

6-2021

**INSIGHTS INTO THE PACKAGING OF MOUSE MAMMARY TUMOR
VIRUS (MMTV) GENOMIC RNA BY IDENTIFYING PR77^{GAG}
BINDING SITES INVOLVED DURING ITS SELECTIVE
ENCAPSIDATION**

Akhil Chameettachal

Follow this and additional works at: https://scholarworks.uaeu.ac.ae/all_dissertations



Part of the [Medical Immunology Commons](#), and the [Medical Microbiology Commons](#)

United Arab Emirates University
College of Medicine and Health Sciences

INSIGHTS INTO THE PACKAGING OF MOUSE MAMMARY
TUMOR VIRUS (MMTV) GENOMIC RNA BY IDENTIFYING
PR77^{GAG} BINDING SITES INVOLVED DURING ITS SELECTIVE
ENCAPSIDATION

Akhil Chameettachal

This dissertation is submitted in partial fulfilment of the requirements for the degree
of Doctor of Philosophy

Under the Supervision of Professor Tahir A. Rizvi

June 2021

Declaration of Original Work

I, Akhil Chameettachal, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this dissertation entitled “*Insights into the Packaging of Mouse Mammary Tumor Virus (MMTV) Genomic RNA by Identifying Pr77^{Gag} Binding Sites 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 A. Rizvi, in the College of Medicine and Health Sciences at UAEU. This work has not previously 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 Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my dissertation have been properly cited and acknowledged in accordance with appropriate academic conventions.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: June 15, 2021

Copyright © 2021 Akhil Chameettachal
All Rights Reserved

Advisory Committee

1) Advisor: Tahir A. Rizvi

Title: Professor

Department of Microbiology

College of Medicine and Health Sciences

2) Co-advisor: Farah Mustafa

Title: Associate Professor

Department of Biochemistry

College of Medicine and Health Sciences

3) Member: Gulfaraz Khan

Title: Professor

Department of Microbiology

College of Medicine and Health Sciences

4) Member: Ahmed Al-Marzouqi

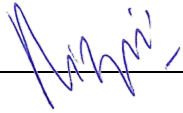
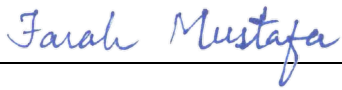

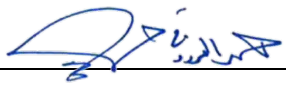
Title: Associate Professor and Chair

Department of Biochemistry

College of Medicine and Health Sciences

Approval of the Doctorate Dissertation

This Doctorate Dissertation is approved by the following Examining Committee Members:

- 1) Advisor (Committee Chair): Tahir A. Rizvi
Title: Professor
Department of Microbiology
College of Medicine and Health Sciences
Signature  Date June 15, 2021
- 2) Member: Farah Mustafa
Title: Associate Professor
Department of Biochemistry
College of of Medicine and Health Sciences
Signature  Date June 15, 2021
- 3) Member: Gulfaraz Khan
Title: Professor
Department of Microbiology
College of Medicine and Health Sciences
Signature  Date June 15, 2021
- 4) Member: Ahmed Al-Marzouqi
Title: Associate Professor and Chair
Department of Biochemistry
College of Medicine and Health Sciences
Signature  Date June 15, 2021

5) Member (External Examiner): Michael F. Summers


Title: Professor and Investigator

Howard Hughes Medical Institute

Department of Chemistry and Biochemistry

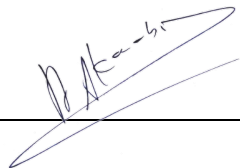
College of Natural and Mathematical Sciences

Institution: University of Maryland at Baltimore, United States

Signature  Date June 15, 2021

This Doctorate Dissertation is accepted by:

Acting Dean of the College of Medicine and Health Sciences: Professor Juma Al
Kaabi

Signature  _____ Date 26 July 2021

Dean of the College of Graduate Studies: Professor Ali Al-Marzouqi

Signature  _____ Date 26 July 2021

Copy ____ of ____

Abstract

Selective encapsidation and/or packaging of retroviral genomic RNA (gRNA) by Gag during retrovirus assembly is a crucial step for generating infectious virus particles. Despite having been studied extensively, the mechanism by which the retroviral Gag precursor selects and packages the retroviral genome remains largely unclear. Therefore, to understand the molecular mechanism(s) of mouse mammary tumor virus (MMTV) gRNA packaging, as a first step, expression of full-length recombinant Pr77^{Gag}-His₆-tag fusion protein in bacteria was done. The recombinant Gag protein was then purified from the soluble fractions of bacterial cultures using immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC). The purified recombinant Pr77^{Gag}-His₆-tag protein retained the ability to assemble *in vitro* into virus-like particles (VLPs). In parallel, the VLPs made *in vivo* following expression of the recombinant Pr77^{Gag}-His₆-tag fusion protein in eukaryotic cells could package MMTV subgenomic RNAs. Next, RNA binding and footprinting assays using the purified protein and *in cell* gRNA packaging experiments identified two critical, non-redundant Pr77^{Gag} binding sites. These binding sites include: i) a stretch of purines in a hairpin loop immediately adjacent to the dimerization initiation site (DIS) hairpin, thus forming a bifurcated stem-loop structure and ii) the primer binding site (PBS). Despite the presence of the packaging signals on both unspliced and spliced RNAs, Pr77^{Gag} specifically bound to unspliced RNA, which is the only one that can adopt the native bifurcated stem-loop structure. Together this study demonstrates the minimal packaging elements at both sequence and structural levels required to initiate MMTV gRNA packaging. Unlike purine rich regions, the direct involvement of PBS in retroviral gRNA packaging has not been documented in retroviruses. These findings add to the knowledge of retroviral gRNA packaging and assembly, making it a potential target for novel therapeutic approaches as well as the development of safer gene therapy vectors.

Keywords: Retrovirus, mouse mammary tumor virus (MMTV), RNA-protein interaction, virus assembly, protein expression, RNA packaging, RNA-Gag interactions, Pr77^{Gag}, footprinting, purine-rich sequence.

Title and Abstract (in Arabic)

فهم آليات تجميع الحمض النووي الريبوزي لفيروس ال MMTV من خلال تحديد مواقع ربط بروتين ال Pr77^{Gag} والمتضمن أثناء عملية التجميع الانتقائي

الملخص

يعد التجميع الانتقائي و / أو تغليف الحمض النووي الريبوزي الجيني للفيروسات الارتجاعية (gRNA) بواسطة بروتين ال Gag أثناء عملية تكوين الفيروس، خطوة حاسمة لتوليد جزيئات الفيروس المعدية. على الرغم من دراستها على نطاق واسع، إلا أن الآلية التي من خلالها يختار بروتين ال Gag، الحمض النووي الريبوزي الخاص بالفيروس من بين الاحماض النووية الريبوزية الأخرى المتواجدة في الخلية، غير واضحة إلى حد كبير. لذلك، لفهم الآلية (الآليات) الجزيئية لتجميع الحمض النووي الريبوزي (gRNA) لفيروس ال (MMTV)، كخطوة أولى، تم التعبير عن بروتين الاندماج Pr77^{Gag}-His₆-tag كامل الطول في البكتيريا. ثم بعد ذلك، تم تنقية البروتين من الأجزاء القابلة للذوبان باستخدام كروماتوغرافيا تقارب ايونات المعدن الثابت (IMAC) وكروماتوغرافيا الاستبعاد عن طريق اختلاف الحجم (SEC).

احتفظ البروتين المنقى Pr77^{Gag}-His₆-tag بالقدرة على التجميع في المختبر في جزيئات تشبه الفيروسات (VLPs). وفي نفس السياق الموازي، يمكن لجزيئات ال VLPs المصنوعة بعد التعبير عن بروتين الاندماج ال Pr77^{Gag}-His₆-tag في الخلايا حقيقية النواة أن تقوم بتجميع الحمض النووي الريبوزي الخاص بفيروس ال MMTV. بعد ذلك، حددت اختبارات ربط الحمض النووي الريبوزي (gRNA) وتجارب البصمة باستخدام البروتين المنقى وتجارب تجميع ال gRNA الخلوية موقعين مهمين وغير متكررين لربط ال Pr77^{Gag}. تشمل مواقع الربط هذه: (1) امتداداً من البيورينات في شكل حلقة مجاورة مباشرة لحلقة موقع البدء (DIS)، وبالتالي تشكل بنية حلقة جذعية متشعبة و (2) موقع الربط التمهيدي (PBS). على الرغم من وجود إشارات التجميع على كل من RNAs المقسمة وغير المقسمة، فإن Pr77^{Gag} مرتبط على وجه التحديد بال RNA غير المقسم، وهو الوحيد الذي يمكنه التعرف والارتباط ببنية الحلقة الجذعية الأصلية المتشعبة.

توضح هذه الدراسة الحد الأدنى من عناصر التجميع الانتقائي على المستويين التسلسلي والهيكلية المطلوب لبدء تجميع ال gRNA الخاص بفيروس ال MMTV. على عكس المناطق

الغنية بالبيورين، لم يتم توثيق المشاركة المباشرة للـ PBS في تجميع الحمض النووي الريبوزي في الفيروسات الارتجاعية. تضيف هذه النتائج إلى المعرفة بعمليات التجميع الانتقائي في الفيروسات الارتجاعية، مما يجعلها وسيلة محتملة للطرق العلاجية الجديدة بالإضافة إلى القدرة على تطوير نوافل للعلاج الجيني أكثر فعالية وأمان.

مفاهيم البحث الرئيسية: الفيروسات الارتجاعية، فيروس الاورام الثديية للفئران (MMTV)، تفاعل البروتين مع الـ RNA، تجميع الفيروس، التعبير عن البروتين، تجميع وتغليف الـ RNA، تفاعلات الـ RNA-Gag، Pr77^{Gag}، البصمة، التسلسل الغني بالبيورين.

Acknowledgements

It is with a great sense of gratitude that I present this dissertation, which has witnessed my assiduous hard work and now, marks my egress from the framework of all my well-wishers. I have had stimulating discussions with several scholarly individuals and friends, and the essence of these discussions has brought forth a finesse to my dissertation. Hence, I would like to express my unwavering gratitude and greatest appreciation to all those who have steadfastly stood by me through this journey.

Hereby I sign the hallmark of gratitude to my guide, Prof. Tahir A. Rizvi, for his continued guidance and motivation enabling me to successfully submit an innovative and erudite dissertation. I am indebted to Prof. Rizvi for understanding my strengths and weaknesses and thus, constantly helping me improve. He has had the exceptional ability to always read a student's mind.

My gratitude never ends without thanking my team members at Prof. Rizvi's laboratory, Dr. Lizna M. Ali, Ms. Vineeta N. Pillai, Dr. Fathima Nuzra Nagoor Pitchai, Ms. Anjana Krishnan, and Mr. Suresha G. Prabhu for providing me a sound laboratory atmosphere to carry out my work. The highly professional, yet-fun filled environment at the lab has played an important role in molding me into a competent scientist. Being a part of this team for five years has inculcated in me the qualities of a great team player and provided me a novel perspective towards research.

I would like to extend my heartfelt thanks to Dr. Roland Marquet, Dr. Valérie Vivet-Boudou and their entire team for the excellent training I gained at their lab

and, appreciate all their efforts to make my stay in Strasbourg and training in their lab comfortable and enjoyable.

I thank my beloved family and friends for their constant support, encouragement, and care that has always propelled me forward in life. Last but not the least, I thank the Almighty for his blessings.

Dedication

To my parents, siblings, my wife, and my son

Table of Contents

Title	i
Declaration of Original Work	ii
Copyright.....	iii
Advisory Committee	iv
Approval of the Doctorate Dissertation	v
Abstract	viii
Title and Abstract (in Arabic)	ix
Acknowledgements	xi
Dedication	xiii
Table of Contents	xiv
List of Tables.....	xix
List of Figures	xx
List of Abbreviations.....	xxiii
Chapter 1: Introduction	1
1.1 Retroviruses: Brief History and Classification.....	1
1.2 Virion Structure and Genome Organization.....	2
1.2.1 Non-Coding Regions	4
1.2.2 Coding Regions	6
1.3 Retroviral Life Cycle.....	9
1.3.1 Attachment, Membrane Fusion and Entry.....	9
1.3.2 Uncoating and Nuclear Import	10
1.3.3 Reverse Transcription.....	11
1.3.4 Integration.....	14
1.3.5 Transcription and Processing of Viral RNAs	16
1.3.6 Nuclear Export.....	18
1.3.7 Translation	19
1.3.8 Assembly and RNA Packaging	21
1.3.9 Release and Maturation	23
1.4 RNA Packaging in Retroviruses	25

1.4.1 RNA Binding in Context of Gag	25
1.4.2 The Ψ RNA	27
1.4.3 Ψ RNA Acts as the Nucleation Point for Gag Multimerization.....	32
1.4.4 Nucleus as the Initial Site of gRNA Selection	34
1.4.5 Role of Nuclear Export Pathway in RNA Packaging.....	35
1.4.6 Role of Assembly Intermediates in gRNA Packaging	36
1.5 Retroviral RNA Dimerization	37
1.6 Mouse Mammary Tumor Virus.....	43
1.6.1 The Pr77 ^{Gag} Precursor Polyprotein.....	43
1.6.2 Two Frameshifts are Required for Maintaining Ratio of MMTV Gag, Gag-Pro and Gag-Pro-Pol	44
1.6.3 MMTV Envelope (Env) Protein.....	45
1.6.4 Accessory Proteins Encoded by MMTV Genome	46
1.6.5 MMTV Pathogenesis.....	47
1.6.6 Genomic RNA Packaging in MMTV.....	48
1.7 Objectives	51
1.7.1 Specific Aim I: Expression, Purification and Characterization of MMTV Pr77 ^{Gag}	51
1.7.2 Specific Aim II: Identification and Characterization of Specific Pr77 ^{Gag} Binding Site(s) on MMTV gRNA	52
1.7.3 Specific Aim III: Establish Biological Correlation between the Pr77 ^{Gag} Binding site(s) and MMTV gRNA Packaging.....	52
Chapter 2: Biochemical and Functional Characterization of Mouse Mammary Tumor Virus Full-Length Pr77 ^{Gag} Expressed in Prokaryotic and Eukaryotic Cells.....	53
2.1 Abstract	53
2.2 Introduction	54
2.3 Materials and Methods	58
2.3.1 Nucleotide Numbers	58
2.3.2 Full-Length Recombinant Gag Prokaryotic Expression Plasmids	58
2.3.3 Full-Length Recombinant Gag Eukaryotic Expression Plasmids	60

2.3.4	<i>Escherichia coli</i> Strains and Growth Media.....	61
2.3.5	Expression of Recombinant Full-Length Pr77 ^{Gag} -His ₆ - Tagged Protein in Bacteria.....	62
2.3.6	Affinity Purification and Gel Filtration Chromatography.....	62
2.3.7	Expression of Recombinant Full-Length Pr77 ^{Gag} -His ₆ - Tagged Protein in Eukaryotic Cells.....	63
2.3.8	Estimation of RNA Packaging Potential by Reverse Transcriptase PCR.....	64
2.3.9	Sodium dodecyl sulfate-Polyacrylamide Gel Electrophoresis and Western Blotting.....	65
2.3.10	Detection of Prokaryotically-Expressed Virus-Like Particles Using Transmission Electron Microscopy.....	65
2.3.11	<i>In Vitro</i> Assembly of Virus-Like Particles from Bacterially Expressed Recombinant Full-Length Pr77 ^{Gag} - His ₆ -Tag Protein.....	66
2.4	Results and Discussion.....	66
2.4.1	Successful Expression of Full-Length Recombinant MMTV Pr77 ^{Gag} -His ₆ -Tagged Protein in Bacteria.....	66
2.4.2	Full-Length MMTV Pr77 ^{Gag} -His ₆ -Tagged Fusion Protein is Expressed in the Soluble Form in Bacteria.....	68
2.4.3	Immobilized Metal Affinity Chromatography Purification of the Soluble Fraction Containing Recombinant Full- Length Pr77 ^{Gag} -His ₆ -Tagged Fusion Protein.....	73
2.4.4	Gel Filtration Chromatography Purification of the IMAC- Purified Recombinant Full-Length Pr77 ^{Gag} -His ₆ -Tagged Fusion Protein.....	74
2.4.5	<i>In Vitro</i> Assembly to Form Virus-Like Particles by the Recombinant Full-Length Pr77 ^{Gag} -His ₆ -Tagged Fusion Protein.....	75
2.4.6	Recombinant Full-Length MMTV Pr77 ^{Gag} -His ₆ -Tagged Protein Expressed in Bacteria Can Form Virus-Like Particles.....	77
2.4.7	Eukaryotically-Expressed, Full-Length Recombinant Pr77 ^{Gag} His ₆ -Tagged Fusion Protein Can Form Virus-Like Particles Competent to Package Unspliced Subgenomic RNA.....	79
2.5	Conclusions.....	84

2.6 Funding.....	84
Chapter 3: A Purine Loop and the Primer Binding Site are Critical for the Selective Encapsidation of Mouse Mammary Tumor Virus Genomic RNA by Pr77 ^{Gag}	85
3.1 Abstract	85
3.2 Introduction	86
3.3 Materials and Methods	91
3.3.1 Nucleotide Numbers	91
3.3.2 Expression and Purification of Full-Length MMTV Pr77 ^{Gag}	92
3.3.3 Physical Characterization of Pr77 ^{Gag} by Dynamic Light Scattering (DLS)	92
3.3.4 Cloning, Mutagenesis, <i>In Vitro</i> Transcription and RNA Purification.....	93
3.3.5 Band-Shift and Band-Shift Competition Assays	95
3.3.6 Filter-Binding Assay.....	96
3.3.7 Footprinting Experiments using hSHAPE in the Presence and Absence of Pr77 ^{Gag}	97
3.3.8 In Cell Genetic Complementation Assay	99
3.3.9 Nucleocytoplasmic Fractionation, Isolation of RNA and cDNA Preparation.....	102
3.3.10 Quantitative RT-PCR (RT-qPCR) for Transfer Vector RNA Packaging Efficiency	102
3.3.11 <i>In Vitro</i> Dimerization Assay.....	103
3.4 Results	103
3.4.1 Physical Characterization of Functional Pr77 ^{Gag} Polyprotein	103
3.4.2 Pr77 ^{Gag} Binds Specifically to Unspliced MMTV gRNA.....	104
3.4.3 Single-Stranded Purines (ssPurines) are Crucial for Pr77 ^{Gag} Binding <i>In Vitro</i>	109
3.4.4 ssPurines in the Spliced RNAs are Base-Paired.....	113
3.4.5 Pr77 ^{Gag} Attenuates the Reactivity of ssPurines towards SHAPE Reagents	116
3.4.6 Pr77 ^{Gag} Attenuates Reactivities of Nucleotides Other than ssPurines.....	120

3.4.7 Role of ssPurines during in Cell Packaging and Transduction of ψ -Containing RNAs.....	120
3.4.8 Nucleotides in the PBS Region Play a Crucial Role in RNA Packaging	125
3.4.9 The role of PBS in MMTV gRNA Packaging is at the Primary Sequence Level	132
3.5 Discussion	137
3.6 Funding.....	146
Chapter 4: Conclusions and Future Directions	147
4.1 Conclusions	147
4.2 Future Directions.....	148
References	151
List of Publications	191
Appendices.....	192
Appendix A	192
Appendix B	199
Appendix C	218

List of Tables

Table 1: Classification of retroviruses based on genera and morphology	2
--	---

List of Figures

Figure 1: Schematic cross section through a retroviral particle.....	3
Figure 2: Classification based on morphology of mature virions.....	4
Figure 3: Organization of retroviral genomic RNA.....	6
Figure 4: Genomic organization of retroviruses.....	8
Figure 5: Reverse transcription: process of reverse transcription of the retroviral genome.....	14
Figure 6: Splicing events in simple and complex retroviruses.....	18
Figure 7: Organization of the immature and mature HIV-1 virions.....	24
Figure 8: Different mechanisms of selective packaging of unspliced RNA.....	28
Figure 9: The two proposed structures for HIV-1 packaging signal RNA.....	30
Figure 10: Stem-loops in MLV ψ	32
Figure 11: The Kissing-Loop Model of HIV-1 RNA Dimerization.....	38
Figure 12: Model showing the putative conformational switch that is proposed to regulate translation and packaging of the HIV-1 genomic RNA.....	41
Figure 13: Schematic overview of the role of RNA dimerization in the retroviral life cycle.....	42
Figure 14: Schematic illustration of the gag, pro and pol genes in MMTV provirus.....	45
Figure 15: Coding regions in the MMTV proviral genome and transcripts.....	47
Figure 16: SHAPE constrained RNA structure model of MMTV packaging signal RNA.....	49
Figure 17: Mfold RNA secondary structure predictions of wild type and mutant MMTV packaging signal RNAs.....	51
Figure 18: Construction of the recombinant full-length Pr77 ^{Gag} bacterial expression vector.....	60
Figure 19: Expression of recombinant Pr77 ^{Gag} -His ₆ -tag fusion protein in <i>Escherichia coli</i> (<i>E. coli</i>).....	67
Figure 20: Recombinant Pr77 ^{Gag} -His ₆ -tag fusion protein expressed in soluble fraction of <i>E. coli</i>	69
Figure 21: Silent mutations in the Shine-Dalgarno-like sequence and the predicted relative translation rates from the first and the second in-frame ATGs.....	71
Figure 22: Expression of recombinant Pr77 ^{Gag} -His ₆ -tag fusion protein from AK7 in soluble fractions of <i>E. coli</i> before and after immobilized metal affinity chromatography (IMAC) purification.....	72
Figure 23: Resolution of IMAC-purified recombinant Pr77 ^{Gag} -His ₆ -tag fusion protein by size exclusion chromatography and western blot analysis.....	75

Figure 24: Transmission electron micrographs showing virus-like particles (VLPs) following in vitro assembly	77
Figure 25: Formation of VLPs by recombinant Pr77 ^{Gag} -His ₆ -tag fusion protein in <i>E. coli</i> BL21(DE3)	79
Figure 26: Schematic representation of the two-plasmid genetic complementation assay to demonstrate VLPs formation following Pr77 ^{Gag} -His ₆ -tag protein expression in eukaryotic cells and their ability to package MMTV subgenomic RNA	83
Figure 27: Schematic representation of MMTV genome, organization of different domains of full-length Gag (Pr77 ^{Gag}) and higher order structure of MMTV packaging signal RNA	90
Figure 28: Design and rationale of the MMTV three-plasmid genetic complementation assay	101
Figure 29: Dynamic light scattering of purified Pr77 ^{Gag} in the RNA binding buffer	104
Figure 30: The MMTV packaging signal RNA binds to Pr77 ^{Gag}	105
Figure 31: Unspliced genomic and spliced RNAs binds differentially to MMTV Pr77 ^{Gag}	107
Figure 32: Binding of unspliced genomic and spliced RNAs to MMTV Pr77 ^{Gag} analyzed by filter-binding assay	108
Figure 33: Differential binding abilities of ssPurines and its mutant RNAs to MMTV Pr77 ^{Gag}	110
Figure 34: Binding of ssPurines and its mutant RNAs to MMTV Pr77 ^{Gag} analysed by filter-binding assay	112
Figure 35: In vitro dimerization ability of the WT (SA35) and the ssPurine mutant (AK18, AK62 & AK63) RNAs	113
Figure 36: SHAPE-validated RNA secondary structure of first 712 nts of <i>env</i> and <i>sag</i> RNAs	115
Figure 37: Band-shift assay showing the formation of a single high-affinity Pr77 ^{Gag} - radiolabeled gRNA complex	116
Figure 38: SHAPE-validated secondary structure and footprints of Pr77 ^{Gag} on MMTV packaging signal RNA	118
Figure 39: Histograms showing the Pr77 ^{Gag} -induced attenuation of the SHAPE reactivities of nucleotides in MMTV gRNA	119
Figure 40: Role of ssPurines in MMTV gRNA packaging and transduction efficiencies	122
Figure 41: Role of nucleotides outside ssPurines that showed attenuation of SHAPE reactivities in the presence of MMTV Pr77 ^{Gag} on MMTV gRNA packaging and transduction efficiencies	126
Figure 42: Control PCR amplifications used for RNAs with mutations that have been introduced in regions other than ssPurines	127
Figure 43: Band-shift competition assays for the clones containing mutations in PBS region and basal part of SL3	129

Figure 44: SHAPE-validated secondary structure of wild type (SA35) and mutant containing substitutions in the bottom part of SL3 RNAs	130
Figure 45: SHAPE-validated RNA secondary structure of WT and AK74 RNAs	131
Figure 46: Histograms showing SHAPE reactivities of the mutation (AK74) introduced into the PBS region in the absence and presence of Pr77 ^{Gag}	132
Figure 47: Role of PBS in dimerization and packaging of MMTV gRNA	134
Figure 48: Representative gel picture showing in vitro dimerization of WT gRNA (SA35) and PBS mutant RNAs in native (TBM) and denaturing (TBE) conditions	135
Figure 49: Control PCR amplifications used for RNAs with mutations that have been introduced in PBS region.....	136

List of Abbreviations

ALV	Avian Leucosis Virus
BMH	Branched Multiple Hairpin
CFU	Colony Forming Unit
CTE	Constitutive Transport Element
DIS	Dimerization Initiation Site
DLS	Dynamic Light Scattering
DNA	Deoxy Ribonucleic Acid
DTT	Dithiothreitol
Gag	Group Specific Antigen
gRNA	Genomic Ribonucleic Acid
HEK	Human Embryonic Kidney
His ₆	Hexa-Histidine
HIV	Human Immunodeficiency Virus
HRP	Horse Radish Peroxidase
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β - d-1-Thiogalactopyranoside
LDI	Long Distance Interaction
LRI	Long Range Interaction
LTR	Long Terminal Repeat
MA	Matrix
MLV	Murine Leukemia Virus
MMTV	Mouse Mammary Tumor Virus

MPMV	Mason-Pfizer Monkey Virus
mSD	Major Splice Donor
NC	Nucleocapsid
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
Pol	Polymerase
Pr77 ^{Gag}	Precursor Polyprotein 77 Gag
Pro	Protease
Psi	Packaging Signal
RNA	Ribonucleic Acid
RPE	Relative Packaging Efficiency
RT	Reverse Transcriptase
RT-qPCR	Quantitative Reverse Transcriptase PCR
SDS	Sodium Dodecyl Sulphate
SEAP	Secreted Alkaline Phosphatase
SHAPE	Selective 2'Hydroxyl Acylation Analyzed by Primer Extension
SL	Stem Loop
SU	Surface
TM	Transmembrane
U3	3' Unique
U5	5' Unique
UTR	Untranslated Region
VLP	Virus Like Particle

WT

Wild Type

Chapter 1: Introduction

1.1 Retroviruses: Brief History and Classification

Retroviruses are the members of the family *Retroviridae*; a unique family of viruses, which are enveloped viruses containing two copies of single-stranded positive sense RNA as their genome which is often referred as genomic RNA (gRNA). Their life cycle includes extraordinary steps of reverse transcription of the gRNA into DNA, which is then imported into the nucleus where it integrates into the host chromosome (Coffin et al., 1997). Retroviruses infect large number of host species and cause a variety of diseases including immunodeficiency, tumors and leukemia both in humans and animals. The first identified retrovirus was the causative agent of leukemia in chickens (avian leukosis virus; ALV) by Ellermann and Bang in 1908. In 1911, Peyton Rous reported the transmission of sarcomas in the chicken through cell-free filtrates and subsequently named Rous sarcoma virus (RSV; Rous, 1911). Human T-lymphotropic virus 1 (HTLV-1), the first discovered human retrovirus was described in 1980 (Poiesz et al., 1980; Yoshida et al., 1982). Three years later, Montagnier and co-workers isolated a virus from the lymph nodes of patients with acquired immunodeficiency syndrome (AIDS) and in 1984, the link between human immunodeficiency virus type 1 (HIV-1) and AIDS was established by Gallo and colleagues (Barré-Sinoussi et al., 1983; Gallo, 1988; Gallo et al., 1984).

According to International Committee on Taxonomy of Viruses (ICTV), the family *Retroviridae* further divided into two subfamilies; *Orthoretrovirinae* and *Spumavirinae* and based on their evolutionary relatedness, they are further classified into 7 genera (Knipe & Howley, 2013; Table 1). The *Orthoretrovirinae* contains

Alpharetroviruses, *Betaretroviruses*, *Gammaretroviruses*, *Deltaretroviruses*, *Epsilonretroviruses* and *Lentiviruses*. The genus *Spumaviruses* is the single member of the subfamily *Spumavirinae* (Table 1).

Table 1: Classification of retroviruses based on genera and morphology (adapted from Knipe & Howley, 2013)

Name	Examples	Morphology
Alpharetrovirus	Avian leukosis virus (ALV) Rous sarcoma virus	C type
Betaretrovirus	Mouse mammary tumor virus (MMTV) Mason-Pfizer monkey virus (M-PMV) Jaagsiekte sheep retrovirus	B, D type
Gammaretrovirus	Murine leukemia viruses (MuLV) Feline leukemia virus (FeLV) Gibbon ape leukemia virus (GaLV) Reticuloendotheliosis virus (RevT)	C type
Deltaretrovirus	Human T-lymphotropic virus type 1, 2 Bovine leukemia virus (BLV) Simian T-lymphotropic virus type 1, 2, 3	Rod-shaped core
Epsilonretrovirus	Walleye dermal sarcoma virus Walleye epidermal hyperplasia virus 1	—
Lentivirus	Human immunodeficiency virus type 1 Human immunodeficiency virus type 2 Simian immunodeficiency virus (SIV) Equine infectious anemia virus (EIAV) Feline immunodeficiency virus (FIV) Caprine arthritis encephalitis virus (CAEV) Visna/maedi virus	Rod/Cone-shaped cores
Spumavirus	Human foamy virus	Immature

1.2 Virion Structure and Genome Organization

Retroviruses are spherical in shape measuring approximately 80- 120 nm in diameter, comprising of approximately 2000-5000 molecules of structural protein called Group-specific antigen (Gag; Briggs et al., 2004; Carlson et al., 2008; Coffin et al., 1997). The arrangement of structural proteins in the mature virion is depicted in Figure 1.

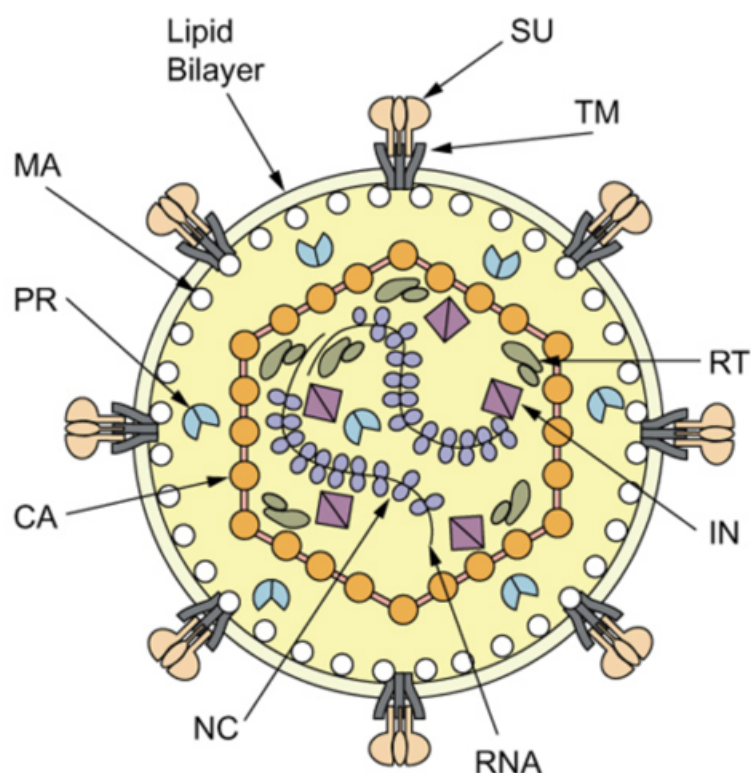


Figure 1: Schematic cross section through a retroviral particle

The viral envelope is formed by a cell-derived lipid bilayer into which proteins encoded by the *envelope (env)* region of the viral genome are inserted. These consist of the transmembrane (TM) and the surface (SU) components linked together by disulfide bonds. Internal non-glycosylated structural proteins are encoded by the *gag* region of the viral genome. They are the matrix (MA) protein, capsid (CA) protein and nucleocapsid (NC) protein. Major products of the *polymerase (pol)*-coding region are reverse transcriptase (RT) and integrase (IN). The protease (PR) is derived from the *protease (pro)* gene between *gag* and *pol*. Figure and legend adapted from Coffin et al., 1997; Voisset & Andrawiss, 2000.

Historically, retroviruses were classified based on the structure and location of their nucleocapsid core within the mature particle (Table 1, Figure 2). This classification includes: an intracellular membrane-lacking forms (intracytoplasmic A particle), type-B: an extracellular eccentric, spherical core, type-C: spherical/cone shaped core and type-D: a cylindrical/bar shaped core.

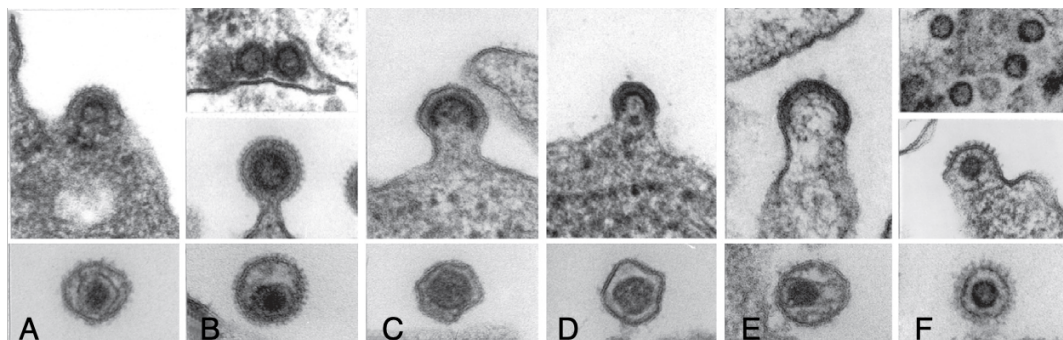


Figure 2: Classification based on morphology of mature virions

(A) *Alpharetrovirus*: avian leukosis virus (ALV), type-C morphology; (B) *Betaretrovirus*: mouse mammary tumor virus (MMTV), type-B; (C) *Gammaretrovirus*: murine leukemia virus (MLV), type-C; (D) *Deltaretrovirus*: bovine leukemia virus (BLV), type-C; (E) *Lentivirus*: human immunodeficiency virus 1 (HIV-1), type-C and (F) *Spumavirus*: the human foamy virus (HFV), immature/partially mature. The bar represents 100 nm. The figure and legend adapted from Hunter, 2008.

The infectious retrovirus particle consists of two copies of single stranded, 7-12 kilobase (kb) long, linear, non-segmented, positive-sense RNA as their genome which are capped at the 5' end and poly adenylated at the 3' end. The key sequences and features of the retroviral gRNA are depicted in Figure 3 and include:

1.2.1 Non-Coding Regions

- 1) R (Repeat): A short sequence (15-250 nucleotides (nts); 15 nts in case of MMTV and approximately 250 nts in HTLV-1) that is repeated twice in the RNA: extreme 5' end just after the cap and at the 3' end immediately upstream of the poly(A) tail. In lentiviruses, R also consists of highly structured region called *trans*-activation response element (TAR) that recruit transcriptional *trans*-activator (Tat) protein required for efficient transcription of the gRNA (Coffin et al., 1997; Hunter, 2008; Knipe & Howley, 2013).
- 2) U5 (Unique to 5'): Sequence of 70-250 nts located downstream of the R and before the primer binding site (PBS). U5 contains one of the attachment sites

(U5 *att*) necessary for proviral integration into the host chromosome (Coffin et al., 1997; Hunter, 2008; Knipe & Howley, 2013).

- 3) PBS (primer binding site): 18 nts sequence complementary to the 3' terminus of host tRNA which acts as the primer for the initiation of reverse transcription (Coffin et al., 1997; Hunter, 2008; Knipe & Howley, 2013).
- 4) PPT (polypurine tract): During reverse transcription, PPT serve as the primer for synthesis of the plus (+) strand of viral DNA. As the name indicates, it consists of a stretch of 'A' and 'G' and usually 7-18 nts long located just upstream of U3 region (Coffin et al., 1997; Hunter, 2008; Knipe & Howley, 2013).
- 5) U3 (Unique to 3'): this region contains *cis*-acting elements such as promoter and enhancer sequences required for viral RNA transcription and the second *att* site (U3 *att*) required for integration. U3 also may contain the coding sequences (e.g., portion of *nef* in HIV-1, *sag* and *rem* in MMTV; Coffin et al., 1997; Hunter, 2008; Knipe & Howley, 2013; Ross, 2010).
- 6) Poly(A) tract: Downstream to the 3' R region, is the post-transcriptionally added poly(A) tail of 50-200 nts. The most common polyadenylation signal is AAUAAA and either found in R region (e.g., HIV and Mo-MLV) or U3 region (e.g., HTLV-1, ASLV and MMTV; Coffin et al., 1997; Hunter, 2008; Knipe & Howley, 2013).

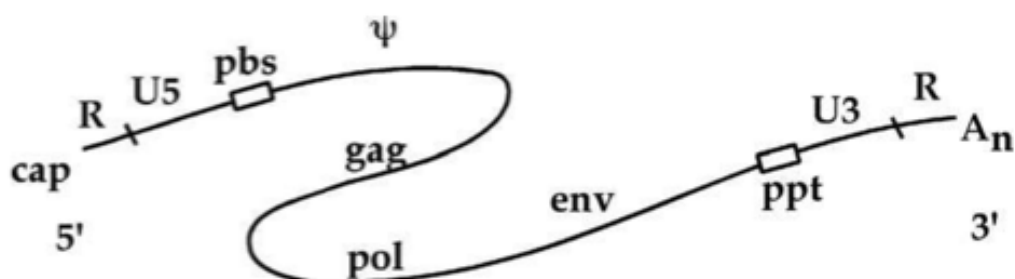


Figure 3: Organization of retroviral genomic RNA

The single-stranded RNA genome is depicted as a curved line. From 5' to 3' along the RNA, the features include a 5' cap structure; R, a sequence block repeated at both 5' and 3' ends; U5, a unique 5' sequence block; pbs, the primer binding site and site of initiation of minus strand DNA synthesis; Ψ the major recognition site for the packaging of the viral RNA into the virion particle; the *gag*, *pol* and *env* genes; ppt, the polypurine tract and site of initiation of the plus strand DNA synthesis; U3, a unique 3' sequence block; the second copy of the R sequence; and finally, a 3' poly(A) sequence. Figure and legend adapted from Knipe & Howley, 2013.

1.2.2 Coding Regions

All known retroviruses encode four canonical proteins which are necessary for viral replication: group specific antigen (Gag), protease (Pro), polymerase (Pol) and envelope (Env). Based on the genetic complexity, retroviruses are classified into simple and complex retroviruses; simple retroviruses encode only Gag, Pro, Pol and Env while complex retroviruses in addition to these canonical proteins also encode a number of small regulatory proteins (accessory proteins) which are encoded from singly or alternatively spliced mRNAs (Figure 4; Coffin et al., 1997; Hunter, 2008; Knipe & Howley, 2013).

- 1) Gag: Gag is expressed as a precursor polyprotein and cleaved by viral protease during maturation into three structural proteins: matrix (MA), capsid (CA) and nucleocapsid (NC).
- 2) Pro and Pol: Like Gag, retroviral Pol is also a polyprotein which is cleaved by the viral encoded protease (PR). The constituent proteins generated after PR

mediated cleavage depends on the retroviral species. In lentiviruses, the PR, reverse transcriptase (RT) and integrase (IN) are encoded in the *pol* gene and expressed as Gag-Pol fusion protein, while in *Alpharetroviruses*, the PR is expressed as part of *Gag* open reading frame (ORF). In *Betaretroviruses*, the PR is expressed from a separate ORF as Gag-Pro and Gag-Pro-Pol fusion proteins (Coffin et al., 1997; Hunter, 2008; Knipe & Howley, 2013).

- 3) Env: Env is translated from a singly spliced mRNA and is a membrane targeting protein which is necessary for retroviral attachment, fusion and entry. The host cellular protease furin cleaves the Env into the surface (SU) and transmembrane (TM) domains (Coffin et al., 1997; Hunter, 2008; Knipe & Howley, 2013).

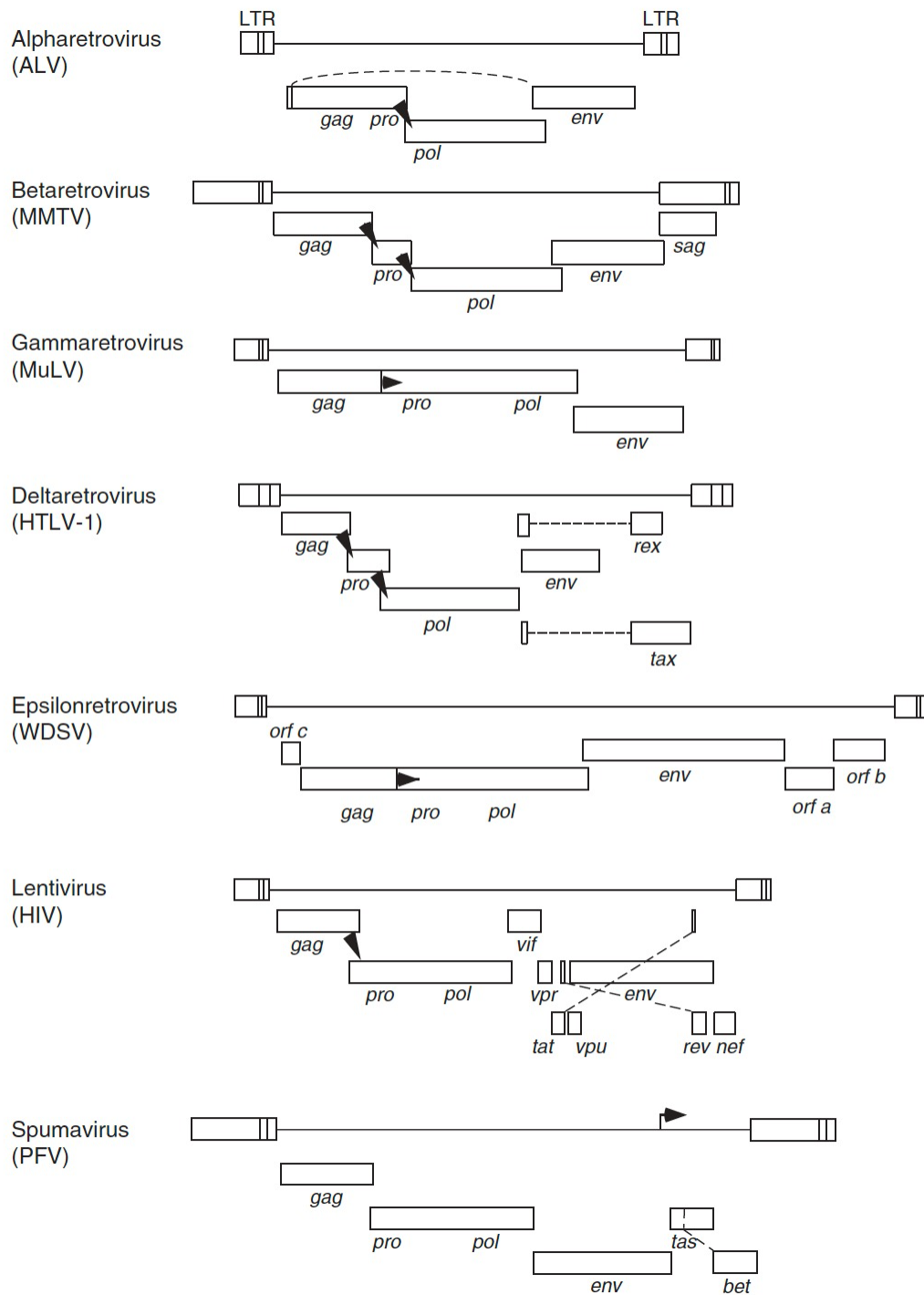


Figure 4: Genomic organization of retroviruses

The different prototypical provirus genomes for each genus are shown indicating the positions of the LTRs and encoded structural genes (*gag*, *pro*, *pol*, *env*) and certain other nonstructural genes (e.g., *tax* and *rex* in the deltaretroviruses) as well as their reading frames (ribosomal frameshift or ribosomal readthrough sites: arrowheads). LTR, long terminal repeat. Figure and legend adapted from Hunter, 2008.

1.3 Retroviral Life Cycle

1.3.1 Attachment, Membrane Fusion and Entry

This earliest step in retroviral replication involves multiple stages; initial binding followed by a conformational change in the Env protein that results in the membrane fusion. Initial binding occurs through cell receptor and SU subunit of the Env protein. This interaction leads to a drastic conformational change in Env trimer allowing the TM domain penetration into the target cell membrane which results in the fusion of viral and cell lipid bilayers (Doms & Moore, 2000; White et al., 2008). In many viruses such as orthomyxoviruses, this conformational change is aided by the acidic environment within the endosomes and takes place after receptor-mediated endocytosis. But in most retroviruses, this fusion is pH-independent; and thus can happen at the cell surface (Coffin et al., 1997). The classical experiments to study the effect of pH on fusion has been carried out using lysosomotropic reagents such as chloroquine and ammonium chloride that raises the endosomal pH. These experiments showed that HIV-1 fusion is insensitive to wide range of pH (McClure et al., 1988; Sinangil et al., 1988; Stein et al., 1987). On the other hand, in the case of ALV, it has been shown that the low pH-mediated activation is required for virus entry following the conformational change that occurs after Env-receptor binding (Mothes et al., 2000).

The type of receptor that the Env can interact with determines the retroviral cell tropism. Retroviruses utilizes various cell surface molecules as receptors, reflecting the wide range of host species that they can infect. HIV-1 uses the CD4 (cluster of differentiation 4) receptor which is expressed on the helper subset of T lymphocytes, macrophages and dendritic cells, but engineered expression of CD4

alone on non-primate and rodent cells was not sufficient for productive infection (Dalglish et al., 1984; McDougal et al., 1986). Moreover, it has been suggested that some HIV-1 variants and HIV-2 are able to infect cells in CD4-independent manner (Borsetti et al., 2000; Dragic & Alizon, 1993; Weiss, 1993). Further studies identified two members of the chemokine receptor family, CXCR4 and CCR5 as co-receptors for HIV-1 entry (Bleul et al., 1997). A population with a mutated allele of CCR5 (a 32 bp deletion) exhibited least susceptibility towards HIV-1 infection (McNicholl et al., 1997). The co-receptors for other retroviruses have not been studied extensively. In HTLV-1, the primary entry receptor is unclear. Though initially it was shown that the glucose transporter molecule GLUT-1 is the primary receptor, heparan sulfate proteoglycans (HSPGs) and neuropilin-1 (NRP-1) also play a critical role in entry. Additionally, the vascular cell adhesion molecule-1 (VCAM-1) has been identified as the potential co-receptor for HTLV-1 entry (Ghez et al., 2006; Hildreth et al., 1997; Jones et al., 2005; Manel et al., 2003).

1.3.2 Uncoating and Nuclear Import

After fusion and entry, the mature core enters the cytoplasm, leaving SU, TM and MA that remains associated with endosomal membrane. Within the capsid core, the gRNA remains attached to NC and other enzymatic proteins. The intracytoplasmic trafficking of the virus core to the nucleus occurs through cytoskeletal fibers (Kizhatil & Albritton, 1997; McDonald et al., 2002). In case of MLV, an active nuclear import mechanism has not been identified, since the entire viral core enter nucleus during mitosis (Roe et al., 1993). However, HIV-1 uses an active import mechanism to enter nucleus through nuclear pores irrespective of the fact whether the host cells are dividing or not (Bukrinsky et al., 1992; Katz et al.,

2003). Hence, due to the bigger size of the HIV-1 core (~60 nm) compared to the diameter of nuclear pore (~30 nm), initially it was assumed that the uncoating occurs within the cytoplasm and only the pre-integration complex (PIC) crosses the nuclear membrane. Consistent with this, the CA is not present in the PIC (Farnet & Haseltine, 1991; Levy et al., 2013). However, the intracellular location where the HIV-1 uncoating occurs is still debatable. Several studies suggest that the initial hypothesis is still correct: that the uncoating happens in the cytoplasm after the entry, but before nuclear import (Bukrinsky, 2004; Dvorin & Malim, 2003; Lehmann-Che & Saïb, 2004; Suzuki & Craigie, 2007). Another model proposes that until the core reaches the nuclear membrane, it remains intact and the uncoating occurs at the nuclear pore. This model also suggests that uncoating happens only after the completion of reverse transcription (Li et al., 2000). However, now it is becoming increasingly clear that the HIV-1 uncoating occurs in the nucleus. A recent study demonstrated the nuclear import of intact HIV-1 core and uncoating near the integration sites that occurs only after the completion of reverse transcription and formation of PIC (Burdick et al., 2020; Li et al., 2021). This study also showed that the CA interacts with cellular cleavage and polyadenylation specificity factor 6 (CPSF6) in order to facilitate the nuclear import of intact viral core. A number of recent other studies also underline this model (Dharan et al., 2020; Francis et al., 2020; Selyutina et al., 2020).

1.3.3 Reverse Transcription

The process of reverse transcription completes before uncoating that occurs within the nucleus. Treating with inhibitors of reverse transcription results in delayed uncoating, suggesting a check-point that ensures the release of properly formed PIC

(Burdick et al., 2020). The functional reverse transcription complex (RTC) contains the dimeric RNA, RT, CA, NC and IN (and some accessory proteins, e.g., Vpr in HIV-1). The RT possess two activities: a DNA polymerase activity that uses either RNA or DNA template to synthesize DNA strand and an RNase H activity that degrades RNA in the RNA:DNA hybrid (Coffin et al., 1997). The viral RNA has U5 and U3 at corresponding to 5' and 3' ends, respectively. On the other hand, the proviral DNA contains both U3 and U5 on either end. The U3, R and U5 at either end of the proviral DNA together is known as long terminal repeats (LTRs). Since the viral promoter is located in the U3 region, duplication of U3 to the 5' end of the DNA is necessary for further steps of viral gene expression. Hence duplication of LTR on both ends during reverse transcription is mandatory (Coffin et al., 1997).

Like other polymerases, the RT requires a free 3' hydroxyl group to initiate reverse transcription. The tRNA annealed to the PBS at the 5' untranslated region (UTR) acts as the primer. The 18-nts long PBS binds to the 3' acceptor arm and T Ψ C arms of the tRNA and determines which tRNA to bind (Das et al., 1995; Wakefield et al., 1996). HIV-1 and MMTV uses tRNA^{Lys3} as primer, RSV uses tRNA^{Trp} and tRNA^{Pro} for HTLV-1. The specific packaging of tRNA is achieved by enrichment of the corresponding charged aminoacyl tRNA synthetase in the assembling virions (Jin & Musier-Forsyth, 2019).

Retroviral reverse transcription of viral RNA into double stranded DNA (dsDNA) involves several steps (Figure 5). In the first step, the short stretch of minus strand DNA synthesis occurs copying 5' U5 and 5' R, using annealed tRNA as the primer. RNase H activity of the RT degrades the RNA from this hybrid, exposing the 5' R region which is in turn complementary to the 3' R region. Complementary base

pairing of the R region on either end leads to the first template switching known as minus strand transfer. This template switching has a significant role in the retroviral recombination since it can occur intra- or inter-molecularly, especially given the fact that retroviruses contain two copies of the gRNA (Hu et al., 1997; Jones et al., 1994). After successful minus strand transfer, RT further extends the minus strand DNA, including the U3 located at the 3' end. In the next step, the RNase H activity cleaves the RNA located between the PPT and 3' end of the viral RNA, leaving a small stretch of PPT, which is resistant to RNase H degradation and thus acts as primer for positive DNA strand synthesis. Also, degradation of tRNA by RNase H creates an overhanging PBS region at the 3' end of positive strand DNA, which is complementary to the PBS at the 5' end of the minus strand DNA, which leads to the second template switching or plus DNA strand transfer.

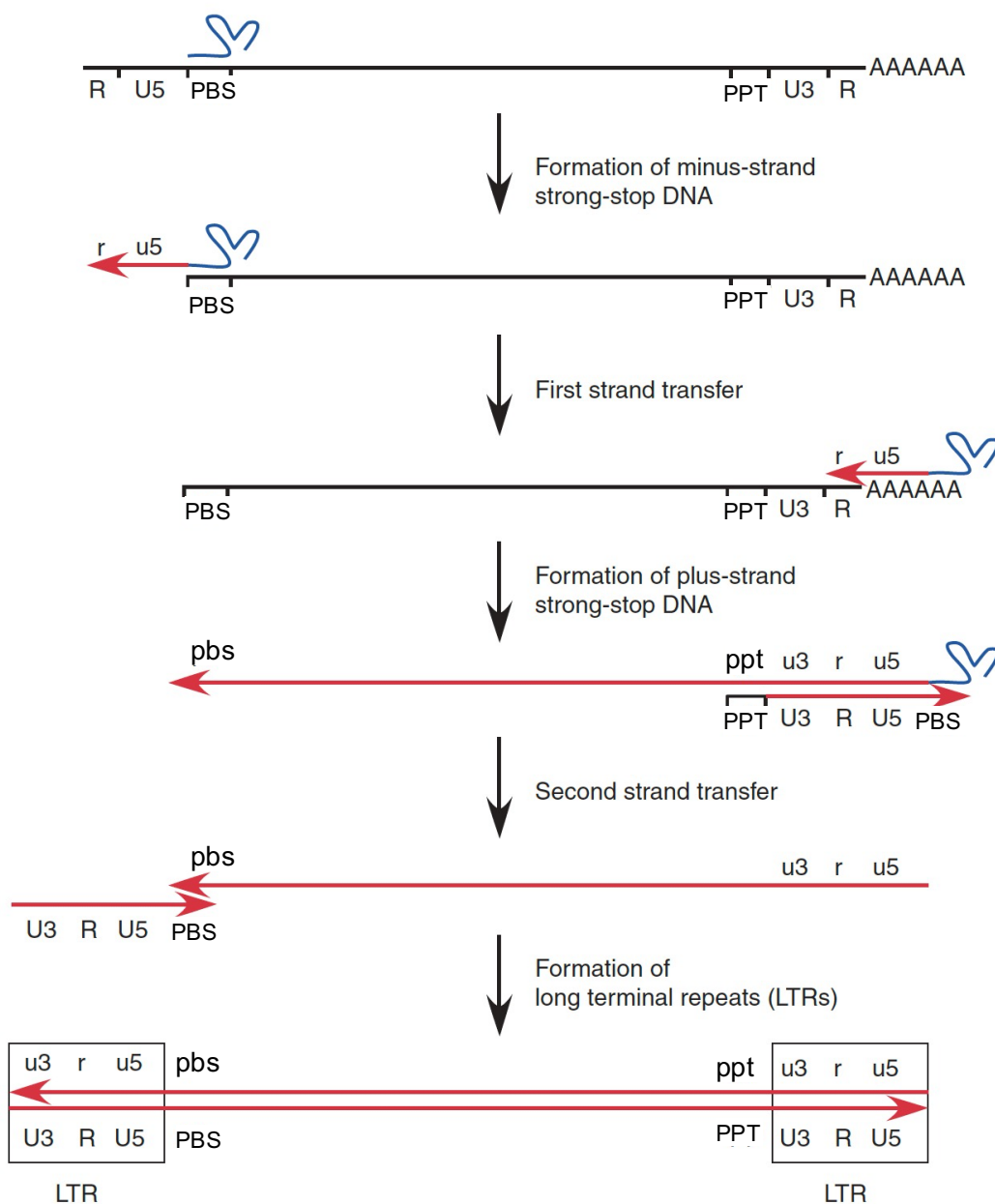


Figure 5: Reverse transcription: process of reverse transcription of the retroviral genome

Black lines represent RNA and DNA newly synthesizing DNA is represented in red line. Figure adapted from Hunter, 2008.

1.3.4 Integration

Integration of retroviral dsDNA generated through reverse transcription into the host chromosome is necessary for the continuity of retroviral life cycle. Within

the PIC (or intasome), the dsDNA is complexed primarily with IN enzyme. The integrated viral DNA is called *provirus* and serves as the template for transcription of viral RNAs. The initial seminal studies describing the fundamental steps and mechanisms involved during integration of the reverse transcribed DNA were carried out in MLV (Brown et al., 1989).

It has been shown that the selection of integration site within the host chromosome is non-random though no strict sequence specificity exists. However, based on analysis of several retroviral integration site sequences, studies have shown that virus-specific weak preference for consensus palindromic sequence does exist during retroviral integration (Holman & Coffin, 2005; Stevens & Griffith, 1994; Wu et al., 2005). In addition, the PIC preferentially targets to DNA associated with histones (nucleosomes) over naked DNA (Maskell et al., 2015; Naughtin et al., 2015). Different retroviruses have evolved to prefer different integration sites based on the chromatin state. It has also been shown that HIV-1, MLV and ASLV select transcribing regions. Within the transcribing regions, it has been demonstrated that MLV prefer the 5' end of the transcribing regions, while ASLV rarely prefer the 5' end of transcribing regions (Mitchell et al., 2004; Shih et al., 1988). The IN and CA proteins have also been implicated in the integration site selection (Lewinski et al., 2006; Schaller et al., 2011).

The *att* sites in the reverse transcribed retroviral DNA are located in the U5 and U3 regions, often as imperfect inverted repeats. The conserved CA and TG dinucleotide pairs in the U5 and U3 *att* sites play an important role in integration (Coffin et al., 1997). However, the sequences of *att* sites between different retroviruses are not highly conserved (reviewed in Ali et al., 2016). In the initial step,

the endonuclease activity of IN cleaves 3' *att* canonical CA sequences resulting in the release of dinucleotides and generating staggered ends. In the second step, a transesterification reaction joins the 3' OH groups of the viral DNA to the host DNA by forming a phosphodiester linkage. The resulting gaps in the unjoined 5' ends of viral DNA are repaired by host cell machinery, though involvement of RT and IN has also been proposed (Acel et al., 1998; Yoder & Bushman, 2000). This repair results in the duplication of a short sequence in the target sites. The number of duplicated nucleotides vary from 4-6 depending upon the virus (Lesbats et al., 2016).

1.3.5 Transcription and Processing of Viral RNAs

Expression of viral RNAs from the integrated *provirus* is facilitated by the host enzymes. The U3 region contains a promoter for RNA polymerase II and other regulatory elements. The newly transcribed RNAs are 7-methyl guanosine capped at the 5' end and polyadenylated at the 3' end. Retroviruses often encode multiple proteins from a single mRNA using the frameshift mechanism. Additionally, they increased their coding capacity by generating subgenomic RNAs by splicing and alternate splicing (Coffin et al., 1997). Apart from the promoter, the U3 region also contains several regions that binds to cellular transcription factors. For example, in the case of HIV-1, the cellular transcription factor NF κ B induces expression from the 5' LTR in T lymphocytes (Nabel & Baltimore, 1987). The MMTV U3 region contains binding sites for glucocorticoid receptors which considerably improve RNA expression and regulates MMTV infection in a tissue specific manner (Günzburg & Salmons, 1992).

Apart from the cellular transcription factors, complex retroviruses encode for *trans*-acting regulators of transcription (transcriptional *trans*-activators). *Trans*-activators expressed at very low levels are sufficient to drive high levels of viral RNA expression. The best studied retroviral *trans*-activators are Tax and Tat proteins of HTLV and HIV, respectively. The Tax protein does not directly bind to the Tax responsive element (TRE) on the HTLV LTR; instead it enhances the binding of host transcription factor cAMP response element binding/activating transcription factor (CREB/ATF) to TRE in the LTR of the provirus (Suzuki et al., 1993). On the other hand, in the case of HIV, the Tat protein bind to a stem-loop structure located in the R region (transactivation response element, TAR) of the newly transcribing RNA and increases the efficiency of transcriptional elongation (Feinberg et al., 1991).

Like eukaryotic mRNAs, the termination of transcription of retroviral mRNAs involves polyadenylation at the 3' end. The signal for polyadenylation is a highly conserved *cis*-acting sequence AAUAAA located 10-30 nts upstream of polyadenylation site which binds to cellular factor CPSF. The GU-rich and U-rich elements 10-30 nts downstream of polyadenylation site may also act as a signal for polyadenylation (Coffin et al., 1997; Guntaka, 1993). In the case of HIV-1, the AAUAAA signal sequence is present in the R and it exists in both 5' and 3' LTRs. However premature transcriptional termination due to the presence of 5' signal sequence is negatively regulated by the promoter proximity and splice site proximity (Ashe et al., 1995; Iwasaki & Temin, 1990; Weichs an der Glon et al., 1991).

After the completion of the transcription, a portion of the primary transcripts undergo splicing to increase their coding capacity. In simple retroviruses, only single

splicing occurs while in complex retroviruses, both single and multiple splicing events occur in order to generate subgenomic RNAs (Figure 6; Coffin et al., 1997).

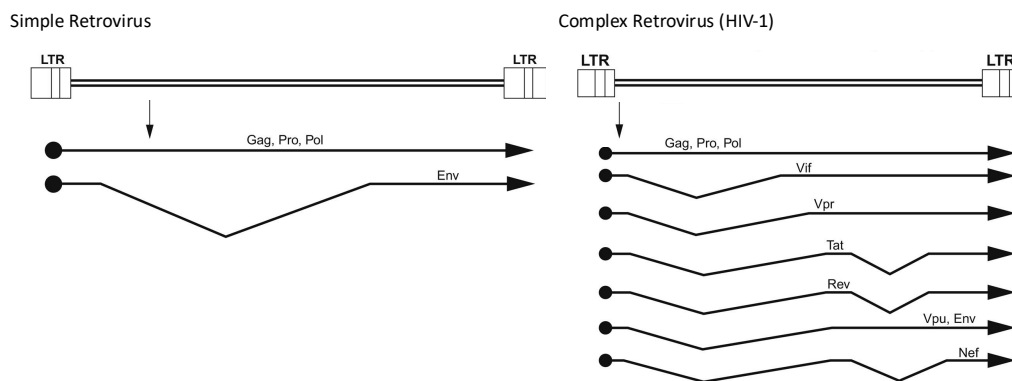


Figure 6: Splicing events in simple and complex retroviruses

Figure modified from MacLachlan & Dubovi, 2017.

