
UNIVERSITY OF PÉCS

Doctoral School of Biology and Sport Biology

**Insights into Mason-Pfizer monkey virus (MPMV) genomic
RNA packaging: Identification of minimal structural motifs
involved during its selective encapsidation**

PhD thesis

Lizna Mohamed Ali

PÉCS, 2021

UNIVERSITY OF PÉCS

Doctoral School of Biology and Sport Biology

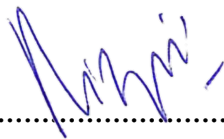
Insights into Mason-Pfizer monkey virus (MPMV) genomic RNA packaging: Identification of minimal structural motifs involved during its selective encapsidation

PhD thesis

Lizna Mohamed Ali

Supervisor:

Tahir A Rizvi
PhD



.....
Supervisor's signature

Professor Tahir A. Rizvi



.....
Head of Doctoral school's signature

Professor Robert Gabriel

PÉCS, 2021

Declaration of Original Work

I, Lizna Mohamed Ali, the undersigned, a graduate student at the University Of Pécs, and the author of this dissertation entitled “*Insights into Mason-Pfizer Monkey Virus (MPMV) Genomic RNA Packaging: Identification of Minimal Structural Motifs Involved During its Selective Encapsidation*”, hereby, solemnly declare that this dissertation is my own original research work that has been done and prepared by me under the supervision of Professor Tahir Rizvi, in the College of Medicine and Health sciences at UAEU and Professor Ferenc Jakab, in the Szentágothai Research Centre at the University Of Pécs. This work has not previously been presented or published or formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my dissertation have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this dissertation.

Student’s Signature: _____



Date: 24.08.2021

Copyright © <2021> <Lizna Mohamed Ali>
All Rights Reserved

Abstract

A distinguishing feature of the Mason-Pfizer monkey virus (MPMV) packaging signal RNA secondary structure is a single-stranded purine-rich sequence (ssPurines; U¹⁹¹UAAAAGUGAAAGUAA²⁰⁶) in close vicinity to a palindromic stem loop (Pal SL) that functions as MPMV dimerization initiation site (DIS). However, unlike other retroviruses, MPMV contains a partially base-paired repeat sequence of ssPurines (bpPurines) in the adjacent region. Both purine-rich sequences have earlier been proposed to act as potentially redundant Gag binding sites to initiate the process of MPMV genomic RNA (gRNA) packaging. The objective of this study was to investigate the biological significance of ssPurines and bpPurines in MPMV gRNA packaging by systematic mutational analysis using an *in vivo* packaging and propagation assay combined with biochemical RNA structure probing analysis. Deletion of either ssPurines or bpPurines individually had no significant effect on MPMV gRNA packaging, but it was severely compromised when both sequences were deleted simultaneously. Selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) analysis of the mutant RNAs revealed only mild effects on structure by deletion of either ssPurines or bpPurines, while the structure was dramatically affected by the two simultaneous deletions. This suggests that ssPurines and bpPurines play a redundant role in MPMV gRNA packaging process. Additionally, this study also investigates the role of a single stranded GU rich region, (G²⁵⁴UGUU²⁵⁸; present immediately downstream of bpPurines) recently identified to function as a redundant Gag binding site along with the ssPurine loop. Biological data presented here confirms the role of both these regions (ssPurine and the GU-rich region) functioning as redundant *cis* acting elements required for MPMV gRNA packaging. Finally, the deletion of bpPurines revealed no significant effect on MPMV gRNA packaging. However, this deletion indicated a severe defect on RNA propagation that was independent of the presence or absence of ssPurines or the gRNA structure of the region. These findings further suggest that the bpPurines play a vital role in the early steps of MPMV replication cycle that is yet to be identified.

Keywords: Mason-Pfizer monkey virus, RNA packaging, RNA secondary structure, single-stranded purines, base paired purines, RNA-Gag interaction, retroviruses, SHAPE.

Összefoglaló

A Mason-Pfizer majom vírus (MPMV) becsomagolást elindító másodlagos RNS szerkezet jellegzetesége az MPMV dimerizáció kezdő helyeként (DIS) szolgáló, palindromikus hajcsat hurok (palindromic stem loop (Pal SL)) közelében elhelyezkedő egyszálú, purinban gazdag (ssPurinok; U191UAAAAGUGAAAGUAA206) szekvencia. Azonban, egyéb retrovírusoktól eltérően, a szomszédos régióban az MPMV rendelkezik egy további szekvenciával is, ami csak részlegesen bázis-párosított (bpPurines) ismétlődő ssPurinokból áll. Korábban mindkét purinban gazdag szekvenciával kapcsolatban felmerült, hogy potenciálisan redundáns Gag kötőhelyként szolgálnak az MPMV genomiális RNS (gRNS) becsomagolásának megkezdésében. Munkánk célja az volt, hogy *in vivo* becsomagolási és propagálási tesztekkel biokémiai RNS szerkezet analízissel kombinálva megvizsgáljuk az ssPurinoknak és bpPurinoknak az MPMV gRNS becsomagolásában játszott szerepét. Az ssPurinok, vagy a bpPurinok egyenkénti deléciója nem volt jelentős hatással a MPMV gRNS becsomagolására, de mindkét szekvencia deléciója súlyosan károsította azt. Az ssPurinok vagy bpPurinok deléciója során a mutáns RNS-ek primer extenziós analízissel (SHAPE) vizsgált szelektív 2' hidroxil-acilálási analízise csak mérsékelt szerkezeti változásokat tárt fel az RNS szerkezetében, azonban, amennyiben mindkét struktúra deléciója létrejött, a hatás jelentős volt. Mindez azt sugallja, hogy az ssPurinok és bpPurinok redundáns szerepet játszanak az MPMV gRNS becsomagolásában. Ezen túlmenően vizsgáltuk egy egyszálú, a bpPurinokat közvetlenül követő, GU-ban gazdag régió (G 254UGUU258) biológiai szerepét is, melyről a közelmúltban derül ki, hogy a ssPurin hurok mellett egy redundáns Gag kötőhelyként működik. Eredményeink megerősítik, hogy mindkét régió (az ssPurin és a GU-ban gazdag régió) cisz-hatású elemként szükséges az MPMV gRNS becsomagolásához. Végezetül, a bpPurinok deléciója, bár nem gyakorolt szignifikáns hatást az MPMV becsomagolására, de az RNS propagációjában jelentős defektust eredményezett, mely független volt az ssPurinok hiányától vagy meglététől, illetve a gRNS másodlagos helyi szerkezetétől. Mindezek az adatok felvetik, hogy a bpPurinok fontos szerepet játszanak az MPMV replikáció még feltáratlan korai lépései során.

Acknowledgements

I would begin this special section with utmost gratitude to the Almighty for presenting me with this wonderful opportunity to advance my career and attain my life goals.

I am forever grateful to my supervisor, Professor Tahir A. Rizvi, Professor of Molecular Virology, for his persistent support and guidance throughout my term in his lab where I started my career as a research staff. It is my honor and privilege to become a student under his mentorship which molded me into a better person both personally and professionally.

Words are not enough to thank Dr Farah Mustafa, Associate Professor of Biochemistry, who has been always there for me to drive my project forward whenever I encountered an obstacle. Her immense knowledge and patience with her students make her an idol for any student to emulate. She has been my 'SOS' system whose timely intervention revived me to move ahead until I see the results.

A unique space in my PhD journey is occupied by my lab colleagues, who have been wonderful friends with their endless support, understanding and assistance whenever needed. I will cherish all the witty times and the intellectual discussions we had and how we converted all the worries into hopes and happiness. My heartfelt appreciation goes to our current team: Ms Vineeta N. Pillai, Dr Akhil Chameettachal, Dr Fathima Nuzra Nagoor Pitchai, Ms Anjana Krishan and Mr Suresha G. Prabhu, and my former lab members: Ms Pretty Philip, Dr Akela Ghazawi, Ms Aayesha Jabeen, Dr Suriya Aktar, Ms Rawan Kalloush, Dr Jimsheena V.K and Dr Jaleel Kizhakkayil.

I also impart my gratefulness to Dr Tibor Pal, Professor of Microbiology, for introducing me to the University of Pécs as a prospective graduate student. Much gratitude and thanks to Dr Robert Gabriel, Dean and Head of Doctoral School for accepting me into the PhD program and guiding me throughout my study. My special thanks to Dr Ferenc Jakab, Professor of Virology, my internal consultant, for all the assistance and hospitality he provided at the time of my interview and thereafter.

I am forever indebted to my Mom and Dad for raising me as a person with values, inspiring me to dream high and never give up and, for their infinite love and care. I want to thank my brother Fawaz, my all-time supporter and companion in all stages of

my life and my little sister, Fahiya, as my motivator. I cannot exclude my in-laws from the list of my well-wishers, and I really appreciate their support provided during this trying period.

Just as importantly, I express my eternal love and thankfulness to my dear husband, Jali, for being the pillar of my life; his immense patience and love was my strength to face any challenge and attain my aspirations. Last but not the least, I want to abundantly thank my lovely boys, Zehal, Zayan and Zidan, who cheered me up during my low spirits and for their precocious ‘maturity in understanding’ and helping their mom accomplish her dream.

I could not have had a better life without you all, Thank you!

Dedication

*I dedicate this thesis to my parents,
To my husband
&
My boys*

Table of Contents

Title	i
Declaration of Original Work	iii
Copyright	iv
Abstract	v
Összefoglaló.....	vi
Acknowledgements	vii
Dedication	ix
Table of Contents	x
List of Tables.....	xii
List of Figures	xiii
List of Abbreviations.....	xvi
CHAPTER 1: INTRODUCTION	1
1.1 Overview	1
1.2 Viruses as Gene Therapy Vectors	1
1.3 Retroviruses as Gene Therapy Tool.....	4
1.4 Understanding Retroviruses	5
1.5 An Overview of Retroviral Life Cycle.....	12
1.6 The MPMV Packaging Determinants for gRNA Packaging	48
1.7 Objectives.....	54
CHAPTER 2: MATERIALS AND METHODS	56
2.1 MPMV Strain, Nucleotide Designations and Plasmid Construction	56
2.2 Cell Culture, Transfection, and Infection.....	57
2.3 RNA Extraction and Reverse-Transcriptase Polymerase Chain Reaction (RT PCR)	58

2.4 Determination of the Relative Packaging Efficiencies of Transfer Vector by Real-Time Quantitative PCR (RT-qPCR).....	59
2.5 Statistical Analysis	59
2.6 <i>In Silico</i> Analysis of MPMV RNA Secondary Structure.....	60
2.7 <i>In vitro</i> RNA Transcription.....	60
2.9 <i>In vitro</i> RNA Dimerization Assays	62
CHAPTER 3: RESULTS	63
3.1 Experimental Approach to Analyze the Effects of ssPurines and bpPurines Mutants on RNA Packaging and Propagation	63
3.2 Role of ssPurines in gRNA Packaging and Propagation	69
3.3 Role of bpPurines in gRNA Packaging and Propagation.....	69
3.4 Structure-Function Analysis of ssPurines Mutants	73
3.5 Structure-Function Analysis of bpPurines Mutants	78
3.6 <i>In vitro</i> Dimerization Capability of the Mutant RNAs	82
3.7 Role of a GU-rich Sequences of Region B in gRNA Packaging and Propagation	87
CHAPTER 4: DISCUSSION.....	89
CHAPTER 5: CONCLUSIONS	94
REFERENCES.....	95
LIST OF PUBLICATIONS	118
APPENDIX A	120
APPENDIX B	124

List of Tables

Table 1. Viruses that are currently applied in gene therapy trials.....	2
Table 2. Retroviruses and its receptors	14
Table 3. Viral and cellular factors affecting retroviral gene expression	35
Table 4. Description of the mutations introduced in ssPurines and bpPurines	65

List of Figures

Figure 1. Structure of retroviral particle	7
Figure 2. Phylogenetic classification of retroviruses	8
Figure 3. Genome organization of retroviruses.....	10
Figure 4. Steps in the retroviral life cycle	12
Figure 5. The HIV entry process as an example for receptor-mediated endocytosis	15
Figure 6. Receptor-mediated signaling pathways for endocytosis.....	16
Figure 7. Delineating various steps involved in reverse transcription.....	18
Figure 8. Schematic representation of viral gRNA and proviral DNA after reverse transcription.....	19
Figure 9. Major events occurring during nuclear import and export of HIV-1	21
Figure 10. Steps in the synthesis of retroviral DNA and its integration into host DNA	22
Figure 11. Overview of retroviral transcription and RNA processing.....	24
Figure 12. Properties of selected retrovirus transcripts.....	25
Figure 13. Regulation of primary transcript in simple retroviruses	29
Figure 14. Regulation of primary transcript in complex retroviruses	30
Figure 15. Organization and features of the 5' untranslated regions of the major HIV-1 transcripts	32
Figure 16. Ribosomal frameshifting in the viral protein synthesis. Eg: HIV-1	34
Figure 17. Proposed mechanism for HIV-1 in packaging and translation.....	37
Figure 18. HIV-1 assembly, budding, and maturation.....	41
Figure 19. Positioning of the primary packaging determinants of various retroviruses	43
Figure 20. Genetic organization of the 5'-untranslated region (UTR) of the retroviral genomic RNA- example of SIVmac	44
Figure 21. Retroviral RNA dimerization and packaging	47

Figure 22. Minimal free-energy and SHAPE (selective 2' hydroxyl acylation analyzed by primer extension)-validated models of the MPMV packaging signal RNA	51
Figure 23. Role of ssPurines in MPMV genomic RNA packaging	53
Figure 24. Illustration of the vectors used in the <i>in vivo</i> packaging and <i>in vitro</i> transcription assays and SHAPE-validated structure of the MPMV packaging signal RNA	64
Figure 25. Illustration of the 3-plasmid <i>in vivo</i> packaging and propagation assay	66
Figure 26. RT-PCR of viral and cytoplasmic cDNA fractions with appropriate controls.....	67
Figure 27. β -actin mRNA is packaged into viral particles at the same levels irrespective of the amount of viral RNA present in the virions	68
Figure 28. RT-qPCR and virus titer analysis of ssPurines and bpPurines mutants to establish their role in MPMV RNA packaging and propagation	72
Figure 29. Mfold structural predictions for the wild type (SJ2) and mutant packaging signal RNAs	74
Figure 30. Selective 2' hydroxyl acylation analyzed by primer extension-validated structures of the wild-type and mutant LA-I/FN-I, LA-II/FN-II, and LA-III/FN-III packaging signal RNAs.....	77
Figure 31. Selective 2' hydroxyl acylation analyzed by primer extension-validated structures of the wild-type and mutant LA-IV/FN-IV and LA-V/FN-V packaging signal RNAs	79
Figure 32. Selective 2' hydroxyl acylation analyzed by primer extension-validated structures of the wild-type and mutant LA-VI/FN-VI and LA-VII/FN-VII packaging signal RNAs	81
Figure 33. <i>In vitro</i> RNA dimerization assay on the wild-type and mutant SHAPE-interrogated packaging signal RNAs	83
Figure 34. RT-qPCR and virus titer analysis of bpPurines and its complementary region mutants to establish their role in MPMV RNA packaging and propagation ..	85
Figure 35. Selective 2' hydroxyl acylation analyzed by primer extension-validated structures of the wild-type and mutant LA-VIII/FN-VIII and LA-IX/FN-IX packaging signal RNAs.....	86

Figure 36. RT-qPCR and virus titer analysis of GU-rich region mutants to establish their role in MPMV RNA packaging and propagation 88

Figure 37. Structural model of MPMV packaging signal RNA containing two purine-rich sequences (ssPurines and bpPurines) which play a redundant role during MPMV gRNA packaging 90

List of Abbreviations

AS	Antisense
bpPurines	Base-paired Purines
BzCN	Benzoyl Cyanide
CA	Capsid
cDNA	Complementary Deoxyribonucleic acid
CREB/ATF	cAMP Response Element Binding Protein/Activating Transcription Factors
CRM1	Chromosome Region Maintenance
CTE	Constitutive Transport Element
DIS	Dimerization Initiation Site
DEAE	Diethylaminoethyl
DNA	Deoxyribonucleic acid
Env/ <i>env</i>	Envelope
Gag	Group Antigen
gRNA	Genomic Ribonucleic Acid
hSHAPE	High-Throughput Selective 2'Hydroxyl Acylation by Primer Extension
HIV	Human Immunodeficiency Virus
LRI	Long Range Interaction
LTR	Long Terminal Repeat
MA	Matrix
MMTV	Mouse Mammary Tumour Virus
MPMV	Mason-Pfizer Monkey Virus
mRNA	Messenger Ribonucleic acid

mSD	Major Splice Donor
NC	Nucleocapsid
PCR	Polymerase Chain Reaction
<i>Psi</i>	Packaging Signal
PBS	Primer Binding Site
RNA	Ribonucleic Acid
S	Sense
SEAP	Secreted Alkaline Phosphatase
SL	Stem Loop
SOE	Splice Overlap Extension
ssPurines	Single Stranded Purines
UTR	Untranslated Region
U5	5' Unique
U3	3' Unique
WT	Wild Type

CHAPTER 1: INTRODUCTION

1.1 Overview

Over the years increased understanding of molecular biology have opened avenues for using vectors from several viruses to deliver the therapeutic genes in order to treat various gene disorders. These disorders include metabolic, cardiovascular, muscular, hematologic, ophthalmologic, infectious diseases, and different types of cancer. Nearly all human diseases have a genetic basis; some inherited from the parents, for example, sickle cell anemia, Down's syndrome, and other rare diseases, while others such as cancers are caused due to acquired genetic alterations (Genetic Disorders, n.d.). Conventionally, genetic disorders were assumed to be not amenable to any treatment or cure until the advent of gene therapy. Gene therapy is a therapeutic intervention where the "faulty" or the "missing" gene is replaced by the functional copy of the gene to obviate the disease complications. The success of gene therapy resides on many factors such as knowing the correct pathogenesis of the disease, an effective gene(s) and efficient delivery system targeting the specific cells or tissues, and an animal model mimicking the disorder for preclinical trials. In recent years, viruses have been utilized as effective vectors of the corrective nucleic acids. Significant innovations in vector engineering and safe delivery systems have rendered viral vector-based therapy at the forefront in the modern-day medicine.

1.2 Viruses as Gene Therapy Vectors

Viruses owing to their inherent property of highly efficient nucleic acid delivery to the target cells while evading the immune system have been extensively utilized in human gene therapy protocols (Robbins & Ghivizzani, 1998). Different types of viruses, such as retrovirus, adenovirus, adeno-associated virus (AAV), and herpes simplex virus (HSV), have been manipulated in the laboratory for use in gene therapy applications. Each of these vector systems has its advantages and disadvantages, and considerable developments in vector designing have proved its suitability for treating a specific disease. Until 2017, it has been estimated that nearly 70% of the clinical trials for human gene therapy are based on utilizing viral vectors (reviewed in Beitelshes et al., 2017; Lundstrom, 2018). Some of the main groups of viruses currently used in gene therapy applications are reviewed in the table1, details retrieved from Lundstrom, 2018.

Table 1. Viruses that are currently applied in gene therapy trials

Virus vector	Genome & Insert size	Features	Disease
Adenoviruses Ad5	dsDNA <7.5 kb	broad host range transient expression strong immunogenicity	pancreatic CA neuroendocrine CA breast CA glioma
Adeno-associated virus AAV2, 3, 5, 6, 8, 9	ssDNA <4 kb	relatively broad host range slow expression onset chromosomal integration immune response	Hemophilia A RTT HD DMD CT
Herpes simplex virus HSV1, HSV	dsDNA >30 kb	broad host range latent infection long-term expression	Glioblastoma PDN SN Cancer
Retroviruses MMSV, MSCV	ssRNA 8 kb	transduces only dividing cells long-term expression random integration	X-CGD HGG
Lentiviruses HIV-1, HIV-2	ssRNA 8 kb	transduces dividing and non-dividing cells broad host range low cytotoxicity integration long-term expression	Hemophilia A Hemophilia B CF PKD PD AD HIV SHIV
Alphaviruses SFV, SIN, VEE, M1	ssRNA 8 kb	broad host range extreme transient expression low immunogenicity neuron- and glial-specific mutants	Prostate CA Lung CA Glioma Liver CA
Flaviviruses Kunjin, West Nile, Dengue virus	ssRNA 8 kb	relatively broad host range transient expression packaging system	Colon CA
Rhabdoviruses Rabies, VSV	ssRNA 6 kb	relatively broad host range high transient expression low immunogenicity	Vaccines against HIV-1/SIV Brain tumour
Measles virus MV-Edm	ssRNA 6 kb	transient expression oncolytic strains	ATC Ovarian CA HCC
Newcastle disease Virus	ssRNA 6 kb	replication in tumor cells improved oncolytic vectors	Prostate CA Colorectal CA Melanoma Solid tumors
Poxviruses VV	dsDNA >30 kb	broad host range large inserts replication-competent vectors	Colorectal CA Glioblastoma Pancreatic CA HCC Prostrate CA
Picornaviruses Coxsackievirus	ssRNA 6 kb	oncolytic strains	Melanoma IN Breast CA Prostrate CA

AAV, adeno-associated virus; HIV, human immunodeficiency virus; HSV, herpes simplex virus; MMSV, Moloney murine sarcoma virus; MSCV, murine stem cell virus; SFV, Semliki Forest virus; SIN, Sindbis virus; VEE, Venezuelan equine encephalitis virus; VSV, vesicular stomatitis virus; VV, vaccinia virus; dsDNA, double stranded DNA; ssDNA, single stranded DNA; CA, carcinoma; RTT, Rett's syndrome; HD, Huntington's disease; DMD, Duchenne muscular dystrophy; CF, Cystic fibrosis; PDN, Painful diabetic neuropathy, SN, Sensory neuropathy; X-CGD, X-linked granulomatous disease; HGG, High-grade glioma; PKD, Pyruvate kinase deficiency; PD, Parkinson's disease; AD, Alzheimer's disease; SHIV, Simian/Human Immunodeficiency virus; ATC, Anaplastic thyroid cancer; HCC, Hepatocellular carcinoma; IN, Ischemic necrosis. Details in the table from Lundstrom, 2018.

1.3 Retroviruses as Gene Therapy Tool

Retroviruses were the first group of viruses utilized for the classical gene therapy trials as in the case of severe combined immune deficiencies (SCIDs) (Anderson, 1990; Hacein-Bey-Abina et al., 2002). Retroviruses are well known for their ability to accommodate the gene of interest of ~8kb in length which can be integrated into the host genome along with retroviral vectors making it part and parcel of the host genome hence contributing to its long-term expression. Most retroviruses infect dividing cells however certain retroviruses such as the lentiviruses (HIV-1 and FIV) and betaretrovirus (MMTV) have the additional advantage of infecting non-dividing cells with low cytotoxicity (Bukrinsky et al., 1993; Kenyon & Lever, 2011; Konstantoulas & Indik, 2014; Poeschla et al., 1998; Kay et al., 2001; Vigna & Naldini, 2000). Thus, retroviruses are the vector of choice for gene delivery, and it is not surprising that approximately a quarter of all human gene therapy trials have used retroviral vectors (reviewed in Ginn et al., 2018; Hu et al., 2011).

Although the first clinical trials for the treatment of SCIDs have been successful in the patients, the development of T-cell leukemia due to insertional mutagenesis in two out of ten cases, later raised the safety concern for its application in the disease treatment (Gaspar & Thrasher, 2005; Hacein-Bey-Abina et al., 2008; McCormack & Rabbitts, 2004). This shortcoming has called for the development of safer vectors with directed chromosomal integration. Another caveat in using retroviral vectors is the possibility of generating replication competent viruses through recombination with endogenous retroviral genome in the host that could lead to inadvertent effects in the patients. This could be negated by using vectors from phylogenetically distant non-human retroviruses that are also self-inactivating. For any virus to be used as a carrier for a therapeutic gene, the sequences that lead to pathogenicity should be avoided while retaining the region required for efficient gene transfer along with its persistent expression. Therefore, it is important to study the retroviral lifecycle for a comprehensive understanding of the necessary elements involved in replication to circumvent the limitations of the current vectors used.

The thesis work being presented here is a step towards defining the packaging events of a non-human phylogenetically distant, monkey (simian) retrovirus, Mason-Pfizer monkey virus (MPMV) that causes immunodeficiency syndrome (Bohl et al.,

2005) and tumors (De las Heras et al., 1991) in Old World monkeys, family *Cercopithecidae* (Fine & Schochetman, 1978). MPMV is a type D *Betaretrovirus*, evolved as a simian retrovirus (SRV) and is different from the simian immunodeficiency virus (SIV) (Kestler et al., 1990), despite the similarity in the disease condition caused by these viruses. Type D viruses, infect a variety of mammalian hosts, however no sequence homology was observed between the genome of MPMV and of primates such as the new world monkeys, apes including humans suggesting that cellular DNAs of these primates are free from endogenous MPMV (Fine & Schochetman, 1978) and could possibly be safe to be used as a gene therapy tool.

1.4 Understanding Retroviruses

Retroviruses belong to the family *Retroviridae* and are taxonomically classified according to their structure, composition, and replicative properties (Coffin et al., 1997a). Retroviral virions generally vary in size from 80-100 nm in diameter which are enveloped with the lipid bilayer membrane derived from the host cell harboring virally encoded envelope glycoproteins and include surface (SU) and the transmembrane (TM) components. The genome of retrovirus is a linear, single-stranded (ss), non-segmented, 7-12 kb RNA of positive polarity. The RNA genome is enclosed as dimers within the protein core encoded by the *gag* gene and forms the internal structural components of the virions. Figure 1A below schematically depicts the structural organization of a retroviral particle. The shape and location of the protein core determined the conventional grouping of these viruses into four main groups namely A, B, C, and D-type retroviruses. Based on the ultrastructural appearances as shown in figure 1B, retroviral particles can be characterized based on the shape of the core. For example, type A virions form an intracisternal double membrane particle which lack budding, type B virions are characterized by an eccentric, spherical core; type C virions are characterized by a central, spherical core; and type D virions contain a cylindrical core.

A

Env proteins

Surface protein (SU)

Transmembrane protein (TM)

Membrane

Gag proteins

Matrix protein (MA)

Capsid protein (CA)

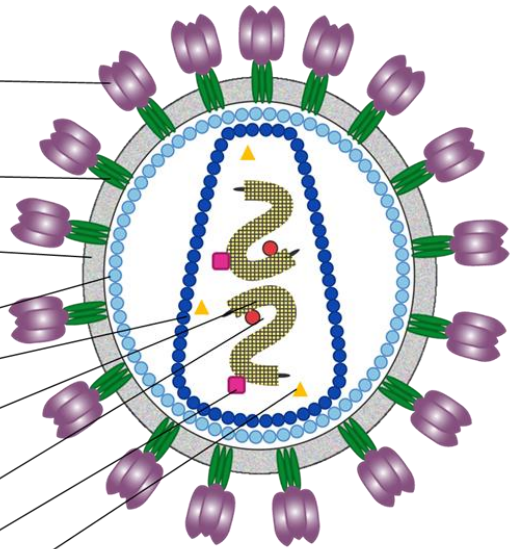
Nucleocapsid (NC)

Pol proteins

Reverse transcriptase (RT)

Integrase (IN)

Protease (PR)



B

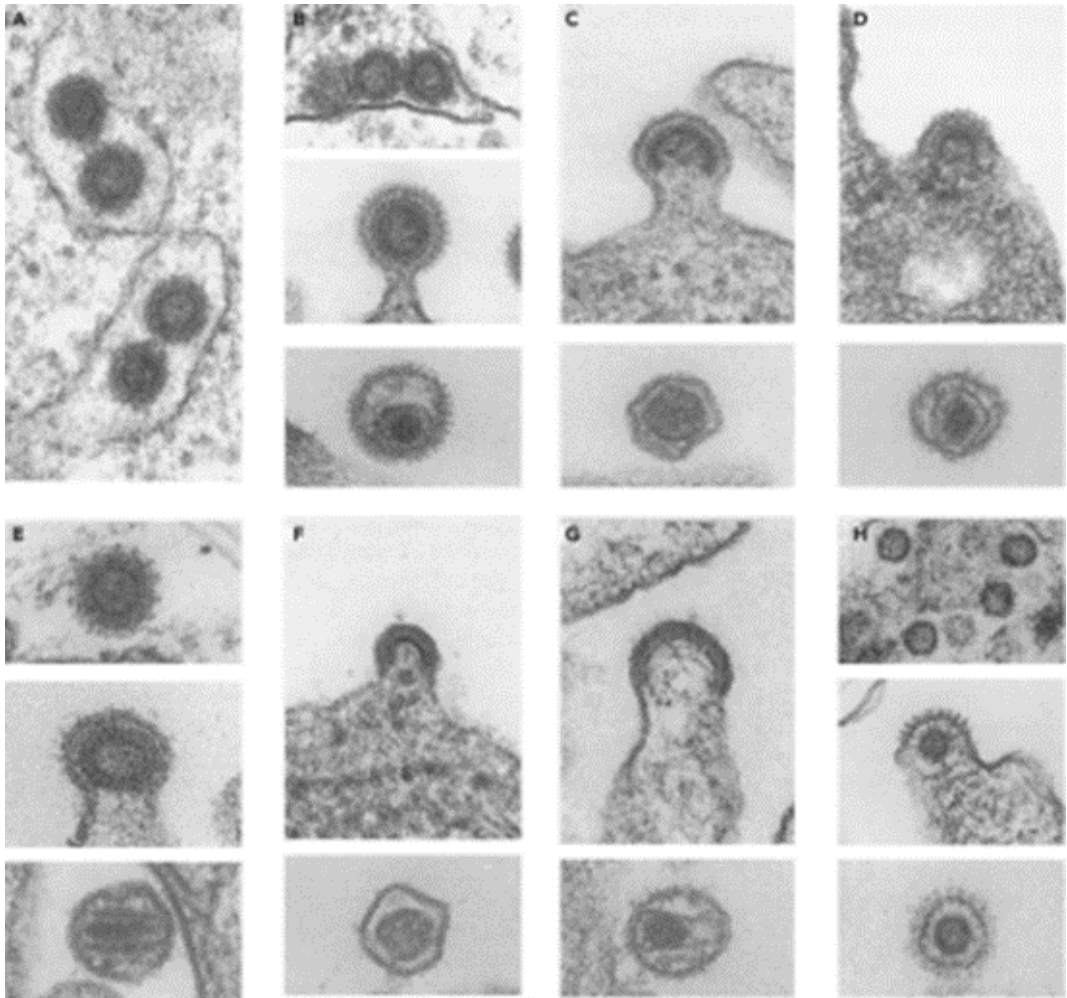


Figure 1. Structure of retroviral particle

(A) Schematic representation of retroviral particle. The gag domain encodes the structural components of the viral core the capsid (CA), matrix (MA) and nucleocapsid (NC) proteins. The pro domain encodes the viral protease (PR), while pol encodes the reverse transcriptase (RT) and integrase (IN) enzymes. The env domain encodes the surface (SU) and transmembrane (TM) glycoproteins responsible for viral attachment and penetration.

(B) Electron micrographs of representative virion particles. The diameters of all the particles are approximately 100 nm. **A:** Type A particles. Intracisternal A particles in the endoplasmic reticulum. **B:** Betaretrovirus. Mouse mammary tumor virus, MMTV; type B morphology (top, intracytoplasmic particles; middle, budding particles; bottom, mature extracellular particles). **C:** Gammaretrovirus. Murine leukemia virus, MLV; type C morphology (top, budding; bottom, mature extracellular particles). **D:** Alpharetrovirus. Avian leukosis virus; type C morphology (top, budding; bottom, mature extracellular particles). **E:** Betaretrovirus. Mason-Pfizer monkey virus, MPMV; type D morphology (top, intracytoplasmic A-type particles; middle, budding; bottom, mature extracellular particles). **F:** Deltaretrovirus. Bovine leukemia virus, BLV (top, budding; bottom, mature extracellular particles). **G:** Lentivirus. Bovine immunodeficiency virus (top, budding; bottom, mature extracellular particles). **H:** Spumavirus. Bovine syncytial virus (top, intracytoplasmic particles; middle, budding; bottom, mature extracellular particles). (Micrographs courtesy of Dr. Matthew Gonda, and reproduced from Coffin et al., 1997a.

Currently, the International Committee on Taxonomy of Viruses (ICTV) has classified retroviruses based on their 1) phylogeny and 2) genome organization into two subfamilies: orthoretroviruses (orthoretroviranae) and spumaviruses (spumaretroviranae) (MacLachlan & Dubovi, 2017). Each of these are then classified into seven genera based on their evolutionary relatedness (Figure 2). Five of these groups, except for lentiviruses and spumaviruses are associated with the development of cancers in the infected animals, whereas lentiviruses cause immunodeficiency and neurological symptoms. The spumaviruses contain only one genus, the foamy virus, which is not related to any disease, but it differs from other retroviruses since its RNA genome is already reverse transcribed in the virions before infecting the host cell (Yu et al., 1999).

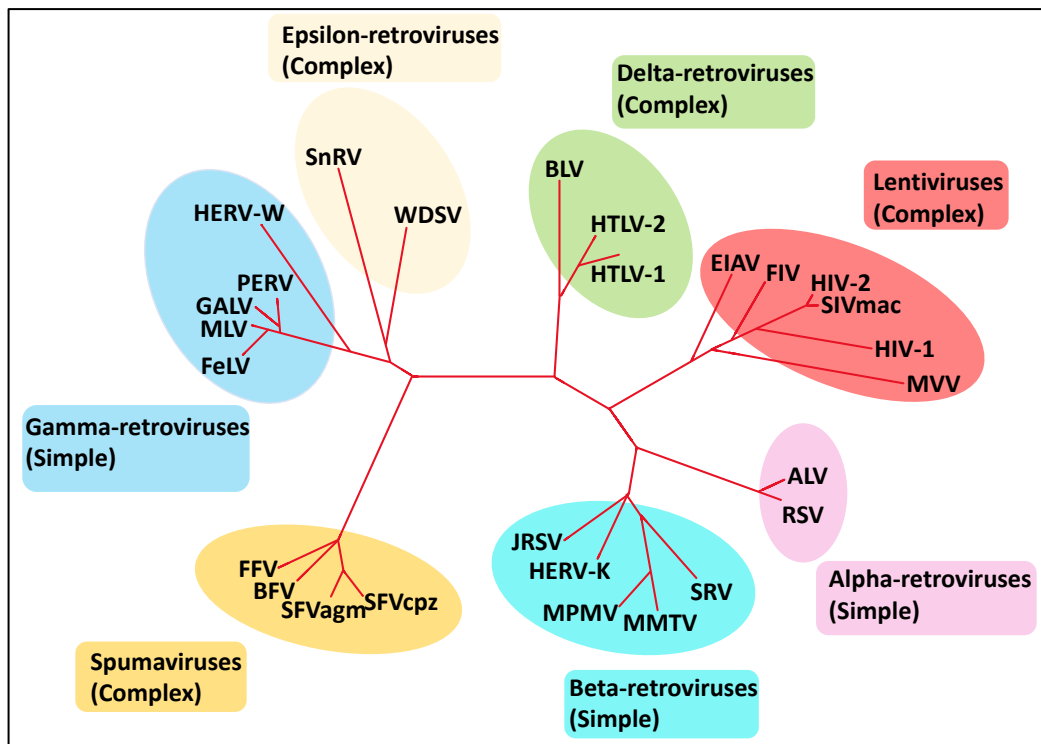


Figure 2. Phylogenetic classification of retroviruses

ALV, Avian leukosis virus; RSV, Rous sarcoma virus; SRV, simian retrovirus; MMTV, mouse mammary tumor virus; MPMV, Mason-Pfizer monkey virus; HERV-K, human endogenous retrovirus type K; JRSV, Jaagsiekte sheep retrovirus; FFV, feline foamy virus; BFV, bovine foamy virus; SFVagm, simian foamy virus- African green monkeys; SFVcpz, simian foamy virus- Chimpanzee; FeLV, feline leukemia virus; MLV, murine leukemia virus; GALV, gibbon-ape leukemia virus; PERV, porcine endogenous retrovirus; HERV-W, human endogenous retrovirus type W; SnRV, snakehead retrovirus; WDSV, walleye dermal sarcoma virus; BLV, bovine leukemia virus; HTLV-1, human T-cell leukemia virus-1; HTLV-2, human T-cell leukemia virus-2; EIAV, Equine Infectious Anemia Virus; FIV, feline immunodeficiency virus; HIV-1, human immunodeficiency virus-1; HIV-2, human immunodeficiency virus-2; SIV-mac, simian immunodeficiency virus- rhesus macaque; MVV, Maedi-visna virus. Figure adapted and modified from MacLachlan & Dubovi, 2017.

All retroviruses possess three open reading frames *gag*, *pol* and *env* encoding the essential viral structural and enzymatic proteins. *Group-specific antigen (gag)* gene encode the structural proteins, the matrix (MA), capsid (CA) and the nucleocapsid (NC). The *polymerase (pol)* gene encodes for the reverse transcriptase and integrase enzymes vital for the viral replication; and the *envelope (env)* gene codes for the host receptor attachment proteins, SU and TM on the viral envelope (Figure 3A). Furthermore, the virus also expresses a vital proteolytic enzyme, protease encoded by the *pro* gene located between *gag* and *pol* (Figure 3A). Retroviruses containing only these standard genes in their genome are termed as simple retroviruses (Figure 3A). On the other hand, retroviruses belonging to delta, epsilon, lenti and spumaviruses have additional genes that are regulatory or accessory in function and are classified as complex retroviruses (Figure 3B). These additional mRNA transcripts take control over the cellular functions and enable them to infect the immunocompetent hosts more frequently than simple retroviruses. The structural genes in the viral RNA abut sequences known as untranslated regions (UTRs) at the 5' and 3' ends of the genome. The UTRs consist of a terminal identical repeat sequence (R) and a unique sequence (U5 and U3) each at the 5' and 3' ends respectively (Figure 3A). The properties and the genome organization of retroviruses belonging to each group is illustrated in figure 3B.

A

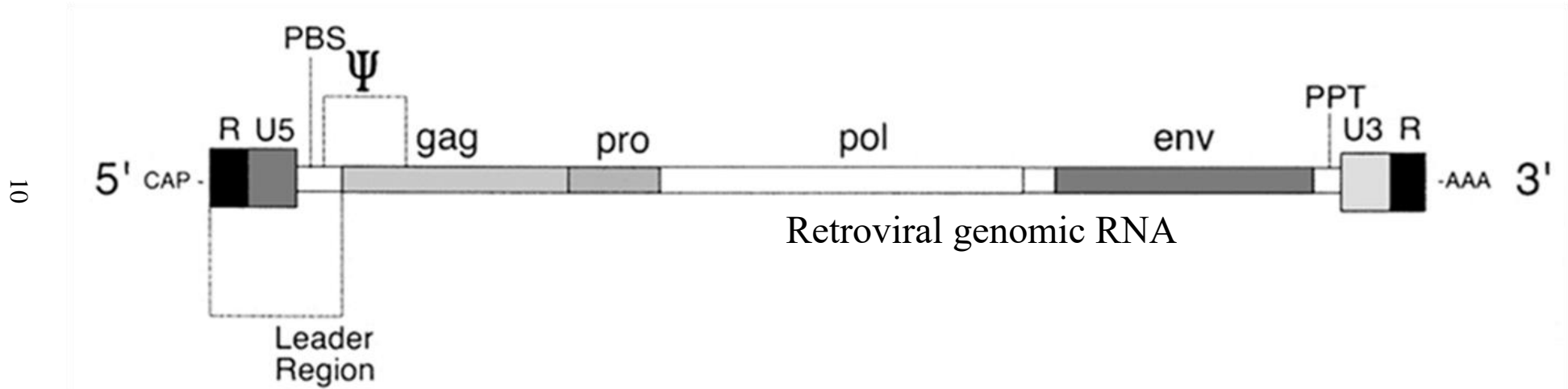
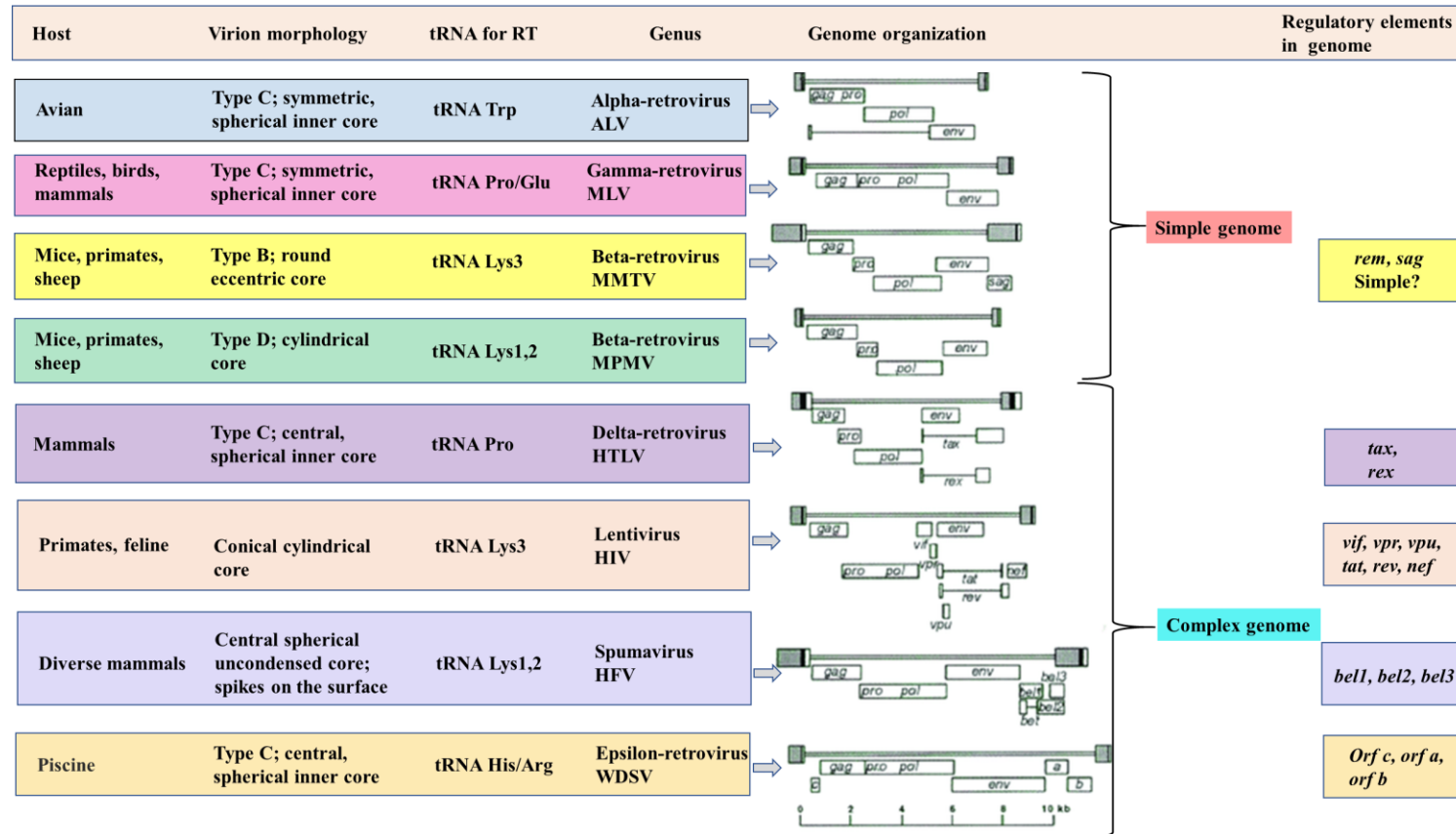


Figure 3. Genome organization of retroviruses

(A) **Retrovirus genome structure.** Four essential coding domains; Gag, Pro, Pol, and Env, are found in the same relative positions in all replication-competent retroviruses. The genomic RNA is a single-stranded molecule of positive polarity. The 5' and 3' ends are flanked by the repeat (R) sequence and the unique sequences (U5 and U3) to each end. Sequences important for replication are marked on in the approximate positions in which they are found; (PBS) primer-binding site; (ψ) packaging sequences; (PPT) polypurine tract; (PA) polyadenylation signal; (AAA) poly(A) tail. Figure and legend modified from Gifford & Tristem, 2003.

B



11

Figure 3 (B) Classification and properties of retroviruses. Host, Virion morphology, tRNA selected for reverse transcription (RT) initiation, level of genome organization and the additional regulatory genes expressing different mRNA transcripts in each group with complex genome is shown. MMTV, conventionally grouped as a simple retrovirus now appears to be complex due to virally encoded Sag and Rem proteins. ALV, Avian leukosis virus; MLV, murine leukemia virus; MMTV, mouse mammary tumor virus; MPMV, Mason-Pfizer monkey virus; HTLV, human T-cell leukemia virus; HIV; human immunodeficiency virus; HFV, human foamy virus; WDSV, walleye dermal sarcoma virus. Figure for the genome organization adapted from Coffin et al., 1997a.

1.5 An Overview of Retroviral Life Cycle

Retroviruses once inside the cell, reverse transcribe their RNA genome into DNA utilizing the reverse transcriptase enzyme. Reverse transcribed DNA is then imported into the nucleus where it is integrated into the host genome, process that is facilitated by the virally encoded integrase enzyme. This integrated retroviral genome is called the provirus which uses cellular machinery to transcribe RNA for making viral proteins as well as to generate extra copies of full-length RNA to be served as the genome for its progeny (Fields et al 1996). The transcribed viral mRNAs (both full-length and spliced RNAs) are exported out of the nucleus into the cytosol for translation. During assembly viral proteins assemble around its genomic RNA (gRNA) which is packaged into the nascent virions that bud off from the cell membrane as an immature virus particle and maturation occurs following the release from the cell surface (Figure 4).

