Identification of Pr78\textsuperscript{Gag} binding sites on the Mason-Pfizer monkey virus genomic RNA packaging determinants

Fathima Nuzra Nagoor Pitchai, Akhil Chameettachal, Valérie Vivet-Boudou, Lizna Mohamed Ali, Vineeta N. Pillai, Anjana Krishnan, Serena Bernacchi, Farah Mustafa, Roland Marquet, Tahir A. Rizvi

PII: S0022-2836(21)00122-4
DOI: https://doi.org/10.1016/j.jmb.2021.166923
Reference: YJMBI 166923

To appear in: Journal of Molecular Biology

Received Date: 29 November 2020
Revised Date: 1 March 2021
Accepted Date: 1 March 2021

Please cite this article as: F. Nuzra Nagoor Pitchai, A. Chameettachal, V. Vivet-Boudou, L. Mohamed Ali, V.N. Pillai, A. Krishnan, S. Bernacchi, F. Mustafa, R. Marquet, T.A. Rizvi, Identification of Pr78\textsuperscript{Gag} binding sites on the Mason-Pfizer monkey virus genomic RNA packaging determinants, Journal of Molecular Biology (2021), doi: https://doi.org/10.1016/j.jmb.2021.166923

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 The Author(s). Published by Elsevier Ltd.
Identification of Pr78\textsuperscript{Gag} binding sites on the Mason-Pfizer monkey virus genomic RNA packaging determinants

Fathima Nuzra Nagoor Pitchai\textsuperscript{1}, Akhil Chameettachal\textsuperscript{1}, Valérie Vivet-Boudou\textsuperscript{2}, Lizna Mohamed Ali\textsuperscript{1}, Vineeta N. Pillai\textsuperscript{1}, Anjana Krishnan\textsuperscript{1}, Serena Bernacchi\textsuperscript{2}, Farah Mustafà\textsuperscript{3,4}, Roland Marquet\textsuperscript{2},*, Tahir A. Rizvi\textsuperscript{1,4,*}

\textsuperscript{1}Department of Microbiology & Immunology, College of Medicine and Health Sciences (CMHS), United Arab Emirates University (UAEU), Al Ain, United Arab Emirates (UAE)

\textsuperscript{2}Université de Strasbourg, CNRS, Architecture et Réactivité de l’ARN, UPR 9002, Strasbourg, France

\textsuperscript{3}Department of Biochemistry, College of Medicine and Health Sciences (CMHS), United Arab Emirates University (UAEU), Al Ain, United Arab Emirates (UAE)

\textsuperscript{4}Zayed bin Sultan Center for Health Sciences, United Arab Emirates University

*To whom correspondence should be addressed:

TAR, Department of Microbiology & Immunology, College of Medicine and Health Sciences (CMHS), United Arab Emirates University (UAEU), P.O. Box 17666, Al Ain, UAE.

Tel: (+971) 3-713-7514; Fax: (+971) 3-767-1966; E-mail: tarizvi@uaeu.ac.ae

RM, Université de Strasbourg, CNRS, Architecture et Réactivité de l’ARN, UPR 9002, IBMC, 2 Allée Konrad Roentgen, 67084 Strasbourg cedex, France.

Tel: (+33) 388 41 70 54. Fax: (+33) 388 60 22 18; E-mail: r.marquet@ibmc-cnrs.unistra.fr
Abstract

How retroviral Gag proteins recognize the packaging signals (Ψi) on their genomic RNA (gRNA) is a key question that we addressed here using Mason-Pfizer monkey virus (MPMV) as a model system by combining band-shift assays and footprinting experiments. Our data show that Pr78Gag selects gRNA against spliced viral RNA by simultaneously binding to two single stranded loops on the MPMV Ψi 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 major splice donor (mSD) 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.

Key words:
Retroviruses; Gag-RNA interactions; Purines; Footprinting; hSHAPE.
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 [1–6]. 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 [7–13]. 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 [7,9–15].

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* [1,2,4–6,13,14,16–21]. These *Psi* sequences harbor unique Gag binding sites important for selective gRNA packaging in retroviruses [21,22]. 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 their primary sequence [15,20,23–27].

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 [6,12–14]. 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_{2-4}-C-X_{4-6}-H-X_{4-6}-C; where C = Cys, H = His, X_n = n number of other amino acids) that sequester zinc ions required for specific gRNA binding [6,12–14].

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 [28–37]. 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 [28,30,37]. However, use of full-length HIV-1 Pr55\textsuperscript{Gag} has revealed the presence of a high affinity binding site on the internal (G//AGG) loop of SL1 of the HIV Psi RNA [9]. 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, the role of domains other than NC cannot be excluded. In fact, a number of studies have implicated other domains of Gag in specific binding to gRNA, including the MA [38–42], CA [43], p1 [44], p2 [45,46] and p6 [47,48] 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 betaretrovirus. It was first isolated from the breast adenocarcinoma of a rhesus monkey (Macaca mulatta) and is known to cause fatal immunodeficiency in macaques [49,50]. 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 [51–54]. 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 [55]. 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 [24,27,56–63]. These sequences span from the 5’ UTR into the gag open reading frame, similar to most other retroviruses [14]. 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 1A; 59,60]. Consistent with other retroviruses, the Psi sequences on MPMV gRNA fold into a higher order structure comprising of various structural motifs [24,60]. Among these structural motifs, two purine-rich motifs, the single stranded purines (ssPurines; U\textsuperscript{191}UAAAAGUGAAAGUAA\textsuperscript{206}) and the base paired purines (bpPurines; G\textsuperscript{246}AAAGUAA\textsuperscript{253}), have been identified as unique regions.
in their composition and positioning that may contribute to Gag binding [24,60]; Figure 1B).

Presence of these 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 [3,9,15,17,18,64–69].

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 1B). 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 [9]. Interestingly, the latter half of the ssPurines sequence is found repeated downstream in a base paired manner to form the bpPurines (Figure 1B; [60]. 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 [63].

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\textsuperscript{Gag}) binding sites (Figure 1B). Our results indicate that the MPMV polyprotein, Pr78\textsuperscript{Gag} binds to two loops: 1) the ssPurines loop (U\textsuperscript{191}UAAAAGUGAAAGUAA\textsuperscript{206}) and 2) a second loop (A\textsuperscript{252}AGUGU\textsuperscript{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\textsuperscript{Gag}. 


Results

Characterization of recombinant full-length purified MPMV Gag polyprotein (Pr78\textsuperscript{Gag-His\textsubscript{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\textsuperscript{Gag} was performed and the protein characterized for its biological function, revealing that it could assemble \textit{in vitro} to form virus-like particles (VLPs), and also form VLPs in bacteria, and that the VLPs produced in eukaryotic cells could encapsidate MPMV RNA containing the \textit{Psi} region [62].

The bacterially expressed and full-length Pr78\textsuperscript{Gag} was characterized by DLS, which revealed that the purified protein did not contain any aggregates. The mean hydrodynamic radius (R\textsubscript{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\textsuperscript{Gag} trimers (Supplemental figure 1).

