
UNIVERSITY OF PÉCS

Doctoral School of Biology and Sport Biology

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

PhD thesis

Lizna Mohamed Ali

Supervisor:

Tahir A Rizvi
PhD

PÉCS, 2021

INTRODUCTION

Over the years, the advancement in the field of molecular biology have opened avenues for using vectors from several viruses to deliver the therapeutic genes to treat various rare or genetic disorders. Different types of viruses, such as retrovirus, adenovirus, adeno-associated virus (AAV), and herpes simplex virus (HSV), have been manipulated in the laboratory for such applications. Until 2017, it has been estimated that nearly 70% of the clinical trials for human gene therapy are based on utilizing viral vectors. Among these, retroviruses were the first group of viruses used in gene therapy.

Retroviruses are well known for their ability to accommodate the gene of interest of ~8kb in length that can be integrated into the host genome making it part and parcel of the host genome hence contributing to its long-term expression. Most retroviruses infect dividing cells, however certain retroviruses such as the lentiviruses (HIV-1 and FIV) and betaretrovirus (MMTV) have the additional advantage of infecting non-dividing cells with low

cytotoxicity. Thus, retroviruses are the vector of choice for gene delivery, and it is not surprising that approximately a quarter of all human gene therapy trials have used them.

Despite having these advantages, gene therapists are posed with several challenges when it comes to bringing their research from the laboratory to the bedside. Some of the major concerns are:

1. Efficient packaging of the therapeutic viral vectors into the viral assemblages and precise targeting of the therapeutic viral vectors into the host cells.
2. Random vector integration in the host genome that may lead to unprecedented events of proto-oncogene activation and/or insertional mutagenesis.
3. Risk of generating replication competent retroviruses upon transduction of the host cells, that harbors endogenous retroviral remnants or are previously infected with retroviruses.

Therefore, it is important to study the molecular mechanisms of retroviral life cycle to design and construct

safe and efficient therapeutic vector systems. Here, we investigate the genomic RNA (gRNA) packaging mechanism of a betaretrovirus, Mason-Pfizer monkey virus (MPMV) as they are good candidates for use in human gene therapy. For example: i) they are phylogenetically distinct from human retroviruses, thus avoiding/limiting issues related to recombination with related viruses, ii) MPMV despite being non-human primate virus can be expressed efficiently in human cells and iii) presence of post-transcriptional regulatory elements (CTE) in MPMV vectors may facilitate efficient cytoplasmic expression of the therapeutic genes.

Retroviruses are enveloped viruses that generally vary in size from 80-100 nm in diameter. The lipid bilayer membrane derived from the host cell harbor virally encoded envelope glycoproteins, surface (SU) and the transmembrane (TM) components and determine the host cell tropism. The genome of retrovirus is a linear, single-stranded (ss), non-segmented, 7-12 kb RNA of positive polarity, enclosed as dimers within the protein core encoded

by the *gag* gene that forms the internal structural components of the virions.

The 5' and 3'ends of the retroviral gRNA are flanked by the repeat (R) sequence and the unique sequences (U5 and U3) to each end. Intervening these sequences are the *gag*, *pol* and *env* genes that code for the structural, enzymatic, and surface proteins respectively. Once inside the cell, retroviruses reverse transcribe their RNA genome into DNA utilizing the viral reverse transcriptase enzyme which is then imported into the nucleus where it is integrated into the host genome by the virally encoded integrase enzyme. This integrated retroviral genome called the provirus can now use the cellular machinery to transcribe RNA for making viral proteins as well as to generate extra copies of full-length RNA to be served as the genome for its progeny. The unspliced or full-length RNA plays a dual role as a messenger RNA (mRNA) where it codes for the Gag and Gag/Pol polyproteins and as the gRNA to be encapsidated into newly synthesized virus particles. The transcribed viral mRNAs (both full-length

and spliced RNAs) are exported out of the nucleus into the cytosol for translation. During assembly, viral Gag proteins assemble around its gRNA which is packaged into the nascent virions that bud off from the cell membrane as immature virus particles. Following the release from the cell surface, maturation of the nascent virions occurs by the viral protease rendering it infectious.

