SHAPE analysis of the 5′ end of the Mason-Pfizer monkey virus (MPMV) genomic RNA reveals structural elements required for genome dimerization

SURIYA J. AKTAR,[1] **AYESHA JABEEN,**[1] **LIZNA M. ALI,**[1] **VALÉRIE VIVET-BOUDOU,**[2] **ROLAND MARQUET,**[2,3] **and TAHIR A. RIZVI**[1,3]

[1]Department of Microbiology and Immunology, College of Medicine and Health Sciences (CMHS), United Arab Emirates University (UAEU), Al Ain, United Arab Emirates
[2]Architecture et Réactivité de l'ARN, Université de Strasbourg, IBMC, CNRS, 67084 Strasbourg, France

ABSTRACT

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

Keywords: Mason Pfizer monkey virus (MPMV); RNA dimerization; dimerization initiation site (DIS); RNA packaging signal; long-range interaction (LRI)

INTRODUCTION

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

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

It is becoming increasingly clear that palindromic (pal) sequences play a crucial role in gRNA dimerization: For human immunodeficiency type 1 and type 2 viruses (HIV-1 and HIV-2), feline immunodeficiency virus (FIV), Moloney murine leukemia virus (MoMuLV), and murine sarcoma

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

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

A stretch of 5′ end sequences of the prototype β-retrovirus Mason-Pfizer monkey virus (MPMV), which causes an immunodeficiency disease in newborn rhesus monkey, has been shown to be involved in RNA packaging (Vile et al. 1992; Harrison et al. 1995; Guesdon et al. 2001). Employing a combination of genetic and structural prediction analyses, we have recently shown that MPMV packaging determinants comprise two discontinuous core regions ("A" and "B") within the 5′ untranslated region (UTR) (Schmidt et al. 2003; Jaballah et al. 2010). In addition to these two regions in the UTR, a stretch of ∼120 nt of *gag* have also been shown to be needed for optimal MPMV RNA packaging (Schmidt et al. 2003; Jaballah et al. 2010). Based on these mutational studies and structural prediction analyses, a minimal free energy RNA secondary structure model has been proposed for the 5′ end (388 nt starting from R) of MPMV gRNA (Fig. 1A; Jaballah et al. 2010). This model suggested that the packaging determinants fold into several stable stem–loop (SL) structures and also revealed a 14-nt GC-rich pal sequence (5′-UCGCCGGCCGGCGA-3′) that folds into a hairpin with a canonical "GC" dyad in the loop, a 16-nt stretch of single-stranded purine (ssPurine)–rich region (8 nt of this ssPurine-rich region is duplicated as a base-paired sequence in an adjacent region), and LRIs between U5 and *gag* sequences (Jaballah et al. 2010). The 14-nt pal has been proposed to play a role in the initiation of gRNA dimerization (Jaballah et al. 2010). The ssPurine-rich region (or its repeat in region "B" when predicted to refold as the ssPurine-rich region) has been shown to be essential for RNA packaging, possibly functioning as a potential NC binding site. Two LRIs between U5 and *gag* sequences could potentially play a role in MPMV RNA packaging by maintaining the overall RNA secondary structure as has recently been shown in the case of FIV (Kenyon et al. 2008, 2011; Rizvi et al. 2010). Even though the deletion analysis of 5′ end of MPMV gRNA provides a plausible explanation for the discontinuous nature of MPMV packaging signals, the existence of the different structural motifs of the predicted RNA secondary structure of this region has not been validated experimentally.

RESULTS AND DISCUSSION

To validate the predicted structure (Jaballah et al. 2010) and to generate a more precise structural model of the MPMV gRNA packaging signals, we employed selective 2′ hydroxyl acylation analyzed by primer extension (SHAPE), which examines the flexibility of the RNA backbone at each nucleotide position (Merino et al. 2005; Mortimer and Weeks 2007, 2009). SHAPE reagents target and modify the ribose moiety of all 4 nt in single-stranded regions of a folded RNA molecule, whereas the nucleotides that are base-paired or architecturally constrained do not react or show reduced reactivity.

SHAPE data validated the major structural motifs predicted in the 5′ end of MPMV gRNA

Our SHAPE data validated the overall predicted structure (Jaballah et al. 2010) of the MPMV packaging signal RNA.
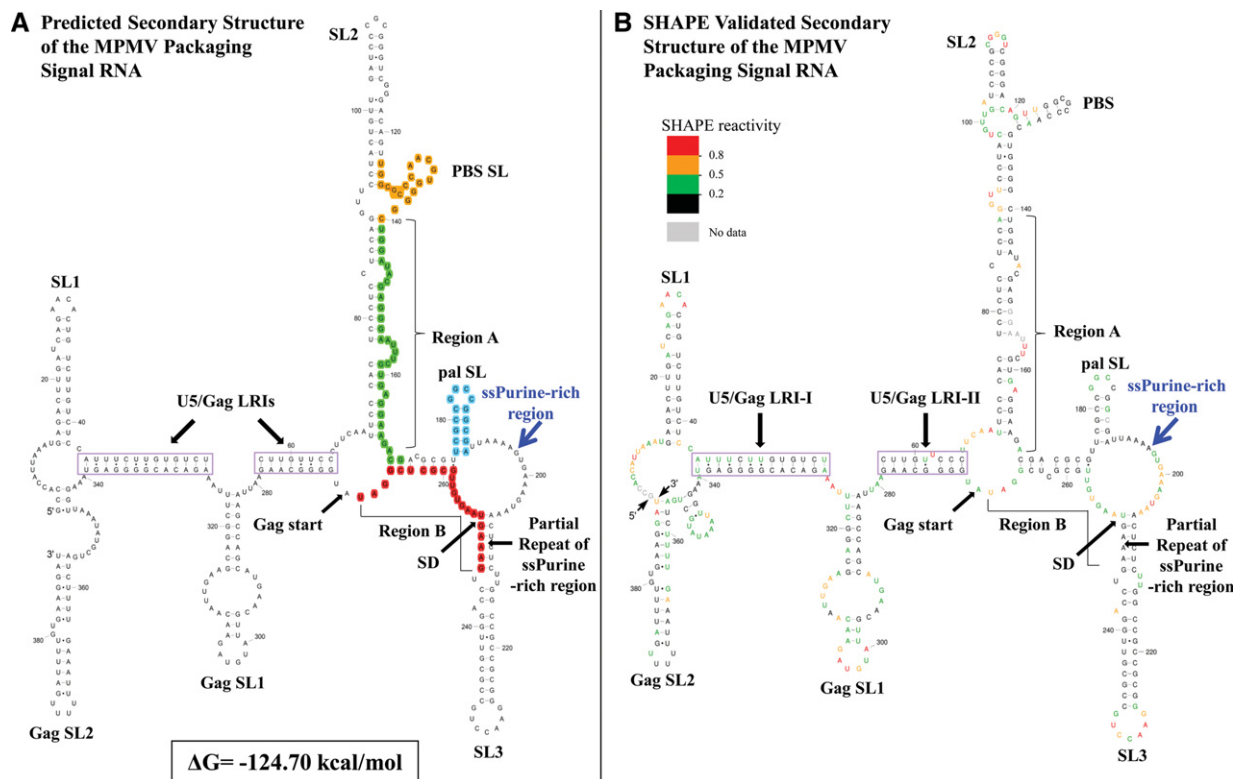
FIGURE 1. Minimal free-energy and SHAPE (selective 2′ hydroxyl acylation analyzed by primer extension)-validated models of the MPMV packaging signal RNA. The region used for analysis by Mfold and SHAPE included sequences from R up to 120 nt of *gag*. (*A*) MPMV packaging signal RNA secondary structure predicted earlier (Jaballah et al. 2010) using Mfold (Mathews et al. 1999; Zuker 2003). Sequences in orange, green, red, and blue represent the primer binding site (PBS), regions "A" and "B" (that have been shown to be important in gRNA packaging), and pal sequences, respectively. Boxed areas in purple show the predicted LRIs between U5 and *gag*. (*B*) SHAPE-constrained RNAstructure (Reuter and Mathews 2010) model of MPMV packaging signal. Nucleotides are color annotated as per the SHAPE reactivities key. SD indicates splice donor.

