Analysis of the impact of codon choice on gene expression of HIV-1

DISSERTATION ZUR ERLANGUNG DES

DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)

DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN

DER UNIVERSITÄT REGENSBURG

vorgelegt von

Thomas Schuster

aus

Kronstadt

im Jahr 2019

Das Promotionsgesuch wurde eingereicht am:

26.04.2019

Die Arbeit wurde angeleitet von:

Prof. Dr. Ralf Wagner

Unterschrift:

Thomas Schuster

A Contents

A Contents ..... - 4 -

B Abstract ..... - 8 -

C Introduction ..... - 9 -

C.1 The human immunodeficiency virus -1 (HIV-1) ..... - 9 -

C.1.1 Epidemiology ..... - 9 -

C.1.2 Structure of HIV-1 ..... - 10 -

C.1.3 The life cycle of HIV-1 ..... - 11 -

C.1.4 Genome organization and viral gene expression ..... - 12 -

C.1.5 The HIV-1 Gag Polyprotein ..... - 13 -

C.1.6 The role of the Rev protein and the Rev-responsive element in the life cycle of HIV- 14 -

C.1.7 Usage of a subgenomic *gag* reporter system to analyze the effects of codon adaptation on viral gene expression ..... - 16 -

C.1.8 The role of the Tat protein in the transcription of viral genes by the cellular RNA polymerase II. - 16 -

C.2 RNA export from the nucleus to the cytoplasm ..... - 17 -

C.3 The genetic code and its role in protein biosynthesis ..... - 18 -

C.4 Codon usage and codon usage bias ..... - 20 -

C.5 Codon optimization and deoptimization ..... - 21 -

C.6 Nucleotide composition of HIV ..... - 22 -

C.7 Objective ..... - 23 -

D Material and methods ..... - 25 -

D.1 Cell lines ..... - 25 -

D.1.1 Prokaryotic cell lines ..... - 25 -

D.1.2 Eukaryotic cell lines ..... - 25 -

D.2 Nucleic acids ..... - 25 -

D.2.1 Oligonucleotides ..... - 25 -

D.2.2 Plasmids ..... - 31 -

D.2.3 Antibodies ..... - 32 -

D.2.4 Commercial Kits ..... - 32 -

D.2.5 Standards ..... - 33 -

D.2.6 Computer programs and databases ..... - 33 -
D.3 Cell culture techniques ..... - 34 -
D.3.1 Cultivation of eukaryotic cells ..... - 34 -
D.3.2 Transfection of eukaryotic cells ..... - 34 -
D.4 Microbiological methods ..... - 35 -
D.4.1 Growth and selection of prokaryotic cells ..... - 35 -
D.4.2 Transformation of chemically competent cells ..... - 35 -
D.5 Molecular biology methods ..... - 35 -
D.5.1 Photometric quantification of nucleic acid concentrations ..... - 35 -
D.5.2 Isolation of plasmid DNA ..... - 36 -
D.5.3 Polymerase chain reaction ..... - 36 -
D.5.4 Fusion PCR ..... - 37 -
D.5.5 Standard cloning procedures ..... - 38 -
D.5.6 Cloning with exocutter BsmBI ..... - 38 -
D.5.7 Agarose gel electrophoresis ..... - 38 -
D.5.8 Detection of potential cryptic splicing products ..... - 39 -
D.5.9 Generation of completely and partially codon adapted *gag*-variants ..... - 39 -
D.6 Working with RNA ..... - 39 -
D.6.1 Isolation of total RNA ..... - 40 -
D.6.2 Nuclear Run on assay ..... - 40 -
D.6.3 Determination of mRNA half-life ..... - 41 -
D.6.4 Reverse transcription (synthesis of copy DNA) ..... - 42 -
D.6.5 Quantification of mRNA expression levels by quantitative real-time PCR (RT-qPCR) ..... - 42 -
D.7 Protein biochemistry techniques ..... - 47 -
D.7.1 Lysis of HEK293T cells ..... - 47 -
D.7.2 Quantification of total protein amount by Bradford analysis ..... - 47 -
D.7.3 Quantification of Gag expression by p24-Enzyme-linked Immunosorbent Assay (ELISA) ..... - 48 -
D.8 Analysis of GFP expression by flow cytometry ..... - 49 -
D.9 DNA Sequencing ..... - 50 -
E Results ..... - 51 -

E.1 Generation of partially humanized *gag* variants ..... - 51 -
E.2 Sequence properties of partially humanized *gag* variants ..... - 52 -
E.3 Influence of humanization on *gag* protein expression of 5'-3' adapted variants ..... - 54 -
E.4 Effect of Rev inhibition by addition of Leptomycin B on the *gag* expression of the 5'- 3' adapted *gag*-variants ..... - 55 -
E.5 Influence of humanization on Gag protein expression of 3'-5' adapted variants ..... - 57 -
E.6 Characterization of the impact of the 5' end of the HIV-1 *gag* on gene expression of codon adapted variants ..... - 58 -
E.6.1 Analysis of the impact of the first 423 bp at the 5' end of HIV-1 *gag* ..... - 58 -
E.6.2 Analysis of the importance of the first 100 base pairs at the 5' end of HIV-1 *gag* ..... - 59 -
E.6.3 Localization of a potential inhibitory sequence motif at the 5' end of HIV-1 *gag* ..... - 61 -
E.7 Confirmation of the position of the inhibitory sequence motif ..... - 63 -
E.8 Influence of humanization on mRNA expression levels of 5'-3'- and 3'-5'-adapted variants ... . - 64 -
E.9 Analysis of transcription efficiency for selected partially humanized variants ..... - 66 -
E.10 Determination of mRNA stability of selected *gag* variants ..... - 69 -
E.11 Detection of potentially generated cryptic splicing variants ..... - 72 -
E.12 Impact of the Splice Donor SD1 on *gag* expression ..... - 73 -
E.13 Analysis of expression of partially humanized *gag* variants under control of the LTR promoter ..... - 75 -
E.14 Analysis of the impact of codon adaptation on a quasi-lentiviral GFP reporter ..... - 76 -
E.15 Influence of the identified inhibitory HIV-1 *gag* motif on *egfp* expression ..... - 79 -
F Discussion ..... - 81 -
F.1 Generation of partially humanized variants and analysis of gene expression ..... - 82 -
F.1.1 Generation and analysis of partially humanized *gag* variants ..... - 82 -
F.1.2 Generation and analysis of partially humanized *egfp* variants ..... - 83 -
F.2 Localization and examination of the inhibitory motif within the 5' part of HIV-1 *gag* ..... - 84 -
F.3 Molecular mechanisms that contribute to altered expression of partially codon-adapted genes ..... - 85 -
F.3.1 Molecular mechanisms that contribute to the general dependency of protein production from the amount of humanized codons ..... - 85 -

F.3.2 Molecular mechanisms that contribute to the function of the inhibitory motif on protein production ..... - 87 -
F.3.3 Influence of changed mRNA secondary structures on altered gene expression ..... - 88 -
F.3.4 Influence of splicing on altered gene expression ..... - 91 -
F.3.5 Impact of the promoter on altered gene expression ..... - 92 -
F.4 Future prospects ..... - 92 -
G Appendix ..... - 94 -
G.1 Northern Blot analysis of 5'-3'-adapted variants ..... - 94 -
G.2 Comparison of p24 levels in the supernatant and intracellular p24 levels of transfected HEK293T cells. ..... - 94 -
H List of References ..... - 96 -
I Danksagung ..... - 105 -

B Abstract

The degeneracy of the genetic code with its 61 codons encoding for only 20 amino acids is the basis for a phenomenon known as codon usage bias. That means that different organisms show differences in the frequency of occurrence of synonymous codons. Despite the necessity of HIV to use the host's translational machinery for viral gene expression, the virus exhibits an adenine-rich nucleotide composition which differs clearly from the GC-rich coding regions of humans. Changing the nucleotide composition by choosing synonymous codons at certain positions might therefore impact viral replication by affecting viral gene expression. Previous studies of our group showed that adapting the *gag* gene to human codon usage (huGag) led not only to a significantly increased protein production but also caused independency of Rev, an accessory protein of HIV that exports intron-containing viral mRNAs. The aim of this work was to gain insight into the effects of codon adaptation, especially regarding length and position, and the associated impact on *gag* expression. For this, subgenomic *gag* reporter constructs were generated that systematically varied the humanized part of the gene. Those constructs were then transfected into HEK293T cells. Gag expression was investigated on protein level by p24 ELISA as well as on RNA level by qPCR. Furthermore, transcription efficiency as well as RNA stability were analyzed using nuclear run-on and actinomycin D assays. It was observed that increasing the length of the humanized sequence starting from the 5' end directly correlated with p24 and *gag* mRNA levels. Contrary to that, such a correlation was lacking for constructs humanized progressively in 3' to 5' direction. It became apparent that humanization of the 5' end of *gag* is necessary for enhanced protein production and Rev-independent expression. In addition, it was found that a short sequence, surrounding the nucleotides 61-75 of *gag* in the 5'-part of the gene contains a specific inhibitory motif which affects transcription rate as well as RNA stability. Moreover, the expression pattern of the different variants was analyzed under the control of a heterologous CMV promoter as well as an LTR promoter of HIV-1. For both systems, comparable expression patterns were observed. By PCR analysis of reverse transcribed RNA from transfected cells, as well as northern blot analysis, the generation of cryptic splicing variants was excluded. Further, a second reporter gene was used in order to investigate whether the function of the identified inhibitory motif could be transferred. For this, a quasi-lentiviral system was used to express *egfp* after its adaptation to human and HI-viral codon usage. It became apparent that the inhibitory effect of the identified motif was only transferable when it was embedded in a larger part of wild-type *gag*. The influence of the inhibitory motif on HIV *gag* expression as well as the missing inhibitory effect on *egfp* could be based on characteristics of the mRNA secondary structures. For *gag*, but not for *egfp* variants, presence of the motif had a clearly destabilizing effect in mRNA folding predictions, which might influence protein binding of the RNA degradation machinery. In future experiments an examination of this hypothesis would be eligible and helpful to gain further insight in the connection between altered codon usage and gene expression.

C Introduction

C.1 The human immunodeficiency virus -1 (HIV-1)

C.1.1 Epidemiology

Despite intensive research for more than 30 years [2] and accompanying major advances in medical treatment, the infection with the immunodeficiency virus (HIV) still represents a life threatening situation for millions of people. Untreated, an HIV infection leads to the *Acquired Immune Deficiency Syndrome* known as AIDS. The course of this disease is typically characterized by three phases. Firstly, infected people show influenza-like symptoms caused by a high viremia and a rapid decline in CD4-positive T-cells [3]. The second phase is a symptom-free state, often mediated by a T-cell driven immune response against HIV [4]. Lastly, the continuous depletion of CD4-positive T-cells leads to an increase in viral load. This phase is generally linked to opportunistic infections and tumors, which are typically for the disease pattern of AIDS [5].

The adequate availability of *Antiretroviral Therapy* (ART) still is a problem in some areas of the world. So, despite constantly increasing ART coverage over time, in 2017 only about 21 million patients of the estimated 37million infected people are receiving an adequate therapy, which accounts for only around 60%. Especially in Western and Central Africa and the Eastern Mediterranean, ART coverage is not satisfactory with 29% or even 18% of HIV positive men and 48% and 19% of HIV positive women receiving ART, as shown in Figure C-1[1].

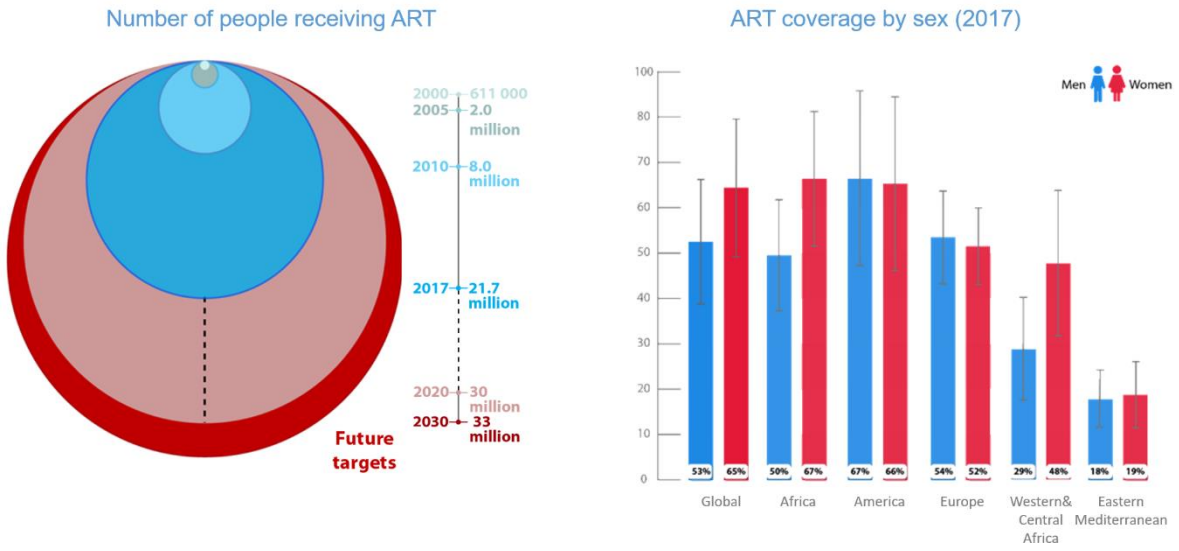

Figure C-1: Accessibility of antiretroviral treatment to HIV positive individuals
Development of the numbers of ART-receiving people worldwide (left) and ART coverage for males and females in different WHO regions. Modified after [1]. In 2017 only 21.7 million of the estimated 37 million HIV-infected people received ART. Especially people in Western & Central Africa as well as the Eastern Mediterranean were not supplied with an adequate therapy.

HIV-1 can be classified into the four Groups M, N, O and P which arose through zoonotic transmissions of the chimpanzee simian immunodeficiency virus (SIV_{CPZ}) or gorilla immunodeficiency virus (SIV_{GOR}) to humans [6]–[8]. Group M, which causes most of the infections, can be divided into nine clades (A-D, F-H, J and K). Besides HIV-1, there is also the human immunodeficiency virus type 2 (HIV-2), which is predominantly found in West African nations, though increasing numbers of infections have been recognized also in other parts of the world. The modes of transmission are equal to those of HIV-1, namely sexual contact, blood-borne exposure (blood transfusion, shared needles), and perinatal transmission [9].

C.1.2 Structure of HIV-1

HIV-1 is a member of the lentiviruses which belong to the viral family *retroviridae*. Because of the intensive research over the last decades most questions regarding the general structural features as well as the viral replication have been solved and can be reviewed in different excellent publications, books and reviews from which the following information are derived [10]–[15]. A characteristic feature of retroviruses is the reverse transcription of their RNA genome in double-stranded DNA, followed by the integration into the host's genome. Infectious HIV-1 particles contain a membrane which is derived from the host's cell membrane and encircles a conic capsid. Furthermore, the viral membrane contains the envelope protein (Env), which is the only viral protein that is present at the particle's surface and consists of gp120 and gp41. The matrix protein (MA, p17) is bound inside of the membrane by an N-terminally attached myristoyl group. The conic capsid is formed only by the p24 capsid protein (CA) and contains the viral enzymes reverse transcriptase (RT), protease (PR), integrase (IN) as well as the viral genome [10]–[15].

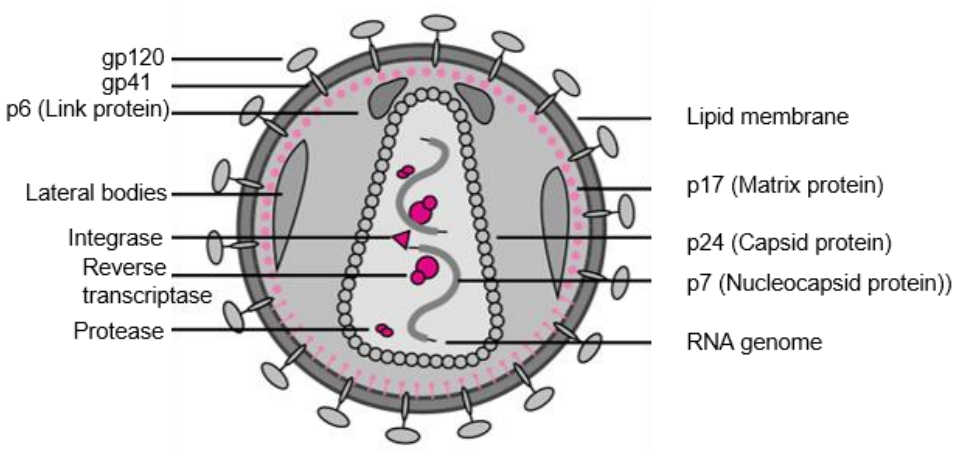

Figure C-2: Schematic structure of an HIV particle (modified after [10])
The lipid membrane of HIV is derived from the host's plasma membrane by budding of the virus and contains the only viral surface molecules gp41 and gp120. Inside of the lipid membrane the matrix protein (MA, p17) is attached. Inside of the conic capsid which is built by the capsid protein (CA, p24) the two identical RNA copies are complexed with nucleocapsid proteins (NC, p7). Furthermore, the capsid contains the viral enzymes reverse transcriptase (RT), integrase (IN) and protease (PR).

C.1.3 The life cycle of HIV-1

The initial adsorption of the viral particle to the cells is mediated by an interaction of the external glycoprotein gp120 to the CD4 receptor of the target cell e.g. T-helper cells, monocytes, macrophages and dendritic cells. This event is followed by a conformational change of the gp120 protein which allows the interaction with either the CCR5 or CXCR4 chemokine receptor. These processes enhance the binding affinity between virus and cell surface which allows the fusion of the membranes and therefore enables the entry into the cell. This is followed by uncoating and reverse transcription of the viral RNA genome in double-stranded DNA, which generates the long terminal repeats at both ends of the genome (see below). The RNA part gets degraded during this process by the RNase-H function of the reverse transcriptase. The double-stranded DNA molecule is transported into the nucleus as a pre-integration complex (PIC) together with associated cellular and viral proteins like the integrase, Vpr and matrix protein (p17). A phosphorylated form of p17 stays connected with the

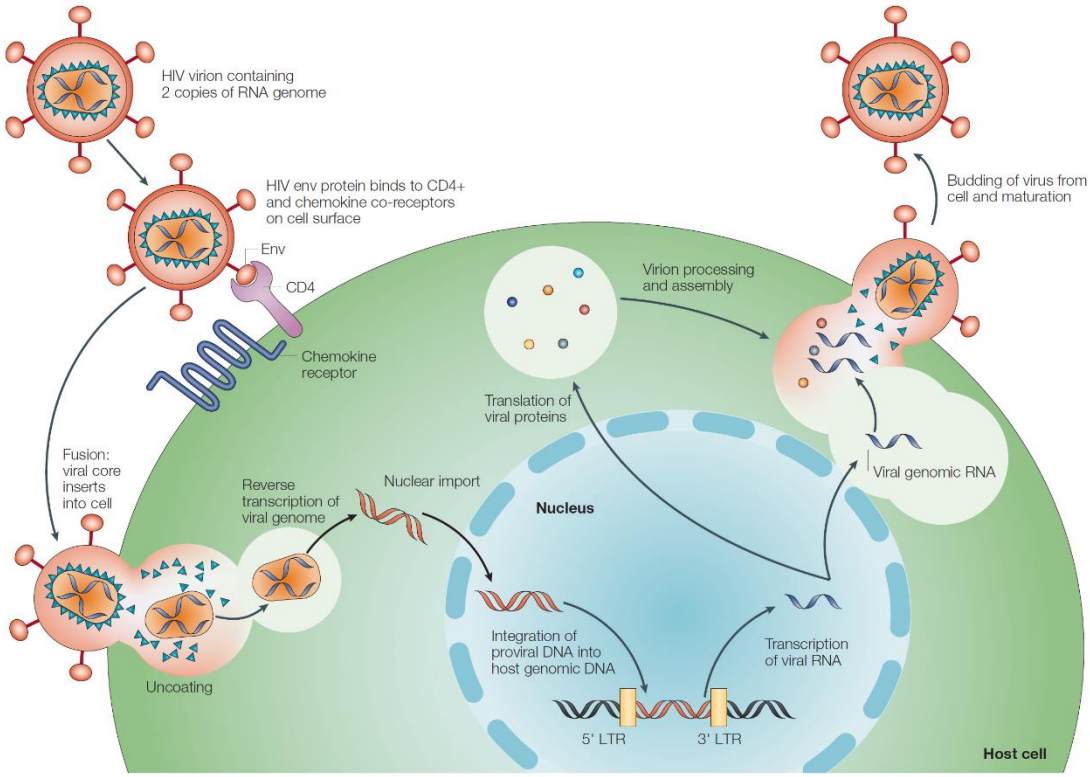

Figure C-3: Replication cycle of HIV-1 (from [16])
Initial binding of HIV to the target cell is mediated by interaction of gp120 and the CD4 receptor of the host cell. After a conformational change an additional interaction with the chemokine receptor CXCR4 or CCR5 occurs. This allows the fusion of the membranes of the virus and the host cell, which is followed by uncoating and reverse transcription of the viral genome in double-stranded DNA. The reverse-transcribed genome is transported with cellular and viral proteins as a pre-integration complex into the nucleus. The viral protein integrase mediates the integration of the viral genome into the host's genome, followed by transcription. The transcribed viral RNAs are used for protein synthesis and as viral genomes for progeny virions.

pre-integration complex and allows, together with the Vpr protein and cellular nuclear import factors, the transport of the reverse-transcribed genome through the nuclear pores, which enables

the infection of non-dividing cells. Inside the nucleus the viral integrase mediates the integration of the genome at an unspecific position, followed by transcription of viral RNAs. The different classes of transcripts are used either for protein synthesis or as viral genomes for progeny virions [16]–[20].

C.1.4 Genome organization and viral gene expression

HIV-1 contains two identical copies of (+)-sense, single-stranded RNA molecules in a complex with nucleocapsid proteins (NC, p7). Since the two RNA molecules have a 5' cap as well as 3' polyadenylation, they show typical features of a eukaryotic mRNA. The genome of HIV-1 (as shown in Figure C-4) has a size of about 9700 nucleotides which encode for structural genes (*gag, pol* and *env*) as well as regulatory and accessory genes (*tat, rev, nef, vif, vpr* and *vpu*). After reverse transcription of the RNA into double-stranded DNA, and following degradation of the viral RNA, the proviral DNA is generated by integration into the human genome. During that process, long terminal repeats (LTRs) are generated which contain all cis-active sequences as well as promoter and enhancer elements which control the retroviral gene expression. The viral genes get transcribed by the cellular DNA-dependend RNA polymerase and are expressed from partially overlapping open reading frames. Lastly, after alternative splicing, the mRNAs get translated by the cellular translation machinery.[10], [21].

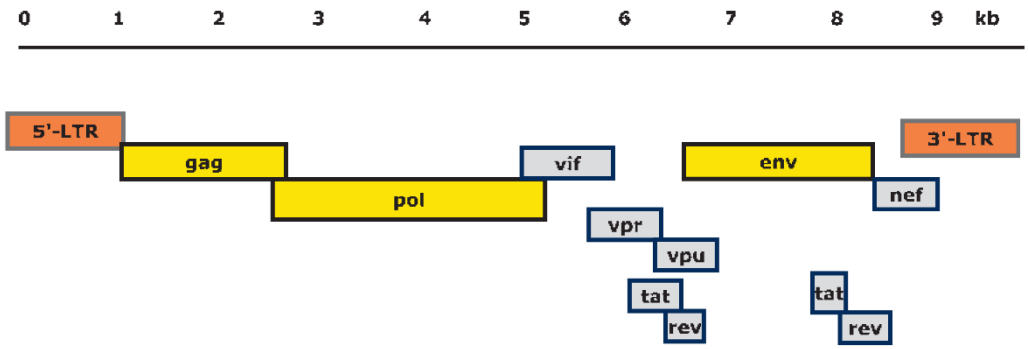

Figure C-4: Schematic structure and organization of the HIV-1 genome:
During integration, long terminal repeats (LTRs) are generated. The structural genes gag, pol and env are shown in yellow, whereas regulatory and accessory genes are depicted in grey. The open reading frames of tat and rev are encoded by two exons each.

The produced transcripts of HIV during infection can be classified into three groups. Shortly after infection, only short, completely spliced mRNAs are produced which encode the regulatory viral proteins Tat, Nef and Rev. Over time, the transcription rate gets strongly enhanced and incompletely spliced mRNAs are produced. Those mRNAs encode the envelope protein (Env) as well as the accessory proteins Vif, Vpr and Vpu. Additionally, the last group of transcripts, the unspliced mRNAs are used for translation of the Gag-Pol polyprotein and as new genome for progeny virions. This complex and strongly regulated gene expression pattern is significantly regulated by the two proteins Tat and Rev. Tat enhances transcription by binding to a regulatory RNA element called TAR and

recruits elongation factors to the transcription complex resulting in strong enhancement of transcription. The viral protein Rev, which role is described in detail in C.1.6, plays an essential role in mediating the cascade-like gene expression [22]–[24]. Among others, HIV Gag that was mainly used to analyze the influence on codon usage on viral gene expression in this work, depends on the Rev-mediated mRNA export.

C.1.5 The HIV-1 Gag Polyprotein

The Gag (group-specific antigen) polyprotein of HIV-1 is the main structural protein of the virus and is translated from an unspliced 9 kb transcript. Gag plays also an important role in the immune response against HIV, since the initial HIV-specific immune response is directed against Tat, Rev and Gag. Because the regulatory and accessory proteins show a higher mutation rate, the immune response against Gag plays a key role to fight the virus.[25], [26]. Intensive research has shown the importance of Gag for the viral life cycle. Initially assumed as a simple scaffold protein, Gag meanwhile is known to specifically recognize genomic RNA as well as viral and cellular proteins [27]. Furthermore, Gag is necessary and sufficient for budding of HIV from the host's plasma membrane [28]. Therefore, Gag has to package two copies of viral genomic RNA per particle [29] and interact with cellular trafficking proteins to hijack the endosomal-sorting complexes required for transport (ESCRT) system [30]–[32].

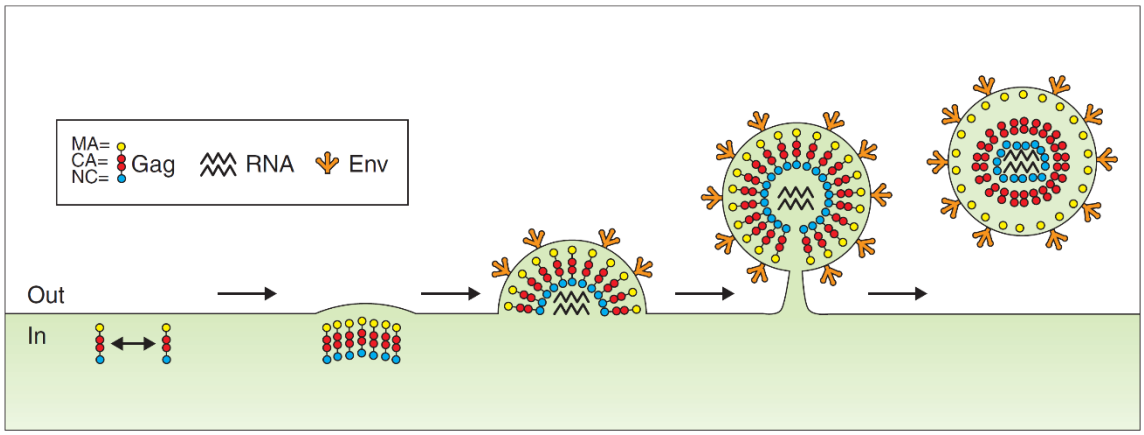

Figure C-5: Schematic overview of the budding process of HIV-1 mediated by Gag (from [33])
Gag proteins move towards the plasma membrane, interact with each other and induce extrusion of the lipid bilayer, followed by the membrane fusion to pinch-off the virion. Lastly, proteolytic processing by the viral protease leads to the maturation of the viral particle [33], [34].

Structurally, Gag can be divided into four major domains: The matrix, capsid, nucleocapsid and link protein, each playing a different role during the viral life cycle.

The matrix protein consists of 128 amino acids and plays an essential role for plasma membrane targeting and viral assembly. For both processes, an N-terminal myristoylation at Gly-2 plays a key role [35], [36], since it allows the binding of Gag to cholesterol- and sphingomyelin-rich micro domains within the plasma membrane [37].

The capsid protein is responsible for forming the core of the virus and can be divided into a C-terminal and N-terminal domain (CTD and NTD), joined by a flexible linker [38], [39] and a hexameric protein lattice is produced by CTD interaction[40]. Since formation of the mature and immature virus is strongly dependent on the CA-CA interaction, inhibitors targeting different sites of the capsid protein can have severe effects on viral replication, making CA an interesting target for an antiviral therapy [27], [41].

The nucleocapsid protein (NC) has several functions despite being one of the smallest parts of Gag with only 55 amino acids. During viral maturation and assembly, NC is involved in Gag-Gag interaction as well as in the recognition of the nucleic acid [27]. An important feature are two zinc finger motifs which are separated by a functionally important basic domain [42]. Furthermore, NC acts as an RNA chaperone and supports reverse transcription as well as the integration in the host's genome [10]. The last component of Gag is the 52 amino acid long link protein (p6). It gets translated as two different forms, the in-frame Gag p6 and the -1 frameshifted Gag–Pol p6*. It plays an important role for the release of new virions from the plasma membrane because of the so-called late domain of HIV, a sequence which seems to be responsible for the separation of the viral envelope and the cellular plasma membrane [43]–[46]. As mentioned above, the expression of Gag is strongly dependent on the accessory protein Rev and its interaction with a conserved sequence called the Rev-responsive element.

C.1.6 The role of the Rev protein and the Rev-responsive element in the life cycle of HIV

The regulatory protein Rev is a 13-kDa phosphoprotein which allows the export of unspliced and incompletely spliced viral mRNAs. Rev consists of 116 amino acids which are encoded by two exons and consists of several distinct domains [10], [47]–[50]. At the N-terminus an arginine rich motif functions as a nuclear localization sequence (NLS) [51], [52]. The interaction with a conserved RNA secondary structure called Rev-responsive element (RRE), which is present in all unspliced and incompletely spliced HI-viral mRNAs, is mediated by a arginine rich domain between amino acid 35 and 50 [10], [52], [53]. The RRE is a ~ 350 nucleotide long, highly structured element inside of the *env* RNA. It forms a conserved secondary structure with five stems around a central loop, according to

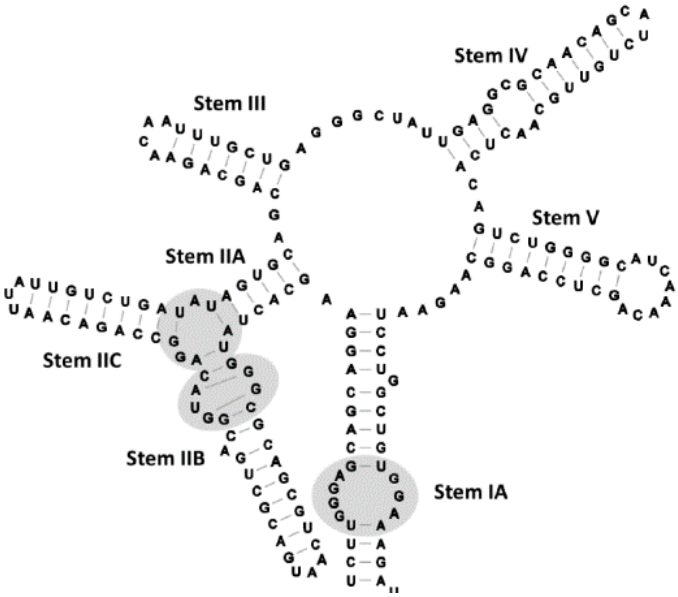

Figure C-6: Five stem model of the Rev-responsive element from [34]: Rev binding sides are shown in grey

the original model. Later, a 4-stem model suggested that stem loop III and IV form a hybrid III/IV stem loop together with the intervening loop region [54]–[56]. Initially Rev binds to the IIB site, followed by a multimerization of Rev by five additional molecules. The bi-directional transport is mediated by a nuclear localization signal (NLS) and a distinct nuclear export signal (NES). The NLS overlaps with the RNA-binding domain, while a leucine-rich sequence acts as the NES, allowing the interaction with cellular mRNA export proteins [47], [57]. Especially the amino acids between position 73 and 84 are essential for the Rev-mediated nuclear export, since they mediate the interaction with exportin 1, also known as CRM1 (chromosomal maintenance gene 1) which binds to the GTP-loaded form of the Ran protein. Thereby, the mRNA-Rev-protein-Crm1-Ran-GTP complex is exported out of the nucleus through the nuclear pores. By hydrolyzation of GTP, Crm1 and Ran disengage themselves from the export complex and Rev dissociates from the mRNA which now gets translated. By this mechanism HIV ensures that all viral gene products can get translated in a highly regulated way (see Figure C-7).

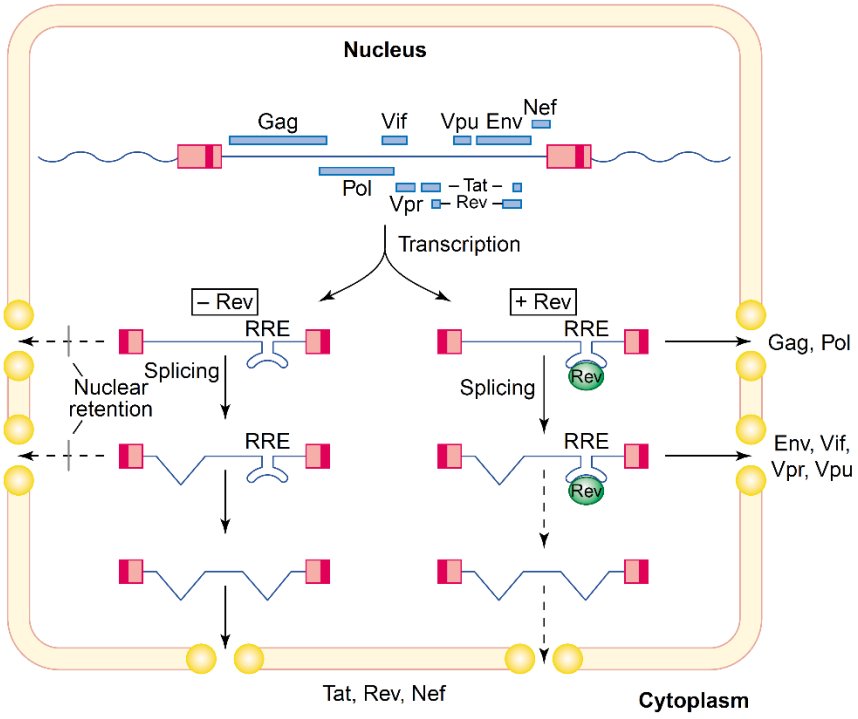

Figure C-7: Role of Rev in the HIV-1 life cycle (from [58])
HIV transcripts can be classified into unspliced, incompletely spliced and fully spliced mRNAs. At the beginning of infection, in the absence of Rev, only fully spliced mRNAs can be exported out of the nucleus. From those mRNAs, Tat, Nef and Rev get translated. Unspliced and incompletely spliced mRNAs are retained inside of the nucleus until Rev migrates back, binds to a secondary structure known as Rev-responsive element, which is present in all variants of unspliced and incompletely spliced transcripts and mediates the export of those mRNAs. For this, Rev interacts with exportin 1, also known as CRM-1 which mediates the export through the nuclear pore [58].

The fully spliced transcripts that get exported even in the absence of Rev use cellular proteins for the mRNA export. The correct transport of all different classes of RNA in general is a highly regulated process and is described in the following chapter.

C.1.7 Usage of a subgenomic *gag* reporter system to analyze the effects of codon adaptation on viral gene expression

To analyze the effects of altered codon usage on *gag* expression independently of other processes in the life cycle of HIV like infectivity, uncoating, reverse transcription or budding, a simplified *gag* reporter system was used, which was established in previous studies of our group. For this, the fact was used that HIV-1's late gene expression depends on cis-acting elements as well as the interaction of Rev with the RRE. For mimicking of this situation the 5' UTR of HIV-1 carrying the highly functional major splice donor SD1 was fused upstream and a fragment known to carry the RRE was fused downstream to the *gag*-encoding open reading frame. Adapting the *gag* gene to human codon usage resulted in constitutive nuclear export allowing high levels of Gag expression independently of the Rev/Rev-responsive element system because of the modification of intragenic regulatory elements [59]. Later, it was shown that it is possible to transfer this reporter system also to other report genes, like the green fluorescent protein (GFP). Adapting the open reading frame of GFP to HIV-1's codon usage was sufficient to turn this hivGFP RNA into a quasi-lentiviral message following the rules of late lentiviral gene expression. Again, cis-active elements which are known to influence HIV-1gene expression, like the 5'UTR as well as the RRE were added either upstream (5'UTR) or downstream (RRE) of the open reading frame [60]. This principle was used in several studies of our group and also in this thesis, in order to analyze the influence of partial codon adaptation on viral gene expression.

C.1.8 The role of the Tat protein in the transcription of viral genes by the cellular RNA polymerase II

The transactivator of transcription (Tat) plays a central role in the life cycle of HIV, since it mediates a more than hundred enhanced transcription of viral genes from the LTR promoter by binding to the TAR (transactivation response) element. The TAR element is a secondary structure which is built across the first 59 nucleotides and is present at the 5' part of all viral mRNA species. The binding of Tat is followed by the recruitment of the positive transcription elongation complex (P-TEFb), which consists of CDK9 and Cyclin T1. CDK9 then phosphorylates the carboxyl terminal domain of RNAP II. Transcription by RNAP II is a highly complex process from initiation to termination. Transcriptional initiation requires the assembly of the preinitiation complex (PIC), composed of the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH; the Mediator complex; and RNAP II with an unphosphorylated carboxy terminal domain (CTD). But, throughout the process of transcription, the CTD becomes highly phosphorylated which enhances the transcriptional activity [61]–[63]. This process is mediated by the Tat protein in the gene expression of HIV-1.

C.2 RNA export from the nucleus to the cytoplasm

As already mentioned above, the fully spliced transcripts are exported without the help of viral proteins like Rev. Those mRNAs are transported by the same mechanisms that are responsible for the export of cellular mRNAs. In general, for different RNA species that get exported out of the nucleus through the nuclear pore, different export pathways are used. Small RNAs, like tRNAs and micro RNAs get exported by relatively simple export pathways. Their transport is mediated by a direct binding to their respective export receptors. For example, the export of tRNAs is mediated by the export receptor exportin-t, a member of the karyopherin superfamily, which binds directly to tRNAs in a RanGTP-dependent manner. After transport of the tRNA-exportin-t-RanGTP complex to the cytoplasm, RanGAP stimulates GTP hydrolysis on Ran, inducing the release of the tRNA cargo from its receptor (see below) [64]–[66]. Large RNAs, like rRNAs and mRNAs build quite complex ribonucleoprotein (RNP) particles, including specific adaptor proteins to the given export factors [65].

