






Universitat Autònoma de Barcelona

**ADVERTIMENT.** L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  [http://cat.creativecommons.org/?page\\_id=184](http://cat.creativecommons.org/?page_id=184)

**ADVERTENCIA.** El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

**WARNING.** The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

**Universitat Autònoma de Barcelona**  
Facultat de Medicina  
Departament de Biologia Cel·lular, Fisiologia I Immunologia

**Synonymous changes in the  
Human Immunodeficiency Virus genome  
as a strategy to study virus biology**

Ana Jordán de Paiz

September 2019

Institut de Recerca de la SIDA (IrsiCaixa)  
Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IGTP)

Doctoral thesis to obtain a PhD degree in Advanced Immunology  
from Universitat Autònoma de Barcelona

Director: Dr. Miguel Ángel Martínez de la Sierra  
Tutor: Dra. Dolores Jaraquemada Pérez de Guzmán

Thesis director

PhD candidate

**Dr. Miguel Ángel Martínez de la Sierra**

**Ana Jordán de Paiz**





El Dr. Miguel Ángel Martínez de la Sierra, investigador del Institut de Recerca de la SIDA (IrsiCaixa) y del Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IGTP) del Hospital Germans Trias i Pujol,

Certifica:

Que el trabajo experimental y la redacción de la memoria de la Tesis doctoral titulada **“Synonymous changes in the Human Immunodeficiency Virus genome as a strategy to study virus biology”** se ha realizado por Ana Jordán de Paiz bajo su dirección y que es apta para ser presentada y optar al grado de Doctora en Inmunología Avanzada por la Universitat Autònoma de Barcelona.

Y para que quede constancia, firma este documento.

Badalona, 25 de septiembre de 2019.

Dr. Miguel Ángel Martínez de la Sierra



This work has been supported by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO) and the European Regional Development Fund (FEDER) under agreement (BES-2014-068111/SAF2013-41421-R/SAF2016-75277-R/EEBB-I-18-12809), and by Grifols Grant miRNA 2018.

Cover design: Fran Rodríguez Barrios and Ana Amaya Hernández, 2019

The printing of this thesis was possible by the financial support of the Universitat Autònoma de Barcelona.



***“One is not born, but rather becomes, a woman”***

Simone de Beauvoir

***“When you know you’re right, you don’t care what others think.  
You know that sooner or later, it will come out in the wash”***

Barbara McClintock





***A mi madre***



---

<b>Summary</b> .....	<b>15</b>
<b>Resumen</b> .....	<b>17</b>
<b>Abbreviations</b> .....	<b>19</b>
<b>Introduction</b> .....	<b>23</b>
Human Immunodeficiency Virus 1 (HIV-1).....	25
HIV-1 particle structure .....	25
HIV-1 genome: genes and their protein products .....	26
Replication cycle.....	27
Temporal regulation of the genes.....	29
HIV-1 protease .....	31
HIV-1 envelope .....	33
Rev Response Element (RRE) .....	34
Codon usage.....	35
HIV-1 genome composition and innate immune system.....	37
Codon pair usage.....	39
Synthetic synonymous genome recoding .....	40
<b>Hypothesis and Objectives</b> .....	<b>43</b>
<b>Materials and Methods</b> .....	<b>47</b>
Cell lines .....	49
Plasmids.....	49
Gp160-based mutants.....	52
Protease-based mutants .....	54
Transfections.....	57
MT-4 cells transfection.....	57
293T cells transfection .....	59
Real-time quantitative PCR .....	60
Immunoblot analyses .....	60
HIV-1 drug susceptibility tests.....	60
Selection of ATV- and DRV- resistant viruses.....	61
Ultradeep sequencing .....	61
RNA secondary structures: Computational analyses .....	63
ViennaRNA Package 2.4.10 .....	63
VARNA .....	64

CD-HIT-EST.....	64
MAFFT and LocARNA.....	64
LRIScan.....	64
Statistical Analyses.....	65
<b>Chapter One. Codon pair recoding of the protease gene and its evolvability .....</b>	<b>67</b>
Recoding of the protease gene .....	69
ATV and DRV selective pressure .....	70
Ultradeep sequencing .....	71
Exploring resistance substitutions .....	73
Replication capacity assays .....	74
Phylogenetic analyses .....	75
<b>Chapter Two. Codon recoding of the envelope gene .....</b>	<b>77</b>
Envelope recoding: changes in codon usage .....	79
Viral stability of the Recoded-env .....	81
New mutants based on Recoded-env .....	81
Viral stability of the new designed mutants .....	83
Reversion of mutations in the 3' region.....	86
Viral stability of all mutants .....	90
Infectivity capacity .....	93
Effects on translation: protein expression .....	94
Effects on transcription: mRNA levels.....	96
RNA secondary structures.....	97
Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) .....	104
<b>Chapter Three. Codon pair recoding of the envelope gene: effect of CpG dinucleotides .....</b>	<b>107</b>
Envelope recoding: Changes in codon pair bias.....	109
Viral stability of the viruses.....	111
Replication capacity of MinCpG.2 .....	112
Viral stability to neutralizing antibody .....	113
Reversion of the 3' env region to WT.....	113
Viral stability of the viruses.....	114
Replication capacity of the viruses.....	116
Effects on translation: protein expression .....	117
Effects on transcription: mRNA levels.....	117
Analyses of the CpGs in the 5' env region.....	119

Analyses of the 3' <i>env</i> region .....	121
RNA secondary structures .....	123
<b>Discussion .....</b>	<b>127</b>
<b>Conclusions .....</b>	<b>141</b>
<b>Bibliography .....</b>	<b>145</b>
<b>Publications .....</b>	<b>161</b>
<b>Annex .....</b>	<b>165</b>



## Summary

Synonymous genome recoding has been widely used to study different aspects of viral biology. It is based on modifying the nucleotide sequence of a gene without changing the amino acid sequence, thus the protein sequence is not affected. Previous studies have demonstrated viral attenuation by reduction in protein expression after synonymous recoding. Likewise, this tool has allowed the identification of RNA secondary structures, the description of viral factors or interactions with the innate immune system, as well as unravelling the temporal regulation of viral genes. Here, we synonymously recoded the Human Immunodeficiency Virus 1 (HIV-1) protease and envelope genes. We compared the development of HIV-1 resistance to protease inhibitors (PIs) between wild-type (WT) virus and a synthetic virus (MAX) carrying a codon pair re-engineered protease sequence including 38 (13%) synonymous mutations. WT and MAX viruses showed indistinguishable replication. Both viruses were subjected to serial passages in MT-4 cells with selective pressure from the PIs atazanavir (ATV) and darunavir (DRV) and they both developed phenotypic resistance to PIs. Ultra-deep sequence clonal analysis revealed that both viruses harbored previously described resistance mutations to ATV and DRV. However, the WT and MAX virus proteases showed different resistance variant repertoires. Our results indicate that HIV-1 recoded protease showed similar mutational robustness and evolvability to WT, but the position in sequence space delineates the evolution of each mutant spectra. To study how synonymous mutations affected virus replication, Env mutants were designed by changing its codon usage and codon pair usage. In addition, the number of CpG dinucleotides were also altered. The synthetic Recoded-env virus variant was lethal for HIV-1 after changing its codon usage by synonymously altering 39 nucleotides (1,5%). Protein expression analyses revealed that translation was modified in Recoded-env. Several mutants were designed based on Recoded-env to further investigate the lethality of this mutant. One of them, Recoded\_env\_3'Bwt, only reverted to WT two mutations of the same codon, named codon 34, located in gp41 coding region. Interestingly, this virus variant did not show significant differences in replication capacity both in MT-4 cells and PBMCs nor a dramatic decrease in protein production, when compared to WT. Analyzing sequences of gp41 in which codon 34 is located, we concluded that the WT RNA secondary structure was severely disrupted in Recoded-env. Similarly, after changing *env* codon pair usage, different virus variants which optimized, deoptimized or maintained neutral the codon pair bias (CPB) confirmed the relevance of synonymous substitutions in gp41 coding region. Only one of the virus variants, MinCpG.2, was replicative without reverting gp41 to WT. Differences in this region were found between MinCpG.2 and the rest of the designed CPB variants that



explained their lethality. We also observed that two virus variants, Max-3'wt and MaxCpG-3'wt, had significantly lower replication capacity when compared to WT although they had an optimized CPB. Moreover, the CPB virus variants with different number of CpGs allowed us to conclude that, in our model, increasing the number of CpGs did not attenuate the virus. Overall, our results confirmed that synonymous recoding a gene or genome is a useful tool to unravel different aspects of the virus biology.

## Resumen

La recodificación sinónima de los genomas se ha usado ampliamente para estudiar diferentes aspectos de la biología de los virus. Esta herramienta se basa en modificar la secuencia de nucleótidos de un gen sin cambiar la secuencia de aminoácidos, es decir, la secuencia de la proteína no se ve afectada. Estudios previos han demostrado atenuación viral debida a una reducción en la expresión proteica después de llevar a cabo la recodificación sinónima. De la misma manera, esta herramienta ha permitido la identificación de estructuras secundarias del ARN, la descripción de factores virales o interacciones con el sistema inmune innato, así como desentrañar la regulación temporal de los genes virales. En este trabajo hemos recodificado sinónimamente los genes de la proteasa y de la envuelta (Env) del Virus de Inmunodeficiencia Humana 1 (HIV-1). Comparamos el desarrollo de resistencias del HIV-1 a inhibidores de la proteasa (IPs) entre el virus salvaje (WT) y un virus sintético (MAX) que lleva una secuencia de la proteasa recodificada en el uso de pares de codones incluyendo 38 mutaciones sinónimas (13%). Los virus WT y MAX no muestran diferencias replicativas. Ambos virus fueron propagados en células MT-4 con la presión selectiva a dos IPs atazanavir (ATV) y darunavir (DRV). Tanto el virus WT como el MAX desarrollaron resistencias fenotípicas a IPs. Análisis clonales de secuenciación masiva revelaron que ambos virus desarrollaron mutaciones de resistencia a ATV y DRV previamente descritas. Sin embargo, las resistencias que desarrollaron las proteasas WT y MAX eran diferentes entre ambas variantes. Nuestros resultados indican que la proteasa recodificada del HIV-1 mostró una robustez y evolvabilidad similar al WT, pero la posición en el espacio de secuencias delinea la evolución de cada espectro mutante. Para estudiar cómo las mutaciones sinónimas afectan a la replicación viral, se diseñaron mutantes de Env que cambiaban el uso de codones y el uso de pares de codones; el número de dinucleótidos CpG también se alteró. La variante sintética del virus Recoded-env resultó letal para la viabilidad del VIH-1 después de cambiar su uso de codones mediante la modificación de 39 nucleótidos (1,5%). Análisis de expresión proteica revelaron que la traducción se había modificado en Recoded-env. Varios mutantes se diseñaron basados en Recoded-env para investigar la letalidad de este mutante. Uno de ellos, Recoded\_env\_3'Bwt, sólo revirtió a WT dos mutaciones de un mismo codón, el codón 34, localizado en la región codificante gp41. Es de resaltar que esta variante del virus no mostró diferencias en replicación en células MT-4 o PBMCs ni una reducción dramática en expresión proteica en comparación con el WT. El análisis de secuencias de gp41, donde se encuentra localizado el codón 34, demostró que la estructura secundaria del ARN estaba severamente interrumpida en Recoded-env. De la misma forma, después de cambiar el uso de pares de codones de *env*, se generaron

diferentes variantes que optimizaban, deoptimizaban o mantenían neutro el sesgo de pares de codones (CPB). Estas variantes confirmaron la importancia de las sustituciones sinónimas en la región codificante de gp41. Solo una de estas variantes, MinCpG.2, era viable con la región gp41 mutada. Se encontraron diferencias en esta región entre MinCpG.2 y el resto de variantes CPB que explicaban la letalidad. También observamos que dos variantes virales que tenían un CPB optimizado, Max-3'wt y MaxCpG-3'wt, mostraban una capacidad replicativa significativamente reducida cuando se comparó con el WT. Además, las variantes de CPB con diferentes proporciones de CpGs nos permitieron concluir que, en nuestro modelo, un número elevado de CpGs no atenuaba el virus. Nuestros resultados confirman que la recodificación sinónima de un gen o genoma es una herramienta muy útil para descifrar diferentes aspectos de la biología viral.

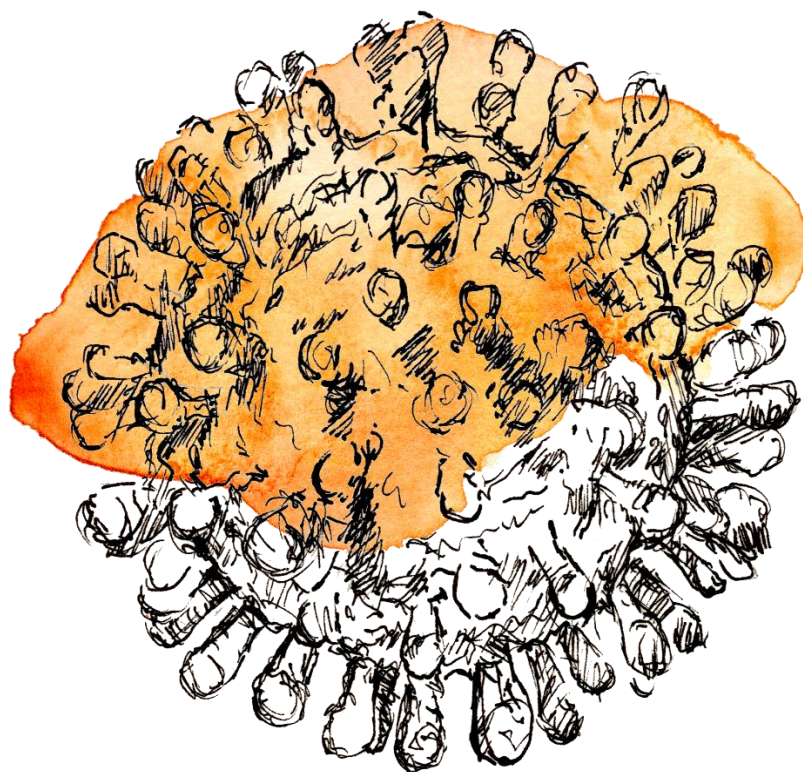
# Abbreviations

<b>µg</b>	Microgram
<b>µl</b>	Microliter
<b>Ab</b>	Antibody
<b>AIDS</b>	Acquired Immune Deficiency Syndrome
<b>ATCC</b>	American Type Culture Collection
<b>ATV</b>	Atazanavir
<b>CAI</b>	Codon Adaptation Index
<b>cDNA</b>	Complementary DNA
<b>CPB</b>	Codon Pair Bias
<b>CpG</b>	5'-Cytosine-phosphate-Guanine-3'
<b>CPS</b>	Codon Pair Score
<b>Ct</b>	Threshold cycle
<b>DMEM</b>	Dulbecco's Modified Eagle's medium
<b>DNA</b>	Deoxyribonucleic Acid
<b>DRV</b>	Darunavir
<b>E. coli</b>	Escherichia coli
<b>EGFP</b>	Enhanced Green Fluorescent Protein
<b>Env</b>	HIV-1 Envelope
<b>FBS</b>	Fetal Bovine Serum
<b>Fig</b>	Figure
<b>Fw</b>	Forward
<b>gp120</b>	Glycoprotein 120
<b>gp160</b>	Glycoprotein 160

<b>gp41</b>	Glycoprotein 41
<b>HAART</b>	Highly Active Antiretroviral Therapy
<b>HIV-1</b>	Human Immunodeficiency Virus type 1
<b>IC50</b>	Half maximal Inhibitory Concentration
<b>IN</b>	Integrase
<b>Kb</b>	Kilobase
<b>kcal</b>	Kilocalorie
<b>log</b>	Logarithm
<b>LTR</b>	Long Terminal Repeats
<b>MFE</b>	Minimum Free Energy
<b>Min</b>	Minute
<b>ml</b>	Milliliter
<b>mRNA</b>	Messenger RNA
<b>ng</b>	Nanogram
<b>nM</b>	Nanomolar
<b>nt</b>	Nucleotide
<b>°C</b>	Degree Celsius
<b>PAMP</b>	Pathogen-Associated Molecular Pattern
<b>PBMC</b>	Peripheral Blood Mononuclear Cell
<b>PBS</b>	Phosphate Buffered Saline
<b>PCR</b>	Polymerase Chain Reaction
<b>PI</b>	Protease Inhibitor
<b>pmol</b>	Picomole
<b>PMSF</b>	Phenylmethylsulfonyl Fluoride
<b>PR</b>	Protease

<b>qPCR</b>	Quantitative Polymerase Chain Reaction
<b>RER</b>	Rough Endoplasmic Reticulum
<b>RNA</b>	Ribonucleic Acid
<b>RPMI</b>	Roswell Park Memorial Institute
<b>RRE</b>	Rev Response Element
<b>RRV</b>	Rhesus monkey Rhadinovirus
<b>RT</b>	Reverse Transcriptase
<b>RT-PCR</b>	Reverse Transcription Polymerase Chain Reaction
<b>Rv</b>	Reverse
<b>SD</b>	Standard Deviation
<b>Sec</b>	Second
<b>SLFN11</b>	Schlafen 11
<b>TAR</b>	Transactivation Response Element
<b>Tat</b>	Trans-activator of Transcription
<b>TCID50</b>	Tissue Culture Dose for 50% of Infectivity
<b>TLR9</b>	Toll-like receptor 9
<b>U</b>	Units
<b>Vif</b>	Viral Infectivity Factor
<b>Vpr</b>	Viral Protein r
<b>Vpu</b>	Viral Protein u
<b>WT</b>	Wild-type





## Introduction

---





## **Human Immunodeficiency Virus 1 (HIV-1)**

The Human Immunodeficiency Virus type 1 (HIV-1) is a virus that belongs to the *Retroviridae* family, more specifically to the *Lentivirus* genus. These viruses are characterized by possessing a single-stranded, positive-sense RNA which is retrotranscribed to DNA. Following retrotranscription, the DNA molecule integrates into the genomic DNA of the host cell (Coffin 1992).

The HIV-1 preferentially infects CD4+ T cells. When the infection becomes chronic, a severe reduction in CD4 + T cells is produced. Without treatment, the infection can result in Acquire Immunodeficiency Syndrome (AIDS). The first evidence of the existence of the virus was reported in 1981, when young homosexual men in the United States were diagnosed with Kaposi's sarcoma and other opportunistic infections (Friedman-Kien 1981). However, it was not isolated and identified until 1983 (Barre-Sinoussi et al. 1983).

In 2018, there were 37.9 million living with HIV. Moreover, 1.7 million became newly infected. More than half of the infected people are from eastern and southern Africa. Among these Africans, only 67% had access to treatment. During 2018, 700.000 people died from AIDS-related illnesses. Overall, around 74.9 million people have become infected since the start of the epidemic (UNAIDS 2018).

Phylogenetic analyses divide the HIV-1 into four different groups: M (Major), O (Outlier) (Maucière et al. 1997), N (Non-M, non-O) (Simon et al. 1998) and P (Putative) (Plantier et al. 2009). Group M is the responsible for most of the HIV-1 infections around the world and it can be divided into several subtypes (A-H). The subtypes represent different lineages of HIV.

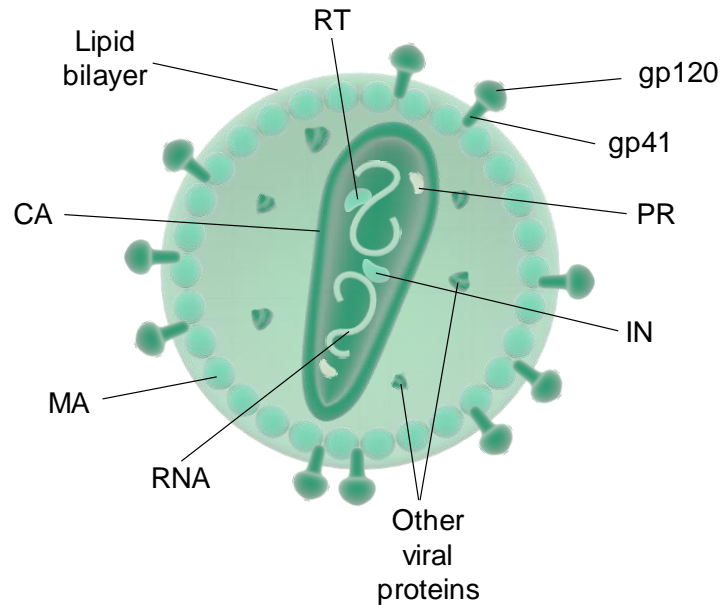
### *HIV-1 particle structure*

The HIV-1 virion particle is a sphere of about 100-120nm. In the outermost region there is a lipid bilayer obtained from the host cell membrane during virion formation. Attached to this lipid bilayer, about 10-14 HIV envelope particles are found (Figure 1). These envelope particles are formed by trimers of gp120 (surface protein) and trimers of gp41 (transmembrane protein).

The matrix is found below the envelope. It is composed of an association of p17 proteins, which are responsible for the integrity of the virion particle (Figure 1).

## Introduction

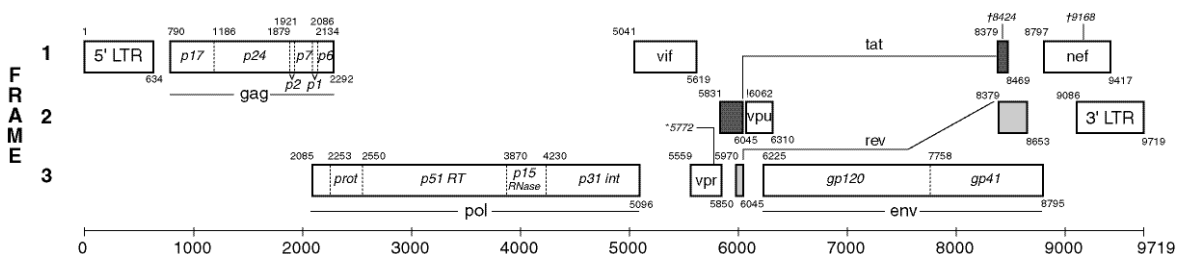
A conic capsid is formed of p24 proteins enclosing two copies of positive single-stranded RNA. Nucleocapsid proteins (p7, p6, RT and IN) are tightly attached to the RNA particles. Other proteins, such as Nef, Vif, Vpr and PR are also inside the viral particle (Figure 1).



**Figure 1. Schematic representation of the HIV-1 viral particle structure.** The RNA is located inside the capsid, tightly attached to nucleocapsid proteins. CA: capsid; MA: matrix; IN: integrase; PR: protease; RT: reverse transcriptase; gp120 and gp41 form the envelope.

### *HIV-1 genome: genes and their protein products*

The HIV-1 genome has a length of almost 10kb containing nine genes which encode for fifteen proteins. These genes are located within the three different reading frames (Figure 2).



**Figure 2. Schematic representation of the HXB2 strain genome.** The three frames are indicated on the left of the illustration. The nine reading frames are shown in rectangles. The long terminal repeats (LTRs) are also shown in rectangles at both sides of the genome. Illustration from [www.hiv.lanl.gov](http://www.hiv.lanl.gov).

HIV-1 genes can be divided into three groups, depending of the function of the proteins they encode for:

Structural and enzymatic proteins

- Gag. This gene encodes for the proteins of the capsid (CA). The polyprotein precursor is cleaved into the matrix (MA) (p17), capsid (p24), nucleocapsid (NC) (p7) and p6 proteins. This cleavage process is accomplished by the viral protease (PR).
- Pol. The precursor polyprotein Gag-Pol is cleaved by the viral PR into the viral enzymes PR, reverse-transcriptase (RT) and integrase (IN).
- Env. The precursor obtained from this gene is called gp160, which is processed to give the glycoproteins gp120 and gp41.

Regulatory proteins

- Tat. The first of the two regulatory factors of the gene expression of the HIV; Tat is the trans-activator, which binds to the transactivator response element (TAR) to enhance the transcription process.
- Rev. This is the second regulatory factor and it is responsible for the export to the nucleus of the large mRNAs containing the Rev Response Element (RRE).

Accessory proteins

Vif, Vpr, Vpu and Nef: In general, these proteins are not essential for viral replication in tissue cultures. However, they are highly conserved among different isolates, suggesting a very important role *in vivo*.

Non-coding long terminal repeats (LTRs) are found in the extremes of the HIV-1 genome. They are essential for the transcription initiation (5'LTR) and the polyadenylation (3'LTR).

*Replication cycle*

The HIV-1 infects human CD4+ T cells to use its machinery to generate more virion copies. All the processes, from the viral entry until the exit of the new virions, are known as the life cycle of the HIV-1. This replication cycle has several stages (Figure 3):

- I. Virus entry: binding, fusion and uncoating.

The first step that is needed for the entry of the virus is to bind to the host cell. This is produced when gp120 binds the cell CD4 receptor. After that, a structural changed

is produced in gp120 that allows its interaction with other cellular co-receptors (mainly chemokine receptor CXCR4 or/and chemokine receptor CCR5).

This interaction provokes a conformational change in gp41 that allows the fusion of the host cell membrane with the viral lipid bilayer. The uncoating process ends with the entry of the viral core to the cytoplasm of the host cell.

### II. Reverse-transcription.

Viral RNA retro-transcription is catalyzed by the RT enzyme. DNA synthesis starts from the 5' end of the RNA, and an intermediate DNA/RNA hybrid molecule is formed (Freed 2001). After the entire process, a double-strand DNA is synthesized and the nuclear import takes place.

### III. Integration.

The newly formed viral DNA is integrated into the host genome. This reaction is catalyzed by the viral protein IN. This integrated DNA or provirus may stay latent for several years.

### IV. Gene expression.

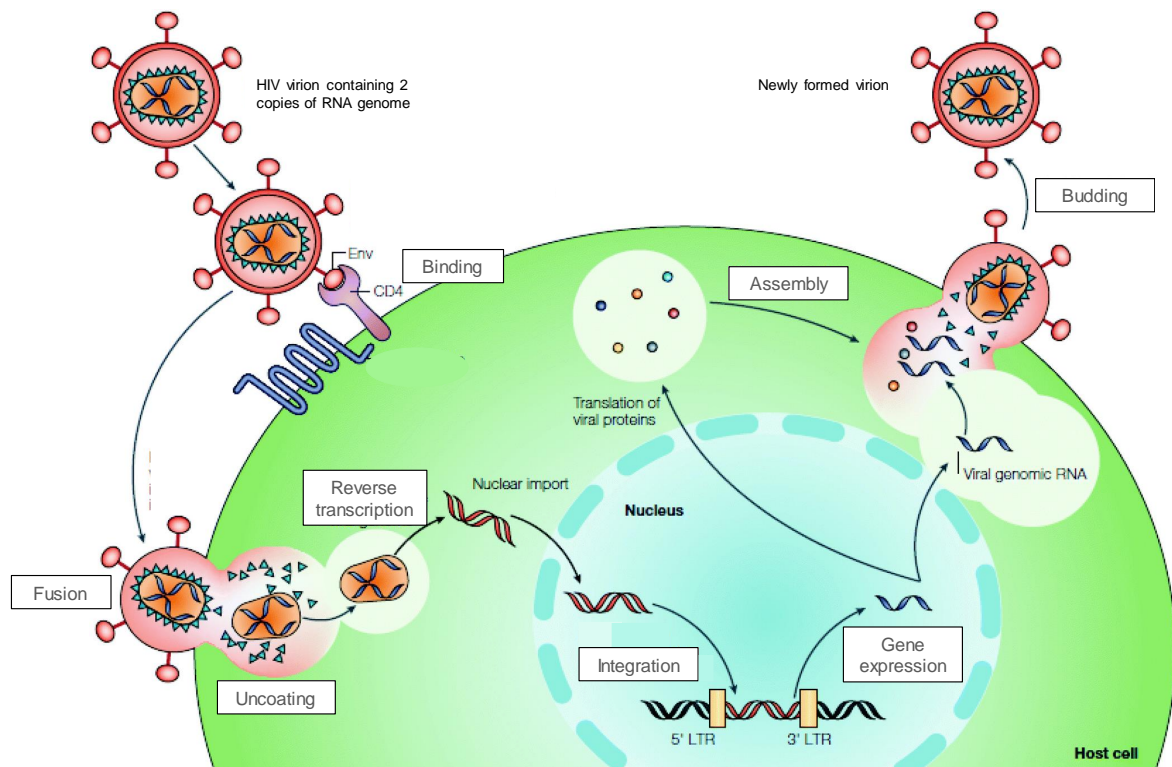
The non-coding 5'-LTR serves as transcription initiation. The transcription process is the step where the mRNAs for the different proteins and the genomic RNA are synthesized. It has been described that the LTR transcription is very low (Freed 2001) and the presence of the viral protein Tat increases the RNA synthesis (Dayton 1986; Fisher et al. 1986). After transcription, the mRNAs are transported to the cytoplasm, where translation and the production of viral proteins are performed.

### V. Assembly.

After the production of all viral proteins, the assembly process begins. It takes place at the plasma membrane of the host cell. This step consists on the assemble of all the proteins into immature HIV particles. Here, the genomic RNA is also encapsidated.

### VI. Budding.

The final step of the life cycle in which the immature particles bud out from the host cell plasma membrane. Soon after budding, the PR acts cleaving gag-pol precursors and maturation is finished.



**Figure 3. Schematic illustration of the HIV-1 replication cycle.** The virus binds the CD4 receptor and then the fusion of the membranes takes place. After the uncoating of the viral core, the reverse transcription of the viral genome begins. The newly synthesized double stranded-DNA enters the nucleus and integrates in the host cell genome. Generation of the viral mRNA occurs in the nucleus, and production of the proteins takes place in the cytoplasm, after release of the mRNAs. Still inside the cytoplasm the assembly of all viral particle occurs. The final steps are budding and maturation of the new virions. Adapted from Rambaut *et al.* (Rambaut *et al.* 2004)

### *Temporal regulation of the genes*

The HIV has a temporal expression of its genes that can be divided into two phases (Figure 4), the early and the late phases.

#### Early phase

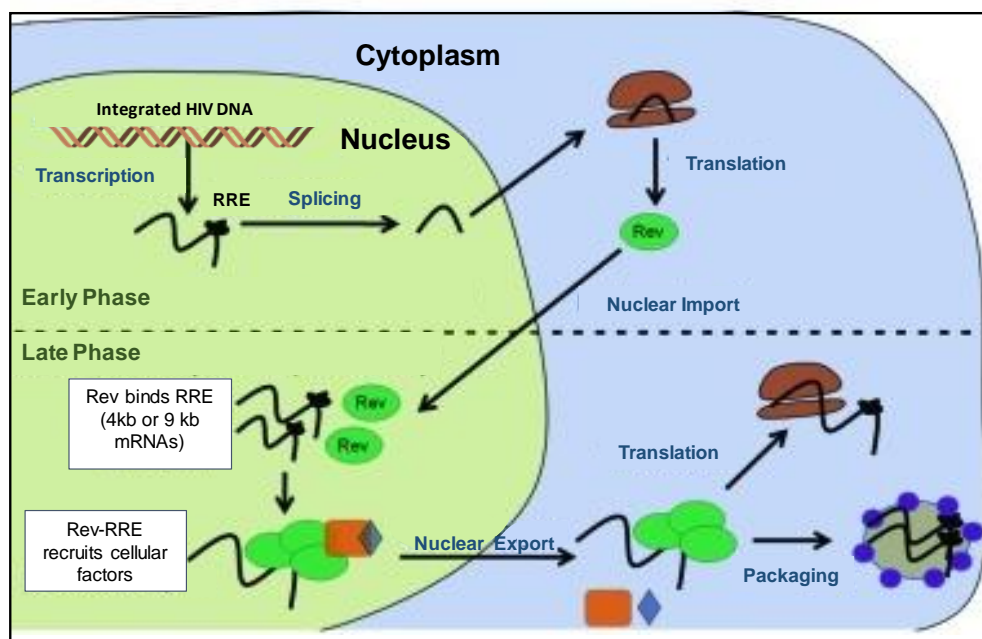
To synthesize all the proteins for the formation of new virion particles, the production of all the corresponding mRNAs is necessary. Thus, the primary transcripts of the HIV-1 suffer alternative splicing (Karn and Stoltzfus 2012). Among the mRNAs we can find unspliced mRNAs (~9kb) which correspond the genomic RNA and/or the Gag-Pol mRNA; single spliced mRNAs (~4kb) which correspond to Vif, Vpr and Env/Vpu mRNAs; and completely spliced mRNAs (~1.8kb) which correspond to Tat, Rev and Nef mRNAs (Bohne, Wodrich, and Kräusslich 2005).

## Introduction

From the beginning of transcription, all three types of mRNAs are expressed at both early and late phases in the nucleus of the host cell (Karn and Stoltzfus 2012). All these mRNAs must be transported to the cytoplasm to be translated into proteins. However, in the early phase, the completely spliced mRNAs (~1.8kb) are the only capable to be exported to the cytoplasm using a cellular pathway (Rev-independent RNAs lacking an intron) (Bohne et al. 2005). The bigger mRNAs (9kb and 4kb) are degraded or spliced in the nucleus in the early phase.

### Late phase

When enough amount of Rev is produced in the cytoplasm, Rev can enter the nucleus by binding to a nuclear import factor. Inside the nucleus, Rev protein joins to the RRE of the unspliced and single spliced mRNAs (Malim et al. 1989) (Rev-dependent RNAs containing at least one functional intron) (Bohne et al. 2005). This joining leads to a change in conformation that allows the binding of more monomers of Rev (Malim and Cullen 1991; Mann et al. 1994; Zapp et al. 1991) and cellular factors, creating a complex with a high ability to transport the mRNAs out of the nucleus, which occurs via the nuclear pore complexes (Karn and Stoltzfus 2012). Once the larger mRNAs are in the cytoplasm, their translation begins and the viral replication cycle continues.

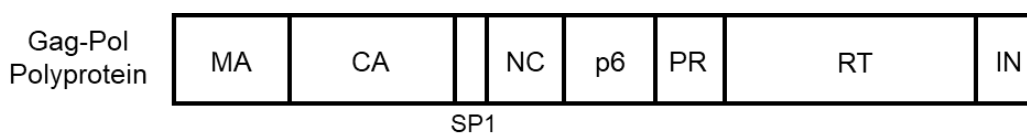


**Figure 4. Temporal expression of the HIV-1 genes.** In the early phase, the small mRNAs exit the nucleus to be translated in the cytoplasm. Transinductors are produced in this phase. When enough amount of Rev is produced, this protein suffers a nuclear import and joins to the RRE of the bigger

mRNAs (4kb and 9 kb). This is when the late phase begins. The joining of Rev to RRE attracts cellular factors that facilitate the exit of these mRNAs to be translated in the cytoplasm. Adapted from Fernandes *et al.* (Fernandes, Jayaraman, and Frankel 2012).

## HIV-1 protease

The HIV-1 protease (PR) is an aspartic protease responsible for the maturation of the newly formed virion. Gag and Gag-Pol polyproteins are hydrolyzed by the PR to produce the mature structural and functional proteins of the virus (Darke *et al.* 1988; Dunn *et al.* 2002). Without PR action, the virus life cycle is not completed and the virions do not mature. PR protein is synthesized as part of the Gag-Pol polyprotein, which is also the precursor of other proteins (Figure 5). In the Gag-Pol precursor, PR is flanked by p6 and by the RT (Figure 5).



**Figure 5. HIV-1 proviral Gag-Pol polyprotein.** MA: matrix; CA: capsid; SP1: spacer peptide 1; NC: nucleocapsid; PR: protease; RT: reverse-transcriptase; IN: integrase.

The C-terminal and the N-terminal of PR tend to be very conserved, while the rest of the sequence shows sequence variability (Weber 1989). The mature PR is constituted by two 99-amino acids monomers which form an active homodimer. To become active, two Gag-Pol precursors need to dimerize. Proteolytic reactions are performed by the Gag-Pol polyprotein itself to release the mature PR in a process called autoprocessing (Huang and Chen 2013; Pettit *et al.* 2004). Following autoprocessing, the mature PR is ready to catalyze the reactions for the rest of the viral proteins.

Among the different steps and proteins involved in the HIV-1 cycle, PR has been a main target for anti-HIV drugs due to its role in viral maturation (Debouck 1992; Eder *et al.* 2007; Huang and Chen 2013; Louis *et al.* 2000; Prejdová, Soucek, and Konvalinka 2004). As a result, from the 26 anti-HIV compounds approved by the US FDA, 10 are PR inhibitors (PIs) (Table 1) (Lv, Chu, and Wang 2015; Wensing, van Maarseveen, and Nijhuis 2010). All approved PIs have similar structure and binding pattern. The mechanism they use is to bind to the PR with high affinity, competing with the natural substrates. Consequently, virion maturation is interrupted and non-infectious particles are formed.



## ***Introduction***

---

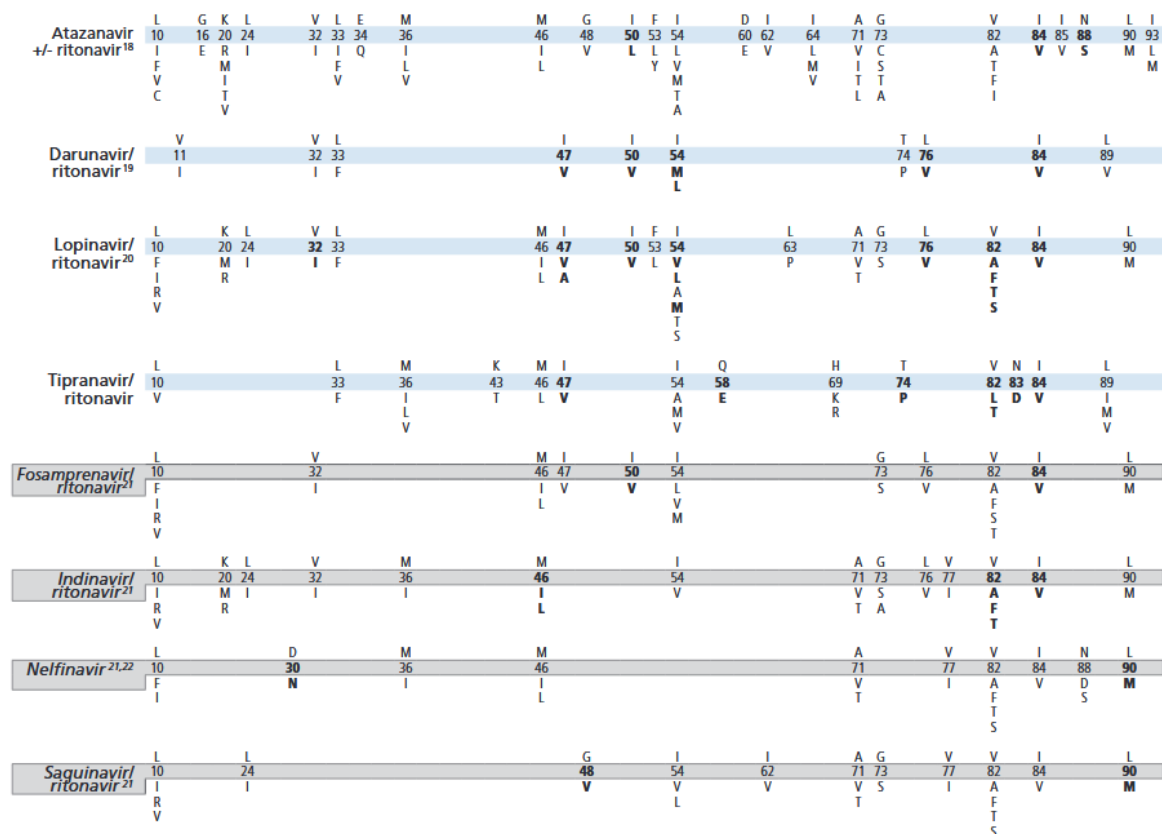
The most common and effective treatment against HIV is known as highly active antiretroviral therapy (HAART), and it is based on prescribing three different drug molecules that in some cases include PIs (Lv et al. 2015). Indeed, PIs can be one of the most important drugs within the combination of HAART compounds.

**Table 1. Protease inhibitors approved by the US FDA as anti-HIV drugs.**

<b>Protease inhibitor</b>	<b>Brand name</b>	<b>FDA approval</b>
Saquinavir mesylate	Invirase	1995
Indinavir	Crixivan	1996
Ritonavir	Norvir	1996
Nelfinavir	Viracept	1997
Amprenavir (no longer marketed)	Agenerase	1999
Lopinavir	Kaletra	2000
Atazanavir	Reyataz	2003
Fosamprenavir	Lexiva	2003
Tipranavir	Aptivus	2005
Darunavir	Prezista	2006

Resistance mutations have been described for all PIs (Figure 6). The generation of viral resistance mutations benefits from the absence of proofreading activity of the viral RT. Since developing protease resistance mutations is a stepwise process (Kaplan et al. 1994; Molla et al. 1996), major and minor mutations can be found. Usually, the first resistance mutations to PIs are generated near the substrate-binding cleft of the viral protease. As a consequence, the structure of the substrate-binding cleft is altered and the high binding affinity to PIs is reduced (Chen et al. 1995; Gulnik et al. 1995; Hong et al. 1997; Mahalingam et al. 1999). These mutations are known as major resistance mutations. Minor mutations are those developed later. These minor changes do not confer resistance by themselves. However, when found with major mutations, the replication capacity of the virus is improved (Borman, Paulous, and Clavel 1996; Carrillo et al. 1998; Nijhuis et al. 1999). When major mutations are generated and the structure of the substrate-binding cleft is changed, the replication capacity of the virus is reduced as a consequence of a reduced capacity to cleave the PR natural substrates. It is believed that compensatory mutations, probably minor mutations, are generated to either increase the resistance or increase the replication capacity of the virus (Borman et al. 1996; Mammano et al. 2000; Nijhuis et al. 1999).

To avoid viral escaping as a consequence of the generation of resistance mutations, PI levels are increased in plasma after administration. Accordingly, most PIs are co-administered with a low dose of ritonavir (Kempf et al. 1997), an inhibitor of the cytochrome P450-mediated metabolism of PIs.



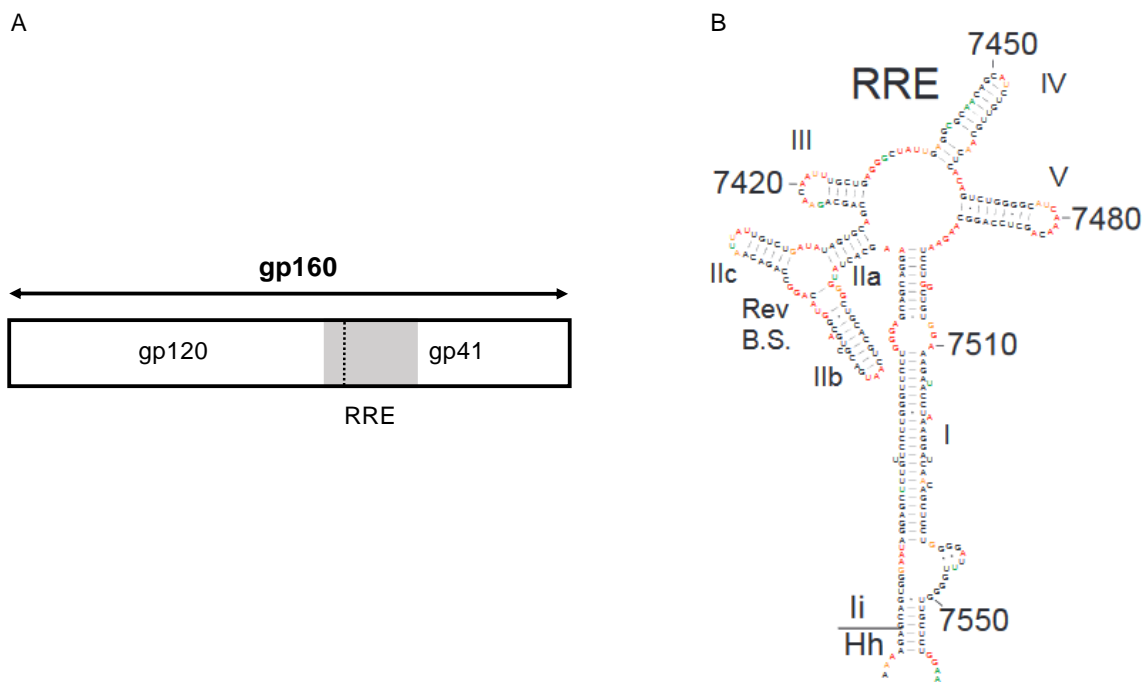
**Figure 6. Mutations in the PR gene associated with resistance to protease inhibitors.** Adapted from Wensing and colleagues (Wensing et al. 2019).

## HIV-1 envelope

The HIV-1 envelope protein joins to the CD4 cell receptor to start the infection cycle. Env is synthesized as the precursor gp160, which is proteolytically cleaved into gp120 and gp41. This protein suffers high glycosylation while it is being produced, developing a glycan shield that makes it one of the most glycosylated proteins in nature (Mao et al. 2012). Eventually, this glycan shield is one of the weapons that the virus uses to evade the immune system, since only 2% of the surface is recognized by antibodies (Pancera et al. 2014).

*Rev Response Element (RRE)*

The RRE is a conserved structure of about 350 nucleotides (nts) located in the Env gene. Rev binds to the RRE with high affinity (Cullen 1991) to transport the largest mRNA to the cytoplasm. The extremely conserved sequence of the RRE is located from position 7709 to position 8063 in the HXB2 genome, thus beginning at the end of gp120 with the vast majority of the sequence located in gp41 (Figure 7 A).



**Figure 7. HIV-1 RRE location and structure.** A) The RRE is located between positions 7709-8063 of the HXB2 strain, beginning its sequence about 50 nucleotides before the end of gp120. The RRE location is shown in grey. B) RNA secondary structure of the RRE. Five loops are shown in this structure. Adapted from Watts *et al.* (Watts *et al.* 2009).

The RNA secondary structure of the RRE is also very conserved (Figure 7 B), formed by several loops and bulges (Mann *et al.* 1994; Watts *et al.* 2009). Rev does specifically join to these loops of the RRE. Previous studies described how the RRE forms a five-stem loop in which the 5' end and the 3' end base-pair to include the main structure in the middle (Dayton, Powell, and Dayton 1989; Kjems *et al.* 1991; Watts *et al.* 2009). However, recent studies show how the RRE can adopt different conformations and form secondary structures with four loops, although the four-loop conformation promotes worse functional Rev/RRE activity compared to the five-stem loop (Sherpa *et al.* 2015).

## Codon usage

To produce the proteins of any organism, mRNAs need to be translated. To perform that process, nature has selected a genetic code which is common to all organisms. The genetic code is formed by four bases. The combination of three bases form the codons. In total, there are 61 coding codons and 3 stop codons. The 61 coding codons are translated to a total of 20 amino acids. Some of these amino acids are encoded by up to six different codons, meaning that the genetic code is redundant (Figure 8). The codons that encode for the same amino acid are synonymous. Normally, codons encoding for the same amino acid differ in the third base.

		Second base							
		U	C	A	G				
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U C A G
	UUC		UCC		UAC		UGC		
	UUA	Leu	UCA	UAA	STOP	UGA	STOP		
	UUG		UCG	UAG	STOP	UGG	Trp		
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U C A G
	CUC		CCC		CAC		CGC		
	CUA	Leu	CCA	CAA	Gln	CGA	Arg		
	CUG		CCG	CAG		CGG			
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U C A G
	AUC		ACC		AAC		AGC		
	AUA	Met	ACA	AAA	Lys	AGA	Arg		
	AUG		ACG	AAG		AGG			
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U C A G
	GUC		GCC		GAC		GGC		
	GUA	Val	GCA	GAA	Glu	GGA	Gly		
	GUG		GCG	GAG		GGG			

Ala = alanine	Gln = glutamine	Leu = leucine	Ser = serine
Arg = arginine	Glu = glutamate	Lys = lysine	Thr = threonine
Asn = asparagine	Gly = glycine	Met = methionine	Trp = tryptophan
Asp = aspartate	His = histidine	Phe = phenylalanine	Tyr = tyrosine
Cys = cysteine	Ile = isoleucine	Pro = proline	Val = valine

**Figure 8. Standard genetic code.** Only methionine and tryptophan are encoded for just one codon. There are three stop codons in the genetic code.

In a gene or a genome, not all codons are used with the same frequency. For the same amino acid, some codons are more widely used than others and, in contrast, other codons appear less frequently than expected. This phenomenon is called codon usage bias and it is species-specific. For example, the codon usage from *Escherichia coli* (*E. coli*) is different from that of the human.

A plausible explanation for the codon usage bias is the translational pressure. This consists on the correlation of the codon usage with the abundance of isoaccepting tRNAs, and this correlation between them relates with protein production levels of the genes (Duret 2000; Ikemura 1981, 1985; dos Reis, Savva, and Wernisch 2004). This hypothesis is based on the supply of tRNAs, which can lead into a slower translation elongation which reduces translation rates. Nevertheless, other studies did not corroborate these facts (Gustafsson, Govindarajan, and Minshull 2004) and several works have helped to unravel why codon usage affects translation or other mechanisms (Brule and Grayhack 2017). It has been described that codon usage can influence transcription modifications (Findlay et al. 2018; Zhou et al. 2016), translation initiation, elongation or accuracy (Akashi 1994; Boël et al. 2016; Goodman, Church, and Kosuri 2013; Sørensen, Kurland, and Pedersen 1989), RNA stability (Presnyak et al. 2015), RNA structure and folding (Presnyak et al. 2015; Shabalina, Ogurtsov, and Spiridonov 2006), RNA splicing (Takata et al. 2018), etc.

Viral quasispecies, also called viral mutant clouds, are the set of genetically variable populations generated after replication of RNA viruses. Genetic variability is generated by accumulation of mutations during replication and by genome recombination (Andino and Domingo 2015). RNA virus accumulation of mutations is favored by their high mutation rate (Domingo et al. 1978). It has been suggested that the incorporation of high rates of mutations help creating mutant clouds where beneficial mutations provide a greater probability to adapt to new environments (Coffin 1995). Accumulation of mutations in RNA viruses is due to the lack of proofreading activity of their polymerases, generating between  $10^{-3}$  to  $10^{-5}$  mutations per nucleotide in one replication cycle (Más et al. 2010). How codon usage may affect virus evolution and population diversity in absence of inhibitors has recently been studied. To that end, poliovirus was synonymously recoded, and the quasispecies of the recoded and WT viruses were compared. Both viruses explored different sequence spaces and generated different mutant clouds (Lauring et al. 2012). Thus, codon choice is also influencing genetic variability of RNA viruses.

The HIV-1 virus does also have a high rate of mutations and quasispecies are found in viral infections. Interestingly, although HIV-1 has a high mutation rate, its nucleotide composition has maintained stable over the decades (van der Kuyl and Berkhout 2012). This virus has a nucleotide composition bias based on an adenine (A)-rich, cytosine (C)-low genome (Kypr and Mrázek 1987). Although the total number of Cs and guanidines (G) together is not low, the total amount of CpG dinucleotides in the HIV-1 is very reduced (Bronson and Anderson 1994; Kypr, Mrázek, and Reich 1989). On the contrary, GpC amounts are not suppressed. This phenomenon is also observed in other RNA viruses (Rima and McFerran 1997).

The reasons why the HIV-1 has an A-rich genome have not been elucidated, although evolutionary pressure and mutational rate by the viral polymerase have been pointed (van der Kuyl and Berkhout 2012). Similarly, studies about the CpG suppression are not decisive. It has been suggested that CpG suppression in viral genomes is due to avoid DNA methylation by the host cellular machinery (Shpaer and Mullins 1990). However, analyses in other viral genomes revealed that small virus genomes (<30kb) are CpG suppressed, while large virus genomes are not (Karlin, Doerfler, and Cardon 1994). CpG suppression is thus observed in RNA viruses without a DNA intermediate, so DNA methylation must not be the only reason why low CpG amounts are observed.

