REGULATION OF HIV-1 SPLICING

Ann Emery

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology.

Chapel Hill
2017

Approved by:

Ronald Swanstrom

Corbin Jones

Piotr Mieczkowski

Nathaniel Moorman

Jan Prins

Kevin Weeks

ABSTRACT

Ann Emery: Regulation of HIV-1 Splicing
(Under the direction of Ronald Swanstrom)

The Human Immunodeficiency Virus Type 1 (HIV-1) has a single primary transcript – full-length genomic RNA. Left unspliced, it serves as either genomic RNA or as mRNA for the viral reverse transcriptase, protease, integrase, and structural proteins. The mRNAs for all other viral proteins require splicing of the full-length transcript. HIV-1 undergoes a complex program of splicing and suppression of splicing to make more than 50 transcript types. Since these complex splicing patterns are essential for viral replication, splicing disruption could be a point of vulnerability given a detailed understanding of the steps involved.

In order to assess the regulation of splicing, the products of splicing have to be quantifiable. This dissertation describes two Primer ID-tagged deep sequencing assays developed to quantify HIV-1 splicing in the context of viral infection/transfection and in the context of a full-length viral genome. The depth of sequencing allows quantification of even rare splicing events.

Using these deep sequencing assays I examined splicing across HIV-1 subtypes and between HIV-1 and SIVmac239 and found that while patterns of splicing are well conserved, wide variation in acceptor usage is tolerated among different HIV-1 strains. Correct splicing depends on cis-acting control sequences interacting with cellular splicing factors. I quantified the effects of mutations that alter these sequences or the secondary RNA structures containing them. I reevaluated the effects of previously characterized splicing regulatory sequences, validating the functions of some but redefining others. In collaborative efforts I showed that mutation of splicing factor binding sites or knock down of cellular splicing proteins produces specific splicing phenotypes. Analysis of a set of global silent mutations across the HIV-1 genome found regions of the viral genome that interfere with suppression of splicing and carry a high fitness cost.

These experiments show that HIV-1 splicing regulation involves more complex patterns of factor binding and cooperative interactions than previously described, suggesting that existing models are overly simplistic, and my studies contribute to a more accurate description and understanding of HIV-1 splicing.

PREFACE

Much of the work described here was done in collaboration with others. Chapter 2 describes the splicing assay that is the basis for my thesis work. I am the lead author and did the experiments and wrote the data analysis code. Two members of the Swanstrom lab, Elizabeth Pollom and Shuntai Zhou, provided background work and help with deep sequencing respectively. This paper has been published with the following citation:

> Emery A, Zhou S, Pollom E, Swanstrom R. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. Journal of virology. 2017;91(6). Epub 2017/01/13. doi: 10.1128/jvi.02515-16. PubMed PMID: 28077653; PubMed Central PMCID: PMCPMC5331825.

Chapter 3 was done in collaboration with members of Paul Bieniasz's lab at The Rockefeller University: Matthew Takata, Steven J. Soll, and Daniel Blanco-Melo. This study involved large mutations across the HIV-1 genome and my part in this work was to identify splicing changes in the mutants and revertants. This involved adapting the splicing assay protocol and data analysis program to each mutant. I also looked for cryptic splicing and this required development of another program that identifies splice sites from sequencing data. The manuscript was written by Matthew Takata and Paul Bieniasz. I assisted with editing. This paper is currently under revision at Plos Pathogens.

Chapter 4 was a collaboration with Sebla Kutluay's lab at Washington University, Blanton Tolbert's lab at Case Western Reserve University, and Paul Bienaisz's lab at The Rockefeller University, with contributions from the following people: Srinivas Penumutchu, Dana Townsend, Michaela Madison, Amanda Stukenbroeker, Chelsea Powell, David Jannain. My contribution was adapting the splicing program to the strains used and the analysis of splicing for siRNA knockdowns of cellular splicing factors and in hnRNP H1 binding site mutants (Figure 4, B and C; Figure 7 B-D). The manuscript was primarily written by Sebla Kutluay, with input from other authors. My part of the writing was to describe the splicing assay and some aspects of the splicing data. This manuscript is ready for submission pending the results of some additional experiments.

Chapter 5 is a bioinformatics analysis of a subset of mutations and their effects on splicing, and is my own work.

Chapter 6 is a collaboration with the Tolbert (Case Western) and Telesnitsky (University of Michigan) labs. Jeff Levengood and John Collins designed and constructed the mutations to the splicing regulatory elements. I transfected the constructs and grew the virus, and then prepared the samples for deep sequencing. I developed the random reverse primer assay to use with these samples, and also adapted the splicing assay and splice site finder program to process the data.

The paper in the appendix has been previously published with the following citation:

> Miller CM, Akiyama H, Agosto LM, Emery A, Ettinger CR, Swanstrom RI, Henderson AJ, Gummuluru S. 2017. Virion-Associated Vpr Alleviates a Postintegration Block to HIV-1 Infection of Dendritic Cells. J Virol 91.

My contribution was the analysis of splicing in cells infected with wild type or mutant Vpr viruses, with the results shown in Figure 8.5 E-G.

TABLE OF CONTENTS

LIST OF FIGURES xiii

LIST OF TABLES xvi

LIST OF ABBREVIATIONS xvii

CHAPTER 1: INTRODUCTION 1

Human Immunodeficiency Virus 1

Splicing 3

Alternative Splicing 11

Splicing in HIV-1 15

Thesis Overview 21

REFERENCES 23

CHAPTER 2: CHARACTERIZING HIV-1 SPLICING USING NEXT GENERATION SEQUENCING 33

Overview 33

Importance 33

Introduction 34

Materials and Methods 36

Viruses 36

Cells 37

RNA extraction, cDNA synthesis, and Illumina MiSeq library preparation 38

Sequencing 41

Splicing analysis 41

Results 42

Description of a new splicing assay 42

Quantification of acceptor splice site usage in NL4-3 43

Differential effects of temperature on splicing 45
Changes in secondary structure can affect splicing 47
Combined effects of temperature and structure 48
Splicing in transfected cells 49
Comparison of splicing in transmitted/founder viruses 50
Detection of cryptic/rare splicing events 52
Splicing in subtype C 55
Splicing quantification in in SIVmac239 57
Discussion 60
REFERENCES 67
CHAPTER 3: GLOBAL SYNONYMOUS MUTAGENESIS IDENTIFIES *CIS*-ACTING RNA ELEMENTS THAT REGULATE HIV-1 SPLICING AND REPLICATION 72
Overview 72
Introduction 73
Results 76
Design and synthesis of synonymously mutated HIV-1 strains 76
Protein expression and replication properties of synonymously mutated HIV-1 mutants 76
Assays for splicing perturbations in replication-defective HIV-1 mutants 79
Activation of cryptic splice sites in mutants A and B (Group 2a) 81
Overuse of canonical splice acceptor sites in mutants I, J and K (Group 2b) 81
Viruses with blocks of synonymous mutations with three different phenotypes that map to distinct regions of the HIV-1 genome 82
Mapping elements responsible for splicing perturbations in Group 2 viruses 84
Redundant HIV-1 RNA sequences mediate suppression of cryptic splice sites that are activated in Group 2a mutant viruses 85
Multiple novel *cis*-acting RNA elements suppress HIV-1 splicing at A1 and A2 88
A novel cis-acting RNA element inhibits HIV-1 splicing at A3 93
Discussion 95

Material and Methods 100
Cell lines, viruses and infections 100
Construction of mutant proviral plasmids 100
PCR quantification of unspliced HIV-1 RNA 100
Antibodies and western blotting 101
Analysis of HIV-1 splicing with fluorescent primer PCR 101
Analysis of HIV-1 splicing using Primer ID-based deep sequencing 102
REFERENCES 103
CHAPTER 4: CLIP-SEQ REVEALS A KEY ROLE FOR hnRNP H1 IN REGULATION OF HIV-1 ALTERNATIVE mRNA SPLICING 106
Overview 106
Introduction 107
Results 109
Generation of 3xHA-tagged hnRNP variants 109
Binding sites of hnRNPs on viral RNAs 111
HnRNPs bind primarily to intronic sequences on cellular RNAs 113
Direct effect of hnRNP RNAi on HIV-1 splicing 115
hnRNP H1 binds to purine-rich sequences on viral RNAs 117
Calorimetric and NMR titrations validate hnRNP H1 RNA binding preferences 119
Effect of mutations within the hnRNP H1 binding sites 120
Discussion 123
Materials and Methods 125
Cell culture, transfection and infections 125
Antibodies 126
Plasmids 126
CLIP-seq experiments 128
Nucleic acid isolation, reverse transcription, Q-PCR 129
Splicing assays 129

Protein expression and purification 130
Isothermal titration calorimetry 130
NMR experiments 130
REFERENCES 132
CHAPTER 5: USING APOBEC3G AS A MUTAGEN TO PROBE HIV-1 SPLICING SIGNALS 136
Background 136
Sequencing and Bioinformatics Methods 139
Results 139
Mutations are not equally represented in the two HIV-1 spliced transcript size classes 139
G-to-A and A-to-G mutations are skewed in size class distribution 142
Mutations between A5 and D4 affect A5 usage 143
Discussion 144
REFERENCES 149
CHAPTER 6: A SURVEY OF KNOWN HIV-1 SPLICE REGULATORY ELEMENTS (SRE): A RE-EVALUATION 153
Background 153
Materials and Methods 154
Cells and constructs 154
Sequencing and bioinformatics 154
The Random Reverse Primer Assay (RRPA) 155
Results 157
Mutations to ESEVif, ESEM1, and ESEM2 alter splicing to acceptor A1 157
Mutation of ESSV activates acceptor A2 usage and also activates D1 to cause oversplicing 159
Mutation of ESS2 and ESS2p synergistically activate A3 and D1 to cause oversplicing 161
Combined mutation of ESSV, ESS2p, and ESS2 caused oversplicing to both A2 and A3 163
ISS suppresses pseudo-splice acceptors upstream of A7 164

Mutations to the ESS3 stem-loop cause a small increase in splicing from D4 to A7 165

Discussion 167

REFERENCES 169

CHAPTER 7: CONCLUSIONS AND FUTURE DIRECTIONS 171

REFERENCES 179

APPENDIX: VIRION ASSOCIATED VPR ALLEVIATES A POST-INTEGRATION BLOCK TO HIV-1 INFECTION OF DENDRITIC CELLS 181

Overview 181

Importance 181

Introduction 182

Results 183

Vpr-deficient viruses display a replication defect in DCs 183

Defects in Vpr infection are independent of viral glycoprotein expression 186

Infection with Vpr-deficient HIV-1 does not induce type 1 IFN 188

Infection with ΔVpr viruses results in decreased infection in a single round of replication and is rescued by virion-associated Vpr 190

Proviral LTR-mediated transcriptional activity is attenuated in Vpr-deficient virus infection in DCs 191

Mutations in the C-terminal end of Vpr or those that disrupt binding to $CRL4^{DCAF1}$ ubiquitin ligase attenuate viral replication in DCs 193

Discussion 197

Materials and Methods 200

Plasmids 200

Cells and viruses 200

Drug treatments 201

Quantitative western blotting 201

Quantitative RT-PCR 201

Quantification of viral integration 202

Splicing assay 202

Cytokine measurements 203

FACS 203

REFERENCES 204

LIST OF FIGURES

Figure 1.1. Splicing Signal Sequences. 4

Figure 1.2. Spliceosome E Complex. 5

Figure 1.3. Spliceosome A Complex. 5

Figure 1.4. Spliceosome B Complex. 6

Figure 1.5. Spliceosome B* Complex 7

Figure 1.6. Spliceosome C Complex. 7

Figure 1.7. Post-Splicing Products. 8

Figure 1.8. Exon Recognition and Inclusion. 12

Figure 1.9. Intron Retention. 12

Figure 1.10. Undefined Exon Skipped. 12

Figure 1.11. Alternative Acceptor Usage. 13

Figure 1.12. Inclusion or Skipping of an Alternative Exon. 14

Figure 1.13. Splicing in HIV-1. 15

Figure 1.14. Major and Minor HIV-1 Splice Sites. 17

Figure 1.15. HIV-1 from and Exon/Intron Perspective. 18

Figure 1.16. Known HIV-1 SREs. 18

Figure 2.1. HIV-1 Splice Patterns and Primer Locations. 35

Figure 2.2 Quantification of HIV-1 Splicing Patterns. 44

Figure 2.3. Temperature Dependent Splicing Regulation. 46

Figure 2.4. Effect of SLSA1 Structural Mutations on Splicing. 48

Figure 2.5. Combined Effects of Temperature and SLSA1 Mutation. 49

Figure 2.6. Splicing in Transmitted/Founder Viruses. 51

Figure 2.7. Cryptic Acceptor Usage as a Percentage of Major Acceptor Usage in Transmitted/Founder Viruses and NL4-3. 53

Figure 2.8. Quantification of Splicing in Subtype C Strain pZM247Fv2. 56

Figure 2.9. SIVmac239 Splice Patterns and Primer Locations. 57

Figure 2.10. Quantification of Splicing in SIVmac239. 58

Figure 3.1. Design and analysis of panel of synonymously mutated HIV-1 viruses. 75

Figure 3.2. Spreading replication properties of mutant viruses. 78

Figure 3.3. Analysis of HIV-1 splicing in WT and synonymously mutated HIV-1. 80

Figure 3.4. Phenotypes of synonymously mutated HIV-1 viruses. 83

Figure 3.5. Activation of cryptic splice sites by synonymous mutations in Gag. 87

Figure 3.6. Activation of canonical splice acceptor sites (A1 and A2) by synonymous mutations in mutant I. 90

Figure 3.7. Activation of canonical splice acceptor site A2 by synonymous mutations in mutant J. 92

Figure 3.8. Activation of canonical splice acceptor site A3 by synonymous mutations in mutant K. 94

Figure 3.9. Summary of splicing control in HIV-1. 97

Figure 4.1. Generation of cell lines stably expressing HA-tagged hnRNPs and summary of CLIP-seq. 110

Figure 4.2. CLIP-seq of hnRNPs reveal their binding sites on viral RNAs. 112

Figure 4.3. Classification of hnRNP-bound cellular RNAs and the binding motifs identified within them. 114

Figure 4.4. RNAi of hnRNP H1 decreases Gag expression and alters viral splicing in cells. 116

Figure 4.5. HnRNP H1 binds to purine-rich sequences on the viral genome. 118

Figure 4.6. HnRNP H1 exhibits specificity for G-rich sequences in vitro. 119

Figure 4.7. Mutation of hnRNP H1 binding sites alters HIV-1 replication and alternative splicing. 122

Figure 5.1. Target sequence between A5 and D4. 138

Figure 5.2. Mutation rates for WT or SLSA1 (low *vif*) in the 1.8 kb or 4 kb size classes. 140

Figure 5.3. G-to-A mutations as a proportion of total mutations. 141

Figure 5.4. Skew of G-to-A mutations. 142

Figure 5.5. Skew of A-to-G mutations. 143

Figure 5.6. Mutations affecting the use of acceptor A5. 144

Figure 6.1. Relative positions of known HIV-1 SREs. 154

Figure 6.2. Random reverse primer outcomes. 156

Figure 6.3. Splicing regulatory elements mutated. 157

Figure 6.4. Usage of acceptor A1 in mutated SREs between A1 and D2. 158

Figure 6.5. Acceptor usage in ESSV mutant relative to WT. 160

Figure 6.6. Percent of unspliced transcripts in the WT and the ESSV mutant. 161

Figure 6.7. Acceptor usage for mutations to ESS2 and ESS2p. 162

Figure 6.8. Effects of ESS2 and ESS2p mutation on percentage of unspliced transcripts. 163

Figure 6.9. Two transcript types make up the majority of transcripts in the ESS2/ESS2p/ESSV combined mutations. 164

Figure 6.10. ISS suppresses cryptic splicing near A7. 165

Figure 6.11. The ESS3 stem-loop structure located downstream of splice acceptor A7. 166

Figure 6.12. Ratio of 1.8 kb transcripts to 4 kb transcripts in each ESS3 mutant and its internal NL4-3 control. 167

Figure 7.1. A Proposed Model for hnRNP Binding. 177

Figure 8.1. Infection with Vpr-deficient HIV-1 results in attenuated virus replication in MDDCs and MDDC-T co-cultures. 185

Figure 8.2. Vpr does not regulate Env expression in infected MDDCs or incorporation of Env into MDDC-derived virions. 187

Figure 8.3. Vpr-deficiency does not result in enhanced type I IFN production in productively infected MDDCs. 189

Figure 8.4. Infection of MDDCs with Vpr-deficient viruses results in block to HIV-1 replication in single round infection analysis. 191

Figure 8.5. Viral transcription is attenuated in ΔVpr virus infected MDDCs. 192

Figure 8.6. Vpr mutants deficient for interaction with DCAF1/DDB1/E3 ubiquitin ligase and inducing G2 cell cycle arrest are attenuated in a single cycle of replication analysis in MDDCs. 196

LIST OF TABLES

Table 2.1. Primers Used in This Study. 40

Table 2.2. Primers Used for Each Sample..... 41

LIST OF ABBREVIATIONS

4-SU	4-Thiouridine
A	Adenine
A3G	APOBEC 3G family
ABI	Applied Biosystems
AX	Splice Acceptor Number X
ADAR	Adenosine Deaminase Acting on RNA
AIDS	Acquired Immunodeficiency Syndrome
APC	Allophycocyanin
APOBEC	Apolipoprotein B mRNA Editing Enzyme, Catalytic Polypeptide-Like
Arg	Arginine
ASF/SF2	Serine Arginine Rich Splicing Factor 1
ATM	Ataxia-Telangiectasia Mutated Kinase
ATP	Adenosine Triphosphate
AZT	Azidothymidine
BST-2	Bone Marrow Stromal Cell Antigen 2 or Tetherin
C	Cytosine
CA	HIV-1 Capsid Protein
CCR5	C-C Motif Chemokine Receptor 5
CD4	Cluster of Differentiation 4 Receptor
CD11c	Integrin, AlphaX (complement component 3 receptor 4 subunit)
cDNA	Complementary DNA
cERMIT	(Conserved) Evidence-Ranked Motif Identification Tool Algorithm
CLIP-seq	Crosslinking Immunoprecipitation and Sequencing
$CRL4^{DCAF1}$	Cul4A/DCAF/DDB1 E3 Ubiquitin Ligase Complex
Crm1	Chromosomal Maintenance 1 or Exportin 1
C-terminal	Carboxyl Terminal
Cul4A	Cullin 4A

CXCR4	CXC Motif Chemokine Receptor
D	Splice Donor
DC	Dendritic Cells
DCAF	DDB1 and CUL4-Associated Factors
DC-SIGN	Dendritic Cell-Specific Intercellular Adhesion Molecule-3-Grabbing Non-Integrin Receptor
DDB1	DNA Damage-Binding Protein 1
DDR	DNA Damage Response
DGCR8	DiGeorge Syndrome Critical Region 8
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphates
E3	Ubiquitin Ligase
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
Env	HIV-1 Envelope Protein
ER	Endoplasmic Reticulum
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
FACS	Fluorescence Activated Cell Sorting
FSC	Forward Scatter
F/T	Founder/Transmitted
G	Guanine
G2/M	Cell Cycle Interphase and Mitosis
Gag	HIV-1 Group Specific Antigen
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GAR ESE	Guanine-Adenine Rich Exonic Splicing Enhancer
GE	General Electric

GFP	Green Fluorescent Protein
Gly	Glycine
GPP	Gag-Pro-Pol polyprotein
HA	Hemagglutinin
HD	Phosphohydrolase Protein Family
HIS_6	Histadine Tag
HIV-1	Human Immunodeficiency Virus Type 1
HLTF	Helicase-like Transcription Factor
hnRNP	Heterogeneous Nuclear Ribonucleoproteins
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
IFN	Interferon
Ig	Immunoglobulin
IL-2	Interleukin-2
IN	HIV-1 Integrase Protein
IRF	Interferon Regulatory Factor
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
ITC	Isothermal Titration Calorimetry
Kb	Kilobase
LTR	Long Terminal Repeat
MA	HIV-1 Matrix Protein
MDDC	Monocyte-Derived Dendritic Cells
MDM	Monocyte-Derived Macrophages
MHC	Major Histocompatibility Complex
MHz	Megahertz
MOI	Multiplicity of Infection
mRNA	Messenger Ribonucleic Acid
Mus81-Eme1	Holliday Junction Resolvase

Nef	Viral Negative Regulatory Factor
NF-κB	Nuclear Factor Kappa-light-chain-enhancer of Activated B Cells
NGS	Next Generation Sequencing
NIAID	National Institute of Allergy and Infectious Diseases
NIH	National Institute of Health
Ni-IDA	Nickle-iminodiacetic Acid
NMD	Nonsense Mediated Decay
NMR	Nuclear Magnetic Resonance
NT	No Treatment
ORF	Open Reading Frame
p24	24kDa HIV-1 Capsid Protein
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEI	Polyethylenimine
PFA	Paraformaldehyde
PHA	Phytohaemagglutinin P
Pol	HIV-1 Reverse Transcriptase Polymerase
PPT	Polypyrimidine Tracts
PR	HIV-1 Protease
P-TEFb	Positive Transcription Elongation Factor
Q-PCR	Quantitative Polymerase Chain Reaction
qRRM	Quasi-RNA Recognition Motif
Rev	Regulator of Expression of Virion Proteins
RGG boxes	Arginine-Glycine-Glycine repeats
RNA	Ribonucleic Acid
RNAi	RNA Interference
RNAseq	RNA sequencing

RPMI	Roswell Park Memorial Institute Medium
RRE	Rev Response Element
RRPA	Random Reverse Primer Assay
RT	Reverse Transcriptase
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SAM	Sterile Alpha Motif
SAMHD1	SAM and HD Domain Containing Deoxynucleoside Triphosphate Triphosphohydrolase 1
SDS	Sodium Dodecyl Sulfate
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SERINC	Serine Incorporator
SC35	Serine Arginine Rich Splicing Factor 35
SEM	Standard Error of the Mean
SHAPE	Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension
SIV	Simian Immunodeficiency Virus
SLSA1	Stem-Loop Containing Splice Acceptor A1
SLX1, SLX4	Structure Specific Endonuclease Subunits
SLX4com	Structure-Specific Endonuclease Complex
SMN	Survival Motor Neuron Gene
snRNP	Small Nuclear Ribonucleoproteins
SR Proteins	Serine Arginine Protein Family
SRA	Sequence Read Archive Database
SRE	Splicing Regulatory Element
SRp40	Serine and Arginine Rich Splicing Factor 5
SSC	Side Scatter
SX	Small Exon
T	Thymine
TAR	Trans-Activation Response Element
Tat	Viral Trans-Activator of Transcription

TBE	Tris Borate EDTA Buffer
TCEP	Tris(2-Carboxyethyl)Phosphine
TET	Ten-Eleven Translocation Methylcytosine Dioxygenase
T/F	Transmitted/Founder
TLR4	Toll Like Receptor 4
U2AF65	Splicing Factor 65 kDa Subunit
UNG2	Uracil DNA Glycosylase
Vif	Viral Infectivity Factor
Vpr	Viral Protein R
Vpu	Viral Protein U
Vpx	Viral Protein X
VSV-G	Vesicular Stomatitis Virus
WT	Wild Type

CHAPTER 1: INTRODUCTION

Human Immunodeficiency Virus

The human immunodeficiency virus is the cause of the current AIDS pandemic. There are two virus types, HIV-1 and HIV-2, both resulting from multiple cross-species transmission events of simian immunodeficiency viruses (SIV) (1). HIV-1 is the primary and most virulent of the two types and the one studied in this thesis.

HIV-1 is a retrovirus and like all retroviruses, it has a positive single-stranded RNA genome. Two copies of this genome are packaged into a viral protein capsid and surrounded by a membrane derived from the host cell. HIV-1 targets CD4+ cells and the initial step of HIV-1 infection is recognition of the host cell CD4 receptor by a trimer of the viral envelope (Env) surface antigen. After binding the CD4 receptor Env becomes competent to bind a co-receptor, either CCR5 or CXCR4, and undergoes a conformational change that fuses the viral membrane with the host cell membrane and releases the viral contents into the cell (2). After entry the viral capsid largely disassembles, allowing the viral reverse transcriptase to make a double-stranded DNA copy from the genomic RNA template. The viral integrase protein (IN) forms a complex with the ends of the linear viral DNA and inserts it into the host cell genome. High levels of transcription of viral RNA begin after integration, leading to high levels of viral protein production.

There are eleven HIV-1 proteins. The Gag precursor polyprotein is cleaved by the HIV-1 protease (PR) to form the components of the capsid structure with Env trimers studding the membrane surface of the particle. The Integrase (IN) protein and reverse transcriptase (RT) are cleaved from the C-terminal end of the Gag-Pro-Pol polyprotein (3). The other viral proteins – Vpu, Vif, Vpr, Tat, Rev, and Nef – are regulatory proteins. Only one type of viral transcript is made – a full-length positive strand that can be used as genomic RNA of progeny viruses or as mRNA for synthesis of Gag and Gag-Pro-Pol. Other viral transcripts are made by alternative splicing of this full-length transcript (4, 5).

The viral proteins have multiple known functions and ongoing studies continue to discover and characterize additional functions (6). Protein function is linked to the regulation of its expression and alternative splicing is a major feature of gene regulation for HIV-1.

Env and Vpu have a common transcript. Although the Vpu open reading frame is upstream of Env, leaky or discontinuous ribosomal scanning may bypass the Vpu start codon and continue to the Env start codon (7). Vpu stands for “viral protein U” (8) and it functions to downregulate CD4 from the cell surface and degrade it in the endoplasmic reticulum (ER). This may facilitate viral particle production by preventing Env from binding to CD4 in the ER, directing it instead to virus particles (9). Vpu also functions to downregulate BST-2 (also known as tetherin), a cellular membrane-associated protein that stalls release of viral particles. In the absence of Vpu, tetherin inhibits viral particle release by tethering the budding virus to the host cell membrane (10).

Viral infectivity factor (Vif) counteracts the cellular restriction factor APOBEC3G (A3G) and 3F and is indispensable for viral replication in cells expressing these proteins. A3G and A3F proteins are packaged in HIV-1 virions and cause mutations during reverse transcription. Vif targets A3G/F for proteosomal degradation (6). In the absence of Vif, the viral genome becomes hypermutated (11, 12).

Viral Protein R (Vpr) has several known functions. It induces cell cycle arrest in the G2 stage (13). Vpr interacts with the ubiquitin ligase complex to target cellular proteins for degradation and is involved in the nuclear import of the pre-integration complex (14). Vpr-defective virus also has a transcription defect (15).

The HIV-1 Tat protein is required for efficient elongation of viral transcripts. It recruits the transcription factor P-TEFb from a cellular storage complex and brings it to the TAR loop, a region of viral RNA secondary structure. This Tat/P-TEFb/TAR complex interacts with the elongating RNA polymerase to greatly increase processivity (4). Viral transcripts abort prematurely without sufficient Tat (16) and it is thought that this may be associated with viral latency (17).

Generally, only completely spliced transcripts are transported from the nucleus to the cytoplasm (18). HIV-1 unspliced and partially spliced transcripts retain all or one intron respectively and would normally be targeted for degradation. The viral protein Rev connects these incompletely spliced RNAs to the Crm1 nuclear export pathway. Rev binds both a highly structured RNA sequence called the Rev

response element (RRE) and Crm1 and in this way transcripts with retained introns are exported to the cytoplasm (19-21).

Negative factor (Nef) modulates the immune response of the infected cell to accommodate viral replication by downregulating CD4 and MHC Class I from the cell surface (22, 23). Nef also suppresses the apoptotic response and counteracts the anti-viral effect of SERINCs, a family of host proteins (24) that are included in viral particles and block entry into the cell (25).

Viral assembly takes place at the cellular membrane. Multiple domains of Gag target it to the cell membrane, bind two copies of the viral genome, and make numerous protein-protein interactions. These proteins are incorporated into the viral capsid and include the viral reverse transcriptase, Vif, and integrase. Trimers of envelope proteins are embedded in the cell membrane so as to be on the outer surface of the virion as it buds from the cell (3).

Splicing

Most eukaryotic genes contain short coding sequences (exons) interspersed with long non-coding sequences (introns). In 1977 reports of splicing in adenovirus were published (26, 27) and soon splicing of discontinuous genes was recognized as a common feature across eukaryotic species (28). After a pre-mRNA is transcribed it must be completely processed before it can be exported to the cytoplasm and translated. This processing includes adding a poly A sequence at the 3' end, attaching a 5' cap, and splicing to excise the introns and join the exons.

The terms exon and intron can be confusing so it helps to know where these names originated. Some parts of the pre-mRNA were ex<u>ex</u>pressed in protein so they were called exons. Exons were flanked by <u>in</u>tervening sequences, hence called introns. Exons are typically 100-200 bases in length. The average length of a vertebrate exon is 137 bases with less than 1% of exons longer than 400 bases (29). Introns are longer – generally 1000-2000 bases although some are much longer. The 5' end of an intron is called the donor and the 3' end is the acceptor. Splicing cuts out the sequence between the donor and acceptor.

Correct splicing requires recognition of the splice sites. Every intron has three core splicing signal consensus sequences that are recognized by the splicing machinery. These are the 5' donor sequence (recognized by U1), and the 3' acceptor sequence and branch point/poly pyrimidine tract

sequences (both recognized by U2). The 5' and 3' consensus sequences are located at each end of the intron (Figure 1.1). The branch point and polypyrimidine tract (PPT) sequences are about 20 bases long and begin about 25 to 30 bases upstream of the 3' acceptor site. Almost without exception intronic sequences begin with GU and end with AG and the branch point is an A. HIV-1 is an exception and has some non-adenosine branch points (30, 31). The terminal exons obviously need a different set of recognition signals. For the 5' terminal exon, the 7-methyl-guanosine cap may be essential for exon recognition, as the cap and its binding proteins are required for splicing in experimental systems (32). The 3' terminal exon may be recognized by polyadenylation. Splicing and polyadenylation are coupled and splicing of an upstream intron upregulates polyadenylation but mutation of the final 3' splice site interferes with the polyadenylation process (33).

Figure 1.1. Splicing Signal Sequences.

Splicing requires two transesterification reactions catalyzed by the spliceosome, a huge cellular splicing machine (34). This two-step process is catalyzed by a set of five ribonucleoproteins called snRNPs (35), numbered U1, U2, U4, U5, and U6. These snRNPs and a large number of associated proteins combine to create the spliceosome. A proteomic analysis was done to identify spliceosomal proteins and 145 distinct proteins were identified (36), although the actual number may be even larger. A collective review of splicing factor analyses lists almost 300 proteins, making the spliceosome the biggest macromolecular machine in the cell (37, 38).

The first step of splicing is the binding through complementary base-pairing of snRNPs U1 to the 5'splice donor site and of U2 to the 3' acceptor site and branch point sequence. U2 pairs with the branch point sequence is such a way as to cause the branch point adenosine to bulge out and expose its 2' hydroxyl group in preparation for nucleophilic attack. These weak base pairings are stabilized by other protein factors, and in particular by the cellular SR proteins. These U1 and U2 binding events form the E

complex (Figure 1.2) (35, 39, 40) and are often depicted as U1 and U2 binding to opposite ends of an intron. In most eukaryotic genes, however, U1 and U2 actually pair up on opposite ends of an exon. This is called exon definition (29).

Figure 1.2. Spliceosome E Complex.

In the next step, spliceosome proteins assist U1 and U2 to interact across the intron to form the A complex (Figure 1.3). The transition from exon definition to intron definition is not as yet understood. This U1-U2 interaction puts the bulged-out branch point A into proximity with the 5' donor site (35, 39, 40).

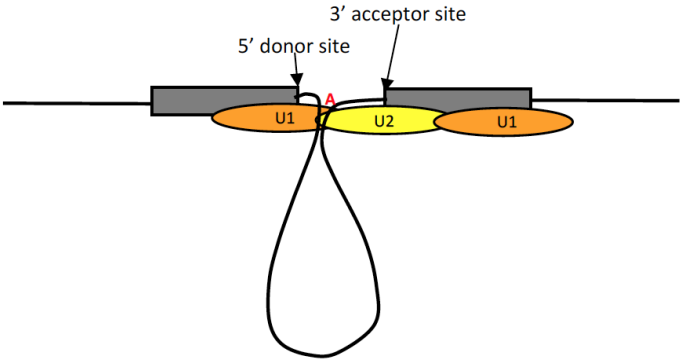

Figure 1.3. Spliceosome A Complex.

At this point a trimer of the remaining three snRNPs - U4, U5, and U6 – is recruited to the pre-spliceosome to form the B complex (Figure 1.4). A joint structure containing U4 and U6 binds to U2, and U5 interacts with both of the exons about to be spliced, perhaps holding them in proximity for the catalytic step ahead (41).

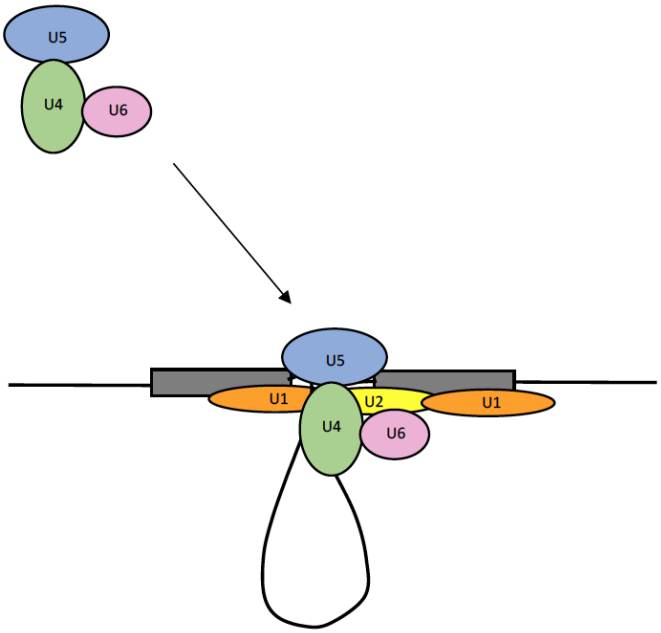

Figure 1.4. Spliceosome B Complex.

The B complex undergoes rearrangement (42). Assisted by helicases, the snRNPs remodel their base pairings. U6 unwinds from U4 and replaces U1 binding at the 5' splice site while remaining bound to U2 at the 3'splice site. U1 and U4 are released. U5 continues to bind both exons. This set of rearrangements remodels U2 and U6 into a catalytically active structure (43, 44) and the first splicing reaction occurs, joining the branch point A to the 5' splice site in a lariat formation. This stage of the spliceosome is called complex B* (Figure 1.5).

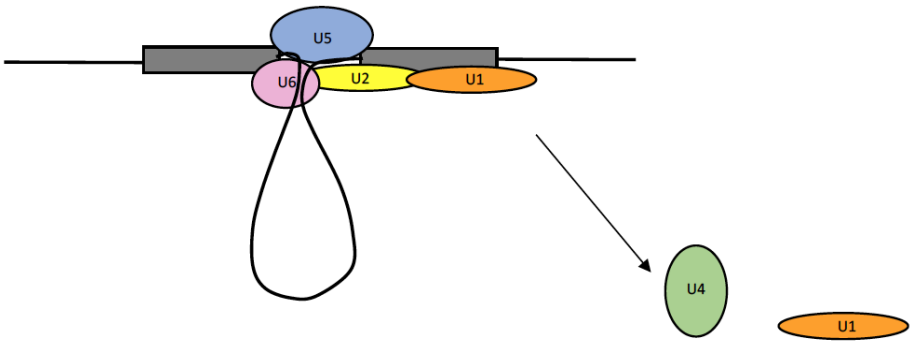

Figure 1.5. Spliceosome B* Complex

The resulting partially spliced products form complex C (Figure 1.6). Another set of rearrangements leads to catalysis of the second splicing nucleophilic attack –the 3'hydroxyl at the free 5' donor splice site attacks the phosphate at the end of the intron (the 3' splice site), displacing the lariat and connecting the exons (35, 39, 40).

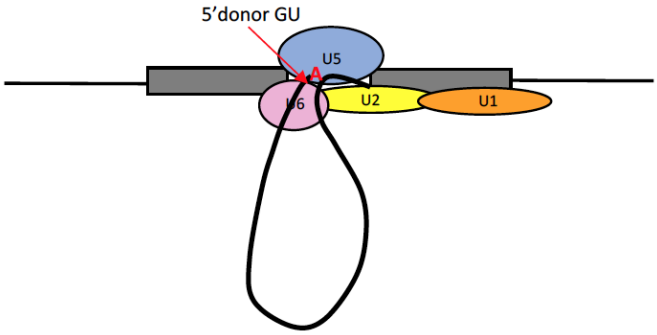

Figure 1.6. Spliceosome C Complex.

Once splicing is completed the snRNPs dissociate from the spliced exons to be recycled and the lariat is rapidly degraded (Figure 1.7).

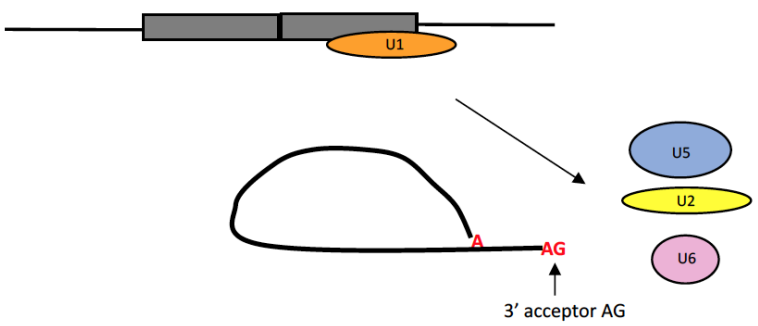

Figure 1.7. Post-Splicing Products.

Splicing requires ATP. The multiple remodeling steps involve a collection of core splicing factors, many of them ATP-dependent RNA helicases (37). The stages of spliceosome assembly are reversible *in vitro*, and this may give the splicing machinery the ability to form, release, and reform on optimal or correct sites to increase splicing fidelity (45, 46).

The spliceosome does not processively travel down the pre-mRNA, but appears to be formed de novo for each intron excision, meaning that a transcript with multiple introns will require a new spliceosome to assemble for each splicing event (47). The spliceosome can self-organize *in vitro* and does so in the precise series of steps described above. These steps are well documented and agreed on (37, 39, 40, 48, 49).

Recent studies describe the structural conformations of the spliceosome intermediate and catalytic states (50-57). These structural studies have used *in vitro* one intron systems that block the splicing reaction at a specific point. The results are instructive but may not fully represent the situation *in vivo* where cellular transcripts have multiple introns and also undergo additional pre-mRNA processing.

Although the spliceosome is generally thought of as the “splicing machine,” work by the Sperling laboratory has found that in nuclei isolated in physiological conditions, spliceosomes are assembled into a tetrameric form called the supraspliceosome (58-61). A supraspliceosome consists of four individual spliceosomal subunits multimerized around a single pre-mRNA. Whereas studies involving individual spliceosomes indicate that a spliceosome forms on each intron/exon, catalyzes splicing, and then disassembles; studies of supraspliceosomes suggest that a supraspliceosome assembles in a 1:1 ratio to pre-mRNAs and that splicing of the entire transcript takes place within this supraspliceosome, regardless of the number of introns in the unspliced transcript (62). This would require some

movement/rearrangement of the transcript through the supraspliceosome structure - that mechanism has not yet been described. Although studies of individual spliceosomes detail a complex temporal association of the five splicing snRNPs, all five are continually associated with the supraspliceosome.

Additional factors involved in pre-mRNA processing have been found in the supraspliceosome structure. hnRNP and SR proteins control alternative splicing and are both found in the supraspliceosome. Additionally, Drosha and its cofactor DGCR8, components of pre-miRNA processing, have also been found in the supraspliceosome, suggesting that miRNA processing and splicing may be connected. mRNAs must be capped at the 5' end and polyadenylated at the 3' end, and components required for end processing have also been found in supraspliceosomes (63). ADAR editing is another component of pre-mRNA processing. ADAR proteins make A to I edits in double stranded RNA. Supraspliceosomes were found to contain both ADAR1 and ADAR2 (64). The collective presence of all of these pre-mRNA processing elements suggests that the supraspliceosome is a cellular machine designed to carry out multiple pre-mRNA processing steps in a common location (65). Much work has been done to characterize the structure of the supraspliceosome (40). The resulting structures suggest that a single supraspliceosome is associated with a single pre-mRNA and fragmenting the transcript results in dissolution of the supraspliceosome into four separate spliceosomes, which can be reassembled by adding an intact transcript. This hints that the pre-mRNA itself is an integral part of the supraspliceosome structure. Each spliceosome directs one exon to a central cavity of the supraspliceosome, bringing the exons into close proximity with the introns looped out around the exterior of the structure. Each spliceosome subunit can splice out the intron looped out around it, thus the supraspliceosome could splice out four introns at a time, though it is not known if splices occur in a 5' to 3' order within the supraspliceosome. It has been hypothesized that an exon that got included in the external intron loop would be skipped (61), one possible pathway for alternative splicing.

Splicing happens with high accuracy but the information in the core splicing signals is not sufficient to define splice junctions (66). Numerous false splicing signals exist in introns, sometimes even in exon-like pairs, and yet these cryptic sites are seldom used (67). An additional layer of control is needed to correctly identify genuine exons, so there are cis-acting splicing regulatory elements (or SREs). Splicing regulatory elements are short sequence elements of 5 to 20 bases that bind trans-acting

proteins. This helps identify valid splice sites and enhance or repress splicing. These SREs can occur in introns, where they are called intronic splicing suppressors (ISS) or intronic splicing enhancers (ISE). More often they are found in exons, called exonic splicing silencers (ESS) or exonic splicing enhancers (ESE) (68).

The two classes of cellular proteins that act as trans-factors are the SR proteins and the hnRNP proteins. SR proteins are a family of cellular proteins named for their serine-arginine rich motifs (RS motifs) (69). Although SR proteins are a related family of proteins, the heterogeneous nuclear ribonucleoproteins (hnRNP) are a heterogeneous group of RNA binding proteins that associate with pre-mRNAs in the nucleus, as the name implies. In general, SR proteins activate splicing and hnRNP proteins suppress splicing, but their activity is context dependent (70-72). This context-dependent behavior of SREs initially seems random but in fact follows some highly predictable, if complex, patterns and rules (73).

It's known that SR and hnRNP proteins bind to splicing regulatory elements (74) and available evidence suggests that this binding recruits other proteins, either to construct a spliceosome or to make a large enough protein concatenation to block splice sites. This idea is further supported by experiments that tether an hnRNP or SR protein near a splice site, producing the effect of an ESS or an ESE respectively (75-78). Efforts have been made to discover the binding motifs of hnRNP and SR proteins using SELEX (to identify optimal binding sequences (Tian, 1995 #196)) and computational approaches (79, 80) to predict SREs, including an HIV-1 specific screening algorithm (81). These approaches have met with variable success. The redundancy of the proteins and the degeneracy of binding motifs make it difficult to say for certain which protein or proteins control a particular splicing event.

Some SREs may act at a distance from the affected splice site. For example, hnRNP A1 binds specifically to SREs but also to RNA generally. Once bound, hnRNP H1 multimerizes along the RNA to occlude binding sites of SR proteins and spliceosome components, and also loops out to simultaneously bind to more distal SREs, which may explain how ESS elements outside the splicing area can suppress splicing (82, 83).

ESEs are important for exon recognition and constitutive exons and all exons have an ESE (84, 85). An ESE and an ESS in the same exon can competitively bind to SR or hnRNP proteins respectively to enhance or repress spliceosome formation across the exon (86).

There is evidence that cellular SREs are associated with RNA secondary structures, in particular the loops of stem-loop structures (89, 90), and that perturbation or stabilization of structural features may alter splicing by masking or unmasking splicing sites or regulatory elements (91). Secondary structures control some HIV-1 splicing events (92), as will be demonstrated in Chapter 2.

As with the core splicing signals, splicing motifs are highly overrepresented relative to valid usage. Splicing regulatory elements are short sequences and thus any given element will be present in far more locations than actual SREs exist. Presumably all these candidate sites are not recognized by trans-acting cellular factors, just as all pseudo-splice sites are not recognized. Splicing factors can recognize multiple sequence motifs, and a given sequence motif can be recognized by redundant splicing factors. Any one exon can be controlled by multiple SREs. The possibilities that feed into splicing decisions are numerous and complex, and many co-operating/competing factors play into the final splicing outcome (35). This suggests that some complex combinatorial code exists to selectively identify genuine splice sites distinct from the background noise, a code that as yet remains uncracked (68, 85, 87, 88).

Alternative Splicing

Alternative splicing is a mechanism in which exons can be used in different arrangements to give specific activities and functions to the resulting proteins (93, 94). There are various types of alternative splicing seen in cellular transcripts. The following figures illustrate the three types of alternative splicing also seen in HIV-1: intron retention, exon skipping, and multiple acceptor sites (68, 95).

A typical cellular pre-mRNA transcript is represented by Figure 1.8 - short exons interspersed with long introns. Exons are recognized by the splicing signal sequences on each end that identify donor and acceptor sites spaced at an appropriate exon length. Spliceosome elements assemble at both ends of the exon, the exons are brought together, the spliceosomes assemble across the introns, and splicing occurs.

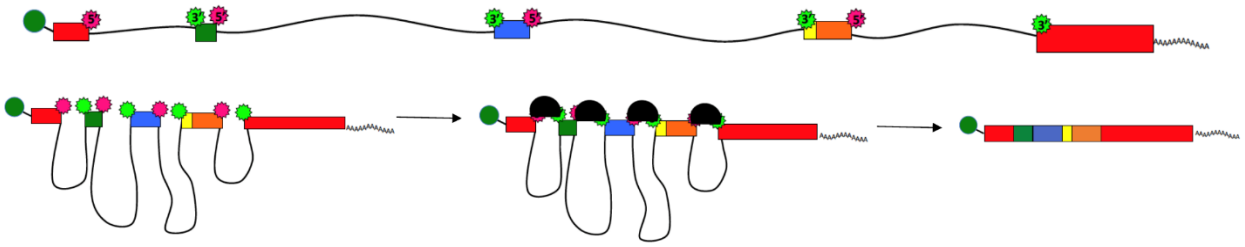

Figure 1.8. Exon Recognition and Inclusion.

It is interesting that in cellular transcripts, all the introns are spliced out – intron retention is a rare type of alternative splicing (Figure 1.9). In HIV-1, introns are retained and splicing is so heavily suppressed that the majority of HIV-1 transcripts remain completely unspliced. The mechanism for intron retention is not known.

Figure 1.9. Intron Retention.

Constitutive exons are present in every transcript but there are also alternate exons that are sometimes skipped. If the exon does not get defined, as shown in Figure 1.10, then it is considered part of the intron and is left out.

Figure 1.10. Undefined Exon Skipped.

Some exons have more than one possible 3' acceptor site, as illustrated by the orange and yellow exon in Figure 1.11. Depending on where the spliceosome components bind, the yellow sequence will be included or left out.

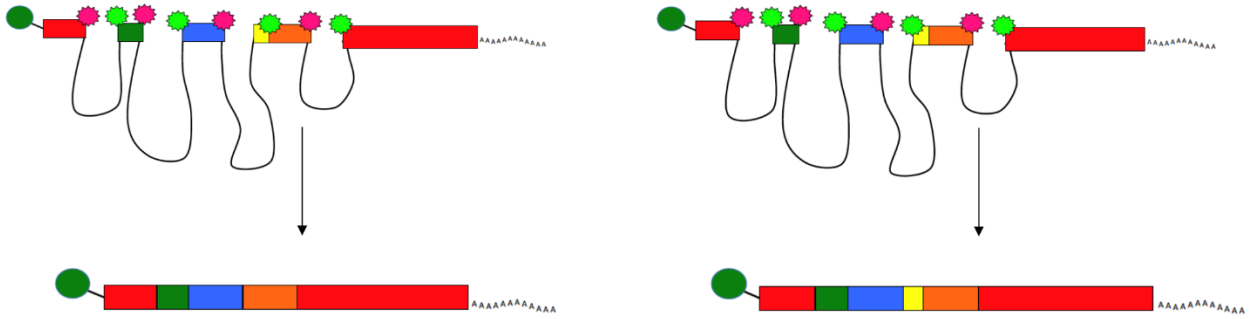

Figure 1.11. Alternative Acceptor Usage.

More than one-third of alternate splicing involves alternate exons, and more than one-fourth involves the use of alternate donor or acceptor splice sites (95). Other types of alternative splicing include multiple 5' donor sites, mutually exclusive exons, and alternative leader or terminal sequences (68).

SREs and the SR and hnRNP trans-acting proteins act competitively to control alternate splicing. A constitutive exon has the splice acceptor and donor signal sequences at each end of the exon and at least one exonic splicing enhancer (ESE) (Figure 1.12). This ESE is bound by one of the family of SR proteins. The SR proteins recruit and stabilize the formation of the early spliceosome spanning the exon (96). This defines the exon allowing it to be included. An alternate exon also has the splicing signal sequences at the exon boundaries, but it has two splicing regulatory elements - an enhancer (ESE) and a silencer (ESS). What happens depends on which protein gets there first. If an SR protein binds to the enhancer then the spliceosome begins to assemble, the exon is defined, and the exon gets included. However, if an hnRNP protein binds to the silencer first then spliceosome assembly is blocked and the exon isn't recognized and gets skipped. Thus every exon must have at least one enhancer and every alternative exon must also have a silencer.

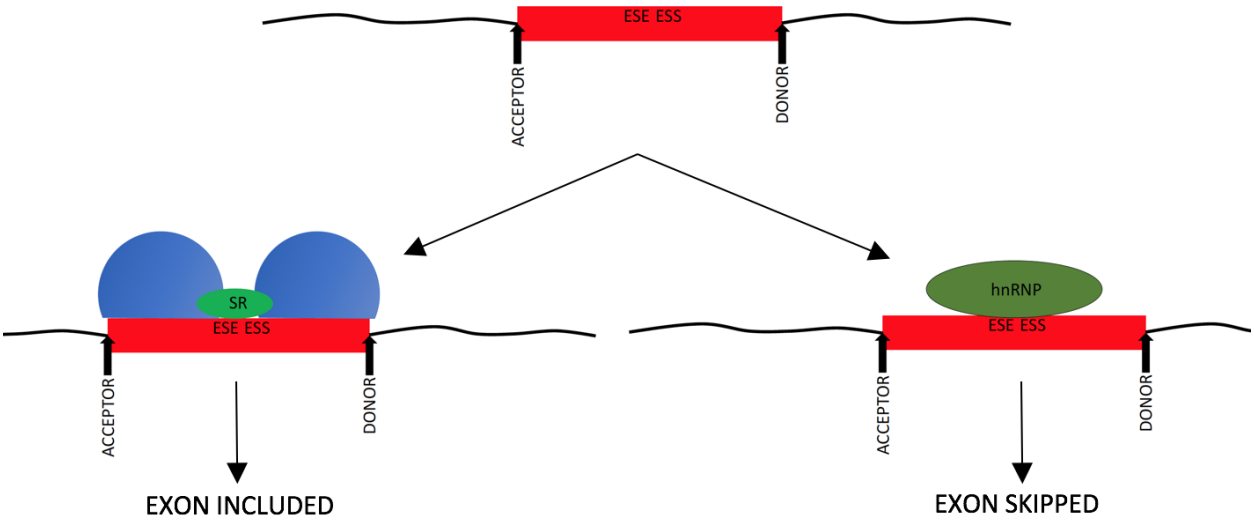

Figure 1.12. Inclusion or Skipping of an Alternative Exon.

Pseudo/cryptic splice sites typically have silencer elements, particularly those near a constitutive splice site. This may help ensure the usage of correct splice sites, and this type of silencer is seen in HIV-1 (76, 95, 97).

It is not always clear what function, if any, results from many alternatively spliced cellular transcripts. One interesting idea is that some transcripts may include exons that bind and sequester protein factors, making them unavailable (98). Most alternatively spliced transcripts are present at low levels. One evolutionary hypothesis is that these low levels give the cell a chance to try out a new protein isoform while maintaining adequate production of the previous isoform (83). Only about 10% of alternate splicing products have a clearly predictable functional effect but many would be predicted to alter protein surfaces and thereby alter interaction partners (95).

Some functional roles for alternative splicing have been discovered in development and immune response. One well-studied area of alternative splicing is its role in neuronal development (99). The mechanisms are not yet known, but the splicing differences are specific to cell types. Alternative splicing has also been documented in activated T-cells, the primary host cells for HIV-1 (100, 101). HIV-1 infection of macrophages was shown to change the balance of SR and hnRNP proteins over time (102).

Many alternatively spliced transcripts cause frame shifts and introduce a termination codon. This allows the cell to recognize and remove aberrant transcripts by a process called nonsense-mediated

decay (NMD) (103). It's been predicted that up to one-third of alternative splicing events introduce premature stop codons that subject them to NMD (104).

Changes to splicing can have pathological effects. A well-known example is mutation of the SMN (survival motor neuron) gene. A point mutation damages an ESE, causing an exon to be skipped. The shorter protein is rapidly degraded and this loss of function is the cause of spinal muscular atrophy (105). Interestingly, spinal muscular atrophy is one of the first diseases to be successfully treating by controlling a specific splicing event with an antisense oligonucleotide (106). Perhaps as many as half of genetic diseases are mutations that cause aberrant alternative splicing (95).

Splicing in HIV-1

Figure 1.13 is a simplified illustration of splicing in HIV-1. There are four donor sites and six acceptor sites. This figure is not drawn to scale – splice sites A1 through A5 actually occupy a 1 kilobase sequence in the middle of the genome. Only one mRNA type is transcribed – a full length genomic RNA, represented by the black bar. If this RNA is not spliced it serves as either a *gag/pro/pol* mRNA, or as genomic RNA (5, 107).

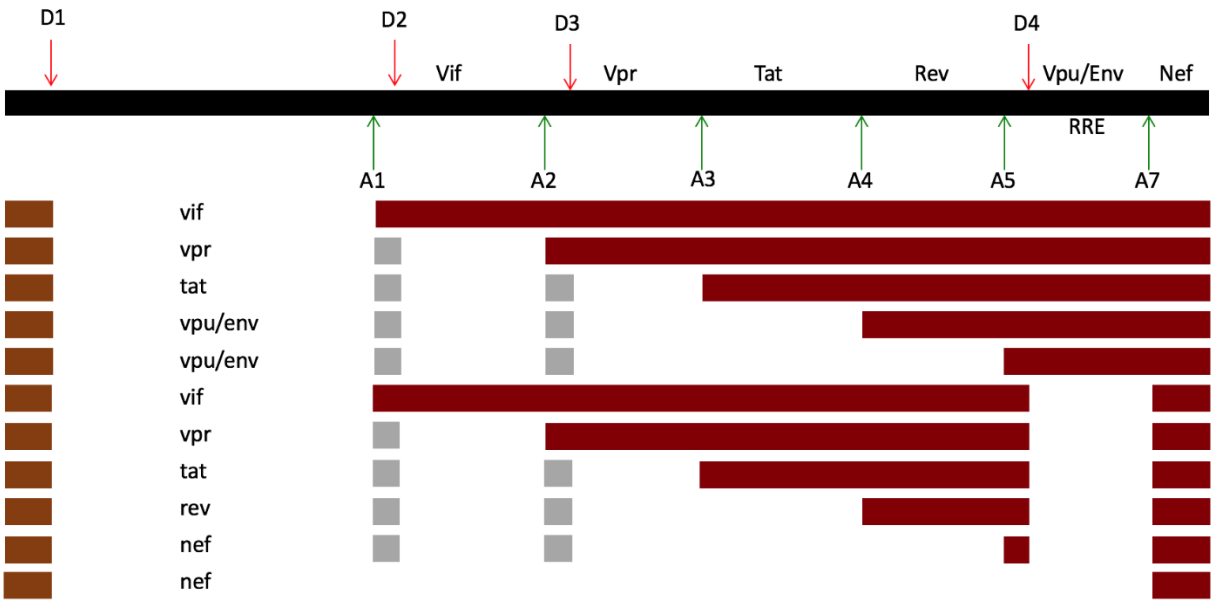

Figure 1.13. Splicing in HIV-1.

If HIV-1 mRNA splices there are two basic splicing events. The first event is when D1 is used to splice to one of the six downstream acceptor sites. The spliced transcript generally codes for the first open reading frame after the splice, so a splice from D1 to A1 creates an mRNA for Vif, a splice to A2 for Vpr, and Tat for A3. The transcript type for splicing to A4, and A5 is determined by the second splicing event.

The second major splicing event is from D4 to A7. This splice removes the *env* intron and the RRE, and joins together the two exons of *tat* and *rev*. All spliced transcripts first splice from D1 to an acceptor A1 through A5, and having done so, they may or may not splice out the D4 to A7 sequence. Spliced transcripts are divided into two size classes, depending on whether or not the D4-A7 splice happens. Retaining the *env* intron makes a larger class of transcripts collectively known as the 4 kb class or incompletely spliced transcripts, while transcripts that splice from D4 to A7 are called the 1.8 kb class or completely spliced transcripts. As discussed above, it is not usual for cellular transcripts to remain unspliced or to retain introns. HIV-1 unspliced or incompletely spliced transcripts keep all introns or the *env* intron respectively, and this prevents their nuclear export, so they need the RRE and Rev to connect to the Crm1 export pathway (19).

There are two additional donor sites, D2 and D3. Use of D2 or D3 creates small non-coding exons. The red bars in Figure 1.13 illustrate the constitutive exons in HIV-1 transcript types. The grey bars represent the small alternative exons created by use of D2 and/or D3. Either, both, or none can be included in a transcript. These small exons greatly increase the number of possible splice variants (5, 107). For each acceptor there are four possible types of transcripts with variable inclusion of the small exons, but they are not made in equal amounts. In general, direct splices, with no small exons, make up 75-85 % of transcripts. The exception is *nef*. The direct splice from D1 to A7 makes a *nef* transcript but this is an uncommon splicing event. The majority of *nef* transcripts splice to A5 (with or without including the two small upstream exons) and thus the region between A5 and D4 becomes a 3rd small exon.

The purpose of the small exons has thus far eluded discovery. It's been proposed that D2 and D3 exist to provide exon definition for A1 and A2 respectively to ensure adequate levels of Vif and Vpr, though there is not yet sufficient evidence to support this claim (108). Another possible but not yet studied purpose is that binding of U1 to D2 and D3 may stabilize *vif* and *vpr* transcripts (108), as has

already been observed for D4 and *env* transcripts (described below). One study claimed the first small exon stimulated protein production while the second small exon reduced it, with the improbable claim that this effect was activated after transcription but before splicing (when the small exons come into existence) (109). This notion was refuted in a later study that claimed the small exons do not affect transcript stability or protein production (108). Small exon inclusion does not change any known coding sequences or create a new ORF, yet the small exons are well conserved across HIV and SIV.

It is of interest that certain splicing events have not been seen. One such event is the inclusion of the small exons in *nef* transcripts that splice to A7. A second more interesting omission is transcripts that fail to splice at D1 but later splice from D4 to A7. This transcript type has never yet been reported in otherwise fairly comprehensive splicing analyses (5, 107), and I have not found it in quantities above what might be caused by PCR artifacts. This and other data presented later suggest that there are mechanisms to suppress certain types of HIV-1 splicing early and completely.

Splicing of cellular genes is an accurate process but splicing in HIV-1 is much less precise. Splicing is primarily from the major donors to acceptors but there are a number of lesser used sites, most often in close proximity to major sites as illustrated in Figure 1.14. This diagram shows some of the splice sites commonly found across different HIV-1 strains but each strain also has its individual minor splice sites, sometimes leading to idiosyncratic splicing and protein products (110-112) and even a novel transcript size class (107).

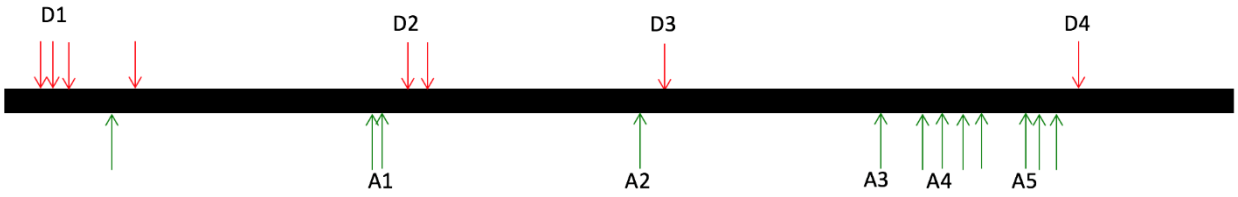

Figure 1.14. Major and Minor HIV-1 Splice Sites.

Even though splicing in HIV-1 does some things that cellular transcripts do not do, it is helpful to think of HIV from an exon/ intron perspective (Figure 1.15). The two small exons are alternative exons, thus they would be expected to have both enhancers and silencers. The exon defined by D4 is

constitutive but it uses alternate acceptors, so some control elements regulating acceptor usage would be predicted.

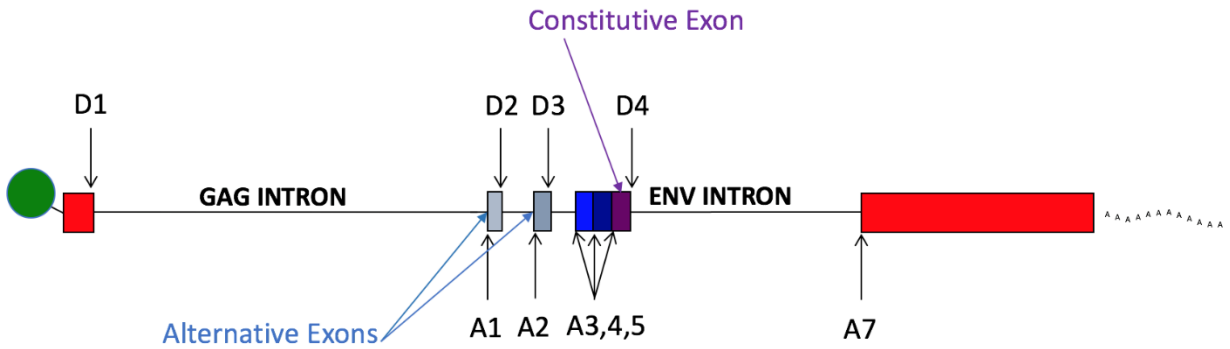

Figure 1.15. HIV-1 from and Exon/Intron Perspective.

HIV-1 donor sites D1 and D4 are efficient, D2 and D3 less so (108). HIV-1 acceptor sites are described as "weak" compared to the cellular splicing consensus sequence; nevertheless, they must still be suppressed to achieve balanced splicing (113-115). It is certain that all known HIV-1 SREs are not yet discovered and characterized but Figure 1.16 shows those that have been described to date.

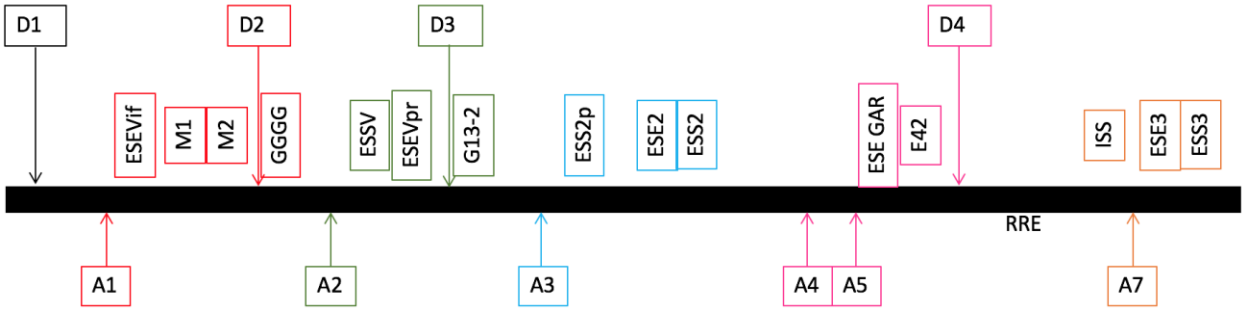

Figure 1.16. Known HIV-1 SREs.

D1 is the major splice donor and is used in every spliced transcript, yet it is the least understood. Studies that look for regulatory elements near D1 are confounded by other proximal structural and functional sequences. Presumably this would account for the high conservation of the D1 splice recognition sequences. Perturbations of splicing at D1 activate cryptic sites (5, 116) and prevent viral replication, but thus far credible sites of specific regulatory elements have not been published though the importance of the regional secondary structure is known (72, 77, 117-121).

There are numerous SREs in and immediately downstream of the first small exon from A1 to D2. This exon is only 50 bases long, and such short exons are particularly likely to be skipped, possibly due to steric hindrance between the two portions of the spliceosome that must form on each end. Such tiny exons may need additional enhancers to be included (29). Three ESEs have been characterized between A1 and D2: ESEVif (122), and M1 and M2- repeats of the same enhancer sequence (123). A GGGG sequence just downstream of D2 suppresses D2 use (122). Additional elements have been found downstream of D2 that suppress splicing from a cryptic donor site (97, 124). The strength of splice site D2 also controls splicing, and mutations that strengthen or weaken its ability to bind U1 snRNP increase or decrease recognition of the exon and thereby splicing to A1 (108, 125). Attempts have been made to identify the SR and hnRNP binding proteins that bind HIV-1 SREs and confer splicing control. ESEvif was found to bind SRp75, M1 and M2 bind ASF/SF2, but the binding protein for GGGG is not yet known.

The second small exon between A2 and D3 has an enhancer, ESEVpr (126-128) and a silencer ESSV. ESSV binds the hnRNP A/B proteins and prevents binding of U2AF65 to the viral PPT, preventing U2AF65 from stabilizing the binding of U2 and initiating spliceosome formation (126, 129). ESEVpr may bind the SR proteins Tra2 alpha/beta to support binding of U1 to D3 (128). Another poly-G run called G13-2 just downstream of D3 suppresses D3 and thereby splicing to A2. G13-2 binds hnRNP F/H and hnRNP A2/B1 (130), which is typical of G runs (131). As with D2 and the first small exon, mutations that enhance or reduce D3 complementarity to U1 snRNP increase or reduce splicing to A2 and inclusion of the second small exon. Mutations to ESSV or G13-2, or upregulation of D3 cause activation of splicing from D1 to A2 such that the number of unspliced transcripts is dramatically reduced. This situation is known as oversplicing and has a severe fitness cost. Interestingly, ESSV is not well conserved across HIV groups (127). Although splicing branch points are almost invariably A residues, the branch point for A2 is a G. This unusual branch point does not appear to make A2 a less effective acceptor (31).

Two SREs downstream of A3 have overlapping binding sites. ESE2 binds SC35 and ESS2 binds hnRNP A1 (132) to either up or down regulate splicing to A3 (115, 133-135). Both ESE2 and ESS2 are not immediately downstream of A3. A second suppressor even more distal to A3, ESS2p, binds hnRNP H (136). ESE2 and ESS2 are in a long stem-loop structure SLS3. NMR studies have shown that hnRNP A1 binds UAG triplets and forms a stable multimer on the SLS3 structure (137). A3 is in the exon defined

by D4 and yet little is known about the interaction of these two sites on each other. Like the silencer ESSV that represses use of acceptor A2, ESS2 is also not well conserved across HIV groups (138).

