In vitro characterization of a base pairing interaction between the primer binding site and the minimal packaging signal of avian leukosis virus genomic RNA

Igor Kanevsky, Natalya Vasilenko, Hélène Dumay-Odelot and Philippe Fossé*

Unité Mixte de Recherche 8113 du CNRS, LBPA-Alembert, Ecole Normale Supérieure de Cachan, 94235 Cachan cedex, France

Received October 9, 2003; Revised and Accepted October 30, 2003

ABSTRACT
The 5' leader region of avian sarcoma-leukosis viruses (ASLVs) folds into a series of RNA secondary structures which are involved in key steps in the viral replication cycle such as reverse transcription, dimerization and packaging of genomic RNA. The O3 stem and three stem–loops (O3SLa, O3SLb and O3SLc) form the minimal packaging signal that is located downstream of the primer binding site (PBS). The U5±PBS region contributes to packaging via a mechanism that remains unknown. In this in vitro study, we have investigated the possibility of interactions between the R±U5±PBS region and the minimal packaging signal using chemical and enzymatic probing, antisense oligonucleotides and site-directed mutagenesis. We have identified a base pairing interaction between the PBS sequence and the terminal loop of O3SLa. It was found that the PBS/O3SLa interaction was intramolecular since it occurred not only in dimeric RNA but also in monomeric RNA. This interaction probably corresponds to a pseudoknot interaction. The PBS/O3SLa interaction may be formed in vivo since the sequences are highly conserved in ASLV strains. The PBS/O3SLa interaction may regulate the processes of primer tRNA annealing, packaging and initiation of Gag translation through its involvement in leader tertiary structure. Interestingly, we found that in other retroviruses the PBS sequence can also base pair with a terminal loop of the stem–loops involved in RNA packaging.

INTRODUCTION
The 5' leader sequences of all genomic retroviral RNAs fold into a series of RNA secondary structures that regulate key steps in the retroviral life cycle (1). These structures are involved in initiation of reverse transcription, specific packaging of genomic RNA into the virion, genomic RNA dimerization and splicing of viral mRNA. The specificity of RNA packaging is determined by the interactions between the packaging signal, called Ψ or E, and the viral polyprotein Gag (2). The core of the packaging signal of most retroviruses corresponds to a cluster of stem–loops that are located downstream of the primer binding site (PBS) and the start of gag (3–6). However, other sequences are required for optimal RNA packaging. For example, mutations in the 5' R region of human immunodeficiency virus type 1 (HIV-1), reduces the efficiency of RNA packaging (7–10). In HIV-1, Moloney murine leukemia virus (M-MuLV) and Rous sarcoma virus (RSV), mutations in the U5-PBS region cause significant RNA packaging defects (10–13). To date, the mechanism of action of the R–U5–PBS region on RNA packaging remains unresolved.

A feature of the avian sarcoma-leukosis virus (ASLV) 5' leader is that it contains three short open reading frames (uORF1, uORF2 and uORF3) upstream of the Gag initiation codon (14). The first and second of these short ORFs, uORF1 (nucleotides 41–61) and uORF2 (nucleotides 82–129), are located in the R–U5–PBS region that is able to form stable secondary structures (Fig. 1A). Mutations targeting uORF1 markedly reduce the packaging efficiency of genomic RNA (15). Extensions of the upper stem of the U5 region (nucleotides 82–88 and 94–100 within uORF2) cause significant RNA packaging defects (12). Mutations in uORF3 (nucleotides 198–224) strongly decrease the RNA packaging efficiency (15,16). Some studies suggest that translation of uORF1 and uORF3 is necessary for encapsidation of ASLV genomic RNA into the virion (15,17). However, it has not been demonstrated that the peptides encoded by these ORFs are synthesized in cells infected by replication-competent viruses. In addition, Sonstegard and Hackett (16) were unable to find a direct functional coupling between translation of uORFs and RNA packaging. Their data and other studies (6,12,18,19) strongly suggest that mutations in the uORFs affect packaging because they disrupt the RNA structure of Ψ interacting with Gag.

The Ψ region of ASLVs, located between the PBS and the splice donor site (SD), plays an essential role in genomic RNA packaging (20–22). Phylogenetic analysis predicts that this
region is highly structured (14). Deletion of nucleotides 207–270 within mΨ strongly reduces genomic RNA packaging (22).

