Regulation of HIV-1 replication by CpG dinucleotides

Antzin Anduetza, Irati

Awarding institution:
King's College London

Regulation of HIV-1 replication by CpG dinucleotides

Irati Antzin Anduetza

A thesis submitted to the University of London for the degree of Doctor of Philosophy

Department of Infectious Diseases

King's College London

School of Immunology & Microbial Sciences

Declaration

I, Irati Antzin Anduetza, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been referenced in the thesis.

London, 29th November 2018

Abstract

Characterizing how RNA elements and nucleotide biases in the HIV-1 genome regulate replication is necessary for a complete understanding of the viral life cycle. To characterise how the RNA sequence in *gag* controls HIV-1 replication, we codon modified an ~350nt region in this gene. These synonymous mutations inhibited viral replication by decreasing genomic RNA (gRNA) abundance, gRNA stability, Gag expression, virion production and infectivity. Using RNA-seq, we identified that the synonymous mutations activated a cryptic splice site that is the cause of the reduction in viral replication.

There is a much lower than expected frequency of CpG dinucleotides in HIV-1 and codon modification introduced a substantial increase in CpG abundance. To determine if CpG dinucleotides are necessary for inhibition of HIV-1 replication, codons introducing CpGs were mutated back to the wild type codon, which restored efficient Gag expression and infectious virion production. To determine if they are sufficient to inhibit viral replication, CpG dinucleotides were inserted into *gag* in the absence of other point mutations. The results showed that while viral replication is greatly decreased, Gag expression, and virion production levels are similar to wild type HIV-1. Interestingly, gRNA abundance was decreased in the cell lysate and the media.

When we introduced CpG dinucleotides into nucleotides 22-1188 in *gag* we showed a reduction in viral replication that was the result of a decrease in Gag expression, virion production and infectivity. In addition, we showed that the inhibitory mechanism required the RNA-binding protein ZAP.

In summary, we have used HIV-1 as a model system to understand how CpG dinucleotides can inhibit viral replication. Understanding the mechanisms underlying the reduction in viral fitness may be extrapolated to different RNA viruses and could result in many applications.

Table of Contents

Table of Figures

Table of Tables

Abbreviations

AIDS	Acquired immune deficiency syndrome
AGS	Aicardi-Goutières syndrome
ART	Antiretroviral treatment
APOBEC	Apolyprotein-B mRNA editing enzyme catalytic polypeptide-like
ALIX	Apoptosis-linked gene-2-interacting protein X
CCR5	C-C chemokine receptor type 5
CTD	C-terminal domain
CA	Capsid
CCD	Catalytic core domain
cGAS	cGAMP synthase
CHMP	Charged multivesicular body proteins
CRM1	Chromosomal maintenance 1
CPSF6	Cleavage and polyadenylation specific factor 6
CM	Codon modified
CTE	Constitutive transport element
CUL5	Cullin 5
CXCR4	CXC chemokine receptor type 4
CycT1	Cyclin T1
CDK9	Cyclin-dependent kinase 9
cGAMP	Cyclin guanosine monophosphate-adenosine monophosphate
CypA	Cyclophilin A
SOCS	Cytokine suppressor signalling
CT	Cytoplasmic tail
CTL	Cytotoxic cell
DC	Dendritic cell
DENV	Dengue virus
dNTP	Deoxynucleotide triphosphate
DIS	Dimerization initiation signal
EBOV	Ebola virus

ELOB	Elongin B
ELOC	Elongin C
ER	Endoplasmic reticulum
ESCRT	Endosomal sorting complexes required for transport
Env	Envelope
EIAV	Equine infectious anemia virus
E. Coli	Escherichia coli
EDTA	Ethyleneadiaminetetraacetic acid
eiF4F	Eukaryotic initiation factor 4F
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FIV	Feline immunodeficiency virus
FBS	Fetal bovine serum
FS	Frameshift
FV1	Friend virus susceptibility 1
gRNA	Genomic RNA
GPI	Glycosylphosphatidylinositol
Hsp90	Heat shock protein 90
HCV	Hepatitis C virus
HSV-1	Herpes simplex virus 1
HAART	Highly active antiretroviral treatment
HIV	Human immunodeficiency
IAV	Influenza A virus
IN	Integrase
IFN	Interferon
IFITM	Interferon inducible transmembrane proteins
ISG	Interferon-stimulated genes
IRES	Internal ribosome binding site
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
KO	Knock out
L-domain	Late-budding domain
LEDGF	Lens epithelium-derived growth factor

LTR	Long terminal repeat
LNTP	Long-term non-progressor
MHC	Major histocompatibility complex
MA	Matrix
MX2	Myxovirus resistance 2
NTD	N-terminal domain
NELF	Negative elongation factor
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NES	Nuclear export signal
NF-κB	Nuclear factor kappa B
NLS	Nuclear localization signal
NPC	Nuclear pore complex
NXF1	Nuclear RNA export factor 1
NC	Nucleocapsid
NUP153	Nucleoporin 153
NUP358	Nucleoporin 358
NRTI	Nucleoside reverse transcriptase inhibitor
ORF	Open reading frame
OD	Optical density
Pts	Pan troglodytes schweinfurthii
Ptt	Pan troglodytes troglodytes
PAMP	Pathogen associated molecule pattern
PRR	Pathogen-recognition receptor
PEI	Polyethylenimine
P-TEFb	Positive transcription elongation factor
PIC	Pre-integration complex
PR	Protease
RELIK	Rabbit endogenous lentivirus type K
RRE	Rev-responsive element
RT	Reverse transcriptase
RTC	Reverse transcription complex
RNAP II	RNA polymerase II
RSV	Rous Sarcoma virus

SAMHD1	SAM domain and HD domain-containing protein 1
SERINC	Serine incorporator
SIV	Simian immunodeficiency virus
siRNA	Small interfering RNA
SP1	Spacer peptide 1
SP2	Spacer peptide 2
SSE	Structure-specific endonuclease
SU	Surface
SAVE	Synthetic attenuated virus engineering
TCC	Target capture complex
TLR	Toll-like receptor
TAR	Trans-activation region
TM	Transmembrane
TF	Transmitted founder
TNPO3	Transportin 3
TRIM5α	Tripartite motif-containing protein 5
TSG101	Tumor susceptibility gene 101
UTR	Untranslated region
UBER	Uracil base excision repair
UNG2	Uracil DNA glycosylase 2
VSV-G	Vesicular stomatitis virus G protein
Vpr	Viral protein R
Vpu	Viral protein U
ZAP	Zinc-finger antiviral protein P

Acknowledgements

Firstly, I would like to thank my supervisor Chad Swanson for his support during these years. His motivation and immense knowledge are admirable. I am very grateful for having had the opportunity to learn lab skills directly from him in the very beginning; I think it made a difference. He has been excellent in guiding me through the process and very patient in difficult periods. I am lucky to have had him as a supervisor.

I would like to thank my second supervisor Michael Malim for his scientific advice and guidance over the course of my PhD. Thank you to the entire Malim lab for always being available to help, especially Luis Apolonia and Gilberto Betancor.

I would like to thank my thesis committee, Eugene Makeyev, Julie Fox and especially Juan Martin-Serrano, for the very valuable advice during the meetings and support chats.

I would like to thank members of the Swanson lab. A special thank you goes to Charlotte Mahiet for all the teaching and fun from the very beginning and to Laura Hidalgo Lumbreras for the emotional support in the end. Thank you to Helin Sertkaya, Kristina Schierhorn and Mattia Ficarelli for making the lab a nice place to work.

Last but not least, I would like to thank my parents Karlos and Maria Jose for all their hard work so I could always have the best opportunities. Thank you to Federico's infinite patience and hugs, especially in the end, I know it has not been easy.

Chapter 1: Introduction

Since the first clinical observations more than 3 decades ago, the human immunodeficiency virus (HIV) pandemic has caused more than 35 million deaths. Although highly active antiretroviral treatment (HAART) has reduced the number of acquired immune deficiency syndrome (AIDS) - related mortality, 37 million people are currently living infected with HIV-1 (1). Access to therapy has improved but is not universal, and prospects of curative treatments and an effective vaccine are uncertain. Thus, AIDS continues to pose one of the most challenging threats to public health (1-3).

1.1 Origins of HIV-1 and AIDS

AIDS was first recognised as a new disease in 1981 when increasing numbers of men having sex with men that contracted opportunistic infections with similar clinical observations were reported in the United States (4). In 1983, a new human retrovirus was isolated from a patient and, within a year, similar viruses had been isolated from patients with AIDS, which confirmed that HIV causes AIDS (5-7).

HIV-1 is a retrovirus. Reverse transcription and integration are defining features of retroviridae. These viruses use the virally encoded enzyme, reverse transcriptase (RT), to convert their RNA genomes into DNA and integrase (IN) to integrate it into the host genome. Retroviruses are classified as simple and complex viruses: simple viruses encode for Gag, protease (Pro), polymerase (Pol) and envelope (Env) gene products and include alpharetroviruses, betaretroviruses and gammaretroviruses. Complex retroviruses comprise deltaretroviruses, epsiloretroviruses, spumaviruses and lentiviruses and they encode for an array of regulatory proteins in addition to the already mentioned gene products. HIV-1 belongs to the lentivirus genus.

Lentiviruses cause chronic persistent infection in various mammalian species such as horses (equine infectious anemia virus, EIAV), felines (feline immunodeficiency virus, FIV), and primates (simian immunodeficiency virus, SIV) (2). In addition, some lentiviruses have become endogenous in the past, such as the rabbit endogenous lentivirus type K (RELIK), which infiltrated the germ-line about 12 million years ago (8, 9). These endogenous lentiviruses are of particular interest to understand the timescale of lentivirus evolution. Studies carried out on SIV-infected monkeys showed that while most of the SIV transmissions occur within members of the same species, there is numerous evidence of cross-species transmission from SIV (2, 10, 11).

It is now well established that HIV-1 is the result of a zoonotic transfer of viruses infecting primates in Africa, where SIV, which naturally infects primates, infected humans (2). While viral sequence comparisons between SIVmac (isolated from rhesus macaque) and HIV-1 did not show great similarity, comparisons to HIV-2 showed that these viruses are closely related. HIV-2 also induces AIDS-like syndrome in their host. Later, HIV-1 specific antibodies were isolated from two wild common chimpanzees (Pan troglodytes troglodytes, p.t troglodytes) in Gabon (12-15). This virus is named SIVcpz, and the overall genetic organisation is closely related to HIV-1 (12, 13). Determining Pan troglodytes as the natural reservoir of SIVcpz has been challenging. This is because common chimpanzees are an endangered species and studies were limited to captive-born apes (16, 17). SIVcpz was rarely found among captive chimpanzees because most of them had been imported and were members of the Pan troglodytes verus subspecies, which is not infected by this virus. This suggested that chimpanzees acquired the infection after their divergence into different subspecies (18). These findings also led to hypothesise that a cross-species transmission from SIVcpz resulted in HIV-1 infection in humans. HIV-1 comprises 4 different lineages, each of which is the result of an independent cross-species transmission: group M, N, O and P. Group M was the first discovered and is the pandemic form of HIV-1 (19). Group O was isolated later in Cameroon and represents less than 1% of the HIV-1 infections worldwide (20). Following this lineage, infection caused by the group N (21)

was documented, and infection by group P has only been identified in two persons (22). The origins of group M and N are different from the origins of group O, and P. SIVcpz strains are now classified into two subspecies-specific lineages, named SIVcpzPtt and SIVcpzPts. Phylogenetic analysis of the sequences has confirmed that M and N originated from SIVcpzPtt while O and P originated from cross-species transmission of SIVgor viruses from gorillas that are evolutionary related to SIVcpzPtt (2, 12-14) (Figure 1.1). Transmission of SIVcpzPts to humans has not been identified. HIV-1 group M is further classified into 9 subgroups (A-D, F-H, J and K). While subtype B is the cause of most of the infections in Europe and America, A, C and D predominate in Africa (2).

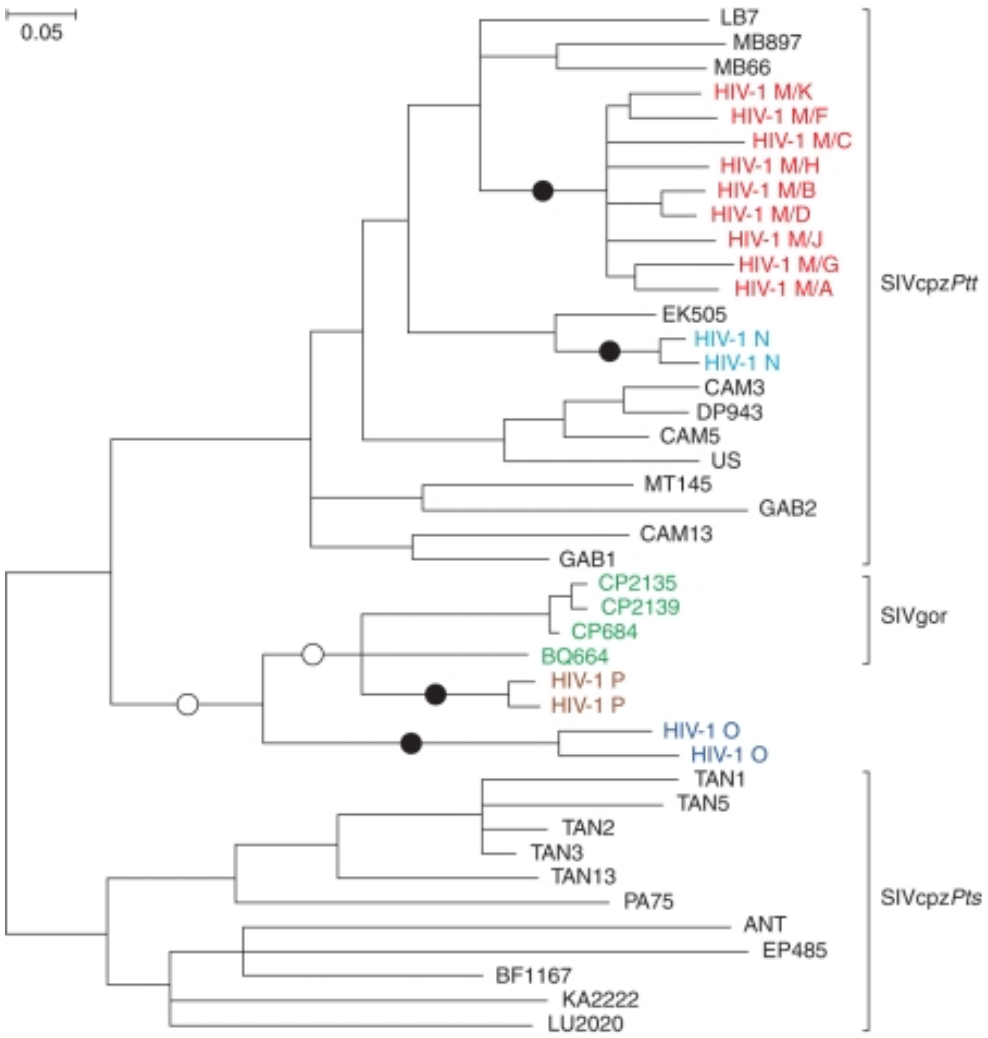


Figure 1.1. Origins of the different HIV-1 strains and the phylogenetic relationships of representative SIVcpzPtt, SIVcpzPts and SIVgor strains. HIV-1 M and N pandemic strains originated from SIVcpzPtt, and HIV-1 O and P originated from cross-species transmission of SIVgor. Figure taken from (2) with the copyright permission from Cold Spring Harbor Laboratory Press.

1.2 AIDS pathogenesis

HIV-1 is mainly a sexually transmitted disease although infection can also occur when the virus is in direct contact with the bloodstream (23, 24). The early events of HIV-1 infection are difficult to study in humans and therefore most of the knowledge describing this critical period has been studied in SIV infection in macaques. The initial target for viral infection is CD4+ T cells on the site of the exposure, usually cells in the damaged mucosa. In the case of vaginal transmission in SIVmac model, the virus specifically targets the Th17 lineage (25). Although macrophages have been reported to express low levels of CD4 (26), infection of peripheral monocytes has been identified in vivo (27, 28). Infection starts with a single target cell being infected with a single virion. The SIVmac model has shown that when the viral concentration is high enough it diffuses to the adjacent cells and tissues such as CD4+ T cells in the gut-associated lymphoid tissue (GALT) (29) and then to systemic infection (30). The same model has also shown that productive infection in CD4+ T cells can be detected only after 2 days from the exposure and spread to the distal lymph nodes and other tissues occurs within a 2-week period (23, 24).

The course of HIV-1 infection can be classified in three stages and is depicted in Figure 1.2. During the first stage or eclipse phase (weeks 1-3) the virus replicates locally within the mucosa, and the viral load is undetectable. In this phase, the viral reservoir is established, and IFN response is detected (31). Subsequently, viral dissemination via lymph nodes into tissues and organs leads to the acute infection (24, 32). Here, high levels of viremia (more than 10^7 copies of viral genomes per millimetre of blood) are detected, and a severe initial loss of CD4+ T cells is observed. This phase is accompanied by flu-like symptoms, where the severity of the symptoms correlates with the peak of viral load. An adaptive cytotoxic CD8+ T cell response results in partial control of viral replication, at the end of the acute phase, a sharp decline in viral load is observed (33, 34). This is followed by the chronic infection or clinical latency phase where a constant or slowly

increasing level of viremia is detected. Antiretroviral treatment (ART) prevents new cells from becoming infected but does not eliminate the virus from already infected cells. While patients are asymptomatic during this phase, a large number of CD4+ cells are being infected. In patients without ART, CD4+ T cell count declines to <200 cells per microliter and the immune system is no longer able to fight against pathogens. This phase is accompanied by opportunistic infections that can be fatal (23, 24).

Infection of CD4+ T cells results in cell death, and this subsequently leads to AIDS. The mechanism by which the virus kills the infected cells, though, is still controversial. One possible mechanism suggests that CD4 T cells block viral replication, which results in an accumulation of incomplete DNA transcripts (35). These cytosolic DNA transcripts activate the innate immune response that triggers cell death. The host defence mechanism promotes proapoptotic and proinflammatory response, which involves caspase-3 and caspase-1 activation. Another proposed mechanism involves CD8+ T cell lymphocyte (CTL)-mediated killing of infected CD4+ T cells (36).

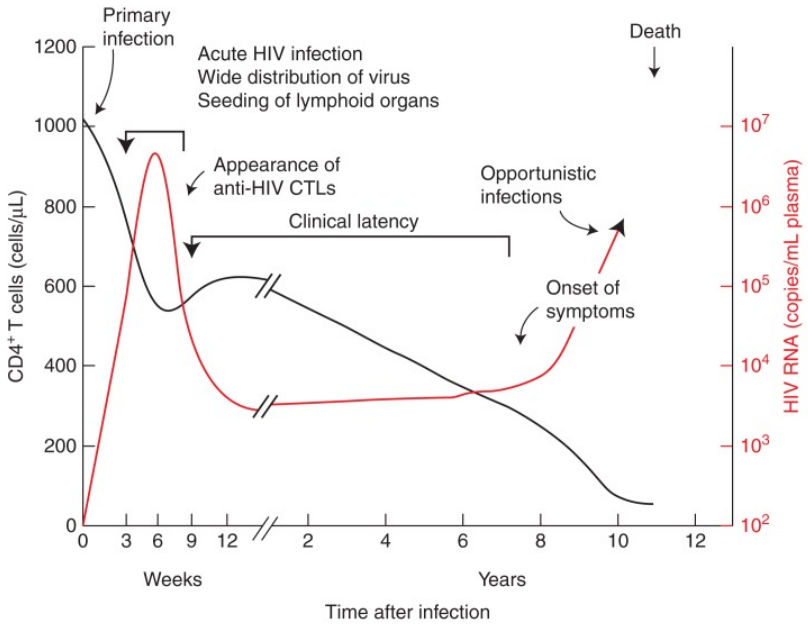


Figure 1.2. Typical time-course on an HIV-1 infected patient. After infection the virus rapidly replicates in the first weeks and is accompanied by a large loss of CD4+ T cells. Following this phase viremia declines rapidly and also the rate at which CD4+ T cells are depleted. Figure taken from (36) with the copyright permission from Cold Spring Harbor Laboratory Press.

The innate immune response to viral infection is triggered upon sensing of the pathogen through pathogen-associated molecular patterns (PAMPs) in viral products by pathogen-recognition receptors (PRRs). As a result, the host cells initiates innate immune signalling, which leads to virus restriction (37). Virus-host interactions and sensing occurs in HIV-1-infected CD4+ T cells and culminates with the transcriptional activation and production of interferons (IFNs) (37, 38). IFNs are a family of pro-inflammatory and immune-modulatory cytokines that trigger an antiviral response through the upregulation of multiple genes named interferon-stimulated genes (ISGs) (37). There are 3 types of IFNs, but the antiviral state is mainly induced by type I IFN.

Several host proteins have been identified as PRRs for HIV PAMPs. IFI16, one of the hundreds IFN-stimulated genes (ISGs), has been shown to sense HIV-1 through recognition of the viral reverse transcriptase products (35). IFI16 can recognise and bind both, DNA products of HIV reverse transcription and a segment of the HIV-1 long terminal repeat region (LTR). This results in IFN-stimulatory DNA response, which is dependent on the adaptor STING, the protein kinase TBK1 and the transcription factors IRF3 and IRF7 (39). Interestingly, IFI16 has also been shown to produce inflammasome activation and inflammatory cell death of infected CD4+T cells (40). cGAS has also been identified as a PRR against HIV-1 infection and other retroviruses. This cytosolic protein is known to bind double-stranded DNA (dsDNA), which results in the production of cyclic GMP-AMP (cGAMP) (41). This dinucleotide product acts a second messenger that binds STING and thus, also results in TBK1 activation and downstream IRF3 and IRF7 activation (42). While production of cGAMP is inhibited upon treatment of reverse transcriptase inhibitors, this does not happen upon treatment with inhibitors of later stages of the HIV-1 cycle, which suggests that cGAS also acts as PRR to DNA products of the reverse transcription. The result of HIV-1-sensing by both IFI16 and cGAS is induction of type-1 IFN-mediated proinflammatory cytokines, characteristic of the acute phase of the HIV-1 infection (38, 43). TLR7, on the other hand, is essential for HIV-1 sensing by pDCs (44). Recognition of viral genomic RNA by TLR7 occurs in endosomes

and leads to its activation (44, 45). This results in the recruitment of myeloid differentiation primary response gene 88 (MYD88), which induces IFNs and other cytokines through activation of IRF7 and NF-κB (46). Interestigly, depletion of pDCs prior to HIV-1 infection abolishes induction of IFN-I and ISGs during acute infection (47).

1.2.1 HIV-1 prevention and treatment

The use of condom has been a cornerstone of HIV prevention, although its effectiveness is limited to 70-80% (24, 48). Other prevention strategies also exist. Three randomised control trials have reported a 50-60% reduction in HIV-1 acquisition in circumcised men, although this measure does not reduce the risk of transmission to the female partners (24, 49). Mother-to-child transmission is possible during pregnancy, labour, delivery and breastfeeding, but antiretroviral treatment (ART) of pregnant women can reduce transmission to <5% (24, 50). People with low levels of viremia are less likely than people with high viral loads to transmit the virus, and therefore, treatment with ART can greatly reduce HIV-1 transmission (24, 51). Post-exposure prophylaxis, when started immediately or no longer than 72h after exposure also maximises the chances of prevention. For individuals at high risk of becoming infected, pre-exposure prophylaxis can also provide protection (24). Regarding the development of a protective vaccine, several trials have reported some evidence of a reduction in HIV-1 acquisition (52), but new strategies that provide higher levels of protection are needed (24).

The current treatment strategy is focused on slowing disease progression, and ART is very effective in suppressing viremia levels (24). Nevertheless, many patients fail to restore a normal immune function, some develop drug resistance, and toxicities can have a cumulative effect. These drugs target essential steps in the HIV-1 life cycle (described in section 1.4). Nucleoside reverse transcriptase inhibitors (NRTIs) are analogues of nucleosides and nucleotides that are preferentially used during reverse transcription, leading to the termination of the cDNA synthesis (24). Non-nucleoside reverse transcriptase inhibitors (NNRTIs) inhibit reverse transcription by producing a

conformational change in the reverse transcriptase (RT). Integrase inhibitors prevent the integration of the viral genome in the host cell (24). Protease inhibitors block maturation of the viral particle, resulting in a block of the HIV-1 life cycle at a late stage. Entry inhibitors prevent the viral particle from entering the cell by binding to the co-receptor CCR5 or CXCR4 and fusion inhibitors such as T20 (enfurvitide) block membrane fusion by binding to gp41 (24, 53).

1.3 Genomic organisation of the HIV-1 genome

Expression of the HIV-1 genome is controlled by the 5' long terminal repeat (LTR), which includes the HIV-1 promoter (54, 55). The 3'LTR, on the other hand, promotes polyadenylation of the newly synthesised transcripts, a process that is necessary for nuclear export and subsequent translation (54).

The HIV-1 genome is a ~9kb single stranded RNA virus that encodes nine open reading frames (ORF)(56). HIV-1 extensively relies on polyprotein processing, alternative splicing, frameshifting, and leaky scanning in order to express all the proteins. The initial HIV-1 transcript is processed into completely spliced mRNAs and partially spliced mRNAs (57). The regulatory proteins Tat and Rev and the accessory protein Nef, are encoded in fully spliced mRNAs. Vif, Vpr, Vpu and Env are translated from incompletely spliced mRNAs and, finally, Gag-Pol polyprotein is encoded in an unspliced mRNA (58) (Figure 1.3).

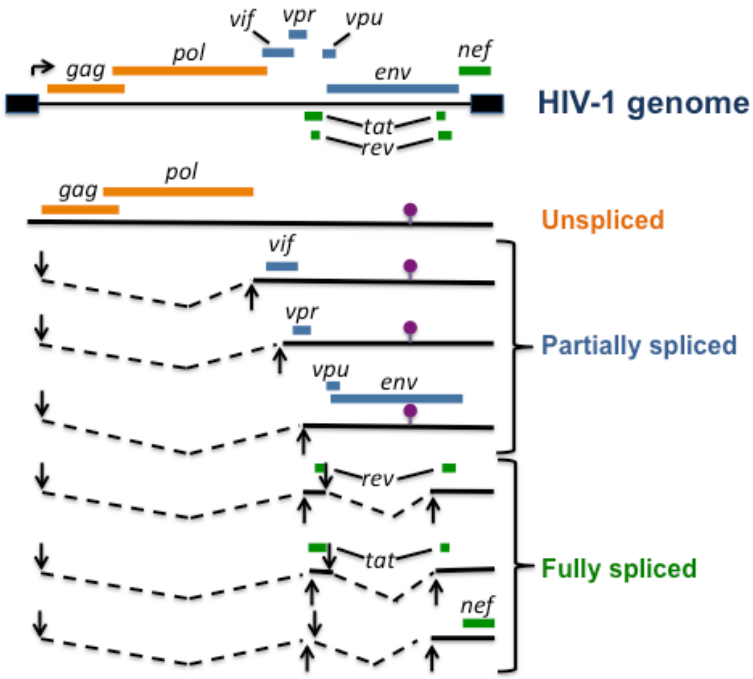


Figure 1.3. The HIV-1 genome is processed into completely spliced mRNAs and partially spliced mRNAs.

1.3.1 HIV-1 open reading frames

HIV-1 *pol* encodes for several enzymes that are essential for HIV-1 replication. Reverse transcriptase (RT) is responsible for the conversion of the genomic viral single-stranded RNA to double stranded cDNA. Integrase (IN) mediates the integration of the proviral DNA into the host genome. On the other hand, protease (PR) is required for the cleavage of the Gag precursor and is necessary for virion maturation (54). During or shortly after budding the 55 kDa Gag polyprotein is cleaved into the mature proteins matrix (MA), capsid (CA), nucleocapsid (NC), p6 and the spacer peptides SP1 and SP2. Gag contains all the domains necessary to assemble the viral particle (59, 60). The structural gene *env* encodes for Env. Env is synthesised in the rough endoplasmic reticulum (ER) from a singly spliced Vpu/Env bicistronic mRNA. Env is a 160 kDa polyprotein that is cleaved into a gp120 surface (SU) and gp41 transmembrane (TM) products and contains

the determinants to bind to the cellular receptor CD4, which together with the co-receptors CCR5 and CXCR4 enables the fusion with the plasma membrane.

In addition to encoding for the structural proteins, the HIV-1 genome also encodes for the regulatory proteins Tat and Rev. Tat regulates transcriptional elongation by binding to TAR. HIV-1 must export intron-containing mRNAs to express the proteins encoded in the partially spliced (Vif, Vpu, Vpr, Env) or unspliced (Gag-Pol) transcripts. Eukaryotic cells, however, block the export of intron-containings mRNAs (61). Translation of Rev in the early stages of the HIV-1 life cycle is the mechanism by which the virus overcomes this restriction. After translation, Rev shuttles back to the nucleus and binds to the Rev response element (RRE) present in the unspliced transcripts facilitating their export to the cytoplasm (62). HIV-1 also encodes for four accessory proteins: Vif, Vpr, Vpu and Nef. The accessory proteins play an essential role in the regulation of host and adaptive immune response and evasion and will be discussed in more detail in section 1.5.

Upon translation of the proteins described above, the HIV-1 viral particle is assembled to ensure the progression of the infection. The mature virion has a spherical shape and around 100nm in diameter and contains the conical capsid core where the ssRNA, the virally encoded enzymes (RT, PR and IN) and the accessory protein Vpr are harboured (63-67).

1.3.2 HIV-1 *cis*-acting elements

In addition to coding for the amino acids that are translated into proteins, the HIV-1 genome has incorporated several sequence elements that regulate the production of progeny virions at a different level. Many of these elements are included within the untranslated region of the genome: the 5' untranslated region (5'UTR) is highly structured and regulates transcriptional elongation, splicing, genomic RNA dimerisation, packaging of full-length viral RNA, and reverse transcription (68). The 5´UTR contains the repeat region (R), U5

region and the leader region. The R region folds into the transactivation response region (TAR), that serves as binding site for Tat, and the polyA hairpins (54). The 3´UTR contains the U3 region and another R region. Long terminal repeats (LTR) are generated during reverse transcription and contain multiple regulatory elements that serve as binding for transcription initiation factors that regulate RNA synthesis (54). In addition to containing the HIV-1 promoter, the 4 stem-loop structures present in the 5'UTR (SL1, SL2, SL3 and SL4) regulate packaging of the gRNA into the virion. The dimerisation initiation signal (DIS) is also located in SL1 and provides a complementary sequence for the formation of a diploid RNA that is packaged into the virion (69).

Other regulatory elements, however, are located in the translated region of the genome: there is a heptameric sequence where the ribosomal frameshifting occurs (FS). The Rev-responsive element (RRE), on the other hand, plays a major role in nuclear export and it will be discussed in detail in section 1.4.8 (70).

In addition to these highly structured elements, the HIV-1 genome also contains other *cis*-acting signals that regulate viral replication and remain to be further characterised. *Cis*-acting repressive sequences (CRSs) are AU-rich sequences that are present throughout the HIV-1 genome and contribute to the degradation of intron-containing mRNAs in the absence of Rev, but no consensus sequence has been defined yet (71-73). The nucleotide composition of the viral RNA may also play an important role in the regulation of HIV-1 replication (74). The HIV-1 genome is significantly enriched in adenosine (A). This feature contributes not only to under- and over-representation of nucleotide frequencies but also to an unusual codon usage (74-76).

1.3.3 Viral epitranscriptome

Post-transcriptional modifications of the viral RNA transcripts have also been reported to be functionally relevant in regulating viral replication. This area of

study has gained interest in recent years and is called viral epitranscriptomics (77, 78). These nucleotide modifications do not change the sequence of the RNA. Some are located in the 5' end of the transcript and include the 7-methylguanosine (m^7G) (functional relevance is described in more detail in section 1.4.9), N^6-methyladenosine (m^6A), 2'-O-methyladenosine (Am), N^6-2-O-methyladenosine (m^6Am), pseudouridine, and 5-methylcytosine (77, 78). Other modifications such as the poly(A) tail are found in the 3'end of transcripts (78). One of the most extensively studied modifications is the m^6A and several recent studies have reported its functional relevance in mRNA function (77).

m^6A was first identified on cellular mRNAs, and it is now known to be more abundantly present in highly regulated genes (79, 80). Following this reports the presence of m^6A in a wide range of different viral RNA species was characterised. RNA viruses such as Influenza A virus (IAV), Rous sarcoma virus (RSV) and HIV-1 and DNA viruses such as adenovirus and herpes simplex virus 1 (HSV-1) have been reported to contain m^6A residues in their mRNAs (77). Regarding HIV-1, 10-14 m^6A residues have been identified and suggested to play a role in gene expression and nuclear export (81, 82).

The addition of m^6A is mediated by the enzyme methyltransferase-like 3 (METTL3) and it mainly occurs in the nucleus. Several cofactors such as METTL14, WTAP and KIAA1429 have also been reported to be involved in the inclusion of this nucleotide modification (77). After mRNAs with m^6A residues are exported from the nucleus, the cytoplasmic proteins YTHDF1, YTHDF2 and YTHDF3 bind the RNA modification and mediate its effects in mRNA stability and translation (79, 80). The existence of an m^6A-remover nuclear protein (ALKBH5) suggests that m^6A can function as a dynamic mRNA modification that can be added or removed depending on the cellular environment (79, 80). Importantly, removal of m^6A in plant embryos is lethal and has also been shown to be aberrant in the determination of sex in *Drosophila*, highlighting its potential functional relevance (77).

1.4 HIV-1 Life Cycle

The HIV-1 life cycle starts with the binding of the viral particle to several proteins in the cellular membrane. After entry, the viral particle is partially uncoated and reverse transcribed. The viral cDNA is then imported to the nucleus where it is integrated into the host genome. Subsequently, the provirus serves as a template for the transcription of a full-length transcript that is processed into fully spliced, partially spliced, and unspliced RNAs. Since Rev is encoded in a fully spliced transcript is exported to the cytoplasm where it is translated into a protein. Once translated, it shuttles back to the nucleus and binds to intron-containing RNAs, permitting their export to the cytoplasm. The unspliced gRNA serves as a template for the translation of Gag and is also the viral genomic RNA (gRNA) that is packaged into the nascent virion. During or shortly after budding, the HIV-1 protease catalyses virion maturation, rendering it fully capable of propagating the infection (Figure 1.4).

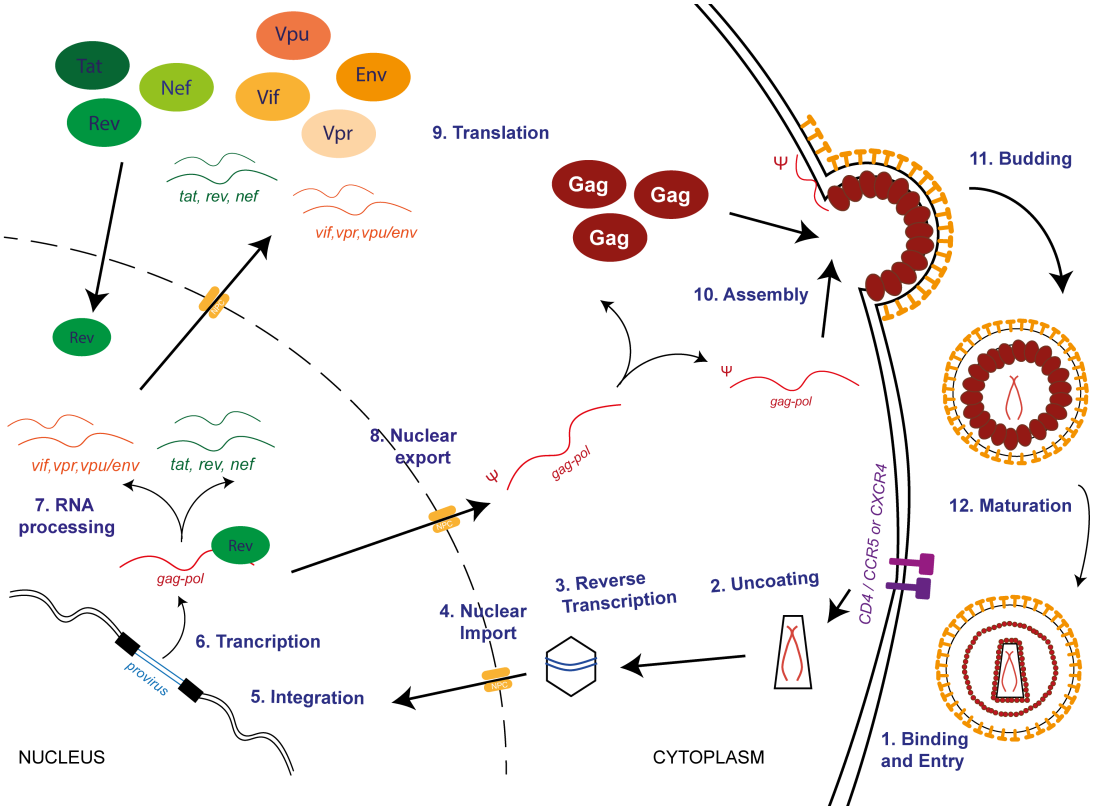


Figure 1.4. Schematic representation of the HIV-1 life cycle. Steps 1-10 are described in detail in sections 1.4.1-1.4.10.

1.4.1 Receptor binding and virus entry

During viral entry, the genome is transferred from the viral particle into the host cell. Enveloped viruses, such as HIV-1, achieve this by encoding and expressing proteins in their membrane that can undergo conformational changes that cause the cellular membrane and the viral membrane to fuse (83). The entry process can be classified in three stages: attachment, a triggering event, and membrane fusion (Figure 1.4, step 1).

During attachment, binding of HIV-1 to the primary receptor is fundamental. CD4 is the primary HIV-1 receptor and therefore this virus mainly infects CD4+ T cells or macrophages (84, 85). The primary receptor was discovered following observations that HIV-1 preferentially infected CD4+ cells and that expression of CD4 in non-permissive cells rendered cells susceptible to HIV-1 infection (84, 85). Viral attachment to the cellular membrane is mediated by the HIV-1 Env protein. Env is present in the HIV-1 particle surface as a trimer that contains three highly glycosylated gp120/gp41 heterodimers called Env “spikes” (86, 87). Studies in murine cells expressing CD4 showed that HIV-1 was able to bind the cell surface but unable to proceed with infection (85). These results led to the hypothesis that Env requires another protein to fuse with the cellular membrane (34). The gp120 subunit mediates the binding of the virus to CD4 in the cellular membrane, which leads to a conformational rearrangement that enables the second fundamental stage of binding, the co-receptor binding (88, 89). Co-receptor use is dependent on the V3-loop region within the gp120 subunit, and the most commonly found co-receptor in primary strains is CCR5, but some viruses also evolve to use CXCR4 (88). As a consequence of co-receptor binding, the gp41 subunit exposes a hydrophobic amino-terminal fusion peptide that interacts with the cell membrane through the formation of a 6-helix bundle, which brings the viral and cellular membranes together (53). Folding of the fusion peptide (gp41), therefore, enables the fusion and formation of the pore from which the virus accesses the cytosol of the host cell (87, 90). In order to the fusion

pore to form, multiple CD4 receptors, 4-6 CCR5 co-receptors and 3-6 Env trimmers are needed (88).

1.4.2 Uncoating

After internalisation of the virus into the host cell the viral particle is partially uncoated (Figure 1.4, step 2). This process is necessary for the subsequent import of the reverse transcription complex (RTC) into the nucleus, as the capsid core is 50-60 nm (91) and does not fit through the nuclear pore complex (NPC), which has an upper limit of around 39 nm (92, 93). There are several models by which core disassembly may occur, and they might not be mutually exclusive since it may vary by cell type or activation state of the infected host cell (64).

One mechanism supports the idea of a rapid core disassembly that happens immediately after viral entry. Several biochemical studies have shown absence of CA in comparison to MA or RT (94). Recent observations, however, have raised uncertainty around this model, given that the capsid core has been suggested to provide protection to the viral genome from antiviral cDNA sensor cGAS (41, 42). Besides, recent studies have shown that during uncoating a small amount of CA remains associated with the RTC (95) (Figure 1.5 A).

Another model supports the lost of CA from the viral RTC while reverse transcription is being carried out and traffics towards the nucleus. The cytoplasmic uncoating has been suggested after studies that have demonstrated that HIV-1 gradually becomes insensitive to TRIM-Cyp restriction, suggesting that CA is gradually disassembled from the conical core (96, 97). This idea, however, does not take into account the function of the capsid core on protecting the viral genome from the host antiviral mechanisms (Figure 1.5 B).

The last model supports that disassembly of the capsid core occurs during nuclear import of the viral RTC. This mechanism is mainly supported by the

fact that the viral cDNA generated during reverse transcription is not recognised by cytosolic DNA sensors (41, 42). Studies using electron micrographs have shown intact capsid cores at the NPC, supporting this mechanism (98). Nevertheless, these intact cores were found 48h post infection, a time-point that is later than core disassembly and reverse transcription, which suggests that this model requires further research to address this problem (Figure 1.5 C). In support with this model, single HIV-1 live imaging has revealed that CA protects HIV-1 complexes from degradation and enables docking at the nuclear pore (99).

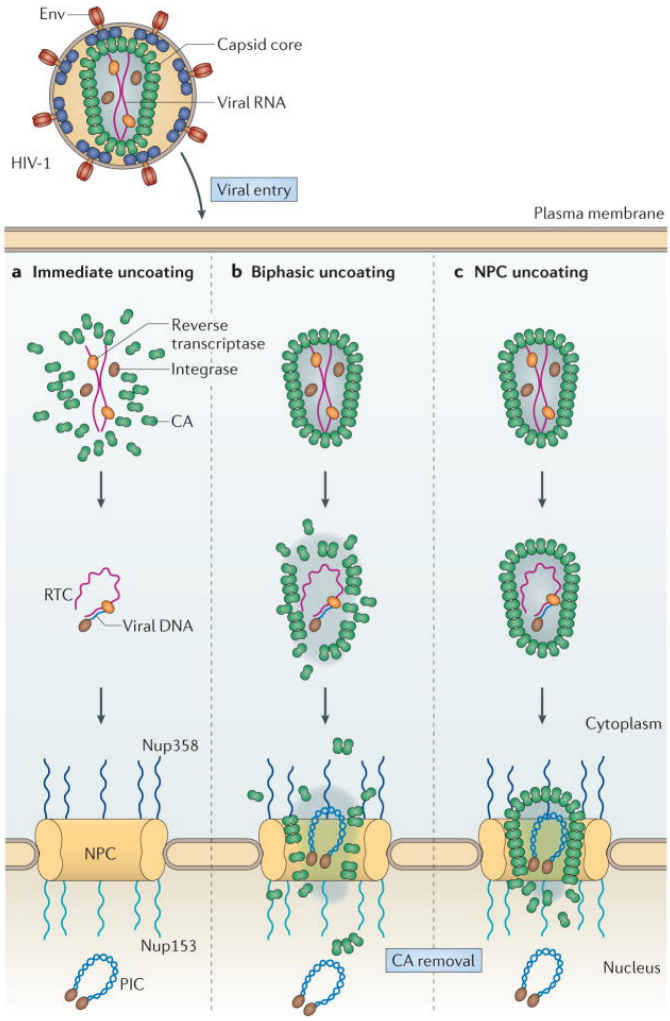


Figure 1.5. Models of uncoating. Schematic representation of the models described in section 1.4.2. **A)** Immediate uncoating **B)** Cytoplasmic uncoating **C)** NPC uncoating. Figure taken from (64) with the permission of Springer Nature, license number 4474701043924.

1.4.3 Reverse transcription

Reverse transcription is carried out by the virally encoded RT and converts the viral RNA genome into DNA (Figure 1.4, step 3). HIV-1 RT is formed upon cleavage of the Gag-Pol polyprotein by the viral protease (100). Structurally, RT is composed of two heterodimer subunits p66, which has the two enzymatic activities and p51, which plays an important role in maintaining the structural composition of the enzyme (101). Within the p66 subunit, the DNA polymerase activity copies either a DNA or RNA template and an RNase H degrades RNA when it is part of an RNA-DNA duplex (100). Several viral and cellular factors that facilitate the process, but the RT possesses the two enzymatic activities that are sufficient to complete the process. Although RT is sufficient to carry out reverse transcription *in vitro*, *in vivo* RT is part of the reverse transcription complex (RTC) (100). Several viral and cellular proteins have been proposed to be present in the RTC, but their relevance remains unknown (100, 102).

HIV-1 uses $tRNA^{Lys,3}$ as a primer, which is complementary to primer binding site (pbs) in the 5' end of the viral genome and reverse transcription starts towards the 5' end (Figure 1.6). Next, the RNase H degrades the viral RNA, exposing the newly synthesised minus-strand RNA. The ends of the viral RNA are direct repeats called R. These repeats allow the transfer of the newly synthesised minus-strand DNA to the 3' end of the viral RNA, where the minus-strand synthesis continues along the length of the genome. As DNA synthesis proceeds, so does RNase H degradation. While the tRNA determines the origin of the minus strand, the origin of the plus-strand is determined by a polypurine tract (ppt), which is resistant to RNase H activity. HIV-1 possesses one near the 3' end of the RNA and another one in the middle of the genome, where only the first one is essential for viral replication (103, 104). Synthesis of the plus-strand also leads to the generation of the first 18 nucleotides of the $tRNA^{Lys,3}$ primer. In addition, reverse transcription produces a DNA product that is longer than the viral genome from which it has been transcribed, called long terminal repeats (LTR). The LTR is formed

of the U3 (in the 3' end) and the U5 (in the 3'end), and after integration, they flank either end of the ssgRNA provirus (U3-R-U5) (100) (Figure 1.6).

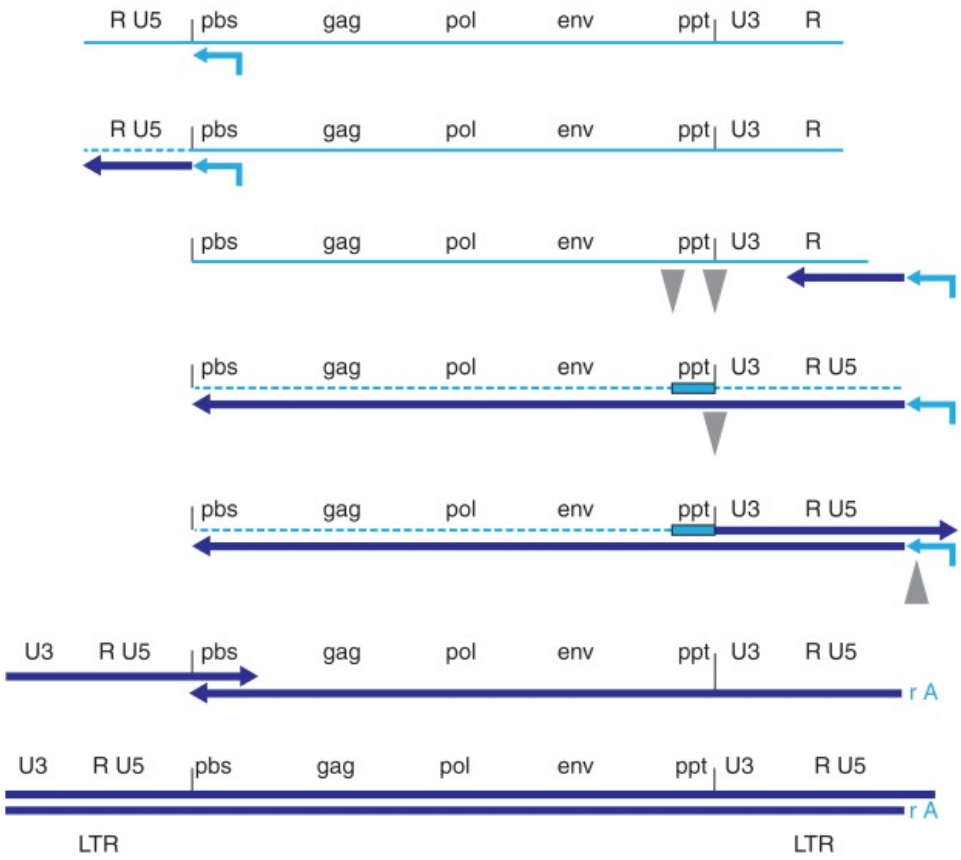


Figure 1.6. Schematic representation of HIV-1 reverse transcription. The tRNA primer base pairs near the 5'end and initiates conversion of the single-stranded RNA genome while the RNase H activity degrades the RNA template (dashed line). Subsequently, minus-strand transfer occurs between the R sequences, allowing minus-strand DNA synthesis and further RNA degradation. The ppt next to the U3 is resistant to the RNase H activity and serves as a primer for the synthesis of the plus-strand. Figure taken and adapted from (100) with the copyright permission from Cold Spring Harbor Laboratory Press.

1.4.4 Nuclear import

While most of the retroviruses require the breakdown of the nuclear envelope to infect cells, HIV-1 holds the capacity to access non-dividing cells (105) (Figure 1.4, step 4). Due to the size of the pre-integration complex (PIC) nuclear import is an energy-dependent active process that is facilitated by different viral and cellular factors (106, 107). Recent data shows that CA plays an important role in the nuclear import of the reverse transcribed viral genome (108). Moreover, multiple cellular proteins, including some components of the NPC have been shown to interact with CA (109-111). Not only this demonstrates that CA is necessary to drive critical steps during viral replication but confirms that CA remains associated with the RTC during nuclear import (112). Cellular factors such as the cleavage and polyadenylation specificity factor 6 (CPSF6), transportin 3 (TNPO3) and components of the NPC nucleoporin 358 (NUP358) and NUP153 have been described to be implicated in the process (102, 110).

Genetic evidence demonstrates that CA is the major viral determinant in nuclear import. As for the cellular factors, NUP258, transportin 3 (TNPO3) and NUP153 have been identified in siRNA screens and shown to participate in nuclear import (109, 110, 113, 114). NUP358 (also called RANBP2) and NUP153 are components of the nuclear pore (115, 116). While NUP358 projects towards the cytoplasm, NUP153 is localised towards the nuclear side of the NPC. NUP358 depletion inhibits viral replication by preventing efficient nuclear import of the RTC (106, 115, 116). This nucleoporin also contains a CypA-homology domain, which has been demonstrated to interact with CA. Cyclophilin A (CypA) is a peptidyl-prolyl isomerase that binds to an exposed loop on the surface of the capsid, and it has been suggested to direct the viral DNA into nuclear entry. The interaction between CA and NUP153, on the other hand, is mediated by binding pockets in CA that enable interactions with other nuclear-import pathway factors (117). Depletion of NUP153 has been associated with a reduction of 2-LTR circles, which are considered a sign of the arrival of PIC to the nucleus (117, 118).

TNPO3 is a member of the importin-β family of proteins. HIV-1 replication has also been shown to be dependent on TNPO3 by inhibiting 2-LTR circle formation (119), and CA is also a determinant of its function in nuclear import (120). CPSF6 is a mRNA-processing protein that shuttles between the nucleus and the cytoplasm. The role of CPSF6 in HIV-1 replication was identified on a study where a truncated form of CPSF6 was shown to attenuate infection by HIV-1 (121). HIV-1 acquires resistance to CPSF6 action through a mutation in CA that impairs binding (122).

1.4.5 Integration

Integrase (IN) is encoded in the *pol* gene as a result of the Gag-Pol polyprotein cleavage by the viral protease. It is composed of three structurally and functionally distinct domains: the amino-terminal zinc-binding domain (NTD), the central catalytic core domain (CCD) and the C-terminal domain (CTD). As explained above, upon reverse transcription in the cytoplasm, the newly generated viral cDNA forms a PIC, which includes viral DNA, integrase, capsid proteins and some cellular proteins and is imported to the nucleus for its subsequent integration (Figure 1.4, step 5).

Integration of retroviral genomes can theoretically happen throughout the host genome; nevertheless, certain regions are preferentially targeted. High gene density and transcriptionally active genes, proximity to specific genomic elements (such as transcription start regions) and the primary DNA sequence itself are some of the determinants in the process of integration (123-125). Some host factors such as LEDGF (126) and CPSF6 (127) have also been shown to play an active part in the process by guiding integration towards highly spliced transcriptional units and intron-rich genes (123). Palindromic consensus sequences have also been proposed to present a target for integration (128-130). Conversely, a statistical analysis of palindromic consensus sequences throughout a population has shown that these sequences are not really representative of the population (131). It has been suggested that these palindromic sequences are instead sequences that

arise in ‘forward’ and ‘reverse complement’ orientations in the viral genome that form an integration site motif (131).

Integrase carries out two essential enzymatic reactions: 3’end processing of the viral DNA and strand transfer reaction. In the PIC, IN multimerizes on the newly synthesised DNA to form the intasome, a functional integrase-viral DNA complex (132, 133). In the first step each end of the HIV-1 LTR is cleaved and in a region next to the invariant dinucleotide sequence d(C-A), generating a linear viral DNA and with reactive 3’-hydroxyl ends. After import into the cell nucleus, the intasome binds onto the target DNA to form a target capture complex (TCC) and proceeds with the second step of integration (133, 134). Integrase uses the 3’- hydroxyl groups on the ends of the viral DNA to attack a pair of phosphodiester bonds on opposite strands of the target chromosomal DNA, across the major groove. Complementary strands of viral DNA are then joined with the 5’phosphates at the ends of the target DNA. The resulting DNA recombination intermediate contains single-strand gaps and the two-nucleotide overhang at the 5’ends of the viral DNA, which must be repaired by cellular enzymes to complete integration (133, 134).

1.4.6 Transcription

The HIV-1 long terminal repeat (LTR) acts as the viral promoter and is segmented into the U3, R and U5 regions (54). The LTR contains several upstream DNA regulatory elements that serve as binding sites for cellular transcription initiation factors (54). The core promoter is a powerful and highly optimised promoter formed of three tandem SP1 binding sites (135), an efficient TATA element (136), and a highly active initiator sequence (137) (Figure 1.4, step 6).

The requirement of Tat for HIV-1 gene expression came following experiments in which they showed that expression of reporter genes under the LTR promoter was dependent on a trans-acting factor, which was later named Tat (54, 138, 139). Deletion analysis of the LTR showed that, in order

to function, Tat had to bind to a regulatory element named the transactivation-responsive region (TAR), a ~50bp element located adjacent to the 3' to the transcription start site. Moreover, other studies showed that mutations in TAR, while not interfering with Tat binding, they inhibited transactivation. These experiments led to hypothesise that Tat required an essential cofactor called the positive transcription elongation factor b (P-TEFb(140)) (54).