Pr78\textsuperscript{Gag} discriminates between full-length, unspliced sub-genomic \textit{Psi} RNA and spliced \textit{env} RNA

We first attempted to establish if Pr78\textsuperscript{Gag} specifically binds to the full-length, unspliced gRNA (RCR001) over spliced \textit{env} RNA (FN42; Figure 2A). 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, we performed band shift assays to determine the optimal concentration of Pr78\textsuperscript{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\textsuperscript{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 2B). The band shift gel was then quantitated, and the data plotted to fit Hill’s equation (Figure 2C). The Hill’s coefficient was estimated to be 1.216 ± 0.422 (mean ± SD), suggesting that binding of Pr78\textsuperscript{Gag} to gRNA is weakly cooperative or non-cooperative. The apparent K\textsubscript{d} obtained was 216.2 ± 76.10 nM (mean ± SD).

Having determined the optimal concentration of Pr78\textsuperscript{Gag} for a complete shift of the dimeric gRNA, we proceeded to perform competitive band shift assays to determine the differential ability of Pr78\textsuperscript{Gag} to bind to unspliced gRNA (RCR001) versus spliced \textit{env} RNA (FN42) of equal length (Figure 2A). 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) \textit{env} RNA (Figure 2A & D). The experiments 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 2D). 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 \textit{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 2D). This is further supported by the statistically significant difference observed for the percentage of bound RNA between the unspliced gRNA (RCR001)
and the spliced \textit{env} RNA (FN42). These results clearly indicate the preferential binding of Pr78\textsuperscript{Gag} to unspliced MPMV gRNA rather than to its spliced \textit{env} RNA.

**Pr78\textsuperscript{Gag} binds redundantly to both the ssPurines and bpPurines.**

Next, we investigated where Pr78\textsuperscript{Gag} binds on the full-length MPMV gRNA. Our earlier genetic and biochemical analyses had identified two structural elements crucial for MPMV gRNA packaging, ssPurines and bpPurines [24,60]. Mutations introduced in these purine-rich regions had shown that both these motifs were responsible for efficient gRNA packaging in a redundant fashion [63]. Thus, competitive band shift assays were performed on a number of mutant clones in these motifs to determine whether binding of Pr78\textsuperscript{Gag} to these regions could be observed (Figure 3A).

Using a similar experimental design as described for \textit{env} RNA, the mutant RNAs were \textit{in vitro} transcribed, and increasing concentrations of the mutant RNAs (from 0 to 400 nM; Figure 3A) 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 \textit{versus} increasing competitor RNA (Figure 3B & C). The data thus obtained was compared to that obtained from the WT unspliced gRNA (RCR001) to determine the competing (or non-competing) ability of these mutants to bind Pr78\textsuperscript{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 3B). 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\(^{\text{Gag}}\) 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\(^{\text{Gag}}\) binding to the FN26 mutant RNA, we proceeded to investigate the effect of protein binding to several bpPurines mutants (FN16, FN19, 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 3A). 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 3B & C). These results reveal that the bpPurines are dispensable for Pr78\(^{\text{Gag}}\) binding in the presence of ssPurines and implies the presence of other Pr78\(^{\text{Gag}}\) 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.

Having observed no significant effect on the \textit{in vitro} binding of Pr78\(^{\text{Gag}}\) to these individual ss- and bp-Purines mutants, we next investigated the combined effect of the deletion of both these purine-rich regions on \textit{in vitro} Pr78\(^{\text{Gag}}\) binding. 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 3A) revealed poor competition for Pr78\(^{\text{Gag}}\) binding against the labeled WT RNA with a statistically significant difference between the levels of their bound RNA (Figure 3C).

The \textit{in vitro} Pr78\(^{\text{Gag}}\) binding to these gRNA mutants correlated well with their \textit{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 3A; [63]).

**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, we combined hSHAPE, a technique that allows structural investigation at single nucleotide resolution following structure-dependent modification of the RNA, with footprinting assays. In these experiments, RNA-protein interactions were monitored by modifying the WT RNA with an 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 [9] and MMTV [Chameettachal A. et al., to be published] 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 \(\mu\)M 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 \(\leq 0.05\),
between the samples treated with and without Pr78\textsuperscript{Gag}, were considered significant (Supplemental table 1). Any significant attenuation in reactivities of nucleotides were regarded as Pr78\textsuperscript{Gag} binding nucleotides or 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 [24]; Figure 4A). In the presence of Pr78\textsuperscript{Gag}, a significant attenuation of reactivities was observed in two major regions of the \textit{Psi}, including: (1) the ssPurines (U\textsuperscript{191}UAAAAGUGAAAGUAA\textsuperscript{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\textsuperscript{254}UGU\textsuperscript{257}) immediately downstream of the bpPurines (Figure 4B-D). The nucleotides A193, A194, A196, U198, A200 and A201 in the ssPurines showed a significant attenuation in reactivities (with \textit{p} values \leq 0.05 and reactivities with a 1.5- to 2-fold reduction or more), indicating Pr78\textsuperscript{Gag} binding to these nucleotides (Figure 4D & E). However, a few other nucleotides (A195, G199 and A205) within this region also showed a 2-fold reduction in reactivity with \textit{p} values \leq 0.06-0.08 (Figure 4D & E). These results suggest significant binding of Pr78\textsuperscript{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 4C-E).

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 adenine nucleotides (A252 and A253) located at the 3’ end of the bpPurines (Figure 4C-E). Quite unexpectedly, this Pr78\textsuperscript{Gag} footprinting extended into an immediately adjacent cluster of guanosine and uracil nucleotides and hence is referred to here as the “GU-rich region”
(Figure 1B, 4C & D). This GU-rich region is also located within the region determined earlier to be indispensable for in vivo MPMV RNA packaging [59,60]. 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 4C & D). Our footprinting results show 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 4C-E). In addition, the nucleotide G259 located immediately after the GU-rich region showed a significant attenuation in hSHAPE reactivity (Figure 4C-E). 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.

**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, we performed footprinting experiments 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 5A; [63]). The ssPurines deletion had only local structural effects while the bpPurines retained their partially base-paired nature (Figure 5A). 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 \( \leq 0.05 \), between the RNA treated with and without protein, were considered significant (Supplemental table 2).

Footprinting results obtained for the ssPurines mutant, FN26, revealed significant attenuation in hSHAPE reactivities in a few of the nucleotides of the bpPurines and the GU-rich region, including nucleotides G230, A232 and A237 of the bpPurines, suggesting Pr78\textsuperscript{Gag} binding to this region (Figure 5B & D); note that the nucleotide numbers of this region differ due to the deletion in the mutant; Figure 5C-E). 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\textsuperscript{Gag}, presumably due to conformational changes in the structure of this region due to Pr78\textsuperscript{Gag} binding (Figure 5C-E). Surprisingly, here too we observed significant reduction in reactivities of the nucleotides pertaining to the GU-rich region (G\textsuperscript{238}UGUU\textsuperscript{242}; Figure 5B-E). All the nucleotides in this region, from G238 to U242, indicated a significant reduction in hSHAPE reactivities; hence, Pr78\textsuperscript{Gag} binding (Figure 5C-E). 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 pair upon Pr78\textsuperscript{Gag} binding (Figure 5C & D).
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\textsuperscript{Gag} on the MPMV \textit{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 \textit{env} RNA (FN42) to determine whether the ssPurines adopt an alternative confirmation compared to the unspliced, full length \textit{Psi} RNA (RCR001) using hSHAPE. This analysis revealed that nucleotides A\textsubscript{196}GUGA\textsubscript{200}, forming the ssPurines loop in the unspliced RNA SHAPE validated structure (Figure 4A & C), are involved in partial base pairing with nucleotides U\textsuperscript{209}CUCU\textsuperscript{213}, downstream of the ssPurines (Figure 6). Interestingly, four of the five nucleotides forming this partially base paired conformation in the spliced RNA, were found to be involved in significant Pr78\textsuperscript{Gag} binding in the unspliced RNA, as shown by footprinting (Figure 4D & E).