A4 has 3 or 4 subsites (depending on the strain) named A4a, A4b, A4c, and A4d and A5 has two subsites, A5 and A5b. All the A4 and A5 sites lie within a 45 base sequence and are alternative acceptors in the constitutive exon defined by D4 (see Figure 1.15). A4a, A4b, A5 and A5b are within a 26 base sequence. They share some common regulatory sequences. Their polypyrimidine tracts overlap and are short and interspersed with purines. There are eight branch points in the A4/5 region and they may be used by more than one acceptor. In the NL4-3 strain, the splice site A4b is also a branch point used by A5 (138, 139). An enhancer named ESE GAR (GAR for guanine/adenine rich) is located just downstream of D4 and binds to SF2/ASF and/or SRp40 to activate splicing to A4 and A5 sites. The SR protein used may influence the acceptor site that gets defined (140, 141). D4 is the donor involved in the second splicing event – removal or retention of the *env* intron. While binding of U1 does not always lead to splicing at D4, binding of U1 is necessary for the stability of the partially spliced *env* transcripts (140, 142), similar to U1 binding in the *gag-pro-pol* gene region that confers stability to *gag* transcripts, also mediated by SF2/ASF (143). E42 is the sequence between ESE GAR and D4. It is required for D4 exon definition but the specific sequence elements are not yet defined. In the partially spliced 4 kb transcript class, transcripts that splice to either an A4 or A5 acceptor make a *vpu/env* transcript. All of the A4 and A5 transcripts support Env protein production, but only transcripts that splice to A5 are efficiently translated into Vpu. The presence or absence of the small exons has no differential effect (144).

Although D4 is a strong donor, A7 is a weak acceptor and therefore it is A7 efficiency that determines removal or retention of the *env* intron (30, 113). Optimizing the A7 splicing signal sequence activates D4-A7 splicing even if D4 is weakened (123). The region surrounding A7 is highly structured and contains three SREs that control splicing to A7. The ISS, an intronic splicing silencer, is just upstream of A7 inside the *env* intron and overlaps with the A7 branch point (145-147). An enhancer and silencer, ESE3 and ESS3, are immediately downstream of A7 and also have partial overlap (134). ESS3 is composed of two silencing elements – ESS3a and ESS3b (148). SF2/ASF and hnRNP A1 control splicing by competitive binding to the ISS, ESE3 and ESS3 elements. hnRNP A1 has been shown to

bind to the specific ESS3a/b and ISS sequences (147, 149) and cooperatively multimerize to prevent binding of SF2/ASF to ESE3 (74, 145, 146).

Thesis Overview

Research is ongoing to correctly identify and characterize HIV-1 SREs, regions of regulatory RNA secondary structure, and the trans-acting SR and hnRNP factors that control splicing. This research requires methods to quantify splicing and splicing changes. Previous work has used artificial sub-genomic constructs to measure splicing. These sub-genomic constructs are transfected or used in cellular extracts and analyzed using agarose gels or Northern blots. As splicing and the effects of SREs are context and structure dependent (68), these artificial constructs, while useful in discovering SREs, may not accurately report their effects on splicing in the context of viral infection.

Chapter 2 describes a new deep sequencing assay I developed to quantify splicing in HIV-1. It combines Primer ID technology with Illumina paired-end deep sequencing to tally the number of all the different splice types in each of the two HIV-1 transcript size classes. This assay was used to measure splicing changes caused by changes in temperature and mutation of a secondary structure. Splicing was quantified in a panel of founder and transmitted viruses. The assay was adapted to measure splicing in SIVmac239. An additional set of programs was written to scan for cryptic splicing and trans-splicing was found. These experiments demonstrate the utility of this splicing assay.

The project in Chapter 3 was done in collaboration with the Bieniasz lab at The Rockefeller University. Blocks of synonymous mutations were made across the HIV-1 genome, some of which had serious replication defects. My part in the collaboration was to characterize and quantify changes in splicing and to identify new donor and acceptor sites in the original mutations and compensating revertant mutants. Previously unknown splicing control elements were discovered.

Chapter 4 is a collaboration with the Kutluay lab at Washington University and the Tolbert lab at Case Western University. This paper presents a study of binding proteins and their effects on HIV-1 splicing, with a particular emphasis on the role of hnRNP H1. My contribution quantified the effects of cellular splicing factor knock downs and the effect of mutations to hnRNP H1 binding sites.

Chapter 5 is a genetic screen of mutations between A5 and D4 and their effects on splicing both to upstream acceptors and to downstream splicing from D4 to A7.

Chapter 6 was done in collaboration with the Tolbert lab (Case Western University) and the Telesnitsky lab (University of Michigan). This chapter describes a new random reverse primer and deep sequencing assay used to distinguish fully spliced, partially splice, and unspliced transcripts. Mutations to a subset of known HIV-1 SREs were analyzed for splicing.

Chapter 7 summarizes the results of my splicing research and discusses directions for future research.

REFERENCES

1. Sharp PM, Hahn BH. 2011. Origins of HIV and the AIDS Pandemic. Cold Spring Harb Perspect Med 1.

2. Wilen CB, Tilton JC, Doms RW. 2012. HIV: cell binding and entry. Cold Spring Harb Perspect Med 2.

3. Sundquist WI, Krausslich HG. 2012. HIV-1 assembly, budding, and maturation. Cold Spring Harb Perspect Med 2:a006924.

4. Karn J, Stoltzfus CM. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med 2:a006916.

5. Purcell DF, Martin MA. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67:6365-6378.

6. Malim MH, Bieniasz PD. 2012. HIV Restriction Factors and Mechanisms of Evasion. Cold Spring Harb Perspect Med 2:a006940.

7. Schwartz S, Felber BK, Fenyo EM, Pavlakis GN. 1990. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. J Virol 64:5448-5456.

8. Cohen EA, Terwilliger EF, Sodroski JG, Haseltine WA. 1988. Identification of a protein encoded by the vpu gene of HIV-1. Nature 334:532-534.

9. Estrabaud E, Le Rouzic E, Lopez-Verges S, Morel M, Belaidouni N, Benarous R, Transy C, Berlioz-Torrent C, Margottin-Goguet F. 2007. Regulated degradation of the HIV-1 Vpu protein through a betaTrCP-independent pathway limits the release of viral particles. PLoS Pathog 3:e104.

10. Neil SJ, Zang T, Bieniasz PD. 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature 451:425-430.

11. Sheehy AM, Gaddis NC, Choi JD, Malim MH. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418:646-650.

12. Lecossier D, Bouchonnet F, Clavel F, Hance AJ. 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science 300:1112.

13. He J, Choe S, Walker R, Di Marzio P, Morgan DO, Landau NR. 1995. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J Virol 69:6705-6711.

14. Vodicka MA, Koepp DM, Silver PA, Emerman M. 1998. HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection. Genes Dev 12:175-185.

15. Miller CM, Akiyama H, Agosto LM, Emery A, Ettinger CR, Swanstrom RI, Henderson AJ, Gummuluru S. 2017. Virion-Associated Vpr Alleviates a Postintegration Block to HIV-1 Infection of Dendritic Cells. J Virol 91.

16. Kao SY, Calman AF, Luciw PA, Peterlin BM. 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature 330:489-493.

17. Lassen KG, Bailey JR, Siliciano RF. 2004. Analysis of human immunodeficiency virus type 1 transcriptional elongation in resting CD4+ T cells in vivo. J Virol 78:9105-9114.

18. Jacob AG, Smith CW. 2017. Intron retention as a component of regulated gene expression programs. Hum Genet doi:10.1007/s00439-017-1791-x.

19. Malim MH, McCarn DF, Tiley LS, Cullen BR. 1991. Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. J Virol 65:4248-4254.

20. Hope TJ. 1999. The ins and outs of HIV Rev. Arch Biochem Biophys 365:186-191.

21. Hammarskjold MH, Rekosh D. 2011. A long-awaited structure is rev-ealed. Viruses 3:484-492.

22. Harris M. 1999. HIV: a new role for Nef in the spread of HIV. Curr Biol 9:R459-461.

23. Azad AA. 2000. Could Nef and Vpr proteins contribute to disease progression by promoting depletion of bystander cells and prolonged survival of HIV-infected cells? Biochem Biophys Res Commun 267:677-685.

24. Usami Y, Wu Y, Gottlinger HG. 2015. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. Nature 526:218-223.

25. Trautz B, Wiedemann H, Luchtenborg C, Pierini V, Kranich J, Glass B, Krausslich HG, Brocker T, Pizzato M, Ruggieri A, Brugger B, Fackler OT. 2017. The host-cell restriction factor SERINC5 restricts HIV-1 infectivity without altering the lipid composition and organization of viral particles. J Biol Chem 292:13702-13713.

26. Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell 12:1-8.

27. Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A 74:3171-3175.

28. Flint SJ. 1979. Spliced viral messenger RNA. Am Sci 67:300-311.

29. Berget SM. 1995. Exon recognition in vertebrate splicing. J Biol Chem 270:2411-2414.

30. Dyhr-Mikkelsen H, Kjems J. 1995. Inefficient spliceosome assembly and abnormal branch site selection in splicing of an HIV-1 transcript in vitro. J Biol Chem 270:24060-24066.

31. Damier L, Domenjoud L, Branlant C. 1997. The D1-A2 and D2-A2 pairs of splice sites from human immunodeficiency virus type 1 are highly efficient in vitro, in spite of an unusual branch site. Biochem Biophys Res Commun 237:182-187.

32. Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, Mattaj IW. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell 78:657-668.

33. Niwa M, Rose SD, Berget SM. 1990. In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes Dev 4:1552-1559.

34. Brody E, Abelson J. 1985. The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science 228:963-967.

35. Matera AG, Wang Z. 2014. A day in the life of the spliceosome. Nat Rev Mol Cell Biol 15:108-121.

36. Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. Nature 419:182-185.

37. Jurica MS, Moore MJ. 2003. Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 12:5-14.

38. Nilsen TW. 2003. The spliceosome: the most complex macromolecular machine in the cell? Bioessays 25:1147-1149.

39. Rino J, Carmo-Fonseca M. 2009. The spliceosome: a self-organized macromolecular machine in the nucleus? Trends Cell Biol 19:375-384.

40. Sperling R. 2017. The nuts and bolts of the endogenous spliceosome. Wiley Interdiscip Rev RNA 8.

41. Madhani HD, Guthrie C. 1994. Dynamic RNA-RNA interactions in the spliceosome. Annu Rev Genet 28:1-26.

42. Makarov EM, Makarova OV, Urlaub H, Gentzel M, Will CL, Wilm M, Luhrmann R. 2002. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science 298:2205-2208.

43. Valadkhan S, Mohammadi A, Jaladat Y, Geisler S. 2009. Protein-free small nuclear RNAs catalyze a two-step splicing reaction. Proc Natl Acad Sci U S A 106:11901-11906.

44. Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. 2013. RNA catalyses nuclear pre-mRNA splicing. Nature 503:229-234.

45. Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, Ramirez N, Cornish VW, Gelles J, Moore MJ. 2011. Ordered and dynamic assembly of single spliceosomes. Science 331:1289-1295.

46. Yang F, Wang XY, Zhang ZM, Pu J, Fan YJ, Zhou J, Query CC, Xu YZ. 2013. Splicing proofreading at 5' splice sites by ATPase Prp28p. Nucleic Acids Res 41:4660-4670.

47. Staley JP, Guthrie C. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 92:315-326.

48. Smith DJ, Query CC, Konarska MM. 2008. "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. Mol Cell 30:657-666.

49. Valadkhan S. 2007. The spliceosome: caught in a web of shifting interactions. Curr Opin Struct Biol 17:310-315.

50. Galej WP, Wilkinson ME, Fica SM, Oubridge C, Newman AJ, Nagai K. 2016. Cryo-EM structure of the spliceosome immediately after branching. Nature 537:197-201.

51. Kosmyna B, Query CC. 2016. Structural biology: Catalytic spliceosome captured. Nature 537:175-176.

52. Wan R, Yan C, Bai R, Huang G, Shi Y. 2016. Structure of a yeast catalytic step I spliceosome at 3.4 A resolution. Science 353:895-904.

53. Wan R, Yan C, Bai R, Wang L, Huang M, Wong CC, Shi Y. 2016. The 3.8 A structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. Science 351:466-475.

54. Yan C, Wan R, Bai R, Huang G, Shi Y. 2016. Structure of a yeast activated spliceosome at 3.5 A resolution. Science 353:904-911.

55. Bertram K, Agafonov DE, Liu WT, Dybkov O, Will CL, Hartmuth K, Urlaub H, Kastner B, Stark H, Luhrmann R. 2017. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature 542:318-323.

56. Fica SM, Oubridge C, Galej WP, Wilkinson ME, Bai XC, Newman AJ, Nagai K. 2017. Structure of a spliceosome remodelled for exon ligation. Nature 542:377-380.

57. Yan C, Wan R, Bai R, Huang G, Shi Y. 2017. Structure of a yeast step II catalytically activated spliceosome. Science 355:149-155.

58. Muller S, Wolpensinger B, Angenitzki M, Engel A, Sperling J, Sperling R. 1998. A supraspliceosome model for large nuclear ribonucleoprotein particles based on mass determinations by scanning transmission electron microscopy. J Mol Biol 283:383-394.

59. Cohen-Krausz S, Sperling R, Sperling J. 2007. Exploring the architecture of the intact supraspliceosome using electron microscopy. J Mol Biol 368:319-327.

60. Sperling J, Azubel M, Sperling R. 2008. Structure and function of the Pre-mRNA splicing machine. Structure 16:1605-1615.

61. Sperling J, Sperling R. 2017. Structural studies of the endogenous spliceosome - The supraspliceosome. Methods doi:10.1016/j.ymeth.2017.04.005.

62. Azubel M, Habib N, Sperling R, Sperling J. 2006. Native spliceosomes assemble with pre-mRNA to form supraspliceosomes. J Mol Biol 356:955-966.

63. Raitskin O, Angenitzki M, Sperling J, Sperling R. 2002. Large nuclear RNP particles--the nuclear pre-mRNA processing machine. J Struct Biol 140:123-130.

64. Raitskin O, Cho DS, Sperling J, Nishikura K, Sperling R. 2001. RNA editing activity is associated with splicing factors in lnRNP particles: The nuclear pre-mRNA processing machinery. Proc Natl Acad Sci U S A 98:6571-6576.

65. Shefer K, Sperling J, Sperling R. 2014. The Supraspliceosome - A Multi-Task Machine for Regulated Pre-mRNA Processing in the Cell Nucleus. Comput Struct Biotechnol J 11:113-122.

66. Lim LP, Burge CB. 2001. A computational analysis of sequence features involved in recognition of short introns. Proc Natl Acad Sci U S A 98:11193-11198.

67. Sun H, Chasin LA. 2000. Multiple splicing defects in an intronic false exon. Mol Cell Biol 20:6414-6425.

68. Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. Rna 14:802-813.

69. Graveley BR. 2000. Sorting out the complexity of SR protein functions. Rna 6:1197-1211.

70. McCullough AJ, Berget SM. 1997. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol Cell Biol 17:4562-4571.

71. Chen CD, Kobayashi R, Helfman DM. 1999. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. Genes Dev 13:593-606.

72. Mueller N, Klaver B, Berkhout B, Das AT. 2015. Human immunodeficiency virus type 1 splicing at the major splice donor site is controlled by highly conserved RNA sequence and structural elements. J Gen Virol 96:3389-3395.

73. Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. Nature 444:580-586.

74. Mayeda A, Screaton GR, Chandler SD, Fu XD, Krainer AR. 1999. Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol Cell Biol 19:1853-1863.

75. Pongoski J, Asai K, Cochrane A. 2002. Positive and negative modulation of human immunodeficiency virus type 1 Rev function by cis and trans regulators of viral RNA splicing. J Virol 76:5108-5120.

76. Wang Z, Xiao X, Van Nostrand E, Burge CB. 2006. General and specific functions of exonic splicing silencers in splicing control. Mol Cell 23:61-70.

77. Erkelenz S, Mueller WF, Evans MS, Busch A, Schoneweis K, Hertel KJ, Schaal H. 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. Rna 19:96-102.

78. Hillebrand F, Peter JO, Brillen AL, Otte M, Schaal H, Erkelenz S. 2017. Differential hnRNP D isoform incorporation may confer plasticity to the ESSV-mediated repressive state across HIV-1 exon 3. Biochim Biophys Acta 1860:205-217.

79. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31:3568-3571.

80. Zhang XH, Leslie CS, Chasin LA. 2005. Computational searches for splicing signals. Methods 37:292-305.

81. Erkelenz S, Theiss S, Otte M, Widera M, Peter JO, Schaal H. 2014. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic Acids Res 42:10681-10697.

82. Zhu J, Mayeda A, Krainer AR. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell 8:1351-1361.

83. Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72:291-336.

84. Schaal TD, Maniatis T. 1999. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol 19:261-273.

85. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing enhancers in human genes. Science 297:1007-1013.

86. Pozzoli U, Sironi M. 2005. Silencers regulate both constitutive and alternative splicing events in mammals. Cell Mol Life Sci 62:1579-1604.

87. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. 2004. Systematic identification and analysis of exonic splicing silencers. Cell 119:831-845.

88. Chasin LA. 2007. Searching for splicing motifs. Adv Exp Med Biol 623:85-106.

89. Hiller M, Zhang Z, Backofen R, Stamm S. 2007. Pre-mRNA secondary structures influence exon recognition. PLoS Genet 3:e204.

90. Warf MB, Berglund JA. 2007. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. Rna 13:2238-2251.

91. Donahue CP, Muratore C, Wu JY, Kosik KS, Wolfe MS. 2006. Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. J Biol Chem 281:23302-23306.

92. Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM, 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 Pathog 9:e1003294.

93. Kopelman NM, Lancet D, Yanai I. 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat Genet 37:588-589.

94. Su Z, Wang J, Yu J, Huang X, Gu X. 2006. Evolution of alternative splicing after gene duplication. Genome Res 16:182-189.

95. Blencowe BJ. 2006. Alternative splicing: new insights from global analyses. Cell 126:37-47.

96. Shen H, Kan JL, Green MR. 2004. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell 13:367-376.

97. Brillen AL, Walotka L, Hillebrand F, Muller L, Widera M, Theiss S, Schaal H. 2017. Analysis of competing HIV-1 splice donor sites uncovers a tight cluster of splicing regulatory elements within exon 2/2b. J Virol doi:10.1128/jvi.00389-17.

98. Kanadia RN, Urbinati CR, Crusselle VJ, Luo D, Lee YJ, Harrison JK, Oh SP, Swanson MS. 2003. Developmental expression of mouse muscleblind genes Mbnl1, Mbnl2 and Mbnl3. Gene Expr Patterns 3:459-462.

99. Liu J, Geng A, Wu X, Lin RJ, Lu Q. 2017. Alternative RNA Splicing Associated With Mammalian Neuronal Differentiation. Cereb Cortex doi:10.1093/cercor/bhx160:1-7.

100. Ip JY, Tong A, Pan Q, Topp JD, Blencowe BJ, Lynch KW. 2007. Global analysis of alternative splicing during T-cell activation. Rna 13:563-572.

101. Martinez NM, Pan Q, Cole BS, Yarosh CA, Babcock GA, Heyd F, Zhu W, Ajith S, Blencowe BJ, Lynch KW. 2012. Alternative splicing networks regulated by signaling in human T cells. Rna 18:1029-1040.

102. Dowling D, Nasr-Esfahani S, Tan CH, O'Brien K, Howard JL, Jans DA, Purcell DF, Stoltzfus CM, Sonza S. 2008. HIV-1 infection induces changes in expression of cellular splicing factors that regulate alternative viral splicing and virus production in macrophages. Retrovirology 5:18.

103. Baker KE, Parker R. 2004. Nonsense-mediated mRNA decay: terminating erroneous gene expression. Curr Opin Cell Biol 16:293-299.

104. McGlincy NJ, Smith CW. 2008. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends Biochem Sci 33:385-393.

105. Laird AS, Mackovski N, Rinkwitz S, Becker TS, Giacomotto J. 2016. Tissue-specific models of spinal muscular atrophy confirm a critical role of SMN in motor neurons from embryonic to adult stages. Hum Mol Genet 25:1728-1738.

106. Ottesen EW. 2017. ISS-N1 makes the First FDA-approved Drug for Spinal Muscular Atrophy. Transl Neurosci 8:1-6.

107. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res 40:10345-10355.

108. Madsen JM, Stoltzfus CM. 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.

109. Krummheuer J, Lenz C, Kammler S, Scheid A, Schaal H. 2001. Influence of the small leader exons 2 and 3 on human immunodeficiency virus type 1 gene expression. Virology 286:276-289.

110. Benko DM, Schwartz S, Pavlakis GN, Felber BK. 1990. A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. J Virol 64:2505-2518.

111. Caputi M, Zahler AM. 2002. SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. Embo j 21:845-855.

112. Wentz MP, Moore BE, Cloyd MW, Berget SM, Donehower LA. 1997. A naturally arising mutation of a potential silencer of exon splicing in human immunodeficiency virus type 1 induces dominant aberrant splicing and arrests virus production. J Virol 71:8542-8551.

113. Staffa A, Cochrane A. 1994. The tat/rev intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3' splice site. J Virol 68:3071-3079.

114. O'Reilly MM, McNally MT, Beemon KL. 1995. Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. Virology 213:373-385.

115. Zahler AM, Damgaard CK, Kjems J, Caputi M. 2004. SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. J Biol Chem 279:10077-10084.

116. Borg KT, Favaro JP, Arrigo SJ, Schmidt M. 1999. Activation of a cryptic splice donor in human immunodeficiency virus type-1. J Biomed Sci 6:45-52.

117. Abbink TE, Berkhout B. 2008. RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. J Virol 82:3090-3098.

118. Lenz C, Scheid A, Schaal H. 1997. Exon 1 leader sequences downstream of U5 are important for efficient human immunodeficiency virus type 1 gene expression. J Virol 71:2757-2764.

119. Asang C, Erkelenz S, Schaal H. 2012. The HIV-1 major splice donor D1 is activated by splicing enhancer elements within the leader region and the p17-inhibitory sequence. Virology 432:133-145.

120. Mueller N, van Bel N, Berkhout B, Das AT. 2014. HIV-1 splicing at the major splice donor site is restricted by RNA structure. Virology 468-470:609-620.

121. Mueller N, Berkhout B, Das AT. 2015. HIV-1 splicing is controlled by local RNA structure and binding of splicing regulatory proteins at the major 5' splice site. J Gen Virol 96:1906-1917.

122. Exline CM, Feng Z, Stoltzfus CM. 2008. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol 82:3921-3931.

123. Kammler S, Otte M, Hauber I, Kjems J, Hauber J, Schaal H. 2006. The strength of the HIV-1 3' splice sites affects Rev function. Retrovirology 3:89.

124. Widera M, Erkelenz S, Hillebrand F, Krikoni A, Widera D, Kaisers W, Deenen R, Gombert M, Dellen R, Pfeiffer T, Kaltschmidt B, Munk C, Bosch V, Kohrer K, Schaal H. 2013. An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA. J Virol 87:2707-2720.

125. Mandal D, Exline CM, Feng Z, Stoltzfus CM. 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. J Virol 83:6067-6078.

126. Bilodeau PS, Domsic JK, Mayeda A, Krainer AR, Stoltzfus CM. 2001. RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element. J Virol 75:8487-8497.

127. Madsen JM, Stoltzfus CM. 2005. An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. J Virol 79:10478-10486.

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

129. Domsic JK, Wang Y, Mayeda A, Krainer AR, Stoltzfus CM. 2003. Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts. Mol Cell Biol 23:8762-8772.

130. Widera M, Hillebrand F, Erkelenz S, Vasudevan AA, Munk C, Schaal H. 2014. A functional conserved intronic G run in HIV-1 intron 3 is critical to counteract APOBEC3G-mediated host restriction. Retrovirology 11:72.

131. Schaub MC, Lopez SR, Caputi M. 2007. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. J Biol Chem 282:13617-13626.

132. Del Gatto-Konczak F, Olive M, Gesnel MC, Breathnach R. 1999. hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. Mol Cell Biol 19:251-260.

133. Amendt BA, Hesslein D, Chang LJ, Stoltzfus CM. 1994. Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. Mol Cell Biol 14:3960-3970.

134. Amendt BA, Si ZH, Stoltzfus CM. 1995. Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. Mol Cell Biol 15:4606-4615.

135. Si Z, Amendt BA, Stoltzfus CM. 1997. Splicing efficiency of human immunodeficiency virus type 1 tat RNA is determined by both a suboptimal 3' splice site and a 10 nucleotide exon splicing silencer element located within tat exon 2. Nucleic Acids Res 25:861-867.

136. Jacquenet S, Mereau A, Bilodeau PS, Damier L, Stoltzfus CM, Branlant C. 2001. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. J Biol Chem 276:40464-40475.

137. Hallay H, Locker N, Ayadi L, Ropers D, Guittet E, Branlant C. 2006. Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site. J Biol Chem 281:37159-37174.

138. Bilodeau PS, Domsic JK, Stoltzfus CM. 1999. Splicing regulatory elements within tat exon 2 of human immunodeficiency virus type 1 (HIV-1) are characteristic of group M but not group O HIV-1 strains. J Virol 73:9764-9772.

139. Swanson AK, Stoltzfus CM. 1998. Overlapping cis sites used for splicing of HIV-1 env/nef and rev mRNAs. J Biol Chem 273:34551-34557.

140. Kammler S, Leurs C, Freund M, Krummheuer J, Seidel K, Tange TO, Lund MK, Kjems J, Scheid A, Schaal H. 2001. The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. Rna 7:421-434.

141. Caputi M, Freund M, Kammler S, Asang C, Schaal H. 2004. A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. J Virol 78:6517-6526.

142. Lu XB, Heimer J, Rekosh D, Hammarskjold ML. 1990. U1 small nuclear RNA plays a direct role in the formation of a rev-regulated human immunodeficiency virus env mRNA that remains unspliced. Proc Natl Acad Sci U S A 87:7598-7602.

143. Lutzelberger M, Reinert LS, Das AT, Berkhout B, Kjems J. 2006. A novel splice donor site in the gag-pol gene is required for HIV-1 RNA stability. J Biol Chem 281:18644-18651.

144. Anderson JL, Johnson AT, Howard JL, Purcell DF. 2007. Both linear and discontinuous ribosome scanning are used for translation initiation from bicistronic human immunodeficiency virus type 1 env mRNAs. J Virol 81:4664-4676.

145. Damgaard CK, Tange TO, Kjems J. 2002. hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. Rna 8:1401-1415.

146. Marchand V, Mereau A, Jacquenet S, Thomas D, Mougin A, Gattoni R, Stevenin J, Branlant C. 2002. A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding. J Mol Biol 323:629-652.

147. Levengood JD, Rollins C, Mishler CH, Johnson CA, Miner G, Rajan P, Znosko BM, Tolbert BS. 2012. Solution structure of the HIV-1 exon splicing silencer 3. J Mol Biol 415:680-698.

148. Si ZH, Rauch D, Stoltzfus CM. 1998. The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. Mol Cell Biol 18:5404-5413.

149. Jain N, Morgan CE, Rife BD, Salemi M, Tolbert BS. 2016. Solution Structure of the HIV-1 Intron Splicing Silencer and Its Interactions with the UP1 Domain of Heterogeneous Nuclear Ribonucleoprotein (hnRNP) A1. J Biol Chem 291:2331-2344.

CHAPTER 2: CHARACTERIZING HIV-1 SPLICING USING NEXT GENERATION SEQUENCING[1]

Overview

Full-length HIV-1 RNA serves as the genome or as an mRNA, or this RNA undergoes splicing using four donors and ten acceptors to create over 50 physiologically relevant transcripts in two size classes (1.8 kb and 4 kb). We developed an assay using Primer ID-tagged deep sequencing to quantify HIV-1 splicing. Using the NL4-3 lab strain we found that A5 (*env*/*nef*) is the most commonly used acceptor (about 50%) with A3 (*tat*) the least used (about 3%). Two small exons are made when a splice to acceptor A1 or A2 is followed by activation of donor D2 or D3, and the high-level use of D2 and D3 dramatically reduces the amount of *vif* and *vpr* transcripts. We observed distinct patterns of temperature sensitivity of splicing to acceptors A1 and A2. In addition, disruption of a conserved structure proximal to A1 caused a ten-fold reduction in all transcripts that utilized A1. Analysis of a panel of subtype B transmitted/founder viruses showed that splicing patterns are conserved, but with surprising variability of usage. A subtype C isolate was similar, while an SIV isolate showed significant differences. We also observed trans-splicing from a downstream donor on one transcript to an upstream acceptor on a different transcript which we detected in 0.3% of 1.8 kb RNA reads. There were several examples of splicing suppression when the *env* intron was retained in the 4 kb size class. These results demonstrate the utility of this assay and identify new examples of HIV-1 splicing regulation.

Importance

During HIV-1 replication over 50 conserved spliced RNA variants are generated. The splicing assay described here uses new developments in deep sequencing technology combined with Primer ID-tagged cDNA primers to efficiently quantify HIV-1 splicing at a depth that allows even low frequency splice variants to be monitored. We have used this assay to examine several features of HIV-1 splicing

[1] This chapter was previously published in the Journal of Virology. The original citation is as follows:
Emery A, Zhou S, Pollom E, Swanstrom R. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. Journal of virology. 2017;91(6). Epub 2017/01/13. doi: 10.1128/jvi.02515-16. PubMed PMID: 28077653; PubMed Central PMCID: PMCPMC5331825.

and to identify new examples of different mechanisms of regulation of these splicing patterns. This splicing assay can be used to explore in detail how HIV-1 splicing is regulated and, with moderate throughput, could be used to screen for structural elements, small molecules, and host factors that alter these relatively conserved splicing patterns.

Introduction

HIV-1 is a plus-stranded RNA virus that replicates through an obligatory integrated DNA intermediate. The virus encodes 10 genes and uses several different strategies to allow expression of these genes, given that eukaryotes typically translate only the first open reading frame (ORF) of an mRNA. Unspliced full-length transcripts code for the Gag and Gag-Pro-Pol polyproteins. The remaining seven viral proteins are translated from over 50 mRNA variants generated by the host cell splicing machinery (1).

HIV-1 splicing has been extensively studied, and the splice variants and splice regulatory sequences are well characterized (2, 3), mostly using the laboratory strain NL4-3. Approximately one-half of the initial full-length transcripts remain unspliced and serve as either genomic RNA or as mRNA for the Gag and Gag-Pro-Pol polyproteins (Figure 2.1). The remaining half of the full-length transcripts are spliced (1). All spliced RNAs use donor site 1 (D1) in the 5' noncoding region, and one of ten downstream acceptor sites (A1, A2, A3, A4a, A4b, A4c, A4d, A5, A5b, A7) (2, 4). About half of the spliced RNAs also undergo further splicing to remove the D4 – A7 *env* intron (2). These smaller RNAs are referred to as the completely spliced or 1.8 kb size class. Spliced RNAs that retain the D4 – A7 intron are known as the incompletely spliced or 4 kb size class. Completely spliced transcripts are exported from the nucleus using cellular pathways; however, the incompletely spliced and full-length transcripts require a viral protein, Rev, for nuclear export. The D4-A7 intron contains the Rev binding sequence, the RRE (Rev Response Element). With increasing concentrations of Rev the expression pattern is altered by allowing incompletely spliced RNAs to exit the nucleus (5, 6). As a rule, the major splice products are produced from a direct splicing event between D1 and one of the acceptor sites (Figure 2.1). The ensuing transcript is generally translated from the first ORF downstream of this direct splice junction. Thus D1-A1 splicing makes a *vif* transcript, D1-A2 a *vpr* transcript, and so on. The fact that D1 is engaged as a splice donor in only a fraction of the full-length transcripts, that D1 splicing must singly

engage a number of different downstream splice acceptor sites, and that splicing of the D4-A7 intron is itself regulated all point to a highly orchestrated pattern of splicing to allow adequate expression of each of the viral proteins.

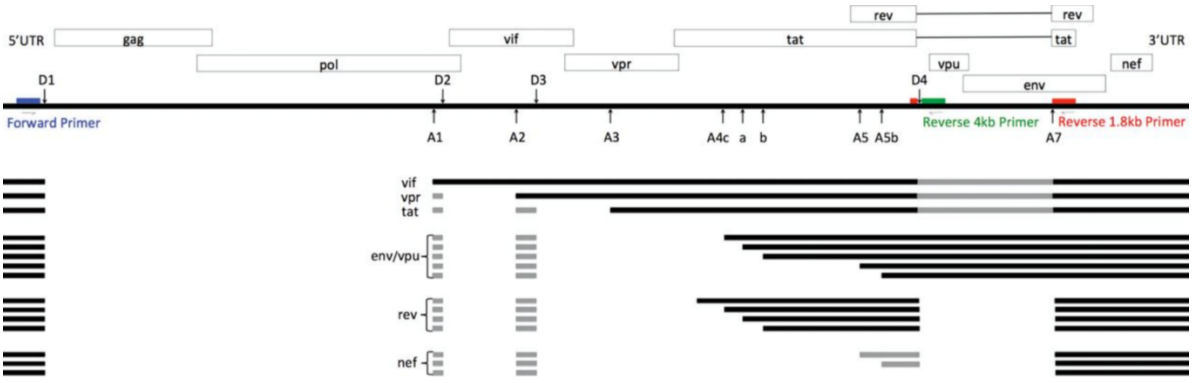

Figure 2.1. HIV-1 Splice Patterns and Primer Locations.
Blue = forward primer. Red = 1.8 kb class Primer ID tagged reverse primer. Green = 4 kb class Primer ID tagged reverse primer. Grey boxes are small exons or sequences that may or may not be present in the respective transcripts. Adapted from Purcell (2).

Even greater splicing complexity comes from the use of two additional donor sites, D2 and D3, that create two small exons, A1-D2 and A2-D3, one or both of which may be included in splice variants for the downstream genes (2). The purpose of these small exons is unknown and they are present in only a small fraction of RNAs. They do not code for any known regulatory elements or alternative proteins, and mutations that prevent their use have shown contradictory results regarding fitness, transcript stability, and splicing effects (1, 7-10). Other studies have searched for regulatory sequence elements around the major splice acceptors, and different types of regulatory elements have been identified that control the utilization of splice sites (10-13). Sequence elements outside of the specific splice site sequences bind cellular factors that enhance or suppress donor or acceptor usage, often in a concerted or competitive manner (14, 15). Mutations in these elements have been reported to alter splicing (1, 2).

The major splicing elements of HIV-1 subtypes B and C were originally defined using assays that require RT-PCR amplification of splice variants followed by cloning and sequencing of clones. Another approach to assaying splicing has been to identify the RT-PCR products of the individual splice variants by size (2, 16). Studies assessing the effects of control element mutations on splicing have often been

done using incomplete viral RNAs, such as single intron mini-genes, simplifying the analysis but placing splicing outside of the context of the entire viral genome. More recently a deep sequencing approach used the PacBio platform with its long read capability to characterize HIV-1 splicing in extensive detail. This approach confirmed the previously described splice variants and discovered new ones (3).

Here we report a new assay to analyze HIV-1 splicing based on the Illumina MiSeq deep sequencing platform. We used the assay to quantify the splicing patterns within each mRNA size class in HIV-1_{NL4-3} infection and after transfection of an infectious clone. We quantified acceptor usage and the differential usage of minor splice donors D2 and D3 in the two mRNA size classes, one of several examples where splicing was differentially regulated between the two size classes. We observed that two silent mutations designed to disrupt a conserved element of secondary structure, SLSA1 (stem-loop containing splice acceptor A1), caused a 10-fold decrease in splicing to A1, linking this feature of secondary structure to regulation of A1 usage. We examined the stability of splicing as a function of temperature and observed two distinct patterns of temperature sensitivity, with splicing to A2 increased at the lower temperature but inhibited at the higher temperature while splicing to A1 was just inhibited at the higher temperature. We found the combined effects of the SLSA1 mutation and temperature were additive in their effects on A1. We also examined the extent to which complex splicing patterns are conserved by assaying splicing in a panel of transmitted/founder viruses, which showed that most features of HIV-1 splicing are highly conserved, although examples of variability in the frequency of use were seen. We adapted the assay to quantify splicing in a subtype C isolate and found high conservation compared to subtype B. Analysis of splicing in an SIV isolate revealed many conserved features but also significant differences. We detected trans-splicing between different transcripts and even between transcripts from different proviruses. Collectively these results provide greatly enhanced quantification of splicing in the context of viral infection and offer new insights into HIV-1 splicing and its regulation.

Materials and Methods

Viruses

A panel of full-length transmitted/founder (T/F) HIV-1 infectious molecular clones was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr. John Kappes (17-20). Clones are identified by the last two digits of their identification number. The infectious clone pNL4-3 was

obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr. Malcolm Martin (21), and the SLSA1 mutant virus was made as described in Pollom et al. (22). Virus was produced by transfecting 293T cells using 4 µg of a viral DNA clone/well of a 6 well plate for the SLSA1 experiment and 1 µg/well for the T/F viruses, using FuGene (Promega) transfection reagent as per the FuGene protocol database. At 48 hours after transfection, the medium was removed and filtered through a 45 µm filter, then frozen in aliquots at -80˚ C for later use. pZM247Fv2 was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr. Beatrice Hahn (23). A partial *env* deletion of this clone, pZM247v2∆env, was obtained from Dr. Sarah Joseph. VSV G-pseudotyped pZM247v2∆env virus was produced by cotransfecting 293T cells with 5 µg each of the pZM247v2∆env plasmid and a VSV G expression plasmid in a T75 flask, as per the FuGene protocol database, and harvested and filtered as above. A molecular clone of SIVmac239 was obtained from Dr. Ronald Desrosiers (24) and used to produce virus by transfection.

Cells

CEMx174 cells were used for HIV-1 infections. Cells were cultured in RPMI plus 10% fetal bovine serum (FBS) with added penicillin and streptomycin. For the SLSA1 mutant/temperature gradient experiment, 2 x 10^6 cells were pelleted and resuspended in 800 µl of medium with either the SLSA1 or NL4-3 virus and incubated for 45 minutes, then brought to a total volume of 50ml of RPMI with FBS. The cells were then incubated at 37˚ C for 4 days, until syncytia were evident. An aliquot of 10 ml of each culture (with cells) was put into each of 3 flasks and incubated for 6 hours at 33˚, 37˚, or 41˚ C, then the cells were harvested for RNA extraction. The T/F virus molecular clones were transfected into 293T cells plated the day before at 5 x 10^5 cells/well of a six well plate. Cells were harvested 48 hours post transfection for RNA extraction. 8E5 cells [8E5 cells were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, 8E5/LAV from Dr. Thomas Folks (25)], a clonal cell line producing virus from a single defective provirus, were super-infected with VSV G-pseudotyped subtype C pZM247v2∆env virus. 5 x 10^7 8E5 cells were pelleted and resuspended in 40 ml of medium with the pseudotyped virus and aliquots were harvested between 1 and 3 days post super-infection.

RNA extraction, cDNA synthesis, and Illumina MiSeq library preparation

Total cellular RNA was extracted using the RNeasy Plus Mini Kit (Qiagen), with the QIAshredder to homogenize cell lysates (Qiagen) and including 1% β-mercaptoethanol in the lysis buffer. Primers used for cDNA synthesis for NL4-3 and 8E5 cells, the T/F strains, pZM247v2Δenv, and SIVmac239 are shown in Tables 1 and 2. All primers were resuspended at a concentration of 10 mM. Indexed primers were obtained from Integrated DNA Technologies custom oligos, with hand mixing to optimize random base incorporation. For the SLSA1 and temperature experiments, 2 µg of total cellular RNA from each sample was mixed with 1 µl of 10 mM dNTPs, 0.5 µl of the indicated reverse primer, and dH_2O to 13 µl. The samples were heated to 65˚ C for 5 minutes, then held at 4˚ C. To each tube was added 1 µl of 0.1 M dithiothreitol, 1 µl RNaseOUT (ThermoFisher), 4 µl of 5x SuperScript III RT buffer, and 1 µl SuperScript III Reverse Transcriptase (ThermoFisher), then heated to 55˚ C for 1 hour followed by 70˚ C for 15 minutes. An aliquot of 1 µl Ribonuclease H (ThermoFisher) was added to each sample which was then incubated at 37˚ C for 20 minutes. cDNA synthesis for the T/F virus samples and SIVmac239 was the same except 4 µg of total cellular RNA was used. cDNA synthesis for the pZM247v2Δenv-infected 8E5 cells was done using two reverse primers for each size class, one specific to the 8E5 virus (the NL4-3 primers) and one specific to the pZM247Fv2 virus. cDNAs were purified and the Primer ID cDNA primer removed using Agencourt RNAClean XP beads (Beckman Coulter) as per protocol with a 2:1 ratio of beads to sample volume. Illumina libraries for the SLSA1 mutant and temperature gradient samples, VSV G-pseudotyped pZM247v2Δenv, and SIVmac239 were made using two successive rounds of PCR amplification. For the first PCR, 1 µg cDNA template, 1 µl dNTPs, 2.5 µl ADPT_2a reverse primer, 2.5 µl idx_fp forward primer, 0.4 µl KAPA Robust Hot Start Polymerase (KAPA Biosystems), 7.5 % DMSO, 10 µl KAPA A Buffer, 10 µl KAPA Enhancer 1, and dH_2O to 50 µl were combined and cycled with an initial denaturation at 95˚ C for 5 minutes, followed by 3 cycles each consisting of 95˚ for 30 seconds, progressively decreasing annealing temperatures of 64˚, 61˚, 58˚, 55˚, and 52˚ C for 15 seconds, and 72˚ C for 2 minutes, then 5 more cycles with an annealing temperature at 50° C. The final extension was 72˚ C for 10 minutes. PCR products were purified and primers removed using AMPure XP Beads (Beckman Coulter) as per protocol, using a 1.6:1 ratio of beads:PCR volume. An aliquot of 3 µl of the cleaned product from this first PCR was used as template for the second PCR, added to 10 µl KAPA A buffer, 1 µl

dNTPs, 2.5 µl UniAdpt forward primer, 2.5 µl IIIIndAdpt forward primer, 0.4 µl KAPA Robust Hot Start Polymerase, 5% DMSO, 10 µl KAPA Enhancer 1, and dH_2O to 50 µl. Cycling conditions were as for the first PCR but with annealing temperatures of 66˚, 63˚, and 60˚ C each for 3 cycles, followed by 25 cycles with an annealing temperature of 58˚ C. "N"s in the IIIIndAdpt forward primer are specific sequences that bar code each sample to allow multiplexing in the Illumina sequencing reaction. PCR products were visualized on a 2% agarose gel, then cleaned as before using AMPure XP Beads. T/F virus samples: 4 µg of total cellular RNA was used as input to the cDNA reactions as described above. cDNAs were forward tagged as follows. An aliquot of 1 µg of purified cDNA, 0.5 µl forward primer Iidx_fp_I (or the appropriate sample-specific forward primer), 1 µl dNTPs, and dH_2O to 14.8 µl were heated to 75˚ C and then cooled in 1˚ C increments for 30 seconds until reaching 45˚ C. To each sample was added 0.2 µl Phusion HF, 4 µl Phusion 5x buffer, and 1 µl DMSO. Samples were then heated to 45˚ C and the temperature was increased in 1 degree increments every 10 seconds until reaching 72˚ C, where it was held for 10 minutes. Following forward tagging, the samples were purified using DNA XP beads as described above but eluted in 23 µl. An aliquot of 21.1 µl of the purified tagged forward strand was input into a second PCR as above, with the following cycle changes: initial denaturation at 95˚ C for 5 minutes; then 3x each touchdown cycles using progressively cooler annealing temperatures – 95˚ C for 30 seconds, annealing at 64˚, 61˚, 58˚, 55˚, and 52˚ C for 15 seconds, extension for 2 minutes at 72˚ C; then 20 cycles with annealing temperature at 50˚ C; and a final extension at 72˚ C for 10 minutes.

Table 2.1. Primers Used in This Study.

Primer use and name	Sequence
SLSA1, temp, T/F NL4-3, 8E5 primers	
idx_fp	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAGNNNNTGCTGAAGCGCGCACGGCAAG
l_idx_4r	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)GTACTATAGGTTGCATTACATGTACTACTTAC
l_idx_1.8r	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)CAGTTCGGGATTGGGAGGTGGGTTGC
T/F 4-kb reverse primers	
l_idx_39r	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TAAAGGTTGCATTACATATATTACTTAC
l_idx_42r	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TAAAGATTGCATTACAGATGCTACTTAC
l_idx_43r	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TAAAGAGTTCATTACATGCACTACTCAC
l_idx_46r	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TAAAGTTTGCATTACAAGCACTACTTAC
l_idx_48r	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TAAAGCTTGCATTACATATACTACTTAC
T/F 1.8-kb reverse primers	
idx_18_43	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TCCCCTCTGGGCTGGGGGGCGGATTGC
idx_18_44	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TCGGGGCTGGGAAGCGGGTTGC
idx_18_45	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TCGGGGCTGGGAGGCGGGTTGC
T/F forward primers	
lidx_fp_l	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG(N×14)TGCTGAAGCGCGCACGGCAAG
TF_42_fp	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG(N×14)TGCTGAAGCGCGCACAGCAAG
TF_47_fp	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG(N×14)TGCTGAAGCGCGCGCGGCAAG
Subtype C pZM247Fv2	
idx_C3_4rp	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TTCAAGCATTACATCTACTATTTGGTACTTAC
idx_C3_18rp	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)TCGGGTCCCCTCGAGTTTGGGATAAGGGTTGC
idx_C3_fp	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAGNNNNTGCTGAAGTGCACGCAGCAAG
SIVmac239	
SIV_f_DO	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAGNNNNGCCCTGGGAGGTTCTCTCCA
SIV_f_D1	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAGNNNNAGGAAGAGGCCTCCGGTTGCA
SIV_rp_18	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)CGGGTCCTGTTGGATATGGGTTTG
SIV_rp_4k	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)ATTCCCAAGACATCCCATACTTAC
Random reverse	
uni_rp	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT(N×14)
Illumina library	
UniAdpt	AATGATACGGCGACCACCGAGATCTACACGCCTCCCTCGCGCCATCAGAGATGTG
IIIIndAdpt	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTC
ADPT_2a	GTGACTGGAGTTCAGACGTGTGCTC

Table 2.2. Primers Used for Each Sample.

Sample type or ID no.	Forward primer	Reverse 1.8-kb primer	Reverse 4-kb primer
SLSA1 and temp expt	idx_fp	idx_1.8rp	idx_4rp
T/F NL4-3	lidx_fp_l	l_idx_1.8r	l_idx_4r
40	lidx_fp_l	l_idx_1.8r	l_idx_39r
42	TF_42_fp	l_idx_1.8r	l_idx_42r
43	lidx_fp_l	idx_18_43	l_idx_43r
44	lidx_fp_l	idx_18_44	l_idx_4r
45	lidx_fp_l	idx_18_45	l_idx_4r
46	lidx_fp_l	l_idx_1.8r	l_idx_46r
47	TF_47_fp	l_idx_1.8r	l_idx_4r
48	lidx_fp_l	idx_18_45	l_idx_48r
56	lidx_fp_l	idx_18_45	l_idx_39r

Sequencing

Libraries were mixed/multiplexed and sequenced using the 300 base paired-end read Illumina Miseq platform. Reads were sorted using the Illumina bcl2fastq pipeline (v.1.8.4) to separate the multiplexed samples. All samples were loaded at a concentration of 8 pM with 15 % PhiX double-stranded DNA fragments to optimize cluster formation. Sequencing data files are available in the SRA database, BioProject ID: PRJNA324601. Because of quality control features that are intrinsic to each sequencing run, comparisons of relative abundance of spliced RNAs between samples are most reliably accomplished when the samples are multiplexed in the same sequencing run.

Splicing analysis

Data analysis and splicing quantification were done using an in-house pipeline written in Ruby. Using the combined data from the forward and reverse reads, this program identifies the specific splice variant for each paired-end read and bins the reads by Primer ID to prevent skewing from PCR resampling. The same sorting program was used for both the SLSA1 mutant and temperature gradient experiments. Separate programs were adapted to the specific primers and sequence of each T/F virus, as well as for pZM247Fv2 and SIVmac239. Rare 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. Specific rare acceptor site usage was quantified using a modification to the basic splicing identification program. These programs are available at https://github.com/SwanstromLab/SPLICING.

Results

Description of a new splicing assay

We developed a new HIV-1 splicing assay that combines deep sequencing with the previously developed Primer ID technology (26). In this assay, total cellular RNA from infected or transfected cells is extracted and used in two different cDNA reactions, each one specific to the two different viral mRNA size classes (1.8 kb and 4 kb). The 3' end of the cDNA primer used to capture the 4 kb size class of viral RNAs is placed near the 5' end of the D4 – A7 *env* intron (Figure 2.1, shown in green). This intron is removed from the 1.8 kb size class RNAs so this primer is specific for the 4 kb class. The cDNA primer for the 1.8 kb class is shown in red. The 3' end of this primer spans the D4 – A7 splice junction, which forms only in the 1.8 kb RNAs. Each of the cDNA primers includes a common sequence at its 5' end that serves as the sequence of the reverse primer in PCR. The forward PCR primer, shown in blue, is upstream of D1 and is used in the PCR amplification of both size classes.

The Primer ID portion of the cDNA primer is an internal random sequence tag in the primer that is used to identify and quantify individual viral RNA templates. The cDNA primers are synthesized with this internal random nucleotide sequence block. This creates random combinations of primer tags. Given a sufficient excess of random sequence tags relative to sample templates, any one viral RNA template will be primed with a unique random sequence in the cDNA primer that is then incorporated into the cDNA. The same sequence tag, present in the cDNA primer, will be present on all PCR products that come from this one viral RNA template. The cDNAs are then PCR amplified, made into an Illumina library and sequenced using Miseq paired-end sequencing. Each sample produces millions of sequence reads. In the Illumina paired-end method, each read contains approximately 300 bases starting from the forward primer, and similarly, a corresponding 300 base reverse sequence beginning with the Primer ID and reverse primer that is paired to the read from the other end. The combined information from the paired-end reads is sufficient to distinguish all splice variants except *nef* 1 (a direct splice from D1 to A7), which

lies outside of the primer range. The inclusion of Illumina indexed linkers allows multiplexing of multiple samples in a single sequencing run.

A sorting program first reads the forward sequence and identifies the donor and acceptor sites in the spliced RNA template, then does the same for the reverse read, and combines the information to give a pattern of splice junctions. These patterns are sorted according to Primer ID. All reads with the same Primer ID sequence are condensed into a single splice junction pattern and that pattern is matched to a known splice variant. Regardless of the number of times a specific Primer ID sequence occurs in the sequencing output, each unique Primer ID tag is counted as a single observation of an mRNA molecule. For the reported experiments the sequencing depth typically ranged from 100,000 to 500,000 unique Primer IDs indicating this number of independent mRNA templates queried, allowing the accurate quantification of the normal viral spliced RNAs and the reliable detection of rare or cryptic events.

Quantification of acceptor splice site usage in NL4-3

CEMx174 cells were infected with HIV-1_{NL4-3} and the cells grown in culture until syncytia formed, as evidence of a robust but asynchronous infection. Total cellular RNA was extracted and used as input for the sequencing and quantification of spliced viral mRNAs. An overview of the data is as follows. Essentially all spliced viral transcripts use D1 and splice to a downstream acceptor site. Although there are multiple A4 and A5 acceptors, they are grouped together (as A4 or A5) in the following figures. Relative usage of each acceptor for both size classes is shown in Figure 2.2A. The patterns are somewhat similar in both size classes, with *env* A5 as the most commonly used acceptor, and *tat* A3 the least used acceptor. A previous study suggested HIV-1 splicing is sequential (27) and the relative proportions in each size class suggests that the choice of upstream splice acceptor site differentially influences downstream splicing (i.e. of the D4-A7 intron). A transcript spliced from D1 to A1, A3, or A4 has an increased probability of splicing from D4 to A7, and so being a 1.8 kb transcript. In contrast, a splice between D1 and A2 does not affect the use of D4 and A7, while splicing from D1 to A5 skews toward suppressing the use of D4 and A7 thereby including the *env* intron (a 4 kb transcript). Thus the use of D4 and A7 appears to be differentially impacted by the choice of the upstream splice acceptor by D1.

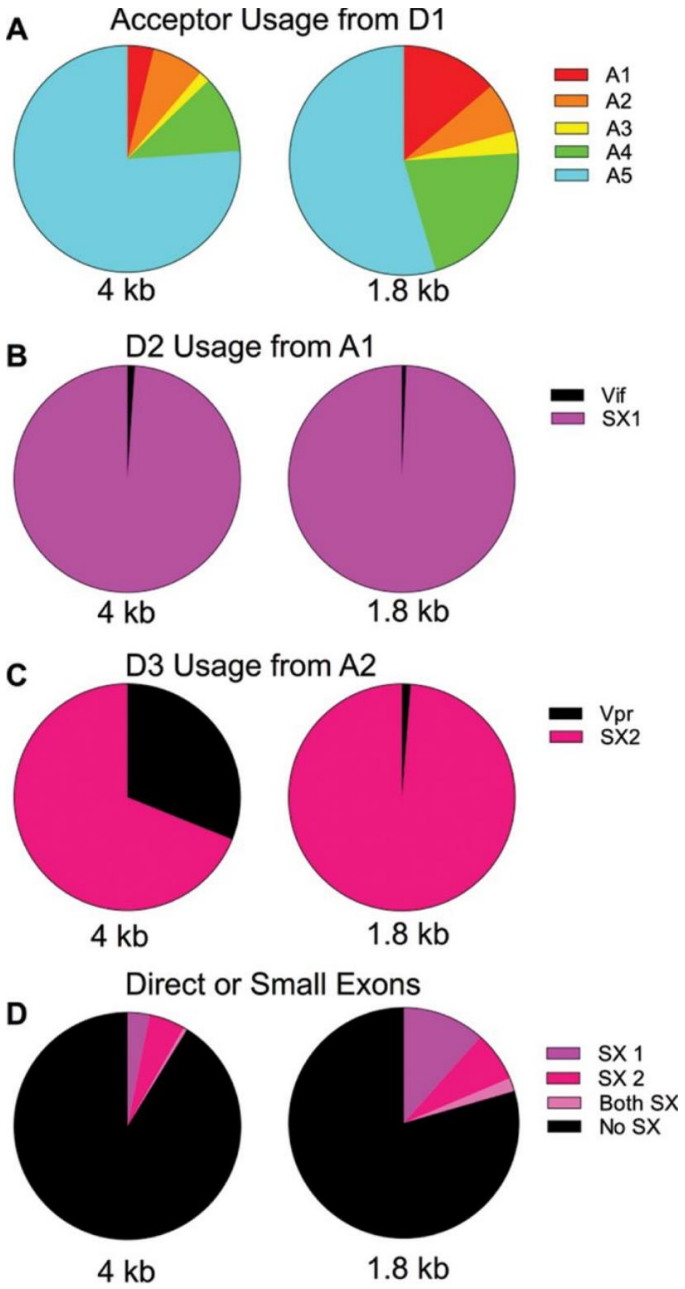

Figure 2.2 Quantification of HIV-1 Splicing Patterns.
A) Acceptor usage from D1 in the two size classes. Circles represent all transcripts in size class. B) Circles represent the total splices from D1 to A1. Shown are proportions that splice again at D2 and those that remain vif transcripts. C) Circles represent the total splices to A2, and the proportions that splice again at D3 compared to those that remain *vpr* transcripts. D) Circles represent all transcripts in size class. Shown are proportions of transcripts that contain either one, both, or no small exons. SX1 = small exon 1, SX2 = small exon 2.

Splicing events from D1 to A1 or A2 initially make a *vif* or *vpr* mRNA, respectively. However, a splice to A1 most often splices again by utilizing the nearby downstream donor D2 to create small exon 1 (SX1), with D2 splicing to one of the other downstream acceptor sites, A2-A5 (Figure 2.2B). Thus the

initial splicing to A1 results is a relatively low number of *vif* transcripts in both size classes, with the use of D2 reducing the potential number of *vif* transcripts by over 95%.

The splicing events from D1 to A2 have a different outcome in the two size classes with respect to the utilization of the adjacent D3 donor to generate small exon 2 (SX2) (Figure 2.2C). In the 1.8 kb class, most A2 splices utilize D3 to create the small exon 2 by splicing to A3-5; however, in the 4 kb class, 31% of splices to A2 remain as *vpr* transcripts. This suggests that D3 usage affects downstream D4 usage, such that the inclusion of small exon two (i.e. the use of D3) increases the probability of splicing from D4 to A7 with removal of the *env* intron. Alternatively, the same mechanism that suppresses D4 utilization to create the 4 kb *env* mRNA may also suppress utilization of D3.

Transcripts with either or both small exons are present at twice the frequency in 1.8 kb transcripts compared to 4 kb transcripts (Figure 2.2D), an observation that is partially driven by the disproportionate inclusion of small exon 2 in the 1.8 kb mRNAs (Figure 2.2C), but primarily driven by the increased use of A1 (and the small exon from A1-D2) in the 1.8kb mRNAs (Figure 2.2A and 2.2B). This suggests that inclusion of small exons up-regulates splicing from D4-A7. It is curious that in the vast majority of cases, splicing to A1/*vif* and A2/*vpr* are followed by subsequent splicing at D2 and D3, respectively; however, the majority of the A3-A5 transcripts are the result of a direct splice from D1 and do not include the small exons (Figure 2.2D). This suggests that the main function of D2 and D3 are to down regulate the expression of *vif* and *vpr* beyond what can be accomplished through differential splicing among the splice acceptor sites. The high frequency use of D2 and D3 would also be expected in exon definition (discussed below), however this highlights the unexpected difference in D3 use between the size classes. In this analysis we have likely underestimated A1 and A2 usage due to differences in the length of cDNA product needed to see splicing to the proximal splice acceptor sites A3-A5 (within 300 nucleotides) compared the distance from the cDNA primer to A1 (1100 nucleotides); however, the use of Primer ID corrects for any subsequent skewing in amplification efficiency of longer versus shorter PCR products.

Differential effects of temperature on splicing

HIV-1 splicing must function through the presence of sub-optimal splice donor and acceptor sites to allow for a vast array of splicing outcomes. Inefficient pairing of these sub-optimal sites with the cellular splicing machinery has been shown to affect donor and acceptor usage (10, 15, 28-30). We

reasoned that some of these interactions with the cellular splicing machinery could be temperature sensitive. Temperature changes might also affect the stability of RNA secondary structures that work in cis to regulate splicing. To test the effects of temperature, a culture of CEMx174 cells infected with HIV-1_{NL4-3} was split into 3 flasks that were then incubated for 6 hours at 33°, 37°, or 41° C. At the end of 6 hours RNA was extracted from the cells and the viral splice products were measured (Figure 2.3). In general, increasing temperature modestly up-regulated splice variants without small exons 1 or 2. However, splice acceptors A1 and A2 were sensitive to temperature and showed different responses to temperature changes. Splicing to A1 decreased with increasing temperature, and this was seen for both the *vif* transcript and all transcripts that include SX1 defined by A1 and D2 usage. Similarly, all splices to A2 also decreased with increasing temperature, but were additionally significantly up-regulated at lower temperature, including *vpr* transcripts and transcripts including SX2 between A2 and D3. Transcripts that use both A1 and A2 showed the more dominant effect of A2 alone. Thus we see distinctive temperature sensitivities of the splice acceptor sites with the most sensitive being the sites associated with the generation of the small exons.

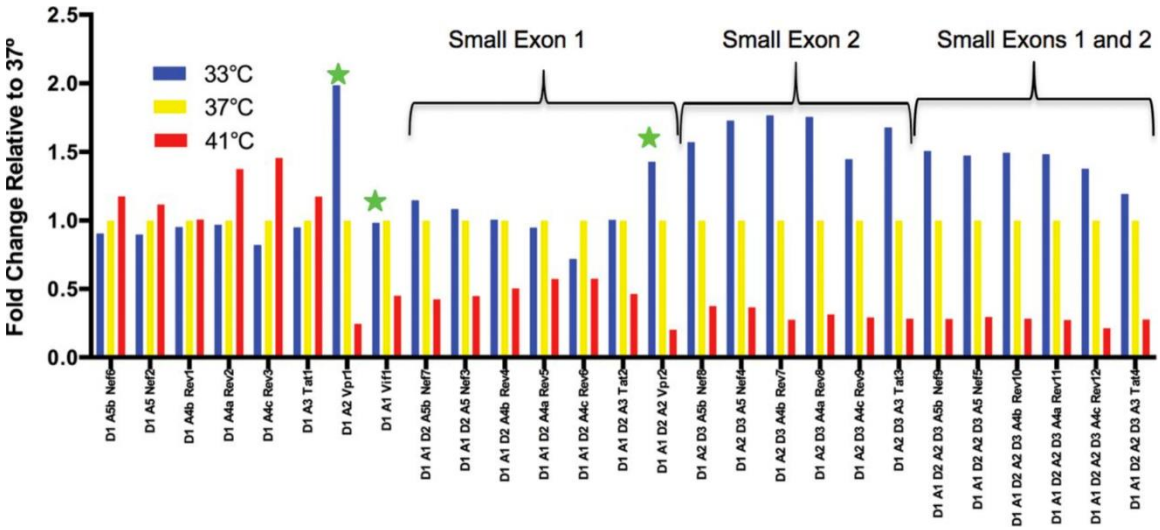

Figure 2.3. Temperature Dependent Splicing Regulation.
Cells were infected with the NL4-3 strain of HIV-1 and after several days the culture was divided and place at the indicated temperatures for the last 6 hours. Whole cell RNA was extracted and viral RNA assessed for splicing. The histogram shows the ratio of the percentage of each 1.8 kb transcript type relative to the 37o sample (always at a 1:1 ratio with itself). Splice variants are grouped by direct splice to final acceptor or small exon inclusion. Stars indicate splices to A1 that make *vif* transcripts or to A2 that make vpr. X-axis labels show the splice donors and acceptors used and the protein product of the transcript.

Changes in secondary structure can affect splicing

HIV-1 RNA has regions of highly conserved secondary structure and many of these structures have functional relevance (22, 28, 29, 31-35). These include the TAR binding site in the 5' UTR which binds the viral Tat protein and is required for processive transcriptional elongation (36). The major splice donor D1 is also in this region as well as a structure that regulates genome packaging (32), and the dimerization sequence. The Rev Response (binding) Element, RRE, located in the *env* intron, also has features of highly conserved secondary structure (37). Some splicing elements are contained in, or are in close proximity to, conserved structural elements, and some of these structures are well characterized (22, 28, 30, 34). This raises important questions about the role of RNA structural elements in controlling splicing during viral infection, and the importance of such structures for correct splicing.

A previous comparison of RNA secondary structure between HIV-1 NL4-3 and SIVmac239 revealed regions of highly conserved structure (22) one of which was a stem-loop structure termed SLSA1 that contains the *vif* splice acceptor site A1 in the loop. We previously engineered two silent point mutations to disrupt the SLSA1 stem without changing the A1 splice acceptor sequence or nearby reported splice regulatory sequences and noted a change in downstream splicing using an older, PCR product size-based assay (22). We have re-evaluated this mutant with the new deep sequencing assay and confirm an effect on splicing but can now identify it as an effect on the proximal *vif* A1 splice acceptor site. RNA from CEMx174 cells infected with the SLSA1 mutant virus was processed in parallel with the RNA from cells infected with the wild-type/parental NL4-3 strain. These two point mutations resulted in an approximately 10-fold decrease in the utilization of A1 relative to the control, which was reflected in the decrease of the direct splice to create *vif* mRNA and in RNAs containing the first small exon from A1 to D2 (Figure 2.4). Thus the conserved SLSA1 stem loop structure containing the A1 splice site plays an important role in regulating the utilization of that site.

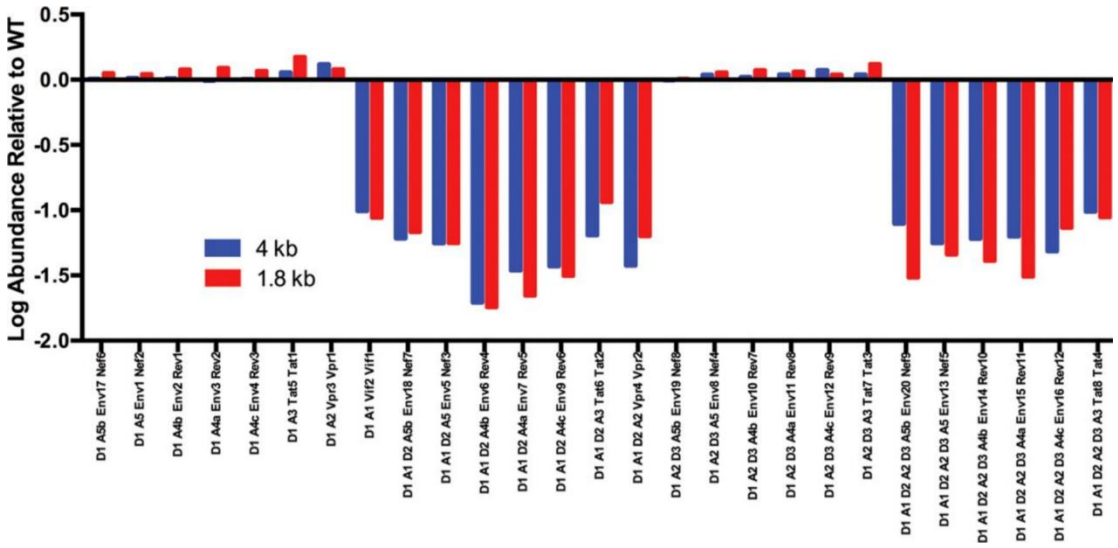

Figure 2.4. Effect of SLSA1 Structural Mutations on Splicing.
Y-axis is a log scale of the fold change of splice type abundance of SLSA1 mutant relative to WT NL4-3. X-axis shows the pattern of splice donors and acceptors and the transcript type in the 4 kb and 1.8 kb class with that pattern. A missing bar indicates that splice type did not occur in the SLSA1 mutant.

Combined effects of temperature and structure

We next reasoned that the temperature sensitivity of A1 might be related to the melting of the proximal conserved stem-loop structure (SLSA1). This would predict that the SLSA1 mutant would not be temperature sensitive for splicing to A1. To test this we compared splicing across the temperature range in the SLSA1 mutant to WT NL4-3, using the same strategy as in the temperature gradient experiment. We saw that the effects of temperature and the SLSA1 mutations were additive when compared to the wild type virus at 37° C, reducing splicing to A1 to very low levels (Figure 2.5A). When we focused on the effect of the different temperatures within the context of reduced splicing of the SLSA1 mutant we saw that the pattern of increases and decreases of splice site usage as a function of temperature were the same as seen with wild type virus (Figure 2.5B). This suggests that there are two separate mechanisms giving rise to the SLSA1 and temperature sensitive phenotypes for splicing to A1. We have also noted enhanced accumulation of APOBEC3 G/F mutations in the deep sequencing data set with the SLSA1 mutant, consistent with reduced *vif* mRNA resulting in reduced levels of Vif (data not shown).