P-TEFb contains a kinase subunit called CDK9 and the cellular factor cyclin T1 (CycT1). Tat recruits P-TEFb by interacting with CycT1 and the kinase Cdk9 (58, 140). Binding of Tat to Cdk9 induces a conformational change that activates the kinase activity and results in the phosphorylation of several cellular factors. Some of the factors include inhibitors of transcription elongation, such as negative elongation factor (NELF), which upon phosphorylation dissociates from TAR enabling elongation (141). P-TEFb also phosphorylates the carboxy-terminal domain of the cellular RNA polymerase II (RNAP II), which transforms RNAP II into transcription competent, and promotes efficient transcription by the cellular RNAP II (58, 142-144). The integrated HIV-1 genome is transcribed into a genome-length RNA (58, 61).

1.4.7 RNA processing and alternative splicing

The full-length ~9 kb viral genome is spliced inefficiently to produce the multiple transcripts that encode all the viral proteins (145-148). HIV-1 employs alternative splicing to produce the transcripts that encode for the accessory proteins Vif, Vpr, Vpu, and Env, which are encoded in intron-containing mRNAs. The structural protein Gag-Pol is encoded in an unspliced mRNA (54, 149) (Figure 1.4, step 7).

Most HIV-1 strains use four different splice donor or 5' splice sites (D1, D2, D3, D4) and eight different acceptor or 3' splice sites (A1, A2, A3, A4a,b,c, A5 and A7) to produce more than 40 different spliced mRNA species in the

infected cell (54). In each of the spliced mRNAs, the major splice donor (D1) is spliced to one of the 3'splice sites. As a result, all HIV-1 mRNAs include the highly structured noncoding exon 1 that extends from the 5'cap to 5'ss D1 while the different acceptor sites are used to generate the diverse pool of mRNAs encoding for the viral proteins Vif (A1), Vpr (A2), Tat (A3), Rev (A4a,b,c), Vpu/Env (A4a,b,c, and A5) and Nef (A5 and A7) (149, 150). Because 50% of the HIV-1 transcripts are required to remain unspliced so that they can encode for Gag and Pol and can provide the viral genomes, splicing is a substantially inefficient process (54, 151). Alternative splicing is carried out by the spliceosome, and the process is controlled by exonic splicing enhancers (ESEs) and intronic splicing enhancers (ISEs), which promote splice site recognition and are selectively bound by RNA binding proteins such as SR proteins family (152, 153). There are also intronic and exonic splicing silencers (ISSs and ESSs), which inhibit splicing and are typically bound by specific members of the heterogeneous ribonuclear protein family (hnRNPs) (54, 153).

1.4.8 Nuclear export

Almost all cellular mRNAs are transcribed as intron-containing pre-mRNAs. Eukaryotic cells, however, do not normally permit the export of intron-containing mRNAs and are typically degraded in the nucleus. These introns are recognised by splicing factors that commit the pre-mRNA to the splicing pathway and prevent the pre-mRNAs from exiting the nucleus until the introns are removed (58, 154).

HIV-1 mRNAs rely entirely on the cellular machinery to carry out splicing and, therefore, intron-containing transcripts are retained in the nucleus. The mechanism that HIV-1 has evolved to enable the export and synthesis of the proteins encoded in intron-containing mRNAs is dependent on Rev (Figure 1.4, step 8), which is encoded in a fully spliced mRNA. After its synthesis in can shuttle back to the nucleus (155) where it binds to unspliced or partially spliced mRNAs to permit their export to the cytoplasm where they can be

translated (58). The binding is mediated by a highly structured RNA element of 351 nt located in the HIV-1 *env* gene called Rev-response element (RRE), and is present in all 9- and 4-kb mRNAs (70, 151, 152). The export to the cytoplasm is mediated by the binding of Rev to a cellular factor called CRM1 through a 10 amino acid leucine-rich nuclear export signal (NES), which directs the HIV-1 transcripts to the nuclear pore complex (58). CRM1 binds its cargo in the nucleus in the presence of a GTP-bound form of the Ran GTPase. After nuclear export hydrolysis of the bound GTP to GDP causes disassembly of the complex and cytoplasmic release of the intron-containing mRNA. Conversely, simple retroviruses encode constitutive transport elements (CTEs) (156) that directly recruit components of the NXF1 mRNA nuclear export pathway. Interestingly, exchanging the canonical RRE HIV-1 export pathway to the CTE/NFX1 pathway has been shown to restore efficient virion production in mouse cells. These results raise the possibility of modulating the cytosolic fate or activity of proteins through modification of the export pathway (157).

1.4.9 Translation

In eukaryotes, translation normally starts upon recognition of the start codon via scanning (Figure 1.4, step 9). This model states that recognition of the mRNA's m^7G cap structure by the initiation factor eIF4F, which consists of the cap-binding protein eIF4E, the RNA helicase eIF4A and eIF4G results in the initiation of translation (158, 159). The role of this third protein is to facilitate the recruitment of the 43S pre-initiation complex (PIC). The PIC is formed by the 40S ribosomal subunit, loaded with the Met-tRNA and other initiation factors (eIFs) and after binding the activated mRNA at the 5'end it migrates along searching for the translation initiation AUG codon (158). After scanning, the 60S ribosomal subunit joins the complex to form the 80S subunit leading to the start of elongation (57, 160). Cap-dependent translation is efficient in mRNAs with short, unstructured 5' untranslated region (UTR). The retroviral RNA can also interact with multiple cellular and viral co-factors that can enhance cap-dependent translation such as the post-

transcriptional control element (PCE). Several proteins have been reported to promote HIV-1 translation through binding to its 5'UTR. Some of these proteins include TRBP, Staufen, DHX9 and DDX3 where the last two are the best characterised of all (159). DHX9 is an RNA helicase that binds to the post-transcriptional control element located in the 5'UTR and promotes polysome association of the HIV-1 gRNA and virion infectivity (161). RNAi knockdown of this protein results in decreased expression of Gag, Vif, Nef and Rev (162).

HIV-1, however, contains a highly structured 5'UTR, which may impede ribosome scanning and reduce the efficiency of translation initiation. As a consequence, HIV-1 has evolved several strategies to enable the synthesis of all the proteins from a primary transcript. DDX3 is a DEAD-box RNA helicase that has been shown to unwind the 5'-terminal RNA structure. DDX3 binds both TAR and the eIF4F complex, which allows the entry of the 43S PIC (159, 163, 164). Internal ribosome entry sequences (IRESes) constitute another mechanism by which HIV-1 may circumvent these obstacles (165). IRESes constitute an internal ribosome binding site that directly recruits the translation machinery (57, 166) and enables the binding of the translation initiation complex to the open reading frame in a cap-independent manner. Indeed, translation of Gag, Vif, Vpu, Vpr and Nef are expressed as a result of internal translation (167). Two different IRESes have been reported in the HIV-1 genome; one is located within the 5'UTR and another one in *gag* (159, 168, 169).

In order to express multiple viral proteins from a single initial transcript, HIV-1 has evolved several mechanisms. Leaky scanning provides HIV-1 Vpu and Env protein translation from a single bicistronic vpu-*env* mRNA, and there are currently two models explaining how these are translated. The first model states that Env is translated through leaky scanning because of a weak Kozak context around the *vpu* AUG and, therefore, initiation codon is passed by the ribosome to reach the AUG in *env* (159). The second model states that Env is the result of a ribosome shunt that induces discontinuous scanning (159). HIV-1 protease, integrase and reverse transcriptase,

however, are encoded in the Pol ORF and synthesise as a result of a frameshift. Both an upstream heptanucleotide slippery sequence (UUUUUUA) and an RNA stem-loop pseudoknot structure downstream of a stimulatory signal are necessary to induce the ribosome to shift to the overlapping pol ORF (a-1 reading frame change) (159) (Table 1.1).

Table 1.1. Schematic summary of the mechanisms involved in the regulation of HIV-1 gene expression

Mechanism	Effect on translation
Leaky scanning	Readthrough of the upstream start codon (AUG) that allows initiation of translation of a downstream ORF.
Frameshifting	Prevents from terminating translation at the end of the open reading frame inducing ribosomes to shift into the overlapping ORF.
Incomplete splicing	Mechanism by which protein expression is regulated. In the process, splicing of the primary transcript occurs forming several distinct mRNAs that generates protein diversity.
IRES	Provides an internal ribosome binding site that circumvents the structural impediment that the HIV-1 5'UTR may present.

1.4.10 Virus assembly, release and maturation

The assembled virion packages all of the components required for infectivity. These include two copies of the genomic RNA, cellular $tRNA^{Lys,3}$, Env, Gag, and the three viral enzymes (protease, integrase and reverse transcriptase). HIV-1 virion assembly occurs at the plasma membrane and Gag plays a central role (170, 171) (Figure 1.4, step 10).

The N-terminal MA domain binds to the plasma membrane where it recruits Env. This protein consists of an N-terminally myristoylated globular head and a largely a-helical stalk, which is responsible for directing Gag to the cell membranes. CA mediates the protein-protein interactions required for immature virion assembly and forms the capsid core (64, 91, 170, 171). CA comprises an N-terminal domain that is positioned on the outer face of the conical core and a C-terminal domain that is oriented in the inner face of the core. The N-terminal domain of CA contains three α-helices that stabilise the hexameric subunits of CA. In addition, CA-CA interactions further stabilize the subunits and lead to the formation of binding pockets that enable the interaction with several cellular factors that take part in viral replication (91, 112). The conical core fulfils two major functions. First, it is essential to maintain the HIV-1 enzymes, the HIV-1 genome and the viral accessory proteins in a closed environment. Second, it protects the viral genome from host factors that recognise cytosolic DNA (64). The NC domain binds the RNA to enable its packaging into the nascent virion. Simultaneous interactions between a single RNA molecule and multiple Gag molecules help to drive Gag multimerisation. P6 contains docking sites for several proteins such as bind TSG101 and ALIX proteins from the ESCRT pathway. On the other hand, SP1 and SP2 help to regulate the conformational changes that accompany viral maturation (170, 171).

A population of Gag monomers or oligomers are distributed in the cytoplasm and can diffuse or be transported to the cytoplasm. However, given that the plasma membrane represents only a small fraction of the total membrane

present in a cell there must be a signal that HIV-1 recognises for its assembly. Indeed, the plasma membrane contains a phosphoinositide called $PI(4,5)P_2$ that is recognised by the globular head of MA (172). The interaction between MA and $PI(4,5)P_2$ leads to a conformational change that favours the exposure of the MA myristate. Both, the binding to $PI(4,5)P_2$ and the myristate exposure favour multimerisation and a series of protein-protein and protein-lipid interactions that stabilise Gag interaction with the membrane. These events result in a growing Gag protein shell at the plasma membrane (170, 171).

In order to complete assembly, L-domains (late-budding domains) directly or indirectly bind to components of the ESCRT pathway, which enables virion scission. In HIV-1, Gag p6 domain contains the L domains and consists of peptide sequences PTAP and YPXL, which bind Tsg101 (173) (a component of ESCRT-1) and ALIX (an ESCRT-I And ESCRT-III-binding protein). L-domain binding leads to a conformational change in ALIX, which induces dimerisation and activates ALIX for membrane binding and ESCRT-II recruitment (174, 175). Upon activation of ALIX, the Bro domain binds carboxy-terminal helices located within the tails of all three human CHMP4 proteins (176). This binding relieves CHMP4 auto-inhibition, which results in polymerisation into filaments within the virion neck (174, 175, 177-179). Subsequently, CMPH4 recruits CHMP2 proteins (180), which as a result exposes the carboxy-terminal tail. This domain contains a helical sequence motif that binds the amino-terminal MIT domains of VPS4 ATPases (181-183). CHMP4 subunits together with CHMP2 and other ESCRT-III proteins form a ring filament within the neck of the budding virus, that as they spiral inward, they constrict the ring resulting in fission (170) (Figure 1.4, step 11).

Finally, during or shortly after budding the HIV-1 protease cleaves the Gag and Gag-Pro-Pol polyproteins at ten different sites producing the fully processed MA, CA, NC, p6 and the viral enzymes PR, RT, IN proteins. Upon maturation, these proteins undergo a dramatic rearrangement that results in the mature infectious virus (Figure 1.7). While MA remains associated with the inner side of the viral membrane, approximately 1500 (65) copies of CA

form the outer conical capsid shell (170, 171). Cleavage of the $p55^{Gag}$ polyprotein by the HIV-1 protease also induces a conformational change in Env, which enables its full fusogenic potential (170, 184) (Figure 1.4, step 12 and Figure 1.7).

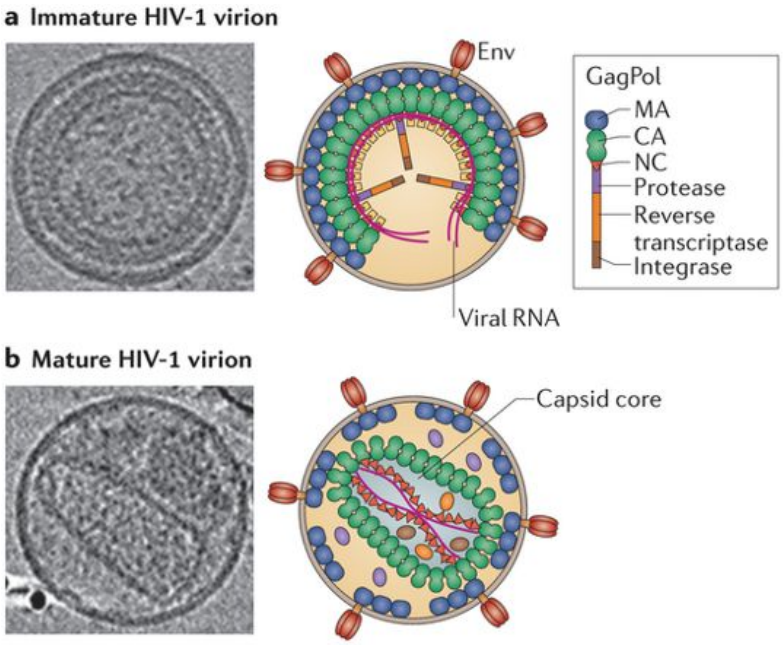


Figure 1.7. HIV-1 virion maturation. Schematic representation of the structural differences between a mature and an immature virion. Upon maturation, MA remains associated with the inner side to the plasma membrane and CA oligomerises to form a conical structure that serves as protection for the two copies of the gRNA, PR, RT, IN, Vpr and some cellular proteins. **A)** Cryoelectron tomogram and illustration of an immature virion **B)** Cryoelectron tomogram and illustration of a mature virion. Figure taken and adapted from (60) with the permission of Springer Nature, license number 4474701466664.

1.5 HIV-1 accessory proteins and restriction factors

HIV-1 requires the assistance of multiple cellular proteins at the different steps of the life cycle to achieve efficient replication (114). Conversely, host cells also express proteins that function as the initial line of defence against infection to suppress virus replication. These proteins are named restriction factors and usually are a component of the innate antiviral responses (185). However, HIV-1 is not capable of replicating in some nonhuman cells (e.g. rhesus) due to the presence of restriction factors, which marks these proteins as cross-species transmission determinants (2). The mechanism by which HIV-1 evades from restriction factors relies mostly on the accessory proteins encoded in the viral genome (Vif, Nef, Vpu and Vpr).

1.5.1 APOBEC3G and Vif

Apolipoprotein-B mRNA editing enzyme catalytic polypeptide-like 3G (APOBEC3G, A3G) is a cellular protein that inhibits HIV-1 replication and is counteracted by the HIV-1 accessory protein Vif (185, 186). A3G was discovered when viruses deficient in Vif (Δvif) were found non-infectious in non-permissive cells (primary human T cells, HUT78 and CEM). On the contrary, many other cell lines, such as SupT1, CEM-SS and 293T, support the production of fully infectious HIV-1 Δvif. These findings suggested the presence of an activity that inhibits HIV-1 replication in the absence of Vif (186, 187) and the hypothesis was later supported by experiments carried out with transient heterokaryons (187). The fusion of a non-permissive and a permissive cell resulted to be non-permissive to HIV-1 Δvif infection, which suggested the presence of a dominant cellular factor in non-permissive cells that was leading to the block in infection (187). This restriction factor was later identified as A3G (188).

A3G is a member of a family of vertebrate proteins with polynucleotide cytidine deaminase activity that edits the nucleotide sequence. Structurally,

A3G contains five α helices including two that form the zinc-coordinating deaminase domain and are located over a platform of five β strands (189). While the C-terminal domain catalyses the hypermutation, the N-terminal domain's activity is necessary to package A3G into HIV-1 virions. Incorporation of A3G in virions occurs during virion production from the host cell (185). Upon viral infection, the optimal location enables A3G to edit up to the 10% of cytidines in the DNA. As a consequence of this activity, during RT, the C-to-U changes translate into the incorporation of A instead of G in the proviral plus-strand DNA genome (190), which can be deleterious to the virus. The excessive mutations result in substantially decreased dsDNA levels and ultimately in the inhibition of viral replication (188, 191). Moreover, other studies showed that A3G also inhibits the translocation of the RT along the RNA template thereby interfering with the processivity of reverse transcription (191, 192). Importantly, a recent study has brought new insight into the mechanism by which reverse transcription is inhibited proposing an interplay between A3G, RT and the cellular DNA repair machinery (193). More specifically, this new data has shown that inhibition of reverse transcription is site- and sequence-independent and that A3G directly interacts with the RT. They also propose that inefficient recognition of the deaminated cDNAs by the cellular uracil base excision repair (UBER) enzymes may suggest its role as a pathogen-associated molecular pattern (193).

The viral accessory protein Vif antagonises the antiviral activity of A3G. Vif interacts with APOBEC3 proteins through its amino terminal domain and directs it to a cellular ubiquitin ligase complex, which results in polyubiquitination and proteasomal degradation (194). The interaction induces the recruitment of cullin 5 (CUL5) through to the cytokine suppressor signalling (SOCS) box and the substrate adaptors elongin B/C heterodimer (ELOB and ELOC) (195, 196), and its Z domain to cullin5 (185, 197, 198). Several members of the APOBEC protein family have been shown to inhibit HIV-1 replication upon overexpression. At more relevant levels of expression, however, only A3F, A3H and A3D have also been shown be packaged into

virions and to provoke a significant suppression of HIV-1 replication (199, 200).

1.5.2 TRIM5α and capsid

Genetic studies of susceptibility of mice to MLV-induced pathogenesis identified Friend-virus susceptibility gene 1 (Fv1), which blocks replication at the early stages of the viral cycle. Although Fv1 is only present in mice, it later became clear the selective restriction activity towards MLV was present in other species, including humans (201). A screen for rhesus macaque genes that could inhibit HIV-1 infection when expressed in human cells resulted in the identification of TRIM5α (202). HIV-1 efficiently enters the cells of Old World monkeys but encounters a block before reverse transcription (203), whereas SIV can efficiently infect these cells. This suggests that TRIM5α proteins are poor inhibitors of retroviruses found naturally in the same host species, but they are quite often active against retroviruses that are found in non-natural species. Thus, TRIM5α functions as a good barrier to cross-species transmission of primate lentiviruses (204). This species-specific restriction is dependent on HIV-1 capsid (205).

TRIM5α is one of a family of tripartite motif (TRIM)-containing proteins. The TRIM domain is composed of a RING domain and a B-box domain in the N-terminal region. The C-terminal domain contains a PRYSPRY motif. Finally, a coiled-coil domain that enables the dimerisation links the amino- and carboxy-terminal domains.

Although it has been shown that TRIM5α inhibits viral replication soon after viral entry, the mechanism by which this restriction is mediated is still unclear (202). It has been proposed that the PRYSPRY domain enables TRIM5α recognition and accelerated fragmentation of HIV-1 CA (206-208). Alternatively, it is possible that TRIM5α competes with cellular factors that promote infection for capsid binding. The abundant host cell chaperone protein, cyclophilin A (CypA) also binds CA and contributes to early events in viral replication. It has been shown that Cyp A is an auxiliary but not

necessary cofactor for the restriction of HIV-1 replication by TRIM5α (209, 210). It has also been reported that TRIM5α accelerates uncoating in the post-entry phase of the viral cycle resulting in the disruption of the RTC architecture and inhibiting reverse transcription (185).

In addition to blocking HIV-1 replication through the mechanisms described above, TRIM5α has also been shown to function as an innate immune sensor (211). Upon retroviral infection, TRIM5α interacts with the capsid lattice which leads to the synthesis of K63-linked ubiquitin chains that activate TAK1 and subsequently induce AP-1 and NFκN signalling. This signalling results in the activation of transcription of pro-inflammatory genes that further enhances the antiviral immune response (211).

1.5.3 Tetherin and Vpu

Viral protein U (Vpu) deficient HIV-1 showed significantly lower infectivity when compared to the WT counterparts (212). In the absence of Vpu, HIV-1 particles undergo normal assembly and ESCRT-protein-mediated fission from the plasma membrane, but the virions remain sequestered at the surface of the infected cells (213). However, this observation was only true for some cell lines while in others Vpu was completely dispensable, which suggested that this phenotype was caused by a cellular protein (185). Tetherin was discovered upon gene expression analyses on RNA extracted from either restrictive or non-restrictive cell lines, in the presence or absence of IFN (214). This restriction factor was also discovered by another group as a result of a proteomics analysis where it had been shown to be the target of the Kaposi's sarcoma-associated herpesvirus (KSHV) (215, 216).

Tetherin is a type II single-pass transmembrane protein that contains a transmembrane anchor domain and a glycophosphatidylinositol (GPI). Tetherin arrests the virion in the cell surface by introducing the TM domain into the assembling virion while the GPI domain remains in the plasma membrane. Both domains are necessary for this inhibitory activity, and it

seems that the target of tetherin is a host-derived lipid envelope rather than a viral protein. The HIV-1 Vpu protein antagonises this restriction factor (214, 216). Although the exact mechanism is still unclear, Vpu has been reported to colocalise and co-immunoprecipitate with tetherin. Although it has been proposed that Vpu reduces the level of tetherin in the cell surface (216) it is not clear whether it is being degraded (by the proteasome or the endolysosomal system) given that this down-regulation is modest in some cell types where Vpu is fully functional (185).

In addition to its role in sequestering the nascent virion, tetherin has been shown to act as a sensor and signal to the immune system. Upon restriction of Vpu-defective HIV-1, tetherin acts as a virus sensor and induces NF-κB-dependent proinflammatory gene expression. Retention of the virions in the cell surfaces results in phosphorylation of Tyr residues in the CT domain of tetherin dimers. This results in the recruitment of spleen tyrosine kinase (Syk) which is required for downstream activation of NF-κB. This kinase triggers a pro-inflammatory signal cascade that involves TAK-1 and TRAF6 and ends with NF-κB activation (217, 218).

In addition to enhancing the release of progeny virions from infected cells, Vpu also induces the degradation of the CD4 receptor. The accessory protein Nef also mediates the removal of CD4 in infected cells, although they function at different stages. While Nef operates in the early stages following infection, Vpu acts in the late stages. Vpu is expressed in the ER where it interacts with a membrane-proximal domain of the cytoplasmic tail of CD4 (219). For degradation to occur, Vpu must be phosphorylated in its cytoplasmic tail, which facilitates the binding of βTrCP (220). These events subsequently result in the recruitment of Skp1-Cullin-F-Box (SCF) E3 ubiquitin ligase complex, a member of the ubiquitination machinery that ubiquitines CD4 directing it to the proteasomes for degradation (62, 221).

More recently, Vpu has also been shown to suppress the immune response to HIV-1 infection. In particular, it has been reported to inhibit the antiviral activity of NK cells and CD8+ cytotoxic T lymphocytes (CTLs) by interfering

with the activation of their cytotoxic function by multiple proteins, such as HLA-C, CD1d and CCR7 (222-224). Vpu mediated downregulation of HLA-C decreases the ability of HLA-C restricted CTLs to suppress viral replication of CD4+T cells (222). Inhibition of the surface expression of CD1d in infected DCs, on the other hand, is the result of a Vpu-mediated block in the recycling of CD1d in the endosomal compartments, which enables evasion of the immune response (223). The chemokine receptor CCR7 is also downregulated from the surface of CD4+ T cells and results in impaired migration and chemotactic signalling within CD4+ T cells (224).

1.5.4 SERINCS and Nef

SERINC3 and SERINC5 are members of a family of transmembrane proteins. SERINC3 was identified in a proteomic analysis of virions produced by cells infected with wild-type and Nef-defective viruses (225). SERINC5, on the other hand, was identified upon infections of 31 human cell lines with wild-type and Nef-defective HIV-1 and using a global transcriptome analysis (226). Subsequently, siRNA and CRISPR-Cas9 experiments resulted in increased infectivity of Nef-defective HIV-1, validating the results.

The precise mechanism by which SERINC proteins mediate HIV-1 restriction remains to be characterised. Both proteins have been shown to mediate this effect since targeting SERINC3 and SERINC 5 enhanced the individual effects. Incorporation of these restriction factors into virions reduces viral infectivity by inhibiting the enlargement of the fusion pore, a process that might happen through inducing functional inactivation of some Env glycoproteins by SERINC5. SERINC5 is targeted to late endosomes together with Nef in an AP-2-dependent manner. Nef efficiently down-regulates both proteins from the cell surface, preventing their incorporation into the nascent virion, and therefore counteracting the inhibitory effect in viral replication (225, 226). In addition, it has been shown that Env proteins from some HIV-1 strains are able to overcome SERINC5 inhibition without preventing SERINC5 incorporation into the virion. Introduction of SERINC5 into the

virions, however, makes HIV-1 more sensitive to broadly neutralising antibodies, which suggests that Nef might only be needed under this pressure (227, 228).

Nef is a 27 kDa myristoylated protein that is abundantly produced during the early phase of the viral replication cycle and is highly conserved among primate lentiviruses (62). In addition to counteracting SERINC-mediated antagonism, Nef modulates the expression of the cell surface proteins CD4, CD8, CD28 (229), major histocompatibility complex class I (MHC-I) (230) and class II (MHC-II) proteins. Downregulation of CD4 and CD28 is mediated by an interaction between the dileucine-based motif in the second disordered loop of Nef and the adaptor protein (AP) which is part of the cellular endocytosis machinery (231). These surface proteins can also be downregulated by accelerated endocytosis via clathrin-coated pits followed by lysosomal degradation. Binding of Nef to the phosphofurin acidic cluster sorting protein 1 (PACS-1), on the other hand, facilitates the degradation of MHC-I (62, 232). This auxiliary protein also interferes with the cellular signal transduction pathways. The myristoyl on the N-terminal domain and the proline-rich SH3 domain mediate the association with lipid rafts, cholesterol-rich microdomains that concentrate potent signalling mediators. Nef has also been found to interact with and activate serine/threonine protein kinase PAK, which might contribute to the activation of the infected cell (62). It has also been found to enhance virion infectivity (233) and viral replication by inducing actin-remodelling and promote the movement of the viral core to a potentially obstructive cortical actin barrier (62, 234). Finally, Nef also regulates cholesterol trafficking in infected cells, a lipid known to be important during HIV-1 assembly, budding and infection (235). Low cholesterol levels in the plasma membrane greatly reduce the formation of new viral particles; Nef can bind newly synthesised cholesterol and transport it to the plasma membrane where budding is occurring (62, 236).

1.5.5 SAMHD1 and Vpx

HIV-1 Infection of monocytes, dendritic cells (DCs) or non-activated CD4+ T cells is non-permissive (237, 238). In contrast, HIV-2 and SIVmac can efficiently infect myeloid cells, and this is associated to the HIV-2 encoded viral protein X (Vpx) (239, 240), since infection with HIV-2 and SIVmac of myeloid cells, substantially increased in the presence of Vpx (43, 241, 242). Moreover, Vpx was shown to associate with an E3 ubiquitin ligase complex, which suggested that, like the HIV-1 accessory protein Vif, Vpx could target a cellular restriction factor to degradation (240). This discovery led to the search of a cellular restriction factor that is present in cells that are inefficiently infected by HIV-1 and is antagonised by Vpx. Sterile alpha motif domain and HD domain-containing protein-1 (SAMHD1) expression in myeloid cells was shown to be the cellular factor preventing these cells from HIV-1 replication. Expression of Vpx or silencing of SAMHD1 relieved the restriction to HIV-1 infection, which confirmed SAMHD1 as a restriction factor (243, 244).

SAMHD1 is a restriction factor that is expressed in cells of the myeloid lineage that inhibits an early step of the viral life cycle (243). More specifically, SAMHD1 is a potent dGTP-stimulated triphosphydrolase that converts deoxynucleoside triphosphates to the constituent deoxynucleoside and inorganic triphosphate. Some studies have reported that as a consequence of the depletion of the majority of the cellular dNTPs, reverse transcription of the viral cDNA is inhibited, resulting in restriction of HIV-1 replication (245, 246).

Later findings, however, revealed that the ability of SAMHD1 to restrict infection is not related to its function on decreasing the cellular dNTP levels (43, 247). SAMHD1 is expressed in both dividing and non-dividing cells but the ability to restrict viral activity is limited to cycling cells. This difference in activity is modulated by the phosphorylation of residue T592 in the cdk1-target motif. Phosphorylation has been shown to interfere with the restriction

of viral infection but does not affect the ability of SAMHD1 to decrease cellular dNTP levels (247).

Mutations in SAMHD1 have been shown to be involved in Aicardi-Goutières syndrome (AGS), a genetic encephalopathy with symptoms mimicking congenital viral infection and it has been proposed to act as a negative regulator of the interferon response (248). AGS patients present mutations in genes involved in nucleic acid metabolism such as SAMHD1 and TREX and are characterised by high levels of pro-inflammatory cytokines (249). Interestingly, TREX has been shown to inhibit the innate immune response to HIV-1 by digesting the excess of viral DNA that would otherwise activate cGAS pathway (250). SAMHD1 restriction of HIV-1 replication in myeloid cells, on the other hand, is regulated by phosphorylation since SAMHD1 phosphorylation-defective mutant was able to restrict viral replication. Moreover, phosphorylation is linked to the antiviral activity since treatment of cells with IFN-I reduced phosphorylation (251).

HIV-2 overcomes SAMHD1-mediated restriction by expressing the accessory protein Vpx. Vpx directs SAMHD1 to degradation through the recruitment of an E3 ubiquitin ligase complex that contains cullin 4 (CUL4), DNA damage-binding protein 1 (DCAF1) and DDB1-CUL4-associated factor 1 (DDB1). The resulting complex regulates the degradation of a large number of cellular DNA repair proteins, transcription factors and replication enzymes (241). Overall, these studies suggest that HIV-2 needs to block SAMHD1 function through Vpx in order to replicate efficiently. Intriguingly, SAMHD1 is not restrictive in activated CD4+ T cells (241), the main target of HIV-1 infection, and HIV-1 does not express an accessory protein that antagonises this restriction factor.

1.5.6 MX2

Myxovirus resistance 2 (MX2 also called MXB) is an IFN-inducible dynamin-like GTPase that localises to the nuclear envelope, NPCs and cytoplasmic punctae and inhibits HIV-1 replication (252-254). Multiple factors have been proposed to inhibit viral replication upon stimulation with type I IFN (255-259). Treatment of macrophages, THP1 cells and primary CD4+ T cells with IFN-α, for example, inhibits infection whereas the block in CD4+ T cell lines, such as CEM-SS and Jurkat is minimal (257). Comparison of the inhibitory activity of IFN-α in the different cell lines and correlation of this phenotype to the interferon-induced expression of genes led to the discovery of MX2. In addition, shRNA –mediated depletion of MX2 also resulted in the loss of HIV-1 inhibition, thus confirming the antiviral activity of this host protein and explaining part of the IFN block phenotype (252, 260).

In contrast to the antiviral mechanisms described above, the restriction caused by MX2 occurs after the cDNA has been synthesised (257). It is known that the generation of HIV-1 minus strand DNA was not affected suggesting that reverse transcription occurs normally. The abundance of the 2-LTR circular forms of HIV-1 DNA, however, decreases in MX2 expressing cells suggesting that MX2 may act by preventing nuclear accumulation of HIV-1 cDNA (252, 253).

Structure-function studies have determined that monomeric MX2 does not possess the antiviral activity, but that lower-order oligomerisation is sufficient for effective viral inhibition (261). Although the mechanism of HIV-1 inhibition by MX2 is not clear, CA has been suggested to play an important role (252, 253). In particular, a conserved single residue close to the N-terminal of MX2 has been reported to determine antiviral specificity (262). Some studies have shown that viruses with mutations in CA preventing CypA binding were not affected by MX2 (260).

Currently, several hypotheses have been proposed to explain the mechanism by which MX2 mediates HIV-1 restriction. One such hypothesis

postulates that MX2 directly interacts with CA and this results in the block of nuclear import of HIV-1 dsDNA (253, 263). On the other hand, it has also been proposed that MX2 may function at a post-nuclear import stage by altering the stability of the HIV-1 DNA or affecting integration (252, 253, 263).

1.5.7 Interferon-inducible transmembrane proteins (IFITMs)

One of the first host responses to HIV-1 infection is characterised by a cascade of systemic inflammatory cytokines driven by type 1 interferons (IFN-I) being detectable 7 days post-infection (31, 264). Both alpha and beta interferons induce transcription of interferon-stimulated genes (ISGs) by activating receptors IFNAR1/2 via the JAK/STAT pathway and induce activation of the systemic innate and adaptive immunity (264, 265). Some ISGs target different stages of viral replication showing direct antiviral activity (266). Importantly, ectopic expression of several ISGs have direct antiviral activity against HIV-1 (43). Treatment of CD4+ T cells or macrophages with IFN-I in culture, for example, reduces HIV-1 replication (43). Treatment of HIV-1-infected patients with pegylated-IFN, on the other hand, results in a temporary decrease in viral loads (259).

The interferon-induced transmembrane (IFITM) proteins are an example of ISGs that target a specific step of viral replication. These proteins are a family of antiviral factors that inhibit viral fusion of not only HIV-1 to the target cells, but also a number of other viruses such as influenza A virus (IAV), Dengue virus (DENV), hepatitis C virus (HCV) and Ebola virus (EBOV) (267-269). These proteins are mainly found in the plasma membrane or on endosomal membranes, which coincides with the entry pathways of most viruses (270). Five members of the IFITM family have been identified but only IFITM1, 2 and 3 response to IFN-I (264).

Although IFITMs are known to impair viral fusion by changing the curvature or fluidity of the membranes, the complete mechanism by which IFITMs restrict viral replication has not been characterised (43, 271). Nevertheless, it

is known that expression of IFITMs in un-infected cells greatly increases resistance to HIV-1 infection (264, 269, 272). It has been suggested that HIV-1 restriction by IFITMs was dependent on the HIV-1 co-receptor usage (272). IFITM1 was shown to better inhibit viral entry of those viruses that required CCR5 co-receptor in the plasma membrane while IFITM2 and 3 were found to mainly inhibit CXCR4-using viruses that were localised in endosomal compartments (272). Moreover, mutation of the AP-2 binding site (YXXF) or knockdown of same protein resulted in a change of the localisation of IFITMs and a switch in the restrictive activity.

Intriguingly, transmitted/founder (TF) viruses exhibit a substantially higher resistance to the effects of IFN-I in their CD4+T cells than viruses isolated in later stages of the disease, showing that as infection progresses IFN-I resistance declines (272, 273). It is now clear that the acquired HIV-1 sensitivity to IFN as the disease progresses is, at least in part, related to the mutations that Env acquires in the TF/6-months period. These mutations have also been shown to enable escape from early autologous neutralization, and if reverted, they prevent IFITIM-mediated HIV-1 restriction (272, 274). Interestingly, the *env* gene in TF HIV-1 viruses can also confer resistance to a late block induced by Type II IFN that is phenotypically different from the block induced by type I IFN (275).

1.5.8 The role of Vpr

Vpr functions in multiple steps of the viral replication and, interestingly, is the only accessory protein for which any counteraction function towards a host restriction factor has not been described. Vpr is also unique in comparison to the other accessory proteins in that is incorporated into the nascent virions, through residues in the p6 domain of Gag (276). The fact that is a virion-associated protein suggests that Vpr is involved in the early stages of viral replication. Several roles have been attributed to HIV-1 Vpr such as apoptosis, nuclear import of viral DNA and the transcriptional regulation of viral and cellular genes (277).

An extensively studied function of Vpr is its ability to induce cell cycle arrest, thus blocking the proliferation of newly infected T cells. G2 arrest involves high levels of viral replication in human T cells, which correlates with HIV-1 genome expression being more active in this phase (277, 278). Thus, Vpr arrests the cell cycle in this phase, where the LTR is most active so that virus production is maximised (278). Although the mechanism behind has not been fully characterised Vpr interacts with the DCAF-1-Cul4-DDB1 E3 ubiquitin ligase complex and may lead to the degradation of a G2 arrest-linked cellular protein (277). In fact, Vpr has been shown to recruit and degrade the human uracil DNA glycosylase 2 (UNG2), a DNA repair enzyme (277). This result led to hypothesise that association of Vpr with this complex induces polyubiquitination of a cellular protein involved in DNA repair leading to its proteasomal degradation, which ultimately results in ATR-mediated G2 cycle arrest (279). More recently, activation of the structure-specific endonuclease (SSE) regulator SLX4 complex (SLX4com) by Vpr has been shown to be essential in driving G2 arrest. The interaction between Vpr and SLX4 results in the recruitment of VPRBP and kinase-active PLK1 which promotes cleavage of DNA by SLX4- associated MUS81-EME1 endonucleases (279, 280). Activation of these endonucleases has been proposed to result into an accumulation of DNA double strand breaks, which might be sensed by ATR and subsequently lead to cell cycle arrest (279, 280). In addition, SLX4 was also shown to inhibit type I IFN-mediated antiviral response, thus functioning as a regulator of innate immunity (279).

1. 6 Aims of the Thesis

HIV-1 Gag is the protein necessary and sufficient to form the viral particle. As described above, Gag plays a major role in multiple steps of the viral cycle, and therefore, understanding the RNA elements that control Gag expression is important. The HIV-1 genome contains a much lower than expected CpG dinucleotide content, and the consequences of introducing additional CpGs have not been studied. Taken all together, the aims of the thesis are:

1. To determine if RNA elements in *gag* regulate Gag expression and HIV-1 replication

2. To characterise whether CpG dinucleotides are necessary and sufficient to induce inhibition of HIV-1 replication and provide mechanistic insight by which this occurs

Chapter 2: Materials and methods

2.1 DNA methodology

2.1.1 Plasmids

The pHIV-1_{NL4-3} constructs used in this study contain the provirus sequence from pHIV-1_{NL4-3} (281) cloned into the KpnI and SalI sites of pGL4.10 (Promega). pHIV-1 CM22-165, pHIV-1 CM166-261 were generated by overlapping PCR using pHIV-1 CM22-261 as template and have the designated sequences from pHDMHgpm2 (282). pGFP and pVSV-G have been previously described (283, 284) (Table 2.2) .

2.1.2 DNA synthesis

cDNAs were synthesised by GeneART (Life Technologies): For pHIV-1 GagCM22-261, pHIV-1 GagCM22-378, pHIV-1 GagΔ22-261 and pHIV-1 GagΔ22-378 the HIV-1 *gag* ORF sequence contained within BssHII and SphI restriction site of HIV-1_{NL4-3} was synthesised by GeneART (Life Technologies) (281) and have the designated region in *gag* replaced with a XbaI site as in Reil *et al* (285) (Table 2.1 and 2.2) (Figure 2.1). For pHIV-1 GagCM 22-261_{lowCpG} and pHIV-1 GagCM22-378_{lowCpG}, pHIV-1 GagCpG22-261 and pHIV-1 GagCpG22-378, the sequences were synthesised by GeneART and cloned into pHIV-1_{NL4-3}, using restriction sites BssHII and SphI (Table 2.1 and 2.2) (Figure 2.1). For pHIV-1 GagDC22-378, the sequence was synthesised by GeneART and cloned into pHIV-1_{NL4-3} using restriction sites BssHII and SphI. For pHIV-1 GagCpG22-1190, pHIV-1 GagCpG22-759 and pHIV-1 GagCpG760-1190, the sequences were synthesised by GeneART and cloned into pHIV-1_{NL4-3} using restriction sites BssHII/ApaI, BssHII/SphI, and SphI/ApaI, respectively.

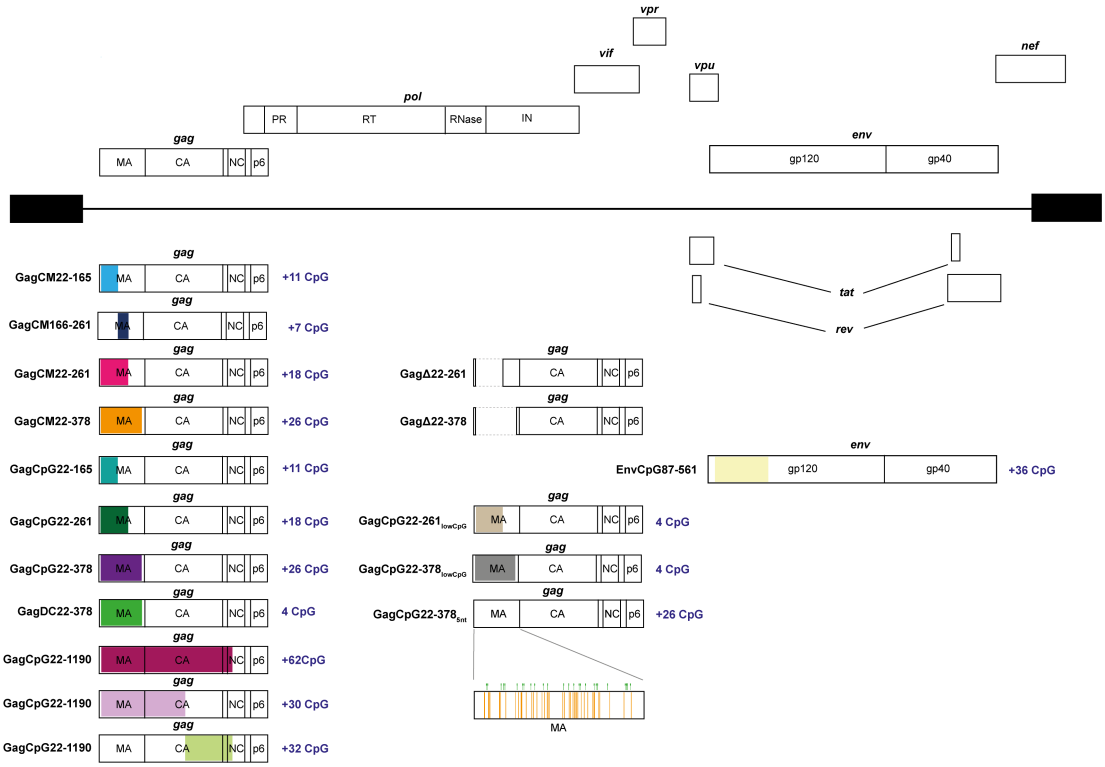


Figure 2.1. Schematic representation of the mutant viruses used in this project. The coloured region represents the codon modified region or the region where CpG dinucleotides have been introduced. The white areas within the grey dash lines represent the deleted regions. The green lollipops represent the CpGs. The number of CpG dinucleotides added is shown in dark blue.

Table 2.1. CpG dinucleotide and mutation number in each construct

	Within the modified region		
	WT	Mutant	
Virus	**Total CpGs**	**Total CpGs**	**Total mutations**
HIV-1 GagCM22-165	4	15	49
HIV-1 GagCM22-261	4	22	80
HIV-1 GagCM22-378	4	30	109
HIV-1 GagCM22-261$_{lowCpG}$	4	4	59
HIV-1 GagCM22-378$_{lowCpG}$	4	4	79
HIV-1 GagCpG22-165	4	15	12
HIV-1 GagCpG22-261	4	22	21
HIV-1 GagCpG22-378	4	30	30
HIV-1 GagCpG22-378$_{5nt}$	4	30	71
HIV-1 GagDC22-378	4	4	23
HIV-1 GagCpG22-1188	9	71	73
HIV-1 GagCpG22-654	4	36	37
HIV-1 GagCpG658-1188	5	35	36
HIV-1 EnvCpG87-561	1	37	43

Table 2.2 List of plasmids used in the project

Plasmid	Description	Cloning/Source
pCMS620	HIV-1_{NL4-3}	HIV-1_{NL4-3} into pGL4 (Swanson Lab)
pCMS308	GFP	GFP into pcDNA3.1 (Swanson Lab)
VSVG	VSVG	From (284)
pIA16	CM22-261	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA24	CM22-165	Overlapping PCR cloned with BssHII/SphI into pCMS620
pIA25	CM166-261	Overlapping PCR cloned with BssHII/SphI into pCMS620
pIA18	CM22-378	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA15	Δ22-261	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA17	Δ22-378	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA30	CM22-261_{lowCpG}	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA32	CM22-378_{lowCpG}	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA35	CpG22-165	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA36	CpG22-261	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA34	CpG22-378	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA37	CpG22-261_{5nt}	Synthesised insert cloned with BssHII/SphI into pCMS620
pMF11	EnvCpG87-561	Mattia Ficarelli (Swanson Lab)
pIA43	GagDC22-378	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA40	GagCpG22-1188	Synthesised insert cloned with BssHII/ApaI into pCMS620

pIA41	GagCpG22-654	Synthesised insert cloned with BssHII/SphI into pCMS620
pIA42	GagCpG658-1190	Synthesised insert cloned with SphI/ApaI into pCMS620

2.1.3 Polymerase chain reaction PCR

pHIV-1 CM22-165 and pHIV-1 CM166-261 were generated in three steps comprising two standard PCR reactions and one overlapping PCR. The first two standard PCR reactions were carried out using a forward primer either in the 5'end of *gag* or in the region anchored in nt 165 of *gag* and reverse primers hybridized in the SphI restriction site and in nt 165 of *gag*, creating PCR products PCR1 and PCR2. The overlapping PCR reaction was performed using both PCR1 and PCR2 products as template and both forward primer containing the BssHII restriction site and the reverse primer containing the SphI restriction site. Phusion High-Fidelity DNA polymerase (New England Biolabs, NEB) was used according to the Phusion PCR protocol: 35µl of MiliQ H_2O (34µl in overlapping PCR), 10µl High Fidelity Buffer, 1µl 10mM 4 dNTPs, 1.25µl 20 mM 5' primer, 1.25µl 20mM 3' primer, 1µl DNA template (1µg/µl) (1µl of DNA template 1a/2a and 1b/2b in overlapping PCR), 0.5µl Phusion High-Fidelity DNA polymerase in a total of 50µl reaction. 35 PCR cycles were performed (Table 2.3).

Table 2.3. PCR Phusion Protocol

	Temperature (°C)	**Time**
Initial denaturation	98	30 s
Denaturation	98	10 s
Annealing	50	30 s
Extension	72	20 s/ Kb
Final extension	72	10 min

2.1.4 Extraction and Purification of DNA fragments

The different DNA fragments were separated based on their molecular weight on a 1% agarose gel. The gel was prepared by mixing 0.5g of agarose (Fisher Scientific) and 1XTBE buffer (89mM Tris acetate, 2mM EDTA, 89mM boric acid, pH8.3) and heating until dissolved. After letting the mix cool down for few minutes, 3µl of ethidium bromide (Sigma) were added, poured into the gel base and left around 20 min to set at room temperature. After this time DNA samples were mixed with 6x loading dye (NEB) and loaded together with a 2-Log DNA or 1Kb DNA ladder (NEB B7025, NEB N3200S, 4µl). The gel was run during approximately 1h at 100V in 1X TBE. Next, separation of the DNA fragments were visualized under UV using the Chemi Doc UV system (Bio-rad). If the fragments were needed for subsequent purification they were visualized at longer wavelength (365 nm) and excised. DNA purification was carried out using the QIAquick Gel Extraction kit (QIAGEN) following manufacturer's instructions.

2.1.5 Restriction enzyme digestion of DNA

DNA or PCR products were digested with different restriction enzyme sites to produce the inserts to clone into the vector pHIV-1_{NL4-3}. The constructs used in this study contain the provirus sequence from pHIV-1_{NL4-3} (281) cloned into the KpnI and SalI sites of pGL4.10 (Promega). All restriction enzymes (NEB) used on this project were incubated with the digestion products for 2h at 37°C (except BssHII, which was incubated at 50°C for 1h) following the reaction mix on Table 2.4. In addition, CIP was added to the digestion mix containing the vector to prevent re-circularization of the plasmid during ligation. Upon digestion fragments were separated by electrophoresis and DNA purified as stated on section 2.1.4.

Table 2.4. Restriction digestion mix

Reagent	Volume (µl)
ddH_2O	47
Buffer	6
Restriction enzyme 1	1
Restriction enzyme 2	1
DNA	5
Total Volume	60

2.1.6 Ligation of inserts and vectors

All the ligations were carried out using T4 DNA ligase in an overnight reaction (Table 2.5) at room temperature. Alongside all ligation reactions a negative control reaction that lacked the DNA insert was carried out. For most of the ligations a 1:5 vector to insert ratio was used.

Table 2.5. Ligation reaction mix

Reagent	Volume (µl)
ddH_2O	2
T4 DNA ligase buffer (NEB)	3
T4 DNA ligase (NEB, 400.000U/mL)	1
Insert DNA	20
Vector DNA	4
Total Volume	30

2.1.7 Transformation of plasmid DNA

First, competent DH10 *Escherichia coli* (E. Coli) were thawed for 15 min on ice. Then, 5µl of ligation product was added into 50µl of competent cells and the mix was incubated in ice for 30min. Once the incubation time was completed the mix was heat-shocked for 45 seconds at 42°C and immediately after the samples were placed back into ice. Third, 300µl of Luria-Bertani (LB) broth (Miller) was added to the mix and incubated for 1h at the shaking incubator at 30°C. Last, 200µl of the culture were added onto LB agar plates containing ampicillin (100µg/ mL) and incubated at 30°C for about 16-18h. The following day individual colonies were picked and cultured in 5mL of LB broth containing ampicillin overnight at 30°C with shaking. The next day 1.5mL of the culture was pelleted and DNA extracted from the competent cells as explained on section 2.1.9.

2.1.8 Preparation of chemically competent DH10 E. Coli

Prior to the generation of competent DH10 E.Coli cells two filter-sterilised buffers were prepared (Table 2.6). E.Coli (Invitrogen) were used to inoculate an agar plate lacking any antibiotic. The following day 4 different colonies were inoculated into 50mL LB broth cultures and incubated at 30°C with shaking. Optical density was measured at 550 nm (OD_{550}) regularly until it reached 0.45. Subsequently, the cultures were place in ice for 10 min after which the competent cells were centrifuged for 10 min at 3000 x g and 4°C. Pelleted competent cells were resuspended in 100 mL Buffer 1 and incubated 10 additional minutes on ice. After this incubation time, cells were newly pelleted following the instructions above and resuspended in 10 mL Buffer 2. Finally, the competent cells were divided into 200µl aliquots and stored at -80°C.

Table 2.6. Buffer composition for the preparation of DH10 E.Coli

Buffer 1	Buffer 2
KAc (30nM)	PIPES (10nm)
RbCl (100nM)	$CaCl_2$ (75nm)
$CaCl_2$ (10nm)	RbCl (10nm)
$MnCl_2$ (50nm)	Glycerol (15%)
Glycerol (15%)	

2.1.9 Miniprep

1.5 mL of culture was centrifuged at 14.000rpm and the pellet resuspended in 150µl resuspension buffer (Macherey-Nagel). Next, 150µl of lysis buffer (Macherey-Nagel) were added to the mix and incubated for 5min after which 150µl neutralization buffer were added (Macherey-Nagel). The mix was pelleted by spinning it for 15 min at 14.000 rpm and the supernatant transferred to a new tube containing 350µl of isopropanol. After mixing it by inverting the tube 5 times the samples were centrifuged for 30min at 14.000rpm. The supernatant was then discarded and the pellet washed with 70% ethanol for 5 min at 14.000rpm. Finally, the pellet was air-dried for 30min and resuspended in 50µl EB buffer (QIAGEN).

2.1.10 DNA sequencing

Following DNA extraction DNA samples were first screened by restriction digestion and bands analysed in a 1% agarose electrophoresis gel. Based on these results the appropriate DNA plasmid was selected and around 50ng were sent to sequence with 10µM of the required forward and reverse primers (Eurofins Genomics or GeneWiz). Upon receiving the sequencing results the sequences were aligned to the maps using MacVector.

2.1.11 Generation of plasmid DNA stocks

The DNA product that resulted from the miniprep and was confirmed to have the correct sequence was retransformed by mixing 1µl of plasmid DNA into 10µl of competent cells (and the protocol on section 2.1.7 was followed). Following re-transformation one of the colonies was inoculated into a starter culture containing 5mL LB and ampicillin (100mg/mL). Approximately after 5h of culture with shaking the started culture was transferred into a 1L Erlenmeyer flask containing 250mL of LB (previously sterilized) and ampicillin and incubated in shaking for at 30°C for ~16h. The following day the culture was centrifuged at 6000 x g for 15 min at 4°C. The supernatant was discarded and DNA extraction from the pellet was carried out using NucleoBond Xtra Maxi kit (Macherey Nagel) following manufacturer's instructions. In the last step, DNA pellet was resuspended in 100-200µl of EB buffer, depending on the size of the pellet. Next, DNA concentration was measured using Nanodrop ND100 spectrophotometer, following manufacturer's instructions (Labtech International).

2.1.12 Sequence analysis of the HIV-1_{NL4-3} gRNA

The "analyze base composition" tool in MacVector was used to calculate the mono- and di-nucleotide frequencies for the HIV-1_{NL4-3} gRNA (NCBI accession number M19921). The dinucleotide frequencies are calculated using the following formula: number of dinucleotide occurrences / (frequency of base 1 in pair * frequency of base 2 in pair) where frequency of base is number of occurrences of base / total number of bases in sequence. WebLogo (286) was used to generate conserved nucleotides surrounding the CpG dinucleotides.

2.2 Cells methodology

2.2.1 Cell culture

Jurkat cells were cultured in Roswell Park Memorial Institute (RPMI) 1640 GlutaMAX Medium (Gibco) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin and maintained at a concentration of $2x10^6$ cells every two days. Hela, TZM-bl and 293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% FBS and 1% penicillin-streptomycin. HeLa and 293T cells were split every other day using a 1:7 or 1:8 dilution, depending on the confluence and TZM-bl cells at a 1:5 or 1:6 dilution . Adherent cells were detached from the plate using 2mL of trypsin (TrypLE Express, Gibco). All cell lines were grown at 37°C in a humidified atmosphere with 5% CO_2.

2.2.2 Generation of HIV-1 virus stocks and quantification of p24 by ELISA

4×10^6 293T cells were seeded in 10cm plates and transfected with 10µg of pHIV-1 and 1.25µg of pGFP using poly(ethlyleneimine) solution (PEI) at a ratio of 10µl PEI per 1µg DNA. The following day media was aspired and replaced with fresh media. Approximately 48-hours post-transfection, the media was harvested, filtered through a 0.45 µm filter and frozen in nine aliquots of 1mL at -80°C. Gag p24 was quantified using a $p24^{Gag}$ enzyme-linked immunosorbent assay (ELISA) (Perkin-Elmer) following manufacturer's instructions.