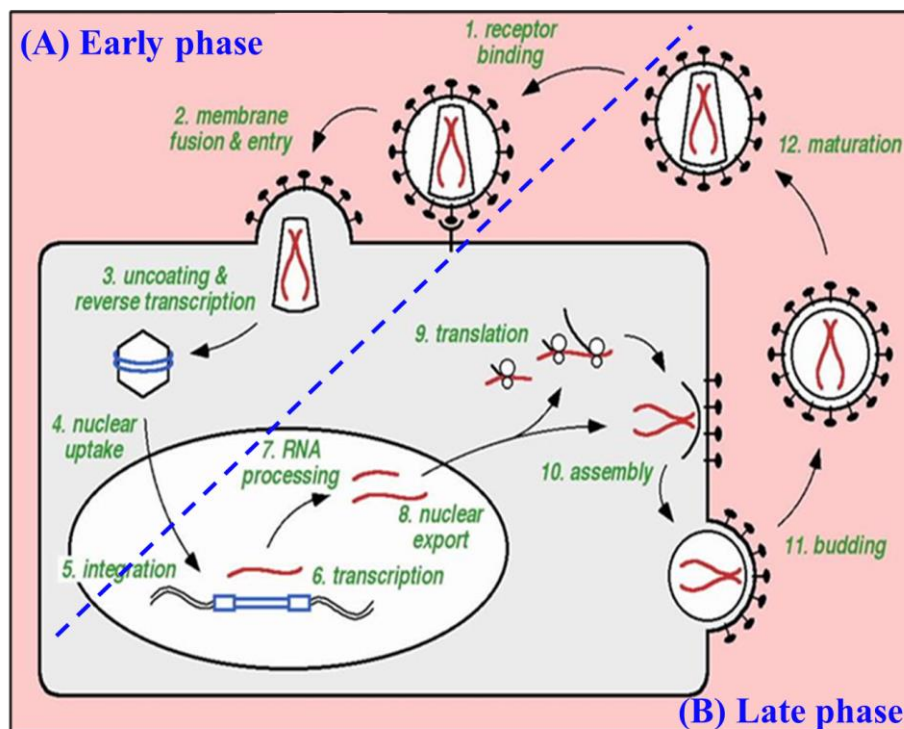


Figure 4. Steps in the retroviral life cycle

Various events in the life cycle of retroviruses are illustrated with HIV as an example. (A) Early phase: 1. Viral entry into cells by binding to a specific receptor on the cell surface; 2. Membrane fusion either at the plasma membrane or from endosomes (not shown); 3. Release of the viral core and partial uncoating and reverse transcription; 4. Transit through the cytoplasm and nuclear entry; 5. Integration into cellular DNA to give a provirus. (B) Late phase: 6. Transcription by RNA polymerase II (RNAPII); 7. Splicing and RNA processing; 8. Nuclear export of viral RNA; 9. Translation of viral proteins; 10. Gag assembly and RNA packaging; 11. Budding through the cell membrane; 12. Release from the cell surface and virus maturation.

1.5.1. Retroviral Receptors

How retroviruses evolved to infect a wide range of cells/tissues of its host is a question of great interest. Although, many of these host-virus interactions remain elusive, in the recent past it is becoming clear that viruses have evolved to recognize different cell surface molecules as receptors, and infection is restricted to only those cells that express these receptors. For example, the distribution of CD4 receptors on the T-helper lymphocytes, macrophages, and dendritic cells present them as potential target for HIV-1 infection (Didigu & Doms, 2012). Similarly, gamma retroviruses like Gibbon-ape leukemia virus (GaLV) and amphotropic murine leukemia virus (MLV) employ Gvrl1 and Ram1 as their receptors respectively (Miller & Chen, 1996). Even though these receptors (Gvrl1 and Ram1) are closely related at the protein sequence level and both function as sodium-dependent phosphate transporters, their interchangeable usage as receptors by either of these viruses is nearly impossible (Miller & Chen, 1996). On the other hand, there is evidence that the virus attachment is not sequence specific and there is redundancy and flexibility in receptor usage among viruses belonging to the same genus. For example, 10A1 MLV, a recombinant derivative of amphotropic MLV (Ott et al., 1990), can enter cells using either of the Gvrl1 and Ram1 receptors (Miller & Chen, 1996; Overbaugh et al., 2001). It is also known that among lentiviruses, HIV and SIV interact with the same CD4 receptor as well as the co-receptors CXCR4 and CCR5 belonging to the chemokine receptor family (Didigu & Doms, 2012; Overbaugh et al., 2001). In the same way, type D retroviruses, RD114 (cat endogenous virus), SRV (simian retrovirus), BaEV (Baboon endogenous retrovirus), HERV-W (human endogenous retrovirus type W) uses retroviral mammalian type D receptors (RDR or RDR2) as their specific entry receptors (Overbaugh et al., 2001). The functional homology of these receptor proteins reflects the wide host range and tissue tropism for different retroviruses. The details of the cloned receptors for different classes of retroviruses have been shown in Table 2.

Table 2. Retroviruses and its receptors

Retrovirus	Receptor	Cellular function
HIV, SIV	CD4 and CCR5 or CXCR4	Immune recognition, G protein-coupled chemokine receptors
FIV	CD9	Signaling transduction Protein
HTLV-1	Neuropilin-1 and GLUT-1	Trans-membrane glycoprotein, Glucose transporter 1
GALV, 10A1, FeLV-B,	Glvr1 (Pit1)	Phosphate transport
A-MLV, 10A1, MLV, FeLV-B	Ram1 (Pit2)	Phosphate transport
RD-114, type D SRV, BaEV, HERV-W	RDR or RDR2	Neutral amino acid transporter
MMTV	Mtvr	Transferrin receptor 1

Contents retrieved from Grove & Marsh, 2011; Overbaugh et al., 2001.

1.5.2 Virus Entry and Fusion

Infection of a cell by a retrovirus can be broadly divided into two phases: early and late (Figure 4). In step 1 of the early phase, the virus attaches to the host cell carrying the specific receptor or ligands followed by the membrane fusion of the virus and the host cell. These processes are brought about by the envelope SU and TM glycoproteins, exploiting either of the two mechanisms: i) receptor-mediated endocytosis (Figure 5) and ii) receptor-mediated cell signaling pathways leading to low-pH dependent endocytosis (Figure 6). SU recognizes the specific receptor on the target cells and therefore is the key determinant for the tropism of the virus. In receptor-mediated endocytosis as in the case of HIV and SIV, SU (gp120)-receptor binding triggers a conformational change and exposes the TM (gp41) ectodomain to induce the fusion of the viral and cellular lipid membranes (Clapham & McKnight,

2001; Dalgleish et al., 1984; Klatzmann et al., 1984) and reviewed in (Didigu & Doms, 2012; Figure 5) facilitating the entry of the viral core into the cytoplasm.

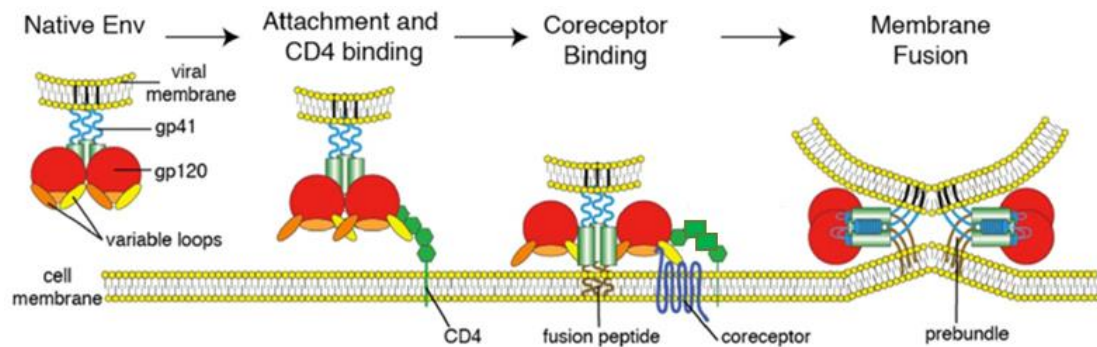


Figure 5. The HIV entry process as an example for receptor-mediated endocytosis

The figure above outlines a model for HIV Entry. The entry process begins with binding of gp120 (red) to its primary cellular receptor CD4 (green). CD4 binding results in conformational changes that allow binding of gp120 to the coreceptor-either CCR5 or CXCR4. Coreceptor binding results in triggering of the fusion machinery and formation of the six-helix bundle required to drive fusion of the viral and host cell membranes. Figure and legend adapted and modified from Didigu & Doms, 2012.

In the receptor-mediated signaling pathway for endocytosis, the virus attachment to its receptors transduces an array of signals to internalize via endosomes and trigger the acidification of the virion-containing intracellular vesicles to fuse with the viral envelope eventually, disposing the viral core into the cytoplasm (Figure 6; Grove & Marsh, 2011; Kubo et al., 2012). Retroviruses like avian sarcoma/leukosis virus (ASLV; Barnard et al., 2004), ecotropic murine leukemia virus (MLV; Kubo et al., 2012) and mouse mammary tumour virus (MMTV; Ross et al., 2002) utilize this mechanism for virus entry (Coffin et al., 1997a; Grove & Marsh, 2011; Kubo et al., 2012).

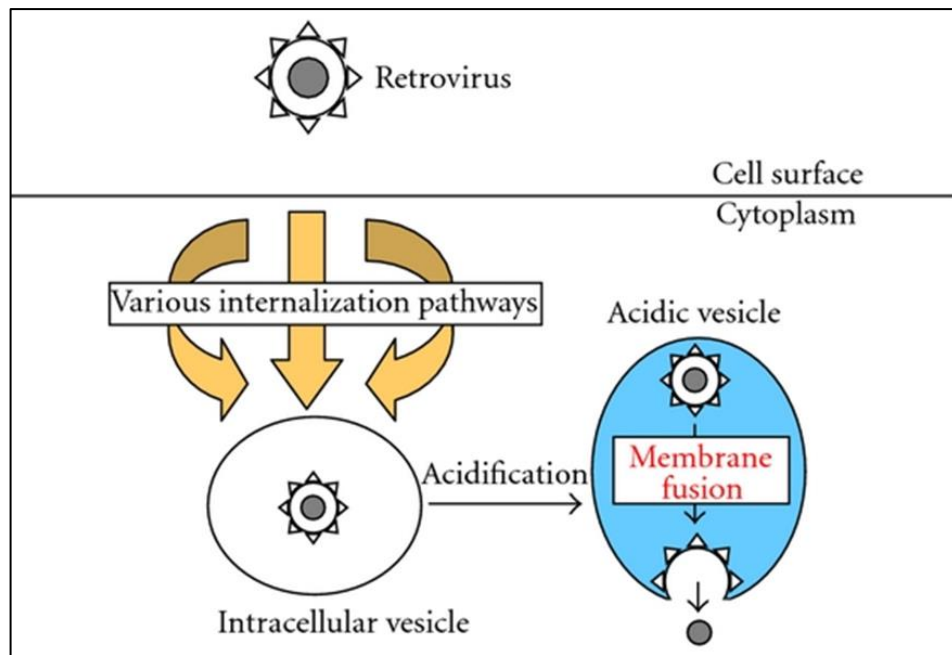


Figure 6. Receptor-mediated signaling pathways for endocytosis

Retrovirus particles are internalized into intracellular vesicles by various pathways, and vesicle acidification is necessary for the infections. Figure and legend adapted from Kubo et al., 2012.

1.5.3 Viral Uncoating, Reverse Transcription, and Integration

Fusion of the viral envelope with the plasma membrane or the endosomes separates the inner matrix from it to be released into the cytoplasm. This exposes the core containing two identical copies of dimeric RNA held together by non-covalent base-pairing as the viral genetic material. This dimer RNA molecule is tightly packaged in the capsid (CA) by the association with the zinc-finger proteins, the nucleocapsid (NC). Soon after the entry of the core into the cytoplasm, it must traverse through the cellular actin meshwork before the viral replication could begin. In the case of HIV-1, the nef protein help to overcome this barrier to promote infectivity (Campbell et al., 2004). Viral entry by endosome acidification is another way for actin cytoskeleton evasion (Albritton, 2018). While in the cytoplasm, the virus prepares to copy its single stranded genomic RNA (ssgRNA) into double stranded DNA (dsDNA) by the process of reverse transcription using the enzyme, reverse transcriptase (RT) that has been shown as bound to the RNA template (Bohlayer & DeStefano, 2006). RT is a heterodimer with various functions: RNA-dependent and DNA-dependent DNA polymerase and RNase H activities. It is perceived that the specific tRNA (transfer RNA) priming at the primer binding site (PBS) on the RNA template and

extension of a few nucleotides by RT begin prior to cell penetration and presumably prevents its elongation due to the shortage of deoxynucleotides (dNTPS) in the virion (Gilboa et al., 1979; Rothenberg et al., 1977; Rothenberg & Baltimore, 1976). Shortly after the porous core encounters ample concentration of dNTPS from the cytosolic environment, it triggers the DNA synthesis (Jacques et al., 2016). The RNA that is being reverse transcribed remains as a large complex with RT and the nascent DNA along with the CA forming the reverse transcriptase complex (RTC) and following the complete DNA synthesis, the complex is termed as preintegration complex (PIC). It has now been documented that CA association with the RTC or PIC is a crucial player in controlling the movement of the core towards the nucleus while uncoating and promotes efficient reverse transcription (Goff, 2018; Li et al., 2021). Details of the steps (A-H) involved during reverse transcription are explained in detail in figure 7. In addition, the precise locations of the *cis*-acting elements (such as R, U5, and U3 regions, the polypurine tract (PPT) as well as PBS) needed during reverse transcription are also shown on the RNA template in figure 7 whereas the structure of long terminal repeats (LTRs) in the proviral DNA is shown in figure 8.

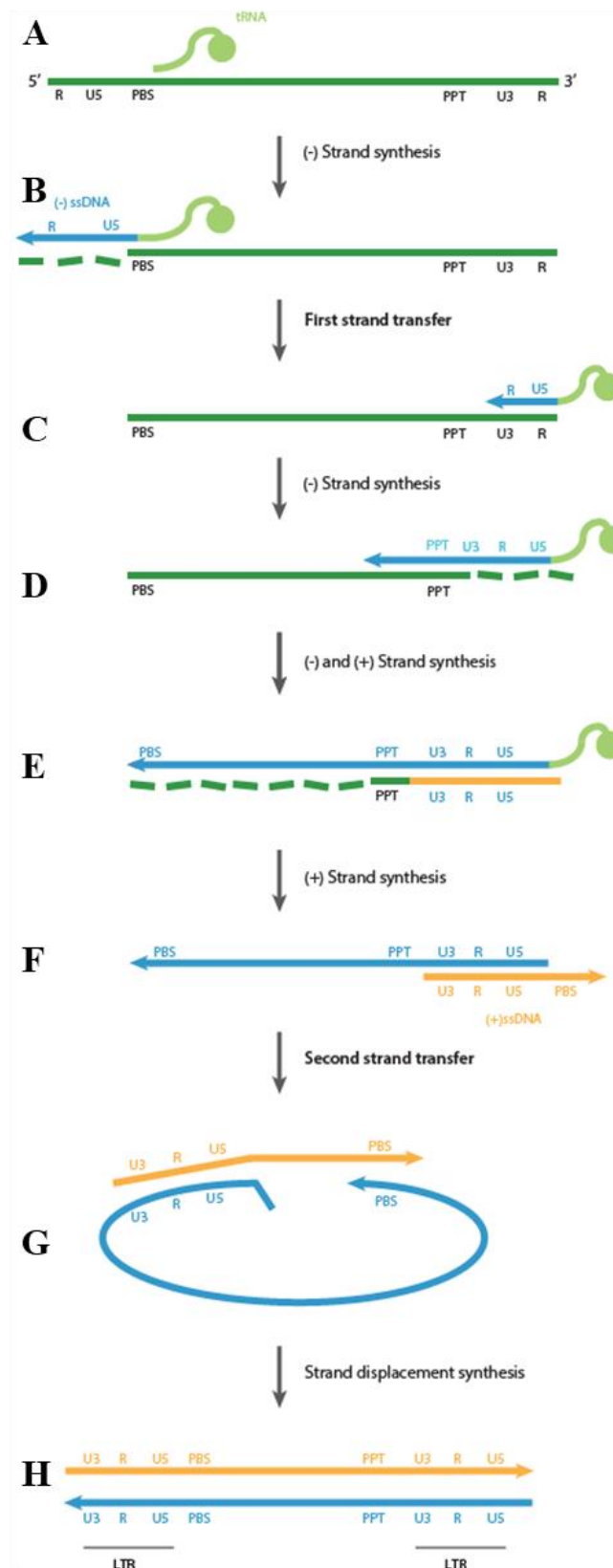


Figure 7. Delineating various steps involved in reverse transcription

A. First strand synthesis starts with the tRNA acting as a primer and partially binding to the PBS present at the 5' end of the RNA template.

B. The tRNA extends through the U5 and R region on the template RNA until it reaches the end of the RNA transcript. This creates a short strand of DNA of 100-150 bp that is termed as minus strand strong-stop DNA ((-) sssDNA).

RNase H cleaves the template RNA so that the newly synthesized cDNA is not annealing to anything, which is when the first strand transfer/jump occurs.

C. The (-) sssDNA R region anneals to the complementary R region on the 3' end of the template RNA.

D. First strand synthesis continues and the (-) sssDNA is extended along the entire length of the template until the PBS on the 5' end of the RNA genome.

E. RNase H continues to degrade the template RNA during this synthesis. However, the RNA template has a part called the PPT region within the sequence that is relatively resistant to the effects of RNase H so it is not degraded and actually acts as the primer for second strand synthesis.

F. Second strand synthesis now begins from the PPT region. The second strand of DNA extends in the opposite direction of the first strand of DNA. Once it copies the end U5 region on the first strand, it creates a short portion of DNA called the plus strand strong-stop DNA ((+) sssDNA).

G. The RNase H causes the tRNA primer on the first strand to fall off so that the PBS on the second strand can now bind to the PBS on the first strand (due to complementarity) and constitutes the second strand transfer.

H. Each strand is now simultaneously a template for one another. They are both extended in opposite directions, the second strand still having the majority to synthesize while the first strand only needs to synthesize the rest of its 5' LTR region.

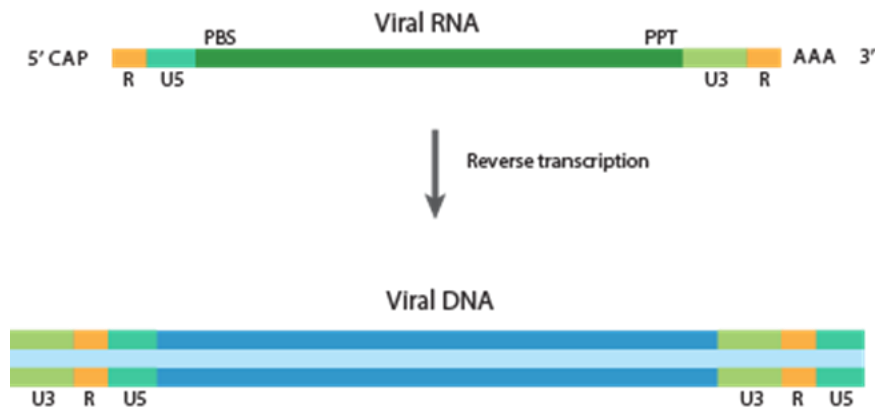


Figure 8. Schematic representation of viral gRNA and proviral DNA after reverse transcription

Note the duplication of unique sites, U3 and U5 at either ends of the reverse transcribed DNA with two identical LTRs making it longer than gRNA making it suitable substrate for integration.

After reverse transcription, the PIC must reach the nucleus for its integration into the host cell chromosomes (Coffin et al., 1997a; Zhang et al., 2000). Retroviruses must cross the nuclear barrier to execute this obligate step during virus replication. Gammaretroviruses like MLV, that infect the dividing cells enter the nucleus during mitosis when the nuclear envelope dissolves, thereby provide access to the host chromosomes (Yamashita & Emerman, 2006). In the case of lentiviruses such as HIV-1 that infect both dividing and non-dividing cells, PIC arrive the nucleus through the nuclear pore complex. Capsid, viral protein R (Vpr) and integrase (IN) possess an arginine-rich nuclear localization signal (NLS) within PIC and interacts with β -Karyopherin transportin1 (TRN-1; Fernandez et al., 2019) along with the help of host factors like the cleavage and polyadenylation specificity factor 6 (CPSF6) play a key role in viral trafficking to the nucleus (Bedwell & Engelman, 2021). Using the host nuclear transport proteins like the transportin 3 and nucleoporins (Nup; as detailed in the figure below), (Endsley et al., 2014; Popov et al., 1998 and reviewed in Matreyek & Engelman, 2013; Shukla & Chauhan, 2019) the viral core in a partially or completely intact form traverses the nucleus and uncoats approximately 1.5 hours before integration. (Figure 9 (a) Nuclear import; (Burdick et al., 2020; Dharan et al., 2020; Zila et al., 2020).

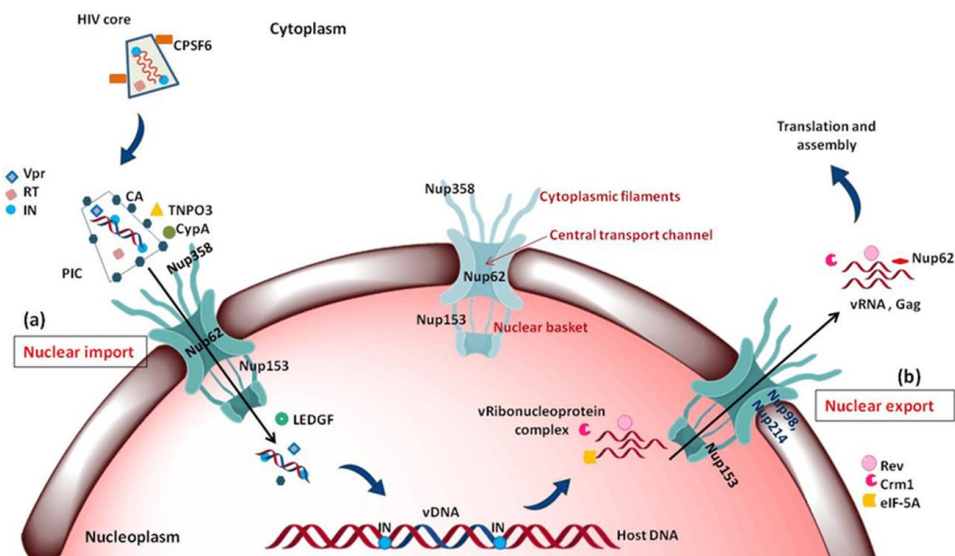


Figure 9. Major events occurring during nuclear import and export of HIV-1

(a) Nuclear import of HIV-1 PIC: CPSF6 interacts with intact HIV core which upon uncoating and reverse transcription forms the PIC. TNPO3, CypA and CypA like domain of Nup358 binds to the leftover CA and helps in docking of PIC on the cytoplasmic face of NPC; while Vpr and IN having NLS, interact with Nups like Nup153 which help in import of the PIC into the nucleus. In nucleoplasm, LEDGF is involved in release of PIC from Nup153 and helps in integration of vDNA into the host DNA; (b) nuclear export of vRNA-protein complex: The unspliced and partially spliced mRNA need viral accessory protein Rev which interacts with various cellular factors like Crm1, eIF-5A and Nups, such as Nup62, Nup214, Nup98 and facilitate the nuclear export of this viral ribonucleoprotein complex. Abbreviations: NPC-nuclear pore complex, CPSF6-cleavage, and polyadenylation specific factor 6, TNPO3-transportin 3, CypA-cyclophilin, CA-capsid, Vpr-viral protein r, IN-integrase, RT-reverse transcriptase, PIC-pre integration complex, LEDGF-lens epithelium-derived growth factor, IN-integrase, Crm1-chromosome region maintenance- 1, Rev-regulator of virion expression, eIF-5A-eukaryotic translation initiation factor-5A. Figure and figure legend adapted from Shukla & Chauhan, 2019.

The completed reverse transcribed viral DNA (vDNA) is rendered integration-competent by the integrase (IN). IN hydrolysis the linear 3'OH ends at the conserved CA/TG dinucleotides located within the viral LTRs of the DNA strands and this processing activity by IN is a pivotal step in facilitating nuclear import of the PIC cargo. Once inside the nucleus, the IN still bound to the processed vDNA catalyzes nicking of the phosphates at the exposed target sites in the host DNA around the nucleosomes (Figure 10; as reviewed in Andrade & Skalka, 2015; Matreyek & Engelman, 2013). Although viral integration is thought to be a random process, recent studies have shown that vDNA may prefer to get incorporated at regions upstream of active genes. These regional hotspots may favor proviral transcription and could be advantageous for viral replication (Wu, 2004).

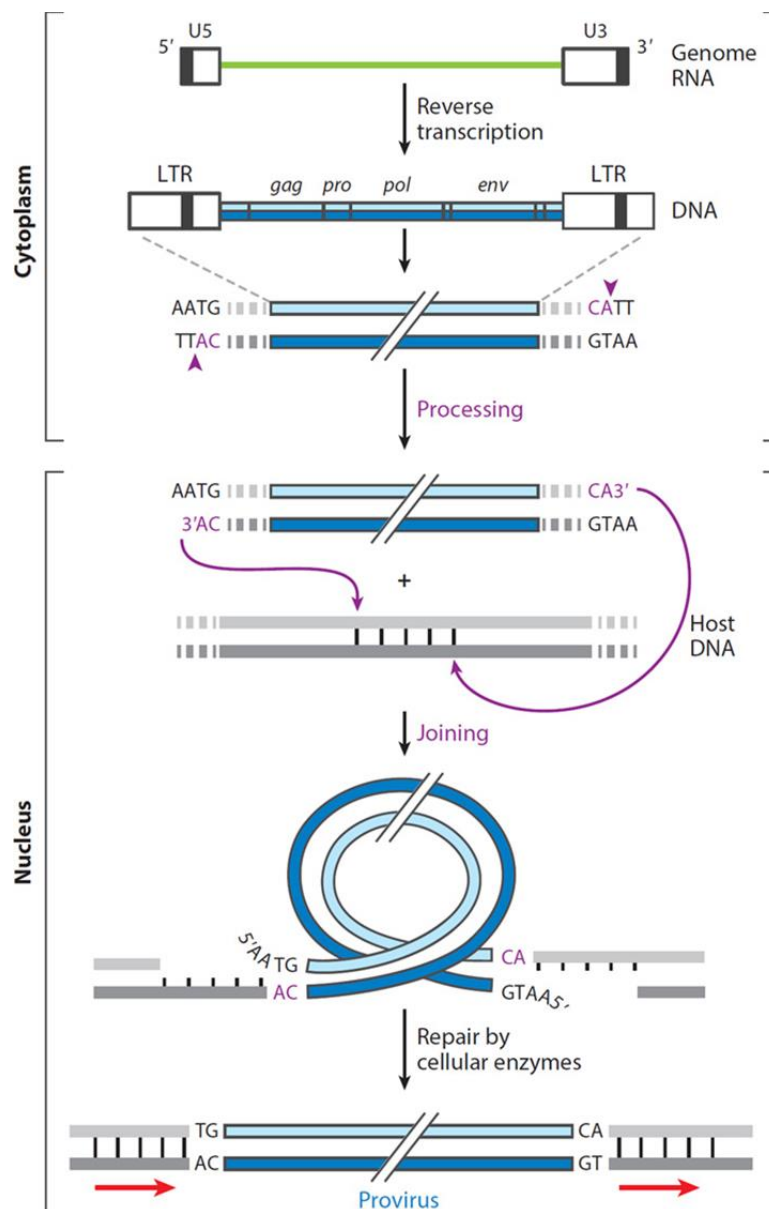


Figure 10. Steps in the synthesis of retroviral DNA and its integration into host DNA

The viral RNA genome (green line) is reverse transcribed in the cytoplasm of the cell within a subviral nucleoprotein structure (called the reverse transcription complex, RTC) to form a duplex DNA containing long terminal repeats (LTRs) of sequences unique to the 5' (U5) and 3' (U3) ends of the viral RNA. Imperfect inverted repeats at the LTR duplex termini are recognized and nicked by cognate integrase (IN) proteins, following a conserved CA dinucleotide at each 3' end (vertical arrows in inset), producing recessed 3'-OH ends. This first reaction catalyzed by IN is called processing and takes place within a nucleoprotein assembly called a preintegration complex. The conserved CA dinucleotide and steps catalyzed by IN are highlighted in magenta. Removal of the noncomplementary 5' nucleotides of vDNA and repair of the gaps in host DNA by host enzymes generate a covalently integrated provirus, shorter by 2 bp on both ends and flanked by duplications of the target site (red arrows). Figure and legend, courtesy from Andrade & Skalka, 2015.

1.5.4 Transcription of the Provirus, Splicing Events, and Nuclear Export

1.5.4.1 Synthesis of the Viral RNA from Proviral DNA

The mechanism for the synthesis of the viral RNA from the proviral DNA occurs in the late phase of viral infection (figure 4B) and is completely dependent on the host cellular machinery. Therefore, the recruitment of the range of cellular transcription factors and activators from a particular cell type by a retrovirus, dictates the features of its own replication mechanism. For example, MMTV that infects the mouse mammary cells utilize the steroid hormones, glucocorticoid, and progestin to stimulate the LTR promoter for the expression of the MMTV proviral genes (Coffin et al., 1997a). Retroviruses that infect the lymphoid cells employs lymphoid-specific transcription factors for transcribing their RNA from proviral DNA. Complex retroviruses like HIV and HTLV encode their own transcription factors like Tat and Tax respectively, and work in collaboration with other cellular transcription factors and take the control of its own and/or the host cell gene expression (Ajasin & Eugenin, 2020; Enose-Akahata et al., 2017; Nyborg et al., 2010; Rice, 2017). For instance, Tax of HTLV-1 activates the NF- κ B and CREB/ATF pathways that are responsible for controlling cellular transcription, proliferation and evade apoptosis and senescence by modulating the expression of a number of cellular genes and leads to adult T-cell leukemia (ATL) in the host (Matsuoka & Jeang, 2011).

1.5.4.2 Processing of the Viral RNA Transcripts

All the viral transcripts, full-length and the sub-genomic sized RNAs undergo post transcriptional processing to resemble the cellular mRNAs. The post transcriptional modifications include the capping of the 5' R region with m⁷G5'ppp5'G_mp and the cleavage and polyadenylation at the U5 region of the 3'LTR which is a usual process for proper protein synthesis within the host cell. Since the transcription initiation site is the same (5'end R region) for all the viral mRNA species, various gene expression from the same transcript demands for multiple unusual post transcriptional regulatory steps. All subsets of mRNAs destined for the synthesis of envelope glycoproteins and other accessory proteins (in complex retroviruses) essential for viral replication and host antiviral evasion are generated from the primary full-length RNA transcripts (Figures 11 & 12). In simple retroviruses like MPMV, MLV, SNV, and RSV only one or two singly spliced classes of mRNA are found

(Figures 11 & 12). However, in the case of complex retroviruses like in HIV and HTLV where the genome consists of additional regulatory or accessory genes, the single mRNA is multiply spliced into several classes of mRNA by alternative RNA splicing mechanisms (Figures 11 & 12; Coffin et al., 1997b; Hunter, 2008 and reviewed in Bolinger & Boris-Lawrie, 2009). In all retroviruses, the unspliced full-length RNA forms the gRNA for progeny virions while also serving as a template for the synthesis of various structural and enzymatic proteins required for viral assembly. The only class of singly spliced mRNA found in all genera of retroviruses is the one that encodes for *env* gene product. All these spliced mRNAs share a common sequence at their 5' ends with an exception in spumaviruses, where they make use of an internal promoter (IP) located in the *env* gene upstream of the accessory protein coding genes (MacLachlan & Dubovi, 2017; Yu et al., 1999). Figure 11 gives an overview of retroviral transcription and RNA processing.

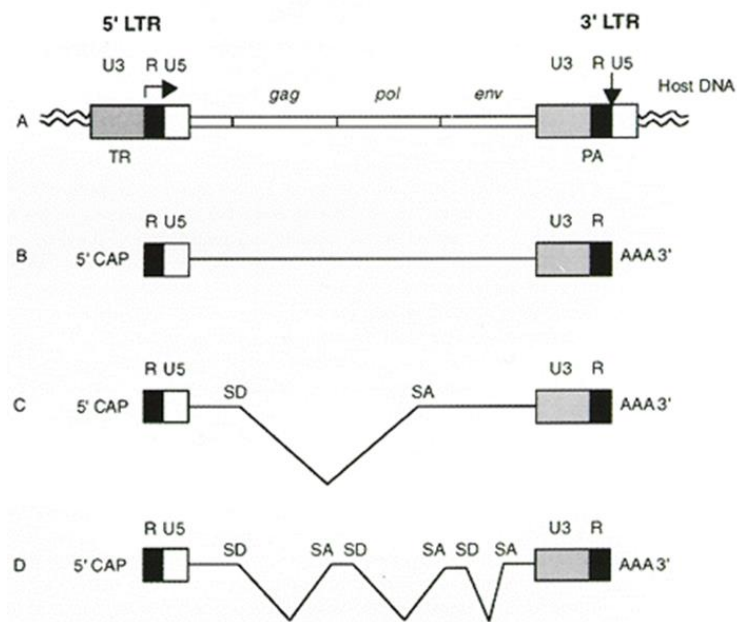


Figure 11. Overview of retroviral transcription and RNA processing

Typical proviral structure (A) with identical 5' and 3' LTRs. Genes illustrated above the line are shared by all replication-competent retroviruses. The horizontal arrow marks the start site of transcription. The vertical arrow denotes the site of 3'-end processing and polyadenylation in the RNA transcript. All viruses synthesize full-length genomic RNA (B). The subgenomic-sized transcripts illustrate some of the possible products of alternative splicing. In simple retroviruses (C), only a single splice donor or, occasionally, two splice acceptors are used. In complex viruses (D), multiple spliced products are found. (TR) Transcriptional control elements that function in DNA; (PA) polyadenylation and 3'-end processing signals that function in RNA; (CAP) posttranscriptional modification of 5' termini of all transcripts; (SD and SA) spliced donor and acceptor signals for splicing (C); (AAA) polyadenylation of 3' termini of all transcripts. Figure and legend from Coffin et al., 1997b.

Different classes of mRNA transcripts in retroviruses

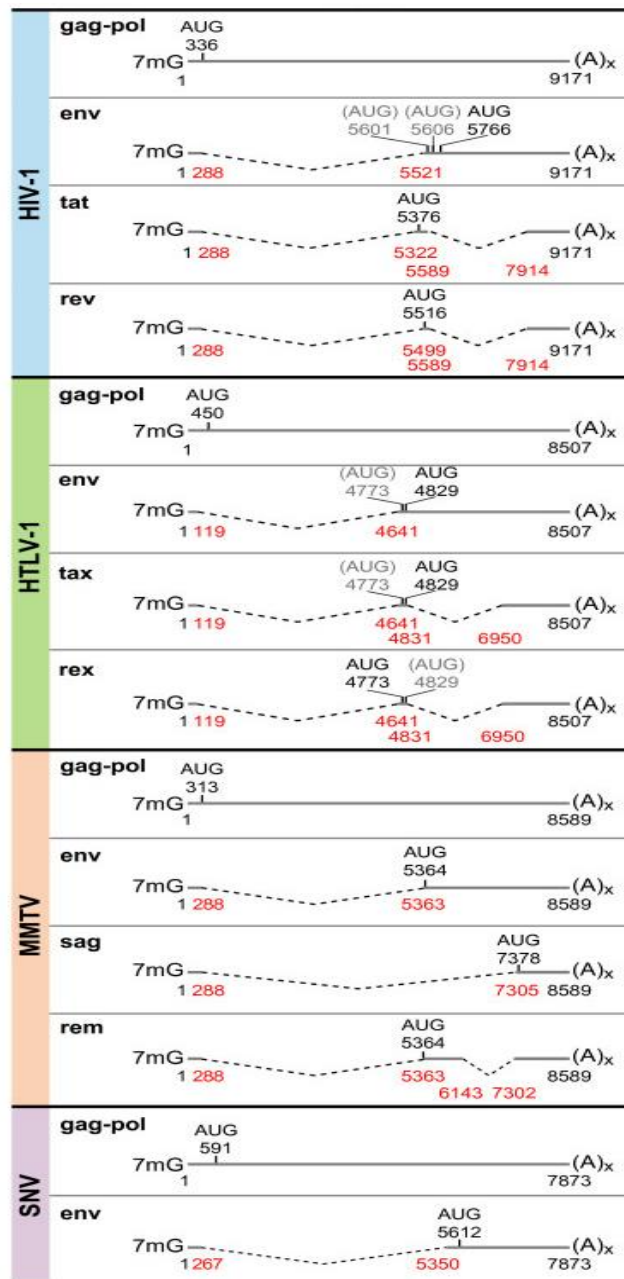


Figure 12. Properties of selected retrovirus transcripts

Human Immuno deficiency virus (HIV-1), human T-cell leukemia virus type 1 (HTLV-1), mouse mammary tumor virus (MMTV), and spleen necrosis virus (SNV) transcripts are depicted, including predominant unspliced and spliced mRNA species. Numbering is in reference to the first nucleotide of R, the RNA start site, as +1. Red numbers below each mRNA indicate the nucleotide position of exon junctions. Dashed lines denote introns. AUG indicates translation initiation codon, and black numbers indicate AUG nucleotide positions. The unused AUG in bicistronic transcripts is depicted in gray parentheses. 7 mG, 5' RNA cap structure; (A)_x, poly A tail. HIV-1 information was derived from (Purcell and Martin, 1993); HTLV-1 information was derived from (Nicot et al., 2005) and GenBank NC_001436; MMTV information was derived from GenBank U40459, DQ223969, and (Coffin et al., 1997a); SNV information was derived from reference sequence pPB101 (Bandyopadhyay & Temin, 1984). Figure and legend courtesy: Bolinger & Boris-Lawrie, 2009.

1.5.4.3 Regulation of the Viral RNA Levels in the Cell

Retroviruses are posed with the challenge of regulating the relative amounts of unspliced and spliced messages that reach the cytoplasm. In complex retroviruses, the HIV *trans*-activating transcription factor (Tat; Kao et al., 1987) and the regulatory protein Rev are expressed at a basal transcription level in the early phase of infection (Lata et al., 2015; Wu & Marsh, 2003), by direct interaction between cellular transcription factors and *cis*-acting elements located in the HIV promoter region. Subsequently, in the late phase of infection Tat then mediate the production of high levels of viral RNA (Kim et al., 1989) by directly interacting with the *cis*-acting *trans*-activation response (TAR) element. TAR is present at the 5' end of the viral RNA and exerts a positive effect on transcriptional elongation (Ajasin & Eugenin, 2020; Rice, 2017). The HTLV Tax protein, also functions in a similar manner by binding to the *cis*-acting Tax-responsive elements (TxREs) in the U3 region (Enose-Akahata et al., 2017; Nyborg et al., 2010). The full length unspliced mRNA which functions as the gRNA as well as the template for encoding the Gag and Pol proteins should remain predominant in the RNA pool so that it can be translated in a fashion that variable amounts of Gag and Pol proteins can be produced as required during the viral assembly progression (Coffin et al., 1997b). As opposed to cellular unspliced RNAs, retroviruses have evolved mechanisms to transport their unspliced or singly spliced mRNAs from the nucleus to the cytoplasm of the infected cell. Unlike cellular mRNAs, these viral RNAs efficiently overcome the nuclear retention by the interaction between the *cis*-acting elements RRE (HIV), RxRE (HTLV) present on these transcripts with virally encoded Rev and Rex proteins respectively along with host nuclear export proteins (see figure 13; reviewed in (Balvay et al., 2007; J. LeBlanc et al., 2013; Shida, 2012). Simple retroviruses modulate their ratio of unspliced to spliced messages either by acquiring variably efficient splice sites or by incorporating *cis*-acting sequences into their genome that regulate the export of full-length gRNA from the nucleus (detailed in figure 13). In fact, in infected cells, only less than 1% of all cellular mRNA is viral RNA (Kutluay et al., 2014; Rulli et al., 2007 and reviewed in Bieniasz & Telesnitsky, 2018).

1.5.4.4 Nuclear Export of the Viral RNAs

The efficient transcription, processing and splicing of the viral mRNA by the host cell machinery generates three classes of mature RNAs-viz. completely spliced, partially spliced and unspliced messages. The latter two RNAs are too large to be exported from the nucleus by the general cellular mRNA pathway. Thus, only the completely spliced mRNAs expressing the viral effector proteins like the Tat, Rev and Nef in HIV (reviewed in Emerman & Malim, 1998), Tax and Rex in HTLV (reviewed in Shida, 2012), src in RSV (reviewed in LeBlanc et al., 2013) and the constitutive *env* encoding mRNAs are readily translocated to the cytoplasm for translation. The regulatory proteins like the Tat/Rev and Tax/Rex that are required abundantly in the early and late phases of viral replication contain an arginine-rich nuclear localization signal (NLS) that help them shuttle back from the cytoplasm to the nucleus through interaction with importin β (Truant & Cullen, 1999). To export unspliced viral RNA from the nucleus to the cytoplasm, retroviruses have co-opted transport mechanisms by either Tap-dependent or Rev/CRM1-dependent routes. Simple retroviruses like MPMV contain a constitutive transport element (CTE) at the 3' end of their genome (Bray et al., 1994; Rizvi et al., 1996; Rizvi et al., 1996) that interact with the host nuclear export factor 1 (NXF1) or Tap that facilitates export and expression of CTE-containing mRNAs (reviewed in Cochrane et al., 2006; Swanson & Malim, 2006). Other simple retroviruses like MLV and RSV also follow the same Tap-dependent transport mechanism by cooperating with its cis-acting sequences (Pessel-Vivares et al., 2014; Ricaña & Johnson, 2021). In complex retroviruses, the Rev protein in HIV and the Rex protein of HTLV interacts with their respective RNA elements (RRE and RxRE) at the 3' end to allow nuclear export of its partially spliced or unspliced viral messages (Groom et al., 2009 and reviewed in Shida, 2012). Both these viruses use the CRM1 pathway and interacts with several cellular proteins like Crm1, eIF-5A and Nups to facilitate the export of Rev-RRE or Rex-RxRE RNA complexes (see figure 9 (b) nuclear export for illustration; reviewed in Shukla & Chauhan, 2019). Similarly, MMTV, which was formerly grouped as a simple retrovirus, also encodes a Rev-like accessory protein, termed Rem which binds to the *cis*-acting Rem responsive element (RemRE) located at the 3' end of the viral transcripts (Akhlaq et al., 2018; Indik et al., 2005; Mertz et al., 2005) to facilitate its nuclear export. Identification of Rem protein synthesized from a doubly spliced mRNA led to the proposition to reclassify MMTV

as the first complex murine retrovirus (Mertz et al., 2005). More details on RNA transport mechanisms of viral transcripts in simple and complex retroviruses can be inferred from figures 13 & 14.

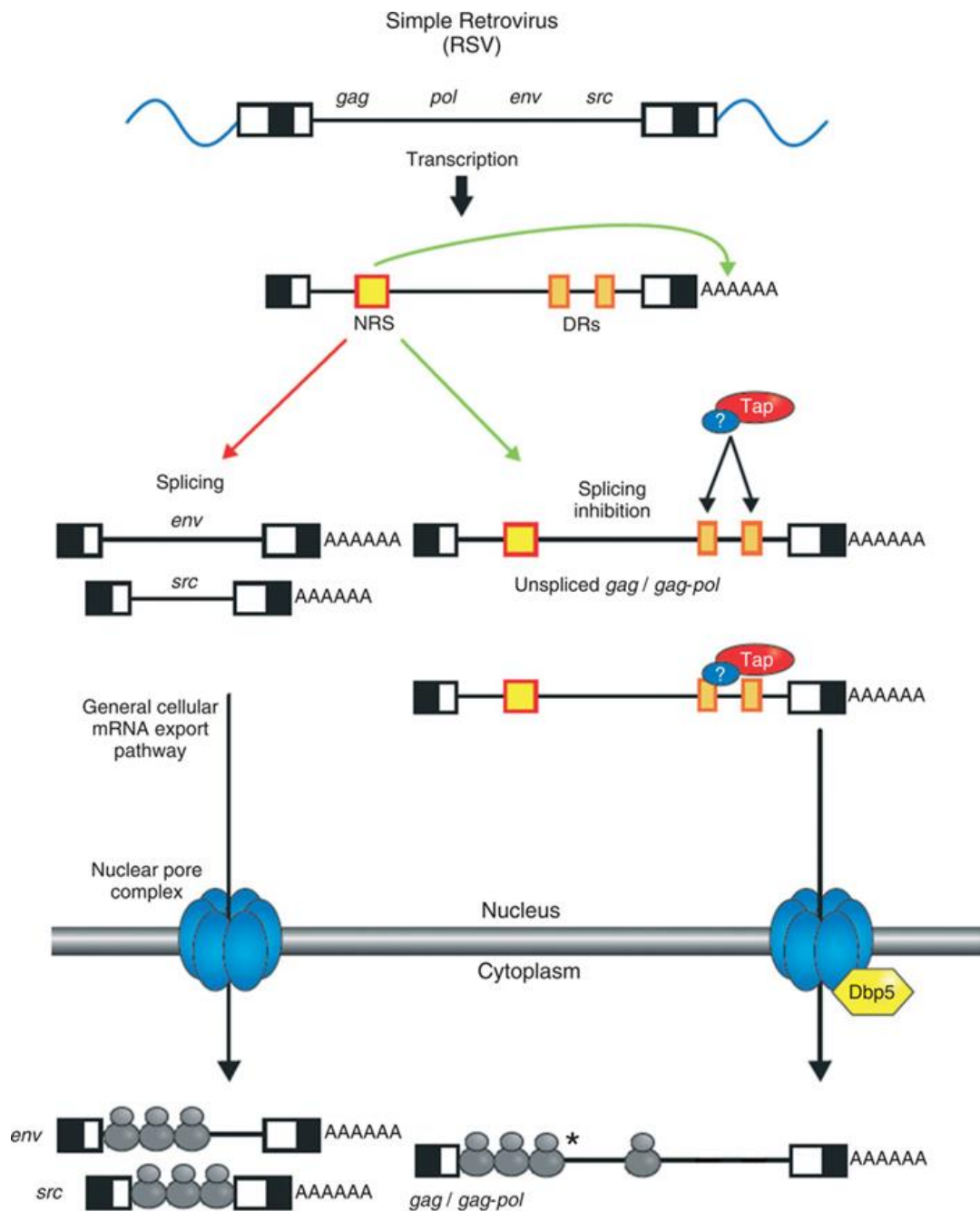


Figure 13. Regulation of primary transcript in simple retroviruses

Simple retroviruses like Rous sarcoma virus transcribe a 9 kb primary RNA transcript. A fraction of these transcripts is spliced to generate sub-genomic env and src mRNAs (left side), which are exported to the cytoplasm like spliced cellular mRNAs. The remaining primary transcripts are not spliced (right side), due to inhibition by the viral NRS (Negative regulator of splicing) element and by inefficient 3' splice sites. The viral DR (Direct repeat) sequences interact with cellular proteins to facilitate export of the unspliced RNA to the cytoplasm. The cytoplasmic translocation of the unspliced viral RNAs through the nuclear pore complex is facilitated by the DEAD box helicase Dbp5. Figure and legend from LeBlanc et al., 2013.

