected with human immunodeficiency virus (HIV) worldwide and accounts for more than 25 million deaths. Since the first case of acquired immunodeficiency syndrome (AIDS) reported in the early 1980s, the number of people infected worldwide has increased rapidly turning the disease in to an epidemic that impacts on the health and economy of many nations. A thorough understanding of functional genomic and structural biology of the virus facilitates the development of therapeutic strategies targeted to the inhibition of viral proliferation.

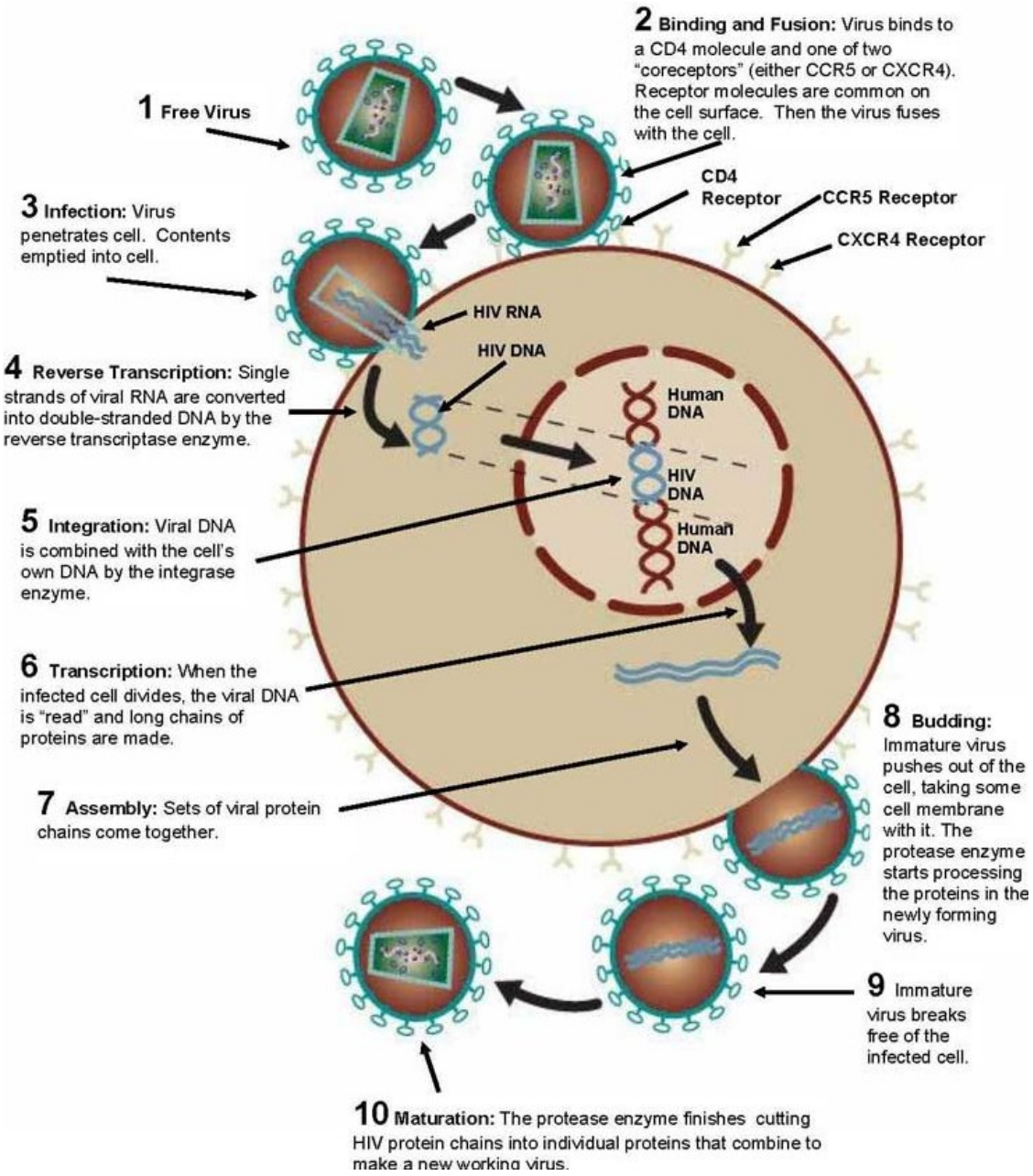
HIV-1 and HIV-2 viruses originated from the simian immunodeficiency viruses (SIVs) of primates. The SIV of chimpanzees (SIV_{cpz}) gave rise to HIV-1 in humans and the SIV of the sooty mangabey monkey (SIV_{sm}) gave rise to HIV-2 in humans[1].They represent two different epidemics. Different strains of HIV-1 have been classified as major (M), new (N) and outlier (O) groups representing the separate zoonotic transfers from chimpanzees. The strains in the group M are diverse and sub classified in to subtypes A-K. These strains are the most responsible for global HIV/AIDS pandemic. Groups N and O are mainly confined to West and Central Africa, though cases of Group O have been found world-wide due to international travel [1].

HIV-1 and HIV-2 are members of the family of retroviruses in the genus of lentiviruses. The viruses are approximately 120nm in diameter enclosed with a lipid envelop and associated with a matrix enclosing a capsid which contains two copies of positive-sense single stranded RNA viral genome and viral encoded enzymes[2].The RNA contains many long and smaller open reading frames (ORFs) and have a 3' poly (A) tail and a 5' cap. The long ORFs encode the viral structural proteins while the short ORFs encode regulators of the viral attachment to the host, membrane fusion, replication and viral particle assembly [2].

HIV and related lentiviruses have growth cycles that are typical of all retroviruses comprising four distinct stages of growth cycle

- Infection
- Reverse transcription and integration
- Viral gene expression
- Virus assembly and maturation

Figure 1 shows the major biochemical steps in the virus growth.



Aids infonet.org

Figure 1. The Major Biochemical Steps in the HIV-1 Virus Growth.

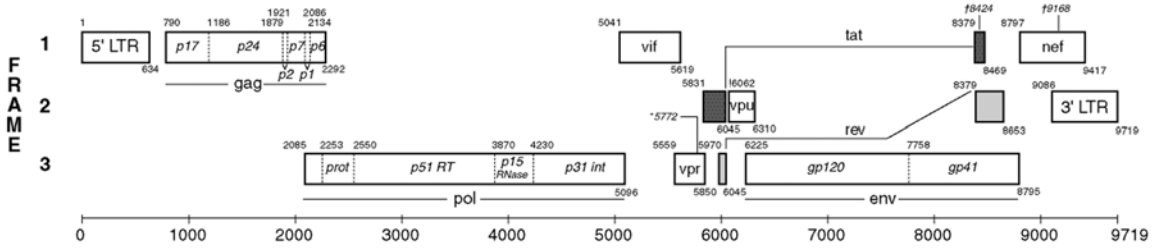
The free virus binds to CD4 receptor molecules that is present on the CD4 positive T (helper) lymphocytes, macrophages and microglial cells and CCR4 or CCR5 core receptors on the cell membrane surface and fuses with the cell leading the penetration of viral genome in to the cell. Then the viral RNA genome is reverse transcribed In to DNA using reverse transcriptase enzyme. Viral DNA is then integrated at random sites in to cellular DNA by the integrase enzyme. Once integrated, proviral genome is subjected to transcriptional regulation by the host cell and also by its own transcriptional control mechanisms. The expression of viral genes produce viral proteins and eventually assemble to make the immature virus and pushed out of the cell along with some cell membrane. The protease enzyme processes the proteins in the newly formed virus particle. Eventually, the immature virus release from the cell and undergo maturation to form the new infectious virus. [1].

1.2 Genome Structure and the Functions of Genes

The HIV genome is tightly compressed. It encodes a total of nine genes which generates 14 different protein products which can be categorized as virion proteins (gag, pol, env gene products), accessory proteins (vpr, vif, vpu gene products) and regulatory proteins (tat, rev, nef gene products) [5].

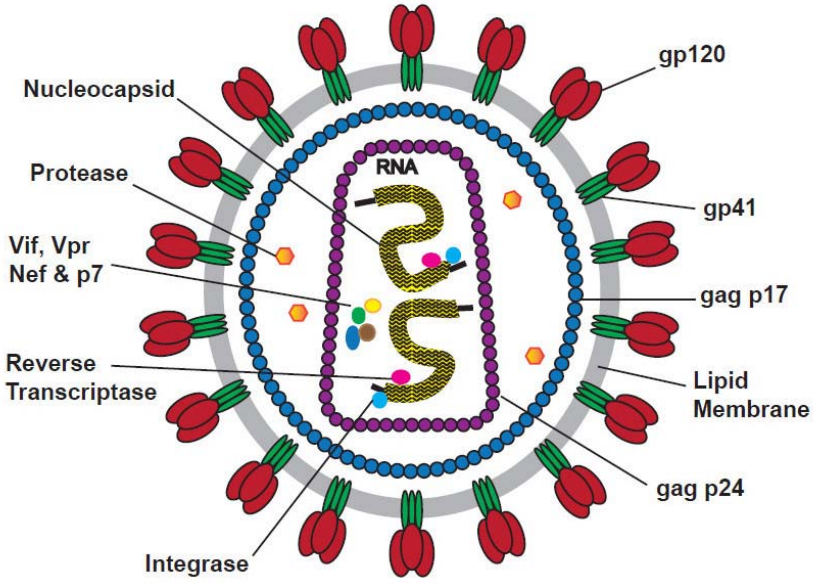
The gag gene produces the virion proteins MA (p17), CA (p24) and NC (p9). P 17 is a matrix protein functional in membrane binding and virus assembly. [6] Capsid protein p24 and neucleocapsid protein p9 involve in virus assembly. The gag and pol genes undergo ribosomal frameshifting. Only 10% of the translated mRNA produces gag-

pol product. The N-terminus of the frameshifted pol gene encodes for the protease gene. Protease is activated in the budding virion particle and cleaves itself, the gag proteins reverse transcriptase (RT) and intigrase (IN). env gene encodes for trimeric transmembrane glycoproteins gp 120 and gp41 which are responsible for the attachment of viral particle to the host cell via CD4 receptor molecules. [7] Nef gene involves in down-regulation of the CD4 and regulates the T cell signaling pathways.



www.hiv.lanl.gov

Figure 2. Human Immunodeficiency Virus – 1 Genome



www.eenzyme.com HIV research tools

Figure 3. Schematic Diagram of the Structure of Human Immunodeficiency Virus

1.3 Extension of Genetic Code

Although the genetic code was "cracked" over 50 years ago, emerging evidence has revealed that there is a deeper level of protein coding information, sometimes called the "second half" of the genetic code. The old maxim of "one gene, one protein" no longer holds, especially with viral genes. Genetic code uses three letter words for amino acids, thus any gene has three different reading frames, a zero reading frame, -1 reading frame +1 reading frame. It is used to believe only one frame is translated and the other frames are artifacts usually containing stop codons, UAA, UAG or UGA. However later discoveries proven that It is possible for one mRNA to encode several proteins of unrelated functions in overlapping reading frames of a single oligonucleotide, or for an

additional protein domain to be added on to a protein by the read through of a stop codon. The question of how and when stop codons do not always mean stop, how slippage from one reading frame into another is controlled, and the factors that trigger those genetic switches, are the subjects of this project.

Due to highly constrained genome size which is limited by their virion structure, specially complex viruses such as retroviruses are under evolutionary pressure to expand the protein coding density.[8] Therefore they have evolved to undergo complex co-translational genetic switching mechanisms to maximize the amount of potential protein coding within the genome. In this project we are discussing about two of those mechanisms, ribosomal frameshifting to produce several different proteins from the same length of oligonucleotide and read through of UGA stop codon to extend the genetic information without stop being a stop.

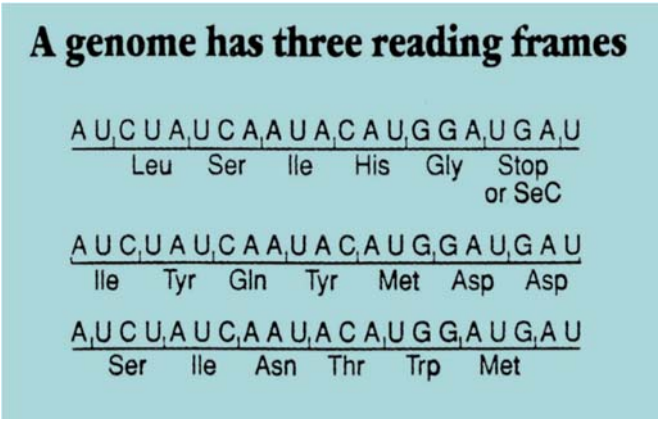
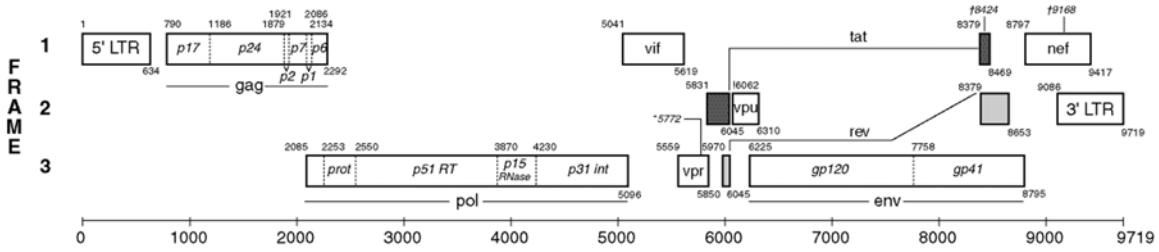


Figure 4. Three Different Reading Frames of a Genome, -1 Frame, +1 Frame and 0 Frame Giving Rise to Three Different Proteins.

Overlapping genes can occur when a single oligonucleotide codes for more than one protein by being translated in multiple reading frames which is seen as a common feature in many RNA viruses such as HIV-1. For example, gag and pol genes, nef and env genes and tat and nef genes are overlapping in HIV-1 genome. These three are three different reading frames which results for different coding proteins upon frameshifting.



www.hiv.lanl.gov

Figure 5. HIV-1 Genome Showing Overlapping Genes.

Ribosomal frameshifting is a way of slipping out of ribosome from one overlapping gene in to another during translation. Apart from overlapping genes, three factors are required for frameshifting which are, a heptameric slippery sequence which should be in the form of XXXYYYZ for optimal frameshifting, a downstream pseudoknot or other large RNA structure like hairpin or stem loop and a spacer sequence between the pseudoknot and slippery sequence. Slippery sequence is the place where frameshifting occurs. The spacer sequence usually should be from 6-12 nucleotides long. Ribosome reads through the gene and when it meets the slippery sequence, the ribosome slips back and therefore the sequence steps back by one base resulting a mismatch in the third base [9]. However, this codon- anti codon mismatch is accepted because it is

occurring only at the third base position. Therefore, the ribosome keeps reading forward in the -1 frame from this point.

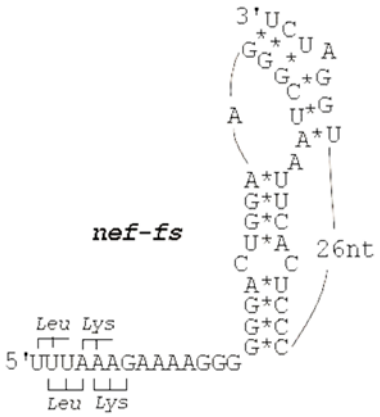


Figure 6. HIV-1 Nef Frameshifting Site in Pseudoknot Form as Predicted by Taylor et al.

The precise mechanism of how pseudoknots stimulates frameshifting is still under debate. When a stem loop is unwound by ribosome, unwinding of the stem forces the loop to rotate. Since stem loop is not restrained, the loop can freely rotate allowing the ribosome to move forward easily. But in the presence of a pseudoknot, the unwinding causes supercoiling and therefore increases the resistance for ribosome to move. This ribosomal pause on slippery sequence stimulates the frameshifting. However, the pseudoknot must be unfolded by the ribosome because as part of the mRNA, its codons must be subsequently decoded by the ribosome. Therefore, the stability and structure are critically important for its function. It has shown that any large RNA hairpin structure can enhance frameshifting to some degree. However, possible roles of non-hairpin structures that affects ribosomal frameshifting has never been explored. In this project we

demonstrate for the first time, the effect of G- quadruplex structure on the -1 frameshifting in HIV-1 nef gene.

