SHAPE Validated Secondary Structures of Mason-Pfizer Monkey Virus (MPMV) and Mouse Mammary Tumor Virus (MMTV) Packaging Signal RNAs Reveal Pal Helix Loops Functioning as Dimerization Initiation Sites (DIS) Controlling their Genomic RNA (gRNA) Packaging

Suriya Jahan Aktar

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

Part of the Medical Microbiology Commons

Recommended Citation
Aktar, Suriya Jahan, "SHAPE Validated Secondary Structures of Mason-Pfizer Monkey Virus (MPMV) and Mouse Mammary Tumor Virus (MMTV) Packaging Signal RNAs Reveal Pal Helix Loops Functioning as Dimerization Initiation Sites (DIS) Controlling their Genomic RNA (gRNA) Packaging" (2014). Dissertations. 28.
https://scholarworks.uaeu.ac.ae/all_dissertations/28

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarworks@UAEU. It has been accepted for inclusion in Dissertations by an authorized administrator of Scholarworks@UAEU. For more information, please contact fadl.musa@uaeu.ac.ae.
SHAPE Validated Secondary Structures of Mason-Pfizer Monkey Virus (MPMV) and Mouse Mammary Tumor Virus (MMTV) Packaging Signal RNAs Reveal Pal Helix Loops Functioning as Dimerization Initiation Sites (DIS) Controlling their Genomic RNA (gRNA) Packaging

Suriya Jahan Aktar

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

Under the direction of Professor Tahir A. Rizvi

April 2014
Declaration of Original Work

I, Suriya Jahan Aktar, the undersigned, a graduate student at the United Arab Emirates University (UAEU) and the author of the dissertation titled “SHAPE validated secondary structures of Mason-Pfizer monkey virus (MPMV) and mouse mammary tumor virus (MMTV) packaging signal RNAs reveal pal helix loops functioning as dimerization initiation sites (DIS) controlling their genomic RNA (gRNA) packaging”, hereby solemnly declare that this dissertation is an original work done and prepared by me under the guidance of Prof. 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 degree, diploma or similar title at this or any other university. The materials borrowed from other sources and included in my thesis/dissertation have been properly acknowledged.

Student’s Signature……………………..                     Date………………………….
Approved by

PhD Examining Committee:

1) Advisor: Prof. Tahir A. Rizvi
   Title: Professor
   Department: Department of Microbiology & Immunology
   Institution: College of Medicine and Health Sciences, UAE University
   Signature ………………………………                Date…………………………

2) Member: Dr. Ahmed Al-Marzouqi
   Title: Associate Professor
   Department: Department of Biochemistry
   Institution: College of Medicine and Health Sciences, UAE University
   Signature ………………………………                Date…………………………

3) Member: Dr. Gulfaraz Khan
   Title: Associate Professor
   Department: Department of Microbiology & Immunology
   Institution: College of Medicine and Health Sciences, UAE University
   Signature ………………………………                Date…………………………

4) Member: Dr. Farah Mustafa
   Title: Assistant Professor
   Department: Department of Biochemistry
   Institution: College of Medicine and Health Sciences, UAE University
   Signature ………………………………                Date…………………………
5) External Examiner: Dr. Wei-Shau Hu

Title: Senior Investigator and Head

Department: Viral Recombination Section, HIV Drug Resistance Program

Institution: Center for Cancer Research/NCI-Frederick

National Institutes of Health (NIH)

Frederick, MD, USA.

Signature ..................................................  Date........................................
Accepted by

Dean of the College of Medicine and Health Sciences: Prof. Tar-Ching Aw

Signature ..................................... Date.................................

Dean of the College of Graduate Studies: Prof. Nagi Wakim

Signature ..................................... Date.................................

Copy____of____
ABSTRACT

The “diploid” genome of retroviruses consists of two strands of RNA that are non-covalently linked as a “dimer” at their 5’ ends. The ubiquitous presence of a dimeric genome among retroviruses suggests that dimerization plays a crucial role in gRNA packaging during viral life-cycle. For almost all retroviruses, determinants of gRNA dimerization and packaging, which are for the most part physically and genetically indistinguishable, reside at the 5’ end of the gRNAs and have been shown to assume higher order structures.

Employing a combination of genetic and structural prediction approaches, we have earlier shown that Mason-Pfizer monkey virus (MPMV) and mouse mammary tumor virus (MMTV) packaging determinants comprise sequences at the 5’ end of the genome, starting from R and extending into the beginning of Gag. Sequences encompassing these regions were predicted to fold into stable RNA secondary structures comprising several structural motifs. In an attempt to establish structure-function relationship of the higher order features of MPMV and MMTV packaging signal RNAs, we first validated their predicted structures employing a novel chemo-enzymatic probing strategy, selective 2’hydroxyl acylation primer extension (SHAPE). The SHAPE-analyzed structures of MPMV and MMTV packaging signal RNAs validated the major structural motifs, including U5/Gag long range interactions (LRIs), a stretch of single-stranded purine (ssPurine)-rich region, and a distinctive GC-rich palindromic (pal) helix loop. Minimum free-energy structure predictions, phylogenetic, and in silico analyses of different MPMV and MMTV strains further suggested the existence of these major structural motifs.
To test the importance of the pal sequences in MPMV and MMTV gRNA dimerization and packaging, we introduced a series of mutations in the pal helix loops. Tests of these mutations employing *in vitro* and *in vivo* complementary approaches, phylogenetic, and structure prediction analyses revealed pal helix loops (5’ CGGCCG 3’ in MPMV and 5’ CGGCCG 3’ in MMTV) containing a canonical “GC” dyad functioning as dimerization initiation sites (DISs) controlling MMTV and MPMV gRNA dimerization and packaging. Furthermore, in MMTV, a second pal within the primer binding site (PBS) was also observed that was found to be involved in gRNA dimerization. Concomitant mutational analysis of pal II and PBS pal suggests that both pals are required for efficient RNA dimerization, packaging and propagation of MMTV gRNA.
الملخص العربي

يتكون المحتوى الجيني للـretroviruses من نوعين من الحمض النووي الريبي التي ترتبط بعضها عن طريق روابط غير تساهمية على شكل Dimer، ويبادأ في النهايات الخاصة بها. يوجد في هذا المحتوى الجيني المتراابط بهذا الشكل بين retroviruses يشير إلى أن gRNA dimerization في النهايات الخصبة بالنسبة للجزء الأكبر لا يمكن التمييز بينها، تقع في النهاية (5') للـgRNAs و التي تثبت أنها تكون هياكل ثانية.

بتطبيق مجموعة من المناهج للتنبؤ الوراثي والبنبوبي، لقد أظهرنا في وقت سابق بأن محددات التجميع للفيروسات موجودة في النهاية 5 من المحتوى الجيني بدأ من R وتمتد إلى بداية MMTV و MPMV. وجد أن التتابعات الممثلة لهذه المناطق تكون هياكل ثانية مستقرة تضم العديد من الزخارف الهيكلية. في محاولة لإيجاد علاقة بين الهيكل والوظيفة للهياكل الثانوية لمحددات التجميع للفيروسات MMTV و MPMV، لقد قمنا بالتحقق من صحة الهياكل المتوقعة مسبقا وذلك عن طريق chemo-enzymatic probing strategy, selective 2’hydroxyl acylation primer extension (SHAPE).

تحليل محددات التجميع للفيروسات MMTV و MPMV بواسطة SHAPe بواسطة MMTV و MPMV الرئيسي، بما في ذلك النماذج طويلة للوبة (LRIs) U5/Gag )التي تمتد إلى نهاية الحركة U5/Gag، السلاسل المختلفة من انتناج المتابعة بالبايروبينات Minimum free-energy structure predictions, phylogenetic, in silico analyses تؤكد وجود هذه الهياكل الكبير.

فهو أن أهمية تسلسل pal في محددات التجميع والـgRNA dimerization في MMTV و MPMV، مما قد يدعم سلسلة من الطفرات في حلقات pal الحلوانية. درسنا هذه الطرفات باستخدام in vitro and in vivo complementary approaches, phylogenetic, and مناهج مختلفة مثل...
(5’ CGGCCG 3’ in the pal structure prediction analyses

which were found in MPMV and 5’ CGGCCG 3’ in MMTV) dimerization

MPMV, MMTV, PBS and the pal II dimerization which is in the MMTV which observed on that, inside of the PBS and the pal II which is sharing also in the gRNA dimerization. Analysis of the fragments of the PBS and pal II which is indicating that both are essential in the dimerization, packaging and propagation of MMTV gRNA.
ACKNOWLEDGEMENTS

First and foremost, I want to thank Allah (the most Merciful, the All-knowing and the most Beneficent) whose constant guidance and support has helped me throughout my life.

I am grateful to Prof. Rizvi for being an extraordinary supervisor who showed me the road and helped me to get started on the path to the PhD degree. His enthusiasm, encouragement and faith in me throughout have been extremely helpful. He was always available for my questions and he was positive and was generous with his time and vast knowledge.

I found a very nice and unforgettable association with Dr. Roland Marquet’s mentorship during our collaboration for my PhD thesis work. I appreciate his and his family’s kindness towards me during my stay in France.

I am grateful to all my committee members (Dr. Farah Mustafa, Dr. Ahmed Al-Marzouqi and Dr. Gulfaraz Khan), who were very helpful to manage time for me from their busy schedules to arrange committee meetings every semester. My sincere appreciation and regards go to Dr. Farah Mustafa who is like a role model. I appreciate her patience that has allowed me to extract her from her busy schedule at any time for discussions.

I am extremely grateful to my current and past lab mates (Arshad, Ayesha, Elsa, Gayathri, Jaicy, Lizna, Pretty, Reka, Sahl, Soumeya, Zahabar) in UAE without whose support I would not have been able to graduate. I am thankful to the people in Dr. Roland Marquet’s lab (Delphine, J.C. Paillart, Julien, Patrick, Redmond, Santiago, Valerie) for their friendliness and for making my work in France quite enjoyable.
My deep appreciation goes to my friends, Varun, Rawan, Nabila, Jasbir, Subash, Waqar and Karima, who supported me during some of my toughest times.

In the end, my special thanks go to my family: my father, Mohd. Saleh Ahmed, who unfortunately did not live long enough to see the fruits of his hard work towards his children’s education, my strong mother, Shamim Akter who supported me in all my decisions, my brother, brother-in-law and sisters who were always available for moral support and to provide all concerned facilities to fulfill my needs by clearing away all the difficulties of my life.

And also to my little sister, Amy, who helped me with my acknowledgement, with mostly minor editing (some of which I neglected to pay any attention to) and who also added this comment.
DEDICATION

I would like to dedicate this dissertation to my family especially my late father.

I am forever indebted to you for your constant love, support and guidance which

have enabled me to bring this work to completion
TABLE OF CONTENTS

Title page i
Declaration of Original Work ii
Copyright Page iii
Signature Page iv
Abstract vii
Abstract in Arabic ix
Acknowledgements xi
Dedication xiii
Table of Contents xiv
List of Figures and Tables xvii
List of Publications xx
CHAPTER 1: Introduction 1
   1.1. Introduction 2
      1.1.1. Retroviruses 2
      1.1.2. Classification of Retrovirus 2
      1.1.3. Genome Organization and Characteristics Features 3
      1.1.4. Retroviral Life Cycle 14
      1.1.5. Retroviral gRNA Packaging 25
      1.1.6. Retroviral gRNA Dimerization 39
   1.2. Objectives 42
CHAPTER 2: SHAPE Analysis of the 5’ end of the Mason-Pfizer Monkey Virus (MPMV) genomic RNA Reveals Structural Elements Required for Genome Dimerization 45
   2.1. Abstract 46
   2.2. Introduction 47
   2.3. Materials and Methods 53
      2.3.1. MPMV RNA Secondary Structure Analyses in silico 53
      2.3.2. Plasmids Used for in vitro Transcription (wild type and pal mutants) 53
      2.3.3. RNA Preparation by in vitro Transcription 58
      2.3.4. SHAPE Methodology 58
      2.3.5. In vitro Dimerization Assay 62
2.4. Results and Discussion

2.4.1. SHAPE Data Validated the Major Structural Motifs Predicted in the 5’ end of MPMV gRNA
2.4.2. Sequences in pal SL Augments MPMV gRNA Dimerization by Functioning as DIS
2.4.3. In vitro Heterodimerization can be Mediated by trans-Complementary Sequences on two RNAs
2.4.4. The SHAPE-Validated Secondary Structural Model is Supported by Sequence Conservation between Different Mpmv Isolates
2.4.5. SL1, SL2, and SL3 are Present in Spliced as well as genomic RNAs

2.5. Acknowledgments

CHAPTER 3: A palindromic stem loop functions as the dimerization initiation site controlling packaging of the genomic RNA of mouse mammary tumor virus (MMTV)

3.1. Abstract

3.2. Introduction

3.3. Materials and Methods

3.3.1. In silico Analyses
3.3.2. Construction of Plasmids
3.3.3. In vitro Transcription Assay
3.3.4. SHAPE Methodology
3.3.5. In vitro Dimerization Assay
3.3.6. In vivo Genetic Complementation Assay
3.3.7. Nucleocytoplasmic Fractionation and Isolation of RNA, and cDNA Preparation
3.3.8. Real Time Quantitative PCR (qPCR) for Transfer Vector Packaging Efficiency

3.4. Results

3.4.1. Predicted Secondary Structure of the MMTV 5’ gRNA
3.4.2. Validation of the Secondary Structure by SHAPE
3.4.3. The SHAPE-Validated RNA Structural Model is Supported by Phylogeny
3.4.4. Identification of the pal sequence Mediating MMTV gRNA Dimerization
3.4.5. Mutational Analysis of pal I
3.4.6. Pal II Functions as the Dimerization Initiation Site (DIS)
3.4.7. Pal II Regulates MMTV gRNA Packaging and Propagation
3.4.8. SHAPE Analyses of pal II and PBS pal Mutants support their role in MMTV
## LIST OF FIGURES AND TABLES

| FIGURE 1.1 | Schematic representation of the viral RNA genome and the viral DNA along with its important cis-acting elements | 5 |
| FIGURE 1.2 | Schematic representation of simple and complex retroviral genomes | 8 |
| FIGURE 1.3 | Schematic representation of a typical retroviral particle depicting major gene products | 10 |
| FIGURE 1.4 | Schematic representation of retroviral life cycle | 13 |
| FIGURE 1.5 | Schematic representation of MPMV and MMTV genomes and their respective RNA transcripts | 17 |
| FIGURE 1.6 | Schematic representation of different pathways employed by retroviruses during nucleocytoplasmic transport of unspliced gRNAs | 19 |
| FIGURE 1.7 | Schematic representation of ribosomal frameshifting phenomenon in retroviruses to express relative amounts of Gag and Pol proteins | 21 |
| FIGURE 1.8 | Role of host restriction factors and their action during retroviral (e.g. HIV-1) replication | 24 |
| FIGURE 1.9 | Schematic representation of the prevailing model of retroviral gRNA packaging/encapsidation | 28 |
| FIGURE 1.10 | Amino acid sequences of the nucleocapsid (NC) proteins of (A) murine leukemia virus (MLV) and (B) human immunodeficiency virus type-1 (HIV-1) | 30 |
| FIGURE 1.11 | A model depicting possible mechanism(s) employed during retroviral RNA dimerization and packaging in a cell co-infected with C-type retroviruses such as human immunodeficiency virus type 1 (HIV-1) and murine leukemia virus (MLV; shown in red at top and green at the bottom) | 32 |
| FIGURE 1.12 | Schematic representation of the packaging determinants of different retroviruses (left panel), which have been universally shown to assume higher order structure schematically represented in the right panel | 35 |
| FIGURE 1.13 | Schematic representation of MLV (A) and HIV-1 (B) genomes and the secondary structure of their | 38 |
packaging signals (Psi; ψ)

FIGURE 2.1 Minimal free-energy and SHAPE-validated models of the MPMV packaging signal RNA

FIGURE 2.2 Sequences in pal SL augments MPMV gRNA dimerization by functioning as DIS

FIGURE 2.3 RNAstructure prediction of the MPMV wild type (RCR001) RNA (nt +1-388) homodimer structure

FIGURE 2.4 In vitro heterodimerization can be mediated by trans-complementary sequences on two RNAs

FIGURE 2.5 Dimer predictions of the mutant RCR005 (A) and RCR006 (B) MPMV RNAs (nt +1-388) using RNAstructure showing that mutant RNAs dimerize via a sequence in the primer binding site

FIGURE 2.6 Predicted homodimer (A-C) and heterodimer (D) structures of MPMV wild type (RCR001) and trans complementary mutants (RCR005 and RCR006)

FIGURE 2.7 Mfold structural predictions of the 5’ end genomes of different MPMV strains (A-E)

FIGURE 2.8 ClustalW sequence alignment of the 5’ end genomes of different MPMV strains

FIGURE 2.9 SL1, SL2, and SL3 are present in genomic as well as spliced RNA

TABLE 3.1 Description of primers used for cloning, sequencing, and conventional and real time PCR

FIGURE 3.1 Design and rationale of three plasmid trans complementation assay for studying mouse mammary tumor virus (MMTV) gRNA packaging and propagation developed earlier (Rizvi et al., 2009)

FIGURE 3.2 Predicted and SHAPE-validated structural models of the MMTV packaging signal RNA

FIGURE 3.3 Phylogenetic conservation of the sequences of major structural motifs of MMTV packaging signal RNA in different strains

FIGURE 3.4 Consensus RNA secondary structure prediction of MMTV packaging signal RNA using sequences from eight different strains

FIGURE 3.5 Contribution of the pal sequences to MMTV gRNA
dimerization. (A) Schematic representations of the predicted RNA secondary structures of the three truncated RNAs (712 nt, 321 nt and 265 nt)

FIGURE 3.6 Role of pal I in MMTV gRNA dimerization, packaging, and propagation. (A) Description of the pal I mutants 122-123

FIGURE 3.7 Role of pal II in MMTV gRNA dimerization, RNA packaging and viral propagation 128-129

FIGURE 3.8 Predicted homodimer structures of (A) MMTV wild type (SA035), (B) pal II deletion (SA042), (C) deletion of the pal sequence within PBS (SA051) and (D) double deletion of both pal II and pal sequence within PBS (SA046) 131

FIGURE 3.9 Effects of pal II substitution mutations on MMTV gRNA dimerization, RNA packaging and viral propagation 134-135

FIGURE 3.10 Structural analysis of the SL4 domain of the pal II and PBS pal mutants 139-140

FIGURE 3.11 Structural analysis of the SL2 domain of the pal II and PBS pal mutants 142

FIGURE 3.12 Predicted homodimer (A-B) and heterodimer (C) structures of MMTV pal II trans-complementary mutants (SA044 and SA045) 145
LIST OF PUBLICATIONS


CHAPTER 1

Introduction
1.1. Introduction

1.1.1. Retroviruses

Retroviruses are enveloped (+) sense ssRNA viruses which are members of the retroviridae family and have been intensely studied because of their association with cancers, leukemias, and a variety of immune system disorders (Coffin, 1992, 1996). Retroviruses have the unique property to replicate by “reverse transcribing” their single stranded RNA genome into a double stranded DNA in the host cell using the virally-encoded polymerase, reverse transcriptase (RT), which is then integrated into the host genome (Baltimore, 1970; Temin & Mizutani, 1970).

1.1.2. Classification of Retrovirus

Using their taxonomic relatedness, retroviruses have been classified in different genuses based on morphogenic features (Goff, 2001) as outlined below:

<table>
<thead>
<tr>
<th>Genus</th>
<th>Features</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpharetrovirus, α</td>
<td>C-type, Simple</td>
<td>Avian leukosis virus (ALV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rous sarcoma virus (RSV)</td>
</tr>
<tr>
<td>Betaretrovirus, β</td>
<td>B-, D-type</td>
<td>Mouse mammary tumor virus (MMTV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mason-Pfizer monkey virus (MPMV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jaagsiekte sheep retrovirus (JSRV)</td>
</tr>
<tr>
<td>Gammaretrovirus, γ</td>
<td>C-type, Simple</td>
<td>Murine leukemia viruses (MuLV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Feline leukemia virus (FeLV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gibbon ape leukemia virus (GaLV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reticuloendotheliosis virus (REV)</td>
</tr>
<tr>
<td>Deltaretrovirus, δ</td>
<td>Complex, Oncogenic</td>
<td>Human T-lymphotropic virus (HTLV)-1, -2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bovine leukemia virus (BLV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simian T-lymphotropic virus types 1-3 (STLV-1, -2, -3)</td>
</tr>
<tr>
<td>Epsilonretrovirus, ε</td>
<td>Complex, Oncogenic</td>
<td>Walleye dermal sarcoma virus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Walleye epidermal hyperplasia virus 1</td>
</tr>
<tr>
<td>Lentivirus</td>
<td>Complex</td>
<td>Human immunodeficiency virus type 1 and 2 (HIV-1 and -2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simian immunodeficiency virus (SIV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Feline immunodeficiency virus (FIV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Equine infectious anemia virus (EIAV)</td>
</tr>
</tbody>
</table>
1.1.3. Genome Organization and Characteristic Features

Retroviral virions range in size between 80-100 nm containing two copies of 7-12 kilo bases (kb) long single stranded RNA molecules as their genome which are non-covalently linked as a “dimer” at their 5’ end (Fig. 1.1; reviewed in Pedersen & Duch, 2006). Following reverse transcription and integration, the two RNA copies lead to only one provirus in cells infected with one virus. Therefore, retroviruses are “pseudodiploid”. The flanking ends of a typical retroviral RNA genome (which is capped and polyadenylated) contain *cis*-acting non-coding controlling sequences needed for virus replication. Between the *cis*-acting controlling sequences are present the *trans*-acting sequences responsible for coding the viral structural and enzymatic proteins. In the retroviral RNA genome, “R” (a repeat sequence) is present at both the 5’ as well as the 3’ ends whereas “U5” (unique sequence at the 5’end) and “U3” (unique sequence at the 3’end) are present at the 5’ and 3’ ends, respectively. During the course of reverse transcription, the “U3” sequences are copied at the 5’ end and the “U5” sequences are copied at the 3’ end, resulting in the formation of a complete long terminal repeat (LTR) consisting of U3-R-U5 sequences at both ends of the viral genome (Fig. 1.1). In close vicinity of *cis*-acting control sequences of the RNA genome, there are short non-coding sequences known as 5’ and 3’ untranslated leader region (UTR) which includes a primer binding site (PBS) and polypurine tract (PPT) playing vital roles during reverse transcription (Fig. 1.1). Sequences in the UTR have been shown to be involved in regulation of many essential steps of
FIGURE 1.1. Schematic representation of the viral RNA genome and the viral DNA along with its important *cis*-acting elements. A typical retroviral particle (A) contains two copies of full length unspliced viral RNA as its genome. Retroviruses have the unique property to reverse transcribe this RNA genome (B) into the DNA (C) that is integrated into the host chromosome to become proviral DNA.
Cis-acting or Controlling Elements:
- mostly are non-coding sequences
- are required for virus replication
- generally are present at either ends of the virus
- include: LTRs, PBS, psi, PPT, and att sites

Trans-acting or Coding Sequences:
- code for viral structural, enzymatic, and regulatory proteins
- can be provided in trans for virus replication
- are found between the controlling regions
viral replication cycle, including genomic RNA (gRNA) dimerization and packaging, splicing, and translation (reviewed in Kuzembayeva et al., 2014).