Figure 1A shows the unconstrained predicted structure, whereas Figure 1B shows the SHAPE-validated MPMV structure. An average of SHAPE reactivities numerical data for each nucleotide that were drawn from two to four experiments were applied as pseudo energy constraints in RNAstructure (Reuter and Mathews 2010; Bellaousov et al. 2013). As shown in Figure 1B, the SHAPE data robustly corroborated the free-energy minimized predicted structure of the MPMV packaging signal RNA presented in Figure 1A. Briefly, the SHAPE-validated model displayed major structural components, including pal SL, ssPurine-rich region, U5-Gag LRIs, SLs 1-3, and Gag SLs 1-2, in extensive conformity with the predicted structure. Consistent with the predicted structure, all the loops and bulges containing unpaired nucleotides showed SHAPE reactivity. A few base-paired nucleotides located at the ends of a helix or near a bulge or a loop also showed reactivity (e.g., nt 22, 25, and 26). Despite such a high conformity between the predicted and the SHAPE-validated structure models, some noticeable differences could be observed. First, SHAPE data indicated that the actual apical loop of SL2 is larger than predicted by Mfold (Fig. 1A,B). Second, this loop is shifted by 2 nt, inducing different base-pairing in the upper part of SL2, including the primer binding site

(PBS). Third, a small SL was observed in the SHAPE-validated structure at the base of Gag SL2, which was not present in the predicted structure model of the MPMV packaging signal RNA. Finally, even though the same pal SL is present in the two structures, it is interesting to observe that the nucleotides in the loop, which contains the "GC" dyad, displayed weak or no reactivity, suggesting that these nucleotides are indeed base-paired. One possibility is that these nucleotides may be base-paired with the same sequence on the other gRNA, which is consistent with the hypothesis that the pal SL might function as the DIS of MPMV gRNA (Jaballah et al. 2010).

A very distinguishing feature of the predicted and the SHAPE-validated structure are the two LRIs (LRI-I and LRI-II) involving sequences from U5 and first 70 nt of *gag* (Fig. 1), which have been shown to be important in MPMV RNA packaging (Schmidt et al. 2003; Jaballah et al. 2010). Except for canonical base pairs at both ends of LRI-I, and wobble base pairs in both LRIs, all nucleotides (except U62) in the complementary sequences of the LRIs were non-reactive to SHAPE reagent and, therefore, were base-paired as initially predicted (Jaballah et al. 2010). U62 is highly reactive to the SHAPE reagent and forms a bulge in LRI-II that was not predicted by Mfold.

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

Earlier studies have shown that deletion of the 14-nt pal SL severely compromised MPMV gRNA packaging (Schmidt et al. 2003; Jaballah et al. 2010). However, these studies did not study the impact of these mutations on gRNA dimerization. The conservation of the pal sequence in different strains of MPMV (Fig. 4F) and the fact that it forms a hairpin with a "GC" dyad in the loop (Fig. 1) makes it a credible candidate to function as MPMV gRNA DIS. Therefore, we introduced a series of mutations in the pal sequence (Fig. 2A) and performed in vitro dimerization assays using in vitro transcribed RNAs. The mutant RCR002 contains the complete deletion of the pal sequence. In order to ascertain whether preservation of the hairpin structure is sufficient to initiate dimerization, mutant RCR003 was created in which the central 4 nt of the loop were substituted with a stable tetra loop sequence. In

order to determine if it is the primary sequence or pal nature of the pal sequence that is required for MPMV gRNA dimerization, another mutant (RCR004) was constructed that contains the substitution of the central 6 nt of MPMV pal sequence with that of HIV-1 pal sequence. Before introducing these mutations, Mfold was used to predict the structure of the mutants in the monomer form to support the premise of our experimental hypothesis.

The data obtained from three independent in vitro dimerization experiments using wild-type and pal mutant RNAs are summarized in Figure 2, B and C. In vitro transcribed wild-type RNA showed ∼87% and 65% dimerization in TBM (50 mM Tris base, 45 mM boric acid, 0.1 mM $MgCl_2$; electrophoresis at 4°C) and TB (50 mM Tris base, 45 mM boric acid; electrophoresis at room temperature) gels, respectively (Fig. 2B). These wild-type (RCR001) values were considered as references for calculating the relative dimerization efficiency (RDE) of the mutants (Fig. 2C). The pal deletion and
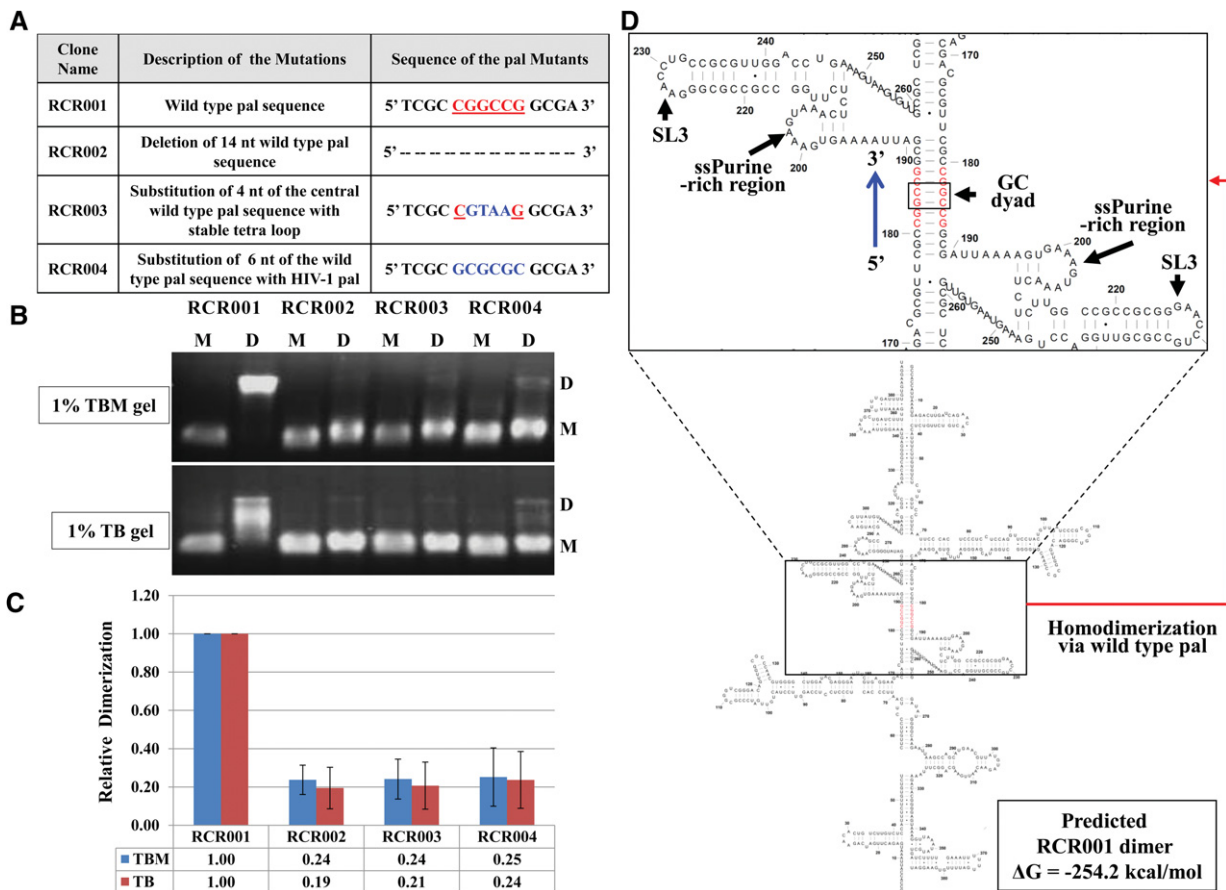