2.2.3 HIV-1 spreading infection assay

A total of 2.5×10^5 Jurkat cells were plated in 1mL of medium per well in 48 well plates and infected with 25ng of $p24^{Gag}$ of each virus. SupT1 cells were infected with 10ng of $p24^{Gag}$ for each virus. Collection of supernatants was initiated when syncitia were first observed in the culture infected with HIV-

1_{NL4-3}. The amount of infectious virus present at each time point was quantified by infecting the TZM-bl indicator cell line (287-289). Infectivity was measured by the induction of β-galactosidase using the Galacto-Star™ System (Applied Biosystems), as in section 2.2.5.

2.2.4 Transient transfection

HeLa cells were seeded in a 6 well plate so that the following day were at a 70% confluence. Cells were transfected using TransIT®-LT1 (Mirus) according to the manufacturer's instructions at the ratio of 3 µL TransIT®-LT1 to 1µg DNA. For each transfection, 0.5µg pHIV-1 and 0.5µg pGFP or pVSV-G was used. First, DNA conditions were prepared. 0.5µg of virus (HIV-1_{NL4-3} or mutant viruses) and 0.5µg of condition (GFP or VSV-G) were mixed. Second, TransIT®-LT1 and OptiMem were mixed in a total volume of 200 µl and incubated for 5 min. Subsequently 200µl of the mix were added to the tubes containing the plasmid DNA and incubated for 30 min at room temperature, after which the volume was added drop-wise to the cells. 24h post-transfection the medium was aspired and replaced. Media was recovered approximately 48h post-transfection and filtered and pelleted through a 20% sucrose cushion for 2h at 20,000 x g. After centrifugation all the supernatant except 50 µl was aspired and another 50 µl of 2x dissociation buffer (60mM Tris-HCl (pH 6.8), was added to it. Cells were lysed in 500µl of radioimmunoprecipitation assay (RIPA) buffer (10mM Tris-HCl, pH 7.5, 150mM NaCl, 1mM EDTA, 0.1% SDS, 1% Triton X-100, 1%sodium deoxycholate). The amount of infectious virus was quantified by using the TZM-bl indicator cell line (287-289) (see section 2.2.5).

2.2.5 Single cycle infectivity assay

Six-well plates of HeLa cells were transfected with 0.5µg pHIV-1 and 0.5µg pGFP or pVSV-G. Media was recovered approximately 48 h post-transfection and frozen at -80°C. TZM-bl cells were seeded in a 24 well plate so that next

day were at a 70% confluence. The following day 250µl of DMEM and 250µl of supernatant containing the viral particles were mixed and added to the cells. 24h post-transfection the medium was aspired and replaced. Cell lysates were collected with 100µl of Tropix Lysis Buffer (Applied Biosystems) approximately 48h post-transfection. 20µl of the cell lysates were then used to determine the infectivity of the virus by mixing them with 100µl of Galacto-Star™ buffer and substrate (99µl and 1µl, respectively). Substrate typically reaches maximum in 60 minutes and exhibits lumino-kinetics for nearly one hour, therefore, we collected data at 4 different time-points and chose the maximum point of the curve for quantification. Quantification of infectious viral particles upon HIV-1 infection spreading assay (see section 2.2.3) was carried out by adding 10µl of media to a 100µl of cells in a 96 well plate. Cells were lysed with 20µl of Tropix Lysis buffer and β-galactosidase activity measeured in the VICTOR™ XLight luminometer (PerkinElmer).

2.2.6 CRISPR-Cas9

CRISPR-Cas9 cells used in this project were generated by Mattia Ficarelli (PhD student, Swanson Lab, KCL).

ZAP targeting guide sequences (Table 2.7), were inserted into a lentivirus-based CRISPR plasmid (LentiCRISPRv2-puro) from Addgene (52961) as described in Sanjana *et al.* 2014 (290) and were provided by Mr. Harry Wilson (Neil Lab, KCL). Bulk ZAP deficient cells were generated by transduction with the lenti-CRISPR vector followed by selection in 1µg ml^{-1} puromycin (Sigma-Aldrich). HeLa single-cell clones were generated by limiting dilution. Cells were selected for puromycin resistance for 4 days after which surviving cells were counted, diluted to 0.3 cell/well in a 96-well plate and seeded in DMEM + 20% FBS. Selected single cell clones were expanded and maintained in the appropriate media with 1µg ml^{-1} puromycin. Loss of ZAP protein expression was confirmed by western blotting.

Table 2.7. CRISPR-Cas9 guides details

Gene Target	Guide name	Guide sequence
ZC3HAV1	ZAP-1	5'-TCTGGTAGAAGTTATATCTG-3'
	ZAP-2	5'-ACTTCCATCTGCCTTACCGG-3'

2.3 Protein methodology

2.3.1 Analysis of protein expression by immunoblotting

Approximately 48-hours post-transfection, HeLa cells were lysed in radioimmunoprecipitation assay (RIPA) buffer (10mM Tris-HCl, pH 7.5, 150mM NaCl, 1mM EDTA, 0.1% SDS, 1% Triton X-100, 1%sodium deoxycholate). The media was clarified using a 0.45µm filter. Virions were pelleted through a 20% sucrose cushion in phosphate-buffered saline (PBS) solution for 2 hours at 20,000 x g. The pellet was resuspended in 2x loading buffer (60mM Tris-HCl (pH 6.8), 10% β-mercaptoethanol, 10% glycerol, 2% sodium dodecyl sulfate (SDS), 0.1% bromophenol blue). Cell lysates and virions were resolved by SDS-polyacrylamide gel electrophoresis and transferred to a nitrocellulose membrane. The primary antibodies used were specific to HIV-1 $p24^{Gag}$ (291), Hsp90 (sc7947: Santa Cruz Biotechnology), phosphoSTAT1 (612132: BD Transduction), IFIT1 (GTX118713-S: Insight Biotechnology) or β-actin (ac-15: Sigma). Dylight™ 800-conjugated secondary antibodies (5151S and 5257S: Cell Signaling) were used to detect the bound primary antibodies with the Li-CoR infrared imaging (LI-COR UK LTD).

2.4 RNA methodology

2.4.1 RNA extraction

HeLa cells were transfected at a confluency of 70% in a 6 well plate (see section 2.2.1) and after 48h were lysed in 350µl RLT lysis buffer (QIAGEN) containing 1:10 β-mercaptoethanol (SIGMA). RNA was then homogenized using QiaShredder column (QIAGEN) and was frozen at -80°C when not extracted immediately after cell lysis. The supernatant was also collected and treated for 3h at 37°C with RQ1 DNase (Invitrogen) to avoid contamination with plasmid DNA. RNA extraction was carried out in the 10% decon-sterilised hood designated for it, which was treated for 10 min with UV light prior to the procedure. RNA from the cell lysates was extracted using Rneasy Kit (QIAGEN) and RNA from the supernatant using the Viral RNA Extraction Kit (QIAGEN) following manufacturer's instructions. RNA from the cell lysates was eluted in 50µl ddH_2O and RNA from the supernatant was eluted by adding two times 40µl of ddH_2O. The concentration of the RNA extracted from the cell lysates was measured with Nanodrop and the concentration of the RNA of the supernatants was measured using Qubit 3.0 (TermoFisher), following manufacturer's instructions.

2.4.2 cDNA synthesis

The RNA extracted in section 2.4.1 was reverse transcribed using high-capacity cDNA reverse transcription kit (ThermoFisher Scientific) following manufacturer's instructions. First, a reverse transcription master mix was prepared for a total volume of 50µl per reaction (Table 2.8). We used 1µg as input for the reverse transcription of RNA from the cell lysates and 20ng of RNA extracted from the supernatant, and the volume made up with ddH_2O to a total of 50µl. Last, 50µl of the reverse transcription mix and 50µl of the samples were mixed and RT performed following the program on Table 2.9. A negative control that lacked MultiScribe reverse transcriptase was carried out alongside all reactions.

Table 2.8. Reverse Transcription reaction master mix

Reagent	Volume (µl)
10X RT buffer	10
25X dNTP Mix	4
10X RT random primers	10
MultiScribe reverse transcriptase	5
RNase-free H_2O	26
Total	50

Table 2.9. Reverse transcription reaction program

	Step 1	Step 2	Step 3	Step 4
Temperature (°C)	25	37	85	4
Time (minutes)	10 min	2h	5 min	∞

2.4.3 Quantitative PCR

qPCR was performed on the cDNA produced in section 2.4.2. For quantification of the RNA abundance of our samples we produced a standard curve formed by a 10-fold serial dilution of HIV-1_{NL4-3} (from 10^1 to 10^9). In this project two different primer and probe set were used, one measuring the gRNA and the other one measuring the total HIV-1 RNA. PCR reactions were performed in triplicate with Taqman Universal PCR mix using the Applied Biosystems 7500 real-time PCR system. HIV-1_{NL4-3} gRNA primers were GGCCAGGGAATTTTCTTCAGA / TTGTCTCTTCCCCAAACCTGA (forward/reverse) and the probe was FAM-

ACCAGAGCCAACAGCCCCACCAGA-TAMRA. HIV-1_{NL4-3} total RNA primers were TAACTAGGGAACCCACTGC/ GCTAGAGATTTTCCACACTG (forward/reverse) and the probe was FAM-ACACAACAGACGGGCACACACTA-TAMRA. The qPCR reaction was performed using 2µl of sample and 8µl of qPCR mix, therefore on a total volume of 10µl (Table 2.10). The slope of the standard curve was used to quantify both, the absolute number of copies and the efficiency of the qPCR reaction (Table 2.11) All reactions carried out in this project were at an efficiency between 95%-105% (Table 2.11).

Table 2.10. qPCR reagent master mix

Reagent	Final Concentration	Volume
2X Taqman universal PCR Mastermix		5
Forward primer	9pM	1
Reverse primer	9pM	1
Probe	2.5pM	1
DNA		2
Total		10

Table 2.11. Calculation for qPCR reaction efficiency

Efficiency of amplification = ((10^(1/E) 1)*100)
E= the slope of the standard curve generated using pHIV-1_{NL4-3} by qPCR

2.5 TLR and IFN stimulations, Sendai virus infection

This experiment was performed by Charlotte Odendall (Lecturer at KCL) (Appendix 1)

HeLa cells were stimulated with synthetic TLR ligands for 5h at the concentrations indicated. Ligands supplied by Invivogen were polyIC : polyIC (tlrl-pic), Gardiquimod (tlrl-gdqs), CL075 (tlrl-c75), R848 (tlrl-r848), Pam3CSK4 (P3C. tlrl-pms), Ultrapure Flagellin (FliC-tlrl-epstfla-5). LPS was supplied by Enzo (ALX-581-012-L002). CpG DNA was synthesised by IDT and 23S ribosomal RNA by Sigma. Sendai Virus (SeV) was obtained from Charles River labs. IFN-β was purchased from Peprotech and was added to the culture for one hour to activate IFN signaling.

2.6 Bioinformatic analysis

The transcriptomic analysis was performed by Rupert Hugh-White (PhD student, Swanson and Schulz Labs, KCL)

Adapter trimming was performed using BBDuk (https://jgi.doe.gov/data-and-tools/bbtools/), and read pairs were merged with BBmerge (https://jgi.doe.gov/data-and-tools/bbtools/) in order to increase read length. Reads were then aligned to the human genome (hg38) and HIV NL4-3 genomic RNA simultaneously using Hisat2 (292). HIV-mapping junction spanning reads were isolated using regtools (https://github.com/griffithlab/regtools) to allow per-junction read counting. To visualise junctions of interest, data from replicates were first merged using the Picard (http://broadinstitute.github.io/picard) MergeSamFiles function, followed by generation of sashimi plots using Gviz (293).

Chapter 3: Synonymous mutagenesis results in inhibition of viral replication

3.1 Introduction

The viral gRNA has three major functions. First, the gRNA is packaged into the nascent virion to ensure progression of the infection (294, 295). Second, it functions as the mRNA that encodes Gag (structural proteins) and Pol (enzymes) gene products (296). Third, it acts as the pre-mRNA for a pool of singly spliced or fully spliced gene products. HIV-1 splicing is incomplete; while the regulatory proteins Tat and Rev and the auxiliary protein Nef are encoded in fully spliced mRNA, the accessory proteins Vif, Vpr, and Vpu and Env are translated from singly spliced mRNAs (296). Around 50% of the initial transcript remains unspliced (151) but it is still exported to the cytoplasm where it is translated into the Gag polyprotein and Pol (159, 297).

The HIV-1 gRNA contains three different regions. First, a 336 nt 5' untranslated region (UTR) that contains multiple *cis*-acting elements that are contained in complex-stem loop structures and are essential for viral replication (298). Second, an 8616 nt region that includes the ORFs (nt numbers according to HIV-1 $_{NL4-3}$ strain (281)), and third, 219 nt 3'UTR. In addition to encoding the amino acids of the viral proteins, the RNA sequence underlying the open reading frames contain RNA elements that are essential to regulate key steps in HIV-1 replication. These include the programmed ribosomal frameshift in *gag*, which enables the translation of the Pol protein (299), splicing signals (151, 300), the Rev-response element in *env,* which permits the export of intron-containing mRNA to the cytoplasm (301) and the polypurine tracts in *pol* and *nef* that play a role in reverse transcription (302). Moreover, some of these elements are highly conserved and under purifying selection (303, 304). As mentioned in the introduction, IRESes and splicing signals also play a major role in the regulation of HIV-1 gene expression (see section 1.4.9). In addition to these extensively studied RNA elements, the

HIV-1 ORFs may contain other *cis*-acting elements that have not been identified yet.

Besides the linear *cis*-acting elements, second order structures have also been proposed to be important for viral replication. In particular, the 5' and 3' ends of the HIV-1 RNA have been shown to interact, resulting in circularisation of the genome (305). The interaction has been defined to occur between complementary RNA sequences in the 600-700nt in the *gag* ORF and the terminal 123nt of the genomic RNA, which correlates with sequences flanking the 3' TAR hairpin. Moreover, phylogenetic analyses have shown that this 5'/3' interaction is conserved among all HIV-1 subtypes, suggesting an important role in the viral cycle (305). Circularisation of the HIV-1 genome has also been shown to be involved in reverse transcription. Mutagenesis experiments demonstrated that altering the extreme 3'end prevented the circularisation while introducing complementary mutations in the 5' *gag* region could restore the interaction. This interaction has been shown to facilitate the first strand transfer during reverse transcription in vitro (306). A combination of a computational and experimental approach suggests an interaction between the region surrounding the AUG in *gag* and an upstream region that correlates with the U5 (307). Importantly, this interaction has been found to be conserved among distantly related human and animal retroviruses, thus suggesting an important role in HIV-1 replication (308-310). RNA interactions within the HIV-1 genome have also been proposed to play a role in the selection of the gRNA during packaging, inhibition of the 5'polyadenylation signal and/or regulation of splicing. In this case, the interaction is mediated by a short sequence of nucleotides directly downstream of the poly(A) signal and a sequence in the MA region of *gag* (311). Other secondary and tertiary structural elements have also been identified, although the functional relevance remains unknown (309, 310, 312, 313).

The studies described above emphasise the importance of the RNA structures in the HIV-1 genome on regulating several steps of the viral cycle. Nevertheless, in addition to these extensively studied RNA elements, the

HIV-1 ORFs may contain other *cis*-acting elements that have not been identified yet. Characterising the full complement of RNA elements that regulate viral replication is necessary to fully understand the HIV-1 life cycle and may help in the development of new antiviral therapies (314).

3.2 Results

3.2.1 Codon modification of nucleotides 22-261 in *gag* inhibits viral replication

To identify the presence of unknown RNA elements in the MA region in *gag*, we codon modified the RNA (CM) by changing the nucleotide sequence without changing the amino acid sequence. We introduced 80 synonymous mutations into nucleotides 22-261 of HIV-1 NL4-3 (HIV-1_{NL4-3}) *gag*, creating HIV-1 CM22-261 (Figure 3.1 A). Taking advantage of the genetic code, 69 of the 80 codons altered. The mutations were derived from pHDMHgpm2, a codon optimized Gag-Pol construct in which many of the HIV-1 codons are replaced with codons that are frequently found in human mRNAs (315, 316). To further map the presence of *cis*-acting elements in *gag* HIV-1 GagCM22-165 and GagCM166-261 were produced in which 49 and 31 synonymous mutations were introduced, respectively (Figure 3.1 A).

We prepared virus stocks by transfecting 293T cells with each proviral DNA construct and we measured $p24^{Gag}$ concentration by ELISA. Figure 3.1 B shows that while HIV-1 CM22-261 had a ~ 65% decrease in $p24^{Gag}$ and HIV-1 GagCM22-165 had a ~ 40% (Figure 3.1 B), which demonstrated a decrease in $p24^{Gag}$ on the producer cell. To analyse the replicative fitness of each virus we normalised the viral inoculum to 25ng so that Jurkat CD4 T cells were infected with the same amount of $p24^{Gag}$. We started collecting supernatants upon observation of syncytia, around day 3 post-infection and we monitored replication for over two weeks. We then measured viral fitness by challenging TZM-bl indicator cells (287-289) with the collected supernatants. HIV-1 GagCM 22-261 replicated at almost background levels

and had >99.9% less infectivity than wild-type (WT) HIV-1_{NL4-3} on the last day of the assay (Figure 3.1 C). HIV-1 GagCM22-165 had slightly higher replication levels and infectivity was also greatly reduced when compared to WT levels on day 12 (>99%). HIV-1 GagCM166-261 although replication was delayed about three days when compared to the replication curve of the wild-type virus, it plateaued at the same level. The representative figure of three independent experiments is shown in Figure 3.1 C, the rest of the experiments are shown in Figure 3.2. In order to confirm that these results were not dependent on the cell line, we carried out the same experiment in SupT1 CD4 T cells (Figure 3.2). In this case, cells were challenged with 10ng of $p24^{Gag}$ of the wild-type and mutated viruses, instead. Very similar results were obtained in both cell lines, and the only experimental difference relied on the delay of the appearance of syncytia (which correlates with the first collected sample) (Figure 3.2).

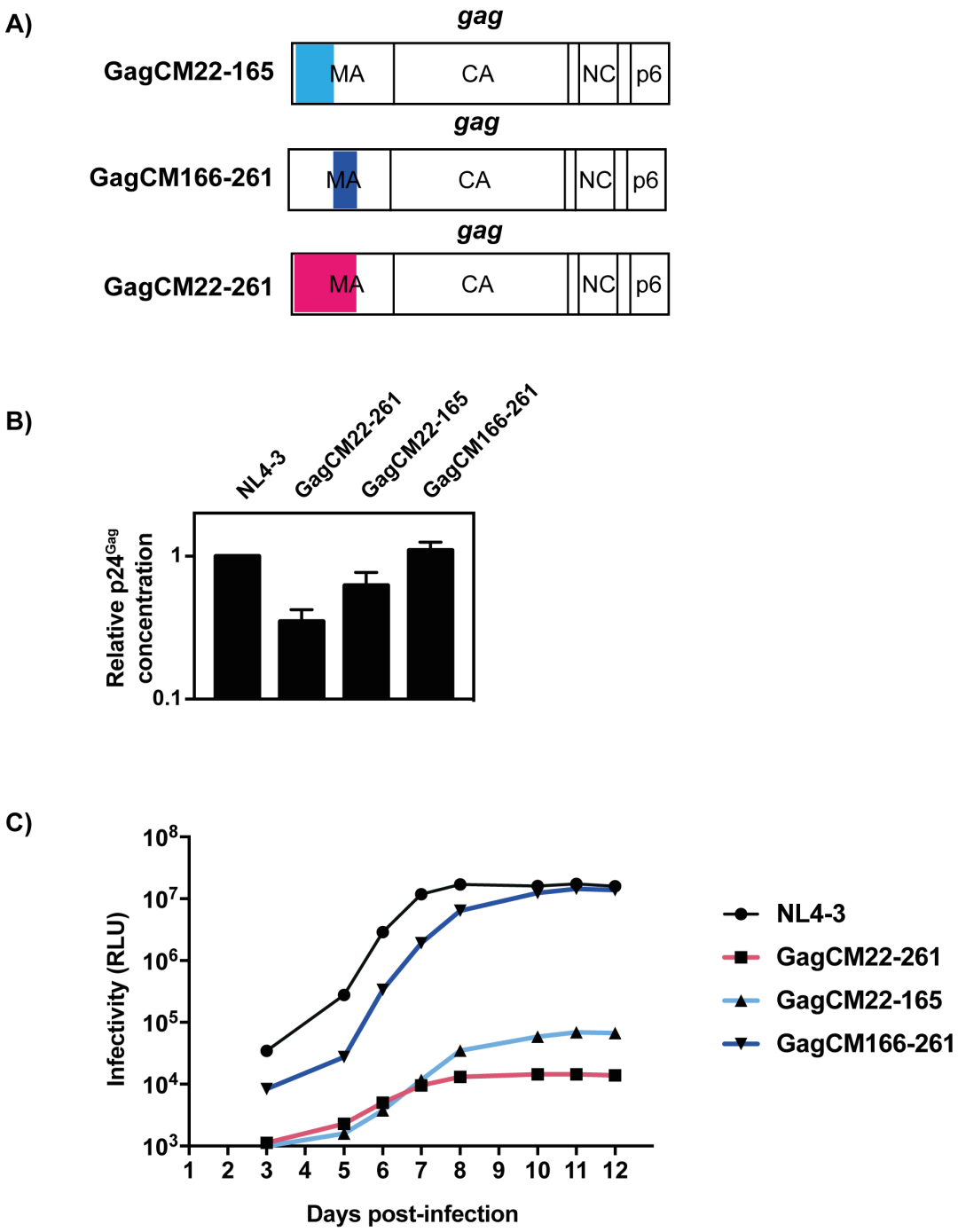


Figure 3.1. Codon modification of nucleotides 22-261 in *gag* inhibits viral replication. A) Schematic representation of HIV-1 GagCM22-261, HIV-1 GagCM22-165 and HIV-1 GagCM166-261. **B)** The amount of HIV-1 CA ($p24^{Gag}$) in supernatants from 293T cells transfected with the viral mutants were quantified by $p24^{Gag}$ ELISA. The bar chart is the average of three independent experiments normalised to $HIV\text{-}1_{NL4\text{-}3}$. Error bars represent standard deviation. **C)** Jurkat cells were infected with 25ng of $p24^{Gag}$ for each indicated virus. The amount of infectious virus present at each time point was measured in TZM-bl cells. This is representative of three independent experiments.

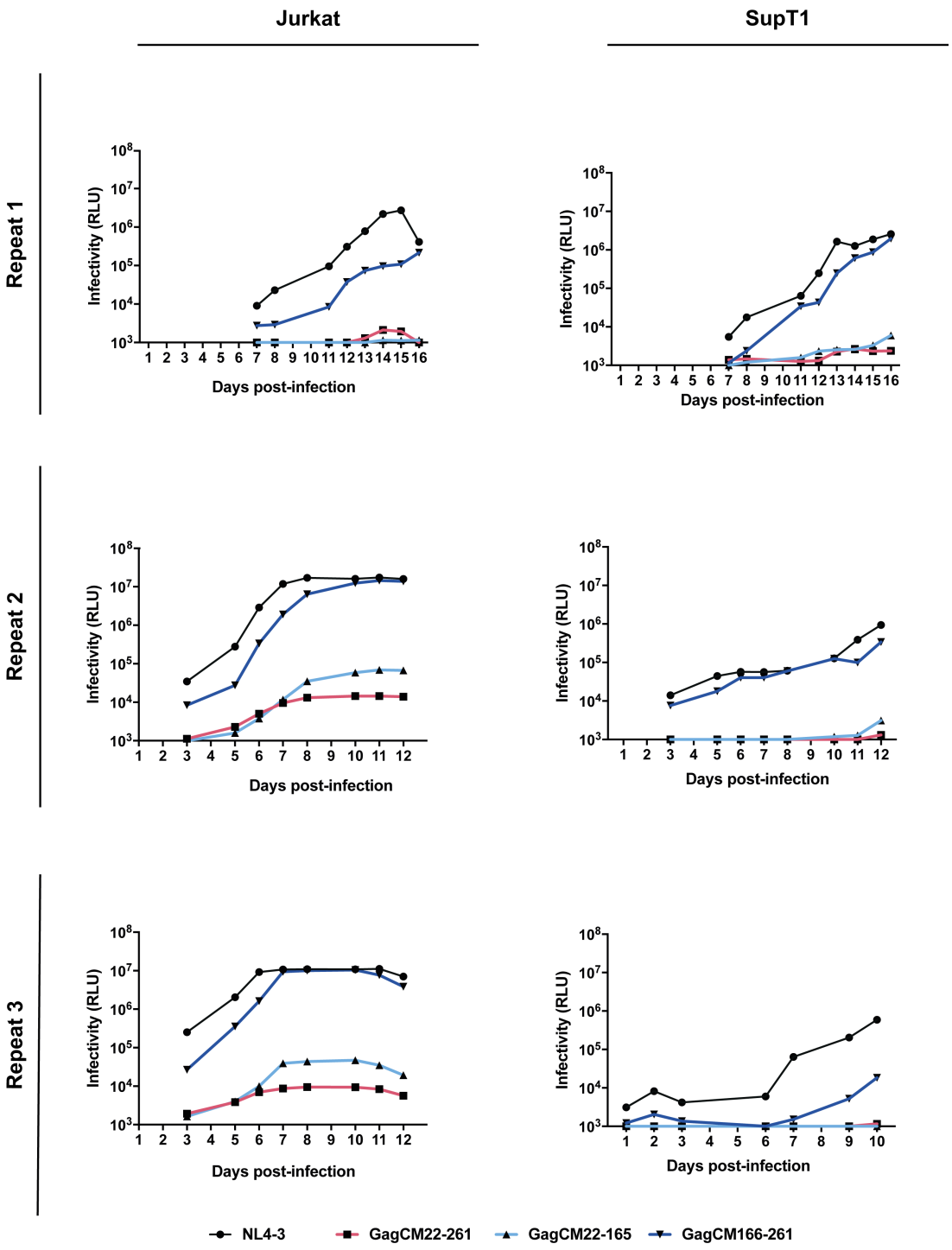


Figure 3.2. Codon modification of nucleotides 22-261 in *gag* inhibits viral replication in Jurkat and SupT1 cells. SupT1 cells were infected with 10ng of $p24^{Gag}$ for each indicated virus. The amount of infectious virus present at each time point was measured in TZM-bl cells. Three independent repeats for each cell line were carried out. Although repeat 2 in Jurkat cells is also shown in Figure 3.1, all repeats have been added to this figure to highlight the reproducibility of the experiments.

3.2.2. Codon modification of nucleotides 22-261 in *gag* inhibits infectious virus production

To further characterise which step of the viral cycle was inhibited upon introduction of synonymous mutations, we used a single cycle infectivity assay. In addition to analysing whether the defect in replication was also visible in a single cycle, with this experiment we studied whether HeLa cells were non-permissive for the mutant viruses. HeLa cells were transfected with pHIV-1_{NL4-3}, HIV-1 GagCM22-261, HIV-1 GagCM22-165 or HIV-1 GagCM166-261 and the supernatant was harvested about 48h post-transfection. Infectivity of the virions in the media was determined by using TZM-bl cells. Compared to the wild-type virus, infectivity of HIV-1 GagCM22-261 was inhibited to the limit of detection of the assay (Figure 3.3 A). Infectivity of HIV-1 GagCM22-165 was decreased, indicating that this virus is also inhibited in HeLa cells. In order to characterise which steps of the viral cycle are inhibited upon introduction of the synonymous mutations, the cell lysate and the virions in the media were also collected to analyse protein expression by quantitative immunoblotting. Both virion production and intracellular Gag expression were decreased around 90% (Figure 3.3 B, C). For HIV-1 GagCM22-165 infectivity was decreased ~98% and intracellular Gag expression and virion production ~50% (Figure 3.3). HIV-1 GagCM166-261 yielded similar amounts of infectivity, intracellular Gag expression and virion production than wild-type HIV-1 (Figure 3.3). Taking all together, for HIV-1 GagCM22-261 we observed a substantial reduction in infectivity, Gag expression and virion production and for HIV-1 GagCM22-165 the reduction is only considerable in the infectivity.

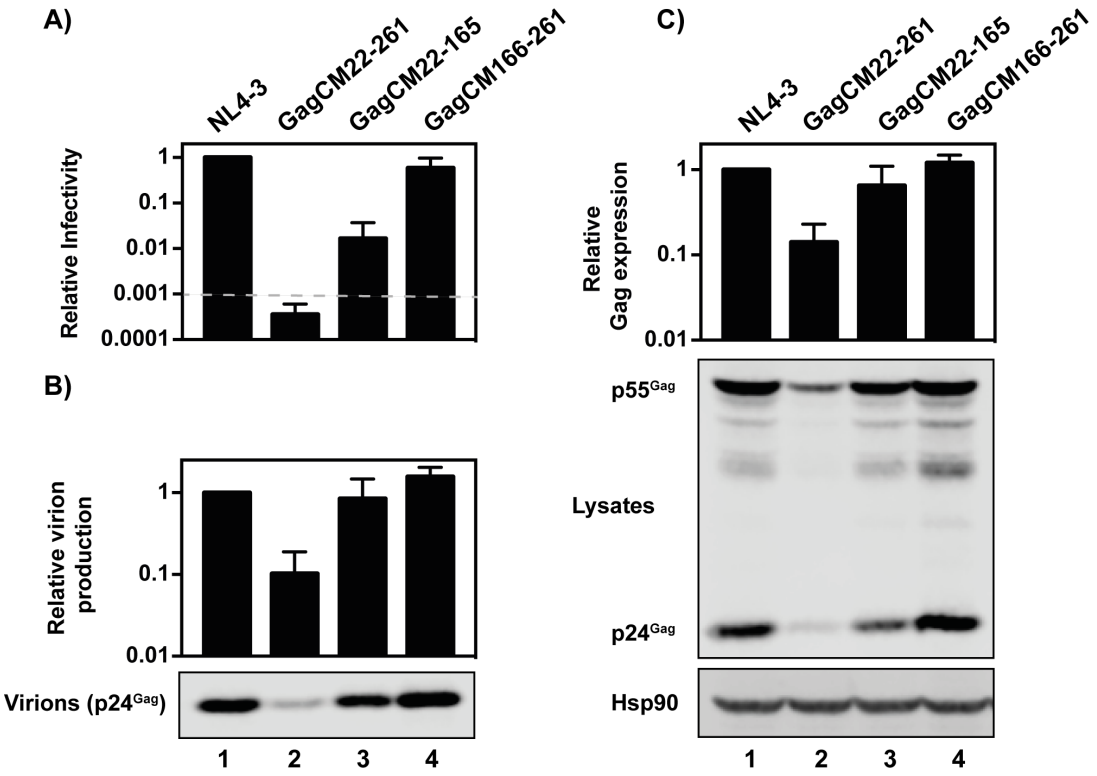


Figure 3.3. Codon modification of nucleotides 22-261 in *gag* inhibits infectious virus production. HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-261, pHIV-1 GagCM22-165 or pHIV-1 GagCM166-261. **A)** Culture supernatants were used to infect TZM-bl reporter cells to measure viral infectivity. The bar charts show the average values of six independent experiments normalised to the value obtained for HIV-1_{NL4-3}. The average relative light units (RLU) for HIV-1_{NL4-3} is 3,680,747. The dashed line represents 1000 RLU, which is approximately the limit of the assay for reproducible differential results. Error bars represent standard deviation. **B)** and **C)** Gag expression in the media (B) and cell lysate (C) was detected using quantitative immunoblotting. The bar charts show the average of four independent experiments normalised to HIV-1_{NL4-3}. Error bars represent standard deviation.

3.2.3 Codon modification of nucleotides 22-261 in *gag* inhibits gRNA abundance in the cell lysates and media

To determine whether the decrease in protein expression was due to a decrease in the gRNA abundance, we carried out quantitative RT-PCR (qRT-PCR) using a primer-probe set that hybridises in a non-mutated region of *gag* (Figure 3.4 A, B). For this experiment, we transfected HeLa cells with pHIV-1_{NL4-3}, pHIV-1 GagCM22-261, pHIV-1 GagCM22-165 and pHIV-1 GagCM166-261 and collected the cell lysate and the supernatant approximately 48h post-transfection. Next, RNA was isolated from both the cell lysate and the media. The gRNA abundance of HIV-1 GagCM22-261 was reduced > 90% in the cell lysate and media compared with the wild type virus (Figure 3.4 A, B). HIV-1 GagCM22-165 gRNA was decreased ~ 70% in the cell and ~ 65% in the media. Regarding HIV-1 GagCM166-261, gRNA abundance in both cell lysates and media was very similar to the levels of wild-type HIV-1. Total RNA abundance was also decreased for HIV-1 GagCM22-165 and HIV-1 GagCM22-261 (Figure 3.4 A).

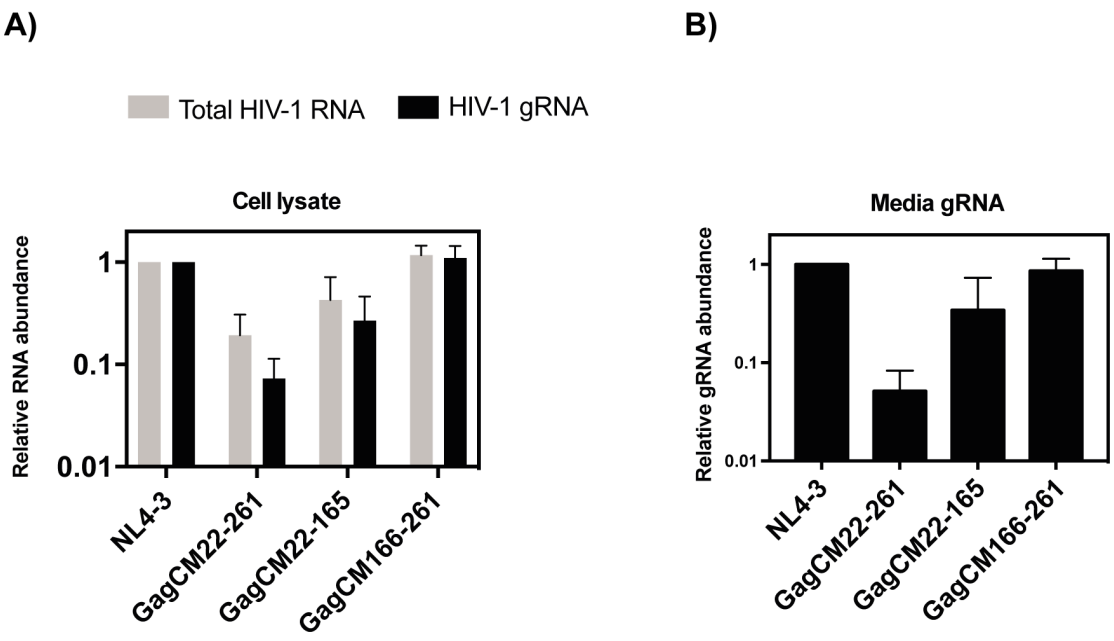


Figure 3.4. Codon modification of nucleotides 22-261 in *gag* inhibits gRNA abundance in the cell lysates and media. HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-261, pHIV-1 GagCM22-165 or pHIV-1 GagCM166-261. The bar charts show the average values of three independent experiments normalised to the value obtained for pHIV-1_{NL4-3}. Error bars represent standard deviation. **A)** RNA was extracted from cell lysates and gRNA and total RNA were quantified by qRT-PCR. **B)** RNA was extracted from the virions in the media and gRNA was quantified by qRT-PCR.

3.2.4 Codon modification of nucleotides 22-261 in *gag* inhibits infectivity when equivalent amount of viral genomes are added

The decrease in infectivity could be explained on the basis of a decrease in the gRNA present in the media. To determine whether the defect in infectivity was persistent when the amount of gRNA was equivalent among all viruses, we measured infectivity/viral genome by infecting TZM-bl cells. For that, the input number of viral genomes was normalised based on the results obtained in Figure 3.4 B. Interestingly, HIV-1 GagCM22-261 infectivity was at the limit of detection of the assay. Infectivity of HIV-1 GagCM22-165 was decreased ~ 98% and no defect was observed for HIV-1 GagCM166-261 (Figure 3.5). These results suggest that the decreased in infectious virus production is not due (or at least fully) to a decrease in the gRNA abundance in the media for HIV-1 GagCM22-261 and HIV-1 GagCM22-165 and it will be further discussed in chapter 5.2.4.

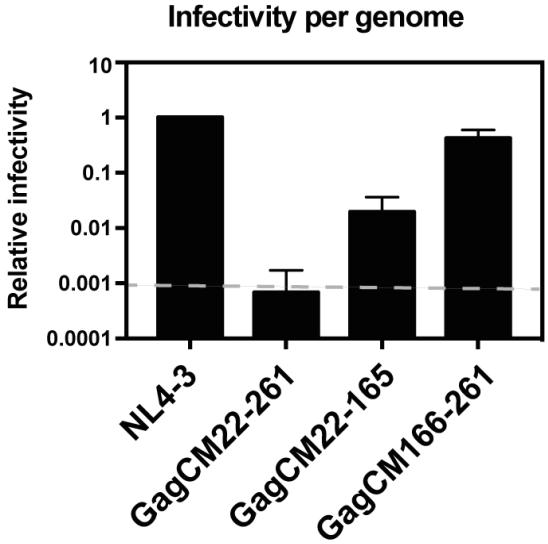


Figure 3.5. Codon modification of nucleotides 22-261 in *gag* inhibits infectivity when equivalent amount of viral genomes are added. Equivalent amounts of HIV-1 genomes were used to infect TZM-bl reporter cells to measure infectivity. The bar charts show the average values of three independent experiments normalised to the value obtained for HIV-1_{NL4-3}. The average RLU for HIV-1NL4-3 is 1,146,196. The dashed line represents 1000 RLU, which is approximately the limit of the assay for reproducible differential results.

3.3 Summary

In this chapter, we have characterised that introduction of synonymous mutations within nucleotides 22-261 produces a large defect in viral fitness (Figure 3.1, HIV-1 GagCM22-261). Importantly, we have shown that these results are not dependent on the cell line (Figure 3.2). We have explained that the decrease in viral replication is caused by a decrease in the gRNA abundance (Figure 3.4), which results in decreased Gag expression and virion production (Figure 3.3). In addition, we have shown that the defect in infectivity is persistent when equivalent amounts of gRNA are used to infect the reporter cell line (Figure 3.5). Taken all together, we have shown that synonymous mutagenesis within nucleotides 22-261 inhibits infectious virus production. In the next chapter, we will focus on understanding the mechanism underlying this inhibition.

Chapter 4: Synonymous mutagenesis results in the introduction of inhibitory sequences

4.1 Introduction

Deletion and mutagenesis studies have widely been used in the literature to map essential regions required for HIV-1 replication (285, 317-321). Previous mutagenesis work on HIV-1 *gag*, for example, has highlighted the importance of the N-terminal region of MA in the regulation of viral replication (284). Indeed, deletion experiments have demonstrated that the N-terminal domain of Gag plays a major role in targeting Gag to the plasma membrane (321). Importantly, efficient HIV-1 replication is obtained even if the entire globular core of MA is missing, provided that the cytoplasmic tail of Env has been removed (285). These results suggest that neither the RNA sequence nor the protein within this region is necessary for efficient HIV-1 replication.

The existence of the functionally relevant interaction between MA and the cytoplasmic tail of gp41 was also shown using site-directed mutagenesis. They showed that the block in the incorporation of Env into the virion upon introduction of mutations in MA could be rescued by deleting several residues in the cytoplasmic tail of (CT) of Env (317). The block in Env incorporation caused by a small alteration within the globular core of MA was also shown to be rescued by envelope glycoproteins with a short CT (322). Furthermore, truncation of the Env CT in mutants with small alterations in MA can replicate in MT4 cells, showing that MA is required for the incorporation of envelope into the nascent virion (285). Of note, pseudotyping with heterologous envelope glycoproteins, such as that from vesicular stomatitis virus (VSV-G), permit viral entry into the host cell. Taken all together, these studies emphasise the importance of the *gag* sequence in regulating multiple steps of the viral cycle that enable efficient viral replication.

Alternative splicing is another essential mechanism that plays a major role in the regulation of HIV-1 replication. HIV-1 can be processed into more than 70 splice forms (323) and the *gag-pol* intron is spliced out through one 5' splice site and six 3' splice sites (150, 151). 3' splice sites of the HIV-1 genome are relatively inefficient compared to those of cellular genes. *tat* coding exon, for example, contains two exonic splicing silencers that suppress A3 in the 3' end, reducing the abundance of both spliced and singly spliced *tat* mRNA. Similarly, low *vpr* mRNA levels are the result of another splicing silencer in the 3' end of the genome (54). Conversely, inefficient splicing signals such as A4c, A4a, A4b and A5 are greatly enhanced by a guanosine-adenosine-rich exonic splicing enhancer within exon 5. Recent studies have highlighted the importance of the HIV-1 RNA sequence and structure surrounding D1 (324) and a novel stem-loop that affects splicing has been defined using SHAPE analysis (309, 325). A tightly regulated splicing pattern is essential in order to maintain a balance between processed transcripts and viral gRNA production. Importantly, silent mutations have previously been shown to positively and negatively alter splicing (54, 149, 326).

4.2 Results

4.2.1 Codon modification results in the introduction of inhibitory sequences in *gag*.

We took advantage of the extensive analysis done on the HIV-1 Gag MA to further characterise the inhibitory phenotype that resulted from codon modification. When studying the importance of the MA sequence in the incorporation of Env into the virion Reil et al. showed that HIV-1 tolerates the deletion of the entire globular core of MA, provided that the cytoplasmic tail of TM has been removed (285). In this study amino acids 8-87, which correlate with nucleotides 22-261 in *gag* were deleted and we used this as a starting point in our analysis.

To characterise whether codon modification of nt 22-261 in *gag* inserted inhibitory sequences or removed *cis*-acting elements that are necessary for HIV-1 replication, we generated an HIV-1_{NL4-3} GagΔ22-261 provirus construct (Figure 4.1 A) and compared it to HIV-1 GagCM22-261 in the presence or absence of VSV-G. In the absence of VSV-G, HIV-1 GagΔ22-261 Gag expression and virion production did not change when compared to wild-type, indicating that the RNA sequence within that region is not essential for regulating those steps in the viral cycle (Figure 4.1). In contrast, codon modification of the same region resulted in inhibited Gag expression and virion production. Infectivity, however, was greatly decreased when nt 22-261 were either deleted or codon modified. This can be explained on the basis of Env not being recruited by ΔMA when the cytoplasmic tail is intact. When the same viruses were pseudotyped with VSV-G the infectivity of HIV-1 GagΔ22-261 was rescued almost to wild-type levels but infectivity HIV-1 GagCM22-261 remained inhibited, with a >99.9% reduction when compared to wild-type (Figure 4.1). These results confirmed that the only defect of the HIV-1 GagΔ22-261 in HeLa cells is the inability to incorporate Env.

Reil et al. (285) also removed amino acids 8-126 in MA to create HIV-1_{HXBH10} Δ8-126/ΔCT, which lacks the globular head. These results showed that Δ8-126/ΔCT, which lacks around 90% of MA, was still able to replicate with only moderate delayed kinetics. Matching this construct, we created HIV-1_{NL4-3} provirus in which nt 22-378 were codon modified (GagCM22-378) or deleted (GagΔ22-378). Correlating with the results obtained for nt 22-261, GagΔ22-378 did not affect Gag expression or virion production but codon modification of the same region (GagCM22-378) greatly inhibited intracellular Gag expression and virion production (Figure 4.1). As expected, in the absence of VSV-G these viruses were defective in infectivity. In the presence of VSV-G, however, infectivity of HIV-1 GagΔ22-378 was rescued but HIV-1 GagCM22-378 had a > 99.9% reduction in infectious virus production. Thus, we concluded that codon modification of these regions in *gag* was more deleterious for viral replication than deleting them, suggesting that by introducing synonymous mutations we were also introducing inhibitory sequences into *gag.*

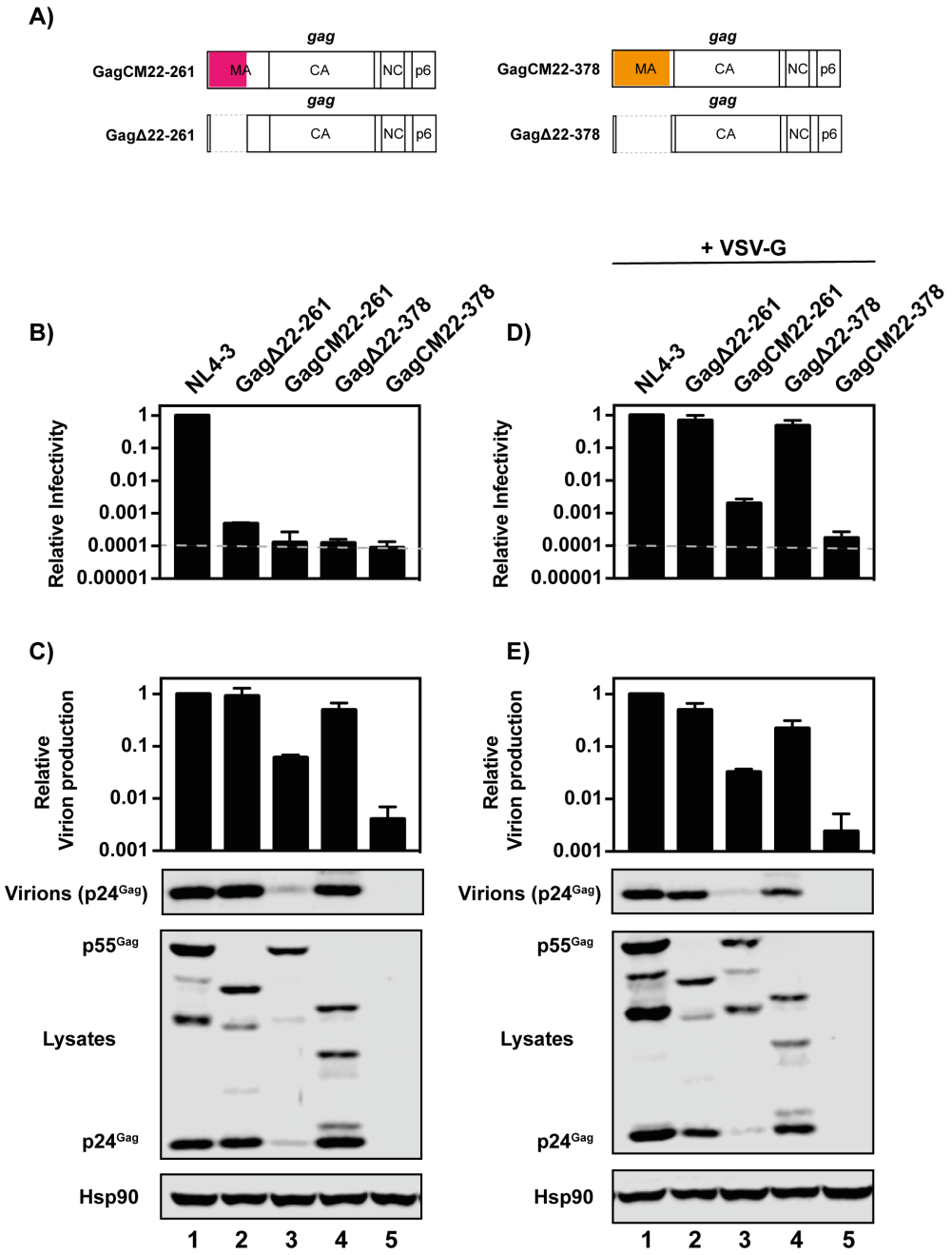


Figure 4.1. Codon modification but not deletion of nucleotides 22-261 or 22-378 in *gag* inhibits infectious virus production. A) Schematic representation of the virus mutants **B)** and **D)** HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-261, pHIV-1 GagΔ22-261, pHIV-1 GagCM22-378 or pHIV-1 GagΔ22-378 and pGFP (B) or pVSV-G (D). The amount of infectious virus in the media was measured in TZM-bl cells. The average RLU for HIV-1_{NL4-3} + GFP and HIV-1_{NL4-3} + VSV-G is 13,701,427 and 16,981,387, respectively. The dashed line represents 1000 RLU, which is approximately the limit of the assay for reproducible differential results. **C)** and **E)** Gag expression in the media and cell lysate were measured by quantitative immunoblotting. (B-E) The bar charts show the average of three independent experiments relative to HIV-1_{NL4-3}. Error bars represent standard deviation.

4.2.2. Codon modification of nucleotides 22-261 and 22-378 in *gag* decreases gRNA abundance and stability

One hypothesis that could explain the decrease in the abundance of viral RNA could be an altered RNA stability. In order to study whether the introduction of synonymous mutations was leading to the degradation of the RNA, we transfected HeLa cells with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-261 or pHIV-1 GagCM22-378. Around 48h post-transfection RNA polymerase II-dependent transcription was inhibited with actinomycin D and RNA was extracted at five different time points. The first time point was 0h, at which RNA was isolated immediately before the addition of the inhibitor into the media. Following this time point RNA was isolated 1h, 2h, 4h and 6h after inhibition of transcription and gRNA abundance was measured by qRT-PCR. The decrease in gRNA abundance of GagCM22-261 and GagCM22-378 was already evident at the 0h time point (Figure 4.2 A). RNA stability of MYC was used as a control, and therefore, its abundance was also analysed at each time point. It is known that MYC mRNA has a very short (<1h) half-life (327), which correlated with the rapid degradation that can be observed in Figure 4.2 B. The relative (compared to 0h) gRNA abundance of pHIV-1_{NL4-3}, HIV-1 GagCM22-261 and HIV-1 GagCM22-378 at the 6h time-point decreased by ~20, ~35 and ~70%, respectively (Figure 4.2 C). The decrease in gRNA abundance could be the result of the inhibitory sequences in the RNA being targeted to a degradation pathway or an increase in post-transcriptional splicing.

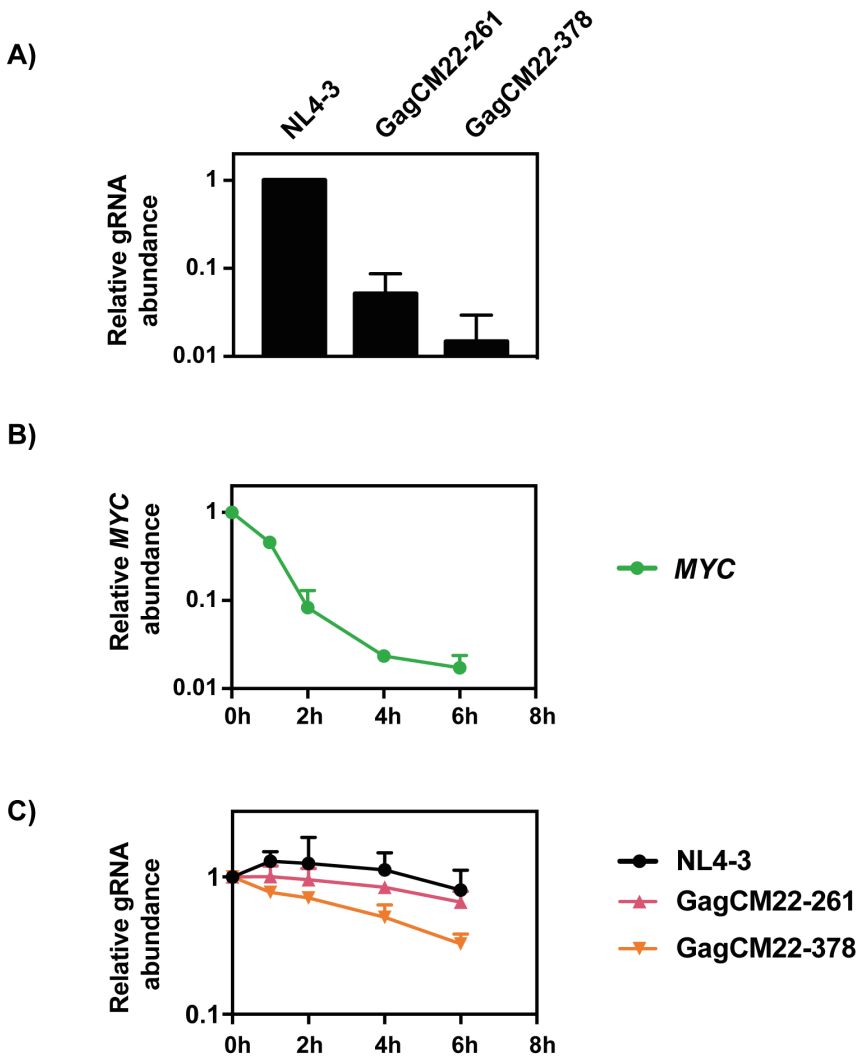


Figure 4.2. Codon modification of nucleotides 22-261 and 22-378 in *gag* decreases gRNA stability. HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-261 or pHIV-1 GagCM22-378. **A)** RNA was extracted from cell lysates at the 0 hour time point and gRNA abundance was quantified by qRT-PCR. The bar charts show the average values of three independent experiments normalised to HIV-1_{NL4-3}. Error bars represent standard deviation. **B)** and **C)** Actinomycin D was added to inhibit RNA polymerase II transcription and the abundance of *MYC* mRNA (B) or gRNA (C) was measured at 0, 1, 2, 4 and 6 hours post-addition. Each value is relative to the 0-hour time point and is an average of three independent experiments. Error bars represent standard deviation.

4.2.3 Codon modification of nucleotides 22-378 results in an altered splicing pattern

We speculated that synonymous mutagenesis could lead to a reduction in the gRNA abundance by introducing an intronic splicing enhancer (ISE) or removing an intronic splicing silencer (ISS) that could alter the splicing pattern. On the other hand, activation of a cryptic splice site (CSS) upon codon modification could also explain the phenotype. CSSs are disadvantageous or inefficient splice sites that are generally dormant unless activated as a result of a splice enhancer in the nearby region or as result of a mutation.

To determine whether the decrease in the gRNA abundance is due to altered splicing, we transfected HeLa cells with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-165, pHIV-1 GagCM22-261 or pHIV-1 GagCM22-378 and harvested the cell lysates ~48h post-transfection. We then quantified total HIV-1 RNA and compared it to HIV-1 gRNA (Figure 4.3). Total HIV-1 RNA was quantified by using a primer-probe upstream of the major 5'splice donor (SD1). HIV-1 GagCM22-261 and HIV-1 GagCM22-165 had an ~80% and ~60% decrease in total HIV-1 RNA abundance (Figure 4.3).