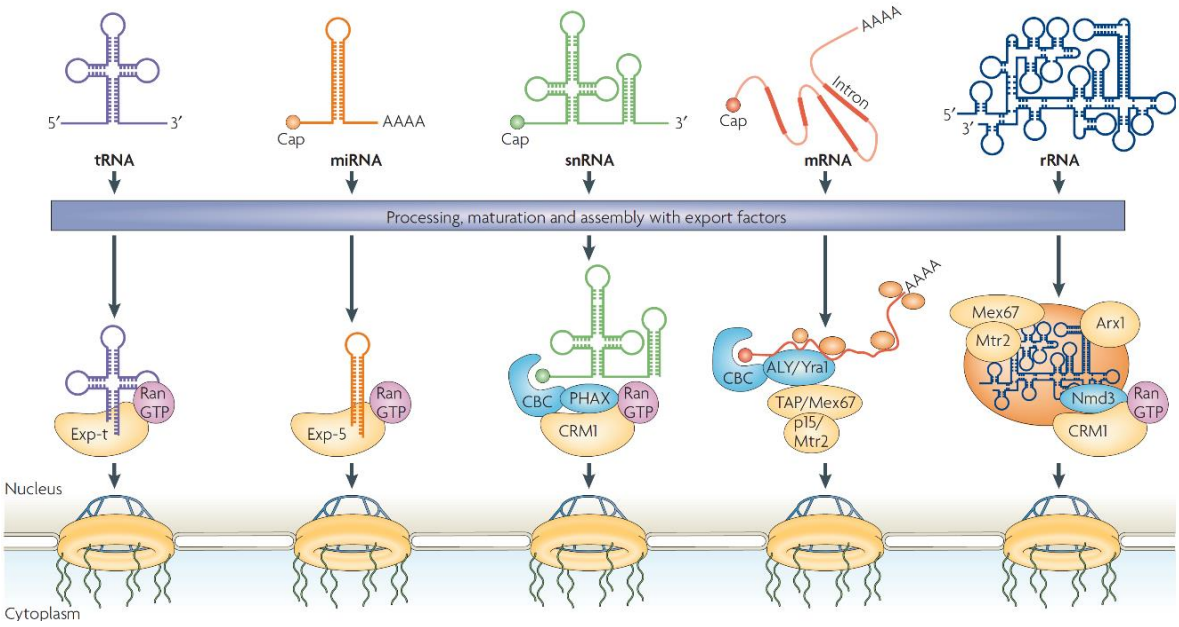

Figure C-8: Schematic illustration of different RNA export pathways (from [65]
The major routes of RNA export pathways are shown. Those include tRNAs, microRNAs (miRNAs), small nuclear (sn)RNAs, messenger RNAs (mRNAs) and ribosomal RNAs (rRNAs). The primary transcripts are shown at the top as well as the molecules after processing, maturation and assembly with export factors (export adaptors are shown in blue, export receptors are shown in yellow). For the mRNA pathway, metazoan and yeast proteins are listed and additional adaptor proteins and RNA binding factors (orange ovals) are depicted.

The precursor mRNA molecules undergo different processing steps like capping, splicing and cleavage/polyadenylation at the 3' end. For this, the mRNA associates with a variety of proteins to form a ribonucleoprotein particle (mRNP). Furthermore, an interaction with nuclear export factors also occurs already during transcription and processing, which ensures that only completely functional transcripts are provided to the translational machinery [67]. In principle the processes during mRNA export are as follows. Transcription of pre-mRNA molecules is carried out by RNA-polymerase II followed by an immediate assembly into RNPs, containing several heterogeneous nuclear ribonucleoproteins (hnRNPs) and splicing commitment factors. Those mRNAs remain inside the nucleus until the completion of splicing, when commitment factors are released and the nuclear mRNA export factors Tap and p15 are recruited to the mature mRNA. Subsequently, the RNP docks at the nuclear pore complex (NPC) and enters the cytoplasm. Nonshuttling hnRNPs have to be released at this point for example by the RNA helicase Dbp5. After the nuclear export, binding of cytoplasmic RNA binding protein occurs followed by translation of the mRNA. The shuttling hnRNPs like hnRNP A1 and nuclear mRNA export factors, including Tap and p15, are released and recycled to the nucleus [68].

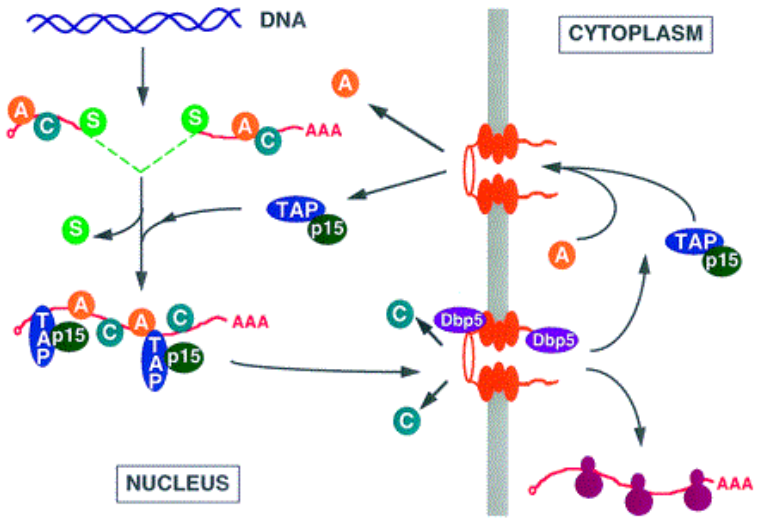

Figure C-9: Export of eukaryotic mRNA from the nucleus (from [61])
Processes and proteins involved in the nuclear export of mRNAs. Letters A-C stand for hnRNPs, S stands for splicing commitment factors. Explanation given in text.

C.3 The genetic code and its role in protein biosynthesis

The genetic code has a central place in biology, since it defines the rules to translate the 4-letter alphabet of nucleic acids into amino acid sequences, the 20-letter alphabet of proteins [69]. This process, known as protein biosynthesis, starts already in the nucleus with the transcription of DNA into a messenger RNA (mRNA). After the export of the mRNA into the cytoplasm (see above), the given nucleotide sequence gets translated according to the genetic code into an amino acid sequence, generating new proteins. These processes underlie complex and highly regulated mechanisms, themselves, like transcription initiation, elongation and termination, as well as post-transcriptional modification. This is also true for nuclear export processes or translational processes like initiation, elongation, transcription and post-translational modifications. Those mechanisms have been studied intensively over the years and were summarized in different excellent reviews

[70]–[74]. A simplified schematic illustration of the translation of a DNA sequence into an amino acid sequence is shown in Figure C-10.

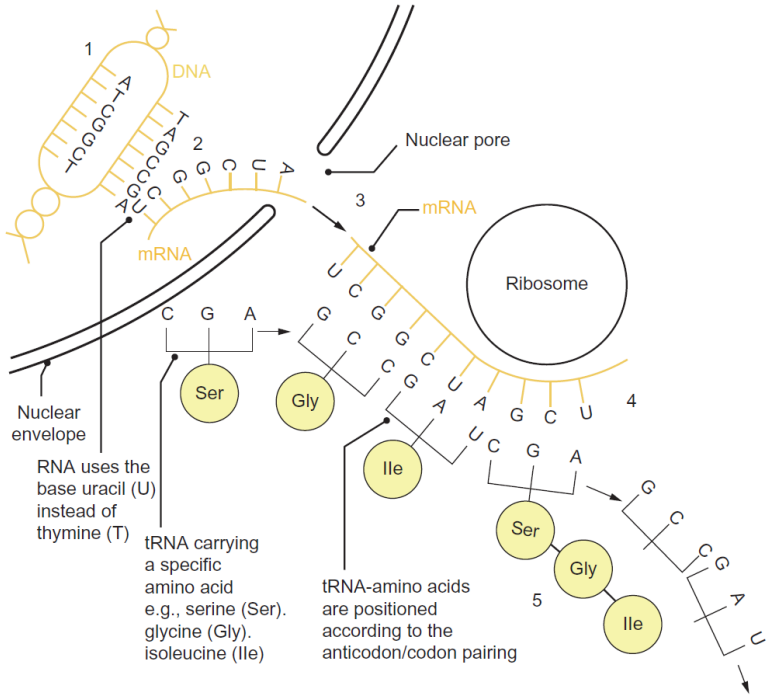

Figure C-10: Schematic and simplified overview of different processes involved in protein biosynthesis (from [69])
(1) Unwinding of the DNA double helix to allow transcription of the desired sequence into an mRNA molecule (2) which gets transported through the nuclear pore into the cytoplasm (3). After that, the mRNA attaches to the ribosome where transfer RNAs (tRNAs), each carrying a specific amino acid, interact with the mRNA trough the tRNA's anticodon (4). Elongation occurs and a growing peptide chain is generated (5) [69].

The genetic code is universal, which means that it is used by all creatures and plays a central role in the processes shown above, since it determines which amino acid sequence gets translated from the codons of the mRNA molecule. After the double helical structure of DNA was deciphered by Watson and Crick in 1953 [75] fundamental findings regarding the genetic code were done during the 1960s. In 1961 Crick and Brenner performed a historic experiment to demonstrate that the genetic code is made up of a series of three base pairs codons by using mutants in the *rII* locus of T4 phage [76]. Later, based on the finding that poly-uracil RNA incorporated only phenylalanine [77], Nirenberg was able to decipher the genetic code [78]. Since the genetic code acts as a fundamental feature all over in biology, it was and still is subject of intensive research that results in four key points, which should always be remembered about the genetic code. Firstly, the genetic code is non-overlapping. That means, consecutive amino acids are determined by consecutive codons. In an overlapping code, consecutive amino acids would be encoded by codons that share some bases. Secondly, an amino acid is always encoded by three bases, which are termed codons. Thirdly, the code is read from a defined starting point to an end point, which determines the coding region. And fourthly, the code is degenerated. That means some amino acids are encoded by more than one codon. This degeneracy

is mediated by a loose kind of base pairing at one end of the codon and anticodon, called wobble position [79]. So, some amino acids can be transported to the ribosome by several tRNAs with different anticodons, whereas for certain other amino acids this can only be done by one specific tRNA. Figure C-11 shows how 61 codons encode the 20 amino acids and three stop codons.

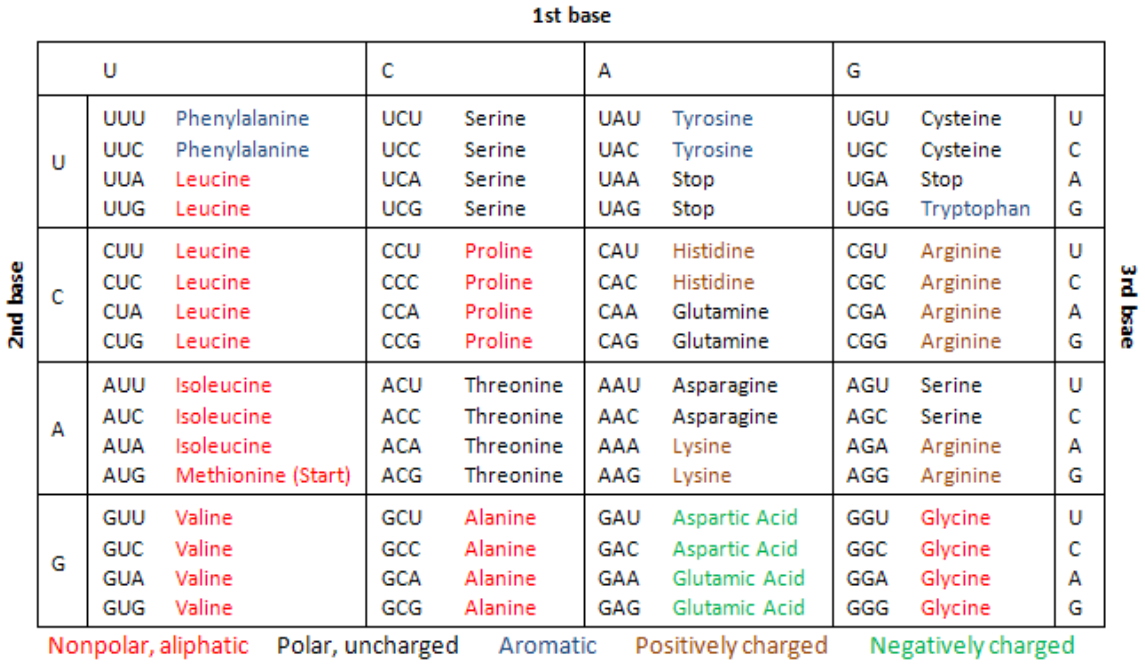

Figure C-11: The genetic code (from [80])
Shown are the 20 amino acids and their chemical features (indicated by different colors) and how they are encoded by the different codons.

The degeneracy of the genetic code is also the basic principle for codon bias and codon optimization, two points that play a central role in this thesis and will be explained in the following chapters.

C.4 Codon usage and codon usage bias

The degeneracy of the genetic code allows the usage of different codons for the same amino acid. Those codons are called synonymous codons and contrary to the first intuition those codons are not used in an equal frequency, but some codons are preferred. This phenomenon is known as codon usage bias and varies between and among species [81]. As explanations for the non-random usage of different codons a combination of mutation, selection and random drift is proposed [82]. In this model a balance between those forces favors specific, superior codons. That means that a mutation in an unpreferred codon could result in the generation of a preferred one when this allows a more efficient way of gene expression [83]–[86]. Furthermore, highly expressed genes show in general a stronger bias in synonymous codon usage [87]–[89]. As explanation for this phenomenon, one of the most intensively supported hypotheses over the years was translational selection. This theory is generally explained by better fitting codons for the most abundant isoaccepting tRNAs [90]–[92]. A

fact, which might be true for prokaryotes, where frequently used codons correlate with abundant cognate isoacceptor transfer RNAs [93].But especially for eukaryotes a variety of factors like GC-content, recombination rate, RNA stability, codon position and gene length can also influence codon usage bias [83], [94]–[98]. Therefore, translational efficiency by itself is not sufficient to account for effects of synonymous codon usage in higher organism. Especially the perception of the influence of synonymous codons on RNA level rose over the last years. Chen *et al.* quantified the impact of synonymous codons on mRNA level by analyzing over 3556 variants of a heterologous gene encoding the green fluorescent protein (GFP) and 523 synonymous variants of the endogenous gene TDH3 in yeast. They could show a positive correlation between mRNA levels and codon usage bias, which points to a direct effect of synonymous mutations on transcript concentration, most likely by influencing mRNA degradation rate [99]. Further studies point into the same direction. Kholiswa *et al* showed in 2008 that transfection of Jurkat cells with codon-optimized *gag* mRNAs led to a small increase in Gag production, whereas transfection of optimized DNA resulted in a very large enhancement of expression, indicating a higher ranking role of mRNA levels [100]. Furthermore, also our group could show that codon-usage-mediated inhibition of HIV-1 *gag* expression in mammalian cells occurs independently of translation [101].

C.5 Codon optimization and deoptimization

The fact that amino acids can be decoded in most cases by more than one codon facilitates the option to change codon choice as a molecular biological tool in form of codon optimization and deoptimization. Codon optimization and deoptimization describes the possibility to change the natural codon usage in two different directions. So, it is possible to introduce synonymous mutations in a coding region that do not affect the amino acid sequence of a protein but substitute specific nucleotides so that either better or worse fitting codons are generated. For this, a robust method had to be used to quantify codon usage bias. For this problem, different methods exist [102] and are favored by different groups. For the analysis in this thesis, the codon adaptation index (CAI) [103] was predominantly used, which is defined as the geometric mean of the relative adaptiveness values (w) of individual codons. The relative adaptiveness (w) of a codon is the ratio of the observed frequency of that codon to the frequency of the most abundant codon for the same amino acid. The most abundant codon is calculated for a reference set of highly expressed genes, such as ribosomal protein genes. Thus, $W = \frac{X_{ij}}{X_{imax}}$ where X_{ij} is the observed frequency of the i^{th} codon for the j^{th} amino acid and X_{imax} is the maximal X value for codons for the same amino acid [83]. The probably widest application for codon optimization is in heterologous gene expression, a well-established method to produce recombinant protein products like drugs, industrial enzymes or biofuels in foreign organisms more efficiently [104]–[108]. On the other hand, it is also possible to worsen the codon usage by using synonymous mutations. As an example, it was already shown that replacement of optimal codons in viral genomes can lead to attenuated viruses by influencing different features like

dinucleotide frequency, GC-content or codon pairs [109]–[111]. In this thesis an optimization of HIV-*gag* was performed in order to analyze the effects of altered codon usage on viral gene expression and the thereby mediated Rev-dependency/independency as already described in several earlier studies of our group [59], [60], [112]–[114]. Further The prerequisite to do this is the different nucleotide composition of HIV compared to its human host.

C.6 Nucleotide composition of HIV

Despite the necessity of HIV to use the host's translational machinery for viral gene expression, the virus exhibits an A-rich nucleotide composition with 36.2% A, 23.9% G, 22.2% U and 17.6% C, which differs clearly from the GC-rich coding regions of humans, where e.g. the GC content of isochores can be up to 60% [115]–[118]. The deviating codon usage is not exclusive for HIV, since also other viruses show an A-rich nucleotide composition or a differing codon usage compared to their hosts in general [119], [120]. From an evolutionary point that raises the question whether or how far the viruses benefit from different codon choice, because after all they rely on cellular processes like transcription or translation and therefore need to take over cellular functions and direct them towards the efficient production of new viruses [121]. The fast adaptation of HIV-1 epitopes to specific human MHC-I molecules validates the importance of viral evolution and adjustment to their hosts [122]. Therefore, there must be reasons why the deviating nucleotide composition of HIV-1 has evolved and seems to be maintained. In general, regarding this question, a model consisting of a combination of mutational activity and/or evolutionary selection is often favored [123]. Mutational activity could be caused either by the enzymatic properties of the error prone reverse transcriptase or by the cellular editing activities of the APOBEC enzyme (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) [124]. The first point can be explained by a biased guanine (G) to adenine (A) transition induced by an imbalanced dNTP pool during reverse transcription [125]. In addition it was shown that cytidine deamination by the restriction factor APOBEC3G/3F also contributes to the hypermutation [126]. APOBEC describes an evolutionary conserved class of enzymes catalyzing C-to-U editing by deamination of cytidine. In this reaction, a coordinated zinc ion in an enzyme active site acts as a Lewis acid to activate a water molecule for hydrolytic, nucleophilic attack of the amide group at the C4 position of cytidine and a conserved glutamic acid acts as a proton shuttle to convert a cytidine base to a uridine with an ammonium leaving group [127]. Despite all these facts, the question remains what effects an altered codon usage would have on viral gene expression and what molecular mechanisms are involved. To gain insight into this question was a central point of this thesis.

C.7 Objective

The question how an altered codon usage affects viral replication has been addressed in different studies for different viruses, but most of them focused on the influence of synonymous codons on viral replication [109]–[111], [128]. All of these studies provide important findings in the field of synonymous codon usage but rely on the complex interplay of several processes of viral replication but not gene expression per se. To address this direct influence of codon adaptation on gene expression, a simplified system of subgenomic *gag* reporter variants with varying codon-adapted(humanized) parts should be used to study the effects on gene expression independently from other biological processes of the replication cycle of HIV. For this, subgenomic *gag* reporter genes with systematically enlarged humanized parts under the control of a heterologous CMV promoter should be generated. In previous studies, our group already established this subgenomic

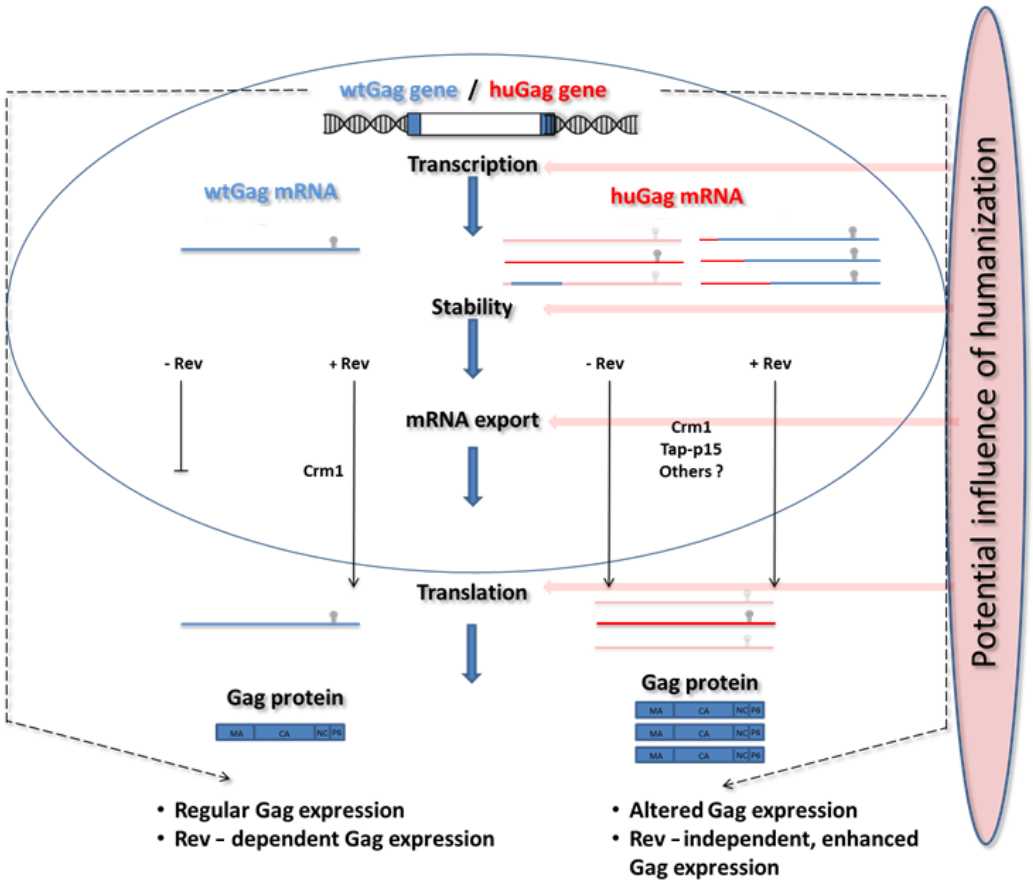

Figure C-12: Schematic illustration of the objective of this thesis:
The Impact of partial humanization (wtGag vs. huGag) should be investigated on different biological stages. Quantification of mRNA and protein levels should be performed as well as experiments analyzing the differences in transcription efficiency, mRNA stability, export pathways and effects on translational processes.

reporter system in order to analyze the effects of codon usage on HIV *gag* gene expression [59]. We could show that a complete humanization of HIV *gag* led to a Rev-independent and constitutive export of *gag* by elimination of cis-acting sequences [59]. Furthermore, humanization significantly

increased *gag* expression in dependency of the ratio of native and optimized codons [101]. To analyze positional and additive effects of codon optimization in a more detailed and systematic way, the subgenomic reporter system should be used in this work to analyze the effects of partially humanized variants. Therefore, the expression of the partially humanized variants should be analyzed on protein as well as mRNA levels, either by ELISA or by quantitative real time PCR. To get insight into the underlying molecular mechanisms, different molecular biological analyses regarding mRNA half-life, or transcriptional efficiency should be performed, as well as the analysis of potential alternative splicing effects. To address the question to what extent the results would be transferable to other genes, *egfp* should be used as an additional reporter gene. For this, a quasi-lentiviral GFP reporter system should be used, which our group established in a previous work [59], [60]. Taken together, the aim of this work should be to address the impact of humanization on the gene expression of HIV-1 in the context of various biological processes, as shown in Figure C-12.

D Material and methods

D.1 Cell lines

D.1.1 Prokaryotic cell lines

Table D-1: Overview of prokaryotic strains

Strain	Genotype
DH5α	F- supE44 ΔlacU169 (ϕ80 lacZΔM15) hsdR1 recA1 endA1 gyrA96 thi-1 relA1 [129]
DH10B	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara,leu)7697 araD139 galU galK nupG rpsL λ- [130]

D.1.2 Eukaryotic cell lines

Table D-2: Overview of eukaryotic cell lines used

Cell line	Description
HEK293T	Ad5-transformed human embryonic kidney cell line [131], expressing the “SV40 large T-antigen” [132]
HeLa	Human epithelial cell line (derived from cervix carcinoma)

D.2 Nucleic acids

D.2.1 Oligonucleotides

Table D-3: Overview of oligonucleotides used

Primer name	Sequence	Usage
Gag_c_wt_hu rev	CTGCAGTGTACTAGTAGTTCCTGCTATGTCACTTC	cloning of *gag* variants
Gag_c_wt_hu fwd	CAGGAACTACTAGTACACTGCAGGAACAGATCG	cloning of *gag* variants
Gag_d_hu_wt rev	GGCCCTGCAGTTCTTGGCAATGTGGC	cloning of *gag* variants
Gag_d_hu_wt fwd	CCAAGAACTGCAGGGCCCCTAGGAAAAAGGG	cloning of *gag* variants
Gag_c_wt_hiv rev	CTCTTGCAGTGTACTAGTAGTTCCTGCTATGTCACTTC	cloning of *gag* variants
Gag_d_hiv_wt fwd	GCAAAGAACTGTAGGGCCCCTAGGAAAAAGGG	cloning of *gag* variants
Gag_c_wt_hiv fwd	GGAACTACTAGTACACTGCAAGAGCAGATAGGAT	cloning of *gag* variants
Gag_d_hiv_wt rev	GGGCCCTACAGTTCTTTGCTATGTGCCC	cloning of *gag* variants
Gag_b_wt_hu rev	GCACCATTTGCCCCTGGAGGTTCTG	cloning of *gag* variants
Gag_b_wt_hu fwd	CCAGGGGCAAATGGTGCATCAGGCCATC	cloning of *gag* variants

Gag_b_wt_hiv fwd	CAGGGGCAAATGGTACACCAGGCAATATCAC	cloning of *gag* variants
Gag_f_hu_wt rev	GATCTTTACTGGCTGCTAGGATCGC	cloning of *gag* variants
Gag_f_hu_wt fwd	CCTAGCAGCCAGTAAAGATCTTCAGACCTGGAGGAG	cloning of *gag* variants
Gag_f_hiv_wt rev	GTCTGAAGATCTTTACTGCTTGGATCGCTTC	cloning of *gag* variants
Gag_f_hiv_wt fwd	CAAGCAGTCAGTAAAGATCTTCAGACCTGGAGGAG	cloning of *gag* variants
Gag_e_hu_wt rev	CCTAAAAAATTGGCCTGCCGCTCG	cloning of *gag* variants
Gag_e_hu_wt fwd	GAGCGGCAGGCCAATTTTTTAGGGAAGATCTGGCCTTCC	cloning of *gag* variants
Gag_e_hiv_wt rev	TCTGAAGAAAATTTGCCTGTCTCTCTGTGCA	cloning of *gag* variants
Gag_e_hiv_wt fwd	CACAGAGAGACAGGCAAATTTTTTAGGGAAGATCTGGCCTTCC	cloning of *gag* variants
Gag_a_wt_hu rev	GGCTCCCATCTCTCTCCTTCTAGCCTCCG	cloning of *gag* variants
Gag_a_wt_hiv rev	CTCTTGCTCCCATCTCTCTCCTTCTAGCCTCCG	cloning of *gag* variants
Gag_a_wt_hu fwd	GAAGGAGAGAGATGGGAGCCAGAGCCTC	cloning of *gag* variants
Gag_a_wt_hiv fwd	GGAGAGAGATGGGAGCAAGAGCAAGC	cloning of *gag* variants
Gag_b_wt_hiv rev	GGTGTACCATTTGCCCCTGGAGGTTCTG	cloning of *gag* variants
Gag_c2_hu_wt fwd	CCACCAGCACCCTTCAGGAACAAATAGGATGG	cloning of *gag* variants
Gag_c2_hu_wt rev	CTGAAGGGTGCTGGTGGTGCCGG	cloning of *gag* variants
ß-actin-Exon3-fwd	CACTGTGCCCATCTACGAGG	cloning of *gag* variants
ß-actin-Exon3-rev	CTCTTGCTCGAAGTCCAGGG	cloning of *gag* variants
Gag_d_wt_hu rev	CTGGGGGCTCTGCAATTTTTGGCTATGTGCCCTT	cloning of *gag* variants
Gag_d_wt_hu_fwd	CAAAAATTGCAGAGCCCCCAGAAAGAAAGG	cloning of *gag* variants
Gag_e_wt_hu_rev	CCAGGAAGTTAGCCTGTCTCTCAGTACAATCTTT	cloning of *gag* variants
Gag_e_wt_hu_fwd	AGAGACAGGCTAACTTCCTGGGCAAGATCTG	cloning of *gag* variants
TS - A Erw1 (A5-A6) Rev	TGCTGGGCCTTTTTCTTACTTTTGTTTTGCTCTTCCTCTATC	cloning of *gag* variants
TS - A Erw1 (A5-A6) For	AAAACAAAAGTAAGAAAAAGGCCCAGCAGGCTGC	cloning of *gag* variants
TS - A Erw2 (A4-A6) Rev	TGCACACAATAGAGGACTGCTATTGTATTATATAATGATC	cloning of *gag* variants
TS - A Erw2 (A4-A6) For	GCAGTCCTCTATTGTGTGCACCAGCGGATCG	cloning of *gag* variants
TS - A Erw3 (A3-A6) Rev	GCTGGCCCAGTATTTGTCTACAGCCTTCTGATGTC	cloning of *gag* variants
TS - A Erw3 (A3-A6) Fwd	GTAGACAAATACTGGGCCAGCTGCAG	cloning of *gag* variants
TS - A Erw4 (A2-A6) Rev	CTGGCCCACACGATATGTTTTAGTTTATATTGTTTCTTTCCCCCTGG	cloning of *gag* variants

TS - A Erw4 (A2-A6) Fwd	CTAAAACATATCGTGTGGGCCAGC	cloning of *gag* variants
TS - A Erw5 (A2-A6) Rev	GACAGCACCGACGCTCTCGCACC	cloning of *gag* variants
TS - A Erw5 (A2-A6) Fwd	GAGCGTCGGTGCTGTCTGGCGGC	cloning of *gag* variants
TS - A Erw6* (A1-A2) Rev	GCTTAATACAGAGGCTCTGGCTCCC	cloning of *gag* variants
TS - A Erw6* (A1-A2) Fwd	AGCCAGAGCCTCTGTATTAAGCGGGGGAGAATTAGATAAATGG	cloning of *gag* variants
TS-BsmBI-MK3-vorne_fwd (105)	GAGAGAGAGACGTGGGTGCGAGAGCG	cloning of *gag* variants
TS-BsmBI-MK3-vorne_rev (105)	CACCCACGTCTCTCTCTCTCCTTCTAGCCTCC	cloning of *gag* variants
TS-BsmBI-MK1_fwd (hinten_285)	GACAAATACCGTCTCTGGGCCAGCTGCAG	cloning of *gag* variants
TS-BsmBI-MK1_rev (hinten_285)	GCCCAGAGACGGTATTTGTCTACAGCCTTCTGATGT	cloning of *gag* variants
TS-BsmBI-MK1_fwd (vorne_225)	AGCTAGAGAGACGACGATTCGCAGTTAATCCTGGC	cloning of *gag* variants
TS-BsmBI-MK1_rev (vorne_225)	CGAATCGTCGTCTCTCTAGCTCCCTGCTTGCC	cloning of *gag* variants
TS-BsmBI-MK2_rev (hinten_225)	CTTTCCAGGAGACGGCTCTCTGCTGGCCCAC	cloning of *gag* variants
TS-BsmBI-MK2_fwd (hinten_225)	AGAGAGCCGTCTCCTGGAAAGATTCGCCGTGAAC	cloning of *gag* variants
TS-BsmBI-MK2_rev (vorne_165)	TGGCCGTCTCCTTAACCGAATTTTTTCCCATTTATCTAATTCTCCC	cloning of *gag* variants
TS-BsmBI-MK2_fwd (vorne_165)	CGGTTAAGGAGACGGCCAGGGGGAAAGAAACAATATAAACTAAAAC	cloning of *gag* variants
TS-BsmBI-MK3_rev (hinten_165)	CGCAGTGAGACGCTGATCTTCTCCCACTTGTCC	cloning of *gag* variants
TS-BsmBI-MK3_fwd (hinten_165)	ATCAGCGTCTCACTGCGGCCTGGC	cloning of *gag* variants
pCMV-LTR_NL4-3 fwd	GTACACGCGTTGGAAGGGCTAATTTGGTCCC	cloning of *gag* variants
pCMV-LTR_NL4-3 rev	GCCGAGTCCTGCGTC	cloning of *gag* variants
pCMV-LTR_NL4-3 fwd (PCR1b)	GACGCAGGACTCGGC	cloning of *gag* variants
TS-wt15-90_PCR1a_Rev	CTTAATACCGAGGCTCTGGCTCCCATC	cloning of *gag* variants
TS-wt15-90_PCR1b_Fwd	AGCCAGAGCCTCGGTATTAAGCGGGGGAG	cloning of *gag* variants

TS-wt30-90_PCR1a_Rev	CTAATTCTCCGCCAGACAGCACAGAGG	cloning of *gag* variants
TS-wt30-90_PCR1b_Fwd	GCTGTCTGGCGGAGAATTAGATAAATGGGAAAAAATTCGGTTAAGG	cloning of *gag* variants
TS-wt45-90_PCR1a_Rev	GAATTTTTTCCCACTTGTCCAGCTCGCCG	cloning of *gag* variants
TS-wt45-90_PCR1b_Fwd	GAGCTGGACAAGTGGGAAAAAATTCGGTTAAGGCCAG	cloning of *gag* variants
TS-wt60-90_PCR1a_Rev	CCCTGGCCTTAATCTGATCTTCTCCCACTTGTCC	cloning of *gag* variants
TS-wt60-90_PCR1b_Fwd	GAGAAGATCAGATTAAGGCCAGGGGGAAAGAAAC	cloning of *gag* variants
TS-wt75-90_PCR1a_Rev	GTTTCTTGCCGCCAGGCCG	cloning of *gag* variants
TS-wt75-90_PCR1b_Fwd	GCCTGGCGGCAAGAAACAATATAAACTGAAGCACATCGTG	cloning of *gag* variants
TS-wt30-75_PCR1a_Rev	CTAATTCTCCGCCAGACAGCACAGAGG	cloning of *gag* variants
TS-wt30-75_PCR1b_Fwd	GCTGTCTGGCGGAGAATTAGATAAATGGGAAAAAATTCGGTTAAGG	cloning of *gag* variants
TS-wt45-60_PCR1a_Rev	GAATTTTTTCCCACTTGTCCAGCTCGCCG	cloning of *gag* variants
TS-wt45-60_PCR1b_Fwd	GAGCTGGACAAGTGGGAAAAAATTCGGTTAAGGCCTG	cloning of *gag* variants
TS-wt15-30_PCR1a_Rev	CTTAATACCGAGGCTCTGGCTCCCATC	cloning of *gag* variants
TS-wt15-30_PCR1b_Fwd	AGCCAGAGCCTCGGTATTAAGCGGGGGC	cloning of *gag* variants
TS-wt30-45_PCR1a_Rev	CTAATTCTCCGCCAGACAGCACAGAGG	cloning of *gag* variants
TS-wt30-45_PCR1b_Fwd	GCTGTCTGGCGGAGAATTAGATAAATGGGAGAAGATCAGAC	cloning of *gag* variants
TS-wt60-75_PCR1a_Rev	CCCTGGCCTTAATCTGATCTTCTCCCACTTGTCC	cloning of *gag* variants
TS-wt60-75_PCR1b_Fwd	GAGAAGATCAGATTAAGGCCAGGGGGCAAG	cloning of *gag* variants
hu60-75_PCR1a_Rev	TTTTAGTTTATATTGTTTCTTGCCGCCAGGCC	cloning of *gag* variants
hu60-75_PCR1b_Fwd	CCTGGCGGCAAGAAACAATATAAACTAAAACATATAGTATGGGCAAGC	cloning of *gag* variants
hu60-90_PCR1a_Rev	CCATACTATATGTTTCAGCTTGTACTGCTTCTTGC	cloning of *gag* variants
hu60-90_PCR1b_Fwd	GCAGTACAAGCTGAAACATATAGTATGGGCAAGCAGG	cloning of *gag* variants
fwd_hu_GFP_FranziW	GACCTACGGCGTGCAATGCTTCAGCCGCTACCC	cloning of *egfp* variants

rev_hu_GFP_FranziW	CATTGCACGCCGTAGGTCAGGGTGGTCACGAGGGTG	cloning of *egfp* variants
fwd_hiv_GFP_FranziW	AGTAACAACATTAACATATGGAGTACAATGTTTTAGCAGATATC	cloning of *egfp* variants
rev_hiv_GFP_FranziW	CTCCATATGTTAATGTTGTTACTAATGTTGGCCAGG	cloning of *egfp* variants
huE_rev_FranziW	CACGGGTCCGTCACCTATTGGTGTATTTTG	cloning of *egfp* variants
huE_fwd_FranziW	AATAGGTGACGGACCCGTGCTGCTGCC	cloning of *egfp* variants
huA_fwd_FranziW	CCTGAAGTTCATCTGTACAACAGGAAAATTACCAGTACCC	cloning of *egfp* variants
huA_rev_FranziW	CCTGTTGTACAGATGAACTTCAGGGTCAGCTTG	cloning of *egfp* variants
huAB_rev_FranziW	TATTGTTCTCTCCTGGACGTAGCCTTC	cloning of *egfp* variants
huAB_fwd_FranziW	CTACGTCCAGGAGAGAACAATATTTTTTAAAGACGACGGAAATTA TAAAACAAGA	cloning of *egfp* variants
huABC_fwd_FranziW	CAAGCTGGAGTACAATTATAATAGCCATAATGTATATATAATGGC AGACAAAC	cloning of *egfp* variants
huABC_rev_FranziW	CATTATGGCTATTATAATTGTACTCCAGCTTGTGCCC	cloning of *egfp* variants
huABCD_rev_FranziW	CTGGTAATAATACTGGGCCGTCGCCGATGG	cloning of *egfp* variants
huABCD_fwd_FranziW	GACGGCCCAGTATTATTACCAGACAATCATTATTTAAGCAC	cloning of *egfp* variants
huBCDE_fwd_FranziW	AATTTATATGCACCACCGGCAAG	cloning of *egfp* variants
huBCDE_rev_FranziW	GGTGGTGCATATAAATTTTAATGTTAATTTTCCATATGTTGCGTCA CC	cloning of *egfp* variants
huCDE_rev_FranziW	GATGGTGCGTTCTTGTACATATCCTTCTGGCATTG	cloning of *egfp* variants
huCDE_fwd_FranziW	GGATATGTACAAGAACGCACCATCTTCTTCAAGGAC	cloning of *egfp* variants
huDE_fwd_FranziW	GGACATAAATTAGAATATAACTACAACAGCCACAACGTC	cloning of *egfp* variants
huDE_rev_FranziW	GTGGCTGTTGTAGTTATATTCTAATTTATGTCCTAATATATTTCCGT CCTCTTTAAAG	cloning of *egfp* variants
TS-ATG-wtA-hueGFP-PCR1a Rev	GCTCACCATTTGCCCCTGGAGGTTCTG	cloning of *egfp* variants
TS-ATG-wtA-hueGFP-PCR1b Fwd	CAGGGGCAAATGGTGAGCAAGGGCGA	cloning of *egfp* variants
ATG-huA-hueGFP-1a Rev	CACCATCTGGCCCTGCAGATTCTG	cloning of *egfp* variants
ATG-huA-hueGFP-1b Fwd	GCAGGGCCAGATGGTGAGCAAGGGCGA	cloning of *egfp* variants
ATT-huA-hueGFP-PCR1a Rev	TCTGGCTCCAATCTCTCTCCTTCTAGCCTCC	cloning of *egfp* variants
ATT-huA-hueGFP-PCR1b Fwd	GGAGAGAGATTGGAGCCAGAGCCTCTG	cloning of *egfp* variants
ATT-wtA-hueGFP-1a Rev	CGCACCAATCTCTCTCCTTCTAGCCTCC	cloning of *egfp* variants
ATT-wtA-hueGFP-1b Fwd	GAAGGAGAGAGATTGGTGCGAGAGCGTC	cloning of *egfp* variants

hu_Gag ohne ATG PCR1a Rev	GCTCTGGCTCCAATCTCTCTCCTTCTAGCCTCC	cloning of *egfp* variants
hu_Gag ohne ATG PCR1b Fwd	GAGAGAGATTGGAGCCAGAGCCTCTG	cloning of *egfp* variants
wt_Gag ohne ATG PCR1a Rev	CTCGCACCAATCTCTCTCCTTCTAGCCTCC	cloning of *egfp* variants
wt_Gag ohne ATG PCR1b Fwd	GGAGAGAGATTGGTGCGAGAGCGTC	cloning of *egfp* variants
Alec_wtA_eGFP_Fusion_FW	GGGGCAAGCCACCATGGTGAGCAAGGGC	cloning of *egfp* variants
Alec_wtA_eGFP_Fusion_RV	CATGGTGGCTTGCCCCTGGAGGTTCTG	cloning of *egfp* variants
Alec_huA_eGFP_Fus_FW	GGCCAGGCCACCATGGTGAGCAAGGGC	cloning of *egfp* variants
Alec_huA_eGFP_Fus_RV	CATGGTGGCCTGGCCCTGCAGATTCTG	cloning of *egfp* variants
TS-5'UTR-wt_C_eGFP_FWD	AGGGCCCCTATGGTGAGCAAGGGCGA	cloning of *egfp* variants
TS- Gag_A_ATT_1a Rev	TCTGGCTCCAATCTCTCTCCTTCTAGCCTCC	mutation of the *gag* start codon
TS- Gag_A_ATT_1b Fwd	GGAGAGAGATTGGAGCCAGAGCCTCTG	mutation of the *gag* start codon
TS-Gag_A_ATT_hueGFP 1a Rev	CTCACCATCTGGCCCTGCAGATTCTG	mutation of the *gag* start codon
TS-Gag_A_ATT_hueGFP 1b Rev	CAGGGCCAGATGGTGAGCAAGGGCGA	mutation of the *gag* start codon
TS-SD1_GT-AA (fwd)	CGACTGAAGAGTACGCCAAAAATTTTGACTAGCG	mutation of the splice donor SD1
TS-SD1_GT-AA (Rev)	CAAAATTTTTGGCGTACTCTTCAGTCGCCGCCCC	mutation of the splice donor SD1
TS_Rev_cloning_Fwd	CCAGGGTACCCTCGAAGCTAGT	cloning of Rev expression plasmid
TS_Rev_cloning_Rev	CTGCTCGAGCTGTGGCATTGAG	cloning of Rev expression plasmid
TS- NADH 1 (mito)_fwd	CCACATCTACCATCACCCTC	analysis nuclear and cytoplasmic RNA fractions
TS- NADH 1 (mito)_rev	CCTAGGAAGATTGTAGTGGTGAG	analysis nuclear and cytoplasmic RNA fractions
TS- U2 small nuclear 1_fwd	GCTAAGATCAAGTGTAGTATCTGTTC	analysis nuclear and cytoplasmic RNA fractions