### *HIV-1 genome composition and innate immune system*

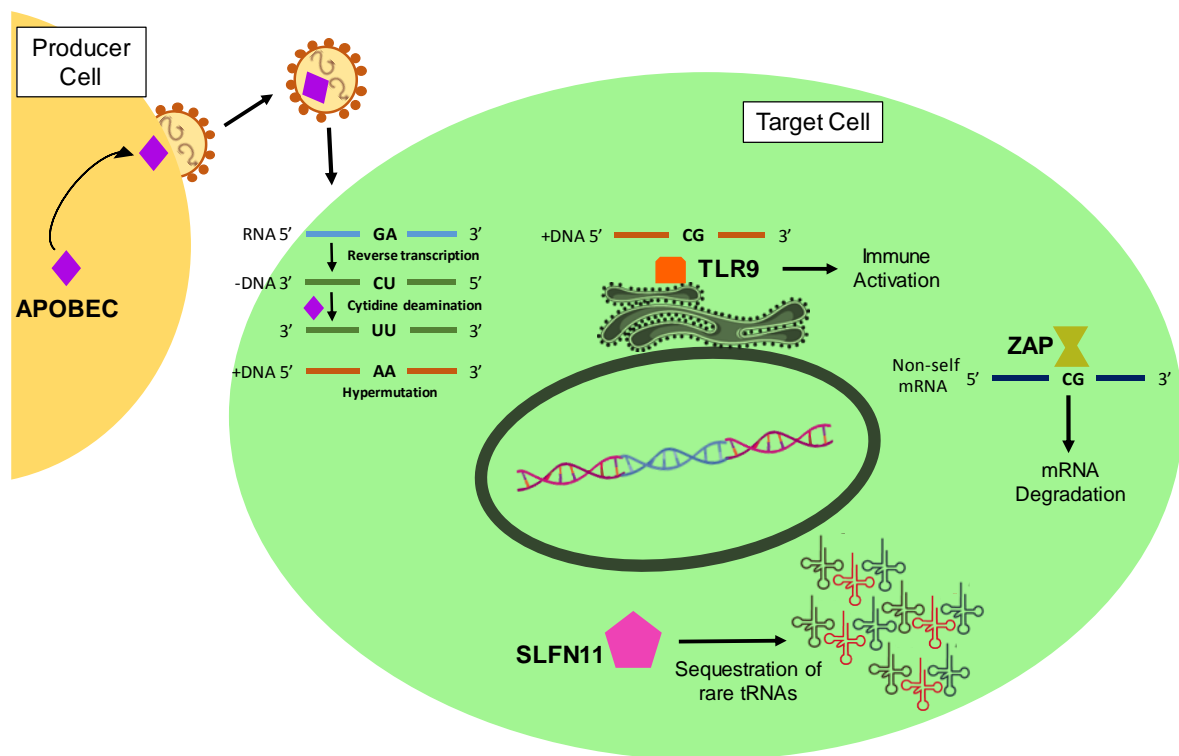
Viral sequences and structures can be recognized by proteins of the innate immune system. Thus, the nucleotide composition of a virus is important to escape from the immune recognition. Several factors of the innate immune response have been described as possible causes for the nucleotide and dinucleotide bias of the HIV-1 (van der Kuyl and Berkhout 2012).

APOBEC proteins are part of the innate immune system. These proteins catalyze cytidine deamination in the HIV-1 single-stranded proviral DNA in the reverse transcription process (Anderson and Hope 2008; Suspène et al. 2004). The result is G-to-A mutations in the newly synthesized viral DNA. It has been described how, in the absence of Vif, APOBEC leads to hypermutagenesis in the viral genome that, finally, lead into the formation of non-infectious virions (Gaddis et al. 2003; Lecossier et al. 2003). This might explain why the HIV-1 genome is A-enriched (Figure 9). However, it has been described how a single “APOBEC-unit” inside the HIV-1 particle would hypermutate the viral genome, in an “all or nothing” phenomenon (Armitage et al. 2012). Thus, it seems not likely that APOBEC is fully responsible for the A-richness in the HIV-1 genome and other factors may be affecting.

Schlafen 11 (SLFN11) is a protein induced via the interferon regulatory factor 3 (IRF3). It has been described that SLFN11 inhibits HIV-1 in a codon-usage dependent manner (Li et al. 2012) by sequestering rare tRNAs induced by HIV-1 and altering the tRNA pool composition (Figure 9). However, it was demonstrated that HIV-1 Vpu depletes IRF3 (Doehle et al. 2012) and, consequently, SLFN11 is not induced. HIV-1 does not have to escape from this immune factor, hence, SLFN11 does not explain the nucleotide bias of the HIV-1.

Toll-like receptors (TLRs) are glycoproteins mainly located in the cell surface, but also in cell organelles, that recognize pathogen-associated molecular patterns (PAMPs) from different microorganisms. Specifically, TLR9 recognizes unmethylated CpGs in DNA sequences (Pandey, Kawai, and Akira 2015) and it is highly expressed in the endoplasmic reticulum of plasmacytoid dendritic cells (pDCs) (Kumagail, Takeuchi, and Akira 2008) (Figure 9). It has been shown that it is probable that HIV-1 does not replicate in pDCs, thus it is not possible that CpG suppression is caused by TLR9. However, it has been suggested an influence of TLR9 in past viral ancestors (van der Kuyl and Berkhout 2012) and the present HIV-1 might have inherited that tendency.

Recently, it has been described that zinc-finger antiviral protein (ZAP) inhibits HIV-1 replication capacity by targeting CG-rich RNAs and promoting their degradation (Takata et al. 2017). When CpGs were added to the viral genome, the replication capacity of the virus decreased. In cells where ZAP was knockout, the same defect viruses recovered their fitness. Moreover, ZAP targets only non-self RNAs (Figure 9). CpG suppression in HIV-1 genome may be a strategy of the virus to escape from the recognition of ZAP. However, more studies need to unravel this hypothesis.



**Figure 9. Factors of the innate immune system that might affect nucleotide composition of the HIV-1 genome.** APOBEC induces cytidine deamination that ultimately leads in hypermutagenesis. SLFN11 sequesters tRNAs in a codon-dependent manner. TLR9 recognizes

unmethylated CpGs in DNA and activates the immune cascade. Finally, ZAP recognizes CpG-rich non-self RNAs and induces their degradation.

## Codon pair usage

As mentioned before, the existence of a codon usage bias is known and is species-specific. Similarly, the frequency of appearance of two adjacent codons is not random, and this is known as codon pair usage. It was observed in *E.coli* that the frequency in which codons are found juxtaposed is not random as compared to the frequency expected if those two codons were randomly placed, independently of the codon usage. In addition, the codon pair usage differs between proteins that are more or less expressed (Gutman and Hatfield 1989).

Not only it has been observed that the frequency of appearance of two adjacent codons is not random, but also some codon pairs are more or less used than other synonymous ones. Analyzing a total of 14795 human genes, it was observed that the frequency of appearance of a specific codon pair was higher or lower than expected (Coleman et al. 2008). For example, the dipeptide Ala-Glu can be encoded by the combination of the synonymous codons of each amino acid. Theoretically, the expected frequency for the codon pair GCCGAA should be the same as the expected frequency of GCAGAG. However, the observed frequency of GCCGAA is very low, even though it contains the most frequent Ala codon. The codon pair GCAGAG is used seven times more than GCCGAA (Martínez et al. 2016) (Figure 10 A).

A	<table border="1"> <thead> <tr> <th>Ala</th> <th>-</th> <th>Glu</th> </tr> </thead> <tbody> <tr><td>GCA</td><td>-</td><td>GAA</td></tr> <tr><td>GCA</td><td>-</td><td>GAG</td></tr> <tr><td>GCG</td><td>-</td><td>GAG</td></tr> <tr><td>GCT</td><td>-</td><td>GAG</td></tr> <tr><td>GCT</td><td>-</td><td>GAA</td></tr> <tr><td>GCC</td><td>-</td><td>GAG</td></tr> <tr><td>GCG</td><td>-</td><td>GAA</td></tr> <tr><td>GCC</td><td>-</td><td>GAA</td></tr> </tbody> </table>	Ala	-	Glu	GCA	-	GAA	GCA	-	GAG	GCG	-	GAG	GCT	-	GAG	GCT	-	GAA	GCC	-	GAG	GCG	-	GAA	GCC	-	GAA	B
Ala	-	Glu																											
GCA	-	GAA																											
GCA	-	GAG																											
GCG	-	GAG																											
GCT	-	GAG																											
GCT	-	GAA																											
GCC	-	GAG																											
GCG	-	GAA																											
GCC	-	GAA																											

**Observed frequency** of GCCGAA in the human genes= 1097

**Expected frequency** of GCCGAA:

$$\frac{(\text{GCC frequency} \times \text{GAA frequency})}{\text{Ala frequency} \times \text{Glu frequency}} \times \text{Ala - Glu frequency} = 6109,5$$

**Figure 10. Codon pair score and codon pair bias.** A) The codon pair GCCGAA was used as an example. Both observed and expected frequencies in the human genes were calculated. B) The codon pair score (CPS) is the log of the observed over the expected frequency, and it indicated if a



codon pair is under or overrepresented in the human genome. Negative values indicate underrepresentation, while positive values indicate overrepresentation. Codon pair bias is the mean of individual CPS of a gene. Adapted from Coleman and colleagues (Coleman et al. 2008).

An algorithm was developed as a measurement of codon pair bias (Coleman et al. 2008). To that end, each codon pair has a calculated codon pair score (CPS). There are 3721 possible codon pairs and each of them has a CPS based on the frequencies observed and expected in the human genes (Figure 10 B). Similarly, the codon pair bias (CPB) of an entire gene is the arithmetic mean of the individual CPS forming that gene. When the CPB value is negative, a predominant number of underrepresented codon pairs constitute the gene. On the contrary, a positive CPB means that the gene is formed by more overrepresented codon pairs.

What determines the codon pair usage is still controversial. It is known that two tRNAs need to accommodate in the ribosome together in sites A and P during translation. A correlation was demonstrated between codon pair usage and the interaction of the nucleotides of the codon-anticodon in P- and A-sites of the tRNA-ribosome complex (Buchan, Aucott, and Stansfield 2006). Moreover, CpG dinucleotides are underrepresented in humans and, interestingly, most underrepresented codon pairs generate CpGs between the third nucleotide of the first codon and the first nucleotide of the second codon. As described above, CpG dinucleotides can be recognized by several immune factors (Pandey et al. 2015; Takata et al. 2017). As a consequence, codon pair usage cannot be distinguished from dinucleotide frequencies (Futcher et al. 2015; Simmonds et al. 2015; Tulloch et al. 2014). Nevertheless, this and the hypothesis of the codon-anticodon interaction (Buchan et al. 2006) are both compatible as possible explanations for the effect.

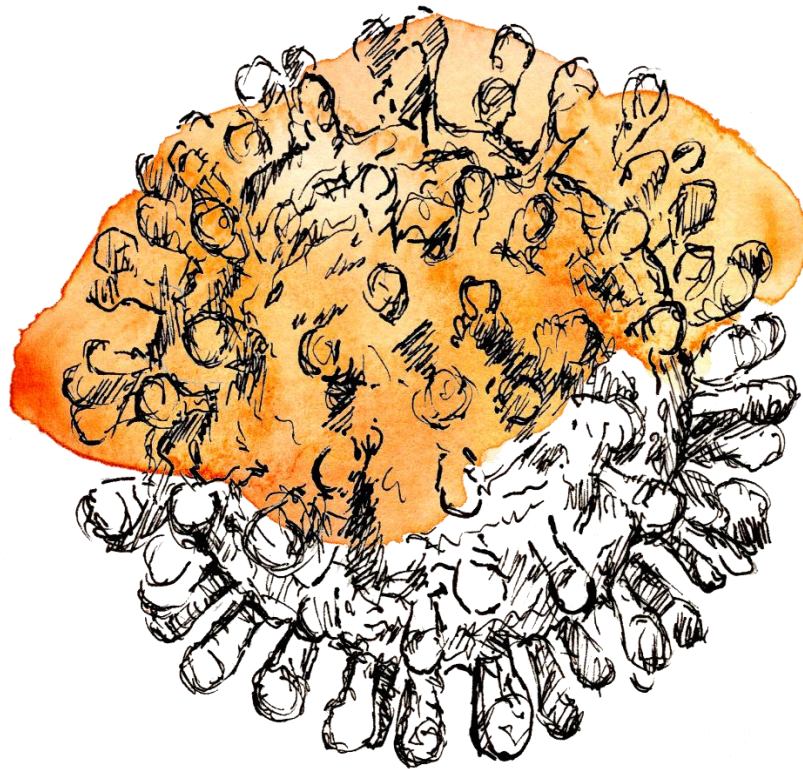
## **Synthetic synonymous genome recoding**

Over the years, gene synthesis has become a widely used tool in biological research. Among the techniques used to generate synthetic genes there is one called genome recoding, in which codons or codon pairs of an organism can be synonymously changed. In this process, codons or codon pairs are changed without altering the amino acid sequence, thus allowing to explore the effect of nucleotide changes. Not only codons and codon pairs can be changed, but also dinucleotides frequencies, maintaining the amino acid sequence.

The vast majority of synthetic recoding has been performed on viral genomes, due to their smaller size (~10kb). Some of the first works using genome recoding were performed on poliovirus, demonstrating how codon deoptimization attenuates the virus replication capacity (Burns et al. 2006; Mueller et al. 2006). The number of mutations introduced was between 500 and 700, making it more difficult for the virus to revert the mutations and, thus, to revert the attenuated phenotype. Since then, several works have been performed changing codon usage and codon pair usage in several RNA and DNA viruses. In all of them, attenuation was observed (Cladel et al. 2013; Martrus et al. 2013; Mueller et al. 2006, 2010; Le Nouen et al. 2014; Vabret et al. 2014; Yang et al. 2013). Recently, altering the CpG/TpA frequencies in several viruses has also demonstrated the capacity of attenuation of the technique (Atkinson et al. 2014; Burns et al. 2009; Gaunt et al. 2016; Tulloch et al. 2014).

In addition to viral attenuation, synonymous genome recoding has allowed to study other features of the viral biology. Synonymous changes performed in poliovirus without altering the codon usage nor the codon pair bias led into the detection of two functionally redundant RNA elements (Song et al. 2012). The temporal regulation of viral gene expression was studied when synonymous mutations were performed on SIV Env gene changing the codon usage to that of the herpesvirus. It was shown how the codon usage limits the expression of late genes (Shin et al. 2015). The influence of a sequence space on viral spectrum, evolutionary trajectory and pathogenicity in absence of viral inhibitors was studied by synonymously recoding the capsid of poliovirus (Lauring et al. 2012). Finally, this technique has been very useful to describe factors involved in the innate immune response. The previously mentioned SLFN11 and ZAP were both described after synonymously altering the codon usage or the amount of CpG dinucleotide, respectively, in HIV-1 genome (Li et al. 2012; Takata et al. 2017).





## Hypothesis and Objectives



Synonymous genome recoding is a powerful and widely used tool to change the nucleotide sequence of an organism without changing the amino acid sequence. Different approaches have been made to synonymous change the different genes or the genome of several organisms: to change the codon usage, to change the codon pair usage or the codon-triplet usage, to randomly alter the codons, and to increase or decrease the frequency of specific dinucleotides (e.g. CpGs/TpAs). This tool has made possible expanding the knowledge on some viral fields, such as the interaction with the immune system, temporal regulation of viral genes, viral adaptability, describing viral genome functional redundant RNA elements, among others.

We aimed here to synonymous recode the protease and the envelope gene of the human immunodeficiency virus type 1 (HIV-1) and to see how virus evolvability and replication capacity is influenced by the codon usage or codon pair usage.

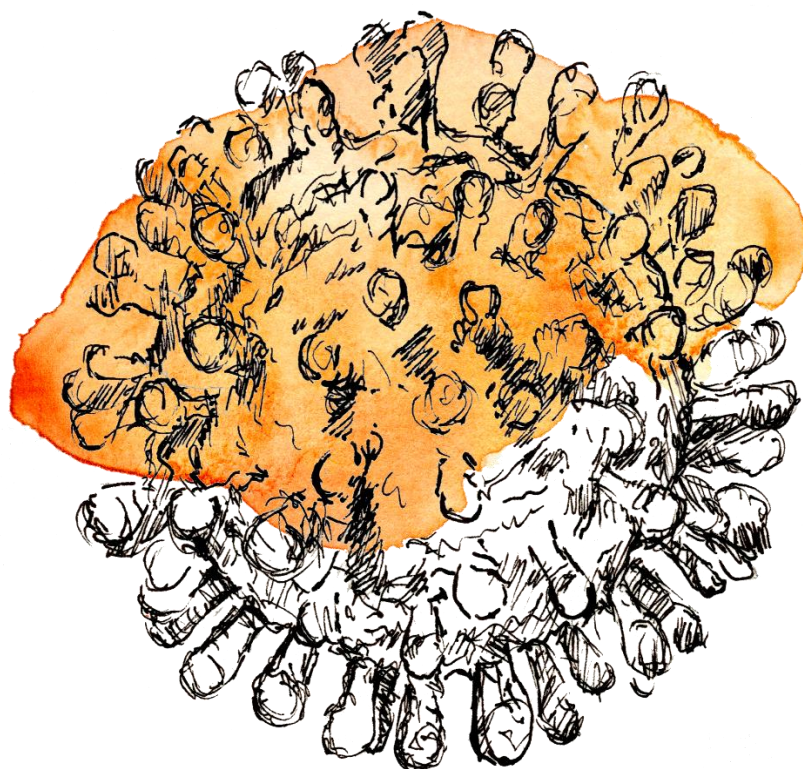
**Objective 1:** To explore how synonymous HIV-1 genome recoding affects both drug resistance development and viral evolutionary capacity.

**Objective 2:** To explore the impact of synonymous codon usage and synonymous codon pair usage on HIV-1 envelope expression and virus replication capacity.

**Objective 3:** To study the impact of codon usage and codon pair usage in RNA folding.

**Objective 4:** To study how increasing the number of CpG dinucleotides affects virus phenotype.





## **Materials and Methods**

---





## **Cell lines**

The cells were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH. MT-4 cells (Gyuris, Vajda, and Földes 1991) were propagated in Roswell Park Memorial Institute 1640 medium (RPMI, Gibco) supplemented with 10% of heat inactivated fetal bovine serum (FBS, Gibco).

293T cells were obtained from the American Type Culture Collection (ATCC) (DuBridge et al. 1987; Pear et al. 1993). 293T cells were grown in Dulbecco's Modified Eagle's medium (DMEM, Gibco) supplemented with 10% of heat inactivated FBS.

TZM-bl is an adherent cell line derived from HELA cells. To generate this cell line, luciferase and  $\beta$ -galactosidase genes were introduced under control of the HIV-1 promoter. TZM-bl cells are highly sensitive to HIV infection by adding DEAE-dextran in the medium. They were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (Derdeyn et al. 2000; Platt et al. 1998, 2009). The growing medium was DMEM supplemented with 10% of heat inactivated FBS.

Peripheral blood mononuclear cells (PBMCs) were obtained from "buffy coats" from healthy donors purchased anonymously from the Catalan Banc de Sang I Teixits (<http://www.bancsang.net/en/index.html>). To isolate the desired fraction of PBMCs, a Ficoll-Paque density gradient centrifugation was performed. The cells were maintained in RPMI supplemented with 20% of heat inactivated FBS and quantified. PBMCs from four different donors were mixed with the same number of cells. From the mixture, aliquots of  $15 \times 10^6$  cells were frozen and stored for later use. After thaw, cells were stimulated and maintained in RPMI supplemented with 20% of heat inactivated FBS.

All cell lines were incubated at 37°C and 5% CO<sub>2</sub>.

## **Plasmids**

The pNL43-dE-EGFP was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (Zhang et al. 2004). This plasmid is an HIV-1 reporter virus with a truncated envelope gene. It contains the gene of enhanced green fluorescent protein (EGFP) in the envelope open reading frame.

The pcDNA3.1D/V5-His-TOPO (Invitrogen) was used to clone the envelopes and used as an expression vector. To clone the envelopes in the vector, these were amplified using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen) and the primers ENV-TOPO-SD

## **Materials and Methods**

---

and env\_pcdna3.1R (Table 2). The ENV-TOPO-SD primer included in its sequence the corresponding splicing donor needed for the expression of the envelopes (Chang and Sharp 1989; Lu et al. 1990). The PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 3 min. All PCRs were purified using QIAquick PCR Purification Kit (Qiagen) and eluted in 50µl of milli-Q water. Briefly, 19 ng of the PCRs were cloned and transform following the TOPO manufacturers' instructions. Positive clones were amplified using *Taq* DNA polymerase (VWR), the primers T7 and BGH and the PCR program: 1 cycle of 94°C for 10 min; 35 cycles of: 94°C for 45 sec, 55°C for 45 sec and 72°C for 3 min; and 1 cycle of 72°C for 10 min. The primers used to sequence the positive clones were T7, BGH and hxb2\_RRE7610-Rv (Table 2).

The pcDNA3.1D/V5-His-TOPO was also used to clone the viral gene Rev and use it as an expression vector (pTOPORev from now on). For the purpose of detecting Rev expression in further analyses, the codon STOP at the end of the gene was not introduced. By doing this, Rev gene is continued by V5 of the pcDNA3.1D/V5-His-TOPO. Briefly, *Rev* was amplified using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen) and the primers TOPO\_rev\_Fw and TOPO\_rev\_Rv (Table 2). The PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 45 sec. Positive clones were amplified using *Taq* DNA polymerase (VWR), the primers T7 and BGH and the PCR program: 1 cycle of 94°C for 10 min; 35 cycles of: 94°C for 45 sec, 55°C for 45 sec and 72°C for 1 min; and 1 cycle of 72°C for 10 min. The primers used to sequence the positive clones were T7 and BGH.

The pDrive cloning vector (Qiagen) was used to clone different fragments from the NL4-3 HIV-1 genome. The amplifications before the cloning were performed using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen), purified by QIAquick PCR Purification Kit (Qiagen) and eluted in 50 µl of milli-Q water. The fragment that extends from the beginning of the genome to the beginning of the envelope was amplified from pNL4-3 using the oligonucleotides pNL43-ltr1-Fw and pNL43-6306-Rv (Table 2). The PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 6 min and 30 sec. The fragment that extends from the end of the envelope to the end of the genome was amplified from pNL4-3 using the oligonucleotides hxb2\_env8369-Fw and pNL43-9709-Rv (Table 2). The PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min and 30 sec. After purification, the fragments were cloned in pDrive following the manufacturer's instructions. Positive clones were amplified with VWR *Taq* DNA polymerase (VWR), using the oligonucleotides T7 and SP6. The primers used to sequence the 1-6306 fragment were T7, pNL43-ltr1-Fw,

pnl43\_769\_Rv, pnl43\_566\_Fw, pnl43\_pol2329\_Rv, pnl43\_pol2134\_Fw, pnl43\_vif5339\_Rv, pnl43\_vif5135\_Fw, pnl43\_vpu6112Fw and SP6 (Table 2). The primers used to sequence the 8369-9709 fragment were T7, hxb2\_env8369-Fw, pNL43-9709-Rv and SP6.

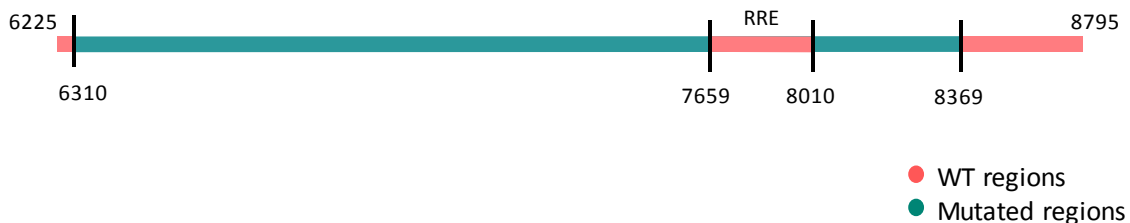
**Table 2. Oligonucleotides used in cloning, PCRs, sequencing and directed mutagenesis of the envelope variants.** Fw: forward primer; Rv: reverse primer.

Primer name	Sequence (5'-3')	Orientation	Method
ENV-TOPO-SD	CACCAACAAGTAAGTATGAGAGTGAAGGAGAAATAT	Fw	Cloning
env_pcdna3.1R	TCATAGCAAAATCCTTTCCAAAGCCC	Rv	Cloning
T7	TAATACGACTCACTATAGGG	Fw	PCR, Sequencing
BGH	TAGAAGGCACAGTCGAGG	Rv	PCR, Sequencing
hxb2_RRE7610-Rv	CTAGCATTCCAAGGCACAGC	Rv	Sequencing, overlapping PCR
TOPO_rev_Fw	CACCATGGCAGGAAGAAGCGGA	Fw	Cloning, sequencing
TOPO_rev_Rv	TTCCTTAGCTCCTGACTCC	Rv	Cloning, sequencing
pNL43-ltr1-Fw	TGGAAGGGCTAATTTGGTCC	Fw	PCR, sequencing
pNL43-6306-Rv	CTACAGATCATCAATATCCC	Rv	PCR
hxb2_env8369-Fw	ACCCACCTCCC AACCCCGAG	Fw	PCR
pNL43-9709-Rv	TGCTAGAGATTTTCCACACT	Rv	PCR
pNL43-6227-Fw	GAGAGTGAAGGAGAAGTATCAGC	Fw	PCR, overlapping PCR
pNL43-8766-Rv	GCAAAATCCTTTCCAAGCCCTGTC	Rv	PCR, overlapping PCR
hxb2_RRE7259-Fw	AGTAGCACCCACC AAGGCAAAAGAG	Fw	Overlapping PCR
env_1877_Fw	CGACCTGGATGGAGTGGGACAGAG	Fw	Site-directed mutagenesis
env_2120_Fw	GAGTTAGGCAGGGATATTCACC	Fw	Site-directed mutagenesis
env_1877_Rv	CTCTGTCCCACCTCCATCCAGGTCCG	Rv	Site-directed mutagenesis
env_2120_Rv	GGTGAATATCCCTGCCTAACTC	Rv	Site-directed mutagenesis
SP6	CATTTAGGTGACACTATAG	Rv	PCR, Sequencing
pnl43_769_Rv	CATCTCTCTCCTTAGCCTCCGC	Rv	Sequencing
pnl43_566_Fw	GTCTGTGTGTGACTCTGGTAAC	Fw	Sequencing
pnl43_pol2329_Rv	CTAATACTGTATCATCTGCTCCTG	Rv	Sequencing
pnl43_pol2134_Fw	CTTCAGAGCAGACCAGAGCC	Fw	Sequencing
pnl43_vif5339_Rv	GAATTAGTTGGTCTGCTAGGTCCAG	Rv	Sequencing
pnl43_vif5135_Fw	CAAGGAAAGCTAAGGACTGG	Fw	Sequencing
pnl43_vpu6112Fw	GCAATAGTTGTGTGGTCCATAG	Fw	Sequencing

### Gp160-based mutants

All the mutant envelopes were based on the HXB2 strain ([www.hiv.lanl.gov](http://www.hiv.lanl.gov)), whereas the rest of the genome of the HIV-1 was based on the NL4-3 strain ([www.hiv.lanl.gov](http://www.hiv.lanl.gov)). HXB2 *env* (gp160) has 2590 nucleotides: 862 nucleotides were maintained WT, while 1708 nucleotides were susceptible of being modified (66.46% of the total number of nucleotides) (Figure 11). Gp160 overlaps at its amino-terminal end with *vpu*, and at the carboxyl-terminal end with *tat*, *rev* and *nef*. In order to avoid misreading and, consequently, to avoid non-viability, these overlapping regions were not altered in *env*. The RRE is a conserved region with a very stable RNA secondary structure to where Rev joins to export the longest mRNA to the cytoplasm. Due to its importance for viral replication, the RRE was also maintained WT. All not modified regions were based on NL4-3.

The mutant envelopes were designed by modifying the nucleotides according to the CPB, the codon usage or the amount of CpGs. The modifications were designed by computer using, in some cases, the SSE program version 1.2 (Simmonds 2012). The SSE tool was used to decrease or increase the number of CpGs of a given sequence. It was also used to change CPB values. Following designing, they were produced by Invitrogen GeneArt Gene Synthesis.

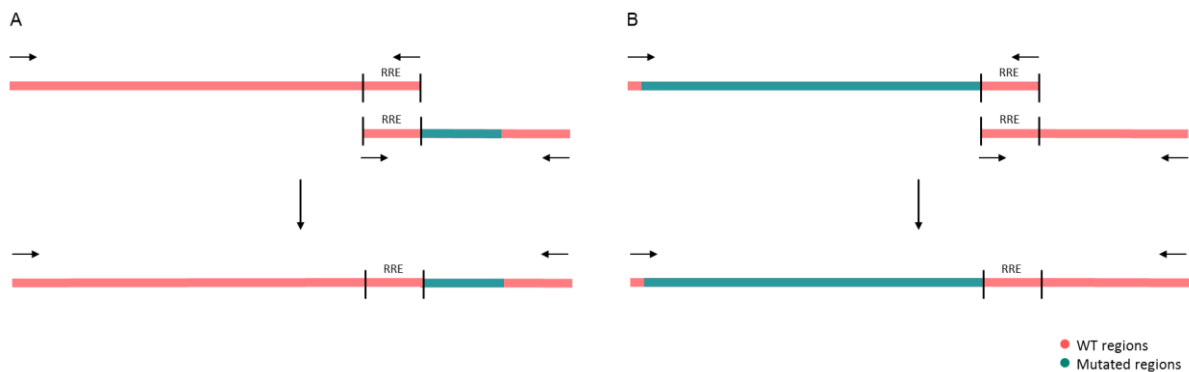


**Figure 11. Schematic representation of the *env* gene.** Not modified regions (WT regions) are shown in pink. Modified regions (mutated regions) are shown in green.

In some variants, the region before the RRE (5' region from now on) or the region after the RRE (3' region from now on) were reverted to WT. To this end, an overlapping PCR was carried out (Figure 12). To revert the mutations in both the 5' region and the 3' region, the fragments were amplified from the infectious plasmid pNL4-3. The modified regions of these variants were amplified from the expression vector containing each corresponding *env* variant.

To amplify the 5' region (WT or mutated), Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen) was used. The primers were ENV-TOPO-SD and hxb2\_RRE7610-Rv (Table 2), and the PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min and 45 sec. The same polymerase was used to amplify the 3' region (WT or mutated). The primers used in this case were hxb2\_RRE7259-Fw and pLN43-8766-Rv (Table 2); and the reaction program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min and 45 sec. Both fragments were purified using QIAquick PCR Purification Kit (Qiagen) and used together in the overlapping PCR as templates. To perform the hybridization PCR, two steps were needed. In the first step, all reaction components were added, except the oligonucleotides. The PCR program used in the first step was 1 cycle of 94°C for 2 min and 20 cycles of: 94°C for 45 sec, 55°C for 45 sec and 68°C for 3 min. In the second step, new *Taq* polymerase and the primers ENV-TOPO-SD and pLN43-8766-Rv were added to the reaction tube. The PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 45 sec, 55°C for 45 sec and 68°C for 3 min.

After joining the fragments, full-length PCRs were purified using QIAquick PCR Purification Kit (Qiagen) and cloned into the pcDNA3.1D/V5-His-TOPO, following the manufacturer's instructions.

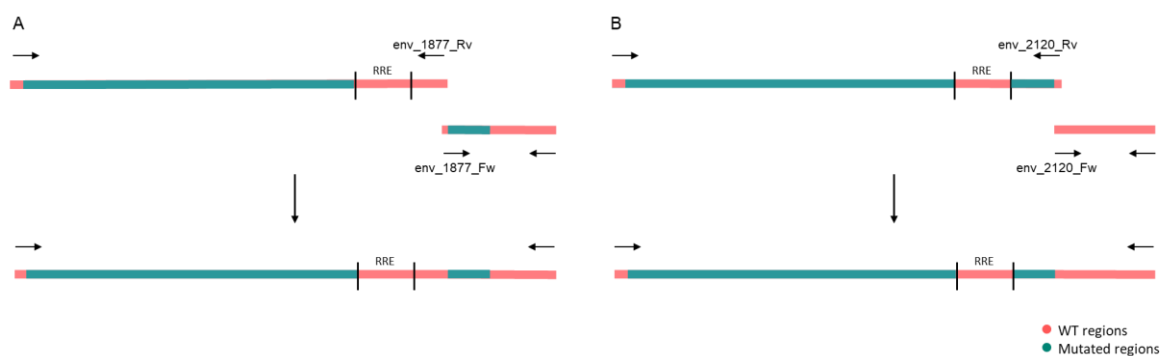


**Figure 12. Illustration of the overlapping PCRs.** (A) Amplification of the 5' region WT and the 3' region mutated. The resulting PCR was a variant with the 5' region reverted to WT. (B) Amplification of the 5' region mutated and the 3' region WT. The resulting PCR was a variant with the 3' region reverted to WT.

In other *env* variants, the 3' region was divided into two smaller fragments (3'A and 3'B) and they were reverted separately (Figure 13). Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen) was used to perform all PCRs. Specific primers were designed to carry out a site-directed mutagenesis, reverting the mutations. To amplify the first fragment (from the

beginning of the envelope to the reverting mutation), the mutant expression vector was used as template, and the primers were ENV-TOPO-SD and env\_1877\_Rv or env\_2120\_Rv (Table 2), depending on reversion of 3'A or 3'B, correspondingly. For both reactions, the PCR program used was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 2 min and 15 sec. To amplify the second fragment (from the reverting mutation to the end of the envelope), the mutant expression vector was also the template, and the primers used were env\_1877\_Fw or env\_2120\_Fw (Table 2), depending on reversion of 3'A or 3'B, correspondingly, and pLN43-8766-Rv. For both PCR reactions, the program used was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min. All fragments were purified using QIAquick PCR Purification Kit (Qiagen) and used in the overlapping PCR as templates. Fragments amplified with env\_1877 oligonucleotides were the templates for the reversion of the 3'A region; fragments amplified with env\_2120 oligonucleotides were the templates for the reversion of the 3'B region. The overlapping PCRs were performed as described above.

After joining the fragments, full-length PCRs were purified using QIAquick PCR Purification Kit (Qiagen) and cloned into the pcDNA3.1D/V5-His-TOPO, following the manufacturer's instructions.

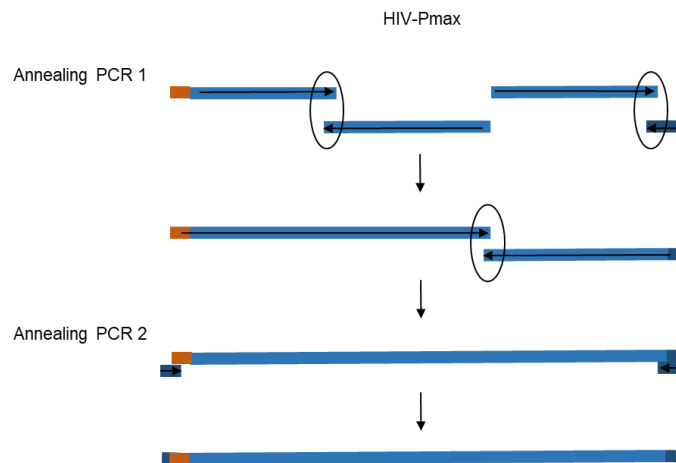


**Figure 13. Illustration of the site-directed mutagenesis reverting PCRs.** For both reverted variants, the 5' region was mutated. (A) Amplification and site-directed mutagenesis to revert the 3'A region. (B) Amplification and site-directed mutagenesis to revert the 3'B region.

## Protease-based mutants

All synthetic HIV-1 protease mutants were based on the HXB2 strain ([www.hiv.lanl.gov](http://www.hiv.lanl.gov)). To generate the protease MAX mutant, three ultramer oligonucleotides (A, B and C), previously designed with the desired MAX sequence (Table 3), were used in an overlapping PCR, as previously described above and elsewhere (Martrus et al. 2013). Briefly, 1,5 pmol of

oligonucleotides A and B, and 1,5 pmol of oligonucleotide C and T7XhoI primer (Table 3) were mixed and hybridized. A double step PCR was performed (Figure 14). For the overlapping PCR, the programs used were: 1 cycle of 94°C for 5 min and 30 cycles of: 94°C for 30 sec, 45°C for 30 sec and 68°C for 30 sec. Finally, 5µl of the PCR were used in a nested PCR using the primers T3proL and T7 (Table 3). PCR was purified using QIAquick PCR Purification Kit (Qiagen).



**Figure 14. Schematic representation of Max protease recoded virus construction.** Two annealing PCRs were performed in order to obtain the final protease fragment. Final PCR was purified.

WT and MAX S37P, G48L, Q58E, I84V, and I89L mutants were generated by site-directed mutagenesis using overlap extension PCR with mutated oligonucleotides (Table 3). To amplify the two fragments that overlapped, WT and MAX were used as templates. For each new mutant, the primers were specific to each corresponding PCR. Reverse oligonucleotides were used together with 5'-prot2 to amplify the first part. Forward oligonucleotides were used together with 3'-prot2 to amplify the second part of the protease. For both reactions (the first and second fragment), the PCR program used was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 30 sec. All fragments were purified using QIAquick PCR Purification Kit (Qiagen) and used in the overlapping PCR as templates.



## Materials and Methods

The procedure to hybridize was described previously. The primers used in the overlapping PCR were 5'-prot2 and 3'-prot2 (Table 3), and the PCR programs used were: 1 cycle of 94°C for 2 min and 20 cycles of: 94°C for 45 sec, 55°C for 45 sec and 68°C for 45 sec.

**Table 3. Oligonucleotides used in cloning, PCRs, sequencing and directed mutagenesis of the envelope variants.** Fw: forward primer; Rv: reverse primer.

Primer name	Sequence (5'-3')	Orientation	Method
5'-prot1	AGGC TAATTTTTAGGGAGATCTGGCC TTCC	Fw	PCR
5'-prot2	TCAGAGCAGCCAGGCCACAGCCCCCA	Fw	PCR
3'-prot1	GCAAATACTGGAGTATTGTATGGATTTTCAGG	Rv	PCR
3'-prot2	AATGCTTTTATTTTTTC TTCTGTCAATGGCC	Rv	PCR
A-Pmax	TCCTTTACTTCCCTCAGATCACITTTATGGCAAGACCAATTAGTACCATCAAGATAGTGGTCACTAAAGAGCTTTTATAGATACAGGAGGAGATGACTGTTTTAGAGAATG	Fw	Overlapping PCR
B-Pmax	CCATTTGCTTTTGTCCACATATTTCTATTAATATTTGATCATATGTC TTACTTTATAAACTCCCAATGCCCTATCATTTTGGCTTCCAGCGGCCGCGAGCGACATTTCTTTGTAACAGATAT	Rv	Overlapping PCR
C-Pmax	TGTGGACATAAAGCAATAGTACTGTATAGTAGGACCCAGCCGGTCAACATCATAGGAAGAAATTTATTACACAATAGGATGTACTTTAAATTTTCCCATTAGCCCTATTGAG	Fw	Overlapping PCR
T7Xol	TAATACGACTCACATAGGGCAATTTGGTACCGGGCCCTCGAGTCAAAAGGCCATCCATTCCTGGC	Rv	Overlapping PCR
T3prot	AAATTAACCCTCAC TAAGGGACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGACTAGTGGATCCCGCCGGGCTGGAGGAATTC TTCTTTAACTTCCCTCAG	Fw	Nested PCR
T3	ATTAACCCTCACTAAAGGGA	Fw	PCR, Sequencing
Q88E WT F	AAAGTAGAGAGTATGATCA	Fw	Directed mutagenesis
Q88E WT R	TGATCACTACTCTTACTTT	Rv	Directed mutagenesis
Q58E MAX F	AAAGTAGAGAGATGATCA	Fw	Directed mutagenesis
Q58E MAX R	TGATCACTACTCTTACTTT	Rv	Directed mutagenesis
I84V WT F	CC TGTCAACGTAATGGAG	Fw	Directed mutagenesis
I84V WT R	CTTCCAATTACGTTGACAGG	Rv	Directed mutagenesis
I84V MAX F	CCGGTCAAGTCAATGGAG	Fw	Directed mutagenesis
I84V MAX R	CTTCCAATTACGTTGACCGG	Rv	Directed mutagenesis
S37P WT F	GAAGAAATGGCTTTTCCAGG	Fw	Directed mutagenesis
S37P WT R	CC TGGCAAGGCATTTCTTC	Rv	Directed mutagenesis
S37P MAX F	GAAGAAATGGCCCTGCCGGG	Fw	Directed mutagenesis
S37P MAX R	CCCGGACGGGCAATTTCTTC	Rv	Directed mutagenesis
I89L WT F	GGAAAGAAATATATTGACTCA	Fw	Directed mutagenesis
I89L WT R	TGAGTCAATATATTCTTCC	Rv	Directed mutagenesis
I89L MAX F	GGAAAGAAATATATTACACA	Fw	Directed mutagenesis
I89L MAX R	TGTGTTAATATTCTTCC	Rv	Directed mutagenesis
G48L F	CAAAAATGACTGGGAATGG	Fw	Directed mutagenesis
G48L R	CCAAATCCCGAGTATCAATTTTG	Rv	Directed mutagenesis
G48L MAX F	CAAAAATGACTGGGAATGG	Fw	Directed mutagenesis
G48L MAX R	CCAAATCCCGAGTATCAATTTTG	Rv	Directed mutagenesis

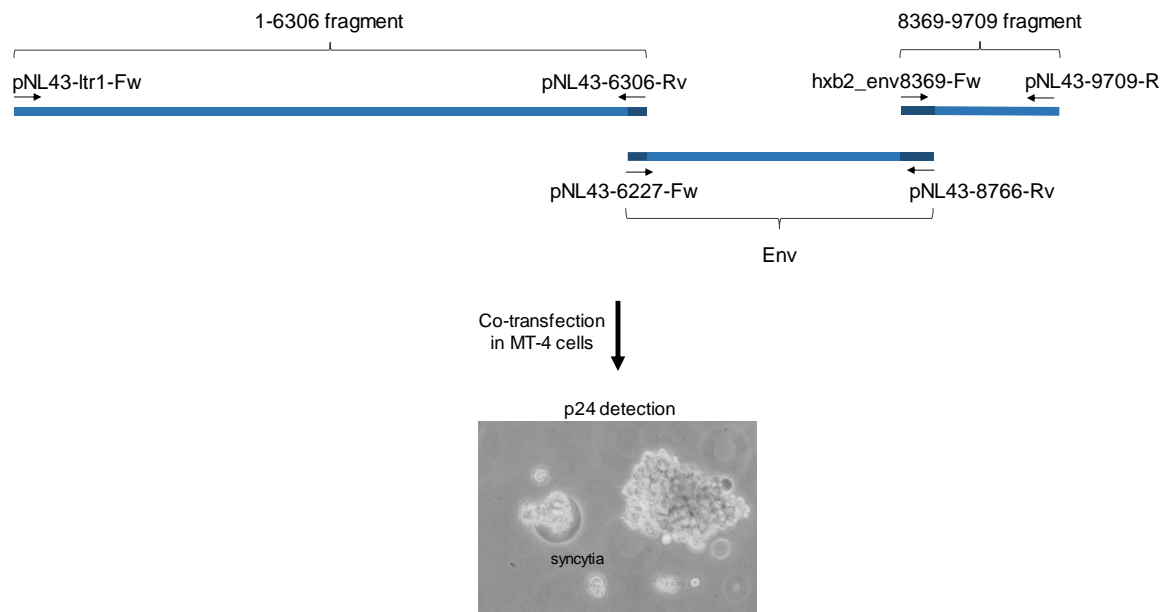
## **Transfections**

### *MT-4 cells transfection*

To obtain gp160-based mutants viral particles, the different parts of HIV-1 genome were PCR amplified and transfected in MT-4 cells. All PCR amplifications used for the transfection were performed using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen). The region of the NL43 genome that extends from nucleotide 1 to nucleotide 6306 was previously cloned in the pDRIVE vector, and this region was amplified using the primers pNL43-ltr1-Fw and pNL43-6306-Rv (Table 2). The PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 6 min and 30 sec. The region of the NL43 genome that extends from position 8369 to position 9709 was previously cloned in the pDRIVE vector and it was amplified using the primers hxb2\_env8369-Fw and pNL43-9709-Rv (Table 2). The PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min and 30 sec. For the amplification of the WT\_env and the different mutants, which were cloned in the expression plasmids, the primers used were pNL43-6227-Fw and pNL43-8766-Rv (Table 2); the PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 3 min.

All PCRs were purified using QIAquick PCR Purification Kit (Qiagen), following the manufacturer's instructions, and eluted in 50µl of milli-Q water. Briefly, 600ng of each purified fragment were transfected into MT-4 cells by electroporation. Of notice, the amplification of the envelope overlaps with the other two fragments of the genome, allowing the virus to form viral particles if viable (Figure 15) (Fujita et al. 2013).

## Materials and Methods

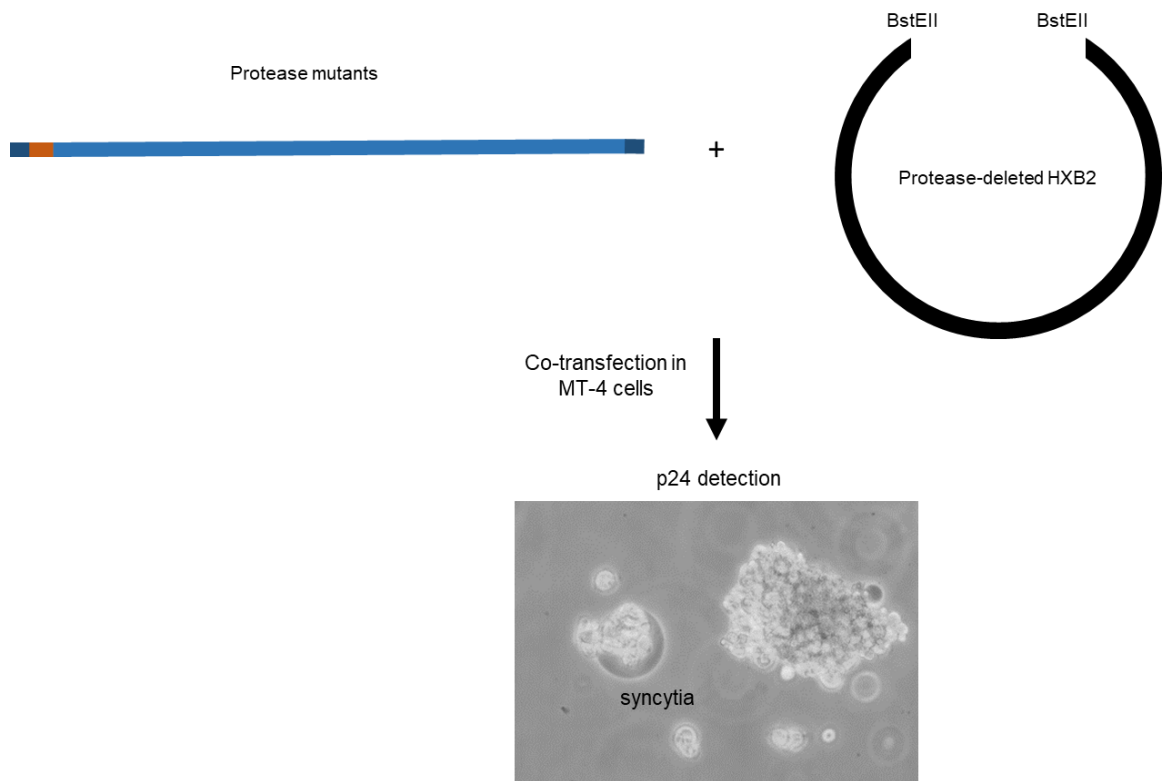


**Figure 15. Representation of the co-transfection in MT-4 cells for the envelope mutants.** The different regions of the HIV-1 genome were PCR amplified: 1-6306 fragment, 8369-9709 fragment, and the different variants of the envelope. They were co-transfected and recombined after culture. Syncytia formation was observed if the viruses were replicative.

To obtain protease-based mutants viral particles, WT protease and mutants were amplified using the primers 5'-prot2 and 3'-prot2, and the reaction program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min. The PCR was purified and 100 ng were cotransfected in MT-4 cells with 1 µg of protease-deleted HXB2 infectious clone that had previously been linearized with 1U of BstEII (New England Biolabs) (Figure 16).

To monitor viral production, HIV-1 p24 antigen was quantified (Genscreen HIV-1 Ag assay, Bio-Rad) from the culture supernatant four and seven days after transfection. If p24 concentration was higher than 500 ng/ml the virus was recovered, aliquoted and stored at -80°C. If after seven days of culture no p24 was detected, blind passages of fresh medium and cells were performed to recover the virus. The mutant was considered as lethal if p24 and syncytia (fusion of an infected cell with neighboring cells (Figure 15, Figure 16) ) were not detected after thirty days post transfection.

Viable viruses were titrated in MT-4 cells and values were given as tissue culture dose for 50% of infectivity per ml (TCID<sub>50</sub>/ml) as previously described (Pauwels et al. 1988).



**Figure 16. Representation of the co-transfection in MT-4 cells for the protease mutants.** The protease mutants were PCR-amplified and co-transfected in MT-4 cells with a protease-deleted HXB2 previously linearized with BstEII. Syncytia formation was observed if the viruses were replicative.

### *293T cells transfection*

To obtain non-infectious pseudoviruses and analyze virus protein expression and mRNA amounts, 293T cells were transfected with the pNL43-dENV-EGFP plasmid (see above) or pTOPORev. Briefly, these cells were seeded in a six-plate at a density of  $5 \times 10^5$  cells/well the day before transfection. The cells were transfected with  $2.5 \mu\text{g}$  of the pNL43-dENV-EGFP plasmid and  $1.25 \mu\text{g}$  of the expression plasmid containing each *env* mutant, or with  $1 \mu\text{g}$  of pTOPORev and  $3 \mu\text{g}$  of the expression plasmid containing each *env* mutant. The co-transfection of the confluent cells was performed with Lipofectamine 3000 Reagent (Invitrogen), as described by the manufacturer. Forty-eight hours after transfection, transfected cells and/or supernatants containing the pseudoviruses were collected.

Supernatant containing the pseudoviruses were filtered and concentrated using Lenti-X™ Concentrator (Takara). Following concentration, pseudoviruses were titrated in TZM-bl cells and values were given as tissue culture dose for 50% of infectivity per ml (TCID<sub>50</sub>/ml) as described elsewhere ((Sarzotti-Kelsoe et al. 2014).

### Real-time quantitative PCR

To determine the amounts of mRNA, real-time quantitative PCR (qPCR) was carried out. A total of  $1 \times 10^6$  cells were collected from the transfection of 293T with the expression plasmids containing each *env* mutant and pTOPORev. Total RNA was extracted from these cells using High Pure RNA Isolation Kit (Roche LifeScience). Complementary DNA (cDNA) was synthesized from 500 ng of RNA using PrimeScript™ RT Master Mix (Perfect Real Time). The obtained transcripts were used to conduct the qPCR. Briefly, qPCRs were performed using TaqMan Universal Master Mix (Applied Biosystems). For the amplification of the WT and mutant *envs*, primers located in the RRE region were used: qPCR-EnvRRE (Table 4). GAPDH (Life Technologies, Applied Biosystems) was used as a cellular control. Rev mRNA was also detected using the primers and probe Rev\_qPCR (Table 4).

**Table 4. Oligonucleotides and probes used in qPCR.**

Primer name	Probe (5'-3')	Primer forward	Primer Reverse
qPCR-EnvRRE	CTTGGGAGCAGCAGGAAGCACTAT	CAGTGGGAATAGGAGCTTTGT	TGTACCGTCAGCGTCATTG
Rev_qPCR	TGGAGAGAGAGACAGACAGATCCA	GCCCGAAGGAATAGAAGAAGAA	GATCGTCCCAGATAAGTGCTAAG

### Immunoblot analyses

To determine the protein expression of the different variants, immunoblot assays were carried out. Briefly,  $1 \times 10^6$  293T transfected cells were lysed using 100  $\mu$ l of Cell Extraction Buffer (ThermoFisher Scientific) supplemented with PMSF and protease inhibitors (cOmplete™ Protease Inhibitor Cocktail, Sigma-Aldrich). 10  $\mu$ g of protein were separated by electrophoresis on NuPage 4-12% Bis-Tris gels (ThermoFisher Scientific) and blotted onto nitrocellulose membranes (ThermoFisher Scientific). Antibody against Hsp90 (C45G5) was obtained from Cell Signaling Technology. Antibody against gp120 (Anti-gp120 HIV-1 Polyclonal Antibody) was obtained from American Research Products (ARP). Antibody against V5 (V5 Tag Monoclonal Antibody) was obtained from Invitrogen.

