United Arab Emirates University

College of Medicine and Health Sciences

DELINEATING THE BINDING SITES OF MASON-PFIZER
MONKEY VIRUS (MPMV) GAG PRECURSOR POLYPROTEIN
(Pr78^{GAG}) ON GENOMIC RNA FOR ITS SELECTIVE PACKAGING

Fathima Nuzra Nagoor Pitchai

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

Under the Supervision of Professor Tahir A. Rizvi

April 2021

Declaration of Original Work

I, Fathima Nuzra Nagoor Pitchai, the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this dissertation entitled "*Delineating the Binding Sites of Mason-Pfizer Monkey Virus (MPMV) Gag Precursor Polyprotein (Pr78^{Gag}) on Genomic RNA for its Selective Packaging*", hereby, solemnly declare that this dissertation is my own original research work that has been done and prepared by me under the supervision of Professor Tahir A. Rizvi, in the College of Medicine and Health Sciences at UAEU. This work has not previously formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my dissertation have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this dissertation.

Student's Signature: _____   Date: 05.10.2021

Advisory Committee

1) Advisor: Tahir A. Rizvi

Title: Professor

Department of Microbiology

College of Medicine and Health Sciences

2) Co-advisor: Farah Mustafa

Title: Associate Professor

Department of Biochemistry

College of Medicine and Health Sciences

3) Member: Gulfaraz Khan

Title: Professor

Department of Microbiology

College of Medicine and Health Sciences

4) Member: Ahmed Al-Marzouqi

Title: Associate Professor and Chair

Department of Biochemistry

College of Medicine and Health Sciences

Approval of the Doctorate Dissertation

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

1) Advisor (Committee Chair): Tahir A. Rizvi

    Title: Professor

    Department of Microbiology

    College of Medicine and Health Sciences
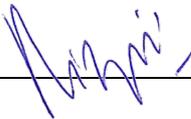
    Signature _____ Date _May 10, 2021_

2) Member: Farah Mustafa

    Title: Associate Professor

    Department of Biochemistry

    College of of Medicine and Health Sciences

    Signature _____ Date _May 10, 2021_

3) Member: Gulfaraz Khan

    Title: Professor

    Department of Microbiology

    College of Medicine and Health Sciences

    Signature _____ Date _May 10, 2021_

4) Member: Ahmed Al-Marzouqi

    Title: Associate Professor and Chair

    Department of Biochemistry

    College of Medicine and Health Sciences

    Signature _____ Date _May 10, 2021_

5) Member (External Examiner): Louis M. Mansky

   Title: Director, Institute for Molecular Virology

   Professor, Department of Basic Sciences (Dentistry) and Microbiology (Medical School)

   Institution: University of Minnesota-Twin Cities, Minneapolis, USA

   Signature _____ Date __May 10, 2021__

This Doctorate Dissertation is accepted by:

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

Kaabi

Signature _____   Date ___06 June 2021_____

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

Signature ___Ali Hassan_____   Date ___06/06/2021_____

Copy ____ of ____

Abstract

A key step in retroviral life cycle is the selective packaging of its dimeric RNA genome (gRNA) from a pool of cellular and spliced viral RNAs into nascent virions. This involves binding of the retroviral Gag polyprotein to sequences at the 5' end of the viral genome, the packaging signal. The aim of this study was to identify full-length Gag polyprotein (Pr78^{Gag}) binding sites on Mason-Pfizer monkey virus (MPMV) gRNA, a promising candidate for the development of safe human gene therapy vectors. Towards this end, recombinant MPMV Pr78^{Gag}-His$_6$-tagged protein was cloned and expressed in bacterial cells, and purified from the soluble fraction using immobilized metal affinity chromatography (IMAC) followed by size exclusion chromatography (SEC). The biological activity of the purified protein was determined by its ability to assemble virus like particles (VLPs), while its ability to package MPMV specific sub-genomic RNAs was confirmed in eukaryotic cells. Competitive band shift assays demonstrated preferential Pr78^{Gag} binding to unspliced over spliced viral RNA. Further competitive band shift assays were performed using mutants in two purine-rich motifs consisting of a 16-nucleotide stretch of single-stranded purines (ssPurines; U^{191}UAAAAGUGAAAGUAA^{206}) and a partially base-paired purine-rich region (bpPurines; G^{246}AAAGUAA^{253}), previously found to be important for MPMV gRNA packaging. To map the precise Pr78^{Gag} binding sites on the MPMV gRNA, *in vitro* Gag-RNA foot-printing experiments followed by high-throughput selective 2' hydroxyl acylation analyzed by primer extension (hSHAPE) were performed. These revealed that Pr78^{Gag} binds to ssPurines, and the A^{252}AGUGUU^{258} loop, corresponding to two unpaired adenosine residues of the bpPurines and the adjacent region called the "GU-rich region" (G^{254}UGUU^{258}), both of which flank the major splice donor. Hence, ssPurines are present on both the genomic and spliced viral RNAs, while the A^{252}AGUGUU^{258} loop is found only on the gRNA, revealing how MPMV discriminates between genomic and spliced RNAs. Collectively, this study reveals how MPMV Pr78^{Gag} binds in a redundant fashion to the two single-stranded loops (ssPurines and the A^{252}AGUGUU^{258} loop) to bring about selective gRNA packaging over spliced viral RNAs. These results should help in understanding virion assembly and facilitate development of safe and efficient retroviral vectors for human gene therapy.

Title and Abstract (in Arabic)

تحديد مواقع الارتباط للبروتين الفيروسي Pr78^{GAG} على الحمض النووي الريبوزي (gRNA) لفيروس القرد مايسون-فايزر (MPMV) ودوره في التجميع الانتقائي

الملخص

يعتبر التجميع الانتقائي هو أحد الخطوات الرئيسية في دورة حياة الفيروسات الارتجاعية لدمج المادة الوراثية للفيروس (الحمض النووي الريبوزي (gRNA)) من بين خليط من الاحماض النووية الريبوزية الخلوية والفيروسية لتكوين فيروسات جديدة. ويتضمن التجميع الانتقائي ارتباط البروتين الفيروسي Gag بالتسلسل الجيني المتواجد في النهاية '5 من المادة الوراثية للفيروس، وهو ما يعرف بمحددات التجميع. ولقد كان الهدف من هذه الدراسة هو تحديد مواقع ارتباط البروتين الفيروسي Gag (Pr78^{Gag}) مع gRNA لفيروس القرد مايسون-فايز (MPMV)، وهوفيروس مرشح للاستخدام في تطوير نواقل آمنة للعلاج الجيني في البشر. لتحقيق هذه الغاية، تم استنساخ البروتين الفيروسي MPMV Pr78^{Gag}-His₆ والتعبير عنه في الخلايا البكتيرية، ثم تنقيته من الجزء القابل للذوبان باستخدام كروماتوغرافيا تقارب أيونات المعدن الثابت (IMAC) متبوعًا بكروماتوغرافيا الاستبعاد عن طريق اختلاف الحجوم (SEC). ولقد تم تحديد النشاط البيولوجي للبروتين المنقى من خلال إبراز قدرته على تجميع جزيئات شبيهة بالفيروسات (VLPs). في حين تم تأكيد قدرة هذا البروتين على جمع وتغليف جزيئات RNA الفرعية التابعة لفيروس MPMV في الخلايا حقيقية النواة. وأظهرت فحوصات تحول النطاق التنافسي ارتباط Pr78^{Gag} بجزيئات RNA الفيروسي غير المقسم على الأخص على الرغم من وجود جزيئات RNA المقسمة. وتم إجراء فحوصات تحول نطاق تنافسية أخرى باستخدام طفرات في مناطق غنية بالبيورين والتي تتكون من امتدادات لحوالي 16 نيوكليوتيد من البيورينات الاحادية (ssPurines; U¹⁹¹UAAAAGUGAAAGUAA²⁰⁶) ومنطقة غنية بالبيورين المقترنة جزئيًا (bpPurines; G²⁴⁶AAAGUAA²⁵³)، والتي تعرف بأهميتها في تجميع gRNA لفيروس ال MPMV. وللتعيين بدقة مواقع ارتباط بروتين ال Pr78^{Gag} على gRNA لفيروس ال MPMV، تم إجراء تجارب بصمة للمركب Gag-RNA متبوعة بتقنية الأسيلة الانتقائية ذات الإنتاج العالية ل '2 هيدروكسيل و تحليلها بواسطة hSHAPE. كشفت نتائج هذه التقنية أن بروتين Pr78^{Gag} يرتبط بـ ssPurines، وحلقة ال A²⁵²AGUGUU²⁵⁸، وهي عبارة عن اثنين من بقايا الأدينوسين غير المتزاوجة من bpPurines والمنطقة المجاورة المسماة "بالمنطقة الغنية بال GU

(G^{254}UGUU^{258})"، وكلاهما يحيط بالمنطقة المانحة الرئيسة (major splice donor). وعلى ذلك، فإن ssPurines موجودة على كل من جزيئات gRNA (غير مقسّمة) وجزيئات RNA مقسّمة تابعة للفيروس. بينما تقع حلقة A^{252}AGUGUU^{258} فقط على gRNA، و هذا يظهر كيفية فرز الفيروس MPMV بين gRNA وجزيئات RNA المقسّمة. بشكل عام، تكشف هذه الدراسة كيف يرتبط بروتين فيروس MPMV Pr78^{Gag} إلى الحلقتين المفردتين (ssPurines وحلقة A^{252}AGUGUU^{258}) للتمييزخلال الجمع و التغليف بين gRNA وجزيئات RNA الأخرى المقسمة. وتكمن أهمية هذه الدراسة في فهم تجميع الفيروسات وتسهيل تطوير ناقلات للجينات العلاجية من الفيروسات الارتجاعية للاستخدام الآمن والفعال في البشر.

كلمات البحث الرئيسية: الفيروسات الارتجاعية، فيروس القرد مايسون-فايزر (MPMV)، بروتين Pr78^{Gag}، الحمض النووي الريبوزي gRNA، ارتباط Gag-RNA، جمع وتغليف RNA، البيورينات، البصمة، تقنية hSHAPE.

Acknowledgements

Dedication

I dedicate this dissertation to my Mum and Dad
To my Husband, Haizam &
My Son, Haadhi <3

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AS	Antisense
bpPurines	Base-Paired Purines
BzCN	Benzoyl Cyanide
CTE	Constitutive Transport Element
DIS	Dimerization Initiation Site
DLS	Dynamic Light Scattering
Env/*env*	Envelope
Gag	Group Antigen
gRNA	Genomic Ribonucleic Acid
HEK	Human Embryonic Kidney
His_6	Hexa-histidine
HIV	Human Immunodeficiency Virus
HRP	Horse Radish Peroxidase
hSHAPE	High-Throughput Selective 2'Hydroxyl Acylation by Primer Extension
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β- d-1-thiogalactopyranoside
LRI	Long Range Interaction
LTR	Long Terminal Repeat
MPMV	Mason Pfizer Monkey Virus
mSD	Major Splice Donor
NC	Nucleocapsid
ORF	Open Reading Frame

PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
Pr78^{Gag}	Precursor Polyprotein 78 Group Antigen
Psi	Packaging Signal
R$_h$	Hydrodynamic Radius
RNA	Ribonucleic Acid
S	Sense
SDS	Sodium Dodecyl Sulphate
SEAP	Secreted Alkaline Phosphatase
SEC	Size Exclusion Chromatography
SL	Stem Loop
SOE	Splice Overlap Extension
ssPurines	Single Stranded Purines
TCA	Trichloro Acetic Acid
UTR	Untranslated Region
U5	5' Unique
U3	3' Unique
VLP	Virus Like Particle
WT	Wild Type

Chapter 1: Introduction

1.1 Retroviruses

Retroviruses comprise of a group of enveloped ribonucleic acid (RNA) viruses categorized into the family *Retroviridae* (Coffin et al., 1997). One important feature of these viruses is that they code for the enzyme reverse transcriptase, capable of reverse transcribing their RNA genome into double stranded deoxyribonucleic acid (dsDNA), and integrates this dsDNA into the host chromosomes leading to the formation of pro-virus in infected host cells (Baltimore, 1970; Temin & Mizutani, 1970). The enzyme reverse transcriptase is also found in other viral families, such as the *Hepadnaviridae* animal viruses and the *Caulimoviridae* plant viruses, and two families of long terminal repeat (LTR) retrotransposons; the *Metaviridae* and *Pseudoviridae* where reverse transcription forms a part of their life cycle. However, integration is a must for the Retroviruses and is rarely found in other families of viruses with the exception of the aforementioned LTR retrotransposons (Menéndez-Arias et al., 2017).

Retroviruses are broadly divided into two types based on their genome organization and complexity; the simple and complex retroviruses. Simple retroviruses, in addition to the *cis*-acting controlling elements contain canonical open reading frames, *gag, pro, pol* and envelope (*env*) expressing both structural and functional proteins required for the continuation of their life cycle. On the other hand, complex retroviruses in addition to these structural and functional proteins also code for regulatory proteins such as Vif, Vpr, Vpu, Tat, Rev, and Nef expressed through singly or multiply spliced messages (Figure 1; Coffin et al., 1997; Murphy et al., 1994).
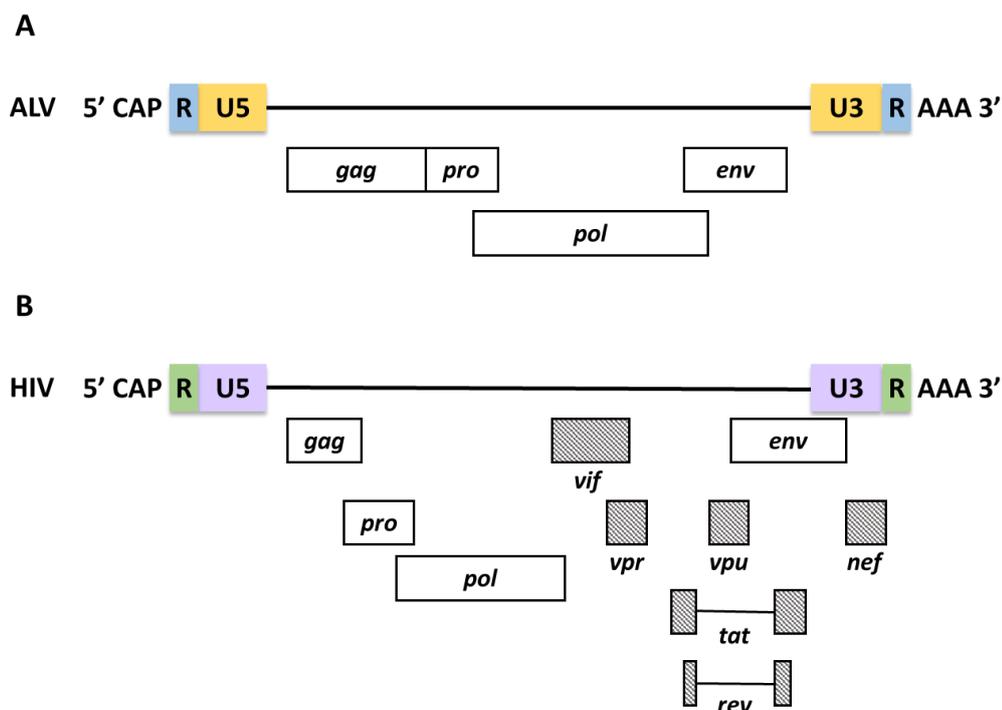
Figure 1: Genome organization of simple and complex retroviruses.

(A) A simple retroviral genome. The genetic map of an avian leukosis virus (ALV) contains four major coding regions, *gag, pro, pol*, and *env*. Different reading frames are indicated by vertical displacement of the coding region. The *pro* gene is encoded in the *gag* reading frame. The terminal noncoding sequences include two direct repeats (R), a U5 (5′ unique), and a U3 (3′ unique) sequence. (B) A complex retroviral genome. The genetic map of human immunodeficiency virus (HIV) contains, besides the major coding domains, information for six other regulatory proteins, Vif, Vpr, Vpu, Nef, Tat and Rev, encoded in the hashed regions joined by RNA splicing. In this case, *gag*, *pro*, and *pol* are all in different reading frames. Figure legend adapted and modified from Coffin et al., 1997.

The *Retroviridae* family has been further sub-categorized into two sub-families and seven genera based on their genetic relatedness of the reverse transcriptase protein (Linial et al., 2005). The two subfamilies are the *Orthoretrovirinae* and the *Spumaretrovirinae*, of which the *Orthoretrovirinae* consists of six genuses; the *Alpharetroviruses*, *Beta-retroviruses*, *Gammaretroviruses*, *Deltaretroviruses*, *Epsilonretroviruses* and *Lentiviruses*. The *Spumaretrovirinae* consists of a single genus the *Spumaviruses* (Table 1).

Table 1: Classification of retroviruses.

Sub-family: *Orthoretrovirinae*		
Genus	**Features**	**Examples**
Alpharetrovirus, α	C-type Simple	Avian leukosis virus (ALV) Rous sarcoma virus (RSV)
Betaretrovirus, β	B and D-type Simple	Mouse mammary tumor virus (MMTV) Mason-Pfizer monkey virus (MPMV) Jaagsiekte sheep retrovirus (JSRV)
Gammaretrovirus, γ	C-type Simple	Murine leukemia viruses (MuLV) Feline leukemia virus (FeLV) Gibbon ape leukemia virus (GaLV) Reticuloendotheliosis virus (REV)
Deltaretrovirus, δ	Complex, Oncogenic	Human T-lymphotropic virus (HTLV)-1, -2 Bovine leukemia virus (BLV) Simian T-lymphotropic virus types 1-3 (STLV-1, -2, -3)
Epsilonretrovirus, ε	Complex, Oncogenic	Walleye dermal sarcoma virus Walleye epidermal hyperplasia virus 1
Lentivirus	Complex	Human immunodeficiency virus type 1 and 2 (HIV-1 and -2) Simian immunodeficiency virus (SIV) Feline immunodeficiency virus (FIV) Equine infectious anemia virus (EIAV)3 Caprine arthritis encephalitis virus (CAEV) Visna/maedi virus
Sub-family: *Spumaretrovirinae*		
Spumavirus	Complex	Human foamy virus (HFV) Simian foamy virus (SFV) Feline Foamy Virus Bovine Foamy Virus

Additionally, retroviruses have also been categorized according to their morphogenic or assembly pathways identified based on electron microscopy (Coffin et al., 1997; Goff, 2001). This classifies members of the *Alpharetrovirus*, *Gammaretrovirus* and *Lentivirus* genera which assemble their immature capsids at the plasma membrane, as C-type viruses. In contrast, members of the *Beta-retrovirus* and *Spumavirus* preassemble their immature cores in the cytoplasm assuming either a B-type (mouse mammary tumor virus, MMTV) or D-type (Mason-Pfizer monkey virus, MPMV) morphology (Figure 2).
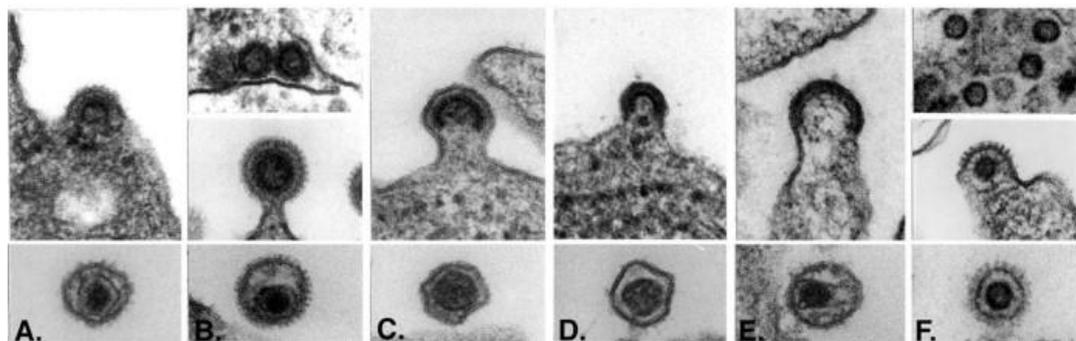
Figure 2: Different morphologies of retroviruses.

(A) Alpharetrovirus, avian leukosis virus (ALV); type "C" morphology. (B) Beta-retrovirus, mouse mammary tumor virus (MMTV); type "B" morphology. (C) Gammaretrovirus, murine leukemia virus (MLV). (D) Deltaretrovirus, bovine leukemia virus (BLV). (E) Lentivirus, human immunodeficiency virus type 1 (HIV-1). (F) Spumavirus, simian foamy virus (SFV; formerly called HFV). Figure and legend adapted from King et al., 2012.

1.2 Virion Structure

Retroviruses are spherical in shape and range in size from 80 to 100 nm in diameter. Their outer lipid envelope is derived from the host cell during budding and harbors the envelope (Env) glycoproteins coded by the *env* gene. The Env glycoproteins are made up of two important polypeptides. One is a hydrophilic surface glycoprotein complex (SU) that interacts with specific receptors on target cells to initiate the entry or infection process, the other is a hydrophobic transmembrane protein (TM) that initiates fusion of the viral envelope with the cell membrane. Beneath this lipid bilayer lies the matrix (MA) which forms a protective layer that encloses the viral core which is in turn made up of the capsid (CA) and the nucleocapsid (NC). The MA, CA and NC are all mature structural protein products of the *gag* gene. The mature core of the virus is also referred to as the viral ribonucleocapsid and consists of the viral genome, which is more specifically a dimeric single stranded positive sense RNA genome, associated with the viral

nucleoprotein, the NC. Apart from the RNA genome of the virus, the mature viral core

also consists of the enzymes reverse transcriptase (RT) and integrase (IN) required for

active viral replication and integration during infection (Figure 3) (Coffin et al., 1997).
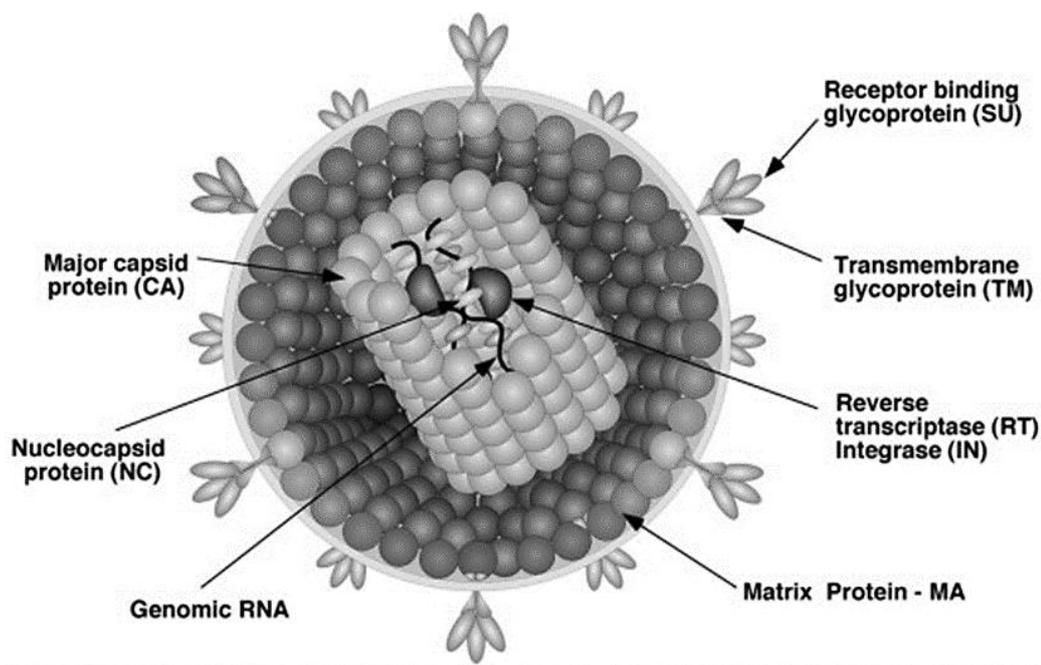
Figure 3: Structure of a retrovirus particle.

Schematic cartoon (not to scale) showing the locations of the various structures and proteins. Figure and legend adapted from King et al., 2012.

The immature viral capsid differs slightly in structure compared to the mature

core, here the Gag (group antigen) polyprotein; a product of the *gag* gene and precursor

of the mature viral structural proteins (MA, CA, and NC) oligomerizes during viral

assembly to form a spherical layer around the viral RNA (Ganser-Pornillos et al.,

2012). Here the C terminal NC domain in context of the Gag polyprotein is found

associated with the viral RNA at the center of the viral particle with the CA and the N

terminal MA organized sequentially from the inside to the outside of the particle

(Figure 4B). The transition of the viral structure from its simplistic immature form to

its complex mature form involves the viral protease (PR) enzyme and is brought about

by the process called 'maturation' during the viral life cycle (Figure 4A). The structural rearrangement that occurs during the maturation process results in unique shapes for their viral cores. The apparently spherical nucleocapsid is eccentric for members of the genus *Beta-retrovirus* (Figure 2B), concentric for members of the genera *Alpharetrovirus*, *Gammaretrovirus*, *Deltaretrovirus* and *Spumavirus* (Figure 2A, C, D and F), and rod or truncated cone-shape for members of the genus *Lentivirus* (Figure 2E; Figure 4B, C, D and E). In contrast, *Spumaviruses* do not undergo maturation and hence retain their immature viral morphology (Figure 2).
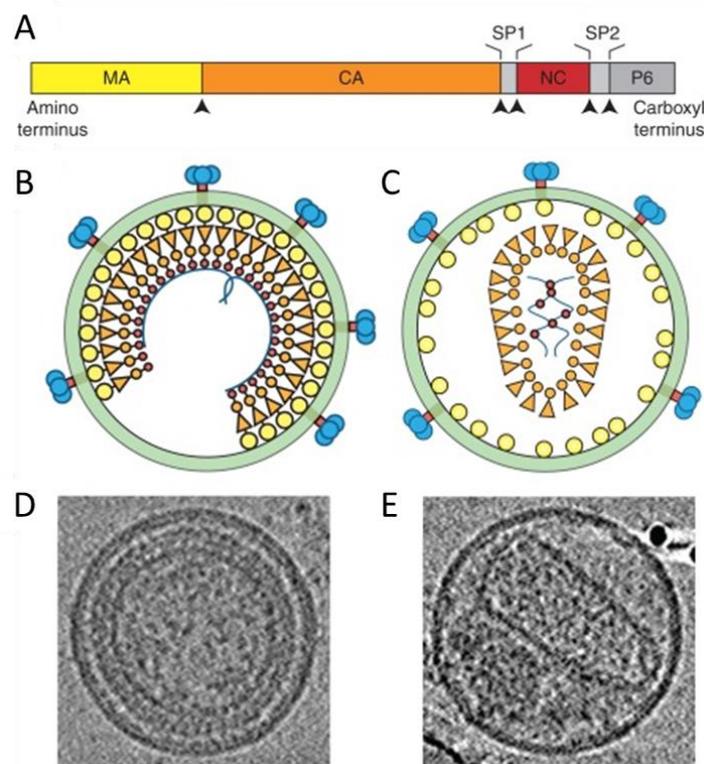
Figure 4: Organization of an immature *versus* mature retroviral particle.

(A) Domain structure of the HIV-1 Gag protein; arrows denote the five sites that are cleaved by the viral PR during maturation. (B) Schematic model showing the organization of the immature HIV-1 virion. (C) Schematic model showing the organization of the mature HIV-1 virion. (D) Central section from a cryo-EM tomographic reconstruction of an immature HIV-1 virion. (E) Central section from a tomographic reconstruction of a mature HIV-1 virion. Figure and legend adapted from Ganser-Pornillos et al., 2012.

1.3 The Retroviral Genome and its Organization

Retroviruses are RNA viruses that consist of two identical copies of single stranded positive sense RNA that are found associated with intermolecular base pairing to form a dimeric genome (Ali et al., 2016; D'Souza & Summers, 2005; Dubois, et al., 2018a; Johnson & Telesnitsky, 2010; Lever, 2007). The RNA genome of retroviruses can be broadly categorized into *cis-* and *trans*-acting elements. Briefly, the *cis*-acting elements are non-coding regulatory sequences while the *trans*-acting elements are sequences coding for the structural and functional proteins of the virus (Coffin et al., 1997).

The *cis*-acting elements being present at the 5' and 3' ends of the viral RNA genome flank the *trans*-acting elements of the genome. The most distinguishing feature of the 5' and 3' *cis*-acting sequences include the 5' unique (U5) and 3' unique (U3) regions respectively. Repeat regions known as R regions of identical sequence are present on both *cis* ends of the RNA genome, found located upstream of U5 and downstream of U3 sequences. The U5 and U3 sequences found on the respective ends of the RNA genome are duplicated at opposite ends in the pro-viral DNA to form long terminal repeats or LTRs (Dhar et al., 1980; Gilboa et al., 1979). This duplication is brought about by the process of reverse transcription, mediated by reverse transcriptase (RT) resulting in the formation of dsDNA, during which RT performs two 'jumps' called first and second strand transfers from one end to the other end of the RNA genome. These jumps are assisted by sequence homology involving both the terminal repeats (R) and the primer binding site (PBS) respectively, enabling the generation of the LTRs at both ends of the dsDNA (Figure 5; Coffin et al., 1997; Hughes, 2015).
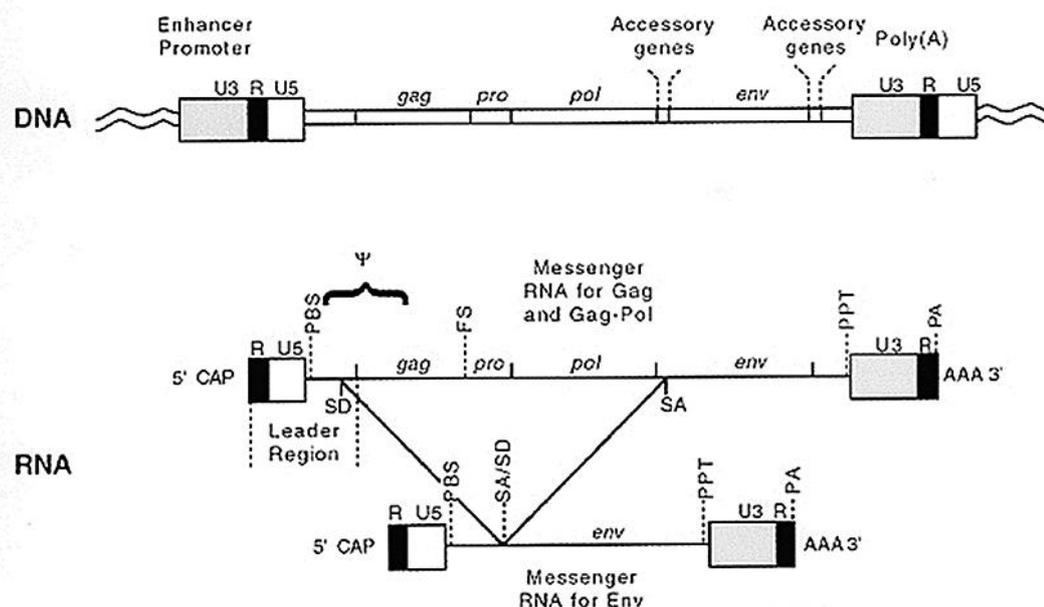
Figure 5: Genetic organization of a generalized provirus.

The proviral DNA as it is inserted into host DNA is shown at the top, with the long terminal repeats (LTRs) composed of U3, R, and U5 elements at each end abutting cellular sequences. Sequences in the LTR that are important for transcription, for example, enhancers, the promoter, and the poly (A) addition signal, are marked. The *gag, pro, pol*, and *env* sequences are located invariably in the positions shown in all retroviruses. Accessory genes are located as shown, and also overlapping *env* and U3 and each other, and occasionally in other locations. The RNA that is the primary transcriptional product is shown on the second line. Sequences that are important for replication and gene expression are shown in the approximate locations in which they are typically found. (PBS) Primer-binding site; (ψ) encapsidation sequence; (SD) splice donor site; (FS) frameshift site; (SA) splice acceptor site; (PPT) polypurine tract; (PA) polyadenylation signal; (AAA) poly(A) tail. The spliced messenger RNA for the Env protein is shown on the third line. Retroviruses with accessory genes have other spliced mRNAs and thus other splice donor and splice acceptor sites as well. Figure and legend adapted from Coffin et al., 1997.

The 5' LTR of the retroviral genome also consists of sequences indispensable for retroviral genomic RNA (gRNA) packaging. These sequences generally span from the R (+1) region at the 5' end into the *gag* open reading frame (ORF) for most of the retroviruses including HIV-1 (with a few exceptions where sequences in the 3' end and elsewhere may be involved) and are referred to as the packaging signals (*Psi*/Ψ) of the retroviral genome and share no consensus in sequence homology among different retroviruses (Figure 6A; Parveen et al., 2004; Al Dhaheri et al., 2009; Al

Shamsi et al., 2011; Ali et al., 2016; Dubois et al., 2018a; Moore et al., 2007, 2009; Motomura et al., 2008; Rizvi & Panganiban, 1993; White et al., 1999; Yin & Hu, 1997). Packaging signals of almost all retroviruses have been found to fold into complex secondary/tertiary structures comprising of various structural motifs such as stem loops (SLs) and occasionally U5-*gag* long range interactions (LRIs) that share structure conservation (Figure 6B; Ali et al., 2016; D'Souza & Summers, 2005; Dubois et al., 2018a; Johnson & Telesnitsky, 2010; Lever, 2007; Lu, Heng, & Summers, 2011; Miyazaki et al., 2011).
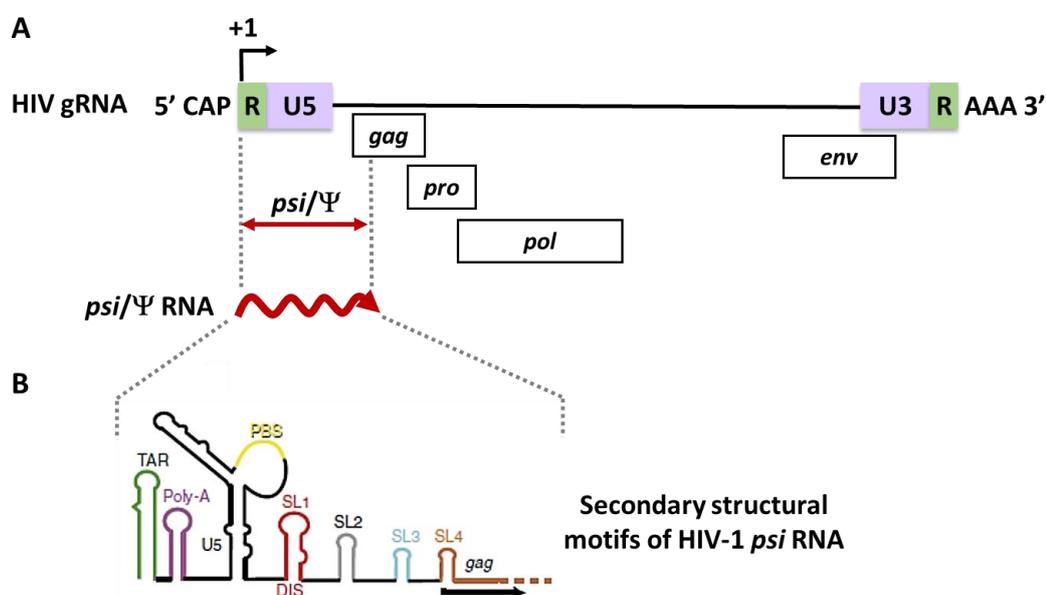
Figure 6: Schematic representation of the HIV-1 genomic RNA, its packaging signal RNA and secondary structural motifs.

(A) Schematic representation of the HIV-1 genomic RNA indicating the +1 transcription start site and the packaging signal region spanning from R to the beginning of *gag*. The red squiggly line represents the packaging signal RNA transcribed from this region. (B) Graphical representation of the secondary structural motifs of the HIV-1 packaging signal RNA showing the stem loop motifs of the TAR element, the poly(A) hairpin, the PBS, and stem-loops 1–4 (SL1- 4) containing the DIS and the *gag* start codon.

This highly structured RNA element also consists of a dimerization initiation site (DIS) which is often present in the form of an apical loop harboring palindromic

or self-complementary sequences that allows for intermolecular base pairing of the RNA genome, a process called RNA dimerization leading to the development of a dimeric genome for packaging into the newly assembling virus particles. The dimerization process is considered as a pre-requisite to retroviral gRNA packaging in most of the retroviruses studied thus far (Paillart et al., 2004; D'Souza and Summers, 2005; Moore et al., 2009; Johnson and Telesnitsky, 2010; Dilley et al., 2011; Ali et al., 2016; Dubois et al., 2018a; Rein, 2019). It is known that the process of dimerization results in conformational change(s) in the RNA secondary structure that may expose the otherwise obscured Gag binding sites on the packaging sequences (Berkhout & Van Wamel, 1996; Johnson & Telesnitsky, 2010; Russell et al., 2004). Gag binding sites that are involved in the specific binding of the Gag polyprotein to the packaging sequences are responsible for augmenting selective packaging of the retroviral gRNA (Ali et al., 2016; Dubois et al., 2018a; Kuzembayeva et al., 2014; Rizvi et al., 2010; Webb et al., 2013). Among the other important structural elements within this packaging signal are the PBS and the major splice donor (mSD). The PBS plays an important role in the priming step of reverse transcription and the mSD plays an important role in the splicing events that generate spliced viral products.

The *trans* acting elements occupy the center of the retroviral/pro-viral genome (Figure 5). These consist of the structural and enzymatic genes *gag, pro, pol* and *env*. These genes are common to both the simple and complex retroviruses. The *gag* gene encodes for the major structural protein Gag. Which is in turn cleaved into the MA, CA and NC proteins with unique roles in viral assembly. The *pro* gene codes for the viral PR whose function allows for viral maturation and the generation of infectious viral particles. The *pol* gene codes for the functional enzymes RT and IN that are responsible for viral DNA synthesis and integration (leading to the formation of pro-

viral DNA) respectively. The *env* gene codes for the TM and SU proteins that form the Env glycoproteins required for successful host cell infection. In addition to these *trans* elements, complex retroviruses such as the lentiviruses, spumaviruses, and the deltaretroviruses; human T cell lymphotrophic virus (HTLV) and bovine leukemia virus (BLV) also consist of genes that code for accessory proteins that aid in the regulation of viral gene expression, virulence, and pathogenicity. These genes are usually found downstream of *pol* and *env* genes (Figure 5). Some oncogenic retroviruses also consist of oncogenic genes responsible for their oncogenic nature (Coffin et al., 1997).

1.4 Retroviral Life Cycle

1.4.1 Fusion and Entry

Retroviruses infect their target cells with the help of their Env glycoproteins. The HIV-1 *env* gene codes for a gp160 which is in turn cleaved by cellular proteases to generate SU (gp120) and the TM (gp41) glycoproteins. These subunits are found associated with each other as gp41/gp120 trimers in the virion structure. This Env complex specifically targets two different receptors on immune T cells, the CD4 receptor and the CXCR4 and CCR5 co-receptors both of which belong to the chemokine receptor family. The specific binding and entry mechanism involve several steps where first the trimers in the complex identifies and binds to CD4 and this brings about a conformational change in both CD4 and gp120, which also allows the condensation of the co-receptors around the binding region (Berger et al., 1999). Secondly, gp120 binds to the co-receptor and this brings about a conformational change that allows for the dissociation of the trimer and allows gp41 to penetrate the plasma membrane and bring about fusion of the viral and cell membranes allowing the

viral core to enter the cytoplasm (Figure 7A) (Gallo et al., 2003; Moore & Doms, 2003; Nisole & Saïb, 2004).
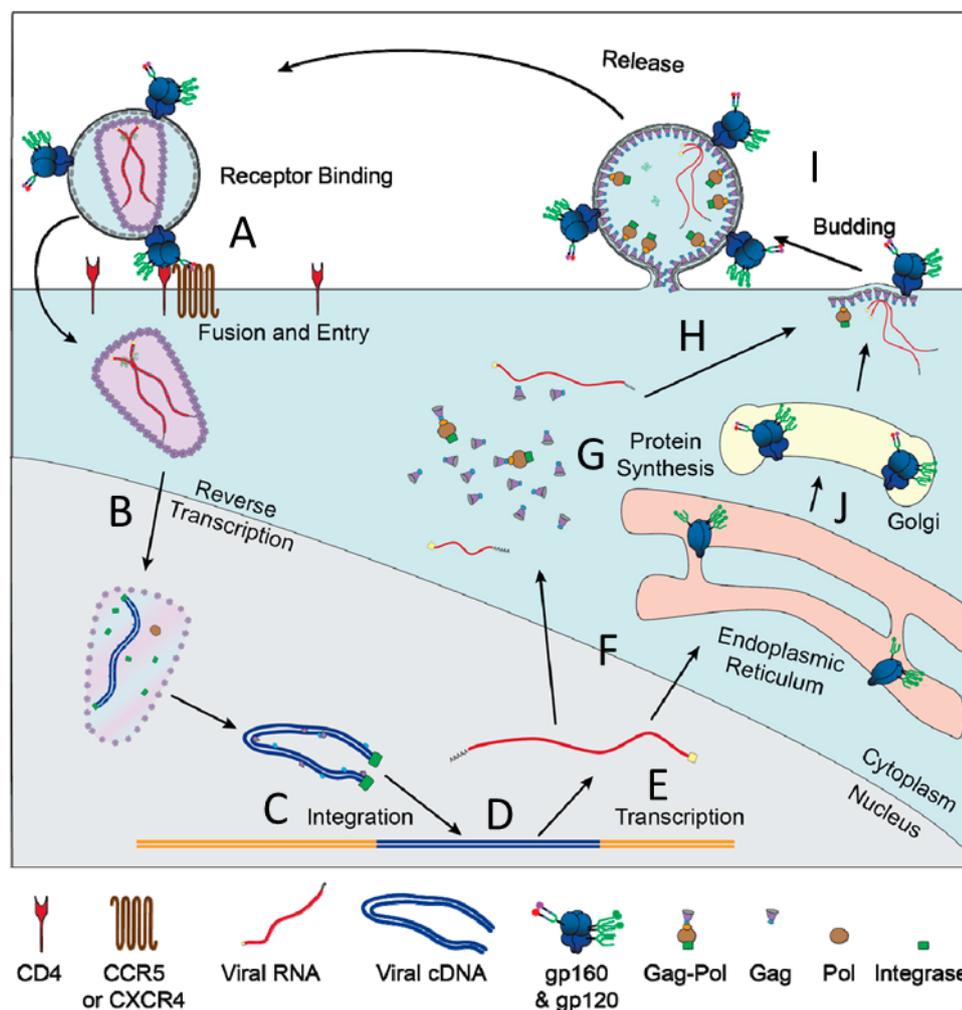
Figure 7: Replication cycle of HIV.

Graphical depiction of the major events of the HIV replication cycle. (A) Viral entry is mediated by binding of Env to CD4 and a co-receptor (CXCR4 or CCR5). After the viral membrane fuses with the plasma membrane the capsid and RNA genome are transported into the nucleus. (B) The viral genome is reversed transcribed by Pol and (C) integrates into the host genome to form the (D) pro-viral DNA. (E) The host RNA polymerase transcribes RNA copies of the genome, which are (F) exported in an unspliced form for (G) protein synthesis to produce Gag and Gag–Pol or various spliced forms to produce the other viral proteins. (H) Unspliced RNA genomes are also packaged into newly (I) assembling virions (J) Most viral protein translation occurs in the cytoplasm, but Env gp160 is translated into the lumen of the rough ER and transported through the secretory pathway, where it is glycosylated and cleaved by furin into gp120 and gp41. Figure and legend adapted and modified from Lubow & Collins, 2020.

1.4.2 Uncoating, Reverse Transcription, and Integration

It was long understood that after the viral core enters the cytoplasm it disassembles into sub-viral particles, namely the reverse transcription complexes (RTCs) responsible for reverse transcription of the RNA genome and the pre-integration complexes (PICs) that are formed following reverse transcription which enter the nucleus for integration into the host chromosomes (Zhang et al., 2000). However more recent studies focusing on the post entry events such as, viral core uncoating and the locality of reverse transcription tend to challenge this understanding. Live cell imaging and microscopy studies carried out by green fluorescent protein (GFP)-CA labelled viral cores in live cells, have identified that the viral core traverses the nucleus with the help of specific host factors (cleavage and polyadenylation specificity factor 6; CPSF6) in a partially or completely intact form, reverse transcription and PIC formation is completed within the intact core (Figure 7B) and it uncoats within the nucleus close to the site of integration, approximately 1.5 hours before integration. (Figure 7C; Burdick et al., 2020; Dharan et al., 2020; Zila et al., 2020). The reverse transcribed DNA within the PIC is then integrated into the target cell chromosome with the help of the virally encoded protein, IN. The integrated viral cDNA is now referred to as pro-viral DNA (Figure 7D). A number of studies have shown that the reverse transcribed retroviral DNA preferentially integrates into actively transcribing genes of the host chromosome and its integration is not sequence specific and instead may be influenced by specific chromatin architecture and interactions that occur between host proteins and viral components (Craigie, 1992; Nisole and Saïb, 2004; Schröder et al., 2002).

1.4.3 Transcription and Nuclear Export

Once integrated into the host cell chromosome the pro-viral DNA is transcribed with the help of the host cell transcriptional machinery. In all retroviruses, the pro-viral DNA transcribes both the full length and spliced mRNA products. Interestingly in retroviruses the full-length mRNA transcript or the unspliced RNA serves as both the packageable gRNA for novel virions during assembly as well as the transcript for Gag-Pro-Pol expression (Figure 7E; Butsch & Boris-Lawrie, 2002). Viral RNA transcripts are generated from the +1 transcription start site on R. After transcription, the retroviral transcripts are treated by the host cell as any other cellular transcript and undergo the ordinary post-transcriptional modifications such as the 5' cap addition and the addition of a 3' poly-A tail. Also like cellular mRNA, viral mRNA transcription is followed by the important event of splicing by the host cell spliceosome that generates the viral spliced mRNA products. The viral transcript undergoes splicing as a means of generating alternate viral products required during the retroviral life cycle. One such spliced product consistently found in all retroviruses is the spliced *env* mRNA that codes for the Env protein however, the number of spliced mRNA and products generated may vary from one retrovirus to another. For example, in the complex lentivirus, HIV-1, six different spliced mRNAs code for seven different products including the Env and other regulatory and accessory proteins; Vif, Vpr, Vpu, Tat, Rev and Nef (Figure 8). Simple beta-retroviruses like MMTV and MPMV generate minimal number of spliced mRNAs. In MPMV, the only spliced mRNA product is that of its Env protein while in MMTV, there exists three different spliced products, the *env, sag,* and *rem* spliced mRNAs that code for the Env, the superantigen (Sag), and Rem proteins respectively (Figure 8).
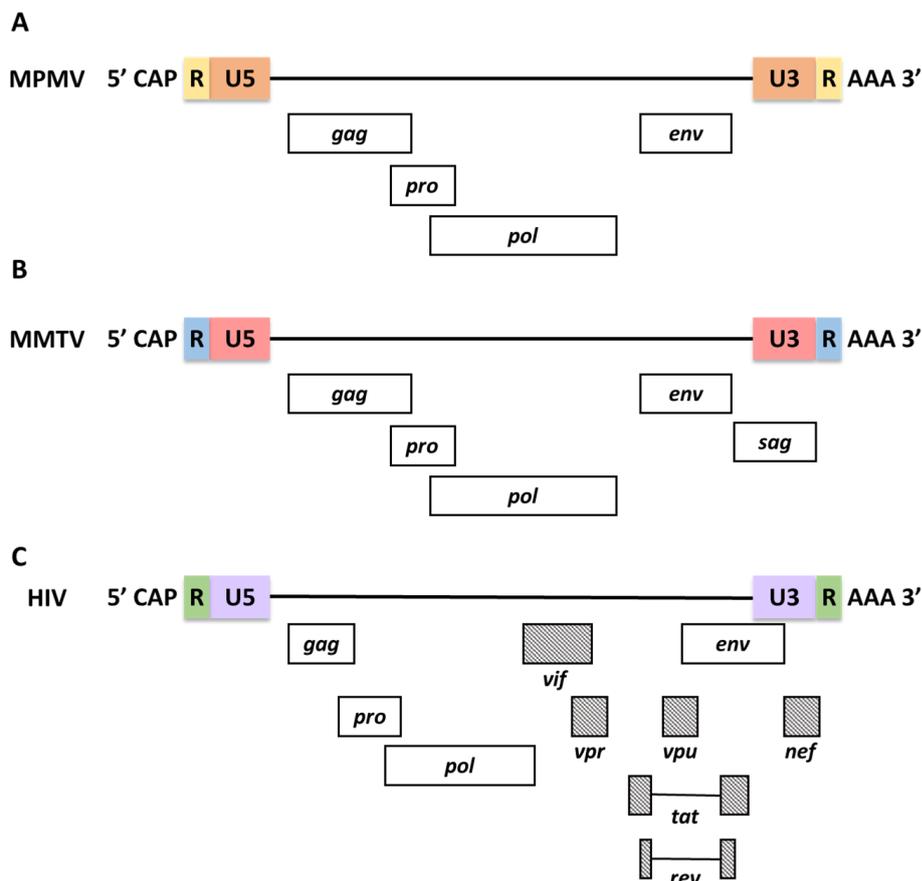
Figure 8: Genetic organization of prototypic beta-retroviruses and lentiviruses.

Prototype examples from the several genera of retroviruses are shown. An open rectangle indicates the open reading frame for the gene marked. Rectangles that appear vertically offset indicate that their reading frames are different. Horizontal lines connecting two rectangles indicate that this segment is spliced out. (A) MPMV, Mason-Pfizer monkey virus (B) MMTV, mouse mammary tumor virus; (C) HIV, human immunodeficiency virus type 1.

The modified full length and spliced mRNAs of the virus are exported out of the nucleus with the help of either virally encoded or cellular export factors depending on the virus (Figure 7F). In the case of HIV-1, the virally encoded spliced product, Rev, interacts with a highly structured element at the 3' end called rev responsive element (RRE, present in both spliced and unspliced RNA transcripts) to bring about their export from the nucleus to the cytoplasm. Rev binds in the form of multimers to a high affinity Rev binding site on the RRE (Fernandes et al., 2012; Pollard & Malim,

1998) . In HTLV-1, the viral protein homologous to Rev is Rex and binds to a Rex Responsive Element (RexRE; Nakano & Watanabe, 2012). Recently, in the case of MMTV it has been reported that a structured element at the 3' end of the genome called Rem responsive element (RmRE) interacts with virally encoded Rem protein to bring about the nuclear export of unspliced viral RNA and efficient expression of viral proteins (Akhlaq et al., 2018). On the other hand, MPMV harbors a constitutive transport element (CTE) located downstream of the *env* and overlapping the 3'LTR. The CTE interacts with the cellular factors TAP (NXF1) to bring about nuclear export (Bray et al., 1994a; Rizvi et al., 1996a; 1996b; 1997).