FIGURE 2. Sequences in pal SL augments MPMV gRNA dimerization by functioning as DIS. (*A*) Table showing the 6-nt wild-type pal sequence (in red) and the different mutations introduced (in blue). (*B*) Gel shift assays of the MPMV wild-type and pal mutants. The *upper* panel shows native 1% TBM (50 mM Tris base, 45 mM boric acid, 0.1 mM $MgCl_2$) gel run at 4°C, and the *lower* panel shows semi-native 1% TB (50 mM Tris base, 45 mM boric acid) gel run at room temperature. M indicates monomer lane or monomer conformer; D, dimer lane or dimer conformer for each sample. (*C*) Relative dimerization of the MPMV wild-type and pal mutants RNAs. Following the gel shift assays, dimerization abilities of the RNAs in each lane were calculated, and the dimerization data were then represented as relative to the wild-type dimerization. (*D*) Structure prediction of the MPMV wild-type RNA +1–388 dimer structure. The site of interaction between the two gRNAs involving the pal sequence is enlarged for the sake of clarity. The central 6 nt of the pal are shown in red.

substitution mutants showed significant degrees of reduction in their abilities to dimerize (RDEs 0.19–0.25 in both semi-native and native conditions; fourfold to 5.27-fold reduction; $P = 0.002$–0.02) (Fig. 2C) compared with the wild-type. Mutants RCR002 and RCR003 showed that not only the pal SL structure but also the nature of the pal sequence are important for MPMV gRNA dimerization. The results obtained with mutant RCR004 are more difficult to explain. The Mfold structural model of RCR004 predicted the formation of a SL similar to MPMV pal SL (data not shown), and therefore, this mutant was expected to restore dimerization to wild-type levels, in contrast with our observation (Fig. 2B,C). In the case of HIV-1, the 5′-GCGCGC-3′ pal is flanked by three purines residues that are crucial for the stability of the RNA dimer (Paillart et al. 1997). Therefore, the inability of RCR004 to restore dimerization could in part be attributed to the absence of flanking purines similar to HIV-1 pal in its native genomic context.

Next we wanted to correlate the results of in vitro dimerization to the SHAPE-validated structural model and establish the structure-function relationships, if any, of MPMV pal SL during gRNA dimerization. Therefore, we folded two molecules of the wild-type 5′ end MPMV gRNA using RNAstructure to predict the folding pattern of a dimer. Folding prediction of the wild-type homodimer showed an overall preservation of the RNA secondary structure of the monomer, with a conformational change in the pal SL, which was predicted to interact and base-pair with the pal sequence on the second MPMV gRNA sequence (Fig. 2D). Thus, this structure prediction points toward the importance of the pal SL sequence in MPMV gRNA dimerization, consistent with our RNA dimerization assay and with our SHAPE data, which showed that the loop of the pal hairpin is only moderately reactive. Furthermore in the homodimer structure, the ssPurine-rich region were predicted to form a helix loop structural motif (Fig. 2D; Supplemental Fig. 1), possibly making it more accessible for possible interaction with NC protein as has earlier been suggested (Jaballah et al. 2010). In sharp contrast, the homodimer structural predictions of mutants RCR002 and RCR003 suggested that these mutants could dimerize using another 6-nt pal sequence (5′-GGCGCC-3′) within the PBS (data not shown). This PBS-mediated homodimerization results in the overall destabilization of the RNA secondary structure and is consistent with our in vitro dimerization results. Taken together, our in vitro dimerization and structure prediction results suggest that the 14-nt GC-rich MPMV pal SL is important in mediating gRNA dimerization by functioning as a DIS.

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

To confirm that the central 6 nt of the pal, indeed, constitute the point of contact between two gRNAs to initiate the process of dimerization, we created two *trans*-complementary mutants maintaining the central "GC" dyads (RCR005 and RCR006) (Fig. 3A). When incubated separately, the RCR005 and RCR006 mutant RNAs formed significant amounts of dimer in TBM, even though dimerization was reduced compared with wild-type RNA (RDEs; 0.69–0.70; 1.43- to 1.45-fold reduction; $P = 0.029$–0.047) (Fig. 3B,C). However, the ability of these mutant RNAs to dimerize was severely compromised in semi-native condition (RDEs; 0.14 and 0.09 for RCR005 and RCR006, respectively; 7.14- to 11.11-fold reduction; $P = 0.062$–0.042) (Fig. 3B, lower panel, C). In contrast, when incubated together, these *trans*-complementary mutant RNAs dimerized to wild-type level, in both native and semi-native conditions (RDEs 0.88–1.09) (Fig. 3B, last lane, C).

These results prove that an intermolecular interaction mediated by the central 6 nt of pal is required for the formation of a stable RNA dimer. In addition, they reveal the existence of a second intermolecular interaction, which allows formation of unstable dimers. Consistent with this interpretation, the unstable homodimers formed by RCR005 and RCR006 RNAs migrated slightly more slowly than the heterodimer or the wild-type RNA homodimer (Fig. 3B, top panel), confirming that they formed different RNA dimers.

In an attempt to identify the residual intermolecular interaction in mutants RNAs, we employed RNAstructure to predict the RCR005 and RCR006 homodimer and heterodimer structures. Structure predictions suggested that the mutant RNAs RCR005 and RCR006 homodimerized via an alternate 6-nt pal sequence within the PBS (Fig. 3D,E; Supplemental Fig. 1B,C). As a consequence, SL2, which is relatively long in the wild-type homodimer and in the RCR005/RCR006 heterodimer, was predicted to be shortened due to PBS-mediated intermolecular interactions; however, its apical loop was maintained (Fig. 3D,E; Supplemental Fig. 1B,C). In the predicted heterodimer structure of mutants RCR005 and RCR006, the two RNAs interact via the substituted *trans*-complementary sequences (Fig. 3F,G; Supplemental Fig. 1D) as expected and as supported by our experimental data (Fig. 3B and C). As a result, the secondary structures of the wild-type RNA homodimer and of the RCR005/RCR006 heterodimer are identical (Figs. 2D, 3G; Supplemental Fig. 1). RNAstructure predictions of the RCR005 and RCR006 homodimers and heterodimer illustrate that not only does the interaction between these *trans*-complementary sequences result in heterodimerization, but also that such an interaction leads to conformational changes in the structure allowing more extensive base-pairing between complementary sequences throughout the genome (Fig. 3G; Supplemental Fig. 1).