### HIV-1 drug susceptibility tests

Protease inhibitors Atazanavir (ATV) and Darunavir (DRV) were obtained from the NIH AIDS Research and Reference Reagent Program. Env monoclonal antibody 2F5 was obtained from Polymun Scientific. Following virus propagation and titration, a tetrazolium-

based colorimetric method was used to determine the HIV-1 drug susceptibility (IC<sub>50</sub>) to ATV, DRV or 2F5 Ab in MT-4 cells as previously described (Betancor et al. 2015). Briefly, HIV-1 drug susceptibility data were collected after infecting  $3 \times 10^4$  MT-4 cells at a MOI of 0.003 and adding various concentrations of each drug or antibody. Cells were incubated for five days and the number of viable cells was determined by the colorimetric method (Pannecouque, Daelemans, and De Clercq 2008).

## **Selection of ATV- and DRV- resistant viruses**

Protease-based viruses were added at a MOI of 0.01 to  $1 \times 10^6$  MT-4 cells. After four days, 1/10 of the culture was transferred, including cells and supernatant, into  $1 \times 10^6$  fresh MT-4 cells. All virus passages were performed in duplicate. Virus production was monitored by measurements of p24 antigen. The starting concentrations were 4 nM of ATV and 3 nM of DRV. Throughout the passages, the drug concentrations were increased until they reached 40 nM of ATV and 25 nM of DRV. In parallel, both viruses were also propagated without either drug.

At passages 1 and 32, 140  $\mu$ l of culture supernatant were collected, from which viral genomic RNAs were extracted using QIAamp viral RNA kits (Qiagen). The purified viral RNAs were then reverse transcribed and PCR amplified using the SuperScript III first-strand synthesis system for RT-PCR (Invitrogen) and 10 pmol (each) of the corresponding protease oligonucleotides: 5'-prot1 and 3'-prot1 (Table 3). The PCR program was 1 cycle of 55°C for 35 min and 94°C for 2 min; 40 cycles of: 94°C for 15 sec, 55°C for 30 sec and 68°C for 1 min; and 1 cycle of 68°C for 7 min. A second PCR (nested PCR) was performed using the oligonucleotides 5'-prot2 and 3'-prot2 and the Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen). The PCR program was 1 cycle of 94°C for 2 min and 35 cycles of: 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min (Capel et al. 2012; Fernandez, Clotet, and Martinez 2007). These PCR products were the starting material for ultradeep sequencing.

## **Ultradeep sequencing**

Massive parallel sequencing was performed with the MiSeq (Illumina) platform. Libraries of 558-nt DNA fragments were ligated to the Illumina adapters by use of a Kapa HyperPrep kit (Roche), a SeqCap adapter kit A (Roche), and a SeqCap adapter kit B (Roche). The products were purified using Kapa Pure beads (Roche). All libraries were quantified using a Qubit dsDNA HS assay kit (ThermoFisher) and a Qubit fluorometer (ThermoFisher) and were qualified using an Agilent DNA 1000 kit (Agilent) and a bioanalyzer (Agilent).

## **Materials and Methods**

---

Sequencing was performed using an Illumina MiSeq reagent kit v3 (600 cycles) (Illumina) following the manufacturer's protocol. Sequencing and paired-end analysis were performed to obtain robust fastq data for bioinformatics analysis. A mean of 50,000 sequences (reads) were obtained per amplicon and per patient sample.

Fastq files were obtained from the MiSeq system by index and pool and submitted to FLASH (Magoc and Salzberg 2011). The 2 × 300 paired-end reads were overlapped to reconstruct the amplicons, with the minimum number of overlapping nucleotides set to 20 and the maximum number of overlapping mismatches set to 10%. The subsequent analysis was performed as previously described (Gregori et al. 2013). Briefly, fastq files were demultiplexed using amplicon oligonucleotides, and oligonucleotides were trimmed at both ends. Each amplicon and strand read was pairwise aligned with respect to the reference WT sequence, insertions were removed, and deletions were repaired if fewer than three gaps were produced. Reads with multiple indeterminations were removed, while reads having a single indetermination were repaired per the reference sequence. Filtered and repaired reads were collapsed into haplotypes with corresponding frequencies. Haplotypes with abundances below 0.1% or that were unique to the forward or reverse strand were removed. Haplotypes common to the forward and reverse strands and with abundances of  $\geq 0.1\%$  were considered consensus haplotypes, and their frequencies were summed. In the final step to remove artifacts, consensus haplotypes with abundances below 0.5% were filtered out. All computations were performed in the R language and platform, using in-house scripts as well as the packages Biostrings (R package, 2.24.1), Ape (Paradis, Claude, and Strimmer 2004), and Seqinr (Charif et al. 2005).

Virus population genetic diversity ( $p$  distance) was determined using the MEGA6 software package (Tamura et al. 2013). To determine possible selective pressures, the MEGA6 software package was used to calculate the proportion of synonymous substitutions per potential synonymous site and the proportion of nonsynonymous substitutions per potential nonsynonymous site. Shannon's entropy values were calculated as follows:  $S_n = -\sum_i (p_i \ln p_i) / \ln N$ , where  $N$  is the total number of analyzed sequences and  $p_i$  is the frequency of each sequence in the viral quasispecies.  $S_n$  values vary from 0 (no complexity) to 1 (maximum complexity) (Wolinsky et al. 1996). The phylogenetic reconstructions were also performed using the MEGA6 software package.

## **RNA secondary structures: Computational analyses**

The RNA secondary structures of the mutants of *env* were predicted by using several computational tools. All bioinformatics analyses were performed in collaboration with Dr. Manja Marz at Friedrich Schiller University Jena (Germany). The employed tools were the following:

### *ViennaRNA Package 2.4.10*

The ViennaRNA Package is formed by several programs that allow the prediction and comparison of RNA secondary structures. The following programs were used:

#### RNAfold

This tool was used to predict secondary structures. RNAfold uses algorithms which calculate the minimum free energy (MFE) structure and the partition function of RNAs (Lorenz et al. 2011). The MFE algorithm yields a single optimal structure (Lorenz et al. 2011; Zuker and Stiegler 1981) and through the partition function of the Boltzmann distribution a thermodynamic ensemble is obtained, which enables to calculate base pair probabilities (Lorenz et al. 2011; McCaskill 1990).

The parameter `-noLP` was set to produce structures without lonely pairs, so pairs that can only occur isolated are discarded.

#### Relplot

The secondary structure plot and the pair probabilities' dot plot produced by RNAfold `-p` (partition function) can be colored by this auxiliary program. The annotation used to colorize was positional entropy (structures with little flexibility have low entropy and the reliability of the prediction is higher), pair probabilities (the probability of a nucleotide of being paired) or accessibility (the probability of a nucleotide of being unpaired) (Lorenz et al. 2011).

#### RNAalifold

This tool was used to predict secondary structures from a group of aligned RNAs. It also calculates the MFE, the partition function and the base pairing probability (Lorenz et al. 2011).



The parameters -p, -noLP and -aln were set to produce the structures.

### *VARNA*

VARNA was used to draw and visualize the secondary structures obtained by RNAfold. It allows to edit and annotate the structures, as well as adding colors to specific nucleotides.

### *CD-HIT-EST*

This tool was used to cluster nucleotide sequences. The aim of this program is to reduce sequence redundancy for further analysis. The percentage of identity between the sequences can be established.

### *MAFFT and LocARNA*

These two programs were used to create multiple sequence alignments (Kato et al. 2002). The difference is that LocARNA performs the alignments not only based on the sequences, but also on the structures of the given RNAs. The version used for LocARNA was 1.8.11.

### *LRIsScan*

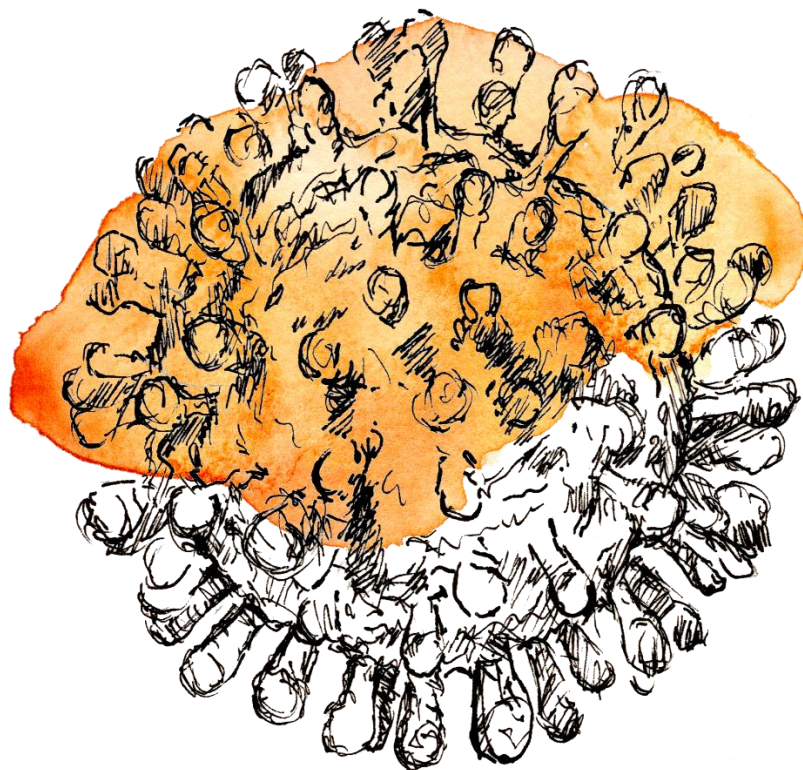
The number of possible secondary structures exponentially increases with the length of a sequence. Therefore, it is not feasible to enumerate and analyze all the possibilities. In addition, one specific nucleotide could base-pair with any corresponding nucleotide up or downstream, increasing the flexibility with the length, and thus, decreasing the reliability of one specific prediction. And lastly, in a completely randomized sequence it can be expected that each third pair of nucleotides can form a base-pairing.

For these reasons, a tool to analyze and predict long range RNA-RNA interactions is necessary. Such interactions can involve at least 10 nucleotides and potentially span the majority of the viral genome. LRIsScan was used to predict those interactions within the HIV-1 genome (Fricke and Marz 2016).

## **Statistical Analyses**

Statistical analyses were performed using GraphPad Prism 7.04 for Windows (GraphPad Software, La Jolla California USA), establishing the confidence interval in 95% ( $p$ -value=0,05). The analyses performed were Student's t-test.





**Chapter One.  
Codon pair recoding of the protease gene  
and its evolvability**

---



## Recoding of the protease gene

The HIV-1 protease gene was synonymously recoded as previously described (Martrus et al. 2013). The resulting protease MAX variant had 38 synonymous substitutions (13% of the protease gene) (Figure 17). This mutant MAX improved the protease CPB from 0.069 (WT) to 0.254. The codon pair usage was changed in the protease without modifying the amino acid sequence.

```

          *      20      *      40      *
WTp1 : CCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGG : 50
MAXp1 : .....T.. : 50

          60      *      80      *      100
WTp1 : GCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAG : 100
MAXp1 : T.....A.....T.....T..T.... : 100

          *      120      *      140      *
WTp1 : AAGAAATGAGTTTGGCAGGAAGATGGAAACCAAAAATGATAGGGGGGAATT : 150
MAXp1 : .....TCGC....G..CC.C....G.....C... : 150

          160      *      180      *      200
WTp1 : GGAGGTTTTATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATCTG : 200
MAXp1 : .....A.....A.....A.....A...T.A.....A.. : 200

          *      220      *      240      *
WTp1 : TGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACA : 250
MAXp1 : .....A.....T.....C..G..G..... : 250

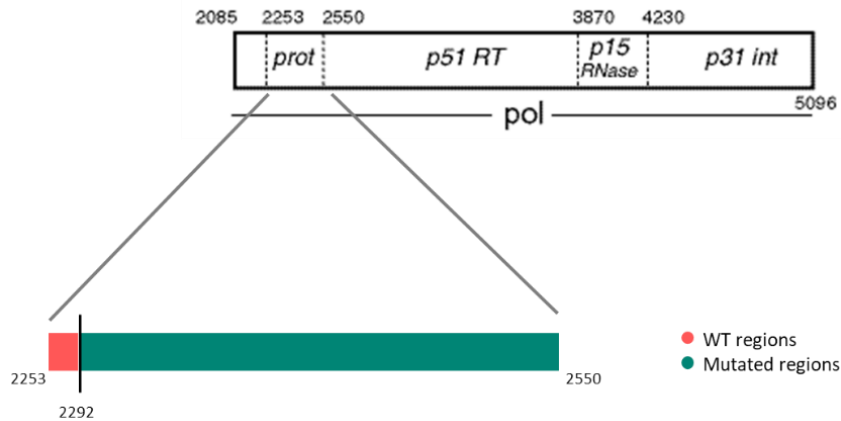
          260      *      280      *
WTp1 : TAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAAATTTT : 297
MAXp1 : .C..A.....T.A..A..A..A..A..T..... : 297

```

**Figure 17.** Alignment of the HIV-1 protease nucleotide sequence of the wild-type (WT) virus, which corresponds to the HIV-1 HXB2 strain (<http://www.hiv.lanl.gov>), and the synthetic MAX variant, which was generated by a PCR combining three overlapping synthetic DNA oligonucleotides as previously described.

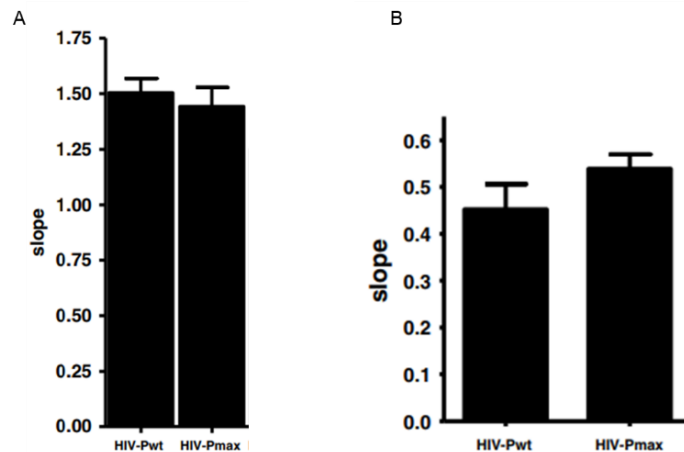
The protease overlaps with *gag* in the first 40 nucleotides (extreme 5') (Figure 2). To avoid non-synonymous changes and/or lethality, those 40 nts were not modified (Figure 17, Figure 18).

## Results



**Figure 18. Schematic representation of the protease MAX.** Pink represents WT regions, and blue represents modified regions. The first 40 nucleotides of the protease gene were not modified due to its overlapping with gag gene.

It was shown by Martrus et al. that the replication capacity of protease MAX was indistinguishable from the WT, both in MT-4 cells and PBMCs (Figure 19) (Martrus et al. 2013).



**Figure 19. Replication capacities of protease WT and protease MAX.** Adapted from (Martrus et al. 2013). A) Replication capacity in MT-4 cells shown as the log of p24 antigen measurements. B) Replication capacity in PBMCs shown as the log of p24 antigen measurements.

## ATV and DRV selective pressure

Both proteases (PR-WT and PR-MAX) were subjected to ATV and DRV selective pressure. To this end, both viruses were propagated in MT-4 cells for 32 serial passages. Viruses and infected cells were passaged every 3-4 days over 128 days of culture. This experiment was

performed in duplicate, with and without drugs. As described in Materials and Methods, the concentration of the PIs increased with the passages.

Virus IC<sub>50</sub> was determined before starting the passages and at passage 32 (WTp1, MAXp1, WTp32 and MAXp32). Before the passages, both viruses showed similar IC<sub>50</sub> (Table 5). At passage 32, for ATV, WTp32 showed a 5-fold increase and MAXp32 showed a 13-fold increase; whereas for DRV, WTp32 showed a 6-fold increase and MAXp32 showed a 10-fold increase. PR-MAX showed more resistance to the drugs than PR-WT, however, these differences were not significant ( $p=0.4816$  for ATV and  $p=0.3451$  for DRV) (Table 5).

**Table 5. Susceptibilities of PR-WTs and PR-MAXs to ATV and DRV before passages and at passage 32.**

Protease	IC <sub>50</sub> (nM) (fold change)	
	ATV	DRV
WTp1	3.1 ± 1.2 (1)	0.9 ± 0.3 (1)
WTp32	14.6 ± 5.3 (5)	5.7 ± 1.0 (6)
MAXp1	1.6 ± 0.1 (1)	0.9 ± 0.2 (1)
MAXp32	21.2 ± 9.0 (13)	9.3 ± 1.9 (10)

Given these results, it was evidenced that MAX can develop resistances to the protease inhibitors (PIs) as the WT. This demonstrated that synonymously changing the CPB of the virus did not alter its capacity to adapt.

## Ultradeep sequencing

Culture supernatant containing the viruses were recovered at passage 32 and the RNAs were extracted. RT-PCRs were performed amplifying the protease region. These PCR products were used as templates for ultradeep sequencing. The aim of this assay was to determine the resistance mutations and to compare their frequencies.

For PR-WT and PR-MAX in the two conditions (presence or absence of ATV or DRV),  $1.9 \times 10^7$  and  $4.1 \times 10^7$  individual nucleotides were sequenced, respectively. This analysis showed that for both viruses, without drugs, there were no resistance-associated substitutions. When the mutations were analyzed for the viruses propagated in the presence of PIs, some substitutions, already described as resistance mutations to ATV and DRV, were found. When the mutations of PR-WT and PR-MAX were compared, it was found that they developed a different repertoire of resistant substitutions (Table 6). In



## Results

particular, the substitution G16E was only observed in the WT in the presence of the two PIs. The reason why this substitution was not observed in MAX was that for this virus, a transition and a transversion were necessary, whereas for the WT only a transition was needed. Other resistance mutations that only appeared in the WT were L23I, V32T, P39Q and V77I (Table 6).

There were some mutations that were only observed with the MAX variant. These substitutions were L33F, K45R/I, G48L, Q58E/K, V82D, I84V, I89L, E21K, H69Y and T91S (Table 6). As it happened with the G16E, the WT needed two mutations to develop the I89L, whereas for the MAX it was only one substitution. However, for other substitutions, K45I, G48L, Q58E and I84V, there was not a clear explanation for their unique appearance in MAX.

**Table 6. Substitutions associated with protease inhibitor resistance detected after MT-4 cell passages in the presence of ATV, DRV, or no drug.**

Drug and substitution	% of sequences			
	Expt 1		Expt 2	
	WTp1	MAXp1	WTp1	MAXp1
<b>ATV</b>				
L10F	1.0	28.0	0	0
G16E	1.7	0	1.8	0
L23I	0.8	0	0	0
V32A	76.3	0	53.3	59.9
V32I	10.1	24.0	2.9	23.8
V32T	0.65	0	0	0
L33F	0	0.8	0	0
P39Q	0.9	0	0.9	0
K45I	0	10.2	6.0	0
K45R	0	0.8	0	0
M46I	12.1	0	1.1	10.4
G48L	0	0	0	22.4
I50L	0	28.6	27.3	0
Q58E	0	20.2	0	0
Q58K	0	0	0	0.8
A71V	88.2	23.2	62.8	71.1
V77I	0	0	11.4	0
V82I	0	0	9.8	0
V82D	0	0	0	1.2
I84V	0	43.0	0	0
N88S	14.0	28.2	19.7	3.4
I89L	0	0	0	59.4
T91S	0	3.6	0	0
<b>DRV</b>				
L10F	0	19.2	0	0
G16E	16.5	0	55.5	0
E21K	0	1.4	0	0
A28S	46.1	45.4	6.6	0
L33F	0	94.2	0	89.2
P39Q	0.6	0	0.6	0
K45I	0	0	0	9.8
M46I	6.0	3.9	94.0	14.6
I50L	49.7	52.3	39.1	100
Q58K	0	0	0	1.2
H69Y	0	0.6	0	0
A71V	0	25.2	0	0
V82I	85.2	0.6	34.8	19.7
I84V	3.4	0	0	0
<b>No drug</b>				
D30N	1.4	2.3		
S37P	0	11.7		
P39Q	2.1	0		
P39T	0.8	0		
R41K	0.8	0		
K43R	0	0.5		
Q58K	0	1.1		
I72T	1.8	0		
V82I	1.3	0		

The substitutions E21K, L23I, P39Q, G48L, H69Y and T91S have not been previously described as PI resistance substitutions; whereas L10F, G16E, L33F, Q58E/K, V77I, I84V and I89L have already been described as drug resistances (Wensing et al. 2019).

Overall, the appearance in the MAX variant of these unusual resistances indicated that this mutant might have explored a sequence space different from that explored by the WT protease.

## Exploring resistance substitutions

To explore why some resistance mutations were favored in the MAX sequence background, these mutations were individually introduced by site-directed mutagenesis in both WT and MAX backgrounds. It was also included the substitution S37P, which only appeared in the MAX variant after passages without drugs.

Following the site-directed mutagenesis, the new variants carrying different resistance substitutions were transfected in MT-4 cells to generate the corresponding viruses. Once collected the supernatants, RNAs were extracted and RT-PCRs were performed to sequence and control the protease region of the newly formed viruses.

The susceptibility to ATV of these variants was measured by determining the IC<sub>50</sub>. The result indicated similar drug resistances for all of them, both in WT and MAX backgrounds. Only WTG48L and MAXG48L displayed higher IC<sub>50</sub> to ATV, but the difference was not significant (Table 7).

**Table 7. Susceptibilities of HIV-1 strains carrying WT or MAX protease variants to ATV.**

Protease	ATV IC <sub>50</sub> (nM) (fold change)
WT	1.8 ± 1.2 (1)
MAX	1.6 ± 0.4 (0.9)
WTS37P	2.2 ± 1.0 (1.2)
MAXS37P	2.2 ± 1.4 (1.2)
WTG48L	14.4 ± 0.5 (8.1)
MAXG48L	14.4 ± 1.4 (8.1)
WTQ58E	4.6 ± 2.5 (2.6)
MAXQ58E	5.6 ± 3.8 (3.2)
WTI84V	1.3 ± 0.9 (0.7)
MAXI84V	1.5 ± 0.8 (0.8)
WTL89I	1.5 ± 0.5 (0.8)
MAXL89I	2.1 ± 0.6 (1.2)

## Replication capacity assays

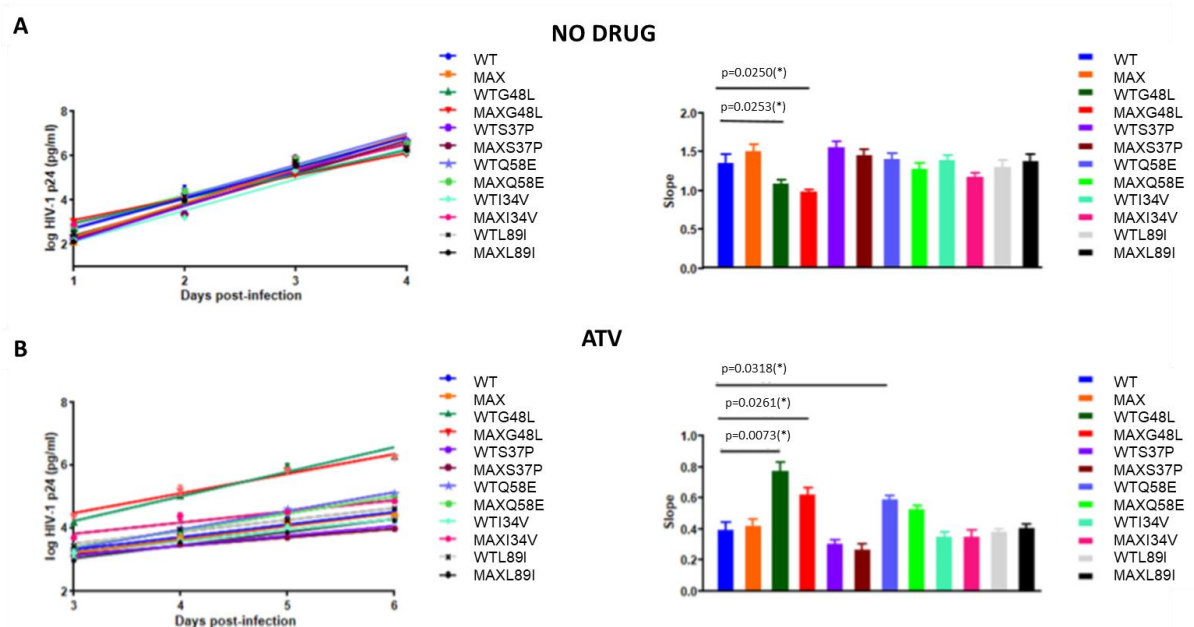
Replication capacity assays were carried out with these 10 new variants (Table 7), together with WT and MAX variants. These assays were performed in the presence and absence of ATV. The cells used were MT-4 and the amount of ATV used was 20 nM.

In the absence of ATV, almost all viruses replicated with a similar capacity as WT virus (Figure 20). Only two of the variants, WTG48L and MAXG48L, showed lower replication capacity.

In the presence of ATV, now WTG48L and MAXG48L variants displayed higher replication capacities, indicating their selective advantage. The Q58E variant also showed selective advantage, although this difference was not significant (Figure 20).

These results indicated that the substitutions G48L, Q58E, I84V and I89L were not intrinsically prohibited in the WT sequence background, and there must be other factors that may explain why these substitutions were not found in the WT protease background.

Similarly, a different mutant spectra was observed in both WT and MAX sequence backgrounds when they were passaged without drugs (Table 6). Only one of the emerging substitutions (D30N) was found in both backgrounds. Moreover, it was not possible to determine if the mutations were adaptive or neutral for both viruses.



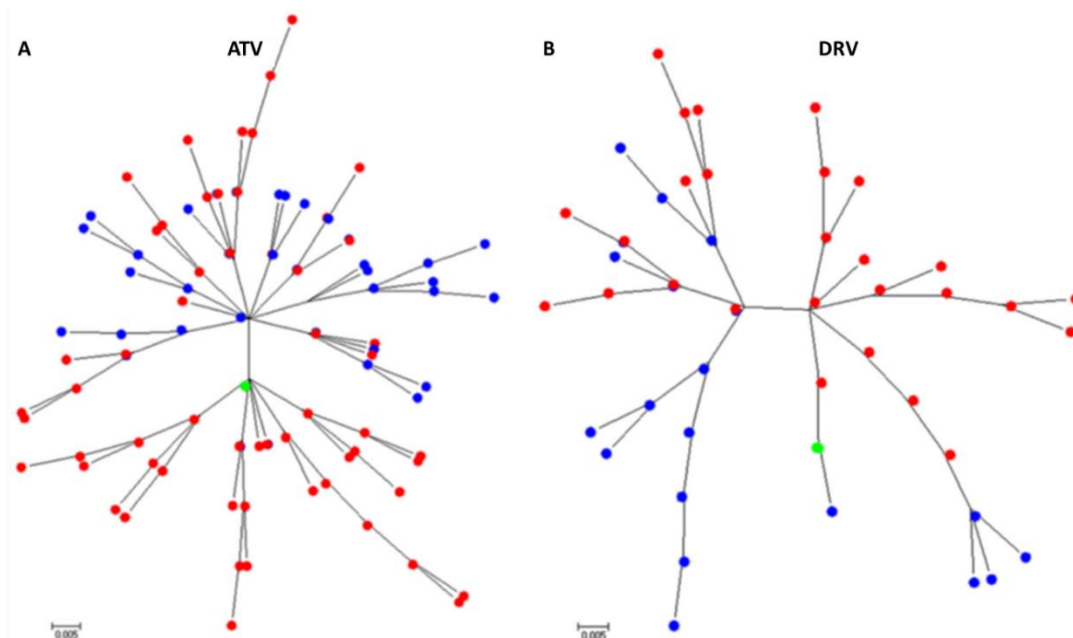
**Figure 20. Replication kinetic assay of wild-type HIV-1 (WT) and the recoded MAX protease variant (MAX) in MT-4 cells.** HIV-1 antigen p24 concentrations in culture supernatants were measured on days 0 to 4 in the absence of drug and days 0 to 6 in the presence of drug. For each virus, the slope of the plot provides an estimate of the viral replication capacity. Bars show the slope

for p24 antigen production from each virus after infection of MT-4 cells. Comparisons of the WT (HXB2) and recoded mutant MAX viruses are shown, as well as comparisons of the corresponding WT and MAX virus variants (i.e., S37P, G48L, Q58E, I84V, and I89L). The significance of the differences between slopes was calculated using an unpaired t test with Welch's correction in GraphPrism v. 7 software. (A) Kinetic assays performed in the absence of drug. All slopes values were statistically tested against the WT value. Only WTG48L and MAXG48L displayed a lower replication capacity than that of the WT. (B) Kinetic assays performed in the presence of 20 nM atazanavir (ATV). All slopes values were statistically tested against the WT value. WTG48L, MAXG48L, and WTQ58E displayed a higher replication capacity than that of the WT in the presence of ATV. Values represent the means  $\pm$  standard deviations (SD) for at least three independent experiments.

## Phylogenetic analyses

The variants that were recovered at passage 32 in the presence of ATV and DRV were used to perform a maximum likelihood phylogenetic reconstruction. Both WT and MAX viruses were included (Figure 21).

The result displayed a different evolutionary trajectory for each of the viruses. Although both of them shared the same amino acid sequence at the beginning, they evolutionary diverged along the passages. Moreover, it can be observed that the MAX protease (red color in Figure 21) generated more amino acid diversity.



**Figure 21. Maximum likelihood phylogram of wild-type (WT) and MAX unique HIV-1 protease amino acid variants selected after 32 passages in MT-4 cells and in the presence of atazanavir (ATV) (A) or darunavir (DRV) (B).** Phylogenetic reconstruction was generated using a Jones-Taylor Thornton (JTT) model as implemented in the MEGA6 software package. Both phylogenetic trees show that the WT and MAX viruses, which shared an identical starting amino acid sequence, followed different evolutionary trajectories. Blue and red labels correspond to WT and MAX variants, respectively. Green labels represent the starting HXB2 protease amino acid sequence.

## Results

For this reason, the nucleotide diversities of the WT and MAX after 32 passages were compared. After analyzing their heterogeneity, it can be concluded that MAX populations had a significantly higher nucleotide diversity when compared with the WT when they were propagated in the presence of ATV ( $0.00946 \pm 0.00005$  vs  $0.00517 \pm 0.00004$ ) ( $p < 0.0001$ ). No differences were observed when the viruses were propagated in the presence of DRV ( $0.00490 \pm 0.00004$  vs  $0.00482 \pm 0.00003$ ).

For both viruses, WT and MAX, there were more non-synonymous mutations than synonymous when the viruses were passaged in the presence of PIs (Table 8). However, there was a higher diversity within the MAX populations ( $p < 0.0001$ ), both in synonymous and non-synonymous mutations. This effect was also observed with MAX populations when they were passaged without PIs ( $0.00095 \pm 0.00001$  vs  $0.00057 \pm 0.00001$ ) ( $p < 0.0001$ ). Interestingly, in this same scenario, MAX populations displayed higher non-synonymous mutations diversity ( $0.00117 \pm 0.00001$  vs  $0.00061 \pm 0.00001$ ) ( $p < 0.0001$ ) than synonymous diversity ( $0.000310 \pm 0.00002$  vs  $0.00053 \pm 0.00002$ ) ( $p < 0.0001$ ). This result suggested that the two viruses, WT and MAX, are under different selective forces when no PI is exerting selective pressure.

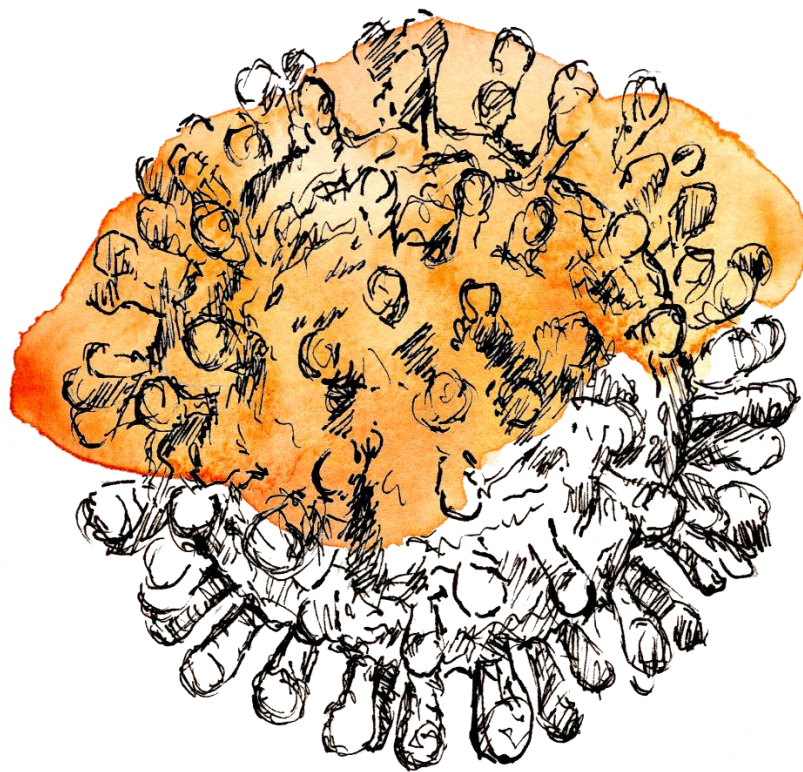
Shannon's entropies were also calculated (Table 8). This measurement is another parameter that indicates the population genetic diversity. When analyzing these values, MAX populations always displayed higher values when compared to WT.

**Table 8. Summary of population metric.**

Protease	No. of bases sequenced	No. (%) of mutations detected	Diversity ( $p$ distance)			Sn value <sup>a</sup>
			Sequence	Synonymous	Nonsynonymous	
WTp1	$8.8 \times 10^6$	2,691 (0.03)	$0.00047 \pm 0.00000$	0	$0.00062 \pm 0.00001$	0.04
MAXp1	$8.5 \times 10^6$	4,257 (0.05)	$0.00085 \pm 0.00004$	0	$0.00113 \pm 0.00006$	0.05
WTp32	$7.2 \times 10^6$	2,465 (0.03)	$0.00057 \pm 0.00001$	$0.00053 \pm 0.00002$	$0.00061 \pm 0.00001$	0.05
MAXp32	$5.2 \times 10^6$	2,924 (0.06)	$0.00095 \pm 0.00001$	$0.000310 \pm 0.00002$	$0.00117 \pm 0.00001$	0.06
WTp32 ATV	$1.9 \times 10^7$	71,206 (0.37)	$0.00517 \pm 0.00004$	$0.00016 \pm 0.00001$	$0.0067 \pm 0.00003$	0.2
MAXp32 ATV	$2.5 \times 10^7$	180,514 (0.73)	$0.00946 \pm 0.00005$	$0.000356 \pm 0.00001$	$0.01252 \pm 0.00005$	0.27
WTp32 DRV	$4.1 \times 10^7$	148,066 (0.36)	$0.00482 \pm 0.00003$	$0.00016 \pm 0.00001$	$0.00624 \pm 0.00003$	0.16
MAXp32 DRV	$1.9 \times 10^7$	95,202 (0.73)	$0.00490 \pm 0.00004$	$0.00026 \pm 0.00001$	$0.00756 \pm 0.00004$	0.17

<sup>a</sup>Shannon's entropy value.

Interestingly, most of the starting synonymous mutations introduced to generate the MAX protease were not reverted after 32 passages, in none of the two conditions (with and without PIs). This and the results mentioned above, indicated that MAX viruses had a higher genetic diversity and a different evolutionary trajectory, after 32 passages, both in the presence or absence of PIs.



## **Chapter Two.** **Codon recoding of the envelope gene**

---

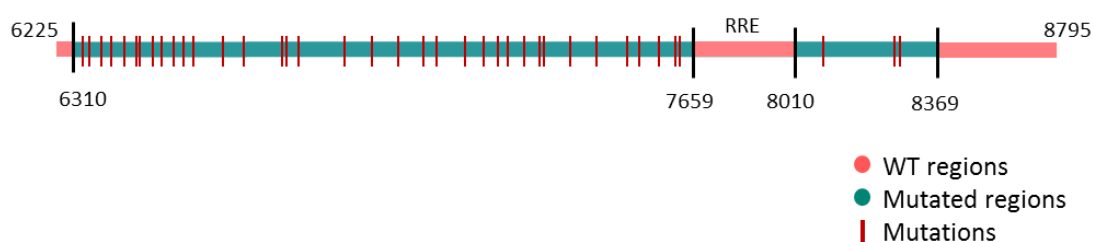


## Envelope recoding: changes in codon usage

The envelope of the HIV-1 HXB2 strain was synonymously recoded by changing these six codons AGG, GAG, CCT, ACT, CTC, GGG to these other six CGT, GAA, CCG, ACG, TTA, GGA, respectively, as previously described by Shin *et al.* (Shin *et al.* 2015). They performed the changes in the SIV *env* with the aim of exchanging the codon usage of the SIV *env* to the codon usage of rhesus monkey rhadinovirus (RRV). After performing the changes, they described a change in the *in vitro* SIV Env gene expression. Our aim here was to explore the *in vivo* effect with the HIV-1 Env.

The altered codons correspond to the amino acids arginine, glutamic acid, proline, threonine, leucine and glycine. To avoid a change in the reading frame of other genes and, therefore, to avoid non-viability, *env* overlapping regions were not altered; that is around 100 nts at the amino-terminal and 400 nts at the carboxi-terminal of the Env gene.

The resulting mutant, named Recoded-*env*, contained 39 mutations (1,5% of gp160; 0,83% of the HIV-1 genome) (Annex Figure 1), which affected a total of 34 codons. The 5' region (the region before the RRE) had approximately 1340 nts susceptible of being synonymously mutated. In Recoded-*env*, the 5' region contained 36 out of the 39 mutations. The last 3 mutations were located in the 3' region (the region after the RRE), which had almost 360 nts susceptible of being synonymously mutated (Figure 22).



**Figure 22. Schematic representation of the Recoded-*env* gene.** WT regions are shown in pink; mutated regions are shown in blue. The distribution of the introduced mutations is shown in red sticks. Each red sticks represents one single mutation: 36 of the introduced mutations were distributed in the 5' region and 3 in the 3' region.

An increased number of CpGs was observed in the Recoded-*env* (Table 9). While the WT-*env* had a total number of 26 CpGs (30% of the total number of CpGs in the HIV-1 HXB2 genome), the Recoded-*env* had 42 CpGs (around 42% of the total number of CpGs in the full-length virus variant genome) (Table 9).



## Results

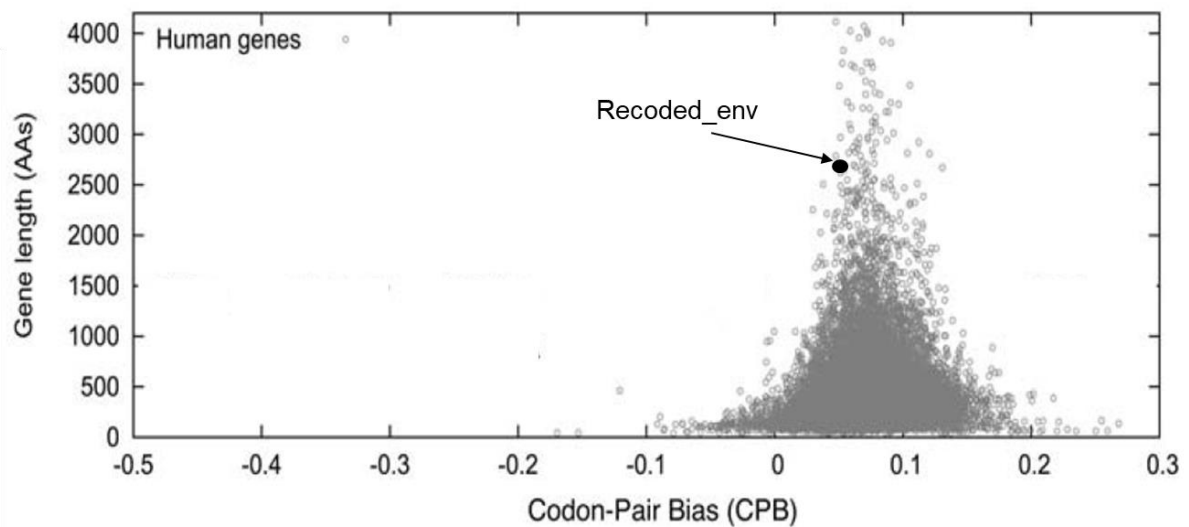
**Table 9. Dinucleotide comparative of the Recoded-env and WT\_env.**

Seq. name	N° mut	CpG	TpA	CPB*	CAI**
WT_env	-	26	194	0,046	0,13689505
Recoded-env	39	42	192	0,048	0,13158661

\*Codon-pair bias

\*\*Codon adaptation index. CAI values were obtained from [www.jcat.de](http://www.jcat.de)

The CPB value is the mean of every codon pair score within the gene. These scores are based on human values from 14.975 human genes (Coleman et al. 2008). While the WT\_env has a CPB value of 0,046, the Recoded-env had a CPB of 0,048 (Table 9). Both values are very similar between them and are annotated as neutral among the human genes (Figure 23).



**Figure 23. Calculated codon pair bias (CPB) score for all 14.795 annotated human genes.** Each dot represents the calculated CPB score of a human gene plotted against its amino acid length. Underrepresented codon pairs yield negative scores. The black spot indicates the CPB value for Recoded\_env. Adapted from Coleman et al. (Coleman et al. 2008).

Three of the introduced codons, ACG, CGT and CCG, are uncommon in the HIV-1 genome. Particularly, in the WT\_env, there are only two CCG codons and one CGT. After introducing the changes in Recoded-env, the number of CCG codons increased to five, and the number

of CGT codons increased to three. In Recoded-env, the frequency of appearance of these codons increased when compared to the WT\_env (Table 10).

**Table 10. Abundances and frequencies of codons in WT\_env and Recoded-env.**

Amino acids Codons	L		P		T		E		R		G	
	CTC	TTA	CCT	CCG	ACT	ACG	GAG	GAA	AGG	CGT	GGG	GGA
WT_env (total n°)	13	21	5	2	12	7	13	35	10	1	13	29
Recoded-env (total n°)	11	23	2	5	2	17	4	44	7	4	6	36
WT_env *	15,19	24,53	5,84	2,34	14,02	8,18	15,19	40,89	11,68	1,17	15,19	33,88
Recoded-env *	12,85	26,87	2,34	5,84	2,34	19,86	46,73	51,40	8,18	4,67	7,01	42,06
HIV-1 genome *	9,3	23,1	15	3,6	15,1	3,1	24,6	42,8	15,6	0,9	18,9	34,2

\* Frequency per thousand

### *Viral stability of the Recoded-env*

The amplified PCR of Recoded-env was co-transfected with the rest of PCR amplified regions of NL4-3 genome (fragments 1-6306 and 8369-9709) in MT-4 cells, as described above. The amplified WT\_env was also co-transfected as a positive control. After more than five blind passages in MT-4 cells without signs of virus replication, the Recoded-env virus was considered as non-viable. No syncytia nor p24 were detected in more than thirty days of culture.

### **New mutants based on Recoded-env**

In order to determine which mutations were responsible for this lethal phenotype, other mutants based on Recoded-env were designed and constructed. All new variants had the same amino acid sequence (Annex Figure 3)

The first approach was to reduce the number of mutations. This new mutant was called Recoded\_env\_mm. Reversions were randomly performed but equally distributed along gp160 (Figure 24). Recoded\_env\_mm had 21 mutations (Table 11, Annex Figure 2).

Next, the number of CpGs of Recoded-env was reduced. As previously suggested (Tulloch et al. 2014; Takata et al. 2017), an increased number of CpGs can impact virus replication capacity. These authors suggested that increasing the number of CpGs allows some factors of the innate immune system to recognize better the virus. This new variant (Recoded\_env\_sCpG) reverted 19 mutations and reduced the number of CpGs to 26 (Table 11, Annex Figure 2). Since this mutant also reduced the total number of mutations, a second

## Results

variant was designed considering both the number of mutations and the total number of CpGs. This new *env* was named Recoded\_ env\_sCpG2. The modifications performed did not include reversions to WT but, instead, changing the nucleotides to different synonymous substitutions (Annex Figure 2) (Figure 24). Recoded\_ env\_sCpG2 had 36 mutations, and the total number of 26 CpGs (Table 11).

Near the beginning of the sequence of the RRE is where gp120 ends and gp41 starts. For that reason, the RRE was considered the middle of our constructs, from which both the 5' and 3' regions were distinguished. We wanted to explore if the sequence alteration of gp120 or gp41, independently, induced the lethality. Thus, the mutations of gp120 (5' region) or gp41 (3' region) were reverted separately. The first variant was Recoded\_ env\_5'wt, which reverted the 36 mutations in the 5' region (Annex Figure 2). This virus had 3 mutations in total (Figure 24). The second variant was Recoded\_ env\_3'wt which reverted the 3 mutations in the 3' region and conserved the 5' region mutated (Figure 24; Annex Figure 2) (Table 11).

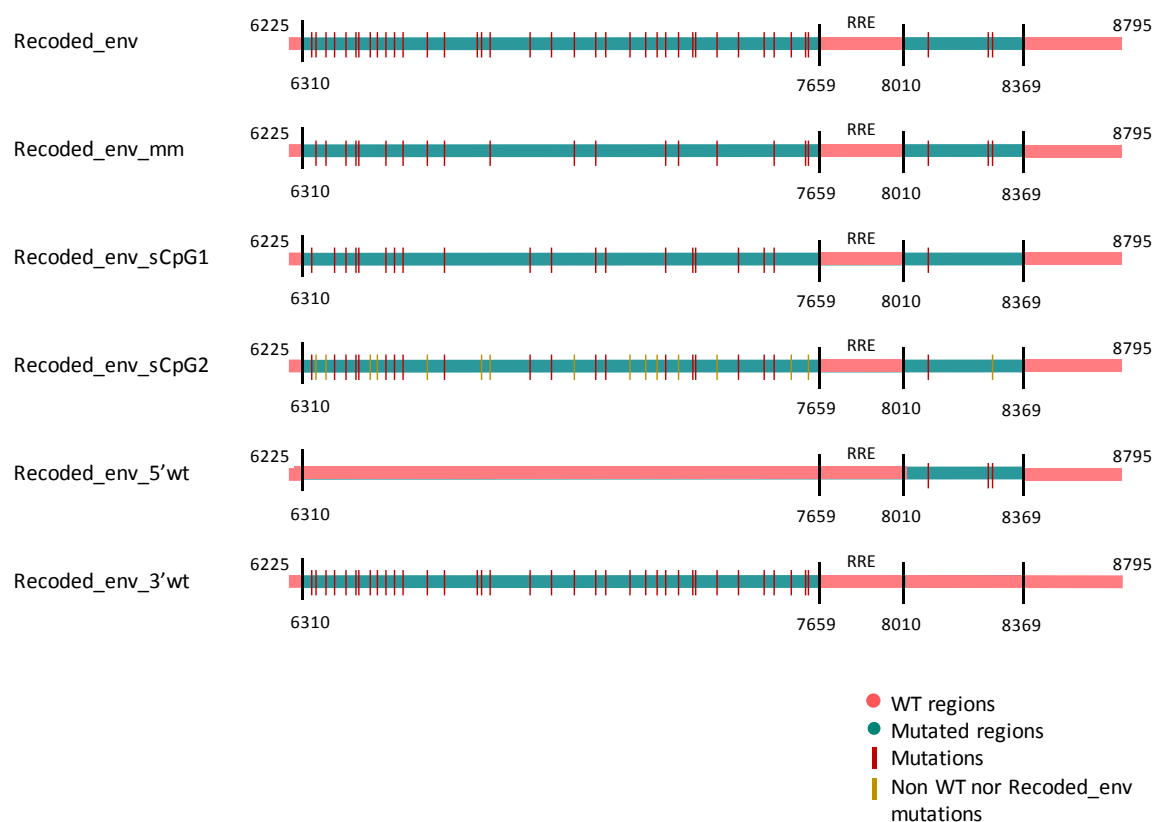
CPB and CAI values were maintained similar to WT in all new variants. The number of TpAs of all the new variants (including Recoded-env) was maintained very similar to WT (Table 11).

**Table 11. Calculated number of mutations, CPB, CAI, and number of CpGs and TpAs of the virus variants based on Recoded-env.**

Seq. name	Nº mut	CpG	TpA	CPB*	CAI**
WT_ env	-	26	194	0,046	0,1368950
Recoded-env	39	42	192	0,048	0,1315866
Recoded_ env_mm	21	35	193	0,046	0,1340826
Recoded_ env_sCpG	20	26	196	0,053	0,1295865
Recoded_ env_sCpG2	36	26	194	0,055	0,1291150
Recoded_ env_5'WT	3	27	193	0,047	0,1360499
Recoded_ env_3'WT	36	41	193	0,046	0,1320015

\*Codon-pair bias

\*\*Codon adaptation index. CAI values were obtained from [www.jcat.de](http://www.jcat.de)



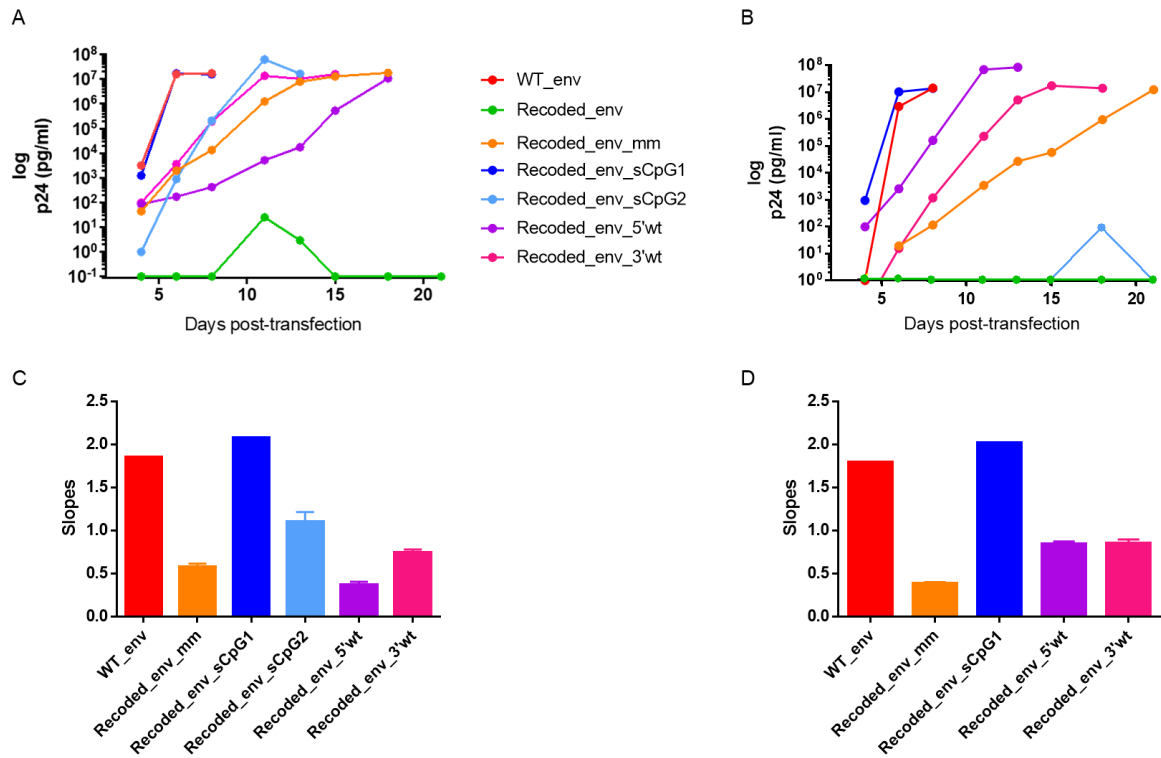
**Figure 24. Schematic representation of the mutants based on Recoded-env.** WT regions are shown in pink. Mutated regions are shown in blue. Red sticks represent each mutation. Yellow sticks represent the mutations in Recoded\_env\_sCpG2 which were different from Recoded-env.