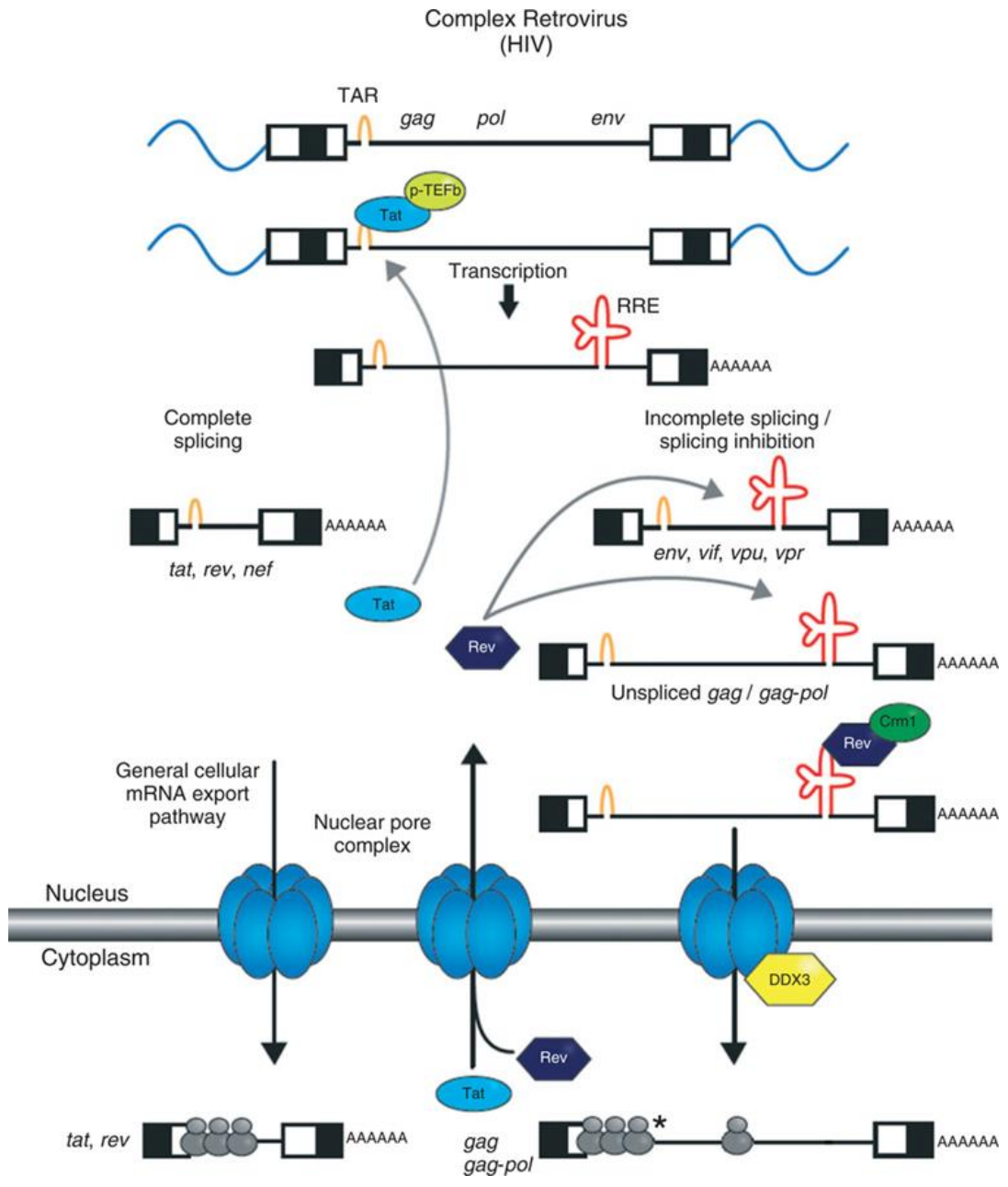


Figure 14. Regulation of primary transcript in complex retroviruses

Complex retroviruses, such as HIV-1, generate 3 classes of mRNA: unspliced (9 kb), singly spliced (4 kb), and completely spliced (2 kb). Early HIV gene expression results in the generation of completely spliced mRNAs (left side), including *tat* and *rev*. Following translation, Tat and Rev are imported into the nucleus. Tat binds the TAR RNA element to promote transcriptional elongation, while Rev binds the RRE of singly spliced and unspliced RNAs (right side) to mediate nuclear export of the viral RNA after recruiting the cellular factor CRM1. The cytoplasmic translocation of the partially spliced/unspliced viral RNA through the nuclear pore complex is facilitated by the DEAD box helicase DDX3. Figure and legend courtesy LeBlanc et al., 2013.

1.5.5 Translation of the Expressed Viral RNAs

After the successful transport of the processed retroviral mRNAs into the cytoplasm, it can finally be translated into its protein components. The polycistronic nature of retroviral primary transcripts helps to maximize its coding capacity and is tightly regulated at the translational level to produce a balanced level of viral proteins necessary for viral replication. This translational control is executed at the initiation step and is the rate-limiting process in protein synthesis (Gingras et al., 1999). In the cytoplasm, the efficiently spliced monocistronic RNAs are translated very much like the cellular mRNAs in a 5' cap-dependent manner where the ribosomes scan for a consensus Kozak sequence (GCC (A/G) CC AUG G) upstream to an authentic AUG initiation codon. In retroviruses, since all the transcripts, spliced or unspliced are transcribed from a common R region at the 5' end, all these mRNAs invariably contain an identical 5' untranslated region (UTR) up to the major splice donor site followed by variable length of UTRs just upstream of the start codon of individual spliced transcripts (Figure 15). This conserved region contains several highly structured *cis*-acting sequences, including TAR in lentiviruses, poly(A) signal, PBS, dimerization initiation site and splice donor site which are part of sequences responsible for packaging the gRNA into the assembling virus particles hence called packaging signal. These uniform long and highly structured motifs at the 5' end of both the precursor and differentially spliced mRNAs render the ribosome scanning inefficient for translation initiation in retroviruses. Another factor that impedes the retroviral translation initiation is the presence of numerous CUG or AUG codons before the authentic start codon which tamper gene expression. Presence of degenerate Kozak consensus sequences in the vicinity of the start codon is another regulatory factor that governs the expression of viral genes (see figure 15, reviewed in Balvay et al., 2007; Bolinger & Boris-Lawrie, 2009; LeBlanc et al., 2013; Yilmaz et al., 2006).

Major transcript		Properties of 5'UTR				
Classes	Structure	Presentation of secondary structure	Length (nt)	Predicted free energy (Kcal/mol)	Number of upstream CUG or AUG	Comparison to Kozak consensus GCC(A/G)CCAUG G
gag-pol	9-kb class mRNA m7G — AUG — (A) _n		336	-119.6	9 / 0	aga <u>G</u> ag AUG <u>G</u>
vif	4-kb class mRNA m7G — AUG — (A) _n		418	-132.0	10 / 0	Ggg <u>G</u> uu AUG <u>G</u>
vpr	4-kb class mRNA m7G — AUG — (A) _n		459	-153.3	11 / 0	Gga <u>ç</u> ag AUG <u>G</u>
vpu	4-kb class mRNA m7G — AUG(AUG) — (A) _n		374	-124.4	9 / 1	cau <u>G</u> ua AUG <u>ç</u>
env	4-kb class mRNA m7G — (AUG)AUG — (A) _n		535	-143.2	9 / 2	Gug <u>G</u> Ca AUG <u>a</u>
tat	2-kb class mRNA m7G — AUG — (A) _n		343	-120.9	9 / 0	caa <u>G</u> aa AUG <u>G</u>
rev	4-kb class mRNA m7G — AUG — (A) _n		305	-112.8	9 / 0	uCu <u>ç</u> Cu AUG <u>G</u>
nef	4-kb class mRNA m7G — AUG — (A) _n		776	-235.2	15 / 1	uau <u>Δ</u> ag AUG <u>G</u>

Figure 15. Organization and features of the 5' untranslated regions of the major HIV-1 transcripts

Structure of the predominant transcript for each open reading frame is shown (Purcell & Martin, 1993). Colored lines highlight sequences that comprise the various 5' UTRs. The blue line depicts the first noncoding exon, which is 289 nts in length and is present in all HIV-1 transcripts. The black line denotes the additional 47 nts that are maintained in the unspliced transcript, which is mRNA template for translation to Gag and Gag-Pol protein and is genomic RNA that is packaged into progeny virions. The gray lines represent coding regions and dashed lines denote introns. AUG indicates the translation initiation codon. Also indicated is the number of CUG and AUG codons that are present upstream of actual initiation codon. Translation to Vpu and Env is achieved from the same transcript. The vpu AUG is subject to leaky scanning and scanning may continue until recognition of the downstream env AUG. The unused AUG in this bicistronic transcript is labeled in parenthesis. Depiction of predicted secondary structure of replication motifs is adapted from (Berkhout, 1996); the motifs include the trans-activation responsive sequence; primer binding site; dimer initiation site; 5' splice site; and core RNA packaging signal. Alternatively, spliced exons, which are depicted in color, may induce variations in the folding of the upstream region, which are not shown. Free energy predictions were calculated using Mfold (Mathews et al., 1999; Zuker, 2003). Deviations from Kozak consensus are indicated in lowercase letters; underlined nts indicate those most critical for efficient cap-dependent translation initiation. Figure and legend from Yilmaz et al., 2006.

Despite these impediments, retroviruses have adopted varied mechanisms in collaboration with the viral and host translational factors to govern its protein synthesis in an efficient manner. HIV, HTLV, SNV, and MPMV contain a cap-dependent translation enhancer called a posttranscriptional control element (PCE) that interacts with RNA helicase A and enhances polysome loading and Gag protein production (Bolinger & Boris-Lawrie, 2009). Another feature that can modulate a bicistronic gene expression is a process termed “leaky scanning” as observed in HIV env expression in which weak Kozak sequence within the vpu start codon promotes translation of its downstream region encoding envelope. Additionally, in at least six retroviruses including HIV-1, HIV-2, SIV, RSV, MLV, and Harvey murine sarcoma virus (HMSV) existence of an internal ribosome entry sequence (IRES) is believed to play a vital role in initiating translation in a cap independent manner (Yilmaz et al., 2006).

As the unspliced transcript is the precursor for Gag and Gag/Pol protein synthesis, it is important for the virus to have an efficient regulatory system to control their expression levels and to package the same mRNA as gRNA during viral assembly (Melinda Butsch & Boris-Lawrie, 2002). Gag being the structural protein, is needed in much more quantity when compared to the enzymatic Gag/Pol proteins. Generally, it is expected that for every 20 Gag molecules that is being generated, one Gag-Pol polyprotein is made (reviewed in Lever, 2007; Pederson & Duch, 2006). In HIV-1 and RSV this is achieved by ribosomal frameshifting (RFS). The frame shift occurs before the gag stop codon when the ribosome encounters a ‘slippery sequence’ (AU rich region) which makes it to slip by one nucleotide backward (-1) and change the gag reading frame to pol open reading frame (ORF; Figure 16) (Brierley et al., 1989; Mardon & Varmus, 1983; O et al., 2006). In some instances, complementary loop sequences, known as ‘pseudoknots’ found near the slippery sequence on the mRNA impose a pause or slowdown in the translating protein allowing robust ribosomal frameshift to promote synthesis of the proteins encoded by the downstream genes (Brierley et al., 1989; O et al., 2006). Other retroviruses, such as HTLV, MMTV, and MPMV, undergo two frameshifting events firstly by shifting gag termination codon out of frame thus placing the ribosomes into the pro (protease) reading frame and the second shift placing the ribosomes into the pol reading frame (reviewed in (Brierley & Dos Ramos, 2006).

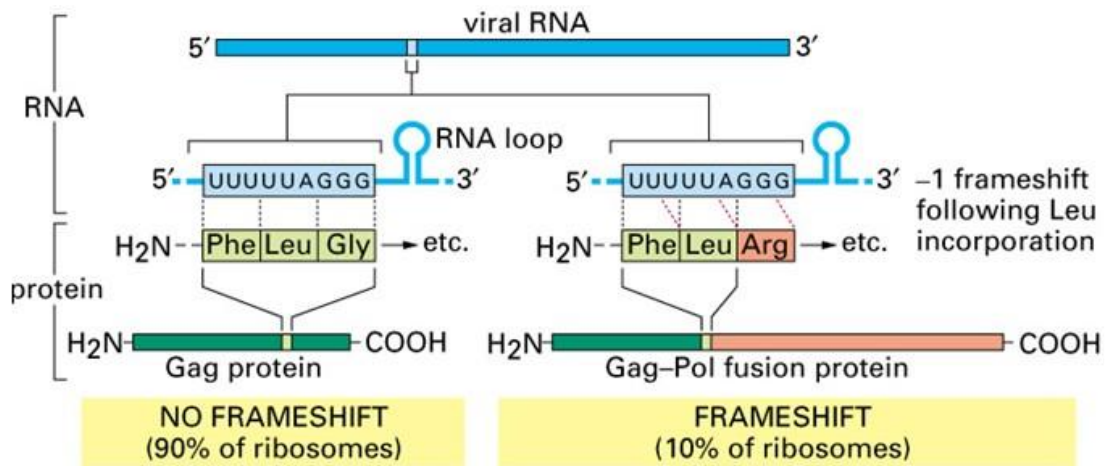


Figure 16. Ribosomal frameshifting in the viral protein synthesis. Eg: HIV-1

Figure adapted from Alberts et al., 2002.

On the other hand, retroviruses like MLV, FeLV, and Walleye dermal sarcoma virus (WDSV) use a termination codon readthrough strategy, by suppressing gag stop codon using a glutamine tRNA during a small fraction of translation events and readthrough the downstream *pol* gene for Gag-Pol polyprotein production. In these cases, *gag* and *pol* genes are in the same reading frame but separated by a termination codon (Honigman et al., 1991; Yoshinaka et al., 1985). Spumaretrovirus is one exception to having the Gag-Pol polyprotein expressed from the same mRNA. Here, the genes are alternatively spliced out into respective discrete mRNAs avoiding the need for ribosomal readthrough or frame shifting (Helps & Harbour, 1997). Table 3 that follows summarizes the known viral and cellular factors affecting retroviral gene expression.

Table 3. Viral and cellular factors affecting retroviral gene expression

Virus	HIV	HTLV	MMTV	MPMV	MLV	RSV
Simple or complex virus	Complex	Complex	Complex	Simple	Simple	Simple
Transcription						
Viral effector	Tat	Tax	None	None	None	None
<i>cis</i> -acting RNA sequence	TAR	TAR	None	None	None	None
Splicing						
Alternative splicing	Yes	Yes	Yes	No	No	Yes
Protein-coding mRNAs ¹	7	4	6 ²	2	2	3
Nuclear export						
Viral effector	Rev	Rex	Rem	None	None	None
<i>cis</i> -acting element	RRE	RxRE	RemRE	CTE	Unknown	DR
Cellular factors	CRM1	CRM1	CRM1	Tap	Tap	Tap
NPC elements	DDX3	Unknown	Unknown	Unknown	Unknown	Dbp5
Translation						
Frame shift sites	One	Two	Two	Two	None ³	One
5' UTR PCE	Present	Present	Unknown	Present	Unknown	Unknown

Tat and Tax, Trans-Activator of Transcription; TAR, *trans*-activating region; RRE, Rev Response Element; RxRE, Rex-responsive element; CTE, Constitutive transport element; DR, Direct repeats; CRM1, Chromosome region maintenance; Tap, Transporter associated with antigen processing; NPC, Nuclear pore complex; DDX3, DEAD-box RNA helicase 3; Dbp5, DEAD-box protein 5; PCE, post transcriptional control element; HIV, human immunodeficiency virus; HTLV, human T-lymphotropic virus; MMTV, mouse mammary tumor virus; MPMV, Mason-Pfizer monkey virus; MLV, murine leukemia virus; RSV, Rous sarcoma virus.

¹Includes only the abundantly expressed mRNAs; minor splice variants are omitted.

²Includes the three distinct *sag* mRNAs, which encode the viral superantigens.

³MLV read through across the *gag-pol* junction occurs via termination codon suppression using a glutamine tRNA.

Contents in table and legend adapted from LeBlanc et al., 2013.

Translation continues until the cellular release factor, eRF1 detects a stop codon and the peptidyl-tRNA is hydrolyzed by the interaction of eRF3 with eRF1 and GTP (Guanosine triphosphate) (Frolova et al., 1996). In MLV, the enzyme reverse transcriptase is reported to have involved to regulate Gag termination and Gag/Pol translation, thereby maintaining an appropriate ratio of Gag:Gag-Pol protein critical for viral infectivity (Felsenstein & Goff, 1988).

1.5.6. The Fate of Precursor mRNA *versus* gRNA

The all-time intriguing question regarding the fate of the unspliced mRNA after the export from the nucleus to the cytoplasm is, what factors trigger the mRNA to be packaged as gRNA *versus* the Gag or Gag/Pol translation? It has been reported that in the case of RSV, the nuclear export pathway opted by the unspliced mRNA determines if the mRNA is encapsidated or translated. The direct repeat (DR) elements (Ogert et al., 1996) dictate the transport of unspliced mRNA through the Tap/Dbp5 route and stimulate Gag translation (refer figure 13; LeBlanc et al., 2007). When enough Gag is produced, it is transiently transported to the nucleus via its NLS and importin associations (Butterfield-Gerson et al., 2006; Gudleski et al., 2010). The imported Gag binds and multimerizes at the packaging signal (D'Souza & Summers, 2005; Maldonado et al., 2020) of the unspliced RNA and this prompts the Gag:RNA complex to choose the CRM1 pathway for its nuclear export (Scheifele et al., 2002, 2005). In other words, Gag binding with gRNA in the nucleus initiates the viral core assembly (Garbitt-Hirst et al., 2009; Scheifele et al., 2007) which is then completed in the cytoplasm (Gudleski et al., 2010; Parent, 2011). In short, the process of translation and gRNA encapsidation in RSV are mutually exclusive (reviewed in Bolinger & Boris-Lawrie, 2009; Shida, 2012). In HIV-1, the gRNA encapsidation occurs at the expense of translation which means that lack of translation does not inhibit Gag binding to its gRNA (Anderson & Lever, 2006; M. Butsch & Boris-Lawrie, 2000; Levin & Rosenak, 1976). In both HIV-1 and MLV, two distinct pools of mRNAs are found, each destined for its role either in Gag production or to serve as gRNA packaging substrate (Figure 17, (Ding et al., 2020; D'Souza & Summers, 2004; Levin & Rosenak, 1976; Levin et al., 1974). The monomeric RNA can exist in two conformers, one in which the DIS is available for a dimer linkage, the Gag AUG is sequestered with a U5 region (AUG:U5) hence inhibiting translation, and promote its packaging. In the other conformational switch, the DIS pairs with U5 (DIS:U5) making the AUG hairpin available for

translation (Figure 17; Ganser & Al-Hashimi, 2016; Keane et al., 2016). Furthermore, recently published studies show that the fate of mRNAs is already established from the starting point of its transcription in the nucleus. The transcripts capped with a single 5'-Guanosine residue favorably form dimers and serve as gRNA whereas the 5' capped transcripts with two or three guanositines form monomers that are destined for splicing and function as mRNAs for translation (Brown et al., 2020; Esquiaqui et al., 2020; Kharytonchyk et al., 2016; Masuda et al., 2015). In HIV-2, however, gRNA packaging and Gag or Gag/Pol translation is not a mutually exclusive process which means that Gag or Gag/Pol polyprotein is produced in a co-translational manner from the same mRNA until enough assembly protein is synthesized to capture the mRNA as its gRNA for packaging into its virions (Kaye & Lever, 1999 and reviewed in Ali et al., 2016).

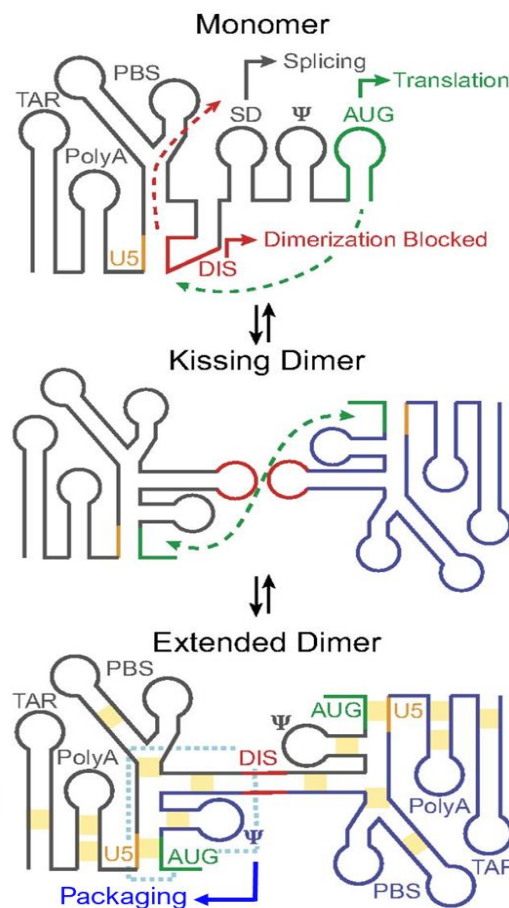


Figure 17. Proposed mechanism for HIV-1 in packaging and translation

In the monomer, the SD and AUG are exposed to promote splicing and translation, respectively. Conversely, the DIS is sequestered by U5, preventing dimerization. In the dimer, DIS dimerization is extended and intermolecular U5:AUG interactions are formed inhibiting Gag translation and promoting packaging. Figure adapted from Ganser & Al-Hashimi, 2016.

1.5.7 Viral assembly, Budding, Release, and Maturation

Following protein synthesis, the process of virion morphogenesis starts in order to produce progeny virions. Retroviral virion morphogenesis consists of three stages: 1) assembly 2) budding and 3) maturation. These stages involve the membrane-targeting (M) and the Gag-Gag interaction (I) for assembly, and the late assembly (L) for the release of the virion from the host cell membrane that are coordinated by the different domains of Gag (Klein et al., 2007, reviewed in Pincetic & Leis, 2009).

1) Assembly: It may begin in the cytoplasm (for MPMV and MMTV), or at the plasma membrane (for HIV and MLV) or even in the nucleus (for RSV) depending on the virus. The virions manage to contain all the essential components for retaining its infectivity and include the dimeric positive sense gRNA, cellular tRNA for priming reverse transcription, the Gag polyprotein and the three viral enzymes: protease (PR), reverse transcriptase (RT), and integrase (IN) along with the viral envelope protein (Figure 18A). The whole events of viral assembly are mediated by the Gag and Gag-Pro-Pol polyprotein. The envelope protein translated in the rough endoplasmic reticulum is N-glycosylated and the SU/TM polyproteins are cleaved by host proteases, furin, in the Golgi apparatus before it is recruited to the host membrane. The envelope proteins reach the plasma membrane independent of Gag, however, the TM domain acts as a mediator to attract the matrix (MA) (Cosson, 1996; Murakami & Freed, 2000; Wyma et al., 2000; Yu et al., 1993), directing Gag nucleation for assembly (Chukkapalli et al., 2010; Datta et al., 2011; Jones et al., 2011). Membrane targeting by Gag also requires myristoylation of the amino-terminal domain of MA and the cooperation of its highly basic region (HBR) with the membrane specific phosphatidyl inositol biphosphate (Freed, 2015; Murphy & Saad, 2020; Olety & Ono, 2014; Ono et al., 2004). This induces a conformational change in the membrane and helps in anchoring of the assembling virions. The capsid (CA), which is the central domain of Gag mediates the protein-protein interaction for the formation of the viral core. Polymerization of the Gag initiates in the cytoplasm and intensifies at the membrane (Kutluay & Bieniasz, 2010). In the cytoplasm, monomers of Gag proteins fold into a compact structure, undergoes conformational changes to facilitate cooperative binding of MA to the membrane, NC:RNA, and Gag:Gag

interactions and all these events occur simultaneously during the assembly process (Chukkapalli et al., 2010; Datta et al., 2011; Jones et al., 2011).

The C-terminal nucleocapsid (NC) domain contain two zinc finger motifs that specifically bind to the packaging signal sequences of the gRNA and captures it into the assemblages. Polymerization of Gag at the Gag-RNA nucleation site (Jouvenet et al., 2009) triggers the budding and release of the progeny virions (Figure 18A).

2) Budding: Retroviruses greatly remodel their host cytoskeleton for its release (Gladnikoff et al., 2009). To facilitate this, short peptide motifs, or L domains encoded by retroviruses in the Gag polyprotein target the cellular machinery known as the Endosomal Sorting Complex Required for Transport (ESCRTs), responsible for vesicular trafficking within the cell (Figure 18A; Demirov & Freed, 2004 and reviewed in Rose et al, 2020; Martin-Serrano & Neil, 2011; Sundquist & Kräusslich, 2012). This terminates the Gag polymerization and trigger the viral release (Carlton & Martin-Serrano, 2007; Hurley & Hanson, 2010; Morita & Sundquist, 2004; Perez-Caballero et al., 2009; Usami et al., 2009).

The short peptides within the L domain are majorly of three distinct core sequences, PPXY, PTAP, and YPXL, identified in different retroviruses (Freed, 2002). The first indication of the role of L domains came from RSV and HIV-1 studies and were mapped to proline-rich sequences, PPPY in the p2b (Wills et al., 1994; Xiang et al., 1996) and PTAP in p6 (Huang et al., 1995) region of Gag, respectively. In some retroviruses tandem repeats of these amino acid motifs are found indicating a redundant or synergistic role in efficient viral particle release from various types of cells infected. For instance, the MPMV (Gottwein et al., 2003) and HTLV-1 (Blot et al., 2004) contain tandem PPXY and PTAP motifs, of which one motif is found dominant over the other in executing its function. In both these viruses, PPPY motif is shown to be dominant. Similarly, HIV-1 Gag contains a secondary YPXL-type motif downstream of its dominant PTAP sequence (Strack et al., 2003). Each L domain interact differently with distinct proteins of E Vps protein family of ESCRT I, II and III complexes that function in the biogenesis of multivesicular

bodies (MVBs) in the eukaryotic cells that are destined for ubiquitination and degradation of protein cargo via endosome formation. Retroviral budding is topologically equivalent to these normal cellular protein degradation process (Hurley & Emr, 2006, reviewed in Pincetic & Leis, 2009).

3) Maturation: The final step in producing infectious viral particles is the cleavage of the Gag and Gag-Pro-Pol polyproteins by viral protease (PR), ultimately making the fully processed MA, CA, NC, PR, RT, and IN proteins (Figure 18B & C; Hill et al., 2005; Swanstrom & Wills, 2011). The immature core proteins rearrange significantly to create the characteristic core morphology. These structural changes render the virus mature with a more compact and stabilized RNA genome within an orderly arranged capsid and prepare the virion to begin the next infection cycle (Figure 18D-G; Sundquist & Kräusslich, 2012).

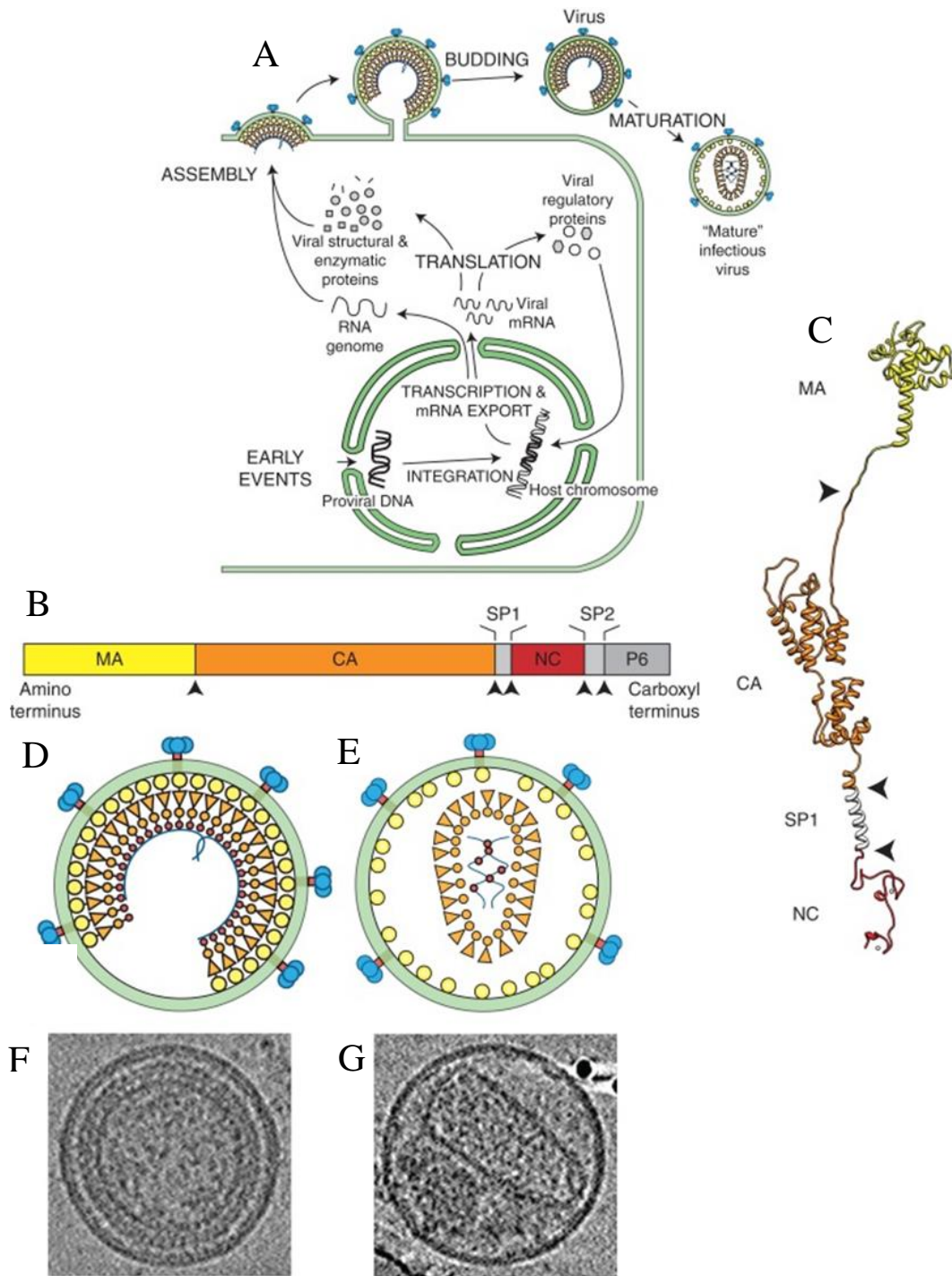


Figure 18. HIV-1 assembly, budding, and maturation

(A) Schematic illustration showing the different stages of HIV-1 assembly, budding, and maturation. (B) Domain structure of the HIV-1 Gag protein; arrows denote the five sites that are cleaved by the viral PR during maturation. (C) Structural model of the HIV-1 Gag protein. (D) Schematic model showing the organization of the immature HIV-1 virion. (E) Schematic model showing the organization of the mature HIV-1 virion. (F) Central section from a cryo-EM tomographic reconstruction of an immature HIV-1 virion. (G) Central section from a tomographic reconstruction of a mature HIV-1 virion. Figure and legend courtesy Sundquist & Kräusslich, 2012.

1.5.8 Retroviral gRNA Packaging

Genomic RNA packaging and assembly of Gag polyprotein occurs simultaneously at the plasma membrane or in the cytoplasm depending upon the type of the virus. Selective packaging of the viral gRNA from a large milieu of cytoplasmic RNAs and spliced viral mRNAs is brought about by the specific interaction of the NC domain of the Gag with the *cis*-acting packaging sequences (*psi*- ψ) at the 5' end of the gRNA (Abd El-Wahab et al., 2014; Ali et al., 2016; Berkowitz et al., 1996; Bernacchi et al., 2017; D'Souza & Summers, 2005; Dubois et al., 2018; Johnson & Telesnitsky, 2010; Kaddis Maldonado & Parent, 2016; Kuzembayeva et al., 2014; Mailler et al., 2016; Miyazaki et al., 2011). These indispensable *psi* (ψ) sequences acquire higher-order structures and comprises the region from R and extend variably into the 5' sequences of the *gag* gene (Figure 19A & B; Ali et al., 2016; Berkhout & van Wamel, 1996; Browning et al., 2003; Clever et al., 2002; Comas-Garcia et al., 2016; Das et al., 1997; D'Souza & Summers, 2005; Dubois et al., 2018; Jewell & Mansky, 2000; Johnson & Telesnitsky, 2010; Jouvenet et al., 2011; Kaddis Maldonado & Parent, 2016; Kenyon et al., 2008, 2011; A. M. L. Lever, 2007; Mailler et al., 2016; Mustafa et al., 2005; Tahir A. Rizvi et al., 2010). The packaging determinants in retroviruses can either be continuous or discontinuous (bipartite or multipartite) at the primary sequence level (Figure 19; reviewed in Ali et al., 2016) and form highly structured RNA motifs that play important roles in gRNA packaging. Any disruption in its native structure abrogates gRNA packaging (Ali et al., 2020; Kalloush et al., 2016; Kharytonchyk et al., 2018). Interestingly, it has been shown that among closely related viruses, these packaging motifs can functionally be exchanged to promote cross/co packaging of each other's genomes implicating a structural role than the primary sequence in viral gRNA packaging (Figure 19; Al Dhaheri et al., 2009; Al Shamsi et al., 2011; Browning et al., 2001; Das et al., 1997; Embretson & Temin, 1987; Rizvi & Panganiban, 1993; White et al., 1999; Yang & Temin, 1994; Yin & Hu, 1997).

necessary elements required for the specific selection and direction of the genomic RNA to the viral assembly machinery (Figure 20; Balvay et al., 2007).

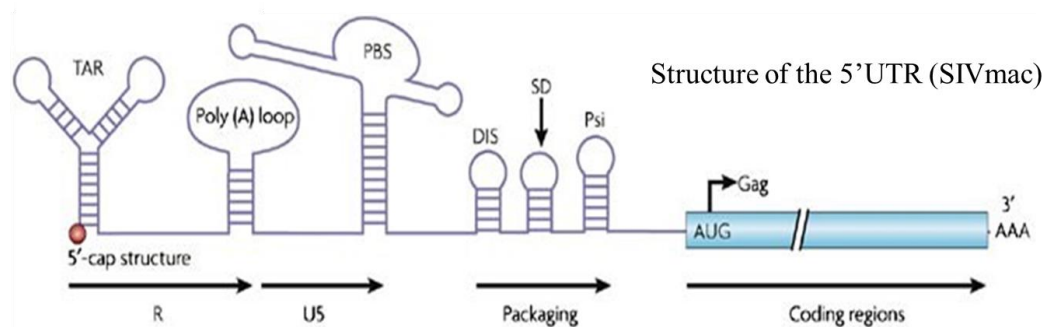


Figure 20. Genetic organization of the 5'-untranslated region (UTR) of the retroviral genomic RNA- example of SIVmac

The 5'-UTR (or leader) of the retroviral genomic RNA contains numerous cis-acting sequences, such as the TAR stem-loop, the poly (A) loop, the primer-binding site (PBS), the dimerization initiation site (DIS), the core packaging signals (*psi*) and the major splice donor (SD). The R-U5, packaging and coding regions are indicated in the figure.

The packaging sequences also contain high affinity binding sites for NC and are predominantly purine rich sequences (adenosine (A) or guanosine (G) (Abd El-Wahab et al., 2014; Lever, 2009; Moore et al., 2009; Moore & Hu, 2009; Paillart et al., 1997) while occasionally uracil (U) residues are also found as part of their binding sites (Gherghe et al., 2010; Zhou et al., 2005, 2007) which may confer to the viral group specificity of NC to its cognate/non cognate gRNA. Nucleocapsid domains of all retroviruses, except spumaviruses, contain one or two zinc finger domains allowing this high affinity interaction with gRNA. These evolutionary conserved, yet diverse domains contain Cys-His boxes (CCHC arrays, C-X₂-C-X₄-H-X₄-C where C = Cys, H = His, X_n = n number of amino acids) that chelate zinc ions allowing for stable NC:gRNA interactions. Spumaviruses contain arginine-rich motifs instead of the zinc fingers. Recent studies to elucidate the structural characteristics of these NC:gRNA complexes reveal that NC binding occurs to a dimeric gRNA, signaling for Gag multimerization and virion assembly thereafter (reviewed in Ali et al., 2016; Johnson & Telesnitsky, 2010; Jouvenet et al., 2011). From the vast biochemical data presented till date, HIV-1 RNA is known to contain more than two dozen of NC binding sites. Nuclear magnetic resonance (NMR) combined with isothermal Titration Calorimetry

(ITC) studies indicate that these binding sites may possess varied affinity and functional implications (Heng et al., 2012; Keane et al., 2015; Lu, Heng, Garyu, et al., 2011). Earlier studies on HIV-1 NC binding showed that its interaction with gRNA is sequence specific (Summers et al., 1992) and requires at least one guanosine residue to the tight binding of NC zinc finger knuckles (South & Summers, 1993). The HIV-1 NC zinc knuckles are found to bind to the SL1(stem loop 1), SDSL (stem loop containing the major splice donor) and ψ SL (SL3) regions of the 5' end of HIV-1 RNA in a similar manner (Ding et al., 2020). However, NC strongly binds to the UUUU:GGAG stem region of the ψ SL with a high affinity *in vitro* and is shown to be required for gRNA packaging in transfected cells and determines the specificity during gRNA selection (Ding et al., 2020). Similar scenario occurs in MLV, where the MLV NC zinc knuckles interact with the exposed guanosines bases but the tight binding is not restricted to a single stem loop as in the case of HIV-1 (D'Souza et al., 2001). MLV NC requires its intact core encapsidation signal (CES) comprising all three SLs (ψ CES) and binding occurs on a dimer following non-covalent interaction of two monomeric RNA molecules. These studies revealed that the dimerization induced shift in the RNA structure exposes the NC-specific binding site UBUG that acts as a linker between the two important adjacent SLs (D'Souza et al., 2001).

However, number of studies with chimeric Gag proteins with its cognate and non-cognate retroviral genomes indicated that though NC binding is specific, it may not be the sole determinant for packaging of the respective viral genomes. Besides NC, other domains of Gag like the MA (Lu, Heng, & Summers, 2011), CA (Guo et al., 2005), p2 peptide (Kaye & Lever, 1998) and p6 (Dubois et al., 2018) in the case of HIV-1, are shown to be involved in selective packaging of its gRNA from the spliced RNAs. Contrastingly, for some delta-retroviruses such as HTLV-1 and HTLV-2, the MA domain has been shown to exhibit a stronger binding affinity than their NC component and may contribute to more than just membrane targeting (Wu et al., 2018). Altogether, these findings suggest that the entire Gag domains play a significant role in the initial events of gRNA recognition and discriminates spliced over unspliced messages during virus assembly.

1.5.9 Retroviral gRNA Dimerization

It is now well understood that retroviral NC binds to two molecules of gRNA that are non-covalently linked at their 5' end. A close observation of the packaging signal of various retroviruses reveal that the NC binding sites lie near a palindromic stem loop (Pal SL) harboring DIS. The DIS has been shown to be involved in initiating dimerization by complementary Watson-crick base pairing between the two RNAs creating a 'kissing loop' interaction (Figure 21; Laughrea et al., 1997; Moore & Hu, 2009; Paillart et al., 1996). It is now becoming vivid that, most retroviruses except spumaviruses (Cain et al., 2001, reviewed in Dubois et al., 2018) require gRNA dimerization prior to RNA packaging. Coupling of dimerization and gRNA packaging is best studied in MLV and HIV-1 systems (Figure 21). The MLV monomeric RNA contains several NC binding motifs (UCUG), which are otherwise occluded, and get exposed only after RNAs have undergone a conformational change following dimerization with a second RNA molecule. Such conformational changes in the RNA and lead to the unpairing of UCUG motifs facilitating binding of NC and thereby the packaging of the dimeric gRNA (Figure 21; D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010). The process of dimerization occurs in the cytoplasm for viruses like HIV-1 and FIV and consequently can form homodimers or heterodimers between the viral genomes. However, MLV and RSV that dimerize in the nucleus and can form only homodimers (Figure 21; reviewed in Ali et al., 2016; Johnson & Telesnitsky, 2010; Jouvenet et al., 2011). Co-packaging of gRNAs as homodimers may be a result of packaging specificity as well as to promote genomic integrity. Moreover, the packaging of two genetically distinct gRNAs as heterodimers promotes reassortments, leading to genetic diversity in viral progeny (reviewed in Ali et al., 2016; Johnson & Telesnitsky, 2010).

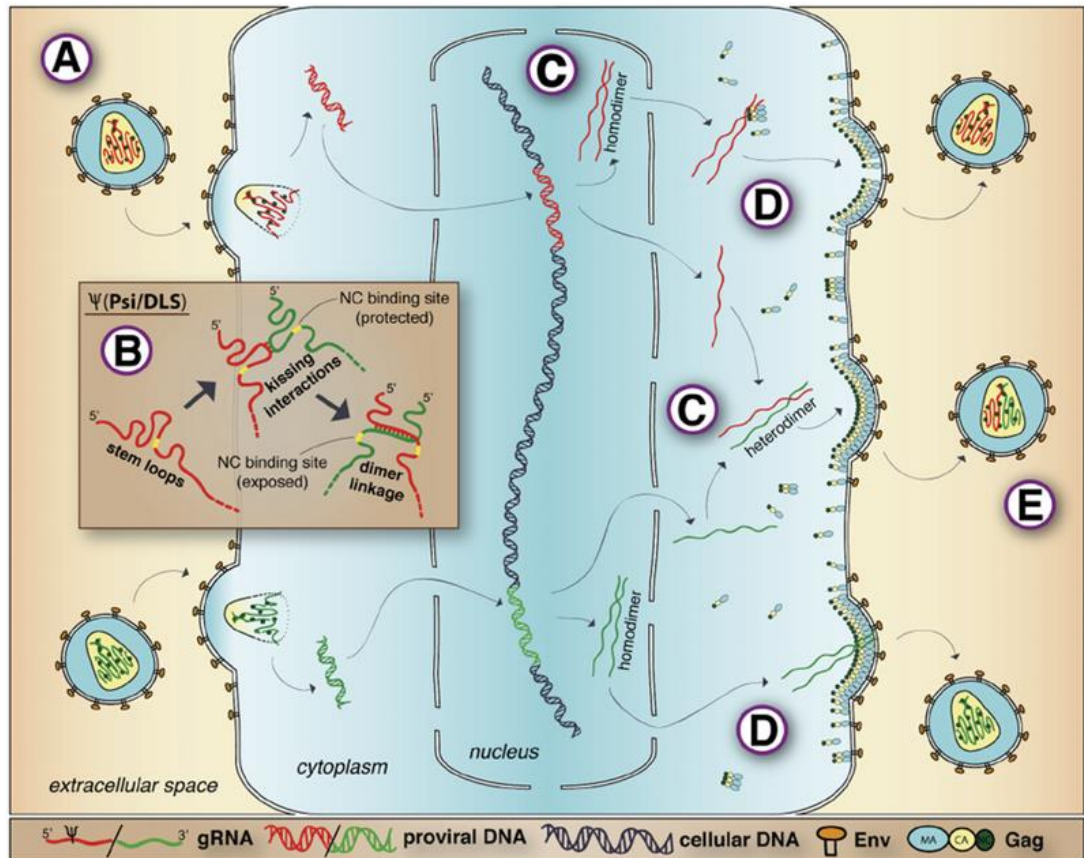


Figure 21. Retroviral RNA dimerization and packaging

The description of retrovirus genomic RNA dimerization and packaging is based on a representative co-infected cell and depicts properties of C-type retroviruses such as HIV-1 and MLV. Note that this figure represents concepts schematically and is not intended to accurately represent structures or scale. **(A)** Two genetically complete but nicked copies of plus sense gRNA (shown in red at top or green at bottom) are packaged within the capsid core and joined by a dimer linkage. The co-packaged gRNAs are condensed in the core and bound by NC (shown as green circles). **(B)** Initial gRNA dimerization occurs via kissing interactions between palindromic stem loops. Subsequent basepairing register-shifts that occur during dimer linkage maturation expose single-stranded NC binding motifs (indicated in yellow) that were previously basepaired and thus sequestered in gRNA monomers, allowing for Gag binding during recruitment. **(C)** The point at which RNA dimerization partners first associate determine the population of homodimers or heterodimers in a virus. MLV gRNA dimers first associate near sites of transcription in the nucleus, which leads to disproportionately large amounts of one homodimer or the other (shown in red at top and green at bottom). HIV-1 gRNA dimers first associate in the cytoplasm, leading to a random assortment of homodimeric and heterodimeric gRNAs. **(D)** gRNAs may form subassemblies with Gag in the cytoplasm (shown at top) or may associate at the plasma membrane (shown at bottom). **(E)** Packaging of gRNAs as homodimers and heterodimers. Figure and legend courtesy (Johnson & Telesnitsky, 2010).