1.4.4 Translation

Once exported out of the nucleus and into the cytoplasm the viral transcripts are translated into their respective proteins using the cellular translational machinery (Figure 7G). The *gag, pr*o and *pol* genes are all transcribed in the form of a single full-length or unspliced mRNA transcript and translated into their precursor proteins by either termination suppression or ribosomal frameshifting. The latter mechanism is more common in most retroviruses (Coffin et al., 1997) . In HIV-1, a single -1 ribosomal frameshift at the 3' end of the *gag* gene shifts the open reading frame from one that codes for the precursor protein Gag to one that codes for Gag-Pol (Coffin et al., 1997; Figure 5). The exact nature of these frameshifts and reading frames differ among some genera of retroviruses. In the case of avian sarcoma leukosis virus (ASLV), this single frameshift is located at the 3' end of *pro* and this gives rise to two fusion products Gag-Pro and Gag-Pro-Pol (Coffin et al., 1997; Figure 9). In MPMV, MMTV (Figure 9) and HTLV-1 two -1 frameshifts; one at the 3' end of *gag* and the

other at the 3' end of *pro*, results in the generation of three fusion proteins: Gag, Gag-Pro and the full-length Gag-Pro-Pol (Coffin et al., 1997; Sonigo et al., 1986).
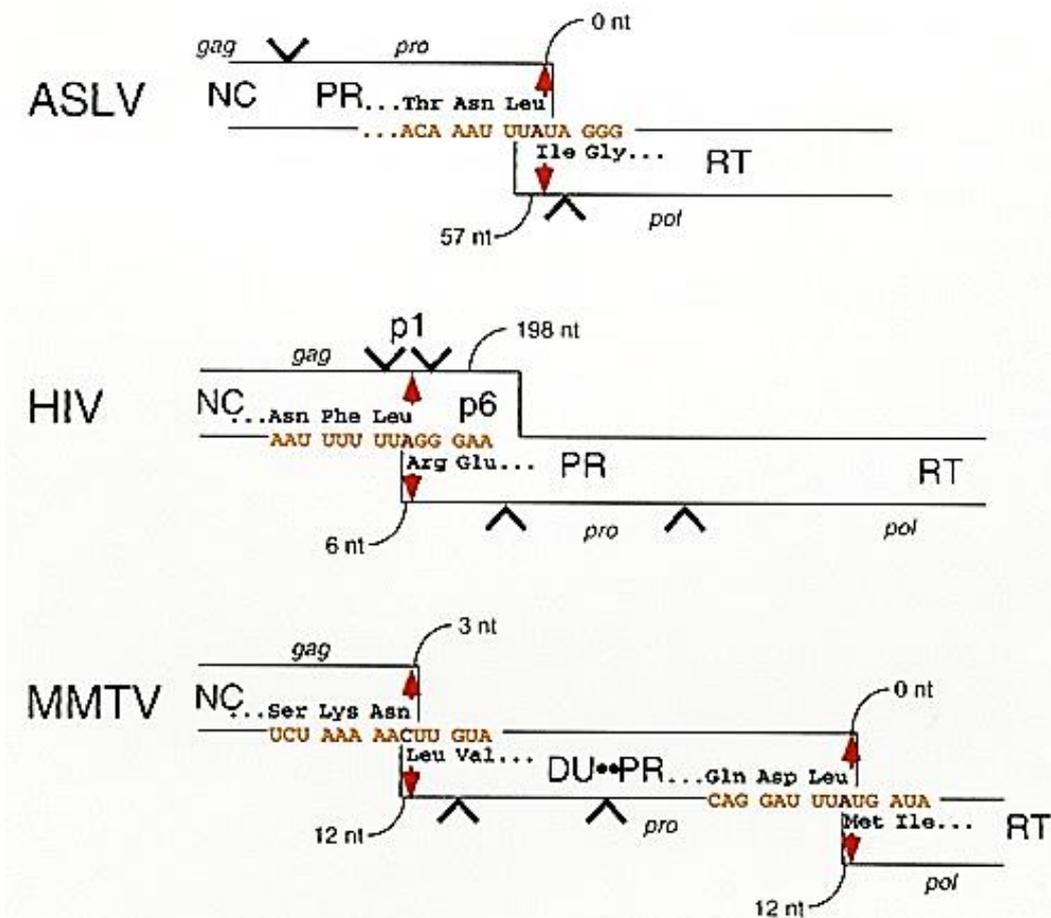
Figure 9: Frameshift suppression in the synthesis of Gag-Pro-Pol.

Shown are the nucleotide sequences at the frameshift site and the amino acids encoded in the Gag-Pro-Pol precursors of the indicated viruses. The upper amino acid sequence is read from either the *gag* or *pro* reading frame and the lower sequence is read from either the *pro* or *pol* reading frame, as shown. The boxes represent the indicated reading frames; the red colored arrows indicate the position of the nucleotide (shown in red) that is read in both reading frames, the Vs represent the positions in the sequence that encode PR processing sites, and the numbers represent the number of nucleotides between the frameshift site and the end of the reading frame (exclusive of the nucleotide that is read in both frames). The nucleotides in the boxes include the beginning and ending codons in the reading frames shown. Figure and legend adapted and modified from Coffin et al., 1997.

1.4.5 Unspliced Full Length RNA: Packaging *versus* Translation

Since the unspliced RNA of the virus functions as both the transcript for Gag-Pro-Pol expression as well as the source of the packageable gRNA during viral assembly, it is thought that there exists some sort of control or mechanism that allows the full-length RNA to fulfill its dual function without any mishaps to the viral life cycle (Figure 7G & H). For some retroviruses such as murine leukemia virus (MLV) its unspliced gRNA is maintained in two different populations with different half-lives within the cytoplasm and they carry out both these functions independently, while in HIV-2 gRNA is packaged co-translationally (Balvay et al., 2007; Butsch & Boris-Lawrie, 2002; Griffin et al., 2001; Kaye & Lever, 1998; Soto-Rifo et al., 2014). However, in HIV-1 these two functions are not exactly mutually exclusive. A number of conformational models of the HIV-1 *Psi* depending on different long-range interactions (LRIs) have been proposed and validated that explain the fate of its full-length transcript. A well-known model is the branched multiple hairpin (BMH) model in which U5 sequences base-paired with Gag AUG making U5-AUG LRI. Such a process results in exposing the DIS so that it can efficiently initiate the process of RNA dimerization and consequently packaging. Such an LRI conformer also makes the Gag AUG unavailable for translation initiation. This model has been experimentally validated and has been shown that the U5-AUG interaction is extended to include sequences of the mSD and hence form a tandem three-way junction (Lu, Heng, & Summers, 2011). An alternative structure that exists in equilibrium with the BMH model is one where there exists a long-distance interaction (LDI) between the U5 and the DIS present on SL1. The base pairing of the DIS with the U5 in LDI model renders the DIS unavailable for RNA dimerization and packaging while the AUG remains

exposed to initiate translation (Lu, Heng, & Summers, 2011; Rein, 2019; Van Bel et al., 2014). A more recent study implies that the number of G residues at the 5' cap of the viral transcript may also play a role in determining the fate of the viral transcript. Interestingly transcripts with only one G residue at its 5' cap adopts the U5-AUG conformation where the 5'cap is sequestered from translation initiation factors and the transcript itself adopts a multiple hairpin structure with an intact DIS for dimerization. On the other hand, transcripts with two or three G residues at the 5' cap adopt the alternative structure with the U5- DIS LRI that exposes both the mSD and the 5' cap for translation initiation (Ali et al., 2016; Brown et al., 2020; Keane et al., 2015; Mailler et al., 2016) . Interestingly, Rous sarcoma virus (RSV) has a different way in determining the fate of its unspliced gRNA for either gRNA packaging or translation of Gag. In RSV it is possible that different populations of unspliced gRNA are 'marked' within the nucleus to undergo different fates. Some of them are co-transcriptionally marked with splicing machinery and hence have their fates sealed for splicing and the production of spliced viral RNAs. Those that are not marked as such would serve as full-length viral gRNAs for encapsidation (Parent, 2011). At first, it is likely that unspliced RNA is trafficked out of the nucleus to serve as mRNA and translated by the cellular translational machinery to produce Gag particles. Interestingly a number of studies have shown that RSV Gag takes an unusual pathway and enters the nucleus (Baluyot et al., 2012; Lochmann et al., 2013; Parent, 2011). Recent single-molecule fluorescent in situ hybridization (smFISH) studies have shown that the initial interaction of RSV Gag and its gRNA may happen in the nucleus. Upon Gag translation in the cytoplasm, a small amount of RSV Gag is shuttled back into the nucleus, binds to the unspliced gRNA, and exports the gRNA out of the nucleus for packaging (Maldonado et al., 2020).

1.4.6 Assembly, Budding, and Release

Once the necessary proteins are translated, viral assembly takes place, this happens at either the plasma membrane as depicted for all lentiviruses/C-type retroviruses or in the cytoplasm/pericentriolar region as depicted for the beta-retroviruses, MMTV and MPMV (Bush & Vogt, 2014; Prchal et al., 2013; Zhang et al., 2015). However, the initiation of viral assembly requires specific sequence identification and interactions between the Gag polyprotein and the full-length, unspliced gRNA. This interaction is brought about by the recognition of specific *cis*-acting packaging sequences present at the 5' end of the dimeric genome by Gag (Ali et al., 2016; Comas-Garcia et al., 2016; Dubois et al., 2018a; Maldonado & Parent, 2016).

In C type retroviruses such as HIV-1, the expressed Gag molecules and Gag-Pol precursor polyproteins, containing MA, CA, NC, p6, spacer peptides (SP), PR, RT, and IN are transported to the plasma membrane for assembly, particle budding and release (Figure 7I). To initiate viral assembly, the NC domain of Gag recognizes and binds to the packaging sequences on the unspliced dimeric viral gRNA. The Gag-RNA complex is then transported to the sites of assembly on the plasma membrane for packaging into the assembling virus particles. At the sites of assembly, Gag binding to the plasma membrane is mediated by the N-terminal myristoyl group (myr) and the highly basic region (HBR) of the MA (Freed, 2015; Murphy & Saad, 2020; Olety & Ono, 2014). Gag, Gag-Pol, and viral RNA at assembly sites nucleate the formation of a lattice like structure, driven by intermolecular Gag-Pol interactions between the CA and SP1 domains of Gag. Gag lattice manifests as a cage of interconnected hexamers that covers most of the inner surface of the viral envelope providing a scaffold for viral particle assembly. Studies have established that non-specific electrostatic and

hydrophobic interactions, protein multimerization, cellular and viral RNA and recognition of specific plasma membrane phospholipids regulate Gag association with membranes. As the lattice grows, the membrane curves outward leading to the formation of an immature virus particle with only a narrow stretch of membrane connecting the viral envelope to the cellular plasma membrane (Alfadhli & Barklis, 2014; Dick & Vogt, 2014; Novikova et al., 2018). The immature lattice forms a spherical shape in which the Gag proteins are arranged radially with MA on one end and p6 on the other end pointing toward the center. The p6 domain recruits the cellular endosomal-sorting complex required for transport (ESCRT), which catalyzes the budding or release of the particle from the membrane (Votteler & Sundquist, 2013).

An addition to the budding viral particle is the Env glycoprotein which is essentially located on the plasma membrane. Spliced *env* mRNA is translated on the rough endoplasmic reticulum (RER) and is shuttled to the plasma membrane via the secretory pathway (Figure 7J). In HIV-1 the Env glycoprotein is first synthesized as a precursor polyprotein gp160, undergoes glycosylation in the Golgi network, and undergoes proteolytic cleavage by cellular proteases into its constitutive SU glycoproteins (gp120) and the TM glycoprotein (gp41) (Checkley et al., 2011; Murphy & Saad, 2020). The two subunits remain non-covalently bound in a trimer of heterodimers that protrudes from the virion surface where they may serve as glycosylated Env spikes during the budding process (Murphy & Saad, 2020). In the beta-retroviruses or the D type retroviruses such as MPMV viral particle assembly occurs in the cytoplasm as opposed to at the plasma membrane as seen in C type retroviruses. Once fully assembled in the cytoplasm the intact viral core is transported to the plasma membrane where budding and release of the enveloped viral particle takes place.

1.4.7 Maturation

After the virus particle is released, it undergoes a process called maturation to become infectious. Viral maturation involves the proteolytic cleavage of the Gag polyprotein into its constituent protein subunits. This is brought about by the viral protease (PR) that is packaged into the viral particle during assembly (Ganser-Pornillos et al., 2012). During or immediately after budding autoproteolytic cleavage of the protease makes it functional. The HIV-1 PR cleaves HIV-1 Gag at five different sites in a sequential order that brings about a sequential morphological change in the virus structure. These sites are cleaved at different rates with the CA being the final domain released (Figure 10). This cleavage event is a highly ordered and controlled process and any mutations in Gag that disrupt these cleavage sites or alter the processing order result in particles that have severely reduced infectivity. Virions arrested at the late budding stage display immature morphology, implying that budding and maturation are intimately coupled events (Ganser-Pornillos et al., 2012).
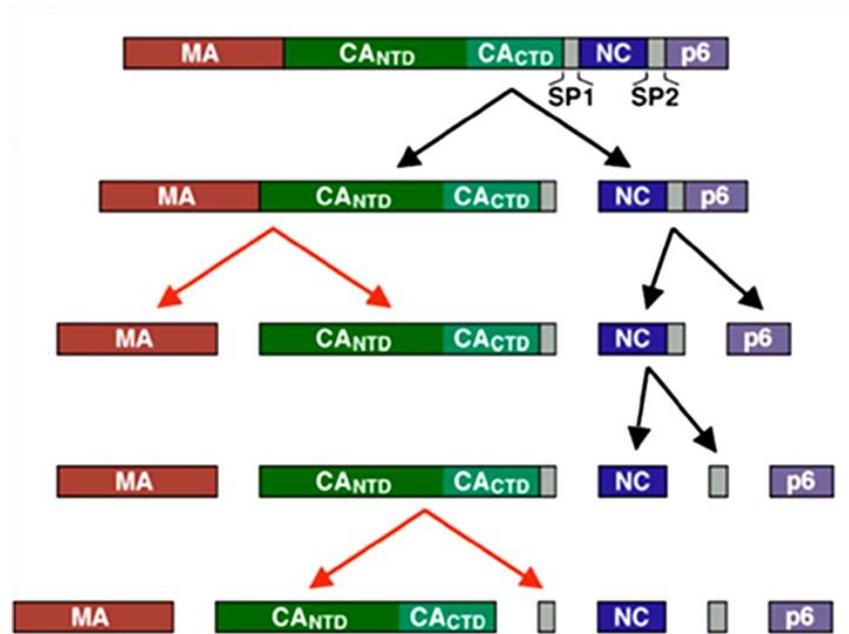
Figure 10: Proteolytic processing of HIV-1 Gag during maturation.

Schematic showing the HIV-1 Gag proteolysis cascade. Cleavage events that generate the mature CA termini are indicated by the *red arrows*. Figure and legend adapted from Ganser-Pornillos et al., 2012.

1.5 Retroviral Gag-gRNA Interactions

Assembling retroviral particles package a dimeric form of their gRNA as a rule and the specific interactions between Gag and gRNA is central to this process. gRNA is specifically selected from a vast array of spliced viral and cellular RNAs within the cytosol, this is a highly selective process since the gRNA comprises only ~1% of the cytosolic RNA. Successful packaging of dimeric gRNAs into retroviral particles depends on the presence of unique Gag binding sites within the sequences that are required for gRNA encapsidation (Abd El-Wahab et al., 2014; Ali et al., 2016; Berkowitz et al., 1996; Bernacchi et al., 2017; Dubois et al., 2018a; Maldonado & Parent, 2016; Kuzembayeva et al., 2014). A number of both *in vivo* and *in vitro* Gag binding studies have been carried out on different retroviruses with a majority of them performed on HIV-1. These studies identify the specific binding sites important to

gRNA packaging and majority of these studies carried out so far have employed either the NC domain of Gag (in its recombinant or mature form) or a truncated form of Gag (Amarasinghe et al., 2000; De Guzman et al., 1998; Ding et al., 2020; D'Souza & Summers, 2004; Gherghe et al., 2010; Kutluay et al., 2014; Olson & Musier-Forsyth, 2019; Wilkinson et al., 2008; Zhou et al., 2007). The NC domain of retroviral Gag harbors the zinc finger motifs that are critical in bringing about specific binding and packaging of gRNA over spliced and cellular RNAs in the cytoplasm. Studies that have assessed this property of the NC domain have shown that tampering with these zinc finger motifs or its N terminal basic residues tend to reduce binding to the packaging sequences (Mitra et al., 2013; Wu et al., 2013; Zhang & Barklis, 1997). While a number of *in vivo* and *in vitro* binding studies have been conducted using the NC domain, which is central to specific gRNA binding and selection, recent studies have shed light on the importance of studying Gag binding to gRNA in the context of full-length Gag. These studies indicate important roles for the other domains of Gag in selective gRNA binding and packaging, these include the MA, the p2 and p6 domains of HIV-1 and the CA-CTD (Alfadhli et al., 2011; Dubois et al., 2018b; Guo et al., 2005; Jones et al., 2011; Lu, Heng, & Summers, 2011; Pachulska-Wieczorek et al., 2016; Parent & Gudleski, 2011; Webb et al., 2013; Wu et al., 2018; Rice et al., 2020). One such domain is the MA domain. The MA is primarily responsible for plasma membrane targeting of the assembled retroviral particles and a number of recent studies have shown that this domain binds to a range of cellular tRNAs which may aid in this function (Olson & Musier-Forsyth, 2019). Interestingly this domain also plays a role in the specific selection of *Psi* RNAs over non-*Psi* RNAs. Retroviral Gag lacking the MA domain have shown a reduction in the selective packaging of gRNA and loss in the basicity of N terminal highly basic region (HBR) of the domain

has been particularly implicated for the reduced RNA packaging in the context of Gag. However, the involvement of HBR domain is not residue specific and heterologous substitution of these residues with other basic amino acid residues retained its ability to discriminate between gRNAs over other (spliced and cellular) RNAs (Olson & Musier-Forsyth, 2019; Rye-McCurdy et al., 2016). This suggests that the HBR domain in Gag MA has a role to play in selective gRNA packaging, however this domain has not been found to specifically bind to gRNA and involves only a non-specific interaction with the gRNA (Alfadhli et al., 2011; Chukkapalli et al., 2010; Comas-Garcia et al., 2017; Kroupa et al., 2020; Lingappa et al., 2014; Rye-McCurdy et al., 2016). Interestingly, while most retroviruses rely on NC as the major selective packaging domain, for some delta-retroviruses such as HTLV-1 and HTLV-2, their MA domains show a greater chaperone/binding affinity than their NC domains and are implicated in contributing directly to gRNA encapsidation rather than membrane targeting (Sun et al., 2014; Wu et al., 2018). The p2 domain of HIV-1 Gag has been shown to significantly improve HIV-1 gRNA packaging in a chimeric viral system where the p2 domain of HIV-1 was added to a chimeric HIV-2 Gag (whose NC was replaced with that of HIV-1), implicating a vital role for the p2 peptide in specific gRNA encapsidation (Kaye & Lever, 1998). The p6 domain of HIV-1 Gag has been shown to be important in viral assembly and regulating specific gRNA packaging and like the MA domain, the p6 domain too contributes by discriminating between spliced and unspliced RNA. Deletion of p6 from the Gag precursor protein indicate a significant decrease in the binding affinities of such mutant to gRNA *versus* spliced viral and cellular RNAs indicating its importance in binding specificity during gRNA packaging (Dubois et al., 2018b). Since p6 domain is known to be acidic with an overall negative charge, it is proposed that this allows it to bend over and interact with

the basic or positively charged zinc fingers partially masking its strong positive charge and exerting some sort of steric selection allowing for gRNA specific packaging (Dubois et al., 2018b).

Packaging sequences form a key element of gRNA encapsidation in retroviruses. It comprises of the unique nucleotide sequences or Gag binding sites that are indispensable for gRNA packaging, and are usually located on the 5' untranslated region (UTR) of the genome and extend into the 5' sequences of *gag*. Packaging sequences on almost all retroviruses, due to their dynamic nature, are found to fold into unique higher order secondary structures comprising of several structural motifs (Ali et al., 2016; Dubois et al., 2018a). Maintaining this secondary structure of the packaging sequences is crucial for efficient gRNA packaging and loss in this higher order structure has often been attributed to loss of gRNA packaging. *In vitro* and *in vivo* NC/Gag binding studies carried out on the retroviral gRNA revealed that these binding sites are prominently purine rich sequences i.e. either adenosine (A) or guanosine (G) (Abd El-Wahab et al., 2014; Lever, 2009; Moore et al., 2009; Moore & Hu, 2009; Paillart et al., 1997). While the popularly studied HIV-1 Gag indicates a preference to G residues the selective binding of Gag exclusively to purine rich regions is however not universal as some retroviruses such as MLV and RSV also identify uracil (U), a pyrimidine, as part of their binding sites (Gherghe et al., 2010; Zhou et al., 2005, 2007) which reflects the difference in Gag binding among different genera of retroviruses. However, i*n vivo* cross-linking immunoprecipitation (CLIP) studies carried out to identify HIV-1 Gag binding to RNA within the infected cellular environment reveals a transition in Gag's nucleotide preference to gRNA during the course of the viral assembly. Interestingly when binding to its gRNA in the cytosol, Gag tends to preferentially bind to G/U regions and when associated with membranes

or in immature form it prefers G/A (Kutluay et al., 2014). Interestingly for most retroviruses studied to date, these NC/Gag binding sites are found located in close proximity to the DIS on their gRNA, this is however not surprising as a number of mutational studies carried out to structurally and functionally mutate their DIS indicate a concomitant reduction in gRNA packaging efficiencies denoting that both these processes are interdependent. Dimerization is often a prerequisite to packaging because kissing loop interactions between the two copies of the RNA leads to conformational changes in the higher order structure exposing the otherwise sequestered Gag binding sites facilitating RNA-protein interaction(s) needed for gRNA packaging (Ali et al., 2016; D'Souza & Summers, 2005; Dubois et al., 2018a; Gherghe et al., 2010; Johnson & Telesnitsky, 2010; Keane et al., 2015; Keane & Summers, 2016; Olson et al., 2015).

Studies carried out on HIV-1 have shed a lot of light on the dynamics of selective Gag-gRNA binding and packaging. The packaging sequences of HIV-1 spans from the 5'UTR (356 nts) to the first 120 nts upstream of the *gag* gene (Figure 6A). A number of mutagenesis studies have identified several motifs within this region to be important in gRNA packaging (Clever et al., 1995; Clever & Parslow, 1997; Hayashi et al., 1992; Luban & Goff, 1994; McBride & Panganiban, 1996, 1997; Skripkin et al., 1994). A number of distal regions have also been shown to affect gRNA encapsidation, such as the Gag-Pol frameshift signal (Chamanian et al., 2013) and the RRE (Cockrell et al., 2011), but to a lesser extent when compared to the 5' packaging sequences. As for other retroviruses the packaging sequences of HIV-1 folds into a hierarchical secondary structure containing a series of stem loops (Figure 11). A number of previous studies have been dedicated to probe and validate the secondary structure of the HIV-1 packaging sequences using a wide array of techniques ranging from

mutagenesis, phylogenetic analyses, biochemical enzymatic and chemical analyses such as SHAPE (Lu, Heng, & Summers, 2011). More recently, nuclear magnetic resonance (NMR) (Keane et al., 2015; Lu, Heng, & Summers, 2011) and techniques such as mass spectrometry (MS) (Yu et al., 2008), fluorescence resonance energy transfer (FRET) (Stephenson et al., 2013), and small angle X-ray scattering (SAXS) (Jones et al., 2014) have also been employed to characterize the 3D structure of the HIV-1 packaging sequences. Biochemical probing studies have been consistent in determining a secondary structure that consists of a series of stem loops in the following order: TAR, Poly A, PBS, and SL1 to SL4 (Figure 11).
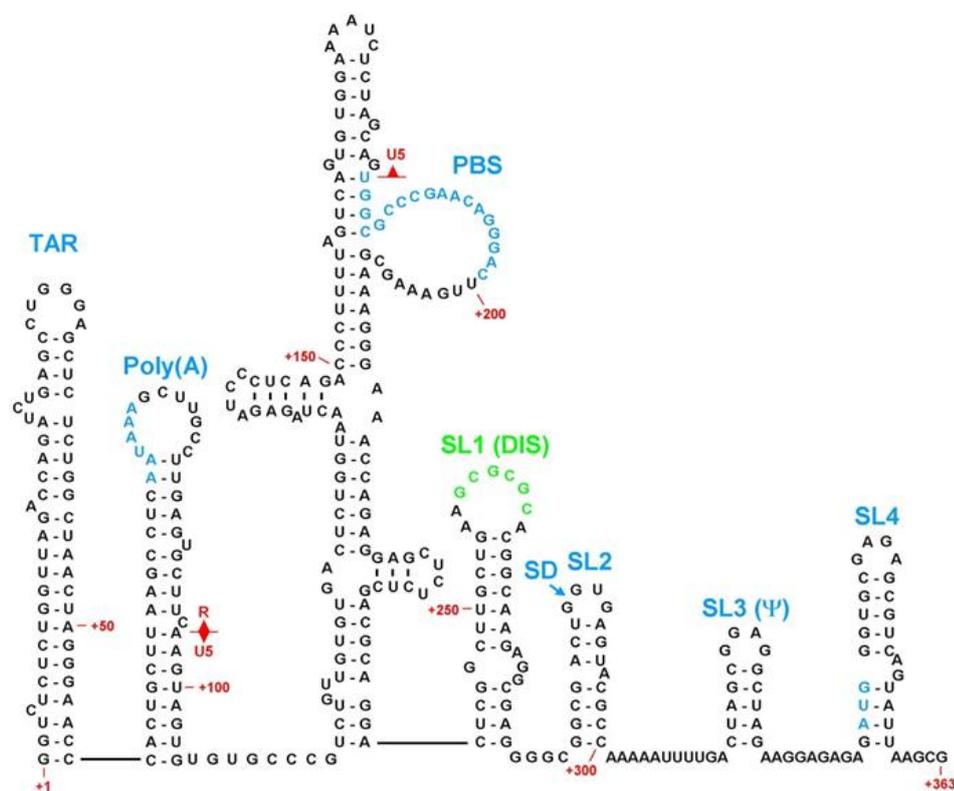
Figure 11: HIV-1 packaging signal RNA structural elements.

Illustration of a structural model of the HIV-1 5' UTR showing the various motifs important for virus replication. These are the TAR element, the poly(A) hairpin, and the U5-PBS complex. Stem-loops 1-4 containing the DIS, the major splice donor, the major packaging signal, and the *gag* start codon, respectively. Nucleotides and numbering correspond to the HIV-1 HXB2 sequence. (Figure and legned adapted from Clever et al. 1995 and Berkhout and Van Wamel, 2000).

Systematic deletion analyses of packaging sequences revealed that the TAR, Poly A and PBS stem loops were dispensable for efficient encapsidation and a (159 nt) core encapsidation signal (CES) that comprised only of the SL1-SL4 stem loops were sufficient for gRNA packaging. This CES was shown to maintain dimerization, NC binding and allow for more than 80% of *in vitro* packaging efficiency compared to the native packaging sequences on the gRNA (Heng et al., 2012).

Initial studies to identify the binding sites of HIV-1 NC to its packaging sequences indicated that NC bound to a number of motifs with the highest binding affinity observed to SL2/mSD and SL3 (Amarasinghe et al., 2000; De Guzman et al., 1998). For a long time SL3 was considered to be the primary Gag binding during gRNA packaging. NMR studies carried out by De Guzman (1998) and colleagues to decipher the NC-SL3 complex identified the G residues of the GGAG apical/tetraloop inserted into the hydrophobic pockets of both the zinc finger motifs of NC indicating a tight binding of SL3 to NC. In the case of SL2, the G residues of its apical GGUG loop were found to interact with NC in the same manner as the GGAG tetraloop of SL3 (Amarasinghe et al., 2000). The apical loop of SL1 that also functions as the DIS and SL4, have been shown to bind to NC with a lower affinity than SL2 and SL3 (Amarasinghe et al., 2000; De Guzman et al., 1998). However, all these studies involved the use of the mature protein NCp7 or the recombinant NC, which has a comparatively lower specificity and cooperativity than HIV-1 full-length Gag (Pr55[Gag]) to gRNA. Recent *in vitro* biochemical studies carried out using the full-length HIV-1 Pr55[Gag] have shown that a unique high affinity binding is located within an internal loop of the SL1 (Figure 11; Abd El-Wahab et al., 2014). This internal loop is made up of just four asymmetric purine nucleotides, GAGG and is interestingly located on SL1 that also harbors the HIV-1 DIS (Figure 11; Abd El-Wahab et al.,

2014). This unique binding site is located upstream of the mSD and hence is found in both spliced and unspliced RNA. However, mutational analysis carried out on both the upstream and downstream regions to this high affinity binding site reveal that the selective binding of Pr55[Gag] to the spliced RNA is negatively regulated by elements upstream to the SLI and that these elements are in turn counteracted by elements downstream of the SL4 (nts 355-400) that suppresses this negative regulation. It is likely that these two regulatory regions may function via a putative LRI that is yet to be identified (Figure 12; Abd El-Wahab et al., 2014).
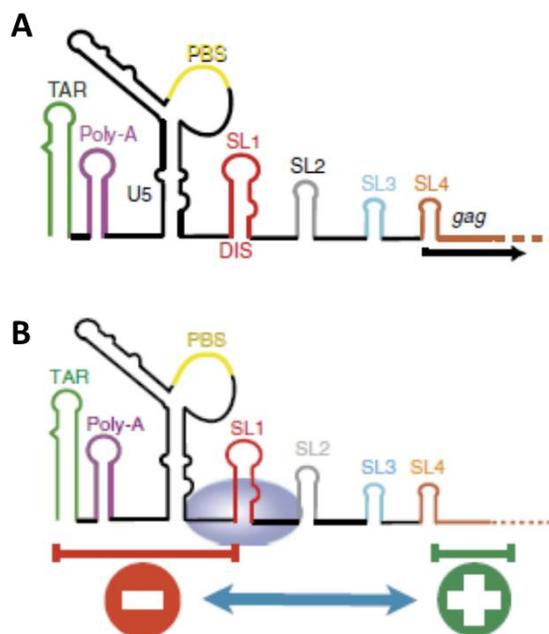
Figure 12: 5'-region of HIV-1 gRNA and how Pr55[Gag] binding to SL1 is regulated by upstream and downstream sequences.

Schematic representation of higher order structural model of the 5'-region of the HIV-1 gRNA. TAR, trans-activation response element; Poly-A, stem-loop containing the 5'-copy of the polyadenylation signal in the apical loop; U5, unique in 5'; PBS, primer binding site; DIS, dimerization initiation site; SL1–4: stem-loops 1–4. Schematic representation of Pr55[Gag] binding to HIV-1 gRNA. Pr55[Gag] binding to the lower part of SL1 is inhibited by the upstream sequences. This negative regulation is counteracted by a short sequence located 3' of SL4. Figure and legend adapted from Abd El Wahab et al., 2014.

1.6 Mason-Pfizer Monkey Virus

Mason Pfizer monkey virus (MPMV) is the most widely studied prototypic D type or beta-retrovirus. It was first isolated from the breast adenocarcinoma of a rhesus monkey (*Macaca mulatta*) and is known to cause an autoimmune deficiency (AID)-like syndrome in Rhesus monkeys (Bryant et al., 1986; Fine et al., 1975). MPMV has been a widely studied simian retrovirus in terms of both retroviral gRNA packaging and assembly (Aktar et al., 2013; Ali et al., 2020; Bush & Vogt, 2014; Guesdon et al., 2001; Harrison et al., 1995; Jaballah et al., 2010; Kalloush et al., 2016, 2019; Schmidt et al., 2003; Vile et al., 1992). It differs distinctly from C-type retroviruses such as HIV-1 by having an intracytoplasmic A-type assembly pathway with spherical capsids. The viral particles acquire a D type morphology with cylindrical core upon viral release and maturation (Bush & Vogt, 2014; Coffin et al., 1997).

Retroviruses have long been extensively studied for their ability to reverse transcribe and integrate their genome into host cells, a feature which makes them promising candidates for the development of gene therapy vectors. Designed carefully they can deliver therapeutic genes into the host genome ensuring their stable long-term expression. Their genomes have large coding capacities allowing the expression of most genes of interest and also importantly the transduction of only the therapeutic gene and not their own genes, thus avoiding host humoral/cellular responses to viral antigens that can eliminate the transduced cells. Consequently, it is not surprising that approximately a quarter of all human gene therapy trials have used retroviral based vectors (Ginn et al., 2013; Hu et al., 2011). Among retroviruses, currently, HIV-1 based vectors are being exploited; however, their use in humans is not very promising due to critical safety concerns such as the generation of replication competent virus through recombination with human endogenous retroviruses. However, MPMV may

tend to serve as an ideal candidate for the development of gene therapy vectors mainly due to its phylogenetically diverse nature from that of human retroviruses such as HIV-1. Importantly the MPMV CTE, a short 219 nt sequence on the 3' end of the genome, analogous to the HIV-1 RRE, allows for efficient cytoplasmic transport of viral RNA uniquely independent of any expressed viral proteins and dependent on only cellular factors (Bray et al., 1994; Pasquinelli et al., 1997; Rizvi et al., 1996a; 1996b, 1997; Zolotukhin et al., 2001). The CTE of MPMV has also been successfully used to significantly increase the titer of MLV based vectors (Zhao et al., 2000). However, it is crucial that the relevant aspects of MPMV replication are fully understood before MPMV based vectors can be utilized for human gene therapy.

1.7 MPMV Gag Polyprotein: Pr78^{Gag}

The retroviral Gag polyprotein, coded by the *gag* gene, consists of the key structural elements of the virus particle and its assembly process. This assembly process is mainly driven by the three major domains of the polyprotein; the MA, CA and NC, found ubiquitously in all retroviruses. Despite the minimal sequence homology among these domains in different retroviruses, they remain structurally and functionally similar to each other (Ali et al., 2016). The Gag of different retroviruses may differ in additional domains that may or may not be present in each of them, such as, the spacer peptides (e.g. SP1 and SP2) in HIV-1, the p domains (e.g. p6 in HIV-1) and the PR domain in RSV. Interestingly HTLV-1 has no extra domains in its Gag. From the time of synthesis in the cytoplasm to viral release and maturation the various domains of Gag drive the assembly process and are responsible for more than one of the key events in the viral life cycle (Bush & Vogt, 2014). Although these domains are mostly studied independent of each other their various functions are achieved within

the context of the polyprotein throughout its existence within the cell and are only liberated from one another post-release from the cell during maturation.

While a multitude of studies have been conducted on HIV-1 Gag and the functionality and structure of its domains widely elaborated, there have not been as many studies on the Gag determinants of other retroviruses, MPMV included. However, the structural and functional conservation among these domains for the few other retroviruses that have been studied allows us to safely translate the data and findings of HIV-1 to other less studied retroviruses such as MPMV. In the following section while most of the functional and structural aspects of the Gag domains refer to those of MPMV, wherever possible similarities and comparisons have been drawn with other retroviruses.

1.7.1 Matrix (MA)

The MPMV Gag precursor polyprotein (Pr78^{Gag}) is composed of six different domains; NH_2-p10 (MA), pp24/16, p12, p27 (CA), p14 (NC), and p4-COOH of which the MA is the foremost N terminal domain (Figure 13A; Bradac & Hunter, 1984; Henderson et al., 1985; Sonigo et al., 1986). Retroviral MA is primarily responsible for plasma membrane targeting and binding the viral ribonucleocapsid to the plasma membrane at the latter stages of viral assembly. This is brought about by the N terminal domain of the protein, which is highly basic owing to the myristic acid moiety, and basic amino acid residues found around its vicinity. The MA is co-translationally myristoylated and during virion assembly, this domain becomes anchored to the inner leaflet of the plasma membrane aiding in Gag oligomerization (Alfadhli & Barklis, 2014; Alfadhli et al., 2009). The MA is also capable of binding to nucleic acid owing to its HBR domain located at its N terminal. Deletion of the HBR or mutation of the

basic lysine (K) residues in the region has shown to abolish MA-nucleic acid interactions (Alfadhli et al., 2011; Cai et al., 2010; Inlora et al., 2014). Interestingly the HBR domain is known to play a role in the selection of gRNA for packaging (Comas-Garcia et al., 2017; Kroupa et al., 2020; Rye-McCurdy et al., 2016).

Like for other retroviruses the MPMV MA also undergoes N terminal myristoylation required for plasma membrane targeting (Figure 13A). The absence of such an N terminal myristic acid has been shown to prevent translocation of assembled virus particles to the membrane. However, the presence of an N terminal myristic acid alone is insufficient for appropriate plasma membrane translocation and requires a second signal which is brought about by the two arginine (R) and K rich helices also located at the N terminal end of the MA protein. The three dimensional solution structure of the MA protein of both HIV and MPMV are almost similar to each other except in this aspect where HIV-1 has only one of these helices (Prchal et al., 2012). Importantly in MPMV one of these helices consists of an 18 amino acid basic region called the cytoplasmic transport/retention signal (CTRS) that is responsible for the typical D type or cytoplasmic assembly of MPMV (Choi et al., 1999). The CTRS is also responsible for the proper transport of the viral CA to the plasma membrane. The mutation of a single amino acid (R55W) within this region has resulted in drastic changes in both of these functions (Figure 13A & B; Rhee & Hunter, 1990). Introduction of this region into the MA protein of MLV resulted in intracytoplasmic assembly of the virus particles as opposed to its typical C type assembly at the membrane (Choi et al., 1999; Rhee & Hunter, 1991; Rhee & Hunter, 1990).

A

RKK
MHR

KKPKR
Np24

PPPY/PSAP
pp16/18

NTD CTD CCHC

Myristic acid

HO

O

MA (p10) | pp24 | p12 | CA (p27) | NC (p14) | p4

O
C
OH

CTRS

ISD

SLD

CCHC

KNKEK

B

CTRS

M G Q E L S Q H E R Y V E Q L K Q A L K T R G V K V K Y A D L L K F F D F V K D T C **P W F P Q E G T I D I K R**[55] **W R R V G** D C F Q D Y Y N T

F G P E K V P V T A F S Y W N L I K E L I D K K E V N P Q V M A A V A Q T E E I L K S N S Q T D L T K T S Q N P D L D L I S L D S D D E G A K S S S L

L domain

Q D K G L S S T **K K P K R** F P V L L T A Q T S K D P E D P N P S E V D W D G L E D E A A K Y H N P D W P P F L T R **P P P Y** N K A T **P S A P** T V

ISD

M A V V N P K **E E L K E K I A Q L E E Q I K L E E L** H Q A L I S K L Q K L K T G N E T V T H P D T A G G L S R T P H W P G Q H I P K G K C C A S

R E K E E Q I P K D I F P V T E T V D G Q Q G A W R H H N G F D F A V I K E L K T A A S Q Y G A T A P Y T L A I V E S V A D N W L T P T D W N T L

V R A V L S G G D H L L W K S E F F E N C R D T A K R N Q Q A G N G W D F D M L T G S G N Y S S T D A Q M Q Y D P G L F A Q I Q A A A T K A

MHR

W R K L P V K G D P G A S L T G **V K Q G P D E P F A D F V H R L I T T A** G R I F G S A E A G V D Y V K Q L A Y E N A N P A **C**[193] Q A A I R P

Spacer like domain

Zinc finger 1

Y **R K K** T D L T G Y I R L **C**[213] S D I G P **S Y Q Q G L A M A A A F S G Q T V K D F L N N** K N K E K G G C **C F K C G K K G H F A K N C**

Zinc finger 2

H E H A H N N A E P K V P G L **C P R C K R G K H W A N E C** K S K T D N Q G N P I P P H Q G N G W R G Q P Q A P K Q A Y G A V S F V P A N

K N N P F Q S L P E P P Q E V Q D W T S V P P P T Q Y

Figure 13: Graphical representation of MPMV polyprotein Pr78[Gag] and its constituent domains and functional motifs important for viral assembly.

(A) Schematic representation of Pr78[Gag] and its domains highlighting the important motifs. MA, matrix; CTRS, cytoplasmic transport retention signal; pp24, phosphoprotein 24; ISD, internal scaffold domain; CA, capsid; NTD, N-terminal domain; CTD, C-terminal domain; MHR, major homology domain; SLD, spacer like domain; CCHC, zinc finger domain. (B) Amino acid sequence of Pr78[Gag] with the important functional motifs annotated. CTRS, cytoplasmic transport retention signal; L, late domain; MHR, major homology domain.

1.7.2 Capsid (CA)

The retroviral CA forms the central nucleocore of the virus particle. It encloses the viral ribonucleoprotein protecting the viral genome. The protein forms essential CA-CA interactions to allow Gag oligomerization and eventually viral core assembly. These interactions also provide stability to the viral structure (Ganser-Pornillos et al., 2008). Despite minimal sequence homology among the CAs of different retroviruses, their secondary and tertiary structures are highly conserved (Campos-Olivas et al.,

2000; Gamble et al., 1997; Gitti et al., 1996; Macek et al., 2009; Momany et al., 1996; Mortuza et al., 2004). Retroviral CAs consist of two independently folded domains, the CTD and the NTD connected by a short linker peptide. CA-CTD structure is known to be conserved among a number of retroviruses while that of the NTD differs substantially (Bharat et al., 2012; Schur et al., 2013, 2015). An important feature of the CA-CTD is the presence of a major homology region (MHR; Figure 13A) which consists of 20 highly conserved amino acids notoriously known to be the most conserved region within the Gag polyprotein (Mammano et al., 1994; Schur et al., 2015). While the exact function of the MHR is not known, mutational studies indicate it to be important for viral assembly (Schwedler et al., 2003; Chu et al., 2006). In HIV-1, it was long thought that the CA-CTD solely contributes to viral assembly and that the NTD did not have much to do with the process. However, recent studies have shown that the involvement of CA-CTD and NTDs appear to transition during viral maturation and involve different interactions for oligomerization in both the immature and mature viral core. The immature viral core seems to be stabilized by intermolecular CA-CTD: CA-CTD interactions whereas the mature viral core involved intramolecular CA-NTD: CA-NTD and CA-NTD: CA-CTD interactions. Mutations in specific residues of both the NTD and CTDs tend to affect viral core stability.

In MPMV, a C terminal basic region referred to as the RKK region in the CA domain is capable of binding to nucleic acid, which possibly functions as scaffolding for cytoplasmic assembly of virus (Figure 13A & B). Mutations in this region was found to severely affect viral particle release, assembly, RNA packaging and nuclear trafficking (Füzik et al., 2016). Apart from this basic motif two cysteine residues in the CTD of the CA are found highly conserved among most retroviruses (apart from alpha and spumaretroviruses) and form intramolecular disulfide bonds in the

polyprotein and are important in maintaining the assembly and stability of both immature and mature cores. In HIV-1 these cysteine residues are C198 and C218 (Figure 13B). Mutating C198 resulted in the reduced infectivity with no effect on viral assembly and release while mutating C218 abrogated particle assembly (McDermott et al., 1996). Mutating these conserved cysteines (C193, C213) in the MPMV CA drastically reduced viral assembly and infectivity (Píchalová et al., 2018).

1.7.3 Nucleocapsid (NC)

The NC protein within the context of the polyprotein is responsible for the selective packaging of gRNA into virus particles. It is a highly basic and hydrophobic protein whose major feature is the presence of two highly conserved zinc finger domains. These domains consist of highly conserved CCHC arrays (C-X_2-C-X_4-H-X_4-C where C = Cys, H = His, X_n = n number of amino acids) that sequester zinc ions required for specific gRNA binding (Figure 13A & B). While both zinger finger domains are indispensable for successful gRNA packaging into virions they do not function equally. Studies have indicated that the proximal zinc finger domain plays a more significant role in selective gRNA packaging and mutations in it result in more drastic effects on packaging compared to that of the second domain, this holds true for MPMV too (Dostálková et al., 2018). The NC proteins and their zinc finger domains bring about specific gRNA packaging by identifying specific encapsidation signals on the gRNA hence acting as a chaperone for gRNA packaging during the viral life cycle. In HIV-1 while other domains such as the MA, CA, p2 spacer peptide and the p6 late domain interact with RNA to a lesser extent, the NC is widely known to be the most important domain that brings about selective gRNA packaging. This has been corroborated by a number of studies where mutations in the NC domain of several

retroviruses abrogates selective packaging (Aldovini & Young, 1990; Dorfman et al., 1993; Gorelick et al., 1988; Méric et al., 1988; Poon et al., 1996). This may also stand true for MPMV where while NC is the major gRNA binding domain it may not be the only domain capable of binding to gRNA as has been shown in HIV-1 and as evidenced in a previous study for MPMV (Füzik et al., 2016).

In addition to the highly basic zinc finger domains, the NC protein of many retroviruses also consist of stretches of basic amino acids flanking these domains that also play a role in specific gRNA binding and Gag oligomerization, proposed to be induced by conformational changes that occur in the polyprotein after Gag gRNA binding. One such cluster of basic amino acids is the $K^{16}NKEK^{20}$ region located upstream of the proximal zinc finger domain (Figure 13A & B; Dostálková et al., 2018). This region is shown to bind to the ssPurines region within the MPMV packaging sequences to bring about successful gRNA encapsidation with this becoming 4 times greater when the basicity of this region is increased by amino acid substitution of K by R. However, the loss of basicity in this region severely affects gRNA incorporation, localization of assembly and also the onset and yield of reverse transcription (Dostálková et al., 2018). In addition to this, the NC protein is also known to act as a chaperone to RT during both of the obligatory strand transfers in viral RNA replication (Muriaux & Darlix, 2010).

MPMV consists of a 'spacer like' sequence similar in functionality to that of the HIV-1 SP1. This region is located on the N terminal domain of the NC, consists of the first 15 amino acids and is found to be specific for MPMV, and is indispensable for viral assembly. Experiments indicate that this region could not be replaced by the HIV-1 SP1 sequence to rescue immature particle assembly (Bohmová et al., 2010) and instead found the Gag particles accumulated at the plasma membrane assembly

(Bohmová et al., 2014) . This region was followed by a non-specific stretch of basic residues (similar to the canonical Interaction (I) domain of retroviral NC) identified to be essential for the assembly function of the spacer like domain (Bohmová et al., 2010). Mutational analysis carried out at a later date identified that this spacer like domain could be extended into the CTD of the CA protein to include the last 8 amino acids of the CA protein and was additionally found to be important in viral release, processing and viral infectivity (Figure 13A & B; Bohmová et al., 2014).

1.7.4 Phosphoprotein 24 (pp24)

The pp24 domain of MPMV is highly conserved among all beta-retroviruses and is located immediately downstream of the MA domain. During viral maturation it is cleaved into two substituent protein components the pp16/18 and the Np24 (Figure 13A).

The pp16/18 comprises of the C-terminal cleavage product of pp24 and constitutes the late (L) budding domains. These domains are highly conserved among all enveloped viruses and are responsible for efficient budding of novel virus particles from infected cells. Unlike HIV-1, MPMV consists of two closely located L domains/ motifs, the PSAP and PPPY motifs, which have been found to be responsible for bringing about efficient budding in MPMV (Figure 13A & B). However, studies have shown that the PPPY motif exerts more of an effect than the PSAP motif with the latter requiring an intact PPPY motif for its efficient functioning. Budding deficient mutants with a PSAP deletion resulted in viral particle accumulation at the plasma membrane with a two fold reduction in budding while no sign of budding was observed for the PPPY mutant (Gottwein et al., 2003; Yasuda & Hunter, 1998).

The Np24 comprises of the N terminal domain and is shown to be important in gRNA packaging and viral replication. Deletion of a conserved basic region (KKPKR) within Np24 reduced packaging by 6 to 8-fold (Bohl et al., 2005). This deletion was also found to affect the Gag localization at the nuclear pore when compared to the wild type Pr78^{Gag}. It is suggested that the KKPKR motif in Np24 could act as a nuclear localization signal or it could simply target Pr78^{Gag} to the nuclear membrane in order to package gRNA prior to CA assembly (Figure 13A & B; Bohl et al., 2005).

1.7.5 p12

The p12 domain of MPMV is unique to the beta-retroviruses and consists of an acidic rich motif majorly made up of glutamic acid residues called the internal scaffold domain (ISD; Figure 13A). This ISD has been shown to act as a scaffold for Gag multimerization, enhancing immature viral assembly in a cell free system and in stable cell lines with levels of low protein in expression (Sommerfelt et al., 1992). However in transfected cell lines with high expression systems this does not seem to be the case, where the ISD deletions had no effect on assembly however this domain is known to be important in viral infectivity (Sommerfelt et al., 1992) and has also shown to allow for successful HIV-1 Gag multimerization and assembly in a cell free system when fused to HIV-1 Gag. (Sakalian et al., 2002; Sakalian & Hunter, 1999).

1.8 The MPMV Packaging Signal RNA

The *cis*-acting sequences of retroviruses have long been studied to identify potential minimal sequences that may be vital for viral gRNA packaging and propagation. These minimal sequences, referred to as the packaging signal (*Psi* /Ψ) have been found to be indispensable for the selective packaging of retroviral gRNA. For most of the retroviruses, these sequences are located in the 5'UTR and are continuous, beginning in the 5' end and spanning all the way into *gag* gene. For MPMV systematic mutational studies have identified essential sequences at its 5' UTR that are responsible for both packaging and dimerization of its gRNA (Aktar et al., 2013; Guesdon et al., 2001; Harrison et al., 1995; Jaballah et al., 2010; Kalloush et al., 2016; Mustafa et al., 2004; Schmidt et al., 2003; Vile et al., 1992). Further genetic studies carried out on the MPMV genome reveal that these sequences are situated in a discontinuous rather than continuous fashion as seen in other retroviruses. The multipartite nature of these sequences was first identified by Schmidt et al., and were found to be located in two distinct regions, namely region "A" and "B". Region "A" includes the first 30 nt of the 5' UTR, boundaries of which were expanded to include additional 20 nt DIS stem loop referred to as the palindromic (Pal) SL. Region "B" comprises the last 23 nt of the 5' UTR, including the mSD and is followed by the first 120 nt of *gag* (Jaballah et al., 2010; Schmidt et al., 2003).

Cross packaging studies carried out between MPMV and the lentiviruses simian immunodeficiency virus (SIV), HIV-1 and the B-type retrovirus MMTV, to investigate the potential ability of these sequences to function as packaging determinants in different viral packaging systems, revealed interesting results. It was found that the packaging signals of these viruses, despite minimal sequence homology,

are quite promiscuous in nature and are capable of being packaged into novel pseudo-viral particles by their reciprocal Gag. However it is worthy to note that despite successful packaging, propagation of these pseudo-viral particles were not observed indicating an effect/s on post packaging events (Al Dhaheri et al., 2009; Al Shamsi et al., 2011). These studies shed light on the fact that the packaging determinants of these viruses are identified at a hierarchical level rather than that at its primary sequence level by Gag. RNA secondary-structure prediction tools such as MFold revealed that these packaging determinants assume stable secondary structure (containing several structural motifs) whose structural conservation may be vital in packaging and propagation of the virus rather than their primary sequence. The predicted secondary structural motifs for MPMV have been experimentally validated via the novel biochemical probing technique called SHAPE and have been found to be in close conformity to its predicted structure (Figure 14; Jaballah et al., 2010; Aktar et al., 2013).