### *Viral stability of the new designed mutants*

All Recoded-env-based new mutants were PCR amplified, as previously described, and co-transfected in MT-4 cells with the rest of NL4-3 genome. The transfection was performed in duplicate. After transfection, culture supernatant was collected every 2-3 days and quantification of HIV-1 p24 was performed.

All of them generated p24, but at different days after transfection and with different concentrations (Figure 25). At day 6 after transfection, the WT\_env and the Recoded\_env\_sCpG1 had similar amounts of p24 in both replicates. The other variants, at day 6 post transfection, started to produce p24, but the concentration was much lower. The rest of the variants reached WT p24 amounts after several days of culture. Some of them needed almost 20 days to reach that same amount (Figure 25 A and B). This result suggested different replication capacities among the different variants.

## Results



**Figure 25. Log of p24 values for the different viruses and days after transfection and the slope for each curve.** A) Representation of p24 amounts obtained at different days from replicate one. B) Representation of p24 amounts obtained at different days from replicate two. C) Slopes obtained from the exponential phase of each growing virus from replicate one. D) Slopes obtained from the exponential phase of each growing virus from replicate two. Recoded-env was not detected in none of the replicates, while Recoded\_env\_sCpG2 was only detected in replicate one.

Similarities and differences were observed between the two replicates. Both WT\_env and Recoded\_env\_sCpG1 had the same slope in both replicates. Moreover, both viruses arrived to plateau phase the same day. Also, Recoded\_env\_3'wt had the same slope in both replicates, although this virus reached plateau phase a few days after WT\_env (Figure 25 C and D). Differences were observed in Recoded\_env\_mm, Recoded\_env\_sCpG2 and Recoded\_env\_5'wt. For Recoded\_env\_mm and Recoded\_env\_5'wt, in one of the two replicates, the slope was steeper. In the case of Recoded\_env\_sCpG2, the virus did not grow in one of the replicates (Figure 25 C and D).

RNAs were extracted from culture supernatant at different days post-transfection and RT-PCR amplification of the different envs was performed. After sequencing, almost all virus variants generated mutations. The only construct that did not generate mutations was Recoded\_env\_sCpG1 from replicate 1 (Table 12). This result suggested that the introduced mutations did not allow proper replication for the viruses and they had to compensate them.

It was also observed that some mutations appeared very soon after transfection (day 6 or day 8) (e.g. Recoded\_env\_sCpG2).

**Table 12. Mutations found after sequencing the viruses at different days post-transfection.**

Replicate 1	Days post transfection	Virus name	Position of Mutation*	Mutation	Syn/Non-syn**
	6	Recoded_env_sCpG2	1050	A/G	Yes
	8	Recoded_env_mm	2299	A/G	S/G
		Recoded_env_sCpG2	1050	A/G	Yes
		Recoded_env_3'wt	453	A/G	Yes
	18	Recoded_env_mm	2299	A/G	S/G
		Recoded_env_5'wt	2518	A/G	I/V
Replicate 2	Days post transfection	Virus name	Position of Mutation*	Mutation	Syn/Non-syn**
	6	Recoded_env_sCpG1	568	A/G	S/G
			1809	T/C	Yes
	8	Recoded_env_sCpG1	568	A/G	S/G
			1809	T/C	Yes
		Recoded_env_3'wt	503	A/G	K/R
	11	Recoded_env_5'wt	678	A/G	Yes
	21	Recoded_env_mm	909	A/C	Yes

\*The positions of the mutations are given related to the beginning of the env.

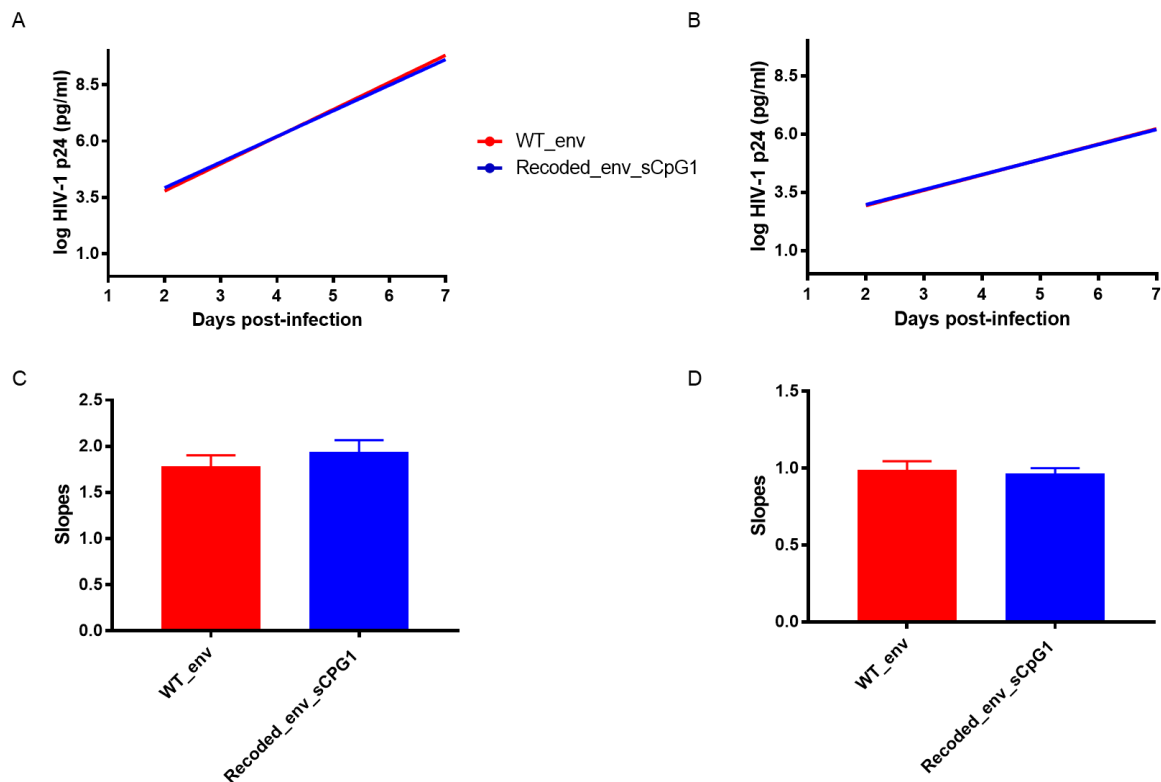
\*\*The change of amino acid is given when the mutation is not synonymous.

### Replication capacity of the recovered viruses

Replication capacity assays were performed with those viruses that surpassed 500ng/ml of p24 at the same day post-transfection as did the WT\_env. Only Recoded\_env\_sCpG1 had enough p24 at day 6 after transfection. At day 8 after transfection (the last day in which WT\_env was collected), none of the other virus variants reached 500ng/ml of p24 nor showed viral titer.

After titration, WT\_env and Recoded\_env\_sCpG1 collected from day 6 post transfection were used to infect MT-4 cells and PBMCs at the same MOI. Recoded\_env\_sCpG1 showed a replication capacity indistinguishable from WT\_env, both in MT-4 cells ( $p = 0,4308$ ) and PBMCs ( $p=0,7556$ ) (Figure 26).

## Results



**Figure 26. Replication capacity assays for WT\_env and Recoded\_env\_sCpG1.** Values represent the means plus SDs of three independent experiments. Lag and plateau phases were not considered in slopes to avoid their influence. A) Replication capacity in MT-4 cells shown as the log of p24 antigen measurements. B) Replication capacity in PBMCs shown as the log of p24 antigen measurements. C) Slopes of the viruses obtained in the exponential phase of growth after infection of MT-4 cells. D) Slopes of the viruses obtained in the exponential phase of growth after infection of PBMCs. P-values (student's t-test) were obtained from comparing the slope of Recoded\_env\_sCpG1 to the slope of WT\_env.

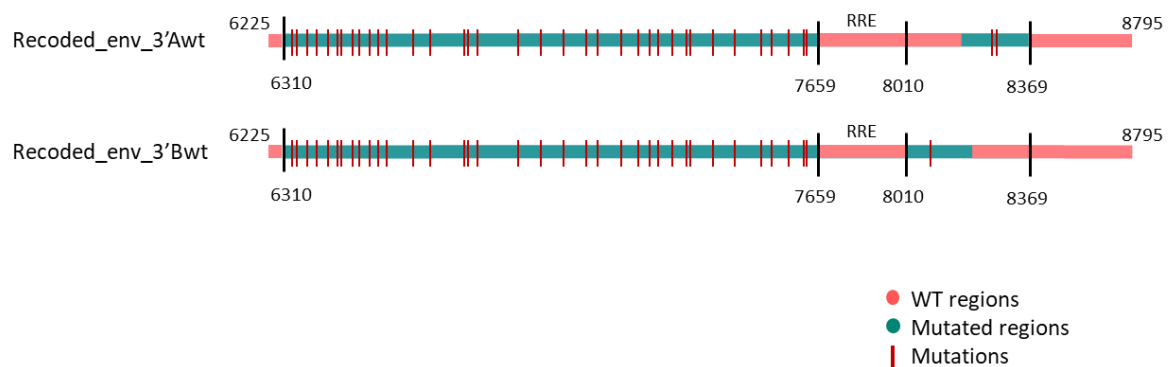
This result demonstrated that reducing the number of CpGs, thus reducing the total number of mutations of Recoded-env, let the virus to replicate as the WT. It also suggested that the introduction of a small number of mutations (20 in total) did not alter the virus phenotype. Nevertheless, the influence of specific mutations on virus replication capacity cannot be discarded (see below).

### *Reversion of mutations in the 3' region*

Both Recoded\_env\_5'wt and Recoded\_env\_3'wt generated compensatory mutations after transfection (Table 12). Of note, Recoded\_env\_5'wt had only 3 mutations, while Recoded\_env\_3'wt had 36 mutations. Moreover, after comparing the growing slopes of both replicates (Figure 25 C and D), Recoded\_env\_3'wt had higher slope than

Recoded\_env\_5'wt. These observations indicated that mutations in the 3' region were important for viral replication.

For this reason, we tested whether the reversion of the mutations in the 3' region may affect the virus replication capacity. To this end, the 3' region was divided into two (3'A and 3'B). The 3'A referred to the first mutation of the 3' region (codon 33), while the 3'B referred to the last 2 substitutions which belonged to the same codon (codon 34). Both 3'A and 3'B were reverted independently, conserving all the changes in the 5' region. These new viruses were named as Recoded\_env\_3'Awt (5' region and 3'B were mutated) and Recoded\_env\_3'Bwt (5' region and 3'A were mutated) (Figure 27; Annex Figure 2).



**Figure 27. Schematic representation of the mutants based on Recoded\_env\_3'wt.** WT regions are shown in pink. Mutated regions are shown in blue. Red sticks represent the mutations. Recoded\_env\_3'Awt reverted the mutation of codon 33, while Recoded\_env\_3'Bwt reverted the two nucleotides of codon 34.

Recoded\_env\_3'Awt had 38 mutations and 42 CpGs, while Recoded\_env\_3'Bwt had 37 mutations and 41 CpGs. Both variants had WT CAI and CPB scores (Table 13). The number of TpAs was also almost the same as WT\_env. Indeed, these two new variants had very similar characteristics as Recoded-env.

These new constructs were PCR amplified and co-transfected in MT-4 cells together with the rest of the NL4-3 genome. Again, WT\_env was transfected as a positive control. As mentioned before, culture supernatants were collected during several days and p24 amounts were quantified. In this case, only Recoded\_env\_3'Bwt replicated (Figure 28 A and B). The slopes of Recoded\_env\_3'Bwt and WT\_env were identical in the two replicates (Figure 28 C and D).



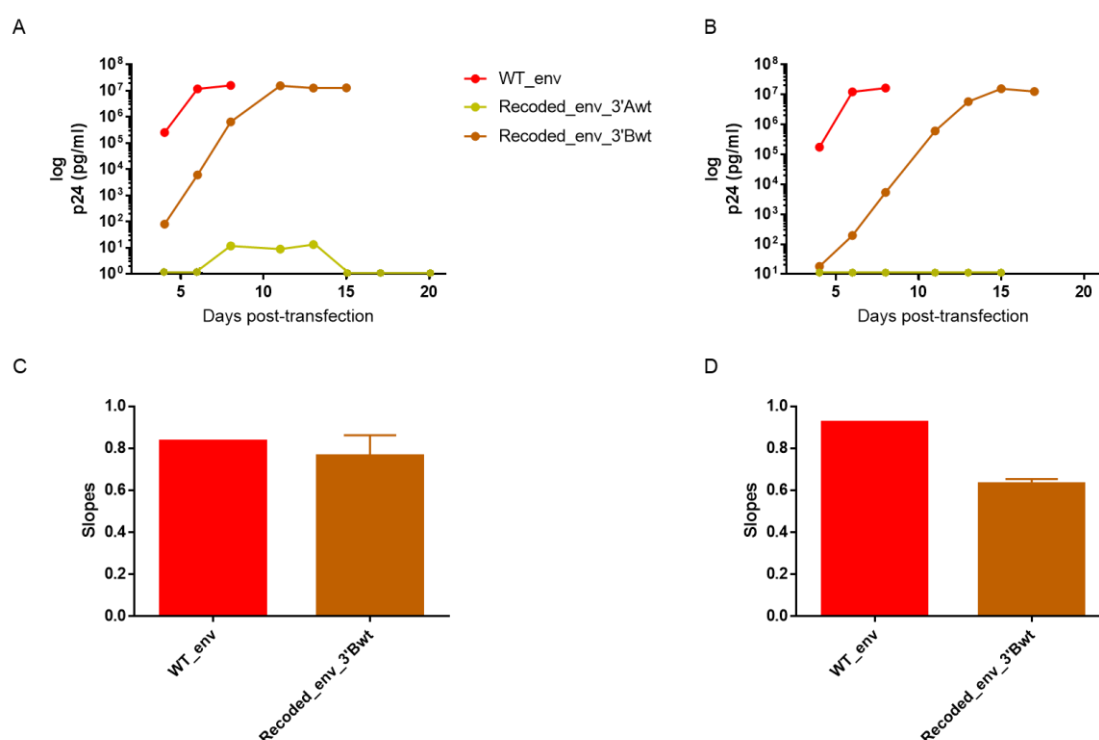
## Results

**Table 13.** Calculated number of mutations, CPB, CAI, and number of CpGs and TpAs of the virus variants Recoded\_env\_3'Awt and Recoded\_env\_3'Bwt.

Seq. name	Nº mut	CpG	TpA	CPB*	CAI**
WT_env	-	26	194	0,046	0,1368950
Recoded-env	39	42	192	0,048	0,1315866
Recoded_env_3'Awt	38	42	192	0,048	0,1319750
Recoded_env_3'Bwt	37	41	193	0,046	0,1316135

\*Codon-pair bias

\*\*Codon adaptation index. CAI values were obtained from [www.jcat.de](http://www.jcat.de)

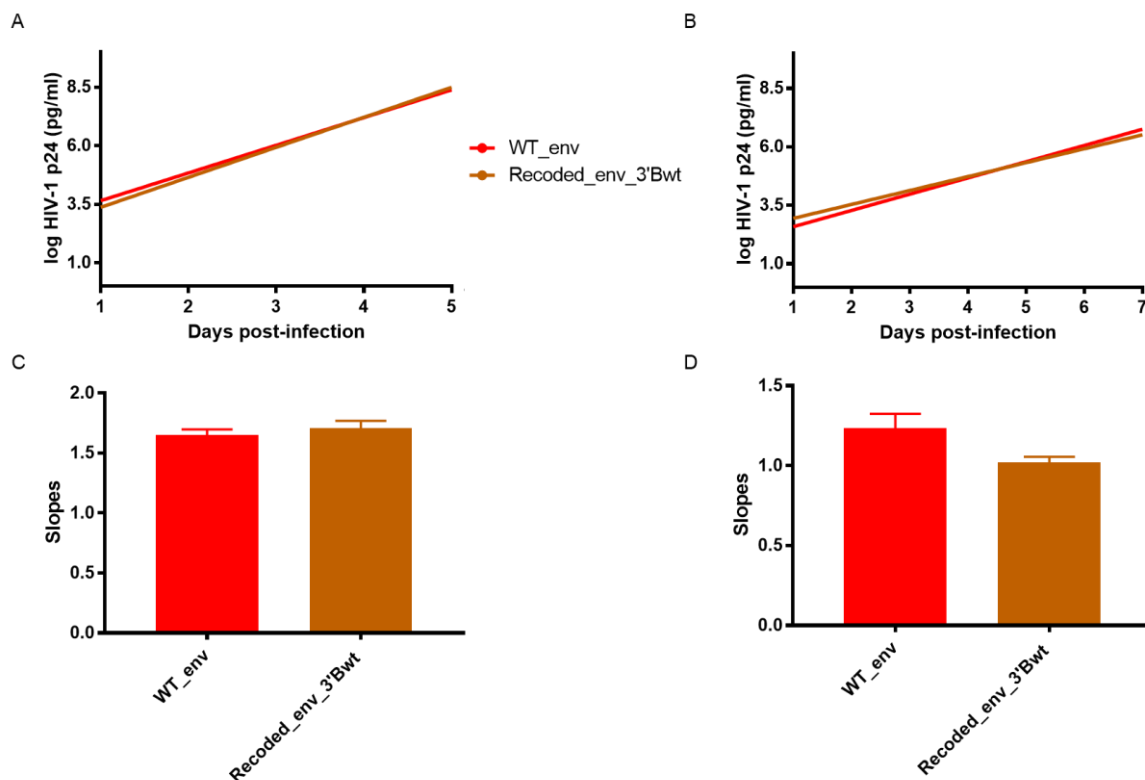


**Figure 28.** Log of antigen p24 values after transfection of the viruses which reverted 3'A or 3'B region and the slopes for each curve. A) Representation of p24 amounts obtained at different days from replicate one. B) Representation of p24 amounts obtained at different days from replicate two. C) Slopes obtained from the exponential phase of each growing virus from replicate one. D) Slopes obtained from the exponential phase of each growing virus from replicate two.

To determine if the viruses grew with the designed sequence, RNAs were extracted from several days post-transfection and RT-PCRs were performed. After sequencing, Recoded\_env\_3'Bwt from replicate one only had one mutation at the beginning of the sequence (position 404, non-synonymous mutation (K/R)), while the same virus from replicate two did not generate any mutation.

### Replication capacity of the obtained virus

Replication capacity assays were performed with the new obtained virus. At day 8 after transfection, only Recoded\_env\_3'Bwt from replicate one surpassed the amount of 500ng/ml of p24. Infections of MT-4 cells and PBMCs were carried out with the same MOI for both viruses. Recoded\_env\_3'Bwt showed identical replication capacity as the WT\_env in MT-4 cells ( $p=0,5372$ ) and PBMCs ( $p=0,0855$ ) (Figure 29).



**Figure 29. Replication capacity assays for WT\_env and Recoded\_env\_3'Bwt.** Values represent the means plus SDs of three independent experiments. Lag and plateau phases were not considered in slopes to avoid their influence. A) Replication capacity in MT-4 cells shown as the log of p24 antigen measurements. B) Replication capacity in PBMCs shown as the log of p24 antigen measurements. C) Slopes of the viruses obtained in the exponential phase of growth after infection of MT-4 cells. D) Slopes of the viruses obtained in the exponential phase of growth after infection of PBMCs. P-values (student's t-test) were obtained from comparing the slope of Recoded\_env\_3'Bwt to the slope of WT\_env.

Since Recoded\_env\_3'Bwt from replicate two had the same slope as Recoded\_env\_3'Bwt from replicate one (Figure 28 C and D), it can be assumed that the mutation found in replicate one was not affecting the replication capacity.

## Results

---

The obtained result suggested that mutating codon 34 was the reason why Recoded-env could not replicate properly. Moreover, mutations in 5' region did not seem to strongly affect virus replication capacity.

### *Viral stability of all mutants*

To corroborate the obtained results, all constructs (including Recoded\_env\_3'Awt and Recoded\_env\_3'Bwt) were transfected together in a third independent experiment. As previously described, all variants were PCR amplified and transfected in MT-4 cells.

After several days post-transfection all variants grew, except Recoded-env and Recoded\_env\_3'Awt, confirming previous results.

To ensure having enough titer for all viruses, they were recovered when p24 amounts surpassed the quantity of 500ng/ml. WT\_env and Recoded\_env\_sCpG1 were recovered at day 4 and 5 after transfection, respectively. The rest of the variants were recovered after day 8 post transfection (Table 14).

**Table 14. Day of recovery and HIV-1 p24 amounts of the variants based on Recoded-env after transfection.**

Seq. name	Day of recovery after transfection	p24 amounts (pg/ml)
WT_env	4	2,3x10 <sup>7</sup>
Recoded-env	-	-
Recoded_env_mm	8	2,1x10 <sup>7</sup>
Recoded_env_sCpG1	5	1,7x10 <sup>7</sup>
Recoded_env_sCpG2	10	1,6x10 <sup>7</sup>
Recoded_env_5'wt	9	9,3x10 <sup>6</sup>
Recoded_env_3'wt	11	1,3x10 <sup>7</sup>
Recoded_env_3'Awt	-	-
Recoded_env_3'Bwt	11	9,9x10 <sup>6</sup>

RNA was extracted from all viruses and RT-PCRs were performed. After sequencing, some of the constructs generated or reverted mutations, as it happened in the previous experiments (Table 12). However, all these changes had something in common: they all were located near or in the two nucleotides of codon 34. Codon 34 is the one reverted in Recoded\_env\_3'Bwt that allowed the replication of the virus (Figure 28). Recoded\_env\_mm

generated a new mutation next to codon 34 that implied a change in amino acid (non-synonymous mutation). Recoded\_env\_5'wt reverted both mutations in codon 34. Interestingly, this mutant had all 5' region WT, so the obtained virus only had 1 mutation when compared to WT\_env. This result demonstrated the importance of the modification of codon 34. The rest of the viruses grew with the designed sequence (Table 15).

**Table 15. Partial alignment of the sequence in 3' region where codon 34 is located.**

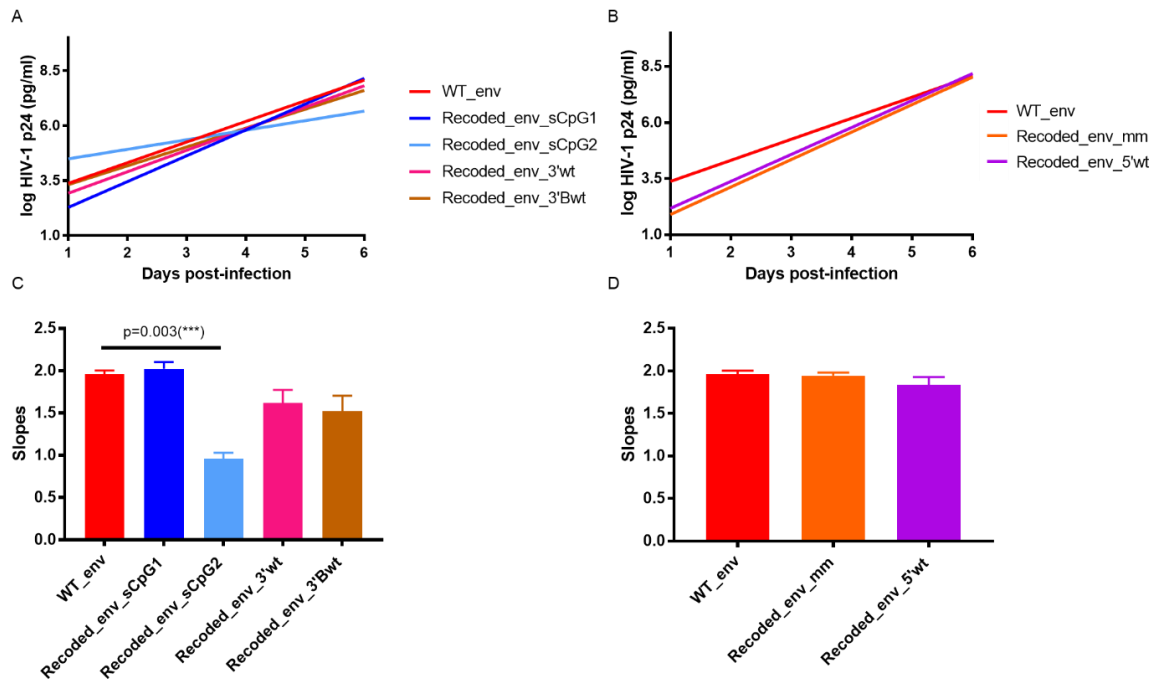
Sequence name	Designed sequence	Sequences in culture*
WT_env	AGGCAGG	AGGCAGG
Recoded_env_mm	C.T....	C.T. <span style="border: 1px solid red; padding: 0 2px;">G</span> ..
Recoded_env_sCpG1	.....	.....
Recoded_env_sCpG2	..A....	..A....
Recoded_env_5'wt	C.T....	<span style="border: 1px solid red; padding: 0 2px;">.</span> <span style="border: 1px solid red; padding: 0 2px;">.</span> ...
Recoded_env_3'wt	.....	.....
Recoded_env_3'Bwt	.....	.....

\*In red, nucleotides of the sequences that mutated after transfection.

After recovering all viruses and titrating them in MT-4 cells, replication capacity assays were carried out in MT-4 cells. In this assay all viruses were included, but analyzed separately.

Recoded\_env\_sCpG1, Recoded\_env\_3'wt and Recoded\_env\_3'Bwt showed a replication capacity identical to WT ( $p=0,5229$ ,  $p=0,1367$  and  $p=0,2644$ , respectively). All the three virus variants grew without modifications. In contrast, Recoded\_env\_sCpG2, which also grew without generating new mutations, had significantly lower replication capacity when compared to WT ( $p=0,003$ ) (Figure 30 A and C). This variant, Recoded\_env\_sCpG2, had almost the same number of mutations as Recoded-env (36 and 39, respectively). The difference between Recoded-env and Recoded\_env\_sCpG2 was that Recoded\_env\_sCpG2 decreased the number of CpGs by generating new, different, synonymous mutations. These changes performed in Recoded\_env\_sCpG2 generated a synonymous mutation in codon 34 (AGA), which is not the sequence observed in WT\_env (AGG) nor in Recoded-env (CGT) (Table 15). Once again, it can be observed that altering codon 34 might have an effect in viral replication.

## Results

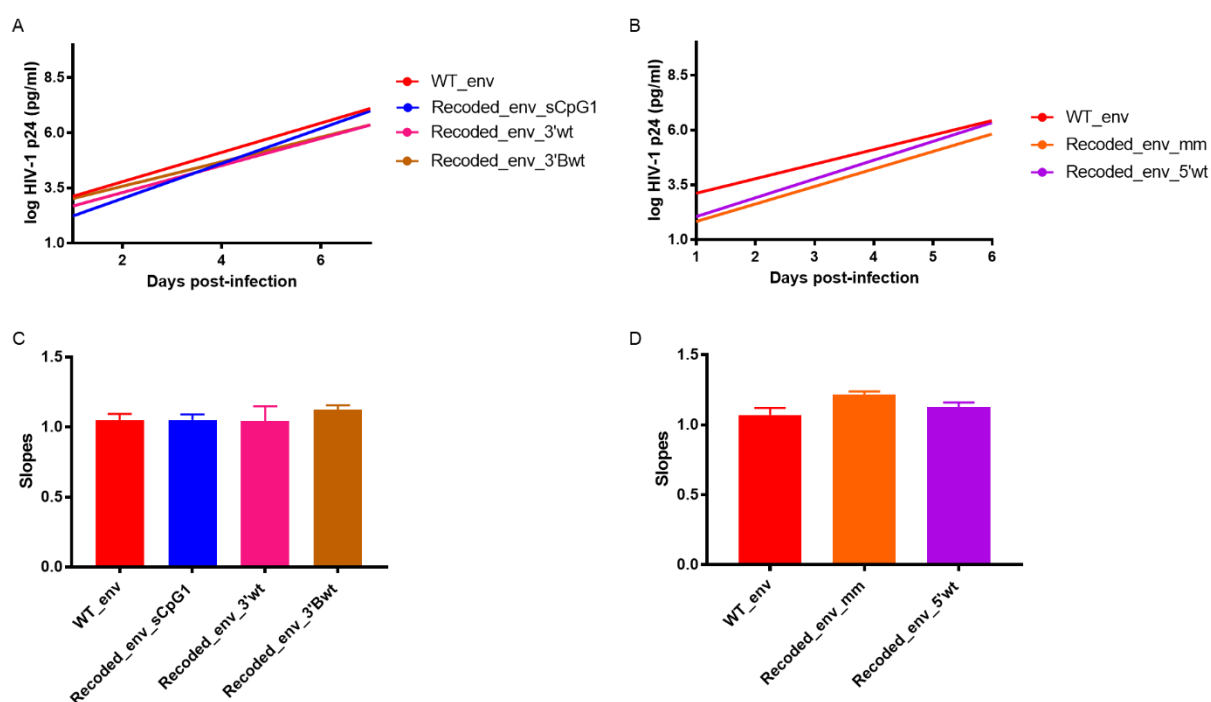


**Figure 30. Replication capacity assays in MT-4 cells.** Values represent the means plus SDs of three independent experiments. Lag and plateau phases were not considered in slopes to avoid their influence. A) Replication capacity shown as the log of p24 antigen measurements for the viruses that did not generate any mutation or reversion in culture after transfection. B) Replication capacity shown as the log of p24 antigen measurements for the viruses that generated or reverted mutations in culture after transfection. C) Slopes of the viruses shown in A) obtained from the exponential phase of growth after infection of MT-4 cells. D) Slopes of the viruses shown in B) obtained from the exponential phase of growth after infection of MT-4 cells. P-values (student's t-test) were obtained from comparing the slope of each virus to the slope of WT\_env.

Recoded\_env\_mm and Recoded\_env\_5'wt, the two virus variants that generated or reverted mutations in culture (Table 15), replicated indistinguishable from WT\_env ( $p=0,7873$  and  $p=0,2129$ , respectively) (Figure 30 B and D). Reverting the nucleotides of codon 34 in Recoded\_env\_5'wt allowed the virus to replicate. The same happened with the Recoded\_env\_mm, which generated a new mutation next to codon 34 that allowed the replication of this virus variant.

The next step was to infect PBMCs with the viruses that did not show differences in MT-4 cells when compared to WT\_env.

As shown in MT-4 cells, Recoded\_env\_sCpG1 ( $p=0,9750$ ), Recoded\_env\_3'wt ( $p=0,9610$ ) and Recoded\_env\_3'Bwt ( $p=0,2375$ ) displayed indistinguishable replication capacities in PBMCs when compared to WT\_env (Figure 31 A and C). The virus variants Recoded\_env\_mm and Recoded\_env\_5'wt neither showed differences in PBMCs ( $p=0,0592$  and  $p=0,4229$ , respectively) (Figure 31 B and D).



**Figure 31. Replication capacity assays in PBMCs cells.** Values represent the means plus SDs of three independent experiments. Lag and plateau phases were not considered in slopes to avoid their influence. A) Replication capacity shown as the log of p24 antigen measurements for the viruses that did not generate any mutation or reversion in culture after transfection. B) Replication capacity shown as the log of p24 antigen measurements for the viruses that generated or reverted mutations in culture after transfection. C) Slopes of the viruses shown in A) obtained from the exponential phase of growth after infection of PBMCs. D) Slopes of the viruses shown in B) obtained from the exponential phase of growth after infection of PBMCs. P-values (student's t-test) were obtained from comparing the slope of each virus to the slope of WT\_env.

The results obtained in PBMCs confirmed the previous results obtained in MT-4 cells. Additionally, replication capacities performed in PBMCs reproduce more faithfully the physiological conditions. This means that all these virus variants may show similar replication capacities *in vivo* as the WT.

## Infectivity capacity

All constructs were cloned in the expression vector pcDNA3.1V5/his TOPO. They were co-transfected with the single-cycle infectious plasmid pNL43-dENV-EGFP in 293T cells to determine the infectivity capacity of the different variants. Forty-eight hours after transfection, supernatants were collected and concentrated. Antigen p24 was determined in order to infect TZM-bl with the same amounts of viral particles (1000ng/ml). The result was given in TCID<sub>50</sub>/ml as the average of three independent experiments (Table 16).

## Results

---

Statistical analyses were performed to the obtained data. In TZM-bl assays, the higher the value, the more infectivity. None of the virus variants displayed significant differences in infectivity when compared with WT\_env, except for Recoded\_env\_mm. This virus variant showed higher infectivity when compared to WT\_env ( $p=0,0318$ ). However, this virus variant had to generate compensatory mutations (Table 12, Table 15), indicating that replication in this mutant was affected in other steps.

Overall, the obtained result suggested that none of the virus variants were affected in infectivity.

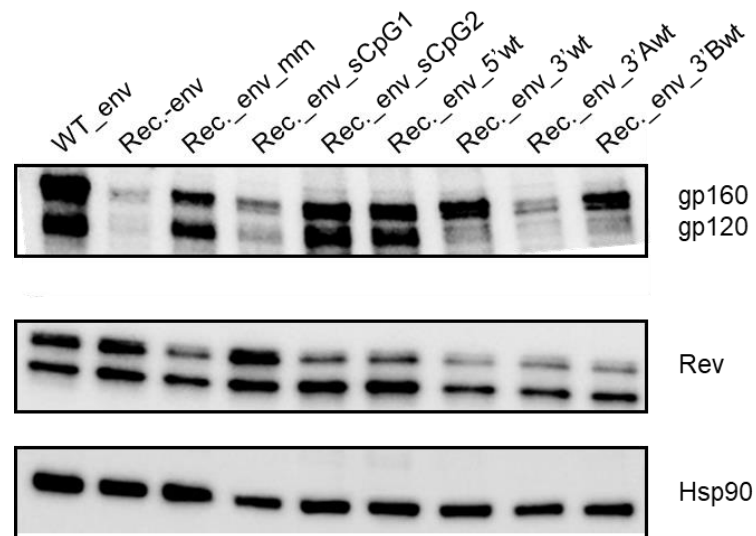
**Table 16. Viral infectivity in TZM-bl.**

Seq. name	TCID50/ml (mean)
WT_env	1809,02
Recoded_env	2696,49
Recoded_env_mm	5590,17
Recoded_env_sCpG1	3526,09
Recoded_env_sCpG2	2868,47
Recoded_env_5'wt	4932,55
Recoded_env_3'wt	1118,03
Recoded_env_3'Aw	1118,03
Recoded_env_3'Bwt	677,49

Values are shown as the mean of three independent experiments

### Effects on translation: protein expression

In order to determine if the protein expression of the different Env variants was affected, expression experiments were carried out. To this end, Env variants were cloned in the expression plasmid pcDNA3.1V5/his TOPO, as described above. Each corresponding mutant plasmid was co-transfected in 293T cells with pTOPORev.



**Figure 32. Immunoblot analyses of the amounts of protein for WT\_env and the mutants based on Recoded\_env.** All mutants were co-transfected with pTOPORev. Rev is necessary for the expression of HIV-1 Env (Cullen 1991; Malim et al. 1989, 1990; Malim and Cullen 1991). Forty-eight hours after transfection cells were collected and proteins were extracted from the cellular pellet. Briefly, 20µg of total protein were loaded in the well and proteins were separated by electrophoresis. Hsp90 was labeled as housekeeping. Env and Rev were also detected. Rev is shown with two bands that correspond to the two isoforms obtained from the alternative splicing of the virus.

A dramatic reduction in Recoded\_env and Recoded\_env\_3'Awt protein expressions were observed. Both gp160 and gp120 of these two virus variants were barely detected by the immunoblot, while Rev levels were very similar to WT (Figure 32). This result matches the lethal phenotype of both virus variants.

Env and Rev protein amounts were normalized to each corresponding Hsp90 amounts. Then, normalized Env amounts were compared to normalized Rev amounts for each virus variant to obtain the differences between the variants and WT\_env. After these calculations were performed, it was shown that Recoded-env amount was reduced a 98% when compared to WT\_env demonstrated that translation of this variant was affected. The same result was observed for Recoded\_env\_3'Awt, whose total protein amount was 5,2% when compared to WT\_env (Table 17).



## Results

Table 17. Total protein percentages of Recoded-env and variants compared to WT.

Seq. name	Env/Rev (normalization)	% compared to WT
WT_env	2,54	100,00
Recoded-env	0,07	2,57
Recoded_env_mm	0,50	19,61
Recoded_env_sCpG1	0,23	9,19
Recoded_env_sCpG2	0,62	24,47
Recoded_env_5'wt	0,52	20,58
Recoded_env_3'wt	1,20	47,24
Recoded_env_3'Awt	0,13	5,20
Recoded_env_3'Bwt	0,99	39,16

### Effects on transcription: mRNA levels

To know whether the transcription of the Env variants was affected, the amounts of each mRNA variant were quantified.

Both *env* and *rev* were relatively quantified to GAPDH by the  $2^{-\Delta\Delta Ct}$  method (Livak and Schmittgen 2001) (Table 18). Following this method, *env* amounts were related to *rev* (*env/rev*). This assay was performed in triplicate and the results are shown as the mean for the three experiments with their standard deviations (SD) (Figure 33).

Table 18. Real-time relative RT-PCR for *env* and *rev* mRNA amounts.

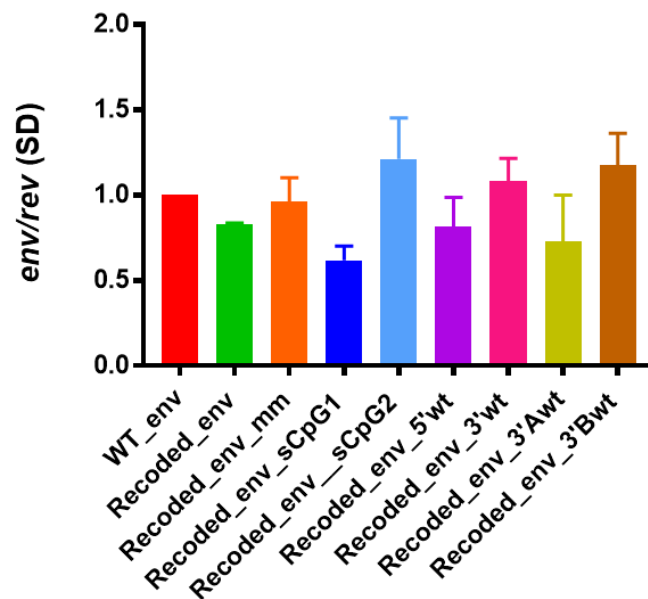
Seq. Name	Replicate 1		Replicate 2		Replicate 3	
	$2^{-\Delta\Delta Ct}$ ( <i>env</i> /GAPDH)	$2^{-\Delta\Delta Ct}$ ( <i>rev</i> /GAPDH)	$2^{-\Delta\Delta Ct}$ ( <i>env</i> /GAPDH)	$2^{-\Delta\Delta Ct}$ ( <i>rev</i> /GAPDH)	$2^{-\Delta\Delta Ct}$ ( <i>env</i> /GAPDH)	$2^{-\Delta\Delta Ct}$ ( <i>rev</i> /GAPDH)
WT_env	1,00	1,00	1,00	1,00	1,00	1,00
Rec._env	1,13	1,36	0,83	0,99	0,54	0,66
Rec._env_mm	1,09	1,15	0,88	1,07	0,72	0,65
Rec._env_sCpG1	0,48	0,91	0,29	0,42	0,55	0,84
Rec._env_sCpG2	0,84	0,81	0,69	0,62	0,98	0,66
Rec._env_5'wt	0,54	0,86	0,58	0,67	0,70	0,73
Rec._env_3'wt	0,84	0,88	0,76	0,70	0,72	0,59
Rec._env_3'Awt	0,50	0,78	0,33	0,32	0,27	0,53
Rec._env_3'Bwt	0,84	0,63	0,68	0,55	0,68	0,70

Viral genes were related to the cellular gene GAPDH and quantitated by the  $2^{-\Delta\Delta Ct}$  method.

The results indicated that the amounts of *env* mRNA were directly related to the amounts of *rev* mRNAs (Table 18). In the case of Recoded\_env\_sCpG1 and Recoded\_env\_3'Awt there was less amount of *env* mRNA when compared to the amounts of *rev* mRNA.

Recoded\_env\_sCpG1 virus variant did not show any replication difference when compared to WT\_env (Figure 26). Thus, the reduced amount of mRNA did not affect virus phenotype.

Overall, these results showed that the loss of protein expression was not due to a reduction in mRNA expression. Thus, the mechanism that was affected in the different recoded HIV-1 Env variants was translation.



**Figure 33. mRNA amounts of the variants, obtained from the division of *env* amounts by *rev* amounts.** The graph shows the mean of three independent experiments with the SDs. All samples were related to WT\_env. Total RNA was extracted from 293T cells transfected with the Env-expression plasmids and pTOPORev. Following extraction, cDNA was synthesized and used as template for a real-time PCR. The cellular gene GAPDH was used as the control gene.

## RNA secondary structures

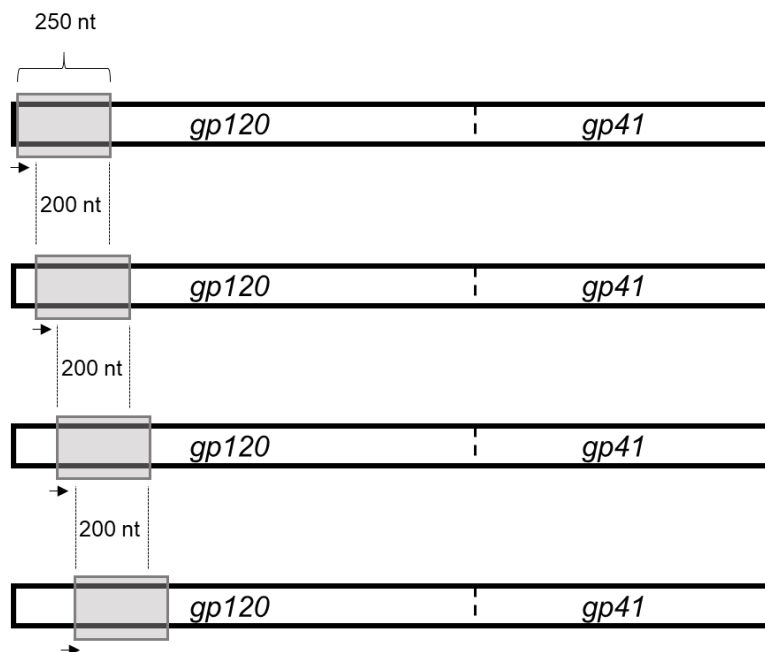
Previous results showed that Recoded-env was lethal for virus replication. However, when two mutations of codon 34 (CGT), located in gp41, were reverted to WT (AGG), the replication capacity was recovered. This suggested that codon 34 was the responsible for the observed phenotype. We hypothesized that this codon may effect Env mRNA secondary structure. Consequently, several bioinformatics tools were used to see if those two mutations were affecting the RNA secondary structure.

The RNAfold tool was used to predict the RNA secondary structures of the WT\_env and different variants. HXB2 envelope has 2590 nucleotides. Obtaining just a single structure for the entire gene is not reliable, due to all possible interactions between sequence nucleotides. Therefore, the best approach is to analyze small local structures. To this end,

## Results

---

a window approach was performed in which the sequence was divided in small regions. Every 50 nucleotides there was a new window of length of 250 nucleotides. Each new window overlapped with 200 nucleotides of the two adjacent ones (Figure 34).



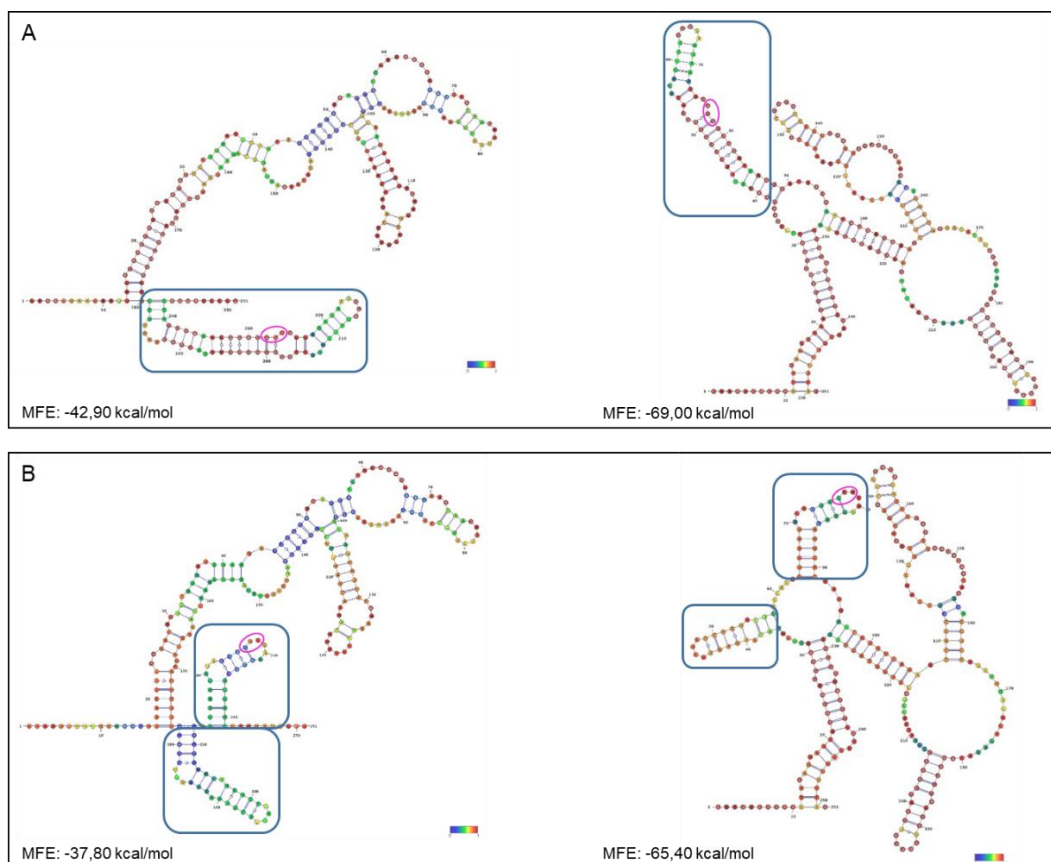
**Figure 34. Schematic representation of the approach performed to generate different windows of partial sequence of the Env coding region.** The square in grey represents the size of the window, which was moved every 50 nts, overlapping with 200 nts of the next section.

Dividing the sequence in different fragments of 250 nts of length resulted in 5 windows that contained codon 34. RNAfold was applied to all 5 windows, in order to know exactly how the distribution of the nucleotides was. Performing RNAfold to just one fragment or window of the sequence may not correctly predict the structure. The reason is that every structure changes within the surrounding nucleotides. This means that, when different windows are analyzed, different structures are obtained. The conformation of codon 34 that is more frequent among the windows and with more stability will be the expected conformation.

After applying RNAfold to WT\_env sequence windows that contained codon 34, it was determined that the A nucleotide was unpaired in the structure, and the two Gs were paired. This organization was seen in almost all structures of the different windows. Moreover, in the windows showing that distribution, the stability and the probability was very high (as shown by the red-orange color in the images) (Annex Figure 4). In two of the windows, the three nucleotides of codon 34 appeared paired. This conformation was discarded because the probability was very low, meaning that being paired was not the most probable

conformation (Annex Figure 4). It was also observed an external loop, located near codon 34, common to the three windows where the conformation was probable and stable.

In contrast, in the Recoded\_env sequence, the structure of codon 34 was not conserved across the windows. The C nucleotide appeared in most of the obtained structures as paired, but the probability of this conformation was very low (color blue/purple), as shown in Annex Figure 5. The other two nucleotides (G and U) appeared as unpaired with a higher probability (Annex Figure 5). Although the conformation of those two unpaired nucleotides was highly probable, none of the nucleotides and conformations surrounding codon 34 had high probability. This suggested that all of them were unstable. Moreover, in most of the windows, it was observed that the change of the codon induced a division of a single conformation into two or more new different and differentiated branches (Figure 35). Overall, it can be confirmed that mutating codon 34 originated a dramatic change in the secondary structure of the WT\_env RNA. In addition, mutated codon 34 disrupted the WT external loop.



**Figure 35. VARNAs visualization of some of the structures obtained from two different window fragments.** The color code represents the base-pairing probability of the individual bases predicted by RNAfold. Purple means low probability; red means high probability. On the left, windows 1900-2150 are represented. On the right, windows 2050-2300 are represented. A) WT structures

## Results

---

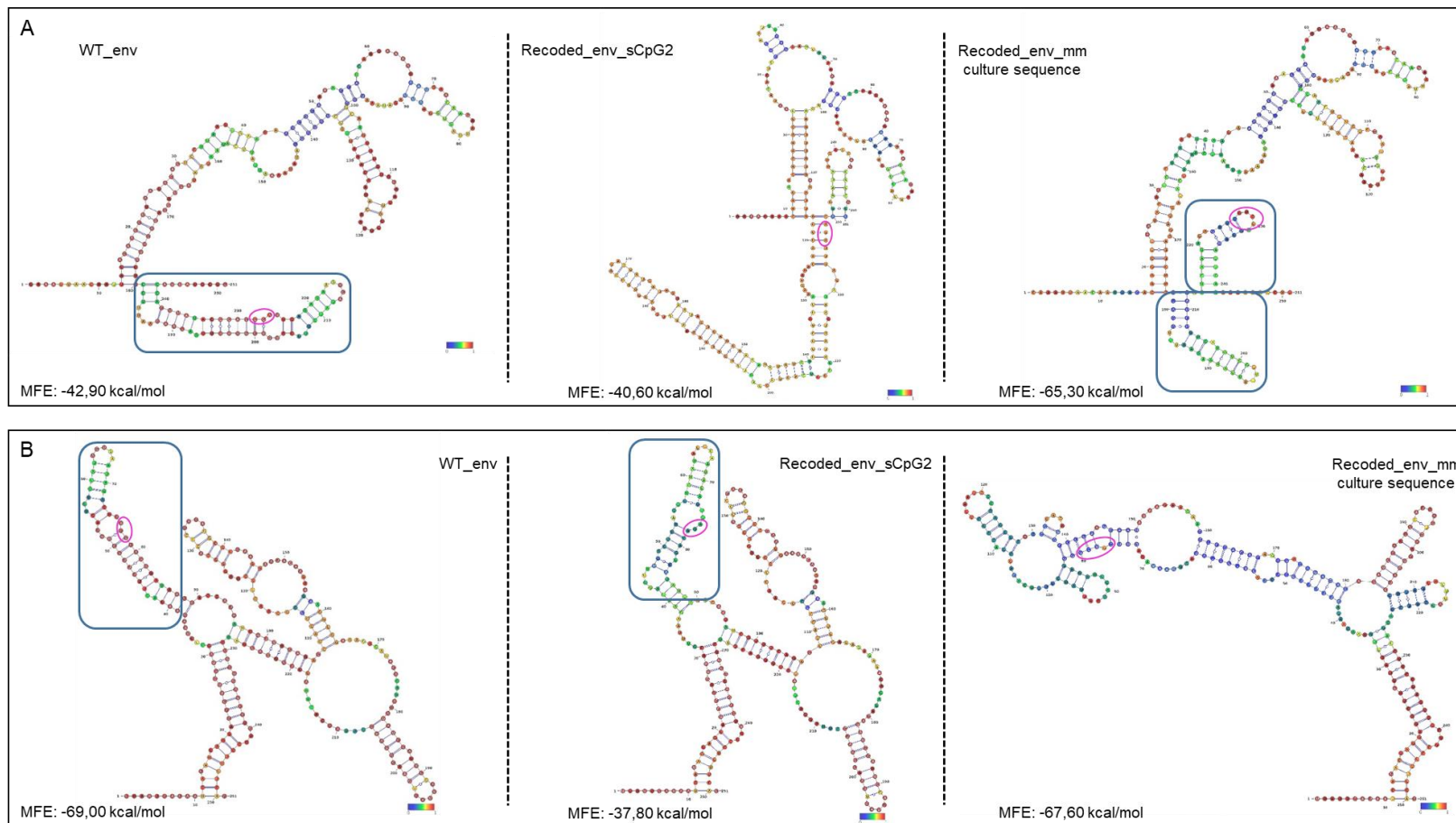
visualized by VARNA. The pink circle surrounds codon 34. Blue square surrounds the conformation that is lost after changing codon 34 nucleotides. B) Structure representations of the mutants that had codon 34. The pink circle shows codon 34 and blue squares surround the new conformations formed after disrupting WT\_env structure.