A classical retrovirus containing three canonical genes or open reading frames (ORFs): gag, pol and env is called a “simple” retrovirus (Fig. 1.2; Coffin, 1992; Murphy et al., 1994). For example, in Mason-Pfizer monkey virus (MPMV), the gag gene encodes for group specific antigen (Gag) precursor structural polyprotein which is cleaved by viral protease during maturation into the matrix (MA), capsid (CA) and nucleocapsid (NC) proteins (Fig. 1.3). The MA protein (forming the inner shell) is involved in mediating a stable association of Env glycoproteins to the viral particles, whereas NC interacts with the full length unspliced gRNA to selectively encapsidate itself into the assembling viral particles (reviewed in Kuzembayeva et al., 2014 and references within). The CA proteins form the viral core. The pol gene encodes for the RT and integrase (IN) enzymes. The RT has retroviral-specific reverse transcriptase activity (polymerase and RNase H) to convert the RNA genome into DNA which is then integrated into the host chromosome, a process that is facilitated by the IN enzyme. In addition, a small open reading frame, pro, encodes for viral protease, which carries out auto cleavage and processing of the viral polyproteins, and is located either at the 3’ end of the gag gene, the 5’ end of the pol ORF or as an ORF between gag and pol, depending on the virus type (Fig.1.3; reviewed in Pedersen & Duch, 2006). The CA, RNA + protein (RNP) and the viral enzymes (RT, RNase, and IN) form the inner core of the viral particle (Fig.1.3). The env gene encodes for the envelope (Env) precursor protein which is cleaved into the transmembrane (TM) and surface (SU) glycoproteins by cellular proteases for their incorporation into viral phospholipid bilayer membrane (Fig. 1.3). For
FIGURE 1.2. Schematic representation of simple and complex retroviral genomes. (A) Mason-Pfizer monkey virus (MPMV; simple), (B) human immunodeficiency virus type-1 (HIV-1; complex), and (C) mouse mammary tumor virus (MMTV; complex?) retroviruses. All retroviruses contain long terminal repeats (LTRs) and 3-4 canonical genes (*gag*, *pol* *pro*, and *env*). Apart from these, complex retroviruses such as HIV-1 contain 6 additional accessory genes, which play important roles in the completion of their life cycle. Complex retroviruses contain *cis*-acting sequences (CTE, constitutive transport element; RRE, rev responsive element, and RmRE-Rem-responsive element) that, after interacting with cellular or viral factors, facilitate the nuclear export of unspliced genomic RNA. Compared to HIV-1, MMTV contains only three accessory genes (*dut*, *sag* and *rem*) and therefore has recently been proposed to be re-classified as a murine complex retrovirus. *sag*, superantigen gene.
A
MPMV (simple retrovirus)

B
HIV-1 (complex retrovirus)

C
MMTV (complex retrovirus?)
FIGURE 1.3. Schematic representation of a typical retroviral particle depicting major gene products. The two copies of the viral RNA genome (red spirals) along with the packaging proteins form the ribonucleoprotein (RNP) inside the inner core (yellow). The core is present inside a protein matrix shell (inner shell; green) that is enclosed in a phospholipid bilayer membrane (blue) containing envelope. The envelope contains glycoprotein spikes made up of gp120 and gp41 in the case of HIV-1. The inner core also contains viral enzymes such as reverse transcriptase (RT) and integrase (In) which are responsible for reverse transcribing and integrating the viral genome into the host genome, resulting in the generation of the proviral DNA. Figure and legends adapted from website: http://cronodon.com/BioTech/Virus_Tech_3.html.
example, in human immunodeficiency virus type 1 (HIV-1), transmembrane glycoprotein gp41 is responsible for the fusion between viral and cell membranes (Allan et al., 1985; Kowalski et al., 1987; Veronese et al., 1985; Willey et al., 1988) and surface glycoprotein gp120 mediates the binding of HIV virions to CD4 receptor (McDougal et al., 1986). The glycoprotein spike formed by gp120 and gp41 enables the virus to attach to and fuse with target cells to initiate infection.

Complex retroviruses such as HIV-1 (Fig. 1.2) have been shown to code for additional accessory and regulatory proteins that are derived from multiply spliced messages (Fig. 1.4). In the case of HIV-1, apart from the canonical gag, pol and env genes, HIV-1 contains two regulatory (tat, rev) and four accessory (vif, vpr, vpu and nef) genes (reviewed in Balvay et al., 2007; Cullen & Greene, 1990; Pavlakis & Felber, 1990). These viral proteins in complex retroviruses play an important role in viral replication. For example, tat encodes for transcriptional activator (Tat) protein by binding to a structured region in the RNA called transactive responsive element (TAR) that is important for replication by mediating transactivation of transcription and regulation of gene expression (Dayton et al., 1986; Fisher et al., 1986; Garcia et al., 1989; Jakobovits et al., 1988). The rev gene encodes Rev protein (regulator of viral gene expression) which regulates nuclear export of viral RNA into the cytoplasm (Meyer & Malim, 1994). Accessory genes vif, vpr, vpu, and nef encode for Vif (viral infectivity factor), Vpr (viral protein R), Vpu (viral protein U) and Nef (negative effector) proteins which perform important functions during viral replication (Frankel & Young, 1998). For example, Vif is crucial for inhibiting the antiviral activity of a host factor, APOBEC3 (Mehle et al., 2004; Strebel et al., 1987). Nef and Vpu proteins are involved in the degradation of CD4 cells (Mangasarian & Trono,
FIGURE 1.4. Schematic representation of retroviral life cycle. (A) The virus binds to a specific receptor on the cell membrane and undergoes membrane fusion (at the cell surface or endosome). The viral core is then released which undergoes partial uncoating during which viral RNA genome (blue) is reverse transcribed by reverse transcriptase (RT; orange circle) into a double stranded DNA (blue). The reverse transcribed DNA is transported through the cytoplasm and imported into nucleus where it is integrated into the host chromosome (green) forming a provirus. (B) The integrated provirus is transcribed classically into two types of RNAs, full length (unspliced) genomic mRNA and envelope spliced mRNA, which are then exported to the cytoplasm. In the cytoplasm, full length (unspliced) genomic mRNA is translated into Gag/Pol proteins (shown in blue circles), whereas the envelope proteins are expressed from spliced mRNA. Furthermore, the gRNA also functions as the substrate for packaging/encapsidation into the assembling virion particles at the cell membrane (or into exocyte) to continue its life cycle. The virion particles undergo maturation after budding. Env, envelope protein; Gag, group-specific-antigen protein; Gag-Pol, Gag-polymerase; PIC, pre-integration complex; RTC, reverse transcription complex. Figure and legends adapted from Stoye (2012).
1997; Tiganos et al., 1997; Willey et al., 1992) during the infection process. Recently, due to the presence of a number of accessory genes (Fig. 1.2), it has been suggested that mouse mammary tumor virus (MMTV) be re-classified as a complex murine retrovirus (Mertz et al., 2005). For example, MMTV codes for the superantigen protein (Sag; Choi et al., 1991; Frankel et al., 1991; Woodland et al., 1991) which mediates the infection of mammary glands in mice. *Rem* gene expresses the rem protein which has been shown to interact with a structural element called Rem-responsive element (RmRE) on the MMTV gRNA to facilitate its nuclear export (Fig. 1.2; Indik et al., 2005; Mertz et al., 2005, 2009a, 2009b; Müllner et al., 2008).

### 1.1.4. Retroviral life cycle

Following entry into the host cell, various events that take place in the life cycle of a typical retrovirus leading to the production of progeny virions are explained in figure 1.4 (reviewed in Stoye, 2012). The life cycle of a retrovirus can be divided into entry and exit phases. The viral Env proteins interact and bind to a specific receptor on the host cell surface to mediate entry into the cell either at the plasma membrane or from endosomes *via* membrane fusion. The viral core is then released into the cytoplasm which then undergoes partial uncoating. Reverse transcription of the viral RNA genome into DNA by RT takes place in the cytoplasm and the pre integration complex (PIC) is imported into the nucleus (nuclear entry), resulting in integration into cellular DNA - a process that is mediated by virally encoded IN. The integrated retroviral DNA is now called a “provirus”. In the exit phase of the viral life cycle, cellular RNA polymerase II (RNAPII) transcribes viral RNA from the proviral DNA. In the case of simple retroviruses like MPMV, two classes of RNAs are generated: full length unspliced
gRNA and spliced envelope RNA (Fig. 1.5). These RNAs are exported from the nucleus to the cytoplasm for translation into viral proteins (Fig. 1.4). In sharp contrast to cellular unspliced RNAs (which are not exported out of the nucleus), retroviruses have evolved several different mechanisms involving both viral and cellular factors for ensuring that sufficient levels of unspliced gRNA are transported to the cytoplasm (Fig. 1.6).

In the cytoplasm, while the spliced RNA is responsible for expressing envelope precursor proteins, the unspliced gRNA plays a dual function: i) It acts as a template for both the translation of Gag/Pol polyproteins and ii) works as a substrate for packaging into the assembling virus particles to continue viral life cycle (Fig. 1.4). In many retroviruses, ribosomal frameshifting plays an important role in expressing Gag/Pol polyproteins. Ribosomal frameshifting is a process in which ribosomes move to a different reading frame and continue translation in that reading frame. The ribosomal frameshift signal (RFS) that makes a ribosome shift comprises two elements: a “slippery sequence” where the ribosome switches the reading frame, and an adjacent stimulatory signal, usually a specific RNA secondary structure called “pseudoknot” (Fig. 1.7) downstream of the RFS (reviewed in Brierley & Dos Ramos, 2006). The slippery sequence has been shown to be shifty on its own in vitro, up to 2%, but is strongly stimulated, up to 40-fold, by the presence of a hairpin pseudoknot located downstream of the slippery sequence (Fig. 1.7). This mRNA secondary structure causes ribosomes translating the message to pause at the position of the slippery sequence, slowing or pausing the ribosome during translation and increasing the frequency at which frame shifting occurs. This boosts the relative amounts of proteins encoded by the downstream reading frames (Fig. 1.7). The NC domain of the Gag precursor
FIGURE 1.5. Schematic representation of MPMV and MMTV genomes and their respective RNA transcripts. Simple retroviruses like MPMV transcribe a full length (unspliced) gRNA which also translates into Gag/Pol polyprotein, and a spliced mRNA which codes for envelope protein. MMTV genome is transcribed into *gag/pol, env, sag,* and *rem* mRNAs by different splicing mechanisms. *Sag* gene uses either a promoter in the 5’ LTR or the *env* promoter for its transcription. *sag,* superantigen gene. MMTV part of the figure is adapted from Mertz et al. (2005).
MPMV

MMTV

gag-pol mRNA

env mRNA

sag mRNA

sag mRNA from env promoter

rem mRNA
**FIGURE 1.6.** Schematic representation of different pathways employed by retroviruses during nucleocytoplasmic transport of unspliced gRNAs. Diverse retroviruses have evolved different mechanisms for successfully transporting unspliced gRNA for use as a template for translation of structural and enzymatic proteins as well as for packaging into the assembling virion particles.  

(A) Virally encoded *trans*-acting Rev protein (in the case of complex retrovirus HIV-1) binds to the *cis*-acting structural Rev responsive element (RRE) to facilitate the efficient transport of unspliced RNAs from the nucleus to the cytoplasm.  

(B) The *cis*-acting constitutive transport element (CTE) in the case of MPMV (a simple retrovirus) interacts with cellular proteins to mediate the nuclear export of the gRNA.  

(C) The *cis*-acting Rem-responsive element (RmRE) at the 3’ of the MMTV gRNA interacts with the virally encoded *trans* proteins, Rem, facilitating the nuclear export of the gRNA.  

Figure and legends adapted from Jaballah (2010).
FIGURE 1.7. Schematic representation of ribosomal frameshifting phenomenon in retroviruses to express relative amounts of Gag and Pol proteins. Retroviral gRNAs contain a slippery sequence that allows 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., *Molecular Biology of the Cell*. 4th Edition, 2002.
NO FRAMESHIFT (90% of ribosomes)

FRAMESHIFT (10% of ribosomes)
protein interacts with the RNA dimer and packages it into budding virions at the cell membrane or exocyte (as in the case of MMTV). Retroviral particles assembly can occur at any of two subcellular locations based on the retroviral classification. For example, type C retroviruses (e.g., MLV) assemble at the cell membrane, while types B and D retroviruses (e.g., MMTV and MPMV) preassemble at a perinuclear location forming the so called type “A” particle that are subsequently translocated to the cell membrane for budding (reviewed in Goff, 2001). These budding particles acquire the Env glycoproteins at this step since they are incorporated into the lipid bilayer of the host earlier after synthesis. The virion is then released from the cell surface and undergoes maturation to become infectious by further proteolytic cleavages and genomic RNA dimer stabilization (Bender et al., 1978; Fu & Rein, 1993; Murti et al., 1981).

Understanding the various steps involved in the viral life cycle has also helped us to learn about a number of cellularly-encoded genes that have been shown to inhibit retrovirus infection by blocking the virus life cycle at different stages (Fig. 1.8; reviewed in Goff, 2004; Pedersen & Duch, 2006; Santa-Marta et al., 2013). These anti-retroviral defense mechanisms have been observed in virally-infected producer and/or target cell of the virus. Cytidine deamination of viral single stranded DNA during reverse transcription can hinder retrovirus replication by the APOBEC family of proteins (Apolipoprotein B-editing Catalytic Polypeptide 3 Proteins; APOBEC3; A3; Fig. 1.8). These proteins have been shown to be packaged by HIV-1 and MPMV in the virus particles (Yu et al., 2004; Zhang et al., 2003). It has also been reported that retroviruses may develop resistance to such anti-retroviral host defense mechanisms. For example, HIV-1Vif protein can account for the resistance to APOBEC due to possible
FIGURE 1.8. Role of host restriction factors and their action during retroviral (e.g. HIV-1) replication. (A) Producer cell infected with virus. (B) Target cell for infection. The symbols and the associated text represent the point of restrictions during the retroviral life cycle and the name of the responsible host gene or gene product. Red ovals represent cellular restriction factors whereas gray hexagons represent viral factors to overcome the host factors. The course of viral replication and actions are shown by the black arrows, while the course of inhibition mechanism is shown by broken arrows. Unknown mechanisms are represented by question marks (?). Figure and legends adapted from Santa-Marta et al. (2013).
interactions between the two proteins (Yu et al., 2003). Some of the other
intracellular restriction factors include TRIM and SAMHD1. Several TRIM
family members have been identified as HIV-1 restriction factor acting at different
steps in the HIV-1 life cycle (reviewed in Santa-Marta et al., 2013). The TRIM5
proteins have been shown to block HIV-1 infection by binding to the HIV-1 CA
and inducing its premature disassembly before reverse transcription can occur.
Vectors that combine anti-CCR5 chemokine receptors short hairpin RNA with
TRIM5a therapeutic targets are being tested to maximize HIV-1 replication
blockage. The sterile alpha motif (SAM) and histidine-aspartate (HD) domain-
containing protein 1 (SAMHD1; reviewed in Santa-Marta et al., 2013) have been
shown to reduce the intracellular dNTP pool below levels that support HIV-1
reverse transcription thus blocking viral replication.

Host microRNAs have also been shown to be responsible to restrict
viruses such as spumavirus, primate foamy virus type 1 (PFV-1), as part of the
RNA interference machinery in human cells (Lecellier et al., 2005). The complete
mechanism(s) of some of these factors involved in restricting the retrovirus life
cycle is yet to be fully understood. For example, in a number of retroviruses such
as HIV-1, SIV and MuLV, a ZAP protein in addition to TRIM5α have been
observed to block infection (Gao et al., 2002; Yap et al., 2004; Ylinen et al.,
2005). These restriction systems provide understandings that can be utilized into
the design of novel anti-retrovirals and unravel the yet to be identified steps of the
virus life cycle (reviewed in Goff, 2004).

1.1.5. Retroviral gRNA packaging

For a retrovirus to continue its life cycle and spread infection, it requires
packaging/encapsidation of its gRNA into the assembling viral proteins before it
leaves a host cell by budding (Fig. 1.4). Therefore, the efficient and specific packaging of retroviral gRNA by the assembling virion particles is considered as one of the hallmarks of retroviral life cycle (Fig. 1.4). During this process, full-length, unspliced gRNA is preferentially packaged, whereas spliced viral and cellular RNAs are generally excluded from nascent virus particles despite the fact that retroviral genomic RNAs constitute 1% or less of the total cellular mRNAs (Coffin et al., 1997). The specificity towards RNA packaging is conferred by the recognition of specific cis-acting sequences, the packaging signal (Ψ), present at the 5’ end of the viral genome, which interacts with the NC domain of the Gag precursor protein (Fig. 1.9; reviewed in D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007; Lu et al., 2011a; Miyazaki et al., 2011).

Specifically, NC contains zinc knuckles with one or two conserved CCHC arrays (C-X2-C-X4-H-X4-C; C = Cys, H = His, X = variable amino acid) that are required for high affinity NC-RNA interactions in vitro (Fig. 1.10; Green and Berg, 1989; Henderson et al., 1981; Jouvenet et al., 2011; Summers et al., 1992). The interaction between NC and the packaging signal results in the incorporation of the gRNAs into the virus particles. Furthermore, it has been shown that packaging can occur when the NC domains and the packaging signals are incorporated into heterologous retroviral Gag and RNA constructs, respectively. For example, chimeric HIV-1 virions that contain the Moloney murine leukemia virus (MoMuLV) NC domain preferentially package MoMuLV RNA and vice versa (Berkowitz et al., 1995; Zhang & Barklis, 1995). The sites of initial interaction between gRNA and NC domain can occur in the cytoplasm, resulting in the formation of subassemblies with Gag (Fig. 1.11D, top) or they can directly associate at the plasma membrane (Fig. 1.11D, bottom) after gRNA is actively
FIGURE 1.9. Schematic representation of the prevailing model of retroviral gRNA packaging/encapsidation. (A) Mason-Pfizer monkey virus (MPMV) genome showing the location of packaging sequences (ψ) at the 5' end of the genome which has been shown to assume higher order structure and is depicted as a cartoon. (B) During retroviral life cycle, full-length, unspliced genomic RNA by virtue of the presence of the ψ is preferentially packaged, whereas spliced viral RNA (in which ψ is excised out during splicing) and cellular RNAs are generally excluded from nascent viral particles.
Specific encapsidation of retroviral RNA occurs **ONLY** if it contains a packaging signal $\psi$.

Specific encapsidation of RNA **CANNOT** take place if there is **NO** packaging signal $\psi$. 

---

A  MPMV  
U3  R  U5  
PBS  
$\psi$  
gag  
pol  
env  
U3  R  U5  
CTE

B  Unspliced RNA  
Spliced RNA  
SD  
SA
FIGURE 1.10. Amino acid sequences of the nucleocapsid (NC) proteins of (A) murine leukemia virus (MLV) and (B) human immunodeficiency virus type-1 (HIV-1). The C-terminus of the Gag polyprotein contains the NC domain that selectively recruits gRNA for packaging by binding to the packaging sequences (ψ) at the 5’ end of gRNA. The zinc fingers (ZFs) with the C-C-H-C consensus sequence (in red) are represented with the central zinc atom (Zn). MLV contain one ZF domain, while HIV-1 contains two. The ZF enables the NC domain to bind and anneal to nucleic acid. Figure and legends adapted from Jouvenet et al. (2011).
A
Murine leukemia virus (MLV)

B
Human immunodeficiency virus type-1 (HIV-1)
FIGURE 1.11. A model depicting possible mechanism(s) employed during retroviral RNA dimerization and packaging in a cell co-infected with C-type retroviruses such as human immunodeficiency virus type 1 (HIV-1) and murine leukemia virus (MLV; shown in red at top and green at the bottom).  

(A) Retroviral particles containing two plus sense gRNAs joined by dimer linkage within the capsid core infect the host cell. The gRNAs are condensed in the core and bound by nucleocapsid (NC; shown as green dots).  

(B) Two copies of gRNA interact (kissing interactions) via stem loops containing palindromic sequences, resulting in initiation of dimerization and changes in structural conformation due to base pairing between the two RNA strands. This further result in the formation of a dimer linkage structure that has exposed NC binding sites (shown in yellow) which were hidden (protected) in a monomeric gRNA structure. Such a conformational change facilitates the binding of NC domain of Gag polyproteins to the NC binding site, allowing the recruitment of gRNA for packaging.  

(C) The cellular site of dimerization initiation is different for different retroviruses. For example, MLV gRNA dimers first interact near the transcription sites in the nucleus which can result in favoring of homodimer formation over heterodimer formation. HIV-1 gRNAs have been shown to undergo kissing interaction in the cytoplasm which can result in homodimers and/or heterodimers.  

(D) The site of gRNA recruitment by Gag can be found in the cytoplasm (shown at top) or at the plasma membrane (shown at bottom) after separate active transport.  

(E) The viral particle buds out after retroviral gRNA packaging. Packaging of two gRNAs as homodimer helps the virus in maintaining packaging specificity and genomic integrity. On the other hand, packaging of a heterodimer (virion at the center) promotes genetic recombination and genetic diversity. Figure and legends adapted from Johnson and Telesnitsky (2010).
transported separately from all or most Gag (reviewed in Johnson & Telesnitsky, 2010; Jouvenet et al., 2011).

The sequences that are responsible for gRNA packaging in retroviruses have been shown to be present in the form of a continuous region or could be in the form of a discontinuous region (multipartite sequence) residing at the 5’ end of the genome (Fig. 1.12; reviewed in Jaballah et al., 2010; Mustafa et al., 2005, 2012). Over the years, we have learned that retroviral packaging sequences are usually located downstream of the major spliced donor (mSD). Such a unique position of the packaging sequences facilitates the discrimination between spliced and unspliced viral RNAs in most of the retroviruses with the exception of Rous sarcoma and avian leukosis viruses. A classic example of such discrimination between spliced and unspliced viral transcripts is observed in the case of HIV-1 during RNA packaging. The stem loop 2 (SL2), a major packaging determinant containing the major splice donor (mSD) has been shown to binds with high affinity to NC. However, because of the presence of mSD during splicing, the SL2 undergoes conformational changes resulting in the destabilization and absence of SL2 in the spliced RNA thereby limiting its packaging potential (reviewed in D’Souza & Summers, 2005). On the other hand, in the case of HIV-2, since the major packaging determinants are present upstream of the mSD, therefore both spliced and unspliced transcripts contain packaging sequences making the discrimination between spliced and unspliced transcripts difficult during packaging. HIV-2 has overcome this disparity by using a unique sorting mechanism during which only those transcripts capable of translating gag in cis can be packaged into the virus particle (Balvay et al., 2007; Griffin et al., 2001; Kay & Lever 1999). In HIV-1, in vivo and in vitro studies have shown that
FIGURE 1.12. Schematic representation of the packaging determinants of different retroviruses (left panel), which have been universally shown to assume higher order structure schematically represented in the right panel. 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 are compared as per published data. The references and other details of the gRNA packaging determinants are provided in the table below. Figure and legends adapted from Mustafa et al. (2012).
<table>
<thead>
<tr>
<th>Virus</th>
<th>Length of 5' UTR</th>
<th>5' UTR Required</th>
<th>Gag Required</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPMV (Type D)</td>
<td>128 bp</td>
<td>First 50 bp + last 23 bp</td>
<td>~ 100 bp</td>
<td>Schmidt et al., 2003; Jaballah et al., 2010</td>
</tr>
<tr>
<td>FIV (Type C)</td>
<td>270 bp</td>
<td>First 150 bp</td>
<td>~ 100 bp</td>
<td>Browning et al., 2003a, 2003b; Mustafa et al., 2005a; Kenyon et al., 2008; Rizvi et al., 2010; Kenyon et al., 2011</td>
</tr>
<tr>
<td>MMTV (Type B)</td>
<td>160 bp</td>
<td>Entire UTR</td>
<td>~120 bp</td>
<td>Rizvi et al., 2009</td>
</tr>
<tr>
<td>HIV-1 (Type C)</td>
<td>152 bp</td>
<td>Entire UTR</td>
<td>?</td>
<td>Berkowitz et al., 1995; McBride &amp; Panganiban, 1997; Clever &amp; Parslow, 1997; Houzet et al., 2007; Heng et al., 2012</td>
</tr>
<tr>
<td>SIV (Type C)</td>
<td>231 bp</td>
<td>Entire UTR</td>
<td>?</td>
<td>Rizvi &amp; Panganiban, 1993; Guan et al., 2000, 2001; Patel et al., 2003; Strappe et al., 2003; Whitney &amp; Wainberg, 2006</td>
</tr>
</tbody>
</table>
translation and packaging takes place in competition, driving the gRNA to translate proteins following which the transcripts undergo changes in their structural conformation, facilitating packaging (Anderson & Lever, 2006; Butsch & Boris-Lawrie, 2000; Dorman & Lever, 2000).