1.3.6 Nuclear Export

Both unspliced and spliced RNAs must be exported out the nucleus into the cytoplasm to translate the viral proteins and ensuring packaging of the unspliced RNA (gRNA) into the assembling particles. HIV-1 and the other complex retroviruses encode an accessory protein which binds to the cis-acting elements that facilitates the nuclear export. In case of HIV-1, the Rev protein bind to the Rev responsive elements (RRE), a highly structured, conserved *cis*-acting sequence located in the envelope region consequently is present in the unspliced and in few spliced RNAs (Hadzopoulou-Cladaras et al., 1989; Malim et al., 1989). The Rev-RRE complex binds to the cellular nuclear export proteins CRM1 (XpoI) and Ran-GTP (reviewed in (Dayton, 2004; Neville et al., 1997)). In the absence of Rev protein, the unspliced and singly spliced RNA have been shown to accumulate in the nucleus (Malim et al., 1989). Similarly, in HTLV-1, the Rex protein binds to the RexRE located in the 3' UTR and hence present in all HTLV-1 RNAs (Grassmann et al.,

1991; Hanly et al., 1989) and facilitate the export of unspliced and singly spliced mRNAs (Nakano & Watanabe, 2012). In case of MPMV, a simple retrovirus, the constitutive transport elements (CTE) located near the 3' UTR acts as the *cis*-acting sequences for the viral RNA nuclear export by directly interacting with host nuclear export proteins (Bray et al., 1994).

1.3.7 Translation

It is becoming increasingly clear that the ribosome utilizes an internal ribosome entry site (IRES) located near upstream of Gag initiation codon in many retroviruses (Balvay et al., 2007). This cap-independent translation initiation has been well characterized in MLV (Berlioz & Darlix, 1995). Although use of IRES for the translation initiation in HIV-1 has been controversial, HIV-1, HIV-2 and SIV possess an IRES within the Gag codon resulting in the translation of N-terminally truncated Gag isoforms (Balvay et al., 2007).

Retroviruses encode Gag-Pol in the form of precursor polypeptide, which is processed by virally encoded protease to generate structural and enzymatic proteins. The major structural precursor polypeptide Gag is also expressed from the same ORF. Hence the expression of Gag and Gag-Pol is highly regulated in order to maintain Gag and Gag-Pol ratio. The Gag uses a stop codon within the Gag ORF but skipping this stop codon leads to the expression of Pol from its ORF resulting in Gag-Pol fusion protein. The first mechanism for such a skipping of Gag termination codon is referred as translational read-through (Balvay et al., 2007; Coffin et al., 1997). In MLV and feline leukemia virus, the amber termination codon (UAG) is occasionally mis-decoded as glutamine (CAG) by the tRNA (Yoshinaka et al.,

1985a, 1985b). In most other retroviruses, the frameshift mechanism exists, where ribosome slip one nucleotide backwards just before the Gag termination codon resulting in a change in ORF, consequently skipping the termination codon and continue translating Pol ORF only (Balvay et al., 2007; Coffin et al., 1997). The ribosomal frameshifting is based on two signal sequences; a hepta-nucleotide sequence, termed as slippery sequence of shift site followed by a higher order RNA structure called pseudoknot which is required for the efficient frameshifting (Chamorro et al., 1992). In general, the slippery sequence follows X XXY YYZ pattern (in case of HIV-1, the slippery sequence is U UUU UUA; Jacks et al., 1988). The structured pseudoknot presented in the form of a stem loop present downstream of the frameshift site results in ribosome stalling and aids the tRNA slippage (Balvay et al., 2007). In case of MMTV, FIV and simian retrovirus-1 (SRV-1), the pseudoknot assumes a more complex structure suggesting the mechanism is more complicated than just ribosome stalling (Coffin et al., 1997). The ribosomal frameshifting in HIV-1 results in expression of Gag and Gag-Pol in a ratio of 20:1 (i.e., 5% of translational events; Jacks et al., 1988). It has been shown that altering this ratio is detrimental to viral infectivity (Shehu-Xhilaga et al., 2001).

The Gag of most of the retroviruses undergoes a number of post-translational modifications (PTMs) which are critical for different stages of virus life cycle. The well-studied PTM in context of retroviral assembly is myristoylation (Göttlinger et al., 1989). Most of the retroviral Gag undergoes myristoylation except some including RSV, EIAV and FV. This PTM involves the addition of a saturated fatty acid (myristic acid) to the second amino acid glycine in the context of a consensus sequence. MA trimerization and membrane binding expose the myristic acid

(presented in a sequestered conformation in the MA), which is required for the anchoring of Gag to the plasma membrane (Bussienne et al., 2021; Coffin et al., 1997). The other common PTM is phosphorylation. The phosphorylation of a serine residue in HIV-1 MA has been shown to regulate the plasma membrane binding along with myristoylation (Yu et al., 1995). Ubiquitination and sumoylation are the other known PTMs of retroviral Gag proteins (Bussienne et al., 2021).

The envelope protein is synthesized from a singly spliced RNA and is processed in a same manner as for membrane and secretory proteins. In the endoplasmic reticulum (ER), the glycosylation occurs followed by the proteolytic removal of leader peptide. Within the ER, they are folded and oligomerized and then transported to Golgi, where the furin mediated cleavage results in the generation of SU and TM domains (Coffin et al., 1997).

1.3.8 Assembly and RNA Packaging

The specific selection of full length unspliced gRNA and assembly into the virus particle are highly intricate and coupled process during retroviral replication. The selective encapsidation of gRNA is explained in detail in a separate section (see Section 1.4). Gag precursor proteins are the building blocks of retroviral particles, and each virus particle is made up of approximately 2000-5000 Gag molecules (Briggs et al., 2004; Carlson et al., 2008). Gag alone is sufficient to make non-infectious virus like particle (VLPs) that can be successfully released out of the cell and are morphologically indistinguishable from the ~100 nm diameter immature virions produced by infected cells (Larson et al., 2005).

The type-C retroviruses assemble on the plasma membrane, facilitated by the highly basic region in the MA domain and the myristic acid (Ono et al., 2000). The hydrophobic and highly basic N-terminal region of MA is often referred to as membrane binding domain (M-domain). Gag is targeted to the membrane area which is rich in phosphoinositide phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂), sphingolipids and cholesterol termed 'lipid rafts'. Binding of MA to PtdIns(4,5)P₂ results in the insertion of myristic acid to the membrane that probably plays a role in anchoring Gag on the membrane (Freed, 2015; Ono et al., 2004; Saad et al., 2006). In case of B and D type retroviruses, assembly occurs intracytoplasmically and pre-assembled particles target to the plasma membrane for budding and release (Coffin et al., 1997). It has been shown that in MMTV, a single amino acid substitution in MA prevents membrane targeting and results in intracytoplasmic accumulation of virus particles (Zábranský et al., 2009; Zhang et al., 2015). In MPMV, the single amino acid mutation in MA region changes the assembly location from cytoplasm to plasma membrane (Rhee & Hunter, 1990, 1991). Thus, the role of MA in targeting the location of viral assembly is well established in retroviruses.

The region that involved in the Gag-Gag interaction is termed as interaction domain (I- domain). The major domain that participates in Gag multimerization is the C-terminal of CA domain (Freed, 2015). Additionally, it has been suggested that the basic residue in NC also plays a critical role in Gag multimerization (Dawson & Yu, 1998). Several reports have shown that the interaction of RNA with NC is required for efficient CA-CA interaction, probably by exposing the I-domains in CA (Burniston et al., 1999; Ott et al., 2009; Tanwar et al., 2017; Yang et al., 2018). It has

been shown that the HIV-1 Gag form oligomers on the gRNA, transported to the plasma membrane where the higher order multimerization occurs (Yang et al., 2018).

1.3.9 Release and Maturation

The third assembly domain which facilitates release of the fully assembled immature virions from the plasma membrane is referred late domains (L-domains). In HIV-1, the p6 domain is considered as L-domain as its deletion results in the accumulations of virus particles which are joined to the membrane by a stalk (Göttlinger et al., 1991). Within the p6 domain, the region required for release is located to a PTAP motif at the C-terminus. This motif binds to the host tumor susceptibility gene 101 (TSG101), a part of endosomal sorting complex required for transport I (ESCRT I) complex (Garrus et al., 2001). The other motif YPXL, located in the p6 domain interacts with Alix/AIP-1, a part of ESCRT II also participate in virion release, but the role of ESCRT I is dominant (Freed, 2015). Additionally, the Vpu protein of HIV-1 enhances the virus particle release from the membrane (Göttlinger et al., 1993). In contrast to HIV-1 L- domain, the L-domain of RSV is located near the N-terminus of the Gag and contains a PPPPY motif that interact with neuronal precursor cell-expressed developmentally downregulated 4 (Nedd4; a class of ubiquitin ligases) leading to the recruitment of TSG101 (Xiang et al., 1996).

Following release, the immature virions undergo proteolytic processing and extensive structural rearrangements referred to as maturation (Figure 7). The Gag alone or PR deficient Gag-Pol are able to produce immature non-infectious virions (Kato et al., 1985; Kohl et al., 1988). Retroviral proteases belongs to the category of aspartyl protease and possess the conserved active site Asp-Thr-Gly (DTG; in HIV-

1, HIV-2, SIV, FIV, MMTV) or Asp-Ser-Gly (DSG in RSV; Konvalinka et al., 2015; Wlodawer & Gustchina, 2000). In the immature virus particle, the Gag molecules are arranged radially, but following maturation, the CA protein reassembles to different shaped cores depends on the virus type. The full-length CA composed of an N-terminal domain (NTD) in the exterior of the core and C-terminal domain (CTD) faced towards interior of the core. The HIV-1 CA core acquires cone shape with approximately 250 CA hexamers and 12 pentamers (Zhao et al., 2013). The initiation of PR activation is highly regulated in retroviruses and occurs shortly after the virus budding from the cell. The functional PR is in its dimer form; however, in the Gag-Pol protein, the enzymatic activity is minimal, and PR undergoes self-cleavage at the N-terminus and folds into a stable dimeric form (Freed, 2015). However, the exact mechanisms that trigger PR activation after budding is unclear.

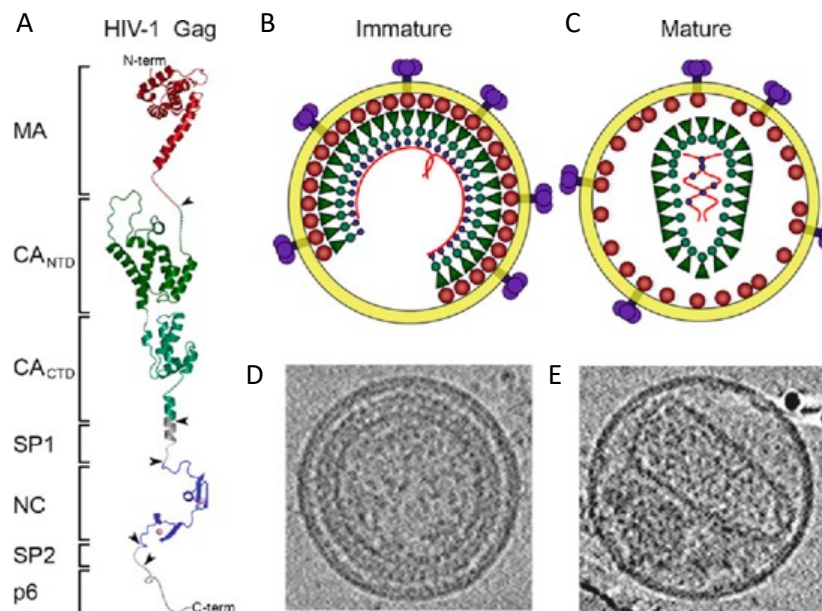


Figure 7: Organization of the immature and mature HIV-1 virions

(A) Schematic tertiary structural model of full-length HIV-1 Gag. Individual domains are in different colors and are labeled on the *left*. This color scheme is maintained throughout the chapter. (B) Schematic model of the immature virion. (C) Schematic model of the mature virion. Images of (D) immature and (E) mature virions preserved in vitreous ice. Figure and legend adapted from Ganser-Pornillos et al., 2012.

1.4 RNA Packaging in Retroviruses

Selective encapsidation of gRNA during the initial stages of retroviral particle assembly is a crucial step in retroviral life cycle. This selective packaging is mediated by the interaction between the Gag protein and the particular sequence(s) of the gRNA located at its 5' region termed "packaging signals" (ψ or *psi*). Despite the fact that gRNA packaging is a universal step in all retroviruses and the location of the packaging signal is conserved within the retroviral gRNA, no sequence conservation has been found between the packaging signals of different retroviruses (Ali et al., 2016; Comas-Garcia et al., 2016; D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007; Rein, 2019). Conventional and novel biochemical techniques have shown that packaging sequences of retroviruses assume higher order structures comprising of different structural motifs (Aktar et al., 2013, 2014; D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Paillart, Dettenhofer et al., 2004; Wilkinson et al., 2008).

1.4.1 RNA Binding in Context of Gag

The Gag precursor polyprotein is the basic building block of the virus assembly and accounts for 50% mass of the virus particle (Bell & Lever, 2013). Gag, in addition to being the major structural protein, plays a crucial role in selective packaging of the RNA genome (D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007).

The MA and NC domains of retroviral Gag has been shown to have the ability to bind nucleic acids. The identification of ψ in the gRNA is facilitated by the zinc finger domain of the NC of Gag polyprotein (D'Souza & Summers, 2005;

Johnson & Telesnitsky, 2010; Lu et al., 2011; Miyazaki et al., 2011). NC contains either one to two evolutionarily conserved, but distinct Cys-His boxes that can sequester Zn^{2+} ions, allowing for high affinity NC-gRNA interactions. The Cys-His boxes contain conserved CCHC arrays (C-X_n-C-X_n-H-X_n-C where C = Cys, H = His, X_n = n number of amino acids) that are variable among different retroviruses. Mutations of the zinc fingers drastically reduce gRNA packaging and production of non-infectious virions (Aldovini & Young, 1990; Wang & Barklis, 1993). In HIV-1, apart from zinc fingers of NC, the flanking regions of Cys-His boxes have also shown to be important for gRNA packaging (Cimarelli et al., 2000; Housset et al., 1993). Interestingly, it has also been shown that the 15 basic amino acids distributed throughout the NC region nonspecifically can bind to RNA resulting in the incorporation of cellular RNAs into the virion (Aronoff & Linial, 1991; Poon et al., 1996).

Many studies have shown that MA domain of several retroviruses also binds to RNA, although the precise role of this interaction in the viral life cycle remains largely unclear (Alfadhli & Barklis, 2014; Chukkapalli et al., 2010). It has been shown that the RNA-binding to MA was necessary for Gag oligomerization when two-thirds of NC was deleted from Gag (Burniston et al., 1999). Similarly, the basic amino acids in either of MA or NC is sufficient to package RNA without significantly reducing virion production. Moreover, in BLV, *in vitro* studies showed that the specific selection of gRNA is conferred by the MA protein and NC binds to non-specifically to any RNAs (Katoh et al., 1991; Wang et al., 2003).

In addition, the p6 domain of HIV-1 has also been shown to be important in selective encapsidation of gRNA and discrimination between gRNA and spliced viral

RNAs since its deletion results in reduced affinity to gRNA binding (Bernacchi et al., 2017; Dubois et al., 2018; Tanwar et al., 2017). In the prototype *Betaretovirus*, MPMV, a KKPKR sequence located in the pp24 domain of Pr78^{Gag} has been implicated to play a role in viral RNA packaging though the direct binding of pp24 to RNA has not yet been established (Bohl et al., 2005). It has been proposed that the presence of these basic residues enhances the RNA packaging mediated by NC. Cryo-electron tomography studies have shown that the RKK residues at C-terminus of MPMV CA has a potential role in nucleic acid binding probably not by directly interacting with RNA, rather stabilizing the Gag-viral RNA complex, or possibly due to structural arrangement of the NC region which is required for packaging (Füzik et al., 2016). Replacing the RSV NC domain with that of MLV or HIV-1 NC with that of MMTV NC, did not completely abrogated the RNA packaging efficiency, rather up to 50% of RNA packaging was retained (Dupraz & Spahr, 1992; Poon et al., 1996; Poon et al., 1998) insinuating the involvement of other Gag domains in gRNA packaging. These studies clearly demonstrate that the packaging of gRNA is an intricate and multifaceted phenomenon that occur in the context of the whole Gag polyprotein, especially in the case of HIV-1.

1.4.2 The Ψ RNA

The 5' end region of retroviral gRNA assumes higher order (secondary and tertiary) structures with several base-paired, unpaired nucleotides including stem loops and internal loops. Many of these structural motifs have been shown to be critical for various steps in the virus life cycle, including reverse transcription, transcriptional activation, splicing, translational regulation, dimerization and packaging. The packaging determinants for retroviruses are located within the 5' end

of the gRNA primarily between R and 5' end of the *gag* gene which is often referred to as Ψ and has been shown to interact with Gag (Ali et al., 2016; Comas-Garcia et al., 2016; D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007). In spliced RNAs, either the Ψ is spliced out (in case of HIV-1, the SL3 which is identified as a primary NC binding site; discussed elsewhere in this dissertation) or architecturally disrupted (Figure 8A; as in the case of MPMV; Pitchai et al., 2021). In the case of HIV-1, where an internal loop of SL1 which is present both in unspliced and spliced RNAs functions as primary Gag binding site. It has been shown that the downstream sequence of Ψ in unspliced RNA acts as positive regulator for packaging to compensate the negative regulatory effect exerted by sequences upstream of Ψ (Figure 8B; Abd El-Wahab et al., 2014). In case of HIV-2, in which the Ψ is located upstream of SD, a *cis*-acting co-translational packaging mechanism has been shown to accomplish gRNA packaging (Kaye & Lever, 1999).

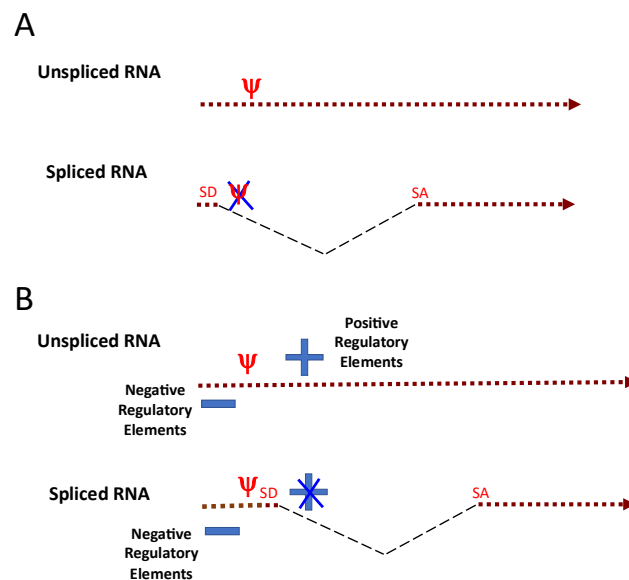


Figure 8: Different mechanisms of selective packaging of unspliced RNA

(A) The Ψ is spliced out/structurally disrupted in the spliced RNAs. (B) The positive regulatory elements present in unspliced RNAs enhance its packaging, while they are absent in spliced RNAs.

Using the biochemical approach selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) and nuclear magnetic resonance (NMR) studies elucidated the secondary structure of HIV-1 Ψ RNA. Two structural models have been demonstrated for the dimerization and encapsidation competent gRNA. They include the branched multiple hairpin and three-way junction models (Figure 9; Lu et al., 2011; Paillart, Shehu-Xhilaga et al., 2004; Wilkinson et al., 2008). In addition, multiple regions within the packaging determinants have been shown to be required for optimal gRNA packaging by Gag. For example, in the case of HIV-1, studies have shown that TAR, the Poly(A) hairpin, basal region of PBS, internal loop within the SL1 and the SL3 apical loop are required for its efficient packaging by Pr55^{Gag} (Clever et al., 2002; Didierlaurent et al., 2011; Helga-Maria et al., 1999). However, within the 5' end region of the gRNA, the core encapsidation signals/determinants (CES) were initially mapped to be within a stretch of 159 nts harboring SL1- SL3 and a few nucleotides of *gag*, but excluded the role of TAR, poly(A) and PBS in RNA packaging (Heng et al., 2012). A later study showed that the poly(A) loop is critically required for the efficient packaging; but the direct interaction with full-length HIV-1 Gag could not be identified (Smyth et al., 2015, 2018). Hence it is possible that the other regions within the RNA are involved in recruiting Gag or stabilizing the Gag-gRNA interaction, rather than acting as direct specific binding sites. Furthermore, in HIV-1, several studies have shown that the purine rich regions, particularly guanosines (G) are involved in interaction with Gag (Figure 9B; Abd El-Wahab et al., 2014; Comas-Garcia et al., 2018; Keane & Summers, 2016; Nikolaitchik et al., 2020).

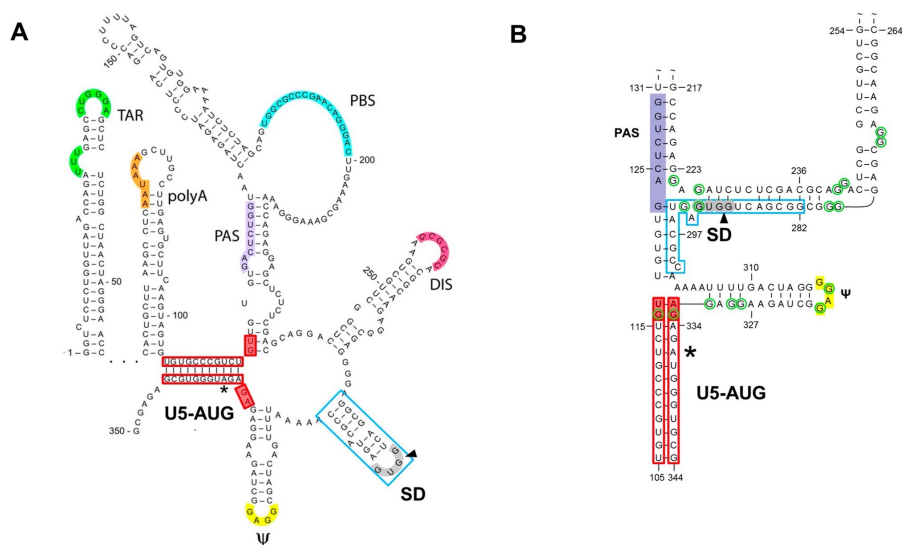


Figure 9: The two proposed structures for HIV-1 packaging signal RNA

(A) The branched multiple hairpin (BMH) structure of the entire packaging signal RNA and (B) the three-way junction structure of the core encapsidation signal with 17 ‘G’ residues that have been implicated in NC binding by different studies are marked with green circles. Figure adapted and modified from Mueller et al., 2016.

A GGAG tetraloop of the SL3, which is present downstream of mSD has been shown to bind to NC domain (De Guzman et al., 1998; Keane et al., 2015; Kenyon et al., 2015). Additionally, a recent study using NMR spectroscopy has identified a weakly base-paired [UUUU]:[GGAG] helix in SL3 stem as the NC binding site (Ding et al., 2020). However, these studies used only the NC domain of the protein and certain other *in vitro* Gag-gRNA interaction studies were performed with Gag Δ p6 (Comas-Garcia et al., 2017; Webb et al., 2013). Therefore, it is reasonable to suggest that the specificity and number of the nucleotide interaction could vary in the full-length structural/conformational context of Gag (Rein, 2019). It was noted that in MLV, full length Gag binds more selectively than NC to the CES (Gherghe et al., 2010). Using HIV-1 full length Pr55^{Gag}, results of *in vitro* binding and footprinting experiments have shown that SL1 is the primary Pr55^{Gag}-binding site and the purine rich internal loop (G//AGG) in SL1 is a key determinant for

Pr55^{Gag} interaction during RNA packaging (Abd El-Wahab et al., 2014; Bernacchi et al., 2017). Similar to HIV-1, in HIV-2, a 5' GGRG 3' motif located upstream of dimerization initiation site (DIS) was found to be important for RNA packaging and has been suggested as Gag binding site (Baig et al., 2009). A recent study also recognized same residues (GGAG motif) as the primary nucleotides for packaging in HIV-2 (Umunnakwe et al., 2021). The preference of NC to bind to G residues has also been demonstrated in HTLV-1 by UV crosslinking experiment. But later analysis of protection by bound NC from the SHAPE reagent showed that only 4 'G' residues among the total 9 being protected (Wu et al., 2018).

In case of the prototypic simple retrovirus, MLV, the core packaging determinants have been mapped to a 101 nt region comprising three stem loops (D'Souza & Summers, 2004; Mann et al., 1983; Miyazaki, Garcia et al., 2010; Mougel & Barklis, 1997). Structural studies using the NC domain of MLV Gag and the CES RNA identified a UCUG motif which is base paired in monomeric conformation of RNA and unpair following dimerization. Moreover, the interaction of 'G' residue within this UCUG motif with the zinc knuckle has been well-characterized (Figure 10; D'Souza & Summers, 2004; Miyazaki, Irobalieva et al., 2010). Footprinting experiments using MLV NC and Gag which required using a longer RNA compared to the NMR study; identified that two UCUG-UR-UCUG motifs both are needed for high affinity NC binding (Gherghe et al., 2010). Interestingly, within these UCUG-UR-UCUG motifs, the 'G's have been shown to be crucial for *in vitro* binding (Gherghe et al., 2010). In RSV, another simple retrovirus, only an 82 nt long region in the 5' UTR and no sequences in *gag* has been implicated in gRNA packaging. The NMR analysis revealed a small four-way

junction structure and the N-terminal zinc knuckle of NC preferentially interacts with UGCG tetraloop with the second ‘G’ making an extensive contact with NC, while the interaction of C-terminal zinc knuckle with an ‘A’ is also required for high affinity binding (Zhou et al., 2005, 2007).

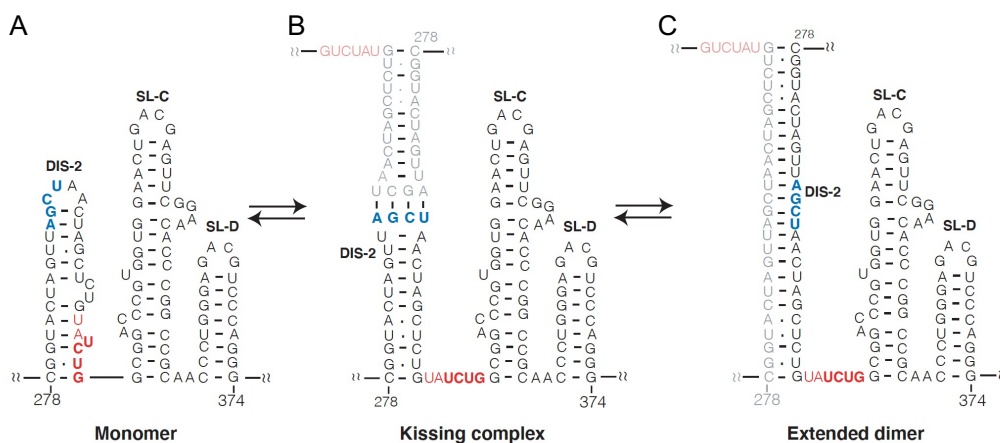


Figure 10: Stem-loops in MLV ψ

(A) The secondary structure of MLV bases 278–374 in monomeric form. (B) The kissing complex, in which DIS-2 has undergone a register shift, exposing the palindrome AGCU in the loop (blue). DIS-2 from two monomers is shown with intermolecular base-pairing between the loops. Note that the register shift has also exposed bases 304–309 (red). (C) The extended dimer, in which the stems, as well as the loops, of the DIS-2 elements of the two monomers are base-paired intermolecularly. Figure and legend adapted from Rein, 2004.

1.4.3 Ψ RNA Acts as the Nucleation Point for Gag Multimerization

The retroviral Gag proteins have the ability to assemble and form virus like particles even in the absence of ψ -containing viral gRNA. The specific encapsidation from large excess of cellular and spliced viral RNAs is achieved by the ψ present on the viral gRNA. Several studies suggested that at physiological salt concentrations, HIV-1 Gag binds to both ψ -containing and non- ψ RNAs with almost equal affinity, this includes both specific and nonspecific interactions. However, the nonspecific

binding is significantly reduced when the ionic strength of the experimental conditions are increased, thereby identifying on the specific-high affinity binding site (Abd El-Wahab et al., 2014; Comas-Garcia et al., 2017; Pitchai et al., 2021; Rein, 2019). Hence it was suggested that the Gag binds to all the RNAs, but the nucleation and assembly following Gag-RNA binding occurs more rapidly on ψ -containing RNAs than nonspecific RNAs, probably in a two-stage process (Rein, 2019).

A study based on single virion analysis showed that when Gag is expressed in cells at levels similar to those in cells containing one provirus, the presence of HIV-1 RNA greatly enhances the viral particle production. However, when Gag is overexpressed, this RNA-specific enhancement disappears, although the packaging of other non-specific RNAs has not been mentioned in this study (Dilley et al., 2017). As previously observed, this study also underlines the fact that RNA with specific packaging signal act as the “nucleation point” for Gag assembly (Dilley et al., 2017). A later study based on single molecule localization microscopy proposed a simplistic model of gRNA interaction with the Gag polyprotein, preceding to Gag multimerization and assembly at the plasma membrane (Yang et al., 2018). It has been shown that the assembly process starts in the cytoplasm, where a small number of Gag molecules interacts with the gRNA, which is then trafficked to the plasma membrane, the place at which higher order multimerization of Gag takes place. The study clearly demonstrated that the initial interaction with gRNA is the key step for entire assembly process since the Δ NC-Gag freely migrate rather than clustering (Yang et al., 2018). This result is previously reported by other group, were the deletion of NC resulted in a delayed particle production, without any morphological differences. Interestingly the delayed particle production is explained by the weaker

affinity nucleic acid interacting capacity of MA (Ott et al., 2009). Thus, Gag-RNA interactions play an essential role in Gag assembly, using RNA as the nucleation point.

1.4.4 Nucleus as the Initial Site of gRNA Selection

The intracellular location where the initial interaction of gRNA with Gag occurs remains a fundamental question in the retroviral life cycle. It has been considered for a long time that this interaction occurs either in the cytoplasm or at the plasma membrane. Many retroviral Gag proteins, including RSV, HIV-1, MMTV, MPMV, FIV and MLV have been identified in the nucleus, raising the possibility that the initial selection of gRNA occurs in the nucleus. Rous sarcoma virus has been considered as the prototypic model for studying the mechanisms and implications of Gag nuclear trafficking. It has been shown that the nuclear trafficking of RSV Gag protein plays an important role in the efficient packaging of RSV RNA. Recently, it has also been established that within the nucleus, the site of active viral RNA transcription is the place where initial interaction between gRNA and RSV Gag takes place (Garbitt-Hirst et al., 2009; Maldonado et al., 2020). However, retroviruses in general are poorly studied towards understanding the role of nuclear shuttling of Gag protein in their RNA packaging.

Despite the fact that several studies have explored the nuclear trafficking of HIV-1 Pr55^{Gag}, the precise location where the initial interaction between HIV-1 Gag and gRNA occurs still remains largely controversial. Some studies have shown that the nuclear localization signal (NLS) of HIV-1 is a part of MA domain of Pr55^{Gag}, the same domain is involved in recruiting Gag-RNA complex into the plasma

membrane (Bukrinsky et al., 1993; Kräusslich & Welker, 1996). Deletion of NLS has been shown to result in the accumulation of gRNA in the nucleus and reduced RNA packaging efficiency alluding that the initial Gag-RNA interaction commences in the nucleus (Dupont et al., 1999). However, live cell imaging studies have shown that HIV-1 Gag-RNA complex is not associated with nuclear envelope, while another study observed the co-localization of RNA and Gag at the perinuclear region (Kemler et al., 2010; Poole et al., 2005). A recent study proposes that the HIV-1 Gag interacts with newly transcribed RNA leading to the formation of ribonucleoprotein (RNP) complex in the nucleus. However, this study does not confirm that nuclear Gag-RNA RNP complex formation is necessary for RNA packaging (Tuffy et al., 2020).

1.4.5 Role of Nuclear Export Pathway in RNA Packaging

Binding of Rev protein to the RRE is a critical requirement for the nuclear export of the unspliced and partially spliced viral RNAs through CRM1 export pathway. In addition to its nuclear export function, RRE has also been suggested to play an important role in RNA packaging (Brandt et al., 2007; Grewe et al., 2012; Kharytonchyk et al., 2018). Additionally, binding of Gag to the RRE located in the 3' UTR has also been demonstrated (Kutluay et al., 2014), although a role of this interaction in packaging still remains elusive. It was suggested that the Rev-RRE interaction is not directly involved in RNA packaging, rather indirectly by influencing the nuclear export pathway (Liu et al., 2017). Interestingly, replacement of RRE with the constitutive transport elements (CTE) of MPMV, that mediates the nuclear export via NXF1 pathway, can efficiently complement the nuclear export function of Rev-RRE (Blissenbach et al., 2010; McBride et al., 1997). However, a

recent study showed that the RRE-containing RNAs packaged more efficiently when compared to CTE-containing RNAs in a competitive experimental set up (Kharytonchyk et al., 2018). Hence further studies are required to elucidate the precise role of Rev-RRE system in the retroviral RNA encapsidation.

1.4.6 Role of Assembly Intermediates in gRNA Packaging

Studies pertaining to the interaction of unspliced gRNA and HIV-1 Gag in the assembly intermediates have been carried out using proximity ligation assays. These assays have demonstrated that the Gag-RNA complex is present only in the second assembly intermediate, suggesting that association of HIV-1 Gag and unspliced HIV-1 RNA occurs within a host RNA granule and may act as assembly precursors (Barajas et al., 2018). These granules also contain the cellular proteins ABCE1 and DDX6 and non-translating mRNAs. However further studies are required to confirm whether the host-derived RNA granules are the initial packaging complexes or not. Also, it is not clear that whether Gag makes direct contact with gRNA in these complexes. Similar to HIV-1, assembly intermediates derived from ABCE1 and DDX6 has also been detected in the case of FIV (Reed et al., 2018). Another host protein Staufen1 was identified to interact with HIV-1 Gag and found to colocalize in the viral RNP complex suggesting its role during HIV-1 gRNA encapsidation (Abrahamyan et al., 2010; Cochrane et al., 2006; Mouland et al., 2000). Altogether these data indicate that the host derived factors in the RNP complex play some roles in encapsidation and assembly during retroviral life cycle.

1.5 Retroviral RNA Dimerization

The retroviral gRNA dimerization is an essential step during the retroviral life cycle during which gRNAs are packaged as dimers, non-covalently joined through their 5' ends. The first observation of existence of a retroviral RNA in a dimer form in the virus particle has been made in RSV (Canaani et al., 1973; Mangel et al., 1974). The dimerization is initiated by a palindromic sequence known as the dimerization initiation site (DIS) which allows interaction between 5' end of the two RNA genomes (Skripkin et al., 1994). Following initial kissing-loop interaction mediated by the DIS, a dimer linkage structure (DLS) conformation evolves involving the base pairing of 50 to a few hundred nucleotides (Figure 11; Murti et al., 1981). The RNA dimers that are formed soon after the formation of kissing-loop complex are often considered as weak dimers. The dimeric RNA isolated from PR(-) HIV-1 particles have lower thermal stability and different conformation, suggesting that further stabilization of the kissing-loop complex was achieved by the nucleic acid chaperone activity of the NC, which is generated as a result of proteolytic maturation of Gag (Feng et al., 1996; Fu et al., 1994). *In vitro* dimerization studies and the analysis of virion RNAs have been widely used to understand the mechanisms involved during dimerization. Several studies showed that the *in vitro* derived RNA dimers have similar characteristics with those derived from virions, hence making *in vitro* dimerization as a reliable and simple technique to study retroviral dimerization (discussed in Marquet et al., 1991; Paillart, Shehu-Xhilaga et al., 2004). It has been shown that the presence of Magnesium ions (Mg^{2+}), Gag or NC strongly influence the dimer formation *in vitro* (Dubois et al., 2018; Huthoff & Berkhout, 2002; Marquet et al., 1991; Paillart, Shehu-Xhilaga et al., 2004; Rein, 2010).

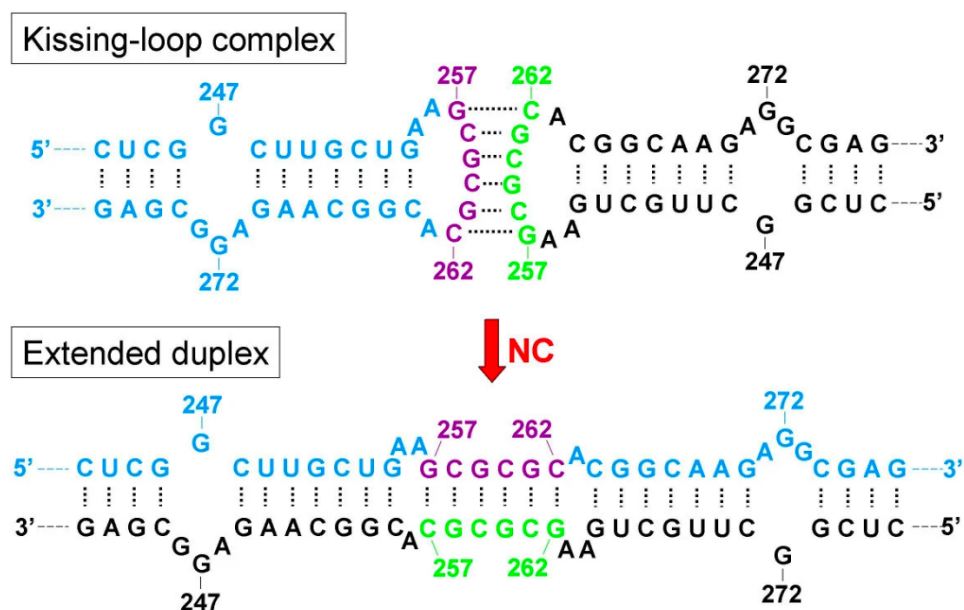


Figure 11: The Kissing-Loop Model of HIV-1 RNA Dimerization

HIV-1 RNA dimerization is initiated by a Watson-Crick base-pairing interaction between two palindromes in the loops of SL1 on two monomeric genomic RNAs. This interaction forms the loose unstable kissing-loop complex. Coincident with virus particle maturation, this unstable dimer is rearranged to form a more stable extended duplex that involves a mechanism whereby the base-pairs in the stems melt and then re-anneal to their complementary sequences on the opposite strand. Nucleotides and numbering correspond to the HIV-1 HXB2 sequence. Figure and legend adapted from Russell et al., 2004.

In case of HIV-1, MPMV and MMTV, a 6 nucleotide GC rich region has been identified which functions as the DIS, while in FIV, a 10 nts long palindromic sequence acts as the DIS (Aktar et al., 2013, 2014; Clever et al., 1996; Kenyon et al., 2011; Muriaux et al., 1995). In the case of MLV, multiple stem loops are involved in the dimer initiation (Ly & Parslow, 2002). It has also been shown that non-palindromic tetranucleotides located in the apical loop of two different stem loops interacts with each other and contribute to dimerization initiation in MLV (Kim & Tinoco, 2000; Konings et al., 1992). Two 13-nt and 14-nt long imperfect palindromes have been identified as the DIS in the case of HTLV-1 (Greatorex et al., 1996; Monie et al., 2001; Wu et al., 2018), however the exact mechanism of the

dimerization initiation is still unknown for HTLV-1. A recent study showed that a bipartite DIS (one is a 6-nt GC rich palindrome while the other is not a palindrome) in RSV 5' UTR, which are functionally redundant and neither of these region are necessary for virus replication in cell culture (Liu et al., 2020) a scenario similar to HIV-1 in which the DIS has been shown to be dispensable for the virus replication in PBMCs (Hill et al., 2003), suggesting the role of other elements during dimerization.

The process of DIS mediated dimerization and gRNA encapsidation are highly interlinked as evident by defective packaging of DIS mutated clones (Laughrea et al., 1997). In HIV-1, the gRNA containing additional DLS has been shown to be successfully packaged as monomer, indicating interaction between two DLS is important for packaging; irrespective of whether it is intra-or intermolecular interaction (Sakuragi et al., 2001). The proximity or overlap between the DIS and the packaging sequences also suggest the existence of an interplay between dimerization and packaging of gRNA (Paillart, Shehu-Xhilaga et al., 2004; Sakuragi et al., 2003). The exact mechanism involved in this selective encapsidation of RNA dimer by Gag is remains unknown, however it is assumed that following dimerization, conformational changes result in the proper exposure of the Gag binding sites within the gRNA which are otherwise remains obscured in the monomeric RNAs. This scenario is best understood in case of MLV, in which the dimerization induces a conformational change that exposes the high affinity NC binding UCUG motifs which are base-paired in monomeric RNA state (D'Souza & Summers, 2005). Ensuring the encapsidation of RNA in a homodimer state has several implications in the retroviral life cycle. For example: i) dimeric RNA greatly enhances the reverse transcription rate (Parent et al., 2000), ii) maintain the functional genome during the

replication by surpassing the single-strand nicks or RNA damage by template switching (Onafuwa-Nuga & Telesnitsky, 2009; Toyoshima et al., 1980) and iii) recombination resulting from the co-packaging of heterodimers (of two different genotypes; in case of HIV-1) leads to genetic variabilities that contributes to increased replicative fitness and drug resistance (Konings et al., 2006; Kozal et al., 1996; Njai et al., 2006).

In HIV-1, it has been shown that an RNA-conformational switch exists that regulates the fate of the gRNA, whether to undergo dimerization and packaging or translation. *In vitro* experiments revealed the existence of two structures with functional implications: the long-distance interaction (LDI) and the branched multiple hairpin (BMH; Figure 12; Abbink et al., 2005; Abbink & Berkhout, 2003). The LDI conformation involves the base-pairing of sequences several hundred nucleotides apart. The DIS in this conformation is base-paired with the poly(A) region, thus not well-suited for dimerization. On the other hand, in the BMH confirmation, several stem loops exist including the SL1 (containing the DIS) and SL3 (containing NC binding motifs) thus facilitating both dimerization and packaging (Figure 12). Structural study based on NMR on short RNAs demonstrated either Gag AUG or DIS base-pairs with U5; the U5: Gag AUG interaction in the three-way junction structure favors the dimerization competent structure like in BMH; except the existence of SL2 (Keane et al., 2015; Lu et al., 2011). Since U5 interacts with Gag start codon in BMH conformation, it was initially proposed that this structure only supports dimerization and packaging and do not serve as translating mRNA (Huthoff & Berkhout, 2001; Lu et al., 2011).

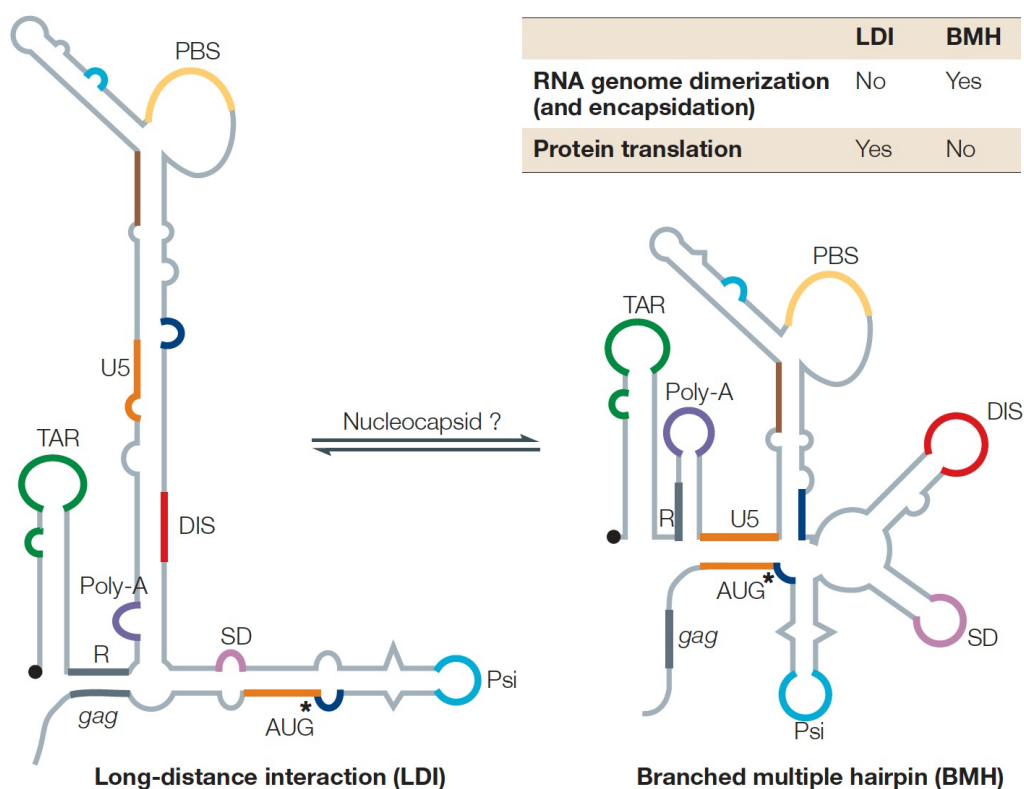


Figure 12: Model showing the putative conformational switch that is proposed to regulate translation and packaging of the HIV-1 genomic RNA

The long-distance interaction (LDI) secondary structure is proposed to be the translation-competent form, whereas a conformational change to the branched multiple hairpin (BMH) secondary structure would allow the genome to be encapsidated through interactions between its Psi and dimerization initiation site (DIS) domains and Gag (mainly its nucleocapsid domain). The regulatory motifs are shown in different colours. PBS, primer-binding site; SD, splice-donor site. Figure and legend Adapted from Paillart, Shehu-Xhilaga et al., 2004.

However, a study using the RNAs with either LDI or BMH stabilized structures did not alter the level of translation; on the other hand, shifting equilibrium towards LDI resulted in lower dimer yields. Hence it was proposed that two different pools of unspliced RNA exist with two functional implications; that are one acts as mRNA and other as gRNA for packaging (Abbink et al., 2005). Consistent with this hypothesis, recently, study using NMR showed that two pools of unspliced RNA exists, based on the transcription initiation site. If the transcript has only one G at the 5' end, the RNA undergoes packaging and translation is inhibited by adopting a

structure in which the 5' cap is sequestered from eIF4E interaction. On the other hand, if either 2 or 3 'G's, the RNA used as mRNA (Brown et al., 2020). The theory of heterogeneity in the RNA transcripts further supported by the results that shows HIV-1 Gag preferentially encapsidate non-translating RNA into the assembling virion (Figure 13; Chen et al., 2020).

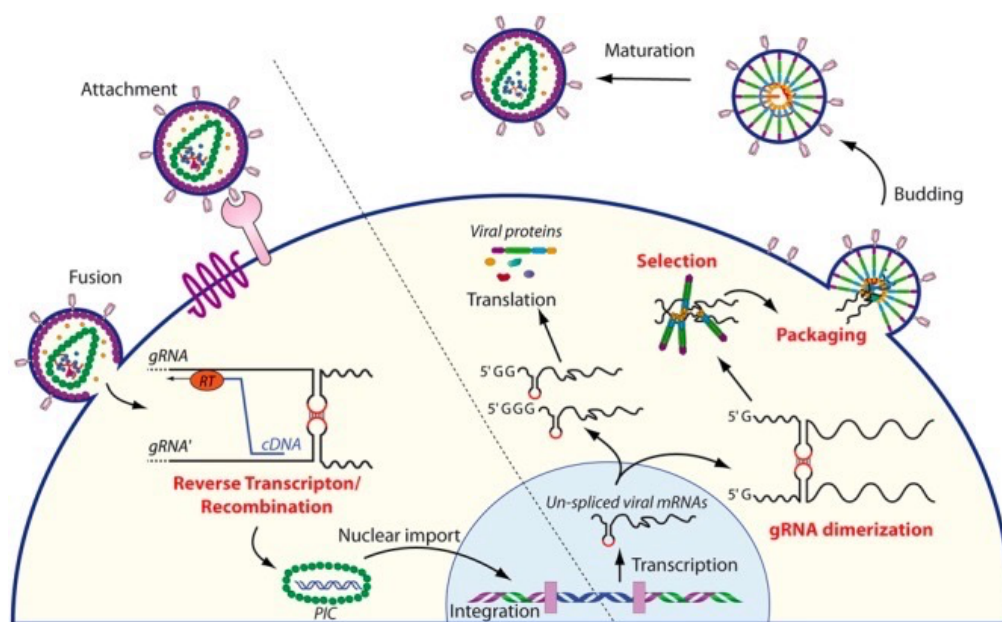


Figure 13: Schematic overview of the role of RNA dimerization in the retroviral life cycle

The cycle begins with the entry of the retrovirus within the target cell, followed by reverse transcription of the RNA genome into cDNA. During this step, gRNA dimerization plays an important role since RT may switch between strands, thus allowing genome repair and/or shuffling. The pre-integration complex (PIC) is then translocated into the nucleus where it is integrated in the genome of the target cell. The unspliced mRNAs are transcribed by the host machinery from the integrated provirus and transported to the cytoplasm. There, the single 5' capped mRNAs serve as genomic RNAs that dimerize and are subsequently selected and packaged into the nascent virions, while mRNAs beginning with two or three guanines are translated by the host machinery. After budding, immature particles follow a maturation step initiated by the viral protease to produce infectious virions (Figure and legend adapted from Dubois et al., 2018).

1.6 Mouse Mammary Tumor Virus

Mouse mammary tumor virus, member of *Betaretrovirus*, is the causative agent of mammary gland tumors in mice, which is transmitted through breast milk from infected mothers to pups (Bittner, 1936). It can be transmitted vertically through the germline as endogenous viruses (Bittner, 1936; Cardiff & Kenney, 2007; Varmus et al., 1972) MMTV has been widely used as a model for studying breast cancer in humans (Callahan & Smith, 2008). MMTV genome is approximately 9 kb and like any other retrovirus it encompasses 5' and 3' UTRs and encodes Gag, Pro, Pol and Env.

1.6.1 The Pr77^{Gag} Precursor Polyprotein

The precursor polypeptide Pr77^{Gag} which is encoded by the *gag* gene and processed by the viral protease into mature proteins NH₂-p10 (matrix), pp21, p3, p8, pn, p27 (capsid), p14 (nucleocapsid)-COOH (n represents 17 amino acids predicted by the DNA sequence but not identified among the purified proteins and peptides (Hizi et al., 1989). The domains pp21 and p3 are considered as the 'late domains' as these are crucial for particle release, while the p8 and pn are involved in maintaining the morphology of the virus particles (Zábranský et al., 2010). The MA domain of Gag contain the cytoplasmic targeting and retention signal (CTR) which is necessary for intracytoplasmic assembly of virus particles. A single amino acid substitution in this CTR has been shown resulted in plasma membrane localization of Gag (Zhang et al., 2015). Like other retroviruses MMTV MA domain also undergoes myristoylation which is necessary for targeting the assembled virus particles to plasma membrane for release. It has been shown that mutation of N-terminal glycine of MA leads to the

blocking of myristoylation and resulted in intracytoplasmic accumulation of virus particles (Zábranský et al., 2009; Zhang et al., 2015).

NC domain of Pr77^{Gag} possess two conserved and highly basic CCHC zinc finger domains for nucleic acid binding (Klein et al., 2000). Based on NMR studies it has been showed that the structure of MMTV distal zinc finger knuckle is significantly different from that of Mason-Pfizer monkey virus (MPMV), another member of *Betaretrovirus* genus (Klein et al., 2008). Interestingly, the proximal zinc finger domains of MMTV folds very similarly to those of HIV-1 and MLV (Klein et al., 2008). An interesting study by *Poon et al* showed that chimeric HIV-1 in which either entire NC is replaced by MMTV NC or only the two zinc finger domains are replaced, still able to package MMTV RNA with high specificity (Poon et al., 1998). Surprisingly, unlike HIV-1, the role of other domains in specific selection and packaging of MMTV gRNA has not yet been studied.

1.6.2 Two Frameshifts are Required for Maintaining Ratio of MMTV Gag, Gag-Pro and Gag-Pro-Pol

The three overlapping open reading frames in the MMTV unspliced RNA encodes for Gag, Pro and Pol respectively. The expression of the Pro and Pol are regulated by the frameshift suppression of termination codons. Two -1 frameshift signals are present within the each overlapping regions in order to facilitate the expression of Gag (Pr⁷⁷), Gag-Pro (Pr¹¹⁰) and Gag-Pro-Pol (Pr¹⁶⁰) polyproteins and translated in ratio of 30:10:1 respectively (Hizi et al., 1987; Jacks et al., 1987). A heptanucleotide sequence, termed as slippery sequence followed by a higher order RNA structure known as pseudoknot is required for the efficient frameshifting process (Chamorro et al., 1992). As mentioned in replication section, in general, the

slippery sequence follows X XXY YYZ pattern. In case of MMTV, for the first frameshifting between *gag* and *pro*, the slippery sequence is A AAA AAC, however the proposed second slippery sequence for frame shift between *pro* and *pol* does not follow this pattern, instead it is a tetranucleotide sequence, UUUA (Figure 14; Jacks et al., 1987). The MMTV pseudoknot acquires a more complex structure when compared to the simple stem-loop in the case of HIV-1 (Chamorro et al., 1992; Coffin et al., 1997). Initially it was suggested that the ribosome stalling at the pseudoknot aids the tRNA slippage but due to the complex structure of the pseudoknot, it was later proposed that the unpaired A in the apical loop of pseudoknot contributes to the frameshifting as it reduces the coaxial stacking of helices (Chen et al., 1995; Coffin et al., 1997; Jacks et al., 1987).

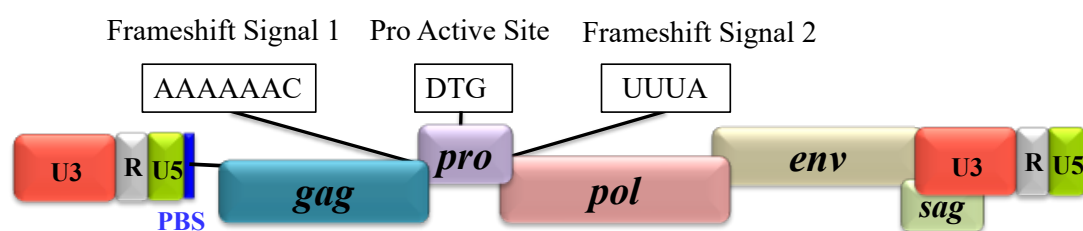


Figure 14: Schematic illustration of the *gag*, *pro* and *pol* genes in MMTV provirus

1.6.3 MMTV Envelope (Env) Protein

Like other retroviruses, MMTV Env processed by cellular furin that cleaves the 73 kD polyprotein to SU (52 kD) and TM domains (36 kD). The SU peptide binds to the transferrin receptor and following receptor mediated endocytosis, TM domain facilitates the membrane fusion within the acidic endosomes (Ross et al., 2002). It has been shown that an immunoreceptor tyrosine activation motif (ITAM)

within the Env is involved in transforming the mammary epithelial cells and thus suggests a role of Env in tumorigenesis (Ross et al., 2006).

1.6.4 Accessory Proteins Encoded by MMTV Genome

In addition to structural and enzymatic proteins, MMTV genome also encodes at least three accessory proteins (Figure 15), one of which is expressed from the full-length unspliced RNA, the *trans*-frame protein p30 or functionally it is called deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase), an enzyme that maintains low dUTP/dTTP ratio in the infected cell (Hizi & Herzig, 2015; Köppe et al., 1994). The other two proteins, superantigen (Sag) and regulator of export of MMTV (Rem) are expressed from spliced RNAs. The major portion of *sag* gene is located in the U3 region and is expressed from two transcripts. One uses the promoter from the 5' LTR, while the second transcript is initiated from a promoter located in the envelope. The Sag protein plays an important role in the dissemination of MMTV during infection (Reuss & Coffin, 1995; Ross, 2010; Xu et al., 1997). MMTV also encodes a protein called Rem, which is encoded from a doubly spliced mRNA and has functions similar to Rev in HIV-1 in the nuclear export of viral transcripts to the cytoplasm by binding to the Rem responsive elements (RemRE). Though MMTV has all along been considered as a simple retrovirus but now after the identification of the aforementioned Rem and RemRE regulatory pathway analogous to HIV-1 Rev/RRE pathway it has been suggested that MMTV be categorized as a complex retrovirus (Indik et al., 2005; Mertz et al., 2005, Mertz, Chadee et al., 2009; Mertz, Lozano et al., 2009; Müllner et al., 2008).

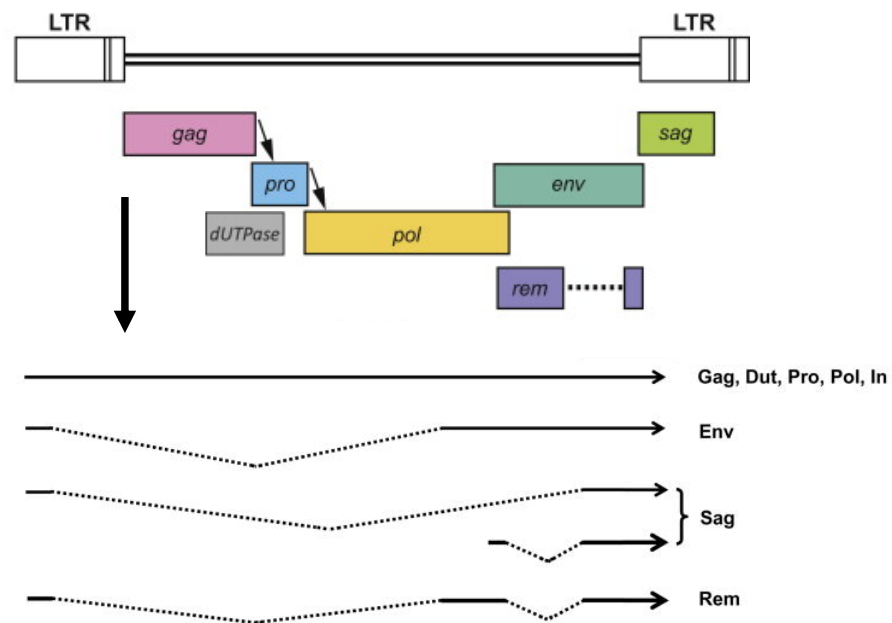


Figure 15: Coding regions in the MMTV proviral genome and transcripts

The arrows between the *gag-pro* and *pro-pol* reading frames indicate ribosomal frameshift signals. Figure modified from King et al., 2012; Ross, 2008.

1.6.5 MMTV Pathogenesis

The MMTV transmitted through milk, carried by dendritic cells (DC) to Peyer's patches, where they infect the T and B-lymphocytes. The presentation of viral protein Sag on MHC-II of DCs and B cells results in the rapid proliferation of T-cells which is necessary for establishing the reservoir of infection as well as the spreading of virus to mammary gland. The MMTV infection of mammary epithelial cells occurs only when they start to divide and high titers of virus are produced during lactation and pregnancy as a result of hormone inducible promoter in the MMTV 3' LTR (Ross, 2008, 2010). MMTV is a non-acutely transforming virus and often oncogenesis is associated with insertional activation of oncogenes; the most identified oncogenes include members of *wnt*, *fgf*, *notch* and *rspo* families (Ross, 2008, 2010).

1.6.6 Genomic RNA Packaging in MMTV

Unlike HIV-1, which is studied most extensively in terms of its RNA packaging, little is known about the molecular mechanisms of MMTV gRNA packaging. During assembly of MMTV particles, the viral Pr77^{Gag} protein must specifically select the viral gRNA from a variety of cellular and viral spliced RNAs. The precise mechanism(s) by which Pr77^{Gag} accomplishes this specific selection is yet to be established. For example, it remains unclear whether the binding of Pr77^{Gag} to gRNA is based on an intrinsic capability of the polyprotein that allows discrimination between gRNA and its spliced variants, or other steps in retroviral life cycle such as the nucleo-cytoplasmic transport and cellular compartmentalization of gRNA are also involved in discriminating gRNA during encapsidation process, as has been suggested for HIV-1 (Barajas et al., 2018; Becker & Sherer, 2017; Behrens et al., 2017; Brandt et al., 2007; Jouvenet et al., 2011; Moore et al., 2009). Recently, employing an *in vivo* packaging and transduction assay (Rizvi et al., 2009), it has shown that the 5' UTR and the first 120 nts of the *gag* gene are required for efficient MMTV gRNA packaging and propagation (Mustafa et al., 2012). To establish the structural basis of MMTV gRNA packaging, these sequences were folded using minimum free energy algorithm programs like Mfold and RNAstructure (Reuter & Mathews, 2010; Zuker, 2003). The folding predictions of these sequences revealed a higher order structure comprising of several structural motifs, which could be involved during MMTV gRNA packaging (Aktar et al., 2014). Later, this structure was validated by SHAPE (selective 2' hydroxyl acylation analyzed by primer extension) and the structure function relationship of various structural motifs during MMTV gRNA packaging was established (Figure 16; Aktar et al., 2014).