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 \textit{gag}, a region that assumes a complex secondary RNA structure (Figure 1B; [24,59]). Interestingly, while most of the structural motifs of MPMV \textit{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 [60,63]. Thus, using purified full-length MPMV Gag protein in footprinting assays, our results provide direct evidence for Pr78\textsuperscript{Gag} binding to two sites within the MPMV \textit{Psi} RNA: 1) the ss-Purine loop (U\textsuperscript{191}UAAAAGUGAAAGUAA\textsuperscript{206}), and 2) a second loop (A\textsuperscript{252}AGUGUU\textsuperscript{258}) corresponding to the last two nucleotides of the bpPurines.
region and extending immediately into the adjacent GU-rich region (Figure 7). 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 [70]. 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 3A) were capable of competing well against the labeled WT unspliced gRNA (Figure 3B & C), revealing the presence of other Pr78Gag 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 Pr78Gag binding (Figure 3C). 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 [63]. This perhaps is responsible for the poor Pr78Gag 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 [69]. 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 [69].

Interestingly, the MPMV Pr78Gag 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 4B-E). Interestingly, among the two regions showing Pr78Gag footprints, ssPurines contained the longer Pr78Gag footprint, which is consistent with the large purine-rich loop in the Psi RNA structure. Notably these Pr78Gag 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 Pr78Gag did not show significant changes in the bpPurines because of their base-paired confirmation (G246AAAGU251; Figure 4A-D). 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 Pr78Gag binding (Figure 4B-E). The GU-rich region (G254UGUU258; Figure 4A-D), 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\textsuperscript{Gag} binding (Figure 4C-E). Additionally, Pr78\textsuperscript{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\textsuperscript{Gag} binding (Figure 4C & D). Considering that A252 and A253 of the bpPurines form a continuous single-stranded loop of seven nucleotides along with the GU-rich region (A\textsuperscript{252}AGUGUU\textsuperscript{258}), five of these nucleotides showed Pr78\textsuperscript{Gag} binding (A252, A253, G254, G256 and U258). Thus, the overall binding of Pr78\textsuperscript{Gag} to the MPMV \textit{Psi} RNA can be categorized into two important regions: (1) the ssPurines and (2) the single-stranded A\textsuperscript{252}AGUGUU\textsuperscript{258} loop that may act redundantly in bringing about gRNA packaging (Figure 7).

In the absence of the ssPurines (FN26 RNA), Pr78\textsuperscript{Gag} binding could be mapped to: (1) bpPurines and (2) the single-stranded GU-rich region (G\textsuperscript{238}UGUU\textsuperscript{242}; Figure 5). Analysis of the footprinting data in the absence of ssPurines revealed a number of nucleotides with a significant increase in SHAPE reactivity upon Pr78\textsuperscript{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 5B-D). This significant increase in reactivity, particularly in base paired-nucleotides, implies a conformational change of this region or nucleotides upon Pr78\textsuperscript{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\textsuperscript{Gag} binding to augment gRNA packaging (Figure 4), also implying that ssPurines may act as the dominant or primary Gag binding site.
A crucial observation from our study is that Pr78\textsuperscript{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 2D). 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 [9,22,70–72, Chameettachal A. et al., to be published]. Interestingly, the mSD in MPMV gRNA is located within one of these Pr78\textsuperscript{Gag} binding sites, the bpPurines, (G\textsuperscript{246}AAA\textsubscript{GU}AA\textsuperscript{253}; Figure 1B), thus making it part of the unspliced RNA. This is unlike the ssPurines (another Pr78\textsuperscript{Gag} binding site), which is located upstream of the mSD and thus, present on both the unspliced and spliced env RNA (Figure 1B). Footprinting on the WT gRNA structure also indicates binding of the polyprotein to two unpaired adenosines (A\textsubscript{252} and A\textsubscript{253}) 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 RNA (Figure 4B-E). Notably, these sequences form a continuous single-stranded loop of six single-stranded nucleotides (A\textsuperscript{252}AGUGU\textsuperscript{257}; Figure 4C & D). The presence of more than one Pr78\textsuperscript{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 crucial Pr78\textsuperscript{Gag}-specific binding region 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 [68,69,74]. Analysis of the Pr78\textsuperscript{Gag} binding data
with the Hill equation indicates that its binding is weakly or not cooperative (Figure 2C). This is in contrast with the situation prevailing in MMTV, since similar experiments showed a high degree of cooperativity of MMTV Pr77\textsuperscript{Gag} to its cognate gRNA (Hill coefficient = 3; Chameettachal A. \textit{et al}, to be published). In HIV-1, while binding of Gag\Delta p6 to gRNA is cooperative [41], binding of the full-length Pr55\textsuperscript{Gag} is not cooperative, but there are two classes of Pr55\textsuperscript{Gag} binding sites with different binding affinities [15]. Irrespective, our data suggest that the ssPurines and the single-stranded GU (A\textsuperscript{252}AGUGU\textsuperscript{257}) region function redundantly, leading to the provocative question as to how the presence of a Pr78\textsuperscript{Gag} binding sites upstream of the mSD in the spliced \textit{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 [9] and MMTV [Chameettachal A. \textit{et al}, to be published]. In the case of HIV-1, it has been suggested that the specific binding site for its precursor polyprotein (Pr55\textsuperscript{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 [9]. 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 [9,15]. 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 [25,74, Chameettachal A. \textit{et al}, to be published] Similar to HIV-1 and MMTV, MPMV too relies on the U5-Gag LRIs to maintain the overall structure of the \textit{Psi} region responsible for initiating and augmenting gRNA packaging [24,25,61, Chameettachal A. \textit{et al}, to be published]. In fact, the
MPMV gRNA harbors two such U5-Gag LRIs, LRI-I and LRI-II (Figure 1B). The loss of these LRIs during splicing results in disruption of the native structure of the \( \text{Psi} \) RNA [61], including the native ssPurines conformation, preventing \( \text{Pr78}^{\text{Gag}} \) binding to it. As is evident from our footprinting data on the WT gRNA, \( \text{Pr78}^{\text{Gag}} \) binding to the ssPurines is limited to the nucleotides at the apical end of the loop (Figure 4B-D), suggesting the requirement of a spatially-accessible loop structure. Secondary structural analysis of the spliced \( \text{env} \) 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 \( \text{Pr78}^{\text{Gag}} \) binding. Such a conformational change results in unpairing of 6 of the 9 nucleotides of the ssPurines which showed Gag footprints (Figure 6). This suggests partial accessibility of \( \text{Pr78}^{\text{Gag}} \) to the spliced \( \text{env} \) RNA, which may explain the 30% bound RNA being observed upon competitive band shift assays performed for \( \text{env} \) spliced RNA (Figure 2D). Furthermore, since spliced \( \text{env} \) RNA is not packaged, these results suggest that potential limited Gag binding to spliced \( \text{env} \) RNA may allow its initial capture from the distinctly non-specific cellular RNAs (Figure 7). However, such inadequate binding may not be enough for Gag multimerization on the spliced \( \text{env} \) RNA necessary for encapsidation. On the other hand, the full length unspliced RNA, having an additional Gag binding GU-rich region (\( \text{G}^{254}\text{UGUU}^{258} \)) downstream of the mSD (Figure 4B-D), may facilitate Gag multimerization onto the unspliced RNA, resulting in its preferential packaging over the spliced RNA (Figure 7). Thus, the overall conformational change of ssPurines in the spliced \( \text{env} \) RNA, the restriction exerted by the partially base paired nucleotides for \( \text{Pr78}^{\text{Gag}} \) 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.

