5-2015

OPTIMAL PACKAGING OF MASON-PFIZER MONKEY VIRUS (MPMV) GENOMIC RNA DEPENDS ON CONSERVED LONG RANGE INTERACTIONS BETWEEN U5 AND GAG COMPLEMENTARY SEQUENCES

Rawan M. Khaleel Kalloush

United Arab Emirates University

College of Medicine and Health Sciences

Department of Microbiology

OPTIMAL PACKAGING OF MASON-PFIZER MONKEY VIRUS (MPMV) GENOMIC RNA DEPENDS ON CONSERVED LONG RANGE INTERACTIONS BETWEEN U5 AND GAG COMPLEMENTARY SEQUENCES

Rawan M. Khaleel Kalloush

This thesis is submitted in partial fulfilment of the requirements for the degree of Master of Medical Sciences (Microbiology & Immunology)

Under the Supervision of Professor Tahir A. Rizvi

May 2015

Declaration of Original Work

I, Rawan M. Khaleel Kalloush, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled "O*ptimal packaging of Mason-Pfizer Monkey Virus (MPMV) genomic RNA depends on conserved long range interactions between U5 and Gag complementary sequences*", hereby, solemnly declare that this thesis is an 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 been previously formed as the basis for the award of any academic degree, diploma or a similar title at this or any other university. The materials borrowed from other sources and included in my thesis have been properly cited and acknowledged.

Student's Signature_____ Date _____

Approval of the Master Thesis

This Master Thesis is approved by the following Examining Committee Members:

1) Advisor: Prof. Tahir A. Rizvi

   Title: Professor

   Department of Microbiology & Immunology

   College of Medicine and Health Sciences, UAE University

   Signature _____ Date _____

2) Member: Dr. Ahemd Al-Marzouqi

   Title: Associate Professor

   Department of Biochemistry

   College of Medicine and Health Sciences, UAE University

   Signature _____ Date _____

3) Member: Dr. Farah Mustafa

   Title: Assistant Professor

   Department of Biochemistry

   College of Medicine and Health Sciences, UAE University

   Signature _____ Date _____

4) Member: Dr. Gulfaraz Khan

Title: Associate Professor

Department of Microbiology & Immunology

College of Medicine and Health Sciences, UAE University

Signature _____ Date _____

5) Member (External Examiner): Prof. Alice Telesnitsky

Title: Professor

Department of Microbiology and Immunology

Institution: University of Michigan Medical School

Signature _____ Date _____

This Master Thesis is accepted by:

Dean of the College of Medicine and Health Sciences: Professor Dennis Tempelton

Signature _____ Date _____

Dean of the College of the Graduate Studies: Professor Nagi T. Wakim

Signature _____ Date _____

Copy ____ of ____

Abstract

Employing a combination of genetic and biochemical approaches, we have recently shown that the 5' end of MPMV genome (from the first nucleotide in R up to 120 nt of Gag) contains a number of sequence and structural motifs important for MPMV gRNA packaging and dimerization. A distinguishing feature of the higher order structure of MPMV packaging signal RNA is two long-range interactions (LRI) between U5 and Gag complementary sequences, LRI-I and LRI-II, which are phylogenetically conserved among different MPMV strains. These LRIs have been suggested to play a role in stabilizing the RNA secondary structure of the 5' UTR sequences that are important for MPMV RNA packaging. The overall RNA secondary structure of this region is further architecturally held together by three other stem loops (SL3, Gag SL1, and Gag SL2) comprising of sequences from distal parts of the 5' UTR and Gag, excluding Gag sequences involved in forming U5-Gag LRIs.

To provide functional evidence for the biological significance of U5-Gag LRIs and the three stem loops to the MPMV life cycle, a series of mutations were introduced in these structural motifs and their effects on MPMV RNA packaging and propagation tested in a genetic trans-complementation assay. Test of LRIs mutants revealed that disrupting the complementary base pairing of the LRI structural motifs affected both gRNA packaging and propagation, confirming their functional role during MPMV life cycle. Our results further revealed that the two LRIs function at different levels. Specifically, LRI-I functions at the secondary structural level, whereas LRI-II functions at both the primary sequence as well as in its native structural context levels. Finally, Mutational analysis of the sequences involved in

forming SL3, Gag SL1, and Gag SL2 revealed that they do not play crucial role at individual levels during MPMV gRNA packaging and propagation. These findings suggest that U5-Gag LRIs have a more important architectural role in stabilizing the structure of the 5' UTR sequences, while the three stem loops (SL3, Gag SL1, and Gag SL2) may have a more secondary role in stabilizing the overall RNA secondary structure, providing a better understanding of the molecular interactions that take place during MPMV gRNA packaging.

Keywords: Retroviruses; Mason-Pfizer monkey virus; RNA secondary structure; RNA packaging; Long range interactions (LRI).

Title and Abstract (in Arabic)

التجميع الأمثل للمادة الوراثية **(RNA)** لفيروس **Mason-Pfizer Monkey Virus (MPMV)** يعتمد على التفاعلات المحفوظة بعيدة المدى بين سلسلتي الحمض الأميني المكونة ل **U5** و **Gag**

لقد أظهرنا مؤخرا من خلال توظيف مزيج من الأساليب الجينية والبيوكيميائية، أن الطرف '5 من جينوم MPMV (من النوكليوتيدات الأولى في R تصل إلى 120 في جين Gag) يحتوي على عدد من الهياكل العليا و سلاسل نيوكليوتيدات ذات أهمية في عمليتي dimerization و التعبئة والتغليف packaging لجينوم MPMV . ومن السمات المميزة للهياكل العليا المكونة لإشارة التعبئة والتغليف في جينوم MPMV RNA هو اثنين من التفاعلات طويلة المدى (LRI) بين U5 وجينوم Gag،هذه التفاعلات ملقبة ب LRI-I وLRI-II، والتي وجد أنها محفوظة phylogenetically بين عدد كبير من سلالات MPMV المختلفة. هذا ما اقترح أن هذه التفاعلات LRIs قد تلعب دورا في استقرار و متانة الهيكل الثانوي ل RNA بداية من الطرف '5 في تسلسل UTR التي تعتبر مهمة لعملية التعبئة والتغليف packaging لجينوم MPMV RNA. كما أن متانة هذا الهيكل الثانوي تعتمد بشكل عام على وجود ثلاثة حلقات جذعية stem loops أخرى تسمى (SL3، Gag SL1 ، Gag SL2) و تتألف من اقتران سلاسل نيوكليوتيدات من أجزاء من الطرف '5 UTR وجينوم Gag، باستثناء الأجزاء من جينوم Gag المشاركة في تشكيل تفاعلات LRIs.

لتقديم أدلة وظيفية للأهمية البيولوجية لتفاعلات LRIs والحلقات الجذعية الثلاثة في دورة الحياة MPMVخاصة عمليتي تكوين ثنائي الجينوم RNA dimerization و التعبئة و التغليف packaging، أدخلت سلسلة من الطفرات في هذه الزخارف الهيكلية الثانوية و تم فحص آثارها على عمليتي تكوين ثنائي الجينوم RNA dimerization و التعبئة و التغليف packaging باستخدام genetic trans-complementation assay. وكشف هذا الاختبار أن الطفرات التي تسببت بتعطيل تكون تفاعلات اقتران هياكل ال LRIs قد أثر سلبيا على كل من عمليتي تغليف الجينوم RNA packaging و التكاثر propagation مؤكدا الدور الوظيفي للهياكل العليا الثانوية LRIs أثناء دورة حياة MPMV. وكشفت النتائج التي توصلنا إليها كذلك إلى أن كل هيكل من LRIs تؤدي وظيفتها بطريقة تختلف عن الأخرى. فعلى وجه التحديد، تؤدي تفاعلات اقتران هيكل LRI-I وظيفتها على المستوى الهيكلي الثانوي، في حين أن وظائف تفاعلات اقتران هيكل LRI-II تؤدي وظيفتها بالاعتماد على التسلسل الأساسي للنوكليوتيدات المشاركة في هذا الاقتران بالإضافة إلى الهيكل الثانوي على حد سواء.

أخيرا، كشف تحليل الطفرات أن تسلسل النوكليوتيدات المشاركة في تكوين الحلقات الجذعية الثلاثة المسماة ب SL3، Gag SL1، و Gag SL2 أنها لا تلعب دورا حاسما على المستوى الفردي خلال عمليتي تغليف الجينوم RNA packaging و التكاثرpropagation ل MPMV gRNA. هذه النتائج تشير إلى أن تفاعلات اقتران النوكليوتيدات من Gag-U5 المكونة ل LRIs لها الدور المعماري الأكثر أهمية في تحقيق الاستقرار في هيكل تسلسل UTR '5، في حين أن الحلقات الجذعية الثلاثة (SL3، Gag SL1، و Gag SL2) قد يكون لها دور أكثر ثانوية في استقرار الهيكل الثانوي ل MPMV RNA بشكل عام، و بالتالي فإن هذه النتائج توفر فهم أفضل للتفاعلات الجزيئية التي تحدث أثناء عملية تغليف MPMV gRNA مما يساعد على استخدامها بشكل آمن في المستقبل في علاج الأمراض الوراثية gene therapy.

ع بارات م ف تاح ية :

ت فاعلات طويل ة امدمل، الامراض الوراث ية ، الهـ يكل الـ ثانوي ، ت غ ل يف الـ ج ي نوم

Acknowledgements

In the name of Allah, Most Gracious, Most Merciful

All praise and glory to Almighty Allah (Subhanahu Wa Taalaa) who gave me courage and patience to carry out this work. Peace and blessing of Allah be upon last Prophet Muhammad (Peace Be upon Him).

My deepest regards goes to my esteemed supervisor prof. Tahir A. Rizvi for giving me the opportunity to follow my dream and enroll into the master's program. He was one of a kind supervisor; he taught me the meaning of dedication to science, how to be organized and to take care of the finest details and always dream big. His continuous support and constructive criticism were very fruitful in shaping my ideas and research.

I would like also to express my sincere gratitude to the passionate scientist, the talented, the patient and on the top of it, the positive energy booster, Dr. Farah Mustafa. I was really lucky to be taught by such a phenomenal teacher whom without I would have never been able to master my experimental techniques. Thank you for always believing in me, appreciating my hard work and for being a role model to Muslim-girls scientists.

Special thanks go to my thesis advisory committee members, Dr. Ahmed Al Marzouqi and Dr. Gulfaraz Khan for their distinguished efforts to build up my scientific foundation in the field and for spending their precious time reading my thesis.

I want to express my deep thanks to my lab family members, Ms. Lizna Ali, my MPMV-mate, who had always listened to my non-stop talks and helped me throughout my experiments. To Ms. Ayesha Jabeen who had hard time between

finishing her work and teaching me when I first joined the lab. To my senior, Suraya Aktar, whom I had learnt a lot through helping her to finalize her PhD dissertation.

To Mr. Arshad who used to set up everything for me on weekends so I spend the minimum time in lab. To the lab ancestors, Ms. Akela Ghazzawi and Ms. Pretty Roni, who were always there to advise me and to draw a smile on my face when things do not work. To the adorable couple, Waqar and Shaima'a, who never fail to make me smile and have been always there for scientific discussions, moral boosts and support.

I am greatly thankful for the sisters I have gained throughout this journey, each one of them have added a special taste to my life here: Dr. Salma, the kind elder sister, thanks for never letting me down. Deena, my journey mate, from classes to midnight experiments, from bachelorette life to motherhood, thanks for being with me through this bumpy ride. Heba, the hyperactive, whom despite all the stress was my daily reminder to smile and keep going. Nour yahfoufi, my classmate and co-presenter who taught me to work hard and be patient. Amal, the multitasking, who taught me never to give up. My hostel sisters: Hasna, Shaima, Haya, Razan, Hanan, Maisam for being there in my best and worst.

My heartfelt gratitude goes to my family, my dad who taught me that the girl's most powerful weapon is her knowledge, my mom, whom without her prayers, emotional support and hot packed meals I would have never reached this stage, and my siblings who were encouraging me overseas and cheering me up with their skype calls.

Keeping the best for the last, I, Rawan Kalloush, wouldn't have been writing this acknowledgement statement today without the believe seeded in my heart by my all times supporter, my husband, Taher, whose consideration, help, listening,

consolation, feedback, patience and enthusiasm made me the researcher who I am today. No thanking word can do your efforts justice. I am thankful for all the help from A to Z especially in formatting and reading my thesis.

Dedication

To my husband, best friend, and all time supporter Taher Abu Seer and to my beloved family

Table of Contents

Title i
Declaration of Original Work ii
Copyright iii
Approval of the Master Thesis iv
Abstract vii
Title and Abstract (in Arabic) ix
Acknowledgement x
Dedication xiii
Table of content xiv
List of tables xvii
List of figures xviii
Chapter 1: Introduction 1
1.1 Retroviruses in a Nutshell 1
1.2 Retroviral Genome Organization 2
1.3 Retrovirus Life Cycle 9
1.3.1 Attachement and Entry 14
1.3.2 Reverse Transcription 15
1.3.3 Integration 19
1.3.4 Transription 23
1.3.5 Retroviral RNA Nuclear Export and Translation 24
1.3.6 Translation and Packaging of Retroviral gRNA 32
1.3.7 Virus Assembly 32
1.3.8 Retroviral Genomic RNA Dimerization 35
1.3.9 Retroviral Genomic RNA Packaging 38
1.3.10 Possible Link Between Retroviral Genomic RNA Packaging and Dimerization 45
1.4 Genomic RNA Packaging and Dimerization Determinants of MPMV 46
1.5 Objectives 55
Chapter 2: Materials and Methods 56

2.1 Genome Nucleotide Numbering System 56
2.2 Plasmid Construction..... 56
2.2.1. MPMV Packaging Construct..... 56
2.2.2. MPMV Transfer Vector: MPMV Packaging Construct..... 56
2.2.3. Envelope Expression Plasmid 59
2.2.4. Transfer Vectors Containing Different Mutations: MPMV Packaging Construct..... 59
2.2.5. Transfer Vectors Containing Substitutions of Uridine (U) in G-U base pairs with Cytosine (C) in the Upper (X-I and X-II) sequence of both LRI-I and LRI-II MPMV Packaging Construct..... 65
2.2.6. Transfer Vectors Containing Deletions of Sequences Involved in U5-Gag LRI-I MPMV Packaging Construct..... 65
2.2.7. Transfer Vectors Containing Substitutions of Sequences Involved in U5-Gag LRI-I 66
2.2.8. Transfer Vectors Containing Substitutions of Sequences (8 nucleotides of the predicted structure) Involved in U5-Gag LRI-II. 67
2.2.9. Transfer Vectors Containing Deletions of Sequences (SHAPE-validated 9 nucleotides) Involved in U5-Gag LRI-II 68
2.2.10. Transfer Vectors Containing Substitutions of Sequences (SHAPE-validated 9 nucleotides) Involved in U5-Gag LRI-II 69
2.2.11. Transfer Vectors Containing Deletions of Sequences Involved in Forming Stem Loops at the 3’ End of the SHAPE-Validated Structure 69
2.3. Three-plasmid *Trans*-complementation Assay 71
2.3.1. Transfection of Producer Cells 74
2.3.2. Infection of Target Cells 75
2.3.3. Luciferase Assay and the Transfection Efficiency 76
2.3.4. Nucleocytoplasmic RNA Fractionation from the Transfected Cells 77
2.3.5. Virus Isolation 78
2.4. RNA Isolation 78
2.5. Reverse Transcriptase PCR (RT-PCR)..... 79

2.6. Real time quantitative PCR (qPCR) for estimation of mutant RNA packaging efficiency 80
2.7. MPMV RNA Secondary Structure Analyses *in silico* 82
2.8. Statistical Analysis 82

Chapter 3: Results 83
3.1. Experimental approach and three plasmid trans-complementation assay to determine MPMV RNA packaging and propagation efficiencies 83
3.2. Development of an MPMV custom-designed qPCR assay for measuring the relative packaging efficiency of mutant viral RNAs 85
3.3. Role of the wobble guanine-uracil (G-U) base-pairing in U5-Gag complementary sequences of LRI-I and LRI-II during MPMV RNA packaging and propagation 89
3.4. Role of the U5-Gag complementary sequences in maintaining LRI-I structure during MPMV RNA packaging and propagation 97
3.5. Role of the U5-Gag complementary sequences in maintaining LRI-II during MPMV RNA packaging and propagation 100
3.6. Role of sequences involved in forming stem loops at the 3' end of the SHAPE-validated structure during MPMV RNA packaging and propagation 108
3.7. Structural analyses of the mutants involving sequences forming the LRIs and stem loops (SL3, Gag SL1, and Gag SL2) at the 3' end of the SHAPE-validated structure 115
3.7.1. Structure-function relationship of the sequences involved in U5-Gag LRI-I 115
3.7.2. Structure-function relationship of the sequences involved in U5-Gag LRI-II 124
3.7.3. Structure-function analysis of the mutations introduced in sequences involved in forming stem loops at the 3' end of the SHAPE-validated structure 128

Chapter 4: Discussion 138
Chapter 5: Conslusion and Recommendations 154

5.1. Conclusion 154

5.2. Recommendations 157

Bibliography..... 158

List of Tables

Table 1: List of deletion/substitution mutations that have been introduced in sequences involved in the formation of LRI-I and LRI-II 63

Table 2:Table outlining the deletion mutations introduced into the Gag sequences . 64

List of Figures

Figure 1.1: Schematic representation of the retrovirus genome 3
Figure 1.2: Schematic representation of a retrovirus particle. 7
Figure 1.3: Schematic representation of different retroviruses genomes..... 10
Figure 1.4: Overview of retroviruses life cycle. 12
Figure 1.5: Schematic representation of the steps involved in reverse transcription. 16
Figure 1.6: Retroviral DNA integration..... 21
Figure 1.7: Nucleo-cytoplasmic export pathways of unspliced or partially spliced retroviral RNAs 26
Figure 1.8: Schematic representation of ribosomal frameshifting phenomenon in retroviruses to express relative amounts of Gag and Pol proteins. 29
Figure 1.9: Translational control of retroviral genomic RNA packaging..... 33
Figure 1.10: Model depiction the steps leading to dimerization and packaging processes among retroviruses..... 36
Figure 1.11: Model depicting specific and selective unspliced genomic RNA packaging among retroviruses 39
Figure 1.12: Summary of the packaging and dimerization determinants of a number of different retroviruses 42
Figure 1.13: Comparison of the predicted and the SHAPE (selective 2' hydroxyl acylation analyzed by primer extension)-validated structure of MPMV packaging signal RNA. 47
Figure 1.14: Mfold predictions of the 5' end of genomic RNA of five different strains of MPMV showing phylogenetically conserved LRIs involving U5 and Gag sequences highlighted in red boxes. 51
Figure 1.15: Sequence alignment of 5' end of genomic RNA from five different isolates of MPMV using Clustal W. 53
Figure 2.1: Schematic representation of MPMV genome, its sub-genomic transfer vector, the packaging construct and the envelope expression plasmid. 57

Figure 2.2: Schematic illustration of the splice overlap extension (SOE) PCR strategy used to introduce mutations. 60
Figure 2.3: Illustration of the 3-plasmid *trans* complementation assay..... 72
Figure 3.1: Validation of the custom-made MPMV real time PCR assay 87
Figure 3.2: Design and test of the deletion/substitution mutations introduced into the complementary U5-Gag LRI-I sequences. 90
Figure 3.3: Effect of the deletion/substitution mutations introduced into U5-Gag LRI-I on MPMV gRNA packaging and propagation. 95
Figure 3.4: Design and test of the deletion/substitution mutations introduced into the U5-Gag LRI-II sequences 101
Figure 3.5: Effect of the deletion/substitution mutations introduced into U5-Gag LRI-II in MPMV gRNA packaging and propagation. (A)..... 104
Figure 3.6: Design and test of the systematic deletion mutations introduced in sequences at the 3' end of the SHAPE-validated structure forming SL3, Gag SL1 and Gag SL2. 109
Figure 3.7: Effect of deletion mutations introduced into sequences at the 3' end of the SHAPE-validated structure forming SL3, Gag SL1 and Gag SL2 on RNA packaging and propagation 112
Figure 3.8: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing stabilizing mutations in the U5 sequences of the LRI-I and LRI-II. 116
Figure 3.9: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing systematic deletions in the LRI-I complementary stem. 119
Figure 3.10: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs disrupting the complementarity of the native LRI-I and recreating an artificial LRI. 122
Figure 3.11: Mfold secondary structure predictions of the LRI-II mutant transfer vector packaging signal RNAs. 126
Figure 3.12: Mfold secondary structure predictions of the LRI-II mutant transfer vector packaging signal RNAs. 129

Figure 3.13: Mfold secondary structure predictions of the LRI-II mutant transfer vector packaging signal RNAs 131
Figure 3.14: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing individual deletions of SL3, Gag SL1, and Gag SL2..... 134
Figure 3.15: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing multiple deletions of SL3, Gag SL1, and Gag SL2. 136
Figure 4.1: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing mutations in LRI-I. 141
Figure 4.2: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing mutations in LRI-II. 145
Figure 4.3: Mfold secondary structure predictions of the hypothetical mutant transfer vector packaging signal RNAs containing stem loop substitution mutations in the packaging signal. 149

Chapter 1: Introduction

1.1. Retroviruses in a Nutshell

Since the early 1900's, substantial insights pertaining a new family of viruses have been introduced to the scientific community. These viruses received more attention when they were implicated as the causative agents for leukemia in chicken by Peyton Rous in 1911. Later on in 1970, the revolutionary discovery of the unique enzyme, reverse transcriptase (RT) from this group of viruses by Howard Temin and David Baltimore shattered the central dogma of molecular biology. The presence of RT in these viruses was able to reverse the flow of genetic information and thus they were named as "Retroviruses". In brief, retroviruses are a large group of viruses belonging to the *Retroviridae* family. They are enveloped viruses and their single stranded RNA genome of positive polarity is linear, non-segmented, and approximately 7-12 kilobases (kb) in size. Over the years, retroviruses have been associated with a variety of diseases, mainly cancers and immunological deficiencies, both in animals and humans.

As their name indicates, retroviruses are characterized by their replication strategy which involves the activity of the virally-encoded RT enzyme which reverse transcribes the viral RNA into double stranded DNA. This viral DNA is then integrated into the host chromosomes by the virally-encoded enzyme, integrase (IN), allowing its stable expression and making it part of the host genome. This is why retroviruses are retroviral for human gene therapy trials (reviewed in Verma and Weitzman, 2005).

1.2. Retroviral Genome Organization

Among the various types of RNA viruses, retroviruses are exclusively known for packaging two copies of their RNA genome (diploid) as a dimer linked at the 5' end with non-covalent interactions (reviewed in Pedersen and Duch, 2006). Retroviral genomic RNA (gRNA) harbors two types of sequences: non-coding sequences that facilitate and control many essential steps during the virus replication (reviewed in Kuzembayeva et al., 2014), and coding sequences that encode for the structural and enzymatic proteins of the virus (Figure 1.1). A typical retrovirus genomic RNA contains a unique sequence at the 5' end called U5 and one at the 3' end called U3. Both the U5 and U3 are juxtaposed with the same repeated sequence "R" forming the *cis*- controlling elements of the viral genome. Retroviral genome organization changes when the RNA genome is reverse transcribed into DNA during which the U5 is copied to the 3' end of the genome and the U3 sequence gets copied to the 5' end, forming a long terminal repeat (LTR) at both ends of the viral DNA (Figure 1.1).

Next to the U5 and U3 sequences is a stretch of non-coding sequences at the end called 5' end and sometimes at the 3' end called untranslated region (UTR) flanking the protein-coding sequences. The 5' UTR is characterized by the presence of primer binding site (PBS), whereas the 3' UTR is characterized by the presence of polypurine tract (PPT) both of which play pivotal roles during reverse transcription of the RNA genome. Over the years, it has been shown that that sequences within the 5' UTR have an essential role in the viral gRNA (gRNA) dimerization and packaging (reviewed in Paillart et al., 2004; Russell et al., 2004; D'Souza &

Figure 1.1: Schematic representation of the retrovirus genome.

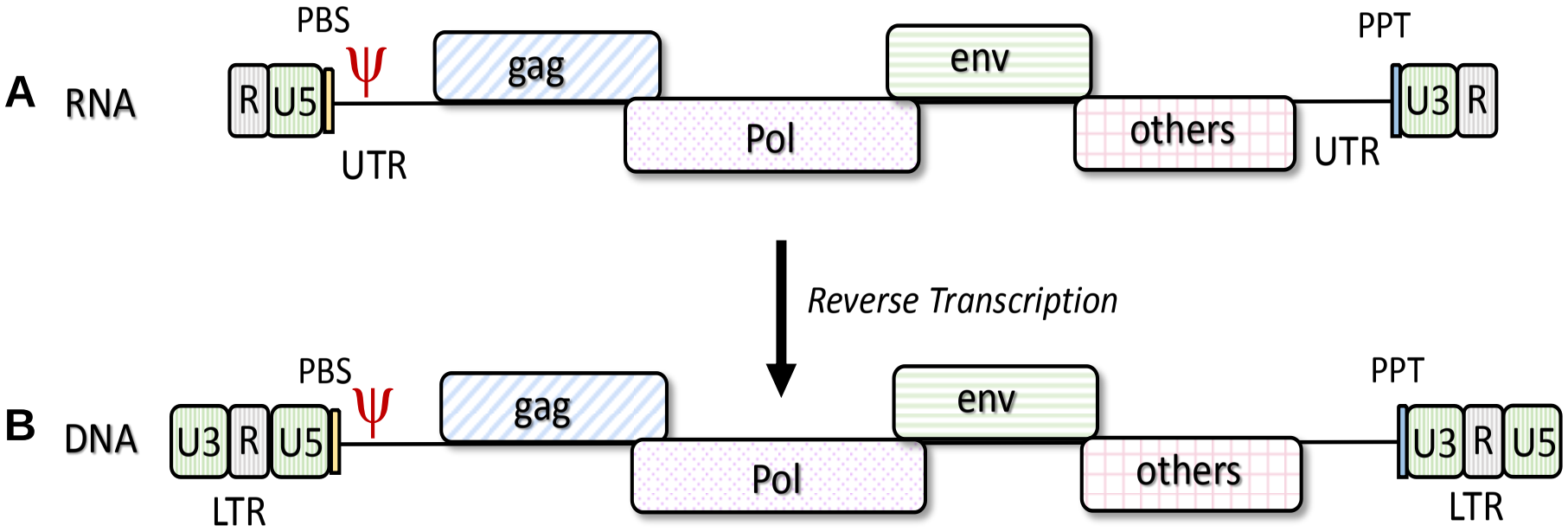


Figure 1.1. Schematic representation of the retrovirus genome. (A) The retrovirus genomic RNA contains both coding and non-coding sequences. The non-coding sequences are located at the 5' and 3' ends of the genome. These sequences contain essential recognition sites for DNA synthesis, integration and polyadenylation. The primer binding site (PBS) and the polypurine tract (PPT) sequences play an important role during reverse transcription. The coding sequences of the viral genome comprises *gag* gene which encodes group-specific structural antigens, the *pol* gene which encodes for the reverse transcriptase, integrase and protease enzymes, and the *envelope (env)* gene which encodes for the envelope structural proteins. As the complexity of the retrovirus genome increases, it acquires additional genes that help in the completion of its life cycle. (B) The viral DNA genome after being reversed transcribed from the RNA template. It acquires an extra copy of the unique 3' sequence at its 5' terminal (U3) end and an extra copy of the unique 5' sequences at its 3' terminal end (U5), forming an identical flanking sequence called long terminal repeat (LTR).

Summers, 2005; Johnson & Telenitsky, 2010; Lever, 2007; Lu et al., 2011; Kuzembayeva et al., 2014).

The structural genes of the virus are encoded in-between these controlling elements. For example, a simple retrovirus like Mason-Pfizer monkey virus (MPMV) harbors three open reading frames (ORFs): *gag*, *pol,* and *envelope* (*env;* Figure 1.1). The *gag* gene encodes for group specific antigen (Gag) precursor structural polyprotein which, upon maturation of the virus particle, is cleaved by the viral protease into structural proteins; capsid (CA), matrix (MA), and nucleocapsid (NC) (Figure 1.2). The capsid (CA) proteins form the viral core which contains the viral gRNA as well as other viral enzymes and proteins, while the matrix (MA) mediates the stable association of the envelope proteins into the virus particle. The role of the nucleocapsid (NC) protein is to selectively bind the unspliced gRNA, facilitating its packaging into the virus particle. The *pol* ORF encodes for the viral enzymatic proteins, namely RT, IN and protease (PR; Figure 1.2). As mentioned above, RT mediates the reverse transcription of the RNA genome into DNA, IN facilitates the integration of the reverse transcribed DNA genome into the host chromosomes, and PR is responsible for cleaving the Gag/Pol polyproteins into mature structural and enzymatic proteins. The *env* gene encodes for the envelope precursor protein which gets cleaved by cellular proteases into the transmembrane (TM) and surface (SU) glycoproteins (gps) which are then incorporated into the phospholipid membrane that is acquired from the host cell membrane through the process of budding (Figure 1.2). Throughout the evolutionary process, as the complexity of the

virus increases, it sometimes becomes necessary for the virus to acquire more genes to fulfill its survival needs in a new and perhaps more complex hosts. For example, human immunodeficiency virus type-1 (HIV-1) has been shown to have six additional and/or accessory genes (*vif*, *vpr*, *vpu* and *nef*, *tat*, and *rev*; Figure 1.3; reviewed in Balvay et al., 2007; Cullen & Green, 1990; Pavlakis & Felber, 1990).

Figure 1.2: Schematic representation of a retrovirus particle.

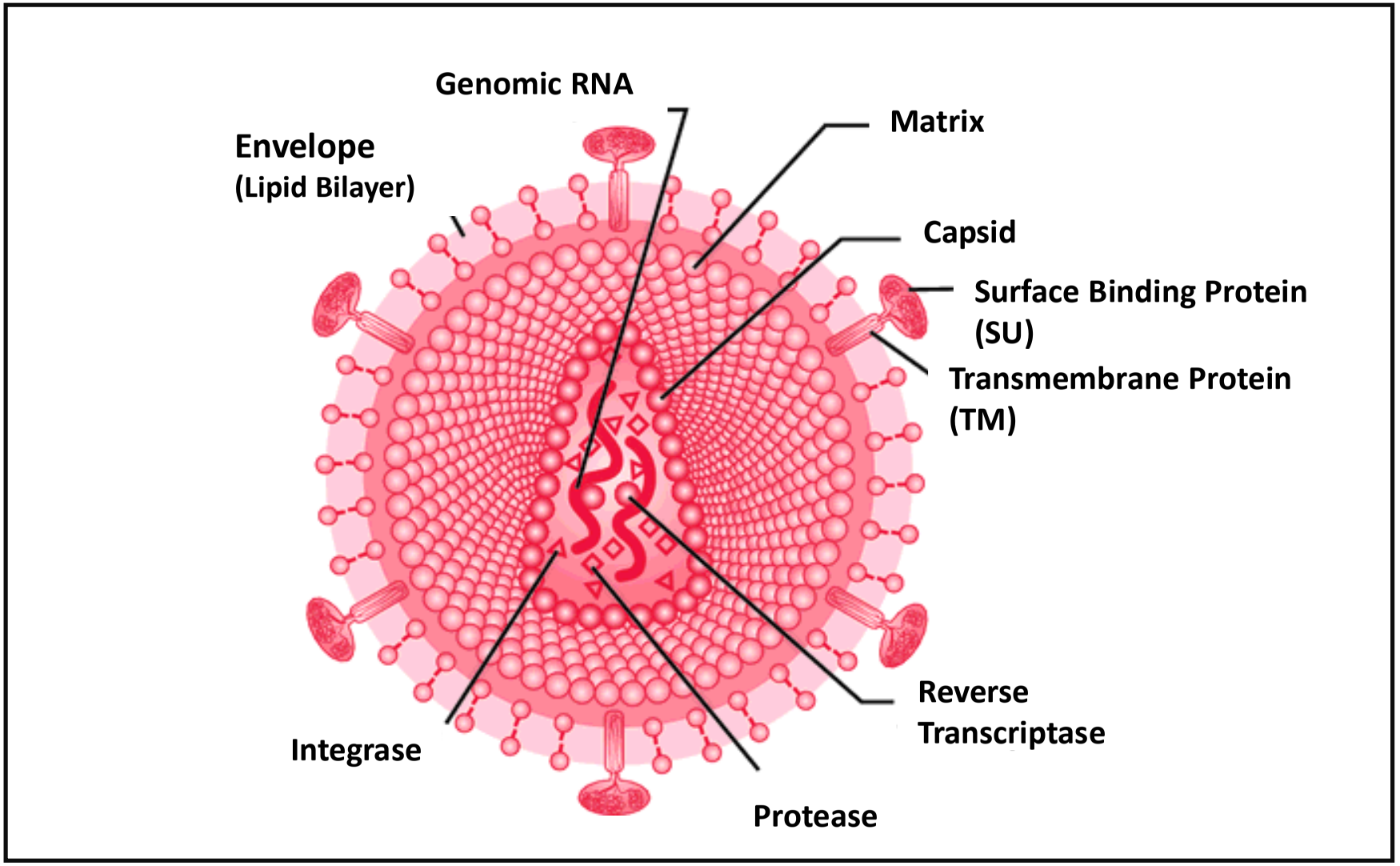


Figure 1.2. Schematic representation of a retrovirus particle. The retroviral genome consists of a diploid single stranded RNA wrapped by structural proteins encoded by the *gag* gene, including the nucleocapsid, capsid and matrix proteins. In addition, non-structural enzymatic proteins encoded by the *pol* gene (reverse transcriptase, protease and integrase) are also packaged into the virus particle. During the process of budding, the virus acquires lipid bilayer from the host into which the virally-encoded envelope glycoproteins (TM and SU) are anchored.

1.3. Retrovirus Life Cycle:

The life cycle of a retrovirus can be divided into two phases, namely extracellular and intracellular (Figure 1.4). In the extracellular phase, the virus recognizes and binds to its host cell via ligand-receptor interaction between the virus envelope proteins and its specific receptor on the host cell membrane (Figure 1.4A). Upon binding, the virus membrane fuses with the host cell plasma membrane thus initiating the intracellular phase of the life cycle. Binding may also require a co-receptor to strengthen the binding and bring the two membranes closer, thus facilitating the fusion. Once inside the cell, RT starts reverse transcription of the gRNA and converting it into a double stranded DNA. The reverse transcribed viral DNA then moves to the nucleus, where IN mediates the process of its integration into the host genome. At this stage, the integrated retroviral DNA is called a "provirus". Now the double stranded DNA is transcribed by the cellular transcription machinery into two different categories RNAs, namely spliced and unspliced RNAs. Both of these types of RNAs are then exported out of the nucleus to the cytoplasm for two different functions (Figure 1.4B). The spliced env mRNA is translated to make the envelope glycoproteins. On the other hand, the unspliced gRNA plays dual functions: 1) it acts as a template for translation of Gag/Pol precursor polyproteins, and 2) it acts as the genome for incorporation into the assembling virus particles.

Figure 1.3.:Schematic representation of different retroviruses genomes

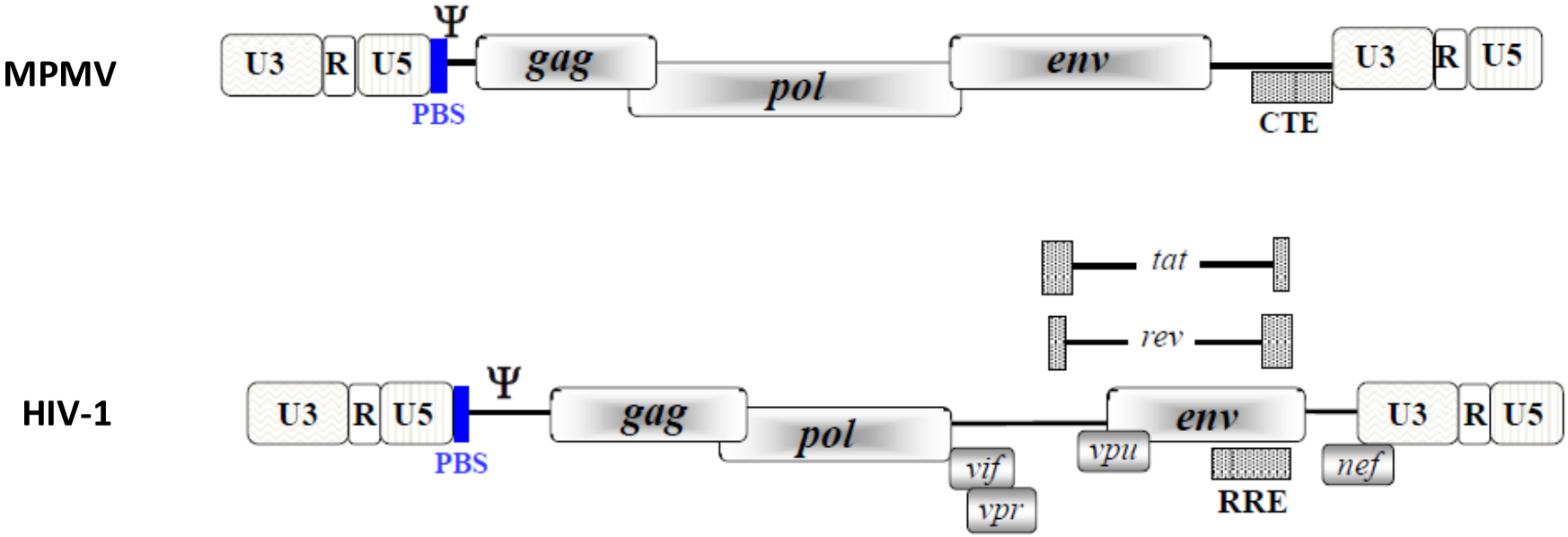


Figure 1.3. Schematic representation of different retroviruses genomes. Typical simple retroviruses like Mason –Pfizer Monkey Virus (MPMV) has three canonical coding genes: *gag*, *pol* and *env.* They maintain a constitutive transport element (CTE) at the 3' untranslated region (UTR) which helps in the nucleo-cytoplasmic export of their RNAs. Complex retroviruses like HIV-1 code for additional accessory and regulatory proteins. It also contains the Rev protein which binds to specific sequences called Rev-Responsive Element (REE) to allow its unspliced and singly-spliced RNAs to get exported from the nucleus.

Figure 1.4: Overview of retroviruses life cycle

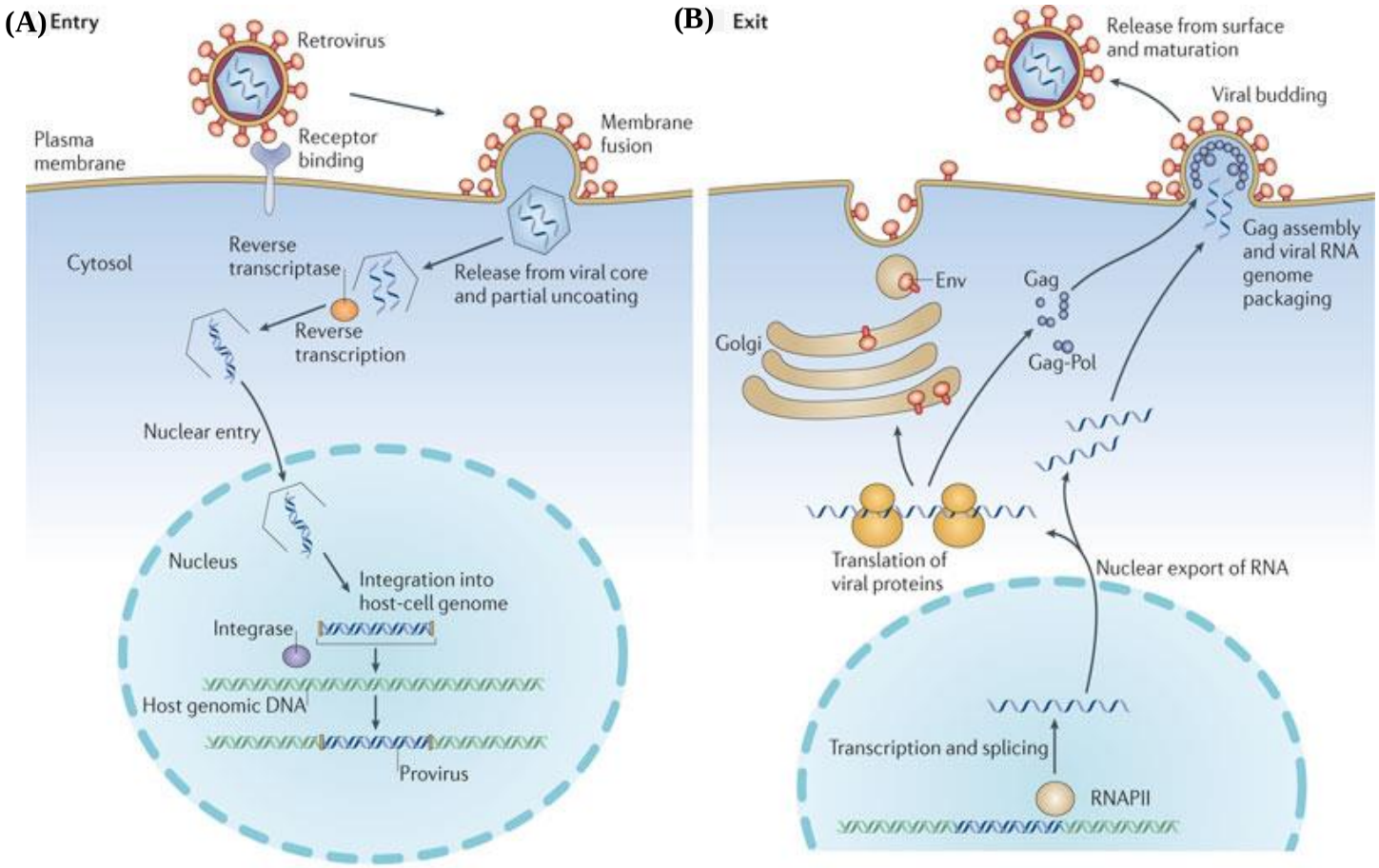


Figure 1.4. Overview of retroviruses life cycle. Retroviruses recognize and bind to specific host cells via their envelope proteins. After entry, the genomic RNA gets reverse transcribed to form double stranded DNA which gets imported into the nucleus and integrates into the host genome to form the provirus. Once integrated, the viral DNA gets transcribed using the host transcription machinery. Transcription results in at least two different types of RNAs, full-length unspliced and singly/multiply spliced RNA, both of which get exported to the cytoplasm where the unspliced genomic RNA express the Gag/Pol proteins and the spliced RNA expresses the envelope and other proteins during the process of translation. Finally, the full length unspliced genomic RNA gets packaged into the budding virus particle. Figure adapted from Stoye, 2012.

The Gag/Pol polyproteins resulting from the expression of the unspliced RNA are cleaved by a virally-encoded protease to give CA, MA, and NC proteins, whereas the envelope precursor glycoprotein is cleaved by cellular proteases into SU and TM proteins. These structural proteins then assemble to form the virus particle (Figure 1.4B). The unspliced full-length gRNA interacts with the NC portion of the Gag protein near the plasma membrane and the two copies of the gRNAs are encapsidated as a dimer. While budding from the host cell, the virus particles acquire part of the host-plasma membrane along with the incorporated envelope glycoproteins. After being released, the virus particle matures and becomes infectious by further proteolytic cleavage and gRNA dimer stabilization (Bender et al., 1978; Murti et al., 1981; Fu and Rein, 1993).

