2019

Non-Coding RNA Features Critical to the Replication of HIV-1

Matthew A. Takata

Non-coding RNA features critical to the replication of HIV-1

A thesis presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

By

Matthew A. Takata

June 2019

Non-coding RNA features critical to the replication of HIV-1

Matthew A. Takata, PhD.

The Rockefeller University 2019

The HIV-1 genome contains RNA sequences and structures that control many aspects of viral replication including, but not limited to transcription, splicing, nuclear export, translation, packaging and reverse transcription. Despite this extensive existing catalogue of RNA sequences that are critical to its replication, chemical probing and targeting mutagenesis studies suggest that the HIV-1 genome may contain many more RNA elements of unknown important function. To determine whether there are additional, undiscovered cis-acting RNA elements in the HIV-1 genome that are important for viral replication, we conducted a global synonymous mutagenesis experiment.

Sixteen mutant proviruses containing clusters of ~50 to ~200 synonymous mutations covering nearly the entire HIV-1 protein coding sequence were designed and synthesized. Analyses of these mutant viruses resulted in their division into three phenotypic groups. Group 1 mutants exhibited near wild-type replication, Group 2 mutants exhibited replication defects accompanied by perturbed RNA splicing, and Group 3 mutants had replication defects in the absence of obvious splicing perturbation. The three phenotypes were caused by mutations that exhibited a clear regional bias in their distribution along the viral genome, and those that caused replication defects all caused reductions in the level of unspliced RNA. We characterized in detail the underlying defects for Group 2 mutants. Second-site revertants that enabled viral replication could be derived for Group 2 mutants, and generally contained point mutations that reduced the utilization of proximal

splice sites. Mapping of the changes responsible for splicing perturbations in Group 2 viruses revealed the presence of several RNA sequences that apparently suppressed the use of cryptic or canonical splice sites. Some sequences that affected splicing were diffusely distributed, while others could be mapped to discrete elements, proximal or distal to the affected splice sites. This data from the Group 2 mutants indicates complex negative regulation of HIV-1 splicing by RNA elements in various regions of the HIV-1 genome that enable balanced splicing and viral replication.

In silico analysis of the Group 3 mutants revealed that our mutagenesis had significantly increased the frequency of CG dinucleotides in sections of the viral genome to that of random sequence. This is important due to the remarkable CG suppression in both the HIV-1 and human genomes, and we had therefore disrupted the dinucleotide congruence that exists between HIV-1 and the genome of its host. We recoded these mutants to selectively remove either only the CG dinucleotides or only remove the mutations that did not encode a CG dinucleotide. Analysis of these mutants clearly demonstrated that the addition of CG dinucleotides were the causative mutations entirely responsible for the observed replication defects. qPCR analysis and smFISH microscopy revealed that the addition of CG dinucleotides to HIV-1 resulted in a depletion of the cytoplasmic mRNA molecules where the CG-dinucleotides were encoded as exons. A targeted siRNA screen for proteins that destabilize cytoplasmic RNA identified the Zinc-finger Antiviral Protein (ZAP) as responsible for the restriction of the CG-high HIV-1, specifically by targeting CG-high viral RNA. CLIP-Seq experiments demonstrate that ZAP binds directly to CG dinucleotides in both cellular and viral RNA. Collectively these studies implicate ZAP as

a cellular protein that can recognize CG-high viral RNA and is possibly a cellular mechanism for determining self from non-self RNA based on the CG composition.

TRIM25 has previously been identified as a cofactor for two cytosolic RNA binding proteins that have antiviral functions, RIG-I where it is an essential cofactor, and ZAP where it functions as an enhancing cofactor. The mechanism by which TRIM25 enhances the antiviral activity of ZAP currently remains unclear. Through CLIP-Seq experiments in cells knocked out for TRIM25, we determined that ZAP does not require TRIM25 to recognize CG-high RNA. Using full length mutants of TRIM25 that are deficient for either RNA binding, E3 ligase activity, or formation of higher order multimers, our data suggest that the key biological activity required for TRIM25 to enhance ZAP is the formation of higher order multimers. Analyzing the replication of CG-high HIV-1 in different cell lines indicates that ZAP is not equally potent across all cell lines. The degree of potency ZAP possess against CG-high HIV-1 does not correlate with TRIM25 expression, suggesting the possibility of an additional ZAP cofactor that is heterogeneously expressed in varying cell lines. siRNA screens have been used in an attempt to identify a yet undiscovered cofactor, but so far these experiments have not yielded any such factor.

Dedicated to Barbara and Dennis Takata.

Acknowledgements

First and foremost, I would like to thank Paul Bieniasz. He has been a thoughtful mentor who is responsible for training me into a molecular retrovirologist from a marine biologist. He possesses an incredible scientific acumen and analytical capabilities that allow him to find coherence out of confusion. Working with him has had a profound impact on how I think about and structure problems; something I will carry with me for the rest of my life. Sohail Tavazoie and Charlie Rice were superb members of my faculty advisory committee. They whipped me into shape to make sure I abide by the proper administrative deadlines. More importantly though, they supported me as I pursued a risky PhD project trying to characterize a perplexing group of mutants.

I would also like to thank Melissa Kane. From my first day in the lab, Melissa has graciously taught me nearly everything I know at the bench. Not only this, but from her I learned a practical framework that allowed me to acquire the rest of skills and techniques I needed for the completion of this PhD. More than this, she is a dear friend who always encouraged me to keep pressing on even when science and life are at their most frustrating. I thank Steven Soll for the work he did to design and initially characterize the synonymous mutagenesis in the HIV-1 genome. The whole work of my PhD was made possible because of his labor.

My two scientific mentors as an undergraduate, John Stegeman and Lisanne Winslow were instrumental in leading me to Rockefeller University as a PhD student. Both of them

saw something in me beyond my sparse resume and took a chance investing in my career as a scientist. Lisanne helped me to see the glory in science. John taught me to fundamentals of working as a scientist. Both were absolutely critical to my success.

My friends in the city and beyond helped support me along the way in the ups and downs of life. While there are too many to detail, two are of particular importance. I'm grateful for my years at Rockefeller eating Chipotle, Texas burgers and General Tso's chicken with Jaime Medina-Manresa. He has always been a faithful friend with whom I have bountiful memories within the city and abroad. This dissertation could have been filled with times Josh Lagan has been there for me and memories we've shared traveling the world. I couldn't have kept afloat and kept going without his unwavering friendship.

Finally I would like to thank my mom and dad. There is nothing I could have done to deserve such wonderful parents as them. They have been there for me through every valley and mountain top. It is their belief in me and support of me that has brought me to the completion of my PhD. It is their love and affection for me that colors the beauty of this accomplishment and makes it so sweet.

Disclaimer

Dr. Steven Soll and Dr. Paul Bieniasz originally conceived of the idea to conduct a synonymous mutagenesis of the HIV-1 genome, and Dr. Steven Soll designed the mutants and conducted the initial characterization. The next generation sequencing analysis of the splicing mutants was performed and analyzed by Dr. Ann Emery under the supervision of Dr. Ronald Swanstrom. The smFISH experiments were performed by Trinity Zang. Dr. Daniel Goçalves-Carneiro made the luciferase constructs with different viral sequences as a 3' UTR and analyzed the CG odds ratio in the human ORFs, 3' UTR and ZAP CLIP clusters. Dr. Owen Pornillos provided us with the mutants in TRIM25. Portions of Chapters 2, 3, 4, and 5 are adapted from manuscripts that were co-written with Dr. Paul Bieniasz.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

-sssDNA	Minus-strand strong stop DNA
+sssDNA	Plus-strand strong stop DNA
4SU	4-Thiouridine
5’-OH	5'-hydroxyl
5’ppp	5’-triphosphorylated
APOBEC3	Apolipoprotein B editing catalytic subunit-like 3
ARD	Argenine rich domain
BDV	Borna disease virus
bp	Base pairs
C-terminus	Carboxyl terminus
CA	Capsid protein
CC	Coiled-coil
CCR	C-C chemokine receptor
CDK	cyclin-dependent kinase
cDNA	Complimentary DNA
CLIP-seq	Crosslinking immunoprecipitation
CsA	CyclosporinA
CXCR	C-X-C chemokine receptor
cypA	Cyclophillin A
DAPI	4',6-diamidino-2-phenylindole
DDX17	p72 DEAD box RNA helicase
DHX30	DEXH-Box protein
DI	doxycycline inducible
dNTPs	Deoxynucleoside triphosphates
dpi	Days post infection
dpt	Days post transfection
dsDNA	Double stranded DNA
EBOV	Ebola virus
ESE	exonic splicing enhancers
ESS	exonic splicing silencers
FACS	Fluorescence-activated cell sorting
fluc	Firefly Luciferase
GFP	Green fluorescent protein
GPI	glycophosphatidylinositol
HA-tag	Human influenza hemagglutinin derived tag
HCV	Hepatitis C Virus
HIV	Human immunodeficiency virus
hnRNP	heterogenous nuclear RNP

HSV-1	Herpes Simplex Virus 1
HTNV	hantavirus
IAV	Influenza A virus
IFN	Interferon
IN	Integrase
ISE	intronic splicing enhancers
ISG	Interferon stimulated gene
ISS	intronic splicing silencers
IU	Infectious units
JEV	Japanese Encephalitis Virus
KB	Kilobases
kDa	Kilodaltons
KO	knockout
LGP2	laboratory of genetics and physiology 2
LTR	Long terminal repeats
MA	Matrix
MDA5	melanoma differentiation association gene 5
MFI	Mean fluorescent intensity
MLV	Murine leukemia virus
MOIT	Multiplicity of infection
mRNA	Messenger RNA
NC	Nucleocapsid
NMR	Nuclear Magnetic Resonance
NP	Nucleoprotein
NPC	Nuclear Pore Complex
NS1	non-structural protein 1
nt	Nucleotide
Nup	nucleoporin
ORF	Open reading frame
PAMPS	pathogen associated molecular patterns
PBS	Primer binding site
PCR	Polymerase chain reaction
PIC	Pre-integration complex
PPT	Polypurine tract
PR	Protease
PTM	Post translational modification
RER	rough endoplasmic reticulum
RLR	RIG-I like receptors
RNA-Seq	RNA sequencing
RNAi	RNA interference

RNAP II	RNA polymerase II
RRE	Rev response element
RRM	RNA recognition motif
RRV	Ross River virus
RSE	RNA stabilizing element
RSV	Rous Sarcoma Virus
RT	Reverse transcriptase
RT-qPCR	Reverse transcription quantitative PCR
SA	Splice acceptor
SD	Spice donor
SELEX	systematic evolution of ligands by exponential enrichment
SFV	Semliki Forest virus
shRNA	short hairpin RNA
SINV	Sindbis virus
siRNA	short interfering RNA
SIV	Simian immunodeficiency virus
smFISH	small molecule fluorescent in situ hybridization
snRNP	small nuclear ribonucleoprotin particle
SR-Protein	Serine-Argenine Protein
ss	Splice site
SU	Surface protein
SX	Short exon
TAF	TBP-associated factor
TAR	Transactivating response
TCR	T cell receptor
TF	transcription factor
TM	Transmembrane protein or domain
TNPO3	transportin 3
TRAF	TNF receptor –associated factor
TRIM	Tripartite motif
U2AF	U2 auxiliary factor
UTR	Un-translated region
VEEV	Venezuelan equine encephalitis virus
vgRNA	Viral genomic RNA
VSV	Vesicular stomatitis virus
VSV-g	Vesicular stomatitis virus glycoprotein
WT	Wild-type
YF	Yellow Fever
ZAP	Zinc-finger Antiviral Protein

Chapter 1. Introduction

Biology is the study of all areas pertaining to bacteria, archaea and eukaryotes, or the tree of life. Viruses, because they are not considered to be alive, do not occupy a branch in the tree of life, yet have unquestionably been an important area of study within biology. Whether or not viruses should be considering living, they have significantly shaped and impacted every area of life, as evidenced by the fact that examples of viruses have been found to infect each domain of life (Koonin et al., 2006). If they were considered alive, they would make up the most abundant life form on earth. It is estimated that there are 4.8×10^{31} phages on earth, while prokaryotes, the most abundant cellular organisms, only total 4.1×10^{30} cells (Cobian Guemes et al., 2016). In addition to being the most abundant biological entities, they also pose one of the greatest threats to life, capable of killing single cellular prokaryotes and organisms as complex as humans which are composed of trillions of cells. A brief survey of the last decade reveals many different viral pandemics that have plagued human life in the world ranging from swine flu (H1N1), Zika virus, Ebola virus, West Nile Virus, and chikungunya. For the last 40 years, illnesses related to Human Immunodeficiency Viruses (HIV) have persisted as a global public health concern, resulting in over 30 million human deaths globally and nearly one million deaths still occurring annually.

Human Immunodeficiency Viruses: structure and replication

HIV-1 is a member of the *Retroviridae* family of viruses, commonly referred to as retroviruses. The two defining features of retroviruses are (i.) the synthesis of a double

stranded DNA intermediate from a positive strand RNA genome followed by (ii.) the integration of the newly synthesized genome into the host chromosomal DNA. The family of *Retroviridae* is composed of two subfamilies: *Orthoretrovirinae* and *Spumaretrovirinae*. Within *Orthoretrovirinae* are six genera: Alpharetroviruses, Betaretroviruses, Gammaretroviruses, Deltaretroviruses, Epsilonretroviruses, and Lentiviruses (Sharp and Hahn, 2011). HIV is a member of the Lentivirus genus and viruses in this genus are broadly associated with long incubation times in their host as well as causing progressive diseases that often result in severe immunological defects (Swanson and Malim, 2008). Acquired immunodeficiency syndrome (AIDS) in humans is caused by two types of HIV, type 1 and type 2 (HIV-1 and HIV-2, respectively) (Sharp and Hahn, 2011). HIV-1 is the more prevalent and pathogenic of the two types, and therefore will be the focus in this thesis.

The HIV-1 genome is approximately 10kb and encodes nine genes, including gag, pol and env, which are core components of all functional retroviruses (Figure 1.1A). The coding region of the genome is flanked by long terminal repeats (LTRs). The function of the 5' LTR is to act as a promoter that initiates the transcription of the viral genomes after integration. The 3' LTR is important for the addition of the polyadenylation signal (poly(A) tail). The HIV-1 pol gene encodes three viral enzymes critical to its replication: protease (PR), reverse transcriptase (RT), and integrase (IN). The genome also encodes two structural genes; gag and env. The Gag polyprotein consists of the three main structural domains: matrix (MA), capsid (CA), and nucleocapsid (NC) as well a p6 domain and two spacer peptides (SP1 and SP2) (Freed, 2015). The env gene encodes two subunits: a

transmembrane domain (TM) gp41 and the surface domain (SU) gp120 (Figure 1.1B). In addition to these proteins found in all retroviruses, HIV encodes six additional accessory/regulatory genes: vif, vpr, vpu, nef, rev and tat. The proteins Vif, Vpr, Vpu, and Nef function as accessory genes that are not essential to replication *per se*, but counteract the innate and adaptive immune system of the host. The Rev protein binds to the Rev response element (RRE), which is a cis-acting sequence located in the env gene, and mediates the efficient nuclear export of the unspliced and partially spliced viral messenger RNA (vmRNA). The Tat protein binds to the trans-activating response (TAR) sequence in the viral RNA during transcription to promote transcription elongation. The mature virion has a diameter of approximately 100 nm and contains two copies of the positive stranded RNA genome (Figure 1.1B). The conical shaped capsid encases the copies of the nucleocapsid-coated genome as well as molecules of IN, PR, and RT. Surrounding the capsid is a layer of matrix surrounding the inner lipid membrane taken from the lipid bilayer of the host cell's membrane (Sundquist and Kräusslich, 2012).

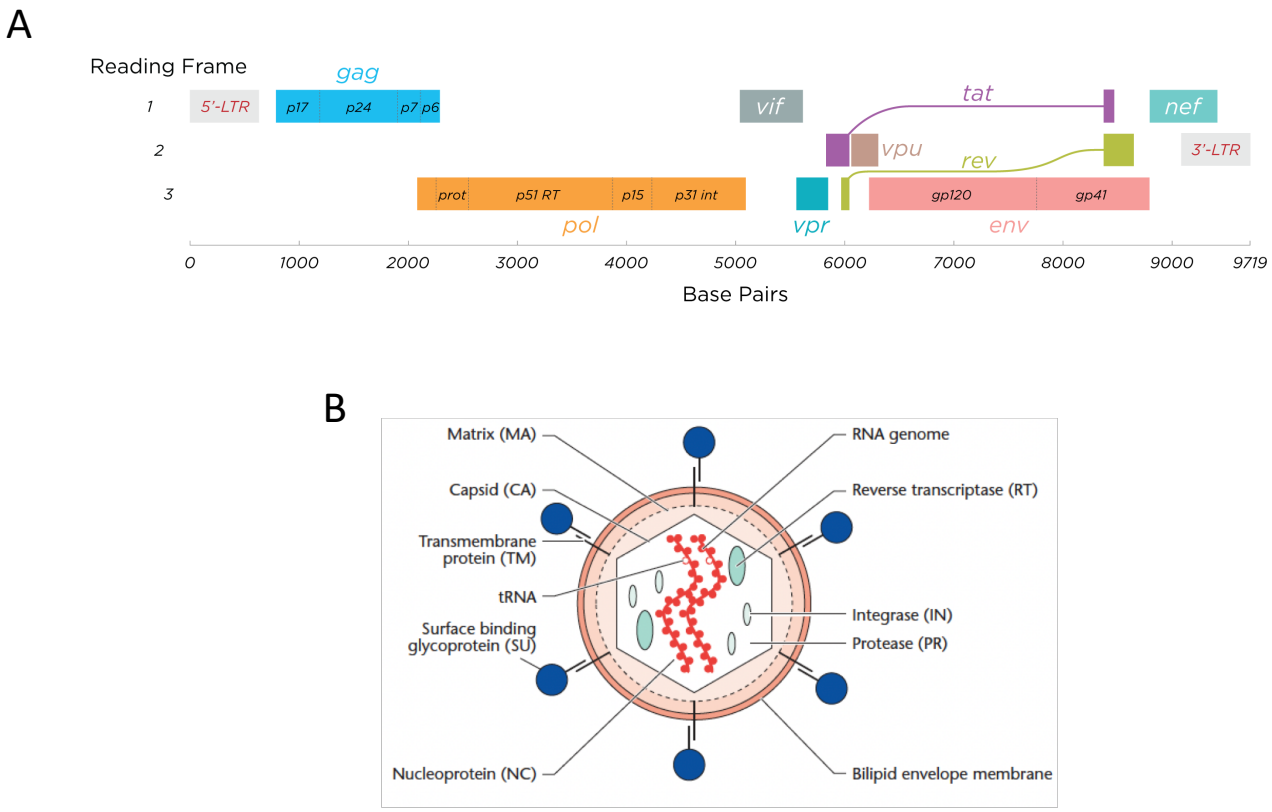


Figure 1.1: Organization and structure of the HIV-1 genome

(A) Organizational schematic of the open reading frames in HIV-1. Polyproteins Gag, Pol and Env are further subdivided into their respective individual proteins. Approximate length in base pairs represented on the X-axis. **(B)** Structure of a mature HIV-1 particle with labeled proteins. (Adapted from Petersen et al. 2006).

Replication Cycle

The first step in the replication cycle of any virus is the attachment of the virion to the host cell. The specificity of this interaction defines the viral tropism, or what cell type the virus

will infect. For HIV-1, the Envelope protein gp120 is restricted to only functionally binding the human protein CD4, which is a transmembrane protein expressed on the cell surface of a subset of T cells (CD4+ T cells) (Kowalski et al., 1987). As a member of the immunoglobulin family, the physiological role of CD4 is the enhance the T cell receptor (TCR)-medicated signaling. Beyond binding to CD4, for HIV-1 to infect a cell it must also engage with one of two different chemokine receptors, CCR5 or CXCR4, present on the surface of different CD4+ T cells (Berger et al., 1999; Wilen et al., 2012) (Figure 1.2). A recent structure of the CD4-gp120-CCR5 complex suggests that the coreceptor functions by stabilizing the conformation change in Env induced by binding to CD4 (Shaik et al., 2019). Ultimately, this co-receptor interaction allows gp120 to insert its hydrophobic fusion peptide into the target cell membrane, which effectively crosslinks the viral and host membranes (Wilen et al., 2012). This insertion initiates the folding of the gp41 subunits and brings together the amino terminal helical region and the carboxyl terminal helical region of gp41, resulting in the formation of the six-helical bundle (Chan et al., 1997) (Figure 1.3).

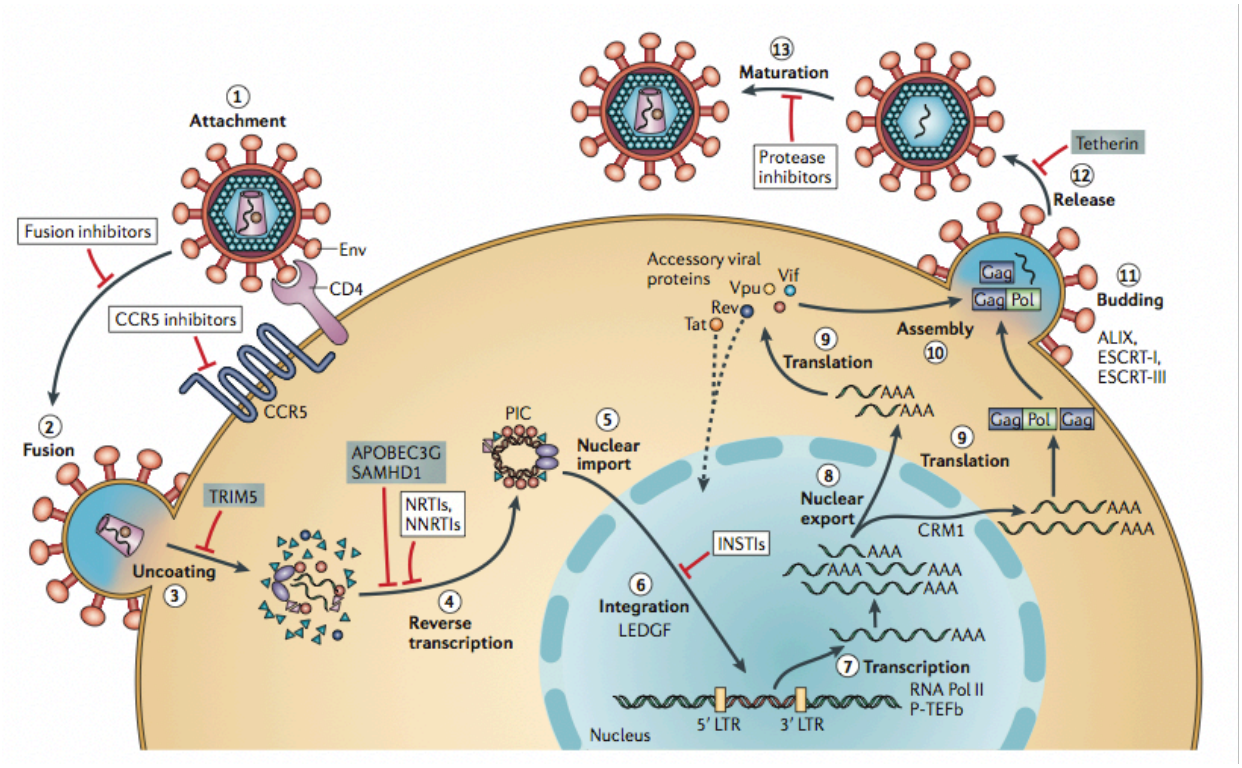


Figure 1.2: Overview of HIV-1 replication cycle

Schematic representation of the HIV-1 replication cycle from entry to production and maturation. (1) Receptor interaction and attachment, (2) membrane fusion and entry, (3) uncoating, (4) reverse transcription, (5) nuclear import, (6) integration, (7) mRNA and gRNA expression, (8) nuclear export of mRNA, (9) protein synthesis, (10) virion assembly, (11) budding, and finally (12) maturation (Adapted from Engelman and Cherepanov 2012).

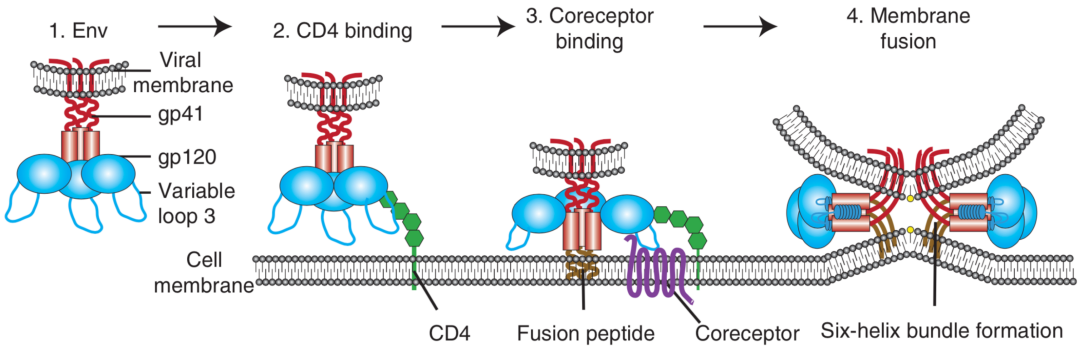


Figure 1.3: Mechanism of HIV-1 attachment and fusion to a CD4+ t cell

Schematic representation of HIV-1 entry. (1) HIV-1 Env is made up of gp120 containing variable loops for attachment and gp41, which is the transmembrane domain, (2) gp120 protein binds to the cellular transmembrane protein CD4, which initiates a conformation change in Env, (3) this change allows the co-receptor to bind and the insertion of the gp41 fusion peptide into the host cell membrane, (4) formation of the six helical bundle finally results in membrane fusion (adapted from Wilen et al. 2012).

Traditionally, after fusion the next stages of HIV-1 replication are in the discrete order of uncoating, reverse transcription, and nuclear import (Campbell and Hope, 2015). There is some evidence to suggest that the core remains intact until it docks at the nuclear pore, while other studies have found uncoating occurs during transport of the genome to the nucleus and is stimulated by reverse transcription (Arhel et al., 2007). Additionally, there is clear evidence that indicates CA is a key determinant in the ability for HIV to infect non-dividing cells, indicating nuclear import requires more than the reverse transcribed genome (Yamashita et al., 2007). The confluence of data surrounding this process of uncoating, reverse transcription, and nuclear import belie the emergence of a clear model concerning the timeline of events. Though, the conflicting data surrounding this topic demonstrates that HIV is a dynamic and adaptable virus and is therefore difficult to broadly characterize in a way that is applicable to all scenarios. Additionally, the limitations of techniques available (e.g. live cell imagining or experimentation on bulk cell populations) to study this process potentially stands in the way of a more clear and robust model.

It is well established that after fusion of the envelope to the target cell, the viral core is released into the cytoplasm with a single stranded RNA (ssRNA) genome that must serve as the template for the synthesis of the double stranded DNA (dsDNA) genome by the viral enzyme RT (Baltimore, 1970; Telesnitsky and Goff, 1997)(Figure 1.4). The process of reverse transcription has been well characterized and broadly follows these steps: (i.) Initially, a short section of the minus-strand DNA must be synthesized from the 3'-OH of the tRNA bound to the primer binding site (PBS). (ii.) The RNaseH domain of the viral RT

then digests the RNA portion of the newly synthesized RNA-DNA hybrid, thus freeing the short single stranded DNA fragment, also known as the minus strand strong-stop DNA (-sssDNA). (iii.) This –sssDNA is transferred to the 3' end of the genome by hybridizing to the repeated region (R region). (iv.) Upon hybridization to the R region, minus strand synthesis occurs using the –sssDNA as a primer. This is followed by the digestion of the RNA from the RNA–DNA hybrid by RNaseH, intentionally excluding digestion of the polypyrimidine tract (PPT). (v.) The remaining RNA bound to the minus-strand DNA at the PPT then serves as a RNA primer for the plus-strand DNA synthesis to occur until the tRNA primer. (vi.) Following plus-strand synthesis of the tRNA primer, the tRNA is removed by RNaseH, exposing the PBS at the 3' end of the plus-strand DNA to hybridize at the 3' end of the minus-strand DNA, which is known as the second-strand transfer. (vii.) Finally, plus-strand and minus-strand DNA synthesis results in the completion of the double stranded DNA provirus that is capable of integrating into the host cell genome (Telesnitsky and Goff, 1997). The synthesis activity of RT is highly error prone making approximately 2×10^{-5} substitutions per base and allows for the virus to test a large number of different sequences over a small number of replication cycles, contributing the ability of HIV to quickly adapt and find escape mutants under different cellular pressures (Hu and Hughes, 2012).

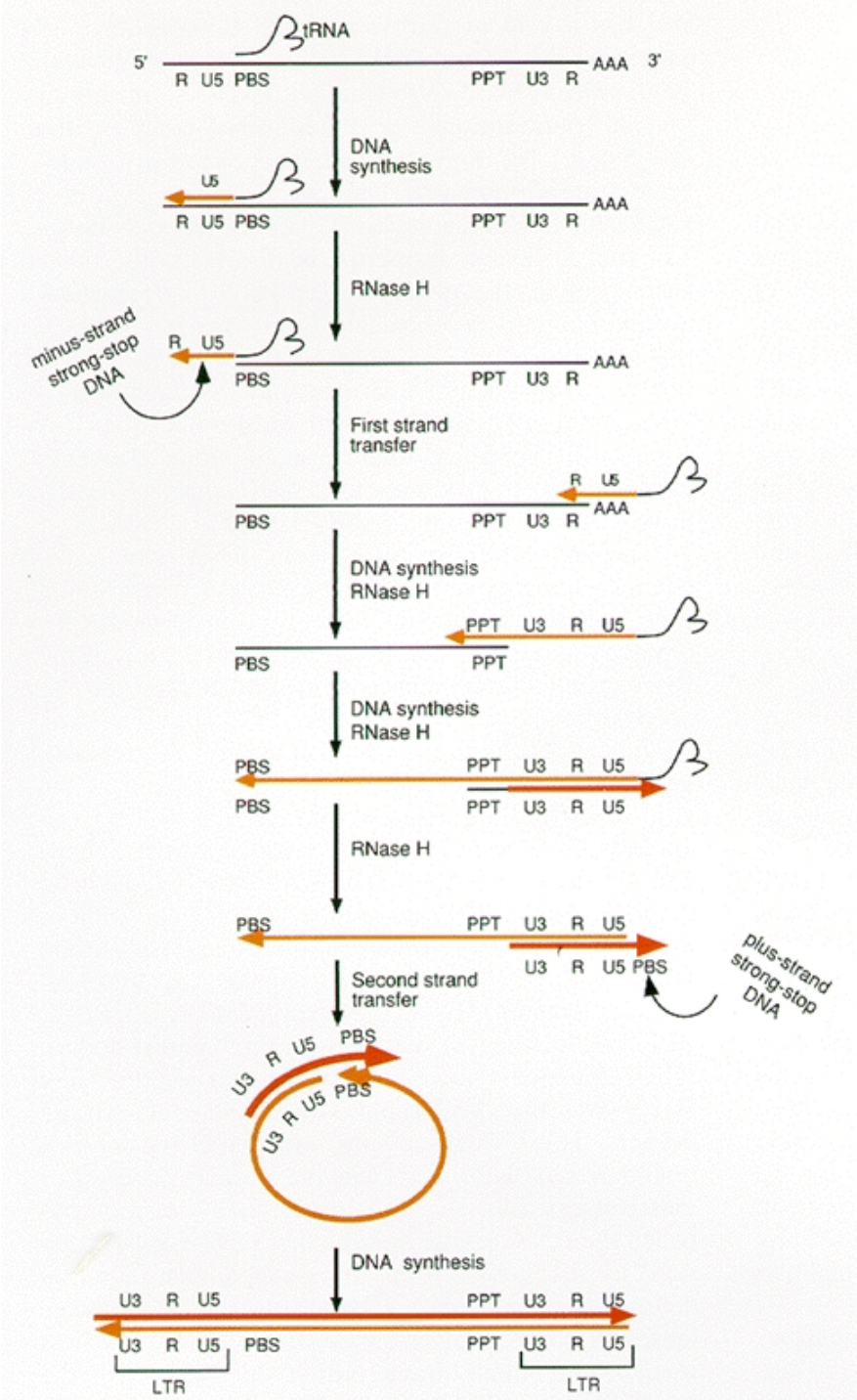


Figure 1.4: Stepwise process of reverse transcription

Schematic representation of the reverse transcription of a retroviral genome. Steps outlining the RT-dependent synthesis of the dsDNA are described in the main text. Black line: RNA; Orange line: minus-strand DNA; Red line plus-strand DNA.

Upon completion of reverse transcription, the double-stranded viral DNA must associate with IN and other viral and cellular proteins to form the pre-integration complex (PIC) (Craigie and Bushman, 2014). The lentivirus genera are capable of replicating in non-dividing cells by translocated through nuclear pores, a process that is dependent on CA (Yamashita and Emerman, 2004). While it is not completely understood how the PIC transverses the nuclear pore, there is a clear consensus on some of the cellular proteins that are required. Multiple genome-wide short interfering RNA (siRNA) screens have confirmed the importance of transportin 3 (TNPO3) and nucleoporins (Nup) 358 and 153 in nuclear entry (Brass et al., 2008; Konig et al., 2008). Consistent with the opacity and complexity surrounding the process of uncoating/nuclear import, a recent study demonstrated how nuclear pore heterogeneity of different cell lines, an innate immune restriction factor that can block nuclear import (MX2), and CA mutations to circumvent these two factors are all obstacles to HIV gaining entry to the genome of a cell (Kane et al., 2018). This *tour de force* study clearly demonstrates that a myriad of factors can profoundly affect how the PIC can adapt and respond to different conditions in order the complete infection.

Once inside the nucleus the viral enzyme IN is responsible for integrating the dsDNA viral genome into the host DNA genome. Initially, IN removes two nucleotides from the 3' end of the viral dsDNA, generating a reactive 3'-hydroxyl group at both ends. These newly formed 3'-hydroxyl ends will ligate to the host DNA through binding to a pair of phosphodiester bonds five nucleotides apart and on opposing stands of the host DNA, creating the integration intermediate. The final step of integration requires trimming the

two remaining nucleotides of the 5' viral DNA and the extension from the 3' end of the genomic DNA (Craigie and Bushman, 2014). High-throughput sequencing has provided insights regarding the integration site preference, suggesting that HIV-1 prefers to integrate into active transcription units (Mitchell et al., 2004; Schroder et al., 2002). The viral genome targets these sites of active transcription through interacting with the host transcription factor LEDGF, which was initially discovered through an affinity-based screen for proteins that interact directly with IN (Cherepanov et al., 2003). LEDGF is capable of recruiting the PIC to integration sites even at very low levels, and only a complete knockout of LEDGF will result in decreased levels of infectivity (Llano et al., 2006). Because LEDGF is a transcription factor and is only in the nucleus when it is at site of active transcription, the strong affinity of IN to LEDGF ensures that the PIC will be recruited to sites of active transcription.

After the viral genome has integrated into the host cell genome, it must initiate transcription. Transcription initiation of the viral mRNA is remarkably efficient and initiated by the U3 regions in 5' LTR, which consists of DNA regulatory elements that recruit cellular transcription factors (Rittner et al., 1995) (Figure 1.5) . The LTR core promoter is finely tuned to effectively recruit the cellular transcription factors and is made up of three tandem Sp1 binding sites (Jones et al., 1986), a TATA box (Garcia et al., 1989), and a highly active initiator (INR) sequence (Zenzie-Gregory et al., 1993). The non-core promoter consists of the transactivation-response region (TAR) and the enhancer region with two NF-KB binding motifs (Nabel and Baltimore, 1987). RNA polymerase II (RNAP II) is recruited to the 5' LTR through an interaction with transcription factors IID (TFIID),

IIH (TFIIH), and the TBP-associated factor (TAF). Along with TFIIH, the cyclin-dependent kinase (CDK7) is recruited to this complex in order to partially phosphorylate the RNAPII. Progression of this complex will stall at TAR due to the presence of the negative elongation factor (NELF) (Yamaguchi et al., 1999). The transcription elongation of this complex is inefficient when it is only comprised of cellular factors, and for this reason HIV encodes the transcriptional transactivator protein Tat that binds TAR and stimulates RNAP II processivity, thus increasing the elongation of the transcripts to the end of the virus (Sodroski et al., 1986). Concurrent with transcription of the full viral genome, HIV uses alternative splicing to create three distinct species of mRNA: (i.) fully spliced transcripts that are efficiently exported from the nucleus by cellular factors, (ii.) partially spliced and (iii.) unspliced transcripts that are not exported from the nucleus by cellular factors. For the partially and unspliced transcripts to be exported from the nucleus the fully spliced rev transcripts must be exported and translated. The Rev protein is then imported into the nucleus to bind the rev response element (RRE) present in the partially and unspliced mRNA for proper nuclear export (Karn and Stoltzfus, 2012; Legrain and Rosbash, 1989).

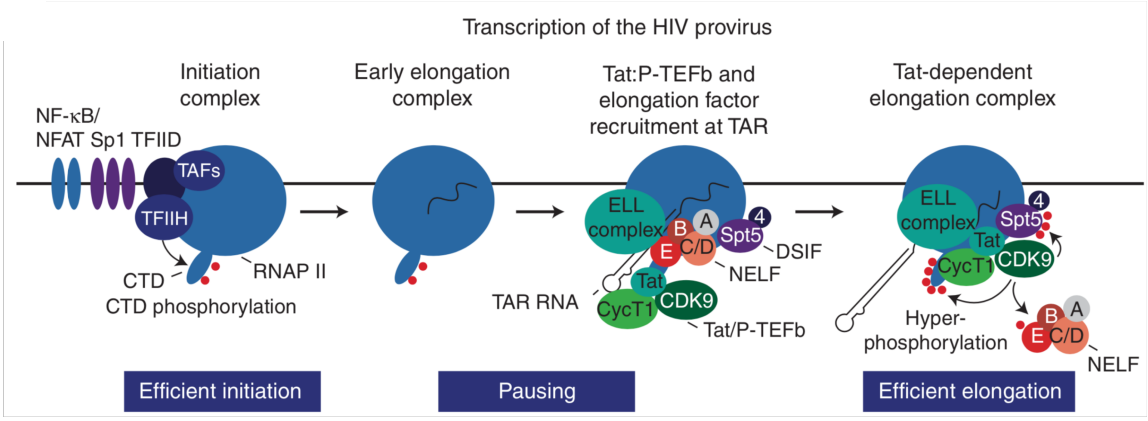


Figure 1.5: Mechanism of HIV-1 transactivation by Tat

Initiation: RNAP II is recruited to the 5' :LTR by TFIID, TFIIH and the cofactor TAF. TFIIH phosphorylates the CTD of RNAP II to initiate elongation. **Pausing**: Once transcription encounters the TAR element NELF and DSIF are recruited to the complex and the whole complex pauses. **Elongation**: Tat/P-TEFb bind to TAR and CDK9 initiates the phosphorylation of NELF resulting in its release as well as phosphorylation of Spt5 and RNAP II, finally allowing for efficient elongation to proceed (adapted from Karn Stoltzfus 2012).

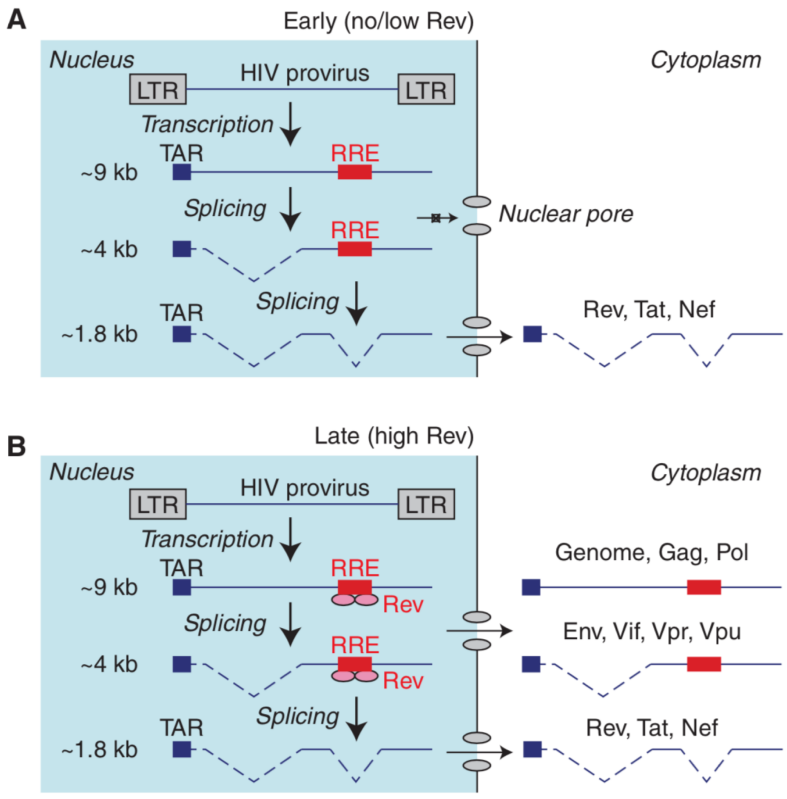


Figure 1.6: HIV-1 mRNA splicing

(A) Early phase of RNA expression: Only the fully spliced 1.8kb mRNA (tat, rev and nef) are capable of being exported from the nucleus into the cytoplasm efficiently for translation. Unspliced and partially spliced transcripts are retained in the nucleus where they will undergo degradation. **(B)** Late phase of mRNA expression: The Rev protein is expressed and translocates into the nucleus, where it binds to the RRE of the fully and partially spliced mRNA, facilitating their export into the cytoplasm (adapted from Karn and Stoltzfus 2009).

HIV-1 protein synthesis relies entirely on the host cell translation machinery. This is made possible through that addition of a 5' 7-methyl-guanosine cap and a 3' poly(A) tail. As with eukaryotic protein synthesis, the synthesis of HIV proteins can be divided into four discrete phases: (i.) recruitment of the small 43S ribosomal complex to the mRNA to scan until it locates an initiating AUG start codon; (ii.) the 60S ribosomal subunit joins this complex and forms the 80S ribosomal complex which begins elongation where the mRNA is decoded and the nascent polypeptide chain begins to form; (iii.) when the ribosome comes to a termination codon it releases the polypeptide chain; (iv.) and finally the ribosome is recycled for the translation of another mRNA (de Breyne et al., 2013). With the exception of Env and Vpu, all of the other viral proteins are synthesized on free ribosomes in the cytoplasm. The Env and Vpu are synthesized in the rough endoplasmic reticulum (RER) for proper insertion into the plasma membrane, decoration with sugar groups, and targeting the cells plasma membrane (Checkley et al., 2011). Expression of the Gag-Pol polyprotein requires translational suppression of the Gag stop codon, and this is accomplished by the presence of the Gag-Pol frame shift RNA sequence that sits between the gag and pol open reading frames. The frameshift sequence is a series of six uracil nucleotides followed by an adenosine codon upstream of the gag stop codon. This "slippery" sequence results in a ribosome shift that skips the gag termination codon and proceeds through to translation of the pol gene 5% of the time (Karn and Stoltzfus, 2012). Upon completion of translation the virus must bring all of the components together to assemble a nascent virion. The assembly of the virion occurs at the plasma membrane of the host cell and is primarily driven by the Gag polyprotein, which is capable of forming

particles in the absence of pol or env (Campbell and Vogt, 1997). Gag is critical to the formation of what will become an infectious virion, coordinating the recruitment of Pol, which is accomplished through the targeting of the Gag-Pol polyprotein, as well as recruiting both copies of the viral genomic RNA (vgRNA) through the NC domain (Kutluay and Bieniasz, 2010; Kutluay et al., 2014). Gag itself is targeted to the plasma membrane through the MA domain, specifically by the myristate group and basic amino acid motif that facilitate a direct interaction with specific phospholipids at the plasma membrane (Saad et al., 2006). As stated previously, Envelope is trafficked to the plasma membrane through the secretory pathway, but the exact process by which the Env protein is incorporated into the nascent particle is still not fully understood. It has been demonstrated that MA interacts with gp41, but this interaction is not strictly necessary for Env incorporation (Sundquist and Kräusslich, 2012). It remains possible that incorporation of the HIV envelope to the viral particle is not highly specific and is supported by the ability easily psuedotype HIV virions with other envelopes. However, the consistent incorporation of 7 to 14 Env trimers into each viral particle might suggest some level of control surrounding the process of Envelope recruitment (Chertova et al., 2002; Zhu et al., 2006).

Once all of the necessary components are present at the plasma membrane, the virus particle needs to detach from the host cell. For budding the occur, the cellular ESCRT machinery must be recruited to the site of viral egress to stop Gag multimerization as well as catalyze the excision of the cell and viral membranes (Bieniasz, 2009). HIV-1 contains two late domains in the p6 domain of Gag, and these late domains are responsible for the

recruitment of early acting ESCRT factors (Carlton and Martin-Serrano, 2009). Specifically, the primary late domain contains a PTAP motif that binds the TSG101 subunit of the ESCRT-I complex for recruitment to the budding virion (Garrus et al., 2001). The presence of PTAP motifs in other cellular proteins suggests that the HIV-1 late domains are a mimicry of the cellular mechanism to recruit the ESCRT pathway (Ren and Hurley, 2011). After the virion is released from the cell via the ESCRT pathway, the final stage is virion maturation. During a late stage of virion assemble the viral PR is cleaved by autoproteolysis and freed to cleave the Gag-Pol polyproteins upon virion release. Pol is cleaved by PR into the viral enzymes RT and IN, while Gag is cleaved into the structural proteins MA, CA, NC, p6, p1 and p2. The cleavage of these proteins produces a dramatic conformational rearrangement. Specifically, CA monomers assemble into hexamers to form the structural core, while the spontaneous formation of capsid pentamers results in the irregularities in capsid organization and produces the cone structure emblematic of the HIV particle (Ganser et al., 1999; Pornillos et al., 2009). The fully matured virion is now poised to infect a target cell that it comes in contact with, thus starting the process over again.

RNA elements in HIV-1

As briefly discussed above, HIV-1 genome contains a variety of RNA elements that have important *cis*-acting function. Some of these RNA sequences are multi-functional in that they lie in open reading frames and therefore encode proteins as well as performing functions as RNA that are critical during viral replication. Additionally, there are several

genes in the HIV genome that overlap with adjacent genes and it has been suggested that this multifunctional coding actually helps constrain the evolution of the virus and provides a fitness advantage (Fernandes et al., 2016). Known *cis*-acting RNA elements in the HIV-1 genome that lie within protein-coding sequences include splice donors, acceptors, and branch points (Purcell and Martin, 1993), splicing regulatory elements that enhance or inhibit the use of proximal splice site (Madsen and Stoltzfus, 2006), the Rev-responsive element (Heaphy et al., 1990; Malim et al., 1990), the central polypurine tract and termination sequence (Charneau and Clavel, 1991), the Gag-Pro-Pol ribosomal frameshift regulatory element (Parkin et al., 1992) and components of the viral genome packaging signal (Kutluay et al., 2014; Kuzembayeva et al., 2014).

Because the replication of HIV is constrained to a human cell and much of the replication machinery in the cell, it must efficiently recruit the set of required cellular factors while avoiding cellular proteins that are deleterious to its replication. This is often studied in the context of protein-protein interactions, more specifically with cellular proteins that have evolved a specific role in actively antagonizing a stage in the HIV-1 life cycle, referred to as restriction factors. This is no less true when considering the fate of the HIV-1 mRNA. Of the proteins encoded in the human genome, approximately 8% of them are RNA binding proteins that accomplish a variety of tasks in the post-transcriptional regulation of RNA in the cell (Gerstberger et al., 2014). All of the HIV-1 mRNA molecules must properly recruit the necessary cellular factors for post-transcriptional regulation while avoiding any cellular factors that might actively seek to recognize the non-self RNA or process it through an aberrant pathway. To accomplish this, the HIV-1 genome has either (i.)

mimicked RNA sequences in cellular pathways to recruit the required cellular factors, or (ii.) uses its own viral RNA binding proteins that have a strong affinity for both an RNA sequence in its genome and also the cellular protein factors that it needs to recruit. Little is known about whether HIV-1 mRNA avoids cellular detection, but this is also possible if not likely. Additionally, a recent study described how the mRNA of HIV-1 might be recognized and induce the production of IFN through a yet undiscovered pathway (Akiyama et al., 2018; McCauley et al., 2018). Here we will review some of the cis-acting RNA elements in the HIV-1 genome that recruit either trans-acting cellular and/or viral factors to ensure the proper replication of the virus.