Conclusion

In conclusion, our study demonstrates that Pr78\textsuperscript{Gag} binds to two single-stranded loops positioned oppositely on the unspliced MPMV Psi RNA consisting of the ssPurines, and the A\textsuperscript{252}-AGUGUU\textsuperscript{258} loop which includes two nucleotides of the bpPurines and an adjacent GU-rich motif (G\textsuperscript{254}-UGUU\textsuperscript{258}; Figure 7). Based on structural differences, our findings also demonstrate that Pr78\textsuperscript{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.

Materials & Methods

Nucleotide numbering system

Nucleotide numbers in this study refer to the MPMV genome with the Genbank accession number M12349.1 [76].

Expression and purification of Pr78\textsuperscript{Gag}

Pr78\textsuperscript{Gag} was expressed with a C-terminal hexa-histidine (His\textsubscript{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 [62].

**Physical characterization of Pr78**

Prior to the *in vitro* assays, purified Pr78 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 [9]. 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 $\mu$ is the viscosity of the solvent. Before sample acquisition, the buffer was filtered through 0.02 µm filters (Millex ®) and the offset of the solvent was measured for subsequent sample data treatment.

**Plasmid construction for spliced env, ss- and bp-Purines mutant RNA production**

The wild type (WT) plasmid (RCR001; Figure 2A) was used for the *in vitro* transcription of the MPMV unspliced full-length *Psi* RNA, as previously described [24]. Plasmids for the *in vitro* transcription of the MPMV spliced *env* RNA (FN42; Figure 2A) and the ss- and bp-Purines mutants were created using spliced overlap extension (SOE) PCR, as previously described (Figure 3A; [24,27,61,77]). For cloning FN42, PCR A was performed using the outer forward
primer or sense (S) primer OTR 1004 (Supplemental table 3) and the inner reverse primer or anti-sense (AS) primer OTR 1458 (Supplemental table 3) using MPMV sub-genomic transfer vector (SJ2) as the template [24,27,60,61]. PCR B was performed employing the inner forward primer (S) OTR 1378 (Supplemental table 3) and the outer reverse primer, (AS) OTR 1379 (Supplemental table 3), on the MPMV full-length molecular clone KAL01 as the template [53]. 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 HindIII and Xmal and ligated into the similarly-digested pUC-based vector, pIC19R [24,27,60,61,63]. 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 Supplemental table 3.

**In vitro transcription and purification of unlabeled and [α-32P]-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 SmaI and used for in vitro transcription (MEGAscript™ T7 Transcription Kit, Thermo Fisher Scientific). Briefly, the linearized plasmids were incubated at 37°C 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 [9,78–82]. 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-32P]-ATP was performed to prepare internally-labeled RNAs, as described previously [9,79–82]. 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 4°C, followed by ethanol precipitation, and resuspension in 10 μl Milli-Q water.

**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 MgCl₂, 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, Pr78Gag (diluted in 30 mM Tris (pH 8.0), 300 mM NaCl, 5 mM MgCl₂, 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 MgCl₂) 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.

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

In the Hill’s equation, \(B_{\text{max}}\) 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 10 \(\mu\)l, and incubated at 37\(^\circ\)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\(^\circ\)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.
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 [9,24,27,61,63,75]. 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\textsuperscript{Gag} footprints on MPMV gRNAs, RNA-protein interactions were monitored by modifying WT RNA with BzCN, both in the absence and presence of Pr78\textsuperscript{Gag}. The nucleotides showing attenuated hSHAPE reactivity in the presence of Pr78\textsuperscript{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 [9] and MMTV [Chameettachal A. \textit{et al.}, to be published] 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\textsuperscript{Gag}, as previously described [9]. \textit{In vitro} transcribed WT gRNA (RCR001; 1 pmol) and a four-molar excess of competitor RNA (spliced \textit{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\textsuperscript{Gag} protein samples were thawed on ice and spun at 13,500 rpm for 15 min. The 20 µM Pr78\textsuperscript{Gag} stock was diluted in 5X HEPES buffer and mixed with the refolded RNA at a final
concentration of 6 µM in 20 µl. RNA in the folding buffer without any Pr78\(^{\text{Gag}}\) was used as a control. The RNA-protein complexes were incubated at 37ºC 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 [9,24].

Briefly, reverse transcription of the modified and unmodified RNAs was carried out, using two sets of primers (OTR 18/19 and 22/23). One primer within each set was labeled with either VIC (OTR18 and OTR22) or NED (OTR19 and OTR23) (Supplemental table 3). For elongation of both the modified and unmodified samples, 1 µl of each of 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 60 ºC for 10 min. For the ddG sequencing ladder, 2 pmol of untreated WT RNA (RCR001) and 1 µl of the 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 of 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 90°C 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 [83] 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 (Supplemental tables 1 and 2). Reactivity data obtained from samples treated in the absence of Pr78\textsuperscript{Gag} was applied as constraints to either the WT MPMV sequence (nt 232 to nt 1171) or FN26 sequence in RNAstructure (version 6.1; [84]). 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; [85]) where the validated structure was redrawn. The reactivity data obtained in the presence of Pr78\textsuperscript{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\textsuperscript{Gag} was determined using a paired two tailed \( t \)-test and a decrease or increase in hSHAPE reactivity with a \( p \)-value less than \( \leq 0.05 \) was considered significant (Supplemental tables 1 and 2).

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; Supplemental table 3). QuShape analysis was performed as above and mean reactivity data from three independent experiments were obtained for secondary structure analysis (Supplemental table 4).