A minimal packaging sequence has been identified that is sufficient for the packaging of a heterologous RNA (6). This sequence, named μΨ, is located between nucleotides 156 and 237. Phylogenetic and mutational analyses indicate that the 160–235 sequence forms the O3 stem and three stem–loops termed O3SLa, O3SLb and O3SLc (Fig. 1A) (6,14). The O3 stem, the uORF3 initiation codon, and the bottom part of O3SLb and O3SLc stems are the most important elements for efficient RNA packaging and Gag–RNA interactions (6,15,16,19,23–25). It is possible that access of Gag to these elements is promoted by the R–U5–PBS region. In other words, specific interactions between the R–U5–PBS and mΨ sequences might play a role in mΨ tertiary structure formation.

The main goal of the work presented below was to determine whether the R–U5–PBS region interacts with the minimal packaging signal in ASLV. To identify RNA–RNA interactions occurring in vitro between the R–U5–PBS and mΨ sequences, analysis of RNA complexes by agarose gel electrophoresis and chemical and enzymatic probing were used. We showed that there is a specific interaction between the R–U5–PBS and mΨ regions. To map the sequences involved in the R–U5–PBS/mΨ interaction, antisense DNA oligonucleotides and site-directed mutagenesis were used. We also investigated the role of RNA dimerization in this interaction using RNAs which are able and unable to dimerize.

MATERIALS AND METHODS

Construction of plasmid DNAs

Nucleotides of the avian leukemia virus of subgroup A (ALV-RSA termed ALV in the text) are numbered according to the Pr-C strain of RSV (26). In the oligonucleotide sequences, the upper-case letters indicate the bases that are complementary to the DNA sequence of ALV. Plasmids pCG44 and pLAD4, a generous gift from J.-L. Darlix (ENS Lyon, France), contain the 1–629 sequence and the complete sequence with truncated LTRs of the ALV genome, respectively (27,28). Plasmid pPFSA266 differs from pCG44 by a single-base substitution at position 266 (29). Mutations were made by PCR-based mutagenesis using the expand™ high fidelity PCR system (Roche Molecular Biochemicals). Plasmid pNVBS1 was constructed by PCR amplification of pLAD4 by using oligonucleotides 5’-gtctagagctctCTCTGAGCCGC-3’ (629–617) and 5’-gccccGGGTGACCaGACGTGATGTT-AGG-3’ (103–126). The PCR product was cloned into the pGEM-TEasy vector (Promega) to generate the BstEII–XhoI insert (103–629). The insert was ligated into pCG44 digested with BstEII and XhoI. Plasmid pIK156 was generated by PCR amplification of pLAD4 by using oligonucleotides 5’-gaccttgactcatGACTCCGTCCTCATC-3’ (156–170) and 5’-gcctgacgCTGCTCCCCTAAGC-3’ (225–222). The resulting PCR product was digested with EcoRI and XhoI and ligated into pCG44 digested with the same enzymes. Cloned sequences and mutations were verified by DNA sequencing.

In vitro RNA synthesis and purification

Plasmids pCG44, pPFSA266 and pNVBS1 were linearized with XhoI to generate templates for the in vitro synthesis of RNAs 1–626wt, 1–626SA266 and 1–626PBS1, respectively. Plasmids pCG44 and pNVBS1 were digested with Sau3AI to generate RNAs 1–159wt and 1–159PBS1, respectively. Plasmid pIK156 was digested with PvuII to generate RNA 156–235. Five micrograms of the linearized plasmid was transcribed with T7 RNA polymerase under the conditions stipulated by the RibomAX™ large scale RNA production system (Promega). RNAs 1–626wt, 1–626SA266 and 1–626PBS1 were purified as described previously (27). RNAs 1–159wt, 1–159PBS1 and 156–235wt were purified by electrophoresis on a 12% denaturing polyacrylamide gel as described previously (29). RNA 156–235wt was treated with alkaline phosphatase from Escherichia coli (Amersham Biosciences) and 5’-end labeled using T4 polynucleotide kinase (New England Biolabs) and [γ-32P]ATP (ICN). The 5’-end labeled RNA was purified by electrophoresis on a 12% denaturing polyacrylamide gel and isolated by elution and ethanol precipitation.