TAR-Tat

As discussed previously, HIV uses its LTR as the promoter, which is a common feature among all retroviruses. Initially, the viral LTR was cloned into a simple plasmid as the promoter for chloramphenicol actyl-transferase (CAT) gene, but high CAT expression was only detected in cells that were infected with HIV-1, suggesting that there must be some viral factor supplied in trans that activates the LTR (Sodroski et al., 1985b). This factor was quickly discovered through iterative cloning regions of the HIV-1 genome into this plasmid by the same group and named Tat for its homologous function to other retroviral proteins (Sodroski et al., 1985a). Through deletion studies of the viral LTR, Tat was found to bind to the transactivation-response region (TAR), and this sequence is the first fifty-nine nucleotides after the initiation site for transcription. TAR does not work like most transcriptional elements, but is only functional with placed in a precise location and

orientation in the LTR (Muesing et al., 1987). Genetic studies revealed that TAR is as a RNA stem-loop with a highly stable structure (Berkhout et al., 1989; Selby et al., 1989). Biochemical assays revealed a direct and specific interaction between Tat and TAR, and later work mapped this interaction to the uridine-rich (U-rich) bulge present within the stem of TAR (Dingwall et al., 1990; Dingwall et al., 1989). Binding of Tat to TAR results in a conformational change in the RNA structure that stabilize the RNA-DNA hybrid (Puglisi et al., 1992). This interaction between Tat the U-rich bulge of TAR blurs the lines between protein recognition of an RNA molecule by sequence or structure, cumulatively suggesting that the sequence, structure and function of this interaction are inextricably linked (Figure 1.5).

The apical loop of TAR is inconsequential in its affinity for Tat, yet mutational analysis of this loop revealed that it is critical to the function of TAR. (Feng and Holland, 1988). This finding suggested that some additional factor binds the apical loop of TAR and is critical to its function. This hypothesis was confirmed by a multitude of studies identifying many more factors that are recruited to this complex and are essential to function. Concomitantly, further investigation revealed that Tat regulates transcriptional elongation rather than initiation. These studies demonstrated that in the absence of Tat a majority of RNA polymerases that initiate transcription are recruited and stalled near the promoter, while when Tat is present the polymerases are more frequently found downstream the promoter (Kao et al., 1987). This idea that Tat is involved in transcriptional elongation was further confirmed when the well characterized positively acting elongation factor, CDK9, was found to strongly interact with Tat. While critical to function, the CDK9 complex was

not a fit for the factor that bound the apical stem-loop of TAR. Finally, the cyclin component CycT1 was discovered to bind the CDK9 complex, which then recruits Tat. Only when Tat is bound to this CycT1-CDK9 complex can it then efficiently bind the TAR stem-loop (Wei et al., 1998).

A variety of mechanisms are observed to be employed by HIV to ensure processive transcription. Transcription initiation is recruited intrinsically through the HIV LTR, but this is not sufficient for full gene expression. Importantly, this weak transcription is viable enough for low levels of Tat production by purely cellular mechanisms, which is then necessary for transcriptional elongation of the full HIV genome through further recruitment of additional cellular machinery.

Splicing

HIV-1 requires alternative splicing to generate all the mRNAs needed for proper replication. The execution of proper mRNA splicing is highly complicated and requires dozens of cellular proteins in a proper order and location on the mRNA. To accomplish this, HIV-1 has mimicked many cis-acting RNA elements that control splicing in human cells in order to process its genome though alternative splicing pathways in the nucleus. Alternative splicing is executed by a macromolecular complex commonly referred to as the spliceosome (Chen and Manley, 2009) (Figure 1.6). Spliceosome assembly is initiated by (i.) the recognition of the 5' splice donor by the small nuclear ribonucleoprotein particle U1 (snRNP U1) and by the binding of the splicing factor 1 (SF1) to the branch point sequence in the RNA, forming the E' complex (Berglund et al., 1997). (ii.) The U2 auxiliary

factor (U2AF) binds to the polypyrimidine tract to create the E complex ((Nelson and Green, 1989; Zamore and Green, 1989). (iii.) SF1 dissociates from the branch point and is replaced by the U2AF to form the A complex. (iv.) Additional recruitment of the U4/U6-U5 tri-snRNP complex forms the B complex, which is still inactive. (v.) Finally, drastic conformational rearrangements of these components form the catalytically active spliceosome, the C complex. In the human genome, exons have an average length of 50-250 nucleotides. This is much shorter than the introns separating them, which are often thousands of nucleotides in length. For this reason, spliceosome assembly usually occurs around exons (Sterner et al., 1996). This process is referred to as “exon definition” and involves some interaction occurring between the SF1/U2AF complex at the 5’ splice site and the U1 protein at the 3’ splice site of the same exon (Berger et al., 1999). Eventually, exon definitely must be complemented by “intron definition”, and this is thought to occur through interactions between the U1 and U2 snRNPs (Kotlajich et al., 2009; Lim and Hertel, 2004). The ability of the splice site to recruit the U1 or U2 snRNPs has the largest impact on the recognition on that particular slice site.

The ability of a splice site to recruit U1 and U2 snRNP proteins is widely thought to depend on two factors: the intrinsic affinity of the snRNP proteins for a particular splice site and the cis-acting elements proximal to the splice site (Graveley, 2000). These cis-acting elements can either enhance or suppress the recruitment of snRNP proteins and are found in either introns or exons and are divided into four categories: exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs), and intronic splicing silencers (ISSs) (Long and Caceres, 2009; Tacke and Manley, 1999).

The silencing elements are thought to be frequently bound by heterogenous nuclear RNPs (hnRNPs), while the enhancing elements are bound by Serine-Arginine (SR) proteins (Dreyfuss et al., 2002; Smith and Valcárcel, 2000). A small number of SR proteins have a known RNA sequence that they bind to, primarily thought a RNA recognition motif (RRM), yet the majority of SR proteins do not have a specific RNA sequence that they have an affinity for (Chen and Manley, 2009). Similarly, there are some members of the hnRNP family of proteins that have a specific sequence known that they recognize through their RRM domains. Most known hnRNPs though are thought to recognize sequences that are rich in a particular nucleotide, while some examples such as hnRNP L or LL are thought to recognize RNA that is rich in cytosine and adenosine (Chen and Manley, 2009). In light of the limited data supporting the definition of cis-acting RNA sequences that recruit hnRNP and SR proteins, it remains unclear exactly how these proteins are recruited to splice acceptors and donors, and the precise mechanism of their regulation of splice sites remains incompletely understood. Yet, these and likely many other factors work in coordination to properly regulate the inclusion of correct exons and frequencies that are required for the particular cell type dependent splicing.

In the case of HIV-1, the virus employs four salient splice donors (D1, D2, D3, D4) and eight acceptors (A1, A2, A3, A4a,b,c, A5 and A7, S1 Table) to generate a large number of mRNAs that enable expression of nine viral open reading frames in a temporally biphasic manner (Holmes et al., 2015; Purcell and Martin, 1993). Additional, 'cryptic' splice sites may be used at very low frequency, and are not required for expression of the viral ORFs . HIV-1 splicing is, by necessity, inefficient as a substantial fraction of the viral

transcripts must remain unspliced so as to provide viral genomes and Gag-Pol mRNA (Martin Stoltzfus, 2009).

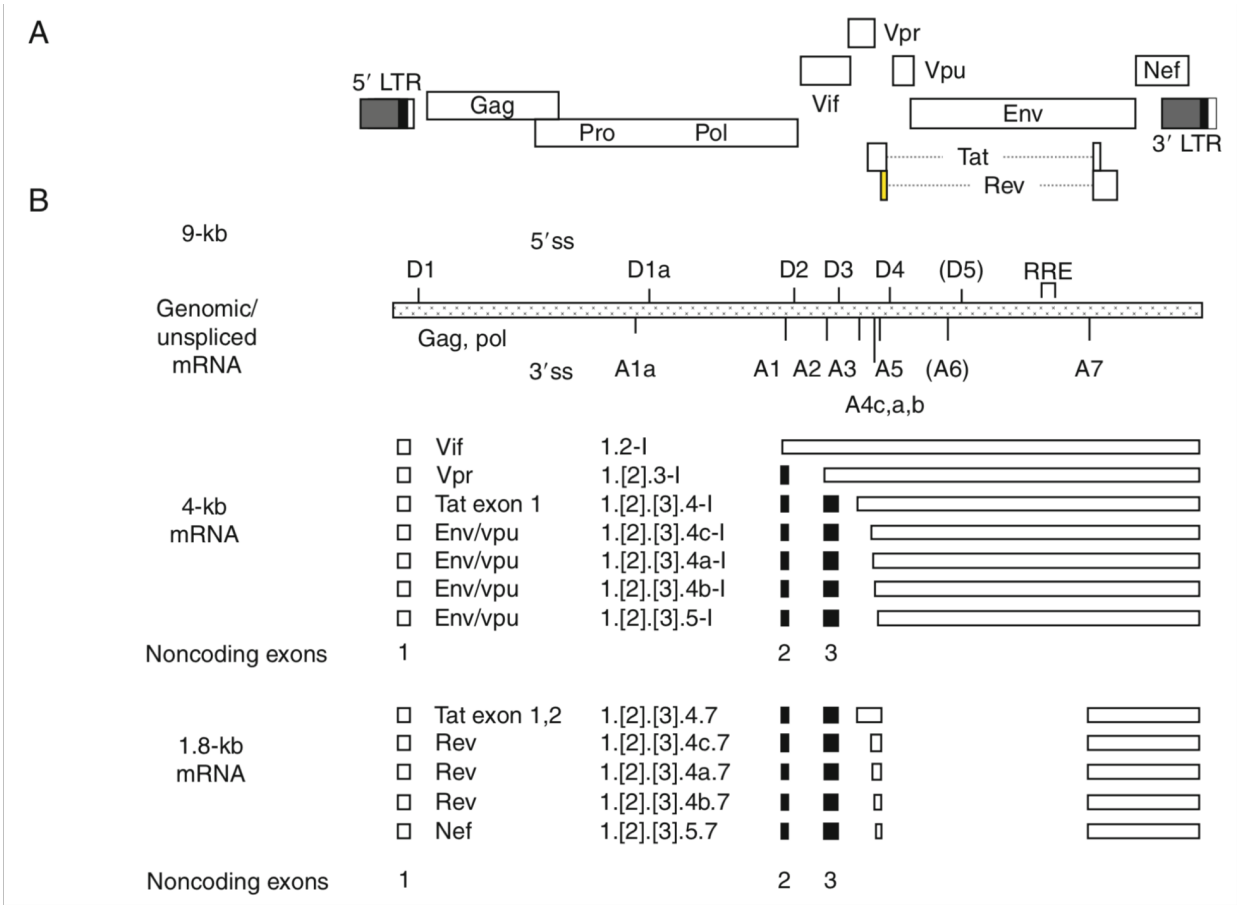


Figure 1.7: Representation of alternatively spliced mRNA in HIV-1

(A) Schematic representation of the HIV-1 genome and open reading frames. **(B)** Locations of the 5' and 3' splice sites in the HIV-1 genome (Genomic/unspliced mRNA). Exons present in the partially spliced (4-kb mRNA) and fully spliced (1.8-kb mRNA) and the corresponding genes they encode. Non-coding exons are show in black rectangles. **(C)** Locations of some of the known splicing regulatory elements shown in their relation to the exons of which they affect the inclusion/exclusion (adapted from Stoltzfus 2009).

All HIV-1 splicing involves D1, while splice acceptor sites 5' to each initiation codon are used to generate mRNAs encoding the HIV-1 proteins Vif (A1), Vpr (A2), Tat (A3) Rev (A4a,b,c), Vpu/Env (A4a,b,c and A5) and Nef (A5 and A7) (Figure 1.7) (Madsen and Stoltzfus, 2006; Purcell and Martin, 1993). In addition, some HIV-1 mRNAs include short noncoding exons (SX1 (A1-D2)) and (SX2 (A2-D3)) positioned 5' to the expressed open reading frame. The relative frequencies that the various splice sites are used, which can be measured using next-generation sequencing approaches (Ocwieja et al., 2012), likely contributes to ensuring that the optimal levels of viral proteins are synthesized for viral replication.

RNA sequence and structure must play a key role in alternative splicing. RNA secondary structure surrounding the major 5' splice donor affects splicing and SHAPE analysis revealed a novel stem loop that influences splicing (Emery et al., 2017; Pollom et al., 2013). With a nearly 10kb genome, it is almost certain that the RNA secondary and tertiary structures of the HIV-1 genome are dynamic and complex, making them difficult to study. Secondary structures could function either as sites for RNA binding proteins such as hnRNP or SR proteins, or as mechanisms for occluding nucleotide sequences that are sites of binding for splicing regulator proteins. Taken together, whether in the human genome or HIV-1 genome, characterizing the direct contributions of sequence homology, RNA structure, and splicesome recruitment remains complex and therefore incompletely understood.

RRE-Rev

As discussed previously, of the three classes of HIV-1 mRNA, only one is capable of export from the nucleus by only cellular factors. The unspliced and incompletely spliced transcripts, if retained in the nucleus, will be recognized as aberrant cellular mRNA molecules and degraded. To overcome this, HIV-1 encodes the viral protein Rev to bind the RRE present in the unspliced and partially spliced transcripts and efficiently export them through the nuclear pore. The initial discovery of Rev described it as an additional transactivating protein that acted post transcriptionally and relieves a negative regulatory pressure on the capsid and envelope protein expression (Sodroski et al., 1986). It was later discovered that Rev acts on these transcripts through binding the highly-structured 351 nucleotide stem-loop in the envelope gene, RRE (Malim et al., 1989). Subsequent biochemical studies found that when Rev binds to the high affinity site in RRE at the Rev-binding stem-loop (SLIIB), the binding event induces a conformational change that stabilizes the complex, reminiscent of the interaction between Tat and TAR (Daly et al., 1989; Heaphy et al., 1990). This binding of Rev to the RRE results in the recruitment of additional Rev monomers, with the degree of Rev oligomerization correlating to efficient nuclear export (Malim et al., 1990). NMR studies reveal that Rev inserts its alpha-helix of the arginine rich domain (ARD) binds along the major grove of the SLIIB. This is mediated by several nucleotide specific contacts and the arginine's of the ARD (Battiste et al., 1996; Tan et al., 1993).

Once Rev is bound and oligomerized to the RRE, it transports the RNA through the nuclear pore complex (NPC) (Kohler and Hurt, 2007). Rev interacts with the cellular

protein Crm1, a member of the karyopherin family, through its nuclear export signal at the carboxyl terminus of Rev (Fornerod et al., 1997). Crm1 will only bind the Rev/RRE complex in the presence of the GTP-bound form of Ran GTPase, after which they go through the NPC the GTP is hydrolyzed to GDP and the destabilizing the complex and disassociating it (Fischer et al., 1995). Rev returns to the nucleus alone through an interaction between a nuclear import signal overlapping with its RNA binding domain, recruiting the nuclear import factor, importin-B (Henderson and Percipalle, 1997).

Other RNA Considerations

That additional RNA sequences function in cis may exist in the HIV-1 genome is suggested by studies employing chemical probing approaches. For example, SHAPE experiments indicate that individual nucleotides in HIV-1 RNA have widely divergent tendencies to be base-paired (Wang et al., 2008; Watts et al., 2009; Wilkinson et al., 2008). These findings, along with phylogeny-based approaches, strongly suggest that secondary structures form in HIV-1 RNA that are conserved between strains, and might therefore serve a function in HIV-1 replication. One example of a recently suggested function for HIV-1 RNA secondary structure is the regulation of translational rate, whereby translation is periodically slowed to enable folding of one protein domain of the multidomain HIV-1 Gag protein before synthesis of the next (Watts et al., 2009).

Despite the suggestion that novel RNA secondary structures may be important for HIV-1 replication, targeted mutagenesis of putatively important individual stem loops has not generally yielded evidence that is strongly supportive of a role for these potential

structures in HIV-1 replication (Knoepfel and Berkhout, 2013). Conversely, some studies in which portions of the HIV-1 genome were synonymously mutated have suggested a role for RNA (as opposed to protein) sequence or structure in HIV-1 replication (Keating et al., 2009; Martrus et al., 2013; Schwartz et al., 1992). However, the precise nature of defects induced by synonymous mutations are unclear, and possibly pleiotropic.

It is not known whether the aforementioned represent a complete catalogue of cis-acting RNA elements, or whether additional RNA-based functionality exists in the HIV-1 genome. Of the known elements listed above, they mostly function through the direct recruitment of a specific RNA binding protein by either their RNA sequence or structure. In addition to the RNA sequence recruiting cellular or viral proteins, there are also elements that exist to avoid detection by cellular RNA binding proteins. One such example is in the RNA stabilizing element (RSE) in Rous Sarcoma Virus (RSV) that recruits PTBP1 to avoid detection from Upf1 and the nonsense mediated decay pathway (Ge et al., 2016). It is likely that more of these types of viral RNA elements exist to avoid cellular detection, but have been thus far difficult to discover.

Innate Antiviral Immunity

Vertebrate genomes have a variety of mechanisms to detect and respond to viral and bacterial infections, broadly classified as either the innate or adaptive immune systems. Within the innate immune system, antiviral proteins act as the first molecular lines of defense against viral infections. Broadly, there are five common features among antiviral proteins (Blanco-Melo et al., 2012; Malim and Bieniasz, 2012): (i.) they are dominantly

acting proteins that exhibit their antiviral activity in cell-culture assays, (ii.) they are often constitutively expressed and/or are further induced by interferon (IFN) stimulation, (iii.) they use unique mechanisms to inhibit specific steps in the viral life cycle, (iv.) they possess an unusually diverse amino acid sequence as a consequence of antagonistic co-evolution of viruses and (v.) they can be antagonized by viral proteins. The first restriction factor was identified in the early 1970's and was shown to specifically protect cells against MLV infection (Lilly, 1970). Currently, more than 20 proteins have been identified as restriction factors against a large variety of viruses across animal species, with more being added each year (Kluge et al., 2015). The reminder of this chapter will focus on a selected range of antiviral proteins that execute their function through a variety of mechanisms.

Tetherin

It had been long known that HIV-1 required the accessory protein Vpu for efficient particle release from CD4+ T cells (Bour and Strebel, 2003). Curiously, electron microscopy reveals that a Vpu-deficient HIV-1 remained attached to the cell surface as mature viral particles, suggesting some defect in viral release (Klimkait et al., 1990). Further characterization through different cell type, species specific, and heterokaryon experiments suggested that Vpu-deficient HIV-1 was due to the expression of a yet undiscovered dominant acting restriction factor (Neil et al., 2006; Varthakavi et al., 2003). Experiments demonstrated cells permissive to Vpu-decificent HIV-1 were made restrictive upon the addition of interferon and that virions trapped on the cell surface could

be released through treatment with protease, together indicating the presence of an IFN induced protein tether that prevented virus release (Neil et al., 2007). Finally, comparative gene expression of permissive and non-permissive cell type identified the protein BST-2 as the restriction factor antagonized by Vpu that can tether budding virions to the cellular membrane, appropriately named “tetherin” (Neil et al., 2008).

Tetherin is a transmembrane glycoprotein with a highly unusual domain architecture. It is comprised of a short N-terminal cytoplasmic tail, a single pass transmembrane helix (TM), a helical coiled-coil ectodomain (CC) that drives homodimeriation and a C-terminal glycophosphatidylinositol membrane anchor (GPI). Remarkably, elegant experiments replacing each of the domains of tetherin demonstrated that it is the domain architecture rather than primary amino acid sequence that are responsible for the ability of tetherin to directly tether virions to cells (Perez-Caballero et al., 2009). It remains unclear exactly how tetherin might target budding virions beyond homogenous expression along cellular membrane, though some evidence might suggest that it is recruited to sites of Gag-induced membrane curvature during budding (Grover et al., 2013). As mentioned, tetherin is antagonized by the HIV-1 accessory gene Vpu, a small N-terminally anchored transmembrane protein. Vpu antagonizes tetherin through a direct interaction in their respective transmembrane domains that a species specific (Kobayashi et al., 2011; McNatt et al., 2009). This interaction between tetherin and Vpu results in in ESCRT-dependent degradation of tetherin in lysosomal compartments (Sauter, 2014).

TRIM5α

Early studies demonstrated that the infection of non-human primate cells by HIV-1 was inefficient in an envelope-independent matter (Hofmann et al., 1999). A following study screened for rhesus macaque genes that restricted HIV-1 infection when exogenously expressed in human cells and identified TRIM5α (Mahadeo et al., 1994). TRIM5α is a member of the tripartite motif (TRIM) family of proteins, characterized by a RING domain with E3 ubiquitin ligase activity, B-box domains, a coiled-coil domain for dimerization, with some also encoding a PRY/SPRY domain that are involved in protein-protein interaction (Rahm and Telenti, 2012; Zheng et al., 2012). TRIM5α recognizes HIV through a direct interaction with CA on incoming viral cores, on which TRIM5 assembles into a three-dimensional lattice mediated through its SPRY domain (Ganser-Pornillos et al., 2011; Mahadeo et al., 1994). While the precise mechanism of this are not fully understood, this interaction results in a premature destabilization of the viral CA and degradation of its components before reverse transcription can complete (Stremlau et al., 2006).

Unlike tetherin, TRIM5α is not known to be antagonized by a viral protein in HIV-1. Rather, HIV-1 evades TRIM5α restriction through sequence variations within CA as well as the recruitment of the cellular protein CyclophilinA (CypA) that together allow efficient replication in human cells. The initial discovery that HIV-1 GagPR55 interacted with CypA was significant because CypA was known to bind tightly to a known immunosuppressive drug, CyclosporinA (CsA) (Handschumacher et al., 1984; Luban et al., 1993). CypA was later found the be incorporated into HIV-1 virions and is necessary for the production of infectious virus (Franke et al., 1994).

Very shortly after the discovery of TRIM5 studies demonstrated that disruption of the interaction between CypA and CA rescued infectivity in owl monkey, the opposite effect it has in human cells (Towers et al., 2003). This contradictory observation was rectified by siRNA experiments against CypA in owl-monkey cells, which rescued replication of HIV-1 and lead to the identification of a new chimeric protein (Sayah et al., 2004). Surprisingly, LINE-1 mediated retrotransposition in owl-monkeys created a CypA-TRIM5α fusion protein, aptly named TRIMCyp, that potently inhibits HIV-1 infection. This remarkable adaption of both vertebrate cells and the virus to coopt cellular proteins to either evade or enhance restriction of HIV represent a salient example of how protein-protein interactions drive the evolution host and viral genomes.

RIG-I

Beyond protein-protein interactions that recognize non-self viral proteins, cells have also evolved mechanisms of detecting non-self RNA. One of the most studied class of these proteins are the RIG-I like receptors (RLRs), which are a family of DExD/H box RNA helicases. RLRs are characterized by their recognition of pathogen associated molecular patterns (PAMPS) in viral RNA followed by a downstream signaling to transcription factor activation that drives the production of type 1 IFN (Loo and Gale, 2011). The RLRs are melanoma differentiation association gene 5 (MDA5), and laboratory of genetics and physiology 2 (LGP2), and RIG-I, the founding member for which the group is named for and therefore the member of which this section will primarily focus (Chow et al., 2018). RIG-I was initially described as a double stranded RNA (dsRNA) binding protein that

induced IFN production in response to synthetic dsRNA poly(IC) (Yoneyama et al., 2004). Subsequently, RIG-I was shown to act as a major factor that controlled permissiveness to the replication of hepatitis C virus (Sumpter et al., 2005). RIG-I recognizes RNA sequences that contain a 5'-triphosphorylated (5'ppp) end (Hornung et al., 2006). Beyond dsRNA with a 5'ppp end, it is thought that RIG-I prefers shorter RNA fragments and can also bind single stranded RNA (ssRNA) (Kato et al., 2008). Multiple studies have further characterized other factors in RNA that might contribute to the activation of RIG-I, indicating that sequence composition is an important determinant, with a preference for polyuridine motifs that are interspersed with cytosines (poly-U/UC) (Saito et al., 2008). In support of this, the deletion of the poly-U/UC in the hepatitis C virus genome abolished activation of RIG-I signaling, despite the inclusion of a 5'ppp in the virus, suggesting that the 5'ppp alone is not sufficient for activation in certain viral contexts (Saito et al., 2008). In the absence of an activating RNA molecule, RIG-I is held in an autorepressed state that is inactive. Upon binding RNA, RIG-I undergoes a conformational change where the CARDs and RD domains are dephosphorylated, allowing the RD domain to become ubiquitinated at K63. The Ubiquitination is essential for function and catalyzed through interactions with the E3 ubiquitin ligase TRIM25 and the RING finger protein Riplet (RNF135) (Cadena et al., 2019; Gack et al., 2007; Oshiumi et al., 2009). The K63-linked ubiquitination keeps RIG-I in its "open" conformation through oligomerization with other RIG-I molecules. This homo-oligomerization allows RIG-I to associate with the molecular adaptor MAVS, which subsequently recruit additional E3 ubiquitin ligases and the downstream proteins TNF receptor associated factor 2 (TRAF2), TRAF3, and TRAF6,

thus creating an active “signalosome” (Cai et al., 2014; Schmitz et al., 2014). The signaling cascade that follows induces the phosphorylation and nuclear translocation of IRF3, IRF7, and NF-kB, the transcription factors that drive the expression of type I and III IFN as well as other inflammatory genes that limit viral replication (Loo et al., 2006; Seth et al., 2005). There is no evidence to suggest that RIG-I degrades the bound RNA, but the RIG-I complex bound to its activating RNA might undergo proteasomal degradation as a negative regulation of the activation of IFN signaling (Arimoto et al., 2007).

It is reported that a RIG-I is capable of detecting and becoming activated by a host of viral genera including but not limited to: paramyxoviradae, rhabdovirade, orthomyxoviridae, filoviridae, and coronoaviridae (Chow et al., 2018). Considering such a diverse group of viruses are sensitive to RIG-I it is no surprise that there are a variety of different ways that viruses have avoided detection. Mechanisms of avoidance range from (i.) viral RNA modifications, (ii.) viral RNA sequestration, (iii.) post translational modification of RIG-I, (iv.) targeting of RIG-I/MAVS for proteasomal degradation, (v.) and modulation of downstream signaling components (Liu et al., 2017).

Because the key hallmark of RNA detected by RIG-I is the presence of a 5’ppp end, a set of viruses has evolved to posttranscriptionally remove the 5’ppp group from their genome. This feature of viral RNA modification has been reported in the Crimean-Congo hemorrhagic fever virus, Borna disease virus (BDV), and hantavirus (HTNV) (Garcin et al., 1995; Marq et al., 2010; Schneider et al., 2005; Wang et al., 2011). In the case of BDV, the 5’ppp is converted in to a monophosphate, which has a significantly lower affinity for RIG-I (Schneider et al., 2005), while the poliovirus genomic RNA is processed through

the cellular RNA-capping pathways, capping the 5'ppp dsRNA with a 7-methyl guanosine and methylated at the 5'-OH to make the viral RNA indistinguishable from the cellular RNA (Decroly et al., 2012; Devarkar et al., 2016; Lee et al., 1977). In contrast the modifications of the viral RNA that make it less distinguishable from host cellular RNA, viruses can also sequester their RNA so that is it not detectable by RIG-I. Ebola virus (EBOV) uses the viral protein VP35 to sequester the viral RNA by competing with RIG-I for the 5'ppp and forms a protein "end-cap" complex with the dsRNA, preventing recognition by RIG-I (Cardenas et al., 2006; Leung et al., 2010). Similarly, Influenza A virus (IAV) uses the non-structural protein 1 (NS1) to shield the viral RNA from RIG-I (Donelan et al., 2003). Further avoidance of RNA recognition is through RNA the re-localization of the viral RNA into a sub-cellular compartment, exemplified by DENV, which is thought to conceal its RNA in the intracellular membranes during replication ((Uchida et al., 2014)). There are multiple reports of how changes in the poly-U/UC motif present in the HCV RNA contributes to the activation or avoidance of RIG-I (Schnell et al., 2012; Uzri and Gehrke, 2009), yet more extensive study of this motif and how the presence or absence of the motif in other viruses contributes to recognition to RIG-I sensing does not exist and could be an additional mechanism of avoidance. The remaining mechanisms of RIG-I avoidance involve viral protein interactions with RIG-I other proteins downstream the cascade pathway. These mechanisms range from modulating the PTM of either RIG-I or MAVS that alter signaling efficiency (Kathum et al., 2016; Rajsbaum et al., 2012), Targeting RIG-I and MAVS for proteasomal degradation (Barral et al., 2009; Feng et al.,

2014), and negatively regulating the JAK-STAT pathways that drive the expression of IFN (Grant et al., 2016; Stevenson et al., 2013).

The remarkable diversity of viruses that RIG-I can antagonize is matched by the numerous ways a variety of viruses have evolved mechanism of evading sensing by RIG-I. Most of the viral mechanisms circumventing RIG-I are based on protein-protein interactions, primarily because of the technologies readily available to study these types of interaction. While there are some viral mechanisms employed to avoid detection of their RNA by RIG-I, the technical limits in RNA biology have hindered a more complete characterization. Therefore, it would not be surprising to discover that novel viral countermeasures exist between viral RNA and cellular proteins.

ZAP

As compared to the other proteins discussed in this section, the Zinc-finger Antiviral Protein (ZAP) is a less well characterized innate antiviral protein with RNA binding activity. A cDNA screen for factors that inhibit MLV replication discovered a protein fragment with four CCCH-type zinc-fingers that had homology to protein in the mouse genome. Simply, the protein was named rZAP, for rat Zinc-finger Antiviral Protein due to its restriction of MLV (Gao et al., 2002). ZAP expression had no effect on the initiation of reverse transcription, plus-strand DNA synthesis, or nuclear entry, but displayed a potent block to viral gene expression. As a control, different reporter genes were tested for ZAP sensitivity, but ZAP had only a modest effect on their expression compared to the potent restriction of MLV gene expression, suggesting that the MLV genome was somehow

specifically targeted by ZAP. Northern blot analysis of MLV gene expression in whole cells, nuclear extracts, and cytoplasmic extracts revealed that ZAP does not affect nuclear levels of MLV mRNA, but specifically targets the cytoplasmic mRNA.

This initial discovery elegantly described ZAP as a cytosolic sensor that specifically targeted the expression of MLV mRNA, but had no effect on the expression of cellular genes or a small effect on report genes. How ZAP distinguished between viral RNA and cellular RNA was an obvious next question to address. This was addressed in part by comparing a panel of viruses for ZAP sensitivity. The viruses tested in this initial characterization were primarily alphaviruses, consisting of Sindbis virus (SINV), Semliki Forest virus (SFV), Ross River virus (RRV) and Venezuelan equine encephalitis virus (VEE), but also included Herpes Simplex Virus 1 (HSV-1) from the Herpesviridae, Yellow Fever from the Flaviviridae family, and Vesicular stomatitis virus (VSV) from the Rhabdoviridae family (Bick et al., 2003). Infecting cells over expressing the rZAP construct originally used to identify ZAP with the alphaviruses revealed a strong inhibition of viruses from this genus, ranging from over 10-fold to 10,000-fold restriction in the case of SINV, even when infected at an MOI as high as 5.0. Equally clear was how the remainder of the viruses tested were not sensitive to ZAP. Even at MOI far below 1.0, none of these viruses displayed any effect on replication in the presence of ZAP. The findings of this study were particularly perplexing due to the groups of viruses sensitive to ZAP. Like SINV, both YF and poliovirus are positive stranded RNA viruses that replicate entirely in the cytoplasm, yet YF and poliovirus are both clearly insensitive to ZAP while SINV is profoundly sensitive. Similarly perplexing was the lack of any obvious sequence

homology that SINV shares with MLV, the virus used to discover ZAP. Importantly, this study described different viruses that are sensitive and insensitive to ZAP, yet no common feature emerged among the sensitive viruses that indicated how ZAP might recognize their RNA as non-self (Bick et al., 2003).

Using a different approach, a study came shortly after attempting to determine what are the viral RNA determinants that drive ZAP recognition. While the pervious experiments set out to broadly survey viruses sensitivity to ZAP with the goal of understanding a common feature among them, this group set out on a much more brute force approach, determined to unearth some sequence motif that ZAP recognizes (Guo et al., 2004). To do this, they used a luciferase reporter gene expression construct transfected in the presence of rZAP. As a 3' UTR to the luciferase were different segments of the SINV genome used as a bait for ZAP because SINV was previously shown to be exquisitely ZAP sensitive. The genome was originally unevenly divided into seven regions, tested as 3'UTR, and ZAP inhibition was measured by luciferase activity relative to cells not expressing ZAP. The regions found to be most sensitive were further divided into smaller segments and cloned into the luciferase construct, all of this done in both the sense and anti-sense orientation. This process continued until the genomic fragment was as small as 110 nucleotides from the original fragments of ~3,000 nucleotides. While this experimentation led to a fragment of 653 nucleotides in length that was significantly inhibited by ZAP, overall the strongest correlation between ZAP sensitivity and the genomic fragments was with the length of the fragment. The conclusion from these experiments was either that ZAP might recognize a complex tertiary structure in large

RNA sequences that computer programs were unable to predict, or that there might exist a small unidentified motif that only conferred sensitivity when the when the cumulative addition of the motif crossed a threshold that allowed ZAP to bind to the RNA (Guo et al., 2004).

Additional biochemical studies were carried out in the following years in an attempt to identify an element in RNA that ZAP recognized. A systematic evolution of ligands by exponential enrichment (SELEX) approach was used to identify the RNA sequence or structure that ZAP bound in RNA (Huang et al., 2010). Through 21 rounds of selection, the group identified a ZAP-binding aptamer that was guanosine rich and predicted top form stem loops. Despite the identification of these sequences, when they were placed in the context of luciferase reporter construct as done previously they did not confer sensitivity to ZAP, suggesting that this identified element might not be a bona-fida ZAP target. A structure of an N-terminal rZAP fragment was determined with the first 225 residues (Chen et al., 2012). This structure identified two RNA-binding cavities in ZAP, but a co-crystal structure with RNA bound was not solved. The researchers proposed key residues that are important for RNA binding and specific recognition of viral RNA, but without a co-crystal structure they could not further determine how ZAP is able to distinguish viral RNA from cellular RNA, other than through the complex RNA structures yet unsolved (Chen et al., 2012).

Mechanistically, once ZAP binds to a viral mRNA, it was clearly shown in the initial discovery that the RNA is degraded and the protein that it encodes is not expressed (Gao et al., 2002). Some work has gone into trying to characterize how the RNA is degraded

yet a clear consensus on how this is accomplished has not been reached. Co-sedimentation assays identified a number of components in the RNA exosome were pulled down with ZAP (Guo et al., 2007) The group went on to show that specifically the exosome component EXOSC4 (hRrp46p) directly interacts with ZAP in *in vitro* pull-down assays and that knock down of this component correlated with a significant decrease in ZAP activity against a Luciferase-SINV construct. Notably, the decrease in ZAP activity was only 2-fold, indicating that knockdown only confers a partial rescue to the RNA stability (Guo et al., 2007). A follow up studies similarly identified the protein p72 DEAD box RNA helicase (DDX17) and DEXH-Box protein (DHX30) as cofactors for ZAP (Chen et al., 2008; Ye et al., 2010). As with EXOSC4, when overexpressed, DDX17 and DEXH30 co-immunoprecipitated with ZAP, and when overexpressed they decrease the expression of a luciferase construct with a SINV fragment as a 3'UTR. Furthermore, when these proteins are knocked down the both decrease the effect ZAP has on the expression of the same reporter construct by 2-fold. While a luciferase is used as a control, crucially, these experiments do not include an experiment where these proteins are overexpressed or knocked down in the absence of ZAP to more precisely implicate their effects as being entirely ZAP dependent. The most recent study on the how ZAP degrades the RNA target came from a study characterize the relationship between ZAP degradation and translational repression (Zhu et al., 2012). The predominant view in the field is that translational repression of mRNA precedes RNA degradation by translocating the mRNA into processing bodies (p-bodies) where the ribonucleases degrade the mRNA (Iwakawa and Tomari, 2015). This experiments from this study are in agreement with this model,

demonstrating that ZAP excludes the target mRNA from polysomes and that this translational repression is independent from the mRNA degradation (Zhu et al., 2012). Taken together, these studies suggest that the exosome might be an important factor in the degradation of mRNA targeted by ZAP, but because key experiments are missing to definitely implicate them in the ZAP pathway combined with the lack of other studies from different groups that verify these results it remains unclear how exactly ZAP degrades the mRNA it targets.

In addition to cofactors that are downstream of ZAP and promote mRNA degradation, there has been interest in identifying a cofactor that is either necessary for or enhance the activity of ZAP. The first study that presented evidence that ZAP might require a cofactor for maximal activity came testing a variety of different cell lines ZAP activity in the presence or absence of IFN treatment. They found that BHK (a hamster kidney cell line) and mouse embryonic fibroblasts (MEF) have significantly less capable of restricting SINV replication when IFN is either not present or IFN signaling inhibited (MacDonald et al., 2007). Crucially, short hairpin RNA (shRNA) silencing of ZAP abrogated the effects of IFN on the replication of SINV, supporting the conclusion that the IFN effects observed were ZAP dependent. This study was followed up ten years later by the identification of a cofactor that fits this description through a genome wide RNA interference (RNAi) based screen. TRIM25, the E3 ubiquitin ligase previously shown to be a necessary cofactor for RIG-I, was found the enhance the antiviral activity of ZAP against SINV (Li et al., 2017). ZAP and TRIM25 were shown to directly interact via the SPRY domain in TRIM25, and TRIM25 demonstrated no antiviral activity against SINV in the absence of ZAP, with both

the RING domain, Coiled-coil domain, and SPRY domain all required for activity. Further, evidence is presented that TRIM25 ubiquitinates ZAP, yet ubiquitination of ZAP does not affect its activity. This compelling evidence presented in the initial study, was shortly followed up and confirmed by an additional group, further supporting the role of TRIM25 as an enhancer of the antiviral activity of ZAP (Zheng et al., 2017).

Since the initial discovery of ZAP and the following description of the viruses that ZAP restricts, additional viruses have been shown to be sensitive to ZAP. Viruses sensitive to ZAP have included Hepatitis B, specific spliced transcripts of HIV-1, Influenza A virus (IAV) and Japanese Encephalitis Virus (JEV) (Chiu et al., 2018; Mao et al., 2013; Tang et al., 2017; Zhu et al., 2011). Despite a broader understanding of the viruses that are sensitive to ZAP, this had not yielded a clearer understanding of how ZAP might restrict the replication of such a diverse group of viruses. Complementing these studies, there is some suggestion that viral proteins counteract ZAP. A study of HSV1 presented evidence the viral protein UL41 degrades the ZAP mRNA and significantly decreases its expression, thus allowing HSV1 the replicate. Removal of UL41 activity from HSV-1 sensitizes the virus to ZAP, while the wild-type virus can replication with normal kinetics (Su et al., 2015). Interestingly, this same group has published multiple reports of UL41 similarly degrading the mRNA of other known restriction factors such as cGAS, Viperin, and IFIT3 (Jiang et al., 2016; Shen et al., 2014; Su and Zheng, 2017). Additionally, two studies have demonstrated that ZAP might have some intracellular function beyond viral restriction. One study suggested that ZAP might inhibit retrotransposition, while another study

presented evidence that ZAP regulate the expression of the cellular mRNA TRAILR4 (Goodier et al., 2015; Moldovan and Moran, 2015; Todorova et al., 2014).

Due to the limited study of ZAP compared to other restriction factors, as of recently, there are still many outstanding questions related to how it recognizes self from non-self RNA, are there other cofactors that are required for activity, how is the RNA bound by ZAP degraded, and does ZAP possess some other cellular function beyond its role as a cytosolic viral sensing protein.

CG dinucleotide suppression

Over ten years ago, an interesting study was published that demonstrated radical alteration of the replication of poliovirus by simple re-coding the genome with synonymous mutations (Coleman et al., 2008). To do this, the authors took advantage of the redundancy of tRNAs that encode the same amino acid, whereby an Alanine can be encoded in RNA by either GCC or GCG, with the former occurring four times more frequently than the later in the human genome. Similarly, pairs of codons in the human genome occur at remarkably different frequencies. For example, the amino acid pair Ala-Glu could be encoded by either GCCGAA or GCAGAG, but the former codon pair is dramatically underrepresented in the human genome compared to the later. The group tested whether adding these rare codon pairs might affect the replication of poliovirus, and synthesized viral mutants with recoded capsid sequences with increasing amounts of rare codon pairs. Remarkably, the replication of the recoded virus was decreased by a 1,000-fold compared to the wild-type virus (Coleman et al., 2008). They concluded that

the cumulative addition of these rare codon pairs added up to dramatically inhibit the replication of the virus, and replicated these findings in other viral systems, notable with respiratory syncytial virus (Le Nouen et al., 2014).

Other groups attempted to recode RNA viruses with a similar rational, i.e. using synonymous mutations at the codon boundaries, yet they claimed they were changing different aspects of the viral sequence. Rather than the rare codon pair, per se, affecting the viral replication, they claimed what they were actually changing was the dinucleotide composition of the viral genome, specifically increase the CG and UA frequency (Atkinson et al., 2014). By specifically increasing the CG and UA Content of an echovirus, they observed a severe attenuation in the replication compared to the wild-type virus. The authors believed that the mechanism of attenuation was the same as the group that was recoding their viruses by codon pair basis. They published a follow up study that directly challenged codon pair biases, claiming that codon pair deoptimization is merely an artifact of increasing the CG and UA frequencies in the viral genomes (Tulloch et al., 2014). This drew a swift and critical response from the group who put forth the idea of codon pair deoptimization, with a letter in PNAS defending their theory over the competing dinucleotide enrichment theory (Futcher et al., 2015). Both sides presented informatics analysis that suggested that their theory was correct, but a third parties analysis finally brought resolution to the debate. Admittedly, they had set out to prove that codon pair bias was the driving factor in viral genomes rather than dinucleotide composition. To do this, they conducted an in-depth analysis of the codon pair basis of viral genomes and the genomes of the species the viruses infect to demonstrate a correlation between the

two. Remarkably, they did not find a correlation between codon pairs usage, but rather a strong correlation in dinucleotide composition, specifically in the frequency of CG and UA dinucleotides (Kunec and Osterrieder, 2016).

Vertebrate genomes exhibit marked CG-suppression, that is lower than expected numbers of 5'-CG-3' dinucleotides (Karlin et al., 1994). This feature is likely due to C-to-T mutations that have accumulated over hundreds of millions of years, driven by CG-specific DNA methyl transferases and spontaneous methyl-cytosine deamination. Remarkably, many RNA viruses of vertebrates that are not substrates for DNA methyl transferases mimic the CG-suppression of their hosts (Karlin et al., 1994; Karlin and Mrazek, 1997; Rima and McFerran, 1997). This striking property of viral genomes is unexplained. The HIV-1 genome is also particularly suppressed for CG dinucleotides and matches the CG dinucleotide composition of its host (Cheng et al., 2013; Futcher et al., 2015; Greenbaum et al., 2008). The evidence presented above suggests that this property of RNA viruses might be essential to their replication, but the reason for this suppression had previously not been identified.

Chapter 2. Materials and Methods

Plasmid construction

The mutated regions of the HIV-1 genome were synthesized (Genewiz) and cloned into HIV-1_{NHG} using restriction digest sites that were proximal to the mutated regions. Division of the original mutants blocks into two new derivative mutants was achieved using overlap extension PCR based approaches with mutant and WT templates. Revertant mutations acquired through passage of the virus were reconstituted into the original mutant provirus from which they arose through site directed mutagenesis and overlap extension PCR. A HIV-1_{NHG} reporter provirus was used in the initial construction of the synonymous mutants (GeneBank: JQ585717.1). This proviral construct contains a GFP reporter in the place of the non-essential gene, nef. This provirus is primarily comprised of NL4.3 sequence, with small segments of the genome derived from HXB2.

A ZAP exon 1-targeting guide sequence: 5'-GGCCGGGATCACCCGATCGGTGG-3' was inserted into a lentiviral based CRISPR plasmid from Addgene (52961) to generate $ZAP^{-/-}$ cells. A TRIM25 exon 1-targeting guide sequence: 5'-GAACACGGTGCTGTGCAACGTGG-3' was inserted into the same Addgene vector to generate TRIM25 knock out cell lines.

A ZAP-S and TRIM25 cDNA that was rendered resistant to the CRISPR resistant through introduction of synonymous mutations in the guide target sequence was generated by overlap extension PCR and inserted into a tetracycline inducible HIV-1 based vector (pLKO.dCMV.TetO/R). An epitope tagged (ZAP-3xHA) cDNA used for CLIP was inserted into the MLV expression vector, LHCX. A Firefly luciferase cDNA (*fluc*) was designed to

remove CG dinucleotides through synonymous substitution, reducing the bringing the total number of CG dinucleotides from 97 to 8. This CG-low *fluc* cDNA was inserted into the expression vector pCR3.1 using EcoRI and NotI. Various sequences were then inserted 3' to the *fluc* cDNA using NotI and XhoI. Specifically, sequences from the Indiana strain of VSV-G and P, and the Influenza A/WSN/1933 NP open reading frames were inserted, as were derivatives with synonymous mutations that maximized CG-dinucleotide content. A CXCR4-2A-CD4 cassette was generated by overlap-extension PCR and inserted into LHCX using HindIII and HpaI. The TRIM25 constructs were cloned into pCR3.1 using EcoRI and NotI restriction sites, and point mutations were introduced using overlap PCR.

Cell culture

The adherent cells 293T, HOS and HeLa were grown in DMEM with 10% fetal bovine serum. MT4 cells were cultured in RPMI supplemented with 10% fetal bovine serum. Primary lymphocytes were isolated from human blood by Ficoll-Paque gradient centrifugation and removal of the plastic adherent fraction. Cells were activated with phytohaemagglutinin (Sigma, 5 μg ml^{-1}) and cultured in the presence of interleukin-2 (50 U ml^{-1}) in RPMI with 10% fetal bovine serum.

ZAP-deficient cells were generated by transduction with the lenti-CRISPR vector followed by selection in 5 μg ml^{-1} blasticidin. Single-cell clones were derived by limiting dilution and maintained in the appropriate media with 5 μg ml^{-1} blasticidin. Loss of ZAP protein and gene was confirmed by PCR amplification and sequencing the genomic locus and by

western blotting. In some CRISPR knockout clones, protein species that reacted with an anti-ZAP antibody arose after extended passage and likely represent truncated forms of ZAP-L whose translation initiated at methionine codons 3' to the CRISPR target site (that was near the ZAP N-terminus). The appearance of these protein species did not affect the ability of the cells to support replication of WT or mutant viruses. Doxycycline-inducible ZAP-S expression in MT4 cells was reconstituted by stable transduction with the LKO ZAP-S expression vector followed by selection in 2.5 $\mu g\,ml^{-1}$ puromycin. These reconstituted cells were used as a pool. Cells (293T) stably expressing either ZAP-S 3xHA or ZAP-L 3xHA were generated by transduction with the LHCX vector containing followed by selection in 50 $\mu g\,ml^{-1}$ hygromycin. A single cell clone was derived by limiting dilution and maintained in DMEM with 50 $\mu g\,ml^{-1}$ hygromycin. HOS cells were stably transduced with LHCX CXCR4-2A-CD4 followed by selection in 25 μg/ml hygromycin. A single cell clone was derived by limiting dilution and maintained in the appropriate media with 25 μg/ml hygromycin.

Virus production

All HIV-1 virus stocks were generated by transfection of 293T cells 10cm dishes with 10μg of proviral plasmid using polyethyleneimine (Polysciences). HIV-1_{WT} and mutant viruses usually contained a GFP reporter and were generated by transfection with HIV-1_{NHG}-derived proviral plasmids. The yields of infectious virus from transfected 293T cells for each of the mutants was similar, despite their differing properties in spreading replication assays. This is very likely due to gross overexpression of the viral genome in

transfected 293T cells. Viruses used in primary lymphocyte replication and CLIP assays were generated by transfection with HIV-1_{NL4-3}. Viruses used to infect CD4-negative HeLa cells in the single cycle replication siRNA screen, or 293T cells in the CLIP assays, were generated by transfection with 10µg of proviral plasmid and 1µg of VSV-G expression plasmid.

Titers of viral stocks were determined by performing 3-fold serial dilutions in a 96 well plate and infecting $5x10^4$ MT4 cells per well. At 16-18 hours post infection, dextran sulfate was added to each well at a concentration of 50 µg ml^{-1} to prevent reinfection by nascent virions. At 48 hours after infection, cells were fixed in 4% PFA and enumerated by FACS analysis using a CyFlow Space cytometer (Partec) coupled to a Hypercyte Autosampler (Intellicyt).