The 5' end of the retroviral full-length RNA harbors a highly structured untranslated region (5′ UTR) or leader region with distinct and well characterized RNA motifs that are involved in many steps of the viral replication cycle. This higher order structure includes the *trans*-activation response element (TAR; found only in lentiviruses) essential for viral RNA (vRNA) transcription, the poly adenylation stem loop that act as Poly (A) tail addition site at the 3'end of the RNA during vRNA processing, however, is ignored when located at the 5'end. Next, follows the primer binding site (PBS), which is complementary to the 3'portion of the cellular tRNA specific for each virus to function as the priming point for the initiation of reverse

transcription. Downstream to the PBS lies the dimerization initiation site (DIS), which is involved in the establishment of the kissing-loop complex between two genomic RNA molecules so as to package as dimers; the major splice donor (mSD), which is the prime splice site used in the production of viral subgenomic RNAs; and finally, the core packaging sequences/signals (*psi*, Ψ) consisting the necessary elements required for the specific selection and direction of the gRNA to the viral assembly machinery. All these motifs function either during RNA splicing and processing (mSD) or during gRNA dimerization and encapsidation (DIS and Ψ) into the progeny viruses. These higher order structures may vary in sequence but are highly conserved for their function among various retroviruses. Apart from these structural motifs discussed above, it is now being shown that the presence of a purine rich sequence in the Ψ region of retroviral gRNA is associated with the initial recognition and interaction of the Gag polyprotein with its cognate gRNA from a large milieu of cellular and viral spliced RNAs furthering its selective packaging during viral assembly.

This thesis work identifies and illustrates the role of a similar purine-rich motif in MPMV that may act as a possible Gag binding site during its assembly process. The packaging signal important for genome encapsidation into MPMV particles, like in all retroviruses, lies at the 5' end of the gRNA that include sequences from R to the first 50 nucleotides (nts) of the 5' UTR in region "A" and the last 23 nts of the 5' UTR in region "B" along with the first 120 nt of the *gag* ORF. Based on RNA structure prediction and chemoenzymatic probing methods, this region has been found to fold into a higher-order RNA structure comprising several stable structural motifs that play key roles in MPMV replication, such as the PBS, DIS and the U5-Gag long range interactions (LRIs) that maintain the overall structural stability of the MPMV packaging signal. A stretch of single stranded purine-rich sequence (ssPurines; 5'UUAAAAGU GAAAGUAA3') was identified in proximity to the DIS and its partly duplicated sequence in base-paired form (named here as "bpPurines"; 5'GAAAGUAA3') was discovered located downstream of the mSD in region "B".

During the assembly of MPMV particles, the precursor Gag protein must specifically select the dimerized viral gRNAs from a variety of cellular and spliced viral RNAs. However, not much is known about the Gag-RNA interaction site(s) within the MPMV packaging determinants. Therefore, the overall goal of this thesis was to gain better understanding of the MPMV gRNA packaging process by delineating the mechanism(s) involved in the initial recognition of the gRNA by addressing the following specific aims:

Specific Aim I*: *Establish the role of putative ssPurines sequence in MPMV gRNA packaging. To test the function of this purine-rich motif as potential Gag binding site in MPMV gRNA packaging process by mutational analysis.

Specific Aim II: Establish the role of duplicated base-paired sequence of ssPurines (bpPurines) sequence in MPMV gRNA packaging. There have been some suggestions from earlier studies that bpPurines may compensate for the loss of ssPurines by also functioning as potential Gag binding site in a redundant fashion.

Therefore, the importance of bpPurines both in the absence as well as in the presence of ssPurines was investigated.

Specific Aim III: Establish structure-function relationship of the mutations introduced into ssPurines as well as in its partially duplicated bpPurines during MPMV RNA packaging. Structure-function correlation was established by biochemically validating the predicted structure of the mutant RNAs. These approaches provide clues on the role of the structure of ssPurine region as well as its partially duplicated base-paired counterpart in the context of the overall higher order structure during MPMV RNA packaging.