As previously described, Recoded\_env\_sCpG2 also had codon 34 mutated (AGA) and its replication capacity was significantly lower when compared to WT (Figure 30).

Similarly, Recoded\_env\_mm was designed with the mutated codon 34 (CGT). After transfection, this virus variant generated either a mutation next to codon 34 (non-synonymous mutation) (Table 15) or other mutations located elsewhere (Table 12), suggesting that changes were needed in order to improve its fitness.

To see how these mutations in or near codon 34 affected the RNA secondary structure of these two mutants, RNAfold was also applied to Recoded\_env\_sCpG2 and to Recoded\_env\_mm. Annex Figure 6 shows every structure for the different 5 windows that contained codon 34 of Recoded\_env\_sCpG2. It was observed that three of the structures were very similar to WT (Annex Figure 6 C and D, mainly, but also E). However, in the conformations in which the three nts of codon 34 were base-paired, the probability was higher (Annex Figure 6 A, B and C). That suggested a different distribution for codon 34 when compared to WT conformation. Moreover, some of the structures of Recoded\_Env\_sCpG2 showed that the external loop of the WT was lost. Instead, other external loops were generated in a similar way (Figure 36 A and B, structures in the middle). Overall, describing the conformation of codon 34 in this virus variant was difficult. The result suggested that the conformation may have changed, allowing the virus to replicate but with lower capacity.

Annex Figure 7 shows the structures for the 5 windows containing codon 34 and the adjacent mutation of Recoded\_env\_mm generated after transfection. In this case, most of the structures were very similar to Recoded-env (Annex Figure 5), except for one that changed dramatically and had low probability (Figure 36 B, structure on the right). In these structures it was observed that when codon 34 nucleotides were paired, their probability and the probability of the surrounding structures was very high. The most probable structures (Annex Figure 7 B and D) were very similar but not identical to Recoded-env (Annex Figure 5 B and D). The structures obtained from Recoded-env were unstable. On the contrary, the new mutation in Recoded\_env\_mm made the structure have more stability. This result suggested that different interactions may happened to allow Recoded\_env\_mm to replicate. Moreover, the mutation generated in culture after transfection was a non-synonymous mutation.



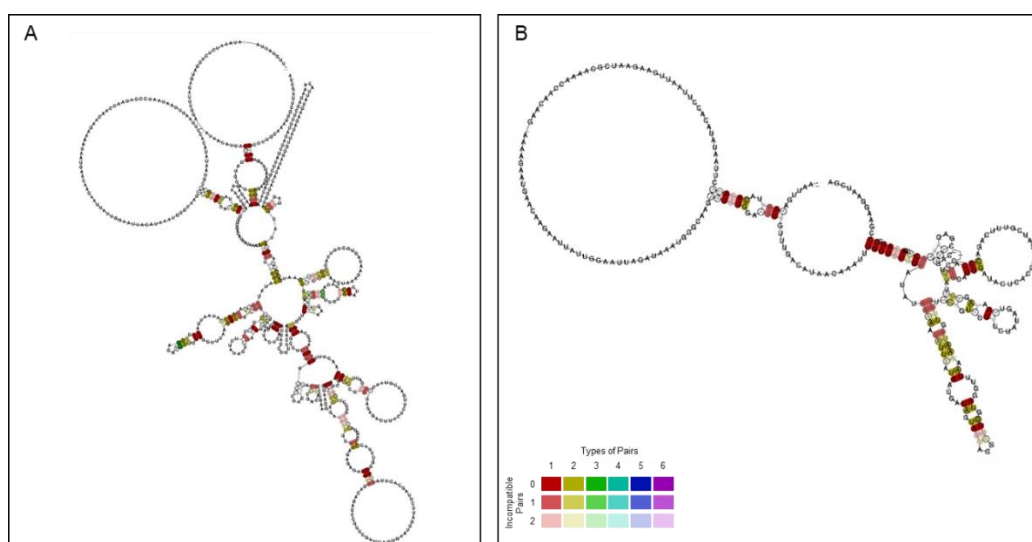
**Figure 36. Varna visualization of some of the structures obtained from two different *env* window fragments.** The color code represents the base-pairing probability of the individual bases predicted by RNAfold. Purple means low probability; red means high probability. The pink circle shows codon 34 and adjacent mutations if corresponding. Blue squares surround the structure in which codon 34 is involved and, if corresponding, the structure that is generated. When the entire structure is new, no blue squares are shown. On the left, structures of WT\_ env. In the middle, structures of Recoded\_ env\_sCpG2. On the right, structures of Recoded\_ env\_mm after transfection. A) Windows 1900-2150 are represented. B) Windows 2050-2300 are represented.

## Results

MAFFT tool was used to determine if that WT *env* RNA secondary structure surrounding codon 34 is conserved among different HIV-1 isolates. To this end, sequences of about 3000 isolates were downloaded from Los Alamos Database and aligned. First, gp120 sequences were deleted from the alignment to better analyze gp41 region that contains codon 34. Then, a BLAST (Basic Local Alignment Search Tool) was applied to gp41, grouping the gp41 sequences of the different isolates. After all gp41 sequences were grouped, they were clustered to avoid redundancy. The percentage of identity was established to 95%, reducing the number of sequences from 3000 to 153. Following clustering, they were aligned using MAFFT. RNAalifold was performed to obtain the secondary structure of the alignment (Figure 37 A).

The same BLAST and cluster tools were applied to three different windows that contained codon 34 (windows from positions 1901-2201, 1950-2200 and 1951-2251). In these cases, the percentages of identity were 95%, 98% and 98%, and the number of resulting sequences were 115, 157 and 152 respectively (Figure 37 B shows one of those windows).

After applying RNAalifold to gp41 aligned sequences and to different shorter windows, the resulting secondary structures were not conclusive. There were local fragments of the structures that could not be located in any specific conformation (big loops shown in Figure 37). The reason was that this kind of alignment was based on diverse different sequences. When there is sequence diversity, the tool cannot properly define a structure.

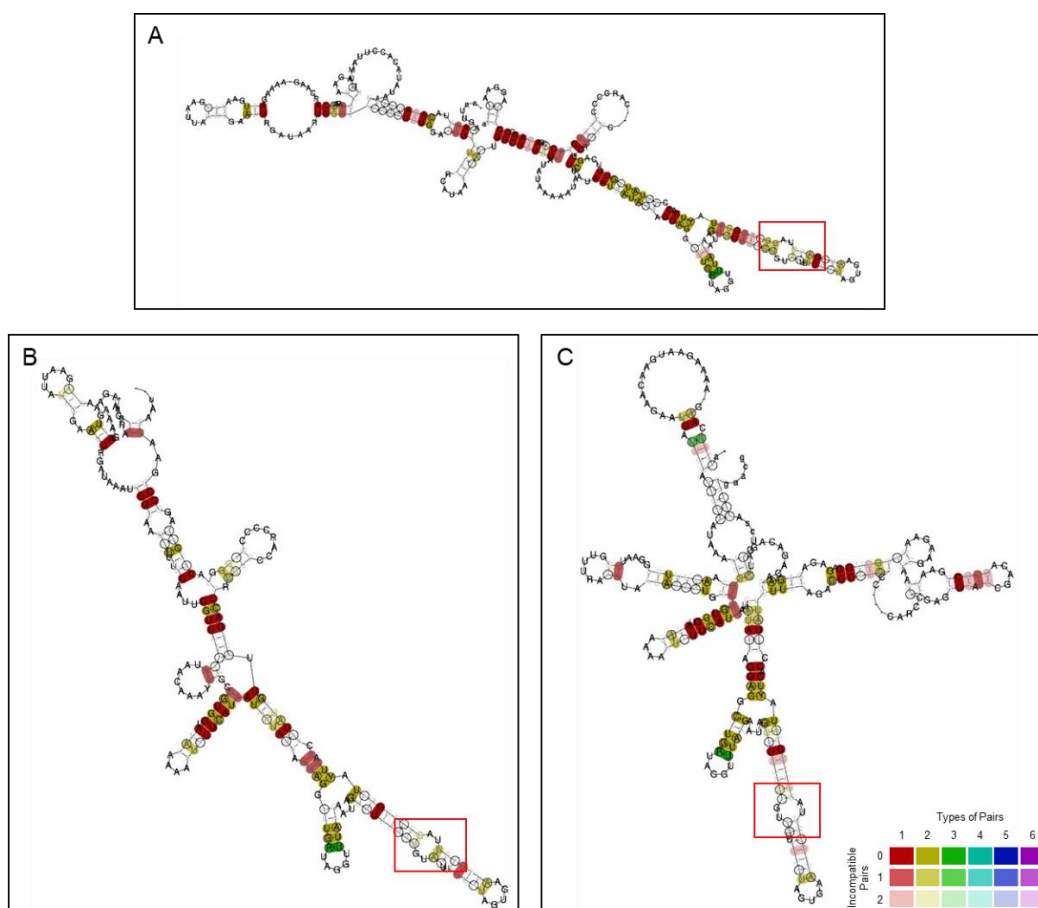


**Figure 37. Secondary structures from the alignment of HIV-1 gp41 isolates applying MAFFT and RNAalifold.** The color code represents the sequence conservation of the base pair. A) Structure of the alignment of all gp41 containing sequences of 153 isolates. B) Structure of the alignment for the window 1901-2201 containing sequences of 115 isolates.

locARNA tool was employed to try to obtain a better alignment of the different gp41 sequences extracted from Los Alamos database. This tool was applied to the different windows that were previously generated.

Overall, locARNA is more accurate because it takes into account not only the sequence but also the RNA secondary structure of each sequence to perform the alignment.

After performing locARNA to the clustered windows mentioned above, three different structures were obtained (Figure 38). In all three window approaches, codon 34 conformation was identical to the one obtained for WT\_env (A unpaired and Gs paired). It was also observed that the sequence and the structure were conserved. The structure was formed by an external loop located before codon 34 (Figure 38). That external loop was lost in Recoded-env (Figure 35). This observation strongly suggests an important role of that loop region (e.g. interaction with other RNA structures, recognition by viral proteins, etc.).



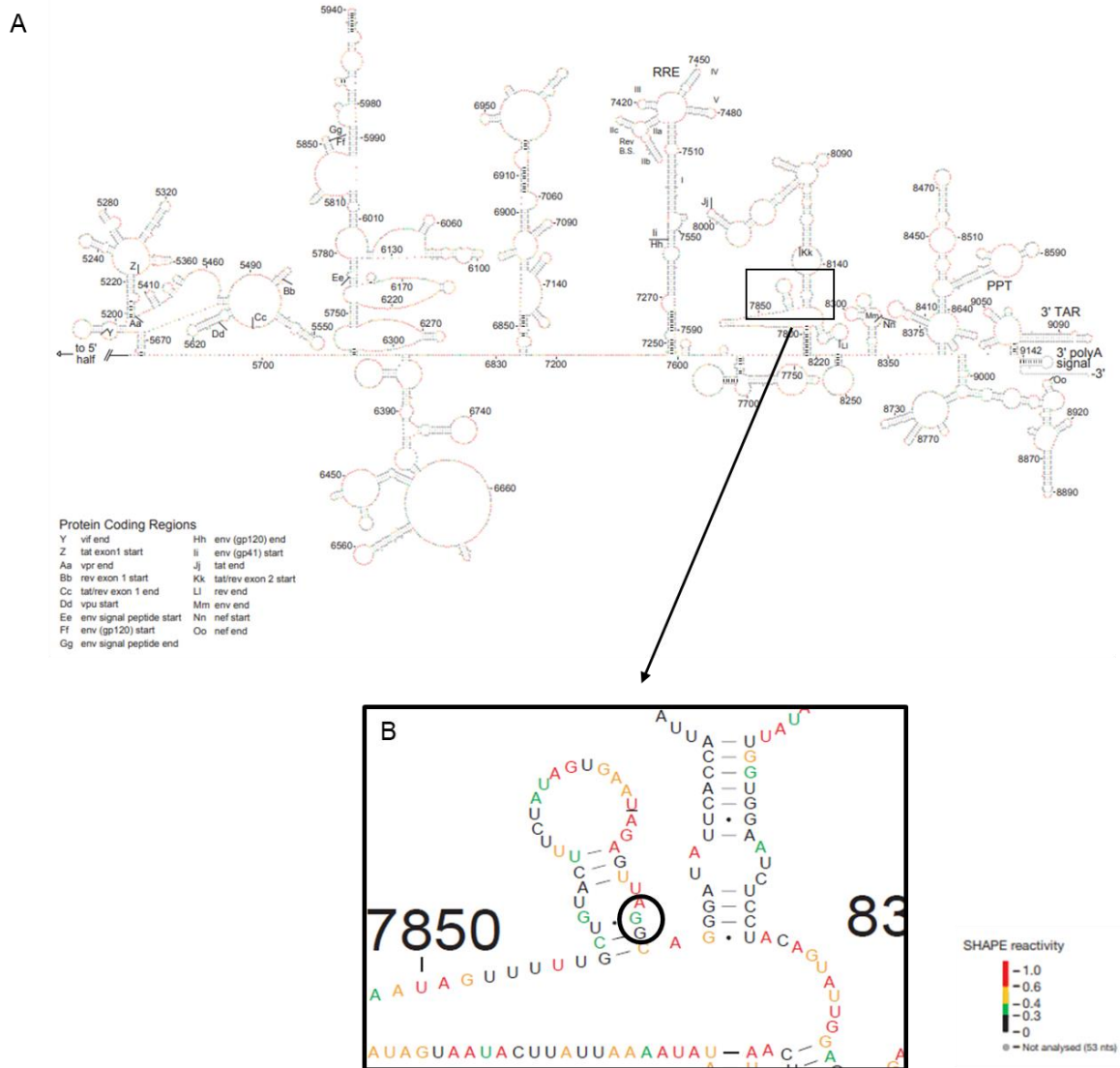
**Figure 38. Secondary structures of different window approaches from locARNA HIV-1 gp41 alignment.** The color code represents the sequence conservation of the base pair. The red square shows the region where the three nucleotides of interest are located. A) Sequences aligned from the window 1901-2201. B) Sequences aligned from the window 150-2200. C) Sequences aligned from the window 1951-2251.



### Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE)

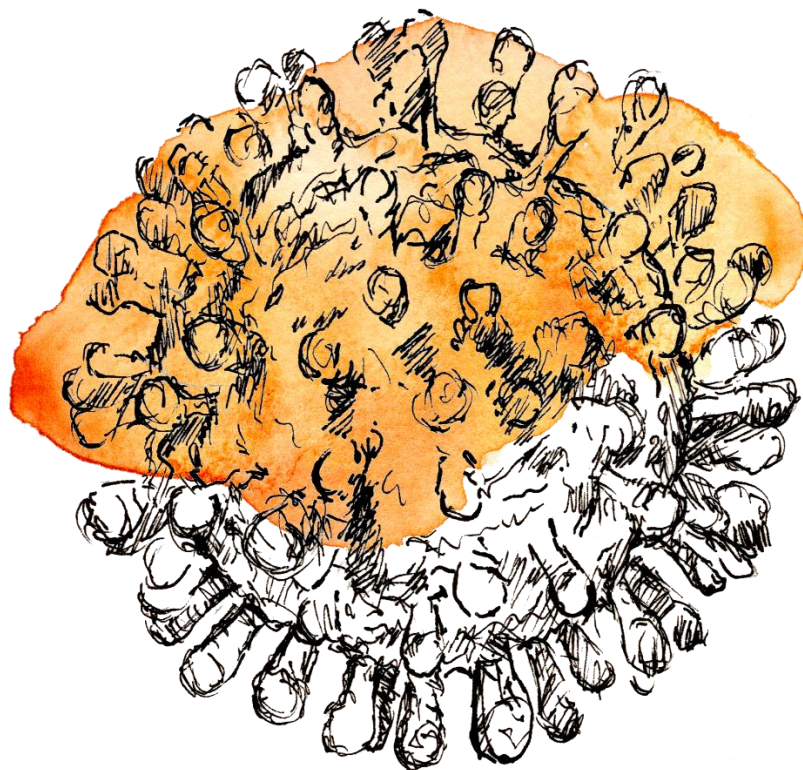
SHAPE is a biochemical experimental procedure to determine RNA secondary structures. SHAPE chemistry is based on the sensitivity of 2'-hydroxyl group to acylation. Non-paired nucleotides are more sensitive to hydroxyl-selective electrophiles than base-paired nucleotides. Since all RNA nucleotides have a 2'-hydroxyl group, the flexibility of every single nucleotide in a RNA sequence can be analyzed (Wilkinson, Merino, and Weeks 2006). Regions with low reactivity (below 0.25) indicate that the nucleotides are base-paired, while regions with reactivity of 0.5 or higher indicate unstructured nucleotides.

SHAPE has already been performed on HIV-1 genome (Watts et al. 2009). When codon 34 was searched within the HIV-1 SHAPE reactivity map (Figure 39) it was observed that this codon is structured as previously described by the computational tools shown above. The A nucleotide is unpaired, while GG nucleotides are paired. In addition, there is also an external loop observed in SHAPE just before codon 34. Although the external loop does not have the same length as the one predicted by RNAfold, it confirmed the existence of this organized structure. Moreover, the sequence used by Watts *et al.* (NL4-3) was aligned with WT\_env sequence. Both sequences are identical in the region where codon 34 is located. Only one discrepancy is found 40 nts after codon 34. Overall, SHAPE reactivity does confirm the results described with our *in silico* tools. It also suggests that codon 34 in Recoded-env may disrupt this stable and conserved conformation.



**Figure 39. SHAPE reactivity map.** Adapted from Watts *et al.* (Watts *et al.* 2009) A) RNA SHAPE reactivity for all nucleotides in HIV-1 genome. B) Partial SHAPE reactivity of the structure containing gp41 codon 34. Codon 34 is located as predicted by RNAfold.





**Chapter Three.**  
**Codon pair recoding of the envelope gene:**  
**effect of CpG dinucleotides**

---



## Envelope recoding: Changes in codon pair bias

The envelope of the HIV-1 was synonymously recoded changing its codon pair usage. This usage was changed in order to optimize, deoptimize or maintain neutral the CPB. To this end, human values and tables from Coleman and colleagues (Coleman et al. 2008) were used. As described before, the RRE and the overlapping regions were maintained WT to avoid non-viability. After performing the changes, an optimized **Max** sequence, a neutral **Neu** sequence and a deoptimized **Min** sequence were designed and constructed. All the changes were equally distributed between the *env* 5' region and the 3' region. These three new mutants increased the CAI value (Table 19), meaning that human rare codons were not introduced.

The total number of mutations introduced in Max were 361 (13,9% of gp160). Max *env* increased the CPB value from 0,046 to 0,176 (optimization of the CPB). The number of TpAs of this variant decreased, whereas the number of CpGs changed from 26 to 43, increasing a 165,4% within *env* (Table 19).

Neu sequence had 396 mutations (15,2% of gp160). This variant also reduced the number of TpAs, while the number of CpGs were incremented up to 50, an increase of 192%. The CPB value of this mutant was 0,035, very similar to the WT\_Env value (0,046) (Table 19).

The number of changes introduced in Min was 427 (16.4% of gp160). As it happened with Max and Neu, this mutant variant also decreased the number of TpAs. The CPB value was reduced to -0,160 (deoptimization of the CPB). On the other hand, the number of CpGs within this mutant was 111, which supposed an increment of 426% of the total number of CpGs of gp160 (Table 19).

**Table 19.** Calculated number of mutations, CPB, CAI, CpG number and TpA number of the viruses based on changes in codon pair usage.

Seq. name	Nº mut	CpG	TpA	CPB*	CAI**
WT_Env	-	26	194	0,046	0,1368950
Max	361	43	152	0,176	0,1579858
Neu	396	50	156	0,035	0,1832075
Min	427	111	144	-0,160	0,2127256

\*Codon-pair bias

\*\*Codon adaptation index. CAI values were obtained from [www.jcat.de](http://www.jcat.de)

## Results

As described above, the number of CpGs increased in the three new variants. For that reason, three new ones were designed based on Max, Neu and Min but with reduced number of CpGs. MaxCpG reverted some CpGs to WT or generated new mutations in order to diminish the effect of the high number of CpG dinucleotides. NeuCpG reverted almost all new CpGs to WT. And MinCpG reverted and generated new mutations to reduce the number of CpGs (Annex Figure 8).

MaxCpG had a total number of 21 CpGs, which were actually 5 CpGs less than the WT\_env. Its CPB value was 0,160 (very similar to Max) and the total number of mutations was reduced to 348. NeuCpG had 29 CpGs and the CPB value was 0,041 (both values were very similar to WT\_env), while the number of mutations was 380. Finally, MinCpG had 23 CpGs, which reduced the number of changes to 374, whereas the CPB value increased to -0,085 (Table 20).

To generate a more CPB deoptimized variant, a new MinCpG.2 was designed, trying to take into account both the CPB values and the number of CpGs introduced. Codon pairs with the most negative values tend to have CpGs located in the third and first positions of the codon pair. Trying to avoid the introduction of new CpGs meant to avoid the most negative codon pairs. As a result, MinCpG.2 did not differ much from MinCpG in the CPB value, the number of CpGs and the number of mutations (Table 20). MinCpG.2 had 354 mutations when compared to WT\_env, whereas the total number of differences between MinCpG and MinCpG.2 was 171 nucleotides (Annex Figure 8).

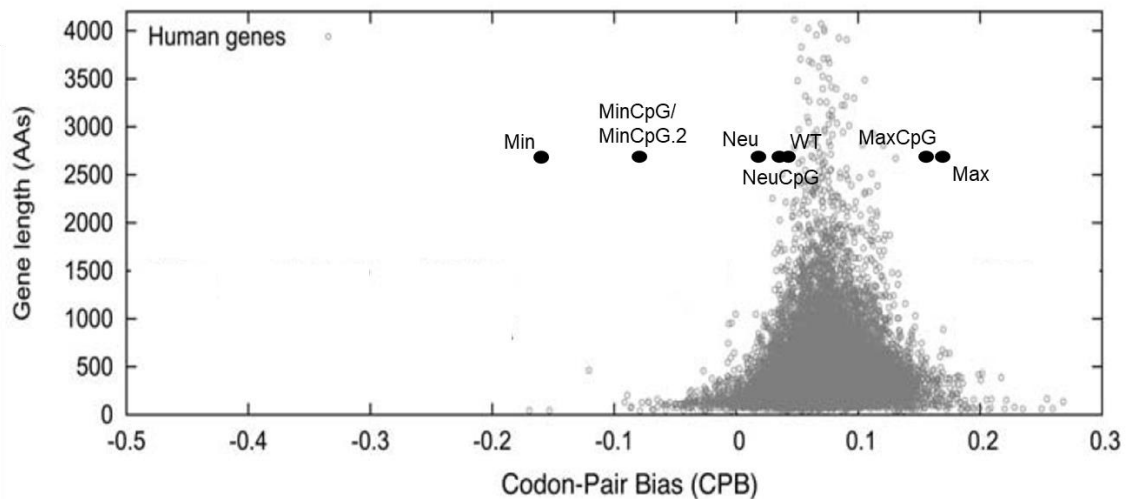
**Table 20. Calculated number of mutations, CPB, CAI, CpG number and TpA number of the viruses based on Max, Neu and Min with reduced number of CpGs.**

Seq. name	Nº mut	CpG	TpA	CPB*	CAI**
WT env	-	26	194	0,046	0,1368950
MaxCpG	348	21	155	0,160	0,1562661
NeuCpG	380	29	159	0,041	0,1801597
MinCpG	374	24	151	-0,085	0,1894470
MinCpG.2	354	23	158	-0,086	0,1905763

\*Codon-pair bias

\*\*Codon adaptation index. CAI values were obtained from [www.jcat.de](http://www.jcat.de)

CPB values of Max and MaxCpG were similar to some small genes of the human genome, meaning that the codons added to these mutants are overrepresented. Both Neu and NeuCpG had CPB values which were in the mean of the human gene values. For Min, MinCpG and MinCpG.2 their CPB was very low, so the codon pairs that were changed were rather rare and had a low frequency of appearance in the human genome (Figure 40).



**Figure 40. Calculated codon pair bias (CPB) score for all 14,795 annotated human genes.** Each dot represents the calculated CPB score of a human gene plotted against its amino acid length. Underrepresented codon pairs yield negative scores. The black spots indicate the CPB value for the variants generated changing the CPB scores. Adapted from (Coleman et al. 2008).

### *Viral stability of the viruses*

As previously described, all mutants were PCR amplified and co-transfected with the rest of the amplified genome of NL4-3 in MT-4 cells in order to know if the introduced mutations affected virus replication capacity.

After transfection, MinCpG.2 was the only recovered virus. Culture supernatant was collected and aliquoted. RNA was extracted from this supernatant and RT-PCR was performed to amplify *env*. After sequencing this RNA, it was concluded that there were no changes or additional mutations generated in MinCpG.2 after transfection.

More than thirty days following transfection and after more than five blind passages, no p24 or syncytia was observed in the other variants. Max, MaxCpG, Neu, NeuCpG, Min and MinCpG were transfected in triplicate to confirm the result. These six *env* constructs were considered lethal.

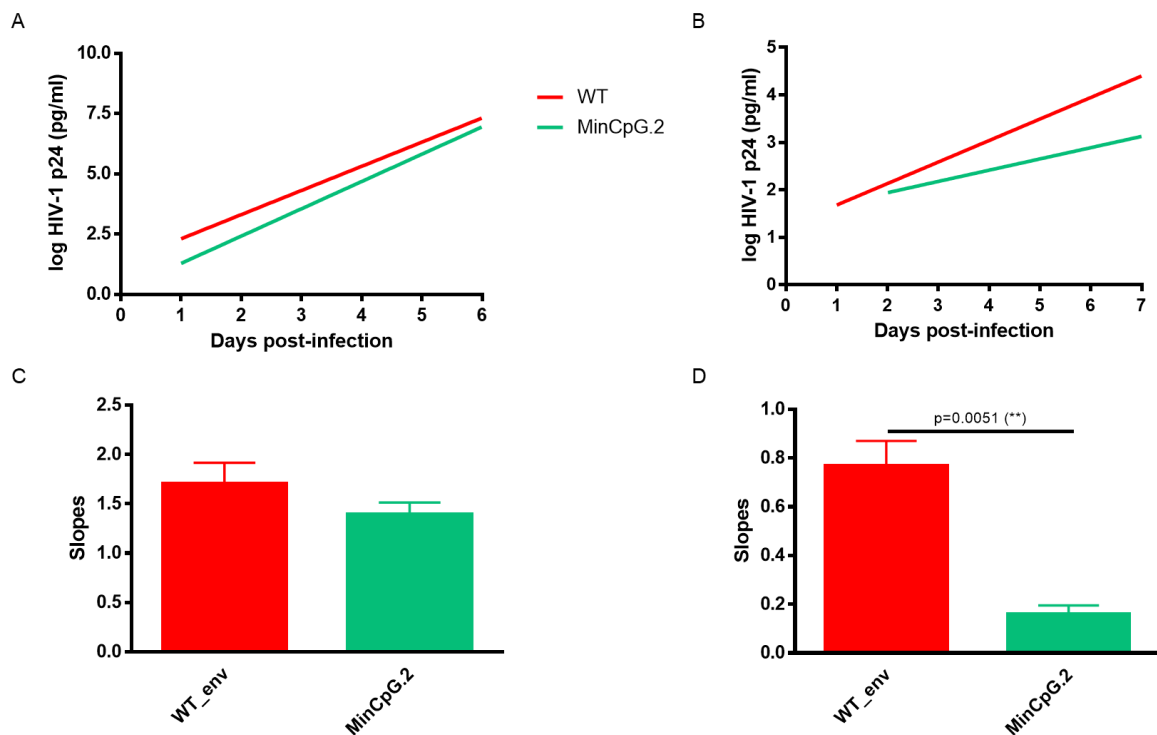


## Results

### Replication capacity of MinCpG.2

MinCpG.2 *env* was the only virus that could grow after transfection in MT-4 cells. To explore the capacity of propagation of this virus, replication capacity assays were carried out with the WT virus as the positive control. This assay was performed in MT-4 cells and PBMCs.

The obtained result demonstrated that MinCpG.2 did not have significant differences when compared to the WT virus ( $p=0.3265$ ) in MT-4 cells (Figure 41 A and C). However, when PBMCs were infected with these two viruses, the replication of MinCpG.2 displayed significantly lower capacity ( $p=0.0051$ ). This suggested that in *in vivo* conditions, this virus would show a reduced efficiency of propagation (Figure 41 B and D).



**Figure 41. Replication capacity assays for WT\_env and MinCpG.2.** Values represent the means plus SDs of three independent experiments. Lag and plateau phases were not considered in slopes to avoid their influence. A) Replication capacity in MT-4 cells shown as the log of p24 antigen measurements. B) Replication capacity in PBMCs shown as the log of p24 antigen measurements. C) Slopes of the viruses obtained in the exponential phase of growth after infection of MT-4 cells. D) Slopes of the viruses obtained in the exponential phase of growth after infection of PBMCs. P-values (student's t-test) were obtained from comparing the slope of MinCpG.2 to the slope of WT\_env.

MinCpG.2 was a recoded *env* that reduced the number of CpGs of Min. That number of CpG dinucleotides was very similar to the amount of that of WT\_env. The main difference (apart from the number of mutations) between WT\_env and MinCpG.2 is that the CPB score

of MinCpG.2 is lower than that of the WT\_env (Table 20). From this result it was shown that the introduction of rare codon pairs in *env* affected viral replication capacity in PBMCs.

#### *Viral stability to neutralizing antibody*

The antibody 2F5 is a broadly neutralizing monoclonal antibody that neutralize HIV-1 by joining to a conserved region in gp41. This antibody was used to explore the susceptibility of MinCpG.2 and see whether its lower replication capacity correlates with a greater susceptibility. The IC50 values obtained did not show any statistical difference between the WT\_env and MinCpG.2 ( $p=0,3406$ ) (Table 21).

**Table 21. Viral susceptibility of WT\_env and MinCpG.2 to 2F5 Ab.**

<b>Seq. Name</b>	<b>IC50 (nM) (Mean (SD))*</b>
<b>WT_env</b>	0,106 (0,06)
<b>MinCpG.2</b>	0,066 (0,006)

\*IC50 values are shown as the mean of, at least, three independent experiments with the SDs.

### **Reversion of the 3' *env* region to WT**

Results obtained from Recoded\_env shown above demonstrated that the 3' sequence region of *env* influenced the replication capacity of the virus. Since Max, MaxCpG, Neu, NeuCpG, Min and MinCpG did not grow after transfection, the 3' *env* region of these variants was reverted to WT. In all of them, the 5' *env* region was maintained as previously designed. The new variants were named as their templates, but adding the suffix "3'wt" to their names (Table 22) (Annex Figure 9).

## Results

Table 22. Calculated number of mutations, CPB, CAI, CpG number and TpA number of the CPB viruses that reverted the 3' *env* region.

Seq. name	Nº mut	CpG	TpA	CPB*	CAI**
WT <i>env</i>	-	26	194	0,046	0,1368950
Max-3'wt	294	39	161	0,147	0,1532634
MaxCpG-3'wt	284	24	163	0,134	0,1515851
Neu-3'wt	319	47	164	0,030	0,1725310
NeuCpG-3'wt	305	30	166	0,035	0,1703344
Min-3'wt	350	103	156	-0,129	0,1935816
MinCpG-3'wt	297	26	163	-0,059	0,1709487

\*Codon-pair bias

\*\*Codon adaptation index. CAI values were obtained from [www.jcat.de](http://www.jcat.de)

After the reversion were performed, Max-3'wt, MaxCpG-3'wt, Neu-3'wt and NeuCpG-3'wt reduced their scores of CPB (Table 22). On the contrary, Min-3'wt and MinCpG-3'wt increased those values. In all of them, as it happened before, the CAI was improved when compared to WT (Table 22).

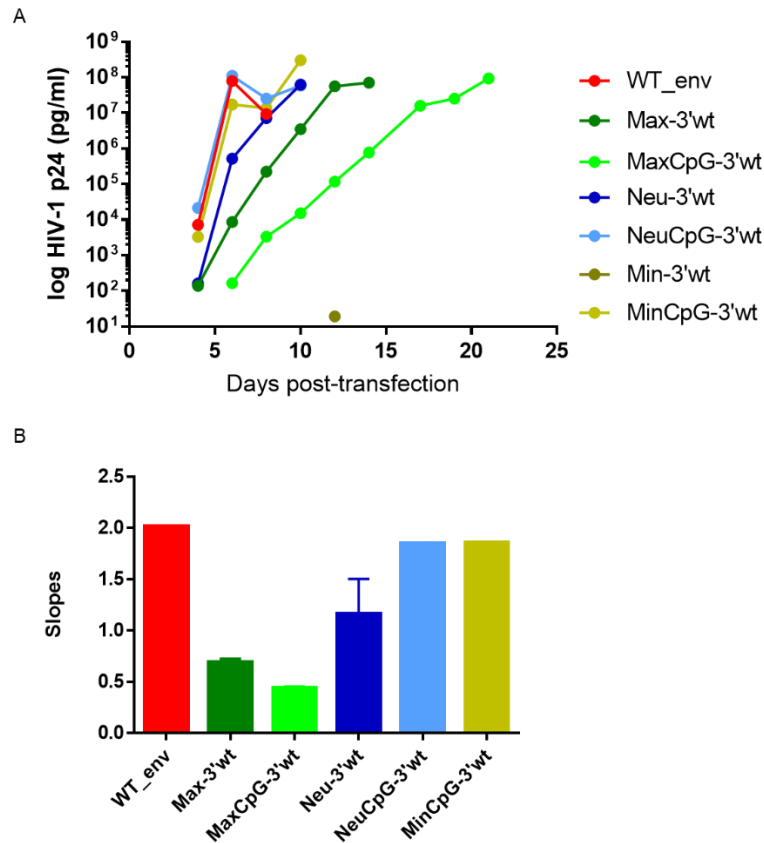
The total amount of mutations was reduced in all variants when compared to the previous mutants. Max-3'wt had 294 mutations (11,3%), MaxCpG-3'wt had 284 (10,9%), Neu-3'wt had 319 (12,3%), NeuCpG-3'wt had 305 (11,7%), Min-3'wt had 350 (13,5%) and MinCpG-3'wt had 297 (11,4%). Within the six new mutants, Min-3'wt is the variant that contained more changes in the 5' *env* region.

In MaxCpG-3'wt, NeuCpG-3'wt and MinCpG-3'wt the number of CpGs increased to 24, 30 and 26, respectively, when compared to the previous mutants. This happened because when the mutants were designed, hundreds of mutations were introduced at the same time. Some of these changes introduced new CpGs, but some others eliminated the CpGs that already existed in the WT\_Env. When the 3' region was reverted, all these modifications got back to WT.

### *Viral stability of the viruses*

These six new variants with the 3' *env* region WT were co-transfected with the other regions of the HIV-1 in MT-4 cells. WT\_Env was also transfected as positive control. Every 2-3 days

after transfection, supernatant culture was collected per each of the mutants and detection of HIV-1 p24 was performed in order to determine if the new constructs were replicative.



**Figure 42. Log of antigen p24 values after transfection and the slopes for each curve of the modified CPB viruses with reverted 3' env region.** A) Representation of p24 amounts obtained at different days after transfection. Only Min-3'wt was not able to replicate. B) Slopes obtained from the exponential phase of each growing virus.

HIV-1 p24 amounts were not detected in Min-3'wt. The rest of the virus variants were detected at different days post-transfection. Three viruses (Neu-3'wt, NeuCpG-3'wt and MinCpG-3'wt) displayed similar amounts of p24 as WT\_env. Both Max-3'wt and MaxCpG-3'wt needed several days more to surpass the amount of 500ng/ml (Figure 42 A). As previously described, the slopes obtained from the transfection anticipated us the replication capacity of the virus variants. Both Max-3'wt and MaxCpG-3'wt showed lower slopes than WT\_env, which might indicate a lower replication capacity (Figure 42 B).

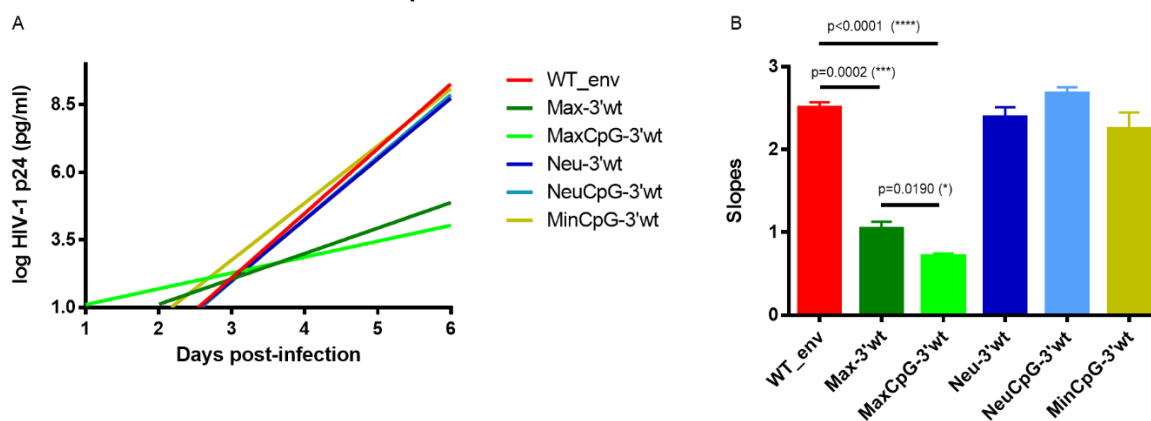
RNAs were extracted from the last day of collection of supernatants, and RT-PCRs amplifying the *envs* were performed. After sequencing these RT-PCRs, no mutations were found, meaning that all viruses grew with the designed sequence.

## Results

### Replication capacity of the viruses

To elucidate how CPB was affecting the replication capacity of the new reconstructed viruses, replication capacity assays were performed in MT-4 cells. Following titration of the viruses, MT-4 cells were infected with the same MOI, including WT\_env as positive control.

After determining the slopes in the exponential phase of growth, it was determined that Max-3'wt and MaxCpG-3'wt had statistically lower replication capacities ( $p=0.0002$  and  $p<0.0001$ , respectively), when compared to WT\_env. Additionally, when the slopes of Max-3'wt and MaxCpG-3'wt were compared between them, it was observed that Max-3'wt had higher replication capacity than MaxCpG-3'wt ( $p=0.0190$ ) (Figure 43). Of note, Max-3'wt had more CpG dinucleotides than MaxCpG-3'wt (39 and 24, respectively), and Max-3'wt did also have more mutations (Table 22). The total number of different nucleotides between them was 17. Among those 17 different nts, 14 were reversions to WT in MaxCpG-3'wt to avoid a high CpG content. The other 3 different nts were new synonymous mutations.



**Figure 43. Replication capacity assays for the CPB mutants with the WT 3' env region.** Values represent the means plus SDs of three independent experiments. Lag and plateau phases were not considered in slopes to avoid their influence. A) Replication capacity in MT-4 cells shown as the log of p24 antigen measurements. B) Slopes of the viruses obtained in the exponential phase of growth. P-values (student's t-test) were obtained from comparing the slope of the variants to the slope of WT\_env, or from comparing the slopes of two specific viruses.

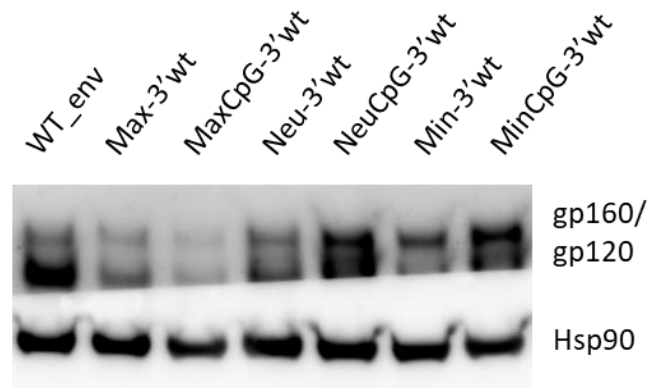
On the other hand, Neu-3'wt ( $p=0,4568$ ), NeuCpG-3'wt ( $p=0,1666$ ) and MinCpG-3'wt ( $p=0,2888$ ) did not show statistically different replication capacities when their slopes were compared to WT\_env. Moreover, when the two slopes of Neu-3'wt and NeuCpG-3'wt were compared, they did not display any difference ( $p=0,1146$ ) (Figure 43).

### *Effects on translation: protein expression*

To explore if there were alterations in protein expression, all the virus variants with WT 3' *env* region and the control WT\_*env* were cloned in the expression plasmid pcDNA3.1/V5-His-TOPO. These expression plasmids were co-transfected in 293T cells with pTOPORev. Forty-eight hours after transfection, cells were collected and proteins were extracted from the pellet.

After performing the WB, it was observed that NeuCpG-3'wt and MinCpG-3'wt generated amounts of protein similar to WT\_*env*. Neu-3'wt displayed less amount than NeuCpG-3'wt, but still the quantity was high (Figure 44).

The amounts of Min-3'wt were not very low, meaning that translation was not affected, and suggesting that another mechanism was being altered to justify its lethality.



**Figure 44.** Immunoblot analyses of the amounts of protein for WT\_*env* and the mutants generated after changing the CPB and reverting their 3' *env* region. All mutants were co-transfected with pTOPORev.

In the case of Max-3'wt and MaxCpG-3'wt the amounts of protein were much lower when compared to WT\_*env* (Figure 44). These amounts correlated with the replication capacity of these viruses, in which Max-3'wt replicated better than MaxCpG-3'wt (Figure 43). In both cases, translation seemed to be affected by the synonymous changes.

### *Effects on transcription: mRNA levels*

To see if the alterations in translation were accompanied by a defect in transcription, real-time PCR was performed on RNAs extracted from 293T cells transfected with the expression *env* plasmids and pTOPORev. Three independent experiments were performed.

## Results

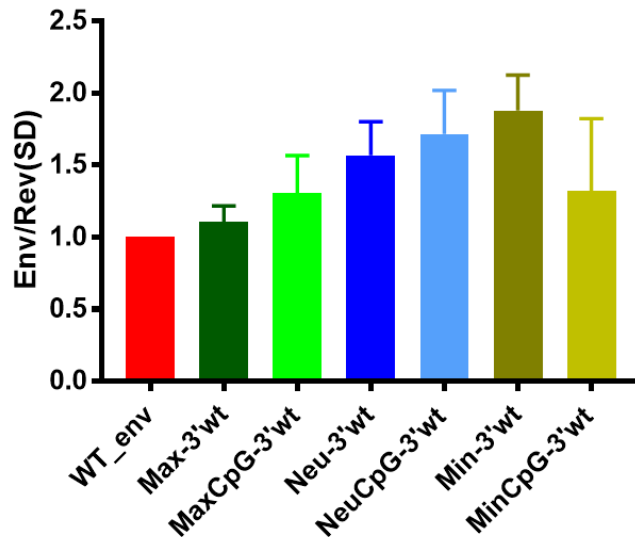
As it was observed with Recoded\_env and related mutants, there was a correlation between the amount of Env mRNA and the amount of Rev mRNA (Table 23), confirming that without Rev, Env cannot be expressed.

**Table 23. Real-time relative RT-PCR for env and rev mRNA amounts.**

Seq. Name	Replicate 1		Replicate 2		Replicate 3	
	2 <sup>-ΔΔCt</sup> (Env/GAPDH)	2 <sup>-ΔΔCt</sup> (Rev/GAPDH)	2 <sup>-ΔΔCt</sup> (Env/GAPDH)	2 <sup>-ΔΔCt</sup> (Rev/GAPDH)	2 <sup>-ΔΔCt</sup> (Env/GAPDH)	2 <sup>-ΔΔCt</sup> (Rev/GAPDH)
WT_env	1,00	1,00	1,00	1,00	1,00	1,00
Max-3'wt	0,49	0,41	0,52	0,46	0,24	0,25
MaxCpG-3'wt	0,37	0,24	0,27	0,21	0,22	0,20
Neu-3'wt	0,32	0,24	0,13	0,07	0,22	0,15
NeuCpG-3'wt	0,16	0,12	0,27	0,13	0,25	0,14
Min-3'wt	0,15	0,09	0,58	0,27	0,61	0,34
MinCpG-3'wt	0,17	0,09	0,15	0,17	0,35	0,30

Viral genes were related to the cellular gene GAPDH and quantitated by the 2<sup>-ΔΔCt</sup> method.

After analyzing the relationship between all mutants and WT\_env, it was observed that there were no significant differences. All mRNA amounts were higher than WT, which meant that transcription was not being affected (Figure 45).



**Figure 45. mRNA amounts of the CPB modified virus variants.** The graph shows the mean of three independent experiments with the SDs. All samples were related to WT\_env.

## Analyses of the CpGs in the 5' *env* region

As it has been described before, only Min-3'wt was lethal for virus replication. The other variants in which the 3' *env* region was reverted to WT were able to replicate. Max-3'wt and MaxCpG-3'wt had a better CPB score when compared to WT\_env and the other designed variants. However, their replication capacity is significantly lower when compared to WT\_env (Figure 43). Thus, CPB was not affecting the replication capacity of these two virus variants.

Apart from the CPB scores and the number of mutations, the number of CpG dinucleotides differed between the different virus variants with the 3'region WT. The distribution and number of the CpGs of all CPB mutants were analyzed in order to see if there were discrepancies that explained the discrepancies in replication capacity.

To start with, the number of CpGs in the 5' *env* region was annotated. Since the 3' *env* region was reverted, the CpGs located in this region were not taken into account. Interestingly, the WT\_env had 8 CpGs in the 5'region, 4 more were located in the RRE and the other 14 CpGs were located in the 3'region. When the other mutants were analyzed, almost all of them had higher number of CpGs in the 5' *env* region than WT\_env (Table 24).

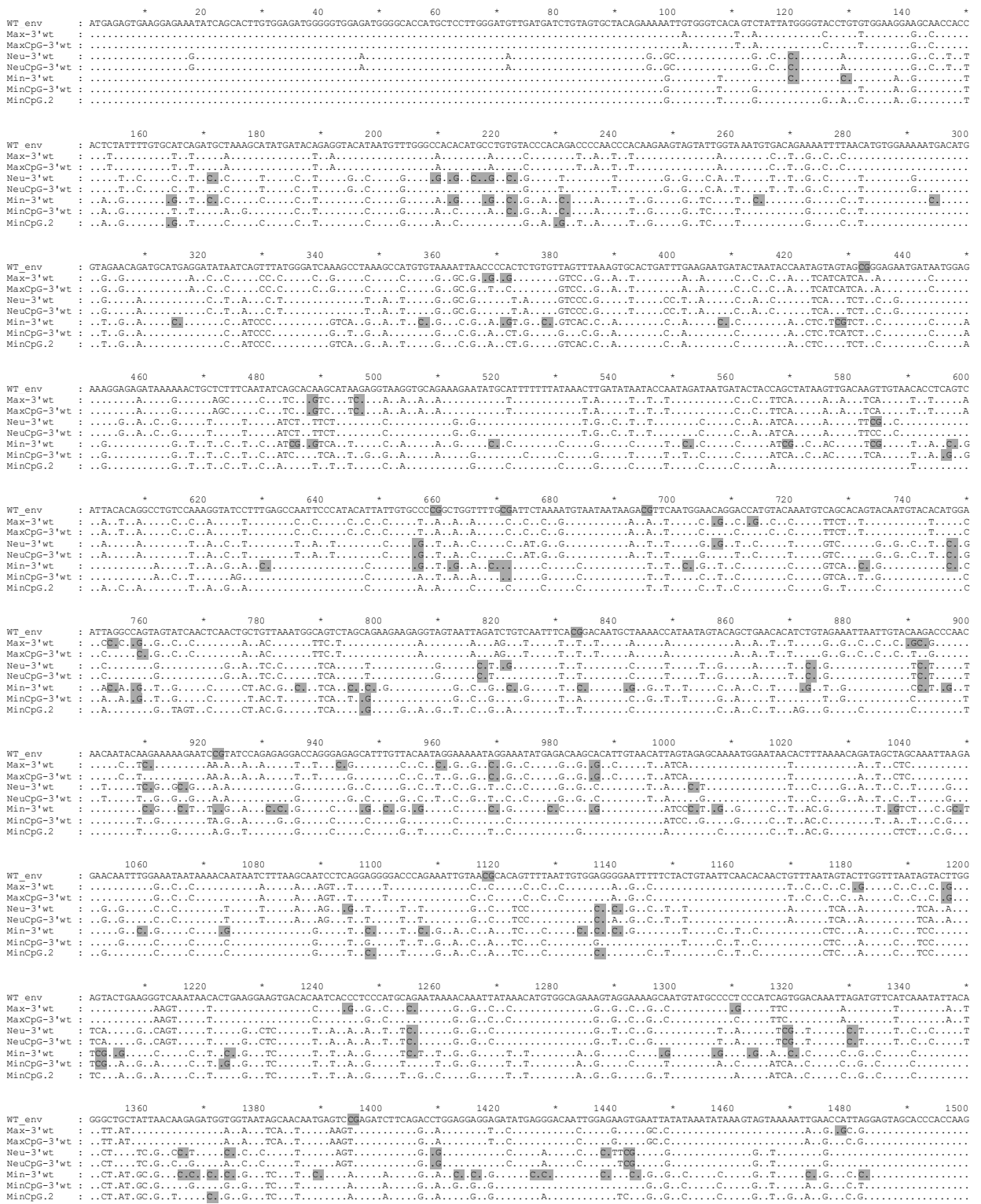
**Table 24. Number of CpGs located in the 5'region of CPB mutants in which the 3' *env* region was reverted to WT.**

Seq. name	N° CpGs in the 5'region
WT_env	8
Max-3'wt	21
MaxCpG-3'wt	6
Neu-3'wt	29
NeuCpG-3'wt	12
Min-3'wt	85
MinCpG-3'wt	8
MinCpG.2	6

MaxCpG-3'wt and MinCpG.2 had only 6 CpGs in the 5'region, while MinCpG-3'wt had the same number of CpGs as the WT\_env. From this result, it can be interpreted that the number of CpGs in the 5' *env* region was not affecting virus replication capacity.



# Results



**Figure 46. Nucleotide alignment of the 5' env region of WT\_env and CPB mutants. CpGs are labeled in grey for each sequence.**

Since the number of CpGs seemed not to affect viral fitness, we decided to explore if the distribution of them was altering the replication capacity of Max-3'wt and MaxCpG-3'wt.

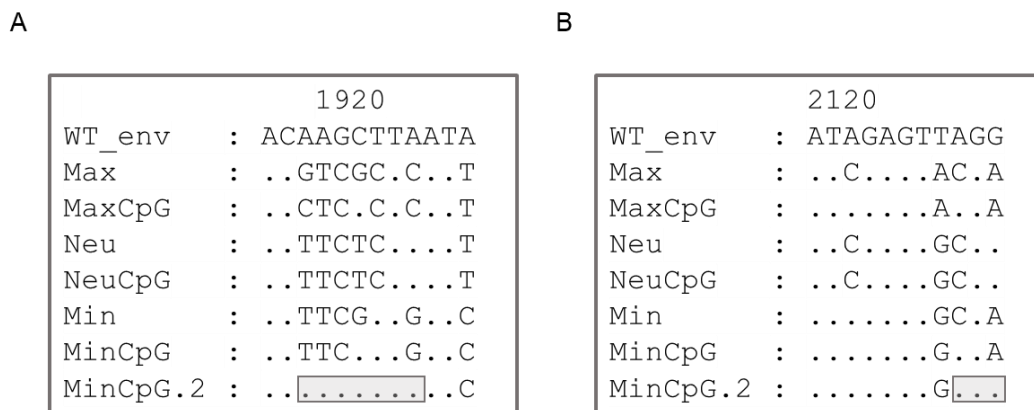
Max-3'wt had its first CpG located at position 368 of *env*, whereas MaxCpG-3'wt had its first at position 489. In WT\_Env, the first CpG is located at position 431. The location of the first CpGs of these three viruses is very similar. Moreover, the rest of the variants had CpGs located in earlier positions (Figure 46). This result suggested that, at least at the beginning of the sequence, the distribution of the CpGs were not affecting viral viability.