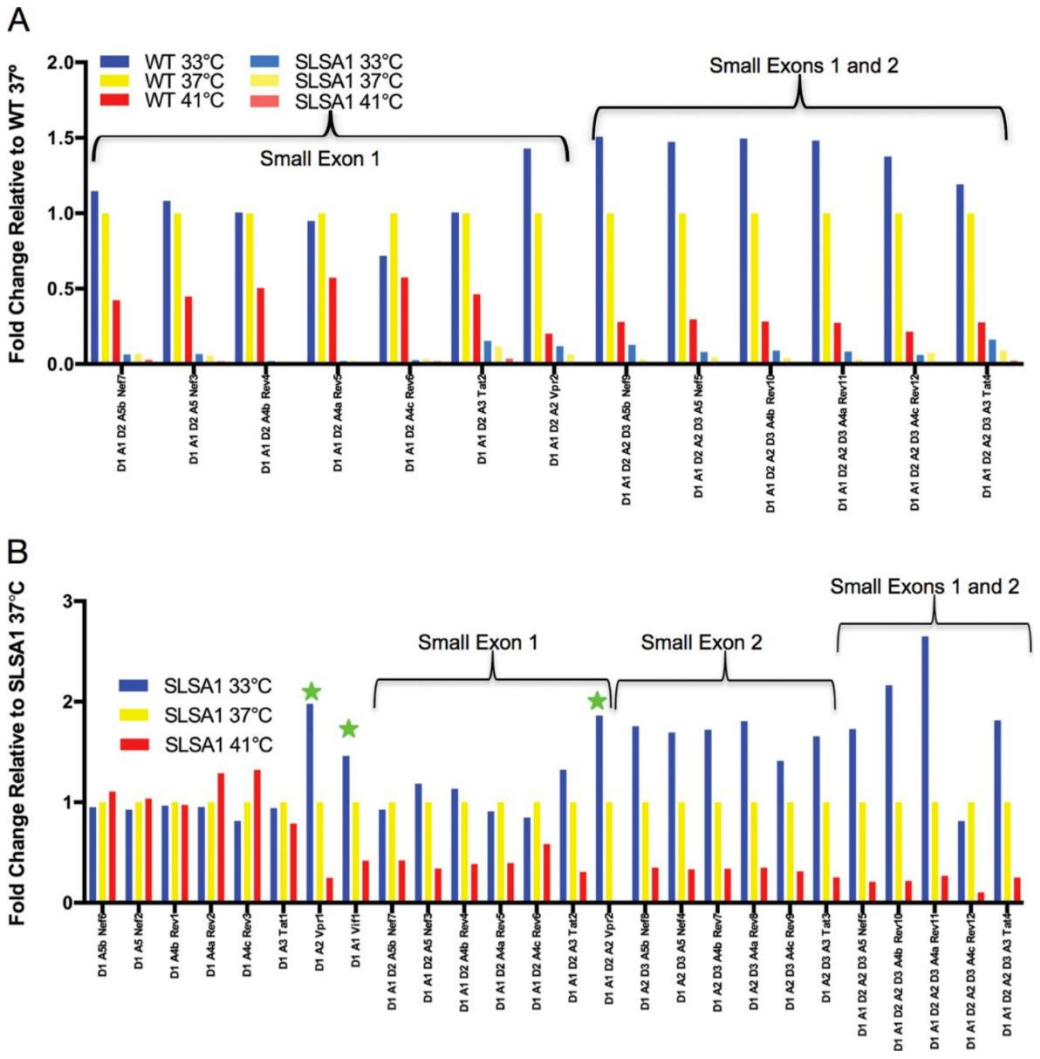

Figure 2.5. Combined Effects of Temperature and SLSA1 Mutation.
A) The histogram shows the ratio of the percentage of each 1.8 kb transcript that utilizes A1 relative to the 37˚C WT sample. B) Ratio of the percentage of all 1.8 kb transcripts from SLSA1 mutant samples relative to 37˚ C SLSA1 mutant sample.

Splicing in transfected cells

The comparison between the SLSA1 mutant and WT NL4-3 was repeated using transfected 293T cells, and essentially identical splicing patterns were observed (data not shown). This and other experiments showed that the splicing quantification assay for HIV-1 gives similar results in comparing asynchronous viral replication in CEMx174 cells or after transfection of 293T cells. Performing the assay in transfected cells has the potential to allow splicing analysis of mutant viral genomes that are incapable of infection and replication in a permissive cell line.

Comparison of splicing in transmitted/founder viruses

Most HIV-1 splicing studies have used the NL4-3 strain, a subtype B chimeric virus and a laboratory strain (21). We asked if splicing in HIV-1_{NL4-3} is representative of splicing in clinical isolates, and if conserved splicing is critical for infection and spread of HIV-1. To address these questions we analyzed splicing in a panel of subtype B T/F viruses (17-20). We found that overall splicing patterns were similar among all of the isolates in that the same splice site acceptors and donors were used, including for the small exons, but with examples of surprising ranges of variability. In these analyses we used values obtained from three different experiments with the NL4-3 strain to account for experimental variability.

The range of A1 usage in the T/F viruses is comparable to NL4-3, while splicing from D1 to A2 varies from near zero to about sixty percent of total transcripts (Figure 2.6A). D2 and D3 are utilized in proportions comparable to NL4-3, and in percentages that correlate closely with the use of upstream A1 or A2. Thus the high frequency use of D3 results in most of the high A2 utilization turning into downstream transcripts, with the level of residual *vpr* transcripts showing only modest variability in spite of the variability in the use of A2 (Figure 2.6B).

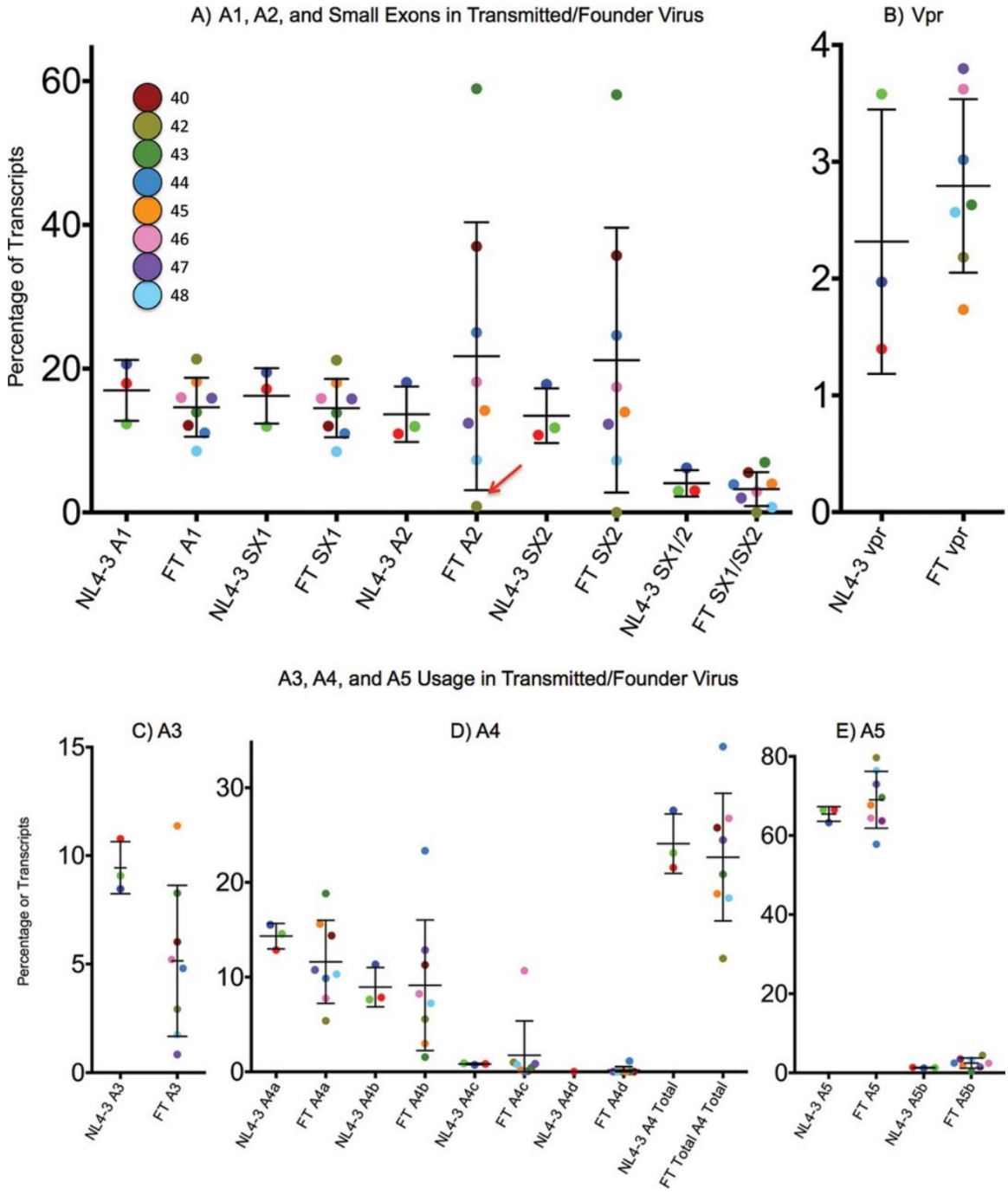

Figure 2.6. Splicing in Transmitted/Founder Viruses.
A) Ranges of A1, A2, and small exon usage in eight transmitted/founder viruses compared to NL4-3. SX1 = small exon 1, SX2 = small exon 2. NL4-3 samples are from three separate experiments, the red dot indicates the sample run in the transmitted/founder experiment. Results shown are for 1.8 kb transcripts. Red arrow indicates sample #42, which has a mutated D3 sequence. B) *vpr* transcripts in transmitted/founder C-E) Usage of A3, A5, and A4 variants in transmitted/founder virus, 1.8 kb transcript class.

One isolate, sample 42 indicated by the red arrow in Figure 2.6A, has a G-to-A transition at the donor site D3 that disrupts the donor consensus sequence. It would be expected that with the knockdown of D3 usage, transcripts spliced to A2 would accumulate resulting in increased *vpr* transcripts and very few transcripts with small exon 2. However, this was not the case. The change to D3 corresponded with greatly decreased splicing to A2 (Figure 2.6A). There are no obvious sequence changes in this virus that would explain reduced use of A2. This suggests either that there is a connection between the recognition of D3 as a splice donor and the use of A2 as a splice acceptor, as predicted by the concept of exon definition, or that some unrecognized sequence change has independently down regulated A2 use in this isolate. Thus in spite of a mutated D3 site and low usage of A2 in this virus, *vpr* transcript levels are near a normal level (Figure 2.6B).

Splicing to A3/*tat* spanned a range from less than 1% to greater than 11% (Figure 2.6C). There are four A4 splice acceptors, named A4a to A4d (Figure 2.6D). There was greater variety in A4 usage than in any other acceptor, and different viruses showed different A4 splice site preferences. There was also a higher degree of sequence variability around these acceptors. The differences in use were partly compensatory in that total A4 usage varied only around 3-fold while the use of any one site varied up to 10-fold. The relative usage of the two A5 acceptors (A5 much greater than A5b) was similar between the different strains, with A5 always being the predominant splice acceptor but varying from less than 60% to more than 80% (Figure 2.6E). Thus the relative amounts of the mRNAs for the different viral proteins were generally similar, with the RNAs for Tat and Rev showing the greatest variability. These results confirm that there is a core set of just over 50 major spliced RNA products that define the conserved HIV-1 RNA expression program.

Detection of cryptic/rare splicing events

While the main donor and acceptor sites are well defined, we identified mis-splicing, particularly to the nearby bases around all of the major donor and acceptor sites. In addition, there was a low-level background of rarely used acceptor sites across the genome, interspersed with hotspots. It has been reported previously that blocking a canonical splice acceptor or donor site leads to activation of cryptic sites (2, 38, 39). While we did not test this phenomenon directly, we will note that the virus with the defective D3 donor site (sample 42) did not activate a substitute donor site for D3 or an alternative

acceptor for A2. Even the T/F viruses in the extremes of low or high acceptor usage did not activate significant cryptic donor or acceptor sites. A scan for rarely used acceptor sites in all of the T/F viruses identified sites used at various frequencies, and the proportion of proximal alternative acceptor sites compared to use of the canonical acceptor sites for the each of the transcripts for viral proteins are shown in Figure 2.7. Rare acceptors within the normally unspliced *gag/pro/pol* (GPP) sequence were also compared to total transcripts. Transcripts with rare splice sites made up a very small proportion (less than 1%) of total transcripts in both NL4-3 and in this panel of T/F viruses.

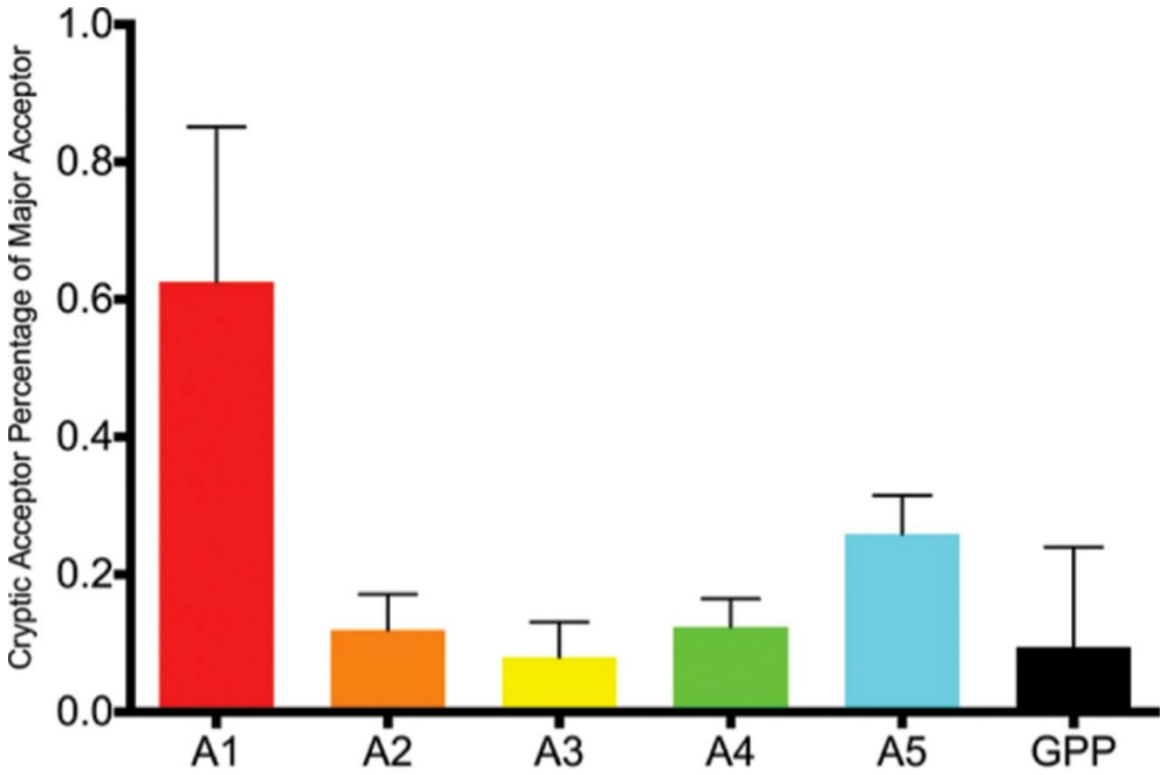

Figure 2.7. Cryptic Acceptor Usage as a Percentage of Major Acceptor Usage in Transmitted/Founder Viruses and NL4-3.
Transcripts with cryptic acceptor sites are shown as a percentage of non-cryptic transcripts using the comparable major acceptor site with the same open reading frame. Cryptic acceptors in the *gag/pro/pol* (GPP) sequence are compared to total transcripts.

An unusual set of rare splicing events was observed in both HIV-1_{NL4-3} and all T/F viruses in which the apparent order of splicing was reversed. Splicing from D2 to A1 and from D3 to A1 or A2 was observed, as well as from D4 to all upstream acceptors. This out-of-order acceptor use suggests trans-splicing, where a donor from one transcript splices to an acceptor in a different transcript. In the 1.8 kb

size class of the NL4-3 isolate, trans-splicing to A1 made up 0.006% of total transcripts, and 0.3% of all reads had a D4 trans-splice to an upstream acceptor. Acceptor usage in these D4 trans-splices was similar to acceptor usage from D1, as shown in Figure 2.2A. In these cases the region between A5 and D4 becomes yet another small exon. These additional small exons skew towards the 1.8 kb transcript class, at a frequency six times greater than that of the 4 kb transcripts for trans-splices from D2 or D3 to A1, and three times greater for trans-splices from D4. Remarkably, consecutive trans-splicing events can also be detected in rare transcripts such as D4 to A2 followed by D3 to A2.

It has been observed that reverse transcriptase and/or PCR template switching can create trans-splicing artifacts between regions of sequence identity (40, 41). We do not believe this is the cause of the trans-splices we have observed between these canonical splice sites. We looked for evidence of artifactual apparent trans-splicing. There are regions of exact identity to portions of several canonical acceptor and donor sites in some of the T/F viruses. Isolates 43 and 47 have a duplicate post-A1 sequence 166 nucleotides upstream of the actual A1 site. In addition, isolate 42 has a duplicate pre-D2 sequence 24 nucleotides upstream. All have 10 nucleotides of exact sequence identity and yet they were never found in combination with a canonical splice donor or acceptor. Thus the trans-splicing is linked to the functional donor and acceptor sites and not regions sharing sequence identity with these sites. Additionally, we reasoned that if template switching were random in regions of sequence similarity, we should see a high proportion of other random events near the trans-spliced sites; however, when we scanned for nearby trans-splices we found none. Taken together, this leads us to conclude that the trans-splicing we observed is not artifactual but rather a measure of errors in splicing by the host cell machinery.

We asked if trans-splicing occurs only in cis, i.e. between transcripts in close proximity as they are consecutively transcribed from a single provirus, or if it could occur in trans between transcripts from proviruses at different integration sites. To address this question we pseudotyped HIV-1 subtype C Δenv particles with the VSV G envelope protein and used this virus to infect 8E5 cells. 8E5 cells have a single stably integrated HIV-1 subtype B genome. 8E5 and the subtype C virus generated from pZM247Fv2 have sequence dissimilarities that can identify a change from a donor of one viral RNA to an acceptor of another viral RNA. We harvested total cellular RNA from the dually infected 8E5 cells and used it in the

splicing assay, with forward and reverse primers specific for 8E5 and pZM247Fv2. Our results found trans-splicing from both the 8E5 virus to the pZM247Fv2 virus and vice versa, suggesting that trans-trans-splicing between transcripts from different proviruses is possible.

Splicing in subtype C

We adapted the assay to quantify splicing in this subtype C strain and found that splicing for virus generated from the pZM247Fv2 clone is very similar to virus generated from the pNL4-3 clone (subtype B). Donor and acceptor sites were precisely conserved. The relative abundance of splice acceptor usage is also very similar (Figure 2.8) as is usage of the small exons. The splicing differences seen between the 4 kb and 1.8 kb sizes in NL4-3 are also seen in pZM247Fv2. A previous study found high variation in A4 splice site usage among subtypes and samples (16, 42). In this single sample, we did not see any new splice site usage, although the pZM247Fv2 sequence has a point mutation that completely inactivates A4d.

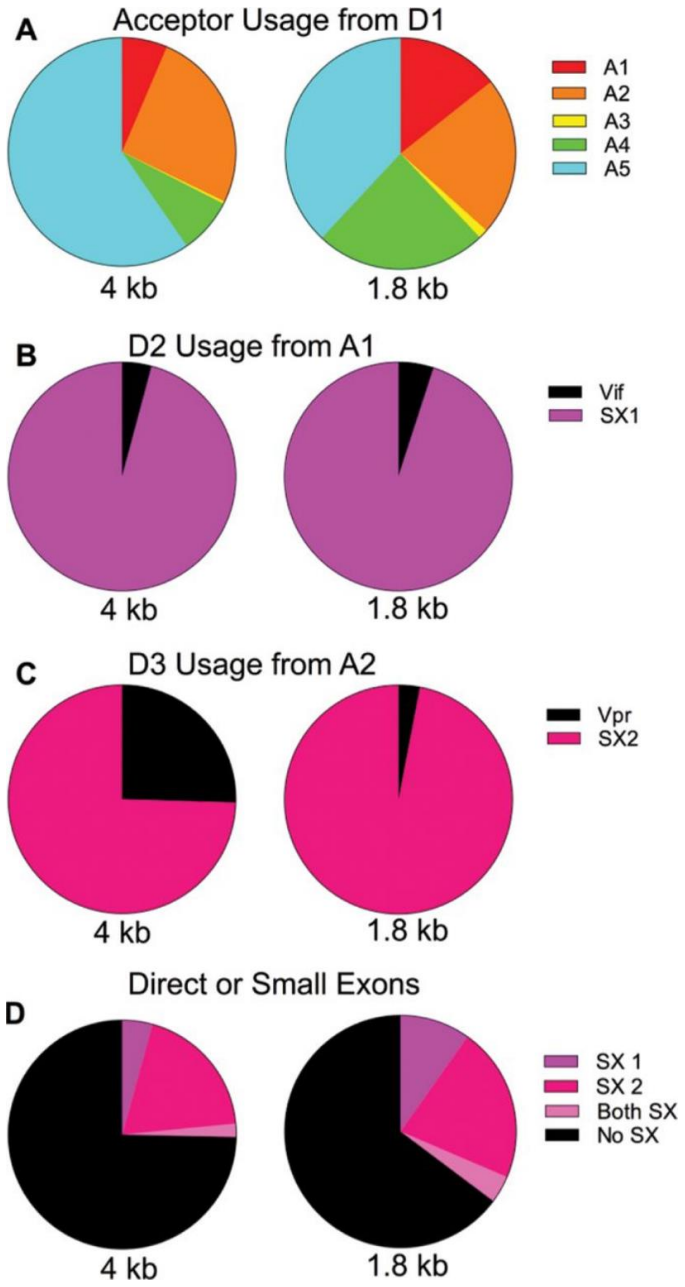

Figure 2.8. Quantification of Splicing in Subtype C Strain pZM247Fv2.
A) Acceptor usage from D1 in the two size classes. Circles represent all transcripts in size class. B) Circles represent the total splices from D1 to A1. Shown are proportions that splice again at D2 and those that remain *vif* transcripts. C) Circles represent the total splices to A2, and the proportions that splice again at D3 compared to those that remain *vpr* transcripts. D) Circles represent all transcripts in size class. Shown are proportions of transcripts that contain either one, both, or no small exons. SX1 = small exon 1, SX2 = small exon 2.

Splicing quantification in in SIVmac239

We next adapted the assay to quantify splicing in SIVmac239. SIVmac239 has a more complicated splicing pattern than HIV-1 (Figure 2.9). Though generally very similar to HIV-1, SIVmac239 includes an expanded 5' untranslated region (UTR) with an optional intron upstream of D1, here called small exon 0 (SX0). We began by using a random reverse primer to find the splice junctions, then designed primers that would pick up the 1.8 and 4 kb size class equivalents, as in the HIV-1 assay. Like the HIV-1 assay, the SIVmac239 assay uses Primer ID-tagged reverse primers specific to each size class and a forward primer 5' of D1, which is also the major splice donor. Additionally, a second forward primer 5' of D0 was used in separate reactions with each reverse primer to quantify splicing from D0.

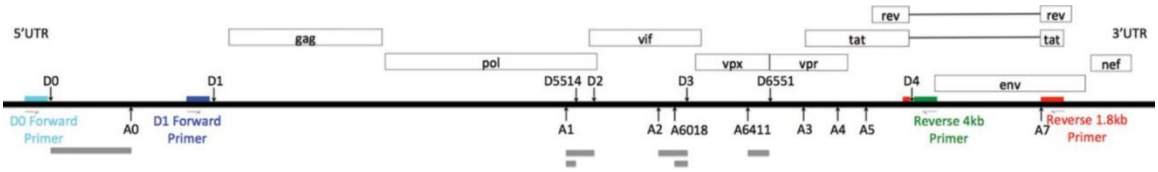

Figure 2.9. SIVmac239 Splice Patterns and Primer Locations.
Light Blue = D0 forward primer. Blue = D1 forward primer. Red = 1.8 kb class Primer ID tagged reverse primer. Green = 4 kb class Primer ID tagged reverse primer. Grey boxes are small exons or sequences that may or may not be present in the respective transcripts. Not drawn to scale but relative locations are correct.

Most SIVmac239 acceptors and donors align with and act analogously to HIV-1 acceptors and donors (Figure 2.10). Relative acceptor usage is mostly similar to HIV-1, except that in SIVmac239 A5 was used even more frequently in both size classes, such that most spliced transcripts use A5 as the final acceptor, whether by direct splicing from D1 or by including the small exons (Figure 2.10B).

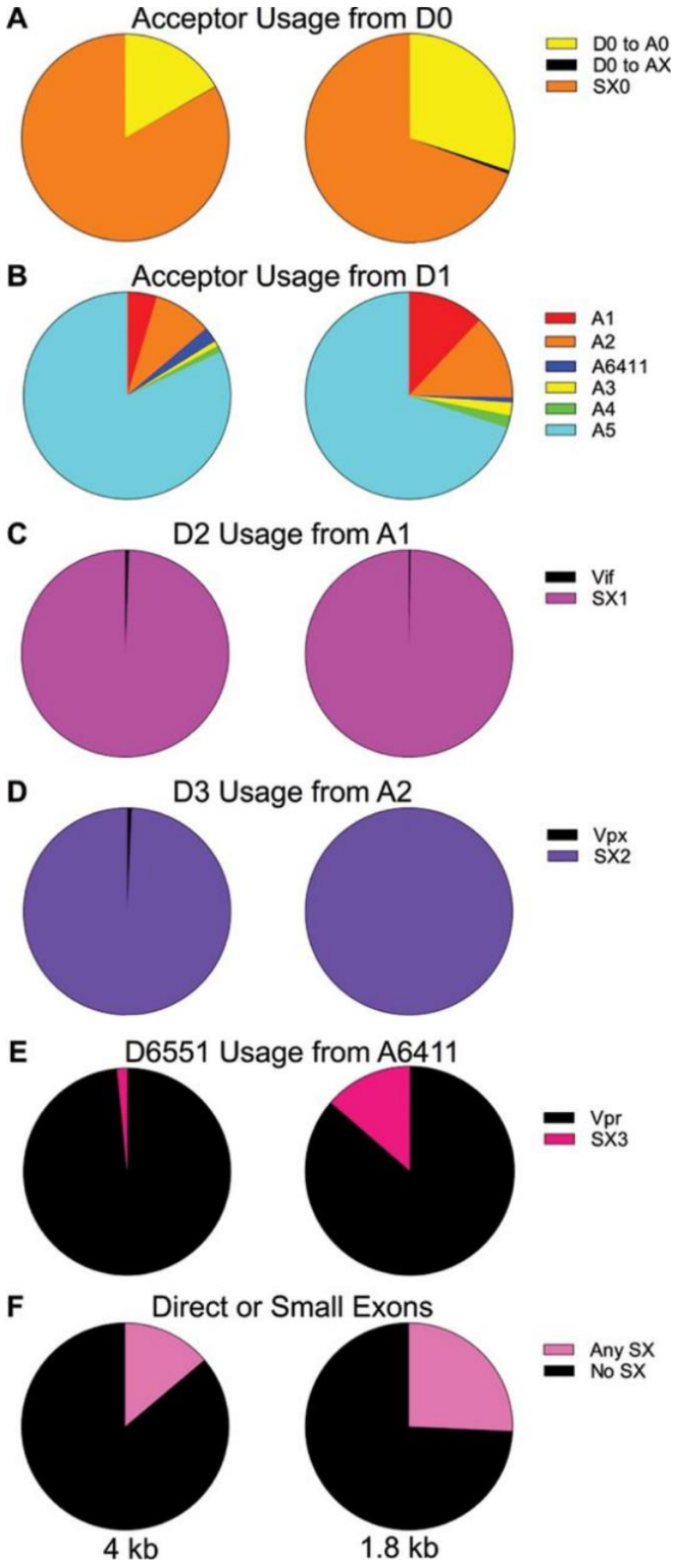

Figure 2.10. Quantification of Splicing in SIVmac239.
A) Acceptor usage from D0 in the two size classes. Yellow represents the proportion of transcripts that splice from D0 to A0, black represents the proportion of transcripts that splice directly from D0 to an acceptor other than A0,

orange represents the transcripts that do not splice from D0 to A0. B) Acceptor usage from D1 in the two size classes. Circles represent all transcripts in size class. C) Circles represent the total splices from D1 to A1. Shown are proportions that splice again at D2 and those that remain *vif* transcripts. D) Circles represent the total splices to A2, and the proportions that splice again at D3 compared to those that remain vpx transcripts. E) Circles represent the total splices to A6411. Shown are proportions that splice again at D6551 compared to those that remain *vpr* transcripts. F) Circles represent all transcripts in size class. Shown are proportions of transcripts that contain any small exons compared to those that splice directly to the final acceptor.

The two small exons, SX1 and SX2 are conserved, though in SIVmac239 additional donors and acceptors create a third small exon (called here SX3) due to the presence of the extra gene *vpx*. Two of these small exons overlap with SX1 (using A1 and either D2 or a donor at 5514) and SX2 (using either A2 or an acceptor at 6018 and D3). The third small exon spans 6411 to 6551. Thus an SIVmac239 transcript can have up to three of these small exons, and all possible combinations of the three small exons are observed. In SIVmac239, A2 (and SX2) are linked to *vpx*, a gene not present in HIV-1 and placed between *vif* and *vpr*. An additional splice acceptor site, A6411, is used to make *vpr* transcripts and this splicing event activates D6551 to make SX3. In both SIVmac239 and HIV-1 a greater fraction of splices to the *vpr* acceptor remain as *vpr* transcripts in the 4 kb class, while in the 1.8 kb size class use of A6411 more often activates D6551 to make SX3; however, in contrast to the high use of D3 to make SX2 in HIV-1, in SIVmac239 SX3 is seldom made even in the 1.8 kb class – most splices to A6411 remain *vpr* transcripts (compare Figure 2.10E and Figure 2.2C). It is worth noting that splicing to *vpx* is similar to the upstream *vif* and not the downstream *vpr* in the high use of the proximal donor site (D3) in both size classes leaving little *vpx* mRNA (Figure 2.10C, D).

There is no splicing analog to SX0 in HIV-1. In contrast to the creation of the small exons SX1 and SX2, which are generated after splicing to A1 or A2 followed by splicing from D2 and D3 (respectively), inclusion of SX0 results from not splicing. Quantification using the primer upstream of D0 shows that the 85 nucleotide long D0-A0 intron is retained in most transcripts – 73.6 % in the 1.8 kb class and 84.6 % in the 4 kb class (Figure 2.10A). Although D0 can splice to acceptors other than A0 it only rarely does so (0.51 % in the 1.8 kb class, 0.075% in the 4 kb class). This suggests that activation of D0, unlike D1, is tightly coupled to the use of the proximal splice acceptor A0.

There are some quantification limitations in the SIVmac239 splicing assay due to the limited read length using the MiSeq platform. Although quantification of splicing from D1 is similar in HIV-1 and SIVmac239 with most splice variants identifiable, the increased length of SIVmac239 between donors and acceptors makes some spliced variants too long to fully catalog, so some long transcripts (A3 with

small exons) are grouped together. Measuring splicing from D0 has additional limitations in that the length of the forward read just reaches to D1, so all splicing from D1 must be quantified in the reverse read and this is not sufficient to separate splices from D1 to A1, A2, and A3. Spliced RNAs quantified from D0 must either splice from D0 directly to some downstream acceptor other than A0 or must also splice at D1 to be short enough to be sequenced. It is possible that there are transcripts that splice from D0 to A0 but then remain full-length downstream, an RNA variant that would not be detected using the current approach.

Discussion

We have developed a new HIV-1 splicing assay that utilizes the MiSeq deep sequencing platform and Primer ID technology to efficiently and accurately quantify HIV-1 splicing. The depth of sequencing data combined with template indexing (Primer ID) make it possible to identify and quantify even low abundance variants. This assay is unique in its ability to simultaneously quantify multiple splicing events in HIV-1 RNA transcripts in the context of viral infection with a total of just three primers. It is noteworthy that this assay also works in transfected 293T cells, allowing quantification of mutated viral genomes that are incapable of productive infection. The assay has been used to measure splicing in infected primary T cells and dendritic cells, and the results were comparable to the splicing seen in CEMx174 cells (C. Miller, R. Gummuluru, A.E., R.S. in preparation), validating the cell line model. This is a moderate throughput assay – 24 samples can be multiplexed in a single deep sequencing reaction – that will allow it to be used to screen and compare the activity of mutations or compounds for their effects on splicing.

It is useful to compare this new assay using the MiSeq platform to previously described assays. Given the complexity of HIV-1 splicing we will limit this comparison to assays that are capable of capturing most of that complexity, specifically the original RT-PCR/gel sizing assay of Purcell and Martin (2), and the more recent use of the PacBio next gen sequencing platform described by Bushman and colleagues (3). While the RT-PCR/gel sizing assay is the simplest it relies on size for the identification of the splicing events, some of which can co-migrate, and has limited sensitivity to detect minor variants. Also, the number of mRNA templates queried is not known making the procedure prone to PCR skewing. For these reasons, next gen sequencing platforms offer the opportunity to capture much more information

about splicing and represent a natural evolution where new technology allows re-examination of longstanding questions to gain a greater depth of understanding.

The MiSeq assay and the PacBio assay share the features of an initial cDNA synthesis reaction followed by amplification and sequence analysis. MiSeq sequencing is sufficiently accurate to allow the use of Primer ID which allows correction for PCR skewing and quantifies the number of mRNA templates actually sequenced. PacBio is not yet amenable to Primer ID due to its higher level of error and therefore it is necessary to do end-point dilution PCR of isolated templates (in droplets) to avoid PCR skewing but this system cannot account for the number of mRNA templates that are sequenced which results in PCR resampling during sequencing. Both systems suffer from the low processivity of RT which results in an under-estimation of splicing events that are detected based on a longer cDNA; in our assay cDNA products range from about 300 to 1100 nucleotides. We have focused our assay on the large majority of splicing events which we detected with just three primers; the PacBio assay included the use of more primers which included the detection of an unusual 1 kb size class of spliced RNAs (discussed below).

Within the individual 1.8 and 4 kb size classes, relative transcript abundance is similar in both assays but both assays share two important weaknesses - the inability to quantify the 1.8 kb transcript class relative to the 4 kb transcript class, or to quantify spliced transcripts versus unspliced transcripts. Size class quantification is a complicated problem as there are no sequences in the 4 kb class of transcripts that are not also present in either the 1.8 kb class and/or unspliced transcripts. Given the inherent, but different, limitations of the MiSeq and PacBio platforms it has been necessary to add a complementary assay to each data set to obtain a more complete view of viral populations. The PacBio approach has been linked to an RNAseq data set where oligo dT-primed cDNAs of whole-cell RNA are analyzed by short reads over their entire length (3, 43). A feature of using RNAseq is that the lack of processivity in the cDNA synthesis step by reverse transcriptase requires a correction of observed to calculated frequency of the longer cDNAs (up to 9 kb) making it difficult to measure small differences with confidence. To complement the use of the MiSeq platform we have explored the use of a random cDNA primer linked to amplification with the same upstream primer just 5' to D1. This allows a comparison of all transcripts, spliced or unspliced. In this analysis we estimate that about 22% of the spliced transcripts are in the 1.8 kb size class with 78% in the 4 kb size class (unpublished observation). Estimation of the ratio

of unspliced to spliced viral RNA appears to be confounded in this system by abortive or incompletely transcribed and processed reads (44, 45), giving a higher-than-expected size of the unspliced population. It is likely that the estimation of the ratio of different size classes and spliced to unspliced will be sensitive to the bias of whatever method is used in which case it will be important to compare several methods or to focus on the use a sensitive and reproducible method that allows the detection of differences as experimental conditions are changed.

In its current form the MiSeq assay relies on just three primers; we have chosen not to include another primer to detect either the direct splice from D1 to A7/*nef* or to look for the 1 kb spliced RNA size class, although additional primers could be included to capture these events. Regarding direct splicing from D1 to A7/*nef*, these are splicing events that are part of the canonical splicing program but that occur at a low percentage. Our data using the random reverse primer suggests that D1 to A7 splices represent only about 0.4 percent of all spliced transcripts. Therefore, we conclude that while not exhaustively inclusive, this new MiSeq assay captures virtually all essential canonical splicing events and can be used to sensitively measure changes in splicing.

Similarly, we have chosen not to use an additional primer to probe for the 1 kb spliced RNA. These smaller RNAs have been previously described (46, 47) and were seen with the HIV-$1_{89.6}$ isolate but generally thought not to be conserved among isolates (3). We did not detect similar spliced RNAs in either the NL4-3 isolate or in the SIVmac239 isolate using the random reverse primer approach. For this reason we have chosen not to include another primer (which would require a separate amplification) to detect what appear to be poorly conserved splicing events outside of the canonical splicing program. However, the existence of the previously uncharacterized 1 kb RNAs does suggest that using an approach like the random reverse primers with the transmitted/founder virus set of isolates would allow a more systematic examination of the question of any as yet undetected conserved splicing events, which must exist at a low level if at all.

Using our assay we quantified splicing patterns in the NL4-3 laboratory strain and observed that usage of splice acceptors A1 and A2 is temperature sensitive. We linked a specific RNA secondary structure, SLSA1, to the efficiency of splicing at its proximal acceptor site (A1). Finally, we used this new assay to analyze splicing in a panel of subtype B transmitted/founder viruses, a subtype C isolate, and an

SIVmac isolate, and found that splice donor and acceptor sites are well conserved, though frequency of use varied. Following we discuss observations and issues raised given these very precise measures, based not on the total number of sequence reads but rather on the actual number of mRNA templates sequenced, of relative splice site usage.

We quantified splice acceptor site usage in NL4-3 and found that of the two small exons made by using D2 or D3, D2 usage was consistent in both transcript size classes but D3 (D6551 in SIVmac239) was not – in the 1.8 kb size class most splices to A2 activate splicing from the adjacent D3, but in the 4 kb size class about half of the D1-A2 spliced RNAs remain as *vpr* transcripts without activating D3 (Figure 2.2). In contrast splicing to A1/*vif* activates D2 most of the time in both RNA size classes. This suggests a link between D3 usage and downstream D4 usage, which has not yet been explored. This difference is even greater in SIVmac239 where splicing to A2/*vpx* activates D3 in both size classes similar to what happens with splicing to A1/*vif*, but splicing to the new downstream A6411/*vpr* acceptor activates the adjacent D6551 donor in only a minority of cases and hardly at all in the 4 kb size class. Thus there is a strong gradient effect of linked donor activation across the center of the viral RNA. Nonsense mediated decay (48) could play a role in degrading 1.8 kb *vpr* transcripts thus leaving mostly 4 kb *vpr* transcripts, but this raises the question of why the same rules don't apply to 1.8 kb and 4 kb *vif* and *vpx* transcripts. Alternatively, the mechanism that suppresses splicing from D4 (allowing the formation of the 4 kb size class) may also influence the use of the proximal upstream donor site in both HIV-1 and SIV. There are several other examples where splicing is differentially suppressed in the 4 kb size class that support this notion (see below).

The purpose of the conserved small exons remains to be clearly elucidated. They are conserved across subtypes and in SIVmac239 (including for the additional *vpx* gene) (49, 50) (Figure 2.1 and 2.9), as well as in the T/F viruses analyzed (Figure 2.6), consistent with strong selective pressure for their presence. Previously cited studies to define their functional importance are, as mentioned, inconclusive (1, 6-9). It is not clear whether donors D2 and D3 exist to down-regulate the amount of *vif* and *vpr* mRNAs or if D2 and D3 are required for exon definition (51) to activate A1 and A2 respectively. The transmitted/founder virus isolate 42 with the mutant D3 argues for exon definition, i.e. knockdown of D3 reduced A2 usage, a result that has also seen in model systems (10, 14, 52); however, in another study

with a different HIV-1 isolate we saw that a reduction in D3 usage led to accumulation of *vpr* transcripts (data not shown). Viruses that differ dramatically in their use of *vpr*/A2 still produce similar levels of *vpr* mRNA (Figure 2.5A), consistent with the idea that splicing regulation can be accomplished by different mechanisms even for the same site. More work is needed to clarify the roles of D2 and D3 and the purpose of the small exons given their conservation across clades and in at least some nonhuman primate lentiviruses.

There are 62 spliced mRNA species that are well conserved among the different isolates. Given the conservation of these alternative splicing pathways for HIV-1 we were interested in using temperature shifts to probe features of their control and found that the use of A1 and A2 are each, but differentially, sensitive to temperature changes. The differential response to temperature in the use of A1 and A2 are consistent with suboptimal interactions of these splice acceptor sites with the host cell splicing machinery, but in different ways. It is worth noting that these same two splice sites are indirectly down-regulated with the high frequency use of D2 and D3, suggesting the use of A1 and A2 have been selected to be suboptimal for interactions near the point of instability. A secondary structure proximal to A1 was seen to be critical to splicing based on the observation that two mutations designed to decrease its structural stability reduced splicing in a way that was additive with the effect of temperature (Figure 2.5). These data do not rule out the possibility that the mutations affect other features of splicing, such as binding by an unknown factor, and further work is required to fully understand the role of this hairpin in regulating A1 usage. It would be also be interesting to use deep sequencing technologies to see if temperature shifts reveal similar changes among alternative splicing pathways of cellular RNAs.

Our data suggest that the SLSA1 structure is important to A1 usage, and other splice acceptors and splice control elements are also located in or near structural features and many of these regions have been studied for their ability to bind cellular splice factors (53, 54). It will be informative to continue to characterize the role these and other structures play in splicing regulation by creating mutations that alter either structure or cellular factor binding sites. One such study has involved the knockdown of the cellular factor hnRNP H1 with the observation that splicing to A1/*vif* is most affected by limiting this factor (S. Kutluay, P. Bieniasz, A.E., R.S., in preparation).

As most previous HIV-1 splicing studies have used HIV-1_{NL4-3} as a model, it is significant that we observed HIV-1_{NL4-3} splicing to be comparable to the mean splicing patterns seen in the T/F virus panel and in a subtype C isolate (Figure 2.6 and 2.8). This finding validates use of HIV-1_{NL4-3} in other studies of splicing regulation. In addition, some general conclusions can be drawn from the T/F virus splicing data. Overall, splicing patterns are the same. Acceptors and donors remain constant, but there is surprisingly high variability in their usage. In addition, there are examples of splicing events within a strain that are not conserved between strains, with some examples being the reported 1 kb size class in the 89.6 strain (3), the SX0 exon in the SIVmac239 strain, and greater than 50% *tat* transcripts in the NDK strain (data not shown). The idea that splicing is a tightly regulated and carefully balanced process (1) needs reexamination. The variance in transcripts in these different strains of virus suggests that if a virus meets a threshold level for each transcript type, it can be transmitted and continue to replicate.

Previous studies suggested that cryptic splicing will be activated if/when usage of a regular acceptor is changed or blocked (1, 2, 38, 39). Our data do not support this concept. Cryptic splicing remained negligible across the entire panel of T/F viruses, in spite of functionally important changes in acceptor and donor consensus sequences. In no case was a cryptic donor or acceptor substitute activated above background levels. HIV-1 acceptors have been classified as weak (55, 56) and clearly they diverge from the cellular splicing consensus sequence and branch point rules (1, 57, 58); however, they are still consistently recognized as major splice sites above the background noise. This suggests that many as yet unexplained factors contribute to HIV-1 splicing regulation.

We found evidence of trans-splicing in NL4-3 and all of the transmitted/founder viruses. Our assay can detect trans-splicing up to a maximum length of four small exons so the full extent of trans-splicing may be greater than what we have seen. The overall frequency of trans-splicing was surprisingly high. At a frequency of 0.3% in the 1.8 kb class, total trans-splicing occurred more often than several of the canonical spliced transcript types. Trans-splicing in HIV-1 has been previously reported (59, 60), though not between HIV-1 RNAs, which we infer must have occurred given the altered order of the exon segments and between transcripts of different proviruses. Whether the cryptic splicing we observed here has biological significance or is merely a stochastic occurrence between two viral transcripts in close proximity is unknown. By observing trans-splicing between transcripts from two different proviruses we

were able to show that this phenomenon occurs beyond a flurry of co-transcriptional splicing between adjacent transcripts. Also, the greater level of trans-splicing in the 1.8 kb class compared to the 4 kb class provides yet another example of where splicing is partially suppressed in the context of retaining the D4/A7 intron. The effects on suppressing the use of these viral splice donor sequences is reminiscent of the observation that there is a general suppression of splicing (i.e. intron retention) in cellular transcripts during HIV-1 infection (43). One model to explain these phenomena is that binding of the RNA by Rev to move the RNA to the Crm1 export pathway may suppress proximal splice donor activation, and that Rev may fortuitously bind to some cellular pre-RNAs.

In conclusion, developments in deep sequencing technology now make it possible to quantify the complexity of HIV-1 splicing in depth and to study effectors of splicing with greater precision than previously possible. Here we report some examples of its utility and suggest potential uses in further studies.

REFERENCES

1. Stoltzfus CM. 2009. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. Adv Virus Res 74:1-40.

2. Purcell DF, Martin MA. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67:6365-6378.

3. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res 40:10345-10355.

4. Cullen BR, Greene WC. 1990. Functions of the auxiliary gene products of the human immunodeficiency virus type 1. Virology 178:1-5.

5. Fischer U, Huber J, Boelens WC, Mattaj IW, 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.

6. Chang DD, Sharp PA. 1989. Regulation by HIV Rev depends upon recognition of splice sites. Cell 59:789-795.

7. Krummheuer J, Lenz C, Kammler S, Scheid A, Schaal H. 2001. Influence of the small leader exons 2 and 3 on human immunodeficiency virus type 1 gene expression. Virology 286:276-289.

8. Schwartz S, Felber BK, Benko DM, Fenyo EM, Pavlakis GN. 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J Virol 64:2519-2529.

9. Muesing MA, Smith DH, Capon DJ. 1987. Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. Cell 48:691-701.

10. Madsen JM, Stoltzfus CM. 2005. An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. J Virol 79:10478-10486.

11. Amendt BA, Hesslein D, Chang LJ, Stoltzfus CM. 1994. Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. Mol Cell Biol 14:3960-3970.

12. Amendt BA, Si ZH, Stoltzfus CM. 1995. Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. Mol Cell Biol 15:4606-4615.

13. Asang C, Hauber I, Schaal H. 2008. Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. Nucleic Acids Res 36:1450-1463.

14. Bilodeau PS, Domsic JK, Mayeda A, Krainer AR, Stoltzfus CM. 2001. RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element. J Virol 75:8487-8497.

15. O'Reilly MM, McNally MT, Beemon KL. 1995. Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. Virology 213:373-385.

16. Delgado E, Carrera C, Nebreda P, Fernandez-Garcia A, Pinilla M, Garcia V, Perez-Alvarez L, Thomson MM. 2012. Identification of new splice sites used for generation of rev transcripts in human immunodeficiency virus type 1 subtype C primary isolates. PLoS One 7:e30574.

17. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, Derdeyn CA, Farmer P, Hunter E, Allen S, Manigart O, Mulenga J, Anderson JA, Swanstrom R, Haynes BF, Athreya GS, Korber BT, Sharp PM, Shaw GM, Hahn BH. 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol 82:3952-3970.

18. Salazar-Gonzalez F MS, B.F. Keele, G.H. Learn, J.M. Decker, S. Wang, J. Baalwe, M. Kraus, B. Guffey, C. Ochsenbauer, J.C. Kappes, M. Saag, et al. 2009. HIV-1 Sequences in Acute and Early Infection Reveal the Genetic Identity, Biological Phenotype, and Precise Evolutionary Pathways of Transmitted/Founder Viruses and Their Progeny. J Exp Med 206:1273-1289.

19. Lee HY, Giorgi EE, Keele BF, Gaschen B, Athreya GS, Salazar-Gonzalez JF, Pham KT, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Hahn BH, Shaw GM, Korber BT, Bhattacharya T, Perelson AS. 2009. Modeling sequence evolution in acute HIV-1 infection. J Theor Biol 261:341-360.

20. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping LH, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105:7552-7557.

21. Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, Martin MA. 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J Virol 59:284-291.

22. Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM, 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 Pathog 9:e1003294.

23. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, Decker JM, Wang S, Baalwa J, Kraus MH, Parrish NF, Shaw KS, Guffey MB, Bar KJ, Davis KL, Ochsenbauer-Jambor C, Kappes JC, Saag MS, Cohen MS, Mulenga J, Derdeyn CA, Allen S, Hunter E, Markowitz M, Hraber P, Perelson AS, Bhattacharya T, Haynes BF, Korber BT, Hahn BH, Shaw GM. 2009. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J Exp Med 206:1273-1289.

24. Regier DA, Desrosiers RC. 1990. The complete nucleotide sequence of a pathogenic molecular clone of simian immunodeficiency virus. AIDS Res Hum Retroviruses 6:1221-1231.

25. Folks TM, Powell D, Lightfoote M, Koenig S, Fauci AS, Benn S, Rabson A, Daugherty D, Gendelman HE, Hoggan MD, et al. 1986. Biological and biochemical characterization of a cloned Leu-3- cell surviving infection with the acquired immune deficiency syndrome retrovirus. J Exp Med 164:280-290.

26. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A 108:20166-20171.

27. Bohne J, Wodrich H, Krausslich HG. 2005. Splicing of human immunodeficiency virus RNA is position-dependent suggesting sequential removal of introns from the 5' end. Nucleic Acids Res 33:825-837.

28. Jain N, Morgan CE, Rife BD, Salemi M, Tolbert BS. 2016. Solution Structure of the HIV-1 Intron Splicing Silencer and Its Interactions with the UP1 Domain of Heterogeneous Nuclear Ribonucleoprotein (hnRNP) A1. J Biol Chem 291:2331-2344.

29. Ajamian L, Abrahamyan L, Milev M, Ivanov PV, Kulozik AE, Gehring NH, Mouland AJ. 2008. Unexpected roles for UPF1 in HIV-1 RNA metabolism and translation. Rna 14:914-927.

30. Rollins C, Levengood JD, Rife BD, Salemi M, Tolbert BS. 2014. Thermodynamic and phylogenetic insights into hnRNP A1 recognition of the HIV-1 exon splicing silencer 3 element. Biochemistry 53:2172-2184.

31. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460:711-716.

32. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, Dorjsuren B, Kulandaivel G, Jones S, Hiremath A, Divakaruni SS, LaCotti C, Barton S, Tummillo D, Hosic A, Edme K, Albrecht S, Telesnitsky A, Summers MF. 2011. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. Science 334:242-245.

33. Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, Barton S, Hosic A, Florwick A, Santos J, Bolden NC, McCowin S, Case DA, Johnson BA, Salemi M, Telesnitsky A, Summers MF. 2015. RNA structure. Structure of the HIV-1 RNA packaging signal. Science 348:917-921.

34. Levengood JD, Rollins C, Mishler CH, Johnson CA, Miner G, Rajan P, Znosko BM, Tolbert BS. 2012. Solution structure of the HIV-1 exon splicing silencer 3. J Mol Biol 415:680-698.

35. Buratti E, Baralle FE. 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol 24:10505-10514.

36. Berkhout B. 1992. Structural features in TAR RNA of human and simian immunodeficiency viruses: a phylogenetic analysis. Nucleic Acids Res 20:27-31.

37. Sherpa C, Rausch JW, Le Grice SF, Hammarskjold ML, Rekosh D. 2015. The HIV-1 Rev response element (RRE) adopts alternative conformations that promote different rates of virus replication. Nucleic Acids Res 43:4676-4686.

38. Borg KT, Favaro JP, Arrigo SJ, Schmidt M. 1999. Activation of a cryptic splice donor in human immunodeficiency virus type-1. J Biomed Sci 6:45-52.

39. Wentz MP, Moore BE, Cloyd MW, Berget SM, Donehower LA. 1997. A naturally arising mutation of a potential silencer of exon splicing in human immunodeficiency virus type 1 induces dominant aberrant splicing and arrests virus production. J Virol 71:8542-8551.

40. Cocquet J, Chong A, Zhang G, Veitia RA. 2006. Reverse transcriptase template switching and false alternative transcripts. Genomics 88:127-131.

41. Houseley J, Tollervey D. 2010. Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. PLoS One 5:e12271.

42. Vega Y, Delgado E, de la Barrera J, Carrera C, Zaballos A, Cuesta I, Marino A, Ocampo A, Miralles C, Perez-Castro S, Alvarez H, Lopez-Miragaya I, Garcia-Bodas E, Diez-Fuertes F, Thomson MM. 2016. Sequence Analysis of In Vivo-Expressed HIV-1 Spliced RNAs Reveals the Usage of New and Unusual Splice Sites by Viruses of Different Subtypes. PLoS One 11:e0158525.

43. Sherrill-Mix S, Ocwieja KE, Bushman FD. 2015. Gene activity in primary T cells infected with HIV89.6: intron retention and induction of genomic repeats. Retrovirology 12:79.

44. Lassen KG, Bailey JR, Siliciano RF. 2004. Analysis of human immunodeficiency virus type 1 transcriptional elongation in resting CD4+ T cells in vivo. J Virol 78:9105-9114.

45. Kao SY, Calman AF, Luciw PA, Peterlin BM. 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature 330:489-493.

46. Caputi M, Zahler AM. 2002. SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. Embo j 21:845-855.

47. Benko DM, Schwartz S, Pavlakis GN, Felber BK. 1990. A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. J Virol 64:2505-2518.

48. Baker KE, Parker R. 2004. Nonsense-mediated mRNA decay: terminating erroneous gene expression. Curr Opin Cell Biol 16:293-299.

49. Viglianti GA, Sharma PL, Mullins JI. 1990. Simian immunodeficiency virus displays complex patterns of RNA splicing. J Virol 64:4207-4216.

50. Reinhart TA, Rogan MJ, Haase AT. 1996. RNA splice site utilization by simian immunodeficiency viruses derived from sooty mangabey monkeys. Virology 224:338-344.

51. Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. Rna 14:802-813.

52. Mandal D, Exline CM, Feng Z, Stoltzfus CM. 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. J Virol 83:6067-6078.

53. Saliou JM, Bourgeois CF, Ayadi-Ben Mena L, Ropers D, Jacquenet S, Marchand V, Stevenin J, Branlant C. 2009. Role of RNA structure and protein factors in the control of HIV-1 splicing. Front Biosci (Landmark Ed) 14:2714-2729.

54. Mueller N, Berkhout B, Das AT. 2015. HIV-1 splicing is controlled by local RNA structure and binding of splicing regulatory proteins at the major 5' splice site. J Gen Virol 96:1906-1917.

55. Madsen JM, Stoltzfus CM. 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.

56. Exline CM, Feng Z, Stoltzfus CM. 2008. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol 82:3921-3931.

57. Dyhr-Mikkelsen H, Kjems J. 1995. Inefficient spliceosome assembly and abnormal branch site selection in splicing of an HIV-1 transcript in vitro. J Biol Chem 270:24060-24066.

58. Damier L, Domenjoud L, Branlant C. 1997. The D1-A2 and D2-A2 pairs of splice sites from human immunodeficiency virus type 1 are highly efficient in vitro, in spite of an unusual branch site. Biochem Biophys Res Commun 237:182-187.

59. Poddar S, Eul J, Patzel V. 2014. Homologous SV40 RNA trans-splicing: Special case or prime example of viral RNA trans-splicing? Comput Struct Biotechnol J 10:51-57.

60. Caudevilla C, Da Silva-Azevedo L, Berg B, Guhl E, Graessmann M, Graessmann A. 2001. Heterologous HIV-nef mRNA trans-splicing: a new principle how mammalian cells generate hybrid mRNA and protein molecules. FEBS Lett 507:269-279.

CHAPTER 3: GLOBAL SYNONYMOUS MUTAGENESIS IDENTIFIES *CIS*-ACTING RNA ELEMENTS THAT REGULATE HIV-1 SPLICING AND REPLICATION[2]

Overview

The ~9.5 kilobase HIV-1 genome contains RNA sequences and structures that control many aspects of viral replication, including transcription, splicing, nuclear export, translation, packaging and reverse transcription. Nonetheless, chemical probing and other approaches suggest that the HIV-1 genome may contain many more RNA secondary structures of unknown importance and function. To determine whether there are additional, undiscovered *cis*-acting RNA elements in the HIV-1 genome that are important for viral replication, we undertook a global silent 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 novel splicing inhibitory RNA signals that suppressed the use of cryptic or canonical splice sites. Some sequences that regulated splicing were diffusely distributed, while others could be mapped to discrete elements, proximal or distal to the affected splice site(s). Overall, our data indicate 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.

[2] This chapter is under revision at Plos Pathogens with the above title and the following authors: Matthew Takata, Steven Soll, Ann Emery, Daniel Blanco-Melo, Ronald Swanstrom and Paul D. Bieniasz.

Introduction

The HIV-1 genome contains a variety of RNA elements that have important cis-acting functions (1, 2). 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. Known cis-acting RNA elements in the HIV-1 genome that lie within protein-coding sequences include splice donors, acceptors, and branch points (3), splicing regulatory elements that enhance or inhibit the use of proximal splice sites (4), the Rev-responsive element (5, 6), the central polypurine tract and termination sequence (7), the Gag-Pro-Pol ribosomal frameshift regulatory element (8) and components of the viral genome packaging signal (9, 10). 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. 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 (11-13). 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 useful function in HIV-1 replication (11, 13). 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 (11). However, no evidence for such a phenomenon was found in an analysis of SIVmac (14).

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 (15). 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 (16, 17). However, the precise nature of defects induced by synonymous mutations are unclear, and possibly pleiotropic.

RNA sequence and structure must play a key role in a particular aspect of HIV-1 replication, namely alternative splicing. Indeed SHAPE analysis revealed a novel stem loop that influences splicing

(14, 18)**.** Nevertheless, the regulatory mechanisms controlling alternative splicing are incompletely understood (4, 19). HIV-1 employs four salient splice donors (D1, D2, D3, D4) and eight acceptors (A1, A2, A3, A4a, b, c, A5 and A7) to generate a large number of mRNAs that enable expression of nine viral open reading frames in a temporally biphasic manner (20). 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 (19, 21). Splice-site utilization is regulated by deviation from consensus splicing signals as well as regulatory elements consisting of exonic and intronic splicing silencer (ESS and ISS) elements, and exonic splicing enhancer (ESE) elements (4, 19, 21-30). 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 3.1) (3, 4). In addition, some HIV-1 mRNAs include short noncoding exons (SX1 (A1-D2)) and (SX2 (A2-D3)) 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 (18, 31), have a profound impact on the levels of viral proteins that are synthesized and thus on viral replication.

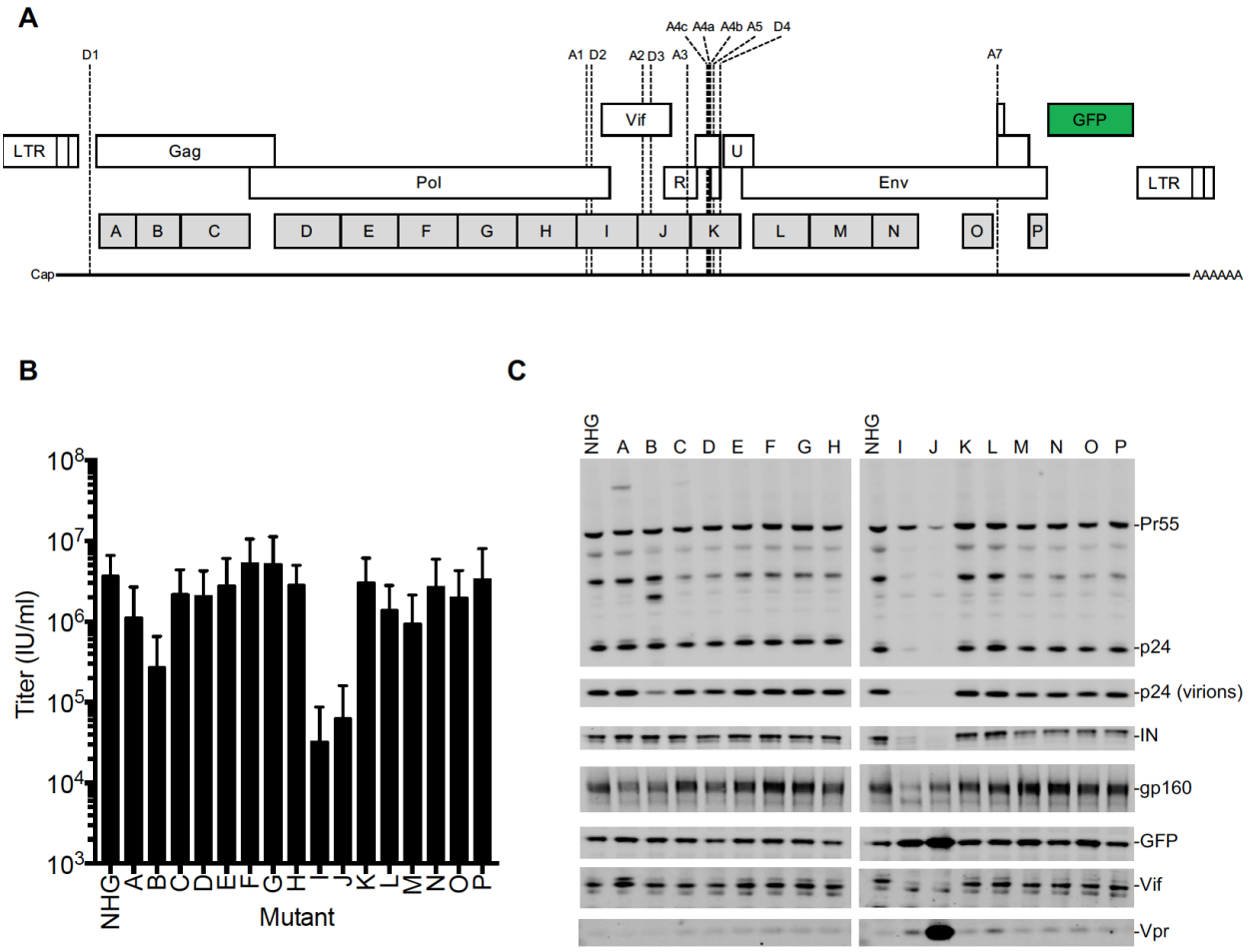

Figure 3.1. Design and analysis of panel of synonymously mutated HIV-1 viruses.
(A) Schematic of HIV-1 proviral DNA, indicating open reading frames, splice sites, and blocks of nucleotides that were synonymously mutated in the 16 proviral plasmids (A-P) (B) Single-cycle infectious titers (measured using MT4 cells) 48h following transfection of 293T cells with each of the WT(HIV-1_{NHG}) and mutant (A-P) proviral plasmids, mean,+/-sd. (n=3). (C) Western blot analysis of protein expression and virion release 48h after transfection of 293T cells with each of the WT (HIV-1_{NHG}) and mutant (A-P) proviral plasmids.

To identify and catalog regions of the HIV-1 genome that contain critical but unidentified cis-acting RNA elements that impact splicing and other functions, we employed a global synonymous mutagenesis strategy. Sixteen mutant proviruses, each containing blocks of synonymous mutations covering nearly the entire HIV-1 protein coding sequence, were examined. Despite encoding identical proteins, and despite the fact that all known *cis*-acting sequences were maintained in an intact form, some viruses were completely incapable of a spreading infection. Other mutants displayed apparently normal fitness. Some defective mutants, termed Group 3 mutants, displayed near normal splicing but were defective in spreading replication assays. In these cases, second site revertants could not be recovered. In this study we focused on Group 2 mutants that displayed aberrant splicing. We found that

replication competence could be recovered by second-site revertant point mutations that were often, but not always, proximal to splice sites. Sequences responsible for aberrant splicing were sometimes diffusely distributed, but sometimes could be mapped to discrete elements of 20- 120 nucleotides and suggest the existence of novel forms of splicing-suppressive RNA signals. Overall, this study demonstrates that novel elements in the HIV-1 genome determine the fate and splicing of HIV-1 RNA and thus the ability of HIV-1 to replicate.

Results

Design and synthesis of synonymously mutated HIV-1 strains

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. 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.

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-1_{NHG}) (32) 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).

Protein expression and replication properties of synonymously mutated HIV-1 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). 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 3.2). 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.

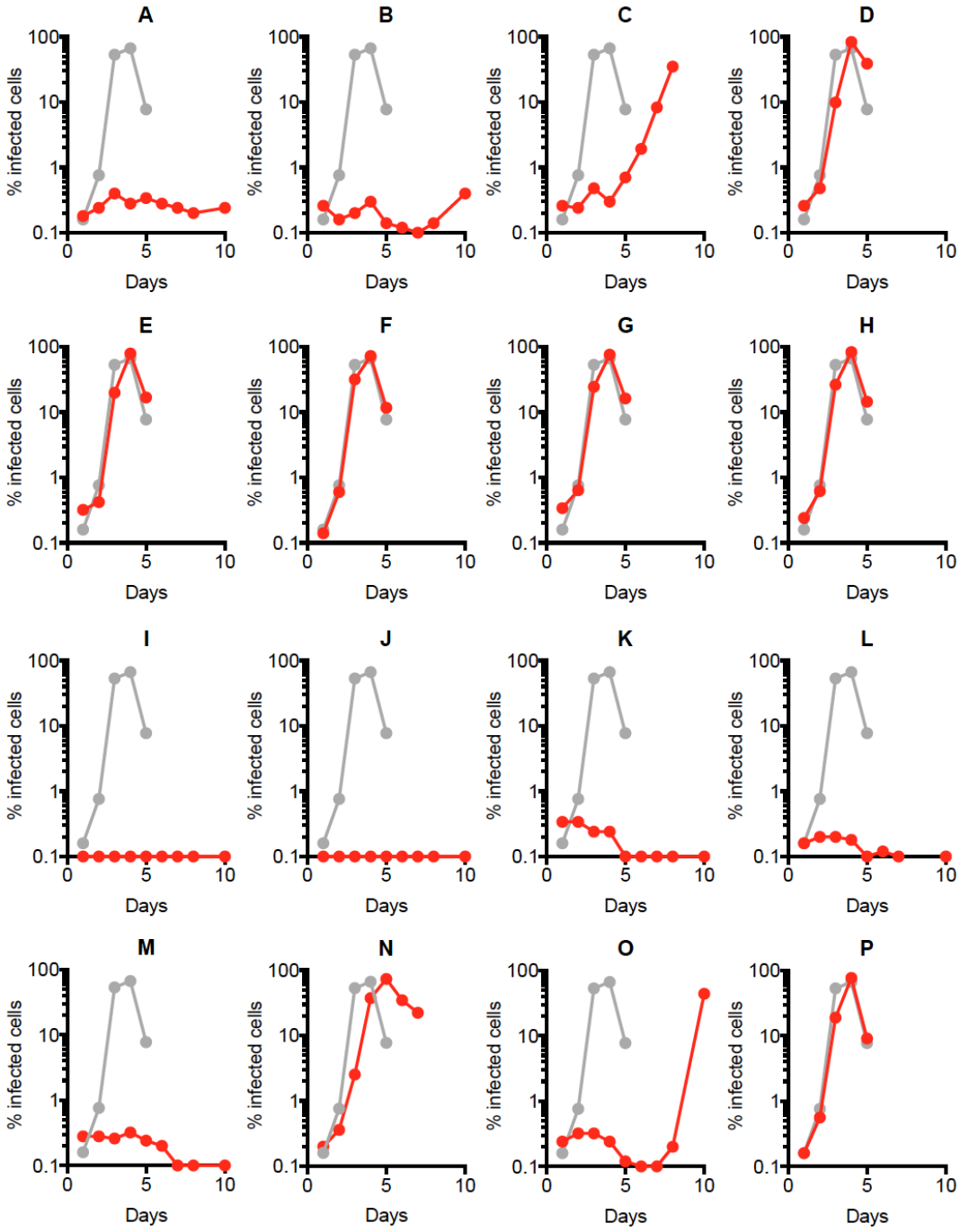

Figure 3.2. Spreading replication properties of mutant viruses.
(A-P) MT4 cells were infected with the indicated virus (harvested from the supernatant of 293T cells transfected with each of the WT (HIV-1_{NHG}) mutant (A-P) proviral plasmids at an MOI of 0.002. Aliquots of infected cells were withdrawn each day, fixed in 4% PFA and the proportion of infected cells determined by FACS analysis of GFP expression.