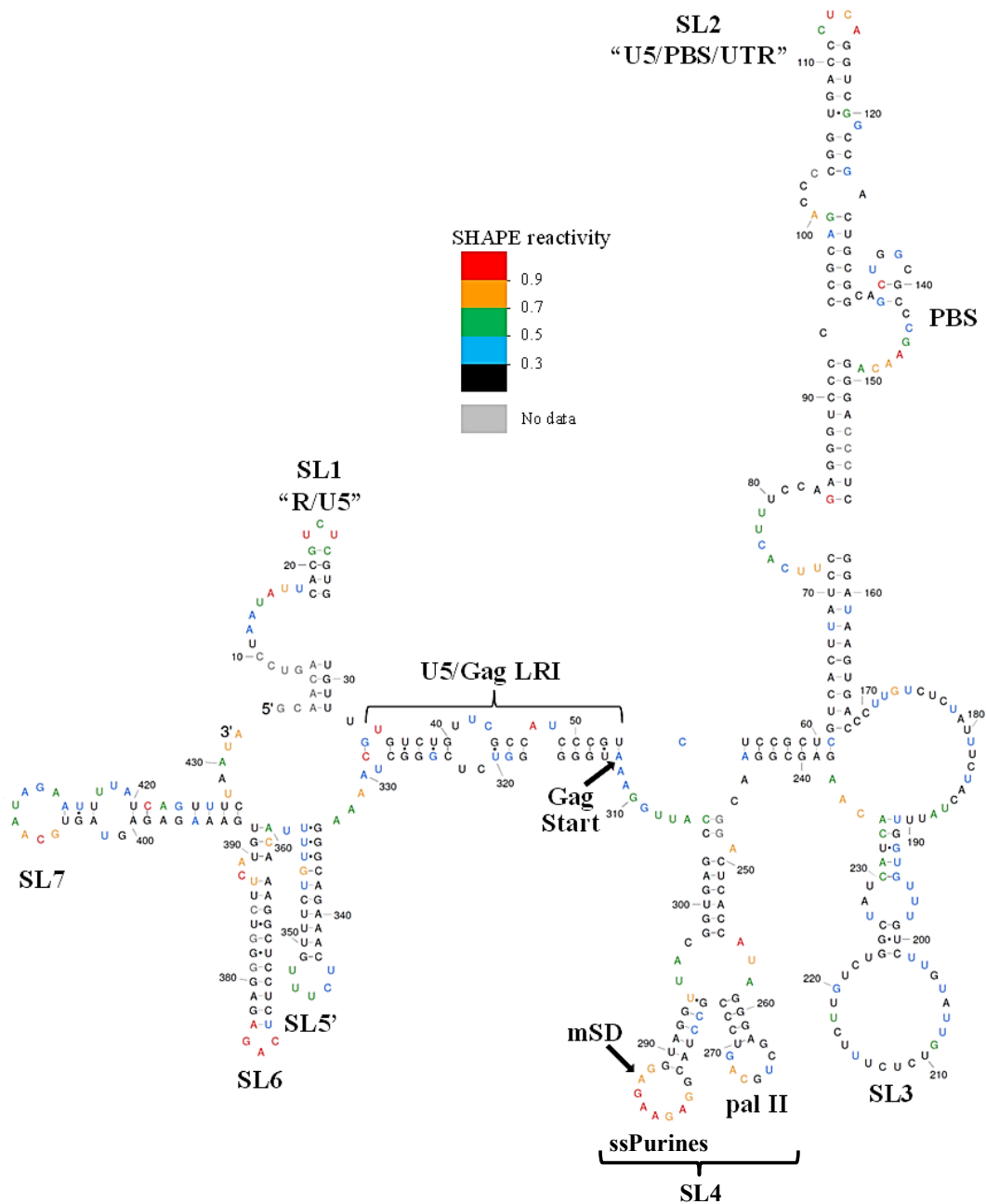


Figure 16: SHAPE constrained RNA structure model of MMTV packaging signal RNA

Nucleotides are color annotated as per the SHAPE reactivities key. Figure adapted from Aktar et al., 2014.

A distinguishing feature of the SHAPE validated structure of the MMTV packaging signal RNA is the presence of a phylogenetically conserved 9 nt stretch (5'GGAGAAGAG 3') of single-stranded purines (ssPurines) in the form of an apical

loop (Aktar et al., 2014). It is interesting to note that this purine apical loop is part of a bifurcated stem loop 4 (SL4) that places the purine loop adjacent to the pal II helix loop (Figure 16), reminiscent to the situation found in MPMV (Aktar et al., 2013; Jaballah et al., 2010). The pal II helix loop has been shown to function as dimerization initiation site (DIS) for MMTV gRNA (Aktar et al., 2014). Interestingly, any manipulation (deletion and/or substitution) in ssPurines sequence dramatically reduces MMTV gRNA packaging and propagation (Mustafa et al., 2018). A mutant containing a complete deletion of ssPurines (AJ006) showed an interesting folding prediction in which the overall RNA secondary structure of MMTV packaging signal RNA is maintained except for replacing the ssPurines loop with a shorter non-purine loop (Figure 17; Mustafa et al., 2018). Furthermore, substitution of ssPurines sequence with a pyrimidine sequence (AJ009) severely impinges gRNA packaging and propagation while maintaining the overall secondary structure of MMTV packaging signal RNA (Mustafa et al., 2018). Since the overall structure of MMTV packaging signal RNA is maintained in these mutants, the abrogation of gRNA packaging and propagation could be directly attributed to the absence of ssPurines and not the destabilization of the overall structure of this region.

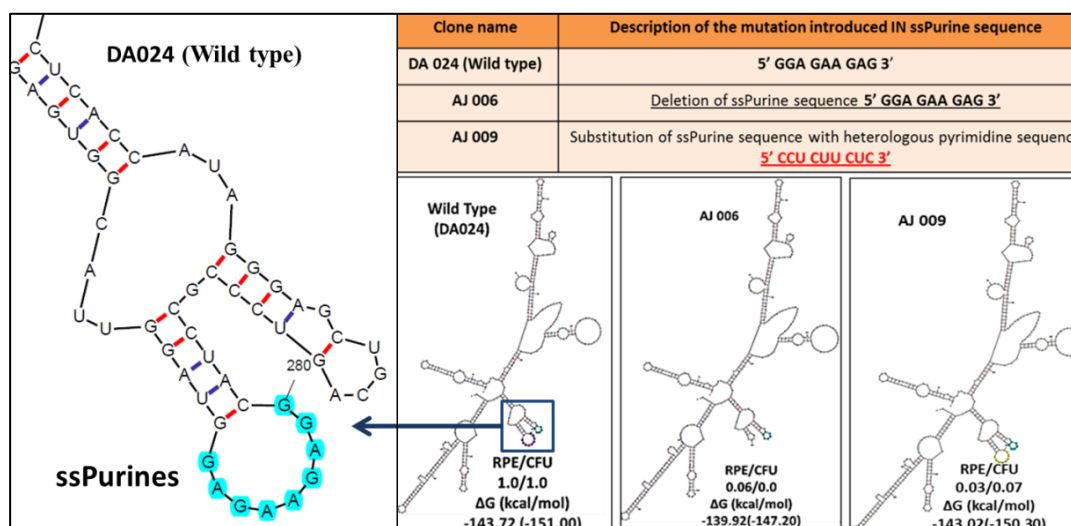


Figure 17: Mfold RNA secondary structure predictions of wild type and mutant MMTV packaging signal RNAs

DA024 is the wild type transfer vector RNA containing the single stranded purines (ssPurines) which are enlarged on the left side for the sake of clarity. AJ006 and AJ009 are mutant transfer vector RNAs in which ssPurines either have been deleted or substituted with pyrimidine sequence. RPE- relative RNA packaging efficiency.

Considering these observations, the hypothesis that was made is ssPurines located in the MMTV packaging signal RNA function either at the sequence or structural level in mediating gRNA packaging, possibly by functioning as a potential Gag binding site. Therefore, the overall goal of this dissertation to express and purify MMTV Pr77^{Gag} and subsequently establish the Gag binding site on MMTV gRNA employing a combination of *in vitro* and *in vivo* experimental approaches.

1.7 Objectives

1.7.1 Specific Aim I: Expression, Purification and Characterization of MMTV Pr77^{Gag}

The mechanism by which the MMTV Pr77^{Gag} specifically recognize the gRNA still remain unknown due to the unavailability of purified full-length Pr77^{Gag}. Thus, large amounts of full-length Pr77^{Gag}-His₆-tag fusion protein was expressed in

bacteria and purified using immobilized metal affinity chromatography (IMAC) followed by size exclusion chromatography (SEC). Further, the purified protein was used to investigate Pr77^{Gag}-His₆-tag fusion protein's differential binding and packaging of MMTV gRNA over spliced RNA. The background and results obtained regarding this specific aim have been discussed in Chapter 2 of this dissertation.

1.7.2 Specific Aim II: Identification and Characterization of Specific Pr77^{Gag} Binding Site(s) on MMTV gRNA

Recently, a stretch of ssPurines in MMTV gRNA packaging signal that may act as a potential Gag binding site either at the sequence level or at the structure level was identified (Aktar et al., 2014). Therefore, a series of mutations in the ssPurines were introduced and cloned into a T7 expression vector. RNA binding assays were performed on the wild type and mutant *in vitro* transcribed RNAs, together with footprinting assay to identify the Pr77^{Gag} binding site on MMTV gRNA. The background and results obtained regarding this specific aim have been discussed in Chapter 3 of this dissertation.

1.7.3 Specific Aim III: Establish Biological Correlation between the Pr77^{Gag} Binding site(s) and MMTV gRNA Packaging

After the identification of the Pr77^{Gag} binding site(s) on MMTV packaging signal RNA, the direct biological role these sequences play in MMTV gRNA packaging and propagation was established. For this purpose, the mutations used for *in vitro* biochemical studies were cloned into a subgenomic transfer vector, DA024 and tested for their packaging efficiency using a biologically relevant *in vivo* packaging and propagation assay (Rizvi et al., 2009). The background and results obtained regarding this specific aim have been discussed in Chapter 3 of this dissertation.

Chapter 2: Biochemical and Functional Characterization of Mouse Mammary Tumor Virus Full-Length Pr77^{Gag} Expressed in Prokaryotic and Eukaryotic Cells

2.1 Abstract

The mouse mammary tumor virus (MMTV) Pr77^{Gag} polypeptide is an essential retroviral structural protein without which infectious viral particles cannot be formed. This process requires specific recognition and packaging of dimerized genomic RNA (gRNA) by Gag during virus assembly. Most of the previous work on retroviral assembly has used either the nucleocapsid portion of Gag, or other truncated Gag derivatives — not the natural substrate for virus assembly. In order to understand the molecular mechanism of MMTV gRNA packaging process, expression and purification of full-length recombinant Pr77^{Gag}-His₆-tag fusion protein from soluble fractions of bacterial cultures was performed. The purified Pr77^{Gag}-His₆-tag protein retained the ability to assemble virus-like particles (VLPs) *in vitro* with morphologically similar immature intracellular particles was shown. The recombinant proteins (with and without His₆-tag) could both be expressed in prokaryotic and eukaryotic cells and had the ability to form VLPs *in vivo*. Most importantly, the recombinant Pr77^{Gag}-His₆-tag fusion proteins capable of making VLPs in eukaryotic cells were competent for packaging subgenomic MMTV RNAs. The successful expression and purification of a biologically active, full-length MMTV Pr77^{Gag} should lay down the foundation towards performing RNA-protein interaction(s), especially for structure-function studies and towards understanding molecular intricacies during MMTV gRNA packaging and assembly processes.

2.2 Introduction

The mouse mammary tumor virus (MMTV) is an oncogenic retrovirus that causes both breast cancer and lymphoma/leukemia in mice. It can be transmitted to the progeny exogenously through the breast milk or vertically through the germline as endogenous viruses (Bittner, 1936; Cardiff & Kenney, 2007; Duesberg & Blair, 1966; Varmus et al., 1972). This makes MMTV a suitable model for studying the mechanism of oncogenesis and the genetics involved in the development of mammary tumors (Dudley et al., 2016; Martin et al., 2017). Furthermore, it has unique genetic properties that make MMTV a desirable vector system for delivering therapeutic genes in human gene transfer studies. The advantages of MMTV-based vectors include: (i) being phylogenetically distinct from human and primate retroviruses, reducing the chances of recombination with endogenous human viruses; (ii) an ability to transduce non-dividing cells, the main target cells of human gene therapy (Konstantoulas & Indik, 2014); (iii) containing multiple promoters and steroid responsive elements, allowing inducible and tissue-specific gene expression (Ham et al., 1988; Klein et al., 2008; Konstantoulas & Indik, 2014; Rouault et al., 2007); and (iv) encoding a unique post transcriptional regulatory system that can enhance gene expression (Indik, 2016; Indik et al., 2005a, 2005b; Mertz et al., 2005, Mertz, Chadee et al., 2009; Mertz, Lozano et al., 2009; Müllner et al., 2008).

Unlike the lentiviruses that are assembled at the plasma membrane, MMTV is a *Betaretrovirus* that displays a type B morphology (Coffin et al., 1997) during replication where intracellular virus particles can be observed. Little is known about how the virus particle is assembled and in particular, the molecular mechanisms of MMTV genomic RNA (gRNA) packaging—a process that allows the virus to

incorporate two copies of its single-stranded RNA genome into the assembling virus particle (Ali et al., 2016; Comas-Garcia et al., 2016; D'Souza & Summers, 2005; Dubois et al., 2018; Johnson & Telesnitsky, 2010; Kaddis Maldonado & Parent, 2016; Lever, 2007; Mailler et al., 2016). Retroviral RNA packaging requires specific interactions between both the gRNA and viral structural proteins, in particular Gag (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Smyth et al., 2015, 2018). Like most retroviruses, an early study suggested that MMTV harbors sequences responsible for gRNA packaging at the 5' end of its genome (Salmons et al., 1989). Recently, employing an *in vivo* packaging and transduction assay (Rizvi et al., 2009), it was shown that the 5' untranslated region (5' UTR) and the first 120 nucleotides (nts) of the *gag* gene are required for efficient MMTV gRNA packaging and propagation (Mustafa et al., 2012). To establish the structural basis of MMTV gRNA packaging, these sequences were folded using minimum free energy algorithm programs like Mfold and RNAstructure (Reuter & Mathews, 2010; Zuker, 2003). The folding predictions of these sequences revealed a higher order structure comprising of several structural motifs, which could be involved during MMTV gRNA packaging (Aktar et al., 2014). Later, this structure was validated by SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) and the structure-function relationship of various structural motifs during MMTV gRNA packaging and dimerization was established (Aktar et al., 2014). Furthermore, it was established that there is a structural motif known as single-stranded purines (ssPurines) in the form of an apical loop which has been proposed to be the potential primary Gag binding site during the process of MMTV gRNA packaging (Aktar et al., 2014).

Retroviral Gag polyproteins comprise of several domains that form the structural elements of the viral particle. Of these, the major domains are the matrix (MA), capsid (CA) and nucleocapsid (NC). The NC serves as the key player in selective gRNA packaging. It is a highly basic protein containing zinc finger motifs to facilitate protein-RNA interactions (Ali et al., 2016; Jewell & Mansky, 2000). By conducting mutational analysis, it has been established that the Gag NC domain of number of retroviruses is the most vital protein domain involved in the gRNA packaging process (Aldovini & Young, 1990; Dorfman et al., 1993; Gorelick et al., 1988; Méric et al., 1988; Poon et al., 1996). However, other Gag domains may also be important for facilitating Gag-RNA interactions, such as the MA (Lu et al., 2011), CA (Kutluay & Bieniasz, 2010), the p2 spacer peptide between CA and NC (Kaye & Lever, 1998; Roy et al., 2006; Russell et al., 2003) and in the case of HIV-1, the terminal p6 late domain (Tanwar et al., 2017). Additionally, it has been suggested that NC probably recognizes dimeric genomes, since dimerization is a prerequisite for RNA packaging (D'Souza & Summers, 2004; Miyazaki, Garcia et al., 2010; Miyazaki, Irobalieva et al., 2010). This interaction is thought to initiate a cascade of events that leads to the oligomerization/multimerization of the Gag polyprotein using gRNA as the substrate, which eventually leads to packaging of the gRNA into the newly forming virus particles. A number of studies have shown that specific selection of gRNA over the cellular and spliced RNAs is a multifaceted phenomenon that has been shown to occur in the context of the whole Gag polyprotein, especially in the case of HIV-1 (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Smyth et al., 2015, 2018). Hence, it is not surprising that the limited understanding towards the highly selective packaging of gRNA by retroviral particles is predominantly due to the unavailability of biologically active full-length Gag polyprotein.

The MMTV Pr77^{Gag}, encoded by the *gag* gene, is a precursor polypeptide, processed by the viral protease into its constituent domains NH₂-p10 (MA), pp21, p3, p8, n, p27(CA) and p14(NC)-COOH (Hizi et al., 1987; Hizi et al., 1989). Like most retroviruses, the MMTV Pr77^{Gag} assembles into an immature capsid and the proteolytic maturation takes place coupled with release from the cell (Smith, 1978; Tanaka et al., 1972). The Pr77^{Gag} polyprotein plays a key role in selectively packaging the full length unspliced gRNA from a pool of cellular and spliced RNAs during viral assembly. The precise mechanism(s) by which Pr77^{Gag} accomplishes this specific selection is yet to be established. For example, it remains rather ambiguous whether the binding of Pr77^{Gag} to gRNA is based on an intrinsic capability of the polyprotein that allows specific selection of gRNA over its spliced variants, or whether other steps in the retroviral life cycle such as the nucleo-cytoplasmic transport and cellular compartmentalization of gRNA are also involved in discriminating gRNA during encapsidation process, as has been suggested for HIV-1 (Barajas et al., 2018; Becker & Sherer, 2017; Behrens et al., 2017; Brandt et al., 2007; Jouvenet et al., 2011; Moore et al., 2009). Thus, to delineate the molecular mechanism of MMTV gRNA packaging, it is critical to understand the biophysical and biochemical properties of full length Pr77^{Gag}. However, the expression and purification of the biologically active MMTV full-length Pr77^{Gag} has not been accomplished in bacteria, although certain other MMTV proteins have been successfully purified after being expressed in bacteria, such as the *gag-pro* transframe protein p30 and reverse transcriptase (Köppe et al., 1994; Taube et al., 1998).

This study reported the successful expression and purification of large amounts of full-length Pr77^{Gag} in soluble fractions of *Escherichia coli* (*E. coli*) containing a hexa-histidine (His₆) tag at C-terminus. The purified MMTV Pr77^{Gag} could assemble *in vitro* into virus-like particles (VLPs), form intracellular VLPs in bacteria and eukaryotic cells and most importantly, successfully package MMTV RNA. Thus, efficient expression and purification of full-length Pr77^{Gag} should allow us to investigate the differential binding ability of this protein to the unspliced full-length gRNA during selective RNA packaging process. This should additionally widen the understanding of the mechanisms of RNA-protein interaction(s) involved in gRNA packaging during MMTV life cycle.

2.3 Materials and Methods

2.3.1 Nucleotide Numbers

All nucleotide numbers in this study refer to the MMTV genome pertaining to Genbank accession number AF228550.1 (Hook et al., 2000).

2.3.2 Full-Length Recombinant Gag Prokaryotic Expression Plasmids

MMTV full-length *gag* gene (Pr77^{Gag}; nucleotides 1485–3260) was commercially synthesized (Macrogen, South Korea) with flanking *NcoI* and *XhoI* sites and cloned into the bacterial expression vector pET28b(+) (Figure 18). Since the *gag* gene contain an inherent *NcoI* site at nt 2389, a silent mutation was created at this site by introducing a one nucleotide modification (*ACCATGG* was changed to *ACTATGG*, while maintaining the amino acid threonine, encoded by the italicized codon), resulting in only one *NcoI* site, to facilitate the cloning process. Using *NcoI*

and *Xho*I sites during cloning resulted in placing the *gag* sequences in-frame with a hexa-histidine sequence that allowed the addition of a His₆-tag at the C-terminus of recombinant Pr77^{Gag} protein with a predicted molecular weight of 65,890 Da, creating a molecular clone, AK1 (Figure 18). AK1 was further modified in such a fashion that a potential N-terminal-truncated protein from a second in-frame AUG, located at nts 1674–1676, is not expressed. Towards this end, a region harboring the Shine-Dalgarno-like sequence (underlined; 5' AAAAGGGTAGGAAGAGAAATG 3'), located four nucleotides upstream of the second in-frame AUG (Bray et al., 1994) was silently mutated and the substituted nucleotides are underlined (5' AAGCGCGTGGGCCGCGAGATG 3') without disrupting the amino acid sequence. These modifications generated the prokaryotic expression plasmid, AK7. AK7 was further modified by introducing a stop codon at the end of the *gag* sequence to create AK31 in a fashion that it expressed the full-length MMTV Pr77^{Gag} without the His₆-tag. The resultant clones were sequenced to ensure that they were devoid of any mutations.

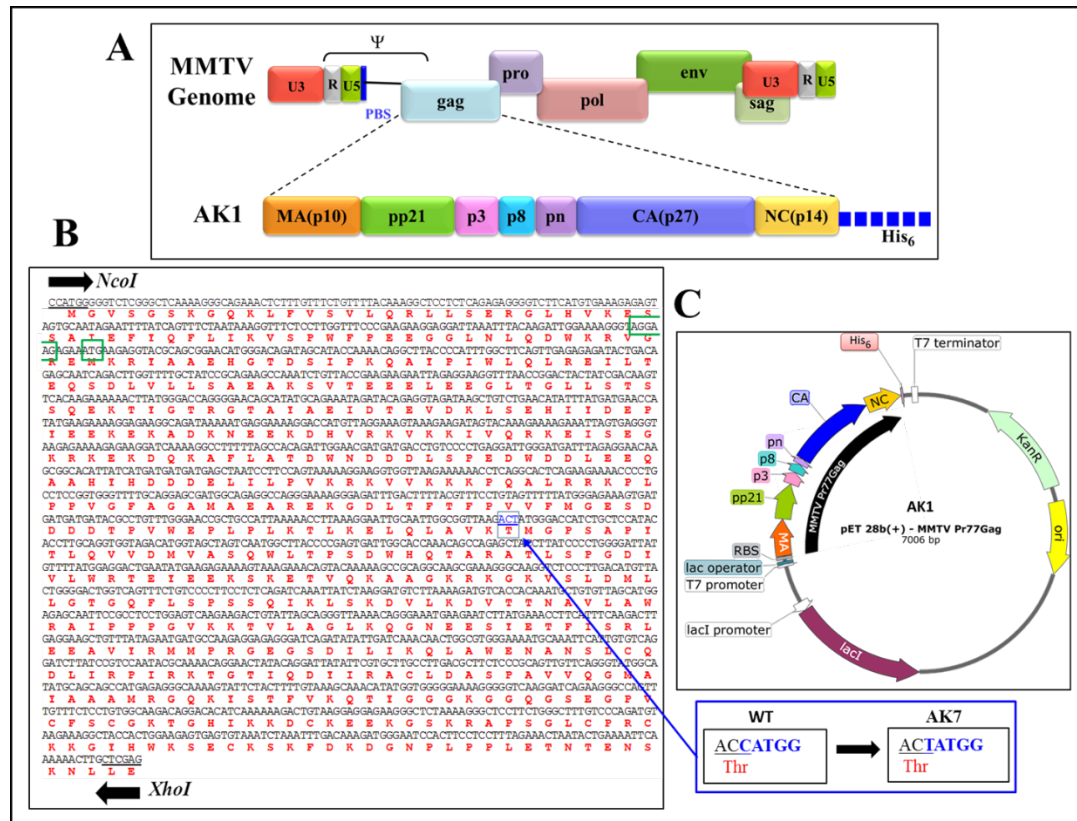


Figure 18: Construction of the recombinant full-length Pr77^{Gag} bacterial expression vector

(A) Domain organization of the mouse mammary tumor virus (MMTV) Gag precursor with His₆-tag; (B) Nucleic acid and amino acid sequences of full-length MMTV gag gene. An internal *NcoI* site (boxed in blue color) was removed by introducing a silent mutation (shown in the inset) that preserved the threonine (Thr) amino acid. The Shine-Dalgarno-like sequence and second in-frame ATG are highlighted by green color; (C) Schematic representation of bacterial expression plasmid AK1 containing full-length MMTV Pr77^{Gag} gene cloned into the *NcoI* and *XhoI* sites of the pET28b(+) vector.

2.3.3 Full-Length Recombinant Gag Eukaryotic Expression Plasmids

The MMTV gag genes cloned into the prokaryotic expression vector (both with and without the His₆-tag) was re-cloned into the eukaryotic expression vector, pCDNA3, using polymerase chain reaction (PCR). Towards this end, the same forward primer OTR1333 was used along with the reverse primer, OTR1335 that introduced the His₆-tag at the end of gag gene using the template AK7. In order to introduce an *XhoI* restriction site (italicized) for cloning purposes and the Kozak

sequences (underlined) at the 5' end of the *gag* gene, an OTR1333 (5' CCG *CTCGAGGCCGCCACCATGGGGGTCTCGGGCTCAAAA* 3') forward primer was employed. Employing similar strategy for introducing the His₆-tag (underlined) just upstream of the *gag* stop codon, followed by an *Xho*I endonuclease site (italicized) OTR1335 (5' CCG*CTCGAGTTA GTGGTGGTGGTGGTGGTGGT*GCAAGTTTTTTTGA ATTTTCAGTATTAGTTTC 3') was used. In order to create a eukaryotic Gag expression plasmid which did not contain the His₆-tag but did contain an *Xho*I restriction site (italicized) immediately downstream of the stop codon, the reverse primer OTR1322 (5' CCG*CTCGAGTTA CAAGTTTTTTTGA* 3') was used. PCR was performed as follows: Initial denaturation at 98°C for 30 s, then 15 cycles of denaturation at 98°C for 10 s, primer annealing at 62°C for 30 s, followed by primer extension at 72°C for 30 s and a final extension at 72°C for 10 min. The amplified products were digested with *Xho*I endonuclease and cloned into pCDNA3 previously digested with the *Xho*I to create clones AK9 and AK10 (with and without His₆-tag, respectively). Finally, to ensure proper nuclear export and translation of Gag mRNA, a PCR-amplified fragment containing the MPMV constitutive transport elements (CTE; Bray et al., 1994; Rizvi et al., 2009) with flanking *Xba*I sites was cloned into AK9 and AK10 previously digested with *Xba*I generating clones, AK13 and AK14. All clones were confirmed by sequencing.

2.3.4 *Escherichia coli* Strains and Growth Media

The cloning of different expression plasmids was performed using the DH5 α strain of *E. coli* using the conventional heat shock protocol with appropriate antibiotics (kanamycin; 50 μ g/mL, ampicillin; 100 μ g/mL). For prokaryotic protein expression, clones AK1, AK7 and AK31 were transformed into the BL21(DE3)

strain of *E. coli*, cultured in Luria-Bertani (LB) medium (1% (w/v) peptone, 0.5% (w/v) yeast extract and 0.5% NaCl) supplemented with kanamycin (50 µg/mL).

2.3.5 Expression of Recombinant Full-Length Pr77^{Gag}-His₆-Tagged Protein in Bacteria

For large scale expression of recombinant Pr77^{Gag}-His₆-tag protein, a single colony of transformed *E. coli* BL21(DE3) cells was inoculated into 50 mL of LB media containing kanamycin antibiotic (50 µg/mL), then cultured at 37°C overnight with agitation at 200 rpm. The overnight culture was sub-cultured into 500 mL LB supplemented with the same concentration of kanamycin and 1% glucose in 2-L baffled flasks. The cultures were allowed to grow at 28°C till an OD₆₀₀ of 0.6 was achieved. Cultures were then induced with 0.4 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) and the cells were grown for an additional 4 h at 28°C. Cells were pelleted by centrifugation at 4°C for 15 min at 6300× *g* and stored at –80°C until processed.

2.3.6 Affinity Purification and Gel Filtration Chromatography

Purification of the recombinant Pr77^{Gag}-His₆-tag protein was carried out as previously described (Bewley et al., 2017; McKinstry et al., 2014; Tanwar et al., 2017). Frozen bacterial pellets were lysed in chilled CelLytic B buffer (Sigma-Aldrich, Saint Louis, MO, USA) supplemented with 5 U/mL Benzonase (Merck, Kenilworth, NJ, USA), 0.2 mg/mL of lysozyme (Sigma-Aldrich) and 1× ethylenediaminetetraacetic acid (EDTA) free protease inhibitor (Roche, Basel, Switzerland). The lysate was then centrifuged at 48,000× *g* for 1 h at 4°C and the 4× binding buffer (0.2 M Tris-HCl of pH 8.0, 4.0 M NaCl, 40 mM β-mercaptoethanol, 10 mM dithiothreitol, 100 mM imidazole, 0.4% (w/v) Tween-20) was used to dilute

the supernatant to a final concentration of 1×. Prior to loading onto 5 mL HisTrap fast flow (FF) column (GE Healthcare, Little Chalfont, UK) which was pre-equilibrated with a buffer (with 50 mM Tris-HCl (pH 8.0), 1.0 M NaCl, 10 mM β-mercaptoethanol, 2.5 mM dithiothreitol, 25 mM imidazole, 0.1% (w/v) Tween-20 and 10% (v/v) glycerol), the lysate was filtered through a 0.4 μm polyethersulfone (PES) syringe filter. After loading the filtrate, the column was washed with the same buffer except for increasing the concentration of imidazole to 50 mM and the protein that bound to the column was eluted with buffer containing 250 mM imidazole.

The HisTrap FF eluted Pr77^{Gag}-His₆-tag protein was concentrated using Amicon Ultra 15 column (with a 30 kDa cut-off; Merck) for fractionation by gel filtration/size exclusion chromatography using a Superdex 200 increase 10/300 GL column (GE Healthcare) which was equilibrated with 50 mM Tris-HCl (pH 8.0) and 1.0 M NaCl. Peak fractions were analyzed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and fractions containing Pr77^{Gag}-His₆-tag protein were pooled and stored at -80°C for downstream processing. The purity of the protein was established by measuring the absorbance ratio at 260 and 280 nm.

2.3.7 Expression of Recombinant Full-Length Pr77^{Gag}-His₆-Tagged Protein in Eukaryotic Cells

The expression of Gag in eukaryotic cells was monitored in transient transfections using calcium phosphate kit (Invitrogen, Carlsbad, CA, USA) in HEK 293T cells. The transfections were carried out in triplicates in 6-well plates using 4 micrograms (μg) of full-length Gag eukaryotic expression plasmids (AK13 or AK14) along with 2 μg of MMTV-based transfer vector, DA024 (Rizvi et al., 2009). To

monitor transfection efficiencies, a secreted alkaline phosphatase expression plasmid (pSEAP, at a concentration of 100 ng per well) was also included in the transfections. Approximately 72 h post transfection, supernatants from the transfected cultures were harvested and subjected to low-speed centrifugation (3700× g for 10 min) to clear cellular debris. The clarified supernatants were then filtered through 0.2 µm cellulose acetate syringe filters and subjected to ultracentrifugation at 70,000× g with a 20% (w/v) sucrose cushion to pellet the VLPs. The pelleted VLPs were then resuspended in TN buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl) and subjected to RNA extraction (TRIzol) and western blotting.

2.3.8 Estimation of RNA Packaging Potential by Reverse Transcriptase PCR

Packaging of MMTV RNA into the Gag-VLPs was tested by RT-PCR. Both cytoplasmic and viral RNA preparations were DNase-treated with TURBO DNase (Invitrogen) and amplified using transfer vector (DA024)-specific primers OTR671 (5' GTCCTAATATTCACGTCTCGTGTG 3') and OTR672 (5' CTGTTTCGGGCGCCAGCTGCCGCAG 3') to confirm the successful removal of contaminating plasmid DNA in the extracted RNA preparations. The cDNA synthesis from the DNased-RNAs was accomplished using random hexamers (5' NNNNNN 3') and MMLV reverse transcriptase (Promega, Madison, WI, USA) as previously described (Ghazawi et al., 2006; Mustafa et al., 2005). To monitor the ability of Pr77^{Gag} VLPs to package transfer vector (DA024) RNA, complementary DNAs (cDNAs) were amplified using the same vector-specific primers (OTR671 and OTR672).

2.3.9 Sodium dodecyl sulfate-Polyacrylamide Gel Electrophoresis and Western Blotting

SDS-PAGE and western blotting were used to monitor the expression and purification of recombinant Pr77^{Gag}-His₆-tag protein. Protein samples were analyzed on 4–12% ExpressPlus™ PAGE gel (GenScript, Piscataway, NJ, USA), electrophoresed under reducing conditions using 3-(N-morpholino)propanesulfonic acid (MOPS) buffer (GenScript) and stained with Coomassie Brilliant Blue. For western blot analyses, duplicate gels were transferred onto nitrocellulose membranes and probed with α -MMTV p27 CA monoclonal antibody Blue 7 (Purdy et al., 2003) and an α -His₆ monoclonal antibody- horseradish peroxidase (HRP) conjugate (Sigma-Aldrich).

2.3.10 Detection of Prokaryotically-Expressed Virus-Like Particles Using Transmission Electron Microscopy

To observe VLPs formed by recombinant Pr77^{Gag}-His₆-tag protein in bacterial cells (following IPTG induction), the pelleted cells were washed with 0.1 M phosphate buffered saline (PBS) and fixed in Karnovsky's fixative overnight. Cell pellets were then stained with 1% osmium tetroxide and subjected to graded ethanol dehydration. The pellets were then embedded in epoxy resin (Agar 100). Ultrathin (95 nm) sections of the embedded samples were fixed on 200 mesh copper (Cu) grids and negatively stained with 1% uranyl acetate and lead citrate as a double stain. The sections were analyzed using a FEI Tecnai Biotwin Spirit G2 transmission electron microscope.

2.3.11 *In Vitro* Assembly of Virus-Like Particles from Bacterially Expressed Recombinant Full-Length Pr77^{Gag}-His₆-Tag Protein

The purified recombinant Pr77^{Gag}-His₆-tag protein expressed in bacteria was resuspended in assembly buffer (50 mM Tris (pH 7.4), 1.0 M NaCl) at a concentration of 2 mg/mL and incubated with yeast tRNA at a nucleic acid to protein ratio of 4% (w/w). This mix was placed in a Slide-A-Lyzer 10K dialysis cassette G2 (Thermo Scientific, Waltham, MA, USA) and dialyzed against dialysis buffer (20 mM Tris (pH 7.4), 150 mM NaCl and 10 mM dithiothreitol) overnight at 4°C. Following overnight dialysis, ~8–10 µL of the dialyzed solution was spotted onto a carbon-coated formvar grid (Proscitech, Kirwan, Australia), air dried and stained with 1% uranyl acetate for observation using transmission electron microscopy (TEM).

2.4 Results and Discussion

2.4.1 Successful Expression of Full-Length Recombinant MMTV Pr77^{Gag}-His₆-Tagged Protein in Bacteria

For the expression of full-length MMTV Pr77^{Gag} which contained a C-terminus His₆-tag, a recombinant bacterial expression plasmid (AK1; Figure 18) was generated. High level expression of His-tagged MMTV Pr77^{Gag} with a predicted molecular weight of ~65,890 Dain BL21(DE3) bacterial cells was achieved by induction with IPTG (Figure 18).

The expression of the recombinant Pr77^{Gag}-His₆-tagged protein in total bacterial lysates was monitored at 0, 2, 4 and 18 h post induction, by SDS-PAGE (Figure 19). Bands corresponding to the size of recombinant Pr77^{Gag}-His₆-tagged protein (70 kDa) were observed only in the IPTG induced cultures at 2, 4 and 18 h

(Figure 19, lanes 5, 7 and 9) but not in cultures at 0 h (lane 3) or un-induced cultures at 2, 4 and 18 h (Figure 19, lanes 4, 6 and 8), as well in cultures transformed with only pET28b(+) expression vector (Figure 19, lane 2). This is despite the presence of the “poison sequences” present in Gag that are presumably “toxic” for bacteria (Brookes et al., 1986). It is possible that the absence of effect of these poison sequences due to the inducible and suboptimal nature of the expression system currently used, allowing the bacteria to survive for short periods under these conditions.

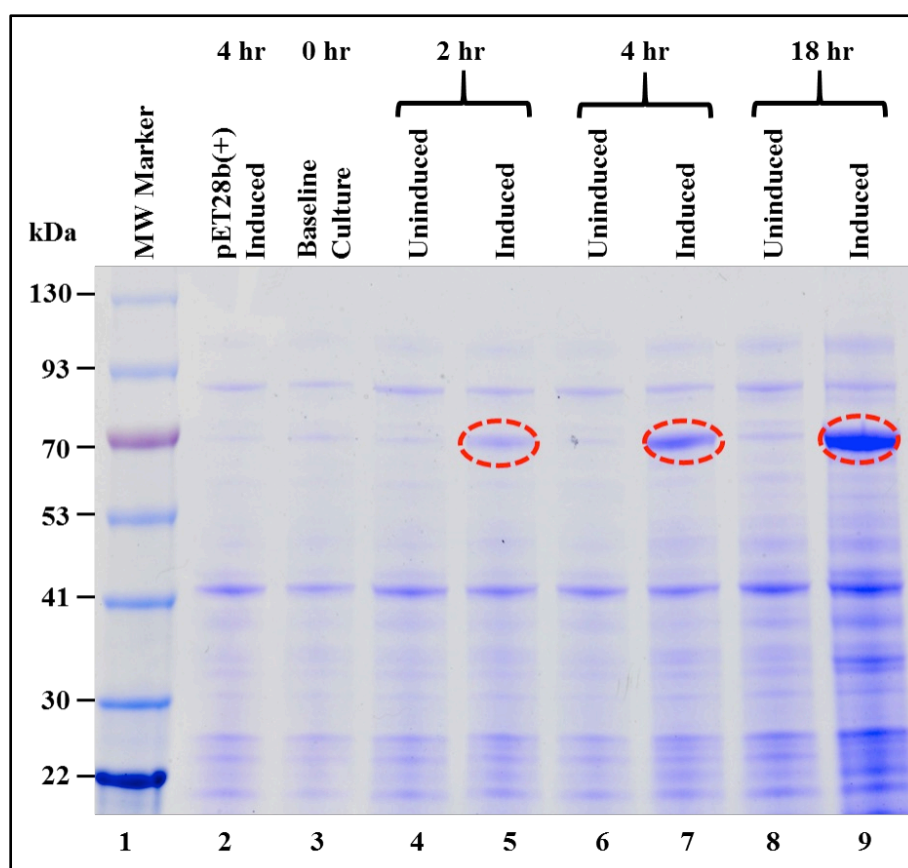


Figure 19: Expression of recombinant Pr77^{Gag}-His₆-tag fusion protein in *Escherichia coli* (*E. coli*)

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis showing full length Pr77^{Gag}-His₆-tag fusion protein expressed from AK1 in total bacterial cell lysates at 0, 2, 4 and 18 h post IPTG-induction and un-induced BL21(DE3) bacterial cells. The bacterial cells were grown at 37°C overnight, but following IPTG induction, cultures were grown sub-optimally at 28°C. MW: molecular weight.

2.4.2 Full-Length MMTV Pr77^{Gag}-His₆-Tagged Fusion Protein is Expressed in the Soluble Form in Bacteria

Next, to establish whether the recombinant MMTV Pr77^{Gag}-His₆-tagged fusion protein was expressed in the soluble bacterial fraction so that it could be purified, the large-scale expression of recombinant MMTV Pr77^{Gag} was performed at sub-optimal conditions such as low temperature (28°C) and shorter duration (4 h only) as described in Materials and Methods. This was based on the earlier observations that in case of Mason-Pfizer monkey virus (MPMV) Pr78^{Gag}, culturing bacteria at 37°C post-induction resulted in the confinement of MPMV Gag polyprotein in the inclusion bodies containing aberrantly assembled spiral like structures (Klikova et al., 1995). Removal of insoluble material (cell debris and/or inclusion bodies) was accomplished by centrifugation of the bacterial lysates. The soluble fractions from different cultures were analyzed for the expression of full-length MMTV Pr77^{Gag}-His₆-tag fusion protein.

As expected, SDS-PAGE analysis of the soluble fraction revealed a very distinctive band of ~70 kDa which corresponded to the size of recombinant MMTV Pr77^{Gag}-His₆-tag fusion protein (Figure 20A; lane 4). Immunoblotting on AK1 un-induced culture lysates using α -His₆ monoclonal antibody as well as α -MMTV p27 monoclonal antibody also showed low level expression of MMTV Pr77^{Gag}-His₆-tag fusion protein (Figure 20B, C; lane 3). Such low-level expression in the un-induced culture could possibly be due to the leaky nature of the bacterial promoter (discussed later).

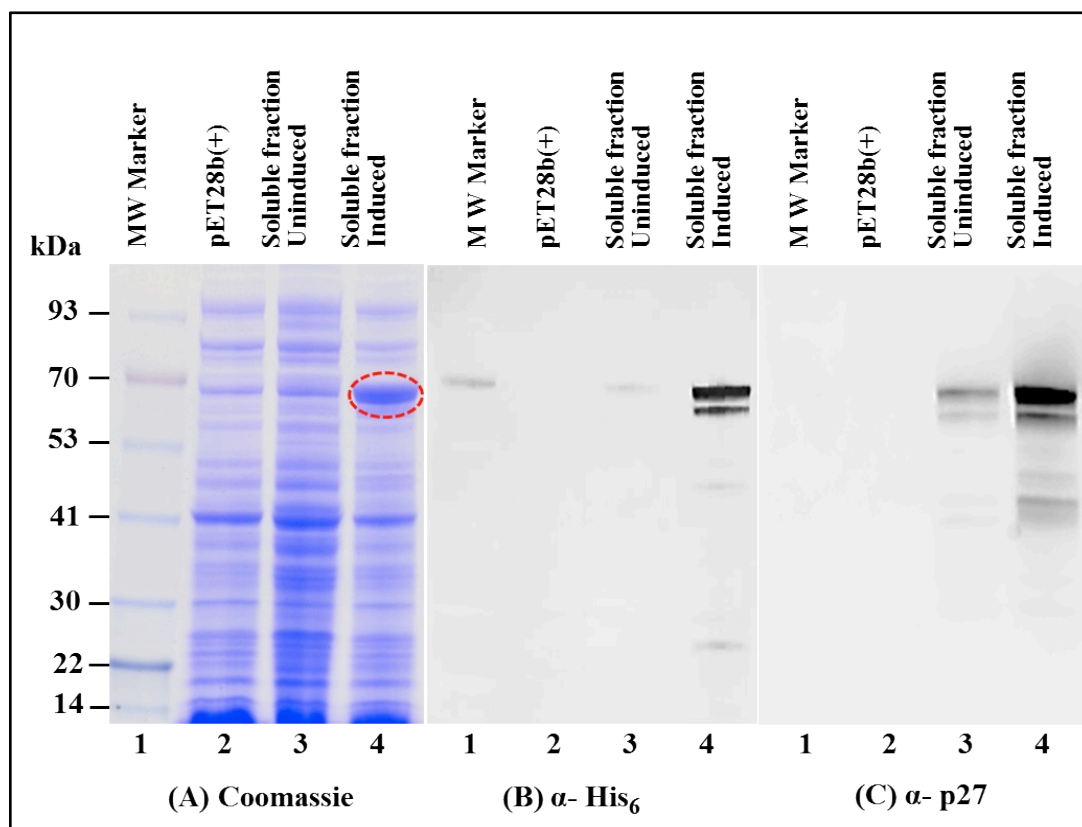


Figure 20: Recombinant Pr77^{Gag}-His₆-tag fusion protein expressed in soluble fraction of *E. coli*

(A) SDS-PAGE analysis showing recombinant MMTV Pr77^{Gag}-His₆-tag fusion protein expression in the bacterial soluble fraction (lane 4) transformed with AK1; (B) western blot analysis of MMTV Pr77^{Gag} expression by AK1 in soluble fraction analyzed with an α -His₆ monoclonal antibody (lane 4); and (C) with an α -p27 monoclonal antibody (lane 4), respectively.

Immunoblotting on IPTG induced soluble fractions using HRP-conjugated α -His₆ monoclonal antibody (Figure 20B; lane 4) as well as α -MMTV p27 monoclonal antibody (Figure 20C; lane 4) confirmed the identity of the ~70 kDa band; however, another band of a slightly lower molecular weight was also observed. Careful analysis of the MMTV full-length *gag* sequence suggested that the second band could be due to the expression of a truncated protein from a second in-frame AUG (nts 1674–1676) located 189 nts downstream from the canonical AUG (nts 1485; Figure 21A). Expression of the truncated Gag protein from this internal AUG was possibly facilitated by the presence of a Shine-Dalgarno-like sequence (AGGAAG;

Figure 21A), located 4 nts upstream of the in-frame second AUG (Shine & Dalgarno, 1974). This was confirmed by calculating the relative translation rates from both the first AUG as well as from the second in-frame AUG employing an online program called RBS Calculator v2.0 (Espah Borujeni et al., 2014; Salis, 2011). Relative translational rate analysis revealed that the predicted translation rate from the first AUG was 2960 arbitrary units (au), whereas that from the second in-frame AUG was 1198 au (Figure 21B). Such a predicted translation rate corroborated well with the level of intensities of the bands following immunoblotting with the truncated protein being expressed approximately 1/3rd of the level of the full-length Gag (Figure 20B,C; lane 4). To eliminate expression from the second AUG, silent mutations were introduced in the 18 nts region (1656–1673) that included the Shine-Dalgarno-like sequence (AGGAAG), resulting in its disruption (from AGGAAG to GGGCCG), but without changing the amino acids sequence (Figure 21A). These changes reduced the predicted translation rate from the second in-frame AUG to almost negligible levels (from 1198 to 7.3 au: Figure 21B).

sequence was sufficient to disable the expression of the truncated Pr77^{Gag}-His₆-tag fusion protein (Figure 22B; lane 3 and Figure 22C; lane 3). Finally, in addition to the desired Pr77^{Gag}-His₆-tag fusion protein, some spurious bands were also observed in immunoblots using α -MMTV p27 monoclonal antibody (Figure 22C; lane 3). These nonspecific bands could be due to the possible degradation of this recombinant protein and could be removed following size exclusion chromatography (described later). Taken together, these results clearly demonstrate that the MMTV recombinant Pr77^{Gag}-His₆-tag protein could be expressed primarily in the soluble fraction of bacterial lysates.

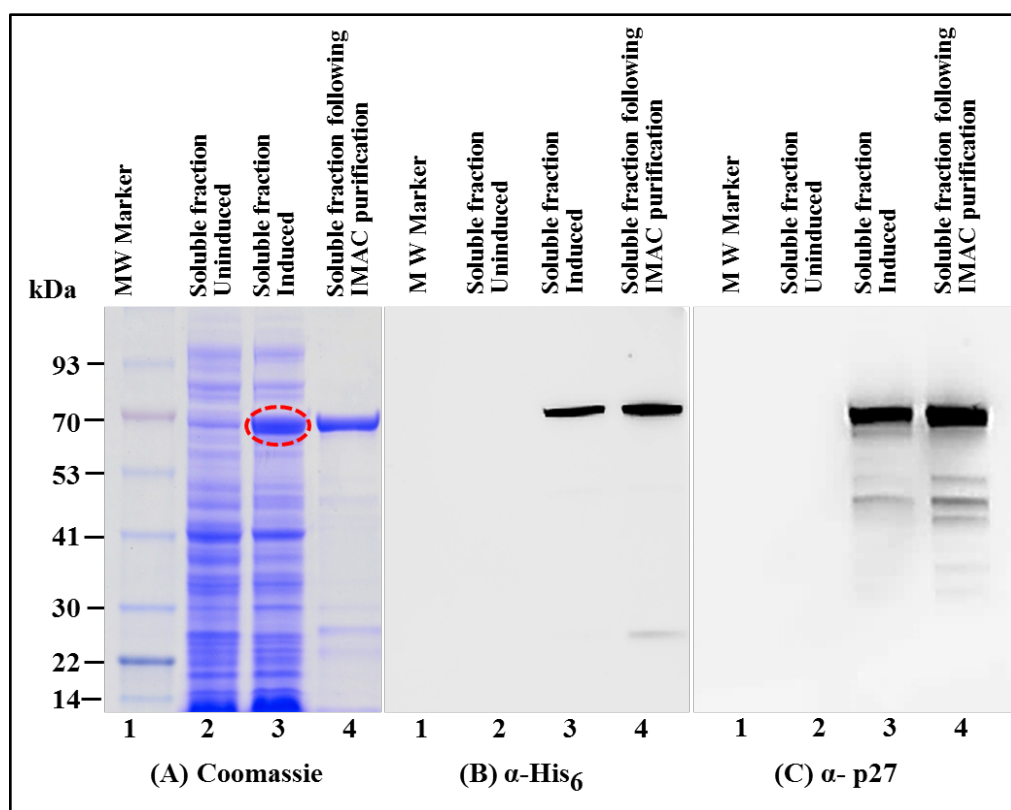


Figure 22: Expression of recombinant Pr77^{Gag}-His₆-tag fusion protein from AK7 in soluble fractions of *E. coli* before and after immobilized metal affinity chromatography (IMAC) purification

(A) SDS-PAGE analysis showing recombinant MMTV Pr77^{Gag}-His₆-tag fusion protein expressed in the bacterial soluble fractions transformed with AK7 (lane 3), followed by IMAC purification (lane 4); (B) western blot analysis of MMTV Pr77^{Gag}-His₆-tag fusion protein analyzed with an α -His₆ monoclonal antibody; and (C) with α -p27 monoclonal antibody, respectively.

2.4.3 Immobilized Metal Affinity Chromatography Purification of the Soluble Fraction Containing Recombinant Full-Length Pr77^{Gag}-His₆-Tagged Fusion Protein

After establishing that the expressed recombinant Pr77^{Gag}-His₆-tag protein was present in the soluble fraction, purification of the recombinant Pr77^{Gag}-His₆-tag protein from the bacterial lysate was performed by employing immobilized metal affinity chromatography (IMAC), as described in Materials and Methods. The non-denaturing buffering conditions (especially the presence of 1.0 M NaCl) were used to facilitate the binding of the protein to the column and to circumvent protein aggregation and precipitation. Following IMAC purification the purity of the recombinant MMTV Pr77^{Gag}-His₆-tag protein was established by SDS-PAGE and immunoblotting. Coomassie Brilliant Blue staining of the gels revealed that the bacterial proteins that were largely present in the soluble fraction before IMAC purification were removed during IMAC purification (Figure 22A; compare lane 3 with lane 4). Immunoblotting of IMAC-purified protein with α -His₆ (Figure 22B; lane 4) and α -MMTV p27 monoclonal antibodies (Figure 22C; lane 4) still showed several additional bands, possibly due to degradation of the fusion protein (Figure 22C) which were successfully removed following size exclusion chromatography (described later). These results confirm that the presence of His₆-tag at the C-terminus of MMTV full-length Gag not only allowed its binding to the HisTrap column but also facilitated elution in a much purer form (Figure 22A; compare lane 3 with lane 4).

2.4.4 Gel Filtration Chromatography Purification of the IMAC-Purified Recombinant Full-Length Pr77^{Gag}-His₆-Tagged Fusion Protein

The protein eluted after IMAC purification was concentrated using an Amicon Ultra 15 centrifugal columns (30 kDa cut-off membrane). Further purification of the protein was carried out by size exclusion chromatography under non-denaturing conditions using a Superdex 200 10/300 GL column. The high salt concentration in the gel filtration buffer (1.0 M NaCl) prevented any possible protein aggregation and precipitation. 500 μ L fractions were collected for several hours (Figure 23A) and protein fractions showing strong absorbance at 280 nm (fractions 23–27) were subjected to SDS-PAGE and western blot analyses.

As shown in Figure 23B, fractions corresponding to the sharp peak consisted of pure MMTV Pr77^{Gag}-His₆-tag fusion protein with varying amounts of protein. Since fractions 26 and 27 showed an additional band of a smaller size, only fractions representing the highest amount of pure protein (peaks 24 and 25; Figure 23B) were pooled, concentrated, and further analyzed by immunoblotting using α -MMTV p27 and α -His₆ monoclonal antibodies. Figure 23C, in close corroboration with the observed SDS-PAGE analysis (Figure 23B), demonstrated that the pooled protein fractions contained pure MMTV Pr77^{Gag}-His₆-tag fusion protein. The purity of the protein was assessed to be greater than 95% as measured by the A_{260}/A_{280} ratio by spectrophotometry (giving a value of 0.6), confirming that the purified protein contains only an insignificant level of nucleic acid contamination. From one liter of bacterial culture, ~4.4 mg protein was obtained after IMAC purification. When IMAC purified protein (4.4 mg) was subjected to gel filtration/size exclusion chromatography ~1.4 mg purified protein was recovered.

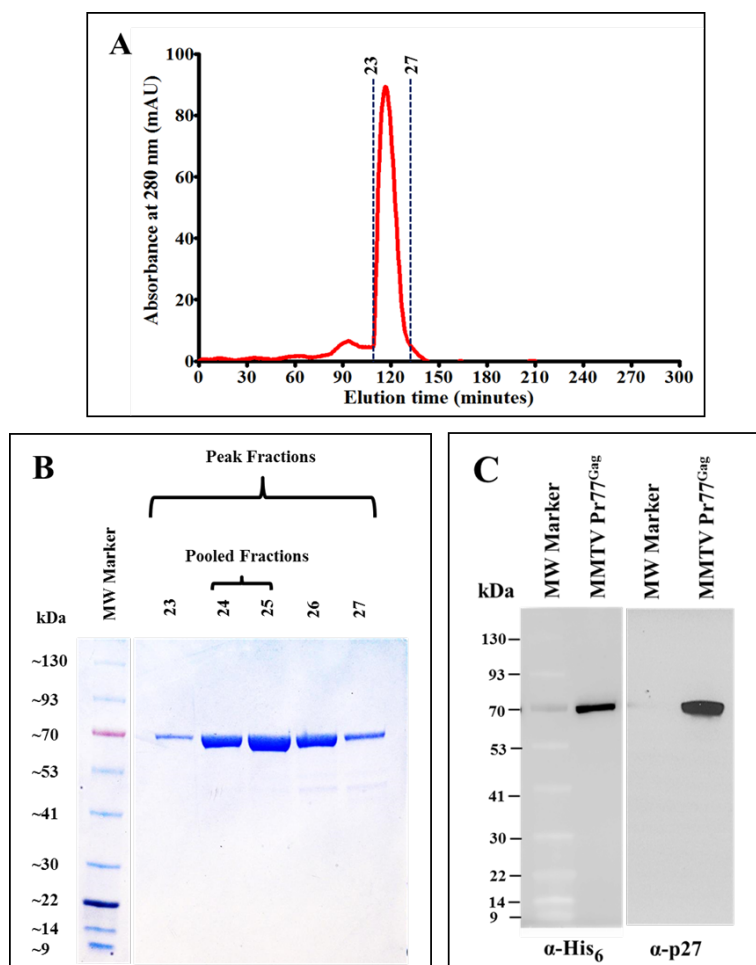


Figure 23: Resolution of IMAC-purified recombinant Pr77^{Gag}-His₆-tag fusion protein by size exclusion chromatography and western blot analysis

(A) Absorbance versus elution time chromatogram of eluted fractions after size exclusion chromatography; (B) Coomassie Brilliant Blue-stained SDS-PAGE analysis of peak fractions 23 to 27, showing the resolution of purified recombinant MMTV Pr77^{Gag} expressed from AK7; (C) western blot analysis of pooled peak fractions of purified MMTV Pr77^{Gag}-His₆-tag fusion protein analyzed with α -His₆ and α -p27 monoclonal antibodies, respectively.

2.4.5 *In Vitro* Assembly to Form Virus-Like Particles by the Recombinant Full-Length Pr77^{Gag}-His₆-Tagged Fusion Protein

The ability of a number of retroviral recombinant full-length Gag proteins to assemble *in vitro* to form VLPs have already been established (Affranchino & González, 2010; Campbell & Rein, 1999; McKinstry et al., 2014; Tanwar et al., 2017). Thus, the *in vitro* assembling ability of purified recombinant MMTV Pr77^{Gag}-His₆-tag fusion protein expressed from AK7 was analyzed. The *in vitro* assembly

experiment was carried out in the presence of yeast tRNA since the presence of nucleic acids along with purified Gag protein has been shown to be a prerequisite for *in vitro* assembly of VLPs (Affranchino & González, 2010; Campbell & Rein, 1999; McKinstry et al., 2014; Tanwar et al., 2017). The protein-RNA mixture in a higher salt concentration buffer (1 M NaCl) was then subjected to dialysis against a buffer with physiological salt concentration. A sample with only yeast tRNA was also dialyzed in the same manner as a control. After overnight dialysis, the protein-RNA mixture was recovered from the dialysis cassette and 10 μ L (~1/40th of the suspension) was spotted on a formvar carbon coated grids and processed for TEM.

VLPs of approximately 62–66nm in size resembling immature virus particles were observed in various electron micrographs taken from different fields (Figure 24A–D). In contrast, as expected, yeast tRNA alone without any purified MMTV full-length Gag did not show any VLP-like structure (Figure 24E, F). Earlier studies have reported a smaller size VLPs (~20–30 nm) obtained following *in vitro* assembly using purified full-length Gag from HIV-1 and feline immunodeficiency virus (FIV; Affranchino & González, 2010; Campbell & Rein, 1999; McKinstry et al., 2014; Tanwar et al., 2017) in contrast to the larger size of full-length Gag particles that have been observed *in vivo* in eukaryotic cells. However, in this case, the size of *in vitro* assembled VLPs are comparable to those observed in *E. coli* (discussed below) and suggest a unique property of MMTV, making this virus particularly interesting for *in vitro* assembly studies.

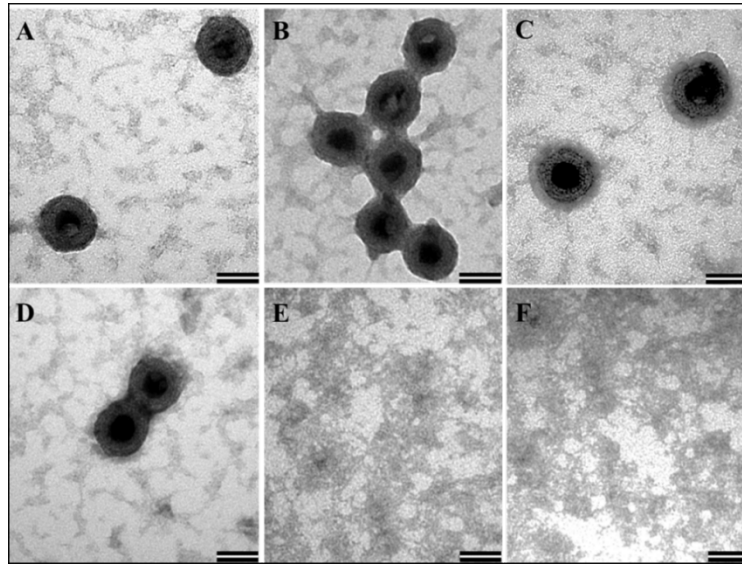


Figure 24: Transmission electron micrographs showing virus-like particles (VLPs) following *in vitro* assembly

(A–D) *In vitro* assembled VLPs from purified recombinant Pr77^{Gag}-His₆-tag fusion protein expressed from AK7 in the presence of yeast tRNA; and (E, F) negative controls consisting of assembly buffer and yeast tRNA in the absence of any protein (Scale bar = 50 nm, 135,000× magnification).

Prior to *in vitro* assembly, the protein was frozen and then thawed. The efficient formation of VLPs suggests that the biophysical activity of purified protein remained intact following a freeze-thaw cycle. From these experiments, it is clear that the purified MMTV recombinant full-length Gag-His₆-tag fusion protein maintained its multimerizing/oligomerizing ability and assembled *in vitro* to form VLPs as reported previously in the case of HIV-1 and FIV (Affranchino & González, 2010; Campbell & Rein, 1999; McKinstry et al., 2014; Tanwar et al., 2017).

2.4.6 Recombinant Full-Length MMTV Pr77^{Gag}-His₆-Tagged Protein Expressed in Bacteria Can Form Virus-Like Particles

The formation of VLPs by several retroviral Gag proteins expressed in bacteria has already been established (Campbell & Vogt, 1997; Ehrlich et al., 1992; Klikova et al., 1995; M. Sakalian et al., 1996; Michael Sakalian & Hunter, 1999).

Therefore, the ability of this recombinant MMTV Pr77^{Gag} proteins (both with and without the His₆-tag), to assemble into VLPs was analyzed. TEM was performed on bacterial samples transformed with full-length MMTV Gag recombinant clone AK7 (with His₆-tag) and AK31 (without His-tag) and cultured at 28°C post-IPTG induction. Ultrathin sections (95 nm) of the bacterial pellets were negatively stained with 1% uranyl acetate followed by lead citrate and visualized on TEM. Electron micrographs of *E. coli* BL21(DE3) cells transformed with both AK7 and AK31 showed intra-cytoplasmic electron dense rings of ~55–70 nm in size, closely resembling immature VLPs (Figure 25A, B). As expected, no VLPs were observed in un-induced bacterial cells that were transformed with either the AK7- or in AK31 expression plasmids (Figure 25C, D). Similarly, no such VLP structures were observed when the cloning vector by itself was transformed into BL21(DE3) cells and induced with IPTG (data not shown). These results suggest that when expressed in bacteria, full-length MMTV Gag proteins either with or without the His₆-tag could form morphologically indistinguishable VLPs (Figure 25A, B).