1.6 The MPMV Packaging Determinants for gRNA Packaging

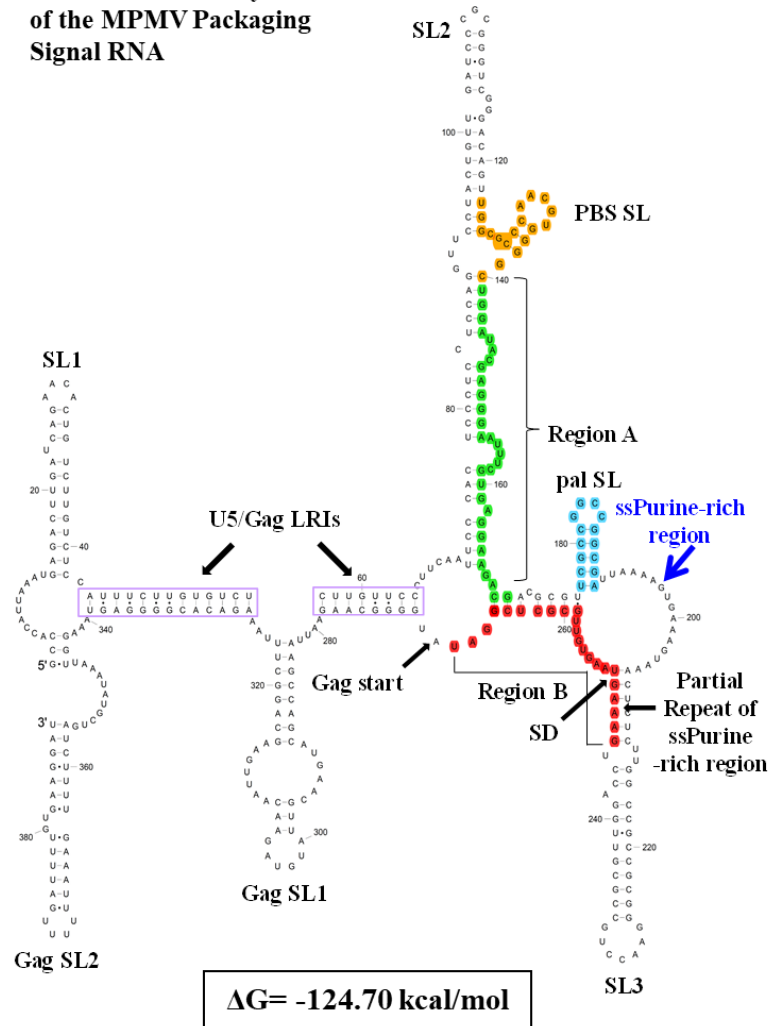
Among the *Betaretroviruses*, MPMV is perhaps the most well-investigated in terms of its gRNA packaging and dimerization. This is because MPMV-based vectors are good candidates for use in human gene therapy since they offer several advantages. For example: i) they are phylogenetically distinct from human retroviruses, thus avoiding/limiting issues related with recombination with related viruses, ii) MPMV despite being non-human primate virus can be expressed efficiently in human cells which is an extremely important criterion for human gene therapy, and iii) presence of post-transcriptional regulatory elements (CTE) in MPMV vectors may facilitate efficient cytoplasmic expression of the therapeutic genes (Bray et al., 1994; Pasquinelli et al., 1997; Rizvi et al., 1996a; Rizvi et al., 1996b; Zolotukhin et al., 2001). Because of the possible use of MPMV-based vectors in human gene therapy, its replicative biology has been investigated extensively with major emphasis on gRNA dimerization, packaging, and RNA propagation (Aktar et al., 2013; Guesdon et al., 2001; Harrison et al., 1995; Jaballah et al., 2010; Kalloush et al., 2016, 2019; Mustafa et al., 2004; Pitchai et al., 2018; Schmidt et al., 2003; Vile et al., 1992). Initial studies on MPMV by Vile et al. located a 624 nt region downstream of PBS to be important for packaging (Vile et al., 1992). Further extension of work towards establishing the packaging signal for MPMV, identified a stretch of sequences at the 5' end of the viral gRNA important for genome packaging that included sequences from R to the first 120 nt of the *gag* ORF, (Guesdon et al., 2001; Harrison et al., 1995; Mustafa et al., 2004; Schmidt et al., 2003). Later, a more systematic mutational analysis of the 5' end of the MPMV genome revealed that the MPMV packaging determinants were bipartite and resided within two distinct regions: region "A" which included the first 50 nt of the 5' untranslated region (UTR) and region "B" that encompassed the last 23 nt of the 5' UTR as well as the first 120 nt of *gag* ORF (Jaballah et al., 2010; Schmidt et al., 2003). When sequences between R and the first 120 nt of *gag* ORF harboring MPMV packaging determinants were folded (using Mfold) the RNA secondary structure predicted several stable stem loop structures (Figure 22A; Jaballah et al., 2010).

To further validate the predicted RNA secondary structure, a novel chemoenzymatic probing strategy known as Selective 2' Hydroxyl Acylation by Primer Extension (SHAPE) methodology was employed. SHAPE analysis validated the overall predicted structure and confirmed that this region folds into a higher-order

structure comprising several stable structural motifs (Figure 22A & B; Aktar et al., 2013; Jaballah et al., 2010). Distinguishing features of the structure included a palindromic stem loop (Pal SL) harboring DIS (Aktar et al., 2013) and U5-Gag long range interactions (LRIs) that were found to be important for anchoring the structure of the entire 5' leader region (Figure 22A & B; Kalloush et al., 2016). The SHAPE-validated higher order structure also revealed a stretch of single stranded purine-rich sequence (ssPurines; UUAAAAGU GAAAGUAA) in close proximity of the Pal SL (Aktar et al., 2013; Jaballah et al., 2010). Interestingly, part of the ssPurine-rich sequence was observed to be duplicated as a base-paired sequence at the base of SL3 (named here as "bpPurines"; GAAAGUAA) in region "B" (Figure 22A & B; Aktar et al., 2013; Jaballah et al., 2010).

A

Predicted Secondary Structure of the MPMV Packaging Signal RNA

**B**

SHAPE Validated Secondary Structure of the MPMV Packaging Signal RNA

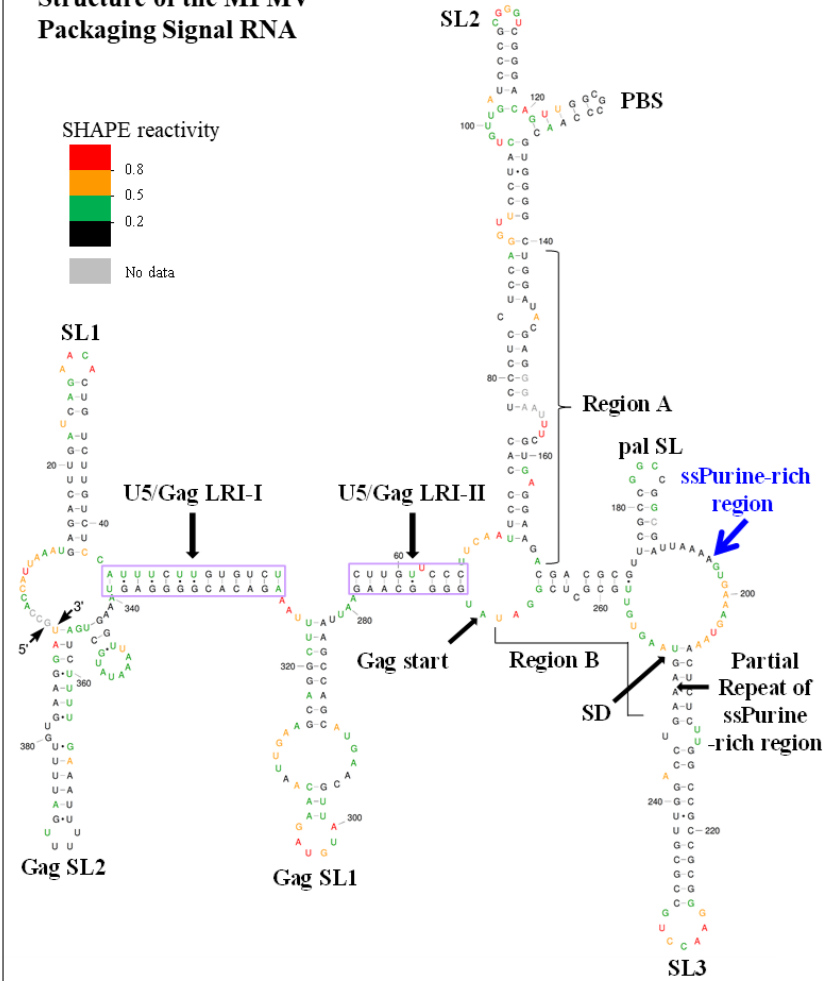


Figure 22. Minimal free-energy and SHAPE (selective 2' hydroxyl acylation analyzed by primer extension)-validated models of the MPMV packaging signal RNA

The region used for analysis by Mfold and SHAPE included sequences from R up to 120 nt of gag. **(A)** MPMV packaging signal RNA secondary structure predicted earlier (Jaballah et al., 2010) using Mfold (Mathews et al., 1999; Zuker, 2003). Sequences in orange, green, red, and blue represent the primer binding site (PBS), regions “A” and “B” (that have been shown to be important in gRNA packaging), and pal sequences, respectively. Boxed areas in purple show the predicted LRIs between n U5 and gag. **(B)** SHAPE-constrained RNA structure (Reuter & Mathews, 2010) model of MPMV packaging signal. Nucleotides are color annotated as per the SHAPE reactivities key. SD indicates splice donor. Figure adapted from Aktar et al., 2013.

The presence of purine-rich sequences in retroviral packaging signal RNA has been proposed to facilitate gRNA packaging by functioning as a potential Gag binding site. In HIV-1 the internal loop (G//AGG) in SL1 (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Smyth et al., 2015, 2018), GGRG motif in HIV-2 (Baig et al., 2009; Damgaard et al., 1998), and GGAGAAGAG purine rich sequence in MMTV (Mustafa et al., 2018; Aktar et al., 2014; Chameettachal et al., 2021) have been shown to function as high affinity Gag binding sites. The precise role of ssPurines in MPMV life cycle has not been validated, despite indirect evidence for their role in MPMV gRNA packaging (Jaballah et al., 2010). Deletion of a 20 nts region which included 16 nts of ssPurines (mutant SJ44; (Jaballah et al., 2010) predicted an interesting conformation in which the bpPurines assumed a single stranded conformation to restore packaging to a level more than 50% of the wildtype. This led to the assumption that the bpPurines may compensate for loss of ssPurines and may potentially function as the Gag binding site in the absence of ssPurines to augment gRNA packaging (Figure 23; Jaballah et al., 2010).

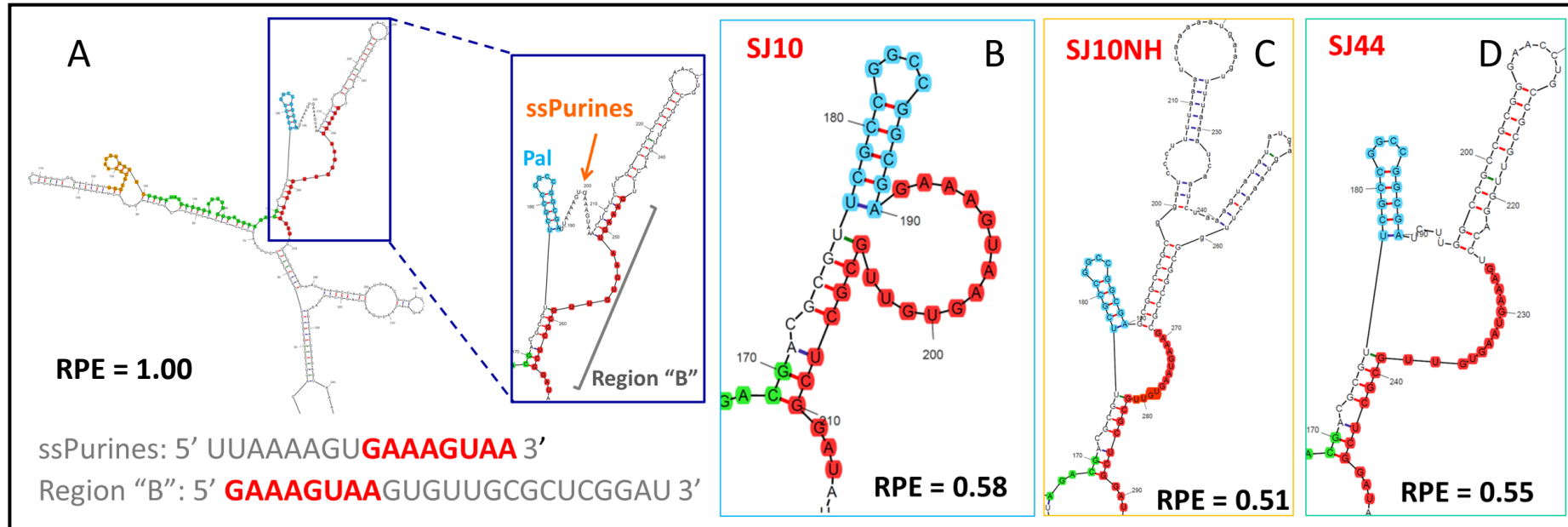


Figure 23. Role of ssPurines in MPMV genomic RNA packaging

(A) Mfold structure prediction of the 5' end of MPMV RNA showing Pal (blue), ssPurines and region B (red) in the enlarged inset. B, C and D- Mutants containing the deletion of ssPurines and respective predicted structures showing the opening of bpPurines that may have contributed to the packaging restoration. RPE, Relative packaging efficiency. Figure adapted and modified from (Jaballah et al., 2010).

1.7 Objectives

Understanding the molecular mechanisms involved in MPMV gRNA packaging process is vital for developing MPMV-based vectors for human gene therapy. Recognizing the critical segments of genomic RNA involved in its selective packaging from a pool of diverse cellular and viral spliced mRNAs is the foremost step towards achieving this goal. During the assembly of MPMV particles, the precursor Gag protein must specifically select the dimerized viral gRNAs from a variety of cellular and spliced viral RNAs. However, not much is known about the Gag-RNA interaction site(s) within the MPMV packaging determinants. Therefore, the overall goal of this thesis was to gain better understanding of the MPMV gRNA packaging process by delineating the mechanism(s) involved in the initial recognition of the gRNA by addressing the following specific aims:

Specific Aim I: Establish the role of putative ssPurines sequence in MPMV gRNA packaging. It is well known in retroviruses that a stretch of purines may act as a potential Gag binding site during the early recognition of its cognate gRNA for packaging. Along the same lines, MPMV is shown to harbor a stretch of 16 nucleotides (ssPurines) upstream of its mSD and in close proximity to the DIS (Aktar et al., 2013; Jaballah et al., 2010). Therefore, to test the function of this purine-rich motif, a series of mutations were introduced and cloned into a subgenomic transfer vector. These ssPurines mutations were then tested employing a biologically relevant *in vivo trans* complementation assay that has been previously described (Browning et al., 2001) to monitor their effects on transfer vector RNA packaging and propagation.

Specific Aim II: Establish the role of duplicated base-paired sequence of ssPurines (bpPurines) sequence in MPMV gRNA packaging. An earlier study has shown that a deletion of a 20 nts region which included 16 nts of ssPurines did not abrogate MPMV gRNA packaging and propagation which suggested that bpPurines may compensate for loss of ssPurines by functioning as potential Gag binding site in a redundant fashion (Jaballah et al., 2010). Therefore, a series of mutations were introduced in bpPurines both in the absence as well as presence of ssPurines. The effects of these mutations in bpPurines on MPMV gRNA packaging and propagation was monitored using already established *in vivo trans* complementation assay (Browning et al., 2001).

Specific Aim III: Establish structure-function relationship of the mutations introduced into ssPurines as well as in its partially duplicated bpPurines during MPMV RNA packaging. To better understand the effects of the mutations introduced in ssPurines and in its partial duplicated sequence (bpPurines), structure-function correlation was established by validating the predicted structure of the introduced mutations. Towards this end mutations in these ssPurines as well as in its duplicated base-paired purines were cloned into a T7 expression plasmid to *in vitro* transcribe mutant and wild type RNAs that were used in a novel biochemical probing technique that examines the flexibility of each nucleotide (a surrogate marker for base-pairing) called SHAPE. Together, these approaches should provide clues on the role of the structure of ssPurine region as well as its partially duplicated base-paired counterpart in the context of the overall higher order structure during MPMV RNA packaging.

Specific Aim IV: Correlate the obtained biological data to the Gag precursor protein (Pr78^{Gag}) binding site(s) in MPMV gRNA packaging determinants. Finally, the results obtained towards the investigation of the specific aims for this thesis were correlated to the recently published work on finding the specific Gag precursor binding site(s) on MPMV gRNA (Pitchai et al., 2021). The findings from Pitchai, 2021 corroborated well with the biological data and helped in proposing a model on how MPMV discriminates the full-length gRNA from its spliced or cellular mRNAs during genome encapsidation.

CHAPTER 2: MATERIALS AND METHODS

2.1 MPMV Strain, Nucleotide Designations and Plasmid Construction

Expression plasmids and MPMV-based vectors were derived from pSHRM15 plasmid (a gift from Dr. Eric Hunter, Emory University, United States) containing sequences of the molecular clone MPMV/6A. The nucleotide numbers refer to the GenBank accession number M12349 (Sonigo et al., 1986).

Site-directed mutations were introduced into the *psi* region (from R in the 50 LTR to 282 nt of gag) of the MPMV sub-genomic vector, SJ2 (Jaballah et al., 2010). SJ2 expresses the hygromycin B phosphotransferase gene from an internal simian virus 40 early promoter (SV-Hygr) that helps to monitor the effect of mutations introduced into this transfer vector upon propagation of the packaged viral RNA. Splice overlap extension (SOE) PCR (Gibbs et al., 1994) was employed to introduce the desired mutations. Briefly, two rounds of PCR were conducted using SJ2 as a template to generate intermediate products containing the introduced mutations. Cloning of the mutants was achieved by using an outer sense (S) primer, OTR787 (containing an XhoI site at the 5' end) and an outer antisense (AS) primer, OTR788 (containing a BamHI restriction site at the 3' end) along with internal overlapping primers containing the desired mutations, as previously described (Jaballah et al., 2010; Kalloush et al., 2016, 2019). This resulted in the generation of mutant *psi* fragments containing appropriate flanking cloning restriction sites (XhoI at the 5' end and BamHI at the 3' end), which were used to clone the fragments containing the mutations into SJ2, resulting in clones LA-I to XI. Details of the oligonucleotides and the templates used for making the mutations are listed in Appendix A while the nature of mutations introduced into each clone are described in Table 4.

For the *in vitro* dimerization and SHAPE assays, the *psi* region harboring the wild-type and the above-mentioned mutations was also cloned into a pUC-based cloning vector, pIC19R (Marsh et al., 1984) to generate the corresponding *in vitro* RNA transcribing clones. These clones were created by using the corresponding SJ2-based mutants as templates which were amplified using the outer primers, OTR1004 (S) (containing a HindIII and the T7 RNA polymerase promoter sequence, upstream to the MPMV R sequence) and OTR1005 (AS; containing MPMV gag sequences nt

1171-1151 with XmaI/SmaI site at the 3'end). The PCR products were cleaved with HindIII and XmaI and ligated to pIC19R that had been previously digested with the same restriction enzymes, resulting in the wildtype (RCR001; (Aktar et al., 2013; Kalloush et al., 2016)) and mutant clones capable of *in vitro* transcription from the T7 promoter. *In vitro* RNA transcribing clones containing the mutations were named as FN-I to IX. All the clones were screened by restriction digestion and confirmed by sequencing (refer to Appendix A for oligos used in sequencing).

2.2 Cell Culture, Transfection, and Infection

Human embryonic kidney cells 293 (HEK 293T) were used for virus production. These cells were maintained at 37°C in the presence of 5% CO₂ in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) heat-inactivated fetal bovine serum. To monitor the propagation efficiency of the transfer vector RNAs, the human cervical cancer cell line HeLa T4 was used and maintained at 37°C in the presence of 5% CO₂ in DMEM supplemented with 7% (v/v) heat-inactivated fetal calf serum.

The SJ2-based mutants were tested *in vivo* to observe the effect of the introduced mutations on gRNA packaging and propagation. Toward this, the MPMV sub-genomic wild-type and mutant transfer vectors that contained the cis-acting sequences needed for genome replication, including transcription, polyadenylation, encapsidation, reverse transcription, and integration, along with the packaging construct, TR301 which expresses Gag-Pol proteins (Browning et al., 2001) and the vesicular stomatitis virus glycoprotein G envelope expression plasmid, MD.G (Naldini et al., 1996), were used in a three-plasmid *trans*-complementation assay, as has been described earlier (Browning et al., 2001; Jaballah et al., 2010; Kalloush et al., 2016, 2019). Transfections of HEK 293T cells with the aforementioned plasmids were carried out using the calcium phosphate transfection method (Invitrogen, United States) according to the manufacturer's protocol along with pSEAP2-Control vector. The pSEAP2- Control vector expresses secreted alkaline phosphatase (SEAP) and was used to normalize for transfection efficiency of the assay using Great EscAPE SEAP Chemiluminescence kit 2.0 (Clontech, United States), as described previously (Chameettachal et al., 2018; Krishnan et al., 2019; Pitchai et al., 2018). 72h post-transfection, the pseudotyped virus particles produced from HEK 293T cells were

harvested and clarified of cellular debris using low-speed centrifugation, and a portion of it was used to infect HeLaT4 target cells in the presence of 1 mg/ml DEAE dextran, a polycation polymer, to enhance infection efficiency. 48 h post-infection; cells were selected with DMEM medium containing 200 mg/ml of hygromycin B antibiotic (Hyclone, United States) for 10–12 days in order to monitor transfer vector propagation efficiency. The number of resulting hygromycin-resistant colonies obtained per milliliter of the virus supernatant (CFU/ml) is a measure of the propagation efficiency of each mutant transfer vector RNA, which was further normalized to the transfection efficiency using SEAP values. Finally, the resulting normalized CFU/ml was reported relative to the wild-type viral titers. Such *in vivo* packaging and propagation assays allowed us to quantify the effects of the introduced mutations on both RNA packaging as well as its propagation without any uncertainty, since the defective nature of the virions produced limited the assay to a single round of replication.

The remaining virus supernatant was clarified of cellular debris by passing through a 0.22-um cellulose acetate syringe filter and then ultracentrifuged at 70,000 x g for 2 h at 4°C on a 20% (w/v) sucrose cushion to concentrate the virus particles, as described previously (Jaballah et al., 2010; Kalloush et al., 2016, 2019; Pitchai et al., 2018). The pelleted virus particles were resuspended in TNE buffer [0.25 M Tris (pH 8.0), 0.1 M NaCl, 0.001 M EDTA] and stored in Trizol Reagent (Invitrogen, United States) for virion RNA isolation and subsequently used for real-time quantitative PCR (RT-qPCR).

2.3 RNA Extraction and Reverse-Transcriptase Polymerase Chain Reaction (RT PCR)

HEK 293T cells transfected with different mutant and/or wild-type vectors were harvested from the six-well plates, and a portion of cells were fractionated into nuclear and cytoplasmic fractions, as described previously (Figure 25; Akhlaq et al., 2018). The cytoplasmic and packaged viral RNAs were isolated from Trizol following manufacturer's instructions (Invitrogen, United States). Two micrograms of cellular RNA and the entire virion RNA prep from a six well plate was subjected to DNase-treatment (Turbo Dnase, Ambion, United States), followed by PCR amplification with MPMV-vector specific primers OTR1161 and OTR1163 (Appendix A) to confirm the lack of any carryover plasmid contamination, as described previously (Kalloush et al.,

2016; Pitchai et al., 2018). The samples were then converted into cDNAs using M-MLV reverse transcriptase (Promega, United States), and random hexamers, as described earlier (Jaballah et al., 2010; Kalloush et al., 2016). The cDNA samples thus generated were subjected to multiplex PCR using primers specific for unspliced β -actin (OTR582/OTR581; Appendix A) and 18S ribosomal RNA primer/competimer (QuantumRNA 18S Internal Standards, Ambion) to monitor amplifiability of the cDNAs as well the integrity of the nuclear membrane during the fractionation process. Further PCRs were conducted to analyze spliced β -actin on cytoplasmic cDNAs (OTR582/OTR580; Appendix A) and MPMV-vector specific primers (OTR1161 and OTR1163) to confirm the expression of cytoplasmic and viral-specific cDNAs.

2.4 Determination of the Relative Packaging Efficiencies of Transfer Vector by Real-Time Quantitative PCR (RT-qPCR)

The RPE of the transfer vector RNAs was determined by conducting RT-qPCR on cDNAs from the cytoplasmic and viral RNA samples. Toward this end, we used a pre-validated MPMV-based custom Taqman gene expression assay developed for this purpose (Kalloush et al., 2016, 2019). Relative quantification of the cytoplasmic and viral packaged RNAs was obtained after normalizing to the endogenous control for which a predesigned human β -actin assay (MGB-FAM) was used, as described previously (Kalloush et al., 2016, 2019; Mustafa et al., 2012). Briefly, equal amounts of cytoplasmic and viral cDNA samples were tested for MPMV and β -actin expression in triplicates for 50 cycles using the QuantStudioTM7 Flex System (Applied Biosystems, Foster city, CA, United States). The relative quantitation (RQ) values for MPMV expression in the mutants were compared to the wild type to get the relative expression of each mutant vector RNA, both of which were normalized to the β -actin expression levels in each sample. Finally, to determine the packaging efficiency of each vector RNA, the ratio of the viral RQ to the corresponding cytoplasmic RQ for each sample was estimated, and the values represented relative to the wild-type levels.

2.5 Statistical Analysis

The significance of the observed results for the packaging and propagation efficiencies between the wild type and the mutants was established using the standard, paired, two-tailed Student's t-test. A P-value of <0.001 was statistically significant in both the packaging and propagation assays.

2.6 *In Silico* Analysis of MPMV RNA Secondary Structure

To establish structure–function relationship of each mutant and wild-type RNA during its RNA packaging process, the online software Mfold was used to predict the RNA secondary structure for each mutant and wild-type MPMV packaging signal RNA (Mathews et al., 1999; Zuker, 2003). Mfold predicts all optimal and suboptimal RNA secondary structures based on energy matrices by considering the minimum free energy of the provided RNA sequence. Next, the predicted structure for each mutant was validated by SHAPE (Aktar et al., 2013, 2014; Kalloush et al., 2016, 2019; Merino et al., 2005; Mortimer & Weeks, 2007, 2009; Mustafa et al., 2018).

2.7 *In vitro* RNA Transcription

The wild-type (RCR001) as well as mutant plasmids created for *in vitro* RNA transcription (FN series of clones; Table 4) were linearized with SmaI restriction enzyme located at the 3' end of the MPMV sequence. The linearized DNA fragments were subjected to *in vitro* transcription using the bacteriophage T7 RNA polymerase (MEGAscript T7 Transcription kit, Thermo Fischer Scientific) as described earlier (Kalloush et al., 2019; Mustafa et al., 2018). A portion of the *in vitro* transcribed RNA was analyzed on 8% acrylamide/8M urea gels to confirm the absence of abortive transcripts, followed by DNase treatment (Turbo DNA, Thermo Fischer Scientific) of the resulting RNA. After phenol/chloroform extraction, the RNAs were further purified on a TSK gel column (TSK Gel G4000SW column, TOSOH Bioscience, Griesheim, Germany) by Fast Protein Liquid Chromatography (FPLC) (AKTA; GE Healthcare Life Sciences, United States) in the presence of a buffer containing 200 mM sodium acetate (pH 6.5) and 1% (v/v) methanol. RNA fractions corresponding to the relevant peaks were pooled and ethanol precipitated.

The purified RNA pellets were resuspended in Milli-Q water, quantified by nanodrop (Thermo Fischer Scientific), and 800 ng of each RNA sample was electrophoresed on an 8% acrylamide/8M urea gel to check for its quality.

The *in vitro* transcribed RNAs were subjected to SHAPE analysis, as described before (Akhlaq et al., 2018; Aktar et al., 2013, 2014; Kalloush et al., 2016, 2019). Benzoyl cyanide (BzCN) was used to acylate the 2'-hydroxyl group of the unconstrained nucleotides in the RNA structure, followed by interrogation of each

nucleotide using two sets of identical but differentially labeled primers: one set (OTR18 and OTR19) contained the MPMV sequence 5' AGTTACTGGGACTTTCTCCG-3' (complementary to MPMV nt 1105-1123) and the second set (OTR22 and OTR23) corresponded to the 5'-CTTACTTTCAGGT CCAACGC-3' sequence (complementary to MPMV nt 857-875). One primer within each set was labeled with either VIC (OTR18 and OTR22) or NED (OTR19 and OTR23). The NED-labeled primers from each set were used to prepare a ddG sequencing ladder from the untreated RNA samples. The VIC-labeled primers were used for reverse transcription of the modified RNA.

Briefly, the sample preparation for SHAPE is as follows: BzCN-modified RNAs were annealed to 1uM VIC-labelled elongation primers for 2 min at 90°C and 2 min on ice. The samples were then incubated for 10 min at room temperature after adding 2ul 5X RT buffer (Life Science). The samples were incubated in elongation buffer (2ul 5XRT buffer, 0.6ul of 25mM dNTPs mix, 0.1ul of 20U/ul RT AMV enzyme from Life Science and made up to 10ul with water) for 20 min at 42°C and at 50°C for 30 min and then finally at 60°C for 10 min. The unmodified RNA samples were also treated in the same manner with the primers as aforementioned. The sequencing ladders were prepared from 2 pmol of untreated native RNA corresponding to each mutant, to which 1ul of 2uM NED labelled primer (either OTR 19 or OTR23) was added. After setting the reaction at 90°C for 2 min and cooling on ice for another 2 min, the volume of the reaction was made up to 10ul by adding 2ul of 5X RT buffer and incubated for 10min at room temperature. To each of the RNA sample 2ul of 5X RT buffer, 6ul of G10 (0.25mM dGTP, 1mM dATP, 1mM dCTP, 1mM dTTP) and 2ul of 100uM ddG and 0.2ul of 20U/ul RT AMV enzyme was added to prepare the ddG ladder. The incubation steps were followed in the same fashion as mentioned earlier. cDNAs from each were extracted by phenol-chloroform method. Each modified and unmodified cDNA samples elongated with the respective VIC-labelled primers were pooled with the corresponding ddG sequencing ladders (eg: samples elongated with OTR18 were pooled with the ladders prepared with OTR19) into a single tube containing 20ul 3M sodium acetate and 600ul of absolute ethanol for cDNA precipitation. After incubating on dry ice for 30 min, the samples were centrifuged at 13,000g for 20 min at 4°C, washed twice with 1ml 80% ethanol, pellets retrieved upon centrifugation for 5 min and air dried. The pellets resuspended in 10ul HiDi formamide

(ABI) were heat denatured at 90°C, cooled on ice for 5 min before loading to the 96-well plate for sequencing (Applied Biosystems 3130xl genetic analyzer).

The results obtained as electrophoretograms from the capillary electrophoresis of SHAPE-modified RNAs were converted to SHAPE reactivity data by the QuShape algorithm (Karabiber et al., 2013). Reactivity data (Appendix B) were applied as constraints to the mutant RNA sequence in RNAstructure (version 6) (Reuter & Mathews, 2010) to obtain the validated RNA secondary structure and were redrawn using VARNA software (Darty et al., 2009).

2.9 *In vitro* RNA Dimerization Assays

In vitro RNA dimerization was performed on the wild-type (RCR001) and FN series of mutant clones (Table 4) according to the method described previously (Aktar et al., 2013, 2014; Kalloush et al., 2019). Briefly, 300 nM of purified mutant or wild-type RNAs were incubated in dimer (50 mM sodium cacodylate pH 7.5, 300 mM KCl, 5 mM MgCl₂) or monomer (50 mM sodium cacodylate pH 7.5, 40 mM KCl, 0.1 mM MgCl₂) buffer for 30 min at 37°C. This was followed by electrophoresis in native 1% agarose TBM (50 mM Tris base, 45 mM boric acid, 0.1 mM MgCl₂) gel at 4°C. The gels were stained with ethidium bromide and visualized for dimeric or monomeric bands using ultraviolet (UV) transillumination. Band intensities were quantified using ImageLab software (BioImager, Biorad), and the percentage of dimerization was calculated for each RNA employing the following formula: $(\text{Intensity of dimer band} - \text{Intensity of background}) / [(\text{Intensity of monomer band} - \text{Intensity of background}) + (\text{Intensity of dimer band} - \text{Intensity of background})]$. Results for each mutant RNA were presented relative to the wild type to determine the effect of mutations on their dimerizing ability.

CHAPTER 3: RESULTS

3.1 Experimental Approach to Analyze the Effects of ssPurines and bpPurines Mutants on RNA Packaging and Propagation

In order to investigate the role of ssPurines and bpPurines during MPMV gRNA packaging, a series of deletion and/or substitution mutations were introduced in the respective regions of MPMV gRNA as shown in Figure 24A-C and the nature of mutations introduced are described in table 4. These mutants are tested in a biologically relevant single round of replication assay as described earlier and the details of the steps involved in the methodology are depicted in Figure 25 (Aktar et al., 2014; Browning et al., 2001; Jaballah et al., 2010; Kalloush et al., 2016, 2019; Mustafa et al., 2004; Schmidt et al., 2003). The virus particles produced were analysed for the packaged RNA by real time quantitative PCR (RT-qPCR), and viral supernatants were used to infect target cells (HeLa T4) to study the propagation of packaged RNA. Such *in vivo* packaging and propagation assays allowed us to quantify the effects of the introduced mutations on both RNA packaging as well as its propagation without any uncertainty, since the defective nature of the virions produced limited the assay to a single round of replication (Figure 25).

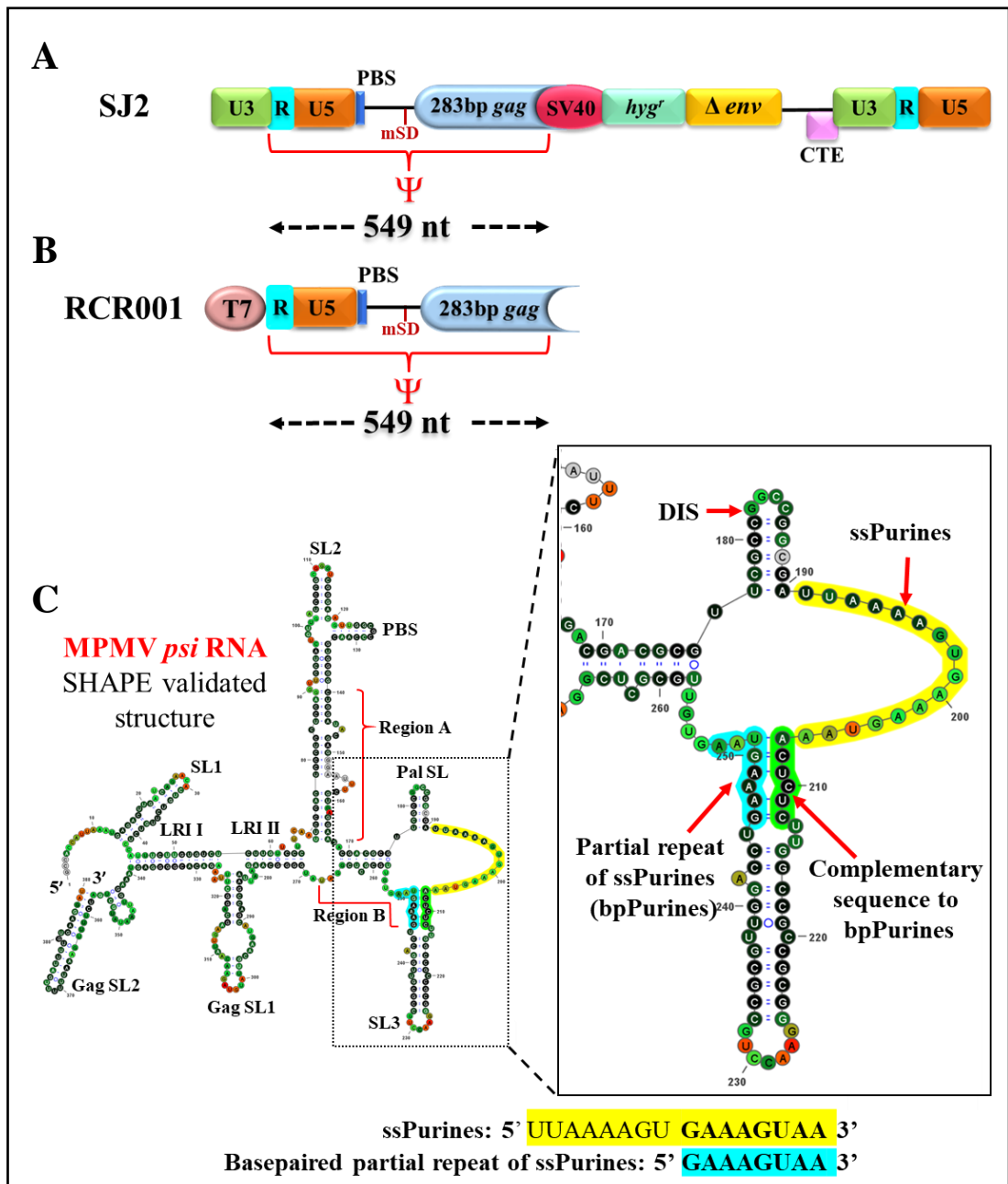


Figure 24. Illustration of the vectors used in the *in vivo* packaging and *in vitro* transcription assays and SHAPE-validated structure of the MPMV packaging signal RNA

(A) Schematic representation of the MPMV wild-type sub-genomic transfer vector, SJ2 used in the *in vivo* RNA packaging and propagation assay along with demarcation of boundaries of the RNA packaging determinants denoted as ψ . PBS; primer binding site; CTE; constitutive transport element. (B) Schematic representation of the wild-type plasmid used for *in vitro* transcription of the determinants of MPMV RNA packaging. (C) SHAPE-validated structure of MPMV RNA packaging determinants. ssPurines, single-stranded purines (yellow); bpPurines, base-paired repeat purines (blue). The green region highlights the sequences complementary to bpPurines. The RNA structure was redrawn in VARNA using SHAPE reactivity data (Aktar et al., 2013). Figure adapted from (Ali et al., 2020).

Table 4. Description of the mutations introduced in ssPurines and bpPurines

Sub-genomic transfer vectors	<i>In vitro</i> transcription vectors	ssPurines sequence and complementary sequence to basepaired purines (bpPurines)	Partial repeat of ssPurines in the form of bpPurines
SJ2 (WT)	RCR001	5' UUAAAAGU GAAAGUAA ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-I	FN-I	5' UUA Δ GAAAGUAA ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-II	FN-II	5' Δ ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-III	FN-III	5' AAUUUUCACUUUCAUU ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-IV	FN-IV	5' Δ ACUCUC 3'	5' Δ GUGUU 3'
LA-V	FN-V	5' UUAAAAGU Δ ACUCUC 3'	5' Δ GUGUU 3'
LA-VI	FN-VI	5' UUAAAAGU GAAAGUAA ACUCUC 3'	5' Δ GUGUU 3'
LA-VII	FN-VII	5' UUAAAAGU GAAAGUAA Δ 3'	5' GAAAGUAA GUGUU 3'

Table describing the nature of the mutations introduced and cloned both in the context of the sub-genomic transfer and *in vitro* transcription vectors. ssPurines, single-stranded purines (yellow); bpPurines, base-paired repeat purines (blue). The green region highlights the sequences complementary to bpPurines.

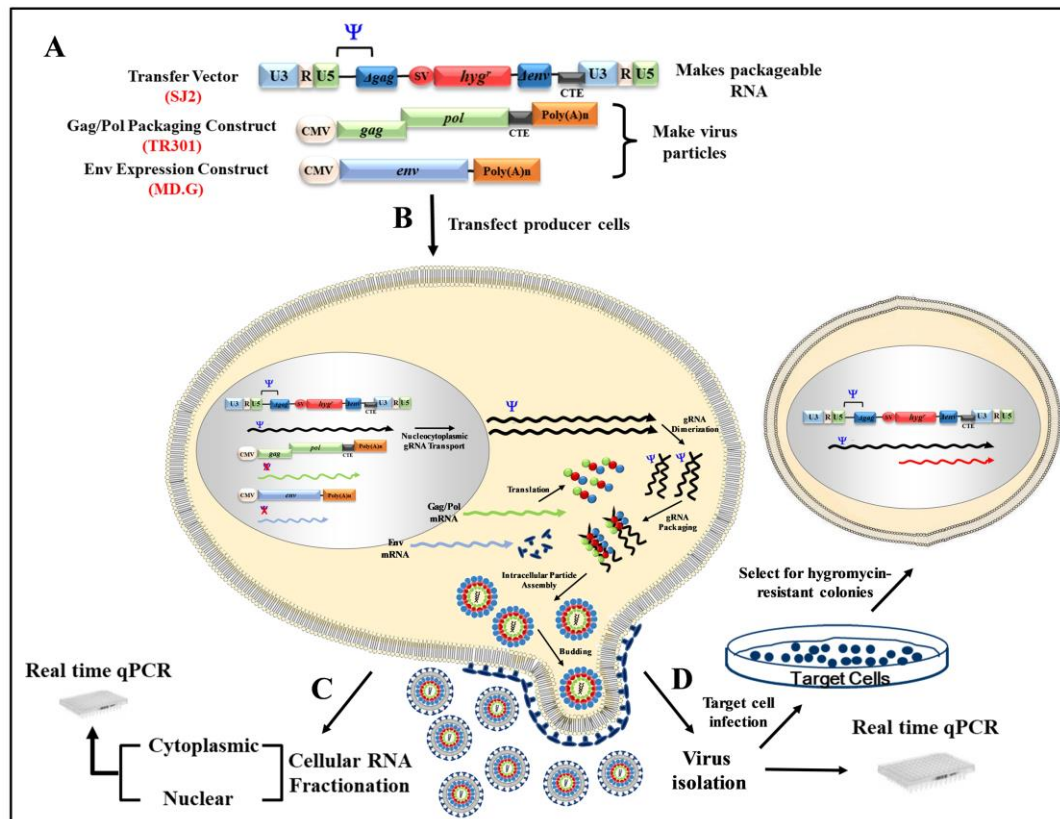


Figure 25. Illustration of the 3-plasmid *in vivo* packaging and propagation assay

(A) Graphical representation of the plasmids used to produce pseudotyped virus particles using VSV-G envelope expression vector. (B) Schematic depiction of an HEK 293T cells co-transfected with the three plasmids to produce infectious virus particles, which can only replicate to a single round. Unspliced sub-genomic RNA transcribed from the wild type vector, SJ2, and/or mutant transfer vectors can be packaged into the virions owing to the presence of intact packaging signal, while excluding RNAs transcribed from TR301 and MD.G. (C) Transfected cells are fractionated into nuclear and cytoplasmic fractions and analyzed for transfer vector RNA transport and expression. (D) Viral particles produced are tested for the amount of RNA packaged by RT-qPCR. Viral supernatants are also used to infect target cells (HeLa T4) to study RNA propagation. After infection, target cells are selected with media containing hygromycin B, allowing only those cells to survive which have been successfully infected since the packaged RNA contains the hygromycin resistance gene. Parts of the figure adapted from Kalloush et al., 2019; Pitchai et al., 2018).

The reliability of our *in vivo* RNA packaging assay was ensured by quality control measures. To ensure the integrity of the nuclear-cytoplasmic fractionation process, end point PCRs specific for unspliced β -actin were performed multiplexed with 18S ribosomal RNA (rRNA) on the cytoplasmic cDNA of the samples (Figure 26; Panel I). Unspliced β -actin mRNA should be restricted to the nuclear compartment of the cell unless the integrity of the nuclear membrane was compromised during the fractionation process (Tan et al., 1995). Lack of amplification of unspliced β -actin confirmed that our cytoplasmic cDNAs were indeed free from any contaminating

nuclear fraction (Figure 26; Panel I). Amplifications of 18S ribosomal and spliced β -actin-specific PCRs were conducted in parallel to confirm the presence of amplifiable cDNAs prepared from the cytoplasmic fractions (Figure 26; Panels I and II, respectively). Next, MPMV transfer vector-specific amplifications were conducted on cDNAs prepared from cytoplasmic fractions as well as from packaged viral RNAs. Such an approach confirmed both efficient nucleocytoplasmic transport as well as stable cytoplasmic expression of the transfer vector RNAs and differential RNA packaging of the mutants (Figure 26; Panel III & IV, respectively).

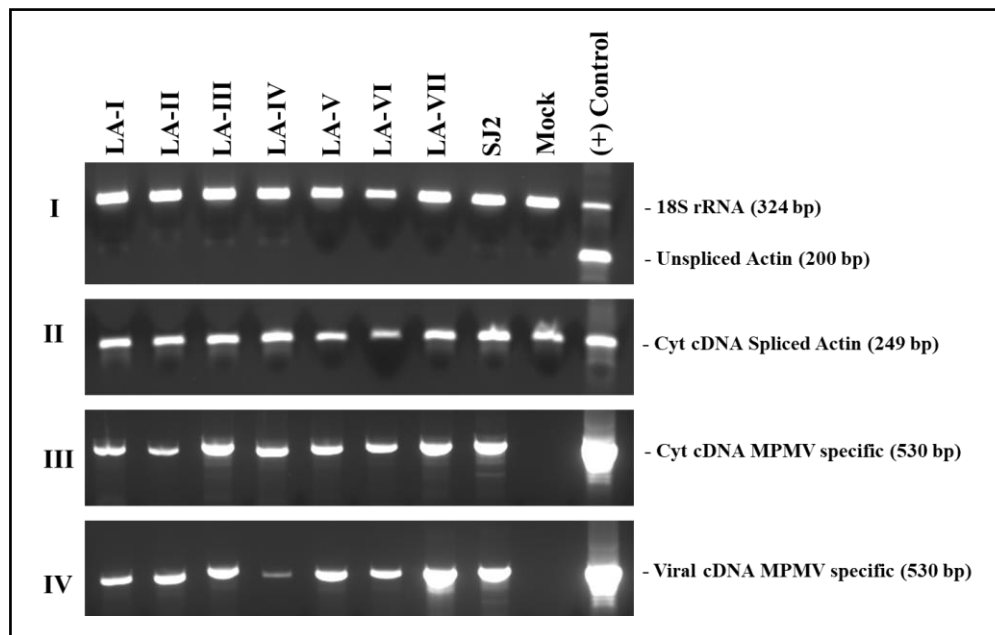


Figure 26. RT-PCR of viral and cytoplasmic cDNA fractions with appropriate controls

Panel I. PCR amplification of unspliced β -actin mRNA from cytoplasmic cDNAs. Multiplex amplifications were conducted in the presence of primers/competimer for 18S ribosomal RNA; Genomic DNA (gDNA) used as (+) Control. **Panel II.** PCR amplification of spliced β -actin mRNA from cytoplasmic cDNAs; gDNA used as (+) Control. **Panel III.** PCR amplification of MPMV transfer vector (SJ2)-specific primers from cytoplasmic cDNAs; plasmid DNA for SJ2 used as (+) Control. **Panel IV.** PCR amplification of MPMV transfer vector (SJ2)-specific sequences from cDNAs prepared from the viral RNAs; plasmid DNA for SJ2 used as (+) Control.

Finally, having confirmed the integrity of our RNA preparations and the amplifiability of cDNA preparations, we performed RT-qPCR to assess the amount of packaged RNA in the virions in relation to the RNA transcripts expressed in the cytoplasm of the producer cells, as described previously (Kalloush et al., 2016, 2019).

As previously reported, the viral particles were observed to have an equivalent amount of β -actin mRNA irrespective of the amount of viral vector RNA that was packaged, thus providing a good proxy for the amount of virions produced (Figure 27; Mustafa et al., 2012). Briefly, qPCRs were conducted for β -actin mRNA packaged into the virus particles, a cellular mRNA which previously has been shown to be packaged passively into the virions at the same levels irrespective of the amount of viral RNA packaged (Mustafa et al., 2012), thus providing an efficient internal control for normalization purposes. Consistently, MPMV particles produced in repeat experiments were observed to have packaged equivalent amounts of β -actin mRNA despite the fact that mutant viral vector RNAs were packaged at different levels, providing an internal control for the amount of virions produced (Figure 27).