The other genetic switching mechanism in this study is, stop codon readthrough. This occurs when the UGA stop codon is translated as an amino acid via codon-anti codon non- Watson Crick base pairing, for example, G-U base pair would permit stop codon UGA to pair with an Arginine tRNA with a UCG anti codon. Another possible way for readthrough is involvement of a suppressor tRNA that exist in cells and also by selenocysteine insertion at the UGA stop codon. [9]. This involves $tRNA_{sec}$ and require specific RNA structures and protein factors.

The focus of this project is the HIV-1 nef gene, which has examples of both of these types of co-translational switching events (translational frameshifting and stop codon readthrough). Nef is a myristoylated protein expressed in the early stage of the HIV-1 life cycle, which functions as a fundamental factor for efficient viral replication and pathogenesis. Since deletion of the nef gene is found to prevent HIV-1 infected primates from progressing to AIDS, nef appears to have a central role in pathogenesis. In order to promote viral infectivity, nef down regulates CD4 and MHC1 expression on the cell surface.[9] However, it may exert other functions not yet fully understood, some of which may be a result of novel protein isoforms that can only be expressed by the unusual genetic mechanisms proposed here.

1.4 Functional Significance of Extended Readthrough Isoforms of Nef

Nef is an early gene of HIV-1, expressed from a spliced mRNA along with the tat and rev gene products. At this early point in cellular infection, nef is expressed at fairly high levels and is even secreted as a soluble form. Even a very low efficiency of SeC incorporation as a rare isoform, e.g. as < 1% of total HIV nef protein production, could significantly deplete Sec in the host cell. Results presented here (Chapter 3) suggest the 3'-UGA codon of nef may be suppressed at fairly high levels, as much as around 20% or more, but much of tat could be conventional protein product, with the UGA being decoded as any one (or all) of amino acids with related codons, like Cys, Trp, Arg or Ser (the amino acid load onto the selenocysteine tRNA before addition of Se to form tRNASec. If the level of nef expression was high enough this could cause an antioxidant defect in cells expressing viral proteins at high levels, which could contribute to the observed correlations between Se status and outcome in HIV infection.

This could also help to explain a number of observations such as the fact that an Australian cohort of "long term non-progressors" have a natural HIV-1 strain with significant deletions in the nef region. Furthermore, Nef gene deletion is the basis of live attenuated vaccines, and transgenic mice bearing an HIV-1 nef-3'-LTR transgene developed severe immunodeficiency and CD4+ cell depletion, "strikingly similar to retrovirus-induced murine AIDS" - all caused by the nef gene alone without HIV (Lindemann et al., J. Exp. Med. 179:797-807, 1994).

In essence, the results presented here support the hypothesis that excessive Nef gene expression with consequent Se depletion may contribute to HIV pathology.

CHAPTER II

READTHROUGH OF THE 3' UGA STOP CODON OF HIV-1 NEF VIA ANTISENSE TETHERING INTERACTIONS BETWEEN HIV-1 AND HOST SELENOPROTEIN MRNA

2.1 Abstract

One of the notable features of nef is the highly conserved 3'-UGA stop codon, and the potential for the protein to be extended by about 30 amino acids if readthrough of that stop codon can occur. We hypothesized that antisense tethering interactions (ATIs) between viral mRNA and host selenoprotein mRNA captures the host selenocysteine insertion sequence (SECIS) element to enable the expression of virally encoded selenoprotein modules via translation of in frame UGA stop codons as selenocysteine (SeC). This mRNA hijack mechanism was predicted theoretically using computational analysis and was experimentally supported at DNA level by gelshift assay. Readthrough of UGA was proved at the mRNA level by Fluorescence microscopy image analysis and flow cytometry of transfected HEK 293 cells with engineered reporter gene plasmid vector constructs which can only be expressed if the UGA is translated. SiRNA knockdown of tethered thioredoxin 1 mRNA in host-virus mRNA tethering interaction resulted in decrease of GFP expression due to the blockage of the components involved in selenocysteine incorporation for the stop codon readthrough, thus proving the proposed antisense tethering interactions.

Furthermore, quantitative analysis of TR 1 mRNA knockdown demonstrated using RT-PCR resulted in an approximate 20% knockdown.

2.2 Introduction

2.2.1 Selenium and HIV

Selenium has shown to be an essential trace mineral which plays important roles in the human health and development which can be incorporated in to selenoproteins as the rare amino acid selenocysteine, the selenium analog of cysteine. This fascinating element was first described as a toxin and a carcinogen until 1957 when Schwarz and Folts found that it was toxic at high levels but was an essential dietary nutrient at low levels. In 1930s it was reported in Nebraska and Dakotas that the horses eating seleniferous plants which grow in the soil with high levels of selenium, suffered from loss of hair in main and tail and also a necrotic hoof malady. [10] But subsequently, an extensive amount of evidence demonstrated that selenium has been linked to many health benefits. Such as decreasing the incidence of cancer, treating muscle disorders, protecting against cardiovascular diseases, control of oxidative stress and inflammatory responses, boosting immune function and most importantly, of all the health benefits attributed to selenium, the one that has received most significance is the correlations of selenium deficiency associated with number of viral infections. Many studies have demonstrated that selenium has chemoprotective effects against several viral diseases. Low selenium status has been associated with Keshan disease myocarditis caused by Coxsackie B3 virus (CVB3), a classical Se- deficiency disease found in rural areas of

China where the selenium level of the soil is very low. [11] People infected with the virus undergo cardiomyopathy leading to death by heart failure. At low levels of selenium in the body, even benign strains of Coxackie B3 virus can become virulent. [11] Other examples for viral diseases associated with low dietary selenium level are, mammary tumors caused by MMTV [10,11], liver cancer which is linked to hepatitis B infection, Asian viral hemorrhagic fever and most importantly, selenium has proven to delay the onset of acquired immunodeficiency syndrome (AIDS) associated with infection by human immunodeficiency virus (HIV) due to its critical antioxidant defense and aspects of particular functions in cellular immunity. Selenium is critical for antioxidant defenses because it is an essential component of many glutathione peroxidases (GPx) and thioredoxin reductases (TR). [17] Glutathione peroxidase is a eukaryotic selenoprotein with the amino acid selenocysteine (SeC) at the enzyme active site encoded by the UGA stop codon in RNA. Most forms of glutathione peroxidase enzymes are essential for fighting harmful lipid peroxidation processes which can be increased under oxidative stress. Lipid peroxidation can cause apoptosis due to the destruction of cell membrane. Therefore, free radicals are not always harmful for the cells. Antioxidant deficiencies can weaken the host immune defenses, thus leading to create a more susceptible environment for viral infection and replication. This could also result in mutant virus strains. Taylor et al have identified GPx related sequences in many RNA viruses, including HIV-1, Coxsackie virus, HIV-2, measles virus and hepatitis C virus. [12] The sequence analysis of these viruses has shown that putative GPx related features are highly conserved within the viral subtypes or genotypes. Moss and co-workers have found that the DNA virus

Molluscum contagiosum encodes a functional selenium dependent GPx enzyme in which the open reading frame (ORF) of encoded sequence has a high homology to mammalian selenoprotein GPx. However, viral selenoprotein encoding would have to occur at a low level in order to not severely deplete cellular Se level and induce free radical mediated cell death, since the virus depends on the host cell for their replication and survival.

2.2.2 Proposed mechanism of Readthrough of 3' UGA stop codon of HIV-1 nef

The genomes of bacteria and eukaryotic organisms are known to encode selenoproteins using the UGA codon, usually a stop codon for selenocysteine (SeC). These organisms use a complex co-translational mechanism to incorporate SeC into polypeptide chain using RNA stem-loop structures. This structure is called a selenocysteine insertion sequence element (SECIS), generally located in the 3' untranslated region (3' UTR) of the selenoprotein mRNA. By recruiting various protein factors, including SECIS binding protein2 (SBP2) and elongation factor Sec (EF), the SECIS element enables delivery of tRNA Sec to the ribosome for Sec incorporation at the UGA codon, preventing it from being a stop codon.

Taylor et al showed that based on an analysis of the genomic structure of the HIV-1, several regions overlapping known HIV genes potentially encode selenoproteins, including one that is a homologue of glutathione peroxidase. This finding confirms the extensive amount of evidence that selenium and other antioxidants such as glutathione (a co-factor of GPx) and/or selenium deficiency has been linked to the incidence, disease progression or virulence associated with HIV-1 infection.

However, a functional SECIS element encoded by an RNA virus has never been demonstrated. The hypothesis presented here explains how this is possible, because a virus would not need its own SECIS if it could hijack one from the host. Since SECIS elements have the ability to function even when carried by a separate mRNA [13], it is probable that the presence of a SECIS in a tethered mRNA could also decode UGA as selenoprotein on the tethered mRNA partner. Here we hypothesize that antisense tethering interactions between viral mRNA and host selenoprotein mRNAs would capture host SECIS to enable the expression of virally encoded selenoprotein modules via translation of in frame UGA stop codons as selenocysteine.

The -1 frameshift site within the nef coding region has several well conserved UGA codons in the overlapping reading frame. In addition, the 3' end of the nef coding sequence terminates in a UGA codon that is highly conserved in global HIV-1 isolates. Computational analysis predicted that there is a potential antisense tethering region located immediately overlapping the conserved in frame UGA codon. In addition, there is a shorter antisense binding region immediately downstream of the conserved UGA codons. The focus of this project is find out the potential readthrough of that UGA codon, potentially making an isoform of nef that is extended by about 30 residues, and incorporating selenocysteine.

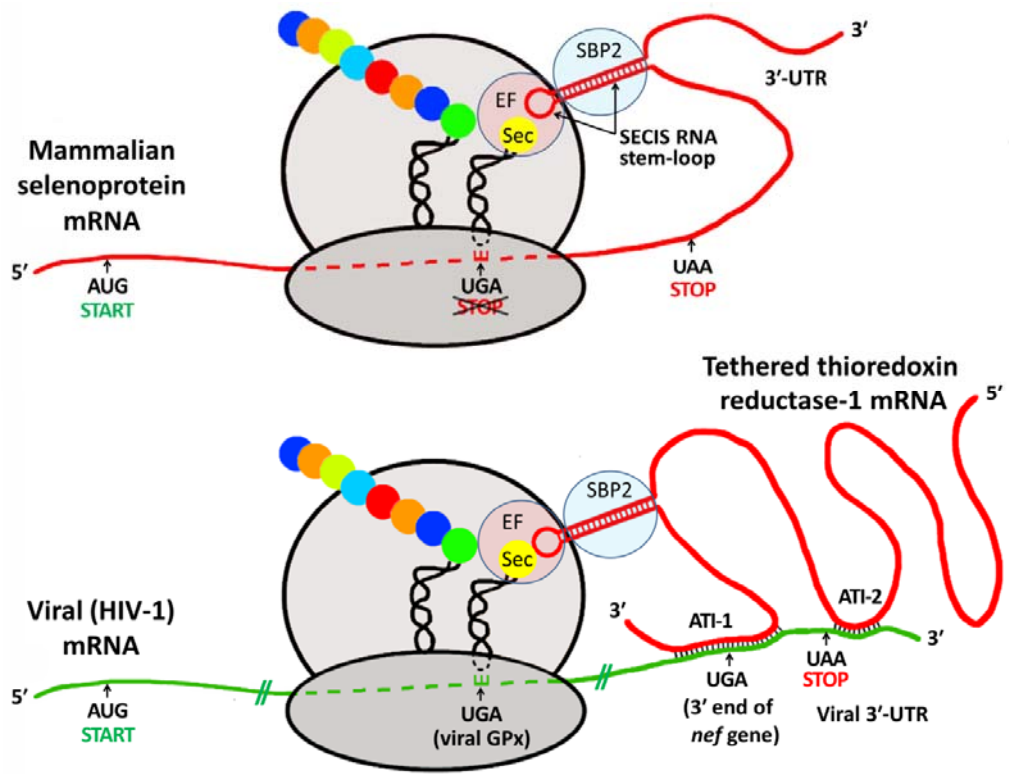


Figure 7. Proposed Mechanism of Sec Incorporation into Viral Proteins via Hijacking of a SECIS Element from a Tethered Host Selenoprotein mRNA (Taylor et al., 2016). Both panels show schematic ribosomes with bound tRNAs. One carrying the Sec, the other a growing peptide chain. The panel above shows the mechanism of insertion of SECIS during mammalian selenoprotein biosynthesis. The panel below shows how the HIV-1 mRNA could hijack the host SECIS element via ATI to decode UGA to synthesize viral selenoprotein such as HIV-1 encoded GPx. ATI-1 is the interaction shown in structure A. This spans the highly conserved 3' UGA codon of the nef gene. ATI-2 is a second shorter antisense region consisting of 13 consecutive base pairs near the end of viral mRNA.

2.3 Materials and Methods

All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). HEK 293T cells were purchased from NIH AIDS reagent program.

2.3.1 Computational analysis of the predicted antisense interactions between regions of mRNAs of human Thioredoxin Reductases and HIV-1 mRNA

To present theoretical evidence for the possibility of antisense tethering interactions (ATIs) between host selenoprotein-mRNAs and mRNA of HIV-1 nef gene. We have used the mRNA of human thioredoxin reductase 1 (TR1) and a region of HIV-1 genomic mRNA, in the end of nef gene at the 3' - long terminal repeat (LTR). These anti sense interactions were discovered via BLAST searches to identify potential helical interactions of 15 or more base pairs. The best matches were then refined by RNA folding methods and the RNA Hybrid 2.2 program (http://bibiserve2.cebitec.uni-bielefeld.de/rnahybrid) to establish the extent and binding energies of RNA: RNA interactions.

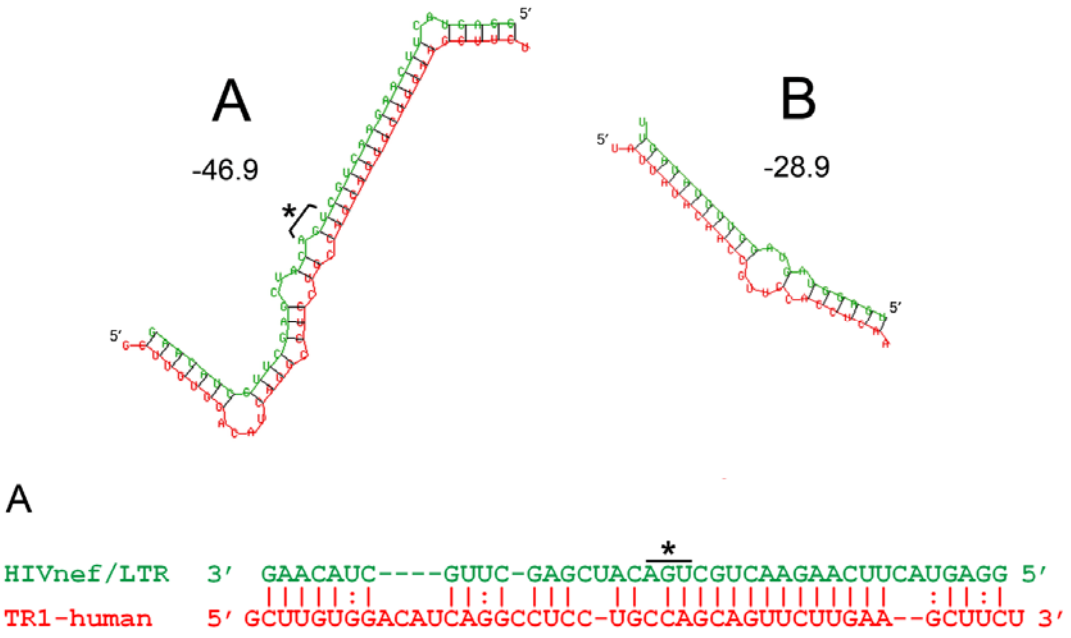


Figure 8. Predicted Antisense Interaction between Regions of mRNA of TR1 and HIV-1 mRNA. The computed free energies of the interactions are shown in kcal/mol. A. HIV-1 nef LTR (green) vs TR1-human (red). The highly conserved 3'-UGA stop codon of nef is indicated by asterisk. B. An imperfect miRNA interaction. (Let-7 vs. a cellular target, the default example in RNAHybrid). This is included to show that the HIV nef ATI (structure A) is both more extensive and thermodynamically stable than a typical imperfect miRNA binding interaction.