The following sections discuss the retrovirus replication steps in more details.

1.3.1. Attachment and Entry

In order for the retrovirus to establish intracellular aspect of its life cycle inside the host cell, it should attach to the cell plasma membrane, which is the rate limiting step of the infection process. Attachment to the target cell membrane is a distinct step from the entry and uptake of the retrovirus. This process requires an entry receptor and may require co-receptors to assist with the entry process. For example, in HIV-1, in addition to the CD4 receptor, another co-receptor is required that differ in different cell types. The T-cells have been shown to use CXCR4 while macrophage-tropic HIV-1 viruses use CCR5 as the co-receptor. (Reviewed in Clapham PR., and McKnight A., 2001). The binding itself doesn't allow for the virus entry, however, it is the viral envelope fusion with the host cell plasma membrane

that facilitates its entry. Retroviral envelope glycoproteins are arranged as oligomers of the globular SU domain anchored to the viral membrane by the TM domain. Upon attachment it has been shown that the oligomeric nature of the envelope glycoproteins favors the clustering of cell receptors where soluble monomeric glycoproteins can each interact with the receptor to allow for an efficient entry. Such has been observed when the monomeric mouse mammary tumor virus (MMTV) gp51 binding was found to be insufficient to promote receptor internalization (we should cite the appropriate reference here). In addition to the clustering strategy, sequential binding of envelope glycoproteins may result in receptor aggregation and thus competent attachment. Conformational changes in the SU domain upon binding expose a hydrophobic peptide in the TM domain that is essential for catalyzing the fusion between viral and cellular membranes. Even though the steps involved in the fusion process are still not fully understood, it seems that the end step includes the formation of a fusion-pore (Muno z-Barroso et al., 1998). As the virus envelope gets fused with the cellular membrane, the virus releases its internal component into the cell cytoplasm where it becomes ready for the first step of the intracellular phase of the virus life cycle that is the reverse transcription.

1.3.2. Reverse Transcription

A unique characteristic of retroviruses is the non-conventional way they continue their life cycle, exploiting the exceptional activity of RT which can convert the viral diploid single stranded RNA genome into double stranded DNA (Figure 1.5A). RT is capable of “reverse” replication because of its two distinct activities: RNA dependent DNA polymerase (RDDP) and DNA dependent DNA polymerase

Figure 1.5: Schematic representation of the steps involved in reverse transcription.

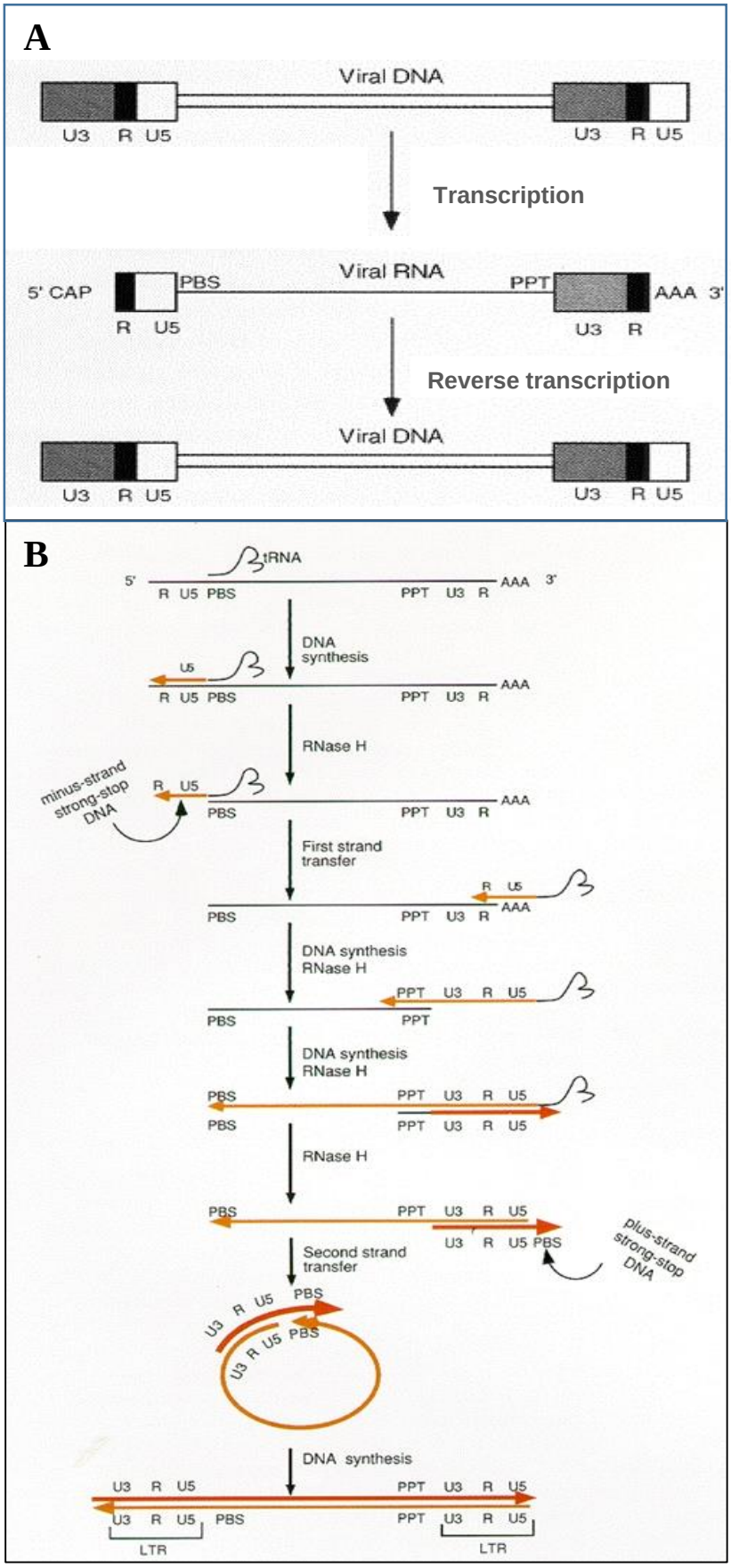


Figure 1.5. Schematic representation of the steps involved in reverse transcription. **(A)** Schematic representation of the flow of the genetic information of a retrovirus genome . **(B)** The reverse transcription process is mediated by a virally encoded enzyme called reverse transcriptase (RT) and involves two transfer steps. It begins with binding of tRNA primer to the primer binding site (PBS) at the 5' end. DNA synthesis begins till the end of the template is reached. This initiates the first of the two transfer steps in the reverse transcription process.. The first intermediate (-) strand DNA synthesized at the 5' end is transferred to the 3' end during (-) strong-stop template switch. Then RNA template from the DNA:RNA-hybrid gets degraded by RNase H as the (-) strand synthesis proceeds, except polypurine tract (PPT) which is RNase resistant and functions as a primer for (+) strand DNA synthesis. During the second transfer, the (+) strand product that results from synthesis into the tRNA primer is transferred to the 5' end of nascent (-) strand DNA during (+) strand strong stop transfer. The resulting DNA contains two terminal repeats (LTRs) each consisting of the terminal block of U3-R-U5. Figure adapted from Telestinsky & Goff , Reverse Transcriptase and the Generation of Retroviral DNA. In: Coffin JM, Hughes SH, Varmus HE, editors. Retroviruses.,1997.

(DDDP). RT also contains a nuclease activity termed Ribonuclease H (RNase H) that degrades the RNA strand from the RNA: DNA duplexes. A role of some of the viral proteins like the nucleocapsid protein in the reverse transcription process has also been shown to be important for an efficient RT process (Levin et al., 2010). The reverse transcription process initiates once the virus enters the host cell, except for the genus *Spumavirinae* in which a mature virion was found to already contain reverse transcribed DNA in addition to RNA (Yu et al., 1996). The key players in the reverse transcription process that act in *cis* are PBS, PPT, and the R sequences on the termini of retroviral gRNA. Various steps involved during the course of reverse transcription process are delineated in (Figure 1.5) and described in detail below.

Once the virus capsid is released into the host cell cytoplasm, a cell-derived tRNA primer binds to its complimentary sequence on retroviral PBS while still inside the virion capsid. Next, the RDDP activity of RT starts making a copy of the viral DNA from the PBS and stops at the 5' end followed by degradation of the RNA portion of the RNA: DNA hybrid by the RNase H activity of RT. This results in the formation of the minus strand strong stop DNA (-sss DNA) (Figure 1.5B). Now RT is ready for the minus strong stop template switch where it jumps to the 3' end of the gRNA and anneals to the redundant terminal R sequences (Figure 1.5B). This primes the synthesis of the minus strand DNA which takes place as the RNA template is being degraded by RNAse H. To initiate the synthesis the DNA plus strand, another primer at the 3' end is needed for the priming purposes. Since PPT sequence is resistant to the hydrolysis by the RNAse H activity of RT, therefore it plays the role of a primer for the synthesis of plus strand DNA (Figure 1.5B). At this stage the RT utilizes its DDDP activity and the minus strand DNA as a template. The synthesis proceeds upto the 5' end of the genome and stops at the first nucleotide

at the 3’ end of the tRNA sequence that anneal to the PBS generating a new PBS sequence. At this point the tRNA is digested by the RNase H activity of the RT and the resulting +sssDNA will have a copy of the 3’ end U3 at the 5’ end and a copy of the 5’ end U5 at the 3’ end. During reverse transcription, recombination among retroviruses can occur during the synthesis of both minus and plus strand DNA synthesis (Hu and Temin, 1990). The recombination between retroviral RNAs during reverse transcription is considered a serious problem, especially if the cell is co-infected with two different strains of viruses, where the recombination could result in the formation of a new virus of unknown biology and pathogenic potential. Emergence of such new variants among retroviruses poses a major threat towards the development of antiviral agents and viral vaccines. Therefore, it is not at all surprising that the first approved anti-HIV drug, AZT, targeted RT and of the 26 currently approved drugs to treat HIV-1 infections, 14 are RT inhibitors (Hu and Hughes, 2012).

1.3.3. Integration

Following reverse transcription, the newly formed retroviral DNA must be imported to the host cell nucleus for integration into the host genome in order to continue its life cycle. The integration process is a hallmark of retroviruses making them good candidates for gene therapy since they can to maintain long-term expression of the integrated therapeutic gene.

The integration process is primarily mediated by the virally encoded IN enzyme. IN, with the help of some other viral proteins that differ from one group of retroviruses to the other, forms the pre-integration complex (PIC). Migration of PIC

to the nucleus occurs in two different ways. The first route is observed in oncoretroviruses, viruses that require the cell to be in the dividing state so that the newly reverse transcribed DNA can gain access to the nuclear DNA when the nuclear membrane is disassembled during mitosis. However, since HIV-1 in particular and the lentiviruses in general can infect nondividing cells, such a strategy cannot be exploited by HIV-1 or other lentiviruses. Consequently, there must be a second route through which the HIV-1 PICs can enter the nucleus. It has been shown that HIV-1 PICs can enter the nucleus during the interphase in an energy-dependent import process through the nuclear pore apparently via nuclear localization signals (NLS) found in the viral MA and the Vpr proteins (Bukrinsky et al., 1993).

Once inside the nucleus, integration into the host chromosome involves a series of nucleophillic attacks, the first of which removes the terminal 2 bases from the 3′ ends of the LTRs that are recognized by viral IN (Fig 1.6), whereas the second attack inserts the viral DNA into the host genome. The site of integration is recognized by the viral IN and is characterized by duplication of the host target site following integration. In addition to the viral proteins, at least two host cellular proteins, high mobility group [HMG-I(Y)] and barrier to autointegration factor [BAF], have been shown to increase the efficiency of the integration reaction (Van Maele et al., 2006). The viral proteins that are used for integration might have other roles like directing the assembly of transcription factors or other cellular proteins onto the newly integrated provirus. The integration process of retroviral DNA into the host genome while being a critical step for the virus to continue its life cycle, at the same time, it poses a great concern to the scientific community when it comes to

Figure 1.6. Retroviral DNA integration

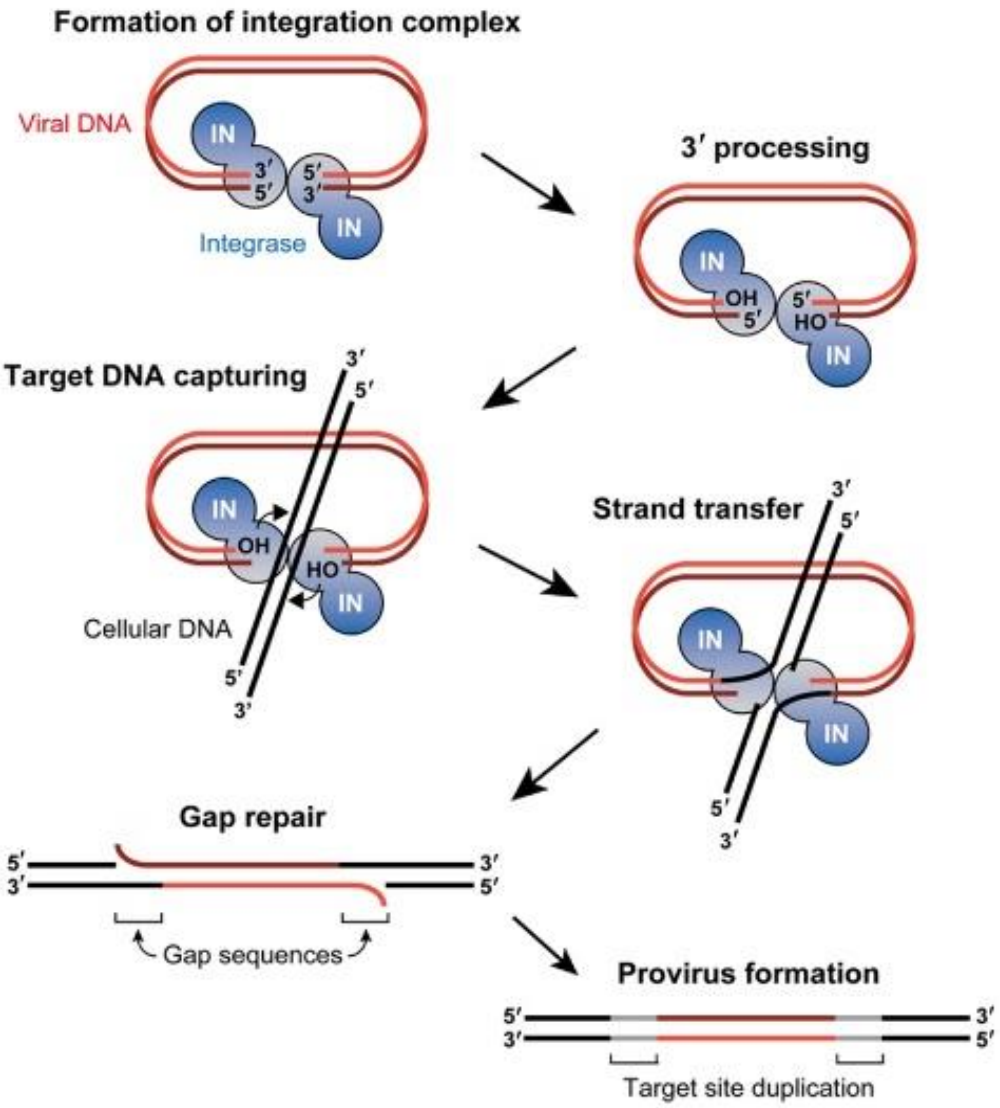


FIGURE 1.6. Retroviral DNA integration. The virally encoded integrase enzyme recognizes and nicks both ends of the viral DNA. It removes 2 or 3 nucleotides from each of the 3' ends of the double stranded DNA. It also introduces random staggered nicks in the host DNA via the exposed hydroxy (OH) groups on the viral DNA ends. This results in simultaneous joining of the 3′ end of the viral DNA and the 5′ end of the target DNA. This results in an integrated retroviral genome called a "provirus". A consequence of this process is duplication of the "target site" which is signature of a successful integration process. Figure adapted from Yasutsugu et al.,2012.

the use of viral vectors in gene therapy since their integration could cause oncogenesis. Retroviral integration is a random process and lentiviral vectors integrate more randomly than the viral vectors throughout the human genome; however, it has been shown that retroviral vectors favor promoters and enhancer regions as preferred integration sites (Schroder et al., 2002). Swapping IN between closely related retroviruses has been shown to change the integration pattern suggesting that the catalytic core of the IN play a role in choosing the preferred site of insertion (Shibagaki et al., 1997). Recently, a lot work is being undertaken to achieve site directed retroviral integration which could possibly avoid the risk of oncogenesis by retroviral vector insertional activation (reviewed in Youngsuk et al., 2011).

1.3.4. Transcription

Using the cellular transcription machinery, the integrated proviral DNA is transcribed along with the cellular DNA to generate mRNAs needed for viral protein production. Specifically, transcription of the proviral DNA produces full length unspliced gRNA that plays a dual role (as mentioned earlier): as an mRNA for the translation of the Gag/Pol polyproteins, and also as a substrate to be packaged into the newly assembled virus particle as gRNA. It has been shown that cellular RNA Polymerase II is responsible for proviral DNA transcription since it has *cis*-acting sequences that direct RNA polymerase II in the production of mRNAs. The polymerase II processivity has shown to be enhanced by the binding of some viral transcriptional transactivators, such as Tat protein in the case of HIV-1 where it binds to TAR on the viral mRNA, enhancing the binding of transcription factors to DNA (reviewed in Jeang, 1999). The resulting transcribed viral RNA experiences

the same processing events that cellular RNAs go through, including RNA splicing, to yield different lengths of unspliced and spliced mRNAs. For example, in all retroviruses, singly spliced mRNA is used for translating the envelope glycoproteins.

1.3.5. Retroviral RNA Nuclear Export and Translation

After being transcribed, processed, and spliced, the retroviral RNA is ready to be exported from the nucleus to the cytoplasm. However, the splicing machinery inhibits the unspliced or singly spliced RNA, from being exported out of the nucleus, a phenomenon called nuclear retention. To overcome such nuclear retention, retroviruses have evolved mechanisms that allow them to successfully export, especially the unspliced RNAs, out of the nucleus. One way by which retroviruses overcome nuclear retention is by the presence of a regulatory pathway in which *cis*-acting RNA structural elements found at the 3' end of retroviral gRNA interact either with the virally encoded or cellular proteins to facilitate the successful export of the unspliced gRNA. One example of such a structure is the constitutive transport element (CTE) in the case of the simple retrovirus MPMV (Figure 1.7; Bray et al., 1994; Rizvi et al., 1996a and 1996b; Ernst et al., 1997). MPMV CTE interacts with cellularly-encoded nuclear export factor 1 (NXF1) or Tap that facilitates export and expression of CTE-containing mRNAs (reviewed in Cochrane et al., 2006; Swanson and Malim, 2006). In complex retroviruses, a virally-encoded protein, such as Rev in the case of HIV-1 interacts with an RNA element at the 3' end called Rev responsive element (RRE), to allow nuclear export of viral messages (reviewed in Groom et al., 2009; Figure 1.7). Similarly, MMTV has evolved the Rem/RemRE regulatory pathway in which the Rem protein interacts with *cis*-acting RemRE sequence located at the 3' end of the viral transcripts (Figure 1.7). Recently, it has

been shown that gamma retroviruses also contain a post transcription element (PTE), a *cis*-acting sequence which overlaps the *pro-pol* open reading frame and is responsible for the regulation of *gag* gene expression (Pilkington et al., 2014). The PTE function resembles that of the CTE function in the case of MPMV, but the location is different from the usual one at the 3' end.

Once successfully exported out of the nucleus, the retroviral RNA is translated into viral proteins, a process that is mediated by the host translational machinery (reviewed in Balvay et al., 2007). The Gag/Pol viral proteins are translated from the unspliced gRNA by the free cellular polyribosomes, whereas the envelope glycoproteins are translated from the spliced mRNA by the rough endoplasmic reticulum-bound ribosomes. The translated envelope precursor protein undergoes post translational modifications. It undergoes N-glycosylation in the lumen of the endoplasmic reticulum and is cleaved at its SU-TM border by host proteases within the Golgi apparatus as it is transported for incorporation into the host cell membrane.

Figure 1.7: Nucleo-cytoplasmic export pathways of unspliced or partially spliced retroviral RNAs

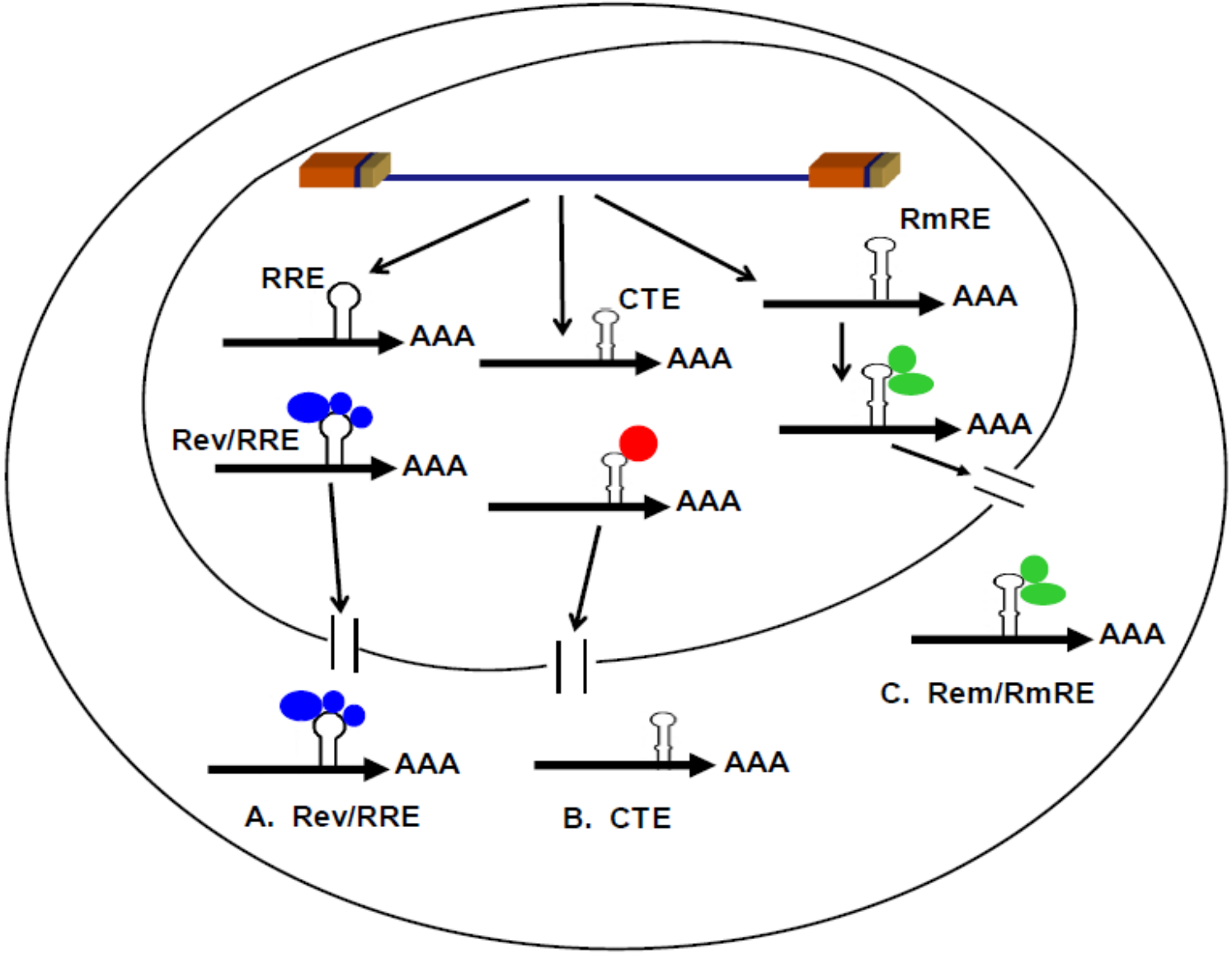


Figure 1.7. Nucleo-cytoplasmic export pathways of unspliced or partially spliced retroviral RNAs. Sufficient amounts of retroviral unspliced RNA needs to be exported to the cytoplasm to get translated into the structural and enzymatic proteins needed for virus survival as well as to be packaged as genomic RNA into the newly assembled virus particles. Number of retroviruses have evolved different mechanisms to facilitate this nuclear export. **A.** In complex retroviruses, such as HIV-1, a *cis* acting structural element called rev responsive element (RRE) binds to virally encoded *trans* acting proteins (REV) to facilitate the efficient export of unspliced RNAs from the nucleus to the cytoplasm. **B.** Simple retroviruses such as MPMV, harbor a *cis*-acting constitutive transport element (CTE), which interacts with cellular proteins to export full length viral RNAs from the nucleus to the cytoplasm. **C.** "semi-complex" retroviruses such as MMTV have been shown to contain a *cis*-acting Rem-Responsive Element (RmRE) at the 3' end of the genome, which interacts with virally encoded *trans* proteins called Rem, to efficiently export MMTV RNA to the cytoplasm. Figure adapted from Jaballah, 2010.

The viral Gag and Pol ORFs overlap and the Gag and/or Gag/Pol polyproteins are translated from the unspliced mRNA in a ratio of 20 Gag products to one Gag-Pol product on free ribosomes (reviewed in Pederson and Duch, 2006; Lever, 2007). Due to the overlapping nature of Gag and Pol ORFs, the translation process is paused at a certain stage to allow the ribosome to shift from translating Gag ORF to Pol ORF. Such a translational switch has been shown to occur in two different ways. In the first way, as in the case of gamma retrovirus, the Moloney murine leukemia virus (MoMLV), translation termination is suppressed by misreading the stop codon. This process shifts the reading from Gag to Pol ORF and thus translating the downstream POL ORF (Yoshinaka et al., 1985; Honigman et al., 1991). The second mechanism (used by most retroviruses) is a frameshift mechanism, in which the ribosomes slips backward by one nucleotide, thus changing the reading frame it is translating in- a process called ribosomal frameshifting (Figure 1.8). This process involves the

Figure 1.8: Schematic representation of ribosomal frameshifting phenomenon in retroviruses to express relative amounts of Gag and Pol proteins.

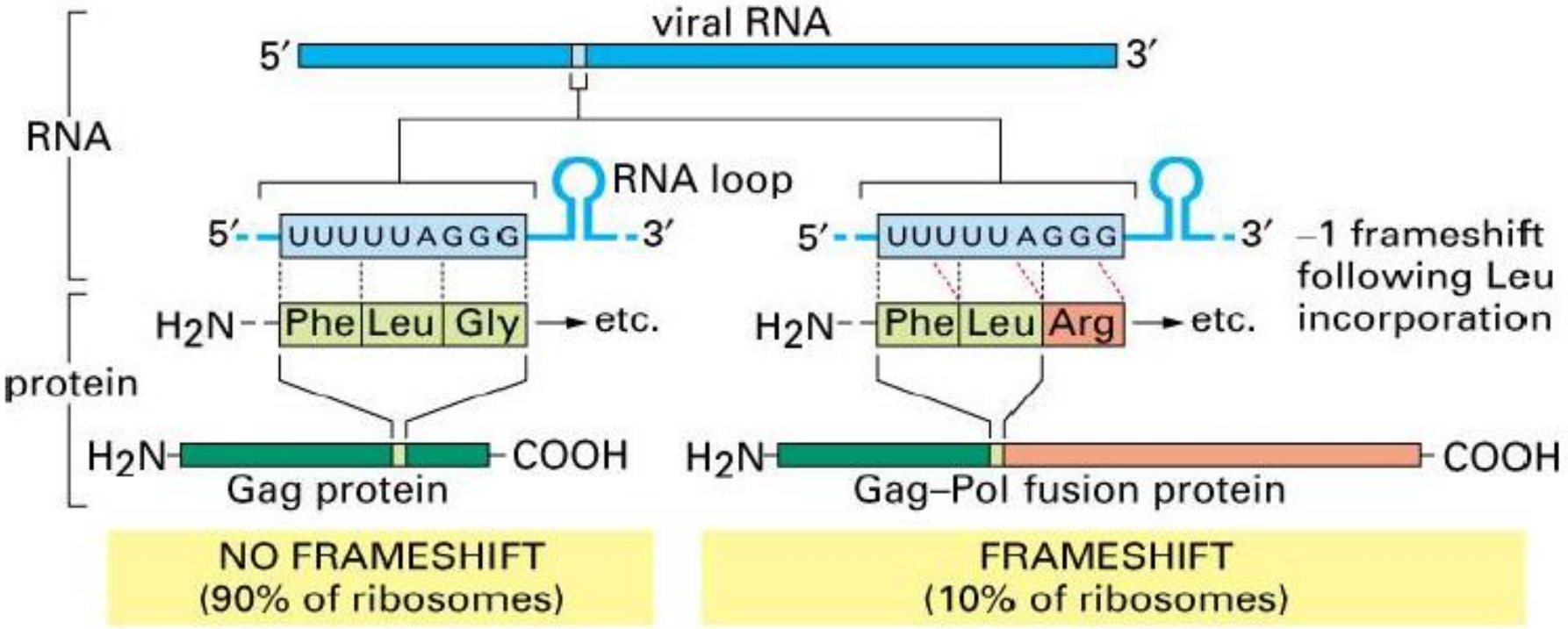


Figure 1.8. Schematic representation of ribosomal frameshifting phenomenon in retroviruses to express relative amounts of Gag and Pol proteins. Retroviral gRNAs contain "slippery" sequences that allow ribosomes to move to a different reading frame and continue translation by a process known as "ribosomal frameshifting". Briefly, the mRNA folds into a RNA secondary structure containing a stem loop which is located just downstream of the slippery sequence. This stem loop causes ribosomes to pause at the position of the slippery sequence during translation, which in turn results in increased frequency of ribosomal frameshifting. Frameshifting enables a higher translation rate of proteins encoded by the downstream reading frame. Figure adapted from Alberts et al., 2002.

ribosomal frameshift signal (RFS) comprised of a reiterated sequence and an RNA secondary structure, the pseudoknot present downstream of the RFS (reviewed in Brierley and Dos Ramos, 2006). Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res.,* 119**:** 29-42).

Most ribosomes encountering RFS translate it without difficulty and continue along the transcript until a translation stop codon is reached. During this process, a proportion of the ribosomes which attempt to translate this sequence slip back by one nucleotide before continuing to translate the message in a different reading frame. Because of this, the sequence has been termed the "slippery sequence" (Figure 1.8). The slippery sequence alone only results in a low frequency of frame shifting, inadequate to produce the amount of protease and reverse transcriptase protein required by the virus. There are additional sequences that further regulate this system and increase the frequency of frame shift events. A short distance downstream of the slippery sequence is an inverted repeat which allows the formation of a stem-loop structure in the mRNA (Figure 1.8). There is an additional sequence complementary to the nucleotides in the loop which allows base-pairing between these two regions of the RNA, allowing the formation of what is known as an RNA "pseudoknot". This secondary structure in the mRNA causes ribosomes translating the message to pause at the position of the slippery sequence and this slowing or pausing of the ribosome during translation increases the frequency at which frame shifting occurs, boosting the relative amounts of the proteins encoded by the downstream reading frames (Figure 1.8).

1.3.6. Translation and Packaging of Retroviral gRNA

A close relationship between retroviral gRNA translation and packaging exists as a result of the dual role of the full length unspliced gRNA since it functions as the precursor for Gag/Pol protein expression as well as the substrate to be packaged into the virus particles. Two possible scenarios describe what happens at this stage (Figure 1.9) (reviewed in Balvay et al., 2007). The first scenario suggests that the unspliced RNA exclusively gets translated into Gag/Pol precursors until a sufficient amount of proteins start to accumulate in the cytoplasm. At this stage, the Gag polyproteins start interacting with the packaging signal within the 5' UTR of the unspliced gRNA. This process results in the progressive occlusion of the 5' UTR as the ribonucleoprotein scaffold is formed. Since, the same RNA that is used for translation is to be packaged into the assembling virus particle therefore in an alternative scenario, the viral Gag/Pol precursor proteins are translated from the unspliced gRNA constitutively. The resulting Gag proteins bind to another unspliced gRNA strand in the pool of unspliced RNA present in the cytoplasm. The two mechanisms are believed to be mutually exclusive (Balvay et al., 2007).

1.3.7. Virus Assembly

The viral gRNA represents only a small fraction of the total RNAs present in the infected cell. However through a very selective and discrete process, only two copies of the viral gRNA as a dimer are packaged in the newly assembled virus. To unravel the mechanism(s) by which the virus encapsidate its gRNA in such a discrete manner one must understand the intricacies involved during gRNA dimerization and packaging.

Figure 1.9: Translational control of retroviral genomic RNA packaging.

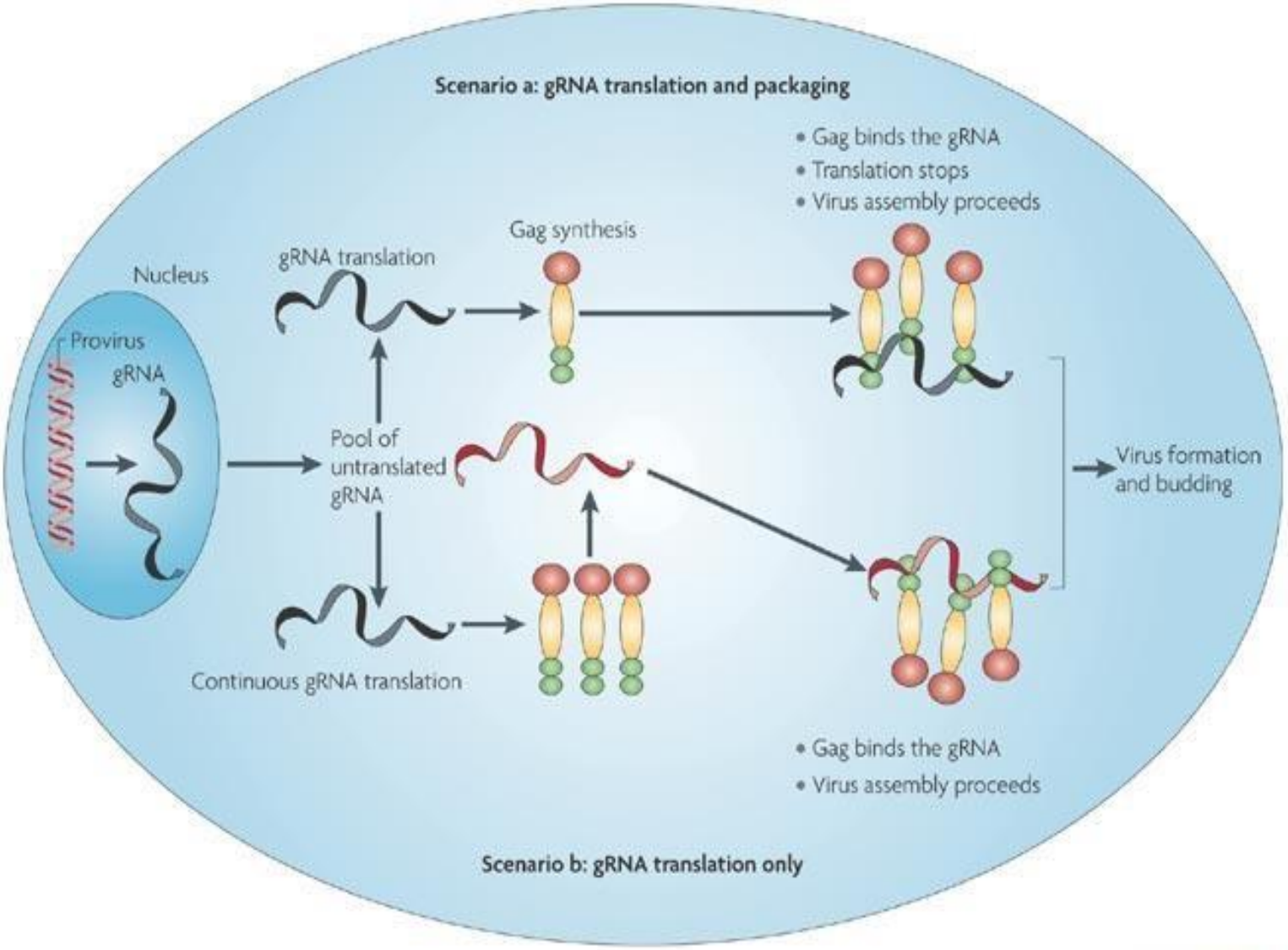


Figure 1.9. Translational control of retroviral genomic RNA packaging. After having integrated into the host genome, the retroviral genome undergoes transcription and splicing followed by the transport of both full length and spliced RNAs to the cytoplasm. The full length unspliced RNA has two roles to play and the mechanism by which its fate is decided can be described in two scenarios. In scenario (a), the genomic RNA (gRNA) undergoes translation to form Gag and Gag-pol proteins (not shown). Upon the accumulation of Gag molecules, a switch from translation to packaging happens, allowing the Gag proteins to bind the gRNA (shown in black) at the 5′ end to the packaging signal and proceed to the assembly step. Whereas in scenario b, the gRNA is translated continuously and the resulting Gag proteins bind the gRNA which was not translated (shown in red). Figure adapted from Balvay et al., 2007.

1.3.8. Retroviral Genomic RNA Dimerization

The dimerization process is considered to be one of the most critical steps during the retroviral life cycle as it is essential to form the non-covalently linked RNA dimers that can be packaged in the assembling virus particles (reviewed in D'souza & summers, 2005; Johnson & Telestnitsky, 2010; Lever, 2007). Dimerization begins at the dimerization initiation site (DIS), which is invariably a G-C rich palindromic (pal) sequence and often present in the form of a loop (at the 5' end of gRNA), facilitating its interaction with the DIS loop on the second gRNA (Laughrea et al., 1996; Paillart et a., 1996 and further reviewed in D'souza & summers, 2005; Johnson & Telestnitsky, 2010; Lever, 2007; Figure 1.10). Three DIS motifs consisting of 6 nucleotides pal sequences (GCGCGC, GTGCAC and GTGCGC) have been reported in multiple HIV-1 isolates (Berkhout, B., and van Wamel, 1996; Clever et al., 1996; Laughrea et al., 1997; Laughrea et al., 1999). Among these, a DIS motif consisting of a characteristic 6 nt pal (GCGCGC) has been found to be phylogenetically conserved in over 50 HIV-1, HIV-2 and SIV isolates (Russell et al., 2004). In addition, it has been shown that a 10 nt pal sequence in the 5' UTR, which is phylogenetically conserved in HIV-2 and macaque and sooty mangabey SIVs (Leitner et al., 2005), is crucial for HIV-2 RNA packaging and dimerization (Lanchy et al., 2003; Baig et al., 2007; Lanchy and Lodmell 2007). Finally, pal sequences that have been shown to function as DIS in augmenting gRNA dimerization and packaging have also been observed in other retroviruses such as Feline immunodeficiency virus (FIV) and MMTV (Kenyon et al., 2008; Rizvi et al., 2010; Aktar et al., 2014). The base pairing between the two strands is mainly maintained by the auto-complementarity between the palindromic sequences

Figure 1.10: Model depiction the steps leading to dimerization and packaging processes among retroviruses.

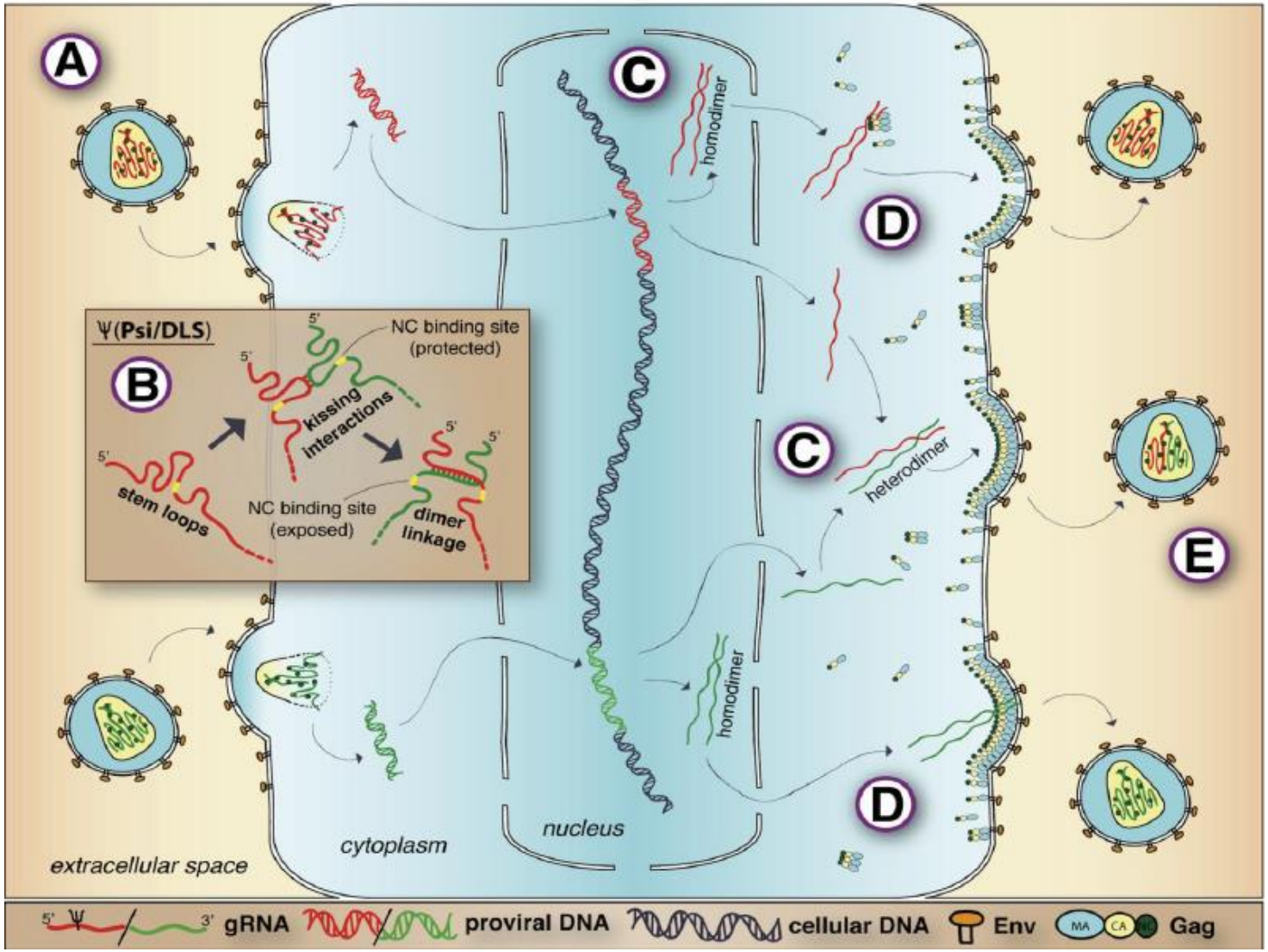


Figure 1.10. Model depiction the steps leading to dimerization and packaging processes among retroviruses. A schematic simulation of dimerization and gRNA packaging in a co-infected cell with two different viruses. (A) Each virus particle has a dimer-linked two identical RNAs packaged in its capsid. (B) The gRNA inside the cytoplasm undergoes dimerization via "kissing" interactions between palindromic stem loops. Once linked, a conformational change occurs to expose single-stranded nucleocapsid (NC) binding motifs (indicated in yellow) which is recognized by the nucleocapsid portion of the Gag polyprotein and is needed for selecting the gRNA for packaging. (C) Different viruses form their dimers at different sites; some form their dimers in the nucleus where the probability of forming homodimers is higher, whereas other retroviruses associate in the cytoplasm and thus random assortment of homodimeric and heterodimeric gRNAs occurs. (D) The binding between the Gag proteins and gRNA dimer can also happen in two sites within the cell. gRNA can form subassemblies with Gag in the cytoplasm (shown at the top) or may associate at the plasma membrane (shown at the bottom). (E) Finally two gRNAs are packaged into the newly budding virus particles. Figure adapted from: Johnson and Telesnitsky, 2010.

in an antiparallel direction to initiate “kissing loop” complex that ultimately leads to a stable dimer formation. The initiation site of gRNA dimerization has been shown to be different in different retroviruses (reviewed in Johnson & Telesnitsky, 2010; Jouvenet et al., 2011; Figure 1.10). Consequently, various types of dimers have been observed in different retroviruses. Viruses that dimerize in the cytoplasm have the possibility to form both homodimers as well as heterodimers, including HIV-1 and feline immunodeficiency virus (FIV), whereas viruses undergoing dimerization in the nucleus usually form homodimers as in the case of MLV and Rous sarcoma virus (RSV) (reviewed in Johnson & Telesnitsky, 2010; Jouvenet et al., 2011; Figure 1.10).