Mutational analysis of the structural motifs of the MPMV packaging signal RNA secondary structure revealed that several regions play vital roles in MPMV life cycle. One of these regions consists of a 6 nts palindromic sequence in Pal SL which functions as dimerization initiation site (DIS; Aktar et al., 2013). Dimerization is an important process in the retroviral life cycle and is known to be a prerequisite to gRNA packaging in most retroviruses. Deletion analysis of this region in MPMV revealed a significant reduction in the *in vitro* dimerization capabilities of the RNA (Aktar et al., 2013.
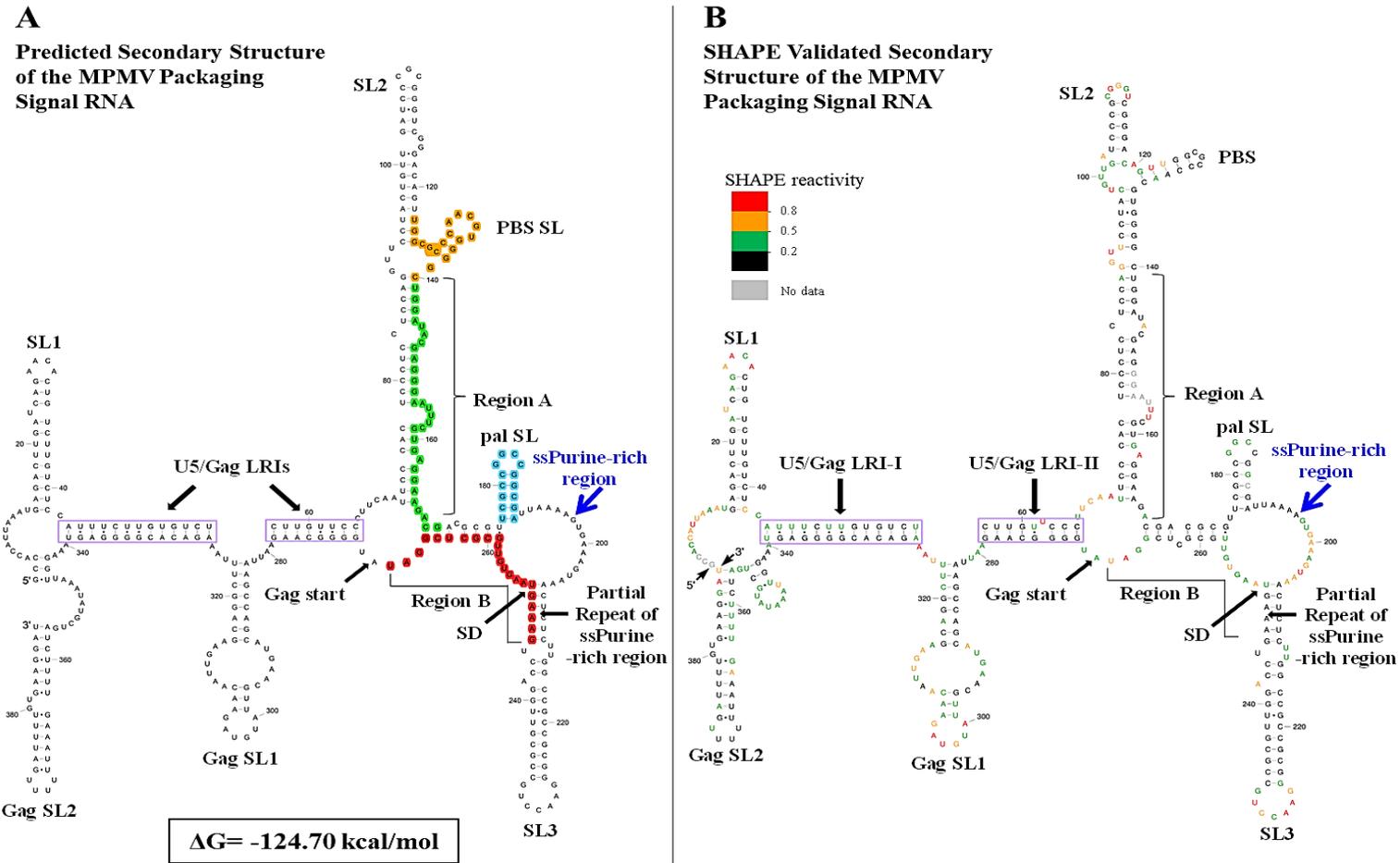
Figure 14: Minimal free-energy and SHAPE validated models of the MPMV packaging signal RNA.

The region used for analysis by Mfold and SHAPE included sequences from R up to 120 nts 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 and Mathews 2010) model of MPMV packaging signal. Nucleotides are color annotated as per the SHAPE reactivities key. SD indicates splice donor. Figure and legend adapted and modified from Aktar et al., 2013.

The other important structural motifs of the RNA secondary structure are two phylogenetically conserved LRIs that form between the U5-*gag* sequences. LRIs as such are a common feature of most retroviral RNA packaging sequences and unlike other retroviruses there exist two such LRIs for MPMV. The LRIs function by bringing about two distal regions of the packaging signal together to form essential SLs/structural motifs that contribute to various aspects of the packaging signal. They may also function by bringing together (or to bring in close vicinity) the essential Gag binding sites for packaging, or by forming structures important in maintaining the stability of these regions and the overall RNA secondary structure. Mutational and structure probing experiments carried out on MPMV LRIs reveal that proper structural maintenance or proper base pairing within these LRIs are important for successful packaging and propagation of the virus. While it was found that LRI I functions at a structural level with no significance at its sequence level, LRI II functions at a sequence level to augment gRNA packaging and propagation (Kalloush et al., 2016). While the LRIs are responsible for maintaining the overall stability of the RNA secondary structure, the three distal stem loop structures i.e. the SL3, Gag SL1, and Gag SL2, have been found to be dispensable for both maintaining the stability and overall structure of the packaging signal as well as for gRNA packaging and propagation (Kalloush et al., 2019).

Closer analysis of the secondary structural motifs of the MPMV packaging sequences identified two purine rich regions of importance. One, a 16 nts single-stranded purine-rich motif referred to here as the ssPurines (U^{191}UAAAAGU GAAAGUAA^{206}) located immediately downstream the Pal SL (Figure 15; Jaballah et al., 2010). Careful observation revealed that, the purine sequence at the latter 3' end of the ssPurines was found repeated in region "B", downstream of the ssPurines.

Though this latter region was identical in sequence it was found partially base paired and hence referred to as the base paired Purines or the bpPurines ($G^{246}AAAGUAA^{253}$) (Figure 15; Jaballah et al., 2010). The purine richness of the ssPurines and its close vicinity to the DIS and availability as a single stranded region in the secondary structure of the packaging signal indicated the possibility of it functioning as a potential Gag binding site and hence allowing for selective gRNA packaging during MPMV life cycle.
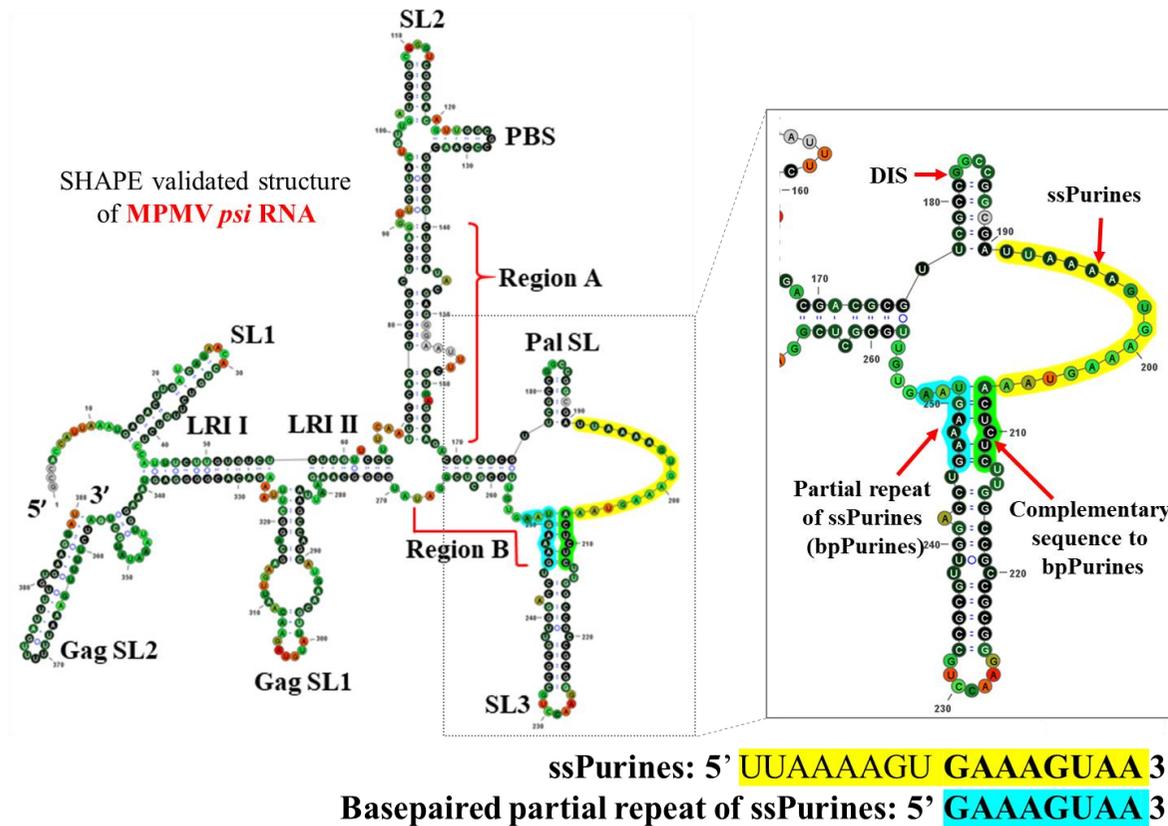
Figure 15: SHAPE-validated secondary structure of the MPMV packaging signal RNA.

SHAPE-validated secondary structure of MPMV RNA packaging determinants indicating the major structural motifs. SL, stem loop; LRI, long range interaction; Pal SL, palindromic stem loop; ssPurines, single-stranded purines (yellow); bpPurines, base-paired repeat purines (blue). The green region highlights the sequences complementary to bpPurines. The RNA structure was redrawn in VARNA using SHAPE reactivity data (Aktar et al., 2013). Figure and legend adapted and modified from Ali et al., 2020.

Deletion analysis of a 20 nts region inclusive of the 16 nts ssPurines region (mutant SJ44; Jaballah et al., 2010) revealed an interesting predicted Mfold conformation with the base paired region of the bpPurines assuming a single stranded conformation which led to suggest a possible role for the bpPurines in Gag binding and gRNA packaging as well, possibly by functioning as a potential Gag binding site in the absence of the ssPurines (Figure 16; Jaballah et al., 2010). Recent more targeted genetic studies carried out to investigate the role of both these purine rich regions (ss- and bp-Purines) in gRNA packaging and propagation revealed that deletion of both the ssPurines and bpPurines independent of each other did not affect gRNA packaging significantly, suggesting that both these regions may act redundantly in bringing about gRNA packaging. Additionally in contrast to the predicted Mfold structure for the clone SJ44, the SHAPE validated structure of a similar clone with the precise deletion of the 16 nts ssPurines (LA-II; Ali et al., 2020), did not reveal unpairing of the otherwise base paired bpPurines (Figure 16). While individual deletions of these purine rich regions (ss- and bp-Purines) did not affect gRNA packaging, the combined deletion of both the regions resulted in abrogation of gRNA packaging. SHAPE validation of the RNA secondary structure of this double deletion mutant indicated a complete loss of its secondary structure hence affecting gRNA packaging. This observation suggested that both the purine rich regions (ss- and bp-Purines) are important in maintaining the overall structural stability of the MPMV packaging sequences that may in turn allow for successful gRNA packaging (Ali et al., 2020).
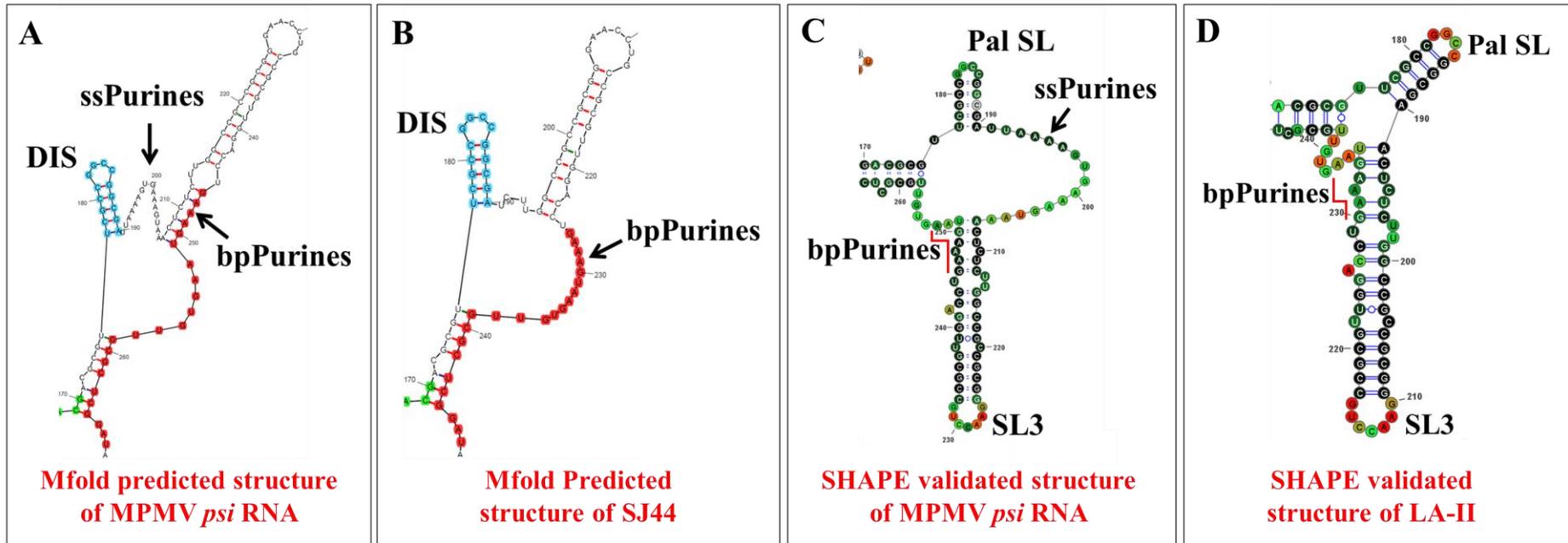
Figure 16: Predicted Mfold structures of the MPMV *Psi* RNA, deletion mutant SJ44 and SHAPE validated structures of the MPMV *Psi* RNA and deletion mutant LA-II.

(A) Mfold predicted structure of the ss- and bp-Purines region of the wild type MPMV *Psi* RNA. (B) Mfold predicted structure of the 20 nts deletion mutant SJ44 (Jaballah et al., 2010) predicting unpairing of the bpPurines. (C) SHAPE validated structure of the the ss- and bp-Purines region of the wild type MPMV *Psi* RNA (D) SHAPE validated structure of the mutant LA-II (Ali et al., 2020) comprising the precise 16 nts deletion of the ssPurines indicating no conformational change or unpairing of the bpPurines.

1.9 Objectives

Several unique features of MPMV have led to an increasing interest in developing MPMV-based vectors for human gene therapy. One of the key steps towards the development of such vectors is the identification of the packaging sequences present in the MPMV gRNA and deciphering the molecular mechanisms underlying the gRNA packaging process. During assembly of MPMV particles, the viral Pr78^{Gag} protein must specifically select the viral gRNA from a variety of cellular and viral spliced RNAs. However, not much is known about how Pr78^{Gag} achieves this selection. Therefore, the overall goal of this study was to gain a better understanding of the MPMV gRNA packaging process by delineating the mechanism(s) involved in the initial recognition of the gRNA by Pr78^{Gag} by addressing the following specific aims:

Specific Aim I: Expression, purification, and characterization of MPMV Pr78^{Gag}.
 Due to the unavailability of purified full-length MPMV Pr78^{Gag}, the process by which the gRNA is recognized by Pr78^{Gag} and the mechanism(s) by which this Gag precursor discriminates against spliced RNAs has not been investigated until now. To overcome this caveat, the first aim was to express large amounts of Pr78^{Gag} in *Escherichia coli* (*E. coli*) and purify it to homogeneity employing immobilized metal affinity chromatography (IMAC) followed by size exclusion chromatography (SEC). The background, methodology employed, and results obtained regarding this specific aim have been discussed in Chapter 2 of this dissertation.

Specific Aim II: Identification and characterization of Pr78^{Gag} binding site(s) on MPMV gRNA.

During MPMV replication, Pr78^{Gag} must select the viral gRNA from a variety of cellular and viral spliced RNAs. However, there is no consensus on how Pr78^{Gag} achieves this selection. Recently two purine-rich regions, i.e., ssPurines and bp-Purines were identified which could function as potential Gag binding site(s) either at the sequence or the structural levels (Jaballah et al., 2010; Aktar et al., 2013; Ali et al., 2020). Therefore, these purine-rich motifs were interrogated via a series of mutations that were introduced and cloned into a T7 expression plasmid. RNA binding assays on the wild type and mutant *in vitro* transcribed RNAs, together with footprinting followed by SHAPE on the wild type RNAs were performed to identify the specific/high affinity Pr78^{Gag} binding site(s) on MPMV gRNA. The background, methodology employed, and results obtained regarding this specific aim have been discussed in Chapter 3 of this dissertation.

Specific Aim III: Establish biological correlation between the Pr78^{Gag} binding site(s) and MPMV gRNA packaging and propagation.

After mapping the Gag binding site(s) on the genomic RNA, in the third aim of the study, the biological role these sequences play in MPMV gRNA packaging and propagation were assessed. This was achieved by comparing the results obtained in Specific Aim 2 to those recently published employing a biologically relevant gRNA packaging and propagation assay (Ali et al., 2020). The background and results obtained regarding this specific aim have been discussed in Chapter 3 of this dissertation.

Chapter 2: Expression, Purification and Characterization of Biologically active Full-length Mason-Pfizer Monkey Virus (MPMV) Pr78^{Gag}

2.1 Abstract

MPMV precursor polypeptide Pr78^{Gag} orchestrates assembly and packaging of genomic RNA (gRNA) into novel MPMV virus particles. Recombinant full-length Pr78^{Gag} was expressed either with or without -His$_6$-tag in bacterial cells and purified from soluble fractions. Successful expression of Pr78^{Gag} with or without -His$_6$-tag was also demonstrated in eukaryotic cells. Here it was demonstrated that recombinant Pr78^{Gag}, expressed and purified from bacterial cells, has the intrinsic ability to assemble *in vitro* in the presence of nucleic acid to form virus like particles (VLPs). Consistent with this observation, the recombinant protein expressed in both prokaryotic and eukaryotic cells established formation of VLPs. VLPs formed in eukaryotic cells by recombinant Pr78^{Gag} either with or without His$_6$-tag were capable of encapsidating MPMV transfer vector RNA, suggesting that the presence of a His$_6$-tag on full-length Pr78^{Gag} did not interfere with its expression or ability to package transfer vector RNA, demonstrating intact biological functioning of the recombinant protein. This study demonstrates the expression and purification of a biologically and functionally active, recombinant Pr78^{Gag}, suitable to study RNA-protein interactions involved in the MPMV gRNA packaging process.

2.2 Introduction

Retroviruses are a group of viruses that require packaging/encapsidation of their "full-length", unspliced, single-stranded, gRNA into assembling viral particles for the continuity of their life cycle. During this process, two copies of the gRNA

dimerize and are preferentially packaged into the assembling virions compared to the spliced viral RNA and the large pool of cellular RNAs of the infected host cell (D'Souza & Summers, 2005; Lever, 2007; Johnson & Telesnitsky, 2010; Ali et al., 2016; Comas-Garcia et al., 2016; Maldonado & Parent, 2016; Dubois, Marquet, et al., 2018a). Such specificity towards packaging of gRNA is a result of intricate interaction(s) between the *cis*-acting sequences on the gRNA and the *trans*-acting viral Gag protein. Retroviral *cis*-acting sequences which interact with Gag polyprotein are generally located at the 5' end of the gRNA and have been designated as the "packaging signal" or "*Psi* /ψ". For almost all retroviruses, the *Psi* sequences required for gRNA packaging have been identified as a structurally-conserved region generally present both upstream and downstream of the mSD within the 5'UTR and often extending into the 5' end of the *Gag* ORFs (Ali et al., 2016; Comas-Garcia et al., 2016; D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Maldonado & Parent, 2016; Lever, 2007; Mailler et al., 2016).

Among the proteins implicated in selective gRNA packaging into virus particles, the NC region of the retroviral Gag polyprotein is a primary candidate, as this highly basic protein contains Cys-His boxes that can interact with Zn^{2+} ions to facilitate protein/RNA interactions (Ali et al., 2016; Jewell & Mansky, 2000). Mutational analysis of the NC domain of several retroviral *gag* genes has shown that it is one of the most critical proteins involved in gRNA packaging (Aldovini & Young, 1990; Dorfman et al., 1993; Gorelick et al., 1988; Méric et al., 1988; Poon et al., 1996). However, additional lines of evidence indicate that NC may not be the only determinant of specific gRNA packaging, and other Gag domains may also be involved, including matrix (Lu et al., 2011), capsid, the p2 spacer peptide between CA and NC (Kaye & Lever, 1998; Roy et al., 2006; Russell et al., 2003), and the terminal

p6 late (L) domain (Tanwar et al., 2017). Furthermore, it is thought that rather than recognizing monomeric RNA substrates, NC probably recognizes dimeric genomes, an interaction that is thought to initiate the multimerization of the Gag polyprotein on the RNA templates, eventually leading to encapsidation of the gRNA into the assembling virus particle (D'Souza & Summers, 2004; Ferrer et al., 2016; Miyazaki et al., 2010). Together, these observations suggests that specific selection of gRNA from cellular and spliced RNAs is a complex phenomenon that happens in the context of the whole Gag polyprotein, as has recently been shown for HIV-1 (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Smyth et al., 2015, 2018).

Based on these observations, a simplistic model shown in Figure 17A suggests that the gRNA is preferentially packaged by virtue of the presence of the *cis*-acting Gag binding site on the structured RNA packaging determinant. In the case of the spliced RNAs, part of the packaging sequence is spliced out, thereby excluding them from encapsidation into the newly formed viral particles, a process that also disrupts the structure of the packaging determinants (Figure 17A). Such a model offers a possible mechanism for discriminating between spliced and unspliced viral mRNAs (D'Souza & Summers, 2005) as has recently been shown in the case of HIV-1 where the SL1 is located in the region harboring the packaging determinants of HIV-1 and is capable of binding HIV-1 Pr55[Gag] with high affinity (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Didierlaurent et al., 2011; Smyth et al., 2015, 2018). Furthermore, the sequences downstream of this region enhance HIV-1 RNA packaging, while the sequences upstream inhibit packaging efficiency (Abd El-Wahab et al., 2014). Such RNA binding studies that challenge the earlier observations where SL3 was thought to contain the primary packaging determinants (Kuzembayeva et al.,

2014) have not been accomplished in most retroviruses owing to the unavailability of respective purified full-length Gag precursor proteins.
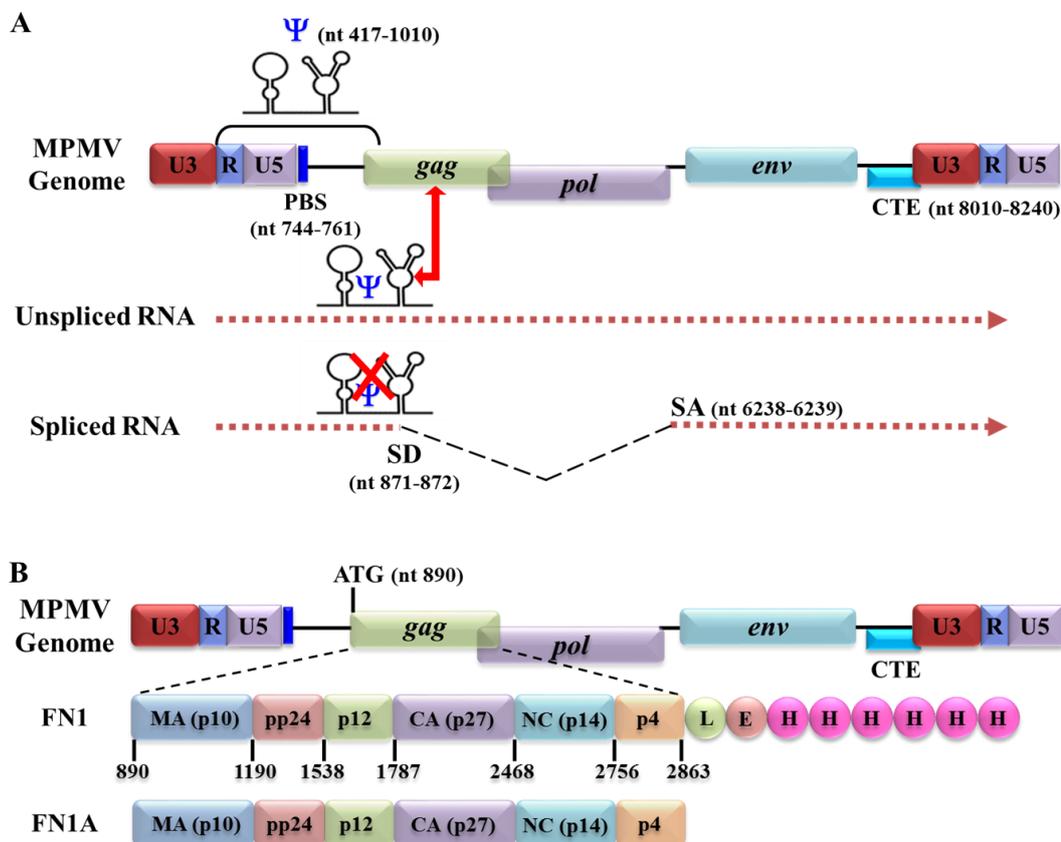
Figure 17: Graphical representation of the mode of selective packaging of gRNA in MPMV and its precursor polyprotein expressed from vectors FN1 and FN1A.

(A) Schematic representation of a simplistic model of retroviral genomic RNA packaging. The presence of an intact RNA secondary structure of the packaging signal (*Psi*; ψ) facilitates genomic RNA packaging, whereas loss of its structure in the spliced RNA excludes its packaging. (B) Illustration of the MPMV Gag proteins expressed from the prokaryotic expression vector either with (FN1) or without (FN1A) the His6- tag (6xH). The amino acid residues LE were introduced upstream of the His6-tag into the FN1 vector during the cloning process. Figure & legend adapted from Pitchai et al., 2018.

The Mason-Pfizer monkey virus (MPMV) is a non-transforming, prototypic simple, type D retrovirus, which has been shown to be involved in causing immunodeficiency in infected new-born Rhesus monkeys (Bryant et al., 1986; Fine et

al., 1975). Among type D retroviruses, MPMV RNA packaging is the most thoroughly investigated (Aktar et al., 2013; Bharat et al., 2012; Bohl et al., 2005; Bray et al., 1994; Füzik et al., 2016; Guesdon et al., 2001; Harrison et al., 1995; Jaballah et al., 2010; Kalloush et al., 2016; Kroupa et al., 2016; Montiel, 2010; Mustafa et al., 2004; Prchal et al., 2013; Schmidt et al., 2003; Schur et al., 2015; Sonigo et al., 1986; Vile et al., 1992), primarily because MPMV-based vectors are considered as potential tools for delivering therapeutic genes in human gene transfer studies. MPMV-based vectors are good candidates for human gene transfer studies because: i) MPMV promoter is transcriptionally active in human cells, thereby obviating the need of replacing MPMV promoters with those of other human viral promoters, and ii) the presence of MPMV constitutive transport element (CTE) should greatly facilitate the nuclear export of the therapeutic genes for their effective expression in the target cells (Bray et al., 1994; Rizvi et al., 1996a; 1996b; 1997b).

A number of studies have focused on identifying the MPMV sequences (at the sequence as well as secondary RNA structure levels) responsible for gRNA packaging and dimerization (Aktar et al., 2013; Guesdon et al., 2001; Harrison et al., 1995; Jaballah et al., 2010; Kalloush et al., 2016; Mustafa et al., 2004; Schmidt et al., 2003; Vile et al., 1992). There is now a consensus that sequences that are responsible for MPMV packaging are highly structured, bipartite in nature, and divided into two regions both upstream and downstream of the major splice donor site (Aktar et al., 2013; Jaballah et al., 2010; Kalloush et al., 2016). However, not much is known as to how Pr78[Gag] selects gRNA. For instance, it remains largely unclear whether discrimination between gRNA and spliced RNA is mediated by the initial binding step to Pr78[Gag], or whether other pathways such as the gRNA nuclear export and subcellular localization are also involved, as has been proposed for HIV-1 (Barajas et al., 2018;

Becker & Sherer, 2017; Behrens et al., 2017; Brandt et al., 2007; Jouvenet et al., 2011; Lingappa et al., 2017; Moore et al., 2009). The void in understanding selective packaging of gRNA among retroviruses is largely due to the unavailability of biologically active full-length Gag polyprotein which has been proposed to interact with the packaging sequences on the full-length, unspliced, and dimerized gRNA (Abd El-Wahab et al., 2014).

The *gag* gene of MPMV encodes a polypeptide, Pr78[Gag] that is the precursor of the viral structural proteins responsible for formation of MPMV particles. Pr78[Gag] is proteolytically cleaved into six proteins (Figure 17B): namely NH2-p10 (MA), pp24 (and its C-terminal cleaved product, referred as pp24/16), p12, p27 (CA), p14 (NC), and p4-COOH (Bradac & Hunter, 1984; Henderson et al., 1985; Sonigo et al., 1986). Cleavage of the polyprotein is achieved by a protease (PR) encoded for by the virally-encoded *pro* gene (Bradac & Hunter, 1984; Henderson et al., 1985; Hrusková-Heidingsfeldová et al., 1995). Like most retroviruses, MPMV Pr78[Gag] assembles to form an immature capsid intracytoplasmically, and expression of the *pro* gene results in the maturation of the virus particles upon budding and release from the infected cell (Rhee & Hunter, 1990).

Since Pr78[Gag] is a critical component of the packaging process, understanding the biochemical and biophysical properties of MPMV Pr78[Gag] is of paramount importance to understand MPMV biology. Overexpression and purification of Pr78[Gag] in bacteria has been reported before; however, the protein was mainly found within inclusion bodies and had to be solubilized and denatured for purification and then refolded for further analysis (Klikova et al., 1995). Furthermore, its suitability for RNA binding assays was never established. Therefore, to overcome this caveat, large amounts of recombinant Pr78[Gag] was expressed in soluble fractions of *Escherichia coli*

(*E. coli*) containing a C-terminal hexa-histidine (His$_6$) tag to facilitate protein purification. This was followed by immobilized metal affinity chromatography (IMAC) to purify the protein to homogeneity employing high-pressure liquid chromatography (HPLC) (Bewley et al., 2017; McKinstry et al., 2014; Tanwar et al., 2017). The availability of purified MPMV full-length Gag polyprotein should allow us to investigate how Pr78^{Gag} is involved in selectively packaging gRNA over spliced viral and cellular RNAs which will further enhance the understanding of molecular intricacies involved in MPMV gRNA packaging, especially in delineating RNA-protein interactions that take place during MPMV replication.

2.3 Materials and Methods

2.3.1 Nucleotide Numbering System

Nucleotide numbers in this study refers to the MPMV genome with the Genbank accession number M12349 (Sonigo et al., 1986).

2.3.2 Construction of Prokaryotic MPMV Gag Expression Plasmids

MPMV *gag* gene (Pr78^{Gag}) harboring sequences spanning nucleotides 890 to 2863 was commercially synthesized as a double-stranded DNA fragment (Macrogen, South Korea). Two restriction sites, *Nco*I and *Xho*I, were incorporated at either ends of the gene to facilitate cloning into the prokaryotic expression vector pET28b (+) (Figure 18A). Use of the *NcoI* site pET28b(+) vector kept the *gag* open reading frame (ORF) intact and allowed the introduction of a hexa-histidine (His$_6$-tag) at the C-terminus of the Pr78^{Gag}, allowing purification of the expressed recombinant Gag polyprotein using IMAC (Bewley et al., 2017; McKinstry et al., 2014). Since the

MPMV *gag* gene contained an additional internal *NcoI* site, a silent mutation was introduced into the internal *NcoI* site at the gene synthesis step at nt 2462, changing C<u>C</u>ATGG to C<u>T</u>ATGG, which resulted in the loss of the *NcoI* site; while maintaining the same amino acid (Figure 18A & C). The resultant clone FN1 was sequenced to ensure that the MPMV *gag* gene did not contain any point mutations (Figure 18B). Such an expression plasmid should produce a recombinant fusion protein comprising of full-length MPMV Pr78^{Gag} and a LEHHHHHH tag at the C-terminus (Pr78^{Gag}-His$_6$-tagged), with a predicted molecular weight of 74101 daltons (Da). Employing a similar strategy, another bacterial expression plasmid (FN1A) was constructed expressing full-length MPMV Pr78^{Gag} without the His$_6$-tag (Figure 17B).
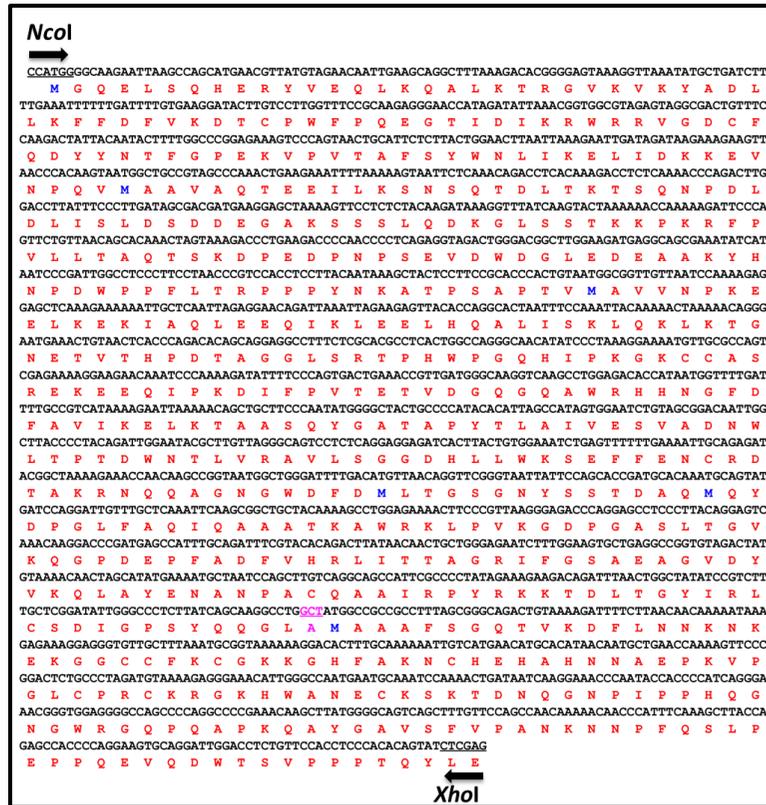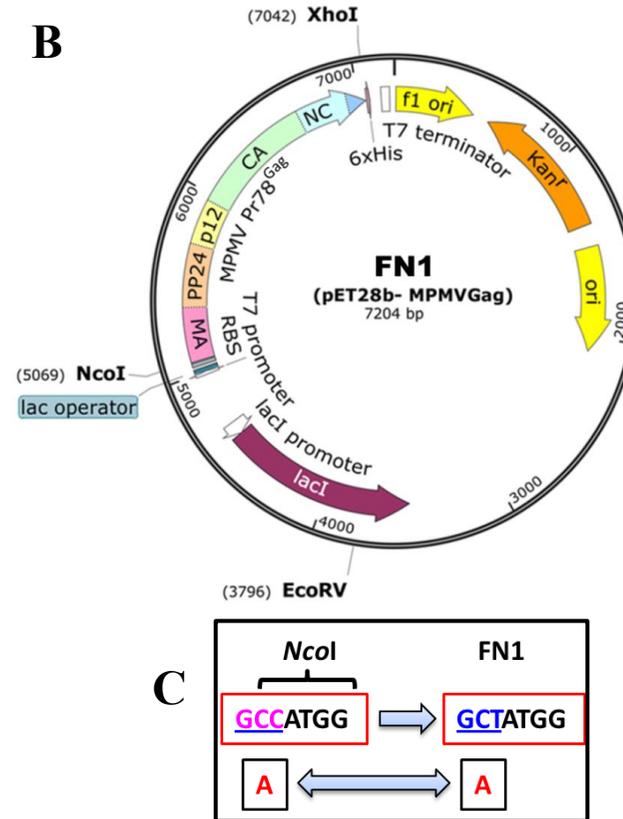
Figure 18: Schematic representation of the construction of the recombinant Pr78^{Gag}.

(A) Full length nucleic acid and amino acid sequence of MPMV Pr78^{Gag}. (B) Design of the modified pET28b (+) vector expressing the full length MPMV Pr78^{Gag} (FN1) cloned into *Nco*I and *Xho*I sites and expressed from the bacteriophage T7 promoter. (C) A silent mutation introduced into the *gag* gene to remove an internal *Nco*I site at nt 2462 for the ease of cloning. Figure & legend adapted from Pitchai et al., 2018.

2.3.3 Construction of Eukaryotic MPMV Gag Expression Plasmids

The full length Pr78^{Gag} sequences both with and without the His$_6$-tag were also cloned into the eukaryotic expression vector pcDNA3. Towards this end, the FN1 plasmid was used as a template for amplification of *gag* gene sequences using the forward primer, OTR1330 and the reverse primers, OTR1331 or OTR1320 (Appendix 1). OTR1330 (5' CCG *CTC GAG* GCC GCC ACC ATG GGG CAA GAA TTA AGC CAG G 3') introduced an *Xho*I restriction site (italicized) followed by Kozak sequence (underlined) at the 5' end of the *gag* gene to enhance gene expression. OTR1331 (5' CAA GGT GGA GGG TGT GTC ATA GTG GTG GTG GTG GTG GTG ATT *GAG CTC* GCC 3') on the other hand, created a His$_6$ tag (underlined) just upstream of the *gag* stop codon, followed by the *Xho*I restriction site (italicized). To create a clone without His$_6$ tag, the reverse primer OTR1320 (5' CCG *CTC GAG* TTA ATA CTG TGT GGG AG 3') was used, which did not contain the His$_6$ tag sequences but did contain an *Xho*I restriction site (italicized). Polymerase chain reaction (PCR) was performed using an initial denaturation at 98°C for 30 seconds, then 15 cycles of denaturation at 98°C for 10 seconds, primer annealing at 62°C for 30 seconds, followed by primer extension at 72°C for 30 seconds. A final extension step at 72°C for 10 minutes was also included. The PCR-amplified products were cleaved with *Xho*I endonuclease and cloned into pCDNA3 which had previously been cleaved with the same restriction endonuclease to create clones FN3 and FN5, with and without His$_6$-tag, respectively. Finally, to facilitate efficient nuclear export of *gag* mRNA, a PCR-amplified fragment containing the MPMV CTE (Bray et al., 1994) with flanking *Xba*I sites was cloned into FN3 and FN5 which had already been cleaved with *Xba*I,

resulting in clones FN7 and FN9 respectively. All clones were confirmed by sequencing.

2.3.4 Bacterial Strains and Media

During the course of cloning, all transformations were performed in the DH5α strain of *E. coli* using standard heat shock protocol and allowed to grow on Luria-Bertani (LB) agar plates in the presence of appropriate antibiotics (kanamycin at 50 µg/ml and ampicillin at 100 µg/ml) depending on the clones. To monitor the recombinant protein expression in bacteria, the prokaryotic expression clones were transformed into T7 Express (New England Bio Labs, USA), a BL21(DE3) strain of *E. coli*. For bacterial expression studies, cells were cultured in LB medium [1% (w/v) peptone, 0.5% (w/v) yeast extract, and 0.5% NaCl] in the presence of kanamycin (50 µg/ml) antibiotic.

2.3.5 Large Scale Expression of Recombinant Pr78^{Gag}-His$_6$-tagged Protein

Large scale recombinant Pr78^{Gag}-His$_6$-tagged protein expression was performed by inoculating a single colony into 25 ml LB media containing kanamycin antibiotic (50 µg/ml), and cultured at 37$^°$C, overnight with shaking at 200 rounds per minute (rpm). The overnight culture was used to inoculate 500 ml LB media containing kanamycin (50 µg/ml) in a 2-liter baffled Erlenmeyer flask. The culture was grown at 28$^°$C until an OD$_{600}$ nm of approximately 0.6 was reached. Protein expression was induced with the addition of 0.4 mM Isopropyl β- d-1-thiogalactopyranoside (IPTG), and the cells were grown for a further 4 h at 28$^°$C. Cells were harvested by centrifugation (6,300 *g*) at 4$^°$C for 15 minutes and cell pellets were stored frozen at -80$^°$C.

2.3.6 IMAC Protein Purification and Size Exclusion Chromatography

Recombinant Pr78^{Gag}-His$_6$-tag protein was purified as has been described previously (Bewley et al., 2017; McKinstry et al., 2014; Tanwar et al., 2017). Briefly, IPTG-induced bacterial pellets were lysed in cold CelLytic B (Sigma-Aldrich) supplemented with 500 units of Benzonase (Merck), 0.2 mg/ml lysozyme (Sigma-Aldrich) and 1X concentration of EDTA-free protease inhibitor tablet (Roche). The soluble fraction was collected after centrifugation (48,000 g for 1 hour) at 4°C and the supernatant was diluted with 4X binding buffer (0.2 M Tris-HCl (pH 8.0), 4.0 M NaCl, 40 mM ß-mercaptoethanol, 10 mM dithiothreitol (DTT), 100 mM imidazole, 0.4% (w/v) Tween-20) to a final concentration of 1X. The lysate was then filtered through a 0.4 µm Polyethersulfone (PES) syringe filter and loaded onto a 5 ml HisTRAP^{TM} FF (fast flow) cartridge (GE Healthcare) that had been equilibrated with equilibration buffer (50 mM Tris-HCl (pH 8.0), 1.0 M NaCl, 10 mM ßmercaptoethanol, 2.5 mM DTT, 25 mM imidazole, 0.1% (w/v) Tween-20, and 10% (v/v) glycerol). After loading the filtrate, the column was washed with the same equilibration buffer except it contained 50 mM imidazole and the bound proteins were eluted with equilibration buffer containing 250 mM imidazole.

Following HisTRAP^{TM} column elution, Pr78^{Gag}-His$_6$-tag protein was concentrated using Amicon$^{®}$ Ultra 15 (30,000 molecular weight cut-off membrane) and was further fractionated by gel filtration/size exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare) previously equilibrated with 50 mM Tris-HCl (pH 8.0) and 1.0 M NaCl. Following SDS-PAGE analysis, peak fractions containing Pr78^{Gag}-His$_6$-tag protein were pooled and stored for long term usage in 2 µg/µl aliquots at -80°C.

2.3.7 SDS-polyacrylamide Gel Electrophoresis and Western Blotting

Expression and purification of recombinant $Pr78^{Gag}$-His_6-tag protein was monitored by SDS-PAGE and western blotting. Briefly, protein samples were mixed with 6X SDS dye, boiled for 5 minutes before loading onto a 4-12% ExpressPlus^{TM} PAGE gel (GenScript), electrophoresed under reducing conditions using MOPS buffer (GenScript), followed by their staining with Coomassie Brilliant Blue. Recombinant $Pr78^{Gag}$ expression and purification was further monitored by transferring non-stained gels onto a nitrocellulose membrane and blotting with anti-rabbit MPMV Gag/Pol Pr78 polyserum (kindly provided by Dr. Eric Hunter, Emory University, Atlanta, GA) and with an anti-His_6 monoclonal antibody-HRP conjugate (Sigma-Aldrich).

2.3.8 Eukaryotic Expression of Recombinant $Pr78^{Gag}$-His_6-tag Protein

Transient transfections of the expression vectors (4 µg) of full-length Gag eukaryotic expression plasmids (FN7 and FN9) along with (2 µg) of MPMV-based transfer vector SJ2 (Jaballah et al., 2010) were carried out in HEK 293T cells in triplicates using a calcium phosphate kit (Invitrogen) following manufacturers' recommendations. The resulting supernatants from the transfected cultures containing virus particles were subjected to low-speed centrifugation (bench-top centrifuge, 3,700 *g* for 10 minutes) to clear cellular debris. Next, supernatants were filtered using 0.2 µm surfactant free cellulose acetate (SFCA) syringe filters and subjected to ultracentrifugation at 70,000 *g* to pellet virus like particles (VLPs) using a 20% (w/v) sucrose cushion. The pelleted VLPs were resuspended in TN buffer (20 mM Tris-HCl (pH 7.4), 150 mM NaCl) and processed for RNA extraction and western blotting. RNA isolation was performed using TRIzol®.

2.3.9 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

Viral RNA preparations were treated with Turbo DNase (Invitrogen) and amplified using transfer vector (SJ2)-specific primers OTR1161 and OTR1163 (Appendix 1) to ensure that the RNA preparations were not contaminated with any plasmid DNA that may have been carried over from the transfected cultures. Next, the DNAsed-RNAs were converted into cDNAs using random hexamers (5' NNNNNN 3') and MMLV reverse transcriptase (Promega, USA) as described previously (Ghazawi et al., 2006; Mustafa et al., 2005). cDNAs were amplified using the same vector-specific primers (OTR1161 and OTR1163; Appendix 1) to monitor the ability of Pr78^{Gag} VLPs to package transfer vector (SJ2) RNA, as described previously (Jaballah et al., 2010).

2.3.10 VLP Production in Prokaryotic Cells and Transmission Electron Microscopy (TEM)

To monitor the formation of VLPs by recombinant Pr78^{Gag}-His$_6$-tag protein in bacterial cells (following induction with IPTG), cells were pelleted, washed with 0.1 M PBS and fixed in Karnovsky's fixative. Next, cell pellets were stained with 1% osmium tetroxide and subjected to dehydration using graded ethanol solutions. Finally, cell pellets were fixed in epoxy resin (agar 100). Ultrathin sections on 200 mesh copper (Cu) grids were negatively stained with 1% uranyl acetate followed by lead citrate double stain and analyzed using a FEI Tecnai Biotwin Spirit G2 transmission electron microscope (TEM).

2.3.11 *In Vitro* Assembly of Recombinant Pr78^{Gag}-His$_6$-tag Protein to form VLPs

To observe *in vitro* assembly of VLPs, purified recombinant Pr78^{Gag}-His$_6$-tag protein (in 20 mM Tris (pH 7.4) containing 1.0 M NaCl and 10 mM DTT) was mixed

with yeast tRNA at a nucleic acid to protein ratio of 4% (w/w), placed in a Slide-A-Lyzer® 10 kDa dialysis cassette G2 (Thermo Scientific), and dialyzed against 20 mM Tris (pH 7.4) containing 150 mM NaCl and 10 mM DTT overnight at 4°C. Following dialysis, ~ 400 µl was recovered that was concentrated using Amicon® Ultra 15 (30,000 molecular weight cut-off membrane) to a final volume of ~250 µl of which 8-10 µl was spotted on a carbon coated formvar grid (Proscitech, Australia), air dried, and stained with 1% uranyl acetate for TEM observation.

2.4 Results and Discussion

2.4.1 Bacterial Expression of Recombinant MPMV Pr78^{Gag}-His$_6$-tag Protein

In order to express full-length MPMV Pr78^{Gag}, a recombinant bacterial expression plasmid (FN1) was designed in a fashion that this plasmid should produce a fusion protein comprising of full-length MPMV Pr78^{Gag} and a LEHHHHHH tag at the C-terminus (Pr78^{Gag}-His$_6$-tag), with a predicted molecular weight of ~74101 Da (Figure 17B). Such a T7 RNA polymerase promoter-based expression plasmid (FN1) facilitated high level expression of His$_6$-tagged MPMV Pr78^{Gag} by BL21(DE3) bacterial cells when induced with IPTG because of the presence of a chromosomal copy of the T7 *RNA polymerase* gene under the dependency of the lac promoter (Figure 18B).

Expression of the recombinant Pr78^{Gag}-His$_6$-tagged protein was confirmed by growing FN1-transformed BL21(DE3) cultures at 37°C followed by IPTG induction and were further grown at 28°C. At 0, 2, 4, 6, and 8 hours post induction, total protein lysates were prepared from both induced as well as un-induced cultures and the expression of the recombinant MPMV Pr78^{Gag}-His$_6$-tag was monitored based on size

on SDS-PAGE. As shown in Figure 19, a distinct band of ~ 78 kDa corresponding to the expected size of the recombinant full-length Pr78^{Gag}-His$_6$-tag was observed in the induced cultures at 2, 4, 6, and 8 hours (lanes 4-7) but not in un-induced cultures (lane 3) or cultures containing only pET28b(+) vector without any MPMV Gag sequences (lane 2). These results reveal successful expression of the recombinant full-length MPMV Pr78^{Gag}. However, whether this recombinant full-length Gag protein was capable of making VLPs or was present in the soluble bacterial fraction was yet to be investigated. Klikova and colleagues have reported earlier that culturing bacteria at 37$^{°}$C post induction resulted in the confinement of MPMV Gag polyprotein in the inclusion bodies containing aberrantly assembled spiral like structures (Klikova et al., 1995). Keeping these observations in mind, the recombinant MPMV Pr78^{Gag} was expressed employing sub-optimal conditions such as low temperature in all subsequent experiments. Furthermore, although the expression of recombinant MPMV full-length Pr78^{Gag}-His$_6$-tag was observed between 2-8 hours post IPTG induction (Figure 19; lanes 4-7), for subsequent experiments to protein was purified from cultures that were induced only for 4 hours sub-optimally at 28$^{°}$C following IPTG induction in order to avoid any possible aggregation and/or sequestration of Gag in inclusion bodies which has been reported earlier (Klikova et al., 1995).
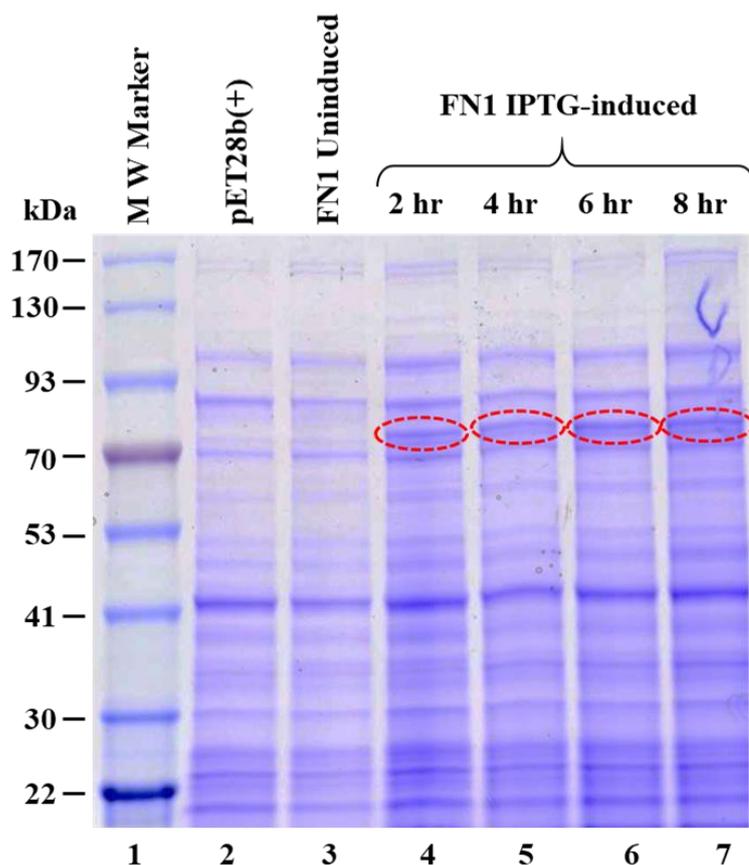
Figure 19: Expression of recombinant Pr78^{Gag} in *Escherichia coli* lysates.

Coomassie Brilliant Blue-stained SDS polyacrylamide gel showing expression of recombinant full-length Pr78^{Gag} prepared from total cell lysates from un-induced and IPTG-induced BL21(DE3) bacterial cells which were cultured for 0, 2, 4, 6, and 8-hours at 28°C. Figure & legend adapted from Pitchai et al., 2018.