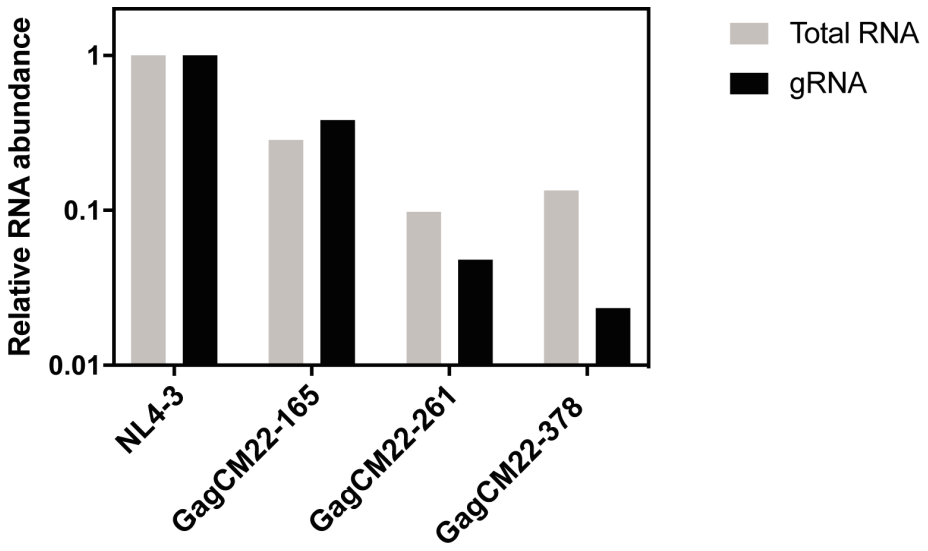


Figure 4.3. Codon modification of nucleotides 22-261 in *gag* decreases gRNA and total HIV-1 RNA abundance. HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-165, pHIV-1 GagCM22-261 or pHIV-1 GagCM22-378. The bar charts show the average values of two independent experiments normalised to HIV-1_{NL4-3}. RNA quantification of GagCM22-165 and GagCM22-261 is also shown in Figure 3.4.

The RNA of the cell lysates were also sent for RNAseq and the reads were analysed using several bioinformatic programs (see section 2.6 for detailed description). The analysis showed that as a result of the synonymous mutagenesis a cryptic splice donor (CD1) is activated in position 718 (nucleotide number correlates with the start of the 5'UTR), which leads to the incorporation of the *gag* start codon in every transcript (Figure 4.4). Table 4.1 shows a switch in the percentage of splice sites used that correlates with the length of the codon modified region in *gag*: while the wild type virus does not use CD1 this percentage increases the longer the region codon modified. This start codon can result in inefficient translation of HIV-1 proteins, which can affect multiple steps of the HIV-1 cycle. In addition to inducing translocation of intron-containing mRNAs into the cytoplasm, Rev has been shown prevent nuclear degradation of pre-mRNAs (328). Quantification of mRNA in HIV-1 infected T cells expressing or lacking Rev showed that the half-life of the transcripts was unstable in the absence of Rev (328). Thus, a decrease in Rev abundance can lead to a decrease in RNA stability and therefore gRNA abundance. In summary, we hypothesise that the decrease in gRNA levels observed in Figure 3.3 may be caused by low Rev abundance as a consequence of the activation of CD1 (Table 4.1).

Table 4.1 Percentage of junction observation for splice donor 1 (D1) or cryptic splice donor (CD). Nucleotide number correlates with the start of the 5'UTR. The cryptic donor is located in nucleotide 718.

start	end	Splice sites	WT.1	WT.2	CM22-165.1	CM22-165.2	CM22-261.1	CM22-261.2	CM22-378.1	CM22-378.2
291	4460	D1-A1	11.92%	11.95%	3.05%	3.42%	2.45%	1.47%	0.18%	0.00%
291	4937	D1-A2	4.43%	4.44%	3.25%	3.32%	2.91%	2.12%	0.00%	0.00%
291	5324	D1-A3	2.98%	3.14%	3.42%	3.49%	3.99%	3.15%	0.00%	0.02%
291	5483	D1-A4c	1.19%	1.36%	0.63%	1.12%	0.06%	0.37%	0.00%	0.00%
291	5501	D1-A4a	4.11%	3.68%	1.91%	1.97%	1.60%	0.91%	0.00%	0.00%
291	5507	D1-A4b	2.22%	2.21%	1.65%	1.63%	1.08%	1.10%	0.27%	0.08%
291	5523	D1-A5	25.71%	26.17%	30.41%	33.30%	29.70%	26.17%	0.67%	0.90%
291	7916	D1-A7	0.94%	0.95%	0.90%	0.76%	0.00%	0.40%	0.00%	0.02%
718	4460	CD-A1	0.00%	0.00%	0.58%	0.41%	2.00%	1.47%	7.81%	7.43%
718	4937	CD-A2	0.00%	0.00%	0.00%	0.03%	0.51%	0.84%	1.72%	1.37%
718	5324	CD-A3	0.00%	0.00%	0.14%	0.13%	1.43%	1.15%	1.63%	1.81%
718	5501	CD-A4a	0.00%	0.00%	0.00%	0.00%	0.00%	0.13%	0.99%	1.16%
718	5507	CD-A4b	0.00%	0.00%	0.00%	0.00%	0.40%	0.38%	3.17%	3.02%
718	5523	CD-A5	0.00%	0.00%	0.20%	0.44%	4.68%	4.45%	24.94%	23.07%

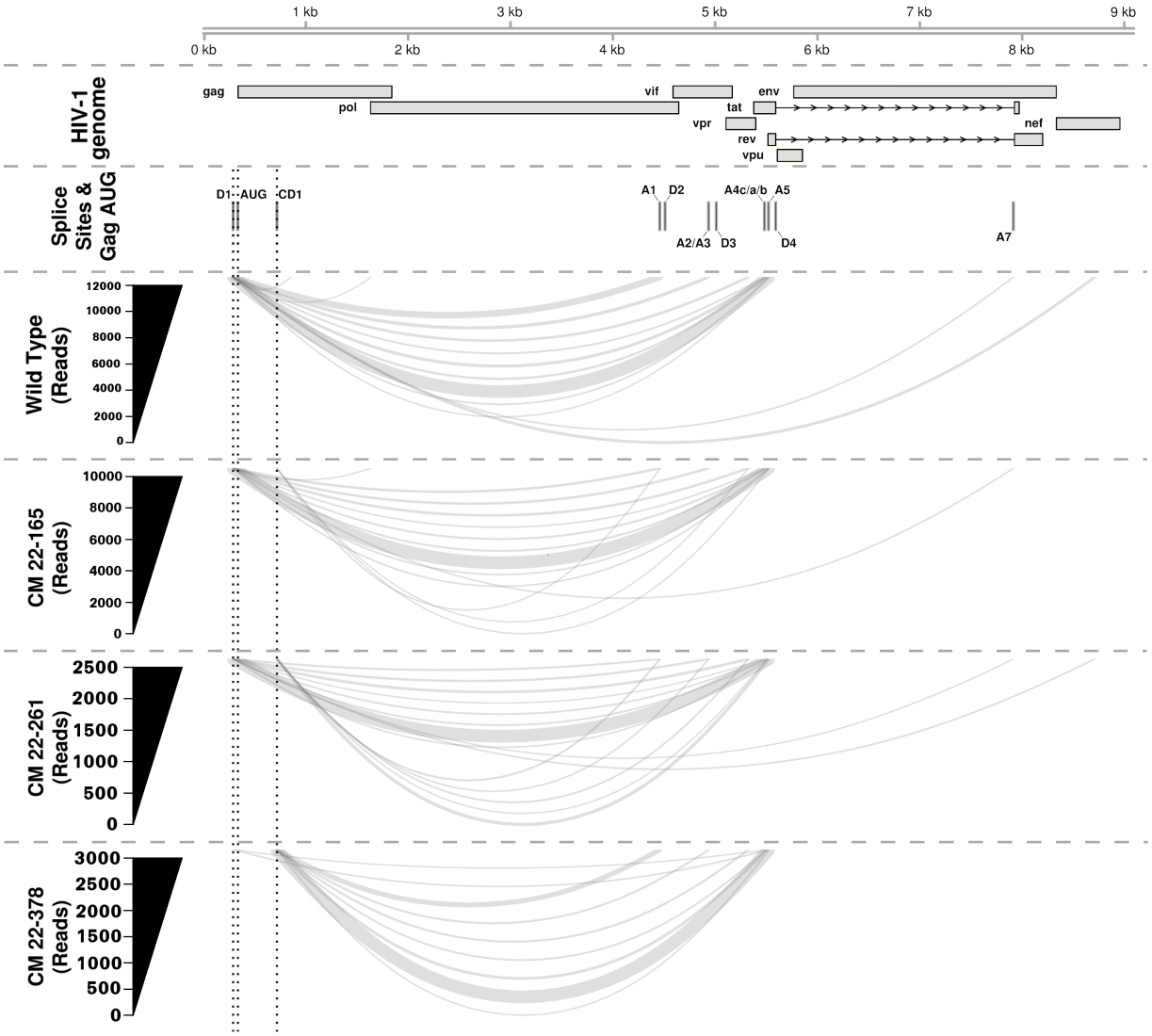


Figure 4.4. GagCM22-378 results in the activation of a cryptic splice donor. Relative usage of splice donor 1 and cryptic donor 1 upon codon modification of HIV-1. The 9173 nt HIV-1 genomic RNA and features are depicted in the "HIV-1 genome" track. Canonical donors (D1-4) and acceptors (A1-7), codon modification-induced-cryptic donor (CD1), and gag start codon (AUG) are shown in the "Splice Sites and Gag AUG" track. The numbers of reads supporting use of D1 (nt 291) or CD1 (nt 718) paired with any canonical or non-canonical acceptor and summed across duplicate samples is shown; depicted by line width (y-axis). Line height is arbitrary. Analysis carried out by Rupert Hugh-White (PhD student, Schulz and Swanson Labs, KCL).

4.3 Summary

In this chapter, we have shown that synonymous mutagenesis results in the introduction of an inhibitory element within nucleotides 22-261 and 22-378 (GagCM22-261 and GagCM22-378) (Figure 4.1). In addition, we have shown that the mechanism by which the inhibition occurs involves a reduction in the RNA levels (Figure 4.2). Importantly, we have characterised that codon modification leads to the activation of a cryptic splice site (CD1) that results in aberrant splicing (Figure 4.4), which we suggest explains the defect in infectivity. In the next chapter, we are going to focus on characterising the nature of the inhibitory element.

Chapter 5: CpG dinucleotides are necessary for the inhibition of infectious virus production

5.1 Introduction

The nucleotide composition of HIV-1 differs from the human in that is disproportionally A-rich and C-low. This distinctive feature is common in HIV-1 groups M, N and O and is generally found in lentiviruses, but not in all retroviruses (74-76, 329-333). In the HIV-1 genome, A-rich content ranges from 35.1% to 40% (75, 330, 334, 335) while the C-count content is ~18%. The A-rich codon usage bias is apparent in each of HIV-1's three major coding regions (*gag, pol* and *env*) but it is most pronounced in the *pol* gene (74, 76). Although HIV-1 is one of the most variable viruses known with regard to its nucleotide substitution rate, the base composition of the genome has surprisingly remained stable over the time of the epidemic (331). This suggests that the A-rich HIV-1 genome is preserved during evolution.

Human APOBEC3 proteins are, at least in part, responsible for the A-richness of the viral genome (336, 337). APOBEC3G is incorporated into the virion and triggers the deamination of cytosine to uracil on the negative sense single-stranded DNA that is generated upon reverse transcription, which results in inhibition of HIV-1 cDNA synthesis. In addition, human APOBEC3 proteins have also been proposed to induce subtle G-to-A mutations that contribute to the HIV-1 sequence variation. These alterations have been proposed to be beneficial for the virus as they could contribute to viral adaptation and evolution (338).

This biased nucleotide composition affects the dinucleotide composition of the HIV-1 genome (339). While the most common dinucleotide is ApA (340), CpG is found at the least abundance (332, 334, 341-343) (Table 5.1) and

has been proposed to be under negative selection and linked to disease progression (344, 345). The fact that CpG dinucleotides are underrepresented in the HIV-1 genome suggests that there might be deleterious for viral replication. Importantly, it has been shown that the low CpG frequency is not due to the high abundance of A-nucleotide in the genome since CpG dinucleotides are suppressed but not GpC (346). This large bias in dinucleotide composition has been shown to be the result of a selection against CpGs according to a statistical codon-based model (347). They argue that the estimates of expected numbers of substitutions or synonymous to non-synonymous mutations that lead to CpG suppression are very low in comparison to the CpG-generating substitutions (347). In addition, it has been hypothesised that the negative selection against CpG dinucleotides might be related to DNA methylation. The newly reverse transcribed cDNA is not methylated previous to integration. Low CpG frequency, therefore, could be a protection mechanism against methylation, since this would lead to repression of the integrated provirus (342). Transcriptionally latent HIV-1 provirus, however, is methylated at two CpG motifs near the transcription start site, suggesting that methylation can repress transcription even when CpG abundance is low (348). Nevertheless, CpG suppression is also present in other RNA viruses that lack a DNA phase in their life cycle, which hints that DNA methylation is not the only reason that leads to CpG suppression (339, 343, 346, 349, 350). HIV-1 is not the only CpG-low RNA virus that replicates in vertebrates.

Interestingly, although RNA or DNA viruses are obligate intracellular parasites, they do not share common features with the genomes of their host organism (343, 346, 351-353). It has been suggested that the CpG suppression in vertebrate viruses is the result of an evolutionary event where the viruses that have derived from host sources have acquired this feature (343). Intriguingly, evolutionary analysis of influenza A and influenza B viruses demonstrate the existence of evolutionary pressure that shapes the viral RNA towards patterns found in the host genome (349). Influenza A, which was originated in avian hosts and has been replicating in humans for around 90 years, has been shown to evolve towards a reduction in the CpG

content of the genome. Influenza B, a virus that has infected the human population for a longer time, has already adapted to the evolutionary pressure received from the human genome and contains a very low CpG dinucleotide content (349).

Another RNA dinucleotide pattern that has previously shown to inhibit RNA virus replication is UpA. In the HIV-1_{NL4-3} gRNA the UpA dinucleotide frequency deviates from what is randomly expected and codon modification of the regions 22-261 and 22-378 greatly decreased the UpA abundance compared to the wild-type sequence (Table 5.1) (354). This shows that the inhibitory phenotype that we observe upon codon modification is not due to a change in the UpA content. Given that CpG dinucleotides are the only dinucleotide substantially suppressed we did not perform any experiments to study the UpA content of the HIV-1 genome (Table 5.1).

The biased nucleotide composition also affects the codon bias, which refers to the different frequency that an organism has to use synonymous codons (355). The redundancy of the genetic code enables 61 codons to specify for 20 amino acids. The usage of the synonymous codons is non-random and is mainly determined by the specific tRNA levels available during protein synthesis. The unusual nucleotide composition of the HIV genome results in a codon bias that is quite different from that of the human genome (74-76).

Table 5.1. HIV-1_{NL4-3} genomic RNA mononucleotide and dinucleotide frequencies

Mononucleotide frequencies

Base	#	Freq.	%	Obs./Exp.
A	3281	0.358	35.8	1.43
C	1635	0.178	17.8	0.71
G	2216	0.242	24.2	0.97
T	2041	0.223	22.3	0.89
G+C	3851	0.420	42.0	0.84
A+T	5322	0.580	58.0	1.16
Total Positions = 9173				

Dinucleotide frequencies

Base	#	Freq.	%	Obs./Exp.
AA	1093	0.119	11.9	0.93
AC	522	0.057	5.7	0.89
AG	962	0.105	10.5	1.21
AT	703	0.077	7.7	0.96
CA	758	0.083	8.3	1.30
CC	371	0.040	4.0	1.27
CG	82	0.009	0.9	0.21
CT	424	0.046	4.6	1.17
GA	762	0.083	8.3	0.96
GC	424	0.046	4.6	1.07
GG	625	0.068	6.8	1.17
GT	405	0.044	4.4	0.82
TA	668	0.073	7.3	0.92
TC	318	0.035	3.5	0.87
TG	546	0.060	6.0	1.11
TT	509	0.055	5.5	1.12
Total Positions = 9172				

Interestingly, increasing the CpG dinucleotide abundance has previously been shown to inhibit replication of several viruses. Replicative fitness of poliovirus, for example, was greatly decreased upon addition of CpG dinucleotides in the capsid region of the genome, which constitutes only a 9% of the total genome (356). In addition, alteration of CpG dinucleotide abundance in the picornavirus echovirus 7 (E7) has also been shown to affect viral fitness. While increasing the CpG frequency in a segment of the E7 viral genome impaired replication kinetics, removal of CpGs turned in an enhance viral fitness (357). Viral attenuation upon altered codon pair (CP) frequencies leads to an unintended increase in the CpG frequencies in picornavirus. Importantly, reduced viral fitness was shown to be the result of increased CpG dinucleotides and not due to the use of disfavoured CPs or impaired translation efficiency (339). Inhibition of IRF3, an intermediary between pattern recognition receptors and interferon production pathways, had no effect on CpG-mediated reduction of viral replication. These results suggested that the IFN pathway is not involved in the enhanced innate immune response to viruses with elevated CpG frequencies (354). Similar results were obtained when maximised frequencies of the CpG dinucleotide were introduced into segment 5 of influenza A virus (IAV). Interestingly, this attenuation was also observed *in vivo*, where they found a 10-100 reduction in viral load in the lungs of mice infected with CpG-high mutants (358). Some deeper mechanistic analysis was carried out by Fros et al. (359), where they showed that the defect in replication upon artificially increasing the CpG abundance in E7 was the result of a block immediately after viral entry. Shortly after entry, high-CpG containing mutants showed a delay in the formation of replication complexes. They found no evidence of mutant viruses producing a reduction in translation efficiency since mutants in which CpGs were added in non-coding regions also showed a defect in replication. Similarly, the viral RNA was not sequestered in stress granules. Although none of the conventional antiviral mechanisms seems to be involved, they

propose that the inhibited replication of CpG-containing mutants is not an intrinsic defect of the virus but a consequence of the host cell defence (359).

5.2 Results

5.2.1 Decreasing the CpG abundance within HIV-1 GagCM22-261 and HIV-1 GagCM22-378 restores infectious virus production

HIV-1_{NL4-3} contains 4 CpGs within nt 22-261 in *gag* and the abundance raises up to 22 in the same region upon codon modification (Table 5.2). Similarly, for the region within nt 22-378 the CpG abundance increases from 4 (in the HIV-1_{NL4-3}) to 30 in GagCM22-378 (Table 5.2).

To test whether the increased CpG abundance is the responsible for the inhibition of the viral fitness we synthesised a HIV-1 *gag* sequence containing all of the synonymous mutations present in pHIV-1 GagCM22-261 with exception of the codon changes that introduced CpG dinucleotides and inserted it into pHIV-1_{NL4-3} to produce pHIV-1 GagCM22-261_{lowCpG}. This provirus contains the same four CpG dinucleotides than the wild-type virus within this region and 59 other mutations, compared to HIV-1 GagCM22-261 that has 22 CpGs and 80 mutations (Table 5.2). We transfected HeLa cells with pHIV-1_{NL4-3}, HIV-1 GagCM22-261 and pHIV-1 GagCM22-261_{lowCpG} and analysed infectivity in single round replication assays, Gag expression and virion production. Interestingly, infectious virion production was rescued upon lowering the CpG abundance to 4 in HIV-1 GagCM22-261_{lowCpG} (Figure 5.1 B, C). We also used the same strategy to produce GagCM22-378_{lowCpG}, which contains 4 CpG dinucleotides and 79 mutations compared to the 30 CpG dinucleotides and 109 mutations in HIV-1 GagCM22-378 (Table 5.2). Similar results were obtained when infectivity, Gag expression and virion production were measured, where infectious virion production of GagCM22-378_{lowCpG} was very similar to that of HIV-1_{NL4-3} (Figure 5.1 D, E). Of note, although codon modification of both regions in *gag* substantially decreased the A-rich content, removing the CpG dinucleotides in GagCM22-261 and

GagCM22-378 did not restore the A-rich bias for GagCM22-261$_{lowCpG}$ and GagCM22-378$_{lowCpG.}$ These data suggest that the decrease in infectious virus production is due to an increased CpG abundance and not to an altered A-rich nucleotide bias. In summary, changing the codons that resulted in addition of CpGs into the region of interest back to the wild-type codon rescues Gag expression, virion production and infectivity, and determines that CpG dinucleotides are necessary to inhibit replication.

Table 5.2. Changes in nucleotide composition and total number of mutations for codon modification of regions 22-261 and 22-378 in *gag*

Construct	% A	% C	% G	% T	Total UpA	Total CpG	Total number mutations relative to wild type
WT 22-261	38	17	23	22	23	4	
GagCM22-261	18	36	32	15	3	22	80
GagCM22-261$_{lowCpG}$	24	27	34	15	5	4	59
GagCpG22-261	32	25	22	21	19	22	21
WT 22-378	42	17	24	18	29	4	
GagCM22-378	22	35	32	12	3	30	109
GagCM22-378$_{lowCpG}$	28	26	33	13	7	4	79
GagCpG22-378	36	25	22	17	23	30	30

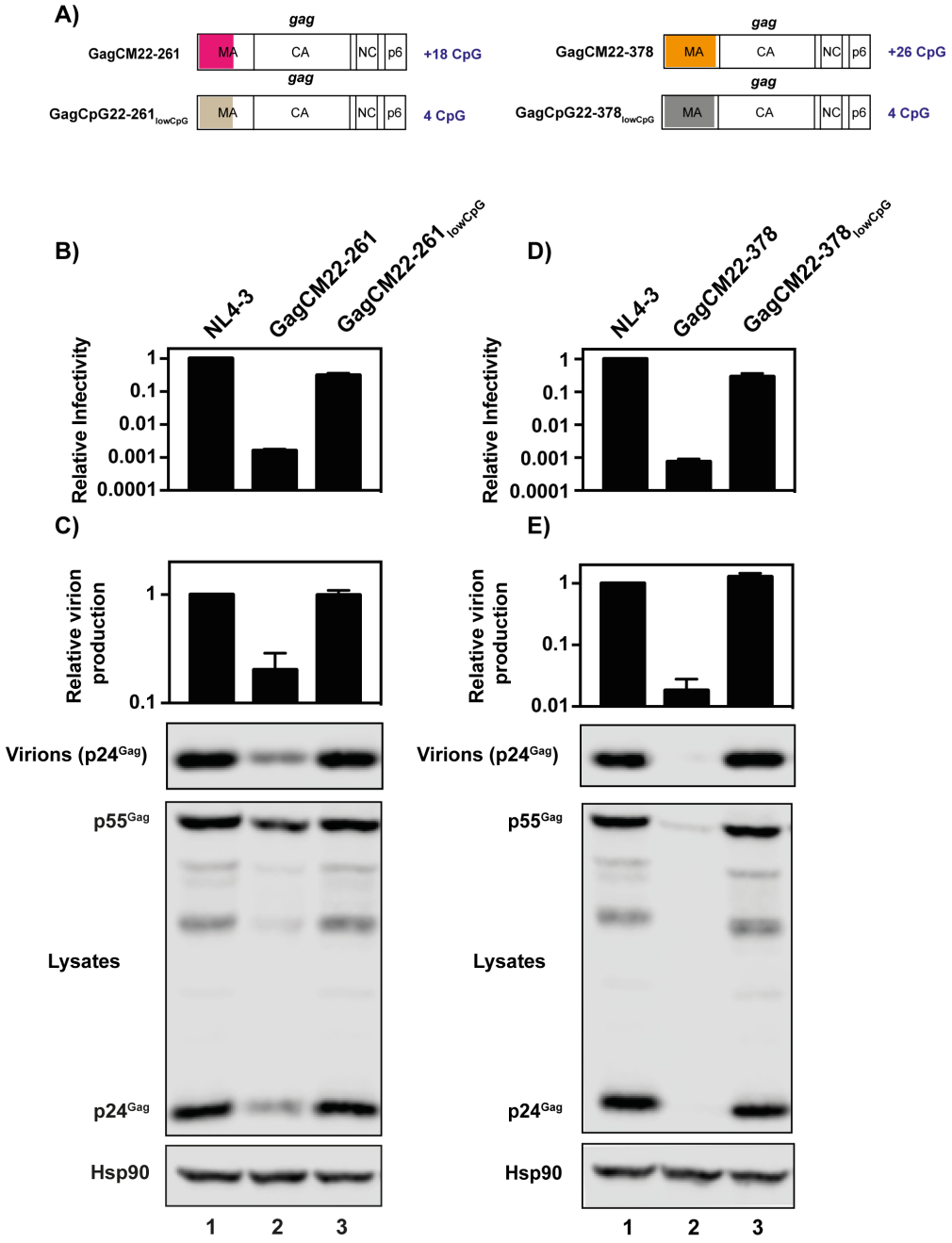


Figure 5.1. Decreasing the CpG abundance within HIV-1 GagCM22-261 and HIV-1 GagCM22-378 restores infectious virus production. HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 GagCM22-261, pHIV-1 GagCM22-261_{lowCpG}, pHIV-1 GagCM22-378 and pHIV-1 GagCM22-378_{lowCpG}. **A)** Schematic representation of the virus mutants **B)** and **D)** The amount of infectious virus present in the media was measured in TZM-bl cells. **C)** and **E)** Virion production and intracellular Gag expression was measured using quantitative western blotting. The bar charts show the average of three independent experiments relative to HIV-1_{NL4-3}. Error bars represent standard deviation.

5.2.2 Introduction of CpG dinucleotides into *gag* inhibits HIV-1 replication in Jurkat cells

To determine if the increased CpG abundance is sufficient to inhibit HIV-1 replicative fitness we synthesised *gag* sequences in which only the codons that led to an addition of CpG dinucleotides in the codon modified sequence were changed. These sequences were also inserted into pHIV-1_{NL4-3} to generate pHIV-1 GagCpG22-165, pHIV1 GagCpG22-261 and pHIV-1 GagCpG22-378 and viral fitness of the new mutant viruses was tested using a spreading infection assay in Jurkat cells. For that, we produced viral stocks by transfecting 293T cells with each proviral construct and we measured abundance of $p24^{Gag}$ by ELISA (Figure 5.2 B). While HIV-1 GagCp22-165 and HIV-1 GagCpG22-261 produced similar amounts of virus than HIV-1_{NL4-3}, HIV-1 GagCpG22-378 virus production was decreased by around 70%. The viral volume used to infect Jurkat cells was normalised so that 25ng of $p24^{Gag}$ were added and replication was monitored over 13 days (Figure 5.2 B). The replicative fitness of GagCpG22-165, which contains an additional of 11 CpGs, was substantially decreased with >90% less infectivity on the last day of the spreading assay compared to HIV-$1_{NL4-3.}$ Neither HIV-1 GagCpG22-261, which contains an additional of 18 CpGs nor HIV-1 GagCpG22-378, which contains an additional of 26 CpGs was able to replicate in Jurkat cells. Therefore, we concluded that addition of CpG dinucleotides is sufficient to attenuate HIV-1 replication in Jurkat cells.

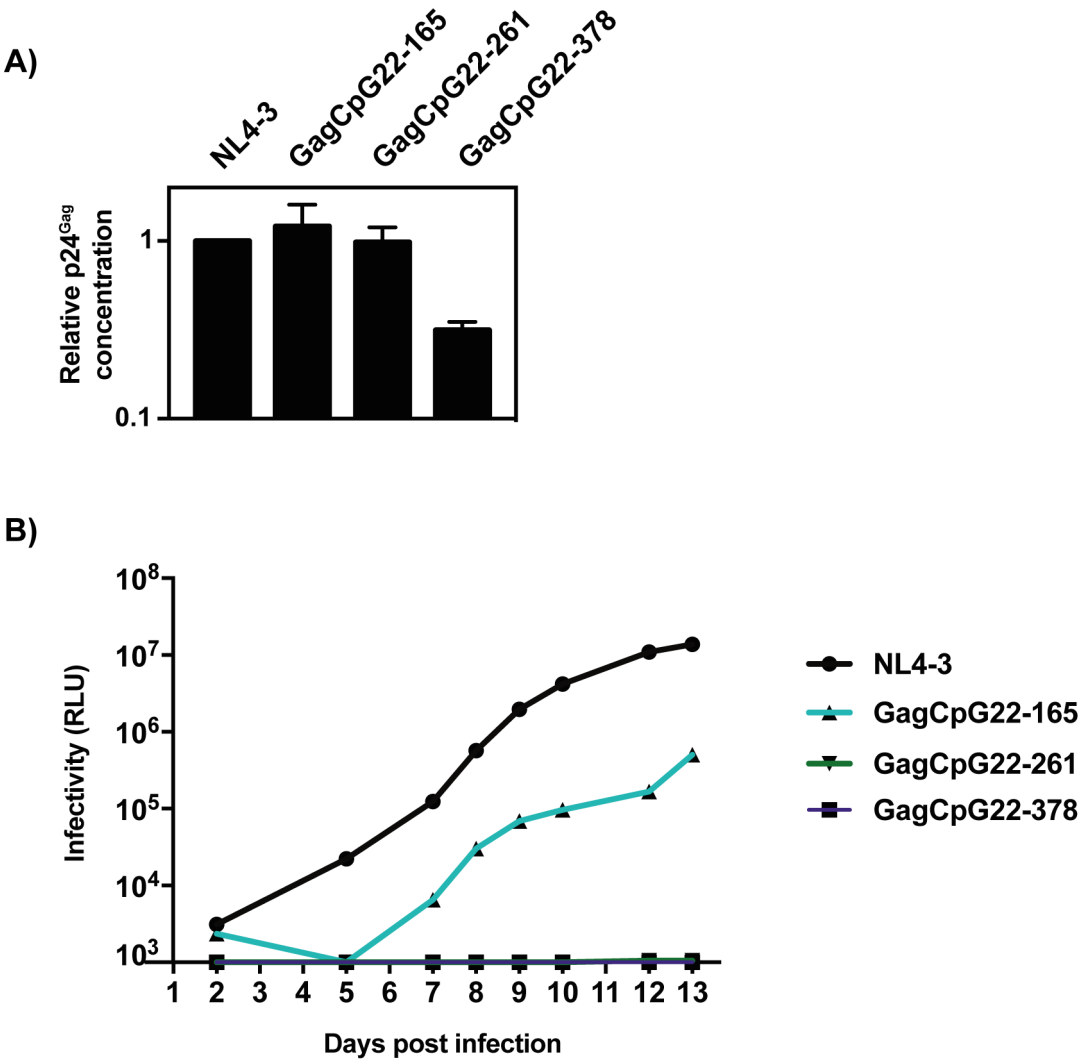


Figure 5.2. Introduction of CpG dinucleotides into *gag* inhibits HIV-1 replication in Jurkat cells. **A)** The amount of HIV-1 CA ($p24^{Gag}$) in supernatants from 293T cells transfected with pHIV-1_{NL4-3}, pHIV-1 GagCpG22-165, pHIV-1 GagCpG22-261 or pHIV-1 GagCpG22-378 were quantified by $p24^{Gag}$ ELISA. The bar chart is the average of three independent experiments normalised to HIV-1_{NL4-3}. Error bars represent standard deviation. **B)** Jurkat cells were infected with 25ng of $p24^{Gag}$ for each indicated virus. The amount of infectious virus present at each time point was measured in TZM-bl cells. This is representative of three independent experiments.

5.2.3 Introduction of CpG dinucleotides into *gag* inhibits infectivity but not Gag expression or virion production

To study which steps of the viral life cycle could be affected by the increased number of CpG dinucleotides and result in inhibited viral replication, HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCpG22-165, pHIV1 GagCpG22-261 and pHIV-1 GagCpG22-378. 48h post-transfection cell lysates and media were harvested to analyse infectivity and protein and RNA abundance. Although there was no change in Gag expression or virion production for HIV-1 GagCpG22-165, GagCpG22-261 and GagCpG22-378, the latter one had a 99% decrease in infectivity in single round replication assays (Figure 5.3 B). HIV-1 GagCpG22-261 had a ~75% decrease in infectivity and infectivity of HIV-1 GagCpG22-165 was very similar to wild-type (Figure 5.3 B). gRNA abundance of GagCpG22-165 and GagCpG22-261 were also similar to HIV-$1_{NL4-3.}$ HIV-1 GagCpG22-378, however, had a ~40% reduction in the cell lysate and ~80% reduction in the media (Figure 5.4). In addition, we have also identified a processing defect upon introduction of CpGs within the nucleotides 22-378, where the p40 band is more intense when compared to HIV-1_{NL4-3} (Figure 5.3 C, * denotes a processing defect). In conclusion, introduction of 26 CpG dinucleotides into nt 22-378 of HIV-1_{NL4-3} greatly inhibits infectivity but does not change the intracellular Gag expression or virion production. Moreover, it is important to mention that in comparison to GagCM22-378, introduction of CpG dinucleotides in the absence of the other synonymous mutations produces a smaller reduction in infectivity.

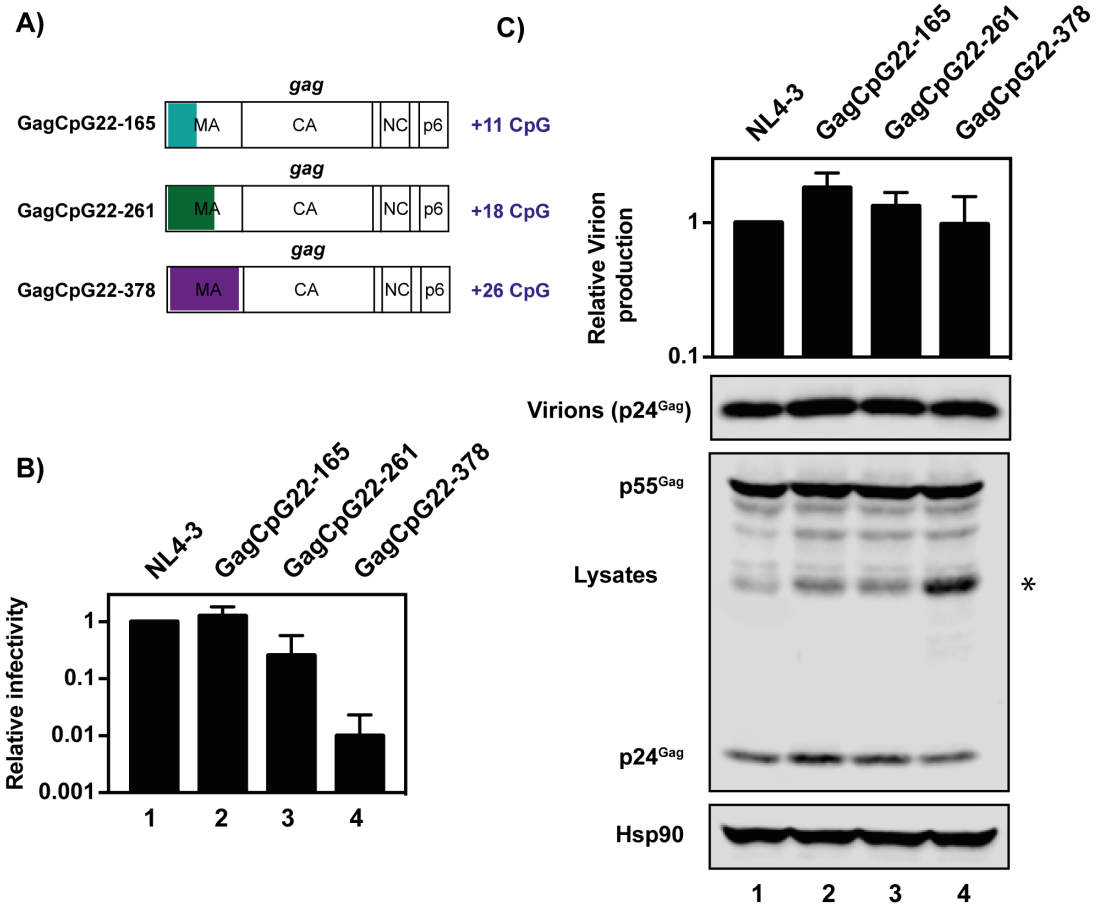


Figure 5.3. Introduction of CpG dinucleotides into *gag* inhibits infectivity in HeLa cells. HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCpG22-165, pHIV-1 GagCpG22-261 or pHIV-1 GagCpG22-378. **A)** Schematic representation of the virus mutants. **B)** Culture supernatants were used to infect TZM-bl reporter cells to measure viral infectivity. The bar charts show the average values of four independent experiments normalised to the value obtained for HIV-1_{NL4-3}. **C)** Gag expression in the media and cell lysate was determined by quantitative immunoblotting. The star (*) denotes a defect in processing. The bar charts show the average of three independent experiments normalized to HIV-1_{NL4-3}. Error bars represent standard deviation.

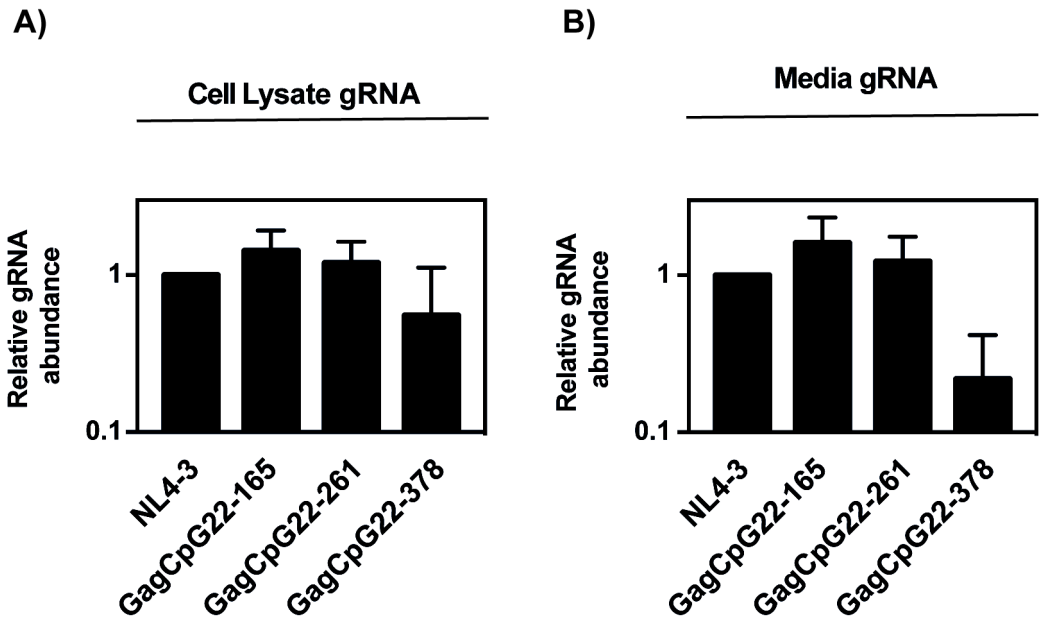


Figure 5.4. Introduction of CpG dinucleotides into nt 22-378 in *gag* inhibits gRNA abundance in HeLa cells. A) RNA was extracted from cell lysates and **B)** media and gRNA abundance was quantified by qRT-PCR. The bar charts show the average of three independent experiments normalised to HIV-1_{NL4-3}. Error bars represent standard deviation.

5.2.4 The context surrounding the CpG dinucleotides determines the inhibition in Gag expression and virion production

While HIV-1 GagCM22-378 and HIV-1 GagCpG22-378 have the same 26 introduced CpGs, there is a much larger decrease in gRNA abundance for HIV-1 GagCM22-378 than HIV-1 GagCpG22-378. Thus, we hypothesised that the milder inhibition of viral replication could be explained on the basis of a context-dependent RNA binding site. Inhibition could be the consequence of an RNA-binding protein that recognises and binds CpG dinucleotides to mediate splicing. Importantly, the binding region might comprise not only the CpG but also the surrounding nucleotides. To test this hypothesis we used WebLogo (286) to align the sequences surrounding the CpG dinucleotides of HIV-1 GagCM22-378 and GagCpG22-378. We generated a 12nt sequence that encompassed the five nucleotides 5' and 3' of the CpG and the alignment showed that the region surrounding is more G/C-rich in the codon modified sequence than in the wild-type HIV-1 sequence (Figure 5.5).

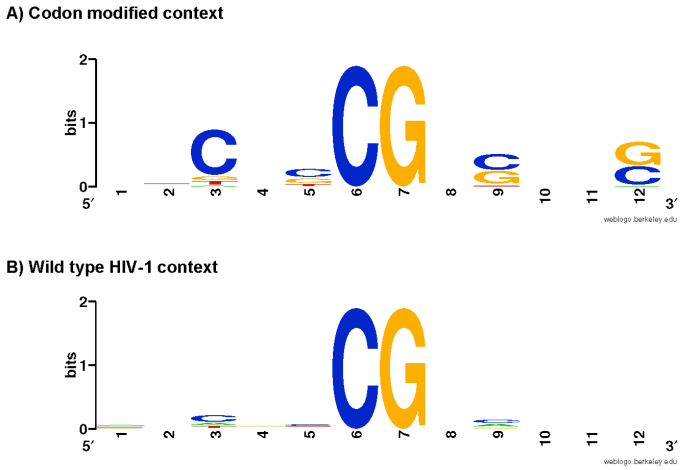


Figure 5.5. The CpG dinucleotide in the codon modified sequence is in a G/C-rich context. The sequence of the five nucleotides 5' and 3' to the CpG dinucleotides introduced into HIV-1 GagCM22-387 and HIV-1 GagCpG22-387 were aligned and a graphical representation of the sequence conservation was generated by WebLogo.

In order to better understand whether the sequence immediately neighbouring the CpG dinucleotides was necessary to drive inhibition, we synthesised a new construct in which only around 5 nucleotides upstream and downstream of the CpGs were codon modified (sequence details on Appendix 3). HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 GagCpG22-378, pHIV-1 GagCM22-378 and pHIV-1 GagCpG22-378_{5nt} and 48h post-transfection cell lysates and media were harvested to analyse infectivity and protein expression. Of note, all HIV-1 mutants contained 26 CpG dinucleotides. Interestingly, infectivity, virion production and intracellular Gag expression of pHIV-1 GagCpG22-261_{5nt} was comparable to that obtained for HIV-1 GagCM22-378 (Figure 5.6). These results show that codon modifying the region surrounding the CpG dinucleotides in addition to the CpG change, is sufficient to produce a similar level of inhibition in Gag expression than the codon modified construct. To further study the mechanism by which pHIV-1 GagCpG22-378_{5nt} inhibits infectious virus production, we measured Env expression and showed that either GagCM22-378 or GagCpG22-378_{5nt} greatly inhibit Env expression. This indicates that the region surrounding the CpG dinucleotide contributes to the mechanism by which HIV-1 replication is inhibited and the defect in infectious virus production of GagCpG22-378_{5nt} is also probably due to a defect in splicing.

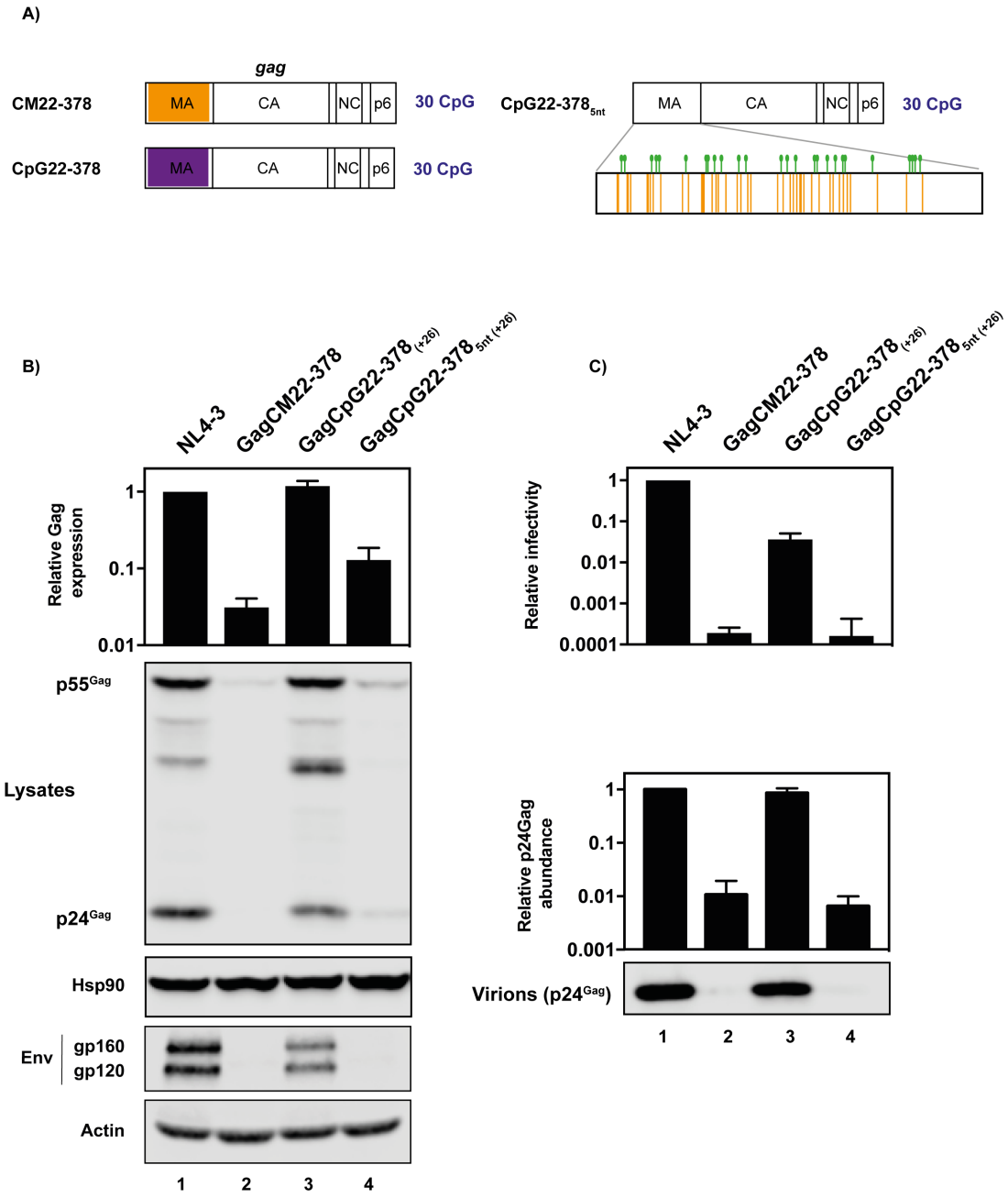


Figure 5.6. Codon modification of nucleotides surrounding the CpG dinucleotides inhibits infectious virus production. HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-261, pHIV-1 GagCpG22-261 or pHIV-1 GagCpG22-261_{5nt}. **A)** Schematic representation of the HIV-1 mutants constructs. The green lollipops represent CpG dinucleotides and the orange regions the codon modified nucleotides **B)** Gag and Env expression in the media (D) and cell lysate (B) was detected using quantitative immunoblotting. **C)** Culture supernatants were used to infect TZM-bl reporter cells to measure viral infectivity. The bar charts show the average values of three independent experiments normalised to the value obtained for HIV-1_{NL4-3}. Error bars represent standard deviation.

5.3 Summary

In this chapter, we have identified CpG dinucleotides as the inhibitory element that is inducing the defect in viral replication. First, we have shown that as result of synonymous mutagenesis we have unintentionally added CpG dinucleotides into nucleotides 22-261 and 22-378 (Table 5.2). Upon lowering the CpG abundance to wild-type levels we have rescued infectious virus production (Figure 5.1). Moreover, adding CpG dinucleotides in the absence of other synonymous mutations has shown that CpGs are necessary to inhibit viral replication (Figure 5.2). Importantly, we have demonstrated that while adding 26 CpG dinucleotides greatly inhibits viral replication, this does not occur as consequence of a reduced gRNA abundance (Figure 5.4) and Gag expression or virion production (Figure 5.3), suggesting that the inhibitory mechanism is different from the one observed in chapter 4. Moreover, we have determined that the nucleotide sequence surrounding the CpGs is important to induce an inhibition in viral replication that involves a decrease in Gag expression and virion production (Figure 5.6).

Chapter 6: The mechanism by which CpGs inhibit viral replication is position-dependent

6.1 Introduction

It is known that CpG dinucleotides are suppressed in RNA viruses (332, 341, 342) and we have shown that introducing CpGs into the MA region of Gag inhibits viral replication. In addition, analysis of the sequences of HIV-1-infected patients has highlighted the importance of CpG dinucleotides *in vivo* (344, 345). Mutations that lead to an increase in the CpG dinucleotide abundance in *pol* has been shown to negatively affect viral fitness, in comparison to mutations that do not increase CpG dinucleotide abundance (345). On the other hand, patients with high-CpG content in HIV-1 *env* appear to present a slower-progression clinical course of the disease (344).

The mechanism by which CpG dinucleotides repress viral fitness, however, is still unclear. Burns et al. (356) demonstrated that viral fitness of poliovirus could be reduced to the threshold of viability upon introduction of CpG dinucleotides in the capsid region of the genome, which constitutes only a 9% of the total viral genome. Infectivity was shown to be inversely correlated with the number of CpGs that were added and they stated this was not a result of a defect in translation (356). Atkinson et al. (357) showed that decreasing CpG frequency of picornavirus echovirus 7 (E7) could increase viral replication. They showed a close correlation between the viral fitness and the number of CpGs removed. In addition, decreasing the number of CpG dinucleotides in different regions could lead to differences in enhanced replication as a reflection of the natural dinucleotide suppression of the region, where increasing the number of nucleotides could have a greater effect in those regions where CpG abundance was naturally lower (357). It is known that E7 effectively blocks IFNβ-mediated cellular responses in the infected cells. Treatment of the mutant viruses with C16, a generally considered protein kinase inhibitor, however, restored viral fitness levels. This results suggested that CpG-containing viruses are not defective itself,

but are likely to induce an innate immune response. Although initial experiments led to hypothesis that PKR was mediating the reduction in viral fitness, they failed to reproduce this results using siRNA or shRNA (357). Tulloch et al. (354) showed that attenuation of the replication of picornavirus with elevated CpG dinucleotide abundance was not the result of a defect in translation or altered codon pairs. Instead, they proposed to be mediated through an enhanced innate immune responses. They suggest that the viral attenuation is not due to the high-CpG viruses being defective but are more recognisable by the cell, which prevents replication (354). Gaunt el al. (358) showed that elevated CpG frequencies in segment 5 of influenza A also led to a decrease in viral fitness. Interestingly, viral attenuation was also observed *in vivo* where the viral load of mutant viruses was greatly reduced in the lungs of infected mice, which had a substantially milder clinical course. They stated that this difference in the induced pathogenicity suggests the existence of host factors that can induce different immune responses (358).

Recently Takata et al. (360) showed that the zinc-finger antiviral protein (ZAP) inhibits replication of cells infected with high-CpG HIV-1. Specifically, the analysis of multiple mutant viruses resulted in the division of three phenotypic groups. Group 1 mutants showed near wild-type replication, Group 2 mutants showed severe splicing defect that resulted in attenuated replication and Group 3 mutants also showed defective viral replication but in the absence of an altered splicing pattern. Importantly, high-CpG content in *env* was shown to be essential to suppress HIV-1 replication (360). Moreover, they proved that ZAP selectively binds RNA sequences containing CpG dinucleotides and this recognition results in the degradation of high-CpG viral RNAs. Given that most cellular mRNAs are not affected by ZAP (361), they suggested that this antiviral protein exploits the low CpG abundance of the cellular genes to identify foreign RNA.

ZAP was first discovered as a restriction factor in MLV (Obs/Exp CpG ratio 0.51) as a result of a screen that targeted genes that prevent retrovirus infection (362). Later studies have also shown the involvement of ZAP in viral restriction of certain filoviruses (Ebola virus (Obs/Exp CpG ratio 0.60) and

Marburg virus, (Obs/Exp CpG ratio 0.53)) and alphaviruses (Sindbis virus (SINV Obs/Exp CpG ratio 0.90) and Semliki Forest virus (Obs/Exp CpG ratio 0.89)) (363-366). ZAP inhibits viral replication by preventing accumulation of viral RNAs in the cytoplasm and binding of this cellular factor to viral RNA seems necessary to drive the restriction (367-369). ZAP contains four CCCH zinc finger motifs that enable the binding to RNA, but only the second and the fourth appear necessary for ZAP's antiviral activity (367, 369). Depletion of the hRrp46p subunit of the exosome by siRNA substantially decreases ZAP's restriction activity suggesting that ZAP might alter RNA stability by recruiting the exosome (370, 371). ZAP-induced RNA degradation starts with the recruitment of the cellular poly(A)-specific ribonuclease (PARN) which shortens the poly(A)A tail of the mRNA allowing the exosome to start degrading from the 3' end (372). In addition, it has also been shown that ZAP enables 5' degradation by recruiting the cellular decapping complex via its cofactor RNA helicase p72 (372). Although the human ortholog of ZAP is expressed as a short (ZAP-S) or a long (ZAP-L) isoform, ZAP-L shows stronger restriction against MLV and Semliki forest virus (373). The short isoform lacks the C-terminal poly(ADP-ribose) polymerase (PARP)-like domain. Interestingly, there are some viruses such as yellow fever virus and herpes simplex virus type 1 (HSV-1) and poliovirus that have been shown not to be restricted by ZAP (363). ZAP has also been proposed to inhibit wild-type HIV-1 (372) although these results remain controversial since other groups have not been able to reproduce these results (360, 373).

6.2 Results

6.2.1 Inhibition of infectious virus production upon introduction of CpG dinucleotides is ZAP-independent

To analyse whether the CpG dinucleotides in HIV-1 GagCM22-378 and HIV-1 GagCpG22-378 induce inhibition of Gag expression and infectious virus production through a mechanism related to ZAP, we generated CRISPR-Cas9 ZAP knock out HeLa cells (Figure 6.1 A) and transfected them with pHIV-1_{NL4-3}, pHIV1 GagCpG22-378, pHIV-1 GagCM22-378 and pHIV-1 EnvCpG87-561 (Figure 6.1). Mutant L in Takata et al. (360) contains 37 CpGs in *env* and these transcripts are recognised by ZAP and depleted. As a positive control we used EnvCpG87-561, which contains 36 CpGs and is an almost identical version of the L mutant that was shown to have a great reduction in viral replication (360). Around 48h post-transfection we collected the cell lysates to analyse the protein expression and the media to analyse virion production and infectivity in single round replication assays. In control HeLa cells Gag expression, virion production and infectivity were inhibited for HIV-1 EnvCpG87-561 and HIV-1 GagCM22-378 and also for HIV-1 GagCpG22-378 (Figure 6.1). In ZAP KO cells, however, Gag expression, virion production and infectivity were rescued to wild-type levels for HIV-1 EnvCpG87-561 but not for HIV-1 GagCM22-378. Gag expression and virion production of HIV-1 GagCpG22-378 was slightly increased in the ZAP-KO cells but not infectivity (Figure 6.1 B, C). These results suggest that introduction of CpG dinucleotides into *gag* leads to inhibition of viral replication in a ZAP-independent manner.