Analysis of RNA–RNA interactions by agarose gel electrophoresis

Labeled RNA 156–235wt (38 pmol) at 2 × 10⁹ c.p.m./pmol was dissolved in 6 µl of water, heated at 90°C for 2 min and chilled for 2 min on ice. Then, 4 µl of 2.5-fold concentrated L buffer was added (final concentrations: 50 mM Tris–HCl, pH 7.5, 125 mM KCl, 5 mM MgCl₂) and the sample was incubated for 15 min at 30°C. Unlabeled RNA 1–159wt (38 pmol) underwent the same heat denaturation and renaturation treatment and was then added, or not, to RNA 156–235wt. Ten microliters of 1× L buffer was added to RNAs 1–159wt and 156–235wt alone. Reaction mixtures were incubated for 60 min at 37°C. For the assays with oligonucleotides dPBS, A25 (5’-CCAATGTGGTGAATGGT-CA-3’) and A187 (5’-TCCCGATA-3’), the same protocol was used, except that 190 pmol of oligonucleotides dPBS and A25 or 380 pmol of oligonucleotide A187 was added after 30 min of the incubation at 37°C, and the incubation was then continued for 30 min at 37°C.

RNAs 1–626wt, 1–626PBS1 and 1–626SA266 were tested for their ability to form loose homodimers as described previously (29). Namely, 7.6 pmol of RNA, in 12 µl of double-distilled water, was heated at 90°C for 2 min and chilled for 2 min on ice. Eight microliters of 2.5-fold concentrated L buffer was added and the sample was incubated for 15 min at 20°C followed by incubation for 60 min at 37°C. For the assay with oligonucleotide dPBS (5’-ATCACGTGGGCTCA-CCA-3’), the same protocol was used, except that 38 pmol of dPBS was added after 30 min of the incubation at 37°C, and then the incubation was continued for 30 min at 37°C. At the end of incubations at 37°C, all samples were put on ice, mixed with 5 µl of loading buffer (50% w/v glycerol, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol). For preparation of heat-denatured RNA, RNA (7.6 or 38 pmol) in 12 µl of double-distilled water, was heated at 90°C for 2 min and chilled for 2 min on ice, and mixed with 5 µl of loading buffer. Each sample was divided into two aliquots, which were analyzed by agarose gel electrophoresis at 4°C in TBM buffer.
(45 mM Tris–borate, pH 8.3, 0.1 mM MgCl₂) and at 25°C in 0.5× TBE buffer (45 mM Tris–borate, pH 8.3, 1 mM EDTA). The percentage of dimerization was determined as described previously (27). The labeled nucleic acids were detected by autoradiography as described elsewhere (27).

**Enzymatic RNA probing with RNase T1**

Labeled RNA 156–235wt (3.8 pmol) at 2 × 10⁴ c.p.m./pmol was dissolved in 12 μl of water, heated at 90°C for 2 min and chilled for 2 min on ice. Eight microliters of 2.5-fold concentrated LC buffer was added (final concentrations: 50 mM sodium cacodylate, pH 7.5, 125 mM KCl, 5 mM MgCl₂ and the sample was incubated for 15 min at 20°C followed by incubation for 30 min at 37°C. Then, RNA 1–159 was added, or not, and the sample was incubated for 30 min at 37°C. The sample was then incubated for 5 min at 20°C and 2 μg of a tRNA mixture from *E.coli* MRE 600 (Roche Molecular Biochemicals) was added. This was done in order to equalize the quantity of RNA in each assay. At this point, 0.16 U of RNase T1 (Life Technologies) was added, and the sample was incubated at 20°C for 10 min. The digestion was stopped by phenol–chloroform extraction followed by ethanol precipitation. Samples were analyzed by electrophoresis on a 12% denaturing polyacrylamide gel. Cleavage positions were identified by running in parallel RNA sequence markers generated by RNase T1 digestion and alkaline hydrolysis (30).