2.3.2 Gel Shift Assay

The predicted interactions were supported by a gel shift assay using synthetic DNA oligos to show strong in vitro DNA hybridization between the predicted regions of the HIV-1 mRNA and human Thioredoxin Reductase 1 (TR1) mRNA, shown in Figure 8A. Gel shift assay was done using 40 mer synthetic simple ssDNA oligos. Lanes were loaded with either a single oligo that runs as ssDNA or an incubated pair of oligos. Briefly, oligos (~1 µg each in 10 µl PBS) were incubated at 37^0C for 15 hours, and separation was done on a 4% agarose gel with ethidium bromide visualization.

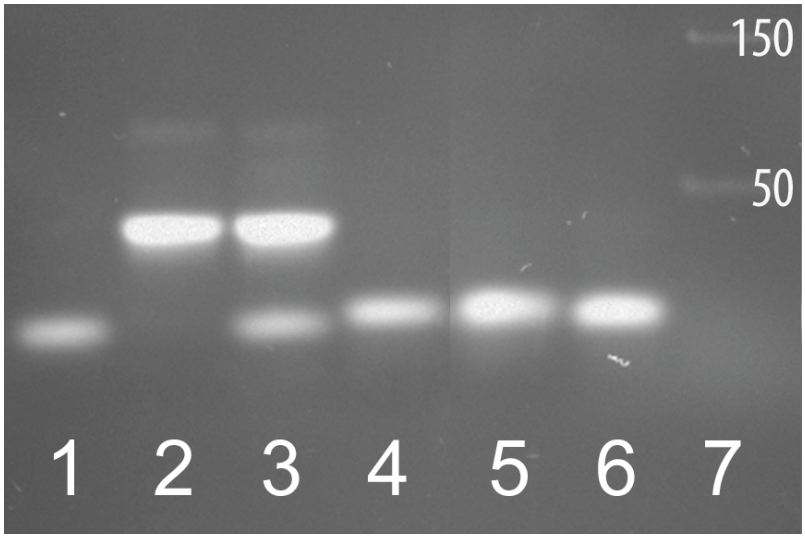


Figure 9. HIV-1 Nef vs. Human Thioredoxin Reductase 1 Antisense Interaction Demonstrated at the DNA level. Target-specific *in vitro* DNA hybridization of the virus-selenoprotein pair is shown by gel shift assay using ~40mer synthetic single stranded ssDNA oligos. The matched pair of oligos from the viral and host mRNAs, corresponding to the sequence regions involved in the antisense interactions shown in Fig. 8A, were the HIV-1 nef/LTR region (Hnf) vs. TR1. Using ~40mer fragments from TR3 and an Ebola virus fragment (Enp), lanes 5 and 6 were included as negative controls to show the specificity of the HIV nef vs TR1 interaction. Lanes have either a single oligo that runs as ssDNA (the lowest bands) or an incubated pair of oligos, as follows: **1.** Hnf; **2.** 1:1 Hnf+TR1; **3.** 2:1 Hnf+TR1; **4.** TR1; **5.** 1:1 Hnf+TR3; **6.** 1:1 Hnf+Enp; **7.** DNA size markers. The bright bands at the size of ~40mer double stranded dsDNA correspond to the expected Hnf+TR1 hybridization at 1:1 (lane 2) and 2:1 molar ratios (lane 3), whereas the HIV nef fragment does not hybridize to either the TR3 or Ebola sequences used as controls (lanes 5 and 6).

2.3.3 Developing plasmid vector constructs to monitor HIV-1 nef Antisense Tethering Interactions

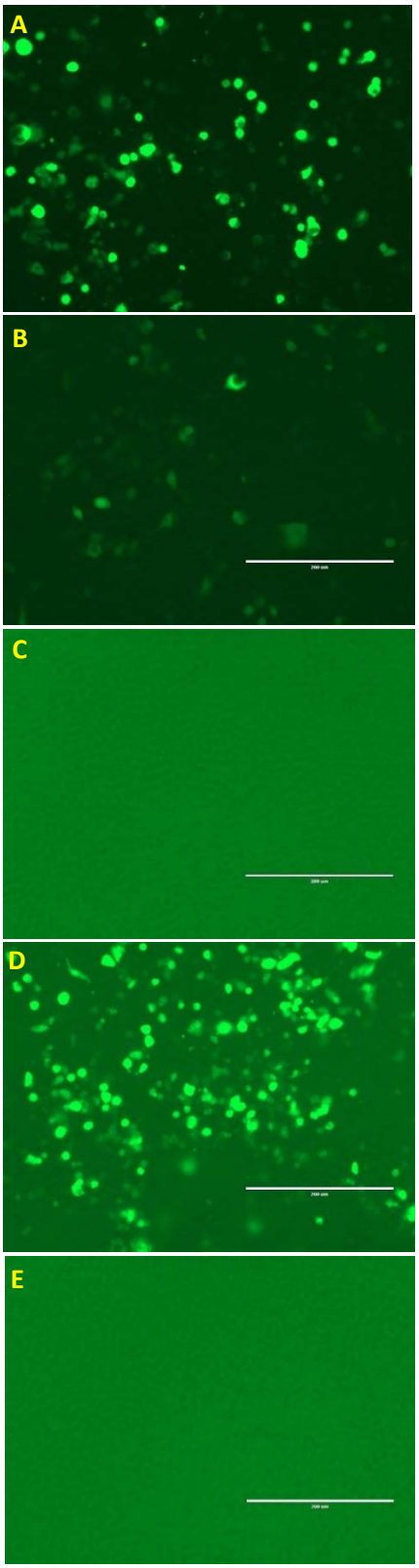
In order to confirm our hypothesis in vivo, 3 different plasmid vector constructs were developed that contain a Green fluorescence protein (GFP) reporter gene downstream of the complete HIV-1 nef coding gene including the 3'- LTR region. pEGFP-N3 parent plasmid was obtained from Clontech laboratories inc. GFP can only be produced only if the UGA stop codon is translated as an amino acid. Three forms of the

construct were made, one containing the entire HIV-1 nef coding region with both Antisense tethering interaction ATI-1 and ATI-2 (pNef-ATI1+2 EGFP-N3). The second construct was only extended far enough to include the HIV-1 nef gene and ATI-1 region, but do not include ATI-2 region (pNef-ATI1 EGFP-N3). A third control vector construct was made including HIV-1 nef, ATI1, ATI2 and an extra stop codon was introduced at the end of 3'LTR region so that GFP will not be expressed (pNef-ATI1+2+stop EGFP-N3). These constructs were generated by PCR amplification of the HIV-1 LAI clone, using 3 different primers. Reverse Primer 1; mutated stop codon at 9107 bp at 3' end of nef (5'-CCA GAG GGA TCC AGT ACA GGC AAA AAG CAG CTG CTT GTA TGC AGC ATC-3') , Reverse primer 2; Leaving in stop codon at 9107 bp and adding an additional stop codon at 3' end of nef (5'-GAG GGA TCC ACT ACA GGC AAA AAG CAG CTG CTT ATA TGC AGC ATC-3') and forward primer for 5' end of nef (5'-CAG GGC TAG CAA AGG ATT TTG CTA TAA CAT GGG TGG CAA G-3') (Integrated DNA technologies, Coralville, IA) which introduced Nhe1 and BamH1 restriction sites for subsequent insertion in the pEGFP-N3 vector. The obtained vector constructs were confirmed by sequencing (Eurofins MWG Operon).

2.3.4 Transfection of HEK cells to determine HIV-1 nef 3' UGA stop codon Readthrough

HEK 293 cells were seeded (20×10^3 cells/well) in a costar 96 well clear bottom black plate and incubated for 24 hours at 37 ^{0}C. Cells were then treated with 20 nM Selenium and incubated for 48 hours at 37 ^{0}C. Next, the cells were transfected with the vector constructs, pNef-ATI1 EGFP-N3, pNef-ATI1+2 EGFP-N3 and pNef-

ATI1+2+Stop EGFP-N3 using lipofectamine 2000 transfection reagent. After 12 hours of incubation at 37^0C, fresh media was added and incubated for 4 days while observing the GFP expression of the cells each day using EVOS cell imaging system. Fluorescence images were taken from each well and each treatment at 10 different locations of the wells. The GFP intensity was counted manually using NIH ImageJ software.



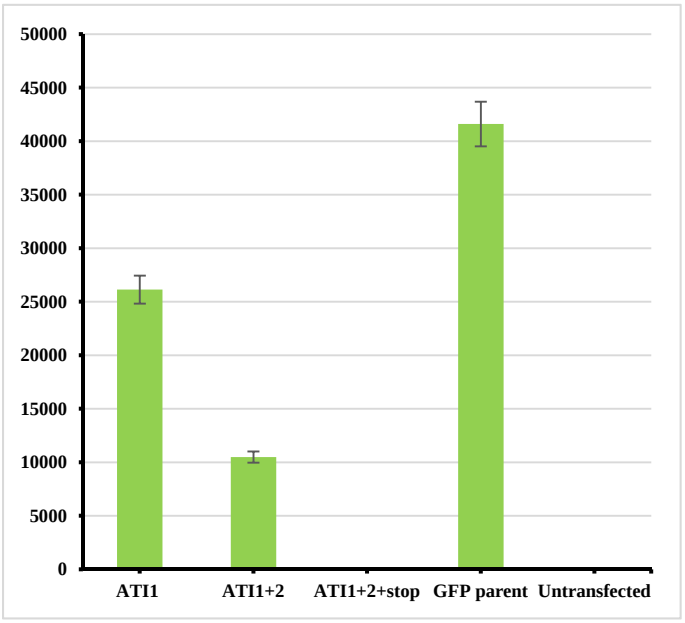


Figure 10. Readthrough of HIV-1 nef 3' UGA Codon(1) Fluorescent microscopic images of HEK 293T cells transfected with 3 ATI plasmid vector constructs and controls. (A)pNef-ATI1 EGFP-N3 (B)pNef ATI1+2 EGFP-N3 (C) pNef-ATI1+2+stop EGFP-N3 (D) EGFP-N3 (E)Untransfected cells.(2) GFP intensity of each transfection condition calculated using NIH ImageJ software.

2.3.5 Selenium dependence of stop codon readthrough

In order to determine the effect of selenium on stop codon read through, HEK 293 cells were pretreated with different concentrations of selenium as sodium selenite (0, 20, 50 and 80 nM), and then were transfected with plasmid vector constructs pNefATI-1 EGFP-N3 and pNefATI-1+2 EGFP-N3 with untransfected cells as a control. The GFP expression was observed under EVOS fluorescent microscope and was measured using Biotek synergy plate reader on the 4th day post transfection.

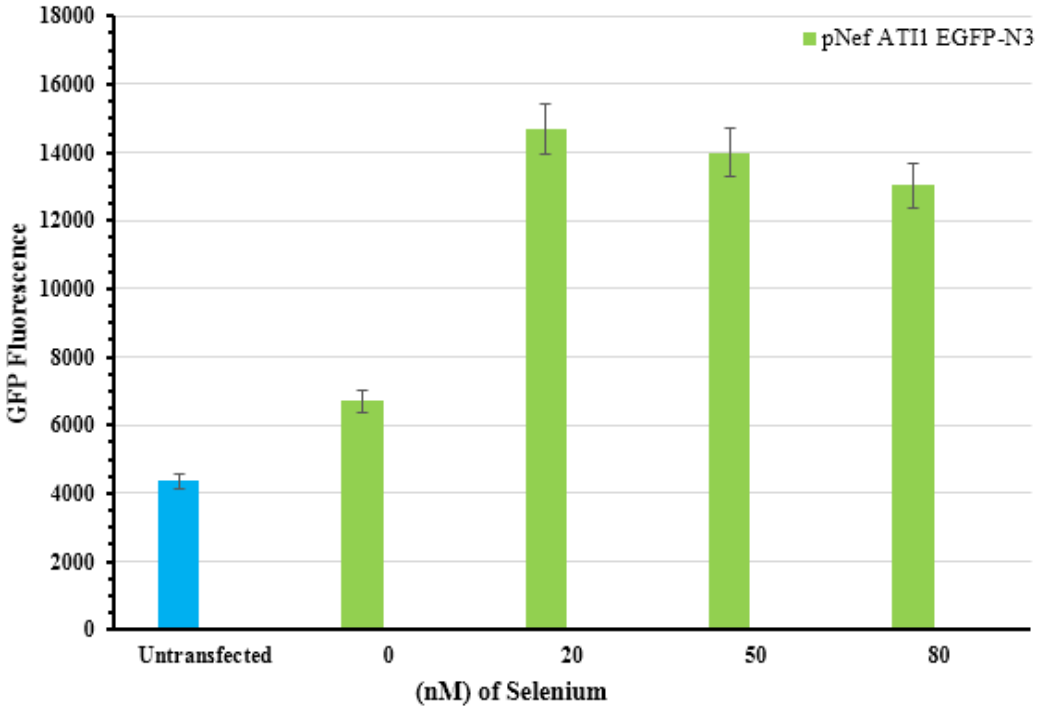


Figure 11. Added Selenium Enhances Stop Codon Readthrough of ATI-1 Plasmid Construct.

The results show that selenium has a significant effect on increasing the stop codon readthrough, resulting higher level of GFP expression. However, the maximal readthrough was observed at a lower level of selenium (20nM) and the readthrough did not change significantly at higher selenium concentrations. Even at 20 nM concentration, addition of sodium selenite essentially doubles UGA stop codon readthrough as measured by GFP production.

2.3.6 Flow Cytometry Analysis

For flow cytometry analysis, HEK 293T cells were seeded (0.5×10^6 cells/ well) in a 6 well plate in 1 ml of Selenium enriched (20 nM) DMEM with 10% FBS medium and incubated overnight. Next, the cells were transfected with the pNefATI1 EGFP-N3 construct and EGFP-N3 parent plasmid using lipofectamine 2000 transfection reagent. After 12 hours, wells were replaced with fresh selenium enriched media and incubated for 4 days. Cells were then trypsinized, washed with 1% PBS and resuspended in 400 µl of 0.1% BSA in PBS. GFP expression was evaluated using FACS DIVA version 6.1.3 software.