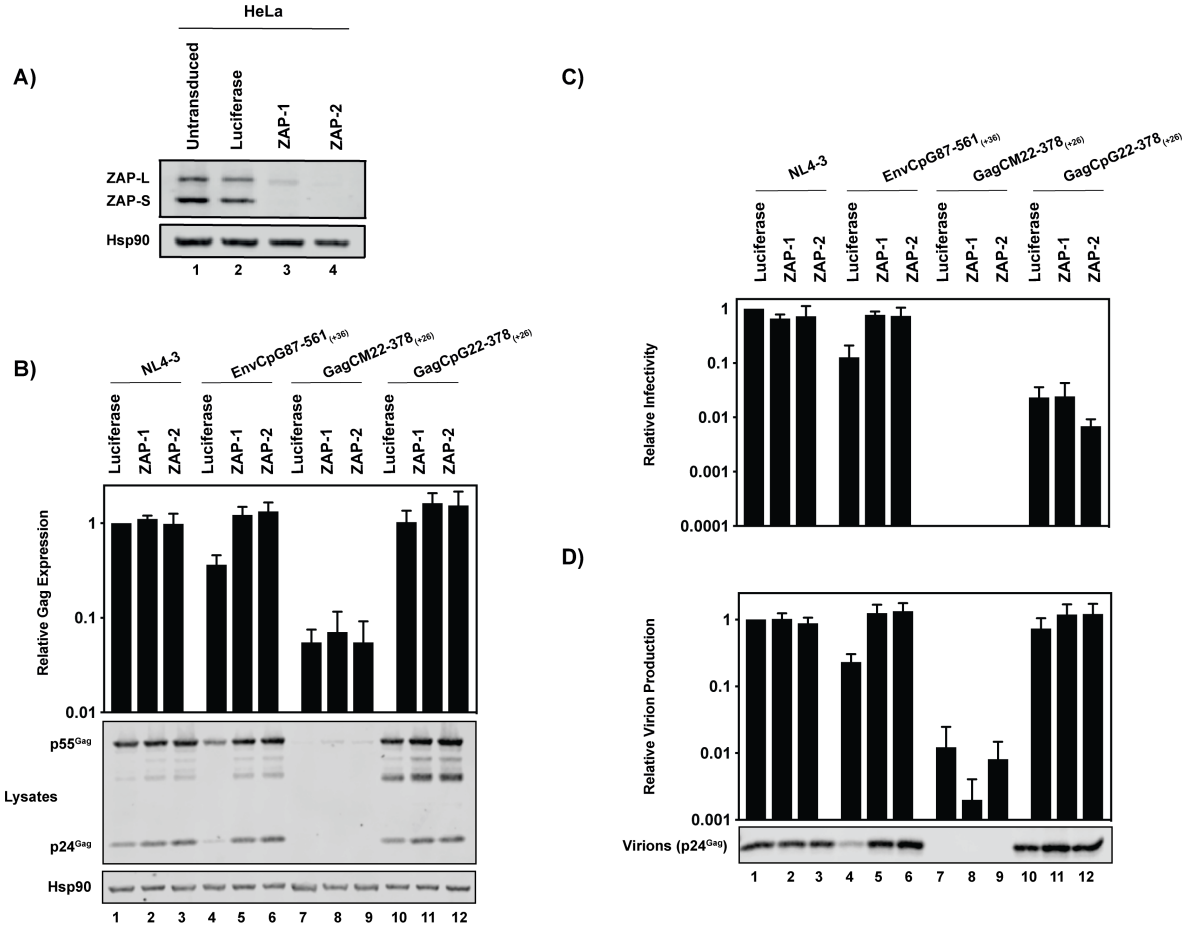


Figure 6.1. Inhibition of infectious virus production upon introduction of CpG dinucleotides is ZAP-independent. ZAP-KO HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 EnvCpG87-561, pHIV-1 GagCM22-261 or pHIV-1 GagCpG22-261. **A)** HeLa ZAP-KO cells with two different guides **B)** and **D)** Gag expression in the media (D) and cell lysate (B) was detected using quantitative immunoblotting. The bar charts show the average of four independent experiments normalised to HIV-1_{NL4-3}. Error bars represent standard deviation. **C)** Culture supernatants were used to infect TZM-bl reporter cells to measure viral infectivity. The bar charts show the average values of three independent experiments normalised to the value obtained for HIV-1_{NL4-3}. Error bars represent standard deviation. The ZAP KO cell lines and these experiments were generated by Mattia Ficarelli using constructs that I had cloned (PhD student, Swanson Lab, KCL).

6.2.2 Changing the codon to one that does not lead to the introduction of CpGs does not inhibit infectious virus production

We have shown that viral attenuation upon introduction of 26 CpGs into HIV-1 *gag* is driven by a mechanism that is independent of ZAP. One hypothesis that could explain a defect in viral replication is a defect in packaging. For some retroviruses, the packaging signal (ψ) is sufficient to direct encapsidation of the viral genome. HIV-1, however, requires not only ψ but also a minimal region in the 5' leader and other poorly defined additional features such as initial *gag* coding sequences (68, 69, 374, 375). In addition, it has been shown that the presence of the HIV-1 5'UTR together with the 5' end of *gag* can mediate efficient packaging of non-viral RNAs (376). Other groups have also suggested the importance of the *gag* sequence in packaging, showing that alterations of the *gag* sequence such as codon modification could lead to poor packaging efficiency (377). It has been suggested that the introduction of synonymous mutations might induce misfolding of the upstream dimerisation and packaging signals and thus interfere with these processes (377).

To determine if the introduced CpGs specifically induced the decrease in infectious virus production or whether a necessary *cis*-acting element in this region had been mutated, we changed the codon that led to an increase in the CpG abundance to a different codon (DC) producing pHIV-1 GagDC22-378. This virus contains the 23 mutations and 4 CpG dinucleotides, like HIV-1_{NL4-3} within the same region (Figure 6.2 A). We then transfected HeLa cells with pHIV-1_{NL4-3}, pHIV-1 GagCM22-378, pHIV-1 GagCM22-378_{lowCpG}, pHIV-1 GagCpG22-378 and HIV-1 GagDC22-378 and harvested cell lysates and media ~48h post-transfection. Cell lysates were used to analyse protein and RNA abundance and media was used to analyse infectivity, virion production and RNA abundance. Gag expression, virion production and RNA in the cell lysates and media for HIV-1 GagDC22-378 were all shown to be at very similar levels than wild-type HIV-1_{NL4-3}, GagCM22-378_{lowCpG} and pHIV-1 GagCpG22-378 (Figures 6.2 and 6.3). Infectivity of GagDC22-378 was

similar to HIV-1_{NL4-3} and GagCM22-378_{lowCpG} (Figure 6.2 C). These results show that changing the codon in those specific positions does not alter the amount of gRNA that is being packaged into the virion, and therefore confirms that viral attenuation is the result of an increased CpG abundance and not due to altered packaging.

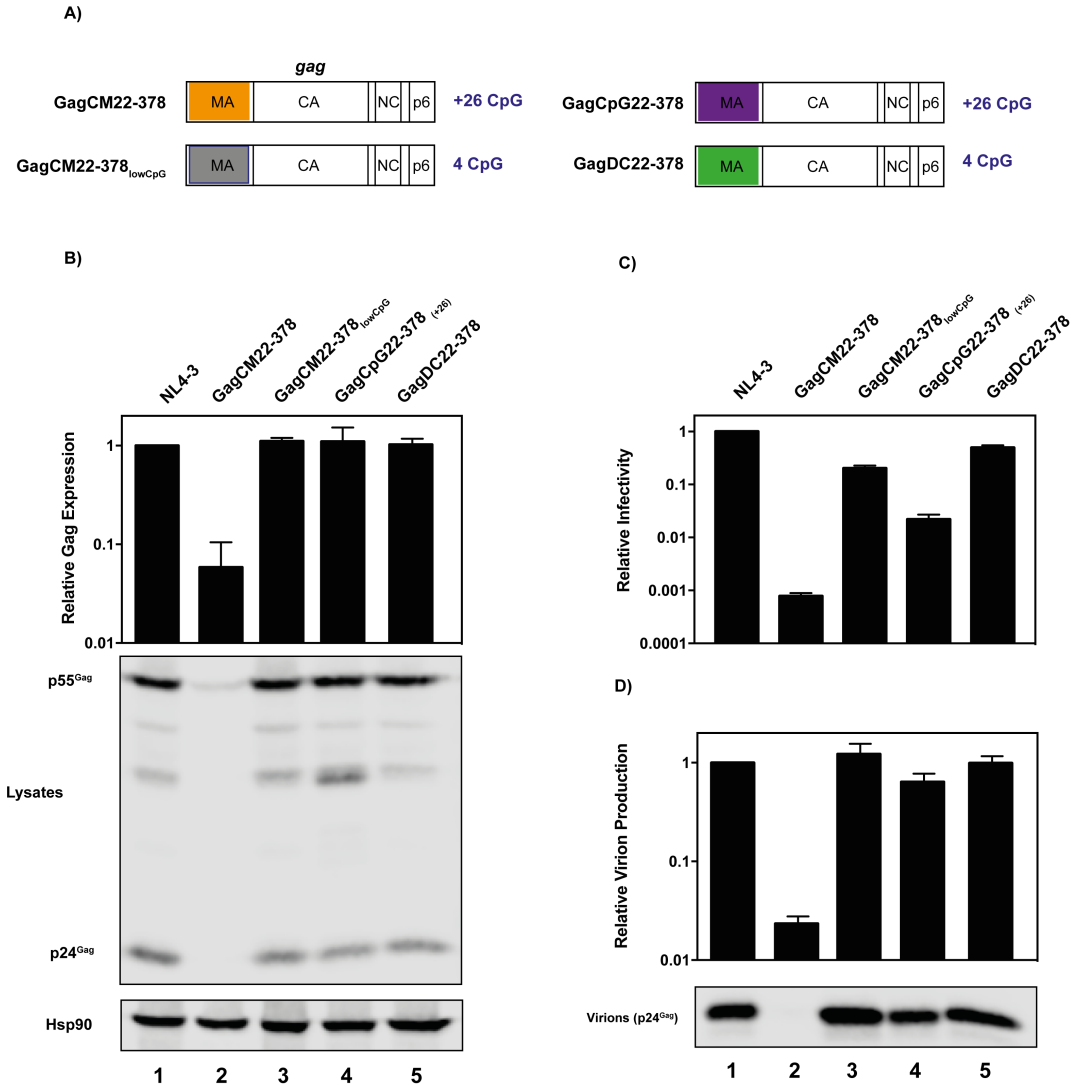


Figure 6.2. Changing the codon to one that does not result in an increased number of CpGs does not inhibit infectious virus production. HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-378, pHIV-1 GagCM22-378_{CpGlow}, pHIV-1 GagCpG22-378 or pHIV-1 GagDC22-378. **A)** Schematic representation of the HIV-1 mutants constructs. **B)** Gag expression in the media **(D)** and cell lysate (B) was detected using quantitative immunoblotting. **C)** Culture supernatants were used to infect TZM-bl reporter cells to measure viral infectivity. The bar charts show the average values of three independent experiments normalised to the value obtained for HIV-1_{NL4-3}. Error bars represent standard deviation.

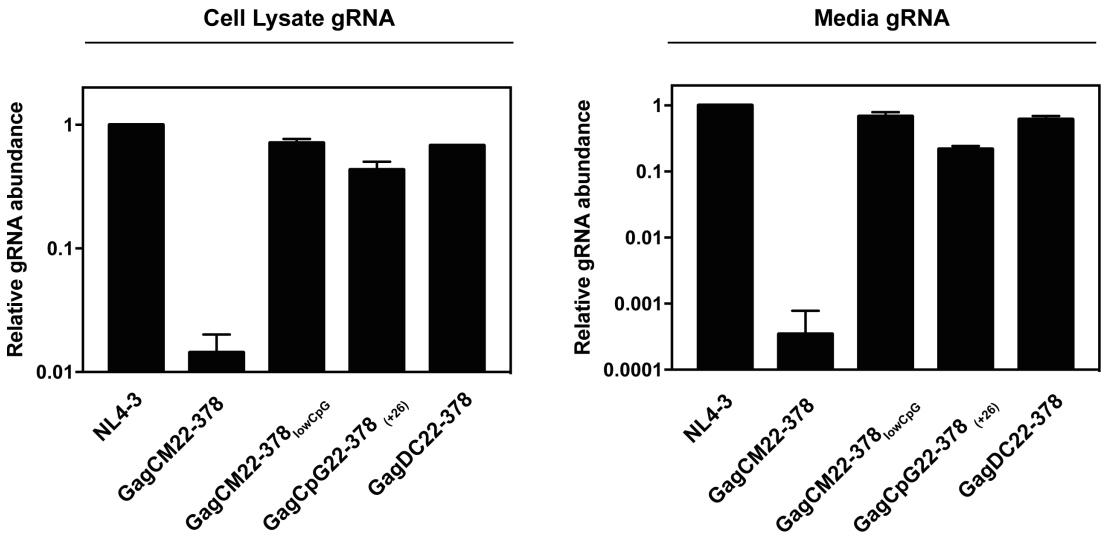


Figure 6.3. Changing the codon to one that does not result in an increased number of CpGs does not decrease gRNA abundance. HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCM22-378, pHIV-1 GagCM22-378_{CpGlow}, pHIV-1 GagCpG22-378 or pHIV-1 GagDC22-378. **A)** RNA was extracted from cell lysates **B)** and media and gRNA abundance were quantified by qRT-PCR. The bar charts show the average of three independent experiments normalised to HIV-1_{NL4-3}. Error bars represent standard deviation.

6.2.3 Introduction of CpG dinucleotides into a longer region of *gag* inhibits infectious virus production in HeLa cells

To better understand the mechanism by which an increased CpG abundance in *gag* can attenuate viral replication we cloned a larger region of high-CpG *gag* into pHIV-1_{NL4-3} to produce pHIV-1 GagCpG22-1188, which contains 62 CpG dinucleotides. In addition, we produced pHIV-1 GagCpG22-654 and pHIV-1 GagCpG658-1188, which contain 30 and 32 CpGs, respectively (Figure 6.4 A). We transfected HeLa cells with HIV-1_{NL4-3} and the newly cloned constructs and collected the cell lysates and media ~48h post-transfection. pHIV-1 GagCpG22-1188 and HIV-1 GagCpG22-654 produced a large inhibition in Gag expression, viral production and infectivity (Figure 6.4). Infectivity of HIV-1 GagCpG658-1188, however, was very similar to HIV-1_{NL4-3} (Figure 6.4 C). Of note, while HIV-1 GagCpG22-654 and HIV-1 GagCpG658-1188 have very similar amount of CpG dinucleotides, the later one does not inhibit viral infectivity, suggesting that the absolute number of CpGs added might not solely drive viral attenuation, and highlights the importance of the CpG dinucleotides position within the HIV-1 genome.

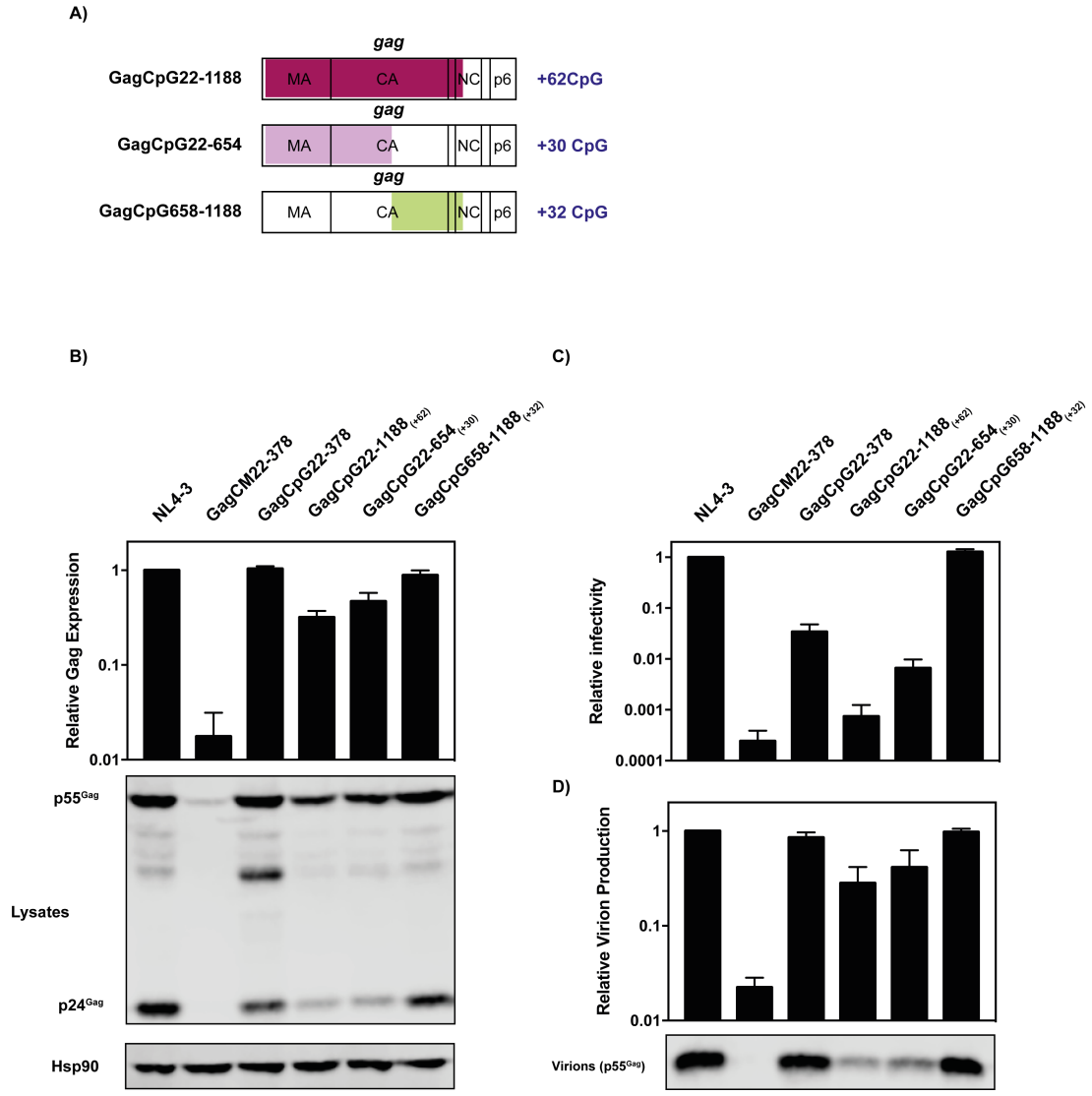


Figure 6.4. Introduction of CpG dinucleotides into nt 22-1188 in *gag* inhibits infectious virus production in HeLa cells. HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 GagCpG22-1188, pHIV-1 GagCpG22-654 or pHIV-1 GagCpG658-1188. **A)** Schematic representation of the mutant viruses. **B)** Gag expression in the cell lysate **D)** and in the media was determined by quantitative immunoblotting. **C)** Culture supernatants were used to infect TZM-bl reporter cells to measure viral infectivity. The bar charts show the average values of three independent experiments normalised to the value obtained for HIV-1_{NL4-3}.

6.2.4 Inhibition of infectious virus production upon introduction of CpG dinucleotides into a longer region in *gag* is ZAP-dependent

While introducing 26 CpGs into GagCpG22-278 did not inhibit HIV-1 replication in a ZAP-dependent manner, reduction in infectious virus production upon introducing 36 CpGs into EnvCpG87-561 was rescued in ZAP KO cells. To determine whether introducing a higher number of CpGs into *gag* could inhibit HIV-1 replication in a ZAP-dependent manner we transfected HeLa ZAP-KO cells with HIV-1_{NL4-3}, pHIV-1 EnvCpG87-561 and pHIV-1 GagCpG22-1188 and measured protein expression and infectivity. As observed previously, addition of 62 CpG dinucleotides into *gag* greatly inhibited Gag expression, virion production and infectivity. Intriguingly, the defect in Gag expression and virion production was greatly rescued in ZAP KO HeLa cells in levels comparable to the control HIV-1 EnvCpG87-561, showing that the replication defect produced by a higher number of CpG dinucleotides is ZAP-dependent (Figure 6.5). Of note, infectivity of HIV-1 GagCpG22-1188 was only slightly higher in ZAP KO cells.

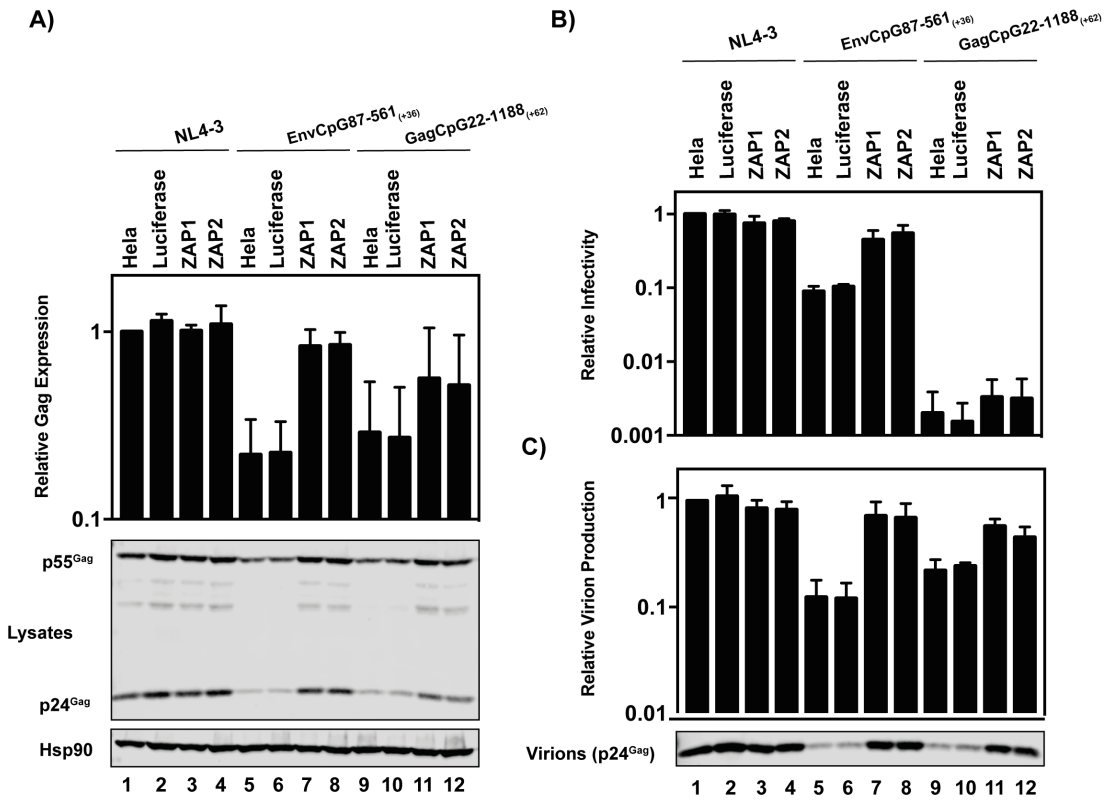


Figure 6.5. Inhibition of infectious virus production upon introduction of CpG dinucleotides into nt 22-1188 in *gag* is ZAP-dependent. ZAP-KO HeLa cells were transfected with pHIV-$1_{NL4-3,}$ pHIV-1 EnvCpG87-561, or pHIV-1 GagCpG22-1188. **A)** and **C)** Gag expression in the media (C) and cell lysate (A) was detected using quantitative immunoblotting. **B)** Culture supernatants were used to infect TZM-bl reporter cells to measure viral infectivity. The bar charts show the average values of three independent experiments normalised to the value obtained for HIV-1_{NL4-3}. Error bars represent standard deviation. The ZAP KO cells were generated by Mattia Ficarelli (PhD student, Swanson Lab, KCL).

6.3 Summary

In this chapter, we have shown that codon modifying (GagCM22-378) or introducing CpGs (GagCpG22-378) within nucleotides 22-278 inhibits HIV-1 replication in a ZAP-independent manner (Figure 6.1). In addition, we have shown that the introduction of synonymous mutations has not altered any necessary RNA element, confirming that CpG dinucleotides specifically induce the defect in replication (Figure 6.2). Importantly, adding a higher number of CpGs (GagCpG22-1188) into *gag* inhibits viral replication through reducing Gag expression and virion production (Figure 6.4). Intriguingly, while adding 26 CpG dinucleotides into GagCpG22-378 does not inhibit Gag expression and virion production, adding a 30 CpGs into a different region of *gag* inhibits infectious virus production through decreasing Gag expression and virion production (GagCpG22-654, Figure 6.4). Moreover, introducing CpGs into nucleotides 22-1188 (GagCpG22-1188) inhibits viral replication in a ZAP-dependent manner (Figure 6.5).

Chapter 7: Discussion

Herein we have used synonymous mutagenesis to experimentally introduce CpG dinucleotides into HIV-1 *gag*. We have shown that CpG-mediated reduction in HIV-1 replication is carried out through at least 3 different mechanisms. First, we have shown that introducing CpG dinucleotides through synonymous mutagenesis (codon modification) inhibits viral replication by generating changes in the splicing pattern (GagCM22-378). Second, we have shown that adding CpG dinucleotides in the absence of other synonymous mutations also reduces viral infectivity through a mechanism that does not affect Gag and virion production but does produce a decrease in the gRNA abundance in the media (GagCpG22-378). Third, we have shown that introducing CpG dinucleotides into a longer region in *gag* inhibits viral replication in a ZAP-dependent manner (GagCpG22-1188).

Synonymous mutagenesis has previously been used to induce viral attenuation. Although the decrease in viral replication was attributed to changes in codon pair bias (378, 379) or a decrease in the A-rich content of the HIV-1 genome (380), it is now known that CpG dinucleotides were unintentionally added into their viral constructs. Moreover, a more recent study has shown that changing the A-content in *pol* does not affect the viral fitness of HIV-1 (381) and CpGs are the underlying mechanism by which viral replication is inhibited in HIV-1 (360, 382).

Coleman et al (379), for example, manipulated the poliovirus genome by synthesising a novel poliovirus with a modified codon pair bias. This group developed a computer algorithm that could recode the viral genome without altering other viral features that might be relevant for viral replication such as codon bias or the folding free energy of the RNA. While the amino acid sequence remained intact, they used rare codons to encode them (379). The strategy was named synthetic attenuated virus engineering (SAVE). Martrus et al. (378) strongly inhibited HIV-1 replication by introducing codon pairs that are underrepresented relative to human mRNAs into *gag* and *pol* (378). The repercussion that the introduction of synonymous mutations could have in

the abundance of CpG dinucleotides, however, was not studied. Indeed, the mutagenesis carried out in the study led to an increase in the number of CpG dinucleotides in *gag* from 15 to 118. Thus, we hypothesise that the inhibition in viral fitness observed could at least to a degree be a result of increasing CpG frequency instead of codon pair deoptimisation. Importantly, recent data has shown that codon pair bias is a direct consequence of nucleotide bias and codon pair deoptimisation increases CpG abundance (339, 354).

Keating et al. (380) stated that HIV-1 relies on an A-rich sequence to support the synthesis of viral cDNA during reverse transcription. The study was carried out by codon modifying regions in *gag* and *pol* where they increased the number of codons that were C- and G-rich. They argued that manipulating the nucleotide content may inhibit reverse transcription through different mechanisms. First, the nucleotide changes could alter the RNA structure and inhibit the reverse transcription complex as a result of impaired interactions between the viral RNA and NC and other factors such as RNA helicase. Second, decreasing the A-richness in *pol could* increase RNA dimer stability and also result in impaired synthesis of viral cDNA. Alternatively, the inhibition of reverse transcription can also be due to changes in the RNA structure that alters the polymerase processivity (380). The source of the codon modified sequence used in this article, however, is phGag-Pol (383), which contains a substantially higher CpG dinucleotide abundance when compared to the wild-type region.

Importantly, and as mentioned previously, lowering the A-content in a 493 nucleotide region in *pol* by 15% or increasing it by a 5% has been demonstrated not to affect the replicative potential of HIV-1 (381). These results demonstrate that the A-rich signature is not essential for HIV-1 replication. In sum, we highlight the importance of CpG dinucleotides as a potent tool to inhibit viral replication.

7.1 Introduction of CpG dinucleotides in the context of codon modification, inhibits HIV-1 replication through altering the splicing pattern

The first inhibitory mechanism was characterised upon codon modifying nucleotides 22-261 in HIV-1 *gag*. Introducing CpG dinucleotides in the context of codon modification within the first ~200 nucleotides in the 5' (GagCM22-261) end of *gag* greatly inhibited viral replication (Figures 3.1 and 3.2). We determined that the decrease in viral fitness was due to a decrease in the gRNA abundance, which leads to a decrease in Gag expression and virion production (Figures 3.3 and 3.4). RNA stability analysis showed that the more synonymous mutations were introduced into *gag* the less stable the RNA was (Figure 4.2). RNAseq analysis revealed that codon modification of nucleotides 22-378 activated a cryptic splice site (CD1) downstream of D1 and altered the splicing pattern (Figure 4.4 and Table 4.1). We showed that as a result of that shift in the use of D1 towards the use of CD1 a new AUG was introduced upstream of *gag* in all the transcripts (Figure 4.4). These results demonstrate that by adding CpGs in the context of CM, we have either introduced an activator or removed a suppressor that enhances the utilisation of the cryptic splice site that plays a major role in regulating splicing. Alternatively, codon modification in *gag* might have altered a protein-binding site or a secondary structure that is normally suppressing the utilisation of the cryptic splice site. Efficient HIV-1 replication requires a balanced splicing, where alternative splicing regulatory signals play an important role in maintaining an adequate level of each pool of spliced and unspliced transcripts so that the necessary viral proteins are available. Importantly, the inclusion of the upstream start codon could reduce the efficiency at which Rev is translated. Rev functions not only in nuclear export but also in increasing the stability of the mRNA pool (328). Indeed, Rev-deficient HIV-1 has previously been related to a decreased RNA stability of unspliced transcripts (328). Taken all together, we hypothesise that the altered splicing patterns can lead to a reduction in Rev expression, which can affect the stability of intron-containing mRNAs such as the gRNA and *env*.

Env abundance might be reduced not only as a consequence of a decrease in the *env* mRNA but also as a result of Env being less efficiently translated due to the AUG upstream. The decrease in relative infectivity/genome (Figure 3.4), therefore, could be at least partially explained on the basis of a decrease in Env expression. Supporting this hypothesis, a recent study has also shown the existence of multiple elements in g*ag* that function to regulate HIV-1 alternative splicing (384). In particular, Group 2a mutants in this study, which exhibited replication defects accompanied by changes in RNA splicing, revealed the presence of an RNA element that is suppressing the use of a cryptic splice site in the 5' end of the viral genome. In addition, they also showed that the introduction of synonymous mutations that targeted the central region of the genome resulted in the overuse of A1, A2 and A3 (384). In this project, in addition to identifying an altered splicing pattern, we have determined that this happens through a mechanism that is dependent on CpG dinucleotides.

7.2 The region surrounding the CpG dinucleotides contributes to the mechanism by which HIV-1 replication is inhibited

Second, we have also demonstrated that introduction CpG dinucleotides in HIV-1 *gag,* inhibits viral replication through a mechanism that is context and position dependent. It is important to highlight that although the same number of CpGs were added into GagCpG22-378 and GagCM22-378, the phenotype caused by these viruses is very different. By comparing the infectious virus production of HIV-1 mutants that lacked a region in *gag* or contained synonymous mutations in *gag* (GagCM22-378 and GagΔ22-378), we demonstrated that none of the RNA elements within nt 22-378 appears necessary for infectious virus production in HeLa cells (Figure 4.1). Moreover, these results showed that silent mutagenesis in the context of codon modification resulted in the introduction of inhibitory sequences. Indeed, we have shown that lowering the CpG-content to match the wild-type count (4 CpG) completely rescues infectious virus production. Besides, we added 18 (GagCpG22-261) or 26 CpG dinucleotides (GagCpG22-378) into

gag and resulted in a large inhibition of viral replication, which demonstrated that the decrease in viral fitness was due to an increase in the CpG abundance. These changes in the dinucleotide composition, however, did not substantially change Gag expression and virion production. As a result, we hypothesised that the sequence neighbouring CpG dinucleotides might play an important role in the CpG-mediated inhibition so we cloned a new virus mutant in which the 5 nucleotides upstream and downstream of the CpG dinucleotides were codon modified (GagCpG22-378_{5nt}). Changing the sequence immediately surrounding the CpGs led to an inhibition that was comparable to that caused by GagCM22-378 (Figure 5.5). Therefore, these results highlighted the importance of the CpG-neighbouring sequence in the CpG-mediated restriction. Because the phenotype caused by GagCpG22-378_{5nt} and GagCM22-378 is very similar, both in regards to Gag expression and Env expression, we hypothesise that inhibition of infectious virus production of GagCpG22-378_{5nt} is also a result of aberrant splicing. While the precise mechanism by which introducing CpG dinucleotides in *gag* (GagCpG22-378) leads to inhibition of HIV-1 replication remains to be fully characterised, we have shown that it does not decrease Gag expression and virion production but it leads to a decrease in the gRNA in the media (Figure 5.3 and 5.4). Recently, the importance of posttranscriptional ribonucleoside modifications in the regulation of RNA structure and function has been highlighted. Acetylation of cytosine by the acetyltransferase NAT10 (ac4C), for example, was shown to enhance translation efficiency. Moreover, depletion of this protein has been shown to result in a decrease of mRNA stability (385). It is possible that the introduction of synonymous mutations has altered the epitranscriptome pattern and resulted in reduced viral fitness.

Several studies have proposed that restriction of viral replication is the result of an antiviral mechanism that is activated in response to a sequence-specific sensing (349, 354, 357, 386). Toll-like receptors (TLRs) are a family of transmembrane receptors that can be found in several immunologically relevant cells such as dendritic cells (DCs), macrophages and B cells (387, 388). Four TLR family members have been implicated in nucleic acid recognition: TLR3, TLR7, TLR8 and TLR9 (389, 390). Generally, activation of

TLRs by recognition of endosomal RNA results in the production of cytokines, and upregulation of MHC and other co-stimulatory molecules than induce both innate immunity and activation of adaptive immunity, contributing to an antiviral mechanism (390).TLR9 is a pattern recognition receptor that can sensor methylated CpGs (349, 391). CpG dinucleotides are also suppressed in RNA viruses that do not have a DNA intermediate (343, 346, 349-351), which indicates that CpG methylation cannot be the mechanism driving CpG-mediated suppression. Although TLR9 is not likely to be expressed in HeLa cells, we confirmed that STAT1 phosphorylation or IFIT1 expression was not detected in response to CpG DNA (Appendix 1).

Other studies carried out in the picornavirus echovirus 7 have also shown CpG-mediated inhibition of viral replication by a mechanism that was not related to an altered translation efficiency (354), an interferon-related pathway, PKR or through conventional pattern recognition receptors (357). Multiple studies have proposed the existence of an innate immune sensor that detects CpG dinucleotides in RNA viruses and decreases their viral fitness (339, 349, 354, 357, 358, 386, 392, 393).

A recent publication has also reported inhibition of viral replication as a result of the introduction of CpG dinucleotides into the HIV-1 genome (360). In this study, CpGs were added into *env* and this resulted in a decrease in the gRNA abundance, Gag expression, Env expression and infectious virus production. Importantly, they showed that KO of the cellular RNA-binding protein ZAP rescues the infectivity of the viruses with a high-CpG count. Although the complete mechanism has not been defined yet, they have shown that ZAP directly binds HIV-1 RNA regions containing CpG dinucleotides. We tested whether the restriction observed upon introduction of CpG dinucleotides in *gag* was also the result of a mechanism mediated by ZAP (360). Infectivity of GagCpG22-378, however, was not rescued upon depletion of the RNA-binding protein, suggesting that the restriction we observe is independent of ZAP activity.

One more possible explanation that can explain the reduction in HIV-1 GagCpG22-378 replication might be inhibition of packaging. We considered the possibility of having inhibited an essential packaging signal as a consequence of the silent mutagenesis. HIV-1 requires both *cis*-acting elements and the *trans*-acting elements in Gag in order to achieve efficient packaging, but the specific necessary sequence has not been defined yet. It is known that the major RNA elements that are necessary for packaging are located in the 5'UTR and they overlap with the dimerisation signal (394-398). Among these elements, the four stem-loops (SL1, SL2, SL3, and SL4) greatly contribute to efficient packaging (398, 399). SL3 is the loop containing the φ signal, also called packaging signal. The minimal necessary and sufficient sequence for packaging has been found in MLV since the insertion of this sequence into heterologous RNA sequence enables packaging (400). In the HIV-1 context, deletions have been made to map the minimal sequence necessary for packaging. This study has resulted in a minimal 159 nucleotide sequence that includes SL1-SL3 through the U5:AUG stem, but does not include TAR, Poly(A) and the upper PBS hairpin (69). In addition, around 300 nucleotides in the beginning of the *gag* ORF has been shown to increase viral titre (294, 295). An approach based on probabilistic evolutionary models that quantify the synonymous mutations rate in the HIV-1 genome has shown that the first 21 nucleotides of *gag* are under purifying selection (303, 304). Moreover, NMR structures have shown interactions between that region and the U5 of the 5'UTR that enable the formation of a dimerisation-competent conformation of the gRNA that is packaged into the nascent virion (298, 401). Although some necessary elements have been mapped, the minimal sequence required for HIV-1 gRNA packaging is unknown.

We have shown that there are no essential *cis*-acting elements within nt 22-378 (GagΔ22-278, Figure 4.1), but altering the RNA sequence could modulate the local RNA structure in ways that deletion does not (402). Therefore, to test whether the introduction of CpG dinucleotides was the cause of the defect in infectious virus production or whether synonymous mutagenesis altered an important *cis*-acting element that plays a major role

in packaging, we cloned a new virus in which the codons that led to CpGs were changed to a different codon (GagDC22-378). Since infectious virus production was of GagDC22-378 was comparable to that of the wild-type HIV-1 we concluded that inhibition of viral replication is not a result of mutating an essential *cis*-acting element in this region.

7.3 Introduction of CpG dinucleotides into a longer region in *gag* inhibits HIV-1 replication in a ZAP-dependent manner

The third mechanism by which CpG dinucleotides inhibit HIV-1 replication is ZAP-dependent. Intriguingly, Gag expression and virion production of the HIV-1 mutant containing 62 CpG dinucleotides (GagCpG22-1188) was greatly rescued when ZAP was depleted from HeLa cells, suggesting that the mechanism by which ZAP restricts HIV-1 replication of high-CpG viruses might be dependent on not only the number of CpGs but also on the position that these nucleotides take in the viral genome. It is important to highlight that introduction of CpG dinucleotides into *env* (360) or and to a lesser extent in *pol* (Appendix 2) leads to a ZAP-dependent restriction of viral replication. Introduction CpGs into *gag*, however, results in a restriction that varies depending on the location of the CpGs in the viral genome. It is worth emphasising on the fact that GagCpG22-378 contains 26 CpGs and GagCpG22-759 30 CpGs. Although GagCpG22-759 has a 50% longer sequence, and the CpG dinucleotide content only differs in 4, it causes a substantially more potent inhibition of Gag expression, virion production and infectivity (Figure 6.4). These results suggest that the frequency at which the dinucleotides are present in a sequence does not play a major role. Instead, the data emphasises the importance of the position of the CpGs in the HIV-1 genome rather than the quantity. One possible explanation could rely on the secondary structure of the HIV-1 genome. We hypothesise that the local sequence surrounding the CpG dinucleotides plays an important role in inducing reduction of viral replication through different mechanisms. Whether inhibition in replication by GagCpG22-759 occurs in a ZAP-dependent or independent manner remains to be characterised.

Although we have not experimentally determined what makes some CpG dinucleotides more recognisable by ZAP than others, it is known that ZAP recognises high-CpG containing RNAs and targets them to a degradation pathway (360). Moreover, ZAP does not alter the abundance of cellular mRNAs, which suggests that specifically targets non-self viral RNAs with increased CpG abundance. Further research on characterising the degradation pathway that leads to a reduction CpG-containing viral RNA is needed.

7.4 Clinical relevance of CpG-mediated inhibition of viral replication

The evolutionary conservation of low CpG abundance in mammalian viral RNAs together with the fact that increasing the nucleotide content severely affects viral fitness suggest that CpG suppression constitutes a very important viral defence mechanism. Thus, understanding and exploiting the mechanisms by which CpG dinucleotides reduce viral replication can provide many possible applications.

There is limited evidence that demonstrates the deleterious effect of CpG dinucleotides in natural isolates. Theys et al. (345) have calculated the fitness cost of individual HIV mutations *in vivo*. They focused their research on calculating the average frequency of transition mutations at 870 sites of the *pol* gene. The study was carried out using viral samples of 160 patients and they determined that mutations that result into the addition of a CpG dinucleotide negatively affect viral fitness (345). Besides, the CpG dinucleotide content has been proposed to be a factor that can predict disease progression (344). In this study, the relative CpG abundance in the HIV-1 *env* sequences from typical-progressors and from slow-progressors were analysed and compared. They showed that HIV-1 *env* sequences from slow-progressors contain a higher number of CpG dinucleotides when compared to the sequences of the typical-progressors. Moreover, slow-

progressors exhibit this feature immediately after seroconversion. They propose that sequencing of patients at the early stage of the disease for analysis of CpG dinucleotide content presents a strong tool to predict HIV-1 disease progression (344).

One possible application is the use of CpG dinucleotides as a marker for prediction of disease progression. While long-term non-progressors (LNTPs) constitute only 1% of the HIV-1-infected patients, they have been widely used to understand the viral, host genetic and immunological factors that underlie this favourable course of the disease (403). These patients are capable of maintaining stable CD4+ T cell counts within the normal range during an extended period of time despite ongoing moderate viral replication. Elite controllers, on the other hand, maintain preserved CD4+ T cell count accompanied by low viral replication levels. These patients generally remain asymptomatic without antiretroviral treatment (ART) (403, 404). Although the mechanism underlying the non-progressive HIV-1 is most likely to be the result of a complex interplay between viral and host factors, multiple studies have identified HIV-1 genomic alterations that contribute to the clinical course of therapy-naïve individuals. In particular, point mutations, amino acid deletions, insertions or substitutions, the introduction of stop codons and polymorphisms in all structural, regulatory or accessory genes of HIV-1 have been associated with non-progressive HIV-1 disease (403). I believe that characterising the CpG content of LNTPs and elite controllers may bring insight into the mechanism underlying a favourable disease progression. In addition, our work supports the hypothesis presented by Theys et al. (345) where identification of high-CpG containing patients by deep sequencing can improve prediction of disease progression. The optimal time to start ART has been debated for many years (24). Some experts support initiation of therapy in people with high CD4+ T cell count, arguing that immediate treatment prevents irreversible harm to the immune system (405). Other groups believe that the potential toxicities associated with treatment could outweigh the benefits (406) and therapy in elite controllers, for example, is usually delayed (24). Therefore, understanding the virus-related factors that drive milder HIV-

1 disease pathogenesis is of great importance and, might help in the development of patient-targeted treatment.

Another possible application is the development of life attenuated viral vaccines. Development of a protective HIV-1 vaccine presents some limitations (404). Many details of the latent HIV-1 reservoir, for example, remain uncharacterised. In addition, clinical trials of vaccine-based strategies have been generally disappointing, with the exception of a combination HIV-1 vaccine in Thailand where 31% reduction in infection acquisition was achieved (24, 52, 404). Nevertheless, CpG dinucleotides have already been shown to inhibit viral replication of a positive strand RNA virus such as picornavirus and negative strand RNA viruses such as influenza A or HIV-1 (356-359, 378, 382, 384). Understanding the mechanisms by which introduction of CpG dinucleotides in HIV-1 reduces viral replication can be used as a model system to explore the possibility of using this as a strategy for attenuation of other RNA viruses. Large-scale synonymous mutagenesis in RNA virus genomes, therefore, can be functionally exploited to develop a novel and non-reverting safe strategy for the development of viral vaccines.

Chapter 8: Conclusions and future directions

It is almost 30 years since it was identified that CpG dinucleotides are underrepresented in the HIV-1 genome. Although the selection against CpGs in the viral genome suggests a potential deleterious effect in viral replication, the consequences of increasing their abundance in the HIV-1 genome had not been studied when we started this project. Herein, we have shown that inhibition of viral replication by increasing the CpG dinucleotide abundance is mediated by at least three different mechanisms but further work needs to be done to completely characterise these.

We demonstrate that introducing CpG dinucleotides by synonymous mutagenesis in HIV-1 GagCM22-378 inhibits viral replication through activating a cryptic splice donor that results in aberrant splicing. We hypothesise that the defect in splicing in addition to introducing an upstream AUG in each RNA transcript, can also affect the relative abundance of *rev* mRNA and in turn affect the abundance of mRNA encoding for Env. In order to further complete the research, we aim to confirm the decrease in Rev by immunoblotting the same cell lysates to quantify and compare the protein abundance in the presence and absence of the synonymous mutations. In addition, to test whether the defect in infectivity is due to a decrease in Env expression, we would also like to quantify the abundance of Env. Further work could also focus on identifying the cellular proteins that cause this effect, such as proteins that constitute the spliceosome.

Inhibition of viral replication upon introduction of CpGs in HIV-1 GagCpG22-378 happens through a different mechanism. Although we have shown that the difference in the phenotype is due to the importance of the nucleotide sequence surrounding the CpGs, further work needs to be done to specifically identify the mechanism by which viral fitness is reduced when only CpGs are introduced. A new line of work could focus on studying the post-transcriptional ribonucleoside modifications that have recently been

shown to add a layer of regulatory complexity to the RNA structure and function.

Introduction of 62 CpGs into nucleotides 22-1188 (GagCpG22-1188) produces a larger inhibition of infectious virus production than introducing 26 CpGs within nucleotides 22-378 (GagCpG22-378). Importantly, introducing 30 CpGs into nucleotides 22-654 (GagCpG22-654) also results in a large reduction of infectious virus production. These results demonstrate that the number of CpG dinucleotides is not as important as the context by which these dinucleotides are surrounded. Future work should also focus on better characterising the position-dependent inhibition of CpG dinucleotides. For this, we have started introducing CpGs in different HIV-1 genes and we have shown that they inhibit viral replication in a position-dependent manner (manuscript under preparation). While we have shown that the defect in viral replication upon addition of 62 CpGs is dependent on ZAP (GagCpG22-1188), we need to test whether adding 30 CpGs (GagCpG22-654) also inhibits viral replication in a ZAP-dependent manner.

Takata et al. (360) showed that ZAP-binding to CpG dinucleotides leads to a decreased RNA cytoplasmic abundance. Although ZAP has been identified as the protein responsible for recognising CpG-containing viral mRNAs, the proteins involved in the degradation of these RNAs has not been characterised yet. We will also focus further research on studying potential candidates that might be involved in the degradation pathway. We have carried out some preliminary work on studying proteins involved in different RNA decay pathways, by siRNA. Some of the proteins studied are involved in the 5'-3' RNA decay machinery. This degradation pathway localises to cytoplasmic P bodies, and has also been shown to inhibit viral replication of some viruses such as flavivirus or poliovirus (407-409). The RNA exosome, on the other hand, mediates 3' to 5' degradation and is localised to the cytoplasm. The complex forms an hexameric barrel that forms an internal channel that, due to its size, can enable the entry of ssRNA but not dsRNAs (409-411). The exosome has also been shown to be involved in antiviral defence. Some RNA-binding proteins, such as DDX17, have been shown to

immunoprecipitate with the exosome, suggesting it recruits and binds it to degrade the bound viral RNAs (409, 412). We have not identified any protein involved in CpG-mediated RNA degradation in this preliminary study yet. Taking all together, I believe that further work should be focused on identifying the proteins and pathways involved in the degradation of high-CpG containing viral RNA.

References

1. UNAIDS. UNAIDS. Global HIV Statistics 2018. Annual Report. 2018.
2. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harbor perspectives in medicine. 2011;1(1):a006841.
3. Barre-Sinoussi F, Ross AL, Delfraissy JF. Past, present and future: 30 years of HIV research. Nature reviews Microbiology. 2013;11(12):877-83.
4. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. The New England journal of medicine. 1981;305(24):1425-31.
5. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (New York, NY). 1983;220(4599):868-71.
6. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science (New York, NY). 1983;220(4599):865-7.
7. Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science (New York, NY). 1984;224(4648):497-500.
8. Katzourakis A, Tristem M, Pybus OG, Gifford RJ. Discovery and analysis of the first endogenous lentivirus. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(15):6261-5.
9. van der Loo W, Abrantes J, Esteves PJ. Sharing of endogenous lentiviral gene fragments among leporid lineages separated for more than 12 million years. Journal of virology. 2009;83(5):2386-8.
10. Sharp PM, Robertson DL, Hahn BH. Cross-species transmission and recombination of 'AIDS' viruses. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 1995;349(1327):41-7.

11. Jin MJ, Rogers J, Phillips-Conroy JE, Allan JS, Desrosiers RC, Shaw GM, et al. Infection of a yellow baboon with simian immunodeficiency virus from African green monkeys: evidence for cross-species transmission in the wild. Journal of virology. 1994;68(12):8454-60.
12. Peeters M, Honore C, Huet T, Bedjabaga L, Ossari S, Bussi P, et al. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. AIDS (London, England). 1989;3(10):625-30.
13. Huet T, Cheynier R, Meyerhans A, Roelants G, Wain-Hobson S. Genetic organization of a chimpanzee lentivirus related to HIV-1. Nature. 1990;345(6273):356-9.
14. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature. 1999;397(6718):436-41.
15. Corbet S, Muller-Trutwin MC, Versmisse P, Delarue S, Ayouba A, Lewis J, et al. env sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. Journal of virology. 2000;74(1):529-34.
16. Sharp PM, Shaw GM, Hahn BH. Simian immunodeficiency virus infection of chimpanzees. Journal of virology. 2005;79(7):3891-902.
17. Prince AM, Brotman B, Lee DH, Andrus L, Valinsky J, Marx P. Lack of evidence for HIV type 1-related SIVcpz infection in captive and wild chimpanzees (Pan troglodytes verus) in West Africa. AIDS research and human retroviruses. 2002;18(9):657-60.
18. Switzer WM, Parekh B, Shanmugam V, Bhullar V, Phillips S, Ely JJ, et al. The epidemiology of simian immunodeficiency virus infection in a large number of wild- and captive-born chimpanzees: evidence for a recent introduction following chimpanzee divergence. AIDS research and human retroviruses. 2005;21(5):335-42.
19. Gurtler LG, Hauser PH, Eberle J, von Brunn A, Knapp S, Zekeng L, et al. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. Journal of virology. 1994;68(3):1581-5.

20. De Leys R, Vanderborght B, Vanden Haesevelde M, Heyndrickx L, van Geel A, Wauters C, et al. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. Journal of virology. 1990;64(3):1207-16.
21. Vallari A, Bodelle P, Ngansop C, Makamche F, Ndembi N, Mbanya D, et al. Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low. AIDS research and human retroviruses. 2010;26(1):109-15.
22. Simon F, Mauclere P, Roques P, Loussert-Ajaka I, Muller-Trutwin MC, Saragosti S, et al. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nature medicine. 1998;4(9):1032-7.
23. Swanstrom R, Coffin J. HIV-1 pathogenesis: the virus. Cold Spring Harbor perspectives in medicine. 2012;2(12):a007443.
24. Deeks SG, Overbaugh J, Phillips A, Buchbinder S. HIV infection. Nature reviews Disease primers. 2015;1:15035.
25. Stieh DJ, Matias E, Xu H, Fought AJ, Blanchard JL, Marx PA, et al. Th17 Cells Are Preferentially Infected Very Early after Vaginal Transmission of SIV in Macaques. Cell host & microbe. 2016;19(4):529-40.
26. Lee B, Sharron M, Montaner LJ, Weissman D, Doms RW. Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(9):5215-20.
27. Zhu T, Muthui D, Holte S, Nickle D, Feng F, Brodie S, et al. Evidence for human immunodeficiency virus type 1 replication in vivo in CD14(+) monocytes and its potential role as a source of virus in patients on highly active antiretroviral therapy. Journal of virology. 2002;76(2):707-16.
28. Fulcher JA, Hwangbo Y, Zioni R, Nickle D, Lin X, Heath L, et al. Compartmentalization of human immunodeficiency virus type 1 between blood monocytes and CD4+ T cells during infection. Journal of virology. 2004;78(15):7883-93.
29. Nawaz F, Cicala C, Van Ryk D, Block KE, Jelicic K, McNally JP, et al. The genotype of early-transmitting HIV gp120s promotes alpha (4) beta(7)-

reactivity, revealing alpha (4) beta(7) +/CD4+ T cells as key targets in mucosal transmission. PLoS pathogens. 2011;7(2):e1001301.
30. Miller CJ, Li Q, Abel K, Kim EY, Ma ZM, Wietgrefe S, et al. Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus. Journal of virology. 2005;79(14):9217-27.
31. Stacey AR, Norris PJ, Qin L, Haygreen EA, Taylor E, Heitman J, et al. Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. Journal of virology. 2009;83(8):3719-33.
32. Kelley CF, Barbour JD, Hecht FM. The relation between symptoms, viral load, and viral load set point in primary HIV infection. J Acquir Immune Defic Syndr. 2007;45(4):445-8.
33. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, et al. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. Journal of virology. 1994;68(7):4650-5.
34. Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science (New York, NY). 1996;272(5263):872-7.
35. Doitsh G, Cavrois M, Lassen KG, Zepeda O, Yang Z, Santiago ML, et al. Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue. Cell. 2010;143(5):789-801.
36. Coffin J, Swanstrom R. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. Cold Spring Harbor perspectives in medicine. 2013;3(1):a012526.
37. Takeuchi O, Akira S. Pattern recognition receptors and inflammation. Cell. 2010;140(6):805-20.
38. Altfeld M, Gale M, Jr. Innate immunity against HIV-1 infection. Nature immunology. 2015;16(6):554-62.
39. Jakobsen MR, Bak RO, Andersen A, Berg RK, Jensen SB, Tengchuan J, et al. IFI16 senses DNA forms of the lentiviral replication cycle and controls HIV-1 replication. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(48):E4571-80.