Although RNA packaging is a ubiquitous process in all retroviruses and the packaging signal always involves sequences at the 5’ end of the retroviral gRNA, no sequence conservation between the packaging signals of different retroviruses has been found. Invariably, it has been shown that the packaging sequences of all known retroviruses assume a higher order structure comprising different structural motifs (Fig. 1.13; D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007; Lu et al., 2011a; Miyazaki et al., 2011;). Therefore, the involvement of such structural RNA motifs on the retroviral genomic transcripts (at times regardless of the primary sequence) have been strongly associated with retroviral RNA encapsidation and could explain the phenomenon of RNA cross-packaging among diverse retroviruses (Al Dhaheri et al., 2009; Al Shamsi et al., 2011; Moore et al., 2007; Parveen et al., 2004; Rizvi & Panganiban, 1993; White et al., 1999; Yin & Hu, 1997). This has further been substantiated by the fact that heterodimers involving RNAs from two divergent retroviruses can also be co-packaged (Moore et al., 2009; Motomura et al., 2008; reviewed in Johnson & Telenitsky, 2010). RNA secondary structures predicted by various approaches, including free energy predictions, phylogenetic analyses, and biochemical probing, and their subsequent validation by biological assays, have further confirmed the function of RNA structural elements in the gRNA packaging for a number of both simple and complex retroviruses (Fig. 1.13; Jouvenet et al., 2011). In addition to the conventional packaging sequences, a
**FIGURE 1.13.** Schematic representation of MLV (A) and HIV-1 (B) genomes and the secondary structure of their packaging signals (Psi; ψ). Interaction between nucleocapsid (NC) domain of Gag and the highly structured packaging determinants located in the 5’ untranslated region (5’ UTR) of gRNA results in its specific selection from a cytoplasmic pool that contains a substantial excess of non-viral and spliced viral RNAs. The higher order structure of the MLV packaging signal contains four stem loops (SLs A-D). The A and B SLs promote gRNA dimerization, while SLs C and D form the core encapsidation signal. In the case of HIV-1, DIS containing SL1 and SL3 form the major packaging determinant. SD, major splice-donor site. Figure and legends adapted from Jouvenet et al. (2011).
A
Murine leukemia virus (MLV)

B
Human immunodeficiency virus type 1 (HIV-1)
recent study has shown that HIV-1 gRNA encapsidation is also dependent on a
*cis*-acting RNA element, genomic RNA packaging enhancer (GRPE) located
within the *gag* open reading frame overlapping the Gag/Pol RFS (Fig. 1.7;
Chamanian et al., 2013). However, further experiments are warranted to establish
the precise role of GRPE in HIV-1 gRNA packaging since another recent study
has suggested that the deletion of the wild-type RFS does not impede the
packaging of the full length HIV-1 RNA (Nikolaitchik & Hu, 2014).

1.1.6. Retroviral gRNA dimerization

Another hallmark of retroviral life cycle is the packaging of two copies of
viral gRNA in the form of a non-covalently linked RNA dimer. The conservation
of this unique genomic structure among retroviruses strongly advocates that a
dimerized genome plays a critical role in the viral life cycle (reviewed in D'Souza
& Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007). In retroviral
replication, the process of RNA dimerization is thought to be closely linked with
RNA packaging, and in some, the former depends on the latter (Hibbert et al.,
2004, reviewed in D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010;
Lever, 2007; Miyazaki et al., 2011; Moore et al., 2009; Paillart et al., 2004).
Therefore, it is not surprising that for almost all retroviruses without any sequence
conservation, determinants of gRNA packaging and dimerization map to the same
~100 to 400 nucleotides (nt) at the 5’ end of the gRNA (Fig. 1.12; reviewed in
D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007). These
gRNA packaging and dimerization sequences are, for the most part, physically
and genetically inseparable and their structural context, almost invariably,
contributes to these processes (D'Souza & Summers, 2005; Johnson &
Telesnitsky, 2010; Lever, 2007). Consistent with this, studies have shown that
mutations affecting dimerization generally also affect packaging (Baig et al., 2007; Berkhout, 1996; Lanchy et al., 2003; Lanchy & Lodmell, 2007; Laughrea et al., 1997; McBride & Panganiban, 1997; Paillart et al., 1996). The fact that packaged RNA remains dimeric, even in cases when RNA packaging was reduced, suggests that the dimerization of the two genomic RNAs may be a prerequisite for packaging (Hibbert et al., 2004; Housset et al., 1993; Levin et al., 1974). Furthermore, single HIV-1 viral RNA molecules containing two copies of 5’ UTR (containing packaging/dimerization sequences) are packaged as monomers, highlighting the importance of the interaction between the 5’ ends of two RNAs as a precondition for RNA packaging (Sakuragi et al., 2001, 2002). In addition, high affinity NC binding residues in the gRNA of MoMuLV, for example, are sequestered between the stem loops of the monomeric structure, and are exposed by the formation of an intermolecular duplex possibly promoted by the chaperone activity of the NC domain of Gag (Darlix et al., 1990, 1995).

Figure 1.11 represents the prevailing model of dimerization-dependent packaging of gRNAs.

Palindromic (pal) and semi-palindromic sequences forming helix loop structural motifs have been strongly implicated in the initiation of the dimerization process between two copies of retroviral gRNAs (reviewed in Greatorex, 2004; Johnson & Telesnitsky, 2010; Moore & Hu, 2009). It has been shown that the dimerization is initiated by a pal sequence in its structural context (pal helix-loop) which is called the dimerization initiation site (DIS; reviewed in D’Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lu et al., 2011a; Lever, 2007; Moore & Hu, 2009; Paillart et al., 1996b, 2004; Russell et al., 2004;). The pal of different retroviruses has been shown to contain a canonical “GC” dyad (reviewed
in Hussein et al., 2010). The autocomplementary property of palindromic sequences allows base pairing between two gRNAs in an antiparallel direction to initiate “kissing loop” interactions (Clever et al., 1996; Haddrick et al., 1996; Lodmell et al., 2000, 2001; Paillart et al., 1994, 1996a, 1997) that ultimately leads to a stable dimer linkage formation exposing the otherwise hidden NC binding sites on the gRNA for interaction with the NC domain of Gag polyprotein (Fig. 1.11; reviewed in Johnson & Telesnitsky, 2010). The site of gRNA dimerization initiation has been shown to be different for different retroviruses (Fig. 1.9C; reviewed in Johnson & Telesnitsky, 2010; Jouvenet et al., 2011). For example, RSV and MLV gRNA dimers are formed in the nucleus near transcription sites which support homodimer formation, whereas in the case of FIV and HIV-1, gRNA dimer formation takes place in the cytoplasm resulting in generation of either homodimers and/or heterodimers (Fig. 1.11C; reviewed in Johnson & Telesnitsky, 2010; Jouvenet et al., 2011).

Employing a combination of genetic and structural prediction analyses, we have recently shown that MPMV and MMTV packaging determinants needed for optimal packaging and propagation of their gRNAs are spread out between R and the first 120 bp of the gag (Fig. 1.12; reviewed in Jaballah et al., 2010; Mustafa et al., 2012). To expand these studies further, MPMV and MMTV gRNA packaging sequences were folded using minimum free energy algorithm programs like Mfold (Mathews et al., 1999; Zuker, 2003). The folding predictions of these sequences revealed a higher order structure comprising several structural motifs, which could be involved in augmenting gRNA dimerization and packaging of these viruses (Jaballah et al., 2010). Therefore, in this thesis, we have developed structure-function relationship of various structural motifs during gRNA dimerization and
packaging processes of these viruses. Delineating the underlying molecular mechanisms in retroviral RNA packaging is important for enhanced understanding of retroviral life cycle.

1.2. Objectives

The advent of recombinant DNA technology has made possible the use of gene therapy to treat, prevent, and even control human diseases such as cardiovascular and pulmonary disorders, diabetes and different kinds of cancers (Herzog et al., 2010; Hu et al., 2011; Ly et al., 2007, 2008; Song et al., 2004; Tian et al., 2004). Until 2012, 1155 clinical trials had been performed for various cancers, accounting for approximately 65% of all human gene therapy trials (Wiley, 2012; Woo et al., 2006). Retroviruses are distinct since they carry the special RT enzyme that converts the RNA genome into a DNA molecule, which is then integrated into the host genome, making it “permanent” (reviewed in Telesnitsky & Goff, 1997). Therefore, modified retroviruses known as “retroviral vectors” are ideal candidates for delivering the therapeutic genes in human gene therapy trials due to their ability to: i) integrate into the host genome, ensuring its stable long-term expression, ii) have large coding capacities, thereby allowing the expression of most genes of interest, and iii) transduce only the therapeutic gene and not the viral genes, thus avoiding host humoral/cellular responses to viral antigens that potentially can eliminate the transduced cells. Consequently, it is not surprising that of all gene therapy trials for human diseases; more than 20% have used retroviral based vectors (Hu et al., 2011; Wiley, 2012).

Among retroviruses, currently HIV-1 based vectors are being exploited; however, their use in humans raises critical safety concerns such as the generation of replication competent virus through recombination with human endogenous
retroviruses. Therefore, vectors from phylogenetically distant non-human retroviruses such as MPMV and MMTV may circumvent such safety concerns. The use of MPMV-based vectors are also being considered ideal since the therapeutic genes to be used may require nuclear export signals like MPMV constitutive transport element (CTE; Bray et al., 1994) for their efficient expression in the target cells. Along the same lines, MMTV has distinguished itself from most other retroviruses in that it contains multiple promoters and its promoters in the LTR are inducible by steroid hormones, making them not only “switchable” but also “tissue-specific” since they would turn on (transcribe RNA) only in cells/tissues with receptors for steroid hormones (reviewed in Ham et al., 1988; Klein et al., 2008; Mustafa et al., 2000). This aspect alone (“targeted” and “inducible” gene expression) has prompted the use of MMTV promoters for “conditional” gene expression in many in vitro systems, especially of those genes whose constitutive expression could be toxic to cells. These unique features have led to increasing interest in developing MPMV and MMTV vectors for human gene therapy. However, before MPMV and MMTV based vectors can be used for human gene therapy, it is crucial that the molecular mechanisms of the pertinent aspects of their life cycle such as dimerization and packaging are fully delineated.

Very little is known about the molecular intricacies involved in the recognition of MPMV and MMTV gRNA that facilitate their dimerization and packaging into the virus particle. The overall goal of this thesis project was to gain a better understanding of MPMV and MMTV gRNA dimerization and packaging mechanisms by establishing structure-function relationship(s) of the sequences involved in these processes. Specifically, this thesis project was undertaken to validate the higher order features of the MPMV and MMTV
packaging signal RNAs required in cis during gRNA dimerization and packaging, employing a newly developed SHAPE (selective 2' hydroxyl acylation analyzed by primer extension) methodology. After having validated the RNA secondary structure by SHAPE, in silico phylogenetic analyses, in vitro RNA dimerization and biologically relevant in vivo RNA packaging and propagation assay were employed to determine the role of pal sequences as potential DIS. To accomplish these, both MPMV and MMTV were studied at the molecular and structural levels with the results of MPMV presented in Chapter 2 and those of MMTV presented in Chapter 3.

The specific objectives that were addressed during the course of the study for both viruses were as follows:

A. To validate the predicted higher order features of the 5’ end of the MPMV and MMTV sequences required for gRNA packaging by SHAPE.

B. To identify gRNA dimerization initiation site (DIS) and establish structure-function relationship during gRNA dimerization and packaging.

C. To perform phylogenetic analyses to determine the level of conservation of different structural motifs in MPMV and MMTV packaging signal RNA.
CHAPTER 2

SHAPE Analysis of the 5’ end of the Mason-Pfizer Monkey Virus (MPMV) genomic RNA Reveals Structural Elements Required for Genome Dimerization

This chapter is a modified version of the manuscript published in the RNA Journal in December 2013. (Aktar, S. J., et al., RNA, 19:1648-1658, 2013).
2.1. Abstract

Earlier genetic and structural prediction analyses revealed that the packaging determinants of Mason Pfizer Monkey Virus (MPMV) include two discontinuous core regions at the 5’ end of its genomic RNA. RNA secondary structure predictions suggested that these packaging determinants fold into several stem loops (SLs). To experimentally validate this structural model, we employed selective 2’ hydroxyl acylation analyzed by primer extension (SHAPE), which examines the flexibility of the RNA backbone at each nucleotide position. Our SHAPE data validated several predicted structural motifs, including U5/Gag long-range interactions (LRIs), a stretch of single stranded purine (ssPurine) rich region, and a distinctive G-C rich palindromic (pal) SL. Minimum free-energy structure predictions, phylogenetic, and in silico modeling analyses of different MPMV strains revealed that the U5 and Gag sequences involved in the LRIs differ minimally within strains and maintain a very high degree of complementarity. Since the pal SL forms a helix loop containing a canonical “GC” dyad, it may act as a RNA dimerization initiation site (DIS), enabling the virus to package two copies of its genome. Analyses of wild type and pal mutant RNAs revealed that disruption of pal sequence strongly affected RNA dimerization. However, when in vitro transcribed trans complementary pal mutants were incubated together showed RNA dimerization was restored authenticating that the pal loop (5’ CGGCCG 3’) functions as DIS.
2.2. Introduction

The “diploid” genome of retroviruses consists of two (usually indistinguishable) strands of RNA that are non-covalently linked as a “dimer” close to their 5’ end (reviewed in D’Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007; Lu et al., 2011a; Moore & Hu, 2009; Paillart et al., 1996b; Paillart et al., 2004). The ubiquitous presence of such a unique dimeric genome among retroviruses suggests that dimerization plays a crucial role in the packaging of the genomic RNA (gRNA), as well as in the subsequent steps of the viral life-cycle (reviewed in Moore & Hu, 2009; Paillart et al., 2004; Russell et al., 2004).

Over the years, we have learnt that for almost all retroviruses, determinants of gRNA dimerization and packaging map to ~100 to 400 nucleotides (nt) at the 5’ end of the gRNAs, which have been shown to assume stable secondary structures (reviewed in Johnson & Telesnitsky, 2010; Paillart et al., 1996b). The gRNA dimerization and packaging sequences are, for the most part, physically and genetically indistinguishable, and consistently, it has been suggested that the retroviral gRNA packaging and dimerization processes are interlinked (reviewed in D’Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007; Lu et al., 2011a; Paillart et al., 2004).

It is becoming increasingly clear that palindromic (pal) sequences play a crucial role in gRNA dimerization: for human immunodeficiency type 1 and type 2 viruses (HIV-1 and HIV-2), feline immunodeficiency virus (FIV), Moloney murine leukemia virus (MoMuLV) and murine sarcoma gamma retroviruses (MuSV). it has been shown that the dimerization process is initiated by a pal sequence in its structural context (pal stem-loop) which is called the dimerization
initiation site (DIS; reviewed in D’Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007; Lu et al., 2011a; Moore & Hu, 2009; Paillart et al., 1996b, 2004; Russell et al., 2004). The pal of different retroviruses has been shown to contain a canonical “GC” dyad (reviewed in Hussein et al., 2010). The pal in the DIS loop of one gRNA has been shown to interact with the complementary sequence in the DIS loop on the second gRNA, resulting in the “kissing loop” interaction (Clever et al., 1996; Haddrick et al., 1996; Lodmell et al., 2000, 2001; Paillart et al., 1994, 1996a, 1997). Due to its central role in gRNA dimerization process that has been shown to be important for gRNA packaging and the continuity of viral life cycle, DIS has been an attractive target for antiretroviral drugs specifically aminoglycosides (Ennifar et al., 2003, 2006). The formation of “kissing loop” complex is likely to be further stabilized by the long-range interactions (LRIs) between the 5’ end (R/U5) and 3’ end (Gag) sequences of the retroviral packaging signal RNA (reviewed in Paillart et al., 1996b). In addition, it has also been proposed that in some retroviruses “kissing loop” interactions and dimer linkage maturation process lead to conformational changes exposing single-stranded nucleocapsid (NC) binding motifs, facilitating Gag binding during recruitment of the gRNA for packaging (reviewed in D’Souza & Summers, 2005; Johnson & Telesnitsky, 2010). Along the same lines, a recent study has shown that in addition to the conventional dimerization/packaging sequences, HIV-1 gRNA packaging is enhanced by a cis-acting RNA element (the genomic RNA packaging enhancer or GRPE) overlapping the Gag/Pol ribosomal frameshift signal (Chamanian et al., 2013).

Despite the fact that gRNA dimerization and packaging processes are universally present in all retroviruses, there is no sequence conservation between
the dimerization and packaging determinants among different retroviruses. Over the years, it is becoming increasingly clear that retroviral gRNA dimerization and packaging are reliant on highly ordered RNA structural motifs of the 5’ end RNA region (reviewed in D’Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Lever, 2007; Lu et al., 2011a). Therefore, it is not surprising that a number of recent cross- and co-packaging studies have elucidated that the specificity of retroviral gRNA packaging can be manipulated by substituting dimerization and packaging sequences at the 5’ end of genome from genetically distinct retroviruses (Al Dhaheri et al., 2009; Al Shamsi et al., 2011). This has further been substantiated by the fact that heterodimers involving RNAs from two divergent retroviruses can also be packaged (Moore & Hu, 2009; Motomura et al., 2008; reviewed in Johnson & Telesnitsky 2010).

A stretch of 5’ end sequences of the prototype betaretrovirus Mason-Pfizer monkey virus (MPMV), which causes an immunodeficiency disease in newborn rhesus monkey, has been shown to be involved in RNA packaging (Guesdon et al., 2001; Harrison et al., 1995; Vile et al., 1992). Employing a combination of genetic and structural prediction analyses, we have recently shown that MPMV packaging determinants comprise two discontinuous core regions (“A” and “B”) within the 5’ untranslated region (UTR; Jaballah et al., 2010; Schmidt et al., 2003). In addition to these two regions in the UTR, stretch of ~120 nt of gag has also been shown to be needed for optimal MPMV RNA packaging (Jaballah et al., 2010; Schmidt et al., 2003). Based on these mutational studies and structural prediction analyses, a minimal free energy RNA secondary structure model has been proposed for the 5’ end (388 nt starting from R) of MPMV gRNA (Jaballah et al., 2010; Fig. 2.1A). This model suggested that the packaging determinants
FIGURE 2.1. Minimal free-energy and SHAPE-validated models of the MPMV packaging signal RNA. The region used for analysis by Mfold and SHAPE included sequences from R up to 120 nt of gag. (A) MPMV packaging signal RNA secondary structure predicted earlier (Jaballah et al., 2010) using Mfold (Mathews et al., 1999; Zuker, 2003). Sequences in orange, green, red, and blue represent the primer binding site (PBS), regions “A” and “B” (that have been shown to be important in gRNA packaging), and pal sequences, respectively. Boxed areas in purple show the predicted LRIs between U5 and Gag. (B) SHAPE-constrained RNAstructure (Reuter & Mathews, 2010) model of MPMV packaging signal. Nucleotides are color annotated as per the SHAPE reactivities key. SD, splice donor.
**Predicted Secondary Structure of the MPMV Packaging Signal RNA**

- Region A
- Region B
- Gag SL1
- Gag SL2
- PBS SL
- U5/Gag LRI-I
- ssPurine-rich region

**SHAPE Validated Secondary Structure of the MPMV Packaging Signal RNA**

- SHAPE reactivity
- U5/Gag LRI-I
- U5/Gag LRI-II
- Partial Repeat of ssPurine-rich region

\[ \Delta G = -124.70 \text{ kcal/mol} \]
fold into several stable stem loop (SL) structures and also revealed a 14 nt GC-rich pal sequence (5’ UCGCCGCGCCGGCGA 3’) that folds into a hairpin with a canonical “GC” dyad in the loop, a 16 nt stretch of single stranded purine (ssPurine) rich region (8 nt of this ssPurine-rich region is duplicated as a base-paired sequence in an adjacent region), and LRIs between U5 and Gag sequences (Jaballah et al., 2010). The 14 nt pal has been proposed to play a role in the initiation of gRNA dimerization (Jaballah et al., 2010). The ssPurine-rich region (or its repeat in region “B” when predicted to refold as the ssPurine-rich region) has been shown to be essential for RNA packaging, possibly functioning as a potential nucleocapsid (NC) binding site. Two long range interactions between U5 and Gag sequences could potentially play a role in MPMV RNA packaging by maintaining the overall RNA secondary structure as has recently been shown in the case of FIV (Kenyon et al., 2008, 2011; Rizvi et al., 2010). Even though the deletion analysis of 5’ end of MPMV gRNA provides a plausible explanation for the discontinuous nature of MPMV packaging signals, the existence of the different structural motifs of the predicted RNA secondary structure of this region has not been validated experimentally.
2.3. Materials and Methods

2.3.1. MPMV RNA secondary structure analyses in silico

MPMV RNA (388 nt and 550 nt) secondary structure was predicted using the Mfold server (Mathews et al., 1999; Zuker, 2003) and validated by applying SHAPE reactivity data in the RNAstructure software (version 5.3, Reuter & Mathews, 2010). The predicted RNA secondary structures were redrawn with XRNA software. Sequences from different strains of MPMV corresponding to packaging signal RNA were employed to predict the RNA secondary structures of the region using Mfold (Mathews et al., 1999; Zuker, 2003). A sequence alignment of the same region for these MPMV isolates was also generated using ClustalW to look for the conservation of sequences in different structural motifs. MPMV nucleotide positions refer to GenBank accession number M12349 (Sonigo et al., 1986).

2.3.2. Plasmids used for in vitro transcription (wild type and pal mutants)

In order to in vitro transcribe the wild type MPMV genomic sequences for SHAPE analysis, SJ2 subgenomic transfer vector (described previously in Jaballah et al., 2010) containing the cis-acting sequences needed for genome replication was used as a template for generating the clone RCR001 by PCR. Primers used for PCR amplification included sense (S) primer OTR1004 (5’ CCC AAG CTT AAT ACG ACT CAC TAT AGG G CCA CCA TTA AAT GAG ACT TGA TC 3’) containing the HindIII restriction site and the T7 RNA polymerase promoter sequence (HindIII site is shown in italics whereas the T7 promoter sequence is underlined) and the anti-sense (AS) primer OTR 1005 (5’ AAA CCC GGG TTC TTT CTT ATC TAT CAA TTC 3’) containing Xmal/Smal site (shown in italics). The resulting PCR product was cleaved at the HindIII and Xmal sites.
(artificially created on either ends of the MPMV sequence) and ligated to a PUC based cloning vector (pIC19R; Marsh et al., 1984), which was previously cleaved with the same restriction enzymes generating wild type plasmid pRCR001. pRCR001 contains nucleotides +1 to 550 (+1 corresponds to the first nucleotide of the R region of the genomic RNA) under the control of a T7 promoter for \textit{in vitro} transcription of the RNA.

To study the effects of the pal sequence (5’ TCGC CGGCCG GCGA 3’) on \textit{in vitro} dimerization, a series of mutations were introduced in the pal sequence. Briefly, these mutations were incorporated through splice overlap extension (SOE) PCR as described previously (Gibbs et al., 1994, Jaballah et al., 2010; Rizvi et al., 2010), using MPMV sub-genomic vector SJ2 as the template requiring two separate amplifications in round one PCR. In brief, both S and AS primers that were used in two separate reactions of round one PCRs were designed in a fashion that the resulting products from the two reactions have overlapping complementary sequences. The overlapping complementary sequences allowed the two products to anneal during the second round of amplification with the outer S and AS primers, resulting in a final product containing the desired deletion mutation. The mutations introduced included deletion of the complete 14 nt pal stem-loop sequence (5’ TCG CCG GCC GGCGA 3’) using S primer OTR 897 (5’ CAC TTT TAA ACG CGT CGT CTT CCT CAC G 3’) and AS primer OTR 898A (5’ CGC GTT TAA AAG TGA AAG TAA ACT CTC TTG 3’) resulting in RCR002 (Fig. 2.2A). Substitution of the 4 nt pal loop 5’ GGCC 3’ with GTAA to generate a stable tetra loop was created using S primer OTR 996 (5’ GTA AGG CGA TTA AAA GTG AAA GTA AAC TC 3’) and AS primer OTR 997 (5’ CAC TTT TAA TAG CCT TAC GGC GAA CGC GTC GTC 3’) generating mutant
**FIGURE 2.2.** Sequences in pal SL augments MPMV gRNA dimerization by functioning as DIS.  **(A)** Table showing the 6 nt wild type pal sequence (in red) and the different mutations introduced (in blue).  **(B)** Gel shift assays of the MPMV WT and pal mutants. The upper panel shows native 1% TBM (50 mM Tris base, 45 mM boric acid, 0.1mM MgCl₂) gel run at 4 °C and the lower panel shows semi-native 1% TB (50 mM Tris base, 45 mM boric acid) gel run at room temperature.  M: monomer lane or monomer conformer; D: dimer lane or dimer conformer for each sample.  **(C)** Relative dimerization of the MPMV wild type and pal mutants RNAs. Following the gel shift assays, dimerization abilities of the RNAs in each lane were calculated and the dimerization data was then represented as relative to the wild type dimerization.
### A

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Description of the Mutations</th>
<th>Sequence of the pal Mutants</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCR001</td>
<td>Wild type pal sequence</td>
<td>5’ TCGCCGGCGCGA 3’</td>
</tr>
<tr>
<td>RCR002</td>
<td>Deletion of 14 nt wild type pal sequence</td>
<td>5’ ------------------------3’</td>
</tr>
<tr>
<td>RCR003</td>
<td>Substitution of 4 nt of the central wild type pal sequence with stable tetra loop</td>
<td>5’ TCGCCGTAAGGCGA 3’</td>
</tr>
<tr>
<td>RCR004</td>
<td>Substitution of 6 nt of the wild type pal sequence with HIV-1 pal</td>
<td>5’ TCGC GGCNGC GCGA 3’</td>
</tr>
</tbody>
</table>

### B

![Image of gel electrophoresis](image)

### C

![Image of bar chart](image)

**Table of Relative Dimerization**

<table>
<thead>
<tr>
<th></th>
<th>RCR001</th>
<th>RCR002</th>
<th>RCR003</th>
<th>RCR004</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBM</td>
<td>1.00</td>
<td>0.24</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>TB</td>
<td>1.00</td>
<td>0.19</td>
<td>0.21</td>
<td>0.24</td>
</tr>
</tbody>
</table>
RCR003 (Figure 2.2A). A mutant containing substitution of the 6 nt pal with HIV-1 pal sequence (5’ GCGCGC 3’) was constructed using S primer OTR 998 (5’ GCG CGC GCG ATT AAA AGT GAA AGT AAA CTC 3’) and AS primer OTR 999 (5’ CAC TTT TAA TCG CGC GCG GAA CGC GTC GTC TTC CTC 3’) generating mutant RCR004 (Fig. 2.2A).