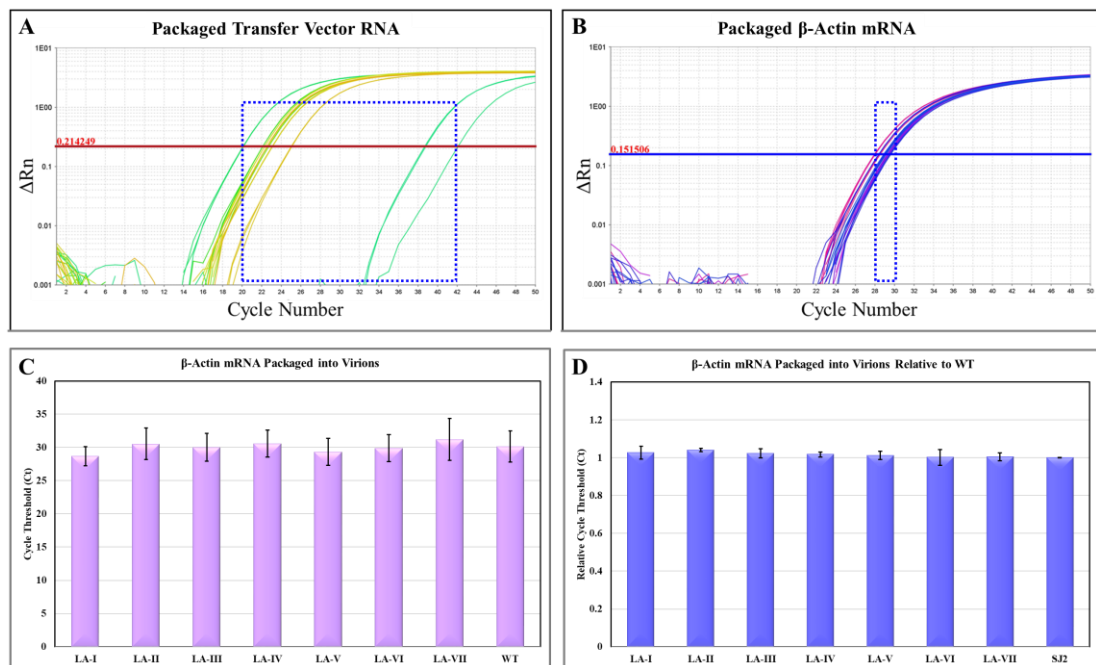


Figure 27. β -actin mRNA is packaged into viral particles at the same levels irrespective of the amount of viral RNA present in the virions

(A) Amplification plots for packaged viral vector RNAs expressed as ΔRn plotted against cycle numbers (Ct) from a representative experiment. (ΔRn is the MPMV or β -actin-specific fluorescence signal normalized to the signal for the internal passive control, ROX (Normalized Reporter or Rn) from which the baseline target fluorescence has been subtracted ($\Delta Rn = \text{Normalized Reporter (Rn)} - \text{baseline}$). The dashed blue boxes show the wide range of Ct values observed for all the mutants and wild type samples. (B) Amplification plots for packaged β -actin mRNA expressed as ΔRn against cycle numbers (Ct) from a representative experiment. The dashed blue boxes show the tight range of Ct values observed for all the mutants and wild type samples, showing that β -actin mRNA is packaged at the same levels in virions irrespective of the amount of vector RNA packaged. (C) Histograms showing the Ct values for β -actin mRNAs observed in all mutants tested. (D) Histograms showing the Ct values for β -actin mRNAs observed in all mutants tested relative to the wild type.

3.2 Role of ssPurines in gRNA Packaging and Propagation

To investigate the role of the ssPurines in MPMV gRNA packaging and propagation, we constructed two deletion mutants, LA-I and LA-II along with a substitution mutant LA-III (Table 4). LA-I comprised a five-nucleotide deletion (AAAGU) in the former part of ssPurines that is unique to this region as opposed to the latter sequence (GAAAGUAA) which is also repeated as bpPurines at the basal part of SL3 (Figure 24C). This deletion resulted in almost a 50% loss of richness in purines of ssPurines and also a reduction in the size of the ssPurine stretch, while the LA-II mutation created a complete deletion of ssPurines. In LA-III, the ssPurines were substituted with a heterologous non-purine sequence to identify if conserving the sequence or maintaining the secondary structure of the region is vital for viral RNA packaging and propagation. Collectively, these mutants should identify the effect, if any, of the ssPurines region on packaging and propagation of the virus, as well as identify if there were any differential effects between the former and latter purines within the ssPurine region.

Estimation of the relative packaging efficiency (RPE) of the transfer vector RNAs revealed that despite partial or complete deletion of ssPurines, RNA packaging of the mutant vectors (LA-I, LA-II and LA-III) was not significantly reduced when compared to the wild-type (WT; SJ2) RNA (LA-I ($P = 0.13$); LA-II ($P = 0.48$); LA-III ($P = 0.04$); Figure 28A). The RNA packaging observation was confirmed by a near identical effect observed on RNA propagation of these mutants, revealing that whatever RNA was packaged by the mutants was propagated successfully to the target cells (compare Figure 28A & B). These results suggest that ssPurines are dispensable for MPMV RNA packaging and propagation. Alternatively, one may also infer that the bpPurines may be compensating for the loss of function of the ssPurines.

3.3 Role of bpPurines in gRNA Packaging and Propagation

To investigate the significance of the redundancy of the purine-rich sequences during MPMV gRNA packaging and propagation processes, two double-deletion mutants involving both the ssPurines and bpPurines were created (LA-IV and LA-V; Table 4) along with a mutant with a complete deletion of only the bpPurines (LA-VI; Table 4). Mutant LA-IV showed severe defects in RNA packaging (RPE of 0.09; $P < 0.001$) compared to the wild-type SJ2 vector RNA (Figure 28A) which

corroborated with a concomitant drastic reduction in the colony-forming units (CFU)/ml observed for this mutant (~100-fold reduction compared to the WT; $P < 0.001$; Figure 28B). Surprisingly, this was not the case for the double deletion mutant, LA-V and the bpPurine deletion mutant, LA-VI, where the packaging capabilities of these mutant RNAs were nearly like that of the wild type (RPE of 0.94 ($P = 0.67$) and 0.74 ($P = 0.12$), respectively, Figure 28A). However, despite efficient packaging, the propagation of these mutant RNAs (LA-V & VI) were drastically reduced (CFU/ml 0.05 and 0.10, respectively, compared to the WT; $P < 0.001$; Figure 28B). Such an ablated propagation despite efficient RNA packaging is probably because these mutations affected post-packaging events in the viral life cycle such as reverse transcription and/or integration. This assertion is since our RNA propagation assay readout is based on successful reverse transcription of the packaged RNA and integration of the reverse transcribed RNA. RNA packaging was drastically affected only by a combined deletion of both the ssPurines and bpPurines sequences (LA-IV; Figure 28A), which could be attributed to structural and/or conformational changes in the RNA of this mutant. Appropriate control experiments, as described above, were conducted to ensure that the wild-type and mutant transfer vector RNAs were expressed efficiently in the cytoplasm and the nucleocytoplasmic fractionation was not compromised in the process (Figure 26; Panel I and III). Considering these controls, effective levels of packaging in both the mutants LA-V and LA-VI suggested that the purine-rich sequences indeed have a redundant role in augmenting gRNA packaging into the assembling virions. Thus, these data demonstrate that ssPurines and its bpPurines function in a compensatory fashion to mediate MPMV gRNA packaging, since their deletions resulted in insignificant effect on RNA packaging of mutants LA-V and LA-VI (Figure 28A).

Finally, mutant LA-VII was designed to study the relevance of maintaining the base pairing in the repeat sequence (bpPurines) at the base of SL3 (Figure 24C; Table 4). This mutant carries a deletion of the sequence (5' ACUCUC 3') complementary to the base paired region (5' GAAAGU 3') in bpPurines; hence, the mutation disrupts base pairing with the repeat region. Interestingly, this mutant did not show any significant defect on gRNA packaging and was encapsidated to nearly the wild-type levels (RPE of 0.87 ($P = 0.44$) relative to the WT; Figure 28A). Such robust RNA packaging was in agreement with the propagation of the packaged RNA (Figure 28B),

keeping in mind all the appropriate controls (Figure 26). Taken together, the data presented here suggest that: (1) the repeat bpPurines are not important for RNA packaging, and (2) base pairing of these repeat Purines (bpPurines) is neither important for MPMV RNA packaging nor RNA propagation. Furthermore, it suggests that the bpPurines play a vital role in viral RNA propagation since its deletion in the mutants LA-V and VI abrogated RNA propagation completely, despite efficient RNA packaging (Figure 28A & B). Next, the structure of these mutants was analyzed to establish structure-function relationship and to determine whether these phenotypes could be explained by changes in the RNA secondary structure of these mutants.

A

Sub-genomic transfer vectors	<i>In vitro</i> transcription vectors	ssPurines sequence and complementary sequence to basepaired purines (bpPurines)	Partial repeat of ssPurines in the form of bpPurines
SJ2 (WT)	RCR001	5' UAAAAAGU GAAAGUAA ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-I	FN-I	5' UUA Δ GAAAGUAA ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-II	FN-II	5' Δ ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-III	FN-III	5' AAUUUUCACUUUCAU ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-IV	FN-IV	5' Δ ACUCUC 3'	5' Δ GUGUU 3'
LA-V	FN-V	5' UAAAAAGU Δ ACUCUC 3'	5' Δ GUGUU 3'
LA-VI	FN-VI	5' UAAAAAGU GAAAGUAA ACUCUC 3'	5' Δ GUGUU 3'
LA-VII	FN-VII	5' UAAAAAGU GAAAGUAA Δ 3'	5' GAAAGUAA GUGUU 3'

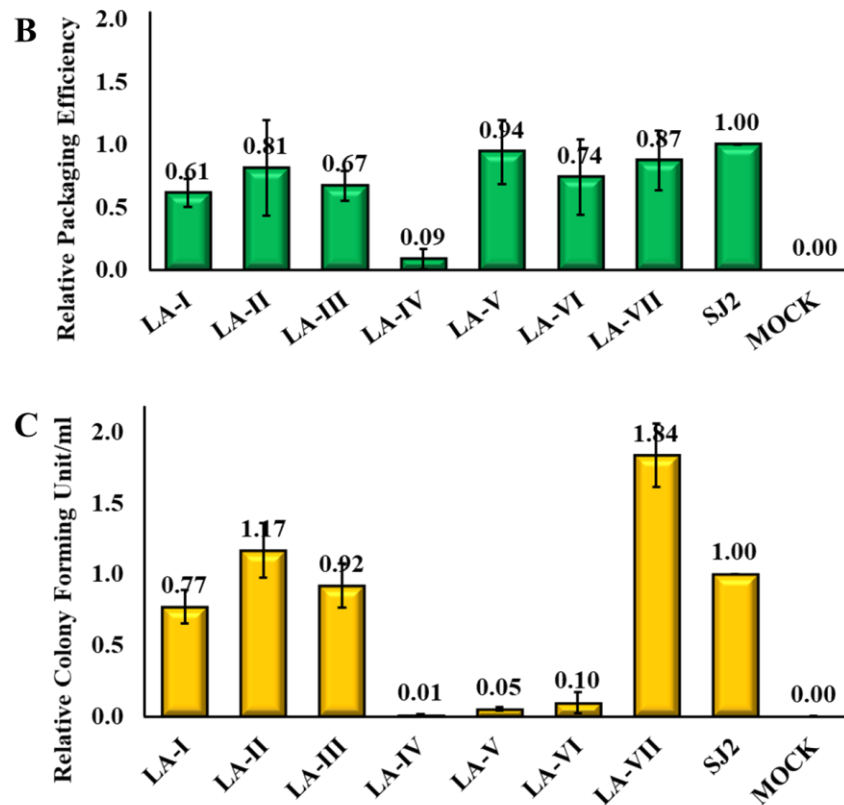


Figure 28. RT-qPCR and virus titer analysis of ssPurines and bpPurines mutants to establish their role in MPMV RNA packaging and propagation

(A) Table 4 describing the mutations introduced in ssPurines and bpPurines has been duplicated here for the ease of reference (B) Relative RNA packaging efficiencies for each of the mutant transfer vector RNAs. (C) Relative RNA propagation efficiencies for each of the mutant transfer vector RNAs after normalization to secreted alkaline phosphatase (SEAP) expression observed in the transfected cultures and represented as colony forming unit per ml (CFU/ml). The histograms represent data from multiple (to a maximum of five) independent experiments (\pm SD).

3.4 Structure-Function Analysis of ssPurines Mutants

The predicted Mfold RNA secondary structures for the ssPurines mutants LA-I to III did not show any noticeable structural disruptions due to deletion or substitution when compared to the wild type (Figure 29A–D). Therefore, to experimentally validate the predicted structures, biochemical probing of the *in vitro* transcribed mutant RNAs was performed employing SHAPE, and their structure–function relationship was established to better understand the results obtained using biological assays.

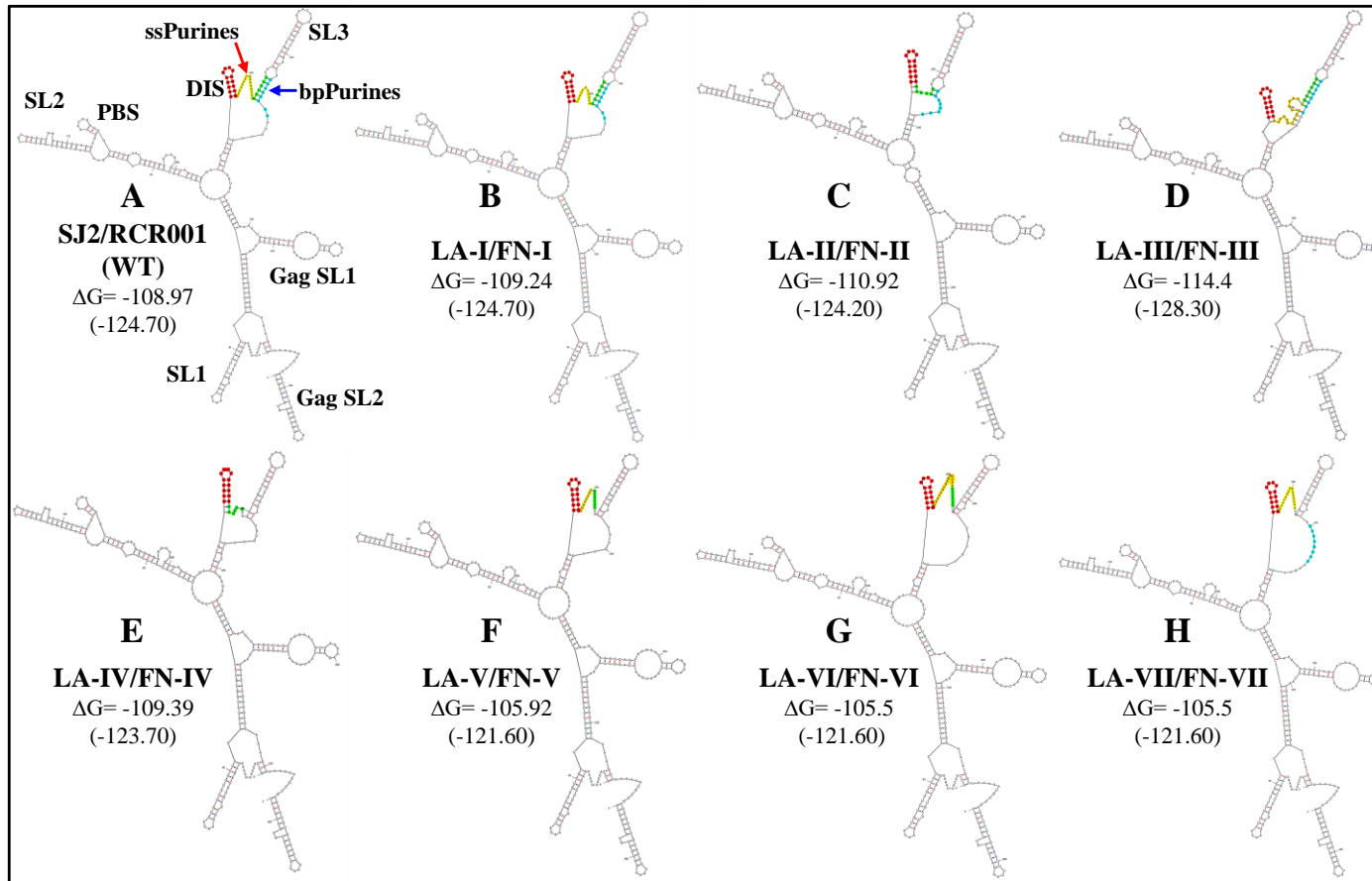


Figure 29. Mfold structural predictions for the wild type (SJ2) and mutant packaging signal RNAs

Color coding scheme for major structural motifs: single-stranded purines (ssPurines) = yellow; base-paired purines (bpPurines) = peacock blue; sequences complementary to bpPurines = green; Pal SL (DIS) = red.

The SHAPE-validated RNA secondary structure for mutant LA-I harboring five-nucleotide deletion in the former part of ssPurines (Figure 24C; Table 4) is depicted in Figure 30B. As predicted by Mfold (Figure 29A & B), the SHAPE-validated structures of LA-I and wild-type RNAs were found to be very similar (compare Figure 30A & B). Briefly, the LA-I SHAPE-validated structure maintained all the major stem loops, including the two U5-Gag LRIs (LRI I and LRI II) and Pal SL, and only minor changes were observed in the close vicinity of the 5-nucleotide deletion in the ssPurine region. The Pal SL was shortened by a single base-pair at the base, and this shortening of Pal SL was compensated by increasing in size of the linker between SL2 and Pal SL by one nucleotide (Figure 30B). These subtle changes did not compromise the overall RNA secondary structure in the mutant and maintained RNA packaging and propagation to wild-type (SJ2) levels (Figure 28A & B).

The SHAPE-validated structure of the LA-II mutant revealed that deletion of the ssPurines, while inducing local remodeling of the RNA secondary structure, did not affect the major structural motifs that have been shown to be important for MPMV gRNA dimerization and packaging (Figure 30C). Briefly, slight nucleotide rearrangements were observed around the base of the SL2 with a noticeable change at the LRI II shortening it by one nucleotide at the U5 region (Figure 30C). Thus, similar to LA-I, the SHAPE-validated structure of LA-II also lends credence to the observation that ssPurines are not crucial for MPMV gRNA packaging process since nearly wild-type levels of RNA packaging and propagation were observed for this mutant (Figure 28A & B). Like LA-I and LA-II, the SHAPE-validated structure of mutant LA-III (substitution of ssPurines with a heterologous non-purine sequence) also revealed conservation of the major structural motifs of the region except for minor changes at the substituted region (only 5 nucleotides remained single-stranded as opposed to 16 in the wild-type structure; Figure 30D vs. Figure 30A). In addition, substitution of ssPurines with a heterologous non-purine sequence in this mutant somehow contributed to a slightly longer stem in both the Pal SL and SL3 (Figure 30D). Corroborating with the SHAPE-validated structure (maintaining all structural motifs needed for RNA packaging and dimerization), this mutant also did not reveal any significant effect on either RNA packaging or propagation (Figure 28A & B). Taken together, analysis of the SHAPE-validated structures of these mutants (LA-I to III) revealed no deleterious changes to the overall RNA secondary structures and

indicated that neither the length of the Pal SL and SL3 stems nor the length or sequence of the ssPurines between these two helices play a significant role in MPMV gRNA packaging.

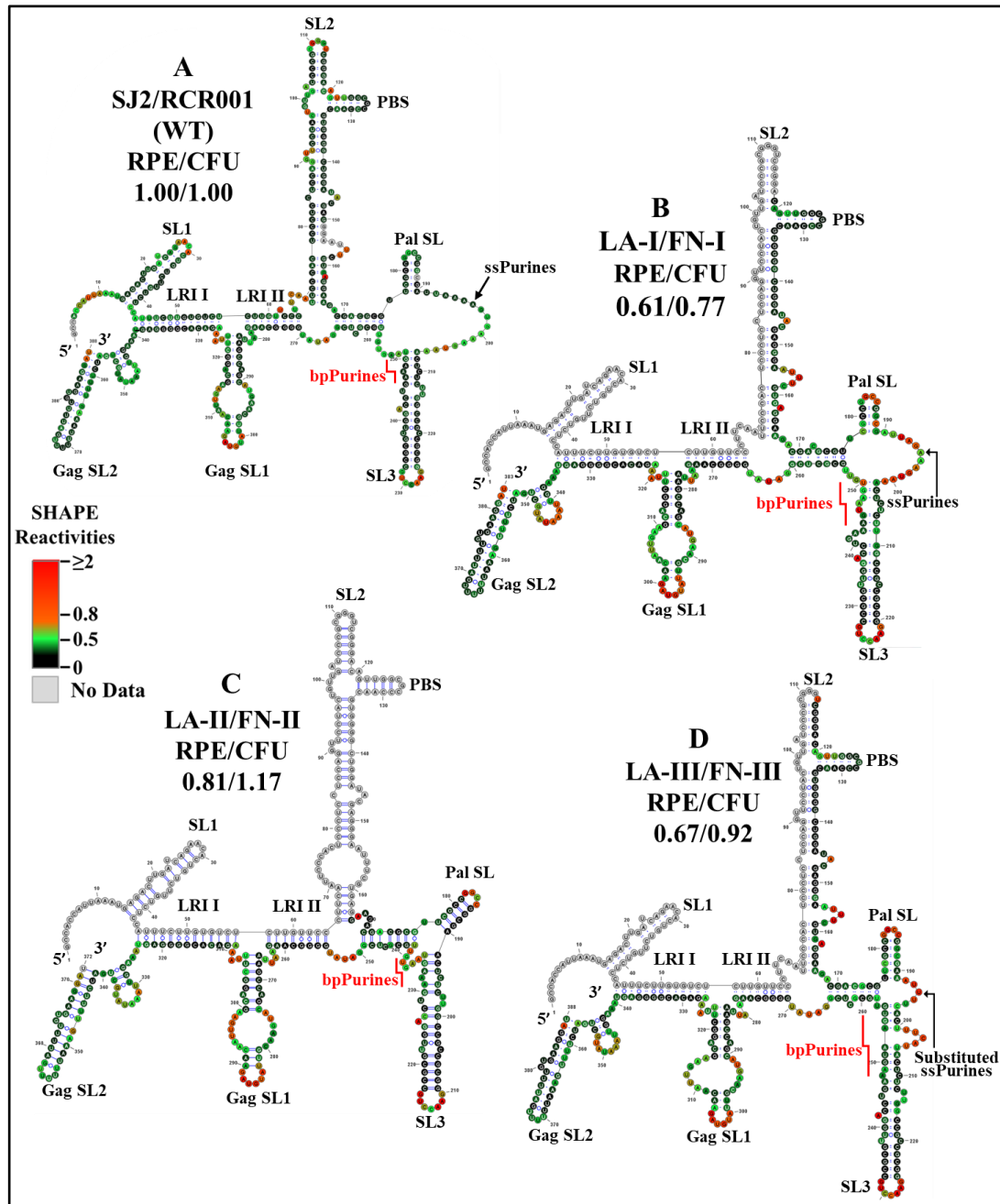


Figure 30. Selective 2' hydroxyl acylation analyzed by primer extension-validated structures of the wild-type and mutant LA-I/FN-I, LA-II/FN-II, and LA-III/FN-III packaging signal RNAs

(A) Wild type (SJ2/RCR001) (B) LA-I/FN-I containing deletion of AAAGU (5' sequence) in ssPurines. (C) LA-II/FN-II containing complete deletion of ssPurines. (D) LA-III/FN-III containing substitution of ssPurines. The SHAPE reactivities from three independent experiments were averaged and applied to RNAstructure program. The structure with the least minimum free energy was selected and redrawn using VARNA software and the major structural elements are annotated. Nucleotides are color annotated as per the SHAPE reactivities key shown.

3.5 Structure-Function Analysis of bpPurines Mutants

Mutants with deletions in bpPurines (LA-IV to VI) revealed a complete abrogation of RNA propagation (Figure 28B), indicating a potential role of this sequence in virus replication. Mfold structure predictions of these mutant RNAs (Figure 29E–G) suggested preservation of overall RNA secondary structure, despite severe effects on propagation. Thus, SHAPE was conducted on the *in vitro* transcribed RNAs of these mutants to identify any structural changes that may have implications for the observed biological results. In contrast with the Mfold prediction, the SHAPE-validated structure of the double deletion mutant LA-IV (ssPurines as well as bpPurines; Table 4) revealed a complete architectural distortion of the region (compare Figure 31A & B). Briefly, crucial structural motifs like U5-Gag LRI I and LRI II that have already been established to be important for MPMV gRNA packaging and propagation (Kalloush et al., 2016) were lost (Figure 31B). Considering that LA-IV was defective for both gRNA packaging and propagation, this observation confirms the importance of maintaining the intact higher-order structure of the packaging signal RNA during MPMV replication. It also reinforces the importance of structural validation of mutants by biochemical probing methods rather than just relying on predictions, especially when the biological data do not correlate with the structure predictions.

The SHAPE-validated RNA structure of mutant LA-V, in which the duplicated sequence (GAAAGUAA) in ssPurines and bpPurines had been deleted and maintained the overall structural motifs critical for gRNA packaging. Noticeably, among the RNA conformers for this mutant, two differently base-paired structures for the primer binding site (PBS) were consistent with the SHAPE data (Figure 31Ca & b). While this double deletion resulted in shortening of SL3, the remaining important structural elements (U5-Gag LRIs and Pal SL) were architecturally maintained (Figure 31Ca & b). These intact RNA domains could have contributed to the efficient packaging of this mutant (Figure 28A). The negative impact on RNA propagation of this mutant could be due to the effect of deletions of the duplicated sequences in the ssPurines and the bpPurines.

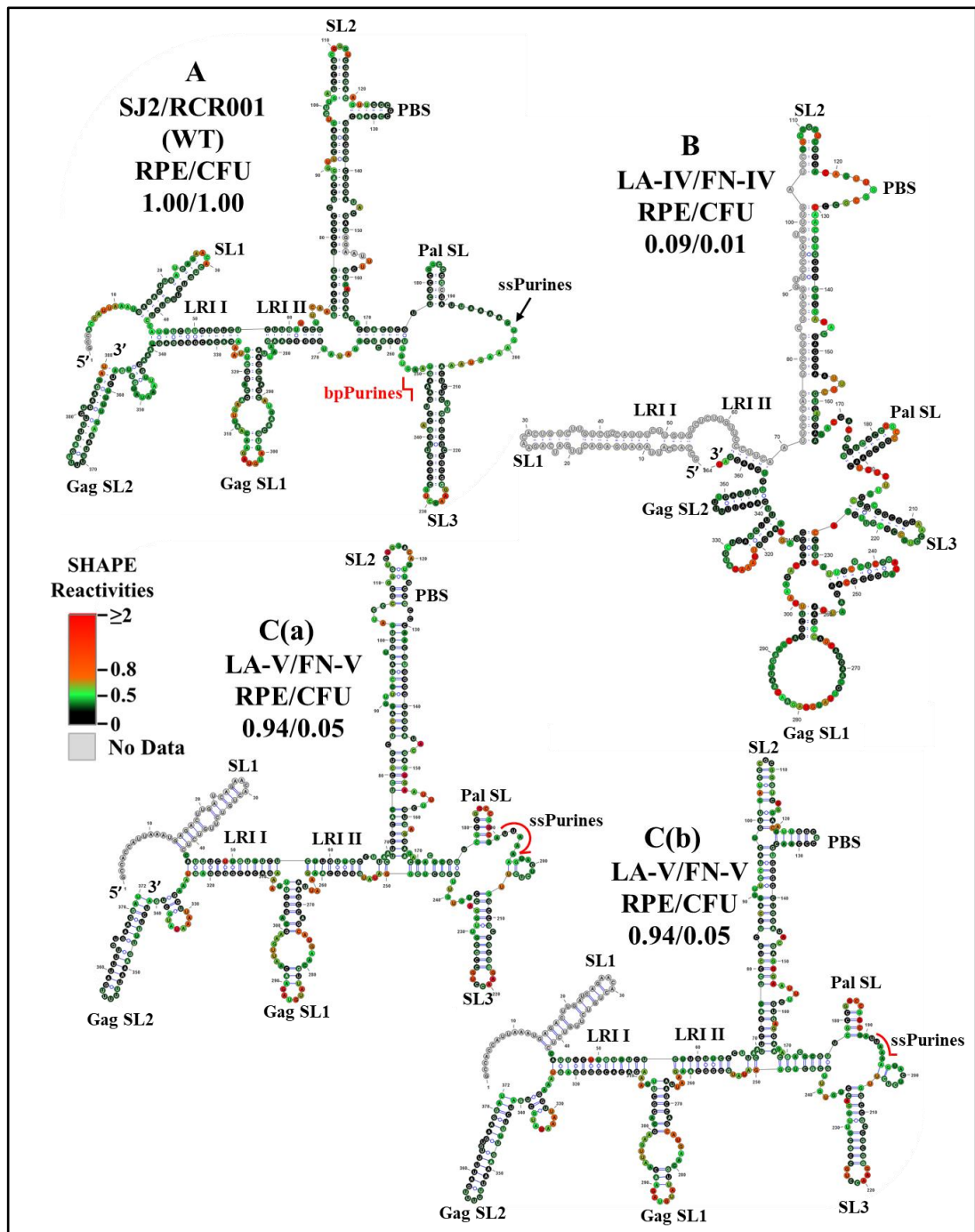


Figure 31. Selective 2' hydroxyl acylation analyzed by primer extension-validated structures of the wild-type and mutant LA-IV/FN-IV and LA-V/FN-V packaging signal RNAs

(A) Wild-type (SJ2/RCR001) (B) LA-IV/FN-IV containing deletion of both ssPurines and bpPurines. (C(a)) Structure number 1 of the mutant LA-V/FN-V containing deletion of 3' sequence (GAAAGUAA) of the ssPurines and a complete deletion of bpPurines. (C(b)) Structure number 2 of the mutant LA-V/FN-V. The SHAPE reactivities from three independent experiments were averaged and applied to RNAstructure program. The structure with the least minimum free energy was selected and redrawn using VARNA software and the major structural elements are marked. Nucleotides are color annotated as per the SHAPE reactivities key shown.

The SHAPE-validated structure for mutant LA-VI (deletion of only the bpPurines) revealed that while SL1, SL2, GagSL1, GagSL2, LRI-I, LRI-II, and PBS structures were maintained, Pal SL and most of SL3 structure were remodeled (Figure 32B). Thus, conservation of LRI I, LRI II, and of the PBS domain structure might have conferred to the efficient packaging of this mutant (Figure 28A), while loss of the native Pal SL and SL3 structure might have caused the propagation defect (Figure 28B). Alternatively, it is possible that bpPurines play a non-structural role in the early steps of the MPMV life cycle that has yet to be elucidated. Analysis of the ssPurines in the SHAPE structure of LA-VI revealed that they were found partially base paired (Figure 32B). This indicates that ssPurines do not need to be fully single stranded to allow efficient RNA packaging.

The SHAPE-validated structure of LA-VII containing deletion of the sequence (5' ACUCUC 3') complementary to the base paired region in bpPurines was not very different from the wild-type structure (Figure 32C), the only differences being slightly shorter Pal SL and SL3 stems, as well as a longer single stranded region immediately downstream of SL3. Based on these structural data, it is not very surprising that the RNA packaging and propagation of this mutant were very efficient (Figure 28A & B). Based on the results from the bpPurines mutants, LA-IV-LA-VII, it is clear that maintaining an intact higher-order structure for the MPMV packaging signal is important for its efficient viral encapsidation. Moreover, the propagation capabilities of the virus were largely influenced by the presence or absence of the bpPurines on an overall mostly intact gRNA structure.

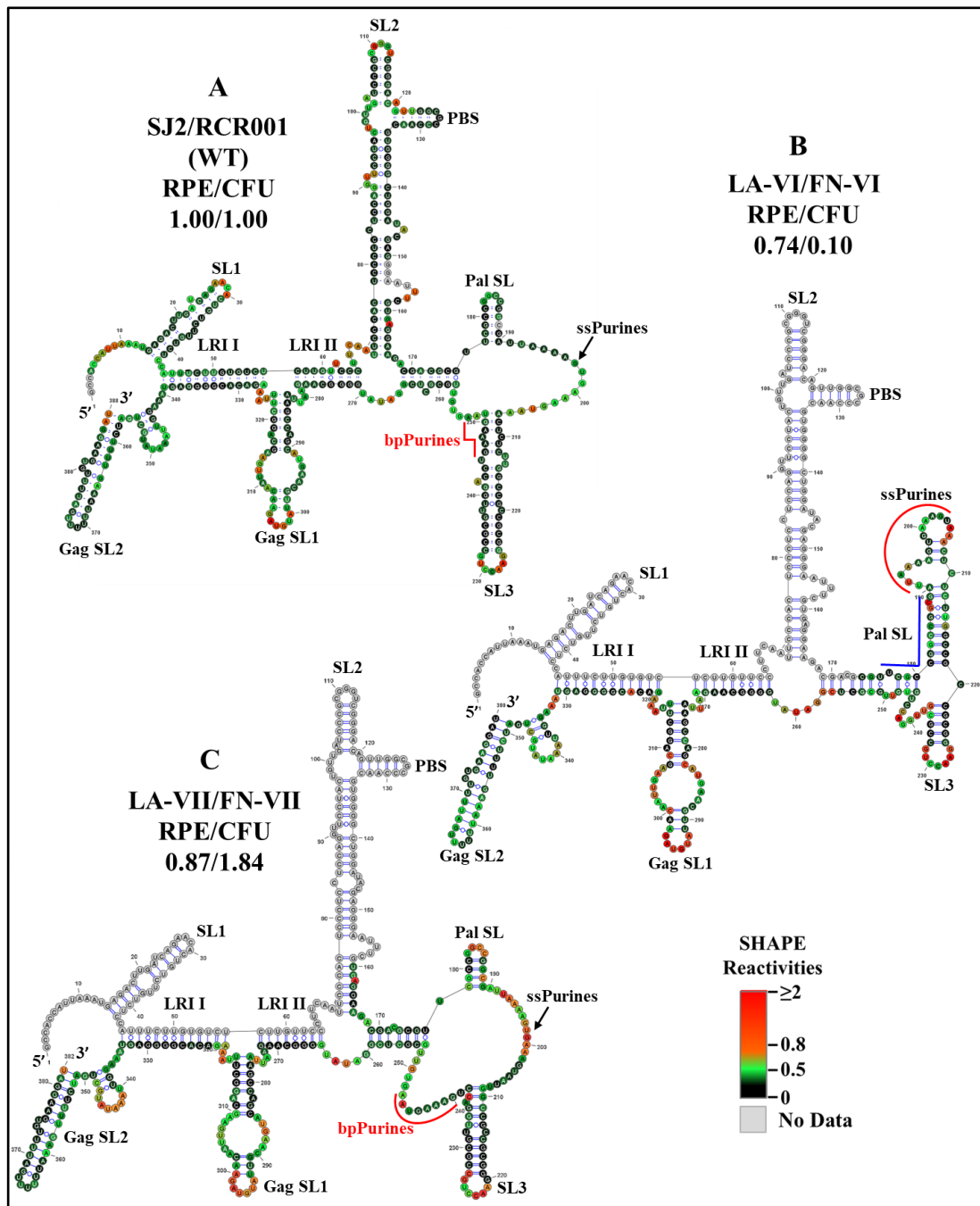


Figure 32. Selective 2' hydroxyl acylation analyzed by primer extension-validated structures of the wild-type and mutant LA-VI/FN-VI and LA-VII/FN-VII packaging signal RNAs

(A) Wild type transfer vector (SJ2/RCR001) (B) LA-VI/FN-VI containing complete deletion of bpPurines only. (C) LA-VII/FN-VII containing deletion of sequences complementary to bpPurines. The SHAPE reactivities from three independent experiments were averaged and applied to RNAstructure program. The structure with the least minimum free energy was selected and redrawn using VARNA software and the major structural elements are marked. Nucleotides are color annotated as per the SHAPE reactivities key shown.

3.6 *In vitro* Dimerization Capability of the Mutant RNAs

Genomic RNA dimerization and packaging are interconnected events in the retroviral life cycle. Since all the mutants considered in this study showed efficient RNA packaging except for mutant LA-IV, it was important to investigate the dimerization abilities of these mutant RNAs. Thus, the *in vitro* transcribed wild-type (RCR001) and mutant RNAs were incubated in a monomer or dimer buffer and analyzed for the percentage of dimerization after running them on agarose gel in TBM buffer at 4°C. The dimers formed in the TBM buffer at 4°C (Figure 33A) served as a reference for quantifying the percentage of dimerization between the WT and the mutant RNAs (Figure 33B) as described earlier (Aktar et al., 2013, 2014; Kalloush et al., 2016). Interestingly, none of the mutant RNAs tested showed any significant difference in their dimerization potential compared to the wild type (Figure 33). This is consistent with SHAPE-derived structures of mutants LA-I to LA-V and LA-VII since all of them maintained Pal SL, including LA-IV, which has the most dramatically affected secondary structure (Figure 31B). In the case of LA-VI, where the Pal SL was found to be base-paired in the SHAPE-validated structure, the results of dimerization showed no pronounced effect on the dimer forming ability of their respective RNAs (77.3 vs. 92.77 of wild type; Figure 32B). It is conceivable that when the Pal SL is structurally not available, its function could possibly be compensated by palindromic sequence within the PBS region, augmenting RNA dimerization, as has been proposed previously for MPMV and MMTV (Aktar et al., 2013, 2014).

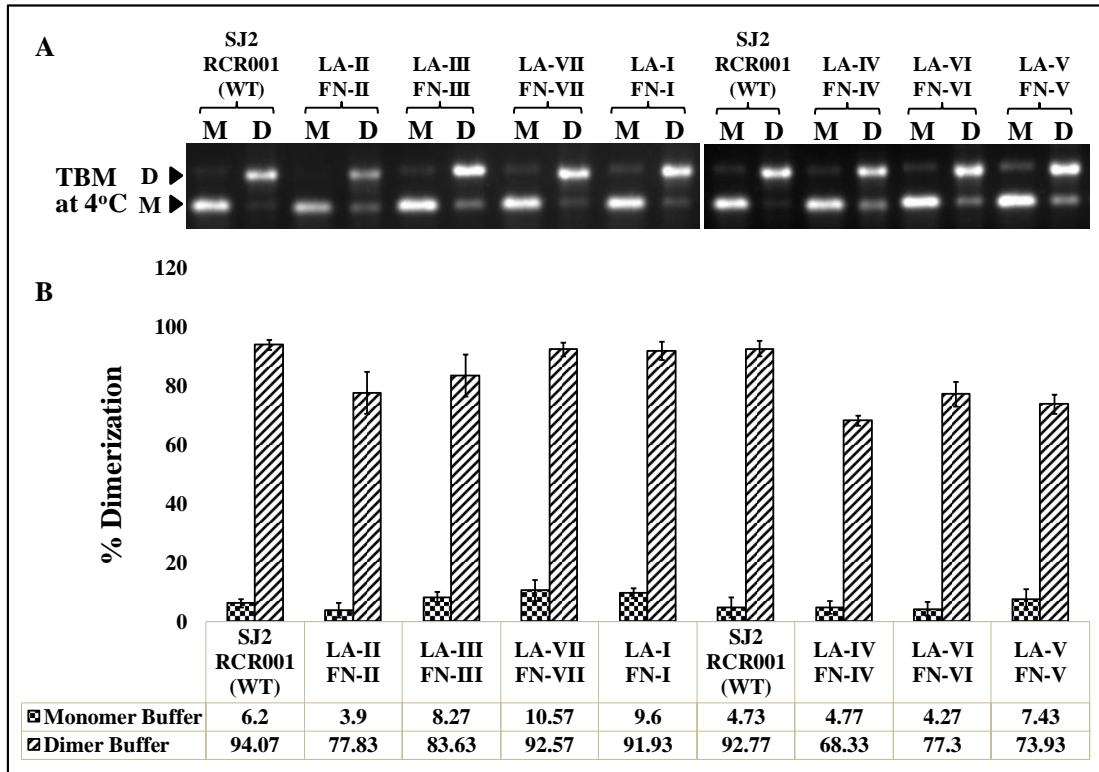


Figure 33. *In vitro* RNA dimerization assay on the wild-type and mutant SHAPE-interrogated packaging signal RNAs

(A) Wild-type and mutant RNAs were incubated in monomer (M) or dimer (D) buffer and analyzed by electrophoresis on 1% agarose gel in TBM buffer at 4°C. (B) Quantification of the relative ability of each mutant RNA to dimerize compared to the wild-type (SJ2/RCR001) RNA. The experiments were performed in triplicates.

Furthermore, two other mutants LA-VIII and LA-IX were made to investigate the importance of spatial location of bpPurines (Figure 34A). In the mutant LA-VIII, the basal part of SL3 that comprise the bpPurines and its complementary region were deleted whereas in the mutant LA-IX the bpPurines and its complementary region of SL3 were switched spatially to ensure the integrity of the gRNA secondary structure. In both these cases, the mutants were able to package its gRNA into the nascent virions further confirming that the packaging elements on gRNA primarily lies on the ssPurines and is present intact in both mutants (Figure 34B). SHAPE validated structures of these two mutants further confirm this assumption that the major packaging determinants are maintained as can be seen in the figure 35A & B below. The shortening of SL3 stem in LA-VIII mutant did not affect its packaging into the virions (RPE-0.88; Figure 34B). As can be seen in the mutant LA-IX, the swapping of the sequences at the base of SL3 did not have a profound effect on its packaging efficiency (RPE-0.73; Figure 34B) and maintains the RNA structure near identical to the wildtype RNA conformation (Figure 35C). The results of these two mutants reiterates the earlier findings that bpPurines are not involved in MPMV gRNA packaging and other motifs in region B could possibly be facilitating in Gag binding to function in a redundant manner. As was observed earlier, the propagation abilities for these two mutant RNAs were lost (Figure 34C); conceivably due to the loss of bpPurines and the results from LA-IX suggest that maintaining the intuitive location of bpPurines in MPMV gRNA is crucial for its replication.

Sub-genomic transfer vectors	<i>In vitro</i> transcription vectors	ssPurines sequence and complementary sequence to basepaired purines (bpPurines)	Partial repeat of ssPurines in the form of bpPurines
SJ2 (WT)	RCR001	5' UUAAAAGU GAAAGUAA ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-VIII	FN-VIII	5' UUAAAAGU GAAAGUAA Δ 3'	5' Δ GUGUU 3'
LA-IX	FN-IX	5' UUAAAAGU GAAAGUAA UGAAAG 3'	5' CUCUCAAA GUGUU 3'

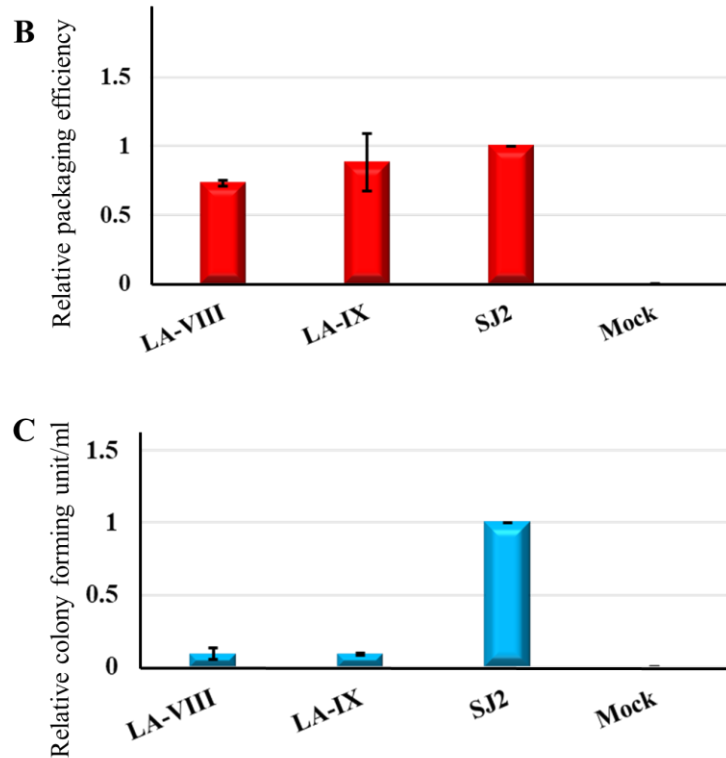


Figure 34. RT-qPCR and virus titer analysis of bpPurines and its complementary region mutants to establish their role in MPMV RNA packaging and propagation

(A) Table describing the mutations introduced in bpPurines and its complementary region in SL3 of MPMV gRNA. (B) Relative RNA packaging efficiencies for each of the mutant transfer vector RNAs. (C) Relative RNA propagation efficiencies for each of the mutant transfer vector RNAs after normalization to secreted alkaline phosphatase (SEAP) expression observed in the transfected cultures and represented as colony forming unit per ml (CFU/ml). The histograms represent data from multiple (to a maximum of two) independent experiments (\pm SD).