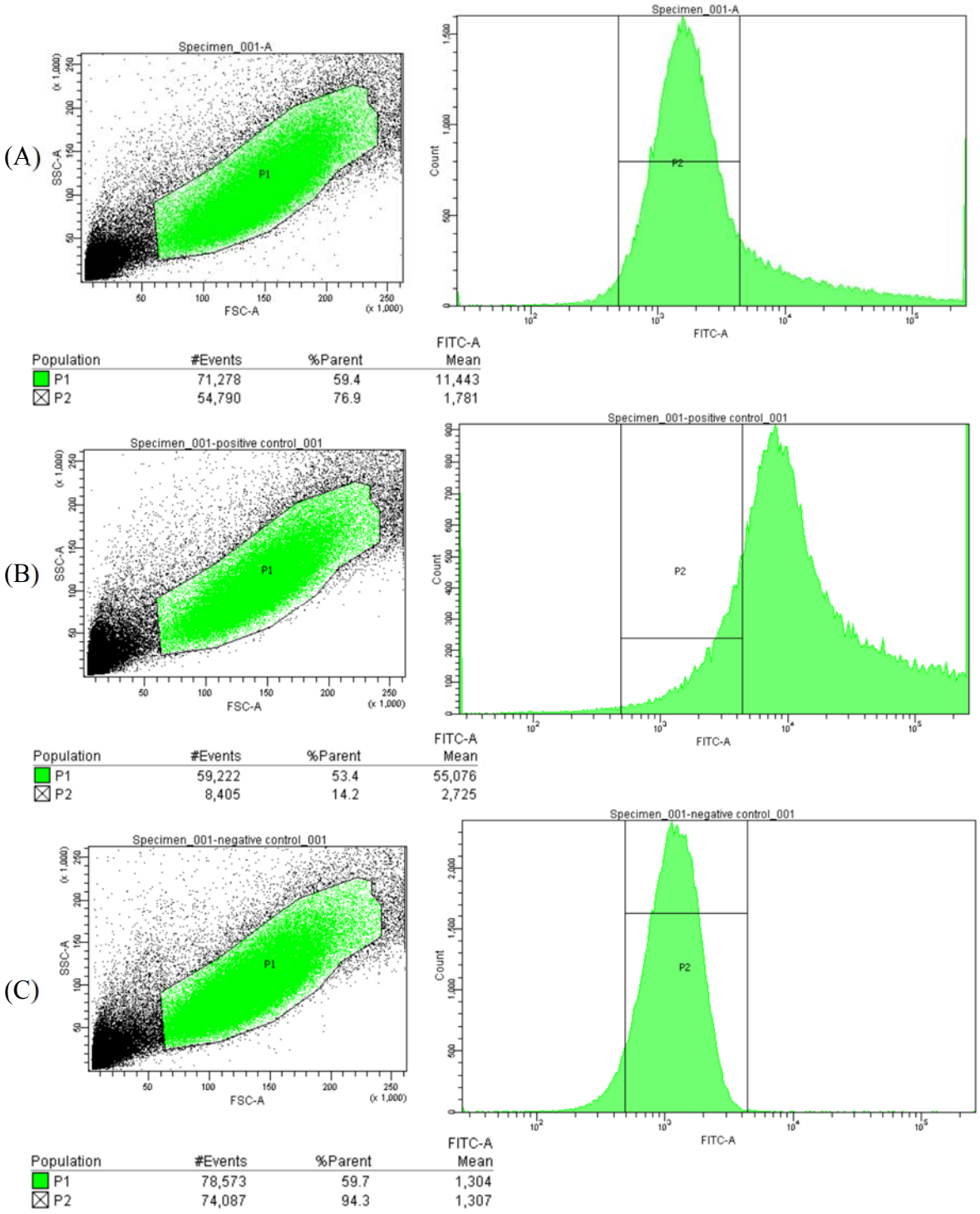


Figure 12. Flow Cytometry Analysis of HIV-1 Nef Stop Codon Readthrough. (A)HEK 293T Cells Transfected with pNef-ATI1 EGFP –N3 vector. P1 population had a mean FITC-A of 11443. (B) Cells transfected with EGFP-N3 plasmid. P1 population had a mean FITC-A of 55076. (C) Untransfected cells. The mean FITC-A of P1 was 1304.

2.3.7 siRNA knockdown of Thioredoxin Reductase 1 (TR1) mRNA

72-hour incubation at 37^0C. For siRNA knockdown of TR1 mRNA, Ambion silencer siRNA transfection II kit was used (Life Technologies Corporation). Selenium treated HEK 293T cells were transfected (50×10^3 cells/ well) with pNef ATI1 EGFP-N3 vector using different volumes of two transfection reagents, siPort Amine and siPort neoFX to select the best transfection reagent and the volume. (0.15 µl, 0.3 µl, 0.6 µl, of siPort Amine and 0.15 µl, 0.5 µl, 1.2 µl of siPort NeoFX). After selecting the best transfection reagent and optimum volume of transfection agent the siRNA knockdown assay was carried out using the silencer siRNA kit. First the HEK 293T cells were trypsinized and resuspended (9×10^6 cells/ml) in selenium treated DMEM/ 10% FBS medium. siPort Amine (0.6 µl/well) was diluted in Opti-MEM 1 medium and incubated at room temperature for 10 minutes. pNEf-ATI1 EGFP-N3 vector was mixed with the transfection reagent at a 0.2 ng/well concentration. Thioredoxin reductase 1 siRNA, GAPDH siRNA, negative control scrambled siRNA were diluted in Opti-MEM 1 medium for a final concentration of 30 nM in the transfection (0.75 µl/well siRNA in 25 µl/well Opti-MEM). The diluted RNA and diluted DNA+transfection reagent complex were mixed and incubated at room temperature for 10 minutes. RNA / DNA + transfection reagent complexes were distributed in to the empty wells of a 24 well culture plate. The HEK 293T cells were transferred to the wells at a 0.5×10^6 cells/ well density in 500 µl of media/well. The plate was incubated at 37^0C for 72 hours. The GFP expression was imaged using EVOS image system and the GFP intensity of 10 images taken from each treatment was calculated using NIH ImageJ software. Another 24 well plate was

transfected with the same conditions and used for qPCR analysis to quantify the siRNA knockdown of TR1 after.

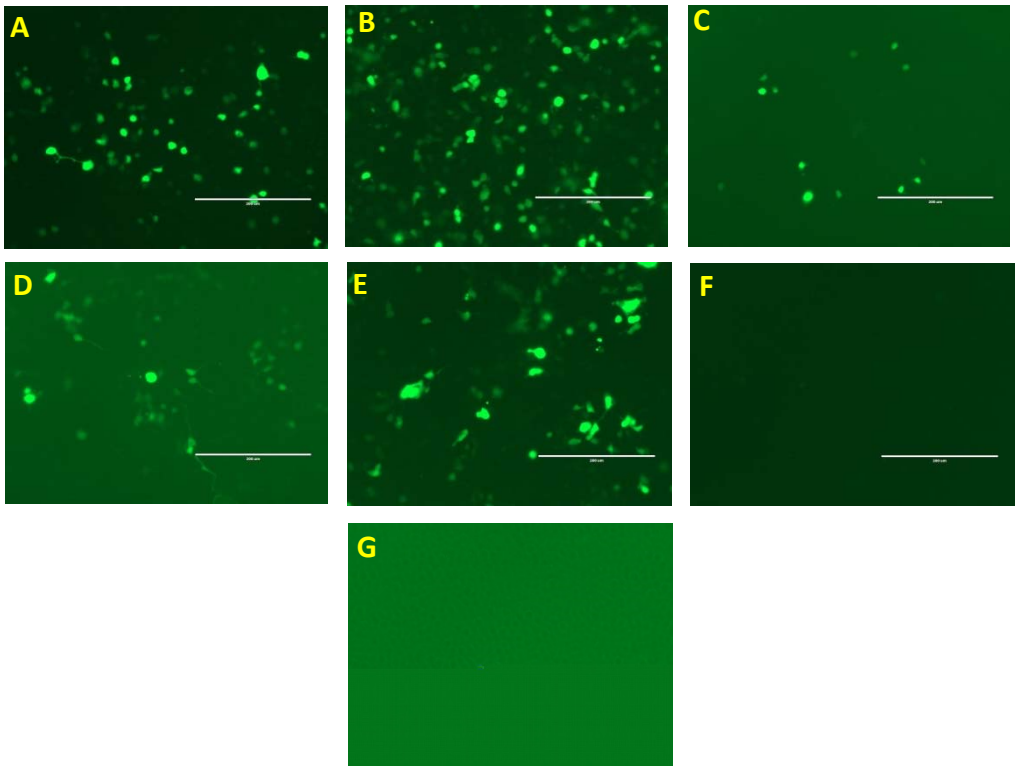
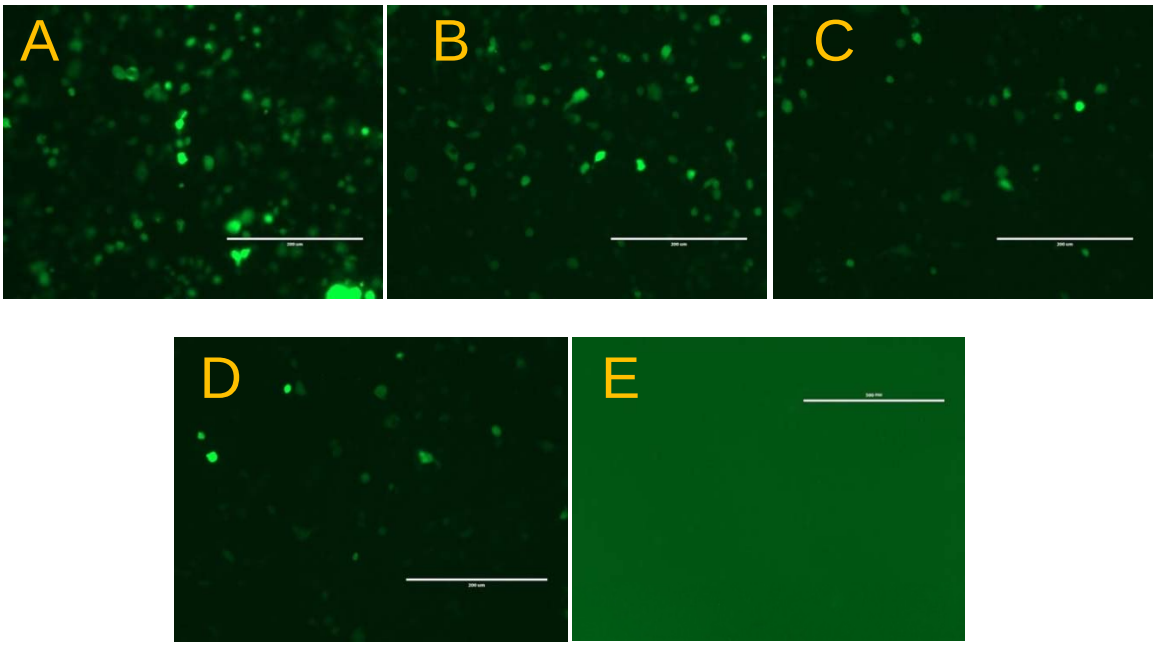


Figure 13. Selection of Transfection Reagent and Optimization of the Transfection Reagent Volume using HEK 293T Cells Transfected with pNef ATI1 EGFP-N3 Vector. EVOS GFP fluorescence Images A,B,C,D,E and F are Cells transfected with 0.3 ul Amine, 0.6 ul amine, 0.15 ul amine, 0.5 ul NeoFX, 1.2 ul NeoFX and 0.15 ul NeoFx respectively. Image F represents untransfected cells.



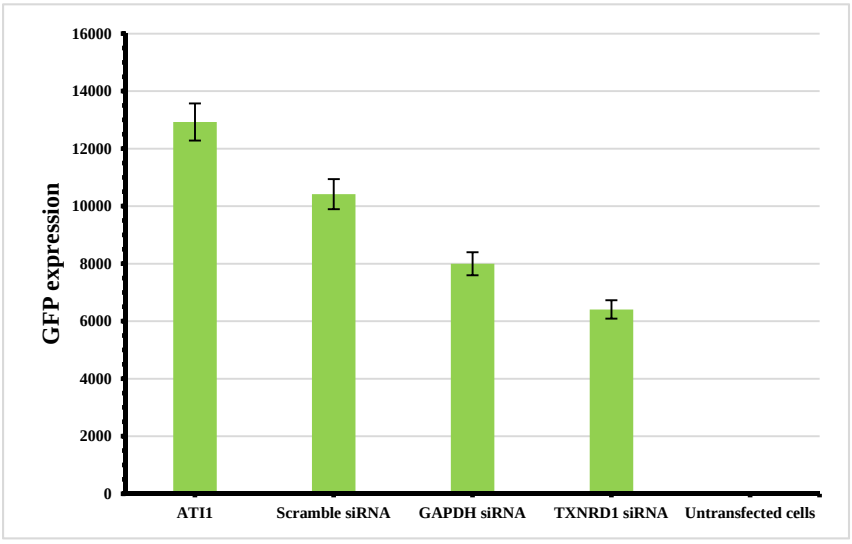


Figure 14. siRNA Knockdown of TR1 mRNA. (1) Evos microscopy images of each transfection condition, A. pNef ATI1 EGFP-N3 with no siRNA. B. Negative control scrambled siRNA. C. Positive control GAPDH siRNA. D. Thioredoxin reductase 1 siRNA knockdown. E. Untransfected cells. (2) GFP expression of 10 images from each treatment calculated using NIH ImageJ software.

2.3.8 Quantitating the TR1 mRNA knockdown using RT-PCR

Total RNA was isolated from each treatment of the siRNA knockdown experiment using Promega total RNA isolation kit. cDNA was synthesized from samples using 300ng RNA as template. A set of reactions were set up without reverse transcriptase for all RNA samples and one reaction without template as a control. Invitrogen superscript III first-strand synthesis mix kit was used for cDNA synthesis. Power SYBR Green PCR master mix (Applied biosystems) was used for qPCR. qPCR was performed on ABI 7500 fast equipment (software version 2.3). The run conditions were 95^0 C for 15 seconds, 58^0C for 15 seconds and 60^0C for 60 seconds. Expression of TXNRD1 mRNA in each sample was calculated relative to TXNRD1 expression in untreated cells. Values for TXNRD1 expression were normalized against expression values for GAPDH housekeeping gene. PfaffI equation was used for calculation of relative expression.

$$\text{ratio} = \frac{(E_{\text{target}})^{\Delta \text{CP}_{\text{target}}(control - sample)}}{(E_{\text{ref}})^{\Delta \text{CP}_{\text{ref}}(control - sample)}}$$

Sample	Relative Expression	SD
Untreated cells	1	0
TXNRD1 siRNA	0.7305093	0.027786
Negative control (scramble siRNA)	0.9135049	0.028464
pNef ATI1 EGFP-N3	0.9188951	0.015494

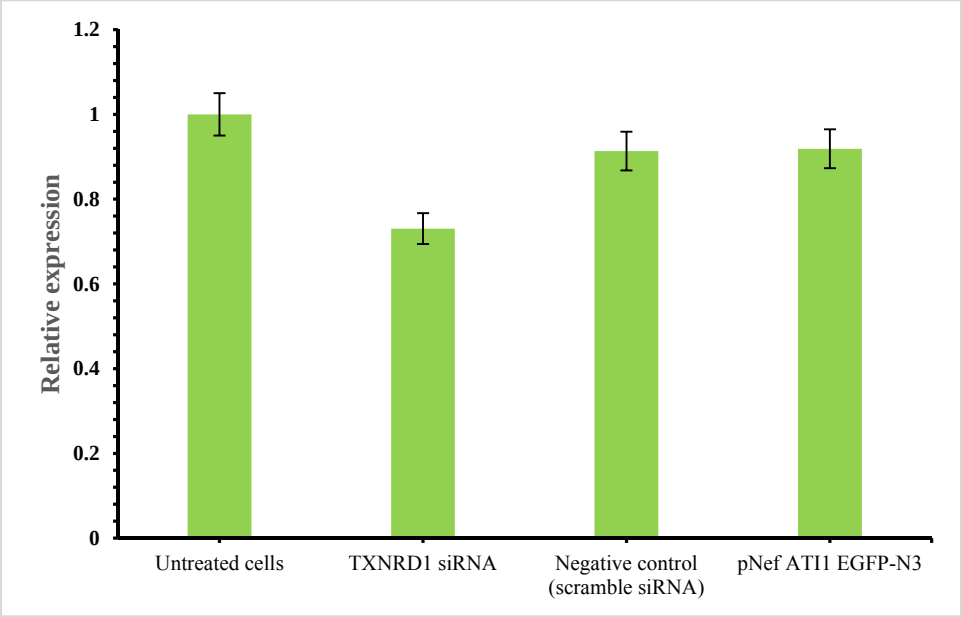


Figure 15. TXNRD1 Expression after qPCR. Compared to untreated cells, an approximate 27% TRXNRD1 knockdown was observed in the sample treated with TXNRD1 siRNA. This confirms the knockdown of TR1 mRNA observed in section 2.3.7.

CHAPTER III

CO-TRANSLATIONAL GENETIC SWITCHING DURING PROTEIN SYNTHESIS IN THE HIV-1 NEF CODING REGION, VIA ALTERNATE RNA STRUCTURES OF THE NEF FRAMESHIFT SITE: PSEUDOKNOT VS G-QUADRUPLEX

3.1 Abstract

HIV-1 nef coding region features a potential -1 frameshift site with a potential overlapping gene region near the middle of the coding sequence. A known -1 frameshifting signal is present in the nef sequence at this point, immediately upstream of a G-quadruplex (QPX) sequence that serves to regulate frameshifting. An in vitro frameshift assay using a dual reporter vector construct in which the putative HIV-1 nef-frameshifting sequence with QPX cloned between two fluorescent reporter genes showed that the -1 frameshifting occurs. Treating the transfected cells with QPX stabilizing synthetic drug TMPYP4 increased the frameshifting efficiency by 27% further demonstrating the occurrence of -1 frameshifting.