In addition two other mutants, RCR005 and RCR006 containing heterologous trans-complementary sequence with a central “GC” dyad were also created to determine if the palindromic nature of the sequence is necessary for in vitro dimerization (Fig. 2.4A). The primers used for RCR005 and RCR006 construction included OTR 1000 (S primer; 5’ ACG CAC GCG ATT AAA AGT GAA AGT AAA CTC 3’) and OTR 1001 (AS; 5’ CAC TTT TAA TCG CGT GCG TGC GAA CGC GTC GTC TTC CTC 3’), OTR 1002 (S primer; 5’ GTG CGT GCG ATT AAA AGT GAA AGT AAA CTC 3’), OTR 1002 (S primer; 5’ GTG CGT GCG ATT AAA AGT GAA AGT AAA CTC 3’) and OTR 1003 (AS primer; 5’ CAC TTT TAA TCG CAC GCA CGC GAA CGC GTC GTC TTC CTC 3’), respectively. For all these mutants, the OTR 1004 (S) and OTR 1005 (AS) were used as outer primers during the second round of amplification and the resulting amplified product containing the T7 promoter and the desired mutation were cleaved and cloned in pIC19R as described above for creating wild type clone RCR001. Both the wild type clone, as well as all pal mutant clones were confirmed by sequencing. PCRs were performed as described previously (Jaballah et al., 2010; Rizvi et al., 2010). Briefly, approximately 3 μg of template DNA was used for each reaction under the following amplification conditions: denaturation at 94°C for 2 minutes followed by 30 cycles of 94°C for 1 minute, annealing at 45-47°C depending on primers used for 90 seconds, extension at
72°C for 90 seconds. Final extension was performed for 10 minutes at 72°C followed by incubation at 4°C.

2.3.3. RNA preparation by in vitro transcription

DNA templates containing the T7 RNA promoter at the 5’ end of the MPMV sequences were linearized by digesting with the Smal restriction enzyme at the site present at the 3’ end of the MPMV sequence in the RCR clone series (RCR001-RCR006). In vitro transcription was performed using bacteriophage T7 RNA polymerase for 2 hours at 37 ºC using conditions described previously (Marquet et al., 1991). A small fraction of each sample was subjected to electrophoresis on an 8% acrylamide/8M urea gel to analyze the in vitro transcription reaction products. The rest of the sample was treated with DNase I for 30 minutes at 37 ºC, then extracted with phenol-chloroform and finally ethanol precipitated. The RNAs were purified by fast protein liquid chromatography (FPLC; Pharmacia) on a TSK 250 column (Biorad) in a buffer containing 200 mM sodium acetate (pH 6.5) and 1% (v/v) methanol. The desired fractions were concentrated in Amicon Ultra-4 10K centrifugal filter device (Millipore) and the RNA concentrations were determined following quantification through nanodrop (ThermoScientific). Finally, the RNA homogeneity was checked by electrophoresis on an 8% acrylamide gel.

2.3.4. SHAPE methodology

To validate the predicted secondary structure of MPMV packaging signal RNA sequence (Jaballah et al., 2010) the purified in vitro transcribed RNAs were subjected to selective 2’ hydroxyl acylation analyzed by primer extension (SHAPE; Mortimer and Weeks, 2009) using benzoyl cyanide (BzCN). Briefly, one picomole of RNA in 8 μL milli-Q water (Millipore) was denatured for 2
minutes at 90 °C and then chilled for 2 minutes on ice, following which 2 μL 5X dimer buffer (250 mM sodium cacodylate, pH 7.5; 1.5 M KCl; 25 mM MgCl₂) was added to the samples which were then incubated at 37 °C for 20 minutes. Next, 2 μg of total yeast tRNA (Sigma Aldrich) was added to each sample and the volume was adjusted to 15 μL by adding 1X dimer buffer and incubated for 10 minutes at room temperature. 3 μL of a 1 μM BzCN solution in anhydrous dimethyl sulphoxide (DMSO) was used to modify the RNA samples for 1 minute and the reaction was stopped by adding 82 μL water. The negative control samples were treated in the same fashion but using only DMSO in the absence of BzCN. All samples were then precipitated using 1 μL of 1 μg/μL glycogen solution, 1/10 volume 3M sodium acetate (pH 6.5) and 3 volume ethanol for 30 minutes on dry ice and the precipitates were collected by centrifugation at 13,000 g for 20 minutes at 4 °C. The RNA pellets were washed twice with 500 μL cold 80% ethanol to remove salts, dried in a vacuum dryer and dissolved in 7 μL milli-Q water. Next, to identify the BzCN modification sites, reverse transcription (RT) was performed on the samples. Towards this end, two types of primer sequence were used to ensure RT of the entire 550 nt MPMV leader RNA. A set of four AS primers labeled with different dyes corresponded to the 5’ AGT TAC TGG GAC TTT CTC CG 3’ sequence (complementary to MPMV nt 483-502) and were termed as OTRs 17, 18, 19 and 20. A second set of AS primers corresponded to the 5’ CTT ACT TTC AGG TCC AAC GC 3’ sequence (complementary to MPMV nt 235-254), and were named OTRs 21, 22, 23 and 24. These primers were labeled with either 6-FAM, VIC, NED or PET. The purpose of using different dyes was that each dye could be detected individually after capillary electrophoresis in a DNA sequencer as described earlier (Merino et al., 2005;
Mortimer & Weeks 2007, 2009). BzCN-modified RNAs were annealed to 1 μM VIC-labeled primers OTR 18 or OTR 22 for 2 minutes at 90 ºC and 2 minutes in ice. After the addition of 2 μL 5X RT buffer (Life Science, USA), the samples were then incubated for 10 minutes at room temperature. Each sample was then divided into two aliquots and elongation reaction was performed for 30 minutes at 42 ºC and for 15 minutes at 50 ºC in elongation buffer (1 μL 5X RT buffer, 3 μL 2.5 mM dNTPs mix, 1 μL 2U/μL RT AMV enzyme from Life Science, USA and water to make the volume to 10.5 μL). For the unmodified RNA samples, 6-FAM labeled primers OTR 17 or OTR 21 were used and RT reaction was performed in the same manner as for the modified RNA samples. A ddA sequencing ladder was prepared using 2 picomoles of untreated RNA and 1 μL of 2 μM NED labeled OTR 19 or OTR 23 in 8 μL milliQ water. Annealing was performed by heating at 90 ºC for 2 minutes and cooling on ice for 2 minutes. The volume was made to 10 μL by adding 2 μL 10X RT buffer and then the samples were incubated for 15 minutes at room temperature. The RNA sample was aliquoted into 2 tubes, and the elongation reaction was performed with 1 μL 10X RT buffer, 3 μL A10 (0.25mM dATP, 1mM dGTP, 1mM dCTP, 1mM dTTP), 1 μL of 100 μM ddA and 1 μL of 1U/μL RT AMV enzyme. A ddG ladder was also prepared in the same fashion by using PET labeled OTR 20 or OTR 24, G10 (0.25mM dGTP, 1mM dATP, 1mM dCTP, 1mM dTTP) and 100 μM ddG. All the reactions were stopped by adjusting the volume 45 μL with water and proteins were extracted with 50 μL phenol-chloroform. For each experiment, the modified and unmodified samples were pooled with the corresponding ddA and ddG sequencing ladders (e.g., samples for which OTRs 17, 18, 19 and 20 were used were pooled together) in a single tube containing 20 μL 3M sodium acetate and 600 μL ethanol.
for cDNA precipitation. The samples were incubated on dry ice for 30 minutes, centrifuged at 13,000 g for 20 minutes at 4 °C, and washed twice with 1 mL cold 80% ethanol. Pellets were spun for 5 minutes at each step to prevent material loss. After drying the pellets in a vacuum dryer, they were resuspended in 10 μL HiDi formamide (ABI) and heat denatured at 90 °C and iced for 5 minutes each. The samples were spun down before loading on the 96 well plates for sequencing on an Applied Biosystems 3130xl genetic analyzer. The results were generated in the form of electropherograms, which were analyzed with the SHAPEfinder program (Vasa et al., 2008; Wilkinson et al., 2008) by following the steps prescribed by program developers such as baseline adjustment, matrixing, cubic mobility shift, signal decay correction, scale factor, alignment and integration to MPMV sequence, and finally fitting. This software enables the calculation of the SHAPE reactivity at each nucleotide. Prior to performing the SHAPE reactions, we calibrated the fluorescent tagged primers for determining the mobility shift (Vasa et al., 2008). The peaks in the electropherograms were normalized by subtracting the peak area of the unmodified RNA sample from the peak area of the SHAPE treated sample (Vasa et al., 2008). The data was then further normalized for outliers by determining 10% of the highest values and excluding the highest 2% of these values. The remaining 8% of the values were averaged and the average value was used to divide the SHAPE data. The SHAPE reactivity data (provided as supplemental data) obtained for each nucleotide were then applied as constraints in the structure prediction program RNAstructure (version 5.3, Reuter and Mathews 2010) in order to obtain the SHAPE-derived RNA structure of the MPMV packaging signal RNA.
2.3.5. **In vitro dimerization assay**

*In vitro* dimerization was performed as described previously (Marquet et al., 1991). Briefly, 300 nM of the purified wild type or pal mutant RNAs were incubated in dimer (250 mM sodium cacodylate, pH 7.5; 1.5 M KCl; 25 mM MgCl$_2$) or monomer (250 mM sodium cacodylate, pH 7.5; 200 mM KCl; 0.5 mM MgCl$_2$) buffer for 30 minutes at 37 °C. Following this incubation, loading dye containing glycerol was added and samples were subjected to electrophoresis in native 1% TBM (50 mM Tris base, 45 mM boric acid, 0.1mM MgCl$_2$) or semi-denaturing TB (50 mM Tris base, 45 mM boric acid) agarose gel at 4 °C or 20 °C, respectively, stained with ethidium bromide and visualized for band shift using ultraviolet (UV) transillumination. Band intensities were measured using Quantity One software and percentage (%) of dimerization was calculated for each mutant employing the formula: (Intensity of dimer band - Intensity of background)/ [(Intensity of monomer band - Intensity of background) + (Intensity of dimer band - Intensity of background)]. Results for each pal mutant were plotted as relative dimerization to the wild type values.
2.4. Results and Discussion

To validate the predicted structure (Jaballah et al., 2010) and to generate a more precise structural model of the MPMV gRNA packaging signals, we employed selective 2’ hydroxyl acylation analyzed by primer extension (SHAPE), which examines the flexibility of the RNA backbone at each nucleotide position (Merino et al., 2005; Mortimer & Weeks, 2007, 2009). SHAPE reagents target and modify the ribose moiety of all four nucleotides in single-stranded regions of a folded RNA molecule, whereas the nucleotides that are base-paired or architecturally constrained do not react or show reduced reactivity.

2.4.1. SHAPE data validated the major structural motifs predicted in the 5’ end of MPMV gRNA

Our SHAPE data validated the overall predicted structure (Jaballah et al., 2010) of the MPMV packaging signal RNA. Figure 2.1A shows the unconstrained predicted structure, whereas Figure 2.1B shows the SHAPE-validated MPMV structure. An average of SHAPE reactivities numerical data for each nucleotide that were drawn from 2-4 experiments were applied as pseudo energy constraints in RNAstructure (Bellaousov et al., 2013; Reuter & Mathews, 2010). As shown in Figure 2.1B, the SHAPE data robustly corroborated the free-energy minimized predicted structure of the MPMV packaging signal RNA presented in Figure 2.1A. Briefly, the SHAPE-validated model displayed major structural components including pal SL, ssPurine-rich region, U5-Gag LRIs, SLs 1-3, and Gag SLs 1-2 in extensive conformity with the predicted structure. Consistent with the predicted structure, all the loops and bulges containing unpaired nucleotides showed SHAPE reactivity. A few base-paired nucleotides located at the ends of a helix, near a bulge or a loop also showed reactivity (for example nt 22, 25, and 26). Despite such a high conformity between the predicted
and the SHAPE-validated structure models, some noticeable differences could be observed. First, SHAPE data indicated that the actual apical loop of SL2 is larger than predicted by Mfold (Fig. 2.1A and 2.1B). Second, this loop is shifted by two nucleotides, inducing different base-pairing in the upper part of SL2, including the primer binding site (PBS). Third, a small stem loop was observed in the SHAPE-validated structure at the base of Gag SL2, which was not present in the predicted structure model of the MPMV packaging signal RNA. Finally, even though the same pal SL is present in the two structures, it is interesting to observe that the nucleotides in the loop, which contains the “GC” dyad, displayed weak or no reactivity, suggesting that these nucleotides are indeed base-paired. One possibility is that these nucleotides may be base-paired with the same sequence on the other gRNA which is consistent with the hypothesis that the pal SL might function as the DIS of MPMV gRNA (Jaballah et al., 2010).

A very distinguishing feature of the predicted and the SHAPE-validated structure is the two LRIs (LRI-I and LRI-II) involving sequences from U5 and first 70 nt of gag (Fig. 2.1), which have been shown to be important in MPMV RNA packaging (Jaballah et al., 2010; Schmidt et al., 2003). Except for canonical base pairs at both ends of LRI-I, and wobble base pairs in both LRIs, all nucleotides (except U62) in the complementary sequences of the LRIs were non-reactive to SHAPE reagent and, therefore, were base-paired as initially predicted (Jaballah et al., 2010). U62 is highly reactive to the SHAPE reagent and forms a bulge in LRI-II that was not predicted by Mfold.

2.4.2. Sequences in pal SL augments MPMV gRNA dimerization by functioning as DIS

Earlier studies have shown that deletion of the 14 nt pal SL severely compromised MPMV gRNA packaging (Jaballah et al., 2010; Schmidt et al.,
2003). However, these studies did not study the impact of these mutations on gRNA dimerization. The conservation of the pal sequence in different strains of MPMV (Fig. 2.8) and the fact that it forms a hairpin with a “GC” dyad in the loop (Fig. 2.1) makes it a credible candidate to function as MPMV gRNA DIS. Therefore, we introduced a series of mutations in the pal sequence (Fig. 2.2A) and performed in vitro dimerization assays using in vitro transcribed RNAs. The mutant RCR002 contains the complete deletion of the pal sequence. In order to ascertain whether preservation of the hairpin structure is sufficient to initiate dimerization, mutant RCR003 was created in which the central 4 nt of the loop were substituted with a stable tetra loop sequence. In order to determine if it is the primary sequence or palindromic nature of the pal sequence that is required for MPMV gRNA dimerization another mutant (RCR004) was constructed, which contains the substitution of the central 6 nt of MPMV pal sequence with that of HIV-1 pal sequence. Before introducing these mutations, Mfold was used to predict the structure of the mutants in the monomer form to support the premise of our experimental hypothesis.

The data obtained from three independent in vitro dimerization experiments using wild type and pal mutant RNAs is summarized in Figure 2.2B and 2.2C. In vitro transcribed wild type RNA showed around 87% and 65% dimerization in TBM (50 mM Tris base, 45 mM boric acid, 0.1mM MgCl₂; electrophoresis at 4 °C) and TB (50 mM Tris base, 45 mM boric acid; electrophoresis at room temperature) gels, respectively (Fig. 2.2B). These wild type (RCR001) values were considered as references for calculating the relative dimerization efficiency (RDE) of the mutants (Fig. 2.2C). The pal deletion and substitution mutants showed significant degrees of reduction in their abilities to
dimerize (RDEs 0.19-0.25 in both semi-native and native conditions; 4-5.27 fold reduction; $P=0.002-0.02$; Fig. 2.2C) compared to the wild type. Mutants RCR002 and RCR003 showed that not only the pal SL structure but also the nature of the pal sequence is important for MPMV gRNA dimerization. The results obtained with mutant RCR004 are more difficult to explain. The Mfold structural model of RCR004 predicted the formation of a stem loop similar to MPMV pal SL (data not shown) and therefore this mutant was expected to restore dimerization to wild type levels, in contrast with our observation (Fig. 2.2B and 2.2C). In the case of HIV-1, the 5’ GCGCGC 3’ pal is flanked by three purines residues that are crucial for the stability of the RNA dimer (Paillart et al., 1997). Therefore, the inability of RCR004 to restore dimerization could in part be attributed to the absence of flanking purines similar to HIV-1 pal in its native genomic context.

Next we wanted to correlate the results of \textit{in vitro} dimerization to the SHAPE-validated structural model and establish the structure-function relationships, if any, of MPMV pal SL during gRNA dimerization. Therefore, we folded two molecules of the wild type 5’ end MPMV gRNA using RNAstructure to predict the folding pattern of a dimer. Folding prediction of the wild type homodimer showed an overall preservation of the RNA secondary structure of the monomer, with a conformational change in the pal SL, which was predicted to interact and base pair with the pal sequence on the second MPMV gRNA sequence (Fig. 2.3). Thus, this structure prediction points towards the importance of the pal SL sequence in MPMV gRNA dimerization, consistent with our RNA dimerization assay and with our SHAPE data, which showed that the loop of the pal hairpin is only moderately reactive. Furthermore in the homodimer structure,
FIGURE 2.3. RNAstructure prediction of the MPMV wild type (RCR001) RNA (nt +1-388) homodimer structure. The site of interaction between the two gRNAs involving the pal sequence is enlarged for clarity. The central 6 nt of the pal sequences are shown in red.
Homodimerization via wild type pal

Predicted RCR001 dimer
\[ \Delta G = -254.2 \text{ kcal/mol} \]
the ssPurine-rich region was predicted to form a helix loop structural motif (Fig. 2.3) possibly making it more accessible for possible interaction with NC protein as has earlier been suggested (Jaballah et al., 2010). In sharp contrast, the homodimer structural predictions of mutants RCR002 and RCR003 suggested that these mutants could dimerize using another 6 nt pal sequence (5’ GGCGCC 3’) within the PBS (data not shown). This PBS-mediated homodimerization results in the overall destabilization of the RNA secondary structure and is consistent with our in vitro dimerization results. Taken together, our in vitro dimerization and structure prediction results suggest that the 14 nt GC-rich MPMV pal SL is important in mediating gRNA dimerization by functioning as a DIS.

2.4.3. In vitro heterodimerization can be mediated by trans-complementary sequences on two RNAs

To confirm that the central 6 nt of the pal indeed constitute the point of contact between two gRNAs to initiate the process of dimerization, we created two trans-complementary mutants maintaining the central “GC” dyads (RCR005 and RCR006; Fig. 2.4A). When incubated separately, the RCR005 and RCR006 mutant RNAs formed significant amounts of dimer in TBM, even though dimerization was reduced compared to wild type RNA (RDEs; 0.69-0.70; 1.43-1.45 fold reduction; $P= 0.029-0.047$; Fig. 2.4B, C). However, the ability of these mutant RNAs to dimerize was severely compromised in semi-native condition (RDEs; 0.14 and 0.09 for RCR005 and RCR006, respectively; 7.14-11.11 fold reduction; $P= 0.062-0.042$; Fig. 2.4B lower panel; Fig. 2.4C). By contrast, when incubated together, these trans complementary mutant RNAs dimerized to wild type level, both in native and semi-native conditions (RDEs 0.88-1.09; Fig. 2.4B last lane, and Fig. 2.4C).
FIGURE 2.4. *In vitro* heterodimerization can be mediated by *trans*-complementary sequences on two RNAs. (A) Table showing the 6 nt wild type pal sequence (in red) and the two complementary substitution mutations (in blue). (B) *In vitro* dimerization assay of the MPMV WT and pal mutants. The upper panel shows native 1% TBM gel run at 4 °C and the lower panel shows semi-native 1% TB gel run at room temperature. M: monomer lane or monomer conformer; D: dimer lane or dimer conformer for each sample. (C) Relative dimerization of the complementary mutants to the wild type RNAs. Dimerization abilities of the RNAs in each lane were calculated and the dimerization data was represented as relative to the wild type dimerization.
### A

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Description of Introduced Mutation</th>
<th>Sequence in Wild Type And Mutants</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCR001</td>
<td>Wild type pal stem loop sequence</td>
<td>5’ TCGCCGGCCGGCGA3’</td>
</tr>
<tr>
<td>RCR005</td>
<td>Substitution of 6 nt of the wild type pal sequence with non palindromic sequence</td>
<td>5’ TCGCACGCACGCGA3’</td>
</tr>
<tr>
<td>RCR006</td>
<td>Substitution of 6nt of the wild type pal sequence with non palindromic sequence</td>
<td>5’ TCGCGTGCGTGCGA3’</td>
</tr>
</tbody>
</table>

### B

**1% TB gel**

![Image of gel showing bands for RCR001, RCR005, and RCR006]

**1% TBM gel**

![Image of gel showing bands for RCR001, RCR005, and RCR006]

### C

![Bar graph showing relative dimerization for RCR001, RCR005, RCR006, and RCR005 + RCR006](image)

<table>
<thead>
<tr>
<th></th>
<th>RCR001</th>
<th>RCR005</th>
<th>RCR006</th>
<th>RCR005 + RCR006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Dimerization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBM</td>
<td>1.00</td>
<td>0.69</td>
<td>0.70</td>
<td>0.88</td>
</tr>
<tr>
<td>TB</td>
<td>1.00</td>
<td>0.14</td>
<td>0.09</td>
<td>1.09</td>
</tr>
</tbody>
</table>
These results prove that an intermolecular interaction mediated by the central 6 nt of pal is required for the formation of a stable RNA dimer. In addition, they reveal the existence of a second intermolecular interaction, which allows formation of unstable dimers. Consistent with this interpretation, the unstable homodimers formed by RCR005 and RCR006 RNAs migrated slightly more slowly than the heterodimer or the wild type RNA homodimer (top panel of Fig. 2.4B), confirming that they formed different RNA dimers.