1.3.9. Retroviral Genomic RNA Packaging

One of the hallmarks of retroviral life cycle is the efficient and specific packaging of retroviral gRNA as dimers by the assembling virus particles. During this process, full-length, unspliced gRNA is preferentially packaged, whereas spliced viral RNA and cellular RNA are generally excluded from being encapsidated into the nascent virus particles. The packaging specificity results from high-affinity interactions between NC and a specific sequence at the 5’ end of the viral genome called the packaging determinant and/or signal (Ψ; reviewed in D'souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007). Determinants of gRNA dimerization and packaging map to the same 100 to 400 nucleotides at the 5' end of the gRNA; therefore, it is not surprising that retroviral packaging signal is located within the dimer linkage structure (DLS) that is formed when two gRNAs dimerize at their 5’ ends. A number of studies have been undertaken to identify the packaging determinants of retroviruses, which have shown that packaging sequences are

Figure 1.11: Model depicting specific and selective unspliced genomic RNA packaging among retroviruses.

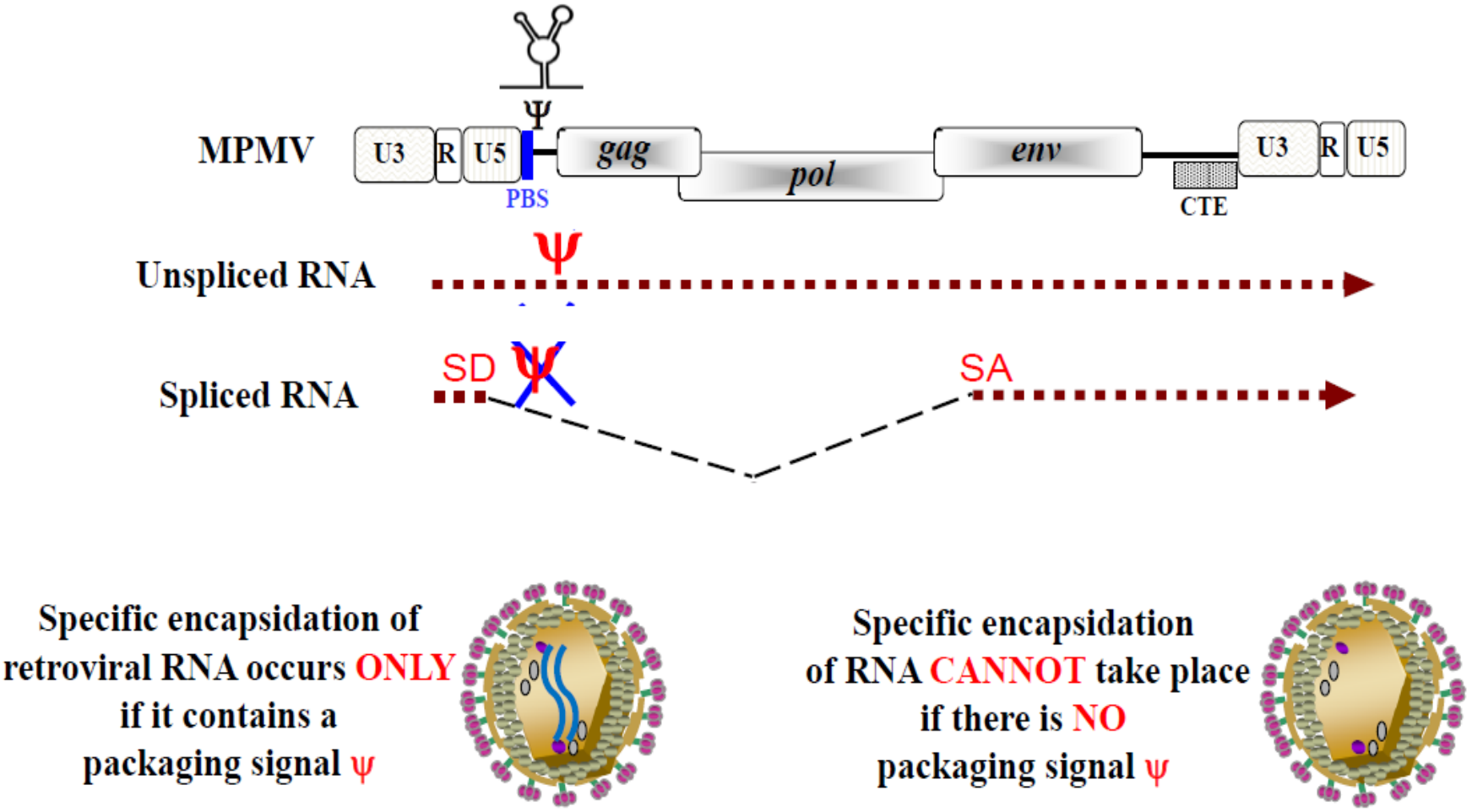


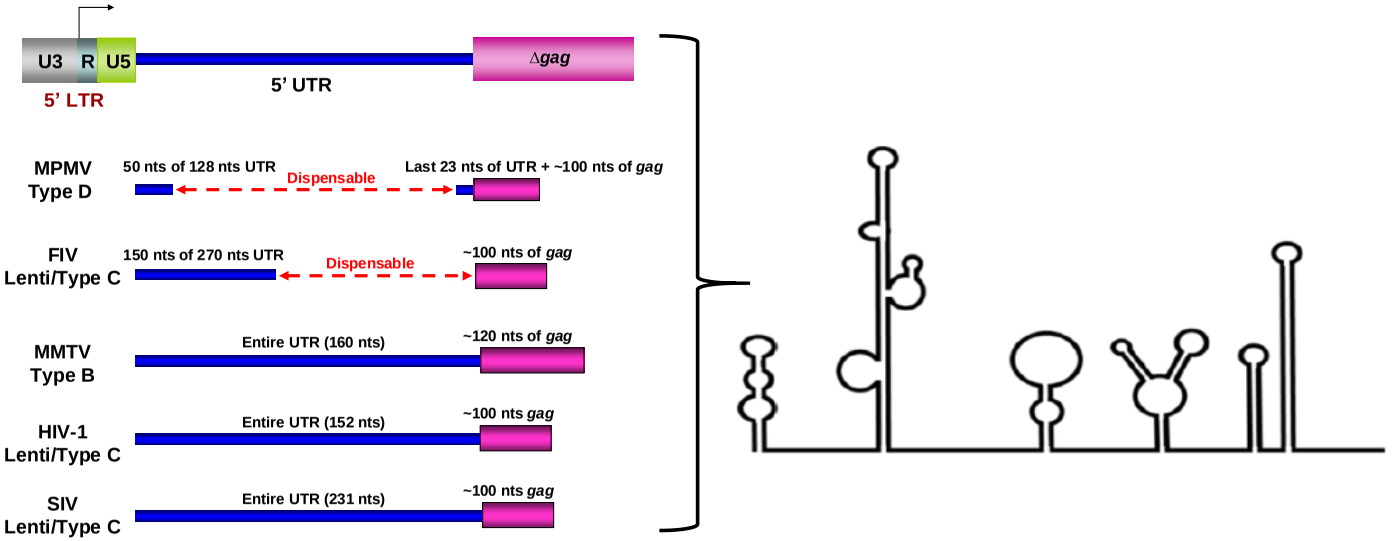
Figure 1.11. Model depicting specific and selective unspliced genomic RNA packaging among retroviruses. Retroviruses specifically and selectively package their gRNA amongst a pool of other spliced viral and cellular RNAs. The packaging strategy requires the recognition of the packaging determinants named the packaging signal (psi or ψ). The location of the packaging signal location is the key for conferring the specificity of the RNA packaging process since in most of the retroviruses it is located after the splice donor site (SD). In the spliced RNAs, the packaging signal is lost during the splicing steps, while in the full length unspliced genomic RNA, the packaging signal is maintained and only those RNAs containing the signal can be encapsidated into the virus particle. Figure adapted from Jaballah thesis, 2010.

generally present downstream of the major splice donor (mSD). Therefore, a very simplistic model (Figure 1.11) suggests that the full-length unspliced genomic RNA is preferentially packaged by virtue of the presence of the packaging determinant following interaction with NC binding site. On the other hand, in the case of the spliced RNAs, the packaging sequences are spliced out, thereby excluding them from encapsidation into the nascent viral particles (Figure 11). Such a model has been proven true in the case of HIV-1 where the stem loop 2 (SL2) is located in the region harboring the packaging determinants of HIV-1 and is capable of binding HIV-1 NC with high affinity. HIV-1 SL2 also contains the mSD; therefore, during splicing it is deleted leading to its absence in spliced transcripts and rendering them incapable of binding to NC. This suggests a possible mechanism for discriminating between spliced and unspliced viral mRNAs (reviewed in D'Souza and Summers, 2005).

The packaging discrimination between spliced and unspliced RNAs in HIV-2 seems to be more complicated as both the spliced and unspliced viral mRNAs contain the packaging determinants, yet only the unspliced messages are encapsidated into the virus particle (reviewed in Balvay et al., 2007). HIV-2 has evolved a distinct strategy to select for the unspliced RNA for packaging by packaging only those mRNAs that are capable of translating Gag in *cis* through a phenomenon called co-translational packaging (Kay and Lever 1999; Griffin et al., 2001).

Despite the fact that RNA packaging is a universal step in all retroviruses and the packaging signal required to accomplish this process is found at the 5' end of the retroviral gRNA (Figure 1.12), no sequence conservation has been found between

Figure 1.12: Summary of the packaging and dimerization determinants of a number of different retroviruses.



Virus	Length of 5' UTR	5' UTR Required	Gag Required	References
MPMV (Type D)	128 bases	First 50 bp + last 23 bp	~ 100 bases	Schmidt et al., 2003; Jaballah et al., 2010; Aktar et al.,2013
FIV (Type C)	270 bases	First 150 bp	~ 100 bases	Browning et al., 2003a and 2003b; Mustafa et al., 2006; Ghazawi et al., 2006
MMTV (Type B)	160 bases	Entire UTR	~120 bases	Rizvi et al., 2009; Mustafa et al., 2012; Aktar et al., 2014
HIV-1 (Type C)	152 bases	Entire UTR	~ 100 bases	McBride and Panganiban, 1997; Clever and Parslow, 1997
SIV (Type C)	231 bases	Entire UTR	~ 100 bases	Rizvi and Panganiban, 1993; Guan et al., 2000; Patel et al., 2003; Strappe et al., 2003; Whitney and Wainberg; 2006

Figure 1.12. Summary of the packaging and dimerization determinants of a number of different retroviruses. Schematic representation of the 5' end of the genomic RNA of different retroviruses with the packaging determinants sequences marked. The packaging signal in almost all retroviruses has been shown to acquire a higher order secondary structure. The table summarizes and compares the sequences starting from nucleotide +1 in R to the beginning of *gag* that have been shown to be required for optimal packaging of retroviral gRNA as per published data. The references and other details of the gRNA packaging determinants are provided in the table below. Figure adapted from Mustafa et al., 2012.

the packaging signals of different retroviruses. Moreover, it is becoming evident from the recent cross-packaging studies that switching the packaging sequences between two different retroviruses that have no primary sequence homology still maintains efficient packaging (Al Daheri et al., 2009; Al Shamsi et al., 2011). As a result, it is plausible to assume that gRNA dimerization and packaging process involves recognition of packaging and dimerization sequences at the secondary structural level rather than the primary sequence level. Additionally, it has been observed that the packaging and dimerization sequences of almost all retroviruses (irrespective of their simple or complex nature) assume higher order structures comprising of various structural motifs (reviewed in D'souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007; Figure 1.12). One of the characteristic features of RNA secondary structure of the retroviral packaging signal RNA is the long range interactions (LRIs) involving sequences of the 5' and the 3' end regions, including sequences in R/U5 that are complementary to a sequence found downstream in the *gag* gene.

Owing to the importance of the RNA secondary structures and long-range base pairings, packaging signal RNA structural models of a number of viruses have been investigated recently by various approaches such as free energy predictions, phylogenetic analyses, and biochemical probing, in addition to their subsequent validation by biological assays. The collective findings are providing new insights into the structural determinants and molecular mechanisms of retroviral genome packaging (Kenyon et al., 2008; Rizvi et al., 2010; Jaballah et al., 2010; Kenyon et al., 2011; Aktar et al., 2013; Aktar et al., 2014).

1.3.10. Possible Link between Retroviral Genomic RNA Packaging and Dimerization

Over the years, the retroviral dimerization has been proposed to be closely related to the viral gRNA packaging, and the dilemma whether dimerization is a pre-request for packaging or they both happen at the same time have been studied extensively (Hibbert et al., 2004; reviewed in D'souza & summers, 2005; Johnson & Telestnitsky, 2010; Lever, 2007; Miyazaki et al., 2011; Moore et al., 2009; Paillart et al., 2004). It has been shown that the packaged RNA remains in the dimeric state, suggesting that dimerization of the two gRNAs may be a prerequisite for packaging (Levin et al., 1974; Housset et al., 1993; Hibbert et al., 2004). Consequently, and not surprisingly, the determinants of gRNA dimerization and packaging map to the same 100 to 400 nucleotides at the 5' end of the gRNA which have been proposed to be physically and genetically inseparable (reviewed in D'souza & summers, 2005; Johnson & Telestnitsky, 2010; Lever, 2007). The understanding of why gRNAs are packaged as dimers has recently been unveiled in the case of MLV. The NC in MLV has been shown to specifically bind to unpaired UCUG motifs abundantly present at the 5’ UTR. However, these motifs in the monomeric gRNA are base paired and thus the NC cannot have an easy access to them. During dimerization, interactions between pal helix loops leads to conformational changes in RNA structural motifs, exposing the high affinity binding sites and making them easily accessible for NC binding, facilitating the packaging of a dimeric gRNA (D’Souza & Summers, 2005; Johnson & Telestnitsky, 2010; Figure 1.10).

1.4. Genomic RNA Packaging and Dimerization Determinants of MPMV

MPMV causes an immunodeficiency syndrome in newborn rhesus monkeys and is a classical type D retrovirus (Bryant et al., 1986; Fine et al., 1975). Retroviruses that resembles very closely to MPMV and cause acquired immunodeficiency syndrome (AIDS) in primates have also been characterized and include simian retrovirus type 1 and type 2 (SRV-1 and SRV-2; Daniel et al., 1984; Desrosiers et al., 1985; Marx et al., 1984). Among these beta retroviruses, MPMV is perhaps the most well-investigated in terms of its gRNA packaging and dimerization. Initially, using MPMV-based retroviral vectors, the first region that was proposed to be important for its RNA packaging was a 624 nt region downstream the PBS (Vile et al., 1992). Later, biochemical probing, free-energy estimations, and phylogenetic analyses were employed to predict the higher order features of part of this region (up to 130 bp of Gag) which revealed a complex structure comprising of eight stem loops (Harrison et al., 1995). A further mutational analysis of the 5' end of MPMV genome revealed that deletion of 61 nt region upstream of the mSD reduced RNA packaging more than 50%, suggesting that MPMV core packaging determinants are present upstream of the mSD (Guesdon et al., 2001).

A more systematic and detailed mutational analysis of the 5' end of MPMV genome was later conducted and the mutant RNAs were tested employing a biologically relevant *in vivo* packaging and transduction assays. These mutational analyses revealed that the packaging determinants of MPMV are multipartite and resides in two distinct regions: region "A" which includes the first 50 nt of UTR and

Figure 1.13: Comparison of the predicted and the SHAPE (selective 2' hydroxyl acylation analyzed by primer extension)-validated structure of MPMV packaging signal RNA.

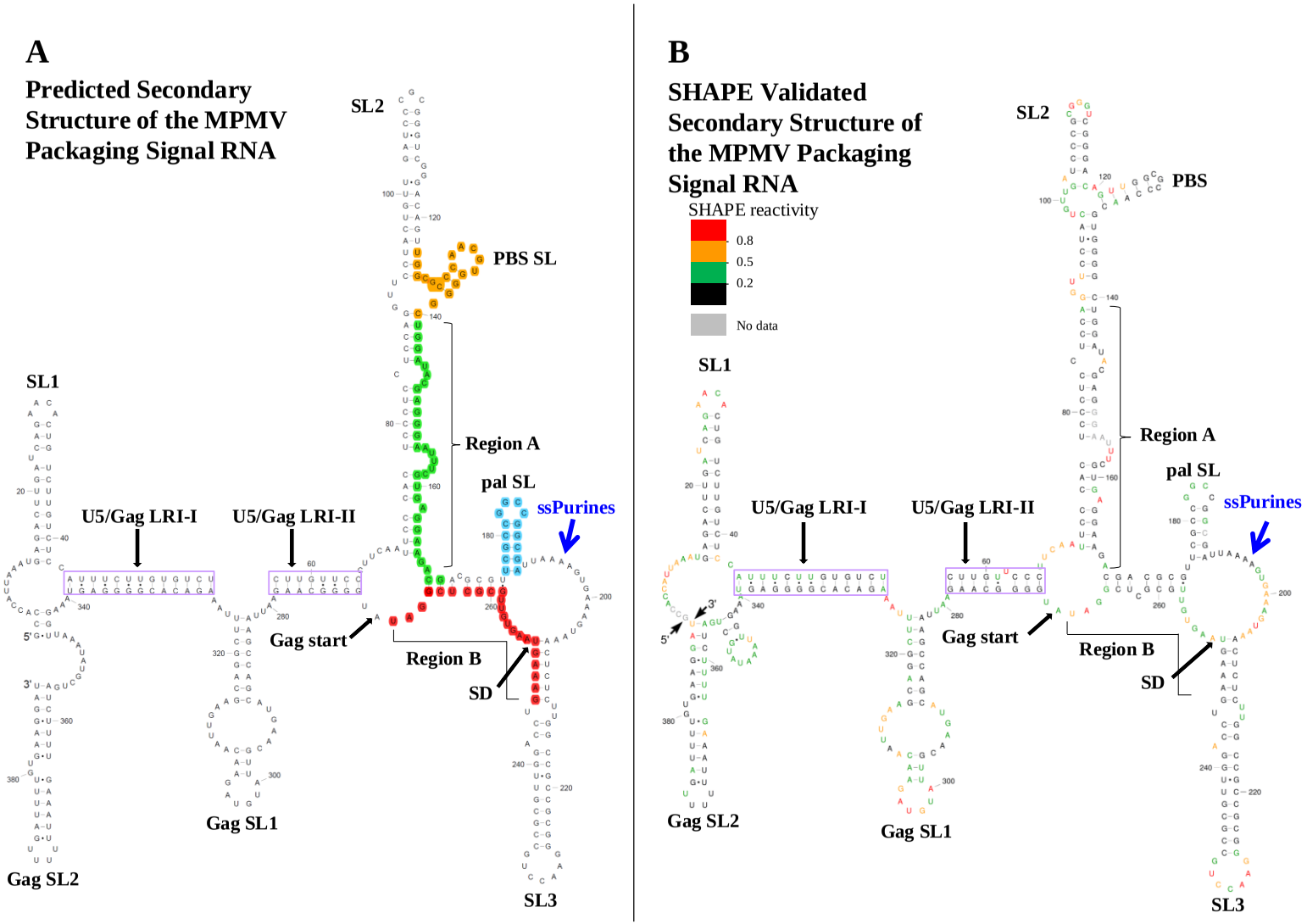


Figure 1.13. Comparison of the predicted and the SHAPE (selective 2' hydroxyl acylation analyzed by primer extension)-validated structure of MPMV packaging signal RNA. (A) Structure of MPMV packaging determinants predicted earlier (Jaballah et al. 2010) by Mfold. Regions that have been shown to be important for MPMV packaging are highlighted in orange, green, red and blue representing the primer binding site (PBS) , region A, region B and pal sequence, respectively. Long range interaction (LRI) -forming sequences are highlighted in purple boxes. **(B)** RNA structure of MPMV packaging signal after applying SHAPE-constrains (Reuter and Mathews 2010). Nucleotides are color coded as per the SHAPE reactivity key. The SHAPE-validated structure corporates well with the predicted structure except that in LRI-II the sequence of the X strand gains one extra cytosine at the 3' end to become 9 nt in length instead of the predicted 8 nt. In addition, a uridine residue becomes un-basepaired and forms a bulge in the LRI-II structure. The region used for the Mfold predictions and the SHAPE analysis included sequences from R up to 120 nt of Gag. Figure adapted from Aktar et al., 2013.

region “B” that encompasses the last 23 nt of UTR as well as the first 120 nt of Gag (Schmidt et al., 2003; Jaballah et al., 2010; Figure 1.13). To establish a structure-function correlation of the effects of the introduced mutations on MPMV RNA packaging, the region between R and the first 120 nt of Gag was folded and RNA secondary-structure predictions showed that the 5’ end of MPMV genomic RNA assumes several stable stem loop structures (Jaballah et al., 2010; Figure 1.13A). The distinctive features of the predicted structures included U5-Gag long range interactions (LRIs), a stretch of single-stranded purine (ssPurine)-rich region, and a distinctive G-C-rich palindromic (pal) stem-loop (Figure 1.13A). Both pal stem-loop and ssPurine-rich region (or its partial repeat region when predicted to refold as ssPurines) have been shown to be essential for MPMV RNA packaging (Jaballah, et al., 2010).

To further validate the predicted RNA secondary structure, a novel chemo-enzymatic probing strategy known as Selective 2' Hydroxyl Acylation by Primer Extension (SHAPE) methodology was employed. SHAPE analysis validated the overall predicted structure (Jaballah et al. 2010) of the MPMV packaging signal RNA (Aktar et al., 2013; Figure 1.13B). Finally, a systematic deletion analysis, minimum free-energy structure predictions, phylogenetic, and *in silico* modeling analyses were undertaken to further develop a structure-function relationship of the various structural motifs of MPMV packaging signal RNA. These analyses revealed that the 6 nt pal (5’-CGGCCG-3’) within the pal stem-loop functions as MPMV dimerization initiation site (Aktar et al., 2013).

One of the characteristic feature of the SHAPE-validated secondary structure of the MPMV packaging signal RNA is the presence of long range interactions (LRIs: LRI-I and LRI-II; Figure 1.13) between U5 and Gag complimentary

sequences in the region that have been shown to be important in MPMV RNA packaging (Schmidt et al., 2003; Jaballah et al., 2010). *In silico* modeling analysis of five different MPMV strains predicted both LRIs formation in all the strains (Aktar et al., 2013; Figure 1.14). Consistent with this the U5 and Gag sequences involved in the LRIs revealed a high degree of conservation and complimentarily within strains and maintain a very high degree of complementarity (Aktar et al., 2013; Figure 1.15). Therefore, it is reasonable to propose that LRIs between U5 and Gag sequences could potentially play a role in gRNA packaging perhaps by maintaining the overall MPMV packaging signal RNA secondary structure which seems to be anchored by complementary and conserved U5 and Gag sequences.

Figure 1.14: Mfold predictions of the 5’ end of genomic RNA of five different strains of MPMV

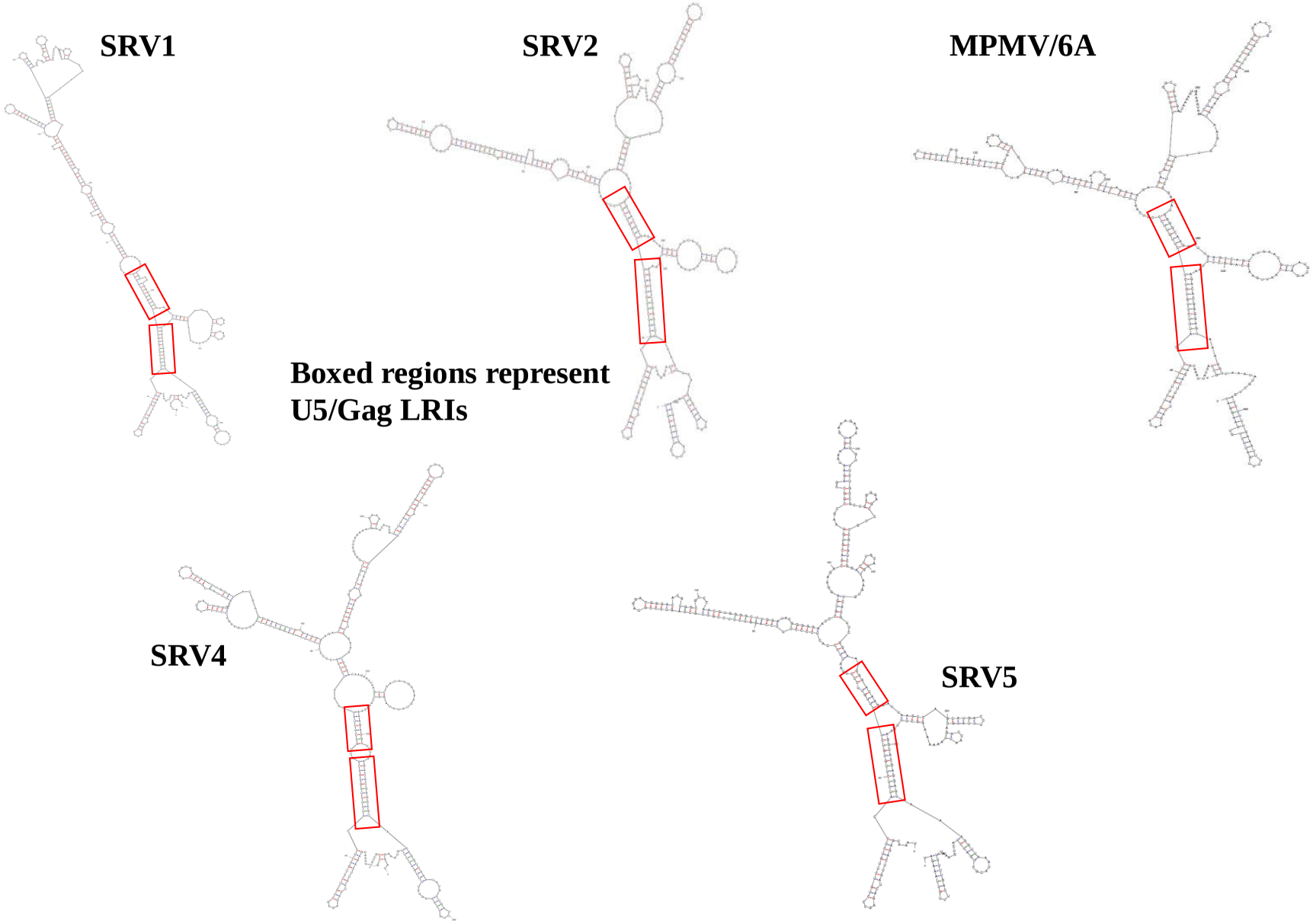


Figure 1.14. Mfold predictions of the 5' end of genomic RNA of five different strains of MPMV showing phylogenetically conserved LRIs involving U5 and Gag sequences highlighted in red boxes. The accession numbers of MPMV/6A, SRV1, SRV2, SRV4 and SRV5 are M12349.1(Sonigo et al., 1986), M11841.1 (Power et al., 1986), AF126467.1 (Marracci et al., 1995), FJ979638.1 (Zao et al., 2010), and AB611707.1 (Takano et al., 2013), respectively. Figure adopted from Aktar et al., 2013.

Figure 1.15: Sequence alignment of 5’ end of genomic RNA from five different isolates of MPMV using Clustal W.

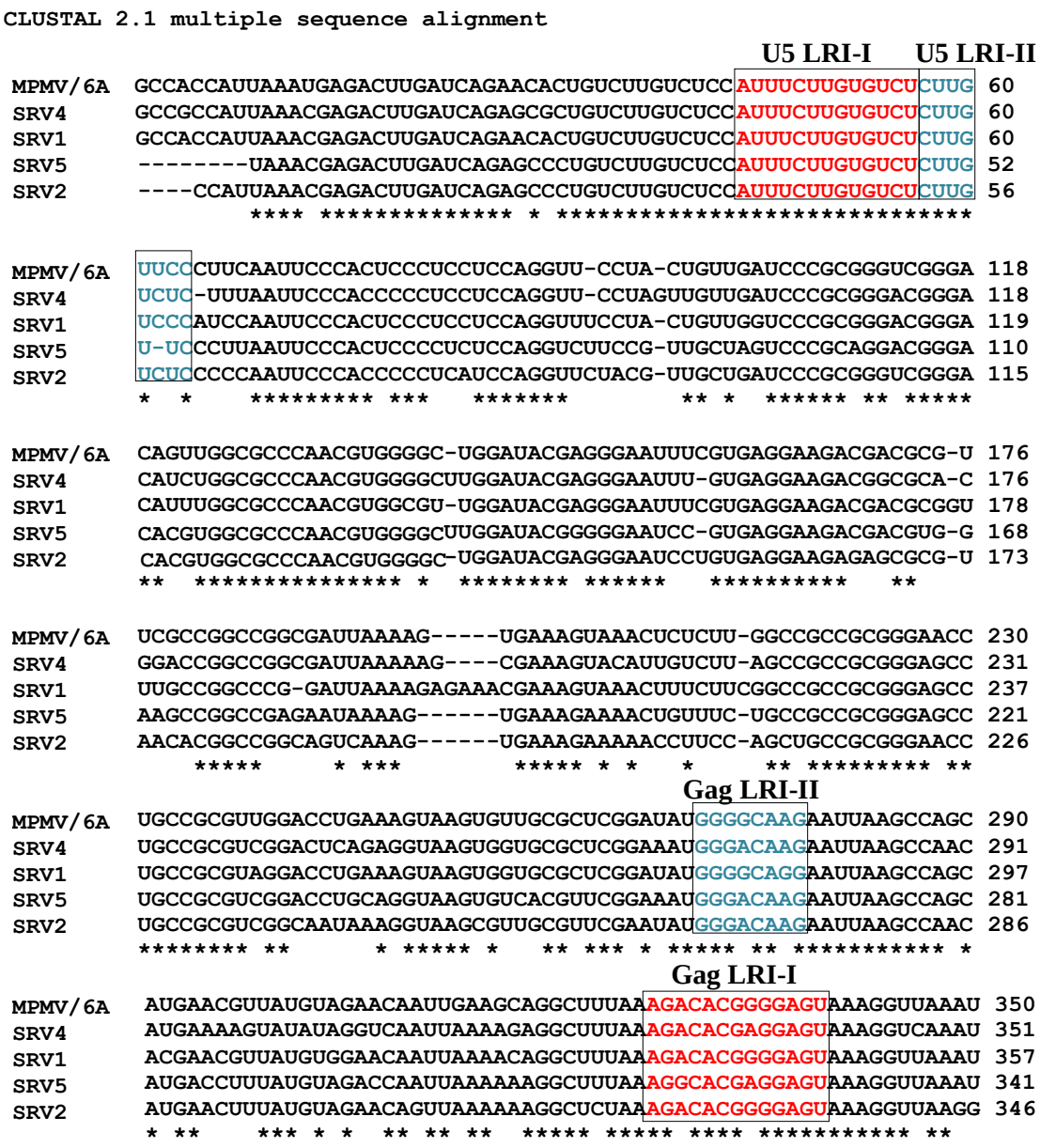


Figure 1.15. Sequence alignment of 5' end of genomic RNA from five different isolates of MPMV using Clustal W. The alignment reveals a high degree of conservation of the sequences involved in the formation of the two LRIs (sequences in the boxes). Even though those sequences show a slight variation, yet they still maintain a high degree of complementarity. The U5/Gag sequences involved in the formation of LRI-I are highlighted in red, whereas the sequences forming LRI-II are highlighted in blue. The accession numbers of MPMV/6A, SRV1, SRV2, SRV4 and SRV5 are M12349.1(Sonigo et al., 1986), M11841.1 (Power et al., 1986), AF126467.1 (Marracci et al., 1995), FJ979638.1 (Zao et al., 2010), and AB611707.1 (Takano et al., 2013), respectively. Figure adapted from Aktar et al., 2013.

1.5. Objectives

Our predicted and SHAPE-validated structures of MPMV gRNA have shown conserved LRIs between the U5 and Gag complimentary sequences in a region that have been shown to be crucial for MPMV gRNA packaging. In addition, several studies from a number of retroviruses have reported the existence of such LRIs which have been found to be crucial for retroviral gRNA packaging. Therefore, we hypothesize that the RNA secondary structures of the MPMV appears to be anchored by complementary sequences in the U5 region and within the Gag open reading frame, providing stability to the overall secondary structure to MPMV packaging signal RNA. We further hypothesize that the U5 and *gag* complementary sequences involved in LRIs could either function at the primary sequence level or at the structure level during MPMV RNA packaging process. Therefore, in this thesis, we tested the following:

1. Ascertain the existence as well as the biological significance of LRIs in MPMV and packaging signal RNA sequences required in *cis* for gRNA packaging

2. Ascertain whether the complimentary U5-Gag sequences involved in maintaining LRIs function at the primary sequence level or the structural level (or both) during gRNA packaging

3. Establish structure-function relationship during the MPMV gRNA packaging process by correlating the biological results with the structural predictions of the LRI mutants

4. Establish the role of Gag sequences other than those involved in U5-Gag LRIs during the MPMV gRNA packaging process.

Chapter 2: Materials & Methods

2.1. Genome Nucleotides Numbering System

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

2.2. Plasmid Construction:

2.2.1. MPMV Packaging Construct:

The packaging construct is the plasmid that provides the viral structural proteins into which retroviral gRNA is packaged. In the case of MPMV, TR301 plays this role and has been described earlier (Browning et al., 2001). In brief, TR301 plasmid expresses the MPMV **gag/pol** genes under the transcriptional control of the human cytomegalovirus (hCMV) intron A promoter/enhancer. In TR301, MPMV CTE has been cloned between *pol* termination codon and the bovine growth hormone (BGH) poly A sequences to ensure proper nuclear export of the unspliced *gag/pol* mRNA (Figure 2.1).

2.2.2. MPMV Transfer Vector:

The MPMV sub-genomic transfer vector, SJ2, has been described previously (Jaballah et al., 2010) and harbors all the required *cis*-acting sequences for transcription, polyadenylation, reverse transcription, integration, dimerization, and packaging (Figure 2.1). In addition to these cis-acting sequences, SJ2 also expresses *hygromycin B phosphotransferase* gene from an internal simian virus 40 early promoter (SV-Hyg^r) that is is used as the selectable marker. The overall size of SJ2

Figure 2.1: Schematic representation of MPMV genome, its sub-genomic transfer vector, the packaging construct and the envelope expression plasmid.

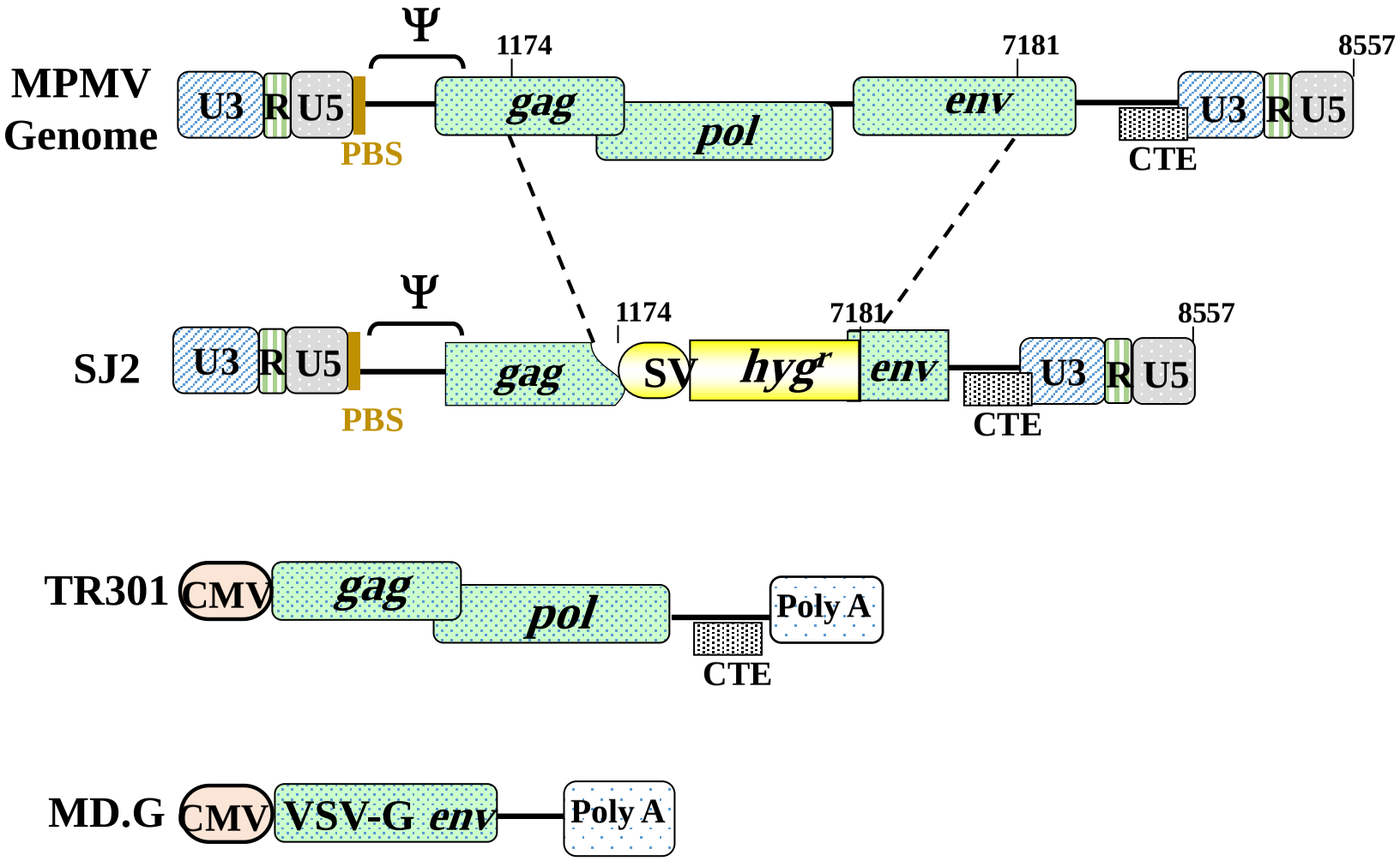


Figure 2.1. Schematic representation of MPMV genome, its sub-genomic transfer vector, the packaging construct and the envelope expression plasmid. The wild type sub-genomic transfer vector, SJ2, contains a deletion in MPMV sequences between nt 1174 and 7181. This region has been replaced with *hygromycin resistance* gene expressed from an internal simian virus 40 promoter (SV-*hyg*^{r} cassette). This transfer vector was used to introduce mutations in the LRI-I, LRI-II and other sequences. TR301 is the packaging construct that expresses MPMV Gag/Pol proteins. MD.G plasmid expresses envelope glycoproteins G from vesicular stomatitis virus (VSV-G). **Ψ**, packaging signal; **CTE**, constitutive transport element; **SV**, Simian virus 40 promoter; ***hyg***^{r}, *hygromycin resistance* gene; **CMV**, human cytomegalovirus promoter; **Poly A**, polyadenylation sequences.

is approximately 6 kb and at its 5' end contain sequences upto 283 nt of Gag (nt 1174) while at the 3' end, it contains sequences from MPMV nucleotide 7181 until the end of MPMV genome (nt 8557).

2.2.3. Envelope Expression Plasmid:

The vesicular stomatitis virus glycoprotein G (VSV-G) expression vector MD.G was used to psuedotype the virus particle produced following transfection. This enables these pseudotyped viruses to infect a number of target cells since receptors for VSV-G are present on many cell types (Figure 2.1). Such pseudotyped particles allowed studying the effect of the mutations on RNA packaging and mutant transfer vector RNA propagation. MD.G has been previously described (Naldini et al., 1996) and was kindly provided by Dr. Dider Trono (Salk Institute,La Jolla, CA).

2.2.4. Transfer Vectors Containing Different Mutations:

To validate the existence and biological significance of the U5-Gag LRIs, a series of mutations were introduced into them. The first series of mutations included deletion of the U5 sequence, the Gag sequence, or both. Mutations were introduced by designing customized primers that can be used in a splice overlap extension PCR (Gibbs et al., 1994) using SJ2 transfer vector as the template DNA as has been represented schematically in figure 2.2 and described previously (Jaballah et al., 2010). Briefly, two rounds of PCR amplifications were performed and each round required a different set of primers. In the first round, the customized primers that were designed flanking the region of mutation were used in two separate PCR reactions, (A) and (B). PCR (A) was performed using the universal outer forward (sense; S) primer OTR787 (5'ccctcgagTGTCCGGAGCCGTGCTGCCCG 3'; first 2

Figure 2.2: Schematic illustration of the splice overlap extension (SOE) PCR strategy used to introduce mutations.