TS- U2 small nuclear 1_rev	GCACCGTTCCTGGAGG	analysis nuclear and cytoplasmic RNA fractions
TS_Rev_seq_1_Fwd	CTACCCTGTCCACCCCTCTG	sequencing primer
TS_Rev_seq_2_Fwd	CGGAGCTGAATGAAGCCATAC	sequencing primer
TS_Rev_seq_3_Fwd	GAAAGGCGGACAGGTATCCG	sequencing primer
CMV-Seq	CGCAAATGGGCGGTAGGCGTG	sequencing primer
TS-LTR Seq.1 (fwd)	GCTGCTTCGCGATGTAC	sequencing primer
TS-LTR Seq.2 (fwd)	GACGCAGGACTCGGC	sequencing primer
TS-LTR Seq.3 (rev)	CAATTGTCCCTCATATCGCCTC	sequencing primer
TS-LTR Seq.4 (fwd)	GATCTTCAGACCTGGAGGAG	sequencing primer

D.2.2 Plasmids

Table D-4: Overview of plasmids used

Plasmid	Description
pc-UTR-wtGag16-30-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-wtGag31-45-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-wtGag46-60-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-wtGag61-75-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-wtGag76-90-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-wtGag16-90-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-wtGag31-90-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-wtGag46-90-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-wtGag61-90-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-wtGag31-75-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, partially humanized *gag* gene, from Alec Geßner (see [133])
pc-UTR-ATT-huA-huEGFP	eukaryotic expression plasmid with CMV-promoter, *gag* 5'UTR, RRE first 423bp of humanized *gag* gene without start codon and humanized *egfp* gene from Alec Geßner (see [133])
pc-UTR-ATT-wtA-huEGFP	eukaryotic expression plasmid with CMV-promoter, *gag* 5'UTR, RRE first 423bp of wild-type HIV-1 *gag* gene without start codon and humanized *egfp* gene from Alec Geßner (see [133])
pc-UTR-wtGag-RRE	eukaryotic expression plasmid with CMV-promoter, wild-type HIV-1 *gag* gene, from Nane Eiber (see [112])
pc-UTR-huGag-RRE	eukaryotic expression plasmid with CMV-promoter, wild-type HIV-1 *gag* gene, from Nane Eiber (see [112])
pc-UTR-huABC-RRE	eukaryotic expression plasmid with CMV-promoter, wild-type HIV-1 *gag* gene, from Nane Eiber (see [112])

pcDNA3.1	eukaryotic expression plasmid (Thermo Fisher, Waltham, USA - V79020)
pcDNA3.1-Rev	eukaryotic expression plasmid with HIV-1 *rev* gene (Wager Lab)
pcDNA3.1-Tat	eukaryotic expression plasmid with HIV-1 *tat* gene (Wagner Lab)

D.2.3 Antibodies

Table D-5:Overview of antibodies used

Antibody	Supplier/Specification
M01-antibody (anti-p24)	Polymun (AB006)
37G12-antibody (anti-p24), biotinylated	Polymun (AB005)
Anti-DIG Antibody (from DIG Northern Starter Kit)	Polyclonal sheep anti-digoxigenin, Fab-fragments, conjugated to alkaline phosphatase

D.2.3.1 Enzymes

Table D-6:

Primer	Sequence
Restriction endonucleases	New England Biolabs
Trypsin/EDTA	Pan Biotech
Phusion DNA-Polymerase	Finnzymes
Calf-Intestine-Phosphate (CIP)	Roche

D.2.4 Commercial Kits

Table D-7:

Kit	Supplier
Plasmid *Plus* Midi Kit	Roche
Plasmid *Plus* Maxi Kit	Roche
RNeasy Kit	Roche

Gel Extraction Kit	Roche
Quantinova Probe PCR Kit	Roche
Reverse Transcription Kit	Roche
Quick Ligation Kit	New England Biolabs
DIG Northern Starter Kit	Sigma-Aldrich

D.2.5 Standards

Table D-8: Overview of DNA and RNA standards used

Primer	Supplier
100 bp DNA ladder	New England Biolabs
1 kb DNA ladder	New England Biolabs
Transcript RNA Markers 0.2-10 kb	Sigma- Aldrich

D.2.6 Computer programs and databases

Table D-9: Overview of programs and databases used

Program/ Database	Specification/Supplier
Kazusa	http://www.kazusa.or.jp/codon/
Codon Usage Exe	Kindly provided from Benedikt Asbach; customized program that calculates CAI according to the formula by Sharp, 1987
Grapd Pad Prism	Used for figures and diagrams
Chromas	Used for sequence analysis
pDRAW 32	Used for in silico simulation of cloning
Ape Plasmid Editor	Used for sequence analysis and in simulation of cloning in silico
Step One Software	Used for evaluation of RT-qPCR results
Corel Draw	Used for figures

D.3 Cell culture techniques

D.3.1 Cultivation of eukaryotic cells

All cell culture techniques were performed under sterile conditions in a class II laminar flow hood. HEK293T cells were cultivated at 37°C and 5% CO_2. Before reaching full confluence, the cells were split 1:10. For this, cells were washed with 10 ml PBS and then incubated with 5 ml trypsin/EDTA for 5 minutes. The reaction was stopped by the addition of $DMEM_{10}$ Medium and cells were transferred into a suitable tube. Cells were then centrifuged at 300g for 5 minutes and the cell pellet was resuspended in 10 ml $DMEM_{10}$. 1 ml of the cell suspension was transferred into a flask with a suitable amount of $DMEM_{10}$. Cell concentrations were determined using trypan blue in a ratio of 1:1 for separation between living and dead cells during counting in a hemocytometer.

Table D-10: Media and reagents for cultivation of eukaryotic cells

Medium/Reagent	Supplier/ Ingredients	Additives
PBS (Phosphate buffered saline)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, adjusted to pH 7.4 with HCl	-
trypsin/EDTA	PAN Biotech	-
$DMEM_0$ (Dulbecco's Modified Eagle Medium)	Thermo Fisher	-
$DMEM_{10}$ (Dulbecco's Modified Eagle Medium)	Thermo Fisher	10% FCS, 1% Pen/Strep
FCS (fetal calf serum)	Sigma-Aldrich	-
Pen/Strep (Penicillin/Streptomycin)	PAN Biotech, 10000 U/ml Penicillin, 10 mg/ml Streptomycin	-
Trypan blue	0.4 %(w/v) trypan blue in PBS	-

D.3.2 Transfection of eukaryotic cells

HEK293T cells were transfected using calcium phosphate precipitation [134]. For this, a defined number of cells were seeded one day before transfection in a suitable dish. In general, $2x10^5$ cells were seeded in a 3.5 cm dish in a total volume of 3 ml $DMEM_{10}$ or $1x10^6$ cells were seeded in a 10 cm dish in a total volume of 4 ml $DMEM_{10}$. The desired amount of DNA was diluted in 135 µL H_2O. Then 15 µL 2.5 M $CaCl_2$ were added and the mixture was slowly added to 150 µL 2x HEBS buffer while constantly vortexing the solution. After an incubation period of 15 minutes the suspension was added to the cells. For transfection of $2x10^5$ cells 2 µg total DNA were used. For transfection of $1x10^6$ cells 10 µg total DNA were used.

Table D-11: Reagents for transfection of eukaryotic cells

Reagent	Supplier/ Ingredients
Calcium chloride	2.5 M in H_2O
2x HEBS buffer	50 mM HEPES (N-2-hydroxyethylpiperazine-N´-2-ethanesulfonic acid), 0.28 M NaCl, 1.5 mM Na_2HPO_4, pH 7.05

D.4 Microbiological methods

D.4.1 Growth and selection of prokaryotic cells

Bacterial cultures were cultivated overnight in an Erlenmeyer flask with LB medium at 37°C and 220 rpm. For selection of transformed cells 100 µg/ml of ampicillin was added.

D.4.2 Transformation of chemically competent cells

For transformation, 100 µl of competent bacterial cells were thawed on ice and mixed with 1 µg of plasmid DNA, followed by an incubation step of 25 minutes on ice. To permeabilize the bacterial cell membrane, cells were incubated for 45 seconds at 42°C followed by an incubation step of 2 minutes on ice. After that, 500 µl LB medium were added, followed by an additional incubation step of 60 minutes at 37°C. At the end, the transformed cells were plated on LB_{amp} agar plates.

Table D-12: Reagents for the transformation of chemically competent cells

Reagent	Supplier/ Ingredients	Additives
LB medium	0,5% (w/v) yeast extract, 1% (w/v) Trypton, 1% (w/v) NaCl, pH 7.4 in H_2O	
LB_{amp} medium	0,5% (w/v) yeast extract, 1% (w/v) Trypton, 1% (w/v) NaCl, pH 7.4 in H_2O	100 µg/ml ampicillin
LB_{amp} agar	LB_{amp} medium, 1 % /w/v) agar	

D.5 Molecular biology methods

D.5.1 Photometric quantification of nucleic acid concentrations

For photometric determination of nucleic acid concentrations as well as determination of purity, the Nanodrop photometer from Peqlab was used. For this, 1.5 µl per sample were loaded onto the photometer and measured against a reference sample (e.g. H_2O) in a spectral range between 220 and 300 nm. Using the Beer-Lambert law ($E= \varepsilon \cdot c \cdot d$), the concentration of the sample could be calculated from the adsorption at 260 nm, at which an adsorption of 1.0 represents an RNA

concentration of 40 ng/ml and a DNA concentration of 50 ng/ml. In the Beer-Lambert Law E is the measured absorbance, ε is a wavelength-dependent absorptivity coefficient, d is the path length, and d is the concentration of the analyte.

D.5.2 Isolation of plasmid DNA

For the isolation of plasmid DNA from bacterial cultures up to 5 ml, the standard method of alkaline lysis was performed [135], [136]. For the preparation of higher amounts of plasmid DNA the DNA extraction kits (QIAGEN Plasmid Plus Midi/Maxi) from Qiagen were used and isolation was performed according to the manufacturer's protocol. For elution H_2O was used.

D.5.3 Polymerase chain reaction

The polymerase chain reaction (PCR) [137] was used to amplify specific DNA fragments. For analytic purposes the *„GoTaq® Green master mix"* (Promega) was used. To calculate the optimal primer annealing temperature, the *T_M Calculator* Software (https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator) from NEB was used. The reaction protocol is shown in table Table D-13: PCR program for reactions with the GoTaq® Green master mix (left) and Phusion HF Polymerase (right). For amplification, 100 ng of DNA were used and mixed with water to reach a reaction volume of 8 µl which was mixed with 10 µl of the 2x *GoTaq® Green master mix* as well as 1 µl of forward and reverse primer. DNA fragments used for cloning were amplified with the *Phusion High fidelity* polymerase (Finnzymes), which has a proofreading function. For those reactions 10 ng of DNA were diluted in a volume of 67 µl of H_2O and mixed with 20 µl of 5x HF-buffer, 2 µl dNTP-mix (10 mM dNTP), 5 µl of forward and reverse primer each as well as 1 µl of Phusion polymerase. The reaction protocol is shown in table Table D-13.

Table D-13: PCR program for reactions with the GoTaq® Green master mix (left) and Phusion HF Polymerase (right)

TaqPolymerase	**Temperature**	**Time**	**Cycle**	**Phusion Polymerase**	**Temperature**	**Time**	**Cycle**
Initial Denaturation	98°C	2 min	1	Initial Denaturation	95°C	2 min	1
Denaturation	95°C	45 s	35	Denaturation	95°C	45 s	35
Annealing	T_{AN}	45 s		Annealing	T_{AN}	45 s	
Elongation	72°C	1 min/kb		Elongation	72°C	30 sec/kb	

Final Elongation	72°C	5 min	1	Final Elongation	72°C	5 min	1

Table D-14: Reagents used for polymerase chain reaction

Reagent	Supplier/ Ingredients
5x HF buffer	New England Biolabs
dNTP-Mix	New England Biolabs (10 mM ATP, 10 mM CTP, 10 mM GTP, 10 mM TTP)

D.5.4 Fusion PCR

The fusion PCR was mainly used to generate partially humanized *gag*-variants. For this, three independent Phusion PCR reactions (see D.5.3) with specifically designed primers were performed. The first reaction amplified the region of the CMV or LTR promoter, the 5' UTR up to the desired position of the transition between wild-type and humanized *gag* sequence (PCR1a). An overlapping sequence with the second PCR product was added during amplification. In the second reaction, the sequence from the desired *gag* position, the remaining *gag* ORF and the RRE were amplified (PCR1b), together with a complementary sequence to the first PCR product. In the third reaction, both variants were fused together (PCR2ab) and subsequently cloned into the pcDNA expression vector.

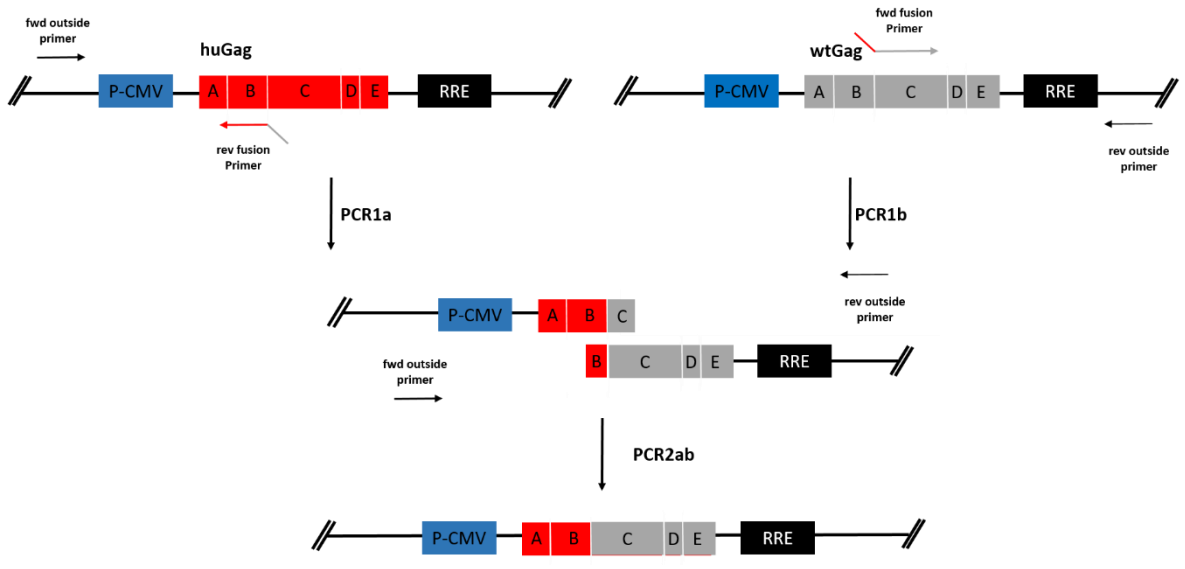

Figure D-1: Schematic process of a fusion PCR to generate partially humanized gag variants
Fusion PCR reactions consist of three independent PCR reactions, as described above (here the generation of huAB-wtCDE-Gag is shown as an example)

D.5.5 Standard cloning procedures

For cloning of genes, standard cloning techniques were used [135]. Vector backbones were digested with the desired restriction endonucleases, followed by dephosphorylation with a calf-intestine-phosphatase (NEB). Purification was performed after agarose gel electrophoresis (see D.5.7) using the QIAquick Gel Extraction Kit (Qiagen). Inserts were generated either by direct digestion of plasmids with specific endonucleases (NEB) or by PCR amplification followed by enzymatic digestion according to the manufacturer's protocol. Purification was also performed by using the QIAquick Gel Extraction Kit after analysis of the DNA by agarose gel electrophoresis. For ligation of inserts and vector backbones, the Quick Ligation Kit™ (NEB) was used according to the manufacturer's protocol. The complete ligation mix was transformed into chemically competent bacterial cells (see D.4.2)

D.5.6 Cloning with exocutter BsmBI

The variants used for the initial localization of the inhibitory motif were generated by the usage of BsmBI. For this, synthetically synthesized oligonucleotides representing the different 5' parts of *gag* were used. The oligonucleotides contained BsmBI cutting sites at their ends, which were used to insert the oligonucleotides in an appropriately prepared expression vector. Digestion and ligation were performed as described above (D.5.5).

D.5.7 Agarose gel electrophoresis

To separate DNA fragments according to their size, agarose gel electrophoresis was performed. For this, 1 % (w/v) agarose was heated in TBE buffer and mixed with 50 ng/ml ethidium bromide. PCR products amplified with the GoTaq® Green master mix (Promega) were loaded directly onto the gel. All other PCR products were mixed with 6x loading dye before loading. For estimation of the size of the different products a DNA ladder (1 kb or 100 bp ladder, NEB) was used. The separation of the single fragments was performed by running the gel at 80 - 200 V for 30 minutes. The visualization was performed by UV-light (302 nm) with the *Alpha Imager* (HP).

Table D-15: Reagents for agarose gel electrophoresis

Reagent	Supplier/ Ingredients
TBE buffer	21.6 g Tris, 11 g boric acid, 8 mL 0.5 M EDTA (pH 8.0) in 1 L H2O
Ethidium bromide	Roth, 10 mg/ml
6x loading dye	0.25 % bromophenol blue, 0.25 % xylencyanol, 30 % glycerol

D.5.8 Detection of potential cryptic splicing products

For the detection of potential cryptic splicing products, specific primers were designed that bind directly upstream of the major splice donor SD1 and downstream of the splice acceptor SA7. Total RNA was isolated and reverse transcribed as described above and subsequently analyzed by agarose gel electrophoresis (see D.5.7). Usage of one of the two splice sites would result in the generation of a shortened amplification product, which could be detected via agaroses gel electrophoresis.

Table D-16: Primer for the detection of potential cryptic splicing products

Primer	Sequence
TS-Splic-Inkl-SD/SA-Fwd	GAGGGGCGGCGAC
TS-Splic-Inkl-SD/SA-Rev	CTTCGGGCCTGTCGG

D.5.9 Generation of completely and partially codon adapted *gag*-variants

The completely humanized *gag* variant was designed by utilization of the KAZUSA codon usage database. For this, generally the most abundant codon for human and HIV codon usage was. In the case of similar results for human and HIV codon usage (e.g. serine), the procedure was slightly changed. For serine, the most abundant codon in human codon usage is AGC, which is also second most frequently used in HIV codon usage. Therefore, in this case the second most abundant codon in human codon usage was used. Methionine, arginine, and tryptophan were not replaced, since there is either only one possible codon, or the order of the codon is usage is identical in both organisms. Subsequently the sequence was checked for detrimental sequences like unwanted splice sites, inverted repeats or RNA-instability motifs with the GeneOptimizer algorithm [138]. The *gag* sequence then was synthetically produced by Geneart and cloned into a pcDNA3.1 expression vector. All partially humanized *gag* variants were generated either by fusion PCR or BsmBI cloning (D.5.4, D.5.6).

D.6 Working with RNA

To minimize the risk for contamination with RNases, all working steps regarding RNA were performed with filter tips and all working places were decontaminated with RNase AWAY (Thermo Scientific) and 1 % SDS (Merck). As far as possible, all solutions and buffers were treated with 1 % DEPC (Sigma Aldrich) and autoclaved after 24 hours of incubation.

D.6.1 Isolation of total RNA

For isolation of total RNA, a defined amount of HEK293T cells (normally $1x10^6$ cells) were seeded in suitable dishes (e.g. 3.5 cm dish / 6 well plate) and transfected with the desired *gag*-variants (see D.3.2). When cells were confluent, cells were washed with PBS and detached by trypsin/EDTA and centrifuged (5 min, 300 g). The cell pellet was resuspended in 1 ml of PBS and transferred into a suitable RNase-free 1,5 ml tube. Cells were partially lysed by the addition of 175 µl RLN buffer and subsequent incubation on ice for 5 minutes. The cytoplasmic fraction was separated from the nuclear fraction by centrifugation (300 g, 2 min, 4°C). Nuclei were then completely lysed by the addition of 350 µl RLT buffer. From that point, RNA was isolated using the RNeasy Kit (Qiagen) according to the manufacturer's protocol. RNA was solved in RNase-free H_2O and stored at -80°C.

Table D-17: Required buffers for the isolation of total RNA

Buffer	Ingredients/Supllier
PBS (Phosphate-buffered saline)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, adjusted to pH 7.4 with HCl
RLN buffer	50 mM Tris-HCl pH 8.0, 140 mM NaCl, 1.5 mM $MgCl_2$, 0.5% Nonidet P-40, 1000 U/ml Rnase Inhibitor
RLT buffer	Qiagen (RNeasy Kit), addition of 10 µl β-mercaptoethanol per 1 ml buffer

D.6.2 Nuclear Run on assay

For determination of transcription efficiency, nuclear run on assays were performed. For this, HEK293T cells were transfected with the desired variants. After 48 h, cells were harvested and washed two times with ice cold PBS. The cell pellet was resuspended in a Nonident P-40 containing lysis buffer and incubated on ice for 5 minutes. Nuclei were then harvested by centrifugation (200 g, 2 min, 4°C) and washed with lysis buffer without NP-40. Nuclei were centrifuged once again (200 g, 2 min, 4°C), resuspended in 100 µl glycerol buffer and mixed with 200 µl transcription buffer as well as 8 µl of biotinylated UTPs (10 mM). After an incubation time of 30 min at 29°C, 6 µl of 250 mM $CaCl_2$ and 6 µl of RNase-free DNase I (10 U/µl; Roche) were added to stop the reaction. Total RNA was isolated using the RNeasy-Kit (Qiagen) according to the manufacturer's protocol and RNA was eluted in 35µl RNase-free H_2O. RNA was mixed with 35 µl M-280 Streptavidinin dynabeads (Invitrogen) binding buffer and incubated at 42°C for 20 minutes and 2 hours at room temperature. After the incubation, dynabeads were isolated using a magnet and washed two times with 500 µl 15% Formamid and 2x SSC. After an additional washing step with 1 ml 2x SSC buffer, the dynabeads

were diluted in 35 µl of RNase-free H_2O.After quantification of the concentration, 200 ng of RNA were used for the generation of cDNA (D.6.4) and quantification by real-time qPCR (D.6.5).

Table D-18: Required reagents and buffers for the execution of nuclear run on assays

Reagent/Buffer	Supplier/ Ingredients
PBS (Phosphate-buffered saline)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, adjusted to pH 7.4 with HCl
Lysis buffer	10mM Tris/HCl pH 7.4, 3mM $MgCl_2$, 10mM NaCl, 0.5% NP-40
Glycerol buffer	50mM Tris/HCl pH 8.3, 5mM $MgCl_2$, 0.1mM EDTA, 40% glycerol
Biotin-16-dUTP	Roche
Binding buffer	10mM Tris/HCl pH 7.5, 1mM EDTA, 2M NaCl
DNase I	Sigma Aldrich
Formamide	Sigma Aldrich
Rneasy Kit	Qiagen

D.6.3 Determination of mRNA half-life

For mRNA half-life determination, $1x10^6$ HEK293T cells were transfected with the desired variants (D.3.2) in duplicates and incubated for 48 hours. Half of the cells were harvested using trypsin and centrifugation (5 min, 300 g) after 48 hours and cell pellets were stored until RNA isolation. To the other half, actinomycin D was added in a final concentration of 2.4µM to block the transcription by intercalating in the double strand of the DNA which leads to inhibition of RNA polymerase II. Cells were incubated with actinomycin D for 24 hours, then were harvested using trypsin and centrifugation (5min, 300g). Thus, for each transfected *gag*-variant an actinomycin D treated and untreated sample could be analyzed. RNA was isolated (D.6.1) and quantification was performed after generation of cDNA as described in D.6.4. Logarithmic plotting of the transcript amounts against the time points led to a linear equation of the form

$$\frac{N(t)}{N(O)} = e^{-kt} \rightarrow lnN(t) = -kt - lnN(0),$$

with decay constant k representing the slope. Half-life of the mRNA can be calculated by the formula

$t_{1/2}$ = -ln2/k.

Table D-19: Required reagents and buffers for the determination of mRNA half-life

Reagent/Buffer	Supplier/ Ingredients
trypsin/EDTA	PAN Biotech
PBS (Phosphate-buffered saline)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, adjusted to pH 7.4 with HCl
Actinomycin D	Sigma Aldrich

D.6.4 Reverse transcription (synthesis of copy DNA)

Synthesis of copy DNA (cDNA) was performed using the reverse transcription Kit from Qiagen. For this, RNA was photometrically quantified (D.5.1). Per reaction, 200 ng of RNA were used while the reaction was performed according to the manufacturer's protocol. The process included also a gDNA removal step which has always been performed.

D.6.5 Quantification of mRNA expression levels by quantitative real-time PCR (RT-qPCR)

For quantification of *gag* expression on mRNA levels, HEK293T cells were transfected as described (D.3.2). After 48 hours, total RNA was isolated using the RNeasy Kit (Qiagen) and cDNA was generated using the reverse transcription kit (Qiagen) as described in D.6.1 and D.6.4 which was then diluted by the factor of 10. For analysis, 1µl of the diluted cDNA was used and quantification was performed using the Bright Green 2x qPCR Mastermix (ABM) or the Quantinova probe Kit (Roche) according to the manufacturer's protocol. For quantification of *gag* expression, RRE-specific primers were designed that bind to the RRE, which was present in all constructs. If the Quantinova probe Kit (Roche) was used, an additionally labeled probe was used that lies between the forward and reverse primer inside the RRE. Here, the detection principle relies on the 5´–3´ exonuclease activity of Taq polymerase. This leads to cleavage of a dual-labeled probe during hybridization to the complementary target sequence and fluorophore-based detection. For the Bright Green 2x qPCR Mastermix (ABM), only forward and reverse primers were used, since quantification is based on the unspecific incorporation of the Bright Green dye in dsDNA molecules. Therefore, a melting point analysis was performed, when using the Bright Green 2x qPCR Mastermix (ABM). For all analyses, the StepOnePlus real-time PCR system (Applied Biosystems) was used. Besides Quantification of *gag* expression, Neomycin expression was analyzed for normalization. For this, specific primers were designed that bind inside the neomycin resistance gene, which was present on all transfected *gag* plasmids. For quantification, a standard curve was generated. For this, Ct values were determined using serial dilutions of a linearized plasmid with calculated total copy numbers (from $1x10^2$ to $1x10^8$). PCR efficiencies (E) were determined for each reaction. E can be calculated from the slope of the

standard curve: $E = 10^{-1/slope}$. Primers were designed such that the E was approximately 2. The C_t value is defined as the cycle number when fluorescence exceeds the background level for the first time.

Table D-20: Required kits and mastermixes for quantification of mRNA expression levels by quantitative real-time PCR

Kits/Mastermix	Supplier
Bright Green 2x qPCR Mastermix-ROX	ABM
QuantiNova Probe PCR-kit	Qiagen

Table D-21: Primers and probes for quantification of mRNA expression

Primer	Sequence
qNeo fwd	GCTATCAGGACATAGCGTTGG
qNeo rev	GAAGGCGATAGAAGGCGATG
TS-qPCR-RRE-fwd	GTTCCTTGGGTTCTTGGGAG
TS-qPCR-RRE-rev	GCCCTCAGCAAATTGTTCTG
TS-qPCR-RRE-probe	ACGTGCAGCCCATAGTGCTTCCT

D.6.5.1 Northern blot analysis

D.6.5.2 Generation of specific RNA probes for northern blot analysis

For the detection of potentially generated cryptic splicing sites, northern blot analysis was performed. For this, specific antisense RNA probes had to be generated. RNA probes were used because of the enhanced stability of RNA-RNA hybrids in comparison to RNA-DNA hybrids and therefore higher detection sensitivity. Those RNA probes were generated using the DIG northern starter kit from Roche. For this a T7- promoter sequence was introduced via the reverse Primer in a PCR reaction, to allow the initiation of the T7 polymerase. During this step, which was performed with 200ng of the PCR product, Digoxin-labeled uracil was incorporated into the generated RNA probe, allowing the following detection. The reaction protocol is shown in Table D-22 and can be seen I detail in the manufacture's instruction of the DIG northern starter kit

Table D-22. Process of generating specific RNA probes for northern blot analysis using the DIG Northern Starter Kit

42°C	1 hour
4°C	Add 2 µl DNase
37°C	15 minutes
4°C	Add 1 µl RNase inhibitor + 1 µl 500mM EDTA

D.6.5.3 Sample preparation

For the analysis, 2 µg of all RNA samples in a volume of 5µl were mixed with 2 volumes of RNA sample buffer as well as 5 volumes of RNA loading buffer and incubated together with the RNA ladder at 65°C for 5 min, followed by immediate transfer on ice.

D.6.5.4 Gel electrophoresis

RNA was then separated according to its size by agarose gel electrophoresis. For this purpose, all materials were cleaned with 1% SDS as well as 3% H_2O_2. The gel was produced by mixing 22.5 ml 10x MOPS buffer with 172.5 ml DEPC treated H_2O and 2.25 g agarose and subsequent boiling using a microwave. As soon as the solution was cooled to 55°C, 30 ml of formaldehyde (37%) were added. As running buffer 1x MOPS buffer was used. The gel run was performed for 2.5 hours at 100V.

D.6.5.5 Blotting of the RNA samples onto a nylon membrane

Prior to blotting, the gel was washed in three different conditions with shaking at 40 rpm according to Table D-23**.**

Table D-23: Washing steps as preparation of blotting RNA samples to a nylon membrane

Washing steps	Time
DEPC - H_2O (300 ml)	3 x 20 min
0.05 M NaOH (250 ml)	20 min
20x SSC Buffer (250 ml)	45 min

To transfer the RNA samples to a positively charged nylon membrane (Sigma), it was positioned on top of the gel. Below the gel, three pieces of Whatman paper were arranged, with the edges of the lowest one dipped into 20x SSC blotting buffer. Two pieces of Whatman paper and one layer of parafilm were placed on top and on the edge of the membrane. Above, pulp was stacked to a height of approximately 15 cm and 2 kg weight was added to ensure overnight blotting (16h) via capillary force.

D.6.5.6 Fixation of RNA

After blotting, the membrane was washed in 5x SSC buffer for 15 minutes and dried. Fixation of the transferred RNA samples was achieved using a UV cross-linking device (Stratalinker 1800) running the programme “Auto Cross Link (1200 kJ)”.

D.6.5.7 Methylene Blue staining

To control the success of blotting, the membrane was stained with 0.03% methylene blue in 0.3M sodium acetate until the ribosomal RNA bands were visible. The membrane was washed in DEPC-H_2O, the rRNA and marker bands highlighted with a pen and a picture was taken with a gel documentation device.

D.6.5.8 Hybridisation

Each hybridisation solution consisted of 7.4 g DIG Easy Hyb Granules from the Northern Starter Kit solubilised in 15 ml DEPC -H_2O. The washed membrane was transferred to a hybridization bottle and pre-incubated with 10 ml of pre-heated hybridisation solution at 68°C for 30 minutes in an oven. For actual hybridisation, the RNA probes were denatured at 99°C for 5 minutes, and 10 µl of the probe were added to the remaining 5 ml of hybridisation solution. The solution was added to each membrane and incubated at 68°C for 16 hours.

D.6.5.9 Washing

Washing of membranes after the hybridisation consisted of two washing steps with pre-heated solutions:

- non-stringent washing: 2 x 5 minutes at 68°C in 2x SSC Buffer + 0.1% SDS (50 ml each)
- stringent washing: 4 x 15 minutes at 68°C in 0.2x SSC Buffer + 0.1% SDS (50 ml each).

D.6.5.10 Detection

The membranes were washed in maleic acid buffer + 0.3% Tween-20 and blocked with 100 ml 1x Blocking Buffer from the Northern Starter Kit for one hour. 5 µl of an anti-DIG antibody with an attached alkaline phosphatase from the kit were diluted in 50 ml of the blocking buffer and the membranes were incubated for 30 minutes with shaking. After two washing steps with maleic acid buffer + Tween-20 for 15 minutes, the membranes were equilibrated in detection buffer for 5 minutes. The detection was performed as described in the manufacturer's protocol, using an. The visualization was performed on the INTAS device (ChemiLuxPro).

Table D-24: Buffer and reagents required for northern blot analysis

Buffer/ Reagent	Supplier/ Ingredients	Additives
10x MOPS buffer (pH 7.0)	400 mM MOPS, 60.5 mM $C_2H_3NaO_2$, 500 mM EDTA pH8.0,	
20x SSC buffer (pH 7.0)	0.3 M $Na_3C_6H_5O_7$, 3 M NaCl,	
RNA sample buffer	65% CH_3NO, 8.5% CH_2O, 6.45% 10x MOPS buffer, 6.45% DEPC-H_2O	
5x RNA loading buffer	20% glycerol (87%), 8mM EDTA (500mM, pH 8.0), 0.4% bromophenol blue, 30.8% CH_3NO, 2.7% CH_2O (37%), 40% 10x MOPS buffer, 1%	
Maleic acid buffer	0.1 M maleic acid, 0.15 M NaCl	
Washing buffer	Maleic acid buffer	0.3% Tween 20
Detection buffer	0.1 M Tris/HCl, 0.1 M NaCl	
Hydrogen peroxide (H_2O_2)	Sigma Aldrich	
Methylene blue	Sigma Aldrich	

Table D-25: Primers required for generation of a complementary RRE probe

Primer	Sequence
TS-RRE-Sonde-NB fwd	GATCTTCAGACCTGGAGGAG
TS-RRE-Sonde mit T7-NB	TAATACGACTCACTATAGGGCTCAGCAAATTGTTCTGCTGC

D.7 Protein biochemistry techniques

D.7.1 Lysis of HEK293T cells

For analysis of intracellular Gag expression, cells were lysed after transfection. For this, cell pellets were washed with Phosphate Buffered Saline (PBS), centrifuged again, resuspended in 150 µl pre-cooled TDLB (RIPA) Buffer and incubated on ice for 20 minutes. Samples were then sonified for 5 minutes using a Bioruptor device (Diagenode) at energy level H. Cellular debris was removed by centrifugation at 20,000g and 4°C for 5 minutes. All cleared lysates were transferred into a fresh Eppendorf tube and total protein amount was quantified by Bradford analysis (D.7.2)

Table D-26: Buffers required for lysis of HEK293T cells

Buffer/Reagent	Supplier/Ingredients
PBS (Phosphate buffered saline)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, adjusted to pH 7.4 with HCl
TDLB (RIPA) buffer	Tris 50 mM, NaCl 150 mM, SDS 0.1 %, Na-desoxycholate 0.5 %, TritonX-100 1 %, pH 8.0

D.7.2 Quantification of total protein amount by Bradford analysis

For quantification of total protein amounts in cellular lysates, a Bradford assay was performed. 2 µl of 1:4 diluted sample lysates were added to 158 µl PBS in triplicates on a 96-well plate (Sarstedt). For generation of a standard curve, defined amounts of bovine serum albumin (BSA) from 0 to 5 µg were used. 40 µl of the Bradford reagent were then added to each well and it was thoroughly mixed. Absorbance at 595nm was detected with an MP Reader 680 photometer (BioRad) and the amount of total protein in each sample was calculated based on the standard curve.