In an attempt to identify the residual intermolecular interaction in mutants RNAs, we employed RNAstructure to predict the RCR005 and RCR006 homodimers and heterodimers structures. Structure predictions suggested that the mutant RNAs RCR005 and RCR006 homodimerized via an alternate 6 nt pal sequence within the PBS (Figs. 2.5A, B, 2.6B, C). As a consequence, SL2, which is relatively long in the wild type homodimer and in the RCR005/RCR006 heterodimer, was predicted to be shortened due to PBS mediated intermolecular interactions; however its apical loop was maintained (Figs. 2.5D, 2.6D). In the predicted heterodimer structure of mutants RCR005 and RCR006, the two RNAs interact via the substituted trans complementary sequences (Figs. 2.5D, 2.6D) as expected, and as supported by our experimental data (Fig 2.4B and C). As a result, the secondary structures of the wild type RNA homodimer and of the RCR005/RCR006 heterodimer are identical (Fig. 2.6A and D). RNAstructure predictions of the RCR005 and RCR006 homodimers and heterodimer illustrate that not only the interaction between these trans complementary sequences result in heterodimerization, but also that such interaction leads to conformational changes in the structure allowing more extensive base pairing between complementary sequences throughout the genome (2.5D).
FIGURE 2.5. Dimer predictions of the mutant RCR005 (A) and RCR006 (B) MPMV RNAs (nt +1-388) using RNAstructure showing that mutant RNAs dimerize via a sequence in the primer binding site. (C) Schematic representation of the expected point of heterodimerization between RCR005 and RCR006 MPMV RNAs. (D) Predicted heterodimer structure of RCR005 and RCR006 RNAs. The complete predicted structures of the homodimers as well as heterodimers are shown in figure 2.6.
A homodimer
\[ \Delta G = -245.6 \text{ kcal/mol} \]

B homodimer
\[ \Delta G = -248.0 \text{ kcal/mol} \]

C predicted RCR005-RCR006 heterodimer
\[ \Delta G = -249.6 \text{ kcal/mol} \]

Non palindrome complementary sequences

Point of dimerization

SS Purine-rich region
FIGURE 2.6. Predicted homodimer (A-C) and heterodimer (D) structures of MPMV wild type (RCR001) and trans complementary mutants (RCR005 and RCR006). The dimers were predicted using RNAstructure and the contact point of the sequences between two RNAs leading to dimerization are boxed. Mutants RCR005 (B) and RCR006 (C) are predicted to homodimerize via the primer binding site (PBS). The heterodimer of RCR005 and RCR006 (D) is predicted to dimerize to exactly the same structural conformation as the WT (RCR001) homodimer (A). The trans-complementary sequences in RCR005 and RCR006 homodimers and heterodimer are shown in blue color and are marked by blue arrows.
BRCR005 homodimer
ΔG = -245.6 kcal/mol

RCR006 homodimer
ΔG = -248.0 kcal/mol

RCR005-RCR006 heterodimer
ΔG = -249.6 kcal/mol

“represent point of homo- and hetero-dimerization”
In summary, our systematic mutational analyses directed towards ascertaining the functional significance of pal SL demonstrated that mutations targeted towards disrupting the formation of the 6 nt pal loop diminished RNA dimerization (Fig. 2.2 and 2.4). This could be attributed to the RNA folding pattern rather than the primary sequence of the pal loop since mutants containing trans complementary sequences in this loop were not only able to heterodimerize at wild type levels, but were also able to re-establish the overall wild type RNA secondary structure (Fig. 2.6). These results strongly suggest that the pal loop is involved in “kissing loop” interactions between two MPMV gRNAs in order to initiate gRNA dimerization, and thus functions as DIS. These observations corroborate with an earlier genetic analysis that showed no effect on RNA packaging when the MPMV native pal sequence was accidentally substituted with a different palindromic sequence (Mustafa et al., 2004).

2.4.4. The SHAPE-validated secondary structural model is supported by sequence conservation between different MPMV isolates

In order to examine the conservation of different structural motifs within the MPMV packaging signal RNA, sequences from different strains of MPMV were used to predict the RNA secondary structure of this region. Mfold predictions of the sequences from different MPMV strains consistently revealed the ability to form both LRIs involving sequences from U5 and Gag (Fig. 2.7A-E). A sequence alignment of this region for these MPMV isolates was generated using ClustalW and revealed a high degree of conservation of the sequences in the two U5-Gag LRIs. The U5 and Gag sequences forming these LRIs slightly differ but maintain a very high degree of complementarity (Fig. 2.7A-E and 2.8).

Based on the RNA free-energy minimized model predicted earlier and validated by SHAPE (Fig. 2.1) in addition to phylogenetic, and in silico modeling
FIGURE 2.7. Mfold structural predictions of the 5’ end genomes of different MPMV strains (A-E). The U5/Gag LRIs and the pal sequence are highlighted by red and green boxes, respectively. The accession numbers for MPMV6/A, 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 2.8. ClustalW sequence alignment of the 5’ end genomes of different MPMV strains. The aligned sequences pertaining to major structural motifs are highlighted by different colors and boxed. The accession numbers for MPMV6/A, 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.
analyses presented here (Fig. 2.7 and 2.8), it is reasonable to hypothesize that such LRIs between U5 and Gag sequences are likely to exist in vivo and have a functional role in MPMV life cycle. Consistent with this, deletion of the gag sequences involved in U5-Gag LRI-II while maintaining the gag sequences that are involved in U5-Gag LRI-I severely diminished MPMV transfer vector RNA packaging ability (Schmidt et al., 2003). On the other hand, inclusion of first 100 nt of gag sequences that would maintain both U5-Gag LRIs in the transfer vector RNA restored packaging efficiency to the wild type level (Schmidt et al., 2003). These results suggest that LRI-I and LRI-II play an important architectural role in stabilizing the RNA secondary structure of the 5’ UTR sequences required for MPMV gRNA packaging (Jaballah et al 2010; Schmidt et al., 2003). Such a scenario could be comparable to that observed in other retroviruses (HIV-1, HIV-2, and FIV) where interactions between the R/U5 regions and the start of gag sequences have been proposed to stabilize the overall RNA secondary structure essential for gRNA dimerization and packaging (Kenyon et al., 2008, 2011; Lu et al., 2011b; Paillart et al., 2002; Rizvi et al., 2010; Song et al., 2008;).

Between SL2 and SL3 of our SHAPE-validated RNA structure is a very distinctive pal SL, which has earlier been shown to be required for optimal packaging of MPMV gRNA (Jaballah et al., 2010). The current study revealed that the central 6 nt of this pal (5’ CGGCCG 3’) functions as DIS, thus probably facilitating the encapsidation of the dimeric genome. Sequence alignment of five different strains of MPMV revealed that the 6 nt central pal is conserved in all but one strain (Fig. 2.8). In one strain (SRV1), the last G is substituted with a C residue, but this strain still contains a 6 nt GC-rich pal (5’ GCCGGC 3’) shifted by two nucleotides compared to the other strains (Fig. 2.8), supporting a functional
role for this pal. Mfold analysis of these strains further revealed that in three strains (MPMV 6/A, SRV1 and SRV2) the pal sequence fold into a stem loop (Fig. 2.7A-C). In the case of SRV4, the pal sequence is base-paired; whereas in the case of SRV5, the pal sequence forms two small bulges (Fig. 2.7D-E). Similar conservation of a DIS comprising a pal sequence has been reported in several retroviruses including HIV-1, HIV-2, SIV, and FIV (Kenyon et al., 2008, 2011; Lever 2007; Paillart et al., 2002; Russell et al., 2004). In a number of retroviruses, sequences augmenting RNA dimerization and packaging have been shown to be intermingled (Lanchy et al., 2003; Lanchy & Lodmell, 2007; Lever, 2007; Paillart et al., 2004; Russell et al., 2004), which is in agreement with the deletion of MPMV pal SL resulting in ablation of RNA packaging (Jaballah et al., 2010).

Earlier studies have identified a prominent stretch of ssPurine-rich region between pal SL and SL3 and a partial base-paired repeat sequence of the ssPurine-rich region was observed in the adjacent region (“B”; Fig. 2.1A; Jaballah et al., 2010). Sequence alignment revealed that out of the 16 nt within the ssPurine-rich region, 10 nt are conserved among all the analyzed MPMV strains (Fig. 2.8). The validation of the ssPurine-rich region and its based-paired repeat structural motifs by SHAPE (Fig. 2.1B) in concurrence with sequence alignment (Fig. 2.8) and earlier mutational and structure prediction analyses (Jaballah et al., 2010) further emphasize the importance of the ssPurine-rich region in MPMV life cycle.

### 2.4.5. SL1, SL2, and SL3 are present in spliced as well as genomic RNAs

To establish the structure-function relationship of structural motifs (especially of those present upstream of the major splice donor (mSD) site) and to address how some of them may help the MPMV packaging machinery in differentiating between genomic and spliced RNA, we predicted the structure of
the 5’ region of the MPMV env spliced RNA (including 251 nt upstream of the mSD and 138 nt downstream of env splice acceptor site). Mfold structural predictions revealed that SL1, SL2, and SL3 are present in both MPMV 5’ end genomic and env spliced RNAs (Fig. 2.9). SL1, SL2, and SL3 were also predicted to be present in the shorter MPMV leader RNA that was truncated at the mSD (+1-251; data not shown). The function of SL1 in the retroviral life cycle has not yet been established, however, Mfold predictions (Fig. 2.7) and sequence alignments (Fig. 2.8) of 5 different strains of MPMV showed that this motif is conserved at the sequence and structural levels.

SL2 is a rather long and stable structure from which PBS protrudes (Fig. 2.1A, B). It is present in the SHAPE-validated structure of the genomic RNA (Fig. 2.1B) and is predicted to exist in the env spliced RNA (Fig. 2.9). In silico analysis revealed a high degree of conservation in the sequences forming SL2 (Fig. 2.8), which was predicted to fold in three out of the five MPMV strains (MPMV 6/A, SRV4 and SRV5) (Fig. 2.7A-E). In addition, SL2 has been shown to include sequences that constitute part of a bipartite packaging signal (Jaballah et al., 2010). All together these data point towards the importance of the overall RNA secondary structure of SL2 in the MPMV life cycle (Jaballah et al., 2010; Schmidt et al., 2003).

Immediately downstream of the ssPurine-rich region is SL3, which like SL1 and SL2 was consistently predicted to fold in both genomic and spliced RNAs (Fig. 2.9). Mfold predictions of different MPMV strains demonstrated that SL3 is formed in all except for one strain (Fig. 2.7A-E) and its sequences are fairly conserved among different strains (Fig. 2.8). However, deletion of the majority of the sequences that are involved in base pairing of SL3 did not affect
FIGURE 2.9. SL1, SL2, and SL3 are present in genomic as well as spliced RNA. (A) Mfold structural predictions of MPMV 5’ end genomic RNA. (B) Mfold structural predictions of MPMV envelope spliced RNA. In both cases, the first 388 nt (from the start of R) of the RNAs were folded.
Mfold predicted structure of the 5' end MPMV genomic
\( \Delta G = -124.70 \text{ kcal/mol} \)

Mfold predicted structure of the MPMV Env spliced RNA
\( \Delta G = -118.80 \text{ kcal/mol} \)

A

Region “A”

Region “B”

SL2

pal SL

mSD

ssPurine-rich region

PBS SL

U5/Gag LRI-II

U5/Gag LRI-I

Gag SL1

SL1

B

Mfold predicted structure of the MPMV Env spliced RNA
\( \Delta G = -118.80 \text{ kcal/mol} \)

ssPurine-rich region

Env SL1

Env SL2

Env SL3

SL3

pal

Region “A”

PBS SL

SL2

SL1
RNA packaging (Jaballah et al., 2010), suggesting that it might play a role in MPMV life cycle that has not yet been established.

In conclusion, employing SHAPE, we validated the overall predicted structure of the MPMV packaging signal RNA (Jaballah et al., 2010; Fig. 2.1). Briefly, MPMV packaging signal RNA comprised five stem loops structures. SL1-SL3 include sequences from R/U5 and UTR whereas Gag SL1 and Gag SL2 employ sequences exclusively from \textit{gag}. SL2 involves sequences from region “A” whereas SL3 in part comprises sequences from region “B”; both of these regions have been shown earlier as major MPMV gRNA packaging determinants (Schmidt et al., 2003; Jaballah et al., 2010). Between SL2 and SL3 is a very distinctive pal SL followed by a stretch of ssPurine-rich region, both of which have also been shown to be required for optimal packaging of MPMV gRNA (Jaballah et al., 2010). The overall structure is anchored by two LRIs (Fig. 2.1A, B) involving sequences from U5 and the first 70 nt of \textit{gag}, which have been shown to be important in MPMV RNA packaging (Schmidt et al., 2003; Jaballah et al., 2010). Mfold predictions of different MPMV strains consistently revealed the ability to form LRIs, suggesting a functional role for these LRIs (Fig. 2.7A-E).

Employing a systematic mutational approach, we showed that the central 6 nt of the pal, which are exposed in the pal SL loop play a crucial role in RNA dimerization, by directly interacting with a second genomic RNA molecule (Fig. 2.2, 2.4 and 2.6). Thus, this sequence likely initiates genomic RNA dimerization by forming a kissing loop complex. Despite some sequence variability, all MMPV strains maintain a 6 nt GC pal at this position, reinforcing its functional role.

The SHAPE-validated structural model, \textit{in vitro} dimerization analysis of pal mutant and the homo- and heterodimer structural models presented here
coupled with earlier published findings on MPMV RNA packaging signals, offer an important functional correlation between gRNA dimerization and the packaging processes. These observations while broadening the understanding of functional regions of MPMV gRNA, further suggest that MPMV may select a dimeric genome for packaging.
2.5. Acknowledgments

This research was funded primarily by a grant from the United Arab Emirates University (UAEU-NRF 26) to TAR and, in part, by a grant from the College of Medicine and Health Sciences (NP 09/18). RM is supported by the Centre National de la Recherche Scientifique (C.N.R.S.). SJA is supported through a scholarship from the Office of the Vice Provost for Research and Graduate Studies of the UAEU. The authors wish to thank Ms. Reka Raghunathan Chakravarthy for her assistance in cloning some of the mutant clones.
CHAPTER 3

A palindromic stem loop functions as the dimerization initiation site controlling packaging of the genomic RNA of mouse mammary tumor virus (MMTV)

This chapter is a modified version of the manuscript submitted for publication in *Nucleic Acids Res.* (Aktar, S. J., et al., submitted March 2014)
3.1. Abstract

Dimerization of genomic RNA plays a key role in the retroviral life cycle, as it is intimately linked to gRNA packaging. Earlier mutational analysis of the MMTV genome indicated that MMTV gRNA packaging determinants comprise sequences both within the 5’ untranslated region and the beginning of gag. To establish structure-function relationships of various stem loops in the packaging determinants, we elucidated RNA secondary structure of this region employing selective 2’hydroxyl acylation analyzed by primer extension (SHAPE). SHAPE analyses revealed presence of a U5-Gag long-range interaction (U5-Gag LRI), not predicted by minimum free-energy structure predictions that potentially stabilizes the global structure of this region. Structure conservation along with base-pair covariations between different strains of MMTV further supported the SHAPE-validated model. RNA dimerization, packaging, and propagation assays on a series of mutants, coupled with SHAPE and structure predictions, revealed that a palindromic sequence containing a canonical “GC” dyad (5’ CGGCCG 3’; pal II) likely constitutes the MMTV RNA dimerization initiation site (DIS). A second pal within the primer binding site (PBS) is also involved in RNA dimerization, and the data presented here suggest that both pals are required for efficient RNA dimerization, packaging and propagation.
3.2. Introduction

An essential step in the retroviral life cycle is the efficient and specific packaging of two copies of plus-strand full length genomic RNA (gRNA) into the virus particle from a large pool of cellular and other viral RNAs in the cytoplasm (reviewed in D’Souza & Summers, 2005; Johnson & Telenitsky, 2010; Lever, 2007; Lu et al., 2011; Paillart et al., 2004; Russell et al., 2004). Retroviral RNA packaging process involves the recognition of packaging signal(s) by the zinc finger(s) in the nucleocapsid (NC) domain of the Gag polyprotein (reviewed in D’Souza & Summers, 2005; Johnson & Telenitsky, 2010; Lever, 2007; Lu et al., 2011). Over the years, it has been shown that the packaging signals of different retroviruses reside at the 5’ end of the gRNA and include continuous or discontinuous sequences from the R, U5, and 5′ untranslated (UTR) regions and extend into the 5′ end of the \textit{gag} gene (reviewed in D’Souza & Summers, 2005; Jaballah et al., 2010, Johnson & Telenitsky, 2010; Lever, 2007; Mustafa et al., 2005, 2012).

The retroviral RNA genome is packaged as a non-covalent dimer, and the processes of dimerization and packaging are closely interlinked (reviewed in D’Souza & Summers, 2005; Johnson & Telenitsky, 2010; Lever, 2007; Lu et al., 2011; Moore & Hu, 2009; Paillart et al., 2004; Russell et al., 2004). The sequences responsible for retroviral gRNA packaging and dimerization of a variety of retroviruses map to the same ~100-400 nucleotides (nt) at the 5’ end of the gRNA, and in most cases are genetically indistinguishable (reviewed in D’Souza & Summers, 2005; Johnson & Telenitsky, 2010; Lever, 2007).

Recent retroviral RNA cross-packaging studies have shown that the specificity of retroviral gRNA packaging can be exchanged by substituting the
packaging signal between genetically diverse retroviruses (Al Dhaheri et al., 2009; Al Shamsi et al., 2011). This has further been substantiated by the fact that heterodimers involving RNAs from two divergent retroviruses with no sequence homology can also be packaged (Moore & Hu, 2009; Motomura et al., 2008; reviewed in Johnson & Telenitsky, 2010). Therefore, the process of gRNA dimerization and packaging is likely to involve recognition of structural motifs rather than primary sequences (Al Dhaheri et al., 2009; Al Shamsi et al., 2011; Moore et al., 2007; Parveen et al., 2004; Rizvi & Panganiban, 1993; White et al., 1999; Yin & Hu, 1997). Consistent with this, the packaging and dimerization sequences of almost all retroviruses have been shown to assume higher order structures comprising various structural motifs that have been shown to mediate RNA-RNA and RNA-protein interactions during retroviral RNA dimerization and packaging (reviewed in D’Souza & Summers, 2005; Johnson & Telenitsky, 2010; Lever, 2007; Lu et al., 2011; Miyazaki et al., 2011).

Retroviral RNA dimerization is usually mediated by a palindromic (pal) sequence known as the dimerization initiation site (DIS) present in the 5’ region of the gRNA. The DIS invariably assumes a hairpin structure containing a canonical “GC” dyad which interacts with the DIS loop on the second gRNA copy, resulting in a kissing loop interaction (Clever et al., 1996; Paillart et al., 1994, 1996a, 1997). Pal sequences involved in RNA dimerization have been identified in HIV-1, HIV-2, SIV (Baig et al., 2007; Berkhout & van Wamel, 1996; Lanchy et al., 2003; Lanchy & Lodmell, 2007; Laughrea et al., 1997; Paillart et al., 1996a; Russell et al., 2004; reviewed in Paillart et al., 2004), feline immunodeficiency virus (FIV; Kenyon et al., 2011; Rizvi et al., 2010) and Mason-Pfizer monkey virus (MPMV) (Aktar et al., 2013; Jaballah et al., 2010). In addition to gRNA
dimerization and packaging, mutations in the DIS have been shown to affect other steps in retroviral life cycle such as reverse transcription (Berkhout & van Wamel, 1996; Houzet et al., 2007; Laughrea et al., 1997; Paillart et al., 1996a; Shen et al., 2000) and recombination (Andersen et al., 2003; Balakrishnan et al., 2001; Chin et al., 2005). Since DIS has been shown to regulate retroviral life cycle, it is therefore an attractive target for antiretroviral drugs, such as aminoglycosides (Ennifar et al., 2003, 2006).

Despite having been studied extensively, little is known about the molecular mechanisms of gRNA dimerization and packaging during the MMTV life cycle (Cardiff & Kenny, 2007; Ross, 2010). An earlier study by Salmons et al., (1989) suggested that MMTV harbors sequences responsible for gRNA packaging in the 5’ region of its genome. Recently, employing a biologically relevant in vivo packaging and transduction assay (Rizvi et al., 2009), Mustafa et al. (2012) identified a continuous region (spanning the beginning of R to 120 nt of gag) critical for MMTV gRNA packaging and propagation. Folding algorithms predicted that these sequences fold into a higher order structure comprising of a number of stable structural motifs.

In this study we employed selective 2’hydroxyl acylation analyzed by primer extension (SHAPE; Merino et al., 2005; Mortimer & Weeks, 2007, 2009) to validate and further refine the predicted higher order features of the MMTV packaging signal. The SHAPE-validated RNA secondary structure revealed three prominent pal sequences (pal I, pal II and pal III) and one of them could potentially function as a possible DIS for MMTV gRNA dimerization. In vitro dimerization, in vivo packaging and transduction assays revealed that RNA dimerization, packaging and propagation processes of pal II mutants were greatly
compromised. Results obtained from these complementary approaches indicate that pal II forms a hairpin structure that plays an important role in MMTV gRNA dimerization and packaging by functioning as DIS.
3.3. Materials and Methods

3.3.1. *In silico* analyses

The secondary structure of the 5’ region of the MMTV gRNA (nt 1-432, i.e. including 120 nt of *gag*) was predicted using the Mfold server (Mathews et al., 1999; Zuker, 2003) and validated by applying SHAPE reactivity data in the RNAstructure software (version 5.3; Reuter & Mathews, 2010). The Mfold-predicted and SHAPE-validated structures were redrawn with the XRNA software (http://rna.ucsc.edu/rnacentral/xrna/xrna.html).

Nucleotides 1-432 of the gRNA of different strains of MMTV were aligned using Clustal Omega (Goujon et al., 2010; Sievers et al., 2011) and a consensus or “conservation annotation” structure for these strains was predicted using RNAalifold server (Bernhart et al., 2008). Accession numbers of MMTV strains are the same as mentioned in figure 3.3A. The number/color scale on the x-axis indicates six different types of base pairings (GC, CG, AU, UA, GU, UG) with 1 (red color) representing completely conserved base pairs to 6 (violet color) representing 6 different types of base pairings at that position among the various strains. The y-axis represents the incompatible nature of the base pairing (non-Watson base pairing) observed for each viral strain. The number (0-2) or color gradation (dark to very pale) represents increasing number of strains showing incompatible (non-Wastson) base pairing within the structure of the various strains. Thus, a dark red color denotes “highest” conservation of base pairings, while a light violet color represents “least” conservation of base pairing among different strains of MMTVs.
3.3.2. Construction of plasmids

The 5’ end of MMTV gRNA corresponding to nt 1-712 was amplified by polymerase chain reaction (PCR) using the MMTV sub-genomic transfer vector DA024 (Rizvi et al., 2009) as the template and the sense (S) OTR 984 (5’ CCGAAGCTTAAATACGACTCTATAGGGCAACAGTCTAATATTCACG 3’) and the antisense (AS) OTR 985 (5’ AAACCCGGGTTCCTGGTCCATAAG 3’) primers. These primers include the T7 promoter sequence (underlined) at the 5’ end along with flanking restriction sites (HindIII and XmaI/SmaI, shown in italics). The HindIII and Xmal sites were utilized to introduce the MMTV sequence into a pUC-based cloning vector (pIC19R; Marsh et al., 1984). The sequence of the resulting clone, SA035, was confirmed by sequencing. SA035 was also used as the wild type clone for in vitro gRNA dimerization assays.