In summary, our systematic mutational analyses directed toward ascertaining the functional significance of pal SL demonstrated that mutations targeted toward disrupting the formation of the 6-nt pal loop diminished RNA dimerization (Figs. 2, 3). This could be attributed to the RNA folding pattern rather than the primary sequence of the pal loop
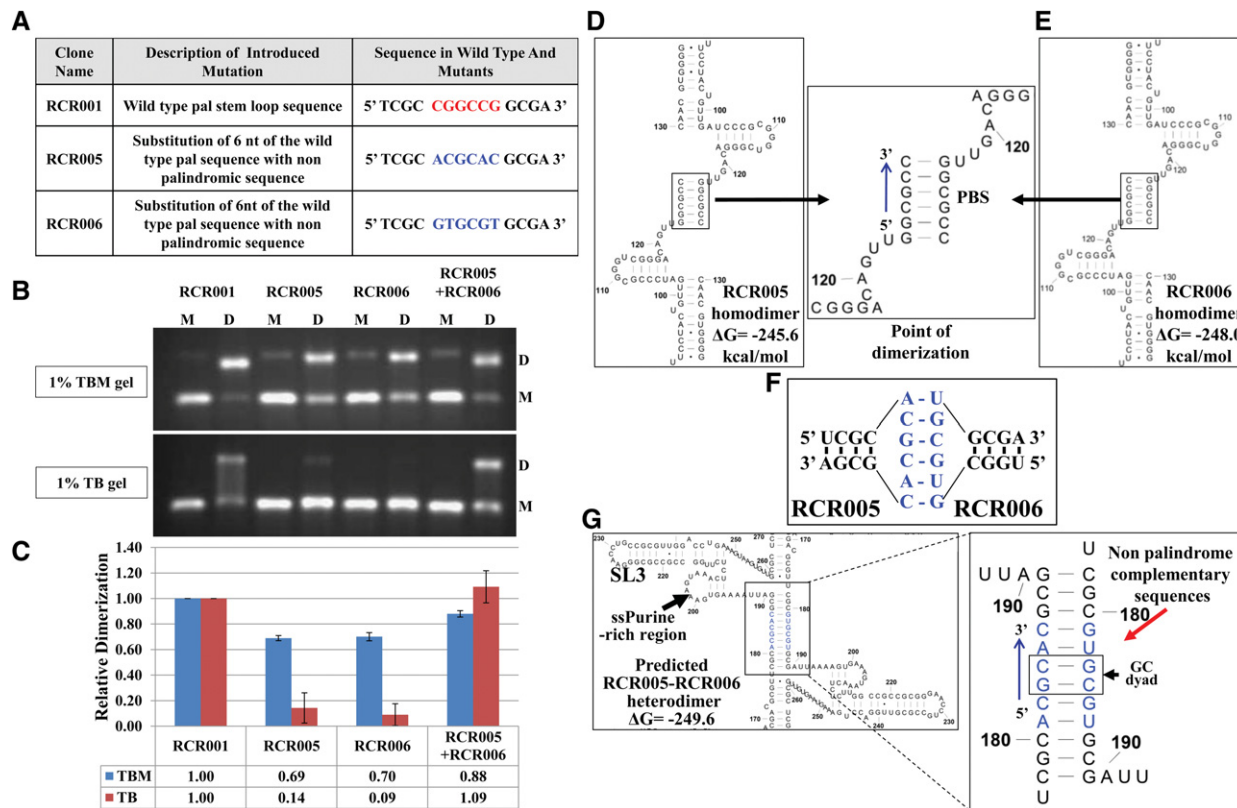
FIGURE 3. In vitro heterodimerization can be mediated by *trans*-complementary sequences on two RNAs. (*A*) Table showing the 6-nt wild-type pal sequence (in red) and the two complementary substitution mutations (in blue). (*B*) In vitro dimerization assay of the MPMV wild-type and pal mutants. The *upper* panel shows native 1% TBM gel run at 4°C, and the *lower* panel shows semi-native 1% TB gel run at room temperature. M indicates monomer lane or monomer conformer; D, dimer lane or dimer conformer for each sample. (*C*) Relative dimerization of the complementary mutants to the wild-type RNAs. Dimerization abilities of the RNAs in each lane were calculated, and the dimerization data were represented as relative to the wild-type dimerization. (*D,E*) Structure predictions of the mutant RCR005 (*D*) and RCR006 (*E*) MPMV RNAs +1–388 using RNAstructure. (*F*) Schematic representation of the expected point of heterodimerization between RCR005 and RCR006 MPMV RNAs. (*G*) Predicted heterodimer structure of RCR005 and RCR006 RNAs.

since mutants containing *trans*-complementary sequences in this loop were not only able to heterodimerize at wild-type levels but also able to re-establish the overall wild-type RNA secondary structure (Figs. 2D, 3G; Supplemental Fig. 1). These results strongly suggest that the pal loop is involved in "kissing loop" interactions between two MPMV gRNAs in order to initiate gRNA dimerization and thus functions as DIS. These observations corroborate with an earlier genetic analysis that showed no effect on RNA packaging when the MPMV native pal sequence was accidentally substituted with a different pal sequence (Mustafa et al. 2004).

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

In order to examine the conservation of different structural motifs within the MPMV packaging signal RNA, sequences from different strains of MPMV were used to predict the RNA secondary structure of this region. Mfold predictions of the sequences from different MPMV strains consistently

revealed the ability to form both LRIs involving sequences from U5 and *gag* (Fig. 4A–E). A sequence alignment of this region for these MPMV isolates was generated using ClustalW and revealed a high degree of conservation of the sequences in the two U5-Gag LRIs. The U5 and *gag* sequences forming these LRIs slightly differ but maintain a very high degree of complementarity (Fig. 4A–F).

Based on the RNA free-energy minimized model predicted earlier and validated by SHAPE (Fig. 1) in addition to the phylogenetic and in silico modeling analyses presented here (Fig. 4), it is reasonable to hypothesize that such LRIs between U5 and *gag* sequences are likely to exist in vivo and have a functional role in MPMV life cycle. Consistent with this, deletion of the *gag* sequences involved in U5-Gag LRI-II while maintaining the *gag* sequences that are involved in U5-Gag LRI-I severely diminished MPMV transfer vector RNA packaging ability (Schmidt et al. 2003). On the other hand, inclusion of first 100 nt of *gag* sequences that would maintain both U5-Gag LRIs in the transfer vector RNA restored packaging efficiency to the wild-type level (Schmidt et al. 2003). These results suggest that LRI-I and LRI-II
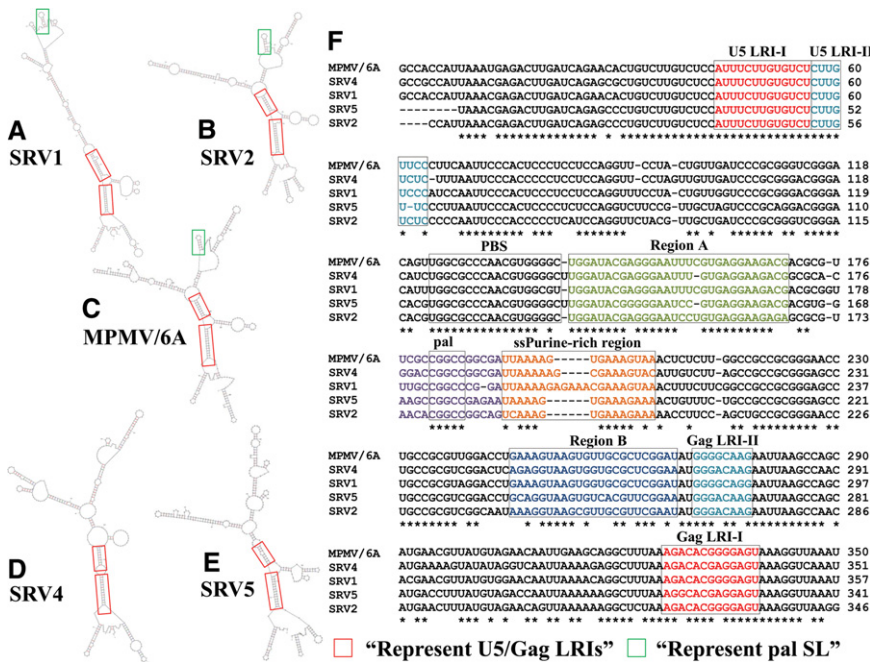
FIGURE 4. Mfold structural predictions (A–E) and ClustalW (F) sequence alignment of the 5′ end genomes of different MPMV strains. (A–E) The U5/Gag LRIs and the pal sequence are highlighted by red and green boxes, respectively. (F) The aligned sequences pertaining to major structural motifs are highlighted by different colors and boxed. The accession nos. for MPMV6/A, SRV1, SRV2, SRV4, and SRV5 are M12349.1 (Sonigo et al. 1986), M11841.1 (Power et al. 1986), AF126467.1 (Marracci et al. 1995), FJ979638.1 (Zao et al. 2010), and AB611707.1 (Takano et al. 2013), respectively.