**Chemical RNA probing with DMS and kethoxal**

Experiments were repeated at least three times. RNAs 1–626wt, 1–626PBS1 and 1–626 SA266 in the absence or presence of oligonucleotide dPBS were incubated under conditions that allowed loose dimerization as described above, except that 50 mM Tris–HCl (pH 7.5) was substituted by 50 mM sodium cacodylate (pH 7.5) in the case of probing with DMS, and was substituted by 50 mM HEPES–KOH (pH 7.5) in the case of probing with kethoxal. After incubation for 60 min at 37°C, each assay was incubated for 5 min at 20°C before 2 μg of a tRNA mixture from *E.coli* MRE 600 was added. At this point, the RNAs were treated with 0.5% of DMS for 4 and 8 min at 20°C or with 0.3 mg/ml of kethoxal for 2, 8 and 12 min at 20°C. The DMS and kethoxal treatments were stopped by adding the DMS termination buffer (final concentrations; 0.2 M Tris–acetate, pH 7.5, 0.2 M β-mercaptoethanol, 0.3 M sodium acetate, 20 μM EDTA) or the kethoxal precipitation buffer (final concentrations; 22 mM boric acid, 0.3 M sodium acetate) following the protocols described by Kjemus et al. (31). The samples were immediately precipitated by ethanol. The precipitated RNA was dissolved in double-distilled water in the case of DMS probing, whereas it was dissolved in 10 mM Tris–HCl (pH 7.8), 50 mM potassium borate, 0.1 mM EDTA in the case of kethoxal probing (31). Modified bases were identified by primer extension with avian myeloblastosis reverse transcriptase (Roche Molecular Biochemicals) following the protocol as described previously (32). Reverse transcriptase stops dNTP incorporation at the residue preceding the site of modification. Primer extension of controls not treated with a chemical probe were run in parallel to detect nicks and pauses in the RNA template during reverse transcription. Oligonucleotides P157 (5'-TCGCCACGCGTCTG-3') and A223 (5'-GCCGCAGC-CCCTCCACCAG-3') were used as primers. The samples were analyzed by electrophoresis on an 8% denaturing polyacrylamide gel.

**Sequence analysis**

The strains and, in parentheses, their DDBJ/EMBL/GenBank accession numbers are: ADOL-7501 (AY027920), AEV (X12707), AEV-ES4 (X06197), ALV (M37980), ALV-J (Z46390), AMV (S74099), AMV-B (X51496), CT10 (Y00302), FuSV (AF033810), HBI (M11784), HPRS-103(A) (AY029745), IC4 (X77628), IC10 (X13744), IC11 (M62407), MAV-1/2 (L10923), MAV-1 (L10922), MAV-2 (L10924), MC29 (V01174), MH2E21 (M14008), PR2257 (M21526), PR2257/16 (L21974), PR2257T (X51863), RAV-2 (K02374), RCAS-J (AF416332), RSV-PrB (J02339), RSV-PrC (J02342), RSV-daPrC (X68524), RSV-SRA (L29198), RSV-SRA-V (U41731), RSV-SRB (AF052428), RSV-SRD (D10652), UR2 (M10455), and Y73 (V01170).

**RESULTS**

**Design of RNA transcripts**

To identify RNA–RNA interactions between the R–U5–PBS region and the μΨ sequence, RNA transcripts representing structural domains of ALV genomic RNA were generated by *in vitro* transcription (Fig. 1B). RNA transcripts used in this study do not contain additional sequences (resulting from DNA plasmid construction), which might induce irrelevant RNA–RNA interactions. The 1–626 RNA contains the complete 5′ leader of ALV genomic RNA that encompasses the R–U5–PBS region, μΨ and the L3 domain (nucleotides 239–298) that is essential for ALV RNA dimer formation *in vitro* (27,29). The first 626 nt at the 5′-end of ALV and RSV SR-A differ only by two nucleotides at positions 512 and 569 (33,34). Therefore, the 1–300 sequence probably forms the same secondary structure in ALV and RSV SR-A. On the basis of a phylogenetic analysis, two conformations for the first 300 nt of the 5′-end of RSV-SRA have been predicted (14). One conformation (Fig. 1A) is slightly more compatible with the RNase T1 cleavage sites (14). The RNA secondary structures formed by nucleotides 1–159 and 239–300 are identical in the two predicted conformations. Enzymatic probing data confirm that nucleotides 58–128 adopt the predicted secondary structure (35). The 1–159 and 156–235 transcripts can form the secondary structures that are predicted for the R–U5–PBS and μΨ regions, respectively.

**Enzymatic probing of RNA 156–235wt in the presence and absence of RNA 1–159wt**

To test the possibility of interactions between RNAs 1–159wt and 156–235wt, we probed the structure of RNA 156–235wt with RNase T1 (specific for unpaired guanines) in the absence and presence of RNA 1–159wt (Fig. 2A). In the absence of RNA 1–159wt, strong and moderate T1 cleavages occurred at G₃₈₄, G₁₈₅, G₁₉₃, G₂₀₇, G₂₀₉ and G₂₁₀, all of which are predicted to lie within a loop or bulge (Fig. 2B). Although, G₁₉₇ is predicted to lie at the end of a stem, it was a moderate T1 target. This observation suggests that base pairing between C₁₉₇ and G₁₆₇ is transient. As expected, the majority of guanine residues in the predicted stems were not cleaved by
Therefore, the RNase T1 cleavage pattern of RNA
156–235wt alone was consistent with the secondary structure
proposed for μY in the ALV leader (Fig. 1A). In the presence
of RNA 1–159wt, the RNase T1 cleavage pattern of RNA
156–235wt did not change, except for G184 and G185. Indeed,
G184 and G185 became markedly protected upon incubation
with increasing amounts of RNA 1–159wt (Fig. 2A, lanes 3–
5). Our data suggest that the terminal loop of O3SLa interacts
with nucleotides of RNA 1–159wt.