3.2 Introduction

G-quadruplexes (QPX) are formed by the stacking of two or more G-quartets (also called G-tetrads), which are tetraplex structures of nucleic acids that form in G-rich DNA and RNA. The GQPX can be either inter or intra molecular four-stranded

structures, with each strand contributing the guanine bases that form the G-quartets, which then stack on top of each other. QPXs are stabilized by Hoogsteen hydrogen bonding between the guanine residues. The structures are further stabilized by electrostatic interactions with cations such as potassium ions, which are located between the tetrads.

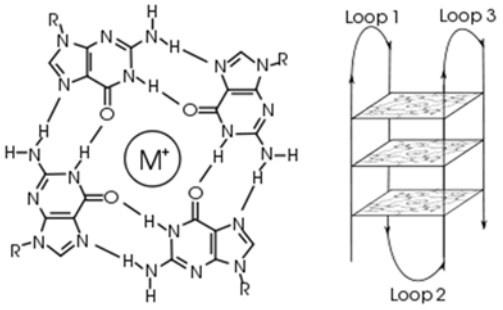


Figure 16. Left: A G-tetrad. Right: An Intramolecular QPX Formed from 3 Stacked G-tetrads.

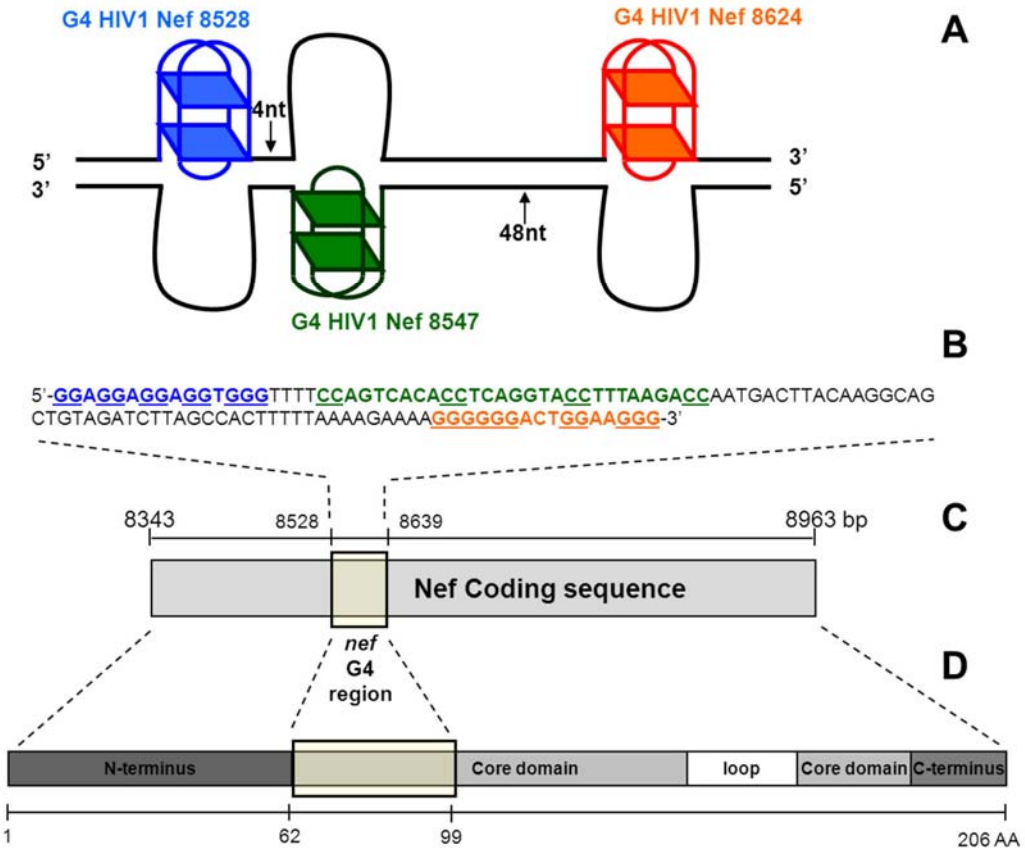
The structures can be described as parallel or antiparallel, depending on the direction of the strands that form the tetrads [14]. If the 5'-3' orientation of all the strands is the same, the quadruplex is termed parallel. If one or more of the runs of guanine bases has a 5'- 3' direction opposite to the other runs of guanine bases, the structure is called antiparallel. Adopted structures are also dependent on factors such as strand concentration, the cations present and the loop size and connectivity.

QPX have been the focus of intensive interest in current nucleic acid research, due to their involvement in number of biological functions. The formation of QPx was initially shown in vivo and in vitro in telomeric sequences in variety of organisms. The

formation of QPX in telomeres inhibits the activity of telomerase, which is responsible for maintaining the length of the telomeres. Stabilization of QPX in tumor cells has been proposed as an effective way to inhibit telomerase activity. QPXs are also found as a target of binding proteins in the coding regions of the human genome including minisatellites, rDNA and immunoglobulin heavy chain switch regions. In DNA, QPX structures are only able to form when at least a local region of the DNA has been separated into single stranded form. Therefore, most likely they form during the DNA replication and transcription. Several QPX stabilizing compounds have been identified up to date. These include perylenes such as PIPER, porphyrins such as TMPyP4, trisubstituted acridines such as BRACO-19. Dr Lijun Zhao, a collaborator of the Taylor lab, has recently shown that resveratrol, an anti-aging compound, can stabilize a QPX structure, which is similar to the previously demonstrated binding of the alkaloid berberine to telomeric QPX, for which an X-ray crystal structure is available. Also, monovalent cations, in particular K+ are known to stabilize QPX. Circular Dichroism has shown that compared to Na+, K+ has a higher effect on stabilizing QPX structure, whereas Na+ may have a destabilizing effect. The roles of QPX structures as regulatory systems have been widely studied in eukaryotic cells and in some prokaryotic systems. But their regulatory roles in viruses and other organisms has not been extensively investigated to date.

Richter et al. demonstrated the presence of a unique cluster of QPX structures in the nef protein coding region of the HIV-1 genome, coding for the nef protein [14]. They demonstrated the presence of 3 putative QPX-forming sequences at Nef 8528, Nef 8624

and Nef 8547. The Nef protein is a fundamental factor for efficient viral replication and pathogenesis in vivo. Therefore, inhibition of the expression of isoforms of nef could lead to novel antiviral therapies.



Ritcher et al. plos one. 2013 vol 8 (open access).

Figure 17. Presence of 3 Putative G-quadruplexes in HIV-1 Nef Gene at Nef 8528, 8624 and 8547. A. Scheme of QPX formation within the double stranded DNA of the nef region. B. Neucleotide sequence of the nef coding region where the QPXs are located. C. Scheme of the amino acidic sequence of the nef protein indicating reported structural domains.

In our study we hope to show for the first time that a QPX can modulate frameshifting, using the predicted site in the HIV-1 nef region as a paradigm. If confirmed, this co-translational genetic switching between the translations of alternate

protein isoforms will be a new biological role of QPX. We are focusing on the G-rich sequence at the nef 8624. This particular G-rich sequence is starting at the beginning of the region that forms a possible RNA pseudoknot in the nef-coding mRNA which is located five bases downstream of a heptameric -1 frameshift sequence, UUUAAAA, which is located about one third of the way into the nef coding region, as measured from the N-terminal Met residue. Pseudoknots are structural elements which forms when bases within the single stranded loop in a secondary structure base pair with bases outside (upstream or downstream) of that stem-loop, and which are known to play a role in a genetic process called ribosomal frameshifting.

Because the genetic code uses three letter words for amino acids and start or stop signals, for any gene read from 5' to 3', a genome has three different reading frames: a zero reading frame, -1 reading frame and +1 reading frame. Overlapping genes can occur when a single oligonucleotide codes for more than one protein, by being translated in multiple reading frames. These are a common feature in many RNA viruses. Viruses have evolved to adopt this feature as a form of genome compression, allowing the virus to maximize the coding potential in order to synthesize and regulate various proteins without increasing the genome length. The proteins coded by overlapped genes can only be produced in relatively small amounts, and thus are more likely to have regulatory roles, thereby potentially influencing viral pathogenicity.

Ribosomal frameshifting is a way of slipping out of one overlapping gene region into another gene region. HIV-1 encodes several overlapping gene regions and utilizes a -

1 frameshifting mechanism at multiple sites to regulate gene expression and optimize its replication relative to the host. There are three physical factors required to induce -1 frameshifting on the mRNA:

1) A heptameric slippery sequence. This is where the frameshift occurs and for optimal efficiency has the sequence X XXY YYZ (which is considered an "ideal" -1 slippery sequence). X, Y, Z denote nucleotide species and triplet groupings indicate the initial (zero) reading frame.

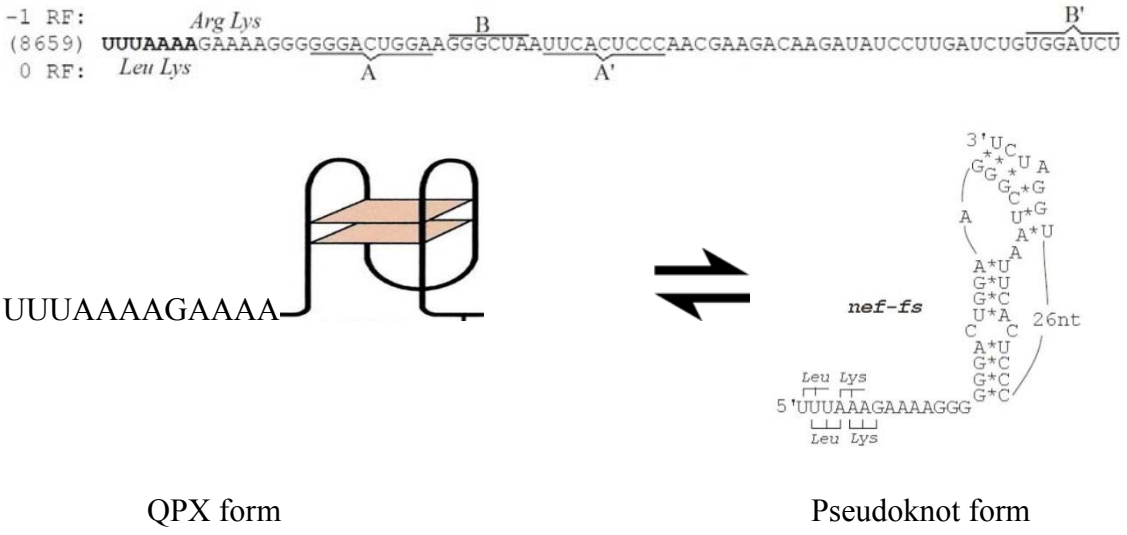
2) A spacer sequence which is usually 6-12 nucleotides long.

3) A downstream pseudoknot structure, or other large RNA structure (hairpin or stem-loop).

The precise mechanism how the pseudoknots stimulate frameshifting is still a subject of debate. It has been shown that when a ribosome encounters the pseudoknot, the ribosome pauses over the slippery sequence, thereby providing enough time for ribosome to remain stalled over the slippery sequence, enhancing the potential for the frameshift to occur. Therefore, ribosomal pausing enhanced by a downstream pseudoknot is necessary for -1 frameshifting. However, the pseudoknot must be unfolded by the ribosome since, as part of the mRNA, its codons must be subsequently decoded by the ribosome. Therefore, the stability and the structure of the pseudoknot are important factors for the efficient function. Furthermore, the geometry and surface charge of the structure may also affect the efficiency of frameshifting.

The putative nef frameshift site that we are focusing on in our study contains an ideal heptameric -1 frameshift sequence, UUUAAAA followed by a downstream RNA pseudoknot structure which is located at a near optimal distance of 7 bases. This is a highly conserved region in HIV-1 because it spans the polypurine tract which begins at the run of a bases in the shift sequence. However, about 8% of the nef isolates in the database show mutation from UUUAAAA [15]. It is mutated to U UUA AGA, which is still be able to function as a -1 FS site because of the role of arginine (codon AGA), which can function as a "hungry codon", facilitating P-site slippage on UUUA. Under arginine deficiency, a ribosomal pause is created similar to that believed to underlie the role of RNA secondary structures like pseudoknots in enhancing frameshift efficiency.

The key observation underlying this part of the project is that the "polypurine tract" which overlaps the potential pseudoknot is also the exact location of one of the QPX identified by Richter et al. in the HIV nef coding region. As shown in the figure 2, it is predicted that the putative QPX and the pseudoknot structure are alternative conformations of the RNA strand that is present in the nef-fs site. Considering the fact that this particular QPX was reported to have the highest innate stability of the 3 QPX identified in the nef region, we hypothesize that it has a potential role in the enhancement of the -1 frameshifting, by presenting an obstruction to the ribosome, creating a ribosomal pause.



QPX form Pseudoknot form

Figure 18. A Region of the HIV Nef Gene mRNA Sequence (top) that can Assume a Possible Pseudoknot or QPX Form (bottom). A: A' and B: B' correspond to the lower and upper stem regions of the pseudoknot secondary structure shown at bottom right. The slippery sequence UUUAAAA is about 7 nucleotides before the beginning of the pseudoknot, and about 5 bases before the beginning of the quadruplex structure. These are alternate molecular conformations of the viral RNA.

The aim of this project is to stabilize the nef QPX using QPX binding synthetic drugs to determine whether QPX formation is a regulator of -1 frameshifting.

3.3 Materials and Methods

Our first step was to construct an expression vector in which the putative HIV-1 nef-fs sequence with QPX region is cloned between two fluorescent reporter genes, to permit the assessment of frameshift efficiency in vitro. The construct was designed to have an upstream Cerulean fluorescent protein reporter gene (CFP) and downstream mOrange Red fluorescent protein reporter gene (RFP). In the nef-fs wild type construct, RFP is in the -1 reading frame to CFP. So RFP is expressed only if the frameshift occurs.

The fluorescence was measured observed under EVOS fluorescence microscope and the intensity was measured using NIH imageJ software.

If our hypothesis is correct, the efficiency of frameshifting is expected to INCREASE in the presence of QPX stabilizing drug TMPyP4.

3.3.1 Plasmid construct design

We used the mCerulean C1 mammalian expression vector (4731bp) which contains the ECFP blue fluorescent protein gene (722bp) upstream of a multi cloning site (obtained from AddGene.org). Since nef has a Bgl11 (A'GATCT) restriction site upstream of frameshift site, we used ECFP as the upstream reporter gene in the construct, by ligating the nef inserts via the Bgl11 site to mCerulean C1, which is digested with Bgl11 restriction enzyme at the multi cloning site. pmOrange N1 mammalian expression vector (obtained from Clontech lab. Inc.) was used as a source for the red fluorescent protein gene, which is located between a BamH1 (G'GATCC) site and a downstream multi cloning site. First ECFP gene was ligated with the nef insert via Bgl11 site and followed with the second ligation of the nef insert + ECFP product into the mOrange vector, using the BamH1 site (nef/morange ligation) and Nhe1 site (mOrange/Cerulean ligation), in order to make the closed complete construct.

3.3.2 Designing of nef insert for plasmid construct

Nef- fs Wild type: ctgtagatcttagccacttttaaaagaaaaggggggactggaagggctaattcactcccaaag
Bgl11

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

ctgtagatcttagccacttttaaaagaaaaggggggactggaagggctaattcggatcccaaag
Bgl11 -1 fs QPX region BamH1

The nef-fs insert was designed and ordered from IDT Inc. Insert was mutated to create a BamH1 site at 3' in order to ligate with pmOrange. Nef sequence originally has a Bgl11 restriction site upstream of QPX region.

3.3.3 In vitro frameshifting assay

The construct was transformed into JM109 cells and grew on LB plates contains Kanamycin. The clones were harvested and plasmids were purified using a plasmid purification kit. And finally the purified plasmids were sent for sequencing to confirm the formation of construct.