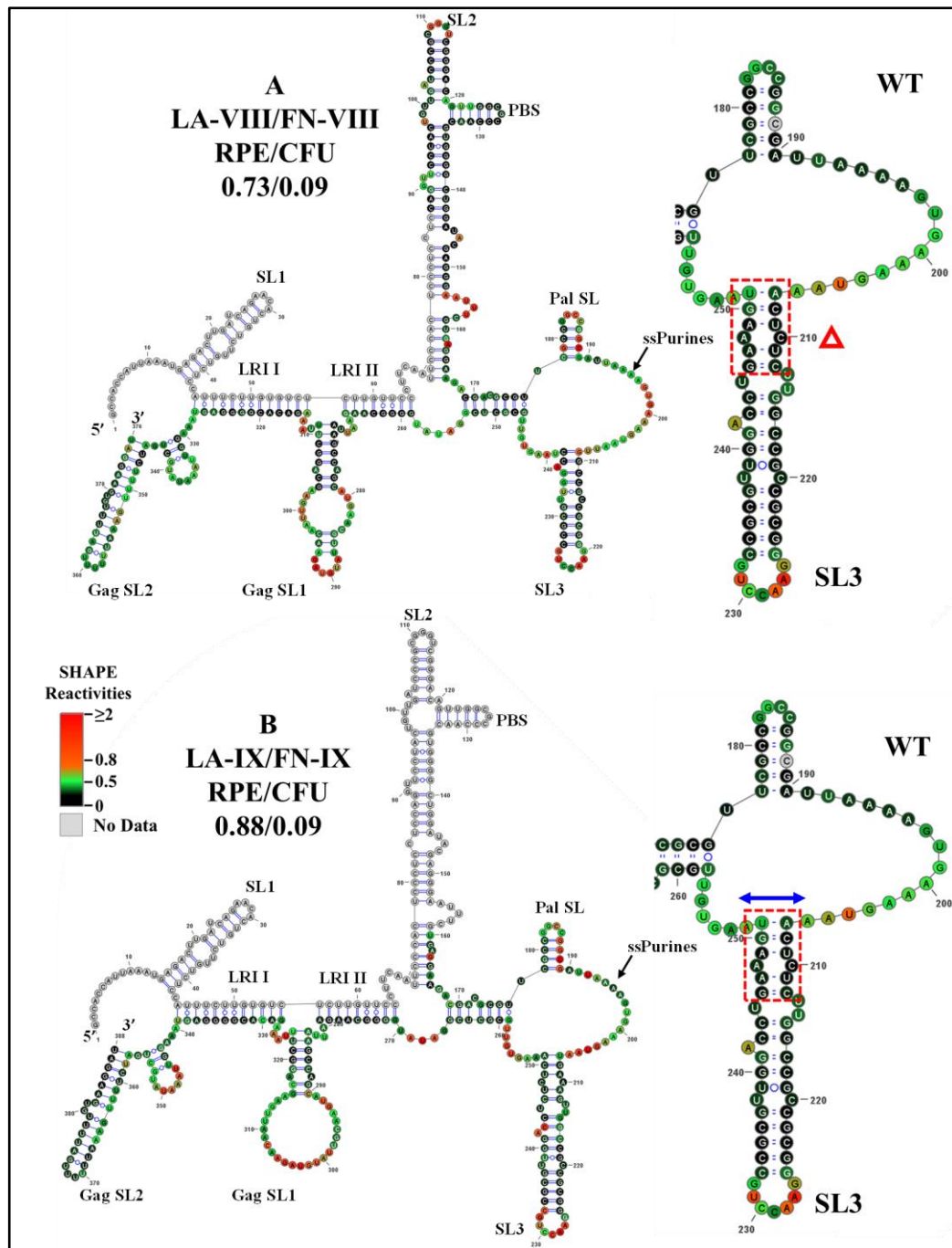


Figure 35. Selective 2' hydroxyl acylation analyzed by primer extension-validated structures of the wild-type and mutant LA-VIII/FN-VIII and LA-IX/FN-IX packaging signal RNAs

(A) SHAPE validated structure for LA-VIII/FN-VIII containing complete deletion of bpPurines and its complementary region in SL3. (B) SHAPE validated structure for LA-IX/FN-IX containing the swapping of bpPurines and its complementary region in SL3. The SHAPE reactivities from three independent experiments were averaged and applied to RNAstructure program. The structure with the least minimum free energy was selected and redrawn using VARNA software and the major structural elements are annotated. Nucleotides are color annotated as per the SHAPE reactivities key shown. Wild type structure for SL3 is shown enlarged in the figure and the red dotted box indicates the nature of mutation introduced.

3.7 Role of a GU-rich Sequences of Region B in gRNA Packaging and Propagation

Recent biochemical data generated in our laboratory have identified two Pr78^{Gag} binding sites on MPMV gRNA which were proposed to act redundantly during MPMV gRNA packaging process (Pitchai et al., 2021). These redundantly acting binding sites include: i) the ss-Purine loop (U¹⁹¹UAAAAGUGAAAGUAA²⁰⁶), ii) a second loop (A²⁵²AGUGUU²⁵⁸) corresponding to the last two nucleotides of the bpPurines region and extending immediately into the adjacent GU-rich region. To further validate the biological role of these Gag binding structural motifs in MPMV gRNA packaging, we introduced a series of mutations and tested them in a biological assay. Briefly, a mutation was created in which both the ssPurine loop along with a GU rich region were deleted simultaneously (LA-X; Figure 36A). In concordance with the biochemical data and as expected, this double deletion mutant showed a drastic defect in RNA packaging (RPE-0.2; Figure 36B), further validating that both the ssPurine loop and the GU rich region are involved in MPMV gRNA packaging. To further confirm this assertion, a mutant, LA-XI (Figure 36A) was made containing the deletion of the GU rich region only while maintaining the ssPurine loop. Consistent with the recent biochemical observations which suggested that ssPurine loop and GU rich region act redundantly during gRNA packaging, this mutant was found to be packaged efficiently (RPE-1.09; Figure 36B) even though it had GU rich region completely deleted. Finally, with no surprise and as expected the RNA packaging data of these two deletion mutants (LA-X & LA-XI) corroborated well with the RNA propagation data (Figure 36C). Taken together results presented in this section further validate that these two Gag binding sites (U¹⁹¹UAAAAGUGAAAGUAA²⁰⁶ & A²⁵²AGUGUU²⁵⁸) act redundantly during MPMV gRNA packaging process.

A

Sub-genomic transfer vectors	ssPurines sequence and complementary sequence to basepaired purines (bpPurines)	Partial repeat of ssPurines in the form of bpPurines
SJ2 (WT)	5' UUAAAAGU GAAAGUAA ACUCUC 3'	5' GAAAGUAA GUGUU 3'
LA-X	5' Δ ACUCUC 3'	5' GAAAGUAA Δ 3'
LA-XI	5' UUAAAAGU GAAAGUAA ACUCUC 3'	5' GAAAGUAA Δ 3'

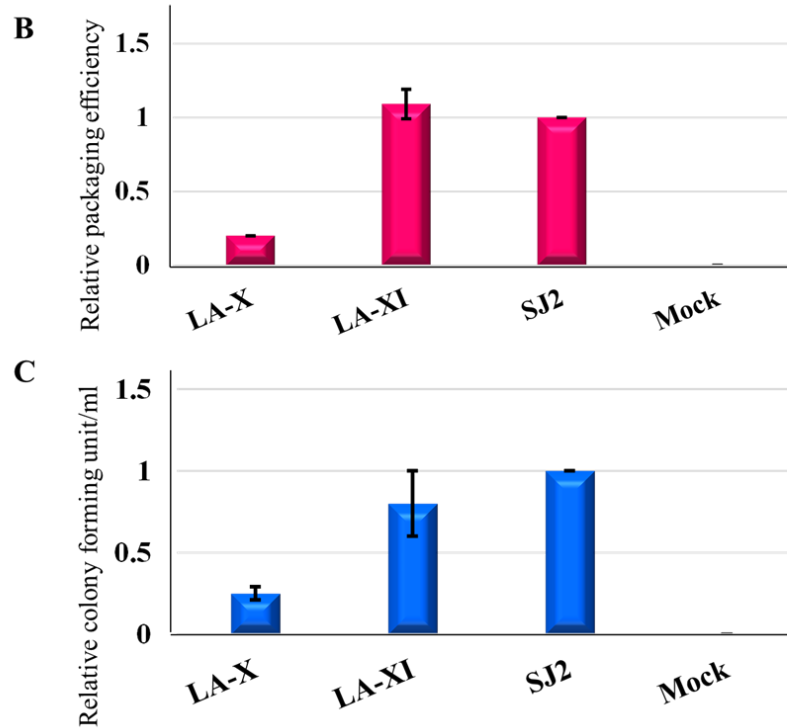


Figure 36. RT-qPCR and virus titer analysis of GU-rich region mutants to establish their role in MPMV RNA packaging and propagation

(A) Table describing the mutations introduced in ssPurines and GU-rich region of region B in MPMV gRNA. (B) Relative RNA packaging efficiencies for each of the mutant transfer vector RNAs. (C) Relative RNA propagation efficiencies for each of the mutant transfer vector RNAs after normalization to secreted alkaline phosphatase (SEAP) expression observed in the transfected cultures and represented as colony forming unit per ml (CFU/ml). The histograms represent data from multiple (to a maximum of two) independent experiments (\pm SD).

CHAPTER 4: DISCUSSION

The current study was undertaken to establish the role of ssPurines and its base-paired partial repeat, bpPurines, in MPMV gRNA packaging and propagation. The presence of a stretch of purines in the packaging sequences on retroviral gRNA has been proposed to facilitate RNA packaging by functioning as a potential Gag binding site (Abd El-Wahab et al., 2014; Bell & Lever, 2013; Didierlaurent et al., 2011; Lever, 2009; Moore et al., 2009; Moore & Hu, 2009; Mustafa et al., 2018; Paillart et al., 1997; Zeffman et al., 2000). In support of this hypothesis, a purine-rich internal loop in the proximity of DIS has recently been shown to bind efficiently to purified full-length HIV-1 Gag (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Smyth et al., 2015, 2018). The SHAPE-validated structure of MPMV packaging signal RNA revealed a stretch of ssPurines (in close proximity of DIS), as well as a partial repeat in the form of bpPurines (Aktar et al., 2013). Therefore, we hypothesized that these stretches of purines could function either at the sequence or structural levels in mediating MPMV gRNA packaging, possibly by functioning as potential Gag binding sites. Based on the data generated while testing this hypothesis, employing mutational, biological, structural, and biochemical analyses allowed us to propose a model shown in Figure 36. According to this model: (1) exclusive deletion of either the ssPurines or bpPurines does not affect MPMV gRNA packaging, (2) deletion of bpPurines, irrespective of the presence or absence of ssPurines affects RNA propagation severely, (3) biochemical probing by SHAPE reveals the structural basis for severe defects in RNA packaging, (4) RNA dimerization is not affected in any of the mutants, and finally, (5) Both the ssPurine loop and GU-rich region downstream of bpPurines play a role in gRNA packaging in a redundant manner. Together, these data and along with the recently published biochemical data (Pitchai, 2021) suggest that binding of full-length Pr78^{Gag} on MPMV gRNA may not be restricted to a single purine-rich domain.

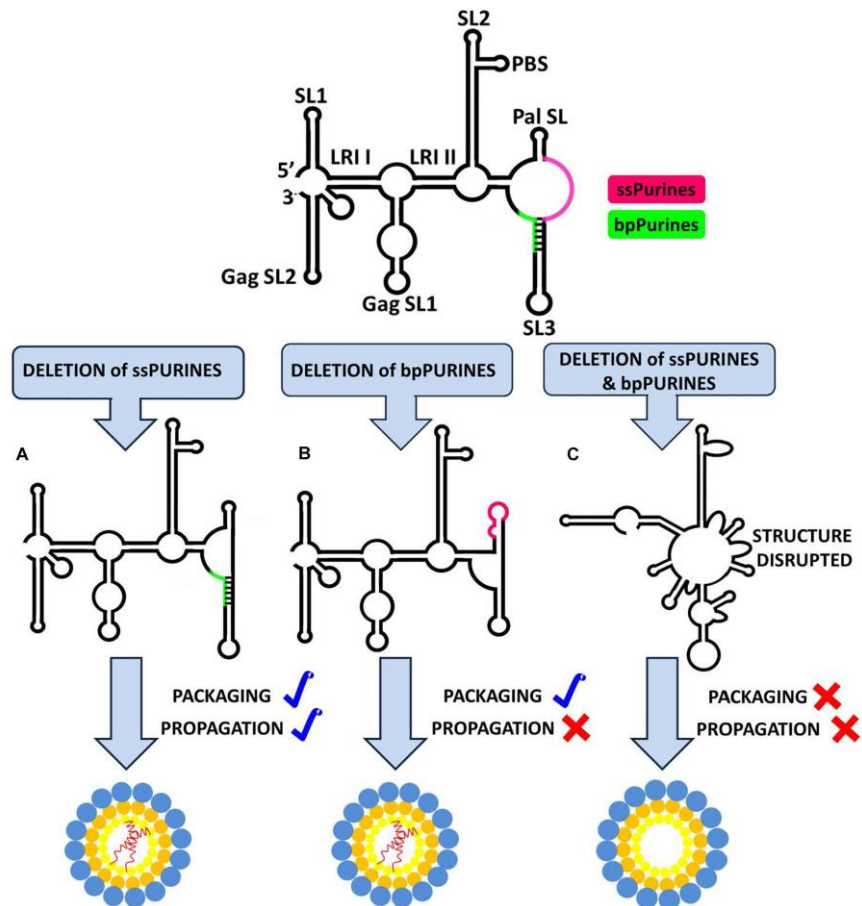


Figure 37. Structural model of MPMV packaging signal RNA containing two purine-rich sequences (ssPurines and bpPurines) which play a redundant role during MPMV gRNA packaging

Single deletion of (A) ssPurines or (B) bpPurines does not disrupt the overall RNA secondary structure. As a result, RNA is packaged efficiently, suggesting that ssPurines & bpPurines could act as potential Gag binding sites during gRNA packaging. Deletion of bpPurines neither affects RNA structure nor its packaging, but RNA propagation is abrogated. (C) Simultaneous deletion of both regions disrupts the overall RNA secondary structure, resulting in abrogation of both RNA packaging and propagation. These data further suggest that in addition to RNA packaging, bpPurines are crucial for some aspects of MPMV replication post-RNA packaging.

Experiments performed employing genetic trans-complementation assays with the ssPurine mutants (LA-I, LA-II, and LA-III, Table 4) did not show any detrimental effect on RNA packaging or propagation (Figure 28A & B). Furthermore, the SHAPE validated structures for these mutants revealed that the major structural motifs of the MPMV packaging signal RNA, such as the SL1, SL2, and Pal SL (Aktar et al., 2013), and U5-Gag LRIs (Kalloush et al., 2016) that are needed to sustain the overall stability of the RNA secondary structures, were maintained except for minor localized changes

due to the introduced mutations (Figure 30B–D). Such a scenario of having localized distortions in the RNA structure not affecting viral replication has been observed earlier, for instance with deletions of the apical part of SL3 that had only a marginal effect on MPMV RNA packaging and propagation (Jaballah et al., 2010; Kalloush et al., 2019). These observations suggest that minor structural changes to the region in and around ssPurines in the MPMV packaging signal RNA are tolerable by the virus; hence, mutations in these regions do not have a profound effect on its packaging or propagation. Taken together, these data suggest that ssPurines by themselves do not play a crucial role in MPMV gRNA packaging and propagation.

Earlier structure prediction analyses of a mutant harboring a deletion of the ssPurines suggested that the stretch of duplicated bpPurines may become single stranded, restoring RNA packaging to almost wild-type levels (Jaballah et al., 2010). However, the SHAPE-validated structure for the replication competent mutant LA-II, containing the precise deletion of ssPurines (Figure 30C) resulted in a structure identical to the wild type, suggesting that the bpPurines may not necessarily be unpaired to allow efficient packaging and propagation of the virus. In agreement with the above observation, the SHAPE-validated structures of mutants containing partial deletion of ssPurines (LA-I) or substitution with heterologous sequence (LA-III) did not reveal any conformational changes leading to unpairing of the bpPurine (Figure 30B-D); yet these mutants maintained efficient RNA packaging and propagation (Figure 28A & B). Mutants containing deletion of bpPurines in the presence as well as absence of ssPurines (LA-IV, LA-V, and LA-VI) were tested to gain insight toward the importance of the bpPurines during gRNA packaging and propagation. Mutant LA-IV (deletion of both ssPurines and bpPurines) was found to be completely incapable of packaging as well as propagation of its RNA (LA-IV; Figure 28A & B). SHAPE validated structure of this mutant RNA revealed a complete disruption of the RNA secondary structure including the loss of U5-Gag LRIs that have been shown to be critical for MPMV RNA packaging (Kalloush et al., 2016; Figure 31B) suggesting that the ssPurines and bpPurines may be important in maintaining the spatial organization of the higher-order structure of MPMV packaging signal RNA in a somewhat similar manner to U5-Gag LRIs (Kalloush et al., 2016). In order to destabilize the RNA secondary structure of the packaging signal, the LA-VII mutant was designed by deleting the complementary sequence to the bpPurines on the other side of the stem

(Figure 24C; Table 4). In contrast to our expectations, the LA-VII mutant maintained its RNA secondary structure very similar to the wild type; hence, LA-VII was not only successful in packaging its gRNA into the virus particles, but it also produced infectious virions that could efficiently transduce the target cells (Figure 28 A & B).

On the other hand, mutants LA-V (double deletion involving only the duplicated sequence in ssPurines and bpPurines; (Figure 24C; Table 4) and LA-VI (deletion of only bpPurines; Figure 24C; Table 4) were found to be defective for RNA propagation, despite their gRNAs being efficiently packaged into the nascent virus particles (Figure 28 A & B). SHAPE-validated structures for both these mutants established the presence of majority of the RNA structural domains critical for MPMV gRNA packaging (LA-V; Figure 31Ca & b; LA-VI; Figure 32B), and further asserting that bpPurines by themselves hold no role in facilitating packaging. However, in mutant LA-V RNA, the PBS stem loop revealed two structures in transition with the native wild-type structure among its SHAPE-compatible conformers. In both structures the PBS was base-paired; hence, the lack of propagation could not be attributed to the alternative conformations of the PBS (compare Figure 31Ca & b). In the mutant LA-VI, the Pal SL, ssPurines and the SL3 region were remodeled. Nevertheless, it is noteworthy that all mutant RNAs (LA-I to LA-VII) were successful in forming stable dimers, even in the case of LA-VI where the Pal SL was observed to be base-paired. This could in part be explained if it is assumed that in the absence of Pal SL, dimerization could be initiated by the palindromic sequence in the PBS Pal. Earlier study on MPMV have shown that structural rearrangements at the Pal SL-SL3 region are well accommodated by the virus during its replication process (Kalloush et al., 2019); hence, the loss of propagation of LA-VI packaged RNA could not be related to these observed structural changes. Additional results from mutants LA-VIII and LA-IX discredited the role of bpPurines in RNA packaging. Considering these observations, we speculate that the lack of propagation in these mutants is likely due to the deletion of bpPurines. Furthermore, it also suggested that retaining bpPurines in its native context in the gRNA is important for successful MPMV RNA propagation.

Consistent with the observations above, recently published work from our lab employing *in vitro* biochemical studies have shown that the MPMV polyprotein Pr78^{Gag} binds to ssPurines and the GU-rich (5'GUGU3') loop in region B which contain two nucleotides (AA) from 3' end of bpPurines (5'GAAAGUAA3'); Pitchai et

al., 2021). Remarkably, the findings from the *in vitro* band shift and foot printing assays primed us to propose a model on how MPMV could possibly discriminate its full-length gRNA from the large milieu of spliced viral and cellular mRNA during viral RNA encapsidation. It is now clear that Gag binds significantly to the ssPurines which is present invariably in all the MPMV transcripts as it is located upstream of mSD; the selection process of gRNA is achieved by the precursor Gag after binding to the GU-rich region present downstream of mSD in the *psi* RNA which is unvaryingly absent in spliced viral RNAs (Pitchai et al., 2021). These results correlate well while comparing the biological data of LA-II, LA-X and LA-XI and where defective RNA packaging was observed when ssPurines and GU-rich region were simultaneously deleted from the MPMV *psi* gRNA.

In summary, this study suggests that the ssPurines, and GU-rich sequences play a redundant role in MPMV life cycle by acting as potential Gag binding sites, while maintaining the overall structure and stability of the MPMV packaging signal RNA. The findings also corroborate well with the mechanism proposed for MPMV gRNA recognition during viral assembly. Additionally, our results also suggest a potential role of the bpPurines in the early steps of the MPMV life cycle that affect propagation of the packaged RNA (Figure 37).

In addition, it would also be interesting to perform *in vivo* or *in virio* probing experiments, as has been done for HIV-1 (Paillart et al., 2004; Wilkinson et al., 2008). Nevertheless, in the case of HIV-1, *in virio* SHAPE data show minimal differences with *in vitro* data, supporting the usefulness of the *in vitro* data.

CHAPTER 5: CONCLUSIONS

According to this study that focused on elucidating the role of purine rich sequences as potential Gag binding sites in selective packaging of the MPMV gRNA during its life cycle draws the following conclusions:

- ❖ Presence of ssPurines in MPMV *psi* RNA establishes packaging suggesting it as the primary Gag binding site during viral RNA encapsidation.
- ❖ Deletion of either the ssPurines or bpPurines does not affect MPMV gRNA packaging and bpPurines may not necessarily be unpaired (as was speculated earlier) to bring about gRNA packaging.
- ❖ Non-responsiveness of bpPurines to RNA packaging suggested the presence of other unpaired regions; the GU-rich loop downstream of bpPurines that facilitate in gRNA packaging.
- ❖ Deletion of bpPurines, irrespective of the presence or absence of ssPurines affects RNA propagation severely.
- ❖ Biochemical probing by SHAPE reveals the importance of structural basis for severe defects in RNA packaging.
- ❖ Biological results presented here further confirm that ssPurines and GU-rich region serves as unique Pr78^{Gag} binding sites in a redundant fashion as has recently been confirmed based on biochemical data (Pitchai, et al., 2021).

REFERENCES

- Abd El-Wahab, E. W., Smyth, R. P., Mailler, E., Bernacchi, S., Vivet-Boudou, V., Hijnen, M., Jossinet, F., Mak, J., Paillart, J.-C., & Marquet, R. (2014). Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nature Communications*, *5*, 4304.
- Ajasin, D., & Eugenin, E. A. (2020). HIV-1 Tat: Role in Bystander Toxicity. *Frontiers in Cellular and Infection Microbiology*, *10*.
- Akhlaq, S., Panicker, N. G., Philip, P. S., Ali, L. M., Dudley, J. P., Rizvi, T. A., & Mustafa, F. (2018). A cis-Acting Element Downstream of the Mouse Mammary Tumor Virus Major Splice Donor Critical for RNA Elongation and Stability. *Journal of Molecular Biology*, *430*(21), 4307–4324.
- Aktar, S. J., Jabeen, A., Ali, L. M., Vivet-Boudou, V., Marquet, R., & Rizvi, T. A. (2013). SHAPE analysis of the 5' end of the Mason-Pfizer monkey virus (MPMV) genomic RNA reveals structural elements required for genome dimerization. *RNA*, *19*(12), 1648–1658.
- Aktar, S. J., Vivet-Boudou, V., Ali, L. M., Jabeen, A., Kalloush, R. M., Richer, D., Mustafa, F., Marquet, R., & Rizvi, T. A. (2014). Structural basis of genomic RNA (gRNA) dimerization and packaging determinants of mouse mammary tumor virus (MMTV). *Retrovirology*, *11*, 96.
- Al Dhaheri, N. S., Phillip, P. S., Ghazawi, A., Ali, J., Beebi, E., Jaballah, S. A., & Rizvi, T. A. (2009). Cross-packaging of genetically distinct mouse and primate retroviral RNAs. *Retrovirology*, *6*, 66.
- Al Shamsi, I. R., Al Dhaheri, N. S., Phillip, P. S., Mustafa, F., & Rizvi, T. A. (2011). Reciprocal cross-packaging of primate lentiviral (HIV-1 and SIV) RNAs by heterologous non-lentiviral MPMV proteins. *Virus Research*, *155*(1), 352–357.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell* (4th ed.). Garland Science.
- Albritton, L. M. (2018). Chapter 1—Retrovirus Receptor Interactions and Entry. In L. J. Parent (Ed.), *Retrovirus-Cell Interactions* (pp. 1–49). Academic Press.
- Ali, Lizna M., Rizvi, T. A., & Mustafa, F. (2016). Cross- and Co-Packaging of Retroviral RNAs and Their Consequences. *Viruses*, *8*(10), 276.
- Ali, Lizna M., Pitchai, F. N. N., Vivet-Boudou, V., Chameettachal, A., Jabeen, A., Pillai, V. N., Mustafa, F., Marquet, R., & Rizvi, T. A. (2020). Role of Purine-Rich Regions in Mason-Pfizer Monkey Virus (MPMV) Genomic RNA Packaging and Propagation. *Frontiers in Microbiology*, *11*.

- Anderson, E. C., & Lever, A. M. L. (2006). Human immunodeficiency virus type 1 Gag polyprotein modulates its own translation. *Journal of Virology*, *80*(21), 10478–10486.
- Anderson, W. F. (1990). September 14, 1990: The Beginning. *Human Gene Therapy*, *1*(4), 371–372.
- Andrake, M. D., & Skalka, A. M. (2015). Retroviral Integrase: Then and Now. *Annual Review of Virology*, *2*(1), 241–264.
- Baig, T. T., Lanchy, J.-M., & Lodmell, J. S. (2009). Randomization and in vivo selection reveal a GGRG motif essential for packaging human immunodeficiency virus type 2 RNA. *Journal of Virology*, *83*(2), 802–810.
- Balvay, L., Lopez Lastra, M., Sargueil, B., Darlix, J.-L., & Ohlmann, T. (2007). Translational control of retroviruses. *Nature Reviews. Microbiology*, *5*(2), 128–140.
- Bandyopadhyay, P. K., & Temin, H. M. (1984). Expression from an internal AUG codon of herpes simplex thymidine kinase gene inserted in a retrovirus vector. *Molecular and Cellular Biology*, *4*(4), 743–748.
- Barnard, R. J. O., Narayan, S., Dornadula, G., Miller, M. D., & Young, J. A. T. (2004). Low pH Is Required for Avian Sarcoma and Leukosis Virus Env-Dependent Viral Penetration into the Cytosol and Not for Viral Uncoating. *Journal of Virology*, *78*(19), 10433–10441.
- Bedwell, G. J., & Engelman, A. N. (2021). Factors that mold the nuclear landscape of HIV-1 integration. *Nucleic Acids Research*, *49*(2), 621–635.
- Beitelshees, M., Hill, A., Rostami, P., Jones, C. H., & Pfeifer, B. A. (2017). Pressing diseases that represent promising targets for gene therapy. *Discovery Medicine*, *24*(134), 313–322.
- Bell, N. M., & Lever, A. M. L. (2013). HIV Gag polyprotein: Processing and early viral particle assembly. *Trends in Microbiology*, *21*(3), 136–144.
- Berkhout, B. (1996). Structure and function of the human immunodeficiency virus leader RNA. *Progress in Nucleic Acid Research and Molecular Biology*, *54*, 1–34.
- Berkhout, B., & van Wamel, J. L. (1996). Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *Journal of Virology*, *70*(10), 6723–6732.
- Berkowitz, R., Fisher, J., & Goff, S. P. (1996). RNA packaging. *Current Topics in Microbiology and Immunology*, *214*, 177–218.

- Bernacchi, S., Abd El-Wahab, E. W., Dubois, N., Hijnen, M., Smyth, R. P., Mak, J., Marquet, R., & Paillart, J.-C. (2017). HIV-1 Pr55Gagbinds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biology*, *14*(1), 90–103.
- Bieniasz, P., & Telesnitsky, A. (2018). Multiple, Switchable Protein:RNA Interactions Regulate Human Immunodeficiency Virus Type 1 Assembly. *Annual Review of Virology*, *5*(1), 165–183.
- Blot, V., Perugi, F., Gay, B., Prévost, M.-C., Briant, L., Tangy, F., Abriel, H., Staub, O., Dokhélar, M.-C., & Pique, C. (2004). Nedd4.1-mediated ubiquitination and subsequent recruitment of Tsg101 ensure HTLV-1 Gag trafficking towards the multivesicular body pathway prior to virus budding. *Journal of Cell Science*, *117*(Pt 11), 2357–2367.
- Bohl, C. R., Brown, S. M., & Weldon, R. A. (2005). The pp24 phosphoprotein of Mason-Pfizer monkey virus contributes to viral genome packaging. *Retrovirology*, *2*, 68.
- Bohlayer, W. P., & DeStefano, J. J. (2006). Tighter binding of HIV reverse transcriptase to RNA-DNA vs. DNA-DNA results mostly from interactions in the polymerase domain and requires just a small stretch of RNA-DNA. *Biochemistry*, *45*(24), 7628–7638.
- Bolinger, C., & Boris-Lawrie, K. (2009). Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. *Retrovirology*, *6*, 8.
- Bray, M., Prasad, S., Dubay, J. W., Hunter, E., Jeang, K. T., Rekosh, D., & Hammarskjöld, M. L. (1994). A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(4), 1256–1260.
- Brierley, I., Digard, P., & Inglis, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. *Cell*, *57*(4), 537–547.
- Brierley, I., & Dos Ramos, F. J. (2006). Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Research*, *119*(1), 29–42.
- Brown, J. D., Kharytonchyk, S., Chaudry, I., Iyer, A. S., Carter, H., Becker, G., Desai, Y., Glang, L., Choi, S. H., Singh, K., Lopresti, M. W., Orellana, M., Rodriguez, T., Oboh, U., Hijji, J., Ghinger, F. G., Stewart, K., Francis, D., Edwards, B., ... Summers, M. F. (2020). Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science (New York, N.Y.)*, *368*(6489), 413–417.

- Browning, M. T., Mustafa, F., Schmidt, R. D., Lew, K. A., & Rizvi, T. A. (2003). Delineation of sequences important for efficient packaging of feline immunodeficiency virus RNA. *The Journal of General Virology*, 84(Pt 3), 621–627.
- Browning, M. T., Schmidt, R. D., Lew, K. A., & Rizvi, T. A. (2001). Primate and Feline Lentivirus Vector RNA Packaging and Propagation by Heterologous Lentivirus Virions. *Journal of Virology*, 75(11), 5129–5140.
- Bukrinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., & Stevenson, M. (1993). A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature*, 365(6447), 666–669.
- Burdick, R. C., Li, C., Munshi, M., Rawson, J. M. O., Nagashima, K., Hu, W.-S., & Pathak, V. K. (2020). HIV-1 uncoats in the nucleus near sites of integration. *Proceedings of the National Academy of Sciences*, 117(10), 5486–5493.
- Butsch, M., & Boris-Lawrie, K. (2000). Translation is not required To generate virion precursor RNA in human immunodeficiency virus type 1-infected T cells. *Journal of Virology*, 74(24), 11531–11537.
- Butsch, Melinda, & Boris-Lawrie, K. (2002). Destiny of Unspliced Retroviral RNA: Ribosome and/or Virion? *Journal of Virology*, 76(7), 3089–3094.
- Butterfield-Gerson, K. L., Scheifele, L. Z., Ryan, E. P., Hopper, A. K., & Parent, L. J. (2006). Importin-beta family members mediate alpharetrovirus gag nuclear entry via interactions with matrix and nucleocapsid. *Journal of Virology*, 80(4), 1798–1806.
- Cain, D., Erlwein, O., Grigg, A., Russell, R. A., & McClure, M. O. (2001). Palindromic Sequence Plays a Critical Role in Human Foamy Virus Dimerization. *Journal of Virology*, 75(8), 3731–3739.
- Campbell, E. M., Nunez, R., & Hope, T. J. (2004). Disruption of the Actin Cytoskeleton Can Complement the Ability of Nef To Enhance Human Immunodeficiency Virus Type 1 Infectivity. *Journal of Virology*, 78(11), 5745–5755.
- Chameettachal, A., Pillai, V., Ali, L., Pitchai, F., Ardah, M., Mustafa, F., Marquet, R., & Rizvi, T. (2018). Biochemical and Functional Characterization of Mouse Mammary Tumor Virus Full-Length Pr77Gag Expressed in Prokaryotic and Eukaryotic Cells. *Viruses*, 10(6), 334.

- Chameettachal, A., Vivet-Boudou, V., Pitchai, F. N. N., Pillai, V. N., Ali, L. M., Krishnan, A., Bernacchi, S., Mustafa, F., Marquet, R., & Rizvi, T. A. (2021). A purine loop and the primer binding site are critical for the selective encapsidation of mouse mammary tumor virus genomic RNA by Pr77Gag. *Nucleic Acids Research*, *49*(8), 4668–4688.
- Chukkapalli, V., Oh, S. J., & Ono, A. (2010). Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(4), 1600–1605.
- Clapham, P. R., & McKnight, A. (2001). HIV-1 receptors and cell tropism. *British Medical Bulletin*, *58*, 43–59.
- Clever, J. L., Daniel Miranda, & Parslow, T. G. (2002). RNA Structure and Packaging Signals in the 5' Leader Region of the Human Immunodeficiency Virus Type 1 Genome. *Journal of Virology*, *76*(23), 12381–12387.
- Cochrane, A. W., McNally, M. T., & Moulard, A. J. (2006). The retrovirus RNA trafficking granule: From birth to maturity. *Retrovirology*, *3*, 18.
- Coffin, J. M., Hughes, S. H., & Varmus, H. E. (Eds.). (1997a). *Retroviruses*. Cold Spring Harbor Laboratory Press.
- Coffin, J. M., Hughes, S. H., & Varmus, H. E. (1997b). Synthesis of Gag and Gag-Pro-Pol Proteins. In *Retroviruses*. Cold Spring Harbor Laboratory Press.
- Comas-Garcia, M., Davis, S. R., & Rein, A. (2016). On the Selective Packaging of Genomic RNA by HIV-1. *Viruses*, *8*(9), 246.
- Cosson, P. (1996). Direct interaction between the envelope and matrix proteins of HIV-1. *The EMBO Journal*, *15*(21), 5783–5788.
- Dagleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F., & Weiss, R. A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, *312*(5996), 763–767.
- Damgaard, C. K., Dyhr-Mikkelsen, H., & Kjems, J. (1998). Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Research*, *26*(16), 3667–3676.
- Darty, K., Denise, A., & Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, *25*(15), 1974–1975.

- Das, A. T., Klaver, B., Klasens, B. I., van Wamel, J. L., & Berkhout, B. (1997). A conserved hairpin motif in the R-U5 region of the human immunodeficiency virus type 1 RNA genome is essential for replication. *Journal of Virology*, *71*(3), 2346–2356.
- Datta, S. A. K., Heinrich, F., Raghunandan, S., Krueger, S., Curtis, J. E., Rein, A., & Nanda, H. (2011). HIV-1 Gag extension: Conformational changes require simultaneous interaction with membrane and nucleic acid. *Journal of Molecular Biology*, *406*(2), 205–214.
- De las Heras, M., Sharp, J. M., Garcia de Jalon, J. A., & Dewar, P. (1991). Enzootic nasal tumour of goats: Demonstration of a type D-related retrovirus in nasal fluids and tumours. *The Journal of General Virology*, *72* (Pt 10), 2533–2535.
- Demirov, D. G., & Freed, E. O. (2004). Retrovirus budding. *Virus Research*, *106*(2), 87–102.
- Dharan, A., Bachmann, N., Talley, S., Zwickelmaier, V., & Campbell, E. M. (2020). Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nature Microbiology*, *5*(9), 1088–1095.
- Didierlaurent, L., Racine, P. J., Houzet, L., Chamontin, C., Berkhout, B., & Mougel, M. (2011). Role of HIV-1 RNA and protein determinants for the selective packaging of spliced and unspliced viral RNA and host U6 and 7SL RNA in virus particles. *Nucleic Acids Research*, *39*(20), 8915–8927.
- Didigu, C. A., & Doms, R. W. (2012). Novel approaches to inhibit HIV entry. *Viruses*, *4*(2), 309–324.
- Ding, P., Kharytonchyk, S., Waller, A., Mbaekwe, U., Basappa, S., Kuo, N., Frank, H. M., Quasney, C., Kidane, A., Swanson, C., Van, V., Sarkar, M., Cannistraci, E., Chaudhary, R., Flores, H., Telesnitsky, A., & Summers, M. F. (2020). Identification of the initial nucleocapsid recognition element in the HIV-1 RNA packaging signal. *Proceedings of the National Academy of Sciences*, *117*(30), 17737.
- D'Souza, V., & Summers, M. F. (2004). Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature*, *431*(7008), 586.
- D'Souza, V., & Summers, M. F. (2005). How retroviruses select their genomes. *Nature Reviews Microbiology*, *3*(8), 643
- Dubois, N., Marquet, R., Paillart, J.-C., & Bernacchi, S. (2018). Retroviral RNA Dimerization: From Structure to Functions. *Frontiers in Microbiology*, *9*, 527.

- Embretson, J. E., & Temin, H. M. (1987). Lack of competition results in efficient packaging of heterologous murine retroviral RNAs and reticuloendotheliosis virus encapsidation-minus RNAs by the reticuloendotheliosis virus helper cell line. *Journal of Virology*, *61*(9), 2675–2683.
- Emerman, M., & Malim, M. H. (1998). HIV-1 regulatory/accessory genes: Keys to unraveling viral and host cell biology. *Science (New York, N.Y.)*, *280*(5371), 1880–1884.
- Endsley, M. A., Somasunderam, A. D., Li, G., Oezguen, N., Thiviyanathan, V., Murray, J. L., Rubin, D. H., Hodge, T. W., O'Brien, W. A., Lewis, B., & Ferguson, M. R. (2014). Nuclear trafficking of the HIV-1 pre-integration complex depends on the ADAM10 intracellular domain. *Virology*, *454–455*, 60–66.
- Enose-Akahata, Y., Vellucci, A., & Jacobson, S. (2017). Role of HTLV-1 Tax and HBZ in the Pathogenesis of HAM/TSP. *Frontiers in Microbiology*, *8*.
- Esquiaqui, J. M., Kharytonchyk, S., Drucker, D., & Telesnitsky, A. (2020). HIV-1 spliced RNAs display transcription start site bias. *RNA (New York, N.Y.)*, *26*(6), 708–714.
- Felsenstein, K. M., & Goff, S. P. (1988). Expression of the gag-pol fusion protein of Moloney murine leukemia virus without gag protein does not induce virion formation or proteolytic processing. *Journal of Virology*, *62*(6), 2179–2182.
- Fernandez, J., Machado, A. K., Lyonnais, S., Chamontin, C., Gärtner, K., Léger, T., Henriquet, C., Garcia, C., Portilho, D. M., Pugnière, M., Chaloin, L., Muriaux, D., Yamauchi, Y., Blaise, M., Nisole, S., & Arhel, N. J. (2019). Transportin-1 binds to the HIV-1 capsid via a nuclear localization signal and triggers uncoating. *Nature Microbiology*, *4*(11), 1840–1850.
- Fields, B.N., David M. Knipe, D.M., and Howley, P.M. (1996). *Fields Virology, Volume 1(3rd Edition) Lippincott-Raven, Philadelphia, PA.*
- Fine, D., & Schochetman, G. (1978). Type D primate retroviruses: A review. *Cancer Research*, *38*(10), 3123–3139.
- Freed, E. O. (2002). Viral Late Domains. *Journal of Virology*, *76*(10), 4679–4687.
- Freed, E. O. (2015). HIV-1 assembly, release and maturation. *Nature Reviews. Microbiology*, *13*(8), 484–496.
- Frolova, L., Le Goff, X., Zhouravleva, G., Davydova, E., Philippe, M., & Kisselev, L. (1996). Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. *RNA (New York, N.Y.)*, *2*(4), 334–341.

- Ganser, L. R., & Al-Hashimi, H. M. (2016). HIV-1 leader RNA dimeric interface revealed by NMR. *Proceedings of the National Academy of Sciences*, *113*(47), 13263–13265.
- Garbitt-Hirst, R., Kenney, S. P., & Parent, L. J. (2009). Genetic evidence for a connection between Rous sarcoma virus gag nuclear trafficking and genomic RNA packaging. *Journal of Virology*, *83*(13), 6790–6797.
- Gaspar, H. B., & Thrasher, A. J. (2005). Gene therapy for severe combined immunodeficiencies. *Expert Opinion on Biological Therapy*, *5*(9), 1175–1182.
- Genetic Disorders*. (n.d.). Genome.Gov. Retrieved April 22, 2021, from <https://www.genome.gov/For-Patients-and-Families/Genetic-Disorders>
- Gherghe, C., Lombo, T., Leonard, C. W., Datta, S. A. K., Bess, J. W., Gorelick, R. J., Rein, A., & Weeks, K. M. (2010). Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(45), 19248–19253.
- Gibbs, J. S., Regier, D. A., & Desrosiers, R. C. (1994). Construction and in vitro properties of SIVmac mutants with deletions in “nonessential” genes. *AIDS Research and Human Retroviruses*, *10*(4), 333–342.
- Gifford, R., & Tristem, M. (2003). The Evolution, Distribution and Diversity of Endogenous Retroviruses. *Virus Genes*, *26*, 291–315.
- Gilboa, E., Mitra, S. W., Goff, S., & Baltimore, D. (1979). A detailed model of reverse transcription and tests of crucial aspects. *Cell*, *18*(1), 93–100. [https://doi.org/10.1016/0092-8674\(79\)90357-X](https://doi.org/10.1016/0092-8674(79)90357-X)
- Gingras, A. C., Raught, B., & Sonenberg, N. (1999). eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. *Annual Review of Biochemistry*, *68*, 913–963.
- Ginn, S. L., Amaya, A. K., Alexander, I. E., Edelstein, M., & Abedi, M. R. (2018). Gene therapy clinical trials worldwide to 2017: An update. *The Journal of Gene Medicine*, *20*(5), e3015.
- Gladnikoff, M., Shimoni, E., Gov, N. S., & Rousso, I. (2009). Retroviral assembly and budding occur through an actin-driven mechanism. *Biophysical Journal*, *97*(9), 2419–2428.
- Goff, S. P. (2018). Chapter 2—Cellular Factors That Regulate Retrovirus Uncoating and Reverse Transcription. In L. J. Parent (Ed.), *Retrovirus-Cell Interactions* (pp. 51–112). Academic Press.

- Gottwein, E., Bodem, J., Müller, B., Schmechel, A., Zentgraf, H., & Kräusslich, H.-G. (2003). The Mason-Pfizer Monkey Virus PPPY and PSAP Motifs Both Contribute to Virus Release. *Journal of Virology*, 77(17), 9474–9485.
- Groom, H. C. T., Anderson, E. C., Dangerfield, J. A., & Lever, A. M. L. (2009). Rev regulates translation of human immunodeficiency virus type 1 RNAs. *The Journal of General Virology*, 90(Pt 5), 1141–1147.
- Grove, J., & Marsh, M. (2011). The cell biology of receptor-mediated virus entry. *Journal of Cell Biology*, 195(7), 1071–1082.
- Gudleski, N., Flanagan, J. M., Ryan, E. P., Bewley, M. C., & Parent, L. J. (2010). Directionality of nucleocytoplasmic transport of the retroviral gag protein depends on sequential binding of karyopherins and viral RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20), 9358–9363.
- Guesdon, F. M. J., Greatorex, J., Rhee, S. R., Fisher, R., Hunter, E., & Lever, A. M. L. (2001). Sequences in the 5' Leader of Mason-Pfizer Monkey Virus Which Affect Viral Particle Production and Genomic RNA Packaging: Development of MPMV Packaging Cell Lines. *Virology*, 288(1), 81–88.
- Guo, X., Roy, B. B., Hu, J., Roldan, A., Wainberg, M. A., & Liang, C. (2005). The R362A mutation at the C-terminus of CA inhibits packaging of human immunodeficiency virus type 1 RNA. *Virology*, 343(2), 190–200.
- Hacein-Bey-Abina, S., Garrigue, A., Wang, G. P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., Asnafi, V., MacIntyre, E., Dal Cortivo, L., Radford, I., Brousse, N., Sigaux, F., Moshous, D., Hauer, J., Borkhardt, A., ... Cavazzana-Calvo, M. (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *The Journal of Clinical Investigation*, 118(9), 3132–3142.
- Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J.-P., Thrasher, A. J., Wulffraat, N., Sorensen, R., Dupuis-Girod, S., Fischer, A., Davies, E. G., Kuis, W., Leiva, L., & Cavazzana-Calvo, M. (2002). Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *The New England Journal of Medicine*, 346(16), 1185–1193.
- Harrison, G. P., Hunter, E., & Lever, A. M. (1995). Secondary structure model of the Mason-Pfizer monkey virus 5' leader sequence: Identification of a structural motif common to a variety of retroviruses. *Journal of Virology*, 69(4), 2175–2186.

- Helps, C. R., & Harbour, D. A. (1997). Comparison of the complete sequence of feline spumavirus with those of the primate spumaviruses reveals a shorter gag gene. *The Journal of General Virology*, 78 (Pt 10), 2549–2564.
- Heng, X., Kharytonchyk, S., Garcia, E. L., Lu, K., Divakaruni, S. S., LaCotti, C., Edme, K., Telesnitsky, A., & Summers, M. F. (2012). Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. *Journal of Molecular Biology*, 417(3), 224–239.
- Hill, M., Tachedjian, G., & Mak, J. (2005). The packaging and maturation of the HIV-1 Pol proteins. *Current HIV Research*, 3(1), 73–85.
- Honigman, A., Wolf, D., Yaish, S., Falk, H., & Panet, A. (1991). Cis acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus. *Virology*, 183(1), 313–319.
- Hu, B., Tai, A., & Wang, P. (2011). Immunization delivered by lentiviral vectors for cancer and infectious diseases. *Immunological Reviews*, 239(1), 45–61.
- Huang, M., Orenstein, J. M., Martin, M. A., & Freed, E. O. (1995). P6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *Journal of Virology*, 69(11), 6810–6818.
- Hunter, E. (2008). Retroviruses: General Features. In B. W. J. Mahy & M. H. V. Van Regenmortel (Eds.), *Encyclopedia of Virology (Third Edition)* (pp. 459–467). Academic Press.
- Hurley, J. H., & Emr, S. D. (2006). The ESCRT complexes: Structure and mechanism of a membrane-trafficking network. *Annual Review of Biophysics and Biomolecular Structure*, 35, 277–298.
- Indik, S., Günzburg, W. H., Salmons, B., & Rouault, F. (2005). A novel, mouse mammary tumor virus encoded protein with Rev-like properties. *Virology*, 337(1), 1–6.
- Jaballah, S. A., Aktar, S. J., Ali, J., Phillip, P. S., Al Dhaheri, N. S., Jabeen, A., & Rizvi, T. A. (2010). A G–C-Rich Palindromic Structural Motif and a Stretch of Single-Stranded Purines Are Required for Optimal Packaging of Mason–Pfizer Monkey Virus (MPMV) Genomic RNA. *Journal of Molecular Biology*, 401(5), 996–1014.
- Jacques, D. A., McEwan, W. A., Hilditch, L., Price, A. J., Towers, G. J., & James, L. C. (2016). HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature*, 536(7616), 349–353.