RNAs 1–159wt and 156–235wt form loose RNA
complexes

If the interaction between RNAs 1–159wt and 156–235wt is
stable enough, then the 1–159wt/156–235wt complex should
be detected by native agarose gel electrophoresis. To test this,
RNA 1–159wt and RNA 156–235wt were incubated alone and
together and analyzed by native agarose gel electrophoresis in
TBM buffer as described in Materials and Methods (Fig. 3A.

Figure 1. Predicted secondary structure and genetic organization for the 5′-end of ALV genomic RNA. (A) Postulated secondary structure for the first 300 nt
of ALV RNA (14). The short ORFs uORF1 (nucleotides 41–61), uORF2 (nucleotides 82–129) and uORF3 (nucleotides 198–224) are indicated by bold letters.
The internal loop of PBS and the terminal loop of O3SLa are boxed in gray. (B) ALV RNA transcripts used in this study. R, terminal repeat; U5, sequence
unique and untranslated at the 5′-end; P, the replication primer iRNA<sup>rep</sup> binding site (PBS); μY, minimal packaging signal composed of the O3 stem and
three stem–loops (O3SLa, O3SLb and O3SLc) (6); L3, domain containing the L3 hairpin required for ALV RNA dimerization in vitro (29); AUG, initiation
codon for Pr76<sup>gag</sup>; SD, splice donor site. Numbering is relative to the genomic RNA cap site (+1). The sequences 107–113 (PBS internal loop) and 263–269
(L3 terminal loop) are indicated at the top of RNAs. Mutations are indicated by lower-case letters.
We found that RNA 156–235wt homodimerized very poorly and the homodimer was only detectable by autoradiography (Fig. 3C, compare lane 4 to lane 5). About two thirds of RNA 1–159 molecules remained in a monomeric form (Fig. 3A, lane 5). However, one third of RNA 1–159wt was in a dimeric form. This observation was unexpected because the dimerization signal in ASLV is the L3 stem–loop formed by nucleotides 258–274 (27,29). Furthermore, a single-base substitution in the L3 stem–loop completely abolished dimerization of RNAs containing the complete 5′ leader (29). We tested oligonucleotides targeting three predicted stem–loops in the 1–159 region (Fig. 1A) and found that oligonucleotide A25 directed against nucleotides 6–25 abolished tight dimerization of RNA 1–159wt (Fig. 3B, lane 8). This result suggests that the subunits of RNA 1–159 dimer are linked by the imperfect autocomplementary sequence (nucleotides 15–27), which can form a terminal stem–loop. As mentioned above, large RNAs containing the complete 5′ leader do not dimerize via the 15–27 sequence. Two explanations can be proposed. First, the three-dimensional folding of large RNAs prevents the stem–loop, formed by nucleotides 15–27, to interact with its counterpart. Second, the secondary structure adopted by the 15–27 sequence is not the same in large RNAs and RNA 1–159 molecules that can dimerize. As shown in Figure 3A (lane 4), two new bands appeared when RNAs 1–159wt and 156–235wt were incubated together. The positions of these bands suggest that RNAs 1–159wt and 156–235wt formed the 1–159wt/156–235wt and (1–159wt)2/156–235wt complexes. The same samples were also analyzed by native agarose gel electrophoresis in 0.5× TBE (Fig. 3B). Under these conditions, the complexes formed by RNAs 1–159wt and 156–235wt were not observed (Fig. 3B, lane 4). These observations suggest that the complexes dissociated during electrophoresis when the electrophoresis buffer was devoid of MgCl2. Therefore, Mg2+ probably stabilized the linkage structure that held the subunits of RNA complexes. These complexes are named loose RNA complexes because they exhibit similar electrophoretic properties to those of loose dimer formed by RNA transcripts derived from ALV (29) and HIV-1 (36,37). It has been shown that a short duplex composed of six Watson–Crick base pairs holds together the subunits of loose dimers (29,38). In contrast, the subunits of tight dimers are associated by an extended duplex involving at least 16 base pairs (29,37,39). Therefore, the subunits of loose RNA complexes formed by RNAs 1–159wt and 156–235wt are probably linked by one or several short duplexes.