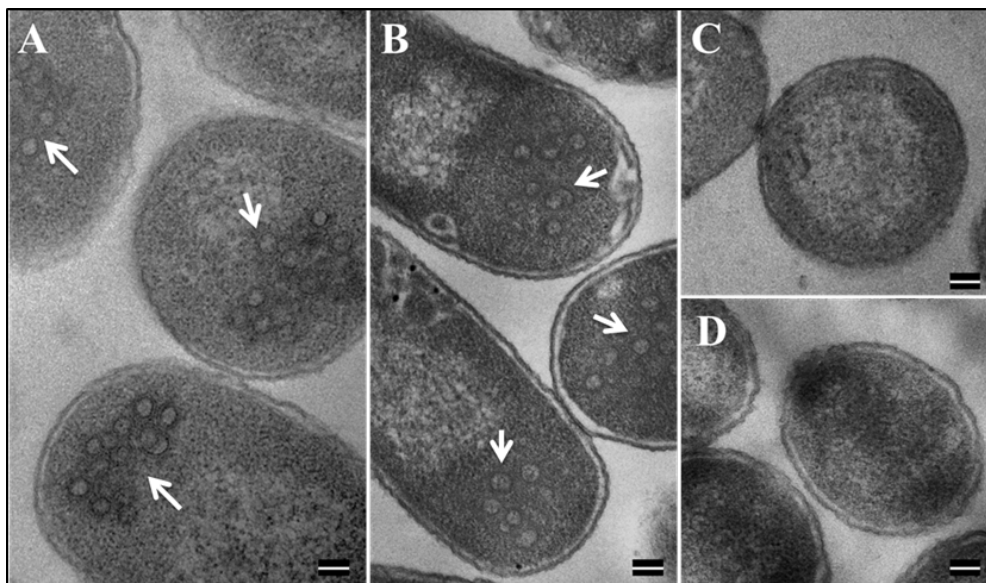


Figure 25: Formation of VLPs by recombinant Pr77^{Gag}-His₆-tag fusion protein in *E. coli* BL21(DE3)

Transmission electron micrographs showing VLPs assembled from (A) recombinant Pr77^{Gag}-His₆-tag fusion protein expressed in *E. coli* BL21(DE3) cells transformed with AK7 and (B) AK31 (without His₆ tag); (C,D) un-induced BL21(DE3) cells transformed with AK7 and AK31, respectively (Scale bar = 100 nm; 60,000 \times magnification).

2.4.7 Eukaryotically-Expressed, Full-Length Recombinant Pr77^{Gag} His₆-Tagged Fusion Protein Can Form Virus-Like Particles Competent to Package Unspliced Subgenomic RNA

Finally, it was determined that the *in vivo* expression and RNA packaging potential of MMTV Gag recombinant proteins in eukaryotic cells. Towards this end, both the His(+) and His(-) versions of the full-length MMTV gag gene were cloned into eukaryotic expression plasmid pCDNA3, creating AK13 (with His₆-tag) and AK14 (without the His₆-tag; Figure 26A; upper panel). To ensure appropriate nuclear export and translation of the MMTV Pr77^{Gag} mRNA, a 231-nucleotide long MPMV CTE was cloned downstream of the MMTV Gag stop codon in both of these clones (Figure 26A upper panel). The MPMV CTE has previously been shown to be required for the successful expression of the MMTV *gag/pol* genes from eukaryotic expression vectors in the absence of a functional MMTV Rem/RmRE transport

system (Rizvi et al., 2009). These full-length Gag expression plasmids were tested for their ability to package MMTV subgenomic RNA expressed from the transfer vector DA024 by co-transfection into the highly efficient HEK 293T cells, as described before (Rizvi et al., 2009).

Western blot analysis of cell lysates revealed successful expression of recombinant MMTV Pr77^{Gag} protein by both the His(+) and His(-) plasmids, AK13 and AK14, respectively (Figure 26B; panel I). These expressed proteins could form VLPs, as detected by their presence in the supernatants pelleted via ultracentrifugation followed by immunoblotting with α -MMTV p27 monoclonal antibody (Figure 26B; panel III). These data reveal that the inclusion of His₆-tag at the C-terminus of MMTV full-length Pr77^{Gag} did not impinge upon the expression of recombinant full-length MMTV Pr77^{Gag}, resulting in the formation of Gag VLPs in the eukaryotic cells.

To establish whether Gag VLPs formed by the recombinant MMTV full-length Pr77^{Gag} can package MMTV subgenomic RNA, RNAs from the cytoplasmic fractions of HEK293T cells as well as from the pelleted viral particles were extracted. The RNA preparations were DNase-treated to remove any contaminating plasmid DNA that may have been carried over from the transfected cultures. After confirming the absence of plasmid DNA by PCR using transfer vector RNA-specific primers (OTR671 and OTR672; data not shown), the DNase-treated-RNAs were converted into cDNAs. Furthermore, the integrity of nucleocytoplasmic fractionation was ensured by testing for the absence of unspliced β -actin mRNA in the cytoplasmic fractions by RT-PCR, as described previously (Rizvi et al., 2009; data not shown). Next, expression of the transfer vector (DA024) RNA was analyzed in the

cytoplasmic fractions by RT-PCR, which ensured that these RNAs were stably expressed and exported to the cytoplasm so that they can function as competent substrates for RNA packaging in the assembling virus particles (Figure 26B; panel IV).

Finally, the ability of the VLPs formed by the recombinant MMTV Pr77^{Gag}-His₆-tag fusion protein (AK13) as well as without His-tag (AK14) to package the DA024 transfer vector RNA was analyzed by preparing cDNAs from the RNA isolated from the pelleted VLPs. Finally, the relative packaging efficiency of the VLPs was assessed by employing the custom-designed Taqman gene expression assay to quantitate MMTV RNA packaging (Aktar et al., 2014; Mustafa et al., 2012). Results shown in Figure 26C confirm that both His(+) and His(-) Gag VLPs could successfully package MMTV subgenomic RNA into the virus particles and in proportion to the amount of corresponding Gag VLPs (Figures 26B; panel III). These results suggest that the recombinant MMTV Pr77^{Gag}-His₆-tag fusion protein expressed in eukaryotic cells is biologically active, resulting in the formation of VLPs with the capability of encapsidating subgenomic MMTV RNA.

A point to note, it was consistently observed that the amount of Gag VLPs from the His(-) vector AK14 was always less compared to the His(+) expression vector, AK13 (Figure 26B; panel III) despite their efficient expression in the cytoplasm (Figure 26B; panels I and II). This was true even when the experiments were repeated multiple times with different preparations of plasmid DNAs and having comparable transfection efficiencies from three independent experiments (SEAP relative luminescence units 1811946 for AK13 versus 1820756 for AK14). The exact reason for this differential Gag VLP formation in current experiments

remains largely unclear. Irrespective of the amount of VLPs produced by the His(+) or His(-) expression vectors, they could package MMTV subgenomic RNA efficiently and correspondingly to the amount of Gag VLPs formed (Figure 26B; panel III & Figure 26C). Thus, from a biological and functional perspective, both proteins seemed to have similar capabilities. Interestingly, when sequences for AK13 (with His₆-tag) and AK14 (without His₆-tag) were analyzed by ExPASy-Compute pI/Mw tool to calculate the theoretical isoelectric point (pI), it predicted a negligible effect (0.18) in these proteins (AK13 with His₆-tag: pI: 6.76 versus AK14 without His₆-tag: pI: 6.58), corroborating similarities in their functional observed capacities.

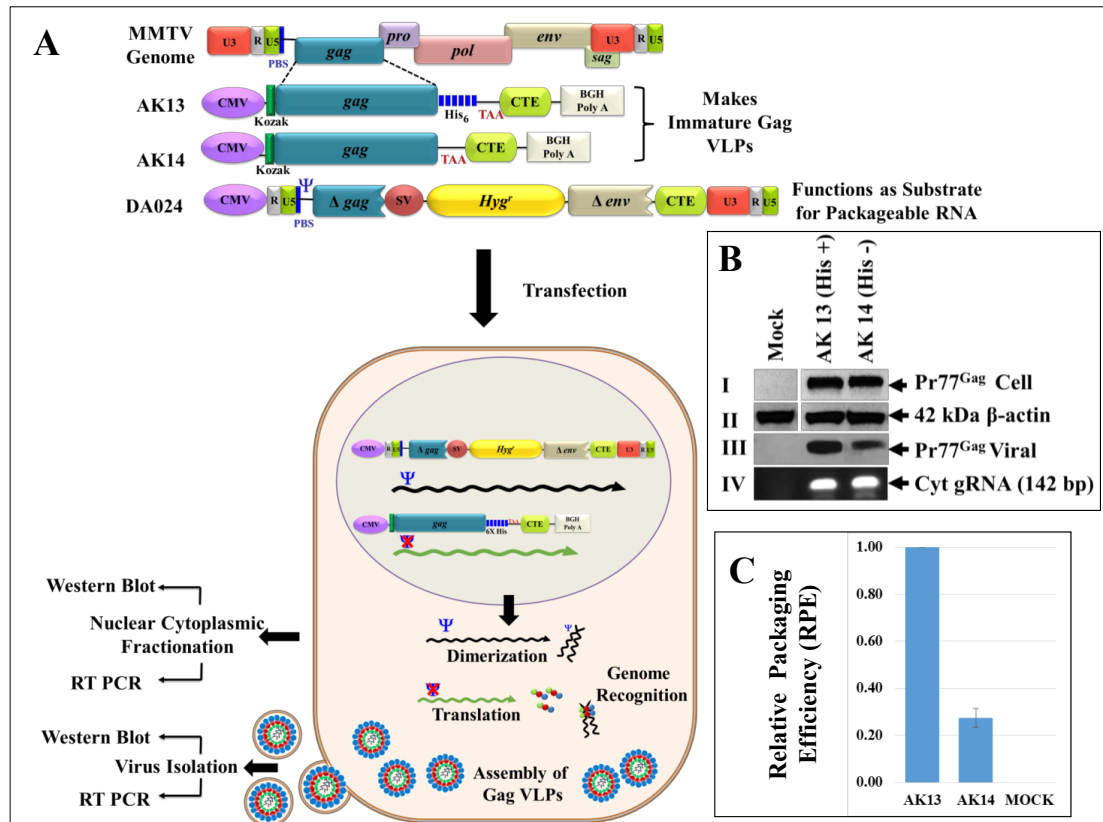


Figure 26: Schematic representation of the two-plasmid genetic complementation assay to demonstrate VLPs formation following Pr77^{Gag}-His₆-tag protein expression in eukaryotic cells and their ability to package MMTV subgenomic RNA

(A) Upper panel; MMTV full-length Gag eukaryotic expression plasmids and MMTV subgenomic transfer vector, DA024 (Rizvi et al., 2009). (A) Lower panel; Graphical representation of the MMTV two-plasmid genetic complementation assay in which VLPs produced by recombinant MMTV Pr77^{Gag} expression plasmids (AK13 and AK14) should package MMTV subgenomic transfer vector (DA024) owing to the presence of the packaging sequences (Ψ). HEK 293T cells co-transfected with the two plasmids were subjected to nucleocytoplasmic fractionation. The cytoplasmic fractions and pelleted VLPs were analyzed for transfer vector RNA expression by RT-PCR; (B) western blots performed on cell lysates and ultracentrifuged transfected culture supernatants using α -MMTV p27 monoclonal antibody (panels I and III) and α - β -actin antibody (panel II), respectively. PCR amplification of cDNAs prepared from cytoplasmic (panel IV) and viral RNA (panel V) demonstrating RNA packaging using MMTV transfer vector (DA024)-specific primers (OTR671/OTR672) to amplify a 142 bp fragment. The RNA packaging experiment was performed more than three independent times followed by its analysis by RT-PCR and a representative blot of the packaged viral RNA is shown in panel V; (C) relative RNA packaging efficiency (RPE) by AK13 and AK14 of one of the representative experiments, as measured by quantitative real time PCR. Briefly, the real time experiments were conducted in triplicates (\pm standard deviation (SD)) and the relative quantification (RQ) values obtained for the packaged viral RNA in the Gag VLPs were normalized to the cytoplasmic expression of the transfer vector RNA (DA024) for the respective clones as described previously.

2.5 Conclusions

Work presented in this study reports the successful cloning and expression of the recombinant full-length Pr77^{Gag} protein of MMTV both with and without a His₆-tag. The protein could be expressed and purified from soluble fractions of bacteria at high levels, had the ability to form VLPs *in vitro* and could also form VLPs in bacterial cells *in vivo*. VLPs formed by the recombinant full-length Gag protein in eukaryotic cells revealed their ability to recognize and encapsidate MMTV subgenomic RNA successfully, despite the presence of His₆-tag at the C-terminus. The availability of pure forms of MMTV Pr77^{Gag} should facilitate structural studies and further biochemical and functional characterization to better understand the molecular interactions that take place during RNA dimerization, packaging and virus assembly steps critical for not only understanding virus replication, but also importantly for the development of MMTV-based vectors for human gene therapy.

2.6 Funding

This research was funded primarily by a grant from the United Arab Emirates University (UAEU) Zayed Bin Sultan Center for Health Sciences (UCBR-31R123) and Program for Advanced Research (UPAR-31M233) and in part by a grant from the College of Medicine and Health Sciences (31M280) to T.A.R. A.C. and F.N.N.P. were supported by UCBR-31R123 and UPAR-31M233, respectively.

Chapter 3: A Purine Loop and the Primer Binding Site are Critical for the Selective Encapsidation of Mouse Mammary Tumor Virus Genomic RNA by Pr77^{Gag}

3.1 Abstract

Retroviral RNA genome (gRNA) harbors *cis*-acting sequences that facilitate its specific packaging from a pool of other viral and cellular RNAs by binding with high-affinity to the viral Gag protein during virus assembly. However, the molecular intricacies involved during selective gRNA packaging are poorly understood. Binding and footprinting assays on mouse mammary tumor virus (MMTV) gRNA with purified Pr77^{Gag} along with in cell gRNA packaging identified two Pr77^{Gag} binding sites constituting critical, non-redundant, packaging signals. These included: a purine loop in a bifurcated stem-loop containing the gRNA dimerization initiation site and the primer binding site (PBS). Despite these sites being present on both unspliced and spliced RNAs, Pr77^{Gag} specifically bound to unspliced RNA, since only that could adopt the native bifurcated stem-loop structure containing looped purines. These results map minimum structural elements required to initiate MMTV gRNA packaging, distinguishing features that are conserved amongst divergent retroviruses from those perhaps unique to MMTV. Unlike purine-rich motifs frequently associated with packaging signals, direct involvement of PBS in gRNA packaging has not been documented in retroviruses. These results enhance the current understanding of retroviral gRNA packaging/assembly, making it not only a target for novel therapeutic interventions, but also development of safer gene therapy vectors.

3.2 Introduction

Viruses consist of a protein shell that encloses a genome that can either be DNA or RNA. Viral structural proteins have the ability to specifically recognize their genome and “package” it into the assembling virus particles. They must incorporate their genomes into the virus particles with high specificity to ensure continuity of their life cycle. Different viral groups employ different mechanisms for packaging their genomes selectively and with high fidelity. Retroviruses belong to a special class of viruses that use RNA as their genome (full length unspliced gRNA), which harbors *cis*-acting packaging sequences (called psi, Ψ), that facilitate specific encapsidation of the gRNA (Ali et al., 2016; Comas-Garcia et al., 2016; D’Souza & Summers, 2005; Dubois et al., 2018; Johnson & Telesnitsky, 2010; Kaddis Maldonado & Parent, 2016; Lever, 2007; Mailler et al., 2016). This is despite the gRNA being in an intense competition with other cellular and viral RNAs for packaging into assembling viral particles formed by the viral protein called “Group specific antigen” (Gag; Ali et al., 2016; Comas-Garcia et al., 2016; D’Souza & Summers, 2005; Dubois et al., 2018; Johnson & Telesnitsky, 2010; Kaddis Maldonado & Parent, 2016; Lever, 2007; Mailler et al., 2016). Gag is sufficient by itself to assemble into virus-like particles (VLPs), as has been shown for a number of retroviruses such as human immunodeficiency virus type 1 (HIV-1), mouse mammary tumor virus (MMTV), Mason-Pfizer monkey virus (MPMV) and feline immunodeficiency virus (FIV (Chameettachal et al., 2018; Krishnan et al., 2019; McKinstry et al., 2014; Pitchai et al., 2018; Tanwar et al., 2017). However, once the viral nucleic acid is present in the cell, Gag can selectively encapsidate the gRNA into the assembling virus particle owing to the presence of *cis*-acting Ψ sequences

(Aronoff & Linial, 1991; Muriaux et al., 2001; Rulli et al., 2007). Such selective and faithful packaging has been linked to the presence of high-affinity Gag binding sites within the Ψ on the gRNA (Abd El-Wahab et al., 2014; Ding et al., 2020; Gherghe et al., 2010; Pitchai et al., 2021; Rein, 2020). However, the precise molecular intricacies during selective packaging of gRNA enabling Gag to bind selectively to gRNA over other viral cellular RNAs are still poorly realized.

Similar to retroviruses, ~ 8% of the human genome consists of LTR-retrotransposon which are retrovirus-like elements also known as endogenous retroviral elements (Jern & Coffin, 2008). LTR-retrotransposon are remnants of retroviruses that have had a profound effect on the evolution of the human genome. They share many functional and structural similarities with retroviruses, such as their basic replication mechanism that includes reverse transcription of gRNA into DNA followed by its integration into the host genome (Boeke & Stoye, 1997; Coffin et al., 1997). Similarly, their unspliced gRNA is used for translation of Gag/Pol proteins as well as for incorporation into the nascent virions (Butsch & Boris-Lawrie, 2002). Among retroviruses, although a *cis*-packaging mechanism had initially been suggested for human immunodeficiency virus type 2 (HIV-2), it is now generally accepted that packaging of retroviral gRNA takes place independently from Gag translation (Ni et al., 2011). It has been proposed that, in HIV-1, two pools of gRNA differing by the RNA structure (Berkhout et al., 2002; Brown et al., 2020) and/or its post-transcriptional modifications (Pereira-Montecinos et al., 2019) co-exist and that only one is packaged. Intracytoplasmic compartmentalization of the Gag and the gRNA and the Rev-dependent nuclear-export mechanism of the gRNA play a critical role in specific packaging of HIV-1 gRNA (Becker & Sherer, 2017; Blissenbach et

al., 2010; Brandt et al., 2007). A recent study suggests that the initial interaction with HIV-1 Gag and a fraction of unspliced viral transcripts occurs in the nucleus forming ribonucleoprotein (RNP) complexes. Formation of RNPs with only a fraction of gRNA could be due to the existence of ‘packageable’ and ‘translatable’ gRNA pools (Tuffy et al., 2020). Similarly, interaction between gRNA and Rous sarcoma virus (RSV) Gag takes place in the nucleus, at the sites of active viral RNA transcription (Maldonado et al., 2020). On the other hand, in *Saccharomyces cerevisiae* Ty1 and Ty3 retrotransposons, the same pool of gRNA undergoes translation and packaging (reviewed in (Pachulska-Wieczorek et al., 2016). Nevertheless, a distinct feature of all retroviruses and retrotransposons is the packaging of two copies of gRNA that are non-covalently associated into a dimer (Dubois et al., 2018; Gumna et al., 2019; Paillart, Shehu-Xhilaga et al., 2004). Indeed, gRNA dimerization and packaging are intricately related processes (Berkhout & van Wamel, 1996; Laughrea et al., 1997; Nikolaitchik et al., 2013; Paillart et al., 1996), for reviews see (Dubois et al., 2018; Paillart, Shehu-Xhilaga et al., 2004), even though the molecular mechanism(s) coupling these events remain unknown.

The specificity towards retroviral gRNA recognition by Gag is conferred by Ψ sequences located within the 5' region of the gRNA (Ali et al., 2016; Comas-Garcia et al., 2016; D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Kaddis Maldonado & Parent, 2016; Lever, 2007). Other than the Ψ , additional regions, not directly interacting with Gag may also facilitate gRNA packaging, perhaps by recruiting Gag indirectly (Smyth et al., 2015, 2018). In Ty3, sequences in both the untranslated regions (UTRs) as well as in the *pol* are necessary for packaging its gRNA, whereas in Ty1, a 144 nucleotide-long region at the 5' end has

been identified as critical for Ty1 gRNA packaging (Clemens et al., 2013; Huang et al., 2013; Malagon & Jensen, 2011). However minimal sequences required for efficient packaging of their gRNA remains largely unclear in both retroviruses and retrotransposons. Nevertheless, it is becoming increasingly clear that specific selection for packaging of retroviral gRNA over cellular and spliced RNAs is a multifaceted phenomenon that occurs in the context of the whole Gag polyprotein (Bernacchi et al., 2017; Dubois et al., 2018). Despite numerous studies, how the Gag precursor specifically recognizes its gRNA and how gRNA dimerization affects this process remains largely unclear. Therefore, in order to identify general rules that govern retroviral gRNA binding to Gag and packaging, it is important to compare the gRNA packaging process in different retroviruses.

The current study involves MMTV, which belongs to the *Betaretrovirus* genus of the *Retroviridae* family and is the etiological agent of breast cancer and at times T-cell lymphomas, in mice (Cardiff & Kenney, 2007; Medina, 2010; Mustafa et al., 2003; Smith, 2005). The biology of MMTV is being intensely studied to design MMTV-based vectors for human gene therapy due to its ability to infect non-dividing cells, the main target cell population for human gene therapy (Konstantoulas & Indik, 2014). Unlike most retroviruses including the well-studied lentiviruses, MMTV assembles intracellularly as spherical particles and migrate to the plasma membrane once assembly is completed; after maturation, MMTV displays type B morphology with eccentric cores (Coffin et al., 1997; Smith, 1978). The MMTV gRNA packaging signal comprises the entire 5' untranslated region (5' UTR) and the first 120 nucleotides (nts) of the *gag* gene (Mustafa et al., 2012; Rizvi et al., 2009; Salmons et al., 1989) and folds into several stem-loops (SLs; Figure 27A & B; Aktar

et al., 2014). Of these, the bifurcated SL4 is of particular interest since one of its apical loops contains a self-complementary sequence that has been shown to act as the dimerization initiation site (DIS), while the other apical loop consists of nine phylogenetically conserved purines (ssPurines; Figure 27B; Aktar et al., 2014) that have been proposed to be important for MMTV gRNA packaging and may constitute a Gag-binding site (Mustafa et al., 2018).

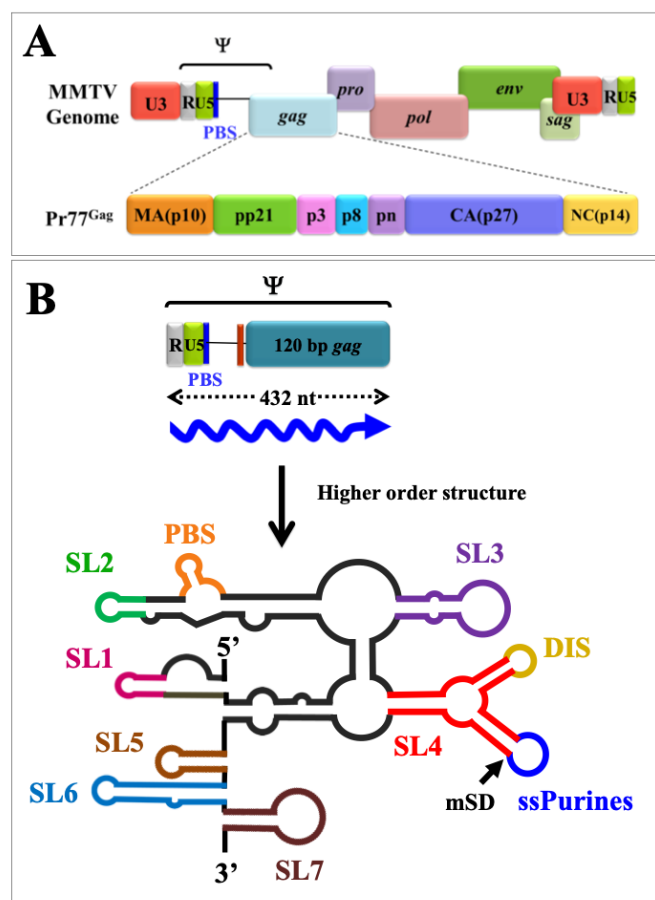


Figure 27: Schematic representation of MMTV genome, organization of different domains of full-length Gag (Pr77^{Gag}) and higher order structure of MMTV packaging signal RNA

(A) Organization of MMTV full-length genome and domain organization of MMTV Gag precursor protein. (B) Schematic representation of RNA secondary structure of MMTV packaging determinants located on the 5' end of the genome. SL1-7, stem-loops 1-7; PBS, primer binding site; DIS, dimerization initiation site; ssPurines, single stranded purines; mSD, major splice donor.

Therefore, to establish whether ssPurines or other sequences within the packaging determinants of MMTV mediate gRNA packaging, possibly by functioning as a Gag binding site(s), a combination of *in vitro* biochemical studies using recombinant full-length Pr77^{Gag} (Chameettachal et al., 2018) and biologically relevant in cell assays were performed. These results demonstrate that while ssPurines are required for high affinity Pr77^{Gag} binding, quite strikingly, the primer binding site (PBS) also critically contributes to Pr77^{Gag} binding. Furthermore, these in cell packaging data also established that ssPurines and PBS are not redundant during the MMTV life cycle, as loss of either of these Pr77^{Gag} binding sites ablates gRNA packaging. To the best of our knowledge, these results demonstrate for the first time a direct role of the PBS in retroviral gRNA encapsidation. Future studies will reveal if this feature is unique to MMTV or if it also exists in other retroviruses and/or LTR retrotransposons. Identifying the structural motifs that allow gRNA/Gag interactions is likely to offer opportunities to develop small molecule-based anti-retroviral therapeutic interventions specifically targeting virus assembly especially given the fact that retroviral Gag is the only gene necessary for virus particle formation and gRNA encapsidation for the perpetuation of the virus life cycle.

3.3 Materials and Methods

3.3.1 Nucleotide Numbers

Nucleotide numbers refer to nucleotide positions of HYB-MTV, a molecular clone created by Shackleford and Varmus (Shackleford & Varmus, 1988).

3.3.2 Expression and Purification of Full-Length MMTV Pr77^{Gag}

Full-length MMTV Gag polypeptide with a C-terminus hexa-histidine tag (Pr77^{Gag}-His₆-tag protein) was expressed, purified and characterized as previously described (Chameettachal et al., 2018).

3.3.3 Physical Characterization of Pr77^{Gag} by Dynamic Light Scattering (DLS)

Pr77^{Gag} was characterized by DLS in the storage buffer (50 mM Tris-HCl pH 8.0 and 1 M NaCl; Chameettachal et al., 2018). Briefly, the intensity of scattered light was measured using a DynaPro Nanostar (100 mW He-Ne laser; Wyatt Technologies) in a 1 μ l quartz cuvette (JC-006, Wyatt Technologies) at 20°C. Variations of the diffused light intensity were recorded at microsecond time intervals and the autocorrelation function was derived, allowing the determination of the translational diffusion coefficients (D). Assimilating the proteins in solution to spheres, the diffusion coefficients were related to the hydrodynamic radius (R_h) of the molecules in solution, *via* the Stokes-Einstein equation:

$$D = \frac{kT}{6\pi\mu R_h}$$

in which k is the Boltzmann constant, T is the temperature and μ is the viscosity of the solvent. Before sample acquisition, buffer was filtered through 0.02 μ m filters (Millex [®]) and the offset of the solvent was measured for subsequent sample data treatment.

3.3.4 Cloning, Mutagenesis, *In Vitro* Transcription and RNA Purification

Desired mutations in the ssPurines and other regions were introduced through splice overlap extension (SOE) PCR using MMTV subgenomic transfer vector DA024 (Rizvi et al., 2009) as the template and OTR 249 and OTR 552 (Appendix A) as outer sense (S) and antisense (AS) primers, respectively. Mutations were incorporated using inner primers bearing the nucleotide changes to be introduced into the template (Appendix A). Following two separate first rounds of amplification, the amplified products were mixed to allow annealing at the complementary region harboring the introduced mutation and amplified using the outer primers OTR 249 and OTR 552. The final amplified products containing flanking *SpeI* sites were cleaved and then cloned into the subgenomic transfer vector DA024 that was previously cleaved with the same restriction endonuclease. All clones were confirmed by sequencing.

Clones for *in vitro* transcription were created as previously described (Aktar et al., 2014). Briefly, the 5' end of the MMTV gRNA corresponding to nts 1-712 (R, +1) was amplified by PCR from the MMTV subgenomic transfer vector DA024 (wild type; WT) or its mutant clones as the template, using primers OTR 984 (S) and OTR 985 (AS) to insert the T7 promoter sequence at its 5' end. The flanking *HindIII* and *XmaI* sites were used to clone the amplified products into a pUC-based cloning vector (pIC19R; Marsh et al., 1984). The resulting clones were confirmed by sequencing. Details of the primers used for cloning are provided in the Appendix A.

Mutations introduced to study in cell biological assays were cloned into the subgenomic transfer vector DA024, referred to as WT. Whereas, for *in vitro*

biochemical assays, the same mutations were cloned into a T7 expression plasmid SA35 (referred to as WT and described earlier; Aktar et al., 2014). To allow ease of understanding, names of the mutations introduced into SA35 were used throughout the manuscript, rather than the actual names of the vectors/clones since the same mutations were introduced into both SA35 and DA024.

For *in vitro* transcription, clones containing the T7 promoter were linearized with *Sma*I, extracted using Roti®Aqua for nucleic acid extraction as per manufacturer's instructions (Carl Roth) and resuspended in 30 µl Milli-Q water. Transcription was performed using MEGAscript T7 transcription kit according to manufacturer's instructions (Thermo Fisher Scientific) and following DNase treatment, RNA was extracted, precipitated and resuspended in 500 µl Milli-Q water. The *in vitro* transcribed RNAs were then purified by gel filtration chromatography using a TSKgel G4000SW column (Tosoh Bioscience) with running buffer containing 200 mM sodium acetate and 1% methanol. The eluted fractions under a same peak were pooled, ethanol precipitated and resuspended in 150 µl of Milli-Q water. The RNA integrity was checked using denaturing polyacrylamide gel electrophoresis (PAGE).

Internally labeled RNAs were prepared by performing *in vitro* transcription in the presence of [α -³²P]-ATP, as described earlier (Paillart, Berthoux, et al., 1996; Paillart et al., 1994; Paillart, Skripkin, et al., 1996; Sinck et al., 2007). Following DNase treatment, the labeled RNAs were electrophoresed on denaturing polyacrylamide gels (6-8%, 8M urea), bands were excised and extracted in 300 µl of buffer containing 500 mM ammonium acetate, 1 mM EDTA and 0.1% SDS

overnight at 4°C. The RNAs were then ethanol precipitated and resuspended in 10 µl Milli-Q water.

3.3.5 Band-Shift and Band-Shift Competition Assays

Samples for band-shift assays were prepared by denaturing 50,000 cpm of internally labeled RNA together with 10 nM of the cognate unlabeled RNA (in order to favor RNA dimerization, which might be critical for Pr77^{Gag} binding) and 0.4 µg of yeast tRNA at 90°C for 2 min followed by chilling on ice for 2 min. The denatured RNAs were next allowed to re-fold by incubation in 1X folding buffer (30 mM Tris pH 7.5, 300 mM NaCl, 5 mM MgCl₂, 5 units of RNasin, 0.01% Triton-X 100, total volume 6 µl) at 37°C for 30 min. Next, increasing concentrations of Pr77^{Gag} were mixed with the refolded RNA (Pr77^{Gag} was diluted in 30 mM Tris pH 7.5, 300 mM NaCl, 5 mM MgCl₂, 10 mM DTT and 0.02 mg/ml BSA in a final volume of 14 µl). The mixture was then incubated at 37°C for 30 min for binding, followed by incubation on ice for 30 min. Samples were separated using 1% agarose gel with electrophoresis performed in TBM buffer (0.5X TB, 0.1mM MgCl₂) at 150 V for 4 hr at 4°C. The gels were fixed in 10% trichloroacetic acid (TCA) for 10 min, dried under vacuum for 1 hr and analyzed using a FLA 5000 (Fuji) scanner. Quantitative analysis of the bands was performed using ImageQuant software (Abd El-Wahab et al., 2014). The Hill slope was identified by applying the values to Hill equation:

$$Y = \frac{B_{max} \times X^h}{(K_d^h + X^h)}$$

using the GraphPad Prism 5 software. B_{max} is the maximum specific binding, K_d is the dissociation constant and h is the Hill slope.

For band-shift competition assays, internally labeled WT RNA (25,000 cpm) and 10 nM unlabeled WT RNA (SA35; Figure 30A) along with increasing concentrations of unlabeled competitor RNAs (up to 400 nM) were denatured at 90°C followed by cooling on ice for 2 min. RNAs were refolded by incubation at 37°C for 30 min in folding buffer (30 mM Tris pH 7.5, 300 mM NaCl, 5 mM MgCl₂, 0.01% Triton X-100 and 5 units RNasin) in a volume of 10 µl. In the meantime, a 20 µM stock of Pr77^{Gag} was diluted in a buffer containing 30 mM Tris pH 7.5, 300 mM NaCl, 5 mM MgCl₂, 10 mM DTT and 0.02 mg/ml BSA in a volume of 5 µl. The refolded RNAs were then mixed with the diluted protein (final concentration of 800 nM) and incubated at 37°C for 30 min for binding and then on ice for 30 min for stabilization. The reaction mixtures were analyzed on 1% agarose gels with electrophoresis performed in TBM buffer (0.5X TB, 0.1 mM MgCl₂) at 150 V for 4 hr at 4°C. The gels were then fixed and analyzed as described above.

3.3.6 Filter-Binding Assay

Samples for filter-binding assays were prepared as described above by denaturing 25,000 cpm of internally labeled RNA, together with 5 nM of the cognate unlabeled RNA and 0.4 µg of yeast tRNA at 90°C for 2 min followed by chilling on ice for 2 min. The denatured RNAs were then incubated at 37°C, mixed with increasing concentrations of Pr77^{Gag} and incubated for an additional 30 min period. The RNA/protein complexes were stabilized at 0°C for 30 min and filtered through a nitrocellulose membrane (0.45 mm, Bio-Rad) using a dot blot apparatus. Membranes

were pre-wetted with Tris buffered saline solution (30 mM Tris pH 7.5, 500 mM NaCl) and pre-washed once with 100 ml of buffer (30 mM Tris pH 7.5, 300 mM NaCl). After sample filtration, wells were washed 3 times with 60 ml of cold buffer (30 mM Tris pH 7.5, 300 mM NaCl), the membranes were removed from the filtration apparatus and dried on air. The filters were exposed with an Imaging Plate (Fujifilm), scanned with a FLA 5000 (Fuji) scanner and quantified with the ImageQuant software (Abd El-Wahab et al., 2014).

3.3.7 Footprinting Experiments using hSHAPE in the Presence and Absence of Pr77^{Gag}

High-throughput selective 2'-hydroxyl acylation analyzed by primer extension (hSHAPE) was performed on the WT (SA35) MMTV RNA either in the presence (4 μ M) or absence of Pr77^{Gag}, as previously described (Aktar et al., 2014). Briefly, the target (100 nM SA35 RNA) and the competitor (400 nM spliced *env* mRNA; AK29) were mixed and denatured at 90°C for 2 min followed by cooling on ice for 2 min. The RNAs were then refolded in a folding buffer (30 mM HEPES-KOH pH 8.0, 300 mM KCl, 5 mM MgCl₂) along with 5 units of RNasin and 2 μ g total yeast tRNA in 10 μ l reaction volume for 30 min at 37°C. The 20 μ M Pr77^{Gag} stock was diluted in the same folding buffer and mixed with the refolded RNA to a final concentration of 4 μ M; RNA in the folding buffer without Gag was used as a control. The mixture was incubated at 37°C for 30 min to allow for protein binding and then stabilized by cooling on ice for 30 min. This was followed by modification of the RNA samples by 100 mM BzCN dissolved in anhydrous DMSO or DMSO alone as a control. The modified and control RNAs were extracted using Roti[®]Aqua, ethanol precipitated and the pellets were resuspended in 7 μ l Milli-Q water.

Finally, the modified and control RNAs were reverse transcribed using VIC-labeled primers (OTR 10 & OTR 14; Appendix A) in the presence of 1X RT buffer, 0.75 mM of dNTPs and 2 units of AMV RT at 42°C for 20 min and 50°C for 30 min, followed by 60°C for 10 min. For sequencing reactions, reverse transcription was performed using two sets of NED-labeled primers OTR 11 and OTR 15 in a reaction mixture containing 1X RT-buffer, 10 μ M ddGTP, 6 μ l of G10 (0.15 mM dGTP, 0.6 mM dATP, 0.6 mM dCTP, 0.6 mM dTTP) and 4 units of AMV RT. The cDNAs thus generated from both reactions were mixed together and sequenced using an Applied Biosystems 3130xl genetic analyzer as described earlier (Aktar et al., 2014; Mustafa et al., 2018). The electropherograms obtained were analyzed with the software QuShape (Karabiber et al., 2013). Normalized SHAPE reactivity data from at least 3-4 independent experiments in the absence of Pr77^{Gag} were used as pseudo-energy constraints to fold the RNA secondary structure of the MMTV packaging signal using the program RNAstructure version 6.1 (Reuter & Mathews, 2010b). RNA structures were then drawn using VARNAv3-93 and SHAPE reactivity was incorporated, as described previously (Darty et al., 2009). The SHAPE reactivity obtained in the presence of 4 μ M Pr77^{Gag} was applied onto the RNA structure obtained using reactivity in the absence of Gag. For determination of statistically significant differences between SHAPE reactivities in the absence and presence of Pr77^{Gag}, the standard paired, two-tailed Student's *t*-test was performed. Variation in SHAPE reactivity by at least 1.5-fold with a *p*-value \leq 0.05 was considered significant.

High-throughput selective 2'-hydroxyl acylation analyzed by primer extension in the absence of Pr77^{Gag} was performed on the spliced RNAs AK29

(*envelope; env*) and AK30 (*superantigen; sag*) as described above except that only yeast tRNA was used as the competitor. For AK29 RNA, Splice_1_1 and Splice_3_1 oligos were used for reverse transcription and Splice_1_2 and Splice_3_2 primers were used for sequencing. Similarly, for AK30 RNA, Splice_1_1 and Splice_2_1 oligos were used for reverse transcription and Splice_1_2 and Splice_2_2 primers were used for sequencing (Appendix A).

3.3.8 In Cell Genetic Complementation Assay

A previously described three-plasmid genetic complementation assay was used to study the gRNA packaging and RNA transduction efficiencies of the ssPurines mutations (Figure 28; Mustafa et al., 2012; Naldini et al., 1996; Rizvi et al., 2009). The MMTV subgenomic transfer vector (DA024) containing the minimum *cis*-acting sequences required for RNA packaging, reverse transcription and integration was used as a source of the RNA to be packaged in the vesicular stomatitis virus envelope glycoprotein (VSV-G) pseudotyped particles produced by JA10 and MD.G (Mustafa et al., 2012; Naldini et al., 1996; Rizvi et al., 2009). Additionally, DA024 also expresses the *hygromycin B phosphotransferase* gene (*hyg^r*), which helps to monitor the effect of the mutations introduced into this transfer vector following transduction of the target cells by the viral RNA in a single round of replication assay (Figure 28).

Virus particles containing the WT or mutated RNAs were produced in human embryonic kidney (HEK) 293T producer cells and purified as described before (Aktar et al., 2014; Mustafa et al., 2018). A fraction of the supernatant was used to infect the human cervical cancer cells (HeLa) to determine the RNA transduction

efficiency of the packaged RNA. This was achieved using the single round of replication assay followed by selecting the transduced cells with medium containing 200 $\mu\text{g/ml}$ of hygromycin B (Hyclone). Transduced cells containing the packageable transfer vector RNA expressing the selectable marker resulted in hygromycin resistant colonies that were counted and represented as colony forming units (CFU/ml). The CFU/ml values were then normalized to the transfection efficiency and divided by the WT values to represent the transduction of the mutant RNAs relative to the WT (relative CFU/ml). For determination of statistically significant differences between the WT and the mutations introduced, the standard paired, two-tailed Student's *t*-test was performed.

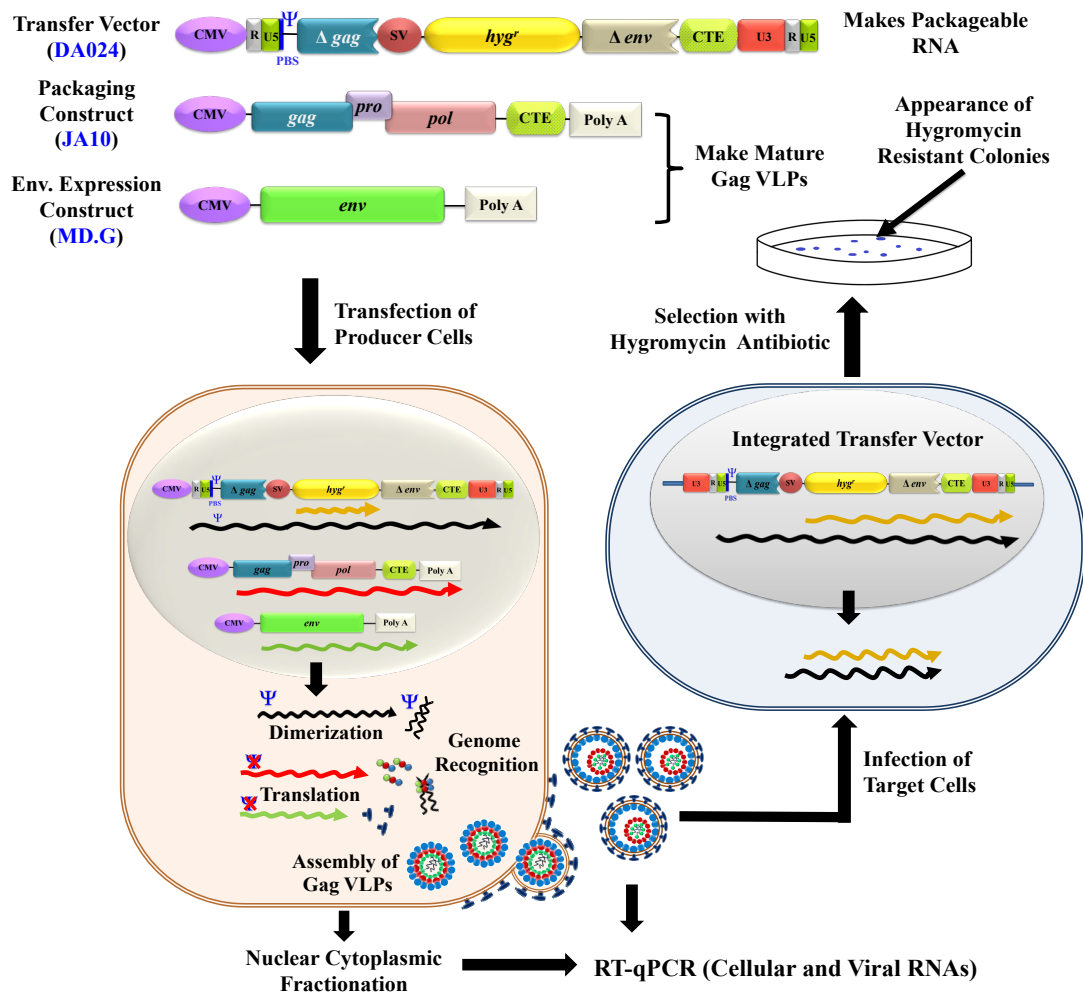


Figure 28: Design and rationale of the MMTV three-plasmid genetic complementation assay

Virus particles produced by the MMTV Gag/Pro/Pol expression plasmid (JA10) and pseudotyped by vesicular stomatitis virus envelope glycoprotein (VSV-G) expressed by MD.G allow packaging of MMTV subgenomic transfer vector (DA024) RNA due to the presence of the packaging signal (Ψ). 293T cells that have been co-transfected with the 3 plasmids produce infectious virus particles and RNAs from these cells are fractionated into nuclear and cytoplasmic fractions. The cytoplasmic fractions are analyzed for proper RNA expression and the virus particles harvested from the transfected 293T cells are tested for the amount of RNA packaged using RT-qPCR. Viral supernatants are also used to infect target HeLa cells to study propagation of the packaged RNA in a single round of replication assay. After infection, target cells are selected with media containing hygromycin B antibiotic, allowing only those cells to survive which have been transduced by the RNA since the packaged RNA contains the *hygromycin resistance* gene cassette.

3.3.9 Nucleocytoplasmic Fractionation, Isolation of RNA and cDNA Preparation

Nuclear and cytoplasmic fractions were separated from transfected cells as described previously (Mustafa et al., 2018). Cellular RNAs from the cytoplasmic fractions and packaged viral RNA from the pelleted viral particles were isolated as described earlier (Mustafa et al., 2018). The extracted RNAs were DNase treated and tested for any residual contaminating plasmids by performing conventional PCR using vector-specific primers OTR 1391(S) and OTR 1392 (AS; Appendix A). DNase free RNA samples were reverse transcribed into cDNA by M-MLV RT (Promega) and amplified to determine the quality of the cDNA samples. A multiplex RT-PCR specific for unspliced β -actin (OTR 582; S and OTR 581; AS; Appendix A) and 18S rRNA (18S Quantum competitor control, Ambion) was performed to ensure the quality of the nucleocytoplasmic fractionation, as described previously (Mustafa et al., 2018). Refer to Appendix A for details of primers.

3.3.10 Quantitative RT-PCR (RT-qPCR) for Transfer Vector RNA Packaging Efficiency

Quantitation of the transfer vector RNAs expressed in the cytoplasm and packaged into the pelleted viral particles was accomplished using a Taqman assay and the relative packaging efficiency (RPE) for each mutation was determined as described previously (Aktar et al., 2014; Mustafa et al., 2012, 2018). Once again, the standard paired, two-tailed Student's *t*-test was performed for determination of statistically significant differences in RPE between the WT and the mutations introduced.

3.3.11 *In Vitro* Dimerization Assay

In vitro dimerization assays were performed using RNAs that were *in vitro* transcribed using MEGAscript™ T7 transcription kit (Invitrogen). The *in vitro* transcribed RNAs were purified using MEGAclean™ kit (Invitrogen) and ethanol precipitated as described previously. Briefly, 300 nM RNAs were denatured at 90°C and incubated on ice for 2 min followed by refolding at 37°C for 30 min in either the dimer (30 mM Tris pH 7.5, 300 mM NaCl, 5 mM MgCl₂) or monomer (30 mM Tris pH 7.5, 300 mM NaCl, 0.1 mM MgCl₂) buffers. The samples were then electrophoresed in TBM buffer (50 mM Tris, 45 mM boric acid, 0.1 mM MgCl₂) at 4°C or TBE buffer (50 mM Tris, 45 mM boric acid, 1 mM EDTA) at 25°C using agarose gels prepared in respective buffers. Density of the dimeric and monomeric RNA species was quantitated using ImageQuant software and dimerization efficiency was calculated by dividing intensity of dimeric RNA band divided by intensity of total RNA bands (i.e., sum of intensities of dimer and monomer bands) and expressed as dimerization relative to the WT for each mutation introduced.

3.4 Results

3.4.1 Physical Characterization of Functional Pr77^{Gag} Polyprotein

Recently, the purification of recombinant MMTV Pr77^{Gag} polyprotein containing His₆ tag at the C-terminus expressed in *E. coli* (Chameettachal et al., 2018). The functionality of the protein was confirmed by the formation of VLPs both in cells and *in vitro* was reported. Moreover, VLPs produced in eukaryotic cells by Pr77^{Gag}-His₆ fusion protein were found to be competent to efficiently package MMTV gRNA (Chameettachal et al., 2018). The protein preparation was >95% pure

based on the ratio of UV absorbance at 260 nm and 280 nm. DLS experiments revealed that the mean hydrodynamic radius (Rh) based on number distribution was estimated to be ~ 6.00 nm which corresponds to Pr77^{Gag} trimers (Figure 29).

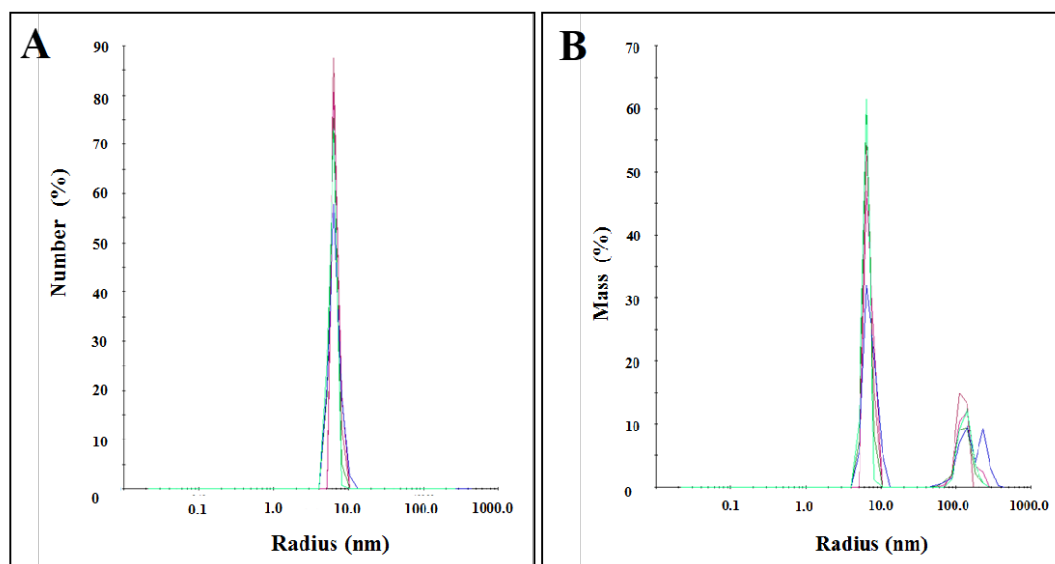


Figure 29: Dynamic light scattering of purified Pr77^{Gag} in the RNA binding buffer

(A) Protein number *versus* hydrodynamic radius (Rh) distribution and (B) Protein mass *versus* hydrodynamic radius (Rh) distribution.

3.4.2 Pr77^{Gag} Binds Specifically to Unspliced MMTV gRNA

First, to check whether the binding of the Pr77^{Gag} to the WT RNA was specific to the unspliced gRNA or whether it could bind to the spliced RNAs produced by MMTV, the differential binding ability of Pr77^{Gag} to unspliced gRNA and spliced *env* and *sag* RNAs using band-shift competition assays was tested. Towards this end, *in vitro* transcription of the first 712 nts of unspliced gRNA from clone SA35 was performed, the spliced *env* and *sag* RNAs were *in vitro* transcribed from clones AK29 and AK30, respectively, while maintaining equal lengths of the RNAs (Figure 30A).

First, binding of gRNA to increasing concentrations of Pr77^{Gag} was analyzed (Figure 30B). Shifted bands corresponding to RNA-protein complexes were clearly visible at 500 nM Pr77^{Gag} and above and the shift increased with increasing protein concentrations, revealing formation of several gRNA-Pr77^{Gag} complexes with different stoichiometry (Figure 30B). After quantification of the gels, the data fit to Hill equation (Figure 30C) and the Hill slope obtained was 3.0 ± 0.4 (mean \pm SD) indicating that binding of Pr77^{Gag} to gRNA is cooperative and that at least three Pr77^{Gag} molecules bind per RNA strand. The apparent K_d obtained was 480 ± 26 nM (mean \pm SD).

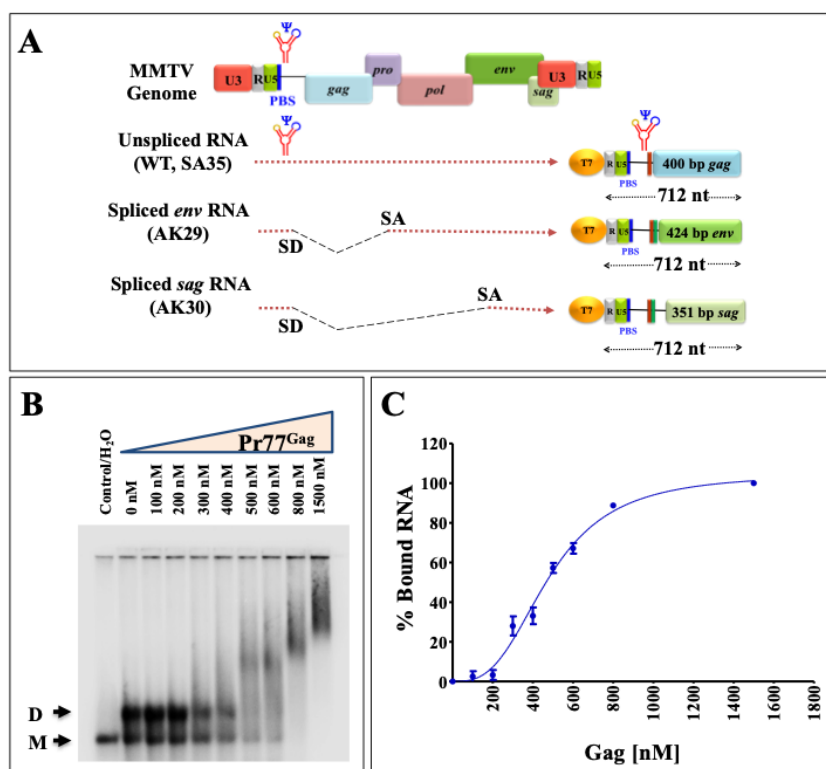


Figure 30: The MMTV packaging signal RNA binds to Pr77^{Gag}

(A) Scheme of the 712 nts long *in vitro* transcribed unspliced gRNA (SA35), spliced *env* (AK29) and spliced *sag* (AK30) RNAs used for band-shift competition assays. In MMTV, the major 5' splice site is located exactly at the 3' end of the ssPurines and the bifurcated SL4 structure can thus not exist in the spliced *env* and *sag* RNAs. (B) Representative gel of a band-shift assay using radiolabeled MMTV gRNA and increasing concentrations of Pr77^{Gag}. (C) Saturation plot fit to Hill equation based on the quantification of bands obtained in the band-shift assay. Best fit was obtained with plateau = 105 ± 6 %, Hill coefficient = 3 ± 0.4 and $K_d = 480 \pm 26$ nM (mean \pm SD); $R^2 = 0.99$.

Next, band-shift competition experiments were carried out in the presence of 800 nM Pr77^{Gag} since an almost complete shift of radiolabeled unspliced gRNA (< 1 nM) was observed at this concentration (Figure 31B). To determine whether the higher mobility-shifted complex was specific to the gRNA, increasing concentrations of unlabeled unspliced gRNA were added to the labeled unspliced gRNA as a competitor (unlabeled SA35; Figure 31A). A gradual reduction and downward shifting of the Gag-bound RNA complex as the concentration of the unlabeled RNA increased, with the reappearance of the dimeric RNA was observed. This indicates that the unlabeled gRNA was able to displace the labeled gRNA from the protein-RNA complex and the displacement was maximum at a concentration of 400 nM of competitor RNA (Figure 31A; lane 8 & Figure 31D). However, when *env* or *sag* spliced RNAs were added as competitors, they could not efficiently compete with the labeled unspliced gRNA for Pr77^{Gag} binding. While the slow-migrating complexes gradually disappeared, an RNA-protein complex migrating immediately above the position of the RNA dimer persisted even at the highest competitor concentrations (Figure 31B & 31C; lanes 4-8). This is clearly visible in Figure 31D that shows quantitation of the experiments shown in Figure 31A, 31B & 31C. Addition of Proteinase K to Gag-RNA complexes resulted in the re-appearance of labeled dimeric RNA (Figure 31A, 31B & 31C; lane 9), confirming that the shift observed when spliced viral RNAs were used as competitors was indeed due to the protein-RNA interactions (Figure 31B, 31C & 31D; lanes 4-8). The interpretation of these experiments is that, similarly to what previously reported for HIV-1 (Abd El-Wahab et al., 2014), both gRNA and spliced viral RNAs harbor unspecific low affinity binding sites and that, in addition gRNA harbors at least one specific high affinity

binding site; spliced viral RNAs are able to displace Pr77^{Gag} from the gRNA low affinity binding sites, but not from the specific high affinity binding site(s).

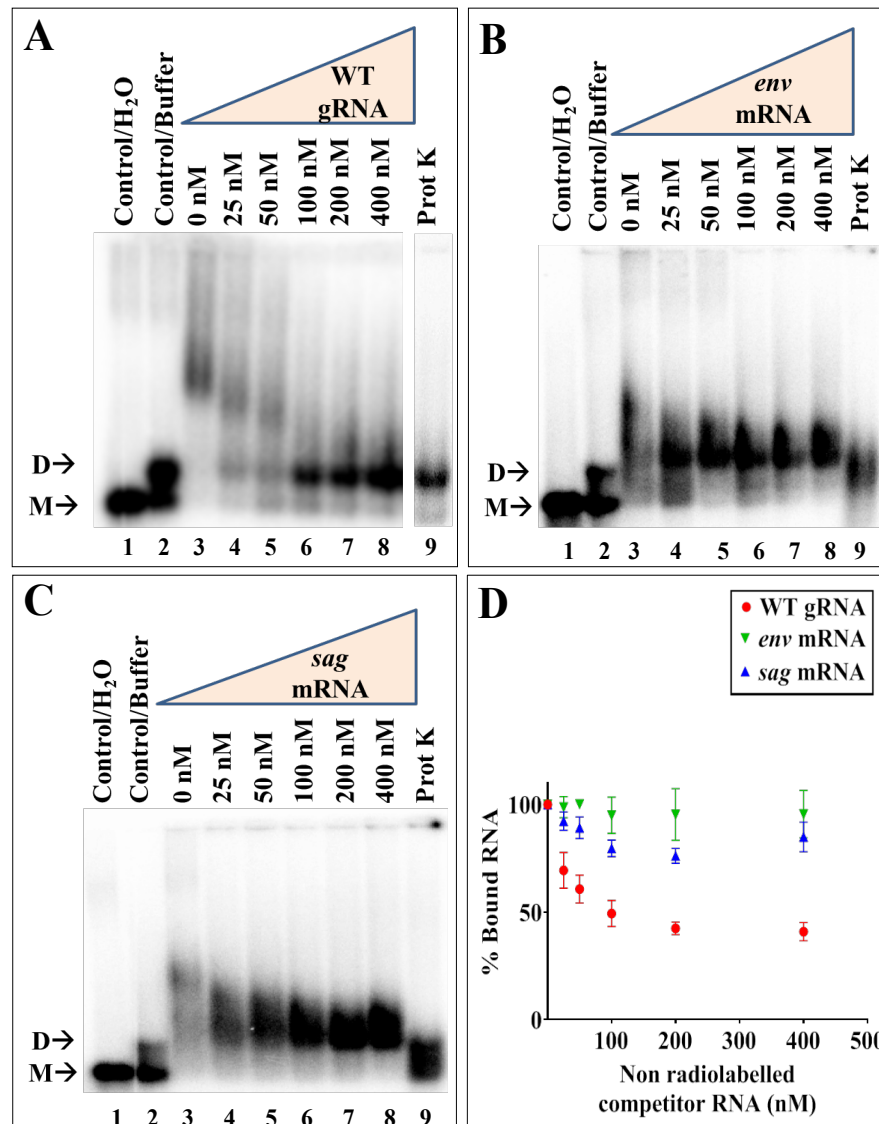


Figure 31: Unspliced genomic and spliced RNAs binds differentially to MMTV Pr77^{Gag}

(A, B & C) Differential binding specificity of unspliced WT gRNA, spliced *env* and *sag* RNAs to full-length MMTV Pr77^{Gag} as analyzed by band-shift competition assay. A radiolabeled 712 nts long unspliced WT gRNA and 800 nM MMTV Pr77^{Gag} were incubated with increasing concentrations of unlabeled unspliced WT gRNA, or spliced *env* or *sag* RNAs as competitors, respectively. The last lane in each gel shows the Proteinase K treated reaction mixture containing unspliced WT gRNA, Pr77^{Gag} and 400 nM of respective competitor RNAs. M and D indicate the monomeric and dimeric RNA forms, respectively. (D) Quantification of the gels showing the bound fraction of RNA in Gag-RNA complexes. WT, wild type.

To confirm the differential binding of Pr77^{Gag} to unspliced genomic and spliced RNAs by a different technique, direct filter-binding assays, not involving competition between several RNA species, were performed using radiolabeled WT (SA35) or spliced (AK29 & AK30) RNAs in the presence of increasing concentrations of Pr77^{Gag}. Compared to WT unspliced gRNA (SA35), the spliced (*env* and *sag*) RNAs showed a drastic reduction in their affinity for Pr77^{Gag}, further stressing the existence of high affinity binding site exclusively on the unspliced gRNA (Figure 32).

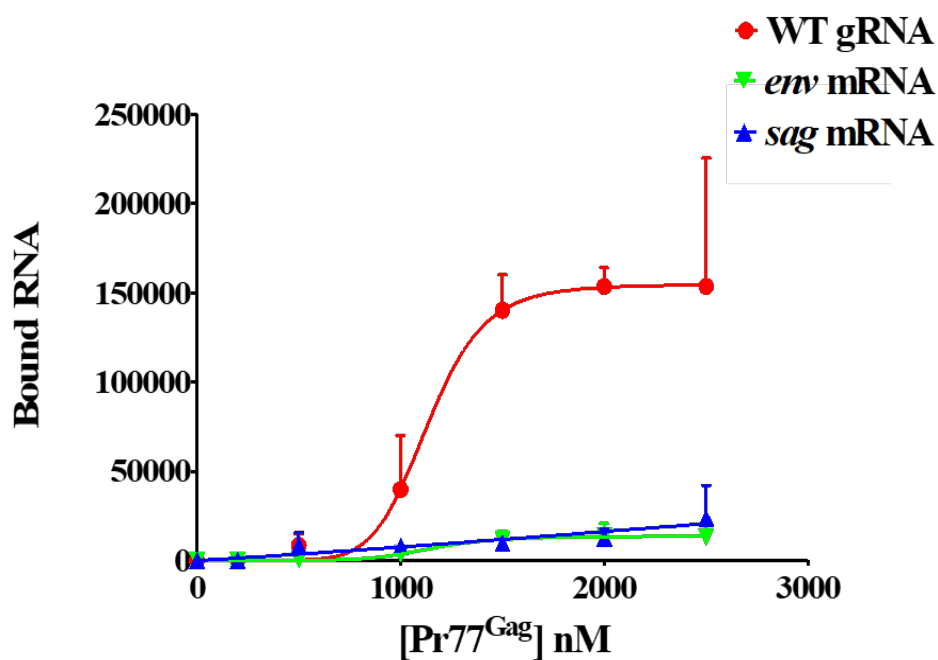


Figure 32: Binding of unspliced genomic and spliced RNAs to MMTV Pr77^{Gag} analyzed by filter-binding assay

The membrane-bound radioactivity of WT gRNA and spliced viral RNAs was quantified at increasing concentrations of MMTV Pr77^{Gag}. The experimental data points were fit with the Hill equation. WT, wild type.