play an important architectural role in stabilizing the RNA secondary structure of the 5′-UTR sequences required for MPMV gRNA packaging (Schmidt et al. 2003; Jaballah et al. 2010). Such a scenario could be comparable to that observed in other retroviruses (HIV-1, HIV-2, and FIV) where interactions between the R/U5 regions and the start of *gag* sequences have been proposed to stabilize the overall RNA secondary structure essential for gRNA dimerization and packaging (Paillart et al. 2002; Kenyon et al. 2008, 2011; Song et al. 2008; Rizvi et al. 2010; Lu et al. 2011b).

Between SL2 and SL3 of our SHAPE-validated RNA structure is a very distinctive pal SL, which has earlier been shown to be required for optimal packaging of MPMV gRNA (Jaballah et al. 2010). The current study revealed that the central 6 nt of this pal (5′-CGGCCG-3′) functions as DIS, thus probably facilitating the encapsidation of the dimeric genome. Sequence alignment of five different strains of MPMV revealed that the 6-nt central pal is conserved in all but one strain (Fig. 4F). In one strain (SRV1), the last G is substituted with a C residue, but this strain still contains a 6-nt GC-rich pal (5′-GCCGGC-3′) shifted by 2 nt compared with the other strains (Fig. 4F), supporting a functional role for this pal. Mfold analysis of these strains further revealed that in three strains (MPMV 6/A, SRV1, and SRV2), the pal sequence fold into a SL (Fig. 4A–C). In the case of SRV4, the pal sequence is base-paired; whereas in the case of

SRV5, the pal sequence forms two small bulges (Fig. 4D,E). Similar conservation of a DIS comprising a pal sequence has been reported in several retroviruses, including HIV-1, HIV-2, SIV, and FIV (Paillart et al. 2002; Russell et al. 2004; Lever 2007; Kenyon et al. 2008, 2011). In a number of retroviruses, sequences augmenting RNA dimerization and packaging have been shown to be intermingled (Lanchy et al. 2003; Paillart et al. 2004; Russell et al. 2004; Lanchy and Lodmell 2007; Lever 2007), which is in agreement with the deletion of MPMV pal SL resulting in ablation of RNA packaging (Jaballah et al. 2010).

Earlier studies have identified a prominent stretch of ssPurine-rich region between pal SL and SL3, and a partial base-paired repeat sequence of the ssPurine-rich region was observed in the adjacent region ("B") (Fig. 1A; Jaballah et al. 2010). Sequence alignment revealed that out of the 16 nt within the ssPurine-rich region, 10 nt are conserved among all the analyzed MPMV strains (Fig. 4F). The validation of the ssPurine-rich region and its based-paired repeat structural motifs by SHAPE (Fig. 1B) in concurrence with sequence alignment (Fig. 4F) and earlier mutational and structure prediction analyses (Jaballah et al. 2010) further emphasize the importance of the ssPurine-rich region in MPMV life cycle.

SL1, SL2, and SL3 are present in spliced as well as gRNAs

To establish the structure-function relationship of structural motifs (especially of those present upstream of the major splice donor [mSD] site) and to address how some of them may help the MPMV packaging machinery in differentiating between genomic and spliced RNA, we predicted the structure of the 5′ region of the MPMV *env* spliced RNA (including 251 nt upstream of the mSD and 138 nt downstream from *env* splice acceptor site). Mfold structural predictions revealed that SL1, SL2, and SL3 are present in both MPMV 5′ end genomic and *env* spliced RNAs (Supplemental Fig. 2). SL1, SL2, and SL3 were also predicted to be present in the shorter MPMV leader RNA that was truncated at the mSD (+1–251; data not shown). The function of SL1 in the retroviral life cycle has not yet been established; however, Mfold predictions (Fig. 4A–E) and sequence alignments (Fig. 4F) of five different strains of MPMV showed that this motif is conserved at the sequence and structural levels.

SL2 is a rather long and stable structure from which PBS protrudes (Fig. 1A,B). It is present in the SHAPE-validated

structure of the gRNA (Fig. 1B) and is predicted to exist in the *env* spliced RNA (Supplemental Fig. 2). In silico analysis revealed a high degree of conservation in the sequences forming SL2 (Fig. 4F), which was predicted to fold in three out of the five MPMV strains (MPMV 6/A, SRV4, and SRV5) (Fig. 4A–E). In addition, SL2 has been shown to include sequences that constitute part of a bipartite packaging signal (Jaballah et al. 2010). All together, these data points toward the importance of the overall RNA secondary structure of SL2 in the MPMV life cycle (Schmidt et al. 2003; Jaballah et al. 2010).

Immediately downstream from the ssPurine-rich region is SL3, which like SL1 and SL2 was consistently predicted to fold in both genomic and spliced RNAs (Supplemental Fig. 2). Mfold predictions of different MPMV strains demonstrated that SL3 is formed in all except for one strain (Fig. 4A–E), and its sequences are fairly conserved among different strains (Fig. 4F). However, deletion of the majority of the sequences that are involved in base-pairing of SL3 did not affect RNA packaging (Jaballah et al. 2010), suggesting that it might play a role in MPMV life cycle that has not yet been established.

CONCLUDING REMARKS

Employing SHAPE, we validated the overall predicted structure of the MPMV packaging signal RNA (Fig. 1; Jaballah et al. 2010). Briefly, MPMV packaging signal RNA comprised five SL structures. SL1–SL3 includes sequences from R/U5 and UTR, whereas Gag SL1 and Gag SL2 employ sequences exclusively from *gag*.