40. Monroe KM, Yang Z, Johnson JR, Geng X, Doitsh G, Krogan NJ, et al. IFI16 DNA sensor is required for death of lymphoid CD4 T cells abortively infected with HIV. Science (New York, NY). 2014;343(6169):428-32.
41. Lahaye X, Satoh T, Gentili M, Cerboni S, Conrad C, Hurbain I, et al. The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. Immunity. 2013;39(6):1132-42.
42. Gao D, Wu J, Wu YT, Du F, Aroh C, Yan N, et al. Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. Science (New York, NY). 2013;341(6148):903-6.
43. Doyle T, Goujon C, Malim MH. HIV-1 and interferons: who's interfering with whom? Nature reviews Microbiology. 2015;13(7):403-13.
44. Lepelley A, Louis S, Sourisseau M, Law HK, Pothlichet J, Schilte C, et al. Innate sensing of HIV-infected cells. PLoS pathogens. 2011;7(2):e1001284.
45. Beignon AS, McKenna K, Skoberne M, Manches O, DaSilva I, Kavanagh DG, et al. Endocytosis of HIV-1 activates plasmacytoid dendritic cells via Toll-like receptor-viral RNA interactions. The Journal of clinical investigation. 2005;115(11):3265-75.
46. Blasius AL, Beutler B. Intracellular toll-like receptors. Immunity. 2010;32(3):305-15.
47. Li G, Cheng M, Nunoya J, Cheng L, Guo H, Yu H, et al. Plasmacytoid dendritic cells suppress HIV-1 replication but contribute to HIV-1 induced immunopathogenesis in humanized mice. PLoS pathogens. 2014;10(7):e1004291.
48. Weller S, Davis K. Condom effectiveness in reducing heterosexual HIV transmission. The Cochrane database of systematic reviews. 2002(1):Cd003255.
49. Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. PLoS medicine. 2005;2(11):e298.
50. de Vincenzi I. Triple antiretroviral compared with zidovudine and single-dose nevirapine prophylaxis during pregnancy and breastfeeding for

prevention of mother-to-child transmission of HIV-1 (Kesho Bora study): a randomised controlled trial. The Lancet Infectious diseases. 2011;11(3):171-80.
51. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. The New England journal of medicine. 2011;365(6):493-505.
52. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. The New England journal of medicine. 2009;361(23):2209-20.
53. Berkhout B, Eggink D, Sanders RW. Is there a future for antiviral fusion inhibitors? Current opinion in virology. 2012;2(1):50-9.
54. Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harbor perspectives in medicine. 2012;2(2):a006916.
55. Pereira LA, Bentley K, Peeters A, Churchill MJ, Deacon NJ. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. Nucleic acids research. 2000;28(3):663-8.
56. Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. Annual review of biochemistry. 1998;67:1-25.
57. Bolinger C, Boris-Lawrie K. Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. Retrovirology. 2009;6:8.
58. Cullen BR. Viral RNAs: lessons from the enemy. Cell. 2009;136(4):592-7.
59. HG G. HIV-1 Gag: a Molecular Machine Driving Viral Particle Assembly and Release. HIV Sequence Compendium. 2001.
60. Freed EO. HIV-1 assembly, release and maturation. Nature reviews Microbiology. 2015;13(8):484-96.
61. Cullen BR. Nuclear mRNA export: insights from virology. Trends in biochemical sciences. 2003;28(8):419-24.
62. Li L, Li HS, Pauza CD, Bukrinsky M, Zhao RY. Roles of HIV-1 auxiliary proteins in viral pathogenesis and host-pathogen interactions. Cell research. 2005;15(11-12):923-34.

63. Jun S, Ke D, Debiec K, Zhao G, Meng X, Ambrose Z, et al. Direct visualization of HIV-1 with correlative live-cell microscopy and cryo-electron tomography. Structure (London, England : 1993). 2011;19(11):1573-81.
64. Campbell EM, Hope TJ. HIV-1 capsid: the multifaceted key player in HIV-1 infection. Nature reviews Microbiology. 2015;13(8):471-83.
65. Briggs JA, Simon MN, Gross I, Krausslich HG, Fuller SD, Vogt VM, et al. The stoichiometry of Gag protein in HIV-1. Nature structural & molecular biology. 2004;11(7):672-5.
66. Welker R, Hohenberg H, Tessmer U, Huckhagel C, Krausslich HG. Biochemical and structural analysis of isolated mature cores of human immunodeficiency virus type 1. Journal of virology. 2000;74(3):1168-77.
67. Accola MA, Ohagen A, Gottlinger HG. Isolation of human immunodeficiency virus type 1 cores: retention of Vpr in the absence of p6(gag). Journal of virology. 2000;74(13):6198-202.
68. Berkhout B. Structure and function of the human immunodeficiency virus leader RNA. Progress in nucleic acid research and molecular biology. 1996;54:1-34.
69. Heng X, Kharytonchyk S, Garcia EL, Lu K, Divakaruni SS, LaCotti C, et al. Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. Journal of molecular biology. 2012;417(3):224-39.
70. Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature. 1989;338(6212):254-7.
71. Schneider R, Campbell M, Nasioulas G, Felber BK, Pavlakis GN. Inactivation of the human immunodeficiency virus type 1 inhibitory elements allows Rev-independent expression of Gag and Gag/protease and particle formation. Journal of virology. 1997;71(7):4892-903.
72. Schwartz S, Campbell M, Nasioulas G, Harrison J, Felber BK, Pavlakis GN. Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. Journal of virology. 1992;66(12):7176-82.
73. Cochrane AW, Jones KS, Beidas S, Dillon PJ, Skalka AM, Rosen CA. Identification and characterization of intragenic sequences which repress

human immunodeficiency virus structural gene expression. Journal of virology. 1991;65(10):5305-13.
74. Kypr J, Mrazek J. Unusual codon usage of HIV. Nature. 1987;327(6117):20.
75. Berkhout B, van Hemert FJ. The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins. Nucleic acids research. 1994;22(9):1705-11.
76. Sharp PM. What can AIDS virus codon usage tell us? Nature. 1986;324(6093):114.
77. Kennedy EM, Courtney DG, Tsai K, Cullen BR. Viral Epitranscriptomics. Journal of virology. 2017;91(9).
78. Li S, Mason CE. The pivotal regulatory landscape of RNA modifications. Annual review of genomics and human genetics. 2014;15:127-50.
79. Meyer KD, Jaffrey SR. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nature reviews Molecular cell biology. 2014;15(5):313-26.
80. Yue Y, Liu J, He C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Genes & development. 2015;29(13):1343-55.
81. Kennedy EM, Bogerd HP, Kornepati AV, Kang D, Ghoshal D, Marshall JB, et al. Posttranscriptional m(6)A Editing of HIV-1 mRNAs Enhances Viral Gene Expression. Cell host & microbe. 2016;19(5):675-85.
82. Lichinchi G, Gao S, Saletore Y, Gonzalez GM, Bansal V, Wang Y, et al. Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells. Nature microbiology. 2016;1:16011.
83. Hernandez LD, Hoffman LR, Wolfsberg TG, White JM. Virus-cell and cell-cell fusion. Annual review of cell and developmental biology. 1996;12:627-61.
84. Klatzmann D, Barre-Sinoussi F, Nugeyre MT, Danquet C, Vilmer E, Griscelli C, et al. Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. Science (New York, NY). 1984;225(4657):59-63.

85. Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell. 1986;47(3):333-48.
86. Allan J, Coligan J, Barin F, McLane M, Sodroski J, Rosen C, et al. Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. Science (New York, NY). 1985;228(4703):1091-4.
87. Wyatt R, Sodroski J. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. Science (New York, NY). 1998;280(5371):1884-8.
88. Doms RW, Trono D. The plasma membrane as a combat zone in the HIV battlefield. Genes & development. 2000;14(21):2677-88.
89. Geijtenbeek TB, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GC, Middel J, et al. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. Cell. 2000;100(5):587-97.
90. Wilen CB, Tilton JC, Doms RW. HIV: cell binding and entry. Cold Spring Harbor perspectives in medicine. 2012;2(8).
91. Ganser-Pornillos BK, Cheng A, Yeager M. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. Cell. 2007;131(1):70-9.
92. Lowe AR, Siegel JJ, Kalab P, Siu M, Weis K, Liphardt JT. Selectivity mechanism of the nuclear pore complex characterized by single cargo tracking. Nature. 2010;467(7315):600-3.
93. Pante N, Kann M. Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. Molecular biology of the cell. 2002;13(2):425-34.
94. Hulme AE, Kelley Z, Foley D, Hope TJ. Complementary Assays Reveal a Low Level of CA Associated with Viral Complexes in the Nuclei of HIV-1-Infected Cells. Journal of virology. 2015;89(10):5350-61.
95. Peng K, Muranyi W, Glass B, Laketa V, Yant SR, Tsai L, et al. Quantitative microscopy of functional HIV post-entry complexes reveals association of replication with the viral capsid. eLife. 2014;3:e04114.
96. Hulme AE, Perez O, Hope TJ. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(24):9975-80.

97. Perez-Caballero D, Hatziioannou T, Zhang F, Cowan S, Bieniasz PD. Restriction of human immunodeficiency virus type 1 by TRIM-CypA occurs with rapid kinetics and independently of cytoplasmic bodies, ubiquitin, and proteasome activity. Journal of virology. 2005;79(24):15567-72.
98. Arhel NJ, Souquere-Besse S, Munier S, Souque P, Guadagnini S, Rutherford S, et al. HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. The EMBO journal. 2007;26(12):3025-37.
99. Francis AC, Melikyan GB. Single HIV-1 Imaging Reveals Progression of Infection through CA-Dependent Steps of Docking at the Nuclear Pore, Uncoating, and Nuclear Transport. Cell host & microbe. 2018;23(4):536-48.e6.
100. Hu WS, Hughes SH. HIV-1 reverse transcription. Cold Spring Harbor perspectives in medicine. 2012;2(10).
101. Jacobo-Molina A, Ding J, Nanni RG, Clark AD, Jr., Lu X, Tantillo C, et al. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(13):6320-4.
102. Hilditch L, Towers GJ. A model for cofactor use during HIV-1 reverse transcription and nuclear entry. Current opinion in virology. 2014;4:32-6.
103. Hungnes O, Tjotta E, Grinde B. Mutations in the central polypurine tract of HIV-1 result in delayed replication. Virology. 1992;190(1):440-2.
104. Charneau P, Alizon M, Clavel F. A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. Journal of virology. 1992;66(5):2814-20.
105. Yamashita M, Emerman M. Retroviral infection of non-dividing cells: old and new perspectives. Virology. 2006;344(1):88-93.
106. Labokha AA, Fassati A. Viruses challenge selectivity barrier of nuclear pores. Viruses. 2013;5(10):2410-23.
107. Strambio-De-Castillia C, Niepel M, Rout MP. The nuclear pore complex: bridging nuclear transport and gene regulation. Nature reviews Molecular cell biology. 2010;11(7):490-501.

108. Yamashita M, Perez O, Hope TJ, Emerman M. Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. PLoS pathogens. 2007;3(10):1502-10.
109. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, et al. Identification of host proteins required for HIV infection through a functional genomic screen. Science (New York, NY). 2008;319(5865):921-6.
110. Konig R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Irelan JT, et al. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell. 2008;135(1):49-60.
111. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, et al. Genome-scale RNAi screen for host factors required for HIV replication. Cell host & microbe. 2008;4(5):495-504.
112. Bhattacharya A, Alam SL, Fricke T, Zadrozny K, Sedzicki J, Taylor AB, et al. Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(52):18625-30.
113. Lusic M, Siliciano RF. Nuclear landscape of HIV-1 infection and integration. Nature reviews Microbiology. 2017;15(2):69-82.
114. Bushman FD, Malani N, Fernandes J, D'Orso I, Cagney G, Diamond TL, et al. Host cell factors in HIV replication: meta-analysis of genome-wide studies. PLoS pathogens. 2009;5(5):e1000437.
115. Di Nunzio F, Danckaert A, Fricke T, Perez P, Fernandez J, Perret E, et al. Human nucleoporins promote HIV-1 docking at the nuclear pore, nuclear import and integration. PloS one. 2012;7(9):e46037.
116. Meehan AM, Saenz DT, Guevera R, Morrison JH, Peretz M, Fadel HJ, et al. A cyclophilin homology domain-independent role for Nup358 in HIV-1 infection. PLoS pathogens. 2014;10(2):e1003969.
117. Matreyek KA, Engelman A. The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid. Journal of virology. 2011;85(15):7818-27.
118. Di Nunzio F, Fricke T, Miccio A, Valle-Casuso JC, Perez P, Souque P, et al. Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. Virology. 2013;440(1):8-18.

119. De Iaco A, Santoni F, Vannier A, Guipponi M, Antonarakis S, Luban J. TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. Retrovirology. 2013;10:20.
120. Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, Li X, et al. The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. Journal of virology. 2010;84(1):397-406.
121. Lee K, Ambrose Z, Martin TD, Oztop I, Mulky A, Julias JG, et al. Flexible use of nuclear import pathways by HIV-1. Cell host & microbe. 2010;7(3):221-33.
122. Lee K, Mulky A, Yuen W, Martin TD, Meyerson NR, Choi L, et al. HIV-1 capsid-targeting domain of cleavage and polyadenylation specificity factor 6. Journal of virology. 2012;86(7):3851-60.
123. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. Cell. 2002;110(4):521-9.
124. Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. Science (New York, NY). 2003;300(5626):1749-51.
125. Melamed A, Laydon DJ, Gillet NA, Tanaka Y, Taylor GP, Bangham CR. Genome-wide determinants of proviral targeting, clonal abundance and expression in natural HTLV-1 infection. PLoS pathogens. 2013;9(3):e1003271.
126. Singh PK, Plumb MR, Ferris AL, Iben JR, Wu X, Fadel HJ, et al. LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. Genes & development. 2015;29(21):2287-97.
127. Sowd GA, Serrao E, Wang H, Wang W, Fadel HJ, Poeschla EM, et al. A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(8):E1054-63.

128. Wu X, Li Y, Crise B, Burgess SM, Munroe DJ. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. Journal of virology. 2005;79(8):5211-4.
129. Holman AG, Coffin JM. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(17):6103-7.
130. Grandgenett DP. Symmetrical recognition of cellular DNA target sequences during retroviral integration. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(17):5903-4.
131. Kirk PD, Huvet M, Melamed A, Maertens GN, Bangham CR. Retroviruses integrate into a shared, non-palindromic DNA motif. Nature microbiology. 2016;2:16212.
132. Ballandras-Colas A, Maskell DP, Serrao E, Locke J, Swuec P, Jonsson SR, et al. A supramolecular assembly mediates lentiviral DNA integration. Science (New York, NY). 2017;355(6320):93-5.
133. Engelman A, Cherepanov P. The structural biology of HIV-1: mechanistic and therapeutic insights. Nature reviews Microbiology. 2012;10(4):279-90.
134. Engelman A, Cherepanov P. Retroviral Integrase Structure and DNA Recombination Mechanism. Microbiology spectrum. 2014;2(6).
135. Jones KA, Kadonaga JT, Luciw PA, Tjian R. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science (New York, NY). 1986;232(4751):755-9.
136. Garcia JA, Harrich D, Soultanakis E, Wu F, Mitsuyasu R, Gaynor RB. Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. The EMBO journal. 1989;8(3):765-78.
137. Zenzie-Gregory B, Sheridan P, Jones KA, Smale ST. HIV-1 core promoter lacks a simple initiator element but contains a bipartite activator at the transcription start site. The Journal of biological chemistry. 1993;268(21):15823-32.

138. Sodroski J, Patarca R, Rosen C, Wong-Staal F, Haseltine W. Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III. Science (New York, NY). 1985;229(4708):74-7.
139. Sodroski J, Rosen C, Wong-Staal F, Salahuddin SZ, Popovic M, Arya S, et al. Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. Science (New York, NY). 1985;227(4683):171-3.
140. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell. 1998;92(4):451-62.
141. Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Molecular and cellular biology. 2004;24(2):787-95.
142. Barboric M, Peterlin BM. A new paradigm in eukaryotic biology: HIV Tat and the control of transcriptional elongation. PLoS biology. 2005;3(2):e76.
143. Kao SY, Calman AF, Luciw PA, Peterlin BM. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature. 1987;330(6147):489-93.
144. Feng S, Holland EC. HIV-1 tat trans-activation requires the loop sequence within tar. Nature. 1988;334(6178):165-7.
145. Muesing MA, Smith DH, Cabradilla CD, Benton CV, Lasky LA, Capon DJ. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. Nature. 1985;313(6002):450-8.
146. Kim SY, Byrn R, Groopman J, Baltimore D. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. Journal of virology. 1989;63(9):3708-13.
147. Robert-Guroff M, Popovic M, Gartner S, Markham P, Gallo RC, Reitz MS. Structure and expression of tat-, rev-, and nef-specific transcripts of human immunodeficiency virus type 1 in infected lymphocytes and macrophages. Journal of virology. 1990;64(7):3391-8.

148. Schwartz S, Felber BK, Benko DM, Fenyo EM, Pavlakis GN. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. Journal of virology. 1990;64(6):2519-29.
149. Purcell DF, Martin MA. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. Journal of virology. 1993;67(11):6365-78.
150. Stoltzfus CM, Madsen JM. Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. Current HIV research. 2006;4(1):43-55.
151. Stoltzfus CM. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. Advances in virus research. 2009;74:1-40.
152. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nature reviews Molecular cell biology. 2009;10(11):741-54.
153. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA (New York, NY). 2008;14(5):802-13.
154. Abovich N, Legrain P, Rosbash M. The yeast PRP6 gene encodes a U4/U6 small nuclear ribonucleoprotein particle (snRNP) protein, and the PRP9 gene encodes a protein required for U2 snRNP binding. Molecular and cellular biology. 1990;10(12):6417-25.
155. Meyer BE, Malim MH. The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. Genes & development. 1994;8(13):1538-47.
156. Rizvi TA, Schmidt RD, Lew KA. Mason-Pfizer monkey virus (MPMV) constitutive transport element (CTE) functions in a position-dependent manner. Virology. 1997;236(1):118-29.
157. Swanson CM, Puffer BA, Ahmad KM, Doms RW, Malim MH. Retroviral mRNA nuclear export elements regulate protein function and virion assembly. The EMBO journal. 2004;23(13):2632-40.
158. Martinez-Salas E, Pacheco A, Serrano P, Fernandez N. New insights into internal ribosome entry site elements relevant for viral gene expression. The Journal of general virology. 2008;89(Pt 3):611-26.

159. Hidalgo L, Swanson CM. Regulation of human immunodeficiency virus type 1 (HIV-1) mRNA translation. Biochemical Society transactions. 2017;45(2):353-64.
160. Kozak M. The scanning model for translation: an update. The Journal of cell biology. 1989;108(2):229-41.
161. Bolinger C, Sharma A, Singh D, Yu L, Boris-Lawrie K. RNA helicase A modulates translation of HIV-1 and infectivity of progeny virions. Nucleic acids research. 2010;38(5):1686-96.
162. Hartman TR, Qian S, Bolinger C, Fernandez S, Schoenberg DR, Boris-Lawrie K. RNA helicase A is necessary for translation of selected messenger RNAs. Nature structural & molecular biology. 2006;13(6):509-16.
163. Soto-Rifo R, Rubilar PS, Ohlmann T. The DEAD-box helicase DDX3 substitutes for the cap-binding protein eIF4E to promote compartmentalized translation initiation of the HIV-1 genomic RNA. Nucleic acids research. 2013;41(12):6286-99.
164. Soto-Rifo R, Rubilar PS, Limousin T, de Breyne S, Decimo D, Ohlmann T. DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. The EMBO journal. 2012;31(18):3745-56.
165. Fitzgerald KD, Semler BL. Bridging IRES elements in mRNAs to the eukaryotic translation apparatus. Biochimica et biophysica acta. 2009;1789(9-10):518-28.
166. Merrick WC. Cap-dependent and cap-independent translation in eukaryotic systems. Gene. 2004;332:1-11.
167. Plank TD, Whitehurst JT, Kieft JS. Cell type specificity and structural determinants of IRES activity from the 5' leaders of different HIV-1 transcripts. Nucleic acids research. 2013;41(13):6698-714.
168. Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF. The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. Journal of virology. 2001;75(1):181-91.
169. Locker N, Chamond N, Sargueil B. A conserved structure within the HIV gag open reading frame that controls translation initiation directly recruits the 40S subunit and eIF3. Nucleic acids research. 2011;39(6):2367-77.

170. Sundquist WI, Krausslich HG. HIV-1 assembly, budding, and maturation. Cold Spring Harbor perspectives in medicine. 2012;2(7):a006924.
171. Bieniasz PD. The cell biology of HIV-1 virion genesis. Cell host & microbe. 2009;5(6):550-8.
172. Saad JS, Miller J, Tai J, Kim A, Ghanam RH, Summers MF. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(30):11364-9.
173. Martin-Serrano J, Zang T, Bieniasz PD. HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. Nature medicine. 2001;7(12):1313-9.
174. Pires R, Hartlieb B, Signor L, Schoehn G, Lata S, Roessle M, et al. A crescent-shaped ALIX dimer targets ESCRT-III CHMP4 filaments. Structure (London, England : 1993). 2009;17(6):843-56.
175. Usami Y, Popov S, Popova E, Inoue M, Weissenhorn W, H GG. The ESCRT pathway and HIV-1 budding. Biochemical Society transactions. 2009;37(Pt 1):181-4.
176. McCullough J, Fisher RD, Whitby FG, Sundquist WI, Hill CP. ALIX-CHMP4 interactions in the human ESCRT pathway. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(22):7687-91.
177. Hanson PI, Roth R, Lin Y, Heuser JE. Plasma membrane deformation by circular arrays of ESCRT-III protein filaments. The Journal of cell biology. 2008;180(2):389-402.
178. Hurley JH, Hanson PI. Membrane budding and scission by the ESCRT machinery: it's all in the neck. Nature reviews Molecular cell biology. 2010;11(8):556-66.
179. Peel S, Macheboeuf P, Martinelli N, Weissenhorn W. Divergent pathways lead to ESCRT-III-catalyzed membrane fission. Trends in biochemical sciences. 2011;36(4):199-210.
180. Lata S, Schoehn G, Jain A, Pires R, Piehler J, Gottlinger HG, et al. Helical structures of ESCRT-III are disassembled by VPS4. Science (New York, NY). 2008;321(5894):1354-7.

181. Obita T, Saksena S, Ghazi-Tabatabai S, Gill DJ, Perisic O, Emr SD, et al. Structural basis for selective recognition of ESCRT-III by the AAA ATPase Vps4. Nature. 2007;449(7163):735-9.
182. Stuchell-Brereton MD, Skalicky JJ, Kieffer C, Karren MA, Ghaffarian S, Sundquist WI. ESCRT-III recognition by VPS4 ATPases. Nature. 2007;449(7163):740-4.
183. Hurley JH, Yang D. MIT domainia. Developmental cell. 2008;14(1):6-8.
184. Murakami T, Ablan S, Freed EO, Tanaka Y. Regulation of human immunodeficiency virus type 1 Env-mediated membrane fusion by viral protease activity. Journal of virology. 2004;78(2):1026-31.
185. Malim MH, Bieniasz PD. HIV Restriction Factors and Mechanisms of Evasion. Cold Spring Harbor perspectives in medicine. 2012;2(5):a006940.
186. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature. 2002;418(6898):646-50.
187. Simon JH, Gaddis NC, Fouchier RA, Malim MH. Evidence for a newly discovered cellular anti-HIV-1 phenotype. Nature medicine. 1998;4(12):1397-400.
188. Sheehy AM, Gaddis NC, Malim MH. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nature medicine. 2003;9(11):1404-7.
189. Chen KM, Harjes E, Gross PJ, Fahmy A, Lu Y, Shindo K, et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature. 2008;452(7183):116-9.
190. Yu Q, Konig R, Pillai S, Chiles K, Kearney M, Palmer S, et al. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nature structural & molecular biology. 2004;11(5):435-42.
191. Bishop KN, Verma M, Kim EY, Wolinsky SM, Malim MH. APOBEC3G inhibits elongation of HIV-1 reverse transcripts. PLoS pathogens. 2008;4(12):e1000231.
192. Iwatani Y, Chan DS, Wang F, Maynard KS, Sugiura W, Gronenborn AM, et al. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic acids research. 2007;35(21):7096-108.

193. Pollpeter D, Parsons M, Sobala AE, Coxhead S, Lang RD, Bruns AM, et al. Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. Nature microbiology. 2018;3(2):220-33.
194. Marin M, Rose KM, Kozak SL, Kabat D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nature medicine. 2003;9(11):1398-403.
195. Bergeron JR, Huthoff H, Veselkov DA, Beavil RL, Simpson PJ, Matthews SJ, et al. The SOCS-box of HIV-1 Vif interacts with ElonginBC by induced-folding to recruit its Cul5-containing ubiquitin ligase complex. PLoS pathogens. 2010;6(6):e1000925.
196. Siu KK, Sultana A, Azimi FC, Lee JE. Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. Nature communications. 2013;4:2593.
197. Guo Y, Dong L, Qiu X, Wang Y, Zhang B, Liu H, et al. Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. Nature. 2014;505(7482):229-33.
198. Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, et al. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science (New York, NY). 2003;302(5647):1056-60.
199. Dang Y, Wang X, Esselman WJ, Zheng YH. Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. Journal of virology. 2006;80(21):10522-33.
200. Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Current biology : CB. 2004;14(15):1392-6.
201. Nisole S, Stoye JP, Saib A. TRIM family proteins: retroviral restriction and antiviral defence. Nature reviews Microbiology. 2005;3(10):799-808.
202. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature. 2004;427(6977):848-53.
203. Shibata R, Sakai H, Kawamura M, Tokunaga K, Adachi A. Early replication block of human immunodeficiency virus type 1 in monkey cells. The Journal of general virology. 1995;76 (Pt 11):2723-30.

204. Hatziioannou T, Princiotta M, Piatak M, Jr., Yuan F, Zhang F, Lifson JD, et al. Generation of simian-tropic HIV-1 by restriction factor evasion. Science (New York, NY). 2006;314(5796):95.
205. Owens CM, Yang PC, Gottlinger H, Sodroski J. Human and simian immunodeficiency virus capsid proteins are major viral determinants of early, postentry replication blocks in simian cells. Journal of virology. 2003;77(1):726-31.
206. Stremlau M, Perron M, Lee M, Li Y, Song B, Javanbakht H, et al. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(14):5514-9.
207. Ganser-Pornillos BK, Chandrasekaran V, Pornillos O, Sodroski JG, Sundquist WI, Yeager M. Hexagonal assembly of a restricting TRIM5alpha protein. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(2):534-9.
208. Wagner JM, Christensen DE, Bhattacharya A, Dawidziak DM, Roganowicz MD, Wan Y, et al. General Model for Retroviral Capsid Pattern Recognition by TRIM5 Proteins. Journal of virology. 2018;92(4).
209. Berthoux L, Sebastian S, Sokolskaja E, Luban J. Cyclophilin A is required for TRIM5{alpha}-mediated resistance to HIV-1 in Old World monkey cells. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(41):14849-53.
210. Stremlau M, Song B, Javanbakht H, Perron M, Sodroski J. Cyclophilin A: an auxiliary but not necessary cofactor for TRIM5alpha restriction of HIV-1. Virology. 2006;351(1):112-20.
211. Pertel T, Hausmann S, Morger D, Zuger S, Guerra J, Lascano J, et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. Nature. 2011;472(7343):361-5.
212. Strebel K, Klimkait T, Martin MA. A novel gene of HIV-1, vpu, and its 16-kilodalton product. Science (New York, NY). 1988;241(4870):1221-3.
213. Neil SJ, Eastman SW, Jouvenet N, Bieniasz PD. HIV-1 Vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane. PLoS pathogens. 2006;2(5):e39.

214. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature. 2008;451(7177):425-30.
215. Bartee E, McCormack A, Fruh K. Quantitative membrane proteomics reveals new cellular targets of viral immune modulators. PLoS pathogens. 2006;2(10):e107.
216. Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC, et al. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. Cell host & microbe. 2008;3(4):245-52.
217. Galao RP, Le Tortorec A, Pickering S, Kueck T, Neil SJ. Innate sensing of HIV-1 assembly by Tetherin induces NFkappaB-dependent proinflammatory responses. Cell host & microbe. 2012;12(5):633-44.
218. Galao RP, Pickering S, Curnock R, Neil SJ. Retroviral retention activates a Syk-dependent HemITAM in human tetherin. Cell host & microbe. 2014;16(3):291-303.
219. Bour S, Schubert U, Strebel K. The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implications for the mechanism of degradation. Journal of virology. 1995;69(3):1510-20.
220. Margottin F, Bour SP, Durand H, Selig L, Benichou S, Richard V, et al. A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. Molecular cell. 1998;1(4):565-74.
221. West CM. Evolutionary and functional implications of the complex glycosylation of Skp1, a cytoplasmic/nuclear glycoprotein associated with polyubiquitination. Cellular and molecular life sciences : CMLS. 2003;60(2):229-40.
222. Apps R, Del Prete GQ, Chatterjee P, Lara A, Brumme ZL, Brockman MA, et al. HIV-1 Vpu Mediates HLA-C Downregulation. Cell host & microbe. 2016;19(5):686-95.
223. Moll M, Andersson SK, Smed-Sorensen A, Sandberg JK. Inhibition of lipid antigen presentation in dendritic cells by HIV-1 Vpu interference with CD1d recycling from endosomal compartments. Blood. 2010;116(11):1876-84.

224. Ramirez PW, Famiglietti M, Sowrirajan B, DePaula-Silva AB, Rodesch C, Barker E, et al. Downmodulation of CCR7 by HIV-1 Vpu results in impaired migration and chemotactic signaling within CD4(+) T cells. Cell reports. 2014;7(6):2019-30.
225. Usami Y, Wu Y, Gottlinger HG. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. Nature. 2015;526(7572):218-23.
226. Rosa A, Chande A, Ziglio S, De Sanctis V, Bertorelli R, Goh SL, et al. HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. Nature. 2015;526(7572):212-7.
227. Beitari S, Ding S, Pan Q, Finzi A, Liang C. Effect of HIV-1 Env on SERINC5 Antagonism. Journal of virology. 2017;91(4).
228. Sood C, Marin M, Chande A, Pizzato M, Melikyan GB. SERINC5 protein inhibits HIV-1 fusion pore formation by promoting functional inactivation of envelope glycoproteins. The Journal of biological chemistry. 2017;292(14):6014-26.
229. Swigut T, Shohdy N, Skowronski J. Mechanism for down-regulation of CD28 by Nef. The EMBO journal. 2001;20(7):1593-604.
230. Schwartz O, Marechal V, Le Gall S, Lemonnier F, Heard JM. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. Nature medicine. 1996;2(3):338-42.
231. Janvier K, Craig H, Hitchin D, Madrid R, Sol-Foulon N, Renault L, et al. HIV-1 Nef stabilizes the association of adaptor protein complexes with membranes. The Journal of biological chemistry. 2003;278(10):8725-32.
232. Piguet V, Wan L, Borel C, Mangasarian A, Demaurex N, Thomas G, et al. HIV-1 Nef protein binds to the cellular protein PACS-1 to downregulate class I major histocompatibility complexes. Nature cell biology. 2000;2(3):163-7.
233. Chowers MY, Spina CA, Kwoh TJ, Fitch NJ, Richman DD, Guatelli JC. Optimal infectivity in vitro of human immunodeficiency virus type 1 requires an intact nef gene. Journal of virology. 1994;68(5):2906-14.
234. Campbell EM, Nunez R, Hope TJ. Disruption of the actin cytoskeleton can complement the ability of Nef to enhance human immunodeficiency virus type 1 infectivity. Journal of virology. 2004;78(11):5745-55.

235. Campbell SM, Crowe SM, Mak J. Virion-associated cholesterol is critical for the maintenance of HIV-1 structure and infectivity. AIDS (London, England). 2002;16(17):2253-61.
236. Zheng YH, Plemenitas A, Fielding CJ, Peterlin BM. Nef increases the synthesis of and transports cholesterol to lipid rafts and HIV-1 progeny virions. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(14):8460-5.
237. Cameron PU, Forsum U, Teppler H, Granelli-Piperno A, Steinman RM. During HIV-1 infection most blood dendritic cells are not productively infected and can induce allogeneic CD4+ T cells clonal expansion. Clinical and experimental immunology. 1992;88(2):226-36.
238. Wu L, KewalRamani VN. Dendritic-cell interactions with HIV: infection and viral dissemination. Nature reviews Immunology. 2006;6(11):859-68.
239. Goujon C, Riviere L, Jarrosson-Wuilleme L, Bernaud J, Rigal D, Darlix JL, et al. SIVSM/HIV-2 Vpx proteins promote retroviral escape from a proteasome-dependent restriction pathway present in human dendritic cells. Retrovirology. 2007;4:2.
240. Srivastava S, Swanson SK, Manel N, Florens L, Washburn MP, Skowronski J. Lentiviral Vpx accessory factor targets VprBP/DCAF1 substrate adaptor for cullin 4 E3 ubiquitin ligase to enable macrophage infection. PLoS pathogens. 2008;4(5):e1000059.
241. Simon V, Bloch N, Landau NR. Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. Nature immunology. 2015;16(6):546-53.
242. Goujon C, Arfi V, Pertel T, Luban J, Lienard J, Rigal D, et al. Characterization of simian immunodeficiency virus SIVSM/human immunodeficiency virus type 2 Vpx function in human myeloid cells. Journal of virology. 2008;82(24):12335-45.
243. Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Segeral E, et al. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature. 2011;474(7353):654-7.
244. Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, et al. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature. 2011;474(7353):658-61.

245. Goldstone DC, Ennis-Adeniran V, Hedden JJ, Groom HC, Rice GI, Christodoulou E, et al. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. Nature. 2011;480(7377):379-82.
246. Lahouassa H, Daddacha W, Hofmann H, Ayinde D, Logue EC, Dragin L, et al. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. Nature immunology. 2012;13(3):223-8.
247. White TE, Brandariz-Nunez A, Valle-Casuso JC, Amie S, Nguyen LA, Kim B, et al. The retroviral restriction ability of SAMHD1, but not its deoxynucleotide triphosphohydrolase activity, is regulated by phosphorylation. Cell host & microbe. 2013;13(4):441-51.
248. Rice GI, Bond J, Asipu A, Brunette RL, Manfield IW, Carr IM, et al. Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response. Nature genetics. 2009;41(7):829-32.
249. Crow YJ, Chase DS, Lowenstein Schmidt J, Szynkiewicz M, Forte GM, Gornall HL, et al. Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. American journal of medical genetics Part A. 2015;167a(2):296-312.
250. Yan N, Regalado-Magdos AD, Stiggelbout B, Lee-Kirsch MA, Lieberman J. The cytosolic exonuclease TREX1 inhibits the innate immune response to human immunodeficiency virus type 1. Nature immunology. 2010;11(11):1005-13.
251. Cribier A, Descours B, Valadao AL, Laguette N, Benkirane M. Phosphorylation of SAMHD1 by cyclin A2/CDK1 regulates its restriction activity toward HIV-1. Cell reports. 2013;3(4):1036-43.
252. Goujon C, Moncorge O, Bauby H, Doyle T, Ward CC, Schaller T, et al. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. Nature. 2013;502(7472):559-62.
253. Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, Wilson SJ, et al. MX2 is an interferon-induced inhibitor of HIV-1 infection. Nature. 2013;502(7472):563-6.

254. Melen K, Keskinen P, Ronni T, Sareneva T, Lounatmaa K, Julkunen I. Human MxB protein, an interferon-alpha-inducible GTPase, contains a nuclear targeting signal and is localized in the heterochromatin region beneath the nuclear envelope. The Journal of biological chemistry. 1996;271(38):23478-86.
255. Asmuth DM, Murphy RL, Rosenkranz SL, Lertora JJ, Kottilil S, Cramer Y, et al. Safety, tolerability, and mechanisms of antiretroviral activity of pegylated interferon Alfa-2a in HIV-1-monoinfected participants: a phase II clinical trial. The Journal of infectious diseases. 2010;201(11):1686-96.
256. Bitzegeio J, Sampias M, Bieniasz PD, Hatziioannou T. Adaptation to the interferon-induced antiviral state by human and simian immunodeficiency viruses. Journal of virology. 2013;87(6):3549-60.
257. Goujon C, Malim MH. Characterization of the alpha interferon-induced postentry block to HIV-1 infection in primary human macrophages and T cells. Journal of virology. 2010;84(18):9254-66.
258. Meylan PR, Guatelli JC, Munis JR, Richman DD, Kornbluth RS. Mechanisms for the inhibition of HIV replication by interferons-alpha, -beta, and -gamma in primary human macrophages. Virology. 1993;193(1):138-48.
259. Pillai SK, Abdel-Mohsen M, Guatelli J, Skasko M, Monto A, Fujimoto K, et al. Role of retroviral restriction factors in the interferon-alpha-mediated suppression of HIV-1 in vivo. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(8):3035-40.
260. Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, et al. The interferon-inducible MxB protein inhibits HIV-1 infection. Cell host & microbe. 2013;14(4):398-410.
261. Dicks MD, Goujon C, Pollpeter D, Betancor G, Apolonia L, Bergeron JR, et al. Oligomerization Requirements for MX2-Mediated Suppression of HIV-1 Infection. Journal of virology. 2016;90(1):22-32.
262. Busnadiego I, Kane M, Rihn SJ, Preugschas HF, Hughes J, Blanco-Melo D, et al. Host and viral determinants of Mx2 antiretroviral activity. Journal of virology. 2014;88(14):7738-52.
263. Fackler OT, Keppler OT. MxB/Mx2: the latest piece in HIV's interferon puzzle. EMBO reports. 2013;14(12):1028-9.

264. Foster TL, Pickering S, Neil SJD. Inhibiting the Ins and Outs of HIV Replication: Cell-Intrinsic Antiretroviral Restrictions at the Plasma Membrane. Frontiers in immunology. 2017;8:1853.
265. Schreiber G. The molecular basis for differential type I interferon signaling. The Journal of biological chemistry. 2017;292(18):7285-94.
266. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. Nature. 2011;472(7344):481-5.
267. Brass AL, Huang IC, Benita Y, John SP, Krishnan MN, Feeley EM, et al. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. Cell. 2009;139(7):1243-54.
268. Huang IC, Bailey CC, Weyer JL, Radoshitzky SR, Becker MM, Chiang JJ, et al. Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. PLoS pathogens. 2011;7(1):e1001258.
269. Lu J, Pan Q, Rong L, He W, Liu SL, Liang C. The IFITM proteins inhibit HIV-1 infection. Journal of virology. 2011;85(5):2126-37.
270. Weston S, Czieso S, White IJ, Smith SE, Kellam P, Marsh M. A membrane topology model for human interferon inducible transmembrane protein 1. PloS one. 2014;9(8):e104341.
271. Compton AA, Bruel T, Porrot F, Mallet A, Sachse M, Euvrard M, et al. IFITM proteins incorporated into HIV-1 virions impair viral fusion and spread. Cell host & microbe. 2014;16(6):736-47.
272. Foster TL, Wilson H, Iyer SS, Coss K, Doores K, Smith S, et al. Resistance of Transmitted Founder HIV-1 to IFITM-Mediated Restriction. Cell host & microbe. 2016;20(4):429-42.
273. Fenton-May AE, Dibben O, Emmerich T, Ding H, Pfafferott K, Aasa-Chapman MM, et al. Relative resistance of HIV-1 founder viruses to control by interferon-alpha. Retrovirology. 2013;10:146.
274. Bar KJ, Tsao CY, Iyer SS, Decker JM, Yang Y, Bonsignori M, et al. Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. PLoS pathogens. 2012;8(5):e1002721.
275. Rihn SJ, Foster TL, Busnadiego I, Aziz MA, Hughes J, Neil SJ, et al. The Envelope Gene of Transmitted HIV-1 Resists a Late Interferon Gamma-Induced Block. Journal of virology. 2017;91(7).

276. Kondo E, Mammano F, Cohen EA, Gottlinger HG. The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. Journal of virology. 1995;69(5):2759-64.
277. Guenzel CA, Herate C, Benichou S. HIV-1 Vpr-a still "enigmatic multitasker". Frontiers in microbiology. 2014;5:127.
278. Goh WC, Rogel ME, Kinsey CM, Michael SF, Fultz PN, Nowak MA, et al. HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo. Nature medicine. 1998;4(1):65-71.
279. Belzile JP, Richard J, Rougeau N, Xiao Y, Cohen EA. HIV-1 Vpr induces the K48-linked polyubiquitination and proteasomal degradation of target cellular proteins to activate ATR and promote G2 arrest. Journal of virology. 2010;84(7):3320-30.
280. Berger G, Lawrence M, Hue S, Neil SJ. G2/M cell cycle arrest correlates with primate lentiviral Vpr interaction with the SLX4 complex. Journal of virology. 2015;89(1):230-40.
281. Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, et al. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. Journal of virology. 1986;59(2):284-91.
282. Lee JS, Gray J, Mulligan R. Packaging cells comprising codon-optimized gagpol sequences and lacking lentiviral accessory proteins. Google Patents; 2006.
283. Swanson CM, Sherer NM, Malim MH. SRp40 and SRp55 promote the translation of unspliced human immunodeficiency virus type 1 RNA. J Virol. 2010;84(13):6748-59.
284. Fouchier RA, Meyer BE, Simon JH, Fischer U, Malim MH. HIV-1 infection of non-dividing cells: evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import. The EMBO journal. 1997;16(15):4531-9.
285. Reil H, Bukovsky AA, Gelderblom HR, Gottlinger HG. Efficient HIV-1 replication can occur in the absence of the viral matrix protein. The EMBO journal. 1998;17(9):2699-708.
286. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome research. 2004;14(6):1188-90.

287. Derdeyn CA, Decker JM, Sfakianos JN, Wu X, O'Brien WA, Ratner L, et al. Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. Journal of virology. 2000;74(18):8358-67.
288. Wei X, Decker JM, Liu H, Zhang Z, Arani RB, Kilby JM, et al. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. Antimicrobial agents and chemotherapy. 2002;46(6):1896-905.
289. Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. Journal of virology. 1998;72(4):2855-64.
290. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nature methods. 2014;11(8):783-4.
291. Chesebro B, Wehrly K, Nishio J, Perryman S. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. J Virol. 1992;66(11):6547-54.
292. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nature methods. 2015;12(4):357-60.
293. Hahne F, Ivanek R. Visualizing Genomic Data Using Gviz and Bioconductor. Methods in molecular biology (Clifton, NJ). 2016;1418:335-51.
294. Kuzembayeva M, Dilley K, Sardo L, Hu WS. Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. Virology. 2014;454-455:362-70.
295. Lu K, Heng X, Summers MF. Structural determinants and mechanism of HIV-1 genome packaging. Journal of molecular biology. 2011;410(4):609-33.
296. Leblanc J, Weil J, Beemon K. Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. Wiley interdisciplinary reviews RNA. 2013;4(5):567-80.

297. Rojas-Araya B, Ohlmann T, Soto-Rifo R. Translational Control of the HIV Unspliced Genomic RNA. Viruses. 2015;7(8):4326-51.
298. Keane SC, Summers MF. NMR Studies of the Structure and Function of the HIV-1 5'-Leader. Viruses. 2016;8(12).
299. Brierley I, Dos Ramos FJ. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. Virus research. 2006;119(1):29-42.
300. Tazi J, Bakkour N, Marchand V, Ayadi L, Aboufirassi A, Branlant C. Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. The FEBS journal. 2010;277(4):867-76.
301. Pollard VW, Malim MH. The HIV-1 Rev protein. Annual review of microbiology. 1998;52:491-532.
302. Le Grice SF. Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise. The Journal of biological chemistry. 2012;287(49):40850-7.
303. Mayrose I, Stern A, Burdelova EO, Sabo Y, Laham-Karam N, Zamostiano R, et al. Synonymous site conservation in the HIV-1 genome. BMC evolutionary biology. 2013;13:164.
304. Ngandu NK, Scheffler K, Moore P, Woodman Z, Martin D, Seoighe C. Extensive purifying selection acting on synonymous sites in HIV-1 Group M sequences. Virology journal. 2008;5:160.
305. Ooms M, Abbink TE, Pham C, Berkhout B. Circularization of the HIV-1 RNA genome. Nucleic acids research. 2007;35(15):5253-61.
306. Beerens N, Kjems J. Circularization of the HIV-1 genome facilitates strand transfer during reverse transcription. RNA (New York, NY). 2010;16(6):1226-35.
307. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS biology. 2008;6(4):e96.
308. Damgaard CK, Andersen ES, Knudsen B, Gorodkin J, Kjems J. RNA interactions in the 5' region of the HIV-1 genome. Journal of molecular biology. 2004;336(2):369-79.
309. Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM, et al. Comparison of SIV and HIV-1 genomic RNA structures reveals impact of

sequence evolution on conserved and non-conserved structural motifs. PLoS pathogens. 2013;9(4):e1003294.
310. Lavender CA, Gorelick RJ, Weeks KM. Structure-Based Alignment and Consensus Secondary Structures for Three HIV-Related RNA Genomes. PLoS computational biology. 2015;11(5):e1004230.
311. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. In vitro evidence for a long range pseudoknot in the 5'-untranslated and matrix coding regions of HIV-1 genomic RNA. The Journal of biological chemistry. 2002;277(8):5995-6004.
312. Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nature methods. 2014;11(9):959-65.
313. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009;460(7256):711-6.
314. Le Grice SF. Targeting the HIV RNA genome: high-hanging fruit only needs a longer ladder. Current topics in microbiology and immunology. 2015;389:147-69.
315. Mostoslavsky G, Kotton DN, Fabian AJ, Gray JT, Lee JS, Mulligan RC. Efficiency of transduction of highly purified murine hematopoietic stem cells by lentiviral and oncoretroviral vectors under conditions of minimal in vitro manipulation. Molecular therapy : the journal of the American Society of Gene Therapy. 2005;11(6):932-40.
316. Lee JS GJ, Mulligan R. Packaging cells comprising codon-optimized gagpol sequences and lacking lentiviral accessory proteins. Google Patents. 2006
.
317. Freed EO, Martin MA. Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. Journal of virology. 1995;69(3):1984-9.
318. Freed EO, Englund G, Maldarelli F, Martin MA. Phosphorylation of residue 131 of HIV-1 matrix is not required for macrophage infection. Cell. 1997;88(2):171-3; discussion 3-4.

319. Fouchier RA, Malim MH. Nuclear import of human immunodeficiency virus type-1 preintegration complexes. Advances in virus research. 1999;52:275-99.
320. Gottlinger HG, Sodroski JG, Haseltine WA. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(15):5781-5.
321. Zhou W, Parent LJ, Wills JW, Resh MD. Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. Journal of virology. 1994;68(4):2556-69.
322. Mammano F, Kondo E, Sodroski J, Bukovsky A, Gottlinger HG. Rescue of human immunodeficiency virus type 1 matrix protein mutants by envelope glycoproteins with short cytoplasmic domains. Journal of virology. 1995;69(6):3824-30.
323. Sherrill-Mix S, Ocwieja KE, Bushman FD. Gene activity in primary T cells infected with HIV89.6: intron retention and induction of genomic repeats. Retrovirology. 2015;12:79.
324. Mueller N, van Bel N, Berkhout B, Das AT. HIV-1 splicing at the major splice donor site is restricted by RNA structure. Virology. 2014;468-470:609-20.
325. Emery A, Zhou S, Pollom E, Swanstrom R. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. Journal of virology. 2017;91(6).
326. Mueller WF, Larsen LS, Garibaldi A, Hatfield GW, Hertel KJ. The Silent Sway of Splicing by Synonymous Substitutions. The Journal of biological chemistry. 2015;290(46):27700-11.
327. Wisdom R, Lee W. The protein-coding region of c-myc mRNA contains a sequence that specifies rapid mRNA turnover and induction by protein synthesis inhibitors. Genes & development. 1991;5(2):232-43.
328. Malim MH, Cullen BR. Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. Molecular and cellular biology. 1993;13(10):6180-9.

329. Bronson EC, Anderson JN. Nucleotide composition as a driving force in the evolution of retroviruses. Journal of molecular evolution. 1994;38(5):506-32.
330. van Hemert FJ, Berkhout B. The tendency of lentiviral open reading frames to become A-rich: constraints imposed by viral genome organization and cellular tRNA availability. Journal of molecular evolution. 1995;41(2):132-40.
331. van der Kuyl AC, Berkhout B. The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. Retrovirology. 2012;9:92.
332. Kypr J, Mrazek J, Reich J. Nucleotide composition bias and CpG dinucleotide content in the genomes of HIV and HTLV 1/2. Biochimica et biophysica acta. 1989;1009(3):280-2.
333. Grantham P, Perrin P. AIDS virus and HTLV-I differ in codon choices. Nature. 1986;319(6056):727-8.
334. Berkhout B, Grigoriev A, Bakker M, Lukashov VV. Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. AIDS research and human retroviruses. 2002;18(2):133-41.
335. Vartanian JP, Henry M, Wain-Hobson S. Sustained G-->A hypermutation during reverse transcription of an entire human immunodeficiency virus type 1 strain Vau group O genome. The Journal of general virology. 2002;83(Pt 4):801-5.
336. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA deamination mediates innate immunity to retroviral infection. Cell. 2003;113(6):803-9.
337. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature. 2003;424(6944):94-8.
338. Kim EY, Lorenzo-Redondo R, Little SJ, Chung YS, Phalora PK, Maljkovic Berry I, et al. Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection. PLoS pathogens. 2014;10(7):e1004281.
339. Kunec D, Osterrieder N. Codon Pair Bias Is a Direct Consequence of Dinucleotide Bias. Cell reports. 2016;14(1):55-67.

340. Lang DM. Circuit assemblages derived from net dinucleotide values provide a succinct identity for the HIV-1 genome and each of its genes. Virus genes. 2008;36(1):11-26.
341. Ohno S, Yomo T. Various regulatory sequences are deprived of their uniqueness by the universal rule of TA/CG deficiency and TG/CT excess. Proceedings of the National Academy of Sciences of the United States of America. 1990;87(3):1218-22.
342. Shpaer EG, Mullins JI. Selection against CpG dinucleotides in lentiviral genes: a possible role of methylation in regulation of viral expression. Nucleic acids research. 1990;18(19):5793-7.
343. Karlin S, Doerfler W, Cardon LR. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? Journal of virology. 1994;68(5):2889-97.
344. Wasson MK, Borkakoti J, Kumar A, Biswas B, Vivekanandan P. The CpG dinucleotide content of the HIV-1 envelope gene may predict disease progression. Scientific reports. 2017;7(1):8162.
345. Theys K, Feder AF, Gelbart M, Hartl M, Stern A, Pennings PS. Within-patient mutation frequencies reveal fitness costs of CpG dinucleotides and drastic amino acid changes in HIV. PLoS genetics. 2018;14(6):e1007420.
346. Rima BK, McFerran NV. Dinucleotide and stop codon frequencies in single-stranded RNA viruses. The Journal of general virology. 1997;78 (Pt 11):2859-70.
347. Pedersen AK, Wiuf C, Christiansen FB. A codon-based model designed to describe lentiviral evolution. Molecular biology and evolution. 1998;15(8):1069-81.
348. Adhya D, Basu A. Epigenetic modulation of host: new insights into immune evasion by viruses. Journal of biosciences. 2010;35(4):647-63.
349. Greenbaum BD, Levine AJ, Bhanot G, Rabadan R. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS pathogens. 2008;4(6):e1000079.
350. Simmonds P, Xia W, Baillie JK, McKinnon K. Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla--selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. BMC genomics. 2013;14:610.

351. Auewarakul P. Composition bias and genome polarity of RNA viruses. Virus research. 2005;109(1):33-7.
352. Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. Journal of molecular evolution. 2006;62(5):551-63.
353. Sewatanon J, Srichatrapimuk S, Auewarakul P. Compositional bias and size of genomes of human DNA viruses. Intervirology. 2007;50(2):123-32.
354. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. eLife. 2014;3:e04531.
355. Hershberg R, Petrov DA. Selection on codon bias. Annual review of genetics. 2008;42:287-99.
356. Burns CC, Campagnoli R, Shaw J, Vincent A, Jorba J, Kew O. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. Journal of virology. 2009;83(19):9957-69.
357. Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. Nucleic acids research. 2014;42(7):4527-45.
358. Gaunt E, Wise HM, Zhang H, Lee LN, Atkinson NJ, Nicol MQ, et al. Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. eLife. 2016;5.
359. Fros JJ, Dietrich I, Alshaikhahmed K, Passchier TC, Evans DJ, Simmonds P. CpG and UpA dinucleotides in both coding and non-coding regions of echovirus 7 inhibit replication initiation post-entry. eLife. 2017;6.
360. Takata MA, Goncalves-Carneiro D, Zang TM, Soll SJ, York A, Blanco-Melo D, et al. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. Nature. 2017;550(7674):124-7.
361. Todorova T, Bock FJ, Chang P. PARP13 regulates cellular mRNA post-transcriptionally and functions as a pro-apoptotic factor by destabilizing TRAILR4 transcript. Nature communications. 2014;5:5362.

362. Gao G, Guo X, Goff SP. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. Science (New York, NY). 2002;297(5587):1703-6.
363. Bick MJ, Carroll JW, Gao G, Goff SP, Rice CM, MacDonald MR. Expression of the zinc-finger antiviral protein inhibits alphavirus replication. Journal of virology. 2003;77(21):11555-62.
364. Mao R, Nie H, Cai D, Zhang J, Liu H, Yan R, et al. Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein. PLoS pathogens. 2013;9(7):e1003494.
365. Zhang Y, Burke CW, Ryman KD, Klimstra WB. Identification and characterization of interferon-induced proteins that inhibit alphavirus replication. Journal of virology. 2007;81(20):11246-55.
366. Muller S, Moller P, Bick MJ, Wurr S, Becker S, Gunther S, et al. Inhibition of filovirus replication by the zinc finger antiviral protein. Journal of virology. 2007;81(5):2391-400.
367. Guo X, Carroll JW, Macdonald MR, Goff SP, Gao G. The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. Journal of virology. 2004;78(23):12781-7.
368. Jeong MS, Kim EJ, Jang SB. Expression and RNA-binding of human zinc-finger antiviral protein. Biochemical and biophysical research communications. 2010;396(3):696-702.
369. Chen S, Xu Y, Zhang K, Wang X, Sun J, Gao G, et al. Structure of N-terminal domain of ZAP indicates how a zinc-finger protein recognizes complex RNA. Nature structural & molecular biology. 2012;19(4):430-5.
370. Guo X, Ma J, Sun J, Gao G. The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(1):151-6.
371. Zhu Y, Gao G. ZAP-mediated mRNA degradation. RNA biology. 2008;5(2):65-7.
372. Zhu Y, Chen G, Lv F, Wang X, Ji X, Xu Y, et al. Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(38):15834-9.