Assays for splicing perturbations in replication-defective HIV-1 mutants

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. 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. 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 (18) to globally quantify the relative utilization of the various splice donors and acceptors in the mutant viruses (Figure 3.3A-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. These two assays yielded results that were in good agreement (see below). Of the canonical splice acceptors in the central portion of the genome (A1, A2, A3 A4a, b, c, and A5, Figure 3.3A), the WT HIV-1_{NHG} most frequently spliced to A5, with lower levels of splicing to A1, A2, A3, A4a, b, c (Figure 3.3A, C, F).

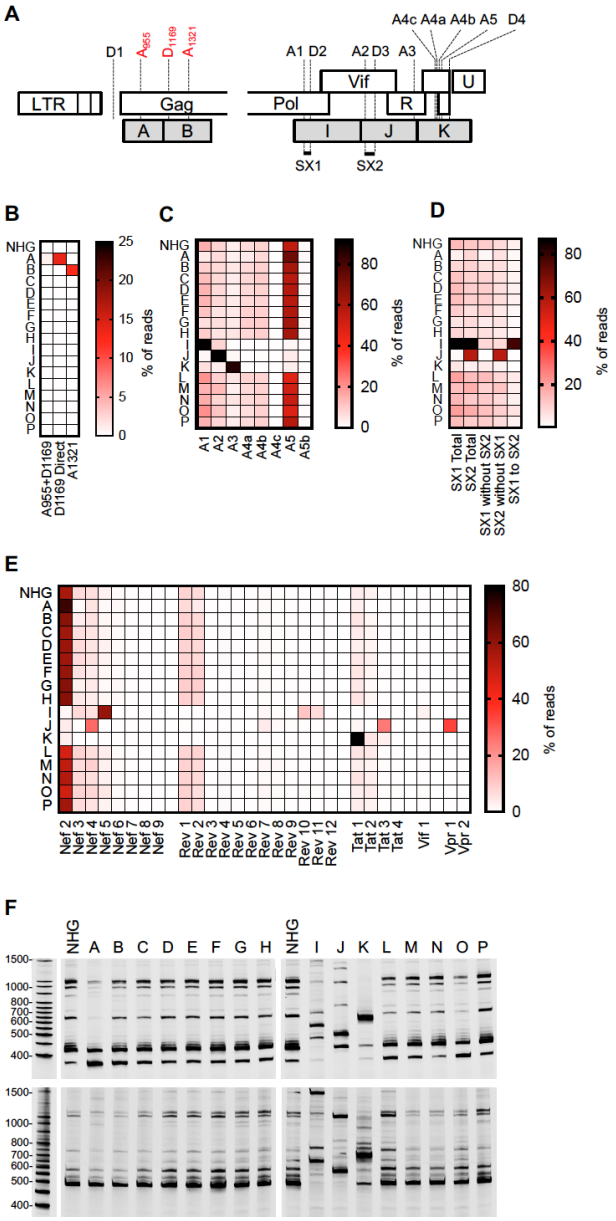

Figure 3.3. Analysis of HIV-1 splicing in WT and synonymously mutated HIV-1.
(A) Schematic of segments of HIV-1 proviral DNA focused on mutants exhibiting perturbed splicing. Canonical splice sites (black) and cryptic splice sites (red) are indicated, as are blocks of nucleotides that were synonymously mutated in the viruses exhibiting perturbed splicing. (B-D) 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 heatmaps (D). For panel (C) only direct splicing to the indicated acceptor sites is indicated in the heatmaps 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, was used along with an antisense primer positioned either 3' to A7 or 3' to D4 (labelled with IRDye 800) was 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.

Activation of cryptic splice sites in mutants A and B (Group 2a)

A splicing defect was observed for mutants A and B, which contained synonymous changes toward the 5' end of the HIV-1. For each of these mutants, the relative levels of canonical splice site utilization was 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 3.3B, C). 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 3.3A, B). While A955 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 3.3B). For unknown reasons, 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 (AE, unpublished observations).

For mutant B, the major perturbation was activation of a cryptic splice acceptor at position 1321 (A_{1321}, Figure 3.3B). This splicing event, involving ~12% of mRNAs in the 1.8 kb class, appeared to enable a cascade of further alternative splicing events involving several downstream cryptic splice sites (AE unpublished observations). However, the most obvious outcome of the aberrant splicing event was the generation of a truncated ~40 kD Gag protein (Figure 3.2B). 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).

Overuse of canonical splice acceptor sites in mutants I, J and K (Group 2b)

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 A1, A2 and A3 respectively, at the expense of direct splicing to downstream (3') acceptors (Figure 3.3C). In the case of mutant I, 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 3.3D). 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 3.3D). 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 3.3D). 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.

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, *nef*2 was the dominant mRNA species (Figure 3.3E). Conversely, in mutant I, *nef*5 was the dominant mRNA, while in mutant J, *nef*4, *tat*3 and *vpr*1 were overrepresented (Figure 3.3E). 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, *tat*1 mRNA was over-represented (Figure 3.3E), but the overall levels of protein expression were not greatly affected in transfected cells.

Viruses with blocks of synonymous mutations with three different phenotypes that map to distinct regions of the HIV-1 genome

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 3.3B-E). These viruses appeared to express a normal complement of viral proteins in transfected cells (Figure 3.1C), thus the underlying basis for their replication defect was initially unexplained. Overall, therefore, analyses of viral replication and RNA splicing led to the classification of the synonymously mutated viruses into three groups (Figure 3.4A): 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 3.4A). 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 the Gag ORF, while Group 2b viruses (I,

J, K) carried mutations or in the central portions of the genome, within the accessory genes (Figure 3.4A). Group 3 viruses (L, M, O) that were defective but exhibited near-normal splicing carried mutations in the *env* gene (Figure 3.4A).

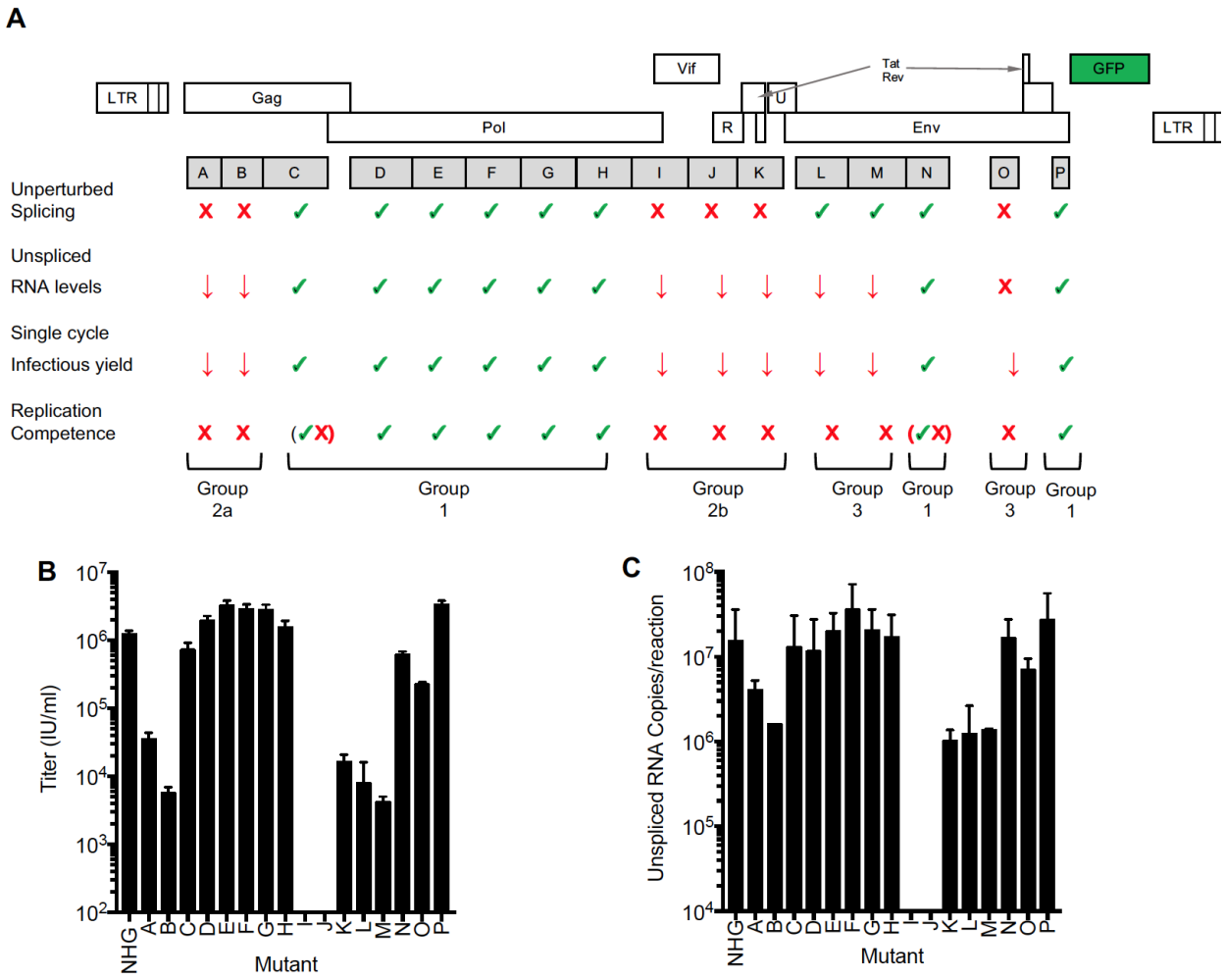

Figure 3.4. Phenotypes of synonymously mutated HIV-1 viruses.
(A) Summary of the properties of HIV-1 viruses carrying blocks of nucleotides that were synonymously mutated (A-P), assessed in transfection (perturbations of splicing), single-cycle replication (unspliced RNA levels and infectious virus yield) and spreading replication (replication competence) 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. (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.

Single-cycle replication assays suggest that reduced levels of unspliced viral RNA underlie replication defects in synonymously mutated HIV-1

We next explored the discrepancy in the abilities of certain mutant viruses to yield infectious virions after transfection of 293T cells, but were unable to spread in MT4 cells (Figure 3.2). We examined this discrepancy by carrying out experiments that analyzed a single cycle of replication in MT4 cells. Specifically, we infected cells with each of the mutant viruses at an MOI of 1 and quantified infectious virion release (Figure 3.4B) as well as levels of full length unspliced viral mRNA, during the first cycle of infection (Figure 3.4C). These experiments revealed significant deficiencies in infectious virion yield from infected cells for mutants A, B, K, L and M, that were not evident or less evident when virions were generated after transfection of 293T cells with a plasmid containing full length viral DNA (Figure 3.4B, Figure 3.1B). Moreover, these deficiencies, and the inability of each mutant virus to support a spreading infection, correlated with reduced levels of unspliced HIV-1 RNA in infected cells (Figure 3.4A, C). These observations suggest that the underlying defect in each of the replication defective mutants is a deficit in maintaining the level of unspliced RNA. This deficiency could result from excessive or aberrant splicing for Group 2 mutants or via unknown mechanisms for Group 3 mutants. It therefore appears likely that the deficit in unspliced RNA can be overcome, for some mutants, through the overexpression that results from transient viral DNA transfection in 293T cells. However, each of the mutants with reduced levels of unspliced RNA in single cycle-infected cells was profoundly defective in a spreading replication assay (Figure 3.4A). For Group 3 viruses, attempts to generate replication competent second-site revertants were unsuccessful, even after long term passage. Moreover, attempts to map discrete elements within the mutated blocks responsible for replication were also unsuccessful for Group 3 viruses.

Mapping elements responsible for splicing perturbations in Group 2 viruses

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.

Redundant HIV-1 RNA sequences mediate suppression of cryptic splice sites that are activated in Group 2a mutant viruses

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 mutant viruses (Figure 3.5A, B). In the case of mutant A, passage in MT4 cells yielded a revertant virus that replicated well and contained two nucleotide substitutions relative to the A mutant parent, one of which (C819T) was responsible for the revertant replication phenotype (Figure 3.5A). 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 defect 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 3.5C). 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 3.5D) 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 through D1, rather than on the cryptic D_{1169} that was activated in mutant A.

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 3.5B) 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 3.5E). Analysis of the G1326A revertant in the NGS splicing assay indicated reduced use of the cryptic acceptor A_{1321} (from 12% to <1%, Figure 3.5C), and both revertants exhibited reduced expression of the truncated Gag protein and increased virion release (not

shown). 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 D_{1169}. Conversely, the G1326 and T1311 mutations (and maybe others) in mutant B virus may have created or revealed a splicing factor binding site that was otherwise limiting for the use of the cryptic D_{1169}.

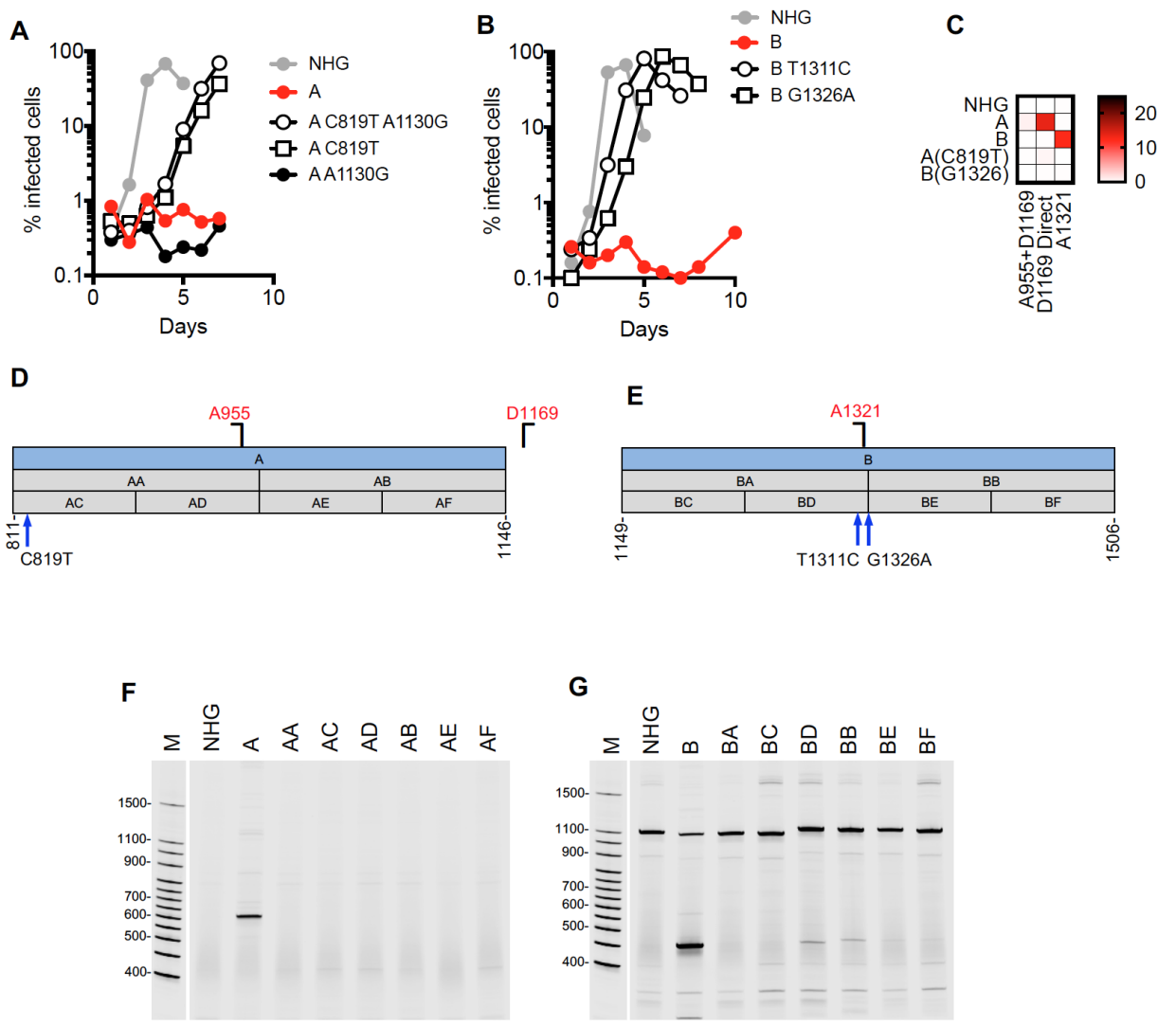

Figure 3.5. Activation of cryptic splice sites by synonymous mutations in Gag.
(A, B) MT4 cells were infected with the indicated virus (harvested from the supernatant of 293T cells transfected proviral plasmids representing each of the WT (HIV-1_{NHG}), mutant (A, B and revertants thereof) at an MOI 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 indicated at the bottom of the heatmap.(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.

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 3.5D, 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 3.5F, 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 3.5F, 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, though reversion of the effect could be observed with single mutations.

Multiple novel *cis*-acting RNA elements suppress HIV-1 splicing at A1 and A2

For mutant I, which exhibits overuse of A1 and A2 (and the corresponding 3' donors D2 and D3), 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 3.6A). Both IA and IB mutants also exhibited splicing defects primarily manifested as overuse of A1, but these defects were less complete, in that some degree of direct splicing to downstream acceptors was present in both IA and IB (Figure 3.6B). Notably, IA exhibited direct oversplicing to A1 only while IB exhibited direct oversplicing to both A1 and (to some degree) to A2 (Figure 3.6B). 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 3.6C, D). In the case of IA, a T4904A mutation occurred in the A1 polypyrimidine tract (Figure 3.6A, C). This caused a profound reduction of splicing to A1, and likely as a consequence, reduction of the inclusion of SX1 (A1-D2) in spliced RNAs (Figure 3.6B). In fact, other than underuse of A1 and reduced inclusion of SX1 (A1-D2), the IA_{T4904A} revertant had a near normal splicing pattern. For IB the situation was more complex; a revertant (G4912A) was recovered after passage, precisely at the A1 acceptor and abolished the use of A1, and consequently the inclusion of SX1 (A1-D2) into spliced mRNA (Figure 3.6A, B, D). However, significant

overuse of A2, and consequent inclusion of SX2 (A2-D3) remained evident in the IB_{G4912A} revertant (Figure 3.6B). Indeed, overuse of A2 was more prominent in the IB_{G4912A} revertant than in IB, perhaps because of the functional removal of A1 (Figure 3.6B). It was therefore apparent that elements within IA suppress splicing at A1, while elements within IB suppress splicing at both A1 and A2.

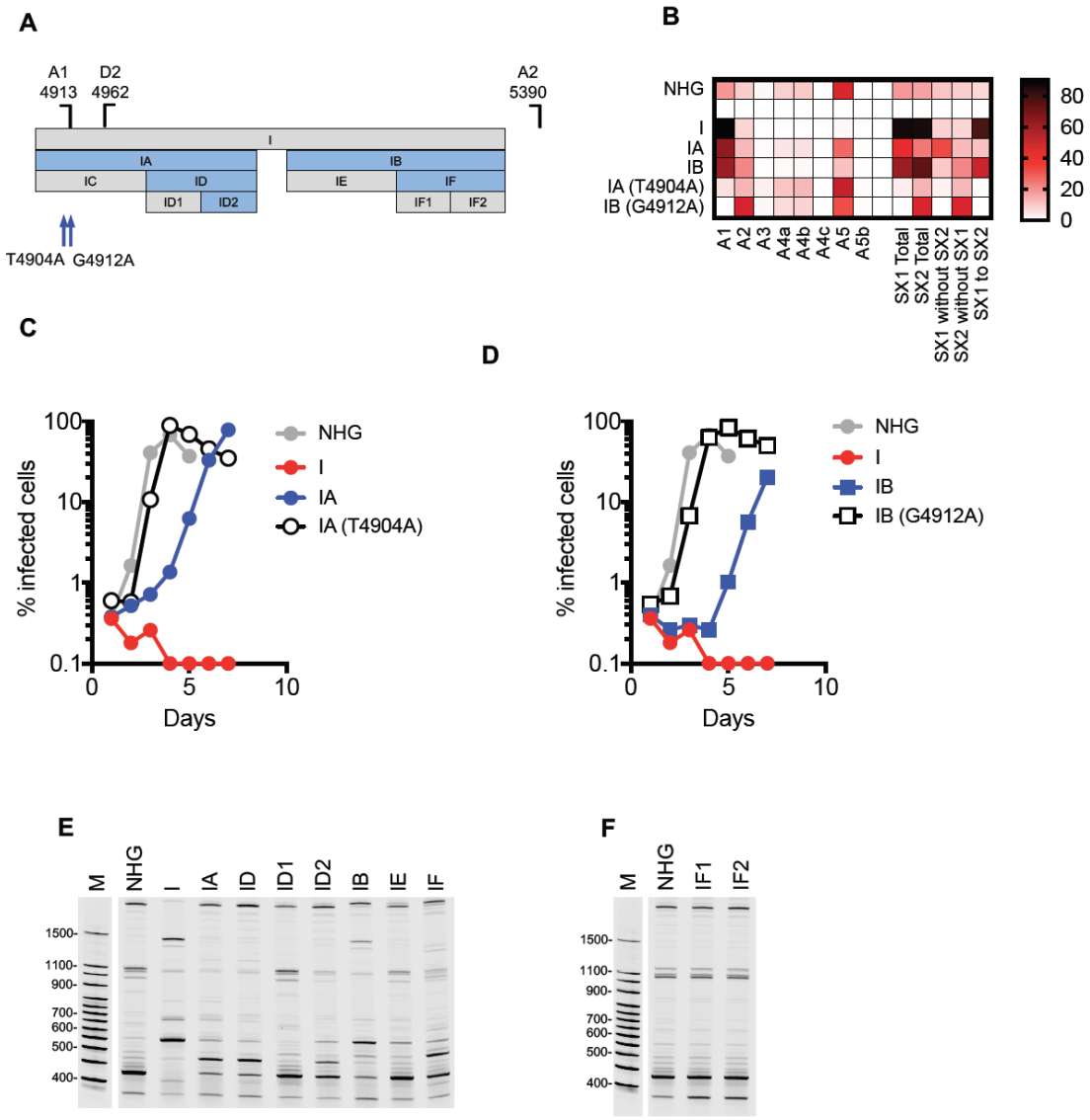

Figure 3.6. Activation of canonical splice acceptor sites (A1 and A2) by synonymous mutations in mutant I.

(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 acceptors or inclusion of the short exons (SX1 and/or SX2) indicated at the bottom of the heatmap. (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 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.

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 3.6A) revealed that aberrant splicing was conferred by the ID element (Figure 3.6E). Thereafter, when ID was subdivided into ID1 and ID2, it was evident that the controlling element resided primarily within ID2 (Figure 3.6E). Thus, this analysis revealed a ~48 nucleotide sequence that contains a novel ESS element that suppresses splicing at A1.

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 3.6A), 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 3.6E). Thus even though the splicing of IE appeared normal (Figure 3.6E), mutations in IE made some contribution to the defects present in IB. Subdivision of IF into IF1 and IF2 yielded viruses with a normal pattern of splicing (Figure 3.6F). Thus, it was evident that multiple elements acting together in IB, distributed over IE, IF1 and IF2 regulate splicing at A1 and A2 and their overall contributions could not be mapped to a small (<100 nucleotide) ESS 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 3.7A). Although there was some degree of splicing perturbation in JB (Figure 3.7B), 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 3.7C). Therefore, we did not attempt to select JB revertants. Conversely, JA recapitulated the splicing perturbation observed in J, and was replication defective (Figure 3.7B, D). Extended passage of mutant JA yielded a revertant mutation (G5463A) that enabled replication (Figure 3.7D). Notably, the reversion mutation was precisely at, and inactivated donor D3 (Figure 3.7A), but also abolished splicing to A2 (Figure 3.7B). Thus, D3 appears to be required for splicing to A2, even though 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 mutant J mutant sequences were distal (~440 to 890 nucleotides) to A1 (Figure 3.7A, B).

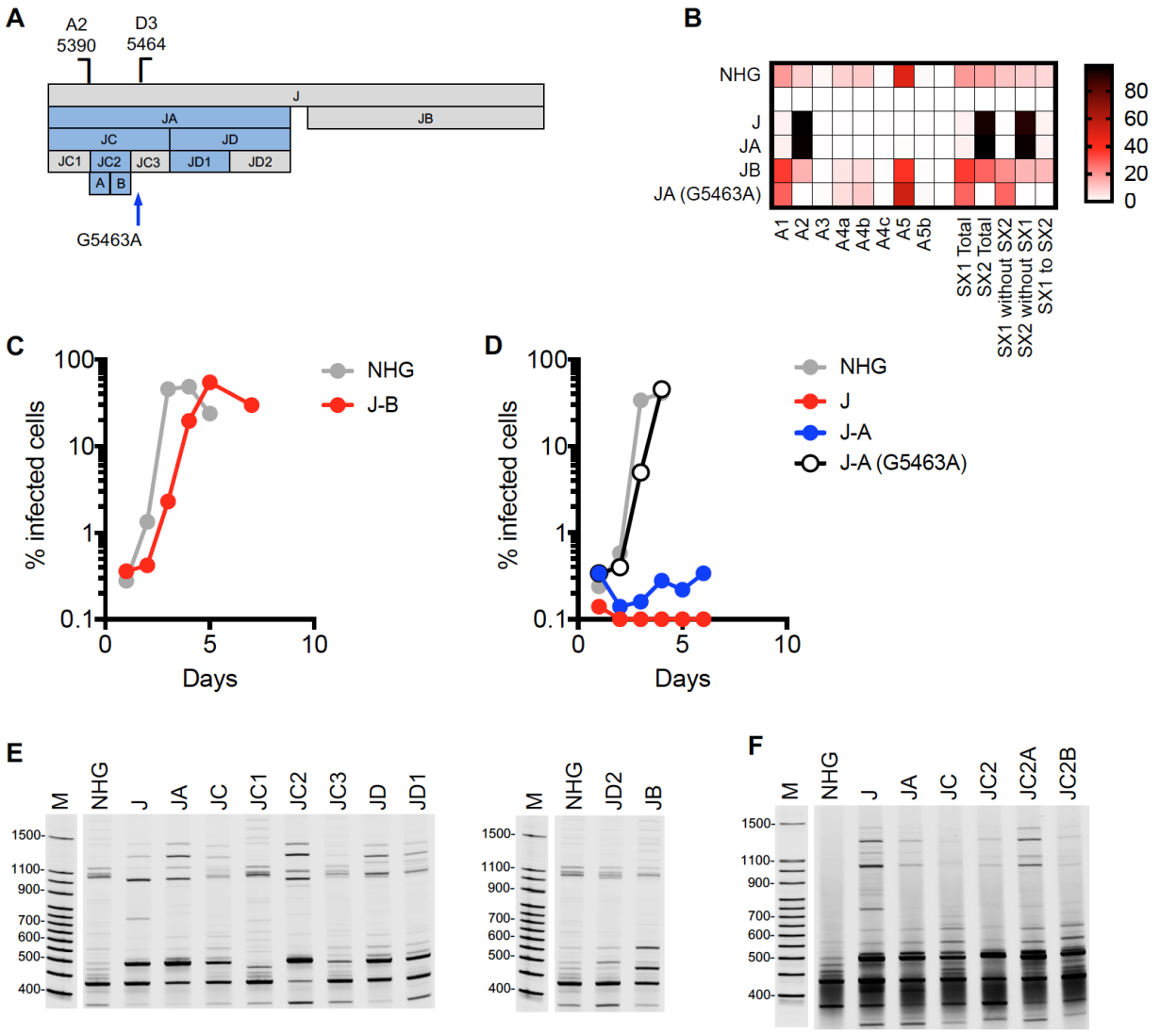

Figure 3.7. Activation of canonical splice acceptor site A2 by synonymous mutations in mutant J. (A) Schematic representation of the mutant blocks of nucleotides in HIV-1 mutant J, indicating positions of mutant derivatives (JA, JB, JC….etc.), and the positions of splice sites and revertant mutant sites (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. (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.

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 3.7A) which both exhibited some degree of perturbed splicing (Figure 3.7E). Further subdivision of JC into JC1, JC2 and JC3 clearly suggested that the 20 nucleotide JC2 segment contained an ESS element whose mutation was primarily responsible for the perturbed splicing in JC (Figure 3.7E), 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 3.7F). Division of the JD segment into JD1 and JD2 clearly revealed another ESS element within the 46 nucleotide JD1 segment, that when mutated yielded a splicing pattern similar to that of the J mutant virus (Figure 3.7E). Thus, multiple mutations within the JA fragment, contained within the segments JC2 and JD1 were capable of causing splicing defects similar to those observed in the J mutant.

A novel cis-acting RNA element inhibits HIV-1 splicing at A3

For mutant K (Figure 3.8A), which exhibits overuse of A3 (Figure 3.8B), it proved straightforward to recover a revertant mutant virus through passage that corrected the splicing defect and replicated well (Figure 3.8B, C). This revertant contained two mutations, one of which (C5774T) was sufficient to restore replication to near WT kinetics (Figure 3.8C). This functional reversion mutation was 3 nucleotides from the 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 3.8B). To map sequences within K that were responsible for causing oversplicing, we divided the K mutant segment into two halves (KA and KB, Figure 3.8A) 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 3.8D). Then, further subdivision of KA (into KC and KD) revealed that KD contained the controlling element (Figure 3.8D). Finally, subdivision of KD (into KD1 and KD2) showed that KD2 contained the ESS element (Figure 3.8E), but further subdivision of KD2 (into KD2A and KD2B) showed that mutations in both of these KD2 components contributed its effect (Figure 3.8F). Thus a 23-nucleotide element (KD2) contained a novel ESS element responsible for inhibiting splicing at A3.

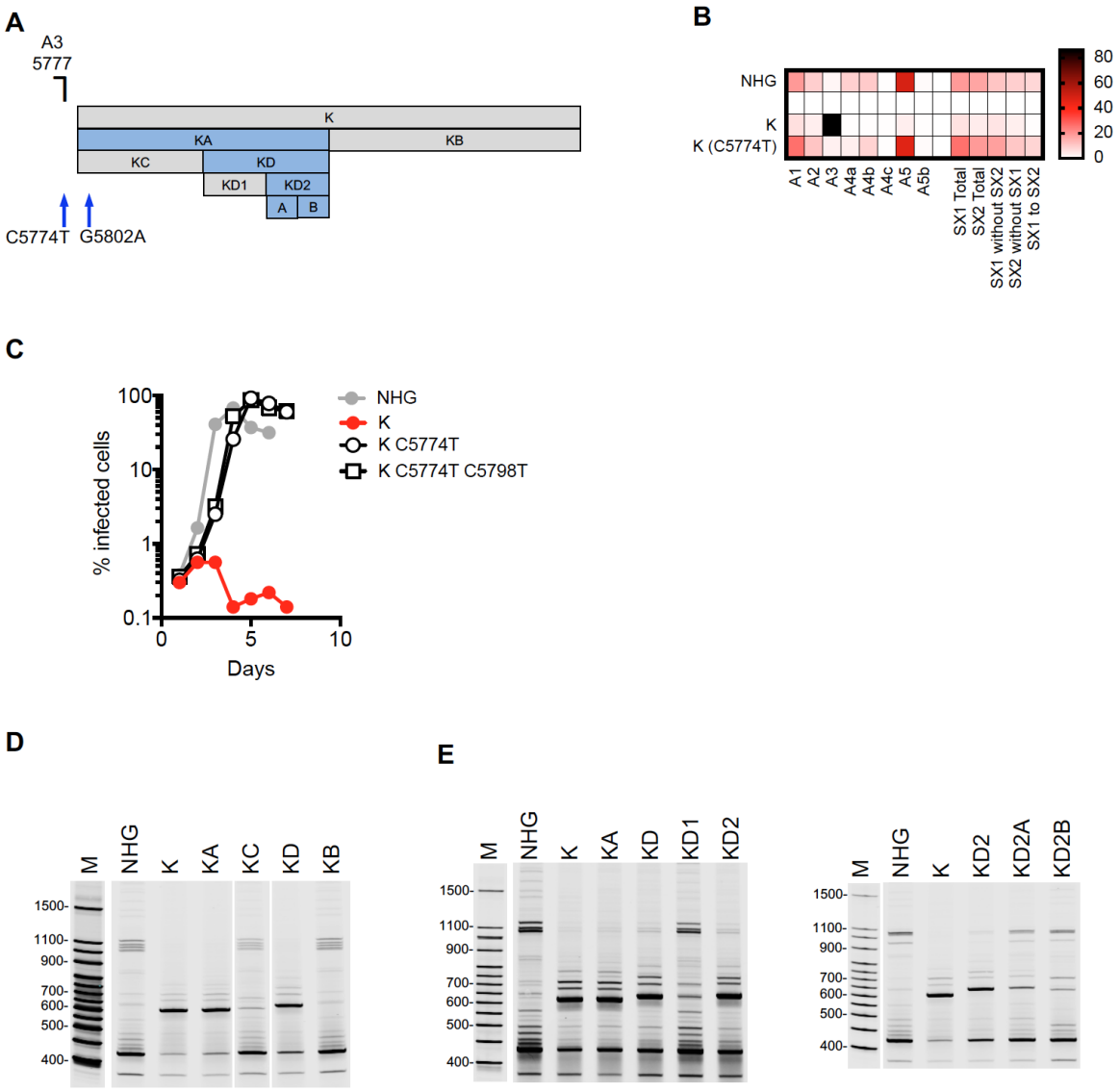

Figure 3.8. Activation of canonical splice acceptor site A3 by synonymous mutations in mutant K. (A) Schematic representation of the mutant blocks of nucleotides in HIV-1 mutant K, indicating positions of mutant derivatives (KA, KB, KC….etc.), and the positions of splice sites and revertant mutant sites (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. (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, 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.

Discussion

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 exist in Group 2 mutants that are essential for replication (however, see discussion of splicing regulation below).

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 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 targeted 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. Potential mechanisms by which cryptic splice sites were activated in Group 2a mutants are not obvious because

discrete elements that led to cryptic splice site suppression or activation could not be mapped. Simply reducing the magnitude of the deviation from WT sequence by 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 the cumulative effect of multiple perturbations to the *gag* nucleotide sequence. Moreover, in the case of mutant A, a single nucleotide revertant mutation that occurred distal to the activated cryptic splice site corrected the splicing defect. 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 (33), 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*. Clearly further work will be required to understand how the WT noncoding sequence avoids utilization of cryptic splice sites.

A detailed analysis of Group 2b mutants, that targeted the central portion of the genome, revealed several previously unidentified elements that apparently act as splicing regulatory signals to inhibit utilization of the canonical splice acceptors A1, A2, and A3. When combined with existing knowledge of splicing regulation at these sites (4, 19, 21-30), these and previous findings indicate a complex regulatory network of functional inputs that govern alternative splicing of the HIV-1 genome (Figure 3.9).

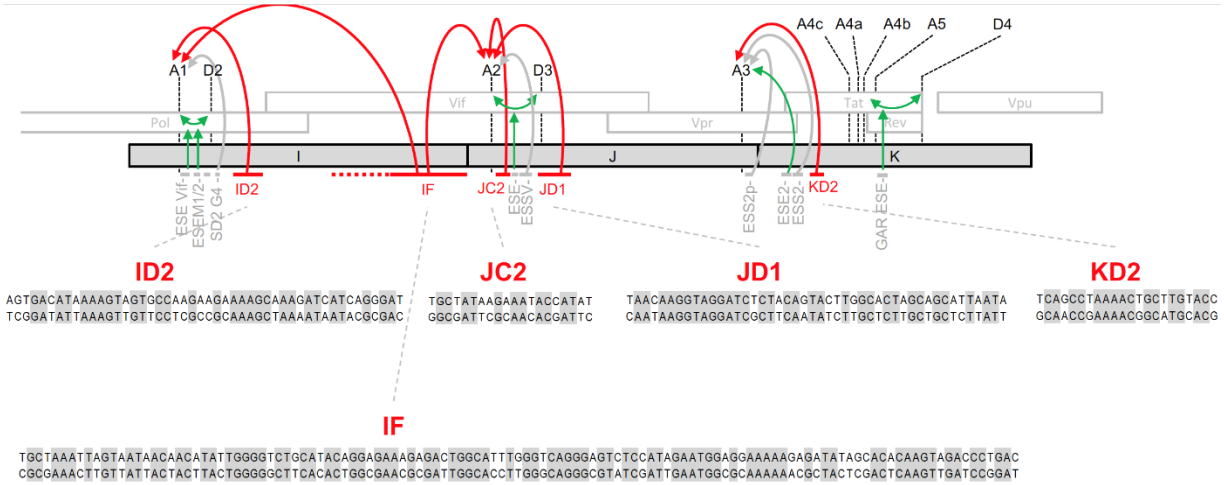

Figure 3.9. Summary of splicing control in HIV-1.
A schematic representation of the central portion of the HIV-1 genome with the positions of canonical splice sites indicated. Previously identified splicing control elements are indicated with grey bars and the splice sites on whose they inhibit and enhance are indicated with grey and green arrows, respectively. Newly identified splicing inhibitor sequences are indicated with red lines and the splice acceptors on which they act are indicated with red arrows. The sequences of the newly identified splicing inhibitory sequences (upper line = WT sequence, lower line = oversplicing mutant sequence) are shown in the lower part of the diagram.

The phenotypes of some mutant and revertant viruses (i.e. viruses with perturbed splicing whose replication was recovered by splice site mutations) is consistent with the notion that exon definition (coordinated recognition of a 5' splice acceptor and a 3' splice donor by the splicing machinery (34)) plays a key role in the regulation of HIV-1 alternative splicing. For example, in a mutant (IA) that exhibits elevated use of A1 and D2, the revertant (IA(4904)) occurs at A1 and exhibits reduced utilization of both A1 and D2. Similarly, in IB, that has elevated use of both A1 and A2, both D3 and D2 are also overused, the revertant mutation (IB(G4912A)) at A1 causes abolition of splicing at both A1 and D2 (no SX1 inclusion) while splicing at A2 and D3 (SX2 inclusion) remains elevated. Consistent with this notion, previous work has also shown that the efficiency with which D2 is recognized can affect the frequency of splicing at A1 (35). Strikingly, in the case of a mutant (JA) that exhibited profound oversplicing at A2, a reversion mutation that occurred at splice site D3 caused abolition of splicing at A2 (as well as D3). 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 (21). The notion that splice acceptor sites compete with each other is also consistent with our data, with 'cascading' effects based on exon definition. For example, in a mutant (JA) that underutilizes A1 and over utilizes A2, the revertant (JA(G54653A)) occurs at D3, abolishes use of A2, and

overutilization of A1 results. In an apparently reciprocal example of A1-A2 competition, abolition of A1 utilization in the revertant IB(G4912A), was accompanied by increased splicing at A2. In another example of apparent acceptor competition, the mutant K, exhibited oversplicing at A3 at the expense of splicing at A1, A2, A4b and A5, but 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 identified herein can have complex effects on overall splicing, through the propagation of their effects from one splice site to another. Clearly the 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 sequence elements that we have identified to inhibit splicing at the various splice acceptors (Figure 3.9) could work directly or indirectly. Specifically, they could act to directly inhibit access of the splicing machinery to the affected splice sites or indirectly through splice donors, inhibiting splicing by blocking 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 (36). 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 (36). 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 5' donors from 3' acceptors thereby inhibiting exon definition. Further work will be required to determine precisely how these elements inhibit HIV-1 splicing.

We note that 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. We note that these perturbations would be expected to lead to under expression or overexpression of Vif and Vpr, neither of which are essential for replication in this particular cell type. Thus, the requirement for a particular balance of HIV-1 mRNAs could be highly context dependent. In our experiments, replication defects that resulted from oversplicing 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.

In the case of another set of replication-defective mutants, that we termed 'Group 3' and had synonymous mutations in the *env* gene, we could not detect significant splicing defects. Nevertheless, reduced levels of unspliced RNA were present in Group 3 virus-infected cells, and reduced levels of infectious particles were generated in single cycle replication assays. It is possible that that some form of splicing perturbation that was not detected in our splicing assays occurred in Group 3 mutants. However, the fact that we could not recover replication competent revertants from Group 3 viruses, while other replication defective mutants with clear splicing defects (Group 2) readily yielded replication revertants, suggests some other defect is responsible for the loss of replication competence in Group 3 mutants. The precise nature of the defect in Group 3 mutants is currently under investigation.

Overall, our global synonymous mutagenesis experiment has revealed several cis-acting RNA sequences that have essential functions in HIV-1 replication. In particular, we have identified several novel signals in the HIV-1 genome that appear to act through suppression of canonical splice sites to regulate alternative splicing and maintain 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 these regulatory signals work in the context of HIV-1 may give insights into the general mechanisms by which alternative splicing is regulated and how splicing regulation evolves, as well as opportunities to intervene therapeutically in HIV-1 infection.

Material and Methods

Cell lines, viruses and infections

293T cells (ATCC) were grown in DMEM with 10% fetal bovine serum (Sigma). MT4 cells (NIH AIDS Reagent Repository) were cultured in RPMI supplemented with 10% fetal bovine serum. HIV-1_{NHG} and mutant derivatives containing a reporter GFP in place of *nef* were produced by transfection of 293T cells with the proviral plasmid using PEI. Virus titers were determined by FACS analysis of target MT4 cells. Dextran sulfate was added to inhibit reinfection at 18 hours post infection and the number of infected cells determined by FACS analysis of GFP expression 2 days post infection.

For spreading replication infections, $2x10^5$ MT4 cells in 2 mL of media were infected 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. For single cycle infections, cells were infected with HIV-1_{NHG} or mutants thereof, at an MOI of 1.0, and harvested 2 days post infection for western blot and qPCR analysis. Serial passage of mutant viruses was started by infecting a culture of $5x10^5$ MT4 cells at an MOI of 0.002 and time points were fixed in 4% PFA. Upon infection of the entire culture, 200 µL of cell free supernatant was filtered and used to inoculate a new culture of $5x10^5$ MT4 cells. After the final passage, the cells were collected and the DNA was extracted (Qiagen Tissue Mini Kit), then the mutated region was sequenced to determine the revertant mutation that had occurred.

Construction of mutant proviral plasmids

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 (Supplemental Table 1). Division of the original mutant 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.

PCR quantification of unspliced HIV-1 RNA

RNA was extracted from infected cells using Trizol and reverse transcribed using the SuperScript III reverse transcriptase (ThermoFisher). The resulting cDNA was used as a temple for quantitative real-time PCR using the ABI Fast RT-PCR system along with the Fast Start TaqMan Probe master mix.

Unspliced viral RNA was detected using a forward primer: 5'-GGACTTGAAAGCGAAAGGGA-3', a reverse primer: 5'-TCTCTCTCCTTCTAGCCTCCG-3' and a TaqMan probe 5'-GGGCGGCGACTGGTGAGT-3' targeting the major splice donor D1. Serial tenfold dilutions of known copy numbers of HIV-1_{NHG} plasmid was used to generate a standard curve.

Antibodies and western blotting

Cells were 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). Blots were probed with an HIV-1 anti-capsid antibody (183-H12-5C) obtained from the NIH AIDS reagent repository, a GFP antibody (G1546, Sigma), and HIV-1 anti-Env gp120 antibody (12-6205-1, American Research Products).

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.

Analysis of HIV-1 splicing using Primer ID-based deep sequencing

Determination of splice site utilization using the Primer ID-based deep sequencing assay was done substantially as described with minor modifications (18). 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 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNCAGTTCGGGATTGGGAGGTGGGTTGC for 1.8 kb spliced transcripts and GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNGCTACTATTGCTATTTGTATAGGTTGCATTACATG 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 (18) 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.

REFERENCES

1. Frankel AD, Young JA. 1998. HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 67:1-25.

2. Cullen BR. 1994. RNA-sequence-mediated gene regulation in HIV-1. Infect Agents Dis 3:68-76.

3. Purcell DF, Martin MA. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67:6365-6378.

4. Stoltzfus CM, Madsen JM. 2006. Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. Curr HIV Res 4:43-55.

5. Malim MH, Tiley LS, McCarn DF, Rusche JR, Hauber J, Cullen BR. 1990. HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. Cell 60:675-683.

6. Heaphy S, Dingwall C, Ernberg I, Gait MJ, Green SM, Karn J, Lowe AD, Singh M, Skinner MA. 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.

7. Charneau P, Clavel F. 1991. A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. J Virol 65:2415-2421.

8. Parkin NT, Chamorro M, Varmus HE. 1992. Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. J Virol 66:5147-5151.

9. Kuzembayeva M, Dilley K, Sardo L, Hu WS. 2014. Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. Virology 454-455:362-370.

10. Kutluay SB, Zang T, Blanco-Melo D, Powell C, Jannain D, Errando M, Bieniasz PD. 2014. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. Cell 159:1096-1109.

11. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460:711-716.

12. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol 6:e96.

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

14. Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM, 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 Pathog 9:e1003294.

15. Knoepfel SA, Berkhout B. 2013. On the role of four small hairpins in the HIV-1 RNA genome. RNA Biol 10:540-552.

16. Keating CP, Hill MK, Hawkes DJ, Smyth RP, Isel C, Le SY, Palmenberg AC, Marshall JA, Marquet R, Nabel GJ, Mak J. 2009. The A-rich RNA sequences of HIV-1 pol are important for the synthesis of viral cDNA. Nucleic Acids Res 37:945-956.

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

18. Emery A, Zhou S, Pollom E, Swanstrom R. 2017. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. J Virol 91.

19. Karn J, Stoltzfus CM. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med 2:a006916.

20. Holmes M, Zhang F, Bieniasz PD. 2015. Single-Cell and Single-Cycle Analysis of HIV-1 Replication. PLoS Pathog 11:e1004961.

21. Stoltzfus CM. 2009. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. Adv Virus Res 74:1-40.

22. Amendt BA, Hesslein D, Chang LJ, Stoltzfus CM. 1994. Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. Mol Cell Biol 14:3960-3970.

23. Jacquenet S, Mereau A, Bilodeau PS, Damier L, Stoltzfus CM, Branlant C. 2001. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. J Biol Chem 276:40464-40475.

24. Caputi M, Freund M, Kammler S, Asang C, Schaal H. 2004. A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. J Virol 78:6517-6526.

25. Bilodeau PS, Domsic JK, Mayeda A, Krainer AR, Stoltzfus CM. 2001. RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element. J Virol 75:8487-8497.

26. Kammler S, Otte M, Hauber I, Kjems J, Hauber J, Schaal H. 2006. The strength of the HIV-1 3' splice sites affects Rev function. Retrovirology 3:89.

27. Exline CM, Feng Z, Stoltzfus CM. 2008. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol 82:3921-3931.

28. Madsen JM, Stoltzfus CM. 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.

29. Asang C, Hauber I, Schaal H. 2008. Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. Nucleic Acids Res 36:1450-1463.

30. Brillen AL, Walotka L, Hillebrand F, Muller L, Widera M, Theiss S, Schaal H. 2017. Analysis of competing HIV-1 splice donor sites uncovers a tight cluster of splicing regulatory elements within exon 2/2b. J Virol doi:10.1128/jvi.00389-17.

31. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res 40:10345-10355.

32. Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, Gifford RJ, Rixon FJ, Bieniasz PD. 2013. Extreme genetic fragility of the HIV-1 capsid. PLoS Pathog 9:e1003461.

33. Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, Barton S, Hosic A, Florwick A, Santos J, Bolden NC, McCowin S, Case DA, Johnson BA, Salemi M, Telesnitsky A, Summers MF. 2015. RNA structure. Structure of the HIV-1 RNA packaging signal. Science 348:917-921.

34. De Conti L, Baralle M, Buratti E. 2013. Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip Rev RNA 4:49-60.

35. Mandal D, Exline CM, Feng Z, Stoltzfus CM. 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. J Virol 83:6067-6078.

36. Saliou JM, Bourgeois CF, Ayadi-Ben Mena L, Ropers D, Jacquenet S, Marchand V, Stevenin J, Branlant C. 2009. Role of RNA structure and protein factors in the control of HIV-1 splicing. Front Biosci (Landmark Ed) 14:2714-2729.

CHAPTER 4: CLIP-SEQ REVEALS A KEY ROLE FOR hnRNP H1 IN REGULATION OF HIV-1 ALTERNATIVE mRNA SPLICING[3]

Overview

Alternative splicing of HIV-1 mRNAs is a complex process that regulates transcript abundance, increases the coding potential of the compact viral genome and allows temporal regulation of viral gene expression. This process is in part regulated by the host heterogeneous nuclear ribonucleoproteins (hnRNPs) that bind to poorly defined sequences on HIV-1 RNAs and subsequently repress alternative splicing at nearby regulatory elements. In this study, we employed crosslinking immunoprecipitation coupled with next-generation sequencing (CLIP-seq) to identify the direct binding sites of several hnRNPs on viral RNAs in cells. We show that hnRNP A1, hnRNP A2 and hnRNP B1 proteins bind promiscuously throughout viral RNAs and require the presence of 'AGG' motifs for binding. In contrast, hnRNP H1 binding took place at well-defined purine-rich sequences comprised of four or more consecutive guanosines that are often surrounded by A-rich elements. Reassuringly, the sequence specificities of hnRNPs on viral RNAs largely mirrored those on cellular introns. In support of these findings, isolated RNA-binding domains of hnRNP H1 exhibited a clear preference for G-rich sequences *in vitro*. Notably, modulation of hnRNP H1 levels by RNAi and mutations within the most prominent hnRNP H1 binding site on viral RNAs decreased usage of splice acceptor A1 as revealed by an unbiased high throughput sequencing-based assay. Overall, our studies reveal novel cis-regulatory elements of HIV-1 splicing and provide the highly needed experimental framework to study an essential step of HIV-1 replication in greater detail.

[3] This chapter is ready for submission for publication pending the addition of some final experimental results. The authors are Sebla B. Kutluay, Ann Emery, Srinivas Penumutchu, Dana Townsend, Michaela Madison, Amanda Stukenbroeker, Chelsea Powell, David Jannain, Blanton S. Tolbert, Ron Swanstrom, and Paul D. Bieniasz.

Introduction

Following its integration into the host genome, genesis of new HIV-1 particles is initiated by the host RNA polymerase II-mediated transcription of a single primary polycistronic viral transcript that undergoes various levels alternative splicing. This process gives rise to over fifty alternative splice isoforms (1, 2), which are classified with respect to their sizes. The completely spliced 1.8 kb class of transcripts encode for Tat, Rev and Nef proteins, the incompletely spliced 4 kb class code for Vif, Vpr, Env, Vpu and a truncated form of Tat, and the incompletely spliced primary transcripts code for Gag and Gag-Pol proteins, and serve as the genomic RNA that is packaged into progeny virions. A novel 1kb class of transcripts was identified in a recent next-generation sequencing-based study (3); however, no functionality has been assigned to this class of transcripts yet. While completely spliced transcripts are easily exported out from the nucleus by canonical cellular pathways, export of incompletely spliced and unspliced transcripts is mediated by the viral Rev protein. Rev binds to a highly structured RNA element referred to as the Rev-responsive element (RRE) on viral RNAs and mediates their export by bridging between viral transcripts and the host karyopherin Crm1 (4). This mode of regulation allows temporal viral gene expression and provides sufficient time for the genesis of infectious virus particles (5).

Alternative HIV-1 splice isoforms are generated by the utilization of at least four splice donor (D1, D2, D3, D4) and eight acceptor sites (A1, A2, A3, A4c, A4a, A4b, A5 and A7) (1, 6). Recent deep-sequencing-based studies have revealed the existence of several other splice donor and acceptor sites that are generally utilized at lower frequencies (3). The intrinsic strengths of the donor and acceptor sites are thought to be governed by the extent of their homology to the consensus donor and acceptor sequences (2). For example, D1 is 100% homologous to the consensus sequence and is the most frequently utilized 5’ splice site. In fact, all downstream splicing events involve the usage of D1. In contrast, many of the other donor and acceptor sites are suboptimal and as such their utilization is thought to be regulated by other nearby cis-acting elements. Exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs) are such cis-acting elements that are bound by members of the serine–arginine-rich protein (SR protein) and heterogeneous nuclear ribonucleoprotein (hnRNP) families, which are generally thought to enhance or repress the usage of nearby splice sites, respectively.

Based primarily on *in vitro* binding and genetic assays, few members of the hnRNP protein family have been proposed to regulate HIV-1 alternative splicing. Among these, members of the hnRNP A/B family have been studied most extensively and were shown to bind to distinct sequences on viral RNAs, including the exonic splice silencer (ESS) ESS2 within *tat* exon 2 (7, 8), ESS3 located within *tat/rev* exon 3 (9, 10), ESSV located in the *vif*-coding exon (11) and an intronic splice silencer element within *tat* intron (10, 12, 13). Similarly, hnRNP H1 has been suggested to bind to ESS2p (14), a cryptic exon within the Env ORF (15, 16), a G-rich motif within *tat* exon 2 (17) and intron 3 (18). In addition, isolated studies have proposed that hnRNP D (19), hnRNP E (20) and hnRNP K (21) can regulate HIV-1 splicing. However, in many of these studies, regulation of HIV-1 splicing has been studied using *in vitro* and *in vivo* splicing reporters wherein viral sub-genomic fragments were moved to foreign genetic environments so as to determine how they influence the distribution of spliced reporter products. In addition, binding of a given hnRNP was assessed largely by *in vitro* binding assays using defined targets. Whether the presumed target sites can be bound by hnRNPs in infected cells and whether other binding sites on viral RNAs exist is unknown.

HnRNPs are modular proteins, a common feature of which is the presence of two or more RNA binding domains. Most hnRNPs also contain RGG boxes (Arg-Gly-Gly repeats) and auxiliary domains with distinct amino acid composition (i.e. glycine-rich, acidic or proline-rich domains) that are believed to be responsible for protein-protein, RNA-protein, and single-stranded DNA-protein interactions (22). As such, most hnRNPs can form homophilic interactions and heterophilic interactions with other hnRNPs.

Crosslinking-immunoprecipitation followed by deep sequencing (CLIP-seq) has emerged as a powerful method to identify the RNA targets of RNA-binding proteins at near-nucleotide resolution in physiological settings. Here we adapted the CLIP-seq methodology to determine viral RNA binding sites of some of the previously studied hnRNP proteins in physiological settings, and determine the effects of these interactions directly on viral splicing using next-generation sequencing-based methods. Our studies reveal that unlike previously suggested, hnRNP A/B group of proteins do not bind to distinct elements but rather bind promiscuously throughout viral RNAs and require the presence of ‘AGG’ motifs within binding sites. In contrast, hnRNP H1 was bound to well defined purine-rich sequences comprised of four or more consecutive guanosines frequently surrounded by A-rich elements. In support of these findings,

biochemical experiments revealed a clear preference for G-rich sequences in isolated RNA-binding domains of hnRNP H1. Overall, the binding specificities of hnRNPs on viral RNAs largely mirrored those on cellular introns. Notably, modulation of hnRNP H1 levels by RNAi and mutations within hnRNP H1 binding sites on viral RNAs decreased usage of splice acceptor A1 as revealed by an unbiased high throughput sequencing-based assay. Overall, our studies reveal novel cis-regulatory elements of HIV-1 splicing and provide the much needed experimental framework to study an essential step of HIV-1 replication in much greater detail.

Results

Generation of 3xHA-tagged hnRNP variants

To determine the binding sites of hnRNP A1, A2, B1, H1 and K proteins on HIV-1 RNAs, we first generated HEK293T-derived cell lines that stably express 3xHA-tagged versions of these proteins at lower than endogenous levels (Figure 4.1A). Importantly, these modifications did not affect HIV-1 Gag expression in cells (Figure 4.1B), particle release (Figure 4.1B) or infectivity (Figure 4.1C). We next determined whether hnRNP-RNA complexes that form in these cells can be isolated by a previously established CLIP-seq protocol (23). To this end, cells were grown in the presence of 4-thiouridine (4-SU) and UV-crosslinked to induce protein-RNA crosslinking at 4-SU-modified nucleotides. hnRNP-RNA adducts were immunoprecipitated, the bound RNA was end-labeled and complexes were visualized by autoradiography following separation by SDS-PAGE and transfer to nitrocellulose membranes. Reassuringly, all tagged hnRNPs were efficiently immunoprecipitated in complex with cellular RNAs (Figure 4.1D). To identify the binding sites of hnRNPs on viral RNAs, similar experiments were conducted in cells transfected with full-length HIV-1_{NL4-3} (subtype B). To account for subtype-specific differences on hnRNP binding, parallel experiments were conducted on cells transfected with HIV-1_{NDK} (subtype D) proviral plasmids. HnRNP-RNA complexes were isolated as above, and bound RNA molecules were further purified and the resulting libraries were subjected to high throughput sequencing. Following trimming of 5' and 3' adapters, the CLIP-seq reads were sequentially mapped to the human and viral genomes. Two independent libraries for each hnRNP and each virus were generated and are represented in Figure 4.1E. In general, the majority of sequencing reads mapped to the human genome and only a small fraction (~1% or less) were derived from viral RNAs (Figure 4.1E). Importantly, these

numbers reflect the relative abundance of viral and cellular mRNAs in similar settings (23), suggesting that hnRNPs bind with similar affinity to viral and cellular RNAs.

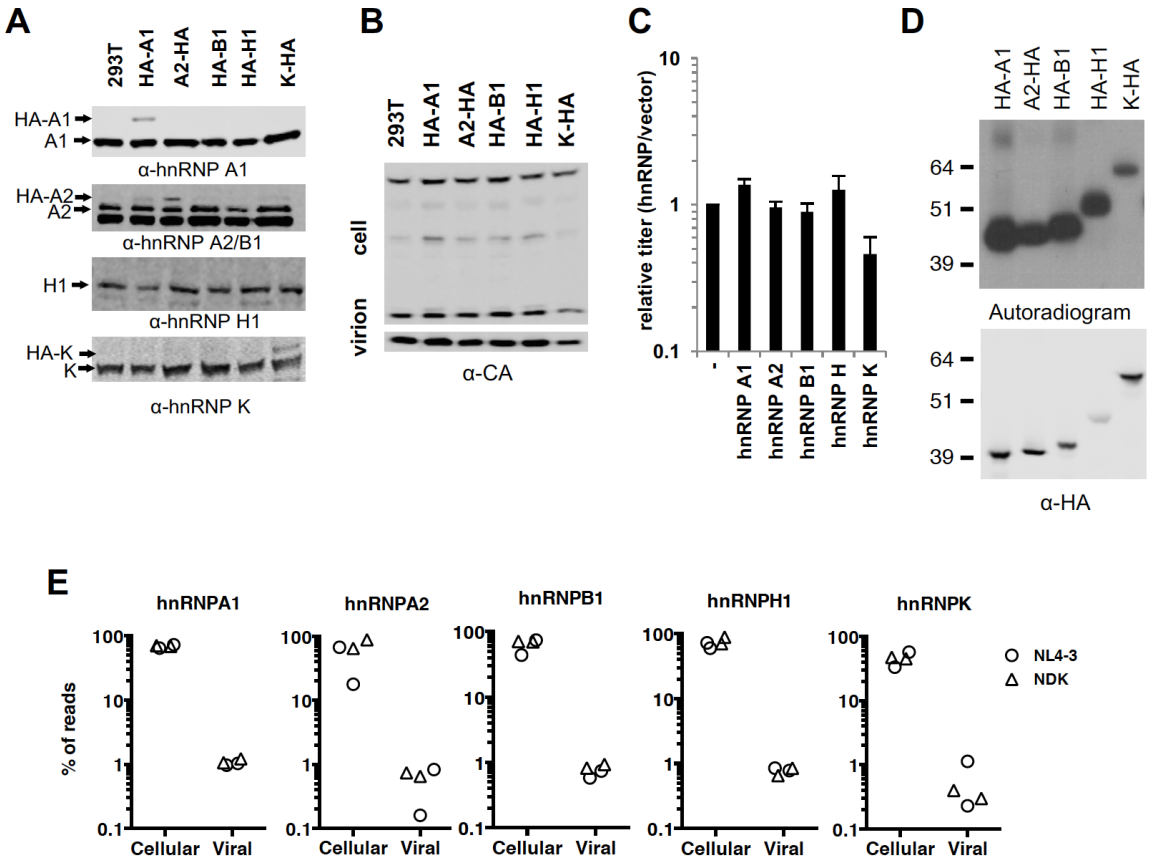

Figure 4.1. Generation of cell lines stably expressing HA-tagged hnRNPs and summary of CLIP-seq.

(A) Western blot analysis of HA-tagged and endogenous hnRNP proteins. (B, C) 293T cells expressing 3xHA-tagged hnRNPs were transfected with HIV-1_{NL4-3}. Western blot analysis of cell lysates and virions using an anti-CA antibody (B). Infectious virion yield was determined using MT4-GFP-reporter cells (C). Data is normalized to vector only control and represents the average of two independent experiments. (D) Immunoprecipitated and end-labeled hnRNP-RNA adducts from 4SU-treated cells is visualized by autoradiography (upper). Western blot analysis of the same membranes (lower). (E) Proportion of raw CLIP-seq reads that map to the human (cellular) and viral genomes obtained from experiments conducted with NL4-3 and NDK viruses. Data represents four independent CLIP-seq experiments (two NL4-3 and two NDK).

Binding sites of hnRNPs on viral RNAs

Analysis of the hnRNP binding sites on viral RNAs revealed unexpected and remarkable differences between the binding patterns of hnRNPs. Although HnRNP A1, A2 and B1 proteins were previously proposed to bind to distinct sequences on viral RNAs as detailed above, CLIP-seq analysis revealed multiple binding sites throughout the viral genome (Figure 4.2A-C). Nevertheless, regions overlapping with splice acceptor A7 (~nt 7860) and downstream from it (~nt 8300) were more frequently bound for both HIV-1_{NL4-3} and HIV-1_{NDK} (Figure 4.2A-C). This pattern of binding may reflect the spreading of hnRNP A1 in 5' and 3' direction through oligomerization following its initial binding to a high affinity binding site (24). Importantly the binding patterns of the highly related hnRNP A1, A2 and B1 proteins were largely similar for both HIV-1_{NL4-3} and HIV-1_{NDK}. We noted that hnRNP B1 binding took place at more distinct locations (Figure 4.2C), which can be explained by the presence of an extended N-terminal domain in hnRNP B1 that contains several basic amino acids that may alter its RNA binding affinity and specificity. In contrast to the promiscuous binding pattern of hnRNP A/B proteins, hnRNP H1 was bound to distinct well-defined sequences on viral RNAs (Figure 4.2D). While the majority of the binding sites were conserved between HIV-1_{NL4-3} and HIV-1_{NDK}, the most frequently bound site within HIV-1_{NL4-3} (~nt 5800) was not present in HIV-1_{NDK} (Figure 4.2D). Importantly, although a previous study indicated a role for hnRNP K in regulation of HIV-1 splicing (21), hnRNP K binding on the viral genome was at background levels (Figure 4.2E) as also apparent in the fairly low percentage of CLIP-seq reads that map to the viral genome as compared to other hnRNPs (Figure 4.1E).

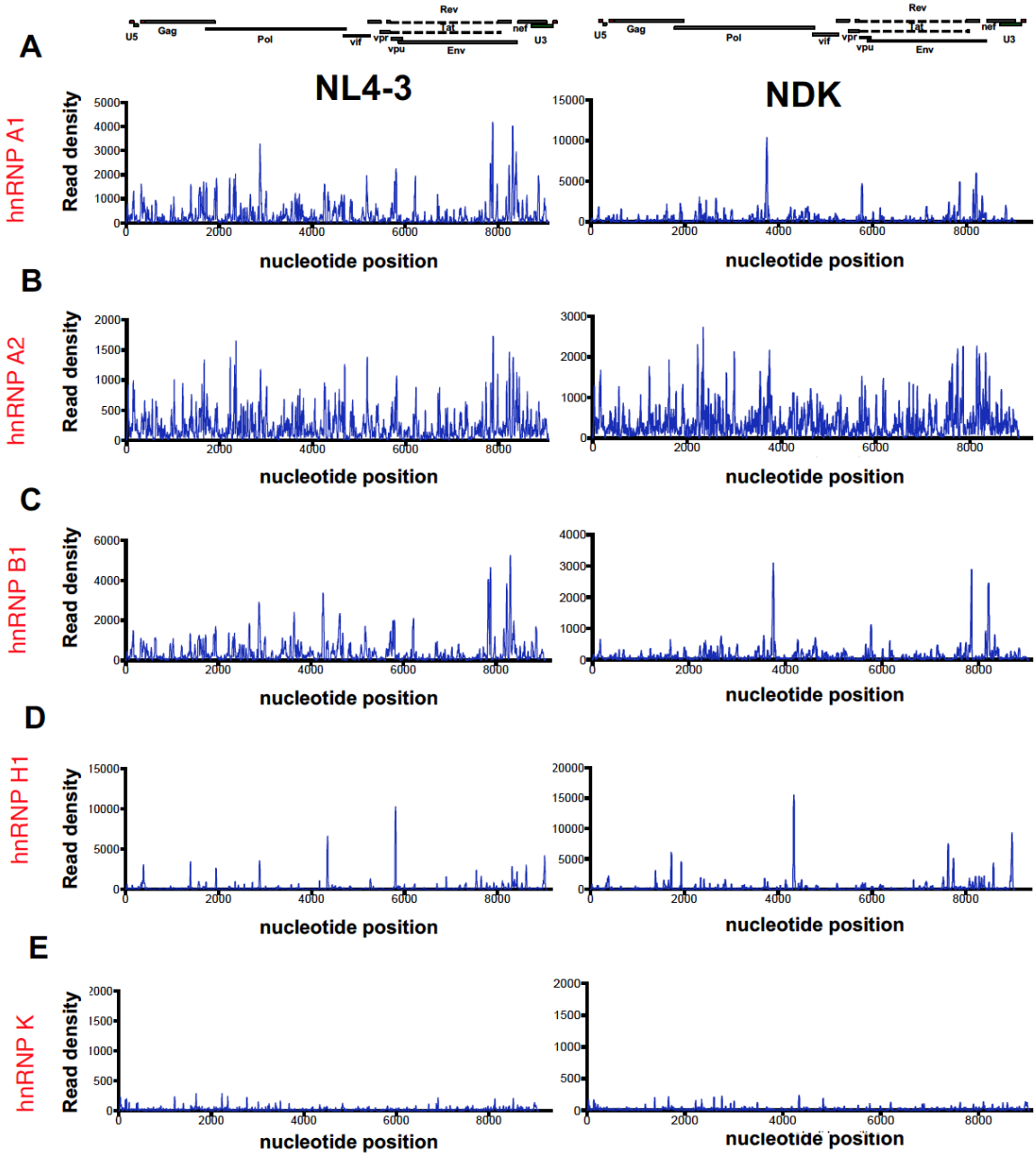

Figure 4.2. CLIP-seq of hnRNPs reveal their binding sites on viral RNAs.
293T cells stably expressing 3xHA-tagged hnRNPs were transfected with full-length HIV-1_{NL4-3} and HIV-1_{NDK} proviral plasmids prior to CLIP-seq. Frequency distribution of nucleotide occurrence (read density) of reads mapping to NL4-3 and NDK genomes are shown for hnRNP A1 (A), hnRNP A2 (B), hnRNP B1 (C), hnRNP H1 (D) and hnRNP K (E). A schematic diagram of HIV-1 genome features shown above is co-linear.