HEK 293 mammalian cells were transfected with the plasmid construct observed the expression levels of fluorescence proteins using EVOS fluorescence microscope and intensity was measured using NIH imageJ software. Cells transfected with pmOrange parent plasmid and mCerulean parent plasmid were used as controls.

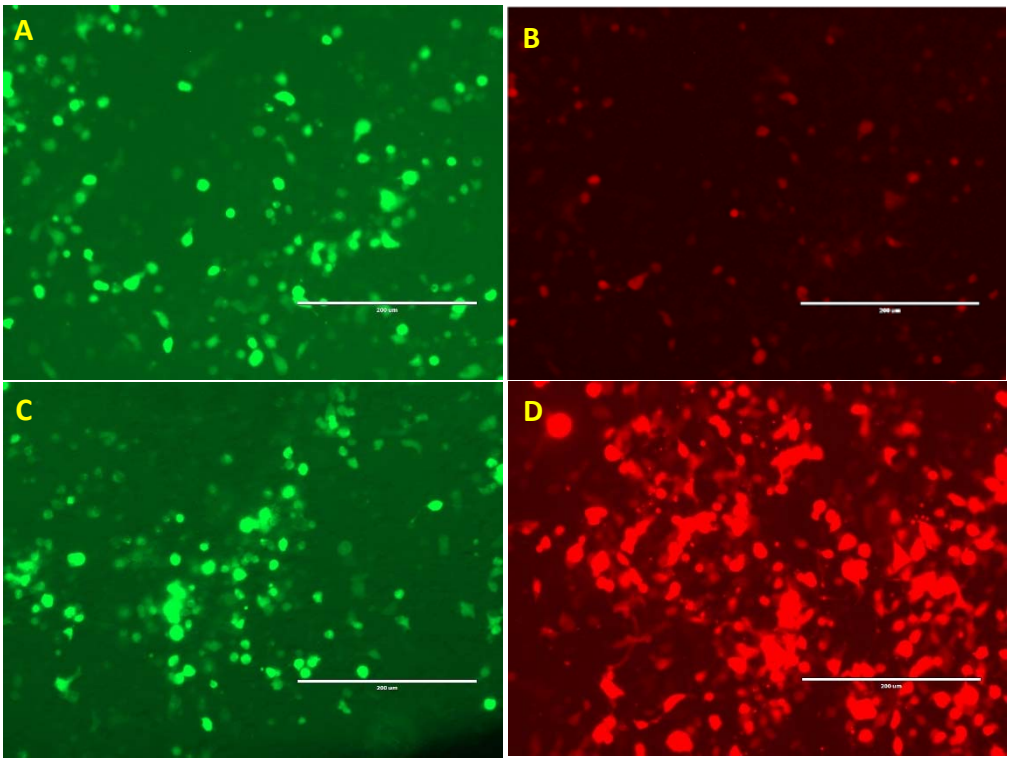


Figure 19. In Vitro Frameshift Assay. A. Upstream cerulean **B**. downstream mOrange. **C**. parent mCerulean. **D**. Parent mOrange plasmid.According to the Evos microscopy images, there is a higher expression of upstream cerulean and a relatively lower expression of downstream mOrange reporter gene. This confirms the -1 frameshifting at HIV-1 nef region. The mOrange parent plasmid expresses a relatively higher RFP (orange fluorescence) than cells transfected with Nef-fs construct.

3.3.4 Effect on frameshifting of stabilization/Destabilization of QPX using QPX binding small drug TMPYP4

Transfected cells were treated with known QPX stabilizing synthetic drug TMPYP4 (10uM) and incubated overnight. Transfected cells which have not treated with any drug were used as a control. Cells transfected with mOrange parent plasmid and mCerulean parent plasmid were used as positive controls. Evaluation of the cytotoxicity of the TMPYP4 in HEK 293 was assessed using MTT assay. Red and blue fluorescence in cells in each transfection was imaged using EVOS fluorescence microscope and the fluorescence intensity was measured using NIH ImageJ software to analyze the effect of TNPYP$ on QPX stabilization thus enhancing the frameshifting.

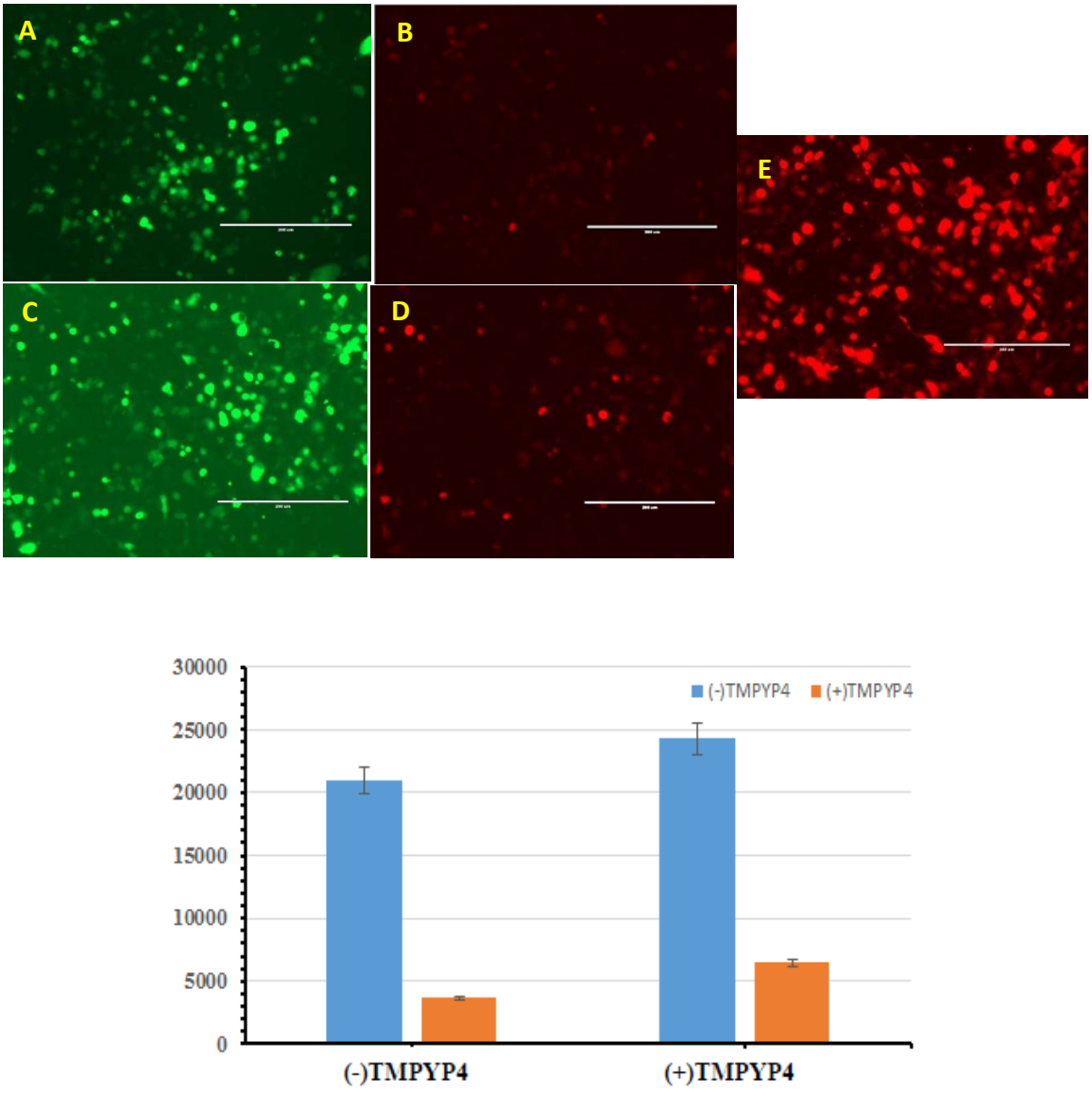


Figure 20. The Effect of QPX Stabilizing Drug TMPYP4 on Nef Frameshifting. A and B: Negative control, HEK 293 cells transfected with Nef-fs construct. C and D: Cells transfected with Nef-fs construct followed by the treatment of TMPYP4 (10uM). E: mOrange parent plasmid transfection. Results confirm the stabilization of QPX region of Nef-fs site by increasing the expression of downstream mOrange reporter gene compared to the control cells transfected without subsequent treatment with known TMYP4 QPX stabilizing synthetic drug.

3.4 Future Research

In order to accurately assess the frameshifting efficiency, appropriate control constructs are needed. This will be achieved by comparing nef-fs wild type test construct to nef-fs mutant positive control as 100% frameshift and a blocked mutant control as 0% frameshift efficiency. All the three constructs express upstream CFP. In the nef-fs wild type construct, RFP is in the -1 reading frame to CFP. So RFP is expressed only if the frameshift occurs. In nef-fs 100% read-through positive control, both RFP and CFP are in the same reading frame thus defined as 100% frameshift efficiency. The third insert is mutated to have two stop codons within the nef-fs preventing the read-through in any reading frame, which therefore should give a background (BG) RFP reading that would be indicative of 0% frameshift efficiency. After subtracting ($Red_{BG}/Blue_{BG}$) from each ratio for background at 0% readthrough, then wild type frameshift efficiency is calculated by:

$$(Red_{WT}/Blue_{WT})/(Red_{control}/Blue_{control})$$

Nef-fs 100% read-through. An additional 'A' is added (purple) to mimic the effect of 100% frameshifting, because both proteins are read from in the same ("zero") reading frame:

ctgtagatcttagccactttttaaaaAgaaaaggggggactggaagggctaattcggatcccaaag

Nef-fs 0% read-through**.** The insert is mutated to create two stop codons within the slippery sequence and in the spacer sequence, so that no downstream protein can be produced:

ctgtagatcttagccacttttaaTaagaTaaaggggggactggaagggctaattcggatcccaaag

In order to determine the effect of K+ / Na+ ratio to control the QPX as a genetic switch, transfected HEK 293 cells grown in media containing different ratios of K+/Na+ and will measure the Red and Blue fluorescence using a plate reader, using K+ and Na+ concentrations that are unable to affect cell proliferation, as measured via the MTT assay. Circular dichroism (CD) will be used to determine the possible binding of TMPYP4 at different ratios of Na+/K+ with and without the drug to demonstrate the function of QPX as a genetic switch.

REFERENCES

[1] Hutchinson, J. F. (2001). "THE BIOLOGY AND EVOLUTION OF HIV." Annual Revue Anthropology

[2] Weiss RA, D. A., Loveday C, Pillay D (2004). "Human Immunodeficiency Viruses." Principles and Practice of Clinical Virology: 721-757.

[3] Montagnier, L. (1999). "Human Immunodeficiency Viruses (Retroviridae) " Encyclopedia of Virology 2: 763-774.

[4] Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody

[5] Lu, K. H., X; Summers, MF (2011). ""Structural determinants and mechanism of HIV-1 genome packaging." Journal of Molecular Biology 410(4): 609-633.

[6] Castelli, J. C. a. L., Jay A. (2002). HIV (Human Immunodeficiency Virus). Encyclopedia of Cancer. 2: 407-415.

[7] j, K. (2001). "HIV." Biochemistry, molecular biology and drug discovery 2.

[8] Taylor, E. W., Nadimpali, RG, Ramanathan CS (1997). Biological trace element research.

[9] Taylor, E. Y., W "A comprehensive analysis and compendium of known potential -1 frameshift sites In HIV-1."

[10] Hatfield DL, T. P., Carlson BA, Gladyshev VN (2014). "Selenium and selenocysteine: roles in cancer, health, and development." Trends Biochem Sci(39): 112-120.

[11] Zhang W, R. C., Nadimpalli RG, Bhat AA, Cox AG, Taylor EW (1999). "Selenium-dependent glutathione peroxidase modules encoded by RNA viruses." (70(2)): 97-116.

[12] Taylor EW, N. R., Zhang W, Bhat A, Zhao L, Ruzicka JA, Dean R (2000). "HIV encoded selenoproteins as a basis for antioxidant therapy. Pharmacology of HIV infection and AIDS."

[13] Taylor EW, R. J., Premadasa LS , Zhao L (2016), "Cellular Selenoprotein mRNA Tethering via Antisense Interactions with Ebola and HIV-1 mRNAs May Impact Host Selenium Biochemistry" Current topics of medicinal chemistry: 1530-153

[14] Perrone, R., et al. (2013). "Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity." 8(8): e73121.

[15] Ramgopal Nadimpalli, J. A. H., Anita Thakur, Roger G. Dean, Ethan W. Taylor and Benjamin M. Blumberg (1996). "In vitro, in situ and theoretical genomic evidence suggests novel isoforms of HIV-1 nef are expressed by frameshifting and suppression of UGA termination codons."

[16] Gasdaska, J. R. (1999). "Regulation of Human Thioredoxin Reductase Expression and Activity by 3'-Untranslated Region Selenocysteine Insertion Sequence and mRNA Instability Elements." Journal of Biological Chemistry 274(36): 25379-25385.

[17] Zhao, L., B. Olubajo, and E.W. Taylor (2006). "Functional studies of an HIV-1 encoded glutathione peroxidase." BioFactors 27(1-4): 93-107.

[18] Zeng, Y.; Yi, R.; Cullen, B. R. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci U S A* 2003, *100*, 9779-84.

[19] Khorkova, O.; Myers, A. J.; Hsiao, J.; Wahlestedt, C. Natural antisense transcripts. *Hum Mol Genet* 2014, *23*, (R1), R54-63.

[20] Mahmoudi, S.; Henriksson, S.; Corcoran, M.; Mendez-Vidal, C.; Wiman, K. G.; Farnebo, M. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. *Mol Cell* 2009, *33*, 462-71.

[21] Berry, M. J.; Banu, L.; Harney, J. W.; Larsen, P. R. Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *EMBO J* 1993, *12*, 3315-22.

[22] Shisler, J. L.; Senkevich, T. G.; Berry, M. J.; Moss, B. Ultraviolet-induced cell death blocked by a selenoprotein from a human dermatotropic poxvirus. *Science* 1998, *279*, 102-5.

[23] Zhang, W.; Ramanathan, C. S.; Nadimpalli, R. G.; Bhat, A. A.; Cox, A. G.; Taylor, E. W. Selenium-dependent glutathione peroxidase modules encoded by RNA viruses. *Biol Trace Elem Res* 1999, *70*, 97-116.

[24] Zhao, L.; Cox, A. G.; Ruzicka, J. A.; Bhat, A. A.; Zhang, W.; Taylor, E. W. Molecular modeling and in vitro activity of an HIV-1-encoded glutathione peroxidase. *Proc Natl Acad Sci U S A* 2000, *97*, 6356-61.

[25] Cohen, I.; Boya, P.; Zhao, L.; Metivier, D.; Andreau, K.; Perfettini, J. L.; Weaver, J. G.; Badley, A.; Taylor, E. W.; Kroemer, G. Anti-apoptotic activity of the glutathione peroxidase homologue encoded by HIV-1. *Apoptosis* 2004, *9*, 181-92.

[26] Rehmsmeier, M.; Steffen, P.; Hochsmann, M.; Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* 2004, *10*, 1507-17.

[27] Busch, A.; Richter, A. S.; Backofen, R. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 2008, *24*, 2849-56.

[28] Wright, P. R.; Georg, J.; Mann, M.; Sorescu, D. A.; Richter, A. S.; Lott, S.; Kleinkauf, R.; Hess, W. R.; Backofen, R. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res* 2014, *42*, (Web Server issue), W119-23.

[29] Taylor, E. W.; Cox, A. G.; Zhao, L.; Ruzicka, J. A.; Bhat, A. A.; Zhang, W.; Nadimpalli, R. G.; Dean, R. G. Nutrition, HIV, and drug abuse: the molecular basis of a unique role for selenium. *J Acquir Immune Defic Syndr* 2000, *25 Suppl 1*, S53-61.

[30] Olubajo, B.; Taylor, E. W. A -1 frameshift in the HIV-1 env gene is enhanced by arginine deficiency via a hungry codon mechanism. *Mutat Res* 2005, *579*, 125-32.

[31] Ramanathan, C. S.; Taylor, E. W. Computational genomic analysis of hemorrhagic fever viruses. Viral selenoproteins as a potential factor in pathogenesis. *Biol Trace Elem Res* 1997, *56*, 93-106.