Between nucleotide 368 and nucleotide 583 of *env*, only Max-3'wt, MaxCpG-3'wt and Min-3'wt had CpGs. Between this region, WT\_Env only had 1 CpG. The other three mutants did not have CpGs located in this region (Figure 46). Interestingly, Max-3'wt, MaxCpG-3'wt and Min-3'wt were the virus variants with lower or none replication capacity. This observation suggested that introducing CpGs between the *env* nucleotides 368 and 583 may affect the replication capacity of HIV-1.

### **Analyses of the 3' *env* region**

As it has been described, when the 3' *env* region was reverted to WT, almost all viruses were able to grow. In addition, MinCpG.2 replicated without reverting its 3' region sequence. In order to see if there were differences in the 3' *env* region that might explain this phenomenon, sequence analyses were performed, as well as applying RNAfold to several sequence fragments.

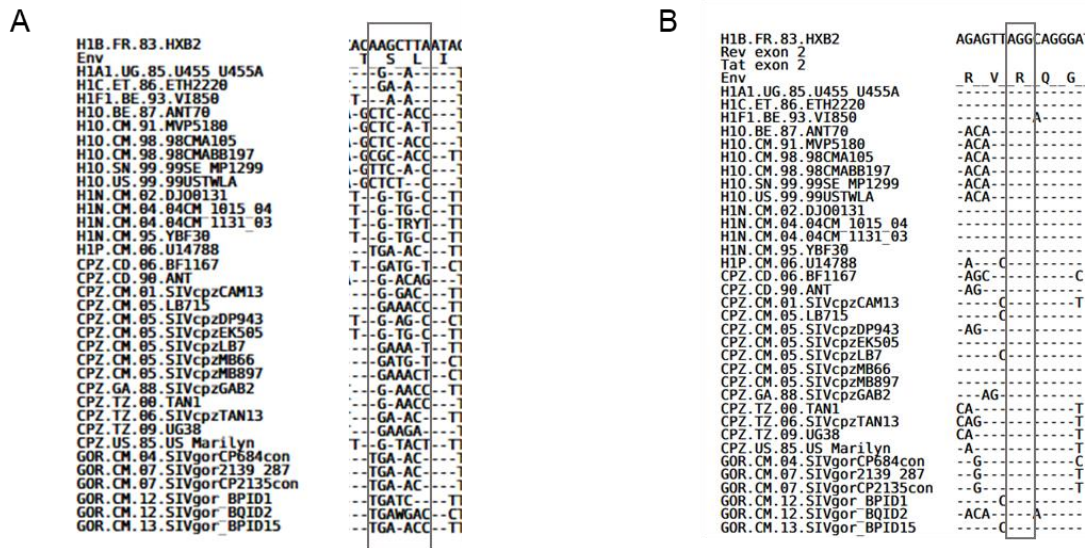
First, the 3' region was compared between all the CPB viruses that were primarily designed (Annex Figure 8). We looked for regions that were exclusive for MinCpG.2, that is, regions that were mutated or WT only in this virus. Two regions were found (Figure 47). The first of those regions began in position 1916 of *env* and had an extension of 7 nucleotides (Figure 47 A). This fragment had WT sequence only in MinCpG.2; in the rest of the virus variants, that region was altered with several mutations. The second region that was exclusively WT for MinCpG.2 was at position 2125 (Figure 47 B). At that position, three nucleotides were WT and the three of them belonged to the same codon. Interestingly, that codon is the same (codon 34) that when mutated in Recoded\_Env, the secondary structure was altered and the virus was not viable.



**Figure 47. Partial alignment of the two regions in the 3' region that remained WT in MinCpG.2 env.** A) Partial alignment of the region in the 3' region that begins in position 1916 of the *env*. All mutants had mutations in that region, except for the MinCpG.2. The mutated sequence extended for 7 nucleotides. B) Partial alignment of the region in the 3' region that begins in position 2125 of the *env*. The codon AGG was maintained WT in MinCpG.2. Additionally, the codon AGG corresponds to the codon 34 of Recoded\_*env*.

When MinCpG.2 is compared with the other virus variants, more discrepancies might be encountered. However, it is important to notice that MinCpG was not able to replicate when the 3' *env* region was mutated. These two virus variants, MinCpG and MinCpG.2, had very similar characteristics: CPB, number of mutations and number of CpGs. Moreover, MinCpG was very similar in sequence to MinCpG.2 (Annex Figure 8). Thus, the differences in the two described regions above (Figure 47) suggested their importance for viral replication. MinCpG had 72 mutations in the 3' region, while MinCpG.2 had 68 mutations. The number of discrepancies in the 3' region between these two viruses was 23 nucleotides. Among those 23 discrepancies, 5 of them were located in the two regions described above.

In order to explore if those regions are conserved among different isolates of the HIV-1 and SIV genome, alignments of different sequences were analyzed. The region that starts in position 1916 and extends for 7 nts is not conserved in *env* gp41. That region highly varies within different HIV-1 and SIV isolates, meaning that it may be a flexible domain (Figure 48 A). However, when the codon AGG (codon 34) was analyzed among different sequences, it was observed that this codon is highly conserved. Importantly, the nucleotides located at both sides of that codon are also highly conserved (Figure 48 B). This means that this region is not mutation flexible and it may not admit changes, probably related to a loss of its RNA secondary structure.



**Figure 48. Partial *env* alignment of different HIV-1 and SIV isolates' sequences.** Alignments were obtained from [www.hiv.lanl.gov](http://www.hiv.lanl.gov). A) Partial alignment of the region in the 3' region that begins in position 1916 of the *env*. B) Partial alignment of the region in the 3' region that begins in position 2125 of the *env*.

### RNA secondary structures

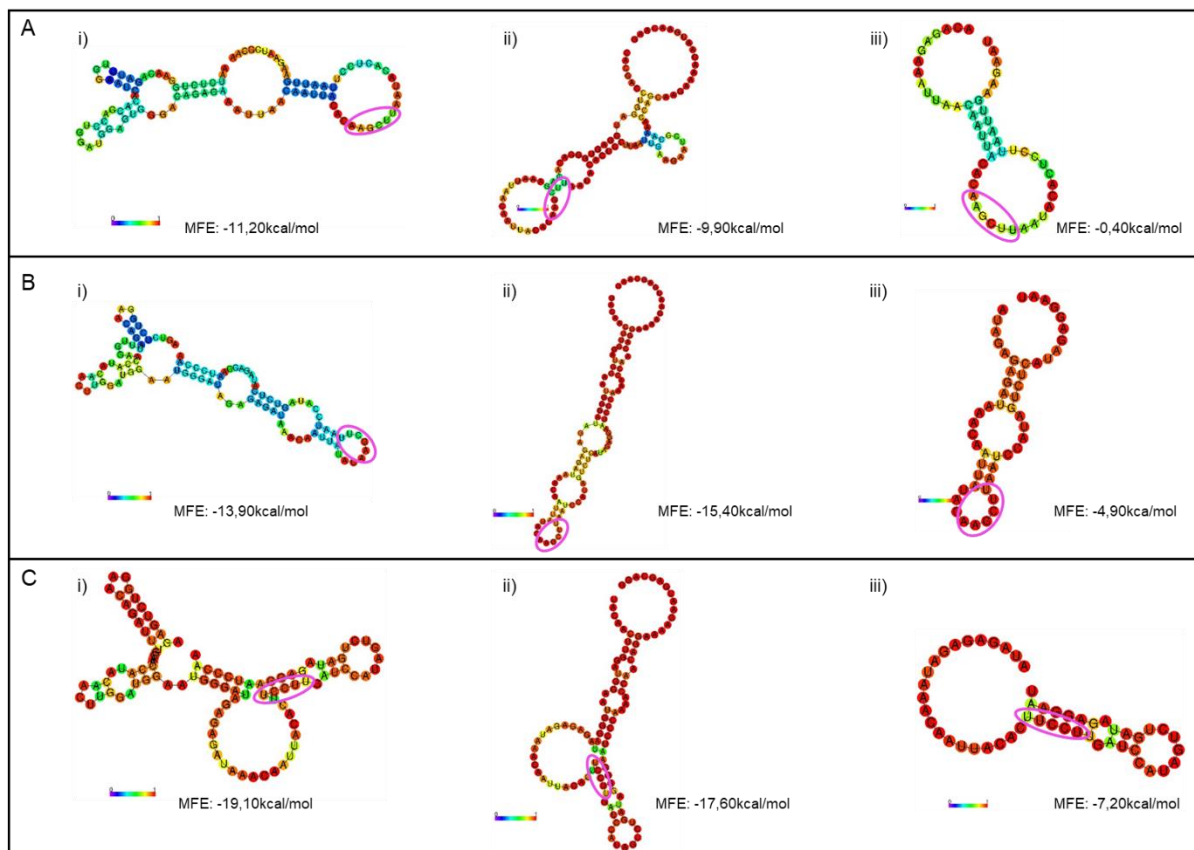
To see if there were differences between the RNA secondary structure of the 3' *env* region of MinCpG and MinCpG.2, RNAfold was applied. The entire 3' *env* region was not introduced in the analysis, because it is very long to be analyzed. For that reason, only the two regions that were different in MinCpG.2 were studied (Figure 47). This computational tool was applied to WT\_*env*, MinCpG and MinCpG.2.

The RNA secondary structure varies depending on the nucleotides of a window. If there are mutations introduced, the number of possibilities and discrepancies increases. Consequently, the window size in these cases was made smaller than 250 nts. Both MinCpG and MinCpG.2 had several mutations surrounding those two regions, so the window size was established in 100 nts. Three windows containing those regions were analyzed. Of note, the target regions are surrounded by other designed mutations of each virus variant. Thus, each 100-nt window contained not only the target region but also the surrounding mutations of each virus. These surrounding mutations were different between MinCpG and MinCpG.2. Moreover, the RNA secondary structures are affected by these mutations too. An extra RNA secondary structure was obtained from a window of 50 nucleotides, in which the target regions that we wanted to analyze were located in the middle of the window.

## Results

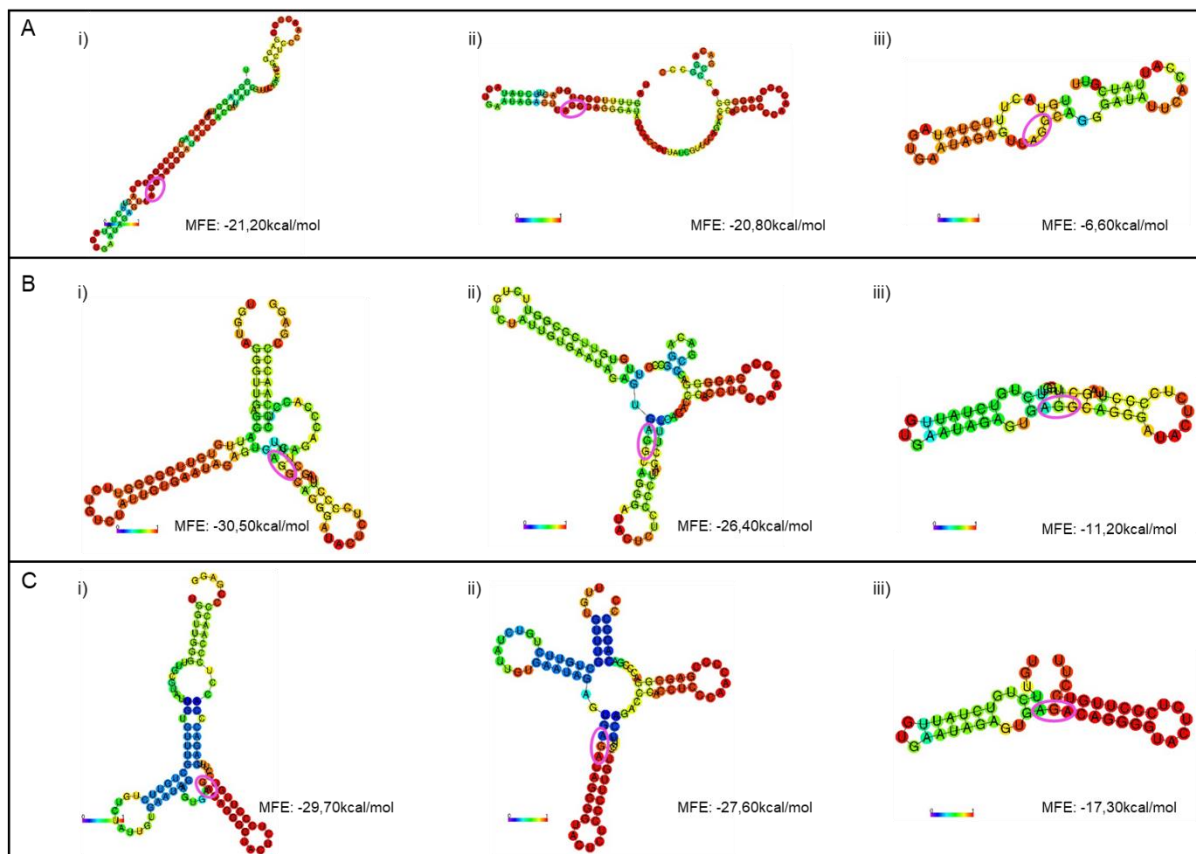
After analyzing the windows that contained the nucleotides surrounding the position 1916, it was observed that in WT\_ env structures, the 7 nucleotides were located, with high probability, into an external loop (Figure 49 A). MinCpG.2 displayed a similar RNA structure when compared to WT. Although the conformation was smaller, the 7 nucleotides were unpaired in an external loop (Figure 49 B). However, MinCpG showed a structure in which these 7 nucleotides were very stably paired (Figure 49 C). The structures obtained from this mutant clearly differed from the ones obtained from WT\_ env and MinCpG.2.

From these results, it can be concluded that MinCpG changed dramatically the conformation, while MinCpG.2 conserved a small external loop. Possibly, MinCpG lost some important interactions for replication, whereas MinCpG.2 may have conserved them due to its *env* RNA secondary structure conformation.



**Figure 49. RNAfold structures of the different windows containing the region that begins at position 1916 of the *env* of WT\_ env and MinCpG and MinCpG.2 variants.** The color code represents the base-pairing probability of the individual bases predicted by RNAfold. Purple means low probability and red means high probability. The nucleotides indicated in Figure 47 A are circled in pink. A) WT\_ env structures. B) MinCpG.2 structures. C) MinCpG structures. i) Structures obtained from the fragment 1850-1950. ii) Structures obtained from the fragment 1875-1975. iii) Structures obtained from the fragment 1895-1945.

On the other hand, when the windows that included the nucleotides that began in position 2125 were analyzed, it was observed that, as previously described (Figure 38, Figure 39 and Annex Figure 4), in WT\_env the A nucleotide was unpaired and the GG dinucleotides were paired, with a previous external loop (Figure 50 A). However, in both MinCpG and MinCpG.2, that structure was lost and all the codon (AGG) appeared as paired (Figure 50 B and C). Although the RNA secondary structure also changed in MinCpG.2, the probability of the distribution of the codon AGG in this virus variant was low, whereas in MinCpG the probability was very high. This suggested that the predicted conformation of MinCpG.2 might not be the expected *in vivo*. On the contrary, the conformation of MinCpG has a high probability of being the predicted one. This difference might be the reason why MinCpG.2 env variant was replicative and MinCpG was not.

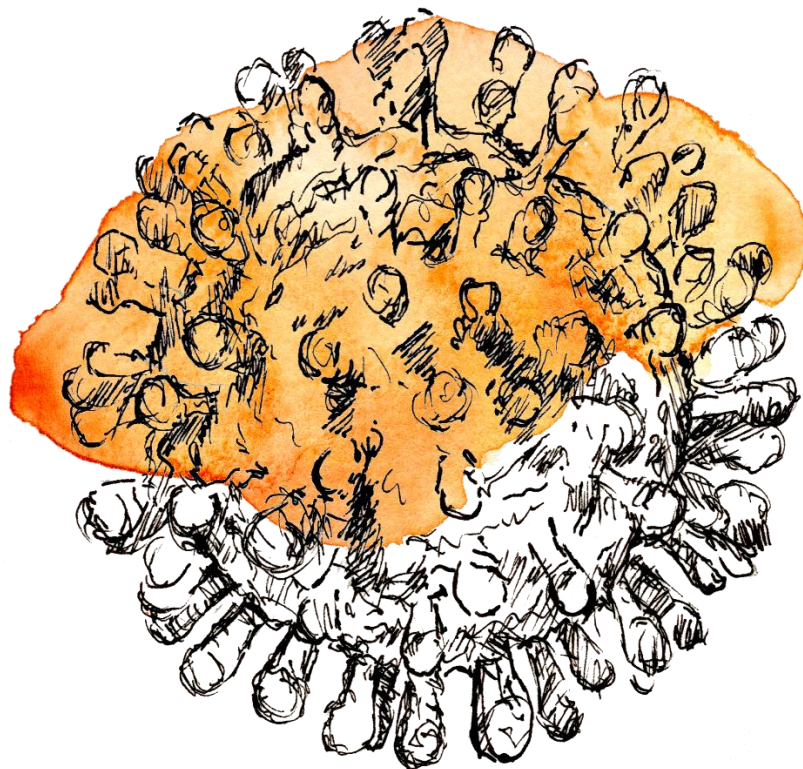


**Figure 50. RNAfold structures of the different windows containing the region that begins at position 2125 of the env of WT\_env and MinCpG and MinCpG.2 variants.** The color code represents the base-pairing probability of the individual bases predicted by RNAfold. Purple means low probability and red means high probability. The nucleotides indicated in Figure 47 A are circled in pink. A) WT\_env structures. B) MinCpG.2 structures. C) MinCpG structures. i) Structures obtained from the fragment 2075-2175. ii) Structures obtained from the fragment 2090-2190. iii) Structures obtained from the fragment 2100-2150.

## ***Results***

---

As mentioned above, the designed windows contained more mutations that were affecting the secondary structure. In addition, it must be taken into account that surrounding the windows there were even more designed mutations that were organized in their own secondary structures, probably generating interactions between them. These interactions may be the same as WT or different. The prediction of these interactions is impossible with the available computational tools.



## Discussion

---





Synonymous recoding of a gene or genome is a powerful tool to explore virus biology. It has been used to study several mechanisms, such as attenuation of viral replication (Cladel et al. 2013; Martrus et al. 2013; Mueller et al. 2006, 2010; Le Nouen et al. 2014; Vabret et al. 2014; Yang et al. 2013), identification of important RNA structures for viral replication (Song et al. 2012), identification of antiviral factors or mechanisms within the innate immune response (Li et al. 2012), unravelling the sequence space and evolution of a virus to attenuate its replication capacity (Moratorio et al. 2017), among others.

Here we show how synonymously recoding the genome of the HIV-1 has allowed us to study the influence of the codon bias in the evolvability and sequence space of a gene, to decipher an important RNA secondary structure for the viral replication, and to contribute to know how codon pair bias and the number of CpGs may affect the fitness of the virus.

Previous research has shown that codon usage can determine the mutational robustness, evolutionary capacity, and virulence of poliovirus (Lauring et al. 2012). This study showed that polioviruses with synonymously mutated capsids were less mutationally robust in tissue culture and displayed an attenuated phenotype in an animal model. However, that study did not focus on how synonymous mutations might affect the development of escape mutations to overcome specific selection pressure targeting a precise virus gene. Here we tested the extent to which a synonymously recoded HIV-1 protease reacted to the specific selective pressure of a PI. Our present study also explored the evolvability of a retrovirus which, in contrast to other RNA viruses, integrates into the host cell genome.

We found that the WT and MAX protease viruses displayed different patterns of resistance mutations after PI treatment. These findings extend those of Lauring et al. (Lauring et al. 2012), confirming that synonymously recoded and WT HIV-1 proteases occupy different sequence spaces. We further demonstrated that although the MAX and WT proteases occupied different sequence spaces, they still showed similar development of phenotypic resistance to PIs. These findings indicate that the recoded protease did not attenuate the virus' capability to develop PI resistance, strongly suggesting that the MAX protease was as robust as the WT protease with regards to this trait. To our knowledge, this is the first study to investigate the evolvability of a synonymously recoded virus enzyme. Our results show convincing evidence that recoded proteins occupy a different sequence space.

In some cases, the different mutant repertoire within the MAX protease background can be easily explained by proximity within the corresponding sequence space (e.g., G16E and L89I). However, with other mutations the explanation for the observed differences is not

## ***Discussion***

---

readily apparent (e.g., S37P, G48L, Q58E and I84V). In particular, the G48L substitution, which is very rarely selected *in vivo* in patients undergoing PI therapy (Gifford et al. 2009). When the above substitutions (S37P, G48L, Q58E, I84V or L89I) were introduced in the WT sequence they showed parallel replication capacities to those observed with a MAX background. We can speculate that the introduced synonymous substitutions affected neighboring residues (e.g., RNA structure). However, it must be noted that the MAX and WT proteases have similar RNA folding free energy (Martrus et al. 2013). Since some of these substitutions were only observed in one of the two independent experiments, it cannot be discarded that, in some cases, a mutational stochastic process was involved in the emergence of PI associated resistance.

One plausible explanation is based on epistatic interactions between protease amino acid substitutions. Epistasis is a phenomenon by which a mutation's impact on protein stability or fitness depends on the genetic background in which it is acquired (Parera and Martinez 2014). Complex mutational patterns often arise during the development of resistance to HIV-1 protease inhibitors. More therapy-associated mutations accumulate under PI therapy than under all other types of antiretroviral therapy (Shafer and Schapiro 2008). Recent findings suggest that the consequences of acquiring primary HIV-1 protease resistance mutations depend on epistatic interactions with the sequence background (Flynn et al. 2017). In our study, over 80% of the MAX protease clones harboring the G48L mutation also had the I89L substitution. As mentioned above, I89L requires two nucleotide substitutions in the WT background and only one in the MAX background. In either case, our results strongly suggest that the MAX protease's sequence position affects its genotypic PI resistance profile. Synonymous codons differ in their propensity to mutate and, as previously suggested (Cambray and Mazel 2008; Dolan, Whitfield, and Andino 2018; Lauring et al. 2012), this differential access to protein sequence space may affect adaptive pathways.

Another intriguing finding of our study is that the MAX virus showed higher population diversity in the recoded and targeted gene following propagation in both the absence and presence of PIs. Again, we can speculate that although the MAX virus shows high fitness in tissue culture, it is subjected to greater pressure to changing or reverting to a WT synonymous background. However, we detected almost no reversions of the synonymous substitutions introduced in the MAX protease following propagation in the presence or absence of PIs. One limitation of our study is that we investigated only one virus enzyme or protein. Further studies should include other virus proteins and other selective pressures (e.g., neutralizing antibodies and cellular virus restriction factors).

It has been suggested that RNA virus synonymous recoding can be used to push a virus to a sequence space region having a low density of neutral mutations (Lauring et al. 2012). Such a lack of access to neutral substitutions could potentially reduce the virus' capacity to generate fit progeny and adaptability to the host's selective pressures, such that this method might serve as a new strategy for development of attenuated vaccines. Our present data suggest that this approach must be developed cautiously, and support a need to evaluate the long-term stability of synonymously recoded viruses and to carefully test individual candidates.

Shin *et al.* demonstrated that the temporal regulation of the SIV viral genes was altered after synonymously changing the codon usage (Shin et al. 2015). To that end, the codon usage of the SIV *env* was modified to the codon usage of the RRV. After performing the changes, they observed that SIV Env expression was lost in the presence of Rev. Since in the former study only *in vitro* protein expression experiments were performed, here we wanted to explore how the same synonymous changes affect the HIV-1 *env* expression and the *ex vivo* replication capacity of the virus. We generated an *env* virus variant in which six codons were synonymously replaced as performed by Shin and colleagues. A total of 39 mutations were introduced in the *env* variant Recoded-*env*. To see if Recoded-*env* was viable, this mutant was transfected in MT-4 cells. After several transfections and blind passages, no HIV-1 p24 was detected. The 39 mutations introduced in Recoded-*env* resulted to be highly deleterious for the virus. We concluded that the mutations were significantly affecting the phenotype of the virus, thus showing how just a small proportion of synonymous mutations (1,5% of *env*) can be lethal for the virus.

To unravel the phenomenon behind this result, several mutants were designed based on Recoded-*env* virus sequence. Some of these new variants were designed by reverting mutations, while others were designed by reducing the amount of CpGs. After transfecting these new variants, we observed that almost all of them were able to replicate, although showing different replication capacities (Figure 30, Figure 31). After sequencing their *env* RNAs, new mutations were generated. Most probably, these new changes were compensatory mutations giving an advantage to these viruses. Compensatory mutations have been extensively described in previous studies as a mechanism of the virus to restore viral replication (González-Ortega et al. 2011; Liang et al. 1998). Interestingly, some of these compensatory mutations were located next to codon 34, in position 2125 of the *env*. Reversions of this codon 34 were also observed. These results suggested a strong relevance of codon 34 sequence for viral replication. This suggestion was supported by the

## Discussion

---

observation that Recoded\_env\_3'Awt, which conserved the mutations in codon 34, did not grow in any of the transfections. In contrast, Recoded\_env\_3'Bwt, with a fully reverted to WT codon 34, did grow without compensatory mutations. Indeed, Recoded\_env\_3'Bwt had no differences in replication when compared to WT\_env, both in MT-4 cells and PBMCs. Interestingly, we observed that the CpGs introduced in the *env* sequence and mutations in the 5' *env* region were not affecting virus replication capacity.

Analyses of the introduced codons in Recoded-env revealed that the number of the rare codon CGT was increased. The frequency of appearance of the codon CGT is 1,17 among different HIV-1 *envs* and 0,9 among different HIV-1 genomes. This codon is also rare in the human codon usage (frequency of 4,5). However, 3 more CGTs were introduced in Recoded-env, raising the frequency to 4,67 in *env*. Another rare codon introduced was CCG. The number of CCGs introduced was also 3, but the codon is not as rare (frequency of 6,9) as CGT in the human genome. Overall, these codon changes generated 9 out of the 39 total mutations introduced in the HIV-1 *env* of Recoded-env. Some previous works related codon usage and protein expression. It was suggested that when the HIV-1 Env coding region was codon-optimized, the expression was improved (Haas, Park, and Seed 1996). It was also described that the expression level of a protein is directly related to the presence of host rare codons, that is, the more rare codons, the less expression (Kane 1995). Our results showed a dramatic decrease of protein expression in Recoded-env and Recoded\_env\_3'Awt, correlating with the lethal phenotype of these virus variants. A possible hypothesis is that the introduction of these rare codons was the reason why a dramatic reduction in protein expression was observed. However, one observation that weakens this hypothesis is that the CAI scores were very similar between WT\_env and Recoded-env, meaning that the codon usage was not severely altered. The second observation that weakens that hypothesis is the fact that Recoded\_env\_3'Bwt could grow and replicate as WT\_env. Its protein expression is also increased when compared to Recoded-env. The difference between Recoded\_env\_3'Bwt and Recoded-env is only the reversion of one rare CGT codon, codon34. Therefore, Recoded\_env\_3'Bwt conserved almost all rare codons introduced in Recoded-env. Our results indicated that the introduction of rare codons was not responsible of the observed low protein expression.

Translation seemed to be affected by other causes. Consequently, we quantified virus mRNA to know whether virus transcription was affected. In several studies, an effect that can be observed after changing the codon usage is an alteration in the mRNA level. In yeast, it has been described that the mRNA stability is related to the amount of optimal or non-optimal gene codons. Those genes with optimal codons are more stable and are less

affected by degradation (Presnyak et al. 2015). Apart from mRNA degradation, it was also suggested that transcription rate was also playing a part in the levels of mRNA after altering the codon usage (Chen et al. 2017; Newman et al. 2016; Zhou et al. 2016). Our results here do not show significant differences in mRNA amounts. The levels of mRNA of Recoded-env and the related virus mutants were very similar to WT\_env mRNA. Moreover, total RNA was extracted forty-eight hours after transfection. If mRNA degradation was taking place, a significant difference may have been observed. Hence, it can be assumed that mRNA degradation was not occurring and transcription rate was not being affected. On the contrary, protein amounts were significantly reduced, pointing that translation, and not transcription, was the only altered process.

We hypothesized here that codon 34 may be affecting virus RNA secondary structures. Tools from Vienna RNA Package were used (Lorenz et al. 2011) to explore the RNA secondary structures of *env* fragments containing codon 34. After several conformations were obtained, we observed that codon 34 (AGG) is positioned into an internal loop in WT\_env, in which the nucleotide A is unpaired and the nucleotides GG are paired. In addition, before codon 34, there is an external loop (hairpin loop). This structure was confirmed by predictions in HIV-1 *env* isolates alignments, suggesting that the predicted structure was not random and it is conserved among different HIV-1 *env* sequences. Importantly, the former RNA structures are also conserved in the SHAPE analysis of HIV-1 genome previously performed (Watts et al. 2009). Analyzing the secondary structure obtained from SHAPE, we observed that the conformation was conserved: AGG codon is located into an internal loop while there is a hairpin loop just before. The information that SHAPE gives reinforces the computational predictions. When the computational tools were applied in Recoded-env, we observed that the WT\_env secondary structure was disrupted and both the internal loop in which codon 34 was located and the hairpin loop were altered. The RNA secondary structure of this variant presented an unstable conformation in which two new branches were formed from the internal loop and hairpin present in WT\_env. In addition, the relevance of this WT RNA secondary structure was highlighted by Recoded\_env\_sCpG2. This virus variant had a mutation in codon 34 different from WT\_env and Recoded-env. This mutant was not lethal but its replication capacity was significantly lower when compared to WT\_env. After analyzing the RNA secondary structures for this virus, new conformations were observed. Some of the new obtained conformations from different window fragments were very similar to WT\_env, although not identical. We hypothesize that the RNA secondary structure of this variant altered WT\_env conformation, but the change was not as severe as the one observed in Recoded-env. This change of

## Discussion

---

structure was not lethal for Recoded\_env\_SCpG2, but instead reduced its replication capacity.

Several studies have described how RNA secondary structures are necessary for viral replication. When these RNA secondary structures are disrupted, the replication capacity of the virus is lowered or lost (Madhugiri et al. 2017; Simon et al. 2019; Song et al. 2012). The relevance of studying the secondary structures of the viral RNAs is increasing, due to the fact that these studies are helping to understand virus biology. An example of this is the very conserved RNA secondary structure of the HIV-1 RRE, which loses its capacity to join to Rev when the conformation is disrupted (Chu et al. 2019). Our results show that when codon 34 was reverted to WT, the structure and the viral capacity of replication were recovered, as seen in Recoded\_env\_3'Bwt. Here we have used computational tools to study why Recoded\_env\_3'Bwt replicated while Recoded-env did not. These tools helped us to describe a plausible important RNA secondary structure located in gp41 coding region of the HIV-1 genome, which confirms the relevance of virus RNA secondary structures.

A possible explanation of why disrupting the former RNA structure led to virus lethality may be the loss of interactions. RNA secondary structures in which external loops are located tend to be important for the interaction or recognition by other molecules or molecular factors (Bannwarth and Gatignol 2005; Dzananovic et al. 2013). By the disruption of the WT\_env structure, we postulate that unknown interactions with the external hairpin were lost in Recoded-env. When Shin *et al.* performed the same changes in SIV *env*, the authors suggested that Rev lost its ability to join to the RRE. This was observed by the lowered protein expression and the reduction of mRNA amounts (Shin et al. 2015). However, our results showed that the amounts of mRNA were maintained as the WT\_env. Thus, the interaction with Rev was not lost in Recoded-env. Long-range RNA-RNA interactions were also examined. However, we cannot detect any long-range interactions between the hairpin or the internal loop containing codon 34. The tool used was also computational and, as described, it is only a predictive tool. We believe that other unknown interactions were altered or new ones leading to lethality were generated.

Even though we have not unraveled which interactions are lost or which new are formed that impede protein production and viral replication, throughout these studies we have described an important and conserved RNA secondary structure that is essential for viral replication and protein translation. Expanding our knowledge on HIV-1 structures and interactions will allow to understand better the replication cycle of the virus. Recent data pointed out the importance of acting against the increasing number of HIV drug resistances (World Health Organization 2017). In the last seven years, it has been observed that the

HIV-1 has started to generate drug resistance mutations at a very fast rate. Consequently, stopping the generation of these resistances has become a very important issue to fight the worldwide pandemic. HAART is based on the combination of three different drugs to avoid viral escape. The combination with a new drug targeting this newly described region would prevent the generation of new resistance mutations. Thus, further studies on the RNA secondary structure that we describe here can lead to the development of a new drug therapy based on targeting this conformation. Indeed, the interest on RNA structures as therapeutic targets is increasing (Connelly, Moon, and Schneekloth Jr 2016; Cooper, Wan, and Dreyfuss 2009). The HIV-1 genome is formed by non-coding, cis-acting functional elements with functional roles at different levels in the HIV-1 cycle. Some of these elements are the transactivation response (TAR) hairpin, target of Tat to enhance transcription, or the Rev response element (RRE), target of Rev for the nucleocytoplasmic transport. Both of these structural elements have been studied as therapeutic targets (Le Grice 2015). The selection of small molecules to target specific RNA structures have proved better absorption, distribution and oral bioavailability than the use of antisense technologies such as siRNAs (Connelly et al. 2016). Actually, promising results were obtained with one compound that selectively binds to TAR, inhibiting HIV-1 replication (Stelzer et al. 2011). Throughout the study presented here, we open the door to study the interactions of the described RNA secondary structure to further exploit them.

As previously discussed, the 3' *env* region (gp41) of the HIV-1 sequence have a relevant role in viral fitness. When the codon pair usage is changed to optimize or deoptimize the CPB score of *env*, none of the designed viruses can grow, except for MinCpG.2. In contrast, when this 3' *env* region was reverted to WT in the lethal virus variants, almost all viruses recovered their replication capacity. Lots of efforts have been made on studying the amino acid sequence conservation of gp41 and its different roles (Ashkenazi et al. 2013; Holguín, De Arellano, and Soriano 2007; Ramakrishnan et al. 2006; Steckbeck et al. 2011), in contrast to the sequence variability that can be observed in gp120 (Hoffman et al. 2002; Yamaguchi-kabata and Gojobori 2000). Here we observed how changing the nucleotide sequence of gp41, but maintaining the amino acid sequence, led to lethality. This result demonstrates how the nucleotide sequence, as the amino acid sequence, can impact viral fitness. As observed in Recoded-*env*, the most probable explanation is that RNA secondary structures were disrupted in gp41 of the CPB variants. However, MinCpG.2 replicated even containing several mutations in its 3' *env* region. Studying the 3' *env* sequence of all the CPB variants, we identified two possible important sequence regions within gp41. Actually, one of the two regions corresponded to codon 34 discussed above. These regions were described based on MinCpG.2 sequence. This virus variant was the only CPB mutant with



## Discussion

---

no mutations in these two regions. To see if RNA secondary structures were altered in these two regions, computational tools were applied. The comparisons were made between MinCpG, MinCpG.2 and WT, because MinCpG shared similar characteristics with MinCpG.2. After obtaining the RNA secondary structures we suspected that changes in the conformation happened. However, we were not able to obtain a stable conformation for each of the analyzed viruses, MinCpG and MinCpG.2. The reason was that several mutations were introduced in gp41 in the different variants, making it difficult to obtain the RNA secondary structure. Predictions with one or two mutations are easy to obtain and to interpret, but when several mutations are introduced, the number of possibilities increases. Thus, deciphering the existing conformation is almost impossible. Nevertheless, although the number of mutations was high and stable structures were not obtained, the MinCpG RNA secondary structures were different from WT\_env, suggesting changes that were probably lethal for the viral replication. In addition, comparisons between RNA secondary structures of MinCpG and MinCpG.2 revealed differences that explained why MinCpG.2 was replicative and MinCpG was not.

Deoptimizing the codon pair usage of the viral genes has been widely used to generate attenuated viruses (Coleman et al. 2008; Khedkar, Osterrieder, and Kunec 2018; Martrus et al. 2013; Mueller et al. 2006; Ni et al. 2014; Le Nouen et al. 2014; Wimmer et al. 2009; Yang et al. 2013). In addition, altering the codon pair bias has not only been used to generate viruses with lower replication capacity, but also to obtain viruses with the same replication capacity as the WT (Martrus et al. 2013; Nevot et al. 2018). Our results show how decreasing or increasing the CPB score does not always attenuate or maintains the virus replication capacity. When codon pairs with statistically overrepresentation were introduced in the *env* gene and, consequently, the CPB score was increased, an improvement was not observed in replication capacity. On the contrary, Max-3'wt and MaxCpG-3'wt had significantly lower replication capacities when compared to WT\_env. Moreover, MinCpG-3'wt, which was designed with higher amount of statistically underrepresented codon pairs, did not show an attenuated phenotype but displayed the same replication capacity as WT\_env. With these results we observe that increasing or reducing the CPB score does not, respectively, improve or worsen the replication capacity of the virus. This remarks the importance of the technique to unravel different aspects of the viral biology.

Protein expression of the CPB variants with the 3' *env* region reverted to WT was analyzed. It was observed that Max-3'wt and MaxCpG-3'wt showed reduced protein expression, corresponding to their replication capacity. These two mutants were CPB-optimized to the

human usage. Also, their CAI increased when compared to WT, meaning that their codon usage was also optimized. Previous research identified an interferon-stimulated protein, Schlafen 11 (SLFN11), as a mechanism of the innate immune system to selectively inhibit viral protein synthesis of the HIV-1 in a codon-usage dependent manner (Li et al. 2012). The authors demonstrated how codon-optimizing the HIV-1 genome for the human usage reduces the effect of SLFN11, and more viral protein is expressed. Here we observe how codon-optimizing two viruses did not produce an increase in protein expression. Thus, our work shows that it is very unlikely that the reduced protein expression and, consequently, the reduced viral fitness, is due to SLFN11. Moreover, all CPB variants had better CAI values than WT\_env. However, none of them improved WT\_env protein expression, suggesting, once more, that SLFN11 was not affecting our virus variants.

Recent data linked the effect of codon pair bias to a dinucleotide bias (Kunec and Osterrieder 2016; Tulloch et al. 2014). Specifically, it was observed that the viral replication capacity is attenuated when the number of CpGs/UpAs is increased in the viral genomes (Atkinson et al. 2014; Burns et al. 2009; Gaunt et al. 2016; Tulloch et al. 2014). In general, in many RNA viruses the frequency of CpGs is lower than expected (Auewarakul 2005; Karlin et al. 1994; Rima and McFerran 1997; Simmonds et al. 2013). This is not an exception for HIV-1 genome, which also shows a much lower than expected amount of CpGs (Bronson and Anderson 1994; Kypr et al. 1989). For this reason, works studying the effect of CpGs in the HIV-1 genome have been performed. As shown with other viruses, some studies in HIV-1 showed that increasing the number of CpGs in the HIV-1 genome inhibits the viral replication (Antzin-Anduetza et al. 2017; Takata et al. 2017). Also, that the proportion of this dinucleotide in the HIV-1 *env* sequence may predict the disease progression (Wasson et al. 2017). Recently, it was measured the cost of generating mutations that create new CpG dinucleotides in *pol*. It was shown that the cost was twice as the cost of mutations that do not create them (Theys et al. 2018). This suggests that creating new CpGs is negatively selected by the virus due to the probability of reducing the replication capacity by recognition of the innate immune system. Nevertheless, our results here do not suggest the same. Both Neu-3'wt and NeuCpG-3'wt show a replication capacity similar to WT\_env (Figure 43), with no significant differences between Neu-3'wt and NeuCpG-3'wt. Both virus variants contain more CpGs than WT\_env (47 and 30, respectively). Interestingly, although Max-3'wt and MaxCpG-3'wt display significantly lower fitness than WT\_env, when the slopes of the two virus variants were compared between them, Max-3'wt showed a significantly better replication capacity than MaxCpG-3'wt (Figure 43). However, Max-3'wt variant have more CpGs in its sequence than MaxCpG-3'wt (39 and 24, respectively). Moreover, the number of CpG dinucleotides in the Neu variants is higher than in the Max variants. Overall, we see

## Discussion

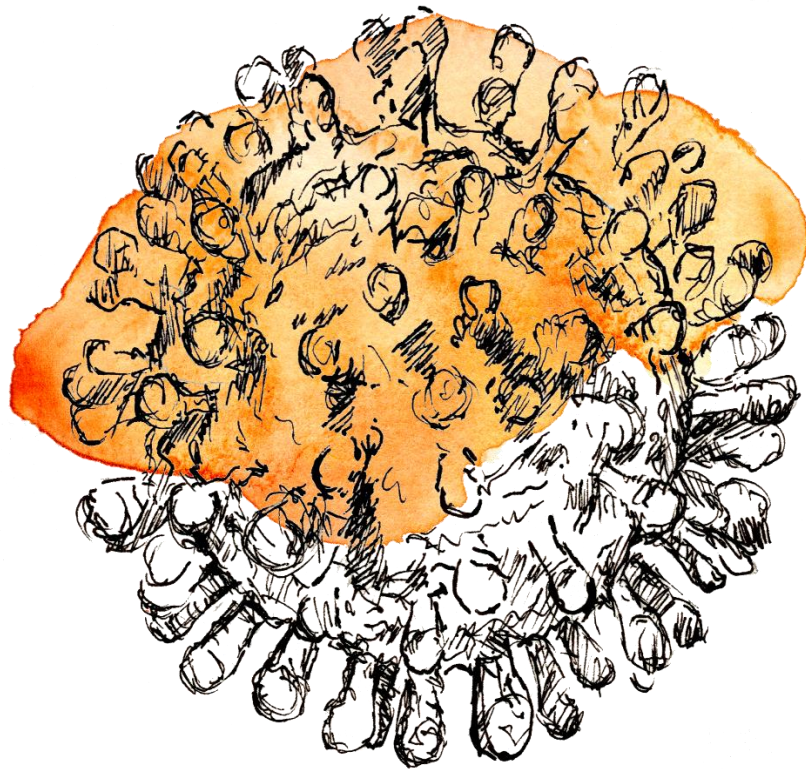
---

that the number of CpGs is not critical for the replication capacity of the viruses. We observe how variants with higher amounts of CpGs, Neu-3'wt and NeuCpG-3'wt, show better replication capacity than variants with lower amounts of CpGs, Max-3'wt and MaxCpG-3'wt.

Several factors of the innate immune response were described as possible causes of the attenuated viral phenotypes observed by others when CpG dinucleotides are incremented. The first of these factors is toll-like receptor 9 (TLR9). TLRs are transmembrane proteins that recognize pathogen-associated molecular patterns (PAMPs). In the case of TLR9, this receptor recognizes unmethylated CpGs in the DNA (Pandey et al. 2015). TLR9 might be responsible for the reduced replication capacity when CpGs are introduced in the genome, since HIV-1 has a DNA intermediate. However, other RNA viruses that do not have a DNA intermediate do also show a suppress frequency of CpGs in their genomes and also suffer attenuation when this dinucleotide is introduced (Atkinson et al. 2014; Gaunt et al. 2016), suggesting that TLR9 might not be the responsible for virus attenuation. The second innate immune factor recently described as responsible for virus attenuation is ZAP (Takata et al. 2017). Takata and colleagues described how virus replication in MT-4 cells was inhibited after introducing CpGs in HIV-1 *env*. They also showed how the replication of the virus was rescued after inhibiting ZAP in the same cells. They suggest that ZAP selectively joins non-self, viral RNAs with high abundance of CpG dinucleotides, promoting RNA degradation. As mentioned above, Neu-3'wt and NeuCpG-3'wt have more CpG dinucleotides than WT\_Env. Despite increasing the frequency of CpGs, attenuation is not observed in MT-4 cells. Max-3'wt and MaxCpG-3'wt, variants that have less CpG number than Neu-3'wt and NeuCpG-3'wt, showed an attenuated replication capacity. When the distribution of the CpGs was analyzed, we observed a region between positions 368-583 of *env* where Max-3'wt and MaxCpG-3'wt did have CpGs, whereas in the same region of the other virus variants no CpGs are observed. What we propose here is that location and distribution of the CpGs may be also crucial for ZAP function, and not only increasing the frequency of CpGs. If it was the number of CpGs, Neu-3'wt would be attenuated, and Max-3'wt would replicate worse than MaxCpG-3'wt. This hypothesis can also be reinforced by the fact that Takata *et al.* saw the effect of ZAP in a mutant that concentrated all the CpGs in a small region of *env*. Moreover, Antzin-Anduetza *et al.* (Antzin-Anduetza et al. 2017), who also described attenuation of the HIV-1 after introducing CpGs, saw the attenuated phenotype after introducing the mutations in small regions of the MA gene. The number of CpGs introduced by these two authors did not differ much from the number that we introduced. The main difference, as mentioned above, is the location and distribution of the dinucleotides. Although our results suggest this phenomenon, further studies need to be

performed to unravel how ZAP or a possible unknown mechanism are affecting our virus variants.





---

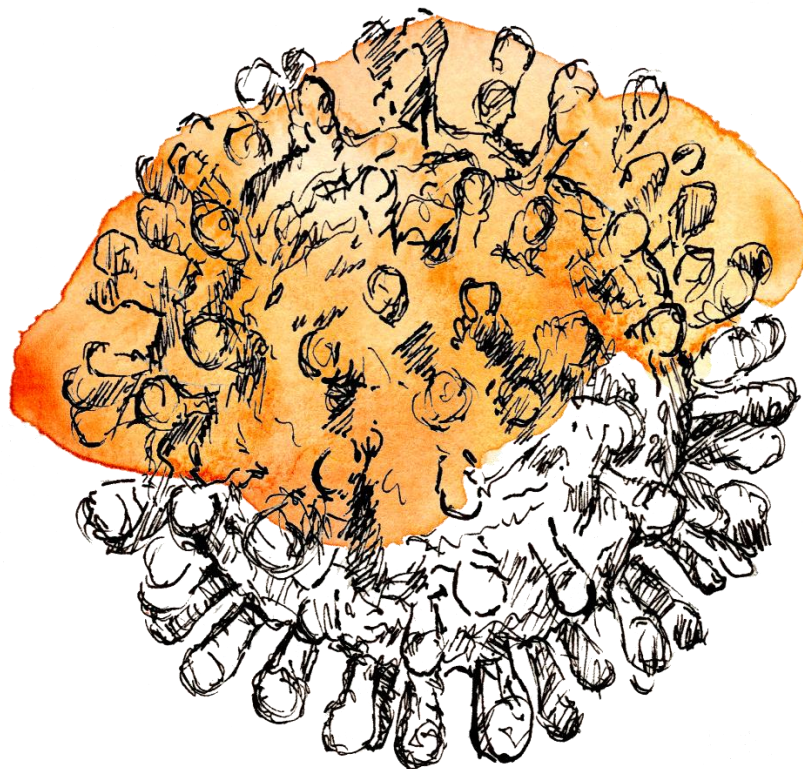
## Conclusions



- ✓ HIV-1 protease position in sequence space delineates the evolution of its mutant spectrum.
- ✓ Wild-type and synonymously recoded protease gene displays different patterns of resistance mutations after treatment of protease inhibitors, confirming that HIV-1 synonymously recoded and wild-type protease occupy different sequences spaces. Although the wild-type and synonymously recoded protease gene occupy different sequence spaces, they still show similar levels of development of phenotypic resistance to protease inhibitors.
- ✓ Changes in the codon usage of the HIV-1 *env* gene strongly impact the replication capacity of the virus.
- ✓ Synonymous recoding of HIV-1 *env* gene has identified, in the gp41 coding region, an evolutionary conserved local RNA secondary structure that may be essential for virus viability. Disruption of this RNA secondary structure lead to severe reduction in mRNA translation and virus replication capacity.
- ✓ Synonymously increasing the number of CpGs in the *env* gene does not reduce, in our model, the replication capacity of the virus.
- ✓ Virus genome synonymous recoding is a powerful tool to impact virus phenotype and to explore virus biology.







---

## **Bibliography**



- Akashi, H.** 1994. "Synonymous Codon Usage in *Drosophila Melanogaster*: Natural Selection and Translational Accuracy." *Genetics* 136(3):927–35.
- Anderson, J. L. and T. J. Hope.** 2008. "APOBEC3G Restricts Early HIV-1 Replication in the Cytoplasm of Target Cells." *Virology* 375(1):1–12.
- Andino, R. and E. Domingo.** 2015. "Viral Quasispecies." *Virology* 479–480:46–51.
- Antzin-Anduetza, I., C. Mahiet, L. A. Granger, C. Odendall, and C. M. Swanson.** 2017. "Increasing the CpG Dinucleotide Abundance in the HIV-1 Genomic RNA Inhibits Viral Replication." *Retrovirology* 14(1):49.
- Armitage, A. E., K. Deforche, C. H. Chang, E. Wee, B. Kramer, J. J. Welch, J. Gerstoft, L. Fugger, A. McMichael, A. Rambaut, and A. K. N. Iversen.** 2012. "APOBEC3G-Induced Hypermutation of Human Immunodeficiency Virus Type-1 Is Typically a Discrete 'All or Nothing' Phenomenon." *PLoS Genetics* 8(3):e1002550.
- Ashkenazi, A., O. Faingold, N. Kaushansky, A. Ben-Nun, and Y. Shai.** 2013. "A Highly Conserved Sequence Associated with the HIV Gp41 Loop Region Is an Immunomodulator of Antigen-Specific T Cells in Mice." *Blood* 121(12):2244–52.
- Atkinson, N. J., J. Witteveldt, D. J. Evans, and P. Simmonds.** 2014. "The Influence of CpG and UpA Dinucleotide Frequencies on RNA Virus Replication and Characterization of the Innate Cellular Pathways Underlying Virus Attenuation and Enhanced Replication." *Nucleic Acids Research* 42(7):4527–45.
- Auewarakul, P.** 2005. "Composition Bias and Genome Polarity of RNA Viruses." *Virus Research* 109(1):33–37.
- Bannwarth, S. and A. Gatignol.** 2005. "HIV-1 TAR RNA: The Target of Molecular Interactions between the Virus and Its Host." *Current HIV Research* 3(1):61–71.
- Barre-Sinoussi, F., J. Chermann, F. Rey, M. Nugeyre, S. Chamaret, J. Gruest, C. Dautuet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. "Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS)." *Science* 220(4599):868–71.
- Betancor, G., M. Álvarez, B. Marcelli, C. Andrés, M. A. Martínez, and L. Menéndez-Arias.** 2015. "Effects of HIV-1 Reverse Transcriptase Connection Subdomain Mutations on Polypurine Tract Removal and Initiation of (+)-Strand DNA Synthesis." *Nucleic Acids Research* 43(4):2259–70.
- Boël, G., R. Letso, H. Neely, W. N. Price, K. H. Wong, M. Su, J. D. Luff, M. Valecha, J. K. Everett, T. B. Acton, R. Xiao, G. T. Montelione, D. P. Aalberts, and J. F. Hunt.** 2016. "Codon Influence on Protein Expression in *E. Coli* Correlates with mRNA Levels." *Nature* 529(7586):358–63.
- Bohne, J., H. Wodrich, and H. G. Kräusslich.** 2005. "Splicing of Human Immunodeficiency Virus RNA Is Position-Dependent Suggesting Sequential Removal of Introns from the 5' End." *Nucleic Acids Research* 33(3):825–37.
- Borman, A. M., S. Paulous, and F. Clavel.** 1996. "Resistance of Human Immunodeficiency Virus Type 1 to Protease Inhibitors: Selection of Resistance Mutations in the Presence and Absence of the Drug." *Journal of General Virology* 77(3):419–26.
- Bronson, E. C. and J. N. Anderson.** 1994. "Nucleotide Composition as a Driving Force in the Evolution of Retroviruses." *Journal of Molecular Evolution* 38(5):506–32.