A number of pal I (5’ GUCGGCCGAC 3’) and pal II (5’ CUGCAG 3’) mutations, including deletions, substitutions and substitution containing heterologous trans-complementary sequences, were introduced through splice overlap extension (SOE) PCR as described previously (Aktar et al., 2013; Gibbs et al., 1994; Jaballah et al., 2010; Rizvi et al., 2010). Briefly, in round one, the MMTV sub-genomic vector DA024 was amplified in two separate PCRs using S and AS primers generating products having overlapping complementary sequences. In round two, the products from round one PCRs which anneal via their overlapping complementary sequences were amplified using outer primers OTR 984 (S) and OTR 985 (AS), generating a final product containing the desired mutation(s). The S and AS primer sequences are provided in Table 3.1, along with the clone names and the necessary description. Conventional PCR was
**TABLE 3.1.** Description of primers used for cloning, sequencing, and conventional and real time PCR. Underlined sequence: T7 promoter. Sequence in small letters: dummy sequence and restriction enzyme sequence. *S, sense; AS, antisense. **HYBMTV, MMTV molecular clone created by Shackleford and Varmus (1988). Table is continued in the following page also.**
<table>
<thead>
<tr>
<th>Oligo Name</th>
<th>S/AS</th>
<th>Clone Name</th>
<th>Sequence</th>
<th>Nucleotide Position (nt)</th>
<th>Virus &amp; Gene (or plasmid)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTR 552</td>
<td>AS</td>
<td>-</td>
<td>5'cgactagtgatatcGTTCCCCTGGTCCCATAAG 3'</td>
<td>1885-1867</td>
<td>HYBMTV** gag</td>
</tr>
<tr>
<td>OTR 984</td>
<td>S</td>
<td>SA035</td>
<td>5'cccaagcttAATACGACTCACTATAGGGCAACAGTCTAATTCTACG 3'</td>
<td>1173-1193</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 985</td>
<td>AS</td>
<td>-</td>
<td>5'aaacccgggTTCCCCTGGTCCCATAAG 3'</td>
<td>1867-1885</td>
<td>HYBMTV gag</td>
</tr>
<tr>
<td>OTR 922</td>
<td>S</td>
<td>SA031/</td>
<td>5'CCCCGCAGATCGCGAGCAGTGGCGC 3'</td>
<td>1264-1313</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 923</td>
<td>AS</td>
<td>SA031P</td>
<td>5'GCTGCGCGATCTGGCGGGGGAACCCG 3'</td>
<td>1258-1307</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 924</td>
<td>S</td>
<td>SA032/</td>
<td>5'TTCGTGCGCGACTGGCGCCGCGCCGGC 3'</td>
<td>1299-1322</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 925</td>
<td>AS</td>
<td>SA032P</td>
<td>5'GCGGCACGAATCTGGCGGGGGAACCCCCTCTGGA 3'</td>
<td>1304-1252</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 926</td>
<td>S</td>
<td>SA033/</td>
<td>5'GCTGCGCGATCTGGCGGGGGAACCCG 3'</td>
<td>1260-1313</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 927</td>
<td>AS</td>
<td>SA033P</td>
<td>5'GCTGCGCGATCTGGCGGGGGAACCCG 3'</td>
<td>1307-1254</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 928</td>
<td>S</td>
<td>SA034/</td>
<td>5'TTCGTGCGCGACTGGCGCCGCGCCGGC 3'</td>
<td>1298-1316</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 929</td>
<td>AS</td>
<td>SA034P</td>
<td>5'GCGGCAGCGAATCTGGCGGGGGAACCCCCTCTGGA 3'</td>
<td>1303-1253</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 986</td>
<td>S</td>
<td>SA041/</td>
<td>5'GTAATCCCGCCTACGGAGAAGAGGTAG 3'</td>
<td>1442-1464</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 987</td>
<td>AS</td>
<td>SA041P</td>
<td>5'GCCGGATTACTCCCTATGGTGAGTCCGTTCC 3'</td>
<td>1434-1414</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 988</td>
<td>S</td>
<td>SA042/</td>
<td>5'CATGCAATGGCAGAAGAGGTAGG 3'</td>
<td>1446-1465</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 989</td>
<td>AS</td>
<td>SA042P</td>
<td>5'GTAGGCTATGGCAGTGGCCTGCCT 3'</td>
<td>1430-1410</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 990</td>
<td>S</td>
<td>SA043/</td>
<td>5'GGCGCGCTCCGCTACCGAGAAAGG 3'</td>
<td>1442-1461</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 991</td>
<td>AS</td>
<td>SA043P</td>
<td>5'GCCGGAGCGCGCCCTCCTATGGTGAGTCCGTTCC 3'</td>
<td>1434-1414</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 992</td>
<td>S</td>
<td>SA044/</td>
<td>5'GACGCCACTCCGCCACGAGGAAGG 3'</td>
<td>1442-1461</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 993</td>
<td>AS</td>
<td>SA044P</td>
<td>5'GCCGGAGGTGCGCTCCCTATGGTGAGTCCGTTCC 3'</td>
<td>1434-1414</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 994</td>
<td>S</td>
<td>SA045/</td>
<td>5'GTTGCCTGTCCGCTACCGAGAAAGG 3'</td>
<td>1442-1461</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 995</td>
<td>AS</td>
<td>SA045P</td>
<td>5'GCCGGAGCGCGCCCTCCTATGGTGAGTCCGTTCC 3'</td>
<td>1434-1414</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 1075</td>
<td>S</td>
<td>SA046/</td>
<td>5'CGACTGGCGGAAAGACGGGACCCCTCGG 3'</td>
<td>1315-1331</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 1076</td>
<td>AS</td>
<td>SA046P</td>
<td>5'CCTGTCGCGGCGACGCTGGCCGACCG 3'</td>
<td>1303-1288</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 1077</td>
<td>S</td>
<td>SA047/</td>
<td>5'AAGCGGCCCATCCTCCGCTAGGGGAAGG 3'</td>
<td>1442-1461</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 1078</td>
<td>AS</td>
<td>SA047P</td>
<td>5'GCCGGATGCGCGCTCCCTATGGTGAGTCCGTTCC 3'</td>
<td>1435-1415</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 930</td>
<td>S</td>
<td>-</td>
<td>5'GCTGCTGTGTTGGAGGTGCTGAG 3'</td>
<td>90-113</td>
<td>pcDNA3</td>
</tr>
<tr>
<td>OTR 725</td>
<td>S</td>
<td>-</td>
<td>5'GTCGCTGTGTTGGAGGTGCTGAG 3'</td>
<td>1325-1347</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 581</td>
<td>AS</td>
<td>-</td>
<td>5'GCCATGGGGAGGCGATCACC 3'</td>
<td>Actin A. Spliced or unspliced mRNA</td>
<td></td>
</tr>
<tr>
<td>Oligo Name</td>
<td>Clone Name</td>
<td>Sequence</td>
<td>Nucleotide Position (nt)</td>
<td>Virus &amp; Gene (or plasmid)</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>----------</td>
<td>--------------------------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td>OTR 552</td>
<td>AS</td>
<td>5' cgactagtgatatcGTTCCCCTGGTCCCATAAG 3'</td>
<td>1885-1867</td>
<td>HYBMTV**</td>
<td></td>
</tr>
<tr>
<td>OTR 984</td>
<td>S</td>
<td>SA035</td>
<td>5'cccaagcttAATACGACTCACTATAGG GCAACAGTCCTAATATTCACGTCTCGTGTG 3’</td>
<td>1173-1193</td>
<td>HYBMTV R (+1) start</td>
</tr>
<tr>
<td>OTR 985</td>
<td>AS</td>
<td>5'aaacccgggTTCCCCTGGTCCCATAAG 3'</td>
<td>1867-1885</td>
<td>HYBMTV**</td>
<td></td>
</tr>
<tr>
<td>OTR 922</td>
<td>S</td>
<td>SA031/SA031P</td>
<td>5'CCCCGCAGATGCGGCAGCTGGCGC 3’</td>
<td>1264-1313</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 923</td>
<td>AS</td>
<td>5' GCTGCCGCATCTGCGGGGGGACCCTCTGGA 3’</td>
<td>1304-1252</td>
<td>HYBMTV U5 region</td>
<td></td>
</tr>
<tr>
<td>OTR 924</td>
<td>S</td>
<td>SA032/SA032P</td>
<td>5' TTCGTGCGGCAGCTGGCGCCCG 3’</td>
<td>1299-1322</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 925</td>
<td>AS</td>
<td>5' GCCGCACGAATCTGCGGGGGGACCCTCTGGA 3’</td>
<td>1304-1252</td>
<td>HYBMTV U5 region</td>
<td></td>
</tr>
<tr>
<td>OTR 926</td>
<td>S</td>
<td>SA033/SA033P</td>
<td>5' CCCCCCGCAGCTGCGGCAGCTGGCGC 3’</td>
<td>1260-1313</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 927</td>
<td>AS</td>
<td>5' GCTGCCGCAGCTGCGGGGGGACCCTCTGGA 3’</td>
<td>1304-1252</td>
<td>HYBMTV U5 region</td>
<td></td>
</tr>
<tr>
<td>OTR 928</td>
<td>S</td>
<td>SA034/SA034P</td>
<td>5' TTCGCTGCGGCAGCTGGCGCCCG 3’</td>
<td>1298-1316</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>OTR 929</td>
<td>AS</td>
<td>5' CCGCAGCGAACTGCGGGGGGACCCTCTGG 3’</td>
<td>1303-1253</td>
<td>HYBMTV U5 region</td>
<td></td>
</tr>
<tr>
<td>OTR 930</td>
<td>S</td>
<td>-</td>
<td>5' GCTTGTGTGTTGGAGGTCGCTGAG 3’</td>
<td>90-113</td>
<td>pcDNA3</td>
</tr>
<tr>
<td>OTR 581</td>
<td>AS</td>
<td>-</td>
<td>5' GGCATGGGGGAGGGCATACC 3’</td>
<td>1325-1347</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 582</td>
<td>S</td>
<td>-</td>
<td>5' CCAGTGGCTTCCCCCAGTG 3’</td>
<td>2211-2230</td>
<td>HYBMTV UTR region</td>
</tr>
<tr>
<td>OTR 671</td>
<td>S</td>
<td>-</td>
<td>5' GTCCTAATATTCACGTCTCGTGTG 3’</td>
<td>1179-1202</td>
<td>HYBMTV R region</td>
</tr>
<tr>
<td>OTR 672</td>
<td>AS</td>
<td>-</td>
<td>5' CTG TTCGGGCAGCTGCGGCAGCTGGCGCAG 3’</td>
<td>1321-1298</td>
<td>HYBMTV PBS region</td>
</tr>
<tr>
<td>MTV-ILTR-SITEM1 FAM</td>
<td>-</td>
<td>-</td>
<td>5'TCGCCATCCCCGTCTCC 3’</td>
<td>1214-1229</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>MTV-ILTR-SITEF</td>
<td>S</td>
<td>-</td>
<td>5'CGTCTCGTGTGTTTGTGTTTGTGTTGTGTGTGTG 3’</td>
<td>1192-1213</td>
<td>HYBMTV U5 region</td>
</tr>
<tr>
<td>MTV-ILTR-SITER</td>
<td>AS</td>
<td>-</td>
<td>5'CCTCTGGAAAGTGGAAGGATGTAAGTGA 3’</td>
<td>1259-1235</td>
<td>HYBMTV U5 region</td>
</tr>
</tbody>
</table>
performed as described previously to amplify the cloning fragments (Jaballah et al., 2010; Rizvi et al., 2010). All clones were confirmed by sequencing.

3.3.3. In vitro transcription assay

After linearization with the Smal restriction enzyme, the wild type (SA035) or mutant (SA031-SA034, SA041-SA045, SA046, SA047 and SA051) clones were transcribed in vitro to generate the RNAs to be used in dimerization assays as previously described (Marquet et al., 1991). Following DNase treatment, RNA extraction, and RNA purification, the RNAs were concentrated using Amicon Ultra-4 10K devices (Millipore) and their concentration was determined using nanodrop (ThermoScientific) as described previously (Aktar et al., 2013). Finally, the quality of the purified RNAs was checked by denaturing polyacrylamide gel electrophoresis.

3.3.4. SHAPE methodology

SHAPE (Merino et al., 2005; Mortimer & Weeks, 2007, 2009) was performed on in vitro transcribed RNAs using benzoyl cyanide (BzCN) following the protocol described recently (Aktar et al., 2013). Briefly, RNAs were modified in 50 mM sodium cacodylate, (pH 7.5), 300 mM KCl and 5 mM MgCl₂ in the presence of 2 µg total yeast tRNA (Sigma Aldrich). The modified RNAs were reverse transcribed using two sets of AS primers: OTRs 9, 10, 11 and 12 (5’ AACAGATTTGCTTCTGCGG 3’; MMTV nt 614-633) and OTRs 13, 14, 15 and 16 (5’ AGTTTCTGCCCTTTTGAGCC 3’; MMTV nt 325-344), labelled with VIC, FAM, NED, or PET, respectively. These two sets of primers enabled the reverse transcription of the entire RNA sequence except for a few nucleotides at the 3’ end due to primer binding. The primer extension products were loaded onto an Applied Biosystems 3130xl genetic analyzer and the electropherograms were
analysed with the SHAPEfinder program (Vasa et al., 2008; Wilkinson et al., 2008). The SHAPE reactivities were determined and normalized as described earlier (Aktar et al., 2013; Vasa et al., 2008). These normalized SHAPE reactivity data were used as constraints to validate the secondary structure of the MMTV packaging signal with RNAstructure (version 5.3) (Reuter & Mathews, 2010) and also to determine the effect of pal mutations on the overall structure.

3.3.5. *In vitro* dimerization assay

*In vitro* dimerization assays of wild type and mutant MMTV RNAs were performed as described previously (Aktar et al., 2013; Marquet et al., 1991). Briefly, RNAs (300nM) were heated at 90 °C and incubated in ice for 2 minutes, incubated at 37 °C for 30 minutes in dimer (50 mM sodium cacodylate, pH 7.5; 300 mM KCl; 5 mM MgCl₂) or monomer (50 mM sodium cacodylate, pH 7.5; 40 mM KCl; 0.1 mM MgCl₂) buffer, and analysed by electrophoresis in TBM (50 mM Tris base, 45 mM boric acid, 0.1mM MgCl₂) or TB (50 mM Tris base, 45 mM boric acid, 1 mM EDTA) agarose gel at 4 °C or 20 °C, respectively. Results for each pal mutant were expressed as dimerization relative to the wild type values.

3.3.6. *In vivo* genetic complementation assay

The *in vivo* genetic *trans*-complementation assay used two expression plasmids, JA10 and MD.G, to provide the structural and enzymatic genes from MMTV Gag/Pol and vesicular stomatitis virus (VSV-G) envelope regions, respectively (Naldini et al., 1996; Rizvi et al., 2009). A third plasmid (MMTV sub-genomic transfer vector- DA024) containing minimum *cis*-acting sequences required for RNA packaging, reverse transcription, and integration, served as a source of the packageable vector RNA (Fig. 3.1; Rizvi et al., 2009). The pal I and
FIGURE 3.1. Design and rationale of three plasmid trans complementation assay for studying mouse mammary tumor virus (MMTV) gRNA packaging and propagation developed earlier (Rizvi et al., 2009). The pseudotyped particles produced following transfection of human embryonic kidney (HEK) 293T cells with MMTV Gag/Pol expression plasmid (JA10) and VSV-G envelope expression plasmid (MD.G) allows packaging of the transfer vector RNA by virtue of the presence of the packaging signal (ψ) on the MMTV transfer vector (DA024). A luciferase expressing control plasmid to monitor transfection efficiencies is also co-transfected into the producer cells. Virus particles produced following transfection contain the packaged transfer vector RNA (DA024) and are used to quantitate RNA packaging by real time PCR or to infect target cells to study RNA propagation. Infected target cells are selected with media containing hygromycin B. Cells containing the integrated transfer vector (DA024) proviral DNA form hygromycin resistant colonies. The number of the resulting hygromycin resistant colonies (colony forming unit/ml; CFU/ml) correlates with the amount of RNA being packaged. In parallel, transfected producer cells are harvested to isolate cytoplasmic and nuclear RNA fractions to ensure the export of the transfer vector RNA and the integrity of the nuclear membrane during RNA fractionation.
Transfect cells for virus production

Producer Cells

Helper-free viral stock

Target Cells

Infect target cells

Select for hygromycin resistant colonies

The number of colonies that form is directly proportional to the infectious virus titer

Gag/Pol Packaging Construct (JA10)

Env Expression Construct (MD.G)

Transfer Vector (DA024)

Real time PCR

RNA Fractionation

CTE

Poly(A)n

env

pol

Δgag

SV

Hyg'

CTE

CTE

Poly(A)n

Makes virus particle

Makes packageable RNA

Real time PCR

Real time PCR
pal II mutations introduced in the T7 promoter-based vectors were inserted in the MMTV sub-genomic transfer vector DA024.

Human embryonic kidney (HEK) 293T producer cells were transfected with 2 µg of each plasmid per well in 6-well plates using a calcium phosphate transfection kit (Invitrogen, USA) to produce VSV-G-pseudotyped-MMTV virus particles containing wild type or mutated RNAs. The luciferase expression vector, pGL3 (Promega, USA), was used as an internal control to measure transfection efficiency (Jaballah et al., 2010; Mustafa et al., 2005, 2012; Rizvi et al., 2010). Supernatants containing viral particles were collected and spun down at 4000 rpm on a table top centrifuge for 10 minutes. A portion of the supernatant was used to infect the human cervical cancer cell line (HeLaT4) to determine the transduction efficiency after selection with medium containing 200 µg/ml of hygromycin B (Hyclone, USA). Cells transduced with virus particles containing the packageable transfer vector RNA express the marker $Hyg^+$ gene, resulting in colonies which are measured as colony forming units (CFU/ml). The CFU/ml values were then normalized to the transfection efficiency. The normalized propagation efficiencies were then divided by the wild type values to represent the propagation of the mutants relative to the wild type (Relative CFU/ml).

3.3.7. Nucleocytoplasmic fractionation and isolation of RNA, and cDNA preparation

Cells from the transfected cultures were fractionated into nuclear and cytoplasmic fractions as described previously (Ghazawi et al., 2006; Mustafa et al., 2005). Cellular RNA from the cytoplasmic fractions and packaged viral RNA were isolated from the pelleted viral particles using Trizol and Trizol LS regents, respectively (Invitrogen), as described earlier (Ghazawi et al., 2006; Mustafa et al., 2005). The extracted RNAs were amplified for 30 cycles by conventional PCR
using vector-specific primers OTR 672 (AS) or OTR 562 (AS) and OTR 671 (S) (Table 3.1) after DNase (Promega) treatment to ensure that there was no contaminated DNA present in the RNA samples. The PCR conditions were as follows: denaturation for 2 minutes at 94 °C, followed by 30 cycles of 45 sec at 90 °C, annealing for 45 sec at 60 °C and extension for 45 sec at 72 °C followed by incubation at 4 °C. The DNase-treated cytoplasmic and viral RNA samples were reverse transcribed into cDNA and PCR amplified to determine the quality of the cDNA samples. To determine the integrity of the nuclear membrane during cytoplasmic RNA extraction, a multiplex PCR was performed on the cDNAs prepared from the cytoplasmic RNA fractions using OTR 582 (S) and OTR 581 (AS) (Table 3.1) designed to amplify unspliced β-actin mRNA (Ghazawi et al., 2006; Mustafa et al., 2005). Amplifications of the cDNAs using primers/competimer for 18S ribosomal RNA (18S Quantum competimer control, Ambion, USA) was also performed as an ancillary control for the presence of amplifiable cDNA in the multiplex PCRs (Ghazawi et al., 2006; Mustafa et al., 2005).

3.3.8. Real Time Quantitative PCR (qPCR) for transfer vector packaging efficiency

To determine the packaging efficiency of the transfer vector RNA, quantitative qPCR was performed on the cDNA prepared from the cytoplasmic and viral RNAs as described earlier (Mustafa et al., 2012). The endogenous control for qPCR was β-actin (Human β-actin Endogenous Control assay, VIC/MGB probe, Applied Biosystems). The primers and probes for MMTV transfer vector RNAs in the FAM-based Applied Biosystems custom expression assay anneal at a 68 nt region in the MMTV U5 (nt 1192-1259) common to wild type and mutant RNAs. The relative amplification efficiencies of the MMTV and
β-actin custom expression assays was determined to be 0.0126 (Mustafa et al., 2012), thus, validating the assays for quantitative PCR analysis using the ΔΔCt method. Equal amounts of viral and cytoplasmic cDNA were tested in triplicates. The cycling conditions used were as follows: Uracil-N-glycosylase incubation 2 minutes at 50 ºC, denaturation for 10 minutes at 94 ºC, followed by 40 cycles of denaturation and annealing/extension steps at 95 ºC for 15 sec and 60 ºC for 1 minutes. The relative packaging efficiencies were derived following a modification of the method described earlier (Mustafa et al., 2012). Briefly, the raw Ct (cycle threshold) values for the MMTV transfer vector were normalized by the Ct values of the β-actin endogenous control. The normalized values were then calculated relative to the mock sample (relative quantification, RQ) that was transfected with all the plasmids, except the transfer vector plasmid forming the empty virus particles without any packaged RNA. For final relative packaging efficiency calculations, the relative viral RQ values were divided by the relative cytoplasmic RQ values, and the resulting values were normalized to the transfection efficiencies. The final values are presented relative to the wild type.
3.4. Results

3.4.1. Predicted secondary structure of the MMTV 5’ gRNA

Earlier studies have shown that the sequences necessary for optimal MMTV gRNA packaging start at R and extend into first 120 nt of \textit{gag} (Mustafa et al., 2012). Structural prediction analyses of this region using Mfold suggested that this region comprises six stem loops 1-6 (SLs1-6, Fig. 3.2A). Two stable stem loops, SL1 and SL2, were predicted in the R/U5/PBS/UTR region in addition to SL5 and SL6 in the \textit{gag} region (Fig. 3.2A). In between these stem loops, SL3 and a bifurcated SL4 were predicted in the UTR region. SL2 was found stably maintained in all predicted structures (Fig. 3.2A). A closer look at the structure revealed a presence of three palindromic sequences, a 10 nt pal I in SL2, a 6 nt pal II in bifurcated SL4, and a 13 nt pal III in SL6 (Fig. 3.2A). Additionally, a 9 nt stretch of single-stranded purines (ssPurines) was also observed in SL4. The first apical loop of the bifurcated SL4 comprised pal II, while the second apical loop contained the ssPurines (Fig. 3.2A).

3.4.2. Validation of the secondary structure by SHAPE

To experimentally validate the predicted structure of the MMTV packaging signal RNA, SHAPE was employed (Merino et al., 2005; Mortimer & Weeks, 2007, 2009). Briefly, RNA transcribed \textit{in vitro} from our wild type construct, SA035 (Fig. 3.2B), was purified and modified with BzCN as described earlier (Aktar et al., 2013). BzCN selectively acylates the 2’hydroxy group of the ribose of nucleotides in flexible (unpaired) regions. Reactivity at each nucleotide was calculated by subtraction of the reverse transcription product of unmodified RNA from that of BzCN-modified RNA (Aktar et al., 2013; Vasa et al., 2008). The SHAPE reactivity data obtained for each nucleotide from 3-6 experiments
FIGURE 3.2. Predicted and SHAPE-validated structural models of the MMTV packaging signal RNA. (A) MMTV packaging signal RNA secondary structure predicted using Mfold. mSD, major splice donor, PBS, primer binding site. (B) Schematic representation of full length MMTV genome, showing the region used to create the wild type T7 expression plasmid, SA035. (C) SHAPE-constrained RNA structure model of MMTV packaging signal. Nucleotides are color annotated as per the SHAPE reactivities key depending on their modification by BzCN. The data shown is an average of 3-6 independent experiments.
Predicted Secondary Structure of the MMTV Packaging Signal RNA

SHAPE-Validated Secondary Structure of the MMTV Packaging Signal RNA

- SHAPE reactivity:
  - 0.9
  - 0.7
  - 0.5
  - 0.3
  - No data
were then applied as pseudo-energy constraints in the structure prediction program RNAstructure in order to develop a refined RNA structural model of the sequences required for MMTV gRNA dimerization and packaging.

Consistent with the predicted structure, all the loops and bulges containing unpaired nucleotides in the MMTV packaging signal RNA showed SHAPE reactivity (Fig. 3.2C). High reactivities were observed for nucleotides forming G-U base pairs, especially at the end of helices (G120, U320, G193, U294, G324, U356, U357, U358, U415, Fig. 3.2C). Some base paired nucleotides (G21, C25, G84) located on either side of a bulge or loop also showed certain amount of reactivity, as expected. The refined structural model corroborated well with the predicted structure, supporting the existence of major structural motifs (SL1-SL6; Fig. 3.2C). Even though a high conformity was observed between the predicted and the SHAPE-validated structural models, some conspicuous dissimilarity could be noted. In the SHAPE validated structure, SL1 and SL6 were shorter, resulting in the formation of SL7 which was not observed in the predicted structure (Fig. 3.2C). In addition, the sequences 5’ to the shortened SL6 form a new SL, which is different from SL5 (Fig. 3.2B) and therefore was labeled as SL5’ (Fig. 3.2C). Furthermore, the SHAPE-validated structure harbored a long-range interaction (LRI) involving complementary sequences from U5 and Gag (U5-Gag LRI; Fig. 3.2C) which was not predicted by Mfold. This U5-Gag LRI is made up of three short helices separated by a bulge and an internal loop (Fig. 3.2C).