For spreading replication assays with GFP reporter viruses, viral stocks generated from transfected 293T cells were adjusted to the same number of single cycle infectious units (determined on MT4 cells as described above). Thereafter, $2x10^5$ MT4 cells were infected at an MOI of 0.002 in 2 mL of RPMI. Aliquots of infected cells were withdrawn each day, fixed in 4% PFA and the proportion of infected cells determined by FACS analysis of GFP expression. For spreading replication assays in PBMCs, cells were infected at an MOI of 0.001. At 18 hours post infection the cells were washed four times with PBS and cultured in RPMI with 50 U ml^{-1} interleukin-2. Supernatants were collected every 24 hours. Viral particle release was determined by measuring the reverse transcriptase activity with a PCR based assay, as previously described (Pizzato et al., 2009).

For single cycle replication MT4 cells were infected at an MOI of 1.0, with HIV-1_{NHG}-derived viruses, washed three times with PBS 18 hours post infection, and resuspended in RPMI with 50 μg ml^{-1} of dextran sulfate to prevent reinfection. At 48 hours post infection, cells and supernatants were collected for analysis. Half of the cells were lysed in SDS sample buffer for western blot analysis and half allocated for RNA extraction and to determine levels of unspliced RNA as described below. The supernatants were filtered with a 0.22 μm filter. An aliquot of filtered supernatant was used to determine infectious viron yield by titration on MT4 cells. The remaining supernatant was centrifuged over a 20% sucrose cushion at 14,000 rpm at 4^{o} C for 90 minutes. Pelleted virions were resuspended in SDS sample buffer for western blot analysis.

Western blot assays

Cells were counted, normalized for cell number, lysed in SDS sample buffer, separated by electrophoresis on NuPage 4-12% Bis-Tris gels (Novex) and blotted onto nitrocellulose membranes (GE Healthcare). Antibodies for PTBP1 (ab5642), Drosha (ab12286), DICR (ab14601), EXOSC6 (ab50910), EXOSC10 (ab50558), and PARN (ab188333) were obtained from Abcam. Antibodies for Upf1 (A300-036A), METTL3 (A301-567A), EXOSC4 (A303-775A), EXOSC5 (A303-887A), and Xrn1 (A300-443A) were obtained from Bethyl Labs. The antibody against ZAP (16820-1-AP) was obtained from Proteintech. The HIV-1 capsid antibody (183-H12-5C) was obtained from the NIH AIDS reagent repository. The GFP (G1546) antibody was obtained from Sigma. The HIV Env (12-6205-1) antibody was

obtained from American Research Products. The HA (HA.11) antibody used in the CLIP assays was obtained from Biolegend.

Cross-linking Immunoprecipitation and Sequencing (CLIP-Seq)

The CLIP method used in this study has been previously described (Kutluay et al., 2014) and was adapted from previously reported HITS-CLIP and PAR-CLIP(Hafner et al., 2010)(Hafner et al., 2010)(Hafner et al., 2010)(Hafner et al., 2010)(Hafner et al., 2010)(Hafner et al., 2010)(Hafner et al., 2010)(Hafner et al., 2010) protocols (Hafner et al., 2010; Licatalosi et al., 2008). In brief, RNA and proteins were cross linked by feeding cells overnight with 4-thiouridine irradiating them at 0.15 J/cm^2 UV (λ = 365 nm) in a Stratalinker 2400 UV crosslinker (Stratagene). Thereafter ZAP-3xHA was immunopurified using Protein G-conjugated magnetic Dynabeads and a mouse monoclonal anti-HA antibody, and the RNA was radiolabeled with 0.5 µCi/µl γ-^{32}P[ATP] ATP. Protein-RNA adducts were separated by SDS-PAGE, transferred to nitrocellulose and detected by autoradiography. Next, sequential 3' and 5' adaptor ligations were performed as previously described (Kutluay 2014) attaching a known sequence that contains primer binding sites for reverse transcription and PCR-amplification of the cDNA library. Sequencing of the cDNA library was done on an Illumina HiSeq 2000 platform.

The analysis pipeline used in this study has previously been described. Processing of raw reads was performed with the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), excluding reads with fewer than 15 nucleotides. Reads were then aligned to the human genome (hg38) concatenated with the HIV-1$_{NL4-3}$ genome or to the viral genome alone. Cluster analysis was performed using PARalyzer (Corcoran et al., 2011).

siRNA transfection assays

ON-TARGET plus siRNAs were ordered from Dharmacon, and contain 4 individual siRNAs against the target gene to ensure efficient knockdown. Stock concentrations were made at 100uM. Working dilutions were prepared at 10uM and stored in small aliquots to prevent freeze thawing. 5uL of the 10uM siRNA was added to 200uL of Opti-MEM (Gibco), separately, 5uL of RNAiMAX (ThermoFisher) was added to a 200uL of Opti-MEM. After mixing, the siRNA and RNAiMAX Dilutions are added to a 6 well plate, mixed, and incubated for 5-10 minutes. After incubation, $1x10^5$ HeLa cells were added in DMEM with 10% FBS. 24 hours post transfection, the cells were trypsanized and reseeded into two wells of a 12-well plate. 48 hours post transfection, the cells were infected with either HIV-1_{NHG} or a CG-high HIV-1_{NHG}. 16-18 hours post infection the cell were washed three times with sterile PBS, and replaced with fresh DMEM with 10% FBS. 48 hours post infection the supernatant was filtered and collected to determine the viral titer released. Additionally, the infected cells were collected for western blot analysis.

The siRNA screens were generated using cherry-pick siRNA libraries from Dharmacon, also containing ON-TARGET plus siRNAs with 0.5nmol per well. Stock concentration plates were made at 10uM and working dilution plates were made at 1uM to avoid freeze thawing. To transfect in 96 well plates, 20 µL of Opti-MEM was added to each well of a 96-well plate with a multichannel. 5uL of the respective siRNA is added to the 96-well plate. Separately, a master mix is made with 23.5 µL Opti-MEM and 1.5 µL RNAiMAX /

well, mixed thoroughly, added to the 96well plates that already contain the siRNA and Opti-MEM, and left of incubate for 5-10 minutes at room temperature. Separately, HeLa cells are prepared and four 96-well plates were seeded with $8x10^4$ cells in 100 µL of DMEM with 10% FBS. After plating Hela cells, 10uL of the siRNA transfection mixture is added, created enough plates for two virtues infections done in technical duplicate. 48 hours post transfection, the HeLa cells are infected at an MOI of 1.0 in 50 µL of DMEM with 10% FBS. 16-18 hours post infection, the media aspirated and replace with fresh DMEM with 10% FBS. 48 hours post infection, the supernatants are collected and added to $5x10^4$ MT4-GFP cells to determine the titer of the produced virus. Additionally, the infected HeLa cells from the screen are trypsinized and fixed in 4% PFA for analysis by flow cytometry to determine what percentage of the cells were infected, using the GFP reported in the HIV-1 viruses in the place of the nef gene.

Analysis of HIV-1 splicing with fluorescent primer PCR

RNA from 293T cells transfected with mutant provirus was extracted using the Nucleospin RNA extraction kit (Machery Nagel). RNA was reverse transcribed using SuperScript III reverse transcriptase (ThermoFisher) with gene specific primers for either fully spliced (8483R: 5'-CCGCAGATCGTCCCAGATAAG-3' and partially spliced (6223R: 5'-CAAGTGCTGATATTTCTCCTTCAC -3') mRNA classes. The cDNA templates were then used in a 10µL PCR reaction with fluorescent reverse primers specific to the splice class (labelled at their 5' ends with IRD800) and a forward primer position 5' to the major splice donor (499F: 5' -CTGAGCCTGGGAGCTCTCTGGC-3') and run for 25 cycles with an

annealing temperature of 54°C. Alternatively, to determine use of the activated cryptic splice site in mutant A, a forward primer, positioned 5' to the mutations (763F 5'-TGACTAGCGGAGGCTAGAAGGAGAGAG -3') and the fluorescent reverse primer for the fully spliced class (8483R) were used in a PCR reaction with the cDNA templates. To determine use of the activated cryptic splice site in mutant B, the forward primer 499F and a fluorescent reverse primer 5' to the mutations in B (1557R 5'-GATAGGTGGATTATGTGTCATCC -3') were used in a PCR reaction with the cDNA template. Then, 10µL of 2x TBE-Urea sample buffer was added to the PCR reaction which was then run on a 6% TBE-Urea gel for 90 minutes at 180V (Novex). A LI-COR Odyssey scanner was used to detect fluorescent signals directly from the gels.

Primer-ID-based deep sequencing analysis

Determination of splice site utilization using the Primer ID-based deep sequencing assay was done substantially as described with minor modifications (Emery et al., 2017). Briefly, RNA was extracted from cells transfected with HIV-1_{NHG} or mutants thereof using the RNeasy Plus minikit (Qiagen). Primers used for cDNA synthesis were GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNCAGTTCG GGATTGGGAGGTGGGTTGC for 1.8 kb spliced transcripts and GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNGCTACTA TTGCTATTTGTATAGGTTGCATTACATG for 4 kb spliced transcripts. Indexed primers were obtained from Integrated DNA Technologies Custom Oligos. Total cell RNA (8 µg) was subjected to cDNA synthesis, purification and cleanup and an initial PCR

amplification. An aliquot of the product from this first PCR was used as the template for the second PCR to add the Illumina adapter and bar codes to allow multiplexing in the Illumina sequencing reaction. PCR products were visualized on a 2% agarose gel and then cleaned. Libraries were mixed/multiplexed and sequenced using the 300-base paired-end read for the Illumina Miseq platform. Reads were sorted using the Illumina bcl2fastq pipeline (v.1.8.4) to separate the multiplexed samples. Subsequent Data analysis and splicing quantification were done using the previously described in-house pipeline (Emery et al., 2017) written in Ruby and adapted to accommodate the mutated sequences. This program uses the combined sequence information from the paired-end reads to identify splice site usage and transcript type. Reads are condensed by Primer ID to prevent skewing in the PCR steps. Cryptic alternative donor and acceptor splice sites were identified using a program that compares data reads to a reference sequence and identifies the base where a splice discontinuity occurs and the base it splices to.

qPCR quantification of unspliced viral RNA

RNA was collected from infected cells using the Nucleospin RNA kit (Macherey Nagel). The RNA concentration was determined using a NanoDrop 2000c (ThermoFisher). Equal amounts of RNA were reverse transcribed using SuperScript III reverse transcriptase with random hexamer priming (ThermoFisher). The cDNA was used as a temple for quantitative real-time PCR using a StepOnePlus RT-PCR machine (Applied Biosystems). Unspliced viral RNA was detected by a TaqMan probe spanning the major splice donor D1, using the Fast Start TaqMan Probe master-mix. Serial tenfold dilutions of known copy

numbers of HIV-1_{NHG} plasmid was used to generate a standard curve. The sequence of the TaqMan probe and primers were: D1 probe: 5'-GGGCGGCGACTGGTGAGT-3'; forward primer: 5'-GGACTTGAAAGCGAAAGGGA-3'; reverse primer: 5'-TCTCTCTCCTTCTAGCCTCCG-3'.

Small molecule fluorescence in-situ hybridization microscopy

HOS CXCR4-2A-CD4 were seeded onto gelatin coated glass bottom 24 well plates (MatTek) and infected at an approximate MOI of <1 with HIV-1_{WT}, $L_{CG\text{-}HI}$, or $L_{GC\text{-}HI}$. Twenty-eight hours after infection the cells were washed with PBS and fixed with 4% paraformaldehyde (Thermo) in PBS for 30 min at RT. After permeabilization with 70% ethanol for 2hr at RT the cells were briefly washed with Stellaris RNA-FISH wash buffer A for 5 min at room temperature. The cells were then incubated with custom Stellaris smFISH probes targeting HIV-1 *gag* or all viral mRNAs (Biosearch Technologies) at a concentration of 0.125 μM in Stellaris RNA FISH hybridization buffer for 16-18 hr at 37°C. The cells were then washed two times for 30 min at 37°C in Stellaris RNA FISH wash buffer A. The second wash contained Hoechst dye at 1 μg/ml. After a 5 minute wash with Stellaris RNA FISH wash buffer B cells were rinsed three times with PBS and imaged by deconvolution microscopy (Deltavision). Image stacks were generated by maximum intensity projection using the Z project function in ImageJ (Version 2.0.0-rc-59/1.51n). RNA spots were counted using Find Maxima function in ImageJ.

Chapter 3. Synonymous mutagenesis of the HIV-1 genome

To determine whether there are important, undiscovered cis-acting elements in the HIV-1 genome, a mutant HIV-1 sequence was designed that contained a maximum number of synonymous mutations in the open reading frames while leaving known RNA elements that are important for virus replication intact (Table 3.1).

Design of global synonymous mutants

Mutations were designed so as to maximize the probability of disrupting base pairing in which a given nucleotide might participate. Thus, where multiple synonymous mutation possibilities were available, transversion mutations (purine to pyrimidine or vice versa) were preferred over transition mutations. To avoid the creation of new splice acceptors and donors, no new AG or GU dinucleotides were introduced. Moreover, sequences encoding overlapping open reading frames were not altered, and all known RNA elements that control HIV-1 splicing, gene expression and reverse transcription remained intact in the mutant viral genome. Mutations were designed so as to maximize the probability of disrupting base pairing in which a given nucleotide might participate. Thus, where multiple synonymous mutation possibilities were available, transversion mutations (purine to pyrimidine or vice versa) were preferred over transition mutations. To avoid the creation of new splice acceptors and donors, no new AG or GU dinucleotides were introduced. Moreover, sequences encoding overlapping open reading frames were not altered, and all known RNA elements that control HIV-1 splicing, gene expression and reverse transcription remained intact in the mutant viral genome (Table 3.1).

This designed HIV-1 sequence contained 1,976 synonymous mutations and was divided into 150-500 nucleotide blocks, that were synthesized separately. Each synthetic mutated fragment was introduced into a replication competent HIV-1 proviral plasmid ($HIV\text{-}1_{NHG}$) that carried GFP in place of the nonessential gene Nef. Thus, sixteen different mutated proviral plasmids, designated A through P, with each carrying a mean of ~125 synonymous mutations were generated (Figure 3.1A).

Table 3.1 : Sequences containing known cis-acting elements left undisturbed in synonymously mutated HIV-1

Site name	Position
Splice sites	
D1	744
A1	4911
D2	4963
D3	5464
A2	5388
A3	5775
A4c	5935
A4a	5952
A4b	5958
A5	5974
D4	6045
A7	8373
Splicing regulators	
ESS2b	4988-5007
ESSV	5422-5436
ESS2	5846-5855
ESS3	8450-8465
ESS2P	5779-5785
ESE2b	5008-5032
ESE vpr	5437-5460
ESE tat	5807-5837
(GAA)3	8418-8428
ESE2	5838-5845
ESEM1	4933-4939
ESEM2	4956-4962
ESE vif	4918-4927
ISS	8331-8353
SD2 G4	4968-4971
Other	
TAR, Poly A, Psi.	1-811
GagPol frameshift	2085-2136
RRE	7705-8059
3'PPT	9553-9586
cPPT	4781-4799
CTS	4883-4898

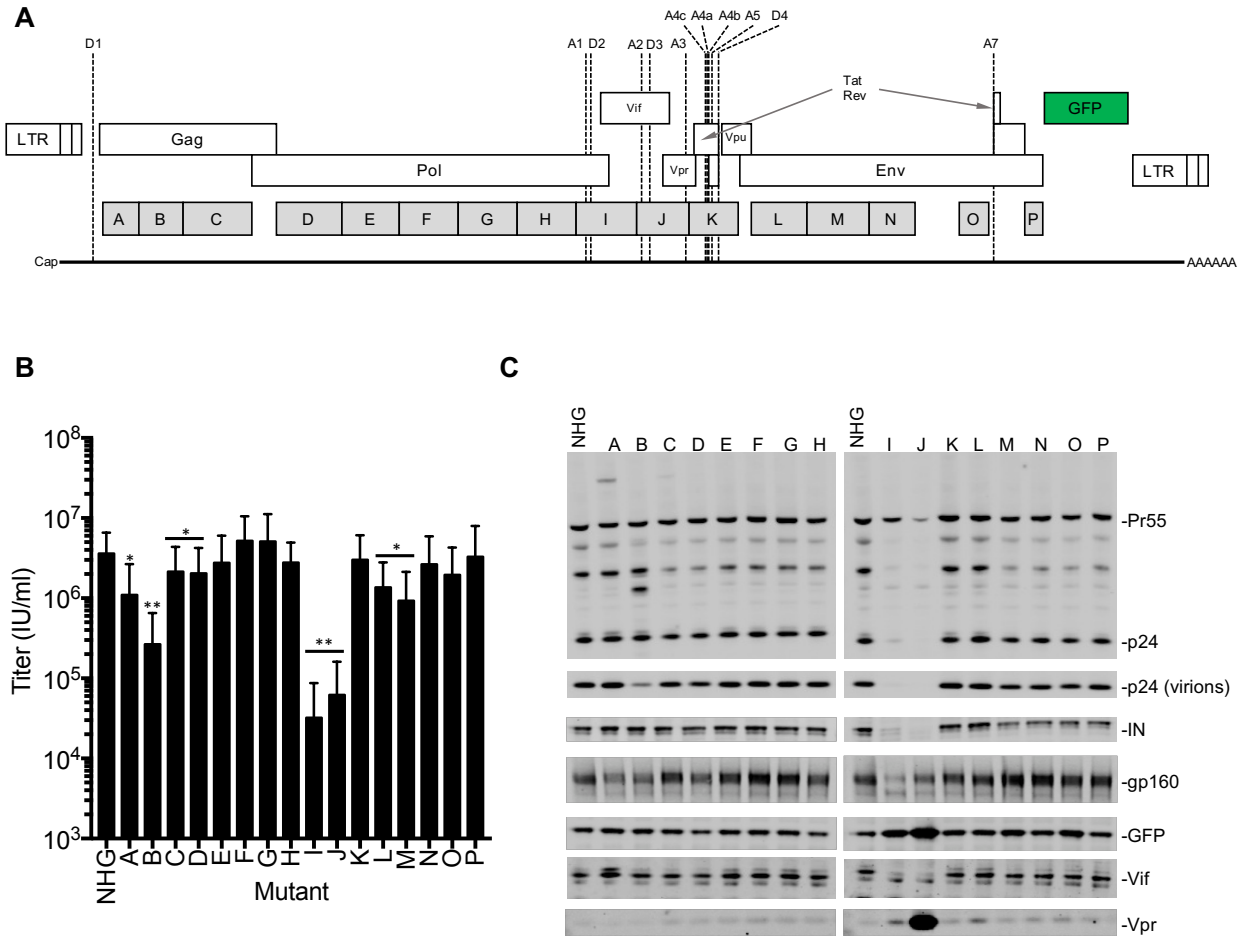


Figure 3.1: Design and analysis of panel of synonymously mutated HIV-1 viruses

(A) Schematic of HIV-1 proviral DNA, indicating open reading frames, splice sites, and blocks of nucleotides that were synonymously mutated in the 16 proviral plasmids (A-P). **(B)** Single-cycle infectious titers (measured using MT4 cells) 48h following transfection of 293T cells with each of the WT(HIV-1_{NHG}) and mutant (A-P) proviral plasmids. Values are given as mean±sd. (n=3) *$p<0.05$, **$p<0.005$ by students t-test, calculated with relative values normalized to WT values in each experiment. **(C)** Western blot analysis of protein levels in transfected cells (and virion particles where indicated) at 48h after transfection of 293T cells with each of the WT(HIV-1_{NHG}) and mutant (A-P) proviral plasmids.

Viral replication analysis of panel of synonymous mutants

Each of the synonymously mutated HIV-1_{NHG} proviral plasmids was transfected into 293T cells and the infectious virion yield was determined in a single-cycle infection assay in MT4 cells (Figure 3.1B). Many of the mutants yielded WT, or close to WT, levels of infectious virions in this transfection/titration assay format. However, mutants A, B, I, and J yielded between 5-fold and 1000-fold fewer infectious virions (Figure 3.1B). Western blot analysis of the transfected 293T cells and extracellular virions showed that mutants A and B expressed additional Gag protein species of unexpected sizes, and mutant B displayed a particle release defect, possibly a consequence of the expression of the aberrant Gag protein (Figure 3.1C). Mutants I and J displayed reduced Gag, Pol, Env and Vif protein, and slightly elevated GFP levels. Mutant J also grossly overexpressed the Vpr protein (Figure 3.1C).

We next examined whether each of the mutants could replicate in MT4 and CEM T-cell lines (Figure 3.2 and Figure 3.3). Seven of the mutants (D, E, F, G, H, N and P) replicated with WT, or close to WT, kinetics while eight other mutants, (A, B, I, J, K, L, M, and O) were replication defective, or highly impaired, in both cell lines (Figure 2.2 and Figure 2.3). Mutant C was somewhat impaired in MT4 cells, but replicated with close to WT kinetics in CEM cells. Thus, an apparent discrepancy was evident in the ability of some of the mutants to generate infectious virions in 293T cells, versus their ability to generate a spreading infection in T-cell lines.

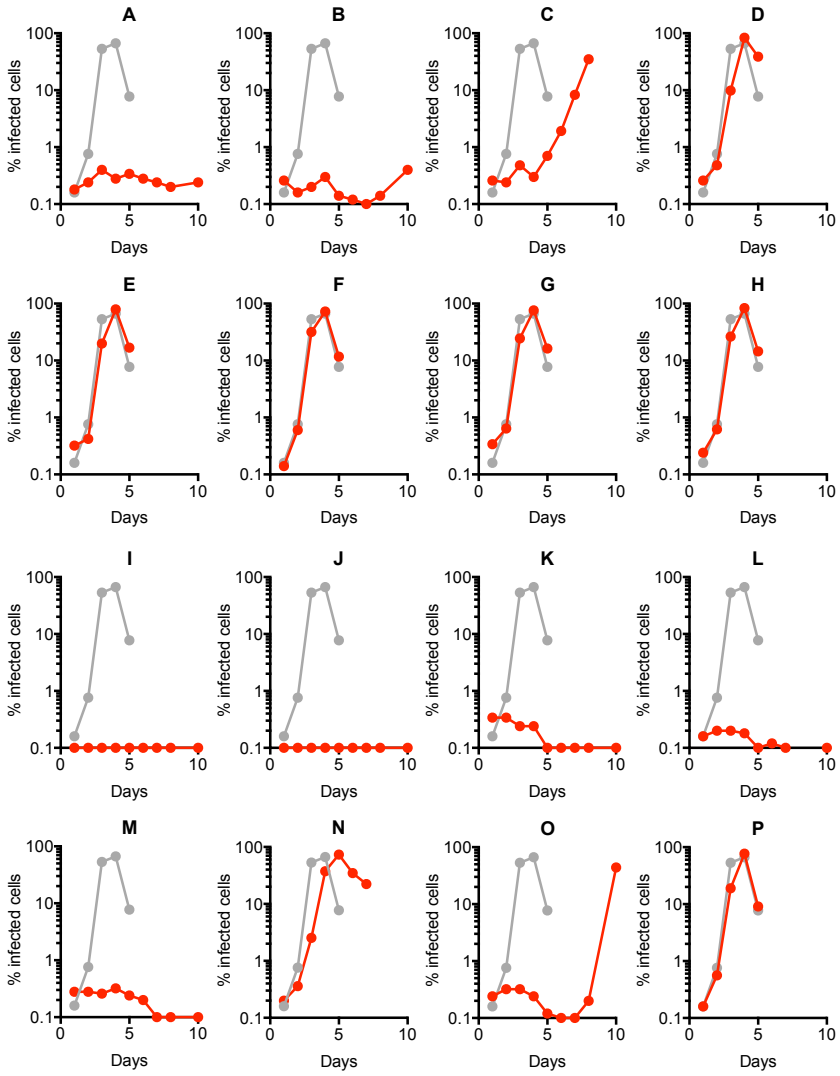


Figure 3.2: Spreading replication properties of mutant viruses

(A-P) MT4 cells were infected with the indicated virus (harvested from the supernatant of 293T cells transfected with each of the WT(HIV-1_{NHG}) or mutant (A-P) proviral plasmids at an MOI of 0.002. Aliquots of infected cells were withdrawn each day, fixed in 4% PFA and the proportion of infected cells determined by FACS analysis of GFP expression. a Representative replication curve for WT(HIV-1_{NHG}) is plotted in each chart as grey symbols and line, while each mutant is plotted using red symbols and lines.

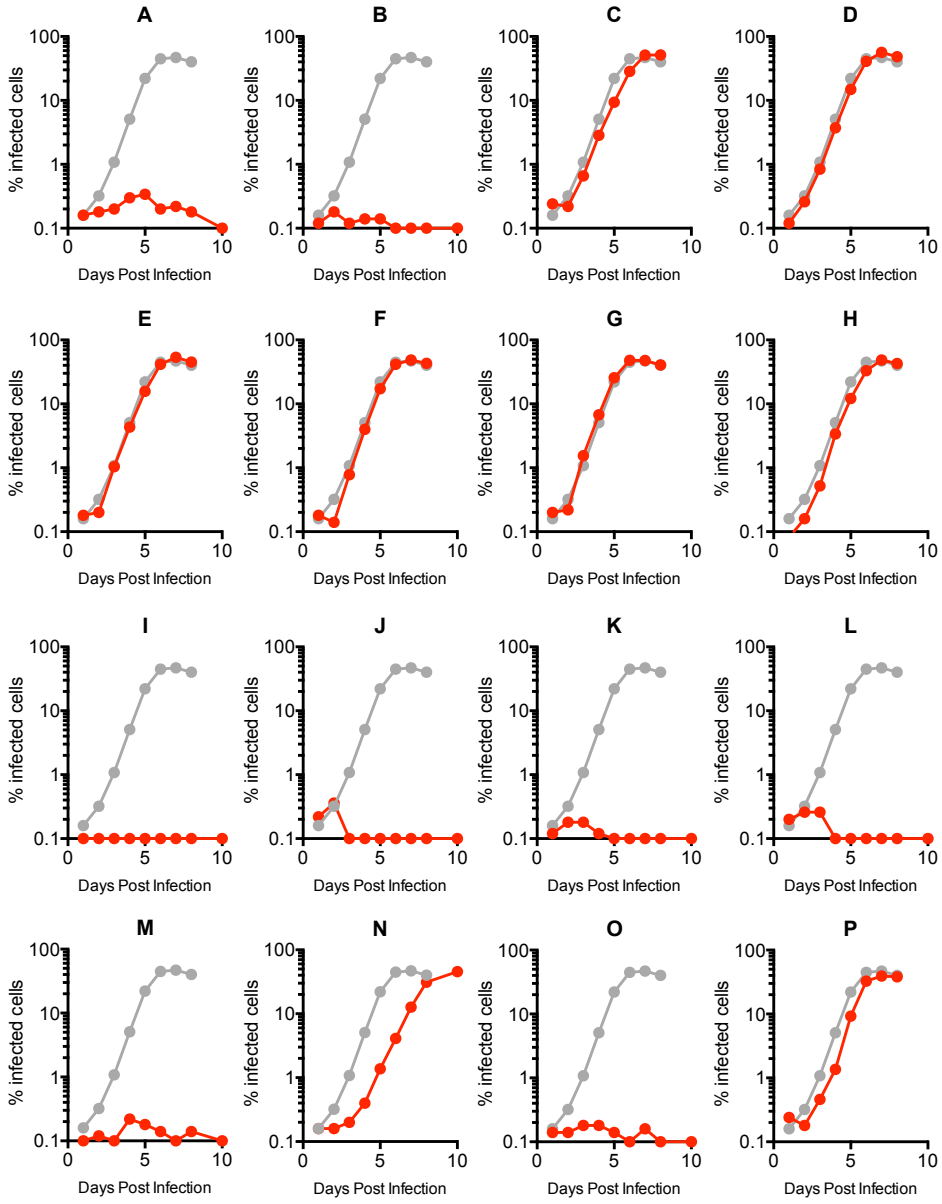


Figure 3.3: Spreading replication properties of mutant viruses in CEMx174 cells

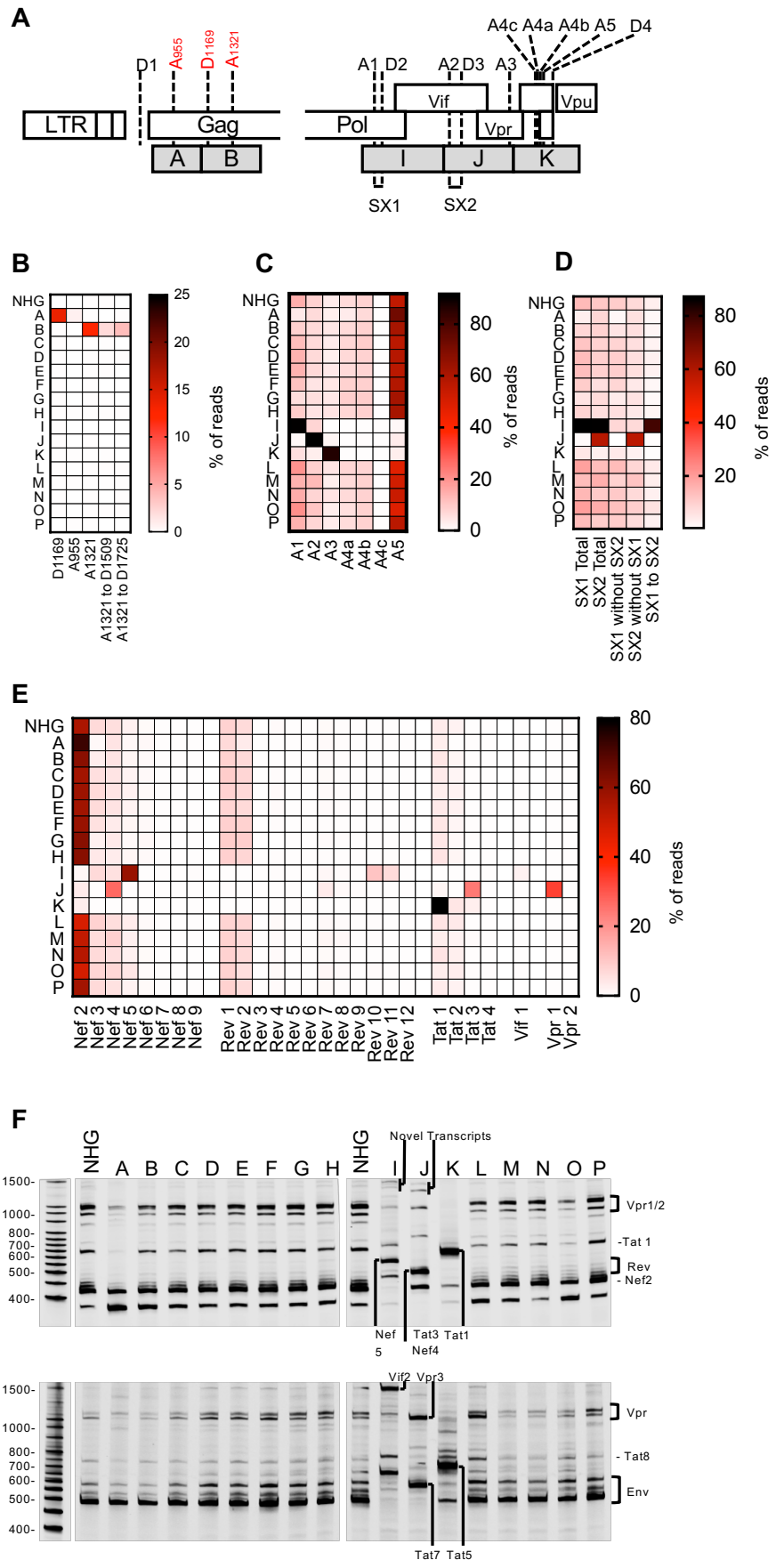
(A-P) CEMx174 cells were infected with the indicated virus (harvested from the supernatant of 293T cells transfected with each of the WT(HIV-1_{NHG}) mutant (A-P) proviral plasmids at an MOI of 0.002. Aliquots of infected cells were withdrawn each day, fixed in 4% PFA and the proportion of infected cells determined by FACS analysis of GFP expression

Splicing defects detected by multiple assays

The viral mutants were designed to avoid altering RNA sequences in the HIV-1 genome that are known to be important for replication, including those that participate in or regulate splicing (Table 3.1). Nevertheless, the aberrant pattern of viral protein expression in two of the synonymously mutated viruses (I, J), and the appearance of novel Gag-related protein species in two others (A, B), suggested that HIV-1 splicing might have been perturbed in at least some of the mutant viruses (Figure 3.1C). Therefore, we next used two approaches to determine whether the mutations affected splicing in each of the mutant viruses. We used a recently described Primer ID-based deep sequencing approach to globally quantify the relative utilization of the various splice donors and acceptors in the mutant viruses (Figure 3.4A-E). We also used a fluorescent primer-based PCR-PAGE assay to more conveniently, albeit less quantitatively, track the generation of the major mRNA species in the canonical spliced 1.8 kb class of HIV-1 mRNAs (Figure 3.4F). These two assays yielded results that were in good agreement. Of the canonical splice acceptors in the central portion of the genome (A1, A2, A3 A4a,b,c, and A5, Figure 3.4A), the WT HIV-1_{NHG} most frequently spliced to A5, with lower levels of splicing to A1, A2, A3, A4a,b,c (Figure 3.4A, C, F).

Figure 3.4: **Analysis of HIV-1 splicing in WT and synonymously mutated HIV-1**

(A) Schematic representation of segments of HIV-1 proviral DNA, focused on mutants exhibiting perturbed splicing. Canonical splice sites (black) and cryptic splice sites (red) are indicated, as are blocks of nucleotides that were synonymously mutated in the viruses exhibiting perturbed splicing. **(B-E)** Nextgen sequencing analysis of HIV-1 splicing, heatmaps indicate relative proportion of sequencing reads that indicate splicing at the sites indicated at the bottom of the heatmaps (B, C), or inclusion of the short exons (SX1 and/or SX2) indicated at the bottom of the heatmap (D). For panel (C) only direct splicing to the indicated acceptor sites is indicated in the heatmap Alternatively, the relative abundance of the various 1.8 kb mRNA species is indicated (E). **(F)** Fluorescent primer PCR analysis of HIV-1 splicing. 293T cells were transfected with the indicated proviruses, RNA extracted and cDNA synthesized. A sense PCR primer situated 5' to the major splice donor, along with an antisense primer positioned either 3' to A7 or 3' to D4 (labelled with IRD800) were used to amplify cDNAs derived from the 1.8 kb (top) or 4 kb (bottom) classes of spliced HIV-1 mRNAs respectively. PCR products were subjected to PAGE and a LI-COR Odyssey scanner was used to detect fluorescent signals directly from the gels.


A
D1
A955
D1169
A1321
A1 D2
A2 D3
A3
A4c A4a A4b A5
D4
LTR
Gag
Pol
Vif
Vpr
Vpu
A
B
I
J
K
SX1
SX2
B
C
D
NHG
% of reads
D1169
A955
A1321
A1321 to D1509
A1321 to D1725
A1
A2
A3
A4a
A4b
A4c
A5
SX1 Total
SX2 Total
SX1 without SX2
SX2 without SX1
SX1 to SX2
E
Nef 2
Nef 3
Nef 4
Nef 5
Nef 6
Nef 7
Nef 8
Nef 9
Rev 1
Rev 2
Rev 3
Rev 4
Rev 5
Rev 6
Rev 7
Rev 8
Rev 9
Rev 10
Rev 11
Rev 12
Tat 1
Tat 2
Tat 3
Tat 4
Vif 1
Vpr 1
Vpr 2
F
NHG
Novel Transcripts
1500-
1000-
800-
700-
600-
500-
400-
Vpr1/2
-Tat 1
Rev
- Nef2
Nef 5
Tat3
Nef4
Tat1
Vif2
Vpr3
Vpr
- Tat8
Env
Tat7
Tat5

A splicing defect was observed for mutants A and B, which contained synonymous changes toward the 5' end of the HIV-1 genome, within *gag*. For each of these mutants, the relative levels of canonical splice site utilization were only marginally perturbed, but the Primer ID-based sequencing assay revealed that cryptic splice sites near the 5' end of the genome were activated (Figure 3.4B, C, D). These mutants were designated Group 2a. For mutant A, a cryptic splice acceptor at position 955 (A_{955}) and a donor at position 1169 (D_{1169}) were activated, neither of which was used at a measurable level in WT HIV-1_{NHG} (Figure 3.4A, B). While A_{955} was used rarely (~1% of 1.8 kb mRNAs) in mutant A, ~14% of the 1.8 kb mRNAs were spliced using D_{1169}. (Figure 3.4B). Splicing events involving D_{1169} were selective with respect to which of the downstream acceptors were used: A3, A4, or A5, were used as acceptors for D_{1169} but A1 or A2 were not. MaxEnt scoring, which employs an *in silico* analysis tool that predicts the intrinsic splicing efficiency of splice acceptors and donor sequences (Yeo and Burge, 2004), indicated that our mutagenesis increased the score of the cryptic acceptor A955 from -9.88 to 1.75, suggesting that it became a stronger splice acceptor (Table 3.2). Thus, the minimal activation of A955 (used in 1% of spliced reads) in mutant A, could possibly be explained by a direct effect of the mutations. However, the predominant defect in mutant A was the activation of D_{1169}, which lies outside the mutated region (Figure 3.4A) and whose MaxEnt score was not altered. Thus, activation of D_{1169} could not be due to increased intrinsic efficiency of this cryptic donor, but rather due to some other mechanism acting via RNA sequences distal to D_{1169}.

Table 3.2: MaxEnt scores of splice acceptors and donors in wild type and mutated sequence

Donor	MaxEnt Score	Sequence
D1	10.1	CTGGTGAGT
D1169 (wt)	7.16	CAGGTCAGC
D1509 (wt)	-4.77	CTAGTACCC
D1725 (wt)	7.3	GAGGTAAAA
D1a (wt)	10.77	CAGGTAAGA
D1a (mut)	5.28	CAGGTTCGC
D2 (wt)	5.73	AAGGTGAAG
D2b (wt)	5.99	CAGGTGATG
D3 (wt)	9.45	AAGGTAGGA
D3 (G5463A mut)	3.63	AAAGTAGGA
D4 (wt)	9.07	GCAGTAAGT
Acceptor	**MaxEnt**	**Sequence**
A955 (wt)	-9.88	CCTTTTAGAGACATCAGAAGGCT
A955 (mut)	1.75	GCTACTTGAAACTTCTGAAGGAT
A1231(wt)	-2.94	GTTTTCAGCATTATCAGAAGGAG
A1231 (mut)	7.03	GTTCTCTGCTCTTTCTGAAGGCG
A1231 (mut, T1311C)	6.09	GTTCTCTGCCCTTTCTGAAGGCG
A1a (wt)	2.47	ATACTTCCTCTTAAAATTAGCAG
A1a_mut	-3.69	TTATTTTCTACTTAAACTTGCTG
A1 (wt, IB mut)	6.41	AATTTTCGGGTTTATTACAGGGA
A1 (I, IA mut)	7.41	AATTTTCGCGTTTATTACAGGGA
A1 (IA mut, T4904A)	6.57	AATTTTCGCGTATATTACAGGGA
A1 (IB mut, G4912A)	-2.34	AATTTTCGGGTTTATTACAAGGA
A2 (wt)	9.43	CTATTTTGATTGTTTTTCAGAAT
A2 (J mut)	9.71	TTACTTCGATTGTTTTTCAGAAT
A3 (wt, K mut)	9.76	CTGCTGTTTATCCATTTCAGAAT
A3 (K mut, C5774T)	10.05	CTGCTGTTTATCCATTTTAGAAT
A5 (wt)	4.01	TTAGGCATCTCCTATGGCAGGAA
A7 (wt)	7.15	ATTCACCATTATCGTTTCAGACC

For mutant B, the major perturbation was activation of a cryptic splice acceptor at position 1321 (A_{1321}, Figure 3.4A, B). This splicing event, involving ~12% of mRNAs in the 1.8 kb class, appeared to enable a cascade of further alternative splicing events, 50% of which subsequently involved cryptic splice donor activation at either D_{1509} or D_{1725}, outside the B mutant region (Figure 3.4B). However, the most obvious outcome of the aberrant splicing event was the generation of a truncated ~40 kD Gag protein (Figure 3.1C). This protein would be the expected translation product of an mRNA in which a D1- A_{1321} splice, and no further splicing, had occurred. Specifically, translation initiation at the second Met codon in the *gag* gene would generate a truncated ~40 kD Gag protein lacking MA, that likely accounts for the aberrant band on the cell-associated Gag western blot as well as the reduced particle yield from cells transfected with the B mutant proviral plasmid (Figure 3.1C).

For mutant B, the activation of A_{1321} may be due to a direct effect of the mutations increasing its MaxEnt score from -2.94 to 7.03, again suggesting it became a stronger splice acceptor (Table 3.2). The other cryptic sites activated, D_{1509} and D_{1725}, are both outside of the mutated region in mutant B, and their use was likely secondary to activation of A_{1321}.

Viruses containing mutations in the central portion of the genome, specifically mutants I, J and K, exhibited a different type of splicing defect, and were designated Group 2b. Specifically, mutants I, J and K exhibited increased direct splicing to canonical splice acceptors, A1, A2 and A3 respectively, at the expense of direct splicing to downstream (3') acceptors (Figure 3.4C). In the case of mutant I, the primary defect (increased use of

A1) was accompanied by increased use of the proximal downstream splice donor (D2) as well as a downstream acceptor–donor pair (A2 and D3) and thus the abundant inclusion of short exons (SX) 1 and 2 (SX1= [A1-D2] and SX2 = [A2-D3]) into spliced viral mRNAs (Figure 3.4D). Some elevation of the use of upstream cryptic splice sites in mutant I (D_{3569}, D_{3969}, D_{4641} and A_{4834}) also generated low levels of novel transcripts (Figure 3.4F). For mutant J, overuse of A2 (which is positioned 3' to SX1) was accompanied by overuse of proximal downstream splice donor D3 and thus overrepresentation of SX2 = [A2-D3] into spliced viral mRNAs (Figure 3.4D). Additionally, some utilization of cryptic splice donors (D_{5052}, D_{5434}, and D_{5478}) generated low levels of novel transcripts (Figure 3.4F). For mutant K, overuse of A3 (which is positioned 3' to SX1 and SX2) did not result in the more frequent inclusion of these short exons (Figure 3.4D). Overall therefore, it appeared that one consequence of the overuse of a given splice acceptor (A1 or A2), was a resultant overuse of the proximal downstream splice donor (D2 or D3), consistent with an 'exon definition' model of splicing control (discussed later).

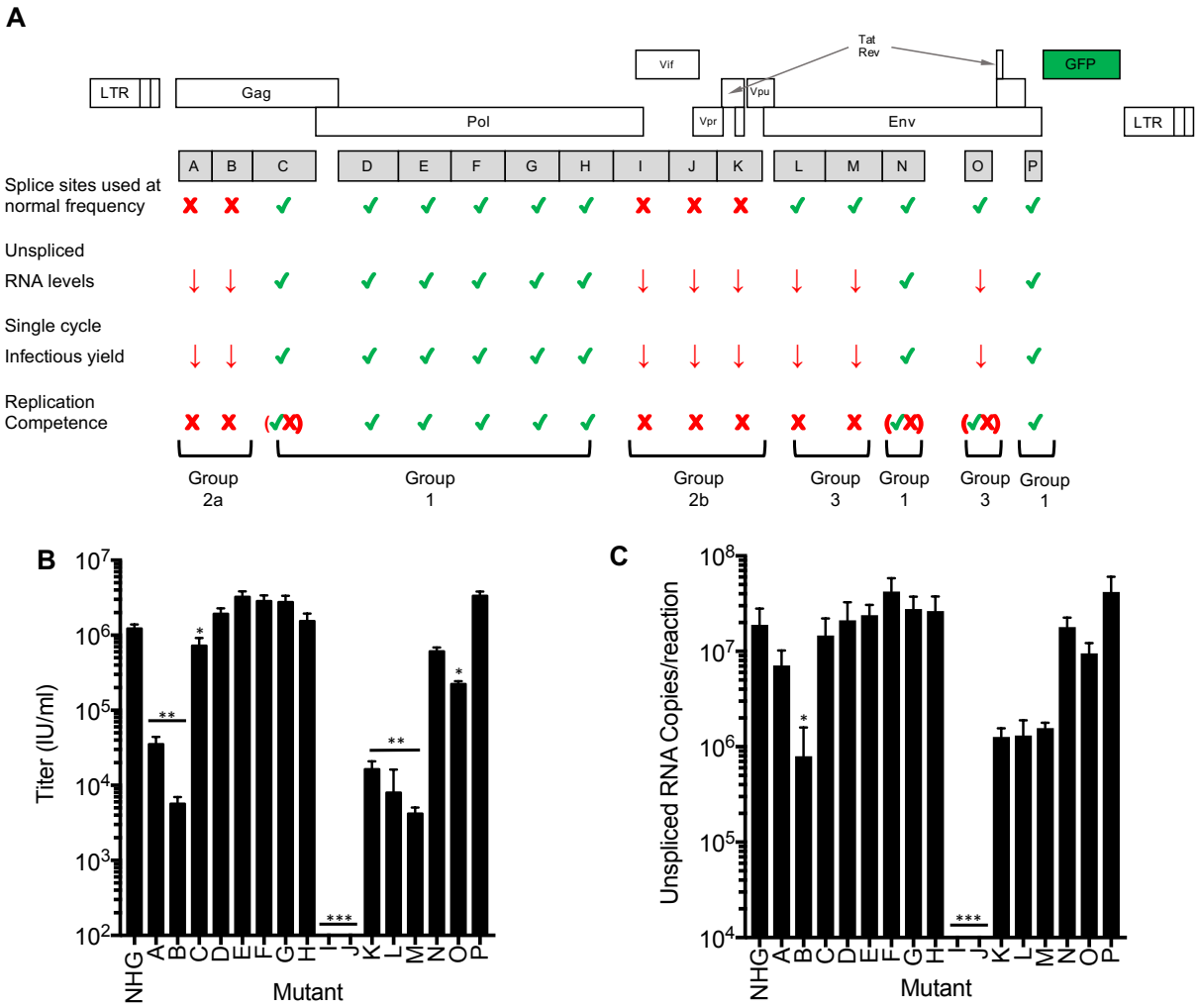
This overuse of canonical splice acceptors in I, J and K resulted in aberrant representation of particular viral mRNAs. Among the 1.8 kb class of mRNAs, for WT HIV-1, Nef2 was the dominant mRNA species in both splicing assays (Figure 3.4E, F). Conversely, in mutant I, Nef5 was the dominant mRNA, while in mutant J, Nef4, Tat3 and Vpr1 were overrepresented (Figure 3.4E, F). These changes were likely responsible for the overexpression of GFP (in the *nef* position) and/or Vpr in these mutants (Figure 3.1C). For mutant K, Tat1 mRNA was over-represented (Figure 3.4E, F), but the overall levels of protein expression were not greatly affected in transfected cells.

Summary

For other replication defective mutants (L, M, and O) that we termed Group 3, the relative uses of splice sites appeared normal, despite obvious replication defects (Figure 3.2, Figure 3.4B-E). These viruses appeared to express a normal complement of viral proteins in transfected cells (Figure 3.1C). Overall, therefore, analyses of viral replication and RNA splicing led to the classification of the synonymously mutated viruses into three groups (Figure 3.5A): Group 1 mutants exhibited near WT fitness, Group 2 mutants exhibited replication defects accompanied by perturbed RNA splicing, while Group 3 mutants had profound replication defects in the absence of obvious splicing perturbation. The three phenotypes were caused by mutations that exhibited a clear regional bias with respect to their distribution along the viral genome (Figure 3.5A). Specifically, Group 1 viruses carried mutations throughout the *pro* (D) and *pol* (E to H) genes or in the 3' portion of the *env* gene (N, P), and replicated indistinguishably from HIV-1_{NHG}. Conversely, Group 2 viruses with obvious splicing defects carried mutations in two distinct genomic regions: Group 2a viruses (A, B) carried mutations toward the 5' end of the genome, within *gag*, while Group 2b viruses (I, J, K) carried mutations or in the central portions of the genome, within the accessory genes (Figure 3.5A). Group 3 viruses (L, M, O) that were defective but exhibited near-normal splicing carried mutations in the *env* gene (Figure 3.5A).