Table D-27:Buffer and reagents required for quantification of total protein amount by Bradford analysis

Buffer/Reagent	Supplier/Ingredients
PBS (Phosphate-buffered saline)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, adjusted to pH 7.4 with HCl
Bradford Reagent	BioRad
Bovine serum albumin	Sigma Aldrich

D.7.3 Quantification of Gag expression by p24-Enzyme-linked Immunosorbent Assay (ELISA)

As a measure of Gag expression, the capsid protein (p24) was quantified by an Enzyme-linked Immunosorbent Assay (ELISA). For this, HEK293T cells were transfected and harvested as described (D.3.2). After incubation of 48 hours either 2 µg of total protein from cell lysates or direct supernatant samples were analyzed as follows:

- Dilutions for supernatants: 1 - 1:10 - 1:100 – 1:1000 in $DMEM_{10}$

- Dilutions for lysates: 1 - 1:10 - 1:100 (2 µg total protein in undiluted sample) in $DMEM_{10}$

For this, a Nunc Maxisorp 96-well plate (Thermo Scientific) was coated with 0.25 µg of primary anti-p-24 antibody (M-01, Polymun) per well in 100 µl of coating buffer the day before analysis and incubated overnight at 4°C. All washing steps have been performed with the tecan hydroflex ELISA washer. Before loading of the dilutions on the coated plate, wells were washed with the tecan hydroflex program p24_3x_A, followed by an incubation time of 1 hour at 37°C. After that, wells were washed with the tecan hydroflex program p24_6x_A and the biotinylated secondary anti-p24-antibody (37G12, Polymun) was added in a concentration of 1:10000 in blocking buffer and incubated for 1 hour at room temperature. Next, wells were washed with tecan hydroflex program p24_10x_A and streptavidin-HRP-conjugate (Roche) was added in a concentration of 1:10000 in blocking buffer, followed by an additional incubation of 30 minutes at room temperature. The streptavidin-HRP-conjugate was washed away by an additional washing step (with tecan hydroflex program p24_10x_A). For detection, a TMB substrate was used. 100 µl of a 20:1 mix (TMB A: TMB B) were added per well and incubated for 5 minutes. The reaction was stopped with 50 µl of 1M H_2SO_4 per well, and the absorbance was measured with a MP Reader 680 photometer (BioRad) at 450 nm in triplicates. Gag amounts in each sample were calculated based on the p24 standard curve. For creation of a standard curve, recombinant p24 was used. Concentration of the standard started at 2.5 ng/ml and was diluted 5 times by the factor of two, to reach a minimal concentration of 0,078 ng/ml.

Table D-28: Buffer, reagents and antibodies required for quantification of Gag expression by p24 ELISA

Buffer/ Reagent/Antibody	Supplier/ Ingredients
M01-antibody (anti-p24)	Polymun (AB006)
37G12- antibody (anti-p24), biotinylated	Polymun (AB005)
Streptavidin-HRP conjugate (500 U/ml)	Roche
Coating buffer	100 mM sodium carbonate, pH 9.5
Blocking buffer	PBS mit 5 % (w/v) BSA
PBS-T	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, adjusted to pH 7.4 with HCl, Tween
TMB-A	30 mM potassium citrate, pH 4.2
TMB-B	10 mM TMB, 10 % (v/v) Aceton, 90 % (v/v) Ethanol, 80 mM H_2O_2

D.8 Analysis of GFP expression by flow cytometry

For analysis of EGFP expression, HEK293T cells were transfected as described (D.3.2). After 48 hours of incubation, medium of cells was removed and cells were washed with 1ml of PBS and harvested using trypsin-EDTA. Cells were pelleted by centrifugation (500g, 3min) and washed again with 1 ml of PBS. Then, cells were centrifuged again (500g, 5min) and the pellet was resuspended in 1 ml of FACS buffer. Analysis of GFP expression was then performed on the FACS Canto II device (BD Biosciences) using a blue laser for excitation of GFP at 488 nm and the emission of EGFP by the FITC filter (fluorescein isothiocyanate-filter, 530/30). For all analyses, 30000 events were recorded.

Table D-29: Buffers and reagents required for analysis of GFP expression by flow cytometry

Buffer/Reagent	Supplier/Ingredients
PBS (Phosphate-buffered saline)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, adjusted to pH 7.4 with HCl
Trypsin/EDTA	PAN Biotech
FACS buffer	1% FCS in PBS

D.9 DNA Sequencing

All sequence analyses were performed either by Seqlab or by GATC. For this, the desired amount of DNA was mixed with the specific sequencing primer in H_2O according to the instructions of the respective company. Results were evaluated with the programs Chromas and Ape.

E Results

E.1 Generation of partially humanized *gag* variants

Several studies already analyzed the influence of codon adaptation on viral gene expression and consequently viral replication [109]–[111], [128]. All of them underlined the importance of codon usage and demonstrated the complex interplay between different biological sequence features and processes. In general, viral replication has been used in those studies as a measure of the impact of codon adaptation. In this thesis, a systematical approach was chosen to analyze the importance of codon usage directly on gene expression. Other biological processes that are involved in viral replication, like viral entry, reverse transcription, budding or release of new virions should be excluded by using subgenomic reporter constructs (see C.1.7). The Gag protein from HIV-1 was chosen as viral model protein for these analyses because of its importance and the availability of data from different preliminary studies [59], [112], [114], [139]. It was already shown that the humanization of *gag* leads to an enhanced and Rev-independent protein production. In order to find out whether there are positional or additive effects of codon adaptation, the *gag*-gene was divided into five parts, which were gradually adapted to human codon usage, a process here referred to as "humanization". The borders of the five parts, named with the letters A-E, were chosen because of

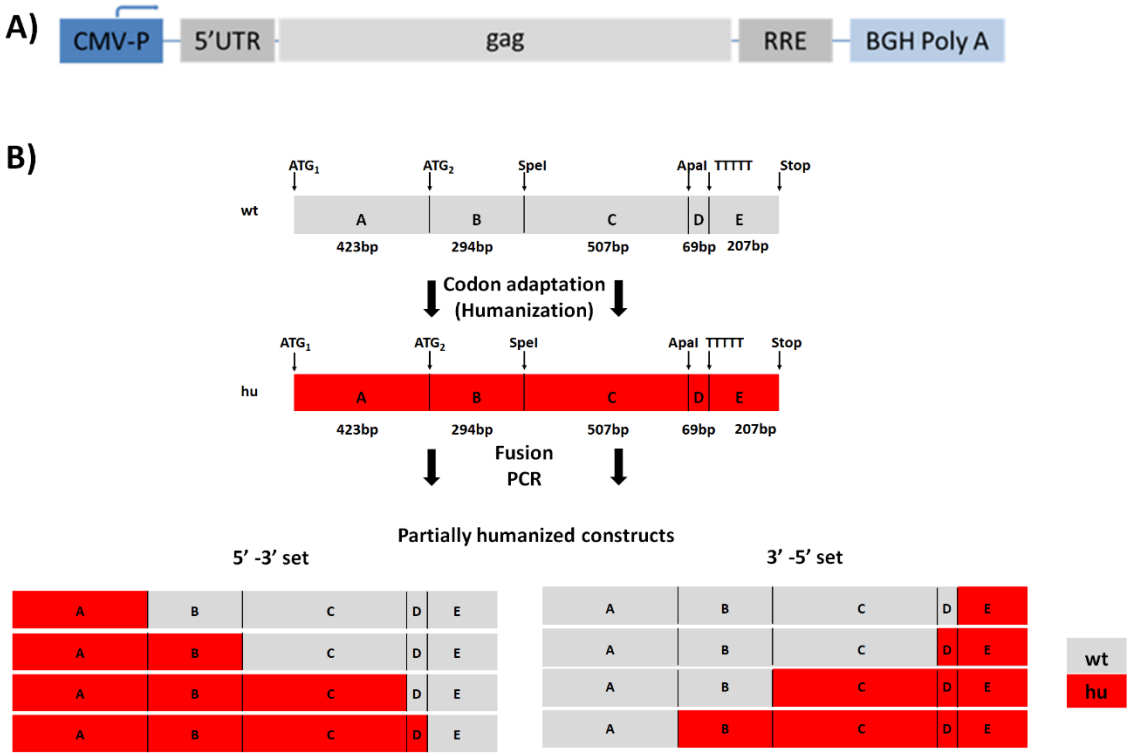

Figure E-1: Schematic illustration of a gag subgenomic construct in general (A) and partially humanized variants (B)
Gag reporter constructs were expressed under control of the heterologous CMV promoter and contained the 5'UTR, the RRE and a polyadenylation site. Based on the wild-type (wt) and fully humanized (hu) gag variant, partially humanized variants were generated that can be divided into a 5'-3' as well as 3'-5' subset.

biological motifs. The boundary between A and B was defined by the second ATG codon within the *gag* sequence, which is might be used for translation inititation in some cases [140]. The C part was confined by two restriction sites, SpeI and ApaI, which were used in previous studies of our group. The end of part D is determined by the slippery site [27], whereas E ends with the stop codon of *gag*. As a starting point for the partially adapted variants, the wild-type *gag* sequence from the HIV B-clade isolate NL4-3 was used, just as the fully humanized version of this *gag* (hu-Gag or huABCDE). Using these two sequences as templates, the partially humanized variants were generated (see Figure E-1) by fusion PCR (D.5.4), all containing the desired *gag* variants under control of a CMV promoter (as shown above), together with the HIV-1 NL4-3 5'UTR, the rev-responsive-element (RRE) and a BGH-polyadenylation site. Initially, eight partially humanized variants were generated which can be divided into a 5'-3' as well as a 3'-5' set. For the 5'-3'set, humanization starts at the 5' part of *gag* (part A) and is enlarged stepwise in 3' direction (to part E). For the 3'-5' set humanization starts at part E and is enlarged into 5' direction towards part A. All variants were cloned into a pcDNA3.1 expression vector.

E.2 Sequence properties of partially humanized *gag* variants

Since codon adaptation necessarily leads to changes of the sequence, variations of specific sequence properties occur at the same time, which are known to impact gene expression, like the codon-

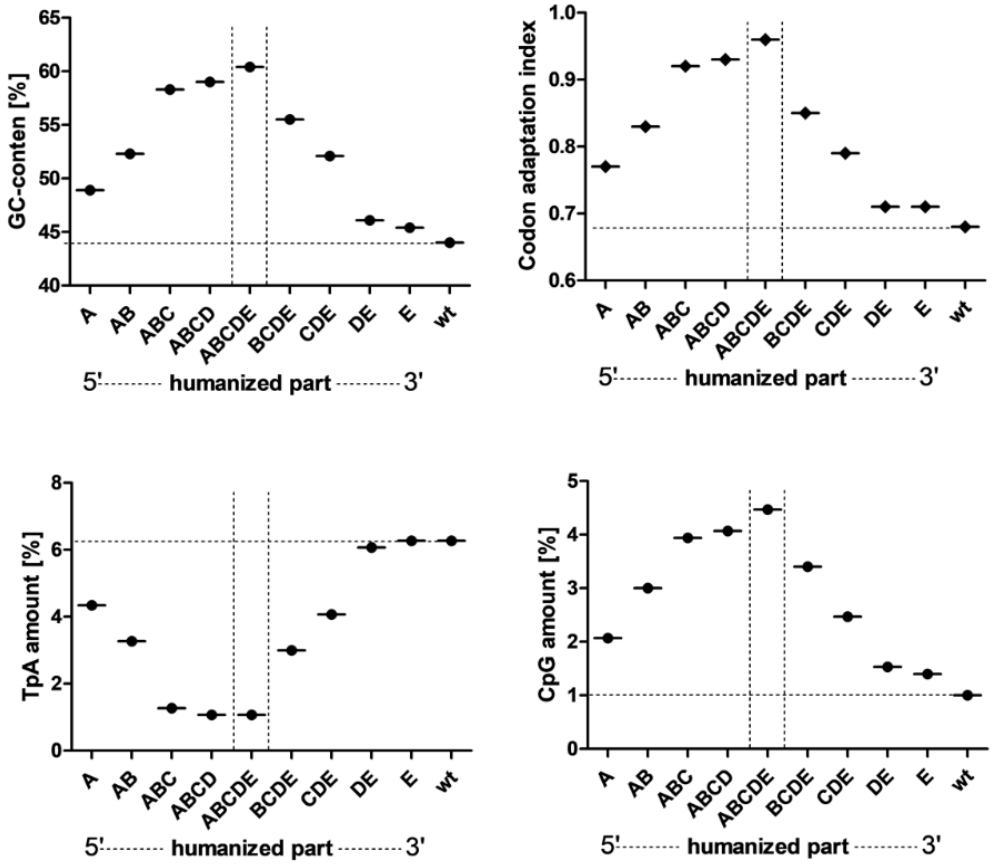

Figure E-2: Sequence properties of partially humanized gag variants regarding CAI, TpG / CpG frequency and GC-content

adaptation-index (CAI), dinucleotide frequency (e.g. CpG and TpA) or GC-content [141]–[143]. To analyze whether possible changes in protein production might possibly be a consequence of deviations in those parameters, these properties were evaluated for all partially humanized variants by bioinformatic analyses to quantify the differences of these parameters for the variants. As shown in Figure E-2, stepwise increase or decrease of humanization by gradual adaptation of *gag* to human codon usage leads to a progressive increase or decrease regarding the analyzed parameters. Accordingly, GC content rises from 54% for huA_wtBCDE to a maximum of 60.4% for the fully humanized *gag* variant (huABCDE) and decreases again to a value of 45.5% for wtABCD_huE. The wild-type *gag* possesses a GC-content of 44.0%. An equivalent course can be seen for the codon adaptation index and the frequency of dinucleotide CpG. According to the A-rich genome of HIV-1,

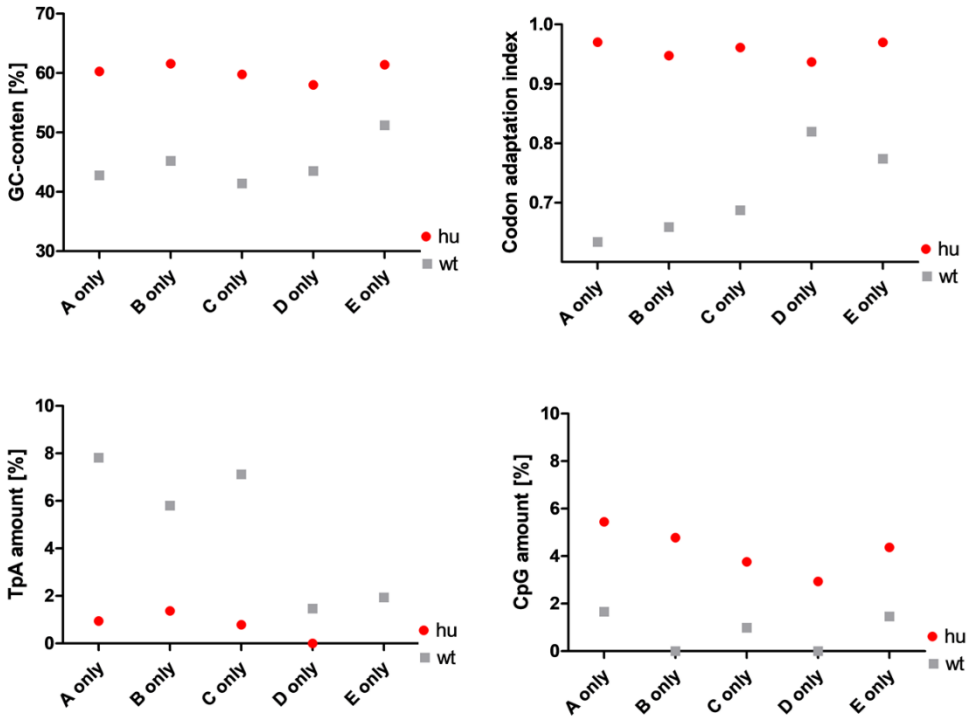

Figure E-3: Sequence properties of single gag parts (A-E) regarding CAI, TpG / CpG frequency and GC-content:

the analysis of the amount of the dinucleotide TpA reaches its maximum for the wild-type *gag* (6.27%) and is minimal for the completely humanized huABCDE variant. Additionally, the five different *gag* parts A-E were also individually analyzed in order to examine how the single parts differ between each other, since positional effects of humanization should be analyzed as well. Here it was seen that humanized and wild-type variants of the single *gag* parts differ strongly from each other, but there was no significant discrepancy between the single humanized or wild-type parts among themselves. The GC- content for example, varies for the humanized sequences between 60.3% (A part) and 61.4% (part E) whereas for the wild-type sequence values lie between 41.4% (part C) and 51.2% (part E). Comparable differences regarding the other analyzed sequence parameters were observed. The wild-type fragment D showed the smallest discrepancy in comparison to its humanized

version which might be because of the small size of the fragment. Further, the wild-type parts D and E showed decreased TpA levels compared to the wild-type fragments A, B or C.

E.3 Influence of humanization on *gag* protein expression of 5‘-3’ adapted variants

Bioinformatic analysis of the 5’- 3’-adapted *gag*-variants showed a systematic increase or decrease in the frequency of the analyzed sequence features (Figure E-2). Of course, it was of interest whether increasing humanization also led to a systematic change in protein production. To analyze this, HEK293T cells were transfected with the 5’-3’-adapted *gag* variants. Gag protein production was quantified in the supernatants of transfected cells after 48 hours by p24-ELISA. Transfection was performed in the presence as well as in the absence of Rev.

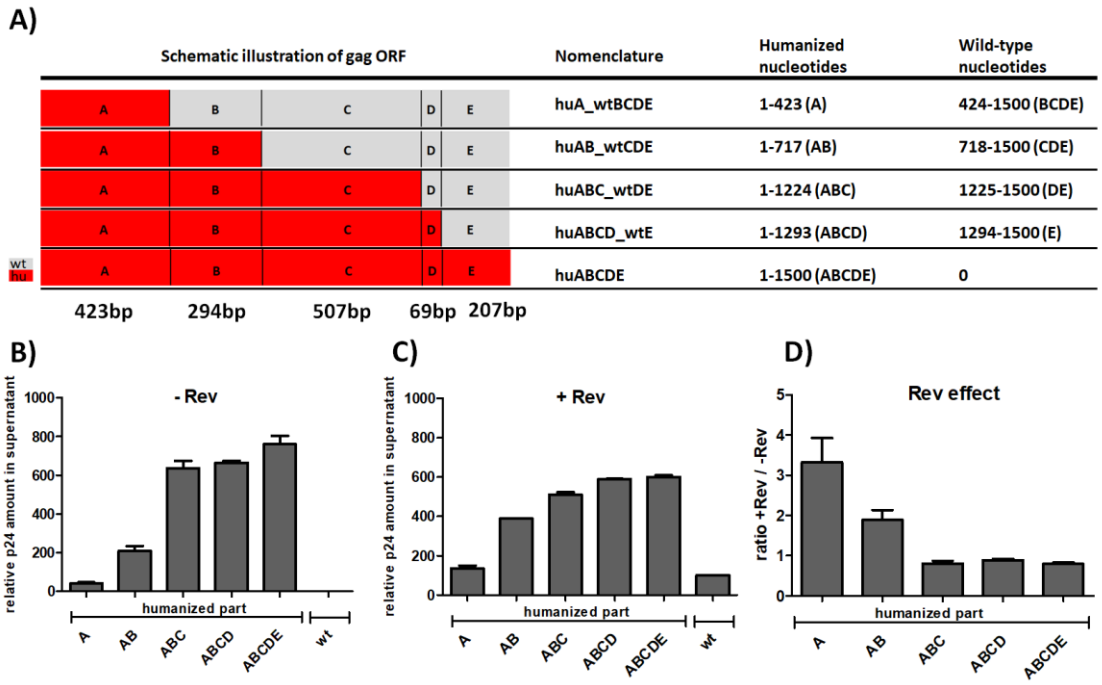

Figure E-4: Quantification of Gag expression of the 5’-3’-adapted variants on protein level measured by p24 ELISA
A) Schematic overview of the 5’-3’ adapted variants and the localization of humanized and wild-type nucleotides. B) Relative p24 expression of 5’-3’ adapted gag-variants in the absence of Rev of transfected HEK293T cells. C) Relative p24 expression of 5’-3’ adapted gag-variants in the presence of Rev of transfected HEK293T cells. D) Influence of Rev on p24 expression determined by calculating the ratio between expression in the presence and absence of Rev. All experiments were performed in biological triplicates.

To allow for a better comparison of all independently performed experiments, where absolute p24 amounts vary due to factors such as transfection efficiency, the expression of wild-type *gag* in the presence of Rev was set to 100%. As shown in Figure E-4A, all 5’-3’ adapted variants were expressed even in the absence of Rev. Furthermore, it became apparent that the numbers of humanized codons directly affected *gag* expression for the 5’-3’ subset on protein level. Increasing numbers of adapted codons accordingly led to an increase in protein production. According to that, for the variant

huA_wtBCDE the weakest Gag expression was measured with a relative p24 value of 41%. Further humanization of an additional part of *gag* (part B) led to an enhanced p24 expression of 209%. This trend was also true for the rest of the 5'- 3' adapted variants with p24 levels of ≈636%, 664% and 760% for the variants huABC_wtDE, huABCD_wtE and huABCDE (huGag). Since the accessory protein Rev is needed for wildtype Gag expression (nearly no expression of wild-type Gag in the absence of Rev), it was of interest to which extent partially humanized variants benefit from the presence of Rev. It turned out that variants with less adapted codons profit to a larger extent from the presence of Rev than stronger humanized variants. Correspondingly, the p24 expression of the subgenomic variants huA_wtBCDE and huAB_wtCDE rise by the factors of 3.3 (huA_wtBCDE) and 1.9 (huAB_wtCDE) to relative p24 levels of 134% and 389%, respectively. Those variants that show already strongly enhanced p24 levels in the absence of Rev (huABC_wtDE, huABCD_wtE and huABCDE), did not benefit by the addition of Rev. As shown in Figure E-4C, the addition of Rev rather led to a slight decrease in p24 expression. Calculating the ratios of the expression in presence and absence of Rev, led to values of 0.8 (huABC_wtDE), 0.9 (huABCD_wtE), and 0.8 (huABCDE) which might be because of the negative effect of transfection of an additional transgene. However, the analysis of *gag* expression of the partially humanized 5'- 3'-adapted variants clearly showed the beneficial effect of humanization since it enabled the enhanced expression of all partially humanized variants even in the absence of Rev.

E.4 Effect of Rev inhibition by addition of Leptomycin B on the *gag* expression of the 5'- 3' adapted *gag*-variants

To confirm the Rev-independent expression of humanized variants and an enhanced beneficial Rev-effect for variants with fewer humanized nucleotides, the expression analysis was performed again for the 5'-3' set in the presence of Rev and Leptomycin B (LMB). LMB is an unsaturated, branched-chain fatty acid, which inhibits the export of proteins containing a nuclear export signal. The mechanism of inhibition is mediated by the direct binding of leptomycin B to CRM1, which blocks the binding of CRM1 to proteins containing the nuclear export signal [144], here especially the Rev protein. According to the data presented in E.3, LMB should have a stronger effect on the expression of *gag*-variants with a smaller amount of humanized nucleotides, since those variants rely stronger on the Rev-mediated expression. To test this, HEK293T cells were transfected as described and LMB was added to the cells in a concentration of 2.5nm immediately after transfection. Since CRM1, which is inhibited by LMB exports not only the Gag protein but mediates the export of several cellular molecules, the optimal concentration of LMB was determined in several toxicity tests to block the export of *gag* mRNA as good as possible without killing of transfected cells (data not shown). Nevertheless, a residual toxicity of LMB could not be prevented. After 48 hours, p24 expression was measured in the supernatants of transfected cells by p24 ELISA (Figure E-5). In general, the expression levels of the different variants correlated with the number of humanized nucleotides. The weakest p24 expression of the partially humanized variants was detected for the huA_wtBCDE (30%)

construct, which is the least humanized one. In contrast, the completely humanized variant huABCDE showed the highest p24 expression with a relative p24 amount of 329%. The remaining variants lie in between with relative p24 levels of 118% (huAB_wtCDE), 224% (huABC_wtDE) and 295% (huABCD_wtE). So, as expected, the addition of LMB lead to general impaired expression levels because of its toxicity. To point out that the addition of LMB did not affect all 5'-3'-adapted variants to the same extent, the impact of LMB treatment was calculated by dividing the relative expression levels in the absence of LMB by the expression levels in the presence of LMB (Figure E-5C). Here it became apparent that the effect of LMB is higher for those variants with smaller numbers of humanized codons, similar to the effect of Rev, shown in Figure E-4D. The LMB ratio of the huA_wtBCDE variant was the highest with 4.6, followed by 3.3 for huAB_wtCDE. The residual variants huABC_wtDE, huABCD_wtE and huABCDE all showed ratios around the factor of 2 (2.3, 2.0 and 1.9) At first glance this might be unexpected, since those variants did not benefit from the addition of Rev as shown in Figure E-4D, hence LMB should also not affect the protein levels of those variants. However, because of the general impaired protein production in the presence of LMB due to toxicity, the reduced p24 level that were used for calculation of the LMB effect lead to values that were higher than those calculated for the Rev effect. Anyway, the lower effect of LMB on expression of the variants huABC_wtDE, huABCD_wtE and huABCDE compared to the variants huA_wtBCDE and huAB_wtCDE indicates the lower beneficial effect of Rev on these variants.

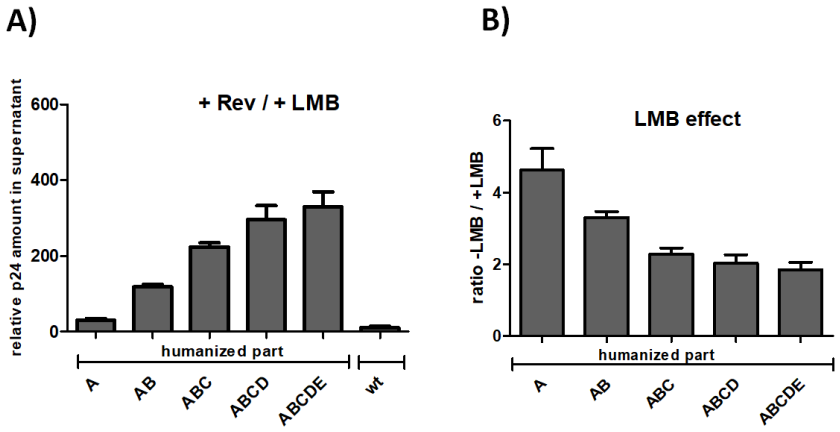

Figure E-5: Quantification of Gag expression of the 5'- 3'-adapted variants on protein level measured by p24 ELISA in the presence of Rev and Leptomycin B (LMB). For a detailed description of the analyzed variants see Figure E-4
A) Relative p24 expression of 5'-3' adapted gag-variants in the presence of Rev and LMB of transfected HEK293T cells. B) Influence of LMB on p24 expression determined by calculating the ratio between expression in the absence of LMB (see Fig. E-5C) and the expression in the presence of LMB. All experiments were performed in biological triplicates.

Therefore, the thesis of a reduced impact of Rev on variants with higher degrees of humanization could be confirmed as well as the cumulative effect of humanization, with higher expression levels for variants with higher numbers of humanized nucleotides.

E.5 Influence of humanization on Gag protein expression of 3‘-5’ adapted variants

To investigate whether besides the cumulative effect of humanization also the position of codon adaptation is important for enhanced and Rev-independent *gag* expression, partially humanized *gag* variants were generated, where humanization was enlarged diametrically opposed to the initial set from 3’ to 5’ direction (3’-5’ set). Expression of these variants (wtABCD_huE, wtABC_huDE, wtAB_huCDE and wtA_huBCDE) was also analyzed in the absence and presence of Rev on protein level. For this, HEK293T cells were transfected with the 3’-5’-adapted *gag* variants. Gag protein production was quantified in the supernatants of transfected cells after 48 hours by p24-ELISA.

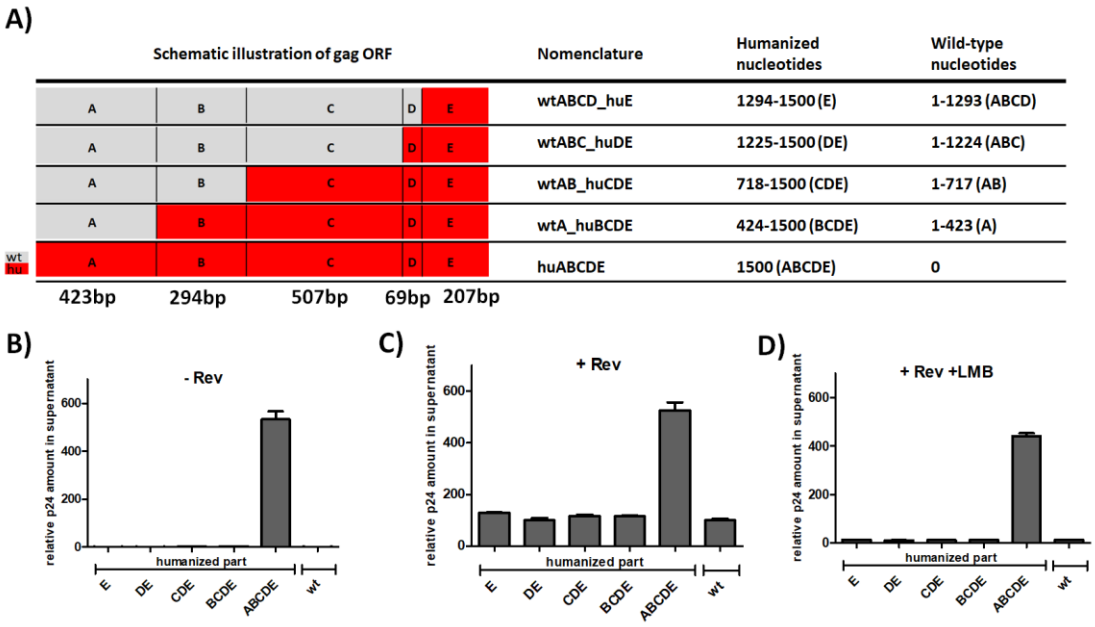

Figure E-6: Quantification of Gag expression of the 3’-5’-adapted variants on protein level measured by p24 ELISA
A) Schematic overview of the 3’-5’-adapted variants and the localization of humanized and wild-type nucleotides. B) Relative p24 expression of 3’-5’-adapted gag-variants in the absence of Rev of transfected HEK293T cells. C) Relative p24 expression of 3’-5’-adapted gag-variants in the presence of Rev of transfected HEK293T cells. All experiments were performed in biological triplicates.

Surprisingly, the expression pattern of 3’-5’-adapted variants differ remarkably from the expression pattern of the 5’-3’ set. All 3’-5’ adapted variants showed expression below or just slightly above the detection limit. So, in the absence of Rev, stepwise humanization did not lead to stepwise enhancement of expression. In contrast to the expression patterns for the 5’-3’ set, no gradual increase of *gag* expression was observed by incremental humanization. Expression of those variants could be rescued by the addition of Rev, but only to a level that was comparable to wild-type *gag* in the presence of Rev, which was set to 100% for all experiments. In the presence of Rev, additional humanization neither had a beneficial effect on *gag* expression, since all variants showed comparable expression levels. The quantified relative p24 levels lie between a minimum of 100% for the

wtABC_huDE variant and a maximum of 129% for the wtABCD_huE. Therefore, analysis of the 3'-5' set underlined the importance of human codon usage at the 5'part of *gag* for our experiments. The direct correlation between Gag protein amount and the number of humanized codons as well as the Rev-independent *gag* expression and enhanced expression levels in general were only observed for a humanized 5'part of *gag*.

E.6 Characterization of the impact of the 5' end of the HIV-1 *gag* on gene expression of codon adapted variants

E.6.1 Analysis of the impact of the first 423 bp at the 5' end of HIV-1 *gag*

The results so far showed that codon adaptation influences gene expression in a positional as well as cumulative way, but humanization at the 5' part of *gag* seemed to be required for enhanced and Rev-independent expression Variants that lack 5' end humanization showed no *gag* expression in the absence of Rev, independently of the number of adapted codons in the rest of the sequence. Therefore, it appeared possible that either the 5' part of wild-type *gag* contains an inhibitory motif or the codon choice of the 5' part influences the protein production level and Rev-dependency in

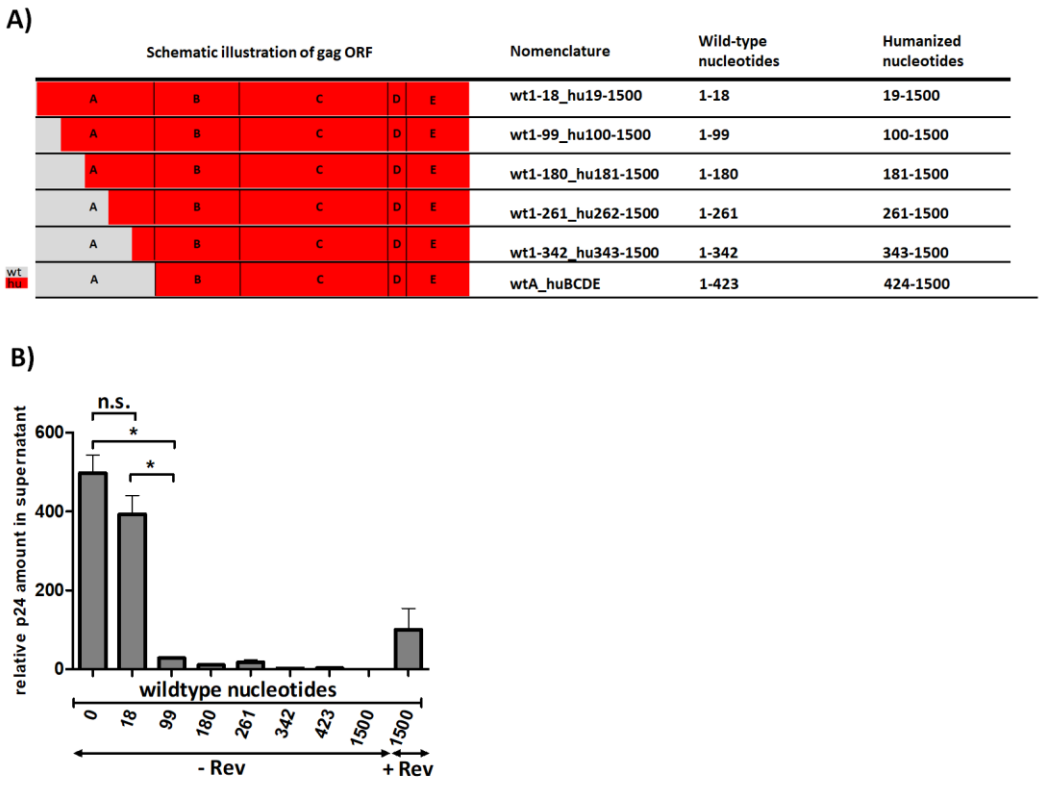

Figure E-7: Quantification of Gag expression of the further subdivided 3'-5' variants on protein level measured by p24 ELISA (modified from [145])

A) Schematic overview of the further subdivided 3'-5' variants and the localization of humanized and wild-type nucleotides B) Relative p24 expression of further subdivided 3'-5' gag-variants of transfected HEK293T cells. Rev was added as indicated. All experiments were performed in biological triplicates.

general. To answer these questions Maria Deichner subdivided the 423 bp part A of HIV-1 *gag* into five additional variants with stepwise increases in the number of humanized nucleotides (Figure E-7) as part of her bachelor-thesis. Starting with the wtA_huBCDE variant which has wt codons from nucleotide 1 to nucleotide 423, for each additional variant 81 nucleotides were additionally humanized. This started at the border between part A and B and was done towards the 5' end of *gag*. In that way, the variants wt1-18, wt1-99, wt1-180, wt1-261, wt1-342, were generated.

Quantification of *gag* expression of the additionally subdivided 3'-5' variants emphasized the importance of the 5' end of HIV-1 *gag* for enhanced and Rev-independent protein production after humanization. The completely humanized variant, which has no wild-type codons (wt0=huABCDE) expectedly showed the strongly enhanced p24 expression even in the absence of Rev (498%), which is in accordance with the previous results so far. A slightly, but not significantly reduced expression level was also seen for the wt1-18_hu19-1500 variant with relative p24 expression of 393%. In contrast, the variant with additional 81 wild-type nucleotides (wt1-99_hu100-1500) already exhibited significantly impaired p24 levels (28%). For the residual variants with additional wild-type nucleotides, *gag* expression was even more impaired with relative p24 levels of 11% for wt1-180_hu181-1500 and 17% for wt1-261_hu262-1500. For the variants wt1-342_hu343-1500 and wt1-423_hu424-1500 (=wtA_huBCDE) p24 expression was just slightly above the detection limit. As expected, the wild-type *gag* variant (without humanization) showed no expression in the absence of Rev. These results suggest that a potential inhibitory motif might lie within the first 100 nucleotides of HIV-1 wild-type *gag* since all variants with wild-type codon usage in this area showed significantly impaired protein production compared to the fully humanized variant. Therefore, codon adaptation of the first 100 nt is required to prevent the dramatic collapse in *gag* expression. Further, just a low number of wild-type nucleotides (18 nucleotides) at the 5'-end do not lead to dramatically impaired protein production.

E.6.2 Analysis of the importance of the first 100 base pairs at the 5' end of HIV-1 *gag*

The aforementioned expression analysis showed that wild-type codon usage at the 5' part of HIV-1 *gag* seems to mediate a negative effect on protein production. If the nucleotides from 1-99 are not humanized, enhanced *gag* expression in the absence of Rev is prevented. In order to find out whether humanization of all of the first 100 nucleotides is required to abolish impaired *gag* expression, variants were generated that further shorten the part at the 5' end with HIV codon usage. For this, variants were generated by fusion PCR that enlarge the number of wild-type nucleotides stepwise by the number of 15 nucleotides (i.e. 5 codons). Starting with a fully humanized variant (without any wild-type nucleotides), seven additional subgenomic constructs were generated with either 15 (wt1-15_hu16-1500), 30 (wt1-30_hu31-1500), 45 (wt1-45_hu46-1500), 60 (wt1-60_hu61-1500), 75 (wt1-75_hu76-1500), 90 (wt1-90_hu91-1500) and 105 (wt1-105_hu1-1500) nucleotides at the 5' end of HIV-1 *gag*. Expression analysis for these variants after transfection of HEK293T cells in the presence

as well as in the absence of Rev was again performed by collecting the supernatants at 48 hours post transfection and performing a p24 ELISA (Figure E-8).