3.4.3. The SHAPE-validated RNA structural model is supported by phylogeny

Sequence alignment of eight strains of MMTV (sequence from +1R-120 nt gag) revealed a high degree of conservation relative to the wild type mtv-1 (AF228550.1) strain, especially of sequences of the major structural motifs such
as U5-Gag LRI, ssPurines, pal II, and PBS (Fig. 3.3A) once superimposed on the SHAPE-validated structure (Fig. 3.3B). Only a few nucleotides showed conservation in less than 6 strains (Fig. 3.3B). A phylogenetic tree of MMTV packaging signal RNA, drawn using CLC Sequence Viewer software (Version 6.8.2; http://www.clcbio.com), further substantiated the high level of sequence conservation and revealed that these strains share a close evolutionary history (Fig. 3.3C). Finally, a consensus structure of this region using RNAalifold was generated (Fig. 3.4). Consistent with the SHAPE-validated structure (Fig. 3.2C), consensus RNA secondary structure predictions (Fig. 3.4) revealed that the major structural motifs were consistently maintained. In addition, the consensus structures predicted by RNAalifold also supported the existence of SHAPE-validated short SL6 and SL5’ over the Mfold predicted long SL6 and SL5 motifs (compare Figs. 3.2A, 3.2C, and 3.4). The consensus structures also revealed the existence of LRI which is very similar to the one observed in the SHAPE-validated structure (Figs. 3.2C, 3.4).

3.4.4. Identification of the pal sequence mediating MMTV gRNA dimerization

The sequence of the MMTV RNA packaging signal revealed three pal sequences: pal I, a 10 nt base paired sequence in the SL2 stem, pal II, a 6 nt sequence present in a loop of the bifurcated SL4, pal III, forming the apical part of SL6 (13 nt) one of these could possibly function as a DIS (Fig. 3.2A, 3.2C). To ascertain which pal sequence may function as the DIS for MMTV gRNA dimerization, dimerization of three RNAs with different truncated 3’ ends containing different number of pals was compared (Fig. 3.5). The longest RNA (712 nt) encompassed all the three pals, the intermediate 371 nt RNA contained only the first two pals, while the shortest RNA included only pal I (Fig. 3.5A).
**FIGURE 3.3.** Phylogenetic conservation of the sequences of major structural motifs of MMTV packaging signal RNA in different strains.  

(A) Clustal Omega sequence alignment of eight different strains of MMTV packaging signal RNA. Sequences of the major structural motifs are highlighted in different colors, whereas the 11 nt pal sequence within PBS is boxed. The accession numbers for the MMTV strains are AF228550.1 (*mtv*-1; Hook et al., 2000), L37517.1 (*mtv*-6; Cho et al., 1995), AF033807.1 (Petropoulos, 1997), M15122.1 (BR6; Moore et al., 1987), D16249.1 (JYG; Moore et al., 1987), AF228551.1 (C3H/HeJ; Hook et al., 2000), X00018.1 (GR; Fasel et al., 1983) and AF228552.1 (C3H; Hook et al., 2000).  

(B) Sequence conservation is superimposed on the SHAPE validated structure. The nucleotides are color-coded based on their conservation level ranging from <75 to 100%.  

(C) Phylogenetic tree of the eight different strains of MMTV packaging signal RNA. Neighbor joining method was used for creating the phylogenetic tree. The horizontal lines are branches and represent evolutionary lineages changing over time. The units of branch length are nucleotide substitutions given as fractional change (the number of changes or 'substitutions' divided by the length of the sequence). Figure continues on the following page.
FIGURE 3.4. Consensus RNA secondary structure prediction of MMTV packaging signal RNA using sequences from eight different strains.

Accession numbers of MMTV strains are the same as mentioned in figure 2A. The sequences of the 8 MMTV strains were aligned using Clustal Omega and the aligned sequences were used as input in the RNAalifold server (Hofacker et al., 2002) to predict the consensus or “conservation annotation” structure. The number/color scale on the x-axis indicates base pairing conservation with 1 (red color) representing completely conserved base pairing to 6 (violet color) representing 6 different types of base pairings at that position among the various strains (see Materials and Methods for details). The y-axis represents the incompatibility score based on non-Watson base pairing within each strain. The number (0-2) or color gradation (dark to very pale) represents increasing number of strains showing incompatible (non-Watson) base pairing within the structure of the various strains. Thus, a dark red color denotes “highest” conservation of base pairings, while a light violet color represents “least” conservation of base pairing among different strains of MMTVs.
FIGURE 3.5. Contribution of the pal sequences to MMTV gRNA dimerization. (A) Schematic representations of the predicted RNA secondary structures of the three truncated RNAs (712 nt, 321 nt, and 265 nt). (B) Summary of the in vitro dimerization data analyzed on TBM or TB gels electrophoresed at 4 °C and at room temperature, respectively, showing the percentage of dimerization of the truncated RNAs. (C) Representative RNA dimerization gels. M: incubation in monomer buffer; D: incubation in dimer buffer. The circles show the loss of dimerization of the 265 nt RNA in TB gels. Wild type HIV-1 RNA (615 nt) was used as a positive control for dimerization.
A

B

<table>
<thead>
<tr>
<th></th>
<th>712 nt</th>
<th>321 nt</th>
<th>265 nt</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBM</td>
<td>78.24%</td>
<td>92.83%</td>
<td>54.93%</td>
</tr>
<tr>
<td>TB</td>
<td>37.65%</td>
<td>57.68%</td>
<td>14.06%</td>
</tr>
</tbody>
</table>

C
RNA samples were incubated in dimer or monomer buffer and analyzed on 1% TBM or TB agarose gels at 4 °C or 20 °C, respectively. The TBM gels stabilized weak dimer conformers, while only stable dimers survive electrophoresis through TB gels. Our results revealed that the longest RNA as well as the intermediate RNA dimerized efficiently, whereas the shortest RNA barely dimerized (Fig. 3.5B, C). These results suggest that pal II may play a crucial role in MMTV gRNA dimerization. Consistently, pal II is the only pal that is observed to adopt a hairpin conformation with a canonical “GC” dyad (Fig. 3.2A, C).

3.4.5. Mutational analysis of pal I

To further test the role of pal I (if any) in dimerization and packaging, a series of mutations were introduced and tested employing a combination of genetic and biochemical approaches (Fig. 3.6A). These mutations included complete deletion of 26 nt of the apical stem loop of SL2 containing pal I, as well as simultaneous substitution of the deleted sequences with a UUCG stable tetraloop in order to maintain a stem loop structure without pal I. It is important to mention that the attachment (att) sites needed for integration of the provirus were not affected in pal I mutants, thus allowing us to test the same mutations in both the in vitro dimerization as well as in vivo packaging and propagation assays (Rizvi et al., 2009). None of the deletion and/or deletion/substitution mutants showed any effects on dimerization when compared to the wild type clone SA035 (Fig. 3.6B). The percentage dimerization of the wild type, SA035 clone and mutant clones (SA031-SA034) was observed to be 81-96% in TBM gels and 39-69% in TB gels (data not shown). Thus, these experiments indicated that pal I is not the DIS, which is in agreement with the analysis of the truncated RNAs in the in vitro dimerization assay (Fig. 3.5).
**FIGURE 3.6.** Role of pal I in MMTV gRNA dimerization, packaging, and propagation. (A) Description of the pal I mutants. The wild type pal I sequence is shown in blue and the mutations are depicted in red. (B) Typical RNA dimerization TBM and TB gels of wild type and mutant MMTV RNAs. M: monomer lane or monomer conformer; D: dimer lane or dimer conformer for each sample. (C) Transfection efficiencies of the mutants and wild type transfer vectors that were used to normalize the packaging efficiency. LUC: Luciferase activity. (D) PCR amplifications of the DNase treated cytoplasmic (panel i) and viral (panel ii) RNAs using virus specific primers. In the third panel (iii), amplification was conducted on the cDNAs obtained from cytoplasmic RNAs using primers that amplify unspliced β-actin mRNA. Multiplex amplifications were conducted in the presence of primers/competimer for 18S ribosomal RNA. The fourth panel (iv) shows PCR of cytoplasmic cDNA using primers that amplify spliced β-actin mRNA. (E) Relative packaging efficiency (RPE) of transfer vector RNAs. (F) Relative hygromycin resistance (Hyg') colony forming unit per ml (CFU/ml) for mutant transfer vectors reflecting the relative RNA propagation efficiencies. In (C), (E), and (F), the histograms represent data from at least three independent experiments (±SD). Figure continues on the following page.
A

<table>
<thead>
<tr>
<th>Vectors for \textit{in vitro} Dimerization</th>
<th>Sub-genomic Transfer Vectors for Packaging and Propagation</th>
<th>Description of the Introduced Mutations</th>
<th>Sequence of the Wild Type and \textit{pal} Mutants</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA035</td>
<td>DA024</td>
<td>Wild type SL2 stem loop with \textit{pal I}</td>
<td>5' ACCCCCCGUGACCCUCAGGUCGGCCGAC 3'</td>
</tr>
<tr>
<td>SA031</td>
<td>SA031P</td>
<td>Deletion of wild type sequence containing \textit{pal I}</td>
<td>5' A---------------------------------------------------------------3'</td>
</tr>
<tr>
<td>SA033</td>
<td>SA033P</td>
<td>Deletion of wild type sequence containing \textit{pal I}</td>
<td>5'---------------------------------------------------------------C 3'</td>
</tr>
<tr>
<td>SA032</td>
<td>SA032P</td>
<td>Substitution of wild type sequence \textit{pal I} with UUCG</td>
<td>5' A-------------------------UUCG----------------------3'</td>
</tr>
<tr>
<td>SA034</td>
<td>SA034P</td>
<td>Substitution of wild type sequence with UUCG</td>
<td>5'--------------------------UUCG----------------------C 3'</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th></th>
<th>SA031</th>
<th>SA032</th>
<th>SA033</th>
<th>SA034</th>
<th>SA035</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>D</td>
<td>M</td>
<td>D</td>
<td>M</td>
<td>D</td>
</tr>
</tbody>
</table>

1% \textit{TB} gel

1% \textit{TB} gel
To evaluate the functional impact of mutations in pal I, we tested these mutants in an in vivo packaging and propagation assay. Wild type and mutant viral particles were produced in HEK 293T cells that were used to: 1) measure the packaged viral RNA using qPCR, and 2) assess the propagation of the packaged RNA in the target cells which was monitored by the appearance of Hygromycin resistant colonies. First, the transfection efficiency of the cultures containing mutant transfer vectors was checked. The transfection efficiencies from at least three independent experiments were observed to be within 2 folds of each other, suggesting that all cultures were efficiently transfected (Fig. 3.6C). These transfection efficiencies were taken into account to normalize and calculate the packaging efficiency and propagation for each mutant relative to the wild type. To determine the RNA packaging efficiency, cDNAs were prepared from cytoplasmic and pelleted virion RNAs. Before making the cDNAs these RNAs were subjected to DNase treatment to ensure that there was no contaminating plasmid DNA, followed by PCR using virus-specific primers. The lack of a positive signal indicated that the DNA contamination in our RNA preparations was below the detection levels (Fig. 3.6D, panels i and ii). After having confirmed this, RNA preparations were reverse transcribed and cDNAs were prepared. Since the relative packaging efficiency data is expressed in relation to the efficiently and stably expressed viral RNAs that are exported from the nucleus to the cytoplasm, the integrity of the cytoplasmic RNA fraction was monitored by checking for the absence of unspliced β-actin mRNA by RT-PCR (Fig. 3.6D, panel iii). To ensure that each cytoplasmic RNA fraction contained amplifiable cDNAs, the unspliced β-actin PCRs were conducted in the presence of primers/competimer for 18S ribosomal RNAs as an internal control (Jaballah et
al., 2010; Mustafa et al., 2012; Rizvi et al., 2010). In a separate amplification reaction, the same cytoplasmic cDNA samples were amplified using spliced β-actin primers (Fig. 3.6D, panel iv). Figure 3.6D (panels iii, and iv) shows that while 18S ribosomal RNA and spliced β-actin mRNA were amplifiable, unspliced β-actin mRNA was not. These results indicated that the fractionation technique employed was robust and that the nuclear membrane integrity was maintained during the fractionation process.

Following normalization of the RNA packaged into the virus particles with the cytoplasmic transfer vector RNA expression and to transfection efficiencies, the in vivo gRNA packaging data revealed that none of the pal I mutations significantly affected gRNA packaging (Fig. 3.6E). This corroborated well with the dimerization data for the same mutants which showed no significant dimerization defect (Fig. 3.6B). Interestingly, when viral particles containing RNAs from these mutants were used to transduce target cells with a hygromycin resistance gene, a drastic reduction of RNA propagation (3-20 fold reduction of the relative CFU/ml, \( P < 0.05 \)) was observed for all mutants (Fig. 3.6F). Lack of RNA propagation of the packaged mutant transfer vector RNAs suggests that the mutations impaired reverse transcription or/and integration since our readout assay is dependent on successful completion of these steps in retroviral life cycle. Consistent with this hypothesis, mutations in close vicinity of the PBS (as has been the case for these pal I mutants) have been shown to greatly impinge reverse transcription (Mikkelsen et al., 1996; Oh et al., 2008; Paillart et al., 1996a).

3.4.6. Pal II functions as the dimerization initiation site (DIS)

To test if pal II modulates MMTV gRNA dimerization and packaging, a series of similar mutations was introduced including a complete deletion of pal II
as well as substitution with a stable tetraloop sequence, in order to maintain the overall RNA secondary structure of this region (Fig. 3.7A). Complete deletion of pal II revealed a significant ~two-fold decrease in RNA dimerization, both in TBM and TB gels (Fig. 3.7B, C, compare mutant SA042 with SA035, \( P<0.001 \)). Deletion of pal II with simultaneous substitution with a stable tetraloop had the same effect on RNA dimerization (Fig. 3.7B, C, mutant SA41; 1.85-2.40-fold reduction in relative RNA dimerization; \( P<0.001 \)).

These results indicated that although pal II deletion mutants could reduce RNA dimerization, they could not completely abrogate it. This suggested the presence of another sequence that is able to facilitate dimerization in the absence of pal II, though not to the wild type level. Therefore, the pal II deletion mutant (SA042) sequence was folded using the RNAstructure software, which revealed that in the absence of pal II, the MMTV RNA packaging signal could potentially dimerize \textit{via} the PBS (Fig. 3.8B), which contains two adjacent pals (5’ \textit{CAGCUGGCGCC} 3’; the first pal is in italics and the second pal is underlined. To determine if this was the case, a mutant (SA051) was generated containing deletion of the 11 nt pal sequence within the PBS. Along the same lines, a double mutant (SA046) was created containing the deletion of pal II sequence as well as the 11 nt pal within the PBS ((Fig. 3.7A). Deletion of the 11 nt pal in the PBS alone resulted in 3 to 6-fold decreased RNA dimerization (SA051; (Fig. 3.7B, C). Deletion of both sequences (SA046) resulted in almost complete abrogation of dimerization in TBM conditions and to a five-fold decrease in TB (Fig. 3.7B, C). These results revealed an additive and/or synergistic role of pal II and the pal sequence within PBS in mediating MMTV gRNA dimerization.
FIGURE 3.7. Role of pal II in MMTV gRNA dimerization, RNA packaging and viral propagation.  (A) Description of the pal II mutants.  
(B) Typical RNA dimerization in TBM and TB gels of wild type and mutant MMTV RNAs.  M: monomer lane or monomer conformer; D: dimer lane or dimer conformer.  (C) Quantification of the relative RNA dimerization.  (D) Transfection efficiencies of the mutants and wild type transfer vectors that were used to normalize the packaging efficiency.  LUC: Luciferase activity.  (E) PCR amplifications of the DNase treated cytoplasmic (panel i) and viral (panel ii) RNAs using virus specific primers.  In the third panel (iii), amplification was conducted on the cDNAs obtained from cytoplasmic RNAs using primers that amplify unspliced β-actin mRNA.  Multiplex amplifications were conducted in the presence of primers/competimer for 18S ribosomal RNA.  The fourth panel (iv) shows PCR of cytoplasmic cDNA using primers that amplify spliced β-actin mRNA.  (F)  Relative packaging efficiency of transfer vector RNAs.  (G) RNA propagation efficiency expressed as Hyg' colony forming unit per ml (CFU/ml).  In (C), (D), (F), and (G) the histograms represent data from at least three independent experiments (±SD).  Figure continues on the following page.
### Description of the Introduced Mutations

<table>
<thead>
<tr>
<th>Vectors for <strong>in vitro</strong> Dimerization</th>
<th>Sub-genomic Transfer Vectors for Packaging and Propagation</th>
<th>Description of the Introduced Mutations</th>
<th>Sequence of the Wild Type and pal II Mutants</th>
<th>Presence or Absence of an Alternate 11 nt pal within PBS 5'CAGCUGGCGCC 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA035</td>
<td>DA024</td>
<td>Wild type SL4 stem loop sequence with pal II</td>
<td>5'GGGA G [CUGCAG] UCCC 3'</td>
<td>+</td>
</tr>
<tr>
<td>SA041</td>
<td>SA041P</td>
<td>Substitution of wild type pal II sequence with stable loop GUAA</td>
<td>5'GGGA GUAA UCCC 3'</td>
<td>+</td>
</tr>
<tr>
<td>SA042</td>
<td>SA042P</td>
<td>Deletion of wild type sequence containing pal II</td>
<td>5'---...---...---...---... 3'</td>
<td>+</td>
</tr>
<tr>
<td>SA046</td>
<td>SA046P</td>
<td>Deletion of an alternate 11 nt pal within PBS and wild type sequence containing pal II</td>
<td>5'---...---...---...---... 3'</td>
<td>-</td>
</tr>
<tr>
<td>SA051</td>
<td>SA051P</td>
<td>Deletion of only 11 nt PBS pal</td>
<td>5'GGGA G [CUGCAG] UCCC 3'</td>
<td>-</td>
</tr>
</tbody>
</table>

### Relative Dimerization

**A**

- **1% TB gel**
- **1% TBM gel**

**B**

- **SA035**
- **SA041**
- **SA042**
- **SA046**
- **SA051**

**C**

- **Graph showing relative dimerization**

<table>
<thead>
<tr>
<th>Vectors</th>
<th>Relative Dimerization</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA035</td>
<td>1.00</td>
</tr>
<tr>
<td>SA041</td>
<td>0.53</td>
</tr>
<tr>
<td>SA042</td>
<td>0.54</td>
</tr>
<tr>
<td>SA046</td>
<td>0.01</td>
</tr>
<tr>
<td>SA051</td>
<td>0.32</td>
</tr>
<tr>
<td>TBM</td>
<td>1.00</td>
</tr>
<tr>
<td>TB</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.17</td>
</tr>
</tbody>
</table>
FIGURE 3.8. Predicted homodimer structures of (A) MMTV wild type (SA035), (B) pal II deletion (SA042), (C) deletion of the pal sequence within PBS (SA051) and (D) double deletion of both pal II and pal sequence within PBS (SA046). The dimers were predicted using RNAstructure and the contact point (point of dimerization) of the sequences between two RNAs leading to dimerization are boxed. PBS, primer binding site; ssPurines, single stranded purines.
A
SA035 (wildtype)
homodimer
\( \Delta G = -311.8 \text{ kcal/mol} \)

B
SA042 (Δpal II)
homodimer
\( \Delta G = -313.0 \text{ kcal/mol} \)

C
SA051 (Δpal within PBS)
homodimer
\( \Delta G = -309.8 \text{ kcal/mol} \)

D
SA046 (Δpal II and Δpal within PBS)
homodimer
\( \Delta G = -309.7 \text{ kcal/mol} \)
To ascertain whether the primary sequence or the palindromic nature of pal II was important for MMTV gRNA dimerization, a few additional mutants were generated. The wild type pal II was substituted with the HIV-1 6 nt pal or with the HIV-1 pal flanked by additional purines on either side (Fig. 3.9A, mutants SA043 and SA047; respectively). Two additional mutants were also created in which the pal II sequence was substituted by non-palindromic trans-complementary sequences (Fig. 3.9A, mutants SA044 and SA045). In all these mutants, the pal sequence within the PBS was maintained. Consistent with the effect of the purines flanking the pal sequence on HIV-1 RNA dimerization (Paillart et al., 1997), replacing pal II by the HIV-1 pal (SA043) reduced MMTV RNA dimerization by ~two-fold, while replacing it by the HIV-1 pal containing flanking purines (SA047) preserved MMTV RNA dimerization to wild type levels (Fig. 3.9B, C). The trans-complementary mutants SA044 and SA045 were incubated either separately or together, allowing them to interact to initiate dimerization due to the trans-complementary nature of the sequences. When incubated separately, the mutants displayed a ~two-fold decrease in dimerization (Fig. 3.9B, C), as expected similar to other pal II substitution and deletion mutants (Figs. 3.7B, C, 3.9B, C). Surprisingly, RNA dimerization was not restored when these mutants were co-incubated (Fig. 3.9B, C), in sharp contrast with other retroviral systems in which dimerization has been restored to wild type levels in the mutants containing trans-complementary sequences (Aktar et al., 2013; Chen J. et al., 2009; Jossinet et al., 2001; Paillart et al., 1996b).

3.4.7. Pal II regulates MMTV gRNA packaging and propagation

Next, the RNA packaging and propagation efficiencies of sub-genomic transfer vectors containing pal II mutations was analyzed by transfecting the HEK
FIGURE 3.9. Effects of pal II substitution mutations on MMTV gRNA dimerization, RNA packaging and viral propagation. (A) Description of the pal II substitution mutants. (B) Typical RNA dimerization in TBM and TB gels of wild type and mutant MMTV RNAs. The last two lanes correspond to the co-incubation of *trans*-complementary mutants SA044 and SA045. M: monomer lane or monomer conformer; D: dimer lane or dimer conformer. (C) Quantification of the relative RNA dimerization. (D) PCR amplifications of the DNase treated cytoplasmic (panel i) and viral (panel ii) RNAs using virus specific primers. In the third panel (iii), amplification was conducted on the cDNAs obtained from cytoplasmic RNAs using primers that amplify unspliced β-actin mRNA. Multiplex amplifications were conducted in the presence of primers/competimer for 18S ribosomal RNA. The fourth panel (iv) shows PCR of cytoplasmic cDNA using primers that amplify spliced β-actin mRNA. (E) Transfection efficiencies of the mutants and wild type transfer vectors that were used to normalize the packaging efficiency. LUC: Luciferase activity. (F) Relative packaging efficiency of transfer vector RNAs. (G) RNA propagation efficiency expressed as Hyg<sup>+</sup> colony forming unit per ml (CFU/ml). In (C), (E), (F), and (G) the histograms represent data from at least three independent experiments (±SD). Figure continues on the following page.
### Vectors for in vitro Dimerization

<table>
<thead>
<tr>
<th>Vectors for in vitro Dimerization</th>
<th>Sub-genomic Transfer Vectors for Packaging and Propagation</th>
<th>Description of the Introduced Mutations</th>
<th>Sequence of the Wild Type and pal II Mutants</th>
<th>Presence or Absence of an Alternate 11 nt pal within PBS 5' CAGCUGGC GCC 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA035</td>
<td>DA024</td>
<td>Wild type SL4 stem loop sequence with pal II</td>
<td>5' GGGAG CUGCAG UCCC 3'</td>
<td>+</td>
</tr>
<tr>
<td>SA043</td>
<td>SA043P</td>
<td>Substitution of 6 nt of the wild type pal II sequence with HIV-1 pal</td>
<td>5' GGGAG GCGCGC UCCC 3'</td>
<td>+</td>
</tr>
<tr>
<td>SA044</td>
<td>SA044P</td>
<td>Substitution of wild type pal II sequence with non palindromic sequence</td>
<td>5' GGGAG ACGCAC UCCC 3'</td>
<td>+</td>
</tr>
<tr>
<td>SA045</td>
<td>SA045P</td>
<td>Substitution of wild type pal II sequence with non palindromic sequence</td>
<td>5' GGGAG GUGCGU UCCC 3'</td>
<td>+</td>
</tr>
<tr>
<td>SA047</td>
<td>SA047P</td>
<td>Substitution of 6 nt of the wild type pal II sequence with HIV-1 pal with additional purines</td>
<td>5' GGGAG GAGCGCGCA UCCC 3'</td>
<td>+</td>
</tr>
</tbody>
</table>

### B

```
<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA035</td>
<td>[image]</td>
<td>[image]</td>
</tr>
<tr>
<td>SA043</td>
<td>[image]</td>
<td>[image]</td>
</tr>
<tr>
<td>SA044</td>
<td>[image]</td>
<td>[image]</td>
</tr>
<tr>
<td>SA045</td>
<td>[image]</td>
<td>[image]</td>
</tr>
<tr>
<td>SA047</td>
<td>[image]</td>
<td>[image]</td>
</tr>
<tr>
<td>SA044+SA045</td>
<td>[image]</td>
<td>[image]</td>
</tr>
</tbody>
</table>
```

### C

```
<table>
<thead>
<tr>
<th></th>
<th>Relative Dimerization</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA035</td>
<td>1.00 0.58 0.51 0.55 1.06 0.50</td>
</tr>
<tr>
<td>SA043</td>
<td>1.00 0.50 0.56 0.54 1.24 0.57</td>
</tr>
</tbody>
</table>
```

1% TBM gel

1% TB gel
293T cells along with the packaging and VSV-G envelope expression constructs. As described previously for pal I mutants, data from several experiments showed that pal II mutants’ transfection efficiencies were within two folds and that the transfer vector RNAs were expressed efficiently (Figs. 3.7D, E, 3.9D, E). When similar amounts of virions were used to isolate packaged RNA, all of the mutants tested except for SA051P and SA043P, were found to be severely impaired for gRNA packaging (27-150 fold reduction, \( P < 0.001 \), Figs. 3.7F, 3.9F). Although packaged, the relative packaging efficiency of SA051P and SA043P mutants were reduced by 9- and 2-fold, respectively, when compared to the wild type (Figs. 3.7F, 3.9F). Propagation of these mutants, as measured by counting the number of hygromycin resistant colonies (CFU/ml) in the infected cultures (Figs. 3.7G, 3.9G), corroborated well with the RNA packaging data, except for mutant SA051P. Lack of propagation of this mutant can be explained by the deletion in the PBS since our assay requires the packaged RNA to be successfully reverse transcribed (initiated by an intact PBS) before integration to allow expression of the hygromycin resistant gene. Altogether, the dimerization, packaging, and propagation results suggest that the pal II hairpin plays an important role in both MMTV gRNA dimerization and packaging; however, RNA dimerization is not sufficient to ensure efficient packaging and propagation. The case of mutant SA047P, containing substitution of pal II with HIV-1 pal along with flanking purines, is particularly intriguing as its RNA packaging and propagation are abolished (Fig. 3.9F, G), despite displaying wild type RNA dimerization levels. These results could indicate that the pal II hairpin is also involved in RNA-protein interactions during genome encapsidation.
3.4.8. SHAPE analyses of pal II and PBS pal mutants support their role in MMTV gRNA dimerization

In order to test whether the dimerization, packaging and propagation data could be explained by effects of the mutations in pal II on the MMTV RNA secondary structure, we performed SHAPE analyses on all the pal II mutants in dimer buffer. Figures 3.10 and 3.11 focus on the SL4 region (~ nt 250-303, encompassing pal II) and the PBS region, respectively. In the wild type RNA (SA035) and in the substitution mutant with the HIV-1 pal containing flanking purines (SA047), the two pals displayed limited or no SHAPE reactivity, consistent with their normal role in RNA dimerization (Fig. 3.10A). Accordingly, pal II substitution mutants with decreased RNA dimerization displayed high reactivity of at least one nucleotide in the mutated region (SA041, SA044, SA045, SA043, Fig. 3.10A). The ssPurines sequence was highly reactive in all RNAs (Fig. 3.10A). Several mutants (SA041, SA044, SA043, and SA047) preserved the bifurcated stem loop structure of SL4 (Fig. 3.10B). Surprisingly, SA045 adopted a totally different structure in which the mutated sequence is mostly base-paired. Misfolding of this RNA likely explains why the trans-complementary mutants SA044 and SA045 were unable to form heterodimers (Fig. 3.9B, C). In these two pal II deletion mutants (SA042 and SA046), SL4 adopted a single hairpin conformation in which most of the nucleotides that formed the bifurcating loop in the wild type structure were base-paired.