Figure 3.5: Phenotypes of synonymously mutated HIV-1 viruses

(A) Summary of the properties of HIV-1 viruses carrying blocks of nucleotides that were synonymously mutated (A-P). The frequency of splice site utilization was assessed in transfected 293T cells (Figure 3.4), Single-cycle replication assays were used to assess unspliced RNA levels and infectious virus yield (see panels B and C below). Replication competence was determined using spreading replication assays. **(B)** Infectious virion yield measured in the supernatant of MT4 cells, infected with each of the mutant viruses at an MOI of 1.0, and harvested 2 days post infection. Values are the mean ±sd n=3 or n=2 experiments, *p<0.05, **p<0.005 by students t-test calculated with relative values compared to wild-type virus (*** Values were below the limit of quantitation). **(C)** Levels of unspliced HIV-1 genomes in RNA extracted from MT4 cells, infected with each of the mutant viruses at an MOI of 1.0, and harvested 2 days post infection, mean ±sd n=3 or n=2 experiments, *p<0.05 by students t-test compared to wild-type virus. (*** Values were below the limit of quantitation).


A
LTR
Gag
Pol
Vif
Vpr
Vpu
Tat
Rev
Env
GFP
LTR
A B C D E F G H I J K L M N O P
Splice sites used at normal frequency
Unspliced RNA levels
Single cycle Infectious yield
Replication Competence
Group 2a
Group 1
Group 2b
Group 3
Group 1
Group 3
Group 1
B
Titer (IU/ml)
NHG A B C D E F G H I J K L M N O P
Mutant
C
Unspliced RNA Copies/reaction
NHG A B C D E F G H I J K L M N O P
Mutant

Chapter 4. Novel cis-Acting Splicing Elements in HIV-1

To determine the sequence elements responsible for the perturbations in splicing in Group 2a and Group 2b viruses we took two approaches. First, we attempted to derive second-site revertant viruses through passage of each viral mutant in MT4 cells. Second, we applied a mapping approach, in which each block of mutations in mutants A, B, I, J and K was divided into roughly equally sized component segments, and the splicing properties of each secondary mutant re-examined. Through an iterative process, sometimes combining mapping and second-site revertant derivation, we could determine the nature of the defects in each Group 2 mutant and map the responsible cis-acting sequences.

Group 2a: Activation of cryptic splice sites in gag

For mutants A and B in which cryptic splice sites within the *gag* gene were activated by the silent mutations, we first attempted to derive revertant viruses by passage in MT4 cells. For both mutants A and B, passage quickly yielded viruses that replicated more rapidly than the parental mutants viruses (Figure 4.2A, B, Figure 4.1). In the case of mutant A, passage in MT4 cells yielded a revertant virus that replicated well, albeit with delayed kinetics relative to WT HIV-1_{NHG} and contained two nucleotide substitutions relative to the A mutant parent. One of these mutations (C819T) was responsible for the revertant replication phenotype (Figure 4.2A). The C819T mutation was synonymous, and while it occurred at a position that differs from the WT in mutant A, the reversion was not to the WT sequence (WT=G819, mutant A=C819, revertant=T819). Thus, if position G819

in the WT virus was involved in a hypothetical RNA secondary structure that was perturbed or induced in mutant A (C819), then the C to T substitution in the revertant would not be expected to affect the perturbation of that secondary structure. The C819T revertant largely corrected the predominant splicing defects in mutant A, reducing the use of the cryptic splice acceptor (A_{955}) from ~ 1% to ~0.1% and the cryptic splice donor (D_{1169}) from ~14% to <1% (Figure 4.2C). Since the reversion mutation C819 was distal to the cryptic splice sites (~140 and ~350 nucleotides 5' to A_{955} and D_{1169}, respectively, Figure 4.2D) the mechanism by which it exerts its effect was unclear. Notably, the revertant mutation occurred within a few nucleotides of the reported secondary structure that includes the HIV-1 packaging sequence, and D1 (which is at position 743). Thus, it may be that the revertant mutation acts by modulating the secondary structure surrounding D1, rather than on the cryptic A_{955} and D_{1169} that were activated in mutant A.

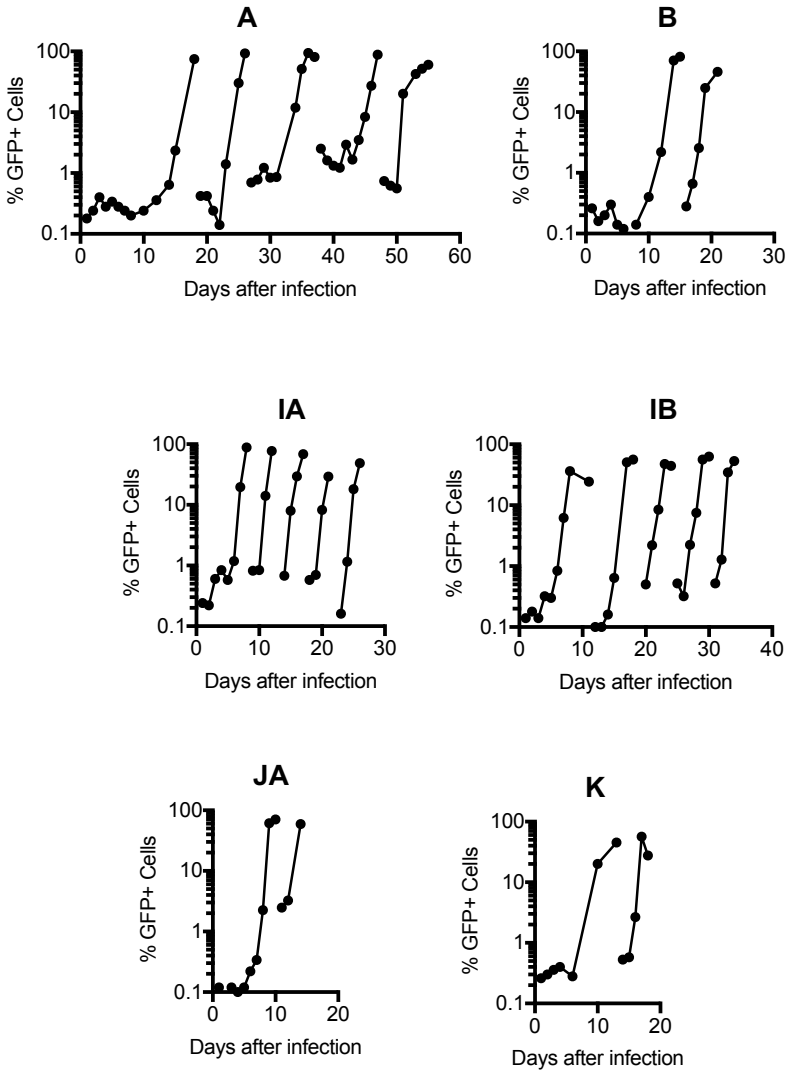
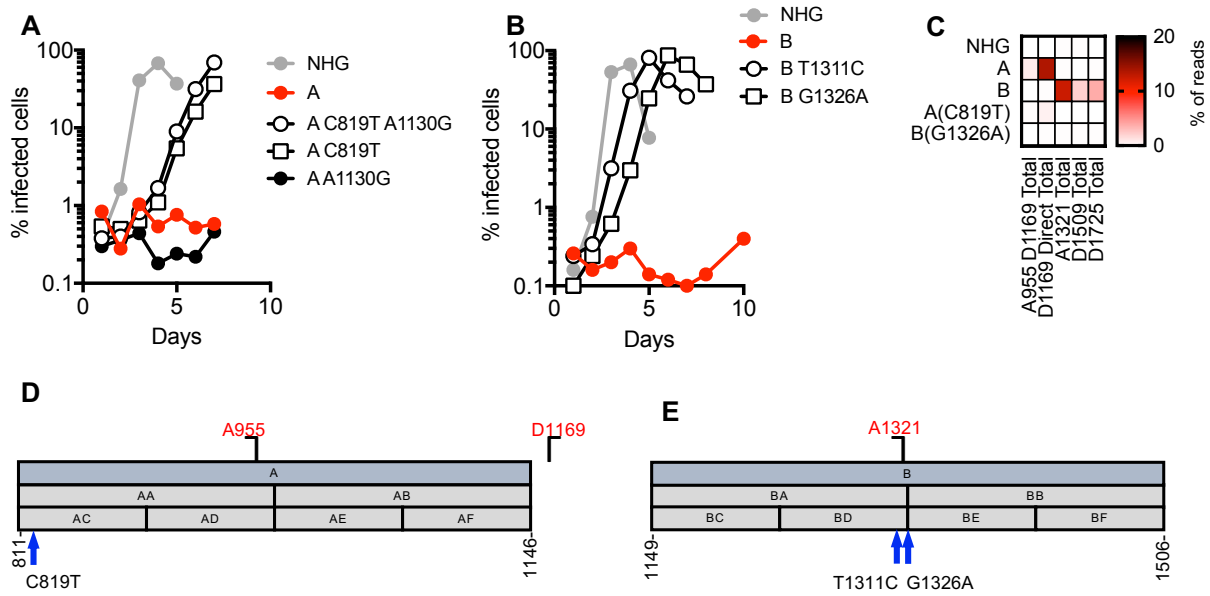


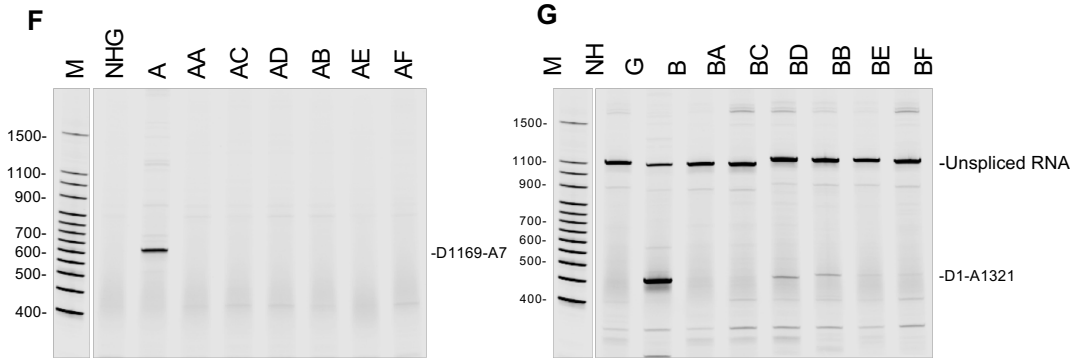
Figure 4.1: Spreading replication experiments to recover second-site revertants of defective mutants

MT4 cells were infected with the mutant viruses (A, B, IA, IB, JA, K, as indicated, harvested from the supernatant of 293T cells transfected with each of the indicated mutant proviral plasmids). Aliquots of infected cells were withdrawn each day, fixed in 4% PFA and the proportion of infected cells determined by FACS analysis of GFP expression.

Figure 4.2: Activation of cryptic splice sites by synonymous mutations in Gag

(A, B) MT4 cells were infected with the indicated virus (harvested from the supernatant of 293T cells transfected proviral plasmids representing each of the WT(HIV-1_{NHG}), mutant (A, B and revertants (A C819T, A1130G and B T1311C, G1326A) thereof) at an MOI of 0.002. Aliquots of infected cells were withdrawn each day and the proportion of infected cells determined by FACS analysis of GFP expression. **(C)** Next gen sequencing analysis of HIV-1 splicing. The heatmap indicates relative proportion of sequencing reads that used the cryptic splice sites for WT(HIV-1_{NHG}), mutant (A, B and revertants thereof) . **(D,E)** Schematic representation of the mutant blocks of nucleotides in HIV-1 mutants A (D) and B (F), indicating positions of mutant derivatives (AA, AB, BA, BB) etc, and the positions of cryptic splice sites and revertant mutant sites. Blocks colored blue are those that conferred overt splicing perturbations when mutated. **(F)** Fluorescent primer PCR analysis of HIV-1 splicing in mutant A. A sense PCR primer situated 5' to the cryptic donor D1169, was used along with an antisense primer positioned 3' to A7 (labelled with IRD800) was used to amplify cDNAs derived from the 1.8 kb class of spliced HIV-1 mRNAs respectively. **(G)** Fluorescent primer PCR analysis of HIV-1 splicing in mutant B. A sense PCR primer situated 5' to D1, was used along with an antisense primer positioned 3' to the mutant B block (labelled with IRD800) was used to amplify cDNAs derived HIV-1 mRNAs. For panels (F) and (G) PCR products were subjected to PAGE and a LI-COR Odyssey scanner was used to detect fluorescent signals directly from the gels.


A
% infected cells
Days
NHG
A
A C819T A1130G
A C819T
A A1130G
B
% infected cells
Days
NHG
B
B T1311C
B G1326A
C
NHG
A
B
A(C819T)
B(G1326A)
A955 D1169 Total
D1169 Direct Total
A1321 Total
D1509 Total
D1725 Total
% of reads
D
A955
D1169
A
AA
AB
AC
AD
AE
AF
811
1146
C819T
E
A1321
B
BA
BB
BC
BD
BE
BF
1149
1506
T1311C G1326A


F
M
NHG
A
AA
AC
AD
AB
AE
AF
1500-
1100-
900-
700-
600-
500-
400-
-D1169-A7
G
M
NH
G
B
BA
BC
BD
BB
BE
BF
1500-
1100-
900-
700-
600-
500-
400-
-Unspliced RNA
-D1-A1321

For mutant B, passage in MT4 or CEM cells yielded two different replicating revertant viruses each of which contained a single nucleotide substitution relative to the B mutant (G1326A in MT4 cells and T1311C in CEM cells) (Figure 4.2B, Figure 4.1) Both revertant mutations were synonymous, at positions that differed in WT and mutant B viruses. Both mutations were proximal to the cryptic splice acceptor (A_{1321}) that was activated by the B mutations (Figure 4.2E).

Analysis of the G1326A revertant in the NGS splicing assay indicated reduced use of the cryptic acceptor A_{1321}, D_{1509}, and D_{1725} (from 12%, 2.9%, and 3.9% to <1% respectively Figure 4.2C). Given the proximity of the reversion mutations to the silenced cryptic splice site, it is likely that these mutations act directly to reduce splicing factor binding, and thereby reduce the use of the cryptic A_{1321}. However, only the T1311C mutation had a marginal effect on the predicted strength of A_{1321}, reducing the MaxEnt score from7.03 to 6.09 (Table 3.2), while G1326A had no effect on the MaxEnt score, yet this mutation abolished the use of A_{1321} as an acceptor. Mutations in mutant B may have created or revealed a splicing factor binding site that was otherwise limiting for the use of the cryptic A_{1321} in a manner that was reversed by the G1326A and T1311C revertant mutations.

To map mutations in A and B that were responsible for activating cryptic splice sites D_{1169} and A_{1321} respectively, we generated a set of mutant viruses (AA, AB, AC, AD, AE and BA, BB, BC, BD, BE) that contained subsets of the synonymous mutations present in mutants A and B (Figure 4.2D, E). We also designed a fluorescent primer-based PCR-PAGE assay in which a PCR primers were positioned to conveniently and specifically monitor the major aberrant splicing event in mutants A and B which, as expected, yielded

PCR product consistent with splicing at the respective cryptic splice sites (D_{1169} and A_{1321}, respectively) (Figure 4.2F, G). Surprisingly, none of the secondary mutants containing subsets of the A and B mutations recapitulated the effects of the A and B mutations (Figure 4.2F, G). Thus, the activation of the cryptic splice sites by the Group 2a mutants A and B was the result of multiple synonymous changes in those mutant viruses. Additionally, the apparent inability of MaxEnt scoring to consistently predict cryptic site utilization in the context of these mutants indicated that splicing defects could not be due solely to direct enhancing effects of mutations on specific cryptic sites.

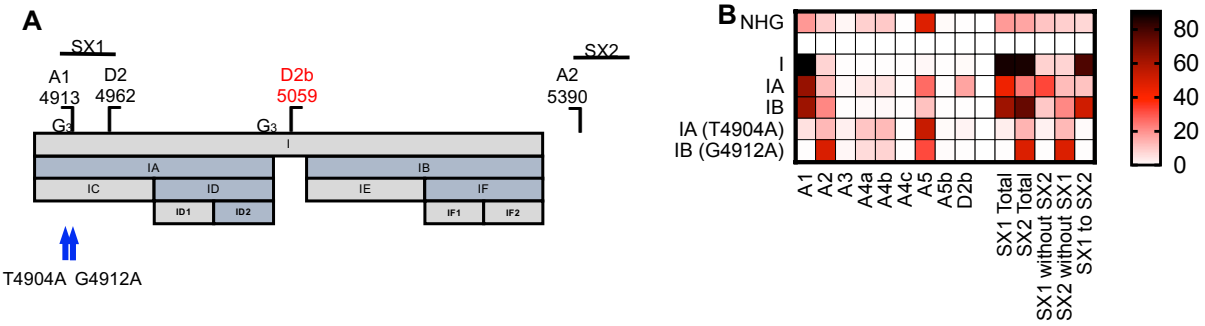
Group 2b: Overuse of canonical splice sites

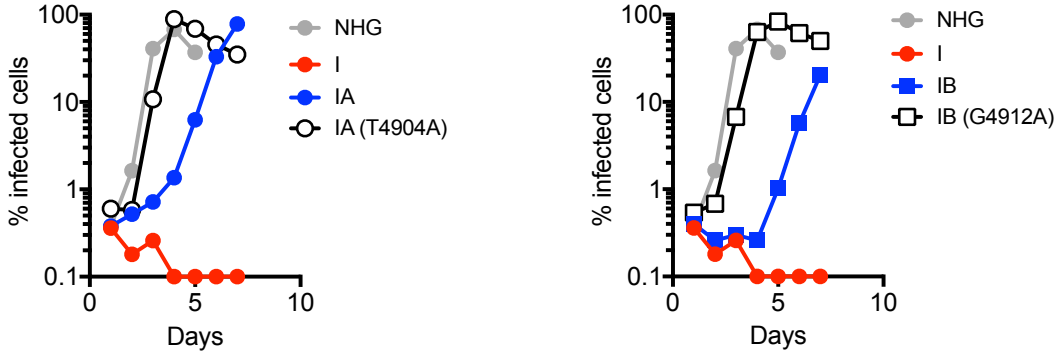
For mutant I, which exhibits overuse of A1 and A2, as well as the corresponding donors (D2 and D3) positioned immediately 3' to A1 and A2, (Figure 4.3A) we failed to obtain revertant replication competent viruses, even after extended passage. Therefore, we divided the I segment into 5' and 3' halves and generated two derivative mutant viruses (IA and IB) each of which had approximately half the of the synonymous mutations that were present in I (Figure 4.3A). Both IA and IB mutants also exhibited splicing defects that were primarily manifested as overuse of A1, but these defects were less complete, in that some degree of direct splicing to downstream acceptors (e.g. A5) was present in both IA and IB (Figure 4.3B). Notably, IA exhibited oversplicing at A1 and (unlike the parent mutant I) the cryptic splice donor D2b. Conversely IB exhibited direct oversplicing at both A1 and, to some degree, to A2 but not D2b (Figure 4.3B).

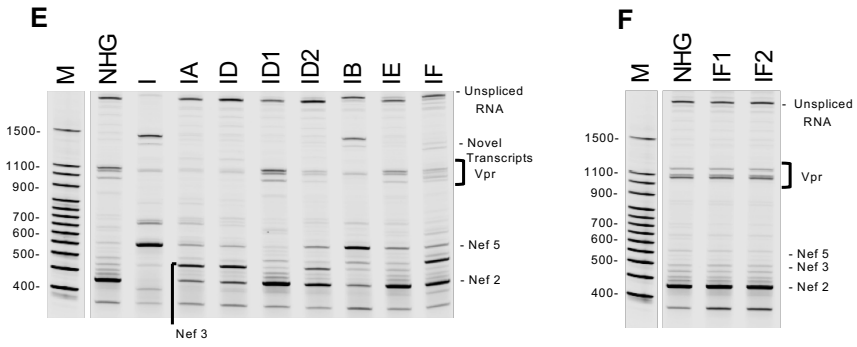
Both IA and IB could replicate with delayed kinetics compared to HIV-1_{NHG}, and extended passage of IA and IB yielded point mutation revertants that could replicate with kinetics close to those of HIV-1_{NHG} (Figure 4.3C, D, Figure 4.1). In the case of IA, a T4904A mutation occurred in the A1 polypyrimidine tract (Figure 4.3A, C). This caused a profound reduction of splicing at A1 and likely as a consequence, reduction of the inclusion of SX1 (A1-D2) in spliced RNAs (Figure 4.3B). Use of the cryptic D2b site was also abolished in the IA(T4904A) revertant. In fact, other than underuse of A1 and reduced inclusion of SX1, the IA(T4904A) revertant had a near normal splicing pattern Thus, activation of canonical and cryptic downstream splice donors in mutant IA appeared to be secondary to activation of A1.

Figure 4.3: Activation of canonical splice acceptor sites (A1 and A2) by synonymous mutations in mutant I

(A) Schematic representation of the mutant blocks of nucleotides in HIV-1 mutant I, indicating positions of mutant derivatives (IA, IB, IC….etc), and the positions of splice sites and revertant mutant sites (blue arrows). Blocks colored blue are those that conferred overt splicing perturbations when mutated. **(B)** Next gen sequencing analysis of HIV-1 splicing in transfected 293T cells. The heatmap indicates relative proportion of sequencing reads that indicate direct splicing to the indicated acceptors or inclusion of the short exons (SX1 and/or SX2) as indicated at the bottom of the heatmap for WT(HIV-1_{NHG}), mutants I, IA, IB and revertants thereof. **(C, D)** MT4 cells were infected with the indicated virus (harvested from the supernatant of 293T cells transfected proviral plasmids representing WT(HIV-1_{NHG}), mutant (I, IA, IB and revertants thereof) at an MOI of 0.002. Aliquots of infected cells were withdrawn each day and the proportion of infected cells determined by FACS analysis of GFP expression. **(E, F)** Fluorescent primer PCR analysis of HIV-1 splicing. 293T cells were transfected with the indicated WT and mutant proviruses, RNA extracted and cDNA synthesized. A sense PCR primers situated 5' to the major splice donor, was used along with an antisense primer positioned either 3' to A7 (labelled with IRD800) to amplify cDNAs derived from the 1.8 kb class of spliced HIV-1 mRNAs. PCR products were subjected to PAGE and a LI-COR Odyssey scanner was used to detect fluorescent signals directly from the gels. Salient mRNA species determined by direct sequencing of extracted gel bands, or inferred from Nextgen sequencing assays are indicated.


A
SX1
A1
4913
D2
4962
D2b
5059
SX2
A2
5390
I
IA
IB
IC
ID
IE
IF
ID1
ID2
IF1
IF2
T4904A G4912A
B
NHG
I
IA
IB
IA (T4904A)
IB (G4912A)
A1
A2
A3
A4a
A4b
A4c
A5
A5b
D2b
SX1 Total
SX2 Total
SX1 without SX2
SX2 without SX1
SX1 to SX2
80
60
40
20
0


C
% infected cells
Days
NHG
I
IA
IA (T4904A)
D
% infected cells
Days
NHG
I
IB
IB (G4912A)


E
M
NHG
I
IA
ID
ID1
ID2
IB
IE
IF
1500-
1100-
900-
700-
600-
500-
400-
- Unspliced RNA
- Novel Transcripts
Vpr
- Nef 5
- Nef 2
Nef 3
F
M
NHG
IF1
IF2
1500-
1100-
900-
700-
600-
500-
400-
- Unspliced RNA
Vpr
- Nef 5
- Nef 3
- Nef 2

For mutant IB, the situation was more complex; a revertant (G4912A) was recovered after passage (Figure 4.1), precisely at the A1 acceptor that abolished the use of A1, and consequently the inclusion of SX1 into spliced mRNA (Figure 4.3A, B, D). However, significant overuse of A2, and consequent inclusion of SX2 (A2-D3) remained evident in the IB(G4912A) revertant (Figure 4.3B). Indeed, overuse of A2 was more prominent in the IB(G4912A) revertant than in IB, perhaps because of the functional removal of A1 (Figure 4.3B). It was therefore apparent that native sequences within IA result in suppression of splicing at A1, while sequences within IB cause suppression of splicing at both A1 and A2.

To map elements within IA and IB that control splicing at A1 and A2, we used the fluorescent PCR-based assay to analyze the pattern of splicing for viruses containing subsets of the IA and IB mutations. For IA, analysis of viruses containing smaller component mutant elements (IC and ID, Figure 4.3A) revealed that aberrant splicing was conferred only by the ID element (Figure 4.3E). Thereafter, when ID was subdivided into ID1 and ID2, it was evident that the controlling element resided primarily within ID2 (Figure 4.3E). Thus, this analysis revealed a 48 nucleotide sequence that appeared to suppress splicing at A1. Notably, a novel ESS/ESE element, termed ESS2b/ESE2b was recently identified that nearly precisely coincides with ID2, indicating that our approach has the potential to identify novel splicing regulatory signals (Brillen et al., 2017).

Notably, the unmutated IA segment contained short runs of three G's (G_3) that have been reported to constitute hnRNP binding sites which suppress the activation of the cryptic splice donor D2b. The mutations in IA disrupted two of these G_3 runs (Figure 4.3A) and

D2b was used in ~17% of IA spliced reads. The revertant mutation IA(4904) was able to inhibit activation of D2b from 17% to 2% (Figure 4.3B), suggesting that the activation of D2b is predominately a secondary effect of A1 overutilization, Nevertheless, one of the G_3 motifs coincides with the ID2 segment and therefore disruption of the G_3 motifs may also contribute to the activation of the cryptic D2b site.

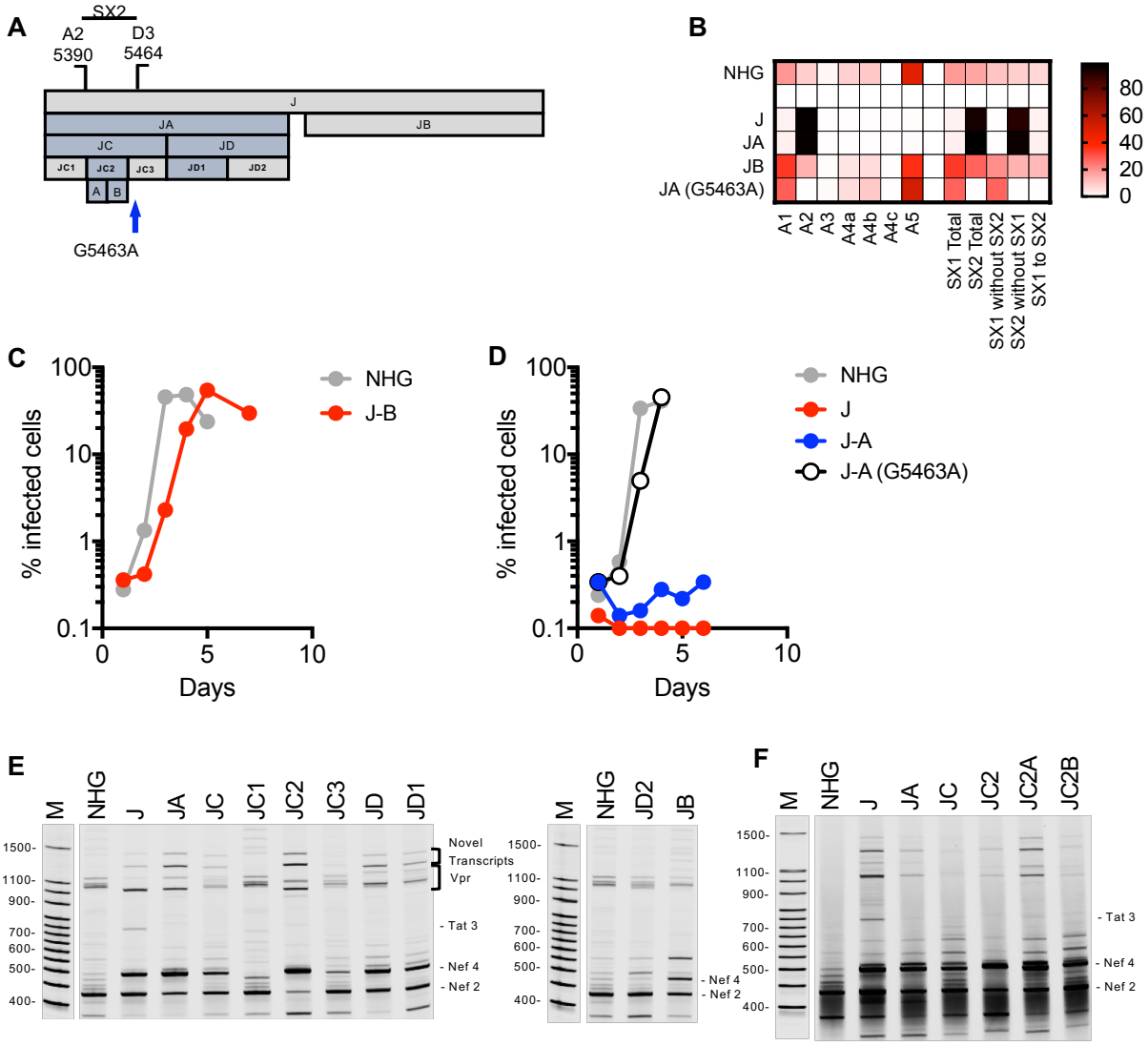
For IB the situation was more complex, as mutations in this segment control splicing at both A1 and A2. Nevertheless, subdivision of the IB mutations into components IE and IF (Figure 4.3A), revealed that the majority the effect of IB mutations were conferred by mutations in IF. However, IF did not exhibit as prominent a degree of perturbation as IB (Figure 4.3E). Even though the splicing of mutant IE appeared normal, mutations in IE made some contribution to the defects present in IB (Figure 4.3E). Subdivision of IF into IF1 and IF2 yielded viruses with a normal pattern of splicing (Figure 4.3F). Thus, it was evident that multiple sequences acting together in IB, distributed over IE, but primarily concentrated in IF1 and IF2, regulate splicing at A1 and A2 and their overall contributions could not be mapped through this approach to a single small candidate regulatory element.

For mutant J, which exhibited overuse of A2 and D3 we also failed to obtain revertant replication competent viruses, even after extended passage. Therefore, we divided the J segment into 5' and 3' halves and generated two derivative mutant viruses (JA and JB, Figure 4.4A). Although there was some degree of splicing perturbation in JB (Figure 4.4B), this perturbation was modest compared to J and JA. Moreover, JB was only marginally delayed in spreading replication assays compared to HIV-1_{NHG}. (Figure 4.4C). Therefore,

we did not attempt to select JB revertants. Conversely, JA recapitulated the splicing perturbation observed in J, and was replication defective (Figure 4.4B, D). Extended passage of mutant JA yielded a revertant mutation (G5463A) that enabled replication (Figure 4.4D, Figure 4.1). Notably, the reversion mutation was precisely at, and inactivated donor D3 (Figure 4.4A), but also abolished splicing to A2 (Figure 4.4B). The enhancing effect of D3 on A2 splicing has previously been demonstrated (Stoltzfus, 2009). Interestingly, even though D3 appeared to be required for splicing at A2 only a fraction of RNAs that are spliced to A2 are also spliced at D3. It was notable that the JB mutants as well as the JA(G5463A) revertant exhibited some oversplicing to A1 (and therefore elevated inclusion of SX1) even though the J mutant sequences were distal (~440 to 890 nucleotides) to A1 (Figure 4.4A, B).

Figure 4.4: Activation of canonical splice acceptor site A2 by synonymous mutations in mutant J

(A) Schematic representation of the mutant blocks of nucleotides in HIV-1 mutant J, indicating positions of mutant derivatives (JA, JB, JC….etc), and the positions of splice sites and revertant mutant sites (blue arrows). Blocks colored blue are those that conferred overt splicing perturbations when mutated. **(B)** Next gen sequencing analysis of HIV-1 splicing in transfected 293T cells. The heatmap indicates relative proportion of sequencing reads that indicate direct splicing to the acceptors or inclusion of the short exons (SX1 and/or SX2) indicated at the bottom of the heatmap for WT(HIV-1_{NHG}), mutants J, JA, JB and revertants thereof. **(C, D)** MT4 cells were infected with the indicated virus (harvested from the supernatant of 293T cells transfected proviral plasmids representing each of the WT(HIV-1_{NHG}), mutant or revertant viruses at an MOI of 0.002. Aliquots of infected cells were withdrawn each day and the proportion of infected cells determined by FACS analysis of GFP expression. **(E, F)** Fluorescent primer PCR analysis of HIV-1 splicing. 293T cells were transfected with the indicated WT (HIV-1_{NHG}) and mutant proviruses, RNA extracted and cDNA synthesized. A sense PCR primer situated 5' to the major splice donor, was used along with an antisense primer positioned either 3' to A7 (labeled with IRD800) to amplify cDNAs derived from the 1.8 kb class of spliced HIV-1 mRNAs. PCR products were subjected to PAGE and a LI-COR Odyssey scanner was used to detect fluorescent signals directly from the gels. Salient mRNA species determined by direct sequencing of extracted gel bands, or inferred from Nextgen sequencing assays are indicated.


A
SX2
A2
5390
D3
5464
J
JA
JB
JC
JD
JC1
JC2
JC3
JD1
JD2
A
B
G5463A
B
NHG
J
JA
JB
JA (G5463A)
A1
A2
A3
A4a
A4b
A4c
A5
SX1 Total
SX2 Total
SX1 without SX2
SX2 without SX1
SX1 to SX2
80
60
40
20
0
C
% infected cells
100
10
1
0.1
0
5
10
Days
NHG
J-B
D
% infected cells
100
10
1
0.1
0
5
10
Days
NHG
J
J-A
J-A (G5463A)
E
M
NHG
J
JA
JC
JC1
JC2
JC3
JD
JD1
1500-
1100-
900-
700-
600-
500-
400-
Novel
Transcripts
Vpr
- Tat 3
- Nef 4
- Nef 2
M
NHG
JD2
JB
1500-
1100-
900-
700-
600-
500-
400-
- Nef 4
- Nef 2
F
M
NHG
J
JA
JC
JC2
JC2A
JC2B
1500-
1100-
900-
700-
600-
500-
400-
- Tat 3
- Nef 4
- Nef 2

To map elements within JA that control splicing, we used the fluorescent PCR-based assay for the 1.8 kb HIV-1 mRNAs to analyze viruses containing subsets of the JA mutations. We divided the JA mutant segment into 5' and 3' halves in two derivative mutant viruses (JC and JD, Figure 4.4A) which both exhibited some degree of perturbed splicing (Figure 4.4E). Further subdivision of JC into JC1, JC2 and JC3 clearly suggested that the 20 nucleotide JC2 segment contained an element whose mutation was primarily responsible for the perturbed splicing in JC (Figure 4.4E), but further division of 20 nucleotide JC2 yielded two mutant segments (JC2A and JC2B) both of which cause perturbed splicing to nearly the same degree as the J, JA, JC and JC2 mutant segments from which they were derived (Figure 4.4F). Division of the JD segment into JD1 and JD2 clearly revealed another element within the 46 nucleotide JD1 segment, that when mutated yielded a oversplicing pattern similar to that of the J mutant virus (Figure 4.4E). Thus, multiple mutations within the JA fragment, contained within the segments JC2 and JD1 were capable of causing oversplicing defects similar to those observed in the J mutant.

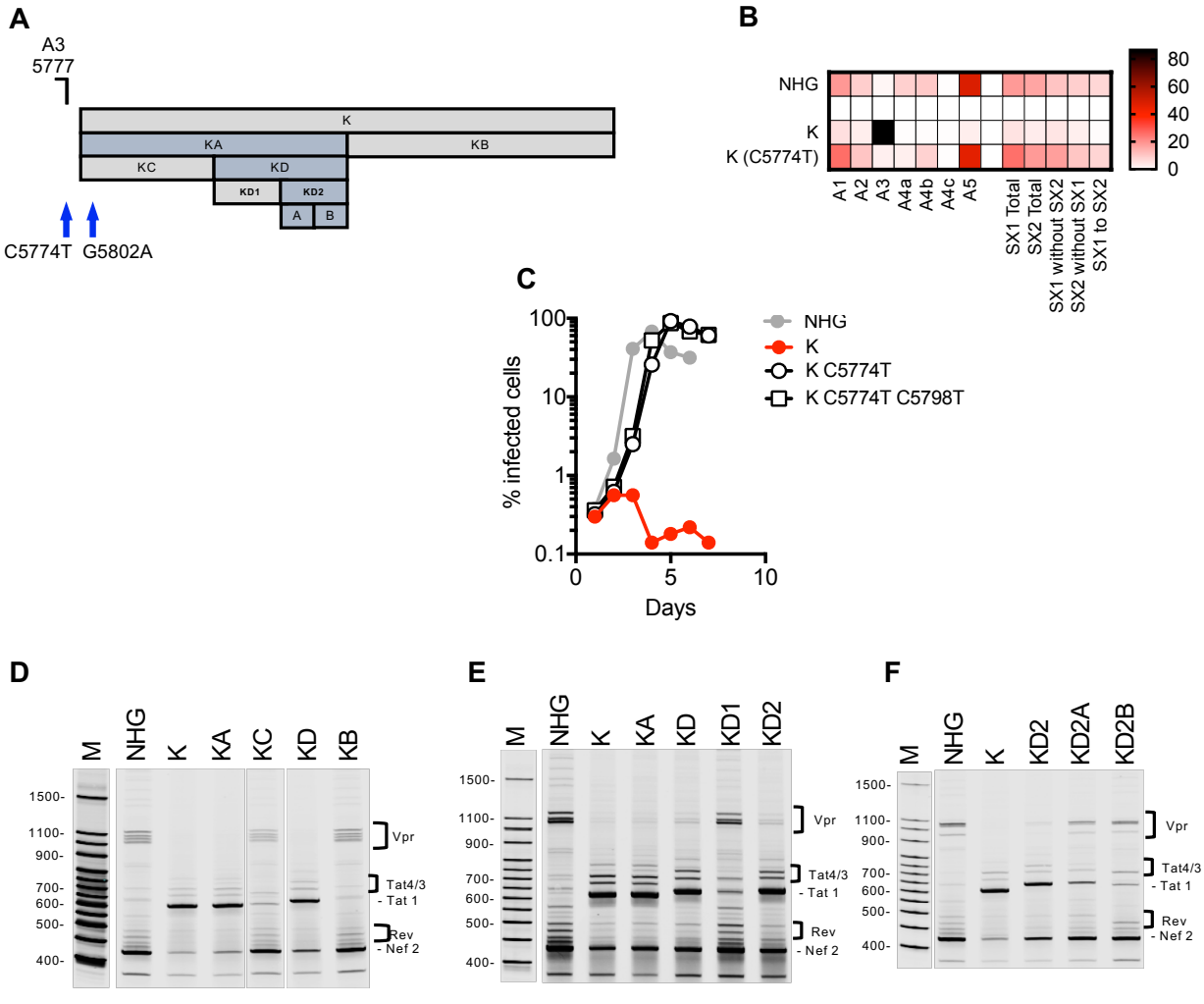
Group 2b: A discrete RNA sequence regulates HIV-1 splicing at A3

For mutant K (Figure 4.5A), which exhibits overuse of A3 (Figure 4.5B), it proved straightforward to recover a revertant mutant virus through passage that corrected the splicing defect and replicated well (Figure 4.5B, C, Figure 4.1). This revertant contained two mutations, one of which (C5774T) was sufficient to restore replication to near WT kinetics (Figure 4.5C). This functional reversion mutation was 3 nucleotides from A3.

Notably the K(C5774T) revertant not only corrected overuse of A3 but exhibited a splicing pattern that was nearly indistinguishable from that of HIV-1_{NHG} (Figure 4.5B). This was surprising because previous studies have reported that CAG and TAG are used at a similar efficiency as 3' splice acceptors in the context of HeLa cell nuclear extracts (Smith 1993). The remarkable diminution of splicing in the K(C5774T) reversion mutant suggests that in the context of A3, the TAG is far less well utilized than CAG. Further, the A3 MaxEnt score of is increased from 9.76 to 10.05 in the context of the K(C5774T) reversion mutant, predicting that A3 is a stronger acceptor in K(C5774T). However, experimentally the reverse is the case, demonstrating the limitation of both *in silico* and *in vitro* analyses to predict splicing phenotypes in the context of a full-length HIV-1 construct in a living cell.

Figure 4.5: Activation of canonical splice acceptor site A3 by synonymous mutations in mutant K

(A) Schematic representation of the mutant blocks of nucleotides in HIV-1 mutant K, indicating positions of mutant derivatives (KA, KB, KC….etc), and the positions of splice sites and revertant mutant sites (blue arrows). Blocks colored blue are those that conferred overt splicing perturbations when mutated. **(B)** Next gen sequencing analysis of HIV-1 splicing in transfected 293T cells. The heatmap indicates relative proportion of sequencing reads that indicate direct splicing to the acceptors or inclusion of the short exons (SX1 and/or SX2) indicated at the bottom of the heatmap for WT(HIV-1_{NHG}), mutant K and the C5774T revertant. **(C)** MT4 cells were infected with the indicated virus (harvested from the supernatant of 293T cells transfected proviral plasmids representing each of the WT(HIV-1_{NHG}), mutant or revertant viruses at an MOI of 0.002. Aliquots of infected cells were withdrawn each day and the proportion of infected cells determined by FACS analysis of GFP expression. **(D-F)** Fluorescent primer PCR analysis of HIV-1 splicing. 293T cells were transfected with the indicated WT (HIV-1_{NHG}) and mutant proviruses, RNA extracted and cDNA synthesized. A sense PCR primer situated 5' to the major splice donor, was used along with an antisense primer positioned either 3' to A7 (labeled with IRD800) to amplify cDNAs derived from the 1.8 kb class of spliced HIV-1 mRNAs. PCR products were subjected to PAGE and a LI-COR Odyssey scanner was used to detect fluorescent signals directly from the gels. Salient mRNA species determined by direct sequencing of extracted gel bands, or inferred from Nextgen sequencing assays are indicated.


A
A3
5777
K
KA
KB
KC
KD
KD1
KD2
A
B
C5774T
G5802A
B
NHG
K
K (C5774T)
A1
A2
A3
A4a
A4b
A4c
A5
SX1 Total
SX2 Total
SX1 without SX2
SX2 without SX1
SX1 to SX2
80
60
40
20
0
C
% infected cells
100
10
1
0.1
0
5
10
Days
NHG
K
K C5774T
K C5774T C5798T
D
M
NHG
K
KA
KC
KD
KB
1500-
1100-
900-
700-
600-
500-
400-
Vpr
Tat4/3
- Tat 1
Rev
- Nef 2
E
M
NHG
K
KA
KD
KD1
KD2
1500-
1100-
900-
700-
600-
500-
400-
Vpr
Tat4/3
- Tat 1
Rev
- Nef 2
F
M
NHG
K
KD2
KD2A
KD2B
1500-
1100-
900-
700-
600-
500-
400-
Vpr
Tat4/3
- Tat 1
Rev
- Nef 2

To map sequences within the K mutant that were responsible for causing oversplicing, we divided the K mutant segment into two halves (KA and KB, Figure 4.5A) and analyzed the pattern of 1.8 kb mRNAs using the fluorescent primer PCR assay. This analysis revealed that mutations responsible for A3 overuse resided in KA (Figure 4.5D). Then, further subdivision of KA (into KC and KD) revealed that KD contained the controlling element(s) (Figure 4.5D). Finally, subdivision of KD (into KD1 and KD2) showed that KD2 contained RNA sequences whose mutation caused oversplicing (Figure 4.5E), but further subdivision of KD2 (into KD2A and KD2B) showed that mutations in both of these KD2 components contributed its effect (Figure 4.5F). Thus, a 23-nucleotide element (KD2) positioned >100nt from A3 contained an RNA element whose sequence influences splicing at A3.

Summary

Through synonymous mutations in the HIV-1 genome , these studies demonstrated the importance and complexity of the non-coding sequences in relation to alternative splicing. We have described a small and discrete element that controls splicing of a single splice site, but more frequently, we found large RNA elements that cannot map to a small sequence and have effects on multiple splice sites. Additionally, we have shown the flexibility of the HIV-1 genome. Our synonymous mutations profoundly attenuated the replication of the viruses, yet acquisition of a single second site revertant mutation was able to completely restore replication to the virus, overcoming the deleterious effects of hundreds of mutations.

Chapter 5. ZAP mediated restriction of CG-enriched HIV

While the precise mechanism driving splicing defects in the previous chapter remain unclear, division of the mutations into smaller segments allowed us to follow which sets of mutations were responsible for inhibiting the replication of the virus. Largely using this strategy, we catalogued regions of the HIV-1 genome that are responsible for control over different splice sites. Pairing this approach with the PCR splicing assay allowed for a fast and reliable read out for mapping perturbations to splicing. Initially, we applied this same approach to the group 3 mutants, but this yielded limited success due to the absence of a clear phenotype beyond fitness in a spreading replication assay. Conversely, initial studies with group 1 mutants yielded no detectable defect in splicing, single cycle infection mRNA or protein expression, or spreading replication assays, and therefore further characterization did not seem relevant.

Inhibitory effects of CG dinucleotides on HIV replication

Group 3 mutants yielded near normal infectious titers when proviral plasmids were transfected in 293T cells and lacked an obvious splicing defect. Unexpectedly the virions collected were only capable of infecting cells in a single cycle and completely defective in spreading replication assays (Figure 5.2A; Figure 5.1B, C). For two defective group 3 mutant viruses, termed L and M, mapping experiments employing derivatives containing smaller mutant segments, termed LA-LF and MA-MD (Figure 5.1A, B, C and D) revealed that their replication defects were not caused by perturbation of a single discrete element. Indeed, mutants LC, LD, LE, LF, MA, MC, and MD, which collectively represented all

mutations in the defective mutants L and M, each replicated with near HIV-1_{WT} kinetics (Figure 5.1 C and D). Moreover, when the mutations in four replication-competent *pol* mutants (E though H, Figure 5.1 A) were combined, the resulting mutant virus (EH) was defective (Figure 5.1 E). Thus, HIV-1 replication defects could be induced by the cumulative effects of synonymous mutations in *pol* or *env*.

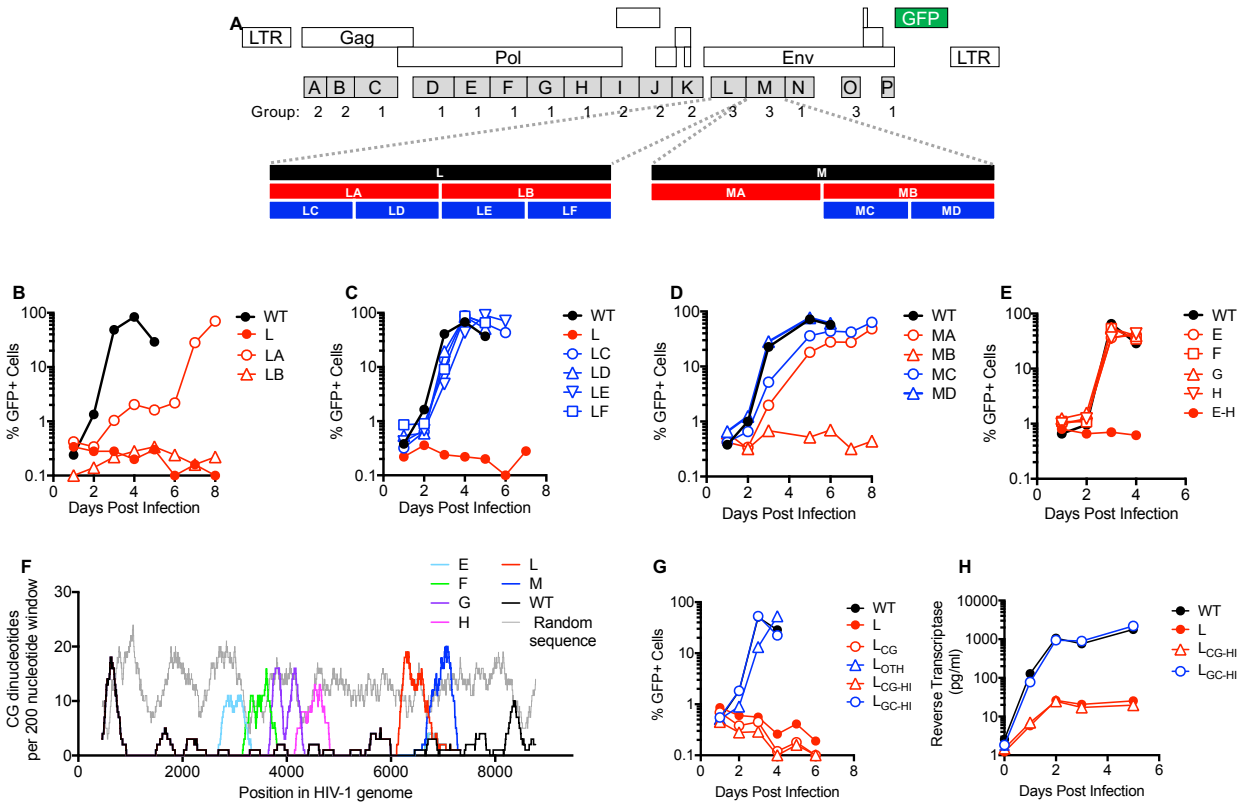


Figure 5.1: Synonymous mutagenesis reveals inhibitory effects of CG dinucleotides on HIV-1 replication

(A) Schematic representation HIV-1$_{NHG}$ GFP provirus, showing locations of synonymous mutant blocks, and corresponding phenotypes. **(B-E)**, Spreading replication of HIV-1 mutants in MT4 cells, as measured by FACS enumeration of infected cells over time. **(F)**, Number of CG dinucleotides present in a 200 nucleotide sliding window in the indicated viral and random sequences. **(G)**, Spreading replication of HIV-1 mutants in MT4 cells, are measured by FACS enumeration of infected cells over time. **(H)**, Spreading replication of HIV-1 mutants in primary lymphocytes, as measured by reverse transcriptase activity in the supernatant of infected cells over time.