2.4.2 The Bacterially-expressed MPMV Pr78^{Gag}-His$_6$-tag Protein Forms VLPs

It has previously been shown that not only MPMV Gag (Klikova et al., 1995; Sakalian et al., 1996; Sakalian & Hunter, 1999), but other retroviral Gag proteins (Campbell & Vogt, 1997; Ehrlich et al., 1992) can form immature VLPs. In order to test whether the His$_6$-tag interfered with VLP formation the recombinant MPMV Pr78^{Gag} either with or without His$_6$-tag at the C-terminus was tested for its ability to assemble into VLPs within bacterial cells. Towards this end, the full-length MPMV Gag recombinant clone FN1 (with His$_6$-tag) and FN1A (without His$_6$-tag; Figure 17B)

were expressed in BL21(DE3) cells at $28^{\circ}C$ and tested for their ability to form immature VLPs using transmission electron microscopy (TEM). The ultrathin sections were negatively stained with 1% uranyl acetate and visualized. Electron micrographs of the IPTG-induced bacterial expression plasmids (FN1 and FN1A) revealed a predominant population of intra-cytoplasmic structures appearing as electron dense rings of ~55-65 nm in size resembling immature VLPs (Figure 20 A-C). This range of size (55-65 nm) of VLPs is consistent with the earlier published observations of MPMV VLPs assembled in bacteria from full-length Gag (Klikova et al., 1995) or mutated Gag (Sakalian & Hunter, 1999b). No such VLP structures were observed when FN1- and FN1A-transformed bacterial cells were not induced with IPTG (Figure 20 D and E). Similarly, no VLP-like structures were observed when the cloning vector, pET28b (+), by itself was transformed in BL21(DE3) and induced employing similar conditions (data not shown). These results suggest that clone FN1 containing full-length MPMV Gag with a His_6-tag at the C-terminus as well as FN1A without the His_6-tag were able to express full-length Gag proteins capable of assembling VLPs in bacteria. Based on these results, it can be further concluded that the presence of His_6-tag at the C-terminus does not interfere with the recombinant full-length MPMV $Pr78^{Gag}$ expression as well as VLP formation. This observation is further strengthened by the fact that when sequences for FN1 (with His_6-tag) and FN1A (without His_6-tag) were subjected to online ExPASy-Compute pI/Mw tool which allows the computation of the theoretical isoelectric point (pI) of proteins, it predicted minimal difference (0.15) in these proteins (FN1-with His_6-tag: pI: 7.07 *versus* FN1A-without His_6-tag: pI: 7.22).
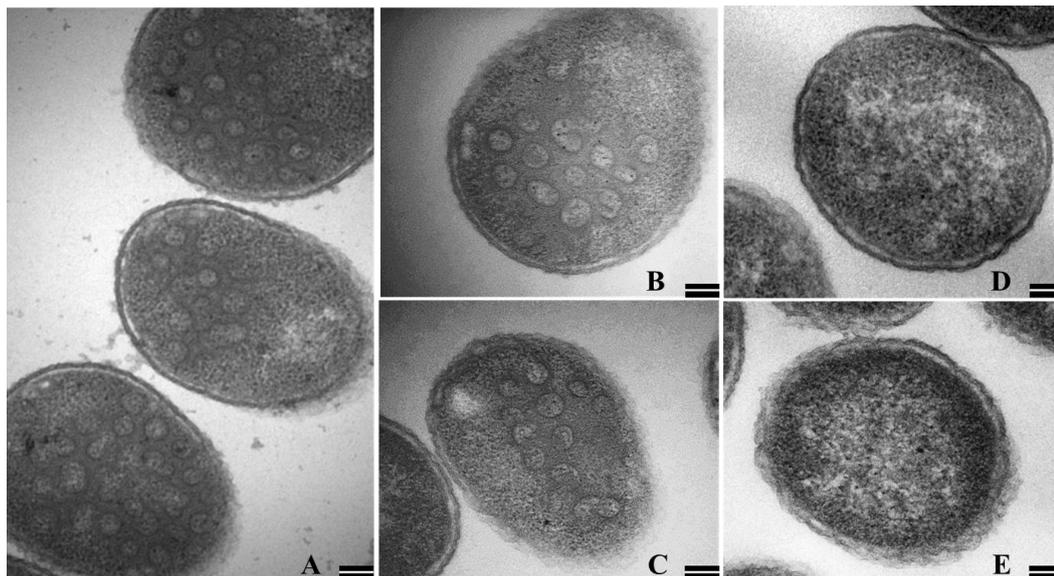
Figure 20: Assembly of virus like particles (VLPs) by recombinant Pr78^{Gag} in *Escherichia coli*.

Transmission electron micrographs showing VLPs assembled in E. coli BL21(DE3) cells transformed with (A) the FN1 clone containing His$_6$-tag and (B and C) with a FN1A clone without the His$_6$-tag. (D and E) BL21(DE3) uninduced BL21(DE3) cells transformed with FN1 and FN1A, respectively. (Scale bar = 100 nm; 60,000X magnification). Figure & legend adapted from Pitchai et al., 2018.

2.4.3 MPMV Pr78^{Gag}-His$_6$-tag Protein is Expressed in the Soluble Fraction in Bacteria

In order to determine whether the recombinant MPMV Pr78^{Gag}-His$_6$-tag protein was present in the soluble bacterial fraction, expression of the full-length MPMV Gag clone in FN1 was induced with IPTG at 28°C for 4 hours and lysed as described in Material and Methods. Insoluble material (containing cell debris and inclusion bodies, if any) was removed by centrifugation and the soluble fraction from different cultures was either stored at -80°C or immediately monitored for the expression of the recombinant MPMV Pr78^{Gag}-His$_6$-tag protein by SDS-PAGE and immunoblotting. Staining with Coomassie Brilliant Blue revealed presence of a distinct band of ~78 kDa corresponding to the expected size of recombinant MPMV Pr78^{Gag}-His$_6$-tag protein (Figure 21A; lane 2). The identity and recombinant nature (Pr78^{Gag}-His$_6$-tag fusion protein) of this band was established by immunoblotting

using HRP-conjugated α-His$_6$ monoclonal antibody (Figure 21B; lane 2) as well as α-MPMV Pr78 polyserum (Figure 21C; lane 2). These results show that under the conditions used, the recombinant Pr78^{Gag}-His$_6$-tag protein was expressed primarily in the soluble fraction in contrast to the inclusion bodies, as reported earlier (Klikova et al., 1995).
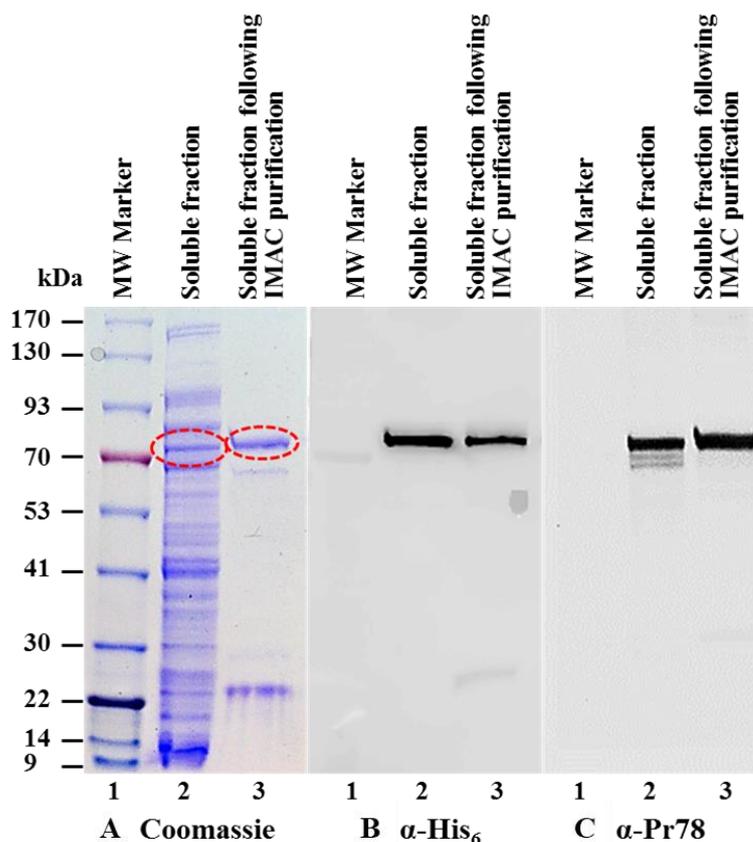
Figure 21: Expression of recombinant Pr78^{Gag} in the soluble fraction of *Escherichia coli*.

(A) Coomassie Brilliant Blue stained SDS-polyacrylamide gel with lysates from the soluble fraction of bacteria transformed with FN1 expressing recombinant full-length MPMV Pr78^{Gag}-His$_6$-tag fusion protein (lane 2), followed by IMAC purification (lane 3). The same lysates were analyzed with (B) a monoclonal anti-His$_6$ monoclonal antibody, and (C) an anti-Pr78^{Gag} polyserum, respectively. Figure & legend adapted from Pitchai et al., 2018.

2.4.4 Further Purification of the Soluble Fraction Containing Pr78^{Gag}-His$_6$-tag Fusion Protein by Immobilized Metal Affinity Chromatography (IMAC)

Having established that the expressed recombinant Pr78^{Gag}-His$_6$-tag protein was soluble, to further increase the purity of the protein since two additional faint bands were observed underneath the primary protein band when detected by the α-MPMV Pr78 polyserum, but not by α-His$_6$ antibody, which could be degraded products (Figure 21B *versus* Figure 21C). Thus, the purified bacterial lysate (containing soluble fraction of recombinant Pr78^{Gag}-His$_6$-tag protein) was further clarified employing IMAC as described in Materials and Methods (Section 2.3.6). The buffering conditions used in these protocols (non-denaturing conditions and especially the presence of 1.0 M NaCl), not only allowed the protein to bind to the column, but also avoided protein aggregation and precipitation. The IMAC-purified protein was then monitored for purity of the recombinant MPMV Pr78^{Gag}-His$_6$-tag protein by SDS-PAGE and immunoblotting. Coomassie Brilliant Blue stain of the SDS-PAGE demonstrated that most bacterial proteins that were present in the soluble fraction prior to IMAC purification were eliminated following IMAC purification (compare lane 2 with lane 3 in Figure 21A). Immunoblotting of IMAC-purified protein with HRP-conjugated anti-His$_6$ monoclonal antibody (Figure 21B; lane 3) and anti-MPMV Pr78 polyserum (Figure 21C; lane 3) further confirmed the purity of the protein as observed by the disappearance of the additional faint bands seen earlier (compare lane 2 with lane 3 in Figure 21C). These results confirm that MPMV full-length Gag is truly fused with the His$_6$-tag, allowing its binding to the HisTRAP^{TM} column and subsequent elution in the purified form, further authenticating the recombinant nature of MPMV Pr78^{Gag}- His$_6$-tag fusion protein (compare Figure 21B; lane 3 with Figure 21C; lane 3).

2.4.5 Concentration and Further Purification of the IMAC-Purified Pr78^{Gag}-His$_6$-tag Protein by Gel Filtration Chromatography

Following IMAC purification bacterially-expressed full-length Pr78^{Gag}-His$_6$-tag protein was concentrated and further purified by gel filtration/size exclusion chromatography under non-denaturing conditions. As in the case of IMAC purification, non-denaturing conditions (especially the presence of 1.0 M NaCl) were employed to prevent protein aggregation and precipitation. Based on a sharp absorbance peak at 280 nm, fractions of 500 µl each were collected over several hours (Figure 22A). Additionally, two other smaller peaks were also noticed that represented the degradation products or unidentified proteins of much lower molecular weight compared to Pr78^{Gag}-His$_6$-tag protein as established by SDS-PAGE analysis (data not shown) and therefore were eliminated for any further downstream applications.

Protein fractions representing the sharp peak (fractions 19-26) were further analyzed by separation on SDS-PAGE. As shown in Figure 22B, fractions collected from the MPMV Pr78^{Gag}-His$_6$-tag peak were characteristically pure, with varying amounts of protein. Fractions representing the highest amount of pure protein (peaks 22-24) were pooled and once again concentrated using Ultra 15 (30,000 molecular weight cut-off membrane) concentrators. To establish the purity of recombinant full-length MPMV Pr78^{Gag}-His$_6$-tag Gag fusion the A260/A280 ratio of the protein was measured by spectrophotometry. Spectrophotometric analysis revealed the A260/A280 ratio to be 0.61, indicating that recombinant full-length MPMV Pr78^{Gag}-His$_6$-tag fusion protein's purity was greater than 95%. The concentrated protein was further analyzed by immunoblotting using anti-MPMV Pr78 polyserum and HRP-conjugated anti-His$_6$ monoclonal antibody. Figure 22C and D, in close corroboration with the SDS-PAGE analysis (Figure 22B), clearly demonstrate that the pooled protein

fractions contained pure MPMV Pr78^{Gag}-His$_6$-tag fusion protein. The protein yield following IMAC purification and gel filtration/size exclusion chromatography was estimated to be 3.8 mg and 0.23 mg per liter respectively.
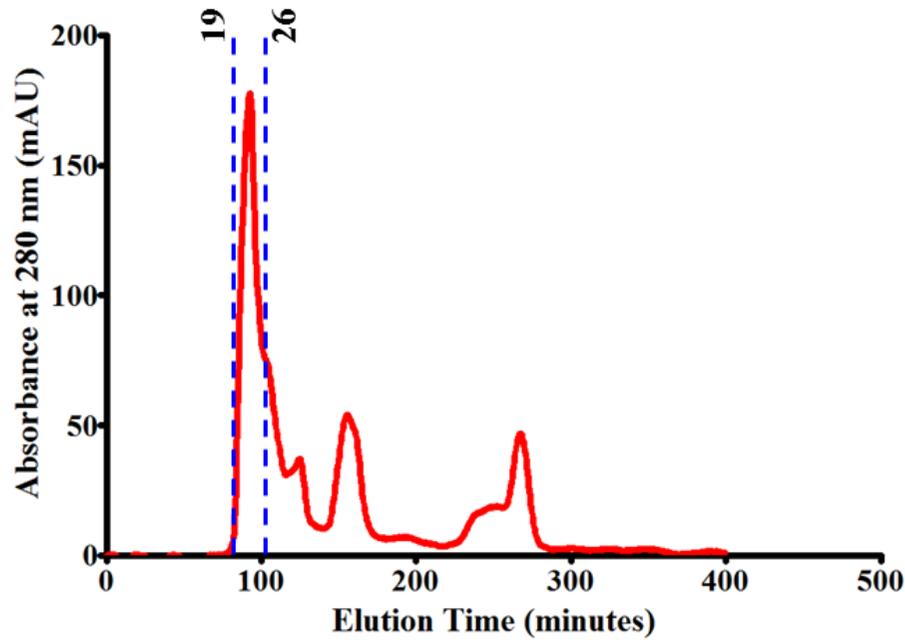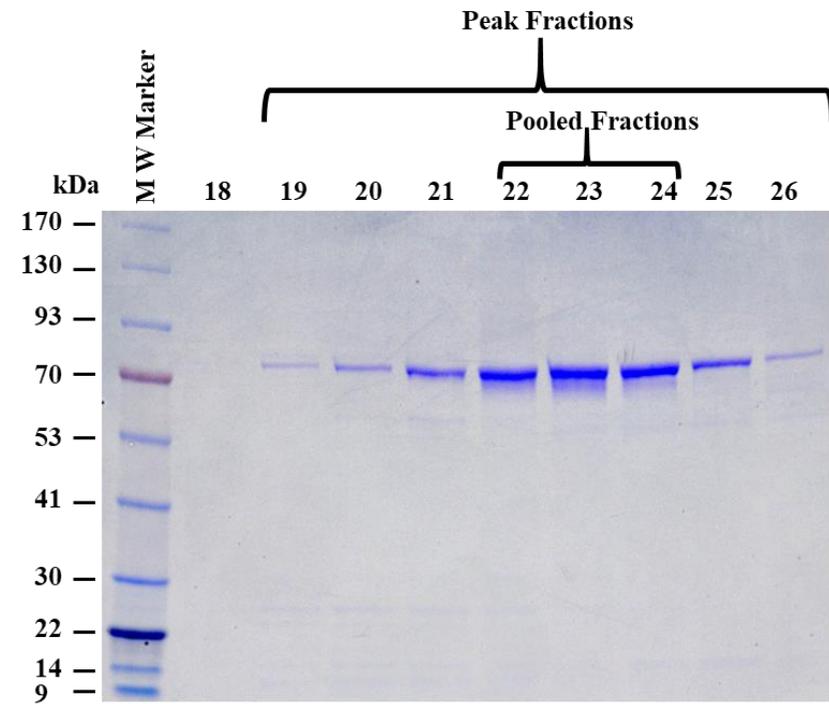
Figure 22: Fractionation of IMAC-purified recombinant Pr78^{Gag} protein by size exclusion chromatography.

(A) Absorbance versus elution time chromatogram plotted from the data obtained from a Superdex 200 column showing peak fractions with maximum absorbance containing purified recombinant full-length MPMV Pr78^{Gag}-His$_6$-tag fusion protein expressed from FN1. (B) Coomassie Brilliant Blue-stained SDS-polyacrylamide gel showing the resolution of purified recombinant full-length MPMV Pr78^{Gag}-His$_6$-tag fusion protein expressed from FN1 in fractions 18–26. Figure & legend adapted from Pitchai et al., 2018.
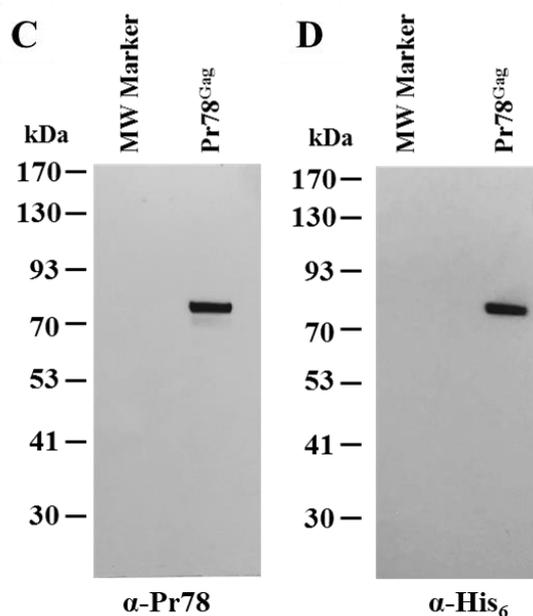
Figure 22: Fractionation of IMAC-purified recombinant Pr78^{Gag} protein by size exclusion chromatography (Continued)

(C) Western blot analysis of pooled peak fractions (22–24) of purified recombinant full-length MPMV Pr78^{Gag}-His$_6$-tag fusion protein analyzed with anti-Pr78^{Gag} polyserum, and (D) anti-6x-His monoclonal antibody, respectively. Figure & legend adapted from Pitchai et al., 2018.

2.4.6 The Recombinant Pr78^{Gag}-His$_6$-tag Protein Can Assemble *In Vitro* to form VLPs

The recombinant full-length Gag proteins from HIV-1 and feline immunodeficiency virus (FIV) have been shown to assemble *in vitro* to form VLPs in the presence of nucleic acid (Affranchino & González, 2010; Campbell & Rein, 1999; McKinstry et al., 2014; Tanwar et al., 2017). Therefore, the ability of the purified recombinant full-length MPMV Pr78^{Gag}-His$_6$-tag fusion protein to assemble *in vitro* to form VLPs was tested. Since the presence of nucleic acids along with purified Gag protein has been shown to be a prerequisite for VLP formation (Affranchino & González, 2010; Campbell & Rein, 1999; McKinstry et al., 2014; Tanwar et al., 2017), purified recombinant Pr78^{Gag}-His$_6$-tag fusion protein was mixed with yeast tRNA as described in Materials and Methods (in an appropriate buffer to avoid formation of

protein aggregates). This mixture was then dialyzed against a buffer that was of low salt concentration compared to the buffer in which the protein-RNA mixture was prepared. As an appropriate negative control, only yeast tRNA was also dialyzed without any protein in the same buffer and manner as the protein-RNA mixture. Following dialysis, the protein-RNA mixture as well as the control tRNA suspension (without any protein) was recovered from the cassette and concentrated after which ~ $1/25^{th}$ of the concentrate was spotted onto carbon coated grids, dried, and stained for observation under an electron microscope.

Analysis of various electron micrographs taken from different fields revealed the assembly of VLPs in the form of compact electron-dense rings of approximately 30-35 nm in size resembling immature VLPs (Figure 23A-F). The size of these *in vitro* assembled VLPs corroborated well with the earlier observations which have reported a similar size (~20-30 nm) obtained following *in vitro* assembly using purified Gag as opposed to the larger size of full-length Gag particles (~55-65 nm) produced *in vivo* in eukaryotic cells (Affranchino & González, 2010; Campbell & Rein, 1999; McKinstry et al., 2014; Tanwar et al., 2017). These VLPs were assembled efficiently despite the fact that the purified protein was frozen and thawed which suggests that the purified protein remained biologically active following freeze-thaw cycle. In contrast, yeast tRNA alone without any purified MPMV full-length Gag, did not show any VLP-like structure (Figure 23G and H). These results suggest that MPMV recombinant full-length Gag His$_6$-tag fusion protein has the inherent property of forming VLPs by virtue of its intrinsic multimerizing ability, as has been reported previously in the case of HIV-1 (Affranchino & González, 2010; Campbell et al., 2001; Campbell & Rein, 1999; McKinstry et al., 2014; Tanwar et al., 2017).
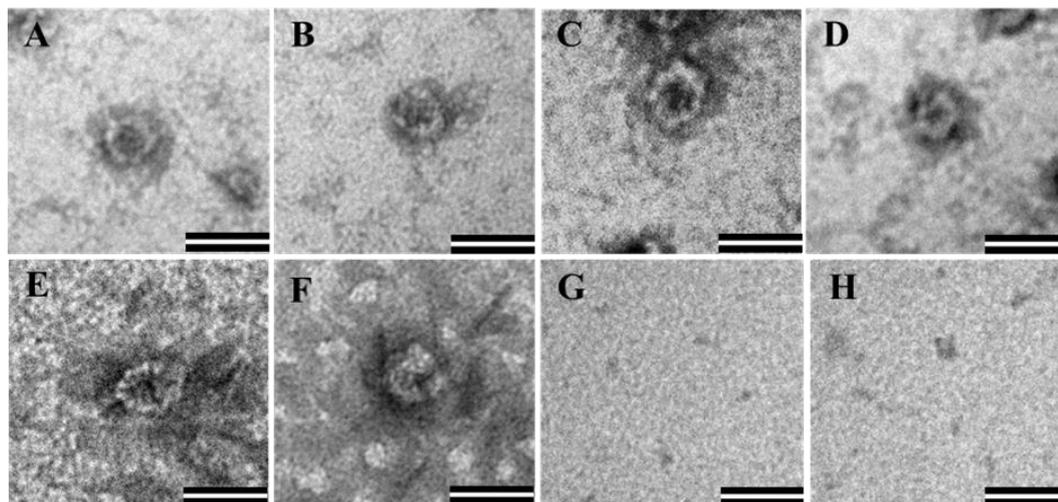
Figure 23: *In vitro* assembly by purified recombinant Pr78^{Gag} protein to form virus like particles (VLPs).

(A-F) Transmission electron micrographs showing *in vitro*-assembled virus like particles from purified recombinant full-length MPMV Pr78^{Gag}-His$_6$-tag fusion protein expressed from FN1 in the presence of yeast tRNA. (G and H) Electron micrographs of negative controls comprising of assembly buffer and yeast tRNA only without any protein. Scale bar = 50 nm; 135,000X magnification. Figure & legend adapted from Pitchai et al., 2018.

2.4.7 Recombinant Pr78^{Gag}-His$_6$-tag Protein Expressed in Eukaryotic Cells can form VLPs Capable of Packaging Unspliced Transfer Vector RNA

Finally, in order to ensure that MPMV full-length Pr78^{Gag}-His$_6$-tag is able to encapsidate transfer vector RNA, two full-length MPMV Gag eukaryotic expression plasmids were created, with or without the His$_6$-tag (FN7 and FN9, respectively (Figure 24A). To ensure proper export of the MPMV Pr78^{Gag}-His$_6$-tag mRNA out of the nucleus, the MPMV CTE was inserted immediately downstream of the MPMV Gag stop codon (Figure 24A). The ability of these full-length Gag expression plasmids to encapsidate MPMV RNA was tested by employing a two-plasmid genetic complementation assay (Figure 24B). In this assay, either FN7 or FN9 were co-transfected with the MPMV transfer vector, SJ2, to test their ability to package the MPMV transfer vector RNA expressed from SJ2 (Jaballah et al., 2010). To monitor

transient transfection efficiencies, a secreted alkaline phosphatase (SEAP) expression plasmid (pSEAP) was also included in the transfection DNA cocktail.

As can be seen in Figure 24, the transfected 293T cells revealed successful expression of the full-length MPMV Pr78^{Gag} proteins using anti-Pr78^{Gag} polyserum and anti-β-Actin monoclonal antibody as a control (Figure 24C; panels I and II, respectively). Western blot analysis of virus particles isolated from transfected culture supernatants further confirmed VLP production by both the His (+) and His (-) Gag-expression plasmids, FN7 and FN9, respectively (Figure 24C; panel III). These results clearly demonstrate that the presence of His$_6$-tag at the C-terminus of MPMV full-length Pr78^{Gag} did not interfere with the expression of recombinant full-length MPMV Pr78^{Gag}-His$_6$-tag fusion protein or its ability to form Gag VLPs in 293T cells.

Next, the ability of VLPs produced by the recombinant MPMV full-length Pr78^{Gag} in eukaryotic cells to package MPMV transfer vector RNA was tested. Towards this end, RNA was extracted from the cytoplasmic fractions as well as the pelleted viral particles and DNase-treated to deplete any contaminating plasmid DNA from the transfected cultures. PCR conducted on the DNased-RNAs using transfer vector RNA-specific primers (OTR 1161 and OTR 1163) confirmed the absence of any contaminating plasmid DNA in the RNA preparations (data not shown) and this was followed by their conversion into cDNAs. To ensure that the fractionation technique was not compromised and there was no leakage of RNA from the nucleus to the cytoplasm we tested for the absence of unspliced actin mRNA in cytoplasmic fractions during this process (data not shown), as described by a group earlier (Aktar et al., 2014; Jaballah et al., 2010; Kalloush et al., 2016; Mustafa et al., 2012; Rizvi et al., 2010). RT-PCR amplification of transfer vector (SJ2)-specific RNA revealed its successful nuclear transport and expression in the cytoplasm (Figure 24C; panel IV).
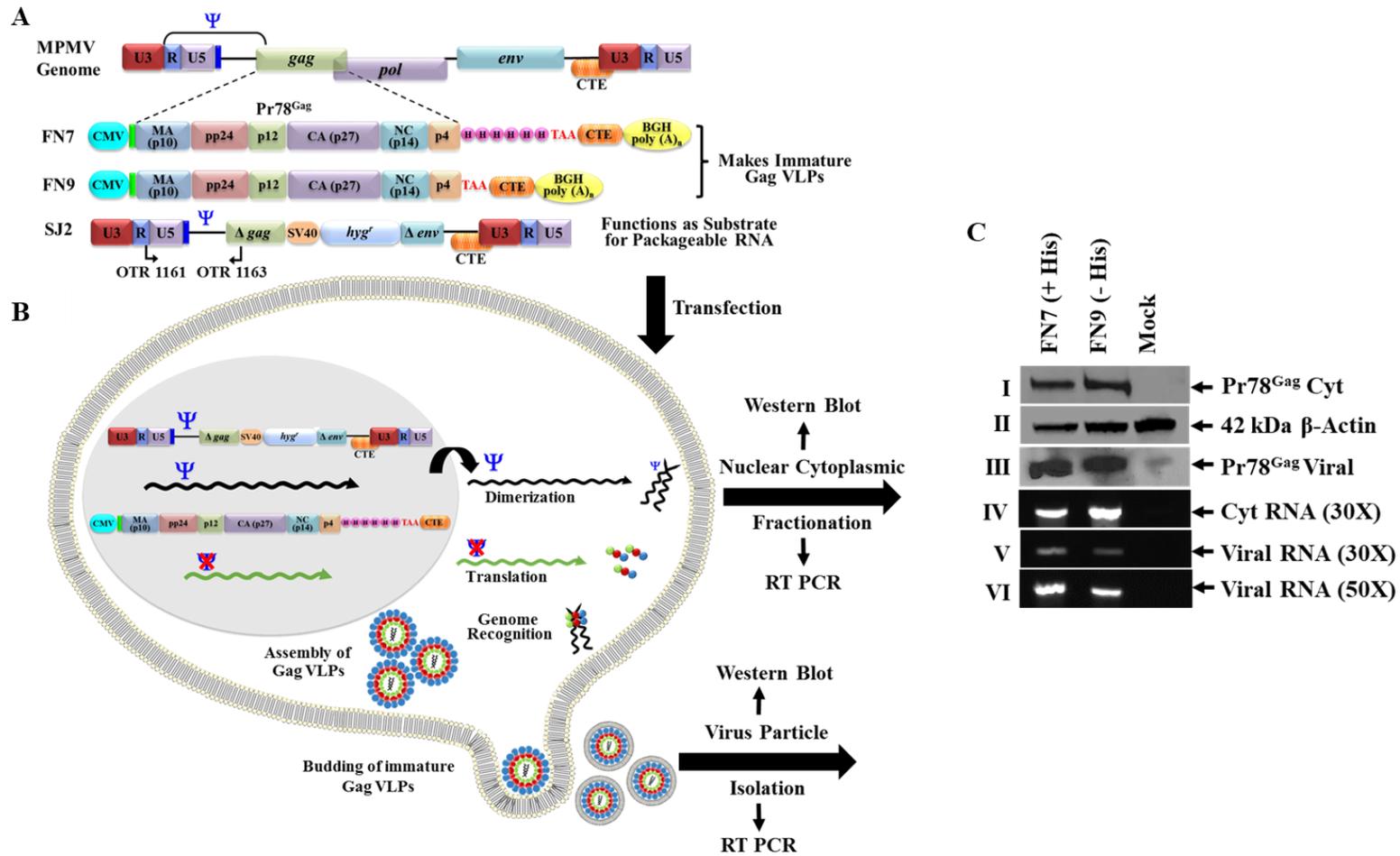
Figure 24: Two-plasmid genetic complementation assay to test the ability of MPMV virus like particles (VLPs) to package transfer vector viral RNA following Pr78^{Gag}-His$_6$-tag expression in eukaryotic cells.

(A) Graphical representation of MPMV full-length Gag expression plasmids with and without His$_6$-tag (FN7 and FN9 respectively) and MPMV transfer vector, SJ2 which provides the substrate for packageable RNA. (B) Design and rationale of the MPMV 2-plasmid genetic complementation assay. VLPs produced by the eukaryotic Gag expression plasmids (FN7 and FN9) should allow packaging of the transfer vector RNA expressed from SJ2 due to the presence of the packaging signal (Ψ) on its RNA. The 293T cells were co-transfected with the either of the two Gag-expression plasmids along with SJ2 and fractionated into nuclear and cytoplasmic fractions. The cytoplasmic fractions were analyzed for transfer vector (SJ2) RNA expression, while the virus particles were tested for their ability to package SJ2 RNA using RT- PCR. (C) MPMV full-length Gag expression plasmids with (FN7) and without His$_6$-tag (FN9) were transfected into 293T cells along with SJ2 and western blots were performed on cell lysates to detect Gag proteins using anti-Pr78^{Gag} polyserum (panel I) and β-actin proteins using a monoclonal antibody as a control (panel II). Western blots on ultracentrifuged virus particles using anti-Pr78^{Gag} polyserum (panel III). RT-PCR using MPMV transfer vector (SJ2)-specific primers (OTR1161 and OTR1163) amplifying a 530 bp fragment from the cDNAs prepared from the cytoplasmic (panel IV) and virion RNAs amplified for 30 (panel V) and 50 (panel VI) cycles, respectively. Experiments were conducted multiple times with reproducible results and representative blots are shown. Figure & legend adapted from Pitchai et al., 2018.

Amplification of the SJ2-specific cDNAs isolated from Gag VLPs formed by either the MPMV Pr78^{Gag}-His$_6$-tag fusion protein (FN7) or without His$_6$-tag (FN9) revealed that transfer vector RNA was efficiently packaged by both types of VLPs (Figure 24C; panels V and VI). Interestingly, the VLPs formed by the C-terminally His$_6$-tagged version of the Gag protein seemed slightly more efficient in encapsidating the transfer vector RNA than its untagged version (Figure 24C; panels V and VI). Since the Gag NC is thought to be the main domain involved in the interaction with the genomic RNA using its positively-charged zinc finger domains, addition of the His$_6$-tag may have increased the basic nature of the polyprotein and stabilized its interaction with the RNA further. Therefore, the slight increase in packaging could be attributed to the increased positive charge added to the full-length MPMV Pr78^{Gag} due to the presence of positively charged His$_6$-tag as has recently been suggested for HIV-1 Pr55^{Gag}- nucleic acid interaction *in vitro* (Bewley et al., 2017).

2.5 Funding

Chapter 3: Identification of Pr78$^{\text{Gag}}$ Binding Sites on the Mason-Pfizer Monkey Virus Genomic RNA Packaging Determinants

3.1 Abstract

How retroviral Gag proteins recognize the packaging signals (*Psi*) on their genomic RNA (gRNA) is a key question and this is addressed in this study using MPMV as a model system by combining band-shift assays and footprinting experiments. Data presented in this chapter shows that Pr78$^{\text{Gag}}$ selects gRNA against spliced viral RNA by simultaneously binding to two single stranded loops on the MPMV *Psi* RNA: (1) a large purine loop (ssPurines), and (2) a loop which partially overlaps with a mostly base-paired purine repeat (bpPurines) and extends into a GU-rich binding motif. Importantly, this second Gag binding site is located immediately downstream of the mSD site and is thus absent from the spliced viral RNAs. Identifying elements crucial for MPMV gRNA packaging should help in understanding not only the mechanism of virion assembly by retroviruses, but also facilitate construction of safer retroviral vectors for human gene therapy.

3.2 Introduction

Specific selection of the retroviral genome is central to the process of virion assembly during which a dimeric form of retroviral genomic RNA (gRNA) is selectively packaged into the nascently forming virions (Dilley et al., 2011; D'Souza & Summers, 2005; Johnson & Telesnitsky, 2010; Moore & Hu, 2009; Paillart et al., 2004; Rein, 2019). Despite the fact that the viral gRNA constitutes only ~1% of the total RNA in the cell milieu it is still specifically selected from a vast array of spliced viral and cellular RNAs (Abd El-Wahab et al., 2014b; Berkowitz et al., 1996; Dubois

et al., 2018a; Maldonado & Parent, 2016; Kuzembayeva et al., 2014; Lever, 2007; Mailler et al., 2016). This highly controlled and selective process is dependent on two important factors: (1) the presence of specific sequences or structures within the gRNA, and (2) the retroviral precursor polyprotein Gag and its ability to identify and bind to these unique sequences or structures (Abd El-Wahab et al., 2014; Ali et al., 2016; Berkowitz et al., 1996; Bernacchi et al., 2017; Dubois et al., 2018; Maldonado & Parent, 2016; Kuzembayeva et al., 2014; Mailler et al., 2016).

All retroviruses harbor *cis*-acting sequences referred to as the packaging signal (*Psi*/Ψ) that are indispensable for the selective packaging of their genome. For most of the retroviruses, these sequences are located at the 5' untranslated region (UTR) and extend into *gag* (Ali et al., 2016; Comas-Garcia et al., 2016; Dilley et al., 2011; D'Souza & Summers, 2005; Dubois et al., 2018a; Gherghe et al., 2010; Johnson & Telesnitsky, 2010; Kenyon et al., 2008; Miyazaki et al., 2011; Moore et al., 2009; Paillart et al., 2004; Rein, 2019; Rizvi et al., 2010). These *Psi* sequences harbor unique Gag binding sites important for selective gRNA packaging in retroviruses (Comas-Garcia et al., 2016; Webb et al., 2013). Owing to the flexibility of RNA, the *Psi* elements are capable of assuming secondary structures, leading to the formation of stem loops (SLs), single-stranded (ss) regions, and long-range interactions (LRIs). These structural motifs are involved in the specific selection of gRNA independent of the primary sequence (Aktar et al., 2013, 2014; Bernacchi et al., 2017; Kalloush et al., 2019; Kenyon et al., 2011; Miyazaki et al., 2011; Smyth et al., 2015).

A key player in the process of specific gRNA packaging is the retroviral Gag polyprotein. The three major domains of the polyprotein are matrix (MA), capsid (CA) and nucleocapsid (NC), found ubiquitously in all retroviruses (Ali et al., 2016; Dubois et al., 2018a; Mailler et al., 2016; Rein, 2019). From the time of its synthesis in the

cytoplasm to virion release and maturation, the various domains of Gag drive the assembly process and are responsible for one or more key events in the retroviral life cycle. One of these events is to identify and bind to specific sequences on the *Psi* RNA and bring about selective packaging of its gRNA. Although Gag domains are most often studied independently of each other, their various functions are achieved within the native context of the polyprotein. These domains are liberated from one another only upon virus maturation post virion assembly and release from the infected cell. Among these domains, the NC is primarily responsible for the selective binding and packaging of gRNA in most retroviruses. It is a highly basic and hydrophobic protein, harboring two highly conserved zinc finger domains. These fingers consist of CCHC arrays (C-X$_2$-C-X$_4$-H-X$_4$-C; where C = Cys, H = His, X$_n$ = n number of other amino acids) that sequester zinc ions required for specific gRNA binding (Ali et al., 2016; Dubois et al., 2018; Mailler et al., 2016; Rein, 2019).

A majority of studies conducted to date to identify the specific high affinity binding sites of Gag on retroviral *Psi* RNA have been restricted to the NC domain alone, either in its immature or mature form (Amarasinghe et al., 2000; De Guzman et al., 1998; D'Souza & Summers, 2004; Fisher et al., 1998; Morcock et al., 2000, 2002; Stewart-Maynard et al., 2008; Urbaneja et al., 2000; Wilkinson et al., 2008; Zhou et al., 2007). Preliminary studies carried out on human immunodeficiency virus type 1 (HIV-1) indicated the high affinity binding site of its NC domain to be located on the apical loop of SL3 (Amarasinghe et al., 2000; De Guzman et al., 1998; Wilkinson et al., 2008). However, use of full-length HIV-1 Pr55^{Gag} has revealed the presence of a high affinity binding site on the internal (G//AGG) loop of SL1 of the HIV *Psi* RNA (Abd El-Wahab et al., 2014). These findings suggest the importance of studying the selective packaging of retroviral gRNA in the context of the full length precursor

polyprotein, Gag. However, since NC plays its role as part of full-length Gag, a role of domains other than NC cannot be excluded. In fact, a number of studies have implicated other regions of the Gag domains in specific binding to gRNA, including the MA (Alfadhli et al., 2011; Comas-Garcia et al., 2017; Kroupa et al., 2020; Lu et al., 2011; Pachulska-Wieczorek et al., 2016), CA (Guo et al., 2005), p1 (Roy et al., 2006), p2 (Kaye & Lever, 1998; Russell et al., 2003) and p6 (Dubois et al., 2018b; Zhang & Barklis, 1997) domains of HIV-1. Results from these studies suggest that all three major domains of Gag (MA, CA, and NC) and the other domains are capable of binding to various structured motifs on the *Psi* RNA with varying affinities, emphasizing the importance of studying selective *Psi* RNA packaging in the native context of full-length Gag.

Mason Pfizer monkey virus (MPMV) is the most widely studied prototypic type D beta-retrovirus. It was first isolated from the breast adenocarcinoma of a rhesus monkey (*Macaca mulatta*) and is known to cause fatal immunodeficiency in macaques (Bryant et al., 1986; Fine et al., 1975). MPMV serves as a potential candidate for the development of gene therapy vectors due to its phylogenetic distance from human retroviruses, such as HIV-1. It harbors promoters that are functional in human cells and also its constitutive transport element (CTE), analogous to the HIV-1 Rev responsive element (RRE), allows for efficient cytoplasmic transport of viral RNA independent of any viral protein (Bray et al., 1994; Rizvi et al., 1996a; 1996b; 1997). MPMV has also been widely studied to decipher the assembly process of retroviruses and thus can be used as an experimental tool to investigate potential inhibitors of retroviral particle assembly (Bush & Vogt, 2014). It differs distinctly from C-type retroviruses such as HIV-1 by adopting an intracytoplasmic A-type morphology with spherical capsids.

The 5' end of MPMV has been extensively investigated to demarcate the boundaries of the minimal packaging sequences required for efficient incorporation of gRNA into viral particles (Aktar et al., 2013; Ali et al., 2020; Guesdon et al., 2001; Harrison et al., 1995; Jaballah et al., 2010; Kalloush et al., 2016, 2019; Pitchai et al., 2018; Schmidt et al., 2003; Vile et al., 1992). These sequences spans from the 5' UTR into the *gag* open reading frame, similar to most other retroviruses (Ali et al., 2016). Systematic mutational analyses carried out on this region has revealed a discontinuous or bipartite signal consisting of the first 50 nts of the 5' UTR, inclusive of the palindromic stem loop (Pal SL) that serves as the dimerization initiation site (DIS) for the gRNA, and the last 23 nts of the 5' UTR followed by the first 120 nts of *gag,* both of which are required for successful MPMV gRNA packaging (Figure 25A; Schmidt et al., 2003; Jaballah et al., 2010). Consistent with other retroviruses, the *Psi* sequences on MPMV gRNA fold into a higher order structure comprising of various structural motifs (Aktar et al., 2013; Jaballah et al., 2010). Among these structural motifs, two purine-rich motifs, the single stranded purines (ssPurines; U^{191}UAAAAGUGAAAGUAA^{206}) and the base paired purines (bpPurines; G^{246}AAAGUAA^{253}), have been identified as unique regions in their composition and positioning that may contribute to Gag binding (Aktar et al., 2013; Jaballah et al., 2010; Figure 25B). Presence of purine-rich sequences in the MPMV *Psi* RNA is consistent with the fact that a stretch of purines in the *Psi* of other retroviral gRNAs has been proposed to facilitate gRNA packaging by functioning as a potential NC binding site (Abd El-Wahab et al., 2014; Bernacchi et al., 2017; Ding et al., 2020; Gherghe et al., 2010; Keane et al., 2015; Lever, 2009; Moore et al., 2009a; Moore & Hu, 2009; Nikolaitchik et al., 2020; Paillart et al., 1997; Zeffman et al., 2000).
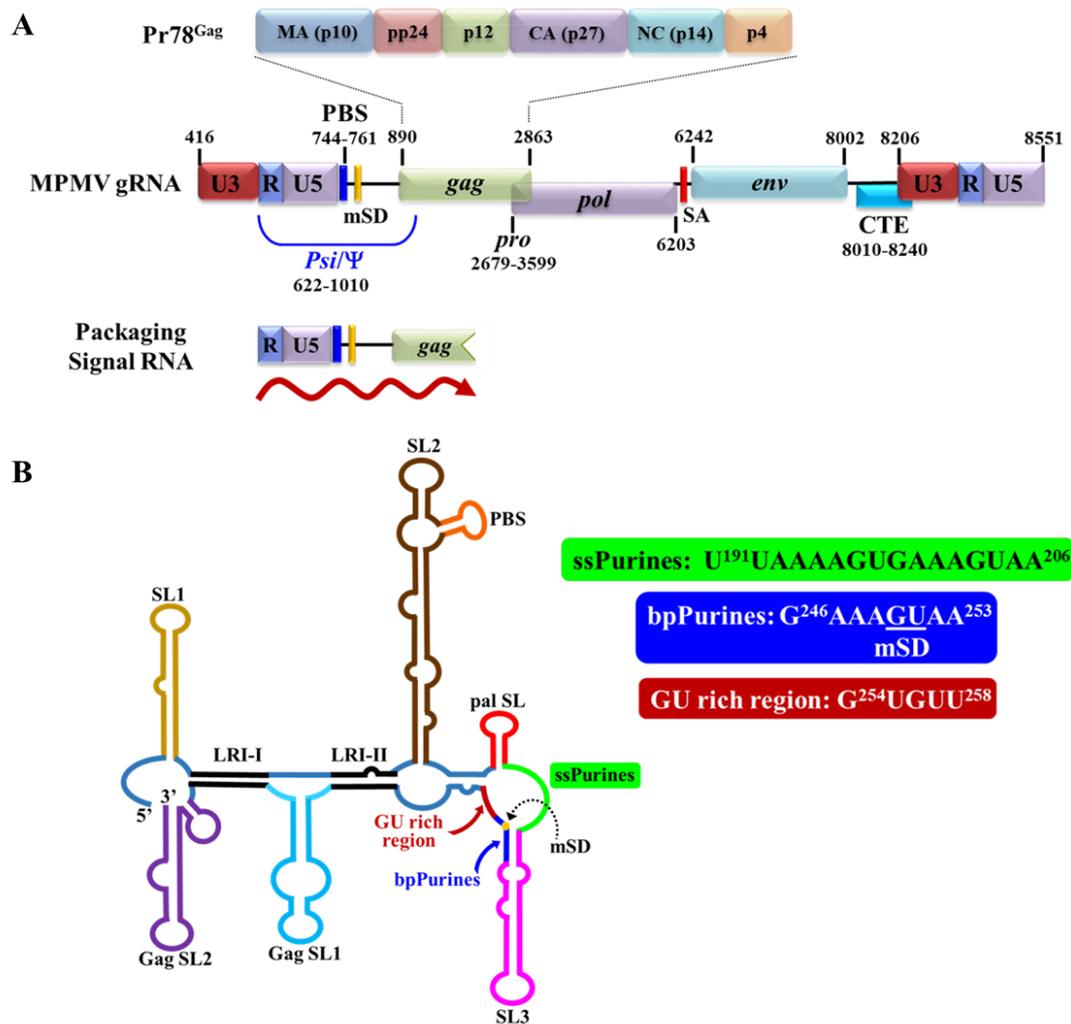
Figure 25: Schematic representations of MPMV genome, organization of the different domains of full-length MPMV Gag precursor (Pr78[Gag]), and higher order structure of MPMV packaging signal RNA.

(A) Organization of the MPMV genome, domain organization of MPMV Gag precursor (Pr78[Gag]), and 5' end of the genome that contains the packaging signals (*Psi*/ Ψ). Essential elements of the genome are annotated and their nucleotide numbers correspond to the MPMV genome with the Genbank accession number M12349.1. The red squiggly arrow denotes the transcript initiating from the R region of the packaging signal RNA. (B) Schematic representation of the RNA secondary structure of MPMV packaging determinants located at the 5' end of the genome showing multiple structural motifs including: stem-loops (SL1-3); LRI, long range interactions (LRIs); PBS, primer binding site; palindromic stem-loop (pal SL); ssPurines & bpPurines, single-stranded and based-purines; mSD, major splice donor; Gag SL1-2, Group antigen stem loops 1-2.

The ssPurines present in MPMV *Psi* RNA by far forms the largest single-stranded purine rich region found in any widely studied retroviral *Psi* RNA (Figure 25B). It consists of 16 single-stranded nucleotides with 75% purines, and more

importantly, is located immediately downstream of the DIS which makes it a potential motif for Gag binding. This observation is consistent with HIV-1 where the primary Gag binding site (G//AGG) is located on the internal loop of SL1, whose apical loop also functions as DIS (Abd El-Wahab et al., 2014). Interestingly, the latter half of the ssPurines sequence is found repeated downstream in a base paired manner to form the bpPurines (Figure 25B; Jaballah et al., 2010). Employing genetic, biochemical, and structure-function approaches, a recent study has pointed towards these two purine-rich regions (ss- and bp-Purines) functioning as redundant packaging motifs and possible Gag binding sites during viral assembly (Ali et al., 2020). Therefore, the current study was undertaken to establish whether these two purine-rich regions (ss- and bp-Purines) on MPMV *Psi* RNA truly function as MPMV Gag precursor polyprotein (Pr78^{Gag}) binding sites (Figure 1B). Results indicate that the MPMV polyprotein, Pr78^{Gag} binds to two loops: 1) the ssPurines loop (U^{191}UAAAAGUGAAAGUAA^{206}) and 2) a second loop (A^{252}AGUGU^{257}) corresponding to the last two purines of the bpPurines and extending into a GU-rich region (Figure 1B). Interestingly, this second binding site is located immediately downstream of the mSD and is thus absent from the spliced viral RNAs (Figure 1B). Finally, we propose a model for the specific selection of full length unspliced MPMV RNA over cellular and viral spliced *env* RNA by Pr78^{Gag}.

3.3 Materials & Methods

3.3.1 Nucleotide Numbering System

Nucleotide numbers in this study refer to the MPMV genome with the Genbank accession number M12349.1 (Sonigo et al., 1986).

3.3.2 Expression and Purification of Pr78^{Gag}

Pr78^{Gag} was expressed with a C-terminal hexa-histidine (His$_6$) tag and purified via immobilized metal affinity chromatography (IMAC) followed by size exclusion chromatography (SEC). The purified protein was characterized using western blot. Methodology employed to express and purify the protein has already been described (Chapter 2).

3.3.3 Physical Characterization of Pr78^{Gag} by Dynamic Light Scattering (DLS)

Prior to the *in vitro* assays, purified Pr78^{Gag} was characterized by DLS using a DynaPro Nanostar (100 mW He-Ne laser; Wyatt Technologies,) in a 1-µl quartz cuvette (JC-006, Wyatt Technologies) at 20°C as previously described (Abd El-Wahab et al., 2014). By assimilating the protein in solution to spheres, the diffusion coefficients (D) were correlated to the hydrodynamic radius (R$_h$) of the molecules in solution by the Stokes-Einstein equation:

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

In this equation, *k* represents the Boltzmann constant, while *T* represents the absolute temperature, and *μ* is the viscosity of the solvent. Before sample acquisition, the buffer was filtered through 0.2 µm filters (Millex ®) and the offset of the solvent was measured for subsequent sample data treatment.

3.3.4 Plasmid Construction for Spliced *env*, ss- and bp-Purines Mutant RNA Production

The wild type (WT) plasmid (RCR001; Figure 26A) was used for the *in vitro* transcription of the MPMV unspliced full-length *Psi* RNA, as previously described (Aktar et al., 2013). Plasmids for the *in vitro* transcription of the MPMV spliced *env* RNA (FN42; Figure 26A) and the ss- and bp-Purines mutants were created using spliced overlap extension (SOE) PCR, as previously described (Aktar et al., 2013; Gibbs et al., 1994; Kalloush et al., 2019, 2016). For cloning FN42, PCR A was performed using the outer forward primer or sense (S) primer OTR 1004 (Appendix A) and the inner reverse primer or anti-sense (AS) primer OTR 1458 (Appendix A) using MPMV sub-genomic transfer vector (SJ2) as the template (Aktar et al., 2013; Jaballah et al., 2010; Kalloush et al., 2016, 2019). PCR B was performed employing the inner forward primer (S) OTR 1378 (Appendix A) and the outer reverse primer, (AS) OTR 1379 (Appendix A), on the MPMV full-length molecular clone KAL01 as the template (Rizvi et al., 1996b). A second round of amplification was carried out using the products of PCRs A & B with primers OTR 1004 and OTR 1379. The resulting product was cleaved with *Hind*III and *Xma*I and ligated into the similarly-digested pUC-based vector, pIC19R (Aktar et al., 2013; Ali et al., 2020; Jaballah et al., 2010; Kalloush et al., 2016, 2019). The sequence of FN42 was then confirmed via sequencing (Macrogen, South Korea) and subsequently used for *in vitro* transcription. Mutations in the ss- and bp-Purines present in the MPMV packaging determinants were also introduced employing the same strategy, but using SJ2 as the template along with primers listed in Appendix A.
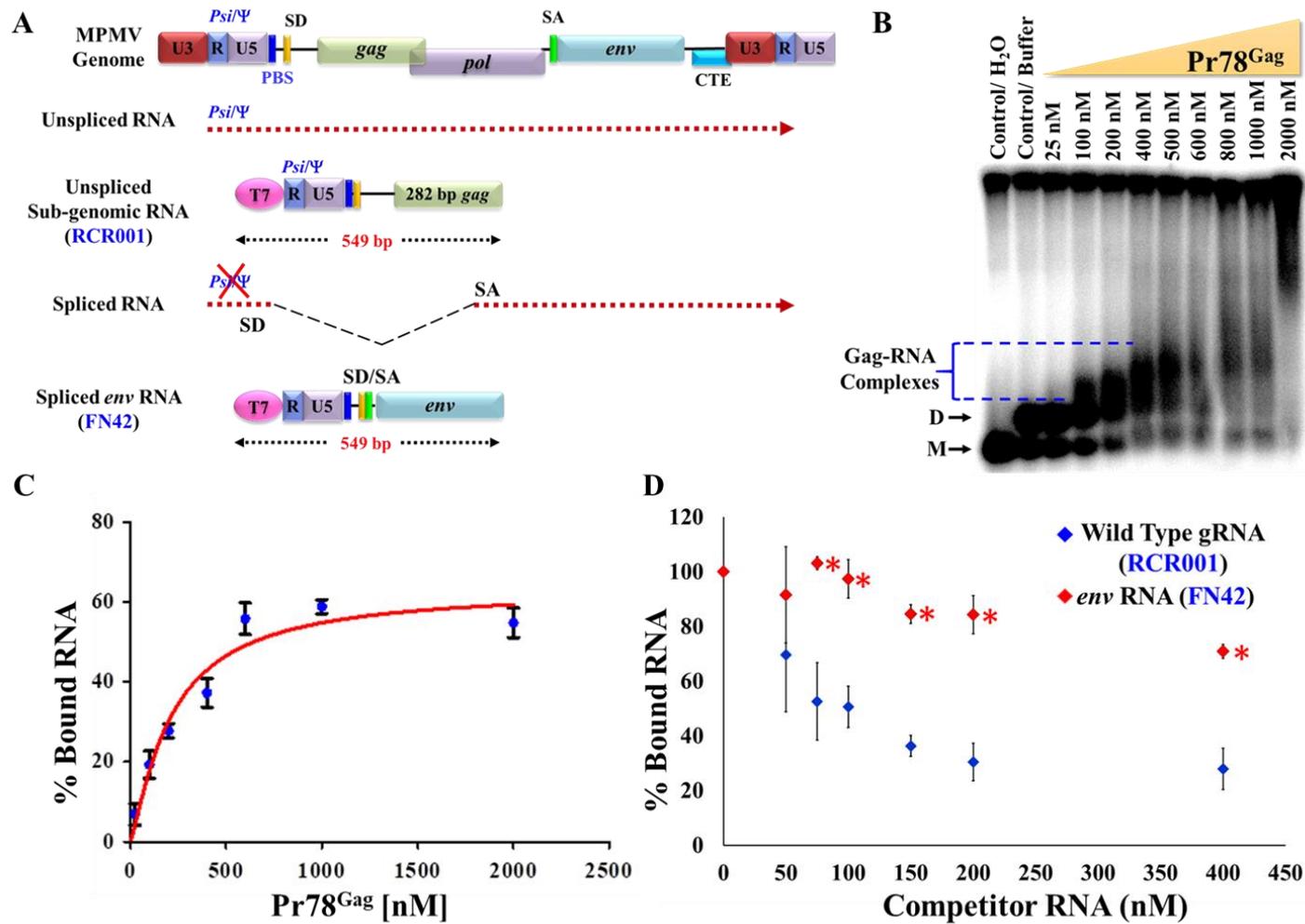
Figure 26: Pr78^{Gag} binds preferentially to MPMV packaging signal RNA.