3.4.3 Single-Stranded Purines (ssPurines) are Crucial for Pr77^{Gag} Binding *In Vitro*

To characterize the role of ssPurines in Pr77^{Gag} binding, a series of mutations were introduced into the ssPurines loop of SL4 (Figure 33A) and the binding specificity of Pr77^{Gag} for each mutation was analyzed by band-shift competition assay, as described above. Competitor RNA transcribed from AK18, in which ssPurines were substituted with complementary pyrimidines (Figure 33A), showed a protein displacement pattern similar to spliced RNAs, indicating that this RNA was unable to displace labeled WT RNA from the Pr77^{Gag}-RNA complex (Figure 33B). Such a loss of Pr77^{Gag} binding ability could not be attributed to structural changes as it has recently been shown that substitution of ssPurines with the pyrimidines does not alter the higher order structure of this RNA (Mustafa et al., 2018).

In order to determine the effects of the sequence of ssPurines (5' GGAGAAGAG 3') on Gag binding, a mutation (AK20) was created in which 6 of the 9 purines were substituted with other purines (5' GAGAAGAGG 3'; the substituted nucleotides are underlined; Figure 33A). In mutation AK50, the internal AAG motif within the ssPurines was substituted with UUC (5' GGAGUUCAG 3'; substituted nucleotides are underlined; Figure 33A). Band-shift competition assays performed using mutant RNAs (AK20 & AK50) showed binding ability of Pr77^{Gag} similar to WT RNA (Figure 33B & D). Substitution of internal GAAG purines with CUUC in AK44 (5' GGACUUCAG 3'; Figure 33A) or maintaining GAAG while substituting the flanking sequences in ssPurines in AK45 (5' CCUGAAGUC 3'; Figure 33A), showed a slightly reduced ability to compete with WT RNA for Pr77^{Gag} binding (Figure 33C).

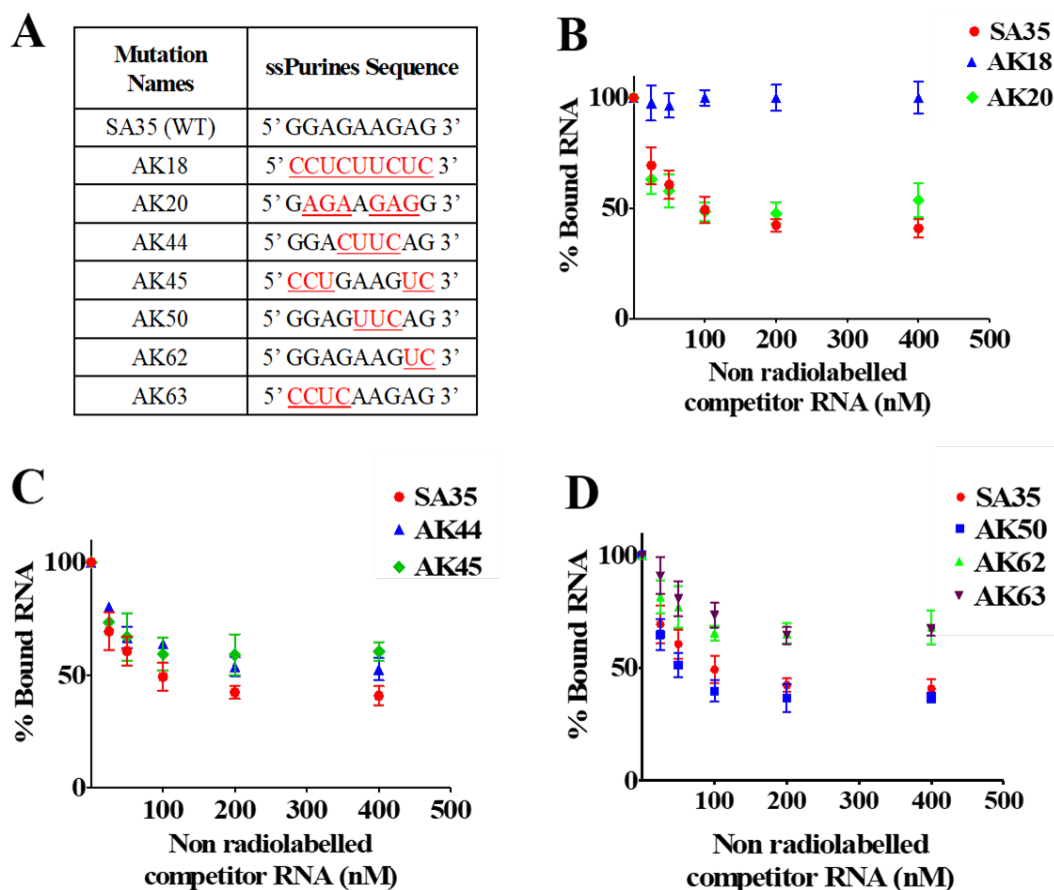


Figure 33: Differential binding abilities of ssPurines and its mutant RNAs to MMTV Pr77^{Gag}

(A) Nature of mutations introduced in ssPurines that were used in band-shift competition assays. The red and underlined nucleotides represent substitutions introduced. (B, C & D) Data generated from quantification of gels analyzing the binding specificity of the listed ssPurine-mutant RNAs to full-length MMTV Pr77^{Gag} analyzed by band-shift competition assays. A radiolabeled 712 nts long unspliced WT gRNA and 800 nM MMTV Pr77^{Gag} were incubated with increasing concentrations of unlabeled competitor mutant RNAs. WT, wild type.

In contrast, substitution of the first 4 purines (AK63, 5' CCUCAAGAG 3'; Figure 33A) or the last 2 purines (AK62, 5' GGAGAAGUC 3'; Figure 33A) with complementary pyrimidines, resulted in RNA substrates with lower ability to displace the Gag protein from the WT RNA, albeit some displacement was still observed (Figure 33D).

To further demonstrate the *in vitro* binding of Pr77^{Gag} to ssPurines in a non-competitive experimental setup, filter-binding assay using radiolabeled WT (SA35; 5' GGAGAAGAG 3') or ssPurine mutant RNAs was performed (Figure 33A). The mutations AK20 (containing 6 purine substitutions with other purines; 5' GAGAAGAGG 3') and AK50 (5' GGAGUUCAG 3'; 3 purines substituted with pyrimidines) revealed slightly reduced levels of binding to Gag compared to the WT gRNA (Figure 34A & 34C). Unexpectedly, mutation AK18 (ssPurines substituted with complementary pyrimidines) showed moderate binding affinity to Pr77^{Gag} (Figure 34A), in contrast to the competition assay which revealed that AK18 is a non-competitor compared to WT gRNA (Figure 33B). Such a discrepancy could be due to the differences in the two experimental conditions employed. On the other hand, mutations AK44 (5' GGACUUCAG 3') and AK45 (5' CCUGAAGUC 3') containing partial substitution of purines to pyrimidines showed a slightly reduced competition to WT gRNA (Figure 33C) and moderate affinity to Pr77^{Gag} in filter-binding assays (Figure 34B). Substitution mutations AK62 (5' GGAGAAGUC 3') and AK63 (AK63, 5' CCUCAAGAG 3') exhibited a drastically reduced ability to bind to Pr77^{Gag} (Figure 34C), further validating results of the competition assays (Figure 33D). This confirmed that the two stretches of purines: i) GGAG at the 5' end and ii) AG at the 3' end of ssPurines are important for RNA-Gag interactions.

To ensure that the introduced mutations in ssPurines somehow did not compromise the ability of the mutant RNAs to dimerize (which in turn may result in poor affinity for Gag), *in vitro* dimerization assays were performed on three crucial mutations (AK18, AK62 and AK63). The results obtained revealed that all the three

mutations maintained dimerization abilities similar to that of the WT (SA35) RNA (Figure 35).

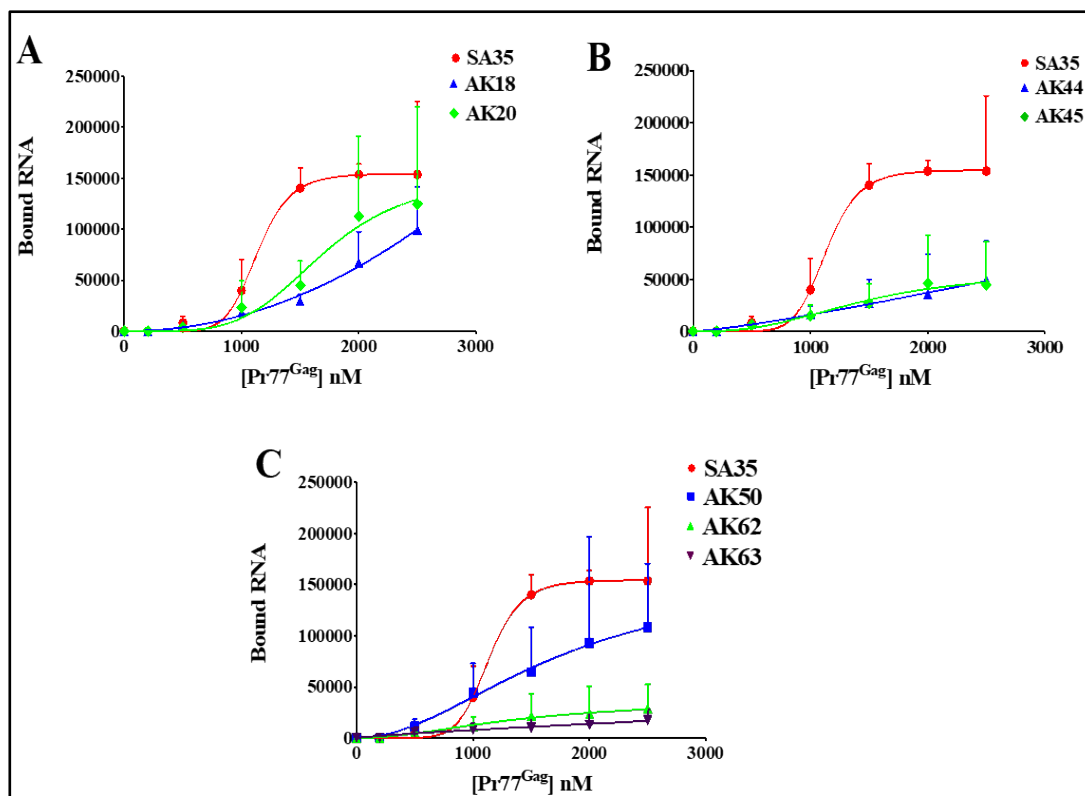


Figure 34: Binding of ssPurines and its mutant RNAs to MMTV Pr77^{Gag} analysed by filter-binding assay

The membrane-bound radioactivity of WT gRNA and selected mutant RNAs was quantified at increasing concentrations of MMTV Pr77^{Gag}. The experimental data points were fit with the Hill equation. WT, wild type.

Together, these results validate the current hypothesis that ssPurines (5' GGAGAAGAG 3') act as a high affinity binding site for Pr77^{Gag} and the results observed were not due to effects of the introduced mutations on gRNA dimerization. Notably, the 5' GGAG and 3' AG nucleotides are important for retaining high affinity binding (mutations AK50, AK62 and AK63, Figure 33D & 34C), while the remaining purines (AAG; AK50; Figure 33D & 34C) are less critical. Furthermore,

comparison of the SA35 (WT) and AK20 sequences suggest that the identity of the first and last nucleotides in the SL4 ssPurine loop is a key determinant for Pr77^{Gag} binding (Figure 33B & 34A).

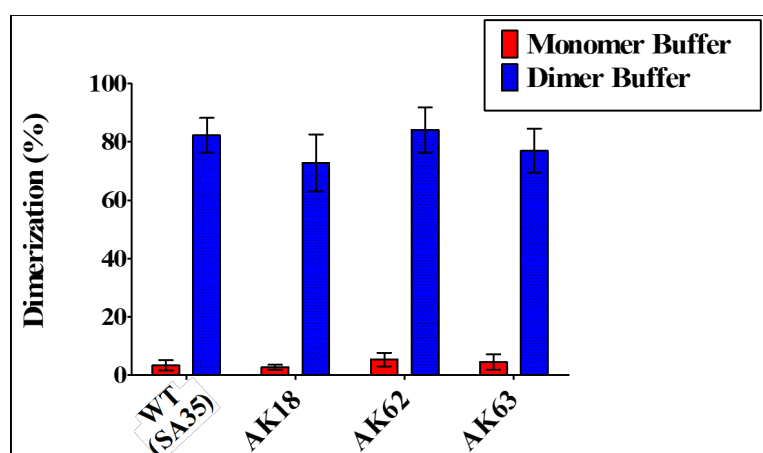


Figure 35: In vitro dimerization ability of the WT (SA35) and the ssPurine mutant (AK18, AK62 & AK63) RNAs

These mutant RNAs show minimal effects, suggesting that the loss of Pr77^{Gag} binding is not due to the impairment in *in vitro* RNA dimerization. WT, wild type.

3.4.4 ssPurines in the Spliced RNAs are Base-Paired

The ssPurines are present upstream of the mSD and thus present in both the unspliced (SA35) as well as spliced mRNAs (*env*; AK29 and *sag*; AK30; Figure 30A). However, *in vitro* binding assays performed on these spliced RNAs showed poor competition efficiency/binding capacity to Pr77^{Gag} compared to unspliced RNA (Figures 31 & 32). These were intriguing results that one could rationalize if the ssPurines in the spliced RNAs were no longer present in an unpaired conformation. Therefore, to test such a possibility, we performed hSHAPE experiments on the gRNA and on both *env* (AK29) and *sag* (AK30) spliced RNAs to elucidate the secondary structure of these RNAs. Results of these experiments revealed that the

ssPurines in both the *env* and *sag* spliced RNAs were present in a base-paired state with neighboring sequences and therefore unavailable for Pr77^{Gag} binding (Figure 36 & Appendix B). This suggests a possible mechanism for the preferential binding of unspliced RNAs over spliced RNAs, governing specific selection of full-length gRNA during the packaging process.

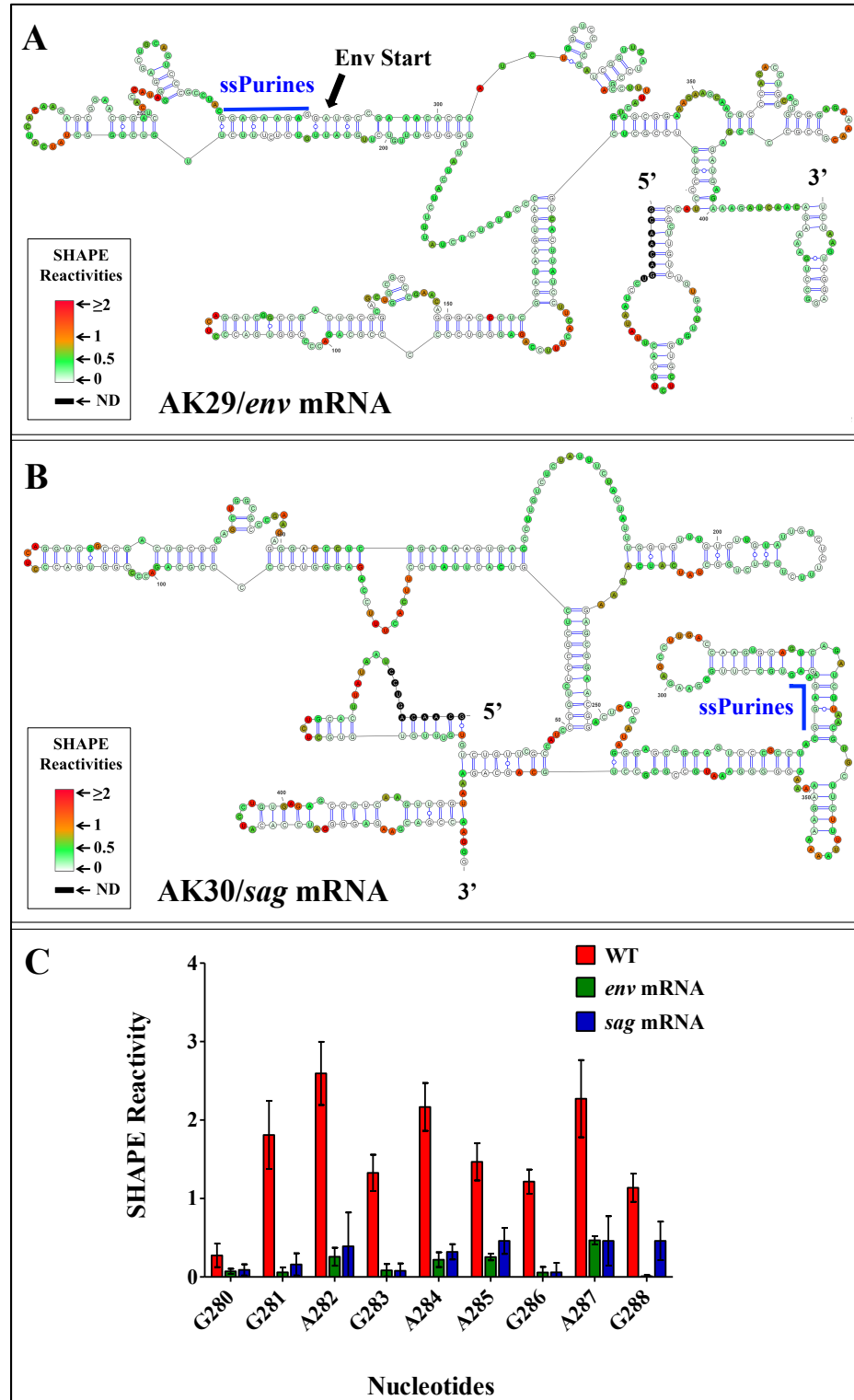


Figure 36: SHAPE-validated RNA secondary structure of first 712 nts of *env* and *sag* RNAs

(A) *env* (AK29), (B) *sag* (AK30) RNAs showing that ssPurines base-pairs with other sequences following RNA splicing and (C) Graphical representation showing SHAPE reactivities of ssPurine region in the WT gRNA, spliced *env* and *sag* RNAs. SHAPE data were used as constraints for all structures. WT, wild type.

3.4.5 Pr77^{Gag} Attenuates the Reactivity of ssPurines towards SHAPE Reagents

To further analyze binding of Pr77^{Gag}, hSHAPE was performed using the benzoyl cyanide reagent (BzCN) and WT ψ -containing RNA, both in the absence as well as presence of Pr77^{Gag}. The nucleotides showing reduced SHAPE reactivity in the presence of the protein correspond to protein binding site(s), as Pr77^{Gag} binding protects these nucleotides from modification by BzCN. To minimize non-specific binding of Pr77^{Gag} to WT RNA, hSHAPE was performed in the presence of excess spliced *env* mRNA (AK29; 4-fold excess molar concentration). Pr77^{Gag} was used at a concentration of 4 μ M (i.e. about 10-fold higher concentration than the K_d value) to ensure complete binding of RNA to Pr77^{Gag}. Under the conditions (protein and competitor RNA concentrations) used for footprinting, the band-shift assay showed formation of only a single complex, indicating that only at the high affinity binding site was targeted in the SHAPE footprinting experiments (Figure 37).

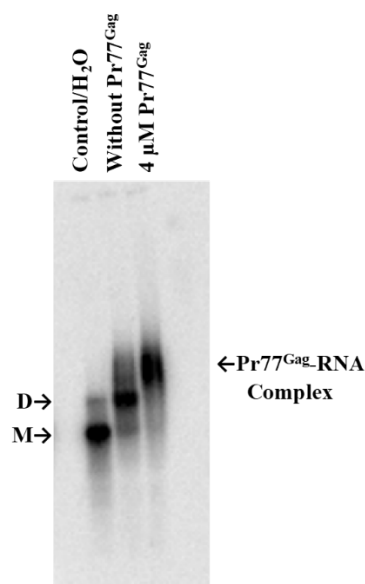


Figure 37: Band-shift assay showing the formation of a single high-affinity Pr77^{Gag}-radiolabeled gRNA complex

The exact conditions used for footprinting experiments were applied.

In the absence of Pr77^{Gag}, the higher order structure of MMTV ψ -RNA (Figure 25A) was obtained that is identical to the one that has been reported earlier (Aktar et al., 2014; Mustafa et al., 2018) except for some minor changes in SL5, SL6 and SL7 located at the 3' end of the packaging signal RNA. SHAPE reactivities obtained in the presence of Pr77^{Gag} were then plotted on the structure obtained using reactivities without Pr77^{Gag} in order to identify the reactivity changes for each nucleotide (Figure 25B). High-throughput SHAPE analysis in the presence of Pr77^{Gag} resulted in 1.5 to 2-fold reduction in reactivity of all ssPurines (p -values ≤ 0.05 ; Figure 39A & Appendix C), indicating that Pr77^{Gag} directly binds to the ssPurines, an observation in agreement with the band-shift competition and filter-binding data.

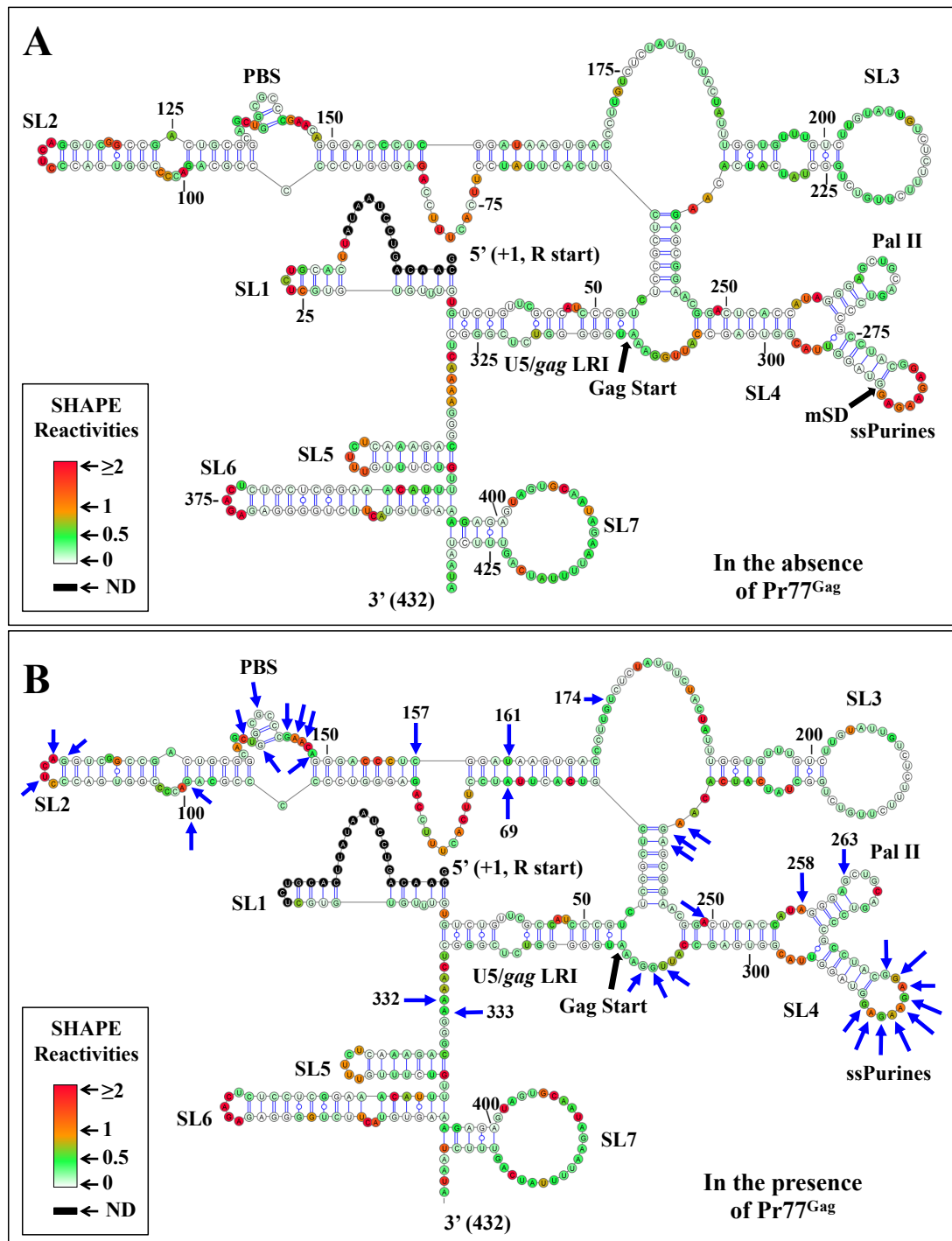


Figure 38: SHAPE-validated secondary structure and footprints of Pr77^{Gag} on MMTV packaging signal RNA

(A) SHAPE validated secondary structure of MMTV packaging signal RNA obtained when hSHAPE was conducted in the absence of MMTV Pr77^{Gag}. (B) The reactivities obtained in presence of Gag were represented on the secondary structure based on the SHAPE reactivities in the absence of Gag. All nucleotides marked by arrows showed a 1.5 to 2-fold reduction in SHAPE reactivities with statistical significance (p -value ≤ 0.05 ; tested by paired, two-tailed Student's t -test).

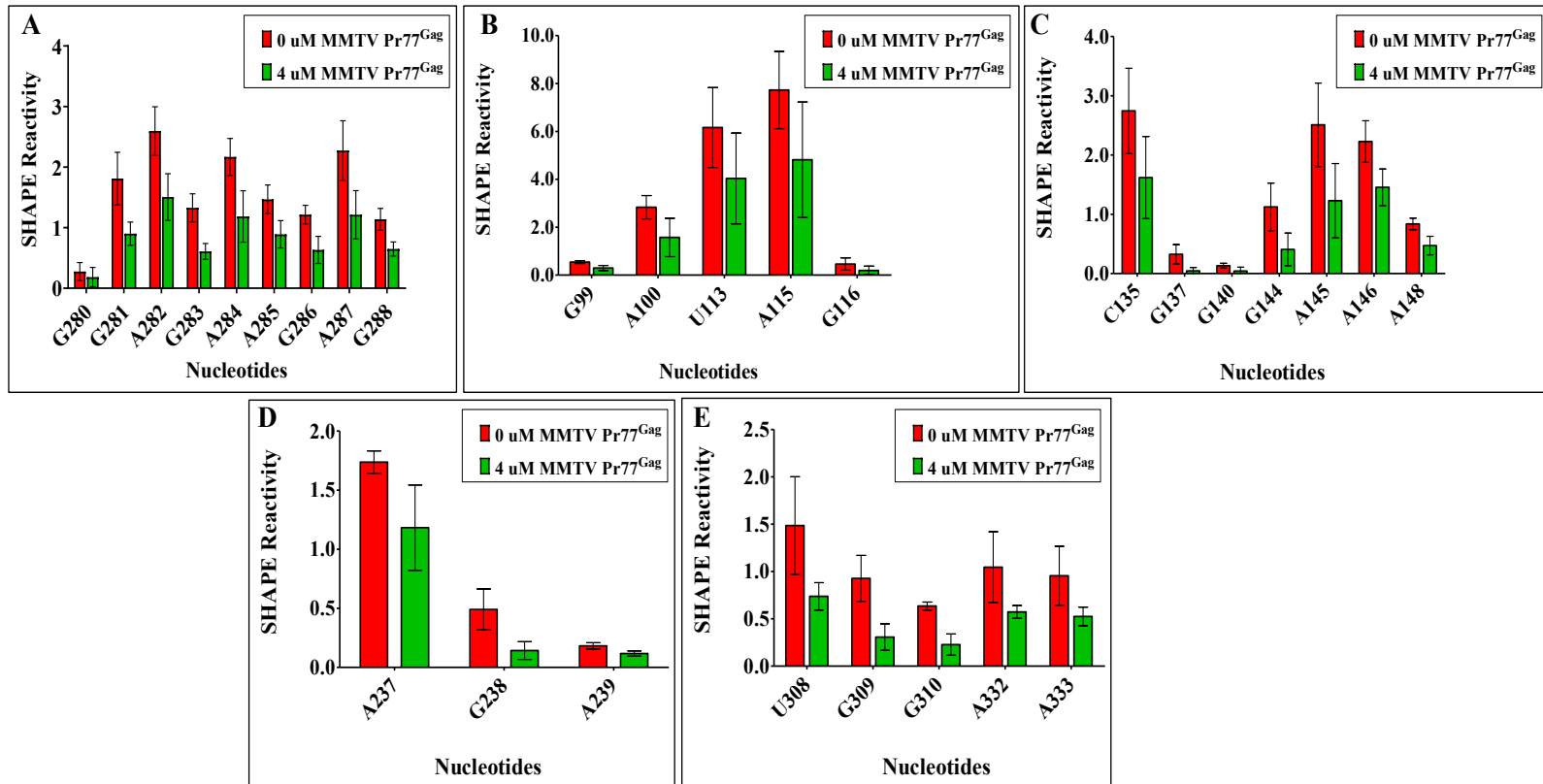


Figure 39: Histograms showing the Pr77^{Gag}-induced attenuation of the SHAPE reactivities of nucleotides in MMTV gRNA

(A) ssPurines, (B) apical part of SL2, (C) PBS region, (D) basal part of SL3 and (E) bottom region downstream of SL4 and unpaired region upstream of SL5 (p value ≤ 0.05 ; tested by paired, two-tailed Student's t -test).

3.4.6 Pr77^{Gag} Attenuates Reactivities of Nucleotides Other than ssPurines

Besides ssPurines, SHAPE analysis of MMTV WT RNA conducted with and without Pr77^{Gag} revealed attenuated reactivity for a few other nucleotides (p -values ≤ 0.05 , Figure 38B). These nucleotides are dispersed as discontinuous short stretches throughout the packaging signal RNA, such as the apical part of SL2 (nts G99, A100, U113, A115 and G116; Figure 38B & Figure 39B), the PBS region (nts C135, G137, G140, G144, A145, A146 and A148; Figure 38B & Figure 39C), the basal part of SL3 (nts A237, G238 and A239; Figure 38B & Figure 39D), the bulge downstream of SL4 (nts U308, G309 and G310; Figure 38B & Figure 39E) and the unpaired region upstream of SL5 (nts A332 and A333; Figure 38B & Figure 39E). In addition, some nucleotides that are distantly located from these regions, A69 and its complementary pair U161, C157 (in SL2), U174 (bulge downstream of SL2), A249, A258 and G263 (in SL4) also showed attenuated SHAPE reactivity (Figure 38B). However, at this point, one cannot ascertain whether these attenuated SHAPE reactivities are due to direct Pr77^{Gag} binding or to Gag induced structural changes in these regions. The role of these nucleotides in gRNA packaging is analyzed below.

3.4.7 Role of ssPurines during in Cell Packaging and Transduction of ψ -Containing RNAs

In order to assess the biological significance of the ssPurines mutations tested in band-shift competition and filter-binding assays (Figure 33A) and which showed attenuated SHAPE reactivity (Figure 38), same mutations were cloned into the subgenomic transfer vector (DA024; Figure 40A) and are listed in figure 40B. These mutations were tested in a biologically relevant single round of replication assay (Rizvi et al., 2009) to investigate their effect on RNA packaging and transduction

efficiencies of the packaged RNAs. The rationale of three-plasmid genetic complementation assay is depicted in Figure 28 and described in the Material and Methods section. Following transfection, the pseudotyped MMTV particles generated were analyzed for the amount of transfer vector RNA packaged into the virions to determine the biological significance of the mutations introduced in ssPurines. In parallel, a portion of the virus particles were used to infect HeLa cells to monitor their ability to transduce the packaged RNA into the target cells by the appearance of hygromycin resistant colonies.

The packaging efficiency of the RNAs was calculated by determining the amount of RNA packaged into the virus particles normalized to the amount of RNA expressed in the cytoplasm. Thus, RNAs from cytoplasmic fractions and pelleted viral particles were extracted and DNase treated to remove any contaminating plasmid DNA. This was confirmed by amplification of the DNase-treated cytoplasmic and viral RNAs using transfer vector-specific primers and showing no amplification, thus confirming that RNA preparations were free of any contaminating plasmids (Figure 40C, Panel I & II respectively). The DNased-RNAs were then converted into cDNAs and the expression of the transfer vector RNAs (WT & mutant RNAs) in the cytoplasm and the packaged RNA in the virus particles were analyzed by RT-PCR (Figure 40C, Panel III & IV). The absence of unspliced β -actin mRNA and presence of 18S rRNA (Figure 40C, Panel V) in the cytoplasmic fractions by RT-PCR ensured that the nuclear membrane integrity was maintained during nucleocytoplasmic fractionation. This was important to demonstrate since the relative packaging efficiency (RPE) is calculated in relation to the stably expressed RNAs that are efficiently exported from the nucleus

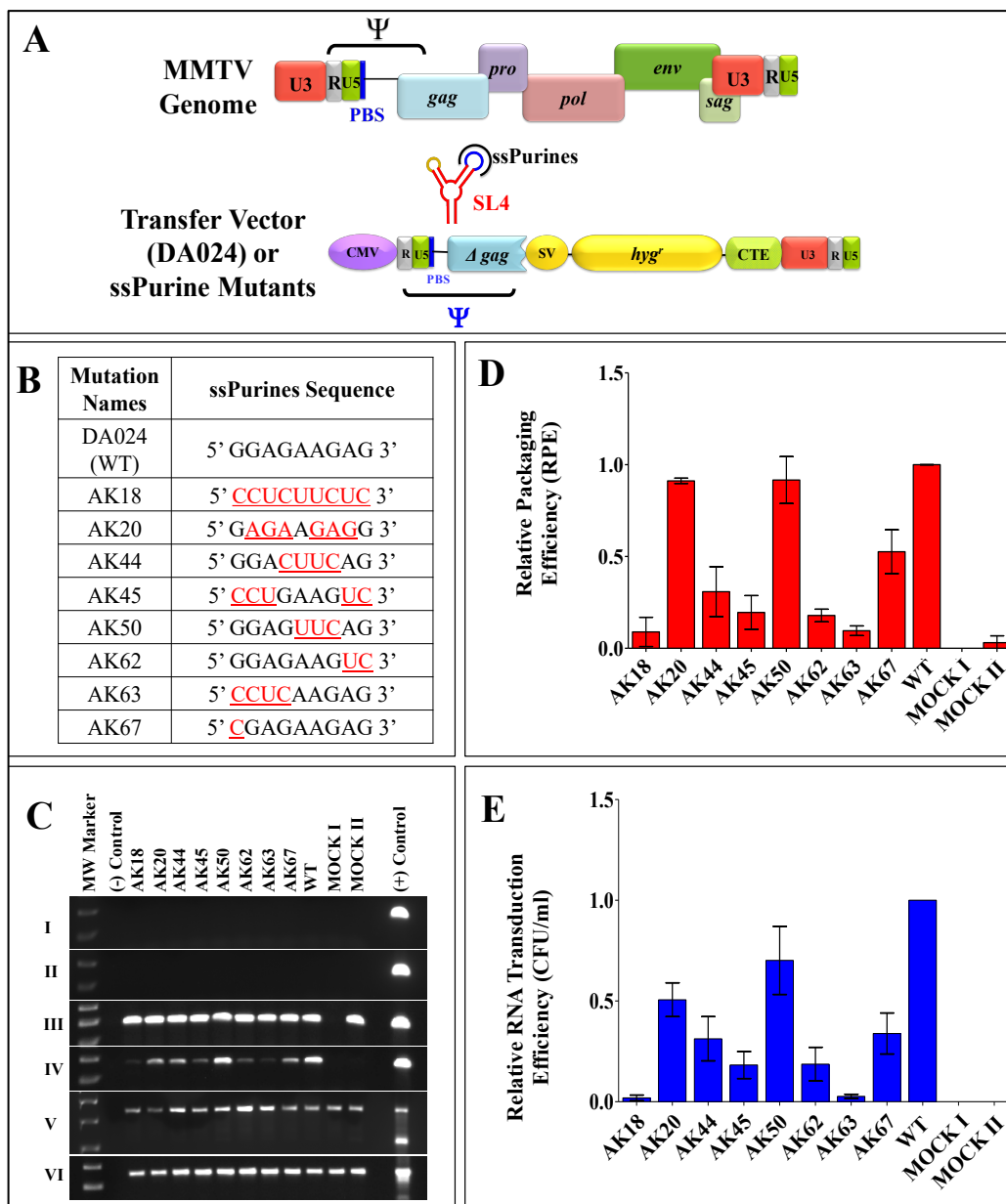


Figure 40: Role of ssPurines in MMTV gRNA packaging and transduction efficiencies

(A) Schematic representation of the MMTV genome and MMTV-based WT transfer vector, DA024. (B) Nature of the mutations introduced in the ssPurines and cloned into the background of subgenomic transfer vector DA024. The red and underlined nucleotides represent substitutions introduced. (C) PCR amplifications of the DNase-treated cytoplasmic (panel I) and viral (panel II) RNAs using virus-specific primers (169 bp). Panels III and IV show PCR amplifications of the cytoplasmic and viral cDNAs using virus-specific primers (169 bp). Multiplex amplifications conducted on cDNAs in the presence of primers/competimer for 18S rRNA (324 bp) and unspliced β -actin mRNA (200 bp; panel V). Panel VI shows PCR amplification of cytoplasmic cDNAs using spliced β -actin mRNA primers (249 bp). (D) Relative packaging efficiencies (RPE) of mutant transfer vector RNAs. (E) Relative RNA transduction efficiencies of the packaged mutant RNA represented as hygromycin resistant (hyg^r) colony forming units per ml (CFU/ml). Mock I contain only the packaging construct (without transfer vector) and Mock II has only transfer vector and no packaging construct.

to the cytoplasm (Aktar et al., 2014; Mustafa et al., 2012, 2018). After having taken into account all the necessary controls described above, the amount of transfer vector RNAs (WT or mutant RNAs) expressed in the cytoplasm and packaged RNA in the pelleted virus particles were quantitated using RT-qPCR (Figure 40D; Aktar et al., 2014; Mustafa et al., 2012, 2018).

A complete substitution of ssPurines with complementary pyrimidines (AK18) resulted in a drastic reduction in the packaging efficiency ($RPE=0.089 \pm 0.079$; Figure 40D) of AK18 compared to the WT, DA024. On the other hand, packaging efficiency of the mutation AK20 (containing a reverse sequence of ssPurines, 5' GGAGAAGAG 3' to 5' GAGAAGAGG 3') was similar to wild type levels (Figure 40D). These results demonstrated that ssPurines in either orientation was necessary for optimal packaging of MMTV gRNA. The mutation in which the GGA at the 5' end and AG at the 3' end of ssPurines were substituted with CCU and UC, respectively (5' CCUGAAGUC 3'; AK45) showed nearly 80% reduction in packageability of the mutant RNA ($RPE = 0.195 \pm 0.091$; Figure 40D) which was in close agreement with Pr77^{Gag} binding ability of this mutation (AK45; Figure 33C & 34B). This further suggested that either both or one of the flanking sides of ssPurines plays an important role in binding to Pr77^{Gag}. The packaging efficiency of mutation in which the central GAAG of ssPurines was substituted with CUUC (5' GGACUUCAG3'; AK44) was found to be slightly increased when compared to AK45 containing flanking substitutions (5' CCUGAAGUC 3'; $RPE = 0.308 \pm 0.135$; Figure 40D). Interestingly, the mutation in which the internal AAG was substituted with UUC (5' GGAGUUCAG 3'; AK50) showed packaging efficiency comparable to WT ($RPE = 1.00 \pm 0.4$; Figure 40D).

Next, to investigate the importance of the nucleotides present at the flanks of ssPurines, individual mutations specific to the GGAG at the 5' end (5' CCUCAAGAG 3'; AK63) and AG at 3' end of ssPurines (5' GGAGAAGUC 3'; AK62; Figure 40B) were tested for the direct effect of these regions on the RNA packaging process. The RPE of AK63 and AK62 revealed 80 to 90 percent reduced packaging efficiency (AK63, RPE = 0.095 ± 0.026 and AK62, RPE = 0.178 ± 0.034 ; Figure 40D). These results are consistent with Pr77^{Gag} binding assays, which showed reduced ability of these mutant RNAs to bind Gag (AK63 & AK62; Figure 33D & 34C). Finally, based on the SHAPE footprinting data, G280 (the first G in ssPurines) showed minimal SHAPE reactivity both in the absence as well as presence of Pr77^{Gag} (0.27 versus 0.18, respectively, $p \leq 0.05$; Figure 38 & Figure 39A). Analysis of the RPE of the substitution mutation (5' CGAGAAGAG 3'; AK67; Figure 40B) revealed about 50% reduction in the packaging efficiency when compared to the WT (DA024).

Finally, the relative RNA transduction efficiency for each of the mutation introduced was calculated by counting the number of hygromycin resistant colonies (CFU/ml) that appeared when viral particles containing these mutants RNAs were used to transduce target cells. For all the introduced mutations, relative transduction of the packaged RNA corroborated well with the relative packaging efficiency except for mutation AK20 (containing a reverse sequence of ssPurines, 5' GGAGAAGAG 3' to 5' GAGAAGAGG 3'). Contrary to the expectations, RNA transduction efficiency of AK20 RNA was observed to be only 50% (compared to WT; DA024) despite the fact that the RNA containing this mutation was packaged almost to the WT (DA024) levels (compare Figure 40D with 40E). This suggests that while the

AK20 mutation did not affect RNA packaging, it may have had some inadvertent effect on subsequent steps of RNA propagation, such as reverse transcription and/or integration. Overall, results from these single round replication assays indicate that ssPurines are crucial for RNA packaging and the 5' GGAG and the 3' AG nucleotides are required for the optimal packaging of MMTV gRNA.

3.4.8 Nucleotides in the PBS Region Play a Crucial Role in RNA Packaging

After having confirmed that the ssPurines are required for MMTV gRNA packaging, it was important to determine the role of other nucleotides that have shown attenuation of SHAPE reactivity in the footprinting assays. It was particularly important to ascertain whether the reduced SHAPE reactivity was due to direct Pr77^{Gag} binding or to Pr77^{Gag}-induced RNA conformational changes. Towards this end, those nucleotides were grouped into five regions and groups of substitution mutations were introduced to test their RNA packaging and transduction abilities using the single round of replication assays. The details and/or nature of these substitution mutations (AK68, AK69, AK70, AK73 and AK74) is schematically shown in Figure 41A. After performing the same necessary controls as described for the other mutations (Figure 42), the packaging efficiency of these mutations was determined. Mutations in the apical part of SL2 (AK68), bulge downstream of SL4 (AK69) and the unpaired region upstream of SL5 (AK70) did not detrimentally affect RNA packaging efficiency; rather, these mutations showed increased RNA packaging (Figure 41B, AK68; RPE = 1.21 ± 0.47 , AK69; RPE = 1.50 ± 0.62 , AK70; RPE = 2.02 ± 0.67).

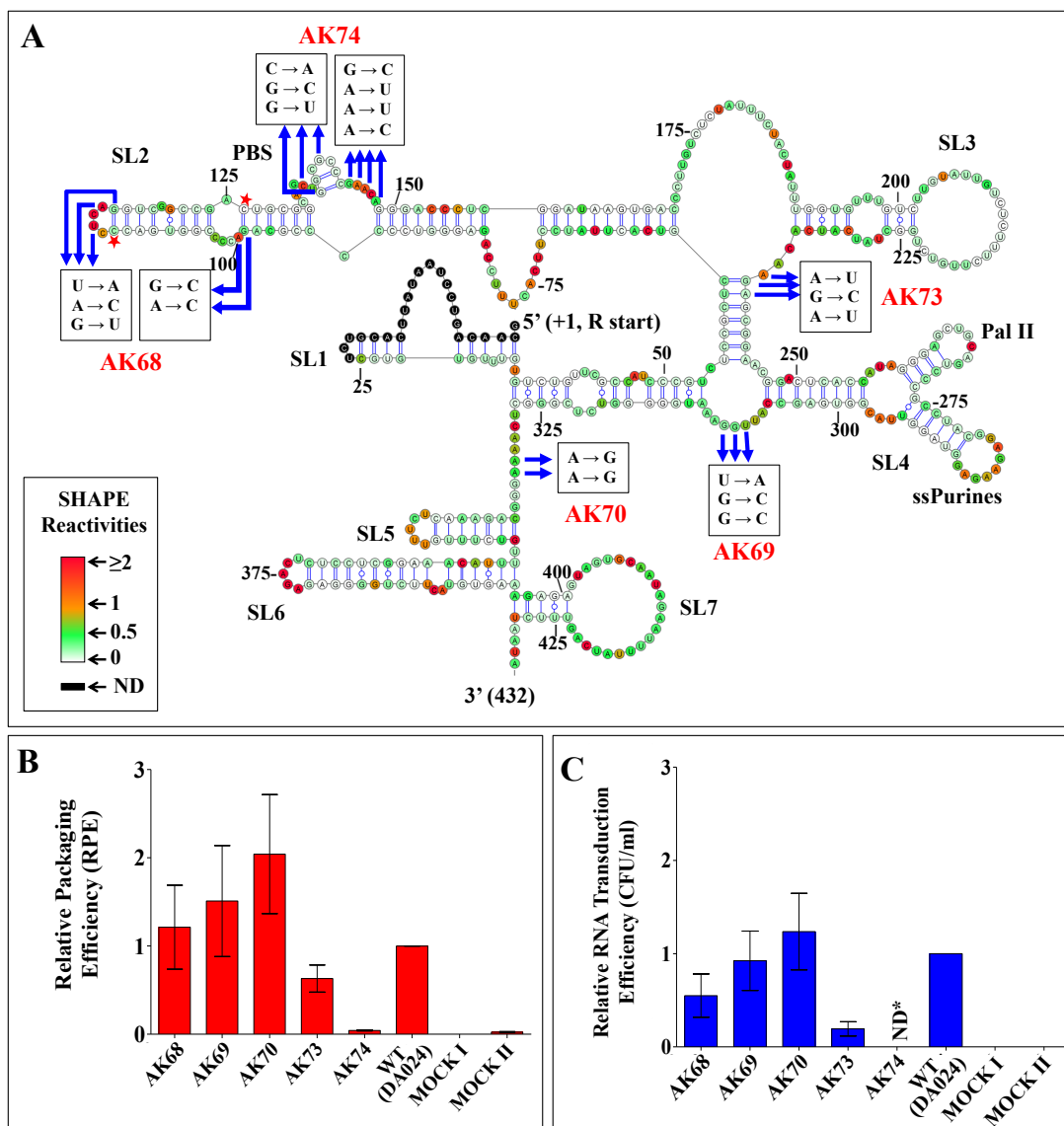


Figure 41: Role of nucleotides outside ssPurines that showed attenuation of SHAPE reactivities in the presence of MMTV Pr77^{Gag} on MMTV gRNA packaging and transduction efficiencies

(A) Nature of mutations introduced in the SHAPE-validated structure of MMTV gRNA showing attenuation of SHAPE reactivities in the presence of MMTV Gag. The nucleotides showing attenuation other than ssPurines are marked by arrows. Note that this figure and figure 38B are the same and has been reproduced to depict the SHAPE reactivities and the locations of the nucleotides identified for introducing mutations. Mutations were introduced in these nucleotides (boxed) to test their role in RNA packaging and transduction. The red star marked nucleotides are the compensatory mutated nucleotides (C111 to A and C126 to G) in AK68 in order to maintain the secondary structure. (B) Relative packaging efficiency (RPE) of mutant transfer vector RNAs. (C) Relative RNA transduction efficiencies (tested using single round of replication assays) of the packaged mutant RNAs represented as hygromycin resistant (Hyg^r) colony forming unit per ml (CFU/ml). Mock I contain only the packaging construct (without transfer vector) and Mock II has only transfer vector and no packaging construct. *ND; not determined.

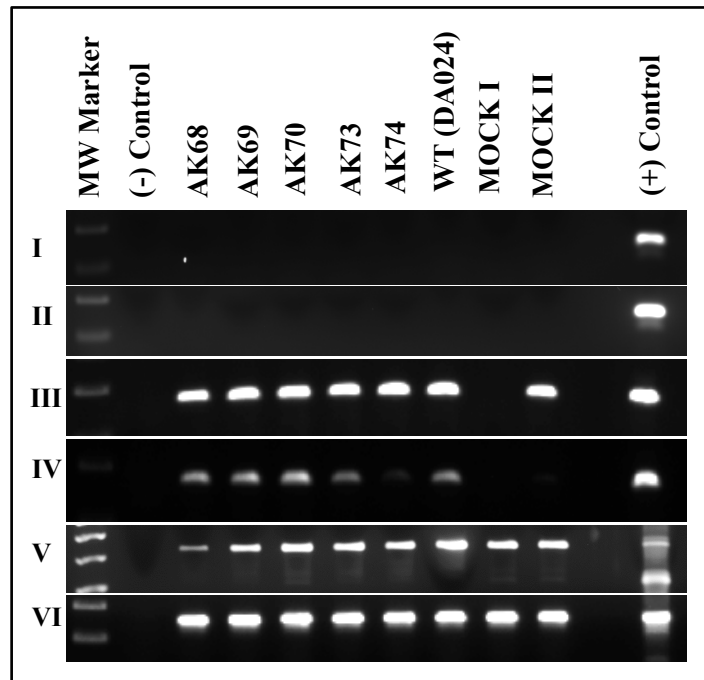


Figure 42: Control PCR amplifications used for RNAs with mutations that have been introduced in regions other than ssPurines

PCR amplifications of the DNase-treated cytoplasmic (panel I) and viral (panel II) RNAs using MMTV-specific primers (169 bp). Panels III and IV show PCR amplifications of the cytoplasmic and viral cDNAs, respectively, using MMTV-specific primers (169 bp). Multiplex amplifications were conducted in the presence of primers/competimer for 18S rRNA (324 bp) and unspliced β -actin mRNA (200 bp). Lack of any amplifications of unspliced β -actin mRNA in panel V confirms that there was no contamination of cytoplasmic RNA fractions with that of nuclear fraction, while amplification of 18S rRNA confirms the quality of the cDNA. Panel VI shows amplification of cytoplasmic cDNA using primers that amplify spliced β -actin mRNA (249 bp). Mock I contain only the packaging construct (without transfer vector) and Mock II has only transfer vector and no packaging construct.

Transduction efficiencies for these mutations revealed a similar pattern except for mutation AK68 (Figure 41C). The mutation AK68 showed almost 50% reduction in its ability to transduce the packaged RNA (Relative CFU/ml = 0.54 ± 0.23 ; compare Figure 41B with 41C). On the other hand, substitution of the three nucleotides at the bottom part of SL3 in mutation AK73 showed an intermediate effect, resulting in 40% reduction in RNA packaging efficiency compared to WT (AK73, RPE = 0.63 ± 0.153) and a correspondingly reduced transduction ability

(CFU/ml = 0.19 ± 0.076 ; compare Figure 41B with 41C). On the contrary, substitution of 7 nucleotides in the PBS region within SL2 (AK74) resulted in a drastic reduction in RNA packaging (96% reduction, RPE = 0.043 ± 0.005 ; Figure 41B), revealing the importance of this region to MMTV gRNA packaging. Transduction of the packaged RNA for this mutation (AK74) was not investigated since mutations in the PBS limits the ability of the packaged RNA to reverse transcribe which is a must for the transduction read out assay (Figure 28). Together, these results demonstrate that the nucleotides in the first three regions tested (apical part of SL2, bulge downstream of SL4 and the unpaired region upstream of SL5) are not important for RNA packaging, nucleotides in the bulge downstream of SL3 have some moderate effect, while nucleotides in the PBS region have the most drastic effect on RNA packaging amongst the regions tested.

The mutational analysis showed that in addition to ssPurines, mutations in the bottom part of SL3 and PBS region (AK73 and AK74) also showed a 40% and 96% reduction in their packaging abilities respectively when compared to WT (DA024). To determine if these nucleotides played a true role in Pr77^{Gag} binding, AK73 and AK74 mutations were tested in band-shift competition assays. Both mutations (AK73 and AK74) showed significantly reduced affinity to Pr77^{Gag} compared to the WT (SA35) in band-shift competition assays (Figure 43). In order to check whether this loss of affinity was due to loss of higher order structure, hSHAPE was performed on these mutants. In the case of mutation AK73, the first two SHAPE-validated structures showed different secondary structures when compared to the WT; however, structure 3 was very similar to WT (SA35), maintaining the conserved structural motifs (part of U5/*gag* LRI, SL2, SL3 and SL4) between nt 48 - 318

(Figure 44). However, the AK74 mutation maintained an overall WT (SA35) structure with a prominent change at the PBS region, resulting in unpairing of the short stem-loop in the PBS structure of WT (SA35) RNA (Figure 45). These data suggest that the 40% packaging defect observed in the mutation introduced in the bottom part of SL3 (AK73) is probably due to a loss of Gag binding to the RNA, attributed to a partial loss of structure of the packaging signal RNA. Alternatively, it is also possible that the mutated nucleotides in this region play some role in Gag binding. On the other hand, substitution mutations in the PBS region (AK74) did not perturb the RNA structure, suggesting that the PBS has a direct role in MMTV RNA packaging, by affecting Gag binding.

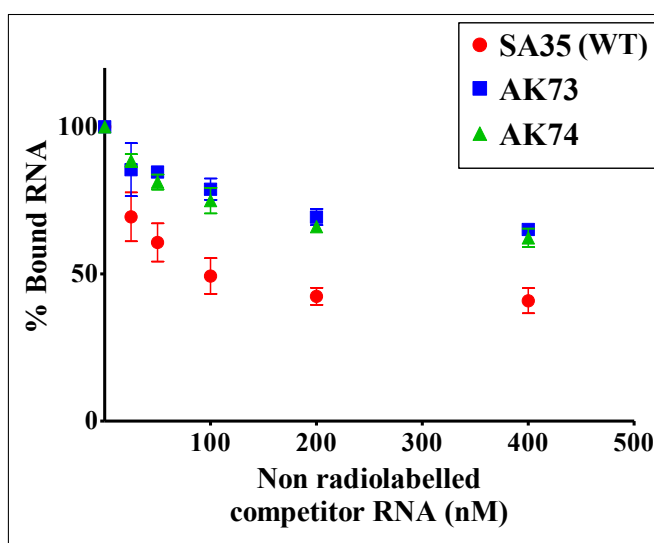


Figure 43: Band-shift competition assays for the clones containing mutations in PBS region and basal part of SL3

Band-shift competition assays for the clones containing mutations in nucleotides showing attenuation in SHAPE reactivities in PBS region (AK74) and basal part of SL3 (AK73; see Figure 41A for details of mutations introduced). Quantification of the gels showing the binding specificity of AK74 and AK73 RNAs to full-length MMTV Pr77^{Gag} analyzed by band-shift competition assay.

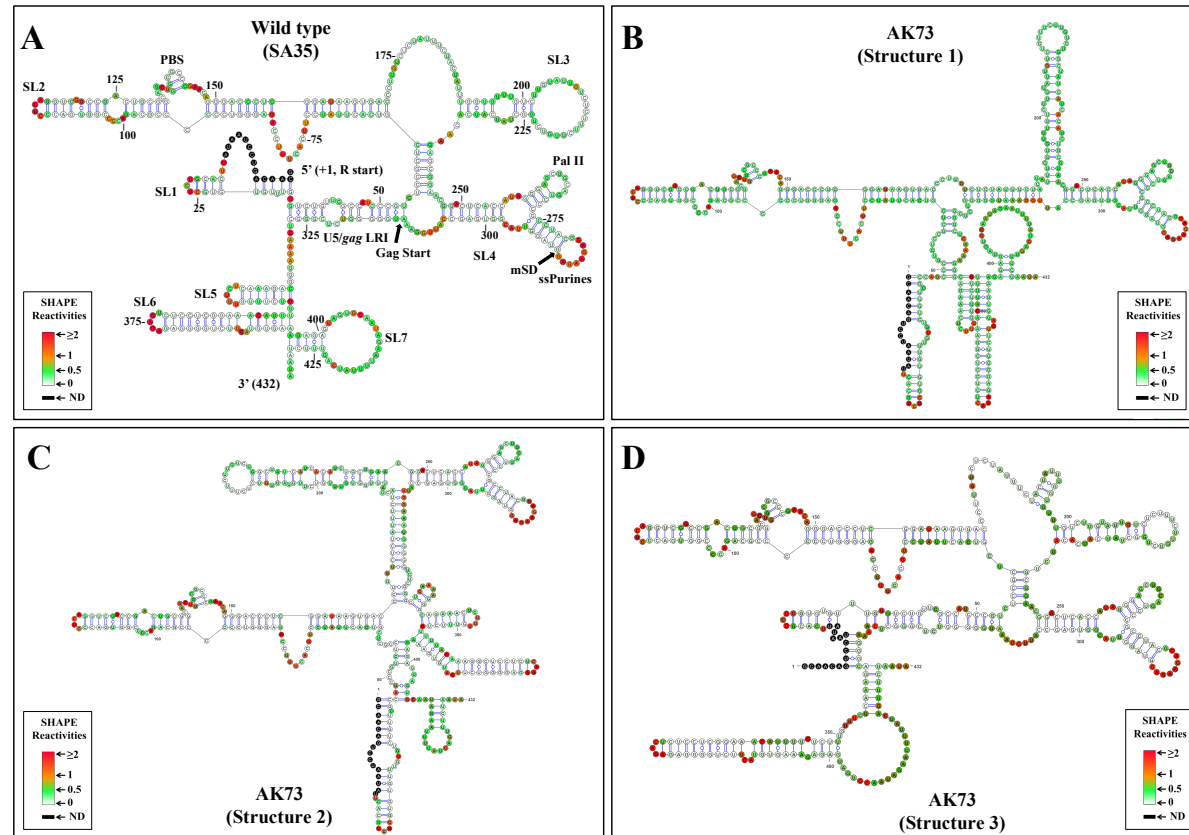


Figure 44: SHAPE-validated secondary structure of wild type (SA35) and mutant containing substitutions in the bottom part of SL3 RNAs

(A) Wild type (SA35) (B) AK73 (A237, G238 and A239 to U237, C238 and U239) structure 1 obtained from the software RNAstructure 6.1, Energy (E) = -274.8 kcal/mol. (C) Structure 2, E = -279.6 kcal/mol. (D) Structure 3, E = -270.6 kcal/mol. SHAPE data were used as constraints for all structures.

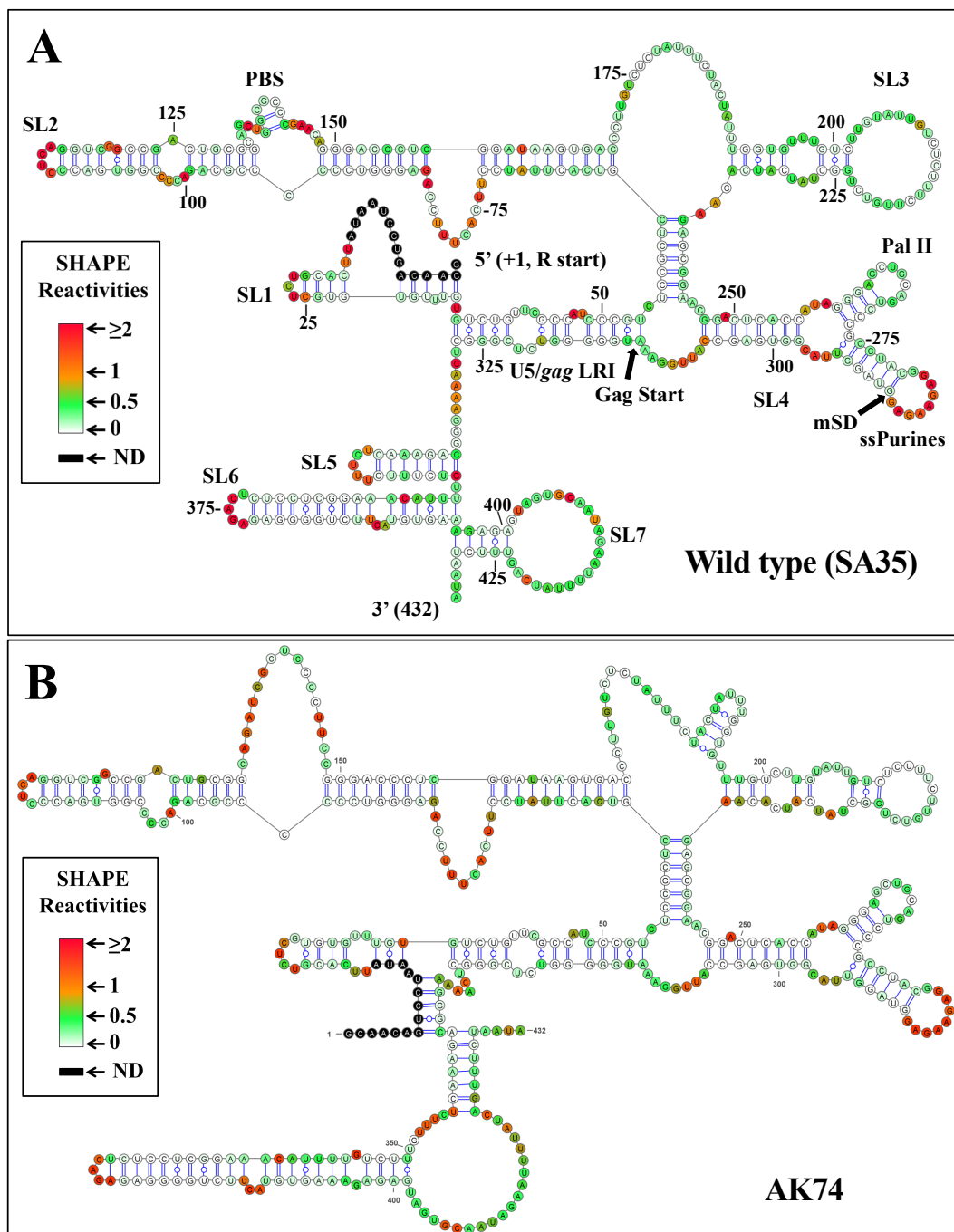


Figure 45: SHAPE-validated RNA secondary structure of WT and AK74 RNAs

(A) Wild type (SA35) and (B) AK74 containing substitution mutations in the primer binding site (PBS) region; structure number 2 from the RNAstructure 6.1, Energy (E) = -280.8 kcal/mol. SHAPE data were used as constraints for all structures.