HnRNPs bind primarily to intronic sequences on cellular RNAs

We next determined the properties of the cellular RNAs bound by hnRNPs to gain further insight into whether the binding specificities of hnRNPs on viral RNAs are similar to those on cellular RNAs. Reassuringly, the majority of binding events for all hnRNPs took place within the transcripts of known genes (Figure 4.3A). Of these, a large fraction of CLIP-seq reads were derived from introns (Figure 4.3B), as also observed previously (25). To identify the sequence motifs within mRNAs and introns bound by hnRNPs, we applied the cERMIT algorithm (26). Representative sequence motifs and the cumulative frequencies of their occurrence are depicted in (Figure 4.3C). For hnRNP A1, A2 and B1 the predicted UNAGG motif resembles the well-characterized UAGG motif within the hnRNP A1 binding site (27, 28). HnRNP H1 binding took place at highly G-rich sequences similar to the previously known binding specificity for GGGA motif (29). Although hnRNP K has been proposed to bind short C-rich tracks as inferred from a SELEX study (30), CLIP-seq revealed the presence of ‘CGAU’ and ‘CCAUAGA’ motifs within a subset of cellular binding sites. Overall, the binding specificities of hnRNPs on viral RNAs mirrored those on cellular RNAs (see below).

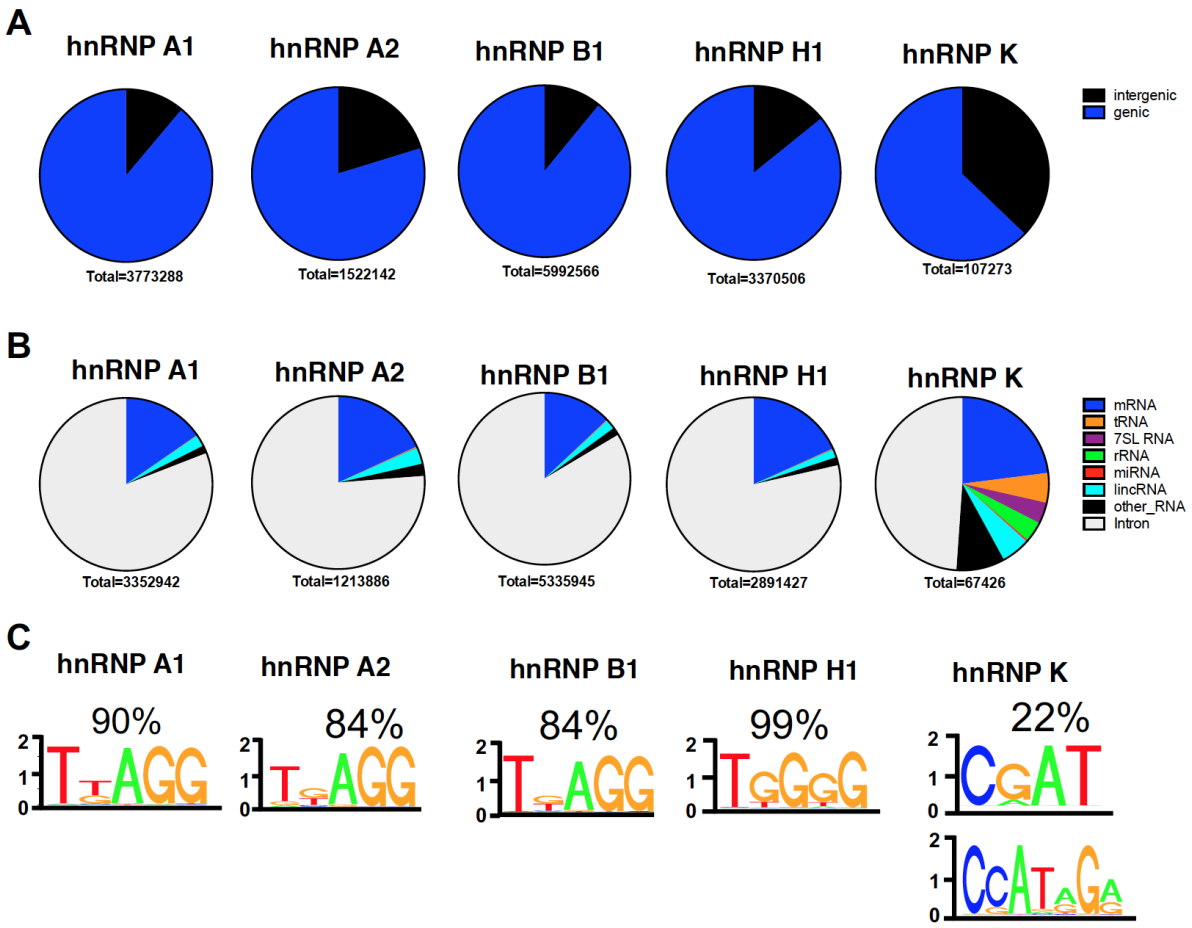

Figure 4.3. Classification of hnRNP-bound cellular RNAs and the binding motifs identified within them.

(A) hnRNP CLIP-seq derived sequencing reads were mapped to the human genome, clustered and classified based on whether they map to RNA-coding genes (genic) or intergenic regions. (B) Further classification of "genic" clusters is detailed below. Total number of sequencing reads in each group is represented below the pie charts. (C) Binding motifs within these clusters and the cumulative percentages of their occurrence within clusters.

Direct effect of hnRNP RNAi on HIV-1 splicing

We next tested how the RNAi-mediated knock down of hnRNPs may alter the global landscape of HIV-1 alternative splicing events as well as viral gene expression and infectivity. To this end, following the transfection of siRNA duplexes targeting hnRNP A1, A2/B1 and H1, cells were infected with HIV-1_{NL4-3}/VSV-G and HIV-1_{NDK}/VSV-G for two days. As expected, expression level of all hnRNPs, in particular hnRNP A2/B1 and hnRNP H1 was largely reduced upon siRNA transfection (Figure 4.4A). Next, we determined the direct effects of hnRNP RNAi on HIV-1 alternative splicing in cells by employing a high-throughput sequencing-based assay. While RNAi of all hnRNPs modestly increased the usage of splice acceptor A3 and decreased the usage of A4 for both 1.8 kb (Figure 4.4B) and 4 kb (Figure 4.4C) classes of transcripts, utilization of A1 was reduced upon RNAi of hnRNP A2/B1 and hnRNP H1 (Figure 4.4B, C). In addition, hnRNP H1 RNAi decreased the usage of splice acceptor A2 for both 1.8 kb and 4 kb classes of HIV-1 mRNAs (Fig, 4E, F). Analysis of cell lysates and cell culture supernatants revealed that RNAi of hnRNP H1 has consistently reduced NL4-3 (Figure 4.4D) and NDK Gag (Figure 4.4E) expression in cells, resulting in a proportional decrease in the relative number of infectious virus particles in cell culture supernatants (Figure 4.4D-F). In general, RNAi of hnRNP A1 and hnRNP A2/B1 modestly affected Gag expression and particle release/infectivity (Figure 4.4D-F), despite altering viral splicing. This in part can be explained by the opposing effects of hnRNP A1 and A2/B1 knock down on *tat* (increase in A3 usage) and *rev* (decrease in A4 usage) expression (Figure 4.4B). Overall, RNAi of each hnRNP had a distinct outcome on viral splicing but relatively modest effects on Gag expression with the exception of hnRNP H1.

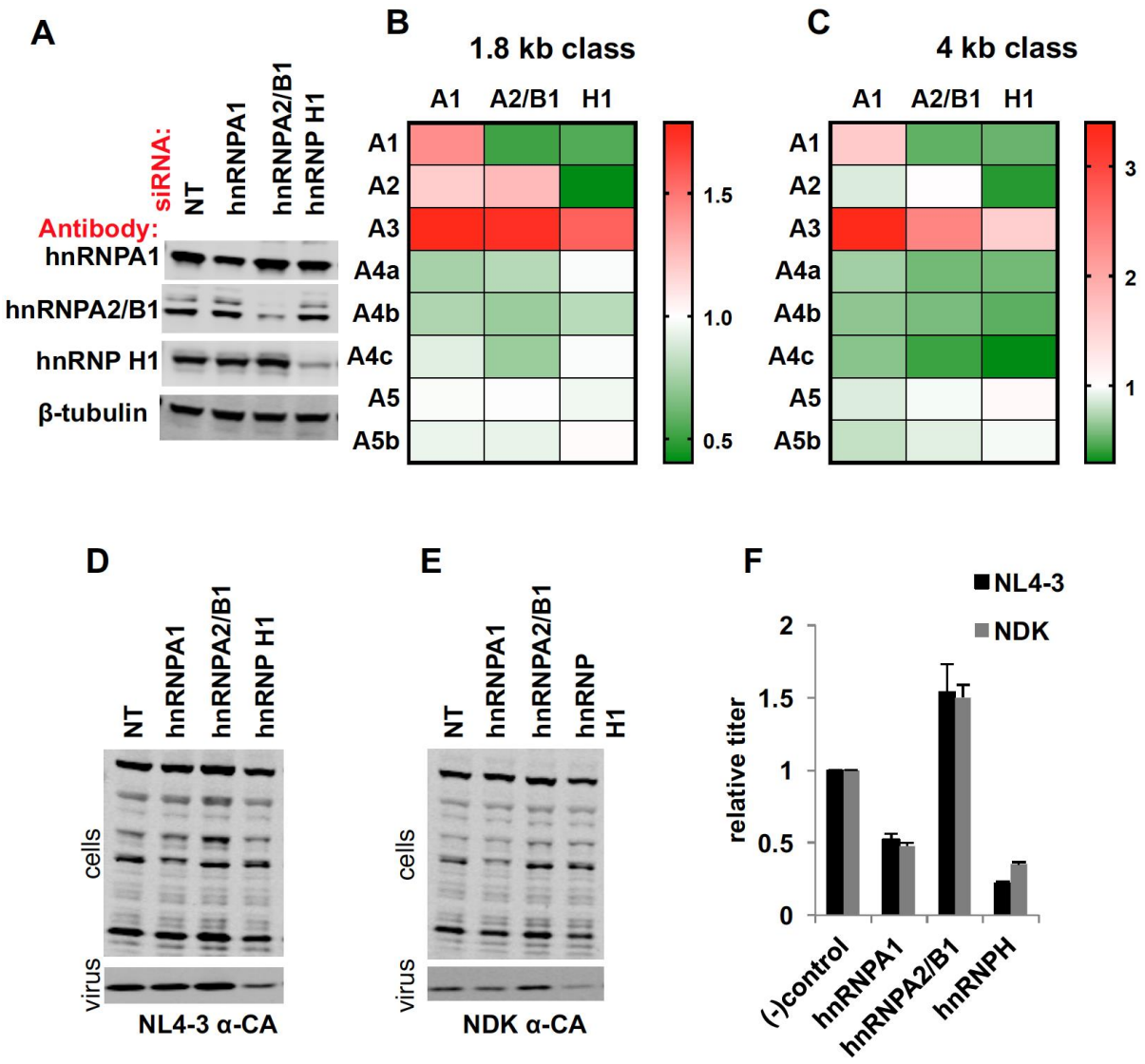

Figure 4.4. RNAi of hnRNP H1 decreases Gag expression and alters viral splicing in cells.
HEK293T cells were transfected with siRNAs targeting the indicated hnRNPs, infected with HIV_{NL4-3}/VSV-G or HIV_{NDK}/VSV-G one day post-transfection, and cells and viruses were harvested two days post-infection for further analysis. (A) Expression of targeted hnRNPs was determined by immunoblotting against the indicated hnRNPs. (B, C) Western blot analysis of cells lysates and virions using an anti-CA antibody is shown for HIV_{NL4-3} (B) or HIV_{NDK} (C)-infected cells. (D) Virus titers in cell culture supernatants was determined on MT4-GFP indicator cells. Data is normalized to non-targeting siRNA-transfected cells and shows the average of two independent experiments. Error bars represent the range. (E, F) Analysis of splicing events that occur at acceptors A1-A5b products for 1.8 kb (E) and 4 kb (F) classes of viral RNAs.

hnRNP H1 binds to purine-rich sequences on viral RNAs

Given that hnRNP H1 was bound to distinct sequences on viral RNAs (Figure 4.2D) and its RNAi had the most prominent effect on Gag expression (Figure 4.4D, E), we next aimed to determine how hnRNP H1 binding on viral RNAs directly affect HIV-1 alternative mRNA splicing. Analysis of the eight most frequently bound sites on the HIV-1 genome revealed that the majority of them consisted of one or more copies of G-rich tracks (four or more guanosines) surrounded by A-rich sequences (Figure 4.5A). Given this preference for G-rich sequences, we conducted hnRNP H1 CLIP-seq experiments in cells grown in the presence of 6SG to avoid any bias towards G-rich sequence elements with nearby uridines. While the overall binding pattern of hnRNP H1 was similar, we noticed that the relative frequencies of binding were different within certain sites. In particular, binding sites near ~3000 nt (site 4) and ~5800 nt (site 6) were relatively less frequently bound in 6SG-based CLIP-seq experiments (Figure 4.5B). Other sites, including site 5, which encompasses the central polypurine track remained as one of the most frequently bound sites (Figure 4.5B). We then tested whether the highly related hnRNP F protein binds to similar sites on viral RNAs by CLIP-seq. Although the majority of the binding sites, including sites 2, 5 and 6, overlapped, many other putative lower-frequency binding sites were observed (Figure 4.5C), suggesting that these highly similar proteins potentially have both overlapping and distinct functions in regulation of HIV-1 splicing. We then tested whether mutations within hnRNP H1 binding sites blocked hnRNP H1 binding on viral RNAs. To this end, mutations within the eight most predominant hnRNP H1 binding sites were introduced such that no more than two consecutive A or G nucleotides remained within the hnRNP H1 binding site. Reassuringly, these mutations abolished hnRNP H1 binding on viral RNAs and did not lead to the emergence of other high frequency binding sites (Figure 4.5D). Overall, these findings corroborate the aforementioned CLIP-seq studies with hnRNP H1 (Figure 4.2D) and support the presence of several high-specificity binding sites for hnRNP H1 on HIV-1 RNAs.

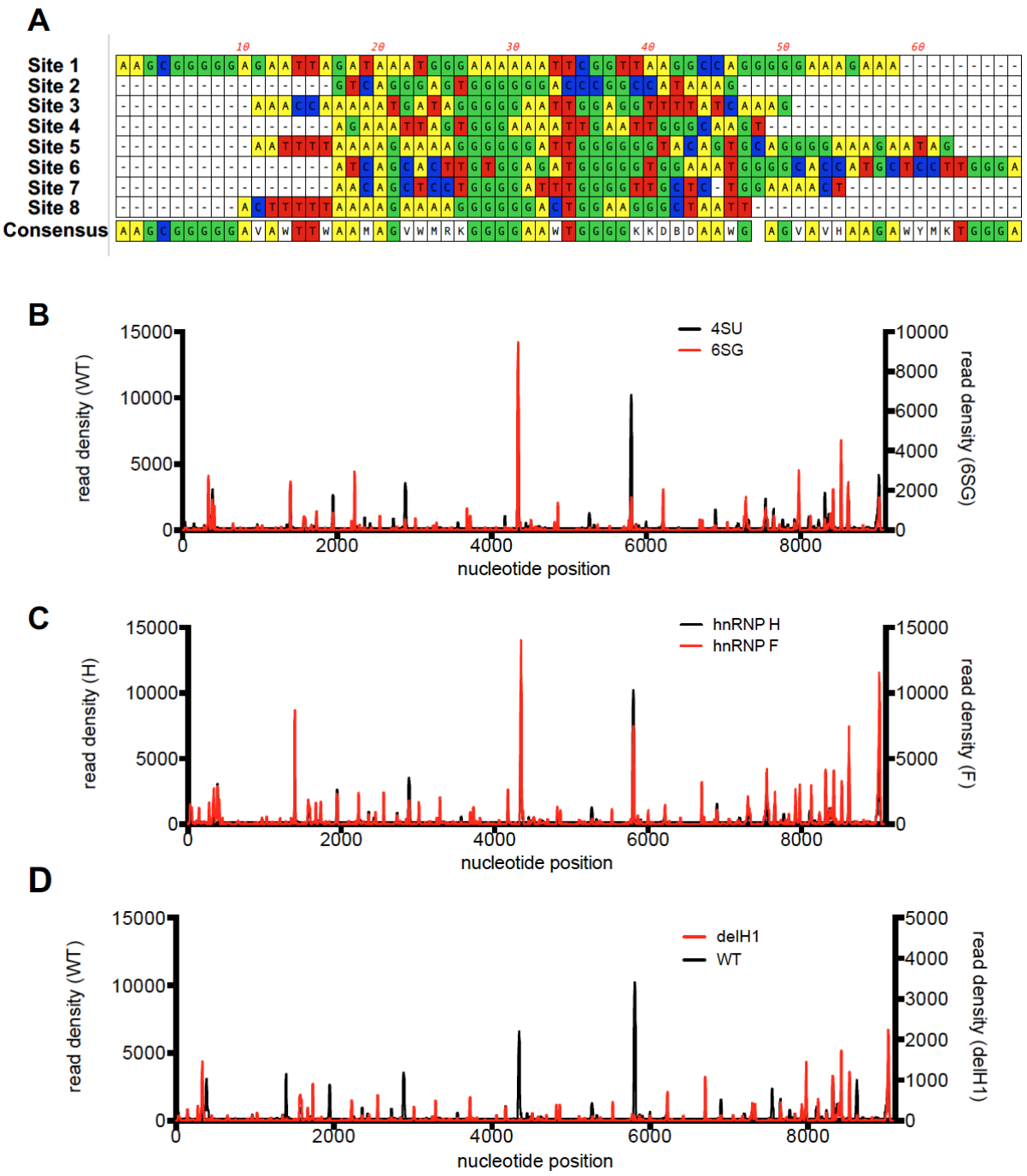

Figure 4.5. HnRNP H1 binds to purine-rich sequences on the viral genome.
(A) Sequence alignment of the eight most frequently hnRNP H1-bound sites as derived from CLIP-seq analysis shown in Figure 4.2. (B-D) Comparison of the frequency distribution of nucleotide occurrence (read density) of reads mapping to the NL4-3 genome derived from CLIP-seq experiments conducted on 4SU vs. 6-SG-treated cells (B), hnRNP H vs. hnRNP F-expressing cell lines (C), and in hnRNP H1-expressing cells transfected with WT vs. hnRNP H1-binding site mutant proviral clones (D).

Calorimetric and NMR titrations validate hnRNP H1 RNA binding preferences

To gain a quantitative description of how hnRNP H1 binds purine rich sequences, we carried out calorimetric and NMR titrations using its RNA binding domains. Human hnRNP H1 consists of three quasi-RNA Recognition Motifs (qRRMs) and two glycine-rich domains (Figure 4.6A). At its N-terminus, qRRM1 and qRRM2 are connected through a ~20 amino acid linker (qRRM12), whereas qRRM3 is embedded between the glycine-rich domains at its C-terminus. While members of the hnRNP H/F family are known to bind G-rich sequences, our CLIP-seq experiments reveal an obvious poly-A preference surrounding these sequences. To quantitatively determine the RNA sequence contribution to hnRNP H1 in recognition of HIV-1 RNAs, we performed calorimetric titrations of its qRRM12 domain with an RNA oligonucleotide that has the same composition as site 5 (Figure 4.6B). The titration showed that qRRM12 binds site 5 with moderate affinity (K_d = 0.62 μM) and with a stoichiometry >2, likely as result of the presence of three G-tracts that are 4-6 nts long within site 5. The poly-A region did not appear to contribute to binding affinity since titrations of qRRM12 with a 5'-AAAAA-3' oligonucleotide showed no binding (Figure 4.6B). Furthermore, titrations with 5'-AGGAC-3' revealed that binding requires more than 2 consecutive guanosines (Figure 4.6B).

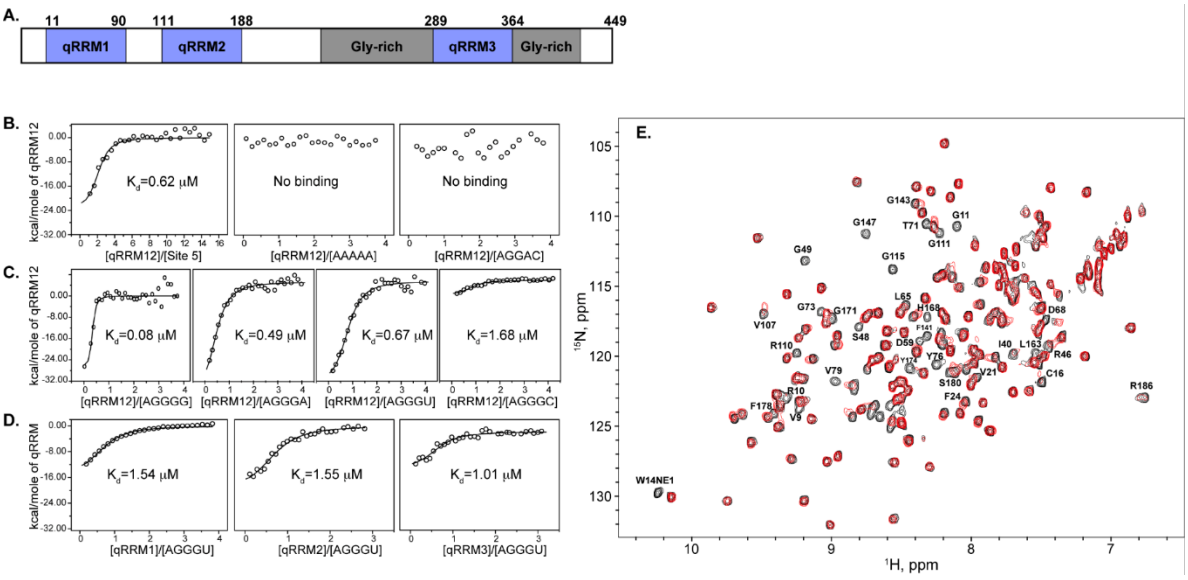

Figure 4.6. HnRNP H1 exhibits specificity for G-rich sequences in vitro.
(A) Schematic diagram of the domain architecture of hnRNP H1. ITC experiments conducted with (B) hnRNP H1-qRRM12 and RNA fragments consisting of Site 5 within the viral genome, 'AAAAA', and 'AGGAC', (C) hnRNP H1-qRRM12 and RNA fragments 'AGGGG', 'AGGGA', 'AGGGU' and 'AGGGC'. (D) ITC experiments were conducted using individual RNA binding domains of hnRNP H1, qRRM1-3 and 'AGGGU'. (E) HSQC titrations were performed with 15N-labled qRRM12 and unlabeled 'AGGGU'.

Having established that qRRM12 does not bind poly-A and minimally requires >2 consecutive guanosines, we next performed calorimetric titrations with model RNA oligonucleotides: 5'-AGGGG-3', 5'-AGGGA-3', 5-AGGGU-3' and 5'-AGGGC-3' (Fig 6C). qRRM12 bound all four oligonucleotides with moderate affinities (K_d = 0.08-1.7 uM) but it bound 5'-AGGGG-3' the tightest (Fig 6B). The binding stoichiometry for the qRRM12-(5'-AGGGG-3') complex is <0.5, which we determined by NMR (not shown) to be a result of a preexisting single-strand to G-quadruplex equilibrium of 5'-AGGGG-3'. The stoichiometric ratios for all other complexes are consistent with 1:1 binding, indicating that these oligonucleotides predominantly adopt a single-strand conformation under the solution conditions. Given that qRRM12 binds G-tracts with 1:1 stoichiometry, we sought to determine if the isolated qRRMs retain a similar mode of interaction. Figure 4.6D shows that all three qRRMs bound 5'-AGGGU-3' with nearly identical affinity and as a 1:1 complex. To further identify the binding interface of qRRM12 with G-tract RNA, we performed HSQC titrations, wherein uniformly ^{15}N-labled qRRM12 was titrated with unlabeled 5'-AGGGU-3' (Fig 6E). The differential line broadening of the NMR signals for qRRM12 in complex with 5'-AGGGU-3' clearly indicate binding specificity through both qRRM1 and qRRM2 domains. Thus, the combined calorimetric and NMR titrations reveal that each qRRM binds G-tract sequences equally well and that the tandem qRRM12 domain does not appear to bind G-tracts cooperatively. However, these results shed light on why Site 5, which contains three consecutive G-rich tracts constitutes the primary hnRNP H1 binding site on viral RNAs.

Effect of mutations within the hnRNP H1 binding sites

We then tested whether individual mutations within hnRNP H1 binding sites can recapitulate the observed effects of hnRNP H1 RNAi on HIV-1 alternative mRNA splicing (Figure 4.4B, C). To this end, mutations within hnRNP H1 binding site were introduced as above. Note that although the majority of these mutations were silent, conserved amino acid substitutions (i.e. lysine to arginine) were introduced within hnRNP H1 binding sites when necessary. Reassuringly, transfection of HEK293T cells with proviral plasmids yielded similar levels of infectious virus for the majority of the mutants (Figure 4.7A), with the exception of mutation within the eighth binding site that encompasses the 3' polypurine tract. To test the effects of hnRNP H1 mutations on viral splicing, HEK293T cells were transfected with NL4-3-derived proviral plasmids that bear the individual hnRNP H1 binding site mutations as well as a mutant plasmid in

which all mutations were combined. Total RNA was analyzed as in Figure 4.4 by deep-sequencing. While the majority of mutations had relatively modest effects on HIV-1 splicing, mutation of the 5th binding site reduced the usage of splice acceptor A1 at levels similar to hnRNP H1 knock down (Figure 4.7C). Note that this site has consistently constituted the most predominant hnRNP H1 binding site in all CLIP-seq experiments and is localized near splice acceptor A1 (Figure 4.7D). Importantly, amongst all of the hnRNP H1 binding sites, site 5 exhibited the strongest splicing silencer activity in a reporter assay system (31) (Figure 4.7E). Combination of all mutations similarly led to a further reduction in A1 usage, suggesting the contribution of other hnRNP H1 binding sites to regulation of A1 usage. Next, we determined whether these changes in viral splicing would affect viral gene expression. To this end, HEK293T cells were infected with VSV-G pseudotyped WT and mutant viruses at equivalent MOIs, and Gag and Vif expression in cell lysates was analyzed two days post infection. While the mutations within sites 1-3 and 5 had no major effect on Gag expression, mutations 4 and 6-8 reduced Gag expression significantly (Figure 4.7F). Note that similar experiments conducted in the presence of reverse transcriptase inhibitor, Nevirapine, did not reveal any Gag signal in western blots, suggesting that the Gag detected in this setting is synthesized *de novo* and does not represent the input viruses. Most notably, while mutation 5 did not affect Gag expression, it led to a significant reduction in Vif expression (Figure 4.7F), in line with the observed reduction in A1 usage that controls Vif expression (Figure 4.7C). Overall, these experiments revealed a novel cis-acting element that regulates splice acceptor A1 usage and subsequent Vif expression.

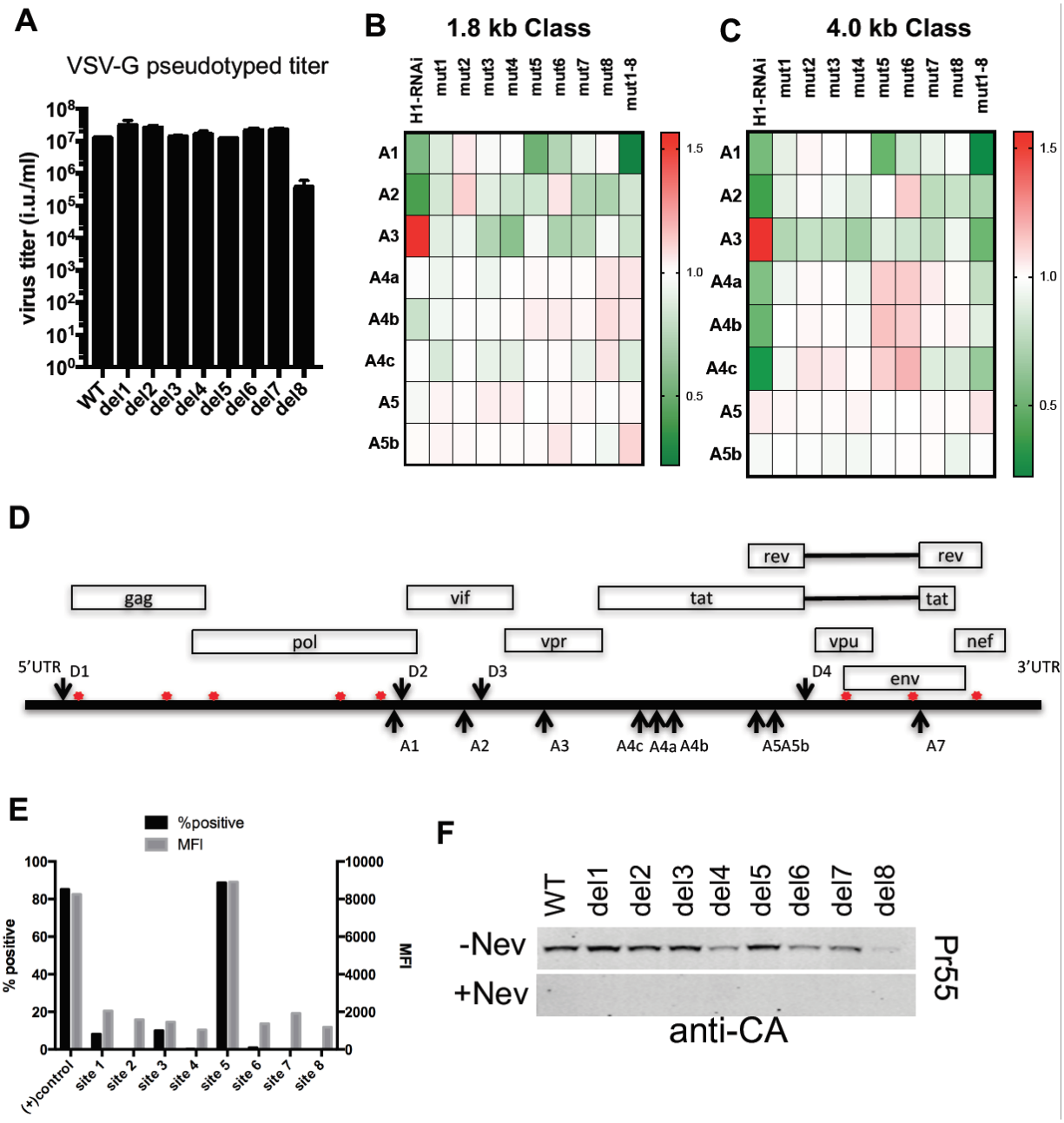

Figure 4.7. Mutation of hnRNP H1 binding sites alters HIV-1 replication and alternative splicing. (A) Viral titers for individual hnRNP H1 mutant clones transfected into 293T cells (B, C) Splice acceptor usage for the 1.8 kb and 4 kb class of HIV-1 RNAs is shown for virus growing in cells where hnRNP H1 was knocked down with siRNA (first column), and in which one or all of the hnRNP H1 binding sites were mutated. (D) Schematic diagram of the HIV-1 genome features and the locations of the mutated hnRNP H1-binding sites. (E) Splicing activity of the hnRNP binding sites in a reporter assay. (F) Gag expression of the hnRNP H1 binding site mutants, with and without the reverse transcriptase inhibitor Nevirapine.

Discussion

Alternative splicing of HIV-1 mRNAs has traditionally been studied in artificial settings whereby a fragment of the viral genome with presumed regulatory activity was taken out of the context of the viral genome. These elements and the trans-acting factors (i.e. hnRNPs or SR proteins) that bind to them were typically analyzed by employing *in vitro* or transfection-based splicing reporter assays. These approaches were typically combined with RT-PCR/gel sizing assays (6) to capture how mutation of a regulatory element alters viral splicing. However, as this method relies on size-separation of viral mRNA isoforms following PCR, it cannot accurately identify differences in co-migrating variants, has limited sensitivity to detect minor variants, and is prone to skewing in the relative quantities of splice isoforms due to PCR artifacts. As such, these approaches are greatly limited in their ability to reveal the physiologically relevant splicing regulatory events.

In this study, we employed next-generation sequencing-based approaches to identify the targets of several hnRNP proteins on HIV-1 RNAs in cells and how these binding events directly affect viral alternative mRNA splicing. In contrast to previous studies, which suggested that members of the hnRNP A/B family bind to distinct sequences on viral RNAs, CLIP-seq experiments revealed that these proteins bind rather promiscuously throughout viral RNAs and require the presence of ‘AGG’ motifs within binding sites. In contrast, hnRNP H1 binding took place at distinct G-rich sequence elements, which often consisted of two or more copies of ‘GGGG’ nucleotides. As such, it is probabilistically expected that the viral genome contains significantly more binding sites for hnRNP A/B than hnRNP H1. Reassuringly, the most predominant hnRNP H1 binding site near splice acceptor A1 exhibited splice enhancer activity, i.e. mutation of this site led to a significant decrease in the usage of A1 and Vif expression. While hnRNPs are generally thought to repress splicing at nearby elements, it is important to note that many act in a context dependent manner and can exhibit splice enhancer or silencer activities (32).

hnRNPs can repress the assembly of spliceosomes through multimerization along exons (33). Therefore, the promiscuous binding pattern of hnRNP A1 and A2/B1 proteins on viral RNAs can be explained by their initial interaction with high affinity binding sites and their subsequent oligomerization throughout viral RNAs. Another possible explanation is that not all binding events identified by CLIP-seq are relevant to splicing regulation, but reflect other metabolic roles of hnRNPs throughout the cell. Note

that the CLIP-seq experiments in this study were conducted on total cell lysates suggesting this possibility. For example, hnRNP proteins have known regulatory roles in retention of unspliced viral genomic RNA in the nucleus (34, 35), nuclear export of viral/cellular RNAs (36), mRNA translation, storage, degradation and stress responses (36-40). While at steady state the majority of hnRNP/SR proteins are localized in the nucleus, many hnRNP/SR proteins shuttle between the nucleus and the cytosol (41, 42).

RNAi of hnRNPs have been widely utilized in the field to determine their roles in HIV-1 splicing regulation. However, we noticed that RNAi of hnRNP A1 and A2/B1 did not affect Gag expression or particle infectivity. This may indicate the functional redundancy between these proteins, as also supported by the CLIP-seq data. However, it is noteworthy that the RNAi of hnRNP A1 and A2/B1 altered the utilization of several splice acceptors, including A3, which controls *tat* expression, and A4, which controls *rev* expression. As such, the opposing effects of hnRNP A1 and A2/B1 RNAi on *tat* and *rev* expression can also, in principle, explain the lack of effect on Gag expression. In contrast, while RNAi of hnRNP H1 decreased A4 usage (and hence likely decreased *rev* expression), it did not cause a corresponding increase in A3 usage (i.e. *tat* expression), which may explain the Gag expression defect.

Several previous studies have identified hnRNP H1 binding sites on viral RNAs. Notably, none of these sites are bound at high levels by hnRNP H1 in the CLIP-seq setting. For example, the proposed hnRNP H1 binding sites within *tat* exon 2 (14) and tev-specific exon 6D in certain HIV-1 isolates (15) consists of only one ‘GGG’ element and no surrounding A-rich sequences. Notably, detailed *in vitro* binding experiments previously revealed a key preference for the presence of multiple copies of three or more G-rich sequences interspersed with other nucleotides for high affinity binding for all members of the hnRNP H family (17). We have observed that the affinity of hnRNP H1-RRM12 towards the CLIP-seq-derived binding site 5, which contains three high affinity binding sites for hnRNP H1, was similar to isolated G-rich sequence elements *in vitro*. This perhaps suggests that the ability of hnRNP H to bind to target sequences is not only driven by the affinity of individual RRMs, but also by avidity, which would explain why site 5 within the viral genome constitutes the primary binding site for hnRNP H1. Future structural studies will be crucial in determining how G-rich repeats affect RNA structure and how hnRNPs gain specificity for target sequences given the apparent low-modest affinity of binding *in vitro*.

Although Site 5 exhibited strong splicing silencer activity in a reporter-based assay, our results indicate that it acts to enhance splicing to acceptor A1 within the context of the viral genome. While this reiterates the importance of studying a splicing regulatory element within its natural context, it is also noteworthy that no other cryptic sites were activated either near A1 or in other locations across the genome by mutations to hnRNP H1 binding sites.

Although we have identified several distinct hnRNP H1 binding sites by CLIP-seq, mutations introduced within the majority of hnRNP H1 binding sites only modestly affected splicing. One explanation for this phenomenon is the other known roles of hnRNPs in RNA metabolism as detailed above, which also is in line with the result that mutation of some hnRNP H1 binding sites led to decreased Gag expression in cells. A further explanation is that these binding sites work in a combinatorial fashion with other hnRNP H1 binding sites. In fact, we noted that mutation of all hnRNP H1 binding sites led to a more significant decrease in splice acceptor usage (Figure 4.7D, E). This would also be in line with the model that hnRNPs can loop out entire exons (43), which would require a minimum of two binding sites surrounding exons.

Overall, our studies reveal precisely where on viral RNAs the hnRNPs bind to but also how these binding events directly regulate HIV-1 splicing. Finally, these studies also provide the key framework to further investigate the poorly understood process of HIV-1 splicing.

Materials and Methods

Cell culture, transfection and infections

HeLa-derived TZM-bl cells (NIH AIDS Reagents) and HEK 293T cells (ATCC CRL-11268) were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. Full-length HIV-1 proviral clones, NL4-3 (NIH AIDS Reagents, (44)), NDK (45), and derivatives thereof were transfected to HEK293T cells either alone or together with a VSV-G expression vector at ratio of 4:1 using polyethylenimine (PolySciences). To generate HEK293T cells stably expressing hnRNPs, cells were transduced with VSV-G pseudotyped viruses that were produced by transfection of 293T cells with plasmids expressing MLV Gag-Pol, LHCX-derived hnRNP expression plasmids and VSV-G at a ratio of 5:5:1, respectively, using polyethylenimine. Following selection in cell culture media supplemented with 50 μg/ml hygromycin, single cell clones were isolated and propagated. For CLIP-seq experiments,

HEK293T cells stably expressing 3xHA-tagged hnRNPs were grown in 15-cm dishes and transfected with 30 μg of the indicated proviral plasmids using polyethylenimine (PolySciences). In other experiments, 293T cells grown in 24-well cell culture dishes were transfected with siRNA duplexes against hnRNPs using Lipofectamine RNAimax reagent (Life Technologies) according the manufacturer's instructions. One day post siRNA transfection, cells were infected with VSV-G-pseudotyped HIV_{NL4-3} or $HIV\text{-}1_{NDK,}$ and viral gene expression was analyzed two days later.

Antibodies

Antibodies used in CLIP-seq and western blot assays were as follows: mouse monoclonal anti-HA (HA.11 Covance), mouse monoclonal anti-hnRNP A1 (SC-32301, Santa Cruz), mouse monoclonal anti-hnRNP A2/B1 (SC-374053, Santa Cruz), goat polyclonal anti-hnRNP H (SC-10042, Santa Cruz), mouse monoclonal hnRNP K (SC-28380, Santa Cruz), and mouse monoclonal anti-HIV-1 p24CA (183-H12-5C, NIH AIDS Reagent Program).

Plasmids

Full-length proviral plasmids pNL4-3 (NIH AIDS Reagents) and pNDK were described before (44, 46). N- and C-terminally 3xHA-tagged hnRNPs were cloned into the pCR3.1 and pLHCX (Clontech) vector backbones by conventional molecular biology tools. HnRNPH1 binding site mutations were generated on the NL4-3 proviral backbone by overlap PCR using the following primer pairs containing the mutations:

356-414 (F): 5' ATT AAG CGG CGG AGA ATT AGA TAA GTT CGA CAA GAT TCG GTT AAG GCC AGG CGG CAA GA 3'

356-414 (R): 5' TCT TGC CGC CTG GCC TTA ACC GAA TCT TGT CGA ACT TAT CTA ATT CTC CGC CGC TTA AT 3'

1373-1424 (F): 5' GAT GAC AGC ATG TCA AGG CGT CGG CGG CCC CGG CCA TAA GGC AAG AGT TTT G 3'

1373-1424 (R): 5' CAA AAC TCT TGC CTT ATG GCC GGG GCC GCC GAC GCC TTG ACA TGC TGT CAT C 3'

1918-1961 (F): 5' GAA GAT GGA AGC CCA AGA TGA TAG GCG GCA TTG GAG GTT TTA TC 3'

1918-1961 (R): 5' GAT AAA ACC TCC AAT GCC GCC TAT CAT CTT GGG CTT CCA TCT TC 3'

2961-2909 (F): 5’ GAC ATA CAG AAG TTA GTC GGC AAG CTC AAT TTC GCA AGT CAG ATT TAT G 3’

2961-2909 (R): 5’ CAT AAA TCT GAC TTG CGA AAT TGA GCT TGC CGA CTA ACT TCT GTA TGT C 3’

4322-4381 (F): 5’ CAC AAT TTT AAG AGG AAG GGC GGC ATT GGC GGC TAC AGT GCA GGC GAG AGA ATA GTA GAC 3’

4322-4381 (R): 5’ GTC TAC TAT TCT CTC GCC TGC ACT GTA GCC GCC AAT GCC GCC CTT CCT CTT AAA ATT GTG 3’

4322-4381 (F): 5’ CAC AAT TTT CGA CGC CGC GGC GGC ATT GGC GGC TAC AGT GCA GGC GAG AGA ATA GTA GAC 3’

4322-4381 (R): 5’ GTC TAC TAT TCT CTC GCC TGC ACT GTA GCC GCC AAT GCC GCC GCG GCG TCG AAA ATT GTG 3’

5785-5836 (F): 5’ TAT CAG CAC TTG TGG AGA TTC GGC TGG CGC TTC GGC ACC ATG CTC CTT GGG A 3’

5785-5836 (R): 5’ TCC CAA GGA GCA TGG TGC CGA AGC GCC AGC CGA ATC TCC ACA AGT GCT GAT A 3’

7524-7575 (F): 5’ GGA TCA ACA GCT CCT CGG CAT TTT CGG TTG CTC TGG CAA GCT CAT TTG CAC C 3’

7524-7575 (F): 5’ GGT GCA AAT GAG CTT GCC AGA GCA ACC GAA AAT GCC GAG GAG CTG TTG ATC C 3’

8587-8645 (F): 5’ AGA TCT TAG CCA CTT TCT CCG CGA GCG CGG CGG ACT GGA AGG CCT AAT TCA CTC CCA AA 3’

8587-8645 (R): 5’ TTT GGG AGT GAA TTA GGC CTT CCA GTC CGC CGC GCT CGC GGA GAA AGT GGC TAA GAT CT 3’

A GFP-based splicing silencer reporter plasmid, pZW4, was published before (31). The following oligonucleotides encompassing the hnRNP H1 binding sites and containing 5’ XhoI- 3’ Apa I restriction sites were annealed *in vitro* cloned into the pZW4 vector. The resulting constructs were transfected into HEK293T cells and GFP expression was analyzed two-days post-transfection by FACS.

362-402 (F): 5’ TCGAGCGGGGGAGAATTAGATAAATGGGAAAAAATTCGGTTAAGGC GGGCC 3’

362-402 (R): 5’ CGCCTTAACCGAATTTTTTCCCATTTATCTAATTCTCCCCCGC 3’

1386-1406 (F): 5’ TCGAG CAGGGAGTGGGGGGACCCGGCGGGCC 3’

1386-1406 (R): 5’ CGCCGGGTCCCCCCACTCCCTGC 3’

1925- 1958 (F): 5’ TCGAGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTAGGGCC 3’

1925-1958 (R): 5’ CTAAAACCTCCAATTCCCCCTATCATTTTTGGTTTC 3’

2867-2899 (F): 5’ TCGAGCAGAAATTAGTGGGAAAATTGAATTGGGCAAGTGGGCC 3’

2867-2899 (R): 5’ CACTTGCCCAATTCAATTTTCCCACTAATTTCTGC 3’

4325-4371 (F): 5’

TCGAGAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGGGGCC 3’

4325-4371 (R): 5’ CCTTTCCCCTGCACTGTACCCCCCAATCCCCCCTTTTCTTTTAAAATTC

3’

5789-5822 (F): 5’ TCGAGAGCACTTGTGGAGATGGGGGTGGAAATGGGGCACGGGCC 3’

5789-5822 (R): 5’ CGTGCCCCATTTCCACCCCCATCTCCACAAGTGCTC 3’

7536-7564 (F): 5’ TCGAGCCTGGGGATTTGGGGTTGCTCTGGAAAACGGGCC 3’

7536-7564 (R): 5’ CGTTTTCCAGAGCAACCCCAAATCCCCAGGC 3’

8604-8631 (F): 5’ TCGAGTAAAAGAAAAGGGGGGACTGGAAGGGCTGGGCC 3’

8604-8631 (R): 5’ CAGCCCTTCCAGTCCCCCCTTTTCTTTTAC 3’

CLIP-seq experiments

CLIP-seq experiments were conducted as explained before (23). Briefly, 293T cells stably expressing 3xHA-tagged hnRNPs and transfected with proviral plasmids were grown in the presence of 4-thiouridine. Cells were washed in 1x phosphate buffered saline and UV-crosslinked at 365nm wavelength. Following lysis, hnRNP-RNA complexes were immunoprecipitated using monoclonal antibodies against the HA-tag. Bound RNA was end-labeled with γ-32P-ATP and T4 polynucleotide kinase. The isolated protein-RNA complexes were separated by SDS-PAGE, transferred to nitrocellulose membranes and exposed to autoradiography films. Protein-RNA adducts were excised from membranes and RNA was purified by proteinase K digestion. RNA was sequentially ligated to 3’ and 5’ adapters, reverse-transcribed, PCR-amplified and the resulting DNA library was subjected to sequencing by

Illumina platforms. The remainder of the data analysis was followed as detailed before (23), using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit), Bowtie (47), PARalyzer (48), with the exception of applying a more stringent T-to-C substitution-based cutoff for defining IN-binding sites (PARpipe). Briefly, sequencing reads were first trimmed off of the 3' adapter, reads shorter than 15 nucleotides were filtered out and the library was demultiplexed based on the 5' adapter sequences using the FASTX toolkit. Sequences were then collapsed into unique reads followed by trimming off the 5' adapter using the FASTX toolkit. The resulting reads were mapped to the human genome (hg19) or viral genome (pNL4-3) containing a single repeat (R) region using the Bowtie algorithm by allowing two mismatches (-v 2 -m 10 --best strata). Reads that map to the human genome were further processed and clustered by PARalyzer and PARpipe pipelines using default settings

Nucleic acid isolation, reverse transcription, Q-PCR

Total cellular RNA was isolated by Trizol extraction (Life Technologies) and the resulting RNA was reverse-transcribed using the Improm-II Reverse Transcription System (Promega) following the manufacturers' instructions. In some experiments, cytosolic and nuclear extracts were isolated using the NE-PER kit (Thermo-Fisher), and RNA and DNA were isolated subsequently by phenol:chloroform extraction. Q-PCR analysis of unspliced and spliced RNAs was conducted using the previously published primer pairs (23).

Splicing assays

A next-generation-sequencing splicing assay was conducted as detailed in Chapter 2. Briefly, total RNA from HEK293T cells transfected with proviral plasmids or infected with the corresponding VSV-G pseudotyped viruses was isolated by Trizol. The resulting RNA was subjected to the previously described library generation protocol (Chapter 2) and sequenced using the 300 base paired-end read Illumina MiSeq platform with samples loaded at 8 pM plus 15 % PhiX double-stranded DNA fragments to maximize cluster separation. Sample reads were demultiplexed using the Illumina bcl2fastq v.1.8.4 pipeline.

Splicing data analysis was done using a customized Ruby program which combines splicing data from the forward and reverse reads to identify and quantify specific transcript types. As previously described, all reads are tagged with a unique Primer ID sequence identifier attached in the cDNA step.

All sequencing reads with the same Primer ID are combined to prevent skewing from PCR resampling. Alternative acceptor splice sites are also identified and quantified. Programs used are available at https://github.com/SwanstromLab/SPLICING.

Protein expression and purification

We cloned HIS_6 tag qRRM1, qRRM2, qRRM3 and qRRM12 domains into pMCSG7 vector. The recombinant proteins were overexpressed in *E. coli* strain BL21 (DE3). The BL21 (DE3) cells were grown to an optical density of 0.8 at 37°C with 220 rpm and through induction at 20°C with 1mM IPTG for 16 hours. Cells were harvested in lysis buffer containing 20 mM Na_2HPO_4, 20 mM imidazole, 500 mM NaCl, 4mM TCEP at pH 8.0 by sonication at 4°C and collected by centrifugation. The cleared lysate was then applied to Ni-NTA agarose and washed with lysis buffer and lysis buffer containing 40mM imidazole, and eluted with buffer containing 300mM imidazole. The elution fractions were concentrated and buffer exchanged into lysis buffer. The HIS_6 tag of the proteins were removed by incubating with TEV Protease at 25°C for 16 hours. The proteins were purified by size exclusion chromatography with NMR buffer containing 20mM sodium phosphate, 20 mM NaCl, 4mM TCEP at pH 6.2. The purity of the proteins was analyzed and estimated > 95% by SDS page analysis.

Isothermal titration calorimetry

The binding affinities of qRRM1, qRRM2, qRRM3 and qRRM12 with the G-tract RNA oligonucleotides were characterized by measuring heat changes on titrating protein domains into each G-tract RNA oligonucleotide solution using a Microcal VP-ITC calorimeter. Protein and RNA solutions were buffer exchanged to 20mM sodium phosphate, 20 mM NaCl, 4mM TCEP at pH 6.2, centrifuged and degassed under vacuum before use. All titrations were performed at 25°C and the data were analyzed using Origin software (Microcal).

NMR experiments

The NMR experiments were performed on a Bruker 800 MHz spectrometer at 32°C equipped with a cryprobe. The ^{15}N-labeled and $^{15}N/^{13}C$-labeled samples for NMR experiments were prepared in a buffer containing 20mM sodium phosphate, 20 mM NaCl, 4mM TCEP and 10 % D2O at pH 6.2. Backbone assignments of qRRM12 were obtained from HNCA, HNCACB, HNCO, and HN(CA)CO. For NMR titrations, the uniformed ^{15}N-labeled qHRRM12 samples were prepared at a concentration of 90 □M in

20mM sodium phosphate, 20 mM NaCl, 4mM TCEP and 10 % D2O at pH 6.2. The unlabeled 5’-AGGGC-3’ oligonucleotide was added to uniformly ^{15}N-labeled qRRM12 at molar ratio of 1:0:33, 1:0:66, 1:1, 1:2. All Spectra were processed with NMRPipe/DRAW and analyzed using Sparky.

REFERENCES

1. Karn J, Stoltzfus CM. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med 2:a006916.

2. Mandal D, Exline CM, Feng Z, Stoltzfus CM. 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. J Virol 83:6067-6078.

3. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res 40:10345-10355.

4. Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. 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.

5. Holmes M, Zhang F, Bieniasz PD. 2015. Single-Cell and Single-Cycle Analysis of HIV-1 Replication. PLoS Pathog 11:e1004961.

6. Purcell DF, Martin MA. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67:6365-6378.

7. Caputi M, Mayeda A, Krainer AR, Zahler AM. 1999. hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. Embo j 18:4060-4067.

8. Del Gatto-Konczak F, Olive M, Gesnel MC, Breathnach R. 1999. hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. Mol Cell Biol 19:251-260.

9. Si ZH, Rauch D, Stoltzfus CM. 1998. The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. Mol Cell Biol 18:5404-5413.

10. Damgaard CK, Tange TO, Kjems J. 2002. hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. Rna 8:1401-1415.

11. Bilodeau PS, Domsic JK, Mayeda A, Krainer AR, Stoltzfus CM. 2001. RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element. J Virol 75:8487-8497.

12. Tange TO, Damgaard CK, Guth S, Valcarcel J, Kjems J. 2001. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. Embo j 20:5748-5758.

13. Asai K, Platt C, Cochrane A. 2003. Control of HIV-1 env RNA splicing and transport: investigating the role of hnRNP A1 in exon splicing silencer (ESS3a) function. Virology 314:229-242.

14. Jacquenet S, Mereau A, Bilodeau PS, Damier L, Stoltzfus CM, Branlant C. 2001. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. J Biol Chem 276:40464-40475.

15. Caputi M, Zahler AM. 2002. SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. Embo j 21:845-855.

16. Jablonski JA, Buratti E, Stuani C, Caputi M. 2008. The secondary structure of the human immunodeficiency virus type 1 transcript modulates viral splicing and infectivity. J Virol 82:8038-8050.

17. Schaub MC, Lopez SR, Caputi M. 2007. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. J Biol Chem 282:13617-13626.

18. Widera M, Hillebrand F, Erkelenz S, Vasudevan AA, Munk C, Schaal H. 2014. A functional conserved intronic G run in HIV-1 intron 3 is critical to counteract APOBEC3G-mediated host restriction. Retrovirology 11:72.

19. Lund N, Milev MP, Wong R, Sanmuganantham T, Woolaway K, Chabot B, Abou Elela S, Mouland AJ, Cochrane A. 2012. Differential effects of hnRNP D/AUF1 isoforms on HIV-1 gene expression. Nucleic Acids Res 40:3663-3675.

20. Woolaway K, Asai K, Emili A, Cochrane A. 2007. hnRNP E1 and E2 have distinct roles in modulating HIV-1 gene expression. Retrovirology 4:28.

21. Marchand V, Santerre M, Aigueperse C, Fouillen L, Saliou JM, Van Dorsselaer A, Sanglier-Cianferani S, Branlant C, Motorin Y. 2011. Identification of protein partners of the human immunodeficiency virus 1 tat/rev exon 3 leads to the discovery of a new HIV-1 splicing regulator, protein hnRNP K. RNA Biol 8:325-342.

22. Han SP, Tang YH, Smith R. 2010. Functional diversity of the hnRNPs: past, present and perspectives. Biochem J 430:379-392.

23. Kutluay SB, Zang T, Blanco-Melo D, Powell C, Jannain D, Errando M, Bieniasz PD. 2014. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. Cell 159:1096-1109.

24. Okunola HL, Krainer AR. 2009. Cooperative-binding and splicing-repressive properties of hnRNP A1. Mol Cell Biol 29:5620-5631.

25. Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S, Ares M, Jr., Yeo GW. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep 1:167-178.

26. Georgiev S, Boyle AP, Jayasurya K, Ding X, Mukherjee S, Ohler U. 2010. Evidence-ranked motif identification. Genome Biol 11:R19.

27. Hutchison S, LeBel C, Blanchette M, Chabot B. 2002. Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. J Biol Chem 277:29745-29752.

28. Bruun GH, Doktor TK, Borch-Jensen J, Masuda A, Krainer AR, Ohno K, Andresen BS. 2016. Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. BMC Biol 14:54.

29. Caputi M, Zahler AM. 2001. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. J Biol Chem 276:43850-43859.

30. Thisted T, Lyakhov DL, Liebhaber SA. 2001. Optimized RNA targets of two closely related triple KH domain proteins, heterogeneous nuclear ribonucleoprotein K and alphaCP-2KL, suggest Distinct modes of RNA recognition. J Biol Chem 276:17484-17496.

31. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. 2004. Systematic identification and analysis of exonic splicing silencers. Cell 119:831-845.

32. Busch A, Hertel KJ. 2012. Evolution of SR protein and hnRNP splicing regulatory factors. Wiley Interdiscip Rev RNA 3:1-12.

33. Zhu J, Mayeda A, Krainer AR. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell 8:1351-1361.

34. Gordon H, Ajamian L, Valiente-Echeverria F, Levesque K, Rigby WF, Mouland AJ. 2013. Depletion of hnRNP A2/B1 overrides the nuclear retention of the HIV-1 genomic RNA. RNA Biol 10:1714-1725.

35. Olsen HS, Cochrane AW, Rosen C. 1992. Interaction of cellular factors with intragenic cis-acting repressive sequences within the HIV genome. Virology 191:709-715.

36. Roy R, Durie D, Li H, Liu BQ, Skehel JM, Mauri F, Cuorvo LV, Barbareschi M, Guo L, Holcik M, Seckl MJ, Pardo OE. 2014. hnRNPA1 couples nuclear export and translation of specific mRNAs downstream of FGF-2/S6K2 signalling. Nucleic Acids Res 42:12483-12497.

37. Vincendeau M, Nagel D, Brenke JK, Brack-Werner R, Hadian K. 2013. Heterogenous nuclear ribonucleoprotein Q increases protein expression from HIV-1 Rev-dependent transcripts. Virol J 10:151.

38. Swanson CM, Sherer NM, Malim MH. 2010. SRp40 and SRp55 promote the translation of unspliced human immunodeficiency virus type 1 RNA. J Virol 84:6748-6759.

39. Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, McKnight SL. 2012. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell 149:768-779.

40. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL. 2012. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell 149:753-767.

41. Caceres JF, Screaton GR, Krainer AR. 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. Genes Dev 12:55-66.

42. Pinol-Roma S, Dreyfuss G. 1993. hnRNP proteins: localization and transport between the nucleus and the cytoplasm. Trends Cell Biol 3:151-155.

43. Martinez-Contreras R, Fisette JF, Nasim FU, Madden R, Cordeau M, Chabot B. 2006. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. PLoS Biol 4:e21.

44. Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, Martin MA. 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J Virol 59:284-291.

45. Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:W686-689.

46. Spire B, Sire J, Zachar V, Rey F, Barre-Sinoussi F, Galibert F, Hampe A, Chermann JC. 1989. Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunodeficiency virus. Gene 81:275-284.

47. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

48. Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U. 2011. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol 12:R79.

CHAPTER 5: USING APOBEC3G AS A MUTAGEN TO PROBE HIV-1 SPLICING SIGNALS

Background

HIV-1 RNA is spliced to produce more than 40 different transcripts, and regulated splicing is required to produce sufficient amounts of each mRNA to make the viral proteins. Full-length genomic RNA is the primary product of transcription, and it may remain unspliced as genomic RNAs or as the *gag* and *gag/pro/pol* mRNAs. All other viral mRNAs are made by splicing of this full-length transcript (1-3). All spliced transcripts first use splice donor D1 to splice to one of the downstream acceptors. A second splicing event may remove the *env* intron by splicing between D4 and A7. Transcripts that retain the *env* intron are called incompletely spliced or 4 kb transcripts, and those that splice out the *env* intron are called completely spliced or 2 kb transcripts.

The HIV-1 Vif protein is made by a splice from the 5' donor D1 to the 3' acceptor A1 without any further splicing at D2(1-3). Vif prevents the mutagenic effects of cellular APOBEC3G (A3G) on viral DNA by targeting A3G for ubiquitination and proteosomal degradation (4-6). In the absence of Vif, A3G is incorporated into progeny virions and causes C to T mutations on the newly synthesized negative strand of viral DNA during reverse transcription (7). It is therefore critical that the virus makes enough Vif to counter A3G hypermutation, but not too much Vif since overexpression would be at the expense of expression of other viral proteins. The jury is still out on the question of the toxicity of too much Vif protein per se. While some researchers have reported that high levels of Vif reduce Gag processing (8, 9), another group showed that the high levels of Vif are the result of mutations that cause oversplicing, and the excessive number of spliced *vif* transcripts reduces the number of unspliced *gag* transcripts (10). A third group finds that in naturally occurring variations, the level of *vif* transcripts does not affect other transcripts but high or low Vif has a fitness cost (11). Either way, the number of *vif* transcripts must be kept within certain parameters for optimal viral fitness. Because HIV-1 splicing is highly suppressed and inherently inefficient, cis-acting splicing regulatory elements (SRE) are required to control splicing and splice acceptor usage. Many studies have found, characterized, and tested the effects of SREs acting on

HIV-1 splice acceptor A1 (12-16). Although even low levels of Vif provide enough protection from A3G to allow a near-normal degree of replication, Vif-minus viruses are completely replication defective in cells that express A3G (13).

Structure also plays an important role in the regulation of *vif* transcripts. The stem-loop structure containing splice acceptor A1 (SLSA1) is highly conserved (17), and naturally occurring variations in the structure result in varied levels of Vif (11). We have previously found that silent mutations that disrupt the SLSA1 structure caused a 10-fold reduction in *vif* transcripts (3). This loss in *vif* transcripts caused a barely noticeable replication defect (17), as is typical for mutations that partially down-regulate Vif (13).

Considering the lack of fitness loss, we asked if the SLSA1 mutant virus had a higher mutation rate than the WT NL4-3, based on our previous splicing data (3). We used Primer ID technology to accurately identify mutations and mutation rates. Primer ID is a random sequence tag in the cDNA primer, which is incorporated in the cDNA synthesis step and all PCR products subsequently made from that tagged cDNA (18). Each cDNA gets a unique Primer ID and after sequencing, the PCR products can be sorted by their Primer IDs. No matter how many times any one Primer ID occurs, each Primer ID sequence is counted only once, and skewing in the PCR steps is filtered out. Additionally, a consensus sequence is made from the multiple reads with the same Primer ID, making it possible to filter out PCR and sequencing errors and identify valid mutations.

For this analysis of mutation rates, we looked at a sequence between splice acceptor A5 and splice donor D4 (Figure 5.1). This is a sixty-one nucleotide sequence that starts 6 bases downstream of A5 and ends 1 base before D4. These boundaries excluded bases that are part of the splicing consensus sequences. This region is present in all HIV-1 transcripts. Additionally, it is located immediately adjacent to the Primer ID-tagged reverse primers used for cDNA synthesis and is in the early and accurate part of the sequencing reads, making it well suited for identifying mutations.

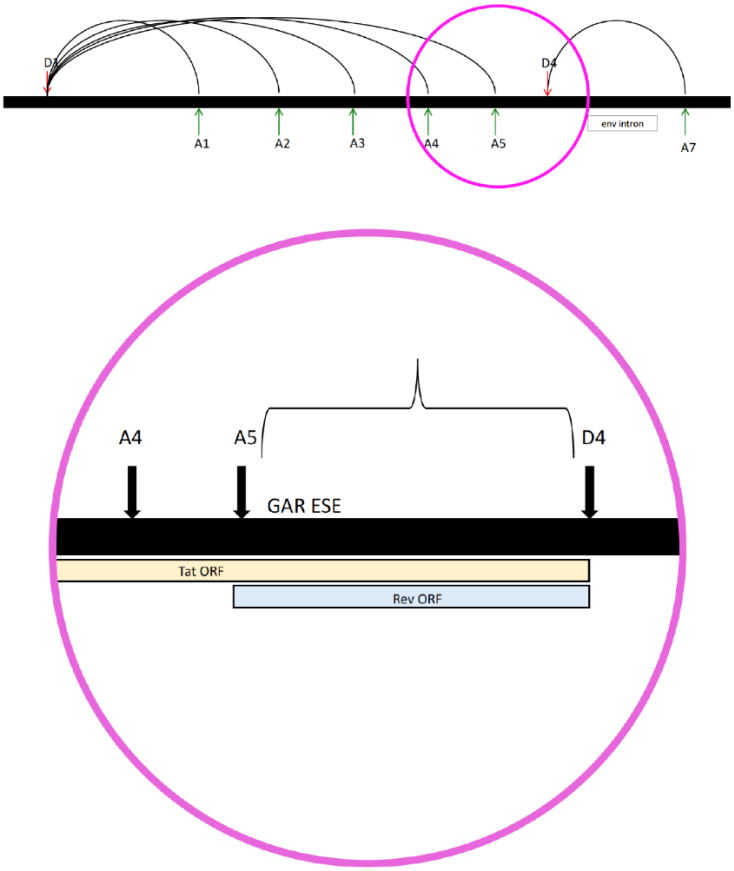

Figure 5.1. Target sequence between A5 and D4.
The sequence used to quantify mutations is in brackets and contains part of the GAR ESE element and is part of both the Tat and Rev ORFs.

This target sequence between A5 and D4 contains part of the ORFs for both Tat and Rev, plus a splicing enhancer named GAR ESE (guanine-adenine rich exonic splicing enhancer) which has been previously identified as necessary for the splicing of *env* transcripts (use of acceptors A4 or A5) (19, 20). Besides the GAR ESE element, the rest of the sequence between A5 and D4 was also found to be essential for splicing but the specific bases involved were not defined (20-22). This previous work used artificial splicing constructs and PCR product/gel quantification. While such artificial constructs are useful for screening potential SREs, they may not represent splicing in the full-length virus. The constructs used to characterize the region between A5 and D4 failed to exhibit normal HIV-1 splicing behavior and therefore the conclusions drawn from them must await confirmation by an in-context demonstration of their validity. Our straightforward assay quantifies HIV-1 RNA splicing patterns in the context of viral infection.

When we analyzed mutation rates, the results were unexpected. Specific mutations collected differentially in the two HIV-1 size classes, suggesting that the mutations themselves impacted splicing behavior and thus created a genetic screen that correlates mutations in the A5 to D4 sequence with both upstream and downstream changes in splicing.

Sequencing and Bioinformatics Methods

As previously described in Chapter 2 for the SLSA1 experiment, CEMx174 cells were infected with either WT NL4-3 or SLSA1 mutant viruses. RNA was harvested after 5 days of spreading infection. Primer ID-tagged reverse primers specific for either the 1.8kb or 4kb splice size classes and a common forward primer were used to tag the 50+ HIV-1 spliced mRNAs with a Primer ID (Figure 2.1). Following amplification and Illumina paired-end deep sequencing, the data files for both the WT and mutant SLSA1 for both size classes were processed using the following algorithm. mRNAs were quantified according to spliced transcript type using the splicing program described in Chapter 2. Reads that occurred 3 or more times in the sequencing data (identified by their matching Primer IDs) were collapsed into consensus sequences spanning the A5 to D4 region shown in Figure 5.1. Each consensus sequence was then aligned with the WT A5 to D4 sequence and mutations were identified and scored. Only consensus sequences with two or fewer mutations were used, to filter out hypermutated or mis-primed reads. The consensus sequences were sorted and quantified according to mutation type. This information provided the data for calculations of total mutation rates and for specific mutation types.

To quantify the proportional representation of mutations in each size class, the consensus sequences were sorted and quantified according to mutation type. For each mutation type we calculated its percentage of the total consensus sequences. These relative proportions were then compared across size classes. The rates of upstream splice acceptor usage were also tabulated for each mutation type.

Results

<u>Mutations are not equally represented in the two HIV-1 spliced transcript size classes</u>

Figure 5.2 shows the mutation rate in the WT vs. SLSA1 mutant for both the 1.8 kb and 4 kb size classes. In the 4 kb transcripts, there was no increase in mutation rate in the SLSA1 sample, but there was an increase in mutation rate in the 1.8 kb samples.