It became apparent that 60 nucleotides with HIV codon usage at the very 5'-end of the *gag* gene are tolerated in the absence of Rev. All variants with 60 wt nucleotides or less at the 5' part of *gag* were expressed even in the absence of Rev. The four variants with this feature (wt1-15_hu16-1500, wt1-30_hu31-1500, wt1-45_hu46-1500, wt1-60_hu61-1500) can be divided into two groups, according to their expression levels. The wt1-15_hu16-1500 variant showed comparable p24 levels (507 %) to the fully humanized variant (535%). The second group, consisting of the three variants wt1-30_hu31-1500, wt1-45_hu46-1500, and wt1-60_hu61-1500 showed reduced expression levels compared to the fully humanized variant, although only wt1-60_hu61-1500 was significantly reduced compared to the fully humanized variant (huABCDE) in an unpaired t-test (with $p < 0.05$ considered significant and $p < 0.01$ considered highly significant). However, these three variants were still expressed better in the absence of Rev than the wild-type construct with Rev (≈ factor of 2), which was set to 100%. The three remaining variants (wt1-75_hu76-1500, wt1-90_hu91-1500 and wt1-105_hu1-1500

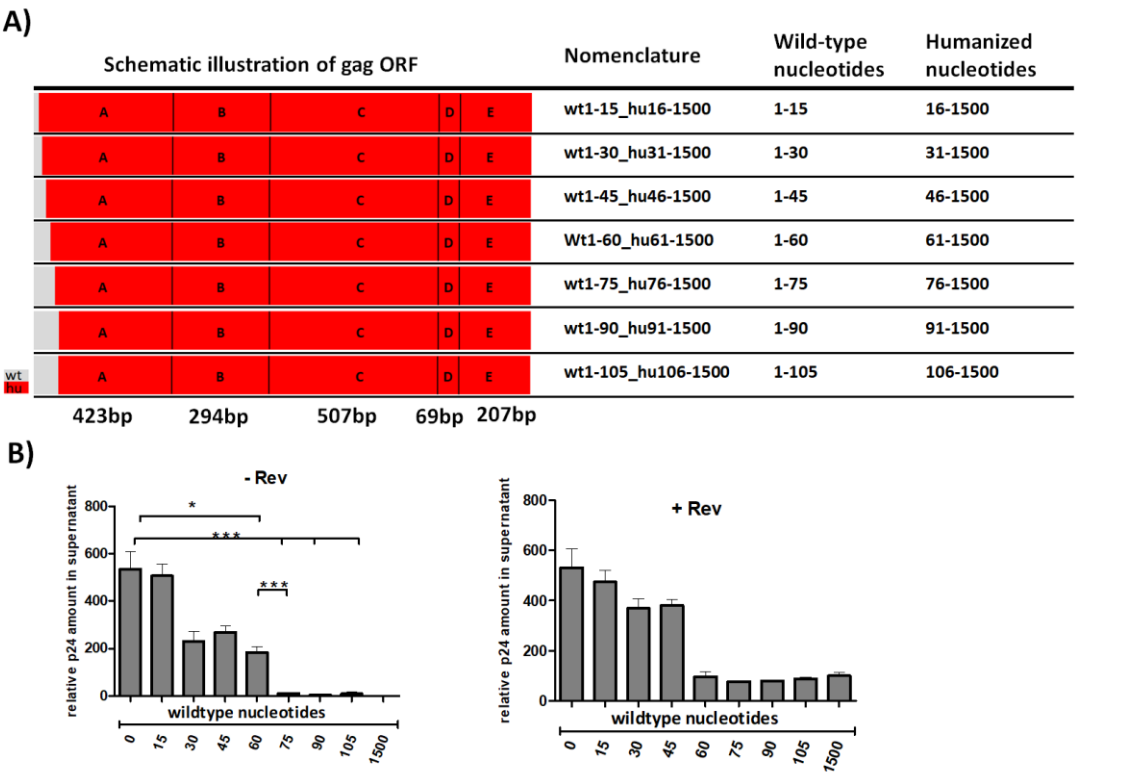

Figure E-8: Quantification of Gag expression of variants further subdivided within the first 100 base pairs
A) Schematic overview of the further subdivided variants with focus on the first 100 nucleotides of HIV-1 gag and the localization of humanized and wild-type nucleotides. B) Relative p24 expression of these gag-variants after transfection into HEK293T cells in the absence of Rev, and C) in the presence of Rev. All experiments were performed in biological triplicates.

showed expression levels in the absence of Rev only slightly above the detection limit and were highly significantly reduced compared to the fully humanized variant. Therefore, it was concluded that the 15 additional wild-type nucleotides of variant wt1-75_hu76-1500 are involved in impairing Gag protein production compared to the variant wt1-60_hu61-1500.

Co-expression of Rev could rescue variants that lacked expression without Rev. However, with a maximum relative p24 level of 89 % (wt1-105_hu106-1500), none of these variants exceeded the wild-type *gag* in the presence of Rev, which was set to 100%. The wt1-15_hu16-1500 variant that showed a strongly enhanced *gag* expression even in the absence of Rev, with more than 5-fold enhanced expression levels compared to the reference wild-type *gag* variant, was unaffected by the addition of Rev (relative p24 level of 530 %). For the two variants wt1-30_hu31-1500 and wt1-45_hu46-1500, the addition of Rev led to an enhanced expression by the factor of 1.6 (relative p24 level of 369 % in the presence vs 230% in the absence of Rev) or 1.4 (relative p24 level of 380 % in the presence vs 286 % in the absence of Rev), respectively. The only variant that showed a reduction in p24 expression after the addition of Rev, was the variant wt1-60_hu61-1500. The expression decreased from 183 % (without Rev) to 97 % (with Rev), which might be a sign for competitive Rev-dependent and Rev-independent pathways.

E.6.3 Localization of a potential inhibitory sequence motif at the 5' end of HIV-1 *gag*

To investigate whether the loss in expression is induced by the continuous usage of wild-type codons from the start ATG over the first 75 nucleotides of *gag* or whether one or more smaller parts within this region is responsible for the effect, further *gag* variants were generated that were designed as follows. All variants were generated with human codons from nucleotide 91 to the end of the gene. In the area between nucleotide 1 and nucleotide 90, different parts with varying length of wild-type codon usage were inserted at several positions (Figure E-9A). The first five variants all contained only 15 wild-type nucleotides, while the rest of *gag* exhibited human codon usage. The position of the 15 nucleotides was shifted within the first 90 nucleotides of the 5'part of *gag* (hu1-15_wt16-30_hu31-1500, hu1-30_wt31-45_hu46-1500, hu1-45_wt46-60_hu61-1500, hu1-60_wt61-75_hu76-1500, and hu1-75_wt76-90_hu91-1500). Additionally, five further variants were generated with at least 30 wild-type nucleotides at different positions within the 5' part of *gag* (hu1-15_wt16-90_hu91-1500, hu_1-30_wt31-90_hu91-1500, hu1-45_wt46-90_hu91-1500, hu1-60_wt61-90_hu91-1500 and hu1-30_wt31-75_hu76-1500). Expression analysis for these variants was performed as before in the presence as well as in the absence of Rev.

As shown in Figure E-9, wild-type codon usage within the first 60 nucleotides of HIV-1 *gag* had no influence on protein production when the residual codons are humanized, consistent with the results described above. The three variants hu1-15_wt16-30_hu31-1500, hu1-30_wt31-45_hu46-1500 and

hu1-45_wt46-60_hu61-1500 exhibited p24 expression levels comparable to the fully humanized variant (huABCDE) in the absence of Rev, indicating that this area does not contain the crucial sequence which potentially mediates the inhibitory effect. However, shifting of the 15 nucleotides into 3' direction led to an impaired protein production in the absence of Rev that was also statistically highly significant ($p < 0.01$in unpaired t-test). For instance, the variant hu1-60_wt61-75_hu76-1500 showed p24 levels slightly above the detection limit in the p24 ELISA, despite the marginally changed position of the 15 nucleotides with wild-type-codon usage. In addition, all further variants that include wild-type codons between the positions 61 and 75 showed weak p24 expression in the absence of Rev, and Gag levels were significantly reduced compared to the fully humanized variant huABCDE. The addition of Rev led to enhanced p24 levels mainly for variants with very low *gag* expression in the absence of Rev. Especially the variants hu1-60_wt61-75_hu76-1500 and

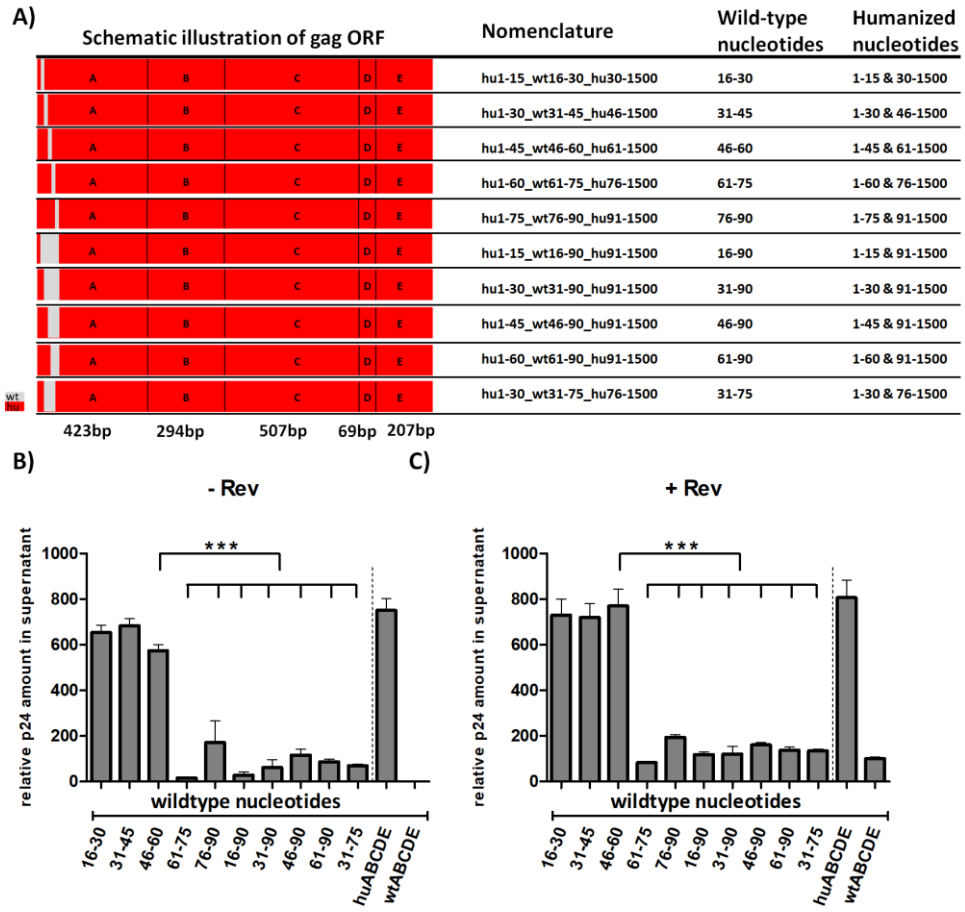

Figure E-9: Quantification of Gag expression by p24 ELISA of the further subdivided variants for localization of a potential inhibitory motif at the 5' part of HIV-1 gag

A) Schematic overview of the further subdivided variants with focus on the first 90 nucleotides of HIV-1 gag and the localization of humanized and wild-type nucleotides B) Relative p24 expression of further subdivided gag-variants of transfected HEK293T cells in the absence of Rev, and C) in the presence of Rev. All experiments were performed in biological triplicates

hu1-15_wt16-90_hu91-1500 showed enhanced relative p24 levels by the factor of 5.3 (hu1-60_wt61-75_hu76-1500) or 4.3 (hu1-15_wt16-90_hu91-1500). All other variants were only marginally affected by the addition of Rev with factors between 1.1 (hu1-30_wt31-45_hu46-1500) and 2.0 (hu1-1-30_wt31-75_hu76-1500). As observed before, addition of Rev can lead to an enhanced *gag* expression, but only to a level which is comparable to wild-type *gag* expression in the presence of Rev. Taken together, expression analysis of the additionally generated variants with only 15 wild-type nucleotides at different locations within the first 90 nt of the 5' end of HIV-1 *gag*, indicated a negative effect of wild-type nucleotides between the positions 61 and 75 of HIV-1 *gag*. Further, the surrounding nucleotides also seem to have some impact on this potential inhibitory motif.

E.7 Confirmation of the position of the inhibitory sequence motif

To confirm the localization of the inhibitory motif in the 5' part of *gag*, variants with inverted codon usage were analyzed. Those variants contained humanized codons between nucleotides 61 and 75 in a wtABCDE or huBCDE *gag* gene. It was of interest whether conversely changing codons from wild-type to humanized codon usage in this area can enhance *gag* expression for variants with otherwise weak expression levels in the absence of Rev. The generated variants were named wt1-60_hu61-75_wt76-1500 and wt1-60_hu61-75_wt76-423_hu424-1500 (Figure E-10). Gag expression was as before alongside the control variants huABCDE, wtABCDE, huBCDE and hu1-60_wt61-75_hu76-1500.

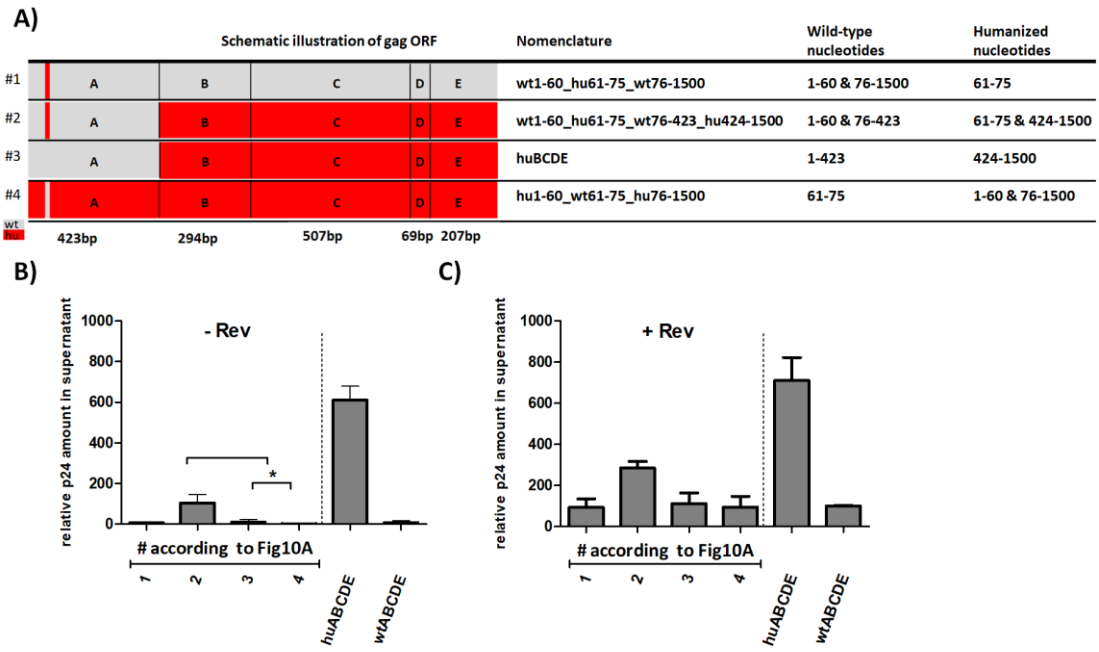

Figure E-10: Quantification of Gag expression on protein level measured by p24 ELISA of variants with inverted codon usage for confirmation of a potential inhibitory motif at the 5' part of HIV-1 gag

A) Schematic overview of the further subdivided variants of HIV-1 gag and the localization of humanized and wild-type nucleotides B) Relative p24 expression of further subdivided gag-variants of transfected HEK293T cells in the absence of Rev, and C) in the presence of Rev. All experiments were performed in biological triplicates

Expression analysis showed that changing the nucleotide composition between nucleotides 61 and 75 from wild-type to human codon usage, indeed led to a significant improvement in *gag* expression for the huBCDE variant. As expected, this variant showed nearly no expression in the absence of Rev (relative p24 level of 2.3 %), whereas the same variant with human codon usage only in the short stretch between nucleotides 61 and 75 (wt1-60_hu61-75_wt76-423_hu424-1500) showed expression levels comparable to the reference variant wtABCDE in the presence of Rev with a relative p24 expression of 105 %. In contrast, the variant wt1-60_hu61-75_wt76-1500, which represents a wtABCDE with human codon usage between nucleotides 61 and 75 did not benefit from the 15 humanized nucleotides, with a relative p24 level of 6.8 % in the absence of Rev. As observed in the previous experiments, the addition of Rev rescued the variants with initially very low p24 levels (wt1-60_hu61-75_wt76-1500, wtABCDE, huBCDE and hu1-60_wt61-75_hu76-1500) to levels comparable to the reference variant (wtABCDE in the presence of Rev). Therefore, it can be concluded that removing the inhibitory motif between nucleotides 61-75 of *gag* by the changing of codons is necessary but not enough for enhanced and Rev-independent protein production. With the inhibitory motif inactivated, further humanization of the rest of the *gag* gene contributes positively to an enhanced protein production.

E.8 Influence of humanization on mRNA expression levels of 5'-3'- and 3'-5'-adapted variants

One crucial question regarding the changes in protein production due to codon adaptation as described above, is which biological processes are mainly affected. Until now, all analyses have been performed on protein level via a p24-ELISA. Therefore, an important issue was, whether changes on protein level can also be seen already on the level of mRNA expression and whether the influence of the localized inhibitory motif is also true for mRNA expression. Thus, the 5'-3' and 3'-5' sets were analyzed regarding their mRNA expression levels. To do this, cells were transfected with the desired variants and cultivated for 48 hours. Then cells were harvested and RNA isolation was performed. The *gag* mRNA was quantified and normalized to neomycin mRNA by real-time quantitative PCR analysis, since all expression plasmids contained a neomycin resistance gene.

As illustrated in Figure E-11, the mRNA expression levels in general reflected the protein expression patterns. For the 5'-3'-adapted variants an enhanced *gag* expression was seen already in the absence of Rev. All variants with 5' humanization (huA, huAB, huABC, huABCD, and huABCDE) showed an enhanced expression compared to the wild-type variant. Further the beneficial influence of humanization rises with the amount of humanized nucleotides, comparable to the protein expression data (Figure E-4). So, the variant with 423 humanized nucleotides (huA) showed a relative *gag* mRNA amount of 57 %, and the expression rises to 291 % for the variant huAB (717 codon-adapted nucleotides). As seen before, the variants huABC (1224 humanized nucleotides), huABCD (1293 humanized nucleotides) and huABCDE (1500 humanized nucleotides) showed even higher *gag* mRNA

levels with 710 %, 858 % and 805 %, respectively. The addition of Rev had a beneficial effect for the variants with lower numbers of humanized nucleotides (huA and huAB). The relative mRNA levels rise to 151 % (huA) and 514 % (huAB) by the addition of Rev. For the variants huABC, huABCD, and huABCDE that showed already strongly enhanced *gag* mRNA expression in the absence of Rev, the beneficial effect of co-transfected Rev was negligible. This became even more clear by calculating the Rev effect for all 5'-3'-adapted variants (Figure E-11D) as ratio of the *gag* mRNA expression levels in

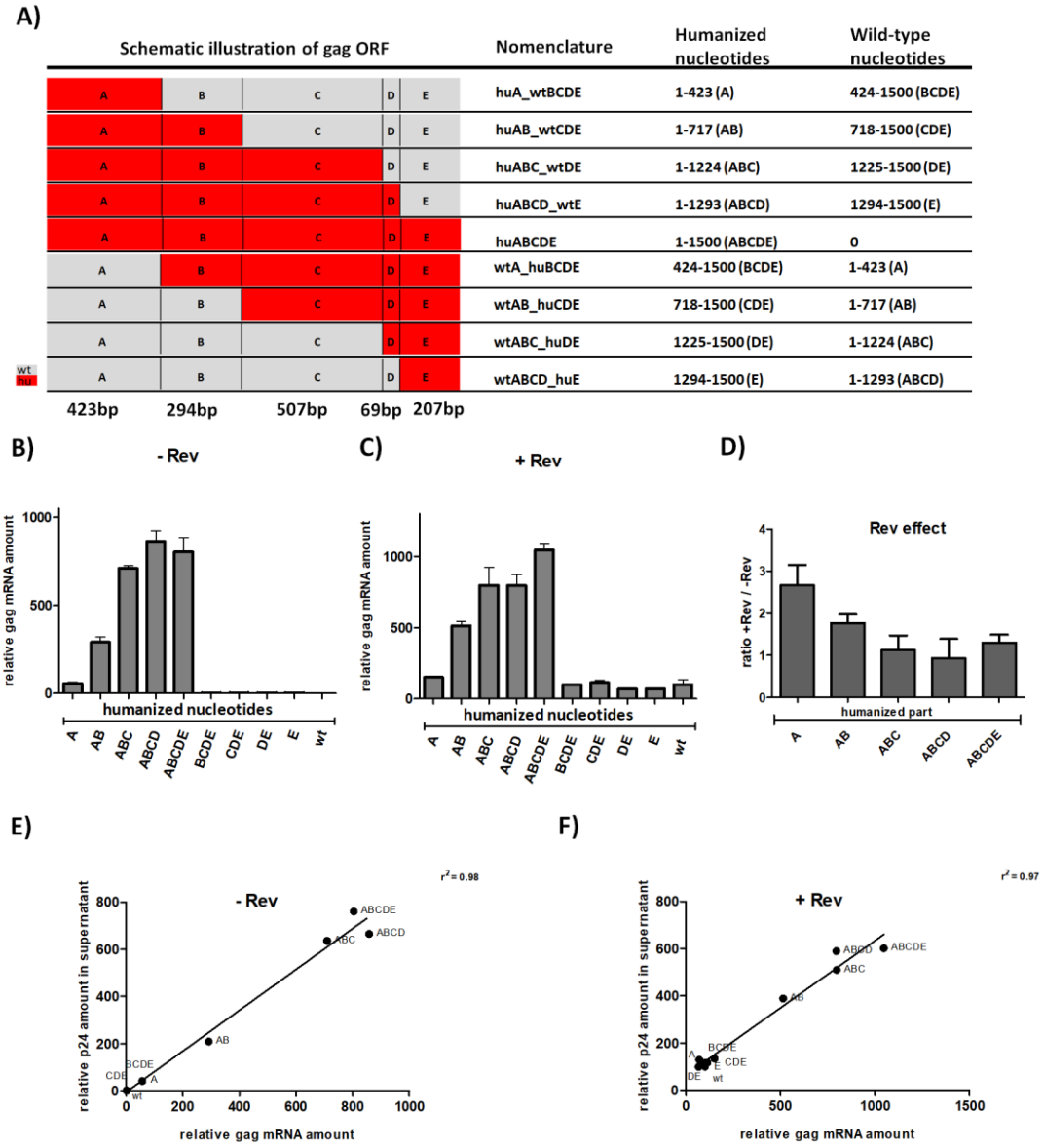

Figure E-11: Quantification of Gag expression of the 5'-3-' and 3'-5'-adapted variants on mRNA level by RT-qPCR
A) Schematic overview of the 5'-3'- and 3'-5'-adapted variants and localization of humanized and wild-type nucleotides. B) Relative gag mRNA expression of 5'-3' and 3'-5'-adapted gag-variants in the absence of Rev in transfected HEK293T cells. C) Relative gag mRNA expression of 5'-3'and 3'-5'-adapted gag-variants in the presence of Rev of transfected HEK293T cells. D) Influence of Rev on gag mRNA expression (ratio between expression in the presence and absence of Rev) for 5'-3'-adapted variants. All experiments were performed in biological triplicates. E) Correlation of gag mRNA and p24 expression in the absence of Rev and F) in the presence of Rev

the presence absence of Rev. The variant huA benefited the most from the addition of Rev with a calculated ratio of 2.7, followed by the variant huAB with a ratio of 1.7. The Rev effect of the variants huABC, huABCD and huABCDE was smaller with values of 1.1 (huABC), 0.9 (huABCD) and 1.3 (huABCDE).

So, in general, the overall trend of protein expression pattern was reflected quite well for the 5'- 3' as well as 3'-5'-adapted variants. As already seen for *gag* protein expression, humanization is also required for enhanced and Rev-independent *gag* mRNA expression on mRNA level. In the absence of Rev, none of the 3'-5'-adapted variants were expressed, independently of the number of humanized codons. For all analyzed variants, *gag* mRNA expression lies between 1.0 % for variant huBCDE and 2.2 % for variant huCDE. As expected, the addition of Rev led to enhanced *gag* mRNA levels for all 3'-5'-adapted variants. However, as already seen for *gag* protein expression, none of those variants could significantly exceed the reference variant (wtABCDE) in the presence of Rev. Taken together, the importance of the humanization of the 5' part of Gag also seems to be true for mRNA expression, since mRNA data correlated very good with *gag* protein expression profiles of 5'- 3' and 3'-5'-adapted *gag* variants in the presence as well as in the absence of Rev (Figure E-11E and Figure E-11F).

E.9 Analysis of transcription efficiency for selected partially humanized variants

Since p24 and *gag* mRNA expression levels correlated very strongly, it seems like codon adaptation influences expression already on the RNA level. This leads to the assumption that humanization directly influences mRNA amounts. The potential mechanisms are manifold and range from altered stability or transcription efficiency to splicing effects. To address this question which processes are affected by synonymous codon usage, different possible biological processes were analyzed that could be responsible for altered expression levels by codon adaptation. First, transcription efficiency was addressed by performing nuclear run-on assays in the absence of Rev. For this, HEK293T cells were transfected with the desired variants. After an incubation period of 48 hours, nuclei were isolated by centrifugation and initiated transcription was resumed by the addition of transcription buffer, containing biotinylated UTP and elongation was allowed to proceed for 30 minutes. After isolating total RNA, newly synthesized molecules were separated by streptavidin-coupled dynabeads (see D.6.2). Lastly, RT-qPCR was performed for quantification of this biotin-labelled mRNA molecules. Importantly, this assay allows quantification of the newly synthesized mRNA molecules independently of other processes that might influence mRNA amount (e.g. stability) [146], [147].

For a general survey of the effect of codon adaptation, transcription rate was investigated for 5'-3' (huA, huAB, huABC, huABCD, huABCDE) as well as 3'-5' (huE, huDE, huCDE, huBCDE, huABCDE)-adapted variants (Figure E-12). As already shown in Figure E-11B, total *gag* mRNA levels increased continuously for the 5'-3' set with the length of the humanized sequence part. Further, the

expression of the fully humanized variant huABCDE was about eight-fold higher compared to wtABCDE in the presence of Rev. In the following experiment, quantification of mRNA levels was performed after the purification of newly synthesized molecules to detect possible differences regarding the mRNA expression levels between total mRNA and newly synthesized mRNA molecules. Indeed, a highly significant difference between the total mRNA amount of the completely humanized *gag* variant (huABCDE) and the purified, newly synthesized mRNA molecules was detected (relative total *gag* mRNA amount 804 % vs relative newly synthesized *gag* mRNA amount 206 % [p=0.0003, unpaired t-test]). The clear difference between the total mRNA levels and the newly synthesized levels in the nuclear run-on assay was also detected for the variants huAB, huABC and huABCD and huABCDE with p values of at least 0.012. Only for the variant huA no significant differences were observed between both assays, most likely due to the generally reduced expression of this variant compared to the other 5'-3'-adapted variants.

In addition, the level of synthesized huABCDE mRNA was only around two times higher compared to wtABCDE. This indicates that enhanced transcription rate seems to influence RNA amounts to some degree but is not the only reason for the observed differences in total *gag* mRNA levels. Accordingly, the variants huABC, huABCD and huABCDE all showed enhanced transcriptional-rates with relative *gag* mRNA levels of 180 (huABC), 218 (huABCD), and 205 (huABCDE), respectively. Further, the variants huA and huAB exhibited comparable levels to wtABCDE, indicating no enhanced transcriptional efficiency for these variants. For 3'-5'-adapted constructs, no variant showed enhanced transcriptional transcriptional-rates. With relative newly synthesized mRNA amounts of 104 (huE), 103 (huDE), 89 (huCDE), and 115 (huBCDE), all 3'-5'-adapted variants showed comparable mRNA levels to wtABCDE.

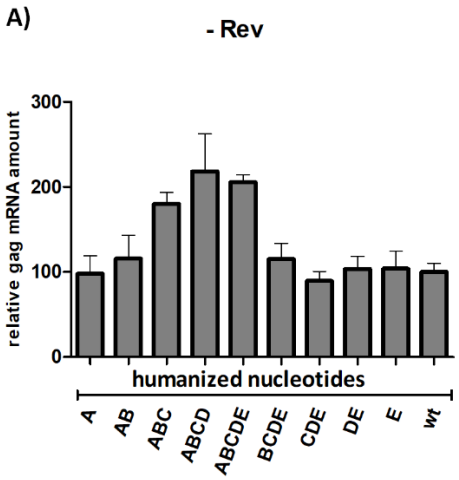

Figure E-12: Analysis of transcription efficiency by nuclear run-on assay for 5'-3' and 3'-5'-adapted variants
A) Schematic overview of the 5' -3'and 3'-5'-adapted variants and localization of humanized and wild-type nucleotides. B) Relative nuclear gag mRNA level after purification of newly synthesized transcripts for 5'-3' (left) and 3'-5' (right)-adapted variants. All experiments were performed in biological triplicates. For detailed description of the analyzed variants see Figure E-11

In prior experiments, a motif surrounding the nucleotides 61 to 75 at the 5' end of HIV-1 *gag* was detected that impaired *gag* protein production (E.6.3) in the absence of Rev. Therefore, it was of interest, whether this motif has a direct influence on the transcriptional transcriptional-rate. To address this question, nuclear run-on assays for the panel of variants that were already used for localization and confirmation of the inhibitory motif (Figure E-9 and Figure E-10) were performed. To obtain knowledge about the impact on transcriptional transcriptional-rates, RT-qPCR analysis was performed before purification of the new synthesized mRNA as well as after the purification.

A)

Schematic illustration of gag ORF	Nomenclature	Wild-type nucleotides	Humanized nucleotides
A B C D E	hu1-15_wt16-30_hu30-1500	16-30	1-15 & 30-1500
A B C D E	hu1-30_wt31-45_hu46-1500	31-45	1-30 & 46-1500
A B C D E	hu1-45_wt46-60_hu61-1500	46-60	1-45 & 61-1500
A B C D E	hu1-60_wt61-75_hu76-1500	61-75	1-60 & 76-1500
A B C D E	hu1-75_wt76-90_hu91-1500	76-90	1-75 & 91-1500
A B C D E	hu1-15_wt16-90_hu91-1500	16-90	1-15 & 91-1500
A B C D E	hu1-30_wt31-90_hu91-1500	31-90	1-30 & 91-1500
A B C D E	hu1-45_wt46-90_hu91-1500	46-90	1-45 & 91-1500
A B C D E	hu1-60_wt61-90_hu91-1500	61-90	1-60 & 91-1500
A B C D E	hu1-30_wt31-75_hu76-1500	31-75	1-30 & 76-1500

wt
hu

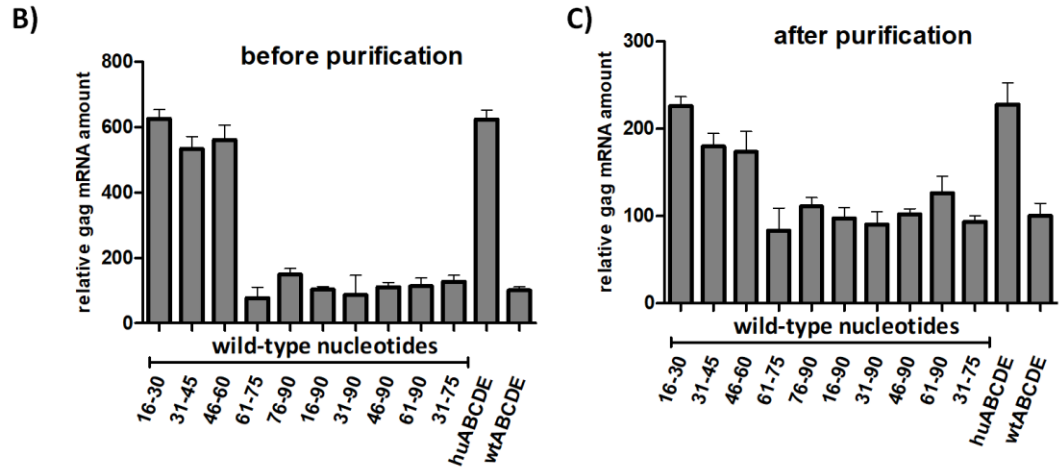

Figure E-13: Analysis of transcription efficiency by nuclear run-on assay for selected adapted variants
A) Schematic overview of the adapted variants and localization of humanized and wild-type nucleotides. B) Relative nuclear gag mRNA levels before purification of newly synthesized transcripts and C) after purification of newly synthesized transcripts. All experiments were performed in biological triplicates.

It became apparent that all analyzed variants can be divided into two groups: Those whose relative *gag* mRNA levels were enhanced (hu1-15_wt16-30_hu31-1500, hu1-30_wt31-45_hu46-1500, hu1-45_wt46-60_hu61-1500 and huABCDE) and those whose relative mRNA levels were comparable to wtABCDE (hu1-60_wt61-75_hu76-1500, hu1-75_wt76-90_hu91-1500, hu1-15_wt16-90_hu91-1500, hu_1-30_wt31-90_hu91-1500, hu1-45_wt46-90_hu91-1500, hu1-60_wt61-90_hu91-1500 and hu1-

30_wt31-75_hu76-1500). For total nuclear RNA (before purification) the maximal differences between enhanced and unaffected variants was 8.2 with the highest mRNA level of 626% (hu1-15_wt16-30_hu31-1500) and the lowest level of 77% (hu1-60_wt61-75_hu76-1500). Quantification of the isolated newly synthesized mRNA showed that all variants can similarly be divided into two groups: Those variants whose transcriptional initiation profits from humanization in the given areas and those whose transcriptional-rates remained unaffected. The variants hu1-15_wt16-30_hu31-1500**,** hu1-30_wt31-45_hu46-1500 and hu1-45_wt46-60_hu61-1500 all were comparable to huABCDE whose relative level of newly synthesized transcripts was at 227 %. In contrast, the variants hu1-60_wt61-75_hu76-1500, hu1-75_wt76-90_hu91-1500, hu1-15_wt16-90_hu91-1500, hu_1-30_wt31-90_hu91-1500, hu1-45_wt46-90_hu91-1500, hu1-60_wt61-90_hu91-1500 and hu1-30_wt31-75_hu76-1500 showed relative synthesis rate of approximately 100%. Nuclear run-on data therefore indicates that the enhanced *gag* expression is only to some degree due to enhanced transcriptional initiation. Furthermore, the localized inhibitory motif in the 5' part of *gag* influences expression level at least to some extent because of altered transcriptional transcriptional-rates. All variants that contain wild-type nucleotides between position 61 and 75 and therefore possess the potential inhibitory motif, showed impaired relative *gag* mRNA levels after the purification of the newly synthesized mRNA.

E.10 Determination of mRNA stability of selected *gag* variants

Since the differences in transcription efficiency were not sufficient to explain the differences in total mRNA levels between the various constructs, additional biological processes must be involved. A possible explanation why transcriptional transcriptional-rates represent the varying Gag mRNA levels only to some degree, could be the additional influence of differential *gag* mRNA stability. This hypothesis was addressed by performing an RNA stability assay that uses actinomycin D, an inhibitor of RNA Pol II, to prevent synthesis of new mRNAs. After transfection of the *gag* variants and addition of actinomycin D, transcription is blocked and existing transcripts get degraded over time in dependency on their stability, which allows mRNA half-life determination. Again, for a broader analysis of the impact of mRNA stability, the initially generated 5'-3' and 3'-5' humanized variants were analyzed first. It was determined that the completely humanized *gag* (huABCDE) mRNA exhibits an mRNA half-life of 6.2 hours. Contrary to that, wild-type Gag (wtABCDE) showed clearly reduced mRNA stability with a half-life of about 1.7 hours and the variant huA showed a half-life of 3.0 hours. Increasing the length of humanization also contributes positively to mRNA stability, since the half-lives of further 5'-3'-adapted variants increased with additional adapted nucleotides (huAB, 3.5 h; huABC, 5.1 h and huABCD, 5.2 h). For the 3'-5'-adapted constructs, all variants without humanized nucleotides in the 5' part of the gene (huBCDE, huCDE, huDE and huE) showed half-lives between 1.7 h and 2.2 h, which is comparable to that of wtABCDE (1.7 h). Therefore, it seems like humanization in the 5' part of Gag contributes positively to mRNA stability and is required for the additive beneficial effect on *gag* expression by increasing humanization.

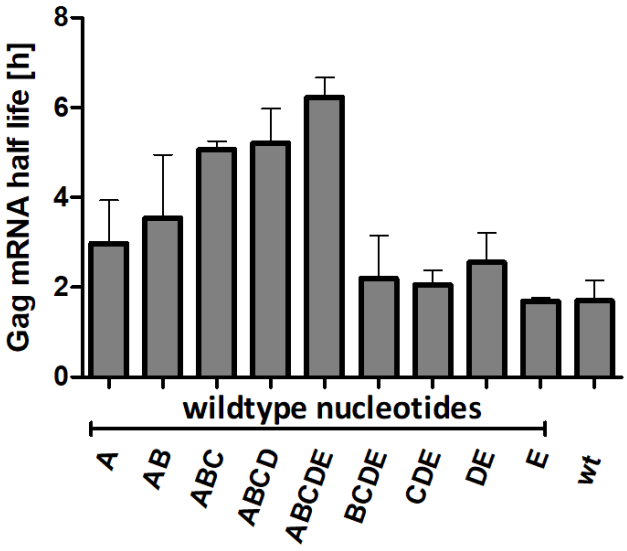

Figure E-14: mRNA stability analysis for the 5'-3' and 3'-5' sets of partially humanized variants.
Half-lives of 5'-3' and 3'-5' adapted variants. Half-lives of gag variants were determined after transfection of HEK293T cells and subsequent treatment of cells with the transcription inhibitor Actinomycin D. gag mRNA levels were quantified by RT-qPCR analysis and used for half-life calculation. All experiments were performed in biological triplicates. For detailed description of the analyzed variants see Figure E-11

Next, it was investigated whether the inhibitory motif in the 5' part of *gag* already influences *gag* mRNA stability on its own. Hence, Actinomycin D experiments were performed also for the subset of variants with small parts of wild type codon usage in a surrounding humanized context (Figure E-15). The variants again can be separated in two groups according to their expression levels. For the first group (hu1-15_wt16-30_hu31-1500, hu1-30_wt31-45_hu46-1500 and hu1-45_wt46-60_hu61-1500) wild-type codon usage had no influence on mRNA stability. Half-lives of those variants varied between 6.3 hours (hu1-30_wt31-45_hu46-1500) and 6.1 hours (hu1-15_wt16-30_hu31-1500), which is approximately the same half-life as for the fully humanized variant huABCDE (6.8 hours) in this experiment. The second group showed significantly impaired mRNA half-lives. All variants that contain wild-type codons surrounding position 75 of Gag (hu1-60_wt61-75_hu76-1500, hu1-75_wt76-90_hu91-1500, hu1-15_wt16-90_hu91-1500, hu_1-30_wt31-90_hu91-1500, hu1-45_wt46-90_hu91-1500, hu1-60_wt61-90_hu91-1500, hu1-30_wt31-75_hu76-1500 and wtABCDE) belong to this group, indicating a clear negative influence of wild-type sequence at this position on mRNA stability. The variant of this group with the longest half-life was the wt46-90 variant with a half-life of 2.7 hours. The variantwt1-60_hu61-75_wt76-1500 showed a half-life comparable to the wild-type variant with 2.2 h. Interestingly. The variant wt1-60_hu61-75_wt76-423_hu424-1500 exhibit an enhanced half-life of 4.4 h.