(A) Graphic representation of both the *in vitro* transcribed unspliced (WT; RCR001; 549 nts) gRNA and spliced *env* RNA (FN42; 549 nts) used in band-shift competition assays (B) Representative gel of a band-shift assay performed using radiolabeled WT MPMV RNA (RCR001) in the presence of increasing concentrations of Pr78^{Gag}. Monomeric (M) and dimeric (D) RNA forms are labeled on the gels. (C) Saturation plot obtained by quantifying the bands from the band shift assay. Experimental data were fit to Hill's equation. The best fit was obtained with plateau = 63±9, Hills coefficient = 1.2±0.4 and Kd = 216±76 nM (mean±SD); R^2= 0.947. (D) Graph showing the quantitation of the gels representing the percentage of radiolabeled bound WT MPMV (RCR001) RNA in the Gag-RNA complexes *versus* increasing concentrations of the respective competitor RNA. The red (FN42) asterisks (*) indicate the statistically significant differences between spliced env RNA and WT (RCR001) RNA at different data points in the quantitated gels, calculated using a paired two tailed *t*-test.

3.3.5 *In vitro* Transcription and Purification of Unlabeled and [α-^{32}P]-Labeled RNA

All plasmids containing the WT (RCR001) and mutant MPMV packaging sequences under the influence of the T7 promoter were linearized by cleaving the plasmid DNAs with *Sma*I and used for *in vitro* transcription (MEGAscript™ T7 Transcription Kit, Thermo Fisher Scientific). Briefly, the linearized plasmids were incubated at 37ºC for 4 hours in the presence of NTPs, 10X reaction buffer and T7 RNA polymerase, followed by a 15-minute incubation at the same temperature with 1µl of TURBO DNase (2U/ µl). Quality of the *in vitro* transcribed RNA was determined by testing 2µl of the product on 8%, denaturing (8M urea) polyacrylamide gels following which the remainder of the product was ethanol precipitated overnight at -20 ºC. The precipitated RNAs were then purified by gel filtration chromatography using TSK Gel G2000SW columns (Tosoh Bioscience) in 0.2 M sodium acetate (pH 6.5) and 1% (v/v) methanol, as previously described (Abd El-Wahab et al., 2014; Marquet et al., 1991; Paillart et al., 1996a; 1996b; 1994; Sinck et al., 2007). Finally, RNAs collected from appropriate fractions were pooled, ethanol precipitated, and examined for purity and integrity using 8M urea denaturing 8% polyacrylamide gels.

In vitro transcription in the presence of [α-^{32}P]-ATP was performed to prepare internally-labeled RNAs, as described previously (Abd El-Wahab et al., 2014; Paillart et al., 1996a; 1996b; 1994; Sinck et al., 2007). Following *in vitro* transcription, the RNA samples were DNase treated, and the labeled RNAs purified by electrophoresis on a 8% polyacrylamide gel under 8M urea denaturing conditions. RNAs from the bands were excised and extracted in 300 µl of buffer containing 500 mM ammonium acetate, 1 mM EDTA and 0.1% SDS overnight at 4ºC, followed by ethanol precipitation, and resuspension in 10 µl Milli-Q water.

3.3.6 Band-Shift and Competitive Band-Shift Assays

For band-shift assays, the radiolabeled unspliced, full-length WT (RCR001) RNA (50,000 cpm) and yeast tRNA (2 μg) were denatured at 90°C (2 minutes) followed by chilling on ice (2 minutes). The denatured RNAs were re-folded in 1X dimer buffer (30 mM Tris pH 8.0, 300 mM NaCl, 5 mM MgCl$_2$, 5 units of RNase inhibitor (RNasin, Promega), 0.01% Triton-X 100, in a total volume 10 μl) at 37°C for 30 min. Next, Pr78^{Gag} (diluted in 30 mM Tris (pH 8.0), 300 mM NaCl, 5 mM MgCl$_2$, 10 mM DTT, and 0.02 mg/ml BSA in a final volume of 10 μl) in increasing concentrations (0 to 2000 nM) was mixed with the refolded RNA. The mixture was incubated at 37°C for 30 minutes to allow for binding, followed by incubation on ice for 30 minutes. Samples were separated on 1% agarose gels using TBM buffer (0.5X Tris-Borate, 0.1 mM MgCl$_2$) at 150 V for 4 hr at 4°C. The gels were then fixed in 10% trichloroacetic acid (TCA) for 10 minutes, dried under vacuum, and analyzed using a FLA 5000 (Fuji) scanner. Bands on the gels were quantified using ImageQuant software. The experimental data were fit with Hill's equation shown below using the GraphPad Prism 5 software.

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

In the Hill's equation, B$_{max}$ represents the maximum specific binding while h is the Hill's coefficient.

For competitive band-shift assays, 50,000 cpm of radiolabeled, unspliced, full-length WT (RCR001) RNA were denatured and refolded in the presence of either spliced *env* RNA (FN42) or increasing concentrations of unlabeled competitor RNAs (0 to 400 nM), as described above. The refolded RNAs were then mixed with Pr78^{Gag}

at a final concentration of 500 nM in a volume of 20 μl, and incubated at 37°C for 30 min to allow for binding and then on ice for 30 min for stabilization. The reaction mixtures were then electrophoresed on 1% agarose gels using TBM buffer (0.5X Tris-Borate, 0.1 mM MgCl$_2$) at 150 V for 4 hr at 4°C, fixed in 10% trichloroacetic acid (TCA) and quantified as described above. The percentage of bound RNA were calculated from triplicate gels in each case and the statistical significance between the WT MPMV RNA (RCR001) and the spliced *env* RNA (FN42) and competitor RNAs were determined using a paired two tailed *t*-test.

3.3.7 RNA Footprinting and High-throughput Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (hSHAPE)

Biochemical analysis of the structures of the mutants was performed using the hSHAPE methodology that allows structural investigation at each nucleotide following structure-dependent modification of the RNA (Abd El-Wahab et al., 2014; Aktar et al., 2013; Ali et al., 2020; Kalloush et al., 2016, 2019; Mustafa et al., 2018). Briefly, a single hSHAPE reagent (such as benzoyl cyanide; BzCN) allows the modifications of all 4 nucleotides that are single-stranded and hence more reactive to hSHAPE reagents, whereas the base-paired nucleotides being structurally constrained are not reactive or much less reactive. To identify the Pr78^{Gag} footprints on MPMV gRNAs, RNA-protein interactions were monitored by modifying WT RNA with BzCN, both in the absence and presence of Pr78^{Gag}. The nucleotides showing attenuated hSHAPE reactivity in the presence of Pr78^{Gag} suggested that the protein binds to these RNA regions to form an RNA-protein complex, preventing these nucleotides from modification by the hSHAPE reagents. Such an approach has been successful in the case of HIV-1 (Abd El-Wahab et al., 2014) and MMTV

(Chameettachal et al., 2021) in order to identify Gag binding sites on their respective gRNAs.

Footprinting experiments were performed on either the WT (RCR001) MPMV gRNA or the mutant RNA, FN26, in the presence (6 µM) or absence of Pr78^{Gag}, as previously described (Abd El-Wahab et al., 2014). *In vitro* transcribed WT gRNA (RCR001; 1 pmol) and a four-molar excess of competitor RNA (spliced *env* RNA; FN42) were denatured at 90°C for 2 min and then renatured on ice for 2 min, followed by the addition of excess yeast tRNA (2 µg), RNasin (5U) and 5X HEPES buffer in a total volume of 10 µl. The RNA concoction was then incubated at 37°C for 30 min. Frozen Pr78^{Gag} protein samples were thawed on ice and spun at 13,500 rpm for 15 min. The 20 µM Pr78^{Gag} stock was diluted in 5X HEPES buffer and mixed with the refolded RNA at a final concentration of 6 µM in 20µl. RNA in the folding buffer without any Pr78^{Gag} was used as a control. The RNA-protein complexes were incubated at 37°C for 30 min and then cooled on ice for 30 min. Following RNA-protein incubations, the complex was modified by 2 µl of 100 mM BzCN by incubating for a minute at room temperature. Similarly, for the control (unmodified RNA-protein sample), 2 µl of anhydrous DMSO was added and incubated in the same manner. The modification reaction was stopped by adding 78 µl of nuclease free water and the chemically modified RNA samples were extracted (Roti®-Phenol/Chloroform/Isoamyl alcohol), ethanol precipitated, air dried, resuspended in 7 µl of nuclease free water, and subjected to primer extension as previously described (Abd El-Wahab et al., 2014; Aktar et al., 2013).

Briefly, reverse transcription of the modified and unmodified RNAs was carried out, using two sets of (OTR 18/19 and 22/23). One primer within each set was labeled with either VIC (OTR18 and OTR22) or NED (OTR19 and OTR23; Appendix

A). For elongation of both the modified and unmodified samples, 1 µl of each of the VIC labelled primers OTR 18 (1 µM) and 22 (2 µM) were added to the resuspended RNA and incubated at 90ºC for 2 min, then cooled on ice for 2 min. 2 µl of 5X RT buffer was added to each of the samples and incubated at room temperature for 10 min. Following this, 10 µl of the elongation mix (2µl of 5X RT Buffer, 0.6 µl of 25 mM dNTP and 2U of AMV RT (Life Science) was added to each of the tubes and incubated at 42ºC for 20 min, 50ºC for 30 min and 60ºC for 10 min. For the ddG sequencing ladder, 2 pmol of untreated WT RNA (RCR001) and 1 µl of the NED labelled primers OTR 19 (1 µM) and 23 (2 µM) were incubated at 90ºC for 2 min and cooled on ice for 2 min. 2µl of 5X RT buffer were added to the tubes and incubated at room temperature for 10 min. Following this, 10.2 µl of the sequencing mix (2 µl of 5X RT Buffer, 2 µl of 100 µM ddGTP, 6 µl of G10 (0.25 mM dGTP, 1mM dATP, 1mM dCTP, 1mM dTTP) and 2U of AMV RT (Life Science) was added to each of the tubes and incubated as above. 80 µl of nuclease free water were added and cDNA was extracted using Roti® Aqua-Phenol/Chloroform/Isoamyl alcohol (Carl Roth). For each experiment, the aqueous phase from the modified or unmodified samples were pooled with the aqueous phase of the ddG sequencing ladder. The samples were then ethanol precipitated and resuspended in 10 µl of HiDi Formamide (ABI). The samples were then incubated at 90ºC for 5 min, then on ice for 5 min, centrifuged at 13,500 rpm for 15 min, and loaded onto a 96-well plate for sequencing (Applied Biosystems 3130xl genetic analyser).

The electropherograms obtained were analyzed with QuShape (Karabiber et al., 2013) to extract reactivity data for each sample. The mean reactivity data from at least three to four independent experiments were obtained for each sample and used to obtain the validated structures of each RNA (Appendices C and D). Reactivity data

obtained from samples treated in the absence of Pr78^{Gag} was applied as constraints to either the WT MPMV sequence (nt 232 to nt 1171) or FN26 sequence in RNAstructure (version 6.1; Reuter and Mathews, 2010). The dot bracket file obtained from RNAstructure was then used to translate the structural data into VARNA version 3-93 (Visualization Applet for RNA secondary structure; Darty et al., 2009) where the validated structure was redrawn. The reactivity data obtained in the presence of Pr78^{Gag} was applied onto the RNA structures obtained in the absence of Gag. The statistical significance between the SHAPE reactivities of each of the nucleotides obtained in the presence and absence of Pr78^{Gag} was calculated using a paired two tailed *t*-test and a significant decrease or increase in hSHAPE reactivity with a *p*-value less than <0.05 was considered significant (Appendices C and D).

hSHAPE experiments to biochemically validate the predicted secondary structure of the spliced *env* RNA were performed using the same methodology as explained above except in the absence of any protein. RNA was subjected to modification with appropriate controls and followed by reverse transcription using two sets of primers labelled with both VIC and NED (OTR_312-322 and OTR_497-518; Appendix A). QuShape analysis was performed as above and mean reactivity data from three independent experiments were obtained for secondary structure analysis (Appendix E).

3.4 Results

3.4.1 Characterization of Recombinant Full-length Purified MPMV Gag Polyprotein (Pr78^{Gag}-His$_6$-tag Fusion Protein) by Dynamic Light Scattering (DLS)

To determine whether the two purine-rich regions (ss- and bp-Purines) important for MPMV gRNA packaging act as binding sites for the Gag precursor polyprotein, large scale expression and purification of Pr78^{Gag} was performed and the protein characterized for its biological function, revealing that it could assemble *in vitro* into virus-like particles (VLPs) in bacteria and the VLPs produced in eukaryotic cells could encapsidate MPMV RNA containing the *Psi* region (*see* Chapter 2; Pitchai et al., 2018).

The bacterially expressed and full-length Pr78^{Gag} was characterized by DLS, which revealed that the purified protein did not contain any aggregates. The mean hydrodynamic radius (R$_h$) based on volume (percent) and number (percent) distribution was estimated to be 6.7 and 5.8 nm, respectively. This corresponded to a molecular weight of 288 and 206 kDa, respectively, indicative of Pr78^{Gag} trimers (Figure 27).
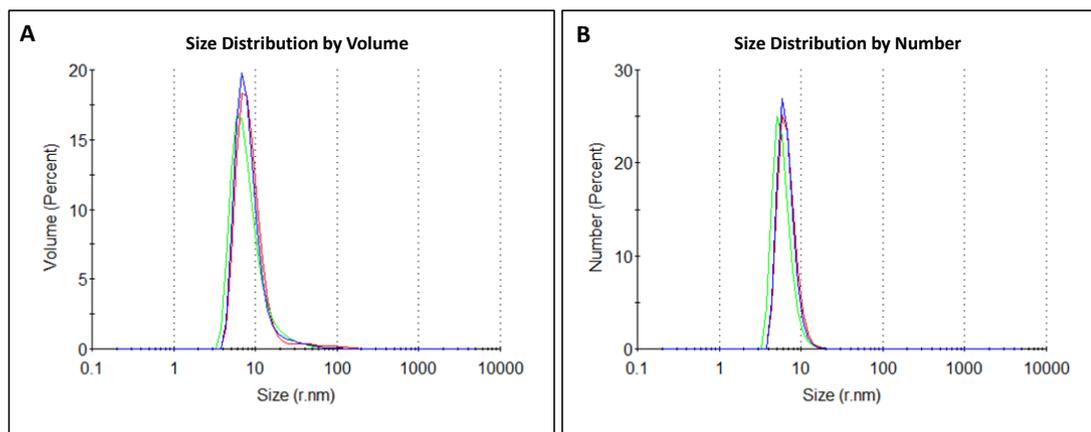
Figure 27: Characterization of MPMV Gag precursor Pr78^{Gag} by dynamic light scattering (DLS) in RNA binding buffer.

(**A**) Volume *versus* size distribution (**B**) Number versus size distribution. The blue, red and green peaks represent three independent Pr78^{Gag} samples tested in binding buffer.

3.4.2 Pr78^{Gag} Discriminates between Full-length, Unspliced Sub-genomic *Psi* RNA and Spliced *env* RNA

We first attempted to establish if Pr78^{Gag} specifically binds to the full-length, unspliced gRNA (RCR001) over spliced *env* RNA (FN42; Figure 26A). The T7-based plasmids expressing these RNAs were designed in a way that they would express RNAs of the same length (549 nucleotides). As a first step, band shift assays were performed to determine the optimal concentration of Pr78^{Gag} at which a complete shift of dimeric gRNA would occur upon successful Gag-RNA complex formation. These band shift assays were performed with a constant amount of radiolabeled unspliced gRNA (50,000 cpm, ~5 nM) against progressively increasing concentrations of Pr78^{Gag} (from 0 to 2000 nM). Results of these band shift assays indicated a complete shift of the dimeric gRNA at a protein concentration of 500 nM (Figure 26B). The band shift gel was then quantitated, and the data plotted to fit Hill's equation (Figure 26C). The Hill's coefficient was estimated to be 1.216 ± 0.422 (mean $\pm$ SD), suggesting that

binding of $Pr78^{Gag}$ to gRNA is weakly cooperative or non-cooperative. The apparent K_d obtained was 216.2 ± 76.10 nM (mean $\pm$ SD).

Having determined the optimal concentration of $Pr78^{Gag}$ for a complete shift of the dimeric gRNA, competitive band shift assays were performed to determine the differential ability of $Pr78^{Gag}$ to bind to unspliced gRNA (RCR001) *versus* spliced *env* RNA (FN42) of equal length (Figure 26A). This was accomplished by using a constant amount of radiolabeled (50,000 cpm) unspliced gRNA (RCR001) against progressively increasing concentrations (from 0 to 400 nM) of either non-labeled competitor RNA, which in this case was the unspliced (RCR001) or the spliced (FN42) *env* RNA (Figure 26A & D). The experiments with *env* RNA were performed in triplicates, the resultant gels were quantified, and the results depicted as a percentage of bound RNA in the protein-RNA complex *versus* competitor RNA (Figure 26D). These experiments revealed that increasing concentrations of the unspliced gRNA was capable of efficiently competing against the labeled unspliced gRNA (RCR001), displacing around ~50% of the bound RNA at a low concentration (75 nM) and ~80% of bound RNA at a maximum concentration of 400 nM. Comparatively, the spliced *env* RNA (FN42) competed poorly with the labeled unspliced gRNA with only around ~30% of the labeled RNA being displaced at the maximum concentration (Figure 26D). This is further supported by the statistically significant difference observed for the percentage of bound RNA between the unspliced gRNA (RCR001) and the spliced *env* RNA (FN42). These results clearly indicate the preferential binding of $Pr78^{Gag}$ to unspliced MPMV gRNA rather than to its spliced *env* RNA.

3.4.3 Pr78^{Gag} Binds Redundantly to both the ssPurines and bpPurines

The next step was to investigate where Pr78^{Gag} binds to on the full-length MPMV gRNA. Earlier genetic and biochemical analyses had identified two structural elements crucial for MPMV gRNA packaging, ssPurines and bpPurines (Aktar et al., 2013; Jaballah et al., 2010). Mutations introduced in these purine-rich regions had shown that both these motifs were responsible for efficient gRNA packaging in a redundant fashion (Ali et al., 2020). Thus, competitive band shift assays were performed on a number of mutant clones in these motifs to determine whether binding of Pr78^{Gag} to these regions could be observed (Figure 28A).

Using a similar experimental design as described for *env* RNA, the mutant RNAs were *in vitro* transcribed, and increasing concentrations of the mutant RNAs (from 0 to 400 nM; Figure 28A) were used as competitors against the radiolabeled unspliced gRNA (RCR001) to assess their ability to displace the radiolabeled unspliced gRNA from the protein-RNA complex. Experiments for each mutant clone were performed in triplicates, the resultant gels were quantified, and the results depicted as percentage of bound radiolabeled RNA in the protein-RNA complex *versus* increasing competitor RNA (Figure 28B & C). The data thus obtained was compared to that obtained from both the WT unspliced gRNA (RCR001) and spliced *env* RNA (FN42) to determine the competing (or non-competing) ability of these mutants to bind Pr78^{Gag}. Mutant FN26, containing a complete deletion of the ssPurines, demonstrated efficient competition against the labeled unspliced gRNA, revealing that this particular mutant RNA was able to efficiently displace the bound protein from the labeled RNA-protein complex (Figure 28B). The mean percentage of bound RNA displaced by this mutant was within the range of that for the WT RNA

(RCR001), indicating the presence of other Pr78^{Gag} binding site(s) within this mutant that were not affected by the absence of the ssPurines. Given the absence of a statistically significant reduction in Pr78^{Gag} binding to the FN26 mutant RNA, the next step was to investigate the effect of protein binding to several bpPurine mutants (FN16, FN19, and FN30). Briefly, these clones included deletion of bpPurines (FN16), deletion of the helix/stem of bpPurines only (FN19), and deletion of the complementary sequence of the bpPurines (FN30; Figure 28A). The competitive band shift data for these bpPurine mutants showed a similar pattern of binding as that of FN26 indicating that they competed well against the labeled WT RNA with no statistically significant difference between their percentage of bound RNA (Figure 28B & C). These results reveal that the bpPurines are dispensable for Pr78^{Gag} binding in the presence of ssPurines and implies the presence of other Pr78^{Gag} binding sites on these mutant RNAs. Furthermore, they suggest a possible redundant Gag binding function for the purine-rich regions (ss- and bp-Purines) in MPMV gRNA packaging.

A

Clone	ssPurines & Complementary sequence to bpPurines	Partial repeat of ssPurines (bpPurines) & GU rich region	Relative RNA packaging efficiency (RPE)
RCR001/WT	5' U^{191}UAAAAGU GAAAGUAA^{206} ACUCUC 3'	5' G^{246}AAAGUAA GUGUU^{258} 3'	1.00
FN15	5' Δ ACUCUC 3'	5' Δ GUGUU 3'	0.09
FN16	5' UUAAAAGU GAAAGUAA ACUCUC 3'	5' Δ GUGUU 3'	0.74
FN19	5' UUAAAAGU GAAAGUAA Δ 3'	5' Δ AA GUGUU 3'	ND
FN26	5' Δ ACUCUC 3'	5' GAAAGUAA GUGUU 3'	0.81
FN30	5' UUAAAAGU GAAAGUAA Δ 3'	5' GAAAGUAA GUGUU 3'	0.87
FN42	MPMV Spliced *env* RNA		

Pal SL

ssPurines

GU rich region

bpPurines

B

RCR001/WT FN42 FN26 FN30

%Bound RNA vs Competitor RNA (nM)

C

RCR001/WT FN42 FN15 FN16 FN19

%Bound RNA vs Competitor RNA (nM)
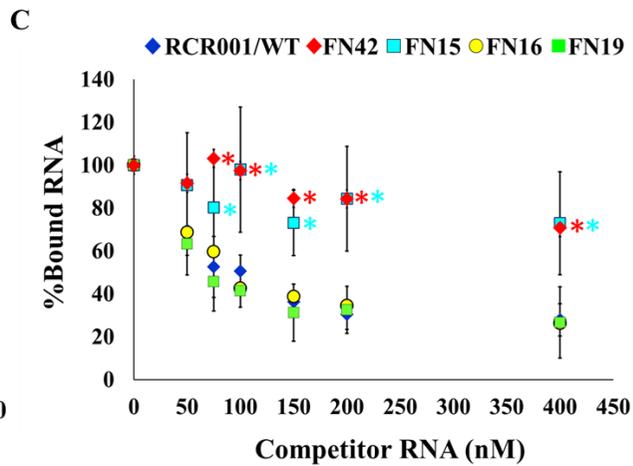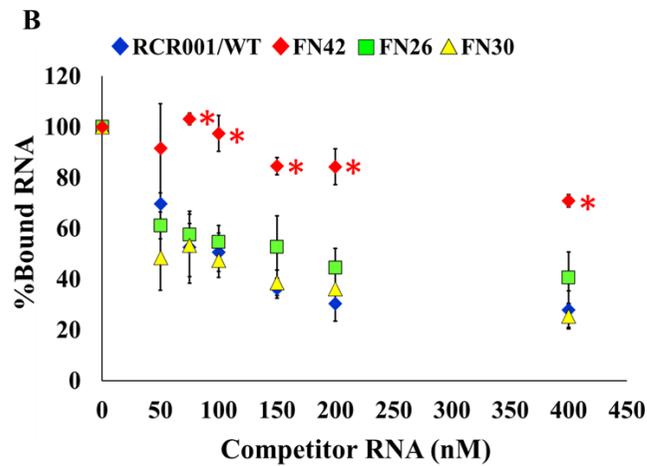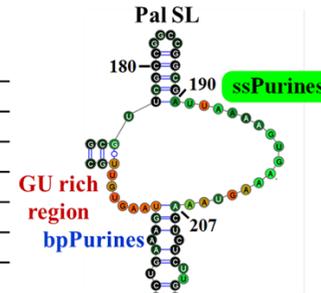
Figure 28: Single-stranded and base-paired purines (ss- and bpPurines) mutants of the MPMV packaging signal RNA bind differentially to Pr78^{Gag}.

(A) Table with the list of molecular clones containing mutations in the ss- and bp-Purines used in band-shift competition assays. ¥The relative packaging efficiencies (RPEs) of the respective molecular clones were previously determined using an in vivo packaging assay (Ali et al., 2020). ND; Not determined. (B and C) Quantitation of the percentage of radiolabeled bound RNA in the Gag-RNA complexes *versus* the concentration of the respective competitor RNA. 500 nM of MPMV Pr78^{Gag} and radiolabeled unspliced wild type (WT; RCR001) RNA were incubated with the respective unlabeled competitor mutant RNAs in increasing concentrations (0-400 nM). The red (FN42) and turquoise blue (FN15) asterisks (*) indicate the statistically significant differences between respective RNAs and WT (RCR001) RNA at different data points in the quantitated gels, calculated using a paired two tailed *t*-test.

Having observed no significant effect on the *in vitro* binding of Pr78^{Gag} to these individual ss- and bp-Purines mutants, as a next step the combined effect of the deletion of both these purine-rich regions on *in vitro* Pr78^{Gag} binding was investigated. As expected for a redundant role of ssPurines and bpPurines, the competitive band shift data for FN15 (containing deletion of both the ssPurines and bpPurines; Figure 28A) revealed poor competition for Pr78^{Gag} binding against the labeled WT RNA with a statistically significant difference between the levels of their bound RNA (Figure 28B).

The *in vitro* Pr78^{Gag} binding to these gRNA mutants correlated well with their *in vivo* packaging, since simultaneous deletion of ssPurines and bpPurines had a dramatic effect on packaging in contrast with deletion of either of these regions alone, which had minimal effects (Figure 28A; Relative RNA Packaging Efficiency (RPE) data from Ali et al., 2020).

3.4.4 MPMV Pr78^{Gag} Binds to the ssPurines, bpPurines, and a Single-Stranded GU-rich Region Located Immediately Downstream of bpPurines

To observe direct binding of Pr78^{Gag} to MPMV *Psi* RNA, hSHAPE a technique that allows structural investigation at single nucleotide resolution, following structure-dependent modification of the RNA, was combined with footprinting assays. In these experiments, RNA-protein interactions were monitored by modifying the WT RNA with a hSHAPE reagent (BzCN) in the absence and presence of Pr78^{Gag}. Any nucleotide showing attenuated hSHAPE reactivity in the presence of the Pr78^{Gag} suggested that the protein can bind to these regions to form an RNA-protein complex, preventing them from modification by BzCN. Such an approach has worked successfully in the case of HIV-1 (Abd El-Wahab et al., 2014) and MMTV

(Chameettachal et al., 2021) for identifying Gag binding sites on their respective gRNAs.

Thus, the WT MPMV packaging signal RNA (RCR001) was tested in the presence of four molar excess of competitor RNA, which in this case was MPMV spliced *env* RNA (FN42) and then incubated with either 6 µM of Pr78[Gag] or no Pr78[Gag]. The protein-RNA complex or RNA alone was then subjected to chemical modification using BzCN, followed by hSHAPE analysis to obtain reactivity data. Experimental triplicates were used to obtain the mean reactivity data for nucleotides in each case and the dataset without Pr78[Gag] was used to obtain the secondary structure of the packaging signal RNA. Mean reactivity data from triplicate experiments with protein was applied onto the secondary structure of the MPMV packaging signal obtained in the absence of Pr78[Gag]. Changes in reactivity data, with *p* values ≤ 0.05, between the samples treated with and without Pr78[Gag], were considered significant (Appendix C). Any significant attenuation in reactivities of nucleotides were regarded as Pr78[Gag] binding nucleotides/sites.

The secondary structure of the MPMV packaging signal RNA obtained via hSHAPE and footprinting in the absence of protein was identical to the previously published hSHAPE structure (Aktar et al., 2013; Figure 29A). In the presence of Pr78[Gag], a significant attenuation of reactivities was observed in two major regions of the *Psi*, including: (1) the ssPurines (U[191]UAAAAGUGAAAGUAA[206]), (2) nucleotides 252 to 257, which correspond to the unpaired A252 and A253 nucleotides at the 3' end of the bpPurines, and a GU-rich region (G[254]UGU[257]) immediately downstream of the bpPurines (Figure 29B & 30A).
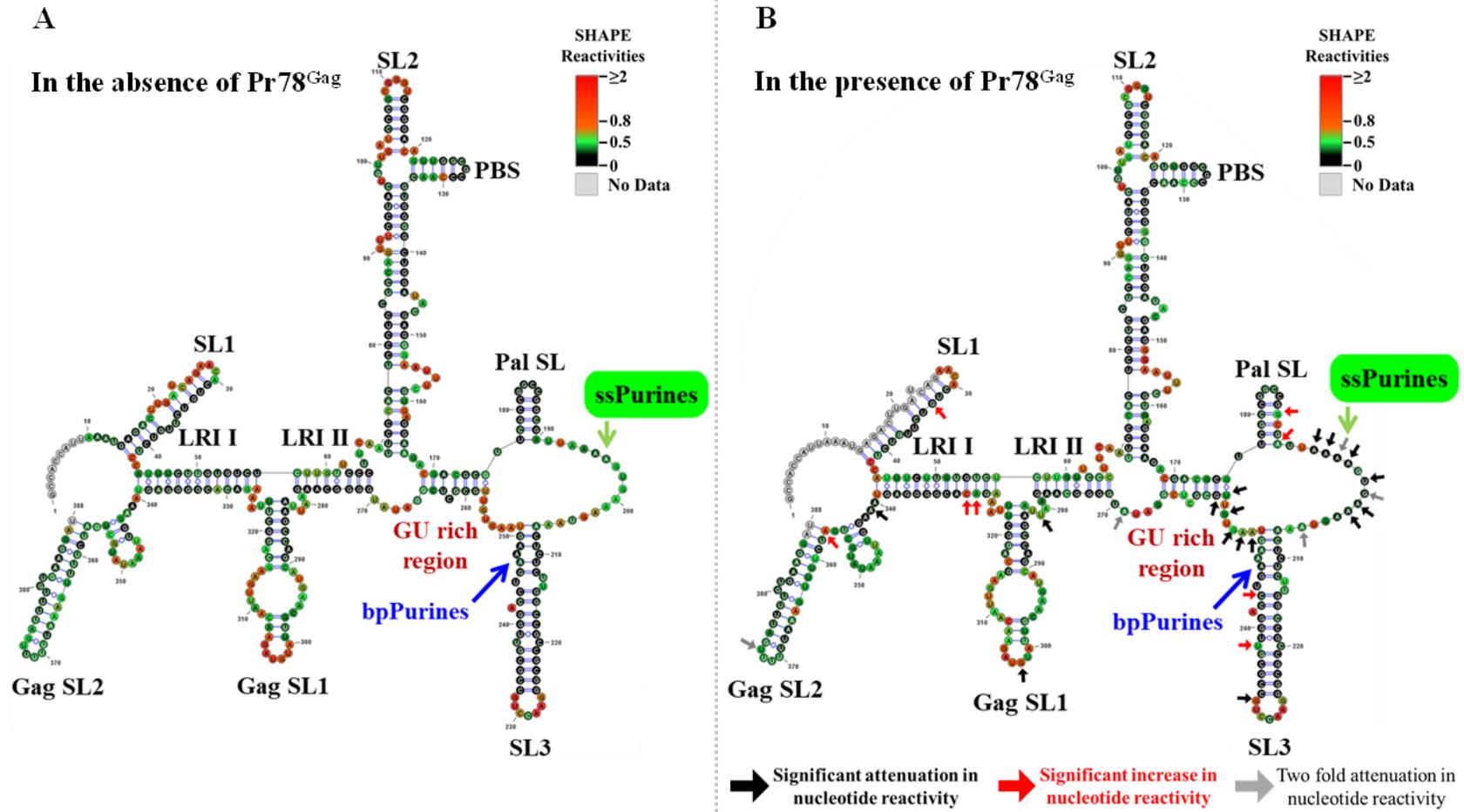
Figure 29: Footprints of Pr78[Gag] on wild type unspliced MPMV packaging signal RNA shows binding to two major regions on the unspliced MPMV packaging signal RNA.

hSHAPE was conducted both in the (A) absence and (B) presence of Pr78^{Gag}. The mean triplicate SHAPE reactivity obtained in the absence of Pr78^{Gag} was used to draw RNA secondary structure of RCR001. This was followed by applying the mean hSHAPE reactivities obtained in presence of Pr78^{Gag} onto the secondary structure obtained in the absence of Pr78^{Gag}. Statistical significance was measured using a paired two tailed t-test and nucleotides with a significant attenuation in SHAPE reactivities ($p \leq 0.05$) are indicated by black arrows. Nucleotides with a significant increase in SHAPE reactivities ($p \leq 0.05$) are indicated by red arrows, and nucleotides with a two-fold attenuation in SHAPE reactivities ($p \leq 0.06\text{-}0.08$) are indicated by grey arrows.

The nucleotides A193, A194, A196, U198, A200 and A201 in the ssPurines showed a significant attenuation in reactivities (with p values ≤ 0.05 and reactivities with a 1.5- to 2-fold reduction or more), indicating Pr78[Gag] binding to these nucleotides (Figure 30B). However, a few other nucleotides (A195, G199 and A205) within this region also showed a 2-fold reduction in reactivity with p values ≤ 0.06-0.08 (Figure 5B). These results suggest significant binding of Pr78[Gag] to nine nucleotides (A193, A194, A195, A196, U198, G199, A200, A201, and A205) out of the 16 nucleotides of the ssPurines with 3 out of the 9 nucleotides (A195, G199 and A205) being less significant yet having a two-fold reduction in reactivity (Figure 30A & B).

In the case of bpPurines, only 2 out of 8 of the nucleotides indicated significant attenuation in reactivities and hence potential Gag binding. These two nucleotides are the unpaired adenosine nucleotides (A252 and A253) located at the 3' end of the bpPurines (Figure 30A & B). Quite unexpectedly, this Pr78[Gag] footprinting extended into an immediately adjacent cluster of guanosine and uracil nucleotides and hence is referred to here as the "GU-rich region". This GU-rich region is also located within the region determined earlier to be indispensable for in $vivo$ MPMV RNA packaging (Jaballah et al., 2010; Schmidt et al., 2003). It comprises of the sequence $G^{254}UGUU^{258}$ and along with the two 3' nucleotides of the bpPurines (A252 and A253) is part of a continuous single-stranded loop $A^{252}AGUGUU^{258}$ (Figure 30A). Footprinting results presented here shows that three nucleotides (G254, G256, and U258) out of these five nucleotides ($G^{254}UGUU^{258}$) showed a significant attenuation in hSHAPE reactivity and hence binding to Pr78[Gag] (Figure 30A & B). In addition, the nucleotide G259 located immediately after the GU-rich region showed a significant attenuation in hSHAPE reactivity. However, since this nucleotide is involved in a G-C base pair, it

is unclear as to whether this is an attenuation due to Pr78^{Gag} binding or due to conformational changes in the structure of the RNA as a result of Pr78^{Gag} binding.
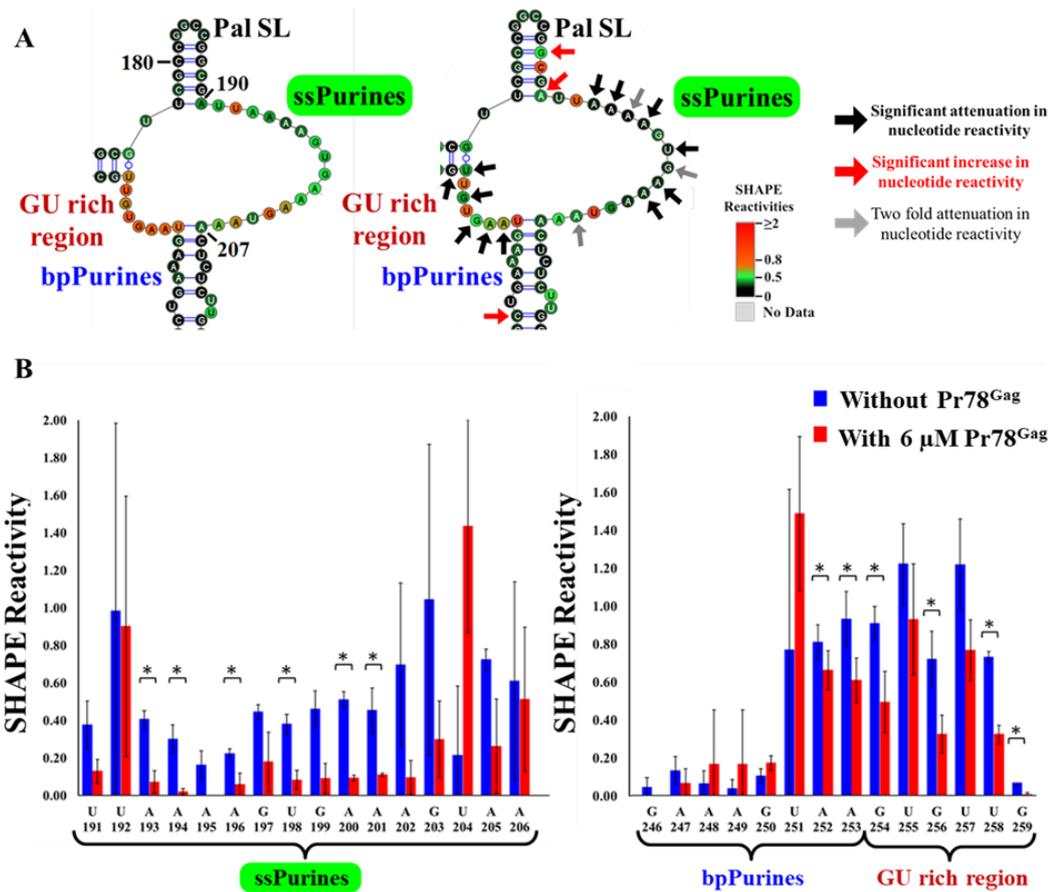
Figure 30: MPMV Pr78^{Gag} binds to two major regions on the unspliced MPMV packaging signal RNA.

(A) hSHAPE conducted both in the absence (left) and presence of Pr78^{Gag} (right) showing footprints in two major regions: the ss-Purines loop and a second loop corresponding to the last two bpPurines and the adjacent GU-rich region in the wild type MPMV RNA (RCR001). Statistical significance was measured using a paired two tailed t-test and nucleotides with a significant attenuation in reactivities ($p \leq 0.05$) are indicated by black arrows. Nucleotides with a significant increase in SHAPE reactivities ($p \leq 0.05$) are indicated by red arrows, and nucleotides with a two-fold attenuation in SHAPE reactivities ($p \leq 0.06\text{-}0.08$) are indicated by grey arrows. (B) Histograms showing the Pr78^{Gag}-induced attenuation of SHAPE reactivities of nucleotides in the ss-, bp-Purines, and the GU-rich region.

3.4.5 Pr78^{Gag} Footprints in the Absence of the ssPurines indicate Binding to the bpPurines and the Single-Stranded GU-rich Region

Given the elaborate binding pattern of Pr78^{Gag} observed at the ssPurines region, we wanted to determine the binding pattern of Pr78^{Gag} to the packaging signal RNA in the absence of the ssPurines. Hence, footprinting experiments were performed on the mutant RNA FN26 which harbors a complete 16-nucleotide deletion of the ssPurines. The secondary structure of the FN26 mutant RNA was determined by hSHAPE in the absence of protein and the resultant structure was observed to be identical to the recently published structure (Figure 31A; Ali et al., 2020). The ssPurine deletion had only local structural effects while the bpPurines retained their partially base-paired nature (Figure 31A). Footprinting experiments were then performed, as mentioned earlier, and hSHAPE reactivity data obtained in the presence of the protein was applied onto the secondary structure of the mutant RNA obtained without protein. Changes in reactivity data, with p values ≤ 0.05, between the RNA treated with and without protein, were considered significant (Appendix D).
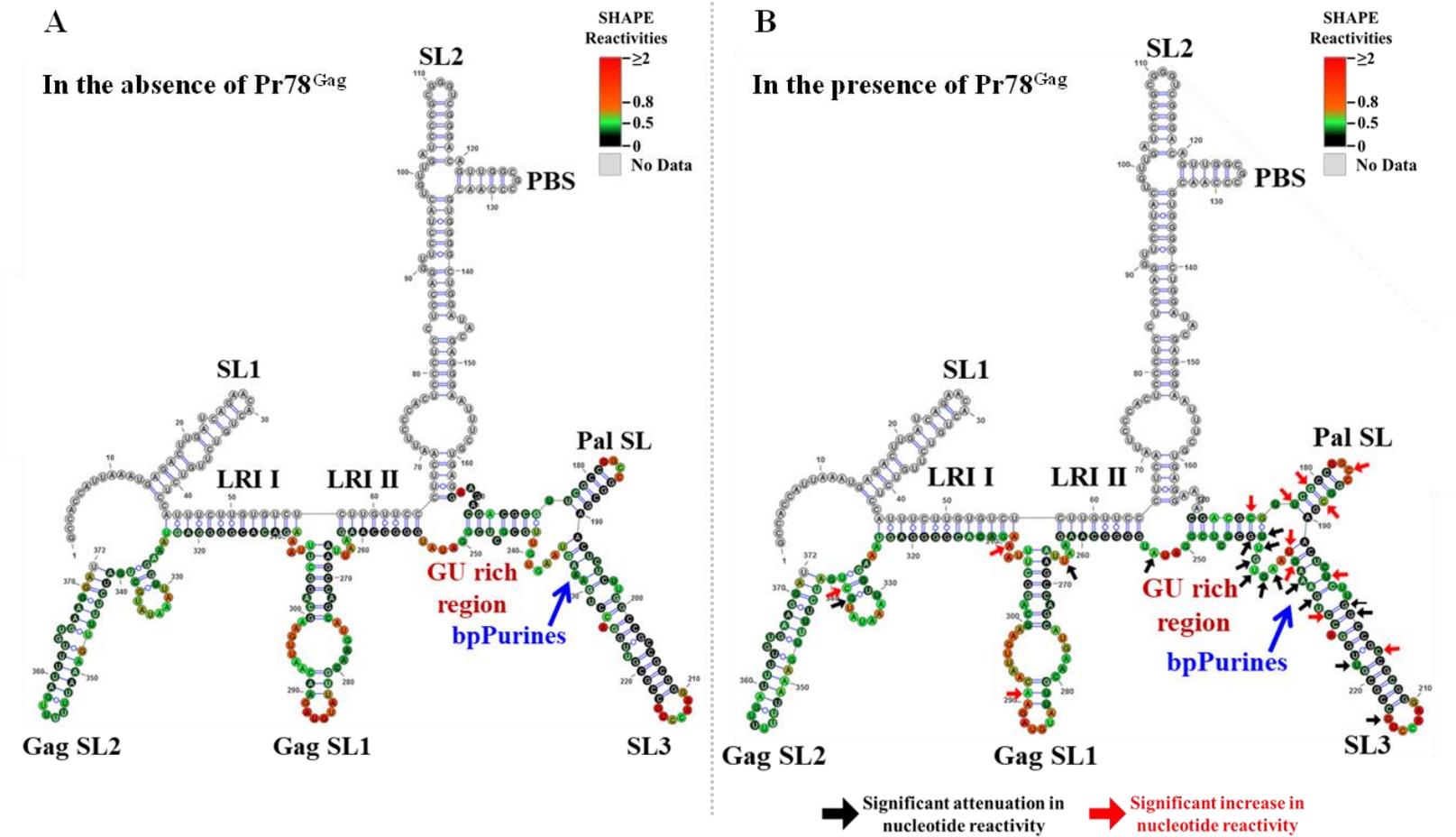
Figure 31: Pr78[Gag] footprints on mutant FN26 RNA containing ssPurines deletion.

(A) Mean triplicate SHAPE reactivity obtained in the absence of Pr78^{Gag} was used to draw RNA secondary structure of FN26. (B) Mean SHAPE reactivities obtained in presence of 6 µM of Pr78^{Gag} were applied onto the secondary structure obtained in the absence of Pr78^{Gag}. Statistical significance was measured using a paired two tailed t-test and nucleotides with a significant attenuation in SHAPE reactivities ($p \leq 0.05$) are indicated by black arrows. Nucleotides with a significant increase in SHAPE reactivities ($p \leq 0.05$) are indicated by red arrows.

Footprinting results obtained for the ssPurine mutant FN26 revealed significant attenuation in hSHAPE reactivities in a few of the nucleotides of bpPurines and the GU-rich region, including nucleotides G230, A232 and A237 of the bpPurines, suggesting Pr78^{Gag} binding to this region (note that the nucleotide numbers of this region differ due to the deletion in the mutant; Figure 32A & B). As anticipated from the hSHAPE-validated structure of this mutant RNA, a majority of the base-paired nucleotides in this region did not show a significant reduction in reactivity due to conformational constraints. However, two other nucleotides (G234 and U235) in this region revealed a significantly higher hSHAPE reactivity in the presence of Pr78^{Gag}, presumably due to conformational changes in the structure of this region due to Pr78^{Gag} binding (Figure 31B & 32A). Surprisingly, here too significant reduction in reactivities were observed in the nucleotides pertaining to the GU-rich region (G^{238}UGUU^{242}; Figure 31 & 32A). All the nucleotides in this region, from G238 to U242, indicated a significant reduction in hSHAPE reactivities; hence, Pr78^{Gag} binding (Figure 32A & B). Nucleotide G243 located immediately after the GU-rich region also indicated a significant decrease in its reactivity. The significant increase in hSHAPE reactivity of the nucleotide C174 that is seen base-paired to the nucleotide G243 in the structure indicates unpairing of this base paired nucleotide upon Pr78^{Gag} binding (Figure 31B & 32A).
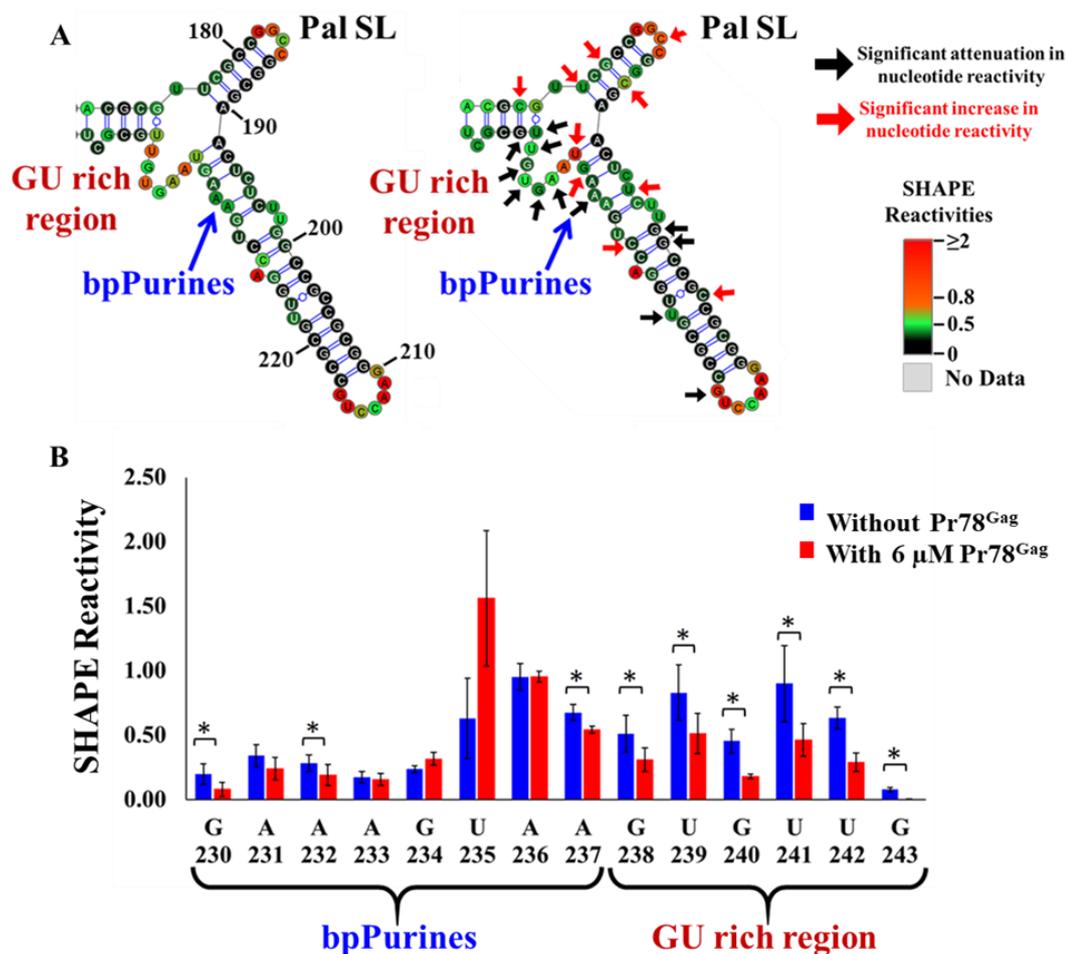
Figure 32: MPMV Pr78[Gag] binds to the last nucleotide of the bpPurines and the adjacent GU-rich region of FN26 mutant RNA in the absence of ssPurines.

(A) hSHAPE conducted both in the absence (left) as well as presence of Pr78[Gag] (right) shows footprints on the last nucleotide of the bpPurines and the adjacent GU-rich regionin FN26. Statistical significance was measured using a paired two tailed t-test and nucleotides with a significant attenuation in SHAPE reactivities ($p \leq 0.05$) are indicated by black arrows. Nucleotides with a significant increase in SHAPE reactivities ($p \leq 0.05$) are indicated by red arrows. (B) Histograms showing the Pr78[Gag]-induced attenuation of hSHAPE reactivities of nucleotides in the bpPurines and the GU rich region. Please note that the nucleotide numbers of the bpPurines and GU-rich region on FN26 RNA are different from that of the wild type (RCR001) RNA due to the 16-nt deletion of ssPurines. Statistical significance was measured using a paired two tailed t-test and nucleotides with a significant attenuation in SHAPE reactivities ($p \leq 0.05$) are indicated by black arrows. Nucleotides with a significant increase in SHAPE reactivities ($p \leq 0.05$) are indicated by red arrows.

