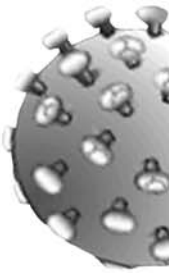
Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture

Gloria Martrus Zapater

Changes in codon-pair bias of HIV-1 have profound effects on virus replication in cell culture

Glòria Martrus Zapater



PhD thesis
2013

UNIVERSITAT DE BARCELONA

FACULTAT

Farmàcia

DEPARTAMENT

Bioquímica i Biologia Molecular

Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture

Glòria Martrus Zapater

Barcelona, 2013

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

DEPARTAMENT
Bioquímica i Biologia Molecular

PROGRAMA DE DOCTORAT de Biotecnologia

Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture

Memòria presentada per Glòria Martrus Zapater per optar al títol de doctor per la Universitat de Barcelona

El director de la tesi	La doctoranda	La tutora de la tesi:
Dr. Miguel Angel Martínez de la Sierra	**Glòria Martrus Zapater**	**Dra.Josefa Badia Palacín**

Glòria Martrus Zapater

2013

"I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale."

Marie Curie, 1933

"Conclusion: Big helix in several chains, phosphates on outside, phosphate-phosphate inter-helical bonds disrupted by water. Phosphate links available to proteins. "

Rosalind Franklin, 7 February 1952

A la meva família
i al Héctor

This work has been supported by the Instituto Carlos III, (PFIS scholarship, FI08/00031) and by the “Ministerio de Economía y competitividad” project BFU2010-15194.

Cover design by Cristina Lorca, Marina Montori and Glòria Martrus, 2013

SUMMARY

Human immunodeficiency virus type 1 (HIV-1) has a biased nucleotide composition different from human genes. This raises the question of how evolution has chosen the nucleotide sequence HIV-1 observed today, or to what extent the actual encoding contributes to virus replication capacity, evolvability and pathogenesis. Prior work has documented the effectiveness of making changes to the codon-pair bias of viral genomes in order to generate attenuated poliovirus and influenza virus. In this thesis, we applied the previously described synthetic attenuated virus engineering (SAVE) approach to HIV-1. Using synonymous codon pairs, we rationally recoded and codon pair–reoptimized and deoptimized different moieties of the HIV-1 gag and pol genes. RNA structures and codon usage of new recoded fragments were not affected by recoding. Deoptimized viruses had significantly lower viral replication capacity in MT-4 cells and peripheral blood mononuclear cells (PBMCs). Various degrees of *ex vivo* attenuation were obtained depending upon the specific deoptimized region and the number of deoptimized codons. After restricting viral replication to a single cycle by using a single-cycle HIV-1 vector, a significant reduction in protein production was observed in the vector carrying an attenuated virus variant. This reduction in protein synthesis was not accompanied by a reduction in the targeted transcript copy number, which strongly suggests that translation, and not transcription, is implicated in the generation of the attenuated phenotype by SAVE technology. A protease reoptimized virus carrying 38 synonymous mutations was not attenuated and displayed a replication capacity similar to that of the wild type virus in MT-4 cells and PBMCs. Although attenuation is based on several tens of nucleotide changes, after serial passages in MT-4 cells, both gag and protease deoptimized HIV-1 reverted to wild-type virulence in MT-4 cells while some maintain a certain attenuation degree in PBMCs. Quasispecies analysis of viral passaged sequences showed that attenuated viruses accumulated either synonymous mutations (reversions to wild-type sequences or novel mutations) or non-synonymous mutations. Recoded viruses explored different space sequences. Remarkably, no important reversion was observed in the reoptimized virus. Thus, these data demonstrate that SAVE is a useful strategy to gradually affect the replicative properties of HIV-1 by a mechanism that involves translation. HIV-1 with different degrees of attenuation can be a useful tool for the development of a safe and effective vaccine as well as the development of safer gene-therapy lentiviral vectors.

RESUM

El virus de la immunodeficiència humana 1 (VIH-1) conté una composició de nucleòtids diferent de la existent en el gens humans. Aquest fet planteja les qüestions de com la evolució ha triat la seqüència nucleotídica del VIH1 observada avui en dia, i de fins a quin punt aquesta seqüència actual contribueix a la capacitat replicativa, evolució i patogènesis virals. S'ha descrit que canvis en el ús de parelles de codons són eficaços per tal de generar virus atenuats de *Poliovirus* i *Influenza*. En aquesta tesi, hem aplicat la tecnologia prèviament descrita, "synthetic attenuated virus engineering" (SAVE) al VIH-1. Emprant parelles de codó sinònimes de manera racional, hem recodificat reoptimizant i desoptimizant per parelles de codó diferents fragments dels gens *gag* i *pol* del VIH-1. Les estructures de ARN i el us de codó dels nous fragments recodificats no es van veure afectades per la recodificació. Els virus desoptimizats van mostrar una replicació viral significativament inferior al virus control en cèl·lules MT-4 i en cèl·lules mononuclears de sang perifèrica (PBMCs). Depenent de la regió específica desoptimizada i del número de codons desoptimizats, es van obtenir diversos nivells d'atenuació *ex vivo.* Una reducció significant en la producció proteica es va observar quan la replicació viral va ser restringida a un sol cicle de replicació emprant un vector VIH-1 d'un sol cicle de replicació. La menor producció proteica no va correlacionar amb una reducció en el número de còpies del transcrit diana. Aquest fet suggereix que la transcripció, i no la traducció, es troba implicada en la generació dels fenotips atenuats produïts per la tecnologia de SAVE. El virus de proteasa reoptimizat que contenia 38 mutacions sinònimes, no es va mostrar atenuat, ans el contrari, mostrava una capacitat replicativa similar a la del virus control en cèl·lules MT-4 i en PBMCs. Encara que l'atenuació dels virus desoptimizats es basava en varies desenes de canvis nucleotídics, després de varis passis seriats en cèl·lules MT-4s, els virus desoptimizats de les regions de gag i proteasa van revertir a la virulència del virus control en cèl·lules MT-4. Alguns virus desoptimizats passats encara van mantenir un cert grau d'atenuació en PBMCs. Els anàlisis de quasiespècies de les seqüències dels virus passats en cultiu van mostrar que els virus atenuats acumulaven o bé mutacions sinònimes (reversions a la seqüència control o noves mutacions) o bé mutacions no-sinònimes. Els virus recodificats per la tecnologia de SAVE exploren diferents espais de seqüència. Singularment, no es va observar cap reversió important al virus reoptimizat passat en cultiu. Per tant, totes aquestes dades demostren que la tecnologia de SAVE és una estratègia útil per a afectar gradualment fenotípicament la capacitat replicativa del VIH-1, mitjançant un mecanisme que implica la traducció. El VIH-1 amb diferents nivells d'atenuació pot ser una eina utilitzable per al desenvolupament d'una vacuna segura i efectiva, així com pel desenvolupament de vectors lenvirals per a teràpia gènica més segurs.

RESUMEN

El virus de la inmunodeficiencia humana 1 (VIH-1) contiene una composición de nucleótidos diferente de la existente en los genes humanos. Este hecho plantea las cuestiones de cómo la evolución ha escogido la secuencia nucleotídica del VIH-1 actual, y de hasta qué punto dicha secuencia actual contribuye a la capacidad replicativa, a la evolución i a la patogénesis viral. Se ha descrito que cambios en el uso de parejas de codones son eficaces para generar virus atenuados de *Poliovirus* i *Influenza*. En esta tesis, hemos aplicado la tecnología previamente descrita, "synthetic attenuated virus engineering" (SAVE) al VIH-1. Mediante parejas de codones sinónimas de manera racional, hemos recodificado reoptimizando y desoptimizando per parejas de codón diferentes fragmentos de los genes *gag* y *pol* del VIH-1. Las estructuras de ARN i el uso de codón de los nuevos fragmentos recodificados no se vieron afectadas por la recodificación. Los virus desoptimizados mostraron una replicación viral significativamente inferior al virus control en células MT-4 y células mononucleares de sangre periférica (PBMCs). Dependiendo de la región específica desoptimizada y del número de codones desoptimizados, se obtuvieron diversos niveles de atenuación *ex vivo.* Una reducción significante en la producción proteica se observó cuando la replicación viral fue restringida a un solo ciclo de replicación mediante un vector VIH-1 de un solo ciclo de replicación. La menor producción proteica no correlacionó con una reducción en el número de copias del tránscrito diana. Este hecho sugiere que la transcripción, y no la traducción, se ve implicada en la generación de los fenotipos atenuados producidos por la tecnología de SAVE. El virus de proteasa reoptimizado que contenía 38 mutaciones sinónimas, no se mostró atenuado, al contrario, mostraba una capacidad replicativa similar a la del virus control en células MT-4 y en PBMCs. Aunque la atenuación de los virus desoptimizados se basaba en varias decenas de cambios nucleotídicos, después de varios pases seriados en células MT-4s, los virus desoptimizados de las regiones de gag y proteasa revirtieron a la virulencia del virus control en células MT-4. Algunos virus desoptimizados pasados aún mantuvieron un cierto grado de atenuación en PBMCs. Los análisis de cuasiespecies de las secuencias de los virus pasados en cultivo mostraron que los virus atenuados acumulaban o bien mutaciones sinónimas (reversiones a la secuencia control o nuevas mutaciones) o bien mutaciones no-sinónimas. Los virus recodificados por la tecnología de SAVE exploran distintos espacios de secuencia. Singularmente, no se observó ninguna reversión importante en el virus reoptimizado pasado en cultivo. Por lo tanto, todos estos datos demuestran que la tecnología de SAVE es una estrategia útil para afectar gradualmente fenotípicamente la capacidad replicativa del VIH-1, mediante un mecanismo que implica la traducción. El VIH-1 con distintos niveles de atenuación puede ser una herramienta útil para el desarrollo de una vacuna segura y efectiva, así como para el desarrollo de vectores lentivirales para terapia génica más seguros.

ABBREVIATIONS

AAV	Adeno-associated Virus
APOBEC	Apolipoprotein B mRNAEediting enzyme, Catalytic polypeptide-like
APV	Amprenavir
ATCC	American Type Culture Collection
ATV	Atazanavir
AZT	Azidothymidine
CA	Capsid
CCDS	Consensus CDS
CCR5	C-C chemokine receptor type 5
CD4	Cluster of Differentiation 4
CPB	Codon Pair Bias
CPS	Codon Pair Score
CXCR4	C-X-C Chemokine Receptor type 4
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
DRV	Darunavir
E.coli	*Escherichia coli*
ENC	Effective Number of Codons
FBS	Fetal Bovine Serum
GFP	Green Fluorescent Protein
HAV	Hepatitis A Virus
HEK 293T	Human Embryonic Kidney 293 Transformed cells
HIV-1	Human Immunodeficiency Virus type 1
IC50	Inhibitory Concentration 50
Il-2	Interleukin 2
IN	Integrase
LTR	Long Terminal Repeat
MA	Matrix
min	minute
MOI	Multiplicity of Infection
mRNA	messenger RNA
NC	Nucleocapsid
NFV	Nelfinavir
ng	nanograms
NIH	National Institutes of Health
nM	nanomolar
NRTI	Nucleoside Reverse Transcriptase Inhibitor
ORF	Open Reading Frame
PBMC	Peripheral Blood Mononuclear Cell
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PHA-P	Phaseolus vulgarus lectin P

PI	Protease Inhibitor
PR	Protease
RNA	Ribonucleic Acid
RPMI	Roswell Park Memorial Institute medium
RSCU	Relative Synonymous Codon Usage
RT	Reverse Transcriptase
RT-PCR	Real-Time PCR
SAVE	Synthetic Attenuated Virus Engineering
SD	Standart Deviation
sec	seconds
SHAPE	Selective 2'-hydroxyl acylation Analyzed by Primer Extension
TCID50	50% Tissue Culture Infectivity Dose
TNF-DF	Tenofovir Disoproxil Fumarate
tRNA	transfer RNA
µg	micrograms
µM	micromolar

Index

INTRODUCTION

1. HIV-1 (HUMAN IMMUNODEFICIENCY VIRUS 1)

Human immunodeficiency virus 1 (HIV-1) belongs to Retroviridae family, specifically to Lentivirus genus, a group of single-stranded, positive-sense, enveloped RNA viruses. This type of viruses replicates via a DNA intermediary, using reverse transcriptase enzyme, and then integrates in the host genome (Coffin *et al.*, 1997). The synthesis of genomic RNA and viral proteins come from integrated proviruses.

Based on phylogenetic analysis, four different groups of HIV-1 have been detected: M (Major), O (Outlier) (Mauclère *et al.*, 1997; Peeters *et al.*, 1997), N (Non-M non-O) (Simon *et al.*, 1998) and a recently discovered P (Putative) (Plantier *et al.*, 2009). Nowadays, group M is the main cause of HIV-1 pandemic. This group contains several subtypes (A-H), subtype B being the predominant strain in Europe, America, Australia and Asia (Keele *et al.*, 2008; Parrish *et al.*, 2013; Plantier *et al.*, 2009; Rambaut *et al.*, 2004; 2008; Simon *et al.*, 1998).

A. VIRION STRUCTURE

The HIV-1 infectious viral particle has a spherical structure of approximately 145nm. It consists of an exterior lipid bilayer deriving from cell host plasmatic membrane with two viral glycoproteins (gp120 and gp41) and an internal cone-shaped nucleoprotein core containing viral genetic material (Figure 1).

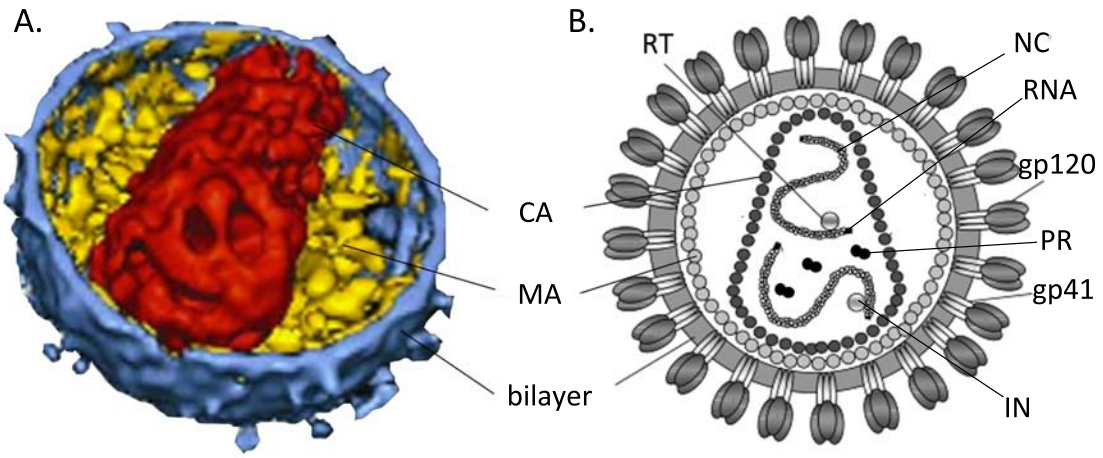


Figure 1. Schematic representation of the structure of HIV. (A) Reconstruction of a mature HIV particle after analysis of viral cryosections. Adapted from (Briggs *et al.*, 2006). (B). Schematic picture showing genomic RNA and viral proteins: MA: structural matrix; CA: structural Capsid; PR: protease; RT: Retrotranscriptase; NC: Nucleocapsid; IN: Integrase; gp120 and gp41 (envelope). Adapted from (Goh *et al.*, 2008).

Directly under the lipid bilayer, a dense layer of matrix protein (MA) provides structural integrity to the virion. The internal core of the viral particle is formed by approximately 2000 copies of the capsid protein p24 (CA). HIV-1 genome is packaged inside this core and involves two copies of RNA genome, each one linked with a tRNA molecule, which serves as an initiator of reverse transcription. RNA is also strongly bound to nucleocapsid (NC) proteins and enzymes, such as the integrase (IN) and the reverse transcriptase (RT) (Briggs & Kräusslich, 2011).

B. GENOME

HIV-1 genome is a single-stranded RNA of positive polarity with two identical copies of approximately 9.8kb. The genome is divided in nine different genes (Figure 2) and distributed in three different frames. It codifies for viral particle compounds (structural genes) and expression, regulatory and enzymatic proteins (enzymatic and regulator genes).

- Structural genes. The virion structure and the envelope are formed by four proteins derived from gag gene, matrix (MA), capsid (CA), nucleocapsid (NC) and p6, as well as by two proteins codified by env gene, gp120 and gp41.
- Enzymatic genes. Pol gene codifies for enzymatic proteins: protease (PR), reverse transcriptase (RT) and integrase (IN).
- Regulator genes. Six other additional or accessory proteins are encoded by HIV-1 genome in order to counteract cellular responses (Vif, Vpu, Nef and Vpr) plus regulate transcription (Tat) and post-transcription (Rev).

The two genome extremes are composed by two non-codifying sequences, named Long Terminal Repeats (LTRs). The ends of LTRs are critical for the provirus integration in the host genome. Once the provirus has been integrated, 5' LTR acts as the promoter for the complete retroviral genome transcription, while 3' LTR adds polyadenilation to new viral RNA.

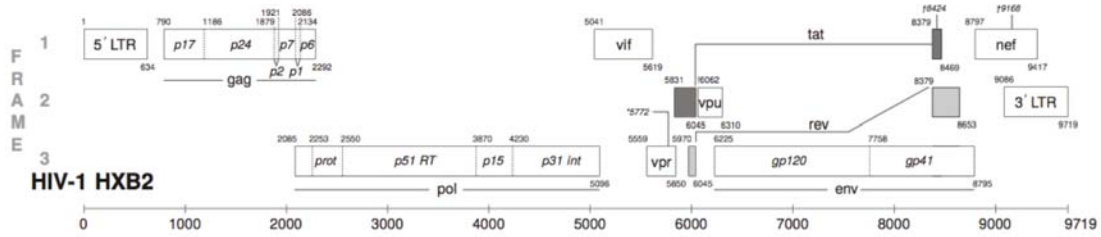


Figure 2. Schematic structure of HIV-1 genome organization. 9 open reading frames, encoding for the main viral proteins are despicted as rectangles. At each extreme, LTRs are flanking the viral genome. Adapted from (Kuiken *et al.*, 2012).

C. *VIRAL LIFE CYCLE*

As shown in Figure 3, HIV-1 cell cycle can be divided in two different phases:

I. Early Phase

a) Viral entry and uncoating: the viral protein gp120 attaches to its primary cell receptor CD4, leading to a structural change in the envelope complex and allowing gp120 to interact with the target chemokine receptor (mainly CXCR4 and CCR5). This more stable attachment allows the fusion of the viral and host membranes mediated and subsequent release of the viral content into the cell cytoplasm.

b) Reverse transcription/Integration: RT enzyme retro-transcribes single-stranded HIV RNA, generating a linear DNA duplex. The pre-integration complex, formed by IN and other accessory proteins like Vpr (Bukrinsky *et al.*, 1992), helps this DNA duplex to enter to the nucleus through nuclear pores. Then, LTRs and IN execute an arbitrary integration into the host genome, where provirus may stay latent for long periods of time.

II. Late phase:

a) Transcription and translation: the cellular machinery transcribes proviral RNA. This RNA is spliced into different transcripts. Later on, transcription of regulatory proteins and polyprotein precursors of structural genes occurs.

b) Virion production and budding: the formation of a new virion needs the assembly of all proteins (both regulatory and structural) and RNA near the host membrane. This stage ends with the budding of viral particles, located in the membrane regions expressing Env.

c) Maturation: PR induces proteolysis of gag-pol precursor. Mature viral particles are ready to infect new cells (Kohl *et al.*, 1988).

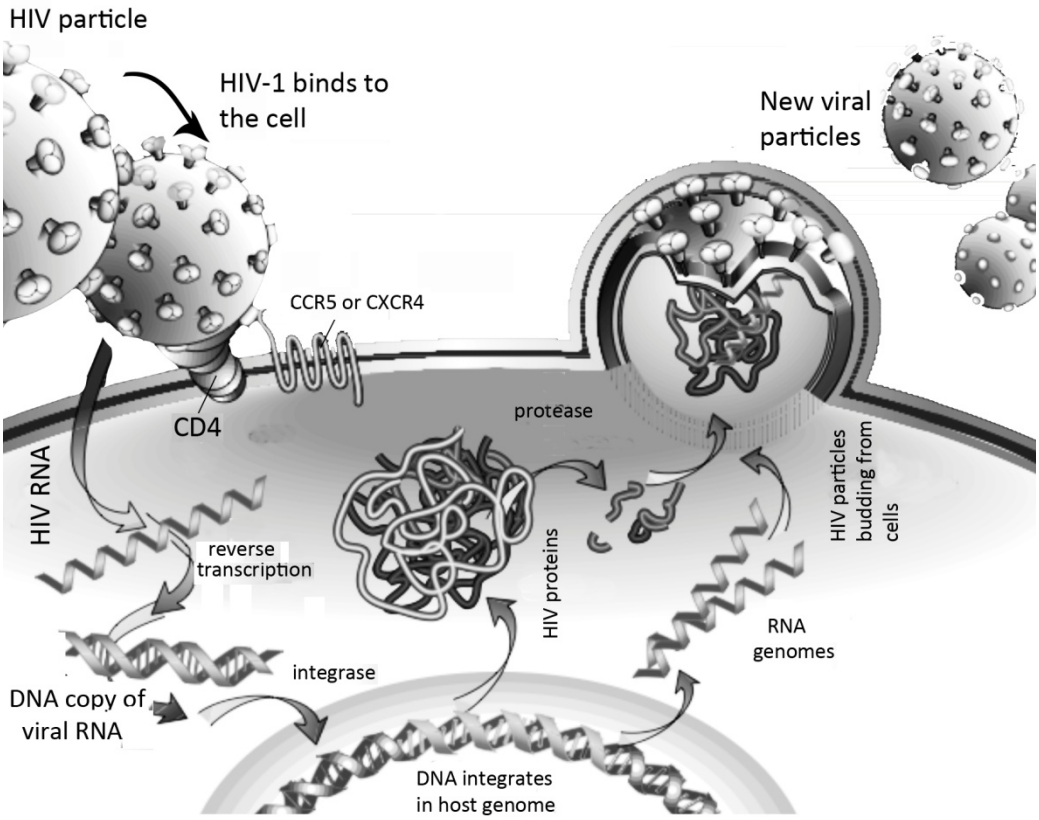


Figure 3. Life cycle of HIV-1 virus. Interaction between CD4+ receptor and co-receptor (CCR5 or CXCR4) is required to allow the viral entry into the cells. Then, once the viral content is released into the cell, RT transcripts the viral single-stranded (+)RNA genome to proviral double-stranded DNA. The newly viral dsDNA enters into the nucleus and IN enzyme integrates its into host cell genome, where it can stay latent for long periods of time. The transcription and translation machinery from the host cells translate the messenger mRNAs into a long polypeptide. The maturation endures after the budding of viral particles from the cells through the action of the PR. Adapted from (Weiss, 2001).

D. VIRAL FITNESS, EVOLUTION AND QUASISPECIES

Viral fitness was originally described as ‘‘the capacity of a virus to produce infectious progeny in a given environment’’ (Domingo & Holland, 1997). HIV-1 viral fitness is principally due to both, viral and host factors. In order to measure fitness *in vitro*, viral replication capacity assays can be performed and are defined as the ability of virus to multiply in a given environment, compared with a reference or control (wild-type) virus.

Viral attenuation, or the decrease of the virus virulence and impaired replicative capacity, can be achieved through different strategies, including sequential passage, deletion of genes or genetic engineering of viral genes encoding immunomodulatory proteins (Jacobs *et al.*, 2009). In *in vitro* experiments, a severe fitness reduction in HIV-1 virus was produced

by serial plaque-to-plaque bottleneck (accumulation of 4 to 28 mutations per genome in HIV-1) (Yuste *et al.*, 1999; 2000)(Figure 4). On the contrary, when large population passages are performed, a fast increase in fitness is generated (Domingo & Holland, 1997).

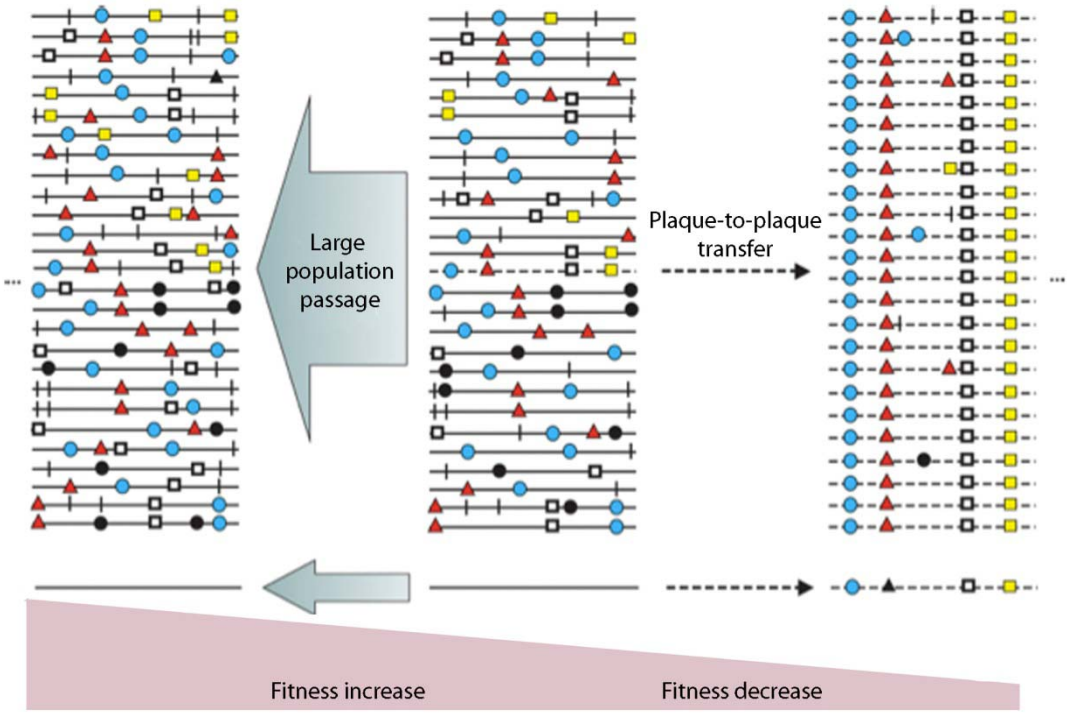


Figure 4. Fitness evolution of RNA virus populations. Typical heterogeneous distribution of RNA virus genomes is represented in the central scheme. Lines represent genome and symbol point mutations. Consensus sequence or median of the distribution is shown above. Initial quasispecies (in the center) can evolve through large population passages to a fitness increase (left panel) or through bottleneck passages or plaque-to-plaque passages to a fitness decrease (right panel). Large passages favor competition between variants and thus, generate an increase in fitness. Plaque-to-plaque passages produce an accumulation of deleterious mutations and therefore a decrease in fitness. Adapted from (Domingo *et al.*, 2012).

Viruses with RNA genome, together with viroids, have genomes that replicate with the highest error rates in nature (one mutation in 10^{-3}-10^{-5} nucleotides incorporated) when compared to DNA cellular error rates (Drake, 1991; 1993; Goodman & Fygenson, 1998; Kunkel & Alexander, 1986). This is due to the absence of proofreading and repair activities in their RNA-dependent RNA polymerases or DNA-dependent DNA polymerases (Drake, 1993; Drake & Holland, 1999; Mansky & Temin, 1995; Preston & Dougherty, 1996; Temin, 1993).

The main causes of HIV-1 variability are due to cellular restriction factors (e.g. APOBEC-3 family) and viral factors (recombination, high mutation and replication rate). Nucleotide misincorporation during HIV-1 reverse transcription has been estimated to be between 10^{-4} and 10^{-5} mutations per nucleotide and per replication cycle (Mansky & Temin, 1995). In an infected person, a rapid viral turnover occurs (10^{9} to 10^{10} viral particles per day) while a large number of cells are infected (10^{7} to 10^{8}). Given that HIV-1 genome is approximately 9.7kb long, there would be 0.9-9.7 mutations in every genome copied. This high variability exists not only between patients (interpatient variability) but also within the patient (intrapatient variability) (Pedroza martins *et al.*, 1992; Yoshimura *et al.*, 1996). Overall, the emergence of mutant clouds and the mixture of different, but highly related viral genomes, origins the phenomenon of viral quasispecies (Eigen, 1971).

2. CODON USAGE

The standard genetic code consists of 64 codons for a set of 20 amino acids and the stop signal, showing its redundancy (except for tryptophan and methionine) and implying that several synonymous triplets encode for the same amino acid (Figure 5). Usually, the position of such synonymous codons is not constant along the protein coding sequences and therefore, their properties are not entirely comparable. Relative amounts of tRNAs isoacceptors have been associated to codon usage patterns in different organisms (Gouy & Gautier, 1982; Gustafsson *et al.*, 2004; Ikemura, 1985; Moriyama & Powell, 1997). For instance, a certain architecture of the ribosome-decoding center, in combination with the specific structure of the individual tRNAs, could be crucial in the codon usage evolution in *E.coli* (Boycheva *et al.*, 2003).

Codon usage bias, the frequency of occurrence of synonymous codons in coding DNA, in human cells is limited by amino acids needed for protein structure/function and by genome signatures (dinucleotide relative abundances). In contrast, translational and transcriptional influences appear to play a minor role in human codon usage (Karlin & Mrázek, 1996).

HIV-1 virus, like other viruses, needs the cellular host translational and replication machinery to produce new viral particles. Analysis of HIV-1 relative synonymous codon usage (RSCU) has shown that HIV-1 RSCU is different to human RSCU (Grantham & Perrin, 1986; Kypr & Mrázek, 1987). This strategy has also been used to scape from host cell defense by other viruses, such as Hepatitis A virus (HAV), which also possesses a

significantly different frequency in codon usage when compared to the human one (Andrea *et al.*, 2011). In HIV-1 genome, the RSCU in early expressed genes (rev, tat and nef) is closer to the human RSCU than the RSCU in late expressed genes (gag, pol, env, vif, vpu and vpr) (Pandit & Sinha, 2011). This fact may be due to viral adaptation to different tRNAs cellular pools (van Weringh *et al.*, 2011). As a consequence of the limited availability of the matching tRNAs, HIV-1 codon composition leads to suboptimal protein expression in infected cells (Haas *et al.*, 1996). Despite HIV-1 early isolates do not differ in nucleotide composition from recent isolates (van der Kuyl & Berkhout, 2012), in infected persons a drift towards human RSCU in late sequences has been described (Meintjes & Rodrigo, 2005).

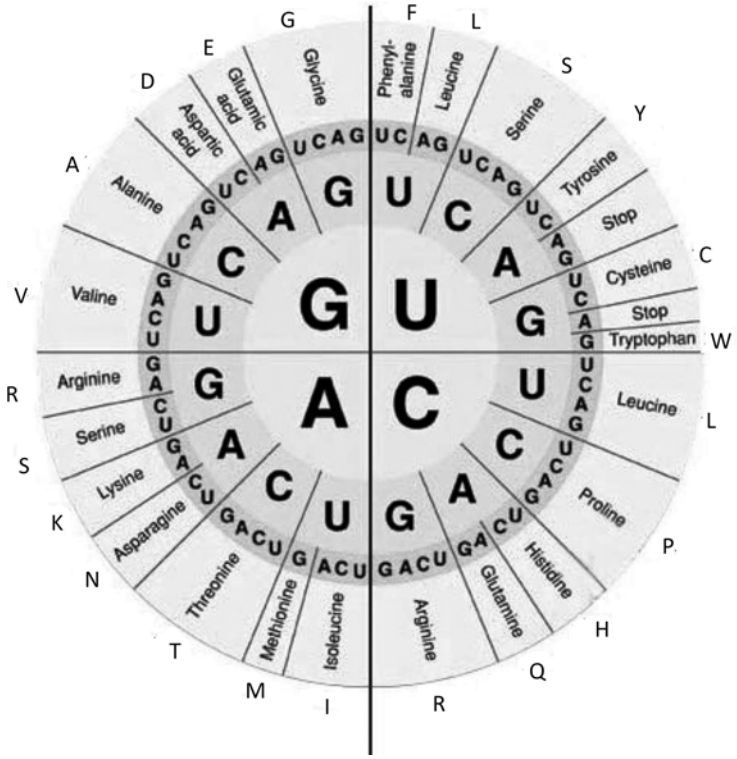


Figure 5. Codon usage bias. The codons are written in the 5' to 3' direction from the center of the circle to the exterior. All amino acids except methionine and tryptophan are represented by more than one codon. Generally, codons that codify for the same amino acid only differ at the third base. Stop codons are represented by UGA, UAA and UAG.

The remarkable nucleotide composition of HIV-1 genome, containing an above average percentage of A nucleotides, results in a codon composition different from the human one (van der Kuyl & Berkhout, 2012). Specifically, the more flexible third codon positions are preferentially occupied by A-nucleotides (Berkhout & van Hemert, 1994) (van Hemert *et al.*, 2013) that usually induce ribosome pausing and inefficient translation (Haas *et al.*, 1996). Still, it is unclear whether HIV-1 codon composition is an adaptive trait or has neutrally evolved through genome drift. In both cases, it may influence HIV-1 evolution as

well as drug resistance development (Mulder *et al.*, 2008; Wood *et al.*, 2009). The frequent G-to-A mutations observed in HIV-1 genome might be associated to the action of APOBEC3 family host enzymes (Anderson & Hope, 2008; Bishop *et al.*, 2007; Newman *et al.*, 2005). Changes in codon composition may have a significant impact on different aspects of HIV-1. For instance, a codon-optimized gag HIV-1 showed a significant increase in protein expression (up to a 60%) compared to wild-type gag in a *in vitro* study (Ngumbela *et al.*, 2008). However, it is unknown whether this codon optimization would increase viral fitness or virulence. Another study showed that recoding with a high level of preferred human codon usage may provoke the disruption of some unknown HIV-1 RNA properties deleterious for other reasons, e.g. disturbing the secondary structure of HIV-1 RNA genome (Sanjuán & Bordería, 2011).

By deoptimizing the codon usage of capsid gene in poliovirus, two groups have shown a reduction of viral fitness, with a decrease in virus-particle-specific infectivity up to 1,000-fold (Mueller *et al.*, 2006) and, genetic and phenotypic stability through 25 passages (approximately 50 replication cycles) (Burns *et al.*, 2006).

3. CODON PAIR USAGE

While the non-random usage of synonymous codons for the same amino acids is defined as codon bias, the codon pair bias means the non-random juxtaposition of codons. Codon pairs occur at irregular levels (Gutman & Hatfield, 1989; Moura *et al.*, 2007). The observed juxtaposition frequency of two codons is different to that expected if those codons were randomly located next to each other. This codon pair usage of two codon pairs can be quantified as a codon pair score (CPS) as well as the mean codon pair bias (CPB) of an through an open reading frame (ORF). Based on the importance of interactions between tRNA molecules binding to adjacent codons, two main hypotheses were proposed to explain the CPB observed *in E.coli*. Firstly, a significantly non-random motif of the 3721(61^2) possible pairs of codons used in protein-coding DNA sequences should exist. Secondly, this motif, if existing, should be correlated in its structure and abundance to tRNAs, and should also affect the levels of gene expression (Gutman & Hatfield, 1989). Codon pairs were found to be highly over-represented or under-represented independently from codon usage (Gutman & Hatfield, 1989), and this was associated with increased or decreased levels of protein expression (Shao *et al.*, 2012).

Codon pair usage is directional in prokaryotes. The hypothetical bias of codon-pair A-B is independent from the bias associated with codon-pair B-A. Since, they do not have the same statistical representation value (Irwin *et al.*, 1995), the CPB is determined to be unidirectional. A broad correlation between CPB and translation was demonstrated: the more under-represented codon pairs were present in a sequence, the faster they were translated, and vice-versa (Irwin *et al.*, 1995).

During translation in the ribosome, the formation of the peptide bond demands the adjustment of two codons and the accommodation of two tRNAs in the A and P ribosomal sites (Figure 6) (Uemura *et al.*, 2010). It has been suggested that, for steric causes, not all the combinations of tRNAs and codons might be equally fitting on the ribosome surface (Smith & Yarus, 1989) (even though ribosomes are only occupied briefly by two tRNAs (Uemura *et al.*, 2010)). Therefore, some codon pairs might have a translational benefit over others.

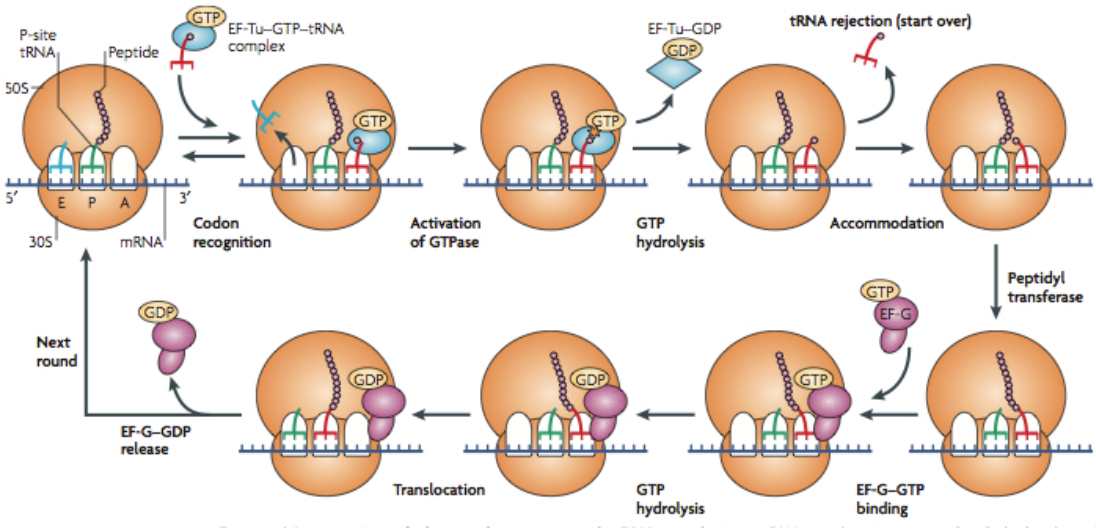


Figure 6. An overview of ribosomal structure and mRNA translation. A ribosome contains three binding sites for RNA molecules: one for mRNA and two for tRNAs. The peptidyl-tRNA binding site (P-site) holds the tRNA molecule that is linked to the growing end of the polypeptide chain and mRNA translation is started when the initial tRNAfmet binds the P-site (not shown). The aminoacyl-tRNA-binding site, or A-site, holds the incoming tRNA molecule charged with a new amino acid in complex with elongation factor (EF)-Tu-GTP. A peptidyl-transferase reaction follow and the peptide chain is transferred onto the A-site tRNA. The ribosome moves exactly three nucleotides along the mRNA in the 3' direction to decode the next mRNA codon. The deacylated P-site tRNA moves to the E-site and the peptidyl-tRNA at the A-site moves to the P-site. A new round of elongation can start. Adapted from (Steitz, 2008).

A bias in A and P ribosomal sites, between cP3 (third position of tRNA in P site) and

cA1/cA2 (first and second position in A site) was described, and was probably caused by the interplay of nucleotides of the codon-anticodon interaction. The third base of the P-site codon had an impact on the first and second base of the A-site codon (Buchan, 2006; Moura *et al.*, 2005). Independent groups described that most under-represented codon pairs possessed a central CG dinucleotide (e.g. the codon-pair Leu-Glu: CTC-GAA). CpG islands are methodically under-represented in mammals and viruses (Buchan, 2006; Curran *et al.*, 1995; Fedorov *et al.*, 2002). Probably evolution lowered CpG content to avoid mammals innate immune system and elude to trigger an immune response induction (Karlin *et al.*, 1994; Sugiyama *et al.*, 2005).

In microorganisms, concrete parts of tRNAs can exert a supplementary effect, disturbing codon pair preference (Buchan, 2006). Eukaryotes and prokaryotes have reversed codon pair usage rules: in eukaryotes pairs of rare codons are under-used while in prokaryotes they are over-used, and vice-versa (Buchan, 2006). General rules about codon pair usage in different organisms have been determined, being the top 10 under-represented or over-represented codon pairs universally conserved (Tats *et al.*, 2008). Furthermore, codon pair context was observed to be species-specific. In eubacteria and archeae, codon-pair usage appears to be determined by limits inflicted by the translational machinery (ribosome), while in eukaryotes the main pressure appears to be driven by the emergence of DNA methylation and tri-nucleotide repeats (Moura *et al.*, 2007). In a multiple ORFeome (a set of ORFs cloned from different species) analysis, translation was suggested to exert a more powerful impact on codon pair biases than molecular mechanisms (Tats *et al.*, 2008). By analyzing in an ORFeome context, evolutionary conservation of codon usage compared to codon pair usage, the trees topologies showed that, in eukaryotes, the resemblance of codon pair usage is higher than codon usage (Figure 7). Those trees differ both in branch lengths and positions, suggesting that codon usage and codon pair usage are independent, due to different molecular mechanisms (Tats *et al.*, 2008).

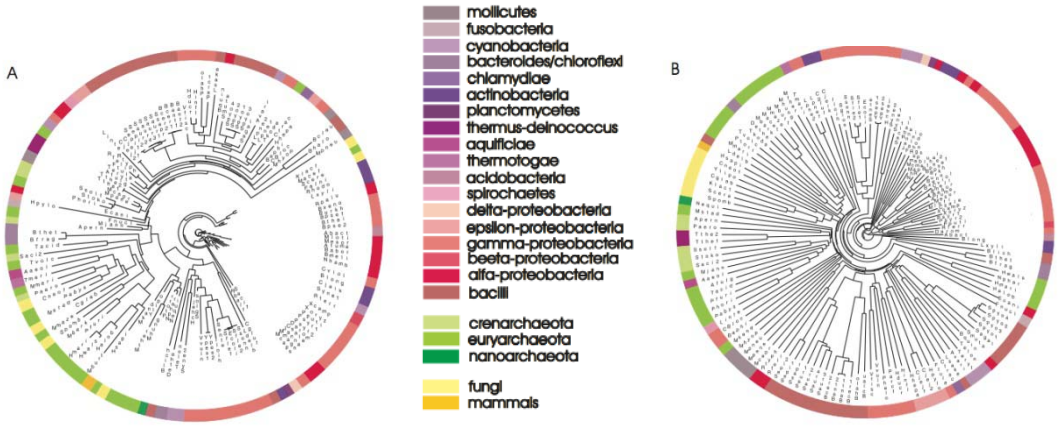


Figure 7. The evolutionary conservation of codon usage (A) and codon pair usage (B) in ORFeomes. Every branch lengths depict the correlation of codon usage and codon pair usage between two organisms-the higher the correlation, the shorter the branch. Generally, branch lengths of codon pair usage tree (B) are larger than those of codon usage tree (A), suggesting a higher similarity of codon usage than codon pair usage between different organisms. Eukaryotes are clustered all together in the codon pair usage tree (B) while in the codon usage tree they are spread all around. Adapted from (Tats *et al.*, 2008).