Specific Aim IV: Correlate the obtained biological data to the Gag precursor protein ($Pr78^{Gag}$) binding site(s) in MPMV gRNA packaging determinants. Finally, the results obtained towards the investigation of the specific aims for this thesis were correlated to the recently published work on finding the specific Gag precursor binding site(s) on MPMV gRNA and helped us in proposing a model on how MPMV could possibly discriminate the full-length

gRNA from its spliced or cellular mRNAs during viral genome encapsidation.

MATERIALS AND METHODS

1. Site-directed mutagenesis and cloning of the mutant RNAs. The desired mutations were designed and introduced at the sequence under investigation by Splice overlap extension (SOE) PCR and cloned into a MPMV-based transfer vector that also serve as a substrate for efficient RNA packaging, if not affected by the subjected mutations, as it contains all the *cis*-acting sequences necessary for viral replication.

2. Three plasmid trans-complementation assay and quantification of the packaging and propagation efficiencies of the mutant RNAs. Each of the mutant transfer vector RNAs was allowed to propagate in an in-cell based, single round replication assay relative to the wild type (WT) MPMV RNA. To generate infectious particles, two other plasmids- 1) that code for the MPMV Gag/Pol polyproteins (structural and enzymatic) to make the virus-

like particles, 2) to encode the vesicular stomatitis virus G envelope glycoprotein to pseudotype the virus particles; are co-transfected along with the transfer vector carrying the mutations or WT RNA. Transfected cells are fractionated into nuclear and cytoplasmic fractions, analyzed, and quantitated for transfer vector RNA nuclear export and expression. Viral particles produced are quantitated for the packaged RNA by RT-qPCR. Viral supernatants are also used to infect target cells (HeLaT4) to study RNA propagation. After infection, target cells are selected with media containing hygromycin B, allowing only those cells to survive which have been successfully infected since the packaged vector RNA contains the *hygromycin resistance* gene. The number of resulting hygromycin-resistant colonies obtained per milliliter of the virus supernatant (CFU/ml) is a measure of the propagation efficiency of each mutant transfer vector RNA, which was normalized to the transfection efficiency and the values reported relative to the WT viral titers.

3. In Silico analysis of MPMV RNA secondary structure, in vitro RNA transcription and SHAPE. To establish structure-function relationship for each mutant in relation to WT RNA during its RNA packaging process, an online software Mfold, was used to predict the RNA secondary structure. Mfold predicts all optimal and suboptimal RNA secondary structures based on energy matrices by considering the minimum free energy of the provided RNA sequence. Furthermore, the predicted RNA structure for each mutant and the WT was validated by a chemical probing method called Selective 2' hydroxyl acylation analyzed by primer extension (SHAPE). SHAPE utilizes a chemical, Benzoyl cyanide (BzCN) to acylate (or modify) the 2'-hydroxyl group of the unconstrained nucleotides in the RNA structure, followed by interrogation of each nucleotide using two sets of identical but differentially labeled primers: one set for reverse transcription extension of the modified RNA and the other set to generate the sequencing ladders specific to the RNA (unmodified) in the reaction. Upon sequencing, the electrophoretograms of the SHAPE-modified RNAs were converted to SHAPE

reactivity data by the QuShape algorithm that were applied as constraints to the mutant RNA sequence in RNAstructure program (version 6) to obtain the validated RNA secondary structure and were redrawn using VARNA software for better representation. For this, the WT and mutant DNA clones were *in vitro* transcribed before SHAPE could be performed on them.

4. In vitro RNA Dimerization Assays. To test the effect of the introduced mutations on gRNA dimerization, *in vitro* RNA dimerization assays were performed. *In vitro* transcribed and purified WT and mutant RNAs were incubated in dimer or monomer buffer, electrophoresed, stained with ethidium bromide and visualized for dimeric or monomeric bands using ultraviolet (UV) transillumination.