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

T.A.R., R.M., and F.M. conceived the concept; T.A.R., F.M. and R.M. supervised the project; F.N.N.P., A.C., V.V-B., S.B., L.M.A., V.N.P. and A.K., performed the experiments; F.N.N.P., A.C., V.V-B., S.B., L.M.A., V.N.P., A.K., R.M. and T.A.R. contributed towards data analysis. F.N.N.P. wrote the manuscript. T.A.R., F.M., R.M., V.V-B., S.B., L.M.A., F.N.N.P. and A.C. contributed towards reviewing and editing the manuscript. All authors discussed the results and commented on the manuscript.

Funding Details

This research was funded by grants from the United Arab Emirates University (UAEU) Program for Advanced Research-UPAR (UPAR-31M233) to T.A.R., as well as by the RetroPack International Research Project from the CNRS to R.M. F.N.N.P. and A.C. were supported by UPAR-31M233 and UAE University Zayed Bin Sultan Center for Health Sciences (UCBR-31R123) grants, respectively.
Figure legends

Figure 1. Schematic representations of MPMV genome, organization of the different domains of full-length MPMV Gag precursor (Pr78Gag), and higher order structure of MPMV packaging signal RNA. (A) Organization of the MPMV genome, domain organization of MPMV Gag precursor (Pr78Gag), and 5’ end of the genome that contains the packaging signal (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; GU rich region; Gag SL1-2, Group antigen stem loops 1-2.

Figure 2. Pr78Gag 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 Pr78Gag. 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); R²= 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.

**Figure 3.** Single-stranded and base-paired purines (ss- and bpPurines) mutants of the MPMV packaging signal RNA bind differentially to Pr78Gag. (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 [63]. 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 Pr78Gag 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.

**Figure 4.** Footprints of Pr78Gag 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 Pr78Gag. The mean triplicate SHAPE reactivity obtained in the absence of Pr78Gag was used to draw RNA secondary structure of WT (RCR001) RNA. This was followed by applying the mean hSHAPE reactivities obtained in presence of Pr78Gag onto the secondary structure obtained in the absence of Pr78Gag. (C) and (D) regions of interests showing SHAPE reactivities in two major regions: (1) the ss-
Purines loop and (2) a loop corresponding to the last two bpPurines and the adjacent GU-rich region on the wild type MPMV RNA (RCR001) both in the absence and presence of Pr78\textsuperscript{Gag} respectively. (E) Histograms showing the Pr78\textsuperscript{Gag}-induced attenuation of SHAPE reactivities of nucleotides in the ss-, bp-Purines, and the GU-rich region. 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 ($p \leq 0.06-0.08$) are indicated by grey arrows.

Figure 5. Pr78\textsuperscript{Gag} footprints on mutant RNA containing ssPurines deletion (FN26) shows binding to the last nucleotide of the bpPurines and the adjacent GU-rich region. (A) Mean triplicate SHAPE reactivity obtained in the absence of Pr78\textsuperscript{Gag} was used to draw RNA secondary structure of FN26. (B) Mean SHAPE reactivities obtained in presence of 6 $\mu$M of Pr78\textsuperscript{Gag} were applied onto the secondary structure obtained in the absence of Pr78\textsuperscript{Gag}. (C) and (D) regions of interests showing SHAPE reactivities on the last nucleotides of the bpPurines and the adjacent GU-rich region in FN26 RNA both in the absence and presence of Pr78\textsuperscript{Gag} respectively. (E) Histograms showing the Pr78\textsuperscript{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.
Figure 6. SHAPE validated secondary structure of the MPMV spliced \textit{env} RNA (FN42).

The ssPurines attain a partially base paired conformation in the secondary structure of the spliced \textit{env} RNA. Nucleotides showing footprints on the wild type unspliced RNA (RCR001) have been superimposed on the SHAPE validated spliced \textit{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 ($p \leq 0.06\text{-}0.08$) are indicated by grey arrows.

Figure 7. Model proposing how MPMV Pr78\textsuperscript{Gag} captures \textit{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 ($Psi/\Psi$) of MPMV gRNA spans from R to the first 120 bp of \textit{gag} that folds into a unique secondary structure. Pr78\textsuperscript{Gag} binds to two single-stranded regions within the MPMV \textit{Psi} RNA: (1) the ssPurines loop (U\textsuperscript{191}UAAAAGUGAAAGUAA\textsuperscript{206}), located upstream the mSD and hence found in all viral RNAs, and (2) the single-stranded A\textsuperscript{252}AGUGUU\textsuperscript{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\textsuperscript{Gag} may first bind to the ssPurines loop to distinguish viral RNAs from cellular mRNAs. This is probably followed by binding to the A\textsuperscript{252}AGUGUU\textsuperscript{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\textsuperscript{252}AGUGUU\textsuperscript{258} loop. This ensures that the MPMV gRNA is packaged preferentially into the viral particles, excluding cellular and spliced RNAs.

**Supplemental Figure 1.** Characterization of MPMV Gag precursor Pr\textsuperscript{78Gag} 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 Pr\textsuperscript{78Gag} samples tested in binding buffer.

**Supplemental Table 1.** Mean hSHAPE reactivity data from triplicate hSHAPE experiments on wild type unspliced RNA (RCR001) both in the absence as well presence of Pr\textsuperscript{78Gag} with standard deviation (SD) and \( p \)-values. Color key: green, single-stranded purines (ssPurines); blue, base-paired purines (bpPurines); brick red, GU rich region.

**Supplemental Table 2.** Mean hSHAPE reactivity data from triplicate hSHAPE experiments on ssPurines deleted mutant RNA (FN26) both in the absence as well presence of Pr\textsuperscript{78Gag} with standard deviation (SD) and \( p \)-values. Color key: blue, base-paired purines (bpPurines) and brick red, GU rich region.

**Supplemental Table 3.** Description of primers and DNA templates used in cloning, sequencing and hSHAPE.

**Supplemental Table 4.** Mean SHAPE reactivity data from triplicate hSHAPE experiments on the spliced \textit{env} RNA (FN42) with standard deviation (SD).
References

[1] J.-C. Paillart, M. Shehu-Xhilaga, R. Marquet, J. Mak, Dimerization of retroviral RNA genomes: an inseparable pair, Nat. Rev. Microbiol. 2 (2004) 461–472. https://doi.org/10.1038/nrmicro903.

[2] V. D’Souza, M.F. Summers, How retroviruses select their genomes, Nat. Rev. Microbiol. 3 (2005) 643. https://doi.org/10.1038/nrmicro1210.

[3] M.D. Moore, W.S. Hu, HIV-1 RNA dimerization: It takes two to tango, AIDS Rev. 11 (2009) 91–102.

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

[5] K.A. Dilley, N. Ni, O.A. Nikolaitchik, J. Chen, A. Galli, W.-S. Hu, Determining the Frequency and Mechanisms of HIV-1 and HIV-2 RNA Copackaging by Single-Virion Analysis, J. Virol. 85 (2011) 10499–10508. https://doi.org/10.1128/JVI.05147-11.

[6] A. Rein, RNA Packaging in HIV, Trends Microbiol. 27 (2019) 715–723. https://doi.org/10.1016/j.tim.2019.04.003.

[7] R. Berkowitz, J. Fisher, S.P. Goff, RNA packaging, Curr. Top. Microbiol. Immunol. 214 (1996) 177–218. https://doi.org/10.1007/978-3-642-80145-7_6.