Finally, a co-evolution between the translational machinery used and a given organism has been suggested. The consequence of that co-evolution is the non-randomness in codon pair usage, probably explaining that, not only codon pair bias, but also codon bias and tRNA isoacceptors are species-specific (Eckard Wimmer, 2012).

CODON PAIR BIAS AND SAVE TECHNOLOGY

Recently a new approach, termed synthetic attenuated virus engineering (SAVE), was used to rationally design live attenuated poliovirus and influenza virus vaccines (Coleman *et al.*, 2008; Mueller *et al.*, 2010). SAVE works by recoding and synthesizing the viral genome. The wild-type amino acid sequence is preserved while the existing synonymous codons are rearranged to create a suboptimal arrangement of codon pairs (Coleman *et al.*, 2008).

For instance, as described by Coleman et al., 8 different codon pairs can encode the amino acid pair Ala-Glu. Thus, the expected frequency of each 8 encodings can be calculated by multiplying the individual frequencies of codon usage of the two individual codons (Figure 8). In order to calculate the individual frequency of each codon, the consensus coding DNA sequence (CCDS) database dating from March 2005 and containing a total of 14 795 human genes was used (http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi). If the ratio observed/expected is lower than one, this codon pair is considered under-represented and vice-versa. As shown in Figure 8, the natural logarithm of the ratio between

observed/expected frequencies of each human codon pair was used to calculate the Codon Pair Score (CPS). A positive CPS implies that the codon pair is statistically over-represented while a negative CPS implies that the pair is statistically under-represented. Considering these calculations in human genes, Ala-Glu pair is expected to be encoded by GCCGAA and GCAGAG approximately as often. However, the codon pair GCCGAA is strongly under-represented (CPS = -1.717) and is used only approximately one-seventh as often as GCAGAG (CPS = 0.411) (Coleman *et al.*, 2008).

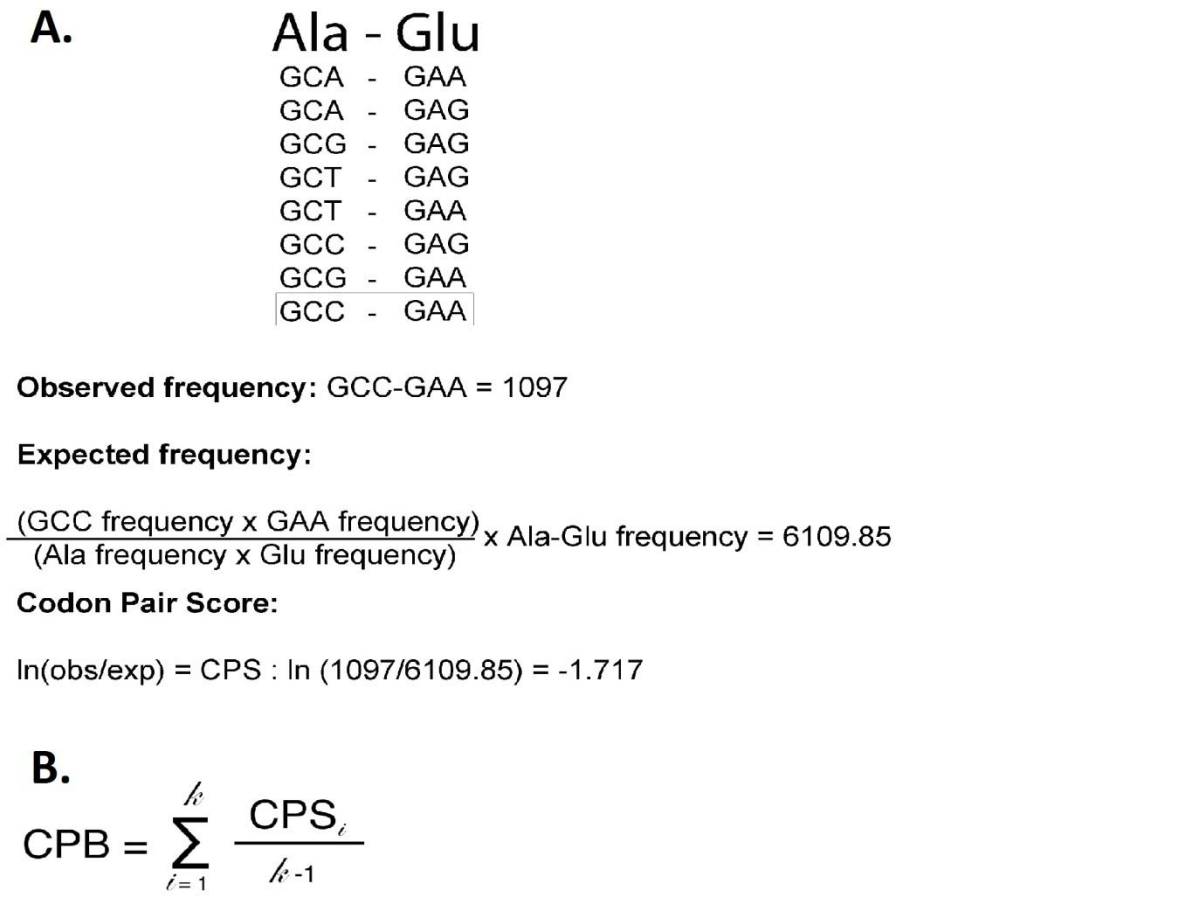


Figure 8. A specific codon-pair example to demonstrate how a CPS and CPB are calculated. A. The amino acid pair Ala-Glu can be encoded by eight possible codon-pairs. The individual CPS for GCC-GAA codon pair [A-E] is used as an example. The observed frequency is the summed appearance of this pair in all annotated human genes. The expected frequency is the calculated value if this pairing was simply random. The expected frequency normalizes for codon and amino acid frequencies. The CPS in this example was determined to have a negative value, indicating this codon-pair is statistically under-represented. B. The equation used to calculate the CPB for an entire gene. CPB is the arithmetic mean of the individual codon pair scores of all pairs making up the ORF. Adapted from (Coleman et *al*, 2008).

For an entire ORF, the CPB is calculated as a mean: the sum of all individual CPS in an ORF divided by the number of pairs present (Figure 8B). If the CPB resulting is positive, the gene

contains over-represented codon-pairs and vice-versa (Figure 9). As CPB score decreased, recoded *Poliovirus* and *Influenza* virus were more deoptimized.

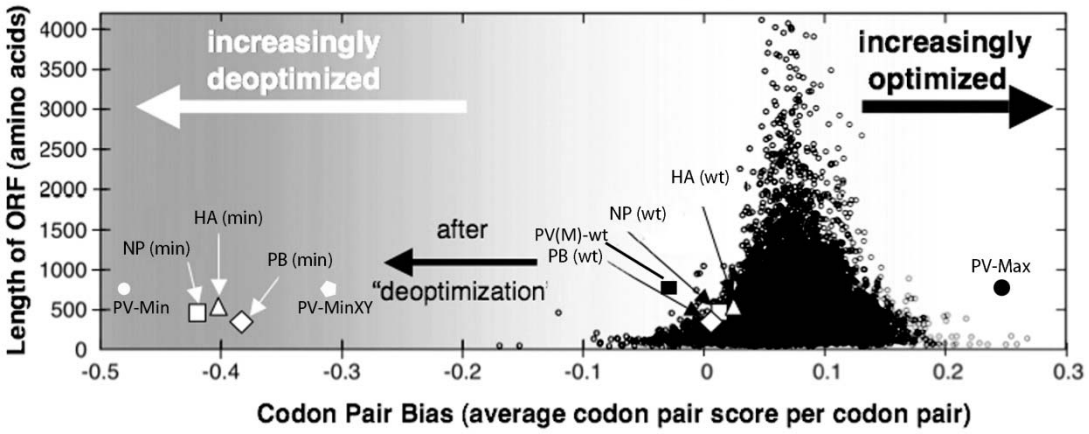


Figure 9. Schematic depiction of a computer analysis of the CPB score of 14,795 annotated human genes, of the poliovirus genome variants and of the influenza genome variants. CPB of the poliovirus P1 region, and HA, NP and PB genes in influenza virus were altered. CPB of a gene reflects the average of codon pair scores of all codon pairs in the gene; predominant use of underrepresented codon pairs yields negative codon bias values. Adapted from (Coleman *et al.*, 2008; Mueller *et al.*, 2010).

SAVE was used previously to rationally recode poliovirus and influenza virus genomes in order to include infrequently used codon pairs, and resulting in the generation of highly attenuated viruses (Coleman *et al.*, 2008; Mueller *et al.*, 2010). When challenged with wild-type virus, mice previously immunized with recoded attenuated viruses, displayed a protective immunity. Thus, recoded viruses were efficiently used in mice as vaccines (Coleman *et al.*, 2008; Mueller *et al.*, 2010). The mechanism of attenuation remains unclear, though it has been suggested that translation might be affected (Coleman *et al.*, 2008; Mueller *et al.*, 2010). Since codon-pair deoptimization is the result of tens, hundreds, or even thousands of nucleotide substitutions, reversion to virulence is unlikely to occur. Attenuation may be fine-tuned by adjusting the extent of codon-pair deoptimization (Coleman et al., 2008).

In addition to explore virus attenuation, synonymous codon pair reoptimization has been used to overcome adeno-associated virus (AAV) Rep gene's inhibitory effects on adenovirus replication (Sitaraman *et al.*, 2011), to identify unknown functional RNA elements in poliovirus coding sequences (Song *et al.*, 2012) and to study poliovirus evolvability and pathogenesis (Lauring *et al.*, 2012).

HYPOTHESIS AND OBJECTIVES

Human Immunodeficiency Virus type 1 (HIV-1) genome has a different codon usage when compared to humans. This difference results in a higher presence of A-nucleotides in the third codon position. SAVE technology allows to recode and synthetize a new viral genome without affecting the amino acid sequence. We aimed here to explore how evolution has chosen the nucleotide HIV-1 sequence existing nowadays, and to determine to what extent the actual encoding contributes to virus replication capacity, evolvability and pathogenesis. To this end, we used SAVE technology to rationally recode HIV-1 codon-pairs in protease and gag genes.

Objective 1: To study the impact of codon-pair usage in the HIV-1 replication capacity.

Objective 2: To study the genetic/phenotypic stability in cell culture of recoded viruses.

Objective 3: To determine the mechanism involved in attenuation by codon pair bias.

Objective 4: To measure the impact of codon-pair usage in RNA structures of protease recoded fragments and gag recoded fragments.

MATERIALS AND METHODS

1. CELL LINES

MT-4 cells (a lymphocytic laboratory-stable cell line) were obtained from the NIH AIDS Research and Reference Reagent Program (https://www.aidsreagent.org/index.cfm). HEK 293T cells were obtained from the American Type Culture Collection (ATCC). MT-4 cells were grown in Roswell Park Memorial Institute (RPMI) 1640 L-Glutamine medium supplemented with 10% of heat-inactivated fetal bovine serum (FBS) (Gibco). 293T cells were grown with Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% of heat-inactivated FBS.

Peripherial blood mononuclear cells (PBMCs) were obtained from healthy donors from the Banc de Sang i Teixits (BST, Hospital Vall d'Hebrón, Barcelona, Spain) and approved by the Ethical Committee of the center. Purification was performed using Lymphoprep^{TM}, a density gradient, to isolate the desired fraction of PBMCs. PBMCs were maintained with in RPMI 1640 L-Glutamine medium supplemented with 20% of heat inactivated FBS. Lastly, using a micro-beads system of a known concentration (Perfect Count MycrospheresTM, Cytognos), PBMCs were quantified following manufacturer's instructions. PBMC from 4 different donors were mixed equally and resuspended at $20 \cdot 10^6$ PBMC/ml in heat-inactivated FBS containing 10% dimethyl sulfoxide (DMSO). 1 ml aliquots were frozen and stored in liquid nitrogen until use. In order to promote HIV-1 replication in T-cells, PMCS were thaw in 20ml of FBS, centrifuged and resuspended at $1x10^6$ PBMC/ml. Finally, they were stimulated with 5µg/ml of PHA-P (Sigma-Aldrich) and 10U/ml of IL-2 (Roche), and maintained in RPMI 1640 L-Glutamine medium supplemented with 20% of heat inactivated FBS for 3 days at 37ºC and 5%CO^2 at a final concentration of $1 x10^6$ cells/ml.

2. GENERATION OF SYNTHETIC HIV-1.

- *PROTEASE*

All synthetic HIV-1 used in this study are based on the HXB2 strain (http://www.hiv.lanl.gov). Protease gene was divided in 3 different moieties, A, B and C. Synthetic HIV-1 protease was generated by combining overlapping synthetic DNA oligonucleotides corresponding to that moieties (Eurogentec, Madrid, Spain) by PCR, using an overlapping protocol (Figure 10). The hybrid constructs were assembled in three different steps (Table 1). First, 1,5pmol of oligonucleotides A (120bp, positions 2241-2361 in HXB2) and B (130bp, positions 2342-2472 in HXB2) (see Figure 10), and 1,5pmol of oligonucleotide C (117 bp, positions 2450-2567) and T7XhoI (5'-TAA TAC GAC TCA CTA TAG GGC GAA TTG GGT ACC GGG CCC CCC CTC

GAG TCA AAG GCC ATC CAT TCC TGG C-3') primer were mixed and hybridized. A double step PCR with Taq High Fidelity (Invitrogen, Life Technologies) was performed with an initial round that contained all the reagents required in a standard PCR mix but without specific primers. This first step involved one cycle of denaturation at 94ºC for 5 minutes and 30 cycles of denaturation at 94ºC for 30 seconds, followed by 30 seconds at 45ºC and 30 seconds at 68ºC of annealing and extension, respectively, in a final volume of 50µl. To proceed to the second step, half of both PCRs were mixed, new Taq polymerase was added and a second round of PCR was performed with one cycle of denaturation at 94ºC for 5 minutes and 30 cycles of: denaturation at 94ºC for 30 seconds, followed by 30 seconds at 45ºC, 1minute at 68ºC of annealing and extension, and 7 minutes at 68ºC for final extension. Finally, 5ul of the PCR were used in a nested PCR (Taq Platinum HF) with T3proL (5'-AAT TAA CCC TCA CTA AAG GGA ACA AAA GCT GGA GCT CCA CCG CGG TGG CGG CCG CTC TAG AAC TAG TGG ATC CCC CGG GCT GCA GGA ATT CTT CCT TTA ACT TCC CTC AG-3') and T7 (5'-TAA TAC GAC TCA CTA TAG GG-3') primers in a final volume of 100µl and 35 cycles: one cycle of denaturation at 94ºC for 5 minutes and 35 cycles of denaturation at 94ºC for 30 seconds, followed by 30 seconds at 55ºC, 1 minute at 68ºC of annealing, and a final extension at 68ºC for 7 minutes. PCRs were purified with the QIAquick PCR Purification Kit (QIAGEN) and eluted in 35µl of milli-Q water. 200ng of purified PCR were digested with 1 unit of FastDigest EcoRI (ThermoFisher) at 37ºC 15 minutes and 1 unit of FastDigest XhoI. (ThermoFisher) at 37ºC 5 minutes following manufacter's instructions. In order to discard undesired fragments and enzyme traces, digested PCRs were run on an electrophoresis gel, subsequently purified with QIAquick Gel extraction kit (QIAGEN) and eluted in 35µl of milli-Q water.

These final constructs were ligated in lambda DNA (Uni-ZAP XR Vector Kit, Stratagene), packaged (Uni-ZAP XR Gigapack Cloning Kit, Stratagene), titered, and amplified according to standard procedures (Parera *et al.*, 2006). Individual phage lambda clones were obtained in XL1-Blue MRF' cells.

The presence of the desired mutations was determined by nucleotide sequencing of the encoded HIV-1 protease gene included in individual phage colonies. Briefly, 1µl of phage supernatant from individual colonies was PCR amplified and sequenced with the flanking oligonucleotides T3 (5'-ATT AAC CCT CAC TAA AGG GA-3') and T7 in a standard sequencing reaction procedure: 1 cycle of 95ºC for 5 min and 35 cycles of: 95ºC for 30 secs, 56ºC for 15 secs and 60ºC for 4 min (BigDye terminator v3.1 cycle sequencing kit (Applied Biosystems). Sequence alignment and editing were

performed with the Sequencer version 4.8 (GeneCodes) and the Genedoc v2.6 programs. Once the sequence was verified, the corresponding clones were PCR-amplified with oligonucleotides 5'-prot2 (5'-TCA GAG CAG ACC AGA GCC AAC AGC CCC A-3') and 3'-prot2 (5'-AAT GCT TTT ATT TTT TCT TCT GTC AAT GGC-3') primers (Table 1) in a standard PCR reaction with Taq Platinum.

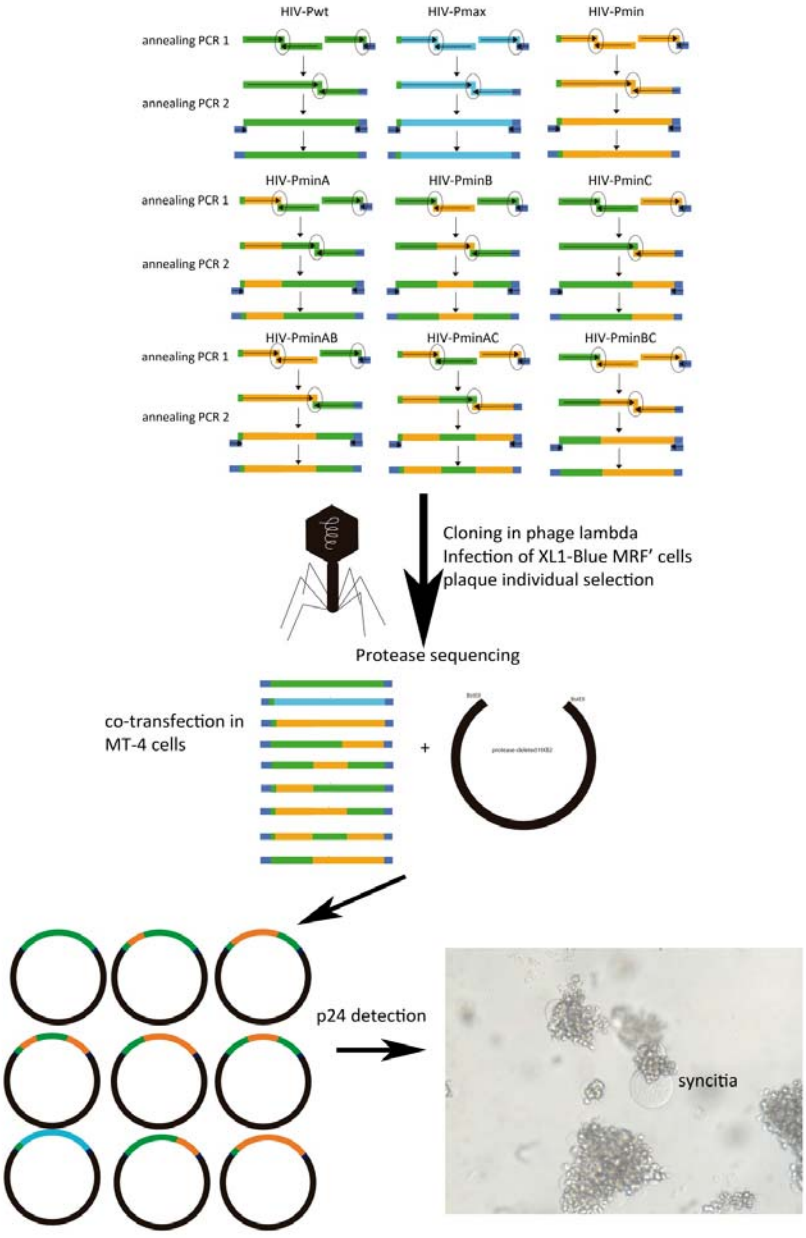


Figure 10. Schematic representation of CPB protease recoded viruses construction. Two annealing PCRs were performed for each construct in order to obtain the 9 final protease fragments. PCR constructs were cloned into phage lambda and individual clones were sequenced. Correct fragments were PCR-amplified and cotransfected in MT-4 cells with a protease-deleted HXB2 clone. Syncitia formation was observed in viable viruses at days or weeks post-transfection.

- *GAG*

Gag gene was divided in 4 different fragments, A, B, C and D. Synthetic HIV-1 gag constructs were generated by combining 8 wild-type and 8 deoptimized overlapping synthetic DNA oligonucleotides corresponding to the 4 different fragments (2 for each fragment) (Figure 11, Table 1). Primers 1-2, 3-4, 5-6 and 7-8 (Table 1) were combined in overlapping PCRs to obtain all the different A, B, C and D respectively constructions (Figure 11). On one hand, minimum or wild type fragments A (primers 1 and 2) and B (primers 3 and 4) and on the other hand minimum or wild type fragments C (primers 5 and 6) and D (primers 7 and 8) were overlapped in a two-step PCR. That involved an initial round containing all the reagents required in a standard PCR mix with the exception of specific primers with one cycle of denaturation at 94ºC for 5 min and 30 cycles of: 94ºC 30 secs, 45ºC 30 secs and 68ºC 1min in a final volume of 100µl. New Taq polymerase and 10pmol of each one of the corresponding primers (Table 2) were added and a second round of PCR was performed with one cycle of denaturation at 94ºC for 5 min and 30 cycles of: 94ºC 30 secs, 55ºC 30 secs and 68ºC 1min and a final extension step at 68ºC for 7 min. The constructs Gag-CD had to be PCR recombined with wild-type HXB2 protease in an overlapping PCR, in a similar way as described before. The correct sizes of the bands were checked with a 1% agarose gel. PCRs were purified with the QIAquick PCR Purification Kit (Qiagen) and sequenced with the suitable primers.

The different products were combined in order to obtain all the final gag constructs (Figure 11) in an another two-step PCR with a preliminary round without primers with one cycle of denaturation at 94ºC for 5 min and 30 cycles of: 94ºC 30 secs, 45ºC 30 secs and 68ºC 2min in a final volume of 100µl. Then, new Taq polymerase and 10pmol of each one of the corresponding primers were added and a final round of PCR was performed with one cycle of denaturation at 94ºC for 5 min and 30 cycles of: 94ºC 30 secs, 55ºC 30 secs and 68ºC 2min and a final extension step at 68ºC for 7 min. The correct sizes of the bands were checked with a 1% agarose gel. PCRs were purified with the QIAquick PCR Purification Kit (QIAGEN) and sequenced with the suitable primers (Table 2).

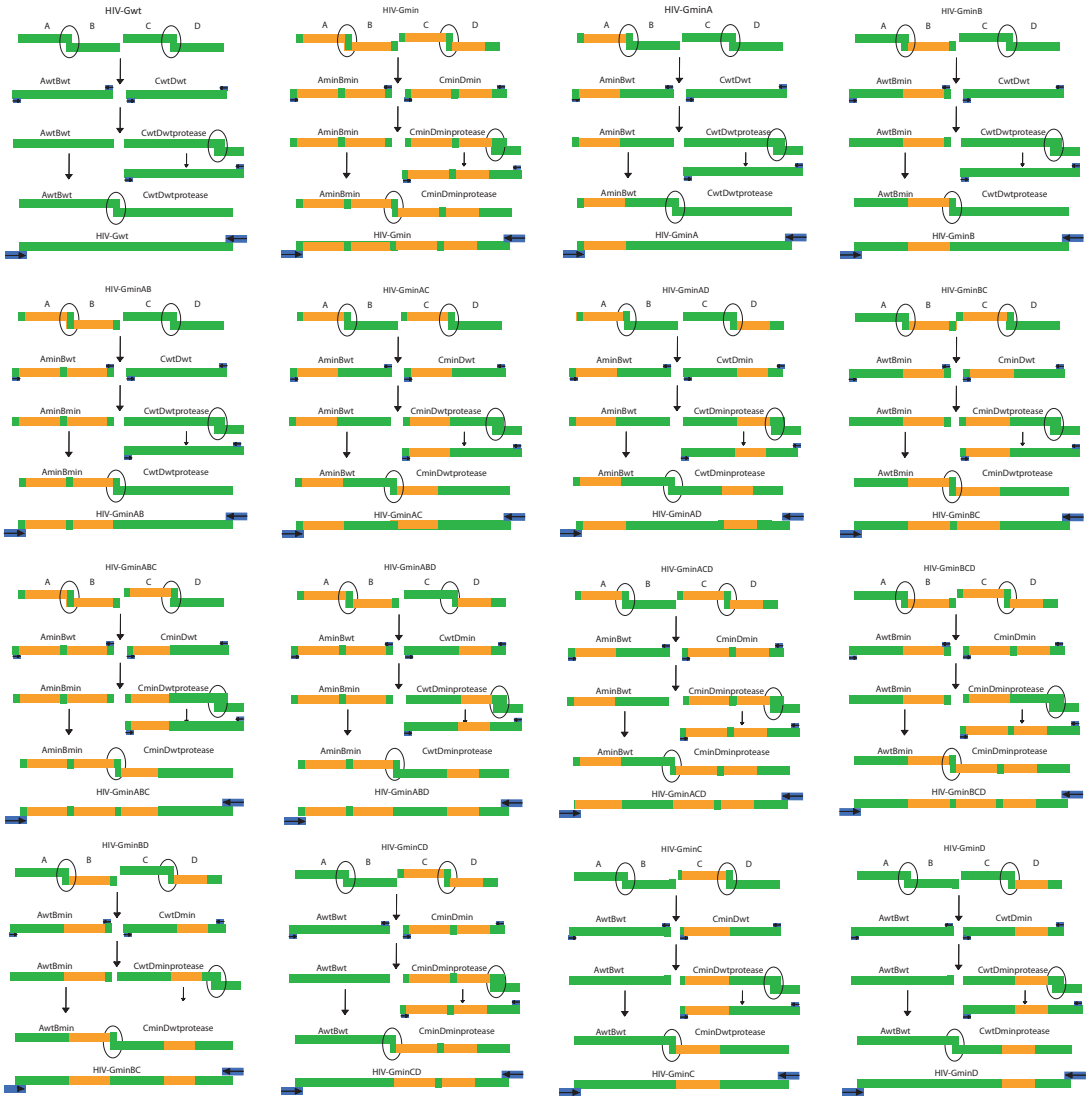


Figure 11. Annealing PCRs of all different gag constructs. In all constructs, a first overlapping PCR was performed to generate fragments AB and fragments CD. CD fragments were PCR-overlapped with wild-type protease. AB and CD-protease fragments were overlapped in a final PCR. Green boxes represent wild-type sequences, orange boxes represent deoptimized sequences and blue boxes represent primers.

Primer	Sequence (5' → 3')	Orientation	HXB2 pos.	CDS	Method
Gagmin1	GCTAGAAGGAGAGAGATGGGGGCTAGGGCTAGCGTACTGTCCGGTGGCGAACTCGATCGGTGGGAGAAGATACGCTTACGCCCCGGTGGTAAGAAAAAAT ATAAGCTTAAGCATATCGTTTGGGCTAGCCGCGAACTCGAACGGTTTGCCGTAAATCCGGGACTGCTCGAAACTAGCGAGGGATGTAGGCAAATCCTCGG	F	775-974	gag	overlapping PCR
Gagmin2	GTGTCAGCTGCCGCTTGTTGCGCTTTTTTTTTCGACTTATTCTGTTCTTCCTCGATTTTATCGAGCGCTTCCTTAGTGTCTTTGATTTCTATGCGTTGGT GTACGCAGTATAGTGTCGCTACGGTGTTATACAGACTGCGTAATTCCTCGCTACCCGTTTGCAGACTCGGTTGCAATTGTCCGAGGATTTGCCTACATCC	R	955-1154	gag	overlapping PCR
Gagmin3	AGGACACAGCAACCAGGTATCGCAAAATTACCCGATAGTGCAAAATATCCAGGGGCAAATGGTGCATCAAGCGATTAGCCCTAGGACACTCAACGCTTGG GTTAAGGTAGTCGAAGAAAAGGCATTCAGTCCGGAGGTAATACCGATGTTCTCCGCTCTTAGCGAGGGAGCGACACCGCAAGATCTGAATACGATGCTCA	F	1155-1354	gag	overlapping PCR
Gagmin4	ATCCTATTTGTTCTTGCAGCGTACTCGTAGTTCCGGCGATATCGCTACCCCTAGGTTCGCGCATTTGACCGGGAGCGATCGGTCCAGCGTGTACGGGATG CACCCTATCCCATTCTGCGGCTTCCTCGTTAATCGTTTCCTTAAGCATTTGCATCGCCGCTTGATGCCCTCCGACCGTATTGAGCATCGTATTCAGATCT	R	1373-1572	gag	overlapping PCR
Gagmin5	GGATGACAAACAATCCGCCGATACCCGTAGGCGAAATATACAAACGCTGGATCATACTGGGACTGAATAAGATCGTTCGTATGTATAGCCCTACGTCGAT CCTCGATATACGCCAGGGGCCGAAAGAACCGTTTCGCGATTACGTCGATCGGTTTTATAAGACACTCCGCGCTGAACAGGCTAGCCAGGAAGTTAAGAAT	F	1573-1772	gag	overlapping PCR
Gagmin6	GGTAGCTGAATTCGTTACCTGGCTCATCGCTTCTGCGAGAACGCGCGCTTTATGCCCCGGTCCTCCGACCCCCTGGCATGCCGTCATCATCTCTTCGAGT GTCGCAGCGGGTCCGAGCGCTTTCAGTATTGTTTTGCAATCCGGATTAGCGTTTTGCACTAACAGCGTTTCCGTCATCCAATTCTTAACTTCCTGGCTAG	R	1753-1952	gag	overlapping PCR
Gagmin7	ATAATGATGCAACGCGGTAACTTTCGGAATCAGCGTAAGATCGTTAAATGCTTTAACTGCGGAAAAGAGGGGCATACCGCTCGTAACTGTAGGGCGCCAC GTAAGAAGGGATGTTGGAAATGCGGTAAGGAGGGACACCAAATGAAAGATTGCACCGAACGCCAAGCGAATTTTTTAGGGAAGATCTGGCCTTCCTACAA	F	1953-2152	gag	overlapping PCR
Gagmin8	TTATTGTGACGAGGGGTCGTTGCCAAAGAGTGACCTGAGGGAAGTTAAAGGATACAGTTCCTTGTCTATCGGCTCCTGCTTCTGAGGGGGAGTTGTTGTC TCTACCCCAGACCTGAAGCTCTCTTCTGGTGGGGCTGTTGGCTCTGGTCTGCTCTGAAGAAAATTCCCTGGCCTTCCCTTGTAGGAAGGCCAGATCTT	R	2133-2330	gag	overlapping PCR
Gagwt1	GCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAAT ATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGG	F	775-974	gag	overlapping PCR
Gagwt2	GTGTCAGCTGCTGCTTGCTGTGCTTTTTTCTTACTTTTGTTTTGCTCTTCCTCTATCTTGTCTAAAGCTTCCTTGGTGTCTTTTATCTCTATCCTTTGAT GCACACAATAGAGGGTTGCTACTGTATTATATAATGATCTAAGTTCTTCTGATCCTGTCTGAAGGGATGGTTGTAGCTGTCCCAGTATTTGTCTACAGCC	R	955-1154	gag	overlapping PCR
Gagwt3	AGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATCCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG GTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAA	F	1155-1354	gag	overlapping PCR
Gagwt4	ATCCTATTTGTTCCTGAAGGGTACTAGTAGTTCCTGCTATGTCACTTCCCCTTGGTTCTCTCATCTGGCCTGGTGCAATAGGCCCTGCATGCACTGGATG CACTCTATCCCATTCTGCAGCTTCCTCATTGATGGTCTCTTTTAACATTTGCATGGCTGCTTGATGTCCCCCCACTGTGTTTAGCATGGTGTTTAAATCT	R	1373-1572	gag	overlapping PCR
Gagwt5	GGATGACAAATAATCCACCTATCCCAGTAGGAGAAATTTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCAGCAT TCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAAT	F	1573-1772	gag	overlapping PCR
Gagwt6	GGTAGCTGAATTTGTTACTTGGCTCATTGCTTCAGCCAAAACTCTTGCCTTATGGCCGGGTCCTCCTACTCCCTGACATGCTGTCATCATTTCTTCTAGT GTAGCCGCTGGTCCCAATGCTTTTAAAATAGTCTTACAATCTGGGTTCGCATTTTGGACCAACAAGGTTTCTGTCATCCAATTTTTTACCTCCTGTGAAG	R	1753-1952	gag	overlapping PCR
Gagwt7	ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGATTGTTAAGTGTTTCAATTGTGGCAAAGAAGGGCACACAGCCAGAAATTGCAGGGCCCCTA GGAAAAAGGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGACAGGCTAATTTTTTAGGGAAGATCTGGCCTTCCTACAA	F	1953-2152	gag	overlapping PCR
Gagwt8	TTATTGTGACGAGGGGTCGTTGCCAAAGAGTGACCTGAGGGAAGTTAAAGGATACAGTTCCTTGTCTATCGGCTCCTGCTTCTGAGGGGGAGTTGTTGTC TCTACCCCAGACCTGAAGCTCTCTTCTGGTGGGGCTGTTGGCTCTGGTCTGCTCTGAAGAAAATTCCCTGGCCTTCCCTTGTAGGAAGGCCAGATCTT	R	2133-2330	gag	overlapping PCR
A-Pwt	TCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAATAAAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATG	F	2241-2361	protease	overlapping PCR
B-Pwt	CCTATAGCTTTATGTCCACAGATTTCTATGAGTATCTGATCATACTGTCTTACTTTGATAAAACCTCCAATTCCCCCTATCATTTTTGGTTTCCATCTTCCTGGCAAACTCATTTCTTCTAATACTGTAT	R	2342-2472	protease	overlapping PCR
C-Pwt	TGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAAATTTTCCCATTAGCCCTATTGAG	F	2450-2567	protease	overlapping PCR
A-Pmax	TCCTTTAACTTCCCTCAGATCACTTTATGGCAAAGACCATTAGTTACCATCAAGATAGGTGGTCAACTAAAAGAAGCTTTATTAGATACAGGAGCAGATGATACTGTTTTAGAAGAAATG	F	2241-2361	protease	overlapping PCR
B-Pmax	CCTATTGCTTTATGTCCACATATTTCTATTAATATTTGATCATATTGTCTTACTTTTATAAATCCTCCAATGCCCCCTATCATTTTTGGCTTCCAGCGGCCCGGCAGCGACATTTCTTCTAAAACAGTAT	R	2342-2472	protease	overlapping PCR
C-Pmax	TGTGGACATAAAGCAATAGGTACTGTATTAGTAGGACCCACGCCGGTCAACATCATAGGAAGAAATTTATTAACACAAATAGGATGTACTTTAAATTTTCCCATTAGCCCTATTGAG	F	2450-2567	protease	overlapping PCR
A-Pmin	TCCTTTAACTTCCCTCAGATCACACTTTGGCAACGCCCTCTCGTTACGATTAAGATCGGCGGACAGTTGAAGGAAGCGTTACTCGATACCGGAGCCGATGATACCGTACTCGAAGAGATG	F	2241-2361	protease	overlapping PCR
B-Pmin	CCTATCGCTTTGTGTCCGCAAATCTCTATCAGAATCTGATCGTACTGCCTAACCTTTATGAACCCTCCGATCCCTCCGATCATTTTCGGTTTCCACCTACCGGGTAAGCTCATCTCTTCGAGTACGGTAT	R	2342-2472	protease	overlapping PCR
C-Pmin	TGCGGACACAAAGCGATAGGCACAGTGTTAGTCGGACCGACACCGGTAAACATAATCGGACGGAATCTGTTAACGCAAATCGGATGCACTCTCAATTTCCCCATTAGCCCTATTGAG	F	2450-2567	protease	overlapping PCR

Table 1. Ultramer oligonucleotides used to construct gag and protease recoded regions. F: Forward primer; R: Reverse primer; HXB2 pos: primer position relative to HXB2 genome; CDS: Coding DNA Sequence.

- *VIRUS GENERATION*

Protease and Gag recombinant infectious viruses were generated as previously described (Capel *et al.*, 2012; Kisic *et al.*, 2011). Briefly, 100 ng of full-length protease or 500 ng of gag purified PCR products were cotransfected into MT-4 cells with 1 µg of protease-deleted HXB2 clone (Maschera *et al.*, 1995) or gag-protease-deleted NL4-3 clone (Miura *et al.*, 2008), respectively, previously linearized with 1 unit of BstEII (New England Biolabs). Cell culture supernatants were harvested on days 3, 5, and 7 after transfection when the concentration of HIV-1 p24 antigen (Genscreen HIV-1 Ag assay, Bio-Rad) surpassed 500 ng/ml. If p24 antigen was not detected after 7 days of culture, five blind passages with a refeeding of the cultures with fresh medium and new MT-4 cells were performed to recover viable virus. After five blind passages without p24 antigen detection in the cultured supernatant, the construct was considered non-viable in MT-4 cell culture. Virus titration was performed in MT-4 cells, and values were expressed as tissue culture dose for 50% infectivity (TCID50), as previously described (Pauwels *et al.*, 1988).

3. REPLICATION CAPACITY ASSAYS

Viral replication kinetics were performed by infecting either 1×10^6 of MT-4 cells with 200 TCID50 (MOI 0.0002) or 2×10^6 PBMCs (mixed from four healthy donors), previously stimulated 3 days before with 5µg/ml of PHA (Sigma-Aldrich) and 10 U/ml of IL2 with 2000 TCID50 of each viral stock (MOI 0.001). After 4 hours of incubation at 37ºC and 5% CO2, cells were washed twice with phosphate buffered saline (PBS) and resuspended in RPMI medium supplemented with 10% FBS or 20% FBS plus 10U/ml of IL2 for MT-4 cells or PMBCs, respectively. Every 24h, 200ul of supernatant was removed for p24 antigen quantification. Viral replication was quantified by measuring HIV-1 capsid p24 antigen production (Genscreen HIV-1 Ag assay, Bio-Rad) in the culture supernatant during 5 days (for MT-4) or 7 days (for PBMCs). Growth kinetics were analyzed by fitting a linear model to the log-transformed p24 data during the exponential growth phase by maximum likelihood methods as we previously described (Betancor *et al.*, 2010).

4. SERIAL PASSAGES EXPERIMENTS

Phenotypic stability of all the attenuated viruses was tested by serially propagating them in MT-4 cells. Briefly, viruses were added at a MOI of 0.002 to $1x10^6$ MT-4 cells. MT-4 cells were maintained in RPMI medium supplemented with 10% of FBS. After 3-

4 days, 200ul of the culture were transferred to $1x10^6$ fresh MT-4 cells. Virus production was monitored by measuring p24 antigen.

Viral genomic RNA from 140µl of culture supernatant was purified using the QIAamp Viral RNA Kit (QIAGEN). After the viral RNA was isolated, 5 µl of resuspended RNA was reverse transcribed and PCR amplified using the SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies, Invitrogen) and 10 pmol of oligonucleotides 5'prot 2 and 3'prot 2, as we previously described (Capel *et al.*, 2012; Fernandez *et al.*, 2007). The amplified DNA was cloned into the pGEM-T Easy Vector System I (Promega) and transformed in OmniMAX™ 2 T1 Phage-Resistant (T1R) *E. coli* strain. DNA inserts were sequenced using the flanking oligonucleotides SP6 (5'- ATT TAG GTG ACA CTA TAG-3') and T7.

5. GROWTH COMPETITION ASSAYS

Dual competition experiments were performed by infecting 1×10^6 MT-4 cells with each pair of competing viruses mixed at different ratios with an overall $TCID_{50}$ of 1000 (MOI of 0.001). HIV-PminB and HIV-PminC were independently competed with the HXB2 wild-type virus. After 4 h at 37°C, cells were washed twice with PBS, resuspended in 5 ml of medium, and cultured in six-well plates. Culture supernatants were collected twice weekly, and 100µl of supernatant was used to infect fresh cells. Viral RNA was extracted and amplified by one-step RT-PCR with oligonucleotides 5'prot 2 and 3'prot 2. Ratios of the two competing variants were estimated on days 0, 7, 14, and 21, based on the relative peak heights of all the different mutations in electropherograms obtained by bidirectional sequencing of the protease coding region.

6. SINGLE-CYCLE INFECTIVITY ASSAY

Recoded HIV-1 proteases and wild-type HXB2 protease were cloned into the GFP reporter HIV-1 infectious clone pNL4-3-deltaE-EGFP (NIH AIDS Research and Reference Reagent Program). Protease sequences were PCR amplified with oligonuclotides Apa1988 (5'- ACA TAG CCA AAA ATT GCA GGG CCC CTA G-3') and Sbf2838 (5'-CTG ATT TTT TCT GTT TTA ACC CTG CAG GAT G-3'), digested with ApaI and SbfI (New England Biolabs), and cloned into pNL4-3-deltaE-EGFP between the ApaI and SbfI sites. Recombinant pNL4-3-deltaE-EGFP plasmids (4µg) were used to transfect 8×10^5 293T cells in the presence of Lipofectamine 2000 (Invitrogen). The

relative replication capacity of the virus was determined by measuring the amount of p24 antigen produced 72 h after transfection. Replication capacity is expressed as the percentage of p24 antigen produced by the vectors containing each of the derived protease sequences compared to the p24 antigen from the vector containing the HIV-1 HXB2 protease reference sequence (100%). Replication capacity measurements were normalized for differences in transfection efficiencies by monitoring the GFP activity (by flow cytometry) generated in transfected cells.

7. REAL- TIME QUANTITATIVE RT-PCR ANALYSIS OF HIV-1 P24 AND GFP GENES.

Real-Time quantitative RT-PCR analysis was used to quantify HIV-1 gag transcripts. A total of 8×10^5 293T cells were transfected with 4µg of recombinant pNL4-3-deltaE-EGFP plasmid. Total cell RNA was isolated (High Pure RNA Isolation Kit, Roche) 72 h post-transfection. Reverse transcription was performed with a SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies, Invitrogen) in 20µl reaction volume containing 1µg of total RNA. Real-time quantitative RT-PCR analysis of p24 was carried out using a TaqMan Universal Master Mix (Applied Biosystems, Spain) and the oligonucleotides p24-Fw (5'-AAG GCC AGG GAA TTT TCT TC-3') and p24_Rv (5'-TAT TGT GAC GAG GGG TCG TT-3') and the probe p24 (FAM 5'-AAG CAG GAG CCG ATA GAC AA-3' TAMRA). Real-time quantitative RT-PCR analysis of GFP was made using the primers rightGFP (5'-CCA TGC CGA GAG TGA TCC-3') and leftGFP (5'-CGA CAA CCA CTA CCT GAG CA-3') and the GFP probe (FAM 5'- ACA TGG TCC TGC TGG AGT TC-3' TAMRA). The endogenous gene GAPDH (Life Technologies, Applied Biosystems) was used as control. Relative quantification of both mRNAs was calculated as described elsewhere (Livak & Schmittgen, 2001).