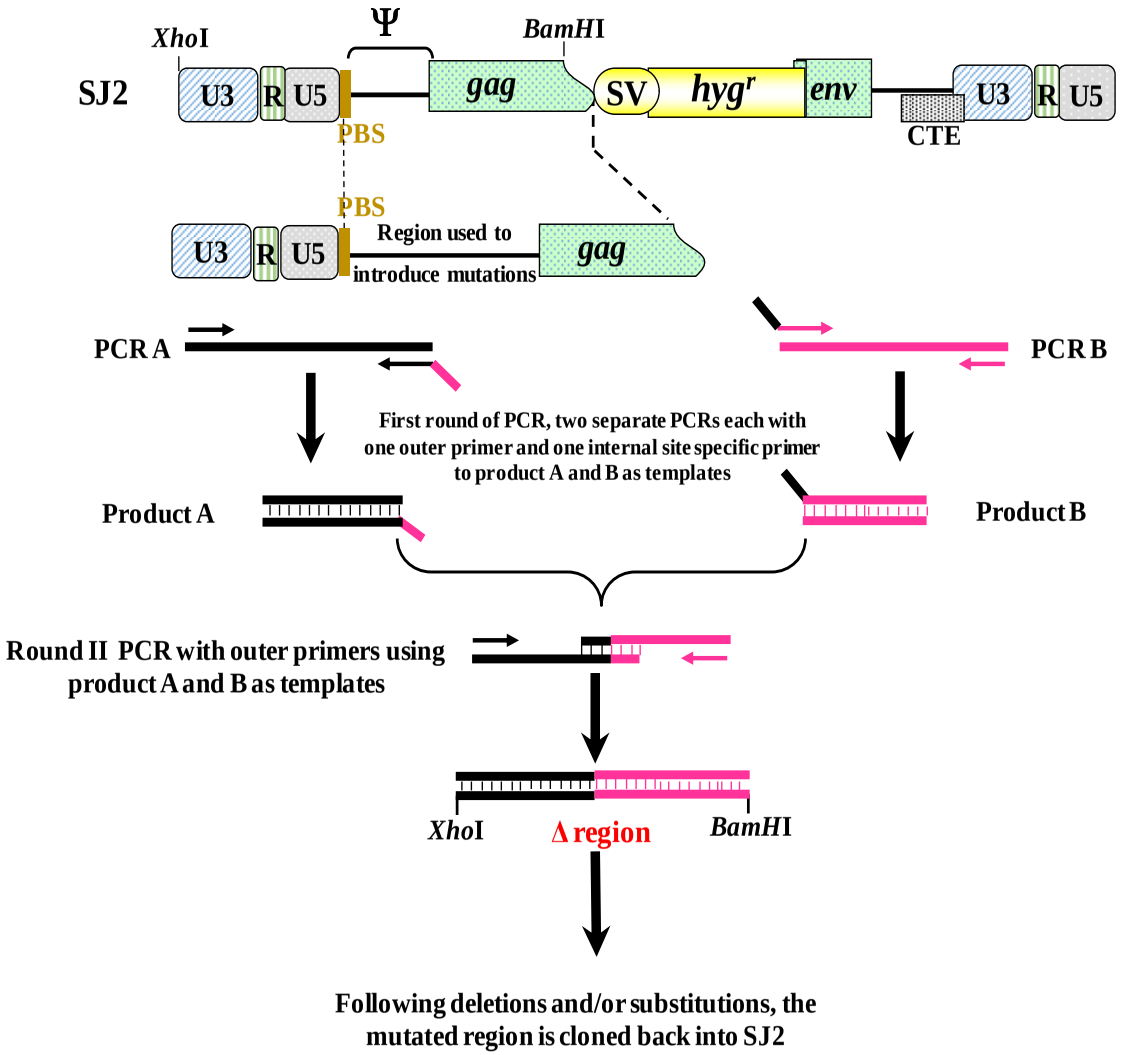


Figure 2.2. Schematic illustration of the splice overlap extension (SOE) PCR strategy used to introduce mutations. In the first round of PCR, two separate PCR reactions were performed (PCR-A and PCR-B) where each has an outer primer that amplifies either the 5' end or 3' end regions of the MPMV sequences and another internal primer that is specific to the mutated sequence. Both inner sense (S) and antisense (AS) primers that were employed in PCRs (B) and (A), respectively, were customized in a way that the resulting PCR products from these two separate reactions (PCR A and PCR B) should have overlapping complementary sequences. The presence of these complementary sequences in PCR products A and B allowed them to anneal in round 2 PCR when performed using outer S and AS primers, generating a final product harboring the deletion introduced. Following round 2 PCR, the final product in addition to containing the desired mutation also acquired flanking distinctive endonuclease restriction sites, namely *BamH*I at the 5'-end and *Xho*I at the 3'-end of the product that was designed to be used to clone the desired mutations. **Ψ**, packaging signal; **CTE**, constitutive transport element; **SV**, Simian virus 40 promoter; ***hyg***^{r}, *hygromycin resistance* gene. nucleotides (nts) are dummies followed by *Xho*I site and MPMV 5' LTR sequences spanning the region between nts 397-417 shown in upper case; sequences shown in lowercase were incorporated for cloning purpose) along with the inner reverse (antisense; AS) primer that varied depending upon the mutation introduced. In PCR (B), the AS outer primer OTR788 (5'cccggatccTTCTTTCTTATCTATCAATTCTTTAATTAAG 3'; first 3 nt are dummies followed by *BamH*I site and MPMV Gag sequences spanning the region between nts 1171-1141 shown in upper case; sequences shown in lowercase were incorporated for cloning purposes) was used along with the inner S primer which also varied depending upon the mutations introduced.

Both inner S and AS primers that were employed in PCRs (A) and (B), respectively, were customized in a way that the resulting PCR products from these two separate reactions (PCR A and PCR B) would have overlapping complementary sequences. The presence of these complementary sequences in PCR products A and B allowed them to anneal in round 2 PCR when performed using outer S and AS primers (OTR787/OTR788), generating a final product harboring the mutation introduced. Following round 2 PCR, the final product, in addition to containing the desired mutation, should also acquire flanking distinctive endonuclease restriction sites namely: *BamH*I at the 5'-end and *Xho*I at the 3'-end of the product (Figure 2.2). The reason of introducing these single cutters (*BamH*I and *Xho*I) was to allow their cloning in a directed manner into the SJ2 vector backbone creating the mutant transfer vectors. Utilizing this methodology, various mutations were introduced into the SJ2 backbone and the mutations were verified by DNA sequencing (Tables 1;2).

The inner S and AS oligos that were designed and synthesized for introducing the specific mutations into SJ2 are described in the following sections along with the nature of the mutation introduced. In these oligos the underlined nucleotides represent complementary sequences of the tail that will not anneal to the template in the first round PCR but will allow the products from PCR A and PCR B to anneal during round 2 PCR (Figure 2.2).

Table 1: List of deletion/substitution mutations that have been introduced in sequences involved in the formation of LRI-I and LRI-II

Clone Name	X-I (U5)	Y-I (Gag)	Description of the Mutation
SJ2	**5' AUUUCUUGUGUCU 3'**	**3' UGAGGGGCACAGA 5'**	**Wild Type SHAPE-validated sequence**
RK1	**5' acucccgugucu 3'**	**3' UGAGGGGCACAGA 5'**	Substitution of 'U' in the upper strand X-I (U5) of 13nt LRI-I with 'C'
RK2	**5' acucccgugucu 3'***	**3' UGAGGGGCACAGA 5'**	Substitution of 'U' in the upper strand X-I and II (U5) of both LRIs with 'C'
RK3	Δ	**3' UGAGGGGCACAGA 5'**	Deletion of the upper X-I (U5) 13 nt of LRI-I
RK4	**5' AUUUCUUGUGUCU 3'**	Δ	Deletion of the lower Y-I (Gag)13 nt of LRI-I
RK5	Δ	Δ	Double deletion of 13 nt upper X-I (U5) and lower Y-I (Gag) strands of LRI-I
RK6	**5' aucucuuauauau 3'**	**3' UGAGGGGCACAGA 5'**	Substitution of the upper strand X-I (U5) of 13 nt LRI-I with heterologous sequence
RK7	**5' aucucuuauauau 3'**	**3' uagagaauauaua 5'**	Substitution of the lower Y-I (Gag) 13 nt LRI with complimentary heterologous sequence to RK6
Clone Name	**X-II (U5)**	**Y-II (Gag)**	**Description of the Mutation**
SJ2	**5' CUUGUUCC 3'****	**3' GAACGGGG 5'**	**Wild Type predicted sequence**
RK8	**5' agauagag 3'**	**3' GAACGGGG 5'**	Substitution of upper 8nt X-II (U5) LRI-II with heterologous sequence
RK9	**5'ggggcaag 3'**	**3' GAACGGGG 5'**	Substitution of upper 8 nt X-II (U5) of LRI-II with the lower 8 nt Y-II (Gag) of LRI-II sequence
RK10	**5' CUUGUUCC 3'**	**3' cuuguucc 5'**	Substitution of lower 8 nt Y-II (Gag) of LRI-II with upper 8 nt X-II (U5) of LRI-II sequence
RK11	**5' ggggcaag 3'**	**3' ccuuguuc 5'**	Substitution of upper 8 nt X-II (U5) of LRI-II with lower 8 nt Y-II (Gag) of LRI-II sequence complimentary to RK10
SJ2	**5' CUUGU^{U}CCC 3'*****	**3' GAACGGGG 5'**	**Wild Type SHAPE-validated sequence**
RK12	Δ	**3' GAACGGGG 5'**	Deletion of upper 9 nt X-II (U5) of LRI-II
RK13	**5' aggaggagg 3'**	**3' GAACGGGG 5'**	Substitution of upper 9 nt X-II (U5) of LRI-II with heterologous sequence
RK14	**5' aggaggagg 3'**	**3' uccuccucc 5'**	Substitution of lower 8 nt Y-II (Gag) of LRI-II with complimentary heterologous sequence to RK13

Table 2. Table outlining the deletion mutations introduced into the Gag sequences

Clone Name	Description of the Mutation
RK15	Deletion of 33 nt of SL3
RK16	Deletion of 30 nt of Gag SL2
RK17	Deletion of 39 nt of Gag SL1
RK18	Double Deletion of SL3 and Gag SL1
RK19	Double Deletion of SL3 and Gag SL2
RK20	Triple Deletion of SL3, Gag SL1 and Gag SL2

2.2.5. Transfer Vectors Containing Substitutions of Uridine (U) in G-U base pairs with Cytosine (C) in the Upper (X-I and X-II) sequence of both LRI-I and LRI-II

RK1: All Us involved in wobble base-pairing within the U5 sequence of the LRI-I were substituted with Cs in order to make a more stable LRI-I using the following primers:

OTR1157: S; 5' ACT CCC CGT GTC TCT TGT TCC CTT CAA TTC CCA C3', MPMV: 678-698nt.

OTR1158: AS; 5' GAA CAA GAG ACA CGG GGA GTG GAG ACA AGA CAG TGT TCT GA 3'; MPMV: 684-643nt.

RK2: All Us involved in wobble base-pairing within the U5 sequences of both LRI-I and LRI-II were substituted with Cs in order to make more stable LRIs using the following primers:

OTR1159: S; 5' ACT CCC CGT GTC TCT TGC CCC CTT CAA TTC CCA CTC CCT CC 3'; MPMV: 686-705 nt.

OTR1160: AS; 5' GAA GGG GGC AAG AGA CAC GGG GAG TGG AGA CAA GAC AGT GTT CTG 3'; MPMV: 664-645 nt.

2.2.6. Transfer Vectors Containing Deletions of Sequences Involved in U5-Gag LRI-I

RK3: Deletion of U5 sequence X-I (13 nt) of the LRI-I through SOE PCR using following primers:

OTR1190: S; 5' CTT GTT CCC TTC AAT TCC CAC 3', MPMV: 678nt-698nt.

OTR1191: AS; 5’ GGA ATT GAA GGG AAC AAG GGA GAC AAG ACA GTG TTC TGA T 3’ MPMV: 695-678 nt Δ 664-642 nt.

RK4: Deletion of Gag sequence Y-I (13 nt) of the LRI-I using following primers:

OTR1192: S; 5’ AAA GGT TAA ATA TGC TGA TCT TTT G 3’; MPMV: 961-985 nt.

OTR1193: AS; 5’ GAT CAG CAT ATT TAA CCT TTT TAA AGC CTG CTT CAA TTG TTC TA 3’; MPMV: 980-961 nt Δ 946-924 nt.

RK5: Deletion of both complementary sequences involved in the formation of U5-Gag LRI-I using the DNA of RK4 as template and OTR1190 and OTR1191 described above.

2.2.7. Transfer Vectors Containing Substitutions of Sequences Involved in U5-Gag LRI-I

RK6: Substitution of U5 sequence X-I (13 nt) of the LRI-I with heterologous sequence (5’-AUCUCUUAUAUAU-3’) using the following primers:

OTR 1153: S; 5’ ATC TCT TAT ATA TCT TGT TCC CTT CAA TTC CCA C 3’; MPMV: 678-698nt.

OTR 1154: AS; 5’ GAA CAA GAT ATA TAA GAG ATG GAG ACA AGA CAG TGT TCT G 3’; MPMV: 684-645 nt.

RK7: Substitution of Gag sequence Y-I (13 nt) of the LRI-I with a heterologous sequence (5'-ATATATAAGAGAT -3') complementary to the sequence introduced in RK6 to restore artificial complementarity using the following primers:

OTR1155: S; 5' ATA TAT AAG AGA TAA AGG TTA AAT ATG CTG ATC TTT TG 3'; MPMV: 961-985 nt.

OTR1156: AS; 5' CCT TTA TCT CTT ATA TAT TTA AAG CCT GCT TCA ATT GTT C 3'; MPMV: 965-926 nt.

2.2.8. Transfer Vectors Containing Substitutions of Sequences (8 nucleotides of the predicted structure) Involved in U5-Gag LRI-II

RK8: Substitution of U5 sequence X-II (8 nt of the predicted structure) of the LRI-II with heterologous sequence (5'-AGA UAG AG-3') using the following primers:

OTR1095: S; 5' AGA TAG AGC TTC AAT TCC CAC TCC CTC C 3'; MPMV 686-705 nt.

OTR1096: AS; 5' GAA TTG AAG CTC TAT CTA GAC ACA AGA AAT GGA GAC AAG 3'; MPMV 694-652 nt.

RK9: Substitution of U5 sequence X-II (8 nt) of the LRI-II with its complementary Gag sequence Y-II (5'-GGG GCA AG-3') to disrupt the complementarity and thereby LRI-II using following primers:

OTR1093: S; 5' GGG GCA AGC TTC AAT TCC CAC TCC CTC C 3', MPMV: 686-705 nt.

OTR 1094: AS; 5' GAA TTG AAG CTT GCC CCA GAC ACA AGA AAT GGA GAC AAG 3'; MPMV: 694-656 nt.

RK10: Substitution of Gag sequence Y-II (8 nt) of the LRI-II with U5 sequence maintaining 3' to 5' orientation thereby disrupting the complementarity using the following primers:

OTR1099: S; 5' CCT TGT TCA ATT AAG CCA GCA TGA ACG TTA TG 3' , MPMV: 900-924 nt.

OTR1100: AS; 5'GCT TAA TTG AAC AAG GAT ATC CGA GCG CAA CAC TTA C 3', MPMV: 891-871 nt.

RK11: Substitution of Gag sequence Y-II (8 nt) of the LRI-II with its complementary U5 sequence X-II (5'-CUU GUUCC-3') using RK9 (in which U5 sequence X-II has already been substituted with its complementary Gag sequence Y-II (5'-GGG GCA AG-3') to disrupt the complementarity as template DNA during PCR creating a double mutant to restore complementarity (by flipping the sequence) in LRI-II using the following primers:

OTR1101: S; 5'CTT GTT CC A ATT AAG CCA GCA TGA ACG TTA TG 3'; MPMV: 900-923 nt.

OTR1102: AS; 5'GCT TAA TT G GAA CAA GAT ATC CGA GCG CAA CAC TTA C 3'; MPMV: 907-900 nt 5'-GGA ACA G 3' 891-872 nt.

2.2.9. Transfer Vectors Containing Deletions of Sequences (SHAPE validated 9 nucleotides) Involved in U5-Gag LRI-II

RK12: Deletion of U5 sequence X-II (SHAPE validated 9 nt) of the LRI-II through SOE PCR using the following primers:

OTR1188: S; 5' TTC AAT TCC CAC TCC CTC CTC3', MPMV: 687-707 nt.

OTR1189: AS; 5’ GAG GGA GTG GGA ATT GAA AGA CAC AAG AAA TGG AGA CAA G 3’ , MPMV: 704-687 nt and 678-669nt.

2.2.10. Transfer Vectors Containing Substitutions of Sequences (SHAPE validated 9 nucleotides) Involved in U5-Gag LRI-II

RK13: substitution of U5 sequence X-II (9 nt of the SHAPE validated structure) of the LRI-II with heterologous sequence (5’-AGG AGG AGG-3’) using following primers:

OTR1178: S; 5’AGG AGG AGG TTC AAT TCC CAC TCC CTC CT 3’; MPMV: 687-707 nt.

OTR1179: AS; 5’ GAA CCT CCT CCT AGA CAC AAG AAA TGG AGA CAA G 3’; MPMV: 677-657 nt.

RK14: Substitution of Gag sequence Y-II (9 nt of the SHAPE validated structure) of the LRI-II with heterologous sequence complementary with the one introduced in RK13 using RK13 DNA as template and the following primers:

OTR1180: S; 5’ CCT CCT CCT AAT TAA GCC AGC ATG AAC GTT A 3’; MPMV: 900-920 nt.

OTR1181: AS; 5’ CTT AAT TAG GAG GAG GAT ATC CGA GCG CAA CAC TTA C3’; MPMV: 906-871 nt.

2.2.11 Transfer Vectors Containing Deletions of Sequences Involved in Forming Stem Loops at the 3’ end of the SHAPE Validated Structure

RK15: Deletion of the 33 nucleotides involved in forming the lower part of stem loop 3 (SL3) using the following primers:

OTR1079: S; 5’ CTC TCG AAA GTA AGT GTT GCG CTC GG 3’, MPMV: 829-8-33 and from 867-886 nt.

OTR1080: AS; 5’ CTT TCG AGA GTT TAC TTT CAC TTT TAA TC 3’, MPMV: 871- 867 and from 833-810 nt.

RK16: Deletion of the 30 nucleotides of Gag sequences involved in forming Gag stem loop 2 (complete deletion of Gag SL2) using the following primers:

OTR1085: S; 5’ GCT GAT ACT TGT CCT TGG TTT CCG C 3’; MPMV: 974-978 and from 1009-1028 nt.

OTR1086: AS; 5’ CAA GTA TCA GCA TAT TTA ACC TTT ACT CC 3’; MPMV: 1014-1009 and from 978-956 nt.

RK17: Deletion of the 39 nucleotides of Gag sequences involved in forming Gag stem loop 1 (complete deletion of Gag SL1) using the following primers:

OTR1194: S; 5’ TTA AAG ACA CGG GGA GTA AAG G 3’; MPMV: 944-965 nt.

OTR1195: AS; 5’ TAC TCC CCG TGT CTT TAA TAA TTC TTG CCC CAT ATC CGA G 3’; MPMV: 961-944nt and from 904-883nt.

RK18: Simultaneous deletion of sequences involved in forming the lower part of stem loop 3 (SL3) and Gag stem loop 1 (Gag SL1) using RK15 DNA as template and the following primers:

OTR1194: S; 5’ TTA AAG ACA CGG GGA GTA AAG G 3’; MPMV: 944-965 nt.

OTR1195: AS; 5’ TAC TCC CCG TGT CTT TAA TAA TTC TTG CCC CAT ATC CGA G 3’; MPMV: 961-944 nt and from 904-883 nt.

RK19: Simultaneous deletion of the sequences involved in forming the lower part of stem loop 3 (SL3) and Gag stem loop 2 (Gag SL2) using RK15 DNA as template and the following primers:

OTR1196: S; 5' TAC TTG TCC TTG GTT TCC GCA A 3'; MPMV: 1009-1030 nt.

OTR1197: AS; 5' GGA AAC CAA GGA CAA GTA TCA GCA TAT TTA ACC TTT ACT CCC 3'; MPMV: 1026-1009 nt and from 978-955 nt.

RK20: Simultaneous deletion of the sequences involved in forming the lower part of stem loop 3 (SL3), Gag stem loop 1 (Gag SL1), and Gag stem loop 2 (Gag SL2) using RK19 DNA as template and the following primers:

OTR1194: S; 5' TTA AAG ACA CGG GGA GTA AAG G 3'; MPMV: 944-965 nt.

OTR1195: AS; 5' TAC TCC CCG TGT CTT TAA TAA TTC TTG CCC CAT ATC CGA G 3'; MPMV: 961-944 nt and from 904-883 nt.

2.3. Three-plasmid *trans*-complementation assay

The U5-Gag LRI mutants were tested for both packaging and propagation employing a previously transcribed three plasmid *trans*-complementation assay developed by our group (Figure 2.3; Browning et al., 2001; Schmidt et al., 2003; Jaballah et al., 2010) to establish the biological significance of U5-Gag LRIs. As the name indicates, this assay utilizes three different plasmids to establish reliable experimental conditions to monitor the effects of U5-Gag LRIs mutants on both MPMV gRNA packaging and propagation. The three plasmids that were used in this assay included: 1) wild type SJ2 or the mutant transfer vector which contains

Figure 2.3: Illustration of the 3-plasmid *trans* complementation assay.

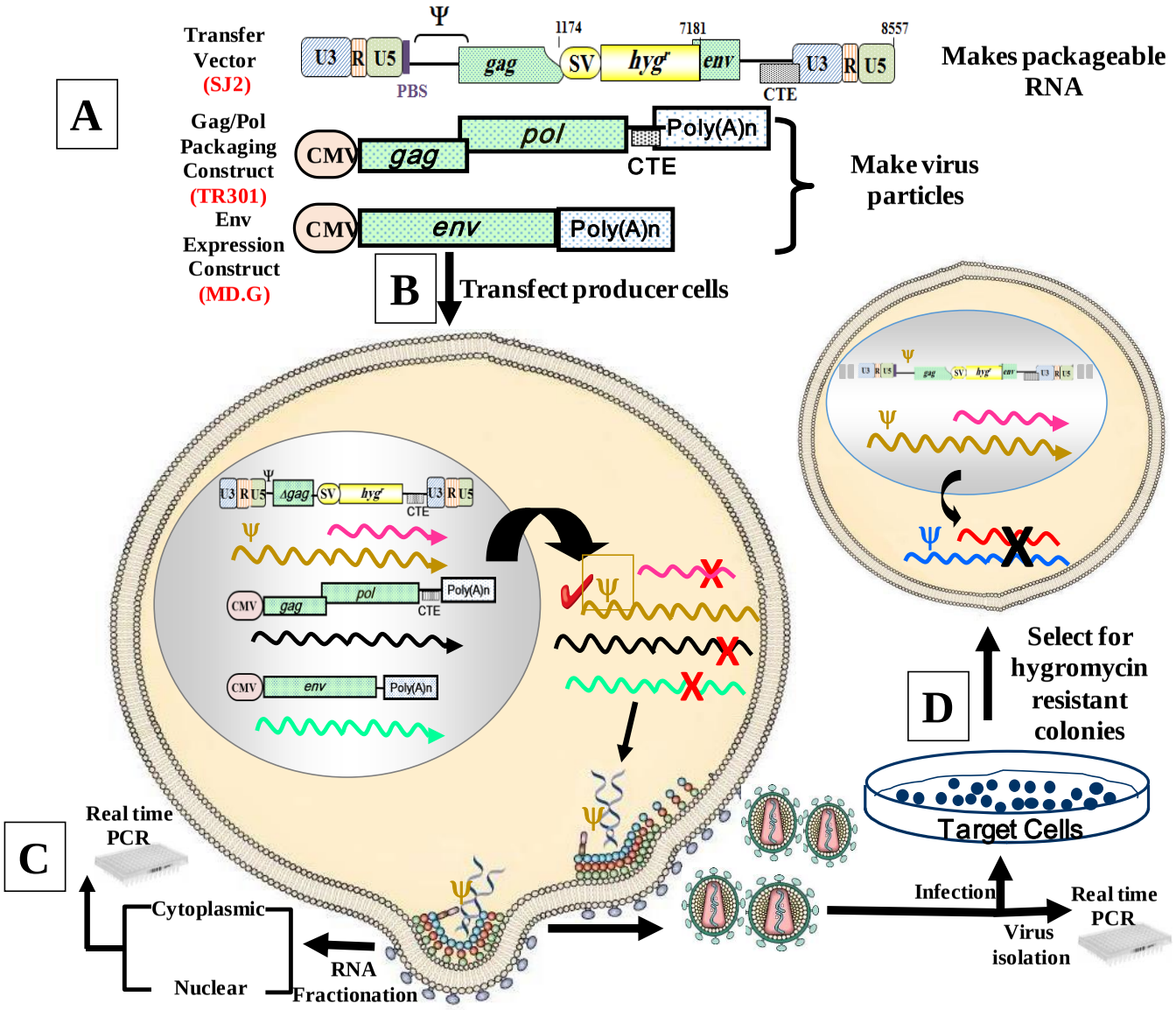


Figure 2.3. Illustration of the 3-plasmid *trans* complementation assay. A. Design o f the three plasmids that were transfected to produce virus particles. **B.** Cartoon depicting a 293T cell that has been co-transfected with the 3 plasmids to produce the replication-defective but infectious virus particles. Inside the cell, virus particles are made by Gag/pol proteins produced by the expression plasmid (TR301). The packageble RNA is produced by the MPMV wild type transfer vector (SJ2), while the vesicular stomatitis envelope expression construct (MD.G) produces the necessary envelope glycoprotein to pseudotype the virus particles and infect the target cells. The only RNA that can be packaged into the budding virus particles is the one that is generated from the transfer vector RNA since only that RNA contains an intact packaging signal. **C.** The transfected producer cells are processed and fractionated into nuclear and cytoplasmic fractions. The cytoplasmic fractions are used to study the cytoplasmic expression of the gRNA in the cells by RT-PCR followed by real time PCR. **D**. Once transfected, the 293T cell produce virus particles containing the packaged transfer vector RNA (SJ2). These particles are used to isolate the packaged viral RNA and quantitate relative RNA packaging in different mutants by RT-PCR followed by real time PCR. In addition, these particles are also used to infect target cells to study RNA propagation. After infection, target cells were selected with media containing hygromycin B antibiotic so that only those cells in which there had been a successful infection allowing them to express *hygromycin resistance* gene would survive. The number of the resulting hygromycin resistant colonies (colony forming unit/ml; CFU/ml) directly correlates with the amount of RNA packaged unless a step in the viral life cycle following RNA packaging has been affected. The virions produced by the 293T cells are restricted to one round of replication inside the target cells as they contain only the transfer vector RNA as their genome restricting the re-infection of the target cells, making the assay quantitative and sensitive. RNAs that cannot be packaged since they 1) do not carry packaging signal on themselves or 2) the cells do not carry RNA to produce viral proteins for their encapsidation. Figure partly adapted from D'Souza and Summers, 2005.

minimum *cis*-acting sequences required for RNA packaging, reverse transcription, and integration and functions as substrate to produce packageble RN, 2) packaging construct TR301 that expresses the MPMV the structural and enzymatic proteins, and 3) vesicular stomatitis envelope glycoprotein expression plasmid MD.G to pseudotype the viral particles produced containing the packaged RNA. An additional plasmid pGL3C (Promega, Madison, WI) which expresses the firefly *luciferase* gene was also added to the DNA cocktail to monitor the transfection efficiencies.

2.3.1. Transfection of Producer Cells

The producer 293T cells were seeded in six-well plates at a density of 4×10^5 cells per well and maintained at 37^0 C in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS from Hyclone (Logan, UT) a day prior to transfection and maintained in a humidified incubator with a constant supply of 5% CO2. After approximately 16 hours of seeding the cells, media was changed 2-4 hours before performing the transfections. Calcium phosphate transfections were performed by preparing a DNA cocktail with a total of 6μg DNA (2 μg of transfer vector + 2 μg of the envelope expression vector MD.G + 2 μg of the packaging construct TR301) + 250 ng of the pGL3C vector expressing the firefly *luciferase* gene. The DNA cocktail was prepared in a buffer containing 250 mM $CaCl_2$, 150 mM NaCl, 10 mM Tris-Cl, pH 7.4, and 1 mM EDTA. In order to form the calcium phosphate precipitate, a 2X transfection buffer (50 mM HEPES, 180 mM NaCl, 4 mM sodium phosphate, pH 7.0) was added drop-wise to the DNA cocktail while bubbling for approximately 2 minutes including, a 15 second vortex to increase the aeration in the tube. Finally, the precipitates were allowed to form

during a 30 minute incubation period at room temperature. The resulting precipitate was added drop-wise in a circular motion to the 293T plates and incubated for 4 hours at 37^0 C. At the end of the 4 hours incubation, transfected cells were washed twice, first wash with the media in the plate itself and the second round with phosphate-buffered saline (PBS). Lastly, 3 ml of fresh medium was added to each well and incubated overnight. The next day, the old media was removed and 1.5 ml fresh medium was added to each well. Lowering the media helps in concentrating the virus particles in the producer cell supernatant. Briefly, approximately 60 hours following transfection, pseudotyped virus particles produced were used to monitor transfer vector RNA packaging and propagation (Figure 2.3). To monitor the efficiency of RNA packaging, viral RNA was isolated from the virions produced in the culture supernatants and quantitated using real time quantitative PCR (qPCR; described in detail below).

2.3.2. Infection of Target Cells

Transfer vector RNA propagation of the packaged RNA was assessed by infecting HeLa T4 target cells with the virions produced by the 293T producer cells (Figure 2.3). RNA propagation was monitored by the successful transduction of the target cells by the *hygromycin resistance* gene present on the packaged transfer vector RNA. Briefly, HeLa T4 cells were maintained in DMEM supplemented with 7% calf serum from Hyclone (Logan, UT). They were seeded approximately 16 hours prior to infection at a density of $4x10^5$ cells/60mm plate. Supernatants from the transfected 293T producer cells were harvested approximately 60 hours post transfection, subjected to low speed centrifugation (2500 rpm) for 10 minutes to remove cellular debris and used to infect Hela T4 cells using DEAE-dextran as

described earlier (Browning et al., 2001). The cleared viral supernatant was also used to isolate virus particles to extract virion RNA (described in detail in section 2.3.5). Approximately 48 hours post infection, infected cultures were selected with media containing 200 µg/ml hygromycin B antibiotic for about 10-12 days following which hygromycin resistant colonies were stained with 0.5% crystal violet dye in 50% methanol as described earlier (Browning et al., 2001). The number of hygromycin resistant colonies represented as colony forming units per milliliter (CFU/ml) are directly proportional to the propagation efficiency of packaged transfer vector RNA (Browning et al., 2001). The colony counts were normalized to the transfection efficiency of different cultures as measured by luciferase expression (described in detail in section 2.3.1).

2.3.3. Luciferase Assay and the Transfection Efficiency:

Following transfection of the producer cells transfection efficiencies in the cultures were monitored by the Dual-luciferase Reporter Assay (Promega, Madison, WI). Briefly, after collecting the culture supernatants, 1 ml of DMEM containing10% FBS and 1 ml of cold PBS was added to each well and the cells were removed without trypsinization. Harvested cells were then washed once in 5 ml cold PBS, centrifuged at 1500 rpm for 5 min, and the resulting cellular pellet was resuspended in 1 ml of cold PBS. Next, 100 µl of cells ($1/10^{th}$ the volume) were pelleted and resuspended with 100 µl of 1X Passive Lysis Buffer, PLB, (Promega, Madison, WI). To prepare the cell lysate, resuspended cells were subjected to three freeze/thaw cycles of (each comprising of 2 minutes in dry ice-ethanol bath and 2 minutes at 37^0 C water bath) followed by microcentrifugation at 4^0 C for 5 minutes at 13k rpm to pellet down cellular debris. Clarified lysates were transferred to new

eppendorf tubes and were used in the Dual-reporter Luciferase Assay. For every 2µl lysate, 25µl of the Luciferase Assay Reagent II (LARII), was added, mixed by tapping gently, and the readings were taken in the Turner TD-20e Luminometer (Turner Design, Inc. Sunnyvale, CA) using a 5 seconds delay and 20 seconds integration time. In order to determine protein concentrations, the same cell lysates were tested in the Bradford Bio-Rad Protein Assay (BioRad, Hercules, CA) against a standard curve that was prepared from known concentrations of bovine serum albumin (BSA). The luciferase readings were then normalized to protein values in order to determine relative transfection efficiencies represented as luciferase activity per µg of protein. Finally, relative transfection efficiencies were used to normalize the transfer vector cytoplasmic RNA expression from the transfected cultures as well as the number of hygromycin resistant CFU/ml.

2.3.4. Nucleocytoplasmic RNA Fractionation from the Transfected Cells:

For nucleocytoplasmic fractionation, 700µl of the total 1 ml transfected cell culture was used by spinning in a microfuge for 2 minutes at 13000 rpm at 4^0 C. The cell pellet was then resuspended in cold fractionation buffer [diethylpyrocarbonate (DEPC)-treated RLN buffer (50 mM Tris pH8.0, 140 mM NaCl, 1.5 mM $MgCl_2$)] supplemented with 0.5% NP40. After 5 minutes of incubation on ice to gently lyse the cells without disrupting the nuclear membrane, lysates were spun down in a microfuge at 300g for 2 minutes and 90% of the supernatant containing the cytoplasmic fraction was carefully transferred to eppendorf tubes containing Trizol (Invitrogen Life Technologies) for subsequent RNA isolation.

2.3.5. Virus Isolation

The clarified supernatants containing virus particles from the transfected cultures were subsequently passed through 0.2-μm cellulose acetate syringe filters to ensure no cellular components were present. The maximum equal volume remaining in all the different supernatants were transferred to the bottom of ultra-clear centrifuge tubes (Beckman) containing a cushion of 2 ml 20% sucrose cushion. The tubes were topped up with fresh medium for balancing and ultracentrifuged in SW41 rotor at a speed of 26,000 rpm for 2 hours at 4^0 C. The resulting virus pellets were resuspended in 126 μl of TNE buffer (50 mM Tris-Cl, pH7.4, 100mM NaCl, and 1 mM EDTA, pH8.0) and lysed in 500 μl Trizol LS reagent (Invitrogen Life Technologies) containing 5 μl of polyacryl (Molecular Research Center, OH) as a carrier to isolate virion RNA.

2.4. RNA Isolation:

The cytoplasmic fractions and viral particles resuspended in Trizol and Trizol LS reagents, respectively, were used for RNA extraction. Towards this end, samples in Trizol were incubated for 5 minutes at room temperature and then extracted with 120μl chloroform. Phase separation was achieved by centrifugation at maximum speed in a microfuge (13000 rpm) for 15 min at 4^0 C. Next, the upper aqueous phase containing the RNA was transferred to the new tubes containing 600μl isopropyl alcohol and incubated at room temperature for 10 minutes. Samples were then centrifuged at maximum speed in a microfuge (13000 rpm) for 10 min at 4^0 C. The supernatant was decanted and resulting RNA pellets were washed with 1 ml of 70% ethanol. The tubes were mixed gently to loosen the pellet and centrifuged in a

microfuge at 10,000 rpm for 5 minutes at 4^0 C. After decanting the ethanol pellets were air dried. Finally, the pellets were resuspended in 100µl RNase-free water, vortexed and incubated at 55^0 C for 10 minutes to dissolve the pellets were stored at -80 ºC.

2.5 Reverse Transcriptase PCR (RT-PCR)

Before embarking for RT-PCR it was imperative that we ensure that our RNA preparations did not have any plasmid DNA that might have been carried over from transfected cultures. Towards this end, 25 µl viral RNA or 1-2 µg of cytoplasmic RNA fractions were DNase-treated using 2 units RQ1 RNase-Free DNase (Promega, Madison, WI) at 37^0 C for 30 minutes in a mixture containing 10X DNase Buffer and 20 units of Recombinant RNasin (Promega, Madison, WI). Reactions were then the reaction was stopped by adding the stop solution containing 20mM EGTA (Promega, Madison, WI) at 1X concentration to the reaction mixture and heat inactivating the samples at 65^0 C for 10 minutes in a shaking Thermomixer. DNase-treated RNA samples were subjected to 30 cycle PCR to test for any residual contaminating plasmid DNA, if any, using MPMV specific primer pair: OTR 1161 (S; 5' GAT CAG AAC ACT GTC TTG TC 3') and OTR 1163(AS; 5' CTT TCT TAT CTA TCA ATT CTT TAA 3').

Once confirmed that RNA preparations were clear of any DNA contamination, DNAse treated RNAs were reverse transcribed to make cDNA. Briefly, DNased-RNA samples were incubated with 3µl of 25mM dNTPs and 300 ng of random hexamers (OTR 603; 5' NNNNNN 3') for 5 minutes at 70^0 C, followed by quick cooling on ice for 5 minutes. Samples were then spun down and the cDNA synthesis was initiated by the addition of a reaction mixture containing

200 units of moloney murine leukemia virus (MoMuLV) reverse transcriptase (Promega, Madison, WI) in the presence of 40 units of Recombinant RNasin (Promega, Madison, WI) to inhibit RNAse activity for one hour at 37^0 C. cDNAs prepared from the cytoplasmic RNAs were tested for the integrity of the RNA fractionation process by amplifying 1µl of cDNA to look for the absence of unspliced β-actin mRNA which is exclusively found in the nucleus (Tan et al., 1995) using OTR582 (S; 5' CCAGTGGCTT CCCCAGTG 3') and OTR581 (AS; 5' GGCATGGGGGAGGGCATACC 3'). For this purpose, a multiplex PCR was performed using the OTRs 581/582 and the primers/competimers for 18S ribosomal RNA (Ambion, TX) as a control for the presence of amplifiable cDNA during the PCR. In order to further check that transfer vector RNAs were efficiently exported out of the nucleus, 1 µl of the cytoplasmic cDNA samples were amplified using an MPMV-specific transfer vector primer pair (OTRs 1161/1163) for 30 cycles. Similarly, cDNA preparations from viral RNA samples were also amplified using the same MPMV specific transfer vector primer pair (OTR 1161/1163) to qualitatively check the transfer vector RNA packaging.

2.6. Real time quantitative PCR (qPCR) for estimation of mutant RNA packaging efficiency

The cytoplasmic and viral cDNA samples were used to quantify the relative expression of the various transfer vector RNAs in the cytoplasm and their packaging efficiency into the nascently-produced virions. Quantification was performed by developing a Taqman quantitative gene expression real time PCR assay (Applied Biosystem). This custom-made assay employed a FAM-labelled probe along with primers within the U5/PBS region of MPMV, a region which was common to the

wild type and all the mutant transfer vector RNAs and away from any site of mutation. This resulted in restriction of where the assay could be designed which specifically consisted of a sense primer (5' CTCCTCCAGGTTCCTACTGTTGA 3'; nt 702 to 724) an antisense primer (5' TCGTATCCAGCCCCACGTT 3'; nt 770 to 752), and a FAM-labeled probe (5' TCGGGACAGTTGGC 3'; nt 734 to 747). The PCR efficiency of the MPMV assay was calculated using the online PCR Efficiency Calculator (http://srvgen.upct.es/index.html) which gave a predicted value of 1.86. This value fell slightly below the level suggested for the use of the ΔΔCT method (≥ 1.9). To ensure that we could use this method for the relative quantification purposes, the assay was empirically tested against the standard curve method (see Results for details) using test samples with varying amounts of MMTV RNA. This comparative analysis validated the use of the ΔΔCT method for the quantification purposes.

Thus, this method was used for all subsequent analyses and necessitated the use of an endogenous control. Towards this end, a pre-designed VIC-labelled human ß-Actin assay (Applied Biosystems #4326315E) was used. Equal amounts of cDNAs from the wild type and mutant samples were tested in triplicates for both the MPMV and β-Actin assays. A 20 µl PCR mix was prepared per sample containing 2µl cDNA and 10 ul of the Taqman Universal Master Mix (Applied Biosystems #4440045). The reaction was amplified for 50 cycles using the qPCR 7500 (Applied Biosystems Inc. CA USA) and analyzed as per protocol. The RQ values thus obtained for MPMV expression in both the cytoplasmic and viral samples were normalized to the luciferase values obtained per µg protein to control for differences in transfection efficiencies. Finally, to estimate the packaging efficiency, the results

obtained for MPMV expression in the virions were divided by those for the cytoplasmic expression and then reported relative to the wild type levels.

2.7. MPMV RNA secondary structure analyses *In Silico*

The 5' end of the MPMV 5’ end genome (region between R and the first 120 nt of Gag) was folded using the RNA folding software “Mfold” (Mathews et al., 1999; Zuker, 2003) to correlate the effects of the introduced mutations on MPMV packaging signal RNA secondary structure.

2.8. Statistical Analysis:

The statistical analysis was performed employing the standard paired 2-tailed student t-test between the wild type and the mutant clones to establish statistical significant differences. A *P*-value of 0.01 was considered to be significant.

Chapter 3: Results

The predicted as well as SHAPE-validated secondary structures of MPMV packaging signal RNA have shown two phylogenetically conserved LRIs (LRI-I and LRI-II) involving U5 and Gag sequences. Additionally, the entire region seems to be anchored by three stem-loops, namely SL3, Gag SL1, and Gag SL2 (Figure 1.13). However, neither the two LRIs nor the three stem-loops have been tested empirically for establishing their biological significance during MPMV RNA packaging and propagation processes. Therefore, to provide functional evidence for the existence of U5-Gag LRIs and the three stem loops, and validate their biological significance to the MPMV life cycle, a series of mutations were introduced and their effects on MPMV RNA packaging and propagation tested in a biological relevant trans-complementation assay.

3.1. Experimental approach and three plasmid *trans*-complementation assay to determine MPMV RNA packaging and propagation efficiencies

The sequences involved in formation of the U5-Gag LRIs fall in a region that has been earlier reported to be important for both MPMV RNA packaging and dimerization (Figure 1.12; Jaballah et al., 2010; Schmidt et al., 2003; Aktar et al., 2013). Therefore, it becomes difficult to test the effects of the introduced mutations in U5-Gag LRIs in the full-length genome context since it is very likely that mutations introduced in this region might affect the flanking packaging determinants. To overcome such a caveat, we took advantage of the three-plasmid *trans*-complementation assay developed earlier in our laboratory for MPMV that

provides the necessary biological components and the environment to generate virus particles containing the packaged RNA, the replication of which is limited to a single round because re-infection of the target cells cannot take place (Figure 2.3). The assay requires co-transfection of 293T producer cells with three different plasmids, two of which produce viral structural and enzymatic proteins Gag/Pol and Env in *trans* to produce viral particles, while the third plasmid (sub-genomic transfer vector) produces the RNA which works as a substrate to be packaged into the resulting virus particles. Such a scheme allowed us to introduce mutations in the U5-Gag LRIs as well as in SL3, Gag SL1, Gag SL2 and study their effects on RNA packaging without affecting the Gag/Pol ORFs since these proteins were provided in *trans* from a separate expression plasmid. Briefly, a wild type (SJ2) or mutant transfer vector, a Gag/Pol packaging construct (TR301), and an envelope expressing plasmid (MD.G), along with a firefly luciferase expression plasmid (pGL3C) were co-transfected into the producer 293T cells. The resulting virus particles were used to test the effect of the mutations on MPMV RNA packaging by quantifying the amount of packaged RNA, using real time PCR (qPCR). Virus particles were also used to monitor the propagation of the packaged transfer vector RNA in the infected HeLa T4 cells following their transduction with the marker *hygromycin resistant* gene. The number of hygromycin resistant colonies should be directly proportional to the packaged viral RNA content, providing an indirect estimate of RNA packaging efficiency (Figure 2.3).