RESULTS

1. Role of ssPurines in gRNA packaging and propagation.

Three distinct mutations were created in the ssPurines region in which, the former region (5 nts; underlined) of ssPurines (5'UUAAAAGU GAAAGUAA3') that is unique to this site was deleted. The second and third mutants comprised of complete deletion and substitution of ssPurines respectively. Collectively, these mutants should identify the effect, if any, of the ssPurines region on packaging and propagation of the virus, as well as identify if there were any differential effects between the former and latter purines within the ssPurines. Results of these mutant RNAs revealed that, despite partial or complete deletion or substitution of ssPurines, RNA packaging and propagation of these transfer vectors were not significantly affected when compared to the WT suggesting that ssPurines are dispensable for MPMV replication. Or else, as speculated earlier, one may also infer that the bpPurines (partial duplication of ssPurines in region 'B') may be compensating for the loss of function of the ssPurines.

2. Role of bpPurines in gRNA packaging and propagation. Here, group of three mutants were created that invariably had deletion of bpPurines (5'GAAAGUAA3') along with any one of the three scenarios with respect to the ssPurines:

1) complete deletion of ssPurines (5' UUAAAAGU GAAAGUAA 3');

2) partial deletion of ssPurines (5' UUAAAAGU GAAAGUAA 3') where the duplicated sequence of bpPurines (underlined) was deleted;

3) in the presence of intact ssPurines.

In the first case of double deletion mutant involving complete deletion of ssPurines and bpPurines, the results showed severe defects in RNA packaging compared to the WT vector RNA which corroborated well with a concomitant drastic reduction in the colony-forming units (CFU)/ml for this mutant. Surprisingly, this was not the case for the second and third mutants (with the presence of partial or intact ssPurines) where the packaging capabilities of these mutant RNAs were nearly like that of the WT.

However, despite efficient RNA packaging, the RNA propagation capabilities were drastically affected for these bpPurines deficient mutants.

3. Structure-function analysis of ssPurines and bpPurines mutants. SHAPE-validated structure for the ssPurines mutant RNAs revealed an almost intact packaging signal RNA relative to the WT structure. This correlated well with the biological results obtained for these mutants which behaved much alike the WT. We also observed that in the absence of ssPurines, the bpPurines still remained base paired as opposed to our earlier speculation that it would unpair and compensate for the loss of ssPurines, an assertion made employing prediction approaches. Among the bpPurines mutants, the double deletion mutant RNA comprising the deletion of ssPurines and bpPurines, SHAPE revealed a complete architectural distortion of the packaging signal RNA which explained severe defects in RNA packaging and propagation of this mutant. This observation confirms the importance of maintaining an intact higher order structure of the packaging signal RNA

for MPMV replication. The validated RNA structures for the other bpPurines deleted mutants, in the presence of partial or complete ssPurines, maintained all the important structural motifs and thus explains the unaffected biological phenotype. On the other hand, the propagation capability of the virus was largely influenced by the presence or absence of the bpPurines on an overall mostly intact gRNA structure.

4. In vitro dimerization capability of the mutant RNAs. Genomic RNA dimerization and packaging are interconnected events in the retroviral life cycle. Since all the mutants considered in this study showed efficient RNA packaging except for one mutant which jointly lacked the ssPurines and bpPuirnes, it was important to investigate the dimerization abilities of these mutant RNAs. Interestingly, none of the mutant RNAs tested showed any significant difference in their dimerization potential compared to the wild type suggesting that any effect on packaging of the mutant RNA could be attributed to the introduced mutation

and not due to the tampering of other aspects of virus replication.