8. HIV-1 DRUG SUSCEPTIBILITY TESTS.

Reverse transcriptase inhibitors azidothymidine (AZT) and Tenofovir Disoproxil Fumarate (TNF-DF) and protease inhibitors Darunavir (DRV), Atazanavir (ATV), Amprenavir (APV) and Nelfinavir (NFV) were obtained from the NIH AIDS Research and Reference Reagent Program. After virus propagation and titration, HIV-1 drug susceptibility data were obtained in MT-4 cells as previously described (Betancor et al. 2010). Briefly, HIV-1 drug susceptibility data were obtained after infecting 30 000 MT-4 cells at a MOI of 0.003, and exposing the HIV-1-infected cultures to various concentrations of each drug (five-fold dilutions). After MT-4 cells had been allowed to proliferate for 5 days, the number of viable cells was determined using a tetrazolium-based colorimetric method (Pannecouque *et al.*, 2008).

9. ENC (Effective Number of Codons) and RNA stability

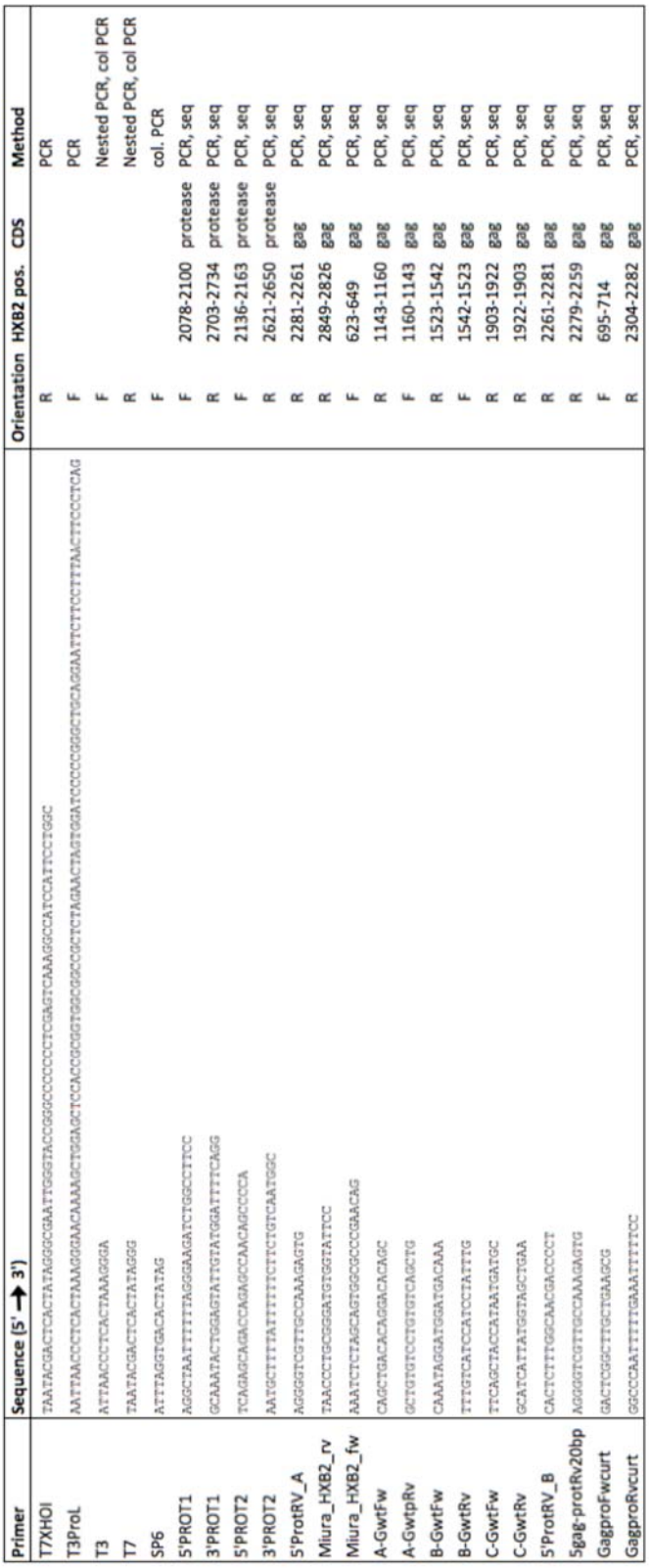
To control the folding free energy of the RNA of the recoded constructs, and confirm that strong secondary structures will not affect translation efficiency, RNA folding energies were determined with the program mFold, as previously described (Coleman *et al.*, 2008). That prevents any large changes in secondary structure and the creation of any new regions, such as stable hairpins or stem loops and assures that the free energy of the RNA of our new constructs is planned to be similar to the free energy of the input sequence. Briefly, 100-base long segments of all our constructs were analyzed with mFold (http://mfold.rna.albany.edu/?q=mfold), with an 80 bases overlap with each other. It has been described that any segments that had a binding energy lower than -30kcal/mol would produce random synonymous substitutions at C-G binding locations. ENC and GC3s were calculated by the use of the Codon W program (http://mobyle.pasteur.fr).

10. Statistical Analysis

All the statistics tests were performed using GraphPad Prism version 4.00 for Windows (San Diego, CA, USA) and a p-value of 0.05 was considered to be significant.

11. Nucleotide sequence accession numbers

The sequences of synthetically recoded HIV-1 variants, have been deposited in the GenBank database [accession numbers JX961631(HIV-Pmin, JX961632 (HIV-Pmax) and JX961633 (HIV-Gmin)].



Primer	Sequence (5' → 3')	Orientation	HXB2 pos.	CDS	Method
T7XHOI	TAATACGACTCACTATAGGGCGAATTGGGGTACCGGGCCCCCCCTCGAGTCAAAGGCCATCCATCCCTGGC	R			PCR
T3ProL	AATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCTTCCTTTAACTTCCCTCAG	F			PCR
T3	ATTAACCCTCACTAAAGGGA	F			Nested PCR, col PCR
T7	TAATACGACTCACTATAGGG	R			Nested PCR, col PCR
SP6	ATTTAGGTGACACTATAG	F			col. PCR
5'PROT1	AGGCTAATTTTTTAGGGAAGATCTGGCCTTCC	F	2078-2100	protease	PCR, seq
3'PROT1	GCAAATACTGGAGTATTGTATGGATTTTTCAGG	R	2703-2734	protease	PCR, seq
5'PROT2	TCAGAGCAGACCAGAGCCAACAGCCCCA	F	2136-2163	protease	PCR, seq
3'PROT2	AATGCTTTTATTTTTTCTCTGTCAATGGC	R	2621-2650	protease	PCR, seq
5'ProtRV_A	AGGGGTCGTTGCCAAAGAGTG	R	2281-2261	gag	PCR, seq
Miura_HXB2_rv	TAACCCTGCGGGATGTGGTATTCC	R	2849-2826	gag	PCR, seq
Miura_HXB2_fw	AAATCTCTAGCAGTGGCGCCCGAACAG	F	623-649	gag	PCR, seq
A-GwtFw	CAGCTGACACAGGACACAGC	R	1143-1160	gag	PCR, seq
A-GwtpRv	GCTGTGTCCTGTGTCAGCTG	F	1160-1143	gag	PCR, seq
B-GwtFw	CAAATAGGGATGGATGACAAA	R	1523-1542	gag	PCR, seq
B-GwtRv	TTTGTCATCCATCCTATTTG	F	1542-1523	gag	PCR, seq
C-GwtFw	TTCAGCTACCATAATGATGC	R	1903-1922	gag	PCR, seq
C-GwtRv	GCATCATTATGGTAGCTGAA	R	1922-1903	gag	PCR, seq
5'ProtRV_B	CACTCTTTGGCAACGACCCCT	R	2261-2281	gag	PCR, seq
Sgag-protRv20bp	AGGGGTCGTTGCCAAAGAGTG	R	2279-2259	gag	PCR, seq
GagproFwcurt	GACTGGACTTGCTGAAGCG	F	695-714	gag	PCR, seq
GagproRvcurt	GGCCCAATTTTTGAAATTTTCC	R	2304-2282	gag	PCR, seq

Table 2. Oligonucleotides used to sequence, PCR, nested PCR or amplify colony DNA. F: Forward primer; R: Reverse primer; HXB2 pos: primer position relative to HXB2 genome; CDS: Coding DNA Sequence.

CHAPTER 1. PROTEASE CPB RECODING

1. PROTEASE LARGE-SCALE MANIPULATION: CPB DESIGN

The HIV-1 protease gene was recoded using codon pairs that are under-represented or over-represented relative to the human genome by using a previously described algorithm (Coleman *et al.*, 2008). To obtain deoptimized protease constructs, protease wild-type codon-pairs were substituted by codon-pairs with a low CPS. Inversely, reoptimized protease construct was recalculated based in the highest scored codon-pairs. Specifically, only synonymous nucleotide substitutions were introduced while conserving amino-acid sequence. Thus, virus protein epitopes and structures were preserved. Recoded segments had the same amino acid sequence as the wild-type segments, but different pairwise arrangements of synonymous codons.

Protease overlaps with gag in the firsts 40 nucleotides in two different frames. In order to avoid non-viability due to the introduction of non-synonymous substitutions in gag, no mutations were designed in the first portion of protease. Otherwise, the stop codon of gag is abolished and, a new p6 protein is created containing 6 extra amino acids.

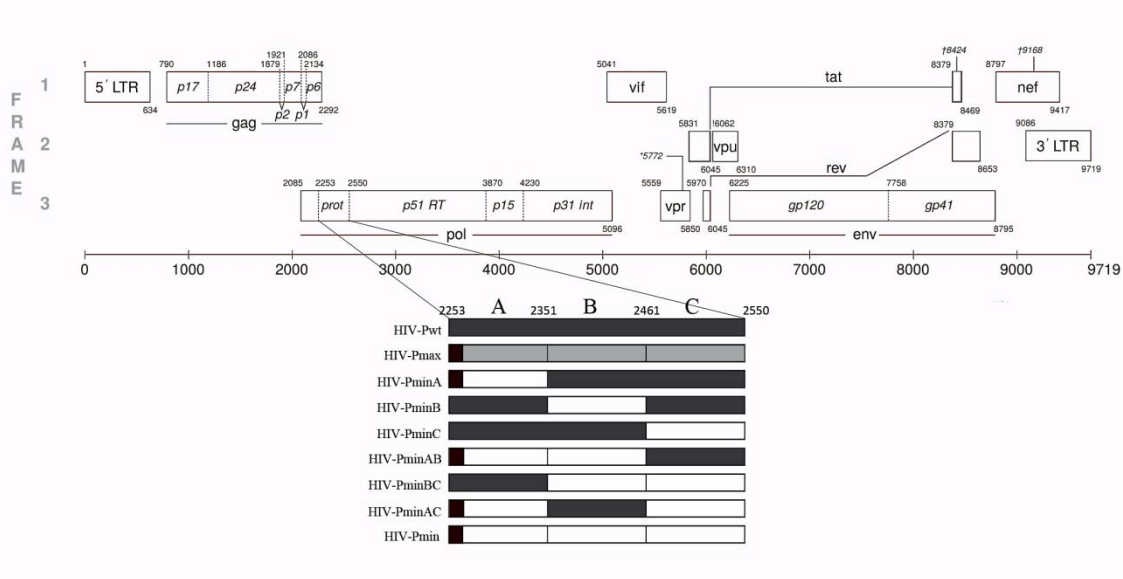


Figure 12. Structures of the various chimeric, partly synthetic protease HIV-1 constructs. Wild type (black), codon pair-deoptimized (White) or reoptimized (greys) sequences are shown. Nucleotide positions are indicated relative to HXB2 sequence. Adapted from (Kuiken *et al.*, 2012).

The first designs included an entire deoptimized protease gene, HIV-Pmin, as well as a entire reoptimized protease gene, HIV-Pmax (Figure 12). Of the 297 nucleotides that encode the HIV-1 protease, HIV-Pmin and HIV-Pmax contained respectively 57 and 38

silent synonymous substitutions (Table 3). HIV-Pmin had a CPB score much lower than the median of the human genes CPB score, -0.480, while HIV-Pmax had a CPB score of 0.254, much higher than the median of the human genes CPB score (Table 3). Both of them are extreme CPB scores, surpassing both the outer limit of naturally existing in human genome's CPB scores (Figure 13).

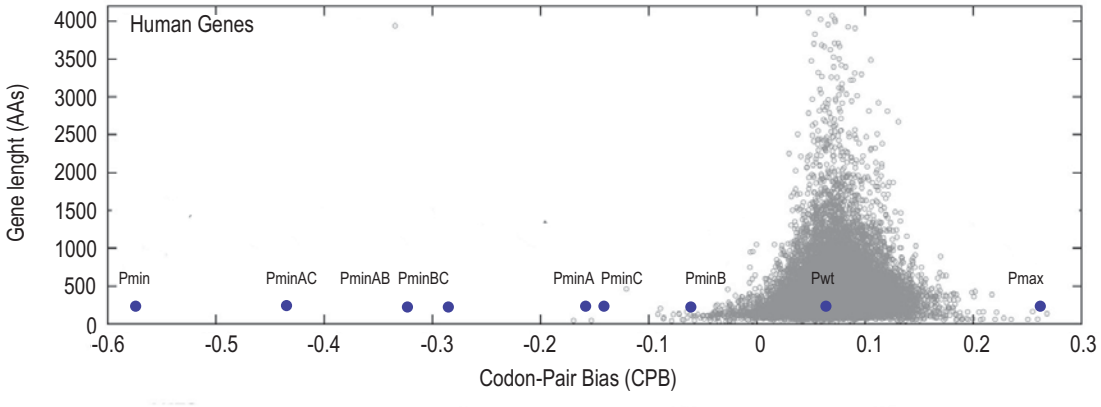


Figure 13. Calculated codon pair bias (CPB) score for all 14.795 annotated human genes. Each dot represents the calculated CPB score of a gene plotted against its amino acid length. Under-represented codon pairs yield negative scores. The protease constructs are represented by symbols; Pwt represent the wild-type HIV-1 protease (CPB=0.069). Adapted from (Coleman *et al.*, 2008).

One objective was to find an attenuated virus that allowed us to understand CPB mechanism in HIV-1. Based on Coleman's work, our hypothesis was that HIV-Pmin, with the lowest CPB score, would probably prevent the recovery of viable HIV-1. Thus, after these CPB score results, HIV-Pmin recoded protease was divided in three different moieties to progressively increase its CPB score (Figure 12). To obtain deoptimized viable virus, the number of under-represented pair codons present in HIV-Pmin was reduced (Figures 12 and 13) and new variants were produced. They had a different number of synonymous substitutions and different codon-pair scores (Table 3), ranging from -0.317 (HIV-PminBC and HIV-PminAC) to -0.094 (HIV-PminA and HIV-PminB). These new CPB scores were widely distributed along the human codon pair usage. HIV-PminAC, HIV-PminAB, HIV-PminBC, as HIV-Pmin did, surpassed the extreme human codon-pair usage while HIV-PminA, HIV-PminC and HIV-PminB were still present in the range of human codon-pair usage (Figure 13). The percentages of mutations compared to the wild-type HXB-2 protease varied from 5.05% (HIV-PminA) to 19.19% (HIV-Pmin).

Table 3. Calculated codon pair bias (CPB) score for each one of the constructs. In blue: viable viruses, in orange: non-viable viruses, even after five blind passages.

	CPB score	# mutations	% mutations
HIV-Pwt	0,069		
HIV-Pmax	0,254	38	12,79
HIV-PminA	-0,094	15	5,05
HIV-PminB	-0,094	22	7,41
HIV-PminC	-0,154	18	6,06
HIV-PminBC	-0,317	41	13,80
HIV-PminAB	-0,256	38	12,79
HIV-PminAC	-0,317	29	9,76
HIV-Pmin	-0,480	57	19,19

Kimura-2 parameters phylogenetic reconstruction was performed for all the nucleotide sequences of new protease constructs to determine the evolutionary relationships of the different variants and, to assess the degree of genetic differences that all synonymous mutations introduced could produce (Figure 14). As expected, HIV-Pmax clustered alone while all sequences containing deoptimized B moiety and deoptimized C moiety formed two independent clusters, suggesting those mutations could have relevance at the phylogenetic level.

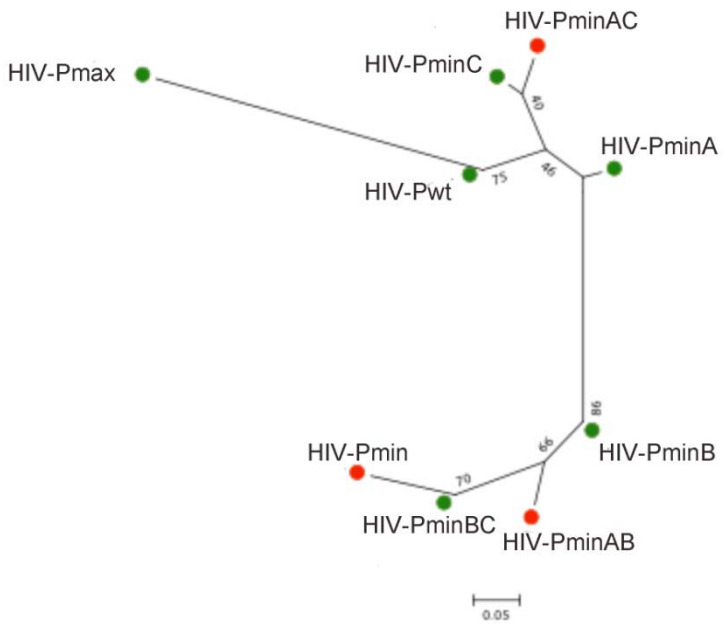


Figure 14. Evolutionary relationships of HIV-1 recoded proteases. The evolutionary history was inferred using the Neighbor-Joining method. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter =0.14). Evolutionary analyses were conducted in MEGA5.

Protease CPB score is the arithmetic mean of the individual codon-pair scores (CPS) of all codon pairs present in recoded protease sequences. While this index can be a significant measurement of protease deoptimization, individual CPS should also be considered as critical for the translation of the protease. A heat map was constructed to determine the degree of deoptimization or reoptimization that individual codon-pairs can introduce (Figure 15). HIV-Pmax construct had, along the sequence, over-represented codon-pairs with individual CPS above 0. All constructs containing deoptimized A fragment had low CPS, specifically the 6 codon-pairs (CP16, CP24, CP26, CP28, CP31 and CP33) with CPS ranging from -1.402 to -2.134. In contrast, in B and C deoptimized moieties only 4 CPS low scores were found: CP39 and CP40 (-1.413 and -1.357) in B deoptimized fragment, and CP67 and CP77 (-1.504 and -1.503) in C deoptimized fragment. Although deoptimized A fragment was the smallest one compared to others fragments (because the first region of protease was conserved), these 6 new low CPS found can be critical for virus viability.

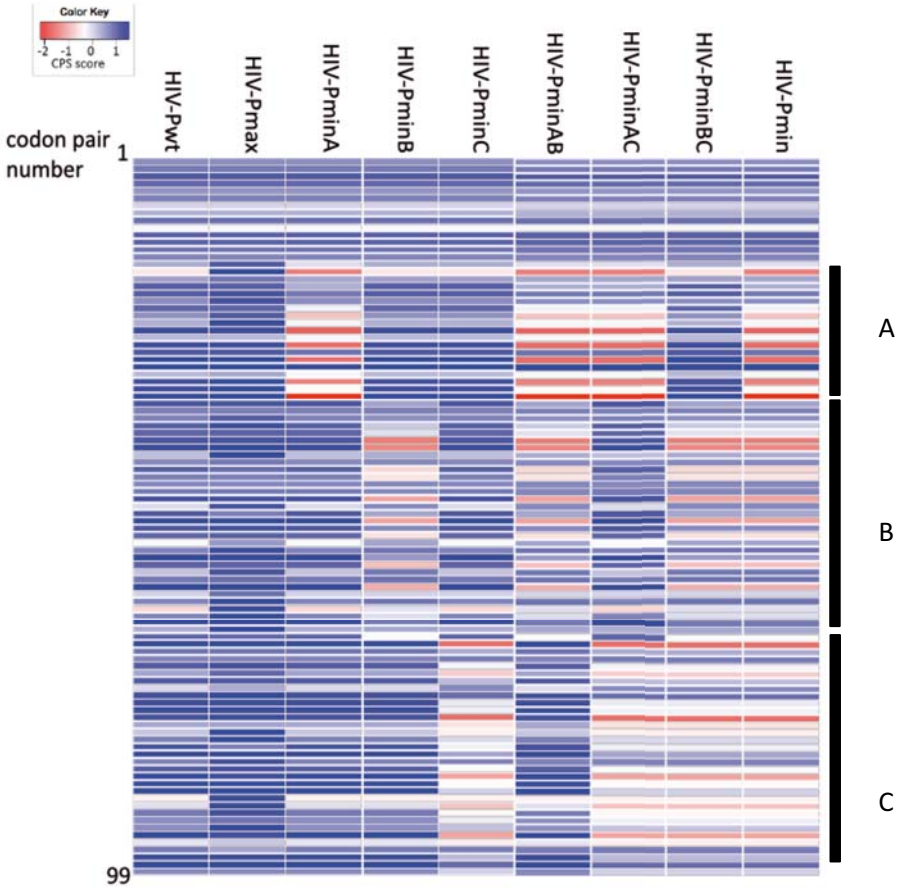


Figure 15. Protease recoded sequences Heat Map. A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors. Blue colors represent codon pairs with high CPS, red colors represent codon pairs with low CPS and white colors represent intermediate CPS (around 0). A, B and C regions are indicated by black boxes.

2. *EX-VIVO* REPLICATION CAPACITIES OF CPB VIRUSES

A. MT-4 AND PBMCS REPLICATION CAPACITY ASSAYS

All protease constructs were co-transfected by electroporation in MT-4 cells with an infectious HIV-1 DNA clone in which protease or gag had been deleted. HIV-Pwt, HIV-Pmax, HIV-PminA, HIV-PminB, HIV-PminC, and HIV-PminBC produced viable virus in MT-4 cells. No viable virus were obtained from HIV-PminAB, HIV-PminAC, HIV-Pmin constructs, even after five blind passages in MT-4 cells with re-feedings. All constructs with the exception of HIV-PminA had a similar TCID50/p24 antigen ratio compared to HIV-Pwt (Table 4). HIV-PminA had a decreased TCID50/p24 antigen ratio compared to HIV-Pwt (4.76%), the lowest infectivity-to-viral particle ratio observed in all variant viruses.

Table 4. HIV-1 contructs infectivity-to-virion particle ratios[a]

Virus	TCID50/ml[a]	p24 antigen (pg/ml)[b]	Relative Specific Infectivity[c]	TCID50/p24 antigen ratio to WT (%)
HIV-Pwt	6,26E+05	1,28E+06	1/2.03	100,00
HIV-Pmax	8,12E+05	1,64E+06	1/2.01	100,94
HIV-PminA	6,77E+04	2,90E+06	1/42.85	4,76
HIV-PminB	9,38E+05	1,42E+06	1/1.51	134,84
HIV-PminC	5,21E+05	1,06E+06	1/2.03	99,99
HIV-PminBC	5,03E+05	1,44E+06	1/2.86	71,20

[a]Infectivity was measured by determining TCID50, [b]virion particles were estimated by HIV-1 p24 antigen measurement from a single clarified supernatant (virus stock) of each HIV variant and [c]relative specific infectivity was calculated by dividing infectivity of viruses by p24 antigen/ml in culture.

Replication capacity assays *in vitro* were performed among viable viruses. Those experiments were conducted in MT-4 cells and in previously stimulated PBMCs deriving from a pool of 4 different donors. The use of primary cells such as PBMCs in *in vitro* experiments can more accurately reflect to physiological conditions. MT-4 cells were infected at a MOI of 0.0002, while PBMCs were infected at a MOI of 0.001, recollecting samples every day. Viral growth was monitorized by measuring the amount of capsid p24 in culture supernatants.

Although HIV-PminB, HIV-PminC, HIV-PminBC have 22, 18 and 42 synonymous codon-pair mutations, their viral replication capacity in MT-4 cells (Figure 16-A) and PBMCs (Figure 16-

B) was indistinguishable from the HIV-1 wild-type virus (p=0.2317, p=0.8742, and p=0.1984, respectively for MT-4 cells and, p=0.8539, p=0.2789 and p=0.1014, respectively for PBMCs). Similarly, HIV-Pmax, with 38 codon-pair synonymous mutations, was also viable and had a similar replication capacity in MT-4 cells compared to HIV-Pwt (Figure 16-A) and in PMBCs (Figure 16-B), suggesting that HIV-1 protease can still have a window of improvement by codon pair usage. In contrast, HIV-PminA, with only 18 synonymous substitutions, produced a virus that was highly attenuated in MT-4 cells ($p < 0.05$) and in PBMCs ($p < 0.01$) (Figure 16-A and Figure 16-B respectively). HIV-PminA had a replication capacity of 80% in MT-4 cells and of 40% in PBMCs (Figure 16-C and 16D respectively) compared to the wild-type virus.

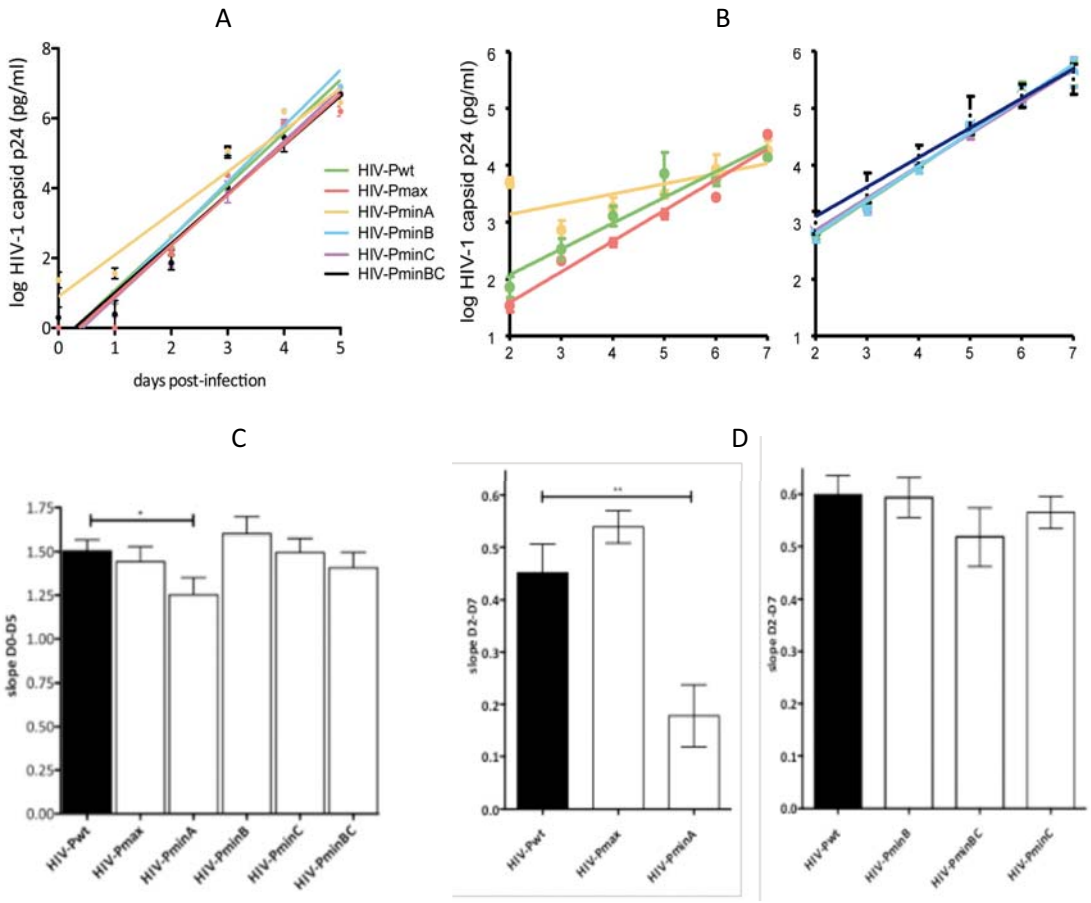


Figure 16. Replication capacity assays. MT-4 cells (A) and PBMCs (B). For each virus, viral replication capacity was measured in MT-4 cells (A) and PBMCs (B) by p24 antigen production. The slope of each virus containing viable CPB protease mutants after infection of MT-4 cells (C) and PBMCs (D) is shown and compared to wild-type HXB2 virus. P-values (student T test) are indicated in each plot. Values represent the means +/- SDs from three independent experiments.

We next tested the susceptibility of recoded HIV-1 variants to different antiretroviral compounds (AZT, TNF-DF, NFV, ATV, APV and DRV). The association between hyper-susceptibility to protease inhibitors and low replication capacity to virus isolated from chronically infected patients (Martinez-Picado *et al.*, 1999) promoted us to test the susceptibility of recoded HIV-1 variants to antiretroviral compounds. Antiviral activities to different protease inhibitors (PIs) and two nucleoside reverse transcriptase inhibitors (NRTIs) were analyzed with all recoded viruses.

Table 5. Susceptibility of HIV-1 constructs to antiviral compounds

	IC50 (nM)					
	NRTIs		protease inhibitors			
	AZT	**TNF-DF**	**NFV**	**DRV**	**ATV**	**APV**
HIV-Pwt	15,33 ± 3,61	27,32 ± 5,41	11,44 ± 4,68	1,69 ± 0,31	3,19 ± 0,74	14,43 ± 1,29
HIV-Pmax	29,02 ± 3,65 (1,89)	36,83 ± 7,31 (1,35)	12,63 ± 3,31 (1,1)	2,93 ± 0,07 (1,73)	3,57 ± 1(1,12)	15,11 ± 2,72 (1,05)
HIV-PminA	18,25 ± 9,5 (1,19)	34,18 ± 1,94 (1,25)	13,73 ± 7,23 (1,2)	2,12 ± 0,96 (1,25)	3,57 ± 0,53 (1,12)	18,47 ± 1,69 (1,28)
Hiv-PminB	14,07 ± 8,10 (0,92)	33,98 ± 3,36 (1,24)	12,01 ± 1,80 (1,05)	2,83 ± 0,20 (1,67)	3,02 ± 0,74 (0,95)	17,68 ± 2,03 (1,22)
HIV-PminC	16,08 ± 6,94 (1,05)	34,28 ± 4,76 (1,25)	12,03 ± 4,32 (1,05)	3,03 ± 0,51 (1,79)	3,62 ± 0,65 (1,13)	17,74 ± 2,29 (1,23)
HIV-PminBC	14,24 ± 5,02 (0,93)	28,81 ± 4,20 (1,05)	9,84 ± 2,71 (0,86)	2,77 ± 0,25 (1,64)	4,41 ± 0,49 (1,38)	16,38 ± 2,91 (1,13)

The IC50 values shown are means ± standard deviations of at least three tests, with each one performed six times. The fold increases in IC50 relative to the wild-type HXB2 virus control carrying the RT sequence of BH10 is shown in parentheses.

No significant differences in drug susceptibility were observed between wild-type virus and CPB recoded viruses (Table 5).

B. GROWTH COMPETITION ASSAYS

Given that HIV-PminB and HIV-PminC had similar replication capacities as HIV-Pwt in both cell models, growth competition experiments between HIV-Pwt and HIV-PminB and between HIV-Pwt and HIV-PminC were performed in MT-4 cells, in order to evaluate if any fitness cost deriving from synonymous mutations could exist. MT-4 cells were chosen due to their higher number of replication rounds (compared to PBMCs).

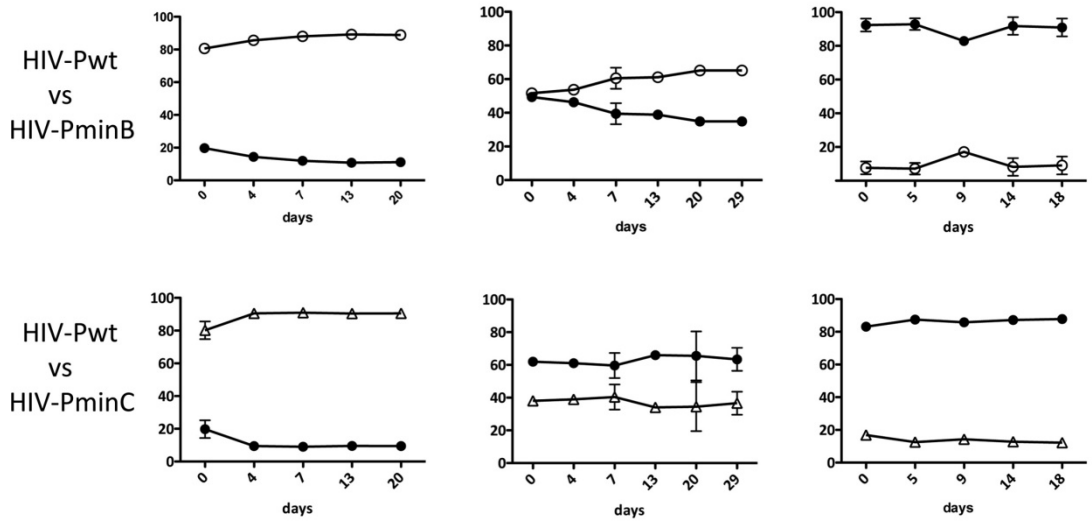


Figure 17. Growth competition assays between HIV-Pwt and HIV-PminB (upper row) or HIV-Pwt and HIV-PminC (lower row). The HIV-Pwt variant is represented with a circle solid symbol, HIV-PminB with a circle open symbol and HIV-PminC with a triangle open symbol. Each pair of viruses was mixed at initial ratios of 20:80 (first column), 50:50 (second column) and 80:20 (third column), and the proportion of each viral variant was determined. The error bars represent the standard error of three independent experiments.

After 29 days in cell culture, quantification of viral RNA sequences in the supernatant confirmed that these codon pair deoptimized protease viruses had a replication capacity similar to the wild-type virus (Figure 17). There was not a significant difference in the competitions between the wild-type virus and the HIV-PminB or HIV-PminC. This experiment confirmed that there were no differences in viruses' replicative capacities.

C. VIRAL STABILITY IN CELL CULTURE

To test the phenotypic stability of the attenuated variant HIV-PminA (containing 18 mutations compared to the wild-type virus) and the reoptimized variant HIV-Pmax (baring 38 mutations), serial large passages in MT-4 cells were performed. Serial passages were started at a multiplicity of infection (MOI) of 0.002 and two independent replicates were maintained. Every 3-4 days, viruses and infected cells were passaged to allow virus-free infection and cell-to-cell infection. After 15 passages, approximately 60 days of cell culture, supernatants were collected, viruses were titrated in MT-4 cells and viral replication capacities assays were performed.

As performed with starting viruses, the passaged viruses ratios of infectivity-to-virion particles were calculated. As can be seen in Table 6, HIV-PminA at passage 15 (HIV-PminAp15) replicate 1 and 2 have recovered infectivity. Their ratio at passage 0 was 4.76%, at passage 15 the ratio increased to 42.90% (replicate 1) and to 50.83% (replicate 2). HIV-Pmax had a similar ratio when compared to the wild-type virus at passage 0. At passage 15, HIV-Pmax (HIV-Pmaxp15) replicate 1 surpassed HIV-Pwt at passage 15 (HIV-Pwtp15) replicate 1 in infectivity-to-virion particle ratio from 103% to 237%. These results suggested that viruses at passage 15 recovered or ameliorated their replication capacities compared to wild-type passaged viruses.

Table 6.HIV-1 passaged viruses' infectivity-to-virion particle ratios[a]

	Virus	TCID50/ml[a]	p24 antigen (pg/ml)[b]	Relative Specific Infectivity[c]	TCID50/p24 antigen ratio to WT (%)
replicate 1	HIV-Pwtp15	1,99E+05	7,86E+06	1/39.45	100,00
	HIV-Pmaxp15	2,83E+05	4,69E+06	1/16.6	237,71
	HIV-PminAp15	7,28E+04	6,70E+06	1/92	42,90
replicate 2	HIV-Pwtp15	1,32E+05	5,96E+06	1/45.31	100,00
	HIV-Pmaxp15	1,02E+05	6,60E+06	1/64.75	70,00
	HIV-PminAp15	5,14E+04	4,58E+06	1/89.16	50,83

[a]Infectivity was measured by determining TCID50, [b]virion particles were estimated by HIV-1 p24 antigen measurement from a single clarified supernatant (virus stock) of each HIV variant and [c]relative specific infectivity was calculated by dividing infectivity of viruses by p24 antigen/ml in culture.

Ex vivo replication capacities in MT-4 cells of serially passaged HIV-1 codon-pair recoded viruses were performed at a MOI of 0.0002. In replicate 1 and 2, after 15 passages, HIV-PminA virus has recovered fitness, and passaged viruses were as potent as the wild-type virus passaged in parallel (Figure 18). Although a small difference in replicate 2 seemed to persist compared to the HIV-Pwtp15, it was not statistically significant different (p=0.2425). Remarkably, both replicates of passaged reoptimized viruses HIV-Pmax were still as fit as the wild-type virus in MT-4 cells.

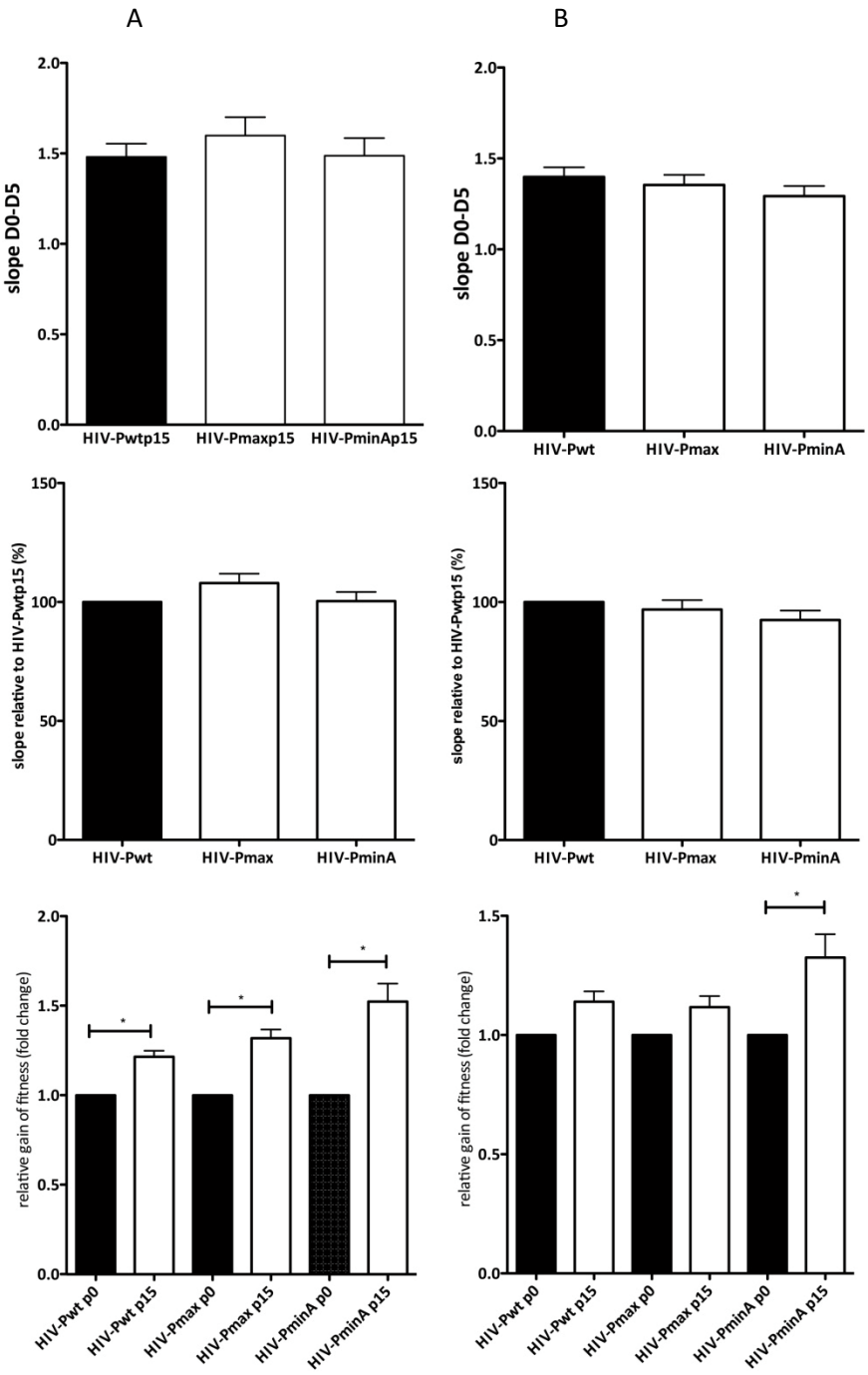


Figure 18. Replication kinetic assay of HIV-1 protease viruses replicate 1 (A) and replicate 2 (B) after 15 serial passages in MT-4 cells. Replication capacity assays were started with a MOI of 0.0002. Two replicates were performed for each variant. Production of the HIV-1 antigen p24 in culture supernatants was determined on days 0–5. (Upper Panel) Each bar represents the slope of antigen production for a given virus after infection of MT-4 cells, which provides an estimate of the viral replication capacity. (Middle panel) Comparisons between wild-type (HXB2) and mutant recoded viruses are shown. (Lower panel) Fold increases are shown relative to passage 0. The significance of the difference between slopes was calculated with GraphPrism v. 4 software (unpaired *t*-test). Values represent the mean ± SD from at least three independent experiments.

While all viruses from the first replicate (HIV-Pwtp15, HIV-Pmaxp15 and HIV-PminAp15) increased their replication capacity compared to passage 0 (p=0.0254, p=0.0221 and p=0.012, respectively), only HIV-PminAp15 among the second replicate improved its fitness (p= 0.0292).

Ex vivo replication capacities in PBMCs of serially passaged HIV-1 codon-pair recoded viruses were also performed at a MOI of 0.001 (Figure 19). As in MT-4 cells, HIV-PminA at passage 15 (HIV-PminAp15) from the first replicate regained fitness compared to HIV-PminA at passage 0 (HIV-PminAp0) (p=0.0358), while no differences between HIV-Pwt at passage 15 (HIV-Pwtp15) and HIV-PminAp15 slopes were observed (p=0.356). HIV-Pmax at passage 15 (HIV-Pmaxp15) had an enhanced replication capacity compared to HIV-Pwtp15 (p=0.0192). However, this improvement was not due to an increase in HIV-Pmaxp15 replication but rather to an HIV-Pwtp15 fitness reduction when compared to HIV-Pwtp0 fitness (p=0.0144).