3.2. Development of an MPMV custom-designed qPCR assay for measuring the relative packaging efficiency of mutant viral RNAs

To determine the effect of the mutations on MPMV RNA packaging, a quantitative qPCR assay needed to be developed that could be used to measure the packaging efficiency of all mutant vector RNAs irrespective of the mutation they displayed. This was achieved by designing a Taqman real time PCR assay that employed a FAM-labelled probe along with primers that bound within the U5/PBS region of MPMV (nt 702 to 770), a region that was common to the wild type as well as all the mutant transfer vector RNAs and away from any of the mutations introduced, ensuring 100% complementary binding efficiency of the primers and probe to the target sites. The PCR efficiency of the MPMV assay was calculated using the online PCR Efficiency Calculator (http://srvgen.upct.es/index.html) (Mallons et al., 2011) which gave a predicted value of 1.86, a value that fell slightly below the level suggested for the use of the ΔΔCT method (≥ 1.9). To ensure that we could use this method for the relative quantification purposes, the assay was empirically tested against the standard curve method using test samples with varying amounts of MPMV RNA.

The ΔΔCT relative quantification method requires > 90% efficiency of amplification of the custom assay since the values obtained from this data are normalized to the amplification of an internal endogenous control (ß-actin in our case) that has a guaranteed 100% efficiency of amplification. Therefore, the first step was to determine the amplification efficiency of the custom assay empirically. This was achieved by testing a series of 10-fold diluted MPMV plasmid starting from 100 pg up to 0.01 pg (5 points). Similarly, a 10-fold serial dilution of a

cytoplasmic cDNA sample expressing ß-actin was tested starting from 100 ng up to 0.01ng (5 points). The amplification efficiency curves obtained from the two assays were compared. The value of the slope of log input amount verses ΔCT should be approximately zero for the two assays to have similar amplification efficiencies. However, we observed a value of 1.83, suggesting that the custom made assay was on the borderline of accepted efficiency level. Therefore, we decided to empirically test whether the assay was valid for use in packaging efficiency calculations.

This was achieved by testing the level of MPMV expression on a set of 12 unknown samples with variable MPMV expression in the cytoplasm and virions by both the standard curve and ΔΔCT method. The standard curve method required running a standard curve of MPMV SJ2 plasmid DNA dilutions as well as dilutions of an actin expression plasmid alongside the unknown samples and estimating the relative amounts of MPMV and actin in the unknown samples compared to the values obtained from the two standard curves. The ΔΔCT method, on the other hand, did not need the two standard curves for estimating relative expression in the unknowns; instead, it used the endogenous β-actin expression to normalize for the amount of cDNA used in estimating the expression of MPMV in each sample relative to a calibrator sample that was "mock" cDNA in our case. Thus, the two distinct methods were used to estimate MPMV expression in the 12 unknown "cytoplasmic" and "viral" samples.

Figure 3.1 panels A and C show the results of MPMV expression in the cytoplasm, while panels B and D show the relative packaging efficiency in the virus particles, respectively, using the relative standard curve and ΔΔCT method. The relative packaging efficiency was calculated by normalizing the RQ values obtained

Figure 3.1: Validation of the custom-made MPMV real time PCR assay

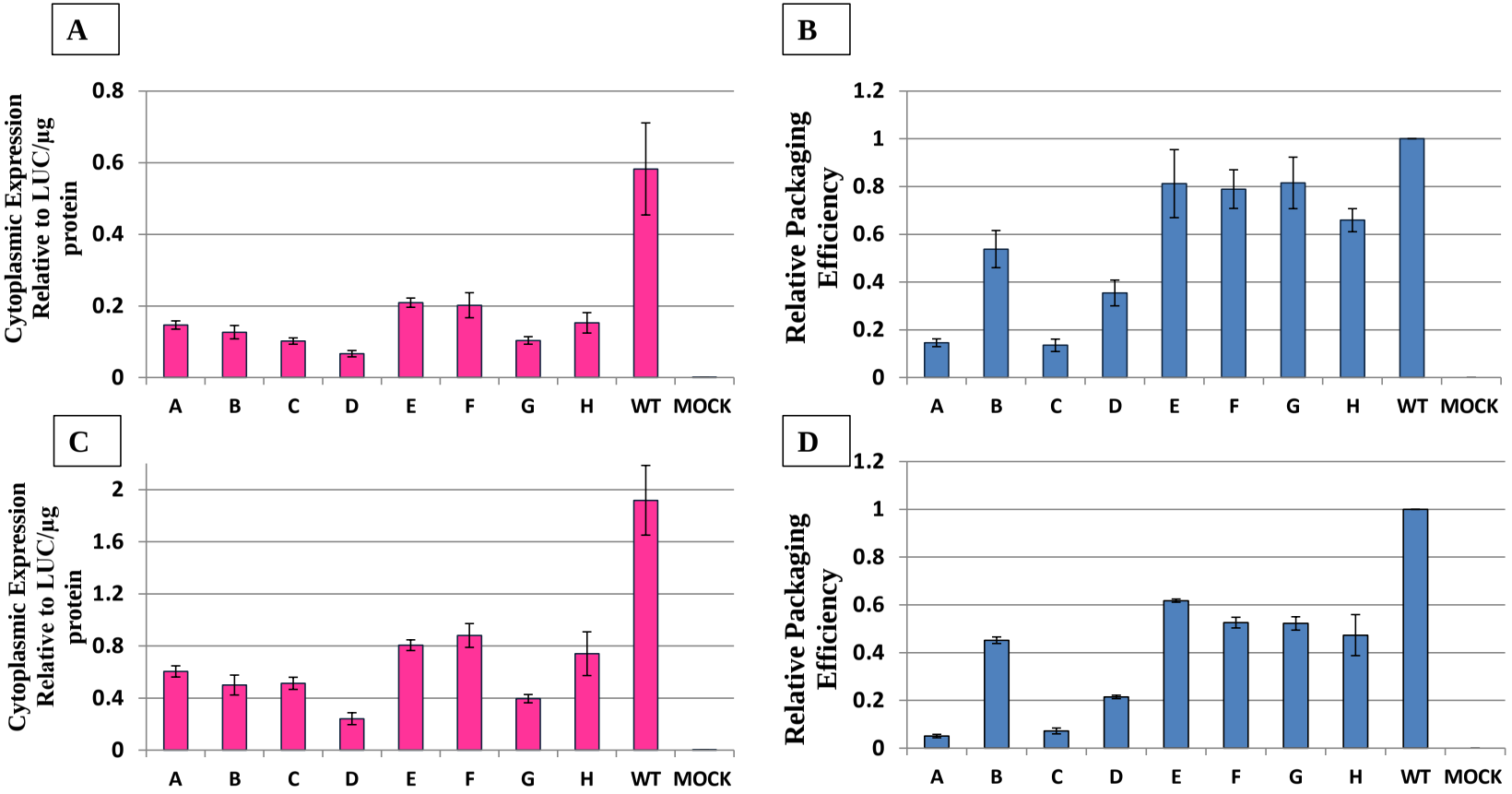


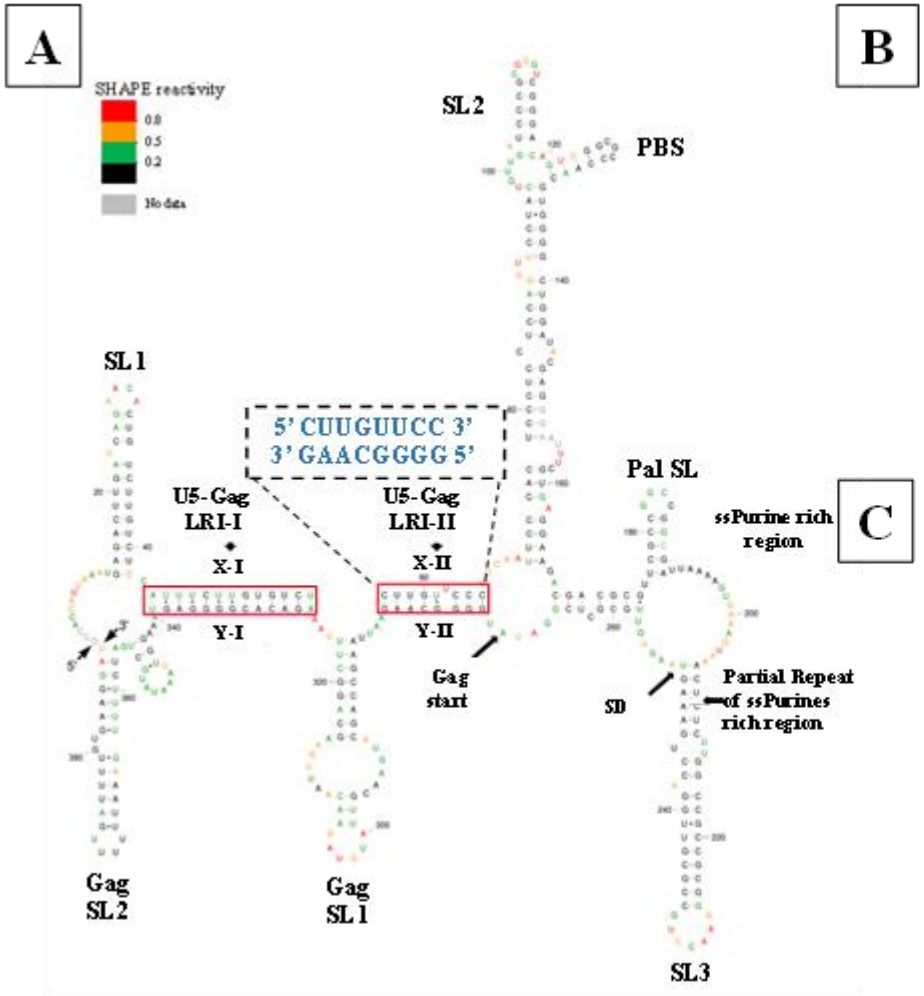
Figure 3.1. Validation of the custom-made MPMV real time PCR assay. In order to use the MPMV real time PCR assay to quantify the effect of the mutations on gRNA packaging, the assay validity was tested using two methods: **(A and B)** the relative standard curve method, and **(C and D)** the ΔΔCT method. The relative standard curve method involved running a standard curve where a 10-fold serial dilution of the MPMV wild type transfer vector plasmid, SJ2, was tested in triplicates as the standard curve for MPMV, while a similar 10-fold serial dilution was made of an SJ2-expressing cDNA as the standard curve for the endogenous β-actin assay. A set of unknown samples from A to H expressing MPMV RNAs at different levels in the cytoplasm **(panels A and C)** and packaged into virions **(panels B and D)** were tested by both methods in addition to SJ2 cDNA as a positive control and a mock sample serving as a negative control as well as the calibrator in the ΔΔCT method. The ΔΔCT method was used to estimate MPMV expression in the unknown samples relative to the endogenous β-actin expression. The relative fluorescence values (RQ values) thus obtained in the unknown samples were subsequently normalized to the transfection efficiency differences as measured by the luciferase assay. These values were used to calculate the packaging efficiency of the unknown samples by dividing the normalized RQ values obtained from the cytoplasmic expression of MPMV with that obtained for the MPMV expression in the virions. Both methods gave a similar pattern of relative expression in the unknown samples, revealing that the ΔΔ CT method could be used for calculating the RNA packaging efficiency of the virions. Finally, the results obtained for MPMV expression in the virions were normalized to those for the cytoplasmic expression and then reported relative to the wild type levels.

for MPMV expression in both the cytoplasmic and viral samples to the luciferase values obtained per ug protein to control for differences in transfection efficiencies. As can be seen, the relative pattern of expression estimated from the two very different analyses yielded very similar results, thus validating the assay for use in the estimation of mutant RNA packaging efficiency (Figure 3.1). Therefore, the ΔΔCT method was used in all subsequent estimations of MPMV expression rather than the standard curve method. This circumvented the need to generate a standard curve for both MPMV and actin each time an assay was done, greatly facilitating the analysis of RNA packaging efficiency.

3.3. Role of the wobble guanine-uracil (G-U) base-pairing in U5-Gag complementary sequences of LRI-I and LRI-II during MPMV RNA packaging and propagation

A distinguishing feature of the MPMV packaging signal RNA secondary structure is the presence of two long range interactions, LRIs (LRI-I and LRI-II) (Figure 3.2A). These LRIs are located within a region that has been shown to be phylogenetically conserved in different MPMV strains and important for MPMV RNA packaging and dimerization (Jaballah et al., 2010; Schmidt et al., 2003; Aktar et al., 2013). We hypothesized that these LRIs could potentially play a role in MPMV RNA packaging by maintaining the overall RNA secondary structure of the whole 5' UTR region that seems to be anchored by complementary and conserved U5 and Gag sequences, (Figure 3.2A). A careful analysis of the sequences involved in U5-Gag LRI-I revealed that out of the 13 complementary nucleotides, four base-paired nucleotides were present as non-Watson-Crick guanine-uracil (G-U) base-

Figure 3.2: Design and test of the deletion/substitution mutations introduced into the complementary U5-Gag LRI-I sequences.



B

Clone Name	X-I (U5)	Y-I (Gag)	Description
SJ2	5' AUUUCUUGUGUCU 3'	3' UGAGGGGCACAGA 5'	Wild type
RK1	5' acuccccgugucu 3'	3' UGAGGGGCACAGA 5'	G-U to G-C on X-I
RK2	5' acuccccgugucu 3' *	3' UGAGGGGCACAGA 5'	4 Us to 6 Cs on X-I
RK3	Δ	3' UGAGGGGCACAGA 5'	Δ X-I
RK4	5' AUUUCUUGUGUCU 3'	Δ	Δ Y-I
RK5	Δ	Δ	Δ X-I & Y-I
RK6	5' aucucuauauauu 3'	3' UGAGGGGCACAGA 5'	Substitution of X-I to disrupt the complementarity
RK7	5' aucucuauauauu 3'	3' uagagaauauauua 5'	Substitution of Y-I to restore the complementarity

C

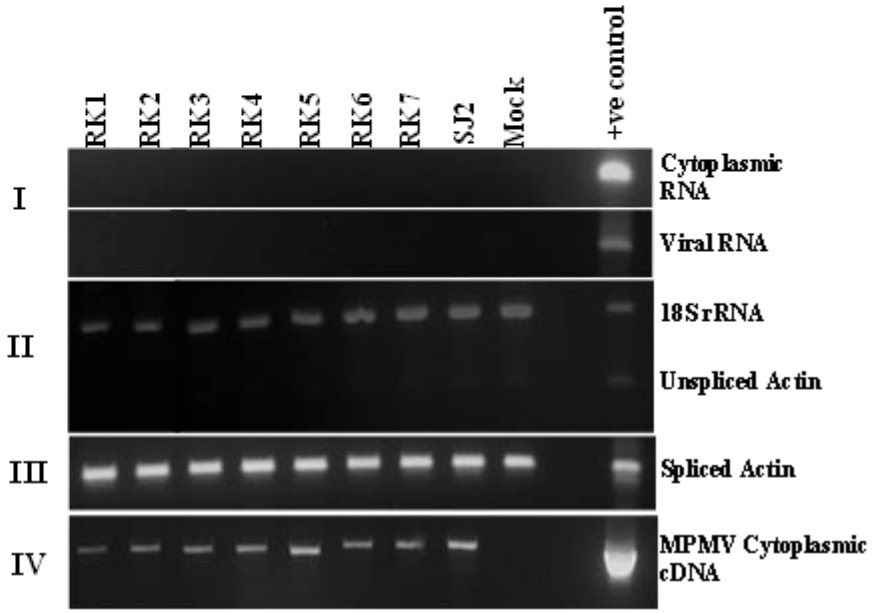


Figure 3.2. Design and test of the deletion/substitution mutations introduced into the complementary U5-Gag LRI-I sequences. (A) Illustration of the SHAPE-validated structure of the MPMV packaging signal RNA with long range interaction regions I and II (LRI-I and LRI-II) highlighted in red boxes. The predicted complementary 8 nucleotide U5-Gag sequence of LRI-II is shown in blue box with hatched lines. **(B)** Table outlining the deletion/substitution mutations introduced into the U5-Gag sequences of LRI-I. Sequences of the mutations introduced are represented in lower case and in red color. The mutant transfer vectors RK1 and RK2 contain LRI-stabilizing substitution mutations where the uridines were substituted with cytosines, forming wobble base pairs in LRI-I only in RK1 or LRI-I and LRI-II in RK2, respectively. RK3, RK4, and RK5 are deletion mutants in LRI-I as described in the table. RK6 contains a substitution of the X-I sequence with heterologous sequence so that the complementarity between U5 and Gag (X-I and Y-I) is lost. In RK7, an artificial LRI-I is re-established by substituting the Gag (Y-I) sequence with a complementary heterologous sequence to the one introduced in RK6. Asterisk (*) denotes that only part of the U5 sequence for both LRIs is shown due to space limitations. **(C)** Representative gel images of the controls needed for validating different aspects of the three plasmid *trans* complementation assay. **(I)** PCR amplification of DNase-treated RNA from the cytoplasmic (upper panel) and viral (lower panel) RNA preparations with MPMV-specific vector primers **(II)** multiplex amplification of unspliced β-actin mRNA and 18S rRNA and **(III)** PCR amplification of spliced β-actin mRNA to check for the nucleocytoplasmic fractionation technique **(IV)** PCR amplification of transfer vector cytoplasmic cDNAs using MPMV vector-specific primers.

pairs (Figure 3.2A). Since we have proposed that the complementary nature of the LRI sequences play a role in anchoring the RNA structure; therefore, we reasoned that further strengthening the base-pairing between the U5 and Gag sequences of the LRI-I should potentially make the structure more stable, resulting in enhanced RNA packaging. With this rationale in mind, we created a mutant transfer vector (RK1) in which the four uridine (U) residues in the U5 sequence (X-I) of the LRI-I involved in the low affinity G-U wobble base-pairing were substituted by cytosines (C) to form strong and stable Watson G-C base pairings (Figure 3.2A and 3.2B). Similarly, in the case of LRI-II where two out of the eight complementary nucleotides have the G-U (non-Watson-Crick, wobble base-pairing), we substituted two uracil (U) residues in the U5 sequence (X-II) of the LRI-II with cytosines (C), creating another mutant transfer vector (RK2). For creating RK2, we used the mutant clone RK1 DNA as a template to substitute the additional two uridine residues, and in doing so converted a total of six G-U (non-Watson-Crick, wobble base-pairs) into Watson stable G-C base pairs (Figure 3.2A and 3.2B).

To test the effects of the introduced mutations on RNA packaging, the wild type (SJ2) and mutant transfer vectors (RK1 and RK2) along with the MPMV packaging construct (TR301), the envelope expression plasmid (MD.G), in the presence of firefly luciferase expressing plasmid, pGL3, were co-transfected into 293T producer cells. Following the experimental protocol described in Materials and Methods, the viral supernatants from transfected cultures were harvested and used to isolate viral RNA to determine RNA packaging and to infect HeLa T4 cells in order to monitor the packaged vector RNA propagation.

Following transfection and prior to successful RNA packaging, it is imperative that mutant transfer vectors RNAs are stably expressed and efficiently

exported to the cytoplasm. Therefore, to ensure these crucial steps in viral life cycle, the transfected cells were fractionated into cytoplasmic and nuclear fractions. Next, RNAs were prepared from both the cytoplasmic fractions and the pelleted viral particles and analyzed by RT-PCR for the integrity of fractionation and MPMV expression. To exclude the possibility of any contaminating plasmid DNA that may have carried over from the transfected cultures, RNA preparations were treated with RNase-free DNase and PCR amplified for 30 cycles using MPMV vector-specific primers. A lack of any demonstrable amplification signal in DNase treated RNA samples suggested that any plasmid DNA contamination, in these RNA preparations was below the detection level (Figure 3.2C.I; upper panel cytoplasmic RNA; lower panel viral RNA). After having ensured this, cDNAs from the DNase-treated RNAs were prepared and integrity of the nucleo-cytoplasmic fractionation technique was tested by ensuring that no RNA physically leaked from the nucleus to the cytoplasm due to rupturing of the nuclear membrane during the fractionation process. This was achieved by testing for the presence of unspliced β-actin mRNA in the cytoplasmic fractions, an mRNA that should remain exclusively nuclear, while the spliced β-actin mRNA should be observed in both the fractions.

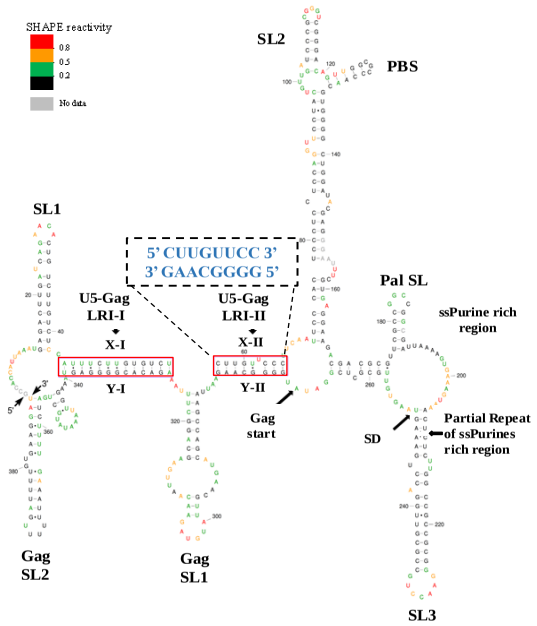
Test of the cytoplasmic cDNAs revealed that the unspliced β-actin mRNA could not be detected in any of the samples prepared even though it could be detected in the nuclear fraction, which also served as a positive control (Figure 3.2C.II). On the other hand, the spliced β-actin mRNA was observed in all the cytoplasmic fractions (Figure 3.2C.III). The absence of any amplifiable signal for unspliced β-actin mRNA which is exclusively found in the nucleus suggested that the nuclear membrane integrity was not compromised during fractionation and our RNA preparations were *bona fide* cytoplasmic (Figure 3.2C.II). Since our

interpretations were based on the inability of unspliced β-actin mRNA to be amplified from the cytoplasmic RNA fractions, it was important to ascertain that each sample during PCR contained amplifiable cDNAs. To establish this, unspliced β-actin amplifications were performed as a multiplex PCR in the presence of primers/competimer for 18S ribosomal RNAs as an internal control. Successful amplification of 18S ribosomal RNAs across all the samples validated the presence of amplifiable cDNAs (Figure 3.2C.II). Finally, cDNAs prepared from cytoplasmic RNA fractions were amplified using MPMV transfer vector-specific primers and the amplification signal across all the samples ensured that the transfer vector RNAs were efficiently and stably expressed and properly transported from the nucleus to the cytoplasm (Figure 3.2C.IV).

Having taken into consideration all the necessary controls, we next analyzed the packaging efficiencies of the transfer vector RNAs transcribed from RK1 and RK2 (containing substitutions in G-U base-pairing in the U5-Gag complementary sequences of LRI-I and LRI-II) in virus particles relative to the wild type (SJ2) transfer vector RNA. Towards this end, the MPMV custom-designed qPCR real time assay was used in combination with a commercially available β-actin Taqman assay as an endogenous control on both the cytoplasmic and virion RNA samples in triplicates as described earlier. The viral RNA packaging results obtained were further normalized to their cytoplasmic expression (Figure 3.3C.I) and the transfection efficiency (data not shown) to determine the packaging efficiency of RK1 and RK2. Next, the ratio of packaged mutant RNA was calculated relative to the wild type RNA to determine the relative packaging efficiency (RPE) of each mutant transfer vector RNA (Figure 3.3C.II). Such an analysis revealed that the RNA packaging efficiency of the LRI-1-stabilizing mutants (RK1 and RK2) was

Figure 3.3: Effect of the deletion/substitution mutations introduced into U5-Gag LRI-I on MPMV gRNA packaging and propagation

A



B

Clone Name	X-I (U5)	Y-I (Gag)	Description
SJ2	5' AUUUCUUGUGUCU 3'	3' UGAGGGGCACAGA 5'	Wild type
RK1	5' acucccgugucu 3'	3' UGAGGGGCACAGA 5'	G-U to G-C on X-I
RK2	5' acucccgugucu 3' *	3' UGAGGGGCACAGA 5'	4 Us to 6 Cs on X-I
RK3	Δ	3' UGAGGGGCACAGA 5'	Δ X-I
RK4	5' AUUUCUUGUGUCU 3'	Δ	Δ Y-I
RK5	Δ	Δ	Δ X-I & Y-I
RK6	5' aucucuuauauau 3'	3' UGAGGGGCACAGA 5'	Substitution of X-I to disrupt the complementarity
RK7	5' aucucuuauauau 3'	3' uagagaauauaua 5'	Substitution of Y-I to restore the complementarity

C

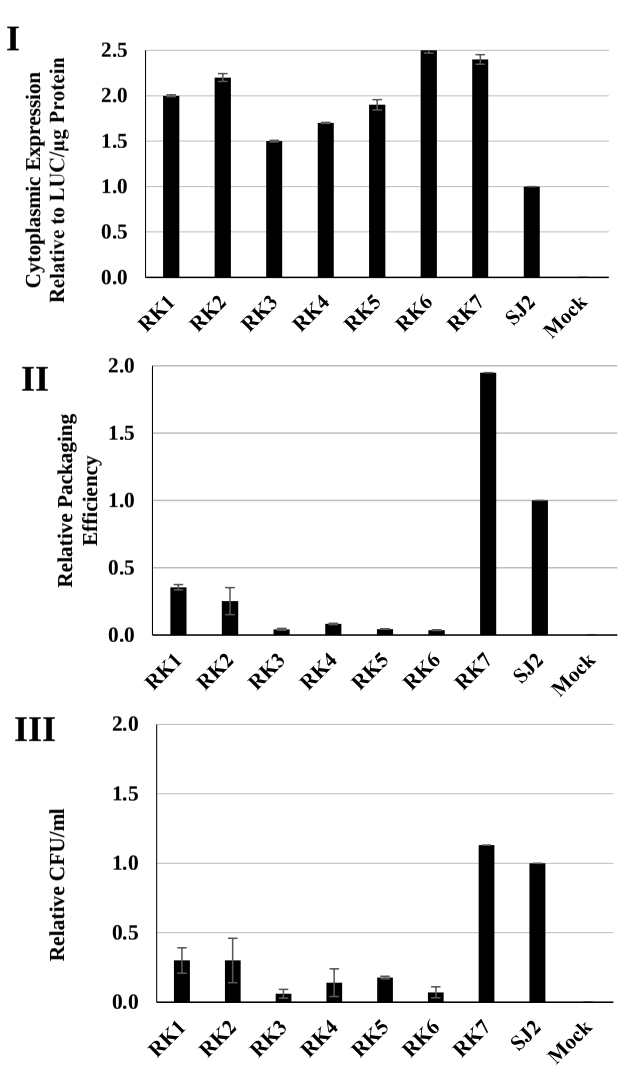


Figure 3.3. Effect of the deletion/substitution mutations introduced into U5-Gag LRI-I on MPMV gRNA packaging and propagation. **(A)** Illustration of the SHAPE-validated structure of the MPMV packaging signal RNA with long range interaction regions I and II (LRI-I and LRI-II) highlighted in red boxes. The predicted complementary 8 nucleotide U5-Gag sequence of LRI-II is shown in blue box with hatched lines. **(B)** Table outlining the deletion/substitution mutations introduced into the U5-Gag sequences of LRI-I. Sequences of the mutations introduced are represented in lower case and in red color. The mutant transfer vectors RK1 and RK2 contain LRI stabilizing substitution mutations where the uridines were substituted with cytosines, forming wobble base pairs in LRI-I only in RK1 or LRI-I and LRI-II in RK2, respectively. RK3, RK4, and RK5 are deletion mutants in LRI-I as described in the table. RK6 contains a substitution of the X-I sequence with heterologous sequence so that the complementarity between U5 and Gag (X-I and Y-I) is lost. In RK7, an artificial LRI-I is re-established by substituting the Gag (Y-I) sequence with a complementary heterologous sequence to the one introduced in RK6. Asterisk (*) denotes that only part of the U5 sequence for both LRIs is shown due to space limitations. **(C)** Relative cytoplasmic expression and packaging efficiencies of MMTV transfer vector RNAs as measure by the ΔΔCT method, and propagation efficiencies of LRI-I mutant transfer vectors. Panel **(I)** Cytoplasmic transfer vector RNA expression in 293T cells relative to the wild type (SJ2 vector) after normalization with the β-actin endogenous control and luciferase expression. Panel **(II)** mutant transfer vector RNA packaging efficiencies relative to the wild type (SJ2) after normalization with β-actin and luciferase expression. Panel **(III)** Relative hygromycin resistant colony-forming units (CFU)/ml for mutant transfer vectors reflecting the relative propagation efficiencies compared to the wild type SJ2 vector. The data represented in histograms correspond to the mean of the samples when tested in triplicates (± SD) following transfection and infection experiments.

affected when compared to the wild type transfer vector SJ2 (2.5- and 2.0-fold reduction in RPEs for RK1 and RK2, respectively; *P<0.01*; Figure 3.3C.II), in sharp contrast to our expectations. In agreement with the RNA packaging data, it was observed that these mutants were also proportionally defective for RNA propagation when compared to the wild type, SJ2 (Figure 3.3C.III). The RNA propagation defects in these mutants were observed by counting the hygromycin resistant colonies (colony forming unit/ml; CFU/ml) that appeared in the transduced HeLa T4 cells following infection with the virus particles containing *hygromycin resistance* gene on packaged transfer vector RNA. Such effects on RNA packaging and propagation were observed despite the fact that the transfection efficiencies for these mutants (RK and RK2) and wild type (SJ2) in multiple experiments were within 2-folds of each other (data not shown) and real time qPCR analysis of mutant transfer vector RNA showed steady-state levels of expression in the cytoplasm relative to the wild type, SJ2 (Figure 3.3C.I). Contrary to our predictions, this data suggest that further strengthening the base-pairing of U5-Gag LRIs adversely affects both RNA packaging and propagation.

3.4. Role of the U5-Gag complementary sequences in maintaining LRI-I structure during MPMV RNA packaging and propagation

To investigate the role of the complementary sequences involved in the U5-Gag LRI-I, a series of mutations were introduced in a fashion that either disrupted the complementarity between U5 and Gag sequences or completely deleted the complementary sequences (Figure 3.2A and 3.2B). In mutant transfer vectors RK3 and RK4, the complementarity between the U5 and Gag sequences was perturbed due to the deletion of U5 (X-I) and Gag (Y-I) complementary sequences,

respectively, while RK5 contained a double deletion in which both the U5 (X-I) and its complementary Gag (Y-I) sequences involved in forming the LRI-I were deleted simultaneously (Figure 3.2A and 3.2B). Such drastic double mutations were introduced to determine whether MPMV could package its RNA in the absence of LRI-I while maintaining only LRI-II. These mutant transfer vectors (RK3-RK5) were tested in our *in vivo* packaging and propagation assay to assess the effects of the introduced mutations on MPMV RNA packaging and propagation.

After having confirmed the absence of any contaminating plasmid DNA in our cytoplasmic and viral RNA preparations (Figure 3.2C.I), we ensured that nuclear membrane integrity was maintained during cytoplasmic RNA fractionation, that cDNA preparations were amplifiable (Figure 3.2C.II and III), and the transfer vector RNAs were stably expressed and efficiently exported to the cytoplasm (Figure 3.2C.II and IV). Next, results obtained from qPCR real time assay were calculated to determine the packaging efficiencies of RK3-RK5 relative to the wild type (SJ2) levels following normalization to the cytoplasmic transfer vector RNA expression (Figure 3.3C.I) as well as to the transfection efficiency (data not shown). Results obtained following these analyses demonstrated that deletion of either one (U5-X-I in RK3 or Gag-Y-I in RK4) or both complementary sequences (RK5) involved in LRI-I almost abolished RNA packaging (>10-folds compared to the wild type; Figure 3.3C.II). Consistent with the RPEs of RK3-RK5, the propagation efficiencies of these transfer vector RNAs were also proportionally reduced (Figure 3.3C.III). Data obtained from these biological tests suggested that maintenance of LRI-I involving complementary U5 and Gag sequences is crucial for MPMV RNA packaging. These results also suggest that the mere presence of only LRI-II is not sufficient to encapsidate MPMV RNA in the budding virus particles.

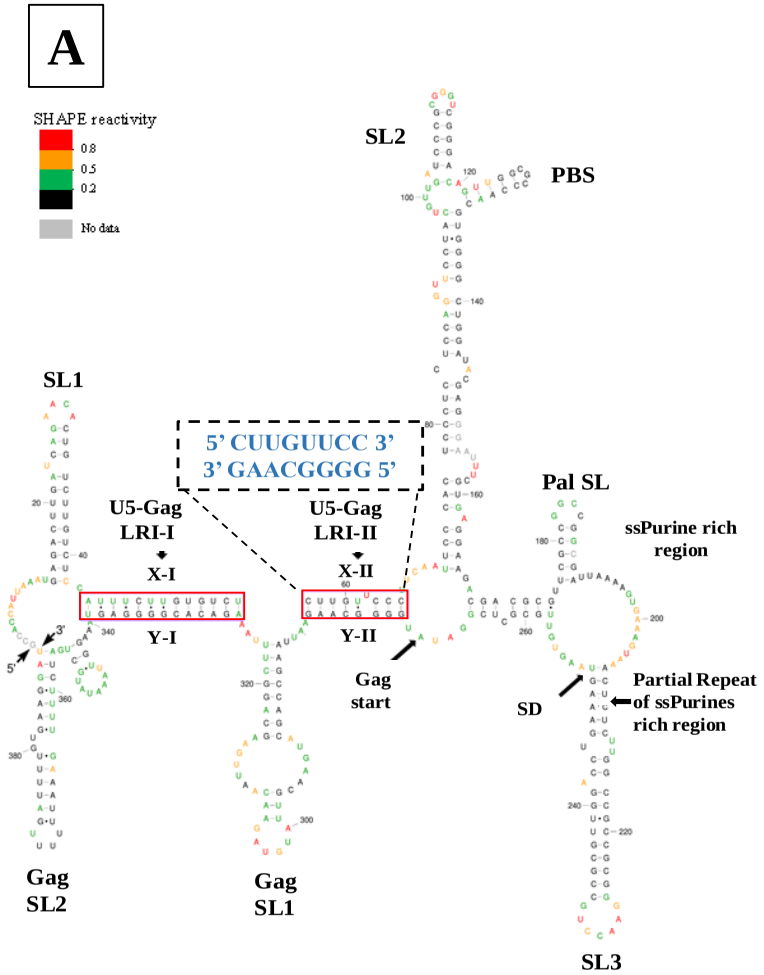
In light of the above observations, it was necessary to establish whether it is the primary sequence of the complementary U5 and Gag nucleotides important for RNA packaging or the non-viral heterologous nucleotides keeping the complementarity between U5 and Gag sequences would be sufficient to maintain the LRI-I and therefore RNA packaging. Thus, two mutants were generated to address these possibilities. In the RK6 mutant clone, one side of the complementary sequence (U5; X-I) was substituted with heterologous sequences of equal length (5' AUCUCUUAUAUAU 3') so that its complementarity with the sequence on the other side (Gag; Y-I) was lost (Figure 3.2A and 3.2B). Specifically, substitution of U5 with such a heterologous sequence resulted in loss of base-pairing for four nucleotides, four nucleotides assumed non-Watson-Crick, wobble (G-U) base-pairing, and five nucleotides maintained their wild type base-pairing with the Gag (Y-I) sequence. We expected this mutant to display severely compromised RNA packaging and propagation if the LRI was required for stabilizing the overall higher order structure and consequently the function of packaging signal RNA sequences. Next, RK7 was constructed that contained a compensatory mutation in which the Gag (Y-I) sequence in RK6 was substituted with the sequence 5' AUAUAUAAGAGAU 3' complementary to the heterologous sequence substituted in RK6, thus restoring the artificial/heterologous LRI-I in place of the original U5-Gag LRI (Figure 3.2A and 3.2B). Consistent with the deletion mutants RK3-RK5, disruption of the LRI-I complementarity in RK6 (due to the substitution of U5 (X-1) sequence with a heterologous one) almost abolished RNA packaging (> 20 fold reduction; $P<0.01$ compared to the wild type SJ2; Figure 3.3C.II). Consistent with our assumption, restoration of LRI-I with heterologous sequences in RK7 not only restored RNA packaging, but it was observed to be more than the wild type levels

(Figure 3.3C.II). The low level of RNA packaging in the case of RK6 corroborated well with the reduced RNA propagation observed for this mutant (25 fold reduction; $P<0.01$ compare to the wild type SJ2; compare Figure 3.3C.II with Figure 3.3C.III). Similarly, restoration of RNA packaging with the re-establishment of LRI-I with heterologous sequences in the case of RK7 also restored RNA propagation of this mutant to wild type levels (compare Figure 3.3C.II with Figure 3.3C.III). These results biologically validate the existence of LRI-I and further suggest its role at the structural level rather than at the primary sequence level since re-establishment of artificial/heterologous LRI-I at its native location restored RNA packaging and propagation to wild type levels.

3.5. Role of the U5-Gag complementary sequences in maintaining LRI-II during MPMV RNA packaging and propagation

A similar mutational approach was employed towards investigating the role of U5-Gag complementary sequences involved in forming LRI-II. Thus, a series of substitution, deletion, and compensatory mutations were first introduced in the eight nucleotide U5 and Gag region involved in forming LRI-II and tested for their effect on MPMV RNA packaging and propagation. In the mutant transfer vector RK8, all eight nucleotides within the U5 (X-II) that were predicted to be paired with Gag (Y-II) sequence were substituted with heterologous eight nucleotides (5’ AGAUAGAG 3’) to lose the U5-Gag complementarity responsible for maintaining LRI-II (Figure 3.4A and 3.4B). In the substitution mutant RK9, the U5 (X-II) sequences were replaced by its complementary sequences in Gag (Y-II), thus duplicating the sequence of Gag (Y-II) on both ends of LRI-II, resulting in the disruption of U5-Gag complementarity in six out of the eight nucleotides (Figure 3.4A and 3.4B). Mutant

Figure 3.4: Design and test of the deletion/substitution mutations introduced into the U5-Gag LRI-II sequences.



B

Clone Name	X-II (U5)	Y-II (Gag)	Description
SJ2	5' CUUGUUCC 3'*	3' GAACGGGG 5'	Wild type, predicted 8 nt LRI-II
RK8	5'agauagag 3'	3' GAACGGGG 5'	Substitution of X-II to disrupt complementarity
RK9	5'ggggcaag 3'	3' GAACGGGG 5'	Substitution of X-II with Y-II to disrupt complementarity
RK10	5' CUUGUUCC 3'	3' cuuguucc 5'	Substitution of Y-II with X-II to disrupt complementarity
RK11	5' ggggcaag 3'	3' ccuuguuc 5'	Swapping of X-II & Y-II to restore complementarity
SJ2	5' CUUGU^{U}CCC 3'**	3' GAACGGGG 5'	Wild type, SHAPE-validated 9 nt LRI-II
RK12	Δ	3' GAACGGGG 5'	Δ X-II
RK13	5' aggaggagg 3'	3' GAACGGGG 5'	Substitution of X-II to disrupt the complementarity
RK14	5' aggaggagg 3'	3' uccuccucc 5'	Substitution of Y-II to restore the complementarity

C

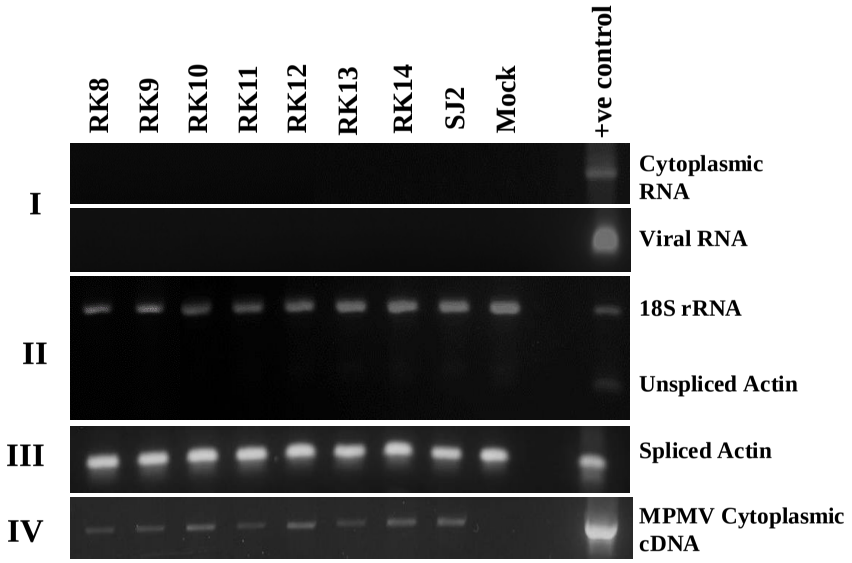
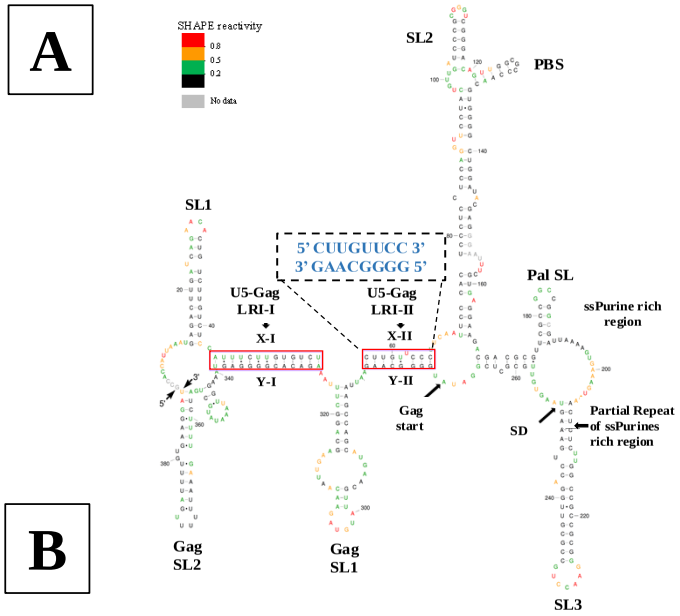


Figure 3.4. Design and test of the deletion/substitution mutations introduced into the U5-Gag LRI-II sequences. **(A)** Illustration of the SHAPE-validated structure of the MPMV packaging signal RNA with long range interaction regions I and II (LRI-I and LRI-II) highlighted in red boxes. The predicted complementary 8 nucleotide U5-Gag sequence of LRI-II is shown in blue box with hatched lines. **(B)** Table outlining the deletion/substitution mutations introduced into the U5-Gag sequences of LRI-II. Sequences of the mutations introduced are represented in lower case and in red color. Two sets of mutants are outlined in the table. The first set of mutations were introduced into the Mfold- predicted 8 nt LRI-II structure, while the second set describes mutations introduced into the SHAPE-validated LRI-II structure. The mutant transfer vectors RK8 and RK9 contain substitutions mutations in the X-II sequence, while RK10 contains substitutions mutations in the Y-II sequence. These substitutions caused a loss of complementarity between U5 and Gag (X-II and Y-II). In RK11, an artificial LRI-I was re-established by flipping the original LRI-II strand sequences while still maintaining the wild type complementarity. In RK12, the SHAPE-validated X-II sequence was deleted completely. In RK13, the deleted sequence in RK12 was substituted with heterologous sequences which resulted in the loss of complementarity. In RK14, the complementarity of LRI-II was restored by substituting the Y-II sequence with heterologous sequences complementary to the X-II sequence of RK13. Asterisk (*) denotes the 8 nucleotide U5 sequence of the predicted LRI-II. Asterisks (**) denote the 9 nucleotide U5 sequence of the SHAPE validated LRI-II. **(C)** Representative gel images of the controls needed for validating different aspects of the three plasmid *trans* complementation assay. **(I)** PCR amplification of DNase-treated RNA from the cytoplasmic (upper panel) and viral (lower panel) RNA preparations with MPMV-specific vector primers **(II)** Multiplex amplification of unspliced β-actin mRNA and 18S rRNA and **(III)** PCR amplification of spliced β-actin mRNA to check for the nucleocytoplasmic fractionation technique **(IV)** PCR amplification of transfer vector cytoplasmic cDNAs using MPMV vector-specific primers.

transfer vector RK10 was designed in a fashion that the U5 (X-II) sequence was duplicated in place of Gag (Y-II) sequence while maintaining the 5' to 3' polarity. Such a strategy resulted in creating the U5 (X-II) sequence (5' CUUGUUCC 3') on both ends of LRI-II (X-II and Y-II) in the same orientation. Thus, none of the eight nucleotides maintained any complementarity with the complementary strand (Figure 3.4A and 3.4B). These substitution mutants (RK8-RK10) were tested in the *in vivo* packaging assay, and the RPEs of these mutant transfer vectors were calculated after taking into consideration all the necessary controls (Figure 3.4C.I-IV). Consistent with our earlier observations as in the case of LRI-I mutants, RNA packaging of these mutants, designed to disrupt the complementarity of U5 and Gag sequences of LRI-II, was observed to be nearly ablated (Figure 3.5C.II). In good concordance with the RNA packaging data, the RNA propagation of these mutant transfer vectors was found to be severely compromised (Figure 3.5C.III).