Analysis of the HIV-1 genome reveals that it is remarkably sparse in C mononucleotides and particularly deficient in CG dinucleotides, like many vertebrate viruses (Figure 5.1F) (Cheng et al., 2013). Our synonymous mutagenesis coincidentally increased the CG dinucleotide content in mutant segment of viruses L an M, to a level similar to that of random sequence, and to a lesser degree in the individual mutants E through H (Figure 5.1F). We generated derivatives of mutant L, termed L_{CG} and L_{OTH}, respectively, containing only mutations that generated new CG dinucleotides (39/145 original mutations) or the 117 other mutations. We also generated mutants that maximized the CG or GC dinucleotide content in the same segment ($L_{CG\text{-}HI}$ and $L_{GC\text{-}HI}$) (Table 5.1). Strikingly, L_{CG} and $L_{CG\text{-}HI}$ were replication defective in MT4 cells, while L_{OTH} and $L_{GC\text{-}HI}$ replicated with near HIV-1_{WT} kinetics (Figure 5.1G). Mutants L and $L_{CG\text{-}HI}$ also replicated at ~100-fold lower levels than HIV-1_{WT} and $L_{GC\text{-}HI}$ in primary lymphocytes (Figure 5.1H, Figure 5.2B, C).This removal a subset of mutations, particularly those that added a CG dinucleotide, decisively links the replication defects observed with the increase CG content of the HIV-1 genome, with 37 synonymous mutation capable of completely inhibiting viral replication in MT4 cells.

Table 5.1: Mutations in the HIV-1 L mutant and its derivatives

Virus	Mutations	CG dinucleotides added	CG dinucleotides total	GC dinucleotides total
WT	none	0	2	18
L	145 synonymous mutations	37	39	25
L_{CG}	37 mutations (subset of L mutations that generate new CG dinucleotides)	37	39	27
L_{OTH}	108 mutations (subset of L mutations that do not generate new CG dinucleotides)	0	2	19
$L_{CG\text{-}HI}$	41 mutations	41	43	18
$L_{GC\text{-}HI}$	9 mutations	0	2	27

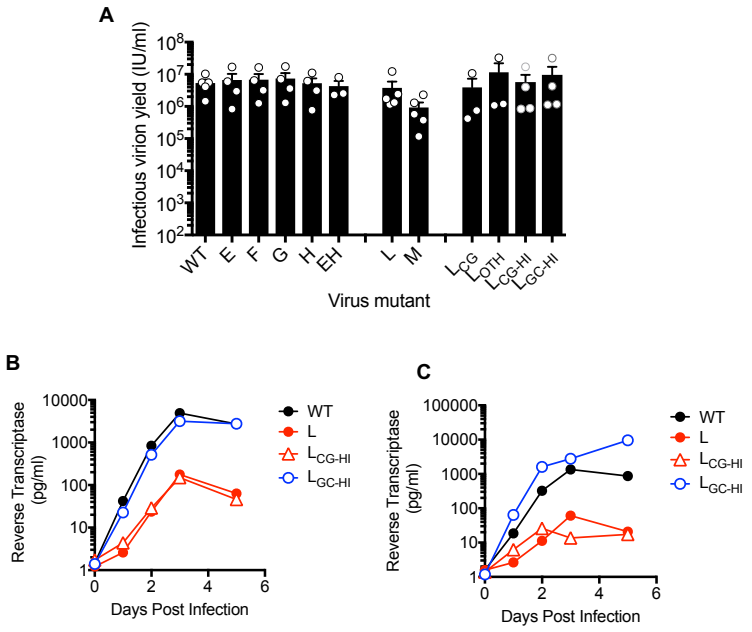


Figure 5.2: CG-enriched HIV-1 clones yield near WT levels of virus from transfected 293T cells but are attenuated in replication in primary lymphocytes

(A) Yield of infectious virus from proviral plasmid transfected 293T cells, as measured by infection of MT4 cells (mean ± sd, n=3 to 5 independent experiments). **(B, C)** Spreading replication of HIV-1 mutants in primary lymphocytes from two additional donors as measured by reverse transcriptase activity in the supernatant of infected cells over time.

CG dinucleotides cause depletion of cytoplasmic RNA

To understand the basis for replication defects in the CG-enriched HIV-1 mutants, we infected MT4 cells with equal titers of each virus in single-cycle replication experiments. Notably L, L_{CG} and $L_{CG\text{-}HI}$ infected cells generated ~1000-fold fewer infectious progeny virions than L_{OTH} and $L_{GC\text{-}HI}$ infected cells (Figure 5.3A). Infectious virion yields from EH infected cells were similarly reduced (Figure 5.4A). Western blot analyses revealed abnormally low levels of Gag and Env proteins in cells infected with L, L_{CG} and $L_{CG\text{-}HI}$, but HIV-1_{WT} levels for L_{OTH} and $L_{GC\text{-}HI}$ (Figure 5.3B). Expression of the *gfp* reporter that was embedded in the *nef* gene and therefore expressed via an mRNA from which the L segment is removed by splicing (Fig 5.3C) was equivalent for each virus, as measured by western blotting or flow cytometry (Fig 5.3B, Figure 5.4B). A deficit in Gag levels also occurred in EH infected cells. However, normal levels of both Env and GFP proteins, whose spliced mRNAs lack the CG-enriched segment, were generated in EH infected cells (Figure 5.4C), suggesting that the inclusion of CG dinucleotides into RNA only affects the corresponding protein expression when they are included as exons.

Unspliced viral RNA levels, measured by RT-PCR, in single-cycle infected MT4 cells were 5 to 10-fold lower in L, L_{CG}, $L_{CG\text{-}HI}$ and EH infected cells and at HIV-1_{WT} levels for L_{OTH}, $L_{GC\text{-}HI}$, E, F, G, or H infected cells (Figure 5.3D, Figure 5.4D). Single molecule fluorescence in situ hybridization (smFISH) experiments using a *gag* probe revealed that the deficit in unspliced viral RNA occurred specifically in the cytoplasm in $L_{CG\text{-}HI}$ infected cells, while nuclear levels were normal (Figure 5.3E, F, Figure 5.5). Similar smFISH experiments employing a probe that detected all spliced and unspliced viral RNAs (Figure

5.3C) revealed a marginal, statistically ambiguous deficit for $L_{CG\text{-}HI}$ (Figure 5.4E, Figure 5.6), consistent with the notion that unspliced RNA (that represent only a subset of total HIV-1 RNAs) were selectively depleted in in $L_{CG\text{-}HI}$ infected cells.

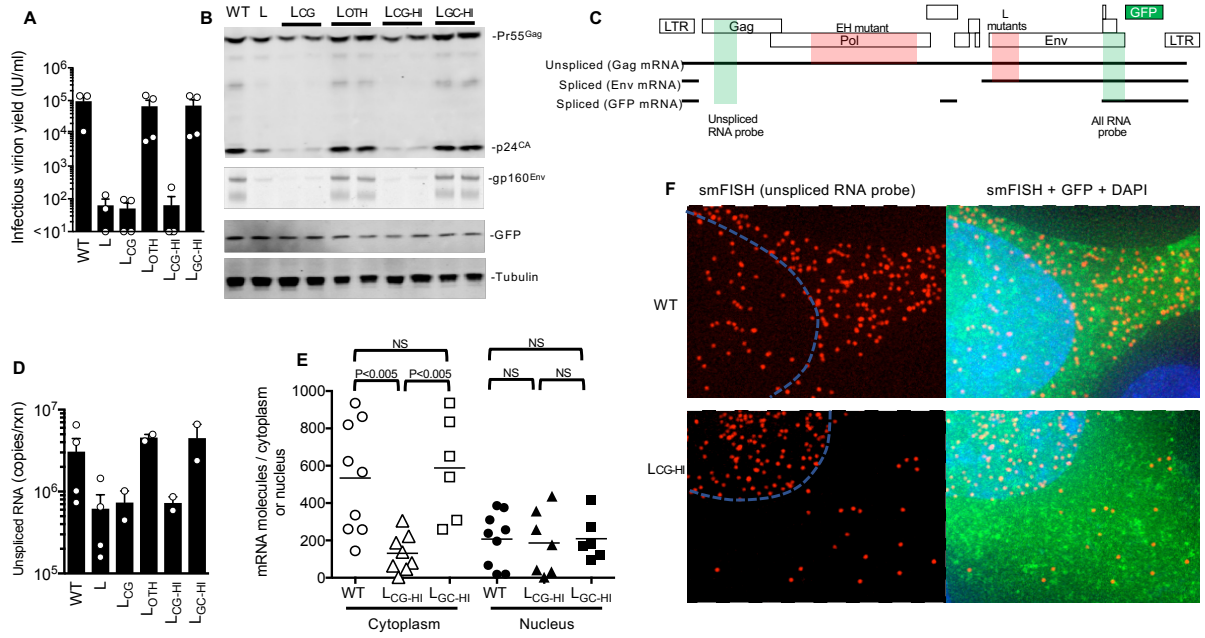
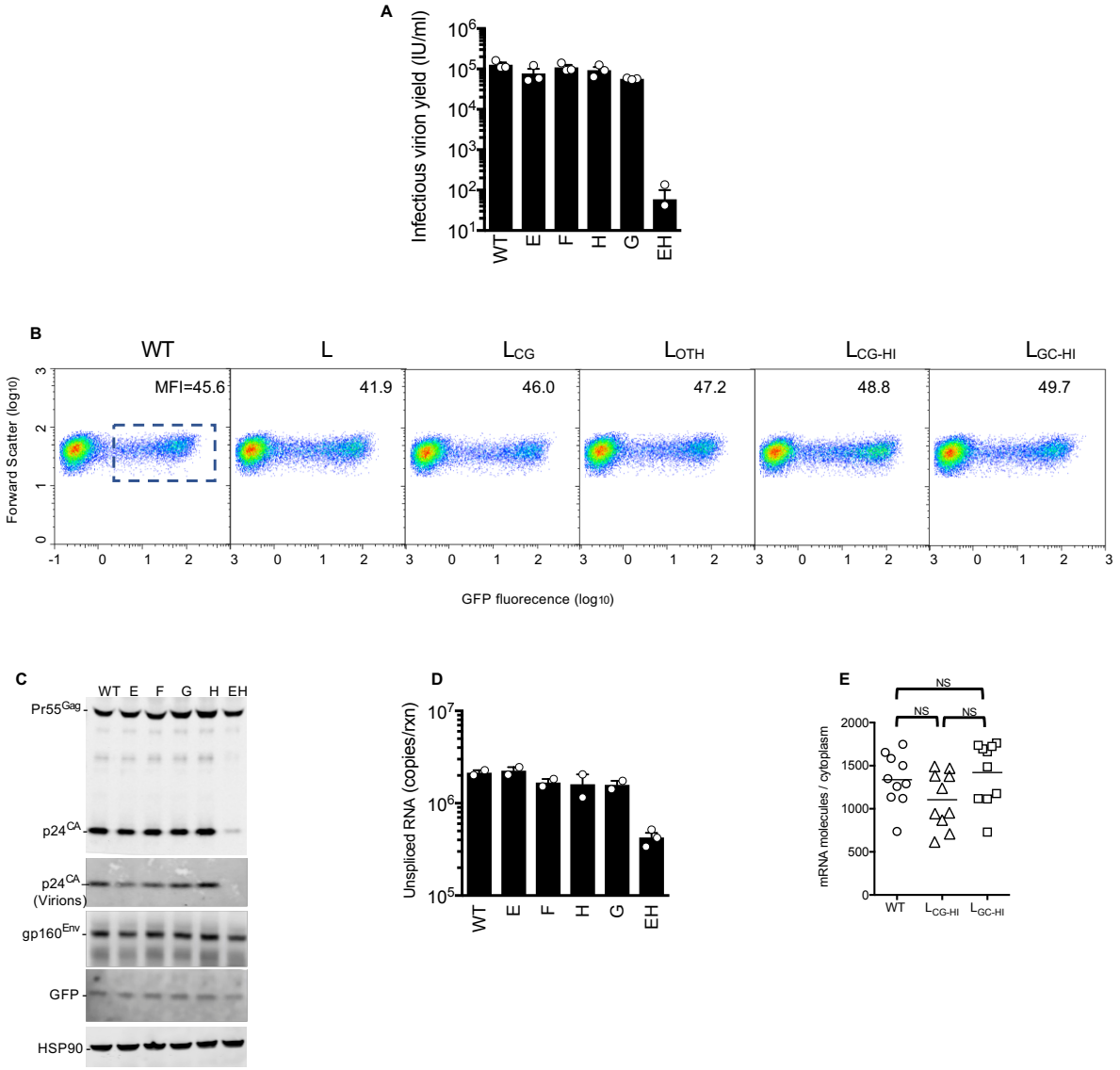


Figure 5.3: CG dinucleotides cause depletion of cytoplasmic RNA

(A) Yield of infectious virus, in a single-cycle of replication following infection of MT4 cells with equal titers of HIV-1$_{WT}$ and mutants (mean ± sd n=3 independent experiments). **(B)** Western blot analysis (anti-Gag, anti-Env anti-GFP and anti-Tubulin) of viral, reporter and cellular protein expression 48h after a single-cycle infection of MT4 cells with WT and synonymous *env* mutant HIV-1. **(C)** Location of exons for salient mRNAs (black lines), mutated segments (red shading) and smFISH probes (green shading) in HIV-1 sequence. **(D)** Q-RT-PCR quantification of unspliced RNA in MT4 cells in a single-cycle infection assay (mean ± sd n= 2-4 independent experiments). **(E)** Quantification of RNA molecules (fluorescent spots) by smFISH in cytoplasm (open symbols) and nucleus (filled symbols) using a probe for unspliced RNA in *gag* after infection of HOS/CD4-CXCR4 cells. Each symbol represents nucleus or cytoplasm of an individual cell. Horizontal lines represent mean values. P-values were determined using Mann-Whitney test. **(F)** Examples of smFISH analysis of an HIV-1$_{WT}$ and mutant infected cell (red=smFISH *gag* probe, green=GFP, blue=Hoescht dye) blue line indicates nucleus/cytoplasm boundary.

Figure 5.4: Effects of CG dinucleotides on the HIV-1 infectious virion yield, RNA and protein levels in a single-cycle replication assays

(A) Yield of infectious virus in a single cycle of replication following infection of MT4 cells with equal titers of HIV-1WT and *pol* mutants (mean ± s.e.m., *n* = 3 independent experiments). **(B)** Expression of *gfp* in MT4 cells, as measured by flow cytometry, 48 h after infection with equal titers of the indicated viruses. Numerical values are mean fluorescent intensity (MFI) of infected cells (indicated by the dotted box). **(C)** Western blot analysis (anti-Gag, anti-Env, anti-GFP and anti-HSP90) of viral, reporter and cellular protein expression, 48 h after a single cycle of infection of MT4 cells with wild-type and synonymous *pol* mutant HIV-1. Representative of three experiments. **(D)** RT–qPCR quantification of unspliced RNA in MT4 cells in a single- cycle infection assay with wild-type and synonymous *pol* mutant HIV-1 (mean ± s.e.m., *n* = 2 or 3 independent experiments). **(E)** Quantification of RNA molecules (fluorescent spots) by smFISH in cytoplasm using a probe targeting all spliced and unspliced HIV-1 RNA species after infection of HOS/CXCR4-CD4 cells. Each symbol represents an individual cell. Horizontal lines represent mean values; *P* values were determined using Mann–Whitney test (*n* = 10).


A
Infectious virion yield (IU/ml)
WT
E
F
H
G
EH
B
WT
L
LCG
LOTH
LCG-HI
LGC-HI
MFI=45.6
41.9
46.0
47.2
48.8
49.7
Forward Scatter (log10)
GFP fluorecence (log10)
C
WT E F G H EH
Pr55Gag
p24CA
p24CA (Virions)
gp160Env
GFP
HSP90
D
Unspliced RNA (copies/rxn)
E
NS
mRNA molecules / cytoplasm
WT
LCG-HI
LGC-HI

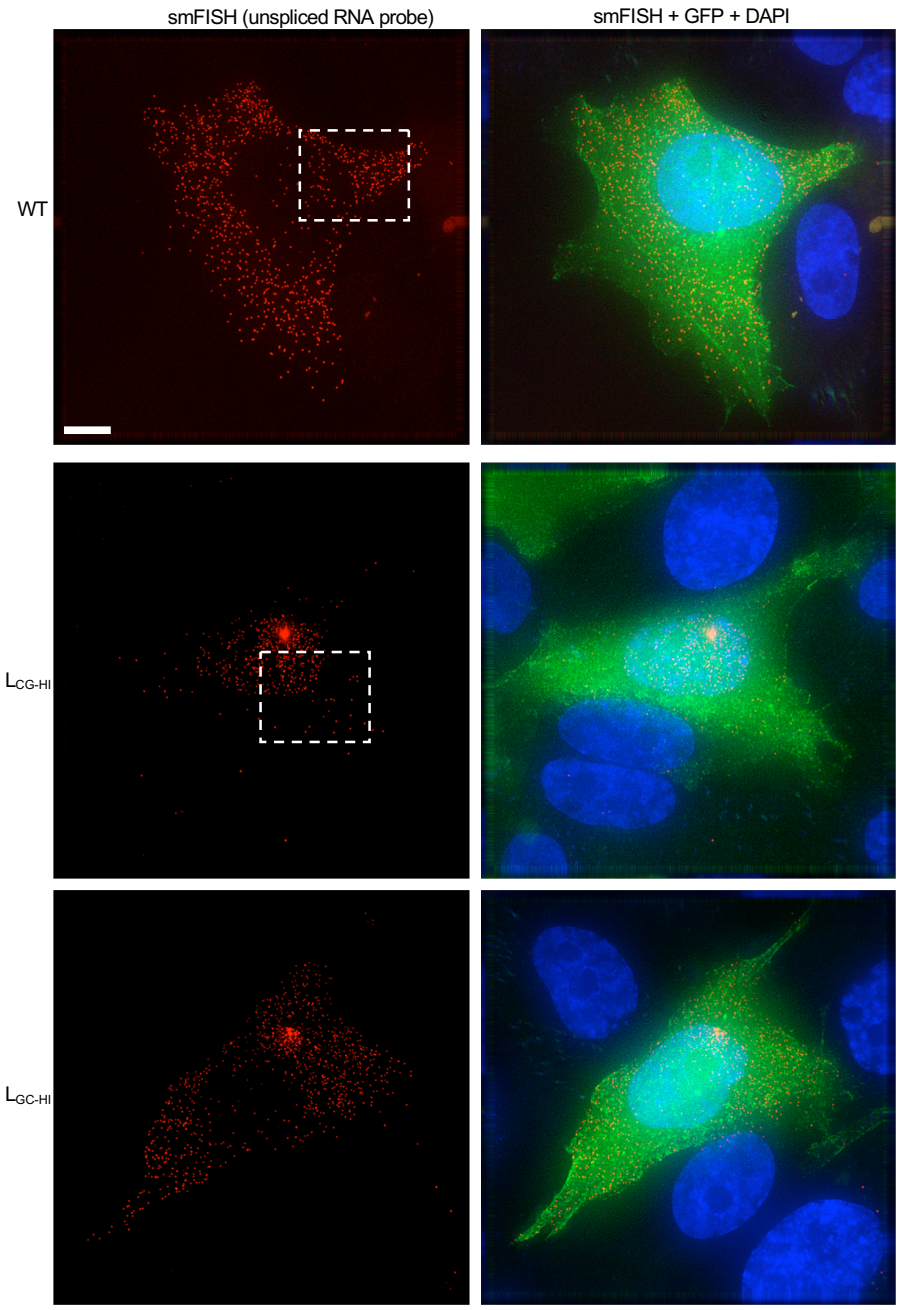


Figure 5.5: smFISH quantification of unspliced HIV-1 RNA in infected cells

Examples of smFISH analysis of wild-type and synonymous mutant HIV-1-infected cells (red, smFISH *gag* probe (see Figure 5.3C); green, GFP; blue, Hoescht dye). The boxed areas indicate regions selected for expanded views in Figure 5.3F Clusters of RNA molecules in the nuclei of some infected cells may represent sites of proviral integration. Representative of three independent experiments. Scale bar, 5 μm.

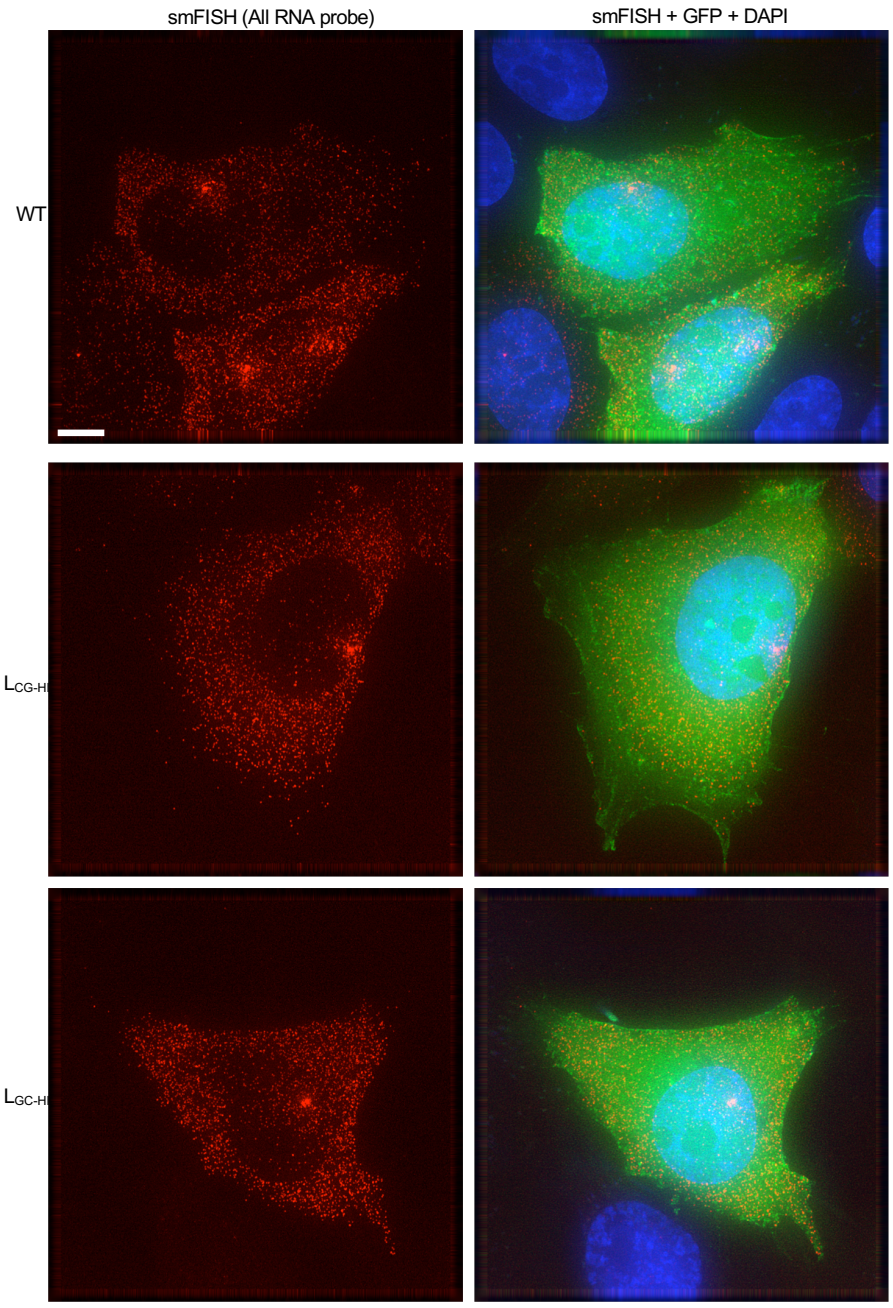


Figure 5.6: smFISH quantification of total HIV-1 RNA in infected cells

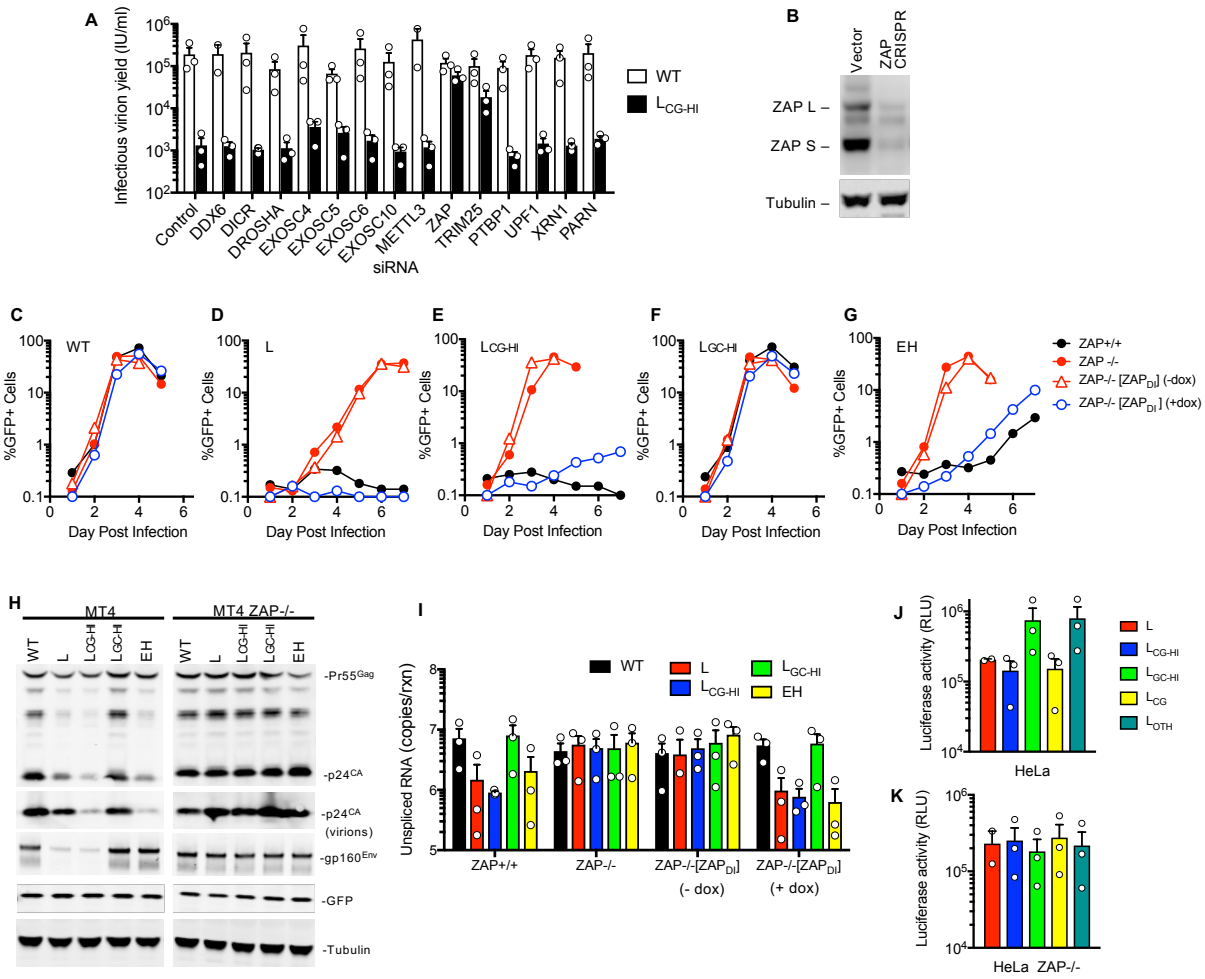
Examples of smFISH analysis of wild-type and synonymous mutant HIV-1-infected cells (red, smFISH probe targeting all viral mRNA species (see Figure 5.3C); green, GFP; blue, Hoescht dye). Clusters of RNA molecules in the nuclei of some infected cells may represent sites of proviral integration. Representative of three independent experiments. Scale bar, 5 µm.

ZAP specifically inhibits CG-enriched HIV-1 replication

A deficit in levels of CG-containing RNA and their protein products appeared be the foundational defect in cells infected with the defective viral mutants L, L_{CG} ,$L_{CG\text{-}HI}$ and EH. Therefore, we conducted a focused siRNA screen targeting proteins implicated in cytoplasmic RNA degradation pathways, e.g. microRNA, nonsense mediated decay, and RNA exosome pathways (Figure 5.7A, Figure 5.8A). Single-cycle replication experiments in HeLa cells revealed that knockdown of zinc finger antiviral protein (ZAP) nearly completely restored infectious virion yield from $L_{CG\text{-}HI}$ infected cells (Figure 5.7A). Knockdown of TRIM25, which enhances ZAP activity, also substantially restored viral yield.

Figure 5.7: ZAP specifically inhibits CG-enriched HIV-1 replication

(A) Yield of infectious virus, in a single-cycle of replication following infection of siRNA transfected Hela cells with equal titers of HIV-1_{WT} and $L_{CG\text{-}HI}$ mutant (mean ± sd n=3 independent experiments). **(B)** Western blot analysis of ZAP expression following CRISPR mutation of ZAP exon 1 in MT4 cells. **(C-G)** Spreading replication of HIV-1 mutants in control (ZAP+/+), ZAP-knockout (ZAP-/-) and doxycycline-inducible ZAP (ZAP_{DI}) reconstituted MT4 cells, as measured by FACS enumeration of infected cells over time. **(H)** Western blot analysis (anti-Gag, anti-Env anti-GFP and anti-Tubulin) of viral, reporter and cellular protein expression in cells and virions 48h after a single-cycle infection of ZAP+/+ and ZAP-/- MT4 cells with WT and mutant HIV-1. **(I)** Q-RT-PCR quantification of unspliced RNA in MT4 cells in a single-cycle infection assay (mean ± sd n=3 independent experiments). (**J, K)** Luciferase expression following transfection of Hela **(J)** or HeLa ZAP-/- **(K)** with reporter plasmids incorporating the indicated HIV-1 RNA segments as 3'UTRs (mean ± sd n=3 independent experiments).


A
Infectious virion yield (IU/ml)
WT
LCG-HI
Control
DDX6
DICR
DROSHA
EXOSC4
EXOSC5
EXOSC6
EXOSC10
METTL3
ZAP
TRIM25
PTBP1
UPF1
XRN1
PARN
siRNA
B
Vector
ZAP CRISPR
ZAP L –
ZAP S –
Tubulin –
C
WT
D
L
E
LCG-HI
F
LGC-HI
G
EH
%GFP+ Cells
Day Post Infection
ZAP+/+
ZAP -/-
ZAP-/- [ZAPDI] (-dox)
ZAP-/- [ZAPDI] (+dox)
H
MT4
MT4 ZAP-/-
WT
L
LCG-HI
LGC-HI
EH
-Pr55Gag
-p24CA
-p24CA (virions)
-gp160Env
-GFP
-Tubulin
I
Unspliced RNA (copies/rxn)
WT
L
LCG-HI
LGC-HI
EH
ZAP+/+
ZAP-/-
ZAP-/-[ZAPDI] (- dox)
ZAP-/-[ZAPDI] (+ dox)
J
Luciferase activity (RLU)
L
LCG-HI
LGC-HI
LCG
LOTH
HeLa
K
HeLa ZAP-/-

We generated ZAP$^{-/-}$ MT4 cells lacking both major ZAP isoforms (ZAP-L and ZAP-S, Figure 5.8B) using CRISPR/Cas9-based approaches. Strikingly, this manipulation restored replication competence to all of the defective CG-enriched viruses, but did not affect HIV-1$_{WT}$ or L$_{GC\text{-}HI}$ replication (Figure 5.7C-G). Indeed, L$_{CG\text{-}HI}$ and EH were defective in unmanipulated MT4 cells, but replicated indistinguishably from HIV-1$_{WT}$ in ZAP$^{-/-}$ MT4 cells. In single-cycle replication experiments, the deficits in Gag and Env protein levels observed with CG-enriched viruses were abolished in ZAP$^{-/-}$ cells (Figure 5.7H). Reconstitution of ZAP$^{-/-}$ MT4 cells with a CRISPR-resistant, doxycycline-inducible ZAP-S construct [ZAP$_{DI}$] enabled doxycycline-dependent inhibition of CG-enriched virus replication, and protein expression in single cycle assays, but did not affect HIV-1$_{WT}$ (Figure 5.7C-G, Figure 5.8B, C). Moreover, the deficit in unspliced viral RNA that was evident in CG-enriched virus-infected cells was abolished in ZAP$^{-/-}$ cells, and reinstated in a doxycycline-dependent manner in [ZAP$_{DI}$] reconstituted ZAP$^{-/-}$ cells (Figure 5.7I). Taken together, suggesting that ZAP is required for the inhibition of the CG-high viruses and through destabilizing the CG-high, with no detectable effect on the replication of the wild-type HIV-1.

We transferred the L-mutant segment and its derivatives into a heterologous context, namely the 3'UTR of a reporter construct encoding a *fluc* gene from which we depleted CG-dinucleotides (Figure 5.9A). When these constructs were transfected into wild-type HeLa cells, he CG-enriched L-derived elements inhibited luciferase expression by ~5-fold (Figure 5.7J). Notably, these inhibitory effects were abolished when ZAP$^{-/-}$ HeLa cells were transfected, consistent with the notion that ZAP is inhibiting the expression of CG-

high RNA. (Figure 5.7K; Figure 5.9B). Similar constructs containing 3'UTR fragments derived from naturally CG-suppressed vesicular stomatitis virus or influenza virus-derived RNA sequences (Cheng et al., 2013; Greenbaum et al., 2008) (Figure 5.9A, C) revealed that elevating CG-dinucleotide content of these sequences conferred sensitivity to inhibition by coexpressed ZAP-L in cotransfection assays (Figure 5.9D). Importantly, these studies demonstrate that ZAP is capable of inhibiting mRNA outside of the context of a replicating virus purely based on their CG-high content.

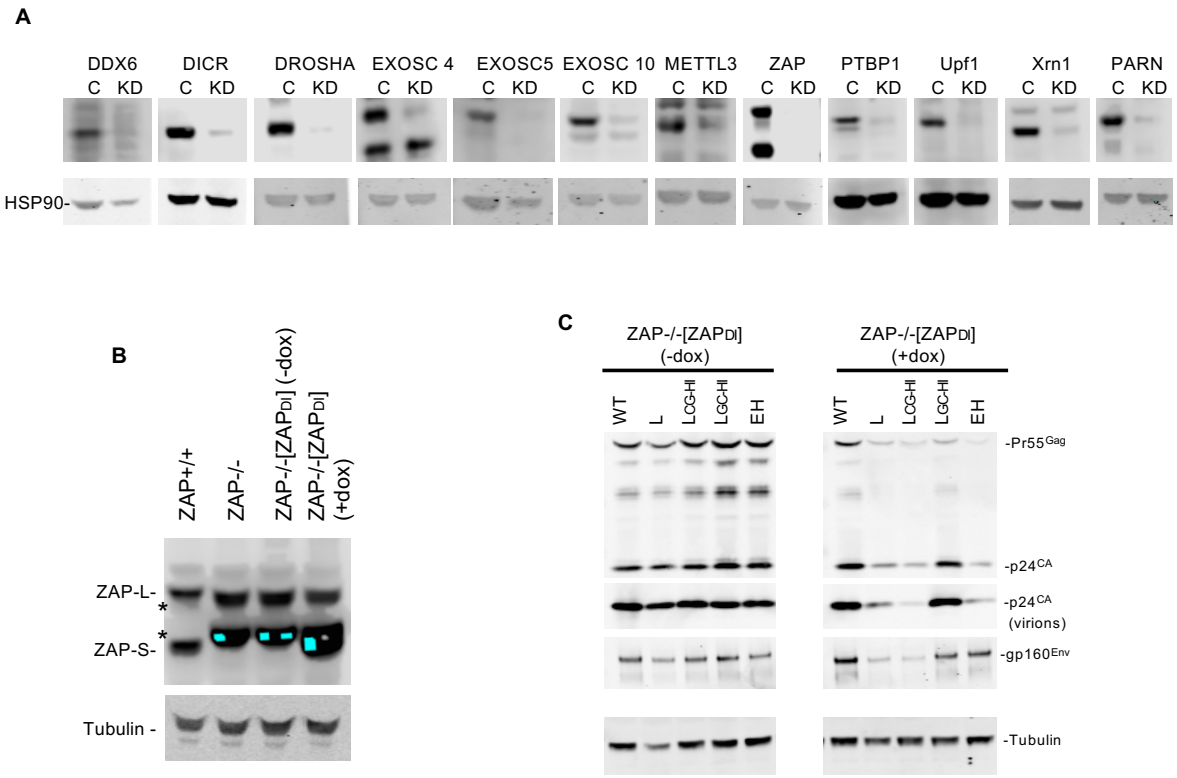


Figure 5.8: ZAP mediates deleterious effects of CG dinucleotides on HIV-1 replication

(A) Western blot analyses, using the indicated antibodies, following transfection of HeLa cells with the corresponding siRNAs, or control siRNAs, in the single-cycle replication assays described in Fig. 3a. Representative of 2 experiments. **(B)** Western blot analysis of ZAP expression in control, CRISPR-knockout MT4 cells and doxycycline-inducible ZAP-S-reconstituted MT4 cells. Asterisks indicate protein species that appeared in some CRISPR knockout clones, reacted with an anti-ZAP antibody and arose after extended passage. These are likely to represent truncated forms of ZAP-L whose translation initiated at methionine codons 3′ to the CRISPR target site (near the ZAP N terminus). Representative of three experiments. **(C)** Western blot analysis (anti-Gag, anti-Env, anti-GFP and anti-tubulin) of viral and cellular protein levels in cells and virions, 48 h after single-cycle wild-type or mutant HIV-1infection of $ZAP^{-/-}$ MT4 cells that had been reconstituted with a doxycycline-inducible ZAP-S expression construct (ZAP_{DI}) and left untreated or treated with doxycycline. Representative of three experiments.

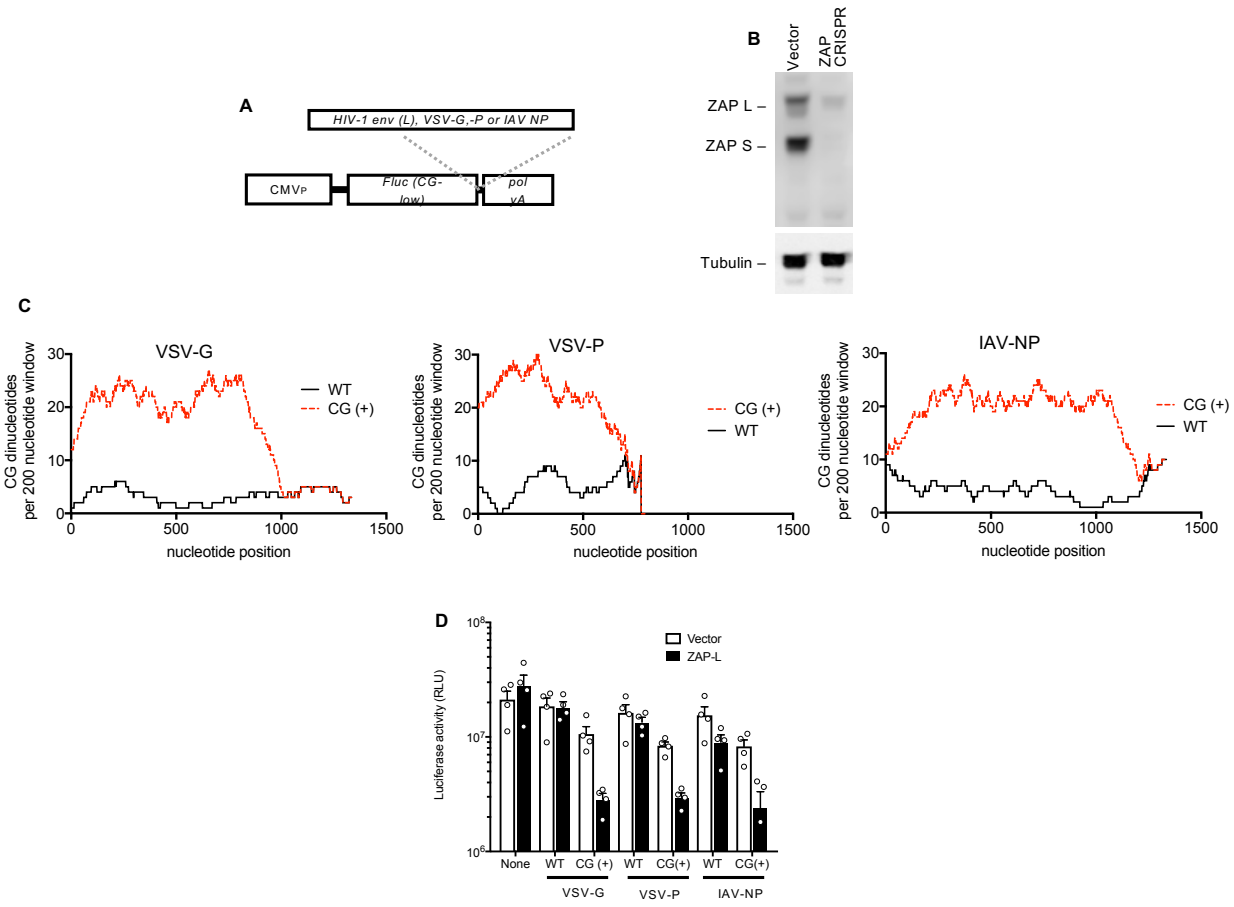


Figure 5.9: CG dinucleotides in 3' UTR confer sensitivity to inhibition by ZAP

(A) Schematic representation of a reporter construct encoding a CG dinucleotide-depleted *fluc* cDNA into which were inserted the indicated sequences as 3′ UTRs. **(B)** Western blot analysis of ZAP expression following CRISPR mutation of ZAP exon 1 in HeLa cells. Representative of three experiments. **(C)** Number of CG dinucleotides present in a 200-nucleotide sliding window in the indicated viral cDNA sequences that were left unmanipulated (WT) or recoded with synonymous mutations to contain the maximum number of CG dinucleotides (CG+). **(D)** Luciferase expression following transfection of 293T ZAP$^{-/-}$ cells with CG dinucleotide-depleted *fluc* reporter plasmids incorporating the indicated VSV or influenza A virus (IAV) RNA sequences as 3′ UTRs, in the presence or absence of a cotransfected ZAP-L expression plasmid (mean ± s.e.m., $n = 4$ independent experiments).

ZAP binds directly and preferentially to CG-dinucleotide containing RNA

ZAP has been reported to bind RNA, but no shared features of its reported target sequences are evident. To determine the RNA binding specificity of ZAP, we used crosslinking-immunoprecipitation-sequencing (CLIP-seq) assays in cells infected with HIV-1_{WT} or mutant L. Remarkably, ZAP bound to the HIV-1 genome predominantly at a location that precisely coincided with the CG-enriched segment in mutant L (Figure 5.10A, B). Conversely, ZAP bound less frequently to HIV-1_{WT} and the unaltered portions of the L genome, but those binding sites coincided with the rare occurrence of CG dinucleotides (Figure 5.10A, B).

Although the L mutant genome was the single most frequently bound RNA in infected cells, ZAP also bound cellular mRNAs (Figure 5.11B). CG-suppression is marked in human mRNA ORF and 3'UTR sequences (Figure 5.11C, D) but was absent in the subset of these sequences that represented the 100 most preferred ZAP binding sites (Figure 5.11E). A more detailed analysis of dinucleotides that are underrepresented (CG and UA) or overrepresented (UG) in ORFs and 3'UTRs as well as an 'inverted CG' control dinucleotide (GC), revealed that ZAP binding sites were highly CG-enriched (Figure 5.10C). Conversely, UA, UG or GC dinucleotides were present in preferred ZAP binding elements at frequencies typical of ORFs and 3'UTRs (Figure 5.10C). A control RNA binding protein showed no preference for CG-enriched elements (Figure 5.11G).