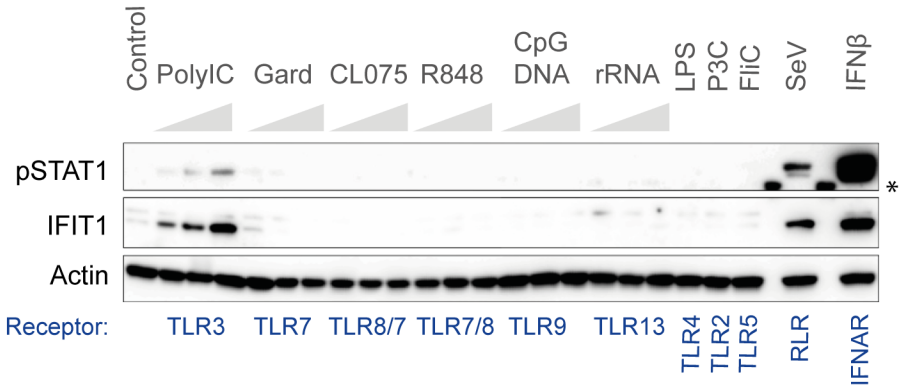
373. Kerns JA, Emerman M, Malik HS. Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. PLoS genetics. 2008;4(1):e21.
374. Lever AM. HIV-1 RNA packaging. Advances in pharmacology (San Diego, Calif). 2007;55:1-32.
375. van Bel N, Ghabri A, Das AT, Berkhout B. The HIV-1 leader RNA is exquisitely sensitive to structural changes. Virology. 2015;483:236-52.
376. Liu Y, Nikolaitchik OA, Rahman SA, Chen J, Pathak VK, Hu WS. HIV-1 Sequence Necessary and Sufficient to Package Non-viral RNAs into HIV-1 Particles. Journal of molecular biology. 2017;429(16):2542-55.
377. Kharytonchyk S, Brown JD, Stilger K, Yasin S, Iyer AS, Collins J, et al. Influence of gag and RRE Sequences on HIV-1 RNA Packaging Signal Structure and Function. Journal of molecular biology. 2018;430(14):2066-79.
378. Martrus G, Nevot M, Andres C, Clotet B, Martinez MA. Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. Retrovirology. 2013;10:78.
379. Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. Virus attenuation by genome-scale changes in codon pair bias. Science (New York, NY). 2008;320(5884):1784-7.
380. Keating CP, Hill MK, Hawkes DJ, Smyth RP, Isel C, Le SY, et al. The A-rich RNA sequences of HIV-1 pol are important for the synthesis of viral cDNA. Nucleic acids research. 2009;37(3):945-56.
381. Klaver B, van der Velden Y, van Hemert F, van der Kuyl AC, Berkhout B. HIV-1 tolerates changes in A-count in a small segment of the pol gene. Retrovirology. 2017;14(1):43.
382. Antzin-Anduetza I, Mahiet C, Granger LA, Odendall C, Swanson CM. Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication. Retrovirology. 2017;14(1):49.
383. Huang Y, Kong WP, Nabel GJ. Human immunodeficiency virus type 1-specific immunity after genetic immunization is enhanced by modification of Gag and Pol expression. Journal of virology. 2001;75(10):4947-51.
384. Takata MA, Soll SJ, Emery A, Blanco-Melo D, Swanstrom R, Bieniasz PD. Global synonymous mutagenesis identifies cis-acting RNA elements that

regulate HIV-1 splicing and replication. PLoS pathogens. 2018;14(1):e1006824.
385. Arango D, Sturgill D, Alhusaini N, Dillman AA, Sweet TJ, Hanson G, et al. Acetylation of Cytidine in mRNA Promotes Translation Efficiency. Cell. 2018.
386. Vabret N, Bhardwaj N, Greenbaum BD. Sequence-Specific Sensing of Nucleic Acids. Trends in immunology. 2017;38(1):53-65.
387. Iwasaki A, Medzhitov R. Toll-like receptor control of the adaptive immune responses. Nature immunology. 2004;5(10):987-95.
388. Kadowaki N, Ho S, Antonenko S, Malefyt RW, Kastelein RA, Bazan F, et al. Subsets of human dendritic cell precursors express different toll-like receptors and respond to different microbial antigens. The Journal of experimental medicine. 2001;194(6):863-9.
389. Krug A, Towarowski A, Britsch S, Rothenfusser S, Hornung V, Bals R, et al. Toll-like receptor expression reveals CpG DNA as a unique microbial stimulus for plasmacytoid dendritic cells which synergizes with CD40 ligand to induce high amounts of IL-12. European journal of immunology. 2001;31(10):3026-37.
390. Barbalat R, Ewald SE, Mouchess ML, Barton GM. Nucleic acid recognition by the innate immune system. Annual review of immunology. 2011;29:185-214.
391. Alinejad-Rokny H, Anwar F, Waters SA, Davenport MP, Ebrahimi D. Source of CpG Depletion in the HIV-1 Genome. Molecular biology and evolution. 2016;33(12):3205-12.
392. Sugiyama T, Gursel M, Takeshita F, Coban C, Conover J, Kaisho T, et al. CpG RNA: identification of novel single-stranded RNA that stimulates human CD14+CD11c+ monocytes. Journal of immunology (Baltimore, Md : 1950). 2005;174(4):2273-9.
393. Tanne A, Muniz LR, Puzio-Kuter A, Leonova KI, Gudkov AV, Ting DT, et al. Distinguishing the immunostimulatory properties of noncoding RNAs expressed in cancer cells. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(49):15154-9.

394. Aldovini A, Young RA. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. Journal of virology. 1990;64(5):1920-6.
395. Clavel F, Orenstein JM. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. Journal of virology. 1990;64(10):5230-4.
396. Hayashi T, Shioda T, Iwakura Y, Shibuta H. RNA packaging signal of human immunodeficiency virus type 1. Virology. 1992;188(2):590-9.
397. Luban J, Goff SP. Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. Journal of virology. 1994;68(6):3784-93.
398. McBride MS, Schwartz MD, Panganiban AT. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. Journal of virology. 1997;71(6):4544-54.
399. Clever J, Sassetti C, Parslow TG. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. Journal of virology. 1995;69(4):2101-9.
400. Hibbert CS, Mirro J, Rein A. mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. Journal of virology. 2004;78(20):10927-38.
401. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, et al. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. Science (New York, NY). 2011;334(6053):242-5.
402. Das AT, Vrolijk MM, Harwig A, Berkhout B. Opening of the TAR hairpin in the HIV-1 genome causes aberrant RNA dimerization and packaging. Retrovirology. 2012;9:59.
403. Wang B. Viral factors in non-progression. Frontiers in immunology. 2013;4:355.
404. Cummins NW, Badley AD. Can HIV Be Cured and Should We Try? Mayo Clinic proceedings. 2015;90(6):705-9.
405. Danel C, Moh R, Gabillard D, Badje A, Le Carrou J, Ouassa T, et al. A Trial of Early Antiretrovirals and Isoniazid Preventive Therapy in Africa. The New England journal of medicine. 2015;373(9):808-22.

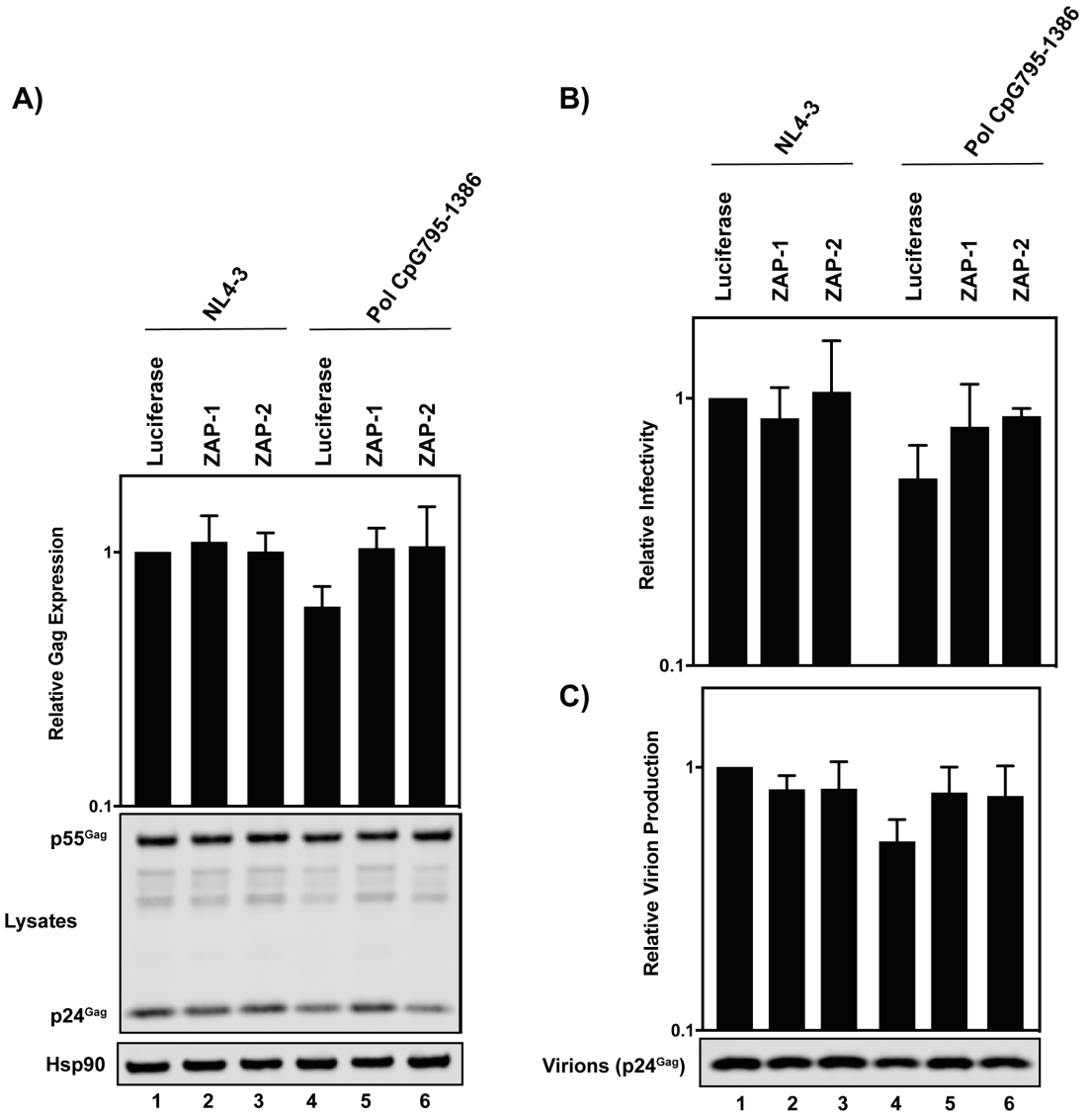
406. Williams I, Churchill D, Anderson J, Boffito M, Bower M, Cairns G, et al. British HIV Association guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2012 (Updated November 2013. All changed text is cast in yellow highlight.). HIV medicine. 2014;15 Suppl 1:1-85.
407. Silva PA, Pereira CF, Dalebout TJ, Spaan WJ, Bredenbeek PJ. An RNA pseudoknot is required for production of yellow fever virus subgenomic RNA by the host nuclease XRN1. Journal of virology. 2010;84(21):11395-406.
408. Dougherty JD, White JP, Lloyd RE. Poliovirus-mediated disruption of cytoplasmic processing bodies. Journal of virology. 2011;85(1):64-75.
409. Molleston JM, Cherry S. Attacked from All Sides: RNA Decay in Antiviral Defense. Viruses. 2017;9(1).
410. Schneider C, Tollervey D. Threading the barrel of the RNA exosome. Trends in biochemical sciences. 2013;38(10):485-93.
411. Makino DL, Baumgartner M, Conti E. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. Nature. 2013;495(7439):70-5.
412. Lubas M, Christensen MS, Kristiansen MS, Domanski M, Falkenby LG, Lykke-Andersen S, et al. Interaction profiling identifies the human nuclear exosome targeting complex. Molecular cell. 2011;43(4):624-37.

Appendix 1



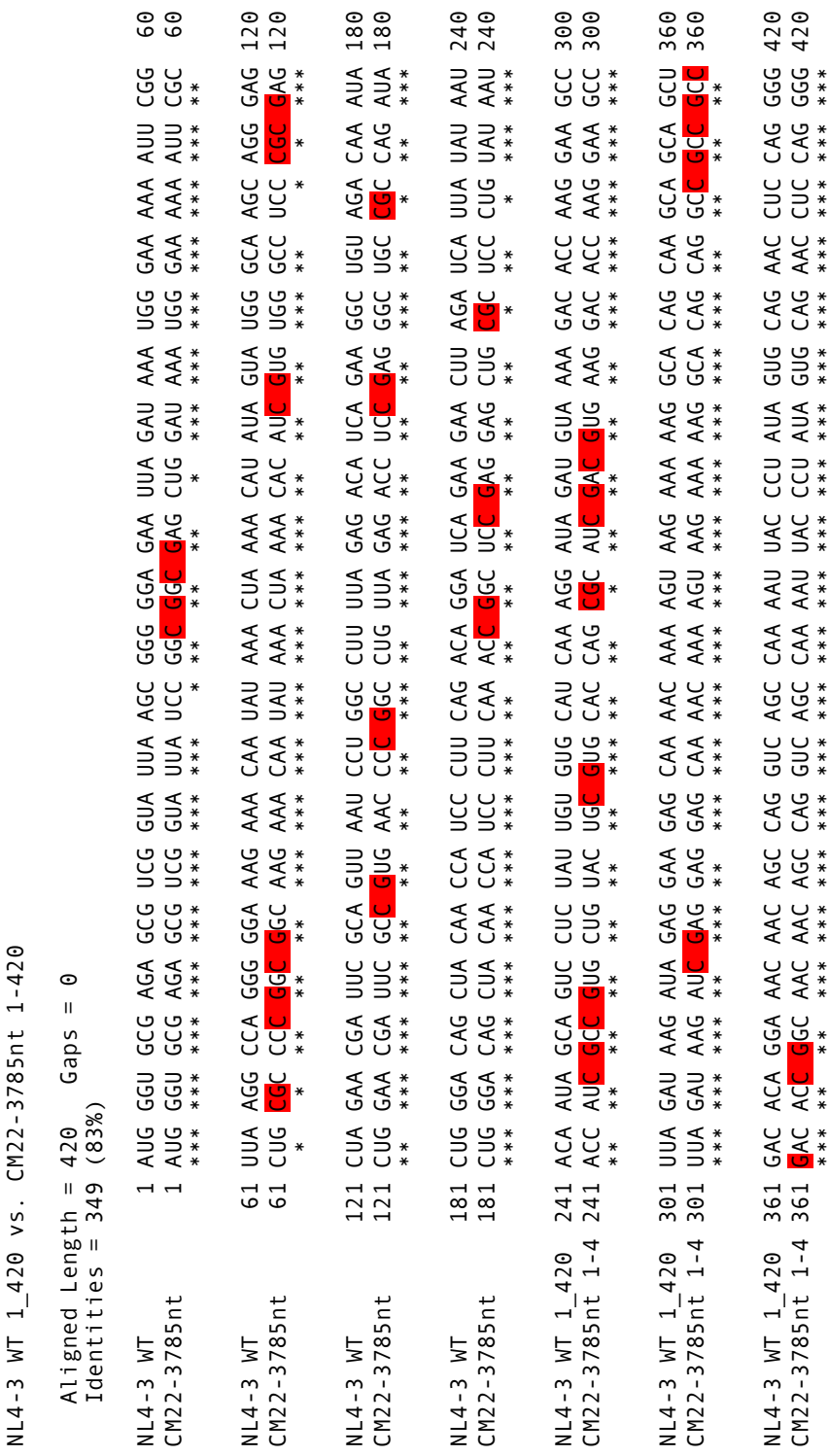
Appendix 1. CpG DNA does not stimulate STAT1 phosphorylation or ISG expression. HeLa cells were stimulated with TLR ligands for 5 h. TLR3 was targeted with 0.1, 1, 10µg/mL polyIC, TLR7 with 0.3, 3, 30 nM Gardiquimod (Gard), TLR8/7 with 0.3, 3, 30nM CL075, TLR7/8 with 0.01, 0.1, 1µg/mL R848, TLR9 with 0.01, 0.1, 1µg/ml CpG DNA, TLR13 with 0.01, 0.1, 1µg/mL 23S ribosomal RNA (rRNA). TLR 4, 2 and 5 were targeted with 0.1 µg/mL LPS, Pam3Cys (P3C) or Flagellin (FliC), respectively. As controls for pattern recognition receptor signaling and JAK-STAT signaling, cells were infected with 50 HAU/mL Sendai virus (SeV) for 5 h or stimulated with 0.01 µg/mL IFN-β for 1 h, respectively. Activation of IFN signaling was monitored by western immunoblotting against phosphorylated STAT1 (pSTAT1) or expression of the ISG IFIT1. Actin was used as a loading control. (* Denotes the molecular weight marker). *Experiment performed by Charlotte Odendall (Lecturer, KLC).*B

Appendix 2



Appendix 2. Inhibition of viral infectivity upon introduction of CpGs into *pol* is ZAP-dependent. Introduction of CpGs into 795-1386 produces a mild inhibition of Gag expression, virion production and infectivity HeLa cells. Depletion of ZAP rescues infectious virus production. *Experiment performed by Mattia Ficarelli (PhD student, Swanson Lab, KCL).*

Appendix 3



Appendix 3. Sequence alignment of HIV-1_{NL4-3} and HIV-1 GagCM22-378_{5nt}. Alignment starts in the *gag* AUG. CpG dinucleotides are highlighted in red.

Appendix 4

Antzin-Anduetza *et al. Retrovirology (2017) 14:49*
DOI 10.1186/s12977-017-0374-1

Retrovirology

RESEARCH Open Access

Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication

Irati Antzin-Anduetza, Charlotte Mahiet, Luke A. Granger, Charlotte Odendall and Chad M. Swanson*

Abstract

Background: The human immunodeficiency virus type 1 (HIV-1) structural protein Gag is necessary and sufficient to form viral particles. In addition to encoding the amino acid sequence for Gag, the underlying RNA sequence could encode *cis*-acting elements or nucleotide biases that are necessary for viral replication. Furthermore, RNA sequences that inhibit viral replication could be suppressed in *gag*. However, the functional relevance of RNA elements and nucleotide biases that promote or repress HIV-1 replication remain poorly understood.

Results: To characterize if the RNA sequence in *gag* controls HIV-1 replication, the matrix (MA) region was codon modified, allowing the RNA sequence to be altered without affecting the protein sequence. Codon modification of nucleotides (nt) 22-261 or 22-378 in *gag* inhibited viral replication by decreasing genomic RNA (gRNA) abundance, gRNA stability, Gag expression, virion production and infectivity. Comparing the effect of these point mutations to deletions of the same region revealed that the mutations inhibited infectious virus production while the deletions did not. This demonstrated that codon modification introduced inhibitory sequences. There is a much lower than expected frequency of CpG dinucleotides in HIV-1 and codon modification introduced a substantial increase in CpG abundance. To determine if they are necessary for inhibition of HIV-1 replication, codons introducing CpG dinucleotides were mutated back to the wild type codon, which restored efficient Gag expression and infectious virion production. To determine if they are sufficient to inhibit viral replication, CpG dinucleotides were inserted into *gag* in the absence of other changes. The increased CpG dinucleotide content decreased HIV-1 infectivity and viral replication.

Conclusions: The HIV-1 RNA sequence contains low abundance of CpG dinucleotides. Increasing the abundance of CpG dinucleotides inhibits multiple steps of the viral life cycle, providing a functional explanation for why CpG dinucleotides are suppressed in HIV-1.

Keywords: HIV-1, Genomic RNA, CpG dinucleotide, Viral replication

Background

The HIV-1 genomic RNA (gRNA) has three major functions in the viral life cycle [1]. First, it serves as the pre-mRNA that is spliced into over 70 different transcripts [2–4]. Second, it acts as the mRNA for the Gag and Gag-Pol polyproteins that comprise the structural and enzymatic proteins, respectively [5, 6]. Third, it is the genome that is packaged into virions and is reverse-transcribed upon infection of a new target cell [7, 8]. The gRNA can be divided into three regions: a 336 nt 5′ untranslated region (UTR), a 219 nt 3′ UTR, and an 8618 nt region that is densely packed with multiple open reading frames (nt lengths reference the HIV-1_{NL4-3} strain [9]). The 5′ UTR contains several *cis*-acting elements in complex stem-loop structures that regulate multiple stages of the viral life cycle including transcription, splicing, gRNA dimerization, encapsidation and reverse transcription [8, 10]. The central 8618 nt region encodes nine open reading frames: *gag*, *pol*, *vif*, *vpr*, *tat*, *rev*, *vpu*, *env* and *nef*.

*Correspondence: chad.swanson@kcl.ac.uk
Department of Infectious Diseases, King's College London, 3rd Floor Borough Wing, Guy's Hospital, London SE1 9RT, UK

In addition to encoding the amino acids of the viral proteins, the RNA sequence underlying the open reading frames could regulate multiple steps of the HIV-1 life cycle including splicing, RNA stability, RNA nuclear export, translation and reverse transcription. Indeed, a large number of *cis*-acting RNA elements within the protein coding regions have been reported to regulate HIV-1 replication, some of which are highly conserved and under purifying selection [11, 12]. These include the programmed ribosomal frameshift sequence in *gag* for Pol translation [13], splicing signals in *pol*, *vif*, *vpr*, *tat*, *rev* and *env* [2, 3], the Rev-response element (RRE) in *env* [14] and the polypurine tracts in *pol* and *nef* that are necessary for reverse transcription [15]. There is also extensive secondary and tertiary RNA structure throughout the gRNA that could regulate viral replication [16–19].

Determining the full complement of *cis*-acting elements in the gRNA that regulate viral replication is necessary for a complete understanding of the HIV-1 replication cycle and may aid in the development of novel antiviral therapies [20]. Furthermore, identifying and characterizing evolutionarily conserved *cis*-acting elements and structures is essential for understanding HIV-1 purifying and positive selection as well as recombination events [11, 12, 21–24]. Gag consists of four protein domains and two spacer peptides that control virion assembly [25]. Matrix (MA/p17) mediates Gag trafficking to the plasma membrane, capsid (CA/p24) forms the structure of the virion core, nucleocapsid (NC/p7) binds the genomic RNA to mediate encapsidation, and p6 recruits the ESCRT complexes necessary for membrane fission during budding. Within the MA open reading frame, there are a large number of proposed *cis*-acting RNA elements that could be necessary for viral replication. These include a hnRNPA1 binding site that may regulate gRNA nuclear export [26], an intronic splice enhancer [27, 28], an internal ribosome entry site [29], instability sequences that lead to RNA degradation in the absence of Rev [30], sequences that base pair with the 5′ and 3′ UTRs [31–35], and elements that regulate encapsidation [7, 8]. However, the functional relevance of most of these elements for viral replication is unclear.

Some nucleotide patterns may also regulate HIV-1 replication and be under evolutionary selection. For example, the base composition of HIV-1 deviates from that of the human genome. HIV-1 RNA has a high percentage of adenine (A, 36%) and low percentage of cytosine (C, 18%) [36–43]. This nucleotide bias is found in groups M, N and O and is a general property of lentiviruses, though not all retroviruses [36, 38, 42, 44]. Even though HIV-1 has a very high nucleotide substitution rate and sequence diversity, the A-rich bias has been conserved during the HIV-1 pandemic [42]. There are two hypotheses for why this has been maintained in the virus. First, the mutational pattern of reverse transcriptase or antiviral APOBEC3 proteins could impose an A-rich nucleotide bias [45–52]. Second, this bias could be required for viral replication and be under purifying selection [38, 53, 54].

In addition to understanding the RNA elements that are necessary for viral replication, it is important to characterize the motifs that are underrepresented and may be deleterious. HIV-1 has a much lower than expected frequency of the dinucleotide CpG [36, 40, 44, 55–57]. This has been proposed to be under negative selection and the CpG dinucleotide abundance in HIV-1 may be linked to disease progression [58]. However, the mechanism by which CpG dinucleotides affect viral replication is unknown.

These nucleotide biases cause the HIV-1 open reading frames to have a codon usage pattern that differs substantially from that of human mRNAs [36–41, 43]. The genetic code is redundant in that there are 61 codons for 20 amino acids and all of the amino acids except methionine and tryptophan are encoded by at least two codons. The preferred codons in cellular mRNAs are thought to correlate with the availability of the aminoacyl-tRNAs but HIV-1 contains many rare codons [36–41].

In this study, we investigated whether RNA elements in the MA region of *gag* positively or negatively regulate HIV-1 replication. We initially focused on this region because of its high content of potential RNA regulatory elements (discussed above). To change the RNA sequence without altering the amino acid sequence, we codon modified this region by introducing large numbers of synonymous mutations. These mutations strongly inhibited viral replication by decreasing gRNA abundance, gRNA stability, Gag expression, virion production and infectivity. We found that CpG dinucleotides introduced during codon modification were necessary and sufficient to attenuate HIV-1 replication. This highlights the functional importance of the suppressed CpG abundance in HIV-1 [36, 40, 44, 55–57] and shows that increasing the number of CpG dinucleotides in the gRNA inhibits multiple steps of the viral life cycle.

Methods

Cell culture and transfections

Jurkat cells were cultured in RPMI 1640 GlutaMAX Medium (Gibco) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin. Hela, TZM-bl and 293T cells were cultured in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 10% FBS and 1% penicillin–streptomycin. All cell lines were grown at 37 °C in a humidified atmosphere with 5% CO_2.

Plasmids

The pHIV-1_{NL4-3} constructs used in this study contain the provirus sequence from pHIV-1_{NL4-3} [9] cloned into the KpnI and SalI sites of pGL4.10 (Promega). pHIV-1 CM22-261, pHIV-1 CM22-165, pHIV-1 CM166-261 and pHIV-1 CM22-378 have the designated sequences from pHDMHgpm2 [59] chemically synthesized by Life Technologies and cloned into pHIV-1_{NL4-3}. For pHIV-1CM 22-261_{lowCpG} and pHIV-1 CM22-378_{lowCpG}, pHIV-1 CpG22-261 and pHIV-1 CpG22-378, the sequences shown in Fig. 6 were synthesized by Life Technologies and cloned into pHIV-1_{NL4-3}. pHIV-1 Δ22-261 and pHIV-1 Δ22-378 have the designated region in *gag* replaced with a XbaI site as in Reil et al. [60]. The modified sequences in these plasmids were verified by DNA sequencing (Eurofins). pGFP and pVSV-G have been previously described [61, 62].

HIV-1 spreading infection assay

4×10^6 293T cells were seeded in 10 cm plates and transfected with 10 μg of pHIV-1 and 1.25 μg of pGFP using poly(ethlyleneimine) solution (PEI) at a ratio of 5 μl PEI per 1 μg DNA. Approximately 48-h post-transfection, the media was harvested, filtered through a 0.45 μm filter and quantified using a $p24^{Gag}$ enzyme-linked immunosorbent assay (ELISA) (Perkin-Elmer). A total of 2.5×10^5 Jurkat cells were plated in 1 mL of medium per well in 48 well plates and infected with 25 ng of $p24^{Gag}$ of each virus. SupT1 cells were infected with 10 ng of $p24^{Gag}$ for each virus. Supernatants were first collected when syncytia were first observed in the culture infected with HIV-1_{NL4-3}. The amount of infectious virus present at each time point was quantified by infecting the TZM-bl indicator cell line [63–65]. Infectivity was measured by the induction of β-galactosidase using the Galacto-Star™ System (Applied Biosystems).

Single cycle infectivity assay

Six-well plates of HeLa cells were transfected using TransIT®-LT1 (Mirus) according to the manufacturer's instructions at the ratio of 3 μL TransIT®-LT1 to 1 μg DNA. For each transfection, 0.5 μg pHIV-1 and 0.5 μg pGFP or pVSV-G was used. Media was recovered approximately 48 h post-transfection and filtered through a 20% sucrose cushion for 2 h at 20,000×*g*. The amount of infectious virus was quantified by using the TZM-bl indicator cell line [63–65].

Analysis of protein expression by immunoblotting

Approximately 48-h post-transfection, HeLa cells were lysed in radioimmunoprecipitation assay (RIPA) buffer (10 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 1% sodium deoxycholate). The media was clarified using a 0.45 μm filter. Virions were pelleted through a 20% sucrose cushion in phosphate-buffered saline (PBS) solution for 2 h at 20,000×*g*. The pellet was resuspended in 2× loading buffer (60 mM Tris–HCl (pH 6.8), 10% β-mercaptoethanol, 10% glycerol, 2% sodium dodecyl sulfate (SDS), 0.1% bromophenol blue). Cell lysates and virions were resolved by SDS–polyacrylamide gel electrophoresis and transferred to a nitrocellulose membrane. The primary antibodies used were specific to HIV-1 $p24^{Gag}$ [66], Hsp90 (sc7947: Santa Cruz Biotechnology), phosphoSTAT1 (612132: BD Transduction), IFIT1 (GTX118713-S: Insight Biotechnology) or β-actin (ac-15: Sigma). Dylight™ 800-conjugated secondary antibodies (5151S and 5257S: Cell Signaling) were used to detect the bound primary antibodies with the Li-CoR infrared imaging (LI-COR UK LTD).

Quantitative RT-PCR

Hela cells were washed with 1xPBS and the RNA was extracted using the RNeasy kit (Qiagen) following the manufacturer's instructions. 1 μg of RNA was reverse transcribed to cDNA using the High Capacity cDNA archive kit (Applied Biosystems). RNA from virions was isolated using QIAamp viral RNA mini kit following the manufacturer's instructions. Because carrier RNA is added to the lysis buffer, the total RNA isolated was quantified using a Qubit 3.0 fluorometer (ThermoFisher) and normalized so that 20 ng of RNA from each sample was reverse transcribed using the High Capacity cDNA archive kit (Applied Biosystems). PCR reactions were performed in triplicate with Taqman Universal PCR mix using the Applied Biosystems 7500 real-time PCR system. HIV-1_{NL4-3} gRNA primers were GGCCAGGGAATTTTCTTCAGA/TTGTCTCTTCCCCAAACCTGA (forward/reverse) and the probe was FAM-ACCAGAGCCAACAGCCCCACCAGA-TAMRA. HIV-1_{NL4-3} total RNA primers were TAACTAGGGAACCCACTGC/GCTAGAGATTTTCCACACTG (forward/reverse) and the probe was FAM-ACACAACAGACGGGCACACACTA-TAMRA. To analyze gRNA stability, 1 μg/ml Actinomycin D (Sigma Aldrich) was added to HeLa cells ~ 45 h post-transfection. RNA was isolated at the designated timepoints and gRNA abundance was measured.

TLR and IFN stimulations, Sendai virus infection

HeLa cells were stimulated with synthetic TLR ligands for 5 h at the concentrations indicated. Ligands supplied by Invivogen were polyIC: polyIC (tlrl-pic), Gardiquimod (tlrl-gdqs), CL075 (tlrl-c75), R848 (tlrl-r848), Pam3CSK4 (P3C. tlrl-pms), Ultrapure Flagellin (FliC-tlrl-epstfla-5). LPS was supplied by Enzo (ALX-581-012-L002). CpG DNA was synthesised by IDT and 23S ribosomal RNA

by Sigma. Sendai Virus (SeV) was obtained from Charles River labs. IFN-β was purchased from Peprotech and was added to the culture for 1 h to activate IFN signaling.

Sequence analysis of the HIV-1_{NL4-3} gRNA

The "analyze base composition" tool in MacVector was used to calculate the mono- and di-nucleotide frequencies for the HIV-1_{NL4-3} gRNA (NCBI accession number M19921). The dinucleotide frequencies are calculated using the following formula: number of dinucleotide occurrences/(frequency of base 1 in pair × frequency of base 2 in pair) where frequency of base is number of occurrences of base/total number of bases in sequence. WebLogo [67] was used to generate conserved nucleotides surrounding the CpG dinucleotides.

Results

Synonymous mutations in *gag* inhibit HIV-1 replication

To analyze the functional relevance of RNA elements and nucleotide bias underlying the MA domain in Gag, we introduced 80 synonymous mutations into nt 22-261 of HIV-1_{NL4-3} *gag* (Fig. 1a). This codon modified (CM) provirus, HIV-1 CM22-261, has 69/80 codons in this region altered without affecting the amino acid sequence. The mutations were derived from pHDMHgpm2, a codon optimized Gag-Pol construct in which many of the HIV-1 codons are replaced with codons used in highly expressed human mRNAs [59, 68]. In addition, nt 22-165 or 166-261 in *gag* were codon modified to produce HIV-1 CM22-165 and HIV-1 CM166-216, which have 49 and 31 synonymous mutations, respectively. Virus stocks were prepared by transfecting 293T cells with each proviral DNA construct and the concentration of viral CA/$p24^{Gag}$ for each stock was measured by ELISA. HIV-1 CM22-261 and HIV-1 CM22-165 had an ~ 65% and ~ 40% decrease in $p24^{Gag}$ concentration, respectively (Fig. 1b). To analyze the fitness of each virus, the viral inoculum was normalized so that Jurkat CD4 T cells were challenged with 25 ng of $p24^{Gag}$ for each virus (Fig. 1c). The amount of infectious virus in the culture supernatant was monitored over 2 weeks using TZM-bl indicator cells [63–65]. HIV-1 CM22-261 replicated at a very low but detectable level and at day 12 had > 99.9% less infectivity than wild type virus. HIV-1 CM22-165 replicated slightly better than HIV-1 CM22-261, but was still > 99% lower than wild type HIV-1 at day 12. HIV-1 CM166-261 plateaued at the same level as wild type HIV-1 but with a delay of ~ 3 days. Similar results were observed when SupT1 CD4 T cells were challenged with the wild type and mutated viruses (Additional file 1).

We then used a single cycle infectivity assay to determine if HeLa cells were non-permissive for HIV-1 CM22-261, HIV CM22-165 or HIV-1 CM166-261 replication as well as to characterize which steps in the viral life cycle are inhibited by the synonymous changes in *gag*. HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 CM22-261, pHIV-1 CM22-165 or pHIV-1 CM166-261 and the media and cell lysates were harvested ~ 48 h later. HIV-1 infectivity in the media was determined using TZM-bl cells and the abundance of virions in the media and Gag in the cell lysate was analyzed by quantitative immunoblotting. Compared with the wild type virus, HIV-1 CM22-261 infectivity was decreased to the limit of detection of the assay (Fig. 2a), indicating that the virus is attenuated in HeLa cells. Virion production and intracellular Gag expression were decreased ~ 90% (Fig. 2b, c). For HIV-1 CM22-165, the amount of infectious virus in the media was decreased ~ 98% with a < 50% decrease in Gag expression and viron production. HIV-1 CM166-261 consistently yielded similar amounts of infectivity, virions and intracellular Gag expression as wild type HIV-1. Overall, there is a substantial reduction in infectivity for HIV-1 CM22-261 and HIV-1 CM22-165. HIV-1 CM22-261 also has a substantial defect in Gag expression and virion production.

To determine if the decrease in infectious virus production was due to a decrease in gRNA abundance, we performed quantitative RT-PCR (qRT-PCR) using a primer–probe set in a region of *gag* that was not mutated (Fig. 3a, b). pHIV-1_{NL4-3}, pHIV-1 CM22-261, pHIV-1 CM22-165 and pHIV-1 CM22166-261 were transfected into HeLa cells. RNA was isolated from the cell lysate and media ~ 48 h post-transfection. HIV-1 CM22-261 gRNA was reduced > 90% in the cell lysate and media compared with the wild type virus. HIV-1 CM22-165 gRNA was decreased ~ 70% in the cell and ~ 65% in the media. HIV-1 CM166-261 gRNA abundance was equivalent to wild type HIV-1 in the cell lysate and media. We then determined the effect of the synonymous mutations on infectivity/viral genome by infecting TZM-bl cells with an equivalent amount of viral genomes for each virus. When the input number of genomes was normalized based on the results in Fig. 3b, HIV-1 CM22-261 infectivity was at the limit of detection of the assay and HIV-1 CM22-165 infectivity was decreased ~ 98% (Fig. 3c). This indicates that the decreased abundance of viral genomes in the media is not fully responsible for the loss of infectivity for HIV-1 CM22-261 and HIV-1 CM22-165.

The HIV-1 gRNA can be spliced into over 70 different transcripts [4] and the *gag-pol* intron can be spliced out through one 5′ splice site and six 3′ splice sites [2]. A potential explanation for the decrease in intracellular gRNA abundance is that the mutations in *gag* could affect intronic splicing silencer (ISS) sequences. If this occurred, gRNA abundance would decrease due to oversplicing but the total amount of HIV-1 RNA would stay the same. To test this, we determined the total

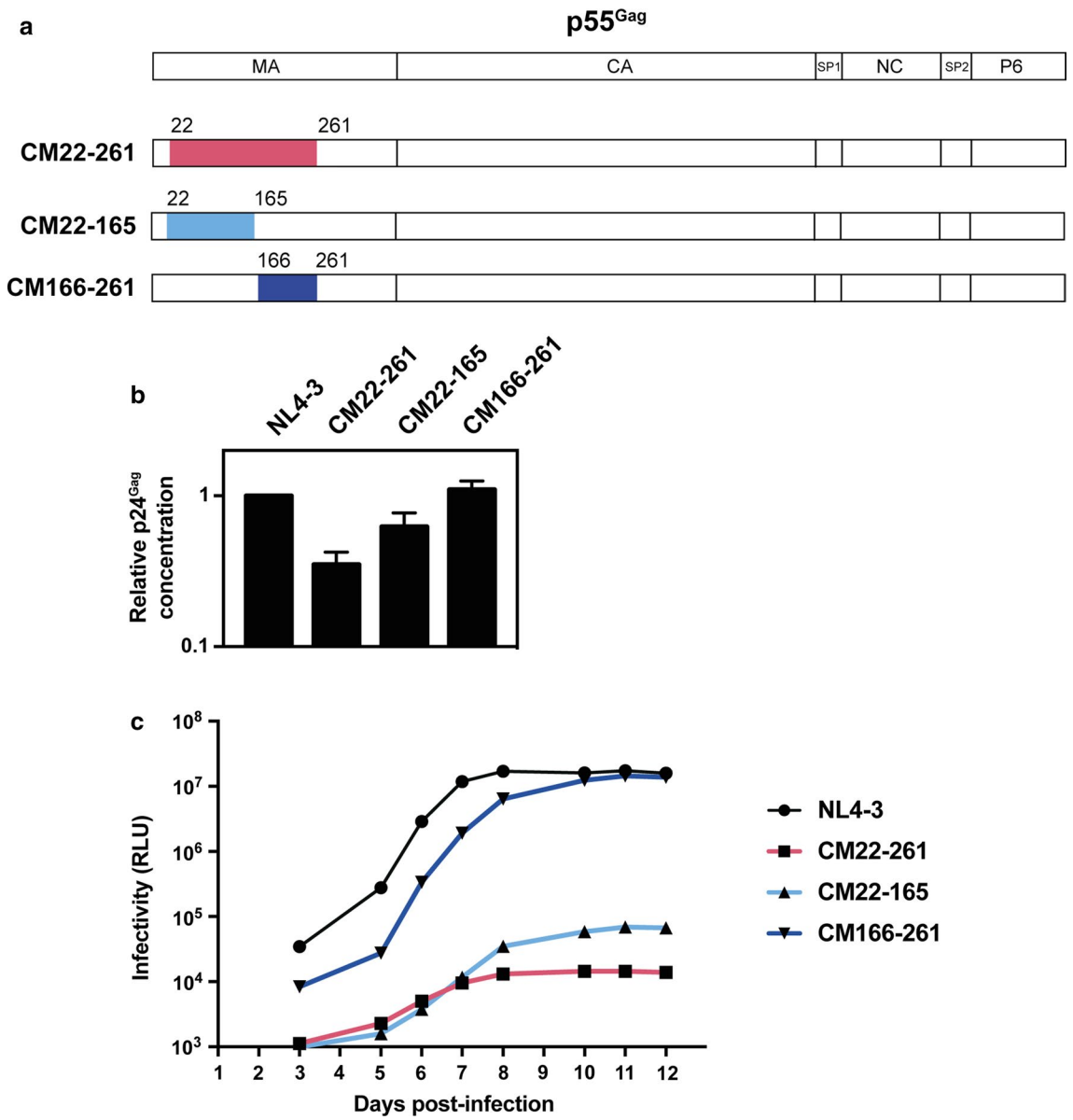


Fig. 1 Codon modification of nucleotides 22-261 in *gag* inhibits viral replication. **a** Schematic representation of $p55^{Gag}$ in HIV-1_{NL4-3}, HIV-1 CM22-261, HIV-1 CM22-165 and HIV-1 CM166-261. **b** The amount of HIV-1 CA ($p24^{Gag}$) in supernatants from 293T cells transfected with pHIV-1_{NL4-3}, pHIV-1 CM22-261, pHIV-1 CM22-165 or pHIV-1 CM166-261 were quantified by $p24^{Gag}$ ELISA. The bar chart is the average of three independent experiments normalized to HIV-1_{NL4-3}. Error bars represent standard deviation. **c** Jurkat cells were infected with 25 ng of $p24^{Gag}$ for each indicated virus. The amount of infectious virus present at each time point was measured in TZM-bl cells. This is representative of three independent experiments

intracellular HIV-1 RNA abundance using a primer–probe set upstream of the major 5′ splice donor (SD1). HIV-1 CM22-261 and HIV-1 CM22-165 had an ~ 80 and ~ 60% decrease in total HIV-1 RNA abundance (Additional file 2). Since ~ 50% of the gRNA remains unspliced [2], this is consistent with a specific reduction in the gRNA and does not appear to be a consequence of oversplicing.

CpG dinucleotides are necessary for the inhibition of infectious virus production

The synonymous mutations introduced into *gag* could inhibit HIV-1 replication by inactivating essential *cis*-acting RNA elements or introducing inhibitory elements. The region mutated in HIV-1 CM22-261 was designed to match the codons previously deleted by Reil et al. [60] in HIV-1_{HXBH10}Δ8-87/ΔCT. In this virus, amino acids 8-87

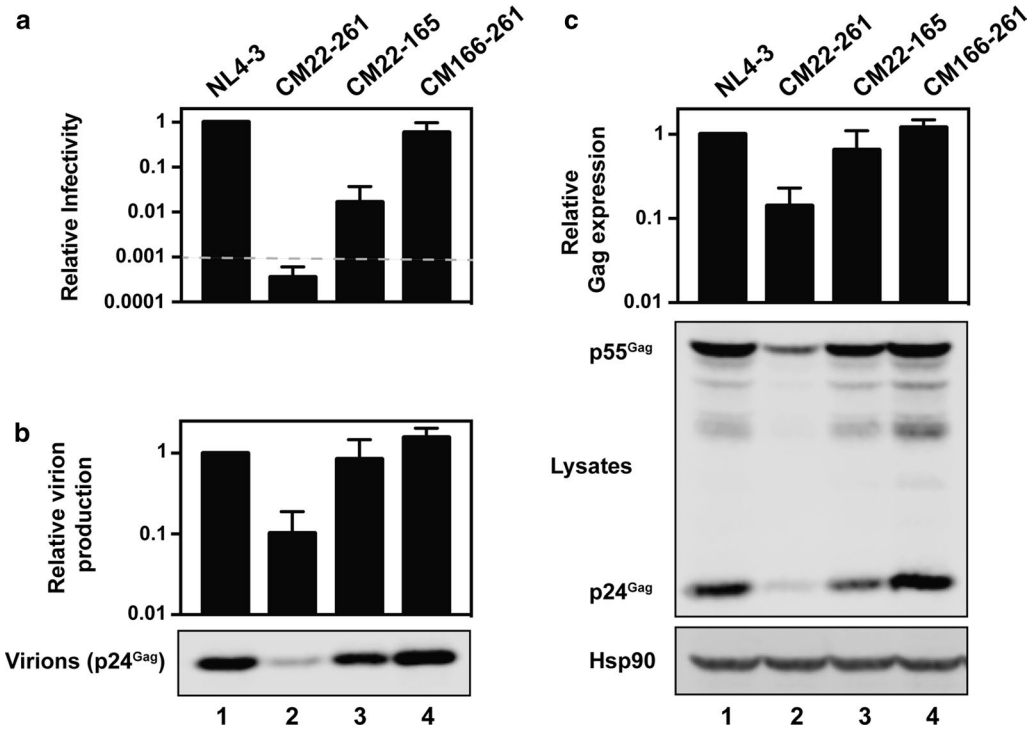


Fig. 2 Codon modification of nucleotides 22-261 in *gag* inhibits infectious virus production. HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 CM22-261, pHIV-1 CM22-165 or pHIV-1 CM166-261. **a** Culture supernatants were used to infect TZM-bl reporter cells to measure viral infectivity. The bar charts show the average values of six independent experiments normalized to the value obtained for HIV-1_{NL4-3}. The average relative light units (RLU) for HIV-1_{NL4-3} is 3,680,747. The dashed line represents 1000 RLU, which is approximately the limit of the assay for reproducible differential results. Error bars represent standard deviation. **b**, **c** Gag expression in the media (**b**) and cell lysate (**c**) was detected using quantitative immunoblotting. The bar charts show the average of four independent experiments normalized to HIV-1_{NL4-3}. Error bars represent standard deviation

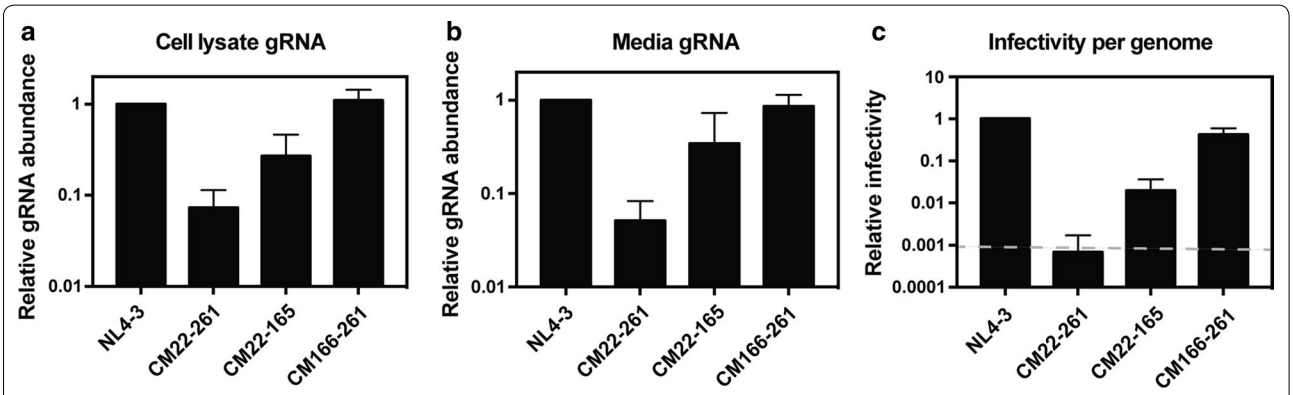


Fig. 3 Codon modification of nucleotides 22-261 in *gag* decreases gRNA abundance. HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 CM22-261, pHIV-1 CM22-165 or pHIV-1 CM166-261. The bar charts show the average values of three independent experiments normalized to HIV-1_{NL4-3}. Error bars represent standard deviation. **a** RNA was extracted from cell lysates and gRNA abundance was quantified by qRT-PCR. **b** RNA was extracted from the media and gRNA abundance was quantified by qRT-PCR. **c** Equivalent amounts of HIV-1 genomes were used to infect TZM-bl reporter cells to measure infectivity. The bar charts show the average values of three independent experiments normalized to the value obtained for HIV-1_{NL4-3}. The average RLU for HIV-1_{NL4-3} is 1,146,196. The dashed line represents 1000 RLU, which is approximately the limit of the assay for reproducible differential results

(nt 22-261) in Gag were deleted and a stop codon in Env removed the cytoplasmic tail domain. Because deletions in the globular core domain of MA prevent incorporation of Env with a full cytoplasmic domain [69], the truncated Env cytoplasmic tail is necessary for virion infectivity. However, pseudotyping with heterologous envelope glycoproteins, such as that from vesicular stomatitis virus (VSV-G), allow viral entry into a target cell. HIV-1_{HXBH10} Δ8-87/ΔCT replicates as well as HIV-1_{HXBH10} ΔCT in the MT4 cell line [60], indicating that neither the protein or RNA sequences in this region are necessary for viral replication in these cells.

To determine whether the synonymous mutations inserted into nt 22-261 of *gag* removed essential *cis*-acting elements or inserted deleterious sequences, we generated a HIV-1_{NL4-3} Δ22-261 provirus construct (Fig. 4a) and compared it with HIV-1 CM22-261 in the absence or presence of VSV-G. In the absence of VSV-G, HIV-1 Δ22-261 produced very low levels of infectious virus (Fig. 4b), which was expected due to the role of MA in recruiting Env with a full-length cytoplasmic tail. Gag expression and virion production were similar for wild type HIV-1 and HIV-1 Δ22-261 (Fig. 4c), indicating that RNA or protein sequences in this region are not necessary for these steps of the viral life cycle. When the viruses were pseudotyped with VSV-G, HIV-1 Δ22-261 infectivity was similar to wild type virus, confirming that the only functional defect for this virus in HeLa cells is Env incorporation. In contrast, HIV-1 CM22-261 was not rescued by VSV-G pseudotyping and had a > 99.9% reduction in infectivity (Fig. 4d).

Reil et al. [60] also deleted amino acids 8-126 (nt 22-378) in MA and found that HIV-1_{HXBH10} Δ8-126/ΔCT replicated with moderately delayed kinetics compared to HIV-1_{HXBH10} ΔCT in MT4 cells. We produced HIV-1_{NL4-3} provirus constructs in which this region was either deleted (HIV-1 Δ22-378) or codon modified (HIV-1 CM22-378). VSV-G pseudotyped HIV-1 Δ22-378 had a small decrease in infectious virus production compared to wild type HIV-1 (Fig. 4d), which correlated with virion production (Fig. 4e). However, VSV-G pseudotyped HIV-1 CM22-378 had a > 99.9% decrease in infectious virus production with Gag expression and virion production near the limit of detection (Fig. 4d, e). Therefore, we concluded that codon modification of nt 22-261 or nt 22-378 of *gag* introduced inhibitory sequences into the HIV-1 genome that reduce Gag expression, virion production and infectivity.

To determine whether the synonymous mutations in *gag* altered the stability of the viral RNA, HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 CM22-261 or pHIV-1 CM22-378 and, ~ 45 h post-transfection, RNA polymerase II-dependent transcription was inhibited by adding actinomycin D. RNA was isolated from the cells immediately before actinomycin D addition (0 h) and then 1, 2, 4 and 6 h thereafter. gRNA abundance at the 0 h timepoint was substantially decreased for HIV-1 CM22-261 and HIV-1 CM22-378 (Fig. 5a) and correlated with the length of codon modified sequence. Since *MYC* mRNA has a half-life of < 1 h [70], we used it as a control for RNA stability and analyzed its abundance at each timepoint. As expected, *MYC* mRNA was rapidly degraded (Fig. 5b). The gRNA abundance for HIV-1_{NL4-3}, HIV-1 CM22-261 and HIV-1 CM22-378 decreased by ~ 20, ~ 35 and ~ 70%, respectively, at the 6 h timepoint relative to its abundance at 0 h (Fig. 5c). This indicates that HIV-1 CM22-261 and HIV-1 CM22-378 gRNA is less stable than HIV-1_{NL4-3} gRNA. Comparing HIV-1 CM22-378 gRNA abundance to that of HIV-1_{NL4-3} at the 6 h timepoint, HIV-1 CM22-378 gRNA was ~ 60% lower than wild type virus gRNA. If the degradation rate is constant, a 50% decrease every 6 h in gRNA abundance for HIV-1 CM22-378 relative to HIV-1_{NL4-3} would be compounded to yield a 98.4% decrease after 36 h. This is consistent with the ~ 98% decrease in steady state gRNA for HIV-1 CM22-378 that we observed ~ 45 h post-transfection (Fig. 5a). Overall, the synonymous mutations introduced into *gag* appear to decrease the stability of the gRNA.

Two types of RNA dinucleotide patterns have previously been implicated in restricting RNA virus replication, UpA and CpG [71–74]. The observed/expected ratio for UpA in the HIV-1_{NL4-3} gRNA is 0.92 (Table 1) and the total number of UpA dinucleotides in nt 22-261 and 22-378 decreased substantially in the codon modified sequence compared with the wild type sequence (Table 2). This indicates that UpA dinucleotide content is not causing the inhibitory phenotype. The CpG dinucleotide observed/expected ratio is 0.21 in HIV-1_{NL4-3} and it is the only dinucleotide substantially suppressed (Table 1). Within nt 22-261 of *gag*, wild type HIV-1 has 4 CpG dinucleotides and the codon modified sequence has 22 (Table 2). Similarly, codon modification of nt 22-378 increased the number of CpGs from 4 to 30. Because CpG dinucleotides are underrepresented in HIV-1 (Table 1) [40, 44, 55–57] and previous reports have shown that increasing the CpG dinucleotide abundance inhibits picornavirus and influenza virus replication [71–74], we hypothesized that the increased number of CpG dinucleotides in HIV-1 CM22-261 and HIV-1 CM22-378 could cause the decrease in HIV-1 infectious virion production.