In order to determine whether the primary sequence of the eight nucleotides in U5 and Gag was essential in its native context in maintaining the predicted LRI-II for RNA packaging, or any heterologous sequence maintaining the complementarity at this location would be sufficient to augment RNA packaging, we created a double complementary mutant, RK11 (Figure 3.4A and 3.4B). RK11 was created from the RK9 mutant clone which already contained the substitution of U5 (X-II) sequence with its complementary sequence in Gag (Y-II). RK9 was used as a template to substitute the original Gag (Y-II) sequence with the U5 (X-II) sequence. The resulting double mutant (RK11) restored the complementarity between the two LRI-II strands (X-II and Y-II) with a 180 degrees flipped orientation, re-establishing an artificial LRI-II structure (Figure 3.4A and 3.4B), similar to RK7 in LRI-1.

Figure 3.5: Effect of the deletion/substitution mutations introduced into U5-Gag LRI-II in MPMV gRNA packaging and propagation



B

Clone Name	X-II (U5)	Y-II (Gag)	Description
SJ2	5' CUUGUUCC 3'*	3' GAACGGGG 5'	Wild type, predicted 8 nt LRI-II
RK8	5'agauagag 3'	3' GAACGGGG 5'	Substitution of X-II to disrupt complementarity
RK9	5'ggggcaag 3'	3' GAACGGGG 5'	Substitution of X-II with Y-II to disrupt complementarity
RK10	5' CUUGUUCC 3'	3' cuuguucc 5'	Substitution of Y-II with X-II to disrupt complementarity
RK11	5' ggggcaag 3'	3' ccuuguuc 5'	Swapping of X-II & Y-II to restore complementarity
SJ2	5' CUUGU^{U}CCC 3'**	3' GAACGGGG 5'	Wild type, SHAPE-validated 9 nt LRI-II
RK12	Δ	3' GAACGGGG 5'	Δ X-II
RK13	5' aggaggagg 3'	3' GAACGGGG 5'	Substitution of X-II to disrupt the complementarity
RK14	5' aggaggagg 3'	3' uccuccucc 5'	Substitution of Y-II to restore the complementarity

C

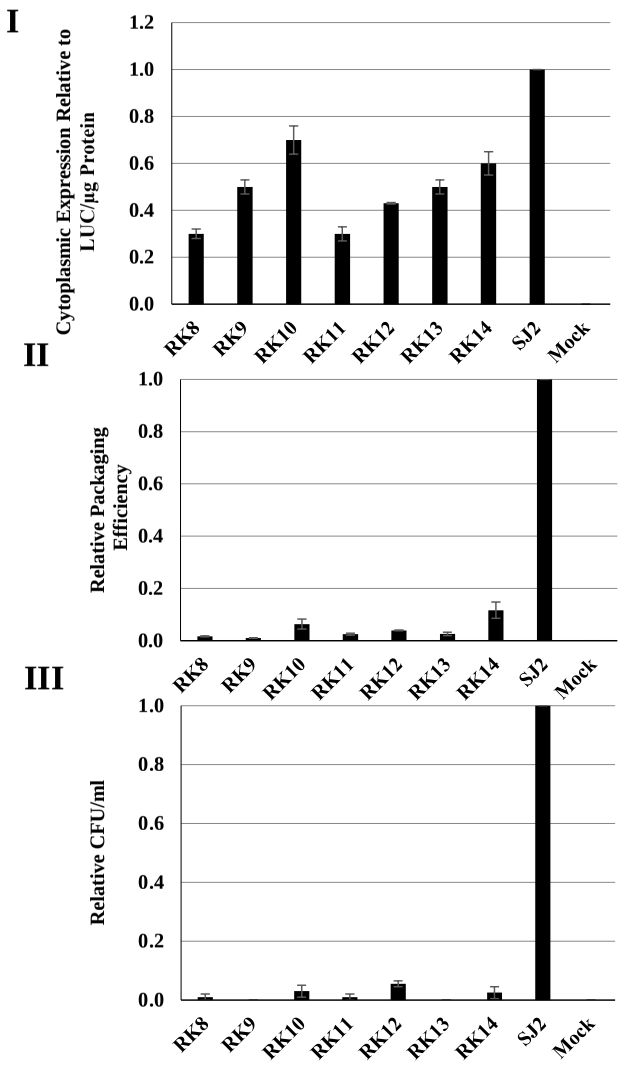


Figure 3.5. Effect of the deletion/substitution mutations introduced into U5-Gag LRI-II in MPMV gRNA packaging and propagation. (A) Illustration of the SHAPE-validated structure of the MPMV packaging signal RNA with long range interaction regions I and II (LRI-I and LRI-II) highlighted in red boxes. The predicted complementary 8 nucleotide U5-Gag sequence of LRI-II is shown in blue box with hatched lines. **(B)** Table outlining the deletion/substitution mutations introduced into the U5-Gag sequences of LRI-II. Asterisk (*) denotes the 8 nucleotide U5 sequence of the predicted LRI-II. Asterisks (**) denote the 9 nucleotide U5 sequence of the SHAPE validated LRI-II. **(C)** Relative cytoplasmic expression and packaging efficiencies of MMTV transfer vector RNAs as measure by the ΔΔCT method, and propagation efficiencies of LRI-II mutant transfer vectors. Panel **(I)** Cytoplasmic transfer vector RNA expression in 293T cells relative to the wild type (SJ2 vector) after normalization with the β-actin endogenous control and luciferase expression. The error bars represent the standard deviation (SD) of triplicates of each clone. Panel **(II)** mutant transfer vector RNA packaging efficiencies relative to the wild type (SJ2) after normalization with β-actin and luciferase expression. Panel **(III)** Relative hygromycin resistant colony-forming units (CFU)/ml for mutant transfer vectors reflecting the relative propagation efficiencies compared to the wild type SJ2 vector. The data represented in histograms correspond to the mean of the samples when tested in triplicates (± SD) following transfection and infection experiments.

Test of RK11 in the *in vivo* packaging and propagation assay revealed that it could not restore RNA packaging and propagation of the mutant vector RNA (Figure 3.5C.II and III), an observation in sharp contrast to that in the case of LRI-I mutant RK7, a double compensatory mutant designed to re-establish artificial/heterologous LRI-I; Figure 3.3C.II and III). These results were rather unexpected, especially given the fact that RK11 retains the primary viral sequences involved in forming LRI-II, however, in an opposite orientation, thus re-establishing an artificial LRI-II. These data therefore suggest that the predicted primary eight nucleotide complimentary sequence of U5 (X-II) and Gag (Y-II) LRI-II in its native context is vital for RNA packaging.

It is worth pointing out that the U5 and Gag sequences involved in forming LRI-II differed somewhat between the predicted and SHAPE-validated secondary structure of MPMV packaging signal RNA (Aktar et al., 2013). Specifically, a uridine residue at position 62 (U62) maintains the G-U wobble base pairing in the predicted structure, whereas in the SHAPE-validated structure, it did not assume the G-U wobble base pairing since it was highly reactive to SHAPE reagents and consequently formed a bulge in the U5 sequence (X-II) of LRI-II (Figure 3.4A; Jaballah et al., 2010; Aktar et al., 2013). In addition, an extra cytosine at position 65 (C65) that was not originally predicted to be part of the U5 (X-II) complementary sequences was included in the SHAPE-validated LRI-II structure and base-paired with a complementary guanine (G) in the complementary Gag (Y-II) sequence. Such a single nucleotide shift made the U5 (X-II) sequences involved in forming LRI-II to be of nine nucleotides instead of eight nucleotides as initially predicted (Jaballah et al., 2010; Aktar et al., 2013; Figure.3.4A). To investigate the role of U5

and Gag sequences of LRI-II in MPMV RNA packaging, we initially introduced mutations in the predicted structure of LRI-II involving complementary eight nucleotides as described above. However, since the SHAPE-validated structure was only recently published, we introduced another set of mutations in the SHAPE-validated nine nucleotides of U5 (X-II) sequences of LRI-II (Figure.3.4A).

Therefore, to further confirm results obtained with the predicted eight nucleotide LRI-II, we created a few more mutants in the nine nucleotide SHAPE-validated LRI-II structure that contained one unpaired U in the U5 (X-II) complementary to eight nucleotides in the Gag (Y-II) sequences (Figure.3.4A). In an attempt to disrupt the complementary U5 (X-II) and Gag (Y-II) sequences, all the nine nucleotides of U5 (X-II) sequence were deleted in RK12 (Figure 3.4A and 3.4B). Consistent with the mutational analysis with the predicted eight nucleotide clones RK8 and RK9, test of RK12 (designed to disrupt the SHAPE-validated LRI-II complementarity) revealed that both RNA packaging and propagation were almost abrogated when compared to the wild type (SJ2; Figure 3.5C.II and III), further confirming the importance of LRI-II to the packaging potential of MPMV.

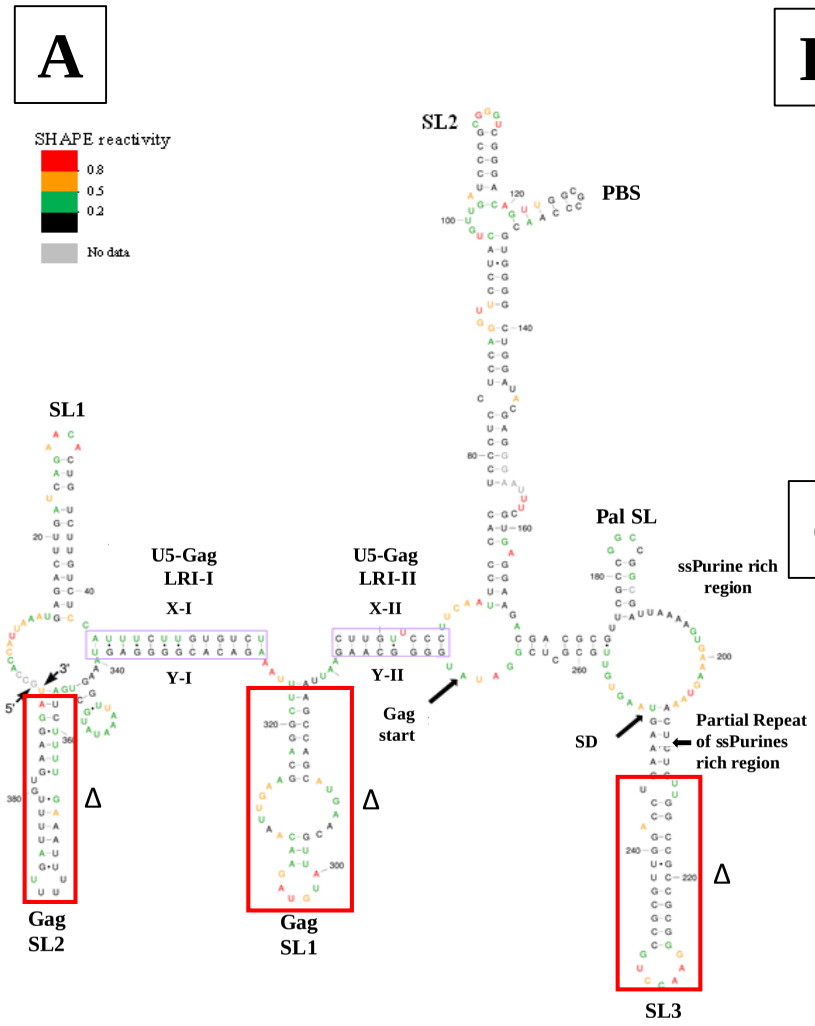
Finally, to examine the importance of complementarity between the 9 nucleotides of U5 and Gag sequences involved in establishing the SHAPE-validated LRI-II at the primary sequence level or at the structural level, two additional mutant clones, RK13 and RK14, were created. RK13 is a substitution mutant in which the U5 (X-II) sequence was substituted with heterologous sequences of an equal length (5’ AGGAGGAGG 3’), resulting in the loss of complementarity between U5 and Gag sequences of LRI-II (Figure 3.4A and 3.4B). Next, in an attempt to create an artificial LRI-II with non-viral heterologous sequences, RK14 was created from

RK13 in which complementarity was restored by substituting the Gag (Y-II) sequence with the heterologous sequence, 5’ CCUCCUCCU 3’ (Figure 3.4A and 3.4B). As expected, RNA packaging and propagation of RK13 (designed to disrupt the complementarity of the SHAPE-validated LRI-II) was essentially abrogated (Figure 3.5C.II and III). Not quite surprisingly, but consistent with RK11 (compensatory mutant designed to re-establish artificial LRI-II based on the predicted structure), RK14 was unable to restore the packaging and propagation of the mutant vector RNA and was observed to be severely impaired in its ability to package and propagate its RNA (Figure 3.5C.II and III). Together, these data unambiguously suggest that it is the primary sequence of LRI-II that is important for efficient MPMV RNA packaging and propagation. These results also show that the primary sequence of LRI-II is critical in its native structural context only to augment MPMV RNA packaging since recreation of the artificial LRI-II using primary viral sequences in RK11 (maintaining complementarity, but in a flipped orientation) failed to restore packaging or propagation to wild-type levels.

3.6. Role of sequences involved in forming stem loops at the 3’ end of the SHAPE-validated structure during MPMV RNA packaging and propagation

While the secondary structure of the 5’ UTR sequences that have been shown to be important for MPMV RNA packaging appear to be anchored by complementary U5 and Gag sequences forming LRI-I and LRI-II (Figure 3.6A) sequences from the distal parts of the 5’ UTR and Gag (but excluding the Gag sequences involved in forming U5-Gag LRIs) forming three stem loops (SL3, Gag SL1, and Gag SL2) seem to provide stability to the overall RNA secondary structure

Figure 3.6: Design and test of the systematic deletion mutations introduced in sequences at the 3’ end of the SHAPE-validated structure forming SL3, Gag SL1 and Gag SL2



B

Clone Name	Description of the Mutation
SJ2	Wild type
RK15	Deletion of 33 nt of SL3
RK16	Deletion of 30 nt of Gag SL2
RK17	Deletion of 39 nt of Gag SL1
RK18	Double Deletion of SL3 and Gag SL1
RK19	Double Deletion of SL3 and Gag SL2
RK20	Triple Deletion of SL3, Gag SL1 and Gag SL2

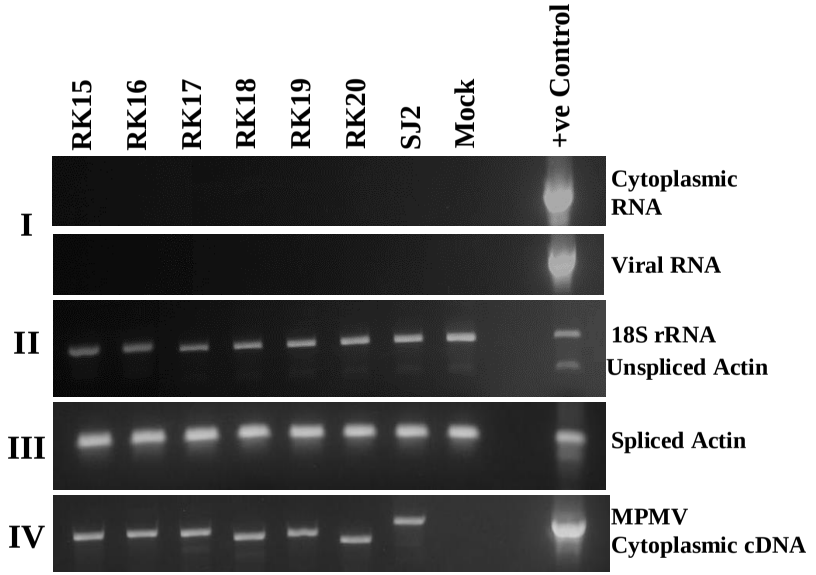


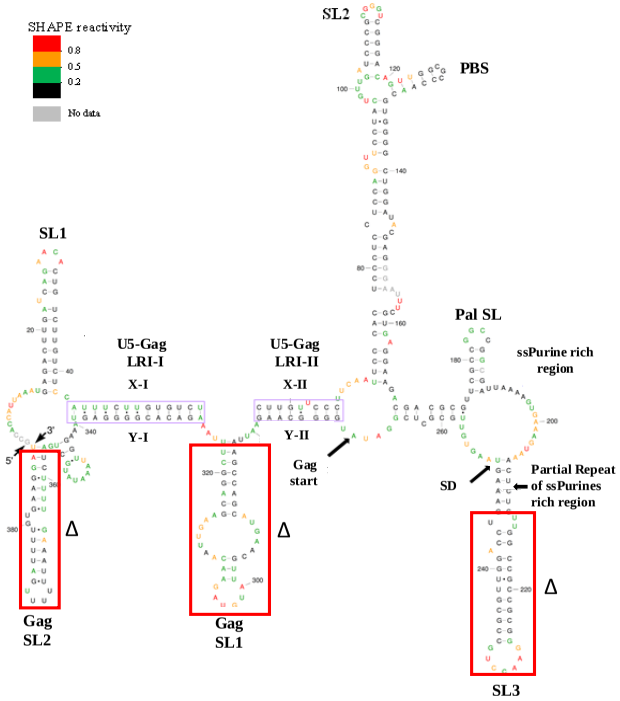
Figure 3.6. Design and test of the systematic deletion mutations introduced in sequences at the 3' end of the SHAPE-validated structure forming SL3, Gag SL1 and Gag SL2. (A) Illustration of the SHAPE-validated structure of MPMV packaging signal RNA with SL3, Gag SL1 and Gag SL2 highlighted in red boxes. **(B)** Table outlining the deletion mutations introduced in SL3 and Gag sequences. In RK15, 16 and 17, single deletion of either SL3 sequence, Gag SL1 or Gag SL2 was deleted respectively. Mutants RK17 and RK18 contain double deletion of SL3 along with either Gag SL1 or Gag SL2, respectively. Mutant RK20 contains a triple deletion of SL3, Gag SL1 and Gag SL2 **(C)** Representative gel images of the controls needed for validating different aspects of the three plasmid *trans* complementation assay. **(I)** PCR amplification of DNase-treated RNA from the cytoplasmic (upper panel) and viral (lower panel) RNA preparations with MPMV-specific vector primers **(II)** Multiplex amplification of unspliced β-actin mRNA and 18S rRNA and **(III)** PCR amplification of spliced β-actin mRNA to check for the nucleocytoplasmic fractionation technique **(IV)** PCR amplification of transfer vector cytoplasmic cDNAs using MPMV vector-specific primers.

of the 5' of MPMV genome. In order to test this, a systematic deletion analysis of the sequences involved in forming SL3, Gag SL1, and Gag SL2 was performed (Figure 3.6A). The first mutant, RK15, contained a 33 nucleotide deletion (from the distal part of the 5' UTR) removing most of the sequences forming SL3 (Figure 3.6A boxed region and 3.6B). Next, a 30 nucleotide deletion was introduced in RK16 to remove Gag SL2 (Figure 3.6A boxed region and 3.6B). Finally, a 39 nucleotide deletion was introduced in RK17 (boxed region in Figure 3.6A and 3.6B) to remove Gag SL1. RNA packaging efficiencies for mutant clones RK15-RK17 were quantified after having taken into consideration all the necessary controls (Figure 3.6C.I-IV). These results revealed that individual deletion of SL3, Gag SL1, and Gag SL2 only marginally affected RNA packaging when compared to the wild type ($\leq$ 1-fold or 0.5, 0.7, and 0.6 relative to the wild type; Figure 3.7C.II). This slight reduction in RNA packaging was in agreement with the RNA propagation data which also exhibited minor reductions in hygromycin-resistant colonies obtained compared to the wild type (Figure 3.7C.III). These results suggested that sequences involved in forming SL3, Gag SL1, and Gag SL2 do not play crucial role at individual levels during MPMV RNA packaging and propagation.

Redundancy is a common characteristic of retroviruses and it is possible that when deleting only one of the three stem loops, the remaining stem loops compensate for the loss of the deleted one for maintaining and stabilizing the major structural motifs of the 5' UTR important for MPMV RNA packaging and propagation. To test this possibility, we created deletion mutant clones in which sequences involved in forming SL3, Gag SL1, and Gag SL2 were deleted simultaneously in multiple combinations. In mutant clone RK18, the same sequences that were deleted individually in RK15 (33 nt of SL3) and RK16 (39 nt of

Figure 3.7: Effect of deletion mutations introduced into sequences at the 3’ end of the SHAPE-validated structure forming SL3, Gag SL1 and Gag SL2 on RNA packaging and propagation.

A



B

Clone Name	Description of the Mutation
SJ2	Wild type
RK15	Deletion of 33 nt of SL3
RK16	Deletion of 30 nt of Gag SL2
RK17	Deletion of 39 nt of Gag SL1
RK18	Double Deletion of SL3 and Gag SL1
RK19	Double Deletion of SL3 and Gag SL2
RK20	Triple Deletion of SL3, Gag SL1 and Gag SL2

C

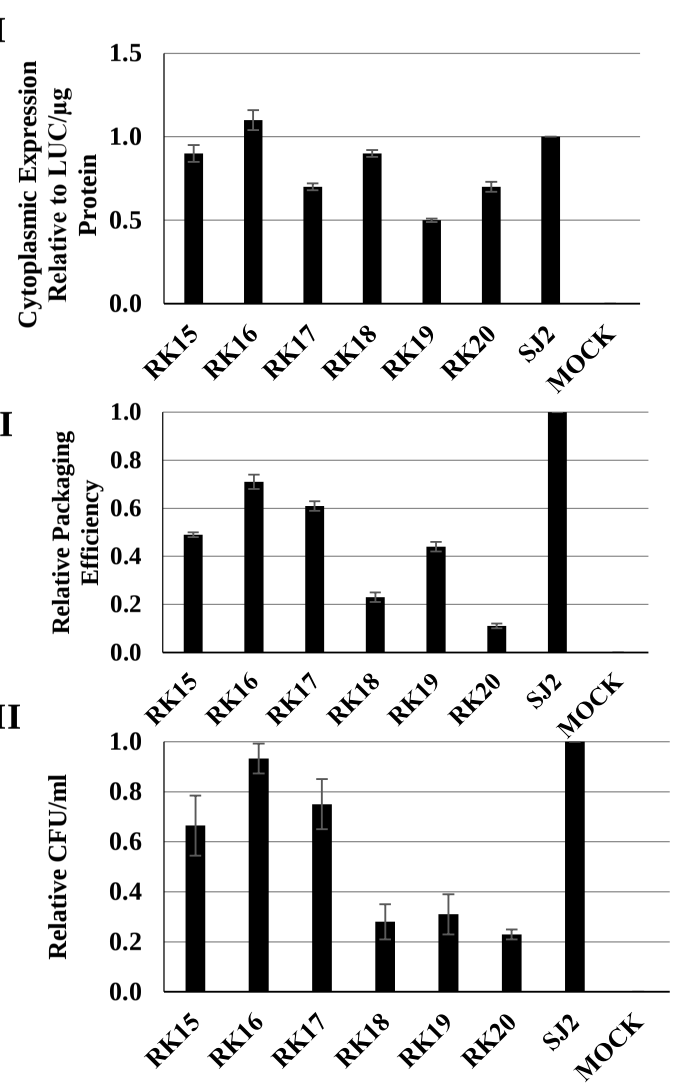


Figure 3.7. Effect of deletion mutations introduced into sequences at the 3' end of the SHAPE-validated structure forming SL3, Gag SL1 and Gag SL2 on RNA packaging and propagation. (A) Illustration of the SHAPE-validated structure of MPMV packaging signal RNA with SL3, Gag SL1 and Gag SL2 highlighted in red boxes. **(B)** Table outlining the deletion mutations introduced in SL3 and Gag sequences. In RK15, 16 and 17, single deletion of either SL3 sequence, Gag SL1 or Gag SL2 was deleted respectively. Mutants RK17 and RK18 contain double deletion of SL3 along with either Gag SL1 or Gag SL2, respectively. Mutant RK20 contains a triple deletion of SL3, Gag SL1 and Gag SL2. **(C)** Relative cytoplasmic expression and packaging efficiencies of MMTV transfer vector RNAs as measure by the ΔΔCT method, and propagation efficiencies of stem-loop mutant transfer vectors. Panel **(I)** Cytoplasmic transfer vector RNA expression in 293T cells relative to the wild type (SJ2 vector) after normalization with the β-actin endogenous control and luciferase expression. The error bars represent the standard deviation (SD) of triplicates of each clone. Panel **(II)** mutant transfer vector RNA packaging efficiencies relative to the wild type (SJ2) after normalization with β-actin and luciferase expression. Panel **(III)** Relative hygromycin resistant colony-forming units (CFU)/ml for mutant transfer vectors reflecting the relative propagation efficiencies compared to the wild type SJ2 vector. The data represented in histograms correspond to the mean of the samples when tested in triplicates (± SD) following transfection and infection experiments.

Gag SL1) were deleted simultaneously, creating a mutant clone with a 72 nucleotide discontinuous deletion removing most of SL3 and all of Gag SL2 (Figure 3.6A and 3.6B). Employing a similar strategy, another double deletion mutant, RK19, was created in which sequences involved in forming SL3 and Gag SL2 were simultaneously deleted (a 63 nt discontinuous deletion; Figure 3.6A and 3.6B). Finally, in a more drastic deletion, sequences involved in forming SL3, Gag SL1, and Gag SL2 were simultaneously deleted (a 102 nt discontinuous deletion) to create RK20 (Figure 3.6A and 3.6B).

There is a possibility such drastic deletion mutations (RK18-RK20) may inadvertently affect transfer vector RNA expression and stability. Therefore, before determining the effects of these mutations on RNA packaging, the cytoplasmic RNA expression of these mutant clones were analyzed to look for their stable expression. Results presented in Figure 3.6C.I-IV demonstrate that these transfer vector RNAs were efficiently and stably expressed and exported to the cytoplasm. Next, the packaging efficiency of the mutant clones relative to the wild type was determined (Figure 3.7C.II) and was found to be significantly impaired (2.5-10 fold less compared to the wild type). The most drastic impairment in RNA packaging was observed in the triple deletion mutant RK20 which could be attributed to the rather large (102 nt) deletion (~10-fold reduction; $P<0.01$; Figure 3.7C.II). RNA propagation in these mutant clones was observed to be proportionally reduced (Figure 3.7C.III). Together, these results suggest that the three stem loops tested above are important for stabilizing the structure of this region in a redundant manner; removal of one of these stem loop results in compensation by the others in stabilizing the structure. Compared to these mutants (RK18-RK20), many of the LRI-I and II mutants involving only few nucleotides mutations resulted in drastic

effects on the packaging and propagation potential of mutant vector RNAs. These results revealed that U5-Gag LRIs have a more important architectural role in stabilizing the structure of the 5' UTR sequences (that have been shown to be important for RNA packaging and dimerization), while the lower three stem loops (SL3, Gag SL1, and Gag SL2) may have a more secondary role in stabilizing the entire region.

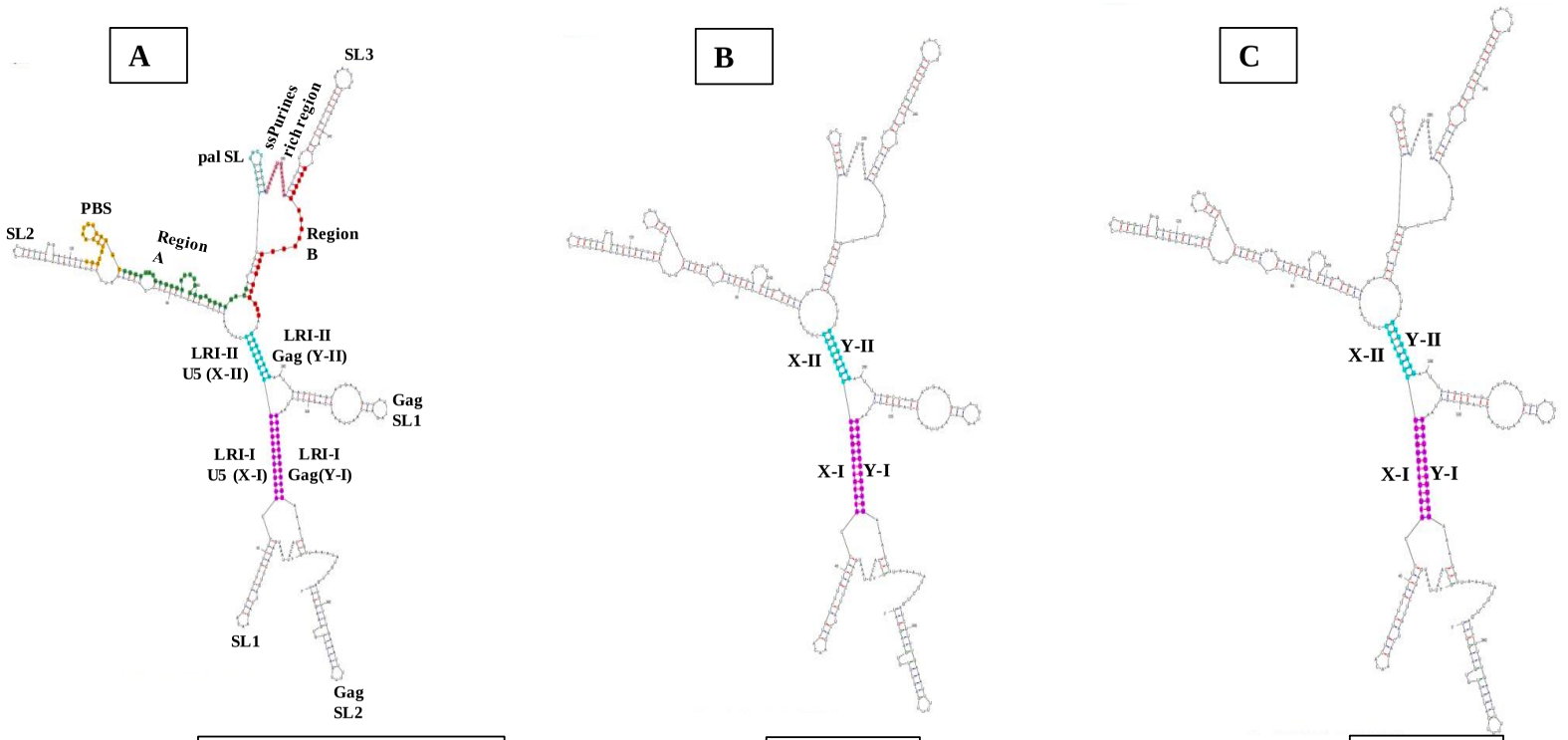
3.7. Structural Analyses of the mutants involving sequences forming the LRIs and stem loops (SL3, Gag SL1, and Gag SL2) at the 3' end of the SHAPE-validated structure

To better understand the results of the mutations introduced into LRI-I, LRI-II, SL3, and Gag SLs, a structural prediction analysis of these mutants was performed to correlate the biological effect of the introduced mutations with their predicted folded secondary structures.

3.7.1. Structure-function relationship of the sequences involved in U5-Gag LRI-I

The secondary structure of the mutant transfer vector RNAs was predicted using Mfold software (Mathews et al., 2003; Zuker, 1999). This algorithm predicts the most stable secondary structure that the RNA can fold by calculating the minimum free energy of each structure. The more negative the free energy, the more stable the structure. These mutants' RNA secondary structure predictions were compared to the wild type structure of the same region in an attempt to establish a structure-function correlation between the biological effect of the introduced mutations in the sequences involved in the formation of U5-Gag LRI-I and their

Figure 3.8: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing stabilizing mutations in the U5 sequences of the LRI-I and LRI-II.



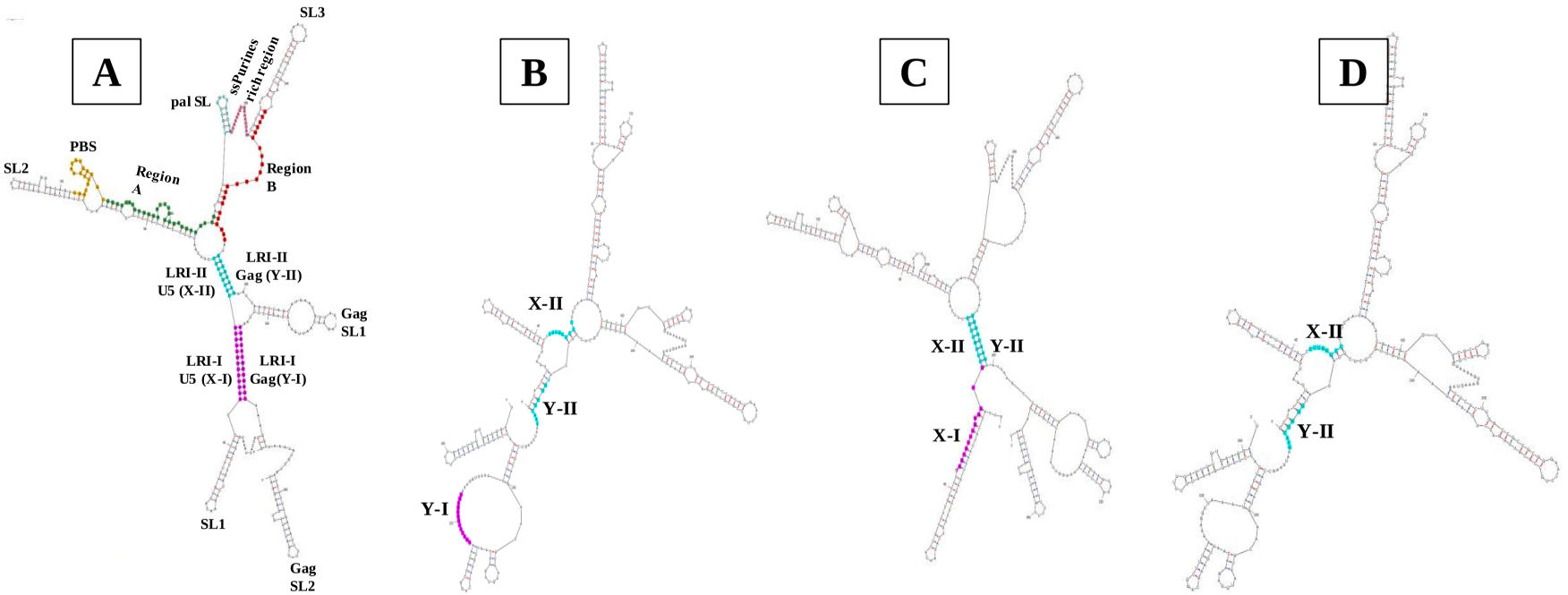
	Wild Type (SJ2)	RK1	RK2
U5 (X-I) **Gag (Y-I)**	**5' AUUUCUUGUGUCU 3'** **3' UGAGGGGCACAGA 5'**	**5' acuccccgugucu 3'** **3' UGAGGGGCACAGA 5'**	**5' acuccccgugucucuugccccc 3'** **3' UGAGGGGCACAGAGAACGGGG 5'**
RPE	1.00	0.36	0.25
ΔG (kcal/mol)	-124.7	-134.9	-140.4

Figure 3.8. Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing stabilizing mutations in the U5 sequences of the LRI-I and LRI-II. (A) MPMV wild type transfer vector SJ2, **Δ**G= -124.7 kcal/mol. **(B)** RK1 containing substitution of four uridines involved in the wobble base pairs with cytosines to stabilize LRI-I; **Δ**G= -134.9 kcal/mol. **(C)** RK2 mutant transfer vector containing substitution of six uridines involved in the formation of wobble base pairs in both LRI-I and LRI-II with cytosines; **Δ**G= -140.4 kcal/mol. Note the significant loss in RPE in both RK1 and RK2 and the increase in the stability of their structure when compared to the wild type due to the formation of G-C base pairs. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

predicted secondary structure. In RK1 and RK2, the substitution of the uridine nucleotides involved in the wobble (G-U) base pairing with cytosines stabilized LRI-I and LRI-II, respectively. As expected, the predicted secondary structure of RK1 and RK2 did not reveal any disruptions in the structure when compared to the wild type (Figure 3.8); however, the overall energy of RK1 and RK2 was observed to be more negative (ΔG = -134.9 and -140.4, respectively) compared to the wild type (ΔG = -124.7), suggesting that these two mutants became energetically more stable. Considering that these mutations had no effect on the structure but resulted in only further stabilization of the LRIs, while they had a significant effect on RNA packaging suggests that the built-in flexibility of this region is important for its function. In the wild type (SJ2) this flexibility is perhaps due to the presence of the wobble G-U base pairs that was lost in RK1 and RK2 due to the introduced mutations. Thus, the loss of the 4 G-U base pairs in RK1 did not only change the primary structure of LRI-I, but also increased the rigidity of the secondary structure since the free energy parameters used to predict the structure are sequence dependent. This means that the fewer numbers of G-U base pairs in the structure, the more rigid it is. Correlating these findings with the ~ 50% loss of packaging and propagation of RK1 and RK2 mutant transfer vector RNAs indicates that the optimal functionality of the LRIs is dependent upon the secondary structure with a moderate flexibility.

The next set of mutants, RK3, RK4 and RK5 represent deletion mutants where U5 (X-I), Gag (Y-I), and both were deleted respectively. The predicted structures of mutants RK3 and RK5 showed that whenever the U5 sequence forming the X-I strand were deleted, both LRI-I and LRI-II structures were lost in addition to the Gag stem loop 1 (Figure 3.9). In RK4, the effect of the deletion of Gag sequence

Figure 3.9: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing systematic deletions in the LRI-I complementary stem.



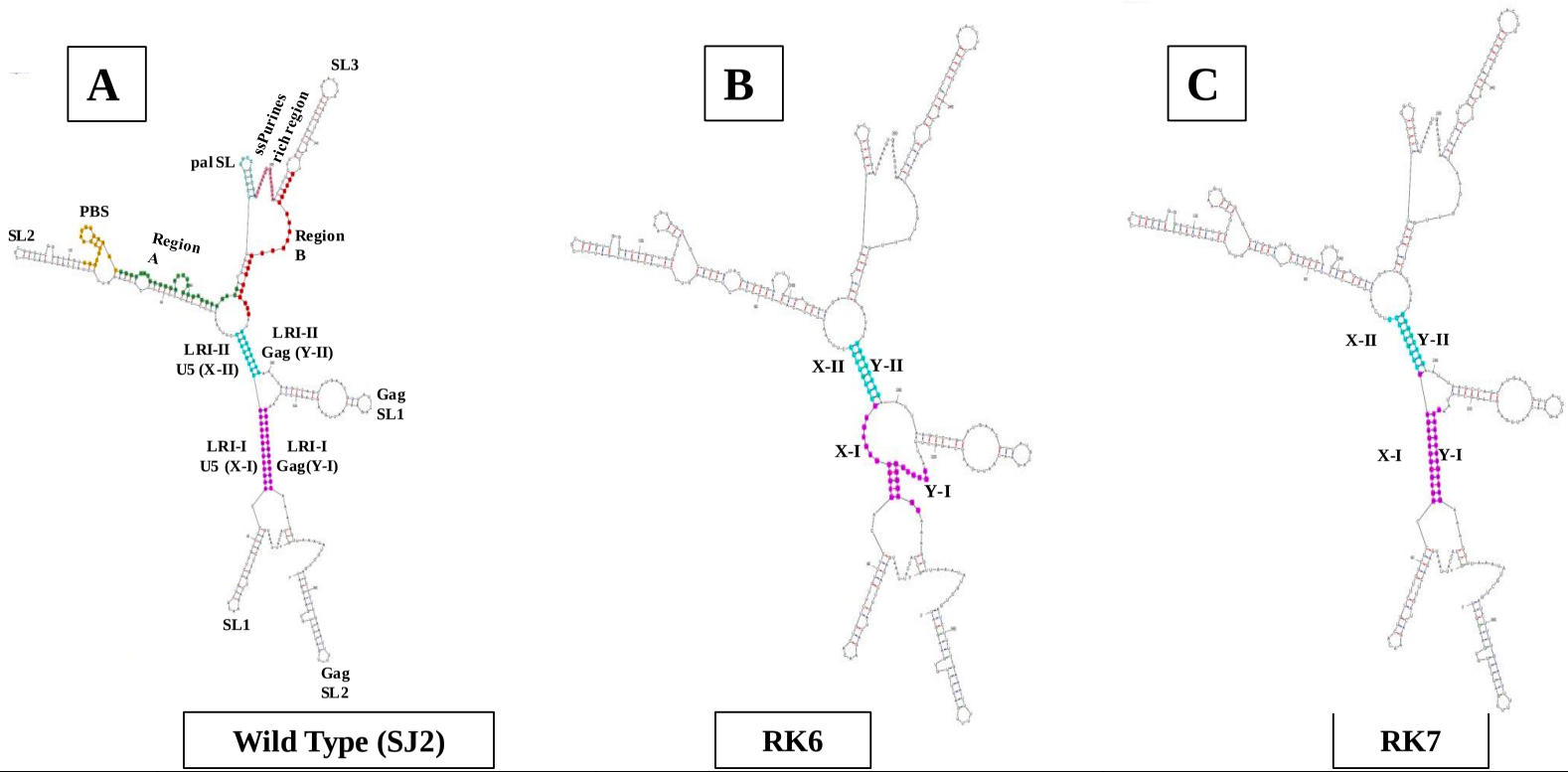
	Wild Type (SJ2)	RK3	RK4	RK5
U5 (X-I) **Gag (Y-I)**	**5' AUUUCUUGUGUCU 3'** **3' UGAGGGGCACAGA 5'**	**5' Δ 3'** **3' UGAGGGGCACAGA 5'**	**5' AUUUCUUGUGUCU 3'** **3' Δ 5'**	**5' Δ 3'** **3' Δ 5'**
RPE	1.00	0.04	0.08	0.05
ΔG (kcal/mol)	-124.7	-112.8	-116.5	-112.8

Figure 3.9. Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing systematic deletions in the LRI-I complementary stem. (A) MPMV wild type transfer vector SJ2; ΔG= -124.7 kcal/mol. **(B)** RK3 where the U5 sequence (X-I) was deleted; ΔG= -112.8. **(C)** RK4 where the Gag (Y-I) sequence was deleted; ΔG= -116.5 kcal/mol. (D) RK5 which contain a double deletion of both U5 and Gad strands of LRI-I; ΔG= -112.8 kcal/mol. Note the loss of LRI-I, Gag SL1 and LRI-II structures in RK3 and RK5 which is consistent with the drop in the free energy as both have ΔG= -112.8 where, as in RK4, only LRI-I structure and Gag SL1 is disrupted, while LRI-II remains intact. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

forming Y-I strand was observed on LRI-I only and Gag SL1 due to its proximity to the deleted sequence, while LRI-II structure remained intact (Figure 3.9). All three mutations had resulted in a complete loss of RNA packaging and propagation, suggesting that the maintenance of LRI-I complementary stem structure is critical for these functions. Additionally, we feel that the total free energy of the structure is important as well for these functions. Thus, although the overall structure of this region was maintained in these mutants, especially the pal SL, ssPurine-rich region, SL3, and Region B (sequences/structures critical for RNA packaging and dimerization), the free energy of the overall structure was lowered due to the effect of the mutations compared to the wild type (ΔG = -112.8, -116.5, and -112.8 compared to -124.7 for the wild type). This perhaps result in fragile structures (structures maintaining the shape but with lower energy), which are less stable than the wild type structure since the lesser negative the free energy, the less stable is the structure (Figure 3.9). This instability can explain the effect of the introduced mutations on the RNA packaging as it probably affects the proposed interaction between the NC domain of Gag protein with sequences augmenting RNA packaging.