SL2 involves sequences from region "A," whereas SL3 in part comprises sequences from region "B"; both of these regions have been shown earlier as major MPMV gRNA packaging determinants (Schmidt et al. 2003; Jaballah et al. 2010). Between SL2 and SL3 is a very distinctive pal SL followed by a stretch of ssPurine-rich region, both of which have also been shown to be required for optimal packaging of MPMV gRNA (Jaballah et al. 2010). The overall structure is anchored by two LRIs (Fig. 1A,B) involving sequences from U5 and the first 70 nt of *gag*, which have been shown to be important in MPMV RNA packaging (Schmidt et al. 2003; Jaballah et al. 2010). Mfold predictions of different MPMV strains consistently revealed the ability to form LRIs suggesting a functional role for these LRIs (Fig. 4A–E).

Employing a systematic mutational approach, we showed that the central 6 nt of the pal, which are exposed in the pal SL loop play a crucial role in RNA dimerization by directly interacting with a second gRNA molecule (Figs. 2, 3; Supplemental Fig. 2). Thus, this sequence likely initiates gRNA dimerization by forming a kissing loop complex. Despite some sequence variability, all MMPV strains maintain a 6-nt GC pal at this position, reinforcing its functional role.

The SHAPE-validated structural model, in vitro dimerization analysis of pal mutant, and the homo- and heterodimer structural models presented here coupled with earlier pub-

lished findings on MPMV RNA packaging signals offer important functional correlation between gRNA dimerization and the packaging processes. These observations while broadening the understanding of functional regions of MPMV gRNA further suggest that MPMV may select a dimeric genome for packaging.

MATERIALS AND METHODS

MPMV RNA secondary structure analyses in silico

MPMV RNA (388 nt and 550 nt) secondary structure was predicted using the Mfold server (Mathews et al. 1999; Zuker 2003) and validated by applying SHAPE reactivity data in the RNAstructure software (version 5.3) (Reuter and Mathews 2010). The predicted RNA secondary structures were redrawn with XRNA software. Sequences from different strains of MPMV corresponding to packaging signal RNA were employed to predict the RNA secondary structures of the region using Mfold (Mathews et al. 1999; Zuker 2003). A sequence alignment of the same region for these MPMV isolates was also generated using ClustalW to look for the conservation of sequences in different structural motifs. MPMV nucleotide positions refer to GenBank accession no. M12349 (Sonigo et al. 1986).

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

In order to in vitro transcribe the wild-type MPMV genomic sequences for SHAPE analysis, SJ2 subgenomic transfer vector (described previously by Jaballah et al. 2010) containing the *cis*-acting sequences needed for genome replication was used as a template for generating the clone RCR001 by PCR. Primers used for PCR amplification included sense (S) primer OTR1004 (5′-CCC*AAG CTT*AATACGACTCACTATAGGGCCACCATTAAATGAGACTTG ATC-3′) containing the HindIII restriction site and the T7 RNA polymerase promoter sequence (HindIII site is shown in italics, whereas the T7 promoter sequence is underlined) and the anti-sense (AS) primer OTR 1005 (5′-AAA*CCCGGG*TTCTTTCTTATCTATCAA TTC-3′) containing XmaI/SmaI site (shown in italics). The resulting PCR product was cleaved at the HindIII and XmaI sites (artificially created on either ends of the MPMV sequence) and ligated to a PUC-based cloning vector (pIC19R) (Marsh et al. 1984), which was previously cleaved with the same restriction enzymes generating wild-type plasmid pRCR001. pRCR001 contains nucleotides +1 to 550 (+1 corresponds to the first nucleotide of the R region of the gRNA) under the control of a T7 promoter for in vitro transcription of the RNA.

To study the effects of the pal sequence (5′-TCGCCGGCCGGCG A-3′) on in vitro dimerization, a series of mutations were introduced in the pal sequence. Briefly, these mutations were incorporated through splice overlap extension (SOE) PCR according to the method described previously (Gibbs et al. 1994; Jaballah et al. 2010; Rizvi et al. 2010) using MPMV subgenomic vector SJ2 as the template requiring two separate amplifications in round-one PCR. In brief, both S and AS primers that were used in two separate reactions of round-one PCRs were designed in a fashion that the resulting products from the two reactions have overlapping complementary sequences. The overlapping complementary sequences allowed the

two products to anneal during the second round of amplification with the outer S and AS primers, resulting in a final product containing the desired deletion mutation. The mutations introduced included deletion of the complete 14-nt pal SL sequence (5′-TCG CCGGCCGGCGA-3′) using S primer OTR 897 (5′-CACTTTT AAACGCGTCGTCTTCCTCAC G-3′) and AS primer OTR 898A (5′-CGCGTTTAAAAGTGAAAGTAAACTCTCTTG-3′), resulting in RCR002 (Fig. 2A). Substitution of the 4-nt pal loop 5′-GGCC-3′ with GTAA to generate a stable tetra loop was created using S primer OTR 996 (5′-GTAAGGCGATTAAAAGTGAAAGTAAACTC-3′) and AS primer OTR 997 (5′-CACTTTTAATAGCCTTACGGCG AACGCGTCGTC-3′), generating mutant RCR003 (Fig. 2A). A mutant containing substitution of the 6-nt pal with HIV-1 pal sequence (5′-GCGCGC-3′) was constructed using S primer OTR 998 (5′-G CGCGCGCGATTAAAAGTGAAAGTAAACTC-3′) and AS primer OTR 999 (5′-CACTTTTAATCGCGCGCGCGCGAACGCGTCGT CTTCCTC-3′), generating mutant RCR004 (Fig. 2A).

In addition, two other mutants, RCR005 and RCR006 containing heterologous *trans*-complementary sequence with a central "GC" dyad were also created to determine if the pal nature of the sequence is necessary for in vitro dimerization (Fig. 3A). The primers used for RCR005 and RCR006 construction included OTR 1000 (S primer; 5′-ACGCACGCGATTAAAAGTGAAAGTAAACTC-3′) and OTR 1001 (AS; 5′-CAC TTT TAA TCG CGT GCG TGCGAACGCGTC GTCTTCCTC-3′), OTR 1002 (S primer; 5′-GTGCGTGCGATT AAAAGTGAAAGTAAACTC-3′) and OTR 1003 (AS primer; 5′-CACTTTTAATCGCACGCACGCGAACGCGTCGTCTTCCTC-3′), respectively. For all these mutants, the OTR 1004 (S) and OTR 1005 (AS) were used as outer primers during second round of amplification, and the resulting amplified product containing the T7 promoter and the desired mutation was cleaved and cloned in pIC19R as described above for creating wild-type clone RCR001. Both the wild-type clone and all the pal mutant clones were confirmed by sequencing. PCRs were performed according to the method described previously (Jaballah et al. 2010; Rizvi et al. 2010). Briefly, ∼3 μg of template DNA was used for each reaction under the following amplification conditions: denaturation for 2 min at 94°C followed by 30 cycles of 1 min at 94°C, annealing at 45°C –47°C depending on primers used for 90 sec, and extension for 90 sec at 72°C. Final extension was performed for 10 min at 72°C followed by incubation at 4°C. The precise details of the nucleotide locations of the primers and the intermediate cloning steps are available from us upon request.

RNA preparation by in vitro transcription

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

and the RNA concentrations were determined following quantification through nanodrop (ThermoScientific). Finally, the RNA homogeneity was checked by electrophoresis on an 8% acrylamide gel.