To test this hypothesis, we synthesized a HIV-1 *gag* sequence containing all of the synonymous mutations present in pHIV CM22-261 with the exception of the codon changes that introduced CpG dinucleotides

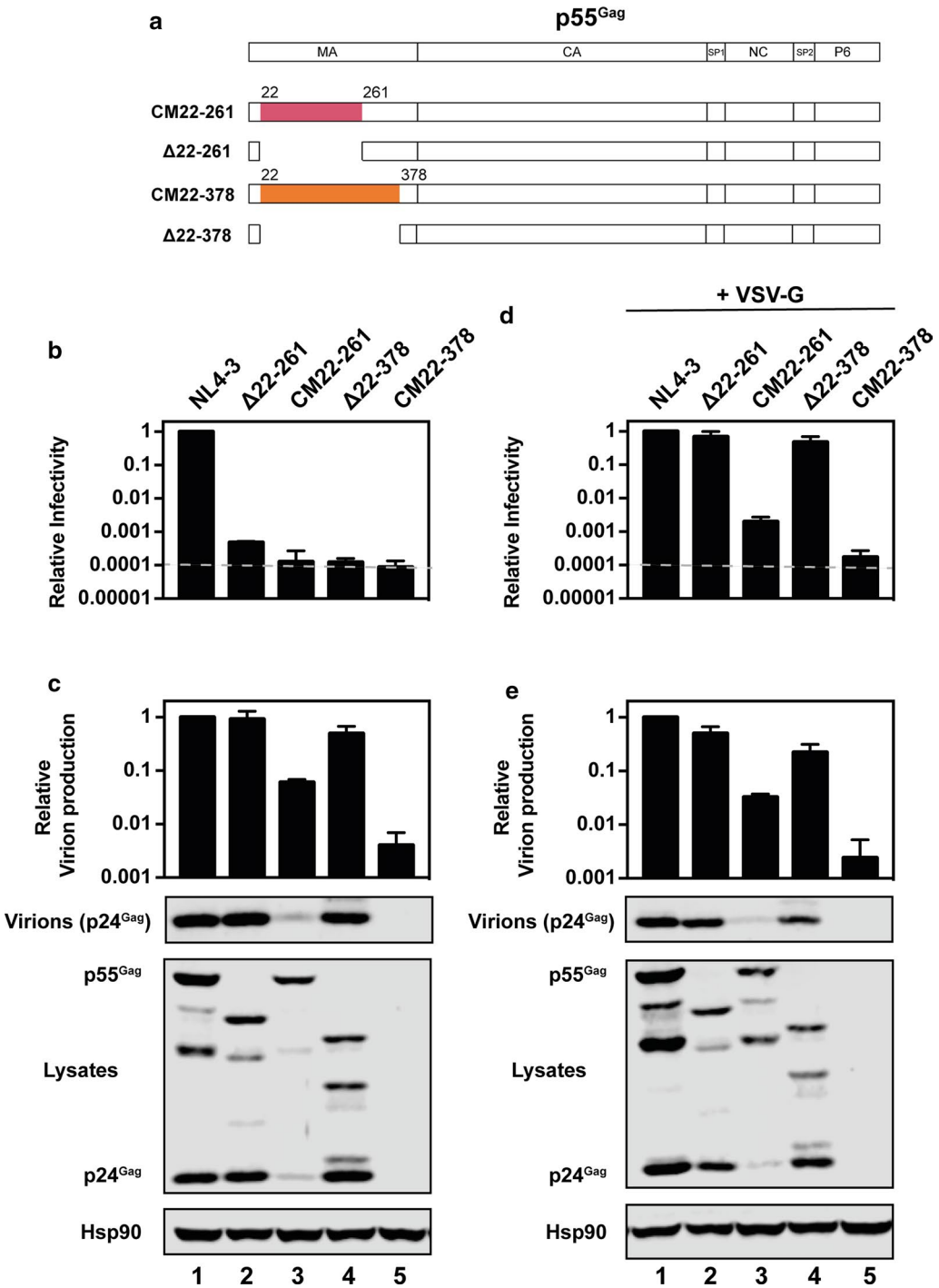


Fig. 4 Codon modification but not deletion of nucleotides 22-261 or 22-378 in *gag* inhibits infectious virus production. **a** Schematic representation of $p55^{Gag}$ in HIV-1_{NL4-3}, HIV-1 CM22-261, HIV-1 Δ22-261, HIV-1 CM22-378 or HIV-1 Δ22-378. **b**, **d** HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 CM22-261, pHIV-1 Δ22-261, pHIV-1 CM22-378 or pHIV-1 Δ22-378 and pGFP (**b**) or pVSV-G (**d**). The amount of infectious virus in the culture supernatants was measured in TZM-bl cells. The average RLU for HIV-1_{NL4-3} + GFP and HIV-1_{NL4-3} + VSV-G is 13,701,427 and 16,981,387, respectively. The dashed line represents 1000 RLU, which is approximately the limit of the assay for reproducible differential results. **c**, **e** Gag expression in the media and cell lysate were measured by quantitative western blotting. **b**–**e** The bar charts show the average of three independent experiments relative to HIV-1_{NL4-3}. Error bars represent standard deviation

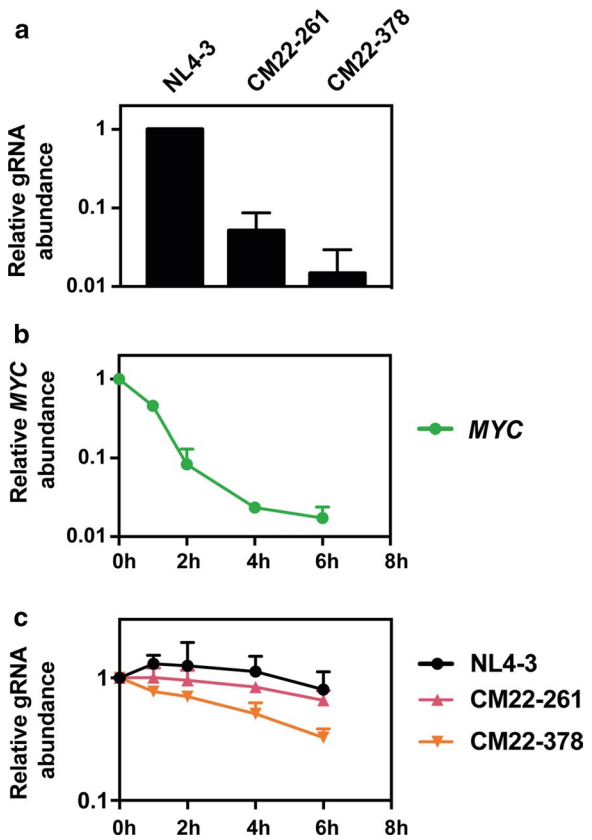


Fig. 5 Codon modification of nucleotides 22-261 and 22-378 in *gag* decreases gRNA abundance and stability. HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 CM22-261 or pHIV-1 CM22-378. **a** RNA was extracted from cell lysates at the 0 h timepoint and gRNA abundance was quantified by qRT-PCR. The bar charts show the average values of three independent experiments normalized to HIV-1_{NL4-3}. Error bars represent standard deviation. **b**, **c** Actinomycin D was added to inhibit RNA polymerase II transcription and the abundance of *MYC* mRNA (**b**) or gRNA (**c**) was measured at 0, 1, 2, 4 and 6 h post-addition. Each value is relative to the 0 h timepoint and is an average of three independent experiments. Error bars represent standard deviation

(Fig. 6) and inserted it into $pHIV_{NL4-3}$ to produce pHIV-1 CM22-261_{lowCpG}. Within nt 22-261 of *gag*, HIV-1 CM22-261_{lowCpG} has the same four CpG dinucleotides as wild type HIV-1 and 59 mutations, compared with HIV-1 CM 22-261 that has 22 CpG dinucleotides and 80 mutations (Fig. 6, Table 2). pHIV-1_{NL4-3}, pHIV-1 CM22-261 and pHIV-1 CM22-261_{lowCpG} were transfected into HeLa cells and single round infectivity assays were performed. In contrast to HIV-1 CM22-261, HIV-1 CM22-261_{lowCpG} infectivity, Gag expression and virion production was similar to HIV-1_{NL4-3} (Fig. 7a, b). We also cloned pHIV-1 CM22-378_{lowCpG}, which has four CpG dinucleotides and 79 mutations compared with the 30 CpG dinucleotides and 109 mutations in HIV-1 CM22-378 (Fig. 6, Table 2). In a single round infectivity assay, HIV-1 CM22-378_{lowCpG} also had similar levels of infectivity, intracellular Gag expression and virion production as HIV-1_{NL4-3} (Fig. 7c, d). While codon modification of nt 22-261 and 22-378 in *gag* substantially decreased the A-rich nucleotide bias of these regions, eliminating only the CpG dinucleotides in the codon modified sequence did not restore the A-rich bias for HIV-1 CM 22-261_{lowCpG} and HIV-1 CM22-378_{lowCpG} (Table 2). This supports the hypothesis that the decrease in infectivity for HIV-1 CM22-261 and HIV-1 CM22-378 is due to the introduced CpG dinucleotides and not changes in the A-rich nucleotide bias. In sum, changing the codons that introduced CpG dinucleotides in nt 22-261 or nt 22-378 back to the wild type HIV-1 codons while maintaining all of the other mutations in these regions restored infectious virus production.

Toll-like receptor 9 (TLR9) recognizes unmethylated CpG DNA motifs in endolysosomes within plasmacytoid dendritic cells, macrophages, and B cells [75]. Upon ligand binding, TLR9 signaling stimulates interferon alpha (IFN-α) production, which binds the interferon alpha and beta receptor (IFNAR) and induces interferon stimulated gene (ISG) expression via the JAK-STAT signaling pathway. To determine whether HeLa cells are responsive to CpG DNA or other TLR ligands, we tested a panel of ligands for TLR2, TLR3, TLR4, TLR5, TLR7, TLR8, TLR9 and TLR13 (Additional file 3). As positive controls, HeLa cells were infected with Sendai virus, which activates the cytoplasmic RNA sensors RIG-I and MDA5 (RIG-I-like receptors, RLR), or were treated with interferon beta (IFN-β), which also binds IFNAR. Both Sendai virus and IFN-β stimulated STAT1 phosphorylation and IFIT1 expression, which is an ISG. CpG DNA did not stimulate STAT1 phosphorylation or IFIT1 expression. The only TLR ligand that stimulated the cells was poly(I:C), which is structurally similar to double stranded RNA and can also signal via RLRs. Therefore, it is unlikely that the additional CpG dinucleotides in HIV-1 CM22-261 and HIV-1 CM22-378 inhibit viral replication via TLR9 or other TLRs in HeLa cells.

Increased abundance of CpG dinucleotides is sufficient for inhibition of viral replication

To determine if increasing the abundance of CpG dinucleotides is sufficient to inhibit HIV-1 replication, we synthesized *gag* sequences in which only the codons that introduced CpG dinucleotides in the codon modified sequence were changed (Fig. 6). These were inserted into $pHIV_{NL4-3}$ to produce pHIV-1 CpG22-165, pHIV-1 CpG22-261 and pHIV-1 CpG22-378. We then used a spreading infection assay in Jurkat cells to analyze the effect of the increased CpG abundance. 293T cells were transfected with each proviral construct to produce stocks of each virus and the abundance of $p24^{Gag}$ was

Table 1 HIV-1_{NL4-3} genomic RNA mononucleotide and dinucleotide frequencies

Base	#	Freq.	%	Obs./Exp.
Mononucleotide frequencies				
A	3281	0.358	35.8	1.43
C	1635	0.178	17.8	0.71
G	2216	0.242	24.2	0.97
T	2041	0.223	22.3	0.89
G + C	3851	0.420	42.0	0.84
A + T	5322	0.580	58.0	1.16
Total positions = 9173				
Base	**#**	**Freq.**	**%**	**Obs./Exp.**
Dinucleotide frequencies				
AA	1093	0.119	11.9	0.93
AC	522	0.057	5.7	0.89
AG	962	0.105	10.5	1.21
AT	703	0.077	7.7	0.96
CA	758	0.083	8.3	1.30
CC	371	0.040	4.0	1.27
CG	82	0.009	0.9	0.21
CT	424	0.046	4.6	1.17
GA	762	0.083	8.3	0.96
GC	424	0.046	4.6	1.07
GG	625	0.068	6.8	1.17
GT	405	0.044	4.4	0.82
TA	668	0.073	7.3	0.92
TC	318	0.035	3.5	0.87
TG	546	0.060	6.0	1.11
TT	509	0.055	5.5	1.12
Total positions = 9172				

Table 2 Changes in nucleotide composition and total number of mutations for codon modification of regions 22-261 and 22-378 in *gag*

Construct	% A	% C	% G	% T	Total UpA	Total CpG	Total number mutations relative to wild type
WT 22-261	38	17	23	22	23	4	
CM 22-261	18	36	32	15	3	22	80
CM 22-261_{lowCpG}	24	27	34	15	5	4	59
CpG 22-261	32	25	22	21	19	22	21
WT 22-378	42	17	24	18	29	4	
CM 22-378	22	35	32	12	3	30	109
CM 22-378_{lowCpG}	28	26	33	13	7	4	79
CpG 22-378	36	25	22	17	23	30	30

measured for each by ELISA. HIV-1 CpG22-165 and HIV-1 CpG22-261 produced similar amounts of virus to HIV-1_{NL4-3} while HIV-1 CpG22-378 virus production was decreased by ~ 70% (Fig. 8a). The viral inoculum was normalized so that Jurkat cells were infected with 25 ng of $p24^{Gag}$ for each virus and replication was monitored over ~ 2 weeks. There are an additional 11 CpG dinucleotides in HIV-1 CpG22-165 and its replication was substantially reduced, with > 90% less infectivity at Day 13 compared to HIV-1_{NL4-3} (Fig. 8b). HIV-1 CpG22-261 has an additional 18 CpG dinucleotides and HIV-1 CpG22-378 has an additional 26 CpG dinucleotides. Neither

Fig. 6 Schematic representation of nucleotides 1-420 in *gag* for HIV-1_{NL4-3}, HIV-1 CM22-261, HIV-1 CM22-165, HIV-1 CM166-261, HIV-1CM22-$261_{low-CpG}$, HIV-1 CM22-378 and HIV-1 CM22-378_{lowCpG}. Mutations that are the result of codon modification and do not introduce CpG dinucleotides are highlighted in turquoise. CpG dinucleotides present in HIV-1_{NL4-3} are highlighted in yellow and CpG dinucleotides introduced by codon modification are highlighted in red

of these viruses replicated in the Jurkat cells (Fig. 8b). Therefore, introducing CpG dinucleotides into HIV-1 *gag* is sufficient to strongly attenuate viral replication in Jurkat cells.

To analyze the steps of the HIV-1 life cycle that were inhibited by the CpG dinucleotides, HeLa cells were transfected with pHIV$_{NL4-3}$, pHIV-1 CpG22-165, pHIV-1 CpG22-261 or pHIV-1 CpG22-378. ~ 48 h later, the media and cell lysates were harvested for infectivity, protein or RNA analysis. In this single cycle assay, HIV-1 CpG22-378 infectivity was decreased ~ 99% (Fig. 9a). While there was no change in Gag expression or virion production (Fig. 9b), HIV-1 CpG22-378 gRNA was decreased ~ 40% in the cell lysate and ~ 80% in the media (Fig. 9c, d). HIV-1 CpG22-261 had a ~ 75% decrease in infectivity with no difference in Gag expression, virion production or gRNA abundance (Fig. 9a–d). There was no decrease in infectious virus production for HIV-1 CpG 22-165. Overall, introducing 26 CpG dinucleotides into HIV$_{NL4-3}$ decreased infectious virus production by ~ 99%, though this is a smaller attenuation than viruses containing CpG dinucleotides in the context of the codon modified sequence (Figs. 2, 4, 7).

One potential explanation for why HIV-1 containing CpG dinucleotides in a codon modified context produce less infectious virus in HeLa cells than HIV-1 in which

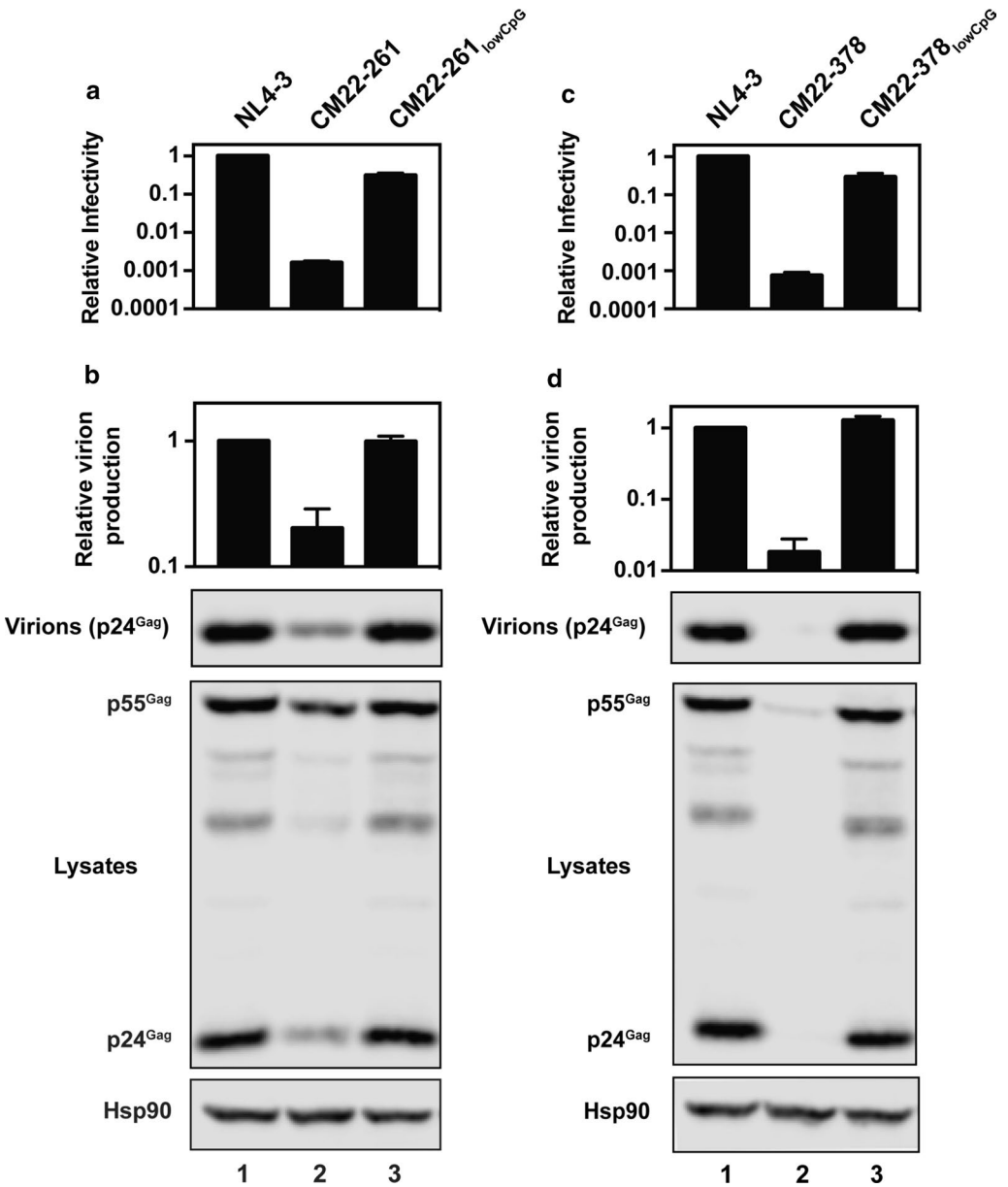


Fig. 7 Decreasing the CpG abundance within HIV-1 CM22-261 and HIV-1 CM22-378 restores infectious virus production. HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 CM22-261, pHIV-1 CM22-261_{lowCpG}, pHIV-1 CM22-378 and pHIV-1 CM22-378_{lowCpG}. The bar charts show the average of three independent experiments relative to HIV-1_{NL4-3}. Error bars represent standard deviation. **a**, **c** The amount of infectious virus present in the media was measured in TZM-bl cells. **b**, **d** Virion production and intracellular Gag expression was measured using quantitative western blotting

only CpG dinucleotides have been added is that a nucleic acid binding protein could bind the CpG dinucleotide to mediate antiviral activity. If a protein does directly bind the CpG dinucleotide, its binding site may encompass more than just the CpG and the surrounding nucleotides may affect binding. Therefore, the five nucleotides 5′ and 3′ of the introduced CpG in HIV-1 CM22-378 and HIV-1 CpG22-378 were used to generate a 12 nt sequence. These were aligned and conserved nucleotides were identified using WebLogo [67]. Interestingly, the nucleotides surrounding the CpG in the codon modified sequence are more G/C-rich than in the wild type HIV-1 sequence (Additional file 4A and B).

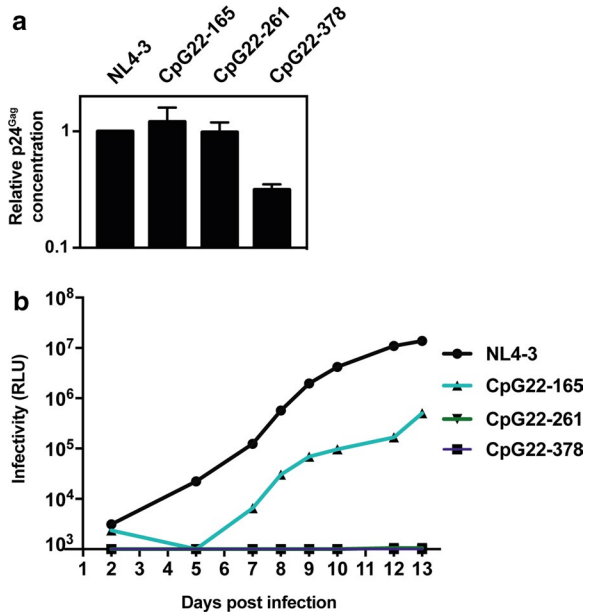


Fig. 8 Introduction of CpG dinucleotides into *gag* inhibits HIV-1 replication in Jurkat cells. **a** The amount of HIV-1 CA ($p24^{Gag}$) in supernatants from 293T cells transfected with $pHIV\text{-}1_{NL4\text{-}3}$, pHIV-1 CpG22-165, pHIV-1 CpG22-261 or pHIV-1 CpG22-378 were quantified by $p24^{Gag}$ ELISA. The bar chart is the average of three independent experiments normalized to $HIV\text{-}1_{NL4\text{-}3}$. Error bars represent standard deviation. **b** Jurkat cells were infected with 25 ng of $p24^{Gag}$ for each indicated virus. The amount of infectious virus present at each time point was measured in TZM-bl cells. This is representative of three independent experiments

Discussion

Herein, we show that introducing CpG dinucleotides into the HIV-1 genome inhibits viral replication. When only 11 CpG dinucleotides are inserted into *gag* in the context of the codon modified sequence (HIV-1 CM22-165), there is a large decrease in infectivity without a substantial loss of gRNA abundance, Gag expression or virion production (Figs. 2, 3). When 18 or 26 CpG dinucleotides are added (HIV-1 CM22-261 and HIV-1 CM22-378, respectively), the intracellular gRNA stability and abundance is decreased which leads to reduced Gag expression and virion production (Figs. 2, 3, 4, 5). In addition, there is a defect in the pre-integration steps of the viral life cycle that is apparent when equivalent numbers of genomes are used to infect target cells (Fig. 3c). Thus, manipulating the CpG abundance in *gag* can impart both producer and target cell defects in replication. Determining whether these deficiencies are underpinned by a common mechanism such as shared host factors, or whether they are distinct from each other, will be an important direction of our future work.

Remarkably, none of the proposed *cis*-acting elements in nt 22-378 of *gag* appear to be necessary for infectious virus production in HeLa cells. We demonstrated this by comparing viruses that have nt 22-261 or nt 22-378 codon modified or deleted (Fig. 4). While codon modification of these regions inhibited HIV-1 infectivity, deletion of the same region did not substantially impair infectivity in a single round assay. Furthermore, removing the introduced CpG dinucleotides in HIV-1 CM22-261 and HIV-1 CM22-378 while maintaining either 59 or 79 other nucleotide changes, respectively, almost completely restored HIV-1 infectious virus production (Fig. 7). This supports the observation by Reil et al. [60] that the globular head of MA can be deleted in the context of Env with a truncated cytoplasmic domain without substantially impairing viral replication in the MT4 T cell line or virion production in HeLa cells.

One of the *cis*-acting elements proposed to be in the *gag* region that we have mutated is the gRNA packaging signal. While a relatively short sequence sufficient for packaging heterologous transcripts into virions has been identified for rous sarcoma virus and murine leukemia virus, delineation of the HIV-1 RNA sequence that is sufficient for packaging has been more controversial [7, 8]. The core packaging signal for HIV-1 is in the 5′ UTR; however, approximately 300 nt of the 5′ end of *gag* has been proposed to improve viral titre [7, 8]. We did not codon modify the first 21 nt of *gag* because this region is under purifying selection [11, 12] and NMR structures have shown that it base pairs with the U5 region of the 5′ UTR to form the dimer promoting conformation of the gRNA that is packaged into virions [10, 76]. However, the relative importance of the sequence in *gag* beyond the first 21 nt for packaging the full-length HIV-1 gRNA, as opposed to heterologous transcripts, is unclear. Mapping Gag binding sites on the gRNA in living cells showed that the RNA elements most frequently crosslinked to Gag were in the 5′ UTR and the RRE [77]. In the context of the full-length virus, our data indicates that nt 22-378 in *gag* appear to have only a small effect on viral infectivity when this region is deleted or mutated without adding CpG dinucleotides (Figs. 4, 7). The requirement for this region for gRNA packaging may be different in the context of lentiviral vector genomes, which contain only a small portion of the HIV-1 gRNA [78].

In principle, codon usage changes in *gag* could affect gRNA translation [79]. However, the translation efficiency of a codon optimized *gag* mRNA is only ~ 1.6-fold higher than wild type *gag* mRNA that contains theoretically suboptimal codon usage [80]. Our data indicate that gRNA translation efficiency is not substantially affected by the changes in codon usage for HIV-1 CM22-261 since Gag expression (Fig. 2c) and intracellular gRNA abundance (Figs. 3c, 5a) were both decreased by ~ 90%. Changing the RNA sequence could also affect

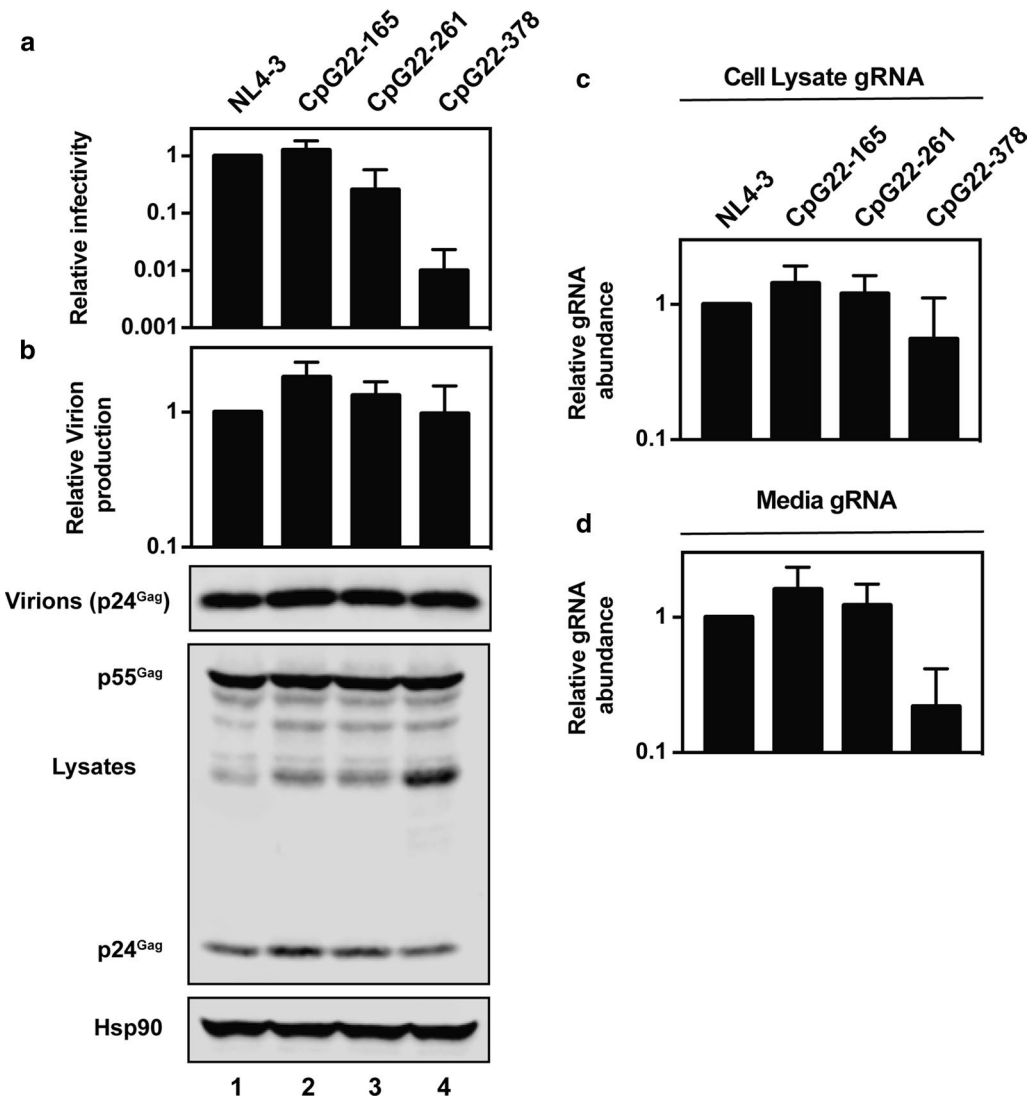


Fig. 9 Introduction of CpG dinucleotides into *gag* inhibits infectious virus production in HeLa cells. HeLa cells were transfected with pHIV-1_{NL4-3}, pHIV-1 CpG22-165, pHIV-1 CpG22-261 or pHIV-1 CpG22-378. **a** Culture supernatants were used to infect TZM-bl reporter cells to measure viral infectivity. The bar charts show the average values of four independent experiments normalized to the value obtained for HIV-1_{NL4-3}. **b** Gag expression in the media and cell lysate was determined by quantitative immunoblotting. The bar charts show the average of three independent experiments normalized to HIV-1_{NL4-3}. Error bars represent standard deviation. RNA was extracted from cell lysates and media (**c**, **d**) and gRNA abundance was quantified by qRT-PCR. **b**–**d** The bar charts show the average of three independent experiments normalized to HIV-1_{NL4-3}. Error bars represent standard deviation

the secondary or tertiary structure of the gRNA. While the amount of RNA structure in the MA region of *gag* is much lower than that of the 5′ UTR [32], we cannot exclude that the synonymous mutations in *gag* have not altered nearby structures. The known structures in the nucleotides that we have mutated are the IRES [29, 53, 81, 82] and the region that base pairs with the 3′ end of the genome [33, 35]. In the context of single cycle infectivity assays in HeLa cells (Fig. 4) and replication in MT4 cells [60], neither the IRES nor circularization of the HIV-1 gRNA appear to be necessary because the region containing these elements can be deleted. However, it is possible that the phenotype for mutating an RNA structure may not be the same as deleting it [83] and these structures could be necessary under conditions not tested in this study, such as cellular stress or the innate immune response [6].

Two previous reports have shown that introducing synonymous mutations into *gag* or *pol* attenuates HIV-1 replication [54, 84]. Martrus et al. [84] introduced codon pairs into *gag* or *pol* that are underrepresented relative to human mRNAs, which strongly inhibited viral

replication. However, codon pair bias in RNA viruses has recently been shown to be due CpG and UpA dinucleotide suppression and codon pair deoptimization increases CpG abundance [74, 85]. While not discussed in their study, the synonymous mutations that Martrus et al. introduced increased the number of CpG dinucleotides in *gag* from 15 to 118. Therefore, we hypothesize that the decrease in viral replication observed in this study is at least partially a result of increasing the CpG frequency instead of codon pair deoptimization. To analyze the role of the A-rich RNA sequence for HIV-1 replication, Keating et al. [54] codon modified regions within *gag* and *pol* by increasing the number of C- and G-rich codons, which also inhibited HIV-1 replication. phGag-Pol [86] was used as the source of codon modified sequence and this has a large increase in CpG dinucleotide abundance compared with the wild type sequence. These additional CpG dinucleotides could be responsible for viral attenuation instead of altering the A-rich codon bias. Supporting this hypothesis, Klaver et al. [87] used phylogeny-instructed mutagenesis to increase or decrease the A-rich codon bias in an ~ 500 nt region of *pol*. In this study, the CpG dinucleotide abundance was decreased by one in the A-Max mutant virus and only increased by four in the A-Min mutant virus. Substantially increasing or decreasing the number of A nucleotides did not affect HIV-1 replication, demonstrating that it is important to avoid introducing suppressed dinucleotides such as CpG in mutagenesis studies analyzing the functional relevance of nucleotide or codon bias.

While CpG dinucleotides appear to be under negative selection in HIV-1 [36, 40, 44, 55–57], the specific selection pressure has been unclear. There are at least four potential causes for the suppressed abundance of CpG dinucleotides in HIV-1. First, this could be due to a mutational bias caused by cytosines in a CpG context in the proviral DNA becoming methylated and then undergoing rapid spontaneous deamination. However, we have shown that increasing the abundance of CpG dinucleotides inhibits viral replication, even in a single round assay (Figs. 1, 2, 8, 9). Second, CpG methylation-induced transcriptional silencing could inhibit HIV-1 gene expression [36, 40, 44, 55–57, 88]. However, this cannot cause the inhibition in the single cycle infectivity assays because the transfected proviral plasmids were amplified in bacteria, which does not methylate CpG dinucleotides. Importantly, when an unmethylated plasmid is transfected into mammalian cells, the CpG dinucleotides are not methylated [89–91]. Third, unmethylated CpGs in the DNA could be recognized by TLR9 [88, 92]. This pattern recognition receptor is expressed in plasmacytoid dendritic cells, macrophages, and B cells [75] and therefore is unlikely to be expressed in HeLa cells. We confirmed that HeLa cells do not induce STAT1 phosphorylation or IFIT1 expression in response to CpG DNA (Additional file 3). In addition, IFN-α does not inhibit wild type HIV-1 infectious virus production in HeLa cells at most concentrations (10–1000 U/ml) and only has moderate inhibition at very high concentrations (10,000 U/ml) [93]. Therefore, the inhibition that we observe in response to introducing CpG dinucleotides into HIV-1 is unlikely to be due to the production of type I interferon.

The fourth possibility is that CpG dinucleotides in the viral RNA could induce an antiviral response that restricts HIV-1 replication. The frequency of CpG dinucleotides is suppressed in many RNA viruses that do not have a DNA intermediate [57, 92, 94–96], indicating that CpG DNA methylation or activation of TLR9 cannot be responsible for CpG suppression in all viruses. Introduction of CpG dinucleotides into picornaviruses or influenza A virus inhibits viral replication [71–74]. It is unclear how CpG dinucleotides restrict RNA virus replication but, for the picornavirus echovirus 7, it is not due to stimulating the interferon pathway, PKR, conventional pattern recognition receptors or altering the translation efficiency of viral proteins [71, 74]. It has been hypothesized that there is an innate immune sensor that detects CpG dinucleotides in viral RNA and leads to the inhibition of viral replication, though the molecular details are unknown [71, 73, 74, 85, 92, 97–99]. We favor the hypothesis that the proposed active restriction pathway targeting CpG dinucleotides in RNA viruses inhibits HIV-1 with an increased CpG abundance. When 26 CpG dinucleotides were added within nt 22-378 of *gag* (HIV-1 CpG22-378), which is < 5% of the HIV-1 gRNA, viral replication was inhibited in Jurkat cells (Fig. 8) and infectivity was decreased by ~ 99% in a single round assay (Fig. 9). This indicates that the CpG dinucleotides induce a potent restriction.

While this manuscript was under review, Takata et al. [100] reported that introducing CpG dinucleotides into *env* inhibited HIV-1 replication by decreasing the abundance of cytoplasmic gRNA, Gag expression, Env expression and infectious virus production. They also demonstrated that depleting the cellular RNA binding protein ZAP rescues replication of HIV-1 with increased CpG abundance and ZAP directly binds HIV-1 RNA regions containing CpG dinucleotides. This indicates that ZAP restricts replication of HIV-1 containing increased CpG abundance, though it is unclear how ZAP promotes viral RNA degradation. Interestingly, we and others have shown that Gag is efficiently expressed from mammalian expression vectors that encode codon-optimized *gag* or *gag-pol* cDNAs containing large numbers of CpG dinucleotides [61, 68, 86, 101, 102]. Therefore, it appears that

CpG dinucleotides in the context of full length HIV-1 is more deleterious for protein expression than CpGs in the context of mammalian expression vectors. How the specific context of CpG dinucleotides affects ZAP binding to RNA or modulates its activity will be an exciting area of future research.

Conclusions

The HIV-1 RNA sequence contains specific nucleotide features such as a low abundance of CpG dinucleotides. Our data shows that introducing CpG dinucleotides into HIV-1 inhibits viral replication by affecting multiple steps of the life cycle. This provides a functional explanation for why CpG dinucleotides are suppressed in HIV-1 and we speculate this dinucleotide is under negative selection to avoid an active restriction system that may require ZAP. Understanding how this restriction system inhibits replication of HIV-1 with increased CpG abundance may provide insight into how other RNA viruses, such as picornaviruses and influenza A virus, are attenuated when CpG dinucleotides are introduced.

Additional files

Additional file 1. Codon modification of nucleotides 22-261 in gag inhibits viral replication in SupT1 cells. SupT1 cells were infected with 10 ng of p24Gag for each indicated virus. The amount of infectious virus present at each time point was measured in TZM-bl cells. This is representative of three independent experiments.

Additional file 2. Codon modification of nucleotides 22-261 in gag decreases total HIV-1 RNA abundance. The RNA that was extracted from cell lysates as described for Fig. 3a was quantified for total HIV-1 RNA by qRT-PCR.

Additional file 3. CpG DNA does not stimulate STAT1 phosphorylation or ISG expression. HeLa cells were stimulated with TLR ligands for 5 h. TLR3 was targeted with 0.1, 1, 10 μg/ml polyIC, TLR7 with 0.3, 3, 30 nM Gardiquimod (Gard), TLR8/7 with 0.3, 3, 30 nM CL075, TLR7/8 with 0.01, 0.1, 1 μg/ml R848, TLR9 with 0.01, 0.1, 1 μg/ml CpG DNA, TLR13 with 0.01, 0.1, 1 μg/ml 23S ribosomal RNA (rRNA). TLR 4, 2 and 5 were targeted with 0.1 μg/ml LPS, Pam3Cys (P3C) or Flagellin (FliC), respectively. As controls for pattern recognition receptor signaling and JAK-STAT signaling, cells were infected with 50 HAU/ml Sendai virus (SeV) for 5 h or stimulated with 0.01 μg/ml IFN-β for 1 h, respectively. Activation of IFN signaling was monitored by western immunoblotting against phosphorylated STAT1 (pSTAT1) or expression of the ISG IFIT1. Actin was used as a loading control. (* Denotes the molecular weight marker).

Additional file 4. The CpG dinucleotide in the codon modified sequence is in a G/C-rich context. The sequence of the five nucleotides 5′ and 3′ to the CpG dinucleotides introduced into HIV-1 CM22-387 and HIV-1 CpG22-387 were aligned and a graphical representation of the sequence conservation was generated by WebLogo.

Authors' contributions

IAA designed and performed all of the HIV-1 experiments under the supervision of CMS. IAA also prepared the figures. CM aided in experimental design, cloning and training. LAG performed preliminary experiments analyzing codon modification on HIV-1 gene expression. CO analyzed the effect of TLR ligands on HeLa cells. The manuscript was written by CMS with all authors contributing to the text. All authors read and approved the final manuscript.

Acknowledgements

The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: TZM-bl from Dr. John C. Kappes, Dr. Xiaoyun Wu and Tranzyme Inc; HIV-1 p24 Hybridoma (183-H12-5C) from Dr. Bruce Chesebro. We thank Jonathan Sumner in Dr. Stuart Neil's lab for assistance in setting up the spreading infection assay and Professor Michael Malim for helpful discussions. We also thank Professor Juan Martin Serrano, Professor Michael Malim and Dr. Hendrik Huthoff for critically reading the manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

CMS was funded by Medical Research Council Grants MR/K000381/1 and MR/M019756/1. IAA and LAG were supported by the King's Bioscience Institute and the Guy's and St Thomas' Charity Prize Ph.D. Programme in Biomedical and Translational Science. CO was supported by a King's College London Prize Fellowship and a Sir Henry Dale Fellowship from the Royal Society and the Wellcome Trust (206200/Z/17/Z). This work was also supported by the Department of Health via a National Institute for Health Research Comprehensive Biomedical Research Centre award to Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust.

Publisher's Note

Received: 12 September 2017 Accepted: 1 November 2017
Published online: 09 November 2017

References

1. Leblanc J, Weil J, Beemon K. Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. Wiley Interdiscip Rev RNA. 2013;4:567–80.
2. Stoltzfus CM. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. Adv Virus Res. 2009;74:1–40.
3. Tazi J, Bakkour N, Marchand V, Ayadi L, Aboufirassi A, Branlant C. Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. FEBS J. 2010;277:867–76.
4. Sherrill-Mix S, Ocwieja KE, Bushman FD. Gene activity in primary T cells infected with HIV89.6: intron retention and induction of genomic repeats. Retrovirology. 2015;12:79.
5. Rojas-Araya B, Ohlmann T, Soto-Rifo R. Translational control of the HIV unspliced genomic RNA. Viruses. 2015;7:4326–51.
6. Hidalgo L, Swanson CM. Regulation of human immunodeficiency virus type 1 (HIV-1) mRNA translation. Biochem Soc Trans. 2017;45:353–64.
7. Kuzembayeva M, Dilley K, Sardo L, Hu WS. Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. Virology. 2014;454–455:362–70.
8. Lu K, Heng X, Summers MF. Structural determinants and mechanism of HIV-1 genome packaging. J Mol Biol. 2011;410:609–33.
9. Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, et al. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J Virol. 1986;59:284–91.

10. Keane SC, Summers MF. NMR studies of the structure and function of the HIV-1 5′-leader. Viruses. 2016;8:338.
11. Mayrose I, Stern A, Burdelova EO, Sabo Y, Laham-Karam N, Zamostiano R, et al. Synonymous site conservation in the HIV-1 genome. BMC Evol Biol. 2013;13:164.
12. Ngandu NK, Scheffler K, Moore P, Woodman Z, Martin D, Seoighe C. Extensive purifying selection acting on synonymous sites in HIV-1 Group M sequences. Virol J. 2008;5:160.
13. Brierley I, Dos Ramos FJ. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. Virus Res. 2006;119:29–42.
14. Pollard VW, Malim MH. The HIV-1 Rev protein. Annu Rev Microbiol. 1998;52:491–532.
15. Le Grice SF. Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise. J Biol Chem. 2012;287:40850–7.
16. Lavender CA, Gorelick RJ, Weeks KM. Structure-based alignment and consensus secondary structures for three HIV-related RNA genomes. PLoS Comput Biol. 2015;11:e1004230.
17. Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM, et al. Comparison of SIV and HIV-1 genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. PLoS Pathog. 2013;9:e1003294.
18. Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat Methods. 2014;11:959–65.
19. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009;460:711–6.
20. Le Grice SF. Targeting the HIV RNA genome: high-hanging fruit only needs a longer ladder. Curr Top Microbiol Immunol. 2015;389:147–69.
21. Mayrose I, Doron-Faigenboim A, Bacharach E, Pupko T. Towards realistic codon models: among site variability and dependency of synonymous and non-synonymous rates. Bioinformatics. 2007;23:i319–27.
22. Sanjuan R, Borderia AV. Interplay between RNA structure and protein evolution in HIV-1. Mol Biol Evol. 2011;28:1333–8.
23. Simon-Loriere E, Martin DP, Weeks KM, Negroni M. RNA structures facilitate recombination-mediated gene swapping in HIV-1. J Virol. 2010;84:12675–82.
24. Snoeck J, Fellay J, Bartha I, Douek DC, Telenti A. Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints. Retrovirology. 2011;8:87.
25. Freed EO. HIV-1 assembly, release and maturation. Nat Rev Microbiol. 2015;13:484–96.
26. Najera I, Krieg M, Karn J. Synergistic stimulation of HIV-1 rev-dependent export of unspliced mRNA to the cytoplasm by hnRNP A1. J Mol Biol. 1999;285:1951–64.
27. Schaub MC, Lopez SR, Caputi M. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. J Biol Chem. 2007;282:13617–26.
28. Asang C, Erkelenz S, Schaal H. The HIV-1 major splice donor D1 is activated by splicing enhancer elements within the leader region and the p17-inhibitory sequence. Virology. 2012;432:133–45.
29. Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF. The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. J Virol. 2001;75:181–91.
30. Schwartz S, Felber BK, Pavlakis GN. Distinct RNA sequences in the gag region of human immunodeficiency virus type 1 decrease RNA stability and inhibit expression in the absence of Rev protein. J Virol. 1992;66:150–9.
31. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. In vitro evidence for a long range pseudoknot in the 5′-untranslated and matrix coding regions of HIV-1 genomic RNA. J Biol Chem. 2002;277:5995–6004.
32. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol. 2008;6:e96.
33. Ooms M, Abbink TE, Pham C, Berkhout B. Circularization of the HIV-1 RNA genome. Nucleic Acids Res. 2007;35:5253–61.
34. Damgaard CK, Andersen ES, Knudsen B, Gorodkin J, Kjems J. RNA interactions in the 5′ region of the HIV-1 genome. J Mol Biol. 2004;336(2):369–79.
35. Beerens N, Kjems J. Circularization of the HIV-1 genome facilitates strand transfer during reverse transcription. RNA. 2010;16:1226–35.
36. Bronson EC, Anderson JN. Nucleotide composition as a driving force in the evolution of retroviruses. J Mol Evol. 1994;38:506–32.
37. Sharp PM. What can AIDS virus codon usage tell us? Nature. 1986;324:114.
38. van Hemert FJ, Berkhout B. The tendency of lentiviral open reading frames to become A-rich: constraints imposed by viral genome organization and cellular tRNA availability. J Mol Evol. 1995;41:132–40.
39. Kypr J, Mrazek J. Unusual codon usage of HIV. Nature. 1987;327:20.
40. Kypr J, Mrazek J, Reich J. Nucleotide composition bias and CpG dinucleotide content in the genomes of HIV and HTLV 1/2. Biochim Biophys Acta. 1989;1009:280–2.
41. Grantham P, Perrin P. AIDS virus and HTLV-I differ in codon choices. Nature. 1986;319:727–8.
42. van der Kuyl AC, Berkhout B. The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. Retrovirology. 2012;9:92.
43. Berkhout B, van Hemert FJ. The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins. Nucleic Acids Res. 1994;22:1705–11.
44. Berkhout B, Grigoriev A, Bakker M, Lukashov VV. Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. AIDS Res Hum Retroviruses. 2002;18:133–41.
45. Vartanian JP, Plikat U, Henry M, Mahieux R, Guillemot L, Meyerhans A, et al. HIV genetic variation is directed and restricted by DNA precursor availability. J Mol Biol. 1997;270:139–51.
46. Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol. 1995;69:5087–94.
47. Martinez MA, Vartanian JP, Wain-Hobson S. Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. Proc Natl Acad Sci USA. 1994;91:11787–91.
48. Vartanian JP, Meyerhans A, Sala M, Wain-Hobson S. G–>A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription. Proc Natl Acad Sci USA. 1994;91:3092–6.
49. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA deamination mediates innate immunity to retroviral infection. Cell. 2003;113:803–9.
50. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature. 2003;424:94–8.
51. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature. 2003;424:99–103.
52. Kim EY, Lorenzo-Redondo R, Little SJ, Chung YS, Phalora PK, Maljkovic Berry I, et al. Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection. PLoS Pathog. 2014;10:e1004281.
53. Deforges J, de Breyne S, Ameur M, Ulryck N, Chamond N, Saadi A, et al. Two ribosome recruitment sites direct multiple translation events within HIV1 Gag open reading frame. Nucleic Acids Res. 2017;45:7382–400.
54. Keating CP, Hill MK, Hawkes DJ, Smyth RP, Isel C, Le SY, et al. The A-rich RNA sequences of HIV-1 pol are important for the synthesis of viral cDNA. Nucleic Acids Res. 2009;37:945–56.
55. Ohno S, Yomo T. Various regulatory sequences are deprived of their uniqueness by the universal rule of TA/CG deficiency and TG/CT excess. Proc Natl Acad Sci USA. 1990;87:1218–22.
56. Shpaer EG, Mullins JI. Selection against CpG dinucleotides in lentiviral genes: a possible role of methylation in regulation of viral expression. Nucleic Acids Res. 1990;18:5793–7.
57. Karlin S, Doerfler W, Cardon LR. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? J Virol. 1994;68:2889–97.

58. Wasson MK, Borkakoti J, Kumar A, Biswas B, Vivekanandan P. The CpG dinucleotide content of the HIV-1 envelope gene may predict disease progression. Sci Rep. 2017;7:8162.
59. Lee JS, Gray J, Mulligan R. Packaging cells comprising codon-optimized gagpol sequences and lacking lentiviral accessory proteins. Google Patents; 2006.
60. Reil H, Bukovsky AA, Gelderblom HR, Gottlinger HG. Efficient HIV-1 replication can occur in the absence of the viral matrix protein. EMBO J. 1998;17:2699–708.
61. Swanson CM, Sherer NM, Malim MH. SRp40 and SRp55 promote the translation of unspliced human immunodeficiency virus type 1 RNA. J Virol. 2010;84:6748–59.
62. Fouchier RA, Meyer BE, Simon JH, Fischer U, Malim MH. HIV-1 infection of non-dividing cells: evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import. EMBO J. 1997;16:4531–9.
63. Derdeyn CA, Decker JM, Sfakianos JN, Wu X, O'Brien WA, Ratner L, et al. Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. J Virol. 2000;74:8358–67.
64. Wei X, Decker JM, Liu H, Zhang Z, Arani RB, Kilby JM, et al. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. Antimicrob Agents Chemother. 2002;46:1896–905.
65. Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. J Virol. 1998;72:2855–64.
66. Chesebro B, Wehrly K, Nishio J, Perryman S. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. J Virol. 1992;66:6547–54.
67. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.
68. Mostoslavsky G, Kotton DN, Fabian AJ, Gray JT, Lee JS, Mulligan RC. Efficiency of transduction of highly purified murine hematopoietic stem cells by lentiviral and oncoretroviral vectors under conditions of minimal in vitro manipulation. Mol Ther. 2005;11:932–40.
69. Mammano F, Kondo E, Sodroski J, Bukovsky A, Gottlinger HG. Rescue of human immunodeficiency virus type 1 matrix protein mutants by envelope glycoproteins with short cytoplasmic domains. J Virol. 1995;69:3824–30.
70. Wisdom R, Lee W. The protein-coding region of c-myc mRNA contains a sequence that specifies rapid mRNA turnover and induction by protein synthesis inhibitors. Genes Dev. 1991;5:232–43.
71. Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P. 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. 2014;42:4527–45.
72. Burns CC, Campagnoli R, Shaw J, Vincent A, Jorba J, Kew O. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. J Virol. 2009;83:9957–69.
73. Gaunt E, Wise HM, Zhang H, Lee LN, Atkinson NJ, Nicol MQ, et al. Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. Elife. 2016;5:e12735.
74. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. Elife. 2014;3:e04531.
75. Pandey S, Kawai T, Akira S. Microbial sensing by Toll-like receptors and intracellular nucleic acid sensors. Cold Spring Harb Perspect Biol. 2014;7:a016246.
76. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, et al. NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. Science. 2011;334:242–5.
77. Kutluay SB, Zang T, Blanco-Melo D, Powell C, Jannain D, Errando M, et al. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. Cell. 2014;159:1096–109.
78. Naldini L, Blomer U, Gage FH, Trono D, Verma IM. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. Proc Natl Acad Sci USA. 1996;93:11382–8.
79. Brule CE, Grayhack EJ. Synonymous codons: choose Wisely for expression. Trends Genet. 2017;33:283–97.
80. Ngumbela KC, Ryan KP, Sivamurthy R, Brockman MA, Gandhi RT, Bhardwaj N, et al. Quantitative effect of suboptimal codon usage on translational efficiency of mRNA encoding HIV-1 gag in intact T cells. PLoS ONE. 2008;3:e2356.
81. Locker N, Chamond N, Sargueil B. A conserved structure within the HIV gag open reading frame that controls translation initiation directly recruits the 40S subunit and eIF3. Nucleic Acids Res. 2011;39:2367–77.
82. Weill L, James L, Ulryck N, Chamond N, Herbreteau CH, Ohlmann T, et al. A new type of IRES within gag coding region recruits three initiation complexes on HIV-2 genomic RNA. Nucleic Acids Res. 2010;38:1367–81.
83. Das AT, Vrolijk MM, Harwig A, Berkhout B. Opening of the TAR hairpin in the HIV-1 genome causes aberrant RNA dimerization and packaging. Retrovirology. 2012;9:59.
84. Martrus G, Nevot M, Andres C, Clotet B, Martinez MA. Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. Retrovirology. 2013;10:78.
85. Kunec D, Osterrieder N. Codon pair bias is a direct consequence of dinucleotide bias. Cell Rep. 2016;14:55–67.
86. Huang Y, Kong WP, Nabel GJ. Human immunodeficiency virus type 1-specific immunity after genetic immunization is enhanced by modification of Gag and Pol expression. J Virol. 2001;75:4947–51.
87. Klaver B, van der Velden Y, van Hemert F, van der Kuyl AC, Berkhout B. HIV-1 tolerates changes in A-count in a small segment of the pol gene. Retrovirology. 2017;14:43.
88. Alinejad-Rokny H, Anwar F, Waters SA, Davenport MP, Ebrahimi D. Source of CpG depletion in the HIV-1 genome. Mol Biol Evol. 2016;33(12):3205–12.
89. Kass SU, Goddard JP, Adams RL. Inactive chromatin spreads from a focus of methylation. Mol Cell Biol. 1993;13:7372–9.
90. Hsieh CL. Dependence of transcriptional repression on CpG methylation density. Mol Cell Biol. 1994;14:5487–94.
91. Hsieh CL. Stability of patch methylation and its impact in regions of transcriptional initiation and elongation. Mol Cell Biol. 1997;17:5897–904.
92. Greenbaum BD, Levine AJ, Bhanot G, Rabadan R. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS Pathog. 2008;4:e1000079.
93. Neil SJ, Sandrin V, Sundquist WI, Bieniasz PD. An interferon-alpha-induced tethering mechanism inhibits HIV-1 and Ebola virus particle release but is counteracted by the HIV-1 Vpu protein. Cell Host Microb. 2007;2:193–203.
94. Simmonds P, Xia W, Baillie JK, McKinnon K. Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla–selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. BMC Genom. 2013;14:610.
95. Rima BK, McFerran NV. Dinucleotide and stop codon frequencies in single-stranded RNA viruses. J Gen Virol. 1997;78:2859–70.
96. Auewarakul P. Composition bias and genome polarity of RNA viruses. Virus Res. 2005;109:33–7.
97. Vabret N, Bhardwaj N, Greenbaum BD. Sequence-specific sensing of nucleic acids. Trends Immunol. 2017;38:53–65.
98. Sugiyama T, Gursel M, Takeshita F, Coban C, Conover J, Kaisho T, et al. CpG RNA: identification of novel single-stranded RNA that stimulates human CD14 + CD11c+ monocytes. J Immunol. 2005;174:2273–9.
99. Tanne A, Muniz LR, Puzio-Kuter A, Leonova KI, Gudkov AV, Ting DT, et al. Distinguishing the immunostimulatory properties of noncoding RNAs expressed in cancer cells. Proc Natl Acad Sci USA. 2015;112:15154–9.
100. Takata MA, Goncalves-Carneiro D, Zang TM, Soll SJ, York A, Blanco-Melo D, et al. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. Nature. 2017;550:124–7.
101. Kotsopoulou E, Kim VN, Kingsman AJ, Kingsman SM, Mitrophanous KA. A Rev-independent human immunodeficiency virus type 1 (HIV-1)-based vector that exploits a codon-optimized HIV-1 gag-pol gene. J Virol. 2000;74:4839–52.
102. Swanson CM, Puffer BA, Ahmad KM, Doms RW, Malim MH. Retroviral mRNA nuclear export elements regulate protein function and virion assembly. EMBO J. 2004;23:2632–40.