3.4.6 The ssPurines is Partially Base Paired in the Spliced *env* RNA

Having established that the ssPurines form one of the significant binding sites of Pr78^{Gag} on the MPMV *Psi* RNA, and because it is located upstream of the mSD, and hence present on both the unspliced and spliced RNAs, we interrogated the structure of the spliced *env* RNA (FN42) to determine whether the ssPurines adopt an alternative confirmation compared to the unspliced, full length *Psi* RNA (RCR001) using hSHAPE. This analysis revealed that nucleotides A^{196}GUGA^{200}, forming the ssPurines loop in the unspliced RNA SHAPE validated structure (Figure 29A & 30A), are involved in partial base pairing with nucleotides U^{209}CUCU^{213}, downstream of the ssPurines (Figure 33). Interestingly, four of the five nucleotides forming this partially base paired conformation in the spliced RNA, were found to be involved in significant Pr78^{Gag} binding in the unspliced RNA, as shown by footprinting (Figure 29B & 30A & B).
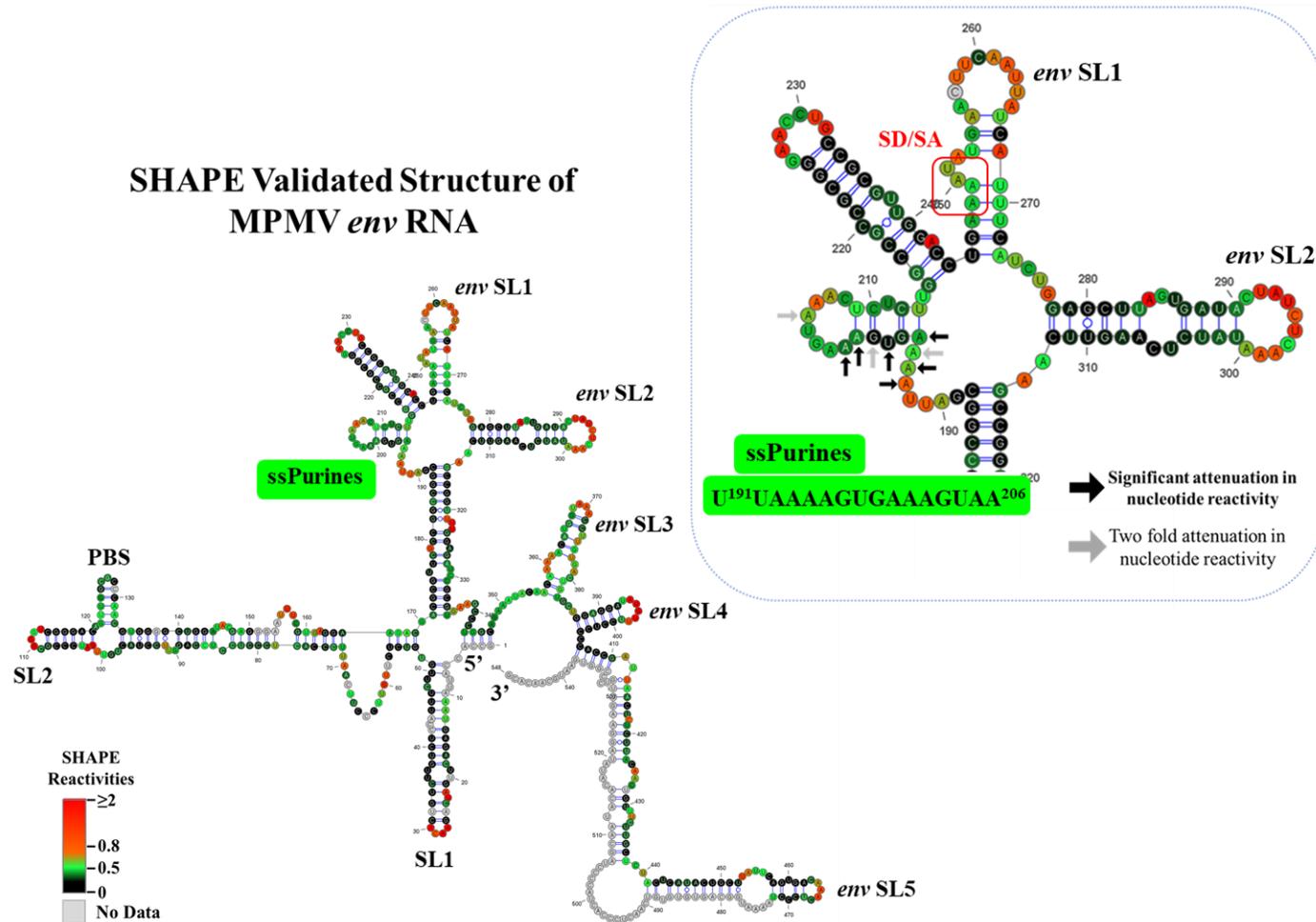
Figure 33: SHAPE validated secondary structure of the MPMV spliced *env* RNA (FN42).

The ssPurines attain a partially base paired conformation in the secondary structure of the spliced *env* RNA. Nucleotides showing footprints on the wild type unspliced RNA (RCR001) have been superimposed on the SHAPE validated spliced *env* RNA indicating that some of these nucleotides are involved in a partially base paired conformation. The splice donor-splice acceptor (SD/SA) junction is demarcated by a red box. Significant attenuation ($p \leq 0.05$) in hSHAPE reactivities of these nucleotides are indicated by black arrows and nucleotides with two-fold attenuation in reactivities ($p \leq 0.06$-0.08) are indicated by grey arrows.

3.5 Discussion

This study aimed to determine the Gag-binding sites on MPMV gRNA important for the specific incorporation of its genome into the assembling virion. Packaging signals for MPMV has been mapped to the 5 'UTR and the first 120 nts of *gag*, a region that assumes a complex secondary RNA structure (Figure 25B; Aktar et al., 2013; Schmidt et al., 2003). Interestingly, while most of the structural motifs of MPMV *Psi* RNA contribute to its structural stability, two purine-rich regions, (1) ssPurines, located immediately downstream of the Pal SL (DIS), and (2) the bpPurines, have been proposed to function redundantly as packaging signals (Ali et al., 2020; Jaballah et al., 2010). Thus, using purified full-length MPMV Gag protein in footprinting assays, results presented here provide direct evidence for $Pr78^{Gag}$ binding to two sites within the MPMV *Psi* RNA: 1) the ss-Purine loop ($U^{191}UAAAAGUGAAAGUAA^{206}$), and 2) a second loop ($A^{252}AGUGUU^{258}$) corresponding to the last two nucleotides of the bpPurines region and extending into the immediately into the adjacent GU-rich region (Figure 34). These regions essentially identify as the minimal *cis* acting sequences required for efficient Gag binding and thereby gRNA packaging in MPMV.

Identifying these crucial Gag binding elements in MPMV gRNA packaging should help in the design of MPMV-based vectors for human gene therapy by allowing only the minimal *cis* acting sequences to be present on the vector, eliminating all unwanted and potentially harmful sequences to make their design safer. Such a design of a gene therapy vector is likely to minimize chances of recombination with endogenous and/or exogenous retroviruses. Additionally, as MPMV assembles in the cytoplasm, *in vitro* studies of its Gag-gRNA interaction are likely to have direct

implications for the selection and packaging of the gRNA. Interestingly, some Gag mutations in MPMV can divert its assembly from the cytoplasm to the plasma membrane (Rhee & Hunter, 1990). Therefore, it would be interesting to determine if re-directing MPMV assembly to the plasma membrane affects Gag-gRNA interactions and thereby RNA packaging.

Competitive band shift assays carried out on ss- and bp-Purines mutants provides a mechanistic rationale for the previously published *in vivo* packaging data for the mutants harboring the same mutations in these regions. Despite the deletion of either of these regions, a majority of the mutants (FN16, 19, 26, and 30; Figure 28A) were capable of competing well against the labeled WT unspliced gRNA, revealing the presence of other Pr78^{Gag} binding sites on these mutant RNAs. As expected, it was only the simultaneous deletion of both the ss- and bp-Purines in mutant FN15 that led to a drastic effect on Pr78^{Gag} binding (Figure 28B). Since both these purine-rich regions contribute to the structural stability of the gRNA, their simultaneous deletion in FN15 resulted in the overall loss of the RNA structure (Ali et al., 2020). This perhaps is responsible for the poor Pr78^{Gag} binding observed for this mutant. The redundant nature of Gag binding to purine residues have been implicated in other retroviruses. For example, in the case of HIV-1 it has been reported that multiple G residues are important during *in vivo* RNA packaging (Nikolaitchik et al., 2020). When these G residues were substituted individually, they did not show a significant change in gRNA packaging, while substitution of these nucleotides collectively resulted in a drastic reduction in packaging, supporting the redundant role of these G residues during Gag binding and gRNA packaging, even though the structure of these HIV-1 mutants has not been studied (Nikolaitchik et al., 2020).
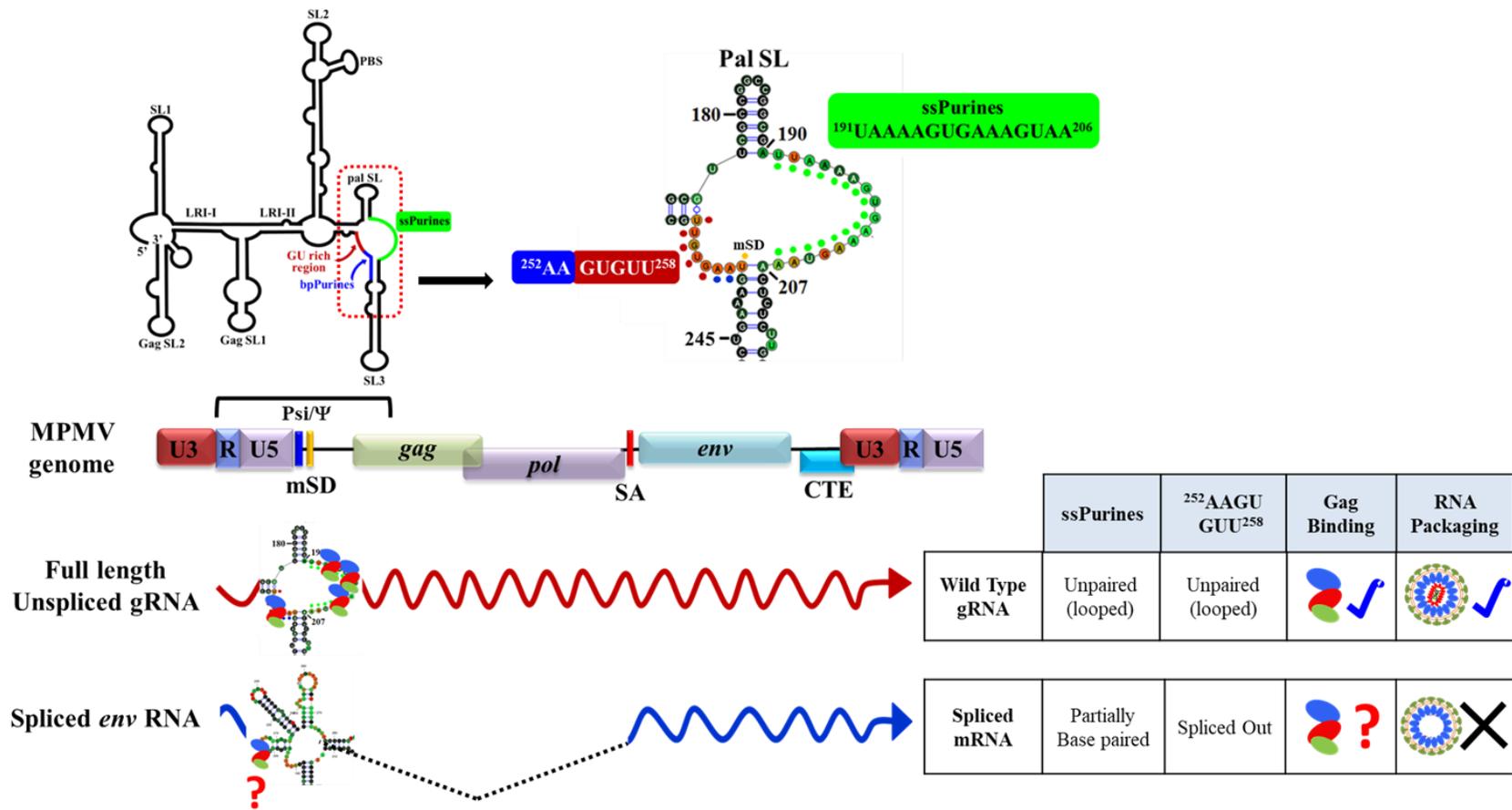
Figure 34: Model proposing how MPMV Pr78^{Gag} captures *Psi* RNA from a mix of cellular and spliced viral RNAs.

Full-length, unspliced transcripts of the MPMV genome serve as the substrate for packageable gRNA into nascent viral particles. The packaging determinants (*Psi*/Ψ) of MPMV gRNA spans from R to the first 120 bp of gag that folds into a unique secondary structure. Pr78[Gag] binds to two single-stranded regions within the MPMV *Psi* RNA: (1) the ssPurines loop (U^{191}UAAAAGUGAAAGUAA^{206}), located upstream the mSD and hence found in all viral RNAs, and (2) the single-stranded A^{252}AGUGUU^{258} loop, which corresponds to the last two purines of the bpPurines and extends into a GU-rich region. The latter is located downstream of the mSD; thus, it is present only in the genomic and not the spliced RNA. Both these regions act redundantly during the gRNA packaging process. We propose that the MPMV Pr78[Gag] may first bind to the ssPurines loop to distinguish viral RNAs from cellular mRNAs. This is probably followed by binding to the A^{252}AGUGUU^{258} loop, multimerizing Gag binding onto the gRNA in a synergistic manner to further enhance the ability to specifically capture the unspliced RNA, containing both single-stranded purine loops, over the spliced RNAs, which lacks the single-stranded A^{252}AGUGUU^{258} loop. This ensures that the MPMV gRNA is packaged preferentially into the viral particles, excluding cellular and spliced RNAs.

Interestingly, the MPMV Pr78^{Gag} footprint on ssPurines revealed a significant attenuation of hSHAPE reactivities for nine out of the 16 nucleotides (A193, A194, A195, A196, U198, G199, A200, A201 and A205) comprising the loop with more than a two-fold reduction in reactivity, out of which six of them had a p-value of ≤ 0.05 and three of them had a p-values ≤ 0.06-0.08 (Figure 29B & 30A). Interestingly, among the two regions showing Pr78^{Gag} footprints, ssPurines contained the longer Pr78^{Gag} footprint, which is consistent with the large purine-rich loop in the *Psi* RNA structure. Notably these Pr78^{Gag} binding nucleotides are found clustered at the apical end of the loop which, in terms of spatial arrangement, may be more accessible for protein binding, especially in the context of the Gag polyprotein which is considerably bulkier than its smaller NC counterpart.

As expected, compared to the ssPurines, hSHAPE on MPMV *Psi* RNA in the presence of Pr78^{Gag} did not show significant changes in the bpPurines because of their base-paired confirmation (G^{246}AAAGU^{251}; Figure 29 & 30). However, the two unpaired A nucleotides (A252 and A253) at the 3' end of the bpPurines showed a significant attenuation in SHAPE reactivities, confirming specific Pr78^{Gag} binding (Figure 31B & 30A). The GU-rich region (G^{254}UGUU^{258}; Figure 29 & 30), on the other hand, demonstrated significant attenuation of SHAPE reactivities at three of the five nucleotides: G254, G256, and U258, where U258 is involved in a wobble base pair, indicating a certain amount of flexibility in this region accounting for Pr78^{Gag} binding (Figure 29B & 30A). Additionally, Pr78^{Gag} binding was also observed at nucleotides G259 with a concomitant increase in reactivity of its complementary base pair C174, indicating unpairing of this base pair upon Pr78^{Gag} binding (Figure 29B & 30A). Considering that A252 and A253 of the bpPurines form a continuous single-stranded loop of seven nucleotides along with the GU-rich region (A^{252}AGUGUU^{258}),

five of these nucleotides showed Pr78^{Gag} binding (A252, A253, G254, G256 and U258). Thus, the overall binding of Pr78^{Gag} to the MPMV *Psi* RNA can be categorized into two important regions: (1) the ssPurines and (2) the single-stranded A^{252}AGUGUU^{258} loop that may act redundantly in bringing about gRNA packaging (Figure 34).

In the absence of the ssPurines (FN26 RNA), Pr78^{Gag} binding could be mapped to: (1) bpPurines and (2) the single-stranded GU-rich region (G^{238}UGUU^{242}; Figure 31 & 32). Analysis of the footprinting data in the absence of ssPurines revealed a number of nucleotides with a significant increase in SHAPE reactivity upon Pr78^{Gag} binding. These nucleotides (C174, U177, G179, C188, U195, G234 and U235) were observed in and around the three-way junction formed by the Pal SL, SL3 and the GU-rich region (Figure 31B & 32). This significant increase in reactivity, particularly in base paired-nucleotides, implies a conformational change of this region/nucleotides upon Pr78^{Gag} binding in the absence of ssPurines. Comparatively, footprinting on the WT RNA, i.e., in the presence of the ssPurines, did not reveal many nucleotides with significant increases in reactivities to suggest distinct conformational changes to the RNA structure, indicating that the WT gRNA, due to the presence of the ssPurines, may not have to undergo as much conformational changes upon Pr78^{Gag} binding to augment gRNA packaging (Figure 29 & 30), also implying that ssPurines may act as the dominant or primary Gag binding site.

A crucial observation from this study is that Pr78^{Gag} is capable of efficiently discriminating between the full-length unspliced gRNA from the spliced *env* RNA, revealing the presence of high affinity binding site(s) in the unspliced gRNA that are recognized early during the selection process (Figure 26D). This corroborates well with studies carried out on other retroviruses, including HIV-1, RSV and MMTV and

explains the selective packaging of unspliced gRNAs over spliced viral RNAs by their respective Gag proteins (Abd El-Wahab et al., 2014; Aronoff and Linial, 1991; Aronoff et al., 1993; Banks et al., 1999; Webb et al., 2013; Chameettachal et al., 2021). Interestingly, the mSD in MPMV gRNA is located within one of these Pr78^{Gag} binding sites, the bpPurines, (G^{246}AAA**GU**AA^{253}; Figure 25B), thus making it part of the unspliced gRNA (Figure 30 & 32). This is unlike the ssPurines (another Pr78^{Gag} binding site), which is located upstream of the mSD and thus, present on both the unspliced full-length and spliced *env* RNA (Figure 29B & 30A). Footprinting on the WT gRNA structure also indicates binding of the polyprotein to two unpaired adenosines (A252 and A253) at the 3' end of the bpPurines and the GU-rich region, both of which are located downstream of the mSD and hence unique to the unspliced gRNA (Figure 29B & 30A). Notably, these sequences form a continuous single-stranded loop of six single-stranded nucleotides (A^{252}AGUGU^{257}; Figure 29 & 30). The presence of more than one Pr78^{Gag}-specific binding site, positioned upstream as well as downstream of the mSD on MPMV gRNA indicates that these regions may act either cooperatively, or as previously suggested, in a redundant manner, in bringing about MPMV gRNA packaging. In the former case, if these regions were to act synergistically, the absence of the two crucial Pr78^{Gag}-specific binding regions downstream of the mSD would explain the selective packaging of the unspliced gRNA over the spliced *env* RNA. Such a requirement of more than one high affinity binding site for selective gRNA packaging is not surprising and has been demonstrated not only in HIV-1 but also in unrelated viruses such as the MS2 phage (Ding et al., 2020; Nikolaitchik et al., 2020; Stockley et al., 2016). Analysis of the Pr78^{Gag} binding data with the Hill equation indicates that its binding is weakly or not cooperative (Figure 26C). This is in contrast with the situation prevailing in MMTV, since similar

experiments showed a high degree of cooperativity of MMTV $Pr77^{Gag}$ to its cognate gRNA (Hill coefficient = 3; Chameettachal et al., 2021). In HIV-1, while binding of GagΔp6 to gRNA is cooperative (Comas-Garcia et al., 2017), binding of the full-length $Pr55^{Gag}$ is not cooperative, but there are two classes of $Pr55^{Gag}$ binding sites with different binding affinities (Bernacchi et al., 2017). Irrespective, this data suggests that the ssPurines and the single-stranded GU ($A^{252}AGUGU^{257}$) region function redundantly, leading to the provocative question as to how the presence of a $Pr78^{Gag}$ binding sites upstream of the mSD in the spliced *env* RNA overcome selective binding to the precursor Gag protein.

The answer probably lies in the effect of splicing on the structure of the Gag-binding sites. The presence of specific Gag binding sites upstream of the mSD has been observed for a number of retroviruses, including HIV-1 (Abd El-Wahab et al., 2014) and MMTV (Chameettachal et al., 2021). In the case of HIV-1, it has been suggested that the specific binding site for its precursor polyprotein ($Pr55^{Gag}$) is located on a purine rich G//AGG motif on the internal loop of SL1; hence, it is present on both the unspliced and spliced viral transcripts (Abd El-Wahab et al., 2014). However, there exists a higher level of selection that is positively controlled by a region downstream of the mSD, likely via a long-range interaction involving sequences both upstream and downstream of the mSD (Abd El-Wahab et al., 2014; Bernacchi et al., 2017). It is proposed that such a mechanism may also come into play for other retroviruses like MMTV, where the putative Gag binding sites present on the bifurcated SL4, in the form of single stranded purines (ssPurines), are located upstream of the mSD (Aktar et al., 2014; Mustafa et al., 2018; Chameettachal et al., 2021) Similar to HIV-1 and MMTV, MPMV too relies on the U5-Gag LRIs to maintain the overall structure of the *Psi* region responsible for initiating and augmenting gRNA packaging (Aktar et al.,

2014, 2013; Kalloush et al., 2016; Chameettachal et al., 2021). In fact, the MPMV gRNA harbors two such U5-Gag LRIs, LRI-I and LRI-II (Figure 25B). The loss of these LRIs during splicing results in disruption of the native structure of the *Psi* RNA (Kalloush et al., 2016), including the native ssPurines conformation, preventing Pr78^{Gag} binding to it. As is evident from the footprinting data on the WT gRNA, Pr78^{Gag} binding to the ssPurines is limited to the nucleotides at the apical end of the loop (Figure 30B), suggesting the requirement of a spatially-accessible loop structure. Secondary structural analysis of the spliced *env* RNA, performed using hSHAPE, indicates that this spatially-accessible loop is lost and instead attains a partially base-paired conformation in which nucleotides in the ssPurines loop are base paired (Figure 6). This is a distinct alteration from its native single-stranded form which would otherwise be completely accessible for Pr78^{Gag} binding. Such a conformational change results in unpairing of 6 of the 9 nucleotides of the ssPurines which showed Gag footprints (Figure 33). This suggests partial accessibility of Pr78^{Gag} to the spliced *env* RNA, which may explain the 30% bound RNA being observed upon competitive band shift assays performed for *env* spliced RNA (Figure 26D). Furthermore, since spliced *env* RNA is not packaged, these results suggest that potential limited Gag binding to spliced *env* RNA may allow its initial capture from the distinctly non-specific cellular RNAs (Figure 34). However, such inadequate binding may not be enough for Gag multimerization on the spliced *env* RNA necessary for encapsidation. On the other hand, the full length unspliced RNA, having an additional Gag binding GU-rich region (G^{254}UGUU^{258}) downstream of the mSD (Figures 29 & 30), may facilitate Gag multimerization onto the unspliced RNA, resulting in its preferential packaging over the spliced RNA (Figure 34). Thus, the overall conformational change of ssPurines in the spliced *env* RNA, the restriction exerted by the partially base paired nucleotides

for Pr78^{Gag} binding, as well as the absence of the GU-rich region following splicing may explain the preferential selection of unspliced *Psi* RNAs over spliced *env* RNA in MPMV gRNA packaging.

In conclusion, this study demonstrates that Pr78^{Gag} binds to two single-stranded loops positioned oppositely on the unspliced MPMV *Psi* RNA consisting of the ssPurines, and the $A^{252}AGUGUU^{258}$ loop which includes two nucleotides of the bpPurines and an adjacent GU-rich motif ($G^{254}UGUU^{258}$; Figure 34). Based on structural differences, findings here also demonstrate that Pr78^{Gag} can effectively discriminate between unspliced MPMV *Psi* RNA from spliced *env* RNA, revealing how MPMV differentiates between the two RNA substrates. Thus, identification of structural elements crucial in MPMV gRNA packaging should help in understanding not only the mechanism of virion assembly by retroviruses, but also facilitate development of safe and efficient retroviral vectors for human gene therapy.

3.6 Funding

Chapter 4: Conclusions & Future Directions

4.1 Conclusions

The main goal of this dissertation was to identify the binding sites of the full-length recombinant Pr78^{Gag} on MPMV gRNA to better understand the mechanism underlying specific selection of gRNA during the packaging process. Owing to the unavailability of recombinant MPMV Pr78^{Gag}, the first part of this study was directed at the successful expression and purification of biologically active Gag precursor polyprotein that could be used for the *in vitro* biochemical experiments (Chapter 2). A recombinant Pr78^{Gag}-His$_6$-tagged protein was expressed in bacteria and purified to homogeneity using IMAC followed by SEC. The Pr78^{Gag}-His$_6$-tagged protein could assemble and form VLPs both *in vitro* as well as in bacterial & eukaryotic cells. No morphological difference could be observed in the VLPs formed in bacteria by Pr78^{Gag} in the presence or absence of the His$_6$-tag, and functionally both proteins could package MPMV gRNA efficiently when expressed in eukaryotic cells, further suggesting that the presence of a His$_6$-tag did not affect VLP assembly or gRNA packaging. In summary, the expressed and purified recombinant Pr78^{Gag}-His$_6$-tag was deemed suitable for use in subsequent *in vitro* biochemical experiments (Pitchai et al., 2018).

In the second part of this study (Chapter 3), a number of *in vitro* biochemical approaches were employed in order to identify essential Pr78^{Gag} binding motifs on the MPMV gRNA. Firstly, competitive band shift assays were performed that revealed that recombinant Pr78^{Gag}-His$_6$-tag was capable of efficiently discriminating between unspliced full-length gRNA and its spliced *env* RNA, a phenomenon that is crucial in the native scenario in bringing about the selective or specific packaging of gRNA into

assembling virus particles. Secondly, competitive band shift assays were performed to identify the binding of Pr78^{Gag} on two purine rich motifs (ssPurines and the bpPurines) on the MPMV gRNA identified earlier to be important for MPMV RNA packaging. Results of such assays carried out using various deletion mutants of the purine-rich motifs revealed that both these regions bind to Pr78^{Gag} in a redundant manner. These observations confirmed the recently published observations that these purine-rich regions are critical for efficient MPMV gRNA packaging and propagation in an *in vivo* biological assay (Ali et al., 2020). Hence, *in vitro* results presented in this study and the published *in vivo* packaging data using the same mutants further validates the redundant role of ssPurines and bpPurines in Pr78^{Gag} binding, augmenting gRNA packaging. Finally, footprinting experiments followed by hSHAPE mapped the Pr78^{Gag} binding to two single-stranded loops: 1) the ssPurines, and 2) A^{252}AGUGUU^{258} loop corresponding to the two 3' unpaired adenosine nucleotides of the bpPurines and the GU-rich region. Additionally, footprinting experiments carried out on FN26, a mutant containing deletion of the major footprinting region the ssPurines, revealed Pr78^{Gag} binding to the nucleotides of the bpPurines and the GU-rich region, reiterating the redundant binding capability of Pr78^{Gag} to this region in the absence of the ssPurines. Notably, the ssPurines are located upstream of the mSD and hence are present on both the unspliced and spliced RNAs and the A^{252}AGUGUU^{258} loop is located downstream of the mSD; hence, present only on the unspliced gRNA.

Based on these findings, this study proposes a model is for the specific selection and packaging of MPMV gRNA over its spliced *env* RNA which proposes that MPMV Pr78^{Gag} first binds to the ssPurines to distinguish viral from cellular mRNAs in the cell *milieu*. This is probably followed by Pr78^{Gag} binding to the A^{252}AGUGUU^{258} region, multimerizing Gag binding onto the gRNA in a synergistic

manner to further enhance the specific capture of the unspliced gRNA that contain both single-stranded loops over the spliced RNAs lacking the $A^{252}AGUGUU^{258}$ loop (Figure 34). Such a scenario ensures that the MPMV gRNA is packaged preferentially into the viral particles, excluding cellular and spliced viral RNAs.

4.2 Future Directions

The current study has examined the expression and purification of a biologically active recombinant full-length MPMV Gag precursor polyprotein and has provided direct evidence of the binding sites of this polyprotein on the MPMV gRNA. While these findings have provided insight into the mechanism that may bring about the selective packaging of MPMV gRNA over spliced RNA during viral assembly, they also open novel areas of future research, some of which are listed below:

1. Given that this study has made available purified full-length recombinant Pr78^{Gag}, one of the interesting research areas that now can be explored are Cryo-EM experiments to determine the 3D structure of the full-length Gag precursor polyprotein which has thus far not been studied. It will also be interesting to perform co-crystallization experiments with full-length Pr78^{Gag} and viral genomic RNA. Having crystal structures of viral RNA-Gag complexes would allow confirmation of proposed MPMV RNA secondary and tertiary structures, as well as proposed Gag-RNA interactions.

2. Having performed *in vitro* footprinting and hSHAPE experiments to determine the binding sites of Pr78^{Gag} on the MPMV gRNA, it will be interesting to determine if these results corroborate in an *in virio* scenario. In this case, MPMV pseudoviral particles can be isolated, purified and treated with Aldrithiol-2 (AT-2) which disrupts NC-RNA interactions. Subsequently, they can be subjected to hSHAPE

where NMIA or IM7 modification are performed *in virio* and then the RNA extracted and subjected to primer extension. Results from such experiments can be compared with *ex virio* control where the pseudoviral particles are non-treated and its RNA is extracted and subjected to hSHAPE in parallel to determine the relevant Gag binding sites on the gRNA. Such studies have been successfully conducted on HIV-1 (Wilkinson et al., 2008) and MLV (Gherghe et al., 2010), and have provided useful insight into the similarities and differences in *in vitro* and *in vivo* hSHAPE structures of their packaging signal RNA and Gag binding sites.

References

Abd El-Wahab, E. W., Smyth, R. P., Mailler, E., Bernacchi, S., Vivet-Boudou, V., Hijnen, M., Jossinet, F., Mak, J., Paillart, J.-C., & Marquet, R. (2014). Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nature Communications*, *5*, 4304. https://doi.org/10.1038/ncomms5304.

Affranchino, J. L., & González, S. A. (2010). In vitro assembly of the feline immunodeficiency virus Gag polyprotein. *Virus Research*, *150*(1), 153–157.

Akhlaq, S., Panicker, N. G., Philip, P. S., Ali, L. M., Dudley, J. P., Rizvi, T. A., & Mustafa, F. (2018). A cis-Acting Element Downstream of the Mouse Mammary Tumor Virus Major Splice Donor Critical for RNA Elongation and Stability. *Journal of Molecular Biology*, *430*(21), 4307–4324.

Aktar, S. J., Jabeen, A., Ali, L. M., Vivet-Boudou, V., Marquet, R., & Rizvi, T. A. (2013). SHAPE analysis of the 5′ end of the Mason-Pfizer monkey virus (MPMV) genomic RNA reveals structural elements required for genome dimerization. *RNA*, *19*(12), 1648–1658.

Aktar, S. J., Vivet-Boudou, V., Ali, L. M., Jabeen, A., Kalloush, R. M., Richer, D., Mustafa, F., Marquet, R., & Rizvi, T. A. (2014). Structural basis of genomic RNA (gRNA) dimerization and packaging determinants of mouse mammary tumor virus (MMTV). *Retrovirology*, *11*, 96. https://doi.org/10.1186/s12977-014-0096-6.

Al Dhaheri, N. S., Phillip, P. S., Ghazawi, A., Ali, J., Beebi, E., Jaballah, S. A., & Rizvi, T. A. (2009). Cross-packaging of genetically distinct mouse and primate retroviral RNAs. *Retrovirology*, *6*, 66. https://doi.org/10.1186/1742-4690-6-66.

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

Aldovini, A., & Young, R. A. (1990). Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *Journal of Virology*, *64*(5), 1920–1926.

Alfadhli, A., & Barklis, E. (2014). The roles of lipids and nucleic acids in HIV-1 assembly. *Frontiers in Microbiology*, *5*, 253. https://doi.org/10.3389/fmicb.2014.00253.

Alfadhli, A., McNett, H., Tsagli, S., Bächinger, H. P., Peyton, D. H., & Barklis, E. (2011). HIV-1 matrix protein binding to RNA. *Journal of Molecular Biology*, *410*(4), 653–666.

Alfadhli, A., Still, A., & Barklis, E. (2009). Analysis of human immunodeficiency virus type 1 matrix binding to membranes and nucleic acids. *Journal of Virology*, *83*(23), 12196–12203.

Ali, L. M., Rizvi, T. A., & Mustafa, F. (2016). Cross- and Co-Packaging of Retroviral RNAs and Their Consequences. *Viruses*, *8*(10), 276. https://doi.org/10.3390/v8100276.

Ali, L. M., Pitchai, F., Vivet-Boudou, V., Chameettachal, A., Jabeen, A., Pillai, V. N., Mustafa, F., Marquet, R., & Rizvi, T. A. (2020). Role of Purine-Rich Regions in Mason-Pfizer Monkey Virus (MPMV) Genomic RNA Packaging and Propagation. Frontiers in microbiology, 11, 595410. https://doi.org/10.3389/fmicb.2020.595410.

Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., & Summers, M. F. (2000). NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *Journal of Molecular Biology*, *301*(2), 491–511.

Aronoff, R., & Linial, M. (1991). Specificity of retroviral RNA packaging. *Journal of Virology*, *65*(1), 71–80.

Aronoff, R., Hajjar, A. M., & Linial, M. L. (1993). Avian retroviral RNA encapsidation: reexamination of functional 5' RNA sequences and the role of nucleocapsid Cys-His motifs. *Journal of Virology*, *67*(1), 178–188.

Bharat, T. A., Davey, N. E., Ulbrich, P., Riches, J. D., de Marco, A., Rumlova, M., Sachse, C., Ruml, T., & Briggs, J. A. (2012). Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy. *Nature*, *487*(7407), 385–389.

Baltimore, D. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature*, *226*(5252), 1209–1211.

Baluyot, M. F., Grosse, S. A., Lyddon, T. D., Janaka, S. K., & Johnson, M. C. (2012). CRM1-Dependent Trafficking of Retroviral Gag Proteins Revisited. *Journal of Virology*, *86*(8), 4696–4700.

Balvay, L., Lopez Lastra, M., Sargueil, B., Darlix, J.-L., & Ohlmann, T. (2007). Translational control of retroviruses. *Nature Reviews. Microbiology*, *5*(2), 128–140.

Banks, J. D., Kealoha, B. O., & Linial, M. L. (1999). An MΨ-Containing Heterologous RNA, but Not env mRNA, Is Efficiently Packaged into Avian Retroviral Particles. *Journal of Virology*, *73*(11), 8926–8933.

Barajas, B. C., Tanaka, M., Robinson, B. A., Phuong, D. J., Chutiraka, K., Reed, J. C., & Lingappa, J. R. (2018). Identifying the assembly intermediate in which Gag first associates with unspliced HIV-1 RNA suggests a novel model for HIV-1 RNA packaging. *PLoS Pathogens*, *14*(4), e1006977. https://doi.org/10.1371/journal.ppat.1006977.

Becker, J. T., & Sherer, N. M. (2017). Subcellular Localization of HIV-1 gag-pol mRNAs Regulates Sites of Virion Assembly. *Journal of Virology*, *91*(6), e02315-16. https://doi.org/10.1128/JVI.02315-16.

Behrens, R. T., Aligeti, M., Pocock, G. M., Higgins, C. A., & Sherer, N. M. (2017). Nuclear Export Signal Masking Regulates HIV-1 Rev Trafficking and Viral RNA Nuclear Export. *Journal of Virology, 91*(3), e02107-16. https://doi.org/10.1128/JVI.02107-16.

Berger, E. A., Murphy, P. M., & Farber, J. M. (1999). Chemokine Receptors as HIV-1 Coreceptors: Roles in Viral Entry, Tropism, and Disease. *Annual Review of Immunology*, *17*(1), 657–700.

Berkhout, B., & Van Wamel, J. L. (1996). Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *Journal of Virology*, *70*(10), 6723–6732.

Berkhout, B., & Van Wamel, J. L. (2000). The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA*, *6*(2), 282–295.

Berkowitz, R., Fisher, J., & Goff, S. P. (1996). RNA packaging. *Current Topics in Microbiology and Immunology*, *214*, 177–218.

Bernacchi, S., Abd El-Wahab, E. W., Dubois, N., Hijnen, M., Smyth, R. P., Mak, J., Marquet, R., & Paillart, J.-C. (2017). HIV-1 Pr55Gagbinds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biology*, *14*(1), 90–103.

Bewley, M. C., Reinhart, L., Stake, M. S., Nadaraia-Hoke, S., Parent, L. J., & Flanagan, J. M. (2017). A non-cleavable hexahistidine affinity tag at the carboxyl-terminus of the HIV-1 Pr55Gag polyprotein alters nucleic acid binding properties. *Protein Expression and Purification*, *130*(Supplement C), 137–145.

Bharat, T. A. M., Davey, N. E., Ulbrich, P., Riches, J. D., de Marco, A., Rumlova, M., Sachse, C., Ruml, T., & Briggs, J. A. G. (2012). Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy. *Nature*, *487*(7407), 385–389.

Bohl, C. R., Brown, S. M., & Weldon, R. A., Jr (2005). The pp24 phosphoprotein of Mason-Pfizer monkey virus contributes to viral genome packaging. *Retrovirology*, *2*, 68. https://doi.org/10.1186/1742-4690-2-68.

Bohmová, K., Hadravová, R., Štokrová, J., Tůma, R., Ruml, T., Pichová, I., & Rumlová, M. (2010). Effect of Dimerizing Domains and Basic Residues on In Vitro and In Vivo Assembly of Mason-Pfizer Monkey Virus and Human Immunodeficiency Virus. *Journal of Virology*, *84*(4), 1977–1988.

Bohmová, K., Spiwok, V., Lepšík, M., Hadravová, R., Křížová, I., Ulbrich, P., Pichová, I., Bednárová, L., Ruml, T., & Rumlová, M. (2014). Role of Mason-Pfizer monkey virus CA-NC spacer peptide-like domain in assembly of immature particles. *Journal of Virology*, *88*(24), 14148–14160.

Bradac, J., & Hunter, E. (1984). Polypeptides of Mason-Pfizer monkey virus I. Synthesis and processing of the gag-gene products. *Virology*, *138*(2), 260–275.

Brandt, S., Blißenbach, M., Grewe, B., Konietzny, R., Grunwald, T., & Überla, K. (2007). Rev Proteins of Human and Simian Immunodeficiency Virus Enhance RNA Encapsidation. *PLOS Pathogens*, *3*(4), e54. https://doi.org/10.1371/journal.ppat.0030054.

Bray, M., Prasad, S., Dubay, J. W., Hunter, E., Jeang, K. T., Rekosh, D., & Hammarskjöld, M. L. (1994). A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(4), 1256–1260.

Brown, J. D., Kharytonchyk, S., Chaudry, I., Iyer, A. S., Carter, H., Becker, G., Desai, Y., Glang, L., Choi, S. H., Singh, K., Lopresti, M. W., Orellana, M., Rodriguez, T., Oboh, U., Hijji, J., Ghinger, F. G., Stewart, K., Francis, D., Edwards, B., … Summers, M. F. (2020). Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science (New York, N.Y.)*, *368*(6489), 413–417.

Bryant, M. L., Gardner, M. B., Marx, P. A., Maul, D. H., Lerche, N. W., Osborn, K. G., Lowenstine, L. J., Bodgen, A., Arthur, L. O., & Hunter, E. (1986). Immunodeficiency in Rhesus Monkeys Associated With the Original Mason-Pfizer Monkey Virus. *JNCI: Journal of the National Cancer Institute*, *77*(4), 957–965.

Burdick, R. C., Li, C., Munshi, M., Rawson, J. M. O., Nagashima, K., Hu, W.-S., & Pathak, V. K. (2020). HIV-1 uncoats in the nucleus near sites of integration. *Proceedings of the National Academy of Sciences*, *117*(10), 5486–5493.

Bush, D. L., & Vogt, V. M. (2014). In Vitro Assembly of Retroviruses. *Annual Review of Virology*, *1*(1), 561–580.

Butsch, M., & Boris-Lawrie, K. (2002). Destiny of Unspliced Retroviral RNA: Ribosome and/or Virion? *Journal of Virology*, *76*(7), 3089–3094.

Cai, M., Huang, Y., Craigie, R., & Clore, G. M. (2010). Structural basis of the association of HIV-1 matrix protein with DNA. PloS one, 5(12), e15675. https://doi.org/10.1371/journal.pone.0015675.

Campbell, S., & Vogt, V. M. (1997). In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: Identification of the p10 domain as a morphological determinant in the formation of spherical particles. *Journal of Virology*, *71*(6), 4425–4435.

Campbell, S., Fisher, R. J., Towler, E. M., Fox, S., Issaq, H. J., Wolfe, T., Phillips, L. R., & Rein,. A. (2001). Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proceedings of the National Academy of Sciences*, *98*(19), 10875–10879.

Campbell, S., & Rein, A. (1999). In Vitro Assembly Properties of Human Immunodeficiency Virus Type 1 Gag Protein Lacking the p6 Domain. *Journal of Virology*, *73*(3), 2270–2279.

Campos-Olivas, R., Newman, J. L., & Summers, M. F. (2000). Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses11Edited by P. E. Wright. *Journal of Molecular Biology*, *296*(2), 633–649.

Chamanian, M., Purzycka, K. J., Wille, P. T., Ha, J. S., McDonald, D., Gao, Y., Le Grice, S. F. J., & Arts, E. J. (2013). A cis-acting element in retroviral genomic RNA links Gag-Pol ribosomal frameshifting to selective viral RNA encapsidation. *Cell Host & Microbe*, *13*(2), 181–192.

Chameettachal, A., Vivet-Boudou, V., Pitchai, F., Pillai, V. N., Ali, L. M., Krishnan, A., Bernacchi, S., Mustafa, F., Marquet, R., & Rizvi, T. A. (2021). A purine loop and the primer binding site are critical for the selective encapsidation of mouse mammary tumor virus genomic RNA by Pr77Gag. *Nucleic Acids Research*, *49*(8), 4668–4688. https://doi.org/10.1093/nar/gkab223.

Checkley, M. A., Luttge, B. G., & Freed, E. O. (2011). HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *Journal of Molecular Biology*, *410*(4), 582–608.

Chu, H. H., Chang, Y. F. & Wang, C. T. (2006). Mutations in the α-helix Directly C-terminal to the Major Homology Region of Human Immunodeficiency Virus Type 1 Capsid Protein Disrupt Gag Multimerization and Markedly Impair Virus Particle Production. *Journal of Biomedical Science,* 13, 645–656.

Chukkapalli, V., Oh, S. J., & Ono, A. (2010). Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(4), 1600–1605.

Clever, J. L., & Parslow, T. G. (1997). Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *Journal of Virology*, *71*(5), 3407–3414.

Clever, J. L., Sassetti, C., & Parslow, T. G. (1995). RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *Journal of Virology*, *69*(4), 2101–2109.

Cockrell, A. S., Van Praag, H., Santistevan, N., Ma, H., & Kafri, T. (2011). The HIV-1 Rev/RRE system is required for HIV-1 5' UTR cis elements to augment encapsidation of heterologous RNA into HIV-1 viral particles. *Retrovirology*, *8*, 51. https://doi.org/10.1186/1742-4690-8-51.

Coffin, J.M., Hughes, S.H., Varmus, H.E. (1997) *Retroviruses*. Cold Spring Harbor New York: Cold Spring Harbor Laboratory Press.

Comas-Garcia, M., Datta, S. A., Baker, L., Varma, R., Gudla, P. R., & Rein, A. (2017). Dissection of specific binding of HIV-1 Gag to the "packaging signal" in viral RNA. *ELife*, *6*, e27055. https://doi.org/10.7554/eLife.27055.

Comas-Garcia, M., Davis, S. R., & Rein, A. (2016). On the Selective Packaging of Genomic RNA by HIV-1. *Viruses*, *8*(9), 246. https://doi.org/10.3390/v8090246.

Craigie R. (1992). Hotspots and warm spots: integration specificity of retroelements. *Trends in genetics*, *8*(6), 187–190.

Darty, K., Denise, A., & Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, *25*(15), 1974–1975.

De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., & Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*, *279*(5349), 384–388.

Dhar, R., McClements, W. L., Enquist, L. W., & Woude, G. F. V. (1980). Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. *Proceedings of the National Academy of Sciences*, *77*(7), 3937–3941.

Dharan, A., Bachmann, N., Talley, S., Zwikelmaier, V., & Campbell, E. M. (2020). Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nature Microbiology*, *5*(9), 1088–1095.

Dick, R. A., & Vogt, V. M. (2014). Membrane interaction of retroviral Gag proteins. *Frontiers in Microbiology*, *5*, 187. https://doi.org/10.3389/fmicb.2014.00187.

Didierlaurent, L., Racine, P. J., Houzet, L., Chamontin, C., Berkhout, B., & Mougel, M. (2011). Role of HIV-1 RNA and protein determinants for the selective packaging of spliced and unspliced viral RNA and host U6 and 7SL RNA in virus particles. *Nucleic Acids Research*, *39*(20), 8915–8927.

Dilley, K. A., Ni, N., Nikolaitchik, O. A., Chen, J., Galli, A., & Hu, W.-S. (2011). Determining the Frequency and Mechanisms of HIV-1 and HIV-2 RNA Copackaging by Single-Virion Analysis▽. *Journal of Virology*, *85*(20), 10499–10508.

Ding, P., Kharytonchyk, S., Waller, A., Mbaekwe, U., Basappa, S., Kuo, N., Frank, H. M., Quasney, C., Kidane, A., Swanson, C., Van, V., Sarkar, M., Cannistraci, E., Chaudhary, R., Flores, H., Telesnitsky, A., & Summers, M. F. (2020). Identification of the initial nucleocapsid recognition element in the HIV-1 RNA packaging signal. *Proceedings of the National Academy of Sciences*, *117*(30), 17737. https://doi.org/10.1073/pnas.2008519117.

Dorfman, T., Luban, J., Goff, S. P., Haseltine, W. A., & Göttlinger, H. G. (1993). Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *Journal of Virology*, *67*(10), 6159–6169.

Dostálková, A., Kaufman, F., Křížová, I., Kultová, A., Strohalmová, K., Hadravová, R., Ruml, T., & Rumlová, M. (2018). Mutations in the Basic Region of the Mason-Pfizer Monkey Virus Nucleocapsid Protein Affect Reverse Transcription, Genomic RNA Packaging, and the Virus Assembly Site. *Journal of Virology*, *92*(10), e00106-18. https://doi.org/10.1128/JVI.00106-18.

D'Souza, V., & Summers, M. F. (2004). Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature*, *431*(7008), 586. https://doi.org/10.1038/nature02944.

D'Souza, V., & Summers, M. F. (2005). How retroviruses select their genomes. *Nature Reviews Microbiology*, *3*(8), 643. https://doi.org/10.1038/nrmicro1210.

Dubois, N., Marquet, R., Paillart, J. C., & Bernacchi, S. (2018). Retroviral RNA Dimerization: From Structure to Functions. F*rontiers in Microbiology*, *9*, 527. https://doi.org/10.3389/fmicb.2018.00527.

Dubois, N., Khoo, K. K., Ghossein, S., Seissler, T., Wolff, P., McKinstry, W. J., Mak, J., Paillart, J.-C., Marquet, R., & Bernacchi, S. (2018b). The C-terminal p6 domain of the HIV-1 Pr55Gag precursor is required for specific binding to the genomic RNA. *RNA Biology*, *15*(7), 923–936.

Ehrlich, L. S., Agresta, B. E., & Carter, C. A. (1992). Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro. *Journal of Virology*, *66*(8), 4874–4883.

Fernandes, J., Jayaraman, B., & Frankel, A. (2012). The HIV-1 Rev response element. *RNA Biology*, *9*(1), 6–11.

Ferrer, M., Clerté, C., Chamontin, C., Basyuk, E., Lainé, S., Hottin, J., Bertrand, E., Margeat, E., & Mougel, M. (2016). Imaging HIV-1 RNA dimerization in cells by multicolor super-resolution and fluctuation microscopies. *Nucleic Acids Research*, *44*(16), 7922–7934.

Fine, D. L., Landon, J. C., Pienta, R. J., Kubicek, M. T., Valerio, M. G., Loeb, W. F., & Chopra, H. C. (1975). Responses of infant rhesus monkeys to inoculation with Mason-Pfizer monkey virus materials. *Journal of the National Cancer Institute*, *54*(3), 651–658.

Fisher, R. J., Rein, A., Fivash, M., Urbaneja, M. A., Casas-Finet, J. R., Medaglia, M., & Henderson, L. E. (1998). Sequence-Specific Binding of Human Immunodeficiency Virus Type 1 Nucleocapsid Protein to Short Oligonucleotides. *Journal of Virology*, *72*(3), 1902–1909.

Freed, E. O. (2015). HIV-1 assembly, release and maturation. *Nature Reviews. Microbiology*, *13*(8), 484–496.

Füzik, T., Píchalová, R., Schur, F. K. M., Strohalmová, K., Křížová, I., Hadravová, R., Rumlová, M., Briggs, J. A. G., Ulbrich, P., & Ruml, T. (2016). Nucleic Acid Binding by Mason-Pfizer Monkey Virus CA Promotes Virus Assembly and Genome Packaging. *Journal of Virology*, *90*(9), 4593–4603.

Gallo, S. A., Finnegan, C. M., Viard, M., Raviv, Y., Dimitrov, A., Rawat, S. S., Puri, A., Durell, S., & Blumenthal, R. (2003). The HIV Env-mediated fusion reaction. *Biochimica Et Biophysica Acta*, *1614*(1), 36–50.

Gamble, T. R., Yoo, S., Vajdos, F. F., Schwedler, U. K. von, Worthylake, D. K., Wang, H., McCutcheon, J. P., Sundquist, W. I., & Hill, C. P. (1997). Structure of the Carboxyl-Terminal Dimerization Domain of the HIV-1 Capsid Protein. *Science*, *278*(5339), 849–853.

Ganser-Pornillos, B. K., Yeager, M., & Pornillos, O. (2012). Assembly and architecture of HIV. *Advances in Experimental Medicine and Biology*, *726*, 441–465.

Ganser-Pornillos, B. K., Yeager, M., & Sundquist, W. I. (2008). The structural biology of HIV assembly. *Current Opinion in Structural Biology*, *18*(2), 203–217.

Ghazawi, A., Mustafa, F., Phillip, P. S., Jayanth, P., Ali, J., & Rizvi, T. A. (2006). Both the 5' and 3' LTRs of FIV contain minor RNA encapsidation determinants compared to the two core packaging determinants within the 5' untranslated region and gag. *Microbes and Infection*, *8*(3), 767–778.

Gherghe, C., Lombo, T., Leonard, C. W., Datta, S. A. K., Bess, J. W., Gorelick, R. J., Rein, A., & Weeks, K. M. (2010). Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(45), 19248–19253.

Gibbs, J. S., Regier, D. A., & Desrosiers, R. C. (1994). Construction and in vitro properties of SIVmac mutants with deletions in "nonessential" genes. *AIDS Research and Human Retroviruses*, *10*(4), 333–342.

Gilboa, E., Mitra, S. W., Goff, S., & Baltimore, D. (1979). A detailed model of reverse transcription and tests of crucial aspects. *Cell*, *18*(1), 93–100.

Ginn, S. L., Alexander, I. E., Edelstein, M. L., Abedi, M. R., & Wixon, J. (2013). Gene therapy clinical trials worldwide to 2012—An update. *The Journal of Gene Medicine*, *15*(2), 65–77.