A)

	Schematic illustration of gag ORF	Nomenclature	Wild-type nucleotides	Humanized nucleotides
# 1	A B C D E	hu1-15_wt16-30_hu30-1500	16-30	1-15 & 30-1500
#2	A B C D E	hu1-30_wt31-45_hu46-1500	31-45	1-30 & 46-1500
#3	A B C D E	hu1-45_wt46-60_hu61-1500	46-60	1-45 & 61-1500
#4	A B C D E	hu1-60_wt61-75_hu76-1500	61-75	1-60 & 76-1500
#5	A B C D E	hu1-75_wt76-90_hu91-1500	76-90	1-75 & 91-1500
#6	A B C D E	hu1-15_wt16-90_hu91-1500	16-90	1-15 & 91-1500
#7	A B C D E	hu1-30_wt31-90_hu91-1500	31-90	1-30 & 91-1500
#8	A B C D E	hu1-45_wt46-90_hu91-1500	46-90	1-45 & 91-1500
#9	A B C D E	hu1-60_wt61-90_hu91-1500	61-90	1-60 & 91-1500
#10	A B C D E	hu1-30_wt31-75_hu76-1500	31-75	1-30 & 76-1500
#11	A B C D E	wt1-60_hu61-75_wt76-1500	1-60 & 76-1500	61-75
wt hu #12	A B C D E	wt1-60_hu61-75_wt76-423_hu424-1500	1-60 & 76-423	61-75 & 424-1500

B)

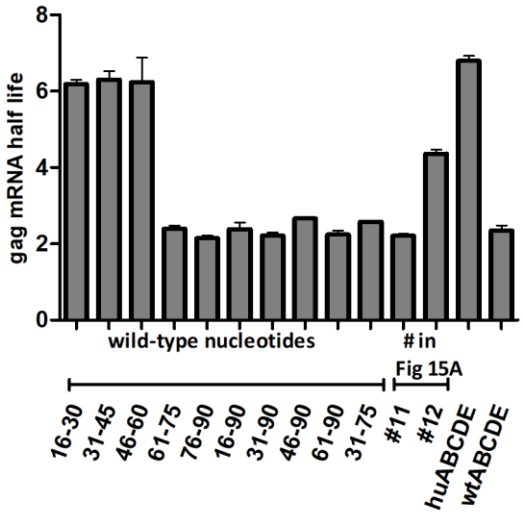

Figure E-15: mRNA stability analysis of selected partially humanized variants.
A) Schematic overview of the adapted variants and localization of humanized and wild-type nucleotides B) Half-lives of selected humanized variants. Half-lives of gag variants were determined after transfection of HEK293T cells and subsequent treatment of cells with the transcription inhibitor Actinomycin D. Gag mRNA levels were quantified by RT-qPCR analysis before Actinomycin D addition and after x h of treatment and used for half-life calculation. All experiments were performed in biological triplicates.

Taken together, it could be shown that the usage of wild-type codons in a specific region of the 5' part of *gag* significantly impairs mRNA stability. It became apparent that even a small part of 15 nucleotides is sufficient to decrease the Gag mRNA half-life from 6.8 hours (for fully humanized *gag* variant) to 2.4 h (hu1-60_wt61-75_hu76-1500) or 2.1 h (hu1-75_wt76-90_hu91-1500). These findings and the differences in transcription rate might together explain the more pronounced differences in total *gag* mRNA abundance.

E.11 Detection of potentially generated cryptic splicing variants

Besides altered transcriptional efficiency and mRNA stability, mRNA level and therefore also protein production can be influenced by splicing effects. All subgenomic reporter constructs were composed of a suitable promoter, followed by the 5'UTR of HIV-1, the *gag* open reading frame as well as the rev-responsive element. That means, all analyzed variants contain HIV-1's major splice donor (SD1), which is important for stabilization of the transcripts and necessary for Rev-dependent export [59], [60], [114] as well as the splice acceptor SA7 which is located in the 3'-part of the RRE. Although those two splicing elements usually do not interact with each other, i.e. there is no canonical SD1SD7 splice isoform known, the altered nucleotide sequence of humanized variants might contain motifs enabling such a splicing pattern or might even have generated additional splicing donor or acceptor sites. Therefore, the creation of cryptic splicing variants that would affect the amount of RNAs coding for full-length Gag had to be excluded. This was done by northern blot analysis as well as PCR analysis after reverse transcription of isolated mRNA. For northern blot analysis a probe was designed that specifically detects the *gag* mRNA and all potential splicing variants due to its complementarity to a short sequence within the 5'UTR upstream of the splice donor SD1. As an additional validation, PCR analysis was performed after RNA isolation and reverse transcription. Primers had been designed in a way to make sure that potential shorter splicing forms would dominate in PCR analysis because those forms would exhibit higher PCR efficiencies. Therefore, the forward primer was designed to bind upstream of the splice donor SD1 and the reverse primer to bind downstream of the splice acceptor SA7 (Figure E-16). For these experiments, the two reference variants huABCDE and wtABCDE, as well as the huBCDE variant were chosen, the latter as a representative for all variants with impaired expression having wild-type nucleotides at the 5' end of *gag*. As shown in Figure E-16B, northern blot analysis showed that no additional splicing forms were generated. In the cytoplasm as well as in the nucleus, only one distinct band was detected for all analyzed variants with a length of approximately 2300bp which represents the expected transcript length including the length of the poly-A-tail. As anticipated, *gag* mRNA was detectable in the nucleus independently of Rev for all variants. Further, the completely humanized variant huABCDE showed the strongest expression. Contrary, the three variants were all detected in the cytoplasm in the presence of Rev, whereas the variants huBCDE and wtABCDE showed very weak signals in the cytoplasm in the absence of Rev. Since the intensity of the detected transcripts was quite low, PCR analysis after RNA isolation and reverse transcription was performed as an additional validation that no cryptic splicing products were generated. As illustrated in Figure E-16C, all variants showed a single band of the expected size. As assumed, cDNA detection was successful in the nuclear fraction for all variants independent of the presence of Rev. In the cytoplasmic fraction, the variants huBCDE and wtABCDE were dominantly expressed in the presence of Rev. Taken together, these experiments showed that no cryptic splicing variants were generated by the humanization of *gag* and the impaired expression of the huBCDE is not affected by altered splicing.

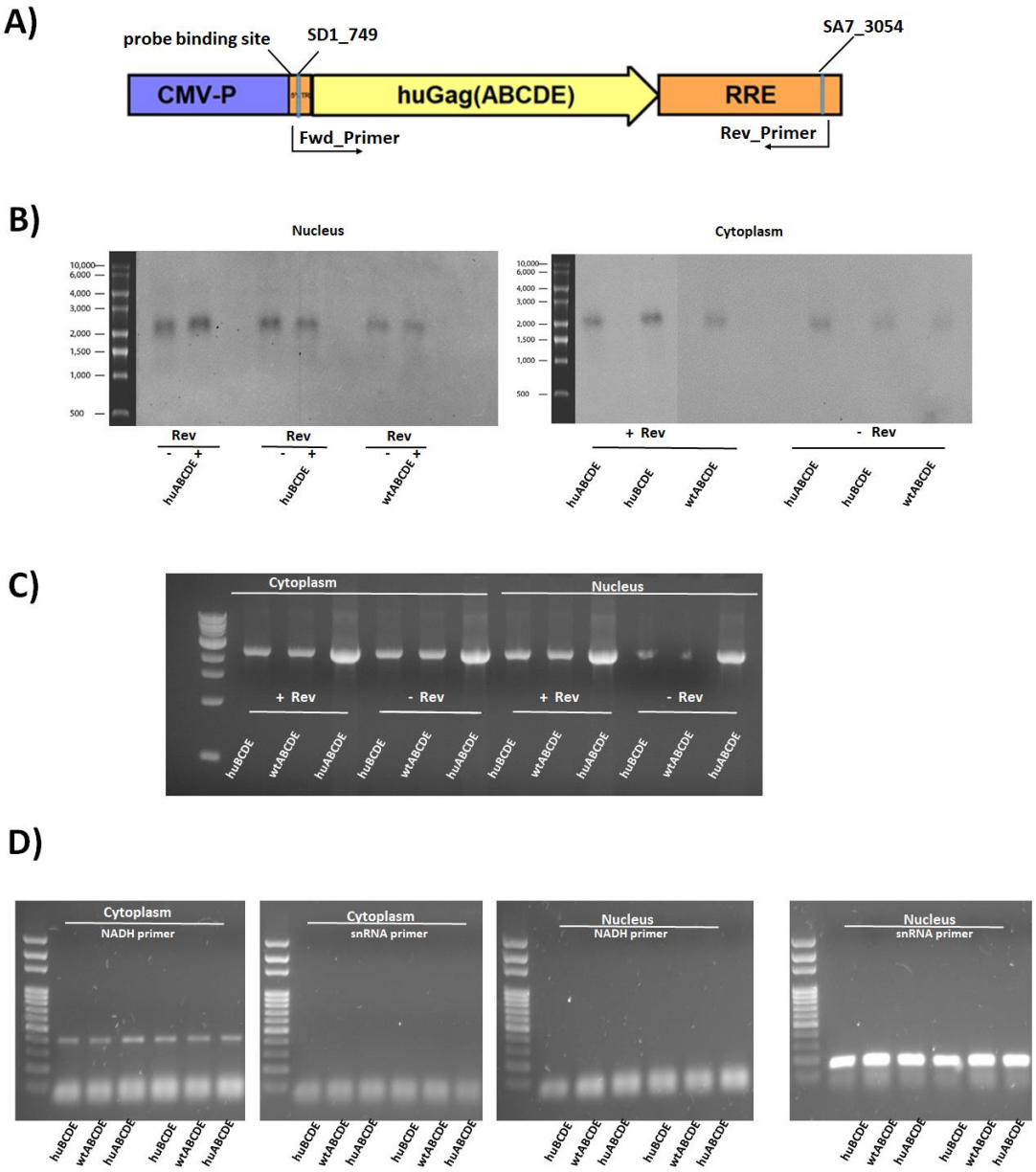

Figure E-16: Analysis of the splicing pattern of the variants huABCDE, wtABCDE and huBCDE
A) Schematic depiction of the subgenomic reporter variants highlighting the locations of the major splice donor SD1 and splice acceptor SA7 from HIV-1 as well as primer binding sites for analysis of splicing isoforms by PCR analysis after reverse transcription. B) Northern blot analysis of selected subgenomic variants by a specifically designed RNA probe. Because northern blot analysis was performed in different experiments and with different arrangement of the variants, different pictures from different blots were put together in order to visualize all relevant results in one picture C) PCR analysis after RNA isolation and reverse transcription of the selected subgenomic variants with the primers shown in A). D) control of the correct separation of cytoplasmic and nuclear RNA fractions. PCR reactions were performed with special primers that amplify only products from cytoplasmic (NADH) or nuclear (snRNA) fraction if separation was performed correctly.

E.12 Impact of the Splice Donor SD1 on *gag* expression

From previous studies it was already known that the presence of the splice donor SD1 can impact *gag* expression. Since all subgenomic variants therefore contained the SD1 as part of the 5'-UTR, it should be analyzed how far expression levels are influenced by the presence of the major splice

donor. To analyze the impact of SD1, three selected variants with mutated and therefore destroyed splicing consensus sequence were generated (Figure E-17B). Since humanization had the strongest beneficial effect on protein production on the 5'-3'-adapted variants, the three constructs huA_wtBCDE, huABC_wtDE and huABCDE (Figure E-17A) were chosen for this analysis as representatives of the 5'-3' set. For this, HEK293T cells were transfected with the three adapted *gag* variants and the *gag* protein amount in the supernatants of transfected cells after 48 hours was quantified by p24-ELISA. Transfection was performed in the presence as well as in the absence of Rev.

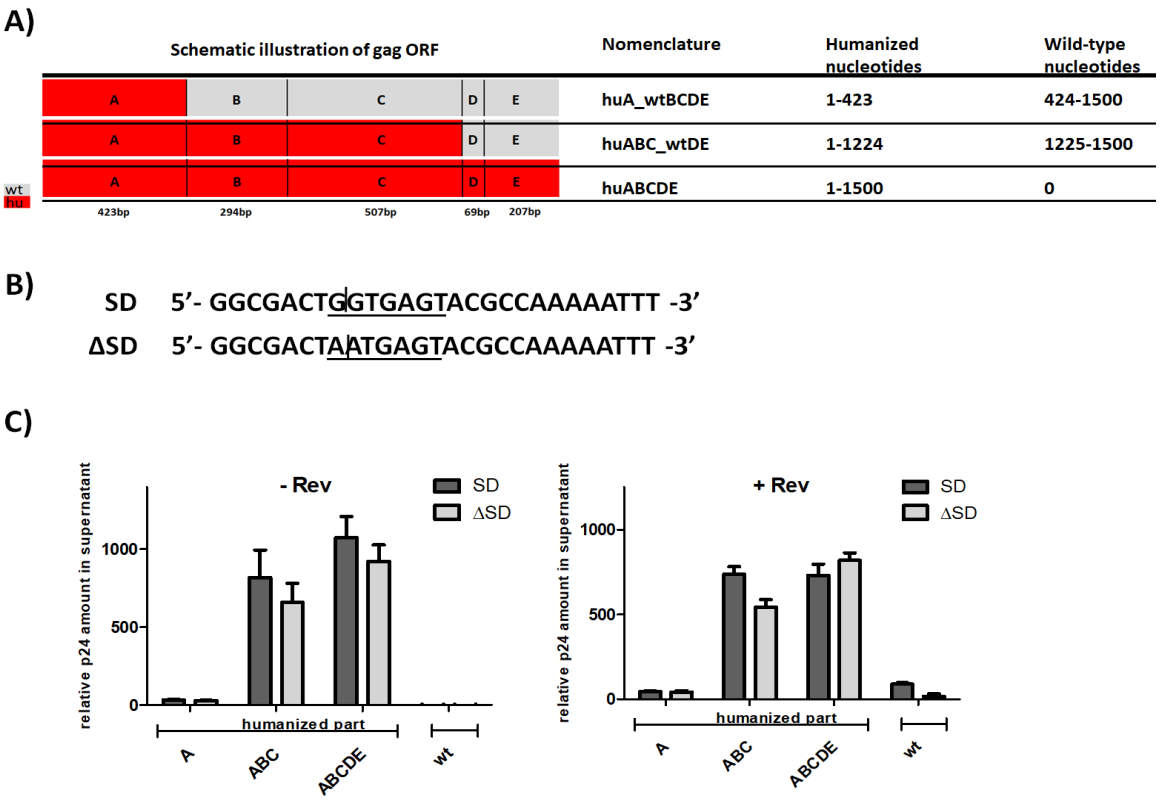

Figure E-17: Analysis of the impact of HIV-1's major splice donor SD1 on protein production
A) Schematic overview of the adapted variants and localization of humanized and wild-type nucleotides. B) Sequence context of the major splice donor SD1 within the 5'UTR as mutated and intact form. Mutation of the SD1 was performed for the representative variants huA_wtBCDE, huABC_wtDE, huABCDE and wt. C) Relative p24 expression in the supernatants of transfected HEK293T cells, in the absence (left) and presence (right) of Rev.

As shown in Figure E-17C, the presence or absence of the splice donor SD1 of HIV had a negligible influence on *gag* protein levels for all partially humanized variants of the 5'-3' set that were analyzed here. In the presence and absence of Rev, comparable p24 levels for all analyzed variants independently of the presence of an intact or mutated splice donor were seen. For none of the variants a significant difference between the SD and ΔSD variant was detectable. Further, as expected, the wild-type variant wtABCDE showed measurable p24 expression only in the presence of Rev. Surprisingly, the expression levels were enhanced for huABC_wtDE and huABCDE in the

absence of Rev compared to the expression levels in the presence of Rev. However, this difference was not statistically significant and might be an experimental variation.

E.13 Analysis of expression of partially humanized *gag* variants under control of the LTR promoter

All analyses described so far were performed with subgenomic *gag* reporter variants under the control of a heterologous CMV promoter. The CMV promoter used is well-known for its strong expression and was used in the experiments because it had been shown to lead to efficient *gag* protein production in previous studies [60], [112]. To analyze whether the observed expression patterns were dependent of the promoter choice, *gag* expression under control of the natural LTR promoter was analyzed. For this, the *gag* open reading frame was set under the control of HIV-1's LTR promoter together with the 5'UTR and the Rev-responsive element. In order to cover partially humanized variants representing all phenotypes described above, various open reading frames were chosen for the analysis. The six variants huA_wtBCDE, huABC_wtDE, huABCDE, wt1-15_hu16-1500, wt1-75_hu76-1500, and wtABCDE were selected because of their different expression levels and striking phenotypes. HEK293T cells were transfected with the adapted *gag* variants under control of

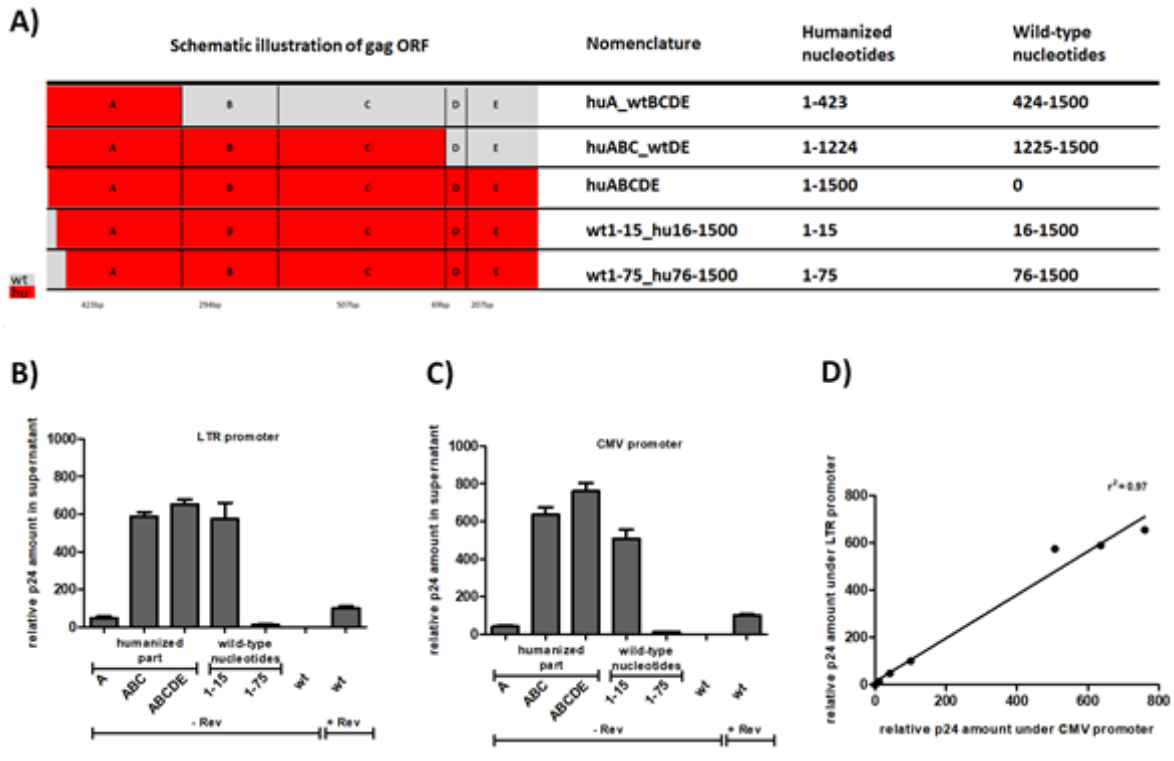

Figure E-18: Comparison of gag expression levels of adapted variants either under control of the CMV or the LTR promoter

A) Schematic overview of the adapted variants and localization of humanized and wild-type nucleotides. B) p24 expression levels of the selected humanized variants under the control of the LTR promoter. C) p24 expression levels of the selected humanized variants under the control of the CMV promoter of HIV-1. Data is collated from Figure E-4 and Figure E-8, etc. D) Correlation analysis of data from B) and C).

the LTR promoter and *gag* protein production was quantified as described before. For easier comparison, the p24 expression levels of the corresponding variants under control of the CMV promoter are also shown in .Figure E-18B. The experiments that had been performed to generate these data are described and already shown in Figure E-4 and Figure E-8.

Expression analysis of the selected variants under control of the LTR promoter of HIV-1 showed very similar results compared to the corresponding variants under control of the CMV promoter. The huA_wtBCDE variant showed the weakest *gag* expression with relative p24 levels of 48 % (LTR promoter) and 41 % (CMV promoter). The variants huABC_wtDE and huABCDE and wt1-15_hu16-1500 showed significantly enhanced expression levels with relative p24 amounts of 589 %, 654 % and 575 % (LTR promoter) or 636 %, 760 % and 507 % (CMV promoter), respectively. The negative impact of wild-type codon usage at the 5' end of *gag* is also true for variants under control of the LTR promoter. The variant wt1-75_hu76-1500 as well as the completely wild-type *gag* variant showed relative p24 levels slightly above the detection limit with values of 11 % and 0.02% (CMV promoter) and 14 % and 0.03 % (LTR promoter). The good comparability between the variants under control of the two promoters becomes even clearer by calculating the correlation between the expression levels. The result is depicted in Figure E-18D. The r^2 value of 0.97 shows a strong correlation and indicates that expression analysis of subgenomic reporter variants under control of the CMV promoter is a suitable model to analyze the impact of codon adaptation on protein production.

E.14 Analysis of the impact of codon adaptation on a quasi-lentiviral GFP reporter

In the previous experiments a positional as well as additive impact of codon adaptation of the HIV-1 *gag* gene on protein production was observed. To find out whether these effects are specific for HIV-1 *gag* or whether these findings can also be seen for other genes, a quasi-lentiviral EGFP reporter system was used. This system was already established in previous studies of our group for the gfp gene [60]. In this work EGFP was used instead of GFP because of the enhanced photo stability and fluorescence mediated by the exchange of threonine by serine at amino acid position 65. For analysis of the impact of codon choice on *egfp* expression, the desired *egfp* variants were cloned into the pcDNA3.1 expression vector and were flanked by the very 5'-UTR and RRE to ensure comparability between the results for *gag* and *egfp* In some cases (see below), a sequence from HIV-1 *gag* was placed in front of the *egfp* open reading frame. In all those cases, the *gag* sequence was used as a 5' untranslated region, downstream of HIV-1's 5'UTR and in front of the start codon of *egfp* (see Figure E-19).

Figure E-19: Schematic illustration of the egfp subgenomic reporter construct
A) All variants were expressed under control of a CMV promoter together with the 5'UTR of HIV-1 and the Rev-responsive element. B) Subgenomic egfp variant as in A with a partial gag sequence (red) as an additional 5'UTR.

To analyze how far the previous results can be transferred to the *egfp* reporter system, partially humanized *egfp* variants were generated. First, the *egfp* sequence was adapted either to HIV codon usage (hiv_*egfp*) or to human codon usage (hu_*egfp*). These two completely adapted variants were then used to generate partially adapted variants, analogous to the 5'-3' and 3'-5' set of *gag*. For this, the 720 bp long *egfp*-sequence was divided into five equal parts of 144 bp named A-E (Figure E-20A). Those variants were then transfected into HEK293T cells and *egfp* expression was quantified after 48 hours by flow cytometry.

Expression analysis of partially codon-adapted *egfp* variants showed remarkable differences compared to expression patterns of partially humanized *gag* variants. Although a stepwise increase in expression with longer humanized sequence parts was observed for the 5'-3'-adapted variants ((GFP)_huA_hivCDE, (GFP)_huAB_hivCDE, (GFP)_huABC_hivDE, (GFP)_huABCD_hivE) similar to *gag*, no positional effects were observed for the 3'-5'-set. In contrast to the 3'-5'-adapted *gag* variants, where no expression was observed for variants lacking humanization at the 5' end in the absence of Rev, 3'-5'-adapted *egfp* variants ((GFP)hivA_huBCDE, (GFP)_ hivAB_huCDE, (GFP)_ hivABC_huDE, (GFP)_ hivABCD_huE) were expressed even without the addition of Rev. So, for all variants, the number of humanized codons directly influenced the expression level. As expected, the strongest signals were detected for the completely humanized *egfp* variant hu_EGFP/ (GFP)_hu_ABCDE with relative MFI values of 1344 (without Rev) and 856 (with Rev). However, the variants (GFP)_huABCD_hivE and (GFP)_ hivA_huBCDE which both had 576 nucleotides with human codon usage, displayed significant differences between each other in the absence as well as presence of Rev, indicating that not the number of codon adapted nucleotides alone defines the level of *egfp* expression. Furthermore, 5'-3'- and 3'-5'-adapted variants benefitted similarly from the addition of Rev. Here, the variants with fewer humanized codons and therefore lower expression levels like (GFP)_ hivA_huBCDE and (GFP)_ hivABCD_huE showed the highest Rev-effect of partially humanized variants. The ratio between *egfp* expression in the presence and absence of Rev was 8.5 for (GFP)_ hivA_huBCDE and 5.5 for (GFP)_ hivABCD_huE. The Rev-effect for the wild-type *egfp* variant was 7.9.

A)

Schematic illustration of gfp ORF	Nomenclature	Humanized nucleotides	HIV adapted nucleotides
A B C D E	(GFP)_huA_hivBCDE	1-144	145-720
A B C D E	(GFP)_huAB_hivCDE	1-288	289-720
A B C D E	(GFP)_huABC_hivDE	1- 432	433-720
A B C D E	(GFP)_huABCD_hivE	1-576	577-720
A B C D E	(GFP)_huABCDE	1-720	0
A B C D E	(GFP)_hivA_huBCDE	145-720	1-144
A B C D E	(GFP)_hivAB_huCDE	289-720	1-288
A B C D E	(GFP)_hivABChutDE	433-720	1-432
A B C D E	(GFP)_hivABCD_huE	577-720	1-576
A B C D E	(GFP)_hivABCDE	0	1-720

hu
hiv

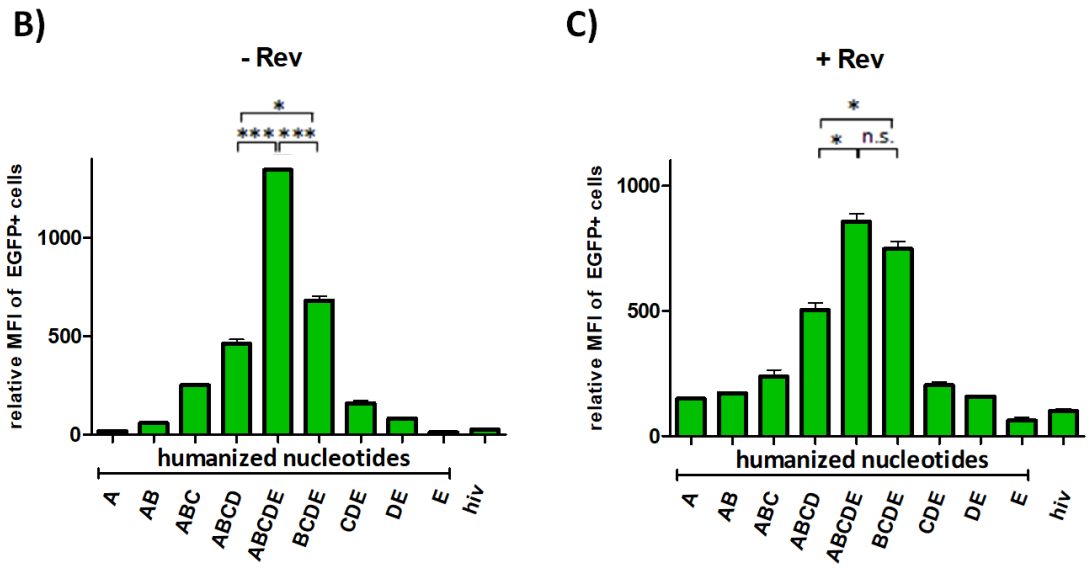

Figure E-20: Analysis of the impact of codon choice on codon-adapted egfp variants in a quasi-lentiviral expression system
A) Schematic overview of the adapted variants and localization of humanized and HIV-like egfp nucleotides. B) Quantification of egfp expression of partially adapted variants in the absence of Rev by flow cytometry. Single cells were gated and GFP positive cells were analyzed 48h post transfection C) Quantification of egfp expression of partially adapted variants in the presence of Rev by flow cytometry.

E.15 Influence of the identified inhibitory HIV-1 *gag* motif on *egfp* expression

In the previous experiments, an inhibitory motif within the 5' part of HIV-1 *gag* was identified that negatively regulated the *gag* expression in general and prevented expression in the absence of Rev (see E.6.3E.6). In the following experiments it was analyzed whether the inhibitory motif might function in a heterologous, dominant manner after transfer to the *egfp* expression system.

A)

Gag sequence (as 5'UTR)	gfp open reading frame	Nomenclature	hiv_gag nucleo-tides(from gag orf as 5'UTR)	hu_gag nucleotides	hiv_gfp nucleotides	hu_gfp nucleotides
	A B C D E	hiv_EGFP	0	0	1-720	0
	A B C D E	hu_EGFP	0	0	0	1-720
A	A B C D E	[Gag_hivA]_hu_EGFP	1-423	0	0	1-720
A	A B C D E	[Gag_huA]_hu_EGFP	0	1-423	0	1-720
A	A B C D E	[Gag_huA$_{hiv46-60}$]_hu_EGFP	1-45	46-423	0	1-720
A	A B C D E	[Gag_huA$_{hiv61-75}$]_hu_EGFP	46-60	1-45 & 61-423	0	1-720
A	A B C D E	[Gag_huA$_{hiv76-90}$]_hu_EGFP	61-75	1-60 & 76-423	0	1-720
	A B C D E	[Gag_hiv45-60]_hu_EGFP	1-45	0	0	1-720
	A B C D E	[Gag_hiv61-75]_hu_EGFP	46-60	0	0	1-720
hu hu hiv hiv	A B C D E	[Gag_hiv76-90]_hu_EGFP	61-75	0	0	1-720

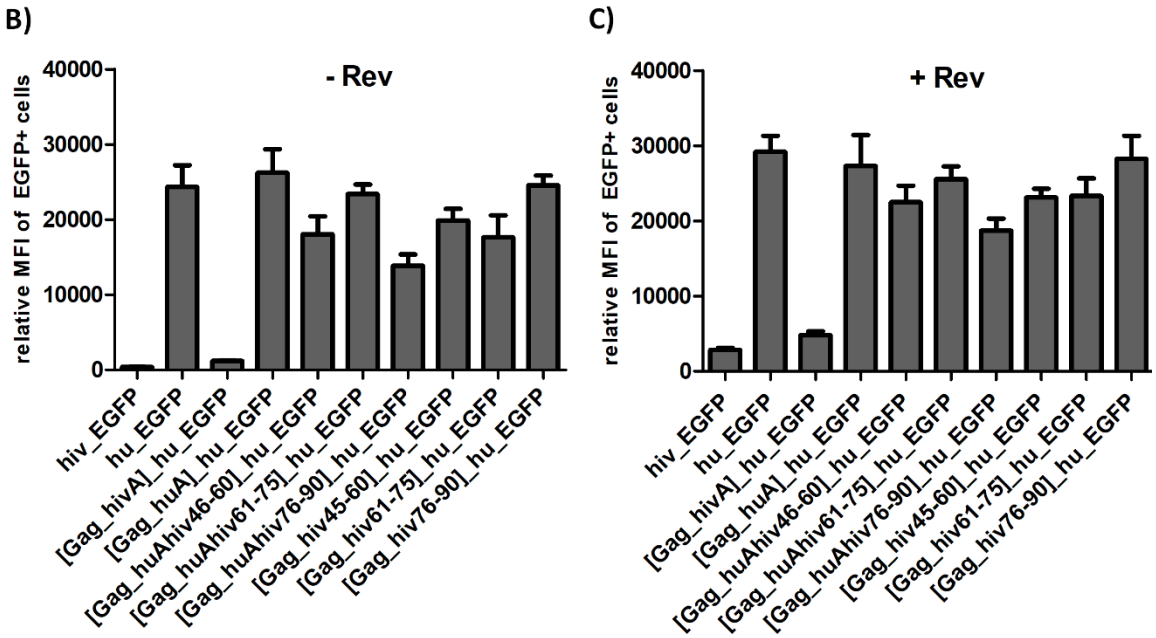

Figure E-21: Analysis of the impact of HIV-1 gag sequences as 5'UTR in the egfp expression system
A) Schematic overview of the adapted variants and localization of humanized and wild-type or HIV-like gag and egfp nucleotides. B) Quantification of egfp expression of partially adapted variants in the absence of Rev by flow cytometry. C) Quantification of egfp expression of partially adapted variants in the presence of Rev by flow cytometry.

Therefore, different *egfp* variants with various sequences from the 5' part of *gag* as 5'-UTR were generated. First, two variants were made with the complete humanized or wild-type A-part of *gag*

(423 bp) in front of the *egfp* open reading frame. Additionally, three variants were generated with either parts of the inhibitory motif (wild-type *gag* nucleotides 61-75 and 76-90) or a non-inhibitory control sequence (wild-type *gag* nucleotides 46-61) in the context of the humanized A part of *gag*. Lastly, the wild-type nucleotides 46-60, 61-75 and 76-90 were added on their own. In all cases, the HIV-1 *gag* sequences were added as an additional 5' untranslated region Figure E-21A. For this, the start codon of *gag* was mutated so that the original EGFP's start constituted the first ATG in the sequence. All variants were used for transfection of HEK293T cells and quantification of *egfp* expression was performed after 48 hours by flow cytometry.

Expression analysis showed that transfer of the function of the inhibitory motif from HIV-1 *gag* is limited. Indeed, hu_*egfp*'s expression was almost completely inhibited when the wild-type part A (all 423 bp) of HIV-1 *gag* was placed in front of the open reading frame in comparison to hu_*egfp* alone, but the inhibitory sequences (61-75 and 76-90) within the otherwise humanized *gag* A fragment showed no specific effect. The inhibitory motifs alone (15 nucleotides) neither showed a specific effect. Apart from the complete humanized A-part of *gag*, which did not influence expression, and the complete wild-type A-part of *gag*, which impaired expression, all variants showed unspecific and only slightly impaired *egfp* levels. Taken together, it was possible to transfer the inhibitory effect of the 5' part of *gag* with wild-type codon usage to the *egfp* system, but only when all 423 bp from wild-type part A were set in front of the *egfp* open reading frame. It remains to be seen whether shorter sequences would exhibit an inhibitory effect. Some of the lastly described experiments were performed by Sebastian Einhauser and can be reviewed in detail in his Bachelor thesis [113].

To sum it all up, it was observed that codon usage at the 5' part of *gag* influences not only general protein amounts but affects also the Rev-dependency. More specifically, a short motif was identified that mediates an inhibitory influence on *gag* gene expression by influencing *gag* mRNA amounts. The impaired levels of mRNA could be explained as a negative effect of hiv wild-type sequence in this special area on mRNA stability as well as transcriptional-rates. The inhibitory effect of this motif was only to some degree present in the additional reporter gene *egfp*. An impaired gene expression was only observed when the motif was embedded within the first 423bp of the hiv wild-type nucleotides of *gag*.

F Discussion

The degeneracy of the genetic code and therefore the possibility to encode a specific amino acid by more than one codon plays a central role in biology as well as in biotechnology. Organisms depend strongly on the utilization of the correct codons, since suboptimal codon choice can result in inefficient gene expression and also in impaired or deregulated replication of organisms [111], [128], [148]. On the other hand, this biological mechanism allows the manipulation of codon choice for a broad spectrum of biotechnological or medical applications. For instance, during the last years, it was shown several times that changing the codon usage can result in attenuated viruses that may potentially be used as vaccines. Nogales and colleagues showed in 2014 that Influenza A virus can be attenuated by codon deoptimization. They generated different variants containing deoptimized synonymous mutations in coding regions comprising the entire nonstructural gene of influenza A virus. They observed that the fitness of the variants was reduced, while immunogenicity was retained [110]. Additionally, several other studies with virus attenuated by changes in codon usage were performed [104], [115], [126], [134], [135]. Besides codon deoptimization in order to achieve attenuation, codon optimization is a common way to enhance protein production and might also be used in this direction for the generation of potential vaccines or antibodies. To use these techniques possibly also for the fight against the human immunodeficiency virus, an exact knowledge about the underlying molecular mechanisms of how codon choice influences viral gene expression and viral replication are required. Different studies have been performed to address this question. Takata et al. performed a global synonymous mutagenesis screen and identified *cis*-acting RNA elements that regulate HIV-1 replication. A key mechanism that affected viral replication in this context were changes in splicing events [151]. Since HIV replication is a highly structured process where chronologically regulated gene expression is inevitable for proper replication, in the present thesis a simplified system was used. By transfecting HEK293T cells with subgenomic *gag* reporter variants the direct influence of codon usage on viral gene expression was analyzed, independently of other viral processes like infection, uncoating, reverse transcription or budding. Intervention in the regulated protein production of HIV could easily lead to the interruption of the replication cycle, which would make it difficult to study the exact effect of codon usage on viral gene expression. Another advantage of the simplified model system of this work is the independency of immunological aspects like cell- or antibody-mediated immune response, which could interfere with the direct effect of humanization on gene expression, although this could become an interesting aspect in further experiments. Takata et al. recently showed the important role of immunological effects induced by the usage of synonymous codons, since CG dinucleotide suppression enables antiviral defense and therefore influences HIV-1 replication. [152].