5. Role of a GU-rich sequence of region "B" in gRNA packaging and propagation. Recent biochemical data from our laboratory have identified two $Pr78^{Gag}$ binding sites on MPMV gRNA which were proposed to act redundantly during MPMV gRNA packaging process. These redundantly acting binding sites include: 1) the ssPurines (5'UAAAAGUGAAAGUAA3') and 2) a second loop (5'AAGUGUU3') corresponding to the last two nucleotides of the bpPurines and extending immediately into the adjacent GU-rich region. To further validate the biological role of these Gag binding structural motifs in MPMV gRNA packaging, a mutant was created including the simultaneous deletion of both the identified Gag binding sites. In concordance with the biochemical data and as expected, this double deletion mutant showed a drastic defect in RNA packaging and concomitant propagation defect further validating that both the ssPurines and the GU rich region are important during MPMV

replication. To further confirm this assertion, another mutant was created, containing the deletion of only the GU rich region while maintaining ssPurines. Consistent with the recent biochemical observations, this mutant was found to be packaged efficiently even though it had GU rich region completely deleted and corroborated well with the RNA propagation data. Taken together, results presented here further validate that these two Gag binding sites act redundantly during MPMV gRNA packaging process. Remarkably, together these findings also primed us to propose a model on how MPMV could possibly discriminate its full-length gRNA which is usually less than 1% in the large pool of spliced viral and cellular mRNA during viral RNA encapsidation. It is now clear that Gag binds significantly to the ssPurines which is present invariably in all the MPMV transcripts as it is located upstream of mSD; the selection process of gRNA is achieved by the precursor Gag after simultaneous binding to the GU-rich region present downstream of mSD that is always absent in spliced viral RNAs.

In summary, this study suggests that the ssPurines, and GU-rich sequences play a redundant role in MPMV life cycle by acting as Gag binding sites, while maintaining the overall structure and stability of the MPMV packaging signal RNA. The findings also corroborate well with the mechanism proposed for MPMV gRNA recognition during viral assembly. Additionally, our results suggest a potential role of the bpPurines in the early steps of the MPMV life cycle that affect propagation of the packaged RNA. It will also be interesting to perform *in vivo* or *in virio* probing experiments to further validate the observed *in vitro* biochemical data presented here.

CONCLUSIONS

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

- Presence of ssPurines in MPMV gRNA establishes packaging which suggests it as the primary Gag binding site during viral RNA encapsidation.
- Deletion of either the ssPurines or bpPurines does not affect MPMV gRNA packaging and bpPurines need not necessarily be unpaired (as was speculated earlier) to bring about gRNA packaging.
- Non-responsiveness of bpPurines to RNA packaging, suggested the presence of other unpaired regions; the GU-rich loop downstream of bpPurines that was shown to facilitate gRNA packaging.

- Deletion of bpPurines, irrespective of the presence or absence of ssPurines affects RNA propagation severely but not RNA packaging.
- Biochemical probing by SHAPE reveals structural implication for severe defects in RNA packaging.
- Biological results presented here further confirm that ssPurines and GU-rich region serves as $Pr78^{Gag}$ binding sites in a redundant fashion as has recently been confirmed based on biochemical data.

LIST OF PUBLICATIONS

Publications Related to thesis:

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

Publications Non-Related to Thesis:

1. Pillai, V. N., **Ali, L. M**., Prabhu, S. G., Krishnan, A., Chameettachal, A., Pitchai, F. N. N., Mustafa, F., Rizvi, T. A. **2021.** A Stretch of Unpaired Purines in the Leader Region of Simian Immunodeficiency Virus (SIV) is Crucial for its Genomic RNA Packaging. ***J Mol Biol*** (Submitted).

2. Chameettachal, A., Pitchai, F. N. N., Vivet Boudou, V., Pillai, V. N., **Ali, L. M**., Mustafa, F., Marquet, R., Rizvi, T. A. **2021.** Selective capture of gRNA by the MMTV Pr77^{Gag} requires single-stranded purines present within the bifurcated stem loop 4. ***Nucleic Acids Research*** 49(8):4668-4688. DOI:10.1093/nar/gkab223

3. Krishnan, A., Pillai, V. N., Chameettachal, A., **Ali, L. M**., Nuzra NagoorPitchai, F., Tariq, S., Mustafa, F., Marquet, R., and A Rizvi, T. **2019.** Purification and Functional Characterization of a Biologically Active Full-Length Feline Immunodeficiency Virus (FIV) Pr50^{Gag}. ***Viruses*** 11(8):689. DOI:10.3390/v11080689