In replicate 2, HIV-Pmaxp15 was as fit as the wild type virus at the same passage (HIV-Pwtp15). Both HIV-Pwtp15 and HIV-Pmaxp15 decreased their fitness significantly compared to their initial virus, HIV-Pwtp0 (p=0.0169) and HIV-Pmaxp0 (p=0.0047), respectively. In contrast, HIV-PminAp15 improved its replication capacity compared to the initial virus (HIV-PminAp0, p=0.0106). Despite this improvement, the difference between HIV-PminAp15 and HIV-Pwtp15 was maintained (p=0.0388), suggesting that HIV-PminAp15 replicate 2 was still statistically attenuated when compared to the wild-type virus passaged in parallel.

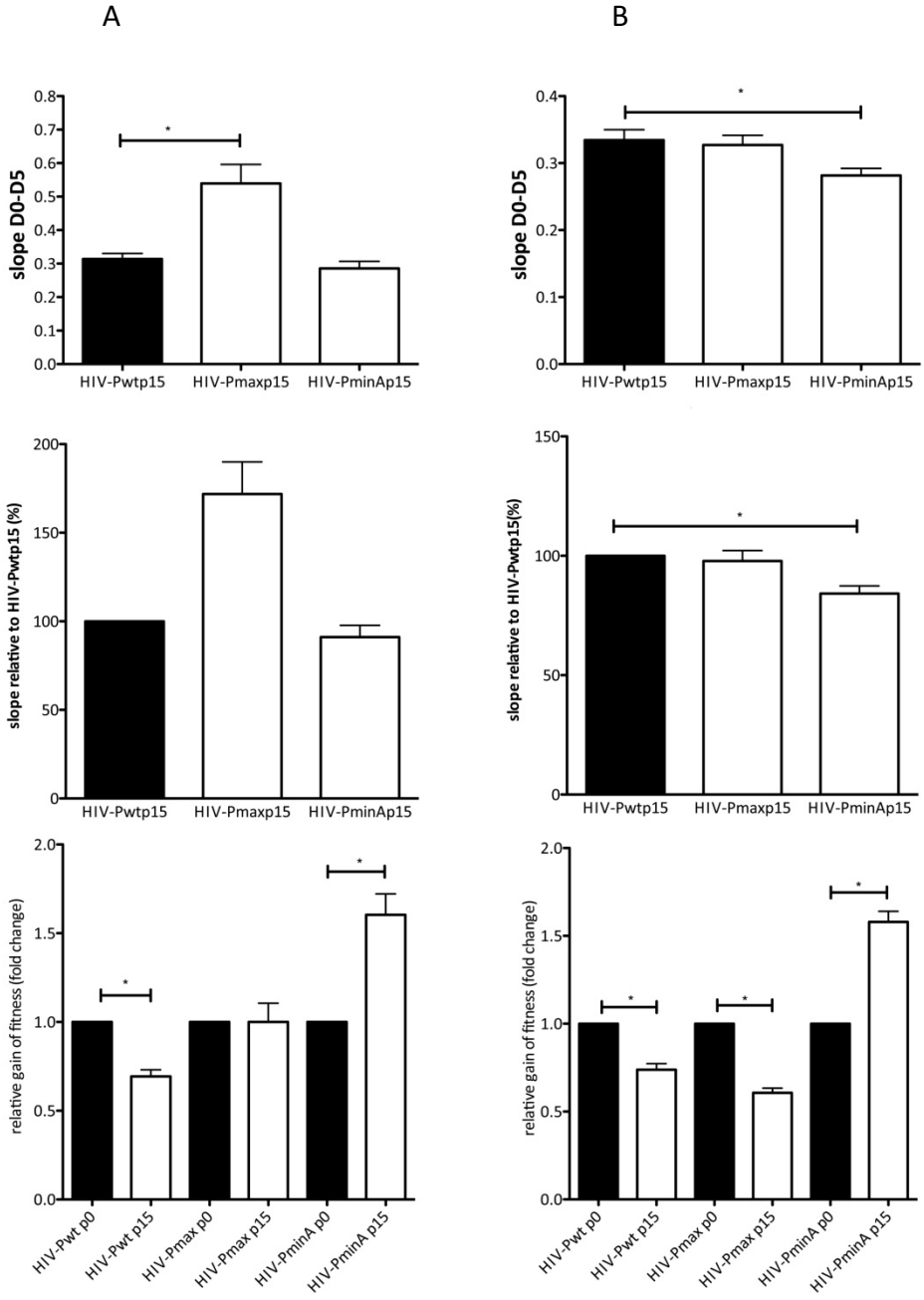


Figure 19. Replication kinetic assay of HIV-1 protease passaged viruses replicate 1 (A and replicate 2 (B) after 15 serial passages in PBMCs. Replication capacity assays were started with a MOI of 0.001. Two replicates were performed for each variant. Production of the HIV-1 antigen p24 in culture supernatants was determined on days 2–7. (Upper Panel) Each bar represents the slope of antigen production for a given virus after infection of MT-4 cells, which provides an estimate of the viral replication capacity. (Middle panel) Comparisons between wild-type (HXB2) and mutant recoded viruses are shown. (Lower panel) Fold increases are shown relative to passage 0. The significance of the difference between slopes was calculated with GraphPrism v. 4 software (unpaired *t*-test). Values represent the mean ± SD from at least three independent experiments.

D. QUASISPECIES ANALYSIS OF PASSAGED VIRUSES.

In order to verify if the observed phenotypic differences of the passaged viruses were due to reversions in protease gene or if they were a consequence of other mutations, we sequenced viruses from both replicates. Bulk sequencing of viral genomic RNA showed that only HIV-PminAp15 replicate 1 had two positions along its sequence with possible reversions of initial synonymous mutations, namely positions 97 and 99.

A clonal sequence analysis was also performed. A median of 22 clones from each virus was analyzed. HIV-Pmaxp15 from replicate 2 had no synonymous mutations along all the clonal sequences, whereas HIV-Pmaxp15 from replicate 1 had two synonymous reversion mutations towards wild-type sequence (Table 7). In contrast to HIV-Pmaxp15, HIV-PminAp15 accumulated more mutation during the passages. HIV-PminAp15 from replicate 1 contained 21 reversion mutations among 24 clones, affecting only 2 positions. The presence of these mutations in most of the clones suggested their relevance in the viral phenotype. As HIV-PminAp15 from replicate 1, HIV-PminAp15 replicate 2 also contained 1 reversion towards wild-type sequence. The lower numbers of synonymous mutations present in HIV-PminAp15 from replicate 2 when compared to HIV-PminAp15 from replicate 1 indicated that the replicate 2 virus was probably not evolving at the same rate as the replicate 1 virus.

Table 7. Mutation frequency and types of mutations in the HIV-1 protease-coding region of infected passaged cultures

	VIRUS	synonymous (reversions)	non-synonymous	number of positions reverted to wild-type	mutations/nucleotides	freq mut	
replicate 1	HIV-Pwtp15	3	9		12/6831	1,76E-03	3,34x
	HIV-Pmaxp15	2 (2)	7	2 in 21 clones	9/6237	1,44E-03	
	HIV-PminA p15	23 (21)	19	3 in 24 clones	42/7128	5,89E-03	
replicate 2	HIV-Pwtp15	2	10		12/6237	1,92E-03	1,57x
	HIV-Pmaxp15	0	5	0 in 17 clones	5/5049	9,90E-04	
	HIV-PminA p15	6 (1)	12	1 in 20 clones	18/5940	3,03E-03	

To test whether HIV-PminAp15 or HIV-Pmaxp15 replicate 1 and 2 differed in mutant spectrum complexity, mutation frequencies for the protease gene were determined and compared to the wild-type passaged viruses. The mutation frequency, determined as the

ratio between the sum of mutations (synonymous and non-synonymous) and total nucleotides analyzed, is shown in table 7. In replicate 1 viruses, mutation frequency rates of HIV-Pwtp15, HIV-Pmaxp15 and HIV-PminAp15 were $1.766x10^{-3}$, $1.44x10^{-3}$ and $5.89x10^{-3}$, respectively. HIV-PminAp15 virus had 3.34 fold more mutation frequency than HIV-Pwtp15. This increase in mutation frequency was statistically significant (p<0.0001; chi-square test) suggesting that HIV-PminAp15 was evolving at a higher rate than the wild-type virus. In contrast, in replicate 2, the corresponding increase between HIV-PminAp15 and HIV-Pwtp15 frequency mutation rates was not statistically significant (p=0,381), only 1.57 fold of difference, supporting the fact that HIV-PminAp15 replicate 2 was not evolving as fast as HIV-PminAp15 from replicate 1.

Table 8. Synonymous mutations and their position in the HIV-1 protease-coding region of infected passaged cultures

		VIRUS	position	starting mutated virus	reversion to wild-type virus	other mutations	nº clones
SYNONYMOUS MUTATIONS	replicate 1	HIV-Pwtp15	207	T		C	1/24
			210	A		G	1/24
			261	A		G	1/24
		HIV-Pmaxp15	81	G	A		1/21
			295	T	C		1/21
		HIV-PminAp15	7	C		T	1/24
			72	C		T	1/24
			97*	C	T		10/24
			99*	C	A		10/24
			265	C		T	1/24
	replicate 2	HIV-Pwtp15	66	T		C	1/21
			285	C		T	1/21
		HIV-PminAp15	30	C		T	1/20
			57	G	A		1/20
			78	C		T	1/20
			93	C		T	1/20
			195	A		G	1/20
			288	T		A	1/20

* linked mutations in all clones.

Each synonymous mutation in HIV-PminAp15 from replicate 1 was a change that reverted a pyrimidine residue to a purine residue (Table 8). Mutations in position 97 and position 99 were determined in the bulk viral genome sequencing as a mixed population. Those changes were linked and present in 10 out of 24 clones. No clone carrying only one of these mutations could be found. Mutations in HIV-PminAp15 from replicate 1 and 2 were

not in the same positions, suggesting that each replicate was exploring different ways to improve replication rates.

Synonymous mutations can affect designed CPS of codons around each single mutation. Thus, the new CPS of each new codon-pair present in the viral quasispecies was recalculated. With the exception of mutation at nucleotide position 9, all mutations present in HIV-PminA virus at passage 15 changed the calculated CPS towards ones with neutral or overrepresentation (Figure 20). Reversions such as in positions 97 and 99 resulted in a double CPS improvement: from -0.691 to 0.397 in codon pairs 32-33 and from -2.134 to 0.573 in codon pairs 33-34. These two ameliorations were probably the cause of the viral capacity replication increase observed in HIV-PminAp15 from replicate 1. On the other hand, in HIV-Pmaxp15 from replicate 1 only two synonymous mutations were reversions towards wild-type sequence. As expected, those two mutations affecting codon pairs 26-27 + 27-28 and 94-95 + 95-96 caused the drop of the reoptimized CPS to the wild-type CPS.

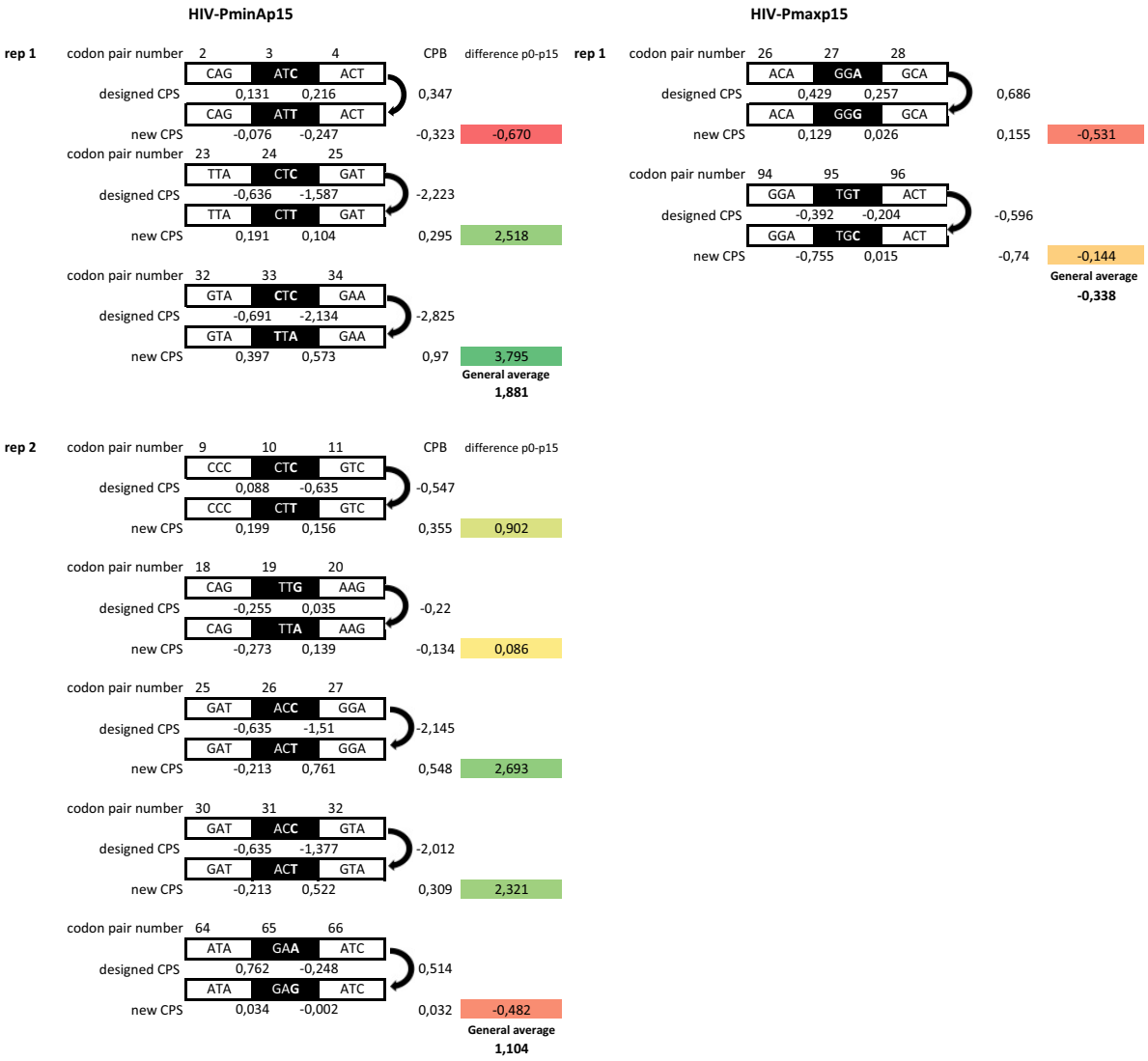


Figure 20. Mutations found in quasispecies of protease redesigned culture-passaged viruses. The mutations are in the dark square. All mutations except mutation in codon 3 alleviate negative CPSs in HIV-PminAp15 or decreased reoptimized positive CPSs in HIV-Pmaxp15. Negative CPS values relate to under-represented codon-pairs, while positive CPS values relate to over-represented codon-pairs.

Non-synonymous mutations were diverse in all evolved viruses (Table 9). Only one mutation, S37P, was present in both replicates of HIV-Pmaxp15 viruses in various clones (4 in replicate 1 and, 3 in replicate 2). Since no other mutations were present in more than one clone in viral quasispecies, differences in replication capacities were probably not due to these non-synonymous mutations.

Table 9. Non-synonymous mutations and their position in the HIV-1 protease-coding region of infected passaged cultures

NON-SYNONYMOUS MUTATIONS PRESENT IN THE QUASISPECIE		VIRUS	mutation	nº clones
	replicate 1	HIV-Pwtp15	D25G	1/24
			M46L	1/24
			G49E	1/24
			V56I	1/24
			P79L	1/24
			V82I	1/24
			G94D	1/24
			N98P	1/24
		HIV-Pmaxp15	G27R	1/21
			G27E	1/21
			S37P	4/21
			L90S	1/21
		HIV-PminAp15	G17R	1/24
			G27R	1/24
			M36I	1/24
			G40E	1/24
			R41K	1/24
			G48R	1/24
			K55Q	1/24
			Y59D	1/24
			C95Y	1/24

NON-SYNONYMOUS MUTATIONS PRESENT IN THE QUASISPECIE		VIRUS	mutation	nº clones
	replicate 2	HIV-Pwtp15	L23V	1/21
			G27R	1/21
			I54M	1/21
			D60N	1/21
			L63V	2/21
			E65K	1/21
			K70I	1/21
			Q92R	1/21
		HIV-Pmaxp15	T26I	1/17
			L33F	1/17
			S37P	3/17
		HIV-PminAp15	D29N	1/20
			R41K	1/20
			stop 42	1/20
			M46I	1/20
			G49E	1/20
			G49R	1/20
			V56I	1/20
			Q61E	2/20
			L63V	1/20
			E65K	1/20
			G68R	1/20

More specifically, looking at the nucleotide level, mutation spectra was different for each quasispecies. For both replicates, viruses HIV-PminAp15 showed a higher frequency of G→ A and C → T transitions than HIV-Pwtp15 (Table 10).

Table 10. Type of mutations in the HIV-1 protease-coding region of infected passaged cultures

Replicate	*Virus*		*Transitions*[a]				*Transversions*[b]							
		G→A +C→T[c]	*A→G*	*G→A*	*T→C*	*C→T*	*A→C*	*A→T*	*G→C*	*G→T*	*C→A*	*C→G*	*T→A*	*T→G*
1	***HIV-Pwtp15***	5	3	4	1	1	2	1						
	HIV-Pmaxp15	2	1 (1)	2	6 (1)									
	HIV-PminAp15	19		6		13 (10)	1				10 (10)			1
2	***HIV-Pwtp15***	5	1	4	1	1		1				4		
	HIV-Pmaxp15	1			3	1		1						
	HIV-PminAp15	13	1	10 (1)		3						3	1	

numbers in parenthesis represent the number of reversions Blank spaces means absence of mutations. Omitted transversions or transitions were not observed

To resolve the complexity of the quasispecies as well as to check if evolved viruses were nearer to the wild type virus (HIV-Pwt), Tamura 3 parameters phylogenetic reconstruction was computed for all the nucleotide sequences of HIV-PminAp15 quasispecies (Figure 21). As expected, HIV-PminAp15 from replicate 1 possessed two clearly differentiated clusters of sequences. One group was clustering together with HIV-PminA while the other group was closer to HIV-Pwt, but not clustering together.

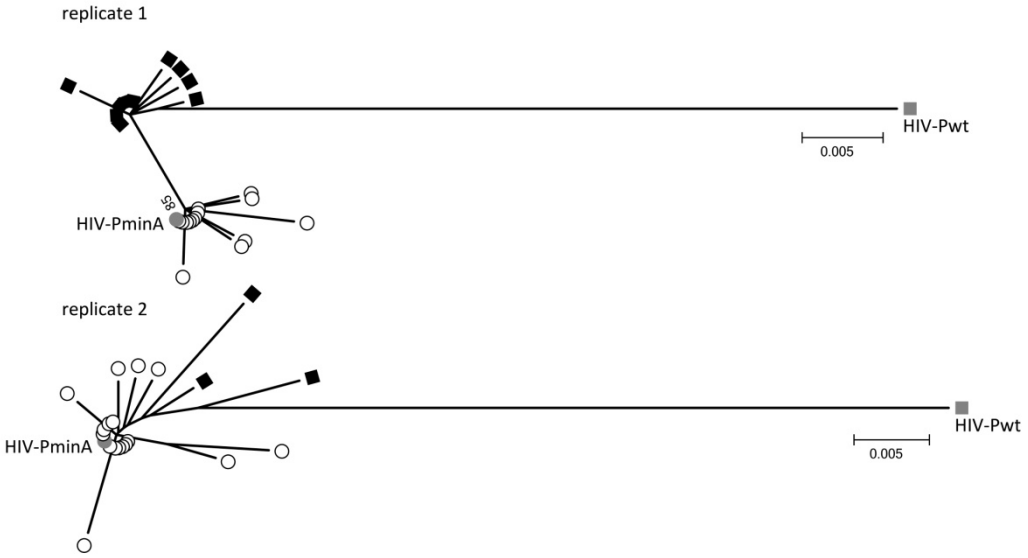


Figure 21. Evolutionary relationships of HIV-PminAp15 quasispecies. The evolutionary history was inferred using the Maximum-likehood method. The evolutionary distances were computed using the Tamura 3-parameters method (boostrap method 1000). Evolutionary analyses were conducted in MEGA5. Grey circle: HIV-PminA parental sequence; open circles: closer HIV-PminAp15 to HIV-PminA sequences; dark squares: HIV-PminAp15 farest from HIV-PminA parental sequence; grey square: HIV-Pwt parental sequence.

As observed in Figure 21, HIV-PminAp15 replicate 2 quasispecies clustered together. Only 3 sequences (dark squares) seemed to get closer to HIV-Pwt sequence confirming that replicate 2 quasispecies evolved towards HIV-Pwt sequence slower than replicate 1 quasispecies.

3. CODON-PAIR MECHANISM: EFFECTS ON TRANSLATION

We aimed to clarify if the mechanism of attenuation was due to pauses in translational machinery, as previously suggested (Coleman *et al.*, 2008; Irwin *et al.*, 1995). To investigate the molecular mechanism of attenuation of codon pair–deoptimized viruses, a GFP

reporter HIV-1 single cycle infectious clone (pNL4-3-deltaE-EGFP) was used. Normalization of GFP production after transfection allowed the quantification of viral growth, viral transcription, and viral translation efficiency because GFP expression depends only on the vector transfection efficiency and not on the viral replication capacity. HIV-1 protease recoded variants were used in these experiments by cloning them into pNL4-3dE-EGFP. Ratio between extracellular p24 and %GFP positive cells revealed the proteases capacity to produce new matures particles (Figure 22).

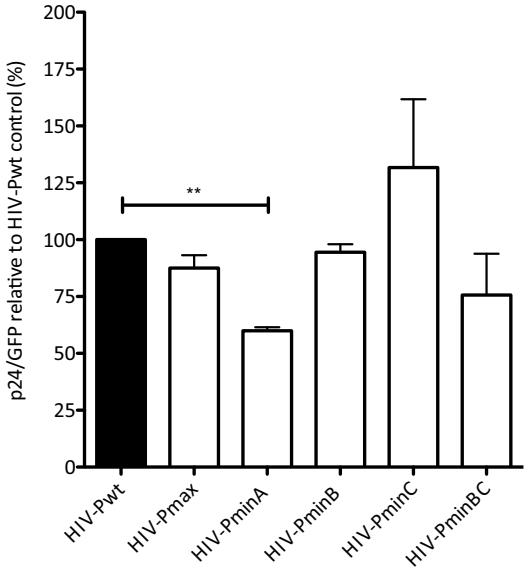


Figure 22. Single-cycle infectivity assay of HIV-1 protease codon pair–recoded synthetic constructs. HIV-1 protease codon pair–recoded synthetic constructs (Figures 12 and 13) were introduced into the GFP reporter HIV-1 infectious clone pNL4-3-deltaE-EGFP. The replication capacities of the different protease constructs are represented as a percentage relative to the wild-type HIV-1 HXB2 strain (100%). The relative replication capacity of the virus was determined by measuring the amount of p24 antigen. A black bar indicates the wild-type sequence. Three replicates were performed for each sample.

After cell vector transfection and GFP normalization, quantification of the production of viral protein p24 confirmed the lower replication capacity of the HIV-PminA deoptimized variant compared to wild-type virus (p=0.0014) (Figure 22). These results are based on the median of three independent experiments. No statistically significant differences were observed in all the other constructions. These results matched the replication capacity results obtained with MT-4 cells.

Our next goal was to determine if these differences were due to transcriptional mechanisms. For this purpose, RNA transcripts were analyzed to measure the amount of gag mRNA produced by transfected cells. The results showed that p24 production

reduction was not accompanied by a decrease in the production of gag mRNA transcript (Table 11), suggesting that the deoptimized codon pairs reduced translation.

Table 11. Real-time quantitative RT-PCR analysis of HIV-1 gag and GFP gene transcripts after one cycle of viral replication.

	HIV-Pwt	HIV-Pmax	HIV-PminA	HIV-PminB	HIV-PminC	HIV-PminBC
GFP Ct[a]	1,85 ± 0,10	16,75 ± 0,13	16,83 ± 0,27	16,48 ± 0,27	16,47 ± 0,17	16,69 ± 0,22
Gag Ct[a]	17,89 ± 0,39	17,40 ± 0,37	18,02 ± 0,36	17,36 ± 0,66	17,53 ± 0,14	17,86 ± 0,19
GAPDH Ct[a]	16,46 ± 0,04	16,25 ± 0,23	16,52 ± 0,30	16,65 ± 0,62	16,49 ± 0,31	16,22 ± 0,04
2^(-ΔΔCt)[b] Gag/GAPDH	1,00	1,22	0,95	1,66	1,32	0,91
2^(-ΔΔCt)[b] Gag/GFP	1,00	1,63	1,78	1,60	1,37	1,12

a. Ct: threshold cycle
b. Livak formula for relative gene expression analysis

4. CODON PAIR BIAS EFFECT ON VIRAL RNA STRUCTURE AND CODON USAGE

A. RNA STRUCTURE

The RNA free energy of the new constructs was designed to be similar to the RNA free energy of the input sequence (HIV-Pwt). To that purpose, 100-base long segments of all constructs, overlapping in 80 bases with each other, were analyzed using mFold. As it has been described, any segments that had a binding energy lower than -30kcal/mol may induce random synonymous substitutions at C-G binding locations because stable stem-loop structures can prevent translation, leading to an mRNA degradation (Doma & Parker, 2006). Therefore, if the protease construction showed low free binding energy, changes in affected synonymous mutations may be needed.

Figure 23 shows the corresponding free energy for every 100-base pair portion tested. Applying the analysis above described to the protease gene, all constructs had folding RNA free energies similar to HIV-Pwt. The wild-type protease mean was ΔG=-16.97kcal/mol, while the lower modified construct (HIV-Pmin) had a mean of ΔG=-21.75kcal/mol. Still, none of the constructs had a free binding energy lower than the threshold of -30kcal/mol. According to that mFold analysis, protease designed constructs did not need any modifications, because the recodification did not create any strong secondary RNA structures.

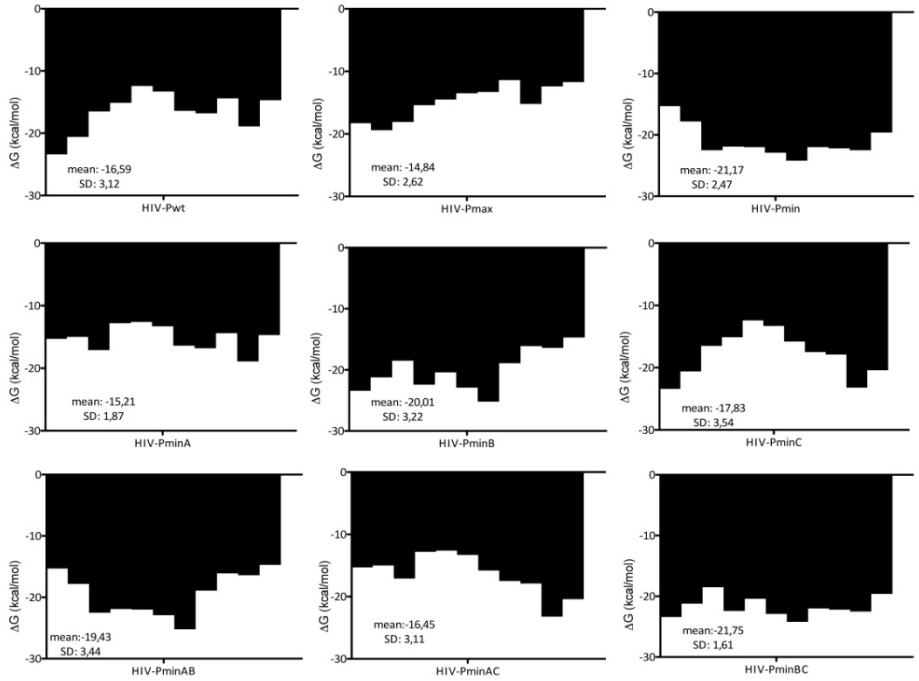


Figure 23. Folding free energy of synthetic and wild type protease constructs. The calculated energy value of each scanned 100-base pair segment was output to produce the graph. The mean folding energy for all constructs were: HIV-Pwt, -16.59; HIV-Pmax, -14.84; HIV-Pmin, -21.17; HIV-PminA, -15.21; HIV-PminB, -20.01; HIV-PminC, -17.83; HIV-PminAB, -19.83; HIV-PminAC, -16.45; HIV-PminBC.

B. SHAPE REACTIVITY IN PROTEASE

A new technique to evaluate RNA structure, SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) was used to measure HIV-1 RNA structure, in windows of 500 nucleotides (Watts *et al.*, 2009). SHAPE quantifies one by one the nucleotide adaptability in RNA by a chemical probing technology. RNA secondary structure can be inferred based on the loci of paired and unpaired nucleotides.

Based on the structure published, the mutations introduced in protease region were located and analyzed. Our aim was to determine if those synonymous changes were positioned in critical structural regions.

Although some punctual differences were observed in the maximum construct (Pmax), the inserted mutations interfered in the same RNA structures as in the minimum construct (Pmin) (Figure 24). Given that HIV-Pmax was viable in cell culture (Table 3), we can

hypothesize that the RNA structure of Pmin was not the main cause of non-functionality. Future work should analyze the real RNA structure *in vivo* of our sequences.

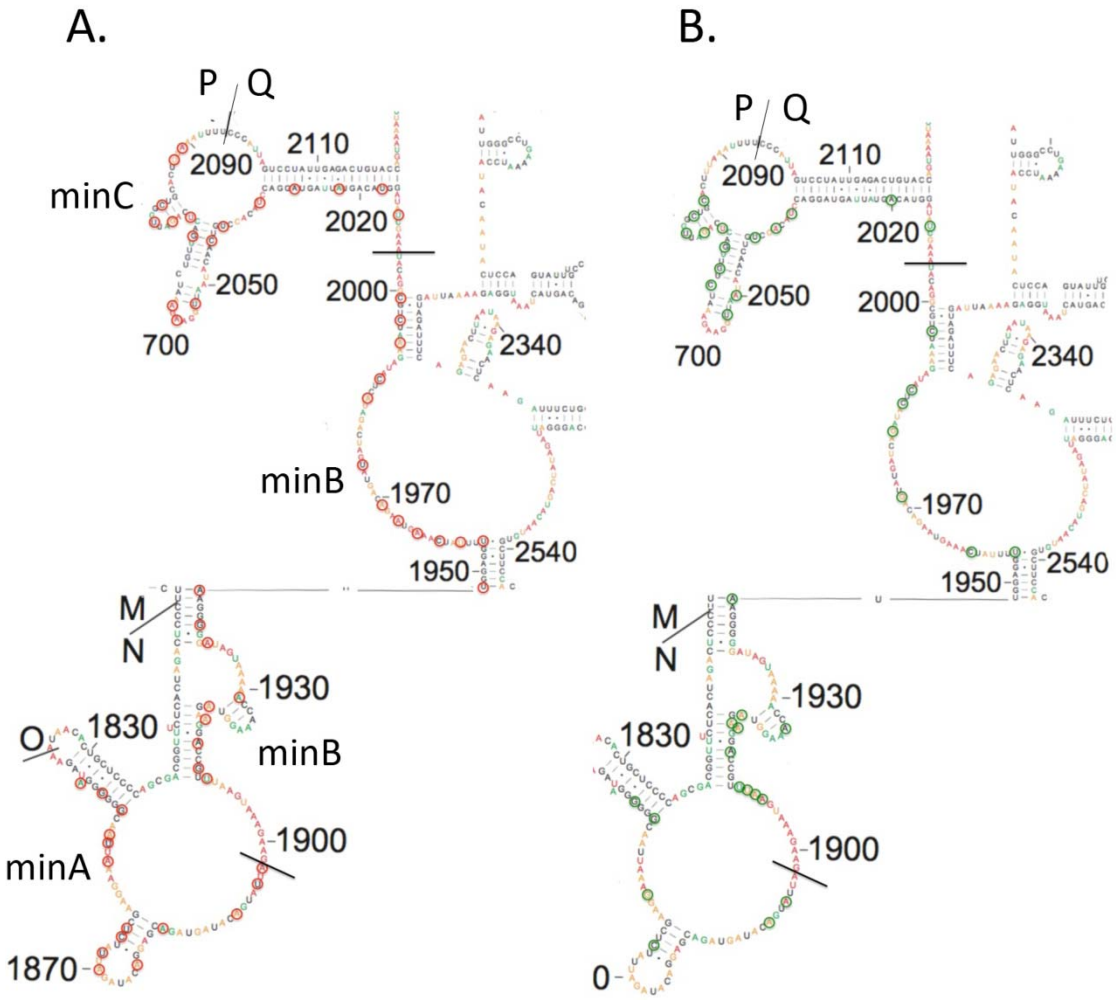


Figure 24. Shape analysis of synthetic protease constructs. Protease construct goes from M to P. A. minimum constructs, mutated nucleotides are shown in red. B. maximum constructs mutated nucleotides are shown in green. Adapted from (Watts et al., 2009).

C. CODON USAGE

When applying SAVE technology to HIV-1 protease coding region, our aim was to recode this region only by affecting codon-pair usage. If mutations are not controlled, codon-pair recoding may also modify codon usage bias. A suitable way to determine that codon usage bias remains untouched is to calculate the Effective Number of Codons (ENC) (Wright, 1990). ENC is one of the best overall measures of absolute synonymous codon usage bias;

the reported values usually range between 20 to and 61. A smaller ENC value indicates a larger extent of codon usage bias in a gene and vice versa. When only one codon is used for each amino acid, the ENC value will be 20; and when codons are used randomly, the ENC value would be 61 (Andrea *et al.*, 2011).

Our results showed that even though ENC is diverse in all constructs, none of the values were under 35, indicating that the codon usage bias remained unaffected in all the new CPB recoded constructs (Table 12).

Table 12. ENC values and GC content in third position and overall in percentages of protease

	ENC	GC3s (%)	GC (%)
HIV-Pwt	40,29	23,2	38,7
HIV-Pmax	39,78	20,0	36,4
HIV-PminA	39,64	29,5	40,7
HIV-PminB	52,32	33,7	42,1
HIV-PminC	54,24	32,6	42,4
HIV-PminAB	57,17	42,1	44,8
HIV-PminAC	53,06	37,9	43,8
HIV-PminBC	53,16	45,3	46,5
HIV-Pmin	54,93	53,7	49,2

A close association between GC% at the third codon position (or GC3%) and GC bias (a significant influence in codon bias) was observed in human and prokaryotic genomes (Palidwor *et al.*, 2010). The main hypothesis to explain this phenomenon is that amino acids (with the exception of methionine and tryptophan) accept at their third position synonymous changes with G or C. Thus, the use of those G/C-ending codons frequency should increase while rising GC bias at the same time. Given that pol gene is one of the richer regions in A nucleotides in the HIV-1 genome, with a low GC content (van der Kuyl & Berkhout, 2012), synonymous changes introduced to alter the CPB may have had an influence in GC bias.

Although the overall GC content remained similar in all constructs to the corresponding wild type region (varying from 38% to 49%), GC_3s% was variable (Table 12). HIV-Pmin, the utmost deoptimized and non-viable virus, doubled the wild type GC_3s% (53.7% versus 23.2%). HIV-PminBC, a highly CPB deoptimized virus (Table 3), had also a high GC_3s% (45.3%) yet its replication capacity was as good as the wild-type virus (Figure 16). The

correlation between $GC_3s\%$ and virus replication capacities was evaluated. In protease viruses, no statistically significant relationship was found between these two values (Figure 25).

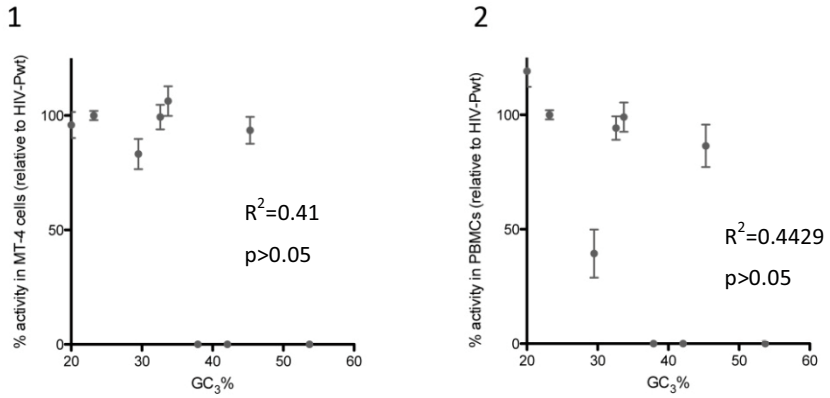


Figure 25. Relation between GCs in third position of the codon and activity in MT-4 cells (1) or PBMCs (2) in protease. Replication capacities relatives to HIV-Pwt corresponding to each protease construct (if the virus was non-viable, the replication capacity was assumed to be 0%) were correlated with $GC_3\%$ found in all sequences. All statistics tests were performed with GraphPrism v. 4 software.

Since one of the most under-represented codon pairs in eukaryotes, and hence in human codon pairs, are CpG islands, correlation between them and replication capacities in MT-4 cells and PBMCs was also analyzed for our protease constructs (Figure 26).

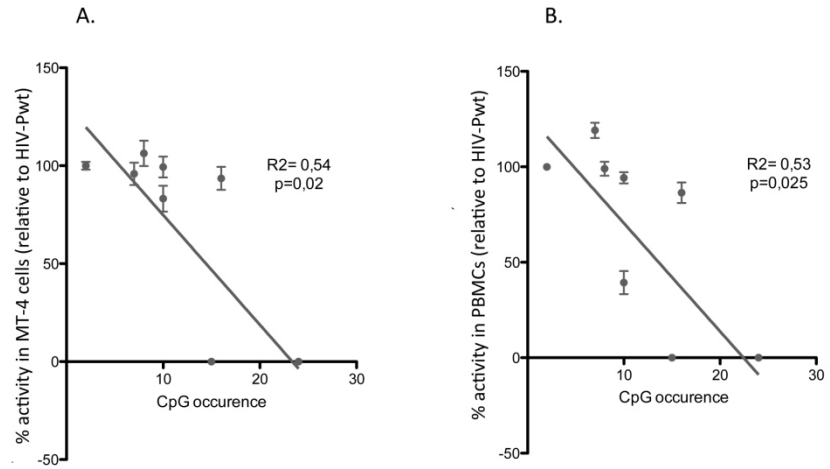


Figure 26. Relation between CpG islands and activity in MT-4 cells (A) or PBMCs (B) in protease constructs. Replication capacities relatives to HIV-Pwt corresponding to each protease construct (if the virus was non-viable, the replication capacity was assumed to be 0%) were correlated with CpG occurrence found in all sequences. All statistics tests were calculated with GraphPrism v. 4 software.

A negative correlation was found between the presence of CpG islands and the reduction of replication capacities when compared to wild type virus.

CHAPTER 2. GAG CPB RECODING

1. GAG LARGE-SCALE MANIPULATION: CPB DESIGN

Gag encodes for viral structural proteins and therefore has a very different role in HIV-1 biology than protease. As it was performed for protease gene, large-scale manipulation of gag region was achieved using the algorithm described by Coleman et al. (Coleman *et al.*, 2008). Gag region was recoded using under-represented codon pairs relative to the human genome. To obtain deoptimized gag constructs, wild-type codon-pairs existing in gag were substituted by codon-pairs with a low CPS. As with protease-recoded viruses, amino-acid sequence was preserved, only synonymous mutations were introduced.

To generate HIV-Gmin (a total deoptimized gag region virus), three hundred and forty-two synonymous substitutions were introduced in the 1503 nucleotides of gag (Figure 27). In order to avoid lethality as well as the introduction of non-synonymous mutations in protease, no substitutions were designed in the p6 portion of gag. Additionally, slippery site AnTTTTTT followed by a stem-loop structure and responsible for the regulation of the -1 ribosomal frameshift out of the Gag reading frame into the Pol reading frame, remained unaffected.

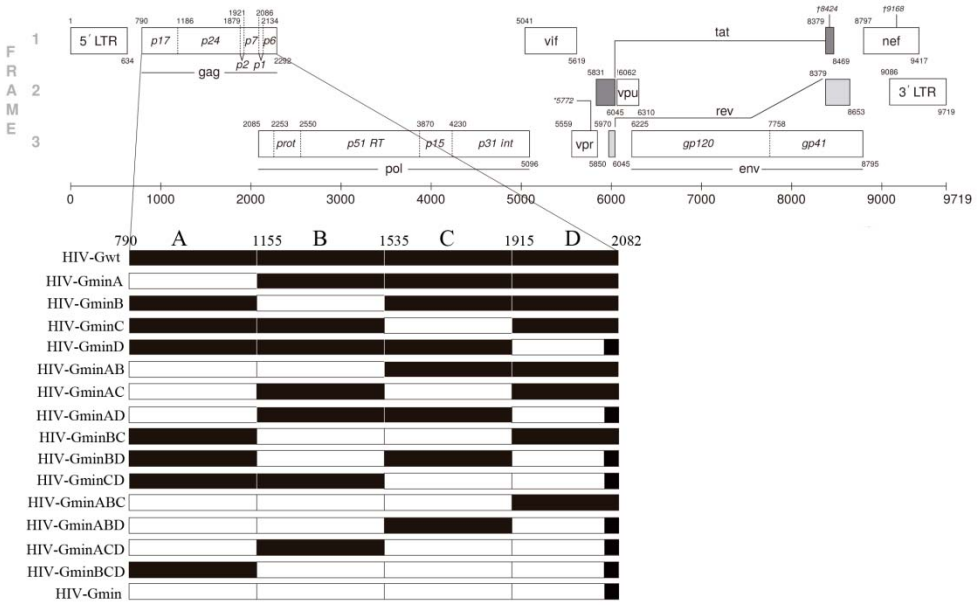


Figure 27. Structures of the various chimeric, partly synthetic gag HIV-1 constructs. Black boxes indicate wild-type regions and white boxes show deoptimized regions. Nucleotide positions are displayed relative to HXB2 sequence.

Although the construct with a total deoptimization of gag (HIV-Gmin) had a CPB score of -0.348, it was not as low as the total deoptmized construct from portease (HIV-Pmin).