- Brule, C. E. and E. J. Grayhack.** 2017. "Synonymous Codons: Choose Wisely for Expression." *Trends in Genetics* 33(4):283–97.
- Buchan, J. R., L. S. Aucott, and I. Stansfield.** 2006. "tRNA Properties Help Shape Codon Pair Preferences in Open Reading Frames." *Nucleic Acids Research* 34(3):1015–27.
- Burns, C. C., R. Campagnoli, J. Shaw, A. Vincent, J. Jorba, and O. Kew.** 2009. "Genetic Inactivation of Poliovirus Infectivity by Increasing the Frequencies of CpG and UpA Dinucleotides within and across Synonymous Capsid Region Codons." *Journal of Virology* 83(19):9957–69.
- Burns, C. C., J. Shaw, R. Campagnoli, J. Jorba, A. Vincent, J. Quay, and O. Kew.** 2006. "Modulation of Poliovirus Replicative Fitness in HeLa Cells by Deoptimization of Synonymous Codon Usage in the Capsid Region." *Journal of Virology* 80(7):3259–72.
- Cambray, G. and D. Mazel.** 2008. "Synonymous Genes Explore Different Evolutionary Landscapes" edited by H. S. Malik. *PLoS Genetics* 4(11):e1000256.
- Capel, E., G. Martrus, M. Parera, B. Clotet, and M. A. Martinez.** 2012. "Evolution of the Human Immunodeficiency Virus Type 1 Protease: Effects on Viral Replication Capacity and Protease Robustness." *Journal of General Virology* 93(Pt\_12):2625–34.
- Carrillo, A., K. D. Stewart, H. L. Sham, D. W. Norbeck, W. E. Kohlbrenner, J. M. Leonard, D. J. Kempf, and A. Molla.** 1998. "In Vitro Selection and Characterization of Human Immunodeficiency Virus Type 1 Variants with Increased Resistance to ABT-378, a Novel Protease Inhibitor." *Journal of Virology* 72(9):7532–41.
- Chang, D. D. and P. A. Sharp.** 1989. "Regulation by HIV Rev Depends upon Recognition of Splice Sites." *Cell* 59(5):789–95.
- Charif, D., J. Thioulouse, J. R. Lobry, and G. Perri re.** 2005. "Online Synonymous Codon Usage Analyses with the Ade4 and SeqinR Packages." *Bioinformatics* 21(4):545–47.
- Chen, S., K. Li, W. Cao, J. Wang, T. Zhao, Q. Huan, Y. F. Yang, S. Wu, and W. Qian.** 2017. "Codon-Resolution Analysis Reveals a Direct and Context-Dependent Impact of Individual Synonymous Mutations on mRNA Level." *Molecular Biology and Evolution* 34(11):2944–58.
- Chen, Z., Y. Li, H. B. Schock, D. Hall, E. Chen, and L. C. Kuo.** 1995. "Three-Dimensional Structure of a Mutant HIV-1 Protease Displaying Cross-Resistance to All Protease Inhibitors in Clinical Trials." *Journal of Biological Chemistry* 270(37):21433–36.
- Chu, C. C., R. Plangger, C. Kreutz, and H. M. Al-Hashimi.** 2019. "Dynamic Ensemble of HIV-1 RRE Stem IIB Reveals Non-Native Conformations That Disrupt the Rev-Binding Site." *Nucleic Acids Research* 47(13):7105–17.
- Cladel, N. M., L. R. Budgeon, J. Hu, K. K. Balogh, and N. D. Christensen.** 2013. "Synonymous Codon Changes in the Oncogenes of the Cottontail Rabbit Papillomavirus Lead to Increased Oncogenicity and Immunogenicity of the Virus." *Virology* 438(2):70–83.
- Coffin, J. M.** 1992. "Structure and Classification of Retroviruses." Pp. 19–49 in *The Retroviridae*, edited by J. A. Levy. VIRS.
- Coffin, J. M.** 1995. "HIV Population Dynamics in Vivo: Implications for Genetic Variation, Pathogenesis, and Therapy." *Science (New York, N.Y.)* 267(5197):483–89.

- Coleman, J. R., D. Papamichail, S. Skiena, B. Futcher, E. Wimmer, and S. Mueller.** 2008. "Virus Attenuation by Genome-Scale Changes in Codon Pair Bias." *Science* 320(5884):1784–87.
- Connelly, C. M., M. H. Moon, and J. S. Schneekloth Jr.** 2016. "The Emerging Role of RNA as a Therapeutic Target for Small Molecules." *Cell Chemical Biology* 23(9):1077–90.
- Cooper, T. A., L. Wan, and G. Dreyfuss.** 2009. "RNA and Disease." *Cell* 136(4):777–93.
- Cullen, B. R.** 1991. "Regulation of HIV-1 Gene Expression." *The FASEB Journal* 5(10):2361–68.
- Darke, P. L., R. F. Nutt, S. F. Brady, V. M. Garsky, T. M. Ciccarone, C. T. Leu, P. K. Lumma, R. M. Freidinger, D. F. Veber, and I. S. Sigal.** 1988. "HIV-1 Protease Specificity of Peptide Cleavage Is Sufficient for Processing of Gag and Pol Polyproteins." *Biochemical and Biophysical Research Communications* 156(1):297–303.
- Dayton, A.** 1986. "The Trans-Activator Gene of the Human T Cell Lymphotropic Virus Type III Is Required for Replication." *Cell* 44(6):941–47.
- Dayton, E., D. Powell, and A. Dayton.** 1989. "Functional Analysis of CAR, the Target Sequence for the Rev Protein of HIV-1." *Science* 246(4937):1625–29.
- Debouck, C.** 1992. "The HIV-1 Protease as a Therapeutic Target for AIDS." *AIDS Research and Human Retroviruses* 8(2):153–64.
- Derdeyn, C. A., J. M. Decker, J. N. Sfakianos, X. Wu, W. A. O'Brien, L. Ratner, J. C. Kappes, G. M. Shaw, and E. Hunter.** 2000. "Sensitivity of Human Immunodeficiency Virus Type 1 to the Fusion Inhibitor T-20 Is Modulated by Coreceptor Specificity Defined by the V3 Loop of Gp120." *Journal of Virology* 74(18):8358–67.
- Doehle, B. P., K. Chang, A. Rustagi, J. McNevin, M. J. McElrath, and M. Gale.** 2012. "Vpu Mediates Depletion of Interferon Regulatory Factor 3 during HIV Infection by a Lysosome-Dependent Mechanism." *Journal of Virology* 86(16):8367–74.
- Dolan, P. T., Z. J. Whitfield, and R. Andino.** 2018. "Mapping the Evolutionary Potential of RNA Viruses." *Cell Host & Microbe* 23(4):435–46.
- Domingo, E., D. Sabo, T. Taniguchi, and C. Weissmann.** 1978. "Nucleotide Sequence Heterogeneity of an RNA Phage Population." *Cell* 13(4):735–44.
- DuBridge, R. B., P. Tang, H. C. Hsia, P. M. Leong, J. H. Miller, and M. P. Calos.** 1987. "Analysis of Mutation in Human Cells by Using an Epstein-Barr Virus Shuttle System." *Molecular and Cellular Biology* 7(1):379–87.
- Dunn, B. M., M. M. Goodenow, A. Gustchina, and A. Wlodawer.** 2002. "Retroviral Proteases." *Genome Biology* 3(4):reviews3006.1-reviews3006.7.
- Duret, L.** 2000. "tRNA Gene Number and Codon Usage in the *C. Elegans* Genome Are Co-Adapted for Optimal Translation of Highly Expressed Genes." *Trends in Genetics* 16(7):287–89.
- Dzananovic, E., T. R. Patel, S. Deo, K. McEleney, J. Stetefeld, and S. A. McKenna.** 2013. "Recognition of Viral RNA Stem-Loops by the Tandem Double-Stranded RNA Binding Domains of PKR." *RNA* 19(3):333–44.
- Eder, J., U. Hommel, F. Cumin, B. Martoglio, and B. Gerhartz.** 2007. "Aspartic Proteases in Drug Discovery." *Current Pharmaceutical Design* 13(3):271–85.

- Fernandes, J., B. Jayaraman, and A. Frankel.** 2012. "The HIV-1 Rev Response Element." *RNA Biology* 9(1):6–11.
- Fernandez, G., B. Clotet, and M. A. Martinez.** 2007. "Fitness Landscape of Human Immunodeficiency Virus Type 1 Protease Quasispecies." *Journal of Virology* 81(5):2485–96.
- Findlay, G. M., R. M. Daza, B. Martin, M. D. Zhang, A. P. Leith, M. Gasperini, J. D. Janizek, X. Huang, L. M. Starita, and J. Shendure.** 2018. "Accurate Classification of BRCA1 Variants with Saturation Genome Editing." *Nature* 562(7726):217–22.
- Fisher, A. G., M. B. Feinberg, S. F. Josephs, M. E. Harper, L. M. Marselle, G. Reyes, M. A. Gonda, A. Aldovini, C. Debouk, R. C. Gallo, and F. Wong-Staal.** 1986. "The Trans-Activator Gene of HTLV-III Is Essential for Virus Replication." *Nature* 320(6060):367–71.
- Flynn, W. F., A. Haldane, B. E. Torbett, and R. M. Levy.** 2017. "Inference of Epistatic Effects Leading to Entrenchment and Drug Resistance in HIV-1 Protease." *Molecular Biology and Evolution* 34(6):1291–1306.
- Freed, E. O.** 2001. "HIV-1 Replication." *Somatic Cell and Molecular Genetics* 26(1–6):13–33.
- Fricke, M. and M. Marz.** 2016. "Prediction of Conserved Long-Range RNA-RNA Interactions in Full Viral Genomes." *Bioinformatics* 32(19):2928–35.
- Friedman-Kien, A. E.** 1981. "Disseminated Kaposi's Sarcoma Syndrome in Young Homosexual Men." *Journal of the American Academy of Dermatology* 5(4):468–71.
- Fujita, Y., H. Otsuki, Y. Watanabe, M. Yasui, T. Kobayashi, T. Miura, and T. Igarashi.** 2013. "Generation of a Replication-Competent Chimeric Simian-Human Immunodeficiency Virus Carrying Env from Subtype C Clinical Isolate through Intracellular Homologous Recombination." *Virology* 436(1):100–111.
- Futcher, B., O. Gorbatsevych, S. H. Shen, C. B. Stauff, Y. Song, B. Wang, J. Leatherwood, J. Gardin, A. Yurovsky, S. Mueller, and E. Wimmer.** 2015. "Reply to Simmonds et Al.: Codon Pair and Dinucleotide Bias Have Not Been Functionally Distinguished." *Proceedings of the National Academy of Sciences* 112(28):E3635–36.
- Gaddis, N. C., E. Chertova, A. M. Sheehy, L. E. Henderson, and M. H. Malim.** 2003. "Comprehensive Investigation of the Molecular Defect in Vif-Deficient Human Immunodeficiency Virus Type 1 Virions." *Journal of Virology* 77(10):5810–20.
- Gaunt, E., H. M. Wise, H. Zhang, L. N. Lee, N. J. Atkinson, M. Q. Nicol, A. J. Highton, P. Klenerman, P. M. Beard, B. M. Dutia, P. Digard, and P. Simmonds.** 2016. "Elevation of CpG Frequencies in Influenza A Genome Attenuates Pathogenicity but Enhances Host Response to Infection." *ELife* 5:e12735.
- Gifford, R. J., T. F. Liu, S. Y. Rhee, M. Kiuchi, S. Hue, D. Pillay, and R. W. Shafer.** 2009. "The Calibrated Population Resistance Tool: Standardized Genotypic Estimation of Transmitted HIV-1 Drug Resistance." *Bioinformatics* 25(9):1197–98.
- González-Ortega, E., E. Ballana, R. Badia, B. Clotet, and J. A. Esté.** 2011. "Compensatory Mutations Rescue the Virus Replicative Capacity of VIRIP-Resistant HIV-1." *Antiviral Research* 92(3):479–83.
- Goodman, D. B., G. M. Church, and S. Kosuri.** 2013. "Causes and Effects of N-Terminal Codon Bias in Bacterial Genes." *Science* 342(6157):475–79.

- Gregori, J., J. I. Esteban, M. Cubero, D. Garcia-Cehic, C. Perales, R. Casillas, M. Alvarez-Tejado, F. Rodríguez-Frías, J. Guardia, E. Domingo, and J. Quer.** 2013. "Ultra-Deep Pyrosequencing (UDPS) Data Treatment to Study Amplicon HCV Minor Variants" edited by O. Schildgen. *PLoS ONE* 8(12):e83361.
- Le Grice, S. F. J.** 2015. "Targeting the HIV RNA Genome: High-Hanging Fruit Only Needs a Longer Ladder." *Current Topics in Microbiology and Immunology* 389:147–69.
- Gulnik, S. V., L. I. Suvorov, B. Liu, B. Yu, B. Anderson, H. Mitsuya, and J. W. Erickson.** 1995. "Kinetic Characterization and Cross-Resistance Patterns Of HIV-1 Protease Mutants Selected under Drug Pressure." *Biochemistry* 34(29):9282–87.
- Gustafsson, C., S. Govindarajan, and J. Minshull.** 2004. "Codon Bias and Heterologous Protein Expression." *Trends in Biotechnology* 22(7):346–53.
- Gutman, G. A. and G. W. Hatfield.** 1989. "Nonrandom Utilization of Codon Pairs in Escherichia Coli." *Proceedings of the National Academy of Sciences* 86(10):3699–3703.
- Gyuris, A., G. Vajda, and I. Földes.** 1991. "Establishment of an MT4 Cell Line Persistently Producing Infective HIV-1 Particles." *Acta Microbiologica Hungarica* 39(3–4):271–79.
- Haas, J., E. C. Park, and B. Seed.** 1996. "Codon Usage Limitation in the Expression of HIV-1 Envelope Glycoprotein." *Current Biology* 6(3):315–24.
- Hoffman, N. G., F. Seillier-Moiseiwitsch, J. Ahn, J. M. Walker, and R. Swanstrom.** 2002. "Variability in the Human Immunodeficiency Virus Type 1 Gp120 Env Protein Linked to Phenotype-Associated Changes in the V3 Loop." *Journal of Virology* 76(8):3852–64.
- Holguín, A., E. R. De Arellano, and V. Soriano.** 2007. "Amino Acid Conservation in the Gp41 Transmembrane Protein and Natural Polymorphisms Associated with Enfuvirtide Resistance across HIV-1 Variants." *AIDS Research and Human Retroviruses* 23(9):1067–74.
- Hong, L., X. J. Zhang, S. Foundling, J. A. Hartsuck, and J. Tang.** 1997. "Structure of a G48H Mutant of HIV-1 Protease Explains How Glycine-48 Replacements Produce Mutants Resistant to Inhibitor Drugs." *FEBS Letters* 420(1):11–16.
- Huang, L. and C. Chen.** 2013. "Understanding HIV-1 Protease Autoprocessing for Novel Therapeutic Development." *Future Medicinal Chemistry* 5(11):1215–29.
- Ikemura, T.** 1981. "Correlation between the Abundance of Escherichia Coli Transfer RNAs and the Occurrence of the Respective Codons in Its Protein Genes: A Proposal for a Synonymous Codon Choice That Is Optimal for the E. Coli Translational System." *Journal of Molecular Biology* 151(3):389–409.
- Ikemura, T.** 1985. "Codon Usage and tRNA Content in Unicellular and Multicellular Organisms." *Molecular Biology and Evolution* 2(1):13–34.
- Kane, J. F.** 1995. "Effects of Rare Codon Clusters on High-Level Expression of Heterologous Proteins in Escherichia Coli." *Current Opinion in Biotechnology* 6(5):494–500.
- Kaplan, A. H., S. F. Michael, R. S. Wehbie, M. F. Knigge, D. A. Paul, L. Everitt, D. J. Kempf, D. W. Norbeck, J. W. Erickson, and R. Swanstrom.** 1994. "Selection of Multiple Human Immunodeficiency Virus Type 1 Variants That Encode Viral Proteases with Decreased Sensitivity to an Inhibitor of the Viral Protease." *Proceedings of the National Academy of Sciences* 91(12):5597–5601.



- Karlin, S., W. Doerfler, and L. R. Cardon.** 1994. "Why Is CpG Suppressed in the Genomes of Virtually All Small Eukaryotic Viruses but Not in Those of Large Eukaryotic Viruses?" *Journal of Virology* 68(5):2889–97.
- Karn, J. and C. M. Stoltzfus.** 2012. "Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression." *Cold Spring Harb Perspect Med* 2(2):a006916.
- Katoh, K., K. Misawa, K. Kuma, and T. Miyata.** 2002. "MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." *Nucleic Acids Research* 30(14):3059–66.
- Kempf, D. J., K. C. Marsh, G. Kumar, A. D. Rodrigues, J. F. Denissen, E. McDonald, M. J. Kukulka, A. Hsu, G. R. Granneman, P. A. Baroldi, E. Sun, D. Pizzuti, J. J. Plattner, D. W. Norbeck, and J. M. Leonard.** 1997. "Pharmacokinetic Enhancement of Inhibitors of the Human Immunodeficiency Virus Protease by Coadministration with Ritonavir." *Antimicrobial Agents and Chemotherapy* 41(3):654–60.
- Khedkar, P. H., N. Osterrieder, and D. Kunec.** 2018. "Codon Pair Bias Deoptimization of the Major Oncogene Meq of a Very Virulent Marek's Disease Virus." *Journal of General Virology* 99(12):1705–16.
- Kjems, J., M. Brown, D. D. Chang, and P. A. Sharp.** 1991. "Structural Analysis of the Interaction between the Human Immunodeficiency Virus Rev Protein and the Rev Response Element." *Proceedings of the National Academy of Sciences* 88(3):683–87.
- Kumagail, Y., O. Takeuchi, and S. Akira.** 2008. "TLR9 as a Key Receptor for the Recognition of DNA." *Advanced Drug Delivery Reviews* 60(7):795–804.
- Kunec, D. and N. Osterrieder.** 2016. "Codon Pair Bias Is a Direct Consequence of Dinucleotide Bias." *Cell Reports* 14(1):55–67.
- van der Kuyl, A. C. and B. Berkhout.** 2012. "The Biased Nucleotide Composition of the HIV Genome: A Constant Factor in a Highly Variable Virus." *Retrovirology* 9(1):92.
- Kypr, J. and J. Mrázek.** 1987. "Unusual Codon Usage of HIV." *Nature* 327(6117):20.
- Kypr, J., J. Mrázek, and J. Reich.** 1989. "Nucleotide Composition Bias and CpG Dinucleotide Content in the Genomes of HIV and HTLV." *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1009(3):280–82.
- Lauring, A. S., A. Acevedo, S. B. Cooper, and R. Andino.** 2012. "Codon Usage Determines the Mutational Robustness, Evolutionary Capacity, and Virulence of an RNA Virus." *Cell Host & Microbe* 12(5):623–32.
- Lecossier, D., F. Bouchonnet, F. Clavel, and A. J. Hance.** 2003. "Hypermutation of HIV-1 DNA in the Absence of the Vif Protein." *Science (New York, N. Y.)* 300(5622):1112.
- Li, M., E. Kao, X. Gao, H. Sandig, K. Limmer, M. Pavon-Eternod, T. E. Jones, S. Landry, T. Pan, M. D. Weitzman, and M. David.** 2012. "Codon-Usage-Based Inhibition of HIV Protein Synthesis by Human Schlafen 11." *Nature* 491(7422):125–28.
- Liang, C., L. Rong, M. Laughrea, L. Kleiman, and M. A. Wainberg.** 1998. "Compensatory Point Mutations in the Human Immunodeficiency Virus Type 1 Gag Region That Are Distal from Deletion Mutations in the Dimerization Initiation Site Can Restore Viral Replication." *Journal of Virology* 72(8):6629–36.
- Livak, K. J. and T. D. Schmittgen.** 2001. "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2<sup>-ΔΔCT</sup> Method." *Methods* 25(4):402–8.

- Lorenz, R., S. H. Bernhart, C. Höner zu Siederdisen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker. 2011. "ViennaRNA Package 2.0." *Algorithms for Molecular Biology* 6(1):26.
- Louis, J. .. M., I. T. Weber, J. Tözsér, G. M. Clore, and A. M. Gronenborn. 2000. "HIV-1 Protease: Maturation, Enzyme Specificity, and Drug Resistance." *Advances in Pharmacology (San Diego, Calif.)* 49:111–46.
- Lu, X. B., J. Heimer, D. Rekosh, and M. L. Hammarskjold. 1990. "U1 Small Nuclear RNA Plays a Direct Role in the Formation of a Rev-Regulated Human Immunodeficiency Virus Env mRNA That Remains Unspliced." *Proceedings of the National Academy of Sciences* 87(19):7598–7602.
- Lv, Z., Y. Chu, and Y. Wang. 2015. "HIV Protease Inhibitors: A Review of Molecular Selectivity and Toxicity." *HIV/AIDS (Auckland, N.Z.)* 7:95–104.
- Madhugiri, R., N. Karl, D. Petersen, K. Lamkiewicz, M. Fricke, U. Wend, R. Scheuer, M. Marz, and J. Ziebuhr. 2017. "Structural and Functional Conservation of Cis-Acting RNA Elements in Coronavirus 5'-Terminal Genome Regions." *Virology* 517:44–55.
- Magoc, T. and S. L. Salzberg. 2011. "FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies." *Bioinformatics* 27(21):2957–63.
- Mahalingam, B., J. M. Louis, C. C. Reed, J. M. Adomat, J. Krouse, Y. F. Wang, R. W. Harrison, and I. T. Weber. 1999. "Structural and Kinetic Analysis of Drug Resistant Mutants of HIV-1 Protease." *European Journal of Biochemistry* 263(1):238–44.
- Malim, M. H. and B. R. Cullen. 1991. "HIV-1 Structural Gene Expression Requires the Binding of Multiple Rev Monomers to the Viral RRE: Implications for HIV-1 Latency." *Cell* 65(2):241–48.
- Malim, M. H., J. Hauber, S. -y. Le, J. V. Maizel, and B. R. Cullen. 1989. "The HIV-1 Rev Trans-Activator Acts through a Structured Target Sequence to Activate Nuclear Export of Unspliced Viral RNA." *Nature* 338(6212):254–57.
- Malim, M. H., L. S. Tiley, D. F. McCarn, J. R. Rusche, J. Hauber, and B. R. Cullen. 1990. "HIV-1 Structural Gene Expression Requires Binding of the Rev Trans-Activator to Its RNA Target Sequence." *Cell* 60(4):675–83.
- Mammano, F., V. Trouplin, V. Zennou, and F. Clavel. 2000. "Retracing the Evolutionary Pathways of Human Immunodeficiency Virus Type 1 Resistance to Protease Inhibitors: Virus Fitness in the Absence and in the Presence of Drug." *Journal of Virology* 74(18):8524–31.
- Mann, D. A., I. Mikaélian, R. W. Zemmel, S. M. Green, A. D. Lowe, T. Kimura, M. Singh, P. Jonathan, G. Butler, M. J. Gait, and J. Karn. 1994. "A Molecular Rheostat. Co-Operative Rev Binding to Stem I of the Rev-Response Element Modulates Human Immunodeficiency Virus Type-1 Late Gene Expression." *Journal of Molecular Biology* 241(2):193–207.
- Mao, Y., L. Wang, C. Gu, A. Herschhorn, S. H. Xiang, H. Haim, X. Yang, and J. Sodroski. 2012. "Subunit Organization of the Membrane-Bound HIV-1 Envelope Glycoprotein Trimer." *Nature Structural & Molecular Biology* 19(9):893–99.
- Martínez, M. A., A. Jordan-Paiz, S. Franco, and M. Nevot. 2016. "Synonymous Virus Genome Recoding as a Tool to Impact Viral Fitness." *Trends in Microbiology* 24(2):134–47.

- Martrus, G., M. Nevot, C. Andres, B. Clotet, and M. A. Martinez.** 2013. "Changes in Codon-Pair Bias of Human Immunodeficiency Virus Type 1 Have Profound Effects on Virus Replication in Cell Culture." *Retrovirology* 10(1):78.
- Más, A., C. López-Galindo, I. Cacho, J. Gómez, and M. A. Martínez.** 2010. "Unfinished Stories on Viral Quasispecies and Darwinian Views of Evolution." *Journal of Molecular Biology* 397(4):865–77.
- Mauclère, P., I. Lousert-Ajaka, F. Damond, P. Fagot, S. Souquières, M. M. Lobe, F. X. M. Keou, F. Barré-Sinoussi, S. Saragosti, F. Brun-Vézinet, and F. Simon.** 1997. "Serological and Virological Characterization of HIV-1 Group O Infection in Cameroon." *AIDS* 11(4):445–53.
- McCaskill, J. S.** 1990. "The Equilibrium Partition Function and Base Pair Binding Probabilities for RNA Secondary Structure." *Biopolymers* 29(6–7):1105–19.
- Molla, A., M. Korneyeva, Q. Gao, S. Vasavanonda, P. J. Schipper, H. M. Mo, M. Markowitz, T. Chernyavskiy, P. Niu, N. Lyons, A. Hsu, G. R. Granneman, D. D. Ho, C. A. Boucher, J. M. Leonard, D. W. Norbeck, and D. J. Kempf.** 1996. "Ordered Accumulation of Mutations in HIV Protease Confers Resistance to Ritonavir." *Nature Medicine* 2(7):760–66.
- Moratorio, G., R. Henningsson, C. Barbezange, L. Carrau, A. V. Bordería, H. Blanc, S. Beaucourt, E. Z. Poirier, T. Vallet, J. Boussier, B. C. Mounce, M. Fontes, and M. Vignuzzi.** 2017. "Attenuation of RNA Viruses by Redirecting Their Evolution in Sequence Space." *Nature Microbiology* 2(8):17088.
- Mueller, S., J. R. Coleman, D. Papamichail, C. B. Ward, A. Nimnual, B. Futcher, S. Skiena, and E. Wimmer.** 2010. "Live Attenuated Influenza Virus Vaccines by Computer-Aided Rational Design." *Nature Biotechnology* 28(7):723–26.
- Mueller, S., D. Papamichail, J. R. Coleman, S. Skiena, and E. Wimmer.** 2006. "Reduction of the Rate of Poliovirus Protein Synthesis through Large-Scale Codon Deoptimization Causes Attenuation of Viral Virulence by Lowering Specific Infectivity." *Journal of Virology* 80(19):9687–96.
- Nevot, M., A. Jordan-Paiz, G. Martrus, C. Andrés, D. García-Cehic, J. Gregori, S. Franco, J. Quer, and M. A. Martinez.** 2018. "HIV-1 Protease Evolvability Is Affected by Synonymous Nucleotide Recoding" edited by G. Silvestri. *Journal of Virology* 92(16):e00777-18.
- Newman, Z. R., J. M. Young, N. T. Ingolia, and G. M. Barton.** 2016. "Differences in Codon Bias and GC Content Contribute to the Balanced Expression of TLR7 and TLR9." *Proceedings of the National Academy of Sciences* 113(10):E1362–71.
- Ni, Y. Y., Z. Zhao, T. Opriessnig, S. Subramaniam, L. Zhou, D. Cao, Q. Cao, H. Yang, and X. J. Meng.** 2014. "Computer-Aided Codon-Pairs Deoptimization of the Major Envelope GP5 Gene Attenuates Porcine Reproductive and Respiratory Syndrome Virus." *Virology* 450–451:132–39.
- Nijhuis, M., R. Schuurman, D. de Jong, J. Erickson, E. Gustchina, J. Albert, P. Schipper, S. Gulnik, and C. A. B. Boucher.** 1999. "Increased Fitness of Drug Resistant HIV-1 Protease as a Result of Acquisition of Compensatory Mutations during Suboptimal Therapy." *AIDS* 13(17):2349–59.

- Le Nouen, C., L. G. Brock, C. Luongo, T. McCarty, L. Yang, M. Mehedi, E. Wimmer, S. Mueller, P. L. Collins, U. J. Buchholz, and J. M. DiNapoli.** 2014. "Attenuation of Human Respiratory Syncytial Virus by Genome-Scale Codon-Pair Deoptimization." *Proceedings of the National Academy of Sciences* 111(36):13169–74.
- Pancera, M., T. Zhou, A. Druz, I. S. Georgiev, C. Soto, J. Gorman, J. Huang, P. Acharya, G. Y. Chuang, G. Ofek, G. B. E. Stewart-Jones, J. Stuckey, R. T. Bailer, M. G. Joyce, M. K. Louder, N. Tumba, Y. Yang, B. Zhang, M. S. Cohen, B. F. Haynes, J. R. Mascola, L. Morris, J. B. Munro, S. C. Blanchard, W. Mothes, M. Connors, and P. D. Kwong.** 2014. "Structure and Immune Recognition of Trimeric Pre-Fusion HIV-1 Env." *Nature* 514(7523):455–61.
- Pandey, S., T. Kawai, and S. Akira.** 2015. "Microbial Sensing by Toll-Like Receptors and Intracellular Nucleic Acid Sensors." *Cold Spring Harbor Perspectives in Biology* 7(1):a016246.
- Pannecouque, C., D. Daelemans, and E. De Clercq.** 2008. "Tetrazolium-Based Colorimetric Assay for the Detection of HIV Replication Inhibitors: Revisited 20 Years Later." *Nature Protocols* 3(3):427–34.
- Paradis, E., J. Claude, and K. Strimmer.** 2004. "APE: Analyses of Phylogenetics and Evolution in R Language." *Bioinformatics* 20(2):289–90.
- Parera, M. and M. A. Martinez.** 2014. "Strong Epistatic Interactions within a Single Protein." *Molecular Biology and Evolution* 31(6):1546–53.
- Pauwels, R., J. Balzarini, M. Baba, R. Snoeck, D. Schols, P. Herdewijn, J. Desmyter, and E. De Clercq.** 1988. "Rapid and Automated Tetrazolium-Based Colorimetric Assay for the Detection of Anti-HIV Compounds." *Journal of Virological Methods* 20(4):309–21.
- Pear, W. S., G. P. Nolan, M. L. Scott, and D. Baltimore.** 1993. "Production of High-Titer Helper-Free Retroviruses by Transient Transfection." *Proceedings of the National Academy of Sciences* 90(18):8392–96.
- Pettit, S. C., L. E. Everitt, S. Choudhury, B. M. Dunn, and A. H. Kaplan.** 2004. "Initial Cleavage of the Human Immunodeficiency Virus Type 1 GagPol Precursor by Its Activated Protease Occurs by an Intramolecular Mechanism." *Journal of Virology* 78(16):8477–85.
- Plantier, J. C., M. Leoz, J. E. Dickerson, F. De Oliveira, F. Cordonnier, V. Lemée, F. Damond, D. L. Robertson, and F. Simon.** 2009. "A New Human Immunodeficiency Virus Derived from Gorillas." *Nature Medicine* 15(8):871–72.
- Platt, E. J., M. Bilska, S. L. Kozak, D. Kabat, and D. C. Montefiori.** 2009. "Evidence That Ecotropic Murine Leukemia Virus Contamination in TZM-BI Cells Does Not Affect the Outcome of Neutralizing Antibody Assays with Human Immunodeficiency Virus Type 1." *Journal of Virology* 83(16):8289–92.
- Platt, E. J., K. Wehrly, S. E. Kuhmann, B. Chesebro, and D. Kabat.** 1998. "Effects of CCR5 and CD4 Cell Surface Concentrations on Infections by Macrophagetropic Isolates of Human Immunodeficiency Virus Type 1." *Journal of Virology* 72(4):2855–64.
- Prejdová, J., M. Soucek, and J. Konvalinka.** 2004. "Determining and Overcoming Resistance to HIV Protease Inhibitors." *Current Drug Targets. Infectious Disorders* 4(2):137–52.

- Presnyak, V., N. Alhusaini, Y. H. Chen, S. Martin, N. Morris, N. Kline, S. Olson, D. Weinberg, K. E. Baker, B. R. Graveley, and J. Collier. 2015. "Codon Optimality Is a Major Determinant of mRNA Stability." *Cell* 160(6):1111–24.
- Ramakrishnan, R., R. Mehta, V. Sundaravaradan, T. Davis, and N. Ahmad. 2006. "Characterization of HIV-1 Envelope Gp41 Genetic Diversity and Functional Domains Following Perinatal Transmission." *Retrovirology* 3:42.
- Rambaut, A., D. Posada, K. A. Crandall, and E. C. Holmes. 2004. "The Causes and Consequences of HIV Evolution." *Nature Reviews Genetics* 5(1):52–61.
- dos Reis, M., R. Savva, and L. Wernisch. 2004. "Solving the Riddle of Codon Usage Preferences: A Test for Translational Selection." *Nucleic Acids Research* 32(17):5036–44.
- Rima, B. K. and N. V. McFerran. 1997. "Dinucleotide and Stop Codon Frequencies in Single-Stranded RNA Viruses Detected Significant Suppression of UpA , Correlating." *Journal of General Virology* 78(Pt 11):2859–70.
- Rima, B. K. and N. V. McFerran. 1997. "Dinucleotide and Stop Codon Frequencies in Single-Stranded RNA Viruses." *Journal of General Virology* 78(11):2859–70.
- Sarzotti-Kelsoe, M., R. T. Bailer, E. Turk, C. -I. Lin, M. Bilaska, K. M. Greene, H. Gao, C. A. Todd, D. A. Ozaki, M. S. Seaman, J. R. Mascola, and D. C. Montefiori. 2014. "Optimization and Validation of the TZM-BI Assay for Standardized Assessments of Neutralizing Antibodies against HIV-1." *Journal of Immunological Methods* 409:131–46.
- Shabalina, S. A., A. Y. Ogurtsov, and N. A. Spiridonov. 2006. "A Periodic Pattern of mRNA Secondary Structure Created by the Genetic Code." *Nucleic Acids Research* 34(8):2428–37.
- Shafer, R. W. and J. M. Schapiro. 2008. "HIV-1 Drug Resistance Mutations: An Updated Framework for the Second Decade of HAART." *AIDS Reviews* 10(2):67–84.
- Sherpa, C., J. W. Rausch, S. F. J. Le Grice, M. L. Hammarskjold, and D. Rekosh. 2015. "The HIV-1 Rev Response Element (RRE) Adopts Alternative Conformations That Promote Different Rates of Virus Replication." *Nucleic Acids Research* 43(9):4676–86.
- Shin, Y. C., G. F. Bischof, W. A. Lauer, and R. C. Desrosiers. 2015. "Importance of Codon Usage for the Temporal Regulation of Viral Gene Expression." *Proceedings of the National Academy of Sciences* 112(45):14030–35.
- Shpaer, E. G. and J. I. Mullins. 1990. "Selection against CpG Dinucleotides in Lentiviral Genes: A Possible Role of Methylation in Regulation of Viral Expression." *Nucleic Acids Research* 18(19):5793–97.
- Simmonds, P. 2012. "SSE: A Nucleotide and Amino Acid Sequence Analysis Platform." *BMC Research Notes* 5(1):50.
- Simmonds, P., F. Tulloch, D. J. Evans, and M. D. Ryan. 2015. "Attenuation of Dengue (and Other RNA Viruses) with Codon Pair Recoding Can Be Explained by Increased CpG/UpA Dinucleotide Frequencies." *Proceedings of the National Academy of Sciences* 112(28):E3633–34.
- Simmonds, P., W. Xia, J. Baillie, and K. McKinnon. 2013. "Modelling Mutational and Selection Pressures on Dinucleotides in Eukaryotic Phyla –Selection against CpG and UpA in Cytoplasmically Expressed RNA and in RNA Viruses." *BMC Genomics* 14(1):610.

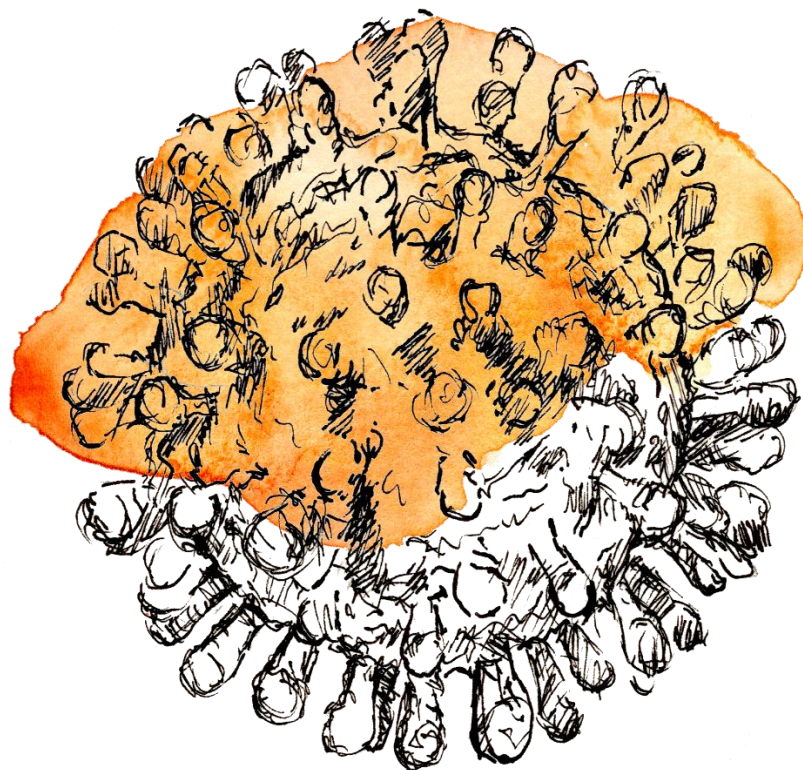
- Simon, F., P. Maucière, P. Roques, I. Loussert-Ajaka, M. C. Müller-Trutwin, S. Saragosti, M. C. Georges-Courbot, F. Barré-Sinoussi, and F. Brun-Vézinet.** 1998. "Identification of a New Human Immunodeficiency Virus Type 1 Distinct from Group M and Group O." *Nature Medicine* 4(9):1032–37.
- Simon, L. M., E. Morandi, A. Luganini, G. Gribaudo, L. Martinez-Sobrido, D. H. Turner, S. Oliviero, and D. Incarnato.** 2019. "In Vivo Analysis of Influenza A MRNA Secondary Structures Identifies Critical Regulatory Motifs." *Nucleic Acids Research* 47(13):7003–17.
- Song, Y., Y. Liu, C. B. Ward, S. Mueller, B. Futcher, S. Skiena, A. V. Paul, and E. Wimmer.** 2012. "Identification of Two Functionally Redundant RNA Elements in the Coding Sequence of Poliovirus Using Computer-Generated Design." *Proceedings of the National Academy of Sciences* 109(36):14301–7.
- Sørensen, M. A., C. G. Kurland, and S. Pedersen.** 1989. "Codon Usage Determines Translation Rate in Escherichia Coli." *Journal of Molecular Biology* 207(2):365–77.
- Steckbeck, J. D., J. K. Craigo, C. O. Barnes, and R. C. Montelaro.** 2011. "Highly Conserved Structural Properties of the C-Terminal Tail of HIV-1 Gp41 Protein Despite Substantial Sequence Variation among Diverse Clades." *Journal of Biological Chemistry* 286(31):27156–66.
- Stelzer, A. C., A. T. Frank, J. D. Kratz, M. D. Swanson, M. J. Gonzalez-Hernandez, J. Lee, I. Andricioaei, D. M. Markovitz, and H. M. Al-Hashimi.** 2011. "Discovery of Selective Bioactive Small Molecules by Targeting an RNA Dynamic Ensemble." *Nature Chemical Biology* 7(8):553–59.
- Suspène, R., P. Sommer, M. Henry, S. Ferris, D. Guétard, S. Pochet, A. Chester, N. Navaratnam, S. Wain-Hobson, and J. P. Vartanian.** 2004. "APOBEC3G Is a Single-Stranded DNA Cytidine Deaminase and Functions Independently of HIV Reverse Transcriptase." *Nucleic Acids Research* 32(8):2421–29.
- Takata, M. A., D. Gonçalves-Carneiro, T. M. Zang, S. J. Soll, A. York, D. Blanco-Melo, and P. D. Bieniasz.** 2017. "CG Dinucleotide Suppression Enables Antiviral Defence Targeting Non-Self RNA." *Nature* 550(7674):124–27.
- Takata, M. A., S. J. Soll, A. Emery, D. Blanco-Melo, R. Swanstrom, and P. D. Bieniasz.** 2018. "Global Synonymous Mutagenesis Identifies Cis-Acting RNA Elements That Regulate HIV-1 Splicing and Replication." *PLOS Pathogens* 14(1):e1006824.
- Tamura, K., G. Stecher, D. Peterson, A. Filipinski, and S. Kumar.** 2013. "MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0." *Molecular Biology and Evolution* 30(12):2725–29.
- Theys, K., A. F. Feder, M. Gelbart, M. Hartl, A. Stern, and P. S. Pennings.** 2018. "Within-Patient Mutation Frequencies Reveal Fitness Costs of CpG Dinucleotides and Drastic Amino Acid Changes in HIV" edited by J. D. Bloom. *PLOS Genetics* 14(6):e1007420.
- Tulloch, F., N. J. Atkinson, D. J. Evans, M. D. Ryan, and P. Simmonds.** 2014. "RNA Virus Attenuation by Codon Pair Deoptimisation Is an Artefact of Increases in CpG/UpA Dinucleotide Frequencies." *ELife* 3:e04531.
- UNAIDS.** 2018. *2018 Global HIV Statistics*.
- Vabret, N., M. Bailly-Bechet, A. Lepelletier, V. Najburg, O. Schwartz, B. Verrier, and F. Tangy.** 2014. "Large-Scale Nucleotide Optimization of Simian Immunodeficiency Virus Reduces Its Capacity To Stimulate Type I Interferon In Vitro." *Journal of Virology* 88(8):4161–72.

- Wasson, M. K., J. Borkakoti, A. Kumar, B. Biswas, and P. Vivekanandan. 2017. "The CpG Dinucleotide Content of the HIV-1 Envelope Gene May Predict Disease Progression." *Scientific Reports* 7(1):8162.
- Watts, J. M., K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess Jr, R. Swanstrom, C. L. Burch, and K. M. Weeks. 2009. "Architecture and Secondary Structure of an Entire HIV-1 RNA Genome." *Nature* 460(7256):711–16.
- Weber, I. T. 1989. "Structural Alignment of Retroviral Protease Sequences." *Gene* 85(2):565–66.
- Wensing, A. M., V. Calvez, F. Ceccherini-Silberstein, C. Charpentier, H. F. Günthard, R. Paredes, R. W. Shafer, and D. D. Richman. 2019. "2019 Update of the Drug Resistance Mutations in HIV-1." *Topics in Antiviral Medicine* 27(3):i–xi.
- Wensing, A. M. J., N. M. van Maarseveen, and M. Nijhuis. 2010. "Fifteen Years of HIV Protease Inhibitors: Raising the Barrier to Resistance." *Antiviral Research* 85(1):59–74.
- Wilkinson, K. A., E. J. Merino, and K. M. Weeks. 2006. "Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE): Quantitative RNA Structure Analysis at Single Nucleotide Resolution." *Nature Protocols* 1(3):1610–16.
- Wimmer, E., S. Mueller, T. M. Tumpey, and J. K. Taubenberger. 2009. "Synthetic Viruses: A New Opportunity to Understand and Prevent Viral Disease." *Nature Biotechnology* 27(12):1163–72.
- Wolinsky, S. M., K. J. Kunstman, J. T. Safrit, R. A. Koup, A. U. Neumann, and B. T. M. Korber. 1996. "Response: HIV-1 Evolution and Disease Progression." *Science* 274(5289):1010–11.
- World Health Organization, WHO. 2017. *HIV Drug Resistance Report 2017*.
- Yamaguchi-kabata, Y. and T. Gojobori. 2000. "Reevaluation of Amino Acid Variability of the Human Immunodeficiency Virus Type 1 Gp120 Envelope Glycoprotein." *Journal of Virology* 74(9):4335–50.
- Yang, C., S. Skiena, B. Futcher, S. Mueller, and E. Wimmer. 2013. "Deliberate Reduction of Hemagglutinin and Neuraminidase Expression of Influenza Virus Leads to an Ultraproductive Live Vaccine in Mice." *Proceedings of the National Academy of Sciences* 110(23):9481–86.
- Zapp, M. L., T. J. Hope, T. G. Parslow, and M. R. Green. 1991. "Oligomerization and RNA Binding Domains of the Type 1 Human Immunodeficiency Virus Rev Protein: A Dual Function for an Arginine-Rich Binding Motif." *PNAS* 88(17):7734–38.
- Zhang, H., Y. Zhou, C. Alcock, T. Kiefer, D. Monie, J. Siliciano, Q. Li, P. Pham, J. Cofrancesco, D. Persaud, and R. F. Siliciano. 2004. "Novel Single-Cell-Level Phenotypic Assay for Residual Drug Susceptibility and Reduced Replication Capacity of Drug-Resistant Human Immunodeficiency Virus Type 1." *Journal of Virology* 78(4):1718–29.
- Zhou, Z., Y. Dang, M. Zhou, L. Li, C. H. Yu, J. Fu, S. Chen, and Y. Liu. 2016. "Codon Usage Is an Important Determinant of Gene Expression Levels Largely through Its Effects on Transcription." *Proceedings of the National Academy of Sciences* 113(41):E6117–25.

**Zuker, M. and P. Stiegler.** 1981. "Optimal Computer Folding of Large RNA Sequences Using Thermodynamics and Auxiliary Information." *Nucleic Acids Research* 9(1):133–48.







## **Publications**

---



Martinez MA, Nevot M, Jordan-Paiz A, Franco S. (2015) Similarities between Human Immunodeficiency Virus Type 1 and Hepatitis C Virus Genetic and Phenotypic Protease Quasispecies Diversity. *Journal of Virology*, 89(19):9758-64.

Martínez MA, Jordan-Paiz A, Franco S, Nevot M. (2016) Synonymous virus genome recoding as a tool to impact viral fitness. *Trends in Microbiology*, 24(2):134-147.

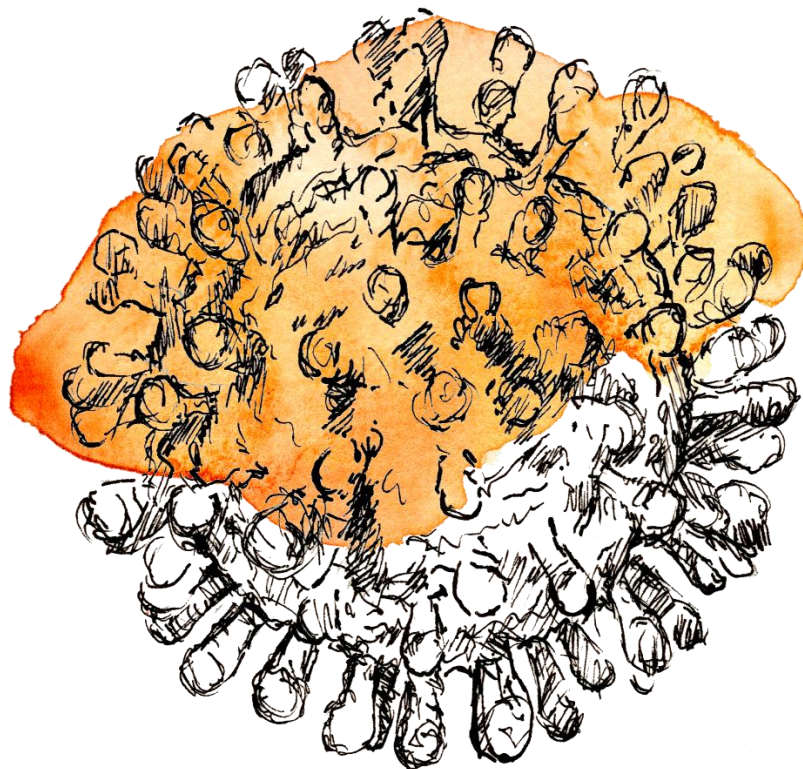
Franco S, Buccione D, Pluvinet R, Mothe B, Ruiz L, Nevot M, Jordan-Paiz A, Ramos L, Aussó S, Morillas RM, Sumoy L, Martinez MA, Tural C. (2018) Large-scale screening of circulating microRNAs in individuals with HIV-1 mono-infections reveals specific liver damage signatures. *Antiviral Research*, 155:106-114

Nevot M, Jordan-Paiz A, Martrus G, Andrés C, García-Cehic D, Gregori J, Franco S, Quer J, Martinez MA. (2018) HIV-1 protease evolvability is affected by synonymous nucleotide recoding. *Journal of Virology*, 92(16):e00777-18.

Martinez MA, Jordan-Paiz A, Franco S, Nevot M. (2019) Synonymous genome recoding: a tool to explore microbial biology and new therapeutic strategies. *Nucleic Acids Research*. In press

Jordan-Paiz A, Nevot M, Lamkiewicz K, Lataretu M, Franco S, Marz M, Martinez MA. (2019) Loss of protein expression and HIV-1 lethality induced by synonymous substitutions in the 3' end of the virus envelope gene. Manuscript in preparation.





**Annex**

---



```

WT_env   : ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCTCTGGGATGTTGATGATCTGTAGTGTCTACAGAAAATTTGGGGTACAGCTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCACCCTCTATTTTGTGCATCAGATGCTAAAACATATGATACAGAGGTACATAA
Recoded_env : .....A.....G.....G.....A.....

WT_env   : TGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTACAGAAAAATTTAACATGTGGAATAATGACATGGTAAAGCAACATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCATAAAGCCATGTGTAATAATTAACCCACTCTGTGTAGTTTAAAGTGCACACTGATT
Recoded_env : .....A.....T.A.....G.....

WT_env   : TGAAGAATGATACTAATACCAATAGTAGTAGCGGAGAAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTCAATATCAGCACAAAGCATAAAGGTAAAGTGCAGAAAAGAAATATGCATTTTTTATAAAGTATATAACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTC
Recoded_env : .....G.....A.....A.....A.....G.....

WT_env   : ATTACACAGGCTGTCCAAGGTATCCTTTGAGCCAATCCCATACATTATTGTGCCCGGCTGGTTTTGGCATTTCAAATGTAAATAAAGACCTTCAATGGAACAGGACCATGTACAAATGTGAGCAGTACAATGTACACATGGAATTAGGCCAGTAGTATCACTCAACTGCTGTTAAATGGCAGCTAGCAGA
Recoded_env : .....A.....C.T.....G.....

WT_env   : AGAAGAGGTAGTAATTAGATCTCTCAATTCACGGCAATGCTAAAACATAATAGTACAGCTGAACACATCTGTAGAAATTAATTGTACAAGACCCAAACAATAACAAGAAAAGAAATCCGTATCCAGAGGAGACCAGGGAGACATTTGTTACAATAGGAAAAATAGGAAATATGAGACAGCACATTGTAACATTAA
Recoded_env : .....A.....A.....

WT_env   : GTAGAGCAAAATGGAATAACACTTTAAACAGATAGCTAGCAAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATCTTTAAGCAATCCCTCAGGAGGGGACCCAGAAAATGTAAACGCACAGTTTAAATTTGGAGGGGAATTTTTCTACTGTAATCAACACAAGTGTAAATAGTACTTGGTTAATAGTACTTGG
Recoded_env : .....G.....A.....A.....G.....G.....

WT_env   : AGTACTGAAGGTCAAATAACACTGAAGGAAGTACACAAATCACCTCCCATGCAGATAAAACAATATAAACAATGTGGCAGAAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGTGGACAAATAGATGTTCAATCAAAATATACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGAT
Recoded_env : .....G.....A.....G.....T.A.....G.....A.....A.....A.....

WT_env   : CTTACAGACTGGAGGAGAGATATGAGGACAAATGGAGAAGTGAATATATAAATAAAGTAGTAAAAATGAACCATTAGGAGTAGCACCACCAGGCAAGAGAGAGTGGTGCAGAGAGAAAAGAGCAGTGGGAATAGGAGCTTTGTTCTCTGGGTTCCTGGAGCAGCAGGAGCAGTATGGGCGAGCCT
Recoded_env : .....G.....C.T.....

WT_env   : CAATGACGCTGACGGTACAGGCCAGACAATTTATGCTGGTATAGTGCAGCAGCAGAAATTTGCTGAGGGCTATTGAGGGCCACAGCATCTGTTGCAACTCAGCAGTCTGGGGCACTCAAGCAGCTCCAGGCAAGAAATCCTGGCTGTGGAAGATACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGA
Recoded_env : .....

WT_env   : AAACCTATTGCAACCCTGCTGTGCCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTGGAAATCACAGCAGCTGGATGGAGTGGGACAGAGAAATTAACAATTCACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACAGCAAGAAAAGAAATGAACAAGAAATTTATGGAATATAGATAAATGGGC
Recoded_env : .....A.....

WT_env   : AAGTTTGTGGAATTTGTTAACAATAACAATTTGGCTGTGGTATATAAAATTTATCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGATAGTTTTTGTGCTACTTTCTATAGTGAATAGAGTTAGGCAGGATATTCACATTATCGTTTCAGACCCACCTCCCAACCCGAGGGGACCCGACAGGCCCAAGGAATAG
Recoded_env : .....C.T.....

WT_env   : AAGAAGAGTGGAGAGAGAGACAGACAGATCCATTTCGATTAGTGAACGGATCCCTAGCACTTATCTGGGACGATCTGCGGAGCCTGTGCTCTCAGTACCACCCCTGAGAGACTTACTCTTGAATGTAACGAGGATTTGGAACCTCTGGACCTGAGGAGGCTGGGAGCCCTCAAAATATTGTTGGAATCTCTCA
Recoded_env : .....

WT_env   : CAGTATTGGAGTCAAGAACTAAGAATAGTCTGTTAACTGCTCAATGCCACAGCATTAGCAGTAGCTGAGGGGACAGATAGGGTTATAGAAGTATTACAGCAGCTTATAGAGCTATTGCCACATACCTAGAAGAATAAGACAGGGCTTGGAAAGGATTTTGTCTA
Recoded_env : .....