Since we observed that the PBS pal played some role in MMTV gRNA dimerization (Fig. 3.7B, C), the structure of the SL2 domain which contains the PBS was analysed in the wild type and mutant RNAs (Fig. 3.11A). Although the PBS is predicted to be unpaired in monomeric RNA, the pal PBS was only weakly reactive, especially the second pal (5’ GGCGCC 3’), consistent with a role in
FIGURE 3.10. Structural analysis of the SL4 domain of the pal II and PBS pal mutants. (A) SHAPE reactivity profile of the wild type and mutant RNAs. Blue boxes indicate pal II, ssPurines, and deletion/substitutions in pal II. Data are an average of at least three independent experiments. (B) Structure of the wild type and mutant SL4 domain in the SHAPE-constrained RNA model by RNAstructure. (C) The RNAstructure predicted site of interaction (point of dimerization) between two wild type gRNAs. The central 6 nt of pal II are shown in blue. The complete predicted structures of the wild type SA035 homodimer is shown in figure 3.8A. Figure continues on the following page.
**Point of Dimerization**

Native pal

pal II

5' CUGCAG 3'

ssPurines

ssPurines

5' CUGCAG 3'

(MMTV Wild Type SA035)

Substitution of pal helix loop with a stable tetraloop "GUAA"

Loss of "native" pal helix loop

Deletion of PBS pal sequence and pal II resulting in loss of "native" pal helix loop

**SHAPE Reactivity**

- 0.9
- 0.7
- 0.5
- 0.3
No data

B

C

A

SA035

SA041

SA042

SA046

**Point of Dimerization**

Native pal

pal II

5' CUGCAG 3'

ssPurines

ssPurines

5' CUGCAG 3'
Loss of "native" pal helix loop

Maintenance of pal helix loop containing HIV-1 pal and flaking purines

SHAPE reactivity

- 0.9
- 0.7
- 0.5
- 0.3
No data
FIGURE 3.11. Structural analysis of the SL2 domain of the pal II and PBS pal mutants. (A) The SHAPE reactivity of the wild type and mutant RNAs are presented as color-coded histograms. The blue boxes highlight the SL2 loop sequences and the 11 nt PBS pal. The dashes on the x-axis of SA046 SHAPE reactivity profile indicate the site of deletion of the PBS pal sequence. (B) The SHAPE-validated wild type SA035 SL2 domain structure and the nucleotides are color-coded according to SHAPE reactivity. (C) The mutant SA046 SL2 structural motif is shown after being constrained by SHAPE data.
SHAPE Reactivity

SA042 SL2

Deletion of 11 nt pal within PBS

SA043 SL2

(pal II substitution with HIV-1 pal; 5' GCGCGC 3')

SA046 SL2

(Double deletion of pal II and 11 nt pal within PBS)

SA047 SL2

(pal II substitution with HIV-1 pal with flanking purines; 5' AAGCGCGCA 3')

SL2 Loop

11 nt pal sequence within PBS

SA046

5' AAGCGCGCA 3'

B

C

SA035 SL2

(Wild type)

SA042 (Deletion of pal II helix loop)

SA043 (pal II substitution with HIV-1 pal; 5' GCGCGC 3')

SA046 (Double deletion of pal II and 11 nt pal within PBS)

SA047 (pal II substitution with HIV-1 pal with flanking purines; 5' AAGCGCGCA 3')
RNA dimerization. The reactivity of this region was not affected by deletion or substitutions in pal II (Fig. 3.11A, mutants SA042, SA043 and SA047), suggesting the second pal within the PBS could perhaps mediate the residual RNA dimerization observed with these mutants (Figs. 3.7, 3.9). In addition, deletion of the 11 nt PBS pal (SA046) did not perturb the global folding of the SL2 domain when compared to the wild type SA035 (compare Fig. 3.11B, C), as the SL2 apical loop is reactive in all RNAs. Thus, the diminished dimerization of mutant SA051 (Fig. 3.7C) could not be attributed to RNA misfolding, but due to the direct deletion of the PBS pal.

3.4.9. Structural prediction of pal II mutant dimers support dimerization data

To establish a structural basis of the in vitro dimerization results, minimal free energy models of the wild type and mutant MMTV RNA dimers were predicted using RNAstructure. This software predicted that wild type RNA dimerizes via the pal II sequence (Figs. 3.8A, 3.10C), while deletion of pal II (SA042) induces dimerization via the PBS pal (Fig. 3.8B), in agreement with our RNA dimerization and SHAPE data (Figs. 3.7, 3.9, 3.10, and 3.11). The predicted structure of SA051 (deletion of 11 nt pal sequence within PBS), revealed that the native structure (like wild type homodimer) was restored except for the loss of the PBS loop (Fig. 3.8C) which could in part explain its ability to dimerize though much less efficiently when compared to the wild type (Fig. 3.7B, C). The mutant with simultaneous deletions of pal II and the PBS pal (SA046) was predicted to dimerize via pal III (Fig. 3.8D), but according to our experimental data, this mutant RNA dimerized weakly (Fig. 3.7B, C). Substitution mutants with non-palindromic sequences in place of pal II (SA044 and SA045) were predicted to dimerize via the PBS pal (Fig. 3.12A, B). Surprisingly and in good agreement
FIGURE 3.12. Predicted homodimer (A-B) and heterodimer (C) structures of MMTV pal II trans-complementary mutants (SA044 and SA045). The dimers were predicted using RNAstructure and the contact point of the sequences (point of dimerization) between two RNAs leading to dimerization are boxed. PBS, primer binding site; ssPurines, single stranded purines.
SA044
(pal II substitution with ACGCAC)
homodimer
$\Delta G = -315.2$ kcal/mol

SA045
(pal II substitution with GUGCNU)
homodimer
$\Delta G = -312.6$ kcal/mol

SA044-SA045 heterodimer
$\Delta G = -313.9$ kcal/mol
with our experimental data, these two mutant RNAs were predicted to form heterodimers *via* the PBS pal rather than *via* their *trans*-complementary mutated sequences (Fig. 3.12C).
3.5. Discussion

In an attempt to enhance our understanding of the gRNA dimerization and packaging processes during MMTV life cycle, we first validated the predicted structure of MMTV packaging signal RNA employing SHAPE. The SHAPE-validated structure was similar, but not identical to the Mfold-predicted structure. An important feature revealed by SHAPE was a U5-Gag LRI, which was also supported by phylogeny and in silico RNA structure predictions (Figs. 3.2, 3.3, and 3.4). The U5-Gag LRI anchors the overall structure of MMTV packaging signal RNA and is consistent with an earlier report showing that the MMTV packaging sequence extends into gag (Mustafa et al., 2012). Similar LRIs exist in the packaging signal RNAs of a number of retroviruses (Aktar et al., 2013; Kenyon et al., 2008, 2011; Paillart et al., 2002; Song et al., 2008; Spriggs et al., 2008) and in several cases, disruption of LRIs have been shown to affect RNA dimerization and packaging (Paillart et al., 2002; Rizvi et al., 2010; Song et al., 2008). In addition to LRI, SHAPE validated the existence of the SL2 and SL4 domains. The SL2 domain contains the PBS and the pal I sequence, and a similar large SL2 domain is found in a number of other retroviral packaging signal RNAs (Aktar et al., 2013; Baudin et al., 1993; Jaballah et al., 2010; Kenyon et al., 2008, 2011). The bifurcated SL4 hairpin contains a pal helix loop (pal II) and an ssPurines loop (Fig. 3.2). Both SL2 and SL4 structures are conserved in all the 8 MMTV strains studied (Figs. 3.3, 3.4).

Retroviral gRNA dimerization has been suggested as a prerequisite for packaging into the budding virions (reviewed in D’Souza & Summers, 2005; Johnson & Telenitsky, 2010; Lever, 2007; Lu et al., 2011; Nikolaitchik et al., 2013; Paillart et al., 2004; Russell et al., 2004). This process is usually initiated
by conserved palindromic sequences that initiate kissing loop interactions (Aktar et al., 2013; Berkhout & van Wamel, 1996; Kenyon et al., 2008, 2011; Lanchy et al., 2003; Lanchy & Lodmell, 2007; Laughrea et al., 1997; Leitner et al., 2005; Lever, 2007; Paillart et al., 1996a, 1996b, 2002; Rizvi et al., 2010; Russell et al., 2004; Skripkin et al., 1994). The MMTV packaging signal RNA initially revealed 3 pal sequences (pal I, pal II and pal III, Fig. 3.2). Our mutational analysis clearly indicated that pal I is not involved in RNA dimerization and packaging (Fig. 3.6). Nevertheless, pal I mutants had significant propagation defects (Fig. 3.6), suggesting it is possible that mutations close to the PBS affect annealing of the primer tRNA or initiation of reverse transcription itself. On the other hand, mutational analysis revealed a key role of pal II in MMTV RNA dimerization (Figs. 3.7, 3.9). However, mutations in pal II only reduced RNA dimerization by twofold, rather than abrogating it. This observation led us to identify an 11 nt sequence in the PBS consisting of two adjacent pals (PBS pal) as also playing an important role in RNA dimerization (Fig. 3.7). Similarly, a pal sequence within the PBS has also been shown to participate in HIV-2 RNA dimerization (Jossinet et al., 2001; Lanchy & Lodmell, 2002). In the case of MMTV, of these two pals, pal II exposed in an apical hairpin and harboring the canonical “GC” dyad, likely functions as the DIS. Interestingly, deletion of either pal II or the PBS pal has a pronounced effect on RNA packaging (even though deletion of pal II is more dramatic), suggesting that the MMTV Gag precursor initially recognizes an RNA dimer with two points of contacts (Fig. 3.7). The overall 3D structure of the packaging domain would thus probably be important for Gag binding. This could explain why deletion of 5’ UTR sequences predicted
to form the basal part of SL2 abrogated MMTV gRNA packaging (Mustafa et al., 2012).

A classical way of demonstrating the existence of any intermolecular interaction between viral RNAs is by using trans-complementary mutants (Aktar et al., 2013; Chen et al., 2009; Gavazzi et al., 2013; Jossinet et al., 2001; Paillart et al., 1996b). In the present study, trans-complementary pal II mutants did not restore either RNA dimerization or packaging (Fig. 3.9). This unexpected result could be explained by misfolding of one of the mutants (SA045; Figs. 3.10B, 3.12B). These two mutants were predicted to form heterodimer via the PBS pal rather than via their complementary mutated sequences (Fig. 3.12C). The results obtained with these mutants stress the importance of performing structural analyses on key RNA mutants; they also demonstrate that while the wild type structure is very stable (Figs. 3.2C, 3.8A), alternative folds with potentially altered biological functions are only marginally less stable. This could provide MMTV the possibility to control key functions by RNA switches, as observed for other retroviruses (Badorrek et al., 2006; Lu et al., 2011; Ooms et al., 2004; Paillart et al., 2004). SHAPE analysis was also important for understanding the results obtained with mutants SA043 and SA047 (in which pal II is replaced by the HIV-1 pal without or with the flaking purines, respectively). Even though these two mutants have a palindromic sequence in place of pal II, SHAPE indicated that the intermolecular interactions are not mediated by the mutated sequences in the absence of the flaking purines (Fig. 3.10), thus corroborating the RNA dimerization data (Fig. 3.9). Although mutant SA047 displayed wild type RNA dimerization levels, its RNA packaging and propagation were abolished (Fig. 3.9),
suggesting that the pal II hairpin could also be involved in RNA-protein interactions during genome encapsidation.

The presence of a stretch of purines in the packaging sequences on retroviral gRNA has been proposed to facilitate RNA packaging by functioning as a potential NC binding site (Lever, 2009; Moore & Hu, 2009; Paillart et al., 1997; Zeffman et al., 2000). Thus, it is interesting to note that the bifurcated SL4 domain contains a purine loop (5’ GGAGAAGAG 3’) adjacent to the pal II helix loop (Fig. 3.2), reminiscent to the situation found in MPMV (Aktar et al., 2013; Jaballah et al., 2010). In the case of MPMV, genetic and structure-prediction analyses have suggested that either ssPurines or its partial repeat base-paired sequence in an adjacent region play a crucial role in gRNA packaging, possibly by functioning as a NC binding site (Jaballah et al., 2010). Therefore, it will be interesting to investigate the role of ssPurines in MMTV gRNA packaging. The result presented in this study enhances our understanding of MMTV gRNA dimerization and packaging processes and the role of structural motifs with respect to RNA-RNA and possibly RNA-protein interactions that might be taking place during MMTV life cycle.
3.6. Acknowledgments

This research was funded primarily by grants from Terry Fox Fund for Cancer Research (2009-06) and the United Arab Emirates University (UAEU-NRF 26) and, in part, by a grant from the College of Medicine and Health Sciences (NP NP/10/22) to TAR. RM is supported by the Centre National de la Recherche Scientifique (C.N.R.S.). SJA is supported through a scholarship from the Office of the Deputy Vice Chancellor for Research and Graduate Studies at the United Arab Emirates University (UAEU).
CHAPTER 4

Conclusions

and

Future Directions
4.1. Conclusions

The main intent of the research presented in this dissertation was to validate the predicted structures of MPMV and MMTV packaging signal RNAs and to identify their DIS that plays an important role in augmenting the gRNA dimerization and packaging processes.

Employing a combination of genetic and structural prediction analyses, our laboratory has earlier shown that MPMV core packaging determinants (5’ UTR and beginning of Gag) are discontinuous (50 bp long region “A” and 23 bp long region “B”; Jaballah et al., 2010) and the intervening sequences are dispensable. Jaballah et al., (2010) also predicted that the 5’ end of the MPMV RNA genome folds into a RNA secondary structure. In the first part of this thesis (Chapter 2), the predicted structure of the MPMV packaging signal RNA was validated using a novel chemo-enzymatic approach known as SHAPE. SHAPE data of this region validated the overall predicted structure comprising of five stem loop structures (SL1-5). A very distinguishing feature of the SHAPE-validated structure of MPMV packaging signal RNA is the two LRIs (LRI-I and LRI-II) involving sequences from the U5 and first 70 nt of gag, which have been shown to be important in MPMV RNA packaging (Jaballah et al., 2010; Schmidt et al., 2003). Therefore, it is reasonable to propose that these two LRIs between U5 and gag sequences could potentially play a role in MPMV RNA packaging by maintaining the overall RNA secondary structure, as has recently been shown in the case of FIV (Kenyon et al., 2008, 2011; Rizvi et al., 2010). Between SL2 and SL3, the SHAPE-validated structure also confirmed the presence of a pal SL followed by a stretch of ssPurine-rich region and these sequences have been shown to be required for optimal MPMV gRNA packaging (Jaballah et al., 2010). The
importance of the pal SL and ssPurine-rich sequences was further supported by their conservation in different MPMV strains (Aktar et al., 2013). Since the pal SL forms a helix loop containing a canonical “GC” dyad, it was hypothesized that it may act as a RNA DIS and initiate the forming of a kissing loop complex, thus facilitating MPMV to encapsidate two copies of its gRNAs. To prove the hypothesis, a series of mutations were introduced into the pal SL and their analyses revealed that any interruption of pal sequences affected RNA dimerization severely, suggesting the crucial role of pal SL in MPMV gRNA dimerization. To further validate these observations, the in vitro transcribed RNAs from trans complementary pal mutants were incubated together to study their effects on RNA dimerization. Results from these tran-complementary approaches revealed that RNA dimerization was restored to wild type levels, further authenticating that the pal loop (5’ CGGCCG 3’) functions as DIS controlling MPMV gRNA dimerization.

In the second part of this thesis, presented in Chapter 3, SHAPE, structure prediction and in silico analyses were employed to validate the higher-order structure of MMTV packaging signal RNA which seems to be architecturally anchored by complementary sequences in U5 and gag. Of these complementary sequences gag sequences have earlier been shown to be crucial for MMTV gRNA packaging (Mustafa et al., 2012), further validating the existence and potential biological role of such LRI. We have also identified a 6 nt palindromic sequence (pal II; 5’ CUGCAG 3’) that folds into a helix loop structural motif that is needed for efficient MMTV gRNA dimerization and packaging. Employing a systematic deletion and/or substitution, and trans-complementary mutational approaches combined with structure predictions, our data convincingly reveal that pal II
functions as DIS, which is capable of initiating and mediating loop-loop intermolecular interaction to augment gRNA dimerization and packaging. Sequence alignment and a consensus structure prediction of 8 MMTV strains revealed that pal II primary sequence and its helix loop nature at the same location is phylogenetically preserved, further authenticating its pivotal role as DIS both at the primary sequence as well as at the structure level. In addition, another pal sequence within the PBS was also identified and its mutational analysis revealed its involvement in gRNA dimerization, suggesting that both pals are required for efficient MMTV gRNA dimerization, packaging, and propagation.

Therefore, results presented in this thesis should enhance our understanding of MPMV and MMTV gRNA dimerization and packaging processes and the role of structural motifs with respect to RNA-RNA and possibly RNA-protein interactions that might be taking place during MMTV life cycle.

4.2. Future Directions

Findings presented in this thesis on pal helix loop in the packaging signal RNA of MPMV and MMTV suggest that these pal helix loops play important roles in regulating MPMV and MMTV gRNA dimerization and packaging, yet the intricacies underlying these mechanisms are not yet fully resolved. Thus, these studies have opened new avenues for research which can be the subject of future research projects. Some of these potential topics that could be addressed are as follows:

1. It has been proposed that during retroviral gRNA dimerization, pal loop-loop intermolecular interactions initiates a kissing loop dimer which leads to further conformational changes in RNA secondary structure, exposing NC binding sites making them accessible for interaction with gRNA to facilitate its packaging
(D’Souza & Summers, 2004; Johnson & Telenitsky, 2010; Lu et al., 2011). The presence of a stretch of purines in the packaging sequences on retroviral gRNA has been proposed to facilitate RNA packaging by functioning as a potential NC binding site (Lever, 2009; Moore et al., 2009; Paillart et al., 1997; Zeffman et al., 2002). Consistent with these observations, a distinguishing feature of the SHAPE-validated structures of MPMV and MMTV packaging signal RNAs is the prominent stretch of ssPurines in the close vicinity of pal helix loops that have been identified as DIS. In the case of MPMV, a partial base-paired repeat sequence of the ssPurines was observed in the adjacent region that had earlier been shown to be important in MPMV gRNA packaging (Jaballah et al., 2010). Deletion mutants containing the region encompassing ssPurines showed an interesting structure folding prediction in which the stretch of duplicated base-paired purines became single stranded, restoring gRNA packaging significantly as well as the RNA secondary structure to that of the wild type (Jaballah et al., 2010). Since the overall predicted secondary structure and packaging was restored only in mutants in which duplicated base-paired purines became single stranded (the stretch of duplicated purines compensated for the conformational structure that would have been lost with the deletion of the ssPurines), it is reasonable to hypothesize that a stretch of native ssPurines or alternate purines in MPMV function either at the primary sequence, structural or both levels in mediating gRNA packaging, possibly by functioning as a potential NC binding site. Therefore, it would be interesting to substitute the single stranded purines with an alternate sequence, with and without altering the purine content, and observe their effects on MPMV and MMTV gRNA packaging both at the primary sequence as well as at the structural levels.
2. In a number of retroviruses such as HIV-1 and FIV, LRIs between complementary U5 and Gag sequences have been identified in their packaging signal RNAs (Aktar et al., 2013; Kenyon et al., 2008, 2011; Paillart et al., 2002; Song et al., 2008; Spriggs et al., 2008). Such an LRI, in the case of FIV, has recently been shown to have biological relevance during FIV gRNA packaging by providing the stability to the overall RNA secondary structure (Rizvi et al., 2010). The SHAPE-validated structures of MPMV and MMTV packaging signal RNAs also predict LRIs between U5 and Gag complementary sequences. Existence of such LRIs in these retroviruses was further supported by in silico analyses of different strains. In light of the results presented in this thesis, it is reasonable to propose that the U5 and gag complementary sequences involved in LRI could either function at the primary sequence level or at a structure level during MPMV and MMTV RNA packaging processes. Therefore, it will be important to ascertain the existence as well as biological significance of the LRIs during MPMV and MMTV gRNA packaging.

3. The SHAPE-validated structural models of MPMV and MMTV packaging signal RNAs revealed the presence of a large SL2 which has been shown to be conserved in a number of other retroviral packaging signal RNAs (Baudin et al., 1993; Kenyon et al., 2008, 2011). Consistent with this observation, Mfold structural predictions of different MPMV and MMTV strains have revealed the conservation of SL2 suggesting that such a structural motif could be of vital importance during MPMV and MMTV replication. Therefore, it will be interesting to mutate SL2 using different using different types of mutations (deletion, substitution, compensatory, etc.) to determine the importance of the
sequences involved in base pairing of SL2 in MPMV and MMTV gRNA packaging.
References


(FIV) are important for efficient RNA encapsidation. *Virus. Res.* 93, 199–209.


167. 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.