Nevertheless, HIV-Gmin CPB score was also an extreme CPB score exceeding the exterior minimum of naturally prevailing in human genome's CPB score. Due to the length of gag, 1503 nucleotides long, this gene was divided in four different moieties to progressively reduce the number of under-represented pair codons present in HIV-Gmin. All combinations were produced, with one, two or three deoptimized regions, generating fifteen viruses with graded levels of deoptimization (Figure 27). These new variants were broadly dispersed along the human codon pair usage. Most of the new recoded viruses (HIV-Gmin, HIV-GminABC, HIV-GminABD, HIV-GminACD, HIV-GminBCD, HIV-GminAB, HIV-GminAC, HIV-GminAD, HIV-GminBC, HIV-GminBD and HIV-GminCD) exceeded the extreme human CPB scores while HIV-GminA, HIV-GminB, HIV-GminC and HIV-GminD were still in the proximity of human CPB range (Figure 28).

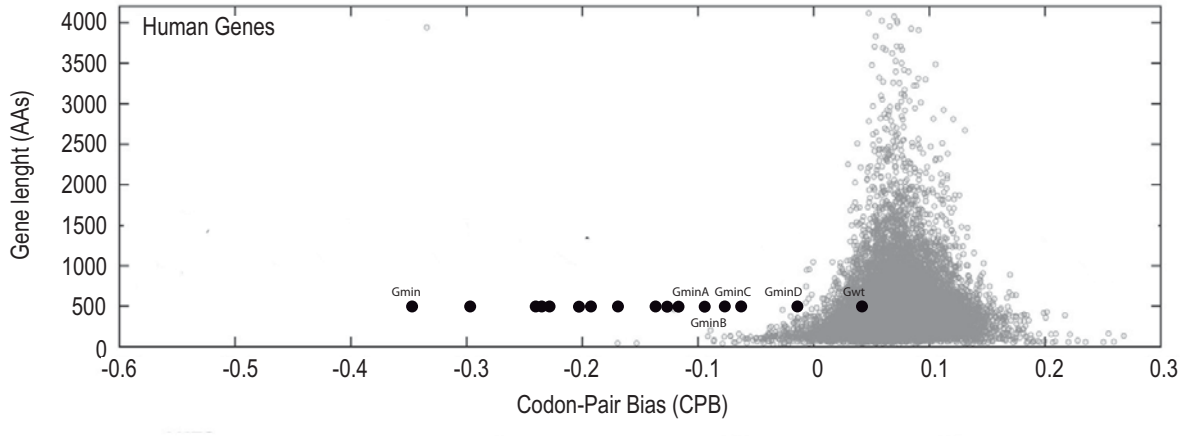


Figure 28. Calculated codon pair bias (CPB) score for all 14.795 annotated human genes. Each dot represents the calculated CPB score of a gene plotted against its amino acid length. Underepresented codon pairs yield negative scores. Various gag constructs are represented by symbols; Gwt represent the wild-type gag HIV-1 virus (CPB=0.035). Adapted from (Coleman *et al.*, 2008).

Wild-type Gag CPB score was once more very similar to the median of all human CPB scores, indicating that probably the HIV-1 has evolved towards a codon pair usage analogous to the human one. All viruses had a different number of synonymous mutations and therefore different CPB scores (Table 13), varying from -0.348 (HIV-Gmin) to -0.011 (HIV-GminD). The percentages of mutations compared to the wild-type HXB-2 gag ranged from 2.26% (HIV-GminD) to 22.77% (HIV-Gmin HIV-GminD generated a construct similar to the wild type virus HIV-Gwt). Contrarily, all the other constructs had a negative CPB score, indicating that there was a general under-representation of the individual codon pairs encoded in those constructs.

Table 13. Calculated codon pair bias (CPB) score for each one of the constructs. In blue: viable viruses, in orange: non-viable viruses, even after five blind passages.

	CPB score	# mutations	% mutations
HIV-Gwt	0,035		
HIV-GminA	-0,091	118,00	7,86
HIV-GminB	-0,077	97,00	6,46
HIV-GminC	-0,070	93,00	6,19
HIV-GminD	-0,011	34,00	2,26
HIV-GminAB	-0,203	215,00	14,31
HIV-GminAC	-0,193	211,00	14,05
HIV-GminAD	-0,133	152,00	10,12
HIV-GminBC	-0,183	190,00	12,65
HIV-GminBD	-0,123	131,00	8,72
HIV-GminCD	-0,113	127,00	8,46
HIV-GminABC	-0,305	308,00	20,51
HIV-GminABD	-0,242	249,00	16,58
HIV-GminACD	-0,235	245,00	16,31
HIV-GminBCD	-0,225	224,00	14,91
HIV-Gmin	-0,348	342,00	22,77

Only constructs HIV-GminA, HIV-GminB, HIV-GminC, and HIV-GminD, which contained 118, 97, 93, and 34 synonymous substitutions, respectively (Table 13), produced viable viruses.

Phylogenetic analyses were conducted to measure the level of genetic differences that could be introduced by those mutations. The genetic distances of recoded HIV-1 gag constructs were computed using the Kimura-2 parameters method. Two robust clusters were formed in a phylogenetic tree, supported by a high number of boostrap (Figure 29). As expected, HIV-GminD clustered together with HIV-Gwt. Moreover, each individual or double mutated sequence formed a group with the corresponding sequence containing D deoptimized region, suggesting that D region (with only 34 mutations and a CPB score of -0.011) was not different enough to create a cluster on its own. All sequences including deoptimized A fragment clustered together, implying that A region could be the one having more evolutionary differences compared to HIV-Gwt.

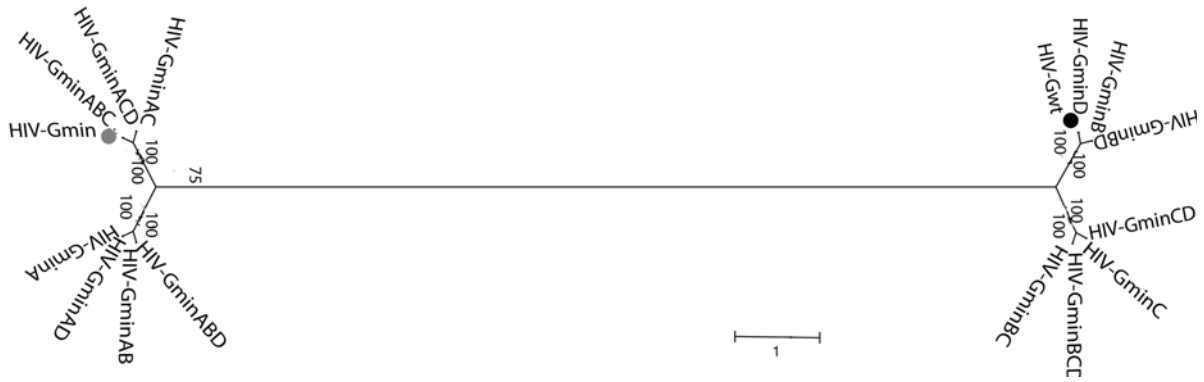


Figure 29. Evolutionary relationships of HIV-1 recoded proteases. The evolutionary history was inferred using the Neighbor-Joining method, Kimura 2-parameters method. The optimal tree with the sum of branch length = 14.30116305 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are indicated next to the branches. The rate variation among sites was modeled with a gamma distribution (shape parameter = 0.05). The analysis involved 16 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1503 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.

Individual CPS were plotted in a heat map, a graphical representation where individual CPS contained in the sequences are represented as colors (Figure 30): red colors showing under-represented codon-pairs and blue colors showing over-represented codon-pairs. On one hand, some highly under-represented codon-pairs were observed in deoptimized A, B and C gag regions. Specifically, region A carried individual CPS as low as -2.134 (CP51 and CP41), -2.191 (CP39) and -1.809 (CP11); region B carried individual CPS as low as -2.191 (CP229) and region C carried individual CPS as low as -1.869 (CP259), -2.134 (CP343) and -1.693 (CP363). On the other hand, deoptimized region D did not carry in its sequence any "on the edge" codon-pairs.

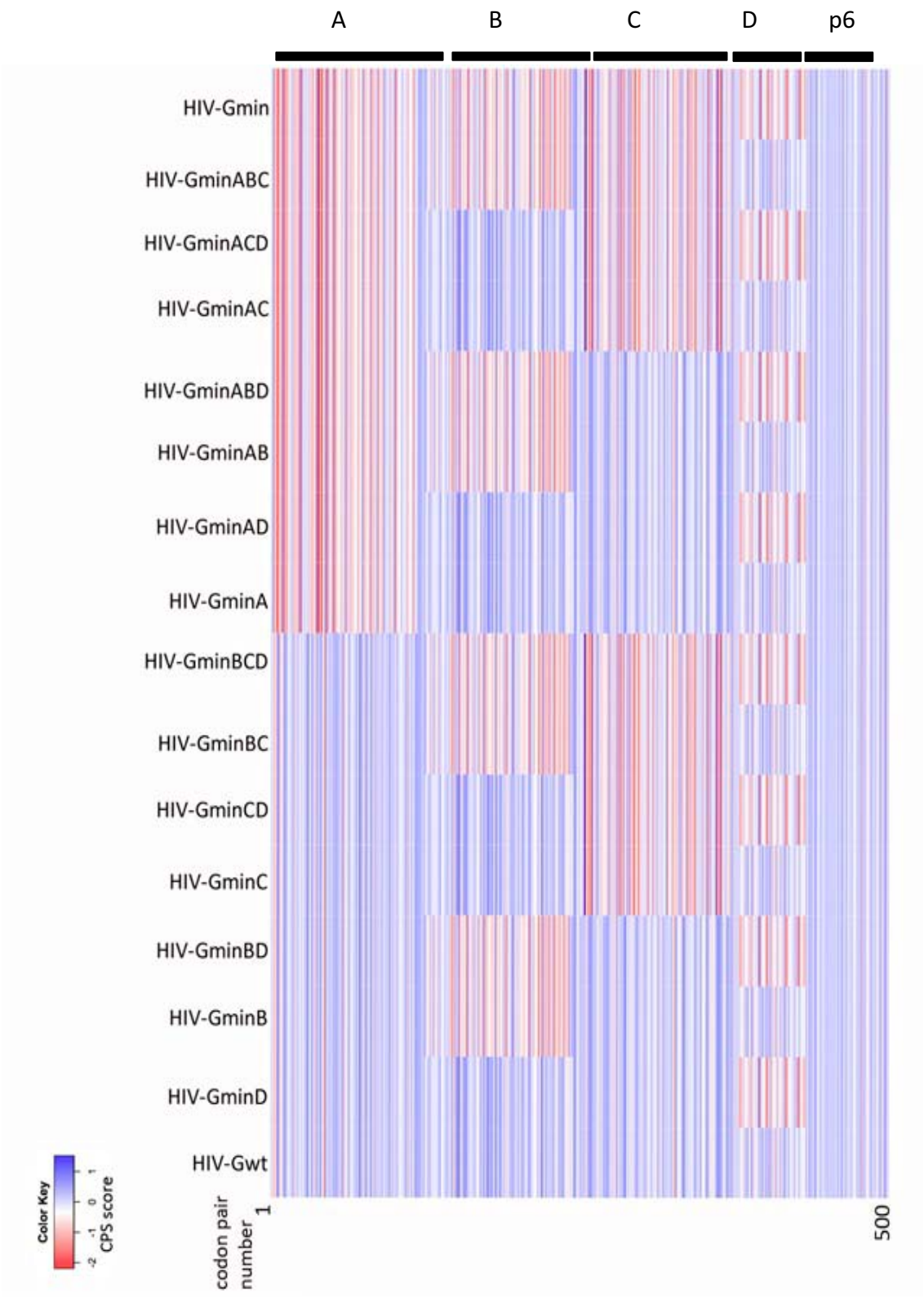


Figure 30. Gag recoded sequences Heat Map. A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors. Blue colors represent codon pairs with high CPS, red colors represent codon pairs with low CPS and white colors represent intermediate CPS (around 0). A, B, C, D and p6 regions are indicated.

2. *EX VIVO* REPLICATION CAPACITY OF RECODED GAG VIRUSES

A. MT-4 AND PBMCS REPLICATION CAPACITY ASSAYS

Recombinant gag infectious viruses were generated by co-transfection of purified recoded gag constructs and gag-protease-deleted NL4-3 clone in MT-4 cells. Only HIV-Gwt, HIV-GminA, HIV-GminB, HIV-GminC and HIV-GminD produced viable viruses. All the other constructs did not produce viable viruses despite the fact that five blind passages were done.

Deoptimized viruses HIV-GminA, HIV-GminB and HIV-GminC (118, 97 and 93 synonymous mutations, Table 12) had a considerably lower TCID50/p24 ratio than HIV-Gwt (Table 14), while HIV-GminD virus (34 synonymous mutations) had a higher TCID50/p24 ratio.

Table 14. HIV-1 contructs infectivity-to-virion particle ratios[a]

Virus	TCID50/ml[a]	p24 antigen (pg/ml)[b]	Relative Specific Infectivity[c]	TCID50/p24 antigen ratio to WT (%)
HIV-Gwt	3,13E+05	3,63E+06	1/11.61	100,00
HIV-GminA	2,97E+04	4,14E+06	1/139.6	8,32
HIV-GminB	1,01E+05	1,04E+07	1/103.44	11,23
HIV-GminC	1,25E+03	6,08E+05	1/486.5	2,39
HIV-GminD	1,12E+06	8,01E+06	1/7.16	162,07

[a]Infectivity was measured by determining TCID50, [b]virion particles were estimated by HIV-1 p24 antigen measurement from a single clarified supernatant (virus stock) of each HIV variant and [c]relative specific infectivity was calculated by dividing infectivity of viruses by p24 antigen/ml in culture.

In vitro viral replicative capacities of viable viruses were determined by infecting both MT-4 cells and previously stimulated PBMCs (deriving from a pool of 4 different donors). Infections in MT-4 cells were performed at a MOI of 0.0002 and in PBMCs at a MOI of 0.001, followed by daily sample recollection. The amount of extracellular capsid p24 protein in culture supernatant was controlled and used to analyze viral growth.

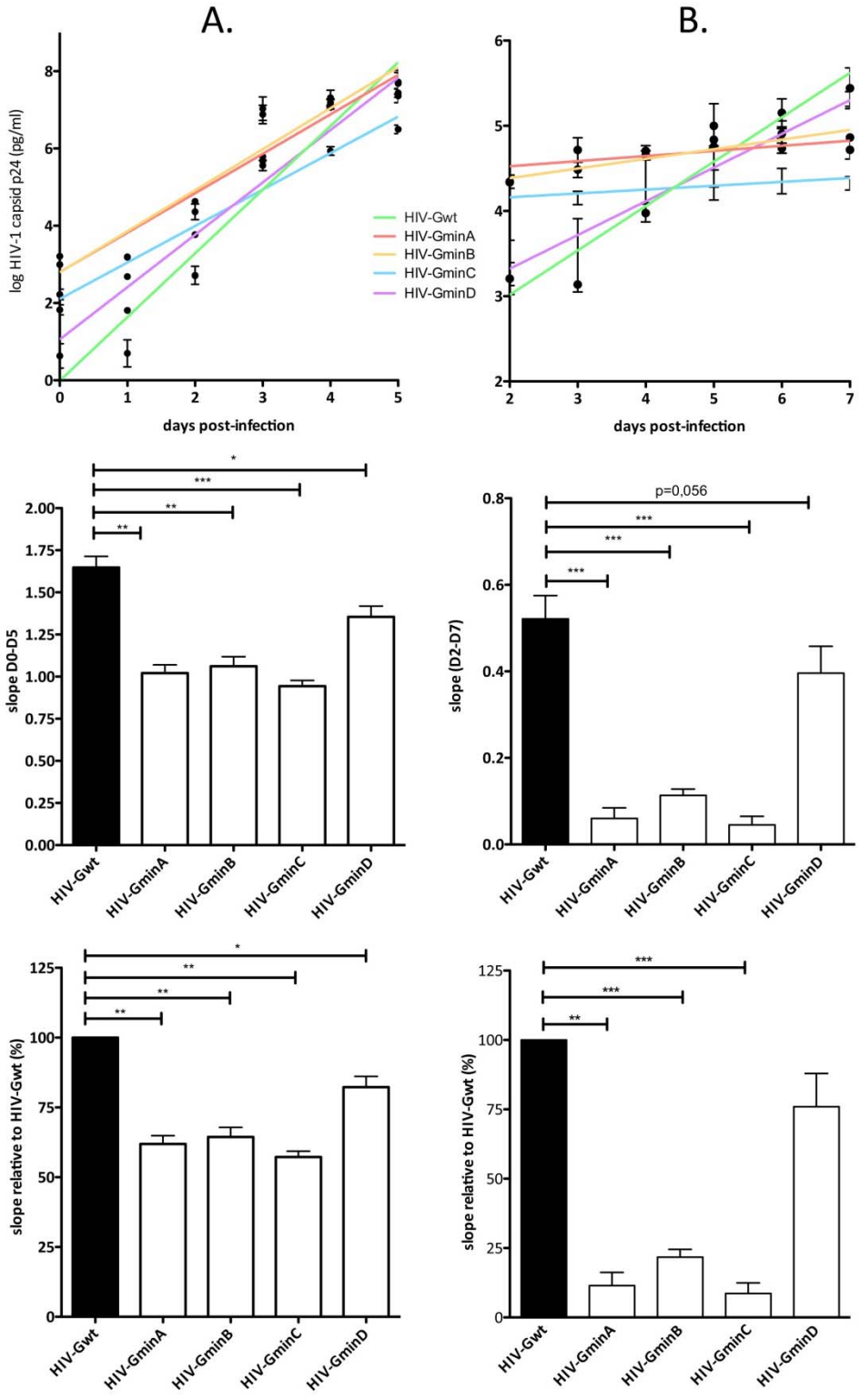


Figure 31. Gag Replication capacity assays. MT-4 cells (A) and PBMCs (B). (Upper panel) For each virus, the slope of the plot provides an estimate of the viral replication capacity. (Middel panel) The slope of p24 antigen production of each virus after infection of MT-4 cells (A) and PBMCs (B) is shown by the bars. (Lower panel) Comparisons of wild-type HXB2 virus versus all the viable CPB gag mutants are shown. The significance of the difference between the slopes was calculated using the GraphPrism v.4 software. Values represent the means +/- SDs from three independent experiments

As in the results previously obtained with the fully-deoptimized protease variant, HIV-Gmin construct (Figure 28) yielded no viable virus. In contrast, constructs HIV-GminA, HIV-GminB, HIV-GminC, and HIV-GminD produced viable viruses that were highly attenuated in MT-4 cells (p=0.047, p=0.025, p=0.019 and p=0.037, respectively) (Figure 31A). HIV-GminA, HIV-GminB and HIV-GminC viable deoptimized viruses also showed a highly significant attenuated phenotype in PBMCs when compared to wild-type virus (p=0.0002, p=0.0002 and p=0.0001, respectively) (Figure 31B). Their replication capacity was lower than a 25% of HIV-Gwt. Furthermore, in PBMCs, HIV-GminD was not statistically significantly different when compared to HIV-Gwt (p=0.0579). Overall, our results demonstrate that the codon-pair bias deoptimization of HIV-1 gag coding regions generates viruses that are attenuated *ex vivo.*

Antiviral activities of codon-pair deoptimized gag viruses were also performed. Drug susceptibility to AZT, TNF-DF, NFV, ATV, APV and DRV (Table 15) was examined. As expected, no differences in drug susceptibility were found when HIV-GminB, HIV-GminA and HIV-GminD viruses were tested. Due to the low viral titer of HIV-GminC virus, we were unable to check for its antiviral activities.

Table 15. Susceptibility of HIV-1 constructs to antiviral compounds

	IC50 (nM)					
	NRTIs		protease inhibitors			
	AZT	TNF-DF	NFV	DRV	ATV	APV
HIV-Gwt	6,06 ± 1,14	18,30 ± 3,41	9,93 ± 3,04	2,37 ± 0,71	4,44 ± 1,28	13,56 ± 1,98
HIV-GminA	6,68 ± 2,12 (1,10)	13,64 ± 8,85 (0,75)	3,39 ± 1,26 (0,34)	1,74 ± 1,24 (0,73)	2,08 ± 0,96 (0,47)	8,11 ± 4,96 (0,60)
HIV-GminB	12,19 ± 1,96 (2,01)	28,73 ± 8,86 (1,57)	10,14 ± 8,03 (1,02)	2,59 ± 0,69 (1,09)	4,25 ± 0,82 (0,96)	14,66 ± 1,63 (1,08)
HIV-GminD	8,64 ± 3,30 (1,42)	27,31 ± 9,23 (1,5)	8,66 ± 4,36 (0,87)	2,20 ± 0,14 (0,93)	3,13 ± 1,54 (0,7)	12,63 ± 2,04 (0,93)

The IC50 values shown are means ± standard deviations of at least three tests. Fold increase in IC50 relative to the wild-type HXB2 virus control is shown in parentheses.

B. VIRAL STABILITY IN CELL CULTURE

In order to investigate the stability of attenuated viable gag variants in cell culture, and given that their attenuation was superior than attenuated protease variant HIV-PminA, serial large passages in MT-4 cells were carried out. Two independent replicates were passaged at an initial MOI of 0.002 in MT-4 cells. Every 3-4 days, viruses and infected cells were passaged to allow virus-free infection and cell-to-cell infection. After 15 passages

(approximately 60 days of cell culture) supernatants were collected, viruses were titrated in MT-4 cells and viral replication capacity assays were performed.

Initially, infectivity-to-virion particle ratios (TCID50/p24) were determined from gag-passaged viruses. As mentioned before, variation in this ratio is a good indicator of passaged viruses replication capacity when compared to passage 0 viruses. As shown in Table 16, in replicate 1, virus HIV-PminB from passage 15 (HIV-GminBp15) recovered infectivity relative wild type virus (HIV-Gwtp15, 32.42%). In contrast, HIV-GminA and HIV-GminC from the last passage (HIV-GminAp15 and HIV-GminCp15, respectively) appeared to maintain their replicative differences from HIV-Gwtp15 (4.25% for passage 15 versus 11.23% for passage 0 for HIV-GminA and 2.56% p15 versus 2.39% p0 for HIV-GminC). Additionally, HIV-GminDp15 decreased its ratio compared to HIV-Gwtp15 (42.60% p15 compared to 162.07% p0).

Table 16. HIV-1 passaged viruses' infectivity-to-virion particle ratios[a]

	Virus	TCID50/ml[a]	p24 antigen (pg/ml)[b]	Relative Specific Infectivity[c]	TCID50/p24 antigen ratio to WT (%)
replicate 1	HIV-Gwt	2,67E+07	8,52E+06	3.13	100,00
	HIV-GminA	1,56E+06	1,17E+07	1/7.5	4,25
	HIV-GminB	1,25E+06	1,23E+06	1.01	32,43
	HIV-GminC	5,19E+04	6,47E+05	1/12.46	2,56
	HIV-GminD	7,81E+06	5,85E+06	1.33	42,60

	Virus	TCID50/ml[a]	p24 antigen (pg/ml)[b]	Relative Specific Infectivity[c]	TCID50/p24 antigen ratio to WT (%)
replicate 2	HIV-Gwt	3,93E+06	5,76E+06	1/1.4	100,00
	HIV-GminA	1,09E+06	4,01E+06	1/3.67	39,84
	HIV-GminB	1,41E+06	5,55E+06	1/3.93	37,19
	HIV-GminC	6,90E+05	1,69E+06	1/2.44	59,88
	HIV-GminD	2,69E+06	5,24E+06	1/1.94	75,18

[a]Infectivity was measured by determining TCID50, [b]virion particles were estimated by HIV-1 p24 antigen measurement from a single clarified supernatant (virus stock) of each HIV variant and [c]relative specific infectivity was calculated by dividing infectivity of viruses by p24 antigen/ml in culture.

In replicate 2, attenuated viruses HIV-GminAp15, HIV-GminBp15 and HIV-GminCp15 improved their TCID50/p24 ratio relative to HIV-Gwtp15 (39.84%, 37.19% and 59.88% respectively). As in replicate 1, HIV-GminDp15 had a lower ratio relative to HIV-Gwtp15 (75.18% p15 compared to 162.07% p0).

These results suggested that viruses at passage 15 had probably improved their replication capacities when compared to wild-type passaged viruses.

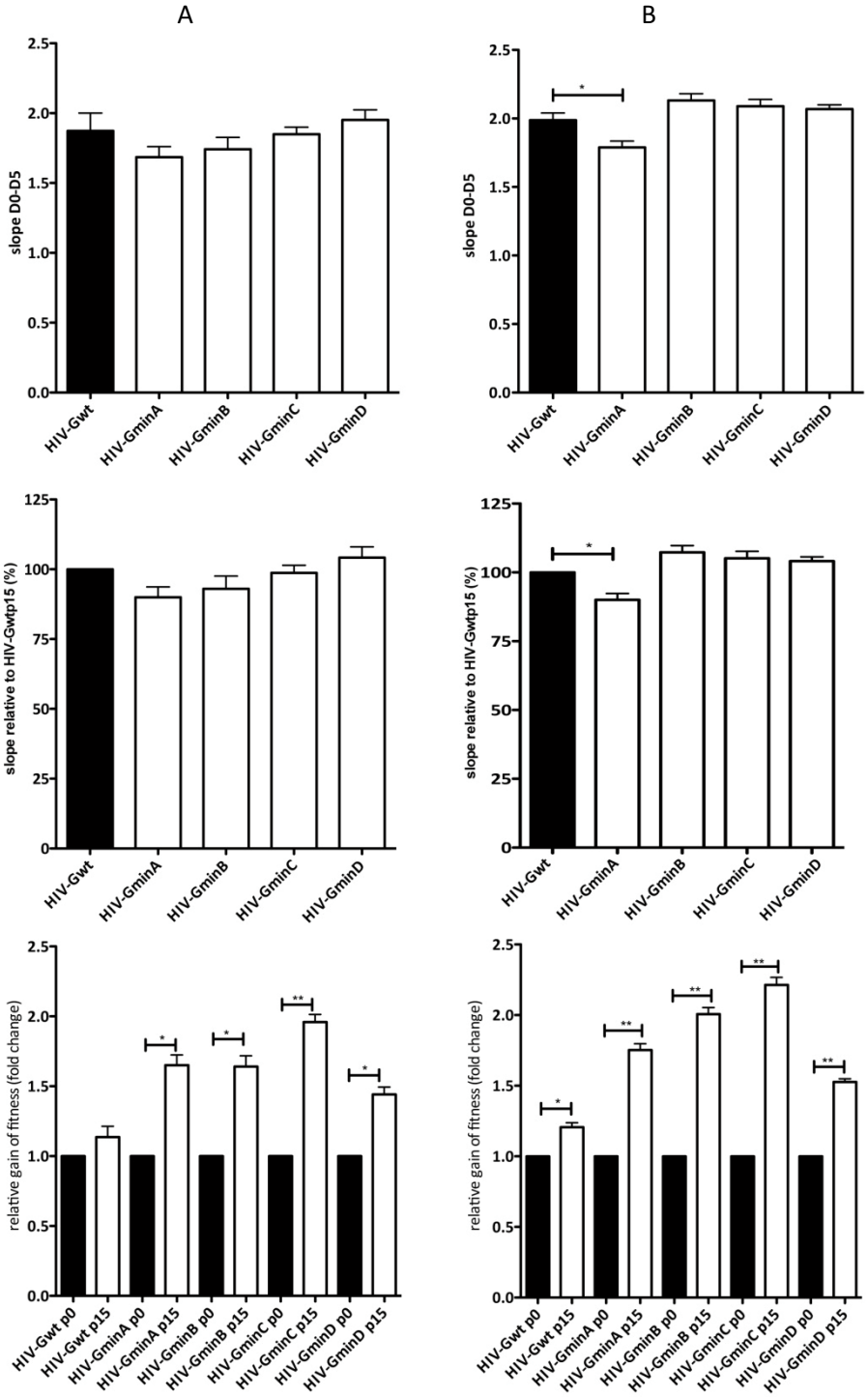


FIGURE 32. Replication kinetic assay of HIV-1 gag passaged viruses replicate 1 (A and replicate 2 (B) after 15 serial passages in MT-4 cells. Replication capacity assays were started with a MOI of 0.0002. Two replicates were performed for each variant. Production of the HIV-1 antigen p24 in culture supernatants was determined on days 0-5. (Upper Panel) Each bar represents the slope of antigen production for a given virus after infection of MT-4 cells, which provides an estimate of the viral replication capacity. (Middle panel) Comparisons between wild-type (HXB2) and mutant recoded viruses are shown. (Lower panel) Fold increases are shown relative to passage 0. The significance of the difference between slopes was calculated with GraphPrism v. 4 software (unpaired *t*-test). Values represent the mean ± SD from at least three independent experiments.

Replication capacities of passaged gag recoded viruses were evaluated in MT-4 cells (Figure 32). While at passage 0 all viruses had a lower slope when compared to the wild-type virus, at passage 15 all viruses from both replicates recovered fitness. All viruses were no longer statistically significantly different from HIV-Gwtp15 *(p*>0.1), with the exception of HIV-GminAp15 replicate 2, which remained less fit than the HIV-Gwtp15 (*p*=0.047). Furthermore, all passaged gag CPB recoded viruses had a statistically significant fold increase (superior to 1.5 fold change) when comparing passage 0 to passage 15 slopes (*p*<0.05) in MT-4 cells.

As shown in Figure 33, phenotypic stability of passaged viruses was also tested in previously stimulated PBMCs. In both replicates, HIV-GminAp15 viruses still displayed a significantly lower replication capacity when compared to HIV-Gwtp15 viruses (*p*=0.0416 for replicate 1 and *p*=0.0317 for replicate 2). Nevertheless, HIV-GminAp15 viruses recovered fitness when compared to passage 0 virus: replicate 1 had an increase of 5.79 fold change (*p*=0.0026) and replicate 2 had an increase of 4.42 fold change (*p*=0.0055). HIV-GminCp15 replicate 1 showed a 70.30% of replication capacity when related to HIV-Gwtp15 (*p*=0.0475). In addition, all other gag-recoded passaged viruses from both replicates had raised their replication capacity compared to HIV-Gwtp15 and were as fit as the wild-type passaged virus (*p*>0.05).

All viruses from both replicates except HIV-Gwtp15 had at least a 3.4 fold increase in slopes when compared to passage 0 viruses (*p*<0.05). This result implies that only gag deoptimized viruses were evolving in cell culture due to a high pressure to revert. Even though HIV-GminAp15 replicate 1 (5.79 fold change) and HIV-GminAp15 replicate 2 (4.42 fold change) had improved in 15 passages and had a better fitness when compared to passage 0, these viruses still were statistically attenuated in PBMCs.

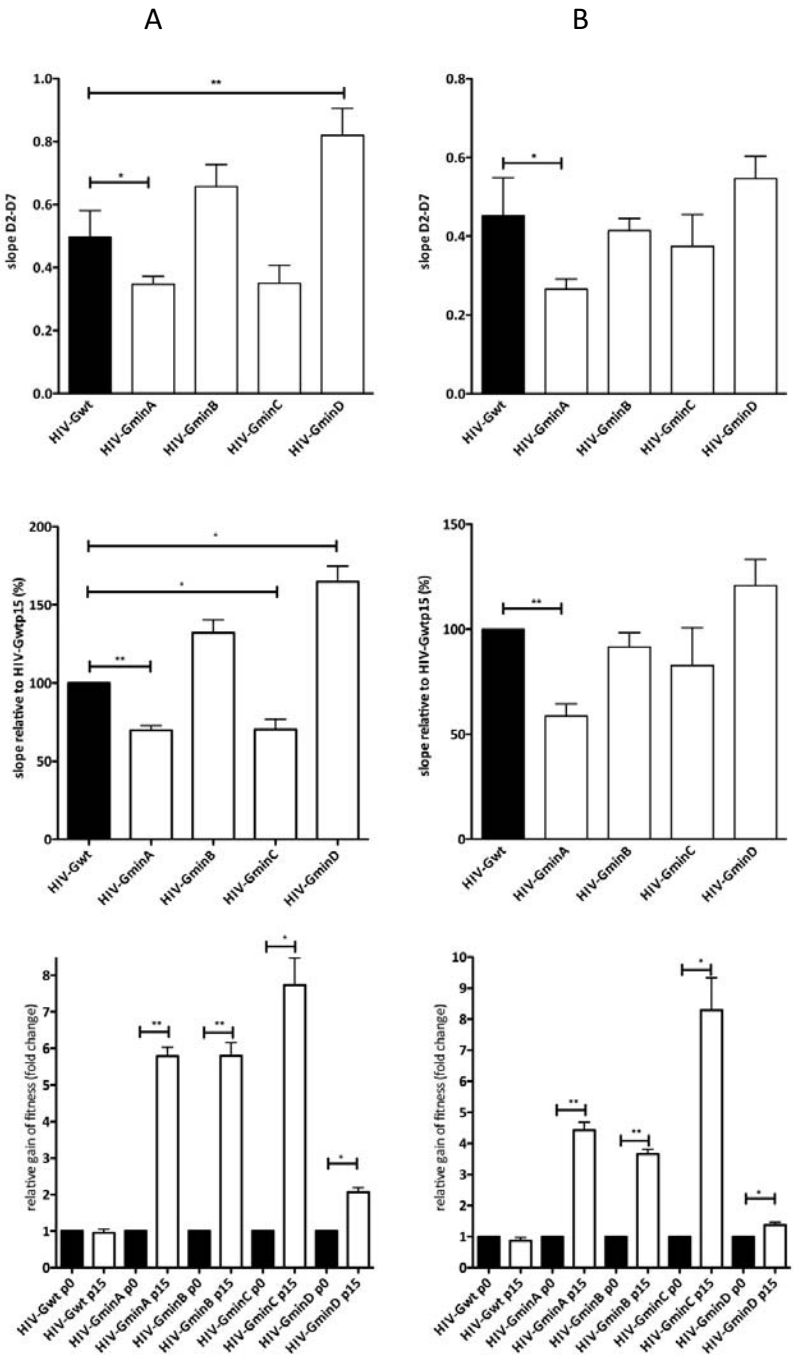


Figure 33. Replication kinetic assay of HIV-1 gag passaged viruses replicate 1 (A and replicate 2 (B) after 15 serial passages in PBMCs. Replication capacity assays were started with a MOI of 0.001. Two replicates were performed for each variant. Production of the HIV-1 antigen p24 in culture supernatants was determined on days 2–7. (Upper Panel) Each bar represents the slope of antigen production for a given virus after infection of PBMCs cells, which provides an estimate of the viral replication capacity. (Middle panel) Comparisons between wild-type (HXB2) and mutant recoded viruses are shown. (Lower panel) Fold increases are shown relative to passage 0. The significance of the difference between slopes was calculated with GraphPrism v. 4 software (unpaired *t*-test). Values represent the mean ± SD from at least three independent experiments.

C. QUASISPECIES ANALYSIS OF PASSAGED VIRUSES.

Genotypic stability of all passaged deoptimized gag viruses was also verified. As in passaged recoded protease viruses, our aim was to know if phenotypic evolutions and gain of replicative capacities observed were due to reversions to wild-type sequences in deoptimized gag viruses. Bulk sequencing of viral extracted genomic RNA revealed that all deoptimized passaged viruses contained mutations in specific positions with possible reversions to initial wild-type sequences.

Clonal sequence analysis showed that all evolved viruses contained synonymous and non-synonymous mutations. Reversion to wild type in replicate 1 and replicate 2 were more abundant in viruses HIV-GminAp15 (7 synonymous and 7 non-synonymous), HIV-GminBp15 (14 synonymous and 15 non-synonymous) and HIV-GminCp15 (6 synonymous and 4 non-synonymous) than in HIV-GminDp15 (1 synonymous and 1 non-synonymous), suggesting again that HIV-GminDp15 virus was not under a high evolutionary pressure (Table 16).

In order to check if all gag passaged deoptimized viruses differed in mutant spectrum complexity, mutation frequencies for the gag gene were analyzed and compared to the wild-type passaged viruses. The mutation frequency, determined as the ratio between the sum of mutations (synonymous and non-synonymous) and total nucleotides analyzed, is shown in Table 17. All CPB recoded viruses except both replicates of HIV-GminDp15, increased their mutation frequency up to a median of 1 mutation every 50 nucleotides (median of $1.95x10^{-2}$). Replicate 1 recoded viruses had a higher mutation frequency than replicate 2 viruses. For instance, HIV-GminBp15 replicate 1 had 11.8 fold times more mutation frequency than HIV-Gwtp15 replicate 1, while HIV-GminBp15 replicate 2 had 6.79 fold times more mutation frequency than its own HIV-Gwtp15. This difference can be explained by a higher mutation frequency of HIV-Gwtp15 in replicate 2 than in replicate 1 .Both replicates of HIV-GminAp15, HIV-GminBp15 and HIV-GminCp15 also had a statistically significant increase in mutation frequency (p<0.0001, Chi square test) when compared to the respective wild-type fragment.

	VIRUS	synonymous (reversions)	non-synonymous	number of positions reverted to wild-type	mutations/nucleotides	freq mut	
replicate 1	HIV-Gwtp15 frA	9	8		17/7875	2,16E-03	x8.65
	HIV-GminAp15	60 (25)	59	7 in 17 clones	119/6375	1,87E-02	
	HIV-Gwtp15 frB	7	11		18/8421	2,14E-03	x11.8
	HIV-GminBp15	186 (172)	5	14 in 19 clones	191/7600	2,51E-02	
	HIV-Gwtp15 frC	24	6		30/8400	3,57E-03	x3.09
	HIV-GminCp15	71 (41)	26	6 in 22 clones	97/8800	1,10E-02	
	HIV-Gwtp15 frD	6	24		30/8190	3,66E-03	x1.63
	HIV-GminDp15	13 (1)	15	1 in 12 clones	28/4680	5,98E-03	
replicate 2	HIV-Gwtp15 frA	9	17		26/6375	4,08E-03	x4.79
	HIV-GminAp15	82 (39)	78	7 in 22 clones	160/8250	1,94E-02	
	HIV-Gwtp15 frB	8	16		24/6000	4,00E-03	x6.79
	HIV-GminBp15	250 (233)	8	15 in 22 clones	250/9200	2,72E-02	
	HIV-Gwtp15 frC	13	8		21/3600	5,83E-03	x2.74
	HIV-GminCp15	45 (20)	19	4 in 10 clones	64/4000	1,60E-02	
	HIV-Gwtp15 frD	19	15		32/5460	5,86E-03	x1.46
	HIV-GminDp15	18 (1)	24	1 in 12 clones	42/4680	8,55E-03	

Table 17. Mutation frequency and types of mutations in the HIV-1 regions of infected passaged cultures. [a] Indicates the number of mutations and nucleotides sequenced for each population. [b] Mutation frequency is the number of total mutations divided by the number of sequenced nucleotides.

The data obtained from the analysis of specific mutations from both replicates showed that some particular positions along gag sequences were more susceptible to revert to wild-type sequence (Table 18 and Table 19). Contrary to what it was observed in protease quasispecies (Chapter 1), evolution of both replicates seemed to explore similar pathways to improve their fitness. For instance, nucleotide positions 378 and 384-387 (6 clear positions in HIV-GminBp15) reverted to wild-type sequence almost in all clones analyzed in both replicates. Although frequency in clones was not exactly the same, both replicates of HIV-GminAp15 shared 8 exact mutations; HIV-GminBp15 replicate 1 and 2 had in common 16 mutations and HIV-GminCp15 possessed 5 exact mutations in replicate 1 and 2. HIV-GminBp15 viruses showed more mutations than other viruses, with 9 reverted mutations. The gag region B seemed to be more prone to reversions to wild-type sequences, probably due to a higher selective pressure to revert. Furthermore, these mutations may explain the increase in fitness observed in MT-4 cells (62% to 95% average in both replicates) and PBMCs (25% to more than 100%) between passage 0 and passage 15 viruses.

Mutations in nucleotide positions 18, 42, 162, 333, 477, 717, 978 and 1110, found in replicates 1 and 2, were synonymous mutations yet not reversions to the wild-type sequence. All those mutations are probably exploring options to recover fitness in a different way than evolving to the wild-type sequence.

In addition, individual CPS in all major mutations (synonymous mutations existing in more than one clone in quasispecies) were recalculated to examine the extent of CPS alterations due to those new mutations. Indeed, all synonymous mutations improved CPS of mutated codon pairs in evolved sequences. All positions analyzed for CPS are highlighted in bold (Table 17 and Table 18).