[32] Baum, M. K.; Shor-Posner, G.; Lai, S.; Zhang, G.; Lai, H.; Fletcher, M. A.; Sauberlich, H.; Page, J. B. High risk of HIV-related mortality is associated with selenium deficiency. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997, *15*, 370-4.

[33] Constans, J.; Pellegrin, J. L.; Sergeant, C.; Simonoff, M.; Pellegrin, I.; Fleury, H.; Leng, B.; Conri, C. Serum selenium predicts outcome in HIV infection. *J Acquir Immune Defic Syndr Hum Retrovirol* 1995, *10*, 392.

[34] Baum, M. K.; Campa, A.; Lai, S.; Sales Martinez, S.; Tsalaile, L.; Burns, P.; Farahani, M.; Li, Y.; van Widenfelt, E.; Page, J. B.; Bussmann, H.; Fawzi, W. W.; Moyo, S.; Makhema, J.; Thior, I.; Essex, M.; Marlink, R. Effect of micronutrient supplementation on disease progression in asymptomatic, antiretroviral-naive, HIV-infected adults in Botswana: a randomized clinical trial. *JAMA* 2013, *310*, 2154-63.

[35] Hurwitz, B. E.; Klaus, J. R.; Llabre, M. M.; Gonzalez, A.; Lawrence, P. J.; Maher, K. J.; Greeson, J. M.; Baum, M. K.; Shor-Posner, G.; Skyler, J. S.; Schneiderman, N. Suppression of human immunodeficiency virus type 1 viral load with selenium supplementation: a randomized controlled trial. *Arch Intern Med* 2007, *167*, 148-54.

[36] Jiamton, S.; Pepin, J.; Suttent, R.; Filteau, S.; Mahakkanukrauh, B.; Hanshaoworakul, W.; Chaisilwattana, P.; Suthipinittharm, P.; Shetty, P.; Jaffar, S. A randomized trial of the impact of multiple micronutrient supplementation on mortality among HIV-infected individuals living in Bangkok. *AIDS* 2003, *17*, 2461-9.

[37] Schrauzer, G. N.; Molenaar, T.; Kuehn, K.; Waller, D. Effect of simulated American, Bulgarian, and Japanese human diets and of selenium supplementation on the incidence of virally induced mammary tumors in female mice. *Biol Trace Elem Res* 1989, *20*, 169-78.

[38] Beck, M. A.; Kolbeck, P. C.; Shi, Q.; Rohr, L. H.; Morris, V. C.; Levander, O. A. Increased virulence of a human enterovirus (coxsackievirus B3) in selenium-deficient mice. *J Infect Dis* 1994, *170*, 351-7.

[39] Yu, S. Y.; Zhu, Y. J.; Li, W. G. Protective role of selenium against hepatitis B virus and primary liver cancer in Qidong. *Biol Trace Elem Res* 1997, *56*, 117-24.

[40] Hou, J. C. Inhibitory effect of selenite and other antioxidants on complement-mediated tissue injury in patients with epidemic hemorrhagic fever. *Biol Trace Elem Res* 1997, *56*, 125-30.

[41] McKenzie, R. C.; Rafferty, T. S.; Beckett, G. J. Selenium: an essential element for immune function. *Immunol Today* 1998, *19*, 342-5.

[42] Steinbrenner, H.; Al-Quraishy, S.; Dkhil, M. A.; Wunderlich, F.; Sies, H. Dietary Selenium in Adjuvant Therapy of Viral and Bacterial Infections. *Adv Nutr* 2015, *6*, 73-82.

[43] Meydani, M. Modulation of the platelet thromboxane A2 and aortic prostacyclin synthesis by dietary selenium and vitamin E. *Biol Trace Elem Res* 1992, *33*, 79-86.

[44] Huang, T. S.; Shyu, Y. C.; Chen, H. Y.; Lin, L. M.; Lo, C. Y.; Yuan, S. S.; Chen, P. J. Effect of parenteral selenium supplementation in critically ill patients: a systematic review and meta-analysis. *PLoS One* 2013, *8*, e54431.

[45] Bray, M.; Mahanty, S. Ebola hemorrhagic fever and septic shock. *J Infect Dis* 2003, *188*, 1613-7.

APPENDIX A

CELLULAR SELENOPROTEIN MRNA TETHERING VIA ANTISENSE INTERACTIONS WITH EBOLA AND HIV-1 MRNAS MAY IMPACT HOST SELENIUM BIOCHEMISTRY

Ethan Will Taylor[1]*, Ph.D.

Jan A. Ruzicka[1], Ph.D.

Lakmini Premadasa[1], B.S.

Lijun Zhao[2], Ph.D.

[1]Dept. of Nanoscience, University of North Carolina at Greensboro,

Joint School of Nanoscience and Nanoengineering,

2907 E. Gate City Blvd., Greensboro, NC 27401 USA

[2]Key Laboratory of Ministry of Education for

Medicinal Plant Resource and Natural Pharmaceutical Chemistry,

College of Life Sciences, Shaanxi Normal University, Xi'an 710062, China.

*Corresponding author.

Email: ewtaylor@uncg.edu

Phone: 1-336-314-884

Running title: Virus-host mRNA antisense interactions and selenium

Key words: Antisense, Ebola, mRNA, selenium, selenoprotein, HIV, tethering, thioredoxin reductase

ABSTRACT

Regulation of protein expression by non-coding RNAs typically involves effects on mRNA degradation and/or ribosomal translation. The possibility of virus-host mRNA-mRNA antisense tethering interactions (ATI) as a gain-of-function strategy, via the capture of functional RNA motifs, has not been hitherto considered. We present evidence that ATIs may be exploited by certain RNA viruses in order to tether the mRNAs of host selenoproteins, potentially exploiting the proximity of a captured host selenocysteine insertion sequence (SECIS) element to enable the expression of virally-encoded selenoprotein modules, via translation of in-frame UGA stop codons as selenocysteine. Computational analysis predicts thermodynamically stable ATIs between several widely expressed mammalian selenoprotein mRNAs (e.g., isoforms of thioredoxin reductase) and specific Ebola virus mRNAs, and HIV-1 mRNA, which we demonstrate via DNA gel shift assays. The probable functional significance of these ATIs is further supported by the observation that, in both viruses, they are located in close proximity to highly conserved in-frame UGA stop codons at the 3' end of open reading frames that encode essential viral proteins (the HIV-1 nef protein and the Ebola nucleoprotein). Significantly, in HIV/AIDS patients, an inverse correlation between serum selenium and mortality has been repeatedly documented, and clinical benefits of selenium in the context of multi-micronutrient supplementation have been demonstrated in several well-controlled clinical trials. Hence, in the light of our findings, the possibility of a similar role for selenium in Ebola pathogenesis and treatment merits serious investigation.

INTRODUCTION

Regulation of mRNA levels and protein synthesis by other RNA species, such as microRNA and siRNA, typically involves either the degradation of target mRNAs, or the inhibition of protein translation in the absence of RNA degradation; these alternative outcomes depend primarily on the degree of complementarity between the two RNAs [18]. For cellular genes having natural antisense transcripts (NATs), protein synthesis can be either downregulated [19] or upregulated [20] by NAT binding to the target mRNA. However, aside from such effects on mRNA degradation vs. stability and translational repression vs. enhancement, the possibility of antisense-based mRNA tethering as a general gain-of-function strategy, via the capture of functional RNA motifs, has not been widely considered, if at all.

Particularly for a small virus with a highly constrained genome size, an ideally suitable function for this hypothetical mechanism would be *the ability to gain additional protein coding potential via the capture of a selenocysteine insertion sequence (SECIS) element*, which would confer the ability for appropriately located UGA stop codons to be "recoded" and translated as selenocysteine (Sec). Since SECIS elements have been shown to function in *trans*, i.e., even when carried by a separate mRNA [21], it is highly probable that the presence of a SECIS in a tethered mRNA would also confer UGA-recoding ability on the tethering mRNA partner, i.e., in this case, the ability to express virally-encoded selenoprotein modules. There is no doubt that selenoprotein-encoding capability can serve a viral agenda, because a functional viral homologue of the prototypical selenoprotein, glutathione peroxidase (GPx), has been identified in the

genome of *Molluscum contagiosum*, a large DNA pox virus [22]. Encoding such an antioxidant selenoprotein could enhance the ability of a virus to respond to and survive oxidant-based immune attacks, thereby increasing viral fitness.

Significantly, our research group has identified GPx-related sequences with in-frame UGA codons in various RNA viruses, including hepatitis C virus and human immunodeficiency virus type one (HIV-1); the UGA-containing active site regions of these GPx-like modules are typically encoded in an overlapping reading frame of another viral protein [23]. The putative HIV-1 GPx was cloned and found to encode functional GPx activity [24] and to exert anti-oxidant and anti-apoptotic effects [25], but only when the viral protein was expressed as a selenoprotein via inclusion of a cellular SECIS element in the expression vector. *A functional SECIS element encoded by an RNA virus has never been demonstrated*. However, a virus would not need its own SECIS if it possessed a mechanism to hijack such a function from the host, which is not only a classic viral strategy, but arguably the essence of the viral life style.

In this study, we present both computational and experimental evidence for *antisense tethering interactions* (ATIs) between host selenoprotein mRNAs (specifically, thioredoxin reductase mRNAs) and certain mRNAs of RNA viruses. We focus here on the two most compelling examples we have identified, one involving the highly pathogenic Zaire strain of Ebola virus (EBOV), and another involving HIV-1. The first, shown as **A** in Fig. 1, is between the mRNA of the human thioredoxin reductase 3 (TR3) and the mRNA encoding the nucleoprotein (NP) of EBOV; the second (**B** in Fig. 1) is

between the mRNA of human thioredoxin reductase 1 (TR1) and a region of HIV-1 genomic mRNA, in the 3'-long terminal repeat (LTR).

There are two major possibilities for how these antisense interactions would be likely to impact host selenium biochemistry during viral infection: either simply by antisense inhibition of cellular selenoprotein synthesis, or by direct competition for a limited pool of selenocysteine if the ATI with a selenoprotein mRNA can enable viral selenoprotein synthesis. We will present a model for the latter, and discuss the implications of these findings for well-established links between dietary selenium status and viral pathogenesis.

MATERIALS AND METHODS

Identification of antisense interactions via computational methods

To evaluate the hypothesis that certain viral mRNAs might engage in antisense interactions with host selenoprotein mRNAs, potential antisense matches were initially identified via nucleotide BLAST searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The HIV-1 genomic mRNA sequence, and individual mRNA sequences of Ebola virus, were used as probes (see below for specific Genbank accession numbers of reference sequences used). Nucleotide BLAST (blastn option) was used with default search parameters, against a search set of the Reference RNA sequence database (refseq.rna), restricted to either Homo sapiens (taxid:9606) or various bat taxa (e.g., Old World fruit bats, taxid:9398), since bats are now believed to be the most probable reservoir species for Ebola virus, and therefore may be hosts to which they have adapted in terms of

potential mRNA interactions. The best candidates for ATI with selenoprotein-encoding mRNAs that were identified using BLAST and selected for further study were the HIV-1 nef region vs. TR1 and the Ebola NP vs. TR3, for both of which continuous 15 base pair antisense matches were identified via Blast, at high significance levels. The RNAHybrid program (http://bibiserv2.cebitec.uni-bielefeld.de/rnahybrid; [26] was then used to establish the extent (in the 5' and 3' directions beyond the core match identified using BLAST) and computed binding energies of the potential RNA-RNA interactions; this generated the hybridizations shown in Fig. 1. The putative ATIs were further validated using another widely used method for accurate prediction of RNA-RNA interactions, IntaRNA (http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp; [27-28]. This method uses a rigorous approach that considers not only the hybridization energy of the interacting pair of RNAs, but makes the assessment in the context of competing *intramolecular* mRNA secondary structures that must be unfolded in order for the intermolecular interaction to occur, using a sliding window to consider all possible competing intramolecular structures. For an interaction to be identified by the program, there must be a significant net interaction energy, after subtraction of the unfolding energies for each of the single strands. Using large input fragments of up to 1500 bases in length from the 3' end of each of the cognate mRNAs as input (HIV-1 vs. TR1 and Ebola NP vs. TR3), with default parameters, IntaRNA identified exactly the same core antisense interactions as previously established using BLAST and RNAhybrid (see Results and Discussion section for details).

Reference sequences used in computational studies and oligonucleotide design

Genbank accession numbers for the viral and host gene reference sequences used, and the relevant sequence ranges shown in Fig. 1, are: 1976 Zaire ebolavirus, NC_002549.1; 2014 Zaire ebolavirus, KJ660346.2 (nucleoprotein, 2350-2388); HIV-1, K02013.1 (nef/LTR region, 8989-9028); human TR1, NM_003330.3 (3612-3655); human TR3, NM_001173513.1 (1663-1703); fruit bat TR3, XM_006911434.

Demonstration of the predicted antisense interactions at the DNA level via gel shift assays.

Target-specific *in vitro* DNA hybridization of the cognate virus-selenoprotein pairs was demonstrated by electrophoretic mobility shift assays, using ~40mer synthetic single stranded ssDNA oligomers (Integrated DNA Technologies, Inc., Coralville, IA). Oligos (~1 μg each in 10 μl PBS), either singly or in cognate or mismatched pairs, were incubated at 37°C for 15 hours, except those in lanes 9 and 10 (Fig. 2), which were heated to 90°C for 10 minutes in the presence of 10 μg of sheered herring sperm DNA (Promega D1811, Madison, WI), followed by cooling to room temperature over 1 hour. The matched pairs of oligos from the respective viral and host mRNAs, corresponding to the sequence regions involved in the antisense interactions shown in Fig. 1, were as follows: EBOV nucleoprotein (Enp) vs. TR3, and HIV-1 nef/LTR region (Hnf) vs. TR1. Mismatched (non-cognate) mRNA pairs, e.g. Enp vs. TR1, were used as negative controls (see legend to Fig. 2 for details). Separation was on a 4% agarose gel with

ethidium bromide visualization. GeneRuler 1 kb Plus DNA ladder (Thermo Scientific, Waltham, MA) was used as a guide.

RESULTS AND DISCUSSION

Discovery of these antisense matches was initially guided via BLAST searches to identify potential core helical interactions of ~15 or more consecutive base pairs, which would correspond to a probability of < 5 x 10-9 in purely random DNA sequences. The best candidates identified using BLAST were the Ebola NP vs. TR3, and the HIV-1 nef region vs. TR1, which both had core antisense matches of 15 consecutive base pairs (Supplemental Material files S1 and S2), with highly significant BLAST "Expect" scores of 0.001 (Ebola NP vs. TR3) and 0.002 (HIV nef vs. TR1) vs. the human selenoprotein mRNA database. However, BLAST is not an ideal tool for discovering antisense interactions, because of the unique non-Watson-Crick base pairings available to RNA. To get a better sense of the significance and stability of these matches, the RNAHybrid program [26] was used to establish the extent and binding energies of the RNA:RNA interactions. The results are shown as **A** and **B** in Fig. 1, rendered as both RNA secondary structures and antisense sequence alignments. Both of these predicted interactions are more energetically stable and extensive than typical microRNA binding interactions, such as that shown as structure **C**, which are inherently limited by the 22 nucleotide size of microRNA.

These results were further validated by the IntaRNA program [27-28], which output essentially identical RNA-RNA interactions as those shown in Fig. 1 for the Ebola

NP vs. TR3 and HIV-1 nef region vs. TR1 (for which the raw IntaRNA output is in Supplemental Material files S3 and S4 respectively). Relative to the RNAhybrid results shown in Fig. 1, the IntaRNA results for the same mRNA pairs, although slightly truncated at either one (HIV-TR1) or both ends (Ebola-TR3), are identical in their core hybridizations. The IntaRNA results extend the BLAST 15-bp core antisense matches to 17/18 consecutive base pairs for the HIV-1 nef region vs. TR1, and 21/22 base pairs, with one single-base insertion, for Ebola NP vs. TR3 (see files S3 and S4). In both cases, even after subtraction of computed unfolding energies for internal mRNA secondary structures, both interacting pairs still have net interaction energies of ~20 kcal/mol, and are the unique results predicted by IntaRNA for these mRNA pairs.