[8] A.M.L. Lever, HIV-1 RNA packaging, Adv. Pharmacol. San Diego Calif. 55 (2007) 1–32. https://doi.org/10.1016/S1054-3589(07)55001-5.
[9] E.W. Abd El-Wahab, R.P. Smyth, E. Mailler, S. Bernacchi, V. Vivet-Boudou, M. Hijnen, F. Jossinet, J. Mak, J.-C. Paillart, R. Marquet, Specific recognition of the HIV-1 genomic RNA by the Gag precursor, Nat. Commun. 5 (2014) 4304. https://doi.org/10.1038/ncomms5304.

[10] M. Kuzembayeva, K. Dilley, L. Sardo, W.-S. Hu, Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles, Virology. 454–455 (2014) 362–370. https://doi.org/10.1016/j.virol.2014.01.019.

[11] R.J. Kaddis Maldonado, L.J. Parent, Orchestrating the Selection and Packaging of Genomic RNA by Retroviruses: An Ensemble of Viral and Host Factors, Viruses. 8 (2016). https://doi.org/10.3390/v8090257.

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

[13] N. Dubois, R. Marquet, J.-C. Paillart, S. Bernacchi, Retroviral RNA Dimerization: From Structure to Functions, Front. Microbiol. 9 (2018) 527. https://doi.org/10.3389/fmicb.2018.00527.

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

[15] S. Bernacchi, E.W. Abd El-Wahab, N. Dubois, M. Hijnen, R.P. Smyth, J. Mak, R. Marquet, J.-C. Paillart, HIV-1 Pr55Gagbinds genomic and spliced RNAs with different affinity and
stoichiometry, RNA Biol. 14 (2017) 90–103.
https://doi.org/10.1080/15476286.2016.1256533.

[16] J.C. Kenyon, A. Ghazawi, W.K.S. Cheung, P.S. Phillip, T.A. Rizvi, A.M.L. Lever, 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 (2008) 2597–2608. https://doi.org/10.1261/rna.1284908.

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

[18] C. Gherghe, T. Lombo, C.W. Leonard, S.A.K. Datta, J.W. Bess, R.J. Gorelick, A. Rein, K.M. Weeks, Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 19248–19253. https://doi.org/10.1073/pnas.1006897107.

[19] T.A. Rizvi, J.C. Kenyon, J. Ali, S.J. Aktar, P.S. Phillip, A. Ghazawi, F. Mustafa, A.M.L. Lever, Optimal packaging of FIV genomic RNA depends upon a conserved long-range interaction and a palindromic sequence within gag, J. Mol. Biol. 403 (2010) 103–119. https://doi.org/10.1016/j.jmb.2010.08.019.

[20] Y. Miyazaki, A. Miyake, M. Nomaguchi, A. Adachi, Structural dynamics of retroviral genome and the packaging, Front. Microbiol. 2 (2011) 264. https://doi.org/10.3389/fmicb.2011.00264.
[21] M. Comas-Garcia, S.R. Davis, A. Rein, On the Selective Packaging of Genomic RNA by HIV-1, Viruses. 8 (2016) 246. https://doi.org/10.3390/v8090246.

[22] J.A. Webb, C.P. Jones, L.J. Parent, I. Rouzina, K. Musier-Forsyth, Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: Implications for viral genomic RNA packaging, RNA. 19 (2013) 1078–1088. https://doi.org/10.1261/rna.038869.113.

[23] J.C. Kenyon, S.J. Tanner, M. Legiewicz, P.S. Phillip, T.A. Rizvi, S.F.J. Le Grice, A.M.L. Lever, SHAPE analysis of the FIV Leader RNA reveals a structural switch potentially controlling viral packaging and genome dimerization, Nucleic Acids Res. 39 (2011) 6692–6704. https://doi.org/10.1093/nar/gkr252.

[24] S.J. Aktar, A. Jabeen, L.M. Ali, V. Vivet-Boudou, R. Marquet, T.A. Rizvi, SHAPE analysis of the 5′ end of the Mason-Pfizer monkey virus (MPMV) genomic RNA reveals structural elements required for genome dimerization, RNA. 19 (2013) 1648–1658. https://doi.org/10.1261/rna.040931.113.

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

[26] R.P. Smyth, L. Despons, G. Huili, S. Bernacchi, M. Hijnen, J. Mak, F. Jossinet, L. Weixi, J.-C. Paillart, M. von Kleist, R. Marquet, Mutational interference mapping experiment (MIME) for studying RNA structure and function, Nat. Methods. 12 (2015) 866–872. https://doi.org/10.1038/nmeth.3490.
[27] R.M. Kalloush, V. Vivet-Boudou, L.M. Ali, V.N. Pillai, F. Mustafa, R. Marquet, T.A. Rizvi, 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 (2019) 612–625. https://doi.org/10.1080/15476286.2019.1572424.

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

[29] R.J. Fisher, A. Rein, M. Fivash, M.A. Urbaneja, J.R. Casas-Finet, M. Medaglia, L.E. Henderson, Sequence-Specific Binding of Human Immunodeficiency Virus Type 1 Nucleocapsid Protein to Short Oligonucleotides, J. Virol. 72 (1998) 1902–1909. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC109481/ (accessed September 14, 2020).

[30] G.K. Amarasinghe, R.N. De Guzman, R.B. Turner, K.J. Chancellor, Z.R. Wu, M.F. Summers, NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition, J. Mol. Biol. 301 (2000) 491–511. https://doi.org/10.1006/jmbi.2000.3979.

[31] Morcock, K. Bp, C.-F. Jr, Fluorescence and nucleic acid binding properties of the human T-cell leukemia virus-type 1 nucleocapsid protein., Biochim. Biophys. Acta. 1481 (2000) 381–394. https://doi.org/10.1016/s0167-4838(00)00181-3.

[32] D.R. Morcock, S. Katakam, B.P. Kane, J.R. Casas-Finet, Fluorescence and nucleic acid binding properties of bovine leukemia virus nucleocapsid protein, Biophys. Chem. 97 (2002) 203–212. https://doi.org/10.1016/S0301-4622(02)00070-4.
[33] M.A. Urbaneja, C.F. McGrath, B.P. Kane, L.E. Henderson, J.R. Casas-Finet, Nucleic acid binding properties of the simian immunodeficiency virus nucleocapsid protein NCp8, J. Biol. Chem. 275 (2000) 10394–10404.

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

[35] J. Zhou, R.L. Bean, V.M. Vogt, M.F. Summers, Solution Structure of the Rous Sarcoma Virus Nucleocapsid Protein:μΨ RNA Packaging Signal Complex, J. Mol. Biol. 365 (2007) 453–467. https://doi.org/10.1016/j.jmb.2006.10.013.

[36] K.M. Stewart-Maynard, M. Cruceanu, F. Wang, M.-N. Vo, R.J. Gorelick, M.C. Williams, I. Rouzina, K. Musier-Forsyth, Retroviral Nucleocapsid Proteins Display Nonequivalent Levels of Nucleic Acid Chaperone Activity, J. Virol. 82 (2008) 10129–10142. https://doi.org/10.1128/JVI.01169-08.