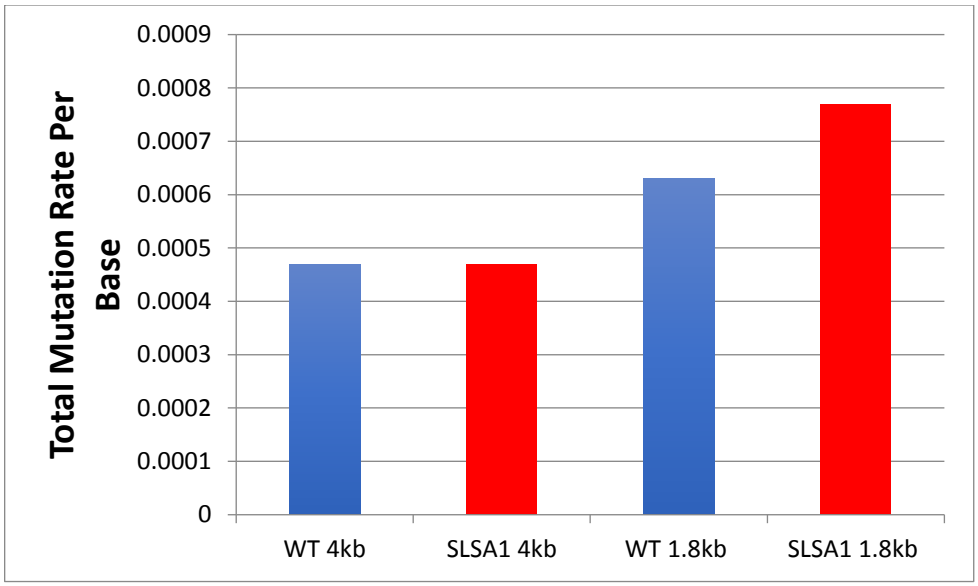

Figure 5.2. Mutation rates for WT or SLSA1 (low *vif*) in the 1.8 kb or 4 kb size classes.

Looking only at the A3G mutations (C-to-T, or G-to-A in the positive strand), we found that they increased in the SLSA1 mutant in both size classes, but that the proportion of these A3G mutations in the 1.8 kb class was about twice that in the 4 kb class (Figure 5.3). This was puzzling because both 1.8 kb and 4 kb transcripts are made from the same full length starting transcripts, which come from the same integrated genomic DNA. If the short transcripts are more mutated, the long ones should be mutated too and at the same rate, unless the mutations themselves increase splicing from D4 to A7.

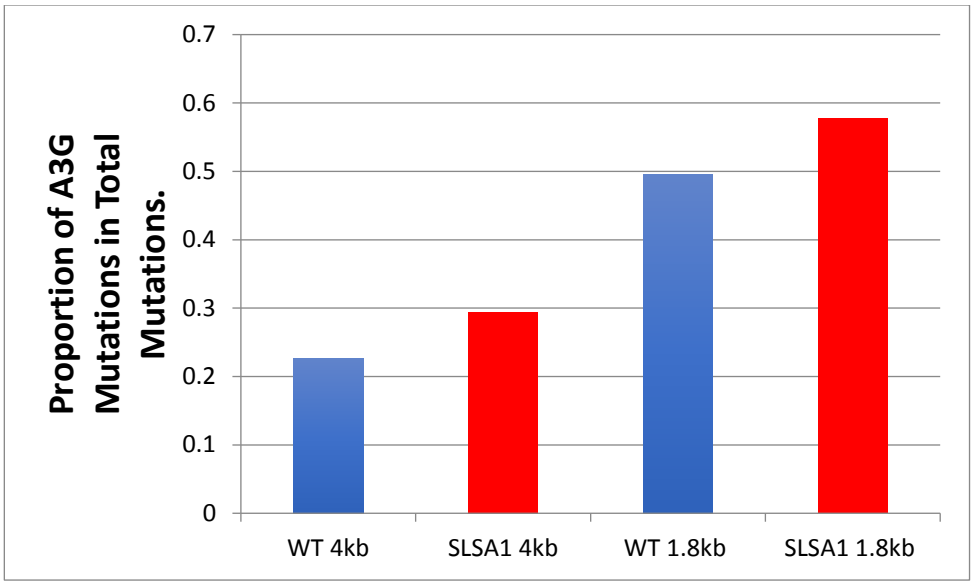

Figure 5.3. G-to-A mutations as a proportion of total mutations.

To avoid confusion between the two strands, the data will be discussed in the positive strand notation, where A3G mutations appear as G-to-A mutations. We looked at the proportional representation of each of the G-to-A mutations in the 1.8 kb and 4 kb size classes and found that they were in fact skewed toward the 1.8 kb class in both the wild type and the mutant SLSA1 samples.

Due to the effect of A3Gs on the G-to-A mutation rate, G-to-A mutations are present at a higher rate than other mutation types. Nevertheless, the depth of sequencing was sufficient to also measure the skew in A-to-G mutations, which are not so overrepresented. This indicated that all types of mutations, if present in sufficient amounts to allow for valid measurements, could be correlated with their effects on splicing. With this data set we were able to analyze two types of potential changes in splicing. First, if skewing was seen between the size classes we could find the mutations that increase or decrease the rate of splicing from D4 to A7. Second, this data set also contains the upstream splicing data which identifies the acceptor used when splicing from D1. With this information we could correlate specific mutations with upstream acceptor usage.

To do this extended mutational analysis we used the WT data only and not the SLSA1 mutant data to avoid introducing any possible additional splicing changes caused by the loss of A1 functionality. Reverse reads were sorted by Primer ID and reads with 3 or more occurrences were collapsed into a consensus sequence for the region between A5 and D4. These consensus sequences were aligned to

the WT sequence and mutations were scored and tabulated along with the splicing data.

Only sequences with 2 or fewer mutations were considered, as we did not want the data confounded by possible hypermutation across the genome outside of the observed area. Nevertheless, it should be noted that there may be unknown mutations outside of the A5 to D4 sequence that could affect splicing outcomes, but such mutations should be random and therefore unlikely to skew the results. Hypermutation was rare and greater than 97% of all consensus sequences were wild type NL4-3. This analysis produced a list of mutations across the A5-D4 sequence and correlated the mutations to both upstream and downstream splicing events.

G-to-A and A-to-G mutations are skewed in size class distribution

Figure 5.4 shows the ratio of proportional representation of the G-to-A mutations in the fully spliced 1.8 kb class compared to the partially spliced 4 kb class. This is a log scale, so an upward bar shows a skew towards short 1.8 kb transcripts and a downward bar shows a skew towards long 4 kb transcripts. G-to-A mutations skew strongly towards D4-A7 splicing (1.8 kb short transcripts). Combinations of G-to-A mutations were seen that skewed at even higher rates, though they are present in low numbers.

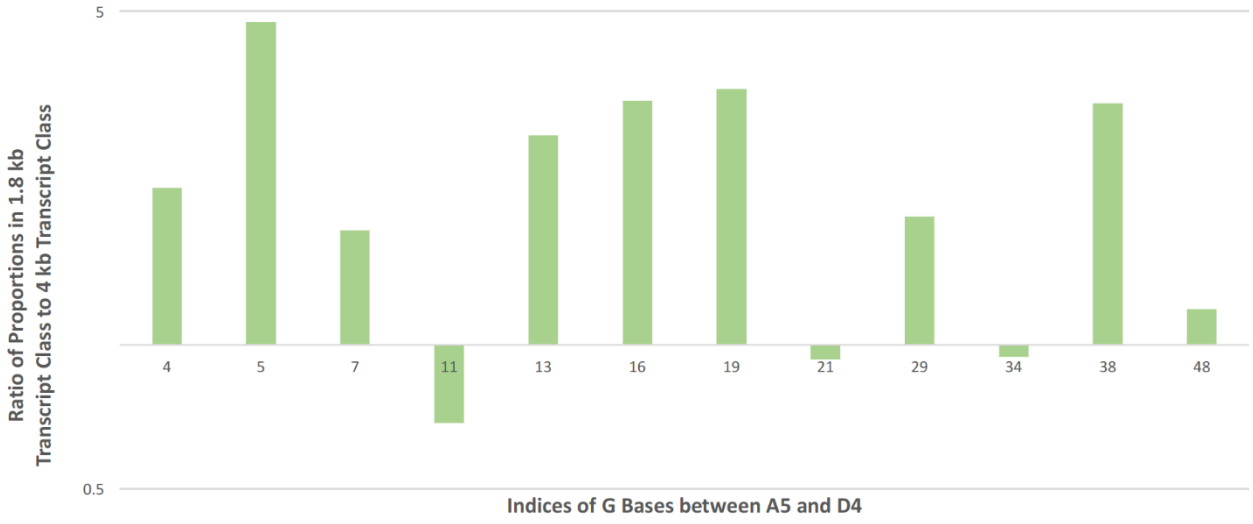

Figure 5.4. Skew of G-to-A mutations.
Y-axis is a log scale of the ratio of proportional representation of each of the mutations in the 1.8 kb size class relative to the 4 kb size class. The numbers on the X axis show the nucleotide position within the 61 nucleotide stretch between A5 and D4.

In contrast to the G-to-A mutations which largely caused skewing toward the 1.8 kb size class, we found the opposite effect for the A-to-G mutations (Figure 5.5). These mutations were more likely to be represented in the partially spliced 4 kb class than in the fully spliced 1.8 kb class. While the skew for any one base was less than seen with the G-to-A mutations, there are more As in this region of the genome – 22 of the 61 bases in the analyzed sequence are As.

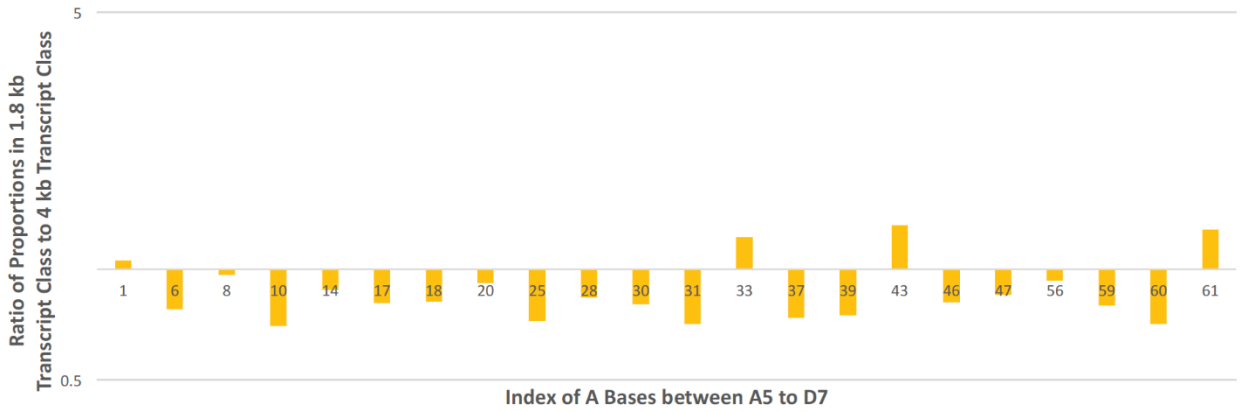

Figure 5.5. Skew of A-to-G mutations.
The Y-axis is a log scale of the ratio of proportional representation of each of the mutations in the 1.8 kb size class relative to the 4 kb size class. The numbers on the X axis show the nucleotide position within the 61 nucleotide stretch between A5 and D4.

After G/A transitions, C/T transitions were the next most numerous mutations in this A5 to D4 sequence. They did not show an obvious size class skew collectively, though some individual mutations have an impact on splicing (not shown). In previous work, mutations of the A or G elements in the GAR ESE sequence were replaced by C or T (21) or used a C and A oligo as a neutral splicing sequence (22). Our data show that the type of base in the mutation as well as the location is important. There are examples of single positions that have two different nucleotide substitutions with different effects on splicing. This suggests that these mutations affect binding of cellular splicing proteins.

Mutations between A5 and D4 affect A5 usage

In addition to splicing from D4 to A7, this data set also identifies mutations that change the choice of upstream acceptor sites. There are four A4 acceptor subsites that are used in different proportions in different HIV-1 strains (3, 23) combined with two A5 acceptor sites. All A4 and A5 sites are included in the group of acceptors defined by D4. They have shared branch points and overlapping polypyrimidine tracts, both involved in splice site selection (24). It is probable that the different A4 and A5 acceptors are

controlled by some of the same splicing regulatory elements.

The usage of A5 is up or downregulated by some of the mutations in the A5-D4 sequence as shown in Figure 5.6. There are hotspots that suggest these may be binding sites for regulatory proteins, some of which overlap with the mutations that control splicing from D4 to A7 and these effects may compound. Some of the bases have multiple mutations in the data set but with results specific to mutation type. For example, the mutation at base 1 shown in Figure 5.6 is an A-to-T mutation and causes a large reduction in A5 usage, but another mutation at base 1 from A-to-G has a smaller effect in the 1.8 kb size class and no effect in the 4 kb size class. This suggests again that these mutations identify sites of SR or hnRNP binding.

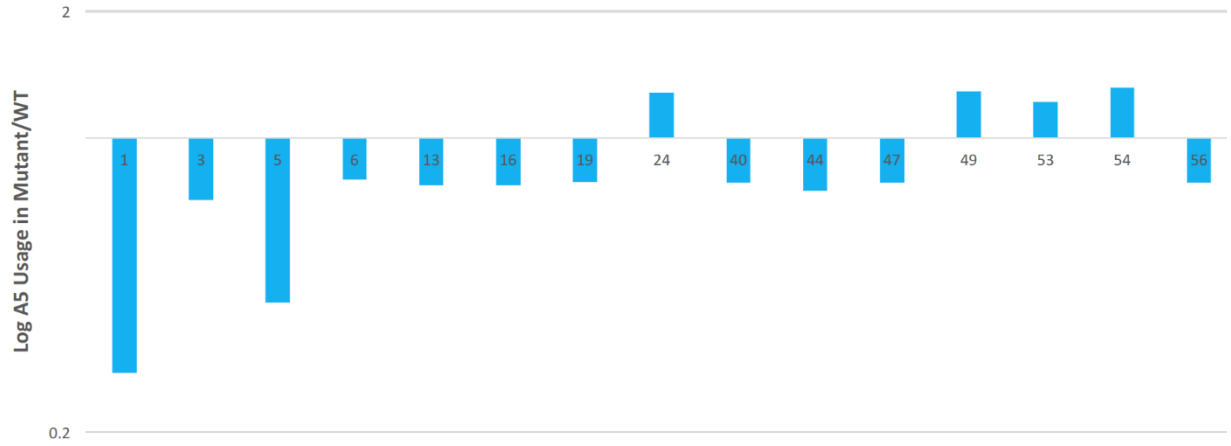

Figure 5.6. Mutations affecting the use of acceptor A5.
The X-axis shows the positions of the mutations within the 61 nucleotide region between A5 and D7. The Y-axis is a log scale of the ratio of A5 acceptor usage in the mutant relative to WT A5 usage.

Discussion

This analysis of the region between A5 and D4 introduces a way to identify and quantify the effects of mutations that affect upstream splicing to A4 and A5 and also affect downstream splicing between D4 and A7. Splicing regulatory elements (SRE) in this area have previously proven difficult to locate due to their modest effects and the confounding results of disturbing the Tat or Rev ORFs (20, 21, 25-27). Studies done thus far have utilized computer algorithms to identify SR protein binding sites and artificial splicing constructs, both with questionable outcomes (20, 21, 28).

The sequence from A5 and D4 is unique because it is the only region between the major splice donor D1 and the final splice acceptor A7 that is included in every HIV-1 transcript. The exon ending in

D4 can utilize the alternative A3 or A4 acceptors instead A5 but the resulting exons still contain the A5 to D4 sequence. This universal inclusion identifies the A5-D4 exon as a constitutive exon. All other HIV-1 exons are alternative, meaning they may or may not be included. Splicing of this constitutive exon seems to require a different kind of regulation from what has been seen with other alternative HIV-1 splicing events. ESE and ESS elements have been identified that control splicing to A1, A2, and A3; however, there are no similar control elements known for A4 or A5 (12-15, 29-39).

Previous work identifies the GAR ESE and E42 sequences as necessary both for splicing to A5 and for splicing between D4 and A7. This is an unsatisfying conclusion since GAR ESE and E42 combined constitute the entire sequence between A5 and D4 (20, 21). The GAR ESE is not a well-defined sequence that controls splicing and the boundary between the GAR ESE and E42 is arbitrary.

Mutation of the regulatory elements that control A1, A2, and A3 have dramatic effects but mutations to the A5 to D4 sequence cause only modest splicing changes (see Chapter 6). It is possible that the regulation of this constitutive exon needs only minimal additional control elements, in contrast to the suppression required for the alternative splices. Elements in the region of A7 have been described that control splicing between D4 and A7 (Chapter 6) (12, 34, 40, 41). These elements are part of extensive pre-mRNA secondary structures and much work has been done to characterize their cooperative/competitive binding to regulatory proteins. Interestingly, like the A5-D4 sequence, effects of mutations to these regions is small compared to the elements controlling splicing at A1, A2, and A3 (27, 42, 43).

Taken together, these modest mutational effects and the spread-out nature of the putative regulatory sequences suggest different mechanisms controlling splicing to A5 and from D4 to A7. We hypothesize that the constitutive nature of the A5-D4 exon requires very little additional control by cis-acting sequences. Like most known constitutive exons, it is likely to have an enhancer element, but perhaps not a powerful one, and no silencer (44, 45). It may be that although splicing to A1, A2, and A3 is controlled and highly suppressed, splicing to A4 and A5 is the default and does not require these multiple layers of control. The same is likely for splicing from D4 to A7 – splicing is the default and will happen unless suppressed. We hypothesize that suppression is the result of cooperative action between many spread-out regulatory elements (the A5-D4 region upstream of D4 and the known splicing control

elements downstream of A7), each relatively weak on its own. The SREs near A7 have already been shown to cooperatively bind hnRNP A1 *in vitro* and to propagate along the mRNA and this cooperativity may account for the modest splicing changes seen in mutation to any one hnRNP binding site.

The search for A5 to D4 control elements is complicated by the requirement for Tat and Rev. This A5 to D4 exon includes parts of both the Tat and Rev ORFs, each in a different reading frame, and both Tat and Rev are critical for splicing regulation, making it difficult to make mutations without confounding effects (46, 47).

One issue that needs to be addressed with this data set is the possibility that these size class effects are due to mutations in Rev. The sequence between A5 and D4 is part of the ORF for both Tat and Rev. Rev is required for nuclear export of unspliced viral RNA and of the partially spliced (4 kb size class) HIV-1 mRNAs (46). Deletions to Rev shift the balance of transcripts to fully spliced (unpublished data).

We do not favor this interpretation of our data. The sequence data used came from NL4-3 virus grown in CEMx174 cells over a 5-day spreading infection. CEMx174 cells form syncytia when infected and we waited to harvest RNA until syncytia were clearly evident and large. Since greater than 97% of all the sequences analyzed were WT NL4-3, we can expect that in multiply infected cells at least one of the infecting viruses was producing functional Rev. When full-length WT and mutant Rev viral plasmids are co-transfected, the WT Rev is sufficient to rescue normal splicing patterns in the mutant *rev* transcripts (unpublished data). Only a small amount of Rev is required for nuclear export, and the amount of Rev in an infected cell is so small that it cannot consistently be accurately measured (Sebla Kutluay- personal communication). An examination of the mutations found in this screen supports the presence of functional Rev, as the following examples illustrate. Some mutation pairs affect the same Rev residue, such as two mutations that change residue 5 from S to C or G. The first mutation is an A-to-T mutation at index 1 that decreases its inclusion in the 1.8 kb class – an unlikely outcome from a Rev mutant. The second mutation is an A-to-G mutation also at index one that causes an insignificant increase 1.8 kb inclusion (see Figure 5.5, index 1). There are several such Rev mutations in this data set with no apparent correlation between mutations and size class skewing. Additionally, some of the sequence mutations that skew representation in the size classes are silent in Rev. The A5-D4 sequence under

analysis here has not been found to harbor critical Rev residues that could have dominant negative effects (48). These facts lead us to believe that the splicing changes seen here are not due to loss of Rev.

A similar argument could be made for Tat, but is unimportant since any cells singly infected by a Tat-defective virus would not produce enough transcripts to skew the quantification. The presence of mutant *tat* transcripts is additional evidence for a multiplicity of infection greater than one.

The region between A5 and D4 is unique in that information for both the mutations and the splicing behavior is in the sequencing data. It is fortuitous that such an analysis works in this specific region where effects are combinatorial and modest and not likely to be identified individually. Splicing to A5 and A4 and from D4 to A7 has been difficult to study due to the bi-directional nature of the proposed SREs. Previous studies used splicing constructs that did not create an equivalent contextual setting to the full-length genome for HIV-1 splicing, and mutations were not studied in the full-length virus. This is the first analysis that measures the effects of mutations between A5 and D4 in the context of viral infection.

Previous studies identified SR binding sites using a computational algorithm, ESEfinder, to locate SR protein binding sites (28) in the ESE GAR region (21, 22). We question the results of this method. SR protein binding sites have degenerate consensus sequences (28) and many more matches to the consensus sequences exist than the valid SR binding sites. We submitted HIV-1 sequences likely to be involved in splicing regulation to ESEfinder and found that more than 50% of the bases in the sequences tested scored as a hit for at least one of the four SR protein binding sites used in ESEfinder. Randomly generated sequences scored equally well. Considering the nature of the known HIV-1 enhancer sequences and their limited locations, it is not reasonable to believe that binding sites identified by ESEfinder have high specificity in that it must be calling many false positive sites. This serves to illustrate the difficulty of identifying precise bases of SREs and the SR or hnRNP elements that bind them *in vivo*.

G-rich elements have been shown to have a splicing suppressive function (49), and two GGGG elements in HIV-1 suppress an upstream acceptor (15, 36). G-rich regions flanked/interspersed by A are preferred hnRNP H1 binding sites (Chapter 4). There are no GGGG regions in the A5 to D4 sequence but our set of mutations for this region suggests that many of the individual G bases, though spread out,

act to suppress downstream D4 to A7 splicing. Though it is not possible to nail down specific factor binding sites, our data identifies hot spots of suppression in regions 4 through 7 and 13 through 19.

While the current data set is informative, an even larger data set would more specifically identify the bases needed for splicing repression and activation. Combinations of mutations skew at even higher rates but the numbers are too low thus far for confident calculations. In addition to information about A5 usage, a larger data set would give sufficient depth of information to quantify A4 and A3 usage as well. The combinatorial nature of the mutations studied here invite a machine learning analysis on a larger data set.

This A5 to D4 location is unique in that the sequencing data contains both mutations and the resulting splicing behavior. This kind of correlation is possible with one other HIV-1 sequence – the area just upstream of D1. Like the area controlling A5 and D4, the region upstream of D1 has been challenging to study (50-56) and this type of deep sequencing analysis may go beyond what has been accomplished using artificial constructs.

REFERENCES

1. Purcell DF, Martin MA. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67:6365-6378.

2. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res 40:10345-10355.

3. Emery A, Zhou S, Pollom E, Swanstrom R. 2017. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. J Virol 91.

4. Sheehy AM, Gaddis NC, Malim MH. 2003. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nat Med 9:1404-1407.

5. Sheehy AM, Gaddis NC, Choi JD, Malim MH. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418:646-650.

6. Malim MH, Bieniasz PD. 2012. HIV Restriction Factors and Mechanisms of Evasion. Cold Spring Harb Perspect Med 2:a006940.

7. Lecossier D, Bouchonnet F, Clavel F, Hance AJ. 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science 300:1112.

8. Akari H, Fujita M, Kao S, Khan MA, Shehu-Xhilaga M, Adachi A, Strebel K. 2004. High level expression of human immunodeficiency virus type-1 Vif inhibits viral infectivity by modulating proteolytic processing of the Gag precursor at the p2/nucleocapsid processing site. J Biol Chem 279:12355-12362.

9. Lu R, Ghory HZ, Engelman A. 2005. Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. J Virol 79:10356-10368.

10. Mandal D, Feng Z, Stoltzfus CM. 2008. Gag-processing defect of human immunodeficiency virus type 1 integrase E246 and G247 mutants is caused by activation of an overlapping 5' splice site. J Virol 82:1600-1604.

11. Nomaguchi M, Doi N, Sakai Y, Ode H, Iwatani Y, Ueno T, Matsumoto Y, Miyazaki Y, Masuda T, Adachi A. 2016. Natural Single-Nucleotide Variations in the HIV-1 Genomic SA1prox Region Can Alter Viral Replication Ability by Regulating Vif Expression Levels. J Virol 90:4563-4578.

12. Kammler S, Otte M, Hauber I, Kjems J, Hauber J, Schaal H. 2006. The strength of the HIV-1 3' splice sites affects Rev function. Retrovirology 3:89.

13. Mandal D, Exline CM, Feng Z, Stoltzfus CM. 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. J Virol 83:6067-6078.

14. Madsen JM, Stoltzfus CM. 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.

15. Exline CM, Feng Z, Stoltzfus CM. 2008. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol 82:3921-3931.

16. Widera M, Erkelenz S, Hillebrand F, Krikoni A, Widera D, Kaisers W, Deenen R, Gombert M, Dellen R, Pfeiffer T, Kaltschmidt B, Munk C, Bosch V, Kohrer K, Schaal H. 2013. An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA. J Virol 87:2707-2720.

17. Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM, 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 Pathog 9:e1003294.

18. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A 108:20166-20171.

19. Kammler S, Leurs C, Freund M, Krummheuer J, Seidel K, Tange TO, Lund MK, Kjems J, Scheid A, Schaal H. 2001. The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. Rna 7:421-434.

20. Caputi M, Freund M, Kammler S, Asang C, Schaal H. 2004. A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. J Virol 78:6517-6526.

21. Asang C, Hauber I, Schaal H. 2008. Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. Nucleic Acids Res 36:1450-1463.

22. Hillebrand F, Peter JO, Brillen AL, Otte M, Schaal H, Erkelenz S. 2017. Differential hnRNP D isoform incorporation may confer plasticity to the ESSV-mediated repressive state across HIV-1 exon 3. Biochim Biophys Acta 1860:205-217.

23. Delgado E, Carrera C, Nebreda P, Fernandez-Garcia A, Pinilla M, Garcia V, Perez-Alvarez L, Thomson MM. 2012. Identification of new splice sites used for generation of rev transcripts in human immunodeficiency virus type 1 subtype C primary isolates. PLoS One 7:e30574.

24. Swanson AK, Stoltzfus CM. 1998. Overlapping cis sites used for splicing of HIV-1 env/nef and rev mRNAs. J Biol Chem 273:34551-34557.

25. Ropers D, Ayadi L, Gattoni R, Jacquenet S, Damier L, Branlant C, Stevenin J. 2004. Differential effects of the SR proteins 9G8, SC35, ASF/SF2, and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA. J Biol Chem 279:29963-29973.

26. Marchand V, Santerre M, Aigueperse C, Fouillen L, Saliou JM, Van Dorsselaer A, Sanglier-Cianferani S, Branlant C, Motorin Y. 2011. Identification of protein partners of the human immunodeficiency virus 1 tat/rev exon 3 leads to the discovery of a new HIV-1 splicing regulator, protein hnRNP K. RNA Biol 8:325-342.

27. Marchand V, Mereau A, Jacquenet S, Thomas D, Mougin A, Gattoni R, Stevenin J, Branlant C. 2002. A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding. J Mol Biol 323:629-652.

28. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31:3568-3571.

29. Zahler AM, Damgaard CK, Kjems J, Caputi M. 2004. SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. J Biol Chem 279:10077-10084.

30. Si Z, Amendt BA, Stoltzfus CM. 1997. Splicing efficiency of human immunodeficiency virus type 1 tat RNA is determined by both a suboptimal 3' splice site and a 10 nucleotide exon splicing silencer element located within tat exon 2. Nucleic Acids Res 25:861-867.

31. Madsen JM, Stoltzfus CM. 2005. An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. J Virol 79:10478-10486.

32. Jacquenet S, Mereau A, Bilodeau PS, Damier L, Stoltzfus CM, Branlant C. 2001. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. J Biol Chem 276:40464-40475.

33. Domsic JK, Wang Y, Mayeda A, Krainer AR, Stoltzfus CM. 2003. Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts. Mol Cell Biol 23:8762-8772.

34. Amendt BA, Si ZH, Stoltzfus CM. 1995. Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. Mol Cell Biol 15:4606-4615.

35. Amendt BA, Hesslein D, Chang LJ, Stoltzfus CM. 1994. Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. Mol Cell Biol 14:3960-3970.

36. Widera M, Hillebrand F, Erkelenz S, Vasudevan AA, Munk C, Schaal H. 2014. A functional conserved intronic G run in HIV-1 intron 3 is critical to counteract APOBEC3G-mediated host restriction. Retrovirology 11:72.

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

38. Bilodeau PS, Domsic JK, Mayeda A, Krainer AR, Stoltzfus CM. 2001. RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element. J Virol 75:8487-8497.

39. Schaub MC, Lopez SR, Caputi M. 2007. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. J Biol Chem 282:13617-13626.

40. Staffa A, Cochrane A. 1994. The tat/rev intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3' splice site. J Virol 68:3071-3079.

41. Si ZH, Rauch D, Stoltzfus CM. 1998. The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. Mol Cell Biol 18:5404-5413.

42. Damgaard CK, Tange TO, Kjems J. 2002. hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. Rna 8:1401-1415.

43. Levengood JD, Rollins C, Mishler CH, Johnson CA, Miner G, Rajan P, Znosko BM, Tolbert BS. 2012. Solution structure of the HIV-1 exon splicing silencer 3. J Mol Biol 415:680-698.

44. Shen H, Kan JL, Green MR. 2004. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell 13:367-376.

45. Wang Z, Xiao X, Van Nostrand E, Burge CB. 2006. General and specific functions of exonic splicing silencers in splicing control. Mol Cell 23:61-70.

46. Pollard VW, Malim MH. 1998. The HIV-1 Rev protein. Annu Rev Microbiol 52:491-532.

47. Karn J, Stoltzfus CM. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med 2:a006916.

48. Malim MH, McCarn DF, Tiley LS, Cullen BR. 1991. Mutational definition of the human immunodeficiency virus type 1 Rev activation domain. J Virol 65:4248-4254.

49. Han K, Yeo G, An P, Burge CB, Grabowski PJ. 2005. A combinatorial code for splicing silencing: UAGG and GGGG motifs. PLoS Biol 3:e158.

50. Mueller N, van Bel N, Berkhout B, Das AT. 2014. HIV-1 splicing at the major splice donor site is restricted by RNA structure. Virology 468-470:609-620.

51. Mueller N, Klaver B, Berkhout B, Das AT. 2015. Human immunodeficiency virus type 1 splicing at the major splice donor site is controlled by highly conserved RNA sequence and structural elements. J Gen Virol 96:3389-3395.

52. Mueller N, Berkhout B, Das AT. 2015. HIV-1 splicing is controlled by local RNA structure and binding of splicing regulatory proteins at the major 5' splice site. J Gen Virol 96:1906-1917.

53. Abbink TE, Berkhout B. 2008. RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. J Virol 82:3090-3098.

54. Lenz C, Scheid A, Schaal H. 1997. Exon 1 leader sequences downstream of U5 are important for efficient human immunodeficiency virus type 1 gene expression. J Virol 71:2757-2764.

55. Erkelenz S, Mueller WF, Evans MS, Busch A, Schoneweis K, Hertel KJ, Schaal H. 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. Rna 19:96-102.

56. Asang C, Erkelenz S, Schaal H. 2012. The HIV-1 major splice donor D1 is activated by splicing enhancer elements within the leader region and the p17-inhibitory sequence. Virology 432:133-145.

CHAPTER 6: A SURVEY OF KNOWN HIV-1 SPLICE REGULATORY ELEMENTS (SRE): A RE-EVALUATION

Background

Splicing in HIV-1 requires cis-acting sequence elements and trans-acting cellular proteins to define and regulate the position and usage of splice sites. The splicing consensus sequences at the splice sites are not sufficient by themselves to correctly distinguish correct splice sites from pseudo splice sites. Cis-acting sequences outside of the splicing consensus sequences bind cellular splicing factors to regulate splicing. These sequences are called splicing regulatory elements (SRE). There are two classes of SREs – splicing enhancers and splicing silencers, which are named according to their locations. Enhancers found in exons are exonic splicing enhancers (ESE) and enhancers in introns are intronic splicing enhancers (ISE). Likewise, silencers are exonic splicing silencers (ESS) or intronic splicing silencers (ISS) (1). In general, the SR family of proteins bind to enhancers to promote the formation of the spliceosome, while hnRNP proteins bind to silencers to occlude spliceosome formation (2, 3).

Chapter 1 summarizes the work done to date to identify and characterize SREs in HIV-1. This chapter describes mutations made to a subset of the known splicing regulatory elements (SRE) shown in Figure 6.1. These sites undoubtedly represent only a subset of regulatory sites with more remaining to be discovered. However, as part of our efforts to gain a fuller view of HIV-1 splicing regulation we used two deep sequencing approaches to characterize splicing in a set of viruses with mutations in these sites. In addition to the previously described splicing assay (Chapter 2), this study required the development of a second new deep sequencing assay to quantify changes in the relative amounts of spliced, partially spliced, and unspliced transcripts. Previous work done to characterize SREs used artificial sub-genomic constructs, but an important feature of these new studies is that they were done in the context of the full length viral genome. As a result of these studies we were able to confirm the contribution of some elements and refute the proposed role of other elements. We were able to quantify the loss of splicing

repression when ESS elements were mutated, leading to oversplicing from D1 and a loss of full-length viral mRNAs.

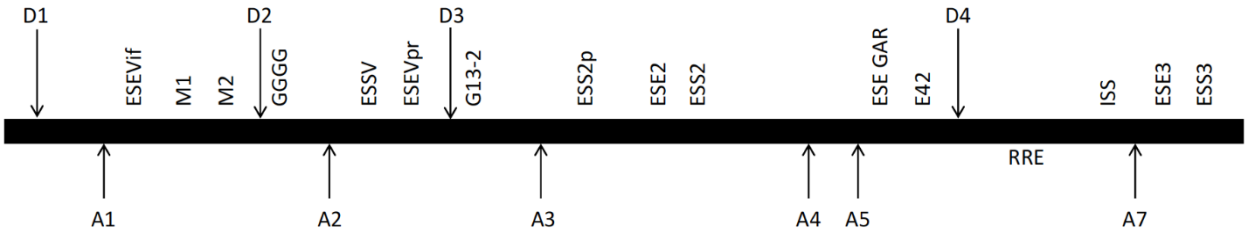

Figure 6.1. Relative positions of known HIV-1 SREs.

Materials and Methods

Cells and constructs

Silent mutations were made to the following SREs: ESEvif, ESEM1, ESEM2, ESSV, ESS2, ESS2p, and ISS. The mutated sequences were inserted into the full-length HIV-1 genome pNL-CH. pNL-CH is a variant of the NL4-3 strain, adapted with two silent mutations for efficient cloning. These full-length constructs were used to make virus and infect CEMx174 cells followed by total cellular RNA extraction, as described in Chapter 2. This RNA was used as input for the splicing assay (Chapter 2) and also for the Random Reverse Primer Assay (RRPA) described below.

A set of 12 mutations were made to the stem-loop structure of ESS3 in a ΔRev construct. ΔRev is a modification of NL4-3 with premature stop codons early in the Rev ORF. These ΔRev/ESS3 constructs were co-transfected with NL4-3 into 293T cells and total cellular RNA harvested and used as input for the Primer ID splicing assay (Chapter 2). Sequence differences between the mutant and wild type genomes allowed the spliced RNAs generated from each genome to be distinguished.

Sequencing and bioinformatics

Sequencing was done using the Ilumina Mi-Seq Platform with 300 base paired-end reads as described in Chapter 2. The program used to analyze the sequence data for spliced RNAs, described in Chapter 2, was adapted to accommodate the silent mutations created to test the role of cis-acting sequences.

A new program was written to process the random reverse primer sequencing reads. The reads were sorted according to spliced or unspliced. If spliced, the acceptor site was identified. Spliced reads

of sufficient length were identified as totally spliced or partially spliced. The number of spliced transcripts was compared to the number of unspliced transcripts of equal length. A minimum read length was used to allow partially spliced and totally spliced transcripts to be distinguished. To enrich for longer sequences, we used the Blue Pippin size separation system to extract reads from 250 to 1500 bases.

For the ESS3 mutations, the ESS3 mutant sequences were separated from the NL4-3 sequences based on the mutations to the Rev ORF. The separated sequence files were used as input to the random reverse primer program, which identifies partially spliced or fully spliced transcripts.

The Random Reverse Primer Assay (RRPA)

The splicing assay described in Chapter 2 can quantify splicing within each size class but cannot compare the relative amounts of the two size classes or detect changes in the amount of spliced to unspliced transcripts. We wanted to develop an assay that could use Primer ID to quantify these other types of changes (4). This is a challenging problem, as there is no single sequence element present in the partially spliced transcripts that is not also present in the unspliced transcripts. Additionally, in order to get reproducible results, a single Primer ID-tagged cDNA primer must be used – all transcripts must be picked up by the same reverse primer in the same reaction.

We used a random reverse primer in which the primer is the Primer ID. The random sequence is 14 nucleotides long, followed by the Illumina adapter sequences. A basic schematic of the HIV-1 genome is shown in Figure 6.2. D1 splices to one of the A1-A5 acceptors shown grouped together, followed by an optional splice from D4-A7. The random primer anneals to and primes all along the RNA. The forward primer is just upstream of D1. If the RNA is unspliced and the random reverse primer primes downstream of the red bracket (top line), the forward sequencing read will show no splice at D1. If the RNA is spliced but not from D4-A7 (middle line), the forward read will show a splice junction from D1 to one of the acceptor sites, and then have some part of the sequence between D4 and A7, which did not get spliced out. If the RNA is completely spliced the sequencing read will contain both the D1 to A1-5 junction and the D4-A7 junction sequences. cDNA synthesis and the Illumina sequencing platform are limited in the length of fragments they can produce. Looking at the subset of reads at least as long as the red bracket gives proportions of unspliced, partially spliced, and fully spliced reads.

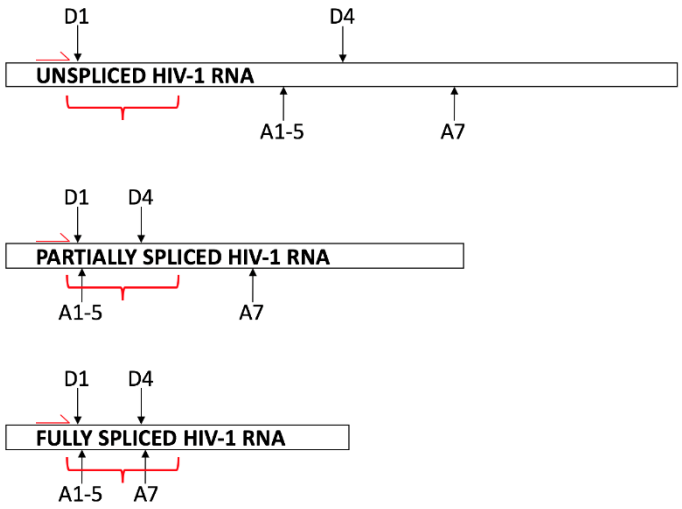

Figure 6.2. Random reverse primer outcomes.

The RRPA has some limitations. It does not prime with equal affinity at all locations, but the pattern of priming is highly reproducible; therefore, ratios of the transcript classes are not absolute counts but the assay can measure changes in the ratios. Additionally, though there may be 67 million possible random sequences in the 14-base reverse primer tube, only a small fraction of those are complementary to the sequence of the HIV-1 genome, and thus many of the Primer IDs are likely to be used more than once (Primer ID resampling). Also, there are many repeated sequences in the HIV-1 genome that could potentially be tagged by the same Primer ID. This led to development the forward tag.

The forward tag is another Primer ID random nucleotide sequence in the forward primer. It is attached after the cDNA synthesis step in a single round of PCR, described in the processing of the founder/transmitted viruses (Chapter 2). This creates a double-stranded DNA with a random sequence tag on both ends. Any repetition of a reverse Primer ID can be distinguished by the Primer ID tag in the forward primer. In effect, the forward primer squares the ~67 million possible Primer IDs. In practice, the forward tag is usually not required as the base-pairing fidelity of the reverse primer is low. The sequencing data shows that the random reverse sequence will prime at many locations that have only minimal complementarity to the genomic RNA at that site, and thus the random reverse primers cover the entire HIV-1 genome (raising some interesting issues concerning the fidelity of PCR in general). Thus the forward tag is an option that can be used in conjunction with the random reverse primer to reduce Primer ID resampling when the sequence depth is high.

Results

Silent mutations to seven known splicing control elements shown in Figure 6.3 were made alone and in combinations and cloned into a complete viral genome, then splicing was quantified using both the Primer ID splicing assay and the RRPA. Additionally, a panel of non-synonymous mutations was made to the most downstream SRE shown in Figure 6.3, ESS3.

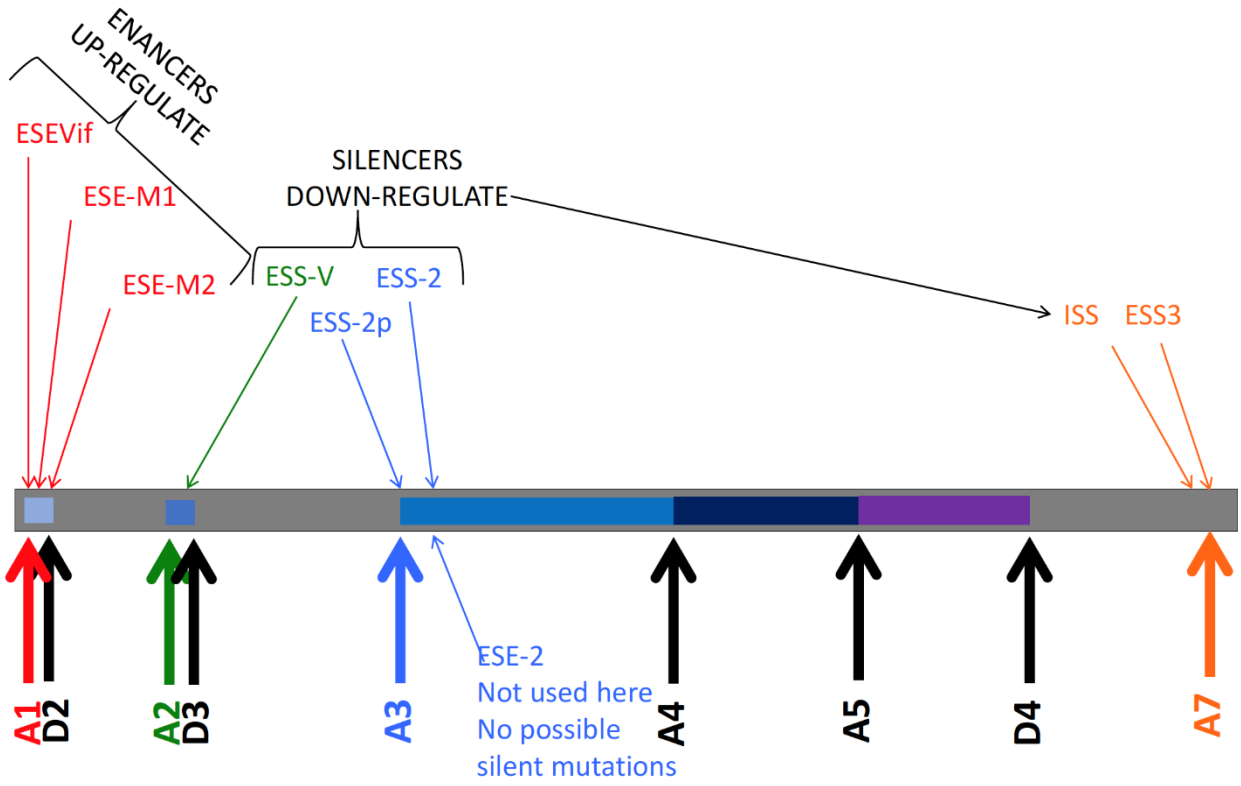

Figure 6.3. Splicing regulatory elements mutated.
SREs are color coded to match the expected affected acceptor site. Enhancer elements (ESE) upregulate splicing while suppressor elements (ESS, ISS) downregulate splicing.

Mutations to ESEVif, ESEM1, and ESEM2 alter splicing to acceptor A1

Figure 6.4 shows the results of the mutations to ESEVif, M1, and M2 alone and in combination. The Y axis is the log of the ratio of splice acceptor A1 usage by the mutants relative to the wild type, with the X axis showing the value for the different mutants in a histogram format. If splicing were unchanged, the ratio would be 1 on the Y axis. Bars that extend up indicate an increase in A1 usage and bars that go down show decreased A1 usage relative to the wild type. For this set of regulatory elements, the only acceptor affected was A1 so only A1 usage is presented. Mutation of M1 or M2 caused a 6-fold or 52-fold

decrease respectively in the number of transcripts that spliced to A1, confirming their roles as splicing enhancers. Unexpectedly, the ESEVif mutant showed a slight increase in A1 usage. This increase was modest in the single mutant, but the effect was consistently seen when combined with mutations of M1, M2 or both. In other words, ESEVif acted like a silencer rather than an enhancer. This suggests that ESEVif was misidentified in previous studies and is actually exonic splicing *silencer*, i.e. ESSVif. This fits with the fact that alternate exons need a silencer as well as an enhancer. It appears as if the mutations that were initially used to find this regulatory element inadvertently created an hnRNP A1 binding site and thus increased its activity as a silencer (5). It is also possible that the mutations we introduced could have created an enhancer instead of mutating a silencer. The RRPA showed no changes in the percentage of unspliced transcripts and no changes in the proportions of the two spliced transcript size classes relative to each other (data not shown).

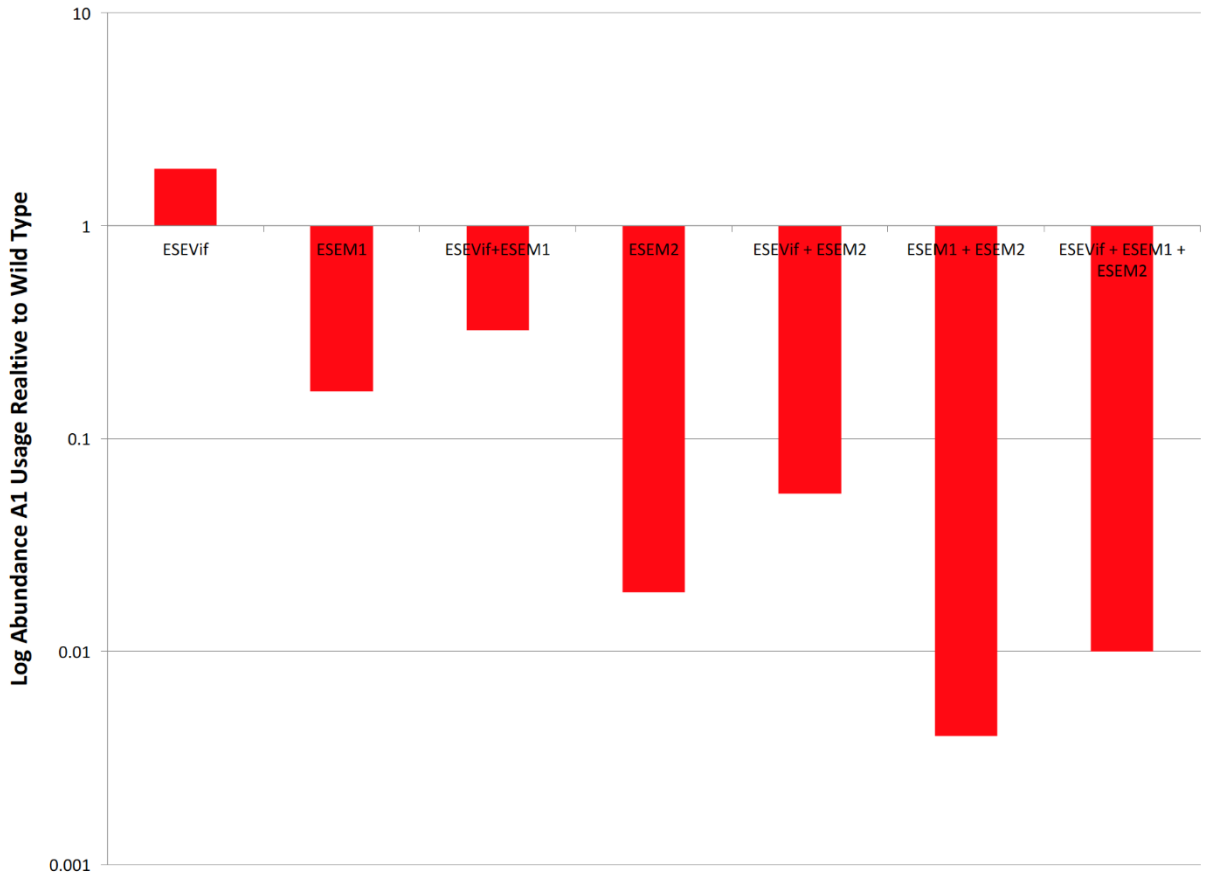

Figure 6.4. Usage of acceptor A1 in mutated SREs between A1 and D2.

It is probable that ESEVif is the silencer predicted to exist in the alternative exon from A1 to D2, and if not, the question arises as to why there might be such a high density of enhancers in this area with no silencer. One possible explanation may be related to the fact that this exon is only 50 bases long. This is short for an exon and short exons may need extra enhancers to be included (6). Thus it is possible that a silencer is not needed for exon skipping. More work and further mutations are needed to definitively characterize ESEVif and its role as either an enhancer or a silencer. ESEVif serves to illustrate some of the difficulties in identifying splicing regulatory elements. The trans-acting cellular factors regulating splicing are generally functionally redundant, bind promiscuously and sometimes cooperatively, and recognize multiple motifs. With so little known about the binding splicing regulatory factors, it is challenging to introduce mutations with confidence that won't create new cis-acting regulatory factor binding sites.

Mutation of ESSV activates acceptor A2 usage and also activates D1 to cause oversplicing

ESSV is a silencer for A2, and Figure 6.5 shows the changes in acceptor usage in splices from D1, again on a log scale. We would expect mutation of ESSV to increase splicing to A2, which it does, and splicing to other acceptors is greatly reduced. This greater than 10-fold increase at A2 represents more than 97% of spliced transcripts – essentially everything that splices now splices to A2.

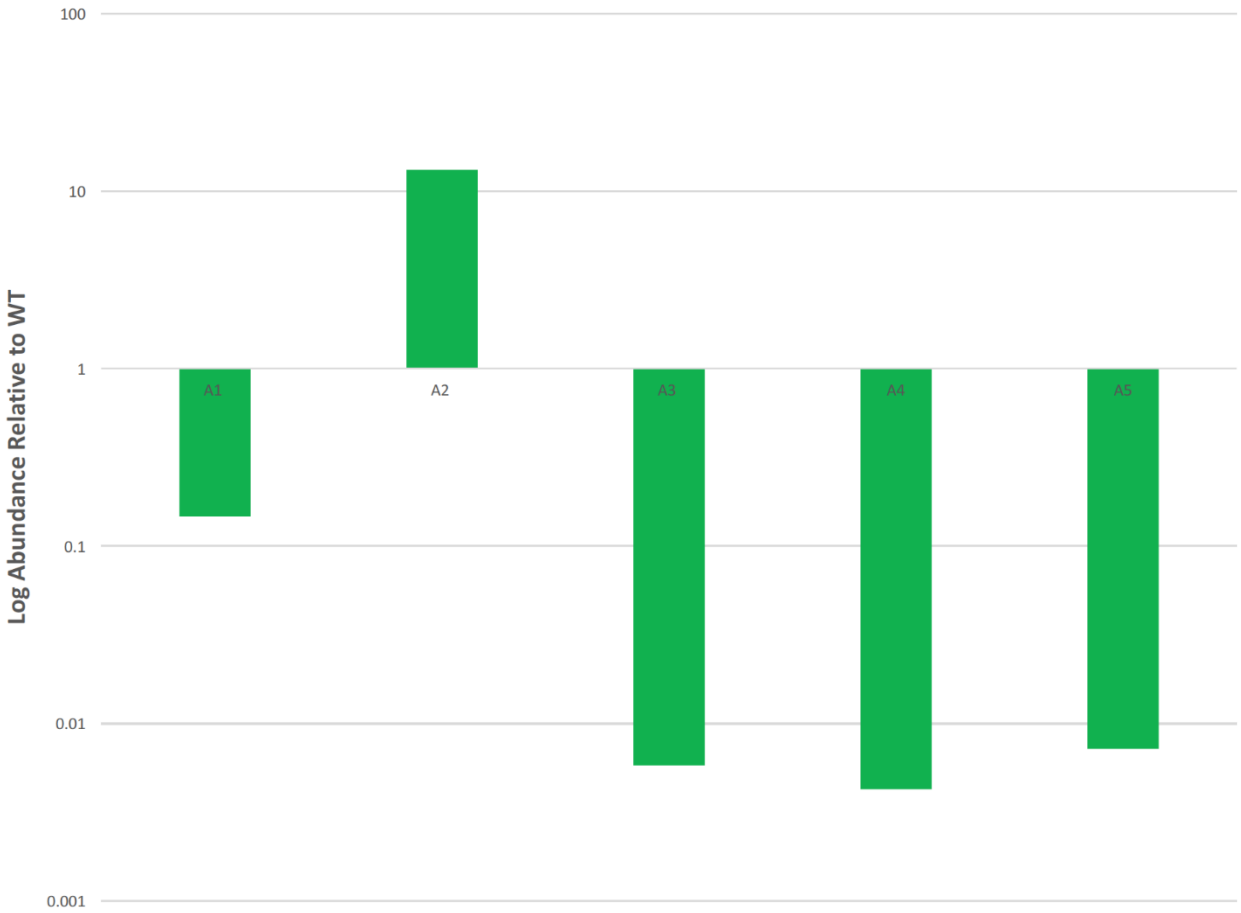

Figure 6.5. Acceptor usage in ESSV mutant relative to WT.

Figure 6.6 shows the random reverse primer data for the ESSV mutant. The percentages of unspliced transcripts for the WT are on the left in yellow and the ESS2 mutant on the right in green. In the WT almost 80% of all transcripts are unspliced. In contrast, the loss of repression at A2 activates D1 so that few transcripts are left unspliced. There is no activation of D4, so the ratio of incompletely spliced transcripts to completely spliced transcripts is the same in the ESSV mutant as the WT (data not shown).

Figure 6.6. Percent of unspliced transcripts in the WT and the ESSV mutant.

Mutation of ESS2 and ESS2p synergistically activate A3 and D1 to cause oversplicing

There are two suppressors that regulate splicing at A3: ESS2 and ESS2p. This area also has an enhancer element but silent mutations to disrupt it were not possible and thus its role in splicing was not tested. Figure 6.7 shows the changes in acceptor usage compared to the WT. The lighter blue bars are with mutations in ESS2 or ESS2p alone, and the dark blue bars represent splicing when both mutations were combined. As expected there was a large increase in A3 usage when these two elements were mutated, and the effects were additive. There was no significant change in splicing to acceptors A1 or A2. This increase in A3 usage came at a loss of A4 and A5 usage, which could be because A3, A4, and A5 are alternate acceptors on the same exon defined by D4. This suggests that to some degree, usage of A3 is in competition with A4 and A5, even though it is farther upstream and does not share the group of common branch points and polypyrimidine tracts (PPT) used in splicing to A4 and A5.

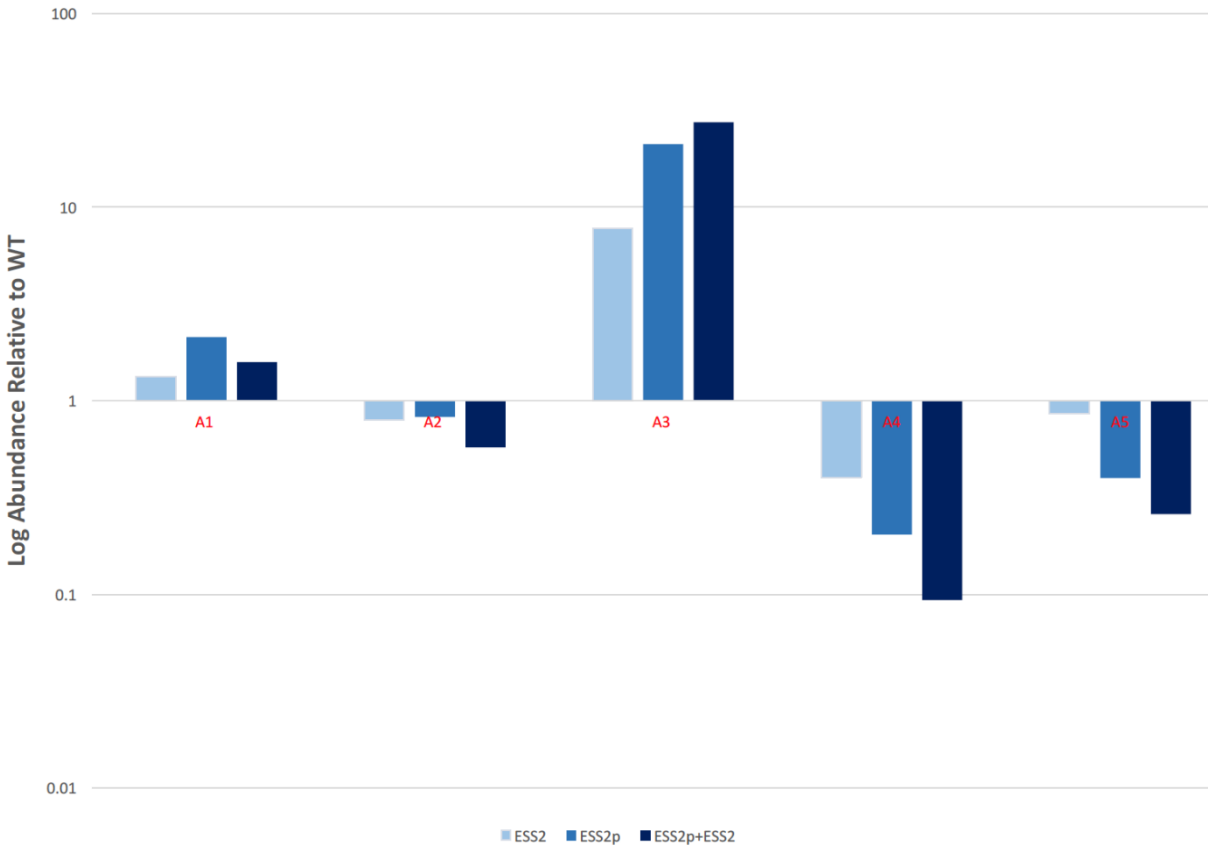

Figure 6.7. Acceptor usage for mutations to ESS2 and ESS2p.

Like mutation of ESSV, these silencer mutations also activate D1 to cause oversplicing, though not to the same degree, and they appear to act synergistically. Figure 6.8 shows the percentage of unspliced transcripts. The WT is in yellow on the left. When comparing the effects of these mutations on unspliced transcripts, ESS2 alone looked like the unmutated WT and had no effect. ESS2p had a significant effect alone, but when combined with ESS2, they created another oversplicing situation, though not as extreme as that seen with the ESSV mutation. A previous study found ESS2p to have a weaker effect than ESS2 and that they act independently, but the results shown in Figure 6.8 suggest the opposite (7). The ratio of partially spliced to completely spliced was not affected (data not shown).

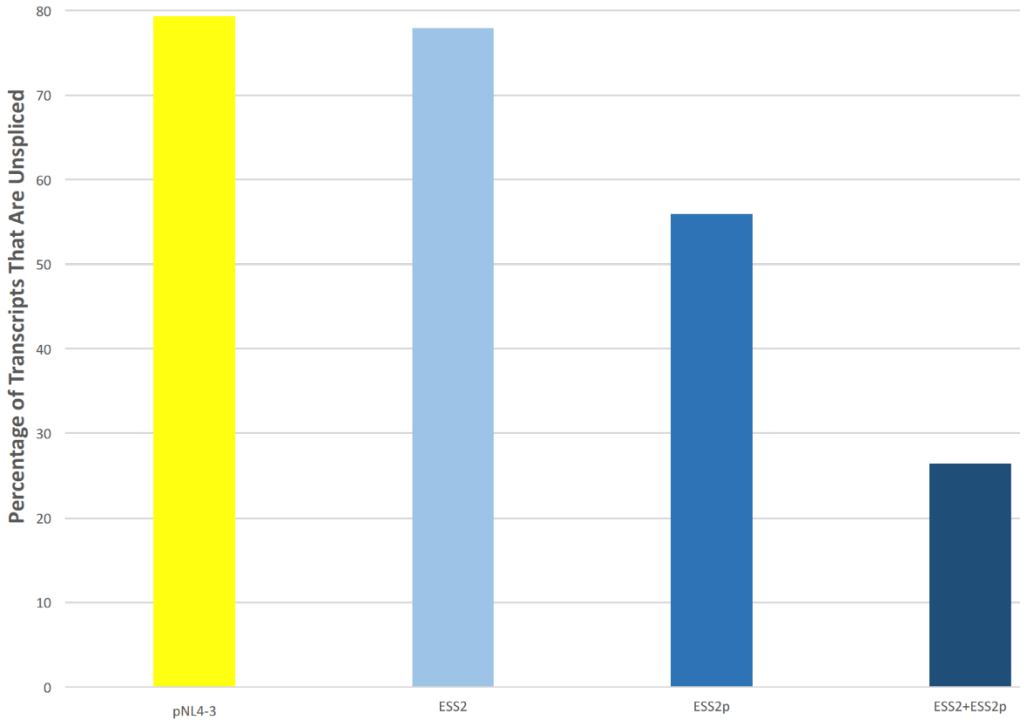

Figure 6.8. Effects of ESS2 and ESS2p mutation on percentage of unspliced transcripts.

Combined mutation of ESSV, ESS2p, and ESS2 caused oversplicing to both A2 and A3

Since mutations of silencers that control both A2 and A3 caused increased use of the specific acceptor and oversplicing, we asked what the combined effect of mutations to all three silencers would be. In the triple mutant the majority of the transcripts were one of two types shown in Figure 6.9 that splice to both A2 and A3. The top transcript spliced from D1 to A2, then again from D3 to A3. The lower transcript spliced first to A1, then to A2, and finally to A3. The level of A1 usage was only modestly affected by these single or combined mutations. This suggests that HIV splicing happens in a sequential manner, based on which acceptor is transcribed first. A1 is transcribed first and there is the normal amount of splicing to it before A2 or A3 are made. Once A2 is available, D1 splices directly to it. We don't see direct splicing to A3, hinting that it is not available for splicing until after A2. This is strong evidence of sequential cotranscriptional splicing.

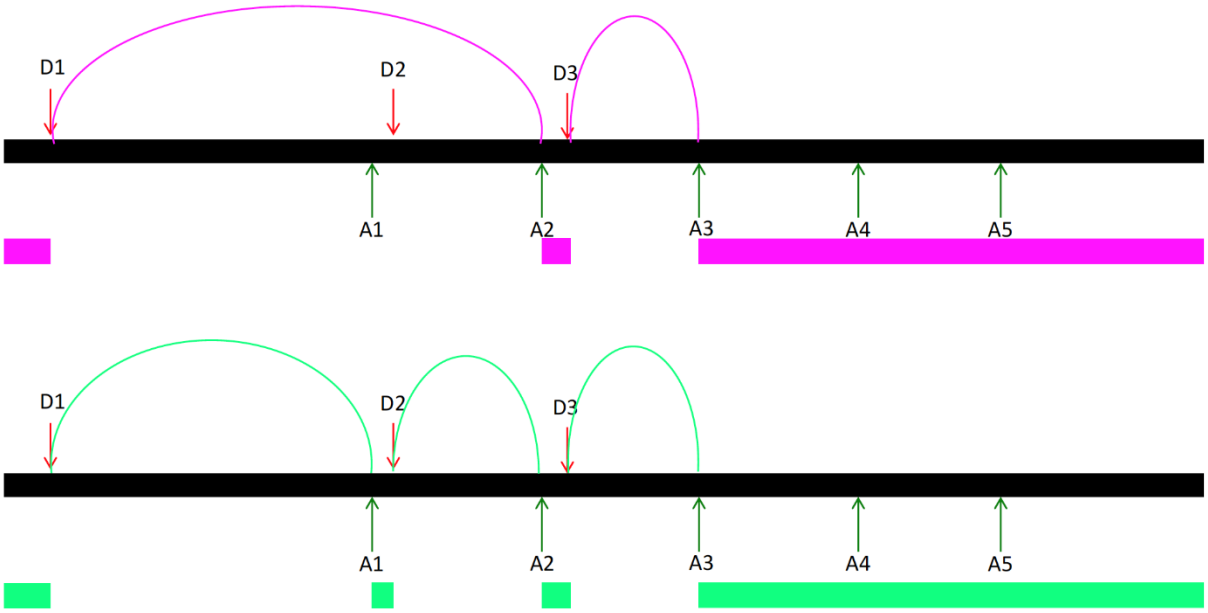

Figure 6.9. Two transcript types make up the majority of transcripts in the ESS2/ESS2p/ESSV combined mutations.

ISS suppresses pseudo-splice acceptors upstream of A7

Only one intronic splicing silencer, ISS, has been described for HIV-1. This ISS is located just upstream of A7 inside the *env* intron and was reported to suppress splicing to A7 though mutation of the element produced only weak splicing enhancement (8). In the current set of experiments, ISS was tested alone and in combination with ESS2, ESS2p, and ESSV. In each case, mutation of ISS caused no changes in acceptor usage or splicing patterns within either the 4 kb or 1.8 kb size classes. In addition, ISS had no effect on the ratio of the size classes or the amount of unspliced transcripts.

Some ISS elements in cellular transcripts have been described that are adjacent to a genuine splice acceptor or donor and they suppress pseudo-donor and pseudo-acceptor sites to correctly define the correct site (9). We looked to see if pseudo-splice acceptor sites next to A7 were used more with the ISS mutation and found they were. Figure 6.10 shows the percentage of splices to the two most used off-target acceptors near A7 (A8345 and A8363). The unmutated control is on the far left and it and all the other mutations to regulators are in yellow. The orange bar on the far right shows the individual ISS mutation and it mis-splices 8-fold more often to a pseudo A7 site. This suggests a function of the ISS element is to define correct A7 usage by suppressing pseudo-splice sites. The shorter orange bar is the ISS mutant combined with the oversplicing mutations. The effect is less when combined with the

oversplicing mutations, which is an interesting example of the importance of context for splicing regulatory elements.