SHAPE methodology

To validate the predicted secondary structure of MPMV packaging signal RNA sequence (Jaballah et al. 2010), the purified in vitro transcribed RNAs were subjected to SHAPE (Mortimer and Weeks 2009) using benzoyl cyanide (BzCN). Briefly, 1 pmol of RNA in 8 μL milli-Q water (Millipore) was denatured for 2 min at 90°C and then chilled for 2 min on ice, following which 2 μL 5X dimer buffer (250 mM sodium cacodylate at pH 7.5; 1.5 M KCl; 25 mM MgCl$_2$) was added to the samples, which were then incubated for 20 min at 37°C. Next, 2 μg of total yeast tRNA (Sigma Aldrich) was added to each sample, and the volume was adjusted to 15 μL by adding 1× dimer buffer and incubated for 10 min at room temperature. Three microliters of a 1 μM BzCN solution in anhydrous dimethyl sulphoxide (DMSO) was used to modify the RNA samples for 1 min, and the reaction was stopped by adding 82 μL water. The negative control samples were treated in the same fashion but using only DMSO in the absence of BzCN. All samples were then precipitated using 1 μL of 1 μg/μL glycogen solution, 1/10 volume 3 M sodium acetate (pH 6.5), and 3 volume ethanol for 30 min on dry ice, and the precipitates were collected by centrifugation at 13,000g for 20 min at 4°C. The RNA pellets were washed two times with 500 μL cold 80% ethanol to remove salts, dried in a vacuum dryer, and dissolved in 7 μL milli-Q water. Next, to identify the BzCN modification sites, reverse transcription (RT) was performed on the samples. Toward this end, two types of primer sequence were used to ensure RT of the entire 550 nt MPMV leader RNA. A set of four AS primers labeled with different dyes corresponded to the 5′-AGTTAC TGGGACTTTCTCCG-3′ sequence (complementary to MPMV nt 483–502) and were termed as OTRs 17, 18, 19, and 20. A second set of AS primers corresponded to the 5′-CTTACTTTCAGGTCC AACGC-3′ sequence (complementary to MPMV nt 235–254), and were named OTRs 21, 22, 23, and 24. These primers were labeled with either 6-FAM, VIC, NED, or PET. The purpose of using different dyes was that each dye could be detected individually after capillary electrophoresis in a DNA sequencer according to the method described earlier (Merino et al. 2005; Mortimer and Weeks 2007, 2009). BzCN-modified RNAs were annealed to 1 μM VIC-labeled primers OTR 18 or OTR 22 for 2 min at 90°C and 2 min in ice. After addition of 2 μL 5X RT buffer (Life Science), the samples were then incubated for 10 min at room temperature. Each sample was then divided into two aliquots, and elongation reaction was performed for 30 min at 42°C and for 15 min at 50°C in elongation buffer (1 μL 5× RT buffer, 3 μL 2.5 mM dNTPs mix, 1 μL 2 U/μL RT AMV enzyme from Life Science and water to make the volume to 10.5 μL). For the unmodified RNA samples, 6-FAM–labeled primers OTR 17 or OTR 21 were used, and RT reaction was performed in the same manner as for the modified RNA samples. A ddA sequencing ladder was prepared using 2 pmol of untreated RNA and 1 μL of 2 μM NED-labeled OTR 19 or OTR 23 in 8 μL milliQ water. Annealing was performed by heating for 2 min at 90°C and cooling for 2 min on ice. The volume was made to 10 μL by adding 2 μL 10× RT buffer, and then the samples were incubated for 15 min at room temperature. The RNA sample was aliquoted into two tubes, and the elongation reaction was performed with 1 μL 10× RT buffer, 3 μL A10 (0.25 mM dATP, 1 mM

dGTP, 1 mM dCTP, 1 mM dTTP), 1 μL of 100 μM ddA, and 1 μL of 1 units/μL RT AMV enzyme. A ddG ladder was also prepared in the same fashion by using PET-labeled OTR 20 or OTR 24, G10 (0.25 mM dGTP, 1 mM dATP, 1 mM dCTP, 1 mM dTTP), and 100 μM ddG. All the reactions were stopped by adjusting the volume 45 μL with water, and proteins were extracted with 50 μL phenol-chloroform. For each experiment, the modified and unmodified samples were pooled with the corresponding ddA and ddG sequencing ladders (e.g., samples for which OTRs 17, 18, 19, and 20 were used were pooled together) in a single tube containing 20 μL 3 M sodium acetate and 600 μL ethanol for cDNA precipitation. The samples were incubated on dry ice for 30 min, centrifuged at 13,000*g* for 20 min at 4°C, and washed twice with 1 mL cold 80% ethanol. Pellets were spun for 5 min at each step to prevent material loss. After drying the pellets in a vacuum dryer, they were resuspended in 10 μL HiDi formamide (ABI) and heat denatured at 90°C and iced for 5 min each. The samples were spun down before loading on the 96-well plates for sequencing on an Applied Biosystems 3130xl genetic analyzer. The results were generated in the form of electropherograms, which were analyzed with the SHAPEfinder program (Vasa et al. 2008; Wilkinson et al. 2008) by following the steps prescribed by program developers such as baseline adjustment, matrixing, cubic mobility shift, signal decay correction, scale factor, alignment and integration to MPMV sequence, and finally fitting. This software enables the calculation of the SHAPE reactivity at each nucleotide. Prior to performing the SHAPE reactions, we calibrated the fluorescent tagged primers for determining the mobility shift (Vasa et al. 2008). The peaks in the electropherograms were normalized by subtracting the peak area of the unmodified RNA sample from the peak area of the SHAPE-treated sample (Vasa et al. 2008). The data were then further normalized for outliers by determining 10% of the highest values and excluding the highest 2% of these values. The remaining 8% of the values were averaged, and the average value was used to divide the SHAPE data. The SHAPE reactivity data (provided as supplemental data) obtained for each nucleotide were then applied as constraints in the structure prediction program RNAstructure (version 5.3) (Reuter and Mathews 2010) in order to obtain the SHAPE-derived RNA structure of the MPMV packaging signal RNA.

In vitro dimerization assay

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

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

REFERENCES

Al Dhaheri NS, Phillip PS, Ghazawi A, Ali J, Beebi E, Jaballah SA, Rizvi TA. 2009. Cross-packaging of genetically distinct mouse and primate retroviral RNAs. *Retrovirology* **6:** 66.

Al Shamsi IR, Al Dhaheri NS, Phillip PS, Mustafa F, Rizvi TA. 2011. Reciprocal cross-packaging of primate lentiviral (HIV-1 and SIV) RNAs by heterologous non-lentiviral MPMV proteins. *Virus Res* **155:** 352–357.

Bellaousov S, Reuter JS, Seetin MG, Mathews DH. 2013. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res* **41:** W471–W474.

Chamanian M, Purzycka KJ, Wille PT, Ha JS, McDonald D, Gao Y, Le Grice SFJ, Arts EJ. 2013. A *cis*-acting element in retroviral genomic RNA links Gag-Pol ribosomal frameshifting to selective viral RNA encapsidation. *Cell Host Microbe* **13:** 181–192.

Clever JL, Wong ML, Parslow TG. 1996. Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol* **70:** 5902–5908.

D'Souza V, Summers MF. 2005. How retroviruses select their genomes. *Nat Rev Microbiol* **8:** 643–655.

Ennifar E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C, Dumas P, Walter P. 2003. HIV-1 RNA dimerization initiation site is structurally similar to the ribosomal A site and binds aminoglycoside antibiotics. *J Biol Chem* **278:** 2723–2730.

Ennifar E, Paillart JC, Bodlenner A, Walter P, Weibel JM, Aubertin AM, Pale P, Dumas P, Marquet R. 2006. Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: From crystal to cell. *Nucleic Acids Res* **34:** 2328–2339.