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

[38] A. Alfadhli, H. McNett, S. Tsagli, H.P. Bächinger, D.H. Peyton, E. Barklis, HIV-1 MATRIX PROTEIN BINDING TO RNA, J. Mol. Biol. 410 (2011) 653–666. https://doi.org/10.1016/j.jmb.2011.04.063.

[39] K. Lu, X. Heng, L. Garyu, S. Monti, E.L. Garcia, S. Kharytonchyk, B. Dorjsuren, G. Kulandaivel, S. Jones, A. Hiremath, S.S. Divakaruni, C. LaCotti, S. Barton, D. Tummillo,
A. Hosic, K. Edme, S. Albrecht, A. Telesnitsky, M.F. Summers, NMR detection of structures in the HIV-1 5’-leader RNA that regulate genome packaging, Science. 334 (2011) 242–245. https://doi.org/10.1126/science.1210460.

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

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

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

[43] X. Guo, B.B. Roy, J. Hu, A. Roldan, M.A. Wainberg, C. Liang, The R362A mutation at the C-terminus of CA inhibits packaging of human immunodeficiency virus type 1 RNA, Virology. 343 (2005) 190–200. https://doi.org/10.1016/j.virol.2005.08.031.

[44] B.B. Roy, R.S. Russell, D. Turner, C. Liang, 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 (2006) 304–314. https://doi.org/10.1016/j.virol.2005.09.011.

[45] J.F. Kaye, A.M.L. Lever, Nonreciprocal Packaging of Human Immunodeficiency Virus Type 1 and Type 2 RNA: a Possible Role for the p2 Domain of Gag in RNA Encapsidation,
J. Virol. 72 (1998) 5877–5885. http://jvi.asm.org/content/72/7/5877 (accessed January 7, 2018).

[46] R.S. Russell, A. Roldan, M. Detorio, J. Hu, M.A. Wainberg, C. Liang, Effects of a Single Amino Acid Substitution within the p2 Region of Human Immunodeficiency Virus Type 1 on Packaging of Spliced Viral RNA, J. Virol. 77 (2003) 12986–12995. https://doi.org/10.1128/JVI.77.24.12986-12995.2003.

[47] Y. Zhang, E. Barklis, Effects of nucleocapsid mutations on human immunodeficiency virus assembly and RNA encapsidation, J. Virol. 71 (1997) 6765–6776. https://doi.org/10.1128/JVI.71.9.6765-6776.1997.

[48] N. Dubois, K.K. Khoo, S. Ghossein, T. Seissler, P. Wolff, W.J. McKinstry, J. Mak, J.-C. Paillart, R. Marquet, S. Bernacchi, The C-terminal p6 domain of the HIV-1 Pr55Gag precursor is required for specific binding to the genomic RNA, RNA Biol. 15 (2018) 923–936. https://doi.org/10.1080/15476286.2018.1481696.

[49] D.L. Fine, J.C. Landon, R.J. Pienta, M.T. Kubicek, M.G. Valerio, W.F. Loeb, H.C. Chopra, Responses of infant rhesus monkeys to inoculation with Mason-Pfizer monkey virus materials, J. Natl. Cancer Inst. 54 (1975) 651–658.

[50] M.L. Bryant, M.B. Gardner, P.A. Marx, D.H. Maul, N.W. Lerche, K.G. Osborn, L.J. Lowenstine, A. Bodgen, L.O. Arthur, E. Hunter, Immunodeficiency in Rhesus Monkeys Associated With the Original Mason-Pfizer Monkey Virus, JNCI J. Natl. Cancer Inst. 77 (1986) 957–965. https://doi.org/10.1093/jnci/77.4.957.
[51] M. Bray, S. Prasad, J.W. Dubay, E. Hunter, K.T. Jeang, D. Rekosh, M.L. Hammarskjöld, A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 1256–1260.

[52] T.A. Rizvi, K.A. Lew, E.C. Murphy, R.D. Schmidt, 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 (1996) 517–532.

[53] T.A. Rizvi, R.D. Schmidt, K.A. Lew, M.E. Keeling, Rev/RRE-Independent Mason–Pfizer Monkey Virus Constitutive Transport Element-Dependent Propagation of SIVmac239 Vectors Using a Single Round of Replication Assay, Virology. 222 (1996) 457–463. https://doi.org/10.1006/viro.1996.0444.

[54] T.A. Rizvi, R.D. Schmidt, K.A. Lew, Mason–Pfizer Monkey Virus (MPMV) Constitutive Transport Element (CTE) Functions in a Position-Dependent Manner, Virology. 236 (1997) 118–129. https://doi.org/10.1006/viro.1997.8728.

[55] D.L. Bush, V.M. Vogt, In Vitro Assembly of Retroviruses, Annu. Rev. Virol. 1 (2014) 561–580. https://doi.org/10.1146/annurev-virology-031413-085427.

[56] R.G. Vile, M. Ali, E. Hunter, M.O. Mcclure, Identification of a generalised packaging sequence for D-type retroviruses and generation of a D-type retroviral vector, Virology. 189 (1992) 786–791. https://doi.org/10.1016/0042-6822(92)90607-Q.
[57] G.P. Harrison, E. Hunter, A.M. Lever, Secondary structure model of the Mason-Pfizer monkey virus 5′ leader sequence: identification of a structural motif common to a variety of retroviruses., J. Virol. 69 (1995) 2175–2186. http://jvi.asm.org/content/69/4/2175 (accessed January 7, 2018).

[58] F.M.J. Guesdon, J. Greatorex, S.R. Rhee, R. Fisher, E. Hunter, A.M.L. Lever, 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 (2001) 81–88. https://doi.org/10.1006/viro.2001.1061.

[59] R.D. Schmidt, F. Mustafa, K.A. Lew, M.T. Browning, T.A. Rizvi, Sequences within both the 5′ untranslated region and the gag gene are important for efficient encapsidation of Mason-Pfizer monkey virus RNA, Virology. 309 (2003) 166–178.

[60] S.A. Jaballah, S.J. Aktar, J. Ali, P.S. Phillip, N.S. Al Dhaheri, A. Jabeen, T.A. Rizvi, A G–C-Rich Palindromic Structural Motif and a Stretch of Single-Stranded Purines Are Required for Optimal Packaging of Mason–Pfizer Monkey Virus (MPMV) Genomic RNA, J. Mol. Biol. 401 (2010) 996–1014. https://doi.org/10.1016/j.jmb.2010.06.043.

[61] R.M. Kalloush, V. Vivet-Boudou, L.M. Ali, F. Mustafa, R. Marquet, T.A. Rizvi, Packaging of Mason-Pfizer monkey virus (MPMV) genomic RNA depends upon conserved long-range interactions (LRIs) between U5 and gag sequences, RNA. 22 (2016) 905–919. https://doi.org/10.1261/rna.055731.115.
[62] F.N.N. Pitchai, L. Ali, V.N. Pillai, A. Chameettachal, S.S. Ashraf, F. Mustafa, R. Marquet, T.A. Rizvi, Expression, purification, and characterization of biologically active full-length Mason-Pfizer monkey virus (MPMV) Pr78 Gag, Sci. Rep. 8 (2018) 11793.