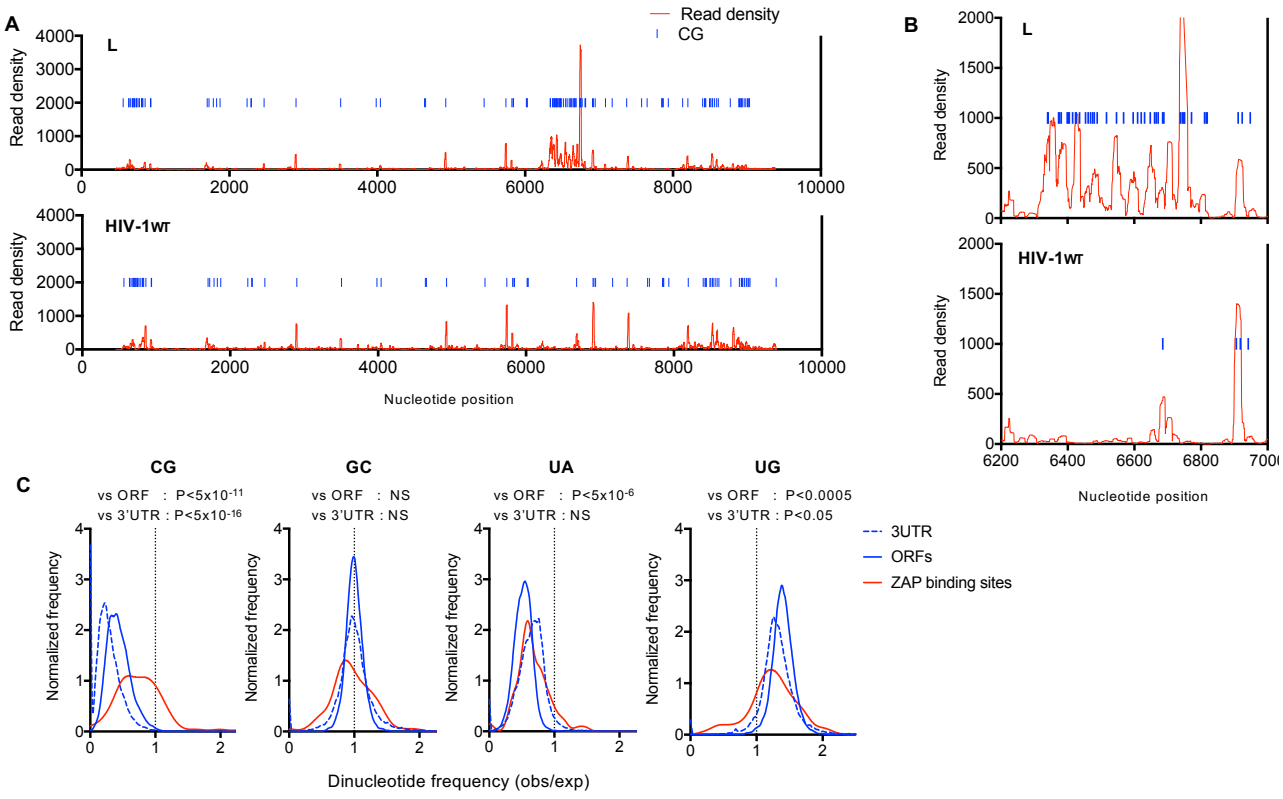
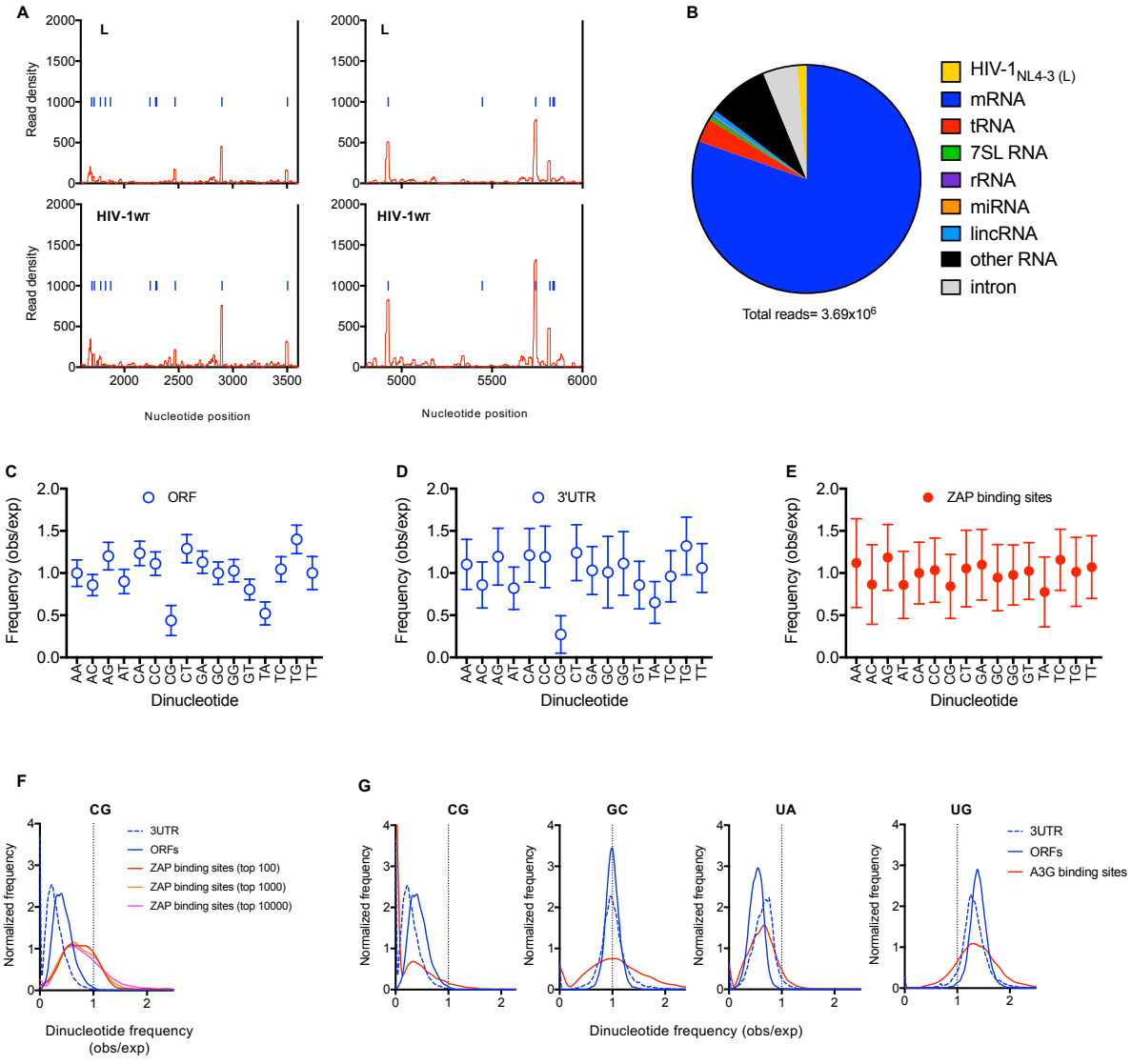


Figure 5.10: ZAP binds directly and preferentially to CG dinucleotide-containing RNA

(A) CLIP analysis of the frequency with which L mutant and HIV-1$_{WT}$ RNA sequences are bound to ZAP in infected cells, versus their position in the viral genome. Positions of CG dinucleotides are indicated as blue lines. The L-mutant segment occupies positions 6307 to 6805. **(B)** Expanded views of the CLIP graphs in (A), showing unaltered portions of the sequence (left, middle) and a portion containing the mutant L segment (right). **(C)** Frequency distributions of CG, GC, UA and UG dinucleotide observed/expected frequencies in human ORFs, 3'UTRs and top 100 ZAP-binding sites. P-values for comparisons of ZAP binding sites and ORFs or 3'UTRs were calculated using Welch's unequal variance t-tests.

Figure 5.11: Dinucleotide composition of ORFs, 3'UTRs, and preferred ZAP binding sites in cellular mRNAs

(A) Sources of RNA reads bound to ZAP in a typical CLIP-seq experiment, done using HIV-1 infected cells **(B-D)**, Ratio of the observed frequency to the expected frequency (obs/exp, based on mononucleotide composition) for each of the 16 possible dinucleotides, in ORFs (**B**), 3' UTR (**C**) sequences as well as the 100 sites in cellular mRNAs that were most frequently bound by ZAP, based on CLIP read numbers (**d**). Plotted values are mean ± sd of all ORF and 3'UTRs in the respective libraries or n=100 most preferred ZAP binding sites. **(E)** Frequency distributions of CG dinucleotide observed/expected frequencies in human ORFs, 3'UTRs and top 100, top 1,000 and top 10,000 ZAP-binding sites in CLIP experiments. The top 100, top 1,000 and top 10,000 ZAP-binding sites account for 6.7%, 18.9% and 46.7% of total reads. **(F)** Frequency distributions of CG, GC, UA and UG dinucleotide observed/expected frequencies in human ORFs, 3'UTRs and the top 100 APOBEC3G-binding sites in CLIP assays.


A
L
L
HIV-1WT
HIV-1WT
Read density
Nucleotide position
B
HIV-1NL4-3 (L)
mRNA
tRNA
7SL RNA
rRNA
miRNA
lincRNA
other RNA
intron
Total reads= 3.69x10^6
C
ORF
D
3'UTR
E
ZAP binding sites
Frequency (obs/exp)
Dinucleotide
F
CG
3UTR
ORFs
ZAP binding sites (top 100)
ZAP binding sites (top 1000)
ZAP binding sites (top 10000)
Normalized frequency
Dinucleotide frequency (obs/exp)
G
CG
GC
UA
UG
3UTR
ORFs
A3G binding sites
Normalized frequency
Dinucleotide frequency (obs/exp)

Summary

Further examination of the group 3 viruses that did not possess a splicing defect, yet had a profound replication defect revealed that it was the coincidental addition of CG dinucleotides that were responsible for their replication defects. The addition of CG dinucleotides resulted in decreased levels of unspliced RNA that was restricted to the cytoplasm and exonic mRNA. We identified ZAP as the cellular restriction factor responsible for specifically inhibiting the replication of these CG-high viruses and destabilizing CG-high viral RNA. CLIP-seq analysis of ZAP reveals that has a strong preference for binding directly to CG dinucleotides in both cellular and viral RNA despite the CG suppression of the human transcriptome, suggesting that ZAP might function as an immune sensor that distinguishes self from non-self RNA based on the CG content. It remains unclear how ZAP destabilizes the CG-high RNA once bound, and what other cofactors might be required in order to do this.

Chapter 6. Mechanistic Insights into TRIM25 and ZAP restriction of CG-enriched viruses

In our initial screen that identified ZAP as the protein responsible for the restriction of a CG-high HIV-1 also identified a protein that significantly rescued the replication of the CG-high virus, TRIM25. As discussed previously, TRIM25 is a member of the tripartite motif family of E3 ubiquitin ligases and is an essential cofactor of RIG-I (Gack et al., 2007). Two groups identified and confirmed that TRIM25 as a cofactor for ZAP and is not required for, but enhances the activity of ZAP (Jiang et al., 2016; Li et al., 2017; Zheng et al., 2017). As a follow-up to our initial studies on ZAP mediated restriction of CG-high viruses, we sought to better understand how TRIM25 might serve as a cofactor for ZAP and attempt to identify other cofactors that are important for ZAP mediated inhibition of CG-high viruses.

RNA binding activity of ZAP in the absence of TRIM25

It is reported that TRIM25 has an RNA binding domain comprised of seven lysine resides N-terminal to the SPRY domain, and these residues are important for the activity as a cofactor with RIG-I (Sanchez et al., 2018). Using CLIP-seq experiments we have shown that ZAP binds directly to CG dinucleotides in both viral RNA and cellular mRNA. Because both ZAP and TRIM25 have RNA binding activity we tested whether TRIM25 contributes to the ability of ZAP to recognize of CG-high RNA. To address this, we performed CLIP-seq experiments either with TRIM25 in the presence and absence of ZAP, or with ZAP in the presence and absence of TRIM25 (Figure 6.1). CLIP experiments with TRIM25

initially sought to determine whether TRIM25 bound any RNA in 293T cells in the presence or absence of ZAP. We also tested these conditions in the presence of either wild-type HIV-1 or CG-high HIV-1 (Figure 6.1A). Surprisingly, our crosslinking immunoprecipitation of TRIM25 with ^{32}P labeled RNA did not bind any detectable RNA under these conditions, indicating that in our experimental conditions, TRIM25 does not detectably bind RNA.

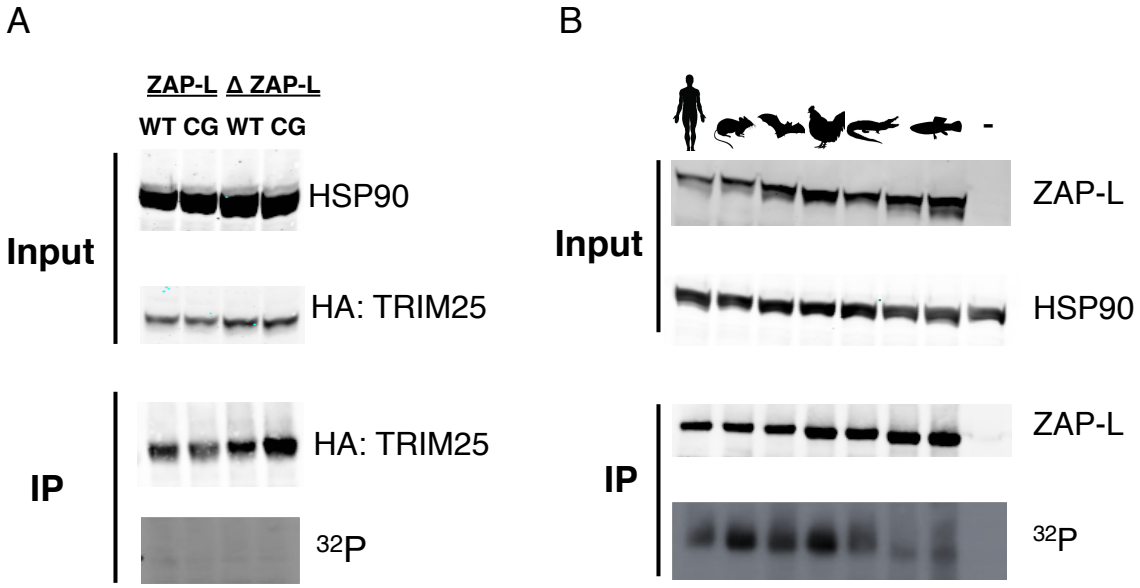


Figure 6.1: RNA binding activities of TRIM25 and ZAP

(A) Input and TRIM25-RNA cross-linked complexes were immunoprecipitated from 293T cells knocked out for ZAP and TRIM25 reconstituted with TRIM25 and ZAP-L (first two lanes) or TRIM25 in the absence of ZAP-L (last two lanes), all cotransfected with either HIV-1 wild type or HIV-1 CG-high (as indicated). Cells were transfected to express 3xHA-tagged-TRIM25 and then fed 4SU and UV-irradiated. Complexes were visualized by autoradiography (bottom) and western blot analysis using a polyclonal anti-HA or monoclonal anti-HSP90. **(B)** Input and ZAP-RNA crosslinked complexes immunoprecipitated from 293T cells knocked out for ZAP and TRIM25, reconstituted with ZAP-L from varies species (Homo sapien, Mus musculus, Eptesicus fuscus, Gallus gallus, Alligator mississippiensis, Danio rerio PARP12A, Danio rerio PARP12B, or empty), co-transfected with HIV-1 CG-high. Cells were transfected to expressed 3xHA-ZAP-L and

then fed 4SU and UV-irradiated. Complexes were visualized by autoradiography (bottom) and western blot analysis using a polyclonal anti-HA or monoclonal anti-HSP90.

Using a similar approach, we tested whether the RNA binding activities of ZAP are dependent on the presence of TRIM25. In a crosslinking immunoprecipitation of ZAP in the absence of TRIM25, ZAP produced a robust RNA signal, indicating that the general RNA binding activity is independent of TRIM25 (Figure 6.1B). Sequencing the RNA pulled down by ZAP in the absence of TRIM25 provided further insights. In comparing the binding profile of ZAP to the CG-high HIV-1 genome it is clear that ZAP is only bound to the HIV-1 genome where a CG dinucleotide is present (Figure 6.2A). There are no peaks along the viral RNA in the absence of a CG dinucleotide and its binding is significantly enriched at the site of the 43 additional CG dinucleotides near the 5' end of envelope. This indicates that ZAP is capable of recognizing CG dinucleotides in viral RNA independent of TRIM25. Analysis of the reads bound by ZAP in open reading frames demonstrated that ZAP maintained a strong preference for CG-high RNA irrespective of whether TRIM25 is present or absent. (Figure 6.2B). Finally, the species of RNA bound by ZAP was virtually identically with or without TRIM25, where the majority of reads going came from mRNA (Figure 6.2C). Collectively, these experiments indicate that ZAP is able to recognize CG-high mRNA independent of TRIM25. This suggests that TRIM25 enhances the antiviral activity of ZAP in a mechanism other than aiding ZAP in recognition of CG-high RNA.

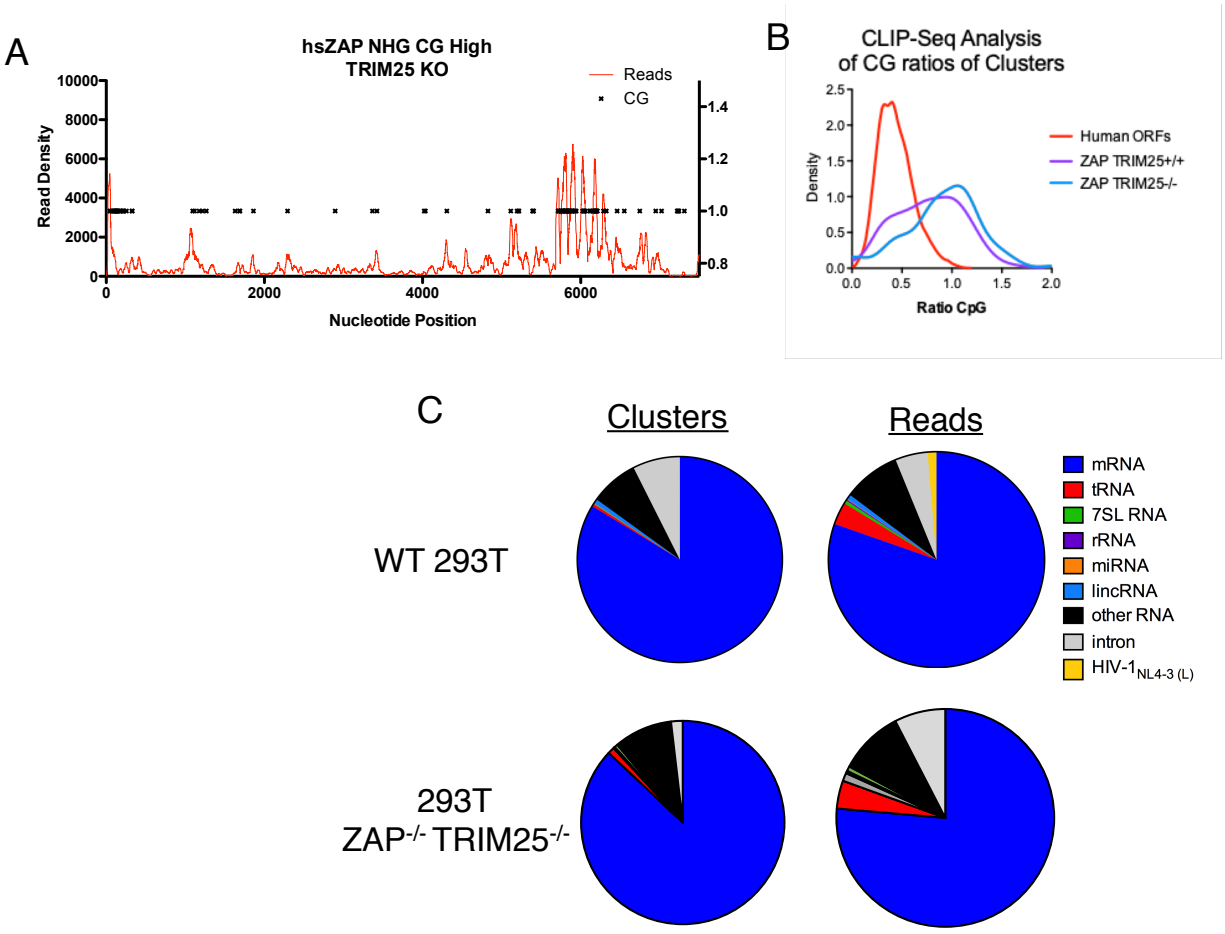


Figure 6.2: RNA targets of human ZAP-L in 293T without TRIM25 expression

(A) Frequency distribution of nucleotide occurrence (read density) in reads that were mapped to the HIV-1_{NHG} CG-high genome. **(B)** Density plot of the CG odds ratio from the top 100 PARalyzer clusters bound by ZAP-L in the presence (magenta) or absence (cyan) of TRIM25, compared against the CG odds ratio of the human genome ORFs (red). **(C)**

Origins of individual reads that map to the human genome in ZAP-L CLIP experiments in the presence or absence of TRIM25 293T cells transfected with HIV-1_{NHG} CG-high.

Biological activities of TRIM25 required for activity with ZAP

Because TRIM25 does not contribute to the ability of ZAP to bind RNA, we sought to test the other mechanisms by which TRIM25 could enhance the antiviral activity of ZAP. TRIM25 is composed of five domains: RING, B-box, coil-coil, RNA binding, and SPRY (Figure 6.3A). The RING domain possess E3 ubiquitin-ligase activity, but also contributes to the formation of higher order multimers (Sanchez et al., 2016). The function of the B-box domain in TRIM25 remains unclear, but in another TRIM protein, TRIM5α, the B-box domain is involved in formation of higher order multimers (Li and Sodroski, 2008). The coil-coil domain allows the formation of elongated antiparallel dimers (Sanchez et al., 2014). The SPRY domain has been shown to directly interact with the two proteins for which TRIM25 acts as a cofactor: ZAP and RIG-I (Gack et al., 2007; Li et al., 2017). Therefore, we tested these four biological activities of TRIM25: the ubiquitin ligase activity, formation of higher order multimers, RNA binding activity, and the requirement of the SPRY domain.

To test each of these activities without disrupting the global structure of the protein we made point mutations that disrupt the E3 ubiquitin-ligase activity (R54A), higher order multimerization (L69A), and RNA binding (7KA), as well as a TRIM25 that does not have a SPRY domain (Sanchez et al., 2016; Sanchez et al., 2018). 293T cells that were knocked out for both ZAP and TRIM25 were co-transfected with plasmids expressing (i.) either a wild-type or CG-high HIV-1, (ii.) ZAP-S, and (iii.) either a wild-type or mutant

TRIM25. The supernatants were filtered and titered on MT4 LTR-GFP cells to determine the viral yield.

When a plasmid expressing either a wild-type or CG-high virus were transfected with plasmids expressing ZAP-S and wild-type TRIM25, 50-fold less virus was produced by the CG-high virus compared to wild-type HIV-1 (Figure 6.3B). This restriction of the CG-high virus was dose dependent on amount of ZAP plasmid transfected. Additionally, ZAP was incapable of inhibiting the CG-high virus in the absence of TRIM25 (Figure 6.3B). The RNA binding mutant of TRIM25 (7KA), which has the patch of seven lysines mutated to alanines (Sanchez et al., 2018) was equally capable of restricting production of the CG-high HIV-1 as compared to the wild-type TRIM25 (Figure 6.3C). The E3 ubiquitin ligase mutant, which is unable to catalyze the addition of ubiquitin (Sanchez 2016), was also not reduced in its ability to restrict the CG-high virus when compared to the wild-type TRIM25 (Figure 6.3D). The TRIM25 mutant that is unable to form higher order multimers (L69A) was 10-fold less functional as a cofactor for ZAP (Figure 6.3E). Finally, the mutant of TRIM25 that had no SPRY domain had no detectable activity in this assay. Taken together, this data suggests that the RNA binding activity and E3 ubiquitin-ligase activity of TRIM25 are dispensable for its activity as a cofactor for ZAP (Figure 6.3F). Further, the SPRY domain and the ability of TRIM25 to form higher order multimers appear essential to its function as a cofactor for ZAP.

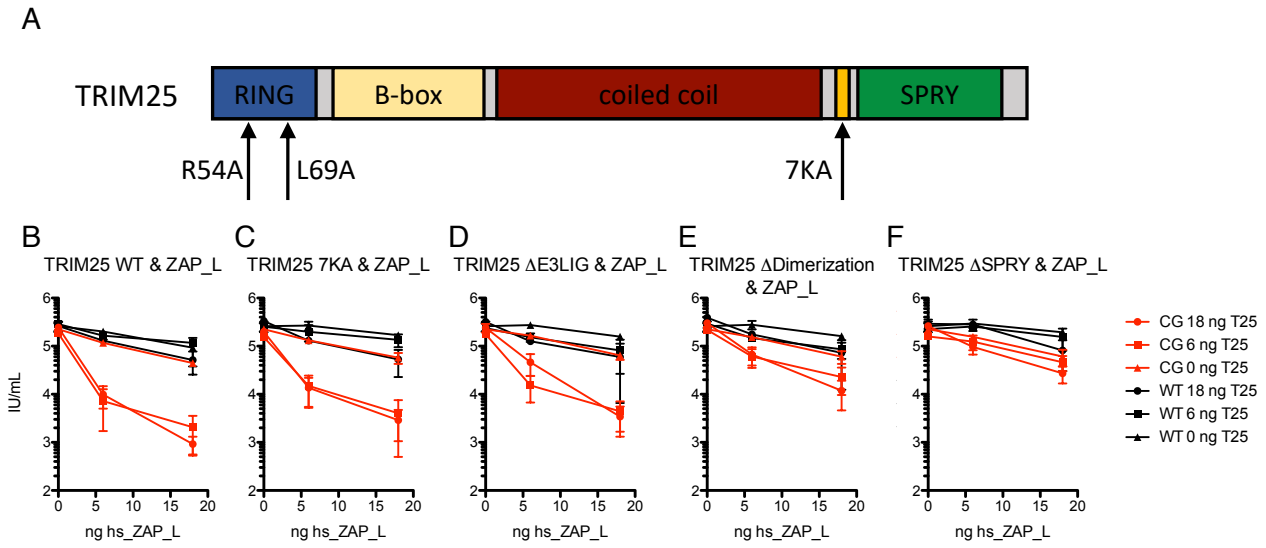


Figure 6.3: Antiviral activity of ZAP with different TRIM25 mutants

(A) Schematic representation of the domains in human TRIM25, arrow and labels below indicate the point mutants or introduced into the protein for analysis below. (B-F) Infectious virion yield measured using MT4-GFP indicator cells following transfection with ZAP-L, TRIM25 mutants, and HIV-1_{NL4-3} wild-type or HIV-1_{NL4-3} CG-high (mean ± s.d., n=2). **(B)** TRIM25 wild-type **(C)** TRIM25 RNA binding mutant 7KA. **(D)** TRIM25 R54A deficient in E3 ubiquitin ligase activity. **(E)** TRIM25 L69A deficient in proper dimerization and formation of high order multimers. **(F)** TRIM25 truncated to exclude the SPRY domain.

Cell type differences of TRIM25 enhancement of ZAP

The experiments shown above with ZAP and TRIM25 suggest that there might be cell type dependence concerning how the activity of ZAP requires TRIM25. siRNA knockdown of TRIM25 in HeLa cells only partially rescued the replication of the CG-high virus (Figure 5.7A). In contrast, when ZAP was expressed without TRIM25 in the 293T double knockout cells, ZAP was inactive (Figure 6.3A). While this discrepancy might be to a due to an incomplete knockdown in HeLa cells compared to a complete knockout of TRIM25 in the 293T cells, we wanted to further investigate whether the activity of ZAP differs based on the cell line.

Using a CRISPR guide RNA against TRIM25, we generated TRIM25 knockout MT4 cells. With these cells, we sought to understand how ZAP knockout and TRIM25 knockout cells compare in restricting the replication of CG-high HIV-1. MT4 wild-type, MT4 ZAP knockout, and MT4 TRIM25 knockout cells were infected with either wild-type virus, a CG-high HIV-1 or a HIV-1 that has half as many CG dinucleotides as the CG-high virus, mutant LB. The MT4 cells were infected at an MOI of 1.0, resulting in approximately 66% of the cells becoming infected. Infected cells were analyzed for viral protein expression while supernatants were filtered and tested for infectious viral yield by titering the supernatants on MT4-GFP cells.

The CG-high virus produced no detectable gp160 protein in wild-type MT4 cells, but there was a complete rescue of gp160 expression when ZAP was knocked out (Figure 6.4A). Interestingly, when TRIM25 was knocked out, the rescue in gp160 expression was modest. When comparing the expression of Gag under these same conditions, ZAP

knock out provided a complete rescue for the CG-high virus. In contrast, TRIM25 knock out did not result in any detectable rescue in Gag expression for the CG-high virus. Similarly, for the virus with half as many CG dinucleotides (mutant LB), ZAP knockout completely rescued all reductions in viral protein expression, while TRIM25 knockout only modestly increased the expression of gp160 and did not increase the expression of capsid (Figure 6.4A). Analysis of infectious virus yield from these infected cells revealed that TRIM25 had only a modest effect on the replication of the CG-high virus (Figure 6.4B). Knockout of TRIM25 increased the infectious virus yield from the CG-high virus by only 3-fold, while ZAP knockout rescued the infectious virus yield by nearly 2,000-fold. For mutant LB, TRIM25 knockout increased the infectious virus yield by 10-fold when compared to the MT4 wild-type cells, while ZAP knockout increased the infectious virus yield by nearly 500-fold as compared to the MT4 wild-type cells. We also tested these cell lines and viruses in a spreading replication assay (Figure 6.4C - E). While ZAP knockout completely rescued the replication of both CG-high viruses, TRIM25 knockout in MT4 cells did not permit either of the CG-enriched viruses to spread in culture. Collectively, these results suggest that TRIM25 is not required for the ZAP mediated restriction of CG-high viruses in MT4 cells.

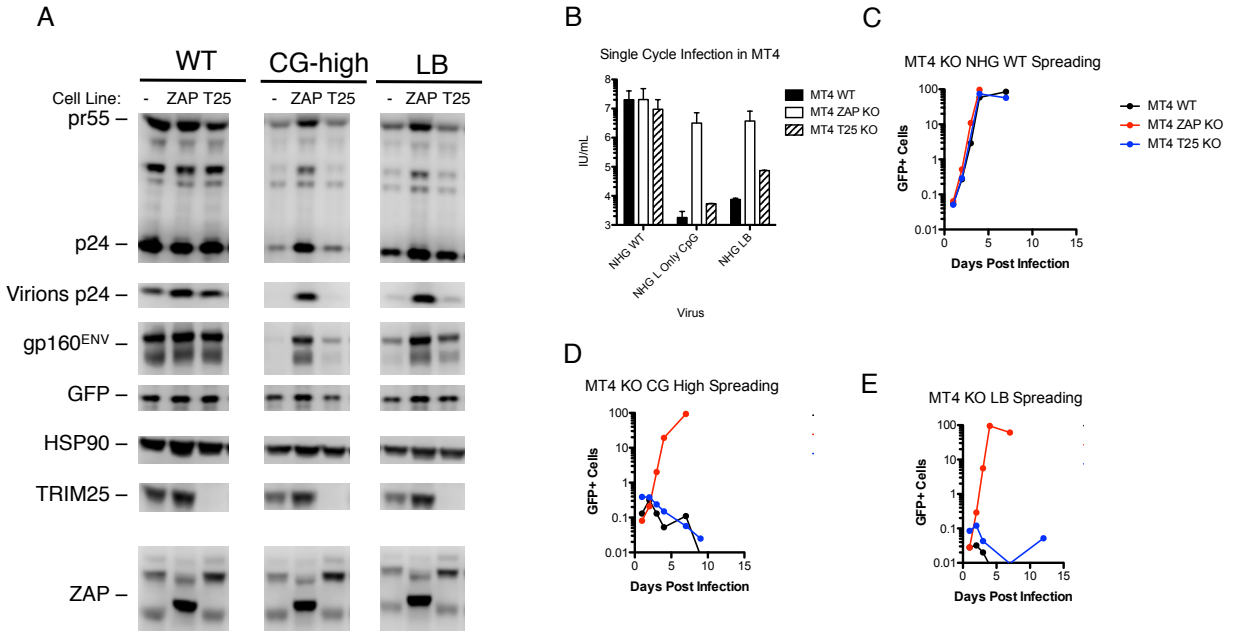


Figure 6.4: Infections of CG-enriched viruses in MT4 cells

(A) Western blot analysis 48 h after a single-cycle infection of MT4 cells with the wild-type of mutant HIV-1, representative of three experiments. **(B)** Single-cycle infectious virus yield, following infection of MT4 cells with equal titers of wild-type HIV-1 and mutants (mean ± s.d., n=3). **(C-E)** Replication of HIV-1 mutants in wild-type MT4 cell, ZAP KO MT4 cells, or TRIM25 KO MT4 cells, as measured by FACS enumeration of infected cells. **(C)** HIV-1 wild-type, **(D)** HIV-1 CG-high, **(F)** HIV-1 LB.

Comparing MT4 and 293T TRIM25 knockout cells revealed that ZAP dependence on TRIM25 to restrict CG-high HIV-1 varies dramatically on the cell line. Previously we noted that cell lines vary in the magnitude with which they inhibited CG-high HIV-1. Comparing the virus yield of a CG-high HIV-1 to a wild-type HIV-1 in a single cycle of replication assay, MT4 cells are capable of restricting the CG-high HIV-1 1,000-fold, HeLa cells capable of restricting 100-fold, and A549 cells capable of restricting only 10-fold. To more precisely quantify this phenomenon, we set out to compare the infectious virus yield from a wild-type virus and the CG-high virus at a variety of MOI to measure the effects of ZAP and TRIM25 in different cell lines. HIV-1_{NHG} was used and cells were infected with virus at different dilutions. Supernatants from the infections were collected and titered on MT4 cells to determine the infectious virus yield. The infected cells were also analyzed by flow cytometry to determine the percentage of cells infected and calculate the MOI . For each cell line, virus and dilution, we calculated the IU/mL in the supernatant and plotted that against the MOI in that well. The adherent cell lines used were A549 wild-type and ZAP knockout (lung adenocarcinoma cells), HOS and U2OS (osteosarcoma cells) and HeLa wild-type and ZAP knockout (cervical cancer cells). The T-cell lines used were H9, SupT1, Jurkat, and MT4 cells (both ZAP knockout and TRIM25 knockout).

The U2OS, HOS and A549 all behaved similarly, as the MOI approached 1.0 the cells infected with the CG-high virus yielded infectious progeny, at a level that was about 10-fold less than cells infected with wild-type HIV-1 (Figure 6.5 A, C, E). At higher MOI the difference in yield between the CG-high virus and the wild-type virus became less pronounced. The HeLa cells behaved differently, even at an MOI of 1.0 they did not

produce any detectable infectious virus from the CG-high HIV-1, while producing more than $1x10^6$ infectious units from the wild-type virus. The HeLa cells only produced detectable infectious virus from the CG-high virus as the MOI approached 10 (Figure 6.5C). For the HeLa and A549 ZAP knockout cells, the CG-high and wild-type virus behaved identically, both producing comparable titers at each of the different MOI (Figure 6.5 D, F). The data collected from the T-cell lines more closely resembled the pattern observed from the HeLa cells. Even at the highest MOI used to infect any of the wild-type T-cell lines, there was hardly any detectable infectious virus produced from the CG-high virus, despite the high titers virus generated by wild-type virus infected T-cell lines at the same MOI (Figure 6.6 A, B, C, D). The ZAP knockout MT4 cell line produced the same infectious virus yield at each of the MOI for the wild-type virus and the CG-high virus (Figure 6.6F). Interestingly, the MT4 TRIM25 knockout cell behaved unlike any of the T-cell lines tested, but rather responded more similarly to the adherent cell lines (Figure 6.6E). This cell line restricted the production of CG-high virus at a MOI < 1.0, but at MOI > 1.0 become strikingly permissive for the production of the CG-high virus. MT4 TRIM25 knockout cells infected with the CG-high virus at an MOI approaching 10 produced nearly $1x10^6$ IU/mL, whereas the same virus in MT4 wild-type cells did not produce any detectable infectious virus even at an MOI above 10.

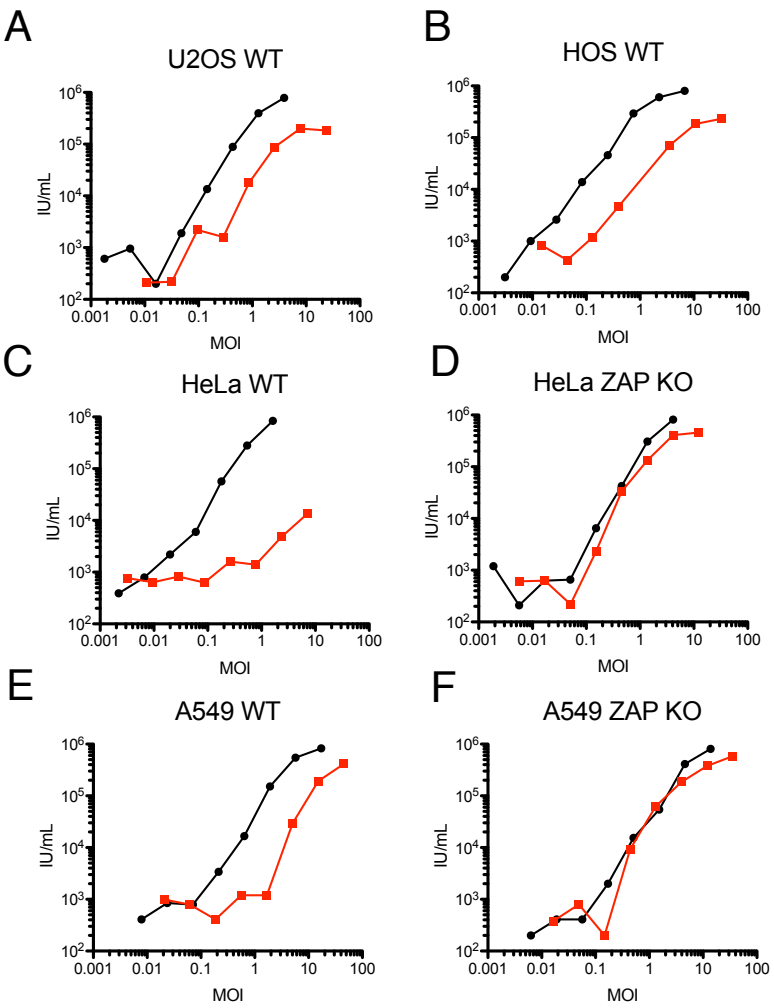


Figure 6.5: Comparison of virus yield and MOI in adherent cell lines by wild-type of CG-enriched HIV-1

(A-F) Single-cycle infectious virus yield, following infection of adherent cell lines at different MOI. MOI was determined FACS infected cells in each well for expression of GFP reported expression from incoming virus. Black lines: wild-type HIV-1; Red lines: CG-high HIV-1. Virus yield was determined by infecting MT4-GFP reporter cells. **(A)** U2OS wild-type, **(B)** HOS wild-type, **(C)** HeLa wild-type, **(D)** HeLa ZAP KO, **(E)** A549 wild-type, **(F)** A549 ZAP KO.

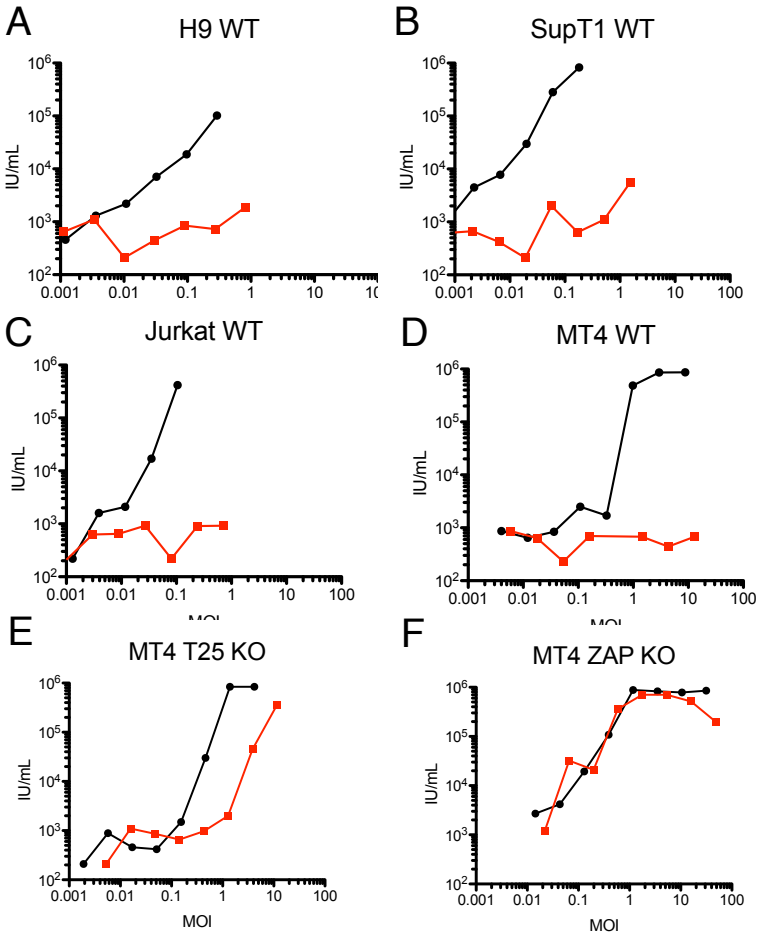


Figure 6.6: Comparison of virus yield and MOI in T-cell lines by wild-type of CG-enriched HIV-1

(A-F) Single-cycle infectious virus yield, following infection of t cell lines at different MOI. MOI was determined FACS infected cells in each well for expression of GFP reported expression from incoming virus. Black lines: wild-type HIV-1; Red lines: CG-high HIV-1. Virus yield was determined by infecting MT4-GFP reporter cells. **(A)** H9 wild-type, **(B)** SupT1 wild-type, **(C)** Jurkat wild-type, **(D)** MT4 wild-type **(E)** MT4 TRIM25 KO, **(F)** MT4 ZAP KO.

The loss of TRIM25 in MT4 cells phenotypically transformed them to behave more similarly to the adherent cell lines tested, suggesting the possibility that the adherent cell lines might be more permissive to CG-high viruses at a higher MOI due to lower levels of TRIM25 (Figure 6.7A, B). To investigate this possibility, we examined at the differential expression of ZAP and TRIM25 in all of the cell lines previously tested. For each cell line, we counted $5x10^5$ cells and lysed them in 0.5 mL of lysis buffer for analysis by western blotting. While the size of the cytoplasm is different for each of these cell lines, the size of their nucleus should be roughly similar (Figure 6.7C). For the reason, we blotted for both tubulin as a surrogate for the approximate size of the cytoplasm and the nuclear envelope marker LaminB1 as a loading control for equal cells lysed (Kane et al., 2018). All of the T cells tested had significantly less tubulin expression, which correlates with their significantly smaller cytoplasm. Comparing the levels of TRIM25 across these cell lines revealed that TRIM25 expression is more closely correlated with the tubulin levels rather than the ability to restrict a CG-high virus production. Notably, all of the adherent cell lines expressed high levels of TRIM25 despite relatively weak restriction. Comparing ZAP expression across all of the cell lines revealed that the short and long isoform are differentially expressed across the different cell lines in a manner that does not correlate with the tubulin levels. The long isoform of ZAP was expressed at relatively similar levels across all cell lines tested, despite the smaller cytoplasm in the T-cells. The short isoform was expressed at low levels in U2OS, HOS and A549 cells, but was expressed at high levels in HeLa cells and in all of the T-cell lines tested. Strikingly, this expression pattern

correlated with the ability to potently restrict CG-high viruses. Quantification of the ZAP-S and tubulin expression across these cell lines further demonstrated that a high ratio of ZAP-S to tubulin, or the ratio of ZAP to the cytoplasmic volume, exhibited a strong correlation with the level of restriction of CG-high HIV-1 (Figure 6.7D).

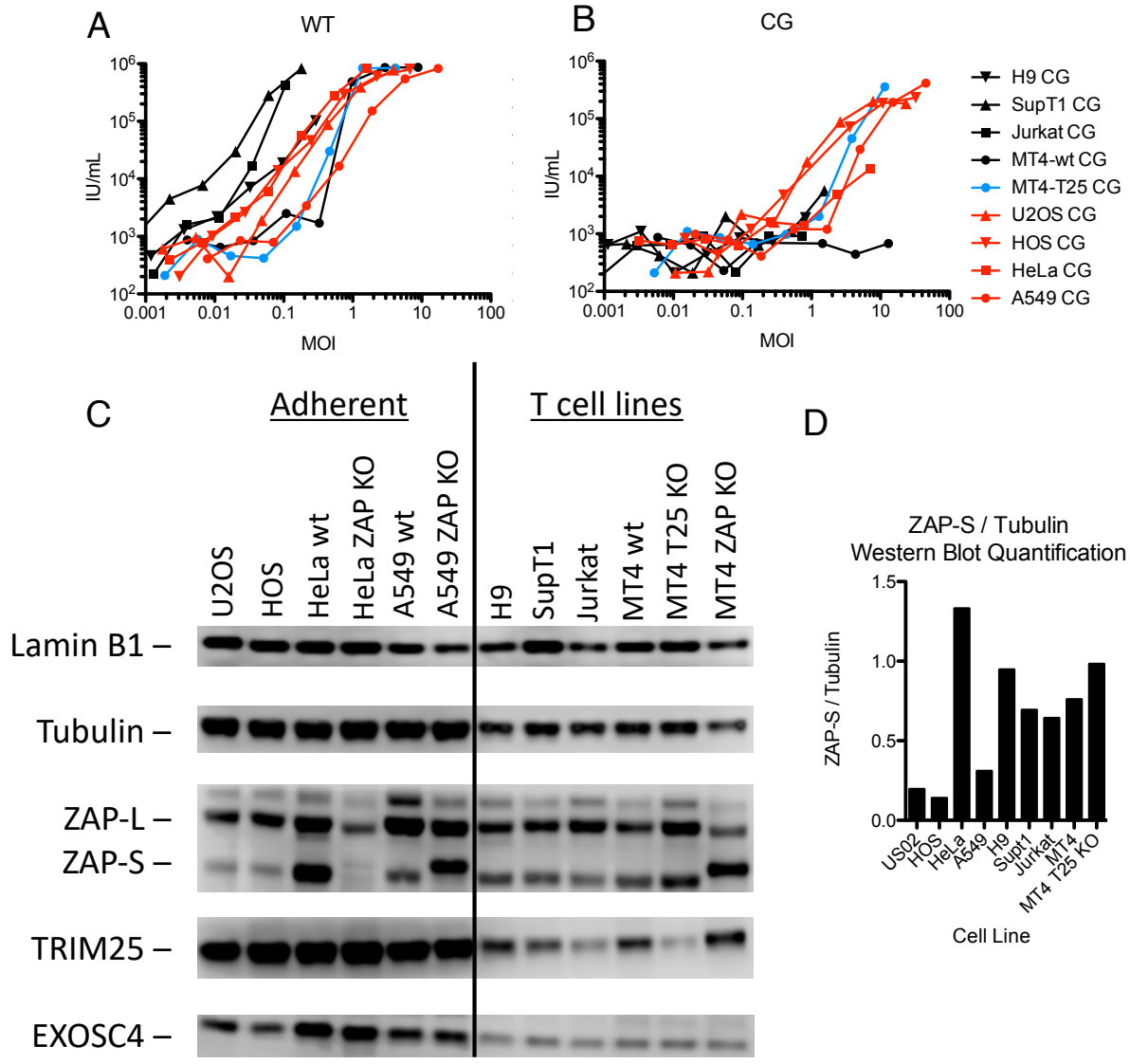


Figure 6.7: Comparison of adherent and t cell lines

(A) Single-cycle infectious virus yield, following infection of adherent cell lines (red), T-cell lines (black) and MT4 TRIM25 KO (blue) tested at different MOI with wild-type HIV-1. **(B)** Single-cycle infectious virus yield, following infection of adherent cell lines (red), T-cell lines (black) and MT4 TRIM25 KO (blue) tested at different MOI with CG-High HIV-1. **(C)** Western blot analysis of protein expression at steady state in adherent cell lines and T-cell lines. **(D)** Quantification of the ratio of ZAP-S to tubulin expression in all of the cell lines from western blots in (C).

To experimentally test whether ZAP-S is either more potent or less dependent on TRIM25 than ZAP-L, we used the previously mentioned cotransfection assay. We transfected (i.) either plasmids expressing ZAP-S or ZAP-L, (ii.) either plasmids expressing TRIM25 or an empty plasmid, and (iii.) different dilutions of plasmids expressing wild-type or a CG-high virus to mimic different MOI. In the absence of TRIM25, both isoforms were inactive against the CG-high virus at all of the proviral concentrations used, suggesting that both isoforms are dependent on TRIM25 in 293T cells (Figure 6.8 A, B). When a plasmid expressing TRIM25 was added to the cotransfection both ZAP isoforms potently restricted the CG-high virus at low concentrations of proviral plasmid, but both ZAP isoforms also became less potent as the amount of cotransfected proviral plasmid was increased. Collectively, this data suggests that in 293T cells both ZAP-S and ZAP-L are TRIM25 dependent and equally potent against the CG-high HIV-1.

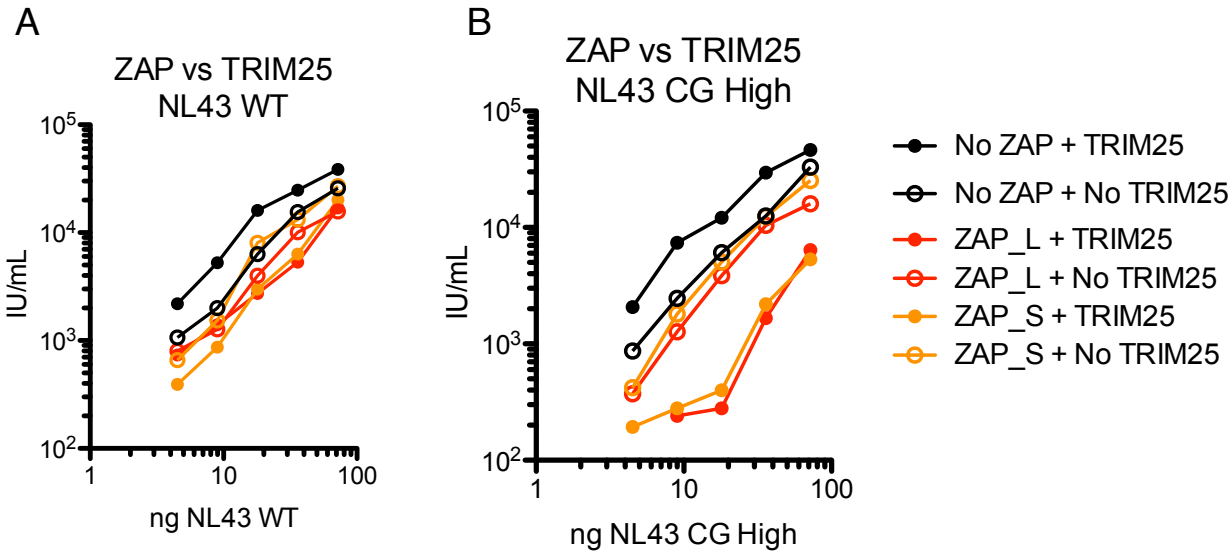


Figure 6.8: Virus yield assay in 293T cells comparing activity of ZAP isoforms

(A,B) Infectious virion yield measured using MT4-GFP indicator cells following transfection with ZAP-L and ZAP-S, TRIM25 or empty plasmid, and HIV-1_{NL4-3} wild-type or HIV-1_{NL4-3} CG-high with different concentrations of virus. **(A)** HIV-1_{NL4-3} wild-type. **(B)** HIV-1_{NL4-3} CG-high.