Gitti, R. K., Lee, B. M., Walker, J., Summers, M. F., Yoo, S., & Sundquist, W. I. (1996). Structure of the Amino-Terminal Core Domain of the HIV-1 Capsid Protein. *Science*, *273*(5272), 231–235.

Goff, S. P. (2001). Intracellular trafficking of retroviral genomes during the early phase of infection: Viral exploitation of cellular pathways. *The Journal of Gene Medicine*, *3*(6), 517–528.

Gorelick, R. J., Henderson, L. E., Hanser, J. P., & Rein, A. (1988). Point mutants of Moloney murine leukemia virus that fail to package viral RNA: Evidence for specific RNA recognition by a "zinc finger-like" protein sequence. *Proceedings of the National Academy of Sciences*, *85*(22), 8420–8424.

Gottwein, E., Bodem, J., Müller, B., Schmechel, A., Zentgraf, H., & Kräusslich, H.-G. (2003). The Mason-Pfizer Monkey Virus PPPY and PSAP Motifs Both Contribute to Virus Release. *Journal of Virology*, *77*(17), 9474–9485.

Griffin, S. D., Allen, J. F., & Lever, A. M. (2001). The major human immunodeficiency virus type 2 (HIV-2) packaging signal is present on all HIV-2 RNA species: Cotranslational RNA encapsidation and limitation of Gag protein confer specificity. *Journal of Virology*, *75*(24), 12058–12069.

Guesdon, F. M. J., Greatorex, J., Rhee, S. R., Fisher, R., Hunter, E., & Lever, A. M. L. (2001). Sequences in the 5′ Leader of Mason-Pfizer Monkey Virus Which Affect Viral Particle Production and Genomic RNA Packaging: Development of MPMV Packaging Cell Lines. *Virology*, *288*(1), 81–88.

Guo, X., Roy, B. B., Hu, J., Roldan, A., Wainberg, M. A., & Liang, C. (2005). The R362A mutation at the C-terminus of CA inhibits packaging of human immunodeficiency virus type 1 RNA. *Virology*, *343*(2), 190–200.

Harrison, G. P., Hunter, E., & Lever, A. M. (1995). Secondary structure model of the Mason-Pfizer monkey virus 5' leader sequence: Identification of a structural motif common to a variety of retroviruses. *Journal of Virology*, *69*(4), 2175–2186.

Hayashi, T., Shioda, T., Iwakura, Y., & Shibuta, H. (1992). RNA packaging signal of human immunodeficiency virus type 1. *Virology*, *188*(2), 590–599.

Henderson, L. E., Sowder, R., Smythers, G., Benveniste, R. E., & Oroszlan, S. (1985). Purification and N-terminal amino acid sequence comparisons of structural proteins from retrovirus-D/Washington and Mason-Pfizer monkey virus. *Journal of Virology*, *55*(3), 778–787.

Heng, X., Kharytonchyk, S., Garcia, E. L., Lu, K., Divakaruni, S. S., LaCotti, C., Edme, K., Telesnitsky, A., & Summers, M. F. (2012). Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. *Journal of Molecular Biology*, *417*(3), 224–239.

Hrusková-Heidingsfeldová, O., Andreansky, M., Fábry, M., Bláha, I., Strop, P., & Hunter, E. (1995). Cloning, Bacterial Expression, and Characterization of the Mason-Pfizer Monkey Virus Proteinase. *Journal of Biological Chemistry*, *270*(25), 15053–15058.

Hu, B., Tai, A., & Wang, P. (2011). Immunization delivered by lentiviral vectors for cancer and infectious diseases. *Immunological Reviews*, *239*(1), 45–61.

Hughes, S. H. (2015). Reverse Transcription of Retroviruses and LTR Retrotransposons. *Microbiology Spectrum*, *3*(2), MDNA3-0027–2014.

Inlora, J., Collins, D. R., Trubin, M. E., Chung, J. Y., & Ono, A. (2014). Membrane binding and subcellular localization of retroviral Gag proteins are differentially regulated by MA interactions with phosphatidylinositol-(4,5)-bisphosphate and RNA. *mBio*, *5*(6), e02202. https://doi.org/10.1128/mBio.02202-14

Jaballah, S. A., Aktar, S. J., Ali, J., Phillip, P. S., Al Dhaheri, N. S., Jabeen, A., & Rizvi, T. A. (2010). A G–C-Rich Palindromic Structural Motif and a Stretch of Single-Stranded Purines Are Required for Optimal Packaging of Mason–Pfizer Monkey Virus (MPMV) Genomic RNA. *Journal of Molecular Biology*, *401*(5), 996–1014.

Jewell, N. A., & Mansky, L. M. (2000). In the beginning: Genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly. *Journal of General Virology*, *81*(8), 1889–1899.

Johnson, S. F., & Telesnitsky, A. (2010). Retroviral RNA Dimerization and Packaging: The What, How, When, Where, and Why. *PLOS Pathogens*, *6*(10), e1001007. https://doi.org/10.1371/journal.ppat.1001007

Jones, C. P., Cantara, W. A., Olson, E. D., & Musier-Forsyth, K. (2014). Small-angle X-ray scattering-derived structure of the HIV-1 5′ UTR reveals 3D tRNA mimicry. *Proceedings of the National Academy of Sciences*, *111*(9), 3395–3400.

Jones, C. P., Datta, S. A. K., Rein, A., Rouzina, I., & Musier-Forsyth, K. (2011). Matrix Domain Modulates HIV-1 Gag's Nucleic Acid Chaperone Activity via Inositol Phosphate Binding. *Journal of Virology*, *85*(4), 1594–1603.

Jouvenet, N., Lainé, S., Pessel-Vivares, L., & Mougel, M. (2011). Cell biology of retroviral RNA packaging. *RNA Biology*, *8*(4), 572–580.

Kalloush, R. M., Vivet-Boudou, V., Ali, L. M., Mustafa, F., Marquet, R., & Rizvi, T. A. (2016). Packaging of Mason-Pfizer monkey virus (MPMV) genomic RNA depends upon conserved long-range interactions (LRIs) between U5 and gag sequences. *RNA*, *22*(6), 905–919.

Kalloush, R. M., Vivet-Boudou, V., Ali, L. M., Pillai, V. N., Mustafa, F., Marquet, R., & Rizvi, T. A. (2019). Stabilizing role of structural elements within the 5´ Untranslated Region (UTR) and gag sequences in Mason-Pfizer monkey virus (MPMV) genomic RNA packaging. *RNA Biology*, *16*(5), 612–625.

Karabiber, F., McGinnis, J. L., Favorov, O. V., & Weeks, K. M. (2013). QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA*, *19*(1), 63–73.

Kaye, J. F., & Lever, A. M. L. (1998). Nonreciprocal Packaging of Human Immunodeficiency Virus Type 1 and Type 2 RNA: A Possible Role for the p2 Domain of Gag in RNA Encapsidation. *Journal of Virology*, *72*(7), 5877–5885.

Keane, S. C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, A., Florwick, A., Santos, J., Bolden, N. C., McCowin, S., Case, D. A., Johnson, B. A., Salemi, M., Telesnitsky, A., & Summers, M. F. (2015). RNA structure. Structure of the HIV-1 RNA packaging signal. *Science (New York, N.Y.)*, *348*(6237), 917–921.

Keane, S. C., & Summers, M. F. (2016). NMR Studies of the Structure and Function of the HIV-1 5'-Leader. Viruses, 8(12), 338. https://doi.org/10.3390/v8120338

Kenyon, J. C., Ghazawi, A., Cheung, W. K. S., Phillip, P. S., Rizvi, T. A., & Lever, A. M. L. (2008). The secondary structure of the 5' end of the FIV genome reveals a long-range interaction between R/U5 and gag sequences, and a large, stable stem-loop. *RNA*, *14*(12), 2597–2608.

Kenyon, J. C., Tanner, S. J., Legiewicz, M., Phillip, P. S., Rizvi, T. A., Le Grice, S. F. J., & Lever, A. M. L. (2011). SHAPE analysis of the FIV Leader RNA reveals a structural switch potentially controlling viral packaging and genome dimerization. *Nucleic Acids Research*, *39*(15), 6692–6704.

King, A. M. Q., Adams M. J., Carstens, E. B., Lefkowitz, E. J. (2012). *Virus Taxonomy* (pp. 477-495). Elsevier.

Klikova, M., Rhee, S. S., Hunter, E., & Ruml, T. (1995). Efficient in vivo and in vitro assembly of retroviral capsids from Gag precursor proteins expressed in bacteria. *Journal of Virology*, *69*(2), 1093–1098.

Kroupa, T., Datta, S. A. K., & Rein, A. (2020). Distinct Contributions of Different Domains within the HIV-1 Gag Polyprotein to Specific and Nonspecific Interactions with RNA. *Viruses*, *12*(4), 394. https://doi.org/10.3390/v12040394.

Kroupa, T., Langerová, H., Doležal, M., Prchal, J., Spiwok, V., Hunter, E., Rumlová, M., Hrabal, R., & Ruml, T. (2016). Membrane Interactions of the Mason-Pfizer Monkey Virus Matrix Protein and Its Budding Deficient Mutants. *Journal of Molecular Biology*, *428*(23), 4708–4722.

Kutluay, S. B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., & Bieniasz, P. D. (2014). Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell*, *159*(5), 1096–1109.

Kuzembayeva, M., Dilley, K., Sardo, L., & Hu, W.-S. (2014a). Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology*, *454–455*(Supplement C), 362–370.

Lever, A. M. (2009). RNA packaging in lentiviruses. *Retrovirology*, *6*(S2), I-13.

Lever, A. M. (2007). HIV-1 RNA packaging. *Advances in Pharmacology (San Diego, Calif.)*, *55*, 1–32.

Lingappa, J. R., Reed, J. C., Tanaka, M., Chutiraka, K., & Robinson, B. A. (2014). How HIV-1 Gag assembles in cells: Putting together pieces of the puzzle. *Virus Research*, *193*, 89–107.

Lingappa, J. R., Tanaka, M., Barajas, B. C., Robinson, B. A., Phuong, D. J., Chutiraka, K., & Reed, J. C. (2017). HIV-1 initiates genomic RNA packaging in a unique subset of host RNA granules. *BioRxiv*, 183855. https://doi.org/10.1101/183855.

Linial, M. L., Fan, H., Hahn, B., Lwer, R., Neil, J., Quackenbusch, S., Rethwilm, A., Sonigo, P., Stoye, J. & Tristem, M. (2005). *Retroviridae*. In Virus Taxonomy, pp. 421–440. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Dessselberger & L. A. Ball. Oxford: Elsevier.

Lochmann, T. L., Bann, D. V., Ryan, E. P., Beyer, A. R., Mao, A., Cochrane, A., & Parent, L. J. (2013). NC-mediated nucleolar localization of retroviral gag proteins. *Virus Research*, *171*(2), 304–318.

Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E. L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A., Divakaruni, S. S., LaCotti, C., Barton, S., Tummillo, D., Hosic, A., Edme, K., Albrecht, S., Telesnitsky, A., & Summers, M. F. (2011). NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science (New York, N.Y.)*, *334*(6053), 242–245.

Lu, K., Heng, X., & Summers, M. F. (2011). Structural determinants and mechanism of HIV-1 genome packaging. *Journal of Molecular Biology*, *410*(4), 609–633.

Luban, J., & Goff, S. P. (1994). Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *Journal of Virology*, *68*(6), 3784–3793.

Lubow, J., & Collins, K. L. (2020). Vpr Is a VIP: HIV Vpr and Infected Macrophages Promote Viral Pathogenesis. Viruses, 12(8), 809. https://doi.org/10.3390/v12080809.

Macek, P., Chmelík, J., Krízová, I., Kaderávek, P., Padrta, P., Zídek, L., Wildová, M., Hadravová, R., Chaloupková, R., Pichová, I., Ruml, T., Rumlová, M., & Sklenár, V. (2009). NMR structure of the N-terminal domain of capsid protein from the mason-pfizer monkey virus. *Journal of Molecular Biology*, *392*(1), 100–114.

Mailler, E., Bernacchi, S., Marquet, R., Paillart, J. C., Vivet-Boudou, V., & Smyth, R. P. (2016). The Life-Cycle of the HIV-1 Gag-RNA Complex. *Viruses*, *8*(9), 248. https://doi.org/10.3390/v8090248.

Maldonado, R. J. K., & Parent, L. J. (2016). Orchestrating the Selection and Packaging of Genomic RNA by Retroviruses: An Ensemble of Viral and Host Factors. *Viruses*, *8*(9), 257. https://doi.org/10.3390/v8090257.

Maldonado, R., Rice, B., Chen, E. C., Tuffy, K. M., Chiari, E. F., Fahrbach, K. M., Hope, T. J., & Parent, L. J. (2020). Visualizing Association of the Retroviral Gag Protein with Unspliced Viral RNA in the Nucleus. *mBio*, *11*(2), e00524-20. https://doi.org/10.1128/mBio.00524-20.

Mammano, F., Ohagen, A., Höglund, S., & Göttlinger, H. G. (1994). Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *Journal of Virology*, *68*(8), 4927–4936.

Marquet, R., Baudin, F., Gabus, C., Darlix, J. L., Mougel, M., Ehresmann, C., & Ehresmann, B. (1991). Dimerization of human immunodeficiency virus (type 1) RNA: Stimulation by cations and possible mechanism. *Nucleic Acids Research*, *19*(9), 2349–2357.

Mathews, D. H., Sabina, J., Zuker, M., & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, 288(5), 911–940.

McBride, M. S., & Panganiban, A. T. (1996). The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *Journal of Virology*, *70*(5), 2963–2973.

McBride, M. S., & Panganiban, A. T. (1997). Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. *Journal of Virology*, *71*(3), 2050–2058.

McDermott, J., Farrell, L., Ross, R., & Barklis, E. (1996). Structural analysis of human immunodeficiency virus type 1 Gag protein interactions, using cysteine-specific reagents. *Journal of Virology*, *70*(8), 5106–5114.

McKinstry, W. J., Hijnen, M., Tanwar, H. S., Sparrow, L. G., Nagarajan, S., Pham, S. T., & Mak, J. (2014). Expression and purification of soluble recombinant full length HIV-1 Pr55Gag protein in Escherichia coli. *Protein Expression and Purification*, *100*(Supplement C), 10–18.

Menéndez-Arias, L., Sebastián-Martín, A., & Álvarez, M. (2017). Viral reverse transcriptases. *Virus Research*, *234*, 153–176.

Méric, C., Gouilloud, E., & Spahr, P. F. (1988). Mutations in Rous sarcoma virus nucleocapsid protein p12 (NC): Deletions of Cys-His boxes. *Journal of Virology*, *62*(9), 3328–3333.

Mitra, M., Wang, W., Vo, M. N., Rouzina, I., Barany, G., & Musier-Forsyth, K. (2013). The N-terminal zinc finger and flanking basic domains represent the minimal region of the human immunodeficiency virus type-1 nucleocapsid protein for targeting chaperone function. Biochemistry, 52(46), 8226–8236.

Miyazaki, Y., Garcia, E. L., King, S. R., Iyalla, K., Loeliger, K., Starck, P., Syed, S., Telesnitsky, A., & Summers, M. F. (2010). An RNA Structural Switch Regulates Diploid Genome Packaging by Moloney Murine Leukemia Virus. *Journal of Molecular Biology*, *396*(1), 141–152.

Miyazaki, Y., Miyake, A., Nomaguchi, M., & Adachi, A. (2011). Structural dynamics of retroviral genome and the packaging. *Frontiers in Microbiology, 2*, 264. https://doi.org/10.3389/fmicb.2011.00264

Momany, C., Kovari, L. C., Prongay, A. J., Keller, W., Gitti, R. K., Lee, B. M., Gorbalenya, A. E., Tong, L., McClure, J., Ehrlich, L. S., Summers, M. F., Carter, C., & Rossmann, M. G. (1996). Crystal structure of dimeric HIV-1 capsid protein. *Nature Structural Biology*, *3*(9), 763–770.

Montiel, N. a. (2010). Review article: An updated review of simian betaretrovirus (SRV) in macaque hosts. *Journal of Medical Primatology*, *39*(5), 303–314.

Moore, J. P., & Doms, R. W. (2003). The entry of entry inhibitors: A fusion of science and medicine. *Proceedings of the National Academy of Sciences*, *100*(19), 10598–10602.

Moore, M. D., Fu, W., Nikolaitchik, O., Chen, J., Ptak, R. G., & Hu, W.-S. (2007). Dimer initiation signal of human immunodeficiency virus type 1: Its role in partner selection during RNA copackaging and its effects on recombination. *Journal of Virology*, *81*(8), 4002–4011.

Moore, M. D., & Hu, W. S. (2009). HIV-1 RNA dimerization: It takes two to tango. *AIDS Reviews*, *11*(2), 91–102.

Moore, M. D., Nikolaitchik, O. A., Chen, J., Hammarskjöld, M.-L., Rekosh, D., & Hu, W.-S. (2009). Probing the HIV-1 Genomic RNA Trafficking Pathway and Dimerization by Genetic Recombination and Single Virion Analyses. *PLOS Pathogens*, *5*(10), e1000627. https://doi.org/10.1371/journal.ppat.1000627

Morcock, Bp, K., & Jr, C.-F. (2000). Fluorescence and nucleic acid binding properties of the human T-cell leukemia virus-type 1 nucleocapsid protein. *Biochimica et Biophysica Acta*, *1481*(2), 381–394.

Morcock, D. R., Katakam, S., Kane, B. P., & Casas-Finet, J. R. (2002). Fluorescence and nucleic acid binding properties of bovine leukemia virus nucleocapsid protein. *Biophysical Chemistry*, *97*(2), 203–212.

Motomura, K., Chen, J., & Hu, W.-S. (2008). Genetic recombination between human immunodeficiency virus type 1 (HIV-1) and HIV-2, two distinct human lentiviruses. *Journal of Virology*, *82*(4), 1923–1933.

Mortuza, G. B., Haire, L. F., Stevens, A., Smerdon, S. J., Stoye, J. P., & Taylor, I. A. (2004). High-resolution structure of a retroviral capsid hexameric amino-terminal domain. *Nature*, *431*(7007), 481–485.

Muriaux, D., & Darlix, J.-L. (2010). Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biology*, *7*(6), 744–753.

Murphy F. A., Fauquet C. M., Bishop D. H. L., Ghabrial S. A., Jarvis A. W., Martelli G. P., Mayo M. A., Summers M. D. (1994) *Virus taxonomy: The classification and nomenclature of viruses, Retroviridae*, Vienna: Springer-Verlag.

Murphy, R. E., & Saad, J. S. (2020). The Interplay between HIV-1 Gag Binding to the Plasma Membrane and Env Incorporation. *Viruses*, *12*(5), 548. https://doi.org/10.3390/v12050548.

Mustafa, F., Amri, D. A., Ali, F. A., Sari, N. A., Suwaidi, S. A., Jayanth, P., Philips, P. S., & Rizvi, T. A. (2012). Sequences within Both the 5′ UTR and Gag Are Required for Optimal In Vivo Packaging and Propagation of Mouse Mammary Tumor Virus (MMTV) Genomic RNA. *PLoS one*, *7*(10), e47088. https://doi.org/10.1371/journal.pone.0047088.

Mustafa, F., Jayanth, P., Phillip, P. S., Ghazawi, A., Schmidt, R. D., Lew, K. A., & Rizvi, T. A. (2005). Relative activity of the feline immunodeficiency virus promoter in feline and primate cell lines. *Microbes and Infection*, *7*(2), 233–239.

Mustafa, F., Lew, K. A., Schmidt, R. D., Browning, M. T., & Rizvi, T. A. (2004). Mutational analysis of the predicted secondary RNA structure of the Mason-Pfizer monkey virus packaging signal. *Virus Research*, *99*(1), 35–46.

Mustafa, F., Vivet-Boudou, V., Jabeen, A., Ali, L. M., Kalloush, R. M., Marquet, R., & Rizvi, T. A. (2018). The bifurcated stem loop 4 (SL4) is crucial for efficient packaging of mouse mammary tumor virus (MMTV) genomic RNA. *RNA Biology*, *15*(8), 1047–1059.

Nakano, K., & Watanabe, T. (2012). HTLV-1 Rex: the courier of viral messages making use of the host vehicle. *Frontiers in Microbiology*, *3*, 330. https://doi.org/10.3389/fmicb.2012.00330.

Nikolaitchik, O. A., Somoulay, X., Rawson, J., Yoo, J. A., Pathak, V. K., & Hu, W. S. (2020). Unpaired Guanosines in the 5' Untranslated Region of HIV-1 RNA Act Synergistically To Mediate Genome Packaging. *Journal of Virology*, *94*(21), e00439-20. https://doi.org/10.1128/JVI.00439-20.

Nisole, S., & Saïb, A. (2004). Early Steps of retrovirus replicative cycle. *Retrovirology*, *1*, 9. https://doi.org/10.1186/1742-4690-1-9

Novikova, M., Adams, L. J., Fontana, J., Gres, A. T., Balasubramaniam, M., Winkler, D. C., Kudchodkar, S. B., Soheilian, F., Sarafianos, S. G., Steven, A. C., & Freed, E. O. (2018). Identification of a Structural Element in HIV-1 Gag Required for Virus Particle Assembly and Maturation. *mBio*, *9*(5), e01567-18. https://doi.org/10.1128/mBio.01567-18

Olety, B., & Ono, A. (2014). Roles played by acidic lipids in HIV-1 Gag membrane binding. *Virus Research*, *193*, 108–115.

Olson, E. D., Cantara, W. A., & Musier-Forsyth, K. (2015). New Structure Sheds Light on Selective HIV-1 Genomic RNA Packaging. *Viruses*, *7*(8), 4826–4835.

Olson, E. D., & Musier-Forsyth, K. (2019). Retroviral Gag protein-RNA interactions: Implications for specific genomic RNA packaging and virion assembly. *Seminars in Cell & Developmental Biology*, *86*, 129–139.

Pachulska-Wieczorek, K., Błaszczyk, L., Biesiada, M., Adamiak, R. W., & Purzycka, K. J. (2016). The matrix domain contributes to the nucleic acid chaperone activity of HIV-2 Gag. *Retrovirology*, *13*, 18. https://doi.org/10.1186/s12977-016-0245-1

Paillart, J. C., Berthoux, L., Ottmann, M., Darlix, J. L., Marquet, R., Ehresmann, B., & Ehresmann, C. (1996a). A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *Journal of Virology*, *70*(12), 8348–8354.

Paillart, J. -C., Marquet, R., Skripkin, E., Ehresmann, B., & Ehresmann, C. (1994). Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *The Journal of Biological Chemistry*, *269*(44), 27486–27493.

Paillart, J. -C., Skripkin, E., Ehresmann, B., Ehresmann, C., & Marquet, R. (1996b). A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(11), 5572–5577.

Paillart, J. -C., Westhof, E., Ehresmann, C., Ehresmann, B., & Marquet, R. (1997). Non-canonical interactions in a kissing loop complex: The dimerization initiation site of HIV-1 genomic RNA. *Journal of Molecular Biology*, *270*(1), 36–49.

Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R., & Mak, J. (2004). Dimerization of retroviral RNA genomes: An inseparable pair. *Nature Reviews. Microbiology*, *2*(6), 461–472.

Parent, L. J. (2011). New insights into the nuclear localization of retroviral Gag proteins. *Nucleus*, *2*(2), 92–97.

Parent, L. J., & Gudleski, N. (2011). Beyond plasma membrane targeting: Role of the MA domain of Gag in retroviral genome encapsidation. *Journal of Molecular Biology*, *410*(4), 553–564.

Parveen, Z., Mukhtar, M., Goodrich, A., Acheampong, E., Dornburg, R., & Pomerantz, R. J. (2004). Cross-packaging of human immunodeficiency virus type 1 vector RNA by spleen necrosis virus proteins: construction of a new generation of spleen necrosis virus-derived retroviral vectors. *Journal of Virology*, *78*(12), 6480–6488.

Pasquinelli, A. E., Ernst, R. K., Lund, E., Grimm, C., Zapp, M. L., Rekosh, D., Hammarskjöld, M. L., & Dahlberg, J. E. (1997). The constitutive transport element (CTE) of Mason-Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. *The EMBO Journal*, *16*(24), 7500–7510.

Píchalová, R., Füzik, T., Vokatá, B., Rumlová, M., Llano, M., Dostálková, A., Křížová, I., Ruml, T., & Ulbrich, P. (2018). Conserved cysteines in Mason-Pfizer monkey virus capsid protein are essential for infectious mature particle formation. *Virology*, *521*, 108–117.

Pitchai, F., Ali, L., Pillai, V. N., Chameettachal, A., Ashraf, S. S., Mustafa, F., Marquet, R., & Rizvi, T. A. (2018). Expression, purification, and characterization of biologically active full-length Mason-Pfizer monkey virus (MPMV) Pr78[Gag]. Scientific reports, 8(1), 11793. https://doi.org/10.1038/s41598-018-30142-0.

Pollard, V. W., & Malim, M. H. (1998). The HIV-1 Rev protein. *Annual Review of Microbiology*, *52*, 491–532.

Poon, D. T., Wu, J., & Aldovini, A. (1996). Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *Journal of Virology*, *70*(10), 6607–6616.

Prchal, J., Kroupa, T., Ruml, T., & Hrabal, R. (2013). Interaction of Mason-Pfizer monkey virus matrix protein with plasma membrane. *Frontiers in Microbiology*, *4*, 423. https://doi.org/10.3389/fmicb.2013.00423

Prchal, J., Srb, P., Hunter, E., Ruml, T., & Hrabal, R. (2012). The Structure of Myristoylated Mason-Pfizer Monkey Virus Matrix Protein and the Role of Phosphatidylinositol-(4,5)-Bisphosphate in Its Membrane Binding. *Journal of Molecular Biology*, *423*(3), 427–438.

Rein, A. (2019). RNA Packaging in HIV. *Trends in Microbiology*, *27*(8), 715–723.

Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, *11*, 129. doi: 10.1186/1471-2105-11-129.

Rhee, S. S., & Hunter, E. (1991). Amino acid substitutions within the matrix protein of type D retroviruses affect assembly, transport and membrane association of a capsid. *The EMBO Journal*, *10*(3), 535–546.

Rhee, S. S., & Hunter, E. (1990). A single amino acid substitution within the matrix protein of a type D retrovirus converts its morphogenesis to that of a type C retrovirus. *Cell*, *63*(1), 77–86.

Rice, B. L., Lochmann, T. L., & Parent, L. J. (2020). RNA-Binding Domains of Heterologous Viral Proteins Substituted for Basic Residues in the RSV Gag NC Domain Restore Specific Packaging of Genomic RNA. *Viruses*, *12*(4), 370. https://doi.org/10.3390/v12040370

Rizvi, T. A., Lew, K. A., Murphy, E. C., & Schmidt, R. D. (1996a). Role of Mason-Pfizer monkey virus (MPMV) constitutive transport element (CTE) in the propagation of MPMV vectors by genetic complementation using homologous/heterologous env genes. *Virology*, *224*(2), 517–532.

Rizvi, T. A., & Panganiban, A. T. (1993). Simian immunodeficiency virus RNA is efficiently encapsidated by human immunodeficiency virus type 1 particles. *Journal of Virology*, *67*(5), 2681–2688.

Rizvi, T. A., Kenyon, J. C., Ali, J., Aktar, S. J., Phillip, P. S., Ghazawi, A., Mustafa, F., & Lever, A. M. L. (2010). Optimal packaging of FIV genomic RNA depends upon a conserved long-range interaction and a palindromic sequence within gag. *Journal of Molecular Biology*, *403*(1), 103–119.

Rizvi, T. A., Schmidt, R. D., & Lew, K. A. (1997). Mason–Pfizer Monkey Virus (MPMV) Constitutive Transport Element (CTE) Functions in a Position-Dependent Manner. *Virology*, *236*(1), 118–129.

Rizvi, Tahir A., Schmidt, R. D., Lew, K. A., & Keeling, M. E. (1996b). Rev/RRE-Independent Mason–Pfizer Monkey Virus Constitutive Transport Element-Dependent Propagation of SIVmac239 Vectors Using a Single Round of Replication Assay. *Virology*, *222*(2), 457–463.

Roy, B. B., Russell, R. S., Turner, D., & Liang, C. (2006). The T12I mutation within the SP1 region of Gag restricts packaging of spliced viral RNA into human immunodeficiency virus type 1 with mutated RNA packaging signals and mutated nucleocapsid sequence. *Virology*, *344*(2), 304–314.

Russell, R. S., Liang, C., & Wainberg, M. A. (2004). Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology*, *1*, 23. https://doi.org/10.1186/1742-4690-1-23

Russell, R. S., Roldan, A., Detorio, M., Hu, J., Wainberg, M. A., & Liang, C. (2003). Effects of a Single Amino Acid Substitution within the p2 Region of Human Immunodeficiency Virus Type 1 on Packaging of Spliced Viral RNA. *Journal of Virology*, *77*(24), 12986–12995.

Rye-McCurdy, T., Olson, E. D., Liu, S., Binkley, C., Reyes, J. P., Thompson, B. R., Flanagan, J. M., Parent, L. J., & Musier-Forsyth, K. (2016). Functional Equivalence of Retroviral MA Domains in Facilitating Psi RNA Binding Specificity by Gag. *Viruses*, *8*(9), 256. https://doi.org/10.3390/v8090256

Sakalian, M., Parker, S. D., Weldon, R. A., & Hunter, E. (1996). Synthesis and assembly of retrovirus Gag precursors into immature capsids in vitro. *Journal of Virology*, *70*(6), 3706–3715.

Sakalian, M., Dittmer, S. S., Gandy, A. D., Rapp, N. D., Zábranský, A., & Hunter, E. (2002). The Mason-Pfizer monkey virus internal scaffold domain enables in vitro assembly of human immunodeficiency virus type 1 Gag. *Journal of Virology*, *76*(21), 10811–10820. https://doi.org/10.1128/jvi.76.21.10811-10820.2002

Sakalian, M., & Hunter, E. (1999). Separate Assembly and Transport Domains within the Gag Precursor of Mason-Pfizer Monkey Virus. *Journal of Virology*, *73*(10), 8073–8082.

Schmidt, R. D., Mustafa, F., Lew, K. A., Browning, M. T., & Rizvi, T. A. (2003). Sequences within both the 5' untranslated region and the gag gene are important for efficient encapsidation of Mason-Pfizer monkey virus RNA. *Virology*, *309*(1), 166–178.

Schröder, A. R. W., Shinn, P., Chen, H., Berry, C., Ecker, J. R., & Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, *110*(4), 521–529.

Schur, F. K. M., Hagen, W. J. H., Rumlová, M., Ruml, T., Müller, B., Kräusslich, H.-G., & Briggs, J. A. G. (2015). Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature*, *517*(7535), 505. https://doi.org/10.1038/nature13838

Schur, F. K., Hagen, W. J., de Marco, A., & Briggs, J. A. (2013). Determination of protein structure at 8.5Å resolution using cryo-electron tomography and sub-tomogram averaging. *Journal of Structural Biology*, *184*(3), 394–400.

Schwedler, U. K. von, Stray, K. M., Garrus, J. E., & Sundquist, W. I. (2003). Functional Surfaces of the Human Immunodeficiency Virus Type 1 Capsid Protein. *Journal of Virology*, *77*(9), 5439–5450.

Sinck, L., Richer, D., Howard, J., Alexander, M., Purcell, D. F. J., Marquet, R., & Paillart, J.-C. (2007). In vitro dimerization of human immunodeficiency virus type 1 (HIV-1) spliced RNAs. *RNA (New York, N.Y.)*, *13*(12), 2141–2150.

Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B., & Ehresmann, C. (1994). Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(11), 4945–4949.

Smyth, R. P., Despons, L., Huili, G., Bernacchi, S., Hijnen, M., Mak, J., Jossinet, F., Weixi, L., Paillart, J.-C., von Kleist, M., & Marquet, R. (2015). Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nature Methods*, *12*(9), 866–872.

Smyth, R. P., Smith, M. R., Jousset, A.-C., Despons, L., Laumond, G., Decoville, T., Cattenoz, P., Moog, C., Jossinet, F., Mougel, M., Paillart, J.-C., von Kleist, M., & Marquet, R. (2018). In cell mutational interference mapping experiment (in cell MIME) identifies the 5' polyadenylation signal as a dual regulator of HIV-1 genomic RNA production and packaging. *Nucleic Acids Research. 46*(9), e57. https://doi.org/10.1093/nar/gky152

Sommerfelt, M. A., Rhee, S. S., & Hunter, E. (1992). Importance of p12 protein in Mason-Pfizer monkey virus assembly and infectivity. *Journal of virology, 66*(12), 7005-7011.

Sonigo, P., Barker, C., Hunter, E., & Wain-Hobson, S. (1986). Nucleotide sequence of Mason-Pfizer monkey virus: An immunosuppressive D-type retrovirus. *Cell*, *45*(3), 375–385.

Soto-Rifo, R., Valiente-Echeverria, F., Rubilar, P. S., Garcia-de-Gracia, F., Ricci, E. P., Limousin, T., Décimo, D., Mouland, A. J., & Ohlmann, T. (2014). HIV-2 genomic RNA accumulates in stress granules in the absence of active translation. *Nucleic Acids Research*, *42*(20), 12861–12875.

Stephenson, J. D., Li, H., Kenyon, J. C., Symmons, M., Klenerman, D., & Lever, A. M. L. (2013). Three-Dimensional RNA Structure of the Major HIV-1 Packaging Signal Region. *Structure (London, England:1993)*, *21*(6), 951–962.

Stewart-Maynard, K. M., Cruceanu, M., Wang, F., Vo, M.-N., Gorelick, R. J., Williams, M. C., Rouzina, I., & Musier-Forsyth, K. (2008). Retroviral Nucleocapsid Proteins Display Nonequivalent Levels of Nucleic Acid Chaperone Activity. *Journal of Virology*, *82*(20), 10129–10142.

Stockley, P. G., White, S. J., Dykeman, E., Manfield, I., Rolfsson, O., Patel, N., Bingham, R., Barker, A., Wroblewski, E., Chandler-Bostock, R., Weiß, E. U., Ranson, N. A., Tuma, R., & Twarock, R. (2016). Bacteriophage MS2 genomic RNA encodes an assembly instruction manual for its capsid. *Bacteriophage*, *6*(1), e1157666. https://doi.org/10.1080/21597081.2016.1157666

Sun, M., Grigsby, I. F., Gorelick, R. J., Mansky, L. M., & Musier-Forsyth, K. (2014). Retrovirus-specific differences in matrix and nucleocapsid protein-nucleic acid interactions: Implications for genomic RNA packaging. *Journal of Virology*, *88*(2), 1271–1280.

Tanwar, H. S., Khoo, K. K., Garvey, M., Waddington, L., Leis, A., Hijnen, M., Velkov, T., Dumsday, G. J., McKinstry, W. J., & Mak, J. (2017). The thermodynamics of Pr55Gag-RNA interaction regulate the assembly of HIV. *PLOS Pathogens*, *13*(2), e1006221. https://doi.org/10.1371/journal.ppat.1006221

Temin, H. M., & Mizutani, S. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature*, *226*(5252), 1211–1213.

Urbaneja, M. A., McGrath, C. F., Kane, B. P., Henderson, L. E., & Casas-Finet, J. R. (2000). Nucleic acid binding properties of the simian immunodeficiency virus nucleocapsid protein NCp8. *The Journal of Biological Chemistry*, *275*(14), 10394–10404.

Van Bel, N., Das, A. T., Cornelissen, M., Abbink, T. E. M., & Berkhout, B. (2014). A Short Sequence Motif in the 5′ Leader of the HIV-1 Genome Modulates Extended RNA Dimer Formation and Virus Replication. *The Journal of Biological Chemistry*, *289*(51), 35061–35074.

Vile, R. G., Ali, M., Hunter, E., & Mcclure, M. O. (1992). Identification of a generalised packaging sequence for D-type retroviruses and generation of a D-type retroviral vector. *Virology*, *189*(2), 786–791.

Votteler, J., & Sundquist, W. I. (2013). Virus budding and the ESCRT pathway. *Cell Host & Microbe*, *14*(3), 232–241.

Webb, J. A., Jones, C. P., Parent, L. J., Rouzina, I., & Musier-Forsyth, K. (2013). Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: Implications for viral genomic RNA packaging. *RNA*, *19*(8), 1078–1088.

White, S. M., Renda, M., Nam, N.-Y., Klimatcheva, E., Zhu, Y., Fisk, J., Halterman, M., Rimel, B. J., Federoff, H., & Pandya, S. (1999). Lentivirus vectors using human and simian immunodeficiency virus elements. *Journal of Virology*, *73*(4), 2832–2840.

Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., Giddings, M. C., & Weeks, K. M. (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology*, *6*(4), e96. https://doi.org/10.1371/journal.pbio.0060096

Wu, H., Mitra, M., McCauley, M. J., Thomas, J. A., Rouzina, I., Musier-Forsyth, K., Williams, M. C., & Gorelick, R. J. (2013). Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication. *Virus Research*, *171*(2), 263–277.

Wu, W., Hatterschide, J., Syu, Y.-C., Cantara, W. A., Blower, R. J., Hanson, H. M., Mansky, L. M., & Musier-Forsyth, K. (2018). Human T-cell leukemia virus type 1 Gag domains have distinct RNA-binding specificities with implications for RNA packaging and dimerization. *The Journal of Biological Chemistry*, *293*(42), 16261–16276.

Yasuda, J., & Hunter, E. (1998). A Proline-Rich Motif (PPPY) in the Gag Polyprotein of Mason-Pfizer Monkey Virus Plays a Maturation-Independent Role in Virion Release. *Journal of Virology*, *72*(5), 4095–4103.

Yin, P. D., & Hu, W.-S. (1997). RNAs from genetically distinct retroviruses can copackage and exchange genetic information in vivo. *Journal of Virology*, *71*(8), 6237–6242.

Yu, E. T., Hawkins, A., Eaton, J., & Fabris, D. (2008). MS3D structural elucidation of the HIV-1 packaging signal. *Proceedings of the National Academy of Sciences*, *105*(34), 12248–12253.

Zeffman, A., Hassard, S., Varani, G., & Lever, A. (2000). The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein. *Journal of Molecular Biology*, *297*(4), 877–893.

Zhang, G., Sharon, D., Jovel, J., Liu, L., Wine, E., Tahbaz, N., Indik, S., & Mason, A. (2015). Pericentriolar Targeting of the Mouse Mammary Tumor Virus GAG Protein. *PLoS ONE*, *10*(6), e0131515. https://doi.org/10.1371/journal.pone.0131515

Zhang, H., Dornadula, G., Orenstein, J., & Pomerantz, R. J. (2000). Morphologic changes in human immunodeficiency virus type 1 virions secondary to intravirion reverse transcription: Evidence indicating that reverse transcription may not take place within the intact viral core. *Journal of Human Virology*, *3*(3), 165–172.

Zhang, Y., & Barklis, E. (1997). Effects of nucleocapsid mutations on human immunodeficiency virus assembly and RNA encapsidation. *Journal of Virology*, *71*(9), 6765–6776.

Zhao, Y., Low, W., & Collins, M. K. L. (2000). Improved safety and titre of murine leukaemia virus (MLV)-based retroviral vectors. *Gene Therapy*, *7*(4), 300–305.

Zhou, J., Bean, R. L., Vogt, V. M., & Summers, M. F. (2007). Solution Structure of the Rous Sarcoma Virus Nucleocapsid Protein:μΨ RNA Packaging Signal Complex. *Journal of Molecular Biology*, *365*(2), 453–467.

Zhou, J., McAllen, J. K., Tailor, Y., & Summers, M. F. (2005). High affinity nucleocapsid protein binding to the muPsi RNA packaging signal of Rous sarcoma virus. *Journal of Molecular Biology*, *349*(5), 976–988.

Zila, V., Margiotta, E., Turonova, B., Müller, T. G., Zimmerli, C. E., Mattei, S., Allegretti, M., Börner, K., Rada, J., Müller, B., Lusic, M., Kräusslich, H.-G., & Beck, M. (2020). Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *BioRxiv*, 2020.07.30.193524. https://doi.org/10.1101/2020.07.30.193524

Zolotukhin, A. S., Michalowski, D., Smulevitch, S., & Felber, B. K. (2001). Retroviral constitutive transport element evolved from cellular TAP (NXF1)-binding sequences. *Journal of Virology*, *75*(12), 5567–5575.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415.

List of Publications

1. Pitchai, F. N. N., Ali, L. M., Pillai, V. N., Chameettachal, A., Ashraf, S. S., Mustafa, F., Marquet, R., Rizvi, T. A. (2018). Expression, purification, and characterization of biologically active full-length Mason-Pfizer monkey virus (MPMV) Pr78[Gag]. Scientific Reports, 8, 11793. https://doi.org/10.1038/s41598-018-30142-0

2. Pitchai, F. N. N., Chameettachal, A., Vivet-Boudou, V., Ali, L. M., Pillai, V. N., Krishnan, A., Bernacchi, S., Mustafa, F., Marquet, R., Rizvi, T. A. (2021). Identification of Pr78[Gag] binding sites on the Mason-Pfizer monkey virus genomic RNA packaging determinants. Journal of Molecular Biology, Mar 11;433(10):166923. doi: 10.1016/j.jmb.2021.166923.

Appendices

Appendix A

Description of primers and DNA templates used for cloning and sequencing.

Oligo Name	*S or AS	Clone Names	Oligo Sequence	DNA template used for SOE PCR	**Nucleotide (nt) Position and/or Reference	Virus Region & Gene
OTR 1004	S	Outer primers for construction of *In vitro* transcription clones	5' CCC aag ctt **AAT ACG ACT CAC TAT AGG** GCC ACC ATT AAA TGA GAC TTG ATC 3'	These oligos were used as outer primers to generate each *in vitro* transcribing clones	3 dummies, *Hind*III , T7 promoter sequence, MPMV nt 622-645	MPMV R
OTR 1005	AS		5' AAA ccc ggg TTC TTT CTT ATC TAT CAA TTC 3'		3 dummies, *Xma*I, MPMV nt 1171-1151	MPMV *gag*
OTR 1378	S	**FN42** (*env* RNA)	5' ATA TGA ACT TCA ATT ATC ATT TCA TCT 3'	KAL01	MPMV nt 6240-6266	MPMV *env*
OTR 1379	AS		5' AAA ccc ggg CGT GTT GCA TTC ACC GGG ACA 3'	KAL01	3 dummies, *Xma*1, MPMV nt 6538- 6522	MPMV *env*
OTR 1458	AS		5' GAA ATG ATA ATT GAA GTT CAT ATT TTC AGG TCC AAC GCG GCA GGT 3'	SJ2	MPMV *env* start to SA (nt 6262- 6240); mSD & upstream (nt 852-869); 852/6240 in this primer forms the mSD/SA junction	MPMV U5

Appendix A

Description of primers and DNA templates used for cloning and sequencing (continued).

Oligo Name	*S or AS	Clone Names	Oligo Sequence	DNA template used for SOE PCR	**Nucleotide (nt) Position and/or Reference	Virus Region & Gene
OTR 1006	S	**FN15**	5' GGA CCT GTG TTG CGC TCG GAT ATG GG 3'	FN26	MPMV nt 861-894 with deletion from nt 867-874	MPMV U5
OTR 1007	AS		5' CAA CAC AGG TCC AAC GCG GCA GGT TC 3'		MPMV nt 880-846 with deletion from nt 874-867	MPMV U5
OTR 1006	S	**FN16**	5' GGA CCT GTG TTG CGC TCG GAT ATG GG 3'	SJ2	MPMV nt 861-894 with deletion from nt 867-874	MPMV U5
OTR 1007	AS		5' CAA CAC AGG TCC AAC GCG GCA GGT TC 3'		MPMV nt 880-846 with deletion from nt 874-867	MPMV U5
OTR 1012	S	**FN26**	5' GGC CGG CGA ACT CTC TTG GCC GCC GCG GG 3'	SJ2	MPMV nt 803-847 with deletion from nt 812-827	MPMV U5
OTR 1013	AS		5' GCC AAG AGA GTT CGC CGG CCG GCG AAC GC 3'		MPMV nt 838-794 with deletion from nt 827-812	MPMV U5
OTR 1139	S	**FN30**	5' GAA AGT AAT TGG CCG CCG CGG GAA C 3'	SJ2	MPMV nt 820-850 with deletion from nt 828-833	MPMV U5
OTR 1140	AS		5' GGC CAA TTA CTT TCA CTT TTA ATC GCC G 3'		MPMV nt 839-806 with deletion from nt 833-828	MPMV U5

Appendix A

Description of primers and DNA templates used for cloning and sequencing (continued).

Oligo Name	*S or AS	Clone Names	Oligo Sequence	DNA template used for SOE PCR	**Nucleotide (nt) Position and/or Reference	Virus Region & Gene
OTR 121	S	**Used for sequencing**	5' GGT TAA ATA TGC TGA TCT TT 3'	NA	MPMV nt 964-983	MPMV *gag*
OTR 198	S		5' ccc gct agc CGC CTA CTC TAC GCC 3'	NA	MPMV nt 1070-1056	MPMV *gag*
OTR 18-VIC/19-NED	AS	**Used in hSHAPE of**	5' AGT TAC TGG GAC TTT CTC CG 3'	NA	MPMV nt 1104-1123/857-875	MPMV U5/*gag*
OTR 22-VIC/23-NED	AS	**RCR001 & FN26**	5' CTT ACT TTC AGG TCC AAC CG 3'	NA	MPMV nt 1104-1123/857-875	MPMV U5/*gag*
OTR_312-322	AS	**Used in hSHAPE of FN42**	5' CAA GCC GGT TTT GGA GAT CCG 3'	NA	MPMV nt 6302-6322	MPMV *env*
OTR_497-518	AS	**Used in hSHAPE of FN42**	5' CTC CCA CTA CCC TAG CAA TAC 3'	NA	MPMV nt 6482-6503	MPMV *env*

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

Sequence in bold: T7 promoter.

*S, sense; AS, antisense.