Table 18. Replicate 1 synonymous mutations in gag-recoded quasispecies.

replicate1	position	starting mutated virus	reversion to wild-type virus	other mutations	nº clones
HIV-GwtfrAp15	87	T	T	C	1/21
	96	A	A	G	1/21
	174	A	A	G	1/21
	201	C	C	T	1/21
	246	A	A	G	1/21
	264	A	A	G	1/21
	279	G	G	A	1/21
	336	G	G	A	1/21
	345	A	A	G	1/21
HIV-GminAp15	**18**	**C**	**A**	**T**	**5/17**
	42	**T**	**T**	**C**	**10/17**
	60	C	G	A	1/17
	66	**C**	**G**		**9/17**
	162	**C**	**A**	**T**	**14/17**
	165	G	A		1/17
	189	**A**	**G**		**2/17**
	201	**T**	**C**		**8/17**
	216	C	A	T	1/17
	258	**C**	**T**		**4/17**
	291	T	C		1/17
	303	C	A		1/17
	333	**G**	**T**	**A**	**3/17**
HIV-GwtfrBp15	417	G	G	A	1/21
	489	T	T	C	1/21
	504	G	G	A	1/21
	621	G	G	A	1/21
	705	C	C	T	1/21
	720	T	T	C	1/21
	755	A	A	C	1/21
HIV-GminBp15	369	A	A	C	1/19
	378	**C**	**T**		**18/19**
	384	**A**	**C**		**18/19**
	385	**T**	**A**		**18/19**
	386	**C**	**G**		**18/19**
	387	**G**	**C**		**18/19**
	399	**G**	**T**		**18/19**
				A	**1/19**
	408	**A**	**G**		**18/19**
	411	**T**	**C**		**18/19**
	429	**G**	**A**		**14/19**
	435	**A**	**G**		**7/19**
	477	C	A	T	1/19
	612	**G**	**C**		**2/19**
				A	**3/19**
	615	T	C		1/19
	687	C	A		1/19
	711	**C**	**A**		**3/19**
	714	A	A	G	1/19
	717	**T**	**T**	**C**	**7/19**
HIV-GwtfrCp15	762	A	A	G	1/21
	870	**G**	**G**	**A**	**21/21**
	903	T	T	C	1/21
	927	T	T	C	1/21
HIV-GminCp15	822	T	A	G	1/22
	843	G	C	A	1/22
	906	G	A		1/22
	921	A	G		1/22
	978	**T**	**G**	**C**	**22/22**
	984	**G**	**A**		**22/22**
	1008	G	A		1/22
	1044	**G**	**A**		**15/22**
	1110	**A**	**T**	**G**	**6/22**
	1113	G	A		1/22
HIV-GwtfrDp15	1146	T	T	C	1/21
	1151	G	G	A	1/21
	1335	A	A	G	1/21
	1380	A	A	G	2/21
	1449	G	G	A	1/21
HIV-GminDp15	1121	C	C	T	1/12
	1170	T	T	A	1/12
	1215	C	C	T	1/12
	1221	G	C	A	1/12
	1224	A	T	C	1/12
	1248	C	T		1/12
	1386	**C**	**C**	**T**	**4/12**
	1395	**T**	**T**	**C**	**3/12**

Table 19. Replicate 2 synonymous mutations in gag-recoded quasispecies

replicate2	position	starting mutated virus	reversion to wild-type virus	other mutations	nº clones
HIV-GwtfrAp15	18	A	A	G	1/17
	144	T	T	G	2/17
	174	A	A	G	1/17
	195	A	A	G	3/17
	294	G	G	A	1/17
	318	A	A	G	1/17
HIV-GminAp15	18	C	A	T	1/22
	42	**T**	**T**	**C**	**16/22**
	51	G	A		1/22
	66	**C**	**G**		**14/22**
	69	C	A	T	1/22
	90	**G**	**A**		**3/22**
	123	C	A	G	1/22
	162	**C**	**A**	**T**	**20/22**
	201	**T**	**C**		**15/22**
	258	C	T		1/22
	291	**C**	**T**		**2/22**
	306	**T**	**C**		**3/22**
	333	**G**	**T**	**A**	**4/22**
HIV-GwtfrBp15	393	T	T	C	1/15
	405	G	G	A	1/15
	420	G	G	A	1/15
	453	T	T	C	1/15
	477	G	G	A	1/15
	495	C	C	T	1/15
	675	A	A	G	1/15
	711	A	A	G	1/15
HIV-GminBp15	**378**	**C**	**T**		
	384	**A**	**C**		**23/23**
	385	**T**	**A**		**23/23**
	386	**C**	**G**		**23/23**
	387	**G**	**C**		**23/23**
	399	**G**	**T**		**23/23**
	408	**A**	**G**		**23/23**
	411	**T**	**C**		**23/23**
	429	**G**	**A**		**22/23**
	435	**A**	**G**		**14/23**
	471	G	A		1/23
	477	C	A	T	1/23
	510	**G**	**C**	**A**	**2/23**
	534	**A**	**A**	**G**	**2/23**
	558	**G**	**C**	**A**	**3/23**
	612	**G**	**C**		**2/23**
				A	**5/23**
	615	**T**	**C**		**3/23**
	639	T	T	C	1/23
	687	**C**	**A**		**6/23**
	711	C	T		1/23
	717	**T**	**T**	**C**	**3/23**

replicate2	position	starting mutated virus	reversion to wild-type virus	other mutations	nº clones
HIV-GwtfrCp15	822	A		G	1/9
	870	G		A	9/9
	915	A		G	1/9
	918	C		T	1/9
HIV-GminCp15	**807**	**A**	**A**	**G**	**10/10**
	864	G	A	C	1/10
	867	G	A	C	1/10
	876	G	C		1/10
	895	C	C	A	1/10
	906	G	A		1/10
	978	**T**	**G**	**C**	**10/10**
	984	**G**	**A**		**10/10**
	1044	**G**	**A**		**8/10**
	1110	**A**	**A**	**G**	**2/10**
HIV-GwtfrDp15	1176	T	T	C	1/14
	1197	G	G	A	1/14
	1405	A	A	G	3/14
	1470	G	G	A	14/14
HIV-GminDp15	1144	A	A	DEL	1/12
	1145	A	A	DEL	1/12
	1146	C	T	DEL	1/12
		C	T		1/12
	1147	T	T	DEL	1/12
	1148	T	T	DEL	1/12
	1149	T	T	DEL	1/12
	1150	C	A	DEL	1/12
	1151	G	G	DEL	1/12
	1152	G	G	DEL	1/12
	1153	A	A	DEL	1/12
	1154	A	A	DEL	1/12
				G	1/12
	1155	T	C	DEL	1/12
	1188	A	C	T	1/12
	1203	C	A	T	1/12
	1221	G	C	A	1/12
	1413	T	T	C	1/12
	1470	**G**	**G**	**A**	**12/12**

Positions selected in bold are the ones present in more than one clone.

Specifically, in HIV-GminAp15 viruses from both replicates, all mutations improved the CPS median of newly recoded codon pairs with the exception of those affecting codon pairs 62-63-64 and 85-86-87 (Figure 34). For the most part, the calculated minimum CPS changed from one with high levels of under-representation to one with neutral or less levels of under-representation. For example, as shown in Figure 34, a single G to A mutation present in codon pair positions 110-111 and 111-112 produced a double amelioration from -0.008 to 0.057 and from -0.788 to 0.168, resulting in a global improvement of 1.018 in the median of both CPS. Once again, and as observed in protease recoded and evolved viruses, HIV-GminAp15 viruses tried to remove and change nucleotides that generated low CPS, as in codon pairs 5-6 and 6-7 or 13-14 and 14-15 (Figure 31). Overall, given the ameliorations

due to a single nucleotide mutation for each codon pair analyzed, the median of enhancement of the sums of CPS was 0.983 for HIV-GminAp15 replicate 1 virus and 1.147 for HIV-GminAp15 replicate 2 virus. These results could imply that these viruses were evolving not always toward a wild-type mutation, but to a recovery of a better CPS.

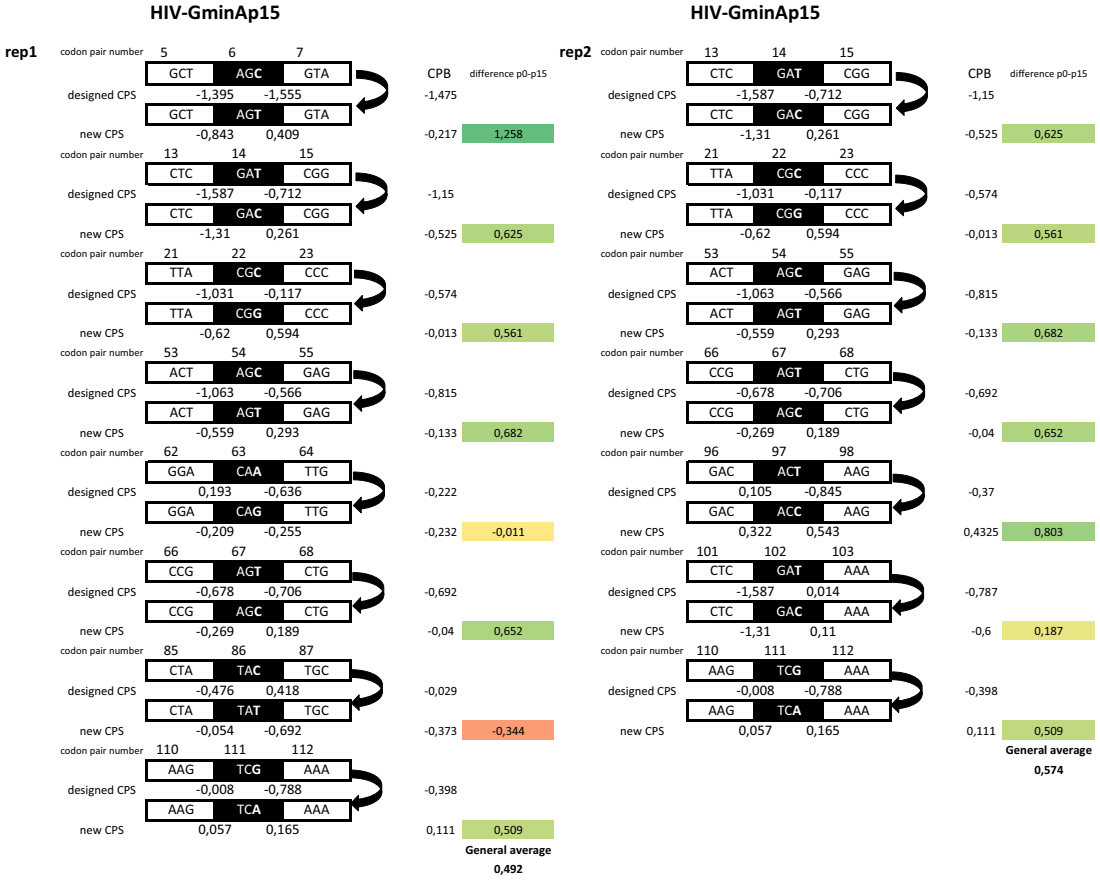


Figure 34. HIV-GminAp15 replicates 1 and 2: mutations found in quasispecies of redesigned culture-passaged viruses. Specific mutations are depicted in the dark square, in white bold. Negative CPS values relate to under-represented codon-pairs, while positive CPS values relate to over-represented codon-pairs. CPB is the average of two CPS. The difference between passage 0 and passage 15 follow a colour gradient: the more greener they are, the more the difference has increased the CPB; the more red they are, the more the difference has decreased the CPB.

As mentioned above, HIV-GminBp15 virus had a higher number of synonymous mutations, implying a greater evolution toward the wild-type sequences. Individual CPS showed that most mutations produced an enhancement of the CPS medians with the exception of codon pairs 125-126 and 126-127 from both replicates. For instance, the increase in codon pairs 238-239-240 was up to 1.235 (from 0.095 and -0.795 to 0.27 and 0.582) (Figure 35 and Figure 36). The median of new CPS due to mutations was 0.239 in replicate 1 and 0.409 in replicate 2. These increases were not as high as the ones observed in HIV-GminAp15 viruses. Nevertheless, it is highly probable that the CPS is not the only factor in

the improvement of these viruses; the number of mutations and the reversion to wild-type sequences should also be considered.

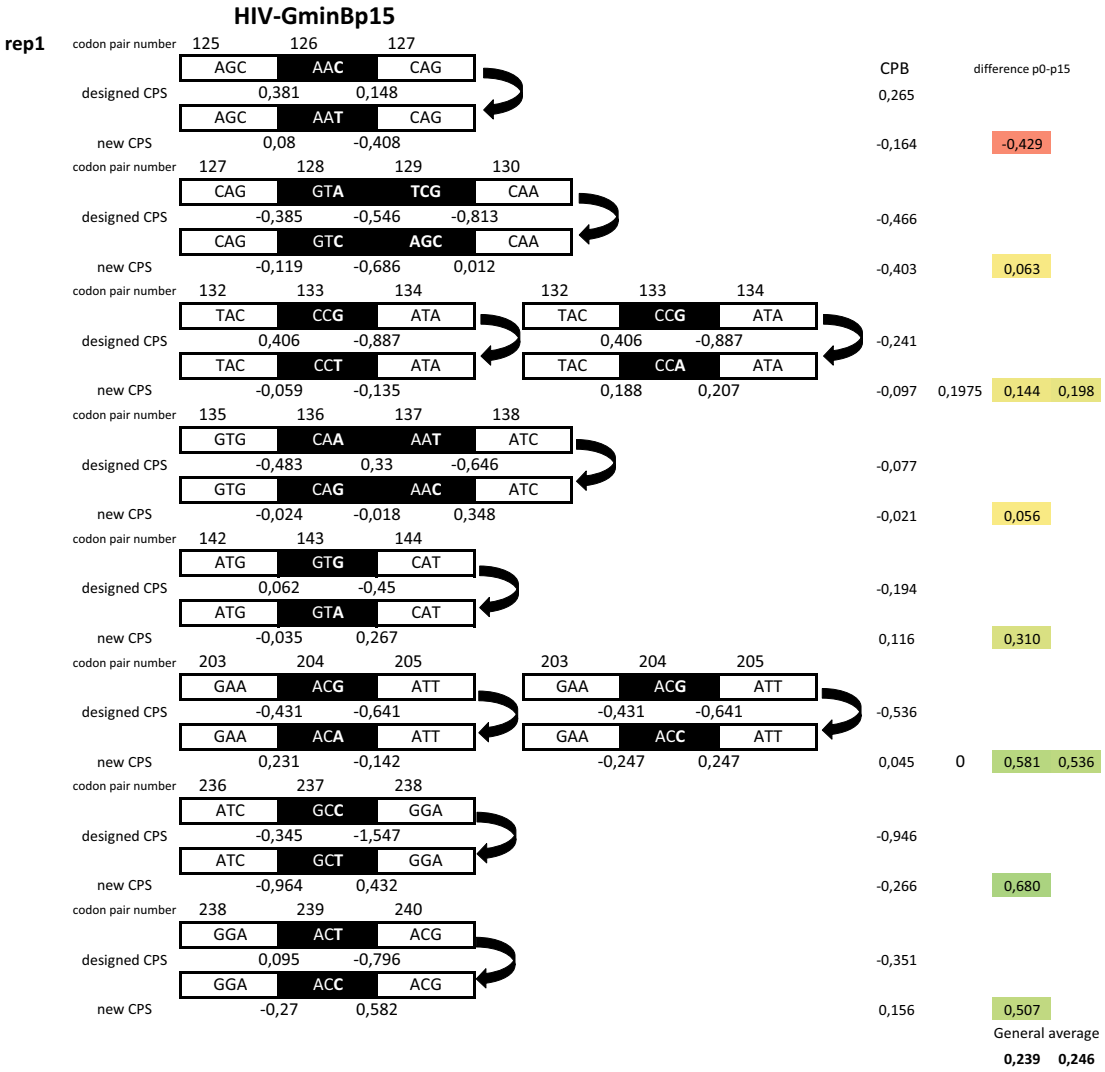


Figure 35. HIV-GminBp15 replicate 1 mutations found in quasispecies of redesigned culture-passaged viruses. Specific mutations are depicted in the dark square, in white bold. Negative CPS values relate to under-represented codon-pairs, while positive CPS values relate to over-represented codon-pairs. CPB is the average of two CPS. The difference between passage 0 and passage 15 follow a colour gradient: the more greener they are, the more the difference has increased the CPB; the more red they are, the more the difference has decreased the CPB.

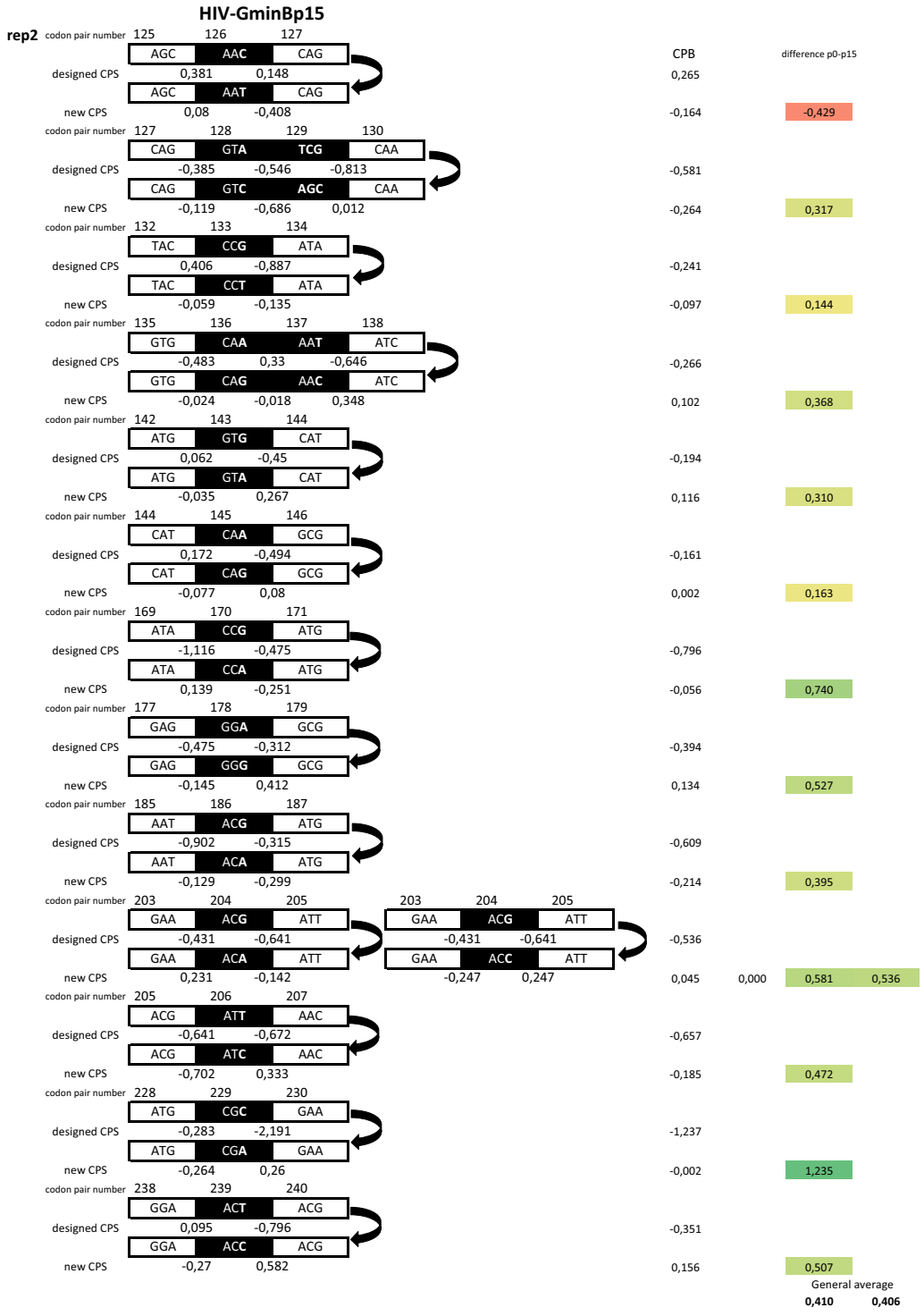


Figure 36. HIV-GminBp15 replicate 2 mutations found in quasispecies of redesigned culture-passaged viruses. Specific mutations are depicted in the dark square, in white bold. Negative CPS values relate to under-represented codon-pairs, while positive CPS values relate to over-represented codon-pairs. CPB is the average of two CPS. The difference between passage 0 and passage 15 follow a colour gradient: the more greener they are, the more the difference has increased the CPB; the more red they are, the more the difference has decreased the CPB.

All synonymous mutations in HIV-GminCp15 replicate 1 and 2 caused increases in CPS affected by them (Figure 37). These mutations were found in most clones, indicating that possibly they were one of the main factors in the fitness improvement of HIV-GminCp15 viruses in MT-4 cells and in PBMCs. The CPS median increase caused by mutations was 0.53 in replicate 1 and 0.60 in replicate 2.

Finally, mutations found in both replicates of HIV-GminDp15 virus were not reoptimizing CPS, but they caused the development of different codon-pairs with a higher degree of under-representation (Figure 34).

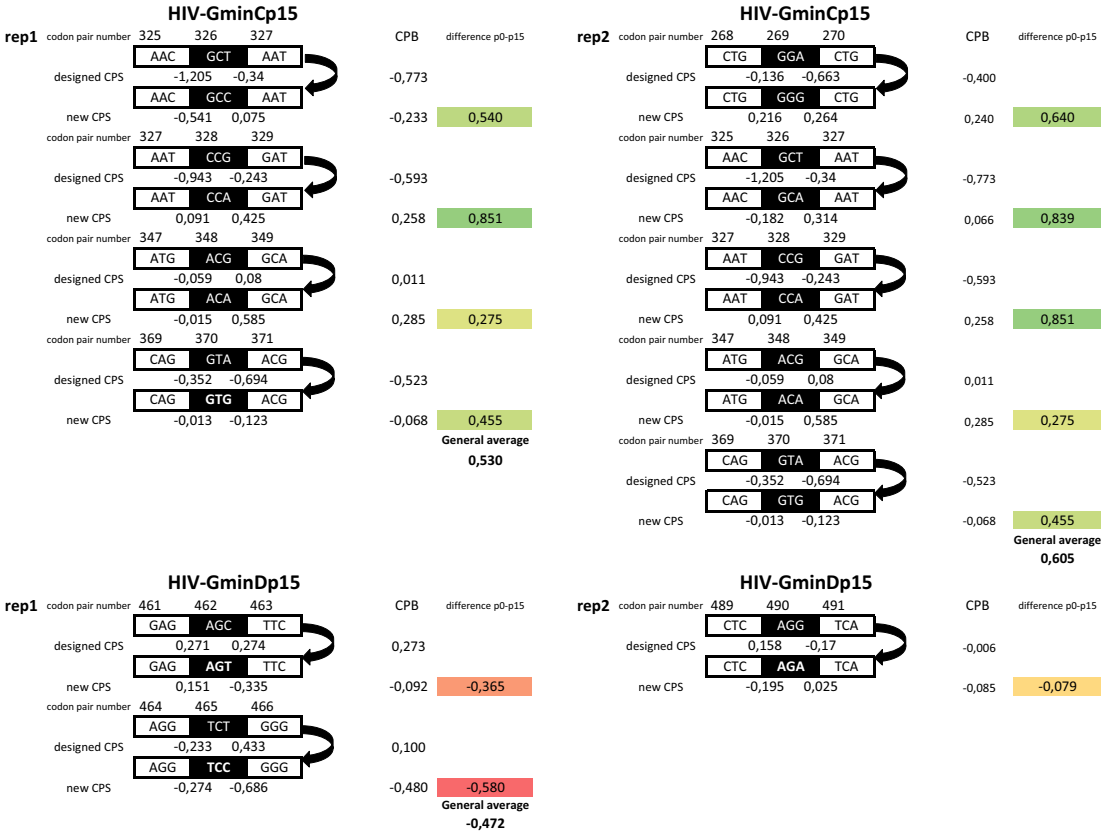


Figure 37. HIV-GminCp15 and HIV-GminDp15 replicate 1 and replicate 2 mutations found in quasispecies of redesigned culture-passaged viruses. Specific mutations are depicted in the dark square, in white bold. Negative CPS values relate to under-represented codon-pairs, while positive CPS values relate to over-represented codon-pairs. CPB is the average of two CPS. The difference between passage 0 and passage 15 follow a colour gradient: the more greener they are, the more the difference has increased the CPB; the more red they are, the more the difference has decreased the CPB.

Non-synonymous mutations may have some influence in replication capacities of evolved viruses (Table 20).

Table 20. Replicate 1 and replicate 2 non-synonymous mutations present in the quasispecies.

replicate1	mutation	nº clones
HIV-GwtfrAp15	G11R	1/21
	G11E	1/21
	K18R	2/21
	G49D	1/21
	C87Y	1/21
	K110E	1/21
	A119T	1/21
HIV-GminAp15	R4G	6/17
	R15Q	1/17
	E40K	1/17
	V46I	3/17
	P48T	14/17
	G49E	1/17
	I60V	5/17
	V82I	1/17
	T84A	7/17
	E105K	2/17
	K114R	17/17
	A115V	1/17
HIV-GwtfrpBp15	N126Y	1/21
	N126S	1/21
	A146T	1/21
	Q199R	1/21
	V215A	6/21
	T251A	1/21
HIV-GminBp15	V164I	1/19
	V173I	4/19
HIV-GwtfrCp15	T251A	1/21
	N525T	1/21
	S281N	1/21
	Q287R	1/21
	T320A	1/21
	A341V	1/21
HIV-GminCp15	R286C	1/22
	T320A	1/22
	S368R	1/22
	V370A	16/22
	T371M	7/22
Hiv-GwtfrDp15	F383I	1/21
	R384K	1/21
	V390G	15/21
	T401I	1/21
	R409K	2/21
	M423L	1/21
	S451N	1/21
	P453T	1/21
	T469A	1/21
HIV-GminDp15	M377T	1/12
	F383S	1/12
	N385S	1/12
	V390I	1/12
	K391R	1/12
	T401I	1/12
	N404S	2/12
	R406G	6/12
	A407V	1/12

replicate2	mutation	nº clones
HIV-GwtfrAp15	E17Q	1/17
	K18R	2/17
	P23A	2/17
	W36G	2/17
	L50R	2/17
	L61R	2/17
	C87W	2/17
	V88G	2/17
	K112R	1/17
HIV-GminAp15	A3T	1/22
	R4G	15/22
	K30R	1/22
	V35I	1/22
	E40G	1/22
	V46I	2/22
	P48T	20/22
	I60V	1/22
	T84A	5/22
	E105K	2/22
	N109S	1/22
	K110R	1/22
	K113T	1/22
	K113N	1/22
	K114R	22/22
	A115V	1/22
	Q117L	1/22
	A119V	1/22
HIV-GwtfrBp15	N126S	1/15
	V128G	2/15
	V135G	2/15
	M142R	2/15
	N153K	2/15
	W155Q	2/15
	Q199A	1/15
	V215A	3/15
HIV-GminBp15	V128I	1/23
	I138T	1/23
	A174T	1/23
	V191I	4/23
	STOP 212	1/23

replicate2	mutation	nº clones
HIV-GwtfrCp15	A341S	1/9
	A341V	1/9
	G357S	2/9
	G357R	1/9
	S373T	1/9
HIV-GminCp15	K290E	1/10
	F293V	1/10
	V297L	1/10
	STOP 301	1/10
	A340V	1/10
	V370A	8/10
	T371M	5/10
	A374T	1/10
HIV-GwtfrDp15	S374T	1/14
	T376I	1/14
	R386K	1/14
	V391A	1/14
	V391G	4/14
	H422P	1/14
	M424L	2/14
	STOP 435	1/14
	L450V	1/14
	E455G	1/14
	T502A	1/14
HIV-GminDp15	N382F	1/12
	N385S	1/12
	I389V	1/12
	V390I	2/12
	E398G	1/12
	N404S	3/12
	R406G	6/12
	A407T	1/12
	R409S	1/12
	E419K	1/12
	E428D	1/12
	N432I	1/12
	A457S	1/12
	F463S	1/12
	S465P	1/12
	T470I	1/12

Especially interesting was the high number of nonsynonymous substitutions found in some deoptimized gag populations (Table 20). For instance, the HIV-GminA construct incorporated as majority variants the nonsynonymous mutations R4G, P48T and K114R in both passage replicates, indicating that these mutations were adaptive. The selection of these mutations or other mutations was not observed in the wild-type virus. The R4 and P48 gag positions are highly conserved in HIV-1. The R4G mutation was rarely observed in nature and the P48T mutation was not found in the alignment of 4644 gag sequences from Los Alamos database (http://www.hiv.lanl.gov/). Most of the nonsynonymous substitutions occurred in codons carrying deoptimized synonymous mutations, showing that the introduction of synonymous substitutions may allow the virus to explore an alternative nonsynonymous sequence space to improve viral fitness(Lauring *et al.*, 2012).

Regarding the base composition of evolved viruses, HIV-GminAp15, HIV-GminBp15 and HIV-GminCp15 from both replicates had an increased number of G to A and C to T mutations when compared to the control viruses (Table 21. These viruses carried rare mutations, as G to C or C to G transversions, unexistent in wild-type passaged viruses. Those mutations were probably fixed in the viral populations due to the pressure being exerted by reversions in gag-recoded viruses. Mutation panels in HIV-GminDp15 virus were similar to the wild type passaged viruses. HIV-GminDp15 virus was not evolving at the same rate than other deoptimized viruses, probably because this variant was not under a selective pressure to revert.

Replicate	*Virus*		*Transitions*[b]				*Transversions*[c]						
		G→A + C→T[a]	***A→G***	***G→A***	***T→C***	***C→T***	***A→C***	***A→T***	***G→C***	***G→T***	***C→A***	***C→G***	***T→A***	***T→G***
1	***HIV-Gwtp15 frA***	9	7	8	1	1	0	0	0	0	0	0	0	0
	HIV-GminAp15	38	37 (2)	13 (1)	19 (9)	25 (4)	0	0	0	0	16 (1)	9 (9)	0	0
	HIV-Gwtp15 frB	5	3	4	8	1	1	1	0	0	0	0	0	0
	HIV-GminBp15	39	27 (25)	21 (14)	26 (19)	18 (18)	18 (18)	0	20 (20)	18 (18)	5 (4)	18 (18)	18 (18)	0
	HIV-Gwtp15 frC	23	4	22	2	1	1	0	0	0	0	0	0	0
	HIV-GminCp15	49	8 (1)	41 (40)	38	8	1	0	0	0	1	0	0	1
	HIV-Gwtp15 frD	17	5	16	2	1	1	2	0	0	0	1	2	4
	HIV-GminDp15	10	10	2	5	8 (1)	1	0	0	0	0	1	1	0
2	***HIV-Gwtp15 frA***	0	9	1	0	0	0	0	0	1	3	0	0	12
	HIV-GminAp15	56	31	30 (4)	34 (18)	26 (3)	1	1	0	0	20	16 (14)	0	0
	HIV-Gwtp15 frB	4	9	1	0	0	0	0	0	0	1	3	0	12
	HIV-GminBp15	65	39 (37)	40 (23)	31 (26)	25 (24)	23 (23)	0	25 (25)	23 (23)	6 (6)	23 (23)	23 (23)	0
	HIV-Gwtp15 frC	14	3	11	0	3	,	0	1	1	0	0	2	0
	HIV-GminCp15	28	13	22 (20)	18	6	0	0	4 (1)	0	1	0	1	1
	HIV-Gwtp15 frD	4	3	3	1	1	1	1	0	0	0	1	2	1
	HIV-GminDp15	9	12	6	1	3	2	2	0	1	0	1	0	0

Table 21. Mutation frequency and types of mutations in the HIV-1 respective gag-coding region of passaged viruses.

3. CODON PAIR BIAS EFFECT ON VIRAL RNA STRUCTURE AND CODON USAGE

A. RNA STRUCTURE

Similarly to protease, the RNA free energy of the new gag constructs was calculated. Analysis of this RNA free energy was performed the same way than in protease constructs.

Figure 38 shows the corresponding free energy for every 100-base pair portion tested. Applying the analysis above described to the protease gene, all of CPB recoded gag constructs had a mean and individual RNA free binding energy higher than -30kcal/mol and were similar to the wild type. The wild-type region already contained one domain with a folding RNA free energy below -30kcal/mol and a mean of ΔG=-18.75 kcal/mol. Since this extreme region already existed in the wild-type gag coding region, no corrections were done to alleviate it, because probably it was important for structure. Recoded Gag constructs did not produce new secondary RNA structures.

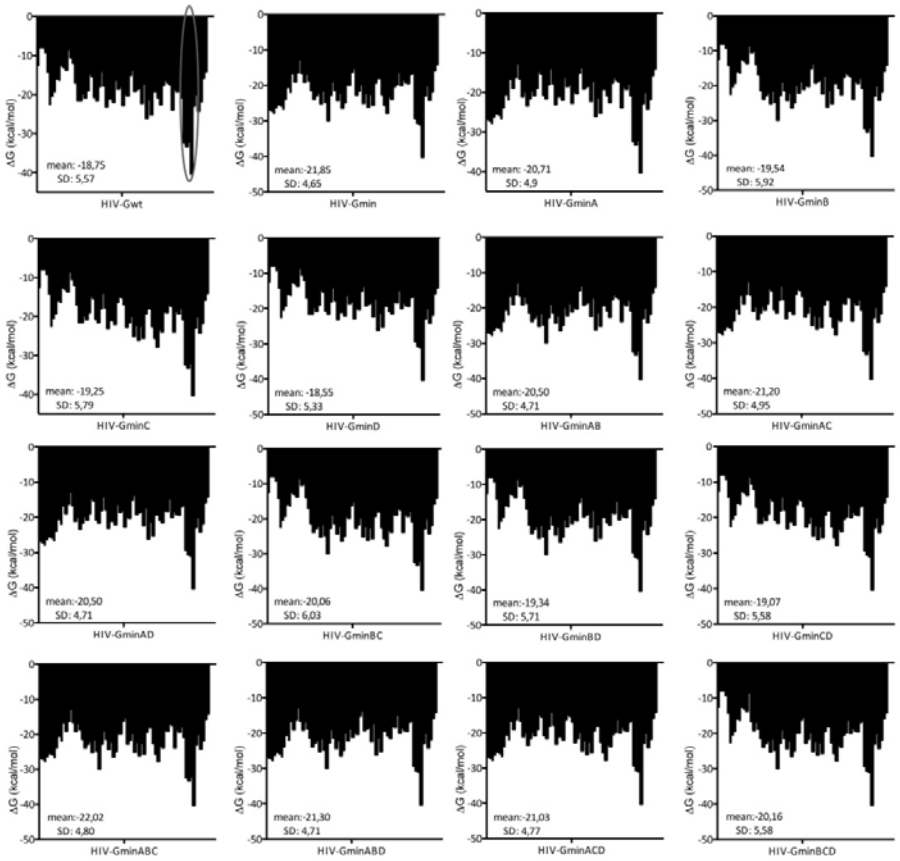


Figure 38. Folding free energy of synthetic and wild type gag constructs. The calculated energy value of each scanned 100-base pair segment was output to produce the graph. The mean folding energy for all constructs were: HIV-Gwt, -18.75; HIV-Gmin, -21.85; HIV-GminA, -20.71; HIV-GminB, -19.54; HIV-GminC, -19.25; HIV-GminD, -18.55; HIV-GminAB, -20.50; HIV-GminAC, -21.20; HIV-GminAD, -20.50; HIV-GminBC, -20.06; HIV-GminBD, -19.34; HIV-GminCD, -19.07; HIV-GminABC, -22.02; HIV-GminABD, -21.30; HIV-GminACD, -21.03; HIV-GminBCD, -20.16

B. CODON USAGE

As it was done with the protease coding region, the application of SAVE technology to HIV-1-gag coding region implies that codon usage bias had to remain unaffected. ENC was used to quantify the changes in codon usage bias.

Our results showed that ENC values were similar in all gag constructs, and none of the values were under 35. These results suggest that codon usage bias in CPB gag recoded constructs persisted without any alteration (Table 22).

Table 22. ENC values and GC content in third position and overall in percentages of and gag constructs

	ENC	GC3s (%)	GC (%)
HIV-Gwt	45,33	35,2	44,2
HIV-GminA	50,56	40,0	46,3
HIV-GminB	49,81	38,8	45,7
HIV-GminC	52,28	41,0	46,8
HIV-GminD	47,94	36,0	44,9
HIV-GminAB	52,27	43,5	47,7
HIV-GminAC	54,68	45,8	48,8
HIV-GminAD	52,53	40,8	46,3
HIV-GminBC	53,95	44,6	48,3
HIV-GminBD	53,03	39,6	46,3
HIV-GminCD	53,26	41,9	47,5
HIV-GminABC	52,44	49,4	50,3
HIV-GminABD	54,15	44,4	48,4
HIV-GminACD	54,10	46,7	49,5
HIV-GminBCD	54,95	45,4	48,9
HIV-Gmin	51,72	50,2	51,0

GC content and GC% at the third codon position (GC3s%) were also analyzed in gag coding regions. Although the overall GC content remained similar in all constructs to the corresponding wild type region (varying from 44.2% to 51%), $GC_3s\%$ was variable (Table 22). HIV-Gmin, the utmost deoptimized and non-viable virus, had the lowest $GC_3s\%$ in its genome (50.2%) compared to the wild type virus (35.2%). The correlation between $GC_3s\%$ and virus replication capacities was tested (Figure 39).

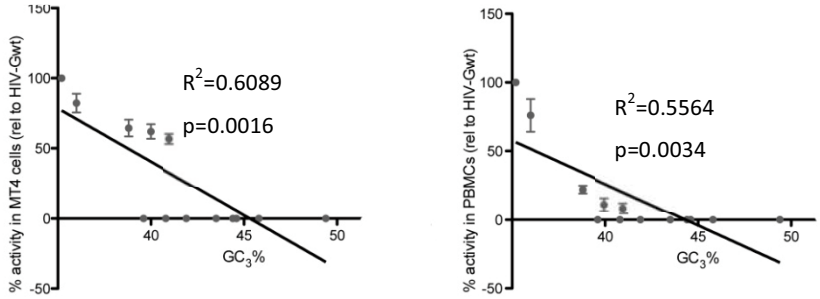


Figure 39. Relation between GCs in third position of the codon and activity in MT-4 cells (1) or PBMCs (2) in gag constructs. Replication capacities relatives to HIV-Gwt corresponding to each protease construct (if the virus was non-viable, the replication capacity was assumed to be 0%) were correlated with GC3% found in all sequences. All statistics tests were performed with GraphPrism v. 4 software.

In the gag gene, a highly statistically significant correlation between GC_3s% and viral replication capacities was found in MT-4 cells and PBMCs (*p*=0.0016 and *p*=0.0034 respectively). Differences between protease and gag results were probably due to the larger length of gag compared to protease, and to the higher number of mutations introduced in gag regions.

CpG islands presence was correlated to replication capacities of gag viable viruses in vitro. (Figure 40).

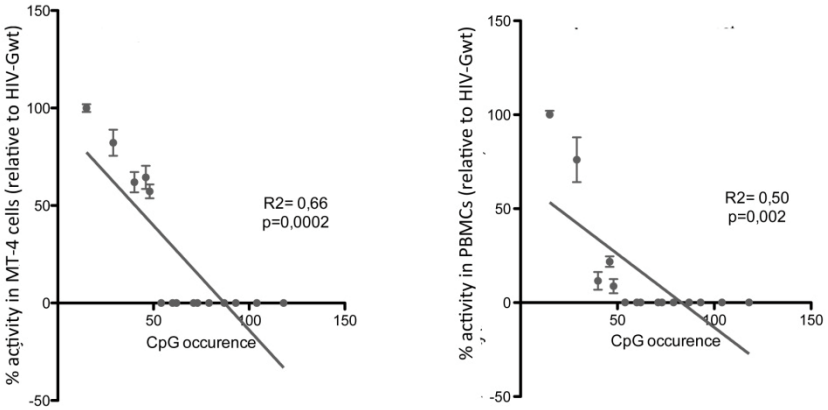


Figure 40. Relation between CpG islands and activity in MT-4 cells (A) or PBMCs (B) in gag constructs. Replication capacities relatives to HIV-Gwt corresponding to each protease construct (if the virus was non-viable, the replication capacity was assumed to be 0%) were correlated with CpG occurrence found in all sequences. All statistics tests were calculated with GraphPrism v. 4 software.

As in protease, a negative correlation was found in gag between the presence of CpG islands and the reduction of replication capacities when compared to wild type virus. This negative correlation was probably due to the exclusion of CpG dinucleotides in eukaryotes, mainly in NNC_3-G_1NN codon pairs, while UpG and CpA islands were strongly preferred (Moura *et al.*, 2007).

DISCUSSION AND PERSPECTIVES

Organism codon pairings are non-random and highly species-specific (Moura *et al.*, 2007). Important differences between eukaryotic and prokaryotic codon pair usage have been described. In eubacteria and archaea, the translational machinery influences codon-pair context, while in eukaryotes, DNA methylation and tri-nucleotide repeats are probably the main influences in codon-pair usage (Tats *et al.*, 2008). Viruses need and use host cell machinery to replicate; however their codon usage might differ from their host. HIV-1, a virus with high replication rate when the infected cell is activated, possesses a significantly different codon usage frequency compared to that of humans, mainly .in late expressing genes (gag, pol, env, vif, vpu and vpr) which carry codons belonging to human rare codon usage (Pandit & Sinha, 2011).

Prior work has documented the effectiveness of making changes to the codon-pair bias of viral genomes in order to generate attenuated *Poliovirus* and *Influenza* virus (Coleman et al., 2008; Mueller et al., 2010). Similarly, recodification of *Streptococcus pneumoniae* pseudolysin gene showed that tuning gene expression by codon-pair bias produced a bacteria live-attenuated vaccine (Coleman *et al.*, 2011). We aimed here the application of this strategy for modifying the replicative capacity of HIV-1. The fact that very different viruses, such as *Poliovirus* and *Influenza* virus are highly sensitive to codon-pair deoptimization suggests that this strategy may operate quite generally. However, no data were available regarding whether this approach would work with lentiviruses, which, in contrast to poliovirus and influenza virus, integrate into the host cell genome. In this study, we tested the extent to which codon-pair deoptimization and reoptimization of the HIV-1 genome could create viruses with altered replication capacity in cell culture.

The extension of codon-pair deoptimization and reoptimization of HIV-1 genome and the resulting generation of viruses with altered replication capacities in cell cultures were tested. Both protease and gag coding regions were re-engineered to change CPB. All those new constructs expressed the identical wild-type amino acid sequence but infrequently used codon pairs were used. We recodified different moieties of HIV-1 gag and pol genes while conserving the CPB and other qualities of the wild-type virus genome. Indeed, the CPB scores of wild-type gag (0.035) and protease (0.069) HXB-2 showed that gag and protease coding regions (both late expressing genes) had a highly similar codon pair usage when compared to the

human one. Probably evolution has driven HIV-1 towards an optimization of its codon pair usage in order to match the human one.

Three protease-deoptimized constructs, HIV-PminAB, HIV-PminAC and HIV-Pmin did not produce viable viruses. All those viruses contained CPB scores below -0.256. Low CPB score was not the only reason of non-viability since HIV-PminBC, a virus with a CPB score of -0.317, was found to be functional. All the non-viable protease deoptimized constructs had in common the deoptimized A region plus one or two of the other deoptimized regions, suggesting that the cause of non-viability might be due to a sum of highly deoptimized codon pairs (with low CPS). In gag, up to 342 synonymous silent mutations were introduced, and we re-engineered 15 gradually deoptimized different constructs, which had CPB scores ranging from -0.011 to -0.348. Apart from the wild-type virus (HIV-Gwt), four deoptimized viruses, each carrying only one recoded region, were viable. Interestingly, in gag region, CPB scores were lower when compared to protease CPB scores. Gag viruses with a CPB score below -0.113 were not viable. In human codon pair usage only a few genes existed with a score below -0.1 (Coleman *et al.*, 2008), confirming that evolution has selectively abrogated those high under-represented codon-pair constructs.

Overall, we demonstrated in this thesis that it was possible to produce viruses with statistically significantly lower viral replication capacity in MT-4 cells and PBMCs by codon pair-deoptimizing different moieties of the HIV-1 gag and protease coding regions. Gag CPB recoded viruses had greater levels of attenuation than protease CPB recoded viruses. This difference can be due to either the higher number of mutations introduced in gag regions or to the higher expression levels of gag compared to protease in infected cells (2000:1). Remarkably, in protease constructs, only HIV-PminA virus was deoptimized. HIV-PminA contained in its sequence 6 critical low CPS. Since viruses HIV-PminAB, HIV-PminAC and HIV-Pmin were not functional, and HIV-PminA virus was the only attenuated virus, we hypothesize that the A region might probably contain too many under-represented codon pairs. CPS of those under-represented codon pairs combined with other deoptimized regions, yielded non-viable viruses. In a laboratory stable cell line, MT-4 cells, and in PBMCs, HIV-Pmax fitness was indistinguishable from wild-type virus, suggesting that a neutral genetic drift may be operating in synonymous nucleotide positions.

Our model indicates that: 1) constructs with too many under-represented codon pairs leaded to non-viable viruses and 2) construct with over-represented codon

pairs produced a similar virus in fitness when compared to the wild-type virus. Therefore, we suggest that not only CPB is different in eukaryotic cells than in prokaryotes, but also that the effects of CPB are reverted. These results are in concordance with previous work from Wimmer groups in *Poliovirus* (Coleman *et al.*, 2008; Mueller *et al.*, 2010) and to the Hatfield groups, which proposed that over-represented codon-pairs reduced translation and vice-versa in *E.coli* (Gutman & Hatfield, 1989; Irwin *et al.*, 1995).