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

[64] J.C. Paillart, E. Westhof, C. Ehresmann, B. Ehresmann, R. Marquet, Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA, J. Mol. Biol. 270 (1997) 36–49. https://doi.org/10.1006/jmbi.1997.1096.

[65] A. Zeffman, S. Hassard, G. Varani, A. Lever, The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein, J. Mol. Biol. 297 (2000) 877–893. https://doi.org/10.1006/jmbi.2000.3611.

[66] A.M. Lever, RNA packaging in lentiviruses, Retrovirology. 6 (2009) I13.

[67] S.C. Keane, X. Heng, K. Lu, S. Kharytonchyk, V. Ramakrishnan, G. Carter, S. Barton, A. Hosic, A. Florwick, J. Santos, N.C. Bolden, S. McCowin, D.A. Case, B.A. Johnson, M. Salemi, A. Telesnitsky, M.F. Summers, RNA structure. Structure of the HIV-1 RNA packaging signal, Science. 348 (2015) 917–921. https://doi.org/10.1126/science.aaa9266.

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

[69] O.A. Nikolaitchik, X. Somoulay, J.M.O. Rawson, J.A. Yoo, V.K. Pathak, W.-S. Hu, Unpaired Guanosines in the 5' Untranslated Region of HIV-1 RNA Act Synergistically to Mediate Genome Packaging, J. Virol. (2020). https://doi.org/10.1128/JVI.00439-20.

[70] S.S. Rhee, E. Hunter, 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 (1990) 77–86. https://doi.org/10.1016/0092-8674(90)90289-Q.

[71] R. Aronoff, M. Linial, Specificity of retroviral RNA packaging, J. Virol. 65 (1991) 71–80.

[72] R. Aronoff, A.M. Hajjar, M.L. Linial, Avian retroviral RNA encapsidation: reexamination of functional 5’ RNA sequences and the role of nucleocapsid Cys-His motifs, J. Virol. USA. (1993). https://agris.fao.org/agris-search/search.do?recordID=US9400227 (accessed September 14, 2020).

[73] J.D. Banks, B.O. Kealoha, M.L. Linial, An MΨ-Containing Heterologous RNA, but Not env mRNA, Is Efficiently Packaged into Avian Retroviral Particles, J. Virol. 73 (1999) 8926–8933. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC112923/ (accessed June 2, 2020).

[74] P.G. Stockley, S.J. White, E. Dykeman, I. Manfield, O. Rolfsson, N. Patel, R. Bingham, A. Barker, E. Wroblewski, R. Chandler-Bostock, E.U. Weiβ, N.A. Ranson, R. Tuma, R. Twarock, Bacteriophage MS2 genomic RNA encodes an assembly instruction manual for its capsid, Bacteriophage. 6 (2016). https://doi.org/10.1080/21597081.2016.1157666.
[75] F. Mustafa, V. Vivet-Boudou, A. Jabeen, L.M. Ali, R.M. Kalloush, R. Marquet, T.A. Rizvi,
The bifurcated stem loop 4 (SL4) is crucial for efficient packaging of mouse mammary
tumor virus (MMTV) genomic RNA, RNA Biol. 15 (2018) 1047–1059.
https://doi.org/10.1080/15476286.2018.1486661.

[76] P. Sonigo, C. Barker, E. Hunter, S. Wain-Hobson, Nucleotide sequence of Mason-Pfizer
monkey virus: An immunosuppressive D-type retrovirus, Cell. 45 (1986) 375–385.
https://doi.org/10.1016/0092-8674(86)90323-5.

[77] J.S. Gibbs, D.A. Regier, R.C. Desrosiers, Construction and in vitro properties of SIVmac
mutants with deletions in “nonessential” genes, AIDS Res. Hum. Retroviruses. 10 (1994)
333–342. https://doi.org/10.1089/aid.1994.10.333.

[78] R. Marquet, F. Baudin, C. Gabus, J.L. Darlix, M. Mougel, C. Ehresmann, B. Ehresmann,
Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and
possible mechanism, Nucleic Acids Res. 19 (1991) 2349–2357.
https://doi.org/10.1093/nar/19.9.2349.

[79] J.C. Paillart, R. Marquet, E. Skripkin, B. Ehresmann, C. Ehresmann, Mutational analysis of
the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic
RNA, J. Biol. Chem. 269 (1994) 27486–27493.

[80] J.C. Paillart, L. Berthoux, M. Ottmann, J.L. Darlix, R. Marquet, B. Ehresmann, C.
Ehresmann, A dual role of the putative RNA dimerization initiation site of human
immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis., J.
Virol. 70 (1996) 8348–8354. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC190922/ (accessed June 3, 2020).

[81] J.C. Paillart, E. Skripkin, B. Ehresmann, C. Ehresmann, R. Marquet, A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 5572–5577.

[82] L. Sinck, D. Richer, J. Howard, M. Alexander, D.F.J. Purcell, R. Marquet, J.-C. Paillart, In vitro dimerization of human immunodeficiency virus type 1 (HIV-1) spliced RNAs, RNA N. Y. N. 13 (2007) 2141–2150. https://doi.org/10.1261/rna.678307.

[83] F. Karabiber, J.L. McGinnis, O.V. Favorov, K.M. Weeks, QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis, RNA. 19 (2013) 63–73. https://doi.org/10.1261/rna.036327.112.

[84] J.S. Reuter, D.H. Mathews, RNAstructure: software for RNA secondary structure prediction and analysis, BMC Bioinformatics. 11 (2010) 129. https://doi.org/10.1186/1471-2105-11-129.

[85] K. Darty, A. Denise, Y. Ponty, VARNA: Interactive drawing and editing of the RNA secondary structure, Bioinformatics. 25 (2009) 1974–1975. https://doi.org/10.1093/bioinformatics/btp250.
Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
CRediT AUTHOR STATEMENT

Tahir A. Rizvi, Roland Marquet, and Farah Mustafa: Conceptualization, Supervision, Writing- Reviewing and Editing. Fathima N.N Pitchai: Investigation, Methodology, Visualization, Software, Validation, Writing- Original draft preparation, Writing- Reviewing and Editing. Akhil Chameettachal, Valérie Vivet-Boudou, Lizna M. Ali, Vineeta N. Pillai, Anjana Krishnan, Serena Bernacchi: Investigation, Methodology, Visualization, Software, Validation, Writing- Reviewing and Editing.

All authors discussed the results and commented on the manuscript.
Research Highlights

- MPMV Pr78Gag captures Psi RNA structure from a mix of cellular/spliced viral RNAs
- Pr78Gag binds to ssPurines and A^{252}AGUGUU^{258} loop within the MPMV Psi RNA
- Both these looped regions act redundantly during the gRNA packaging process
- ssPurines are located upstream of the major splice donor & found in all viral RNAs
- Splicing of A^{252}AGUGUU^{258} in env RNA ensures packaging of only the unspliced gRNA