F.1 Generation of partially humanized variants and analysis of gene expression

F.1.1 Generation and analysis of partially humanized *gag* variants

In order to analyze possible positional and additive effects of codon adaptation, partially humanized variants were generated that differ systematically in their amount and position of humanized codons. Starting from a wild-type *gag* gene (wtABCDE) and a completely humanized *gag* gene huABCDE, two subsets of partially humanized variants were generated by fusion PCR (Figure E-1). For the 5'-3'-set, humanization started at the 5' end of *gag* and was enlarged into 3' direction. For the 3'-5-set, humanization was performed in the opposite direction. The utilization of synonymous codons necessarily goes along with the alteration in certain sequence features. In the last years, different groups showed the importance of specific sequence features like dinucleotide frequency on gene expression or viral replication. For instance, it was shown by different research groups that CpG/UpA dinucleotide frequencies affect viral replication [111], [141]. Therefore, the effect of humanization on dinucleotide frequency, GC-content and the codon adaptation index (CAI) was analyzed for the 5'-3'- and 3'-5'-adapted variants. As expected, a gradual increase or decrease in these parameters for the partially humanized variants was observed, according to the A-rich HIV- and GC-rich human codon usage (Figure E-2). GC- content, CpG dinucleotide frequency rose with increasing number of humanized nucleotides, whereas TpA dinucleotide frequency decreased with increasing humanization. Since the codon adaptation index (CAI) was used as the crucial parameter for the optimization process, it logically rose with increasing humanization. Unexpectedly, these bioinformatics analyses could only predict expression data for one subset (5'-3') and did not reflect the impaired and Rev-dependent expression of *gag* for 3'-5'-adapted variants. Further, the differences between the single Gag parts (A-E) were too small to explain the observed importance of the A-part of Gag (Figure E-3). These findings indicate that sequence features like GC-content, dinucleotide frequency and codon adaptation index does not explain the effects on protein production alone. Gag expression differed significantly between the two main subsets (5'-3' and 3'-5') of partially humanized variants. For the 5'-3' part, a continuous positive effect of humanization on *gag* expression (on protein and mRNA level) was observed. Further, every additional part of *gag* that was humanized resulted in a stepwise increased *gag* expression, until the maximum for the fully adapted version was reached, which is in accordance with the bioinformatic analysis. Additionally, all 5'-3'-adapted variants were already expressed in the absence of Rev (Figure E-4). Further, variants with lower numbers of humanized codons and therefore lower expression levels in the absence of Rev (huA and huAB) benefitted stronger from the addition of Rev than variants with a higher number of humanized codons and consequently higher expression levels. For those less humanized variants, the Rev-Crm1-mediated export pathway for *gag* mRNAs plays a crucial role. However, this export pathway seems to be less efficient for stronger adapted variants because addition of Rev had nearly

no influence on the expression of those variants. It might be that *gag* transcripts are exported out of the nucleus, maybe because of higher binding efficiency to the additional export factor and therefore only a negligible amount of *gag* mRNA remains for the Rev-Crm1 pathway. Surprisingly, the finding of a stepwise increase in expression by expansion of the sequence part with human codon usage was only true for the 5'-3'- but not for the 3'-5'-set (Figure E-6). For the variants that lack 5' humanization, nearly no expression was detectable in the absence of Rev. Furthermore, for these variants it was irrelevant how many codons were humanized. As long as the A-part of *gag* was encoded by wt codons, expression could not be enhanced by additional humanization, i.e. even the variant huBCDE that contains nearly 72% of humanized codons but acts like the wild-type variant. Accordingly, the correlation between amounts of humanized codons and *gag* expression which was observed for the 5'-3'-set was not true for the 3'-5'-set. These findings support the theory that the A-part of HIV *gag* might contain a potential inhibitory motif, which negatively regulates *gag* gene expression and does not allow *gag* expression in the absence of Rev. The phenomenon of the function of the inhibitory motif seems to be superordinate compared to the other effects of codon adaptation and prevents their impact on gene expression. Co-transfection of Rev led to *gag* expression comparable to wild-type *gag* expression, indicating that the export pathway mediated by Rev is not affected by the number of humanized codons if humanization is not present at the 5' part of *gag*. The particular effects of Rev could be reversed by the addition of Leptomycin B (LMB) (Figure E-5). LMB inhibits the export of proteins containing a nuclear export signal by by competing for binding to CRM1. Since Crm1 also mediates the export of several cellular molecules, a general negative effect of LMB on protein production was observed, likely because of an unspecific toxic effect on the transfected cells. For example, a reduction of the factor 1.9 was observed for the completely humanized variant huABCDE.

F.1.2 Generation and analysis of partially humanized *egfp* variants

It was also of interest whether the findings regarding stepwise codon adaptation were specific for the *gag* reporter gene used or whether the effects are similar for a different reporter gene. Therefore, *egfp* was used in the quasi-lentiviral expression system by altering its codon usage to that of HIV genes for further analysis. First, it was tested how stepwise codon adaptation influences gene expression for *egfp* in general. For this, *egfp* was divided into five parts of the same size and adapted stepwise to human or HIV codon usage. It was found that the expression levels for 5'-3' *egfp* adaptation were comparable to *gag* expression data for 5'-3' adaptation. Again, a stepwise increase in humanization enhanced protein production systematically. Starting with a Rev-independent but relatively weak expression of (GFP)_huA_wtBCDE, the MFI was enhanced with increasing numbers of humanized nucleotides. The maximum was reached for the fully humanized *egfp* variant. Contrary to the findings of the necessity of 5' adaptation of *gag* to human codon usage, *egfp* variants with HIV-like codon usage at the 5' part were also expressed in the absence of Rev. In addition, a stepwise decrease in expression by reduction of humanization was observed. These findings strongly support

the theory of an HIV-*gag*-specific phenomenon and contradicts the possibility of a general importance of humanization at the 5' end.

F.2 Localization and examination of the inhibitory motif within the 5' part of HIV-1 *gag*

In order to find the precise localization of the potential inhibitory motif, various additional partially humanized variants were generated (see E.6). Firstly, the 423 bp of the *gag* open reading frame (part A) of HIV-1 were divided into five additional parts with stepwise increase in the number of humanized nucleotides in a 3'-5'-direction. Expression analysis of these constructs showed that the inhibitory motif seems to lie within the first 100 nucleotides of *gag* (Figure E-7). Therefore, further partially humanized variants were generated and their expression was analyzed. It turned out that wild-type codon usage in the region of nucleotides 61-75 led to a nearly complete loss in *gag* expression, whereas a fully humanized *gag* variant with wild-type codon usage between nucleotides 46 and 60 showed no loss in expression. To confirm the findings regarding the position and function of the inhibitory motif, also variants with inverted codon usage were generated. For this, for the variant huBCDE, which showed no *gag* expression in the absence of Rev, the nucleotides 61-75 were changed from wild-type codons to humanized codons. This resulted in a partially abolition of the inhibitory effect by elimination of the potential inhibitory motif (Figure E-10).These findings indicate that a small number of changed nucleotides at this certain position within the 5'UTR of *gag* play a crucial role in gene expression and a change of these nucleotides alters protein levels dramatically. It would be thinkable that the identified motif might mediate e.g. an interaction between the transcript and an RNA binding protein. Dependent of the hypothetical RNA-binding protein this could lead to altered RNA stability or export, which would directly influence protein production. Further, the questions remain whether the effect of the inhibitory motif is specific for *gag* and to what extend the sequence context influences the observed inhibitory effect of the motif.

For further characterization of effect of the identified motif, the wild-type nucleotides as well as other 15 nucleotide-long wild-type sequences either alone or in context of the A-part of Gag, was transplanted as a 5'UTR in front of a humanized *egfp* open reading frame. Expression analysis of those variants showed that the wild-type A-part, but not a humanized A-part, led to impaired expression of the humanized *egfp*, which shows that the negative effect observed for *gag* expression can be dominantly transferred to *egfp*. However, the additional variants with either the potential motif alone, or in the context of a humanized A-part of *gag*, did not lead to a comparably impaired expression as observed for the *gag* reporter gene. An obvious reason for the unsuccessful transfer of the inhibitory effect from *gag* to *egfp* is the importance of a broader sequence context. Only if the

inhibitory motif is surrounded by further wild-type nucleotides, an impaired gene expression is observed. Whether this effect is mediated in this case by the inhibitory motif, or the additional wild-type nucleotides, is unknown. A molecular explanation of this phenomenon might be differences in co-transcriptional mRNA folding of these variants (see F.3.3).

F.3 Molecular mechanisms that contribute to altered expression of partially codon-adapted genes

Besides characterization and localization of the inhibitory motif, possible molecular mechanisms that lead to the observed effects on *gag* expression should be determined Regarding the effects of synonymous codon usage on gene expression and their underlying mechanisms two phenomena must be distinguished. The first aspect is a general dependency of the protein production from the number of humanized codons. Accordingly, humanization of the *egfp* open reading frame lead to enhanced expression levels. This was true for the 5'-3' as well as for the 3'-5'-adapted *egfp* variants. The same effect was observed for 5'-3'-adapted HIV *gag* variants. An exception to this rule was detected for *gag* variants that lack humanization at the 5'-part which seems to represent the second aspect, namely the inhibitory impact of a short sequence of HIV wild-type codons on gene expression. An aspect that seems to be a special feature of HIV *gag.*

F.3.1 Molecular mechanisms that contribute to the general dependency of protein production from the amount of humanized codons

F.3.1.1 Influence of translation on altered gene expression

Different studies regarding the impact of synonymous codon on gene expression showed a clear connection between RNA and protein levels [59], [60], [101], [153] Therefor, RT-qPCR analyses were performed to determine the *gag* mRNA expression for the partially humanized variants. Quantification of mRNA expression showed relative levels comparable to p24 amounts for the 5'-3'- as well as the 3'-5'-set, indicating a major importance RNA levels in the context of codon adaptation. These findings already point to a minor effect of synonymous codon usage on translational effects. The influence of codon adaptation on translation has been under debate for years. One common theory is that synonymous mutations influence protein abundance by changes in translation efficiency [154]–[156], a fact which may hold for many cases in prokaryotes and is often implied as affecting heterologous gene expression e. g in. *E. coli*. However, the influence on translation by itself is not enough to account for effects of synonymous codon usage in higher organisms. Chen *et al*. quantified the impact of synonymous codons on mRNA level by analyzing over 3000 GFP variants in yeast. They could show a positive correlation between mRNA levels and codon usage bias, which points to a direct effect of synonymous mutations on transcript concentration [99]. Further studies point into the same direction. Kholiswa *et al*. showed in 2008 that transfection of Jurkat cells with

codon-optimized *gag* mRNAs only led to small increases in *gag* production over wildtype mRNAs, whereas transfection of optimized DNA resulted in a very large enhancement of expression [100]. Furthermore, also our group could show that codon-usage mediated inhibition of HIV-1 *gag* expression in mammalian cells occurs independently of translation [153]. Additionally, in a recent thesis of our group the influence of codon adaptation on translation was analyzed in detail. For this a Modified Vaccinia Virus Ankara (MVA) T7 based expression system was used. Selected partially humanized *gag* variants with remarkably impaired or enhanced expression levels were cloned into a pT7 expression vector, where the *gag* gene was expressed from a T7 promoter. Since HEK293T cells do not have a T7 RNA Polymerase, expression of *gag* only takes place when a T7 polymerase is added to the system. This was done by infection of transfected cells with the MVA-T7. Since poxviral vectors replicate exclusively in the cytoplasm, an expression system independent from nuclear processes could thus be used, which allowed direct monitoring of the impact of codon adaptation on translation. The analysis of the variants hu_ABCDE, wtABCDE, huA_wtBCDE, wtA_huBCDE and wtABCD_huE which cover all types of partially humanized variants with remarkable phenotypes differed clearly between each other in the subgenomic *gag* reporter system. Interestingly, in the MVA-T7 expression system none of the variants showed significant differences in their expression. In the presence, as well as in the absence of Rev, all variants showed comparable expression levels to each other [113]. These results confirm the theory that codon adaptation primarily influences gene expression on RNA level instead of translation.

F.3.1.2 Influence of transcriptional efficiency on gene expression

The process of RNA surveillance itself is a complex topic in which several processes are involved [157], [158]. Firstly, the export of mRNAs into the cytoplasm is a crucial point for correct gene expression. But, in all analyses, the cytoplasmic RNA levels correlated strongly with the nuclear levels. As shown e. g. in the northern blot analysis (Figure G-1), the enhanced *gag* mRNA expression was observed in the nuclear as well as in the cytoplasmic RNA fraction. Therefore, export effects seemed to play a minor role for the altered expression level of the partially humanized variants. Two other relevant factors, which influence the number of transcripts, are transcriptional activity as well as mRNA stability. To measure the transcriptional activity, nuclear run-on assays were performed. For the two main sets (5'-3' and 3'-5') it was observed that codon usage indeed effects the transcriptional-rate, but only when the 5' part of *gag* was humanized. The variants huA_wtBCDE and huAB_wtCDE showed comparable transcriptional-rates to wtABCDE, indicating that transcriptional efficiency was not influenced for these variants. The constructs huABC_wtDE, huABCD_wtE and huABCDE showed enhanced transcriptional-rates, which points to a positive effect of codon adaptation on transcriptional activity. For the 3'-5' adapted variants, where the 5' part of Gag consists of wild-type codons, further humanization of the remaining part had no effect on the transcription rate. However, the differences between the transcriptional-rate of the adapted variants, especially between wtABCDE and huABCDE were clearly lower (factor ≈2) than the differences in overall mRNA

expression (factor ≈7), indicating that codon adaptation influences transcriptional activity, but further processes must be involved

F.3.1.3 Influence of mRNA stability on altered gene expression

Different studies could show a connection between codon optimality and mRNA stability. For instance, it was shown in 2015 by Presnyak and colleagues that codon optimality is a major determinant of mRNA stability [Ref]. By genome-wide RNA decay analyses, they revealed that stable mRNAs are enriched in optimal codons, whereas unstable mRNAs contain predominantly non-optimal codons. The substitution of optimal codons by suboptimal ones resulted in a clear destabilization of mRNA. Contrary, the converse substitution significantly increased stability [159]. Moreover, Harigaya *et al*. examined the association between codon optimality and mRNA stability in the fission yeast *Schizosaccharomyces pombe*. They observed a genome-wide association between codon optimality and mRNA stability, pointing to an evolutionary conservation of the phenomenon [160]. These findings are in accordance with the result for the constructs analyzed in this thesis regarding mRNA stability. To address this issue, Actinomycin D, an inhibitor of the RNA polymerase was used for mRNA half-life analysis. For the 5'-3'-adapted variants, it was observed that additionally humanized parts of Gag contribute positively to mRNA stability, whereas no effect was observed for the 3'-5'-adapted variants. Again, the data suggests a beneficial effect of humanization under the precondition of humanization at the 5'-end of *gag*. A possible explanation for the general improvement of mRNA stability by humanization is the rising content of CG base pairs and thereby increasing thermodynamic stability. However, since the half lives of all 3'-5'-adapted variants were unaffected regardless of the number of humanized codons, the underlying mechanisms seems to be more complex.

F.3.2 Molecular mechanisms that contribute to the function of the inhibitory motif on protein production

In general, the inhibitory effect could be mediated by one of different cellular processes or, more likely, by an interplay of several mechanisms. For instance, it would be possible that changes in codon usage could affect processes like mRNA export, transcription efficiency, RNA stability or even translation. Experimental data from this thesis as well as earlier studies showed a direct connection between RNA level and protein production in the context of synonymous codons (see E.8 and [153]). Because of that, also the set of variants with a small part of wt codon usage at the 5' part in a humanized *gag* version were analyzed by nuclear run on assays. It became apparent that the identified inhibitory motif negatively influences the transcription rate, since all variants that contain wild-type codon usage surrounding the area of 61-75 exhibits transcriptional activity comparable to wt_ABCDE (see Figure E-13). A possible explanation could be a changed mRNA secondary structure,

which allows better accessibility for proteins involved in transcription elongation. For this, it would be of interest to identify possible interaction partners of the *gag* mRNA in future experiments.

Besides transcription rate, the stability of mRNAs is a second major aspect that influences RNA and therefore also protein levels. So, mRNA stability tests were also perfomed with the variants with small parts of wt-codon-usage at different positions in the 5' part of hu*gag* that were generated for localization of the inhibitory motif. Stability analysis of those variants showed that only 15 wt nucleotides among an otherwise completely humanized *gag* gene could antagonize the positive effect of humanization of the rest of the gene. It was observed that all variants that in some way contain nucleotides from the inhibitory motif showed mRNA stability comparable to wtABCDE, regardless of the codon choice of the rest of the gene. A reason for this phenomenon could be a change in the mRNA secondary structure, which probably make those transcripts more accessible for RNase-dependent degradation. Further, an interaction with a variety of proteins involved in different RNA degradation pathways would be possible. Again, an altered secondary structure could influence these pathways by making the transcripts more or less accessible for the potential interacting proteins.

F.3.3 Influence of changed mRNA secondary structures on altered gene expression

According to the theory above, folding of mRNAs was predicted with the mfold algorithm, a tool which allows the prediction of possible RNA folding forms according to their thermodynamic features [161]–[164]. For all analyses, the 5'-UTR and the first 95 nucleotides were analyzed in order to focus on that part of *gag* that was found as decisive for gene expression. The results are shown in Figure F-1. As expected, the variant wtABCDE showed the least stable form according to the mfold prediction with the highest ΔG-value of the variants assessed. The larger number of A-nucleotides results in more open loops which contributes negatively to the thermodynamical structure (see Figure F-1B). Conversely, the thermodynamically most stable form is the completely humanized variant huABCDE with a free energy value of -65.34 kcal/mol. In line with the enhanced GC content, the predicted structure shows only few regions with open loops. Interestingly, the ΔG value is remarkably reduced for the variant hu1-60_wt61-75_hu76-1500, since the introduction of 15 wild-type nucleotides at this position resulted in the generation of an additional open loop structure between the nucleotides 150 and 184 in Figure F-1D. The ΔG value was clearly lower compared to the completely humanized *gag* variant (ΔG = -54.87 kcal/mol). In contrast, this was not the case for the variant hu1-45_wt46-60_hu61-1500. Despite the same number of wild-type codons the open loop structure is not present in the predicted secondary structure (Figure F-1C), which led to a similar ΔG value of -62.87 kcal/mol in comparison to huABCDE. These results are in accordance with the expression data on protein as well as RNA level, since a thermodynamically less stable structure could result in altered interactions with proteins that might regulate RNA degradation processes.

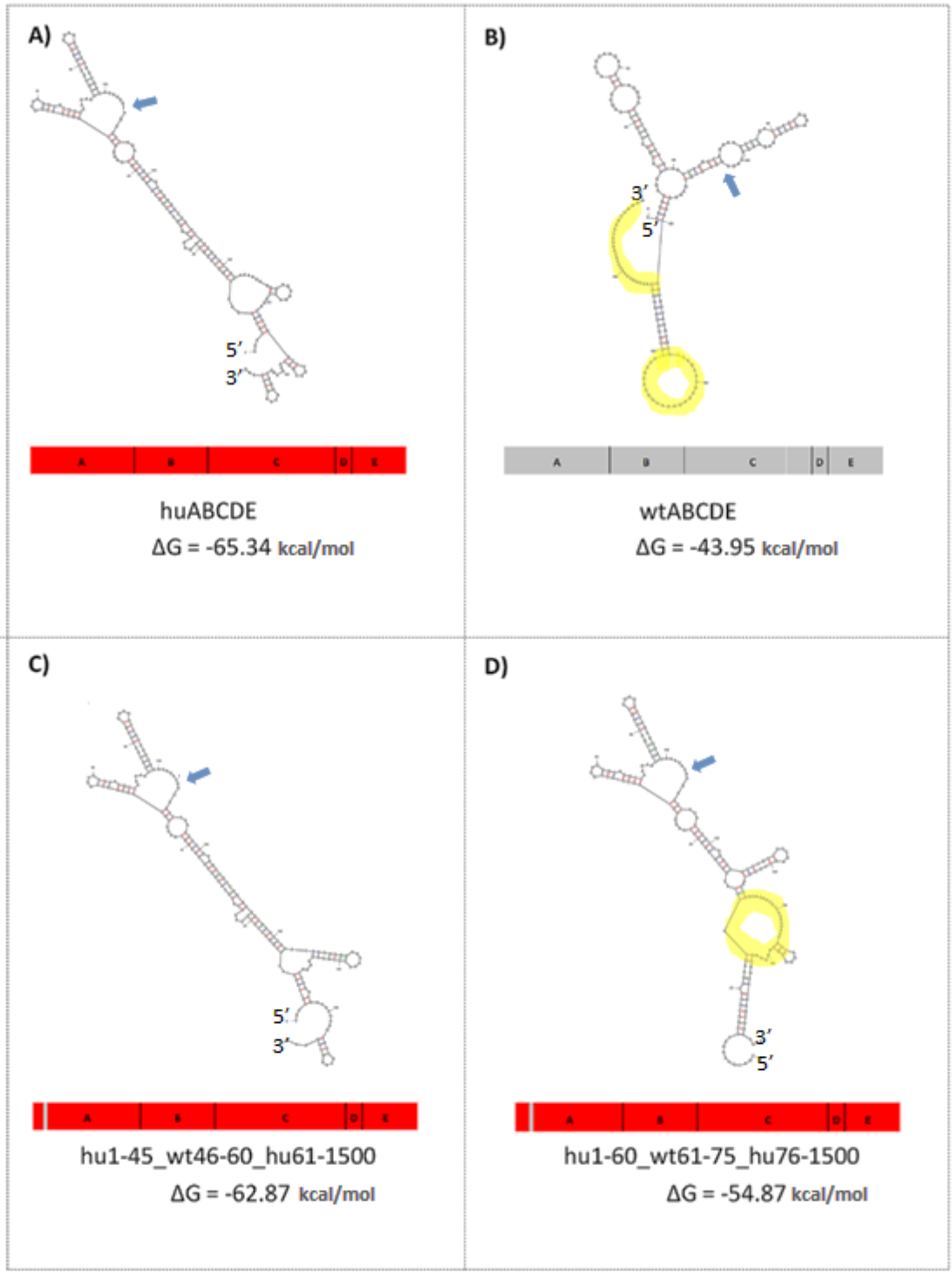

Figure F-1: Prediction of mRNA secondary structures for selected partially humanized variants, based on the mfold algorithm.
The calculated free energy and a schematic illustration of the particular variants are shown. The start codon for each variant is marked with a blue arrow. Also, the 5' and 3' ends of the sequence are indicated. The relevant generated open loops that influence thermodynamical stability are depicted in yellow

As mentioned above, the RNA surveillance machinery is a complex system encompassing several proteins that interact with the RNA molecule like Tap, Adar or Dcp2. To further address and test this

hypothesis, RNA-protein interaction analyses like an RNA pull-down assay could be performed in future experiments.

Interestingly, mRNA secondary structure predictions of *egfp* variants along with their 5'-UTRs, harboring the *gag* sequence parts where applicable, showed indeed a crucial difference to those that were performed for HIV-1 *gag*. The negative effect on thermodynamic stability of wild-type

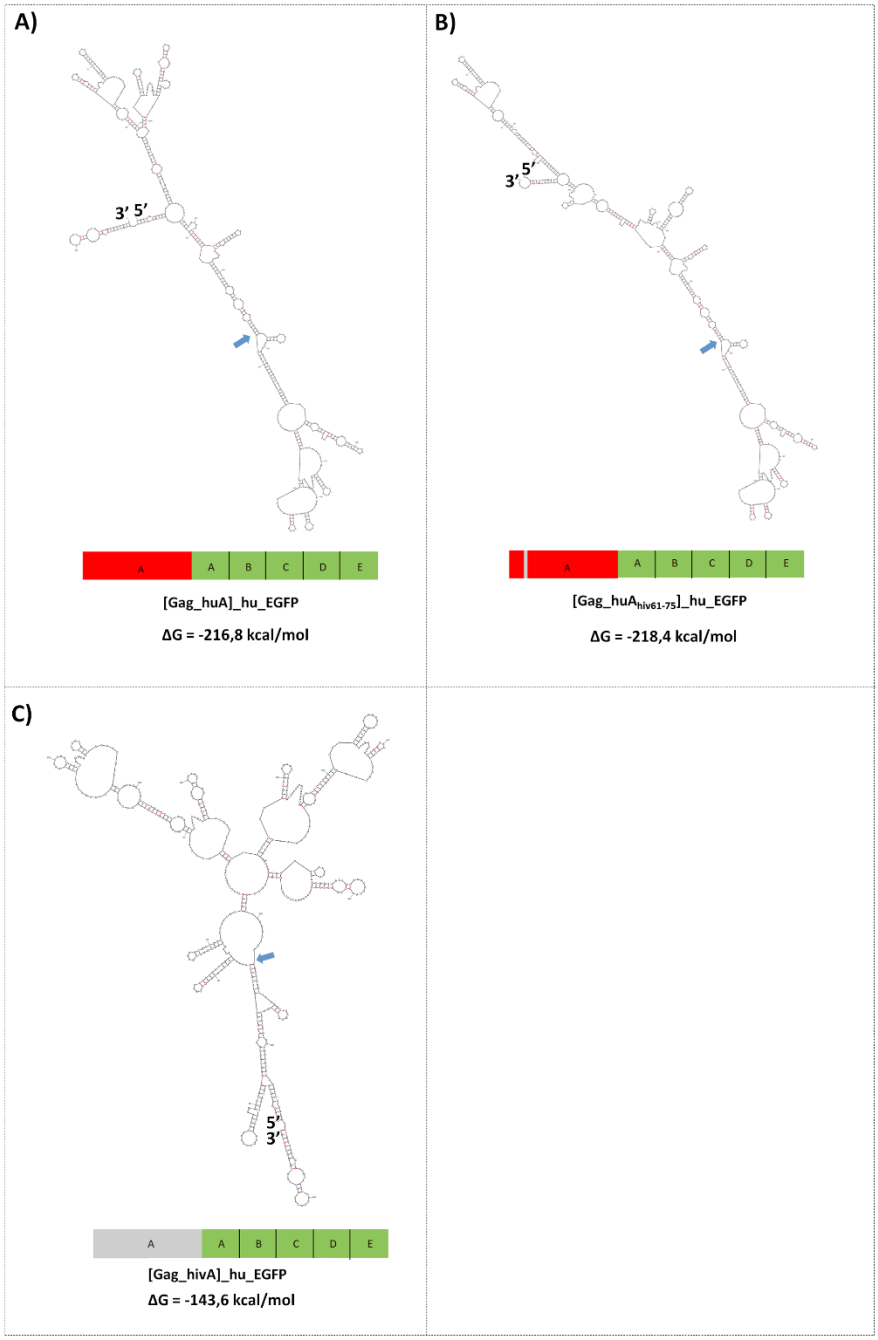

Figure F-2: Prediction of mRNA secondary structures for selected partially humanized egfp variants generated by the mfold algorithm.
The calculated free energy and a schematic illustration of the particular variants are shown. The start codon for each variant is marked with a blue arrow. Also the 5′ and 3′ ends of the sequence are indicated

nucleotides 61 to 75 of HIV-1 was not true for secondary structure predictions of *egfp*. As shown below, the wild-type A part led to a dramatically impaired stability when it was set in front of the *egfp* open reading frame compared to the humanized A part (ΔG = -143.6 kcal/mol vs ΔG = -216.8 kcal/mol). However, this negative impact on stability was not seen if the humanized A-part contains wild-type nucleotides within the positions 61-75 (Gag_huA$_{wt61-75}$_hu_EGFP). Compared to the ∆G value of -216.8 kcal/mol for Gag_huA_hu_EGFP, the *gag* wild-type nucleotides 61-75 within the A-part of *gag* led indeed only to a marginal difference in thermodynamic stability (ΔG = -218.4 kcal/mol). This could be a reason why the negative effect on expression of the inhibitory motif was only seen for the *gag* variants but not for the *egfp* variants.

F.3.4 Influence of splicing on altered gene expression

Besides influencing RNA stability and efficiency, splicing is a key determinant in regulation of transcript amounts. So, splicing is a fundamental process in RNA surveillance and also several studies on synonymous codon usage showed a connection between splicing effects and protein production. Lately, Takata *et al*. showed that the usage of synonymous codons could lead to perturbed RNA splicing, which resulted in defect replication of HIV. In their study, they analyzed sixteen HIV proviral mutants with clusters of synonymous codons, resulting partly in replication defect mutants. One group of those replication incompetent proviruses showed clearly perturbed RNA splicing [151]. Additionally, Mueller *et al*. showed in 2014 that RNA structure restricts HIV-1 splicing at the major splice donor site and that splicing directly influences viral fitness [165]. Although in this present thesis no replication competent virus was used, it had to be examined to what extent splicing effects could influence the eventual *gag* production in the experiments, especially because the subgenomic reporter constructs contained the major splice donor (SD1) in the 5'-UTR and the weak splice acceptor SA7 in the sequence of the RRE (see Figure E-16). To investigate the effect of the major splice donor, variants with mutated SD1 consensus sequence were generated to inactivate the splice donor. For this, point mutations were introduced into selected variants to destroy the splice donor sequence. Expression analysis of these variants showed that the enhanced protein production of the 5'-3' partially adapted variants was independent of the presence or absence of the splice donor, since the p24 expression data for the analyzed variants with or without the major splice donor correlated very strongly. This is in accordance with previous findings of our group that showed that codon-optimized *gag* variants showed enhanced expression levels independently of the SD1 [59].

A second aspect that had to be excluded is the potential generation of cryptic splicing sites. Since the usage of synonymous codons always goes along with an altered nucleotides sequence, there is the potential risk of generating additional splice acceptor or splice donor sites that could interact with each other or with the present SD1 or SA7 sites. Generation of cryptic splicing products could significantly alter protein production and therefore had to be excluded. This was done by two different strategies. Firstly, RNA from transfected cells was isolated and reverse transcribed. After

that, cDNA was analyzed with primers that would allow the detection of potential additional splicing products because of their design to bind upstream of the splice donor SD1 (forward primer) and downstream of the splice acceptor SA7. Analysis was performed with the *gag* variants huABCDE and wtABCDE as well as wtA_huBCDE, since the latter variant showed the unexpected loss in *gag* expression. Secondly, the same variants were analyzed by northern blot analysis, which is a suitable method to visualize mRNA molecules. For this, an RNA probe was designed that binds upstream of the splice donor to make sure that the probe target sequence would not be lost in case of splicing. If cryptic splicing variants would be generated by codon adaptation, additional bands would have been detected in the northern blot analysis. Since this was not the case, the generation of cryptic splicing effects that influence protein production could be largely excluded. If alternative splicing forms are generated at all, their amount lie below the detection limit.

F.3.5 Impact of the promoter on altered gene expression

As illustrated in Figure E-1, all variants that were used for the analysis of the influence of codon adaptation on protein production were expressed under the control of a heterologous CMV promoter. Our group analyzed the effects of synonymous codon usage in earlier studies for different cell lines and identified the CMV promoter as suitable for expression. Therefore, for consistency reasons it was decided to keep to that expression system and express the subgenomic reporter variants under control of the CMV promoter. As this system is of course highly artificial, selected variants were additionally analyzed under the control of the LTR promoter of HIV-1, reflecting the natural context of the *gag* gene. For this, selected variants were cloned into an expression vector with an LTR promoter. As expected, all 5'-3' adapted variants that were analyzed (huA_wtBCDE, huABC_wtDE and huABCDE) showed expression already in the absence of Rev. Further, the beneficial effect of additive humanization was observed to be comparable to the expression of these variants under control of the CMV promoter. The wild-type variant of *gag* wtABCDE showed complete Rev-dependency. The same was true for the variant wt1-75_hu76-1500, which contains the inhibitory motif at the 5' part of *gag*. In general, comparison between the *gag* variants under control of the CMV- and LTR-promoter showed a strong correlation between the two different sets, indicating a negligible effect of the promoter used on observed expression data. Therefore, it can be excluded that the observed effects are attributed to the artificial system of the CMV promoter.

F.4 Future prospects

Taken together, it was seen that codon adaptation influences *gag* gene expression on different levels. First, a positional effect was observed, since humanization at the 5' part of *gag* was a precondition for enhanced expression levels. Especially the nucleotides surrounding the position 61-75 were identified as a core motif of an inhibitory sequence if they are present in their wild-type form. Without the inhibitory motif, an additive effect of humanization was observed on *gag* expression as

it was also observed in the *egfp* experiments. The higher the number of humanized nucleotides, the higher was the expression level. These findings were true on protein as well as RNA level. RNA amounts correlated strongly with protein levels for all analyzed variants, indicating a strong effect of humanization on regulation of transcript amounts, rather than an impact on translation. The RNA amount is controlled by various biological processes. The performed experiments showed that the main reason for altered transcript concentration were RNA stability and transcriptional efficiency. Both processes might be influenced by altered secondary structures induced by codon adaptation. To confirm this hypothesis a detection of a direct RNA-protein interaction would be helpful. RNAs in cells are associated with RNA-binding proteins. These interactions influence different processes like transport, localization and also stability [166] of RNA molecules. Therefore, RNA pull-down assays should be done to identify relevant proteins binding to these RNAs. For instance, Kula *et al*. identified MATR3 as a cellular cofactor of Rev activity by a method to explore the proteome associated with the nuclear HIV-1 RNAs. For this, they generated cell lines harboring an integrated provirus carrying RNA binding sites for the bacteriophage protein MS2. Flag-tagged MS2 was then used for affinity purification of the viral RNA [167]. Another aspect that would be of interest is how the partial codon adaptation would influence gene expression or even HIV-1 replication in a proviral system. As a first approximation to this question, the expression of selected variants was analyzed under control of HIV-1's LTR promoter. Since viral replication clearly is a more complex process that is impacted by several processes like reverse transcription or timely highly regulated gene expression, examination of partially humanized variants in the proviral context would be of interest. However, it might be the case that enhanced and Rev-independent *gag* expression would result in replication-deficient viruses, since HIV-1's gene expression is an extremely regulated processes and especially the timing of expression is crucial. Nevertheless, detailed characterization of the individual replication steps could help to elucidate the mechanisms involved in interpreting the information conveyed by sequences exhibiting different codon compositions.

G Appendix

G.1 Northern Blot analysis of 5'-3'-adapted variants

As an additional validation of correct splicing and mRNA expression, northern blot analysis was performed for all 5'-3'-adapted variants. Additionally, mRNA was isolated from the nucleus as well as from the cytoplasm, to gain insight in export processes.

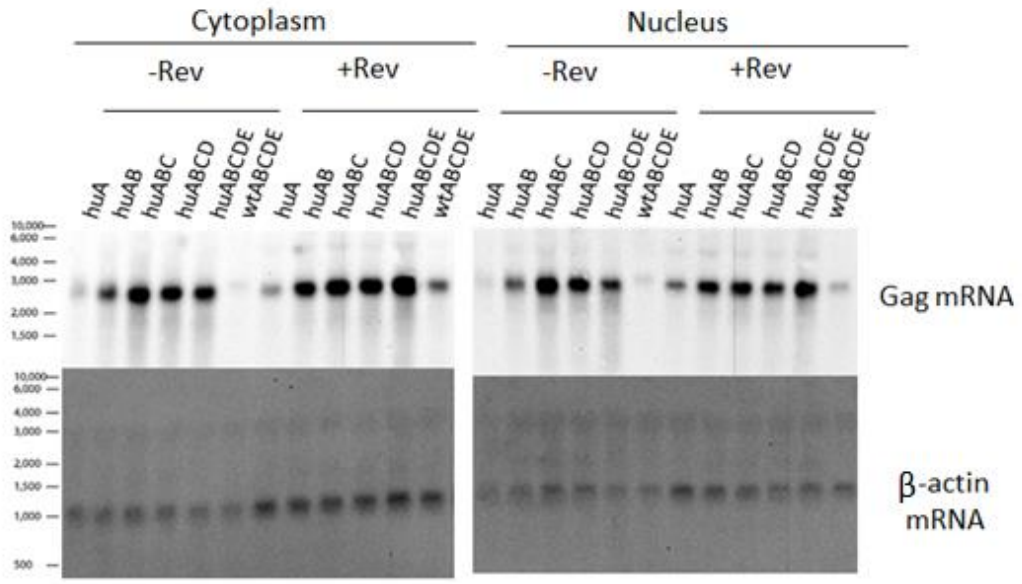

Figure G-1: Northern blot analysis of 5'-3'-adapted variants
HEK293T cells were transfected with the different variants and RNA isolation was performed after separation of nuclear and cytoplasmic fractions. Detection of gag mRNA and β-actin mRNA was performed by the usage of complementary RNA probes.

As shown in Figure G-1 the general expression pattern of the 5'-3' set was confirmed. Additional humanization led to enhanced expression of *gag*. All 5'-3'-adapted variants were expressed in the absence as well as in the presence of Rev. Wild-type *gag* was expressed in the cytoplasm only in the presence of Rev (a minimal band was detected). In the nucleus, an impaired but clearly detectable band was observed. Further, no cryptic splicing variants were detected for any of the partially humanized variants.

G.2 Comparison of p24 levels in the supernatant and intracellular p24 levels of transfected HEK293T cells.

To ensure that analysis of *gag* levels in the supernatants of transfected cells provide reliable insights into *gag* protein production, intracellular *gag* amounts and *gag* release into the supernatant were compared for a selected set of variants. For this, HEK293T cells were transfected with the adapted

gag variants. Gag protein production was quantified in the supernatants as well as cleared cell lysates of transfected cells after 48 hours by p24-ELISA. Transfection was performed with the 5'-3'adapted variants in the absence of Rev which span a broad range of p24 concentrations .

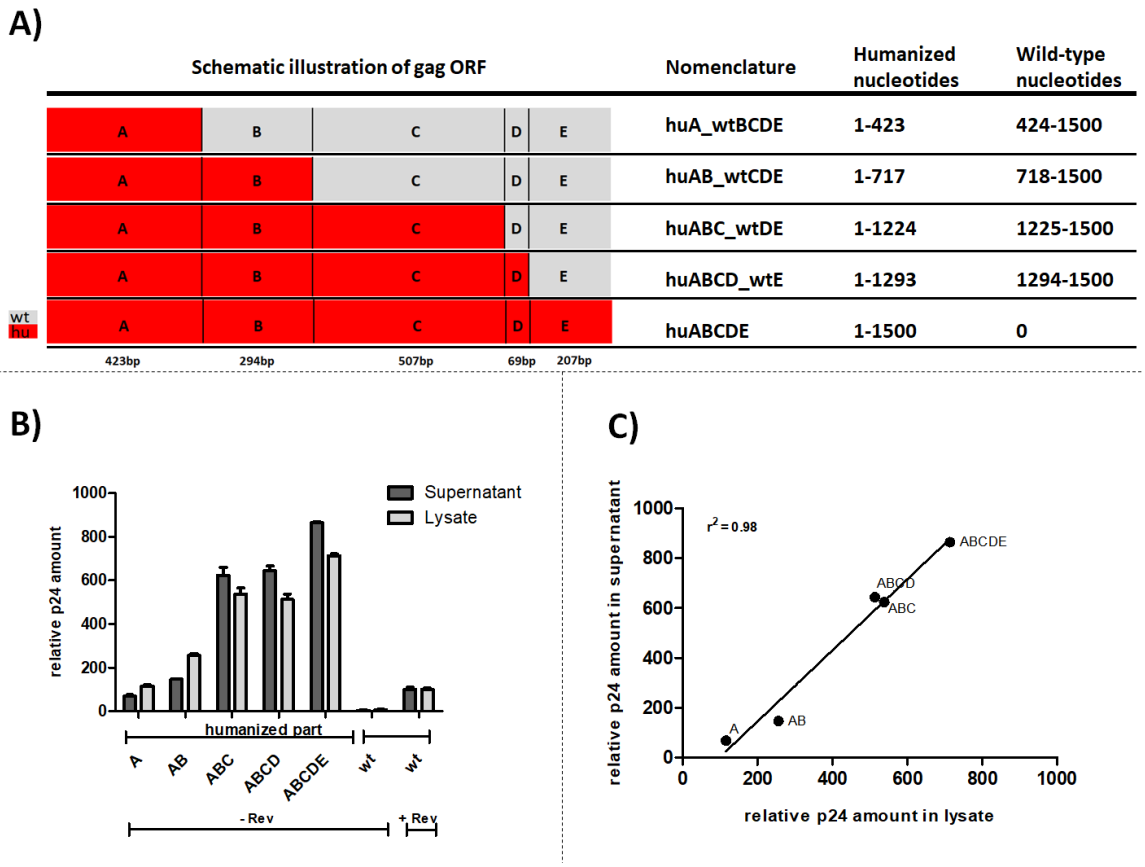

Figure G-2 Comparison of intracellular p24 amounts and p24 levels in the supernatant of transfected HEK293T
A) Schematic overview of the adapted variants and localization of humanized and wild-type nucleotides. B) Gag expression measured by p24-ELISA in supernatants or cell lysates of transfected HEK293T cells for 5'-3'-adapted variants after transfection of HEK293T cells. B) Correlation analysis of intracellular p24 levels (x-axis) and p24 levels in the supernatants (y-axis) of transfected HEK293T cells.