Further specific infectivity analysis demonstrated that a reduction in specific infectivity was occurring in all attenuated viruses, confirming that low-fitness viruses possess lower TCID50/pg p24 ratios (Martinez-Picado *et al.*, 1999). It is assumed that extracellular p24 amounts in cell culture are an indirect measurement of total viral production. HIV-Pwt had a high ratio of TCID50 to p24, implying high levels of viable viruses (1 TCID50/ml: 2pg p24 antigen). In the attenuated HIV-PminA virus, this ratio was lowered to 1 TCID50/ml: 42pg/ml p24 antigen, indicating that the success rate of producing new infections was 20 times lower than the one in HIV-Pwt. HIV-GminA, HIV-GminB and HIV-GminC attenuated viruses had 10-40 times less opportunities to initiate a new infectious cycle when compared to HIV-Gwt; their infectivity ratios ranged from 1 TCID50/ml: 103.44pg p24/ml to 1 TCID50/ml: 486.5pg p24/ml. Therefore, as a result of CPB customization, attenuated viruses had a reduced capacity to spread and start new infection cycles.

The molecular mechanism involved in protease codon-pair deoptimized attenuated viruses was analyzed. All works performed with codon-pair deoptimization suggested that translational machinery was affected (Boycheva *et al.*, 2003; Buchan, 2006; Coleman *et al.*, 2008; 2011; Irwin *et al.*, 1995; Moura *et al.*, 2005; 2007). Our results provided compelling evidence that attenuated viruses displayed decreased rates of protein translation, confirming *Poliovirus* and *Influenza* virus results (Coleman *et al.*, 2008; Mueller *et al.*, 2010). By cloning recoded proteases in a one-cycle virus and checking for mRNA and protein expression, we aimed to test if either transcription or translation was the mechanism involved in HIV-1 codon pair deoptimization. After restricting viral replication to a single cycle, a significant reduction in protein production was observed in the vector carrying the attenuated virus variant. This reduction in protein synthesis was not accompanied by a reduction in the transcript copy number. Amino acid sequence, open reading frames and codon-usage remained constant. These results strongly suggests that translation, and not transcription, is implicated in the generation of the attenuated

phenotype by codon pair recoding. Our results therefore demonstrate that synonymous codon pair reengineering is an effective tool to phenotypically affect the replicative properties of HIV-1.

Phenotypic stability of attenuated viruses and also of reoptimized protease virus HIV-Pmax were tested in serial passages in MT-4 cells. MT-4 cell line was chosen because of its higher viral replication and cell turnover rates compared to PMBCs. Infectivity-to-virion particle ratios of protease and gag passaged viruses improved when compared to passage 0 viruses, suggesting that passaged viruses probably evolved to recover replicative capacity. With the exception of attenuated viruses HIV-PminA replicate 2, HIV-GminA replicate 1 and 2 and HIV-GminC replicate 1, replication capacity assays in MT-4 cells and PBMCs confirmed that viruses regained fitness. In 15 passages (approximately 60 days), viruses that at passage 0 had a 10-20% replication capacity when compared to wild-type virus in PBMCs, improved up to an 80-100% replication capacity relative to wild-type virus. Previously, it was found that a CPB deoptimized poliovirus variant containing 224 synonymous substitutions did not revert its phenotype after 19 passages in HeLaR19 cells (Coleman *et al.*, 2008). In contrast, we show that highly attenuated HIV-1 variants carrying one hundred or more synonymous substitutions reverted their phenotype after 15 passages in MT-4 cells. Life cycle differences between poliovirus and HIV-1 may account for the above discrepancies. Although differences in virus protein or genome mutational robustness and evolvability cannot be discarded (Lauring *et al.*, 2012; 2013).

Quasispecies analyses of passaged viruses were performed to investigate if those phenotype changes were due to reversion of initial mutations or to the appearance of new mutations. In HIV-PminAp15 replicate 1 passaged virus, which evolved to a wild-type phenotype, only 3 key positions were reverted to wild type, which represented the 91% of all synonymous mutations. Specifically, two of those three reversions implied the abrogation of two important under-represented codon pairs. Those 3 positions were not reverted in HIV-PminAp15 virus replicate 2, and this virus still maintained an attenuated phenotype in PMBCs. These two replicates followed two different pathways: HIV-PminAp15 replicate 1 virus reverted to wild-type sequence in 15 passages, while replicate 2 HIV-PminAp15 virus still maintained the recoded under-represented sequence and phenotype in PBMCs. Mutations in culture are stochastic (Drake & Holland, 1999; Manrubia *et al.*, 2005; Read *et al.*,

2012; Ruiz-Jarabo *et al.*, 2003) and probably, more passages should be carried out in order to allow reversion of replicate 2 HIV-PminAp15 to a wild-type phenotype .

Both replicates of gag passaged viruses HIV-GminA, HIV-GminB and HIV-GminC had an increased significant mutation frequency when compared to wild-type virus. These results imply that those viruses were more forced towards overcome their low replication capacities. Although the vast majority of mutations were synonymous mutations reverting to wild type sequence, these viruses also contained a higher number of non-synonymous mutations, suggesting new ways to explore the fitness landscape.

HIV-GminAp15 replicates 1 and 2 and HIV-PminAp15 replicate 2 still displayed an attenuated phenotype in PBMCs and still have a window of phenotype and genotype amelioration. Even though in 60 days of viral passages in MT-4 cells, most of the passaged viruses have recovered wild type replication capacities, some questions still remain to be answered regarding their genetic stability.

Our results also suggest that recoded gag and protease viruses passaged in cell culture may be exploring different sequence spaces. It has been shown that anticancer agents as 5-Azacytidine (5-AZC) or 5-aza-2'-deoxycytidine (decitabine) are mutagenic agents *in vitro* (Clouser *et al.*, 2010; Dapp *et al.*, 2009) (Martrus, unpublished results). It will be interesting to know if those HIV-1 recoded viruses will maintain their stability or will extinguish more quickly than the wild type virus in the presence of 5-AZC or decitabine. Differences in clonal and population fitness could occur within synonymous populations of viruses which have a similar replicative capacity (Coleman *et al.*, 2008; Mueller *et al.*, 2006) and have significantly different distributions of polymorphic amino acid substitutions (Lauring *et al.*, 2012).

Synonymous mutations have significant consequences for cellular processes because they can interfere with the codon usage bias. Thus, this can affect RNA processing, protein translation, and protein folding (Mueller *et al.*, 2010; Plotkin & Kudla, 2010). Work with live *E.coli* cells has shown that silent substitutions predictably alter translation elongation rates and protein folding efficiencies (Leifer *et al.*, 2011; Moura *et al.*, 2007; Spencer *et al.*, 2012). In HIV-1, synonymous substitutions can also affect the architecture and secondary structure of the genomic RNA, and therefore may affect viral replication (Coleman *et al.*, 2011; Sanjuán & Bordería, 2011; Tats *et al.*, 2008; Watts *et al.*, 2009). *In silico* mFold

analysis showed that RNA structures were not altered in recoded viruses. Since complete HIV-1 SHAPE (selective 2´-hydroxyl acylation by primer extension) analysis was reported (Watts *et al.*, 2009), more accurate structural models are arising (Low *et al.*, 2012; Low & Weeks, 2010) and help to understand RNA structures (Sztuba-Solinska & Le Grice, 2012). Comparison of mutated positions in protease gene between a fully reoptimized and functional protease (HIV-Pmax) and different deoptimized fragments showed that positions affected by codon-pair mutations are similar. These results may suggest that RNA structure was probably not the main cause of non-viability. Although some studies have pointed out that SHAPE-directed models cannot be studied as determined structures, but as helpful hypothesis (Kladwang *et al.*, 2011), up to date, SHAPE is the best tool to study RNA structure. It has been showed that gag is highly structured (Schneider *et al.*, 1997; Schwartz *et al.*, 1992a, b). One future direction that may be useful to ascertain that differences in deoptimized gag regions are not due to critical changes in RNA folding, would be to determine the *in vivo* SHAPE structure of different gag and protease constructs.

Recent data have identified a novel antiviral mechanism within the innate immune response, in which human SLFN11 selectively inhibits viral protein synthesis in HIV-infected cells by means of codon-bias discrimination (Li *et al.*, 2012). Synonymous substitutions optimizing for human cell expression reduces the antiviral activity of SLFN11 (Li *et al.*, 2012). Synonymous mutation is thought to be selectively neutral. However, previous work (Coleman *et al.*, 2008; Nevot *et al.*, 2011; Pandit & Sinha, 2011), and the data show in this thesis demonstrate that synonymous substitutions can negatively affect the replication capacity of an RNA virus, including HIV-1. Moreover, our results indicate that, when G/Cs in third position are increased, viruses displayed a stronger attenuated phenotype. Further work should include analysis of SLFN11 protein levels in MT-4 cells and stimulated PBMCs.

Codon-optimized HIV-1 genes significantly increased protein expression (Carnero *et al.*, 2008). However, it was unknown whether HIV-1 codon reoptimization increases viral fitness or virulence. A high level of preferred human codon use can produce the disruption of some unknown property of the HIV-1 RNA that is deleterious for other reasons, e.g. disrupting the secondary structure of the HIV-1 RNA genome (Sanjuán & Bordería, 2011) or increasing the number of CpG islands (Kypr *et al.*, 1989). In wild-type HIV-1, contrary to HTLV 1/2, CpG dinucleotides are underrepresented (Kypr *et al.*, 1989; Meintjes & Rodrigo, 2005), but in CPB-gag recoded constructs higher and statistically significant levels of CpG islands exist.

That was expected, as under-represented codon-pairs are mainly rich in Gs or Cs in codon third position, and the rejection of CpG dinucleotides is general in the three domains of life (Buchan, 2006; Moura *et al.*, 2007). It has been proposed that CpG dinucleotides have been associated with higher methylation levels, although translational efficiency may be also afected. It is not known why CpG dinucleotides are underrepresented; some theories have proposed that it could be an indirect result of DNA methylation in mammalian genomes (Moura *et al.*, 2005; 2007), an indirect effect of genomic forces, or an immune stimulator through activation of TLR9 (Bauer *et al.*, 2001; Fauci *et al.*, 2005; Häcker *et al.*, 2000). Interestingly, as in poliovirus (Coleman *et al.*, 2008), the higher the levels of CpG dinucleotides in sequences were present, the more substantial levels of attenuation were found. To discard DNA methylation, future experiments should include experiments with bisulfite conversion. This technique would allow checking every single methylated nucleotide, because it induces specific changes in the DNA sequence depending on the methylation status of individual cytosine residues.

This study has some limitations that are worth noting. Although we obtained a robust attenuation in PBMCs from healthy donors, our variants should be tested in an animal model, ideally the SIV macaque model, in order to demonstrate *in vivo* attenuation (Alpert *et al.*, 2012;; Koff *et al.*, 2006; Qureshi *et al.*, 2012). Although live attenuated vaccine in macaques offers a reliable long-term protection (Reynolds *et al.*, 2008; Wyand *et al.*, 1996; 1999) they have the potential to revert to wild type phenotype, implying that usage of live attenuated vaccines in humans for HIV-1 has serious safety concerns (Baba *et al.*, 1999; Koff *et al.*, 2006; Murphey-Corb, 1997; Wyand *et al.*, 1997). Animal experiments should include reoptimized viruses to test whether these optimized variants are not attenuated *in vivo*. Proof that lower replication capacities *ex vivo* can be translated into *in vivo* attenuation come from studies of the evolution of resistance to antiretroviral drugs, which is characterized by significant costs in *ex vivo* replication capacity (Martinez-Picado *et al.*, 1999; Martínez-Picado & Martínez, 2008). It has been suggested that drug selection of viruses with decreased *ex vivo* pol replication capacity during antiretroviral therapy may account for the sustained partial suppression of plasma HIV-1 RNA levels in patients with detectable drug-resistant viremia (Barbour *et al.*, 2002; Deeks *et al.*, 2001). Recently, individuals who were infected by an HIV-1 variant with a pol *ex vivo* replication capacity significantly lower than that of wild-type viruses had a significantly higher CD4+ T-cell count, independent of drug resistance and plasma HIV-1 RNA levels. This suggests that resistant viruses had

lowered virulence and were less able to deplete CD4+ T-cell counts (Barbour *et al.*, 2004). Similarly, many HIV-1–infected patients treated with protease inhibitors develop resistant HIV-1 variants and rebounds in viremia, but their CD4+ T-cell counts often do not decrease, suggesting that the resistant virus might be attenuated *in vivo* (Stoddart *et al.*, 2001). Overall, these results suggest a correlation between *ex vivo* HIV-1 replication capacity and *in vivo* HIV-1 pathogenesis. Future work should include the immunization of animal models with different codon pair–deoptimized attenuated HIV-1 strains to investigate whether recoded viruses are also attenuated *in vivo*, to determine their long-term stability, and to evaluate their ability to generate protective immunity. As previously suggested for poliovirus, codon-pair deoptimization could potentially increase the safety of HIV-1 inactivated vaccines while retaining the antigenicity of the wild-type virus. Finally, codon-pair deoptimized HIV-1 variants could be used to develop safer gene-therapy vectors. Obvious safety concerns remain for developing replicating vectors based on HIV-1. However, recent trials have demonstrated that lentiviral vectors, such as those derived from HIV-1, produce less insertional oncogenesis and can transduce non-dividing cells, and allow for more efficient gene transfer into hematopoietic stem cells compared to murine gamma-retrovirus vectors(Cartier *et al.*, 2009; Levine *et al.*, 2006).

CONCLUSIONS

1. Varying degrees of ex vivo attenuation can be obtained, depending upon both the specific deoptimized region and the number of deoptimized codons. This attenuation can be tuned by adjusting the levels and numbers of under-represented codon-pairs.

2. Although attenuation is based on several tens of nucleotide changes, virus evolution is reverting attenuated viruses to wild-type virulence. This improvement in replication capacity is not only due to reversion of initially introduced synonymous mutations but also to the presence of new synonymous and non-synonymous mutations.

3. Codon-pair bias reoptimized protease virus carrying a large number of synonymous mutations was not attenuated. No genetic or phenotypic reversion to the wild-type sequence was observed in viral passages, indicating that synonymous mutations in HIV-1 reoptimized viruses were not affecting its replication capacity.

4. Codon-pair attenuation in HIV-1 protease may be due to a decreased translation of targeted genes.

5. Codon pair reengineering is a useful strategy to phenotypically affect the replicative properties of HIV-1.

REFERENCES

Alpert, M. D., Harvey, J. D., Lauer, W. A., Reeves, R. K., Piatak, M., Carville, A., Mansfield, K. G., Lifson, J. D., Li, W. & other authors. **(2012).** ADCC Develops Over Time during Persistent Infection with Live-Attenuated SIV and Is Associated with Complete Protection against SIVmac251 Challenge. *PLoS Pathog* **8**, e1002890 (A. Trkola, Ed.).

Anderson, J. L. & Hope, T. J. (2008). APOBEC3G restricts early HIV-1 replication in the cytoplasm of target cells. *Virology* **375**, 1–12.

Andrea, L. A. D., Pintó, R. M., Bosch, A., Musto, H. & Cristina, J. (2011). A detailed comparative analysis on the overall codon usage patterns in Hepatitis A virus. *Virus Research* **157**, 19–24. Elsevier B.V.

Baba, T. W., Liska, V., Khimani, A. H., Ray, N. B., Dailey, P. J., Penninck, D., Bronson, R., Greene, M. F., McClure, H. M. & other authors. **(1999).** Live attenuated, multiply deleted simian immunodeficiency virus causes AIDS in infant and adult macaques. *Nature Medicine* **5**, 194–203.

Barbour, J. D., Wrin, T., Grant, R. M., Martin, J. N., Segal, M. R., Petropoulos, C. J. & Deeks, S. G. (2002). Evolution of Phenotypic Drug Susceptibility and Viral Replication Capacity during Long-Term Virologic Failure of Protease Inhibitor Therapy in Human Immunodeficiency Virus-Infected Adults. *Journal of Virology* **76**, 11104–11112.

Barbour, J. D., Hecht, F. M., Wrin, T., Segal, M. R., Ramstead, C. A., Liegler, T. J., Busch, M. P., Petropoulos, C. J., Hellmann, N. S. & other authors. **(2004).** Higher CD4+ T cell counts associated with low viral pol replication capacity among treatment-naive adults in early HIV-1 infection. *J Infect Dis* **190**, 251–256.

Bauer, S., Kirschning, C. J., Häcker, H., Redecke, V., Hausmann, S., Akira, S., Wagner, H. & Lipford, G. B. (2001). Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. *Proc Natl Acad Sci USA* **98**, 9237–9242.

Berkhout, B. & van Hemert, F. J. (1994). The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins. *Nucleic Acids Research* **22**, 1705–1711.

Betancor, G., Puertas, M. C., Nevot, M., Garriga, C., Martinez, M. A., Martinez-Picado, J. & Menendez-Arias, L. (2010). Mechanisms Involved in the Selection of HIV-1 Reverse Transcriptase Thumb Subdomain Polymorphisms Associated with Nucleoside Analogue Therapy Failure. *Antimicrobial Agents and Chemotherapy* **54**, 4799–4811.

Bishop, K. N., Holmes, R. K., Sheehy, A. M. & malim, M. H. (2007). APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci* **305**, 118–128.

Boycheva, S., Chkodrov, G. & Ivanov, I. (2003). Codon pairs in the genome of Escherichia coli. *Bioinformatics* **19**, 987–998.

Briggs, J. A. G., Grünewald, K., Glass, B., Förster, F., Kräusslich, H.-G. & Fuller, S. D. (2006). The Mechanism of HIV-1 Core Assembly: Insights from Three-Dimensional Reconstructions of Authentic Virions. *Structure* **14**, 15–20.

Briggs, J. A. G. & Kräusslich, H.-G. (2011). The Molecular Architecture of HIV. *J Mol Biol* **410**, 491–500. Elsevier Ltd.

Buchan, J. R. (2006). tRNA properties help shape codon pair preferences in open reading frames. *Nucleic Acids Research* **34**, 1015–1027.

Bukrinsky, M. I., Sharova, N., Dempsey, M. P., Stanwick, T. L., Bukrinskaya, A. G., Haggerty, S. & Stevenson, M. (1992). Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc Natl Acad Sci USA* **89**, 6580–6584.

Burns, C. C., Shaw, J., Campagnoli, R., Jorba, J., Vincent, A., Quay, J. & Kew, O. (2006). Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *Journal of Virology* **80**, 3259–3272.

Capel, E., Martrus, G., Parera, M., Clotet, B. & Martinez, M. A. (2012). Evolution of the human immunodeficiency virus type 1 protease: Effects on viral replication capacity and protease robustness. *Journal of General Virology*.

Carnero, E., Li, W., Borderia, A. V., Moltedo, B., Moran, T. & Garcia-Sastre, A. (2008). Optimization of Human Immunodeficiency Virus Gag Expression by Newcastle Disease Virus Vectors for the Induction of Potent Immune Responses. *Journal of Virology* **83**, 584–597.

Cartier, N., Hacein-Bey-Abina, S., Bartholomae, C. C., Veres, G., Schmidt, M., Kutschera, I., Vidaud, M., Abel, U., Dal-Cortivo, L. & other authors. **(2009).** Hematopoietic Stem Cell Gene Therapy with a Lentiviral Vector in X-Linked Adrenoleukodystrophy. *Science* **326**, 818–823.

Clouser, C. L., Patterson, S. E. & Mansky, L. M. (2010). Exploiting Drug Repositioning for Discovery of a Novel HIV Combination Therapy. *Journal of Virology* **84**, 9301–9309.

Coffin, J. M. (1995). HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy
. *Science* **267**, 483–489.

Coffin, J. M., Hughes, S. H. & Varmus, H. E. (1997). Retroviruses. *Cold Spring Harbor Laboratory Press* (J. M. Coffin, S. H. Hughes & H. E. Varmus, Eds.). NY.

Coleman, J. R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E. & Mueller, S. (2008). Virus Attenuation by Genome-Scale Changes in Codon Pair Bias. *Science* **320**, 1784–1787.

Coleman, J. R., Papamichail, D., Yano, M., Garcia-Suarez, M. D. M. & Pirofski, L. A. (2011). Designed Reduction of Streptococcus pneumoniae Pathogenicity via Synthetic Changes in Virulence Factor Codon-pair Bias. *Journal of Infectious Diseases* **203**, 1264–1273.

Curran, J. F., Poole, E. S., Tate, W. P. & Gross, B. L. (1995). Selection of aminoacyl-tRNAs at sense codons: the size of the tRNA variable loop determines whether the immediate 3' nucleotide to the codon has a context effect. *Nucleic Acids Research* **23**, 4104–4108.

Dapp, M. J., Clouser, C. L., Patterson, S. & Mansky, L. M. (2009). 5-Azacytidine Can Induce Lethal Mutagenesis in Human Immunodeficiency Virus Type 1. *Journal of Virology* **83**, 11950–11958.

Deeks, S. G., Wrin, T., Liegler, T., Hoh, R., Hayden, M., Barbour, J. D., Hellmann, N. S., Petropoulos, C. J., McCune, J. M. & other authors. **(2001).** Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia. *N Engl J Med* **344**, 472–480.

Doma, M. K. & Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nat Cell Biol* **440**, 561–564.

Domingo, E. & Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annu Rev Microbiol* **51**, 151–178.

Domingo, E., Sheldon, J. & Perales, C. (2012). Viral Quasispecies Evolution. *Microbiology and Molecular Biology Reviews* **76**, 159–216.

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* **88**, 7160–7164.

Drake, J. W. & Holland, J. J. (1999). Mutation rates among RNA viruses. *Proc Natl Acad Sci USA* **96**, 13910–13913.

Drake, J. W. (1993). Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci USA* **90**, 4171–4175. National Acad Sciences.

Eckard Wimmer, S. S. S. M. B. F. D. P. J. R. C. J. C. (2010, August 19). ATTENUATED VIRUSES USEFUL FOR VACCINES. United States Patent Application Publication.

Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules
. *naturwissenschaften* **10**, 465–523.

Fauci, A. S., Mavilio, D. & Kottilil, S. (2005). NK cells in HIV infection: Paradigm for protection or targets for ambush. *Nat Rev Immunol* **5**, 835–843.

Fedorov, A., Saxonov, S. & Gilbert, W. (2002). Regularities of context-dependent codon bias in eukaryotic genes. *Nucleic Acids Research* **30**, 1192–1197.

Fernandez, G., Clotet, B. & Martinez, M. A. (2007). Fitness Landscape of Human Immunodeficiency Virus Type 1 Protease Quasispecies. *Journal of Virology* **81**, 2485–2496.

Goh, G., Dunker, A. K. & Uversky, V. N. (2008). Protein intrinsic disorder toolbox for comparative analysis of viral proteins. *BMC Genomics* **9**, S4.

Goodman, M. F. & Fygenson, D. K. (1998). DNA polymerase fidelity: from genetics toward a biochemical understanding. *Genetics* **148**, 1475–1482. Genetics Soc America.

Gouy, M. & Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Research* **10**, 7055–7074.

Grantham, R. & Perrin, P. (1986). AIDS virus and HTLV-I differ in codon choices
. *Nature* **319**, 727–728.

Gustafsson, C., Govindarajan, S. & Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in Biotechnology* **22**, 346–353.

Gutman, G. A. & Hatfield, G. W. (1989). Nonrandom utilization of codon pairs in Escherichia coli. *Proc Natl Acad Sci USA* **86**, 3699–3703.

Haas, J., Park, E. C. & Seed, B. (1996). Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Curr Biol* **6**, 315–324.

Häcker, H., Vabulas, R. M., Takeuchi, O., Hoshino, K., Akira, S. & Wagner, H. (2000). Immune cell activation by bacterial CpG-DNA through myeloid differentiation marker 88 and tumor necrosis factor receptor-associated factor (TRAF)6. *J Exp Med* **192**, 595–600.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution* **2**, 13–34.

Irwin, B., Heck, J. D. & Hatfield, G. W. (1995). Codon pair utilization biases influence translational elongation step times. *J Biol Chem* **270**, 22801–22806.

Karlin, S., Doerfler, W. & Cardon, L. R. (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**, 2889–2897.

Karlin, S. & Mrázek, J. (1996). What drives codon choices in human genes? *J Mol Biol* **262**, 459–472.

Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., Sun, C., Grayson, T., Wang, S. & other authors. **(2008).** Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences* **105**, 7552–7557.

Kisic, M., Matamoros, T., Nevot, M., Mendieta, J., Martinez-Picado, J., Martinez, M. A. & Menendez-Arias, L. (2011). Thymidine Analogue Excision and Discrimination Modulated by Mutational Complexes Including Single Amino Acid Deletions of Asp-67 or Thr-69 in HIV-1 Reverse Transcriptase. *Journal of Biological Chemistry* **286**, 20615–20624.

Kladwang, W., VanLang, C. C., Cordero, P. & Das, R. (2011). Understanding the Errors of SHAPE-Directed RNA Structure Modeling. *Biochemistry* **50**, 8049–8056.

Koff, W. C., Johnson, P. R., Watkins, D. I., Burton, D. R., Lifson, J. D., Hasenkrug, K. J., McDermott, A. B., Schultz, A., Zamb, T. J. & other authors. **(2006).** HIV vaccine design: insights from live attenuated SIV vaccines. *Nat Immunol* **7**, 19–23.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M. & Sigal, I. S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci USA* **85**, 4686–4690.

Kuiken, C., Foley, B., Leitner, T., Apetrei, C., Hahn, B. H., Mizrachi, I., Mullins, J. I., Rambaut, A., Wolinsky, S. & Korber, B. (2012). HIV Sequence Compendium 2012 1–450.

Kunkel, T. A. & Alexander, P. S. (1986). The base substitution fidelity of eucaryotic DNA polymerases. Mispairing frequencies, site preferences, insertion preferences, and base substitution by dislocation. *J Biol Chem* **261**, 160–166. ASBMB.

Kypr, J. & Mrázek, J. (1987). Unusual codon usage of HIV. *Nature* **20**.

Kypr, J., Mrázek, J. & Reich, J. (1989). Nucleotide composition bias and CpG dinucleotide content in the genomes of HIV and HTLV 1/2. *Biochimica et Biophysica Acta*.

Lauring, A. S., Acevedo, A., Cooper, S. B. & Andino, R. (2012). Codon Usage Determines the Mutational Robustness, Evolutionary Capacity, and Virulence of an RNA Virus. *Cell Host and Microbe* **12**, 623–632. Elsevier Inc.

Lauring, A. S., Frydman, J. & Andino, R. (2013). The role of mutational robustness in RNA virus evolution. *Nature Reviews Microbiology* 1–10. Nature Publishing Group.

Leifer, I., Hoeper, D., Blome, S., Beer, M. & Ruggli, N. (2011). Clustering of classical swine fever virus isolates by codon pair bias. *BMC Research Notes* **4**, 521. BioMed Central Ltd.

Levine, B. L., Humeau, L. M., Boyer, J., MacGregor, R.-R., Rebello, T., Lu, X., Binder, G. K., Slepushkin, V., Lemiale, F. & other authors. **(2006).** Gene transfer in humans using a conditionally replicating lentiviral vector. *Proc Natl Acad Sci USA* **103**, 17372–17377.

Li, M., Kao, E., Gao, X., Sandig, H., Limmer, K., Pavon-Eternod, M., Jones, T. E., Landry, S., Pan, T. & other authors. **(2012).** Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature* 1–13.

Livak, K. J. & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**, 402–408.

Low, J. T., Knoepfel, S. A., Watts, J. M., Brake, ter, O., Berkhout, B. & Weeks, K. M. (2012). SHAPE-directed Discovery of Potent shRNA Inhibitors of HIV-1. *Mol Ther* **20**, 820–828.

Low, J. T. & Weeks, K. M. (2010). SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150–158. Elsevier Inc.

Manrubia, S. C., Escarmís, C., Domingo, E. & Lázaro, E. (2005). High mutation rates, bottlenecks, and robustness of RNA viral quasispecies. *Gene* **347**, 273–282.

Mansky, L. M. & Temin, H. M. (1995). Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *Journal of Virology* **69**, 5087–5094.

Martinez-Picado, J., Savara, A. V., Sutton, L. & D'Aquila, R. T. (1999). Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *Journal of Virology* **73**, 3744–3752.

Martínez-Picado, J. & Martínez, M. A. (2008). HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: A view from the clinic and ex vivo. *Virus Research* **134**, 104–123.

Maschera, B., Furfine, E. & Blair, E. D. (1995). Analysis of resistance to human immunodeficiency virus type 1 protease inhibitors by using matched bacterial expression and proviral infection vectors. *Journal of Virology* **69**, 5431–5436.

Mauclère, P., Loussert-Ajaka, I., Damond, F., Fagot, P., Souquières, S., Monny Lobe, M., Mbopi Keou, F. X., Barre-Sinoussi, F., Saragosti, S. & other authors. **(1997).** Serological and virological characterization of HIV-1 group O infection in Cameroon. *AIDS* **11**, 445–453.

Meintjes, P. L. & Rodrigo, A. G. (2005). Evolution of relative synonymous codon usage in Human Immunodeficiency Virus type-1. *J Bioinform Comput Biol* **3**, 157–168.

Miura, T., Brockman, M. A., Brumme, C. J., Brumme, Z. L., Carlson, J. M., Pereyra, F., Trocha, A., Addo, M. M., Block, B. L. & other authors. **(2008).** Genetic Characterization of Human Immunodeficiency Virus Type 1 in Elite Controllers: Lack of Gross Genetic Defects or Common Amino Acid Changes. *Journal of Virology* **82**, 8422–8430.

Moriyama, E. N. & Powell, J. R. (1997). Codon usage bias and tRNA abundance in Drosophila. *J Mol Evol* **45**, 514–523.

Moura, G., Pinheiro, M., Arrais, J., Gomes, A. C., Carreto, L., Freitas, A., Oliveira, J. L. & Santos, M. A. S. (2007). Large Scale Comparative Codon-Pair Context Analysis Unveils General Rules that Fine-Tune Evolution of mRNA Primary Structure. *PLOS ONE* **2**, e847

(A. Christoffels, Ed.).

Moura, G., Pinheiro, M., Silva, R., Miranda, I., Afreixo, V., Dias, G., Freitas, A., Oliveira, J. L. & Santos, M. A. S. (2005). Comparative context analysis of codon pairs on an ORFeome scale. *Genome Biol* **6**, R28.

Mueller, S., Papamichail, D., Coleman, J. R., Skiena, S. & Wimmer, E. (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**, 9687–9696.

Mueller, S., Coleman, J. R., Papamichail, D., Ward, C. B., Nimnual, A., Futcher, B., Skiena, S. & Wimmer, E. (2010). Live attenuated influenza virus vaccines by computer-aided rational design. *Nat Biotechnol* **28**, 723–726.

Mulder, L. C. F., Harari, A. & Simon, V. (2008). Cytidine deamination induced HIV-1 drug resistance. *Proceedings of the National Academy of Sciences* **105**, 5501–5506.

Murphey-Corb, M. (1997). Live-attenuated HIV vaccines: how safe is safe enough? *Nature Medicine* **3**, 17–18.

Nevot, M., Martrus, G., Clotet, B. & Martínez, M. A. (2011). RNA Interference as a Tool for Exploring HIV-1 Robustness. *J Mol Biol* **413**, 84–96. Elsevier Ltd.

Newman, E. N. C., Holmes, R. K., Craig, H. M., Klein, K. C., Lingappa, J. R., malim, M. H. & Sheehy, A. M. (2005). Antiviral Function of APOBEC3G Can Be Dissociated from Cytidine Deaminase Activity. *Current Biology* **15**, 166–170.

Ngumbela, K. C., Ryan, K. P., Sivamurthy, R., Brockman, M. A., Gandhi, R. T., Bhardwaj, N. & Kavanagh, D. G. (2008). Quantitative Effect of Suboptimal Codon Usage on Translational Efficiency of mRNA Encoding HIV-1 gag in Intact T Cells. *PLOS ONE* **3**, e2356 (K. C. Ngumbela, K. P. Ryan, R. Sivamurthy, M. A. Brockman, R. T. Gandhi, N. Bhardwaj & D. G. Kavanagh, Eds.).

Palidwor, G. A., Perkins, T. J. & Xia, X. (2010). A general model of codon bias due to GC mutational bias. *PLOS ONE* **5**, e13431.

Pandit, A. & Sinha, S. (2011). Differential Trends in the Codon Usage Patterns in HIV-1 Genes. *PLOS ONE* **6**, 1–10.

Pannecouque, C., Daelemans, D. & De Clercq, E. (2008). Tetrazolium-based colorimetric assay for the detection of HIV replication inhibitors: revisited 20 years later. *Nat Protoc* **3**, 427–434.

Parera, M., Fernandez, G., Clotet, B. & Martinez, M. A. (2006). HIV-1 Protease Catalytic Efficiency Effects Caused by Random Single Amino Acid Substitutions. *Molecular Biology and Evolution* **24**, 382–387.

Parrish, N. F., Gao, F., Li, H., Giorgi, E. E., Barbian, H. J., Parrish, E. H., Zajic, L., Iyer, S. S., Decker, J. M. & other authors. (2013). Phenotypic properties of transmitted founder HIV-1. *Proceedings of the National Academy of Sciences*.

Pauwels, R., Balzarini, J., Baba, M., Snoeck, R., Schols, D., Herdewijn, P., Desmyter, J. & De Clercq, E. (1988). Rapid and automated tetrazolium-based colorimetric assay for the detection of anti-HIV compounds. *Journal of Virological Methods* **20**, 309–321.

Pedroza martins, L., Chenciner, N. & Wain Hobson, S. (1992). Complex intrapatient sequence variation in the V1 and V2 hypervariable regions of the HIV-1 gp120 envelope sequence

. *Virology* **191**, 837–845.

Peeters, M., Gueye, A., Mboup, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P. & other authors. (1997). Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* **11**, 493–498.

Plantier, J.-C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D. L. & Simon, F. (2009). A new human immunodeficiency virus derived from gorillas. *Nature Medicine* **15**, 871–872.

Plotkin, J. B. & Kudla, G. (2010). Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* **12**, 32–42. Nature Publishing Group.

Preston, B. D. & Dougherty, J. P. (1996). Mechanisms of retroviral mutation. *Trends in Microbiology* **4**, 16–21.

Qureshi, H., Ma, Z. M., Huang, Y., Hodge, G., Thomas, M. A., DiPasquale, J., DeSilva, V., Fritts, L., Bett, A. J. & other authors. **(2012).** Low-Dose Penile SIVmac251 Exposure of Rhesus Macaques Infected with Adenovirus Type 5 (Ad5) and Then Immunized with a Replication-Defective Ad5-Based SIV gag/pol/nef Vaccine Recapitulates the Results of the Phase IIb Step Trial of a Similar HIV-1 Vaccine. *Journal of Virology* **86**, 2239–2250.

Rambaut, A., Posada, D., Crandall, K. A. & Holmes, E. C. (2004). The causes and consequences of HIV evolution. *Nat Rev Genet* **5**, 52–61.

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K. & Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615–619.

Read, E. L., Tovo-Dwyer, A. A. & Chakraborty, A. K. (2012). Stochastic effects are important in intrahost HIV evolution even when viral loads are high. *Proceedings of the National Academy of Sciences* **109**, 19727–19732.

Reynolds, M. R., Weiler, A. M., Weisgrau, K. L., Piaskowski, S. M., Furlott, J. R., Weinfurter, J. T., Kaizu, M., Soma, T., Leon, E. J. & other authors. **(2008).** Macaques vaccinated with live-attenuated SIV control replication of heterologous virus. *Journal of Experimental Medicine* **205**, 2537–2550.

Ruiz-Jarabo, C. M., Miller, E., Gómez-Mariano, G. & Domingo, E. (2003). Synchronous Loss of Quasispecies Memory in Parallel Viral Lineages: A Deterministic Feature of Viral Quasispecies. *J Mol Biol* **333**, 553–563.

Sanjuán, R. & Bordería, A. V. (2011). Interplay between RNA structure and protein evolution in HIV-1. *Molecular Biology and Evolution* **28**, 1333–1338.

Schneider, R., Campbell, M., Nasioulas, G., Felber, B. K. & Pavlakis, G. N. (1997). Inactivation of the human immunodeficiency virus type 1 inhibitory elements allows Rev-independent expression of Gag and Gag/protease and particle formation. *Journal of Virology* **71**, 4892–4903.

Schwartz, S., Campbell, M., Nasioulas, G., Harrison, J., Felber, B. K. & Pavlakis, G. N. (1992a). Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. *Journal of Virology* **66**, 7176–7182.

Schwartz, S., Felber, B. K. & Pavlakis, G. N. (1992b). Distinct RNA sequences in the gag region of human immunodeficiency virus type 1 decrease RNA stability and inhibit expression in the absence of Rev protein. *Journal of Virology* **66**, 150–159.

Shao, Z.-Q., Zhang, Y.-M., Feng, X.-Y., Wang, B. & Chen, J.-Q. (2012). Synonymous Codon Ordering: A Subtle but Prevalent Strategy of Bacteria to Improve Translational Efficiency. *PLOS ONE* **7**, e33547 (A. J. Ratner, Ed.).

Simon, F., Mauclère, P., Roques, P., Loussert-Ajaka, I., Müller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F. & Brun-Vézinet, F. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature Medicine* **4**, 1032–1037.

Sitaraman, V., Hearing, P., Ward, C. B., Gnatenko, D. V., Wimmer, E., Mueller, S., Skiena, S. & Bahou, W. F. (2011). Computationally designed adeno-associated virus (AAV) Rep 78 is efficiently maintained within an adenovirus vector. *Proceedings of the National Academy of Sciences* **108**, 14294–14299.

Smith, D. & Yarus, M. (1989). tRNA-tRNA interactions within cellular ribosomes. *Proc Natl Acad Sci USA* **86**, 4397–4401.

Song, Y., Liu, Y., Ward, C. B., Mueller, S., Futcher, B., Skiena, S., Paul, A. V. & Wimmer, E.

(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**, 14301–14307.

Spencer, P. S., Siller, E., Anderson, J. F. & Barral, J. M. (2012). Silent Substitutions Predictably Alter Translation Elongation Rates and Protein Folding Efficiencies. *J Mol Biol* **422**, 328–335. Elsevier Ltd.

Steitz, T. A. (2008). A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* **9**, 242–253.

Stoddart, C. A., Liegler, T. J., Mammano, F., Linquist-Stepps, V. D., Hayden, M. S., Deeks, S. G., Grant, R. M., Clavel, F. & McCune, J. M. (2001). Impaired replication of protease inhibitor-resistant HIV-1 in human thymus. *Nature Medicine* **7**, 712–718.

Sugiyama, T., Gursel, M., Takeshita, F., Coban, C., Conover, J., Kaisho, T., Akira, S., Klinman, D. M. & Ishii, K. J. (2005). CpG RNA: identification of novel single-stranded RNA that stimulates human CD14+CD11c+ monocytes. *J Immunol* **174**, 2273–2279.

Sztuba-Solinska, J. & Le Grice, S. F. J. (2012). Probing Retroviral and Retrotransposon Genome Structures: The "SHAPE" of Things to Come. *Molecular Biology International* **2012**, 1–12.

Tapia, N., Fernandez, G., Parera, M., Gómez-Mariano, G., Clotet, B., Quiñones-Mateu, M., Domingo, E. & Martínez, M. A. (2005). Combination of a mutagenic agent with a reverse transcriptase inhibitor results in systematic inhibition of HIV-1 infection. *Virology* **338**, 1–8.

Tats, A., Tenson, T. & Remm, M. (2008). Preferred and avoided codon pairs in three domains of life. *BMC Genomics* **9**, 463.

Temin, H. M. (1993). Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc Natl Acad Sci USA* **90**, 6900–6903.

Uemura, S., Aitken, C. E., Korlach, J., Flusberg, B. A., Turner, S. W. & Puglisi, J. D. (2010). Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464**, 1012–1017.

van der Kuyl, A. C. & Berkhout, B. (2012). The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. *Retrovirology* **9**, 92.

van Hemert, F. J., van der Kuyl, A. C. & Berkhout, B. (2013). The A-nucleotide preference of HIV-1 in the context of its structured RNA genome. *RNA Biol* **10**, 211–215.

van Weringh, A., Ragonnet-Cronin, M., Pranckeviciene, E., Pavon-Eternod, M., Kleiman, L. & Xia, X. (2011). HIV-1 modulates the tRNA pool to improve translation efficiency. *Molecular Biology and Evolution* **28**, 1827–1834.

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

Weiss, R. A. (2001). Gulliver's travels in HIVland. *Nature* **410**, 1–5.

Wood, N., Bhattacharya, T., Keele, B. F., Giorgi, E., Liu, M., Gaschen, B., Daniels, M., Ferrari, G., Haynes, B. F. & other authors. **(2009).** HIV Evolution in Early Infection: Selection Pressures, Patterns of Insertion and Deletion, and the Impact of APOBEC. *PLoS Pathog* **5**, e1000414 (D. C. Douek, Ed.).

Wright, F. (1990). The "effective number of codons" used in a gene. *Gene* **87**, 23–29.

Wyand, M. S., Manson, K. H., Garcia-Moll, M., Montefiori, D. & Desrosiers, R. C. (1996). Vaccine protection by a triple deletion mutant of simian immunodeficiency virus. *Journal of Virology* **70**, 3724–3733.

Wyand, M. S., Manson, K. H., Lackner, A. A. & Desrosiers, R. C. (1997). Resistance of neonatal monkeys to live attenuated vaccine strains of simian immunodeficiency virus. *Nature Medicine* **3**, 32–36.

Wyand, M. S., Manson, K., Montefiori, D. C., Lifson, J. D., Johnson, R. P. & Desrosiers, R. C.

(1999). Protection by live, attenuated simian immunodeficiency virus against heterologous challenge. *Journal of Virology* **73**, 8356–8363.

Yoshimura, F. K., Diem, K., Learn, G. H., Riddell, S. & Corey, L. (1996). Intrapatient sequence variation of the gag gene of human immunodeficiency virus type 1 plasma virions. *Journal of Virology* **70**, 8879–8887.

Yuste, E., Lopez-Galindez, C. & Domingo, E. (2000). Unusual Distribution of Mutations Associated with Serial Bottleneck Passages of Human Immunodeficiency Virus Type 1. *Journal of Virology* **74**, 9546–9552.

Yuste, E., Sánchez-Palomino, S., Casado, C., Domingo, E. & Lopez-Galindez, C. (1999). Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *Journal of Virology* **73**, 2745–2751.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406–3415.

PUBLICATIONS

Martrus G, Nevot M, Andrés C, Clotet B, Martinez MA. "Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture". Submitted for publication.