Alignment **A** in Fig. 1 also shows differences between 1976 and 2014 strains of EBOV, and between humans and fruit bats, in the antisense-tethered region of TR3. In the mutated positions (black italic letters above or below the alignment), the 2014 EBOV sequence is in all cases a better antisense match to the human TR3. In one location where the mutations align, a C base unique to 1976 EBOV gives a better match to the fruit bat TR3 (GC base pair), whereas the 2014 EBOV has mutated to a U, giving a better match to the human TR3 (AU base pair). This is consistent with the possibility that, in regard to this putative interaction with TR3 mRNA, EBOV may be gradually adapting from bats to humans and other primate hosts such as chimpanzees and gorillas, whose TR3 sequences are identical to humans in this region.

The predicted interactions are unambiguously supported by gel shift assays, which show strong *in vitro* DNA hybridization between the antisense partners. In Fig. 2,

the matched pairs of oligos corresponding to **A** and **B** in Fig. 1 are labeled Enp/TR3 and Hnf/TR1 respectively. Combined at a 1:1 ratio, both cognate antisense pairs form predominantly double-stranded dsDNA, either in buffer solution (lanes 2 and 6), or in the presence of sheared cellular DNA (lanes 9 and 10); at a 2:1 ratio, both single-stranded ssDNA and dsDNA bands are observed (lanes 3 and 7). The interactions are specific, as no visible dsDNA is formed from mismatched pairs like Enp/TR1, Hnf/TR3, or Hnf/Enp under identical conditions (lanes 11-13).

The existence of ATI between the 3' regions of viral mRNAs and host selenoprotein mRNAs suggests a new model for viral selenoprotein synthesis, as a variant of the known mechanism of eukaryotic cellular selenoprotein synthesis (Fig. 3). As shown schematically in Fig. 3B, the "tail-to-tail" antisense interaction between the 3' ends of the TR1 and HIV-1 mRNAs spans the 3' end of the viral *nef* coding sequence, which terminates in a UGA codon that is highly conserved in global HIV-1 isolates [29]. Significantly, the EBOV NP gene also terminates in a UGA codon, conserved in all 1976 through 2014 Zaire EBOV isolates, but not in the much less pathogenic Reston ebolavirus, in which the NP terminates in a UAA. In both HIV-1 *nef* and the EBOV NP mRNAs, recoding of the terminal UGA of these proteins as Sec via a tethered SECIS element (Fig. 3B) would produce a slightly extended isoform of the protein, containing a single selenium atom as Sec, plus a few residues encoded past the UGA. This is analogous to TR, where the UGA is also at the protein C-terminus; in HIV-1 *nef*, the sequence even mimics the TR redox center, with a conserved Cys immediately preceding the UGA [29].

The presence in a viral mRNA of an antisense region capable of tethering a host selenoprotein mRNA, located either immediately overlapping (HIV-1) or within 300 bases (Ebola NP) of a highly conserved in-frame UGA codon, is strong circumstantial evidence favoring the possibility that the UGA can be translated as Sec, even if inefficiently. Additionally, in both viral mRNAs, there is a shorter antisense binding region (labeled ATI-2 in Fig. 3) immediately *downstream* of the conserved UGA codon. Both viral mRNAs also have *upstream* -1 ribosomal frameshift sites: the viral GPx in HIV-1 [24, 30] and predicted sites in Ebola virus [31], that would enable access to other in-frame UGA codons during protein synthesis, providing a further rationale for a need to capture a host SECIS element. Because HIV-nef and Ebola-NP are produced in abundant quantities in infected cells, even if recoding of their terminal UGA codons as Sec was very inefficient, production of only a fraction of a percent of a selenium-containing isoform by this mechanism could perturb the synthesis of cellular selenoproteins, by depletion of a limited pool of Sec.

An alternate hypothesis to viral selenoprotein synthesis is that these antisense interactions may simply lead to downregulation of host selenoproteins via translational inhibition analogous to that by microRNA, without mRNA degradation [18]. In either case, levels of cellular selenoproteins would likely decrease, impairing host defenses in general and antioxidant defenses in particular, and/or helping to create conditions more favorable for viral replication and transfer to new hosts (e.g., coagulopathy in Ebola, because of the role of selenium in the regulation of blood clotting).

In the case of HIV-1, a negative correlation between selenium status and mortality has been firmly established [32-33], and significant clinical benefits of selenium supplementation have been demonstrated in various studies [34-36]. Nor is HIV an isolated example, because there are other established cases of chemoprotective/antiviral effects of dietary selenium, including mammary tumors caused by MMTV, a retrovirus [37], Keshan disease myocarditis, linked to coxsackie virus [38], liver cancer and hepatitis linked to hepatitis B virus [39], and even one example in which an Asian viral hemorrhagic fever was successfully treated with oral sodium selenite, giving an overall 80% reduction in mortality [40]. These results are not entirely unexpected, considering the many essential roles of selenium in the immune system [41].

Evidence for benefits of selenium supplementation in various RNA viral infections, and the host mechanisms involved, was recently reviewed by Steinbrenner et al., who pointed out that “Populations in several countries most afflicted by past and current outbreaks of Ebola fever (e.g., Liberia, Guinea, Democratic Republic of Congo)

exhibit a high risk of selenium deficiency", with Liberia being the lowest ranked African nation for dietary selenium supply [42].

Regarding Ebola, selenium is known to play a significant role in the regulation of blood clotting [43], and thus could play a role in the coagulopathy characteristic of Ebola hemorrhagic fever [31]. Intravenous selenite is also an effective treatment for septic shock [44], which has clinical similarities to hemorrhagic fever [45]. Such observations are consistent with the possibility of a virus-induced selenium defect, resulting from ATIs between Ebola mRNAs and host selenoprotein mRNAs, which could impair host selenoprotein synthesis by antisense translational inhibition, or by competition for Sec, or both. In either case, increasing selenium intake would be expected to reduce the detrimental effects of Ebola on host Se-dependent mechanisms, and selenium deficiency would be predicted to be a risk factor for increased mortality.

If these ATIs are found to be functionally significant at the mRNA level, and their effects confirmed via proteomics studies, it will necessitate a re-evaluation of the multifaceted role of selenium in virus-host interactions, and its clinical significance, particularly in Ebola infections.

Conflict of Interest: The authors have no conflicts of interest to declare.

Acknowledgments: The authors would like to thank Dr. Adam Hall for reviewing the manuscript and for providing helpful suggestions. This research was supported by a gift from the Dr. Arthur and Bonnie Ennis Foundation, Decatur, IL, to E.W.T.

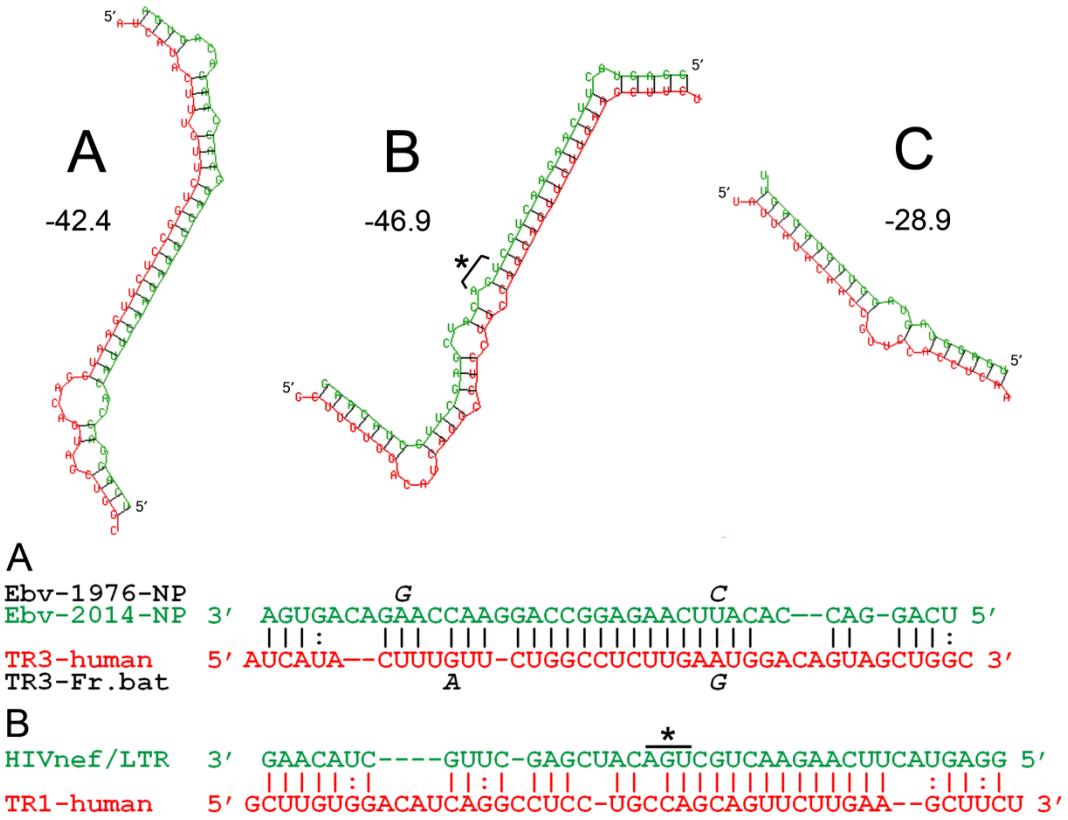


Figure 21. Predicted Antisense interactions between Regions of mRNAs of Human Thioredoxin Reductases and Viral mRNAs. These are shown as both RNA secondary structures and antisense sequence alignments. The RNA structures shown and computed interaction free energies in kcal/mol (numerals next to the structures) were generated using the RNAHybrid 2.2 program (see Methods). The antisense matches are: A. EBOV nucleoprotein mRNA (Ebv-2014-NP, green) vs. TR3 mRNA (TR3-human, red); B. HIV-1 (HIVnef/LTR, green) vs. TR1 (TR1-human, red); C. A typical imperfect microRNA interaction is included for comparison (Let-7 vs. a cellular target, the default example in RNAHybrid). The alignments A and B correspond exactly to the RNA secondary structures above; GU base pairs are indicated by a colon, and the highly conserved UGA stop codon of the HIV-1 *nef* gene is indicated by an asterisk. In alignment A, the letters in italics above and below the sequences correspond to genomic sequence variations between the 2014 EBOV (green) and the earliest 1976 EBOV isolates (black italics), and between human TR3 (red) and fruit bat TR3 (black italics); see text for a discussion of these mutations.

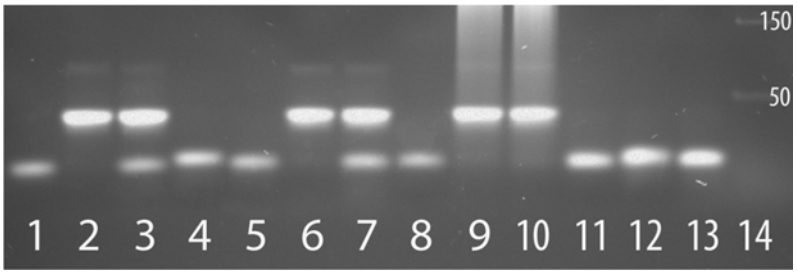


Figure 22. Virus vs. Human Selenoprotein Antisense interactions demonstrated at the DNA level for the Ebola Nucleoprotein and HIV *Nef* Regions shown in Fig. 1. Target-specific *in vitro* DNA hybridization of the cognate virus-selenoprotein pairs is shown by gel shift assay using ~40mer synthetic single stranded ssDNA oligos. The matched pairs of oligos from the respective viral and host mRNAs, corresponding to the sequence regions involved in the antisense interactions shown in Fig. 1, were as follows: EBOV nucleoprotein (Enp) vs. TR3, and HIV-1 nef/LTR region (Hnf) vs. TR1. Lanes have either a single oligo that runs as ssDNA (the lowest bands) or an incubated pair of oligos, as follows: 1. Enp; 2. 1:1 Enp+TR3; 3. 2:1 Enp+TR3; 4. TR3; 5. Hnf; 6. 1:1 Hnf+TR1; 7. 2:1 Hnf+TR1; 8. TR1; 9. 1:1 Enp+TR3 plus sheered herring sperm DNA; 10. 1:1 Hnf+TR1 plus herring DNA; 11. 1:1 Enp+TR1; 12. 1:1 Hnf+TR3; 13. 1:1 Hnf+Enp; 14. DNA size markers. The bright bands at the size of ~40mer double stranded dsDNA correspond to the expected hybridizations (Enp+TR3 or Hnf+TR1) at 1:1 (lanes 2, 6, 9, 10) and 2:1 ratios (lanes 3, 7). The faint bands above the 50 size marker (e.g. lanes 2, 3, 6, 7) are possibly from trimer formation (e.g. TR3+Enp+TR3, etc.).

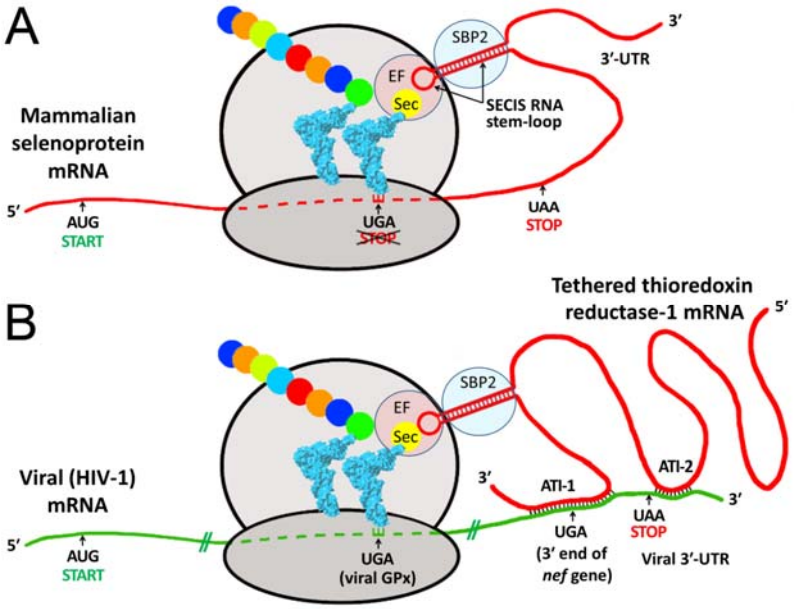


Figure 23. Proposed Mechanism of Selenocysteine (Sec) incorporation into Viral Proteins via hijacking of a SECIS Element from a Tethered host Selenoprotein mRNA. Both panels show schematic ribosomes with bound tRNAs (blue), one carrying the rare selenium-containing amino acid Sec, the other a growing peptide chain (colored beads). A. During biosynthesis of mammalian selenoproteins, Sec is inserted at UGA codons, normally a STOP signal for protein synthesis. This mechanism requires a specialized RNA stem-loop structure, the SECIS element, generally located in the 3' untranslated region (3'-UTR) of the selenoprotein mRNA [4]. By recruiting various protein factors, including SECIS binding protein 2 (SBP2) and elongation factor Sec (EF), the SECIS element enables delivery of $tRNA_{Sec}$ to the ribosome for Sec incorporation at the UGA codon, preventing it from acting as a stop signal. B. Using HIV-1 as an example, the lower panel shows how a viral mRNA, via antisense tethering interactions (ATI), could hijack a host SECIS element for decoding viral selenoprotein modules, such as the HIV-encoded glutathione peroxidase (viral GPx) [7-8, 13]. ATI-1 is the interaction shown as structure B in Fig. 1, and spans the highly conserved 3'-UGA codon of the *nef* gene; ATI-2 is a second shorter antisense region, consisting of 13 consecutive Watson-Crick base pairs near the end of the viral mRNA (bases 9111-9123, CAGCUGCUUUUUG). Similarly, in the Ebola nucleoprotein mRNA, less than 350 bases from the ATI shown as structure A in Fig. 1, there is also a secondary ATI region downstream of the conserved 3'-UGA stop codon (bases 2719-2734, CGACAAAUAGCUAACA), with only one mismatch to TR3 over 16 base pairs (A, opposite a G in TR3).