Characterization of sequences involved in the formation of loose RNA complexes

To identify the sequences that link the subunits of loose RNA complexes, antisense oligonucleotides were added to preformed complexes, and their ability to disrupt the complexes was tested. RNA 1–159wt formed a tight dimer that was observed after migration on both TBM and 0.5× TBE gels (Fig. 3A and B, lanes 5). The upper band (Fig. 3A, lane 4) probably corresponds to the ternary complex formed by one RNA 156–235 molecule bound to dimeric RNA 1–159wt. However, a quaternary complex composed of one RNA 1–159wt molecule and three RNA 156–235wt molecules would migrate at the same position. Oligonucleotide A25 inhibiting

Figure 2. Enzymatic probing of RNA 156–235wt with RNase T1. (A) 5′-end-labeled RNA 156–235wt alone (lanes 1 and 2) or with RNA 1–159wt (lanes 3–5) was incubated in the absence (lane 1) and in the presence of RNase T1 (lanes 2–5) as described in Materials and Methods. RNA 1–159wt/RNA 156–235wt ratios were 1/1, 2/1 and 4/1 (lanes 3–5). T1 and L refer to RNase T1 sequencing and to alkaline ladder, respectively. Arrows point out guanines cleaved by RNase T1. The reactivity changes induced by RNA 1–159wt are indicated by asterisks. (B) Secondary structure diagram displaying the T1 cleavage sites. Black, gray, empty and dotted arrows indicated strong, medium, weak and marginal cleavage sites, respectively.
The dimerization of RNA 1–159wt should prevent ternary complex formation. The absence of a significant mobility shift of monomeric RNA 1–159wt suggests that the oligonucleotide/RNA hybrid dissociated during electrophoresis in 0.5× TBE. Therefore, annealing of oligonucleotide A25 to its target was probably not complete. A band migrating slower than the band corresponding to dimeric RNA 1–159wt was observed when the assay containing RNA 1–159wt and oligonucleotide A25 was subjected to electrophoresis on a TBE gel (Fig. 3A, lane 8). The position of this band corresponds to dimeric RNA 1–159wt bound to oligonucleotide A25. These data suggest that RNA 1–159wt loose dimerized poorly when it was partially annealed to oligonucleotide A25. In the presence of oligonucleotide A25, RNAs 1–159wt and 156–235wt formed only one type of RNA complex corresponding to one RNA 156–235wt molecule bound to monomeric RNA 1–159wt (Fig. 3A, lane 7). The results of these analyses strongly support the assumption that both the monomer and the tight dimer of RNA 1–159wt bound only one RNA 156–235wt molecule unless A25 inhibited both 1–159 dimerization and quaternary complex formation. Dimeric RNA 1–159wt should contain two binding sites for RNA 156–235wt. An explanation is that the tertiary structure of (1–159wt)2/156–235wt complex sterically blocks the access of second RNA 156–235wt molecule to its binding site.

As mentioned previously, the enzymatic probing experiments with RNase T1 suggest that the interaction between RNAs 1–159wt and 156–235wt involves the terminal loop of O3SLa. To test this hypothesis we used oligonucleotide A187 (8 nt long) directed against this loop. It is expected that binding of this short oligonucleotide to its target does not change the folding of RNA 156–235wt. Oligonucleotide A187 almost completely dissociated the loose RNA complexes at an oligonucleotide/RNA ratio of 10/1 (Fig. 3C, lane 1). This result strongly suggests that the terminal loop of O3SLa is a component of the linkage structure that links the subunits of loose RNA complexes. To identify complementary sequences to the O3SLa loop, the 1–159wt sequence was examined. We found that the 109–113 sequence could interact through Watson–Crick base pairing with the 182–186 sequence of the O3SLa loop (see Fig. 7A). The 109–113 sequence forms an internal loop in the PBS sequence (nucleotides 102–119). To demonstrate the interaction between the internal loop and the O3SLa loop, we used oligonucleotide dPBS, which is complementary to the 18 nt of PBS. As expected, oligonucleotide dPBS completely abolished the interaction between RNA 1–159wt and RNA 156–235wt (Fig. 3A, lane 3). Bands migrating slower than the monomer and dimer bands of RNA 156–235wt were observed after migration on both TBE and TBE gels (Fig. 3A and B, compare lanes 3 to lanes 4 and 5). This observation indicates that the majority of oligonucleotide dPBS/RNA 1–159wt hybrids were stable because they did not dissociate during electrophoresis in 0.5× TBE. Therefore, oligonucleotide dPBS and RNA 1–159wt are probably able to form a stable duplex involving 18 bp when they are incubated at 37°C.