Gibbs JS, Regier DA, Desrosiers RC. 1994. Construction and in vitro properties of SIV mac mutants with deletions in "nonessential" genes. *AIDS Res Hum Retroviruses* **10:** 607–616.

Guesdon FMJ, Greatorex J, Rhee SR, Fisher R, Hunter E, Lever AML. 2001. Sequences in the 5′ leader of Mason-Pfizer monkey virus which affect viral particle production and genomic RNA packaging: Development of MPMV packaging cell lines. *Virology* **288:** 81–88.

Haddrick M, Lear AL, Cann AJ, Heaphy S. 1996. Evidence that a kissing loop structure facilitates genomic RNA dimerization in HIV-1. *J Mol Biol* **259:** 58–68.

Harrison GP, Hunter E, Lever AML. 1995. Secondary structure model of the Mason-Pfizer monkey virus 5′ leader sequence: Identification of a structural motif common to a variety of retroviruses. *J Virol* **69:** 2175–2186.

Hussein ITM, Ni N, Galli A, Chen J, Moore MD, Hu WS. 2010. Delineation of the preferences and requirements of the human immunodeficiency virus type 1 dimerization initiation signal by using an in vivo cell-based selection approach. *J Virol* **84:** 6866–6875.

Jaballah SA, Aktar SJ, Ali J, Phillip PS, Al Dhaheri NS, Jabeen A, Rizvi TA. 2010. A G-C rich palindromic structural motif and a stretch of single

stranded purines are required for optimal packaging of Mason-Pfizer monkey virus (MPMV) genomic RNA. *J Mol Biol* **401:** 996–1014.

Johnson SF, Telesnitsky A. 2010. Retroviral RNA dimerization and packaging: The what, how, when, where, and why. *PLoS Pathog* **6:** e1001007. doi: 10.1371/journal.ppat.1001007.

Kenyon JC, Ghazawi A, Cheung WK, Phillip PS, Rizvi TA, Lever AM. 2008. The secondary structure of the 5′ end of the FIV genome reveals a long-range interaction between R/U5 and *gag* sequence, and a large, stable stem–loop. *RNA* **14:** 2597–2608.

Kenyon JC, Tanner S, Legiewicz M, Phillip PS, Rizvi TA, LeGrice S, Lever AML. 2011. SHAPE analysis of the FIV leader RNA reveals a structural switch potentially controlling viral packaging and genome dimerization. *Nucleic Acids Res* **39:** 6692–6704.

Lanchy JM, Lodmell JS. 2007. An extended stem-loop 1 is necessary for human immunodeficiency virus type 2 replication and affects genomic RNA encapsidation. *J Virol* **7:** 3285–3292.

Lanchy JM, Ivanovitch JD, Lodmell JS. 2003. A structural linkage between the dimerization and encapsidation signals in HIV-2 leader RNA. *RNA* **9:** 1007–1018.

Lever AML. 2007. HIV RNA packaging. *Adv Pharmacol* **55:** 1–32.

Lodmell JS, Ehresmann C, Ehresmann B, Marquet R. 2000. Convergence of natural and artificial evolution on an RNA loop–loop interaction: The HIV-1 dimerization initiation site. *RNA* **6:** 1267–1276.

Lodmell JS, Ehresmann C, Ehresmann B, Marquet R. 2001. Structure and dimerization of HIV-1 kissing loop aptamers. *J Mol Biol* **311:** 475–490.

Lu K, Heng X, Summers MF. 2011a. Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol* **410:** 609–633.

Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, Dorjsuren B, Kulandaivel G, Jones S, Hiremath A, et al. 2011b. NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. *Science* **334:** 242–245.

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

Marracci GH, Kelley RD, Pilcher KY, Crabtree L, Shiigi SM, Avery N, Leo G, Webb MC, Hallick LM, Axthelm MK, et al. 1995. Simian AIDS type D serogroup 2 retrovirus: Isolation of an infectious molecular clone and sequence analyses of its envelope glycoprotein gene and 3′ long terminal repeat. *J Virol* **69:** 2621–2628.

Marsh JL, Erfle M, Wykes EJ. 1984. The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. *Gene* **32:** 481–485.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288:** 911–940.

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127:** 4223–4231.

Moore MD, Hu WS. 2009. HIV-1 RNA dimerization: It takes two to tango. *AIDS Rev* **11:** 91–102.

Moore MD, Nikolaitchik OA, Chen J, Hammarskjöld ML, Rekosh D, Hu WS. 2009. Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses. *PLoS Pathog* **5:** e1000627.

Mortimer SA, Weeks KM. 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129:** 4144–4145.

Mortimer SA, Weeks KM. 2009. Time-resolved RNA SHAPE chemistry: Quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc* **4:** 1413–1421.

Motomura K, Chen J, Hu WS. 2008. Genetic recombination between human immunodeficiency virus type 1 (HIV-1) and HIV-2, two distinct human lentiviruses. *J Virol* **82:** 1923–1933.

Mustafa F, Lew KA, Schmidt RD, Browning MT, Rizvi TA. 2004. Mutational analysis of the predicted secondary RNA structure of

the Mason-Pfizer monkey virus packaging signal. *Virus Res* **99:** 35–46.

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

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

Paillart JC, Marquet R, Skripkin E, Ehresmann C, Ehresmann B. 1996b. Dimerization of retroviral genomic RNAs: Structural and functional implications. *Biochimie* **7:** 639–653.

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

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

Paillart JC, Xhilag MS, Marquet R, Mak J. 2004. Dimerization of retroviral RNA genomes: An inseparable pair. *Nat Rev Microbiol* **2:** 461–472.

Power MD, Marx PA, Bryant ML, Gardner MB, Barr PJ, Luciw PA. 1986. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. *Science* **231:** 1567–1572.

Reuter JS, Mathews DH. 2010. RNA structure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11:** 129.

Rizvi TA, Kenyon JC, Ali J, Aktar SJ, Phillip PS, Ghazawi A, Mustafa F, Lever AML. 2010. Optimal packaging of FIV genomic RNA depends upon a conserved long-range interaction and a palindromic sequence within gag. *J Mol Biol* **403:** 103–119.

Russell RS, Liang C, Wainberg MA. 2004. Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology* **1:** 23.

Schmidt RD, Mustafa F, Lew KA, Browning MT, Rizvi TA. 2003. Sequences within both the 5′ untranslated region and the gag gene are important for efficient encapsidation of Mason-Pfizer monkey virus RNA. *Virology* **309:** 166–178.

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

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

Takano J, Leon A, Kato M, Abe Y, Fujimoto K. 2013. Isolation and DNA characterization of a simian retrovirus 5 from a Japanese monkey (*Macaca fuscata*). *J Gen Virol* **94:** 955–959.

Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. 2008. ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14:** 1979–1990.

Vile RG, Ali M, Hunter E, McClure MO. 1992. Identification of a generalized packaging sequence for D-type retroviruses and generation of a D-type retroviral vector. *Virology* **189:** 786–791.

Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6:** e96. doi: 10.1371/journal.pbio.0060096.

Zao CL, Armstrong K, Tomanek L, Cooke A, Berger R, Estep JS, Marx PA, Trask JS, Smith DG, Yee JL, et al. 2010. The complete genome and genetic characteristics of SRV-4 isolated from cynomolgus monkeys (*Macaca fascicularis*). *Virology* **405:** 390–396.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31:** 3406–3415.