To further confirm the direct contribution of nucleotides in the PBS region towards Pr77^{Gag} binding, SHAPE footprinting experiments were performed using the mutation AK74 RNA (in which all the 7 nucleotides that showed attenuation of SHAPE reactivities in the presence of Gag were substituted; Figure 41A) employing the same conditions used for the footprinting of WT (SA35) RNA. Consistent with these expectations, the reactivities of these 7 substituted nucleotides were not significantly reduced in the presence of Pr77^{Gag}, indicating that Gag is not able to bind to these nucleotides (Figure 46), which agrees with the reduced competition ability of the same mutation (AK74) compared to WT (Figure 43).

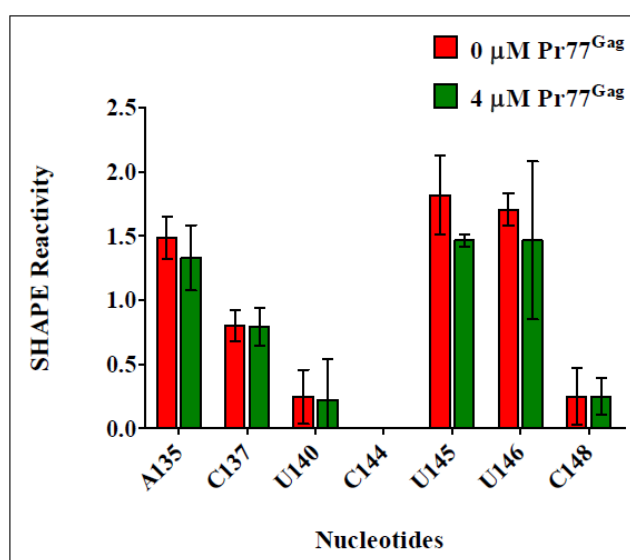


Figure 46: Histograms showing SHAPE reactivities of the mutation (AK74) introduced into the PBS region in the absence and presence of Pr77^{Gag}

3.4.9 The role of PBS in MMTV gRNA Packaging is at the Primary Sequence Level

The PBS harbours overlapping palindromic sequences (PBS pal; 3' CAGCTG*GCGCC* 5'; the underlined nucleotides represent first palindromic sequence, while the second palindromic sequence is italicized) that might affect

RNA dimerization and, indirectly, RNA packaging (Aktar et al., 2014). Thus, at this point, the loss of RNA packaging of the mutation AK74, in which the 7 nucleotides that showed attenuation of SHAPE reactivities in the presence of Gag were substituted, that reported in Figure 41 could be due to a direct effect on Pr77^{Gag} binding, or an indirect effect on RNA dimerization, since the two PBS pals lost their palindromic nature by introducing such a mutation. In order to clarify the precise role of PBS in MMTV gRNA packaging, a series of mutations were introduced in PBS that either disrupted its palindromic nature while maintaining the nucleotides that showed Pr77^{Gag} footprints or *vice versa*, with or without simultaneous mutations in the DIS (Figure 47A), followed by testing their RNA dimerization and encapsidation capabilities.

Briefly, starting with a DIS pal mutation (SA44), a double mutation was created in which the palindromic nature of DIS as well as that of PBS pal was lost (AK84; Figure 47A). In another mutation AK80, the 7 nucleotides in the PBS region that showed attenuation of SHAPE reactivities in the presence of Gag were mutated while maintaining the palindromic nature of the PBS by substituting the 11-nucleotide native overlapping palindromic sequence (CAGCTGGCGCC) with a heterologous palindromic sequence (GACGTCGGCCG). On the other hand, mutation AK81 had disrupted the palindromic nature of the PBS while maintaining the nucleotides that showed attenuation of SHAPE reactivities in the presence of Gag. In the mutation AK82, the overlapping palindromes were maintained but the four downstream nucleotides that showed attenuation of SHAPE reactivities in the presence of Gag were substituted.

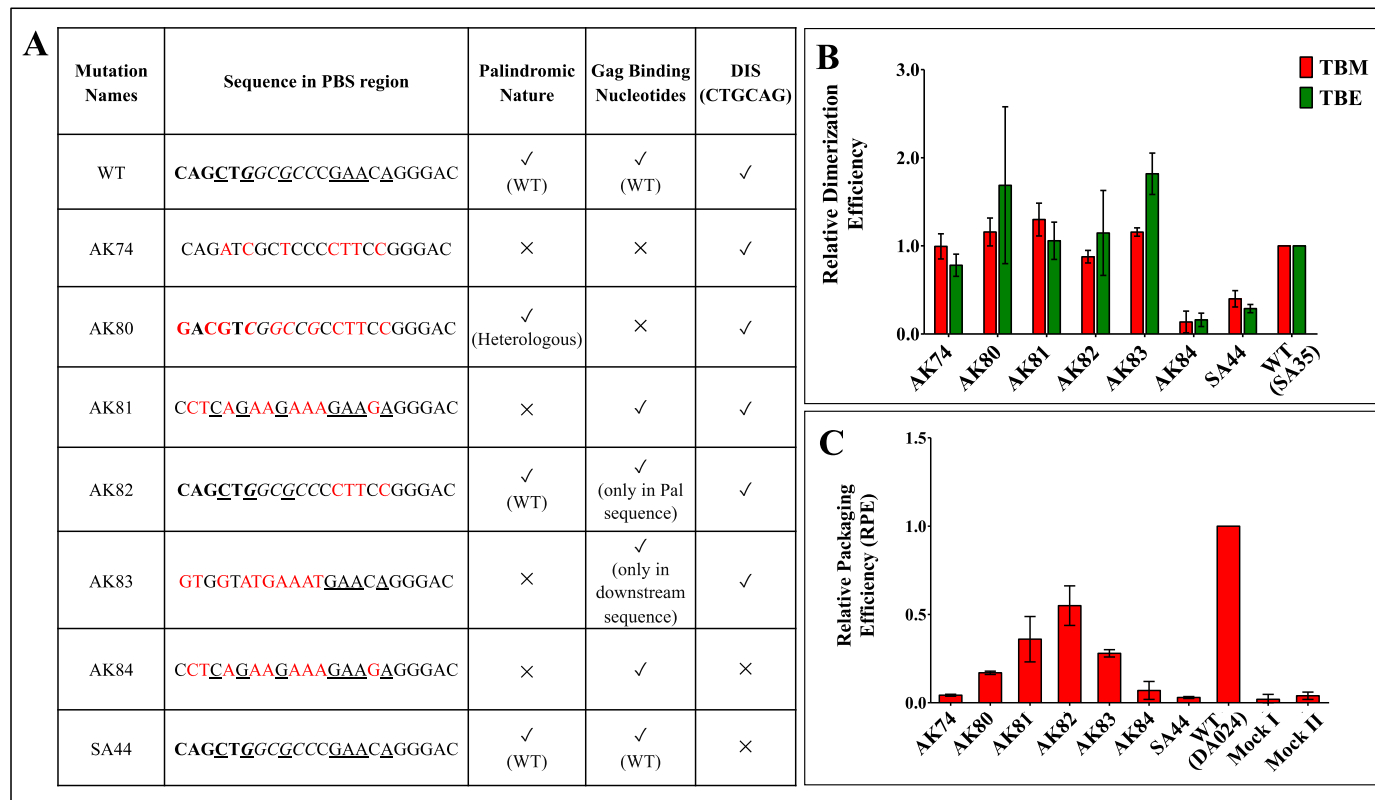


Figure 47: Role of PBS in dimerization and packaging of MMTV gRNA

(A) Table describing the nature of substitution mutations introduced into the PBS region. The bold nucleotides correspond to the first palindromic sequence within the overlapping pals while the second palindromic sequence is italicized. The underlined nucleotides represent the nucleotides which were protected from SHAPE modification by Pr77^{Gag} and substitutions are shown in red. (B) *In vitro* dimerization ability of PBS mutant RNAs related to that of WT (SA35) in native (TBM) and denaturing (TBE) conditions. (C) Packaging efficiencies of mutant transfer vector RNAs relative to the WT (DA024). Mock I contain only the packaging construct (without transfer vector), Mock II has only transfer vector and no packaging construct. WT, wild type.

Finally, in mutation AK83, the palindromic sequences were substituted with non-palindromic sequences concomitantly with substitution of 3 out of 7 nucleotides at the 5' end of the PBS showing Gag footprints, thus restricting the mutations to only the palindromic regions within PBS.

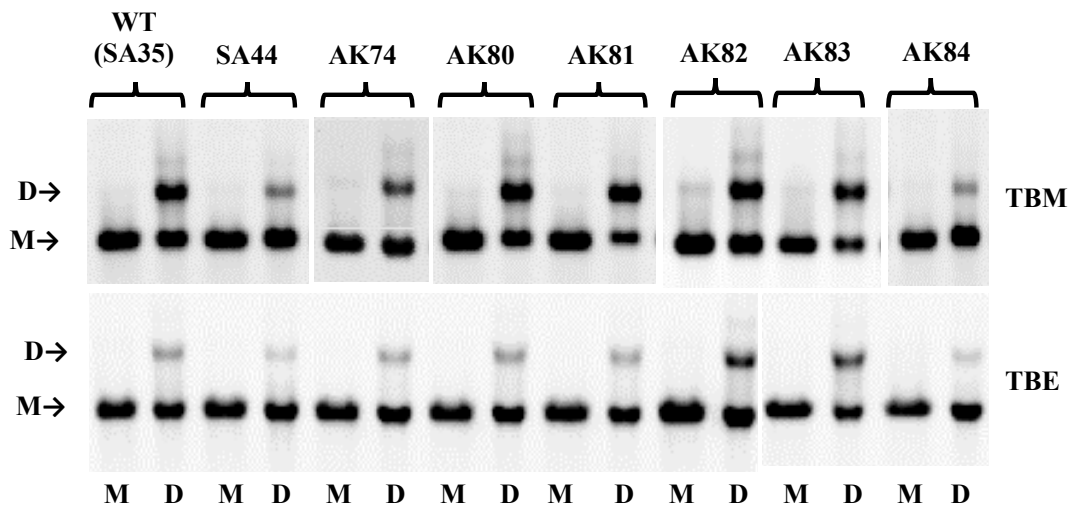


Figure 48: Representative gel picture showing in vitro dimerization of WT gRNA (SA35) and PBS mutant RNAs in native (TBM) and denaturing (TBE) conditions

M & D at the bottom of the gel figure indicates the monomer and dimer buffers in which the dimerization assays were performed. Following dimerization, electrophoresis was performed using both the native (TBM) and denaturing (TBE) gels with respective running buffers. The M and D in the left side of the figure indicates the position of the monomeric and dimeric RNA species, respectively. WT, wild type.

First, the RNA dimerization ability of mutations in which the DIS pal (SA44), the PBS pal (AK74, AK81) or both palindromes (AK84) were mutated (Figure 47A) were compared. While RNA dimerization of SA44 and AK84 was severely compromised, AK74 and AK81 RNAs dimerized with wild type efficiency (Figure 47B & Figure 48), indicating that the PBS pals are not required for efficient RNA dimerization. However, RNA packaging of all the three mutations (AK74, AK81 and AK84) was severely affected (Figure 47C and fractionation and cDNA

controls in Figure 49), suggesting a direct role of PBS sequences in RNA packaging, independent of gRNA dimerization.

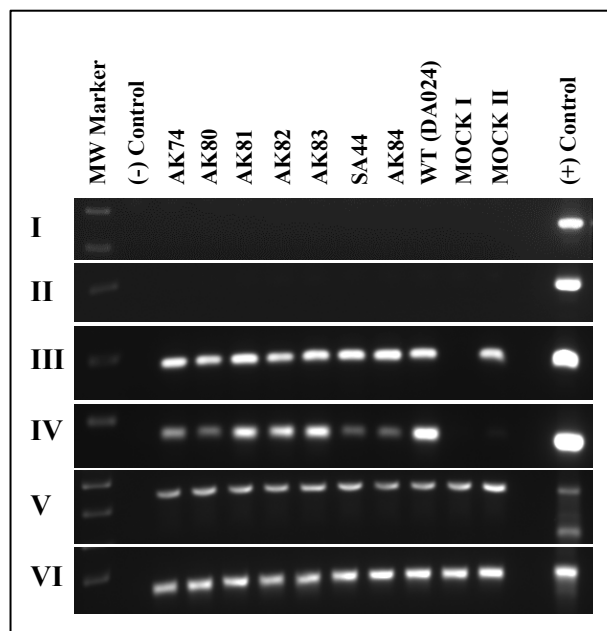


Figure 49: Control PCR amplifications used for RNAs with mutations that have been introduced in PBS region

PCR amplifications of the DNase-treated cytoplasmic (panel I) and viral (panel II) RNAs using MMTV-specific primers (169 bp). Panels III and IV show PCR amplifications of the cytoplasmic and viral cDNAs, respectively, using MMTV-specific primers (169 bp). Multiplex amplifications were conducted in the presence of primers/competimer for 18S rRNA (324 bp) and unspliced β -actin mRNA (200 bp). Lack of any amplifications of unspliced β -actin mRNA in panel V confirms that there was no contamination of cytoplasmic RNA fractions with that of nuclear fraction, while amplification of 18S ribosomal RNA confirms the quality of the cDNA. Panel VI shows amplification of cytoplasmic cDNA using primers that amplify spliced β -actin mRNA (249 bp). Mock I contain only the packaging construct (without transfer vector) and Mock II has only transfer vector and no packaging construct.

Next, the RNA packaging ability of the various mutations introduced in the 7 nucleotides of the PBS region which showed Gag footprints were compared (Figure 38). In the case of AK80 (in which the palindromic nature of PBS was maintained), the RNA packaging efficiency was reduced by 83% (Figure 47C and fractionation and cDNA controls in Figure 49). On the other hand, the mutation AK81 (in which

the palindromic nature of the PBS was lost while maintaining the nucleotides that showed Gag footprints) revealed ~60% reduction in packaging (Figure 47C), less drastic than AK80. Test of mutations AK82 (in which the 5' part of the PBS containing the PBS pals was maintained but the downstream nucleotides showing Gag footprint were substituted) and AK83 (in which the substitutions were restricted only to the PBS pals) revealed a 50% and 86% reduction, respectively, in RNA packaging (Figure 47C). The dimerization ability was not affected in any of these mutations (Figure 47B).

Together, these results indicate that 1) the palindromic sequences in PBS are not required for RNA dimerization (see AK84 and AK81; Figure 47B), 2) the nucleotides in PBS showing Gag footprints are important for RNA packaging (see AK74; Figure 47C), as well as other nucleotides in the PBS that do not show Gag footprints (see AK81; Figure 47C) and 3) the 5' region of PBS apparently plays a greater role in RNA packaging than the 3' region (see AK82 *versus* AK83 in Figure 47C).

3.5 Discussion

The mechanism by which the Gag precursor selects and packages the retroviral genome, a key step in retroviral life cycle, remains largely unclear despite having been studied extensively. Most reported studies on gRNA packaging in the literature have been performed on HIV-1, where purine-rich apical or internal loops in hairpins structures have been proposed to govern gRNA packaging by functioning as Gag binding sites. In this study, to identify the aspect of general rules underlying retroviral gRNA packaging, this aspect is addressed in MMTV, a *Betaretrovirus* that

assembles in the cytosol, before migrating to the plasma membrane and budding. To that aim, expression and purification of the full-length MMTV Pr77^{Gag} was done and compared its binding to WT and mutant gRNA fragments as well as to spliced *env* and *sag* viral mRNAs. Footprinting with SHAPE reagents to identify Pr77^{Gag} binding sites on the WT gRNA was also performed. These results reveal the presence of two specific Gag binding sites of non-redundant nature within the packaging signal RNA consisting of a purine loop and the primer binding site. Despite these sequences being present on both unspliced and spliced RNAs, Gag specifically bound only to unspliced RNA, since it is the only one that could adopt the native bifurcated stem-loop structure containing the looped purines. Thus, results presented in this study have important implications for how MMTV, in particular and retroviruses/retrotransposons, in general, recognize gRNA for specific incorporation into the assembling virions. This study reinforces the hypothesis proposed earlier that specific structural elements in the context of the larger RNA packaging signal act as high affinity binding site(s) for Gag protein recognition (Ding et al., 2020; Gherghe et al., 2010; Rein, 2020). Furthermore, these observations have important ramifications for the development of antiviral therapies that target the virion assembly process in retroviruses.

Previous studies have shown that the major packaging determinants for MMTV gRNA reside in the 5' UTR and 120 nts of *gag* (Mustafa et al., 2012) and fold into a complex secondary structure with several SLs. Among these, the bifurcated SL4 contains the DIS in one of the apical loops and a stretch of ssPurines in the second apical loop; the main 5' splice site that is used to generate the viral spliced RNAs, is located immediately after this ssPurine stretch (Aktar et al., 2014).

In this study, this *in vitro* binding and in cell packaging/transduction assays showed that ssPurines are indispensable for MMTV gRNA packaging and virus replication by directly binding Pr77^{Gag} (Figures 33, 34 & 40). Within the ssPurines, the GGAG at the 5' end and the AG at the 3' end, respectively, are crucial for packaging (Figure 40). Interestingly, a sequence similar to the MMTV 5' GGAG was found to be involved in the Pr55^{Gag} binding in HIV-1 (in the form of an asymmetrical internal loop 5' G/AGG 3'; Abd El-Wahab et al., 2014) and HTLV-1 NC and MA binding (5'GAG 3'; Wu et al., 2018). Similarly, in HIV-2, a 5' GGRG 3' motif located upstream of DIS was found to be important for RNA packaging and has been suggested as Gag binding site (Baig et al., 2009). Altogether, results presented here propose a possible mechanism towards selecting gRNA and suggest that the rather large packaging region on gRNA mapped earlier must maintain specific RNA structural motifs so that the ssPurines can be presented to Gag in a single-stranded manner for recognition during packaging.

Interestingly, this study reveals that the bifurcated SL4 of MMTV gRNA juxtaposes the DIS to the ssPurines that constitutes the primary Pr77^{Gag} binding site. This situation is reminiscent of HIV-1, where the apical and internal loops of a long hairpin (SL1) constitute the DIS (Paillart et al., 1996; Skripkin et al., 1994) and primary Pr55^{Gag} binding site (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Smyth et al., 2015), respectively. This is all the more the case that in MMTV gRNA the two hairpins of the bifurcated SL4 may stack on top of each other (Lescoute & Westhof, 2006). If this is the case, the MMTV DIS and Gag binding sites are collinear, as it has been reported in the case in HIV-1 (Ennifar et al., 2001). The spatial proximity of the DIS and primary Gag binding site may be a general

phenomenon and it would explain the observation that dimerization and packaging of retroviral gRNA are highly interconnected events (Berkhout & van Wamel, 1996; Laughrea et al., 1997; Paillart et al., 1996; Sakuragi et al., 2003). A slightly different situation may prevail in MLV, where dimerization of the gRNA seems to be required to expose a nearby sequence recognized by the nucleocapsid domain of Gag (Miyazaki, Garcia et al., 2010; Miyazaki, Irobalieva et al., 2010; Tounekti et al., 1992).

Packaging of gRNA into newly forming virus particles is highly selective, although cellular and spliced RNAs are also incorporated (Aronoff & Linial, 1991; Didierlaurent et al., 2011; Houzet et al., 2007; Mustafa et al., 2012; Rulli et al., 2007). The binding assays with RNAs corresponding to the first 712 nts of unspliced gRNA and spliced *env* and *sag* mRNAs suggest that selective packaging of unspliced gRNA begins at the initial stages of viral assembly, which involves direct, specific binding of Pr77^{Gag} to unspliced gRNA and not to spliced RNAs. Indeed, despite the presence of the ssPurines in both spliced mRNAs (*env* and *sag*) used in this study, Pr77^{Gag} is unable to bind to those RNAs with high affinity (Figures 31 & 32). This is due to the fact that the major 5' splice donor site is located immediately downstream of the ssPurines and thus, the ssPurine hairpin structure is lost during splicing, an observation that was confirmed by hSHAPE analysis (Figure 36). This analysis revealed that the ssPurines in spliced RNAs are base-paired and unavailable for Gag binding, explaining the selective binding of Pr77^{Gag} to the unspliced gRNA over spliced RNAs. This is in line with previous observations that revealed that the bifurcated SL4 structure is crucial to maintain gRNA packaging (Mustafa et al., 2018). Of note, a previous study reported that the packaging determinants for

MMTV gRNA reside in the entire 5' UTR and extends up to 120 nts of *gag* (Mustafa et al., 2012). Interestingly, the footprinting experiments revealed no Pr77^{Gag} binding sites in the *gag* gene (Figure 38), but the region downstream of the mSD plays a critical function in maintaining the bifurcated structure of SL4. A similar observation has been made in the case of avian leukosis/sarcoma virus (ALSV) and suggest that the ALSV *env* mRNA acquires a packaging incompetent structure (Banks et al., 1999). Selective binding of the Gag precursor to gRNA but not to spliced viral RNAs has also been observed in the case of HIV-1, despite the fact that the major packaging sequence (5'G/AGG 3'), forming the internal loop of SL1, is present in both unspliced and spliced RNAs (Abd El-Wahab et al., 2014). In this case the structure of SL1 is identical in gRNA and spliced viral RNAs, but sequences upstream of SL1 prevent Gag binding if a short region downstream of SL3, which is part of gRNA but not spliced RNAs, is not present (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Smyth et al., 2015). Regions upstream of SL1 and downstream of SL3 maintain a three-dimensional structure that exposes the lower part SL1 for Pr55^{Gag} binding (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Smyth et al., 2015). It has also been shown that in HIV-1, the long-range interactions (LRI) between U5 and the *gag* initiation codon stabilizes a dimerization-competent RNA structure, which in turn, may lead to packaging (Abbink et al., 2005; Abbink & Berkhout, 2003; Brigham et al., 2019; Lu et al., 2011). Further studies will be required to elucidate the precise role of U5-*gag* LRIs (Figure 38), if any, in MMTV gRNA packaging.

The footprinting assays in the current study revealed that Pr77^{Gag} induces attenuation of SHAPE reactivity not only in the ssPurines loop, but also in

nucleotides of other regions (Figure 38 & Figure 39). To determine whether reduced reactivity was due to direct Pr77^{Gag} binding or due to Pr77^{Gag}-induced structural changes, region-wise mutations in these nucleotides were introduced and assayed their effects on in cell RNA packaging and transduction (Figure 41). Mutations in the apical part of SL2, the bulge downstream of SL4 and the unpaired region upstream of SL5 did not have any significant effect on the packaging efficiency. The mutation of 3 nucleotides in the basal part of SL3 (Figure 41), showed a 40% reduction in the packaging efficiency; however, the SHAPE-validated structures suggest that this RNA may exist as a mixture of dynamical structures and a portion of these RNAs may assume a wild type structure (Figure 44). One possibility for reduced packaging, transduction and *in vitro* Gag binding of this mutant RNA could be attributed to the existence of these alternate RNA conformations. Since these mutations resulted in only 40% reduction of packaging efficiency, these nucleotides are not a critical requirement for packaging.

The *in vitro* and in cell assays showed that the PBS region binds to Gag and is a critical element for packaging (Figure 41). The PBS region contains two overlapping palindromic sequences, and a previous study has shown that deletion of these sequences results in a drastic reduction of dimerization ability of the MMTV gRNA (Aktar et al., 2014). By conducting a detailed mutational analysis of the PBS region combined with dimerization and packaging assays of the mutant RNAs, it was observed that the PBS does not play a direct role in the dimerization of MMTV gRNA since loss of its palindromic nature did not show any significant reduction in RNA dimerization (Figure 47B & Figure 48). On the other hand, the PBS sequence is critical for RNA packaging. Of note, a palindrome in the HIV-2 PBS has been

proposed to be involved in its dimerization (Jossinet et al., 2001), but to the best of our knowledge, a role of this PBS palindrome in HIV-2 gRNA packaging has never been demonstrated. In the case of HIV-1, the core Pr55^{Gag} binding domain encompasses the extreme 3' end of the PBS region (Smyth et al., 2015) and secondary Gag binding sites have been identified in this domain by footprinting (Damgaard et al., 1998; Kenyon et al., 2015). Studies using annealing of either tRNA^{Lys3} or oligonucleotides complementary to PBS showed an increase in HIV-1 RNA dimerization by enhancing the dimerization-competent RNA conformation, whereas deletion of PBS resulted in a moderate reduction in packaging (Brigham et al., 2019; Seif et al., 2013).

The precise role of PBS in HIV-1 gRNA packaging is rather elusive, essentially relying on drastic mutations that potentially could have compromised the global RNA secondary structure (Clever et al., 2002). Conversely, systematic point substitutions in the 5' region of the HIV-1 gRNA did not identify sequences in the PBS contributing to RNA packaging (Smyth et al., 2018). In the case of retrotransposon Ty1, the bipartite PBS located at both the 5' and 3' UTRs was observed to be necessary for packaging, possibly by acting as a dimerization site mediated by the hybridization of tRNA_i^{Met} (Gabus et al., 1998). Subsequently, it was shown that the Gag-induced dimerization was not required for annealing of tRNA to PBS (Gumna et al., 2019) and similarly in Ty3, the annealing of tRNA was also not required for RNA dimerization, suggesting that the bipartite PBS is not critical for gRNA packaging of either Ty1 or Ty3 retrotransposons (Clemens et al., 2013; Nymark-McMahon et al., 2002). In light of these observations, the role of the PBS

domain in Gag binding and RNA packaging presented in this study might be specific for MMTV or *Betaretroviruses*.

It has been suggested that in HIV-1, Pr55^{Gag} is involved in the initial placement of tRNA to the PBS, while NCp7 in mature virions facilitates the formation of more stable tRNA-gRNA complexes (Guo et al., 2009; Jin & Musier-Forsyth, 2019). It is also becoming clear that the selective packaging of tRNA in the assembling virions is facilitated by increasing the local concentration of lysyl-tRNA synthetase and is independent of gRNA (Abbink & Berkhout, 2003; Jossinet et al., 2001). During this process, Gag plays a central role in recruiting the tRNA^{Lys3}-lysyl-tRNA synthetase complex through the specific binding of its CA domain to lysyl-tRNA synthetase (reviewed in (Jin & Musier-Forsyth, 2019; Kleiman et al., 2010). Thus, in case of MMTV, it is possible that the Pr77^{Gag} binding to the PBS region may play dual role in MMTV replication by ensuring first selection of the gRNA during the early stages of the viral assembly process, then initial annealing/placement of tRNA^{Lys3} to the PBS once the immature particles are formed.

Taken together, the *in vitro* binding and in cell RNA packaging/transduction assays identify two regions, the architecturally accessible ssPurines loop of SL4 and the PBS domain, critical for the packaging of MMTV gRNA by Pr77^{Gag}. The spatial proximity of ssPurines to the DIS, which is reminiscent to the situation in HIV-1, may provide a molecular explanation for the strong link between gRNA dimerization and packaging that has been described for many retroviruses and retrotransposons. Consistent with this, packaging determinants have mostly been shown to overlap with, or found to be in close proximity to, sequences responsible for augmenting dimerization while having a higher order structure. On the contrary, the role of

MMTV PBS domain in gRNA packaging demonstrated here is most likely not a general phenomenon among retroviruses. Finally, the discrimination between unspliced and spliced RNAs begins at the initial stages of assembly and could primarily rely on maintaining the structural integrity of the bifurcated SL4. These results provide insights into the molecular mechanisms involved in the packaging of a much less studied *Betaretrovirus*, MMTV and allow distinguishing features that are conserved amongst divergent retroviruses, such as MMTV and HIV-1, from those that are virus-specific.

In conclusion, this study demonstrates that MMTV Gag recognizes ssPurines fundamentally on the basis of the higher order structure during gRNA packaging in an infected cell, suggesting that certain structural motifs in secondary or tertiary RNA conformation(s) mediate RNA-protein interaction. Since gRNA packaging during virus assembly is vital for the continuity of viral life cycle, these findings have ramifications towards the development of therapeutic interventions based on unique antiretroviral drugs that target virus assembly, especially given the fact that only retroviral Gag is required for virus particle formation. These drugs could then target specific structural elements (such as ssPurines and PBS) within the packaging signal RNA or its interaction domains within the Gag protein using novel small molecule approaches (Ding et al., 2020; Ingemarsdotter et al., 2018; Solinska, 2014). Unfortunately, a lack of basic understanding of these fundamental processes has prevented the development of such therapeutic modalities that can target retrovirus assembly. Thus, information gleaned from current study significantly adds to the developing literature of how retroviruses recognize their gRNA to initiate the process

of virion particle assembly which can facilitate the development of such novel antivirals into the realm of reality.

3.6 Funding

This research was funded primarily by a grant from the United Arab Emirates University (UAEU) Zayed Center for Health Sciences (UCBR-31R123) and in part by a grant from the College of Medicine and Health Sciences (31M280) to TAR, as well as by the RetroPack International Research Project from the CNRS to RM. AC and FNNP were supported by UCBR-31R123 and UAEU Program for Advanced Research (UPAR-31M233) grants respectively.

Chapter 4: Conclusions and Future Directions

4.1 Conclusions

The primary goal of current study was to delineate the molecular intricacies in the process of selective packaging of MMTV gRNA by Pr77^{Gag} during virus assembly, a crucial step for maintaining the continuity of virus life cycle. Due to the unavailability of MMTV full-length Pr77^{Gag}, as a first step, expression of full-length recombinant-His₆-tag fusion protein in bacteria was done and purified using IMAC and SEC (findings presented in Chapter 2). The purified Pr77^{Gag}-His₆-tag protein was shown to be functionally active by virtue of its assembly into VLPs *in vitro*. The presence of His₆-tag did not interfere with the ability of Gag to selectively bind gRNA as recombinant Pr77^{Gag}-His₆-tag fusion protein expressed in eukaryotic cells could make VLPs which efficiently package MMTV subgenomic RNAs.

The purified Pr77^{Gag}-His₆-tag protein was then used along with a set of *in vitro* transcribed RNAs to conduct band-shift, band-shift competition, filter binding and SHAPE-footprinting assays to identify Gag binding sites on the gRNA. The identified binding sites were then tested in a replication in cell gRNA packaging and transduction assay for their biological significance. These multi-pronged experimental approaches revealed that while ssPurines are required for high affinity Pr77^{Gag} binding, remarkably, the PBS also binds efficiently to Pr77^{Gag}. Additionally, the *in vivo* RNA packaging data further demonstrated that ssPurines and PBS are not redundant during the MMTV life cycle, as loss of either of these Pr77^{Gag} binding sites abrogated gRNA packaging. Findings presented in this dissertation reveal for the first time a direct role of the PBS in retroviral gRNA packaging. Future studies

are warranted to establish whether the involvement of PBS in gRNA packaging is unique to MMTV or such a phenomenon also exists in other retroviruses and/or LTR retrotransposons. Identifying the structural basis of RNA-protein interactions opens avenues towards developing novel anti-retroviral therapeutic interventions that target virus assembly since retroviral Gag is the only protein required to form virus particles and for packaging gRNA for the continuation of the virus life cycle.

4.2 Future Directions

While current study clearly demonstrates that both ssPurines and PBS play a crucial role in the packaging of MMTV gRNA by Pr77^{Gag}, several other questions arise from these observations. For example:

1. Footprinting experiments showed that ssPurines act as direct Gag binding site in addition to the 7 nucleotides in PBS that are also involved in direct Gag binding and are critical for gRNA packaging. However, it was also observed that nucleotides in PBS apart from these 7 Gag binding nucleotides also play a crucial role in RNA packaging. Hence the precise role of PBS region in RNA packaging is yet to be deciphered fully.
2. It is possible that the Gag can bind to nucleotides without blocking the 2'-OH moiety from modification by SHAPE reagents. If such a scenario is happening in the PBS region, additional footprinting experiments using chemicals such as dimethyl sulfate (DMS; so that methylation of unprotected adenosines and cytosines can be identified), kethoxal or CMCT (both modify the base of unpaired guanosine residues) may show the direct evidence of involvement of these nucleotides in Gag binding. Alternatively, high

resolution structural studies such as X-ray crystallography or solution NMR can also be employed for detecting the precise structure of the PBS-Gag complex.

3. Another possibility is that the nucleotides in PBS play an indirect role in RNA packaging through an intramolecular or intermolecular interaction that is necessary for maintaining a higher order structure which is necessary for RNA encapsidation during assembly. Also, one cannot dismiss the possibility that tRNA annealing to the PBS may stabilize the dimerization-competent conformation of the RNA, which in turn could enhance packaging, as has been suggested for HIV-1 (Brigham et al., 2019). Therefore, it will be important to perform further studies towards establishing whether similar mechanism(s) also exists during MMTV gRNA packaging or not.
4. In case of HIV-1, the U5: AUG long-range interaction (LRI) has been shown to be important for stabilizing the dimerization competent structure which could be favored during the RNA packaging process (Abbink et al., 2005; Abbink & Berkhout, 2003). In the SHAPE-validated structure of MMTV packaging determinants, similar U5: AUG interaction was identified. Therefore, additional studies are required to characterize the functional importance of such U5: Gag LRI in MMTV gRNA dimerization and its subsequent packaging.
5. This study has identified the minimal packaging elements needed for MMTV gRNA packaging; however, to prove this further, it would be important to introduce these regions into a heterologous RNA and assess its ability to bind to Pr77^{Gag} and/or its packageability in a biologically relevant system.

6. Finally, Gag protein is the major component that drives gRNA packaging and virus particle assembly. Elucidating near atomic resolution cryo-EM structures of *in vitro* assembled MMTV-like particles, viral assembly intermediates and Gag-RNA complexes would also help to better understand the molecular mechanisms involved in packaging and assembly process which in turn could be beneficial in designing therapeutic compounds that target these steps.

References

- Abbink, T. E. M., & Berkhout, B. (2003). A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *The Journal of Biological Chemistry*, *278*(13), 11601–11611. <https://doi.org/10.1074/jbc.M210291200>
- Abbink, T. E. M., Ooms, M., Haasnoot, P. C. J., & Berkhout, B. (2005). The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry*, *44*(25), 9058–9066. <https://doi.org/10.1021/bi0502588>
- 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. <https://doi.org/10.1038/ncomms5304>
- Abrahamyan, L. G., Chatel-Chaix, L., Ajamian, L., Milev, M. P., Monette, A., Clément, J.-F., Song, R., Lehmann, M., DesGroseillers, L., Laughrea, M., Boccaccio, G., & Mouland, A. J. (2010). Novel Staufen1 ribonucleoproteins prevent formation of stress granules but favour encapsidation of HIV-1 genomic RNA. *Journal of Cell Science*, *123*(3), 369–383. <https://doi.org/10.1242/jcs.055897>
- Acel, A., Udashkin, B. E., Wainberg, M. A., & Faust, E. A. (1998). Efficient Gap Repair Catalyzed In Vitro by an Intrinsic DNA Polymerase Activity of Human Immunodeficiency Virus Type 1 Integrase. *Journal of Virology*, *72*(3), 2062–2071. <https://doi.org/10.1128/JVI.72.3.2062-2071.1998>
- Affranchino, J. L., & González, S. A. (2010). In vitro assembly of the feline immunodeficiency virus Gag polyprotein. *Virus Research*, *150*(1), 153–157. <https://doi.org/10.1016/j.virusres.2010.03.012>
- 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. <https://doi.org/10.1261/rna.040931.113>
- 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. <https://doi.org/10.1186/s12977-014-0096-6>

- Aldovini, A., & Young, R. A. (1990). Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *Journal of Virology*, *64*(5), 1920–1926. <https://doi.org/10.1128/jvi.64.5.1920-1926.1990>
- Alfadhli, A., & Barklis, E. (2014). The roles of lipids and nucleic acids in HIV-1 assembly. *Frontiers in Microbiology*, *5*, 253. <https://doi.org/10.3389/fmicb.2014.00253>
- Ali, L. M., Rizvi, T. A., & Mustafa, F. (2016). Cross- and co-packaging of retroviral RNAs and their consequences. *Viruses*, *8*(10), 276. <https://doi.org/10.3390/v8100276>
- Aronoff, R., & Linial, M. (1991). Specificity of retroviral RNA packaging. *Journal of Virology*, *65*(1), 71–80. <https://doi.org/10.1128/JVI.65.1.71-80.1991>
- Ashe, M. P., Griffin, P., James, W., & Proudfoot, N. J. (1995). Poly(A) site selection in the HIV-1 provirus: Inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes & Development*, *9*(23), 3008–3025. <https://doi.org/10.1101/gad.9.23.3008>
- 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. <https://doi.org/10.1128/JVI.01521-08>
- Balvay, L., Lopez Lastra, M., Sargueil, B., Darlix, J.-L., & Ohlmann, T. (2007). Translational control of retroviruses. *Nature Reviews. Microbiology*, *5*(2), 128–140. <https://doi.org/10.1038/nrmicro1599>
- Banks, J. D., Kealoha, B. O., & Linial, M. L. (1999). An MΨ-containing heterologous RNA, but not env mRNA, is efficiently packaged into avian retroviral particles. *Journal of Virology*, *73*(11), 8926–8933. <https://doi.org/10.1128/JVI.73.11.8926-8933.1999>
- Barajas, B. C., Tanaka, M., Robinson, B. A., Phuong, D. J., Chutiraka, K., Reed, J. C., & Lingappa, J. R. (2018). Identifying the assembly intermediate in which Gag first associates with unspliced HIV-1 RNA suggests a novel model for HIV-1 RNA packaging. *PLoS Pathogens*, *14*(4), e1006977. <https://doi.org/10.1371/journal.ppat.1006977>

- Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dautuet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., & Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, *220*(4599), 868–871. <https://doi.org/10.1126/science.6189183>
- Becker, J. T., & Sherer, N. M. (2017). Subcellular localization of HIV-1 gag-pol mRNAs regulates sites of virion assembly. *Journal of Virology*, *91*(6), e02315-16. <https://doi.org/10.1128/JVI.02315-16>
- Behrens, R. T., Aligeti, M., Pocock, G. M., Higgins, C. A., & Sherer, N. M. (2017). Nuclear export signal masking regulates HIV-1 Rev trafficking and viral RNA nuclear export. *Journal of Virology*, *91*(3). <https://doi.org/10.1128/JVI.02107-16>
- Bell, N. M., & Lever, A. M. L. (2013). HIV Gag polyprotein: Processing and early viral particle assembly. *Trends in Microbiology*, *21*(3), 136–144. <https://doi.org/10.1016/j.tim.2012.11.006>
- Bell, P. A. (2001). E.coli Expression Systems. In A. S. Gerstein (Ed.), *Molecular Biology Problem Solver* (pp. 461–490). John Wiley & Sons, Inc. <https://doi.org/10.1002/0471223905.ch15>
- 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. <https://doi.org/10.1128/JVI.70.10.6723-6732.1996>
- Berkhout, Ben, Ooms, M., Beerens, N., Huthoff, H., Southern, E., & Verhoef, K. (2002). In vitro evidence that the untranslated leader of the HIV-1 genome is an RNA checkpoint that regulates multiple functions through conformational changes. *The Journal of Biological Chemistry*, *277*(22), 19967–19975. <https://doi.org/10.1074/jbc.M200950200>
- Berlioz, C., & Darlix, J. L. (1995). An internal ribosomal entry mechanism promotes translation of murine leukemia virus gag polyprotein precursors. *Journal of Virology*, *69*(4), 2214–2222. <https://doi.org/10.1128/JVI.69.4.2214-2222.1995>
- Bernacchi, S., Abd El-Wahab, E. W., Dubois, N., Hijnen, M., Smyth, R. P., Mak, J., Marquet, R., & Paillart, J.C. (2017). HIV-1 Pr55Gag binds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biology*, *14*(1), 90–103. <https://doi.org/10.1080/15476286.2016.1256533>

- Bewley, M. C., Reinhart, L., Stake, M. S., Nadaraia-Hoke, S., Parent, L. J., & Flanagan, J. M. (2017). A non-cleavable hexahistidine affinity tag at the carboxyl-terminus of the HIV-1 Pr55Gag polyprotein alters nucleic acid binding properties. *Protein Expression and Purification*, 130(Supplement C), 137–145. <https://doi.org/10.1016/j.pep.2016.10.001>
- Bittner, John J. (1936). Some possible effects of nursing on the mammary gland tumor incidence in mice. *Science*, 84(2172), 162–162. <https://doi.org/10.1126/science.84.2172.162>
- Bleul, C. C., Wu, L., Hoxie, J. A., Springer, T. A., & Mackay, C. R. (1997). The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America*, 94(5), 1925–1930. <https://doi.org/10.1073/pnas.94.5.1925>
- Blissenbach, M., Grewe, B., Hoffmann, B., Brandt, S., & Überla, K. (2010). Nuclear RNA export and packaging functions of HIV-1 Rev revisited. *Journal of Virology*, 84(13), 6598–6604. <https://doi.org/10.1128/JVI.02264-09>
- Boeke, J. D., & Stoye, J. P. (1997). Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In J. M. Coffin, S. H. Hughes, & H. E. Varmus (Eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press. <http://www.ncbi.nlm.nih.gov/books/NBK19468/>
- 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(1), 68. <https://doi.org/10.1186/1742-4690-2-68>
- Brandt, S., Blißenbach, M., Grewe, B., Konietzny, R., Grunwald, T., & Überla, K. (2007). Rev proteins of human and simian immunodeficiency virus enhance RNA encapsidation. *PLOS Pathogens*, 3(4), e54. <https://doi.org/10.1371/journal.ppat.0030054>
- 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. <https://doi.org/10.1073/pnas.91.4.1256>
- Briggs, J. A. G., Simon, M. N., Gross, I., Kräusslich, H.-G., Fuller, S. D., Vogt, V. M., & Johnson, M. C. (2004). The stoichiometry of Gag protein in HIV-1. *Nature Structural & Molecular Biology*, 11(7), 672–675. <https://doi.org/10.1038/nsmb785>

- Brigham, B. S., Kitzrow, J. P., Reyes, J.-P. C., Musier-Forsyth, K., & Munro, J. B. (2019). Intrinsic conformational dynamics of the HIV-1 genomic RNA 5'UTR. *Proceedings of the National Academy of Sciences*, *116*(21), 10372–10381. <https://doi.org/10.1073/pnas.1902271116>
- Brookes, S., Placzek, M., Moore, R., Dixon, M., Dickson, C., & Peters, G. (1986). Insertion elements and transitions in cloned mouse mammary tumour virus DNA: Further delineation of the poison sequences. *Nucleic Acids Research*, *14*(21), 8231–8245. <https://doi.org/10.1093/nar/14.21.8231>
- 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. <https://doi.org/10.1126/science.aaz7959>
- Brown, P. O., Bowerman, B., Varmus, H. E., & Bishop, J. M. (1989). Retroviral integration: Structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(8), 2525–2529. <https://doi.org/10.1073/pnas.86.8.2525>
- Bukrinsky, M. (2004). A hard way to the nucleus. *Molecular Medicine (Cambridge, Mass.)*, *10*(1–6), 1–5.
- Bukrinsky, M I, Sharova, N., Dempsey, M. P., Stanwick, T. L., Bukrinskaya, A. G., Haggerty, S., & Stevenson, M. (1992). Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(14), 6580–6584. <https://doi.org/10.1371/journal.ppat.1006977>
- Bukrinsky, Michael I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubei, 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. <https://doi.org/10.1038/365666a0>
- 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. <https://doi.org/10.1073/pnas.1920631117>

- Burniston, M. T., Cimarelli, A., Colgan, J., Curtis, S. P., & Luban, J. (1999). Human immunodeficiency virus type 1 Gag polyprotein multimerization requires the nucleocapsid domain and RNA and is promoted by the capsid-dimer interface and the basic region of matrix protein. *Journal of Virology*, *73*(10), 8527–8540. <https://doi.org/10.1128/JVI.73.10.8527-8540.1999>
- Bussienne, C., Marquet, R., Paillart, J.C., & Bernacchi, S. (2021). Post-translational modifications of retroviral HIV-1 Gag precursors: an overview of their biological role. *International Journal of Molecular Sciences*, *22*(6), 2871. <https://doi.org/10.3390/ijms22062871>
- Butsch, M., & Boris-Lawrie, K. (2002). Destiny of unspliced retroviral RNA: ribosome and/or virion? *Journal of Virology*, *76*(7), 3089–3094. <https://doi.org/10.1128/JVI.76.7.3089-3094.2002>
- Callahan, R., & Smith, G. H. (2008). The mouse as a model for mammary tumorigenesis: History and current aspects. *Journal of Mammary Gland Biology and Neoplasia*, *13*(3), 269. <https://doi.org/10.1007/s10911-008-9094-4>
- Campbell, S., & Vogt, V. M. (1997). In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: Identification of the p10 domain as a morphological determinant in the formation of spherical particles. *Journal of Virology*, *71*(6), 4425–4435. <https://doi.org/10.1128/JVI.71.6.4425-4435.1997>
- Campbell, Stephen, & Rein, A. (1999). In vitro assembly properties of human immunodeficiency virus type 1 gag protein lacking the p6 domain. *Journal of Virology*, *73*(3), 2270–2279. <https://doi.org/10.1128/JVI.73.3.2270-2279.1999>
- Canaani, E., Helm, K. V., & Duesberg, P. (1973). Evidence for 30-40S RNA as precursor of the 60-70S RNA of Rous sarcoma virus. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(2), 401–405. <https://doi.org/10.1073/pnas.70.2.401>
- Cardiff, R. D., & Kenney, N. (2007). Mouse mammary tumor biology: A short history. *Advances in Cancer Research*, *98*, 53–116. [https://doi.org/10.1016/S0065-230X\(06\)98003-8](https://doi.org/10.1016/S0065-230X(06)98003-8)
- Carlson, L.-A., Briggs, J. A. G., Glass, B., Riches, J. D., Simon, M. N., Johnson, M. C., Müller, B., Grünewald, K., & Kräusslich, H.-G. (2008). Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host & Microbe*, *4*(6), 592–599. <https://doi.org/10.1016/j.chom.2008.10.013>

- 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. <https://doi.org/10.3390/v10060334>
- Chamorro, M., Parkin, N., & Varmus, H. E. (1992). An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(2), 713–717. <https://doi.org/10.1073/pnas.89.2.713>
- Chen, J., Liu, Y., Wu, B., Nikolaitchik, O. A., Mohan, P. R., Chen, J., Pathak, V. K., & Hu, W.-S. (2020). Visualizing the translation and packaging of HIV-1 full-length RNA. *Proceedings of the National Academy of Sciences*, *117*(11), 6145–6155. <https://doi.org/10.1073/pnas.1917590117>
- Chen, X., Chamorro, M., Lee, S. I., Shen, L. X., Hines, J. V., Tinoco, I., & Varmus, H. E. (1995). Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: Nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting. *The EMBO Journal*, *14*(4), 842–852. <https://doi.org/10.1002/j.1460-2075.1995.tb07062.x>
- 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. <https://doi.org/10.1073/pnas.0908661107>
- Cimarelli, A., Sandin, S., Höglund, S., & Luban, J. (2000). Basic residues in human immunodeficiency virus type 1 nucleocapsid promote virion assembly via interaction with RNA. *Journal of Virology*, *74*(7), 3046–3057. <https://doi.org/10.1128/jvi.74.7.3046-3057.2000>
- Clemens, K., Bilanchone, V., Beliakova-Bethell, N., Larsen, L. S. Z., Nguyen, K., & Sandmeyer, S. (2013). Sequence requirements for localization and packaging of Ty3 retroelement RNA. *Virus Research*, *171*(2), 319–331. <https://doi.org/10.1016/j.virusres.2012.10.008>
- Clever, J. L., Wong, M. L., & Parslow, T. G. (1996). Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *Journal of Virology*, *70*(9), 5902–5908. <https://doi.org/10.1128/JVI.70.9.5902-5908.1996>

- Clever, Jared 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. <https://doi.org/10.1128/JVI.76.23.12381-12387.2002>
- Cochrane, A. W., McNally, M. T., & Mouland, A. J. (2006). The retrovirus RNA trafficking granule: From birth to maturity. *Retrovirology*, 3, 18. <https://doi.org/10.1186/1742-4690-3-18>
- Coffin, J. M., Hughes, S. H., & Varmus, H. E. (Eds.). (1997). *Retroviruses*. Cold Spring Harbor Laboratory Press. <http://www.ncbi.nlm.nih.gov/books/NBK19376/>
- Comas-Garcia, M., Datta, S. A., Baker, L., Varma, R., Gudla, P. R., & Rein, A. (2017). Dissection of specific binding of HIV-1 Gag to the “packaging signal” in viral RNA. *ELife*, 6, e27055. <https://doi.org/10.7554/eLife.27055>
- Comas-Garcia, M., Davis, S. R., & Rein, A. (2016). On the selective packaging of genomic RNA by HIV-1. *Viruses*, 8(9), 246. <https://doi.org/10.3390/v8090246>
- Comas-Garcia, M., Kroupa, T., Datta, S. A., Harvin, D. P., Hu, W.-S., & Rein, A. (2018). Efficient support of virus-like particle assembly by the HIV-1 packaging signal. *ELife*, 7, e38438. <https://doi.org/10.7554/eLife.38438>
- Dalglish, 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. <https://doi.org/10.1038/312763a0>
- 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. <https://doi.org/10.1093/nar/26.16.3667>
- Darty, K., Denise, A., & Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25(15), 1974–1975. <https://doi.org/10.1093/bioinformatics/btp250>
- Das, A. T., Klaver, B., & Berkhout, B. (1995). Reduced replication of human immunodeficiency virus type 1 mutants that use reverse transcription primers other than the natural tRNA(3Lys). *Journal of Virology*, 69(5), 3090–3097. <https://doi.org/10.1128/JVI.69.5.3090-3097.1995>

- Dawson, L., & Yu, X. F. (1998). The role of nucleocapsid of HIV-1 in virus assembly. *Virology*, *251*(1), 141–157. <https://doi.org/10.1006/viro.1998.9374>
- Dayton, A. I. (2004). Within you, without you: HIV-1 Rev and RNA export. *Retrovirology*, *1*, 35. <https://doi.org/10.1186/1742-4690-1-35>
- De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., & Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*, *279*(5349), 384–388. <https://doi.org/10.1126/science.279.5349.384>
- 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. <https://doi.org/10.1038/s41564-020-0735-8>
- 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. <https://doi.org/10.1093/nar/gkr577>
- Dilley, K. A., Nikolaitchik, O. A., Galli, A., Burdick, R. C., Levine, L., Li, K., Rein, A., Pathak, V. K., & Hu, W.-S. (2017). Interactions between HIV-1 Gag and Viral RNA Genome Enhance Virion Assembly. *Journal of Virology*, *91*(16), e02319-16. <https://doi.org/10.1128/JVI.02319-16>
- 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. <https://doi.org/10.1073/pnas.2008519117>
- Doms, R. W., & Moore, J. P. (2000). HIV-1 membrane fusion. *The Journal of Cell Biology*, *151*(2), 9–14. <https://doi.org/10.1083/jcb.151.2.F9>
- Dorfman, T., Luban, J., Goff, S. P., Haseltine, W. A., & Göttlinger, H. G. (1993). Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *Journal of Virology*, *67*(10), 6159–6169. <https://doi.org/10.1128/JVI.67.10.6159-6169.1993>

- Dragic, T., & Alizon, M. (1993). Different requirements for membrane fusion mediated by the envelopes of human immunodeficiency virus types 1 and 2. *Journal of Virology*, *67*(4), 2355–2359. <https://doi.org/10.1128/JVI.67.4.2355-2359.1993>
- D'Souza, V., & Summers, M. F. (2004). Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature*, *431*(7008), 586. <https://doi.org/10.1038/nature02944>
- D'Souza, V., & Summers, M. F. (2005). How retroviruses select their genomes. *Nature Reviews Microbiology*, *3*(8), 643. <https://doi.org/10.1038/nrmicro1210>
- Dubois, N., Khoo, K. K., Ghossein, S., Seissler, T., Wolff, P., McKinsty, W. J., Mak, J., Paillart, J.C., Marquet, R., & Bernacchi, S. (2018). The C-terminal p6 domain of the HIV-1 Pr55Gag precursor is required for specific binding to the genomic RNA. *RNA Biology*, *15*(7), 923–936. <https://doi.org/10.1080/15476286.2018.1481696>
- Dubois, N., Marquet, R., Paillart, J.C., & Bernacchi, S. (2018). Retroviral RNA dimerization: from structure to functions. *Frontiers in Microbiology*, *9*, 527. <https://doi.org/10.3389/fmicb.2018.00527>
- Dudley, J. P., Golovkina, T. V., & Ross, S. R. (2016). Lessons learned from mouse mammary tumor virus in animal models. *ILAR Journal*, *57*(1), 12–23. <https://doi.org/10.1093/ilar/ilv044>
- Duesberg, P. H., & Blair, P. B. (1966). Isolation of the nucleic acid of mouse mammary tumor virus (MTV). *Proceedings of the National Academy of Sciences of the United States of America*, *55*(6), 1490–1497. <https://doi.org/10.1073/pnas.55.6.1490>
- Dupont, S., Sharova, N., DéHoratius, C., Virbasius, C.-M. A., Zhu, X., Bukrinskaya, A. G., Stevenson, M., & Green, M. R. (1999). A novel nuclear export activity in HIV-1 matrix protein required for viral replication. *Nature*, *402*(6762), 681–685. <https://doi.org/10.1038/45272>
- Dupraz, P., & Spahr, P. F. (1992). Specificity of Rous sarcoma virus nucleocapsid protein in genomic RNA packaging. *Journal of Virology*, *66*(8), 4662–4670. <https://doi.org/10.1128/jvi.66.8.4662-4670.1992>
- Dvorin, J. D., & Malim, M. H. (2003). Intracellular trafficking of HIV-1 cores: Journey to the center of the cell. *Current Topics in Microbiology and Immunology*, *281*, 179–208. https://doi.org/10.1007/978-3-642-19012-4_5

- Ehrlich, L. S., Agresta, B. E., & Carter, C. A. (1992). Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro. *Journal of Virology*, 66(8), 4874–4883. <https://doi.org/10.1128/JVI.66.8.4874-4883.1992>
- Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., & Dumas, P. (2001). Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nature Structural Biology*, 8(12), 1064–1068. <https://doi.org/10.1038/nsb727>
- Espah Borujeni, A., Channarasappa, A. S., & Salis, H. M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, 42(4), 2646–2659. <https://doi.org/10.1093/nar/gkt1139>
- Farnet, C. M., & Haseltine, W. A. (1991). Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *Journal of Virology*, 65(4), 1910–1915. <https://doi.org/10.1128/JVI.65.4.1910-1915.1991>
- Feinberg, M. B., Baltimore, D., & Frankel, A. D. (1991). The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. *Proceedings of the National Academy of Sciences of the United States of America*, 88(9), 4045–4049. <https://doi.org/10.1073/pnas.88.9.4045>
- Feng, Y. X., Copeland, T. D., Henderson, L. E., Gorelick, R. J., Bosche, W. J., Levin, J. G., & Rein, A. (1996). HIV-1 nucleocapsid protein induces “maturation” of dimeric retroviral RNA in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 93(15), 7577–7581. <https://doi.org/10.1073/pnas.93.15.7577>
- Francis, A. C., Marin, M., Prellberg, M. J., Palermino-Rowland, K., & Melikyan, G. B. (2020). HIV-1 uncoating and nuclear import precede the completion of reverse transcription in cell lines and in primary macrophages. *Viruses*, 12(11), 1234. <https://doi.org/10.3390/v12111234>
- Freed, E. O. (2015). HIV-1 assembly, release and maturation. *Nature Reviews. Microbiology*, 13(8), 484–496. <https://doi.org/10.1038/nrmicro3490>
- Fu, W., Gorelick, R. J., & Rein, A. (1994). Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *Journal of Virology*, 68(8), 5013–5018. <https://doi.org/10.1128/JVI.68.8.5013-5018.1994>

- Füzik, T., Píchalová, R., Schur, F. K. M., Strohalmová, K., Křížová, I., Hadravová, R., Rumlová, M., Briggs, J. A. G., Ulbrich, P., & Ruml, T. (2016). Nucleic acid binding by Mason-Pfizer monkey virus CA promotes virus assembly and genome packaging. *Journal of Virology*, *90*(9), 4593–4603.
<https://doi.org/10.1128/JVI.03197-15>
- Gabus, C., Ficheux, D., Rau, M., Keith, G., Sandmeyer, S., & Darlix, J. L. (1998). The yeast Ty3 retrotransposon contains a 5'-3' bipartite primer-binding site and encodes nucleocapsid protein NCp9 functionally homologous to HIV-1 NCp7. *The EMBO Journal*, *17*(16), 4873–4880.
<https://doi.org/10.1093/emboj/17.16.4873>
- Gallo, R. C. (1988). HIV--the cause of AIDS: An overview on its biology, mechanisms of disease induction, and our attempts to control it. *Journal of Acquired Immune Deficiency Syndromes*, *1*(6), 521–535.
- Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., & Safai, B. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science (New York, N.Y.)*, *224*(4648), 500–503. <https://doi.org/10.1126/science.6200936>
- Ganser-Pornillos, B. K., Yeager, M., & Pornillos, O. (2012). Assembly and architecture of HIV. *Advances in Experimental Medicine and Biology*, *726*, 441–465. https://doi.org/10.1007/978-1-4614-0980-9_20
- 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.
<https://doi.org/10.1128/JVI.00101-09>
- Garrus, J. E., von Schwedler, U. K., Pornillos, O. W., Morham, S. G., Zavitz, K. H., Wang, H. E., Wettstein, D. A., Stray, K. M., Côté, M., Rich, R. L., Myszka, D. G., & Sundquist, W. I. (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell*, *107*(1), 55–65.
[https://doi.org/10.1016/s0092-8674\(01\)00506-2](https://doi.org/10.1016/s0092-8674(01)00506-2)
- Ghazawi, A., Mustafa, F., Phillip, P. S., Jayanth, P., Ali, J., & Rizvi, T. A. (2006). Both the 5' and 3' LTRs of FIV contain minor RNA encapsidation determinants compared to the two core packaging determinants within the 5' untranslated region and gag. *Microbes and Infection*, *8*(3), 767–778.
<https://doi.org/10.1016/j.micinf.2005.09.015>

- 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. <https://doi.org/10.1073/pnas.1006897107>
- Ghez, D., Lepelletier, Y., Lambert, S., Fourneau, J.-M., Blot, V., Janvier, S., Arnulf, B., van Endert, P. M., Heveker, N., Pique, C., & Hermine, O. (2006). Neuropilin-1 is involved in human T-cell lymphotropic virus type 1 entry. *Journal of Virology*, *80*(14), 6844–6854. <https://doi.org/10.1128/JVI.02719-05>
- Gorelick, R. J., Henderson, L. E., Hanser, J. P., & Rein, A. (1988). Point mutants of Moloney murine leukemia virus that fail to package viral RNA: Evidence for specific RNA recognition by a “zinc finger-like” protein sequence. *Proceedings of the National Academy of Sciences*, *85*(22), 8420–8424. <https://doi.org/10.1073/pnas.85.22.8420>
- Göttlinger, H. G., Dorfman, T., Cohen, E. A., & Haseltine, W. A. (1993). Vpu protein of human immunodeficiency virus type 1 enhances the release of capsids produced by gag gene constructs of widely divergent retroviruses. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(15), 7381–7385. <https://doi.org/10.1073/pnas.90.15.7381>
- Göttlinger, H. G., Dorfman, T., Sodroski, J. G., & Haseltine, W. A. (1991). Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(8), 3195–3199. <https://doi.org/10.1073/pnas.88.8.3195>
- Göttlinger, H. G., Sodroski, J. G., & Haseltine, W. A. (1989). Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(15), 5781–5785. <https://doi.org/10.1073/pnas.86.15.5781>
- Grassmann, R., Berchtold, S., Aepinus, C., Ballaun, C., Boehnlein, E., & Fleckenstein, B. (1991). In vitro binding of human T-cell leukemia virus rex proteins to the rex-response element of viral transcripts. *Journal of Virology*, *65*(7), 3721–3727. <https://doi.org/10.1128/JVI.65.7.3721-3727.1991>
- Greatorex, J. S., Laisse, V., Dockhelar, M. C., & Lever, A. M. (1996). Sequences involved in the dimerisation of human T cell leukaemia virus type-1 RNA. *Nucleic Acids Research*, *24*(15), 2919–2923. <https://doi.org/10.1093/nar/24.15.2919>

- Grewe, B., Ehrhardt, K., Hoffmann, B., Blissenbach, M., Brandt, S., & Uberla, K. (2012). The HIV-1 Rev protein enhances encapsidation of unspliced and spliced, RRE-containing lentiviral vector RNA. *PLoS One*, 7(11), e48688. <https://doi.org/10.1371/journal.pone.0048688>
- Gumna, J., Purzycka, K. J., Ahn, H. W., Garfinkel, D. J., & Pachulska-Wieczorek, K. (2019). Retroviral-like determinants and functions required for dimerization of Ty1 retrotransposon RNA. *RNA Biology*, 16(12), 1749–1763. <https://doi.org/10.1080/15476286.2019.1657370>
- Guntaka, R. V. (1993). Transcription termination and polyadenylation in retroviruses. *Microbiological Reviews*, 57(3), 511–521. <https://doi.org/10.1128/MR.57.3.511-521.1993>
- Günzburg, W. H., & Salmons, B. (1992). Factors controlling the expression of mouse mammary tumour virus. *Biochemical Journal*, 283(Pt 3), 625–632. <https://doi.org/10.1042/bj2830625>
- Guo, F., Saadatmand, J., Niu, M., & Kleiman, L. (2009). Roles of Gag and NCp7 in facilitating tRNA(Lys)(3) Annealing to viral RNA in human immunodeficiency virus type 1. *Journal of Virology*, 83(16), 8099–8107. <https://doi.org/10.1128/JVI.00488-09>
- Hadzopoulou-Cladaras, M., Felber, B. K., Cladaras, C., Athanassopoulos, A., Tse, A., & Pavlakis, G. N. (1989). The rev (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a cis-acting sequence in the env region. *Journal of Virology*, 63(3), 1265–1274. <https://doi.org/10.1128/JVI.63.3.1265-1274.1989>
- Ham, J., Thomson, A., Needham, M., Webb, P., & Parker, M. (1988). Characterization of response elements for androgens, glucocorticoids and progestins in mouse mammary tumour virus. *Nucleic Acids Research*, 16(12), 5263–5276. <https://doi.org/10.1093/nar/16.12.5263>
- Hanly, S. M., Rimsky, L. T., Malim, M. H., Kim, J. H., Hauber, J., Duc Dodon, M., Le, S. Y., Maizel, J. V., Cullen, B. R., & Greene, W. C. (1989). Comparative analysis of the HTLV-I Rex and HIV-1 Rev trans-regulatory proteins and their RNA response elements. *Genes & Development*, 3(10), 1534–1544. <https://doi.org/10.1101/gad.3.10.1534>
- Helga-Maria, C., Hammarskjöld, M.-L., & Rekosh, D. (1999). An intact TAR element and cytoplasmic localization are necessary for efficient packaging of human immunodeficiency virus type 1 genomic RNA. *Journal of Virology*, 73(5), 4127–4135. <https://doi.org/10.1128/JVI.73.5.4127-4135.1999>