```

**Annex Figure 1. Nucleotide sequence alignment of WT\_env and Recoded-env envs.** Regions maintained WT are shown in blue. Substitutions of Recoded\_env are displayed under WT sequence.



```

                *      20      *      40      *      60      *      80      *      100     *      120     *
WT_env          : ATGAGAGTGAAGGAGAAATATCAGCAC TTGTGGAGATGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTACACAGTCTATATGGGGTACCTG
Recoded_env     : .....A.....G.
Recoded_env_mm  : .....G.
Recoded_env_sCpG1 : .....A.
Recoded_env_sCpG2 : .....A...A.
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....A...G.
Recoded_env_3'Awt : .....A...G.
Recoded_env_3'Bwt : .....A...G.

```

```

                140      *      160      *      180      *      200      *      220      *      240      *      260
WT_env          : TGTGGAAGGAAGCAACCACCCTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGT
Recoded_env     : .....G.....A.....
Recoded_env_mm  : .....G.....
Recoded_env_sCpG1 : .....A.....
Recoded_env_sCpG2 : .....A.....A.....
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....G.....A.....
Recoded_env_3'Awt : .....G.....A.....
Recoded_env_3'Bwt : .....G.....A.....

```

```

                *      280      *      300      *      320      *      340      *      360      *      380      *
WT_env          : AAATGTGACAGAAAATTTTAACATGTGGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCACTCTGTGTAGTTTAAAG
Recoded_env     : .....A.....T.A.....
Recoded_env_mm  : .....A.....T.A.....
Recoded_env_sCpG1 : .....A.....T.A.....
Recoded_env_sCpG2 : .....A.....T.A.....
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....A.....T.A.....
Recoded_env_3'Awt : .....A.....T.A.....
Recoded_env_3'Bwt : .....A.....T.A.....

```

```

                400      *      420      *      440      *      460      *      480      *      500      *      520
WT_env          : TGCCTGATTTGAAGAAATGATACTAATACCAATAGTAGTAGCGGGAGAAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTCAATATCAGCACAAGCATAAGAGGTAAGGTGCAGAAAGAATATG
Recoded_env     : ....G.....G.....A.....A.....A.....
Recoded_env_mm  : .....A.....A.....A.....
Recoded_env_sCpG1 : .....A.....A.....A.....
Recoded_env_sCpG2 : ....A.....A.....A.....A.....A.....
Recoded_env_5'wt : .....
Recoded_env_3'wt : ....G.....G.....A.....A.....A.....
Recoded_env_3'Awt : ....G.....G.....A.....A.....A.....
Recoded_env_3'Bwt : ....G.....G.....A.....A.....A.....

```

```

*      540      *      560      *      580      *      600      *      620      *      640      *
WT_env      : CATT TTTT TATAAA CTTGATATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTA
Recoded_env : .....G.....A.....
Recoded_env_mm : .....G.....A.....
Recoded_env_sCpG1 : .....A.....
Recoded_env_sCpG2 : .....A.....
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....G.....A.....
Recoded_env_3'Awt : .....G.....A.....
Recoded_env_3'Bwt : .....G.....A.....

```

```

        660      *      680      *      700      *      720      *      740      *      760      *      780
WT_env      : TTGTGCCCGGCTGGTTTTGCGATTCTAAAATGTAATAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTG
Recoded_env : .....C.T.....G.....
Recoded_env_mm : .....G.....
Recoded_env_sCpG1 : .....
Recoded_env_sCpG2 : .....A.....A.....
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....C.T.....G.....
Recoded_env_3'Awt : .....C.T.....G.....
Recoded_env_3'Bwt : .....C.T.....G.....

```

```

*      800      *      820      *      840      *      860      *      880      *      900      *
WT_env      : TTAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGTCAATTTACGGACAATGCTAAAACCATAATAGTACAGCTGAACACATCTGTAGAAATTAATGTACAAGACCCAACAACAATACAA
Recoded_env : .....A.....
Recoded_env_mm : .....
Recoded_env_sCpG1 : .....A.....
Recoded_env_sCpG2 : .....A.....
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....A.....
Recoded_env_3'Awt : .....A.....
Recoded_env_3'Bwt : .....A.....

```

```

        920      *      940      *      960      *      980      *      1000      *      1020      *      1040
WT_env      : GAAAAAGAATCCGTATCCAGAGAGGACCAGGGAGAGCATTTGTTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGATAGCTAG
Recoded_env : .....A.....G.....
Recoded_env_mm : .....G.....
Recoded_env_sCpG1 : .....A.....
Recoded_env_sCpG2 : .....A.....A.....
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....A.....G.....
Recoded_env_3'Awt : .....A.....G.....
Recoded_env_3'Bwt : .....A.....G.....

```



```
WT_env      : GGGTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAAATTTGCTGAGGGCTATTGAGGCGCAACAGC
Recoded_env : .....
Recoded_env_mm : .....
Recoded_env_sCpG1 : .....
Recoded_env_sCpG2 : .....
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....
Recoded_env_3'Awt : .....
Recoded_env_3'Bwt : .....
```

```
WT_env      : ATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAGATACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAACCTCATTGACCACTGC
Recoded_env : .....
Recoded_env_mm : .....
Recoded_env_sCpG1 : .....
Recoded_env_sCpG2 : .....
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....
Recoded_env_3'Awt : .....
Recoded_env_3'Bwt : .....
```

```
WT_env      : TGTGCCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAGAATCGCAA
Recoded_env : .....A.....
Recoded_env_mm : .....A.....
Recoded_env_sCpG1 : .....A.....
Recoded_env_sCpG2 : .....A.....
Recoded_env_5'wt : .....A.....
Recoded_env_3'wt : .....
Recoded_env_3'Awt : .....
Recoded_env_3'Bwt : .....A.....
```

```
WT_env      : AACCAGCAAGAAAAAATGAACAAGAATTATTGGAATAGATAAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTTGGCTGTGGTATATAAAATTTATTCATAATGATAGTAGGAGGCTTGGTAG
Recoded_env : .....
Recoded_env_mm : .....
Recoded_env_sCpG1 : .....
Recoded_env_sCpG2 : .....
Recoded_env_5'wt : .....
Recoded_env_3'wt : .....
Recoded_env_3'Awt : .....
Recoded_env_3'Bwt : .....
```

```

*
*           *           2100           *           2120           *           2140           *           2160           *           2180           *           2200
WT_env      : GTTTAAGAATAGTTTTTGGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTACCATTTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGG
Recoded env : .....C.T.....
Recoded env mm : .....C.T.....
Recoded env sCpG1 : .....
Recoded env sCpG2 : .....A.....
Recoded env 5'wt : .....C.T.....
Recoded env 3'wt : .....
Recoded env 3'Awt : .....C.T.....
Recoded env 3'Bwt : .....

*           *           2220           *           2240           *           2260           *           2280           *           2300           *           2320           *           2340
WT_env      : TGGGAGAGAGACAGAGACAGATCCATTTCGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGG
Recoded env : .....
Recoded env mm : .....
Recoded env sCpG1 : .....
Recoded env sCpG2 : .....
Recoded env 5'wt : .....
Recoded env 3'wt : .....
Recoded env 3'Awt : .....
Recoded env 3'Bwt : .....

*           *           2360           *           2380           *           2400           *           2420           *           2440           *           2460           *
WT_env      : ATTTGTGGAACCTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATTGGTGAATCTCCTACAGTATTGGAGTCAGGAACAAAGAATAGTGCTGTTAACTTGCTCAATGCCACAGCCATAGCAGTAGCTG
Recoded env : .....
Recoded env mm : .....
Recoded env sCpG1 : .....
Recoded env sCpG2 : .....
Recoded env 5'wt : .....
Recoded env 3'wt : .....
Recoded env 3'Awt : .....
Recoded env 3'Bwt : .....

*           *           2480           *           2500           *           2520           *           2540           *           2560
WT_env      : AGGGGACAGATAGGGTTATAGAAGTATTACAAGCAGCTTATAGAGCTATTGCCACATACCTAGAAGAATAAGACAGGGCTTGGAAAGGATTTTGCTA
Recoded env : .....
Recoded env mm : .....
Recoded env sCpG1 : .....
Recoded env sCpG2 : .....
Recoded env 5'wt : .....
Recoded env 3'wt : .....
Recoded env 3'Awt : .....
Recoded env 3'Bwt : .....

```

**Annex Figure 2. Nucleotide alignment of WT\_env and all recoded env based on Recoded-env. Regions maintained WT are colored in blue. Substitutions of each mutant are shown within each corresponding sequence.**

```

*      20      *      40      *      60      *      80      *      100     *      120     *      140     *      160     *      180     *      200
WT_env      : MRVKEKYQHLWRGWRGMLLGLMLICSATEKLWVTVYVGVVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFNISTSRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSV
Recoded_env : .....
Recoded_env mm : .....
Recoded_env sCpG1 : .....
Recoded_env sCpG2 : .....
Recoded_env 5'wt : .....
Recoded_env 3'wt : .....
Recoded_env 3'Awt : .....
Recoded_env 3'Bwt : .....

*      220     *      240     *      260     *      280     *      300     *      320     *      340     *      360     *      380     *      400
WT_env      : ITQACPKVSEFPIIHYCAPAGFALLKCNKTFNGTGPCTNVSTVQCTHGIRFVTVSTQQLLNGSLAEVEVIRVSNVFTDPAKTIIVQLMSTVEINCTRFNNTRKRIRIQRGPGRAFVITIGKIGNMRQAHCNISRKWNNTLKQIASKLRQFGNKNKTIIFKQSSGGDPEIVTHSFNCGGEPFYCNSTQLFNSTWFNSTW
Recoded_env : .....
Recoded_env mm : .....
Recoded_env sCpG1 : .....
Recoded_env sCpG2 : .....
Recoded_env 5'wt : .....
Recoded_env 3'wt : .....
Recoded_env 3'Awt : .....
Recoded_env 3'Bwt : .....

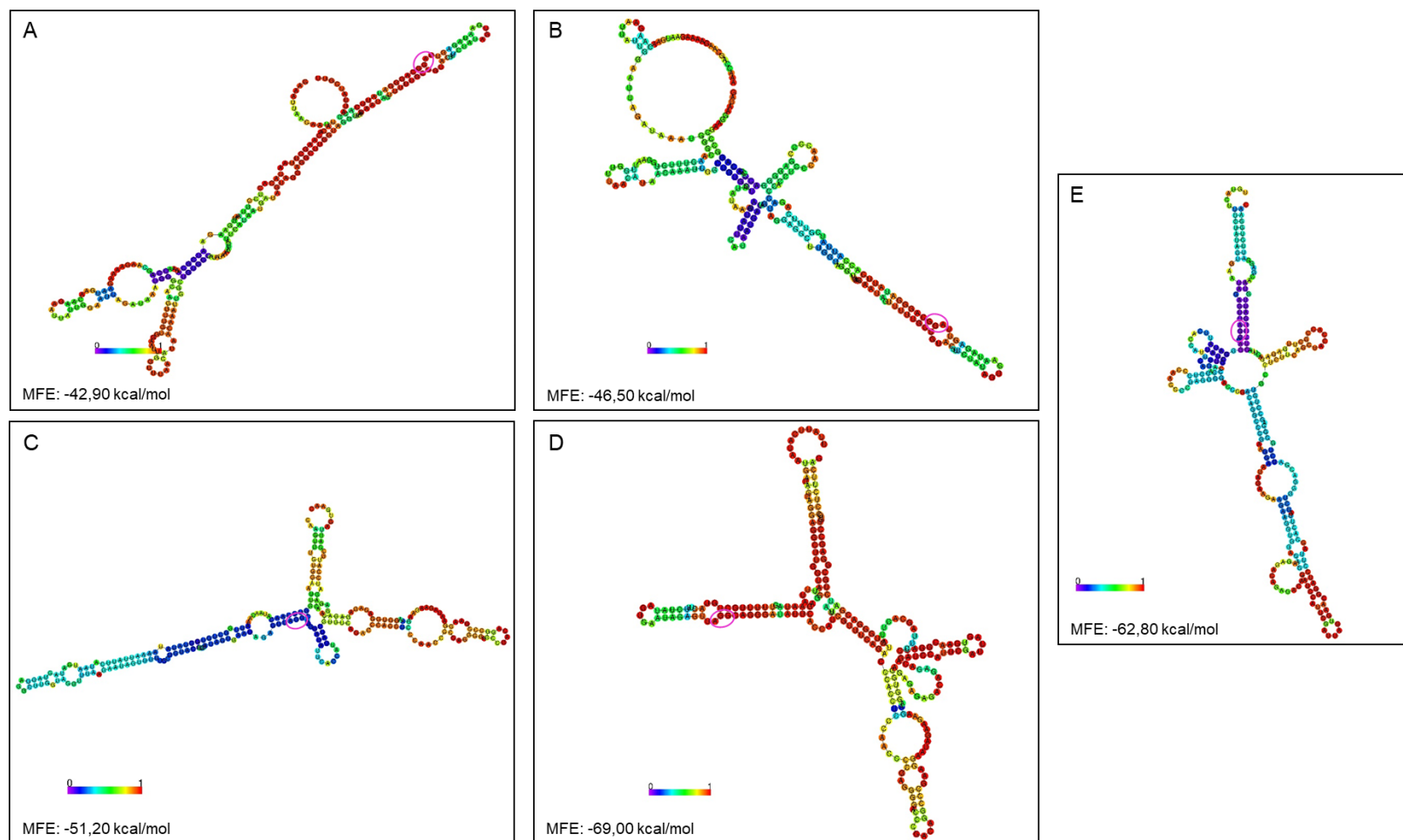
*      420     *      440     *      460     *      480     *      500     *      520     *      540     *      560     *      580     *      600
WT_env      : STEGSNNTGSDTITLPCRIKQIINMWQKVKAMYAPPISGQIRCSNITGLLLTRDGGNSNNESEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVRQEKRAVIGLALFLGLGAAGSTMGAAASMTLVQARQLLSGIVQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQLGIWGCSSG
Recoded_env : .....
Recoded_env mm : .....
Recoded_env sCpG1 : .....
Recoded_env sCpG2 : .....
Recoded_env 5'wt : .....
Recoded_env 3'wt : .....
Recoded_env 3'Awt : .....
Recoded_env 3'Bwt : .....

*      620     *      640     *      660     *      680     *      700     *      720     *      740     *      760     *      780     *      800
WT_env      : KLICTTAVPWNASWSNKSLEQIWNHTTWMEWDREINNYTSLIHSLEESQOQEKNEQELLELDKWSLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFAVLISIVNRVRQGYSPLSFQTHLPTPRGPDPRPEGIEEGERDRDRSIRLVNGSLALIWDLLRSLCLFSYHRLRDLILLIVTRIVELLGRGWALKYWNLL
Recoded_env : .....
Recoded_env mm : .....
Recoded_env sCpG1 : .....
Recoded_env sCpG2 : .....
Recoded_env 5'wt : .....
Recoded_env 3'wt : .....
Recoded_env 3'Awt : .....
Recoded_env 3'Bwt : .....

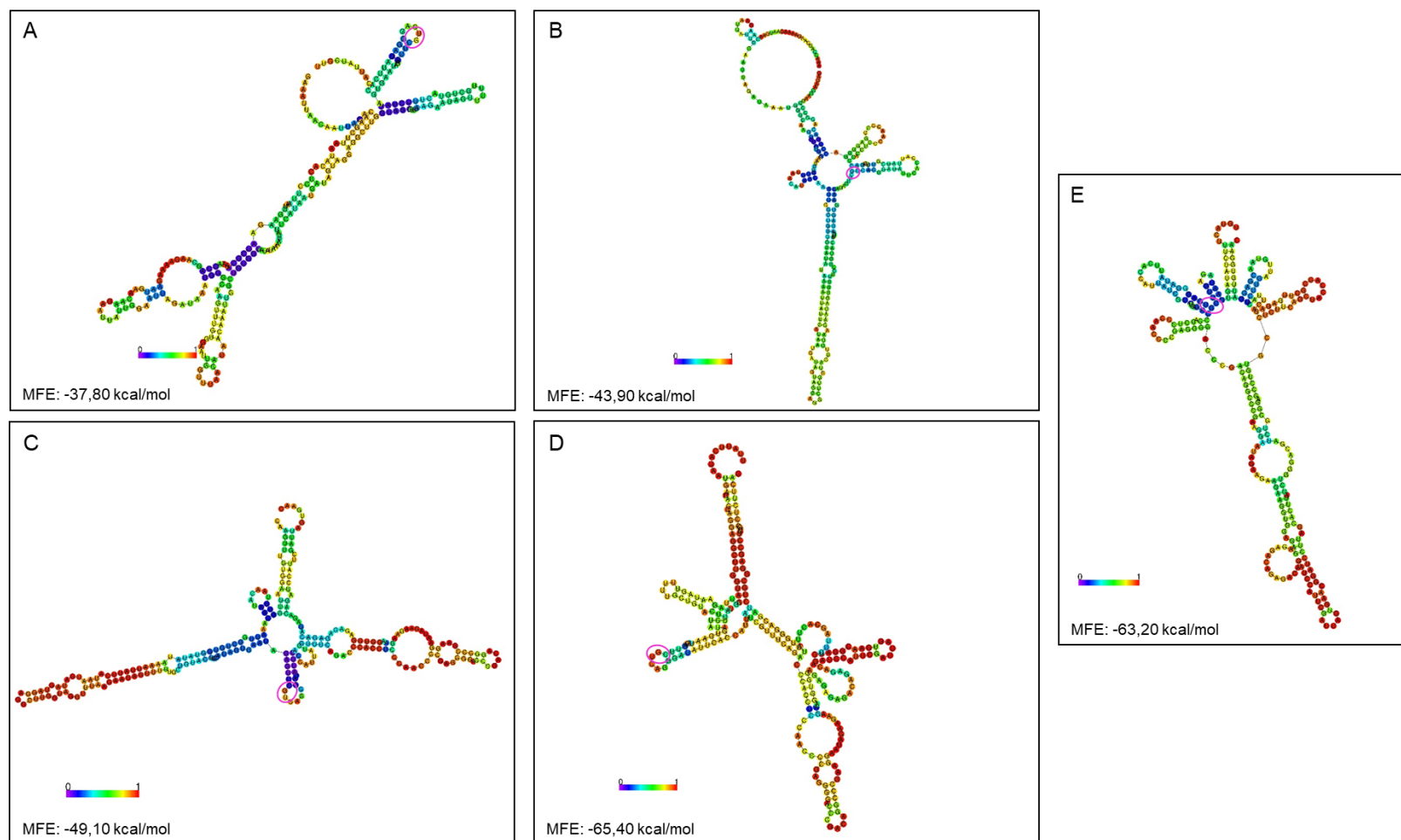
*      820     *      840     *
WT_env      : QYWSQELKNSAVNLLNATAIAVAEGTDRVIEVLQAAYRAIRHPRRIRQGLERILL
Recoded_env : .....
Recoded_env mm : .....
Recoded_env sCpG1 : .....
Recoded_env sCpG2 : .....
Recoded_env 5'wt : .....
Recoded_env 3'wt : .....
Recoded_env 3'Awt : .....
Recoded_env 3'Bwt : .....

```

**Annex Figure 3. Amino acid Env alignment of WT\_env and all recoded envelopes based on Recoded-env. All changes were synonymous.**

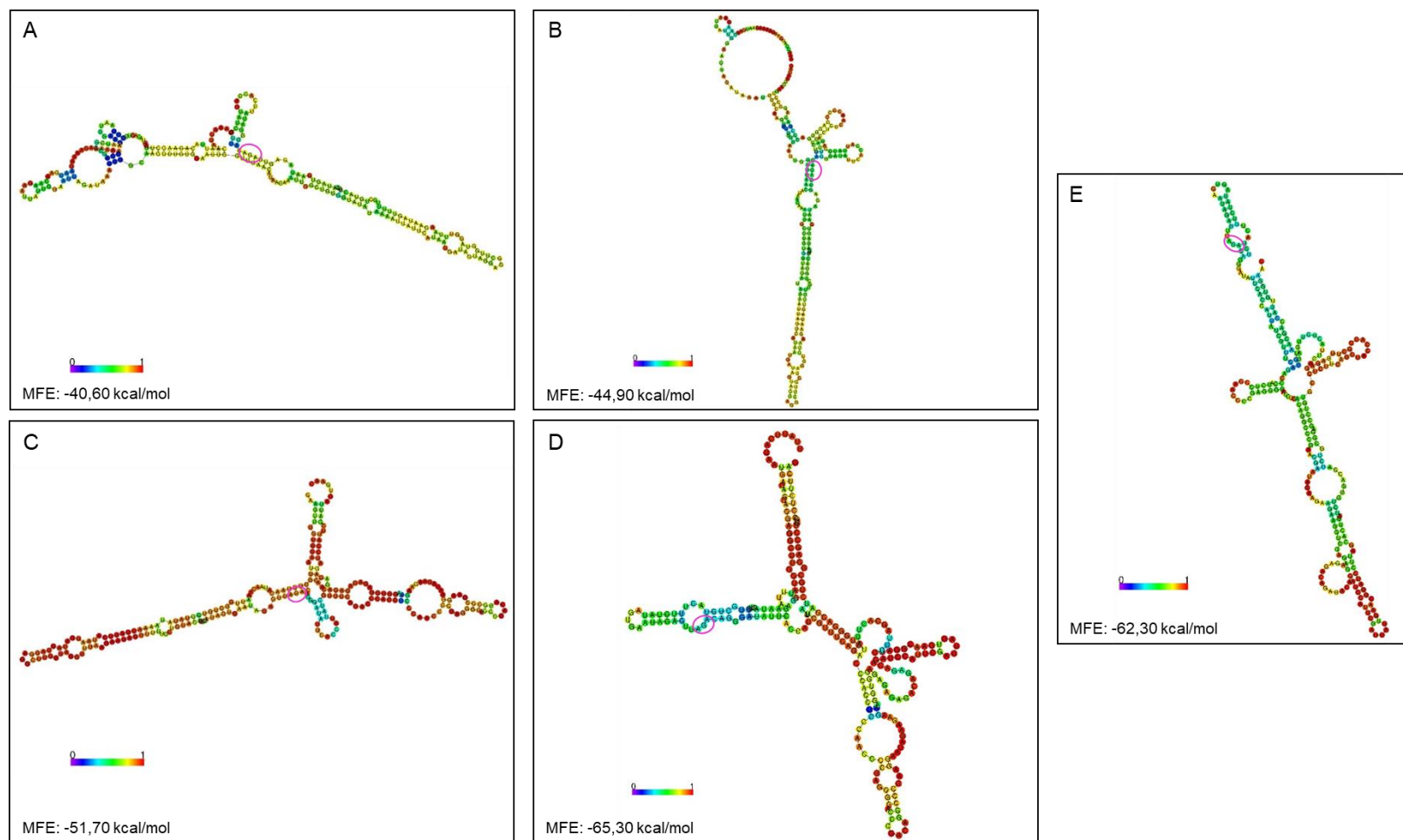


**Annex Figure 4. RNAfold structures of the different windows containing codon 34 in the WT *env* sequence.** The color code represents the base-pairing probability of the individual bases predicted by RNAfold. Purple means low probability; red means high probability. Codon 34 is circled in pink. A) Fragment 1900-2150. B) Fragment 1950-2200. C) Fragment 2000-2250. D) Fragment 2050-2300. E) Fragment 2100-2350.

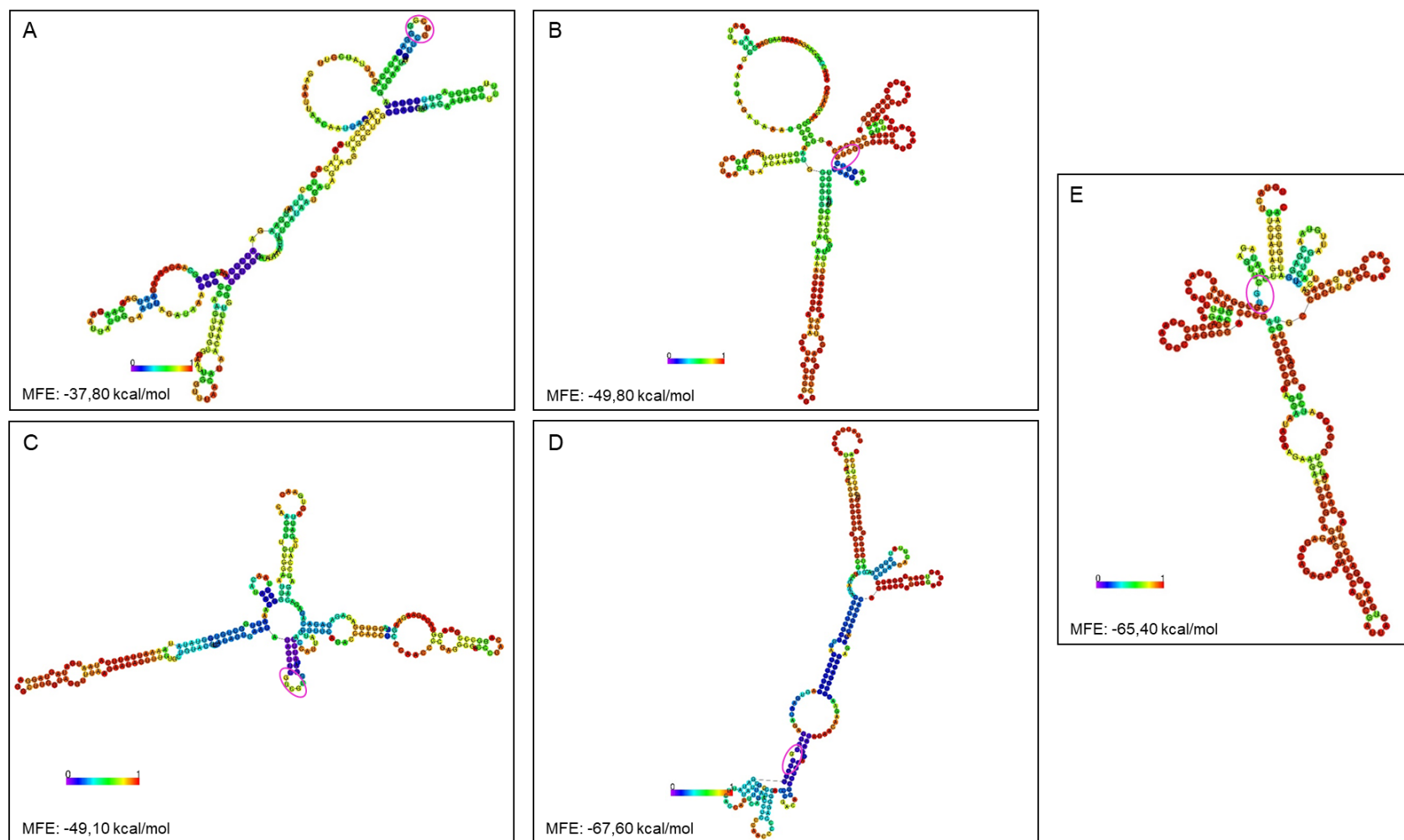


**Annex Figure 5. RNAfold structures of the different windows containing codon 34 in Recoded-env nucleotide sequence.** The color code represents the base-pairing probability of the individual bases predicted by RNAfold. Purple means low probability; red means high probability. Codon 34 is circled in pink. A) Fragment 1900-2150. B) Fragment 1950-2200. C) Fragment 2000-2250. D) Fragment 2050-2300. E) Fragment 2100-2350.





**Annex Figure 6. RNAfold structures of the different windows containing codon 34 in Recoded\_env\_sCpG2 env sequence.** The color code represents the base-pairing probability of the individual bases predicted by RNAfold. Purple means low probability; red means high probability. Codon 34 is circled in pink. A) Fragment 1900-2150. B) Fragment 1950-2200. C) Fragment 2000-2250. D) Fragment 2050-2300. E) Fragment 2100-2350.



**Annex Figure 7. RNAfold structures of the different windows containing codon 34 and the new mutations generated in culture in Recoded\_env\_mm env sequence.** The color code represents the base-pairing probability of the individual bases predicted by RNAfold. Purple means low probability; red means high probability. Codon 34 is circled in pink. Also included in this circle is the new generated mutation. A) Fragment 1900-2150. B) Fragment 1950-2200. C) Fragment 2000-2250. D) Fragment 2050-2300. E) Fragment 2100-2350.





```

*      2180      *      2200      *      2220      *      2240      *      2260      *      2280      *      2300      *      2320      *      2340
WT_env : CTCCCAACCCCGAGGGGACCCGACAGGCCCAGGAAGTATAGAAGAAGAGTGGAGAGAGACAGAGACAGATCCATTCGATTAGTGAAACGGATCCTTAGCACTTATCTGGGACGATCTGGGGAGCCTGTGCCCTCTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGG
Max : .....
MaxCpG : .....
Neu : .....
NeuCpG : .....
Min : .....
MinCpG : .....
MinCpG.2 : .....

*      2360      *      2380      *      2400      *      2420      *      2440      *      2460      *      2480      *      2500      *      2520
WT_env : ATTGTGGAACCTCTGGGACCGAGGGGTGGGAAGCCCTCAAATATGGTGGAAATCTCCTACAGTATGGAGTCAGGAACCTAAAGAAATAGTGCTGTTAACTTGCTCAATGCCACAGCCATAGCAGTAGCTGAGGGGACAGATAGGGTTATAGAAGTATTACAAGCAGCTTATAGAGCTATT
Max : .....
MaxCpG : .....
Neu : .....
NeuCpG : .....
Min : .....
MinCpG : .....
MinCpG.2 : .....

*      2540      *      2560
WT_env : CGCCACATACCTAGAAGAATAAGACAGGGCTTGGAAAGGATTTTGCTA
Max : .....
MaxCpG : .....
Neu : .....
NeuCpG : .....
Min : .....
MinCpG : .....
MinCpG.2 : .....

```

**Annex Figure 8. Nucleotide sequence alignment of WT\_env and all env sequences designed after changing the CPB values. Regions maintained WT are colored in blue. Substitutions of each mutant are shown within each corresponding sequence.**

```

*      20      *      40      *      60      *      80      *      100      *      120      *      140      *      160      *      180
WT_env : ATGAGAGTGAAGGAGAAATATCAGCACCTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTTGATGATCTGTAGTGTCTACAGAAAAATTTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCACCCTCTATTTTGTGCATCAGATGCTAAAGCA
Max-3'wt : .....
MaxCpG-3'wt : .....
Neu-3'wt : .....
NeuCpG-3'wt : .....
Min-3'wt : .....
MinCpG-3'wt : .....

*      200      *      220      *      240      *      260      *      280      *      300      *      320      *      340      *      360
WT_env : TATGATACAGAGGTACATAATGTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAAATTTAACATGTGGAATAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCATAAGCCATGTGTA
Max-3'wt : .....
MaxCpG-3'wt : .....
Neu-3'wt : .....
NeuCpG-3'wt : .....
Min-3'wt : .....
MinCpG-3'wt : .....

```

```

WT_env      : AAATTAACCCACTCTGTGTTAGTTTAAAGTGCACCTGATTTGAAGAATGACTACTAATACCAATAGTAGTAGTCGGGAGAAATGATAATGGAGAAAGGAGAGATAAAAACCTGCTCTTTTCAATATCAGCACACAGCATAAGGTAAGGTGCAGAAAGAAATATGCATTTTTTTTAAACTTGTAT
Max-3'wt    : ..GC.G..G..G.....GTCG..G..A..T.....A..A.....C..C..C..A..TCATCATCA..A.....C.....A.....G.....AGC.....C..TC..GTC..TC..A..A..A..A.....T.....T.....T.....A..
MaxCpG-3'wt : ..GC.G..T..C.....GTCC..G..A..T.....A..A.....C..C..C..A..TCATCATCA..A.....C.....A.....G.....AGC.....C..TC..GTC..TC..A..A..A..A.....T.....T.....T.....T..A..
Neu-3'wt     : ..GC.G.....T..A.....GTCCC..G.....T.....CC.T..A.....C..A..C.....TCA..TCT..C..G.....G..A..C..G.....T..T.....ATCT..TTCT.....C.....G..G.....T..G..C
NeuCpG-3'wt : ..GC.G.....T..A.....GTCCC..G.....T.....CC.T..A.....C..A..C.....TCA..TCT..C..G.....G..A..C..G.....T..T.....ATCT..TTCT.....C.....G..G.....T..G..C
Min-3'wt     : ..C.G..A..GT..G..C..A..GTCAC..C..A.....C..A.....C.....A..CTC.TCCTCT..C.....C.....A..G.....G..T..T..C..T..C..ATCG..GTCA..T.....C..A.....A..G.....C..C.....C.....G.....
MinCpG-3'wt : ..C.G..A..CT..G.....G.....C..G..A.....C..A.....C.....A..CTC.TCCTCT..C.....C.....A..G.....G..T..T..C..T..C..ATC..TCA..T..G..G..A.....A.....G.....C.....C.....G.....

*          380          *          400          *          420          *          440          *          460          *          480          *          500          *          520          *          540

WT_env      : ATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCTGCTCCAAAGGATCCCTTTGAGCCAATTCACATACATTATTGTCGCCCGGCTGGTTTGGGATTCATAAAATGTAATAAAGAGCTCAATGGAAACAGGACCATGTACA
Max-3'wt    : ..T..T..T.....C..C..TCCA.....A..A..TCA.....T..T.....A..A..T..A.....C..C..A.....T.....C..C.....C..C..C.....T..A..A..A.....C..C..C..G.....A..A..T.....C..G..C..G..C..C
MaxCpG-3'wt : ..T..T..T.....C..C..TCCA.....A..A..TCA.....T..T.....A..A..T..A.....C..C..A.....T.....C..C.....C..C..C.....T..A..A..A.....C..C..C..G.....A..A..T.....C..G..C..C..C..C
Neu-3'wt     : ..T..T.....C.....C..A..ATCA.....A.....TTCC..C.....A.....A.....T..A..C..T.....T..A..C..T.....C.....G..T..A..C.....C..AT..G..G.....A..T..T.....G..G..T..C.....T
NeuCpG-3'wt : ..T..T.....C.....C..A..ATCA.....A.....TTCC..C.....A.....A.....T..A..C..T.....T..A..C..T.....C.....G..T..A..C.....C..AT..G..G.....A..T..T.....G..G..T..C.....T
Min-3'wt     : ..C.....T..C..C.....C.....ATCG..C..AC.....TCG.....T..A..C..G.....A.....T..A..G..A..C.....C.....G..T..G..A..C.....C.....C.....T..T..C..G..T..C.....T
MinCpG-3'wt : ..T.....T..T..C.....C.....ATCA..C..AC.....TCA.....T..A..G..G.....A..C..T.....AG.....C.....A..T..A..A.....G.....C.....T..T..C..T..C.....C.....

*          560          *          580          *          600          *          620          *          640          *          660          *          680          *          700          *          720

WT_env      : AATGTCACACAGTACAATGTACACATGGAATTAGGCAGTAGTATCAACTCACTGCCTGTTAAATGGCAGCTAGCAGAAGAAGAGGTAGTAATAGTAGTCTGCAATTTCCAGGACAATGCTAAAACCATATAGTACAGCTGAACACACTCTGTAGAATTAATGTACAGCCCAAC
Max-3'wt    : ....TTCT..T.....T.....C..CC.C..G..G..C..C.....A..AC.....TTC.T.....A.....A.....AG..T.....T..T..T.....A.....A.....A.....A..T..T.....G..G..C..C..C..GC.G..C
MaxCpG-3'wt : ....TTCT..T.....T.....C..C..C.....C..G..C..C.....A..AC.....TTC.T.....A.....A.....AG..T.....T..T..T.....A.....A.....A.....A..T..T.....G..G..C..C..C..T..G.....
Neu-3'wt     : ..GTC.....G..G..C..T..C..G..C.....G.....G..A..TC..C.....TCA.....T.....G.....C..T..G.....T..T.....C.....T..G.....A.....T..C..G.....TC.T.....T
NeuCpG-3'wt : ..GTC.....G..G..C..T..C..G..C.....G.....G..A..TC..C.....TCA.....T.....G.....C..T..G.....T..T.....C.....T..G.....A.....T..C..G.....TC.T.....T
Min-3'wt     : ..GTCA..C..G.....C..C..AC..A..G..T..G.....C.....CT..AC..G..C.....TCA..C..G.....G..C..G..C..G.....T..C.....G..G..T..T.....C..A..C..T.....G..T..G.....CC.T..G..T
MinCpG-3'wt : ..GTCA..T..G.....C..A..A..G..T..G.....C.....T..AC..T.....TCA..T..G.....G..C..G.....G..T..A.....G..G..T..T.....G..A.....T..G.....C.....C.....

*          740          *          760          *          780          *          800          *          820          *          840          *          860          *          880          *          900

WT_env      : AACATACAGAAAAGAAATCCGTATCCAGAGAGACCAGGGAGAGCAATTTTGTACAATAGGAAAATAGGAAATAGAGACAAGCACATTTGTAACATTAGTAGAGCAAAATGGAATAAACCTTTAAAACAGATAGCTAGCAAAATTAAGAGAACAATTTGGAATAATAAAAACAATATC
Max-3'wt    : .....C..TC.....AA..A..A.....T..T..C..G.....C..C..C..G..G..C..G..C.....G..G..G..C..T..ATCA.....T.....A..T..CTC.....G..C..C.....A
MaxCpG-3'wt : .....C..T.....AA..A..A.....T..T.....G.....C..C..T..G..G..C..G..C.....G..G..C..G..C.....T..ATCA.....T.....A..T..CTC.....G..C..C.....A
Neu-3'wt     : ..T.....TC..G..GC..G..A..A.....G.....G..C..G..C..T..C..G..T..C..C.....G..G..C.....T..A..C..T.....T..C..G..A..T..C..T.....G.....G.....C..C.....T..T..T
NeuCpG-3'wt : ..T.....T..G..G..G..A..A.....G.....G..C..G..C..T..C..G..T..C..C.....G..G..C.....T..A.....G.....T..C..G..A..T..C..T.....G.....G.....C..C.....T..T..T
Min-3'wt     : .....C..G.....C..T..T..G..A..C..C..G.....C.....G..C..G..G.....C.....G.....C.....G.....ATCCC..T..G..G.....C..T..AC..G.....T..GTCT..C..GC..T.....G..C..G.....C.....G.....
MinCpG-3'wt : .....T..G.....TA..G..A.....G..G.....C.....C.....G.....C.....C.....C.....ATCCC..G.....G.....C..T..AC..G.....T..A..T..C..G.....G.....C.....C.....

*          920          *          940          *          960          *          980          *          1000          *          1020          *          1040          *          1060          *          1080

WT_env      : TTTAAGCAATCCTCAGGAGGGACCAGAAATTTGTAACGCACAGTTTAAATGTGGAGGGGAATTTTCTACTGTAAATCAACACAACCTGTTTAAATAGTACTTGGTTTAAATAGTACTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCCTCCCATGCGAATA
Max-3'wt    : ....A..AGT..T.....T.....C..C..C..C..C.....A..G..C.....T..C..C..C..G.....C..C..C..G.....AAGT.....T.....C.....G..G..C..C.....
MaxCpG-3'wt : ....A..AGT..T.....T.....C..C..C..C..C.....A..G..C.....T..C..C..C..A.....C..C..C..G.....AAGT.....T.....C.....G..G..C..C.....
Neu-3'wt     : ....A..AG..G..T.....T..T.....G..C..TCC.....C..C..G..C..T..T.....A.....TCA..A.....TCA..A..TCA..A..G..CAGT.....T.....G..CTC.....T..A..A..A..T..TC.....
NeuCpG-3'wt : ....A..AG..T..T..T..T.....G..C..TCC.....C..C..G..C..T..T.....A.....TCA..A.....TCA..A..TCA..A..G..CAGT.....T.....G..CTC.....T..A..A..A..T..TC.....
Min-3'wt     : .....G.....T..C.....T..C..G..A..C..A..TC..C.....C..C..C..G.....C..T..C.....CTC..A.....C..TCC.....TCG..G.....C.....C..T..C..G.....TC.....T..T..A..A..T..TC..T..T
MinCpG-3'wt : .....G.....T..G.....T..T..G..A..C..A..TC..C.....G.....C.....C.....C.....CTC..A.....C..TCC.....TCG..A..G..A.....C..T..G.....TC.....T..T..A..G.....T.....T

*          1100          *          1120          *          1140          *          1160          *          1180          *          1200          *          1220          *          1240          *          1260

WT_env      : TTTAAGCAATCCTCAGGAGGGACCAGAAATTTGTAACGCACAGTTTAAATGTGGAGGGGAATTTTCTACTGTAAATCAACACAACCTGTTTAAATAGTACTTGGTTTAAATAGTACTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCCTCCCATGCGAATA
Max-3'wt    : ....A..AGT..T.....T.....C..C..C..C..C.....A..G..C.....T..C..C..C..G.....C..C..C..G.....AAGT.....T.....C.....G..G..C..C.....
MaxCpG-3'wt : ....A..AGT..T.....T.....C..C..C..C..C.....A..G..C.....T..C..C..C..A.....C..C..C..G.....AAGT.....T.....C.....G..G..C..C.....
Neu-3'wt     : ....A..AG..G..T.....T..T.....G..C..TCC.....C..C..G..C..T..T.....A.....TCA..A.....TCA..A..TCA..A..G..CAGT.....T.....G..CTC.....T..A..A..A..T..TC.....
NeuCpG-3'wt : ....A..AG..T..T..T..T.....G..C..TCC.....C..C..G..C..T..T.....A.....TCA..A.....TCA..A..TCA..A..G..CAGT.....T.....G..CTC.....T..A..A..A..T..TC.....
Min-3'wt     : .....G.....T..C.....T..C..G..A..C..A..TC..C.....C..C..C..G.....C..T..C.....CTC..A.....C..TCC.....TCG..G.....C.....C..T..C..G.....TC.....T..T..A..A..T..TC..T..T
MinCpG-3'wt : .....G.....T..G.....T..T..G..A..C..A..TC..C.....G.....C.....C.....C.....CTC..A.....C..TCC.....TCG..A..G..A.....C..T..G.....TC.....T..T..A..G.....T.....T

*          1280          *          1300          *          1320          *          1340          *          1360          *          1380          *          1400          *          1420          *          1440

WT_env      : AAACAAATATAAACATGTGGCAGAAAGTAGGAAAAGCAATGTATGCCCTCCCATCAGTGGACAATTTAGATGTTTCATCAATATACAGGGCTGCTTAAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATGAGGGACAATTTGGAGA
Max-3'wt    : ..G..G..C..C..G.....G..G..C..G..C.....TTC.....A.....T..TT..AT.....A..T..TCA..T.....AAGT.....G..A.....TCA..T.....T..C.....C.....G
MaxCpG-3'wt : ..G..G..C..C..G.....G..G..C..G..C.....TTC.....A.....T..TT..AT.....A..T..TCA..T.....AAGT.....G..A.....TCA..T.....T..C.....C.....G
Neu-3'wt     : ..G..G..C.....G..T..C..G.....T..A.....TCG..T.....C..T..CT.....TC..G..CC..T..C..C..C.....T..AGT.....G..G.....C.....A.....C..C..T
NeuCpG-3'wt : ..G..G..C.....G..T..C..G.....T..A.....TCG..T.....C..T..CT.....TC..G..C..C..G..A..C..C..C.....T..AGT.....G..G.....C.....A.....C..C..T
Min-3'wt     : ..G..G.....T..T.....A..G.....C..C..G.....G..G..A..C..C..C.....C..G..C.....CT..AT..GC..G.....C..C..C..C..G..TC..T..C.....A.....A.....G..A..C..C..G.....C.....C..
MinCpG-3'wt : ..G..G.....T..T.....A..G.....C..C..T.....A..C.....ATCA..C.....C..G..C.....C.....CT..AT..GC..G.....G.....G..G..TC..T.....A.....A.....G..A..G..G.....G.....

*          1460          *          1480          *          1500          *          1520          *          1540          *          1560          *          1580          *          1600          *          1620

WT_env      : AGTGAATATATAAATATAAAGTAGTAAAAATTTGAACCATTAGGAGTAGCACCACCAAGGCAAGAGAGAGTGGTGCAGAGAGAAAAGAGCAGTGGGAATAGGAGCTTTTCTCTGGGTTCTTGGGAGCAGCAGAGAAGCACTATGGGCGCAGCCTCAATGACGGTACAG
Max-3'wt    : ....GC..C.....A..G..GC..G.....
MaxCpG-3'wt : ....GC..C.....A..G..C..G.....
Neu-3'wt     : TCG.....G.....G..T.....G.....G.....
NeuCpG-3'wt : TCG.....G.....G..T.....G.....G.....
Min-3'wt     : ..C..G..G..C.....C.....G..T.....C..G..C..C.....
MinCpG-3'wt : ..G..G..C.....C.....G..T.....A..G..C..T.....

```

```

*      1640      *      1660      *      1680      *      1700      *      1720      *      1740      *      1760      *      1780      *      1800
WT_env   : GCCAGACAATTATTGTCCTGGTATAGTGCAGCAGCAGACAACAAATTTGCTGAGGGCTATTGAGGGCCAAACAGCATCTGTTGCCAACTCACAGTCTGGGGGCATCAAGCAGCTCCAGGCCAAGAATCCTGGCTGTGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGA
Max-3'wt : .....
MaxCpG-3'wt : .....
Neu-3'wt : .....
NeuCpG-3'wt : .....
Min-3'wt : .....
MinCpG-3'wt : .....

*      1820      *      1840      *      1860      *      1880      *      1900      *      1920      *      1940      *      1960      *      1980
WT_env   : AAACCTCATTTGCACCACCTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTAACAGAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAGAAATGAACAAGAATTA
Max-3'wt : .....
MaxCpG-3'wt : .....
Neu-3'wt : .....
NeuCpG-3'wt : .....
Min-3'wt : .....
MinCpG-3'wt : .....

*      2000      *      2020      *      2040      *      2060      *      2080      *      2100      *      2120      *      2140      *      2160
WT_env   : TTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTAAACATAACAACAAATGGCTGTGGTATATAAAATTAATTCATAATGATAGTAGGAGGCTTGGTAGGTTAAGAATAGTTTTGCTGTACTTCTATAGTGAATAGAGTTAGGCAGGGATATACCATTATCGTTTCAGACCCAC
Max-3'wt : .....
MaxCpG-3'wt : .....
Neu-3'wt : .....
NeuCpG-3'wt : .....
Min-3'wt : .....
MinCpG-3'wt : .....

*      2180      *      2200      *      2220      *      2240      *      2260      *      2280      *      2300      *      2320      *      2340
WT_env   : CTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGTGGAGAGAGACAGACAGATCCATTTCGATTAGTGAACGGATCCTTAGCACTTATCTGGGACGATCTCGGGAGCCTGTGCCCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTAACGAGG
Max-3'wt : .....
MaxCpG-3'wt : .....
Neu-3'wt : .....
NeuCpG-3'wt : .....
Min-3'wt : .....
MinCpG-3'wt : .....

*      2360      *      2380      *      2400      *      2420      *      2440      *      2460      *      2480      *      2500      *      2520
WT_env   : ATTGTGGAACCTTCGGGACGCAGGGGGTGGGAAGCCCTCAAATAATTGGTGAATCTCCTACAGTATTGGAGTCAGGAACAAAGAATAGTGTGTTAACTTGCTCAATGCCACAGCCATAGCAGTAGCTGAGGGGACAGATAGGGTTATAGAAGTATTACAAGCAGCTTATAGAGCTATT
Max-3'wt : .....
MaxCpG-3'wt : .....
Neu-3'wt : .....
NeuCpG-3'wt : .....
Min-3'wt : .....
MinCpG-3'wt : .....

*      2540      *      2560
WT_env   : CGCCACATACCTAGAAGAATAAGACAGGGCTTGAAAGGATTTTGCTA
Max-3'wt : .....
MaxCpG-3'wt : .....
Neu-3'wt : .....
NeuCpG-3'wt : .....
Min-3'wt : .....
MinCpG-3'wt : .....

```

**Annex Figure 9. Alignment of WT\_env and all env sequences designed after reverting the 3' region to WT in the CPB virus variants.** Regions maintained/reverted to WT are colored in blue. Substitutions of each mutant are shown within each corresponding sequence.

```

*      20      *      40      *      60      *      80      *      100     *      120     *      140     *      160     *      180
WT_env  : MRVKEKYQHLWRWGWRTMLLGLMLICSAATEKLWVTVYVYGVVPVWKEATTTFLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTEFNFMKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFNISTSRGKVQKEYAFFYKLD
Max     : .....
MaxCpG : .....
Neu     : .....
NeuCpG : .....
Min     : .....
MinCpG : .....
MinCpG.2 : .....

*      200     *      220     *      240     *      260     *      280     *      300     *      320     *      340     *      360
WT_env  : IIPIDNDTTSYKLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNKTFNGTGPCTNVSTVQCCTHGIRPVVSTQLLNGSLAEEVIVRSVNFDTNAKTIIVQLNITSVEINCTRPNNNTRKRIRIQRGPGRAFVTIGKIGNMRQAHCNISRAKWNNTLKQIASKLREQFGNNKTI
Max     : .....
MaxCpG : .....
Neu     : .....
NeuCpG : .....
Min     : .....
MinCpG : .....
MinCpG.2 : .....

*      380     *      400     *      420     *      440     *      460     *      480     *      500     *      520     *      540
WT_env  : FKQSSGGDPEIVTHSFNCGGEFFYCNSQLFNSTWFNSTWSTEGSNTEGSDTITLPCRIRKQIINMWQKVGKAMYAPPISGQIRCSSNITGLLLRDGGNSNNESEIFRPGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVIGALFLGFLGAAGSTMGAASMTLTVQ
Max     : .....
MaxCpG : .....
Neu     : .....
NeuCpG : .....
Min     : .....
MinCpG : .....
MinCpG.2 : .....

*      560     *      580     *      600     *      620     *      640     *      660     *      680     *      700     *      720
WT_env  : ARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQLQARILAVERYLKDQQLGIWGCSSGLICTTAVPWNASWSNKSLEQIWNHTTWMEWDREINNYTSLIHSLIEESQKNEQELLELDKASLWNNWFNITNWLWYIKLFIMIVGGVLVGLRIVFAVLSIVNRVRQGYSPLSFQTH
Max     : .....
MaxCpG : .....
Neu     : .....
NeuCpG : .....
Min     : .....
MinCpG : .....
MinCpG.2 : .....

*      740     *      760     *      780     *      800     *      820     *      840     *
WT_env  : LPTFRGPDRPEGIEEGGERDRDRSIRLVNGSLALIWDDLRLSCLFSYHRLRDLILLIVTRIVELLGRRGWEALKYWNLLQYWSQELKNSAVNLLNATAIAVAEGTDRVIEVLQAAAYRAIRHIPRRIRQGLERILL
Max     : .....
MaxCpG : .....
Neu     : .....
NeuCpG : .....
Min     : .....
MinCpG : .....
MinCpG.2 : .....

```

**Annex Figure 10. Amino acid Env sequence alignment of WT\_env and all envelopes designed after changing the CPB values. All changes were synonymous.**