To confirm the base pairing interaction between nucleotides 109CCCGA113 and 182UCGGG186, RNA 1–159PBS1 was constructed (Fig. 1B). This mutant contains a 2 nt substitution in the internal loop (107ACCAGA113 instead of 107ACCCGA113) that should disrupt the predicted interaction. RNA 1–159PBS1 exhibits similar properties to those of RNA 1–159wt. First, probing experiments and the mfold program of Zuker (40) predicted that RNA 1–159PBS1 adopts

---

**Figure 3.** Analysis of complexes formed by RNA 1–159wt and RNA 156–235wt. The nucleic acids were visualized by ethidium bromide staining (top) and autoradiography (bottom). (A and B) Lanes 1 and 6, heat-denatured RNAs 156–235wt and 1–159wt, respectively. RNA 156–235wt was unlabeled in lanes 1. RNAs 156–235wt and 1–159wt alone (lanes 2 and 5) or together (lanes 3, 4 and 7), without oligonucleotide (lanes 2, 4–6) or with oligonucleotides dPBS (lanes 3) and A25 (lanes 7 and 8) were incubated at 37°C as described in Materials and Methods. Each sample was divided into two aliquots, which were analyzed by 1.8% agarose gel electrophoresis at 4°C in TBM buffer (A) and at 25°C in 0.5× TBE buffer (B). Lanes MK, 0.16–1.77 kb RNA ladder (Life Technologies). The free oligonucleotides A25 and dPBS are indicated by fo. (C) Lanes 3, 4 and 8, heat-denatured RNAs 1–159wt, 156–235wt and 1–159PBS1, respectively. RNAs alone (lanes 5 and 7) or mixed (lanes 1, 2 and 6) were incubated at 37°C as described in Materials and Methods. Lane 1, RNAs were incubated with oligonucleotide A187. This short oligonucleotide (8 nt) cannot be visualized by ethidium bromide staining. Samples were analyzed by 1.8% agarose gel electrophoresis at 4°C in TBM buffer.
the same secondary structure as RNA 1–159wt (data not shown). Second, RNAs 1–159wt and 1–159PBS1 tightly dimerized moderately when they were incubated at 37°C under conditions that allowed formation of loose RNA complexes (Fig. 3A and B, lanes 5; Fig. 3C, lane 7; and data not shown). However, unlike RNA 1–159wt, RNA 1–159PBS1 was unable to form loose RNA complexes with RNA 156–235wt (Fig. 3C, compare lanes 2 and 6). Taken together, our results demonstrate that the R–U5–PBS region and the minimal packaging sequence can interact via the internal loop of PBS and the terminal loop of O3SLa.

The PBS/O3SLa interaction occurs in the complete 5′ leader of ALV genomic RNA

In light of the above-mentioned results, we wanted to ascertain whether the PBS/O3SLa interaction also occurs in the complete 5′ leader of ALV genomic RNA. For this purpose, we analyzed three RNAs 1–626 (Fig. 1B). RNA 1–626PBS1 cannot form the PBS/O3SLa interaction. To identify secondary structure formations that do not depend on dimerization, RNA 1–626SA266 has been used. Indeed, as previously described (29), RNA 1–626SA266 contains a single-base substitution that completely prevented RNA dimerization (Fig. 4, lane 4). To completely exclude an indirect effect of mutations on the folding of RNA transcripts, we compared the DMS probing patterns of nucleotides 65–120 in RNAs 1–626wt, 1–626SA266 and 1–626PBS1 (Fig. 5). DMS modifies the unpaired A and C residues. Due to the presence of pauses and the absence of a significant dose effect, we could not determine clearly the reactivity of C109