4. Kalloush, R. M., Vivet-Boudou, V., **Ali, L. M**., Pillai, V. N., Mustafa, F., Marquet, R., Rizvi, T. A. **2019.** Stabilizing role of structural elements within the 5´ Untranslated Region (UTR) and gag sequences in Mason-Pfizer monkey virus (MPMV) genomic RNA packaging. ***RNA Biol*** 16: 612–625. DOI:10.1080/15476286.2019.1572424

5. Akhlaq, S., Panicker, N. G., Philip, P. S., **Ali, L. M**., Dudley, J. P., Rizvi, T. A., Mustafa, F. **2018.** A cis-Acting Element Downstream of the Mouse Mammary Tumor Virus Major Splice Donor Critical for RNA

Elongation and Stability. ***J Mol Biol*** 430: 4307–4324. DOI:10.1016/j.jmb.2018.08.025

6. Pitchai, F. N. N., **Ali, L. M**., Pillai, V. N., Chameettachal, A., Ashraf, S. S., Mustafa, F, Marquet, R., Rizvi, T. A. **2018.** Expression, purification, and characterization of biologically active full length Mason-Pfizer monkey virus (MPMV) $Pr78^{Gag}$. ***Sci Rep*** 08 7; 8(1):11793.
7. Chameettachal, A., Pillai, V. N., **Ali, L. M**., Pitchai, F. N. N., Ardah, M. T., Mustafa, F, Marquet, R., Rizvi, T. A. **2018.** Biochemical and functional characterization of mouse mammary tumor virus (MMTV) full-length Pr77Gag expressed in prokaryotic and eukaryotic cells. ***Viruses*** 06 18; 10(6). DOI:10.3390/v10060334
8. Mustafa, F., Vivet-Boudou, V., Jabeen, A., **Ali, L. M**., Kalloush, R. M., Marquet, R., Rizvi, T. A. **2018.** The bifurcated stem loop 4 (SL4) is crucial for efficient packaging of mouse mammary tumor virus (MMTV) genomic RNA. ***RNA Biol*** 15:8, 1047-1059. DOI:10.1080/15476286.2018.1486661
9. **Ali, L. M**., Rizvi, T. A., Mustafa, F. **2016.** Cross- and Co-Packaging of Retroviral RNAs and their Consequences. ***Viruses***, 8(10), 276. DOI:10.3390/v8100276

10. Kalloush, R. M., Vivet-Boudou, V., **Ali, L. M.,** Mustafa, F., Marquet, R., Rizvi, T. A. **2016.** Packaging of Mason-Pfizer Monkey Virus (MPMV) Genomic RNA Depends upon Conserved Long Range Interactions (LRIs) Between U5 and gag Sequences. ***RNA***, 22:905-919. DOI:10.1261/rna.055731.115

11. Al Ahmad, M., Mustafa, F., **Ali, L. M**., Karakkat, J. V., Rizvi, T. A. **2015.** Label-free capacitance-based identification of viruses. ***Sci Rep***, 5:9809. DOI:10.1038/srep09809

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

13. Al Ahmad, M., Mustafa, F., **Ali, L. M**., Rizvi, T. A. **2014.** Virus detection and quantification using electrical parameters. ***Sci Rep***, 4:6831. DOI:10.1038/srep06831

14. Aktar, S. J., Jabeen, A., **Ali, L. M**., Vivet-Boudou, V., Marquet, R., Rizvi, T. A. **2013.** SHAPE analysis of the 5' end of the Mason-Pfizer monkey virus (MPMV) genomic RNA reveals structural elements required for genome dimerization. RNA, 19:1648-58. DOI: 10.1261/rna.040931.113

15. Saeed, I. A., **Ali, L. M**., Jabeen, A., Khasawneh, M., Rizvi, T. A., and Ashraf, S. S. **2013.** Estrogenic activities of ten medicinal herbs from the Middle East**.** ***Journal of Chromatographic Science***, 51:33-39. DOI:10.1093/chromsci/bms101