The last two mutants, RK6 and 7, created to test the LRI-I were designed to destabilize and re-create LRI-I structure, respectively, using non-viral heterologous complementary sequences. As shown in figure 3.10, the structural prediction of RK6 (where the U5 sequence forming X-I strand was substituted with heterologous sequences, thus losing the complementarity between the two strands of the LRI-I) maintained all the important structural motifs including LRI-II; however the secondary structure of LRI-I was lost. The loss of LRI-I structure correlates well with the loss of gRNA packaging and propagation. We believe that this loss also destabilized the surrounding region, making the structure more flexible than needed

Figure 3.10: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs disrupting the complementarity of the native LRI-I and recreating an artificial LRI.



	Wild Type (SJ2)	RK6	RK7
U5 (X-I) Gag (Y-I)	**5' AUUUCUUGUGUCU 3' 3' UGAGGGGCACAGA 5'**	**5' aucucuuauauau 3' 3' UGAGGGGCACAGA 5'**	**5' aucucuuauauau 3' 3' uagagaauauaua 5'**
RPE	1.00	0.04	1.9
ΔG (kcal/mol)	-124.7	-115.4	-123.6

Figure 3.10. Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs disrupting the complementarity of the native LRI-I and recreating an artificial LRI. (A) MPMV wild type transfer vector SJ2; ΔG= -124.7 kcal/mol. **(B)** RK6 where the U5 sequence (X-I) was substituted with heterologous sequence thus disrupting the complementarity of LRI-I; ΔG= -115.4 kcal/mol. **(C)** RK7 where the LRI-1 complementarity was recreated by substituting the Gag sequence (Y-I) with heterologous sequence complementary to the one introduced in RK6; ΔG= -123.6 kcal/mol. Note the restoration of RPE in RK7 to even more than the wild type level while it was abolished in RK6. This correlates with the minimal free energy of RK7; ΔG = -123.6 kcal/mol that is close to the wild type ΔG =-124.7. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

(ΔG = -115.4), thus affecting its ability to function effectively. The mutation introduced in RK7, on the other hand, resulted in the restoration of the LRI-I structure using non-viral heterologous sequences (Figure 3.10) and with it the RNA packaging and propagation. The minimal free energy of the RK7 predicted structure was calculated to be -123.6 which is very close to wild type value of -124.7 (Figure 3.10). Thus, compared to the mutation in RK1, which also created a heterologous non-wild type LRI-I, RK1 was unable to restore RNA packaging and propagation. RK1 has a free energy of -134.9 which we believe makes the structure too rigid for effective function, thus supporting our hypothesis that the flexibility of this region is as important as the structure of LRI-I. This difference in the calculated minimal free energy of both mutants (RK1 and RK7) can be attributed to the number of G-C base pairs in each LRI since none of the artificial LRIs contained G-U base pairing when compared to the wild type (SJ2) LRI-I containing four G-U base pairing. In RK1, eight out of the 13 base pairs were G-C base pairs which resulted in a very stable stem; whereas in RK7, 2 out of 13 base pairs were G-C base pairs, thus lowering the energy of the stem (Figure 3.10). Thus, it is rational to conclude that it is the secondary structure of the LRI-I that is anchoring the overall secondary structure of MPMV packaging signal RNA. In addition, it should have a final embedded energy close to the wild type minimal free energy to maintain a certain level of flexibility which seems to be the characteristic of LRI-I.

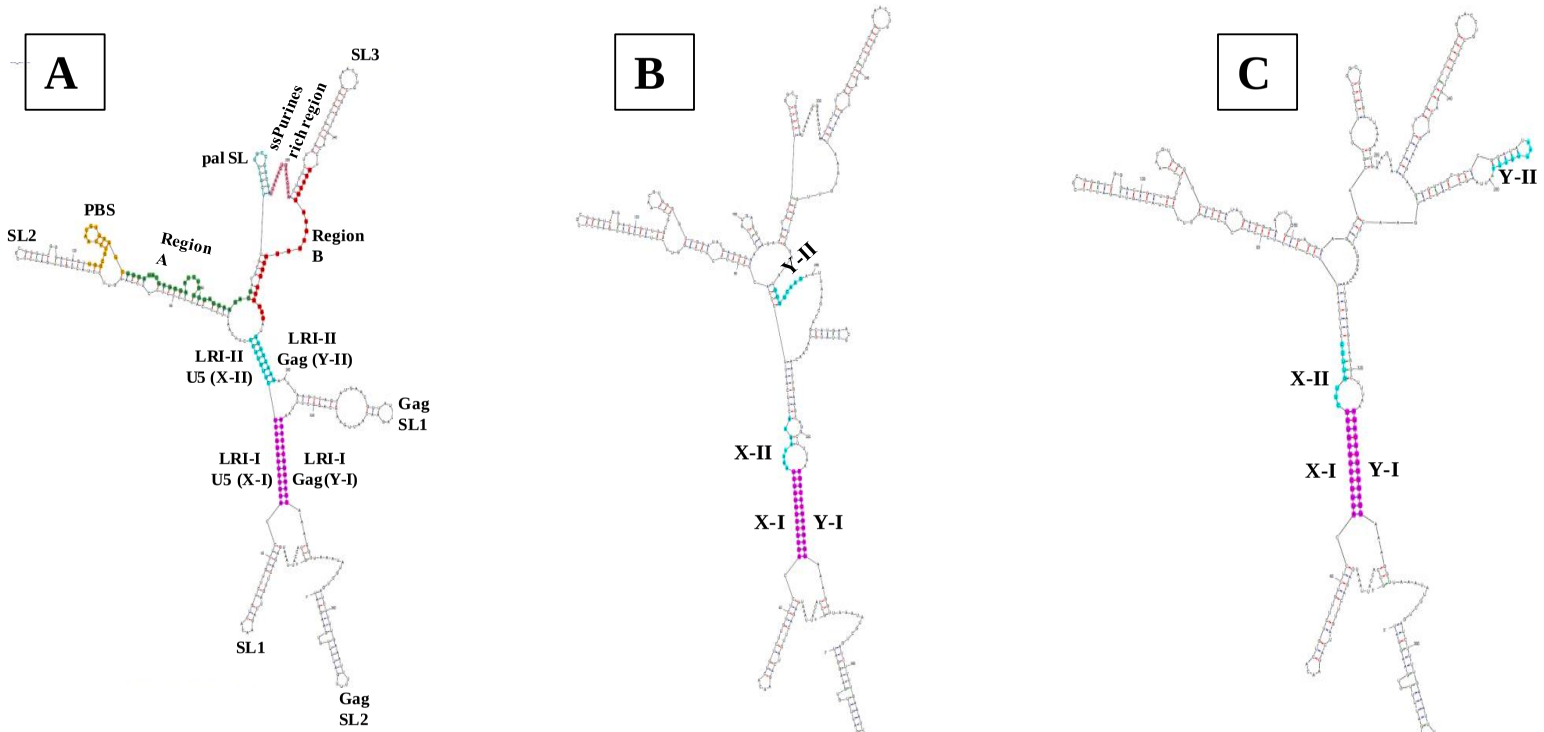
3.7.2. Structure-function relationship of the sequences involved in U5-Gag LRI-II

Next, the structural analysis of the LRI-II mutants was conducted to determine the structure-function relationship of LRI-II with RNA packaging. To

summarize, the mutation in the LRI-II mutant RK8 replaced the predicted eight nucleotides of LRI-II U5 (X-II) with heterologous sequences to disrupt the complementarity with Y-II. Along the same lines the mutation in RK9 replaced U5 (X-II) sequence with a sequence identical to Gag (Y-II) sequence thus duplication Gag (Y-II) sequence and disrupting LRI-II complementarity, while the mutation in RK10 replaced Gag (Y-II) sequence with sequence from U5 (X-II) thus duplicating Gag (Y-II) sequence and disrupting LRI-II complementarity as well. Finally, the mutation in RK11 flipped the LRI-II stem-loop in a 180 degree manner to determine its function relative to the context. As shown in figures 3.4 and 3.5, disruption of LRI-II in any manner resulted in a complete abrogation of both RNA packaging and propagation, suggesting that LRI-II is an important structure to maintain the optimal RNA packaging and propagation of MPMV transfer vector RNAs. However, unlike the LRI-I mutations, the mutational analysis of LRI-II revealed that it is its primary sequence in the native context that is essential for MPMV RNA packaging and propagation.

Such a conclusion pertaining to the role of LRI-II in MPMV RNA packaging and propagation was further supported by the structural analysis of the LRI-II mutants. Mutation of LRI-II by disruption of the X-II/Y-II stem complementarity in RK8 and RK10 actually resulted in structures that led to the creation of stem structures that resembled X-II in LRI-II; however, RNA packaging and propagation of these mutants was still found to be abrogated (Figure 3.11). Similarly, disruption of complementarity LRI-II in RK9 or inversion of LRI-II in RK11 (creating an artificial LRI-II) both resulted in a complete abrogation of packaging and propagation despite the reformation of an LRI-II-like stem in these mutants, separated from LRI-I by a two nucleotide bulge (Figure 3.12). Likewise, a complete

Figure 3.11: Mfold secondary structure predictions of the LRI-II mutant transfer vector packaging signal RNAs.



	Wild Type (SJ2)	RK8	RK10
U5 (X-II) Gag (Y-II)	**5' CUUGUUCC 3' 3' GAACGGGG 5'**	**5'agauagag 3' 3' GAACGGGG 5'**	**5' CUUGUUCC 3' 3' cuuguucc 5'**
RPE	1.00	0.02	0.06
ΔG (kcal/mol)	-124.7	-122.9	-120.6

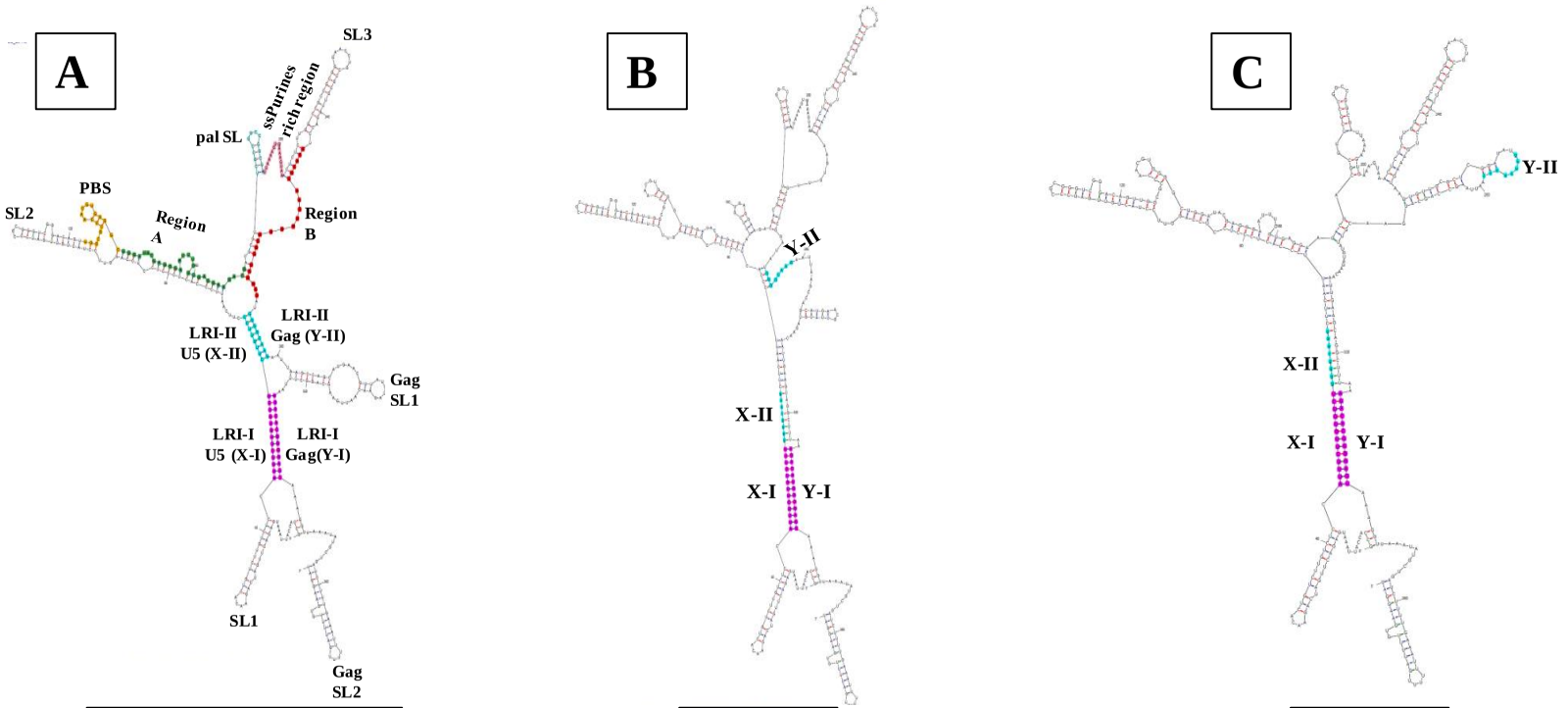
Figure 3.11. Mfold secondary structure predictions of the LRI-II mutant transfer vector packaging signal RNAs. **(A)** MPMV wild type transfer vector SJ2, ΔG= -124.7 kcal/mol. **(B)** RK8 mutant transfer vector structure where the U5 sequence (X-II) was substituted with heterologous sequence, causing the loss of complementarity between the LRI-II strands; ΔG= -122.9 kcal/mol. **(C)** RK10 where the Gag (Y-II) sequence was substituted with X-II sequence of LRI-II which disrupts the complementarity of LRI-II stem; ΔG= -120.6 kcal/mol. Note the abolished packaging efficiency in both mutant transfer vectors which suggests the importance of LRI-II complementary structure. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

deletion of X-II in RK12, or its replacement with a heterologous sequence in RK13, or reestablishment of LRI-II with heterologous sequences in RK14 (again creating an artificial LRI-II) all resulted in complete abrogation of packaging and propagation (Figure 3.5). Structural analysis of these mutants showed that LRI-II stem was indeed disrupted in RK12 and RK13, especially RK13, thus explaining the loss of function. In the case of RK14 even though an artificial LRI-II was re-created by substituting the SHAPE-validated LRI-II sequences, in its native context, yet its structural emergence was insufficient to restore function (Figure 3.13). The structure of RK14 showed that the LRI-II was recreated properly using the heterologous sequences unlike RK11 where the predicted 8 nt LRI-II structure was used to recreate the LRI. This observation further confirms the SHAPE-validated structure of the LRI-II since artificial complementarity (based on predicted structure) in RK11 failed to recreate LRI-II secondary structure (compare structures of RK11 and RK14 in figures 3.12 and 3.13 respectively). Together, these data confirm the earlier conclusion that the primary sequence of LRI-II in its native context is critical for its function.

3.7.3. Structure-function analysis of the mutations introduced in sequences involved in forming stem loops at the 3' end of the SHAPE-validated structure

In contrast to the LRI-I and LRI-II mutants, deletion analysis of the mutant transfer vectors in the more distal 5' UTR or Gag region revealed that single deletion of any of the three targeted stem loops (SL3, Gag SL1 or Gag SL2) individually had a marginal effect on RNA packaging and propagation (Figure 3.7). This can be explained by the structural analysis of RK15, RK16 and RK17 where all the structural motifs remained intact like the wild type except for the deletion of the

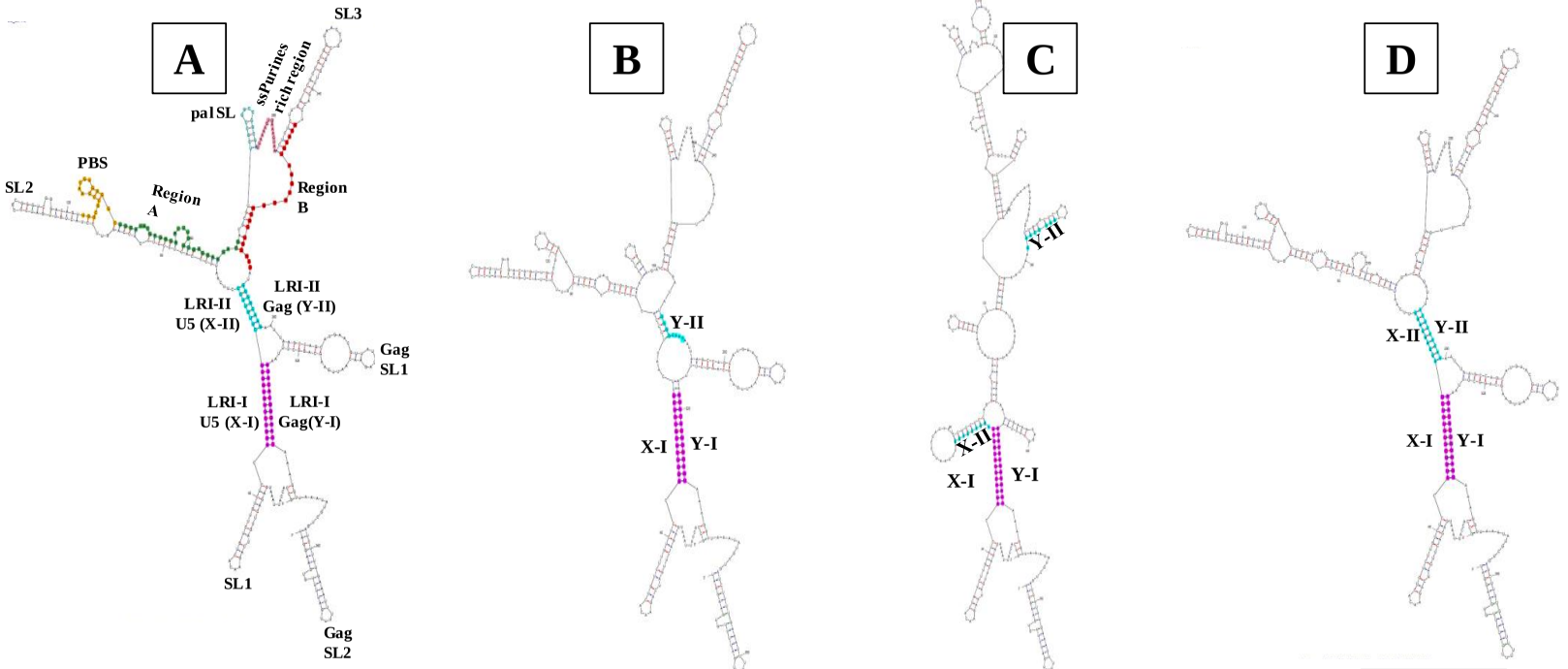
Figure 3.12: Mfold secondary structure predictions of the LRI-II mutant transfer vector packaging signal RNAs.



	Wild Type (SJ2)	RK9	RK11
U5 (X-II) Gag (Y-II)	**5' CUUGUUCC 3' 3' GAACGGGG 5'**	**5'ggggcaag 3' 3' GAACGGGG 5'**	**5' ggggcaag 3' 3' ccuuguuc5'**
RPE	1.00	0.01	0.03
ΔG (kcal/mol)	-124.7	-128.9	-128.6

Figure 3.12. Mfold secondary structure predictions of the LRI-II mutant transfer vector packaging signal RNAs. (A) MPMV wild type transfer vector SJ2, ΔG= -124.7 kcal/mol. **(B)** RK9 mutant transfer vector structure where the U5 sequence (X-II) was substituted with the Gag (Y-II) sequence, causing a loss of complementarity between the LRI-II strands; ΔG= -128.9 kcal/mol. **(C)** RK11 mutant transfer vector contains a flipped LRI-II secondary structure;, ΔG= -128.6 kcal/mol. Note that recreating the LRI-II in RK11 did not restore the packaging efficiency (RPE= 0.025 relative to the wild type) which suggests the importance of the primary sequence of the LRI-II in its native context. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

Figure 3.13: Mfold secondary structure predictions of the LRI-II mutant transfer vector packaging signal RNAs.



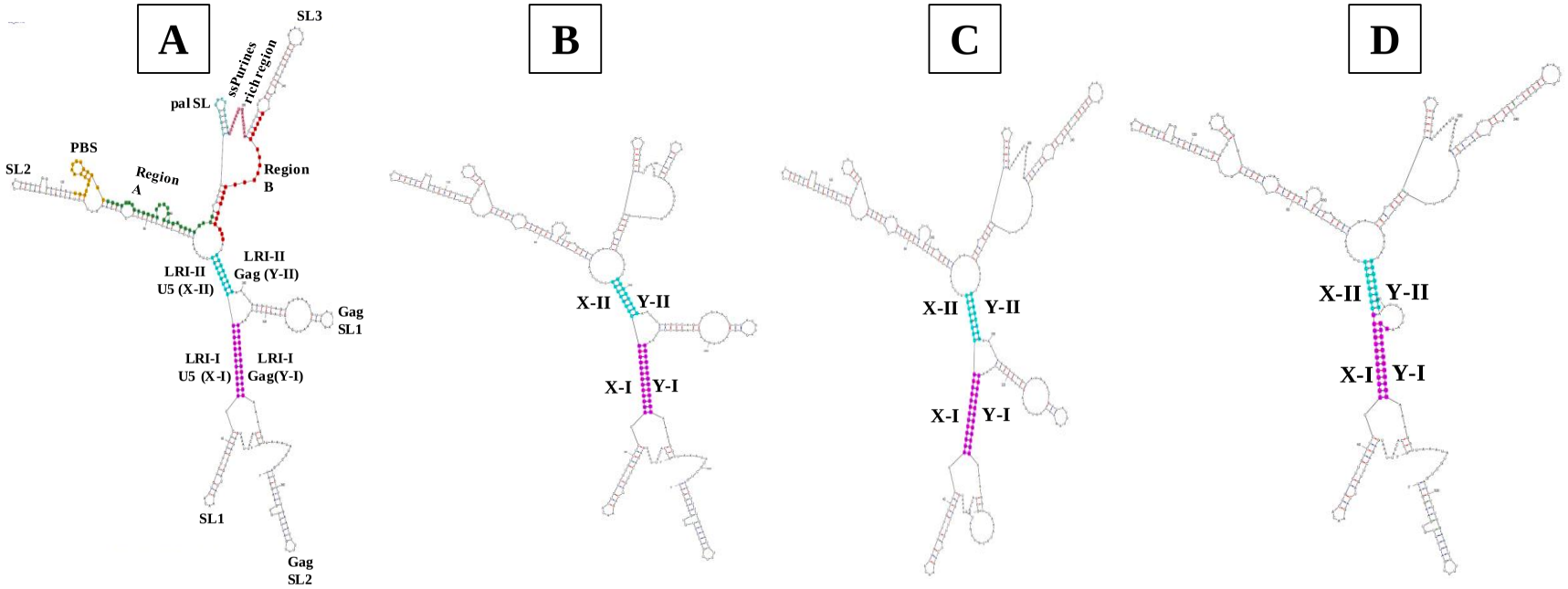
	Wild Type (SJ2)	RK12	RK13	RK14
U5 (X-II) Gag (Y-II)	**5' CUUGUUCC 3' 3' GAACGGGG 5'**	**5' Δ 3' 3' GAACGGGG 5'**	**5' aggaggagg 3' 3' GAACGGGG 5'**	**5' aggaggagg 3' 3' uccuccucc 5'**
RPE	1.00	0.04	0.03	0.12
ΔG (kcal/mol)	-124.7	-119.3	-125.8	-131.8

Figure 3.13. Mfold secondary structure predictions of the LRI-II mutant transfer vector packaging signal RNAs. (A) MPMV wild type transfer vector SJ2; ΔG= -124.7 kcal/mol. **(B)** RK12 where the U5 sequence (X-II) of the SHAPE-validated LRI-II was deleted; ΔG= -119.3 kcal/mol. **(C)** RK13 where the U5 sequence (X-II) of the SHAPE-validated LRI-II was substituted with heterologous sequence to disrupt the complementarity between X-II and Y-II of the LRI-II structure; ΔG= -125.8 kcal/mol. **(D)** RK14 mutant transfer vector where an artificial LRI-II is created by substituting the Y-II strand with heterologous sequence complementary to the one introduced in RK13, thus restoring the complementarity of between X-II and Y-II; ΔG= -131.8 kcal/mol. Note that despite the presence of a complementary LRI in RK14, the RPE was still only 0.116. This means that the primary structure of LRI-II is essential unlike LRI-I. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

targeted stem loop (Figure 3.14). Interestingly, in RK15, the ssPurine-rich region became partially base-paired due to the deletion of SL3. Moreover, six nucleotides in region B which are base-paired in the wild type structure lost their base pairing. Region B has previously been shown to be important for MPMV RNA packaging (Schmidt et al., 2003; Jaballah et al., 2010). This may explain the slight reduction in packaging and propagation observed in this mutant (Figure 3.7).

The double deletions which involved ΔSL3 with either ΔGag SL1 or ΔGag SL2 in separate mutants (RK18 and RK19, respectively) showed a more pronounced effect on RNA packaging and propagation when compared to the wild type (Figure 3.7). Structural predictions of these two mutants revealed the same changes seen in RK15 where the ssPurine-rich region became partially base-paired and in region B the six nucleotides which are part of the ssPurines repeat sequence became single stranded (Figure 3.15). In the case of triple deletion, RK20, which involved ΔSL3, ΔGag SL1 and Δ Gag SL2, the RNA packaging and propagation were further reduced (Figure 3.7). Consistent with secondary structure predictions of RK15, RK18 and RK19, the ssPurine-rich region became partially base paired in RK20 as well, while maintaining the LRIs (Figuers 3.14 and 3.15). This change in structure was due to the deletion of SL3 which is common to all four mutants. Thus, our overall conclusion is that as long as the two LRIs are maintained, deletion of these three stem loops does not have a detrimental effect on the packaging/propagation potential of the MPMV transfer vector RNAs.

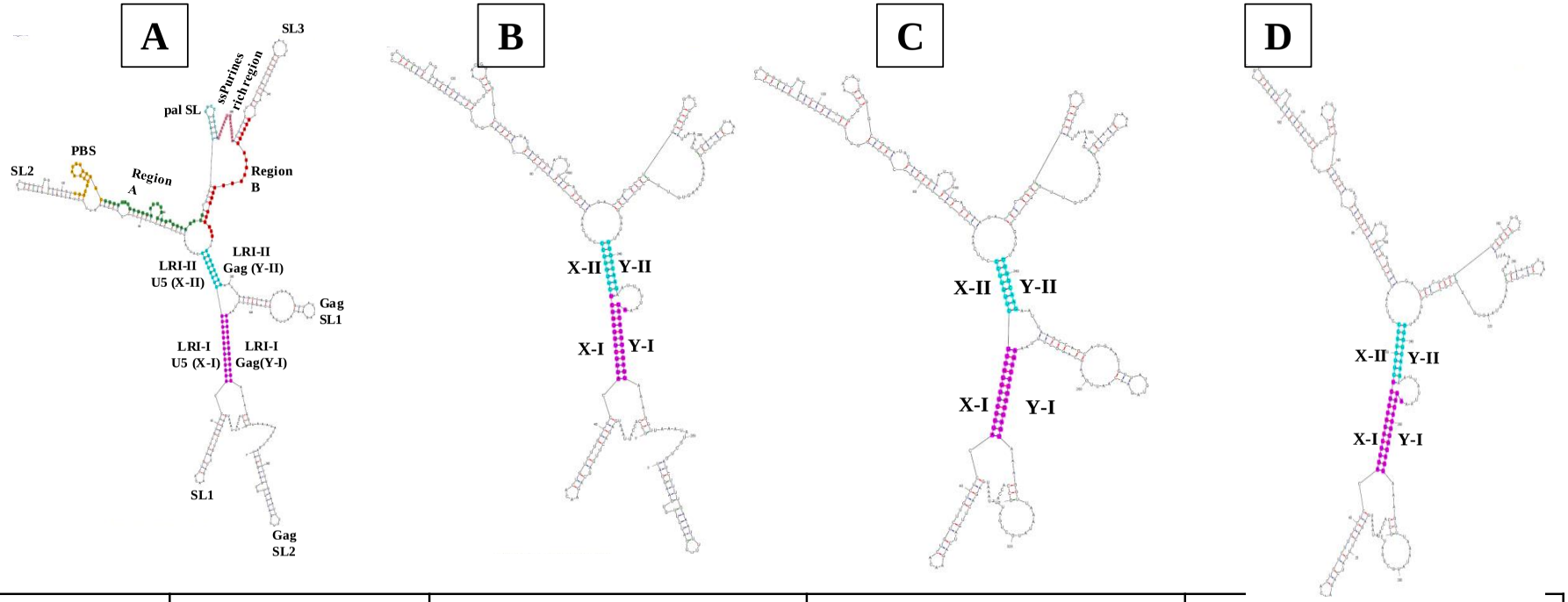
Figure 3.14: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing individual deletions of SL3, Gag SL1, and Gag SL2



Clone	Wild Type (SJ2)	RK15	RK16	RK17
RPE	1.00	0.49	0.71	0.61
ΔG (kcal/mol)	-124.7	-107.9	-119.5	-116.6

Figure 3.14. Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing individual deletions of SL3, Gag SL1, and Gag SL2. **(A)** MPMV wild type transfer vector SJ2; ΔG = -124.7 kcal/mol. **(B)** RK15 containing a deletion of SL3; ΔG = -107.9 kcal/mol. **(C)** RK16 mutant transfer vector containing a deletion of Gag SL2; ΔG = -119.5 kcal/mol. **(D)** RK17 mutant transfer vector containing a deletion of Gag SL1; ΔG=1116.6 kcal/mol. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

Figure 3.15: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing multiple deletions of SL3, Gag SL1, and Gag SL2.



Clone	Wild Type	RK18	RK19	RK20
RPE	1.00	0.23	0.44	0.11
ΔG (kcal/mol)	-124.7	-99.8	-102.7	-194.6

Figure 3.15. Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing multiple deletions of SL3, Gag SL1, and Gag SL2. **(A)** MPMV wild type transfer vector SJ2; ΔG = -124.7 kcal/mol. **(B)** RK18 containing a double deletion of SL3 and Gag SL1; ΔG = -99.8 kcal/mol. **(C)** RK19 mutant transfer vector containing a double deletion of SL3 and Gag SL2; ΔG = -102.7 kcal/mol. **(D)** RK20 mutant transfer vector containing a triple deletion of SL3, Gag SL1, and Gag SL2; ΔG=194.6 kcal/mol. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

Chapter 4: Discussion

A number of findings over the last decade have significantly advanced our understanding of the mechanism of retrovirus RNA packaging process. The fact that packaging signals of different retroviruses share the common characteristic of folding into secondary structural motifs emphasizes the importance of such structural elements to RNA packaging. The 5' end of MPMV genome (from the first nucleotide in R up to 120 bp of Gag) contains a number of sequence and structural elements that have been shown to be important for RNA packaging and dimerization (Schmidt et al., 2003; Jaballah et al., 2010; Aktar et al., 2013; Figure 1.13). One of the distinguishing features of the MPMV packaging signal RNA secondary structure is the long-range interactions of 5' (U5) and 3' (Gag) complementary sequences, forming two LRIs (LRI-I and LRI-II) that are believed to anchor its overall secondary structure (Jaballah et al., 2010; Aktar et al., 2013; Figure 1.13). Both of these LRIs have been found to be phylogenetically conserved among different MPMV strains at the structure level, as well as in maintaining a very high degree of complementarity between U5 and Gag sequences (Aktar et al., 2013; Figure 1.14 and 1.15). LRIs have been found to be conserved in several retroviruses including HIV-1, FIV, and MMTV (Paillart et al., 2002; Abbink and Berkhout, 2003; Kenyon et al., 2008; Kenyon et al., 2011, Aktar et al., 2014). Despite the fact that there are substantial sequence heterogeneities among human, simian, and feline lentiviruses, the preservation of such LRIs in all HIV-1, HIV-2, simian immunodeficiency virus (SIV), and FIV isolates offers additional proof for the functional significance of LRIs in retroviral life cycle (Paillart et al., 2002; Kenyon et al., 2008; Kenyon et al., 2011). Such a pivotal role of LRIs is further evidenced by the fact that mutations

that destabilize the complementarity of these interactions affect several important steps in retroviral life-cycle, including RNA packaging and dimerization (Paillart et al., 2002; Abbink and Berkhout, 2003; Rizvi et al., 2010).

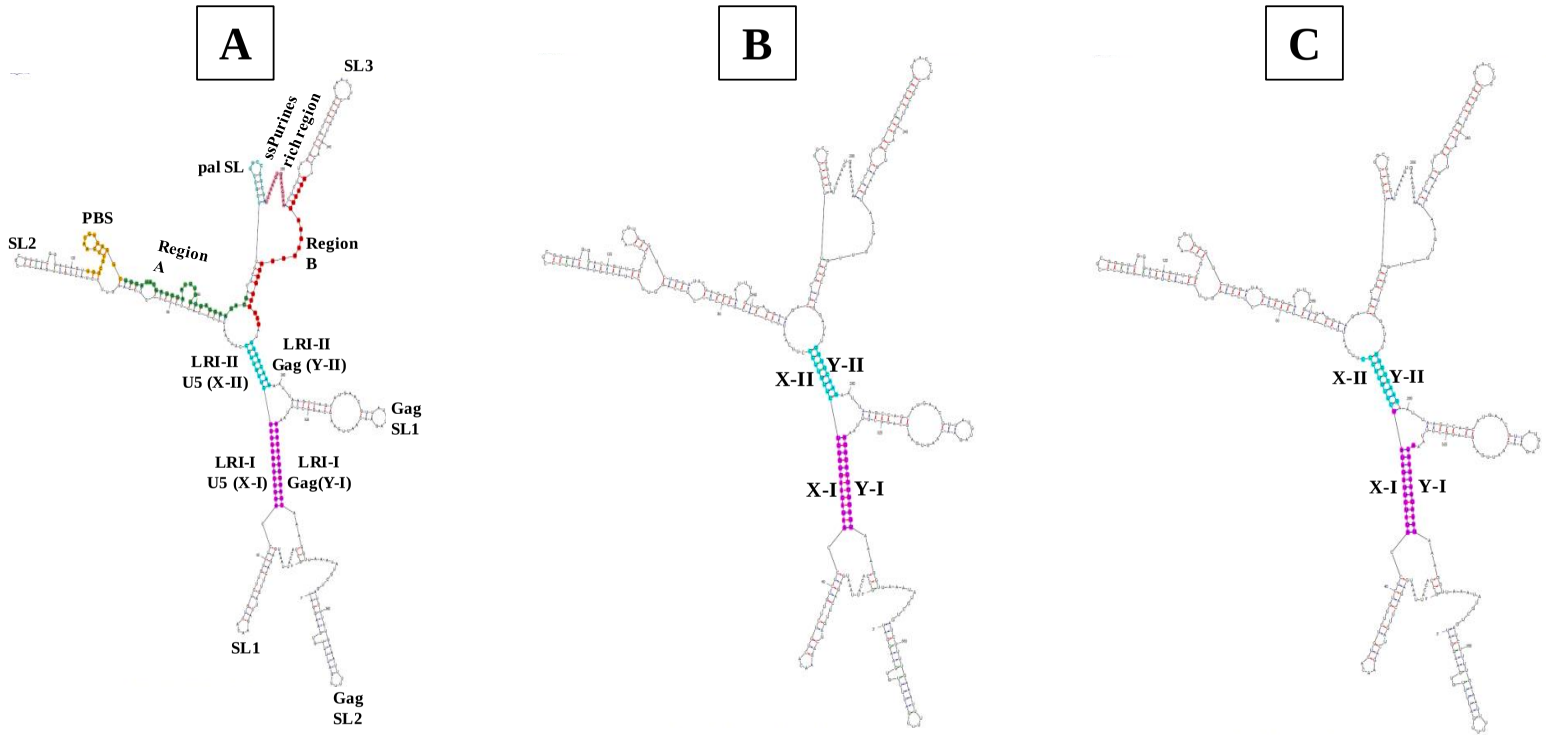
Unlike other retroviruses with one identified LRI, MPMV contains two LRIs (LRI-I and LRI-II; Figure 1.13) both of which maintain the interaction between U5 and Gag sequences (Paillart et al., 2002; Abbink and Berkhout, 2003; Kenyon et al., 2008; Kenyon et al., 2011, Aktar et al., 2014). The work presented here employed mutational and structural analyses of these LRIs (which appears to be anchoring the major structural motifs) as well as three other stem loops (SL3, Gag SL1, and Gag SL2) that are thought to provide architectural support to the overall RNA secondary structure to study their significance during MPMV RNA packaging. The results presented here provide data validating the existence and the biological significance of the two LRIs. These further reveal that the two LRIs function in different manners mechanistically with LRI-I being important at the secondary structural level, while LRI-II being important at both the primary sequence as well as structural levels (Figures 3.3 and 3.5). Finally, deletion and structural analysis of the three additional stem loops targeted reveal that SL3, Gag SL1, and Gag SL2 have secondary significance compared to the LRIs during RNA packaging and propagation and serve mainly to hold the entire secondary structure of this region together with redundant functions such that if one stem loop is deleted, the others can compensate for its absence (Figure 3.7). In summary, our results reveal that the 5’ end of the MPMV genomic RNA is held together by two LRIs and the additional stem loops (SL3, Gag SL1, and Gag SL2) help maintain the proper folding and stability of the secondary structure for function even though none of these stem

loops by themselves are directly involved in MPMV RNA packaging process (Figure 3.7).

Functionally, the U5-Gag LRI-I of MPMV resembles the R/U5-Gag heptanucleotide FIV LRI, where its presence as a structural motif rather than its primary sequences is essential for RNA packaging (Rizvi et al., 2010). However, the MPMV LRI-I differs from the FIV heptanucleotide LRI in the observation that re-establishment of the LRI-I structure using any complementary sequences does not necessarily restore RNA packaging and propagation. Note the lack of RNA packaging in LRI-I re-establishing mutant RK1, while packaging was restored by RK7 (Figure 3.3). RK1 is a LRI-1-stabilizing mutant where the G-U base pairs had been substituted with G-C base pairs to make the LRI-I more stable, without disrupting the RNA structure (Figures 3.3 and 4.1). On the other hand, heterologous complementary sequences were used to substitute the original LRI-I in RK7 (Figures 3.3 and 4.1). Even though in both cases the structure of the region remained identical and the primary sequence had been changed, only RK7 was able to restore RNA packaging and propagation, but not RK1 (Figures 3.3 and 4.1). We believe that this is due to the difference in the free energy (ΔG) of the structures that were created (Figure 4.1). Thus, the re-established heterologous LRIs should have a minimal embedded free energy at a level that is close to that of the wild type (SJ2) level. The free energy values of the predicted secondary structures of the mutant transfer vectors RK1 and RK7 and their relative packaging efficiencies clearly favor this argument (Figure 4.1).

This energy-sensitive phenomenon can be attributed to the fact that the LRI-I structure should have a certain level of flexibility to fulfill its function. This

Figure 4.1: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing mutations in LRI-I.