Effects of interferon on ZAP restriction of CG-high viruses

In the previous section, we detailed how the potency of ZAP mediated restriction of CG-high HIV-1 can vary depending on the cell line infected. MT4 cells are the least permissive to replication of CG-high HIV-1, while 293T and A549 are the most permissive to replication of CG-high HIV-1. One possibility is the less permissive cell lines have an additional cofactor that increases the potency of ZAP. Interferon stimulated genes (ISGs) are known to restrict a wide array of viruses in different cell lines through a diverse range of mechanisms (MacDonald et al., 2007). Therefore, we added interferon to cell lines that are permissive to the replication of CG-high HIV-1 to determine if this would increase the potency of ZAP, a method that has been previously demonstrated to work with ZAP restriction of Sindbis virus before the discovery of TRIM25 as a cofactor (MacDonald et al., 2007). We hypothesized that the addition of IFN might induce the expression of a ZAP cofactor in these cell lines that is not constitutively expressed. For these experiments, we used A549 cells treated with different concentrations of IFNα and IFNγ. To avoid restriction of incoming virus that is not ZAP dependent, we added IFN 18 hours post infection so as to only express genes after the virus had integrated into the cell (Holmes et al., 2015). We assayed these infected cells for viral protein expression, specifically looking for effects on gp160 expression, the viral protein with CG dinucleotides we know the be most sensitive to ZAP. We used 500, 50 and 5 units of IFNα and IFNγ, and blotted for Mx2 as a positive control for IFNα treatment and phosphorylated STAT1 (p-STAT1) as a marker of IFNγ treatment (Figure 6.9 A). Under both IFN conditions we saw a robust and dose-dependent induction of our control genes by western blot. The expression the

gp160 and GFP in the wild-type HIV-1 was unaffected by the interferon treatment, indicating that we had introduced IFN at the proper time as to not have any effect on the wild-type virus. The gp160 expression by the CG-high virus was inhibited by the addition of both IFNα and IFNγ in a dose-dependent manner (Figure 6.9A). Notably, TRIM25 was expressed at slightly higher levels after treatment with either IFN at all doses, while ZAP-S was also expressed at higher levels with either IFN treatment but the expression was dose-dependent. As a control cell line, we conducted the same experiment in A549 cells that were knocked out for ZAP and TRIM25 to demonstrate that these effects are dependent ZAP and TRIM25 activity. In these double knockout cells, IFN treatment had no effect on either the wild-type or CG-high virus (Figure 6.9B). While IFN treatment can increase the restriction of CG-high HIV in a dose-dependent manner, the concomitant increase of ZAP-S expression precludes us from concluding that this is due to an additional factor.

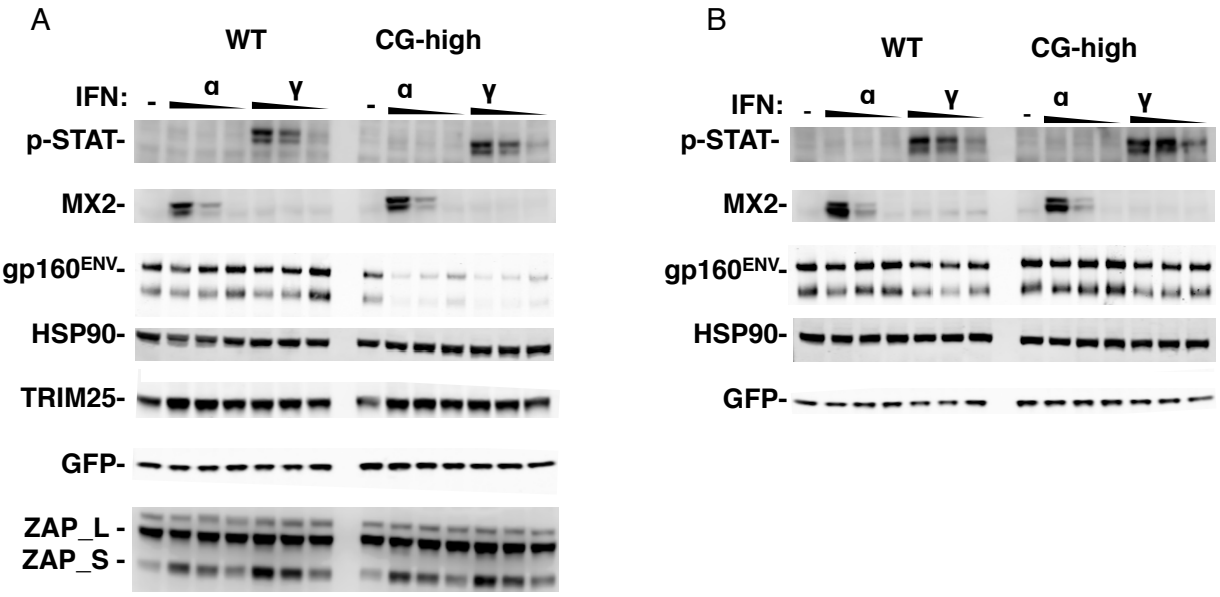


Figure 6.9: Effect of IFN treatment on replication of CG-high virus in A549 cells

(A,B) Western blot analysis 48 h after a single-cycle infection of A549 cells with wild-type or CG-High HIV-1 treated with IFN-α or IFN-γ at different concentrations. IFN was added to the cells 18 hours post infection at the time of replacing the media. MX2 represents the positive control for IFN-α treatment, p-STAT represents the positive control for IFN-γ treatment. **(A)** wild-type A549 cells. **(B)** ZAP and TRIM25 KO A549 cells.

Screening for novel ZAP cofactors

The previous sections detailed the dramatic cell type dependent differences that exist in the restriction of CG-high HIV-1. MT4 cells are capable of restricting the replication of CG-high HIV-1 independent of TRIM25 (at least at low MOI) while 293T cells are completely dependent on TRIM25 to restrict CG-high HIV-1. It is possible that this is due to nuances in the expression of each isoform or heterogeneity of ZAP expression in the cellular population. It is also possible that ZAP might depend on an additional cofactor besides TRIM25 that is not present in all of the cell lines we have tested. To test this latter hypothesis, we conducted a screen with siRNAs directed against possible cofactors. ZAP interacting proteins were identified in wild-type 293T cells in the absence of viral infection (Goodier et al., 2015). ZAP was immunoprecipitated both in the presence and absence of ribonuclease treatment and a total of seventy-eight genes identified, all of which were included in our siRNA library. As performed previously (Figure 5.7A) genes were knocked down with siRNAs, the cells were infected at an MOI of 1.0 with either wild-type or CG-high HIV-1, and the infectious virus yield produced was measured by tittering the supernatants on MT4-GFP cells. We tested this library in HeLa and A549 cells to validate that hits would be supported by results in multiple cell lines. As a positive control, siRNAs against ZAP and TRIM25 were included in the library and their knockdown reproducibly rescued the replication of the CG-high virus. ZAP knockdown completely restored the replication of the CG-high HIV-1 and TRIM25 knockdown partially rescued replication. We sought to ensure that any effects observed were due to an increase in virus yield in

this single cycle replication assay as opposed to an increase in incoming virus. To ensure our siRNA did not have an impact in the susceptibility of the cells to the incoming virus we also measured the percentage of cells that were infected for each knockdown. To do this we used HIV-1_{NHG} so we could measure the percentage of cells infected by the GFP reporter. In the HeLa screen, other than ZAP and TRIM25, knockdown of no other genes significantly in increased the infectious virus yield from the CG-high HIV-1 with the exception of the well containing the siRNA labeled LARP1 (Figure 6.10). Upon further investigation of this gene, it became clear that the well in the siRNA screening plate was contaminated with an siRNA against TRIM25, and LARP1 knockdown alone did not actually rescue the replication of the CG-high virus (data not shown).

We also screened this library with mutant LB, which has half as many CG dinucleotides as the CG-high HIV-1. Because mutant LB is significantly less attenuated than the CG-high HIV-1, the dynamic range of this assay was decreased to 10-fold, but knockdown of ZAP still clearly rescued the replication of LB (Figure 6.11). Surprisingly, knockdown of TRIM25 did not affect the replication of the mutant LB, though previously in MT4 cells, loss of TRIM25 did affect the replication of mutant LB (Figure 6.4B). Similar to the initial screen with the CG-high virus, none of the genes knocked down in the library had large effects on the replication of the mutant LB. The same library tested in A549 cells did produce genes that rescued the replication of the CG-high virus when knocked down, namely USP9X and TUFM, both ubiquitin specific peptidases (Figure 6.12). Follow up work is currently ongoing to validate these hits as potential components to the ZAP pathway.

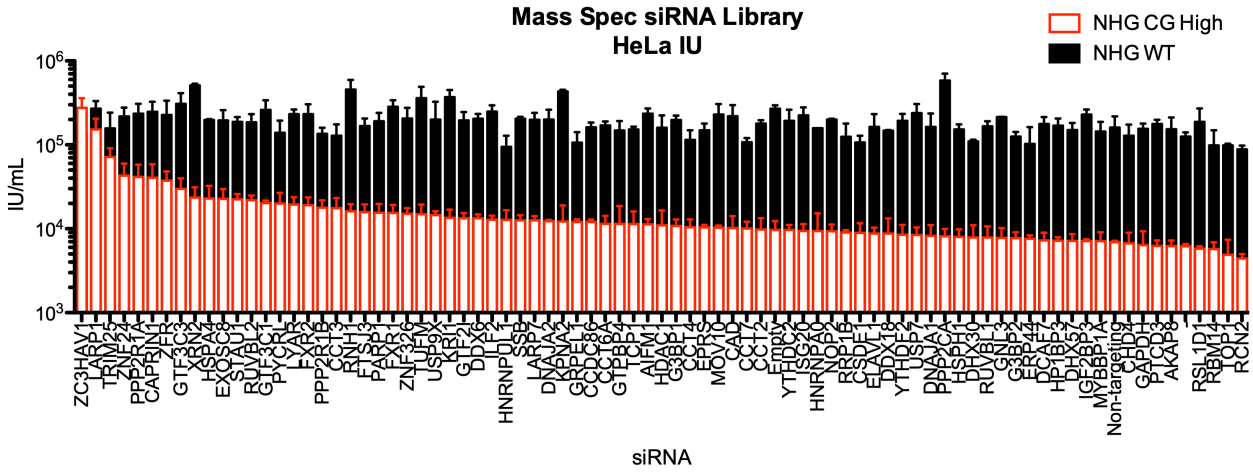


Figure 6.10: siRNA screen from Mass Spectrometry hits in HeLa cells

Single-cycle infectious HIV-1 wild-type and CG-High yield from siRNA-transfected HeLa cells (mean±sd. (n=3)).

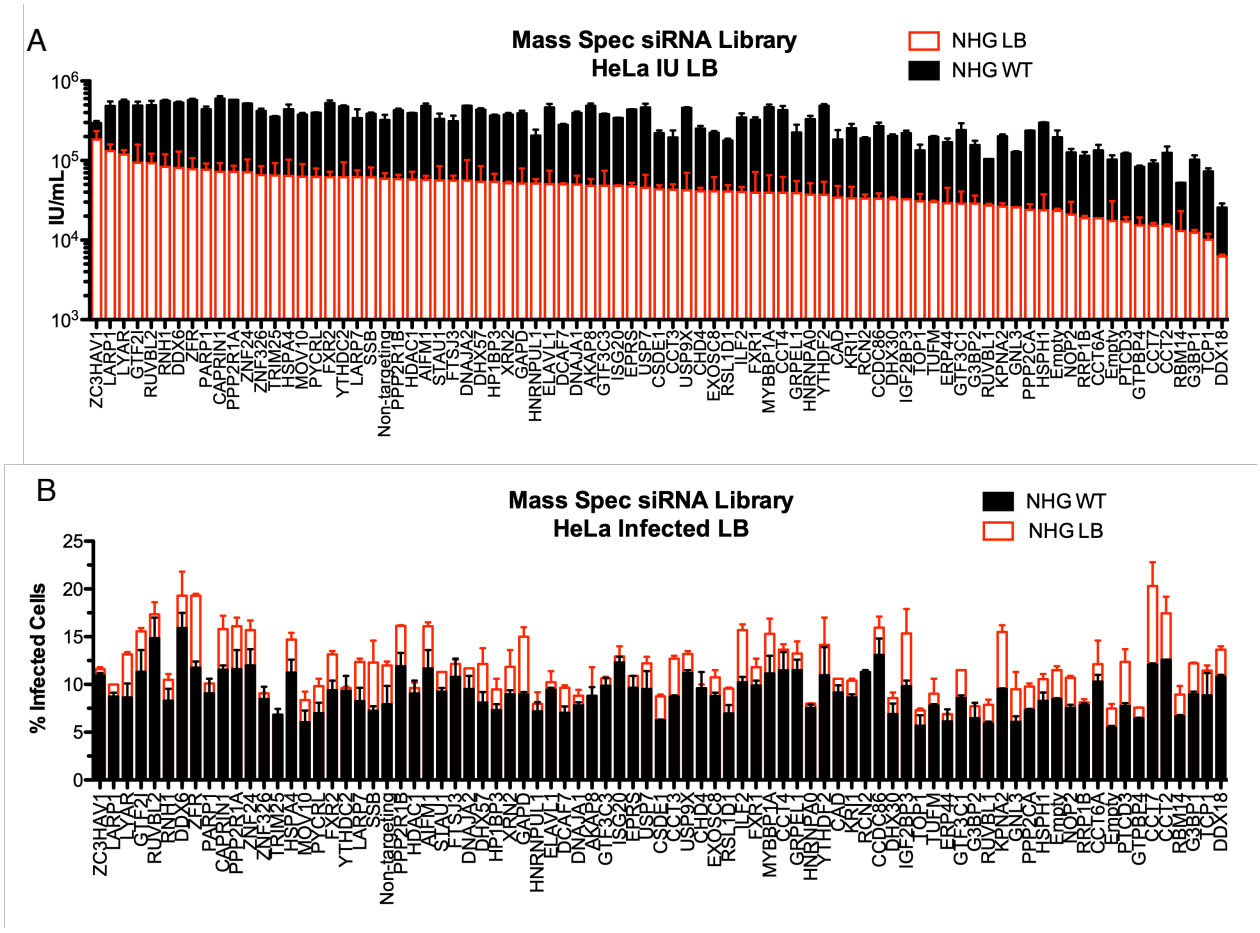


Figure 6.11: siRNA screen from Mass Spectrometry hits in HeLa cells

(A) Single-cycle infectious HIV-1 wild-type and mutant LB yield from siRNA-transfected HeLa cells (mean±sd. (n=3)). (B) Percentage of HeLa cells transfected with siRNA and infected with HIV-1 wild-type or mutant LB.

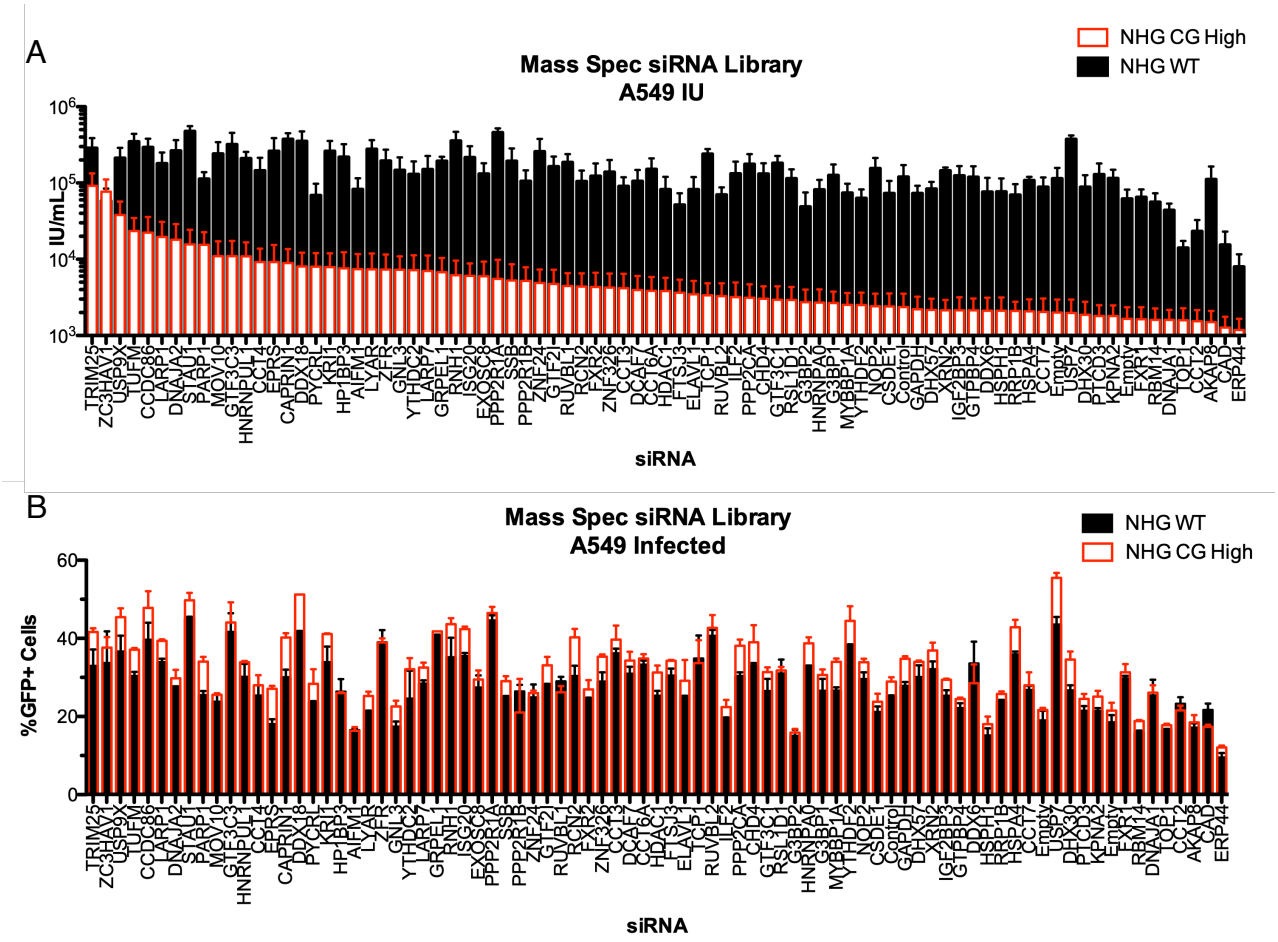


Figure 6.12: siRNA screen from Mass Spectrometry hits in A549 cells

(A) Single-cycle infectious HIV-1 wild-type and CG-High yield from siRNA-transfected A549 cells (mean±sd. (n=3)). **(B)** Percentage of A549 cells transfected with siRNA and infected with HIV-1 wild-type or CG-High.

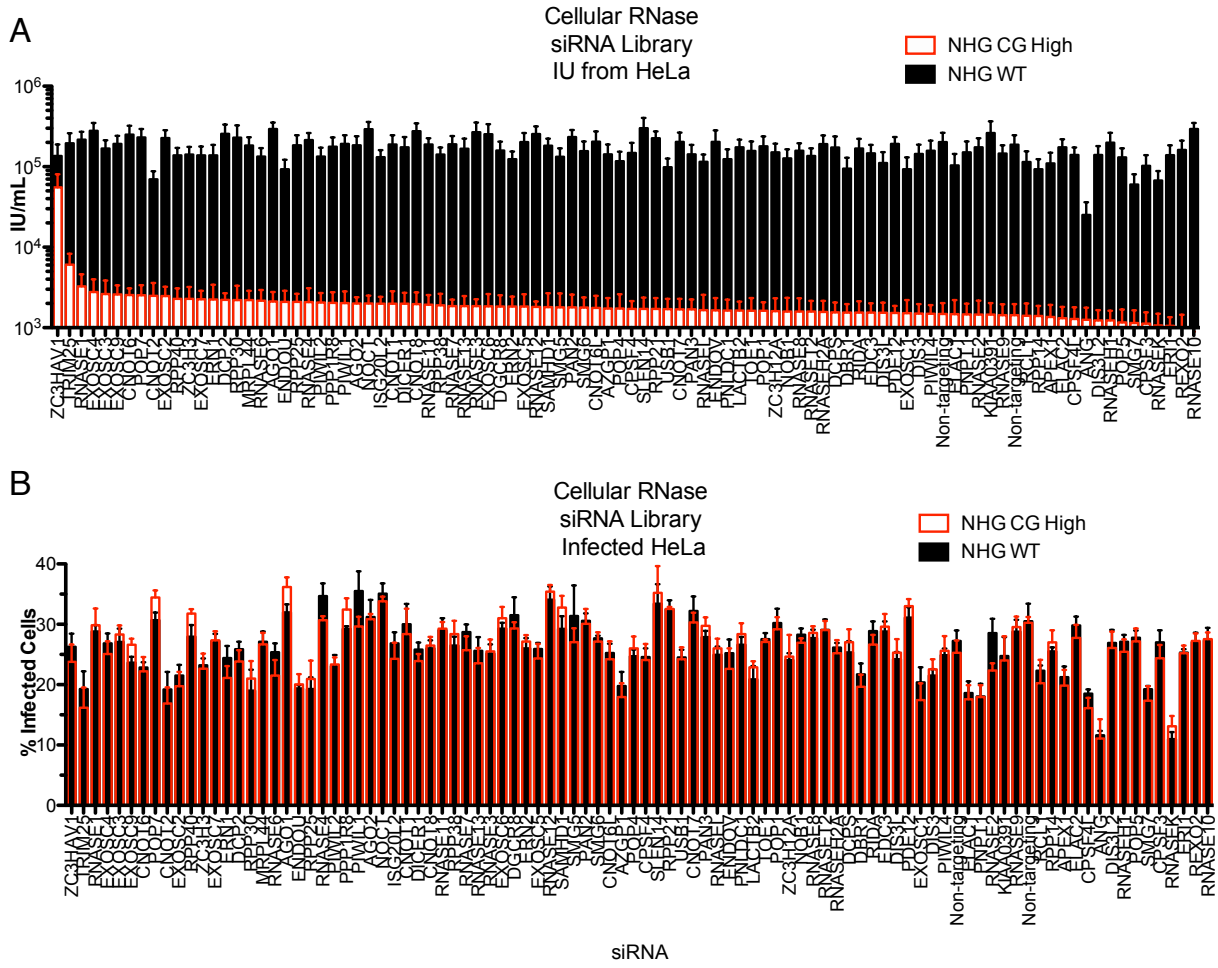


Figure 6.13: siRNA screen against cellular Ribonucleases in HeLa cells

(A) Single-cycle infectious HIV-1 wild-type and CG-High yield from siRNA-transfected HeLa cells (mean±sd. (n=3)). **(B)** Percentage of HeLa cells transfected with siRNA and infected with HIV-1 wild-type or CG-High.

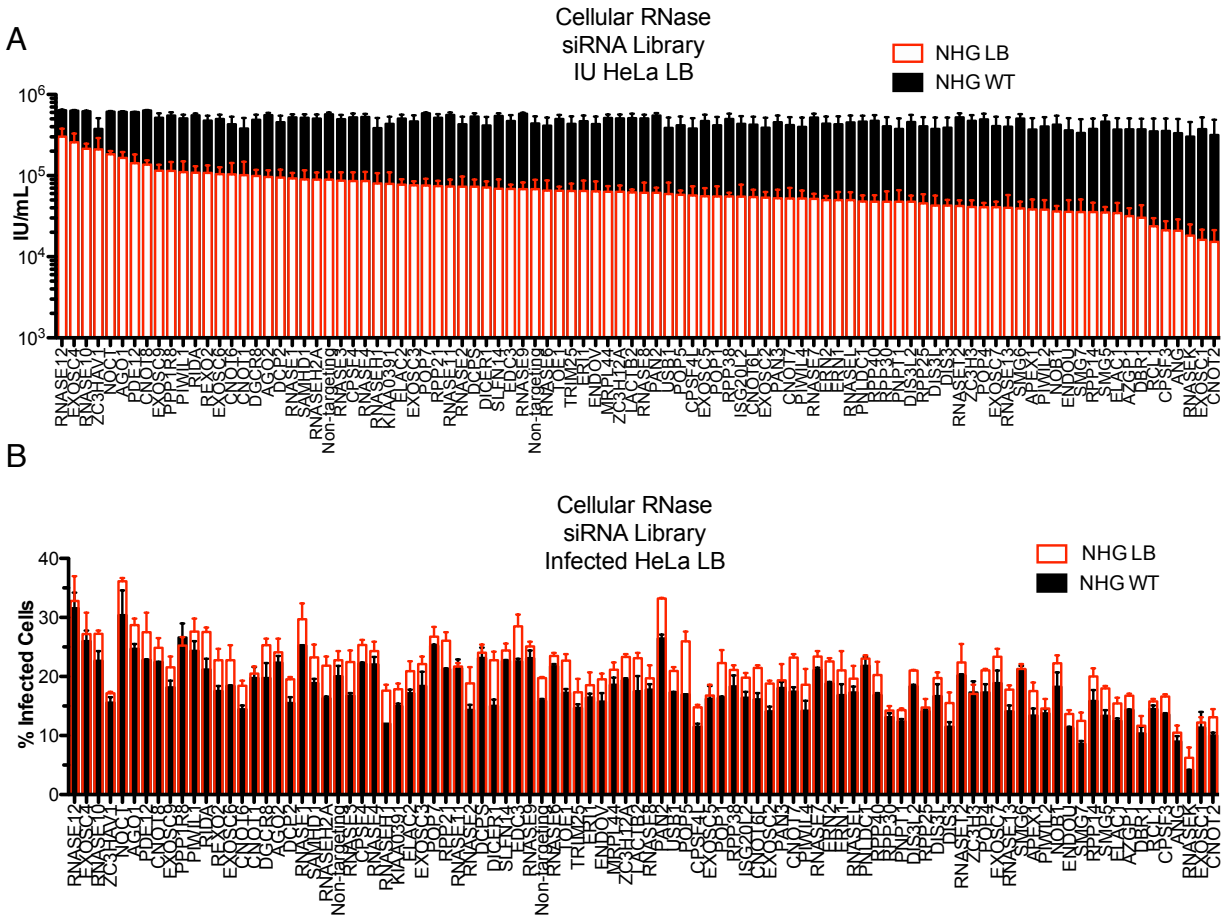


Figure 6.14: siRNA screen against cellular Ribonucleases in HeLa cells

(A) Single-cycle infectious HIV-1 wild-type or mutant LB yield from siRNA-transfected HeLa cells (mean±sd. (n=3)). **(B)** Percentage of HeLa cells transfected with siRNA and infected with HIV-1 wild-type or mutant LB.

We have previously shown that ZAP causes a cytoplasmic depletion of CG-high viral RNA. We hypothesized that this depletion is due to a cellular ribonuclease acting downstream of ZAP binding to CG-high viral RNA. For this reason, we assembled an siRNA library targeting all known cellular proteins catalogued as having a ribonuclease activity based on Uniprot's gene ontologies. The only ribonucleases that were found in Uniprot but excluded from our library on were those catalogued as having a specific activity against tRNAs. This library contained positive control siRNAs targeting ZAP and TRIM25, one negative control against GAPDH, one negative control that was non-targeting, and 92 siRNAs against annotated ribonucleases. We tested this library in wild-type HeLa cells and had a 100-fold dynamic range with our negative control siRNAs. The positive controls against ZAP and TRIM25 significantly rescued the replication of the CG-high virus, while no other siRNAs tested generated a significant rescue (Figure 6.13). As with our previous library, we also used this library in HeLa cells with the less CG-enriched and therefore less attenuated mutant LB. Surprisingly, ZAP did not come up as the top hit in this screen, but rather RNASE12, a poorly characterized member of the Ribonuclease A family that is annotated as inactive (Figure 6.14). Notably, analysis of the infected HeLa cells showed that knockdown of RNASE12 significantly increase the percentage of infected cells, possibly explaining why it also seemingly increased the replication of mutant LB. Another gene that also increased the replication of the LB virus was EXOSC4, a member of the RNA exosome that has previously been implicated as a cofactor for ZAP. EXOSC4 was

also near the top hits when screened against the fully attenuated CG-high virus, but the rescue was not statistically significant.

Summary

We have demonstrated that TRIM25 is an important cofactor for ZAP in the restriction of CG-high HIV-1. Mechanistically, it appears that in order to act as a cofactor for ZAP the RNA binding and ubiquitination activity of ZAP are non-essential. Additionally, ZAP is able to recognize and directly bind to CG-high RNA in the absence of TRIM25. Between the cell lines tested there are dramatic differences in how potent ZAP is against a CG-high HIV-1, but whether this is due to an additional cofactor required for activity or merely reflects the total number of ZAP molecules present in the cytoplasm remains unclear.

Chapter 7. Discussion

Splicing

Through a global synonymous mutagenesis experiment we found that extensive portions of the HIV-1 genome could be synonymously mutated without major effects on viral replication (Group 1 mutants). In particular, synonymous mutations throughout the majority of the *pol* gene had a minimal or no effect on viral fitness, such that their effect was not measurable in our assays. Given the extent of the mutations that were introduced, and the improbability that most RNA secondary structures would be preserved in our mutants, it seems unlikely that undiscovered specific RNA secondary structures essential for replication exist in portions of the viral genome covered by Group 1 mutants. Even among mutants that were replication defective, Group 2 mutants could be restored to replication competence through single nucleotide reversion mutations that suppressed the utilization of cryptic or canonical splice sites, whose use was enabled or elevated in the global mutagenesis experiment. Again, this argues against the notion that undiscovered, specific RNA structures that are essential for replication are prevalent in Group 2 mutants, with the exception of those that regulate splicing.

An important caveat to this conclusion is that replication was measured in permissive cell lines in the absence of competition. It is possible, even likely, that mutants or revertants with WT or near WT replication dynamics, have modest fitness deficits that would be evident in a more stringent environment or a competitive replication assay. For example, some of the HIV-1 accessory genes are important *in vivo*, but non-essential *in vitro*,

therefore defects in their expression would be expected to have minimal effects on replication in our assays. Thus, while we can reasonably conclude that RNA secondary structures in Group 1 and Group 2 revertant mutants are non-essential for replication *per se*, it is nevertheless possible that RNA structures play an accessory role in regulating or fine-tuning the levels or fates of mRNAs encoding certain accessory proteins.

Nevertheless, synonymous mutations in some portions of the HIV-1 genome caused profound, near-lethal defects in these highly permissive T-cell lines (Group 2 and Group 3 mutants). These mutations therefore perturbed essential, non-coding features of the HIV-1 nucleotide sequence. One noncoding feature of the HIV-1 genome that appeared important for replication that was uncovered by our Group 2a mutants was suppression of cryptic splice sites near the 5' end of the RNA genome. For these mutants, the magnitude of the deviation from WT sequence appeared important for cryptic site activation. Simply subdividing the mutant sequence blocks into two mutant blocks approximately equal lengths (e.g. mutants AA, AB and BA, BB) reverted the mutant splicing phenotype. This finding suggests that activation of the existing cryptic splice sites resulted from multiple perturbations to the *gag* nucleotide sequence and that changes in predicted splice site strength were not sufficient to explain their activation. Moreover, in the case of mutant A, a single nucleotide revertant mutation that occurred distal to the activated cryptic splice sites corrected the splicing defect without affecting their MaxEnt scores. These findings might be best explained by the existence of multiple elements in the *gag* gene (secondary structures or protein binding sites) that act redundantly to suppress cryptic splice site utilization. The fact that the revertant mutation for mutant A

occurred at a position proximal to an existing RNA structure that includes D1 (Berkhout and van Wamel, 2000; Keane et al., 2015), may suggest a role for an extended secondary structure involving the 5' leader and the *gag* gene in suppressing the utilization of potential cryptic splice sites in *gag*. The potential for the 5' leader to form alternative structures that could affect splicing appears to be finely tuned (Keane et al., 2015), and therefore could possibly be perturbed by distal mutations in *gag*. Clearly, further work will be required to understand how the WT noncoding RNA sequence is selected to avoid utilization of cryptic splice sites.

A detailed analysis of Group 2b mutants, that targeted the central portion of the genome, revealed several elements whose perturbation resulted in dramatic overuse of the canonical splice acceptors A1, A2, and A3. These mutant elements that resulted in oversplicing could be mapped to sequences of various lengths. Such sequences constitute candidate individual or clustered ESS elements. When combined with existing knowledge of splicing regulation, at A1, A2 and A3 (Madsen and Stoltzfus, 2006; Martin Stoltzfus, 2009), these and previous findings suggest a highly complex regulatory network of functional inputs that govern alternative splicing of the HIV-1 genome, as depicted in (Figure 7.1).

The phenotypes of some mutant and revertant viruses (i.e. viruses with perturbed splicing whose replication was recovered by mutations at splice sites) is consistent with the notion that exon definition (i.e. coordinated recognition of a 5' splice acceptor and a 3' splice donor by the splicing machinery (Kharytonchyk et al., 2016) plays a central role in the regulation of HIV-1 alternative splicing. For example, the mutant (IA) that exhibits elevated

use of A1 and the downstream donor D2 (as well as the cryptic donor D2b) while the revertant (IA(T4904A)) occurred at A1 and reduced utilization of D2, and D2b as well as A1. Similarly, in IB, that exhibits elevated use of both A1 and A2, both downstream donors D3 and D2 are also overused. In this case, the revertant mutation (IB(G4912A)) at A1 caused abolition of splicing at both A1 and D2 while splicing at A2 and D3 remain elevated. Consistent with the notion that splicing at A1 and D2 are coordinated, previous work has shown that the efficiency with which D2 is recognized can affect the frequency of splicing at A1(De Conti 2013). Similar communication occurs between A2 and D3. Indeed, for a mutant (JA) that exhibited profound oversplicing at A2 and D3, a reversion mutation (JA(G5463A)) that inactivated D3 also caused abolition of splicing at A2. Because not all splices at A2 lead to splicing at D3, this observation suggests that recognition of D3 by the splicing machinery is required for splicing at A2 even when D3 is not utilized, as has previously been suggested (Erkelenz et al., 2013; Martin Stoltzfus, 2009). Overall therefore, a major determinant of the use of a given acceptor or donor in the HIV-1 genome was the use of a downstream donor or upstream acceptor, respectively.

Our data is also consistent with the notion that splice acceptor sites compete with each other, with 'cascading' effects based on exon definition. For example, the mutant (JA) under-utilizes A1 and over-utilizes A2. Its revertant (JA(G54653A)) that abolishes splicing at both D3 and A2, causes over-utilization of A1. In a reciprocal example of A1-A2 competition, abolition of A1 utilization in the revertant IB(G4912A), was accompanied by increased splicing at A2. Acceptor competition was also evident in the context of the mutant K, which exhibited oversplicing at A3 at the expense of splicing at A1, A2, A4b

and A5. In this case, the reversion mutation at A3 in K(C5774T) led to restoration of WT splicing frequency at all other sites in the central portion of the HIV-1 genome.

Overall therefore, a key regulator of the use of a particular splice site is the presence and utilization of other splice sites, through coordinated recognition of acceptor and donor sites, along with competition between acceptor sites. Thus, disruption of the splicing regulatory signals whose existence is indicated herein can have complex effects on overall splicing, through the propagation of their effects from one splice site to another. Clearly the overall effect of these perturbations is best appreciated in the nextgen sequencing assay with all splice sites represented in a viral construct.

That being the case, the RNA sequence elements that we have identified as apparent regulators of splice acceptor utilization (Figure 7.1) could work directly or indirectly. Specifically, they could act to directly inhibit access of the splicing machinery to the affected splice acceptors or indirectly through splice donors, inhibiting acceptor utilization by inhibiting exon definition. It is also possible that our mutations affect the ability of splice sites to effectively communicate with each other in the context of exon definition. Existing splicing regulatory sequences have been reported to exert their effect through binding of hnRNP or SR proteins, or through the formation of RNA secondary structures (Mandal 2009). The sequences that we have identified are of varying sizes; some elements that had major effects on splicing appear sufficiently small (e.g. JC2 and KD2) to constitute specific protein binding sites. However, these elements did not appear to be enriched in canonical hnRNP binding motifs, as might be expected for splicing silencers (Mandal et al., 2009). Some of the effects on splicing that we found in our mutants, particularly within

the I and J fragments, appeared complex and not easily mapped to small discrete elements. Perhaps these effects are the result of combinatorial inputs from multiple binding sites or secondary structures that could act to occlude splice sites, or spatially separate 3' donors from 5' acceptors thereby inhibiting exon definition. Further work will be required to determine precisely how these elements inhibit HIV-1 splicing.

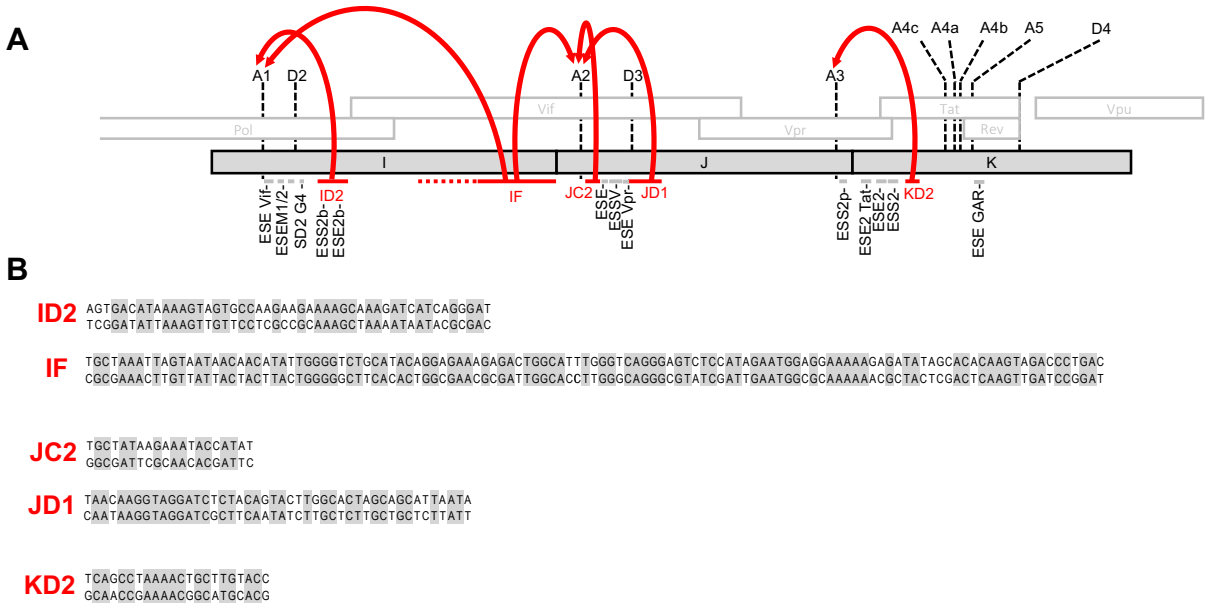


Figure 7.1: Summary of splicing control in HIV-1

(A) Schematic representation of the central portion of the HIV-1 genome with the positions of canonical splice sites indicated. Previously identified splicing control elements are indicated with grey lines. Sequences identified in this study whose mutation enhanced splicing are indicated with red lines and the splice acceptors on which they act are indicated with red arrows. (B) The sequences of the identified elements that affect splicing (upper line = WT sequence, lower line = oversplicing mutant sequence) are shown.

Perturbation of balanced splicing did not always lead to abolition of HIV-1 replication. For example, the mutants JB and JA(5463A) had perturbed splicing (overuse of A1 for JA, and abolition of SX2 utilization for JB) but their replication was only slightly delayed compared to WT. Similarly the IB(G4912A) revertant virus had near WT replication but completely lacked splicing to A1 and therefore abolished inclusion of SX1. These perturbations would be expected to lead to underexpression or overexpression of Vif and Vpr, neither of which are essential for replication in the particular cell type used in our experiments. Thus, the requirement for a particular balance of HIV-1 mRNAs could be highly context dependent. In our experiments, replication defects that resulted from over-splicing were likely the result of depletion of the pool of unspliced RNA, thus leading to lower levels of synthesis of Gag, Pro, Pol, and other viral proteins, and lower levels of viral genomes for packaging. Thus, a key role of splicing inhibitory signals in the HIV-1 genome is to maintain the unspliced RNA pool, as well as adequate levels of necessary viral proteins.

Overall, our global synonymous mutagenesis experiment has revealed several RNA elements whose native sequence is important for HIV-1 replication. In particular, we have identified several RNA elements in the HIV-1 genome whose native sequence appears to be important for suppression of canonical splice sites, regulation of alternative splicing and maintenance of unspliced transcript levels. Additionally, our analysis revealed that some as yet unidentified property of RNA sequences in the *gag* gene suppresses utilization of cryptic splice sites. Understanding how RNA sequence affects splicing in the

context of HIV-1 may give insights into the general mechanisms by which alternative splicing is regulated and how splicing regulation evolve.

ZAP recognition of CG dinucleotides

The synonymous mutagenesis also led to the discovery that ZAP binds CG dinucleotides in RNA. This discovery brings together two previously separate areas of RNA biology: (i.) the outstanding question as to how ZAP recognizes non-self RNA and (ii.) the attenuation of RNA viruses through the addition of CG dinucleotides. While many groups had previously tried to identify an ZAP RNA target sequence through either analysis of sensitive viruses or iterative cloning to find a minimally sensitive element, no common features emerged. Analysis of the viruses previously determined to be sensitive and insensitive to ZAP reveal a clear pattern that differentiates them when comparing the CG dinucleotides odds ratios (Figure 7.2) (Bick et al., 2003; Cheng et al., 2013; Lin et al., 2014; Mao et al., 2013; Muller et al., 2007). Strikingly, the genomes of the viruses that are sensitive to ZAP have a CG dinucleotide odds ratio (the ratio of the observed frequency to the expected frequency based on the mononucleotide composition) of at least 0.54 or higher. The genomes of the viruses insensitive to ZAP have an odds ratio of 0.54 or lower. This analysis, though limited, suggests that the threshold for a ZAP sensitive virus might be somewhere in the range of a 0.54 odds ratio, though a number of additional factors likely also contribute to this sensitivity, including but not limited to: viral proteins that antagonize ZAP activity through re-localization or degradation, additional RNA compositional differences of viral genomes, and mRNA sequestration that

prevents ZAP from gaining access to viral mRNA. Further analysis of the regions of the SINV genome that were used to determine a minimally sensitive ZAP sequence also indicate that ZAP is targeting CG-high RNA (Guo et al., 2004). There is a strong and significant correlation between the odds ratio of CG dinucleotides multiplied by the length of the sequence and the fold inhibition by ZAP. This finding explains why has previously difficult to find a minimally sensitive ZAP fragment, not only did the RNA sequence need to have a certain threshold of CG dinucleotides, but it also needed to be long enough to efficiently recruit ZAP for repression.

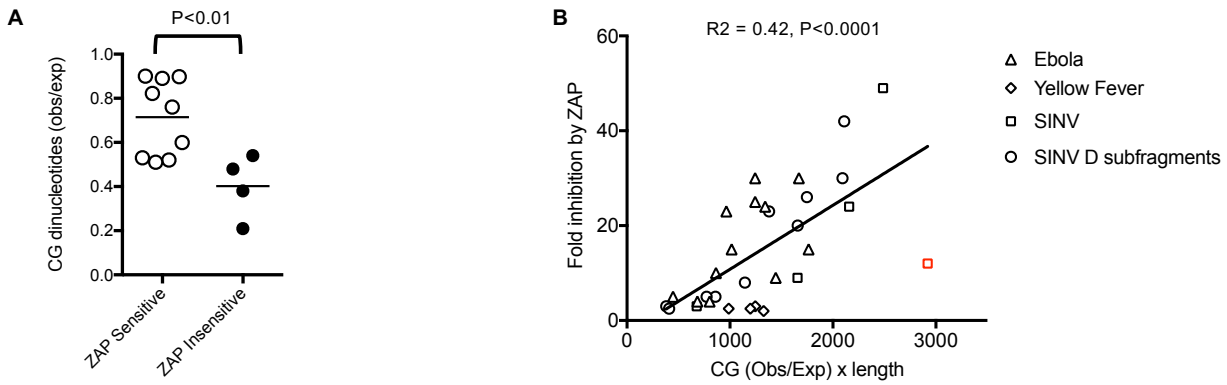


Figure 7.2: Analysis of CG suppression in previously reported ZAP-sensitive and ZAP-resistant viruses and ZAP-sensitizing elements

(A) CG suppression in RNA and reverse transcribing viruses previously reported to be ZAP sensitive (*n* = 9, open symbols) and ZAP resistant (*n* = 4, filled symbols)[7,17–20]. The viruses included in the analysis and their degrees of CG suppression (CG observed/expected) are: ZAP-sensitive: Sinbis virus (0.90), Semliki forest virus (0.89), Venezuelan equine encephalitis virus (0.76), ebolavirus (0.60), hepatitis B virus (0.52), Moloney murine leukemia virus (0.51), Marburg virus (0.53), alphavirus M1 (0.89), Ross river virus (0.82); ZAP-insensitive: HIV-1 (0.21), yellow fever virus (0.38), vesicular stomatitis virus (0.48), poliovirus (0.54). The *P* value was calculated using Student's *t*-test (two-sided, *n* = 9 ZAP-sensitive viruses and *n* = 4 ZAP-resistant viruses). Influenza virus (CG obs/exp = 0.44), which has been reported to be ZAP-resistant owing to the presence of an antagonist and ZAP-L- sensitive via an entirely distinct protein interaction-based mechanism, was excluded from this analysis. **(B)** Previous analysis by Guo et al on ZAP inhibition of reporter gene expression (Guo et al., 2004). Each RNA element derived from the indicated RNA viruses was placed in a 3′ UTR of a luciferase reporter plasmid and fold inhibition by coexpressed ZAP is plotted against the product of CG suppression (CG observed/expected) and length for each RNA element. A data point that is a quantitative outlier from the general trend (indicated in red) is from the Sinbis (SINV) genome, but is nevertheless included in the linear regression analysis. *P* value was calculated using the *F*-test (two-sided, *n* = 32 data points).

While we have demonstrated that ZAP binds to CG rich RNA and results in the cytoplasmic depletion of that RNA, additional analysis of the data presented suggests that a complete understanding of how ZAP targets CG-high RNA is more complex. One example of this is the normal phenotype of Group 1 mutants from the original synonymous mutagenesis. Again, all of the mutants in this group are phenotypically indistinguishable from the wild-type virus in a spreading replication assay in either MT4 cells or CEM cells (Figure 3.2, 3.3). Curiously, *in silico* analysis of their CG enrichment reveals that while they do not have the same total number of CG dinucleotides as mutants L and M, the mutagenesis did significantly increase the number of CG dinucleotides. It was only after combining all of the mutations from mutants E through H, thus creating mutants EH that these CG dinucleotides rendered HIV-1 ZAP sensitive (Figure 5.1E; Figure 5.4 A, C, D). Despite having more than twice as many CG dinucleotides as mutant L, mutant EH is less sensitive to ZAP than mutant L when comparing decreases in levels of unspliced RNA, $Pr55^{Gag}$ expression and fitness in a spreading replication assay (Figure 5.3 B, D; Figure 5.4 C, D). This could be explained by a multitude of factors. One key difference between EH and L is the CG-high region is present in different sections of the coding sequence. For mutant EH, the CG dinucleotides are in the incorporated into the coding sequencing of the gag mRNA, while for mutant L they are in the 3'UTR of the gag mRNA. Another difference might be in the distribution of CG dinucleotides. Mutant L has a higher density of dinucleotides compared to EH, even though it has fewer total. Without further experimentation, we cannot conclude which of factors contribute to the ZAP sensitivity of

a viral mRNA, but we can conclude that there are more aspects of CG-enrichment that determine ZAP mediated repression than we currently understand.

Another observation that suggests ZAP targeting of mRNA is more complex comes from a deeper examination of mutant L expression of the Gag and Env proteins. Comparing the protein expression of these two different proteins during an infection of MT4 cells with mutant L, the Envelope protein is significantly more affected by ZAP compared to Gag. The CG rich region in the coding sequence of env, while it is the 3'UTR of gag, suggesting that ZAP is perhaps more active against coding regions that 3'UTRs. This is further supported by CLIP-seq data that indicates ZAP is more frequently bound to coding regions of mRNA than to 3'UTRs, though this might be explained by the fact that 3'UTR are slightly more suppressed for CG dinucleotides than coding regions (Figure 5.10C). Because ZAP probably destabilizes the mRNA that it binds to, it is possible that the CLIP data is skewed toward mRNA that is less targeted for destabilization. Additionally, the gag and env mRNA are translated in different locations, where env transcripts are translated in the rough ER while the gag transcripts are translated on free ribosomes in the cytoplasm. While microscopy of ZAP indicates that it is expressed in the cytoplasm with a punctate pattern, it is possible that ZAP somehow more efficiently targets mRNA in the rough ER than on free ribosomes. Alternately, it remains possible that ZAP is more active in targeting CG-rich mRNA when the CGs are in coding regions compared to 3'UTRs.