** The MPMV nucleotide numbering system refers to the genome sequence deposited in the Genbank (accession number M12349.1; Sonigo et al. 1986)

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^{Gag}

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		*p*-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
G	1	ND	ND	ND	-	-
C	2	ND	ND	ND	-	-
C	3	ND	ND	ND	-	-
A	4	ND	ND	ND	-	-
C	5	ND	ND	ND	-	-
C	6	ND	ND	ND	-	-
A	7	ND	ND	ND	-	-
U	8	ND	ND	ND	-	-
U	9	ND	ND	ND	-	-
A	10	0.42	0.49	ND	-	-
A	11	0.14	0.14	ND	-	-
A	12	0.17	0.29	ND	-	-
U	13	0.18	0.20	ND	-	-
G	14	0.05	0.09	ND	-	-
A	15	0.01	0.02	ND	-	-
G	16	0.08	0.14	ND	-	-
A	17	0.18	0.16	ND	-	-
C	18	0.39	0.29	ND	-	-
U	19	0.80	0.68	ND	-	-
U	20	1.28	1.04	ND	-	-
G	21	1.37	1.09	ND	-	-
A	22	0.53	0.07	ND	-	-
U	23	0.90	0.31	ND	-	-
C	24	0.66	0.36	ND	-	-
A	25	1.34	1.14	ND	-	-
G	26	1.39	1.06	ND	-	-
A	27	1.01	1.00	1.07	0.92	**0.96**
A	28	1.81	0.33	1.05	0.94	**0.32**
C	29	0.61	0.72	0.74	0.78	**0.87**
A	30	0.44	0.60	1.12	0.86	**0.39**
C	31	0.05	0.06	0.09	0.10	**0.26**
U	32	0.03	0.05	0.04	0.05	**0.85**
G	33	0.03	0.05	0.07	0.06	**0.01**
U	34	0.00	0.00	0.00	0.01	**0.30**

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
C	35	0.03	0.03	0.08	0.10	**0.60**
U	36	0.13	0.12	0.00	0.00	**0.20**
U	37	0.06	0.09	0.00	0.00	**0.36**
G	38	0.07	0.12	0.15	0.08	**0.09**
U	39	0.05	0.07	0.07	0.09	**0.19**
C	40	0.02	0.03	0.07	0.07	**0.18**
U	41	0.17	0.21	0.19	0.24	**0.94**
C	42	1.43	0.53	1.73	0.28	**0.41**
C	43	1.11	1.00	1.43	0.50	**0.65**
A	44	0.30	0.36	0.75	0.45	**0.41**
U	45	0.37	0.21	0.47	0.10	**0.50**
U	46	0.29	0.20	0.29	0.15	**0.96**
U	47	0.29	0.41	0.40	0.15	**0.72**
C	48	0.08	0.12	0.00	0.01	**0.40**
U	49	0.17	0.17	0.41	0.14	**0.01**
U	50	0.24	0.10	0.15	0.16	**0.46**
G	51	0.04	0.07	0.31	0.13	**0.08**
U	52	0.11	0.12	0.17	0.07	**0.62**
G	53	0.04	0.08	0.03	0.02	**0.76**
U	54	0.01	0.02	0.19	0.13	**0.17**
C	55	0.07	0.11	0.18	0.22	**0.25**
U	56	0.05	0.09	0.57	0.21	**0.09**
C	57	0.26	0.23	0.26	0.20	**0.98**
U	58	0.63	0.25	0.60	0.58	**0.95**
U	59	0.63	0.58	0.49	0.21	**0.79**
G	60	0.59	0.28	0.31	0.05	**0.24**
U	61	0.25	0.09	0.29	0.22	**0.83**
U	62	0.75	0.25	0.78	0.25	**0.90**
C	63	0.06	0.10	0.18	0.07	**0.34**
C	64	0.00	0.00	0.24	0.30	**0.30**
C	65	0.01	0.02	0.60	0.75	**0.31**
U	66	0.16	0.28	0.93	0.32	**0.16**
U	67	0.42	0.55	1.06	0.82	**0.47**
C	68	1.33	1.15	1.67	0.58	**0.74**
A	69	0.68	0.62	0.77	0.28	**0.88**
A	70	0.52	0.31	0.66	0.26	**0.71**

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
U	71	0.16	0.14	0.23	0.17	0.38
U	72	0.08	0.06	0.03	0.04	0.51
C	73	0.03	0.03	0.03	0.03	0.68
C	74	0.04	0.06	0.02	0.02	0.68
C	75	0.62	0.80	0.36	0.15	0.67
A	76	0.19	0.09	0.21	0.22	0.89
C	77	0.10	0.04	0.05	0.03	0.26
U	78	0.16	0.07	0.03	0.04	0.20
C	79	0.01	0.02	0.03	0.03	0.52
C	80	0.00	0.00	0.00	0.01	0.30
C	81	0.00	0.01	0.02	0.02	0.46
U	82	0.02	0.03	0.09	0.06	0.31
C	83	0.03	0.05	0.04	0.03	0.90
C	84	0.09	0.08	0.09	0.06	0.90
U	85	0.16	0.18	0.16	0.05	0.97
C	86	0.13	0.11	0.05	0.06	0.27
C	87	0.34	0.42	0.45	0.27	0.77
A	88	0.42	0.37	0.39	0.26	0.84
G	89	0.72	0.44	0.52	0.17	0.41
G	90	0.85	0.40	0.70	0.12	0.53
U	91	1.61	0.66	1.02	0.20	0.28
U	92	1.56	0.64	0.99	0.21	0.27
C	93	0.04	0.04	0.03	0.03	0.69
C	94	0.01	0.01	0.06	0.02	0.10
U	95	0.09	0.08	0.08	0.10	0.85
A	96	0.02	0.03	0.02	0.02	0.18
C	97	0.14	0.12	0.16	0.03	0.78
U	98	1.76	0.42	1.52	0.38	0.42
G	99	0.20	0.05	0.18	0.04	0.44
U	100	0.36	0.11	0.27	0.26	0.71
U	101	1.18	0.40	0.59	0.08	0.13
G	102	1.48	0.59	0.48	0.09	0.11
A	103	1.08	0.99	0.77	0.64	0.77
U	104	0.86	0.98	0.49	0.43	0.67
C	105	0.09	0.15	0.04	0.05	0.66
C	106	0.02	0.04	0.03	0.02	0.97

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
C	107	0.08	0.14	0.02	0.03	0.57
G	108	0.31	0.45	0.21	0.18	0.62
C	109	1.00	0.90	0.56	0.31	0.36
G	110	1.92	0.14	1.64	0.29	0.15
G	111	1.31	0.68	0.97	0.02	0.46
G	112	1.03	0.87	0.31	0.05	0.28
U	113	1.23	1.07	1.29	0.10	0.93
C	114	0.00	0.00	0.04	0.02	0.05
G	115	0.03	0.05	0.13	0.10	0.35
G	116	0.00	0.00	0.12	0.11	0.20
G	117	0.00	0.00	0.08	0.10	0.30
A	118	0.02	0.03	0.10	0.11	0.38
C	119	1.21	0.93	0.70	0.87	0.67
A	120	0.89	0.68	0.94	0.30	0.94
G	121	0.41	0.43	0.15	0.12	0.28
U	122	0.33	0.30	0.13	0.10	0.42
U	123	0.34	0.23	0.27	0.08	0.71
G	124	0.05	0.05	0.01	0.02	0.44
G	125	0.09	0.09	0.01	0.02	0.32
C	126	0.12	0.06	0.11	0.14	0.84
G	127	0.07	0.11	0.00	0.00	0.39
C	128	0.00	0.00	0.03	0.02	0.06
C	129	0.00	0.00	0.23	0.21	0.19
C	130	0.79	1.04	0.50	0.62	0.76
A	131	0.37	0.26	0.13	0.08	0.32
A	132	0.36	0.37	0.24	0.06	0.56
C	133	0.14	0.06	0.03	0.03	0.09
G	134	0.10	0.09	0.01	0.01	0.27
U	135	0.09	0.02	0.06	0.02	0.30
G	136	0.00	0.00	0.02	0.02	0.30
G	137	0.00	0.00	0.04	0.05	0.29
G	138	0.14	0.24	0.55	0.69	0.50
G	139	0.04	0.06	0.25	0.08	0.08
C	140	0.01	0.01	0.19	0.18	0.22
U	141	0.00	0.00	0.04	0.04	0.19
G	142	0.00	0.00	0.01	0.01	0.30

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
G	143	0.01	0.01	0.01	0.01	0.74
A	144	0.05	0.09	0.12	0.13	0.58
U	145	0.73	0.25	0.21	0.15	0.15
A	146	0.41	0.43	0.52	0.24	0.79
C	147	0.20	0.21	0.35	0.21	0.56
G	148	0.03	0.03	0.00	0.00	0.19
A	149	0.00	0.00	0.05	0.06	0.30
G	150	0.00	0.00	0.00	0.00	ND
G	151	0.26	0.46	0.83	1.04	0.51
G	152	0.52	0.69	1.82	0.31	0.15
A	153	0.85	0.68	1.19	0.50	0.66
A	154	1.61	0.46	0.88	0.16	0.12
U	155	1.66	0.32	1.22	0.25	0.30
U	156	1.08	0.66	0.70	0.17	0.48
U	157	1.33	1.15	1.06	0.44	0.79
C	158	0.50	0.50	0.19	0.13	0.43
G	159	0.13	0.12	0.05	0.06	0.53
U	160	0.31	0.11	0.16	0.18	0.13
G	161	0.80	0.93	0.27	0.23	0.39
A	162	0.95	0.85	0.60	0.31	0.64
G	163	0.01	0.02	0.00	0.00	0.42
G	164	0.03	0.04	0.02	0.04	0.88
A	165	0.16	0.20	0.01	0.02	0.34
A	166	0.25	0.14	0.06	0.07	0.17
G	167	0.34	0.19	0.04	0.02	0.13
A	168	0.44	0.39	0.40	0.19	0.90
C	169	0.92	0.95	1.27	0.44	0.70
G	170	0.36	0.10	0.09	0.06	0.04
A	171	0.11	0.14	0.16	0.17	0.77
C	172	0.38	0.43	0.39	0.15	0.98
G	173	0.07	0.07	0.33	0.34	0.35
C	174	0.11	0.11	0.03	0.03	0.39
G	175	0.27	0.35	0.38	0.13	0.74
U	176	0.19	0.32	0.06	0.07	0.60
U	177	0.01	0.01	0.04	0.04	0.23
C	178	0.13	0.14	0.05	0.05	0.53

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		*p*-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
G	179	0.04	0.04	0.10	0.03	0.07
C	180	0.00	0.00	0.12	0.08	0.11
C	181	0.00	0.00	0.01	0.01	0.42
G	182	0.08	0.14	0.06	0.07	0.85
G	183	0.08	0.14	0.13	0.16	0.80
C	184	0.05	0.09	0.14	0.17	0.57
C	185	0.08	0.14	0.02	0.02	0.55
G	186	0.00	0.00	0.04	0.03	0.18
G	187	0.01	0.01	0.50	0.17	0.04
C	188	0.12	0.19	1.29	0.64	0.12
G	189	0.00	0.00	0.11	0.04	0.05
A	190	0.29	0.05	0.26	0.05	0.55
*U	191	0.38	0.13	0.13	0.06	0.15
U	192	0.98	1.00	0.90	0.69	0.94
A	193	0.41	0.05	0.07	0.06	0.02
A	194	0.30	0.08	0.02	0.02	0.02
A	195	0.16	0.08	0.00	0.00	0.06
A	196	0.22	0.03	0.06	0.06	0.04
G	197	0.45	0.04	0.18	0.16	0.12
U	198	0.38	0.05	0.08	0.05	0.03
G	199	0.46	0.10	0.09	0.08	0.06
A	200	0.51	0.04	0.09	0.02	0.00
A	201	0.45	0.12	0.11	0.01	0.04
A	202	0.70	0.44	0.10	0.09	0.15
G	203	1.04	0.83	0.30	0.20	0.29
U	204	0.21	0.37	1.44	0.57	0.09
A	205	0.72	0.06	0.26	0.25	0.08
A	206	0.61	0.53	0.51	0.38	0.85
A	207	0.20	0.02	0.18	0.16	0.86
C	208	0.02	0.04	0.14	0.12	0.15
U	209	0.05	0.00	0.04	0.04	0.80
C	210	0.04	0.05	0.03	0.04	0.86
U	211	0.02	0.00	0.08	0.03	0.07
C	212	0.04	0.06	0.11	0.04	0.33
U	213	0.20	0.08	0.52	0.18	0.14
U	214	0.40	0.02	0.25	0.05	0.06

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78[Gag] (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		*p*-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
G	215	0.06	0.04	0.03	0.02	0.45
G	216	0.05	0.00	0.02	0.03	0.18
C	217	0.06	0.03	0.15	0.06	0.22
C	218	0.01	0.02	0.00	0.00	0.27
G	219	0.03	0.03	0.00	0.00	0.23
C	220	0.05	0.02	0.14	0.03	0.07
C	221	0.00	0.00	0.00	0.01	0.30
G	222	0.00	0.00	0.01	0.01	0.18
C	223	0.00	0.00	0.01	0.01	0.42
G	224	0.00	0.00	0.00	0.00	-
G	225	0.00	0.00	0.00	0.00	-
G	226	0.75	0.19	0.65	0.09	0.59
A	227	2.00	0.00	1.83	0.29	0.42
A	228	2.00	0.00	1.73	0.28	0.24
C	229	0.23	0.08	0.19	0.10	0.73
C	230	0.72	0.06	0.65	0.11	0.49
U	231	2.00	0.00	1.74	0.28	0.25
G	232	1.88	0.21	1.16	0.21	0.02
C	233	0.03	0.02	0.04	0.01	0.27
C	234	0.04	0.07	0.00	0.00	0.42
G	235	0.00	0.00	0.00	0.00	-
C	236	0.00	0.00	0.00	0.00	-
G	237	0.08	0.13	0.12	0.03	0.65
U	238	0.14	0.10	0.40	0.05	0.02
U	239	0.07	0.06	0.00	0.00	0.20
G	240	0.00	0.00	0.00	0.00	-
G	241	0.04	0.00	0.02	0.02	0.19
A	242	2.00	0.00	1.79	0.33	0.38
C	243	0.10	0.02	0.07	0.03	0.46
C	244	0.01	0.02	0.09	0.03	0.04
U	245	0.04	0.05	0.00	0.00	0.32
**G	246	0.04	0.05	0.00	0.00	0.28
A	247	0.13	0.08	0.07	0.08	0.51
A	248	0.06	0.07	0.17	0.29	0.65
A	249	0.04	0.05	0.17	0.29	0.56
G	250	0.10	0.04	0.17	0.04	0.25

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
U	251	0.78	0.86	1.50	0.41	0.41
A	252	0.82	0.09	0.67	0.10	0.03
A	253	0.94	0.15	0.61	0.12	0.00
***G	254	0.92	0.09	0.50	0.16	0.05
U	255	1.23	0.22	0.94	0.30	0.41
G	256	0.73	0.15	0.33	0.10	0.03
U	257	1.23	0.25	0.78	0.16	0.14
U	258	0.74	0.03	0.33	0.05	0.01
G	259	0.07	0.01	0.01	0.01	0.03
C	260	0.08	0.14	0.22	0.05	0.33
G	261	0.15	0.08	0.11	0.02	0.37
C	262	0.05	0.09	0.54	0.14	0.06
U	263	0.07	0.08	0.47	0.11	0.06
C	264	0.28	0.36	0.83	0.30	0.29
G	265	0.26	0.07	0.22	0.07	0.24
G	266	0.22	0.08	0.31	0.04	0.24
A	267	1.30	0.19	1.65	0.33	0.32
U	268	1.21	0.75	1.70	0.52	0.57
A	269	0.66	0.17	0.24	0.02	0.06
U	270	0.64	0.22	0.13	0.11	0.12
G	271	0.00	0.00	0.11	0.14	0.30
G	272	0.00	0.00	0.00	0.00	-
G	273	0.00	0.00	0.00	0.00	-
G	274	0.00	0.00	0.01	0.02	0.42
C	275	0.03	0.05	0.24	0.09	0.11
A	276	0.00	0.00	0.06	0.06	0.22
A	277	0.02	0.03	0.03	0.03	0.63
G	278	0.04	0.05	0.03	0.02	0.76
A	279	0.46	0.18	0.27	0.23	0.25
A	280	0.62	0.14	0.55	0.10	0.35
U	281	0.97	0.16	0.62	0.21	0.04
U	282	0.56	0.12	0.64	0.21	0.71
A	283	0.06	0.02	0.38	0.44	0.35
A	284	0.05	0.02	0.05	0.06	0.96
G	285	0.03	0.02	0.00	0.00	0.08
C	286	0.08	0.07	0.08	0.02	0.95

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^Gag (continued)

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78$^{\text{Gag}}$ (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		*p*-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
C	287	0.11	0.09	0.05	0.06	0.50
A	288	0.05	0.04	0.02	0.01	0.40
G	289	0.06	0.06	0.04	0.03	0.44
C	290	0.39	0.48	0.59	0.21	0.67
A	291	0.60	0.54	1.08	0.18	0.36
U	292	0.79	0.22	0.62	0.01	0.28
G	293	0.65	0.38	0.36	0.05	0.27
A	294	0.50	0.15	0.42	0.07	0.21
A	295	0.41	0.09	0.40	0.07	0.64
C	296	0.30	0.26	0.31	0.05	0.91
G	297	0.30	0.18	0.18	0.04	0.42
U	298	0.19	0.07	0.32	0.08	0.06
U	299	0.30	0.23	0.26	0.13	0.64
A	300	0.94	0.79	1.14	0.21	0.69
U	301	1.39	0.66	0.37	0.06	0.12
G	302	0.95	0.86	0.98	0.16	0.95
U	303	1.67	0.33	1.02	0.22	0.01
A	304	1.81	0.17	1.72	0.33	0.67
G	305	1.73	0.47	1.05	0.25	0.13
A	306	1.02	0.69	0.52	0.13	0.38
A	307	0.32	0.10	0.60	0.17	0.17
C	308	0.44	0.55	1.61	0.52	0.06
A	309	0.81	1.05	0.62	0.09	0.79
A	310	0.36	0.35	0.47	0.09	0.69
U	311	0.60	0.24	0.70	0.09	0.58
U	312	0.85	0.23	0.76	0.12	0.38
G	313	0.89	0.14	0.59	0.09	0.06
A	314	0.74	0.02	0.80	0.15	0.60
A	315	0.90	0.27	0.55	0.18	0.22
G	316	0.50	0.43	0.07	0.10	0.20
C	317	0.21	0.36	0.92	0.33	0.10
A	318	0.48	0.60	0.15	0.07	0.46
G	319	0.11	0.19	0.17	0.29	0.42
G	320	0.00	0.00	0.02	0.04	0.42
C	321	0.10	0.17	0.00	0.00	0.42
U	322	0.17	0.20	0.30	0.11	0.15

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		*p*-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
U	323	0.28	0.36	0.30	0.05	**0.94**
U	324	0.42	0.54	0.93	0.29	**0.09**
A	325	1.07	0.51	0.88	0.28	**0.70**
A	326	0.87	0.43	0.73	0.15	**0.71**
A	327	0.91	0.31	0.63	0.10	**0.25**
G	328	0.54	0.48	0.09	0.12	**0.20**
A	329	0.06	0.10	0.35	0.01	**0.04**
C	330	0.17	0.22	1.40	0.50	**0.04**
A	331	0.62	0.97	0.06	0.02	**0.42**
C	332	0.08	0.14	0.02	0.04	**0.42**
G	333	0.05	0.04	0.00	0.00	**0.14**
G	334	0.06	0.06	0.00	0.00	**0.24**
G	335	0.05	0.05	0.00	0.00	**0.21**
G	336	0.04	0.06	0.00	0.00	**0.46**
A	337	0.13	0.06	0.06	0.03	**0.11**
G	338	0.06	0.02	0.04	0.02	**0.29**
U	339	0.49	0.60	0.89	0.35	**0.17**
A	340	1.02	0.64	0.08	0.03	**0.13**
A	341	0.00	0.00	0.08	0.03	**0.04**
A	342	0.26	0.15	0.03	0.03	**0.14**
G	343	0.37	0.20	0.01	0.02	**0.10**
G	344	0.00	0.01	0.00	0.00	**0.88**
U	345	0.11	0.20	0.36	0.01	**0.15**
U	346	0.55	0.43	0.61	0.16	**0.78**
A	347	1.07	0.32	0.44	0.12	**0.13**
A	348	0.38	0.16	0.40	0.12	**0.81**
A	349	0.55	0.12	0.43	0.12	**0.23**
U	350	0.68	0.20	0.33	0.30	**0.11**
A	351	0.89	0.28	0.31	0.11	**0.10**
U	352	0.30	0.26	0.33	0.08	**0.84**
G	353	0.66	0.32	0.32	0.07	**0.14**
C	354	0.50	0.26	0.33	0.07	**0.30**
U	355	0.29	0.14	0.20	0.07	**0.49**
G	356	0.27	0.19	0.35	0.05	**0.47**
A	357	0.23	0.22	0.96	0.25	**0.01**
U	358	0.21	0.24	0.15	0.08	**0.61**

Appendix B

Mean SHAPE reactivity data of the wild type gRNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
C	359	0.00	0.00	0.07	0.11	**0.42**
U	360	0.21	0.27	0.47	0.01	**0.24**
U	361	0.18	0.10	0.31	0.04	**0.14**
U	362	0.17	0.04	0.30	0.05	**0.06**
U	363	0.35	0.12	0.37	0.08	**0.75**
G	364	0.51	0.13	0.36	0.05	**0.24**
A	365	0.63	0.13	0.79	0.21	**0.16**
A	366	0.50	0.34	0.22	0.06	**0.28**
A	367	0.23	0.35	0.08	0.02	**0.54**
U	368	0.10	0.08	0.05	0.04	**0.58**
U	369	0.26	0.32	0.10	0.14	**0.55**
U	370	0.25	0.11	0.23	0.05	**0.64**
U	371	0.26	0.45	0.23	0.02	**0.92**
U	372	0.17	0.15	0.26	0.15	**0.37**
U	373	0.41	0.17	0.34	0.04	**0.55**
G	374	0.45	0.17	0.22	0.12	**0.07**
A	375	0.38	0.35	0.28	0.06	**0.72**
U	376	0.20	0.17	0.12	0.08	**0.32**
U	377	0.14	0.12	0.07	0.05	**0.44**
U	378	0.12	0.09	0.14	0.15	**0.88**
U	379	0.13	0.07	0.13	0.04	**0.98**
G	380	0.12	0.09	0.10	0.04	**0.62**
U	381	0.14	0.13	0.12	0.03	**0.78**
G	382	0.23	0.23	0.08	0.07	**0.25**
A	383	0.13	0.17	0.09	0.09	**0.82**
A	384	0.09	0.05	0.17	0.02	**0.12**
G	385	0.16	0.11	0.14	0.02	**0.69**
G	386	0.30	0.18	0.34	0.07	**0.71**
A	387	0.69	0.66	ND	-	-
U	388	ND	-	ND	-	-

Appendix B. Mean hSHAPE reactivity data from triplicate hSHAPE experiments on wild type unspliced RNA (RCR001) both in the absence as well presence of Pr78^{Gag} with standard deviation (SD) and p-values. Color key: *green, single- stranded purines (ssPurines); **blue, base-paired purines (bpPurines); ***brick red, GU rich region. ND; no data

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag}

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
G	1	ND	-	ND	-	-
C	2	ND	-	ND	-	-
C	3	ND	-	ND	-	-
A	4	ND	-	ND	-	-
C	5	ND	-	ND	-	-
C	6	ND	-	ND	-	-
A	7	ND	-	ND	-	-
U	8	ND	-	ND	-	-
U	9	ND	-	ND	-	-
A	10	ND	-	ND	-	-
A	11	ND	-	ND	-	-
A	12	ND	-	ND	-	-
U	13	ND	-	ND	-	-
G	14	ND	-	ND	-	-
A	15	ND	-	ND	-	-
G	16	ND	-	ND	-	-
A	17	ND	-	ND	-	-
C	18	ND	-	ND	-	-
U	19	ND	-	ND	-	-
U	20	ND	-	ND	-	-
G	21	ND	-	ND	-	-
A	22	ND	-	ND	-	-
U	23	ND	-	ND	-	-
C	24	ND	-	ND	-	-
A	25	MEAN	-	MEAN	-	-
G	26	ND	-	ND	-	-
A	27	ND	-	ND	-	-
A	28	ND	-	ND	-	-
C	29	ND	-	ND	-	-
A	30	ND	-	ND	-	-
C	31	ND	-	ND	-	-
U	32	ND	-	ND	-	-
G	33	ND	-	ND	-	-
U	34	ND	-	ND	-	-
C	35	ND	-	ND	-	-

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
U	36	ND	-	ND	-	-
U	37	ND	-	ND	-	-
G	38	ND	-	ND	-	-
U	39	ND	-	ND	-	-
C	40	ND	-	ND	-	-
U	41	ND	-	ND	-	-
C	42	ND	-	ND	-	-
C	43	ND	-	ND	-	-
A	44	ND	-	ND	-	-
U	45	ND	-	ND	-	-
U	46	ND	-	ND	-	-
U	47	ND	-	ND	-	-
C	48	ND	-	ND	-	-
U	49	ND	-	ND	-	-
U	50	ND	-	ND	-	-
G	51	ND	-	ND	-	-
U	52	ND	-	ND	-	-
G	53	ND	-	ND	-	-
U	54	ND	-	ND	-	-
C	55	ND	-	ND	-	-
U	56	ND	-	ND	-	-
C	57	ND	-	ND	-	-
U	58	ND	-	ND	-	-
U	59	ND	-	ND	-	-
G	60	ND	-	ND	-	-
U	61	ND	-	ND	-	-
U	62	ND	-	ND	-	-
C	63	ND	-	ND	-	-
C	64	ND	-	ND	-	-
C	65	ND	-	ND	-	-
U	66	ND	-	ND	-	-
U	67	ND	-	ND	-	-
C	68	ND	-	ND	-	-
A	69	ND	-	ND	-	-
A	70	ND	-	ND	-	-
U	71	ND	-	ND	-	-
U	72	ND	-	ND	-	-

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		*p*-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
C	73	ND	-	ND	-	-
C	74	ND	-	ND	-	-
C	75	ND	-	ND	-	-
A	76	ND	-	ND	-	-
C	77	ND	-	ND	-	-
U	78	ND	-	ND	-	-
C	79	ND	-	ND	-	-
C	80	ND	-	ND	-	-
C	81	ND	-	ND	-	-
U	82	ND	-	ND	-	-
C	83	ND	-	ND	-	-
C	84	ND	-	ND	-	-
U	85	ND	-	ND	-	-
C	86	ND	-	ND	-	-
C	87	ND	-	ND	-	-
A	88	ND	-	ND	-	-
G	89	ND	-	ND	-	-
G	90	ND	-	ND	-	-
U	91	ND	-	ND	-	-
U	92	ND	-	ND	-	-
C	93	ND	-	ND	-	-
C	94	ND	-	ND	-	-
U	95	ND	-	ND	-	-
A	96	ND	-	ND	-	-
C	97	ND	-	ND	-	-
U	98	ND	-	ND	-	-
G	99	ND	-	ND	-	-
U	100	ND	-	ND	-	-
U	101	ND	-	ND	-	-
G	102	ND	-	ND	-	-
A	103	ND	-	ND	-	-
U	104	ND	-	ND	-	-
C	105	ND	-	ND	-	-
C	106	ND	-	ND	-	-
C	107	ND	-	ND	-	-
G	108	ND	-	ND	-	-
C	109	ND	-	ND	-	-

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^Gag (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
G	110	ND	-	ND	-	-
G	111	ND	-	ND	-	-
G	112	ND	-	ND	-	-
U	113	ND	-	ND	-	-
C	114	ND	-	ND	-	-
G	115	ND	-	ND	-	-
G	116	ND	-	ND	-	-
G	117	ND	-	ND	-	-
A	118	ND	-	ND	-	-
C	119	ND	-	ND	-	-
A	120	ND	-	ND	-	-
G	121	ND	-	ND	-	-
U	122	ND	-	ND	-	-
U	123	ND	-	ND	-	-
G	124	ND	-	ND	-	-
G	125	ND	-	ND	-	-
C	126	ND	-	ND	-	-
G	127	ND	-	ND	-	-
C	128	ND	-	ND	-	-
C	129	ND	-	ND	-	-
C	130	ND	-	ND	-	-
A	131	ND	-	ND	-	-
A	132	ND	-	ND	-	-
C	133	ND	-	ND	-	-
G	134	ND	-	ND	-	-
U	135	ND	-	ND	-	-
G	136	ND	-	ND	-	-
G	137	ND	-	ND	-	-
G	138	ND	-	ND	-	-
G	139	ND	-	ND	-	-
C	140	ND	-	ND	-	-
U	141	ND	-	ND	-	-
G	142	ND	-	ND	-	-
G	143	ND	-	ND	-	-
A	144	ND	-	ND	-	-
U	145	ND	-	ND	-	-
A	146	ND	-	ND	-	-

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		_p_-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
C	147	ND	-	ND	-	-
G	148	ND	-	ND	-	-
A	149	ND	-	ND	-	-
G	150	ND	-	ND	-	-
G	151	ND	-	ND	-	-
G	152	ND	-	ND	-	-
A	153	ND	-	ND	-	-
A	154	ND	-	ND	-	-
U	155	ND	-	ND	-	-
U	156	ND	-	ND	-	-
U	157	ND	-	ND	-	-
C	158	ND	-	ND	-	-
G	159	ND	-	ND	-	-
U	160	ND	-	ND	-	-
G	161	ND	-	ND	-	-
A	162	ND	-	ND	-	-
G	163	ND	-	ND	-	**0.39**
G	164	0.05	0.05	ND	-	**0.13**
A	165	2.20	0.40	ND	-	**0.26**
A	166	0.08	0.07	ND	-	**0.32**
G	167	0.01	0.01	ND	-	**0.38**
A	168	0.03	0.05	ND	-	**0.34**
C	169	0.12	0.03	0.08	0.02	**0.04**
G	170	0.09	0.06	0.04	0.03	**0.04**
A	171	0.50	0.05	0.51	0.04	**0.87**
C	172	0.08	0.11	0.54	0.41	**0.14**
G	173	0.00	0.01	0.06	0.04	**0.10**
C	174	0.15	0.05	0.40	0.08	**0.05**
G	175	0.26	0.18	0.62	0.32	**0.23**
U	176	0.31	0.14	0.28	0.10	**0.17**
U	177	0.15	0.03	0.31	0.11	**0.05**
C	178	0.19	0.10	0.19	0.10	**0.49**
G	179	0.09	0.05	0.21	0.06	**0.01**
C	180	0.00	0.00	0.02	0.02	**0.19**
C	181	0.06	0.04	0.09	0.11	**0.61**
G	182	1.91	0.06	1.79	0.11	**0.11**
G	183	0.97	0.08	0.96	0.13	**0.69**

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
C	184	0.61	0.19	0.88	0.29	0.04
C	185	1.08	0.13	1.14	0.17	0.73
G	186	0.00	0.00	0.11	0.10	0.11
G	187	0.00	0.00	0.32	0.51	0.30
C	188	0.00	0.00	0.65	0.40	0.05
G	189	0.00	0.00	0.00	0.00	-
A	190	0.00	0.00	0.04	0.05	0.22
A	191	0.00	0.00	0.00	0.00	-
C	192	0.00	0.00	0.00	0.00	-
U	193	0.11	0.07	0.17	0.07	0.12
C	194	0.13	0.09	0.32	0.18	0.12
U	195	0.18	0.05	0.29	0.06	0.03
C	196	0.08	0.09	0.23	0.11	0.17
U	197	0.34	0.05	0.39	0.04	0.14
U	198	0.44	0.05	0.35	0.05	0.15
G	199	0.17	0.03	0.05	0.05	0.03
G	200	0.09	0.03	0.00	0.01	0.01
C	201	0.01	0.01	0.06	0.06	0.15
C	202	0.01	0.02	0.02	0.04	0.55
G	203	0.02	0.02	0.03	0.04	0.76
C	204	0.01	0.01	0.09	0.02	0.00
C	205	0.00	0.00	0.00	0.01	0.39
G	206	0.00	0.00	0.00	0.01	0.39
C	207	0.03	0.06	0.09	0.07	0.09
G	208	0.00	0.01	0.02	0.05	0.39
G	209	0.02	0.03	0.04	0.08	0.75
G	210	0.70	0.43	0.71	0.25	0.95
A	211	2.14	0.29	2.00	0.00	0.39
A	212	2.09	0.18	1.95	0.10	0.21
C	213	0.46	0.47	0.49	0.24	0.90
C	214	0.67	0.16	0.78	0.15	0.22
U	215	2.06	0.13	2.00	0.01	0.35
G	216	2.02	0.17	1.49	0.35	0.02
C	217	0.04	0.03	0.10	0.05	0.26
C	218	0.00	0.00	0.01	0.01	0.39
G	219	0.00	0.00	0.02	0.03	0.39
C	220	0.00	0.00	0.03	0.06	0.39

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
G	221	0.04	0.08	0.12	0.10	0.09
U	222	0.22	0.10	0.32	0.16	0.33
U	223	0.12	0.04	0.08	0.02	0.02
G	224	0.02	0.02	0.00	0.01	0.21
G	225	0.20	0.02	0.16	0.03	0.22
A	226	2.01	1.51	2.00	0.00	0.99
C	227	0.56	0.96	0.10	0.02	0.42
C	228	0.02	0.04	0.12	0.06	0.05
U	229	0.11	0.08	0.20	0.05	0.18
*G	230	0.20	0.08	0.08	0.05	0.02
A	231	0.34	0.09	0.24	0.09	0.06
A	232	0.30	0.08	0.19	0.08	0.13
A	233	0.17	0.04	0.16	0.05	0.16
G	234	0.24	0.03	0.32	0.05	0.05
U	235	0.63	0.31	1.56	0.53	0.04
A	236	0.94	0.09	0.96	0.04	0.95
A	237	0.66	0.05	0.54	0.03	0.04
**G	238	0.51	0.14	0.31	0.09	0.03
U	239	0.83	0.22	0.51	0.16	0.03
G	240	0.46	0.10	0.18	0.02	0.01
U	241	0.90	0.30	0.46	0.13	0.03
U	242	0.63	0.08	0.29	0.07	0.02
G	243	0.08	0.02	0.00	0.01	0.00
C	244	0.04	0.05	0.05	0.06	0.71
G	245	0.39	0.10	0.30	0.11	0.05
C	246	0.08	0.11	0.33	0.28	0.10
U	247	0.18	0.12	0.37	0.28	0.17
C	248	0.10	0.12	0.36	0.17	0.11
G	249	0.44	0.08	0.37	0.05	0.08
G	250	0.40	0.11	0.40	0.12	0.95
A	251	1.51	0.08	1.76	0.21	0.16
U	252	1.10	0.46	1.55	0.16	0.21
A	253	0.83	0.21	0.54	0.18	0.03
U	254	0.75	0.17	0.43	0.22	0.07
G	255	0.02	0.02	0.04	0.09	0.57
G	256	0.00	0.00	0.00	0.00	-
G	257	0.00	0.00	0.00	0.00	-

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
G	258	0.00	0.00	0.00	0.00	-
C	259	0.00	0.00	0.08	0.11	0.26
A	260	0.01	0.01	0.07	0.06	0.09
A	261	0.09	0.03	0.12	0.06	0.17
G	262	0.13	0.04	0.09	0.08	0.70
A	263	0.57	0.09	0.41	0.27	0.40
A	264	0.72	0.06	0.57	0.15	0.08
U	265	1.13	0.10	0.84	0.18	0.03
U	266	0.55	0.16	0.66	0.18	0.10
A	267	0.07	0.03	0.23	0.35	0.41
A	268	0.02	0.02	0.03	0.04	0.59
G	269	0.02	0.04	0.02	0.03	0.50
C	270	0.06	0.03	0.02	0.03	0.35
C	271	0.09	0.08	0.10	0.09	0.49
A	272	0.10	0.07	0.11	0.08	0.40
G	273	0.01	0.01	0.07	0.09	0.70
C	274	0.37	0.48	0.12	0.18	0.19
A	275	0.81	0.16	0.60	0.48	0.95
U	276	0.88	0.45	0.76	0.30	0.98
G	277	0.44	0.22	0.67	0.32	0.51
A	278	0.27	0.09	0.43	0.16	0.54
A	279	0.33	0.04	0.37	0.11	0.50
C	280	0.29	0.12	0.39	0.09	0.24
G	281	0.20	0.07	0.17	0.17	0.63
U	282	0.25	0.07	0.31	0.11	0.56
U	283	0.70	0.67	0.27	0.09	0.38
A	284	1.49	0.63	0.74	0.59	0.45
U	285	0.83	0.70	0.54	0.51	0.35
G	286	1.65	0.18	0.92	0.51	0.38
U	287	1.74	0.24	0.94	0.47	0.05
A	288	1.87	0.21	1.46	0.53	0.27
G	289	1.07	0.59	1.20	0.33	0.65
A	290	0.35	0.23	0.91	0.72	0.39
A	291	0.12	0.05	0.51	0.27	0.06
C	292	0.44	0.20	0.67	0.61	0.50
A	293	0.52	0.09	0.78	0.50	0.13
A	294	0.48	0.40	0.67	0.30	0.61

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
U	295	0.89	0.20	0.72	0.31	**0.82**
U	296	0.95	0.13	0.82	0.30	**0.70**
G	297	0.86	0.28	0.81	0.17	**0.55**
A	298	0.90	0.15	0.93	0.04	**0.26**
A	299	0.50	0.34	0.72	0.19	**0.61**
G	300	0.07	0.15	0.31	0.36	**0.69**
C	301	0.23	0.13	0.38	0.26	**0.38**
A	302	0.18	0.17	0.41	0.32	**0.07**
G	303	0.00	0.00	0.17	0.29	**0.33**
G	304	0.11	0.23	0.00	0.00	**0.39**
C	305	0.47	0.37	0.23	0.28	**0.77**
U	306	0.19	0.21	0.56	0.49	**0.40**
U	307	0.49	0.69	0.49	0.46	**1.00**
U	308	0.87	0.59	0.77	0.78	**0.87**
A	309	1.26	0.24	1.14	0.61	**0.97**
A	310	0.92	0.26	1.23	0.23	**0.02**
A	311	0.60	0.39	0.93	0.10	**0.30**
G	312	0.09	0.14	0.41	0.44	**0.51**
A	313	0.05	0.05	0.27	0.22	**0.17**
C	314	0.14	0.16	0.28	0.28	**0.55**
A	315	0.03	0.02	0.25	0.31	**0.24**
C	316	0.01	0.01	0.13	0.22	**0.36**
G	317	0.04	0.02	0.01	0.02	**0.05**
G	318	0.07	0.07	0.07	0.09	**0.79**
G	319	0.04	0.04	0.04	0.05	**0.57**
G	320	0.10	0.07	0.07	0.08	**0.97**
A	321	0.09	0.09	0.09	0.10	**0.97**
G	322	0.21	0.12	0.12	0.10	**0.69**
U	323	0.50	0.44	0.27	0.34	**0.87**
A	324	0.66	0.53	0.43	0.52	**0.55**
A	325	0.34	0.28	0.80	0.78	**0.43**
A	326	0.26	0.18	0.16	0.03	**0.25**
G	327	0.15	0.13	0.22	0.15	**0.76**
G	328	0.13	0.16	0.28	0.30	**0.49**
U	329	0.29	0.22	0.13	0.16	**0.68**
U	330	0.65	0.42	0.29	0.34	**0.49**
A	331	0.76	0.29	0.51	0.23	**0.26**

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
A	332	0.50	0.19	0.75	0.53	0.53
A	333	0.54	0.10	0.44	0.12	0.04
U	334	0.63	0.23	0.46	0.12	0.18
A	335	0.67	0.23	0.53	0.25	0.20
U	336	0.61	0.13	0.76	0.68	0.60
G	337	0.68	0.19	0.34	0.14	0.10
C	338	0.33	0.23	0.53	0.36	0.86
U	339	0.24	0.22	0.44	0.27	0.23
G	340	0.31	0.26	0.36	0.20	0.93
A	341	0.13	0.17	0.62	0.22	0.04
U	342	0.03	0.07	0.27	0.49	0.52
C	343	0.16	0.26	0.24	0.41	0.30
U	344	0.18	0.14	0.11	0.21	1.00
U	345	0.19	0.15	0.28	0.25	0.23
U	346	0.34	0.11	0.27	0.17	0.57
U	347	0.56	0.21	0.32	0.11	0.30
G	348	0.72	0.33	0.43	0.27	0.42
A	349	0.44	0.29	0.67	0.39	0.36
A	350	0.19	0.20	0.53	0.40	0.36
A	351	0.07	0.08	0.58	0.60	0.21
U	352	0.17	0.19	0.10	0.06	0.72
U	353	0.40	0.34	0.25	0.30	0.96
U	354	0.20	0.18	0.26	0.20	0.89
U	355	0.26	0.08	0.45	0.30	0.38
U	356	0.36	0.23	0.12	0.13	0.20
U	357	0.55	0.32	0.38	0.25	0.78
G	358	0.39	0.13	0.36	0.15	0.20
A	359	0.12	0.13	0.53	0.26	0.08
U	360	0.14	0.10	0.13	0.17	0.94
U	361	0.07	0.07	0.21	0.24	0.63
U	362	0.11	0.07	0.12	0.15	0.70
U	363	0.16	0.09	0.07	0.10	0.51
G	364	0.14	0.09	0.12	0.06	0.25
U	365	0.20	0.12	0.19	0.10	0.62
G	366	0.17	0.05	0.11	0.13	0.45
A	367	0.10	0.12	0.25	0.18	0.32
A	368	0.15	0.16	0.11	0.08	0.78

Appendix C

Mean SHAPE reactivity data of FN26 RNA in the presence and absence of Pr78^{Gag} (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	WITHOUT PROTEIN		WITH PROTEIN		p-value
		MEAN SHAPE REACTIVITY DATA	SD	MEAN SHAPE REACTIVITY DATA	SD	
G	369	0.29	0.34	0.17	0.13	**0.60**
G	370	0.70	0.89	0.30	0.28	**0.57**
A	371	0.68	0.50	0.67	0.79	**0.90**
U	372	-	-	-	-	**-**

Appendix C. Mean hSHAPE reactivity data from triplicate hSHAPE experiments on ssPurines deleted mutant RNA (FN26) both in the absence as well presence of Pr78^{Gag} with standard deviation (SD) and p-values. Color key: *blue, base-paired purines (bpPurines) and **brick red, GU rich region. ND; no data

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
G	1	ND	-
C	2	ND	-
C	3	ND	-
A	4	ND	-
C	5	ND	-
C	6	ND	-
A	7	ND	-
U	8	ND	-
U	9	ND	-
A	10	ND	-
A	11	0.54	-
A	12	0.57	-
U	13	0.56	0.04
G	14	0.04	0.03
A	15	0.02	0.08
G	16	0.00	0.15
A	17	0.10	0.07
C	18	0.21	0.18
U	19	0.04	0.05
U	20	ND	-
G	21	0.00	0.23
A	22	1.49	0.47
U	23	2.00	1.10
C	24	0.21	0.37
A	25	ND	0.41
G	26	0.00	0.34
A	27	2.00	0.62
A	28	2.00	0.60
C	29	0.98	0.22
A	30	2.00	0.53
C	31	0.04	0.22
U	32	ND	-
G	33	0.00	0.21
U	34	0.03	0.14
C	35	0.11	0.10
U	36	0.00	0.14
U	37	0.00	0.26

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
G	38	0.00	0.23
U	39	0.05	0.16
C	40	0.00	0.08
U	41	0.00	0.14
C	42	ND	0.68
C	43	ND	0.77
A	44	ND	0.75
U	45	0.00	0.26
U	46	0.00	0.18
U	47	0.00	0.25
C	48	0.20	0.17
U	49	0.00	0.43
U	50	0.00	0.12
G	51	0.00	0.12
U	52	0.26	0.09
G	53	0.00	0.05
U	54	0.16	0.06
C	55	0.00	0.19
U	56	0.00	0.33
C	57	0.00	0.13
U	58	ND	0.98
U	59	0.97	0.31
G	60	1.45	0.06
U	61	0.62	0.11
U	62	0.50	0.12
C	63	0.00	0.06
C	64	ND	1.13
C	65	0.00	0.43
U	66	0.20	0.06
U	67	0.49	0.24
C	68	ND	-
A	69	0.21	0.22
A	70	0.89	0.03
U	71	0.66	0.18
U	72	0.36	0.15
C	73	0.24	0.09
C	74	0.07	0.08
C	75	0.00	0.22
A	76	0.18	0.24
C	77	0.25	0.13

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
U	78	0.65	0.12
C	79	0.15	0.16
C	80	0.11	0.17
C	81	0.12	0.20
U	82	0.20	0.13
C	83	0.23	0.15
C	84	0.21	0.13
U	85	0.39	0.04
C	86	0.50	0.16
C	87	0.12	0.33
A	88	0.00	0.17
G	89	0.06	0.03
G	90	0.38	0.05
U	91	0.62	0.03
U	92	0.72	0.08
C	93	0.12	0.11
C	94	0.00	0.20
U	95	0.00	0.58
A	96	0.11	0.21
C	97	0.00	0.14
U	98	0.22	0.48
G	99	0.10	0.10
U	100	0.48	0.20
U	101	0.87	0.60
G	102	2.00	0.24
A	103	1.91	0.22
U	104	0.44	0.05
C	105	0.00	0.16
C	106	0.00	0.06
C	107	0.00	0.03
G	108	0.11	0.02
C	109	0.21	0.06
G	110	2.00	0.44
G	111	2.00	0.53
G	112	0.51	0.16
U	113	2.00	0.47
C	114	0.00	0.10
G	115	0.00	0.13
G	116	0.00	0.23
G	117	0.00	0.45

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
A	118	0.00	0.04
C	119	0.00	0.09
A	120	0.21	0.12
G	121	0.38	0.14
U	122	0.50	0.09
U	123	0.38	0.37
G	124	0.30	0.25
G	125	0.08	0.21
C	126	0.22	0.23
G	127	0.25	0.18
C	128	0.00	0.04
C	129	ND	0.58
C	130	0.00	0.30
A	131	0.46	0.07
A	132	0.48	0.16
C	133	0.20	0.06
G	134	0.00	0.03
U	135	0.38	0.11
G	136	0.00	0.08
G	137	0.00	0.13
G	138	ND	-
G	139	0.00	0.30
C	140	0.05	0.16
U	141	0.13	0.19
G	142	0.00	0.20
G	143	0.00	0.09
A	144	0.28	0.02
U	145	0.42	0.19
A	146	0.90	0.28
C	147	0.44	0.10
G	148	0.19	0.25
A	149	0.42	0.60
G	150	0.00	0.50
G	151	ND	2.54
G	152	ND	7.74
A	153	ND	0.97
A	154	ND	1.35
U	155	0.76	0.61
U	156	1.53	0.45
U	157	1.27	0.11

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
C	158	0.28	0.13
G	159	0.13	0.01
U	160	0.45	0.04
G	161	0.65	0.19
A	162	1.24	0.18
G	163	0.04	0.02
G	164	0.01	0.05
A	165	0.12	0.05
A	166	0.39	0.07
G	167	0.47	0.08
A	168	0.55	0.07
C	169	0.16	0.10
G	170	0.38	0.10
A	171	0.23	0.04
C	172	0.02	0.06
G	173	0.00	0.18
C	174	0.02	0.04
G	175	0.00	0.04
U	176	0.07	0.05
U	177	0.00	0.12
C	178	0.17	0.10
G	179	1.11	0.08
C	180	0.05	0.11
C	181	0.01	0.06
G	182	0.04	0.07
G	183	0.03	0.11
C	184	0.22	0.12
C	185	0.06	0.15
G	186	0.00	0.14
G	187	0.00	0.06
C	188	0.00	0.43
G	189	0.00	0.22
A	190	0.67	0.09
U	191	0.89	0.08
U	192	0.96	0.08
A	193	0.82	0.11
A	194	0.61	0.08
A	195	0.54	0.08
A	196	0.33	0.04
G	197	0.10	0.05

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
U	198	0.04	0.05
G	199	0.03	0.07
A	200	0.27	0.05
A	201	0.23	0.07
A	202	0.37	0.13
G	203	0.38	0.07
U	204	0.34	0.33
A	205	0.62	0.10
A	206	0.81	0.19
A	207	0.68	0.18
C	208	0.31	0.14
U	209	0.52	0.09
C	210	0.31	0.09
U	211	0.28	0.10
C	212	0.20	0.10
U	213	0.61	0.07
U	214	0.46	0.12
G	215	0.19	0.10
G	216	0.22	0.12
C	217	0.02	0.07
C	218	0.04	0.07
G	219	0.17	0.08
C	220	0.07	0.04
C	221	0.05	0.05
G	222	0.03	0.07
C	223	0.02	0.10
G	224	0.00	0.08
G	225	0.00	0.04
G	226	0.40	0.09
A	227	1.63	0.13
A	228	1.56	0.08
C	229	0.39	0.13
C	230	0.29	0.12
U	231	1.37	0.11
G	232	1.62	0.09
C	233	0.00	0.03
C	234	0.00	0.05
G	235	0.00	0.10
C	236	0.00	0.05
G	237	0.15	0.09

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
U	238	0.19	0.11
U	239	0.17	0.12
G	240	0.05	0.09
G	241	0.00	0.07
A	242	2.00	0.15
C	243	0.01	0.05
C	244	0.00	0.27
U	245	0.00	0.10
G	246	0.02	0.08
A	247	0.36	0.05
A	248	0.40	0.06
A	249	0.44	0.09
A	250	0.65	0.09
U	251	0.65	0.12
A	252	0.79	0.09
U	253	0.45	0.74
G	254	0.34	0.41
A	255	0.68	0.11
A	256	0.40	0.22
C	257	ND	-
U	258	0.80	0.13
U	259	0.91	0.12
C	260	0.12	0.26
A	261	0.75	0.04
A	262	1.04	0.06
U	263	1.00	0.08
U	264	0.75	0.18
A	265	1.04	0.07
U	266	0.57	0.10
C	267	0.00	0.63
A	268	1.35	0.14
U	269	0.47	0.07
U	270	0.47	0.10
U	271	0.43	0.09
C	272	0.00	0.83
A	273	0.54	0.19
U	274	0.62	0.12
C	275	0.34	0.04
U	276	0.65	0.07
G	277	1.04	0.06

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
G	278	0.14	0.01
A	279	0.05	0.02
G	280	0.00	0.03
C	281	0.00	0.03
U	282	0.08	0.09
U	283	0.25	0.08
A	284	2.00	0.03
G	285	0.37	0.02
U	286	0.09	0.03
G	287	0.11	0.03
A	288	0.13	0.02
U	289	0.11	0.06
A	290	0.24	0.04
C	291	0.38	0.03
U	292	1.47	0.17
A	293	1.81	0.17
U	294	1.73	0.02
C	295	1.01	0.10
U	296	1.45	0.08
C	297	0.46	0.06
A	298	0.98	0.05
A	299	0.98	0.07
A	300	0.63	0.07
U	301	0.19	0.11
A	302	0.19	0.02
U	303	0.15	0.06
C	304	0.08	0.03
U	305	0.09	0.04
C	306	0.00	0.19
A	307	0.02	0.05
A	308	0.09	0.01
G	309	0.06	0.04
U	310	0.04	0.01
U	311	0.08	0.05
C	312	0.00	0.03
A	313	0.47	0.03
A	314	0.93	0.06
G	315	0.22	0.03
C	316	0.00	0.01
C	317	0.00	0.03

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
G	318	0.00	0.02
G	319	0.03	0.02
U	320	0.24	0.04
U	321	1.06	0.03
U	322	1.83	0.08
U	323	2.00	0.13
G	324	0.12	0.02
G	325	0.00	0.02
A	326	0.10	0.05
G	327	0.11	0.02
A	328	0.26	0.03
U	329	0.32	0.03
C	330	0.35	0.06
C	331	0.29	0.07
G	332	0.00	0.03
C	333	0.01	0.01
G	334	0.07	0.05
U	335	0.31	0.06
G	336	0.66	0.06
A	337	0.88	0.07
A	338	0.79	0.08
G	339	0.17	0.04
C	340	0.10	0.02
C	341	0.02	0.02
C	342	0.08	0.03
U	343	0.31	0.08
G	344	0.35	0.05
G	345	0.12	0.04
C	346	0.00	0.07
A	347	0.28	0.03
G	348	0.28	0.01
A	349	0.35	0.03
A	350	0.38	0.03
A	351	0.47	0.02
U	352	0.45	0.05
A	353	0.54	0.07
C	354	0.02	0.10
A	355	0.43	0.05
A	356	0.55	0.07
C	357	0.00	0.13

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
A	358	0.48	0.38
A	359	0.79	0.08
A	360	0.80	0.05
A	361	0.73	0.03
A	362	0.42	0.01
C	363	0.00	0.10
A	364	0.26	0.05
U	365	0.39	0.04
G	366	0.23	0.08
G	367	0.30	0.11
U	368	0.58	0.08
A	369	1.25	0.19
A	370	0.98	0.14
A	371	1.08	0.09
C	372	0.13	0.05
C	373	0.14	0.03
U	374	0.61	0.02
U	375	0.72	0.04
G	376	0.44	0.01
U	377	0.71	0.07
G	378	0.57	0.03
A	379	0.71	0.02
C	380	0.27	0.02
U	381	0.52	0.02
G	382	0.54	0.05
U	383	0.35	0.06
G	384	0.15	0.01
C	385	0.23	0.03
U	386	0.67	0.13
G	387	0.00	0.01
G	388	0.00	0.03
A	389	0.21	0.03
G	390	0.00	0.01
G	391	0.00	0.03
A	392	0.07	0.09
U	393	0.43	0.13
A	394	2.00	0.23
U	395	2.00	0.05
G	396	2.00	0.10
U	397	2.00	0.05

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
U	398	1.08	0.08
U	399	0.26	0.03
C	400	0.02	0.01
C	401	0.00	0.03
U	402	0.00	0.01
C	403	0.00	0.05
C	404	0.00	0.16
C	405	0.00	0.01
C	406	0.00	0.03
A	407	0.04	0.02
C	408	0.02	0.03
C	409	0.03	0.05
G	410	0.13	0.06
A	411	0.70	0.04
U	412	0.54	0.11
U	413	0.97	0.34
A	414	0.53	0.09
A	415	0.20	0.10
C	416	0.01	0.03
U	417	0.17	0.09
C	418	0.83	0.07
U	419	0.29	0.04
C	420	0.07	0.03
U	421	0.13	0.04
U	422	0.22	0.08
A	423	0.32	0.08
C	424	0.04	0.13
A	425	0.91	0.14
A	426	0.63	0.14
C	427	0.30	0.04
U	428	ND	-
G	429	0.00	0.28
U	430	0.10	0.09
U	431	0.46	0.04
U	432	0.73	0.06
C	433	0.25	0.06
U	434	0.16	0.08
U	435	0.18	0.06
G	436	0.09	0.04
C	437	0.00	0.05

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
U	438	0.50	0.02
C	439	0.39	0.10
U	440	0.69	0.18
A	441	0.53	0.05
C	442	0.01	0.03
U	443	0.06	0.05
C	444	0.00	0.11
A	445	0.11	0.03
U	446	0.15	0.10
A	447	0.07	0.07
C	448	0.00	0.02
U	449	0.00	0.03
G	450	0.00	0.09
C	451	0.00	0.25
U	452	0.06	0.07
U	453	1.27	0.14
A	454	0.75	0.04
U	455	0.42	0.04
U	456	0.50	0.09
C	457	0.00	0.33
A	458	0.00	0.22
G	459	0.00	0.05
U	460	0.02	0.06
G	461	0.03	0.08
A	462	0.01	0.06
C	463	0.18	0.08
A	464	0.70	0.05
A	465	1.13	0.09
A	466	1.20	0.06
C	467	0.16	0.09
U	468	0.26	0.03
C	469	0.04	0.05
C	470	0.01	0.03
C	471	0.05	0.02
U	472	0.41	0.09
A	473	ND	-
A	474	ND	-
A	475	ND	-
A	476	ND	-
U	477	ND	-

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
G	478	ND	-
G	479	ND	-
C	480	ND	-
A	481	ND	-
G	482	ND	-
U	483	ND	-
G	484	ND	-
U	485	ND	-
G	486	ND	-
U	487	ND	-
G	488	ND	-
U	489	ND	-
C	490	ND	-
A	491	ND	-
A	492	ND	-
C	493	ND	-
U	494	ND	-
C	495	ND	-
C	496	ND	-
C	497	ND	-
A	498	ND	-
C	499	ND	-
U	500	ND	-
A	501	ND	-
C	502	ND	-
C	503	ND	-
C	504	ND	-
C	505	ND	-
U	506	ND	-
A	507	ND	-
G	508	ND	-
C	509	ND	-
A	510	ND	-
A	511	ND	-
U	512	ND	-
A	513	ND	-
C	514	ND	-
A	515	ND	-
C	516	ND	-
A	517	ND	-

Appendix D

Mean SHAPE reactivity data of the spliced *env* RNA (FN42) (continued)

NUCLEOTIDE SEQUENCE	NUCLEOTIDE NUMBER	MEAN SHAPE REACTIVITY DATA	SD
U	518	ND	-
A	519	ND	-
U	520	ND	-
A	521	ND	-
G	522	ND	-
G	523	ND	-
A	524	ND	-
A	525	ND	-
G	526	ND	-
U	527	ND	-
U	528	ND	-
G	529	ND	-
U	530	ND	-
C	531	ND	-
C	532	ND	-
C	533	ND	-
G	534	ND	-
G	535	ND	-
U	536	ND	-
G	537	ND	-
A	538	ND	-
A	539	ND	-
U	540	ND	-
G	541	ND	-
C	542	ND	-
A	543	ND	-
A	544	ND	-
C	545	ND	-
A	546	ND	-
C	547	ND	-
G	548	ND	-

Appendix D. Mean SHAPE reactivity data from triplicate hSHAPE experiments on the spliced *env* RNA (FN42) with standard deviation (SD). ND; no data.