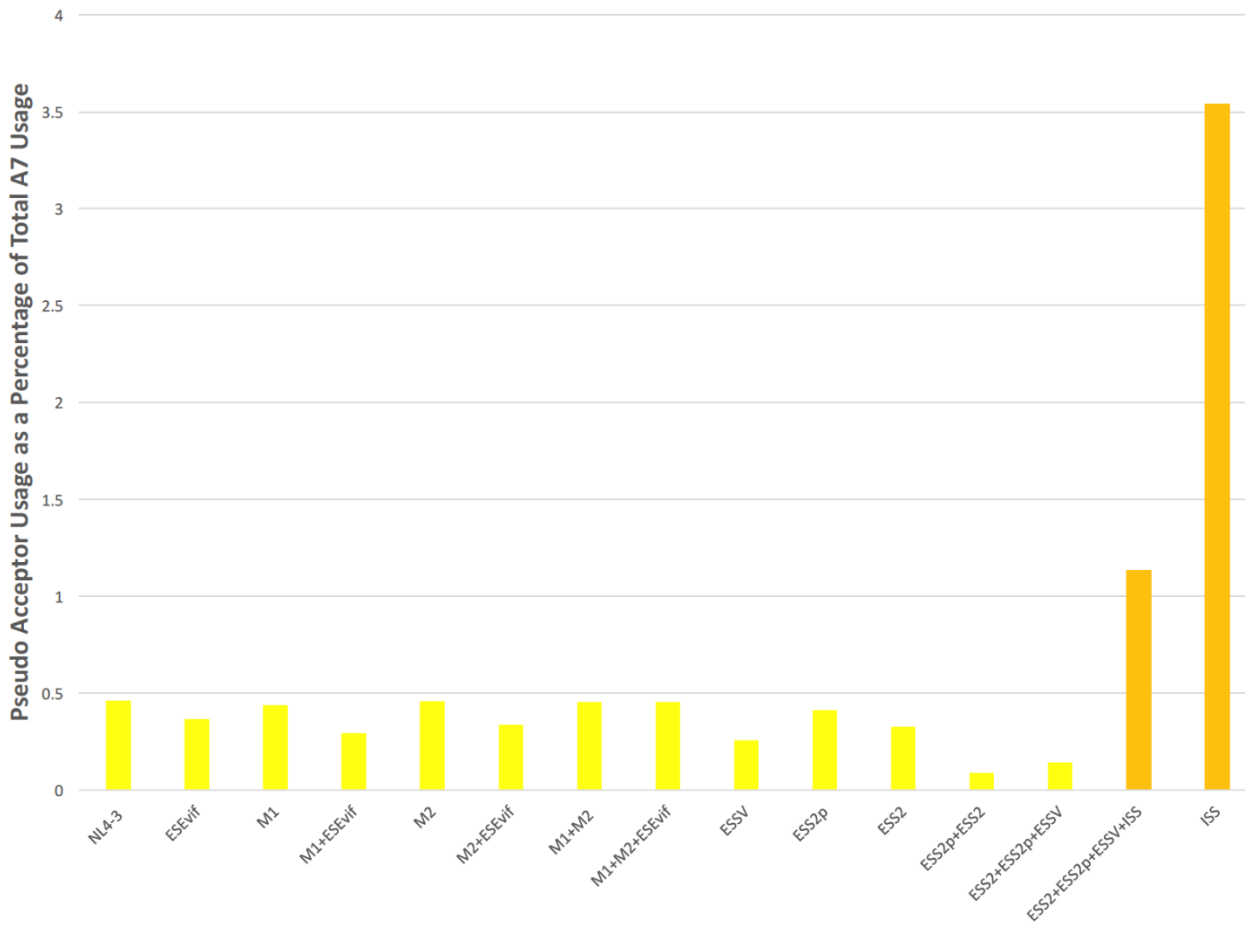

Figure 6.10. ISS suppresses cryptic splicing near A7.
Y axis shows the percentage of splices from D4 to A7 that use the pseudo acceptors A8345 or A8363.

Mutations to the ESS3 stem-loop cause a small increase in splicing from D4 to A7

Some SREs are located in RNA secondary structures, such as the SLSA1 stem loop described in Chapter 2, where the structure is important for SR or hnRNP binding (10). ESS3 is a silencer just downstream of A7. Like ISS and ESE3, two other SREs in the A7 region, ESS3 is highly structured as a stem and loop. This structure has control elements reported to act on D4-A7 splicing (11-15). The conserved nucleotides in the loop bind co-operatively to hnRNP A1 to prevent the D4-A7 splice (16). The loop of the ESS3 stem structure has been extensively studied by the Tolbert lab and the hnRNP A1 binding structure solved (17, 18). Figure 6.11 diagrams the bases of the ESS3 stem-loop structure, which contains hnRNP A1 binding sites UAGU or GAU and without C residues. Mutations were made that

change one or more of the loop bases to a C or alter the stem structure. These mutations were introduced into a full-length NL4-3 construct with a frame shift and a premature stop codon in the Rev ORF to prevent the production of Rev. These infectious clones were then cotransfected with WT NL4-3 into 293T cells, and splicing was analyzed using the splicing assay and the random reverse primer assay. The cotransfected NL4-3 produces WT Rev in sufficient amounts to rescue splicing of the Rev- mutant (data not shown), and also serves as an internal control for normal splicing.

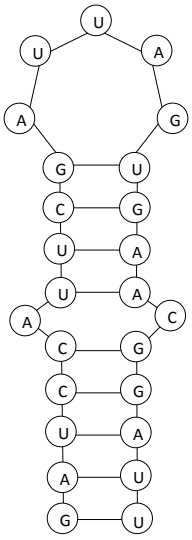

Figure 6.11. The ESS3 stem-loop structure located downstream of splice acceptor A7.
Adapted from Rollins (17).

hnRNP A1 binds cooperatively to prevent D4-A7 splicing, thus mutation of binding site nucleotides to C may enhance splicing. These mutations had minimal or no impact on the spliced transcripts in either of the two size classes (data not shown), but would be expected to alter the balance of the size classes in the mutant compared to NL4-3, increasing the number of transcripts in the 1.8 kb class relative to the 4 kb class. The results are shown in Figure 6.12. In all except mutant #1, the mutant has increased 1.8 kb transcripts (more splicing from D4 to A7). Though consistent, this increase is modest when compared to the Rev-negative parental construct without cotransfected NL4-3, which without Rev has a 1.8 kb to 4 kb ratio of about 31 (data not shown). It has been found that ISS and ESS3 synergistically promote hnRNP A1 multimerization across the extended sequence spanning both silencers (14) so a future experiment will combine mutations to both sites.

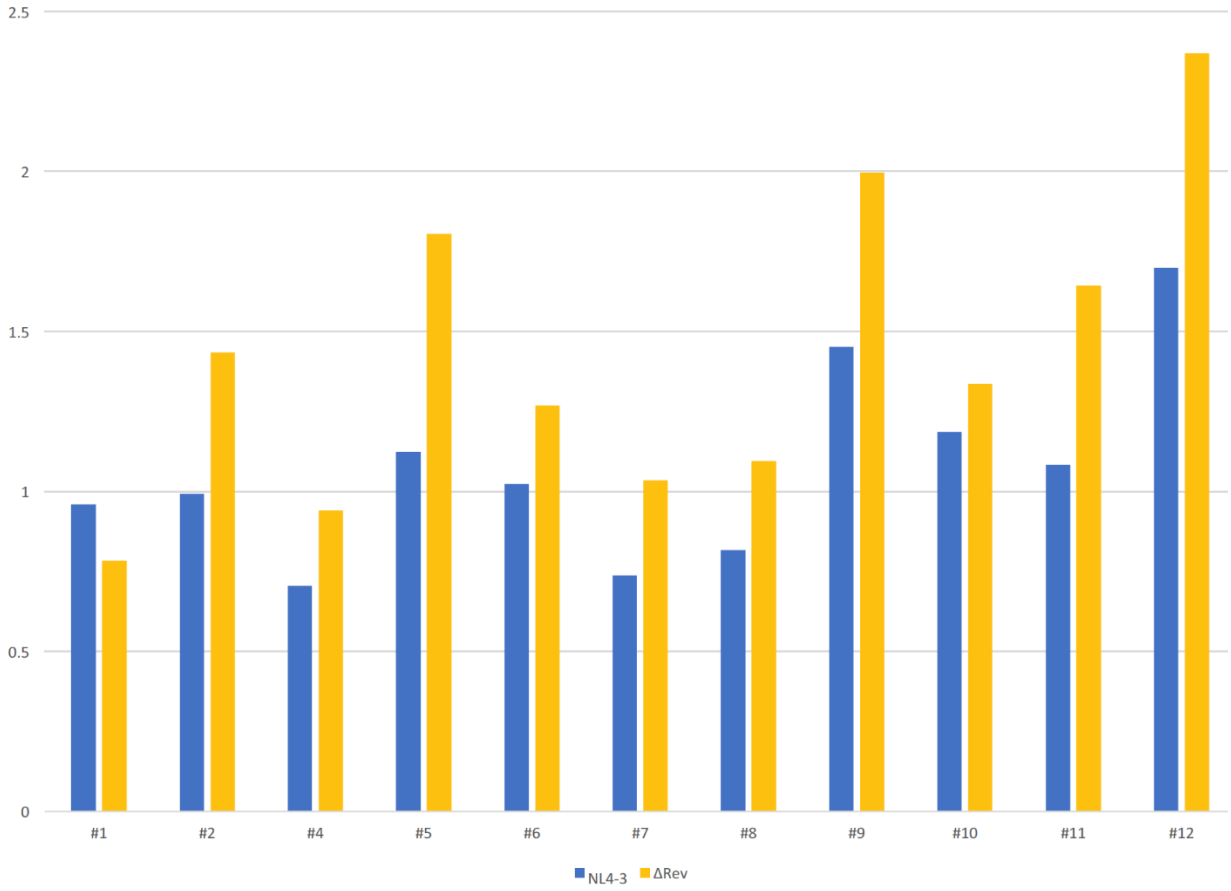

Figure 6.12. Ratio of 1.8 kb transcripts to 4 kb transcripts in each ESS3 mutant and its internal NL4-3 control.

Discussion

Analysis of these SRE mutations demonstrates the utility of the RRPA to measure changes in transcript classes using deep sequencing and in the context of a full-length viral genome. It must be emphasized that these are relative measurements, unlike the splicing assay, as the random reverse primer has preferential but reproducible priming sites.

An additional capability of the random reverse primer assay is its ability to detect aborted transcription. Early in infection, before the viral transcription factor Tat is present in sufficient quantity, many prematurely terminated HIV-1 transcripts are synthesized (19). Though not shown in the data presented here, the random reverse primer assay primes most often in the region just downstream of the transcription start site, suggesting a large number of abortive transcripts even in late infection when Tat levels would be expected to be high. For this reason, the minimum valid read lengths for identifying unspliced transcripts must be long enough to be out of the range of abortive transcripts.

Although all of the mutations tested here have been previously characterized, advances in deep sequencing technology and the development of the two splicing assays presented in this thesis allow in-context measurements of splicing patterns within and between transcript classes. We have found different results from previous studies for ESEVif, ESS2p, ESS2, ISS, and ESS3, and more extensively quantified these known elements. The set of mutations tested here is incomplete and an obvious future direction is to repeat this analysis for the other known SREs.

We have quantified the extent of oversplicing in the ESSV, ESS2 and ESS3p silencer mutations. Oversplicing results in insufficient Gag and consequent replication defects (20). Oversplicing is the most lethal splicing defect we have seen to date, and suggests possible therapeutic applications (21).

Because ISS and ESS3 have previously been characterized as suppressors of splicing from D4 to A7, it is surprising that the effects of mutations were so modest in the context of the full-length virus. This again emphasizes that SREs are context dependent. It's been found that ISS and ESS3 synergistically promote hnRNP A1 multimerization (14) so a future experiment will combine mutations to both sites. It may be that they act in concert, similarly to ESS2 and ESS2p, where the combined effect is much greater than the individual contributions.

Utilizing the recent developments in deep sequencing technology has made it possible to examine splicing regulatory elements in the context of viral infection. These advances in quantification allow the roles of the various SREs to be measured individually and in combination. Further work will more completely characterize SREs that control splicing in HIV-1.

REFERENCES

1. Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. Rna 14:802-813.

2. Graveley BR. 2000. Sorting out the complexity of SR protein functions. Rna 6:1197-1211.

3. Wang Z, Xiao X, Van Nostrand E, Burge CB. 2006. General and specific functions of exonic splicing silencers in splicing control. Mol Cell 23:61-70.

4. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A 108:20166-20171.

5. Exline CM, Feng Z, Stoltzfus CM. 2008. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol 82:3921-3931.

6. Berget SM. 1995. Exon recognition in vertebrate splicing. J Biol Chem 270:2411-2414.

7. Jacquenet S, Mereau A, Bilodeau PS, Damier L, Stoltzfus CM, Branlant C. 2001. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. J Biol Chem 276:40464-40475.

8. Tange TO, Damgaard CK, Guth S, Valcarcel J, Kjems J. 2001. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. Embo j 20:5748-5758.

9. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. 2004. Systematic identification and analysis of exonic splicing silencers. Cell 119:831-845.

10. Buratti E, Baralle FE. 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol 24:10505-10514.

11. Mayeda A, Screaton GR, Chandler SD, Fu XD, Krainer AR. 1999. Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol Cell Biol 19:1853-1863.

12. Tange TO, Kjems J. 2001. SF2/ASF binds to a splicing enhancer in the third HIV-1 tat exon and stimulates U2AF binding independently of the RS domain. J Mol Biol 312:649-662.

13. Zhu J, Mayeda A, Krainer AR. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell 8:1351-1361.

14. Damgaard CK, Tange TO, Kjems J. 2002. hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. Rna 8:1401-1415.

15. Asai K, Platt C, Cochrane A. 2003. Control of HIV-1 env RNA splicing and transport: investigating the role of hnRNP A1 in exon splicing silencer (ESS3a) function. Virology 314:229-242.

16. Marchand V, Mereau A, Jacquenet S, Thomas D, Mougin A, Gattoni R, Stevenin J, Branlant C. 2002. A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding. J Mol Biol 323:629-652.

17. Rollins C, Levengood JD, Rife BD, Salemi M, Tolbert BS. 2014. Thermodynamic and phylogenetic insights into hnRNP A1 recognition of the HIV-1 exon splicing silencer 3 element. Biochemistry 53:2172-2184.

18. Levengood JD, Rollins C, Mishler CH, Johnson CA, Miner G, Rajan P, Znosko BM, Tolbert BS. 2012. Solution structure of the HIV-1 exon splicing silencer 3. J Mol Biol 415:680-698.

19. Karn J, Stoltzfus CM. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med 2:a006916.

20. Mandal D, Feng Z, Stoltzfus CM. 2008. Gag-processing defect of human immunodeficiency virus type 1 integrase E246 and G247 mutants is caused by activation of an overlapping 5' splice site. J Virol 82:1600-1604.

21. Mandal D, Feng Z, Stoltzfus CM. 2010. Excessive RNA splicing and inhibition of HIV-1 replication induced by modified U1 small nuclear RNAs. J Virol 84:12790-12800.

CHAPTER 7: CONCLUSIONS AND FUTURE DIRECTIONS

One important goal for studying HIV-1 is to identify therapeutic targets, and the ultimate purpose of studying HIV-1 splicing is to determine if and how splicing can be targeted to produce a useful therapeutic effect. The ability to quantify splicing is indispensable to this purpose and the work done in this thesis identifies aspects of splicing that are and are not likely to be useful targets.

There is a general assumption in the literature that "balanced" use of the HIV-1 acceptors and thereby a carefully balanced control of protein products is critical to viral replication (1, 2). Tat and Rev are held in particular reverence in this regard (3-5). Without sufficient Tat viral transcription aborts prematurely. Without sufficient Rev, no unspliced transcripts can leave the nucleus. Low levels of either of these proteins would seem to force the cell into a functional, and perhaps actual, state of latency. If some splice acceptors could be made more or less efficient, the balance would be upset with a high fitness cost. The splicing data for the transmitted/founder viruses (described in Chapter 2) challenge the idea of balanced splicing. The data shown in Figure 2.6 show a wide range of both *tat* and *rev* transcript levels in cells expressing the transmitted/founder viruses. This suggests a new paradigm – rather than carefully balanced splicing, some minimum threshold of each protein is all that is required for viral replication. Targeting the relative proportional use of the different acceptors is not likely to have therapeutic value.

Another potential balancing target is the ratio between the two spliced transcript size classes. This ratio is determined by the degree of splicing from D4 to A7, once splicing from D1 has already occurred. In my collective data thus far, the ratio of fully spliced to partially spliced transcripts varied from about 1:2 and 2:1, with this range falling within assay variability as seen in replicates under the same experimental conditions. This is especially evident in the mutations made to ESS3, seen in Figure 6.11. Each of the eleven ESS3 mutations was co-transfected with NL4-3 to provide Rev. The ratio of 1.8 kb to 4 kb RNAs in the NL4-3 controls should be about the same for the different transfections but this ratio varied appreciably across the samples. Although all but mutation #1 show an increase in this ratio

compared to their control, these increases are modest compared to the changes seen in other SRE mutations analyzed in Chapter 6. Mutations to ESS3, ISS and to the GAR ESE region collectively had modest but consistent effects. Thus far I have not seen any mutations that skew D4 to A7 splicing much beyond the range of assay variability, making the D4/A7 splice an unlikely therapeutic target unless a more dramatic regulatory site can be identified. The one factor that did make a big difference in the 1.8 to 4 kb RNA ratio was deleting Rev, which skewed the totally spliced to partially spliced ratio as high as 30 to 1. This suggests there is a race to see if Rev can bind the RRE once it is transcribed before A7 is also transcribed and the RRE gets spliced out. The RRE and A7 could be close enough to sterically exclude either Rev binding or splice factor binding to A7 SREs. Viewed in the context of cellular RNA splicing, the splice from D4 to A7 joins two constitutive exons. The exon containing D4 is always included and D4 is a strong splice donor (6). The layers of control seen in alternative exon splicing may not be required. Conceivably, splicing from D4 to A7 would always occur unless Rev intervenes. The completely spliced/partially spliced balance may be a cycle controlled by Rev. Too much Rev reduces completely spliced transcripts, which in turn reduces the amount of Rev. Less Rev increases completely spliced transcripts, raising Rev levels. The oscillating levels of spliced to unspliced transcripts suggests a continual feedback system operating within loose parameters, regardless of the efficiency of D4-A7 splicing. The role of Rev in splicing is an area for continued study.

Another therapeutic target that appears unlikely to succeed is activation of cryptic splice sites to make aberrant proteins. It has been reported and often repeated in the literature that perturbation of canonical splice sites activates cryptic splice sites (2). My data suggests that cryptic splice site activation is not common. It was not seen in the transmitted/founder viruses, and not it seen generally in the global mutations described in Chapter 3. Some strains do exhibit altered cryptic splicing (7-10) that may give rise to new or combined ORFs, and production of the chimeric protein products has been documented. To date, no useful functionality has been ascribed to any of these hybrid proteins. The splice sites that created them are not generally conserved in HIV-1 strains, but are instead the result of mutations that create new splicing elements rather than arising from suppression of the canonical sites to activate nearby cryptic sites. The hybrid proteins produced by these cryptic sites did not prevent the viruses from replicating.

There is one exception to the cryptic splice site issue. Mutations downstream of D1 (Chapter 3) did cause cryptic splicing. New acceptors and donors and small exons were introduced. Even though this aberrant splicing was only seen in a subset of the transcripts, viruses with these mutations were unable to replicate effectively. Reversion mutants abolished the cryptic sites. While we did not test specifically for this effect, the fitness cost suggests that the truncated fragments of Gag protein encoded by these aberrant transcripts have a dominant negative effect. This is an area for further research. Little is known about regulation of D1, the major splice donor. Studies of D1 are complicated by the density of other nearby sequence and structural elements. The area upstream of D1 and splicing at D1 can be studied using a genetic screen like the one described in Chapter 5. Changes in the D1 area do in fact give rise to cryptic splicing (2) and it will be interesting to probe the mechanism of D1 activation.

Suppression of splicing is an attractive target. HIV-1 is unique in that the majority of transcripts retain introns, and viral infection has been shown to cause intron retention in transcripts of cellular genes (11), hinting that a viral protein is the cause. This suppression of splicing is unique to HIV-1 and thus a potential therapeutic target. Suppression manifests in two different mechanisms.

First, mutations to exonic splicing silencers, explored in Chapter 6, activated the upstream proximal acceptor site (see Figures 6.4 through 6.6). They also lifted suppression of splicing (i.e. activated splicing) at the major 5' splice donor D1 to cause oversplicing, but did not activate splice donor D4. Additional disruptions of silencing were discovered by synonymous mutagenesis of the HIV-1 genome as described in Chapter 3. Loss of suppression caused oversplicing and a concomitant fitness cost due to insufficient Gag particle production.

Second, I have observed that the absence of Rev decreases (but does not eliminate) unspliced or partially spliced transcripts. I looked for but did not find evidence of transcripts that fail to use D1 (i.e. retain the *gag/pro/pol* intron) but use D4 to splice out the *env* intron. Additionally, I found no transcripts that used D2 or D3 without first using D1, making D1 suppression a marker for suppression of all splice donors. This suggests unspliced transcripts are suppressed early and completely from all splicing, long before the other downstream splice donors or the Rev Response Element (RRE) are transcribed. The mechanisms of suppression are another area of ongoing research.

One possible mechanism of splicing suppression could be the secondary structure of the D1 region. A previous study found that this region can fold into two structures, depending on the number of G's included at the variable transcription start site (12). This small change from 1 to 3 G's is sufficient to alter the pre-mRNA secondary structure so that D1 is either presented on the loop of a stem-loop or is part of a double stranded RNA region. This changed access to D1 may be a factor that suppresses splicing early on, before any acceptor sites are transcribed. Blocked access could explain splicing suppression at D1 but not the lack of splicing from D4 to A7, which may also be prevented by participation in as yet unknown secondary structures. Further work will test splicing in these alternate constructs.

Too little is known about the specificities of SR and hnRNP proteins and their roles in HIV-1 splicing to speculate about them being possible therapeutic targets. These proteins are problematic because they are redundant and bind promiscuously across the entire genome. There are multiple motifs reported for binding sites of these proteins, some more convincing than others. I am skeptical of reports of HIV-1 enhancers found by computer algorithms based on binding site consensus sequences. Feeding any HIV-1 sequence that might be associated with splicing, or even randomly generated sequences, into ESEfinder (13) finds that about half of the bases are contained in an enhancer – which cannot be correct. More and better data are needed to find the actual HIV-1 binding sites of SR and hnRNP proteins and their effects, and more studies like those described in Chapter 3 are needed.

The work done in Chapter 3 raises more questions than it answers and calls into question previous models of splicing factor binding. Previous studies have used SELEX to fish out the highest affinity binding sequences out of a pool of random oligos (14). This method has produced degenerate consensus sequences, and different studies report different results. The consensus sequences found using CLIP-seq are not the same (though similar) as previously reported (15, 16). siRNA knock down of cellular splicing factors produced mild changes in splicing phenotypes, but the changes caused by hnRNP H1 knock down are not the same as the changes found by mutation of hnRNP H1 binding sites. This suggests that a partial knock down is not sufficient to cause severe splicing defects, either due to redundancy of splicing factors or a low factor concentration requirement.

We may be looking at the wrong splicing factors. siRNA knockdown was done to target the factors most often reported to play a role in HIV-1 splicing; however, the methods used to determine relevant factors are questionable. In several cases, biotinylated oligos containing splicing regulatory elements were soaked in cellular extracts and they bound hnRNP A/B proteins (17-19), but CLIP-seq data show hnRNP A/B proteins bind promiscuously across the HIV-1 genome (Figure 4.2).

The high-frequency binding sites of hnRNP A/B (Figure 4.2) do not correspond to any of the HIV-1 splicing regulatory elements claimed to bind them. The CLIP-seq experiments in Chapter 4 have negative controls but not a positive control. For this set of experiments negative controls would be the anti-HA Ab alone (which shouldn't bind RNA in the absence of the HA-tagged hnRNP proteins) and the HA-tagged hnRNP proteins without the HA specific Ab (which would bind the RNA but should not be precipitated). Without both the binding protein and the antibody there should be only a very low level of "noise" reads, like that seen with hnRNP K, and therefore we consider the promiscuous binding of hnRNP A/B valid. This promiscuous binding supports previous findings that hnRNP A/B binds at high affinity sites and multimerizes along the RNA (20-22). Nevertheless, it would be reassuring to find that the binding sites found in CLIP-seq actually correspond to known cellular splicing regulatory elements, which we do not see with hnRNP A/B. This suggests that the hnRNP A/B proteins are not the actual factors involved in HIV-1 splicing, but that we have been led astray by these highly promiscuous and abundant splicing proteins. Work is currently in progress to identify other splicing factors by their splicing phenotype rather than binding of a regulatory sequence. In future CLIP-seq experiments, we would also hope to validate the connection between binding sites and splicing phenotype, as we did with hnRNP H1, but were unable to do with hnRNP A/B.

The model of a protein factor binding a specified nucleotide consensus sequence is common in cell biology but is too simplistic for HIV-1 splicing regulation. The hnRNP H1 binding sites in HIV-1 are not a simple consensus sequence of poly-Gs, but a group of poly-Gs. Scanning the HIV-1 genome reveals multiple matches to the TGGGG hnRNP H1 consensus sequence, but only the sites with multiple proximate poly-G tracts make an hnRNP H1 binding site. Additionally, there are no grouped hnRNP H1 consensus sites outside those found by the CLIP-seq data. Mutation of any one of the hnRNP H1 binding sites was insufficient to produce the loss of splicing seen to A1 when all binding sites were mutated. This

hints that there is an additional layer of interaction between the binding sites, as has been shown to occur with hnRNP A1.

This combined data suggests a new model (Figure 7.1) in which binding sites are composed of multiple single sites that collectively make a high probability (high affinity) area for binding of the hnRNP protein to the RNA, followed by cooperative binding between the hnRNP proteins. In Figure 7.1, using hnRNP H1 as an example, the consensus sequences for hnRNP H1 binding sites shown at the top illustrate the multiple poly-G consensus sites grouped together. Site 5 is boxed in pink, and the second row illustrates hypothesized interactions between the hnRNP H1 proteins at site 5. The eight hnRNP H1 binding sites are shown in red on the third line of Figure 7.1, and in this proposed model the individual sites then interact with each other, looping out the intervening sequences. Additional work on binding proteins and their binding sites is underway.

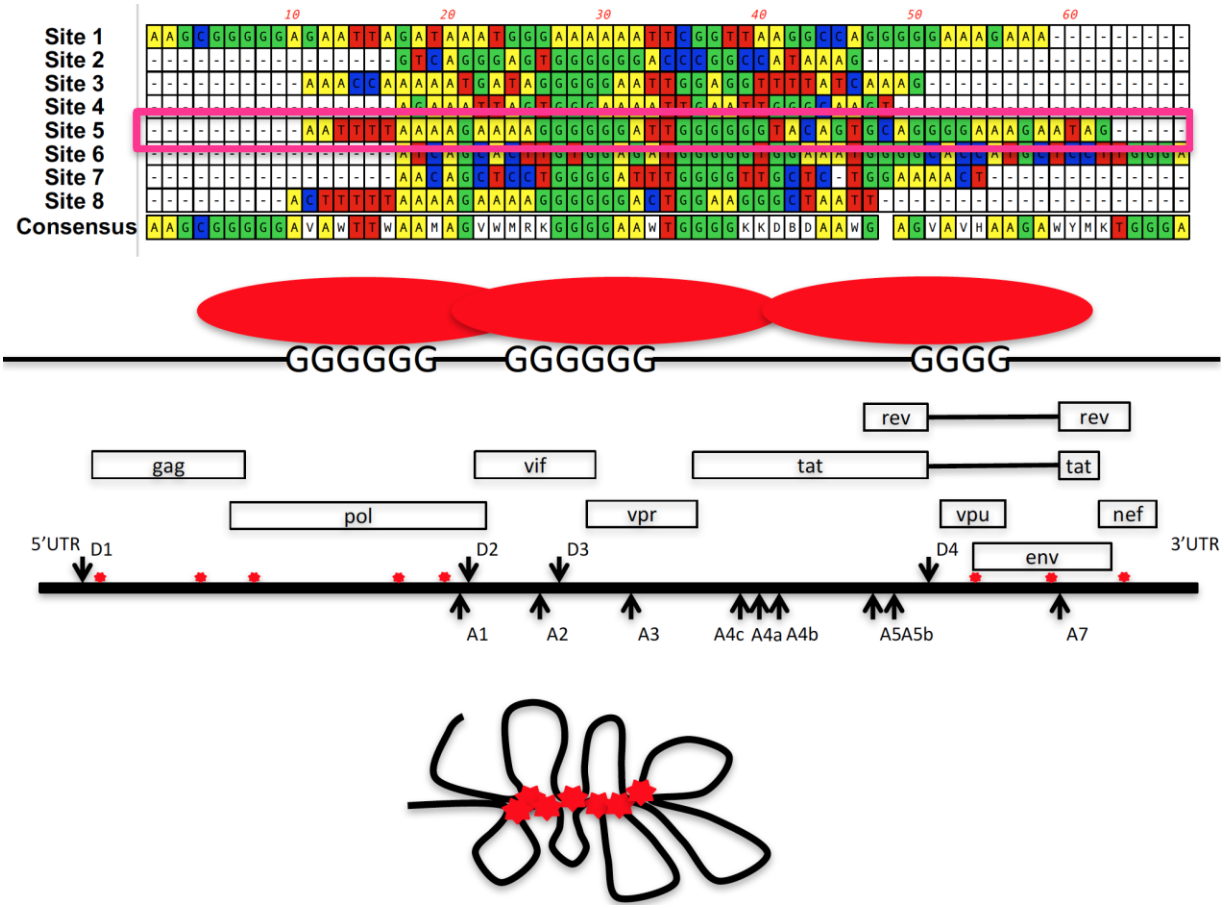

Figure 7.1. A Proposed Model for hnRNP Binding.
Using hnRNP H1 as an example, row 1 shows multiple proximate consensus sites in each binding site. Row 2 illustrates hypothesized interactions between hnRNP proteins to stabilize binding and define the binding site. Row 3 shows the binding sites spread across the genome and row 4 the hypothesized tertiary interactions between the binding sites.

Splicing is essentially the same in infected T cell lines (CEMx174, Jurkat, H1) and transfected 293 T cells. I also found that splicing was the same in T cells and dendritic cells (part of the work done in Miller et al., see appendix) and this validates the use of cell lines and transfection to study splicing. The next step is to quantify splicing in macrophages. Macrophages are reported to have low levels of Tat and slow viral growth, making them long-term viral reservoirs. Splicing patterns in macrophages are also reported to be different following HIV-1 infection (23). Since the virus can adapt to growth in macrophages late in infection it will be important to determine if adjustments in the splicing program are part of this adaptive process.

In Chapter 6, ISS was found to block mis-splicing to pseudo-acceptors upstream of A7. There are other pseudo-splice sites, several of which are detected by the splicing assay. Some come in pairs and can make a small exon but most work together with a canonical acceptor or donor. It's likely that other intronic splicing silencers exist to keep these in check. Elements that control an infrequently used donor site downstream of D2 were recently characterized (19). Elements that suppress other pseudo-sites remain to be found.

Finally, the purpose of the small exons created by donors D2 and D3 remains a mystery. They are highly conserved yet have no clear function. It has been suggested that they are necessary for exon definition to ensure adequate Vif and Vpr expression (24-26), but the data from the transmitted/founder viruses showed little change in *vpr* even when the second small exon was knocked down (Figure 2.6). Also, viruses with low levels of Vif due to reduced splicing to A1 show no fitness costs unless APOBECs are over-expressed (26). This suggests that the small exons aren't critical for Vif and Vpr. They didn't change protein production or transcript stability (24, 27). The reason for their high degree of conservation remains unsolved.

Since the discovery of splicing in adenovirus in 1977, the mechanisms of splicing in cellular genes has been well studied. Although HIV-1 uses the cellular splicing machinery and proteins, it must suppress and control alternative splicing to appropriately express all the viral transcripts from a single 9 kilobase transcript. There are evidently several different regulatory mechanisms at work, some coming into focus while others remain mysterious. Differences in cellular and HIV-1 splicing suggest several areas where the virus may be vulnerable. The advent of successful splicing therapy to treat genetic diseases is creating new technologies that may someday translate knowledge about HIV-1 splicing into clinical usefulness. To understand splicing, it must be quantifiable, and this is made possible by the rapid expansion of sequencing technology and bioinformatics tools. These developing technologies provide opportunities to probe the mechanisms of HIV-1 splicing regulation.

REFERENCES

1. Stoltzfus CM, Madsen JM. 2006. Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. Curr HIV Res 4:43-55.

2. Purcell DF, Martin MA. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67:6365-6378.

3. Mbonye U, Karn J. 2017. The Molecular Basis for Human Immunodeficiency Virus Latency. Annu Rev Virol doi:10.1146/annurev-virology-101416-041646.

4. Pai A, Weinberger LS. 2017. Viral Master Circuits: From Discovery to New Therapy Targets. Annu Rev Virol doi:10.1146/annurev-virology-110615-035606.

5. Malim MH, Cullen BR. 1991. HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. Cell 65:241-248.

6. O'Reilly MM, McNally MT, Beemon KL. 1995. Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. Virology 213:373-385.

7. Benko DM, Schwartz S, Pavlakis GN, Felber BK. 1990. A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. J Virol 64:2505-2518.

8. Wentz MP, Moore BE, Cloyd MW, Berget SM, Donehower LA. 1997. A naturally arising mutation of a potential silencer of exon splicing in human immunodeficiency virus type 1 induces dominant aberrant splicing and arrests virus production. J Virol 71:8542-8551.

9. Caputi M, Zahler AM. 2002. SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. Embo j 21:845-855.

10. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res 40:10345-10355.

11. Sherrill-Mix S, Ocwieja KE, Bushman FD. 2015. Gene activity in primary T cells infected with HIV89.6: intron retention and induction of genomic repeats. Retrovirology 12:79.

12. Kharytonchyk S, Monti S, Smaldino PJ, Van V, Bolden NC, Brown JD, Russo E, Swanson C, Shuey A, Telesnitsky A, Summers MF. 2016. Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. Proc Natl Acad Sci U S A 113:13378-13383.

13. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31:3568-3571.

14. Tian H, Kole R. 1995. Selection of novel exon recognition elements from a pool of random sequences. Mol Cell Biol 15:6291-6298.

15. Hallay H, Locker N, Ayadi L, Ropers D, Guittet E, Branlant C. 2006. Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site. J Biol Chem 281:37159-37174.

16. Schaub MC, Lopez SR, Caputi M. 2007. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. J Biol Chem 282:13617-13626.

17. Widera M, Erkelenz S, Hillebrand F, Krikoni A, Widera D, Kaisers W, Deenen R, Gombert M, Dellen R, Pfeiffer T, Kaltschmidt B, Munk C, Bosch V, Kohrer K, Schaal H. 2013. An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA. J Virol 87:2707-2720.

18. Widera M, Hillebrand F, Erkelenz S, Vasudevan AA, Munk C, Schaal H. 2014. A functional conserved intronic G run in HIV-1 intron 3 is critical to counteract APOBEC3G-mediated host restriction. Retrovirology 11:72.

19. Brillen AL, Walotka L, Hillebrand F, Muller L, Widera M, Theiss S, Schaal H. 2017. Analysis of competing HIV-1 splice donor sites uncovers a tight cluster of splicing regulatory elements within exon 2/2b. J Virol doi:10.1128/jvi.00389-17.

20. Zhu J, Mayeda A, Krainer AR. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell 8:1351-1361.

21. Damgaard CK, Tange TO, Kjems J. 2002. hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. Rna 8:1401-1415.

22. Marchand V, Mereau A, Jacquenet S, Thomas D, Mougin A, Gattoni R, Stevenin J, Branlant C. 2002. A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding. J Mol Biol 323:629-652.

23. Dowling D, Nasr-Esfahani S, Tan CH, O'Brien K, Howard JL, Jans DA, Purcell DF, Stoltzfus CM, Sonza S. 2008. HIV-1 infection induces changes in expression of cellular splicing factors that regulate alternative viral splicing and virus production in macrophages. Retrovirology 5:18.

24. Madsen JM, Stoltzfus CM. 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.

25. Exline CM, Feng Z, Stoltzfus CM. 2008. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol 82:3921-3931.

26. Mandal D, Exline CM, Feng Z, Stoltzfus CM. 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. J Virol 83:6067-6078.

27. Anderson JL, Johnson AT, Howard JL, Purcell DF. 2007. Both linear and discontinuous ribosome scanning are used for translation initiation from bicistronic human immunodeficiency virus type 1 env mRNAs. J Virol 81:4664-4676.

APPENDIX: VIRION ASSOCIATED VPR ALLEVIATES A POST-INTEGRATION BLOCK TO HIV-1 INFECTION OF DENDRITIC CELLS

Overview

Viral protein R (Vpr) is an HIV-1 accessory protein whose function remains poorly understood. In this report, we sought to determine the requirement of Vpr in facilitating HIV-1 infection of monocyte-derived dendritic cells (MDDCs), one of the first cells to encounter virus in the peripheral mucosal tissues. We characterize in this report a significant restriction to Vpr-deficient virus replication and spread in MDDCs alone and in cell-to-cell spread in MDDC – $CD4^+$ T cell co-cultures. This restriction to HIV-1 replication in MDDCs was observed in a single round of virus replication and was rescued by expression of Vpr *in trans* in the incoming virion. Interestingly, infections of MDDCs with viruses that encode Vpr mutants either unable to interact with DCAF1/DDB1 E3 ubiquitin ligase complex or a host factor hypothesized to be targeted for degradation by Vpr also displayed a significant replication defect. While the extent of proviral integration in HIV-1 infected MDDCs was unaffected by the absence of Vpr, transcriptional activity of the viral LTR from Vpr-deficient proviruses was significantly reduced. Together, these results characterize a novel post-integration restriction to HIV-1 replication in MDDCs and that Vpr interaction with the DCAF1/DDB1 E3 ubiquitin ligase complex and the yet-to-be identified host factor might alleviate this restriction by inducing transcription from the viral LTR. Taken together, these findings identify a robust *in vitro* cell culture system that is amenable to addressing mechanisms underlying Vpr-mediated enhancement to HIV-1 replication.

Importance

Despite decades of work, function of the HIV-1 protein Vpr remains poorly understood, primarily due to lack of an *in vitro* cell culture system that demonstrates a deficit in replication upon infection with viruses in the absence of Vpr. In this report, we describe a novel cell infection system that utilizes primary human dendritic cells, which display a robust decrease in viral replication upon infection with Vpr-deficient HIV-1. We show that this replication difference occurs in a single round of infection and is due to decreased transcriptional output from the integrated viral genome. Viral transcription could be rescued by virion-associated Vpr. Using mutational analysis we show that domains of Vpr involved in binding to the DCAF1/DDB1/E3 ubiquitin ligase complex and prevention of cell cycle progression into mitosis are

required for LTR-mediated viral expression, suggesting that the evolutionarily conserved G2 cell cycle arrest function of Vpr is essential for HIV-1 replication.

Introduction

HIV-1 encodes a number of proteins that allow for entry and replication in human cells. In addition to the structural or enzymatic proteins that have well defined functions in the replication cycle, there are also a number of small, accessory proteins. Accessory proteins encoded by HIV-1 are not always necessary for replication *in vitro*, but are absolutely essential for replication *in vivo* (1). These proteins serve to counteract host restriction factors that would normally limit HIV-1 infection (1, 2). Of the accessory proteins encoded by HIV-1, Vpr is the only one whose function remains relatively unclear.

Vpr is a small, 96 amino acid, 14 kDa protein that is packaged into the budding virion through associations with the p6 region of Gag (3–10). This association allows Vpr to be present in the cell at a relatively high quantity (~200-300 molecules/virion) upon initial infection (11). Previous studies have extensively characterized the outcome of Vpr expression in various cell types. In cycling cells, Vpr expression results in G2/M cell cycle arrest which culminates in induction of apoptosis (12–14). It is well established that Vpr-mediated G2/M cell cycle arrest is mediated though its association with the Cul4A/DCAF/DDB1 E3 ($CRL4^{DCAF1}$) ubiquitin ligase complex (15–17). In addition, HIV-1 Vpr has been shown to recruit and degrade a number of DNA-damage response (DDR) proteins, including the SLX4-SLX1/MUS81-EME1 structure-specific endonuclease complex (SLX4com), Uracil DNA glycosylase 2 (UNG2), and helicase-like transcription factor (HLTF) (18–21) via the $CRL4^{DCAF1}$ complex resulting in G2/M cell cycle arrest though it still remains unclear what role this process plays during HIV-1 infection.

Though a number of previous studies have examined the requirement of Vpr on HIV-1 replication in various cell types, including primary $CD4^+$ T cells and monocyte-derived macrophages (MDMs), differences in virus replication have not been consistently observed (18, 20, 22–26). Vpr expression is dispensable for infection in activated $CD4^+$ T cells *in vitro* (22–25, 27), presumably due to the well characterized cytostatic and cytopathic functions of Vpr in cycling cells (12). In contrast, recent studies in MDMs suggest that Vpr is necessary for HIV-1 envelope (Env) expression, and the purported consequence of infection of MDMs with Vpr-deficient viruses was reported to be decreased viral production and reduced cell-to-cell spread to $CD4^+$ T cells (22, 28). Notably, there has been considerable

heterogeneity in replication differences between wild type and Vpr-deficient viruses and host responses to virus infection in MDMs, presumably due to donor and experimental variability between studies (12, 29, 30). Additionally, it has also been reported that Vpr expression in macrophages can both inhibit or induce type I interferon (IFN) responses (18, 28, 31–34).

Dendritic cells (DCs) are sentinel cells that bridge innate and adaptive immunity (35). They actively patrol peripheral tissues, including mucosal sites of HIV-1 transmission, in search of foreign pathogens. Because of this, MDDCs are among the first cells to interact with HIV-1 upon sexual transmission of the virus (36–40). While MDDCs are less susceptible to infection than activated $CD4^+$ T cells and macrophages, they are still able to be infected *ex vivo* at a low but consistent level (41–44). In contrast to work with MDMs and $CD4^+$ T cells, there have been isolated descriptions of effects of Vpr on HIV-1 replicative capacity in MDDCs (41, 44), with no consensus on the mechanisms accounting for Vpr-mediated enhancement of virus replication. In this study, we use MDDCs as a model system to investigate the role of Vpr during infection. We find a robust replication defect of Vpr-deficient HIV-1 in MDDCs and, contrary to previous studies (44), the replication defect was not due to decreased Env expression in Vpr-deficient HIV-1 infected cells. Rather, the block to ΔVpr virus infection was at the step of viral transcription and could be rescued by addition of Vpr *in trans* into the virion in a single round infection analysis. We found that mutations, Vpr-Q65R and Vpr-H71R, which ablate association of Vpr with the $CRL4^{DCAF1}$, or Vpr-R90K which does not induce G2 cell cycle arrest (17, 45–50), displayed similar decreases in replication and viral transcription in single round of infection analysis. Together these data shows a novel post integration block to HIV-1 replication in MDDCs at the point of viral transcription that is alleviated by virion-associated Vpr.

Results

Vpr-deficient viruses display a replication defect in DCs

HIV-1 replication in MDDCs is restricted at the reverse transcription step by SAMHD1 that controls the size of the cytosolic dNTP pools (51, 52). Despite the presence of SAMHD1, MDDCs remain susceptible to HIV-1 infection *in vitro* at a low but measurable level (44, 53–55). We infected MDDCs with replication competent wild type (WT) or Vpr-deficient (ΔVpr) CCR5-tropic Lai-YU2 and harvested cells for intracellular $p24^{Gag}$ expression by flow cytometry analysis 3 days post infection. Input for these infections

was normalized based on infectious titer of the viruses on TZM-bl cells. As expected, $CD11c^+$ $DC\text{-}SIGN^+$ MDDCs were susceptible to viral infection, albeit to low levels (Figure 8.1A and B). Interestingly, Lai-YU2/ΔVpr failed to establish a robust infection in MDDCs (Figure 8.1A and B), and there was a reproducible 3-5 fold decrease in percentage of $p24^{Gag+}$ cells in ΔVpr virus infections as compared to WT virus infections (Figure 8.1B). To determine the functional consequences of Vpr-deficiency on virus spread, DCs and PHA/IL-2-activated $CD4^+$ T cells were infected with infectious viruses (MOI = 1) and cell-free culture supernatants were harvested every 3 days and analyzed for $p24^{Gag}$ content by an ELISA. While there was some donor variability, Lai-YU2/ΔVpr infection of MDDCs derived from 3 independent donors consistently resulted in significantly lower levels of replication than wild type Lai-YU2 infection (Figure 8.1C). In contrast to the substantial attenuation of virus spread in Lai-YU2/ΔVpr infected DCs, both viruses replicated to a similar extent in activated $CD4^+$ T cells (Figure 8.1E), in agreement with previously published studies (22–25, 27). These results suggest that Vpr plays an important role in facilitating HIV-1 infection of DCs.

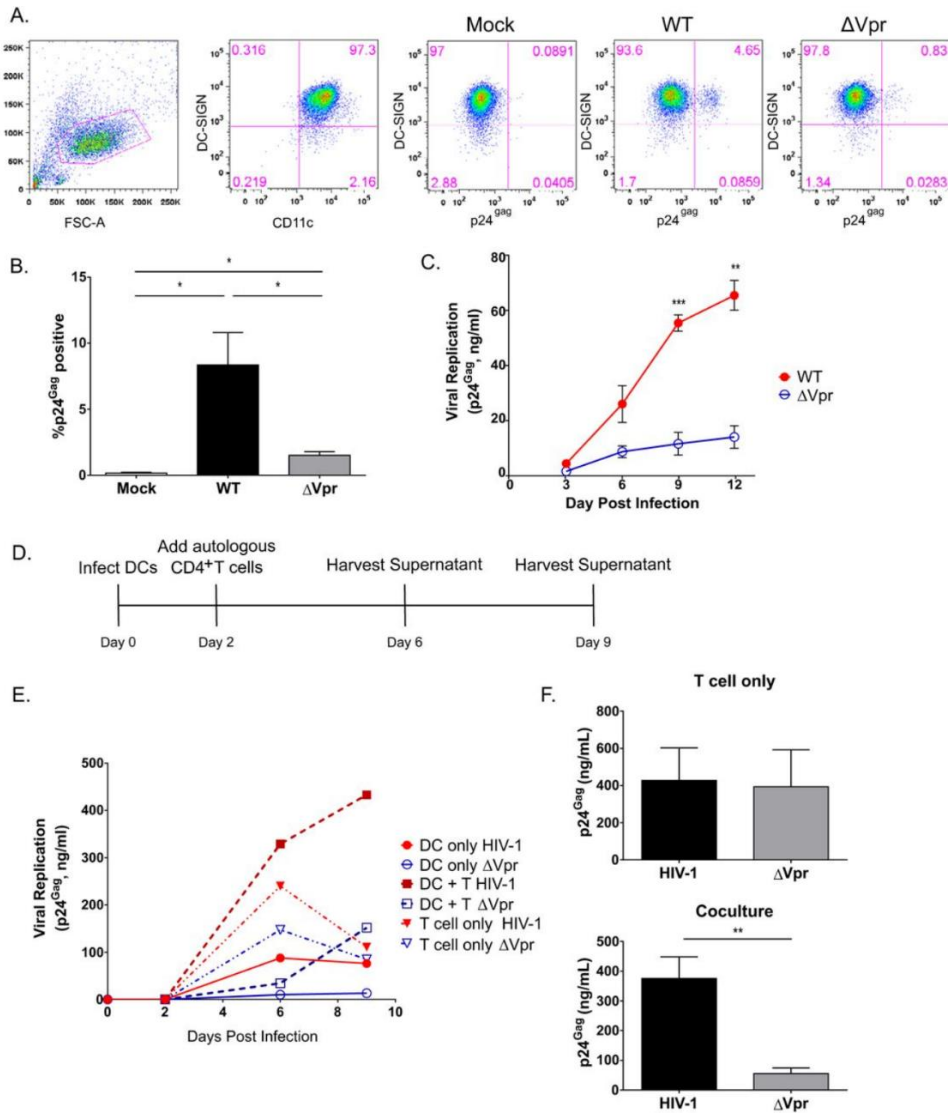

Figure 8.1. Infection with Vpr-deficient HIV-1 results in attenuated virus replication in MDDCs and MDDC-T co-cultures.

(A) FACS profiles of mock infected MDDCs or MDDCs infected with Lai-YU2 or Lai-YU2/ΔVpr (MOI =1) at day 3 post infection. Cells were stained for CD11c, DC-SIGN and p24Gag. From left to right, plots shown depict the gating strategy for the flow cytometry analysis and include plots of forward scatter/side scatter to exclude cellular debris, anti-CD11c/anti-DCSIGN staining to identify MDDC population, and DC-SIGN/p24Gag staining to identify productively infected MDDCs in mock infected, or WT (Lai-YU2) and ΔVpr infected DCs. (B) The mean (± SEM) percentage of DC-SIGN+ intracellular p24Gag positive MDDCs determined from infections of cells derived from three donors infected as in (A). (C) Replication kinetics of Lai-YU2 and Lai-YU2/ΔVpr in MDDCs infected at MOI =1. MDDC supernatants were harvested every three days and analyzed for p24Gag content by an ELISA. Data shown are the mean (± SEM) for three independent experiments with MDDCs derived from three independent donors. (D) Schematic of DC-T cell co-culture set up. MDDCs were infected with Lai-YU2 or Lai-YU2/ΔVpr (MOI = 1). At two days post infection, autologous CD4+ T cells (PHA/IL2 treated) were added at a 2:1 ratio to MDDCs or infected with cell-free virus in parallel (MOI = 1). Supernatants were harvested on day 6 and day 9 post infection (day 3 or 6 for cell-free CD4+ T cell infection), and the p24Gag content in the culture supernatants determined by an ELISA. (E) The data shown is the kinetics of p24Gag production in cell culture supernatants from a representative infection of MDDCs only, CD4+ T cell only or MDDC - CD4+ T cell co-cultures. (F) The mean (±SEM) p24Gag present in the supernatant from five independent donor infections of CD4+ T cells only or DC-CD4+ T cell co-cultures at day 6 post infection (day 3 post infection for cell free CD4+ T cell infections). Significance calculated using paired student's T tests where *p<0.05, **p<0.01, ***p<0.001.

Numerous studies have demonstrated robust HIV-1 replication in DC-T cell co-cultures at levels greater than that observed in infections of either cell type alone, and is dependent on rapid highly efficient transmission of DC-derived progeny virions to $CD4^+$ T cells across infectious synapses (42, 43, 54–60). We sought to determine the effect, if any, of Vpr-deficiency on DC-mediated virus spread to $CD4^+$ T cells. MDDCs were first infected with wild type Lai-YU2 or Lai-YU2/ΔVpr and cultured for two days, prior to initiation of co-culture with autologous activated $CD4^+$ T cells (Figure 8.1D). There was a substantial enhancement of virus replication in co-cultures infected with WT virus, compared to ΔVpr virus infections (Figure 8.1E and F; ~7-fold increase). Interestingly the difference between WT and ΔVpr virus replication in DC-T cell co-cultures was greater than that observed in infections of MDDCs or $CD4^+$ T cells alone (Figure 8.1E and F). Together, these results suggest that the replication defect observed in MDDCs infected with HIV-1/ΔVpr translates to $CD4^+$ T cells during cell-to-cell contact and transmission.

Defects in Vpr infection are independent of viral glycoprotein expression

Previous studies have suggested a requirement for Vpr in maintaining robust HIV-1 gp120 expression in MDMs and MDDCs by counteracting a myeloid cell-intrinsic mechanism of Env degradation (22, 28, 44). To begin to understand the underlying mechanism accounting for the replication defect of HIV-1/ΔVpr in DCs, we examined viral protein expression in MDDCs infected with wild type Lai-YU2 or Lai-YU2/ΔVpr (MOI = 3). Infected cells were lysed 6 days post infection for quantitative western blot analysis. We did not observe any steady-state differences in gp120 expression when normalized to Gag (p55 and p24) levels in MDDCs infected with WT or ΔVpr viruses (Figure 8.2A). Quantification of immunoblots from infected MDDC lysates derived from four independent donors showed no significant differences in gp120 expression (Figure 8.2B). We next sought to determine if Vpr-deficiency might result in decreased gp120 incorporation in virus particles derived from productively infected DCs. MDDC culture supernatants were harvested on multiple days post infection and pooled supernatants were concentrated over a sucrose cushion prior to western blot analysis. We again failed to observe any significant differences in levels of gp120 incorporation between virus particles derived from WT or ΔVpr infected MDDCs (Figure 8.2C and D). The consistency of the replication defect of HIV-1/ΔVpr virus in MDDCs in the absence of any significant differences in gp120 expression suggests that previously hypothesized

Vpr-dependent enhancement of gp120 production is unlikely to account for the observed replication defect in the present study (28, 44).

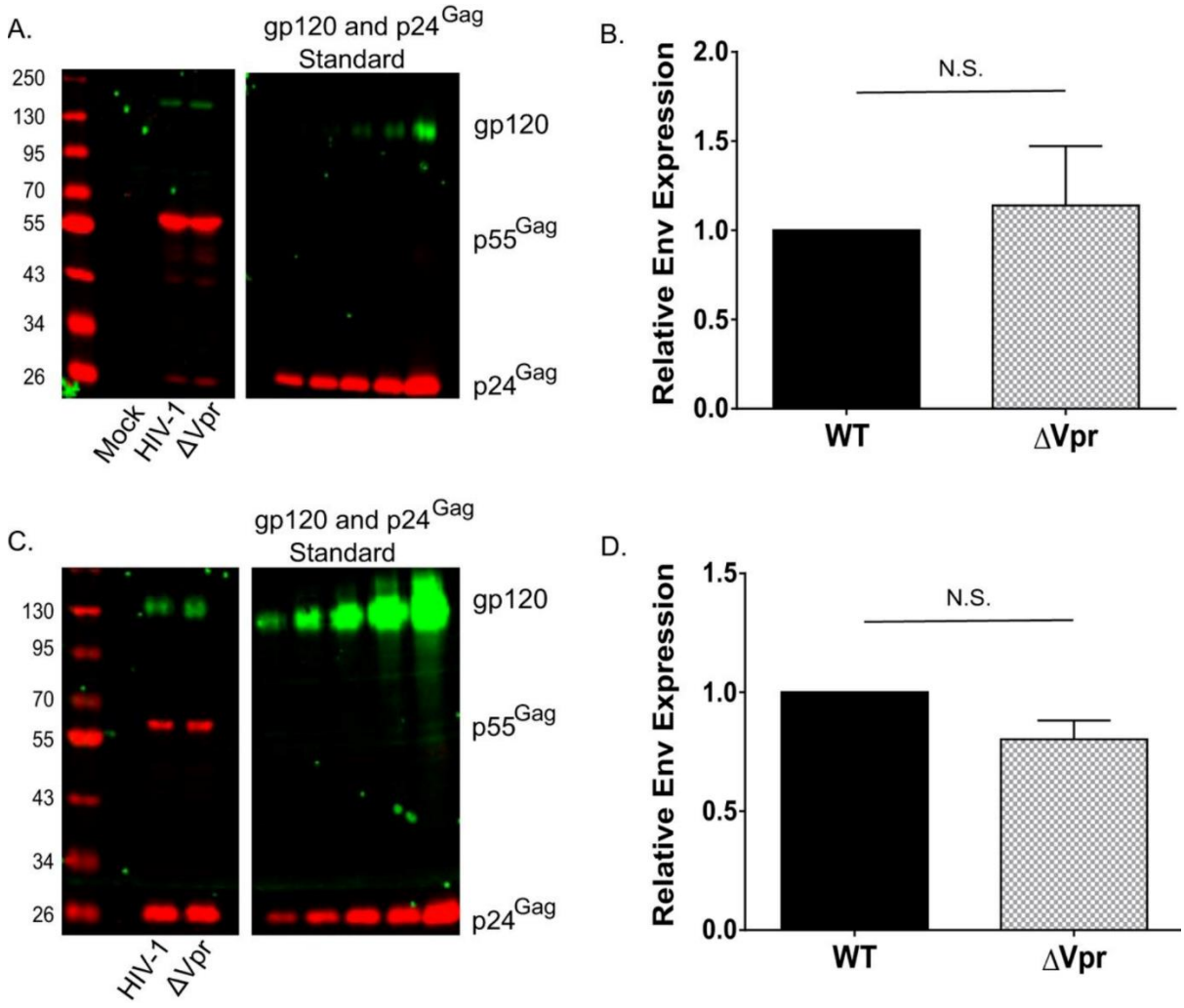

Figure 8.2. Vpr does not regulate Env expression in infected MDDCs or incorporation of Env into MDDC-derived virions.
(A) Western blot analysis of mock infected, Lai-YU2 (WT) or Lai-YU2ΔVpr infected MDDCs (MOI = 3) for p55Gag and gp120 expression at day 6 post infection. (B) Quantification of western blots for p55Gag and gp120 in infected MDDCs as in (A) from four independent experiments. The gp120 band intensity was quantified and normalized to p55Gag from experiments with infected MDDCs derived from 4 donors. Data shown are mean (± SEM). (C) Western blot analysis of p24Gag and gp120 expression in mock infected, Lai-YU2 (WT) or Lai-YU2ΔVpr infected MDDCs (MOI = 5). MDDC culture supernatants were harvested at days 3, 6, and 9 post infection, pooled and concentrated over a 20% sucrose cushion and virus pellets lysed for western blot analysis. (D) Quantification of western blot analysis from MDDC-derived virions from three independent donors. The band intensity for gp120 was quantified and normalized to p24Gag band intensity. Data shown are mean (± SEM). Significance calculated using a one sample T test where N.S>0.05.

Infection with Vpr-deficient HIV-1 does not induce type 1 IFN

Exposure of target cells to type I IFN potently restricts HIV-1 replication *in vitro* (61–69). In addition, recent studies have suggested that infection with ΔVpr virus induces type I IFN (18, 28, 31–34). Hence, we sought to determine if induction of an early type I IFN response in HIV-1/ΔVpr infections of MDDCs accounts for the restricted virus replication and spread. MDDCs infected with wild type Lai-YU2 or Lai-YU2/ΔVpr virus were harvested 48 h post infection, and the mRNA expression levels of IFNβ and the type I IFN-inducible protein, interferon-γ-inducible protein 10 (IP-10) were quantified by qRT-PCR. At 48 h post-virus exposure, we did not detect significant increases in IFN-β mRNA levels in wild type or ΔVpr infected cells compared to mock infected cells (Figure 8.3A). While expression of the ISG, IP-10, was robustly induced by establishment of productive HIV-1 infection of DCs, differences in IP-10 mRNA levels between WT and ΔVpr virus infections were not statistically significant (Figure 8.3B). Note that pre-treatment of cells with azidothymidine (AZT) reduced induction of IP-10 mRNA levels to that observed in mock infected cells, suggesting that induction of IP-10 expression in virus-exposed cells was dependent on *de novo* reverse transcription. In contrast, LPS treatment of MDDCs for 4 hours resulted in robust increases of both IFN-β and IP-10 mRNAs (Figure 8.3A, B). Inability to detect differences in mRNA expression levels of IFN-β in MDDCs infected with WT and ΔVpr viruses was also mirrored with the absence of differences in protein levels in infected MDDC culture supernatants (data not shown). We used a sensitive bioassay to measure type I IFN production in infected MDDC supernatants, and failed to detect any type I IFN production in HIV-1 infected MDDCs over mock infected controls (data not shown) (70). In contrast, IP-10 was robustly secreted in both Lai-YU2 (WT) and Lai-YU2ΔVpr infected MDDC culture supernatants at day 3 post-infection, though the magnitude of IP-10 induction was donor-dependent (Figure 8.3C). Furthermore, we observed a significant increase in IP-10 production upon WT virus infection of MDDCs as compared to mock-infected cells (Figure 8.3C). Again, AZT pre-treatment reduced secretion of IP-10 indicating that IP-10 production is dependent on completion of reverse transcription (Figure 8.3C). Taken together, these results suggest that Vpr deficiency does not result in the induction of type I IFNs during establishment of productive HIV-1 infection of MDDCs and is unlikely to play a role in the restriction of HIV-1/ΔVpr in DCs.

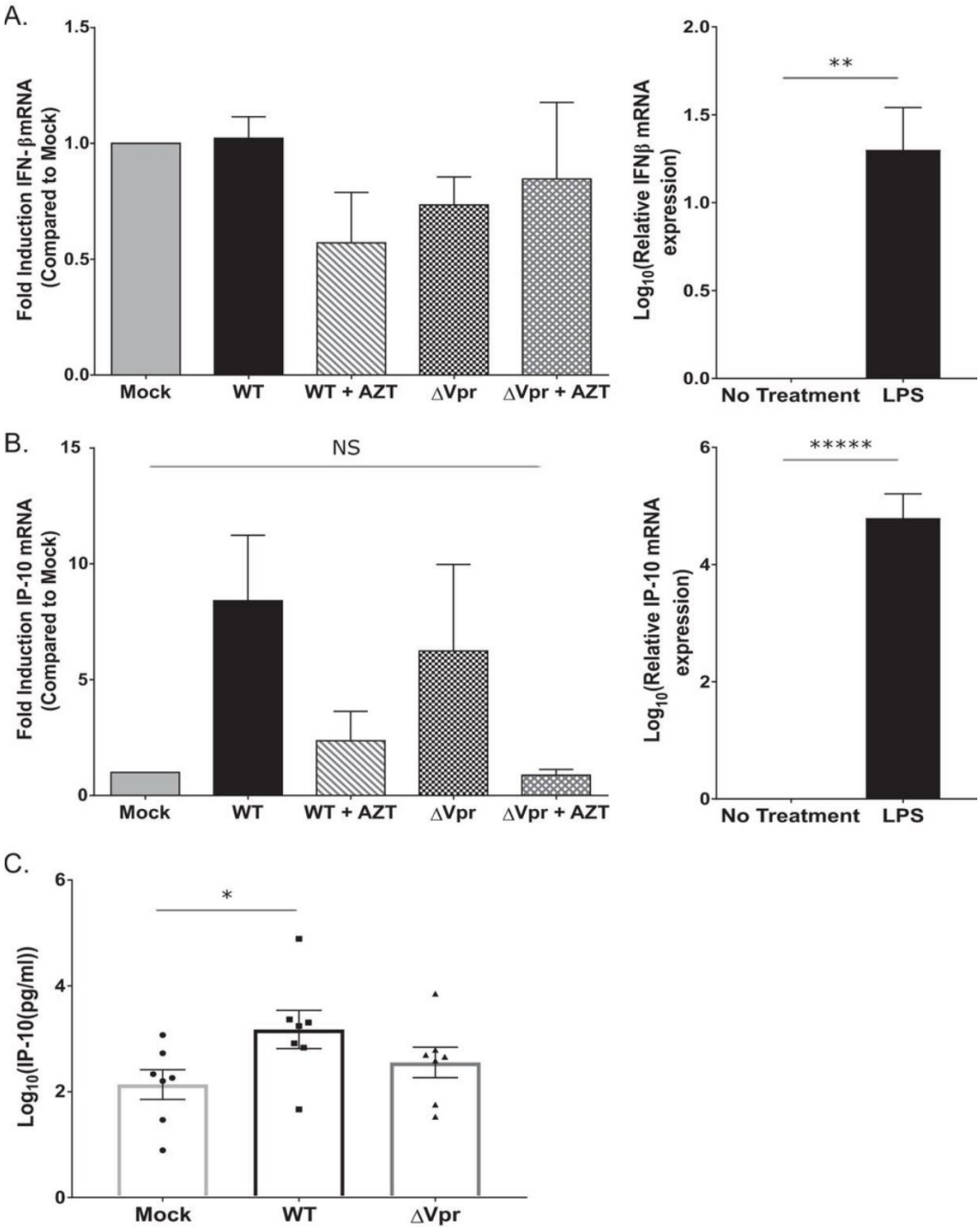

Figure 8.3. Vpr-deficiency does not result in enhanced type I IFN production in productively infected MDDCs.

Quantitative RT-PCR for IFNβ (A) and IP-10 (B) transcripts in infected MDDCs at 48 hours post infection. MDDCs were mock-infected or infected with Lai-YU2 or Lai-YU2/ΔVpr (MOI = 2) in the presence or absence of AZT (10 μM). The amount of IFNβ or IP-10 transcripts in infected MDDCs was normalized to the number of cells using a GAPDH control, and reported as relative to that of mock infected MDDCs (set as 1) for four independent donors. LPS treatment for 4 hours was used as a positive control for IFNβ and IP-10 production. Data is the log-transformed mean (± SEM) of seven donors. (C) Secreted IP-10 in MDDC culture supernatants infected with Lai-YU2 or Lai-YU2/ΔVpr (MOI = 1) at day 3 post infection was measured by an ELISA. The data shown are the log-transformed mean (± SEM) of independent experiments with MDDCs derived from four donors for (A) and (B) and six donors for (C). Significance

calculated using a paired student's T test or a one value T test (when comparing normalized data) where N.S>0.05, *p<0.5, **p<0.1, ***p<0.01, ****p<0.001, *****p<0.0001.

Infection with ΔVpr viruses results in decreased infection in a single round of replication and is rescued by virion-associated Vpr

To identify the step of the virus replication cycle in MDDCs that is affected by Vpr, we next performed single cycle of infection analysis. MDDCs were infected with HIV-1 reporter viruses pseudotyped with VSV-G and expressing luciferase upon establishment of infection that do (Lai-luc Δenv/G or WT) or do not express Vpr (Lai-luc Δenv/G ΔVpr or ΔVpr). Infection with ΔVpr virus resulted in a 3-5 fold decrease in luciferase expression compared to infection with WT virus (Figure 8.4A), suggesting that Vpr acts early in the HIV-1 replication cycle in MDDCs at steps preceding virion assembly and maturation. Since Vpr is a virion-associated protein, we next sought to determine whether incoming virion-associated Vpr was sufficient or if *de novo* synthesized Vpr was required for enhancement of virus replication in DCs. We produced Lai-luc Δenv/G ΔVpr complemented with HA-epitope tagged Vpr *in trans* (Lai-luc Δenv/G Vpr-*trans*) via co-transfection of HEK293T cells with a functional HA-Vpr expression plasmid and the Lai-lucΔenv/G ΔVpr proviral plasmid. HA-Vpr was efficiently incorporated in ΔVpr virus particles to levels similar to that observed in WT virus particles (Figure 8.4B). We then infected MDDCs with Lai-luc Δenv/G-WT, ΔVpr, or Vpr-*trans* viruses and lysed the cells on day 3 post-infection. Incorporation of Vpr *in trans* within incoming virus particles rescued ΔVpr virus infection in a single-round assay (Figure 8.4C), suggesting that virion incorporated Vpr is sufficient for overcoming cell-intrinsic blocks to early steps in HIV-1 replication in DCs.

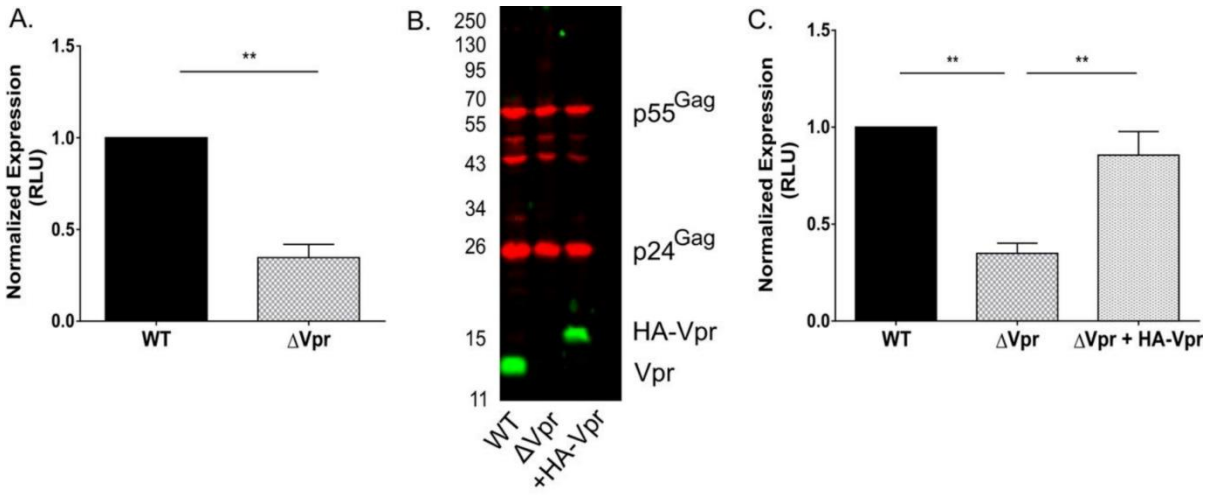

Figure 8.4. Infection of MDDCs with Vpr-deficient viruses results in block to HIV-1 replication in single round infection analysis.
(A) MDDCs infected with 40 ng p24Gag equivalent of VSV-G pseudotyped Lai-luc Δenv (WT or ΔVpr) were lysed 3 days post infection, and viral replication was quantified by measuring luciferase activity in cell lysates. The luciferase activity in ΔVpr infected cell lysates was normalized to that of WT virus-infected MDDC lysates and reported as mean (± SEM) of four independent experiments with MDDCs derived from four independent donors. (B) Western blot analysis of Vpr incorporation in virus particles (Lai-luc Δenv/G, Lai-luc Δenv/G ΔVpr, or Lai-luc Δenv/G Vpr-trans) derived from transient transfection of HEK293T cells. (C) MDDCs were infected with 40 ng p24Gag equivalents of viruses (Lai-lucΔenv, Lai-lucΔenvΔVpr, or Lai-lucΔenvΔVpr + HA-Vpr), and lysed 3 days post infection. Cell lysates were analyzed for luciferase activity and the data reported is normalized to that observed with WT-virus infection and is mean (± SEM) from 4 independent experiments. Significance calculated using a paired student's T test or a one value T test (when comparing normalized data) where *p<0.05, **p<0.01.