- 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. <https://doi.org/10.1016/j.jmb.2012.01.033>
- Hildreth, J. E., Subramaniam, A., & Hampton, R. A. (1997). Human T-cell lymphotropic virus type 1 (HTLV-1)-induced syncytium formation mediated by vascular cell adhesion molecule-1: Evidence for involvement of cell adhesion molecules in HTLV-1 biology. *Journal of Virology*, *71*(2), 1173–1180. <https://doi.org/10.1128/JVI.71.2.1173-1180.1997>
- Hill, M. K., Shehu-Xhilaga, M., Campbell, S. M., Pountourios, P., Crowe, S. M., & Mak, J. (2003). The dimer initiation sequence stem-loop of human immunodeficiency virus type 1 is dispensable for viral replication in peripheral blood mononuclear cells. *Journal of Virology*, *77*(15), 8329–8335. <https://doi.org/10.1128/jvi.77.15.8329-8335.2003>
- Hizi, A., Henderson, L. E., Copeland, T. D., Sowder, R. C., Hixson, C. V., & Oroszlan, S. (1987). Characterization of mouse mammary tumor virus gag-pro gene products and the ribosomal frameshift site by protein sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(20), 7041–7045. <https://doi.org/10.1073/pnas.84.20.7041>
- Hizi, A., Henderson, L. E., Copeland, T. D., Sowder, R. C., Krutzsch, H. C., & Oroszlan, S. (1989). Analysis of gag proteins from mouse mammary tumor virus. *Journal of Virology*, *63*(6), 2543–2549. <https://doi.org/10.1128/jvi.63.6.2543-2549.1989>
- Hizi, Amnon, & Herzig, E. (2015). dUTPase: The frequently overlooked enzyme encoded by many retroviruses. *Retrovirology*, *12*, 70. <https://doi.org/10.1186/s12977-015-0198-9>
- Holman, A. G., & Coffin, J. M. (2005). Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(17), 6103–6107. <https://doi.org/10.1073/pnas.0501646102>
- Hook, L. M., Agafonova, Y., Ross, S. R., Turner, S. J., & Golovkina, T. V. (2000). Genetics of mouse mammary tumor virus-induced mammary tumors: linkage of tumor induction to the gag gene. *Journal of Virology*, *74*(19), 8876–8883. <https://doi.org/10.1128/JVI.74.19.8876-8883.2000>

- Housset, V., De Rocquigny, H., Roques, B. P., & Darlix, J. L. (1993). Basic amino acids flanking the zinc finger of Moloney murine leukemia virus nucleocapsid protein NCp10 are critical for virus infectivity. *Journal of Virology*, *67*(5), 2537–2545. <https://doi.org/10.1128/JVI.67.5.2537-2545.1993>
- Houzet, L., Paillart, J. C., Smagulova, F., Maurel, S., Morichaud, Z., Marquet, R., & Mougel, M. (2007). HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic Acids Research*, *35*(8), 2695–2704. <https://doi.org/10.1093/nar/gkm153>
- Hu, W. S., Bowman, E. H., Delviks, K. A., & Pathak, V. K. (1997). Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference. *Journal of Virology*, *71*(8), 6028–6036. doi: 10.1128/JVI.71.8.6028-6036.1997.
- Huang, Q., Purzycka, K. J., Lusvarghi, S., Li, D., Legrice, S. F. J., & Boeke, J. D. (2013). Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. *RNA (New York, N.Y.)*, *19*(3), 320–332. <https://doi.org/10.1261/rna.035535.112>
- 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. <https://doi.org/10.1016/B978-012374410-4.00677-4>
- Huthoff, H., & Berkhout, B. (2001). Two alternating structures of the HIV-1 leader RNA. *RNA*, *7*(1), 143–157. <https://doi.org/10.1017/s1355838201001881>.
- Huthoff, H., & Berkhout, B. (2002). Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry*, *41*(33), 10439–10445. <https://doi.org/10.1021/bi025993n>
- Indik, S. (2016). Mouse mammary tumor virus-based vector for efficient and safe transgene delivery into mitotic and non-mitotic cells. *Cell Gene Therapy Insights*, *2*(5), 589–597. <https://doi.org/10.18609/cgti.2016.062>
- Indik, S., Günzburg, W. H., Salmons, B., & Rouault, F. (2005a). A novel, mouse mammary tumor virus encoded protein with Rev-like properties. *Virology*, *337*(1), 1–6. <https://doi.org/10.1016/j.virol.2005.03.040>
- Indik, S., Günzburg, W. H., Salmons, B., & Rouault, F. (2005b). Mouse mammary tumor virus infects human cells. *Cancer Research*, *65*(15), 6651–6659. <https://doi.org/10.1158/0008-5472.CAN-04-2609>

- Ingemarsdotter, C. K., Zeng, J., Long, Z., Lever, A. M. L., & Kenyon, J. C. (2018). An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. *Retrovirology*, *15*, 25. <https://doi.org/10.1186/s12977-018-0407-4>
- Iwasaki, K., & Temin, H. M. (1990). The efficiency of RNA 3'-end formation is determined by the distance between the cap site and the poly(A) site in spleen necrosis virus. *Genes & Development*, *4*(12B), 2299–2307. <https://doi.org/10.1101/gad.4.12b.2299>
- 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. <https://doi.org/10.1016/j.jmb.2010.06.043>
- Jacks, T., Madhani, H. D., Masiarz, F. R., & Varmus, H. E. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell*, *55*(3), 447–458. [https://doi.org/10.1016/0092-8674\(88\)90031-1](https://doi.org/10.1016/0092-8674(88)90031-1)
- Jacks, T, Townsley, K., Varmus, H. E., & Majors, J. (1987). Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(12), 4298–4302. <https://doi.org/10.1073/pnas.84.12.4298>
- Jacks, Tyler, Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, *331*(6153), 280–283. <https://doi.org/10.1038/331280a0>
- Jern, P., & Coffin, J. M. (2008). Effects of retroviruses on host genome function. *Annual Review of Genetics*, *42*, 709–732. <https://doi.org/10.1146/annurev.genet.42.110807.091501>
- 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. <https://doi.org/10.1099/0022-1317-81-8-1889>
- Jin, D., & Musier-Forsyth, K. (2019). Role of host tRNAs and aminoacyl-tRNA synthetases in retroviral replication. *The Journal of Biological Chemistry*, *294*(14), 5352–5364. <https://doi.org/10.1074/jbc.REV118.002957>

- Johnson, S. F., & Telesnitsky, A. (2010). Retroviral RNA dimerization and packaging: the what, how, when, where, and why. *PLoS Pathogens*, 6(10), e1001007. <https://doi.org/10.1371/journal.ppat.1001007>
- Jones, J. S., Allan, R. W., & Temin, H. M. (1994). One retroviral RNA is sufficient for synthesis of viral DNA. *Journal of Virology*, 68(1), 207–216. <https://doi.org/10.1128/JVI.68.1.207-216.1994>
- Jones, K. S., Petrow-Sadowski, C., Bertolette, D. C., Huang, Y., & Ruscetti, F. W. (2005). Heparan sulfate proteoglycans mediate attachment and entry of human T-cell leukemia virus type 1 virions into CD4⁺ T cells. *Journal of Virology*, 79(20), 12692–12702. <https://doi.org/10.1128/JVI.79.20.12692-12702.2005>
- Jossinet, F., Lodmell, J. S., Ehresmann, C., Ehresmann, B., & Marquet, R. (2001). Identification of the in vitro HIV-2/SIV RNA dimerization site reveals striking differences with HIV-1. *Journal of Biological Chemistry*, 276(8), 5598–5604. <https://doi.org/10.1074/jbc.M008642200>
- Jouvenet, N., Lainé, S., Pessel-Vivares, L., & Mougél, M. (2011). Cell biology of retroviral RNA packaging. *RNA Biology*, 8(4), 572–580. <https://doi.org/10.4161/rna.8.4.16030>
- 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), 257. <https://doi.org/10.3390/v8090257>
- 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. <https://doi.org/10.1261/rna.036327.112>
- Katoh, I., Kyushiki, H., Sakamoto, Y., Ikawa, Y., & Yoshinaka, Y. (1991). Bovine leukemia virus matrix-associated protein MA(p15): Further processing and formation of a specific complex with the dimer of the 5'-terminal genomic RNA fragment. *Journal of Virology*, 65(12), 6845–6855. <https://doi.org/10.1128/jvi.65.12.6845-6855.1991>
- Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Odaka, T., & Oroszlan, S. (1985). Murine leukemia virus maturation: Protease region required for conversion from “immature” to “mature” core form and for virus infectivity. *Virology*, 145(2), 280–292. [https://doi.org/10.1016/0042-6822\(85\)90161-8](https://doi.org/10.1016/0042-6822(85)90161-8)

- Katz, R. A., Greger, J. G., Boimel, P., & Skalka, A. M. (2003). Human immunodeficiency virus type 1 DNA nuclear import and integration are mitosis independent in cycling cells. *Journal of Virology*, *77*(24), 13412–13417. <https://doi.org/10.1128/jvi.77.24.13412-13417.2003>
- 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. <https://doi.org/10.1128/jvi.72.7.5877-5885.1998>
- Keane, S. C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, 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. <https://doi.org/10.1126/science.aaa9266>
- Keane, S. C., & Summers, M. F. (2016). NMR studies of the structure and function of the HIV-1 5'-leader. *Viruses*, *8*(12), 338. <https://doi.org/10.3390/v8120338>
- Kemler, I., Meehan, A., & Poeschla, E. M. (2010). Live-cell coimaging of the genomic RNAs and Gag proteins of two lentiviruses. *Journal of Virology*, *84*(13), 6352–6366. <https://doi.org/10.1128/JVI.00363-10>
- Kenyon, J. C., Prestwood, L. J., & Lever, A. M. L. (2015). A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Scientific Reports*, *5*(1), 14369. <https://doi.org/10.1038/srep14369>
- 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. <https://doi.org/10.1093/nar/gkr252>
- 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. <https://doi.org/10.1016/j.jmb.2018.05.029>
- Kim, C.-H., & Tinoco, I. (2000). A retroviral RNA kissing complex containing only two G·C base pairs. *Proceedings of the National Academy of Sciences*, *97*(17), 9396–9401. <https://doi.org/10.1073/pnas.170283697>

- King, A. M. Q., Adams, M. J., Carstens, E. B., & Lefkowitz, E. J. (Eds.). (2012). Family—Retroviridae. In *Virus Taxonomy* (pp. 477–495). Elsevier. <https://doi.org/10.1016/B978-0-12-384684-6.00044-6>
- Kizhatil, K., & Albritton, L. M. (1997). Requirements for different components of the host cell cytoskeleton distinguish ecotropic murine leukemia virus entry via endocytosis from entry via surface fusion. *Journal of Virology*, *71*(10), 7145–7156. <https://doi.org/10.1128/JVI.71.10.7145-7156.1997>
- Kleiman, L., Jones, C., & Musier-Forsyth, K. (2010). Formation of the tRNA^{Lys} Packaging Complex in HIV-1. *FEBS Letters*, *584*(2), 359. <https://doi.org/10.1016/j.febslet.2009.11.038>
- Klein, D. J., Johnson, P. E., Zollars, E. S., De Guzman, R. N., & Summers, M. F. (2000). The NMR structure of the nucleocapsid protein from the mouse mammary tumor virus reveals unusual folding of the C-terminal zinc knuckle. *Biochemistry*, *39*(7), 1604–1612. <https://doi.org/10.1021/bi9922493>
- Klein, R., Ruttkowski, B., Schwab, S., Peterbauer, T., Salmons, B., Günzburg, W. H., & Hohenadl, C. (2008). Mouse mammary tumor virus promoter-containing retroviral promoter conversion vectors for gene-directed enzyme prodrug therapy are functional in vitro and in vivo. *Journal of Biomedicine and Biotechnology*, *2008*, 683505. <https://doi.org/10.1155/2008/683505>
- Klikova, M., Rhee, S. S., Hunter, E., & Ruml, T. (1995). Efficient in vivo and in vitro assembly of retroviral capsids from Gag precursor proteins expressed in bacteria. *Journal of Virology*, *69*(2), 1093–1098. <https://doi.org/10.1128/jvi.69.2.1093-1098.1995>
- Knipe, D. M., & Howley, P. (2013). *Fields Virology*. Lippincott Williams & Wilkins.
- Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., & Sigal, I. S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(13), 4686–4690. <https://doi.org/10.1073/pnas.85.13.4686>
- Konings, D. A., Nash, M. A., Maizel, J. V., & Arlinghaus, R. B. (1992). Novel GACG-hairpin pair motif in the 5' untranslated region of type C retroviruses related to murine leukemia virus. *Journal of Virology*, *66*(2), 632–640. <https://doi.org/10.1128/JVI.66.2.632-640.1992>

- Konings, F. A. J., Burda, S. T., Urbanski, M. M., Zhong, P., Nadas, A., & Nyambi, P. N. (2006). Human immunodeficiency virus type 1 (HIV-1) circulating recombinant form 02_AG (CRF02_AG) has a higher in vitro replicative capacity than its parental subtypes A and G. *Journal of Medical Virology*, 78(5), 523–534. <https://doi.org/10.1002/jmv.20572>
- Konstantoulas, C. J., & Indik, S. (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. <https://doi.org/10.1186/1742-4690-11-34>
- Konvalinka, J., Kräusslich, H.-G., & Müller, B. (2015). Retroviral proteases and their roles in virion maturation. *Virology*, 479–480, 403–417. <https://doi.org/10.1016/j.virol.2015.03.021>
- Köppe, B., Menéndez-Arias, L., & Oroszlan, S. (1994). Expression and purification of the mouse mammary tumor virus gag-pro transframe protein p30 and characterization of its dUTPase activity. *Journal of Virology*, 68(4), 2313–2319. <https://doi.org/10.1128/JVI.68.4.2313-2319.1994>
- Kozal, M. J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T. C., Richman, D. D., Morris, D., Hubbell, E., Chee, M., & Gingeras, T. R. (1996). Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nature Medicine*, 2(7), 753–759. <https://doi.org/10.1038/nm0796-753>
- Kräusslich, H. G., & Welker, R. (1996). Intracellular transport of retroviral capsid components. *Current Topics in Microbiology and Immunology*, 214, 25–63. https://doi.org/10.1007/978-3-642-80145-7_2
- 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) Pr50^{Gag}. *Viruses*, 11(8), 689. <https://doi.org/10.3390/v11080689>
- 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. <https://doi.org/10.1371/journal.ppat.1001200>
- 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. <https://doi.org/10.1016/j.cell.2014.09.057>

- Larson, D. R., Johnson, M. C., Webb, W. W., & Vogt, V. M. (2005). Visualization of retrovirus budding with correlated light and electron microscopy. *Proceedings of the National Academy of Sciences*, *102*(43), 15453–15458. <https://doi.org/10.1073/pnas.0504812102>
- 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. <https://doi.org/10.1128/jvi.71.5.3397-3406.1997>
- Lehmann-Che, J., & Saïb, A. (2004). Early stages of HIV replication: How to hijack cellular functions for a successful infection. *AIDS Reviews*, *6*(4), 199–207.
- Lesbats, P., Engelman, A. N., & Cherepanov, P. (2016). Retroviral DNA integration. *Chemical Reviews*, *116*(20), 12730–12757. <https://doi.org/10.1021/acs.chemrev.6b00125>
- Lescoute, A., & Westhof, E. (2006). Topology of three-way junctions in folded RNAs. *RNA*, *12*(1), 83–93. <https://doi.org/10.1261/rna.2208106>
- Lever, A. M. L. (2007). HIV-1 RNA packaging. *Advances in Pharmacology (San Diego, Calif.)*, *55*, 1–32. [https://doi.org/10.1016/S1054-3589\(07\)55001-5](https://doi.org/10.1016/S1054-3589(07)55001-5)
- Levy, N., Eiler, S., Pradeau-Aubretton, K., Crucifix, C., Schaetzel, A., Drillien, R., Parissi, V., Emiliani, S., Mely, Y., Schultz, P., & Ruff, M. (2013). Structural and functional studies of the HIV-1 pre-integration complex. *Retrovirology*, *10*(Suppl 1), 76. <https://doi.org/10.1186/1742-4690-10-S1-P76>
- Lewinski, M. K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., Berry, C. C., Ecker, J. R., & Bushman, F. D. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathogens*, *2*(6), e60. <https://doi.org/10.1371/journal.ppat.0020060>
- 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), e2019467118. <https://doi.org/10.1073/pnas.2019467118>
- Li, S., Hill, C. P., Sundquist, W. I., & Finch, J. T. (2000). Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature*, *407*(6802), 409–413. <https://doi.org/10.1038/35030177>

- Liu, S., Kaddis Maldonado, R., Rye-McCurdy, T., Binkley, C., Bah, A., Chen, E. C., Rice, B. L., Parent, L. J., & Musier-Forsyth, K. (2020). Rous sarcoma virus genomic RNA dimerization capability in vitro is not a prerequisite for viral infectivity. *Viruses*, *12*(5), 568. <https://doi.org/10.3390/v12050568>
- Liu, Y., Nikolaitchik, O. A., Rahman, S. A., Chen, J., Pathak, V. K., & Hu, W.-S. (2017). HIV-1 sequence necessary and sufficient to package non-viral RNAs into HIV-1 particles. *Journal of Molecular Biology*, *429*(16), 2542–2555. <https://doi.org/10.1016/j.jmb.2017.06.018>
- 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., Tummillo, D., Hosic, 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. <https://doi.org/10.1126/science.1210460>
- Ly, H., & Parslow, T. G. (2002). Bipartite signal for genomic RNA dimerization in moloney murine leukemia virus. *Journal of Virology*, *76*(7), 3135–3144. <https://doi.org/10.1128/JVI.76.7.3135-3144.2002>
- MacLachlan, N. J., & Dubovi, E. J. (Eds.). (2017). Chapter 14—Retroviridae. In *Fenner's Veterinary Virology (Fifth Edition)* (pp. 269–297). Academic Press. <https://doi.org/10.1016/B978-0-12-800946-8.00014-3>
- 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), 248. <https://doi.org/10.3390/v8090248>
- Malagon, F., & Jensen, T. H. (2011). T-body formation precedes virus-like particle maturation in *S. cerevisiae*. *RNA Biology*, *8*(2), 184–189. <https://doi.org/10.4161/rna.8.2.14822>
- Maldonado, R. J. K., Rice, B., Chen, E. C., Tuffey, K. M., Chiari, E. F., Fahrback, 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), e00524-20. <https://doi.org/10.1128/mBio.00524-20>
- Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V., & Cullen, B. R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature*, *338*(6212), 254–257. <https://doi.org/10.1038/338254a0>

- Manel, N., Kim, F. J., Kinet, S., Taylor, N., Sitbon, M., & Battini, J.-L. (2003). The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV. *Cell*, *115*(4), 449–459. [https://doi.org/10.1016/s0092-8674\(03\)00881-x](https://doi.org/10.1016/s0092-8674(03)00881-x)
- Mangel, W. F., Delius, H., & Duesberg, P. H. (1974). Structure and molecular weight of the 60-70S RNA and the 30-40S RNA of the Rous sarcoma virus. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(11), 4541–4545. <https://doi.org/10.1073/pnas.71.11.4541>
- Mann, R., Mulligan, R. C., & Baltimore, D. (1983). Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell*, *33*(1), 153–159. [https://doi.org/10.1016/0092-8674\(83\)90344-6](https://doi.org/10.1016/0092-8674(83)90344-6)
- Marquet, R., Baudin, F., Gabus, C., Darlix, J. L., Mougel, M., Ehresmann, C., & Ehresmann, B. (1991). Dimerization of human immunodeficiency virus (type 1) RNA: Stimulation by cations and possible mechanism. *Nucleic Acids Research*, *19*(9), 2349–2357. <https://doi.org/10.1093/nar/19.9.2349>
- 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. [https://doi.org/10.1016/0378-1119\(84\)90022-2](https://doi.org/10.1016/0378-1119(84)90022-2)
- Martin, E. E., Huang, W., Anwar, T., Arellano-Garcia, C., Burman, B., Guan, J.-L., Gonzalez, M. E., & Kleer, C. G. (2017). MMTV-cre; Ccn6 knockout mice develop tumors recapitulating human metaplastic breast carcinomas. *Oncogene*, *36*(16), 2275–2285. <https://doi.org/10.1038/onc.2016.381>
- Maskell, D. P., Renault, L., Serrao, E., Lesbats, P., Matadeen, R., Hare, S., Lindemann, D., Engelman, A. N., Costa, A., & Cherepanov, P. (2015). Structural basis for retroviral integration into nucleosomes. *Nature*, *523*(7560), 366–369. <https://doi.org/10.1038/nature14495>
- McBride, M. S., Schwartz, M. D., & Panganiban, A. T. (1997). Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *Journal of Virology*, *71*(6), 4544–4554. <https://doi.org/10.1128/JVI.71.6.4544-4554.1997>
- McClure, M. O., Marsh, M., & Weiss, R. A. (1988). Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism. *The EMBO Journal*, *7*(2), 513–518. <https://doi.org/10.1002/j.1460-2075.1988.tb02839.x>

- McDonald, D., Vodicka, M. A., Lucero, G., Svitkina, T. M., Borisy, G. G., Emerman, M., & Hope, T. J. (2002). Visualization of the intracellular behavior of HIV in living cells. *Journal of Cell Biology*, *159*(3), 441–452. <https://doi.org/10.1083/jcb.200203150>
- McDougal, J. S., Kennedy, M. S., Sleigh, J. M., Cort, S. P., Mawle, A., & Nicholson, J. K. (1986). Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science (New York, N.Y.)*, *231*(4736), 382–385. <https://doi.org/10.1126/science.3001934>
- McKinstry, W. J., Hijnen, M., Tanwar, H. S., Sparrow, L. G., Nagarajan, S., Pham, S. T., & Mak, J. (2014). Expression and purification of soluble recombinant full length HIV-1 Pr55Gag protein in Escherichia coli. *Protein Expression and Purification*, *100*(Supplement C), 10–18. <https://doi.org/10.1016/j.pep.2014.04.013>
- McNicholl, J. M., Smith, D. K., Qari, S. H., & Hodge, T. (1997). Host genes and HIV: The role of the chemokine receptor gene CCR5 and its allele. *Emerging Infectious Diseases*, *3*(3), 261–271. <https://doi.org/10.3201/eid0303.970302>
- Medina, D. (2010). Of mice and women: a short history of mouse mammary cancer research with an emphasis on the paradigms inspired by the transplantation method. *Cold Spring Harbor Perspectives in Biology*, *2*(10), a004523. <https://doi.org/10.1101/cshperspect.a004523>
- Méric, C., Gouilloud, E., & Spahr, P. F. (1988). Mutations in Rous sarcoma virus nucleocapsid protein p12 (NC): Deletions of Cys-His boxes. *Journal of Virology*, *62*(9), 3328–3333. <https://doi.org/10.1128/JVI.62.9.3328-3333.1988>
- Mertz, J. A., Chadee, A. B., Byun, H., Russell, R., & Dudley, J. P. (2009). Mapping of the functional boundaries and secondary structure of the mouse mammary tumor virus Rem-responsive element. *Journal of Biological Chemistry*, *284*(38), 25642–25652. <https://doi.org/10.1074/jbc.M109.012476>
- Mertz, J. A., Lozano, M. M., & Dudley, J. P. (2009). Rev and Rex proteins of human complex retroviruses function with the MMTV Rem-responsive element. *Retrovirology*, *6*, 10. <https://doi.org/10.1186/1742-4690-6-10>
- 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. <https://doi.org/10.1128/JVI.79.23.14737-14747.2005>

- Mitchell, R. S., Beitzel, B. F., Schroder, A. R. W., Shinn, P., Chen, H., Berry, C. C., Ecker, J. R., & Bushman, F. D. (2004). retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biology*, 2(8), e234. <https://doi.org/10.1371/journal.pbio.0020234>
- Miyazaki, Y., Garcia, E. L., King, S. R., Iyalla, K., Loeliger, K., Starck, P., Syed, S., Telesnitsky, A., & Summers, M. F. (2010). An RNA structural switch regulates diploid genome packaging by moloney murine leukemia virus. *Journal of Molecular Biology*, 396(1), 141–152. <https://doi.org/10.1016/j.jmb.2009.11.033>
- Miyazaki, Y., Irobalieva, R. N., Tolbert, B., Smalls-Mantey, A., Iyalla, K., Loeliger, K., D'Souza, V., Khant, H., Schmid, M. F., Garcia, E., Telesnitsky, A., Chiu, W., & Summers, M. F. (2010). Structure of a conserved retroviral RNA packaging element by NMR spectroscopy and cryo-electron tomography. *Journal of Molecular Biology*, 404(5), 751–772. <https://doi.org/10.1016/j.jmb.2010.09.009>
- Miyazaki, Y., Miyake, A., Nomaguchi, M., & Adachi, A. (2011). Structural dynamics of retroviral genome and the packaging. *Frontiers in Microbiology*, 2, 264. <https://doi.org/10.3389/fmicb.2011.00264>
- Monie, T., Greaux, J., & Lever, A. M. (2001). Oligonucleotide mapping of the core genomic RNA dimer linkage in human T-cell leukaemia virus type-1. *Virus Research*, 78(1–2), 45–56. [https://doi.org/10.1016/s0168-1702\(01\)00283-0](https://doi.org/10.1016/s0168-1702(01)00283-0)
- 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. <https://doi.org/10.1371/journal.ppat.1000627>
- Mothes, W., Boerger, A. L., Narayan, S., Cunningham, J. M., & Young, J. A. (2000). Retroviral entry mediated by receptor priming and low pH triggering of an envelope glycoprotein. *Cell*, 103(4), 679–689. [https://doi.org/10.1016/s0092-8674\(00\)00170-7](https://doi.org/10.1016/s0092-8674(00)00170-7)
- Mougel, M., & Barklis, E. (1997). A role for two hairpin structures as a core RNA encapsidation signal in murine leukemia virus virions. *Journal of Virology*, 71(10), 8061–8065. <https://doi.org/10.1128/JVI.71.10.8061-8065.1997>

- Mouland, A. J., Mercier, J., Luo, M., Bernier, L., DesGroseillers, L., & Cohen, É. A. (2000). The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation. *Journal of Virology*, 74(12), 5441–5451. <https://doi.org/10.1128/JVI.74.12.5441-5451.2000>
- Mueller, N., Das, A. T., & Berkhout, B. (2016). A phylogenetic survey on the structure of the HIV-1 leader RNA domain that encodes the splice donor signal. *Viruses*, 8(7), 200. <https://doi.org/10.3390/v8070200>
- Müllner, M., Salmons, B., Günzburg, W. H., & Indik, S. (2008). Identification of the Rem-responsive element of mouse mammary tumor virus. *Nucleic Acids Research*, 36(19), 6284–6294. <https://doi.org/10.1093/nar/gkn608>
- Muriaux, D., Girard, P. M., Bonnet-Mathonière, B., & Paoletti, J. (1995). Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *The Journal of Biological Chemistry*, 270(14), 8209–8216. <https://doi.org/10.1074/jbc.270.14.8209>
- Muriaux, D., Mirro, J., Harvin, D., & Rein, A. (2001). RNA is a structural element in retrovirus particles. *Proceedings of the National Academy of Sciences of the United States of America*, 98(9), 5246–5251. <https://doi.org/10.1073/pnas.091000398>
- Murti, K. G., Bondurant, M., & Tereba, A. (1981). Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *Journal of Virology*, 37(1), 411–419. <https://doi.org/10.1128/jvi.37.1.411-419.1981>
- 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. <https://doi.org/10.1371/journal.pone.0047088>
- Mustafa, F., Bhadra, S., Johnston, D., Lozano, M., & Dudley, J. P. (2003). The type B leukemogenic virus truncated superantigen is dispensable for T-cell lymphomagenesis. *Journal of Virology*, 77(6), 3866–3870. <https://doi.org/10.1128/JVI.77.6.3866-3870.2003>

- Mustafa, F., Ghazawi, A., Jayanth, P., Phillip, P. S., Ali, J., & Rizvi, T. A. (2005). Sequences intervening between the core packaging determinants are dispensable for maintaining the packaging potential and propagation of feline immunodeficiency virus transfer vector RNAs. *Journal of Virology*, *79*(21), 13817–13821. <https://doi.org/10.1128/JVI.79.21.13817-13821.2005>
- 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. <https://doi.org/10.1080/15476286.2018.1486661>
- Nabel, G., & Baltimore, D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature*, *326*(6114), 711–713. <https://doi.org/10.1038/326711a0>
- Nakano, K., & Watanabe, T. (2012). HTLV-1 Rex: The courier of viral messages making use of the host vehicle. *Frontiers in Microbiology*, *3*, 330. <https://doi.org/10.3389/fmicb.2012.00330>
- 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. <https://doi.org/10.1126/science.272.5259.263>
- Naughtin, M., Haftek-Terreau, Z., Xavier, J., Meyer, S., Silvain, M., Jaszczyszyn, Y., Levy, N., Miele, V., Benleulmi, M. S., Ruff, M., Parissi, V., Vaillant, C., & Lavigne, M. (2015). DNA physical properties and nucleosome positions are major determinants of HIV-1 integrase selectivity. *PloS One*, *10*(6), e0129427. <https://doi.org/10.1371/journal.pone.0129427>
- Neville, M., Stutz, F., Lee, L., Davis, L. I., & Rosbash, M. (1997). The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Current Biology: CB*, *7*(10), 767–775. [https://doi.org/10.1016/s0960-9822\(06\)00335-6](https://doi.org/10.1016/s0960-9822(06)00335-6)
- Ni, N., Nikolaitchik, O. A., Dilley, K. A., Chen, J., Galli, A., Fu, W., Prasad, V. V. S. P., Ptak, R. G., Pathak, V. K., & Hu, W.-S. (2011). Mechanisms of human immunodeficiency virus type 2 RNA packaging: efficient trans packaging and selection of RNA copackaging partners. *Journal of Virology*, *85*(15), 7603–7612. <https://doi.org/10.1128/JVI.00562-11>

- Nikolaitchik, O. A., Dilley, K. A., Fu, W., Gorelick, R. J., Tai, S.-H. S., Soheilian, F., Ptak, R. G., Nagashima, K., Pathak, V. K., & Hu, W.-S. (2013). Dimeric RNA recognition regulates HIV-1 genome packaging. *PLoS Pathogens*, *9*(3), e1003249. <https://doi.org/10.1371/journal.ppat.1003249>
- Nikolaitchik, O. A., Somoulay, X., Rawson, J. M. O., Yoo, J. A., Pathak, V. K., & Hu, W.-S. (2020). Unpaired guanosines in the 5' untranslated region of HIV-1 RNA act synergistically to mediate genome packaging. *Journal of Virology*, *94*(21), e00439-20. <https://doi.org/10.1128/JVI.00439-20>
- Njai, H. F., Gali, Y., Vanham, G., Clybergh, C., Jennes, W., Vidal, N., Butel, C., Mpoudi-Ngolle, E., Peeters, M., & Ariën, K. K. (2006). The predominance of Human Immunodeficiency Virus type 1 (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its replicative fitness. *Retrovirology*, *3*, 40. <https://doi.org/10.1186/1742-4690-3-40>
- Pitchai, F.N.N., Chameettachal, A., Vivet-Boudou, V., Mohamed Ali, L., Pillai, V. N., Krishnan, A., Bernacchi, S., Mustafa, F., Marquet, R., & Rizvi, T. A. (2021). Identification of Pr78^{Gag} binding sites on the Mason-Pfizer monkey virus genomic RNA packaging determinants. *Journal of Molecular Biology*, *433*(10), 166923. <https://doi.org/10.1016/j.jmb.2021.166923>
- Nymark-McMahon, M. H., Beliakova-Bethell, N. S., Darlix, J.-L., Le Grice, S. F. J., & Sandmeyer, S. B. (2002). Ty3 integrase is required for initiation of reverse transcription. *Journal of Virology*, *76*(6), 2804–2816. <https://doi.org/10.1128/JVI.76.6.2804-2816.2002>
- Onafuwa-Nuga, A., & Telesnitsky, A. (2009). The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiology and Molecular Biology Review: MMBR*, *73*(3), 451–480. <https://doi.org/10.1128/MMBR.00012-09>
- Ono, A., Orenstein, J. M., & Freed, E. O. (2000). Role of the Gag matrix domain in targeting human immunodeficiency virus type 1 assembly. *Journal of Virology*, *74*(6), 2855–2866. <https://doi.org/10.1128/jvi.74.6.2855-2866.2000>
- Ono, Akira, 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. <https://doi.org/10.1073/pnas.0405596101>

- Ott, D. E., Coren, L. V., & Shatzer, T. (2009). The nucleocapsid region of human immunodeficiency virus type 1 Gag assists in the coordination of assembly and Gag processing: role for RNA-Gag binding in the early stages of assembly. *Journal of Virology*, *83*(15), 7718–7727. <https://doi.org/10.1128/JVI.00099-09>
- Pachulska-Wieczorek, K., Le Grice, S. F. J., & Purzycka, K. J. (2016). Determinants of Genomic RNA Encapsidation in the *Saccharomyces cerevisiae* Long Terminal Repeat Retrotransposons Ty1 and Ty3. *Viruses*, *8*(7), 193. <https://doi.org/10.3390/v8070193>
- Paillart, J. C., Berthou, L., Ottmann, M., Darlix, J. L., Marquet, R., Ehresmann, B., & Ehresmann, C. (1996). A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *Journal of Virology*, *70*(12), 8348–8354. <https://doi.org/10.1128/JVI.70.12.8348-8354.1996>
- Paillart, J. C., Marquet, R., Skripkin, E., Ehresmann, B., & Ehresmann, C. (1994). Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *The Journal of Biological Chemistry*, *269*(44), 27486–27493. [https://doi.org/10.1016/S0021-9258\(18\)47011-1](https://doi.org/10.1016/S0021-9258(18)47011-1)
- 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. <https://doi.org/10.1073/pnas.93.11.5572>
- 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. <https://doi.org/10.1074/jbc.M408294200>
- Paillart, J.C., Shehu-Xhilaga, M., Marquet, R., & Mak, J. (2004). Dimerization of retroviral RNA genomes: An inseparable pair. *Nature Reviews. Microbiology*, *2*(6), 461–472. <https://doi.org/10.1038/nrmicro903>
- Parent, L. J., Cairns, T. M., Albert, J. A., Wilson, C. B., Wills, J. W., & Craven, R. C. (2000). RNA dimerization defect in a Rous sarcoma virus matrix mutant. *Journal of Virology*, *74*(1), 164–172. <https://doi.org/10.1128/jvi.74.1.164-172.2000>

- Pereira-Montecinos, C., Toro-Ascuy, D., Rojas-Fuentes, C., Riquelme-Barrios, S., Rojas-Araya, B., García-de-Gracia, F., Aguilera-Cortés, P., Ananías-Sáez, C., de Bisschop, G., Chaniderman, J., L. Acevedo, M., Sargueil, B., Valiente-Echeverría, F., & Soto-Rifo, R. (2019, June 19). An epitranscriptomic switch at the 5'-UTR controls genome selection during HIV-1 genomic RNA packaging | *bioRxiv*. <https://www.biorxiv.org/content/10.1101/676031v1>
- 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. <https://doi.org/10.1038/s41598-018-30142-0>
- Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D., & Gallo, R. C. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*, 77(12), 7415–7419. <https://doi.org/10.1073/pnas.77.12.7415>
- Poole, E., Strappe, P., Mok, H.-P., Hicks, R., & Lever, A. M. L. (2005). HIV-1 Gag-RNA interaction occurs at a perinuclear/centrosomal site; analysis by confocal microscopy and FRET. *Traffic (Copenhagen, Denmark)*, 6(9), 741–755. <https://doi.org/10.1111/j.1600-0854.2005.00312.x>
- Poon, D. T. K., Li, G., & Aldovini, A. (1998). Nucleocapsid and matrix protein contributions to selective human immunodeficiency virus type 1 genomic RNA packaging. *Journal of Virology*, 72(3), 1983–1993. <https://doi.org/10.1128/JVI.72.3.1983-1993.1998>
- Poon, D. T., Wu, J., & Aldovini, A. (1996). Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *Journal of Virology*, 70(10), 6607–6616. <https://doi.org/10.1128/JVI.70.10.6607-6616.1996>
- Purdy, A., Case, L., Duvall, M., Overstrom-Coleman, M., Monnier, N., Chervonsky, A., & Golovkina, T. (2003). Unique resistance of I/LnJ mice to a retrovirus is due to sustained interferon gamma-dependent production of virus-neutralizing antibodies. *The Journal of Experimental Medicine*, 197(2), 233–243. <https://doi.org/10.1084/jem.20021499>

- Reed, J. C., Westergreen, N., Barajas, B. C., Ressler, D. T. B., Phuong, D. J., Swain, J. V., Lingappa, V. R., & Lingappa, J. R. (2018). Formation of RNA granule-derived capsid assembly intermediates appears to be conserved between human immunodeficiency virus type 1 and the nonprimate lentivirus feline immunodeficiency virus. *Journal of Virology*, *92*(9). e01761-17. <https://doi.org/10.1128/JVI.01761-17>
- Rein, A. (2004). Take two. *Nature Structural & Molecular Biology*, *11*(11), 1034–1035. <https://doi.org/10.1038/nsmb1104-1034>
- Rein, A. (2010). Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biology*, *7*(6), 700–705. <https://doi.org/10.4161/rna.7.6.13685>
- Rein, A. (2019). RNA packaging in HIV. *Trends in Microbiology*, *27*(8), 715–723. <https://doi.org/10.1016/j.tim.2019.04.003>
- Rein, A. (2020). The heart of the HIV RNA packaging signal? *Proceedings of the National Academy of Sciences of the United States of America*, *117*(33), 19621–19623. <https://doi.org/10.1073/pnas.2013378117>
- Reuss, F. U., & Coffin, J. M. (1995). Stimulation of mouse mammary tumor virus superantigen expression by an intragenic enhancer. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(20), 9293–9297. <https://doi.org/10.1073/pnas.92.20.9293>
- Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, *11*, 129. <https://doi.org/10.1186/1471-2105-11-129>
- Rhee, & Hunter. (1990). A single amino acid substitution within the matrix protein of a type D retrovirus converts its morphogenesis to that of a type C retrovirus. *Cell*, *63*(1), 77–86. [https://doi.org/10.1016/0092-8674\(90\)90289-Q](https://doi.org/10.1016/0092-8674(90)90289-Q)
- Rhee, & Hunter. (1991). Amino acid substitutions within the matrix protein of type D retroviruses affect assembly, transport, and membrane association of a capsid. *The EMBO Journal*, *10*(3), 535–546. <https://doi.org/10.1002/j.1460-2075.1991.tb07980.x>
- Rizvi, T. A., Ali, J., Phillip, P. S., Ghazawi, A., Jayanth, P., & Mustafa, F. (2009). Role of a heterologous retroviral transport element in the development of genetic complementation assay for mouse mammary tumor virus (MMTV) replication. *Virology*, *385*(2), 464–472. <https://doi.org/10.1016/j.virol.2008.12.027>

- Roe, T., Reynolds, T. C., Yu, G., & Brown, P. O. (1993). Integration of murine leukemia virus DNA depends on mitosis. *The EMBO Journal*, *12*(5), 2099–2108.
- Ross, S. R. (2008). MMTV infectious cycle and the contribution of virus-encoded proteins to transformation of mammary tissue. *Journal of Mammary Gland Biology and Neoplasia*, *13*(3), 299. <https://doi.org/10.1002/j.1460-2075.1993.tb05858.x>
- Ross, S. R. (2010). Mouse mammary tumor virus molecular biology and oncogenesis. *Viruses*, *2*(9), 2000–2012. <https://doi.org/10.3390/v2092000>
- Ross, S. R., Schmidt, J. W., Katz, E., Cappelli, L., Hultine, S., Gimotty, P., & Monroe, J. G. (2006). An immunoreceptor tyrosine activation motif in the mouse mammary tumor virus envelope protein plays a role in virus-induced mammary tumors. *Journal of Virology*, *80*(18), 9000–9008. <https://doi.org/10.1128/JVI.00788-06>
- 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. <https://doi.org/10.1073/pnas.192360099>
- Rouault, F., Nejad Asl, S. B., Rungaldier, S., Fuchs, E., Salmons, B., & Günzburg, W. H. (2007). Promoter complex in the central part of the mouse mammary tumor virus long terminal repeat. *Journal of Virology*, *81*(22), 12572–12581. <https://doi.org/10.1128/JVI.00351-07>
- Rous, P. (1911). A sarcoma of the fowl transmissible by an agent separable from the tumor cells. *The Journal of Experimental Medicine*, *13*(4), 397–411. <https://doi.org/10.1084/jem.13.4.397>
- Roy, B. B., Russell, R. S., Turner, D., & Liang, C. (2006). The T12I mutation within the SP1 region of Gag restricts packaging of spliced viral RNA into human immunodeficiency virus type 1 with mutated RNA packaging signals and mutated nucleocapsid sequence. *Virology*, *344*(2), 304–314. <https://doi.org/10.1016/j.virol.2005.09.011>
- 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. <https://doi.org/10.1128/JVI.02833-06>

- Russell, R. S., Liang, C., & Wainberg, M. A. (2004). Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology*, *1*, 23. <https://doi.org/10.1186/1742-4690-1-23>
- Russell, R. S., Roldan, A., Detorio, M., Hu, J., Wainberg, M. A., & Liang, C. (2003). Effects of a single amino acid substitution within the p2 region of human immunodeficiency virus type 1 on packaging of spliced viral RNA. *Journal of Virology*, *77*(24), 12986–12995. <https://doi.org/10.1128/JVI.77.24.12986-12995.2003>
- Saad, J. S., Miller, J., Tai, J., Kim, A., Ghanam, R. H., & Summers, M. F. (2006). Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(30), 11364–11369. <https://doi.org/10.1073/pnas.0602818103>
- Sakalian, M., Parker, S. D., Weldon, R. A., & Hunter, E. (1996). Synthesis and assembly of retrovirus Gag precursors into immature capsids in vitro. *Journal of Virology*, *70*(6), 3706–3715. <https://doi.org/10.1128/JVI.70.6.3706-3715.1996>
- Sakalian, M., & Hunter, E. (1999). Separate assembly and transport domains within the Gag precursor of Mason-Pfizer monkey virus. *Journal of Virology*, *73*(10), 8073–8082. <https://doi.org/10.1128/JVI.73.10.8073-8082.1999>
- Sakuragi, J., Shioda, T., & Panganiban, A. T. (2001). Duplication of the primary encapsidation and dimer linkage region of human immunodeficiency virus type 1 RNA results in the appearance of monomeric RNA in virions. *Journal of Virology*, *75*(6), 2557–2565. <https://doi.org/10.1128/JVI.75.6.2557-2565.2001>
- Sakuragi, J., Ueda, S., Iwamoto, A., & Shioda, T. (2003). Possible role of dimerization in human immunodeficiency virus type 1 genome RNA packaging. *Journal of Virology*, *77*(7), 4060–4069. <https://doi.org/10.1128/JVI.77.7.4060-4069.2003>
- Salis, H. M. (2011). The ribosome binding site calculator. *Methods in Enzymology*, *498*, 19–42. <https://doi.org/10.1016/B978-0-12-385120-8.00002-4>
- Salmons, B., Moritz-Legrand, S., Garcha, I., & Günzburg, W. H. (1989). Construction and characterization of a packaging cell line for MMTV-based conditional retroviral vectors. *Biochemical and Biophysical Research Communications*, *159*(3), 1191–1198. [https://doi.org/10.1016/0006-291X\(89\)92236-5](https://doi.org/10.1016/0006-291X(89)92236-5)

- Schaller, T., Ocwieja, K. E., Rasaiyaah, J., Price, A. J., Brady, T. L., Roth, S. L., Hué, S., Fletcher, A. J., Lee, K., KewalRamani, V. N., Noursadeghi, M., Jenner, R. G., James, L. C., Bushman, F. D., & Towers, G. J. (2011). HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathogens*, 7(12), e1002439. <https://doi.org/10.1371/journal.ppat.1002439>
- Seif, E., Niu, M., & Kleiman, L. (2013). Annealing to sequences within the primer binding site loop promotes an HIV-1 RNA conformation favoring RNA dimerization and packaging. *RNA*, 19(10), 1384–1393. <https://doi.org/10.1261/rna.038497.113>
- Selyutina, A., Persaud, M., Lee, K., KewalRamani, V., & Diaz-Griffero, F. (2020). Nuclear import of the HIV-1 core precedes reverse transcription and uncoating. *Cell Reports*, 32(13), 108201. <https://doi.org/10.1016/j.celrep.2020.108201>
- Shackleford, G. M., & Varmus, H. E. (1988). Construction of a clonable, infectious, and tumorigenic mouse mammary tumor virus provirus and a derivative genetic vector. *Proceedings of the National Academy of Sciences*, 85(24), 9655–9659. <https://doi.org/10.1073/pnas.85.24.9655>
- Shehu-Xhilaga, M., Crowe, S. M., & Mak, J. (2001). Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *Journal of Virology*, 75(4), 1834–1841. <https://doi.org/10.1128/JVI.75.4.1834-1841.2001>
- Shih, C.-C., Stoye, J. P., & Coffin, J. M. (1988). Highly preferred targets for retrovirus integration. *Cell*, 53(4), 531–537. [https://doi.org/10.1016/0092-8674\(88\)90569-7](https://doi.org/10.1016/0092-8674(88)90569-7)
- Shine, J., & Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America*, 71(4), 1342–1346. <https://doi.org/10.1073/pnas.71.4.1342>
- Sinangil, F., Loyter, A., & Volsky, D. J. (1988). Quantitative measurement of fusion between human immunodeficiency virus and cultured cells using membrane fluorescence dequenching. *FEBS Letters*, 239(1), 88–92. [https://doi.org/10.1016/0014-5793\(88\)80551-9](https://doi.org/10.1016/0014-5793(88)80551-9)
- Sinck, L., Richer, D., Howard, J., Alexander, M., Purcell, D. F. J., Marquet, R., & Paillart, J.C. (2007). In vitro dimerization of human immunodeficiency virus type 1 (HIV-1) spliced RNAs. *RNA (New York, N.Y.)*, 13(12), 2141–2150. <https://doi.org/10.1261/rna.678307>

- Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B., & Ehresmann, C. (1994). Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(11), 4945–4949. <https://doi.org/10.1073/pnas.91.11.4945>
- Smith, G. H. (1978). Evidence for a precursor-product relationship between intracytoplasmic A particles and mouse mammary tumour virus cores. *The Journal of General Virology*, *41*(1), 193–200. <https://doi.org/10.1099/0022-1317-41-1-193>
- Smith, Gilbert H. (2005). Stem cells and mammary cancer in mice. *Stem Cell Reviews*, *1*(3), 215–223. <https://doi.org/10.1385/SCR:1:3:215>
- 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. <https://doi.org/10.1038/nmeth.3490>
- Smyth, R. P., Smith, M. R., Jousset, A.-C., Despons, L., Laumond, G., Decoville, T., Cattenoz, P., Moog, C., Jossinet, F., Mougel, 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*, e57. <https://doi.org/10.1093/nar/gky152>
- Solinska, J. S. (2014). Therapeutic targeting of structural RNA motifs in viral RNA genomes. *Int. J. Virol. AIDS*, *1*.
- Stein, B. S., Gowda, S. D., Lifson, J. D., Penhallow, R. C., Bensch, K. G., & Engleman, E. G. (1987). PH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell*, *49*(5), 659–668. [https://doi.org/10.1016/0092-8674\(87\)90542-3](https://doi.org/10.1016/0092-8674(87)90542-3)
- Stevens, S. W., & Griffith, J. D. (1994). Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by L1Hs repetitive elements. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(12), 5557–5561. <https://doi.org/10.1073/pnas.91.12.5557>

- Suzuki, T., Fujisawa, J. I., Toita, M., & Yoshida, M. (1993). The trans-activator tax of human T-cell leukemia virus type 1 (HTLV-1) interacts with cAMP-responsive element (CRE) binding and CRE modulator proteins that bind to the 21-base-pair enhancer of HTLV-1. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(2), 610–614. <https://doi.org/10.1073/pnas.90.2.610>
- Suzuki, Y., & Craigie, R. (2007). The road to chromatin—Nuclear entry of retroviruses. *Nature Reviews. Microbiology*, *5*(3), 187–196. <https://doi.org/10.1038/nrmicro1579>
- Tanaka, H., Tamura, A., & Tsujimura, D. (1972). Properties of the intracytoplasmic A particles purified from mouse tumors. *Virology*, *49*(1), 61–78. [https://doi.org/10.1016/S0042-6822\(72\)80007-2](https://doi.org/10.1016/S0042-6822(72)80007-2)
- Tanwar, H. S., Khoo, K. K., Garvey, M., Waddington, L., Leis, A., Hijnen, M., Velkov, T., Dumsday, G. J., McKinstry, W. J., & Mak, J. (2017). The thermodynamics of Pr55Gag-RNA interaction regulate the assembly of HIV. *PLOS Pathogens*, *13*(2), e1006221. <https://doi.org/10.1371/journal.ppat.1006221>
- Taube, R., Loya, S., Avidan, O., Perach, M., & Hizi, A. (1998). Reverse transcriptase of mouse mammary tumour virus: Expression in bacteria, purification and biochemical characterization. *The Biochemical Journal*, *329* (Pt 3), 579–587. <https://doi.org/10.1042/bj3290579>
- Tounekti, N., Mougél, M., Roy, C., Marquet, R., Darlix, J.-L., Paoletti, J., Ehresmann, B., & Ehresmann, C. (1992). Effect of dimerization on the conformation of the encapsidation Psi domain of Moloney murine leukemia virus RNA. *Journal of Molecular Biology*, *223*(1), 205–220. [https://doi.org/10.1016/0022-2836\(92\)90726-Z](https://doi.org/10.1016/0022-2836(92)90726-Z)
- Toyoshima, K., Niwa, O., Yutsudo, M., Sugiyama, H., Tahara, S., & Sugahara, T. (1980). Sensitivity to gamma rays of avian sarcoma and murine leukemia viruses. *Virology*, *105*(2), 508–515. [https://doi.org/10.1016/0042-6822\(80\)90051-3](https://doi.org/10.1016/0042-6822(80)90051-3)
- Tuffy, K. M., Maldonado, R. J. K., Chang, J., Rosenfeld, P., Cochrane, A., & Parent, L. J. (2020). HIV-1 Gag forms ribonucleoprotein complexes with unspliced viral RNA at transcription sites. *Viruses*, *12*(11), 1281. <https://doi.org/10.3390/v12111281>

- Umunnakwe, C. N., Duchon, A., Nikolaitchik, O. A., Rahman, S. A., Liu, Y., Chen, J., Tai, S., Pathak, V. K., & Hu, W.-S. (2021). Specific guanosines in the HIV-2 leader RNA are essential for efficient viral genome packaging. *Journal of Molecular Biology*, *433*(2), 166718. <https://doi.org/10.1016/j.jmb.2020.11.017>
- Varmus, H. E., Bishop, J. M., Nowinski, R. C., & Sarker, N. H. (1972). Mammary tumour virus specific nucleotide sequences in mouse DNA. *Nature New Biology*, *238*(84), 189. <https://doi.org/10.1038/newbio238189a0>
- Voisset, C., & Andrawiss, M. (2000). Retroviruses at a glance. *Genome Biology*, *1*(3), reports4015.1-reports4015.4.
- Wakefield, J. K., Kang, S. M., & Morrow, C. D. (1996). Construction of a type 1 human immunodeficiency virus that maintains a primer binding site complementary to tRNA(His). *Journal of Virology*, *70*(2), 966–975. <https://doi.org/10.1128/JVI.70.2.966-975.1996>
- Wang, C.-T., & Barklis, E. (1993). Assembly, processing, and infectivity of human immunodeficiency virus type 1 gag mutants. *Journal of Virology*, *67*(7), 4264–4273. <https://doi.org/10.1128/jvi.67.7.4264-4273.1993>
- Wang, H., Norris, K. M., & Mansky, L. M. (2003). Involvement of the matrix and nucleocapsid domains of the bovine leukemia virus Gag polyprotein precursor in viral RNA packaging. *Journal of Virology*, *77*(17), 9431–9438. <https://doi.org/10.1128/jvi.77.17.9431-9438.2003>
- Webb, J. A., Jones, C. P., Parent, L. J., Rouzina, I., & Musier-Forsyth, K. (2013). Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: Implications for viral genomic RNA packaging. *RNA*, *19*(8), 1078–1088. <https://doi.org/10.1261/rna.038869.113>
- Weichs an der Glon, C., Monks, J., & Proudfoot, N. J. (1991). Occlusion of the HIV poly(A) site. *Genes & Development*, *5*(2), 244–253. <https://doi.org/10.1101/gad.5.2.244>
- Weiss, R. A. (1993). Cellular receptors and viral glycoproteins involved in retrovirus entry. In J. A. Levy (Ed.), *The Retroviridae* (pp. 1–108). Springer US. https://doi.org/10.1007/978-1-4899-1627-3_1
- White, J. M., Delos, S. E., Brecher, M., & Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins. *Critical Reviews in Biochemistry and Molecular Biology*, *43*(3), 189–219. <https://doi.org/10.1080/10409230802058320>

- 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. <https://doi.org/10.1371/journal.pbio.0060096>
- Wlodawer, A., & Gustchina, A. (2000). Structural and biochemical studies of retroviral proteases. *Biochimica Et Biophysica Acta*, *1477*(1–2), 16–34. [https://doi.org/10.1016/s0167-4838\(99\)00267-8](https://doi.org/10.1016/s0167-4838(99)00267-8)
- 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. <https://doi.org/10.1074/jbc.RA118.005531>
- Wu, X., Li, Y., Crise, B., Burgess, S. M., & Munroe, D. J. (2005). Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *Journal of Virology*, *79*(8), 5211–5214. <https://doi.org/10.1128/JVI.79.8.5211-5214.2005>
- 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. <https://doi.org/10.1128/JVI.70.8.5695-5700.1996>
- Xu, L., Wrona, T. J., & Dudley, J. P. (1997). Strain-specific expression of spliced MMTV RNAs containing the superantigen gene. *Virology*, *236*(1), 54–65. <https://doi.org/10.1006/viro.1997.8717>
- Yang, Y., Qu, N., Tan, J., Rushdi, M. N., Krueger, C. J., & Chen, A. K. (2018). Roles of Gag-RNA interactions in HIV-1 virus assembly deciphered by single-molecule localization microscopy. *Proceedings of the National Academy of Sciences*, *115*(26), 6721–6726. <https://doi.org/10.1073/pnas.1805728115>
- Yoder, K. E., & Bushman, F. D. (2000). Repair of gaps in retroviral DNA integration intermediates. *Journal of Virology*, *74*(23), 11191–11200. <https://doi.org/10.1128/JVI.74.23.11191-11200.2000>
- Yoshida, M., Miyoshi, I., & Hinuma, Y. (1982). Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(6), 2031–2035. <https://doi.org/10.1073/pnas.79.6.2031>

- Yoshinaka, Y., Katoh, I., Copeland, T. D., & Oroszlan, S. (1985a). Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. *Proceedings of the National Academy of Sciences of the United States of America*, 82(6), 1618–1622. <https://doi.org/10.1073/pnas.82.6.1618>
- Yoshinaka, Y., Katoh, I., Copeland, T. D., & Oroszlan, S. (1985b). Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease. *Journal of Virology*, 55(3), 870–873. <https://doi.org/10.1128/JVI.55.3.870-873.1985>
- Yu, G., Shen, F. S., Sturch, S., Aquino, A., Glazer, R. I., & Felsted, R. L. (1995). Regulation of HIV-1 gag protein subcellular targeting by protein kinase C. *The Journal of Biological Chemistry*, 270(9), 4792–4796. <https://doi.org/10.1074/jbc.270.9.4792>
- Zábranský, Ales, Hadravová, R., Stokrová, J., Sakalian, M., & Pichová, I. (2009). Premature processing of mouse mammary tumor virus Gag polyprotein impairs intracellular capsid assembly. *Virology*, 384(1), 33–37. <https://doi.org/10.1016/j.virol.2008.10.038>
- Zhang, G., Sharon, D., Jovel, J., Liu, L., Wine, E., Tahbaz, N., Indik, S., & Mason, A. (2015). Pericentriolar targeting of the mouse mammary tumor virus Gag protein. *PLOS ONE*, 10(6), e0131515. <https://doi.org/10.1371/journal.pone.0131515>
- Zhao, G., Perilla, J. R., Yufenyuy, E. L., Meng, X., Chen, B., Ning, J., Ahn, J., Gronenborn, A. M., Schulten, K., Aiken, C., & Zhang, P. (2013). Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497(7451), 643–646. <https://doi.org/10.1038/nature12162>
- 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. <https://doi.org/10.1016/j.jmb.2006.10.013>
- 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. <https://doi.org/10.1016/j.jmb.2005.04.046>
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), 3406–3415. <https://doi.org/10.1093/nar/gkg595>

List of Publications

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 Pr77^{Gag} expressed in prokaryotic and eukaryotic cells. *Viruses*, *10*(6), 334. <https://doi.org/10.3390/v10060334>.

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 Pr77^{Gag}. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkab223>.