- Jewell, N. A., & Mansky, L. M. (2000). In the beginning: Genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly. *Journal of General Virology*, *81*(8), 1889–1899.
- Johnson, S. F., & Telesnitsky, A. (2010). Retroviral RNA Dimerization and Packaging: The What, How, When, Where, and Why. *PLOS Pathogens*, *6*(10), e1001007.
- Jones, C. P., Datta, S. A. K., Rein, A., Rouzina, I., & Musier-Forsyth, K. (2011). Matrix Domain Modulates HIV-1 Gag's Nucleic Acid Chaperone Activity via Inositol Phosphate Binding. *Journal of Virology*, *85*(4), 1594–1603.
- Jouvenet, N., Lainé, S., Pessel-Vivares, L., & Mougel, M. (2011). Cell biology of retroviral RNA packaging. *RNA Biology*, *8*(4), 572–580.
- Jouvenet, N., Simon, S. M., & Bieniasz, P. D. (2009). Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proceedings of the National Academy of Sciences*, *106*(45), 19114–19119.
- Kaddis Maldonado, R. J., & Parent, L. J. (2016). Orchestrating the Selection and Packaging of Genomic RNA by Retroviruses: An Ensemble of Viral and Host Factors. *Viruses*, *8*(9).
- Kalloush, R. M., Vivet-Boudou, V., Ali, L. M., Mustafa, F., Marquet, R., & Rizvi, T. A. (2016). Packaging of Mason-Pfizer monkey virus (MPMV) genomic RNA depends upon conserved long-range interactions (LRIs) between U5 and gag sequences. *RNA*, *22*(6), 905–919.
- Kalloush, R. M., Vivet-Boudou, V., Ali, L. M., Pillai, V. N., Mustafa, F., Marquet, R., & Rizvi, T. A. (2019). Stabilizing role of structural elements within the 5' Untranslated Region (UTR) and gag sequences in Mason-Pfizer monkey virus (MPMV) genomic RNA packaging. *RNA Biology*, *16*(5), 612–625.
- Kao, S. Y., Calman, A. F., Luciw, P. A., & Peterlin, B. M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*, *330*(6147), 489–493.
- Karabiber, F., McGinnis, J. L., Favorov, O. V., & Weeks, K. M. (2013). QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA*, *19*(1), 63–73.
- Kay, M. A., Glorioso, J. C., & Naldini, L. (2001). Viral vectors for gene therapy: The art of turning infectious agents into vehicles of therapeutics. *Nature Medicine*, *7*(1), 33–40.
- Kaye, J. F., & Lever, A. M. (1999). Human immunodeficiency virus types 1 and 2 differ in the predominant mechanism used for selection of genomic RNA for encapsidation. *Journal of Virology*, *73*(4), 3023–3031.

- Kaye, Jane F., & Lever, A. M. L. (1998). Nonreciprocal Packaging of Human Immunodeficiency Virus Type 1 and Type 2 RNA: A Possible Role for the p2 Domain of Gag in RNA Encapsidation. *Journal of Virology*, 72(7), 5877–5885.
- Klein, K.C., Reed, J.C., Lingappa, J.R. (2007). Intracellular destinies: degradation, targeting, assembly, and endocytosis of HIV Gag. *AIDS Rev.* 9(3),150-61.
- Keane, S. C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hasic, A., Florwick, A., Santos, J., Bolden, N. C., McCowin, S., Case, D. A., Johnson, B. A., Salemi, M., Telesnitsky, A., & Summers, M. F. (2015). RNA structure. Structure of the HIV-1 RNA packaging signal. *Science (New York, N.Y.)*, 348(6237), 917–921.
- Keane, S. C., Van, V., Frank, H. M., Sciandra, C. A., McCowin, S., Santos, J., Heng, X., & Summers, M. F. (2016). NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome. *Proceedings of the National Academy of Sciences*, 113(46), 13033–13038.
- Kenyon, J. C., Ghazawi, A., Cheung, W. K. S., Phillip, P. S., Rizvi, T. A., & Lever, A. M. L. (2008). The secondary structure of the 5' end of the FIV genome reveals a long-range interaction between R/U5 and gag sequences, and a large, stable stem-loop. *RNA*, 14(12), 2597–2608.
- Kenyon, J. C., & Lever, A. M. (2011). The molecular biology of feline immunodeficiency virus (FIV). *Viruses*, 3(11), 2192–2213.
- Kenyon, J. C., Tanner, S. J., Legiewicz, M., Phillip, P. S., Rizvi, T. A., Le Grice, S. F. J., & Lever, A. M. L. (2011). SHAPE analysis of the FIV Leader RNA reveals a structural switch potentially controlling viral packaging and genome dimerization. *Nucleic Acids Research*, 39(15), 6692–6704.
- Kestler, H., Kodama, T., Ringler, D., Marthas, M., Pedersen, N., Lackner, A., Regier, D., Sehgal, P., Daniel, M., King, N., & Desrosiers, R. (1990). Induction of AIDS in rhesus monkeys by molecularly cloned simian immunodeficiency virus. *Science*, 248(4959), 1109–1112.
- Kharytonchyk, S., Brown, J. D., Stilger, K., Yasin, S., Iyer, A. S., Collins, J., Summers, M. F., & Telesnitsky, A. (2018). Influence of gag and RRE Sequences on HIV-1 RNA Packaging Signal Structure and Function. *Journal of Molecular Biology*, 430(14), 2066–2079.
- Kharytonchyk, S., Monti, S., Smaldino, P. J., Van, V., Bolden, N. C., Brown, J. D., Russo, E., Swanson, C., Shuey, A., Telesnitsky, A., & Summers, M. F. (2016). Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proceedings of the National Academy of Sciences of the United States of America*, 113(47), 13378–13383.

- Kim, S. Y., Byrn, R., Groopman, J., & Baltimore, D. (1989). Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: Evidence for differential gene expression. *Journal of Virology*, *63*(9), 3708–3713.
- Klatzmann, D., Barré-Sinoussi, F., Nugeyre, M. T., Danquet, C., Vilmer, E., Griscelli, C., Brun-Veziret, F., Rouzioux, C., Gluckman, J. C., & Chermann, J. C. (1984). Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science (New York, N.Y.)*, *225*(4657), 59–63.
- Konstantoulas, & Indik. (2014). Mouse mammary tumor virus-based vector transduces non-dividing cells, enters the nucleus via a TNPO3-independent pathway and integrates in a less biased fashion than other retroviruses. *Retrovirology*, *11*, 34.
- Krishnan, A., Pillai, V. N., Chameettachal, A., Mohamed Ali, L., Nuzra Nagoor Pitchai, F., Tariq, S., Mustafa, F., Marquet, R., & A Rizvi, T. (2019). Purification and Functional Characterization of a Biologically Active Full-Length Feline Immunodeficiency Virus (FIV) Pr50Gag. *Viruses*, *11*(8).
- Kubo, Y., Hayashi, H., Matsuyama, T., Sato, H., & Yamamoto, N. (2012). Retrovirus entry by endocytosis and cathepsin proteases. *Advances in Virology*, *2012*, 640894.
- Kutluay, S. B., & Bieniasz, P. D. (2010). Analysis of the Initiating Events in HIV-1 Particle Assembly and Genome Packaging. *PLOS Pathogens*, *6*(11), e1001200.
- Kutluay, S. B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., & Bieniasz, P. D. (2014). Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell*, *159*(5), 1096–1109.
- Kuzembayeva, M., Dilley, K., Sardo, L., & Hu, W.-S. (2014). Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology*, *454–455*(Supplement C), 362–370.
- Lata, S., Ali, A., Sood, V., Raja, R., & Banerjea, A. C. (2015). HIV-1 Rev downregulates Tat expression and viral replication via modulation of NAD(P)H:quinine oxidoreductase 1 (NQO1). *Nature Communications*, *6*(1), 7244.
- Laughrea, M., Jetté, L., Mak, J., Kleiman, L., Liang, C., & Wainberg, M. A. (1997). Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *Journal of Virology*, *71*(5), 3397–3406.

- LeBlanc, J. J., Uddowla, S., Abraham, B., Clatterbuck, S., & Beemon, K. L. (2007). Tap and Dbp5, but not Gag, are involved in DR-mediated nuclear export of unspliced Rous sarcoma virus RNA. *Virology*, *363*(2), 376–386.
- LeBlanc, J., Weil, J., & Beemon, K. (2013). Posttranscriptional regulation of retroviral gene expression: Primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. *Wiley Interdisciplinary Reviews. RNA*, *4*(5).
- Lever, A. M. (2009). RNA packaging in lentiviruses. *Retrovirology*, *6*(S2), I13.
- Lever, A. M. L. (2007). HIV-1 RNA packaging. *Advances in Pharmacology (San Diego, Calif.)*, *55*, 1–32.
- Levin, J. G., & Rosenak, M. J. (1976). Synthesis of murine leukemia virus proteins associated with virions assembled in actinomycin D-treated cells: Evidence for persistence of viral messenger RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *73*(4), 1154–1158.
- Levin, Judith G., Grimley, P. M., Ramseur, J. M., & Berezsky, I. K. (1974). Deficiency of 60 to 70S RNA in Murine Leukemia Virus Particles Assembled in Cells Treated with Actinomycin D. *Journal of Virology*, *14*(1), 152–161.
- Li, C., Burdick, R. C., Nagashima, K., Hu, W.-S., & Pathak, V. K. (2021). HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proceedings of the National Academy of Sciences*, *118*(10).
- Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E. L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A., Divakaruni, S. S., LaCotti, C., Barton, S., Tummlillo, D., Hosis, A., Edme, K., Albrecht, S., Telesnitsky, A., & Summers, M. F. (2011). NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science*, *334*(6053), 242–245.
- Lu, K., Heng, X., & Summers, M. F. (2011). Structural determinants and mechanism of HIV-1 genome packaging. *Journal of Molecular Biology*, *410*(4), 609–633.
- Lundstrom, K. (2018). Viral Vectors in Gene Therapy. *Diseases*, *6*(2).
- MacLachlan, N. J., & Dubovi, E. J. (Eds.). (2017). Chapter 14—Retroviridae. In *Fenner's Veterinary Virology (Fifth Edition)* (pp. 269–297). Academic Press.
- Mailler, E., Bernacchi, S., Marquet, R., Paillart, J.-C., Vivet-Boudou, V., & Smyth, R. P. (2016). The Life-Cycle of the HIV-1 Gag-RNA Complex. *Viruses*, *8*(9).
- Maldonado, R. J. K., Rice, B., Chen, E. C., Tuffy, K. M., Chiari, E. F., Fahrbach, K. M., Hope, T. J., & Parent, L. J. (2020). Visualizing Association of the Retroviral Gag Protein with Unspliced Viral RNA in the Nucleus. *MBio*, *11*(2).

- Mardon, G., & Varmus, H. E. (1983). Frameshift and intragenic suppressor mutations in a rous sarcoma provirus suggest src encodes two proteins. *Cell*, 32(3), 871–879.
- Marsh, J. L., Erfle, M., & Wykes, E. J. (1984). The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. *Gene*, 32(3), 481–485.
- Martin-Serrano, J., & Neil, S. J. D. (2011). Host factors involved in retroviral budding and release. *Nature Reviews. Microbiology*, 9(7), 519–531.
- Masuda, T., Sato, Y., Huang, Y.-L., Koi, S., Takahata, T., Hasegawa, A., Kawai, G., & Kannagi, M. (2015). Fate of HIV-1 cDNA intermediates during reverse transcription is dictated by transcription initiation site of virus genomic RNA. *Scientific Reports*, 5(1), 1–15.
- Mathews, D. H., Sabina, J., Zuker, M., & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, 288(5), 911–940.
- Matreyek, K. A., & Engelman, A. (2013). Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. *Viruses*, 5(10), 2483–2511.
- Matsuoka, M., & Jeang, K.-T. (2011). Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: Viral infectivity, Tax, HBZ and therapy. *Oncogene*, 30(12), 1379–1389.
- McCormack, M. P., & Rabbitts, T. H. (2004). Activation of the T-cell oncogene LMO2 after gene therapy for X-linked severe combined immunodeficiency. *The New England Journal of Medicine*, 350(9), 913–922.
- Merino, E. J., Wilkinson, K. A., Coughlan, J. L., & Weeks, K. M. (2005). RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *Journal of the American Chemical Society*, 127(12), 4223–4231.
- Mertz, J. A., Simper, M. S., Lozano, M. M., Payne, S. M., & Dudley, J. P. (2005). Mouse Mammary Tumor Virus Encodes a Self-Regulatory RNA Export Protein and Is a Complex Retrovirus. *Journal of Virology*, 79(23), 14737–14747.
- Miller, A. D., & Chen, F. (1996). Retrovirus packaging cells based on 10A1 murine leukemia virus for production of vectors that use multiple receptors for cell entry. *Journal of Virology*, 70(8), 5564–5571.

- Miyazaki, Y., Miyake, A., Nomaguchi, M., & Adachi, A. (2011). Structural dynamics of retroviral genome and the packaging. *Frontiers in Microbiology*, 2, 264.
- Moore, M. D., & Hu, W. S. (2009). HIV-1 RNA dimerization: It takes two to tango. *AIDS Reviews*, 11(2), 91–102.
- Moore, M. D., Nikolaitchik, O. A., Chen, J., Hammarskjöld, M.-L., Rekosh, D., & Hu, W.-S. (2009). Probing the HIV-1 Genomic RNA Trafficking Pathway and Dimerization by Genetic Recombination and Single Virion Analyses. *PLOS Pathogens*, 5(10), e1000627.
- Mortimer, S. A., & Weeks, K. M. (2007). A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *Journal of the American Chemical Society*, 129(14), 4144–4145.
- Mortimer, S. A., & Weeks, K. M. (2009). Time-resolved RNA SHAPE chemistry: Quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nature Protocols*, 4(10), 1413–1421.
- Murakami, T., & Freed, E. O. (2000). The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. *Proceedings of the National Academy of Sciences of the United States of America*, 97(1), 343–348.
- Murphy, R. E., & Saad, J. S. (2020). The Interplay between HIV-1 Gag Binding to the Plasma Membrane and Env Incorporation. *Viruses*, 12(5).
- Mustafa, F., Amri, D. A., Ali, F. A., Sari, N. A., Suwaidi, S. A., Jayanth, P., Philips, P. S., & Rizvi, T. A. (2012). Sequences within Both the 5' UTR and Gag Are Required for Optimal In Vivo Packaging and Propagation of Mouse Mammary Tumor Virus (MMTV) Genomic RNA. *PLOS ONE*, 7(10), e47088.
- Mustafa, F., Jayanth, P., Phillip, P. S., Ghazawi, A., Schmidt, R. D., Lew, K. A., & Rizvi, T. A. (2005). Relative activity of the feline immunodeficiency virus promoter in feline and primate cell lines. *Microbes and Infection*, 7(2), 233–239.
- Mustafa, F., Lew, K. A., Schmidt, R. D., Browning, M. T., & Rizvi, T. A. (2004). Mutational analysis of the predicted secondary RNA structure of the Mason-Pfizer monkey virus packaging signal. *Virus Research*, 99(1), 35–46.
- Mustafa, F., Vivet-Boudou, V., Jabeen, A., Ali, L. M., Kalloush, R. M., Marquet, R., & Rizvi, T. A. (2018). The bifurcated stem loop 4 (SL4) is crucial for efficient packaging of mouse mammary tumor virus (MMTV) genomic RNA. *RNA Biology*, 15(8), 1047–1059.

- Naldini, L., Blömer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F. H., Verma, I. M., & Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science*, 272(5259), 263–267.
- Nicot, C., Harrod, R. L., Ciminale, V., & Franchini, G. (2005). Human T-cell leukemia/lymphoma virus type 1 nonstructural genes and their functions. *Oncogene*, 24(39), 6026–6034.
- Nyborg, J. K., Egan, D., & Sharma, N. (2010). The HTLV-1 Tax protein: Revealing mechanisms of transcriptional activation through histone acetylation and nucleosome disassembly. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1799(3), 266–274.
- Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J., Brierley, I. (2006). A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, 441(7090), 244-7.
- Ogert, R. A., Lee, L. H., & Beemon, K. L. (1996). Avian retroviral RNA element promotes unspliced RNA accumulation in the cytoplasm. *Journal of Virology*, 70(6), 3834–3843.
- Olety, B., & Ono, A. (2014). Roles played by acidic lipids in HIV-1 Gag membrane binding. *Virus Research*, 193, 108–115.
- Ono, A., Ablan, S. D., Lockett, S. J., Nagashima, K., & Freed, E. O. (2004). Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proceedings of the National Academy of Sciences of the United States of America*, 101(41), 14889–14894.
- Overbaugh, J., Miller, A. D., & Eiden, M. V. (2001). Receptors and entry cofactors for retroviruses include single and multiple transmembrane-spanning proteins as well as newly described glycoposphatidylinositol-anchored and secreted proteins. *Microbiology and Molecular Biology Reviews: MMBR*, 65(3), 371–389, table of contents.
- Paillart, J. C., Skripkin, E., Ehresmann, B., Ehresmann, C., & Marquet, R. (1996). A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 93(11), 5572–5577.
- Paillart, J. C., Westhof, E., Ehresmann, C., Ehresmann, B., & Marquet, R. (1997). Non-canonical interactions in a kissing loop complex: The dimerization initiation site of HIV-1 genomic RNA. *Journal of Molecular Biology*, 270(1), 36–49.

- Paillart, J.-C., Dettenhofer, M., Yu, X., Ehresmann, C., Ehresmann, B., & Marquet, R. (2004). First Snapshots of the HIV-1 RNA Structure in Infected Cells and in Virions. *Journal of Biological Chemistry*, 279(46), 48397–48403.
- Parent, L. J. (2011). New insights into the nuclear localization of retroviral Gag proteins. *Nucleus*, 2(2), 92–97.
- Pasquinelli, A. E., Ernst, R. K., Lund, E., Grimm, C., Zapp, M. L., Rekosh, D., Hammarskjöld, M.-L., & Dahlberg, J. E. (1997). The constitutive transport element (CTE) of Mason–Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. *The EMBO Journal*, 16(24), 7500–7510.
- Pederson, F. S., & Duch, M. (2006). Retroviral replication. *Encyclopedia of Life Sciences*.
- Pessel-Vivares, L., Ferrer, M., Lainé, S., & Mougel, M. (2014). MLV requires Tap/NXF1-dependent pathway to export its unspliced RNA to the cytoplasm and to express both spliced and unspliced RNAs. *Retrovirology*, 11(1), 21.
- Pincetic, A., & Leis, J. (2009, March 5). *The Mechanism of Budding of Retroviruses from Cell Membranes* [Review Article]. *Advances in Virology*; Hindawi.
- Pitchai, F. N. N., Ali, L., Pillai, V. N., Chameettachal, A., Ashraf, S. S., Mustafa, F., Marquet, R., & Rizvi, T. A. (2018). Expression, purification, and characterization of biologically active full-length Mason-Pfizer monkey virus (MPMV) Pr78 Gag. *Scientific Reports*, 8(1), 11793.
- Pitchai, F. N. N., Chameettachal, A., Vivet-Boudou, V., Ali, L., Pillai, V. N., Krishnan, A., Bernacchi, S., Mustafa, F., Marquet, R., & Rizvi, T. A. (2021). Identification of Pr78Gag binding sites on the Mason-Pfizer monkey virus genomic RNA packaging determinants. *Journal of Molecular Biology*, 166923.
- Poeschla, E. M., Wong-Staal, F., & Looney, D. J. (1998). Efficient transduction of nondividing human cells by feline immunodeficiency virus lentiviral vectors. *Nature Medicine*, 4(3), 354.
- Popov, S., Rexach, M., Zybarth, G., Reiling, N., Lee, M. A., Ratner, L., Lane, C. M., Moore, M. S., Blobel, G., & Bukrinsky, M. (1998). Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. *The EMBO Journal*, 17(4), 909–917.
- Purcell, D. F., & Martin, M. A. (1993). Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *Journal of Virology*, 67(11), 6365–6378.
- Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11, 129.

- Ricaña, C. L., & Johnson, M. C. (2021). *An infectious Rous Sarcoma Virus Gag mutant that is defective in nuclear cycling* [Preprint]. Microbiology.
- Rice, A. P. (2017). The HIV-1 Tat protein: Mechanism of action and target for HIV-1 cure strategies. *Current Pharmaceutical Design*, 23(28), 4098–4102.
- Rizvi, T. A., Lew, K. A., Murphy, E. C., & Schmidt, R. D. (1996). Role of Mason-Pfizer monkey virus (MPMV) constitutive transport element (CTE) in the propagation of MPMV vectors by genetic complementation using homologous/heterologous env genes. *Virology*, 224(2), 517–532.
- Rizvi, T. A., & Panganiban, A. T. (1993). Simian immunodeficiency virus RNA is efficiently encapsidated by human immunodeficiency virus type 1 particles. *Journal of Virology*, 67(5), 2681–2688.
- Rizvi, T. A., Kenyon, J. C., Ali, J., Aktar, S. J., Phillip, P. S., Ghazawi, A., Mustafa, F., & Lever, A. M. L. (2010). Optimal packaging of FIV genomic RNA depends upon a conserved long-range interaction and a palindromic sequence within gag. *Journal of Molecular Biology*, 403(1), 103–119.
- Rizvi, T. A., Schmidt, R. D., Lew, K. A., & Keeling, M. E. (1996). Rev/RRE-Independent Mason–Pfizer Monkey Virus Constitutive Transport Element-Dependent Propagation of SIVmac239 Vectors Using a Single Round of Replication Assay. *Virology*, 222(2), 457–463.
- Robbins, P. D., & Ghivizzani, S. C. (1998). Viral vectors for gene therapy. *Trends in Biotechnology*, 16(1), 35–40.
- Rose, K.M., Hirsch, V.M., Bouamr, F. (2020). Budding of a Retrovirus: Some Assemblies Required. *Viruses*. 12(10),1188.
- Ross, S. R., Schofield, J. J., Farr, C. J., & Bucan, M. (2002). Mouse transferrin receptor 1 is the cell entry receptor for mouse mammary tumor virus. *Proceedings of the National Academy of Sciences of the United States of America*, 99(19), 12386–12390.
- Rothenberg, E., & Baltimore, D. (1976). Synthesis of Long, Representative DNA Copies of the Murine RNA Tumor Virus Genome. *Journal of Virology*, 17(1), 168–174.
- Rothenberg, E., Smotkin, D., Baltimore, D., & Weinberg, R. A. (1977). In vitro synthesis of infectious DNA of murine leukaemia virus. *Nature*, 269(5624), 122–126.
- Rulli, S. J., Hibbert, C. S., Mirro, J., Pederson, T., Biswal, S., & Rein, A. (2007). Selective and nonselective packaging of cellular RNAs in retrovirus particles. *Journal of Virology*, 81(12), 6623–6631.

- Scheifele, L. Z., Garbitt, R. A., Rhoads, J. D., & Parent, L. J. (2002). Nuclear entry and CRM1-dependent nuclear export of the Rous sarcoma virus Gag polyprotein. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6), 3944–3949.
- Scheifele, L. Z., Kenney, S. P., Cairns, T. M., Craven, R. C., & Parent, L. J. (2007). Overlapping roles of the Rous sarcoma virus Gag p10 domain in nuclear export and virion core morphology. *Journal of Virology*, 81(19), 10718–10728.
- Scheifele, L. Z., Ryan, E. P., & Parent, L. J. (2005). Detailed mapping of the nuclear export signal in the Rous sarcoma virus Gag protein. *Journal of Virology*, 79(14), 8732–8741.
- Schmidt, R. D., Mustafa, F., Lew, K. A., Browning, M. T., & Rizvi, T. A. (2003). Sequences within both the 5' untranslated region and the gag gene are important for efficient encapsidation of Mason-Pfizer monkey virus RNA. *Virology*, 309(1), 166–178.
- Shida, H. (2012). Role of Nucleocytoplasmic RNA Transport during the Life Cycle of Retroviruses. *Frontiers in Microbiology*, 3.
- Shukla, E., & Chauhan, R. (2019). Host-HIV-1 Interactome: A Quest for Novel Therapeutic Intervention. *Cells*, 8(10).
- Smyth, R. P., Despons, L., Huili, G., Bernacchi, S., Hijnen, M., Mak, J., Jossinet, F., Weixi, L., Paillart, J.-C., von Kleist, M., & Marquet, R. (2015). Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nature Methods*, 12(9), 866–872.
- Smyth, R. P., Smith, M. R., Jousset, A.-C., Despons, L., Laumond, G., Decoville, T., Cattenoz, P., Moog, C., Jossinet, F., Mougél, M., Paillart, J.-C., von Kleist, M., & Marquet, R. (2018). In cell mutational interference mapping experiment (in cell MIME) identifies the 5' polyadenylation signal as a dual regulator of HIV-1 genomic RNA production and packaging. *Nucleic Acids Research*.
- Sonigo, P., Barker, C., Hunter, E., & Wain-Hobson, S. (1986). Nucleotide sequence of Mason-Pfizer monkey virus: An immunosuppressive D-type retrovirus. *Cell*, 45(3), 375–385.
- South, T. L., & Summers, M. F. (1993). Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Science: A Publication of the Protein Society*, 2(1), 3–19.
- Strack, B., Calistri, A., Craig, S., Popova, E., & Göttlinger, H. G. (2003). AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell*, 114(6), 689–699.

- Summers, M. F., Henderson, L. E., Chance, M. R., Bess, J. W., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder, R. C., & Hare, D. R. (1992). Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Science : A Publication of the Protein Society*, 1(5), 563–574.
- Sundquist, W. I., & Kräusslich, H.-G. (2012). HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harbor Perspectives in Medicine*, 2(7).
- Swanson, C. M., & Malim, M. H. (2006). Retrovirus RNA trafficking: From chromatin to invasive genomes. *Traffic (Copenhagen, Denmark)*, 7(11), 1440–1450.
- Swanstrom, R., & Wills, J. W. (2011). *Synthesis, assembly, and processing of viral proteins*.
- Tan, W., Felber, B. K., Zolotukhin, A. S., Pavlakis, G. N., & Schwartz, S. (1995). Efficient expression of the human papillomavirus type 16 L1 protein in epithelial cells by using Rev and the Rev-responsive element of human immunodeficiency virus or the cis-acting transactivation element of simian retrovirus type 1. *Journal of Virology*, 69(9), 5607–5620.
- Truant, R., & Cullen, B. R. (1999). The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Molecular and Cellular Biology*, 19(2), 1210–1217.
- D'Souza, V., Melamed, J., Habib, D., Pullen, K., Wallace, K., Summers, M. F. (2001). Identification of a high affinity nucleocapsid protein binding element within the Moloney murine leukemia virus Psi-RNA packaging signal: implications for genome recognition. *J Mol Biol.* 314(2), 217-32.
- Vigna, E., & Naldini, L. (2000). Lentiviral vectors: Excellent tools for experimental gene transfer and promising candidates for gene therapy. *The Journal of Gene Medicine*, 2(5), 308–316.
- Vile, R. G., Ali, M., Hunter, E., & McClure, M. O. (1992). Identification of a generalised packaging sequence for D-type retroviruses and generation of a D-type retroviral vector. *Virology*, 189(2), 786–791.
- White, S. M., Renda, M., Nam, N.-Y., Klimatcheva, E., Zhu, Y., Fisk, J., Halterman, M., Rimel, B. J., Federoff, H., & Pandya, S. (1999). Lentivirus vectors using human and simian immunodeficiency virus elements. *Journal of Virology*, 73(4), 2832–2840.

- Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., Giddings, M. C., & Weeks, K. M. (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology*, *6*(4), e96.
- Wills, J. W., Cameron, C. E., Wilson, C. B., Xiang, Y., Bennett, R. P., & Leis, J. (1994). An assembly domain of the Rous sarcoma virus Gag protein required late in budding. *Journal of Virology*, *68*(10), 6605–6618.
- Wu, W., Hatterschide, J., Syu, Y.-C., Cantara, W. A., Blower, R. J., Hanson, H. M., Mansky, L. M., & Musier-Forsyth, K. (2018). Human T-cell leukemia virus type 1 Gag domains have distinct RNA-binding specificities with implications for RNA packaging and dimerization. *The Journal of Biological Chemistry*, *293*(42), 16261–16276.
- Wu, Y. (2004). HIV-1 gene expression: Lessons from provirus and non-integrated DNA. *Retrovirology*, *1*, 13.
- Wu, Y., & Marsh, J. W. (2003). Early Transcription from Nonintegrated DNA in Human Immunodeficiency Virus Infection. *Journal of Virology*, *77*(19), 10376–10382.
- Wyma, D. J., Kotov, A., & Aiken, C. (2000). Evidence for a stable interaction of gp41 with Pr55(Gag) in immature human immunodeficiency virus type 1 particles. *Journal of Virology*, *74*(20), 9381–9387.
- Xiang, Y., Cameron, C. E., Wills, J. W., & Leis, J. (1996). Fine mapping and characterization of the Rous sarcoma virus Pr76gag late assembly domain. *Journal of Virology*, *70*(8), 5695–5700.
- Yamashita, M., & Emerman, M. (2006). Retroviral infection of non-dividing cells: Old and new perspectives. *Virology*, *344*(1), 88–93.
- Yang, S., & Temin, H. M. (1994). A double hairpin structure is necessary for the efficient encapsidation of spleen necrosis virus retroviral RNA. *The EMBO Journal*, *13*(3), 713–726.
- Yilmaz, A., Bolinger, C., & Boris-Lawrie, K. (2006). Retrovirus Translation Initiation: Issues and Hypotheses Derived from Study of HIV-1. *Current HIV Research*, *4*(2), 131–139.
- Yin, P. D., & Hu, W.-S. (1997). RNAs from genetically distinct retroviruses can copackage and exchange genetic information in vivo. *Journal of Virology*, *71*(8), 6237–6242.
- Yoshinaka, Y., Katoh, I., Copeland, T. D., & Oroszlan, S. (1985). Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease. *Journal of Virology*, *55*(3), 870–873.

- Yu, S. F., Sullivan, M. D., & Linial, M. L. (1999). Evidence that the Human Foamy Virus Genome Is DNA. *Journal of Virology*, *73*(2), 1565–1572.
- Yu, X., Yuan, X., McLane, M. F., Lee, T. H., & Essex, M. (1993). Mutations in the cytoplasmic domain of human immunodeficiency virus type 1 transmembrane protein impair the incorporation of Env proteins into mature virions. *Journal of Virology*, *67*(1), 213–221.
- Zeffman, A., Hassard, S., Varani, G., & Lever, A. (2000). The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein. *Journal of Molecular Biology*, *297*(4), 877–893.
- Zhang, H., Dornadula, G., Orenstein, J., & Pomerantz, R. J. (2000). Morphologic changes in human immunodeficiency virus type 1 virions secondary to intravirion reverse transcription: Evidence indicating that reverse transcription may not take place within the intact viral core. *Journal of Human Virology*, *3*(3), 165–172.
- Zhou, J., Bean, R. L., Vogt, V. M., & Summers, M. F. (2007). Solution Structure of the Rous Sarcoma Virus Nucleocapsid Protein:μΨ RNA Packaging Signal Complex. *Journal of Molecular Biology*, *365*(2), 453–467.
- Zhou, J., McAllen, J. K., Tailor, Y., & Summers, M. F. (2005). High affinity nucleocapsid protein binding to the muPsi RNA packaging signal of Rous sarcoma virus. *Journal of Molecular Biology*, *349*(5), 976–988.
- Zila, V., Margiotta, E., Turonova, B., Müller, T. G., Zimmerli, C. E., Mattei, S., Allegretti, M., Börner, K., Rada, J., Müller, B., Lusic, M., Kräusslich, H.-G., & Beck, M. (2020). Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *BioRxiv*, 2020.07.30.193524.
- Zolotukhin, A. S., Michalowski, D., Smulevitch, S., & Felber, B. K. (2001). Retroviral Constitutive Transport Element Evolved from Cellular TAP(NXF1)-Binding Sequences. *Journal of Virology*, *75*(12), 5567–5575.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415.

LIST OF PUBLICATIONS

Publications Related to thesis:

1. **Ali, L. M.**, Pitchai, F. N. N., Vivet-Boudou V., Chameettachal, A., Jabeen, A., Pillai, V. N., Mustafa, F., Marquet, R., Rizvi, T. A. **2020**. Role of purine-rich regions in MasonPfizer monkey virus (MPMV) genomic RNA packaging and propagation. *Front Microbiol.* Nov 5; 11:595410.
[DOI:10.3389/fmicb.2020.595410](https://doi.org/10.3389/fmicb.2020.595410)
2. Pitchai, F. N. N., Chameettachal, A., Vivet Boudou, V., **Ali, L. M.**, Pillai, V. N., Bernacchi, S., Mustafa, F., Marquet, R., Rizvi, T. A. **2021**. Identification of Pr78^{Gag} on the Mason-Pfizer monkey virus genomic RNA packaging determinants. *Journal of Molecular Biology*, 166923.
[DOI: 10.1016/j.jmb.2021.166923](https://doi.org/10.1016/j.jmb.2021.166923)

Publications Non-Related to Thesis:

1. Pillai, V. N., **Ali, L. M.**, Prabhu, S. G., Krishnan, A., Chameettachal, A., Pitchai, F. N. N., Mustafa, F., Rizvi, T. A. **2021**. A Stretch of Unpaired Purines in the Leader Region of Simian Immunodeficiency Virus (SIV) is Crucial for its Genomic RNA Packaging. *J Mol Biol* (Submitted).
2. Chameettachal, A., Pitchai, F. N. N., Vivet Boudou, V., Pillai, V. N., **Ali, L. M.**, Mustafa, F., Marquet, R., Rizvi, T. A. **2021**. Selective capture of gRNA by the MMTV Pr77^{Gag} requires single-stranded purines present within the bifurcated stem loop 4. *Nucleic Acids Research* 49(8):4668-4688.
[DOI:10.1093/nar/gkab223](https://doi.org/10.1093/nar/gkab223)
3. Krishnan, A., Pillai, V. N., Chameettachal, A., **Ali, L. M.**, Nuzra NagoorPitchai, F., Tariq, S., Mustafa, F., Marquet, R., and A Rizvi, T. **2019**. Purification and Functional Characterization of a Biologically Active Full-Length Feline Immunodeficiency Virus (FIV) Pr50^{Gag}. *Viruses* 11(8):689.
[DOI:10.3390/v11080689](https://doi.org/10.3390/v11080689)
4. Kalloush, R. M., Vivet-Boudou, V., **Ali, L. M.**, Pillai, V. N., Mustafa, F., Marquet, R., Rizvi, T. A. **2019**. Stabilizing role of structural elements within the 5' Untranslated Region (UTR) and gag sequences in Mason-Pfizer monkey virus (MPMV) genomic RNA packaging. *RNA Biol* 16: 612–625.
[DOI:10.1080/15476286.2019.1572424](https://doi.org/10.1080/15476286.2019.1572424)
5. Akhlaq, S., Panicker, N. G., Philip, P. S., **Ali, L. M.**, Dudley, J. P., Rizvi, T. A., Mustafa, F. **2018**. A cis-Acting Element Downstream of the Mouse Mammary Tumor Virus Major Splice Donor Critical for RNA Elongation and Stability. *J Mol Biol* 430: 4307–4324. [DOI:10.1016/j.jmb.2018.08.025](https://doi.org/10.1016/j.jmb.2018.08.025)

6. Pitchai, F. N. N., **Ali, L. M.**, Pillai, V. N., Chameettachal, A., Ashraf, S. S., Mustafa, F., Marquet, R., Rizvi, T. A. **2018**. Expression, purification, and characterization of biologically active full length Mason-Pfizer monkey virus (MPMV) Pr78^{Gag}. *Sci Rep* 08 7; 8(1):11793.
[DOI:10.1038/s41598-018-30142-0](https://doi.org/10.1038/s41598-018-30142-0)
7. Chameettachal, A., Pillai, V. N., **Ali, L. M.**, Pitchai, F. N. N., Ardah, M. T., Mustafa, F., Marquet, R., Rizvi, T. A. **2018**. Biochemical and functional characterization of mouse mammary tumor virus (MMTV) full-length Pr77Gag expressed in prokaryotic and eukaryotic cells. *Viruses* 06 18; 10(6).
[DOI:10.3390/v10060334](https://doi.org/10.3390/v10060334)
8. Mustafa, F., Vivet-Boudou, V., Jabeen, A., **Ali, L. M.**, Kalloush, R. M., Marquet, R., Rizvi, T. A. **2018**. The bifurcated stem loop 4 (SL4) is crucial for efficient packaging of mouse mammary tumor virus (MMTV) genomic RNA. *RNA Biol* 15:8, 1047-1059. [DOI:10.1080/15476286.2018.1486661](https://doi.org/10.1080/15476286.2018.1486661)
9. **Ali, L. M.**, Rizvi, T. A., Mustafa, F. **2016**. Cross- and Co-Packaging of Retroviral RNAs and their Consequences. *Viruses*, 8(10), 276.
[DOI:10.3390/v8100276](https://doi.org/10.3390/v8100276)
10. Kalloush, R. M., Vivet-Boudou, V., **Ali, L. M.**, Mustafa, F., Marquet, R., Rizvi, T. A. **2016**. Packaging of Mason-Pfizer Monkey Virus (MPMV) Genomic RNA Depends upon Conserved Long Range Interactions (LRIs) Between U5 and gag Sequences. *RNA*, 22:905-919.
[DOI:10.1261/rna.055731.115](https://doi.org/10.1261/rna.055731.115)
11. Al Ahmad, M., Mustafa, F., **Ali, L. M.**, Karakkat, J. V., Rizvi, T. A. **2015**. Label-free capacitance-based identification of viruses. *Sci Rep*, 5:9809.
[DOI:10.1038/srep09809](https://doi.org/10.1038/srep09809)
12. Aktar, S. J., Vivet-Boudou, V., **Ali, L. M.**, Jabeen, A., Kalloush, R. M., Richer, D., Mustafa, F., Marquet, R., Rizvi, T. A. **2014**. Structural basis of genomic RNA (gRNA) dimerization and packaging determinants of mouse mammary tumor virus (MMTV). *Retrovirology* 11, 96. [DOI:10.1186/s12977-014-0096-6](https://doi.org/10.1186/s12977-014-0096-6)
13. Al Ahmad, M., Mustafa, F., **Ali, L. M.**, Rizvi, T. A. **2014**. Virus detection and quantification using electrical parameters. *Sci Rep*, 4:6831.
[DOI:10.1038/srep06831](https://doi.org/10.1038/srep06831)
14. Aktar, S. J., Jabeen, A., **Ali, L. M.**, Vivet-Boudou, V., Marquet, R., Rizvi, T. A. **2013**. SHAPE analysis of the 5' end of the Mason-Pfizer monkey virus (MPMV) genomic RNA reveals structural elements required for genome dimerization. *RNA*, 19:1648-58. [DOI: 10.1261/rna.040931.113](https://doi.org/10.1261/rna.040931.113)
15. Saeed, I. A., **Ali, L. M.**, Jabeen, A., Khasawneh, M., Rizvi, T. A., and Ashraf, S. S. **2013**. Estrogenic activities of ten medicinal herbs from the Middle East. *Journal of Chromatographic Science*, 51:33-39.
[DOI:10.1093/chromsci/bms101](https://doi.org/10.1093/chromsci/bms101)

APPENDIX A

Appendix A. Description of primers and DNA template used for cloning, sequencing, conventional, and real time PCR

Oligo Name	*S or AS	Clone Names	Oligo Sequence	DNA template used for SOE PCR	**Nucleotide (nt) Position and/or Reference	Virus Region & Gene
OTR 787	S	Outer Primers	5' CC ctcgagT GTC CGG AGC CGT GCT GCC CG 3'	These oligos were used as outer primers to generate each mutant	2 dummies, <i>XhoI</i> , nt 7-21	MPMV 5' LTR
OTR 788	AS		5' CCC gga tcc TTC TTT CTT ATC TAT CAA TTC TTT AAT TAA G 3'		3 dummies, <i>BamHI</i> , MPMV nt 1171-1141	MPMV <i>gag</i>
OTR 1348	S	LA-I	5' GAA AGT AAA CTC TCT TGG CC 3'	SJ2	MPMV nt 820-839	MPMV U5
OTR 1352	AS		5' GCC AAG AGA GTT TAC TTT CTA ATC GCC GGC CGG CGA ACG 3'		MPMV nt 838- 793 with deletion from nt 819-813	MPMV U5
OTR 1012	S	LA-II	5' GGC CGG CGA ACT CTC TTG GCC GCC GCG GG 3'	SJ2	MPMV nt 803-847 with deletion from nt 812-827	MPMV U5
OTR 1013	AS		5' GCC AAG AGA GTT CGC CGG CCG GCG AAC GC 3'		MPMV nt 838-794 with deletion from nt 827-812	MPMV U5
OTR 1014	S	LA-III	5' AAT TTT CAC TTT CAT TAC TCT CTT GGC CGC CGC GGG 3'	SJ2	Substitution, MPMV nt 828- 847	MPMV U5

OTR 1015	AS		5' GCC AAG AGA GTA ATG AAA GTG AAA ATT TCG CCG GCC GGC GAA CGC 3'		MPMV nt 838-794 with substitution from nt827-812	MPMV U5
OTR 1006	S	LA-IV	5' GGA CCT GTG TTG CGC TCG GAT ATG GG 3'	LA-II	MPMV nt 861-894 with deletion from nt 867-874	MPMV U5
OTR 1007	AS		5' CAA CAC AGG TCC AAC GCG GCA GGT TC 3'		MPMV nt 880-846 with deletion from nt 874-867	MPMV U5
OTR 1028	S	LA-V	5' TTA AAA GTA CTC TCT TGG CCG CCG CGG G 3'	LA-IV	MPMV nt 812-847 with deletion from nt 820-827	MPMV U5
OTR 1029	AS		5' GCC AAG AGA GTA CTT TTA ATC GCC GGC CGG C 3'		MPMV nt 838-800 with deletion from nt 827-820	MPMV U5
OTR 1006	S	LA-VI	5' GGA CCT GTG TTG CGC TCG GAT ATG GG 3'	SJ2	MPMV nt 861-894 with deletion from nt 867-874	MPMV U5
OTR 1007	AS		5' CAA CAC AGG TCC AAC GCG GCA GGT TC 3'		MPMV nt 880-846 with deletion from nt 874-867	MPMV U5
OTR 1139	S	LA-VII	5' GAA AGT AAT TGG CCG CCG CGG GAA C 3'	SJ2	MPMV nt 820-850 with deletion from nt 828-833	MPMV U5
OTR 1140	AS		5' GGC CAA TTA CTT TCA CTT TTA ATC GCC G 3'		MPMV nt 839-806 with deletion from nt 833-828	MPMV U5
OTR 1137	S	LA-VIII	5' GGA CCT AAG TGT TGC GCT CGG ATA TG 3'	SJ2	MPMV nt 861-866 with deletion from nt 873-892	MPMV U5

OTR 1138	AS		5' CAC TTA GGT CCA ACG CGG CAG G 3'		MPMV nt 877-873 with deletion from nt 866-850	MPMV U5
OTR 1139	S	LA-VIII	5' GAA AGT AAT TGG CCG CCG CGG GAA C 3'	Deletion of complementary region to bpPurines	MPMV nt 820-827 with deletion from nt 834-850	MPMV U5
OTR 1140	AS		5' GGC CAA TTA CTT TCA CTT TTA ATC GCC G 3'		MPMV nt 839-834 with deletion from nt 827-806	MPMV U5
OTR 1069	S	LA-IX	5' CTC TCA AAG TGT TGC GCT CGG ATA TGG 3'	SJ2	MPMV nt 873-893	MPMV U5
OTR 1070	AS		5' AGC GCA ACA CTT TGA GAG AGG TCC AAC GCG GCA GGT TC 3'		MPMV nt 884-875 with deletion from nt 866-845	MPMV U5
OTR 1065	S	LA-IX	5' TGA AAG TTG GCC GCC GCG GGA ACC 3'	Substitution of complementary region to bpPurines	MPMV nt 834-851	MPMV U5
OTR 1066	AS		5' GCG GCC AAC TTT CAT TAC TTT CAC TTT TAA TCG CCG G 3'		MPMV nt 841-834 with deletion from nt 827-805	MPMV U5
OTR 1234	S	LA-X	5' GCG CTC GGA TAT GGG GCA AG 3'	LA-II	MPMV nt 880-899	MPMV U5
OTR 1235	AS		5' GCC CCA TAT CCG AGC GCT TAC TTT TCA GGT CCA ACG CGG 3'		MPMV nt 896-880 with deletion from nt 874-854	
OTR 1234	S	LA-XI	5' GCG CTC GGA TAT GGG GCA AG 3'	SJ2	MPMV nt 880-899	MPMV U5
OTR 1235	AS		5' GCC CCA TAT CCG AGC GCT TAC TTT TCA GGT CCA ACG CGG 3'		MPMV nt 896-880 with deletion from nt 874-854	

Sequence in lower case: non-viral and/or restriction enzyme sequences that were introduced in the oligos for cloning purposes.

Sequence in bold: T7 promoter.

*S, sense; AS, antisense.

** The MPMV nucleotide numbering system refers to the genome sequence deposited in the Genbank (accession number M12349) by Sonigo et al. 1986.

APPENDIX B

Appendix B. Mean SHAPE reactivity data from triplicate experiments with SD values for wild type (SJ2/RCR001) and LA/FN clones.