Capel E, **Martrus G**, Parera M, Clotet B, Martinez MA. "Evolution of the human immunodeficiency virus type 1 protease: Effects on viral replication capacity and protease robustness." J Gen Virol. 2012 Aug 29.

Parera M, **Martrus G**, Franco S, Clotet B, Martinez MA. "Canine hepacivirus NS3 serine protease can cleave the human adaptor proteins MAVS and TRIF." PLoS One 2012;7(8) :e42481. Epub 2012 Aug 1.

Martinez MA, **Martrus G**, Capel E, Parera M, Franco S and Nevot M. "Quasispecies Dynamics of RNA Viruses". In: Viruses: Essential Agents of Life. G. Witzany (ed). Springer. 2012. ISBN 978-94-007-4899-6.

Nevot M, **Martrus G**, Clotet B, Martinez MA "RNA interference as a tool for exploring HIV-1 robustness." J Mol Biol 2011 Oct 14;413(1):84-96. Epub 2011 Aug 22.

ANNEX

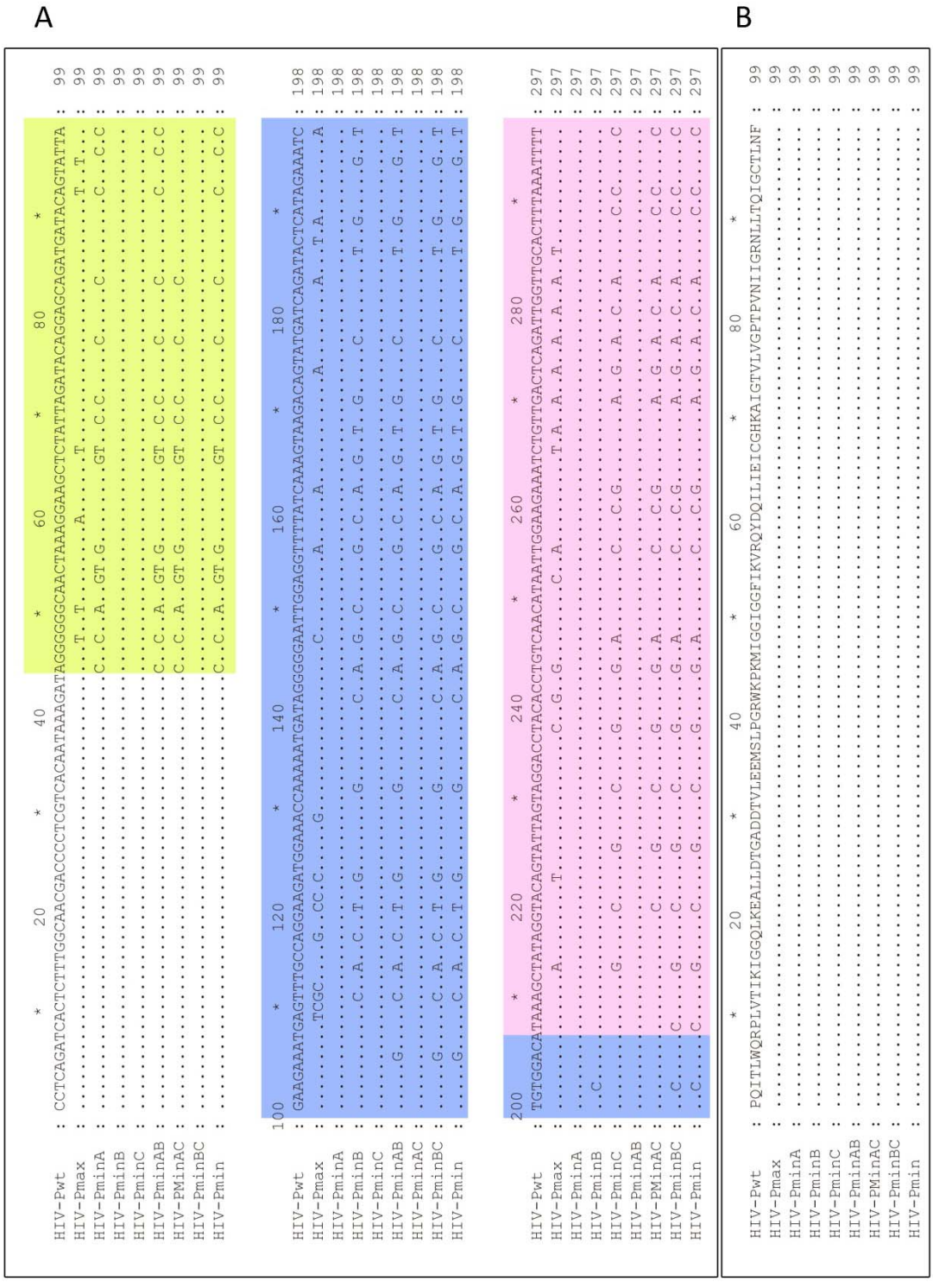
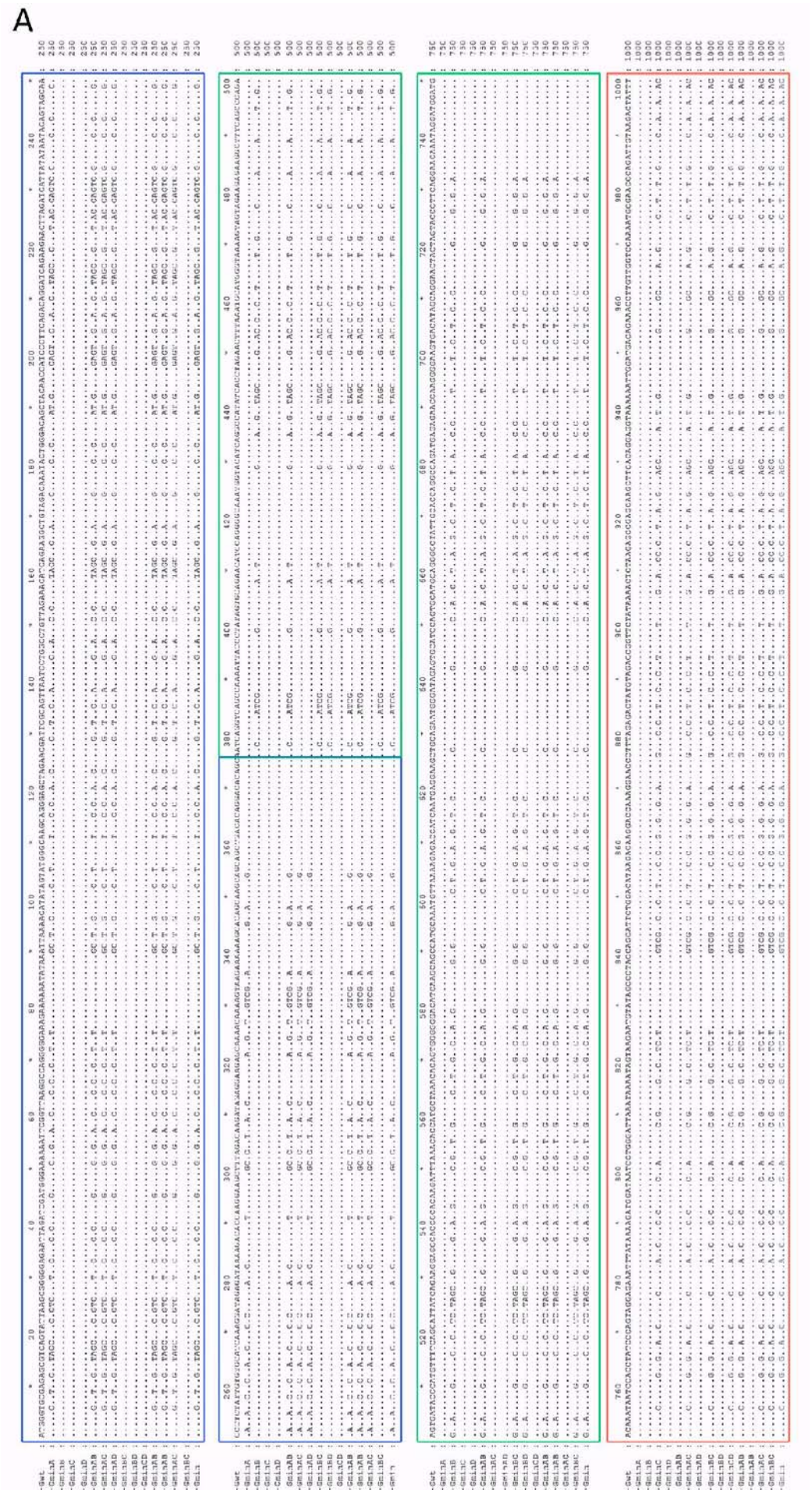


Figure 41. Protease nucleotide (A) and protein (B) sequences. Mutations introduced in nucleotide protease sequence do not affect protein sequence. Yellow square represents A fragment, blue square B fragment and pink square C fragment.

Annex

A



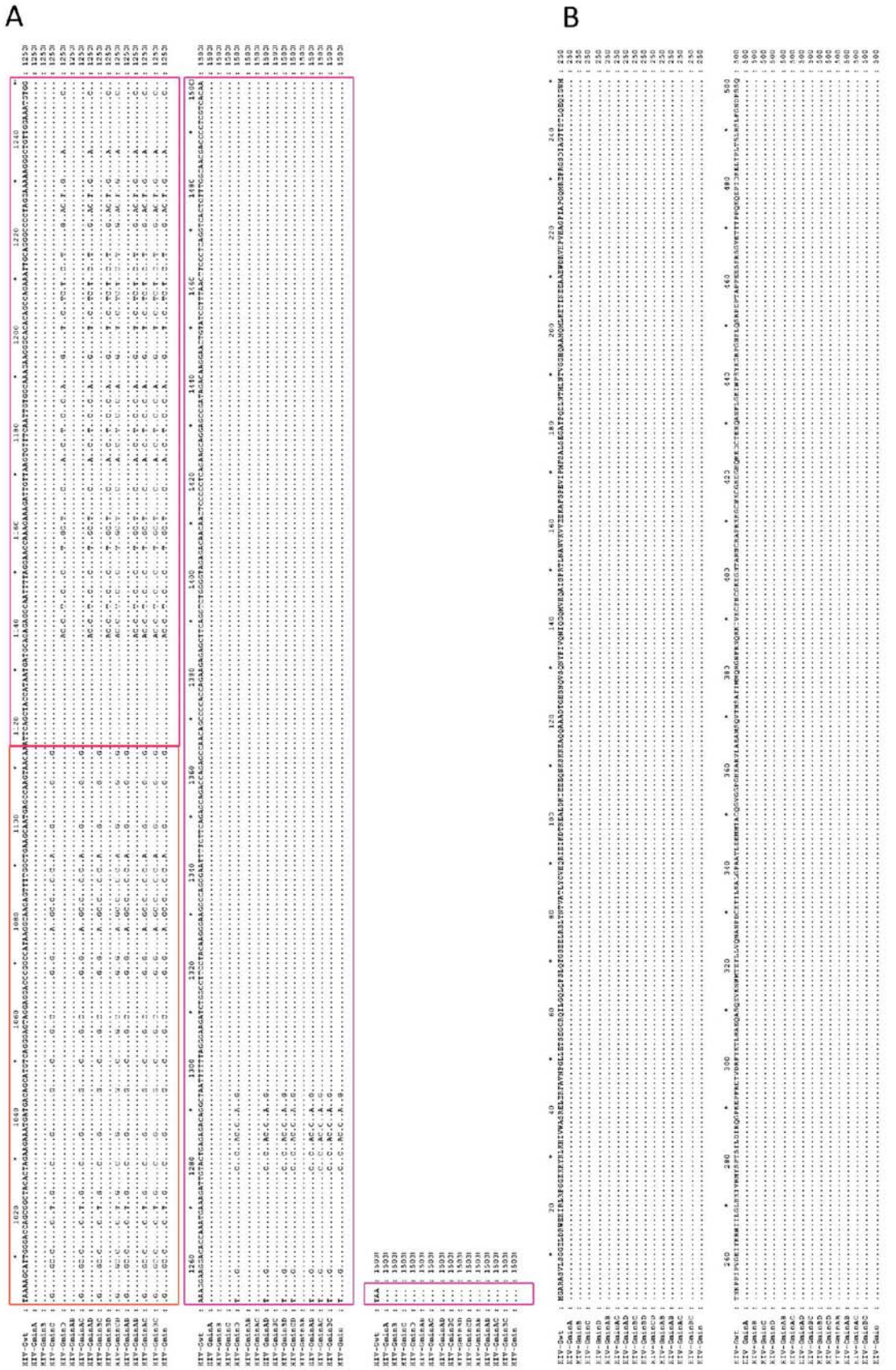


Figure 42. Gag nucleotide (A) and protein (B) sequences. Mutations introduced in nucleotide protease sequence do not affect protein sequence. Blue square represents A fragment, green square B fragment, orange square C fragment and pink square D fragment.

Acknowledgements

ACKNOWLEDGEMENTS

So… that's it! It's the end! Sé ("pillins"!) que aquesta és la única part que es llegirà tothom el dia que porti la tesi impresa, així que intentaré que quedi prou bé. Estic notant que, com tots suposo, no és fàcil d'escriure, fa pena. Pena per què treballar a irsiCaixa amb tots vosaltres ha estat un gran plaer. La sensació que m'enduc és la de tenir al costat sempre a gent disposada a ajudar, a qui consultar dubtes, a repassar problemes i a compartir grans moments dins i fora del lab. En resum, a conèixer, no només a grans professionals, però també a amics que espero que perdurin en la distància. A tots, us vull agrair les hores passades dins i fora de irsi, pels congressos, sopars, dinars, berenars, cafès, esmorzars, partidets de volley platja, i totes les activitats que hem fet junts al llarg d'aquests anys.

No m'ha estat mai fàcil tancar etapes de la meva vida, encara que el que vingui sigui per créixer. Ja aviso aquí, al que em faci plorar amb un vídeo sensiblon, se les té amb mi (hi he estat sempre en contra, no em queda bé el rímel corregut ☺).

Primer de tot, agrair al meu director de tesi, al **Miguel Angel**, per donar-me la oportunitat fa tants anys de començar la tesi amb ell i supervisar-me. Gràcies per confiar en mi, per ajudar-me a fer el salt i superar la por escènica en els congressos nacionals de virologia, en el de la societat catalana de virologia i sobretot, per ajudar-me a anar al CROI. Pels moments fora del lab, on el "jefe" deixa de ser el "jefe" per ser el cuiner que ens fa paelles, fideuas i calçotades memorables. Perquè, per molt que sigui del Madrid, és molt gran. Gràcies per ajudar-me a créixer com a científica i a fer-me prou independent com per a poder arribar a l'etapa de post-doc "preparada".

GRÀCIES:

A les MOLones I!!! Sou molt grans noies! Crec que sense l'ajuda i el suport que he rebut vostre no hauria estat el mateix això! Aniré per ordre alfabètic, així no hi ha "disputes" (;)). A la **Cristina Andrés**, la última incorporació grupal, per la seva innocència que és entranyable, les seves "Hello Kittys" i la seva actualitat "cuoral" i per ajudar-me en aquests últims mesos de bogeria. A la **Ester Aparicio**, de qui he après que la vida es pot veure amb un costat encara més positiu, que si alguna cosa no funciona, no et desesperis que funcionarà (ho faràs funcionar), que em fa enveja tota aquesta vitalitat i energia per treure temps per fer mil projectes alhora. Gràcies per introduir-me en el món del Photoshop fa tants anys ja! ;) A la **Helen, Helencita, Elenita, Elena**… La meva gran companya de congressos al llarg d'aquests anys, la que em va començar a motivar per córrer, la que em feia riure cada cop que ho necessitava! La teva vitalitat i energia positiva, al igual que la de la Ester, son contagiables! Hem fet aquesta "cursa" del PhD juntes, i juntes acabem! Yes, we MADE IT! A la **Sandra**, que em va fer de supervisora de pràctiques i que va confiar en mi quan li va recomanar al Miguel Angel el meu nom per començar la tesi. Moltes gràcies per ensenyar-me quan era una estudiant de carrera i per seguir aconsellant-me sempre durant aquests anys sense mai cansar-te dels dubtes tontos que pogués tenir sobre PCRs, real-times

i lligacions. Sobretot, no enfadar-se quan el "monstre de les 19h" m'assetjava i havia de menjar (si o si) i arrasava amb el seu calaix de kit-kats, galetetes varies i caramels. Tenir-te aquests anys a la taula del costat ha estat un honor i un plaer!!! A la **Maria Nevot**, per ser la gran lluitadora, que no es rendeix mai i m'ha ajudat al llarg d'aquests anys a fer-me de "coach personal", a picar-me per a poder obtenir els resultats finals. Tant tu com el Victor sou un gran exemple a seguir (espero algun dia també ser igual d'aplicada que el Victor planificant experiments). No et podré mai agrair prou el suport científic i personal que m'has donat. Crec que no oblidaré mai les nostres curses matutines a les 7 del matí per l'hospital, com m'arrastrava darrera teu i com, a força d'anar-hi, ens vam fer fortes juntes. Ens queda pujar al famós Mirador abans de que marxi!! AVISO!! Challenge accepted! A la **Mariona Parera**, que crec que totes del grup et devem un monument ENORME. Soc de les que flipa que et poguis "apanyar" amb els tres nens corrents amunt i avall i treballar alhora com una "jabata". Moltes gràcies per ser-hi sempre que et necessito, per la teva memòria i meticulositat alhora d'apuntar-te coses a la llibreta, per posar ordre a tot arreu (tranquil·la, ja recordaré d'ara a endavant a numerar els clons com toca!) i sobretot, per la teva ajuda inestimable al final al suportar-me en el meu pitjor moment. Vull agrair també als estudiants de pràctiques que han estat sota la meva tutela o al laboratori al mateix temps que jo (sota la tutela d'alguna altra del grup) i dels que he après sempre alguna cosa: la **Montse de Castellarnau**, la **Montse Puig**, la **Debora**, la **Rebeca** i el **Gerard Terradas** (el futur australià compartit amb Mol III). M'han fet veure que "ensenyar" el que saps al principi no sempre és fàcil, i que em queda molt de recorregut per poder-ho fer com "mana". Espero haver estat d'ajuda per ells i haver-lis transmès les ganes per seguir en aquest "mundillu".

Als MOLons III!!! Al **Roger Badia**, el periquito de irsiCaixa (a no ser que n'hi hagi algun altre que no hagi confessat), per la seva vitalitat i bon "feeling" que transmet, que des de el primer dia que va aterrar a la biblioteca tots vam sentir que era allà des de sempre. Per la teva meravellosa interpretació del Ferdinand rapero, crec que és insuperable! A la **Ester Ballana**, per estar SEMPRE disponible a ajudar, a millorar un experiment, a planificar noves coses, a ensenyar-te, a corregir textos, a aconseguir acabar "papeleos" varis per mi i sobretot, disposada a escoltar. Una pena la vida posi a prova persones tant bones com ella, i una sort poder tenir el petit Joan (el ajudant al ordenador!). Al **Emmanuel Gonzalez**, per què tots aquests anys tenint-te fins les tantes a la biblio feien que no em sentís tant sola en el nostre "zulo", per les converses metafísiques dins i fora de p3 i per ser un suport emocional quan calia. Espero que et vagi molt bé per Mèxic. A la **Encarna**, **Encarnita**!! Per les teves converses quan estàvem les dues amb el cap com un bombo, el seu accent andalús, el seu "desparpajo" i la teva valentia. Per ser capaç d'haver vist què volia fer a la vida per ser feliç, recapacitar i tornar enrere. Reitero, s'ha de ser molt valenta per fer això. A la **Mari Pau Mena**, la catalana que va marxar a Polònia. Per les tardes de divendres nit cansades a la biblioteca, per fer-me descobrir una gran sèrie ("In therapy"), pel seu riure contagiós i el seu amor als gossos. Al **Edu Pauls**, al retrobat, per què hem pogut recoincidir un temps i pels consells alhora de buscar postdoc. Al **Marc Permanyer** conegut entre els freaks que intentem córrer de tant en tant com "el Killian de irsi". Per les nostres converses metafísiques sobre el futur els caps de setmana, a p3, a les tantes i per despertar-me al mig d'un partit de hockey sobre gel! Relax, que tot arriba (jo ni m'ho crec per això!)! ;) A l'**Alba**

Ruiz, per tenir un cor tant gran que no sé com li cab, pensar sempre en els demés abans que en ella mateixa i per injectar tot i l'agobio de feina que portem (i porta ella) tots a sobre un somriure als demés. Al escocès **Jordi Senserrich**, se't troba a faltar amb les conyes, espero que et tractin bé pel Nord! A la **Eva Riveira**, per què et teu riure es contagia i per què sempre que he necessitat quelcom allà estaves!

Als VICtoriosos… (anda que el nom…). A **Jorge Carrillo**, el Pubmed combinado con la Wikipedia y Google andante! Al trabajador nato, gracias por estar siempre allí dispuesto a ayudar con cualquiera pequeña duda, por tus bromas, por tu salero y por el gran aceite de tu pueblo. Nos vemos a tu vuelta seguro o en NY! Al **Francesc Cunyat**, ciscu, cesc, paco, paquito (no m'he pogut resistir al final, no em matis ;)), per les mil hores compartides, ensanyar-me a fer còpies de seguretat a tot arreu (després de perdre junts el pendrive de 16gb, faig còpia fins i tot al dropbox!), els riures i les conyes, les converses infinites, aguantar el meu mal humor en mals moments, i per saber empentar-me (o picar-me més aviat) en qualsevol tema per treure el millor de mi mateixa. A la **Marta Curriu**, la "alegría de la huerta".. per la teva vitalitat i un gran OLE TU per venir des de Tarragona tants anys cap a Badalona a treballar. Espero que et vagi molt bé en la teva nova etapa! A la **Eli García**, per ser tant franca i directa i per ensenyar-me a veure que a vegades cal ser-ho per saber-te fer sentir. He empezado con dos clases de Yoga y reconozco que me calman los nervios! ;) merci por recomendármelo! A la **Eli Gomez**, felicitats pel carnet de conduir (ja és MOLT, a mi em va costar el pràctic també), per la seva vitalitat treballant i espero que durant la tesi no hagis de treure moltes plaques a les 9 de la nit, i sinó, ANIMS! A la **Silvia Marfil**, la gran patinadora de irsi!! Queda pendent que m'apunti a alguna sortida (a veure si ara que tindré una miiica més de temps abans de marxar m'hi apunto). merci per la organització (conjuntament amb la Mariona) de irsi en general!! la **Marta Massanella**... què puc dir d'ella si la recordo amb 6 anyets? Que marxessis se'm va fer molt dur, per què vaig perdre part del suport moral.. A la curranta sense limits, l'analitzadora de FlowJo, la meva "instructora" en el mòn Apple, en el "pirateo legal" dins i fora de irsi! Jujuju! I miss you darling! Del **Luis Molinos**, no oblidaré la conversa sobre música heavy que vam tenir a p3 i tu manera de derrapar andando, vas corriendo más que yo por el mundo!! A la **Marisa Rodríguez**, que, tot i haver coincidit poc temps al laboratori, he pogut comprovar que també és un pou de sabiduria.. molta sort en la teva etapa a irsi! A la **Lucía Pastor**, la que se nos fue a Mozambique a seguir su sueño! Espero que seas feliz allí y coincidir en algún momento! Finalment, per tancar el grup, a la **Isabella Puigdomenech**, la française d'adoption! Vaig heredar el teu lloc i crec que se'm va pujar al cap la teva "obsessió" pel rosa! ;) Espero que et vagi tot molt bé per Paris!

A las GRECas!! No se me va a olvidar en tiempo vuestra actuación durante la cena de este año! ;) A **Maria José Buzón**, por tu gran trabajo hecho y por hacer, por aconsejarme y ayudarme estos últimos meses. Nos vemos en nada en Boston! A la **Judith Dalmau**, pel desparpajo, per la curranta a les mil, per l'acompanyant al cotxe, per saber-me animar quan ho necessitava i per tenir més piles que el de duracell. Trobaré a faltar els nostres viatjes

amunt i avall! A l'**Itziar Erkizia**, Citómetro-arekiko zure laguntzagatik eskerrik asko eta nik harekin galtzen nuenean barealdia mantentzea! Y... patata-arrautzopil zoragarriengatik eta errezetagatik! (por si acaso el traductor no funciona bien: Gracias por tu ayuda con el citómetro y mantener la calma cuando yo la perdía con él! Y... por las maravillosas tortillas de patatas y la receta!). A **Nuria Izquierdo**, por tu ayuda estadística (que tan bien me fue) y por que parece que hablarle a tus dendris funciona. Creo que voy a empezar a hablarles a mis cultivos! A la **Esther Jimenez** (istar16).. qué puedo escribir que no sepas? Por tus mil horas escuchándome a las mil, por tus consejos y tus ánimos, por hacerme reír cuando lo necesitaba con cierta cancioncilla, por confiar en mi cuando lo has necesitado. Te echaré de menos… Alli estaré igualmente, a solo 2h30 de avión de Barcelona para cuando me necesites! Al **Gerard Minuesa**, al Yeral! Cada cop que escolto la cançó "I need money" no puc evitar recordar com cantaves a ple pulmó a p3, i has aconseguit dollars a NY!! Gràcies pels consells de San Francisco, van fer-me anar a llocs "amazing"! a veure si puc fer-te una visita a NY! A la **Ruth Peña**, per la dolçor que transmet, pels canvis de look de cabell i per aguantar els meus "consells" amb el iPhone i el IG! ;) Al **Dan Ouchi**, per la pèrdua de vergonya bailoteando pel vídeo de la Marta, i per què em "flipa" que et poguis passar el dia fent estadística! ;) A la **Maria Pino**, per la llumeta dels ulls que denota encara l'emoció del principi, per guanyar un premi al millor pòster sense saber-ho, per anar riu avall amb una cadira muntada sobre un neumàtic i per ser tant valenta! A la **MariCarmen Puertas**, l'altra enciclopèdia de protocols i de ciència de irsi! Gracias por tus consejos, tu ayuda, tu gusto compartido por la Nutella matutina (umm!), tus ganas de escucharme y tu cariño. Los echaré de menos. A la **Maria Teresa Rodriguez**, la de les dosis freakis de irsi, que vas confiar en mi! No sabré mai com agrair-te l'ajuda que em vas brindar en la nostra peripècia americana. A distreure'm en el viatge llarguíssim fins al Atlanta i després. Aguantar-me en aquell estat nerviós, ajudar-me a trobar l'hotel i el Ragon per l'entrevista i sobretot, a escoltar-me quan ho necessitava. Us anirà molt bé al Gio i a tu, no ho dubto! ;) finalment, a la **Maria Salgado**, me impresionaste en el CROI del año pasado con tu trabajo, tu desparpajo y tu alegría. Ha sido genial conocerte durante este año y compartir risas contigo.

Als Brander! A la **Ai**, la benretrobada al CROI d'Atlanta d'improvist, thanks for your company during the long hours in p3 and your kind words. You're a really hard hard hard worker! A la **Sandra Silvia Arrieta** (SSA), por sus consejos sobre como vivir en USA, y las charlas que tuvimos en la campana 10-11! A la **Vanessa Bach**, espero que aviat poguis començar a entrenar i còrrer com volguis, i sento haver-te tancat un dimarts que em vaig despistar amb l'alarma a p3 (mea culpa!). Algún dia m'has de passar la recepta dels muffins/cupcakes/etc que veig al IG!!! Al **Pep Coll**, per fer-me de "metge de passadís" i ser un dels bojos del club de bojos de les curses. M'has d'explicar com t'ha anat a la cursa d'orientació! Al **Javi Ibarrondo**, un científico divertido y muy "currante". No olvidaré nuestro viaje de ida y vuelta "flash" a Madrid. Aprendí mucho. Gracias por tu ayuda. A la **Anuska Llano**, gracias por todo, por escucharme cuando lo he necesitado. A la **Mireia Manent**, amb qui he concidit poc però el necessari per veure que és una curranta nata, que es pot "enfrontar" a mil "papeleos" sense parpellejar. A la **Bea Mothe**, la Doctora per partida doble. Gràcies per estar sempre que t'he necessitat disponible, pels consells sense preu i per les consultes de passadís. Per aconsellar-me llocs on anar a sopar el dia que celebri la tesi! ;) Al **Alex Olvera**, el begurenc

de irsi. No sé si donar-te les gràcies per repetir-me la frase de que "els predocs ens pensem que la vida de postdoc és millor! I no ho és!". Al menys, qui avisa no és traïdor! ;) A la **Marta Ruiz**, por compartir penas, por hacerme ver que todo tiene un final y que vamos a escribir un libro. Mucha suerte en tu nueva etapa de mamá. Finalmente, a **Jenny Zamarreño**, por su dulzura, su gran capacidad de trabajo y por haber sido valiente y reconducir la vida profesional para encontrar el camino a la felicidad.

Als GEM o Epidemiologia Molecular!! La **Rocío Bellido**, por tu bondad y por los entrenamientos de la cursa de los bombers que hicimos juntas! ;) fue duro pero divertido! A la **Maria Casadellà**. No sé si és per què jo tinc l'edat de ta germana i tu la de la meva germana petita, però sempre he sentit com un vincle amb tu. Per les converses que hem tingut, per les curses que hem corregut juntes, i per què sento que t'anirà molt bé! Al **Paco Codoñer**, el valencià Culé que ens feia berenar per fer un "descans", per les cerveses amb nervis pre-partit del Barça amb el Francesc fora del camp (amb el Paco al camp, sempre guanyem al Madrid!), per entendre de bioinformàtica (ja em sembla tot un mèrit!) i per les visites tant aquí com a València tot i el poc temps disponible! D'un bioinformàtic a l'altre bioinformàtic, al **Marc Noguera**. Per tenir la gran capacitat d'aparèixer de cop al nostre costat a Seattle i per què mirar la seva pantalla quan passes pel costat fa por de la de "comandos" que hi ha! Al **Christian Pou**, per "matar-nos" un dia corrents pels montes de irsi a la Maria i a mi i no riure-se'n, per les bromes, les conyes, les rises, la "rapidesa mental", pel bon rotllo que transmets i per ser el següent. Et passo el testimoni nanu!! GO GO GO! A la **Cristina Rodríguez**, de la que sempre m'al·lucina la calma que transmets (m'aniria bé una mica de la teva tranquil·litat!), suposo que necessària per poder enfrontar-te a la bioinformàtica! A la **Teresa Sequeros**, pels consells, per l'ajuda corregint mails "políticament correctes", per la paciència, per què transmets bon rotllo i per voler entendre'ns i aprendre quan parlem de les nostres cèl·lules i virus (té molt mèrit, ja que som molt freaks). Finalment, al **Mattia Schiaulini**, siguis on siguis, que et vagi molt bé. No oblidaré mai els glopets de grappa abans de començar a esquiar per "escalfar" motors amb l'Esther i el Francesc. Oju trompades posteriors (no fos per culpa d'això!)!!

A las "chicas de Lidia", les noies de Serveis. A la **Rafi Ayen**, per aguantar-me amb la meva música a p3 als migdies sense fotre'm bronca i per ser la tranquil·litat en persona. Hago una propuesta formal, ¡quiero estudiar tu sistema inmune a todo! A la **Samandy Cedeño**, o la chamita… qué decir.. son mil recuerdos de momentos compartidos que se agolpan en mi (cansado) cerebro…cursas, cocinitas en tu casa, cenas, conciertos, peleas con el HPLC, bailoteos, conversaciones (trascendentales y no tanto), visitas a dani… etc.. en resumen, que te voy a echar de menos y que Hamburgo está a 2h30! Espero una visita vuestra aviso! A la **Susana Esteban**, por tener la paciencia de ayudarnos a colocar o a encontrar material! Y por hacerme ver "el mundo real de la sanidad", porque tu lucha es totalmente necesaria, por ser tan valiente y decir las cosas claras. Ánimos!! A **Lucía Gomez**, por ser tan tan tan tranquila y metódica, quiero aprender a serlo algún día (ya me contarás el secreto! ;)). A la **Tània Gonzalez**, perquè el teu riure s'amplificava per tot p3, per inventar-te paraules que

s'enganxaven, per ballar tant bé i alhora ser vergonyosa (que només et vaig veure en un vídeo), per ser una viatgera de motxilla i "palante" (jo no ho podría fer) i per ser tant "dichaarachera". Espero que tinguis molta sort! A la **Eulàlia Grau**, per ser una curranta i per tenir la paciència d'aguantar una violinista novata a casa! (sé del que parlo! ;)). A la **Teresa Puig**, per tenir teories per tot, per les seves conyes, per escoltar Rac1 (i avisar-me quan surt el meu pare parlant) i per les teves anècdotes. Finalment, a la **Cristina Ramírez**, la veterinària de irsi, pels esmorzars titànics et senten tant bé. Merci per mirar-te les analítiques del Spock en el seu moment i per deixar-me el pàrquing quan vaig al Dani.

I… què faria sense el **Julian Eslava**, l'home que crec que és el més ràpid en recórrer l'espai des dels mòduls fins al meu lloc quan el Mac s'ha espallat, el PC s'ha penjat, el sequencher no reconeix la "contrasenya de xarxa".... Moltes gràcies per estar sempre disposat a ajudar-nos a la mínima i per "aguantar" i fer la "vista grossa" quan veus en els nostres ordenadors (ara en els personals només) programes no del tot legals. La **Matilde Gordero**, per què la vida dóna moltes voltes i ens hem acabat retrobant (des de Noruega a irsiCaixa) i per curranta! I d'una comunicadora a l'altra, la **Rosina Malagrida**. També, sembla que la vida té punts de connexió i aquest cop es fa a Hamburg amb la teva germana (encara em sembla fort). Moltes gràcies per la teva paciència quan vam fer la pàgina web! La **Cristina Mesa**, amb la organització dels viatges, de les reunions amb el jefe i per la seva bona pasta. I, at last but not at least, la **Penélope Riquelme**, que amb el seu somriure cada matí ja fa que entris diferent a irsi. Gràcies per la teva paciència durant aquest temps, per aconsellar-me estèticament i per la reunió de noies a la perruqueria de la teva germana que no oblidaré! ;)

Gràcies també al **Ferdinand**, per la seva sinceritat que em va fer riure, enrojolar-me i em va "xocar" durant el temps que vam coincidir i que em va fer conèixer la seva cultura. Per ser tant futbolero!!

Marta me dijo una vez, pon un bioinformático en tu vida! Descubrí en Atlanta que es de lo más útil saber "comunicarte" con el ordenador. Muchas gracias **Josué** por las gráficas del Heat Map, y por ponerme los colores que quise !! ;) Todavía te debo ese café!

Gràcies al **Ventura**, primer de tot, per haver tingut l'energia i l'empenta per crear un lloc de treball tant punter i agradable. En segon lloc, crec que tota la meva família i jo t'agraïm que no em deixessis baixar amb el meu cotxe el dia de la Gran Nevada (no per les cadenes en les que no confiaves, sinó que, amb la meva mala conducció en gel/neu, m'hauria estampat a mig camí de casa). Moltes gràcies també per haver estat el meu "manager", per haver-me empentat i ajudat a arribar a contactar amb el Marcus i posteriorment per ensenyar-me a ser políticament correcte en els mails! ☺ Et dec mínim un cafè encara! Muchas gracias tambien a **Lourdes**, por ayudarme cuando lo he necesitado (sobretodo al final) por estar siempre disponible para mi y por tomarte con buen humor (y ahora me desvelo) mi amigo invisible del cartelito "Toy" (tuve colaboración, pero eso no lo chivo, aunque no sé si era un secreto guardado a voces, pero soy muy vergonzosa). Me alegro que te gustara.

Finalment, merci a tots els jefes irsiCaixencs, al **Julià Blanco** (per aguantar i tenir paciència per les meves trucades des de p3 els caps de setmana per què falla l'aire, o una màquina o....), a la **Margarita Bofill** (per la seva gran capacitat per apuntar-se a un "bombardeo" als vídeos de les tesis, per la seva gran carrera i espero que poguis descansar una mica ara), al **Christian Brander** (per dedicar-me el seu temps en aconsellar-me i per què em va "flipar" el dia que el vaig veure arribar en bicicleta al hospital), a la **Ceci Cabrera** (por los fines de semana y tus "guarradas" en p3 (jujuju!) con los tejidos y enseñarme a "ver" el tejido linfático (olé tu para verlo!)), al **José Esté** (per les seves felicitacions quan enviava mails de pastissos i els seus consells al CROI de Seattle), a la **Julia G Prado** (por compartir penas del Mac, por enseñarme (por culpa de mi insistencia) una foto preciosa, por darme consejos de postdocs a distancia, por ser tan valiente!), al **Javier Martinez-Picado** (per la seva capacitat excepcional per poder entrenar al mateix temps que publicar de manera excepcional i tenir família!!), al **Roger Paredes** (per la seva música, per les seves bromes, pels seus consells) i a la **Lidia Ruiz** (per tenir la paciència de no ofegar-me quan venia a demanar-li papers a últim dia de la beca i ajudar-me).

A la meva tutora de tesi, la **Josefa Badia**, per la seva paciència sempre que la trucava (encara estic sorpresa que em reconegui per telèfon), per la seva ajuda amb la burocràcia vària des de fa tants anys i el seu suport quan vaig decidir canviar el rumb. Gràcies.

No tot és feina, molta gent m'ha acompanyat en aquest viatge de 5 anys. Als de la Uni, les Pitxurris i els nois. Per què ja fa 12 anys que anem junts i, tot i les distàncies que ens separen a vegades, hi ha una espècie d'imant que ens segueix unint. Gràcies a les noies: a la **Alex (Xandri)**, per ser més despistada que jo i sortir-te'n (ja m'explicaràs, ja que jo no ho aconsegueixo), per què ets un sol i per fer-me riure quan ho necessitava**;** a la **Laura,** que ens vam conèixer en aquell autocar que ens portava de pràctiques al 2001, per estar al meu costat i per què la vida et porta cap a un camí preciós amb el vostre Arnau; a la **Cris P**, la nostra psicòloga de guàrdia, per que m'al·lucina l'energia que tens per fer un Màster, intentar guanyar peles i fer mil pràctiques a llocs diferents, això és lluitar! La pròxima, baixem a les teves terres; i què dir de les "vecinas" **Cris L**., i l'**Anna**, per les cerveses o vinets blancs post-dies durs, pel suport, per les carreres pels montes, per les converses terapèutiques, per estar dia a dia al meu costat i per entendre l'histerisme. Prometo que se'm passarà en breu. Al **Joan**, pels congressos de virologia junts (sembla mentida) i per aquell Sant Joan a Salamanca i al **Juancho** per què les converses al telèfon poden ser llargues, por que por fin acabo y por no robarme mi tesis! ;) Als "desperdigats", a la **Gemma**, al **Benet** (vinc cap a la costa Est un temps! A veure si puc fer escapada a NY!!), al **Edu**, el **Norbert** i al **Meg**, per les seves conyes, pel seu desparpajo i per què quan ens retrobem no sembla que hagi passat el temps. Evidentment, no oblidar-me dels "maschios" alpha i beta de StQ, el **Mariano** i el **Raul**, que fan que les sortides a córrer o a Pals siguin encara més divertides, com tampoc ni del **Joan** (el papa) i la **Maya** (amb les seves calçotades boníssimes i la medecina xinesa, que m'al·lucina).

Acknowledgements

Als del Lycée..gràcies per les calçotades, els cumples G&G, pels casaments, per les festes de cumple, pels 30 que complim aquest any... i aquí copiaré la Marta: "Qui sont les meilleurs?? Evidemment, c'est nous!!" (diga'ns "humils"). A la **Marta** i la **Marina**, per la seva grandíssima paciència aquestes últimes setmanes, per llegir-s'ho tot i no morir d'avorriment, per ajudar-me a millorar, per què us recordo d'enanes (jo encara ho deuria ser més) i per haver crescut juntes!! Us dec un monument enorme, a part d'una festa "comme il faut" (Marta, hauràs de venir cap aquí!!). A les demés nenes, la **Clara** (per ser més que una germana i per fer-me tieta jove!), l'**Ana**, la **Gemma**, la **Paula**, l'**Alèxia** i la **Mar**, pels sopars mensuals que "distreuen" positivament, pel nostre negoci amb el que estic segura que petarem Barcelona, per les festes i les coses que ens queden per viure juntes (nens, comiats de solteres, casaments...), merci! Als nois, al **Gerard**, al **Sergio** (uiii el futuro papa!), al **Jordi** i a l'**Anna** (nois, de nou, felicitats!), al **Albert J**, al **Albert Camps** (ai no, de Azpiazu), al **Luis**, al **Hector G**, al **Gabri C** i la **Estrella**, al **Gabri B**, merci pels moments de distracció, per ajudar-me a no dormir-me a les taules sopant els divendres nit, per aguantar-me histèrica i per fer-nos grans junts.

Als de la família ballenita i ocellet, al **Albert** i la **Silvia**, al **Dani**, al **Juan Daniel**, a la **Vanessa** i al **Sergio**, gràcies per fer-me riure amb els whatsapps de chistes "malos" a les tantes, pels Cavatasts i per entendre que no estic sempre disponible (tot i que ja m'agradaria). Merci! A la meva **família política** en bloc (sois muchos!!), per la seva paciència i dedicació.

Als qui han hagut de suportar-me en els pitjors moments, en les crisis, els que m'han empentat a continuar des de fa tants anys, a lluitar en els moments difícils de malaltia, a no parar i, a aguantar-me en el cansament, amb un somriure. Em sap greu que totes aquestes hores de feina m'hagin pogut treure temps d'estar amb vosaltres. A la meva **família**, als que hi son i als que ja per desgracia no, als meus **pares (**el **Ferran** i la **Isabel)**, a la meva germana, la **Clara** (i al petit/a que ve en camí), i als meus quatre avis, la iaia **Pepita**, l'avi **Martí**, la iaia **Montse** i l'avi **Joan**, per ser tant pacients, transmetre'm els gens necessaris (havia d'haver un punt freak aquí) i sempre trobar les paraules justes per animar-me a continuar lluitant i creixent com a persona. I al **Spock**, que per mi era un més de la família, que sense paraules sabia com arrencar-me un somriure i fer-me desconnectar. Aquesta tesi és gràcies a tots els ànims i l'esforç que m'heu sabut transmetre. Merci.

I finalment, a l'**Héctor**, que fa anys i panys que ha sabut acceptar que això de ser científica no té hores, que té més paciència que un sant, que em fa de company de vida i de laboratori els caps de setmana, no podré mai agrair-li prou el que ha fet per mi. Demano perdó per fer-te perdre temps comú junts (content? ;)). Sense el teu suport incondicional hi hauria hagut moments en els que no hauria aguantat. Gràcies.

IrsiCaixa
Institut de Recerca de la Sida

UNIVERSITAT DE BARCELONA