Appendices

Appendix A

Primers used for introduction of mutations and construction of clones						
Mutation	Primers	Region where mutations were introduced/ description	*S or AS	Sequence	Nucleotide position	Starting plasmid
AK68	OTR 1448	Apical part of SL2	AS	5' TGC CGC ACT CGG CCG ACA GGT GTG TCA CCG GGG GGT GCG GGG GGA CCC TCT GGA A 3'	HYB-MTV 1251-1305	DA024 (Aktar <i>et al.</i> ,2014)
	OTR 1449		S	5' GTC GGC CGA GTG CGG CA 3'	HYB-MTV 1289-1305	
AK69	OTR 1450	Bulge downstream of SL4	AS	5' AGC CCG AGA CCC CCA TTT GGT ATG GCT CAC CGT AAC CTA CCT C 3'	HYB-MTV 1458-1500	DA024
	OTR 1451		S	5' AAA TGG GGG TCT CGG GCT 3'	HYB-MTV 1483-1500	
AK70	OTR 1452	Unpaired region upstream of SL5	AS	5' AGA AAC AAA GAG TTT CTG CCC CCT TGA GCC CGA GAC CCC CAT TT 3'	HYB-MTV 1483-1526	DA024
	OTR 1453		S	5' GGG CAG AAA CTC TTT GTT TCT 3'	HYB-MTV 1506-1526	
AK73	OTR 1454	Basal part of SL3	AS	5' TAT GGT GAG TCC GTT CCG CAG ATG TGA TGA TAG CCA GAC AAG AAA G 3'	HYB-MTV 1385-1430	DA024
	OTR 1455		S	5' GCG GAA CGG ACT CAC CAT A 3'	HYB-MTV 1412-1430	
AK74	OTR 1456	PBS region	AS	5' TCA CTT ATC CGA GGG TCC CGG AAG GGG AGC GAT CTG CCG CAG TCG GCC GAC C 3'	HYB-MTV 1288-1339	DA024
	OTR 1457		S	5' GGG ACC CTC GGA TAA GTG A 3'	HYB-MTV 1321-1339	

Appendix A (continued)

Primers used for introduction of mutations and construction of clones						
Mutation Names	Primers	Region where mutations were introduced/ description	*S or AS	Sequence	Nucleotide position	Starting plasmid
AK80	OTR 1580	PBS region	AS	5' TCA CTT ATC CGA GGG TCC CGG AAG GCG GCC GAC GTC CCG CAG TCG GCC GAC C 3'	HYB-MTV 1288-1339	DA024
	OTR 1457		S	5' GGG ACC CTC GGA TAA GTG A 3'	HYB-MTV 1321-1339	
AK81	OTR 1581	PBS region	AS	5' TCA CTT ATC CGA GGG TCC CTC TTC TTT CTT CTG AGG CCG CAG TCG GCC GAC C 3'	HYB-MTV 1288-1339	DA024
	OTR 1457		S	5' GGG ACC CTC GGA TAA GTG A 3'	HYB-MTV 1321-1339	
AK82	OTR 1582	PBS region	AS	5' TCA CTT ATC CGA GGG TCC CGG AAG GGG CGC CAG CTG CCG CAG TCG GCC GAC C 3'	HYB-MTV 1288-1339	DA024
	OTR 1457		S	5' GGG ACC CTC GGA TAA GTG A 3'	HYB-MTV 1321-1339	
AK83	OTR 1583	PBS region	AS	5' TCA CTT ATC CGA GGG TCC CTG TTC ATT TCA TAC CAC CCG CAG TCG GCC GAC C 3'	HYB-MTV 1288-1339	DA024
	OTR 1457		S	5' GGG ACC CTC GGA TAA GTG A 3'	HYB-MTV 1321-1339	
AK84	OTR 1581	PBS region	AS	5' TCA CTT ATC CGA GGG TCC CTC TTC TTT CTT CTG AGG CCG CAG TCG GCC GAC C 3'	HYB-MTV 1288-1339	SA044P (Aktar <i>et al.</i> ,2014)
	OTR 1457		S	5' GGG ACC CTC GGA TAA GTG A 3'	HYB-MTV 1321-1339	

Appendix A (continued)

Primers used for introduction of mutations and construction of clones						
Mutation Names	Primers	Region where mutations were introduced/ description	*S or AS	Sequence	Nucleotide position	Starting plasmid
AK67	OTR 1460	ssPurines	AS	5' AAT GGC TCA CCG TAA CCT ACC TCT TCT CGG TAG GCG GGA CTG CAG CTC 3'	HYB-MTV 1433-1480	DA024
	OTR 1405		S	5' GTA GGT TAC GGT GAG CCA TT 3'	HYB-MTV 1461-1480	
AK44	OTR 1396	ssPurines	AS	5' AAT GGC TCA CCG TAA CCT ACC TGA AGT CCG TAG GCG GGA CTG CAG CTC 3'	HYB-MTV 1433-1480	DA024
	OTR 1405		S	5' GTA GGT TAC GGT GAG CCA TT 3'	HYB-MTV 1461-1480	
AK45	OTR 1397	ssPurines	AS	5' AAT GGC TCA CCG TAA CCT ACG ACT TCA GGG TAG GCG GGA CTG CAG CTC 3'	HYB-MTV 1433-1480	DA024
	OTR 1405		S	5' GTA GGT TAC GGT GAG CCA TT 3'	HYB-MTV 1461-1480	
AK50	OTR 1402	ssPurines	AS	5' AAT GGC TCA CCG TAA CCT ACC TGA ACT CCG TAG GCG GGA CTG CAG CTC 3'	HYB-MTV 1433-1480	DA024
	OTR 1405		S	5' GTA GGT TAC GGT GAG CCA TT 3'	HYB-MTV 1461-1480	
AK62	OTR 1414	ssPurines	AS	5' AAT GGC TCA CCG TAA CCT ACG ACT TCT CCG TAG GCG GGA CTG CAG CTC 3'	HYB-MTV 1433- 480	DA024
	OTR 1405		S	5' GTA GGT TAC GGT GAG CCA TT 3'	HYB-MTV 1461-1480	

Appendix A (continued)

Primers used for introduction of mutations and construction of clones						
Mutation Names	Primers	Region where mutations were introduced/ description	*S or AS	Sequence	Nucleotide position	Starting plasmid
AK63	OTR 1413	ssPurines	AS	5' AAT GGC TCA CCG TAA CCT ACC TCT TGA GGG TAG GCG GGA CTG CAG CTC 3'	HYB-MTV 1433-1480	DA024
	OTR 1405		S	5' GTA GGT TAC GGT GAG CCA TT 3'	HYB-MTV 1461-1480	
AK18	OTR 1035	ssPurines	AS	5' CCG TAA CCT ACG AGA AGA GGG TAG GCG GGA CTG CAG CTC C 3'	HYB-MTV 1451-1432	DA024
	OTR 1034		S	5' CCT CTT CTC GTA GGT TAC GGT GAG CCA TTG G 3'	HYB-MTV 1471-1461	
AK29	OTR 1371	Amplifying from R to mSD	AS	5' GAT TGG TGT TTC GGC ATC CTC TTC TCC GTA GGC GGG 3'	HYB-MTV 1443-1460 & 6535- 6552	HYB- MTV
	OTR 1370	Amplifying from <i>env</i> splice acceptor to 424 bp of <i>env</i>	S	5' GAT GCC GAA ACA CCA ATC TG 3'	HYB-MTV 6535-6554	
	OTR 1372		AS	5' aaa ccc ggg TAA ACC CGT GAA AGT CAG GC 3'	HYB-MTV 6940-6959	
AK30	OTR 1374	Amplifying from R to mSD	AS	5' AGG CTC TTC GCA AGG CAC TCT TCT CCG TAG GCG GG 3'	HYB-MTV 1443-1460 & 8477- 8493	HYB- MTV
	OTR 1373	Amplifying from <i>sag</i> splice acceptor to 351 bp of <i>sag</i>	S	5' TGC CTT GCG AAG AGC CTT G 3'	HYB-MTV 8477-8495	
	OTR 1375		AS	5' aaa ccc ggg TTT CTG AAG GAC AAA ATC GAT G 3'	HYB-MTV 8880-8901	

Appendix A (continued)

Primers used for introduction of mutations and construction of clones						
Mutation Names	Primers	Region where mutations were introduced/ description	*S or AS	Sequence	Nucleotide position	Starting plasmid
OTR 249		Outer primers used for construction of subgenomic transfer vectors	S	5' CC GCT AGC CTT CGC GAT GTA CGG GCC AGA 3'	pCDNA3 204-224	First round amplification products in the SOE PCR
OTR 552			AS	5' cg act agt gat atc GTT CCC CTG GTC CCA T 3'	HYB-MTV 1885-1867	
OTR 984		Primers used for construction of <i>in vitro</i> transcribing clones	S	5' ccc aag ctt AAT ACG ACT CAC TAT AGG GCA ACA GTC CTA ATA TTC ACG 3'	HYB-MTV 1173-1193	DA024/ its mutants
OTR 985			AS	5' aaa ccc ggg TTC CCC TGG TCC CAT AAG 3'	HYB-MTV 1885-1867	

Appendix A (continued)

Primers used for SHAPE, RT-qPCR and other amplifications				
Primer	Description	*S or AS	Sequence	Nucleotide position or reference
OTR 10	Labelled primers used for SHAPE analysis of WT (SA35)	AS	VIC - 5' AACAGATTTGGCTTCTGCGG 3'	HYB-MTV 1786-1805
OTR 11		AS	NED - 5' AACAGATTTGGCTTCTGCGG 3'	HYB-MTV 1786-1805
OTR 14		AS	VIC- 5' AGTTTCTGCCCTTTTGAGCC 3'	HYB-MTV 1497-1516
OTR 15		AS	NED - 5' AGTTTCTGCCCTTTTGAGCC 3'	HYB-MTV 1497-1516
Splice_1_1	Labelled primers used for SHAPE analysis of <i>sag</i> RNA (AK30)	AS	VIC- 5' GCTCTTGTGATGATAGCCAG 3'	HYB-MTV 1394-1413
Splice_1_2		AS	NED- 5' GCTCTTGTGATGATAGCCAG 3'	
Splice_2_1		AS	VIC- 5' CTATGCCAAGTTTGCAGCAG 3'	HYB-MTV 8693-8712
Splice_2_2		AS	NED- 5' CTATGCCAAGTTTGCAGCAG 3'	
Splice_1_1	Labelled primers used for SHAPE analysis of <i>env</i> RNA (AK29)	AS	VIC- 5' GCTCTTGTGATGATAGCCAG 3'	HYB-MTV 1394-1413
Splice_1_2		AS	NED- 5' GCTCTTGTGATGATAGCCAG 3'	
Splice_3_1		AS	VIC- 5' GGTTTTAAGAACCTCCTCCG 3'	HYB-MTV 6735-6754
Splice_3_2		AS	NED- 5' GGTTTTAAGAACCTCCTCCG 3'	
OTR 580	β -actin spliced mRNA	S	5' TGA GCT GCG TGT GGC TCC 3'	(Aktar <i>et al.</i> ,2014)

Appendix A (continued)

OTR 581	β -actin spliced or unspliced mRNA	AS	5' GGC ATG GGG GAG GGC ATA CC 3'	(Aktar <i>et al.</i> ,2014)
OTR 582	β -actin unspliced mRNA	S	5' CCA GTG GCT TCC CCA GTG 3'	
OTR 1391	Vector specific	S	5' GTC CTA ATA TTC ACG TCT CGT GTG 3'	HYB-MTV 1216-1233
OTR 1392	Vector specific	AS	5' CTG TTC GGG CGC CAG CTG CCG CAG 3'	HYB- MTV 1364-1385
MTV-1LTR - SITEM1 FAM	qPCR probe	Probe	FAM- 5' TCG CCA TCC CGT CTC C 3'	HYB-MTV 1214-1229
MTV- 1LTR -SITEF	qPCR Forward primer	S	5' CGT CTC GTG TGT TTG TGT TTG TGT CTG T 3'	HYB-MTV 1192-1213
MTV-1LTR -SITER	qPCR Reverse Primer	AS	5' CCT CTG GAA AGT GAA GGA TAA GTG A 3'	HYB-MTV 1259-1235

AAT ACG ACT CAC TAT AGG; T7 Promoter

Sequence in lower case; dummy sequences that were introduced in the oligos.

Sequence in lower case and bold; restriction enzyme sequences that were introduced in the oligos.

*S, sense; AS, antisense.

Appendix B

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
1	G	-999	0	1	G	-999	0
2	C	-999	0	2	C	-999	0
3	A	-999	0	3	A	-999	0
4	A	-999	0	4	A	-999	0
5	C	-999	0	5	C	-999	0
6	A	-999	0	6	A	-999	0
7	G	-999	0	7	G	-999	0
8	U	-999	0	8	U	-999	0
9	C	0.3025	0.3391	9	C	-999	0
10	C	0.0000	0.0000	10	C	-999	0
11	U	0.4425	0.2912	11	U	0.6500	0.3054
12	A	0.5100	0.1995	12	A	0.3500	0.1594
13	A	0.9575	0.2722	13	A	0.6000	0.1558
14	U	0.9950	0.2659	14	U	0.8800	0.0787
15	A	2.4625	0.7947	15	A	2.2125	0.6873
16	U	1.5575	0.4891	16	U	1.5550	0.2515
17	U	0.5900	0.1865	17	U	0.6250	0.1190
18	C	0.2825	0.1959	18	C	0.0975	0.1127
19	A	0.2900	0.2230	19	A	0.2650	0.1100
20	C	0.0100	0.0082	20	C	0.0600	0.0392
21	G	0.3975	0.1372	21	G	0.5000	0.1225
22	U	1.9875	0.5611	22	U	2.1050	0.2689

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
23	C	0.6125	0.2445	23	C	0.8150	0.1873
24	U	2.0950	0.5186	24	U	2.1850	0.3074
25	C	0.6100	0.2467	25	C	0.7225	0.1075
26	G	0.0175	0.0287	26	G	0.0725	0.0645
27	U	0.0050	0.0100	27	U	0.0350	0.0436
28	G	0.0125	0.0250	28	G	0.0450	0.0900
29	U	0.3750	0.0843	29	U	0.0600	0.0365
30	G	0.4775	0.1075	30	G	0.1000	0.0688
31	U	0.3375	0.0629	31	U	0.2200	0.0648
32	U	0.6650	0.1498	32	U	0.2175	0.0873
33	U	0.3925	0.0998	33	U	0.2250	0.0733
34	G	0.1575	0.0472	34	G	0.2750	0.0619
35	U	1.1800	0.3455	35	U	1.2125	0.2389
36	G	0.0750	0.0420	36	G	0.1550	0.0387
37	U	0.0075	0.0150	37	U	0.0375	0.0287
38	C	0.0000	0.0000	38	C	0.0000	0.0000
39	U	0.0100	0.0200	39	U	0.0025	0.0050
40	G	0.0025	0.0050	40	G	0.0000	0.0000
41	U	0.0000	0.0000	41	U	0.0000	0.0000
42	U	0.0850	0.1066	42	U	0.0200	0.0283
43	C	0.1700	0.1140	43	C	0.0750	0.0881
44	G	0.0000	0.0000	44	G	0.0000	0.0000
45	C	0.0000	0.0000	45	C	0.0425	0.0850
46	C	0.1000	0.2000	46	C	0.3200	0.6400

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
47	A	1.8675	0.5982	47	A	2.3950	0.6409
48	U	0.9625	0.3280	48	U	1.3250	0.1865
49	C	0.0775	0.0556	49	C	0.2525	0.1526
50	C	0.0000	0.0000	50	C	0.0000	0.0000
51	C	0.0000	0.0000	51	C	0.0025	0.0050
52	G	0.0700	0.1334	52	G	0.0575	0.0465
53	U	0.1225	0.0457	53	U	0.1125	0.0250
54	C	0.2300	0.1111	54	C	0.2825	0.0655
55	U	0.0000	0.0000	55	U	0.0025	0.0050
56	C	0.0000	0.0000	56	C	0.0000	0.0000
57	C	0.0000	0.0000	57	C	0.0000	0.0000
58	G	0.0175	0.0350	58	G	0.0125	0.0250
59	C	0.0250	0.0500	59	C	0.0225	0.0450
60	U	0.5100	0.2340	60	U	0.3250	0.0480
61	C	0.1125	0.0450	61	C	0.2300	0.0548
62	G	0.0000	0.0000	62	G	0.0000	0.0000
63	U	0.0550	0.0971	63	U	0.0525	0.0640
64	C	0.7325	0.5016	64	C	0.5000	0.5788
65	A	0.4425	0.1255	65	A	0.2425	0.1087
66	C	0.0125	0.0250	66	C	0.0175	0.0171
67	U	0.3325	0.1109	67	U	0.2775	0.0330
68	U	0.2600	0.3308	68	U	0.2850	0.2310
69	A	0.4925	0.1678	69	A	0.5025	0.0695

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
70	U	0.2475	0.1100	70	U	0.2700	0.0337
71	C	0.0550	0.0911	71	C	0.0500	0.0577
72	C	0.0475	0.0550	72	C	0.0950	0.0954
73	U	0.9350	0.2173	73	U	1.0275	0.2326
74	U	1.3225	0.2234	74	U	1.0575	0.3489
75	C	0.6975	0.6551	75	C	0.6400	0.7812
76	A	1.4675	0.4583	76	A	1.4950	0.4671
77	C	0.3400	0.1433	77	C	0.4225	0.1830
78	U	1.4075	0.6649	78	U	1.5675	0.4740
79	U	1.7600	0.6639	79	U	1.9550	0.5814
80	U	0.8300	0.3556	80	U	0.9050	0.1942
81	C	0.0725	0.0797	81	C	0.0750	0.0957
82	C	0.4775	0.9550	82	C	0.3475	0.5638
83	A	1.1375	0.9932	83	A	0.5475	0.8918
84	G	1.5475	0.8387	84	G	1.7050	0.2076
85	A	0.4475	0.3435	85	A	0.2925	0.1053
86	G	0.1625	0.2803	86	G	0.2050	0.2762
87	G	0.0000	0.0000	87	G	0.2450	0.3489
88	G	0.0325	0.0650	88	G	0.0825	0.1650
89	U	0.0000	0.0000	89	U	0.0050	0.0100
90	C	0.0025	0.0050	90	C	0.0150	0.0300
91	C	0.0375	0.0450	91	C	0.0450	0.0772
92	C	0.0675	0.0780	92	C	0.1750	0.1546

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
93	C	0.0350	0.0700	93	C	0.0275	0.0550
94	C	0.0175	0.0350	94	C	0.0475	0.0950
95	C	0.0275	0.0550	95	C	0.0750	0.0500
96	G	0.0300	0.0535	96	G	0.1100	0.0141
97	C	0.0600	0.1200	97	C	0.0900	0.1023
98	A	0.3375	0.2089	98	A	0.3675	0.0946
99	G	0.5325	0.1839	99	G	0.5800	0.1102
100	A	1.6725	0.5867	100	A	1.5775	0.2980
101	C	0.1200	0.0606	101	C	0.1475	0.0556
102	C	0.1325	0.1247	102	C	0.2300	0.1829
103	C	0.3150	0.3964	103	C	0.6625	0.3340
104	C	0.0050	0.0100	104	C	0.0500	0.0594
105	G	0.0000	0.0000	105	G	0.0075	0.0150
106	G	0.0000	0.0000	106	G	0.0350	0.0473
107	U	0.1200	0.0920	107	U	0.1400	0.0356
108	G	0.0475	0.0403	108	G	0.0325	0.0359
109	A	0.0650	0.0569	109	A	0.0475	0.0150
110	C	0.0000	0.0000	110	C	0.0050	0.0100
111	C	0.1750	0.0915	111	C	0.1800	0.0796
112	C	0.6875	0.2885	112	C	0.8550	0.1139
113	U	2.7375	0.8160	113	U	2.5350	0.2659
114	C	1.1850	0.3022	114	C	1.0350	0.3795
115	A	3.4175	0.8825	115	A	3.1675	0.5784

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
116	G	0.1375	0.1187	116	G	0.1400	0.0383
117	G	0.0400	0.0497	117	G	0.0800	0.0294
118	U	0.0775	0.0699	118	U	0.1150	0.0191
119	C	0.1575	0.2559	119	C	0.0475	0.0250
120	G	0.5650	0.1790	120	G	0.4925	0.1513
121	G	0.7350	0.5208	121	G	0.9600	0.1249
122	C	0.0125	0.0250	122	C	0.0150	0.0238
123	C	0.0400	0.0800	123	C	0.0025	0.0050
124	G	0.1500	0.1192	124	G	0.1850	0.0614
125	A	0.3250	0.1103	125	A	0.3275	0.0538
126	C	0.0275	0.0263	126	C	0.0625	0.0310
127	U	0.0275	0.0340	127	U	0.0700	0.0337
128	G	0.0250	0.0300	128	G	0.0575	0.0206
129	C	0.0000	0.0000	129	C	0.0000	0.0000
130	G	0.0200	0.0400	130	G	0.0050	0.0100
131	G	0.0025	0.0050	131	G	0.0000	0.0000
132	C	0.3100	0.6200	132	C	0.3125	0.3709
133	A	0.0050	0.0100	133	A	0.1525	0.2035
134	G	0.9975	0.7101	134	G	0.8675	0.7830
135	C	0.7600	0.7869	135	C	0.3300	0.2232
136	U	1.2250	0.6268	136	U	1.5050	0.2999
137	G	0.3100	0.0622	137	G	0.2875	0.0750
138	G	0.1650	0.3300	138	G	0.2050	0.1091

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
139	C	0.1125	0.0096	139	C	0.1425	0.0640
140	G	0.1275	0.0377	140	G	0.1375	0.0287
141	C	0.0000	0.0000	141	C	0.0050	0.0100
142	C	0.0550	0.1100	142	C	0.1750	0.0759
143	C	0.2000	0.1822	143	C	0.3925	0.2103
144	G	0.6975	0.1473	144	G	0.7825	0.1391
145	A	1.0950	0.2745	145	A	1.2475	0.2526
146	A	0.7550	0.5683	146	A	0.7375	0.4988
147	C	0.9625	0.6547	147	C	0.9500	1.0970
148	A	0.0275	0.0550	148	A	0.0625	0.1250
149	G	0.0000	0.0000	149	G	0.0000	0.0000
150	G	0.0000	0.0000	150	G	0.0975	0.1950
151	G	0.0000	0.0000	151	G	0.1275	0.2550
152	A	0.0000	0.0000	152	A	0.0225	0.0450
153	C	0.0475	0.0950	153	C	0.8975	1.6452
154	C	2.1725	2.5086	154	C	0.3475	0.5638
155	C	0.1425	0.2850	155	C	0.7275	0.5416
156	U	0.1650	0.3300	156	U	0.4050	0.4853
157	C	0.1625	0.1752	157	C	0.3950	0.3287
158	G	0.1300	0.2600	158	G	0.2500	0.4609
159	G	0.0525	0.1050	159	G	0.0950	0.1769
160	A	0.0900	0.0883	160	A	0.1500	0.0141
161	U	0.3300	0.2080	161	U	0.3375	0.1905

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
162	A	0.0900	0.0753	162	A	0.0775	0.0763
163	A	0.0475	0.0320	163	A	0.0175	0.0222
164	G	0.0175	0.0350	164	G	0.0075	0.0150
165	U	0.1050	0.2100	165	U	0.0325	0.0340
166	G	0.0250	0.0500	166	G	0.0575	0.0386
167	A	0.0550	0.0802	167	A	0.1950	0.1115
168	C	0.1575	0.1819	168	C	0.4150	0.3373
169	C	0.1450	0.1127	169	C	0.1700	0.1831
170	C	0.1350	0.0926	170	C	0.1825	0.1758
171	U	0.6225	0.0189	171	U	0.5200	0.0707
172	U	0.3675	0.1619	172	U	0.2675	0.1090
173	G	0.3100	0.1407	173	G	0.2825	0.0922
174	U	0.3200	0.0183	174	U	0.2550	0.0705
175	C	0.1325	0.1031	175	C	0.0825	0.0929
176	U	0.4125	0.1750	176	U	0.3625	0.2541
177	C	0.1925	0.1394	177	C	0.2400	0.1757
178	U	0.5425	0.5803	178	U	0.7350	0.5244
179	A	0.8475	0.0709	179	A	0.7350	0.1852
180	U	0.3975	0.0591	180	U	0.4000	0.0949
181	U	0.5775	0.0854	181	U	0.5200	0.0490
182	U	0.3300	0.0638	182	U	0.2750	0.1589
183	C	0.1600	0.1010	183	C	0.1900	0.1719
184	U	0.5275	0.4565	184	U	0.6100	0.4399

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
185	A	0.6025	0.1090	185	A	0.4500	0.3014
186	C	0.2500	0.2501	186	C	0.1725	0.1424
187	U	0.4525	0.3676	187	U	0.3075	0.2341
188	A	0.7000	0.2820	188	A	0.4950	0.2053
189	U	0.2200	0.1030	189	U	0.2325	0.1269
190	U	0.1600	0.1117	190	U	0.2175	0.0896
191	U	0.1700	0.0938	191	U	0.1475	0.0768
192	G	0.0400	0.0497	192	G	0.0600	0.0735
193	G	0.0000	0.0000	193	G	0.0000	0.0000
194	U	0.0025	0.0050	194	U	0.0000	0.0000
195	G	0.0050	0.0100	195	G	0.0050	0.0100
196	U	0.2350	0.1047	196	U	0.1400	0.0787
197	U	0.3275	0.1289	197	U	0.4050	0.0900
198	U	0.3550	0.1678	198	U	0.4350	0.1085
199	G	0.2375	0.1826	199	G	0.0075	0.0150
200	U	0.1875	0.1167	200	U	0.0125	0.0250
201	C	0.0100	0.0200	201	C	0.0050	0.0100
202	U	0.3700	0.1655	202	U	0.0750	0.1136
203	U	0.8625	0.2775	203	U	0.5350	0.1950
204	G	0.1300	0.0408	204	G	0.0350	0.0700
205	U	0.2450	0.3188	205	U	0.3950	0.2124
206	A	0.3275	0.0608	206	A	0.2675	0.1132
207	U	0.2975	0.1184	207	U	0.2575	0.0250

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
208	U	0.3975	0.2211	208	U	0.1400	0.0583
209	G	0.7950	0.3783	209	G	0.0350	0.0507
210	U	0.4025	0.0881	210	U	0.3150	0.2453
211	C	0.0000	0.0000	211	C	0.0000	0.0000
212	U	0.0000	0.0000	212	U	0.0000	0.0000
213	C	0.0000	0.0000	213	C	0.0000	0.0000
214	U	0.0000	0.0000	214	U	0.0475	0.0550
215	U	0.1725	0.1087	215	U	0.3200	0.0739
216	U	0.2750	0.1162	216	U	0.3350	0.0843
217	C	0.0225	0.0450	217	C	0.0200	0.0245
218	U	0.2275	0.0943	218	U	0.2350	0.1066
219	U	0.2800	0.0762	219	U	0.2275	0.1072
220	G	0.3425	0.0780	220	G	0.2250	0.0557
221	U	0.1600	0.0316	221	U	0.1150	0.0850
222	C	0.1525	0.0776	222	C	0.1300	0.0294
223	U	0.3200	0.0638	223	U	0.1275	0.0562
224	G	0.3825	0.0826	224	G	0.2050	0.1269
225	G	0.0850	0.0661	225	G	0.0225	0.0450
226	C	0.1875	0.2156	226	C	0.1050	0.1085
227	U	1.3900	0.3710	227	U	1.2175	0.3657
228	A	1.3700	0.1883	228	A	1.2400	0.1641
229	U	1.3575	0.1209	229	U	1.1275	0.2555
230	C	0.6650	1.0812	230	C	0.6500	0.3599

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
231	A	0.8150	0.2068	231	A	0.7300	0.2162
232	U	0.5375	0.1159	232	U	0.4500	0.1857
233	C	0.7450	1.1778	233	C	0.4900	0.3472
234	A	0.8400	0.2102	234	A	0.7825	0.2870
235	C	1.6475	2.9075	235	C	0.7000	0.5141
236	A	0.6600	0.2789	236	A	0.7775	0.3425
237	A	0.7500	0.1598	237	A	0.8625	0.2783
238	G	0.1450	0.0545	238	G	0.0100	0.0200
239	A	0.1650	0.0289	239	A	0.1400	0.0141
240	G	0.0125	0.0189	240	G	0.0875	0.0299
241	C	0.0300	0.0216	241	C	0.0725	0.1384
242	G	0.0425	0.0723	242	G	0.0600	0.1200
243	G	0.1525	0.0954	243	G	0.1600	0.1071
244	A	0.3800	0.1806	244	A	0.3350	0.1204
245	A	0.1600	0.0294	245	A	0.1175	0.0680
246	C	0.0450	0.0614	246	C	0.0000	0.0000
247	G	0.1025	0.1617	247	G	0.0000	0.0000
248	G	0.0450	0.0714	248	G	0.0000	0.0000
249	A	0.4150	0.1535	249	A	0.5425	0.2207
250	C	0.0475	0.0763	250	C	0.0550	0.1100
251	U	0.2875	0.2500	251	U	0.1275	0.0873
252	C	1.3175	1.9319	252	C	1.0575	0.3951
253	A	0.8225	0.1059	253	A	0.4925	0.1245

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
254	C	0.0900	0.1800	254	C	0.0000	0.0000
255	C	1.8125	3.5191	255	C	0.2700	0.4175
256	A	1.5575	0.6464	256	A	0.8675	0.3398
257	U	1.2975	0.6992	257	U	1.1625	0.4132
258	A	2.1950	0.1921	258	A	1.2675	0.3404
259	G	0.4375	0.0806	259	G	0.4525	0.1424
260	G	0.1200	0.0346	260	G	0.0925	0.0330
261	G	0.0275	0.0427	261	G	0.1425	0.0690
262	A	0.2125	0.0472	262	A	0.4275	0.1218
263	G	0.0650	0.0480	263	G	0.0000	0.0000
264	C	0.0475	0.0330	264	C	0.1250	0.1893
265	U	0.3625	0.1819	265	U	0.1300	0.1061
266	G	0.7850	0.2601	266	G	0.2650	0.1292
267	C	0.4175	0.4007	267	C	0.2150	0.1245
268	A	1.0525	0.3371	268	A	0.3175	0.1799
269	G	0.3025	0.3597	269	G	0.1800	0.0804
270	U	0.3525	0.1678	270	U	0.1550	0.0742
271	C	0.0100	0.0200	271	C	0.0050	0.0100
272	C	0.0250	0.0500	272	C	0.0000	0.0000
273	C	0.0100	0.0200	273	C	0.3275	0.3093
274	G	0.1150	0.0777	274	G	1.2725	0.1181
275	C	0.1875	0.1839	275	C	0.2175	0.1739
276	C	0.3850	0.1870	276	C	0.0800	0.0294

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
277	U	0.5200	0.2920	277	U	0.3675	0.0411
278	A	1.1075	0.1905	278	A	0.5375	0.1112
279	C	0.5000	0.1512	279	C	0.3875	0.1059
280	G	0.0725	0.0330	280	G	0.0900	0.0707
281	G	0.0600	0.0632	281	G	0.1550	0.1406
282	A	0.2575	0.1153	282	A	0.3900	0.4318
283	G	0.0850	0.0790	283	G	0.0750	0.0926
284	A	0.2200	0.0942	284	A	0.3150	0.0975
285	A	0.2550	0.0420	285	A	0.4575	0.1664
286	G	0.0575	0.0704	286	G	0.0600	0.1200
287	A	0.4675	0.0525	287	A	0.4600	0.3151
288	G	0.0100	0.0141	288	G	0.4575	0.2455
289	G	0.0000	0.0000	289	U	0.2200	0.1545
290	A	0.0925	0.1069	290	G	0.3550	0.1038
291	U	0.0450	0.0173	291	C	0.1875	0.1410
292	G	0.0400	0.0163	292	C	0.0250	0.0300
293	C	0.0425	0.0723	293	U	0.0450	0.0835
294	C	0.0550	0.0681	294	U	0.0525	0.0377
295	G	0.2825	0.0699	295	G	0.3725	0.0403
296	A	0.4025	0.1408	296	C	0.4800	0.1669
297	A	0.3675	0.0746	297	G	0.1925	0.0435
298	A	0.4225	0.0350	298	A	0.3050	0.0420
299	C	0.1725	0.3450	299	A	0.1300	0.1010

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
300	A	0.1975	0.0991	300	G	0.2400	0.0408
301	C	0.0175	0.0350	301	A	0.7225	0.1396
302	C	0.1500	0.3000	302	G	0.9575	0.1987
303	A	0.2675	0.3021	303	C	0.2600	0.0909
304	A	1.9200	0.3721	304	C	0.1650	0.0874
305	U	0.6975	0.1081	305	U	1.0400	0.2981
306	C	0.6425	0.1357	306	U	0.8850	0.2469
307	U	1.3425	0.2277	307	G	0.9275	0.4509
308	G	0.5075	0.0896	308	A	1.3350	0.2428
309	G	0.0350	0.0700	309	C	0.0600	0.0952
310	G	0.0875	0.0660	310	C	0.1225	0.2450
311	U	0.1550	0.1012	311	A	0.0475	0.0763
312	C	0.0075	0.0150	312	A	0.1250	0.0819
313	C	0.0275	0.0550	313	G	0.1400	0.0548
314	C	0.1475	0.2950	314	U	0.3650	0.0603
315	C	0.1125	0.2250	315	G	0.0025	0.0050
316	G	0.1875	0.2237	316	C	0.0575	0.0802
317	A	0.8150	0.0835	317	A	1.3425	0.2435
318	U	0.0500	0.0577	318	G	0.6425	0.2855
319	C	0.0000	0.0000	319	U	0.5325	0.2926
320	G	0.0500	0.1000	320	C	0.2000	0.2828
321	G	0.1525	0.1644	321	A	0.2400	0.1691
322	U	0.3650	0.1218	322	G	0.5375	0.3804

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
323	U	0.6950	0.1015	323	A	0.8325	0.2061
324	C	0.6475	1.1844	324	U	0.3725	0.2037
325	A	0.3000	0.1402	325	C	0.1600	0.1236
326	U	0.0675	0.0359	326	U	0.5200	0.1080
327	C	0.0075	0.0150	327	U	1.0325	0.7293
328	C	0.0450	0.0772	328	A	0.7000	0.1645
329	G	0.2625	0.1611	329	A	0.6175	0.1001
330	A	1.5825	0.1914	330	C	0.3700	0.3209
331	C	0.1925	0.0386	331	G	0.3925	0.0544
332	C	0.0725	0.0763	332	U	0.6375	0.0810
333	U	0.7325	0.2920	333	G	0.8225	0.2356
334	U	0.9475	0.2016	334	C	0.1425	0.1590
335	U	1.3300	0.1679	335	U	0.4025	0.3369
336	U	2.7650	0.6488	336	U	0.2675	0.1646
337	A	0.7350	0.1473	337	C	0.1900	0.1647
338	C	0.2700	0.1105	338	U	1.0975	0.6766
339	U	0.6575	0.1473	339	U	0.4275	0.0922
340	G	0.5875	0.1788	340	U	1.3200	0.0529
341	A	0.4050	0.1279	341	U	0.8350	0.4994
342	G	0.0175	0.0206	342	A	0.4800	0.2734
343	C	0.0150	0.0173	343	A	0.9325	0.0411
344	G	0.0000	0.0000	344	A	0.9775	0.2572
345	G	0.0200	0.0400	345	A	0.5550	0.2133

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
346	A	0.4425	0.1491	346	A	0.1200	0.1169
347	A	0.9950	0.1518	347	A	0.2200	0.1068
348	A	1.2075	0.0818	348	G	0.2350	0.0614
349	G	0.8850	0.0806	349	A	0.0650	0.0896
350	A	0.9775	0.2792	350	A	0.3300	0.0852
351	A	0.8200	0.0668	351	A	1.2825	0.4192
352	G	0.2025	0.3299	352	A	0.6775	0.1615
353	C	0.8575	1.7150	353	A	0.8500	0.2586
354	A	0.4700	0.3782	354	A	0.4200	0.1449
355	A	0.4200	0.0606	355	G	0.0050	0.0100
356	C	0.1700	0.0408	356	G	0.0075	0.0150
357	G	0.0850	0.0493	357	G	0.1175	0.2350
358	C	0.0000	0.0000	358	G	0.0775	0.1550
359	C	0.0000	0.0000	359	G	0.1450	0.2900
360	C	0.0000	0.0000	360	A	0.3500	0.7000
361	A	0.8575	0.2935	361	A	0.5300	0.3366
362	C	0.8700	1.5465	362	A	1.6550	0.2931
363	A	1.1975	0.3395	363	U	1.8050	0.3142
364	C	0.0750	0.1500	364	G	0.3650	0.2004
365	C	0.1400	0.1241	365	C	0.0000	0.0000
366	U	0.1725	0.1717	366	C	0.0000	0.0000
367	G	0.0000	0.0000	367	G	0.2700	0.1691
368	G	0.0475	0.0950	368	C	0.3100	0.2189

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
369	C	0.9500	1.9000	369	G	0.5425	0.1338
370	A	0.4175	0.2193	370	C	0.0600	0.0455
371	C	0.1500	0.0913	371	C	0.0975	0.0903
372	U	0.0675	0.0532	372	U	0.2400	0.0245
373	G	0.0550	0.0451	373	G	0.1250	0.0404
374	C	0.0225	0.0287	374	C	1.1750	0.5459
375	G	0.0200	0.0400	375	A	1.4450	0.0173
376	G	0.1625	0.1535	376	G	0.0025	0.0050
377	A	0.6725	0.3041	377	C	0.0000	0.0000
378	G	0.5950	0.5087	378	A	0.0000	0.0000
379	A	1.5150	0.3739	379	G	0.0000	0.0000
380	A	0.0850	0.0532	380	A	0.5100	0.1924
381	A	1.4125	0.3289	381	A	0.7150	0.1515
382	A	1.1650	0.3193	382	A	1.3800	0.1802
383	C	1.3425	0.2707	383	U	1.0175	0.4046
384	G	0.0475	0.0411	384	G	0.0000	0.0000
385	C	0.0000	0.0000	385	G	0.0000	0.0000
386	C	0.0000	0.0000	386	U	0.0450	0.0465
387	G	0.0000	0.0000	387	U	0.0850	0.1179
388	C	0.0100	0.0200	388	G	0.1975	0.0723
389	C	0.0900	0.0589	389	A	0.6350	0.1881
390	G	0.1825	0.0386	390	A	0.7950	0.2034
391	C	0.3525	0.0846	391	C	0.0525	0.0443

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
392	G	0.8300	0.2368	392	U	0.0000	0.0000
393	A	0.3550	0.1072	393	C	0.0800	0.0956
394	G	0.0400	0.0566	394	C	0.0000	0.0000
395	A	0.1275	0.0377	395	C	0.0000	0.0000
396	U	0.1100	0.0469	396	G	0.2950	0.1271
397	G	0.0950	0.0387	397	A	0.4975	0.3597
398	A	0.4350	0.1139	398	G	1.0250	0.2193
399	G	0.5875	0.0954	399	A	2.6075	0.6607
400	A	0.2325	0.0359	400	G	0.8200	1.6400
401	A	0.2375	0.0330	401	U	0.0850	0.0603
402	A	0.3600	0.0455	402	G	0.1900	0.1049
403	G	0.2400	0.0753	403	U	1.7625	0.4518
404	A	0.6775	0.1024	404	C	0.3700	0.3136
405	U	0.6525	0.2090	405	C	0.6050	0.2533
406	C	0.8925	1.7850	406	U	2.7550	1.1024
407	A	0.6775	0.1975	407	A	0.6275	0.0591
408	A	0.4625	0.1994	408	C	0.0000	0.0000
409	C	0.6850	1.3700	409	A	0.0500	0.1000
410	A	0.1225	0.1109	410	C	0.0300	0.0535
411	G	0.0625	0.0793	411	C	0.0450	0.0900
412	A	0.0375	0.0750	412	U	0.2700	0.5400
413	A	0.1225	0.2130	413	A	0.8150	0.1136
414	A	0.2175	0.1159	414	G	1.8875	0.4711

Appendix B (continued)

AK29/ <i>env</i> mRNA				AK30/ <i>sag</i> mRNA			
Nucleotide number & sequence		Mean	SD	Nucleotide number & sequence		Mean	SD
415	A	0.2850	0.0723	415	G	0.0000	0.0000
416	G	0.6175	0.1443	416	G	0.0000	0.0000
417	U	0.2300	0.0337	417	G	0.0000	0.0000
418	C	0.0350	0.0635	418	A	0.2900	0.0966
419	C	0.1425	0.0189	419	G	0.0175	0.0236
420	G	0.3225	0.0532	420	A	1.1550	0.3854
421	G	0.0475	0.0512	421	A	1.1975	0.1473
422	A	0.0550	0.1100	422	G	0.3475	0.1524
423	G	0.0000	0.0000	423	C	0.1200	0.2400
424	G	0.0000	0.0000	424	A	0.1750	0.3500
425	A	0.0000	0.0000	425	G	0.0000	0.0000
426	U	0.1075	0.0918	426	C	0.0000	0.0000
427	G	0.4725	0.3188	427	C	0.0375	0.0519
428	A	0.6750	0.1576	428	A	0.5175	0.3410
429	A	0.8000	0.2525	429	A	1.4025	0.1533
430	U	0.0525	0.0386	430	G	1.8350	0.1179
431	C	0.0550	0.0640	431	G	0.5175	0.2228
432	U	0.0500	0.1000	432	G	0.0000	0.0000

Mean SHAPE reactivities of first 432 nts of spliced *env* (AK29) and *sag* (AK30) mRNAs from four experiments. The boxed and yellow highlighted nucleotides show reduced SHAPE reactivities of single stranded purines (ssPurines; from nucleotides 280 to 288) compared to those of unspliced gRNA.

Appendix C

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4uM Pr77 ^{Gag}	
		Mean	SD	Mean	SD
1	G	-999	0	-999	0
2	C	-999	0	-999	0
3	A	-999	0	-999	0
4	A	-999	0	-999	0
5	C	-999	0	-999	0
6	A	-999	0	-999	0
7	G	-999	0	-999	0
8	U	-999	0	-999	0
9	C	-999	0	-999	0
10	C	-999	0	-999	0
11	U	-999	0	-999	0
12	A	-999	0	-999	0
13	A	-999	0	-999	0
14	U	-999	0	-999	0
15	A	-999	0	-999	0
16	U	2.6225	0.5856	-999	0
17	U	1.21	0.3337	-999	0
18	C	0.1075	0.1295	-999	0
19	A	0.29	0.1802	-999	0
20	C	0.0325	0.0525	-999	0

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4uM Pr77 ^{Gag}	
		Mean	SD	Mean	SD
21	G	0.605	0.3712	-999	0
22	U	3.0475	0.9153	-999	0
23	C	0.7675	0.1493	-999	0
24	U	4.5025	1.5269	-999	0
25	C	1.305	0.3112	0.745	0.4834
26	G	0.1825	0.1034	0.075	0.1500
27	U	0.0375	0.0450	0.0375	0.0519
28	G	0.01	0.0141	0.14	0.1635
29	U	0.1275	0.1438	0.2375	0.1340
30	G	0.08	0.0698	0.1375	0.1187
31	U	0.0425	0.0568	0.135	0.0656
32	U	0.1475	0.0330	0.1425	0.0746
33	U	0.075	0.0465	0.075	0.0835
34	G	0.0475	0.0525	0.065	0.0790
35	U	1.95	0.6578	1.21	0.6981
36	G	0.2725	0.1034	0.14	0.1449
37	U	0.0075	0.0150	0.0175	0.0222
38	C	0.005	0.0100	0	0.0000
39	U	0	0.0000	0.0275	0.0550
40	G	0	0.0000	0	0.0000
41	U	0.0025	0.0050	0	0.0000

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
42	U	0.245	0.3287	0.2325	0.2794
43	C	0.33	0.1479	0.225	0.3009
44	G	0.005	0.0100	0.0125	0.0189
45	C	0	0.0000	0.1475	0.2950
46	C	0	0.0000	0.6375	0.8320
47	A	2.2375	0.7006	1.6675	0.4861
48	U	1.31	0.4790	0.9625	0.3018
49	C	0.205	0.0911	0.28	0.2110
50	C	0	0.0000	0.065	0.1300
51	C	0	0.0000	0.0775	0.1239
52	G	0.075	0.0772	0.07	0.1400
53	U	0.255	0.1967	0.415	0.1603
54	C	0.6025	0.1097	0.47	0.2258
55	U	0.0175	0.0206	0.02	0.0400
56	C	0	0.0000	0.1075	0.0998
57	C	0	0.0000	0.06	0.1200
58	G	0	0.0000	0	0.0000
59	C	0.0175	0.0350	0.15	0.1152
60	U	0.095	0.1308	0.2525	0.1658
61	C	0.235	0.1515	0.27	0.1036
62	G	0	0.0000	0.0625	0.1121

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
63	U	0.1175	0.1138	0.4375	0.3707
64	C	0.195	0.3900	3.4375	3.5259
65	A	0.255	0.0968	0.39	0.1023
66	C	0.0025	0.0050	0.0225	0.0450
67	U	0.21	0.1322	0.46	0.1966
68	U	0.32	0.3947	2.5425	2.4904
69	A	0.9275	0.2763	0.4775	0.1443
70	U	0.4175	0.1357	0.2425	0.0574
71	C	0	0.0000	0.3075	0.3354
72	C	0	0.0000	0.1225	0.1415
73	U	0.985	0.2158	0.93	0.2634
74	U	1.6125	0.5619	1.88	0.1780
75	C	0.045	0.0661	2.5425	2.4623
76	A	1.2225	0.2311	1.18	0.1579
77	C	0.225	0.1542	0.2975	0.0359
78	U	1.3225	0.2439	0.985	0.2844
79	U	1.7775	0.4864	1.19	0.4274
80	U	1.05	0.3473	0.71	0.2824
81	C	0.0225	0.0450	0.185	0.2138
82	C	0.065	0.1300	2.1075	2.2395
83	A	3.105	2.0855	1.92	1.3176

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
84	G	1.7275	3.2096	0.4725	0.8275
85	A	0.2	0.3611	0.0325	0.0340
86	G	0.0225	0.0450	0	0.0000
87	G	0	0.0000	0.0225	0.0263
88	G	0	0.0000	0.055	0.0640
89	U	0.0175	0.0350	0.0175	0.0287
90	C	0.01	0.0115	0	0.0000
91	C	0.0275	0.0222	0.02	0.0337
92	C	0.1125	0.0900	0.165	0.0645
93	C	0.035	0.0635	0.1875	0.1520
94	C	0	0.0000	0.14	0.2668
95	C	0.0275	0.0550	0.035	0.0700
96	G	0.0575	0.0512	0.01	0.0141
97	C	0	0.0000	0.4325	0.4874
98	A	0.26	0.1208	0.195	0.1256
99	G	0.5425	0.0574	0.295	0.1085
100	A	2.835	0.4871	1.575	0.8021
101	C	0.2475	0.1037	0.1225	0.0737
102	C	0.9425	0.1997	0.6025	0.2253
103	C	1.1525	0.4397	0.7275	0.2926
104	C	0.05	0.0935	0.065	0.0943

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
105	G	0.0225	0.0450	0.0075	0.0150
106	G	0.12	0.0770	0.0375	0.0189
107	U	0.2325	0.0411	0.15	0.0497
108	G	0.01	0.0200	0.0225	0.0171
109	A	0.0825	0.0624	0.0175	0.0236
110	C	0	0.0000	0.0425	0.0613
111	C	0.3375	0.2427	0.185	0.1282
112	C	1.595	1.1651	0.9675	0.2834
113	U	6.165	1.6679	4.0375	1.8953
114	C	2.21	0.5687	2.16	0.3789
115	A	7.73	1.6137	4.8175	2.4106
116	G	0.4625	0.2567	0.195	0.1782
117	G	0.0825	0.0806	0.0075	0.0150
118	U	0.085	0.0311	0.085	0.0785
119	C	0.3	0.3369	0.1275	0.2419
120	G	1.3575	1.0648	0.5925	0.1952
121	G	1.66	0.9926	1.2225	0.9326
122	C	0	0.0000	0.065	0.0507
123	C	0.0075	0.0150	0.0775	0.1297
124	G	0.3525	0.1132	0.245	0.0645
125	A	0.705	0.1710	0.26	0.2394

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
126	C	0	0.0000	0	0.0000
127	U	0.0375	0.0263	0.0225	0.0287
128	G	0.11	0.0594	0.025	0.0300
129	C	0	0.0000	0	0.0000
130	G	0	0.0000	0	0.0000
131	G	0	0.0000	0	0.0000
132	C	0.01	0.0200	0.1175	0.1228
133	A	0.2725	0.3308	1.09	1.0723
134	G	0.4675	0.0556	0.46	0.3692
135	C	2.745	0.7189	1.62	0.6926
136	U	1.2675	0.2919	0.6775	0.4871
137	G	0.325	0.1642	0.045	0.0580
138	G	0.3175	0.2287	0.075	0.0311
139	C	0.14	0.0183	0.115	0.0705
140	G	0.1325	0.0411	0.0425	0.0591
141	C	0.0175	0.0350	0.04	0.0616
142	C	0.0225	0.0450	0.055	0.0656
143	C	0.505	0.2356	0.0875	0.1008
144	G	1.125	0.4014	0.4075	0.2755
145	A	2.51	0.7054	1.23	0.6277
146	A	2.2275	0.3514	1.455	0.3091

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
147	C	0.0625	0.1250	2.7475	2.3170
148	A	0.8375	0.0978	0.4725	0.1567
149	G	0.02	0.0245	0.02	0.0245
150	G	0.01	0.0200	0.0075	0.0150
151	G	0	0.0000	0.0575	0.1084
152	A	0.1275	0.2419	0.175	0.3500
153	C	0.0775	0.1484	1.47	1.5422
154	C	0.435	0.5334	1.6975	1.2422
155	C	0.1875	0.2836	0.96	0.4939
156	U	0.1675	0.0350	0.1875	0.1103
157	C	0.455	0.0968	0.28	0.1068
158	G	0	0.0000	0	0.0000
159	G	0	0.0000	0	0.0000
160	A	0.245	0.1678	0.09	0.0762
161	U	1.51	0.5036	0.5875	0.2877
162	A	0.09	0.0942	0.01	0.0200
163	A	0.0775	0.0580	0.0125	0.0189
164	G	0.015	0.0238	0	0.0000
165	U	0.0275	0.0550	0.005	0.0100
166	G	0.08	0.1233	0.02	0.0337
167	A	0.2125	0.2290	0.0525	0.0984

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
168	C	0.155	0.1792	0.19	0.3668
169	C	0.055	0.0666	0.3175	0.3451
170	C	0.0375	0.0750	0.1075	0.1028
171	U	0.18	0.0698	0.17	0.1230
172	U	0.48	0.1968	0.3875	0.1839
173	G	0.925	0.2671	0.4125	0.3764
174	U	0.6425	0.1550	0.3375	0.1982
175	C	0.005	0.0100	0.03	0.0600
176	U	0	0.0000	0.01	0.0200
177	C	0	0.0000	0	0.0000
178	U	0.1925	0.3850	1.4975	1.8414
179	A	0.3575	0.2734	0.3175	0.2198
180	U	0.125	0.0493	0.13	0.1111
181	U	0.12	0.0716	0.155	0.1401
182	U	0.105	0.1156	0.1075	0.1044
183	C	0.01	0.0200	0.1525	0.2122
184	U	0.2475	0.4685	1.1725	1.1405
185	A	0.1	0.0812	0.1825	0.0888
186	C	0.07	0.1400	0.2875	0.2363
187	U	0.5525	1.0459	1.9525	1.4806
188	A	0.6725	0.3582	0.5475	0.3411

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
189	U	0.175	0.0681	0.16	0.1186
190	U	0.07	0.0688	0.09	0.0693
191	U	0.215	0.3129	0.0675	0.1350
192	G	0	0.0000	0.005	0.0100
193	G	0.0375	0.0750	0	0.0000
194	U	0.3275	0.2885	0.175	0.1733
195	G	0.3325	0.2837	0.1275	0.1500
196	U	0.5025	0.2037	0.2225	0.1895
197	U	0.565	0.2640	0.265	0.2659
198	U	0.47	0.2302	0.19	0.1726
199	G	0.035	0.0700	0.0025	0.0050
200	U	0.045	0.0900	0.0125	0.0250
201	C	0	0.0000	0	0.0000
202	U	0.4	0.0726	0.24	0.0775
203	U	0.555	0.2439	0.3775	0.3485
204	G	0.1025	0.2050	0.0775	0.0838
205	U	0.3725	0.3089	1.15	0.9751
206	A	0.295	0.0619	0.165	0.1196
207	U	0.2975	0.1231	0.1525	0.1382
208	U	0.52	0.4262	0.215	0.2307
209	G	0.8975	0.5289	0.485	0.4871

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4uM Pr77 ^{Gag}	
		Mean	SD	Mean	SD
210	U	0.2325	0.1394	0.0975	0.1228
211	C	0	0.0000	0	0.0000
212	U	0.0125	0.0250	0.0125	0.0250
213	C	0.035	0.0700	0	0.0000
214	U	0.0325	0.0650	0.0425	0.0850
215	U	0.165	0.0971	0.0225	0.0263
216	U	0.24	0.1426	0.075	0.0926
217	C	0.0075	0.0150	0.0075	0.0150
218	U	0.37	0.1257	0.2175	0.0746
219	U	0.385	0.1926	0.2125	0.1795
220	G	0.36	0.1720	0.135	0.1121
221	U	0.165	0.1127	0.12	0.0673
222	C	0.0125	0.0189	0.0025	0.0050
223	U	0.4025	0.1922	0.14	0.1208
224	G	0.375	0.2447	0.12	0.0804
225	G	0.0125	0.0250	0	0.0000
226	C	0	0.0000	0.31	0.3404
227	U	0.67	0.6725	1.735	1.4311
228	A	0.705	0.1482	0.38	0.1534
229	U	0.465	0.1109	0.41	0.1953
230	C	0	0.0000	1.4425	1.1923

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
231	A	0.625	0.7664	0.5675	0.3468
232	U	0.475	0.3579	0.64	0.3851
233	C	0.0525	0.1050	1.6925	1.4735
234	A	0.61	0.8398	0.6375	0.6736
235	C	0	0.0000	2.235	2.0859
236	A	0.9325	0.4055	0.7525	0.1394
237	A	1.7375	0.0946	1.1825	0.3608
238	G	0.49	0.1734	0.1425	0.0750
239	A	0.1825	0.0263	0.1175	0.0222
240	G	0.0125	0.0126	0.02	0.0183
241	C	0.015	0.0300	0.0925	0.1269
242	G	0.3125	0.2134	0.0725	0.1450
243	G	0.2225	0.0818	0.1	0.1244
244	A	0.105	0.2100	0.09	0.1052
245	A	0.2425	0.1550	0.085	0.0387
246	C	0.185	0.2014	0.085	0.1387
247	G	0.3575	0.2891	0.0375	0.0750
248	G	0.175	0.2079	0.215	0.2669
249	A	5.125	0.8055	2.71	1.3284
250	C	0.0425	0.0613	0	0.0000
251	U	0.075	0.1500	0	0.0000

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
252	C	0.0225	0.0386	0.0975	0.1127
253	A	0.22	0.1257	0.085	0.1109
254	C	0	0.0000	0.06	0.0712
255	C	0	0.0000	0.57	0.5126
256	A	0.89	1.7800	0.7175	1.3822
257	U	1.3225	0.1601	2.0875	0.8420
258	A	2.19	0.3813	1.2975	0.5881
259	G	0.0425	0.0568	0.02	0.0245
260	G	0.09	0.1485	0	0.0000
261	G	0.0775	0.0866	0.0325	0.0525
262	A	0.6025	0.2485	0.3275	0.2148
263	G	0.36	0.2765	0.085	0.1377
264	C	0.0275	0.0340	0.015	0.0300
265	U	0.4225	0.2850	0.12	0.1010
266	G	0.295	0.2266	0.08	0.0589
267	C	0	0.0000	2.2475	1.9336
268	A	0.1875	0.3750	0.08	0.1600
269	G	0.4425	0.2900	0.285	0.2111
270	U	0.24	0.0627	0.09	0.1068
271	C	0.08	0.1095	0.2425	0.2417
272	C	0	0.0000	0.205	0.2562

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
273	C	0	0.0000	0.0675	0.0670
274	G	0	0.0000	0.05	0.0560
275	C	0.24	0.1541	0.355	0.3117
276	C	0.27	0.1203	0.245	0.2424
277	U	0.2125	0.0988	0.1275	0.1258
278	A	0.0875	0.0680	0.0925	0.0512
279	C	0.145	0.0881	0.13	0.1643
280	G	0.275	0.1502	0.1825	0.1617
281	G	1.81	0.4346	0.9	0.1920
282	A	2.595	0.4022	1.505	0.3861
283	G	1.3275	0.2313	0.6075	0.1312
284	A	2.1675	0.3051	1.185	0.4243
285	A	1.4675	0.2367	0.89	0.2255
286	G	1.215	0.1537	0.6325	0.2240
287	A	2.2725	0.4923	1.2125	0.3981
288	G	1.1375	0.1797	0.6475	0.1164
289	G	0.0925	0.1352	0.16	0.1657
290	U	0.0725	0.0780	0.0475	0.0660
291	A	0.0175	0.0287	0	0.0000
292	G	0.0225	0.0450	0.0025	0.0050
293	G	0.13	0.1467	0.0225	0.0330

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
294	U	0.62	0.3757	0.485	0.3455
295	U	1.3325	0.4555	1.325	0.7713
296	A	1.7375	0.5063	1.29	0.3034
297	C	1.575	0.4132	1.245	0.3478
298	G	0.1025	0.0574	0.085	0.1226
299	G	0.0925	0.0655	0.0225	0.0222
300	U	0.055	0.0681	0.065	0.0695
301	G	0.0025	0.0050	0.0175	0.0287
302	A	0.1075	0.0830	0.1075	0.0629
303	G	0	0.0000	0.1675	0.2558
304	C	0	0.0000	0.3025	0.3505
305	C	1.1275	1.2550	3.3025	3.0092
306	A	0.7325	0.4670	0.755	0.1603
307	U	1.31	0.4285	0.835	0.1420
308	U	1.485	0.5157	0.7375	0.1452
309	G	0.9275	0.2419	0.3075	0.1396
310	G	0.635	0.0420	0.2275	0.1124
311	A	0.37	0.0455	0.2925	0.2854
312	A	0.2875	0.1103	0.2825	0.2965
313	A	0.32	0.1499	0.3025	0.2562
314	U	0.455	0.1702	0.48	0.0716

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
315	G	0.0375	0.0263	0.0625	0.0645
316	G	0.0075	0.0150	0	0.0000
317	G	0.02	0.0400	0	0.0000
318	G	0	0.0000	0.035	0.0700
319	G	0.2275	0.0946	0.26	0.1278
320	U	0.8225	0.2496	0.7	0.3175
321	C	0.085	0.0129	0.205	0.2316
322	U	0.24	0.0622	0.215	0.1150
323	C	0.1025	0.0222	0.1175	0.1193
324	G	0.0775	0.0263	0.035	0.0700
325	G	0.1775	0.1377	0.2175	0.0359
326	G	0	0.0000	0.005	0.0058
327	C	0	0.0000	0.0475	0.0585
328	U	0.3175	0.2855	0.6775	0.3635
329	C	2.755	2.0722	5.32	4.5072
330	A	0.865	0.3478	0.77	0.1530
331	A	1.285	0.5089	0.815	0.2117
332	A	1.045	0.3747	0.5725	0.0675
333	A	0.955	0.3114	0.525	0.0985
334	G	0.19	0.0876	0.1125	0.0806
335	G	0.0225	0.0386	0.0775	0.0665

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
336	G	0.03	0.0476	0.3	0.2680
337	C	0.49	0.1948	0.63	0.2636
338	A	0.0575	0.1150	0.17	0.0455
339	G	0.0425	0.0850	0.1375	0.1406
340	A	0.0975	0.0591	0.1675	0.0660
341	A	0.3025	0.1357	0.29	0.1246
342	A	0.0325	0.0403	0.01	0.0141
343	C	0.0625	0.0776	0.04	0.0616
344	U	1.095	0.2596	0.965	0.2613
345	C	0.4425	0.1422	0.3575	0.0793
346	U	1.5675	0.4700	1.0475	0.1601
347	U	1.2975	0.2893	1.0075	0.1670
348	U	1.445	0.3484	1.0525	0.1008
349	G	0.0125	0.0250	0.0325	0.0377
350	U	0.1225	0.1415	0.1425	0.0995
351	U	0.37	0.0392	0.2725	0.0419
352	U	0	0.0000	0.2075	0.3824
353	C	0	0.0000	0	0.0000
354	U	0.35	0.1017	0.4475	0.2879
355	G	4.2575	1.1943	2.93	0.7032
356	U	0.3175	0.1350	0.1775	0.1374

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4uM Pr77 ^{Gag}	
		Mean	SD	Mean	SD
357	U	0.235	0.0404	0.1775	0.0714
358	U	0.365	0.0635	0.315	0.0635
359	U	0.5925	0.5519	1.03	0.9862
360	A	0.6225	0.2330	0.7525	0.2109
361	C	2.1675	2.4387	5.6125	4.7502
362	A	0.315	0.1526	0.3425	0.1727
363	A	0.13	0.0787	0.1475	0.1162
364	A	0.04	0.0283	0.0225	0.0330
365	G	0.0025	0.0050	0	0.0000
366	G	0.0125	0.0250	0.3525	0.6719
367	C	0.0225	0.0450	0.0825	0.0960
368	U	0	0.0000	0.0325	0.0395
369	C	0.0525	0.0806	0.095	0.1109
370	C	0.005	0.0100	0.155	0.2901
371	U	0.05	0.0469	0.0475	0.0629
372	C	0.07	0.0804	0.255	0.2655
373	U	0.5525	0.1533	0.2825	0.2909
374	C	3.5025	2.6729	2.4225	0.9164
375	A	4.665	1.2064	3.825	0.9533
376	G	3.11	0.9832	2.4075	0.6737
377	A	3.44	1.2007	2.7	0.7877

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
378	G	0.09	0.1052	0.0775	0.0618
379	A	0.0325	0.0320	0	0.0000
380	G	0	0.0000	0.0175	0.0350
381	G	0	0.0000	0	0.0000
382	G	0	0.0000	0.2125	0.2494
383	G	0	0.0000	0.9725	1.6288
384	U	0	0.0000	0.2375	0.1782
385	C	0	0.0000	0.14	0.1689
386	U	0.11	0.0902	0.0175	0.0236
387	U	1.2425	0.4244	1.1025	0.4847
388	C	2.6025	1.9221	6.9	4.7137
389	A	0.8	0.8568	1.49	0.2660
390	U	0.1275	0.0984	0.2725	0.0608
391	G	0.01	0.0200	0.0375	0.0411
392	U	0.05	0.0627	0.0075	0.0096
393	G	0.0225	0.0450	0	0.0000
394	A	0.0375	0.0556	0.0225	0.0263
395	A	0.1425	0.1130	0.1325	0.1544
396	A	0.52	0.1268	0.4425	0.1544
397	G	0.5575	0.1652	0.4925	0.2004
398	A	0.065	0.0480	0.0475	0.0359

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4uM Pr77 ^{Gag}	
		Mean	SD	Mean	SD
399	G	0.0425	0.0723	0.0075	0.0096
400	A	0.055	0.0656	0.0475	0.0629
401	G	0.02	0.0400	0.11	0.1281
402	U	1.49	2.1801	1.965	1.7148
403	A	0.4625	0.1877	0.455	0.1974
404	G	0.3025	0.0780	0.235	0.1348
405	U	0.34	0.0762	0.3225	0.1081
406	G	1.145	0.1008	1.305	0.2424
407	C	2.085	1.6914	5.94	4.5535
408	A	0.3525	0.3077	0.63	0.1098
409	A	0.515	0.0520	0.5575	0.1571
410	U	1.0225	0.9828	2.0425	1.9120
411	A	0.46	0.0424	0.4475	0.1417
412	G	0.315	0.2055	0.305	0.1396
413	A	0.4825	0.0486	0.3925	0.1473
414	A	0.535	0.0619	0.4175	0.1078
415	U	0.3125	0.0650	0.295	0.0545
416	U	0.41	0.0779	0.36	0.1294
417	U	0.5175	0.0991	0.495	0.1121
418	U	0.575	0.3695	0.8925	0.7032
419	A	0.5975	0.1537	0.495	0.1863

Appendix C (continued)

Nucleotide number & sequence		Mean SHAPE Reactivities (from 4 independent experiments)			
		In the absence of Pr77 ^{Gag}		In the presence of 4 μ M Pr77 ^{Gag}	
		Mean	SD	Mean	SD
420	U	0.375	0.1266	0.485	0.0911
421	C	1.4075	1.5318	4.0525	3.0791
422	A	0.28	0.2282	0.47	0.1055
423	G	0.38	0.2652	0.45	0.1822
424	U	0.1925	0.0907	0.215	0.0656
425	U	0.2075	0.0386	0.18	0.0627
426	U	0.0925	0.0250	0.1075	0.1044
427	C	0	0.0000	0.145	0.1150
428	U	0.085	0.1330	1.3875	1.3277
429	A	0.0775	0.0550	0.13	0.0616
430	A	0.3025	0.1147	0.325	0.1797
431	U	0.5925	0.2955	1.6925	1.5118
432	A	0.395	0.1282	0.4	0.1426

Mean SHAPE reactivities from four experiments in the absence and presence of Pr77^{Gag}. The yellow highlighted nucleotides showed ≥ 1.5 -fold reduction of SHAPE reactivities in the presence of Pr77^{Gag} (p -value ≤ 0.05). The boxed and highlighted nucleotides (from nucleotides 280 to 288) represent the sequence of single stranded purines (ssPurines).