* RD: Mean reactivity data for triplicate experiments ** Standard deviation ***Δ Deleted nucleotides in the relevant clones

Nucleotide number	Nucleotides	WT/SJ2 (RCR001)	LA-I/FN-I		LA-II/FN-II		LA-III/FN-III		LA-IV/FN-IV		LA-V/FN-V		LA-VI/FN-VI		LA-VII/FN-VII	
		RD*	RD	SD**	RD	SD	RD	SD	RD	SD	RD	SD	RD	SD	RD	SD
1	G	ND	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
2	C	ND	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
3	C	ND	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
4	A	0.24	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
5	C	0.49	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
6	C	0.67	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
7	A	0.64	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
8	U	1.03	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
9	U	0.78	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
10	A	0.70	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
11	A	0.45	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
12	A	0.53	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
13	U	0.38	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
14	G	0.11	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
15	A	0.08	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
16	G	0.10	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
17	A	0.12	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
18	C	0.07	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
19	U	0.12	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-

20	U	0.17	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
21	G	0.13	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
22	A	0.46	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
23	U	0.53	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
24	C	0.17	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
25	A	0.31	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
26	G	0.35	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
27	A	0.70	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
28	A	0.89	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
29	C	0.48	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
30	A	0.98	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
31	C	0.19	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
32	U	0.17	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
33	G	0.10	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
34	U	0.05	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
35	C	0.06	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
36	U	0.05	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
37	U	0.08	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
38	G	0.15	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
39	U	0.08	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
40	C	0.06	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
41	U	0.10	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
42	C	0.53	ND	-	ND	-	ND	-	ND	-	ND	-	ND	-
43	C	0.45	ND	-	ND	-	ND	-	0.30	0.42	ND	-	ND	-
44	A	0.49	ND	-	ND	-	ND	-	0.07	0.10	ND	-	ND	-
45	U	0.46	ND	-	ND	-	ND	-	0.22	0.28	ND	-	ND	-
46	U	0.33	ND	-	ND	-	ND	-	0.19	0.19	ND	-	ND	-

47	U	0.24	ND -	ND -	ND -	ND -	0.01	0.01	ND -	ND -
48	C	0.07	ND -	ND -	ND -	ND -	0.01	0.01	ND -	ND -
49	U	0.24	ND -	ND -	ND -	ND -	1.47	2.07	ND -	ND -
50	U	0.25	ND -	ND -	ND -	ND -	0.30	0.26	ND -	ND -
51	G	0.03	ND -	ND -	ND -	ND -	0.25	0.35	ND -	ND -
52	U	0.09	ND -	ND -	ND -	ND -	0.18	0.16	ND -	ND -
53	G	0.08	ND -	ND -	ND -	ND -	0.30	0.42	ND -	ND -
54	U	0.05	ND -	ND -	ND -	ND -	0.16	0.22	ND -	ND -
55	C	0.12	ND -	ND -	ND -	ND -	0.00	0.00	ND -	ND -
56	U	0.20	ND -	ND -	ND -	ND -	0.16	0.11	ND -	ND -
57	C	0.08	ND -	ND -	ND -	ND -	0.18	0.25	ND -	ND -
58	U	0.12	ND -	ND -	ND -	ND -	0.29	0.40	ND -	ND -
59	U	0.14	ND -	ND -	ND -	ND -	0.00	0.00	ND -	ND -
60	G	0.17	ND -	ND -	ND -	ND -	0.04	0.06	ND -	ND -
61	U	0.38	ND -	ND -	ND -	ND -	0.14	0.06	ND -	ND -
62	U	1.15	ND -	ND -	ND -	ND -	0.34	0.12	ND -	ND -
63	C	0.05	ND -	ND -	ND -	ND -	0.17	0.23	ND -	ND -
64	C	0.08	ND -	ND -	ND -	ND -	0.05	0.07	ND -	ND -
65	C	0.12	ND -	ND -	ND -	ND -	0.06	0.08	ND -	ND -
66	U	0.40	ND -	ND -	ND -	ND -	0.23	0.27	ND -	ND -
67	U	0.70	ND -	ND -	ND -	ND -	0.12	0.01	ND -	ND -
68	C	0.76	ND -	ND -	ND -	ND -	0.19	0.19	ND -	ND -
69	A	0.73	ND -	ND -	ND -	ND -	0.22	0.16	ND -	ND -
70	A	0.81	ND -	ND -	ND -	ND -	0.19	0.05	ND -	ND -
71	U	0.22	ND -	ND -	ND -	ND -	0.15	0.13	ND -	ND -
72	U	0.11	ND -	ND -	ND -	ND -	0.05	0.06	ND -	ND -
73	C	0.05	ND -	ND -	ND -	ND -	0.00	0.00	ND -	ND -

74	C	0.02	ND -	ND -	ND -	ND -	0.00	0.00	ND -	ND -
75	C	0.04	ND -	ND -	ND -	ND -	0.00	0.00	ND -	ND -
76	A	0.14	ND -	ND -	ND -	ND -	0.19	0.24	ND -	ND -
77	C	0.10	ND -	ND -	ND -	ND -	0.28	0.35	ND -	ND -
78	U	0.08	ND -	ND -	ND -	ND -	0.17	0.19	ND -	ND -
79	C	0.03	ND -	ND -	ND -	ND -	0.24	0.33	ND -	ND -
80	C	0.03	ND -	ND -	ND -	ND -	0.01	0.01	ND -	ND -
81	C	0.05	ND -	ND -	ND -	ND -	0.62	0.87	ND -	ND -
82	U	0.06	ND -	ND -	ND -	ND -	0.01	0.01	ND -	ND -
83	C	0.03	ND -	ND -	ND -	ND -	0.09	0.13	ND -	ND -
84	C	0.07	ND -	ND -	ND -	ND -	0.00	0.00	ND -	ND -
85	U	0.12	ND -	ND -	ND -	ND -	0.18	0.25	ND -	ND -
86	C	0.03	ND -	ND -	ND -	ND -	0.03	0.04	ND -	ND -
87	C	0.05	ND -	ND -	ND -	ND -	0.67	0.94	ND -	ND -
88	A	0.21	ND -	ND -	ND -	ND -	0.09	0.06	ND -	ND -
89	G	0.56	ND -	ND -	ND -	ND -	0.29	0.08	ND -	ND -
90	G	0.58	ND -	ND -	ND -	ND -	0.35	0.13	ND -	ND -
91	U	0.96	ND -	ND -	ND -	ND -	0.44	0.23	ND -	ND -
92	U	0.71	ND -	ND -	ND -	ND -	0.62	0.12	ND -	ND -
93	C	0.07	ND -	ND -	ND -	ND -	0.27	0.37	ND -	ND -
94	C	0.08	ND -	ND -	ND -	ND -	0.54	0.76	ND -	ND -
95	U	0.13	ND -	ND -	ND -	ND -	0.16	0.23	ND -	ND -
96	A	0.02	ND -	ND -	ND -	ND -	0.09	0.12	ND -	ND -
97	C	0.39	ND -	ND -	ND -	ND -	0.12	0.17	ND -	ND -
98	U	0.93	ND -	ND -	ND -	ND -	0.32	0.45	ND -	ND -
99	G	0.21	ND -	ND -	ND -	ND -	0.07	0.10	ND -	ND -
100	U	0.22	ND -	ND -	ND -	ND -	0.10	0.13	ND -	ND -

101	U	0.40	ND	-	ND	-	ND	-	ND	-	0.17	0.10	ND	-	ND	-
102	G	0.48	ND	-	ND	-	ND	-	ND	-	0.40	0.01	ND	-	ND	-
103	A	0.59	ND	-	ND	-	ND	-	ND	-	0.90	0.20	ND	-	ND	-
104	U	0.10	ND	-	ND	-	ND	-	ND	-	0.54	0.64	ND	-	ND	-
105	C	0.03	ND	-	ND	-	ND	-	ND	-	0.16	0.22	ND	-	ND	-
106	C	0.01	ND	-	ND	-	ND	-	ND	-	0.16	0.22	ND	-	ND	-
107	C	0.05	ND	-	ND	-	ND	-	0.27	-	0.64	0.91	ND	-	ND	-
108	G	0.10	ND	-	ND	-	ND	-	0.86	-	0.04	0.06	ND	-	ND	-
109	C	0.46	ND	-	ND	-	ND	-	0.39	-	0.06	0.08	ND	-	ND	-
110	G	1.47	ND	-	ND	-	ND	-	0.31	-	0.59	0.74	ND	-	ND	-
111	G	0.61	ND	-	ND	-	ND	-	0.19	-	0.60	0.85	ND	-	ND	-
112	G	0.37	ND	-	ND	-	ND	-	0.31	-	0.18	0.25	ND	-	ND	-
113	U	1.06	ND	-	ND	-	1.20	0.64	1.18	-	0.59	0.83	ND	-	ND	-
114	C	0.10	ND	-	ND	-	0.05	0.01	0.3	-	0.07	0.09	ND	-	ND	-
115	G	0.06	ND	-	ND	-	0.00	0.00	0.14	-	2.09	2.75	ND	-	ND	-
116	G	0.04	ND	-	ND	-	0.00	0.00	0	-	0.59	0.83	ND	-	ND	-
117	G	0.12	ND	-	ND	-	0.00	0.00	0	-	0.39	0.54	ND	-	ND	-
118	A	0.11	ND	-	ND	-	0.00	0.00	0.31	-	0.07	0.09	ND	-	ND	-
119	C	0.24	0.02	0.03	ND	-	0.00	0.00	5.29	-	0.77	1.08	ND	-	ND	-
120	A	0.99	0.32	0.29	ND	-	0.54	0.12	0.83	-	0.75	0.14	ND	-	ND	-
121	G	0.47	0.35	0.34	ND	-	0.55	0.25	0.32	-	0.30	0.23	ND	-	ND	-
122	U	0.87	0.45	0.38	ND	-	0.78	0.26	0.85	-	0.53	0.28	ND	-	ND	-
123	U	0.58	0.33	0.31	ND	-	0.62	0.17	1.29	-	0.39	0.23	ND	-	ND	-
124	G	0.08	0.07	0.05	ND	-	0.16	0.12	0.49	-	0.05	0.07	ND	-	ND	-
125	G	0.10	0.02	0.02	ND	-	0.06	0.10	0.44	-	0.02	0.01	ND	-	ND	-
126	C	0.10	0.02	0.03	ND	-	0.07	0.02	0.99	-	0.05	0.06	ND	-	ND	-
127	G	0.01	0.04	0.03	ND	-	0.14	0.03	0.63	-	0.19	0.04	ND	-	ND	-

128	C	0.02	0.06	0.09	ND	-	0.03	0.06	0.17	-	0.00	0.00	ND	-	ND	-
129	C	0.17	0.15	0.21	ND	-	0.06	0.10	0	-	0.00	0.00	ND	-	ND	-
130	C	0.05	0.03	0.04	ND	-	0.00	0.00	1.67	-	0.00	0.00	ND	-	ND	-
131	A	0.10	0.04	0.06	ND	-	0.03	0.06	0.5	-	0.28	0.28	ND	-	ND	-
132	A	0.21	0.09	0.10	ND	-	0.16	0.02	0.45	-	0.33	0.32	ND	-	ND	-
133	C	0.03	0.02	0.03	ND	-	0.00	0.00	0.42	-	0.08	0.11	ND	-	ND	-
134	G	0.05	0.03	0.02	ND	-	0.03	0.02	0.22	-	0.08	0.09	ND	-	ND	-
135	U	0.10	0.06	0.06	ND	-	0.07	0.09	0.43	-	0.26	0.13	ND	-	ND	-
136	G	0.06	0.02	0.03	ND	-	0.02	0.03	0.22	-	0.00	0.00	ND	-	ND	-
137	G	0.01	0.05	0.07	ND	-	0.03	0.05	0	-	0.00	0.00	ND	-	ND	-
138	G	0.08	0.06	0.08	ND	-	0.00	0.00	0	-	0.00	0.00	ND	-	ND	-
139	G	0.10	0.14	0.20	ND	-	0.06	0.11	0	-	0.00	0.00	ND	-	ND	-
140	C	0.05	0.07	0.09	ND	-	0.00	0.00	0.23	-	0.08	0.11	ND	-	ND	-
141	U	0.07	0.06	0.08	ND	-	0.01	0.01	0.29	-	0.08	0.11	ND	-	ND	-
142	G	0.05	0.03	0.02	ND	-	0.03	0.04	0	-	0.11	0.16	ND	-	ND	-
143	G	0.06	0.06	0.05	ND	-	0.00	0.00	0.43	-	0.08	0.11	ND	-	ND	-
144	A	0.08	0.08	0.11	ND	-	0.01	0.02	0.65	-	0.18	0.25	ND	-	ND	-
145	U	0.17	0.19	0.26	ND	-	0.08	0.13	2.53	-	0.07	0.09	ND	-	ND	-
146	A	0.67	0.79	0.34	ND	-	0.93	0.25	0.49	-	2.52	2.07	ND	-	ND	-
147	C	0.05	0.18	0.25	ND	-	0.11	0.16	0.26	-	0.60	0.28	ND	-	ND	-
148	G	0.07	0.01	0.01	ND	-	0.02	0.04	0.01	-	0.15	0.04	ND	-	ND	-
149	A	0.00	0	0.00	ND	-	0.00	0.00	0	-	0.10	0.14	ND	-	ND	-
150	G	0.09	0	0.00	ND	-	0.00	0.00	0	-	0.56	0.78	ND	-	ND	-
151	G	ND	0.17	0.24	ND	-	0.00	0.00	0	-	2.70	3.81	ND	-	ND	-
152	G	ND	0	0.00	ND	-	0.22	0.38	0	-	0.63	0.89	ND	-	ND	-
153	A	ND	0.1	0.10	ND	-	0.00	0.00	0	-	3.37	4.76	ND	-	ND	-
154	A	ND	0.68	0.27	ND	-	0.47	0.10	0	-	0.57	0.81	ND	-	ND	-

155	U	ND	1.24	0.41	ND	-	1.44	0.12	0.75	-	1.15	0.60	ND	-	ND	-
156	U	0.83	1.78	0.64	ND	-	2.28	0.41	0.67	-	1.21	0.82	ND	-	ND	-
157	U	0.84	1.5	0.71	ND	-	1.97	0.15	1.15	-	0.45	0.35	ND	-	ND	-
158	C	0.02	0.36	0.27	ND	-	0.30	0.15	0.73	-	0.06	0.08	ND	-	ND	-
159	G	0.10	0.09	0.06	ND	-	0.02	0.04	0.18	-	0.06	0.08	ND	-	ND	-
160	U	0.10	0.18	0.02	ND	-	0.07	0.08	0.23	-	0.06	0.08	ND	-	0.18	0.03
161	G	0.31	0.23	0.05	ND	-	0.30	0.24	0	-	0.16	0.22	ND	-	0.23	0.07
162	A	1.64	2	0.81	ND	-	2.74	0.79	0.62	-	1.15	1.36	0.62	1.03	2.00	1.00
163	G	0.10	0.15	0.02	ND	-	0.09	0.14	0.16	-	0.65	0.78	0.06	0.06	0.15	0.03
164	G	0.05	0.03	0.00	0.09	0.08	0.02	0.03	0	-	0.04	0.06	0.63	1.06	0.03	0.01
165	A	0.11	0.03	0.02	2.20	0.40	0.06	0.05	0	-	0.13	0.18	0.44	0.72	0.03	0.03
166	A	0.16	0.24	0.08	0.10	0.07	0.22	0.09	0.1	-	0.19	0.26	0.09	0.16	0.24	0.10
167	G	0.11	0.34	0.09	0.01	0.01	0.34	0.08	0.27	-	0.40	0.15	0.12	0.21	0.34	0.11
168	A	0.39	0.57	0.12	0.03	0.05	0.56	0.04	0.4	-	0.60	0.16	0.28	0.31	0.57	0.15
169	C	0.05	0.14	0.13	0.13	0.03	0.07	0.08	1.72	0.98	0.19	0.04	0.16	0.15	0.14	0.16
170	G	0.05	0.08	0.09	0.10	0.05	0.09	0.04	0	0.26	0.27	0.11	0.29	0.29	0.08	0.11
171	A	0.19	0.17	0.03	0.51	0.04	0.17	0.04	0.08	0.26	0.36	0.18	0.17	0.24	0.17	0.04
172	C	0.10	0.39	0.28	0.08	0.11	0.26	0.04	1.81	0.57	0.23	0.32	0.28	0.26	0.39	0.34
173	G	0.10	0.19	0.05	0.00	0.01	0.47	0.10	0.17	0.14	0.33	0.35	0.05	0.09	0.19	0.06
174	C	0.00	0.03	0.04	0.18	0.06	0.03	0.04	0.11	0.06	0.30	0.40	0.06	0.10	0.03	0.04
175	G	0.03	0.08	0.10	0.28	0.18	0.07	0.06	0.24	0.15	0.11	0.10	0.25	0.34	0.08	0.13
176	U	0.05	0	0.00	0.33	0.14	0.00	0.01	0.01	0.19	0.22	0.31	0.17	0.24	0.00	0.01
177	U	0.11	0.31	0.05	0.16	0.05	0.39	0.05	0.15	0.19	0.08	0.11	0.33	0.15	0.31	0.06
178	C	0.12	0.62	0.17	0.20	0.12	0.56	0.10	0.25	0.15	0.44	0.16	0.1	0.12	0.62	0.21
179	G	0.06	0.48	0.09	0.09	0.05	0.13	0.02	0	0.03	0.54	0.25	0.23	0.17	0.48	0.11
180	C	0.00	0	0.00	0.00	0.00	0.01	0.02	0.08	0.06	0.00	0.00	0.13	0.22	0.00	0.00
181	C	0.03	0.04	0.03	0.06	0.05	0.10	0.11	0.22	0.11	0.02	0.03	0.18	0.31	0.04	0.04

182	G	0.33	0.42	0.18	1.93	0.08	1.05	0.21	1.38	0.64	0.63	0.14	0.28	0.27	0.42	0.22
183	G	0.43	1.45	0.28	0.98	0.07	2.54	0.37	0.5	0.84	1.74	0.64	0.48	0.35	1.45	0.34
184	C	0.22	0.92	0.34	0.64	0.18	0.98	0.12	0.82	0.38	1.06	0.40	0.2	0.22	0.92	0.42
185	C	0.18	0.98	0.30	1.11	0.07	1.46	0.35	0.75	0.65	1.21	0.40	0.35	0.21	0.98	0.37
186	G	0.05	0.12	0.08	0.00	0.00	0.10	0.06	0	0.06	0.50	0.07	0.53	0.71	0.12	0.10
187	G	0.21	0.34	0.24	0.00	0.00	0.33	0.36	0.05	0.12	2.98	1.89	0.66	0.31	0.34	0.30
188	C	-1.00	0.85	0.60	0.00	0.00	0.39	0.49	0	0.13	1.33	1.03	1.38	0.83	0.85	0.74
189	G	0.00	0.35	0.22	0.00	0.00	0.22	0.38	0	0.06	0.31	0.43	0.3	0.24	0.35	0.27
190	A	0.04	0.54	0.26	0.00	0.00	0.10	0.12	0	0.17	0.36	0.34	0.57	0.34	0.54	0.32
191	U	0.08	0.61	0.10	Δ	Δ	0.22	0.18	Δ	Δ	0.06	0.08	0.57	0.10	0.61	0.13
192	U	0.18	2	1.51	Δ	Δ	1.24	0.36	Δ	Δ	0.00	0.00	1.01	0.91	2.00	1.85
193	A	0.11	1.26	0.06	Δ	Δ	1.53	0.40	Δ	Δ	0.36	0.01	0.57	0.31	1.26	0.07
194	A	0.10	***Δ	Δ	Δ	Δ	1.57	0.40	Δ	Δ	0.38	0.01	0.33	0.19	0.85	0.13
195	A	0.11	Δ	Δ	Δ	Δ	1.20	0.29	Δ	Δ	0.34	0.20	0.07	0.04	0.61	0.04
196	A	0.11	Δ	Δ	Δ	Δ	0.36	0.05	Δ	Δ	0.48	0.07	0.1	0.08	0.55	0.04
197	G	0.33	Δ	Δ	Δ	Δ	0.52	0.91	Δ	Δ	0.52	0.06	0.04	0.05	0.76	0.09
198	U	0.39	Δ	Δ	Δ	Δ	0.14	0.24	Δ	Δ	0.28	0.39	0.23	0.33	1.29	0.25
199	G	0.51	0.85	0.11	Δ	Δ	0.05	0.09	Δ	Δ	Δ	Δ	0.18	0.29	2.00	2.40
200	A	0.54	0.61	0.03	Δ	Δ	0.53	0.08	Δ	Δ	Δ	Δ	0.54	0.24	1.20	0.06
201	A	0.56	0.55	0.03	Δ	Δ	1.02	0.16	Δ	Δ	Δ	Δ	0.71	0.48	1.06	0.12
202	A	0.46	0.76	0.07	Δ	Δ	0.99	0.52	Δ	Δ	Δ	Δ	0.9	0.44	0.32	0.04
203	G	0.58	1.29	0.20	Δ	Δ	1.29	2.23	Δ	Δ	Δ	Δ	0.93	0.28	0.07	0.03
204	U	0.82	2	1.96	Δ	Δ	1.24	0.34	Δ	Δ	Δ	Δ	0.47	0.42	0.14	0.07

205	A	0.67	1.2	0.05	Δ	Δ	0.91	0.21	Δ	Δ	Δ	Δ	0.72	0.15	0.09	0.05
206	A	0.59	1.06	0.09	Δ	Δ	0.41	0.17	Δ	Δ	Δ	Δ	0.52	0.36	0.21	0.08
207	A	0.15	0.32	0.03	0.00	0.00	0.19	0.03	0	0.15	0.39	0.16	0.42	0.29	Δ	Δ
208	C	0.01	0.07	0.02	0.00	0.00	0.01	0.01	0	0.27	0.11	0.06	0.11	0.10	Δ	Δ
209	U	0.04	0.14	0.06	0.11	0.07	0.10	0.06	1.04	0.22	0.32	0.06	0.14	0.05	Δ	Δ
210	C	0.01	0.09	0.04	0.14	0.09	0.04	0.02	1.92	0.66	0.22	0.17	0.13	0.19	Δ	Δ
211	U	0.03	0.21	0.06	0.18	0.04	0.16	0.16	1.75	0.62	0.19	0.04	0.13	0.08	Δ	Δ
212	C	0.08	0.05	0.01	0.08	0.08	0.02	0.02	1.96	0.90	0.24	0.08	0.15	0.23	Δ	Δ
213	U	0.20	0.27	0.02	0.34	0.05	0.35	0.03	0.68	0.28	0.73	0.20	0.08	0.09	0.05	0.02
214	U	0.22	0.41	0.06	0.44	0.05	0.49	0.05	0.52	0.16	0.78	0.13	0.34	0.17	0.27	0.03
215	G	0.15	0.33	0.07	0.17	0.03	0.29	0.24	0.33	0.06	0.47	0.06	0.58	0.19	0.41	0.07
216	G	0.04	0.22	0.07	0.09	0.03	0.44	0.23	0	0.02	0.08	0.04	0.26	0.45	0.33	0.09
217	C	0.05	0.06	0.04	0.01	0.01	0.01	0.02	0.61	0.33	0.03	0.04	0.02	0.03	0.22	0.08
218	C	0.03	0.02	0.03	0.01	0.02	0.03	0.02	0.12	0.06	0.00	0.00	0.06	0.08	0.06	0.06
219	G	0.04	0.07	0.02	0.02	0.01	0.09	0.04	0	0.11	0.16	0.19	0.02	0.02	0.02	0.04
220	C	0.08	0.11	0.02	0.01	0.01	0.08	0.06	0.43	0.20	0.09	0.00	0.03	0.05	0.07	0.03
221	C	0.00	0.01	0.01	0.00	0.00	0.05	0.05	0.05	0.04	0.03	0.04	0.03	0.03	0.11	0.03
222	G	0.01	0.03	0.02	0.00	0.00	0.06	0.06	0.02	0.08	0.11	0.06	0.02	0.03	0.01	0.01
223	C	0.02	0.02	0.02	0.03	0.06	0.02	0.03	0	0.07	0.05	0.07	0.01	0.02	0.03	0.03
224	G	0.02	0	0.00	0.00	0.01	0.02	0.02	0	0.02	0.14	0.19	0	0.00	0.02	0.03
225	G	0.22	0.03	0.05	0.02	0.03	0.10	0.10	0	0.00	0.22	0.18	0	0.01	0.00	0.00
226	G	0.68	0.84	0.34	0.72	0.39	1.32	0.22	0	0.27	0.96	0.27	0.5	0.45	0.03	0.06
227	A	1.49	2	0.38	2.14	0.29	3.30	0.83	0.66	0.42	1.78	0.36	1.43	0.79	0.84	0.41
228	A	0.87	1.93	0.39	2.09	0.18	2.54	0.67	0.6	0.43	1.58	0.33	1.78	0.27	2.00	0.47
229	C	0.28	0.59	0.20	0.47	0.46	0.81	0.40	0.16	0.20	0.18	0.25	0.74	0.79	1.93	0.48

230	C	0.56	0.46	0.35	0.69	0.15	0.66	0.34	0.23	0.16	0.08	0.11	0.35	0.13	0.59	0.24
231	U	1.08	1.72	0.50	2.06	0.13	2.22	0.66	0.61	0.54	1.16	0.15	1.22	0.70	0.46	0.43
232	G	0.42	1.75	0.36	2.02	0.17	2.33	0.58	0.27	0.61	1.00	0.41	1.23	0.54	1.72	0.62
233	C	0.12	0.05	0.04	0.04	0.04	0.02	0.03	0.7	0.39	0.09	0.12	0.32	0.56	1.75	0.44
234	C	0.01	0	0.00	0.00	0.00	0.00	0.00	0.09	0.05	0.01	0.01	0.03	0.05	0.05	0.05
235	G	0.01	0.02	0.02	0.00	0.00	0.03	0.04	0	0.08	0.17	0.18	0.03	0.06	0.00	0.00
236	C	0.01	0	0.00	0.00	0.00	0.00	0.00	0.51	0.27	0.12	0.12	0.22	0.22	0.02	0.03
237	G	0.08	0.04	0.04	0.04	0.08	0.03	0.02	0.17	0.11	0.34	0.04	0.5	0.46	0.00	0.00
238	U	0.13	0.24	0.07	0.22	0.10	0.26	0.08	0.23	0.10	0.42	0.12	0.48	0.42	0.04	0.05
239	U	0.09	0.19	0.00	0.12	0.04	0.19	0.02	0.39	0.05	0.54	0.13	0.68	0.60	0.24	0.09
240	G	0.04	0.1	0.02	0.02	0.02	0.08	0.06	0	0.10	0.38	0.05	0.36	0.20	0.19	0.01
241	G	0.16	0.27	0.03	0.19	0.02	0.19	0.07	0.32	0.05	0.58	0.07	0.61	0.32	0.10	0.03
242	A	0.68	2	0.20	2.01	1.51	2.53	0.63	2.21	1.34	4.04	0.45	2.48	1.46	0.27	0.04
243	C	0.04	0.09	0.03	0.56	0.96	0.03	0.01	0.94	0.44	0.15	0.18	0.7	0.58	2.00	0.24
244	C	0.05	0.04	0.04	0.02	0.04	0.02	0.02	0.06	0.03	0.00	0.00	0.22	0.32	0.09	0.04
245	U	0.11	0.13	0.06	0.11	0.08	0.12	0.06	0.17	0.01	0.31	0.04	0.3	0.30	0.04	0.05
246	G	0.04	0.17	0.04	0.20	0.08	0.07	0.05	Δ	Δ	Δ	Δ	Δ	Δ	0.13	0.07
247	A	0.07	0.18	0.03	0.34	0.09	0.18	0.06	Δ	Δ	Δ	Δ	Δ	Δ	0.17	0.06
248	A	0.09	0.17	0.02	0.28	0.06	0.28	0.09	Δ	Δ	Δ	Δ	Δ	Δ	0.18	0.04
249	A	0.03	0.11	0.07	0.18	0.05	0.38	0.13	Δ	Δ	Δ	Δ	Δ	Δ	0.17	0.03
250	G	0.17	0.28	0.19	0.24	0.03	0.11	0.04	Δ	Δ	Δ	Δ	Δ	Δ	0.11	0.08
251	U	0.38	2	3.21	0.63	0.31	0.14	0.25	Δ	Δ	Δ	Δ	Δ	Δ	0.28	0.23
252	A	0.60	0.59	0.11	0.95	0.10	0.18	0.07	Δ	Δ	Δ	Δ	Δ	Δ	2.00	3.93
253	A	0.32	0.45	0.04	0.68	0.06	0.29	0.10	Δ	Δ	Δ	Δ	Δ	Δ	0.59	0.14
254	G	0.48	0.51	0.03	0.51	0.14	0.24	0.16	0.19	0.17	0.41	0.11	1.14	1.78	0.45	0.06

255	U	0.53	0.9	0.36	0.83	0.22	0.17	0.10	1.65	0.46	0.77	0.48	0.31	0.23	0.51	0.04
256	G	0.39	0.64	0.05	0.46	0.09	0.16	0.16	0.29	0.18	0.45	0.06	0.12	0.11	0.90	0.44
257	U	0.44	0.6	0.06	0.90	0.29	0.32	0.05	0.61	0.27	0.67	0.07	0.56	0.49	0.64	0.07
258	U	0.22	0.52	0.01	0.64	0.08	0.47	0.01	0.49	0.19	0.44	0.11	0.29	0.15	0.60	0.07
259	G	0.01	0.08	0.02	0.08	0.02	0.11	0.14	0.05	0.05	0.14	0.13	0.12	0.07	0.52	0.01
260	C	0.06	0.04	0.04	0.03	0.05	0.04	0.07	0.87	0.34	0.14	0.08	0.34	0.30	0.08	0.02
261	G	0.15	0.21	0.05	0.38	0.11	0.50	0.22	0.25	0.21	0.21	0.04	0.16	0.11	0.04	0.06
262	C	0.13	0.1	0.08	0.08	0.11	0.17	0.17	0.37	0.11	0.09	0.12	0.38	0.65	0.21	0.07
263	U	0.17	0.21	0.03	0.17	0.12	0.14	0.14	0.22	0.16	0.39	0.01	0.19	0.20	0.10	0.10
264	C	0.11	0.17	0.24	0.10	0.12	0.12	0.18	1.57	0.67	0.24	0.05	0.7	0.93	0.21	0.04
265	G	0.35	0.22	0.09	0.44	0.07	0.42	0.11	0.17	0.19	0.45	0.03	0.43	0.37	0.17	0.29
266	G	0.42	0.55	0.06	0.41	0.12	0.69	0.13	0.26	0.13	0.80	0.12	0.45	0.30	0.22	0.11
267	A	0.82	1.05	0.07	1.52	0.07	1.23	0.18	2.03	0.51	0.63	0.22	0.88	0.62	0.55	0.07
268	U	0.66	2	1.20	1.11	0.45	0.99	0.59	5.68	2.86	1.48	2.09	1.9	2.45	1.05	0.09
269	A	0.41	0.57	0.19	0.84	0.22	0.73	0.26	0.29	0.12	0.59	0.11	0.42	0.30	2.00	1.47
270	U	0.40	0.55	0.15	0.77	0.18	0.68	0.28	0.24	0.14	0.74	0.03	0.37	0.23	0.57	0.24
271	G	0.04	0.03	0.02	0.02	0.02	0.05	0.09	0.00	0.03	0.16	0.07	0.16	0.11	0.55	0.19
272	G	0.02	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.55	0.03	0.03
273	G	0.02	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.44	0.00	0.00
274	G	0.01	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.91	0.00	0.00
275	C	0.06	0	0.00	0.00	0.00	0.00	0.00	1.37	0.79	0.00	0.00	0.53	0.88	0.00	0.00
276	A	0.13	0.02	0.02	0.01	0.01	0.08	0.07	0.00	0.06	0.21	0.29	0.26	0.45	0.00	0.00
277	A	0.16	0.08	0.06	0.08	0.03	0.13	0.10	0.00	0.07	0.38	0.28	0.23	0.34	0.02	0.03
278	G	0.18	0.17	0.09	0.11	0.04	0.19	0.23	0.00	0.09	0.64	0.31	0.15	0.14	0.08	0.07
279	A	0.40	0.43	0.08	0.55	0.07	0.52	0.26	0.17	0.14	0.90	0.54	0.29	0.33	0.17	0.11
280	A	0.32	0.6	0.06	0.72	0.06	0.67	0.18	0.34	0.14	0.81	0.34	0.34	0.32	0.43	0.09
281	U	0.49	0.82	0.10	1.14	0.13	0.90	0.11	0.70	0.07	0.93	0.38	0.47	0.42	0.60	0.08

282	U	0.20	0.36	0.20	0.56	0.15	0.20	0.12	1.86	0.94	0.08	0.11	0.32	0.30	0.82	0.12
283	A	0.09	0.05	0.04	0.08	0.03	0.07	0.07	0.00	0.02	0.05	0.06	0	0.00	0.36	0.25
284	A	0.03	0.03	0.01	0.02	0.02	0.04	0.05	0.03	0.02	0.05	0.01	0.08	0.12	0.05	0.05
285	G	0.01	0.01	0.00	0.03	0.04	0.02	0.03	0.06	0.03	0.03	0.04	0.1	0.18	0.03	0.02
286	C	0.09	0.09	0.07	0.06	0.04	0.10	0.08	0.51	0.28	0.01	0.01	0.14	0.16	0.01	0.01
287	C	0.09	0.04	0.06	0.07	0.06	0.05	0.09	0.62	0.32	0.03	0.04	0.35	0.32	0.09	0.08
288	A	0.08	0.05	0.04	0.07	0.07	0.12	0.06	0.04	0.04	0.07	0.05	0.25	0.31	0.04	0.07
289	G	0.02	0.01	0.01	0.05	0.08	0.00	0.00	0.87	0.48	0.00	0.00	0.22	0.38	0.05	0.05
290	C	0.13	0.41	0.29	0.37	0.48	0.17	0.30	3.56	1.90	0.90	1.27	0.73	0.90	0.01	0.02
291	A	0.71	0.8	0.15	0.62	0.44	0.94	0.34	0.10	0.46	0.87	0.47	0.39	0.38	0.41	0.36
292	U	0.48	1.11	0.38	0.76	0.34	1.07	0.10	0.14	0.46	1.33	0.40	0.63	0.53	0.80	0.18
293	G	0.34	0.64	0.15	0.60	0.36	0.65	0.13	0.09	0.26	0.75	0.11	0.24	0.20	1.11	0.46
294	A	0.21	0.47	0.05	0.38	0.15	0.42	0.08	0.12	0.18	0.49	0.23	0.29	0.01	0.64	0.19
295	A	0.19	0.37	0.06	0.34	0.03	0.30	0.04	0.18	0.08	0.37	0.07	0.21	0.05	0.47	0.07
296	C	0.14	0.27	0.08	0.29	0.12	0.25	0.01	0.32	0.05	0.31	0.04	0.17	0.13	0.37	0.07
297	G	0.17	0.15	0.07	0.25	0.16	0.11	0.05	0.06	0.06	0.18	0.12	0.07	0.06	0.27	0.10
298	U	0.26	0.33	0.10	0.25	0.08	0.27	0.06	0.40	0.20	0.00	0.00	0.13	0.13	0.15	0.08
299	U	0.34	1.01	1.04	0.64	0.71	0.24	0.06	2.71	1.52	0.62	0.88	0.12	0.08	0.33	0.12
300	A	1.04	1.09	0.07	1.11	0.75	1.19	0.37	0.72	0.13	1.31	0.32	0.9	0.71	1.01	1.28
301	U	0.41	0.77	0.14	1.12	0.72	0.86	0.33	0.29	0.12	0.67	0.21	0.61	0.16	1.09	0.09
302	G	0.79	1.45	0.07	1.37	0.65	1.45	0.54	1.03	0.11	1.20	0.03	0.91	0.69	0.77	0.18
303	U	1.29	1.87	1.14	1.68	0.22	0.98	0.34	3.15	1.60	0.53	0.75	1.13	0.63	1.45	0.08
304	A	1.80	1.39	0.19	1.82	0.19	1.58	0.52	0.67	0.47	1.47	0.34	1.46	1.04	1.87	1.40
305	G	0.66	1.71	0.43	1.36	0.72	1.89	0.57	0.56	0.61	1.69	0.71	1.35	0.88	1.39	0.23
306	A	0.42	0.37	0.05	0.57	0.31	0.48	0.05	0.22	0.16	0.44	0.27	0.3	0.16	1.71	0.53
307	A	0.36	0.12	0.09	0.16	0.06	0.13	0.12	0.50	0.15	0.23	0.33	0.05	0.08	0.37	0.06
308	C	0.33	0.38	0.50	0.43	0.22	0.10	0.17	6.59	3.64	0.00	0.00	0.36	0.53	0.12	0.12

309	A	0.62	0.26	0.05	0.41	0.19	0.29	0.11	0.24	0.10	0.34	0.27	0.34	0.09	0.38	0.61
310	A	0.20	0.55	0.09	0.48	0.40	0.56	0.05	0.15	0.24	0.60	0.25	0.46	0.08	0.26	0.06
311	U	0.32	0.66	0.06	0.79	0.36	0.70	0.10	0.21	0.30	0.69	0.23	1.05	0.71	0.55	0.11
312	U	0.79	0.62	0.08	0.88	0.07	0.66	0.06	0.29	0.23	0.74	0.25	0.69	0.18	0.66	0.08
313	G	0.59	0.47	0.05	0.95	0.26	0.34	0.10	0.25	0.17	0.58	0.30	0.8	0.90	0.62	0.10
314	A	0.72	0.51	0.07	0.80	0.21	0.55	0.09	0.25	0.23	0.62	0.37	1.13	0.77	0.47	0.06
315	A	0.23	0.54	0.13	0.59	0.42	0.61	0.05	0.29	0.15	0.61	0.23	1.31	1.23	0.51	0.09
316	G	0.01	0.01	0.01	0.24	0.31	0.00	0.00	0.35	0.20	0.01	0.01	0.59	1.00	0.54	0.16
317	C	0.15	0.25	0.36	0.22	0.15	0.13	0.23	4.64	2.46	0.29	0.40	0.97	1.31	0.01	0.02
318	A	0.28	0.22	0.03	0.11	0.16	0.20	0.11	0.10	0.14	0.32	0.45	0.07	0.08	0.25	0.44
319	G	0.01	0	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.16	0.22	0.04
320	G	0.02	0	0.00	0.11	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.31	0.00	0.00
321	C	0.25	0.01	0.01	0.31	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.31	0.00	0.00
322	U	0.42	0.37	0.11	0.26	0.23	0.38	0.21	0.17	0.05	0.22	0.30	0.37	0.44	0.01	0.02
323	U	0.40	0.54	0.13	0.49	0.69	0.57	0.10	0.74	0.22	0.44	0.04	0.53	0.43	0.37	0.13
324	U	0.73	1.35	0.92	0.81	0.69	0.25	0.25	4.45	2.36	0.30	0.42	0.4	0.27	0.54	0.16
325	A	0.89	0.98	0.14	1.14	0.44	0.88	0.03	0.87	0.07	0.91	0.16	0.92	0.20	1.35	1.12
326	A	0.85	0.69	0.12	0.95	0.20	0.64	0.04	0.50	0.19	0.68	0.29	0.71	0.16	0.98	0.17
327	A	0.46	0.55	0.09	0.66	0.42	0.55	0.05	0.38	0.12	0.62	0.30	0.47	0.42	0.69	0.14
328	G	0.08	0.03	0.04	0.26	0.31	0.00	0.01	0.04	0.04	0.03	0.04	0.06	0.07	0.55	0.11
329	A	0.13	0.03	0.04	0.05	0.05	0.02	0.03	0.63	0.33	0.06	0.08	0.11	0.19	0.03	0.05
330	C	0.07	0.08	0.11	0.16	0.15	0.04	0.07	6.13	3.51	0.02	0.03	0.67	1.16	0.03	0.05
331	A	0.09	0.01	0.01	0.03	0.02	0.07	0.04	0.01	0.01	0.01	0.01	0.02	0.03	0.08	0.14
332	C	0.01	0	0.00	0.01	0.01	0.02	0.02	0.05	0.03	0.01	0.01	0.02	0.03	0.01	0.01
333	G	0.00	0	0.00	0.05	0.03	0.04	0.04	0.00	0.00	0.01	0.01	0.17	0.27	0.00	0.01
334	G	0.03	0.03	0.04	0.05	0.08	0.05	0.09	0.00	0.00	0.04	0.05	0.23	0.30	0.00	0.00
335	G	0.01	0	0.00	0.07	0.03	0.03	0.05	0.00	0.00	0.01	0.01	0.22	0.38	0.03	0.04

336	G	0.11	0.02	0.03	0.07	0.08	0.06	0.10	0.07	0.04	0.08	0.11	0.34	0.54	0.00	0.01
337	A	0.08	0.14	0.04	0.09	0.08	0.18	0.14	0.39	0.13	0.42	0.23	0.37	0.32	0.02	0.03
338	G	0.10	0.17	0.03	0.18	0.16	0.16	0.17	0.74	0.38	0.19	0.27	0.38	0.35	0.14	0.05
339	U	0.34	0.26	0.37	0.32	0.33	0.20	0.34	6.23	3.45	0.88	1.24	0.68	0.85	0.17	0.04
340	A	0.20	0.39	0.07	0.67	0.53	0.40	0.18	0.29	0.04	0.57	0.17	0.26	0.23	0.26	0.45
341	A	0.12	0.29	0.07	0.34	0.28	0.28	0.15	0.09	0.08	0.40	0.09	0.15	0.13	0.39	0.09
342	A	0.17	0.35	0.13	0.23	0.13	0.23	0.20	0.07	0.12	0.37	0.21	0.13	0.10	0.29	0.08
343	G	0.05	0.07	0.07	0.26	0.23	0.00	0.00	0.00	0.00	0.04	0.05	0.02	0.03	0.35	0.16
344	G	0.08	0.1	0.07	0.13	0.16	0.00	0.00	0.08	0.05	0.00	0.00	0.03	0.03	0.07	0.09
345	U	0.34	0.39	0.20	0.21	0.26	0.39	0.29	0.82	0.39	0.09	0.12	0.23	0.20	0.10	0.08
346	U	0.53	0.96	0.31	0.44	0.29	0.39	0.31	6.02	3.32	0.77	1.08	0.61	0.56	0.39	0.25
347	A	0.48	0.83	0.19	0.74	0.27	0.71	0.09	0.96	0.26	0.92	0.20	0.37	0.26	0.96	0.38
348	A	0.29	0.8	0.19	0.54	0.15	0.68	0.08	0.85	0.23	0.82	0.23	0.3	0.26	0.83	0.24
349	A	0.38	0.85	0.23	0.57	0.14	0.70	0.09	1.28	0.53	0.59	0.83	0.3	0.28	0.80	0.24
350	U	0.43	1.63	1.51	0.59	0.18	0.37	0.33	4.84	2.67	1.97	2.79	0.43	0.33	0.85	0.28
351	A	0.31	0.68	0.06	0.66	0.21	0.74	0.11	0.26	0.12	0.48	0.18	0.27	0.15	1.63	1.85
352	U	0.27	0.71	0.04	0.48	0.30	0.84	0.18	0.25	0.16	0.61	0.08	0.51	0.05	0.68	0.08
353	G	0.24	0.62	0.11	0.80	0.35	0.70	0.24	0.15	0.28	0.48	0.08	0.56	0.19	0.71	0.05
354	C	0.19	0.1	0.14	0.51	0.25	0.05	0.09	0.48	0.16	0.00	0.00	0.13	0.08	0.62	0.14
355	U	0.33	0.22	0.16	0.16	0.18	0.19	0.16	0.13	0.10	0.10	0.14	0.1	0.07	0.10	0.17
356	G	0.26	0.33	0.03	0.34	0.26	0.24	0.12	0.02	0.20	0.21	0.14	0.07	0.06	0.22	0.19
357	A	0.39	0.47	0.28	0.13	0.17	0.60	0.39	0.02	0.48	0.45	0.45	0.18	0.16	0.33	0.04
358	U	0.03	0.1	0.09	0.07	0.08	0.08	0.07	0.22	0.10	0.05	0.07	0.1	0.16	0.47	0.34
359	C	0.03	0	0.00	0.15	0.26	0.00	0.00	0.17	0.10	0.00	0.00	0.28	0.49	0.10	0.11
360	U	0.32	0.25	0.19	0.11	0.10	0.24	0.21	0.00	0.10	0.04	0.05	0.23	0.24	0.00	0.00
361	U	0.27	0.3	0.08	0.13	0.14	0.35	0.15	0.03	0.17	0.21	0.04	0.23	0.11	0.25	0.23
362	U	0.29	0.39	0.05	0.30	0.11	0.38	0.08	0.10	0.17	0.35	0.06	0.25	0.06	0.30	0.10

363	U	0.33	0.26	0.15	0.49	0.18	0.22	0.19	0.17	0.27	0.17	0.23	0.29	0.03	0.39	0.06
364	G	0.35	0.33	0.12	0.71	0.33	0.14	0.23	0.04	0.29	0.15	0.21	0.24	0.05	0.26	0.19
365	A	0.54	0.47	0.12	0.44	0.29	0.34	0.19	0.03	0.26	0.22	0.25	0.37	0.06	0.33	0.14
366	A	0.13	0.11	0.08	0.29	0.23	0.11	0.10	0.00	0.10	0.02	0.03	0.24	0.26	0.47	0.15
367	A	0.04	0.04	0.05	0.12	0.08	0.02	0.02	0.00	0.03	0.00	0.00	0.01	0.00	0.11	0.10
368	U	0.11	0.09	0.07	0.13	0.21	0.04	0.04	0.00	0.08	0.05	0.07	0.1	0.04	0.04	0.06
369	U	0.15	0.25	0.21	0.26	0.24	0.06	0.10	0.00	0.24	0.19	0.27	0.32	0.27	0.09	0.09
370	U	0.07	0.13	0.09	0.28	0.30	0.26	0.06	0.08	0.16	0.12	0.16	0.26	0.17	0.25	0.26
371	U	0.05	0.2	0.03	0.24	0.11	0.35	0.23	0.07	0.13	0.11	0.16	0.31	0.04	0.13	0.11
372	U	0.15	0.12	0.10	0.36	0.23	0.20	0.15	0.05	0.10	0.11	0.16	0.2	0.12	0.20	0.03
373	U	0.23	0.11	0.08	0.44	0.20	0.23	0.22	0.34	0.20	0.16	0.22	0.29	0.25	0.12	0.13
374	G	0.10	0.06	0.06	0.54	0.22	0.01	0.02	0.05	0.06	0.00	0.00	0.14	0.11	0.11	0.10
375	A	0.23	0.14	0.09	0.09	0.13	0.18	0.25	0.05	0.08	0.00	0.00	0.21	0.09	0.06	0.08
376	U	0.14	0.16	0.11	0.12	0.08	0.22	0.17	0.10	0.10	0.02	0.03	0.21	0.11	0.14	0.12
377	U	0.08	0.06	0.07	0.12	0.12	0.11	0.10	0.07	0.03	0.02	0.02	0.17	0.22	0.16	0.14
378	U	0.15	0.05	0.05	0.09	0.08	0.07	0.11	0.15	0.08	0.01	0.01	0.14	0.12	0.06	0.09
379	U	0.06	0.05	0.05	0.12	0.04	0.05	0.08	0.26	0.13	0.06	0.05	0.22	0.18	0.05	0.06
380	G	0.02	0.07	0.07	0.19	0.07	0.00	0.00	0.04	0.06	0.05	0.06	0.08	0.06	0.05	0.06
381	U	0.15	0.02	0.02	0.13	0.12	0.06	0.10	0.31	0.11	0.01	0.01	0.12	0.09	0.07	0.09
382	G	0.06	0.04	0.06	0.21	0.12	0.00	0.00	0.00	0.05	0.00	0.00	0.06	0.07	0.02	0.03
383	A	0.07	0.03	0.04	0.10	0.12	0.05	0.04	0.00	0.06	0.00	0.00	0.13	0.16	0.04	0.08
384	A	0.08	0.12	0.06	0.13	0.17	0.10	0.06	0.02	0.10	0.05	0.06	0.18	0.16	0.03	0.05
385	G	0.19	0.07	0.04	0.28	0.35	0.00	0.00	0.01	0.11	0.06	0.07	0.12	0.17	0.12	0.07
386	G	0.28	0.21	0.01	0.65	0.92	0.00	0.00	0.15	0.12	0.17	0.11	0.28	0.07	0.07	0.04
387	A	0.83	1.03	0.53	0.59	0.56	1.25	0.75	0.56	0.51	0.39	0.54	0.67	0.11	0.21	0.01
388	U	0.77	1.17	0.58	ND	-	0.19	0.28	2.52	1.31	0.45	0.64	0.34	0.23	1.03	0.65