Analysis of intracellular *gag* production showed very similar levels compared to the quantification of *gag* in the supernatant. The beneficial effect of humanization in dependency on the number of humanized nucleotides was observed in both fractions. All 5'-3'-adapted variants were expressed in the absence of Rev. The lowest p24 values were detected for huA_wtBCDE while the completely humanized variant huABCDE showed the strongest expression pattern. The comparability between intracellular p24 level and p24 expression in the supernatant becomes clear by calculating the correlation with each other. As shown in Figure G-2C the two parameters correlated very strongly (r^2= 0.98) with each other. Therefore, all p24 expression analyses were performed only with supernatant of transfected cells.

H List of References

[1] "WHO | Data and statistics," *WHO*. [Online]. Available: http://www.who.int/hiv/data/en/. [Accessed: 25-Oct-2018].

[2] F. Barre-Sinoussi *et al.*, "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)," *Science*, vol. 220, no. 4599, pp. 868–871, May 1983.

[3] E. S. Daar, T. Moudgil, R. D. Meyer, and D. D. Ho, "Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection," *N. Engl. J. Med.*, vol. 324, no. 14, pp. 961–964, Apr. 1991.

[4] R. A. Koup *et al.*, "Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome," *J. Virol.*, vol. 68, no. 7, pp. 4650–4655, Jul. 1994.

[5] A. T. Haase, "Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues," *Annu. Rev. Immunol.*, vol. 17, pp. 625–656, 1999.

[6] B. H. Hahn, G. M. Shaw, K. M. De Cock, and P. M. Sharp, "AIDS as a zoonosis: scientific and public health implications," *Science*, vol. 287, no. 5453, pp. 607–614, Jan. 2000.

[7] B. Korber *et al.*, "Timing the ancestor of the HIV-1 pandemic strains," *Science*, vol. 288, no. 5472, pp. 1789–1796, Jun. 2000.

[8] P. M. Sharp and B. H. Hahn, "The evolution of HIV-1 and the origin of AIDS," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 365, no. 1552, pp. 2487–2494, Aug. 2010.

[9] O. T. Campbell-Yesufu and R. T. Gandhi, "Update on Human Immunodeficiency Virus (HIV)-2 Infection," *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*, vol. 52, no. 6, pp. 780–787, Mar. 2011.

[10] S. Modrow, Ed., *Molekulare Virologie*, 3. Aufl. Heidelberg: Spektrum Akad. Verl, 2010.

[11] E. O. Freed, "HIV-1 replication," *Somat. Cell Mol. Genet.*, vol. 26, no. 1–6, pp. 13–33, Nov. 2001.

[12] H. R. Gelderblom, E. H. Hausmann, M. Ozel, G. Pauli, and M. A. Koch, "Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins," *Virology*, vol. 156, no. 1, pp. 171–176, Jan. 1987.

[13] A. D. Frankel and J. A. Young, "HIV-1: fifteen proteins and an RNA," *Annu. Rev. Biochem.*, vol. 67, pp. 1–25, 1998.

[14] J. Liu, A. Bartesaghi, M. J. Borgnia, G. Sapiro, and S. Subramaniam, "Molecular architecture of native HIV-1 gp120 trimers," *Nature*, vol. 455, no. 7209, pp. 109–113, Sep. 2008.

[15] J. A. G. Briggs, K. Grünewald, B. Glass, F. Förster, H.-G. Kräusslich, and S. D. Fuller, "The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions," *Struct. Lond. Engl. 1993*, vol. 14, no. 1, pp. 15–20, Jan. 2006.

[16] A. Rambaut, D. Posada, K. A. Crandall, and E. C. Holmes, "The causes and consequences of HIV evolution," *Nat. Rev. Genet.*, vol. 5, no. 1, pp. 52–61, Jan. 2004.

[17] D. Sauter and F. Kirchhoff, "HIV replication: a game of hide and sense," *Curr. Opin. HIV AIDS*, vol. 11, no. 2, pp. 173–181, Mar. 2016.

[18] C. Gomez and T. J. Hope, "The ins and outs of HIV replication," *Cell. Microbiol.*, vol. 7, no. 5, pp. 621–626, May 2005.

[19] S. Nisole and A. Saïb, "Early steps of retrovirus replicative cycle," *Retrovirology*, vol. 1, no. 1, p. 9, May 2004.

[20] N. Levy *et al.*, "Structural and functional studies of the HIV-1 pre-integration complex," *Retrovirology*, vol. 10, no. Suppl 1, p. P76, Sep. 2013.

[21] German Advisory Committee Blood (Arbeitskreis Blut), Subgroup 'Assessment of Pathogens Transmissible by Blood'',' "Human Immunodeficiency Virus (HIV)," *Transfus. Med. Hemotherapy*, vol. 43, no. 3, pp. 203–222, 2016.

[22] J. Karn and C. M. Stoltzfus, “Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression,” *Cold Spring Harb. Perspect. Med.*, vol. 2, no. 2, pp. a006916–a006916, Feb. 2012.

[23] R. J. Pomerantz, D. Trono, M. B. Feinberg, and D. Baltimore, “Cells nonproductively infected with HIV-1 exhibit an aberrant pattern of viral RNA expression: a molecular model for latency,” *Cell*, vol. 61, no. 7, pp. 1271–1276, Jun. 1990.

[24] S. Y. Kim, R. Byrn, J. Groopman, and D. Baltimore, “Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression,” *J. Virol.*, vol. 63, no. 9, pp. 3708–3713, Sep. 1989.

[25] M. M. Addo *et al.*, “The HIV-1 regulatory proteins Tat and Rev are frequently targeted by cytotoxic T lymphocytes derived from HIV-1-infected individuals,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 4, pp. 1781–1786, Feb. 2001.

[26] Y. Rivière *et al.*, “Gag-specific cytotoxic responses to HIV type 1 are associated with a decreased risk of progression to AIDS-related complex or AIDS,” *AIDS Res. Hum. Retroviruses*, vol. 11, no. 8, pp. 903–907, Aug. 1995.

[27] N. M. Bell and A. M. L. Lever, “HIV Gag polyprotein: processing and early viral particle assembly,” *Trends Microbiol.*, vol. 21, no. 3, pp. 136–144, Mar. 2013.

[28] D. Gheysen *et al.*, “Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells,” *Cell*, vol. 59, no. 1, pp. 103–112, Oct. 1989.

[29] A. M. L. Lever, “HIV-1 RNA packaging,” *Adv. Pharmacol. San Diego Calif*, vol. 55, pp. 1–32, 2007.

[30] J. G. Carlton and J. Martin-Serrano, “The ESCRT machinery: new functions in viral and cellular biology,” *Biochem. Soc. Trans.*, vol. 37, no. Pt 1, pp. 195–199, Feb. 2009.

[31] J. H. Hurley, “The ESCRT complexes,” *Crit. Rev. Biochem. Mol. Biol.*, vol. 45, no. 6, pp. 463–487, Dec. 2010.

[32] J. H. Hurley and P. I. Hanson, “Membrane budding and scission by the ESCRT machinery: it’s all in the neck,” *Nat. Rev. Mol. Cell Biol.*, vol. 11, no. 8, pp. 556–566, Aug. 2010.

[33] M. W. Yap and J. P. Stoye, “Bending out and breaking away: host-cell accomplices in retroviral escape,” *J. Biol.*, vol. 3, no. 1, p. 3, 2003.

[34] P. Roingeard and D. Brand, “Budding of Human Immunodeficiency Virus,” *N. Engl. J. Med.*, vol. 339, no. 1, pp. 32–32, Jul. 1998.

[35] V. Chukkapalli and A. Ono, “Molecular determinants that regulate plasma membrane association of HIV-1 Gag,” *J. Mol. Biol.*, vol. 410, no. 4, pp. 512–524, Jul. 2011.

[36] I. B. Hogue, G. N. Llewellyn, and A. Ono, “Dynamic Association between HIV-1 Gag and Membrane Domains,” *Molecular Biology International*, 2012. [Online]. Available: https://www.hindawi.com/journals/mbi/2012/979765/. [Accessed: 09-Nov-2018].

[37] D. H. Nguyen and J. E. Hildreth, “Evidence for budding of human immunodeficiency virus type 1 selectively from glycolipid-enriched membrane lipid rafts,” *J. Virol.*, vol. 74, no. 7, pp. 3264–3272, Apr. 2000.

[38] D. K. Worthylake, H. Wang, S. Yoo, W. I. Sundquist, and C. P. Hill, “Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution,” *Acta Crystallogr. D Biol. Crystallogr.*, vol. 55, no. Pt 1, pp. 85–92, Jan. 1999.

[39] C. Tang, Y. Ndassa, and M. F. Summers, “Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein,” *Nat. Struct. Biol.*, vol. 9, no. 7, pp. 537–543, Jul. 2002.

[40] J. A. G. Briggs, T. Wilk, R. Welker, H.-G. Kräusslich, and S. D. Fuller, “Structural organization of authentic, mature HIV-1 virions and cores,” *EMBO J.*, vol. 22, no. 7, pp. 1707–1715, Apr. 2003.

[41] A. A. Waheed and E. O. Freed, “HIV Type 1 Gag as a Target for Antiviral Therapy,” *AIDS Res. Hum. Retroviruses*, vol. 28, no. 1, pp. 54–75, Jan. 2012.

[42] J. Godet, C. Boudier, N. Humbert, R. Ivanyi-Nagy, J.-L. Darlix, and Y. Mély, “Comparative nucleic acid chaperone properties of the nucleocapsid protein NCp7 and Tat protein of HIV-1,” *Virus Res.*, vol. 169, no. 2, pp. 349–360, Nov. 2012.

[43] B. Müller, T. Patschinsky, and H.-G. Kräusslich, “The Late-Domain-Containing Protein p6 Is the Predominant Phosphoprotein of Human Immunodeficiency Virus Type 1 Particles,” *J. Virol.*, vol. 76, no. 3, pp. 1015–1024, Feb. 2002.

[44] H. G. Göttlinger, T. Dorfman, J. G. Sodroski, and W. A. Haseltine, “Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 88, no. 8, pp. 3195–3199, Apr. 1991.

[45] M. Huang, J. M. Orenstein, M. A. Martin, and E. O. Freed, “p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease.,” *J. Virol.*, vol. 69, no. 11, pp. 6810–6818, Nov. 1995.

[46] X. F. Yu, L. Dawson, C. J. Tian, C. Flexner, and M. Dettenhofer, “Mutations of the human immunodeficiency virus type 1 p6(Gag) domain result in reduced retention of pol proteins during virus assembly,” *J. Virol.*, vol. 72, no. 4, pp. 3412–3417, Apr. 1998.

[47] M. Suhasini and T. R. Reddy, “Cellular proteins and HIV-1 Rev function,” *Curr. HIV Res.*, vol. 7, no. 1, pp. 91–100, Jan. 2009.

[48] M. H. Malim, S. Böhnlein, J. Hauber, and B. R. Cullen, “Functional dissection of the HIV-1 Rev trans-activator--derivation of a trans-dominant repressor of Rev function,” *Cell*, vol. 58, no. 1, pp. 205–214, Jul. 1989.

[49] A. W. Cochrane, A. Perkins, and C. A. Rosen, “Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function,” *J. Virol.*, vol. 64, no. 2, pp. 881–885, Feb. 1990.

[50] T. J. Hope, D. McDonald, X. J. Huang, J. Low, and T. G. Parslow, “Mutational analysis of the human immunodeficiency virus type 1 Rev transactivator: essential residues near the amino terminus,” *J. Virol.*, vol. 64, no. 11, pp. 5360–5366, Nov. 1990.

[51] S. Kubota, H. Siomi, T. Satoh, S. Endo, M. Maki, and M. Hatanaka, “Functional similarity of HIV-I rev and HTLV-I rex proteins: identification of a new nucleolar-targeting signal in rev protein,” *Biochem. Biophys. Res. Commun.*, vol. 162, no. 3, pp. 963–970, Aug. 1989.

[52] A. W. Cochrane, C. H. Chen, and C. A. Rosen, “Specific interaction of the human immunodeficiency virus Rev protein with a structured region in the env mRNA,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 87, no. 3, pp. 1198–1202, Feb. 1990.

[53] M. H. Malim, J. Hauber, S. Y. Le, J. V. Maizel, and B. R. Cullen, “The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA,” *Nature*, vol. 338, no. 6212, pp. 254–257, Mar. 1989.

[54] D. A. Mann *et al.*, “A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression,” *J. Mol. Biol.*, vol. 241, no. 2, pp. 193–207, Aug. 1994.

[55] J. Kjems, M. Brown, D. D. Chang, and P. A. Sharp, “Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 88, no. 3, pp. 683–687, Feb. 1991.

[56] J. Rausch and S. Grice, “HIV Rev Assembly on the Rev Response Element (RRE): A Structural Perspective,” *Viruses*, vol. 7, no. 6, pp. 3053–3075, Jun. 2015.

[57] S. Kubota and R. J. Pomerantz, “A cis-acting peptide signal in human immunodeficiency virus type I Rev which inhibits nuclear entry of small proteins,” *Oncogene*, vol. 16, no. 14, pp. 1851–1861, Apr. 1998.

[58] B. R. Cullen, “Nuclear mRNA export: insights from virology,” *Trends Biochem. Sci.*, vol. 28, no. 8, pp. 419–424, Aug. 2003.

[59] M. Graf, A. Bojak, L. Deml, K. Bieler, H. Wolf, and R. Wagner, “Concerted Action of Multiple cis-Acting Sequences Is Required for Rev Dependence of Late Human Immunodeficiency Virus Type 1 Gene Expression,” *J. Virol.*, vol. 74, no. 22, pp. 10822–10826, Nov. 2000.

[60] M. Graf, C. Ludwig, S. Kehlenbeck, K. Jungert, and R. Wagner, “A quasi-lentiviral green fluorescent protein reporter exhibits nuclear export features of late human immunodeficiency virus type 1 transcripts,” *Virology*, vol. 352, no. 2, pp. 295–305, Sep. 2006.

[61] P. Cramer, “RNA polymerase II structure: from core to functional complexes,” *Curr. Opin. Genet. Dev.*, vol. 14, no. 2, pp. 218–226, Apr. 2004.

[62] Y. Hirose and Y. Ohkuma, “Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression,” *J. Biochem. (Tokyo)*, vol. 141, no. 5, pp. 601–608, May 2007.

[63] T. W. Sikorski and S. Buratowski, “The basal initiation machinery: beyond the general transcription factors,” *Curr. Opin. Cell Biol.*, vol. 21, no. 3, pp. 344–351, Jun. 2009.

[64] U. Kutay *et al.*, “Identification of a tRNA-specific nuclear export receptor,” *Mol. Cell*, vol. 1, no. 3, pp. 359–369, Feb. 1998.

[65] A. Köhler and E. Hurt, “Exporting RNA from the nucleus to the cytoplasm,” *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 10, pp. 761–773, Oct. 2007.

[66] G. J. Arts, M. Fornerod, and I. W. Mattaj, “Identification of a nuclear export receptor for tRNA,” *Curr. Biol. CB*, vol. 8, no. 6, pp. 305–314, Mar. 1998.

[67] J. Katahira, “Nuclear export of messenger RNA,” *Genes*, vol. 6, no. 2, pp. 163–184, Mar. 2015.

[68] B. R. Cullen, “Nuclear RNA Export Pathways,” *Mol. Cell. Biol.*, vol. 20, no. 12, pp. 4181–4187, Jun. 2000.

[69] E. V. Koonin and A. S. Novozhilov, “Origin and Evolution of the Universal Genetic Code,” *Annu. Rev. Genet.*, vol. 51, pp. 45–62, 27 2017.

[70] R. J. Jackson, C. U. T. Hellen, and T. V. Pestova, “The mechanism of eukaryotic translation initiation and principles of its regulation,” *Nat. Rev. Mol. Cell Biol.*, vol. 11, no. 2, pp. 113–127, Feb. 2010.

[71] T. E. Dever and R. Green, “The elongation, termination, and recycling phases of translation in eukaryotes,” *Cold Spring Harb. Perspect. Biol.*, vol. 4, no. 7, p. a013706, Jul. 2012.

[72] I. Jonkers and J. T. Lis, “Getting up to speed with transcription elongation by RNA polymerase II,” *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 3, pp. 167–177, Mar. 2015.

[73] S. Sainsbury, C. Bernecky, and P. Cramer, “Structural basis of transcription initiation by RNA polymerase II,” *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 3, pp. 129–143, Mar. 2015.

[74] O. Porrua and D. Libri, “Transcription termination and the control of the transcriptome: why, where and how to stop,” *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 3, pp. 190–202, Mar. 2015.

[75] J. D. Watson and F. H. Crick, “Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid,” *Nature*, vol. 171, no. 4356, pp. 737–738, Apr. 1953.

[76] F. H. C. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin, “General Nature of the Genetic Code for Proteins,” *Nature*, vol. 192, no. 4809, pp. 1227–1232, Dec. 1961.

[77] M. W. Nirenberg and J. H. Matthaei, “The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides,” *Proc. Natl. Acad. Sci.*, vol. 47, no. 10, pp. 1588–1602, Oct. 1961.

[78] M. Nirenberg *et al.*, “The RNA code and protein synthesis,” *Cold Spring Harb. Symp. Quant. Biol.*, vol. 31, pp. 11–24, 1966.

[79] A. J. Griffiths, J. H. Miller, D. T. Suzuki, R. C. Lewontin, and W. M. Gelbart, “Genetic code,” *Introd. Genet. Anal. 7th Ed.*, 2000.

[80] “Fichier:Codontable1.PNG — Wikipédia.” [Online]. Available: https://fr.m.wikipedia.org/wiki/Fichier:Codontable1.PNG. [Accessed: 13-Nov-2018].

[81] P. M. Sharp, “Codon Usage Bias,” in *Encyclopedia of Genetics*, S. Brenner and J. H. Miller, Eds. New York: Academic Press, 2001, pp. 402–406.

[82] J. B. Plotkin and G. Kudla, “Synonymous but not the same: the causes and consequences of codon bias,” *Nat. Rev. Genet.*, vol. 12, no. 1, pp. 32–42, Jan. 2011.

[83] S. K. Behura and D. W. Severson, “Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes,” *Biol. Rev.*, vol. 88, no. 1, pp. 49–61, Feb. 2013.

[84] M. Bulmer, “The selection-mutation-drift theory of synonymous codon usage,” *Genetics*, vol. 129, no. 3, pp. 897–907, Nov. 1991.

[85] P. M. Sharp, M. Stenico, J. F. Peden, and A. T. Lloyd, “Codon usage: mutational bias, translational selection, or both?,” *Biochem. Soc. Trans.*, vol. 21, no. 4, pp. 835–841, Nov. 1993.

[86] R. M. Kliman and J. Hey, “The effects of mutation and natural selection on codon bias in the genes of Drosophila,” *Genetics*, vol. 137, no. 4, pp. 1049–1056, Aug. 1994.

[87] J. L. Bennetzen and B. D. Hall, “Codon selection in yeast,” *J. Biol. Chem.*, vol. 257, no. 6, pp. 3026–3031, Mar. 1982.

[88] L. Duret and D. Mouchiroud, “Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 8, pp. 4482–4487, Apr. 1999.

[89] S. K. Behura, M. Stanke, C. A. Desjardins, J. H. Werren, and D. W. Severson, “Comparative analysis of nuclear tRNA genes of Nasonia vitripennis and other arthropods, and relationships to codon usage bias,” *Insect Mol. Biol.*, vol. 19 Suppl 1, pp. 49–58, Feb. 2010.

[90] T. Ikemura, “Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system,” *J. Mol. Biol.*, vol. 151, no. 3, pp. 389–409, Sep. 1981.

[91] T. Ikemura, “Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs,” *J. Mol. Biol.*, vol. 158, no. 4, pp. 573–597, Jul. 1982.

[92] S. Kanaya, Y. Yamada, Y. Kudo, and T. Ikemura, “Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis,” *Gene*, vol. 238, no. 1, pp. 143–155, Sep. 1999.

[93] E. Angov, “Codon usage: Nature’s roadmap to expression and folding of proteins,” *Biotechnol. J.*, vol. 6, no. 6, pp. 650–659, Jun. 2011.

[94] H. Akashi, “Codon bias evolution in Drosophila. Population genetics of mutation-selection drift,” *Gene*, vol. 205, no. 1–2, pp. 269–278, Dec. 1997.

[95] J. R. Powell, E. Sezzi, E. N. Moriyama, J. M. Gleason, and A. Caccone, “Analysis of a shift in codon usage in Drosophila,” *J. Mol. Evol.*, vol. 57 Suppl 1, pp. S214-225, 2003.

[96] E. N. Moriyama and J. R. Powell, “Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli,” *Nucleic Acids Res.*, vol. 26, no. 13, pp. 3188–3193, Jul. 1998.

[97] D. B. Carlini, Y. Chen, and W. Stephan, “The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes Adh and Adhr,” *Genetics*, vol. 159, no. 2, pp. 623–633, Oct. 2001.

[98] D. B. Carlini and W. Stephan, “In vivo introduction of unpreferred synonymous codons into the Drosophila Adh gene results in reduced levels of ADH protein,” *Genetics*, vol. 163, no. 1, pp. 239–243, Jan. 2003.

[99] S. Chen *et al.*, “Codon-Resolution Analysis Reveals a Direct and Context-Dependent Impact of Individual Synonymous Mutations on mRNA Level,” *Mol. Biol. Evol.*, vol. 34, no. 11, pp. 2944–2958, Nov. 2017.

[100] K. C. Ngumbela *et al.*, “Quantitative Effect of Suboptimal Codon Usage on Translational Efficiency of mRNA Encoding HIV-1 gag in Intact T Cells,” *PLoS ONE*, vol. 3, no. 6, Jun. 2008.

[101] A. Kofman *et al.*, "HIV-1 gag expression is quantitatively dependent on the ratio of native and optimized codons," *Tsitologiia*, vol. 45, no. 1, pp. 86–93, 2003.

[102] F. Supek and K. Vlahoviček, "Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity," *BMC Bioinformatics*, vol. 6, p. 182, Jul. 2005.

[103] P. M. Sharp and W. H. Li, "The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications.," *Nucleic Acids Res.*, vol. 15, no. 3, pp. 1281–1295, Feb. 1987.

[104] C. Gustafsson, S. Govindarajan, and J. Minshull, "Codon bias and heterologous protein expression," *Trends Biotechnol.*, vol. 22, no. 7, pp. 346–353, Jul. 2004.

[105] E. Leonard, D. Nielsen, K. Solomon, and K. J. Prather, "Engineering microbes with synthetic biology frameworks," *Trends Biotechnol.*, vol. 26, no. 12, pp. 674–681, Dec. 2008.

[106] M. Heinemann and S. Panke, "Synthetic biology--putting engineering into biology," *Bioinforma. Oxf. Engl.*, vol. 22, no. 22, pp. 2790–2799, Nov. 2006.

[107] R. McDaniel and R. Weiss, "Advances in synthetic biology: on the path from prototypes to applications," *Curr. Opin. Biotechnol.*, vol. 16, no. 4, pp. 476–483, Aug. 2005.

[108] C. Elena, P. Ravasi, M. E. Castelli, S. Peirú, and H. G. Menzella, "Expression of codon optimized genes in microbial systems: current industrial applications and perspectives," *Front. Microbiol.*, vol. 5, Feb. 2014.

[109] B. Y. H. Cheng, E. Ortiz-Riaño, A. Nogales, J. C. de la Torre, and L. Martínez-Sobrido, "Development of Live-Attenuated Arenavirus Vaccines Based on Codon Deoptimization," *J. Virol.*, vol. 89, no. 7, pp. 3523–3533, Jan. 2015.

[110] A. Nogales, S. F. Baker, E. Ortiz-Riaño, S. Dewhurst, D. J. Topham, and L. Martínez-Sobrido, "Influenza A Virus Attenuation by Codon Deoptimization of the NS Gene for Vaccine Development," *J. Virol.*, vol. 88, no. 18, pp. 10525–10540, Sep. 2014.

[111] F. Tulloch, N. J. Atkinson, D. J. Evans, M. D. Ryan, and P. Simmonds, "RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies," *eLife*, vol. 3.

[112] N. Eiber, "Characterization of variants of human immunodeficiency virus with adapted codon usage in the gag gene." .

[113] S. Einhauser, "Humaner versus HI-viraler Codongebrauch: Einfluss auf die Expression autologer und heterologer Reportergene." 03-Sep-2018.

[114] R. Wagner *et al.*, "Rev-independent expression of synthetic gag-pol genes of human immunodeficiency virus type 1 and simian immunodeficiency virus: implications for the safety of lentiviral vectors," *Hum. Gene Ther.*, vol. 11, no. 17, pp. 2403–2413, Nov. 2000.

[115] G. Bernardi, "Isochores and the evolutionary genomics of vertebrates," *Gene*, vol. 241, no. 1, pp. 3–17, Jan. 2000.

[116] J. Kypr and J. Mrázek, "Unusual codon usage of HIV," *Nature*, vol. 327, no. 6117, p. 20, May 1987.

[117] J. M. Watts *et al.*, "Architecture and secondary structure of an entire HIV-1 RNA genome," *Nature*, vol. 460, no. 7256, pp. 711–716, Aug. 2009.

[118] G. Bernardi, "The human genome: organization and evolutionary history," *Annu. Rev. Genet.*, vol. 29, pp. 445–476, 1995.

[119] E. C. Bronson and J. N. Anderson, "Nucleotide composition as a driving force in the evolution of retroviruses," *J. Mol. Evol.*, vol. 38, no. 5, pp. 506–532, May 1994.

[120] F. J. van Hemert and B. Berkhout, "The tendency of lentiviral open reading frames to become A-rich: constraints imposed by viral genome organization and cellular tRNA availability," *J. Mol. Evol.*, vol. 41, no. 2, pp. 132–140, Aug. 1995.

[121] I. Bahir, M. Fromer, Y. Prat, and M. Linial, "Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences," *Mol. Syst. Biol.*, vol. 5, p. 311, Oct. 2009.

[122] Y. Kawashima *et al.*, “Adaptation of HIV-1 to human leukocyte antigen class I,” *Nature*, vol. 458, no. 7238, pp. 641–645, Apr. 2009.

[123] F. J. van Hemert, A. C. van der Kuyl, and B. Berkhout, “The A-nucleotide preference of HIV-1 in the context of its structured RNA genome,” *RNA Biol.*, vol. 10, no. 2, pp. 211–215, Feb. 2013.

[124] K. Deforche *et al.*, “Estimating the relative contribution of dNTP pool imbalance and APOBEC3G/3F editing to HIV evolution in vivo,” *J. Comput. Biol. J. Comput. Mol. Cell Biol.*, vol. 14, no. 8, pp. 1105–1114, Oct. 2007.

[125] J. P. Vartanian, A. Meyerhans, M. Sala, and S. Wain-Hobson, “G-->A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 8, pp. 3092–3096, Apr. 1994.

[126] D. Lecossier, F. Bouchonnet, F. Clavel, and A. J. Hance, “Hypermutation of HIV-1 DNA in the Absence of the Vif Protein,” *Science*, vol. 300, no. 5622, pp. 1112–1112, May 2003.

[127] J. D. Salter, R. P. Bennett, and H. C. Smith, “The APOBEC Protein Family: United by Structure, Divergent in Function,” *Trends Biochem. Sci.*, vol. 41, no. 7, pp. 578–594, 2016.

[128] G. Martrus, M. Nevot, C. Andres, B. Clotet, and M. A. Martinez, “Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture,” *Retrovirology*, vol. 10, p. 78, Jul. 2013.

[129] D. Hanahan, “Studies on transformation of Escherichia coli with plasmids,” *J. Mol. Biol.*, vol. 166, no. 4, pp. 557–580, Jun. 1983.

[130] S. G. Grant, J. Jessee, F. R. Bloom, and D. Hanahan, “Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 87, no. 12, pp. 4645–4649, Jun. 1990.

[131] F. L. Graham, J. Smiley, W. C. Russell, and R. Nairn, “Characteristics of a human cell line transformed by DNA from human adenovirus type 5,” *J. Gen. Virol.*, vol. 36, no. 1, pp. 59–74, Jul. 1977.

[132] D. C. Rio, S. G. Clark, and R. Tjian, “A mammalian host-vector system that regulates expression and amplification of transfected genes by temperature induction,” *Science*, vol. 227, no. 4682, pp. 23–28, Jan. 1985.

[133] Geßner, Alec, “Untersuchung eines Sequenzmotivs des HIV-gag-Gens in Bezug auf die Hemmung der Genexpression.” 31-Aug-2017.

[134] F. L. Graham and A. J. van der Eb, “A new technique for the assay of infectivity of human adenovirus 5 DNA,” *Virology*, vol. 52, no. 2, pp. 456–467, Apr. 1973.

[135] “Molecular Cloning: A Laboratory Manual, 3rd ed., Vols 1,2 and 3 J.F. Sambrook and D.W. Russell, ed., Cold Spring Harbor Laboratory Press, 2001, 2100 pp., soft cover | Sigma-Aldrich.” [Online]. Available: https://www.sigmaaldrich.com/catalog/product/sigma/m8265?lang=de&region=DE. [Accessed: 16-Nov-2018].

[136] H. C. Birnboim and J. Doly, “A rapid alkaline extraction procedure for screening recombinant plasmid DNA,” *Nucleic Acids Res.*, vol. 7, no. 6, pp. 1513–1523, Nov. 1979.

[137] R. K. Saiki *et al.*, “Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase,” *Science*, vol. 239, no. 4839, pp. 487–491, Jan. 1988.

[138] D. Raab, M. Graf, F. Notka, T. Schödl, and R. Wagner, “The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization,” *Syst. Synth. Biol.*, vol. 4, no. 3, pp. 215–225, Sep. 2010.

[139] A. Kofman, M. Graf, L. Deml, H. Wolf, and R. Wagner, “Codon usage-mediated inhibition of HIV-1 gag expression in mammalian cells occurs independently of translation,” *Tsitologiia*, vol. 45, no. 1, pp. 94–100, 2003.

[140] C. B. Buck, X. Shen, M. A. Egan, T. C. Pierson, C. M. Walker, and R. F. Siliciano, “The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site,” *J. Virol.*, vol. 75, no. 1, pp. 181–191, Jan. 2001.

[141] N. J. Atkinson, J. Witteveldt, D. J. Evans, and P. Simmonds, “The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication,” *Nucleic Acids Res.*, vol. 42, no. 7, pp. 4527–4545, Apr. 2014.

[142] M. Sémon, D. Mouchiroud, and L. Duret, “Relationship between gene expression and GC-content in mammals: statistical significance and biological relevance,” *Hum. Mol. Genet.*, vol. 14, no. 3, pp. 421–427, Feb. 2005.

[143] A. P. Bauer *et al.*, “The impact of intragenic CpG content on gene expression,” *Nucleic Acids Res.*, vol. 38, no. 12, pp. 3891–3908, Jul. 2010.

[144] “Leptomycin B solution from <I>Streptomyces</I> sp. L2913,” *Sigma-Aldrich*. [Online]. Available: https://www.sigmaaldrich.com/catalog/product/sigma/l2913. [Accessed: 03-Dec-2018].

[145] M. Deichner, “Molekulare Analyse des Einflusses der Codon-Adaptierung des 5‘-Bereichs von HIV-1 gag auf die Genexpression.” 23-Aug-2016.

[146] M. D, “Nuclear run-on analysis of transcription. - PubMed - NCBI.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/21390682. [Accessed: 21-Apr-2019].

[147] S. T. Smale, “Nuclear run-on assay,” *Cold Spring Harb. Protoc.*, vol. 2009, no. 11, p. pdb.prot5329, Nov. 2009.

[148] Z. Zhou *et al.*, “Codon usage is an important determinant of gene expression levels largely through its effects on transcription,” *Proc. Natl. Acad. Sci.*, vol. 113, no. 41, pp. E6117–E6125, Oct. 2016.

[149] C. Le Nouen *et al.*, “Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization,” *Proc. Natl. Acad. Sci.*, vol. 111, no. 36, pp. 13169–13174, Sep. 2014.

[150] R. L. Y. Fan *et al.*, “Generation of Live Attenuated Influenza Virus by Using Codon Usage Bias,” *J. Virol.*, vol. 89, no. 21, pp. 10762–10773, Nov. 2015.

[151] M. A. Takata, S. J. Soll, A. Emery, D. Blanco-Melo, R. Swanstrom, and P. D. Bieniasz, “Global synonymous mutagenesis identifies cis-acting RNA elements that regulate HIV-1 splicing and replication,” *PLoS Pathog.*, vol. 14, no. 1, p. e1006824, 2018.

[152] M. A. Takata *et al.*, “CG dinucleotide suppression enables antiviral defence targeting non-self RNA,” *Nature*, vol. 550, no. 7674, pp. 124–127, Oct. 2017.

[153] A. Kofman, M. Graf, L. Deml, H. Wolf, and R. Wagner, “Codon usage-mediated inhibition of HIV-1 gag expression in mammalian cells occurs independently of translation,” *Tsitologiia*, vol. 45, no. 1, pp. 94–100, 2003.

[154] T. Ikemura, “Codon usage and tRNA content in unicellular and multicellular organisms.,” *Mol. Biol. Evol.*, vol. 2, no. 1, pp. 13–34, Jan. 1985.

[155] C.-H. Yu *et al.*, “Codon usage influences the local rate of translation elongation to regulate co-translational protein folding,” *Mol. Cell*, vol. 59, no. 5, pp. 744–754, Sep. 2015.

[156] R. Hershberg and D. A. Petrov, “Selection on Codon Bias,” *Annu. Rev. Genet.*, vol. 42, no. 1, pp. 287–299, 2008.

[157] A. van Hoof and E. J. Wagner, “A brief survey of mRNA surveillance,” *Trends Biochem. Sci.*, vol. 36, no. 11, pp. 585–592, Nov. 2011.

[158] “RNA Surveillance: Molecular Approaches in Transcript Quality Control and their Implications in Clinical Diseases.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2761007/. [Accessed: 22-Apr-2019].

[159] V. Presnyak *et al.*, “Codon optimality is a major determinant of mRNA stability,” *Cell*, vol. 160, no. 6, pp. 1111–1124, Mar. 2015.

[160] Y. Harigaya and R. Parker, “Analysis of the association between codon optimality and mRNA stability in Schizosaccharomyces pombe,” *BMC Genomics*, vol. 17, no. 1, p. 895, Nov. 2016.

[161] “References | mfold.rit.albany.edu.” [Online]. Available: http://unafold.rna.albany.edu/?q=mfold/mfold-references. [Accessed: 21-Dec-2018].

[162] M. Zuker, “Mfold web server for nucleic acid folding and hybridization prediction,” *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3406–3415, Jul. 2003.

[163] A. Waugh *et al.*, “RNAML: a standard syntax for exchanging RNA information,” *RNA N. Y. N*, vol. 8, no. 6, pp. 707–717, Jun. 2002.

[164] M. Zuker and A. B. Jacobson, “Using reliability information to annotate RNA secondary structures.,” *RNA*, vol. 4, no. 6, pp. 669–679, Jun. 1998.

[165] N. Mueller, N. van Bel, B. Berkhout, and A. T. Das, “HIV-1 splicing at the major splice donor site is restricted by RNA structure,” *Virology*, vol. 468–470, pp. 609–620, Nov. 2014.

[166] T. Glisovic, J. L. Bachorik, J. Yong, and G. Dreyfuss, “RNA-binding proteins and post-transcriptional gene regulation,” *FEBS Lett.*, vol. 582, no. 14, pp. 1977–1986, Jun. 2008.

[167] A. Kula, J. Guerra, A. Knezevich, D. Kleva, M. P. Myers, and A. Marcello, “Characterization of the HIV-1 RNA associated proteome identifies Matrin 3 as a nuclear cofactor of Rev function,” *Retrovirology*, vol. 8, p. 60, Jul. 2011.

I Danksagung

Auf dem Weg der Promotion haben mich einige Menschen auf den verschiedensten Wegen unterstützt, daher möchte ich die Gelegenheit nutzen *Danke* zu sagen.

An erster Stelle danke ich meinem Doktorvater Herrn Prof. Dr. Ralf Wagner. Während der gesamten Zeit konnte ich mich auf deinen fachlichen und menschlichen Rat verlassen. Unsere zahlreichen Gespräche haben nicht nur positiv zu dieser Arbeit, sondern auch zu meiner persönlichen Weiterentwicklung beigetragen. Danke, dass du nicht nur immer ein offenes Ohr, sondern auch die entsprechenden Lösungsvorschläge für meine Probleme hattest.

Ein besonderer Dank gilt auch Herrn Dr. Benedikt Asbach. Die vielen Einzelheiten aufzuzählen, bei denen du mich unterstützt hast, ist hier gar nicht möglich. Dazu gehört die Anleitung im Labor ebenso wie viele fachliche Gespräche und zahlreiche Hilfestellungen bei den verschiedensten Fragestellungen. Deine nette, geduldige und freundliche Art haben entscheidend dazu beigetragen mich permanent wohl zu fühlen.

Auch meinen Mentoren Herrn Prof. Dr. Gunter Meister und Herrn Prof. Dr. Thomas Gramberg möchte ich herzlichen für Ihre Anmerkungen, Tipps und Lösungsvorschläge danken. Sie haben wesentlich dazu beigetragen diese Arbeit voranzubringen.

Meinen Studenten danke ich herzlich für ihre Unterstützung im Labor, besonders natürlich meinen Bacheloranten Maria Deichner und Sebastian Einhauser.

Über die Jahre in der AG-Wagner haben sich auch diverse Freundschaften entwickelt, für die ich sehr dankbar bin. Herr Dr.(!) Johannes Johnny Merga Meier möchte ich zum einen für vielen fachlichen Diskussionen und Hilfestellungen, zum anderen für viele tolle Momente außerhalb des Labors danken. Egal ob bei Fußball, Kino oder der (beinahe) besten Pizza der Welt. Das gleiche gilt für Dr.Richard B. Kiener mit dem ich für einige Zeit das Labor teilen durfte. Auch dir danke ich nicht nur für deinen fachlichen Input, sondern auch für die vielen besonderen Momente, die ich immer positiv in Erinnerung behalten werde. Ich hoffe wir können eines Tages das Mysterium der kaputten Wanduhr lösen. Bei Anja Schütz, der wohl besten MTA überhaupt möchte ich mich ebenfalls für eine tolle Zeit im und außerhalb des Labors bedanken. Ich freue mich schon sehr auf zukünftige Grill- oder Poker Einladungen ;)

Auch bei Benni, David, Miri, Julia, Christina, Meli und (um das Ganze etwas abzukürzen) der gesamten AG Wagner möchte ich mich für eine ganz besondere Zeit bedanken. Das tolle Miteinander im Labor wird mir immer in schöner Erinnerung bleiben.

Von ganzem Herzen danke ich meiner Frau Franzi, die mich immer wieder aufgebaut hat wenn es nötig war, mir zugehört hat wenn ich ein offenes Ohr gebraucht habe und mir seit über 10 Jahren immer den nötigen Rückhalt gibt. Ich bin froh dich immer an meiner Seite zu wissen.

Natürlich danke ich auch meinen Eltern für Ihre Unterstützung auf allen Ebenen und in allen Lebenslagen. Ohne euch wäre weder mein Studium noch meine Promotion überhaupt möglich gewesen.