Proviral LTR-mediated transcriptional activity is attenuated in Vpr-deficient virus infection in DCs

Since the block to HIV-1/ΔVpr infection in MDDCs is evident within a single round of replication, and is independent of the mode of virus entry (VSV-G pseudotyped virus infection was also restricted, Figure 8.4A), we assessed the effect of Vpr-deficiency on HIV-1 reverse transcription (RT) and integration efficiency in DCs. We used qPCR to measure RT-products and the number of proviruses at day 3 post-infection using R-U5 and Alu-Gag primer pairs, respectively (71, 72). Infections were also performed in the presence of AZT to control for contaminating input plasmid DNA. In contrast to previously published findings (41), we saw no decrease in the number of RT products (Figure 8.5A) or integrants (Figure 8.5B, C) upon infection with ΔVpr virus compared to WT virus infections (Figure 8.5A, B and C). Previous studies have suggested that Vpr can modulate HIV-1 LTR transcriptional activity (29, 46, 73–75). We therefore asked if the block to HIV-1/ΔVpr infection occurs at the stage of viral transcript production. To determine the effect of Vpr on LTR-mediated transcription from proviruses, we used qRT-PCR to measure multiply-spliced *tat/rev/nef* transcripts at 48 h post infection (Figure 8.5D). Similar to our findings

with luciferase reporter expression in infected DCs, we observed a 4-fold decrease in the number of multiply-spliced HIV-1 transcripts in HIV-1/ΔVpr-infected cells suggesting that Vpr-deficiency results in inhibition of proviral LTR-mediated transcription in DCs.

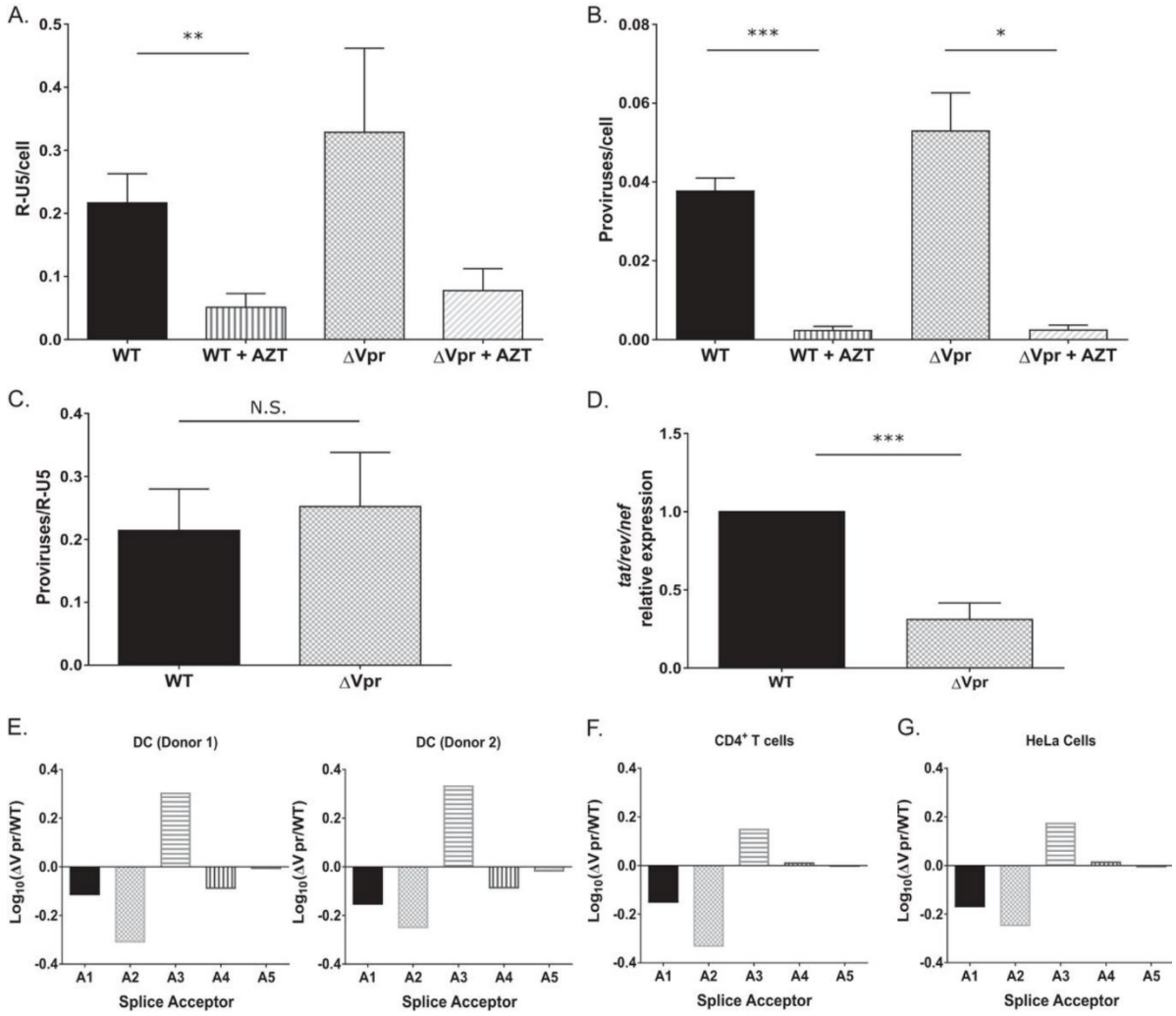

Figure 8.5. Viral transcription is attenuated in ΔVpr virus infected MDDCs.
(A-C) MDDCs infected with WT or ΔVpr viruses (MOI = 2) in the presence or absence of AZT (10 μM) were lysed 72 h post infection, and processed for DNA isolation. Note that infected cells were cultured in the presence of indinavir (1μM) to prevent viral spread. QPCR was used to detect early RT products (A) and integrated proviruses (B) by R-U5 and Alu-PCR primer sets and the number of integrated proviruses normalized to early RT products for each infection is shown in (C). The data reported is the mean (± SEM) of three independent experiments. (D) The numbers of multiply-spliced viral transcripts (*tat-rev-nef*) in MDDCs infected with Lai-YU2 or Lai-YU2/ΔVpr (MOI = 1) was determined at 48 hours post infection by qRT-PCR. Viral transcripts were measured using primers specific to *tat/rev/nef* multiply-spliced transcripts. Data shown are mean (± SEM) of four independent experiments with MDDCs derived from four donors. (E) Quantification of 4 kb class of splice variants for MDDCs infected with Lai-YU2 or Lai-YU2/ΔVpr (MOI = 2) for 72 hours. The data was normalized, log10 transformed, and then graphed according to splice acceptor usage. Histograms show fold changes in splicing from D1 to each of the 5 viral splice acceptor sites A1 through A5 relative to a WT control. Splicing was quantified using a Primer ID-splicing assay for MDDCs from two independent infections (e), productively infected CD4+ T cells (F) and HeLa cells (G) and is data from a single deep

sequencing experiment. Significance calculated used unpaired student's T test where *p<0.05, **p<0.01, and ***p<0.001.

We next sought to determine if the decrease in multiply-spliced viral mRNA levels in HIV-1/ΔVpr-virus infected MDDCs were driven by changes in the pattern of viral mRNA splicing. We used a novel Prime ID-tagged deep sequencing assay (76, 77) to determine the relative abundance of different splice variants in WT and ΔVpr infected DCs, and compared viral splice site usage to that observed in WT or ΔVpr-infected $CD4^+$ T cells and HeLa cells (Figure 8.5E-G). Data depicts the relative quantity of 4 kb singly-spliced mRNA for each splice acceptor and is reflective of the changes observed in the 1.8 kb multiply-spliced mRNA (data not shown). We detected minor differences in splice acceptor usage between WT and ΔVpr infections in MDDCs. We observed small decreases in the use of the Vif [A1] and Vpr [A2] splice acceptors and a small increase in the use of the Tat [A3] splice acceptor, but these differences were well within the normal range of splicing variation seen in productive viral infections (77). These small differences in splice site usage were consistently observed in infections of $CD4^+$ T cells and HeLa cells. Since the differences in splicing are both relatively small and observed in two cell types (primary activated $CD4^+$ T cells and HeLa cells) that do not restrict ΔVpr virus replication, it is unlikely that efficiency of viral mRNA splicing or choice of mRNA splice acceptor sites is a contributing factor to the restricted replication of ΔVpr virus in MDDCs.

Mutations in the C-terminal end of Vpr or those that disrupt binding to $CRL4^{DCAF1}$ ubiquitin ligase attenuate viral replication in DCs

A range of functions have been attributed to Vpr, including G2/M cell cycle arrest, enhancing fidelity of reverse transcription, nuclear import and/or nuclear tethering of the pre-integration complex, and induction of apoptosis (12–14, 30, 41, 46, 78). To clarify which of the known functions of Vpr are important for enhancing HIV-1 replication in DCs, a panel of mutations were introduced in Vpr ORF with previously characterized effects on Vpr functions. HEK293Tderived virus particles were analyzed by quantitative western blotting to assess incorporation of mutant Vpr proteins into viral particles (Figure 8.6A). While all viral mutants expressed and incorporated Vpr in virus particles, the mutants Vpr-F34I and Vpr-H71R had slightly decreased incorporation levels of Vpr compared to wild type viruses (Figure 8.6A), though both wild type and Vpr-mutant viruses were equally infectious on TZM-bl cells on a per particle basis (Figure 8.6B). MDDCs were infected with replication competent HIV-1 (WT or Vpr-mutants) at equal

MOIs and the extent of viral replication was measured by periodic quantification of $p24^{Gag}$ in cell-free culture supernatant by an ELISA (Figure 8.6C). Since there was donor-to-donor variability in the kinetics and extent of virus replication in DCs, we calculated the area under the curve of replication kinetics obtained from four independent infections (Figure 8.6D). As depicted in Figure 8.6C and D, infection with both Vpr-Q65R and Vpr-H71R mutant viruses resulted in significantly attenuated virus replication and spread, similar to what was observed with ΔVpr virus replication in MDDCs (Figure 8.6C and D). In contrast, replication of both Vpr-F34I and Vpr-W54R mutants was not significantly different from that observed with wild type virus infections (Figure 8.6C and D). Cumulative analysis revealed that replication of Vpr-Q65R and Vpr-H71R mutants, which lack the ability to associate with the $CRL4^{DCAF1}$ complex (16–18, 30, 45, 46, 49, 79, 80), was significantly reduced ($p<0.01$), similar to that observed with ΔVpr virus infection (Figure 8.6D). Interestingly, replication of Vpr-F34I mutant which incorporates reduced levels of Vpr in virions (Figure 8.6A), and displays reduced association with the nuclear envelope, (30, 45, 50) was slightly enhanced over that observed with wild type virus replication (Figure 8.6D; $p<0.01$), suggesting a threshold amount of functional Vpr that is still present in the incoming virus particle is sufficient for establishment of productive infections in DCs. The mutation Vpr-W54R, which ablates binding of Vpr to UNG2 (19, 45, 80–82) had a negligible effect on viral replication in DCs.

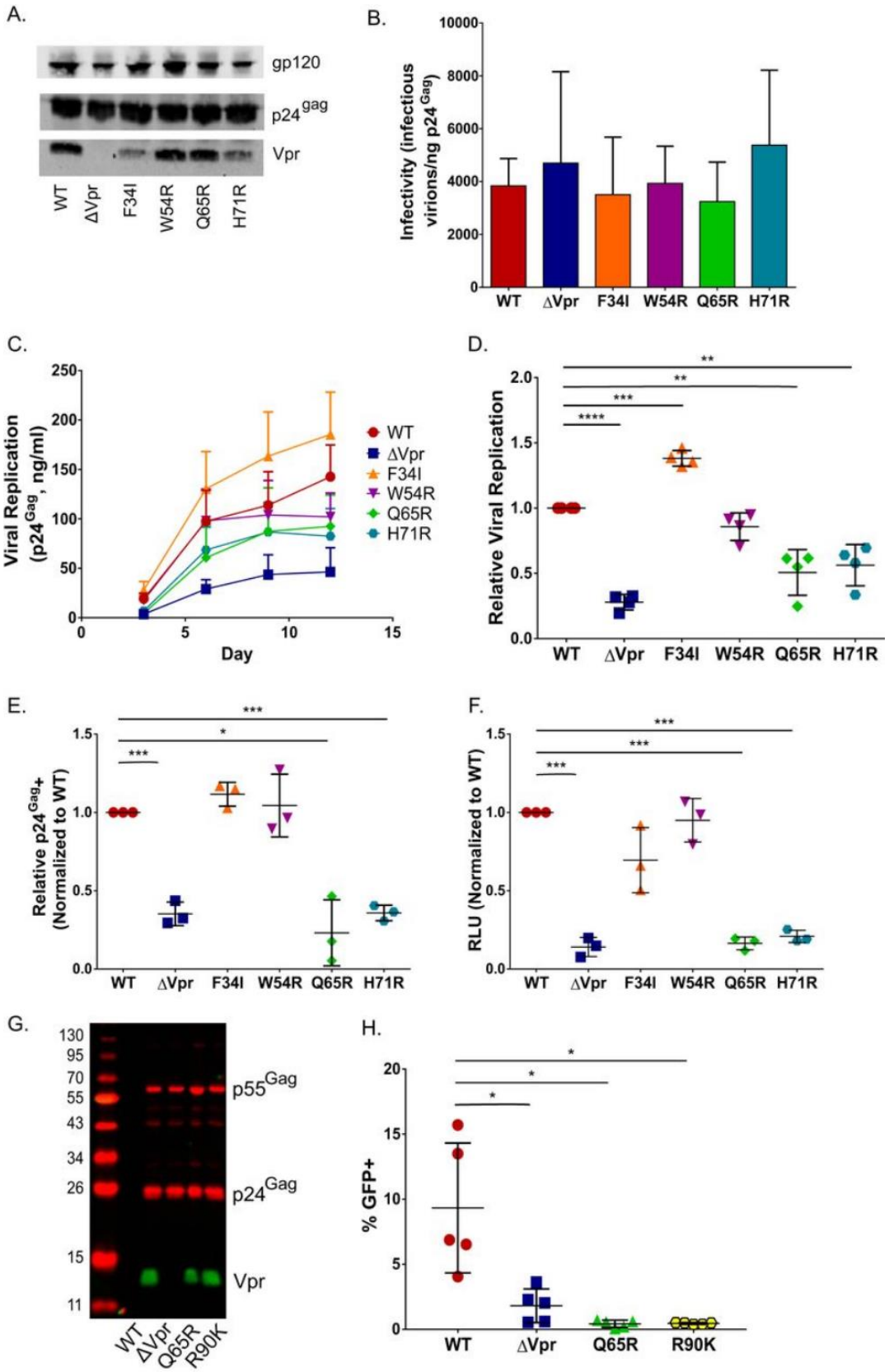
A.
gp120
p24gag
Vpr
WT
ΔVpr
F34I
W54R
Q65R
H71R
B.
Infectivity (infectious virions/ng p24Gag)
10000
8000
6000
4000
2000
0
WT
ΔVpr
F34I
W54R
Q65R
H71R
C.
Viral Replication (p24Gag, ng/ml)
250
200
150
100
50
0
0
5
10
15
Day
WT
ΔVpr
F34I
W54R
Q65R
H71R
D.
Relative Viral Replication
2.0
1.5
1.0
0.5
0.0
**
**
***
****
WT
ΔVpr
F34I
W54R
Q65R
H71R
E.
Relative p24Gag+ (Normalized to WT)
1.5
1.0
0.5
0.0
***
*
***
WT
ΔVpr
F34I
W54R
Q65R
H71R
F.
RLU (Normalized to WT)
1.5
1.0
0.5
0.0
***
***
***
WT
ΔVpr
F34I
W54R
Q65R
H71R
G.
130
95
70
55
43
34
26
15
11
p55Gag
p24Gag
Vpr
WT
ΔVpr
Q65R
R90K
H.
% GFP+
20
15
10
5
0
*
*
*
WT
ΔVpr
Q65R
R90K

Figure 8.6. Vpr mutants deficient for interaction with DCAF1/DDB1/E3 ubiquitin ligase and inducing G2 cell cycle arrest are attenuated in a single cycle of replication analysis in MDDCs. (A) Representative western blot analysis of HEK293T- derived Lai-YU2 (WT) and indicated Vpr mutant viruses used for MDDC infections. Blots were probed with anti-p24Gag, anti-Vpr and anti-gp120 antibodies. (B) Infectivity of Lai-YU2 and corresponding Vpr mutants in TZM-bl cells is reported as the number of infectious units (blue cells) per ng of p24Gag equivalent and are the mean (± SEM) of three independent viral preparations. (C) Viral growth curves of four independent infections of MDDCs with Lai-YU2 and indicated Vpr mutants in DCs. Viral growth was determined by analyzing p24Gag release into cell culture supernatants at days 3, 6, 9 and 12 post infection and determined by ELISA. (D) Area under the curve compiled for four independent MDDC infections represented in (C) normalized to WT virus infection, set as 1 (mean ± SEM). (E) The percentage of p24Gag positive MDDCs at day 3 post infection as measured by intracellular p24Gag staining and FACS analysis. Cells were treated with indinavir (1 μM) post virus exposure to prevent viral spread. The data was normalized to WT virus infection, set as 1, and depicts the mean (± SEM) of three independent infections of MDDCs from three donors. (F) MDDCs infected with 40 ng p24Gag equivalents of Lai-luc Δenv/G (WT or Vpr mutants) were lysed 3 days post infection, and viral replication was quantified by measuring luciferase activity in cell lysates. The luciferase activity in Vpr-mutant infections was normalized to that of WT virus infections, set as 1, and the data shown are the mean (± SEM) for three independent experiments. (G) Western blot analysis of HEK293T-derived Lai-GFP Δenv/G (WT) or indicated Vpr mutant virus particles. (H) MDDCs infected with Lai-GFP Δenv/G (WT) or indicated Vpr-mutants (MOI = 3) were harvested at day 3 post infection and processed for FACS analysis. The data shown is the mean percentage of GFP+ cells (± SEM) of five independent experiments with cells derived from five independent donors. Significance calculated using a paired student's T test or a one value T test (when comparing normalized data) where *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

We next sought to determine which of these Vpr mutants could recapitulate the single cycle of replication defect observed with HIV-1/ΔVpr infection in MDDCs (Figure 8.4A). We infected MDDCs with either replication competent viruses (Lai-YU2, WT or Vpr mutants, MOI = 1) in the presence of a protease inhibitor (indinavir) or with equal amounts of $p24^{Gag}$ equivalents of Lai-luc Δenv/G encoding the various Vpr mutations. Similar to the results observed with replication competent viruses, both the number of $p24^{Gag}$-postitive cells (Figure 8.6E) and luciferase production (Figure 8.6F) from infections with Vpr-Q65R and Vpr-H71R mutants were significantly attenuated in a single round of infection compared to isogenic WT viruses. While the host protein targeted by HIV-1 Vpr to induce G2 cell cycle arrest has not been identified, the C-terminal tail of the protein has been proposed to bind the unknown host factor, and mutations in the C-terminal tail of Vpr abrogate the ability of Vpr to induce G2 cell cycle arrest (45). To determine the role of Vpr-mediated G2 cell cycle arrest on virus infection enhancement in DCs, an additional mutation, Vpr-R90K was introduced in GFP-expressing single-cycle virus (Lai-GFP Δenv/G). The Vpr-R90K mutant can bind $CRL4^{DCAF1}$ complex but fails to induce G2 arrest in cycling cells (45, 50, 80). Despite equivalent incorporation into the virion as WT Vpr (Figure 8.6G), infection of MDDCs with Vpr-R90K mutant resulted in significant infection defect in single round analysis (Figure 8.6H), similar to what was observed in infections with ΔVpr or Vpr-Q65R viruses, suggesting that interaction with a putative host factor whose degradation is critical for the induction of G2 cell cycle arrest is required to enhance HIV-1 infection of DCs. Together, our data suggests that there is a novel block to HIV-1 infection

in MDDCs in the absence of Vpr that is present in a single round of infection and manifests at the stage of viral transcription. Further studies are underway to determine the exact mechanism by which Vpr alleviates the DC-intrinsic block to HIV-1 replication.

Discussion

In the work presented here, we examined the role of Vpr in establishing productive HIV-1 infection of DCs. Previous work in the field suggests that Vpr likely regulates a complex network of host interactions that may vary depending on the cell type infected. We find that, unlike what has been previously observed in activated $CD4^+$ T cells and monocyte-derived macrophages (20, 22–25, 29), infection of MDDCs with ΔVpr viruses was significantly attenuated when compared to WT HIV-1 infections (Figure 8.1), similar to the findings reported by de Silva et al (41). Interestingly, Vpr-mediated enhancement was observed within both a single round viral infection as well as in spreading infections, contrary to what has been reported previously (28, 41). Furthermore, the single round replication defect could be rescued by complimenting back Vpr *in trans* in the incoming virion (Figure 8.4) indicating that incoming virion-associated Vpr is necessary for the establishment of efficient HIV-1 infection of DCs. Initiating infections with the Vpr mutants, Vpr-Q65R, Vpr-H71R and Vpr-R90K that either lack the ability to engage the $CRL4^{DACF1}$ complex or bind the yet-to-identified host factor(s) necessary for inducing G2 cell cycle arrest, displayed similar replication deficits to that observed with ΔVpr virus in both spreading infections and single round infection analysis (Figure 8.6).

Surprisingly, the block to ΔVpr virus replication in MDDCs was evident at a post-integration step and resulted in reduced numbers of viral mRNAs, suggesting that Vpr is acting either directly or indirectly to enhance transcription from the viral LTR (Figure 8.5). It has been reported previously that Vpr can transactivate the viral LTR in a number of cell types and that this function correlates with the ability of Vpr to induce G2 cell cycle arrest (26, 29, 46, 73, 75). Previous studies have also shown that both SIV_{mac} and SIV_{agm} Vpr can also transactivate their respective LTRs (74, 83, 84), suggesting that this is a conserved function among non-human primate lentiviral Vpr proteins. While it is possible that Vpr-mediated transactivation could be more robust in DCs than in $CD4^+$ T cells (Figure 8.1E), another hypothesis is that Vpr is indirectly activating transcription to promote infection in cells that have a higher barrier to infection.

Unlike most of the other lentiviral accessory proteins, Vpr is actively packaged into the budding virion through associations with the p6 region of Gag (3–6, 10). Our work in MDDCs suggests that there may be a novel role for virion-associated Vpr to enhance viral transcription and increase infection of DCs. These findings are at odds with recently published studies on the role of Vpr in modulating *de novo* HIV-1 Env production in productively infected macrophages and MDDCs (28, 44). While we do occasionally see a decrease in viral Env production during infection with ΔVpr virus in MDDCs (one out of four donors tested), infection of MDDCs from most of the donors revealed no differences in Env expression or virion incorporation (Figure 8.2). It is possible that the use of different viral clones, primary cell variation derived from multiple donors, or different infection conditions might play a role in the differences between our results and those described previously. Since we observed infection differences in a single-round infection assay, putative effects of Vpr on Env expression are unlikely to play a role in establishment and spread of virus infection in MDDCs and DC-T cell co-cultures.

HIV-1 is not unique among primate lentiviruses in expressing a protein that functionally allows for infection of DCs. HIV-2 and certain SIV lineages express Vpx, another small accessory protein that targets host restriction factor SAMHD1 for proteosomal degradation by recruiting it to the $CRL4^{DACF1}$ complex, and facilitates infection of MDDCs (51, 52, 85). Interestingly, Vpr-mediated replication enhancement in MDDCs was substantially attenuated upon infection with Vpr mutants (Q65R or H71R; Figure 8.6F) that lack ability to interact with $CRL4^{DCAF1}$ complex, or upon infection with Vpr-R90K mutant (Figure 8.6H), that fails to interact with the host factor(s) hypothesized to be recruited to the $CRL4^{DCAF1}$ complex for proteosomal degradation. Since Vpr is introduced into target cells along with the incoming virion because of its association with the viral capsid, we hypothesize that early interactions of Vpr with a host factor and recruitment of that protein to the $CRL4^{DCAF1}$ complex for proteosomal degradation is essential for promoting HIV-1 replication in DCs, similar to the ability of Vpx from SIV_{mac}/SIV_{smm}/HIV-2 lineages to promote infection of DCs.

Across primate lentiviral Vpr evolution, induction of DDR and G2 cell cycle arrest are conserved functions, and Vpr proteins from diverse primate lentiviruses have been shown to associate with and degrade many DDR regulatory proteins including the SLX4com, HLTF, and UNG2 (18–21, 80, 81, 86–89). While DDR activation may represent a cell-intrinsic antiviral response, it has been suggested that

both RNA and DNA viruses induce DDR signaling to promote cellular conditions that are favorable for viral replication (83, 90–93). For instance, induction of DDR signaling activates ataxia-telangiectasia mutated (ATM) kinase which results in nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) activation (94, 95). Additionally, the DDR pathway also directly activates pro-inflammatory responses through the induction of interferon regulatory factors (IRFs) or through the recruitment of co-activators and chromatin modifying complexes, such as ten-eleven translocation methylcytosine (TET) dioxygenases, which we hypothesize might also activate viral transcription (96). Since the barrier to successful establishment of infection in non-cycling, metabolically quiescent cells like MDDCs is higher than that in activated $CD4^+$ T cells or MDM, Vpr-mediated activation of NF-κB and co-activator recruitment to the viral LTR might be a viral strategy for overcoming the restrictive cellular environment and for optimal production of progeny virions. In line with this hypothesis, numerous studies have documented that Vpr is able to modulate NF-κB activity in different cell lines and primary cells, though these studies rarely agree on the mechanism of regulation or direction of modulation (26, 75, 97–102). Recent work from Höhne, et al has shown similar effects of Vpr on viral replication in non-activated primary $CD4^+$ T cells, which have similar barriers to infection as MDDCs including increased expression of SAMHD1 and low baseline NF-κB activity (26, 103, 104). Some studies have shown virion-associated Vpr-dependent activation of NF-κB occurs via a transforming growth factor-β-activated kinase 1 (TAK1) signaling cascade, while other studies have shown that secreted or synthetic Vpr stimulates NF-κB signaling through a TLR4-dependent mechanism (75, 99, 100, 102). Our data also demonstrates upregulation of IP-10 upon HIV-1 (WT) infection (Figure 8.3C) which is also dependent on NF-κB activation (105–107).These results suggest a link between Vpr-mediated NF-κB activation in MDDCs and enhanced viral gene expression and pro-inflammatory cytokine secretion, which may act *in vivo* to enhance recruitment, activation and infection of $CD4^+$ T cells, resulting in increased viral dissemination (Figure 8.1E) (27, 75, 99).

Studies with peripheral blood myeloid MDDCs and monocyte-derived MDDCs from HIV-1 elite controllers have shown that these cells may be critical for viral control, acting to capture virus and enhance T cell-specific immunity to HIV-1, while being less susceptible to HIV-1 infection compared to MDDCs from healthy controls (108, 109). Understanding the mechanisms that control HIV-1 replication in

MDDCs which are overcome by Vpr, might lead to new insights on viral dissemination and persistence *in vivo*, and development of novel anti-HIV-1 therapeutics.

Materials and Methods

Plasmids

HIV-1 proviral plasmids Lai/YU2 env, Lai/Bal, Lai-luc ∆env (Env deficient HIV-1 containing a luciferase reporter gene in place of *nef*), Lai-GFP ∆env (Env deficient HIV-1 containing GFP in place of *nef*) and the HA-Vpr expression plasmid have been previously described (30, 110). Proviral Lai (CXCR4-tropic) clones containing Vpr mutations, F34I, W54R, and H71R and frame-shift mutation in Vpr (ΔVpr) have been described previously (30, 46, 80, 111). These Vpr mutations were transferred to Lai-YU2 env, Lai-luc ∆env or Lai-GFP ∆env proviral plasmids using Apa I and Sal I restriction sites. To create proviral clones encoding Vpr-Q65R mutation, the Apa I – Sal I fragment of Lai-YU2 env was subcloned into pSL1180 cloning vector (Stratagene) and site directed mutagenesis was preformed using a kit (QuikChange II, Agilent Technologies) and the following primers: 5'-GCCATAATAAGAATTCTGCGACAACTGCTGTTTATCCATTTC-3' and 5'-GAAATGGATAAACAGCAGTTGTCGCAGAATTCTTATTATGGC-3'. The mutated fragment was ligated back into Lai-YU2 env, Lai-luc ∆env or Lai-GFP ∆env using Apa I-Sal I restriction sites. The point mutation Vpr-R90K was derived by sub-cloning the Sal I – BamH I fragment into pSL1180 and via site directed mutagenesis (QuikChange II, Agilent Technologies) using the following primers: 5'-CGTTACTCAACAGAGGAGAGCAAAAAATGGAGCCAGTAGATCCTAGAC-3' and 5'-GTCTAGGATCTACTGGCTCCATTTTTTGCTCTCCTCTGTTGAGTAACG-3'. The mutated fragment was cloned back into Lai-GFP∆env.

Cells and viruses

MDDCs were derived from monocytes isolated from peripheral blood mononuclear cells as previously described (112). TZM-bl and HEK 293T cells have been described previously (112–114). Replication competent Lai-YU2 env viruses were derived using calcium phosphate mediated transient transfection of HEK 293T cells, as described previously (115). HIV-1 vectors were generated from HEK293T cells via co-transfection of Lai-luc∆env or Lai-GFP∆env with VSV-G expression plasmid. Virus-containing cell supernatants were harvested 2 days post-transfection, passed through 0.45 µm filters, and

stored at -80°C until further use. For some experiments, virus particles were concentrated by ultracentrifugation on a 20% sucrose cushion [24,000 rpm at 4˚C for 2 h with a SW28 rotor (Beckman Coulter)] (116). The virus pellets were resuspended in PBS, aliquoted and stored at -80 °C until use. The capsid content of HIV-1 was determined by a $p24^{Gag}$ ELISA while the infectious titer was determined via infecting TZM-bl cells (114, 117). Viral replication in MDDCs and DC- T cell co-cultures was determined by measuring $p24^{Gag}$ content in cell culture supernatants at indicated days post infection by an ELISA (117). Infection of MDDCs using luciferase reporter virus was analyzed using Bright-Glo Luciferase System (Promega) as previously described (118).

Drug treatments

In indicated experiments, cells were pretreated with zidovudine (AZT, 10 µM, NIH AIDS Research and Reference Reagent Program) for 30 minutes at 37°C prior to infection and AZT concentration was maintained for the duration of the cultures, or cells were treated with indinavir (1 µM, NIH AIDS Research and Reference Reagent Program) post-virus exposure.

Quantitative western blotting

To detect Gag and Env in cell and virus particle lysates, cell lysates (normalized to equivalent amounts of cell-associated Gag content) or 100 ng $p24^{Gag}$ virus equivalents (as determined by quantitative ELISA) were run through SDS-PAGE gels and transferred using a semi-dry transfer apparatus to nitrocellulose membranes as previously described (116). Blots were blocked and probed with rabbit anti-gp120 (a gift from Dr. Nancy Haigwood) and mouse anti-$p24^{Gag}$ (clone p24-2, NIH AIDS Research and Reference Reagent Program), followed by goat anti-mouse IgG DyLight 680 (Pierce) and goat anti-rabbit IgG DyLight 800 (Pierce). To determine Vpr incorporation, a polyclonal rabbit anti-Vpr antibody (clone 1-50, NIH AIDS Research and Reference Reagent Program) was used followed by goat anti-rabbit IgG DyLight 700. The membranes were scanned with an Odyssey scanner (Li-Cor).

Quantitative RT-PCR

For the quantitation of IFNβ and IP-10 mRNA, MDDCs ($2\text{-}4x10^6$ cells) were mock infected or infected with Lai-YU2 or Lai-YU2ΔVpr (MOI = 2). At 48 h post infection, cells were harvested for RNA isolation using RNAeasy (QIAGEN) RNA isolation kits and cDNA was synthesized using oligo dT primers and Superscript III RT (Invitrogen). cDNA corresponding to 200 ng of RNA was analyzed by qRT-PCR

using SYBR green (Invitrogen) to quantify mRNA levels for IFNβ (forward primer: 5'-ATTCTAACTGCAACCTTTCG-3' and reverse primer: 5'-GTTGTAGCTCATGGAAAGAG-3'), IP-10 (forward primer: 5'-TCATTGGTCACCTTTTAGTG-3' and reverse primer: 5'-AAAGCAGTTAGCAAGGAAAG-3') and GAPDH (forward primer: 5'-AGGGATGATGTTCTGGAGAG-3' and reverse primer: 5'-CAAGATCATCAGCAATGCCT-3'). The ΔΔCT value relative to GAPDH in the mock-infected cultures was set to 1, and the data from the infected cultures reported as fold enhancements. To determine the extent of *de novo* viral transcription, the number of *tat-rev-nef* multiply spliced transcripts was determined by qRT-PCR using SYBR green as described previously (58), with the following primer set: forward primer, 5'-GCGACGAAGACCTCCTCAG-3' and reverse primer, 5'-GAGGTGGGTTGCTTTGATAGAGA-3'. The data were normalized to GAPDH levels. As a control, MDDCs were treated with AZT (10μM) for 30 min prior to infection and drug levels were maintained during the course of infection.

Quantification of viral integration

To determine the number of proviral integrants, MDDCs ($3x10^6$ cells) were infected with virus (MOI = 3) for 2 h at 37˚C, washed with PBS twice and cultured for 72 h before cells were lysed for DNA extraction with a DNeasy kit (QIAGEN). As a background control, MDDCs were treated with 10 μM AZT for at least 30 min prior to infection. Quantitative *Alu*-PCR was performed using 20 ng of DNA with the following primer sets, as described previously (72). For the first step, the following primers were used: *Alu*-forward 5′-GCCTCCCAAACTGCTGGGATTACAG-3′ and Gag-reverse 5'-GCTCTCGCACCCATCTCTCTCC-3'. For the second step, the following primers were used: R-U5-F: 5′-GCCTCAATAAAGCTTGCCTTGA-3′ and R-U5-R: 5′-TCCACACTGACTAAAAGGGTCTGA-3′ with the following probe: R-U5-Probe: 5′-FAM-CCAGAGTCACACAACAGACG-TAMRA-3′. The data were normalized to a standard curve generated from infected HEK293 cell DNA (71, 72).

Splicing assay

The assay details are described separately in detail (77). Briefly, cDNA primers with an internal random sequence block (Primer ID; (76)) were designed to be within the *env* intron, to measure the 4 kb size class of spliced viral RNAs, or spanning the D4/A7 splice junction to measure the 1.8 kb size class. The reverse primer for the 4 kb size class was: 5'-

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNNNNNN**GTACGCTAATACTTGTAAA GATTGCAGTACATGTACTACTT-3' and the reverse primer for the 1.8 kb size class was: 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNNNNNN**CAGTCTGAGCTGGGAGGT GGGTTGC-3'. Whole cell RNA from infected cells was purified and used in a cDNA reaction. After removal of the cDNA primers, PCR was carried out using a downstream primer encoded in the cDNA primer tail and a forward primer placed just upstream of the D1 major donor site in the 5' noncoding region, 5'-GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG**NNNN**TGCTGAAGCGCGCACGGCAAG-3'. PCR products were sequenced using the MiSeq platform, and sequence reads with the same Primer ID were collapsed into a single read (to correct for skewing during PCR since each unique Primer ID tag represents a separate viral mRNA template). Data were processed using customized scripts that are available on request. The number of unique Primer IDs for each spliced product was used to determine the relative level of splicing from each splice donor to each splice acceptor in the viral genome with the exception of splicing events to the *nef* splice acceptor A7.

Cytokine measurements

Secreted IP-10 in MDDC culture supernatants was measured using a commercially available ELISA kit (Becton Dickinson). Secreted levels of bioactive type I IFN in infected MDDC supernatants was measured using a HEK293 ISRE-luc cell line which expresses luciferase under the control of an IFN-inducible promoter carrying the IFN-stimulated response element (70). Briefly, HEK293 ISRE-luc cells ($8x10^4$) were incubated with MDDC culture supernatants for 21 hours. Cells were lysed and luciferase activity in the cell lysates analyzed with Bright-Glo Luciferase System (Promega), as described above. Serial dilutions of recombinant interferon alpha ranging from 200-0.39 units/ml (PBL Interferon Source) were added to cells in each experiment for generating a standard curve.

FACS

Intracellular fluorescence-activated cell sorter (FACS) analysis for $p24^{Gag}$ was done using FITC-conjugated anti-$p24^{Gag}$ monoclonal antibody (KC57; Coulter) as previously described (117). Surface staining for CD11c was done using APC-conjugated anti-CD11c (Clone B-ly6, Becton Dickinson). Cells were analyzed using either LSRII or FACSCalibur (Becton Dickinson) instruments.

REFERENCES

1. Malim MH, Emerman M. 2008. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell Host Microbe 3:388–98.

2. Strebel K. 2013. HIV accessory proteins versus host restriction factors. Curr Opin Virol 3:692–9.

3. Selig L, Pages JC, Tanchou V, Prévéral S, Berlioz-Torrent C, Liu LX, Erdtmann L, Darlix J, Benarous R, Benichou S. 1999. Interaction with the p6 domain of the gag precursor mediates incorporation into virions of Vpr and Vpx proteins from primate lentiviruses. J Virol 73:592–600.

4. Bachand F, Yao XJ, Hrimech M, Rougeau N, Cohen EA. 1999. Incorporation of Vpr into human immunodeficiency virus type 1 requires a direct interaction with the p6 domain of the p55 gag precursor. J Biol Chem 274:9083–91.

5. Cohen EA, Terwilliger EF, Jalinoos Y, Proulx J, Sodroski JG, Haseltine WA. 1990. Identification of HIV-1 vpr product and function. J Acquir Immune Defic Syndr 3:11–8.

6. Yuan X, Matsuda Z, Matsuda M, Essex M, Lee TH. 1990. Human immunodeficiency virus vpr gene encodes a virion-associated protein. AIDS Res Hum Retroviruses 6:1265–71.

7. Lu YL, Bennett RP, Wills JW, Gorelick R, Ratner L. 1995. A leucine triplet repeat sequence (LXX)4 in p6gag is important for Vpr incorporation into human immunodeficiency virus type 1 particles. J Virol 69:6873–9.

8. Lavallée C, Yao XJ, Ladha A, Göttlinger H, Haseltine WA, Cohen EA. 1994. Requirement of the Pr55gag precursor for incorporation of the Vpr product into human immunodeficiency virus type 1 viral particles. J Virol 68:1926–34.

9. Kondo E, Mammano F, Cohen EA, Göttlinger HG. 1995. The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. J Virol 69:2759–64.

10. Accola MA, Bukovsky AA, Jones MS, Göttlinger HG. 1999. A conserved dileucine-containing motif in p6(gag) governs the particle association of Vpx and Vpr of simian immunodeficiency viruses SIV(mac) and SIV(agm). J Virol 73:9992–9.

11. Müller B, Tessmer U, Schubert U, Kräusslich HG. 2000. Human immunodeficiency virus type 1 Vpr protein is incorporated into the virion in significantly smaller amounts than gag and is phosphorylated in infected cells. J Virol 74:9727–31.

12. Rogel ME, Wu LI, Emerman M. 1995. The human immunodeficiency virus type 1 vpr gene prevents cell proliferation during chronic infection. J Virol 69:882–8.

13. He J, Choe S, Walker R, Di Marzio P, Morgan DO, Landau NR. 1995. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J Virol 69:6705–11.

14. Jowett JB, Planelles V, Poon B, Shah NP, Chen ML, Chen IS. 1995. The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. J Virol 69:6304–13.

15. Belzile J-P, Duisit G, Rougeau N, Mercier J, Finzi A, Cohen EA. 2007. HIV-1 Vpr-mediated G2 arrest involves the DDB1-CUL4AVPRBP E3 ubiquitin ligase. PLoS Pathog 3:e85.

16. DeHart JL, Zimmerman ES, Ardon O, Monteiro-Filho CMR, Argañaraz ER, Planelles V. 2007. HIV-1 Vpr activates the G2 checkpoint through manipulation of the ubiquitin proteasome system. Virol J 4:57.

17. Hrecka K, Gierszewska M, Srivastava S, Kozaczkiewicz L, Swanson SK, Florens L, Washburn MP, Skowronski J. 2007. Lentiviral Vpr usurps Cul4-DDB1[VprBP] E3 ubiquitin ligase to modulate cell cycle. Proc Natl Acad Sci U S A 104:11778–83.

18. Laguette N, Brégnard C, Hue P, Basbous J, Yatim A, Larroque M, Kirchhoff F, Constantinou A, Sobhian B, Benkirane M. 2014. Premature activation of the SLX4 complex by Vpr promotes G2/M arrest and escape from innate immune sensing. Cell 156:134–45.

19. Bouhamdan M, Benichou S, Rey F, Navarro JM, Agostini I, Spire B, Camonis J, Slupphaug G, Vigne R, Benarous R, Sire J. 1996. Human immunodeficiency virus type 1 Vpr protein binds to the uracil DNA glycosylase DNA repair enzyme. J Virol 70:697–704.

20. Lahouassa H, Blondot M-L, Chauveau L, Chougui G, Morel M, Leduc M, Guillonneau F, Ramirez BC, Schwartz O, Margottin-Goguet F. 2016. HIV-1 Vpr degrades the HLTF DNA translocase in T cells and macrophages. Proc Natl Acad Sci U S A 113:5311–6.

21. Hrecka K, Hao C, Shun M-C, Kaur S, Swanson SK, Florens L, Washburn MP, Skowronski J. 2016. HIV-1 and HIV-2 exhibit divergent interactions with HLTF and UNG2 DNA repair proteins. Proc Natl Acad Sci U S A 113:E3921-30.

22. Collins DR, Lubow J, Lukic Z, Mashiba M, Collins KL. 2015. Vpr Promotes Macrophage-Dependent HIV-1 Infection of CD4+ T Lymphocytes. PLoS Pathog 11:e1005054.

23. Dedera D, Hu W, Vander Heyden N, Ratner L. 1989. Viral protein R of human immunodeficiency virus types 1 and 2 is dispensable for replication and cytopathogenicity in lymphoid cells. J Virol 63:3205–8.

24. Eckstein DA, Sherman MP, Penn ML, Chin PS, De Noronha CM, Greene WC, Goldsmith MA. 2001. HIV-1 Vpr enhances viral burden by facilitating infection of tissue macrophages but not nondividing CD4+ T cells. J Exp Med 194:1407–19.

25. Balliet JW, Kolson DL, Eiger G, Kim FM, McGann KA, Srinivasan A, Collman R. 1994. Distinct effects in primary macrophages and lymphocytes of the human immunodeficiency virus type 1 accessory genes vpr, vpu, and nef: mutational analysis of a primary HIV-1 isolate. Virology 200:623–31.

26. Höhne K, Businger R, van Nuffel A, Bolduan S, Koppensteiner H, Baeyens A, Vermeire J, Malatinkova E, Verhasselt B, Schindler M. 2016. Virion encapsidated HIV-1 Vpr induces NFAT to prime non-activated T cells for productive infection. Open Biol 6.

27. Roesch F, Richard L, Rua R, Porrot F, Casartelli N, Schwartz O. 2015. Vpr Enhances Tumor Necrosis Factor Production by HIV-1-Infected T Cells. J Virol 89:12118–30.

28. Mashiba M, Collins DR, Terry VH, Collins KL. 2014. Vpr Overcomes Macrophage-Specific Restriction of HIV-1 Env Expression and Virion Production. Cell Host Microbe 16:722–35.

29. Connor RI, Chen BK, Choe S, Landau NR. 1995. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. Virology 206:935–44.

30. Vodicka MA, Koepp DM, Silver PA, Emerman M. 1998. HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection. Genes Dev 12:175–85.

31. Harman AN, Lai J, Turville S, Samarajiwa S, Gray L, Marsden V, Mercier SK, Mercier SK, Jones K, Nasr N, Rustagi A, Cumming H, Donaghy H, Mak J, Gale M, Churchill M, Hertzog P, Cunningham AL. 2011. HIV infection of dendritic cells subverts the IFN induction pathway via IRF-1 and inhibits type 1 IFN production. Blood 118:298–308.

32. Zahoor MA, Xue G, Sato H, Aida Y. 2015. Genome-wide transcriptional profiling reveals that HIV-1 Vpr differentially regulates interferon-stimulated genes in human monocyte-derived dendritic cells. Virus Res 208:156–63.

33. Zahoor MA, Xue G, Sato H, Murakami T, Takeshima S, Aida Y. 2014. HIV-1 Vpr induces interferon-stimulated genes in human monocyte-derived macrophages. PLoS One 9:e106418.

34. Hong HS, Bhatnagar N, Ballmaier M, Schubert U, Henklein P, Volgmann T, Heiken H, Schmidt RE, Meyer-Olson D. 2009. Exogenous HIV-1 Vpr disrupts IFN-alpha response by plasmacytoid dendritic cells (pDCs) and subsequent pDC/NK interplay. Immunol Lett 125:100–4.

35. Banchereau J, Steinman RM. 1998. Dendritic cells and the control of immunity. Nature 392:245–52.

36. Hladik F, McElrath MJ. 2008. Setting the stage: host invasion by HIV. Nat Rev Immunol 8:447–57.

37. Stieh DJ, Matias E, Xu H, Fought AJ, Blanchard JL, Marx PA, Veazey RS, Hope TJ. 2016. Th17 Cells Are Preferentially Infected Very Early after Vaginal Transmission of SIV in Macaques. Cell Host Microbe 19:529–40.

38. Hladik F, Sakchalathorn P, Ballweber L, Lentz G, Fialkow M, Eschenbach D, McElrath MJ. 2007. Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1. Immunity 26:257–70.

39. Hu J, Gardner MB, Miller CJ. 2000. Simian immunodeficiency virus rapidly penetrates the cervicovaginal mucosa after intravaginal inoculation and infects intraepithelial dendritic cells. J Virol 74:6087–95.

40. Spira AI, Marx PA, Patterson BK, Mahoney J, Koup RA, Wolinsky SM, Ho DD. 1996. Cellular targets of infection and route of viral dissemination after an intravaginal inoculation of simian immunodeficiency virus into rhesus macaques. J Exp Med 183:215–25.

41. de Silva S, Planelles V, Wu L. 2012. Differential effects of Vpr on single-cycle and spreading HIV-1 infections in CD4+ T-cells and dendritic cells. PLoS One 7:e35385.

42. Nobile C, Petit C, Moris A, Skrabal K, Abastado J-P, Mammano F, Schwartz O. 2005. Covert human immunodeficiency virus replication in dendritic cells and in DC-SIGN-expressing cells promotes long-term transmission to lymphocytes. J Virol 79:5386–99.

43. Burleigh L, Lozach P-Y, Schiffer C, Staropoli I, Pezo V, Porrot F, Canque B, Virelizier J-L, Arenzana-Seisdedos F, Amara A. 2006. Infection of dendritic cells (DCs), not DC-SIGN-mediated internalization of human immunodeficiency virus, is required for long-term transfer of virus to T cells. J Virol 80:2949–57.

44. Zhang X, Zhou T, Frabutt DA, Zheng Y-H. 2016. HIV-1 Vpr increases Env expression by preventing Env from endoplasmic reticulum-associated protein degradation (ERAD). Virology 496:194–202.

45. Chen M, Elder RT, Yu M, O'Gorman MG, Selig L, Benarous R, Yamamoto A, Zhao Y. 1999. Mutational analysis of Vpr-induced G2 arrest, nuclear localization, and cell death in fission yeast. J

Virol 73:3236–45.

46. Gummuluru S, Emerman M. 1999. Cell cycle- and Vpr-mediated regulation of human immunodeficiency virus type 1 expression in primary and transformed T-cell lines. J Virol 73:5422–30.

47. Belzile J-P, Abrahamyan LG, Gérard FC a, Rougeau N, Cohen E a. 2010. Formation of mobile chromatin-associated nuclear foci containing HIV-1 Vpr and VPRBP is critical for the induction of G2 cell cycle arrest. PLoS Pathog 6:e1001080.

48. Wen X, Duus KM, Friedrich TD, de Noronha CMC. 2007. The HIV1 protein Vpr acts to promote G2 cell cycle arrest by engaging a DDB1 and Cullin4A-containing ubiquitin ligase complex using VprBP/DCAF1 as an adaptor. J Biol Chem 282:27046–57.

49. Tan L, Ehrlich E, Yu X-F. 2007. DDB1 and Cul4A are required for human immunodeficiency virus type 1 Vpr-induced G2 arrest. J Virol 81:10822–30.

50. Jacquot G, Le Rouzic E, David A, Mazzolini J, Bouchet J, Bouaziz S, Niedergang F, Pancino G, Benichou S. 2007. Localization of HIV-1 Vpr to the nuclear envelope: impact on Vpr functions and virus replication in macrophages. Retrovirology 4:84.

51. Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Ségéral E, Yatim A, Emiliani S, Schwartz O, Benkirane M. 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature 474:654–7.

52. Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, Florens L, Washburn MP, Skowronski J. 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature 474:658–61.

53. Gringhuis SI, van der Vlist M, van den Berg LM, den Dunnen J, Litjens M, Geijtenbeek TBH. 2010. HIV-1 exploits innate signaling by TLR8 and DC-SIGN for productive infection of dendritic cells. Nat Immunol 11:419–426.

54. Turville SG, Santos JJ, Frank I, Cameron PU, Wilkinson J, Miranda-Saksena M, Dable J, Stössel H, Romani N, Piatak M, Lifson JD, Pope M, Cunningham AL. 2004. Immunodeficiency virus uptake, turnover, and 2-phase transfer in human dendritic cells. Blood 103:2170–2179.

55. Aggarwal A, Iemma TL, Shih I, Newsome TP, McAllery S, Cunningham AL, Turville SG. 2012. Mobilization of HIV spread by diaphanous 2 dependent filopodia in infected dendritic cells. PLoS Pathog 8:e1002762.

56. Su B, Biedma ME, Lederle A, Peressin M, Lambotin M, Proust A, Decoville T, Schmidt S, Laumond G, Moog C. 2014. Dendritic cell-lymphocyte cross talk downregulates host restriction factor SAMHD1 and stimulates HIV-1 replication in dendritic cells. J Virol 88:5109–21.

57. Holl V, Xu K, Peressin M, Lederle A, Biedma ME, Delaporte M, Decoville T, Schmidt S, Laumond G, Aubertin A-M, Moog C. 2010. Stimulation of HIV-1 replication in immature dendritic cells in contact with primary CD4 T or B lymphocytes. J Virol 84:4172–82.

58. Gummuluru S, KewalRamani VN, Emerman M. 2002. Dendritic cell-mediated viral transfer to T cells is required for human immunodeficiency virus type 1 persistence in the face of rapid cell turnover. J Virol 76:10692–701.

59. Pope M, Betjes MG, Romani N, Hirmand H, Cameron PU, Hoffman L, Gezelter S, Schuler G, Steinman RM. 1994. Conjugates of dendritic cells and memory T lymphocytes from skin facilitate

productive infection with HIV-1. Cell 78:389–98.

60. Cameron PU, Freudenthal PS, Barker JM, Gezelter S, Inaba K, Steinman RM. 1992. Dendritic cells exposed to human immunodeficiency virus type-1 transmit a vigorous cytopathic infection to CD4+ T cells. Science 257:383–7.

61. Gendelman HE, Baca LM, Turpin J, Kalter DC, Hansen B, Orenstein JM, Dieffenbach CW, Friedman RM, Meltzer MS. 1990. Regulation of HIV replication in infected monocytes by IFN-alpha. Mechanisms for viral restriction. J Immunol 145:2669–76.

62. Dianzani F, Castilletti C, Gentile M, Gelderblom HR, Frezza F, Capobianchi MR. 1998. Effects of IFN alpha on late stages of HIV-1 replication cycle. Biochimie 80:745–54.

63. Cheney KM, McKnight Á. 2010. Interferon-alpha mediates restriction of human immunodeficiency virus type-1 replication in primary human macrophages at an early stage of replication. PLoS One 5:e13521.

64. Shirazi Y, Pitha PM. 1992. Alpha interferon inhibits early stages of the human immunodeficiency virus type 1 replication cycle. J Virol 66:1321–8.

65. Baca-Regen L, Heinzinger N, Stevenson M, Gendelman HE. 1994. Alpha interferon-induced antiretroviral activities: restriction of viral nucleic acid synthesis and progeny virion production in human immunodeficiency virus type 1-infected monocytes. J Virol 68:7559–65.

66. Goujon C, Malim MH. 2010. Characterization of the alpha interferon-induced postentry block to HIV-1 infection in primary human macrophages and T cells. J Virol 84:9254–66.

67. Shirazi Y, Pitha PM. 1993. Interferon alpha-mediated inhibition of human immunodeficiency virus type 1 provirus synthesis in T-cells. Virology 193:303–12.

68. Hartshorn KL, Neumeyer D, Vogt MW, Schooley RT, Hirsch MS. 1987. Activity of interferons alpha, beta, and gamma against human immunodeficiency virus replication in vitro. AIDS Res Hum Retroviruses 3:125–33.

69. Macé K, Duc Dodon M, Gazzolo L. 1989. Restriction of HIV-1 replication in promonocytic cells: a role for IFN-alpha. Virology 168:399–405.

70. Larocque L, Bliu A, Xu R, Diress A, Wang J, Lin R, He R, Girard M, Li X. 2011. Bioactivity determination of native and variant forms of therapeutic interferons. J Biomed Biotechnol 2011:174615.

71. O'Doherty U, Swiggard WJ, Jeyakumar D, McGain D, Malim MH. 2002. A sensitive, quantitative assay for human immunodeficiency virus type 1 integration. J Virol 76:10942–50.

72. Agosto LM, Yu JJ, Dai J, Kaletsky R, Monie D, O'Doherty U. 2007. HIV-1 integrates into resting CD4+ T cells even at low inoculums as demonstrated with an improved assay for HIV-1 integration. Virology 368:60–72.

73. Forget J, Yao XJ, Mercier J, Cohen EA. 1998. Human immunodeficiency virus type 1 vpr protein transactivation function: mechanism and identification of domains involved. J Mol Biol 284:915–23.

74. Philippon V, Matsuda Z, Essex M. 1999. Transactivation is a conserved function among primate lentivirus Vpr proteins but is not shared by Vpx. J Hum Virol 2:167–174.

75. Varin A, Decrion A-Z, Sabbah E, Quivy V, Sire J, Van Lint C, Roques BP, Aggarwal BB, Herbein

G. 2005. Synthetic Vpr protein activates activator protein-1, c-Jun N-terminal kinase, and NF-kappaB and stimulates HIV-1 transcription in promonocytic cells and primary macrophages. J Biol Chem 280:42557–67.

76. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A 108:20166–71.

77. Emery A, Zhou S, Pollom E, Swanstrom R. 2017. Characterizing HIV-1 Splicing Using Next Generation Sequencing. J Virol JVI.02515-16.

78. Bartz SR, Rogel ME, Emerman M. 1996. Human immunodeficiency virus type 1 cell cycle control: Vpr is cytostatic and mediates G2 accumulation by a mechanism which differs from DNA damage checkpoint control. J Virol 70:2324–31.

79. Le Rouzic E, Belaïdouni N, Estrabaud E, Morel M, Rain J-C, Transy C, Margottin-Goguet F. 2007. HIV1 Vpr arrests the cell cycle by recruiting DCAF1/VprBP, a receptor of the Cul4-DDB1 ubiquitin ligase. Cell Cycle 6:182–8.

80. Selig L, Benichou S, Rogel ME, Wu LI, Vodicka MA, Sire J, Benarous R, Emerman M. 1997. Uracil DNA glycosylase specifically interacts with Vpr of both human immunodeficiency virus type 1 and simian immunodeficiency virus of sooty mangabeys, but binding does not correlate with cell cycle arrest. J Virol 71:4842–6.

81. Mansky LM, Preveral S, Selig L, Benarous R, Benichou S. 2000. The interaction of vpr with uracil DNA glycosylase modulates the human immunodeficiency virus type 1 In vivo mutation rate. J Virol 74:7039–47.

82. Schröfelbauer B, Yu Q, Zeitlin SG, Landau NR. 2005. Human immunodeficiency virus type 1 Vpr induces the degradation of the UNG and SMUG uracil-DNA glycosylases. J Virol 79:10978–87.

83. Zhu Y, Gelbard HA, Roshal M, Pursell S, Jamieson BD, Planelles V. 2001. Comparison of cell cycle arrest, transactivation, and apoptosis induced by the simian immunodeficiency virus SIVagm and human immunodeficiency virus type 1 vpr genes. J Virol 75:3791–801.

84. Mueller SM, Lang SM. 2002. The first HxRxG motif in simian immunodeficiency virus mac239 Vpr is crucial for G(2)/M cell cycle arrest. J Virol 76:11704–9.

85. Goujon C, Rivière L, Jarrosson-Wuilleme L, Bernaud J, Rigal D, Darlix J-L, Cimarelli A. 2007. SIVSM/HIV-2 Vpx proteins promote retroviral escape from a proteasome-dependent restriction pathway present in human dendritic cells. Retrovirology 4:2.

86. Ahn J, Vu T, Novince Z, Guerrero-Santoro J, Rapic-Otrin V, Gronenborn AM. 2010. HIV-1 Vpr loads uracil DNA glycosylase-2 onto DCAF1, a substrate recognition subunit of a cullin 4A-ring E3 ubiquitin ligase for proteasome-dependent degradation. J Biol Chem 285:37333–41.

87. Eldin P, Chazal N, Fenard D, Bernard E, Guichou J-F, Briant L. 2014. Vpr expression abolishes the capacity of HIV-1 infected cells to repair uracilated DNA. Nucleic Acids Res 42:1698–710.

88. Fregoso OI, Emerman M. 2016. Activation of the DNA Damage Response Is a Conserved Function of HIV-1 and HIV-2 Vpr That Is Independent of SLX4 Recruitment. MBio 7.

89. Berger G, Lawrence M, Hué S, Neil SJD. 2014. G2/M cell cycle arrest correlates with primate lentiviral Vpr interaction with the SLX4 complex. J Virol 89:230–40.

90. Daniel R, Kao G, Taganov K, Greger JG, Favorova O, Merkel G, Yen TJ, Katz RA, Skalka AM. 2003. Evidence that the retroviral DNA integration process triggers an ATR-dependent DNA damage response. Proc Natl Acad Sci U S A 100:4778–83.

91. Stracker TH, Carson CT, Weitzman MD. 2002. Adenovirus oncoproteins inactivate the Mre11–Rad50–NBS1 DNA repair complex. Nature 418:348–352.

92. Planelles V, Jowett JB, Li QX, Xie Y, Hahn B, Chen IS. 1996. Vpr-induced cell cycle arrest is conserved among primate lentiviruses. J Virol 70:2516–24.

93. Stivahtis GL, Soares M a, Vodicka M a, Hahn BH, Emerman M. 1997. Conservation and host specificity of Vpr-mediated cell cycle arrest suggest a fundamental role in primate lentivirus evolution and biology. J Virol 71:4331–8.

94. Nakai-Murakami C, Shimura M, Kinomoto M, Takizawa Y, Tokunaga K, Taguchi T, Hoshino S, Miyagawa K, Sata T, Kurumizaka H, Yuo A, Ishizaka Y. 2007. HIV-1 Vpr induces ATM-dependent cellular signal with enhanced homologous recombination. Oncogene 26:477–86.

95. Roshal M, Kim B, Zhu Y, Nghiem P, Planelles V. 2003. Activation of the ATR-mediated DNA damage response by the HIV-1 viral protein R. J Biol Chem 278:25879–86.

96. Nakagawa T, Lv L, Nakagawa M, Yu Y, Yu C, D'Alessio AC, Nakayama K, Fan H-Y, Chen X, Xiong Y. 2015. CRL4VprBP E3 Ligase Promotes Monoubiquitylation and Chromatin Binding of TET Dioxygenases. Mol Cell 57:247–260.

97. Ayyavoo V, Mahboubi A, Mahalingam S, Ramalingam R, Kudchodkar S, Williams W V., Green DR, Weiner DB. 1997. HIV-1 Vpr suppresses immune activation and apoptosis through regulation of nuclear factor κB. Nat Med 3:1117–23.

98. Kogan M, Deshmane S, Sawaya BE, Gracely EJ, Khalili K, Rappaport J. 2013. Inhibition of NF-κB activity by HIV-1 Vpr is dependent on Vpr binding protein. J Cell Physiol 228:781–90.

99. Hoshino S, Konishi M, Mori M, Shimura M, Nishitani C, Kuroki Y, Koyanagi Y, Kano S, Itabe H, Ishizaka Y. 2010. HIV-1 Vpr induces TLR4/MyD88-mediated IL-6 production and reactivates viral production from latency. J Leukoc Biol 87:1133–43.

100. Liu R, Tan J, Lin Y, Jia R, Yang W, Liang C, Geng Y, Qiao W. 2013. HIV-1 Vpr activates both canonical and noncanonical NF-κB pathway by enhancing the phosphorylation of IKKα/β. Virology 439:47–56.

101. Roux P, Alfieri C, Hrimech M, Eric A, Tanner JE, Cohen EA. 2000. Activation of Transcription Factors NF- κ B and NF-IL-6 by Human Immunodeficiency Virus Type 1 Protein R (Vpr) Induces Interleukin-8 Expression Activation of Transcription Factors NF- □ B and NF-IL-6 by Human Immunodeficiency Virus Type 1 Protein R (Vp.

102. Liu R, Lin Y, Jia R, Geng Y, Liang C, Tan J, Qiao W. 2014. HIV-1 Vpr stimulates NF-κB and AP-1 signaling by activating TAK1. Retrovirology 11:45.

103. Yan N, Lieberman J. 2012. SAMHD1 does it again, now in resting T cells. Nat Med 18:1611–2.

104. Baldauf H-M, Pan X, Erikson E, Schmidt S, Daddacha W, Burggraf M, Schenkova K, Ambiel I, Wabnitz G, Gramberg T, Panitz S, Flory E, Landau NR, Sertel S, Rutsch F, Lasitschka F, Kim B, König R, Fackler OT, Keppler OT. 2012. SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. Nat Med 18:1682–7.

105. Brownell J, Bruckner J, Wagoner J, Thomas E, Loo Y-M, Gale M, Liang TJ, Polyak SJ, Polyak SJ. 2014. Direct, interferon-independent activation of the CXCL10 promoter by NF-κB and interferon regulatory factor 3 during hepatitis C virus infection. J Virol 88:1582–90.

106. Yeruva S, Ramadori G, Raddatz D. 2008. NF-kappaB-dependent synergistic regulation of CXCL10 gene expression by IL-1beta and IFN-gamma in human intestinal epithelial cell lines. Int J Colorectal Dis 23:305–17.

107. Oslund KL, Zhou X, Lee B, Zhu L, Duong T, Shih R, Baumgarth N, Hung L-Y, Wu R, Chen Y. 2014. Synergistic Up-Regulation of CXCL10 by Virus and IFN γ in Human Airway Epithelial Cells. PLoS One 9:e100978.

108. Martin-Gayo E, Buzon MJ, Ouyang Z, Hickman T, Cronin J, Pimenova D, Walker BD, Lichterfeld M, Yu XG. 2015. Potent Cell-Intrinsic Immune Responses in Dendritic Cells Facilitate HIV-1-Specific T Cell Immunity in HIV-1 Elite Controllers. PLoS Pathog 11:e1004930.

109. Hamimi C, David A, Versmisse P, Weiss L, Bruel T, Zucman D, Appay V, Moris A, Ungeheuer M-N, Lascoux-Combe C, Barré-Sinoussi F, Muller-Trutwin M, Boufassa F, Lambotte O, Pancino G, Sáez-Cirión A, ANRS CO21 CODEX cohort ACC. 2016. Dendritic Cells from HIV Controllers Have Low Susceptibility to HIV-1 Infection In Vitro but High Capacity to Capture HIV-1 Particles. PLoS One 11:e0160251.

110. Yamashita M, Emerman M. 2004. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. J Virol 78:5670–8.

111. Voronin Y, Holte S, Overbaugh J, Emerman M. 2009. Genetic drift of HIV populations in culture. PLoS Genet 5:e1000431.

112. Puryear WB, Akiyama H, Geer SD, Ramirez NP, Yu X, Reinhard BM, Gummuluru S. 2013. Interferon-inducible mechanism of dendritic cell-mediated HIV-1 dissemination is dependent on Siglec-1/CD169. PLoS Pathog 9:e1003291.

113. Puryear WB, Gummuluru S. 2013. Role of glycosphingolipids in dendritic cell-mediated HIV-1 trans-infection. Adv Exp Med Biol 762:131–53.

114. Vodicka MA, Goh WC, Wu LI, Rogel ME, Bartz SR, Schweickart VL, Raport CJ, Emerman M. 1997. Indicator cell lines for detection of primary strains of human and simian immunodeficiency viruses. Virology 233:193–8.

115. Puryear WB, Yu X, Ramirez NP, Reinhard BM, Gummuluru S. 2012. HIV-1 incorporation of host-cell-derived glycosphingolipid GM3 allows for capture by mature dendritic cells. Proc Natl Acad Sci U S A 109:7475–80.

116. Akiyama H, Miller C, Patel H V, Hatch SC, Archer J, Ramirez N-GP, Gummuluru S. 2014. Virus particle release from glycosphingolipid-enriched microdomains is essential for dendritic cell-mediated capture and transfer of HIV-1 and henipavirus. J Virol 88:8813–25.

117. Hatch SC, Archer J, Gummuluru S. 2009. Glycosphingolipid composition of human immunodeficiency virus type 1 (HIV-1) particles is a crucial determinant for dendritic cell-mediated HIV-1 trans-infection. J Virol 83:3496–506.

118. Akiyama H, Ramirez N-GP, Gudheti M V, Gummuluru S. 2015. CD169-Mediated Trafficking of HIV to Plasma Membrane Invaginations in Dendritic Cells Attenuates Efficacy of Anti-gp120 Broadly Neutralizing Antibodies. PLoS Pathog 11:e1004751.