Further evidence for this distinction between CG dinucleotides effecting open reading frames and 3' UTRs comes from the transient transfection assay. These assays used

different sequences as 3'UTR to a CG-depleted luciferase reporter (Figure 5.9). CG-enriched viral sequences from IAV and VSV were used as 3'UTRs to luciferase and transfected into ZAP KO and wild-type HeLa cells. Despite the use of long sequences with CG-enrichment similar to that of mutant L, the expression of the CG high constructs is reduced by less than 10-fold in the presence of ZAP, still orders of magnitude above the limit of detection. This is in contrast to the expression of gp160, where protein expression is hardly detectable. It is possible that the limited restriction of these sequences might be due to compositional differences in either mononucleotides or dinucleotides as well as stability of the encoded protein. Though, it remains possible that these effects are due to the location of the CG dinucleotides in the mRNA, namely in the 3'UTR. Previous studies have observed up to 40-fold restriction of similar constructs with SINV virus genes as the 3'UTR (Guo et al., 2004). In contrast, we have only observed a 10-fold restriction of our luciferase constructs with CG-rich sequences as 3'UTR (Figure 5.9D). One key differentiator between these two experimental designs is that our luciferase gene has been recoded so that it is depleted for CG dinucleotides while the luciferase used in Guo et al. is high in CG dinucleotides, likely contributing to the differences in ZAP sensitivity (Figure 5.9A). In fact, it has been observed that removal of CG dinucleotides from the coding sequence of luciferase increases the expression of luciferase by nearly 10-fold (Atkinson et al., 2014). Taken together, it remains to be determined how the location of CG dinucleotides within an mRNA molecule might sensitize it to ZAP mediated restriction.

ZAP cofactor requirements

Multiple studies have been published to suggest that they have determined a ZAP cofactor (Chen et al., 2008; Erazo and Goff, 2015; Guo et al., 2007), yet to date the only cofactor that has been substantiated by multiple studies is TRIM25 (Li et al., 2017; Zheng et al., 2017). The initial study to describe TRIM25 as a cofactor for ZAP suggested that all of the domains in TRIM25 are important for its function, and specially investigated the ubiquitination of ZAP by TRIM25 (Li et al., 2017). Yet, surprisingly, ablation of the resides in ZAP that are ubiquitinated does not affect its antiviral activity, potentially indicating that ZAP is not the functional target of the ubiqutination activity of TRIM25 (Li et al., 2017). Other groups have described a putative RNA binding domain in TRIM25 that is important for its biological function both in relation to ZAP and to RIG-I (Choudhury et al., 2017; Sanchez et al., 2018). Until this point, the precise mechanisms by which TRIM25 supports the antiviral activity of ZAP is not well understood. Our inability to crosslink any RNA with TRIM25 in a mammalian cell is perplexing considering multiple other groups have claimed that it binds RNA. One possible explanation for this might be purely technical. We are using the modified nucleotide 4SU to more efficiently crosslink our RNA to TRIM25, as we do for all CLIP-seq experiments. If TRIM25 has absolutely no affinity for uridine nucleotides, and therefore is never proximal enough to a 4SU in the RNA, then crosslinking may be inefficient. However, we have done CLIP-seq experiments for multiple RNA binding proteins that do not bind U and always observe a good RNA crosslinking signal (Kutluay et al., 2014; York et al., 2016). Therefore, this possibility seems unlikely, but cannot be entirely discounted. This could also be due to the

differences between biochemical assays with two components (RNA and protein) and the rich heterogeneous mixture of polymers in a mammalian cell. Additionally, TRIM25 might have cellular cofactor that occludes the binding site from interacting with RNA, and this cofactor might be expressed differently based on the cell line used, explaining how some groups have pulled down RNA with TRIM25 (Choudhury et al., 2017).

Regardless of the reason we were unable to pulldown RNA with TRIM25 in 293T cells, it is abundantly clear that ZAP does not require TRIM25 to recognize CG dinucleotides in RNA. Though we have shown that cell type differences have a profound impact on how ZAP and TRIM25 work together, we have done the CLIP-seq studies using ZAP in the absence of TRIM25 in a cell line that requires TRIM25 for proper function. This indicates that even though ZAP requires TRIM25 to restrict CG high viruses in 293T cells, TRIM25 does not change RNA binding preference of ZAP, allowing us to confidently conclude that TRIM25 enhances the antiviral activity of ZAP though some other mechanism. Similarly, cotransfection experiments that reconstitute 293T double knockout cells with ZAP and a TRIM25 that has mutations in the putative RNA binding domain further suggest that RNA binding is not an important biological activity with regards to supporting the antiviral activity of ZAP. Further mutational analysis and reconstitution of TRIM25 in double knock out 293T suggests that ubiquitination is dispensable for function, while the formation of higher order multimers appears to be critical to the function as a cofactor for ZAP.

While in 293T cells we have provided evidence that higher order multimerization is critical to the function of TRIM25, this finding must be taken in light of the evidence that also suggests a cell type dependence on the relationship between TRIM25 and ZAP. The

profound differences we observe between T-cell lines and adherent cell lines (with the exception of HeLa cells) are thus far unexplained. Along with these cell type differences, we do not observe a correlation between ZAP activity and TRIM25 protein expression. One explanation might be that there is an additional positively acting cofactor that is constitutively expressed in T-cells and HeLa cells, but IFN induced in other adherent cells such as A549 cells. It is also possible that there might be a negative regulator of ZAP and TRIM25 present in adherent cells, which is then downregulated by IFN in A549 cells. Historically, restriction factors are mostly positively acting and upregulated, though there are a significant number of genes are downregulated upon IFN treatment, so this possibility cannot be entirely ruled out (Kane et al., 2016; Schoggins et al., 2014).

The failure of previously identified ZAP cofactors to be validated is likely do the how they were identified (Chen et al., 2008; Erazo and Goff, 2015). Often, cellular ribonucleases are overexpressed and determined to have a role in RNA degradation (Chen et al., 2008; Guo et al., 2007; Zhu et al., 2012). Preferably, experiments that knockdown or knockout genes of interest and demonstrate that ZAP is inactive without these genes would be more enlightening. Additionally, these overexpression experiments were never done in the context of a ZAP knockdown or knockout, and therefore not conclusively linked to the ribonuclease activity in the ZAP pathway. Consistent with this, we have not corroborated any of the data that suggests additional factors such as XRN1, DDX17, DIS3 or PARN are not required for the antiviral activity of ZAP. One caveat to this conclusion is that our experiments were done in the context of an siRNA knockdown, which means the protein targets are not completely removed from the cell. If only a very small amount of residual

protein is needed to complement the activity of ZAP, as is the case with LEDGEF and IN (Llano et al., 2006), then our experimental set up would not identify the protein as a cofactor for ZAP. To confidently identify a cofactor would involve overexpressing the proposed cofactor in a cell line that is semi-permissive to the replication of a CG-high virus, resulting in the a greater attenuation of the CG-high virus. Additionally, the same experiment should be performed in a ZAP knockout cell line to confirm that the effects are ZAP dependent and overexpression in the absence of ZAP has no effect on the replication of the CG-high virus. An additional caveat to our approach in primarily focusing on gene knockdown is that there might be multiple ribonucleases that can act as a ZAP cofactor and degrade the CG-high RNA. If multiple proteins can act as the ribonuclease for ZAP, then knocking down only one of them is likely to have a small or no effect. If this is the case, identifying the cofactor might best be carried out using overexpression experiments, but a ZAP knockout would be an important control to confirm specificity.

In this present study, we have identified many new cis-acting RNA elements that control the proper regulation of splicing in HIV-1. In addition, we have explained a primary function for the CG suppression of the HIV-1 genome, namely avoiding detection by ZAP. We have shown that ZAP binds to CG-high RNA, and this binding destabilizes the RNA, but how this destabilization is accomplished remains unclear. We have demonstrated that ZAP recognizes CG-high RNA independently of TRIM25, and that the multimerization activity of TRIM25 is important for its function as a ZAP cofactor. Much of our data indicates that an additional ZAP cofactor is yet undiscovered and heterogeneously expressed among cell lines. Overall, this global synonymous mutagenesis has yielded

many new insights into the non-coding functions within the coding sequence of HIV-1 RNA, and opens the possibility that more elements remain to be discovered.

References

Akiyama, H., Miller, C.M., Ettinger, C.R., Belkina, A.C., Snyder-Cappione, J.E., and Gummuluru, S. (2018). HIV-1 intron-containing RNA expression induces innate immune activation and T cell dysfunction. Nature communications *9*, 3450.

Arhel, N.J., Souquere-Besse, S., Munier, S., Souque, P., Guadagnini, S., Rutherford, S., Prevost, M.C., Allen, T.D., and Charneau, P. (2007). HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. EMBO J *26*, 3025-3037.

Arimoto, K., Takahashi, H., Hishiki, T., Konishi, H., Fujita, T., and Shimotohno, K. (2007). Negative regulation of the RIG-I signaling by the ubiquitin ligase RNF125. Proceedings of the National Academy of Sciences of the United States of America *104*, 7500-7505.

Atkinson, N.J., Witteveldt, J., Evans, D.J., and Simmonds, P. (2014). The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. Nucleic acids research *42*, 4527-4545.

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature *226*, 1209-1211.

Barral, P.M., Sarkar, D., Fisher, P.B., and Racaniello, V.R. (2009). RIG-I is cleaved during picornavirus infection. Virology *391*, 171-176.

Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D., and Williamson, J.R. (1996). α Helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. Science.

Berger, E.A., Murphy, P.M., and Farber, J.M. (1999). Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annual review of immunology *17*, 657-700.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell *89*, 781-787.

Berkhout, B., Silverman, R.H., and Jeang, K.T. (1989). Tat Trans-Activates the Human Immunodeficiency Virus through a Nascent Rna Target. Cell *59*, 273-282.

Berkhout, B., and van Wamel, J.L. (2000). The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. RNA *6*, 282-295.

Bick, M.J., Carroll, J.W., Gao, G., Goff, S.P., Rice, C.M., and MacDonald, M.R. (2003). Expression of the zinc-finger antiviral protein inhibits alphavirus replication. Journal of virology *77*, 11555-11562.

Bieniasz, P.D. (2009). The Cell Biology of HIV-1 Virion Genesis. In Cell Host and Microbe.

Blanco-Melo, D., Venkatesh, S., and Bieniasz, P.D. (2012). Intrinsic cellular defenses against human immunodeficiency viruses. Immunity *37*, 399-411.

Bour, S., and Strebel, K. (2003). The HIV-1 Vpu protein: A multifunctional enhancer of viral particle release. In Microbes and Infection.

Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science *319*, 921-926.

Brillen, A.L., Walotka, L., Hillebrand, F., Muller, L., Widera, M., Theiss, S., and Schaal, H. (2017). Analysis of Competing HIV-1 Splice Donor Sites Uncovers a Tight Cluster of Splicing Regulatory Elements within Exon 2/2b. Journal of virology *91*.

Cadena, C., Ahmad, S., Xavier, A., Willemsen, J., Park, S., Park, J.W., Oh, S.W., Fujita, T., Hou, F., Binder, M.*, et al.* (2019). Ubiquitin-Dependent and -Independent Roles of E3 Ligase RIPLET in Innate Immunity. Cell.

Cai, X., Chen, J., Xu, H., Liu, S., Jiang, Q.X., Halfmann, R., and Chen, Z.J. (2014). Prion-like polymerization underlies signal transduction in antiviral immune defense and inflammasome activation. Cell *156*, 1207-1222.

Campbell, E.M., and Hope, T.J. (2015). HIV-1 capsid: The multifaceted key player in HIV-1 infection. In Nature Reviews Microbiology.

Campbell, S., and Vogt, V.M. (1997). In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles. Journal of virology *71*, 4425-4435.

Cardenas, W.B., Loo, Y.M., Gale, M., Jr., Hartman, A.L., Kimberlin, C.R., Martinez-Sobrido, L., Saphire, E.O., and Basler, C.F. (2006). Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. Journal of virology *80*, 5168-5178.

Carlton, J.G., and Martin-Serrano, J. (2009). The ESCRT machinery: new functions in viral and cellular biology. Biochem Soc Trans *37*, 195-199.

Chan, D.C., Fass, D., Berger, J.M., and Kim, P.S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. Cell *89*, 263-273.

Charneau, P., and Clavel, F. (1991). A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. Journal of virology *65*, 2415-2421.

Checkley, M.A., Luttge, B.G., and Freed, E.O. (2011). HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. In Journal of molecular biology.

Chen, G., Guo, X., Lv, F., Xu, Y., and Gao, G. (2008). p72 DEAD box RNA helicase is required for optimal function of the zinc-finger antiviral protein. Proceedings of the National Academy of Sciences of the United States of America *105*, 4352-4357.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. In Nature Reviews Molecular Cell Biology.

Chen, S., Xu, Y., Zhang, K., Wang, X., Sun, J., Gao, G., and Liu, Y. (2012). Structure of N-terminal domain of ZAP indicates how a zinc-finger protein recognizes complex RNA. Nat Struct Mol Biol *19*, 430-435.

Cheng, X., Virk, N., Chen, W., Ji, S., Ji, S., Sun, Y., and Wu, X. (2013). CpG usage in RNA viruses: data and hypotheses. PLoS One *8*, e74109.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem *278*, 372-381.

Chertova, E., Bess, J.W., Jr., Crise, B.J., Sowder, I.R., Schaden, T.M., Hilburn, J.M., Hoxie, J.A., Benveniste, R.E., Lifson, J.D., Henderson, L.E*., et al.* (2002). Envelope glycoprotein incorporation, not shedding of surface envelope glycoprotein (gp120/SU), Is the primary determinant of SU content of purified human immunodeficiency virus type 1 and simian immunodeficiency virus. Journal of virology *76*, 5315-5325.

Chiu, H.P., Chiu, H., Yang, C.F., Lee, Y.L., Chiu, F.L., Kuo, H.C., Lin, R.J., and Lin, Y.L. (2018). Inhibition of Japanese encephalitis virus infection by the host zinc-finger antiviral protein. PLoS pathogens *14*, e1007166.

Choudhury, N.R., Heikel, G., Trubitsyna, M., Kubik, P., Nowak, J.S., Webb, S., Granneman, S., Spanos, C., Rappsilber, J., Castello, A*., et al.* (2017). RNA-binding activity of TRIM25 is mediated by its PRY/SPRY domain and is required for ubiquitination. BMC Biol *15*, 105.

Chow, K.T., Gale, M., Jr., and Loo, Y.M. (2018). RIG-I and Other RNA Sensors in Antiviral Immunity. Annual review of immunology *36*, 667-694.

Cobian Guemes, A.G., Youle, M., Cantu, V.A., Felts, B., Nulton, J., and Rohwer, F. (2016). Viruses as Winners in the Game of Life. Annu Rev Virol *3*, 197-214.

Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., and Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. Science *320*, 1784-1787.

Corcoran, D.L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R.L., Keene, J.D., and Ohler, U. (2011). PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol *12*, R79.

Craigie, R., and Bushman, F.D. (2014). Host Factors in Retroviral Integration and the Selection of Integration Target Sites. Microbiol Spectr *2*, 1035-1050.

Daly, T.J., Cook, K.S., Gray, G.S., Maione, T.E., and Rusche, J.R. (1989). Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. Nature *342*, 816-819.

de Breyne, S., Soto-Rifo, R., Lopez-Lastra, M., and Ohlmann, T. (2013). Translation initiation is driven by different mechanisms on the HIV-1 and HIV-2 genomic RNAs. Virus Res *171*, 366-381.
Decroly, E., Ferron, F., Lescar, J., and Canard, B. (2012). Conventional and unconventional mechanisms for capping viral mRNA. In Nature Reviews Microbiology.

Devarkar, S.C., Wang, C., Miller, M.T., Ramanathan, A., Jiang, F., Khan, A.G., Patel, S.S., and Marcotrigiano, J. (2016). Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. Proceedings of the National Academy of Sciences of the United States of America *113*, 596-601.

Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M., and Skinner, M.A. (1990). HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. EMBO J *9*, 4145-4153.

Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M., Skinner, M.A., and Valerio, R. (1989). Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. Proceedings of the National Academy of Sciences of the United States of America *86*, 6925-6929.

Donelan, N.R., Basler, C.F., and Garcia-Sastre, A. (2003). A recombinant influenza A virus expressing an RNA-binding-defective NS1 protein induces high levels of beta interferon and is attenuated in mice. Journal of virology *77*, 13257-13266.

Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. In Nature Reviews Molecular Cell Biology.

Emery, A., Zhou, S., Pollom, E., and Swanstrom, R. (2017). Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. Journal of virology *91*.

Erazo, A., and Goff, S.P. (2015). Nuclear matrix protein Matrin 3 is a regulator of ZAP-mediated retroviral restriction. Retrovirology *12*, 57.

Erkelenz, S., Poschmann, G., Theiss, S., Stefanski, A., Hillebrand, F., Otte, M., Stuhler, K., and Schaal, H. (2013). Tra2-mediated recognition of HIV-1 5' splice site D3 as a key factor in the processing of vpr mRNA. Journal of virology *87*, 2721-2734.

Feng, Q., Langereis, M.A., Lork, M., Nguyen, M., Hato, S.V., Lanke, K., Emdad, L., Bhoopathi, P., Fisher, P.B., Lloyd, R.E.*, et al.* (2014). Enterovirus 2Apro targets MDA5 and MAVS in infected cells. Journal of virology *88*, 3369-3378.

Feng, S., and Holland, E.C. (1988). HIV-1 tat trans-activation requires the loop sequence within tar. Nature *334*, 165-167.

Fernandes, J.D., Faust, T.B., Strauli, N.B., Smith, C., Crosby, D.C., Nakamura, R.L., Hernandez, R.D., and Frankel, A.D. (2016). Functional Segregation of Overlapping Genes in HIV. Cell *167*, 1762-1773 e1712.

Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W., and Luhrmann, R. (1995). The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell *82*, 475-483.

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I.W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. Cell *90*, 1051-1060.

Franke, E.K., Yuan, H.E., and Luban, J. (1994). Specific incorporation of cyclophilin A into HIV-1 virions. Nature *372*, 359-362.

Freed, E.O. (2015). HIV-1 assembly, release and maturation. In Nature Reviews Microbiology. Futcher, B., Gorbatsevych, O., Shen, S.H., Stauft, C.B., Song, Y., Wang, B., Leatherwood, J., Gardin, J., Yurovsky, A., Mueller, S.*, et al.* (2015). Reply to Simmonds et al.: Codon pair and dinucleotide bias have not been functionally distinguished. Proceedings of the National Academy of Sciences of the United States of America *112*, E3635-3636.

Gack, M.U., Shin, Y.C., Joo, C.H., Urano, T., Liang, C., Sun, L., Takeuchi, O., Akira, S., Chen, Z., Inoue, S.*, et al.* (2007). TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. Nature *446*, 916-920.

Ganser, B.K., Li, S., Klishko, V.Y., Finch, J.T., and Sundquist, W.I. (1999). Assembly and analysis of conical models for the HIV-1 core. Science *283*, 80-83.

Ganser-Pornillos, B.K., Chandrasekaran, V., Pornillos, O., Sodroski, J.G., Sundquist, W.I., and Yeager, M. (2011). Hexagonal assembly of a restricting TRIM5alpha protein. Proceedings of the National Academy of Sciences of the United States of America *108*, 534-539.

Gao, G., Guo, X., and Goff, S.P. (2002). Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. Science *297*, 1703-1706.

Garcia, J.A., Harrich, D., Soultanakis, E., Wu, F., Mitsuyasu, R., and Gaynor, R.B. (1989). Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. EMBO J *8*, 765-778.

Garcin, D., Lezzi, M., Dobbs, M., Elliott, R.M., Schmaljohn, C., Kang, C.Y., and Kolakofsky, D. (1995). The 5' ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. Journal of virology *69*, 5754-5762.
Garrus, J.E., von Schwedler, U.K., Pornillos, O.W., Morham, S.G., Zavitz, K.H., Wang, H.E., Wettstein, D.A., Stray, K.M., Cote, M., Rich, R.L*., et al.* (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell *107*, 55-65.

Ge, Z., Quek, B.L., Beemon, K.L., and Hogg, J.R. (2016). Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. Elife *5*.
Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat Rev Genet *15*, 829-845.

Goodier, J.L., Pereira, G.C., Cheung, L.E., Rose, R.J., and Kazazian, H.H., Jr. (2015). The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. PLoS Genet *11*, e1005252.

Grant, A., Ponia, S.S., Tripathi, S., Balasubramaniam, V., Miorin, L., Sourisseau, M., Schwarz, M.C., Sanchez-Seco, M.P., Evans, M.J., Best, S.M*., et al.* (2016). Zika Virus Targets Human STAT2 to Inhibit Type I Interferon Signaling. Cell host & microbe *19*, 882-890.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. In RNA.
Greenbaum, B.D., Levine, A.J., Bhanot, G., and Rabadan, R. (2008). Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS pathogens *4*, e1000079.

Grover, J.R., Llewellyn, G.N., Soheilian, F., Nagashima, K., Veatch, S.L., and Ono, A. (2013). Roles played by capsid-dependent induction of membrane curvature and Gag-ESCRT interactions in tetherin recruitment to HIV-1 assembly sites. Journal of virology *87*, 4650-4664.

Guo, X., Carroll, J.W., Macdonald, M.R., Goff, S.P., and Gao, G. (2004). The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. Journal of virology *78*, 12781-12787.

Guo, X., Ma, J., Sun, J., and Gao, G. (2007). 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 *104*, 151-156.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M.*, et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.
Handschumacher, R.E., Harding, M.W., Rice, J., Drugge, R.J., and Speicher, D.W. (1984).

Cyclophilin: a specific cytosolic binding protein for cyclosporin A. Science *226*, 544-547.

Heaphy, S., Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Karn, J., Lowe, A.D., Singh, M., and Skinner, M.A. (1990). Hiv-1 Regulator of Virion Expression (Rev) Protein Binds to an Rna Stem-Loop Structure Located within the Rev Response Element Region. Cell *60*, 685-693.

Henderson, B.R., and Percipalle, P. (1997). Interactions between HIV Rev and nuclear import and export factors: the Rev nuclear localisation signal mediates specific binding to human importin-beta. Journal of molecular biology *274*, 693-707.

Hofmann, W., Schubert, D., LaBonte, J., Munson, L., Gibson, S., Scammell, J., Ferrigno, P., and Sodroski, J. (1999). Species-specific, postentry barriers to primate immunodeficiency virus infection. Journal of virology *73*, 10020-10028.

Holmes, M., Zhang, F., and Bieniasz, P.D. (2015). Single-Cell and Single-Cycle Analysis of HIV-1 Replication. PLoS pathogens *11*, e1004961.

Hornung, V., Ellegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K.K., Schlee, M.*, et al.* (2006). 5'-Triphosphate RNA is the ligand for RIG-I. Science *314*, 994-997.
Hu, W.S., and Hughes, S.H. (2012). HIV-1 reverse transcription. Cold Spring Harbor perspectives in medicine *2*.

Huang, Z., Wang, X., and Gao, G. (2010). Analyses of SELEX-derived ZAP-binding RNA aptamers suggest that the binding specificity is determined by both structure and sequence of the RNA. Protein Cell *1*, 752-759.

Iwakawa, H.o., and Tomari, Y. (2015). The Functions of MicroRNAs: mRNA Decay and Translational Repression. In Trends in Cell Biology.

Jiang, Z., Su, C., and Zheng, C. (2016). Herpes Simplex Virus 1 Tegument Protein UL41 Counteracts IFIT3 Antiviral Innate Immunity. Journal of virology *90*, 11056-11061.

Jones, K.A., Kadonaga, J.T., Luciw, P.A., and Tjian, R. (1986). Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science *232*, 755-759.

Kane, M., Rebensburg, S.V., Takata, M.A., Zang, T.M., Yamashita, M., Kvaratskhelia, M., and Bieniasz, P.D. (2018). Nuclear pore heterogeneity influences HIV-1 infection and the antiviral activity of MX2. Elife *7*.

Kane, M., Zang, T.M., Rihn, S.J., Zhang, F., Kueck, T., Alim, M., Schoggins, J., Rice, C.M., Wilson, S.J., and Bieniasz, P.D. (2016). Identification of Interferon-Stimulated Genes with Antiretroviral Activity. Cell host & microbe *20*, 392-405.

Kao, S.Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature *330*, 489-493.

Karlin, S., Doerfler, W., and Cardon, L.R. (1994). Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? Journal of virology *68*, 2889-2897.

Karlin, S., and Mrazek, J. (1997). Compositional differences within and between eukaryotic genomes. Proceedings of the National Academy of Sciences of the United States of America *94*, 10227-10232.

Karn, J., and Stoltzfus, C.M. (2012). Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harbor perspectives in medicine *2*, a006916.

Kathum, O.A., Schrader, T., Anhlan, D., Nordhoff, C., Liedmann, S., Pande, A., Mellmann, A., Ehrhardt, C., Wixler, V., and Ludwig, S. (2016). Phosphorylation of influenza A virus NS1 protein at threonine 49 suppresses its interferon antagonistic activity. Cell Microbiol *18*, 784-791.

Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T.S., Fujita, T., and Akira, S. (2008). Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. J Exp Med *205*, 1601-1610.

Keane, S.C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, A., Florwick, A., Santos, J.*, et al.* (2015). RNA structure. Structure of the HIV-1 RNA packaging signal. Science *348*, 917-921.

Keating, C.P., Hill, M.K., Hawkes, D.J., Smyth, R.P., Isel, C., Le, S.Y., Palmenberg, A.C., Marshall, J.A., Marquet, R., Nabel, G.J.*, et al.* (2009). The A-rich RNA sequences of HIV-1 pol are important for the synthesis of viral cDNA. Nucleic acids research *37*, 945-956.

Kharytonchyk, S., Monti, S., Smaldino, P.J., Van, V., Bolden, N.C., Brown, J.D., Russo, E., Swanson, C., Shuey, A., Telesnitsky, A.*, et al.* (2016). Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. Proceedings of the National Academy of Sciences of the United States of America *113*, 13378-13383.

Klimkait, T., Strebel, K., Hoggan, M.D., Martin, M.A., and Orenstein, J.M. (1990). The Human Immunodeficiency Virus Type 1-Specific Protein Vpu Is Required for Efficient Virus Maturation and Release. Journal of virology *64*, 621-629.

Kluge, S.F., Sauter, D., and Kirchhoff, F. (2015). SnapShot: antiviral restriction factors. Cell *163*, 774-774 e771.

Knoepfel, S.A., and Berkhout, B. (2013). On the role of four small hairpins in the HIV-1 RNA genome. RNA Biol *10*, 540-552.

Kobayashi, T., Ode, H., Yoshida, T., Sato, K., Gee, P., Yamamoto, S.P., Ebina, H., Strebel, K., Sato, H., and Koyanagi, Y. (2011). Identification of amino acids in the human tetherin transmembrane domain responsible for HIV-1 Vpu interaction and susceptibility. Journal of virology *85*, 932-945.

Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nature reviews Molecular cell biology *8*, 761-773.

Konig, R., Zhou, Y., Elleder, D., Diamond, T.L., Bonamy, G.M., Irelan, J.T., Chiang, C.Y., Tu, B.P., De Jesus, P.D., Lilley, C.E.*, et al.* (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell *135*, 49-60.

Koonin, E.V., Senkevich, T.G., and Dolja, V.V. (2006). The ancient Virus World and evolution of cells. Biol Direct *1*, 29.

Kotlajich, M.V., Crabb, T.L., and Hertel, K.J. (2009). Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex. Mol Cell Biol *29*, 1072-1082.

Kunec, D., and Osterrieder, N. (2016). Codon Pair Bias Is a Direct Consequence of Dinucleotide Bias. Cell reports *14*, 55-67.

Kutluay, S.B., and Bieniasz, P.D. (2010). Analysis of the initiating events in HIV-1 particle assembly and genome packaging. PLoS pathogens *6*, e1001200.

Kutluay, S.B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., and Bieniasz, P.D. (2014). Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. Cell *159*, 1096-1109.

Kuzembayeva, M., Dilley, K., Sardo, L., and Hu, W.S. (2014). Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. In Virology.

Le Nouen, C., Brock, L.G., Luongo, C., McCarty, T., Yang, L., Mehedi, M., Wimmer, E., Mueller, S., Collins, P.L., Buchholz, U.J.*, et al.* (2014). Attenuation of human respiratory syncytial virus by

genome-scale codon-pair deoptimization. Proceedings of the National Academy of Sciences of the United States of America *111*, 13169-13174.

Lee, Y.F., Nomoto, A., Detjen, B.M., and Wimmer, E. (1977). A protein covalently linked to poliovirus genome RNA. Proceedings of the National Academy of Sciences *74*, 59-63.

Legrain, P., and Rosbash, M. (1989). Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. Cell *57*, 573-583.

Leung, D.W., Prins, K.C., Borek, D.M., Farahbakhsh, M., Tufariello, J.M., Ramanan, P., Nix, J.C., Helgeson, L.A., Otwinowski, Z., Honzatko, R.B.*, et al.* (2010). Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. Nat Struct Mol Biol *17*, 165-172.

Li, M.M., Lau, Z., Cheung, P., Aguilar, E.G., Schneider, W.M., Bozzacco, L., Molina, H., Buehler, E., Takaoka, A., Rice, C.M.*, et al.* (2017). TRIM25 Enhances the Antiviral Action of Zinc-Finger Antiviral Protein (ZAP). PLoS pathogens *13*, e1006145.

Li, X., and Sodroski, J. (2008). The TRIM5alpha B-box 2 domain promotes cooperative binding to the retroviral capsid by mediating higher-order self-association. Journal of virology *82*, 11495-11502.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X.*, et al.* (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464-469.

Lilly, F. (1970). Fv-2: identification and location of a second gene governing the spleen focus response to Friend leukemia virus in mice. Journal of the National Cancer Institute *45*, 163-169.
Lim, S.R., and Hertel, K.J. (2004). Commitment to splice site pairing coincides with A complex formation. Mol Cell *15*, 477-483.

Lin, Y., Zhang, H., Liang, J., Li, K., Zhu, W., Fu, L., Wang, F., Zheng, X., Shi, H., Wu, S.*, et al.* (2014). Identification and characterization of alphavirus M1 as a selective oncolytic virus targeting ZAP-defective human cancers. Proceedings of the National Academy of Sciences of the United States of America *111*, E4504-4512.

Liu, Y., Olagnier, D., and Lin, R. (2017). Host and viral modulation of RIG-I-mediated antiviral immunity. In Frontiers in Immunology.

Llano, M., Saenz, D.T., Meehan, A., Wongthida, P., Peretz, M., Walker, W.H., Teo, W., and Poeschla, E.M. (2006). An essential role for LEDGF/p75 in HIV integration. Science *314*, 461-464.

Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem J *417*, 15-27.

Loo, Y.M., and Gale, M. (2011). Immune Signaling by RIG-I-like Receptors. In Immunity.

Loo, Y.M., Owen, D.M., Li, K., Erickson, A.K., Johnson, C.L., Fish, P.M., Carney, D.S., Wang, T., Ishida, H., Yoneyama, M*., et al.* (2006). Viral and therapeutic control of IFN-beta promoter stimulator 1 during hepatitis C virus infection. Proceedings of the National Academy of Sciences of the United States of America *103*, 6001-6006.

Luban, J., Bossolt, K.L., Franke, E.K., Kalpana, G.V., and Goff, S.P. (1993). Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. Cell *73*, 1067-1078.

MacDonald, M.R., Machlin, E.S., Albin, O.R., and Levy, D.E. (2007). The zinc finger antiviral protein acts synergistically with an interferon-induced factor for maximal activity against alphaviruses. Journal of virology *81*, 13509-13518.

Madsen, J.M., and Stoltzfus, C.M. (2006). A suboptimal 5' splice site downstream of HIV-1 splice site A1 is required for unspliced viral mRNA accumulation and efficient virus replication. Retrovirology *3*, 10.

Mahadeo, D., Kaplan, L., Chao, M.V., and Hempstead, B.L. (1994). High affinity nerve growth factor binding displays a faster rate of association than p140trk binding. Implications for multi-subunit polypeptide receptors. J Biol Chem *269*, 6884-6891.

Malim, M.H., and Bieniasz, P.D. (2012). HIV Restriction Factors and Mechanisms of Evasion. Cold Spring Harbor perspectives in medicine *2*, a006940.

Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V., and Cullen, B.R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature *338*, 254-257.

Malim, M.H., Tiley, L.S., McCarn, D.F., Rusche, J.R., Hauber, J., and Cullen, B.R. (1990). HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. Cell *60*, 675-683.

Mandal, D., Exline, C.M., Feng, Z., and Stoltzfus, C.M. (2009). Regulation of Vif mRNA splicing by human immunodeficiency virus type 1 requires 5' splice site D2 and an exonic splicing enhancer to counteract cellular restriction factor APOBEC3G. Journal of virology *83*, 6067-6078.

Mao, R., Nie, H., Cai, D., Zhang, J., Liu, H., Yan, R., Cuconati, A., Block, T.M., Guo, J.T., and Guo, H. (2013). Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein. PLoS pathogens *9*, e1003494.

Marq, J.B., Kolakofsky, D., and Garcin, D. (2010). Unpaired 5' ppp-nucleotides, as found in arenavirus double-stranded RNA panhandles, are not recognized by RIG-I. J Biol Chem *285*, 18208-18216.

Martin Stoltzfus, C. (2009). Chapter 1 Regulation of HIV-1 Alternative RNA Splicing and Its Role in Virus Replication. In Advances in Virus Research.

Martrus, G., Nevot, M., Andres, C., Clotet, B., and Martinez, M.A. (2013). Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. Retrovirology *10*, 78.

McCauley, S.M., Kim, K., Nowosielska, A., Dauphin, A., Yurkovetskiy, L., Diehl, W.E., and Luban, J. (2018). Intron-containing RNA from the HIV-1 provirus activates type I interferon and inflammatory cytokines. Nature communications *9*, 5305.

McNatt, M.W., Zang, T., Hatziioannou, T., Bartlett, M., Fofana, I.B., Johnson, W.E., Neil, S.J., and Bieniasz, P.D. (2009). Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. PLoS pathogens *5*, e1000300.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol *2*, E234.

Moldovan, J.B., and Moran, J.V. (2015). The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. PLoS Genet *11*, e1005121.

Muesing, M.A., Smith, D.H., and Capon, D.J. (1987). Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. Cell *48*, 691-701.

Muller, S., Moller, P., Bick, M.J., Wurr, S., Becker, S., Gunther, S., and Kummerer, B.M. (2007). Inhibition of filovirus replication by the zinc finger antiviral protein. Journal of virology *81*, 2391-2400.

Nabel, G., and Baltimore, D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature *326*, 711-713.

Neil, S.J., Eastman, S.W., Jouvenet, N., and Bieniasz, P.D. (2006). HIV-1 Vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane. PLoS pathogens *2*, e39.

Neil, S.J., Sandrin, V., Sundquist, W.I., and Bieniasz, P.D. (2007). 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 & microbe *2*, 193-203.

Neil, S.J., Zang, T., and Bieniasz, P.D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature *451*, 425-430.

Nelson, K.K., and Green, M.R. (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity. Genes Dev *3*, 1562-1571.

Ocwieja, K.E., Sherrill-Mix, S., Mukherjee, R., Custers-Allen, R., David, P., Brown, M., Wang, S., Link, D.R., Olson, J., Travers, K.*, et al.* (2012). Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic acids research *40*, 10345-10355.

Oshiumi, H., Matsumoto, M., Hatakeyama, S., and Seya, T. (2009). Riplet/RNF135, a RING finger protein, ubiquitinates RIG-I to promote interferon-beta induction during the early phase of viral infection. J Biol Chem *284*, 807-817.

Parkin, N.T., Chamorro, M., and Varmus, H.E. (1992). Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. Journal of virology *66*, 5147-5151.

Perez-Caballero, D., Zang, T., Ebrahimi, A., McNatt, M.W., Gregory, D.A., Johnson, M.C., and Bieniasz, P.D. (2009). Tetherin inhibits HIV-1 release by directly tethering virions to cells. Cell *139*, 499-511.

Pizzato, M., Erlwein, O., Bonsall, D., Kaye, S., Muir, D., and McClure, M.O. (2009). A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. J Virol Methods *156*, 1-7.

Pollom, E., Dang, K.K., Potter, E.L., Gorelick, R.J., Burch, C.L., Weeks, K.M., and Swanstrom, R. (2013). Comparison of SIV and HIV-1 genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. PLoS pathogens *9*, e1003294.

Pornillos, O., Ganser-Pornillos, B.K., Kelly, B.N., Hua, Y., Whitby, F.G., Stout, C.D., Sundquist, W.I., Hill, C.P., and Yeager, M. (2009). X-ray structures of the hexameric building block of the HIV capsid. Cell *137*, 1282-1292.

Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D., and Williamson, J.R. (1992). Conformation of the TAR RNA-arginine complex by NMR spectroscopy. Science *257*, 76-80.

Purcell, D.F., and Martin, M.A. (1993). Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. Journal of virology *67*, 6365-6378.

Rahm, N., and Telenti, A. (2012). The role of tripartite motif family members in mediating susceptibility to HIV-1 infection. In Current Opinion in HIV and AIDS.

Rajsbaum, R., Albrecht, R.A., Wang, M.K., Maharaj, N.P., Versteeg, G.A., Nistal-Villan, E., Garcia-Sastre, A., and Gack, M.U. (2012). Species-specific inhibition of RIG-I ubiquitination and IFN induction by the influenza A virus NS1 protein. PLoS pathogens *8*, e1003059.

Ren, X., and Hurley, J.H. (2011). Proline-rich regions and motifs in trafficking: From ESCRT interaction to viral exploitation. In Traffic.

Rima, B.K., and McFerran, N.V. (1997). Dinucleotide and stop codon frequencies in single-stranded RNA viruses. J Gen Virol *78 (Pt 11)*, 2859-2870.

Rittner, K., Churcher, M.J., Gait, M.J., and Karn, J. (1995). The human immunodeficiency virus long terminal repeat includes a specialised initiator element which is required for Tat-responsive transcription. Journal of molecular biology *248*, 562-580.

Saad, J.S., Miller, J., Tai, J., Kim, A., Ghanam, R.H., and Summers, M.F. (2006). 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 *103*, 11364-11369.

Saito, T., Owen, D.M., Jiang, F., Marcotrigiano, J., and Gale, M., Jr. (2008). Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. Nature *454*, 523-527.

Sanchez, J.G., Chiang, J.J., Sparrer, K.M.J., Alam, S.L., Chi, M., Roganowicz, M.D., Sankaran, B., Gack, M.U., and Pornillos, O. (2016). Mechanism of TRIM25 Catalytic Activation in the Antiviral RIG-I Pathway. Cell reports *16*, 1315-1325.

Sanchez, J.G., Okreglicka, K., Chandrasekaran, V., Welker, J.M., Sundquist, W.I., and Pornillos, O. (2014). The tripartite motif coiled-coil is an elongated antiparallel hairpin dimer. Proceedings of the National Academy of Sciences of the United States of America *111*, 2494-2499.

Sanchez, J.G., Sparrer, K.M.J., Chiang, C., Reis, R.A., Chiang, J.J., Zurenski, M.A., Wan, Y., Gack, M.U., and Pornillos, O. (2018). TRIM25 Binds RNA to Modulate Cellular Anti-viral Defense. Journal of molecular biology *430*, 5280-5293.

Sauter, D. (2014). Counteraction of the multifunctional restriction factor tetherin. Front Microbiol *5*, 163.

Sayah, D.M., Sokolskaja, E., Berthoux, L., and Luban, J. (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature *430*, 569-573.

Schmitz, M.L., Kracht, M., and Saul, V.V. (2014). The intricate interplay between RNA viruses and NF-κB. In Biochimica et Biophysica Acta - Molecular Cell Research.

Schneider, U., Schwemmle, M., and Staeheli, P. (2005). Genome trimming: a unique strategy for replication control employed by Borna disease virus. Proceedings of the National Academy of Sciences of the United States of America *102*, 3441-3446.

Schnell, G., Loo, Y.M., Marcotrigiano, J., and Gale, M., Jr. (2012). Uridine composition of the poly-U/UC tract of HCV RNA defines non-self recognition by RIG-I. PLoS pathogens *8*, e1002839.

Schoggins, J.W., MacDuff, D.A., Imanaka, N., Gainey, M.D., Shrestha, B., Eitson, J.L., Mar, K.B., Richardson, R.B., Ratushny, A.V., Litvak, V.*, et al.* (2014). Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. Nature *505*, 691-695.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.

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

Selby, M.J., Bain, E.S., Luciw, P.A., and Peterlin, B.M. (1989). Structure, Sequence, and Position of the Stem Loop in Tar Determine Transcriptional Elongation by Tat through the Hiv-1 Long Terminal Repeat. Genes & Development *3*, 547-558.

Seth, R.B., Sun, L., Ea, C.K., and Chen, Z.J. (2005). Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. Cell *122*, 669-682.

Shaik, M.M., Peng, H., Lu, J., Rits-Volloch, S., Xu, C., Liao, M., and Chen, B. (2019). Structural basis of coreceptor recognition by HIV-1 envelope spike. Nature *565*, 318-323.

Sharp, P.M., and Hahn, B.H. (2011). Origins of HIV and the AIDS pandemic. Cold Spring Harbor perspectives in medicine *1*, a006841.

Shen, G., Wang, K., Wang, S., Cai, M., Li, M.L., and Zheng, C. (2014). Herpes simplex virus 1 counteracts viperin via its virion host shutoff protein UL41. Journal of virology *88*, 12163-12166.
Smith, C.W.J., and Valcárcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. In Trends in Biochemical Sciences.

Sodroski, J., Goh, W.C., Rosen, C., Dayton, A., Terwilliger, E., and Haseltine, W. (1986). A second post-transcriptional trans-activator gene required for HTLV-III replication. Nature *321*, 412-417.

Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F., and Haseltine, W. (1985a). Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III. Science *229*, 74-77.

Sodroski, J., Rosen, C., Wong-Staal, F., Zaki Salahuddin, S., Popovic, M., Arya, S., Gallo, R.C., and Haseltine, W.A. (1985b). Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. Science.

Sterner, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. Proceedings of the National Academy of Sciences of the United States of America *93*, 15081-15085.

Stevenson, N.J., Bourke, N.M., Ryan, E.J., Binder, M., Fanning, L., Johnston, J.A., Hegarty, J.E., Long, A., and O'Farrelly, C. (2013). Hepatitis C virus targets the interferon-alpha JAK/STAT pathway by promoting proteasomal degradation in immune cells and hepatocytes. FEBS Lett *587*, 1571-1578.

Stremlau, M., Perron, M., Lee, M., Li, Y., Song, B., Javanbakht, H., Diaz-Griffero, F., Anderson, D.J., Sundquist, W.I., and Sodroski, J. (2006). 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 *103*, 5514-5519.

Su, C., Zhang, J., and Zheng, C. (2015). Herpes simplex virus 1 UL41 protein abrogates the antiviral activity of hZAP by degrading its mRNA. Virol J *12*, 203.

Su, C., and Zheng, C. (2017). Herpes Simplex Virus 1 Abrogates the cGAS/STING-Mediated Cytosolic DNA-Sensing Pathway via Its Virion Host Shutoff Protein, UL41. Journal of virology *91*.

Sumpter, R., Jr., Loo, Y.M., Foy, E., Li, K., Yoneyama, M., Fujita, T., Lemon, S.M., and Gale, M., Jr. (2005). Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. Journal of virology *79*, 2689-2699.

Sundquist, W.I., and Kräusslich, H.G. (2012). HIV-1 assembly, budding, and maturation. In Cold Spring Harbor perspectives in medicine.

Swanson, C.M., and Malim, M.H. (2008). SnapShot: HIV-1 proteins. Cell *133*, 742, 742 e741.

Tacke, R., and Manley, J.L. (1999). Determinants of SR protein specificity. In Current Opinion in Cell Biology.

Tan, R., Chen, L., Buettner, J.A., Hudson, D., and Frankel, A.D. (1993). RNA recognition by an isolated alpha helix. Cell.

Tang, Q., Wang, X., and Gao, G. (2017). The Short Form of the Zinc Finger Antiviral Protein Inhibits Influenza A Virus Protein Expression and Is Antagonized by the Virus-Encoded NS1. Journal of virology *91*.

Telesnitsky, A., and Goff, S. (1997). Reverse Transcriptase and the Generation of Retroviral DNA. Retroviruses.

Todorova, T., Bock, F.J., and Chang, P. (2014). PARP13 regulates cellular mRNA post-transcriptionally and functions as a pro-apoptotic factor by destabilizing TRAILR4 transcript. Nature communications *5*, 5362.

Towers, G.J., Hatziioannou, T., Cowan, S., Goff, S.P., Luban, J., and Bieniasz, P.D. (2003). Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. Nat Med *9*, 1138-1143.

Tulloch, F., Atkinson, N.J., Evans, D.J., Ryan, M.D., and Simmonds, P. (2014). RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. Elife *3*, e04531.

Uchida, L., Espada-Murao, L.A., Takamatsu, Y., Okamoto, K., Hayasaka, D., Yu, F., Nabeshima, T., Buerano, C.C., and Morita, K. (2014). The dengue virus conceals double-stranded RNA in the intracellular membrane to escape from an interferon response. Scientific reports *4*, 7395.

Uzri, D., and Gehrke, L. (2009). Nucleotide sequences and modifications that determine RIG-I/RNA binding and signaling activities. Journal of virology *83*, 4174-4184.

Varthakavi, V., Smith, R.M., Bour, S.P., Strebel, K., and Spearman, P. (2003). Viral protein U counteracts a human host cell restriction that inhibits HIV-1 particle production. Proceedings of the National Academy of Sciences of the United States of America *100*, 15154-15159.

Wang, H., Vaheri, A., Weber, F., and Plyusnin, A. (2011). Old World hantaviruses do not produce detectable amounts of dsRNA in infected cells and the 5' termini of their genomic RNAs are monophosphorylated. J Gen Virol *92*, 1199-1204.

Wang, Q., Barr, I., Guo, F., and Lee, C. (2008). Evidence of a novel RNA secondary structure in the coding region of HIV-1 pol gene. RNA *14*, 2478-2488.

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

Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., and Jones, K.A. (1998). 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.

Wilen, C.B., Tilton, J.C., and Doms, R.W. (2012). HIV: cell binding and entry. Cold Spring Harbor perspectives in medicine *2*.

Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., and Weeks, K.M. (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol *6*, e96.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41-51.

Yamashita, M., and Emerman, M. (2004). Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. Journal of virology *78*, 5670-5678.

Yamashita, M., Perez, O., Hope, T.J., and Emerman, M. (2007). Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. PLoS pathogens *3*, 1502-1510.

Ye, P., Liu, S., Zhu, Y., Chen, G., and Gao, G. (2010). DEXH-Box protein DHX30 is required for optimal function of the zinc-finger antiviral protein. Protein Cell *1*, 956-964.

Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol *11*, 377-394.

Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., and Fujita, T. (2004). The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nat Immunol *5*, 730-737.

York, A., Kutluay, S.B., Errando, M., and Bieniasz, P.D. (2016). The RNA Binding Specificity of Human APOBEC3 Proteins Resembles That of HIV-1 Nucleocapsid. PLoS pathogens *12*, e1005833.

Zamore, P.D., and Green, M.R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proceedings of the National Academy of Sciences.

Zenzie-Gregory, B., Khachi, A., Garraway, I.P., and Smale, S.T. (1993). Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. Molecular and Cellular Biology *13*, 3841-3849.

Zheng, X., Wang, X., Tu, F., Wang, Q., Fan, Z., and Gao, G. (2017). TRIM25 Is Required for the Antiviral Activity of Zinc Finger Antiviral Protein. Journal of virology *91*.

Zheng, Y.H., Jeang, K.T., and Tokunaga, K. (2012). Host restriction factors in retroviral infection: Promises in virus-host interaction. In Retrovirology.

Zhu, P., Liu, J., Bess, J., Jr., Chertova, E., Lifson, J.D., Grise, H., Ofek, G.A., Taylor, K.A., and Roux, K.H. (2006). Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature *441*, 847-852.

Zhu, Y., Chen, G., Lv, F., Wang, X., Ji, X., Xu, Y., Sun, J., Wu, L., Zheng, Y.T., and Gao, G. (2011). 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 *108*, 15834-15839.

Zhu, Y., Wang, X., Goff, S.P., and Gao, G. (2012). Translational repression precedes and is required for ZAP-mediated mRNA decay. EMBO J *31*, 4236-4246.