	Wild Type (SJ2)	RK1	RK7
U5 (X-I) Gag (Y-I)	**5' AUUUCUUGUGUCU 3' 3' UGAGGGGCACAGA 5'**	**5' acuccccgugucu 3' 3' UGAGGGGCACAGA 5'**	**5' aucucuuauauau 3' 3' uagagaauauaua 5'**
RPE	1.00	0.36	1.9
ΔG (kcal/mol)	-124.7	-134.9	-123.6

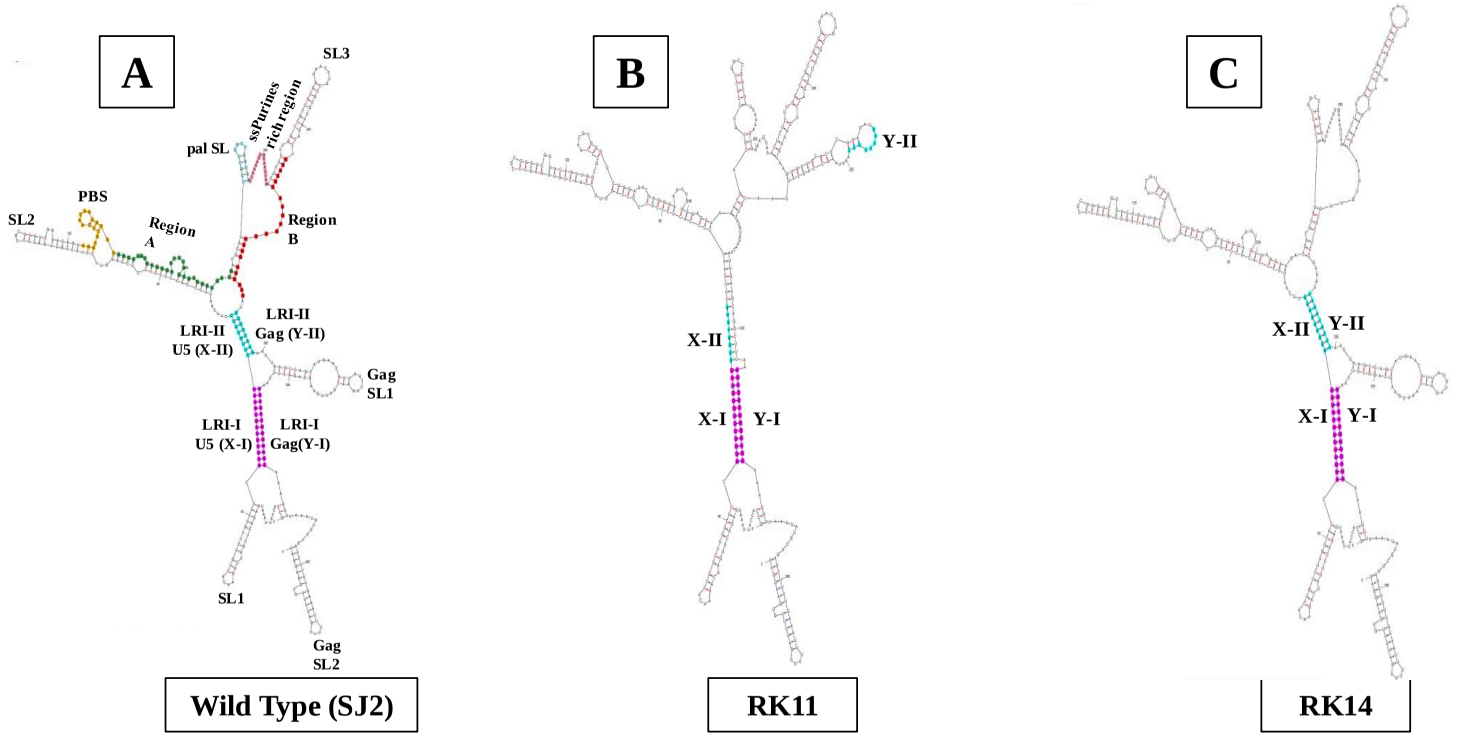
Figure 4.1. Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing mutations in LRI-I. (A) MPMV wild type transfer vector SJ2; ΔG = -124.7 kcal/mol. **(B)** RK1 containing substitution of four uridines involved in the wobble base pairs with cytosines to stabilize LRI-I; **Δ**G= -134.9 kcal/mol. **(C)** RK7 where the LRI-1 complementarity was recreated by substituting the Gag sequence (Y-I) with heterologous sequence complementary to the one introduced in RK6; ΔG= -123.6 kcal/mol. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

flexibility comes from the low affinity base pairing (A-U or G-U; Mathews et al., 1999) being part of the complementary base pairing within the wild type LRI-I structure. The minimal free energy calculated by the Mfold algorithm indicates the flexibility of the RNA structure based on the cumulative energies of the base pairing between the nucleotides involved in the RNA sequence. The secondary structure with the minimum free energy is the most stable structure (Mathews et al., 1999). A careful analysis of the complementary sequences forming wild type (SJ2) and artificial LRI-I mutants RK1 and RK7 revealed an interesting base pairing composition which favors the flexibility argument put forwarded above. For example, in the case of the wild type (SJ2), 9/13 (~70%) were A-U or G-U (5 A-U, 4 G-U, and 4 G-C base pairs), with a free energy (ΔG) of -124.7, while in the case of the LRI-I re-establishing mutant RK1 which could not restore RNA packaging, only 5/13 (38%) were A-U (5 A-U and 8 G-C base pairs) with a free energy (ΔG) of -134.9. Finally, in the case of LRI-I re-establishing mutant RK7 which restored RNA packaging, 11/13 (~85%) were A-U complementary base pairs (11 A-U and 2 G-C base pairs) with a free energy (ΔG) of -123.6 (Figure 4.1). Our interpretation of this data is that the increased embedded energy due to the stable G-C base pairing in RK1 made the LRI-I structures more rigid than needed, resulting in the significant drop in MPMV RNA packaging and propagation observed (Figure 4.1). It is well known that the genomic RNAs are active players in many critical stages in the retroviral life cycle, including processes that are driven mainly by the proper 3-dimensional folding of multiple RNA domains to recruit necessary viral and host factors required for activity. This flexibility, at a moderate level, is of great importance to the RNA as it is a dynamic structure that undergoes several intramolecular and intermolecular interactions (for example, RNA-RNA and Gag-

NC-RNA interactions) during the RNA dimerization and packaging processes. Therefore, even though the LRI-I structure is important, it must maintain a certain level of flexibility (signified by its free energy) to allow further folding or unfolding of the region, as the need may arise during the virus life cycle.

The case of LRI-II, however, was observed to be different as the results obtained suggest that it is its primary sequence in its native structural context that is essential for RNA packaging and propagation. LRI-II stem destabilizing mutations in the LRI-II in mutants RK8-RK10, RK12 and RK13 resulted in a complete loss of RNA packaging and propagation (Figure 3.5). Re-establishment of the LRI-II structure in RK14 using heterologous complementary sequences of equal length also failed to restore packaging or propagation of MPMV transfer vector RNAs (Figure 4.2). Even RK11, the mutant with the primary sequence of LRI-II intact, but in the opposite orientation was unable to restore LRI-II function despite the presence of the primary sequence (Figure 3.5). This can be attributed due to a loss of secondary structure of this mutant, similar to RK9 that contained the heterologous stem disrupting mutation in LRI-II strand (Figures 3.12 and 4.2). Together, these results point to the conclusion that the primary sequence of LRI-II in its native structural context is critical for its function. However, the possibility remains that the energy requirement of the re-established LRI-II may not have been met since the ΔG of RK14 was observed to be -131.8, much lower that of the wild type, making the RNA more rigid than functionally desired (Figure 4.2). It would be interesting to test an LRI-II-restoring mutant that restores the free energy of this region to the wild type levels to confirm or refute this possibility.

Figure 4.2: Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing mutations in LRI-II



	Wild Type (SJ2)	RK11	RK14
U5 (X-II) Gag (Y-II)	**5' CUUGUUCC 3' 3' GAACGGGG 5'**	**5' ggggcaag 3' 3' ccuuguuc5'**	**5' aggaggagg 3' 3' uccuccucc 5'**
RPE	1.00	0.03	0.12
ΔG (kcal/mol)	-124.7	-128.6	-131.8

Figure 4.2. Mfold secondary structure predictions of the mutant transfer vector packaging signal RNAs containing mutations in LRI-II. (A) MPMV wild type transfer vector SJ2; ΔG = -124.7 kcal/mol. **(B)** RK11 mutant transfer vector contains a flipped LRI-II secondary structure; ΔG= -128.6 kcal/mol. **(C)** RK14 mutant transfer vector where an artificial LRI-II is created by substituting the Y-II strand with heterologous sequence complementary to the one introduced in RK13, thus restoring the complementarity of between X-II and Y-II; ΔG= -131.8 kcal/mol. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

Our mutational analysis further confirms the SHAPE-validated structure of LRI-II (Aktar et al., 2013) compared to the predicted one by Mfold (Jaballah et al., 2010). When attempts were made to disrupt LRI-II using the Mfold predicted LRI-II, we observed that the mutation that was supposed to recreate the LRI-II in RK11, was unable to restore the structure (Figure 4.2). However, this was not the case with RK14, the clone with the compensatory heterologous mutation that re-created LRI-II based

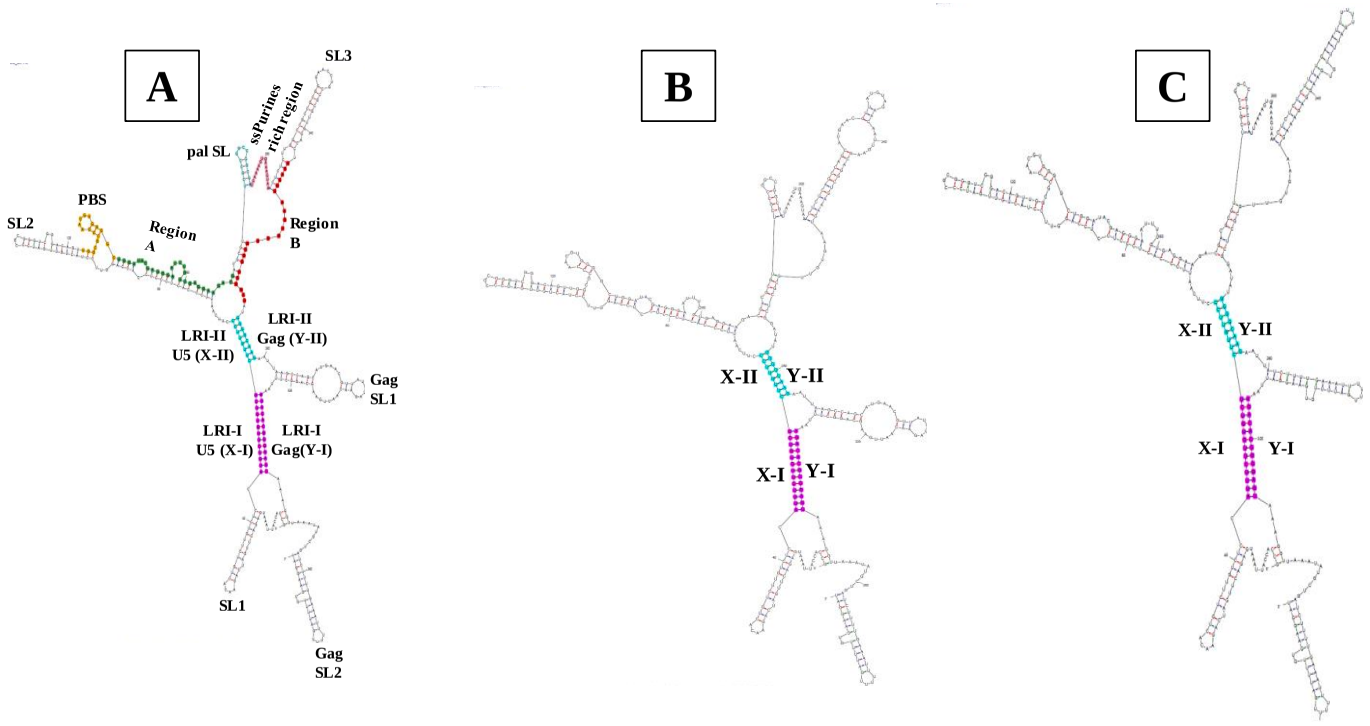
on the SHAPE-validated structure. This structure maintains an unpaired uridine at position 62 (U62), making the length of LRI-II 9 nucleotides, rather than 8 nucleotide as predicted earlier (Jaballah et al., 2010).

As has been shown, the MPMV SHAPE-validated structure appears to be anchored by complementary U5 and Gag sequences forming the LRIs. These LRIs have been suggested to play architectural role in stabilizing the RNA secondary structure of the 5' UTR sequences that are important for MPMV RNA packaging. Based on the spatial organization of the SHAPE-validated structure, it is reasonable to suggest that while LRIs stabilize 5' UTR sequences, the overall RNA secondary structure is further architecturally held together by the three stem loops (SL3, Gag SL1, and Gag SL2) comprising of sequences from the distal parts of the 5' UTR and Gag (but excluding the Gag sequences involved in forming U5-Gag LRIs; Figure 3.7). Deletion of these stem loops either individually or in multiples revealed that they were dispensable for RNA packaging and propagation (Figure 3.7). It required a rather large 102 nucleotide triple stem loop deletion to finally reduce packaging and propagation by five to ten-folds (RK20 in Figure 3.7). These observations support our hypothesis that the MPMV packaging signal RNA secondary structure

can be divided into two parts horizontally. The upper part of the structure comprising of several structural motifs (for example regions A and B, the dimerization initiation site in pal SL, and ssPurine-rich region) that have been shown to be important for MPMV RNA packaging and dimerization (Schmidt et al., 2003; Jaballah et al., 2010; Aktar et al., 2013), and the lower part of the structure containing the non-LRI forming three stem loops, that serve to provide architectural support to the overall RNA secondary structure during the process of RNA packaging and dimerization (Figure 3.7).

Consistent with this hypothesis, our results confirm that the primary sequence forming SL3 found in close vicinity of the ssPurine-rich region is not important since its deletion in RK15 only marginally affected RNA packaging or propagation (Figure 3.14). Substitution of the deleted 33 nt of SL3 with a heterologous sequence maintained the overall structure, including the ssPurine-rich region in its single stranded form (hypothetical Structure A in Figure 4.3). In hypothetical structure A, 33 nt forming SL3 were substituted with the 39 nt sequence of Gag SL1. The structure prediction of the hypothetical clone A revealed an intact structure with Gag SL1 secondary structure being duplicated at the SL3 location while still maintaining the same apical loop of the native Gag SL1 (Figure 4.3). Similarly, in hypothetical Structure B, two of the three stem loops namely SL3 and Gag SL1 sequences were substituted with the 30 nt sequence of Gag SL2. Consistent with hypothetical Structure A, the predicted secondary structure of hypothetical clone B showed a copy of Gag SL2 structure at the site of the deleted SL3 and Gag SL1 while still maintaining the same apical uridine loop of the native Gag SL2 as well as the overall secondary structure of the MPMV packaging signal RNA intact (Figure 4.3). Together, structural analyses of the single and multiple

Figure 4.3: Mfold secondary structure predictions of the hypothetical mutant transfer vector packaging signal RNAs containing stem loop substitution mutations in the packaging signal.



Clone	Wild Type (SJ2)	A	B
RPE	1.00	-	-
ΔG (kcal/mol)	-124.7	-120.6	-117.0

Figure 4.3. Mfold secondary structure predictions of the hypothetical mutant transfer vector packaging signal RNAs containing stem loop substitution mutations in the packaging signal. (A) MPMV wild type transfer vector SJ2; ΔG = -124.7 kcal/mol. **(B)** Mutant A containing a substitution of SL3 with Gag SL1; **Δ**G= -120.6 kcal/mol. **(C)** Mutant B containing a double substitution mutation of SL3 and Gag SL1 with Gag SL2; ΔG= -117.0 kcal/mol. Purple and blue colors highlight the LRI-I and LRI-II regions, respectively.

deletion mutants (RK15-RK20) and hypothetical clones (A and B) confirm the spatial role of the SL3 sequences in MPMV genomic RNA packaging which is consistent, as has been proposed previously (Jaballah et al., 2010).

Mutational analysis of MPMV LRIs involving complementary U5 and Gag sequences is in conformity with earlier observations which have shown the role Gag sequences in RNA packaging, further arguing in favor of the existence of LRIs *in vivo* and their functional role in MPMV life cycle. For example, Schmidt et al., have shown that deletion of the Gag sequences involved in U5-Gag LRIs, while maintaining the U5 sequences that are involved in U5-Gag LRI-I, severely diminished MPMV RNA packaging ability (Schmidt et al., 2003). Moreover, when the Gag sequences were cloned back into the transfer vector, resulting in the maintenance of the now known U5-Gag LRIs, RNA packaging was restored to the wild type levels (Schmidt et al., 2003). These observations are consistent to those that have been shown in other retroviruses (HIV-1, HIV-2, and FIV) where complementary sequences in the R/U5 region and within Gag interact to stabilize the overall RNA secondary structure essential for RNA dimerization and packaging (Paillart et al. 2002; Abbink and Berkhout 2003; Kenyon et al. 2008, 2011; Song et al. 2008; Rizvi et al. 2010; Lu et al. 2011b).

LRI structures have also been shown to be important for RNA dimerization in HIV-1 as the presence of a duplex called U5-AUG forms between nucleotides 105-115 in the U5 region and 334-344 surrounding the AUG initiation codon (Abbink and Berkhout, 2003). This duplex is present in the branched multiple hairpin (BMH) conformation of the HIV-1 gRNA which allows for dimerization as well as packaging. The U5-AUG duplex contains four G-U base pairs out of its 11

base pairs and it has been shown that stabilizing the U5-AUG duplex G-U base pairs by converting them to G-C and A-U base pairs increase the dimer formation rate when compared to the wild type. Such an increase has been attributed to the fact that any mutation introduced into the U5 portion of the duplex destabilizes the long distant interactions (LDI) conformation of HIV-1 RNA and induces a shift from the LDI to the BMH conformation (Abbink and Berkhout, 2003). Moreover, stabilizing the duplex by changing the base pairing pattern seems to hold tightly the structure of BMH conformation, resulting in increase of dimerization more than the wild type (Abbink and Berkhout, 2003). This is in sharp contrast to the observation made in the case of MPMV LRIs since stabilizing the interaction between the U5 and Gag sequences in LRI-I as well as LRI-II severely affected RNA packaging and propagation (Figures 3.3 and 3.8). The interpretation of the negative effect on MPMV RNA packaging and propagation when the LRIs are further stabilized, as discussed above, could be attributed to the differences in the minimal free energy between the stabilized, therefore more rigid LRIs, and the wild type flexible LRI structures (Figure 4.1). Thus, the flexible nature of LRIs is necessary for its efficient function since it ensures that the genomic RNA can properly fold or unfold dynamically during dimerization and packaging.

Besides stabilizing the secondary structure of the packaging signal RNA, the structural role of LRIs in packaging may reside in the specific selection of RNA that is to be packaged into the virus particle. The selection strategy depends on the fact that the LRI-I and LRI-II structures involve sequences within Gag downstream of the mSD which should restrict such a conformation only to the full length unspliced RNA. A similar sort of selection mechanism has also been suggested for HIV-1 gRNA containing U5-AUG duplex because in such a conformation, the Gag AUG

start codon is occluded, downregulating Gag protein translation and thereby allowing the specific packaging of the full length unspliced RNA in the budding virus particles (Abbink and Berkhout, 2003). Similar to this scenario, the MPMV LRIs may also have a role in the translation of the Gag protein since the guanine (G) residue of the Gag AUG start codon is within the U5-Gag complementary sequence of LRI-II. This plausibly could allow the regulation of temporal switch between viral protein synthesis and packaging, conferring high specificity of packaging that is limited exclusively to the full length unspliced RNA.

The work presented here, to the best of our knowledge, is the first study that validate the existence and biological significance of the two LRIs at the 5' end of the MPMV genome that have been shown to be important for its RNA packaging and dimerization (Jaballah et al., 2010; Aktar et al., 2013). While these LRIs play an important role in augmenting MPMV RNA packaging, they function at different levels to facilitate this crucial step in the viral life cycle. Specifically, LRI-I functions at the secondary structural level, whereas LRI-II functions at both the primary sequence as well as in its native structural context levels. Our results also suggest that U5-Gag LRIs have a more important architectural role in stabilizing the structure of the 5' UTR sequences, while the three stem loops (SL3, Gag SL1, and Gag SL2) may have a more secondary role in stabilizing the overall RNA secondary structure of this region. The data presented here should enhance our understanding pertaining to the molecular intricacies involved during the MPMV RNA packaging and dimerization processes.

Chapter 5: Conclusions & Recommendations

5.1. Conclusions

Selective and specific genomic RNA (gRNA) packaging into the budding virion is an essential step in retroviral life cycle. The specificity of gRNA packaging is conferred by recognition of specific *cis*-acting sequences, the packaging signal (Ψ), present at the 5'end of the retroviral genome, which interacts with the nucleocapsid protein. Employing a combination of structural prediction, phylogenetic, biochemical, and genetic analyses, we have earlier shown that MPMV packaging determinants (5' UTR and beginning of Gag) fold into a stable structure comprising of several stable stem loops (Figure 1.13; Schmidt et al., 2003, Jaballah et al., 2010; Aktar et al., 2013). The prevailing structural model reveals a characteristic feature of long-range interactions (LRIs) involving the complementary sequences within the U5- and Gag regions. These LRIs (LRI-I and LRI-II) have been found to be phylogenetically conserved, maintaining a high degree of complementarity between different strains of MPMV both at the secondary structure as well as the primary sequence levels (Figures 1.14 and 1.15; Aktar et al., 2013). Based on these observations, we hypothesized that LRI-II and LRI-II may play architectural role in stabilizing the RNA secondary structure of MPMV packaging signal RNA which is further held together by three stem loops (SL3, Gag SL1, and Gag SL2), comprising of sequences from the distal parts of the 5' UTR and Gag (but excluding the Gag sequences involved in forming U5-Gag LRIs; Figure 3.6).

The higher order features of the MPMV packaging signal RNA have not been investigated empirically for their significance to MPMV replication. Therefore, to establish the biological significance of the LRIs and the other three

stem loops (SL3, Gag SL1, and Gag SL2) within the MPMV packaging signal RNA, we introduced a series of mutations in this region. As a first step, we tested whether thes two LRIs could potentially play a role in MPMV gRNA packaging by maintaining the overall RNA structure or whether this involvement was at the primary sequence level, or both. Towards this end, a series of deletion/substitution mutations were introduced into the U5 and Gag sequences that either disrupted the complementarity or restored it artificially using non-viral sequences (Figures 3.2 and 3.4). Next, a systematic deletion analysis of the sequences involved in forming SL3, Gag SL1, and Gag SL2 was performed (Figure 3.6). These mutants were tested in a biologically relevant *in vivo* packaging and transduction assay for their effect on RNA packaging and propagation (Figures 3.3, 3.5, and 3.7). Finally, an attempt was made to establish structure-function relationships between the biological effects of the introduced mutations and their predicted RNA secondary structure. Results obtained from these multi-pronged approaches have revealed that:

1. It is the secondary structure of the LRI-I, with a moderate level of flexibility in the complementary sequences between U5 and Gag sequences, that is essential for MPMV RNA packaging and propagation. This argument stems from the observation that increasing the embedded energy due to the substitution of stable G-C base pairing in mutant clone RK1 made the LRI-I structures more rigid than what was needed, causing a significant drop in MPMV RNA packaging (Figure 4.1). Consistent with this observation, re-establishment of LRI-I with heterologous sequences in RK7 restored both the minimal free energy embedded in the secondary structure as well as RNA packaging to the wild type levels (Figure 4.1).

2. LRI-II functions at both the primary sequence as well as in its native structural context levels. This is substantiated by the observations that recreation of the artificial LRI-II using primary viral sequences in mutant clone RK11 (retaining the primary viral sequences involved in forming LRI-II, however, in an opposite orientation, thus re-establishing an artificial LRI-II) failed to restore packaging or propagation to wild-type levels (Figure 4.2).

3. Sequences involved in forming SL3, Gag SL1, and Gag SL2 do not play crucial role at individual levels during MPMV RNA packaging and propagation since deleting these stem loops individually only marginally affected RNA packaging when compared to the wild type (Figure 3.7). Furthermore it is only when a more drastic 102 nt deletion of sequences involved in forming SL3, Gag SL1, and Gag SL2 was introduced in mutant clone RK20 that a significant drop in RNA packaging and propagation could be observed (Figure 3.7).

4. U5-Gag LRIs have a more important architectural role in stabilizing the structure of the 5' UTR, while the three stem loops (SL3, Gag SL1, and Gag SL2) seems to have a rather secondary role in stabilizing the overall higher order features of MPMV packaging signal RNA.

5.2. Recommendations

Work presented in this thesis has systematically addressed and unraveled the role of LRIs in MPMV RNA packaging process. As always, this study has also raised new questions which need to be addressed further in order to unravel the molecular

intricacies involved in MPMV RNA packaging. A potential future project to build upon this work could address the following:

1. Validate the mutant transfer vector RNA structure predictions employing biochemical approaches such as SHAPE to correlate the biological results that are obtained with the validated structure rather than the predicted ones.
2. Considering the importance of the embedded energy in RNA secondary structures, create an artificial LRI-II- restoring mutant that maintains the free energy of this region to the wild type levels and monitor the effects of this mutant in on RNA packaging to ascertain whether a certain energy requirement of the re-established LRI-II is required or not to restore function.
3. Mfold predictions have suggested that the primary sequence involved in the formation of SL3 is not essential for the RNA packaging and propagation. Therefore, to further confirm this observation, a mutant clone could be designed in which SL3 is substituted with a heterologous sequence and tested for function.

Bibliography

1. Abbink, T.E., & Berkhout, B. (2003). A novel long distance base-pairing interaction in human immunodeficiency virus type 1RNA occludes the Gag start codon. J. Biol. Chem., 278, 11601–11611.

2. Aktar, S.J., Jabeen, A., Ali, L.M., Vivet-Boudou, V., Marquet, R., & Rizvi, T.A. (2013). SHAPE analysis of the 5' end of the Mason-Pfizer monkey virus (MPMV) genomic RNA reveals structural elements required for genome dimerization. *RNA, 19,* 1648-1658.

3. 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, 14*, 11-96.

4. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). Control of gene expression. *Molecular Biology of the Cell.* 4th Edition. Garland Science Publishing. N. Y. and London.

5. Al Dhaheri, N.S., Phillip, P.S., Ghazawi, A., Ali,J., Beebi, E., Jaballah, S.A., & Rizvi,T.A. (2009). Cross-packaging of genetically distinct mouse & primate retroviral RNAs. *Retrovirology, 6,* 66.

6. Al Shamsi, I.R., Al Dhaheri, N.S., Phillip, P.S., Mustafa, F., & Rizvi, T.A. (2011). Reciprocal cross-packaging of primate lentiviral (HIV-1 & SIV) RNAs by heterologous non-lentiviral MPMV proteins. *Virus Res., 155,* 352-357.

7. Balvay, L., Lopez Lastra, M., Sargueil, B., Darlix, J. L., & Ohlmann, T. (2007). Translational control of retroviruses. *Nat Rev Microbiol., 5*, 128-40.

8. Bender, W., Chein, Y.-H., Chattopadhyay, S., Vogt, P. K., Gardner, M. B. & Davidson, N. (1978). High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. *Journal of Virology, 25*, 888–896.

9. Berkhout, B., & van Wamel, J.L. (1996). Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J. Virol., 70,* 6723-6732.

10. 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. *Proc Natl Acad Sci USA 91,* 1256-1260.

11. Brierley, I., & Dos Ramos, F.J. (2006). Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res., 119,* 29-42.

12. Browning, M.T., Schmidt, R.D., Lew, K.A., & Rizvi, T.A. (2001). Primate and feline lentivirus vector RNA packaging and propagation by heterologous lentivirus virions. *J Virol., 75,* 5129-40.

13. Browning, M.T., Mustafa, F., Schmidt, R.D., Lew, K.A., & Rizvi, T.A. (2003a). Sequences within the gag gene of feline immunodeficiency virus (FIV) are important for efficient RNA encapsidation. *Virus Res., 93,* 199-209.

14. Browning, M.T., Mustafa, F., Schmidt, R.D., Lew, K.A., & Rizvi, T.A. (2003b). Delineation of sequences important for efficient FIV RNA packaging. *J. Gen. Virol., 84,* 621-627.

15. Bryant, M.L., Gardner, M.B., Marx, P.A., Maul, D.H., Lerche, N.W., Osborn, K.G., Lowenstine, L.J., Bodgen, A., Arthur, L.O., & Hunter, E. (1986). Immunodeficiency in rhesus monkeys associated with the original Mason-Pfizer monkey virus. *J Natl Cancer Inst.* 77, 957-65.

16. Bukrinsky, M. I., S. Haggerty, M. P. Dempsey, N. Sharova, A. Adzhubel, L. Spitz, P. Lewis, D. Goldfarb, M. Emerman, and M. Stevenson. 1993. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature, 365,* 666-669.

17. Clever,J. L., Wong,M. L., & Parslow,T. G. (1996) Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J. Virol., 70,* 5902-5908.

18. Clever, J.L., & Parslow T.G. (1997). Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol., 71,* 3407-3414.

19. Clapham, P. R., J. D. Reeves, G. Simmons, N. Dejucq, S. Hibbittsand, and A. McKnight. (1999). HIV coreceptors, cell tropism and inhibition by chemokine receptor ligands. *Mol. Membr. Biol., 16,* 49-55.

20. Cochrane, A. W., M. T. McNally, and A. J. Mouland. (2006). The retrovirus RNA trafficking granule: from birth to maturity. *Retrovirology, 3,* 18.

21. Cullen, B.R., & Greene, W.C. (1990). Functions of the auxiliary gene products of the human immunodeficiency virus type 1. *Virology, 178,* 1-5.

22. Daniel, M. D., N. W. King, N. L. Letvin, R. D. Hunt, P. K. Sehgal, and R. C. Desrosiers. 1984. A new type D retrovirus isolated from macaques with an immunodeficiency syndrome. *Science, 223,* 602-605.

23. Desrosiers, R. C., M. D. Daniel, C. V. Butler, D. K. Schmidt, N. L. Letvin, R. D. Hunt, N. W. King, C. S. Barker, and E. Hunter. 1985. Retrovirus D/New England and its relation to Mason-Pfizer monkey virus. *J. Virol., 54,* 552-560.

24. D'Souza, V., & Summers, M. F. (2005). How retroviruses select their genomes. *Nat. Rev. Microbiol., 8,* 643-655.

25. Ernst, R. K., Bray, M., Rekosh, D., and M. L. Hammarskjöld. 1997. A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. *Mol. Cell Biol. 17,* 135-144.

26. Fu, W. & Rein, A. (1993). Maturation of Dimeric Viral RNA of Moloney Murine Leukemia Virus. *Journal of Virology,* 5443-5449.

27. Fine, D. L., Landon, J. C., Pienta, R. J., Kubicek, M. T., Valerio, M. G., Loeb, W. F. & Chopra, H. C. (1975). Responses of infant rhesus monkeys to inoculation with Mason Pfizer monkey virus materials. *Journal of the National Cancer Institute, 54,* 651–658.

28. 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 Infect., 3,* 767.

29. Gibbs, J. S., Regier, D. A., & Desrosiers, R. C. (1994) Construction and in vitro properties of SIV mac mutants with deletions in "nonessential" genes. *AIDS Res. Hum. Retrovirus, 10,* 607-616.

30. Guesdon, F. M. J., Greatorex, J., Rhee, S. R., Fisher, R., Hunter, E., & Lever, A. M. L. (2001). Sequences in the 5' leader of Mason-Pfizer monkey virus which affect viral particle production and genomic RNA packaging: development of MPMV packaging cell lines. *Virology, 288,* 81-88.

31. Griffin, S.D., Allen, J.F., & Lever, A.M. (2001). The major human immunodeficiency virus type 2 (HIV-2) packaging signal is present on all HIV-2 RNA species: cotranslational RNA encapsidation and limitation of Gag protein confer specificity. *J. Virol., 75,* 12058-69.

32. Groom, H.C., Anderson, E.C., & Lever, A.M. (2009). Rev: beyond nuclear export. *J. Gen. Virol., 90,* 1303-1318.

33. Guan, Y., Whiteny, J.B., Liang, C., & Wainberg, M.A. (2001). Novel, live attenuated simian immunodeficiency virus constructs containing major deletions in leader RNA sequences. *J. Virol., 75,* 2776-2785.

34. Harrison, G. P., Hunter, E., & Lever, A. M. L. (1995). Secondary structure model of the Mason-Pfizer monkey virus 5' leader sequence: identification of a structural motif common to a variety of retroviruses. *J. Virol., 69,* 2175-2186.

35. Hibbert, C. S., Mirro, J., & Rein, A. (2004). mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. *J. Virol., 78,* 10927-10938.

36. Honigman, A., D. Wolf, S. Yaish, H. Falk, and A. Panet. 1991. *Cis*-acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus. *Virology, 183,* 313-319.

37. Housset, V., H. De Rocquigny, B. P. Roques, and J. L. Darlix. 1993. Basic amino acids flanking the zinc finger of Moloney murine leukemia virus nucleocapsid protein NCp10 are critical for virus infectivity. *J. Virol., 67,* 2537-2545.

38. Hu, W.S., & Temin, H.M. (1990). Retroviral recombination and reverse transcription. *Science, 250,* 1227-33.

39. Hu, W.S., & Hughes, S.H. (2012). HIV-1 reverse transcription. *Cold Spring Harb Perspect Med.*

40. Jaballah, Soumaya Ali Jaballah. Delineation of MPMV RNA Packaging Determinants: Structure-Function Correlation of the 5' end of MPMV RNA Genome. Degree of Master of Medical Science in Microbiology & Immunology. United Arab Emirates University (UAE). 2010.

41. 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. *J. Mol. Biol., 401,* 996-1014.

42. Jeang, K. T., H. Xiao, and E. A. Rich. 1999. Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *J. Biol. Chem., 274,* 28837-28840.

43. Johnson, S. F., & Telesnitsky, A. (2010). Retroviral RNA dimerization and packaging: the what, how, when, where, and why. *PLoS Pathog., 6,* e1001007.

44. Jouvenet, N., Lainé, S., Pessel-Vivares, L., & Mougel, M. (2011). Cell biology of retroviral RNA packaging. *RNA Biol., 8,* 572–580.

45. Kaye, J.F., & Lever, A.M.L., (1999). Human immunodeficiency virus types 1 and 2 differ in the predominant mechanism used for selection of genomic RNA for encapsidation. *J. Virol., 73,* 3023- 3031.

46. Kenyon, J. C., Ghazawi, A., Cheung, W. K., Phillip, P. S., Rizvi, T. A., & Lever, A. M. **(**2008). The secondary structure of the 5' end of the FIV genome reveals a long-range interaction between R/U5 and gag sequence, and a large, stable stem loop. *RNA, 14,* 2597.

47. Kenyon, J. C., Tanner, S., Legiewicz, M., Phillip, P. S., Rizvi, T. A., LeGrice, S., & Lever, A. M. (2011). SHAPE analysis of the FIV leader RNA reveals a structural switch potentially controlling viral packaging and genome dimerization. *Nucleic Acids Res., 39,* 6692-6704.

48. Kuzembayeva, M., Dilley, K., Sardo, L., & Hu, W. S. (2014). Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology,* doi: 10.1016/j.virol.2014.01.019. (Epub ahead of print).

49. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., & Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature, 393,* 648-59.

50. Lanchy, J. M., Ivanovitch, J. D., & Lodmell, J. S. (2003). A structural linkage between the dimerization and encapsidation signals in HIV-2 leader RNA. *RNA, 9,* 1007-1018.

51. Lanchy, J. M., & Lodmell, J. S. **(**2007). An extended stem-loop 1 is necessary for human immunodeficiency virus type 2 replication and affects genomic RNA encapsidation. *J. Virol., 7,* 3285-3292.

52. Laughrea, M., Jetté, L. (1996). Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248-271 are dispensable for dimer formation. *Biochemistry,35,* 1589-98.

53. 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. *J. Virol., 71,* 3397-3406.

54. Laughrea, M., Shen, N., Jetté, L., & Wainberg, M.A.(1999). Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: role of stem-loop B in HIV replication and HIV RNA dimerization. *Biochemistry, 5,* 226-34.

55. Lever, A. M. L. (2007). HIV RNA packaging. *Adv. Pharmacol., 55,* 1-32.

56. Levin, J.G., Grimley, P.M., Ramseur, J.M., & Berezesky, I.K. (1974). Deficiency of 60 to 70S RNA in murine leukemia virus particles assembled in cells treated with actinomycin D. *J. Virol., 14,* 152–161.

57. Levin, J.G., Mitra, M., Mascarenhas, A., & Musier-Forsyth, K. (2010). Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol., 7,* 754-774.

58. Leitner, T., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Wolinsky, S., & Korber, B. (2005). HIV sequence compendium. *Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 06-0680.*

59. Lu, K., Heng, X., & Summers, M. F. (2011a). Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol., 410,* 609-633.

60. 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. (2011b). NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science, 334,* 242-245.

61. Mallona, I., Weiss, J., & Marcos, E.C. (2011). pcr Efficiency: a Web tool for PCR amplification efficiency prediction. *BMC Bioinformatics, 12,* 404.

62. Marx, P.A., Maul, D.H., Osborn, K.G., Lerche, N.W., Moody, P., Lowenstine, L.J., Henrickson, R.V., Arthur, L.O., Gilden, R.V., & Gravell, M. (1984). Simian AIDS: isolation of a type D retrovirus and transmission of the disease. *Science, 223,*1083–1086.

63. Mathews, D. H., Sabina, J., Zuker, M., & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol., 288,* 911-940.

64. McBride, M.S., & Panganiban, A.T. (1997). Position dependence of functional hairpins important for human immunodeficiency virus type 1 encapsidation in *vivo*. *J. Virol., 71,* 2050-2058.

65. Miyazaki, Y., Miyake, A., Nomaguchi, M., & Adachi, A. (2011). Structural dynamics of retroviral genome and the packaging. *Frontiers in Microbiology, 2,* 1-9.

66. Moore, M. D., & Hu, W. S. (2009). HIV-1 RNA dimerization: It takes two to tango. *AIDS Rev., 11,* 91-102.

67. Murti, K. G., M. Bondurant, and A. Tereba. 1981. Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *J. Virol. 37,* 411-419.

68. Muñoz-Barroso, I., Durell, S., Sakaguchi, K., Appella, E., & Blumenthal, R. (1998). Dilation of the human immunodeficiency virus-1 envelope

glycoprotein fusion pore revealed by the inhibitory action of a synthetic peptide from gp41. *J Cell Biol., 140,* 315-23.

69. 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 RNA's. *J. Virol., 7,* 13817-13821.

70. Mustafa, F., Al Amri, D., Al Ali, F., Al Sari, N., Al Suwaidi, S., 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, e47088.

71. 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,* 263-267.

72. Paillart, J. C., Berthoux, 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. *J. Virol., 70,* 8348-8354.

73. Paillart, J. C., Skripkin, E., Ehresmann, B., Ehresmann, C., Marquet, R. (2002). *In vitro* evidence for a long range pseudoknot in the 5′-untranslated and matrix coding regions of HIV-1 genomic RNA. *J. Biol. Chem., 277,* 5995–6004.

74. Paillart, J. C., Xhilag, M. S., Marquet, R., & Mak, J. (2004). Dimerization of retroviral RNA genomes: An inseparable pair. *Nat. Rev. Microbiol., 2,* 461-472.

75. Patel, J., Wang, S.W., Izmailova, E., Aldovini, A. (2003). The simian immunodeficiency virus 5′ untranslated leader sequence plays a role in intracellular viral protein accumulation and in RNA packaging. *J. Virol. 77,* 6284–6292. doi: 10.1128/jvi.77.11.6284-6292.2003

76. Pavlakis, G.N., & Felber, B.K. (1990). Regulation of expression of human immunodeficiency virus. *New Biol., 1,* 20-31.

77. Pedersen F.S., & Duch, M. (2006). Retroviral Replication. Encyclopaedia of life sciences, John Wiley & Sons, Ltd. Doi: 10.1038/npg.els.004239.

78. Pilkington, G.R., Purzycka, K.J., Bear, J., Le Grice, S.F., & Felber, B.K. (2014). Gammaretrovirus mRNA expression is mediated by a novel, bipartite post-transcriptional regulatory element. *Nucleic Acids Res., 42,* 11092-106.

79. Rizvi, T.A., & Panganiban, A.T. (1993). Simian immunodeficiency virus RNA is efficiently encapsidated by human immunodeficiency virus type 1 particles. *J. Virol., 67,* 2681-2688.

80. Rizvi, T.A., Schmidt, R.D., Lew, K.A., & Keeling, M.E. (1996a). Rev/RRE-independent Mason-Pfizer monkey virus constitutive transport element-dependent propagation of SIVmac239 vectors using a single round of replication assay. *Virology, 222,*457-463.

81. Rizvi, T.A., Lew, K.A., Murphy, E.C.Jr, & Schmidt, R.D. (1996b). Role of Mason-Pfizer monkey virus (MPMV) constitutive transport element (CTE) in the propagation of MPMV vectors by genetic complementation using homologous/heterologous env genes. *Virology, 224,* 517-532.

82. Rizvi, T.A., Ali, J., Phillip, P.S., Ghazawi, A., & 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,* 464-472.

83. Rizvi, T. A., Kenyon, J. C., Ali, J., Aktar, S. J., Phillip, P. S., Ghazawi, A., Mustafa, F., & Lever, A. M. L. (2010). Optimal packaging of FIV genomic RNA depends upon a conserved long-range interaction and a palindromic sequence within gag. *J. Mol. Biol., 403,* 103-119.

84. Russell, R. S., Liang, C. and Wainberg, M. A. (2004). Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology, 1,* 23

85. Schmidt, R. D., Mustafa, F., Lew, K. A., Browning, M. T., & Rizvi, T. A. (2003). Sequences within both the 5' untranslated region and the gag gene are important for efficient encapsidation of Mason-Pfizer monkey virus RNA. *Virology, 309,* 166-178.

86. Schröder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., & Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell,110,* 521-9.

87. Shibagaki, Y., & Chow, S.A. (1997). Central core domain of retroviral integrase is responsible for target site selection. *J. Biol. Chem., 272,* 8361-9.

88. Song, R., Kafaie, J., & Laughrea, M. (2008). Role of the 5′ TAR stem-loop and the U5-AUG duplex in dimerization of HIV-1 genomic RNA. *Biochemistry, 47,* 3283–3293.

89. Sonigo, P., Barker, C., Hunter, E., & Wain-Hobson, S. (1986). Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell, 45,* 375-385.

90. Stoye, J. P. (2012). Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat. Rev. Microbio.10,* 395-406.

91. Strappe, P.M., Greatorex, J., Thomas, J., P. Biswas, E. McCann, & Lever, A.M. (2003). The packaging signal of simian immunodeficiency virus is upstream of the major splice donor at a distance from the RNA cap site similar to that of human immunodeficiency virus types 1 and 2. *J. Gen. Virol., 84,* 2423-2430.

92. Swanson, C.M., &. Malim, M.H. (2006). Retrovirus RNA trafficking: from chromatin to invasive genomes. *Traffic, 7,* 1440-1450.

93. Tan, W., Felber, B.K., Zolotukhin, A.S., Pavlakis, G.N., & Schwartz, S. (1995). Efficient expression of the human papillomavirus type 16 L1 protein in epithelial cells by using Rev and the Rev-responsive element of human immunodeficiency virus or the cis-acting transactivation element of simian retrovirus type 1. *J. Virol., 69,* 5607–5620.

94. Telestinsky, A. & Goff, S. P. (1997). Reverse Transcriptase and the Generation of Retroviral DNA. In: Coffin JM, Hughes SH, Varmus HE, editors. Retroviruses. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

95. Van Maele, B., Busschots, K., Vandekerckhove, L., Christ, F., & Debyser, Z. (2006). Cellular co-factors of HIV-1 integration. *Trends Biochem. Sci, 2,* 98-105.

96. Verma, I. M., & Weitzman, M. D. (2005). Gene therapy: twenty-first century medicine. *Annu Rev Biochem, 74,* 711-38.

97. Vile, R.G., Ali, M., Hunter, E., & McClure, M. O. (1992). Identification of a generalized packaging sequence for D-type retroviruses and generation of a D-type retroviral vector. *Virology, 189,* 786-791.

98. Whitney J.B., & Wainberg, M.A. (2006). Impaired RNA incorporation and dimerization in live attenuated leader-variants of SIVmac239. *Retrovirology 3,* 96.

99. Yasutsugu, S., Lok, C.M, Youichi, S. (2012). Role of host-encoded proteins in restriction of retroviral integration. *Frontiers in Microbiology,3.*

100. Yoshinaka, Y., Katoh, I., Copeland, T.D., & Oroszlan, S. (1985). Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc. *Natl. Acad. Sci. U.S.A., 82,* 1618-1622.

101. Youngsuk, Yi, Moon Jong, Noh, & Kwan H.L. (2011). Current Advances in Retroviral Gene Therapy. *Curr. Gene Ther., 3,* 218–228.

102. Yu, S. F., Baldwin, D. N., Gwynn, S. R., Yendapalli, S. & Linial, M. L. (1996). Human foamy virus replication: a pathway distinct from that of retroviruses and hepadnaviruses. *Science, 271*, 1579–1582.

103. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res., 31*, 3406-3415.