A113, A118

As expected, DMS modified C68, C73, A89, 91A92, A107, 110A111 and A126, all of which are predicted to lie within a loop or bulge (Fig. 1A). Our data are consistent overall with the predicted secondary structure of the 65–120 sequence (Fig. 1A). Therefore, the 109–113 sequence of the three RNAs 1–626 is probably within an internal loop. The residue A113 was reactive to DMS (Fig. 5), although it was predicted to base pair with U182 in the PBS/O3SLa interaction (see Fig. 7A). It is not surprising because residues that close to a stem are often accessible to chemical probes. The residue C109 was predicted to be involved in the PBS/O3SLa interaction and it was indeed unreactive to DMS (Fig. 5). The reactivity of nucleotides 109–113 to chemical probes should increase when the PBS/O3SLa interaction is disrupted. However, the 109–113 sequence did not become more reactive to DMS and kethoxal when RNAs 1–626wt and 1–626SA266 were incubated in the absence of MgCl2 or in the presence of oligonucleotide A187 (data not shown). Furthermore, the reactivity of nucleotides 109–113 was not increased in RNA 1–159wt, which cannot form the PBS/O3SLa interaction (data not shown). In addition, C108 was unreactive to DMS (Fig. 5). These results suggest that nucleotides 109–113 did not become more accessible to chemical probes when the PBS/O3SLa interaction was disrupted because they form a structured internal loop. Therefore, chemical probing of the PBS sequence cannot demonstrate the existence of the PBS/O3SLa interaction in RNAs 1–626wt and 1–626SA266.

To test the possible base pairing interaction (Fig. 7A) between the internal loop of PBS and the terminal loop O3SLa, we analyzed by chemical probes the 182–186 sequence of RNAs 1–626wt, 1–626SA266 and 1–626 PBS1. As expected (see the discussion above about A113), the residue U182 was slightly reactive to 1-cyclohexyl 3-(2 morpholino-ethyl) carbodiimide metho-p-toluene sulfonate (CMCT) (data not shown). Due to the presence of a pause, we could not determine the reactivity of C183 to DMS (data not shown). The reactivity level of 184GGG186 to kethoxal (specific for unpaired guanines) was very low when RNA 1–626wt was incubated in the presence of MgCl2 (Fig. 6A). We previously showed that MgCl2 stabilized the PBS/O3SLa interaction linking the subunits of loose RNA complexes formed by RNAs 1–159wt and 156–235wt (Fig. 3A and B). Therefore, the putative base pairing interaction between nucleotides 109CCCGA113 and 182UCGGG186 in RNA 1–626wt should be disrupted in the absence of MgCl2. As expected, a significant increase in kethoxal sensitivity for 184GGG186 was observed when the incubation buffer was devoid of MgCl2 (Fig. 6A). In addition, a significant increase in kethoxal sensitivity for 184GGG186 was also observed when oligonucleotide dPBS was annealed to the PBS sequence (Fig. 6A). The reactivity of 184GGG186 to kethoxal was strongly increased by two mutations in the PBS sequence that abolished the PBS/O3SLa interaction (Fig. 6C). These results strongly suggest that 184GGG186 base paired with 109CCCGA113 in RNA 1–626wt. The 184GGG186 sequence of RNA 1–626SA266 also became reactive to kethoxal in the presence of oligonucleotide dPBS or in the absence of MgCl2 (Fig. 6B). These data are fully consistent with the formation of a base pairing interaction between 184GGG186 and 109CCCGA113 in monomeric RNA 1–626SA266.

The PBS/O3SLa interaction is not required for dimerization of ALV RNAs containing the complete 5′ leader

The fact that the PBS/O3SLa interaction occurs in monomeric RNA 1–626SA266 suggests that this interaction does not
play a role in genomic RNA dimerization. To test this hypothesis, RNAs 1–626wt and 1–626PBS1 were incubated under conditions in which the PBS/O3SLa interaction occurred. As previously described (29) and shown in Figure 4 (lane 6), RNA 1–626wt dimerized. Oligonucleotide dPBS hybridized to both monomeric and dimeric RNA and did not dissociate the homodimer of RNA 1–626wt into monomers (Fig. 4, lanes 7 and 8). In addition, RNA 1–626PBS1 dimerized efficiently, although it contained the 110AA111 mutation that prevented the PBS/O3SLa interaction (Fig. 4, lane 2). These results clearly show that the PBS/O3SLa interaction is not required for dimerization of ALV RNA 1–626.

Phylogenetic analysis of the PBS/O3SLa interaction

To examine the functional relevance of the PBS/O3SLa interaction, we searched for its conservation among the ASLV strains. We included defective transforming viruses in our phylogenetic analysis because their packaging signal is probably functional. Indeed, these viruses are transferable by helper virus. In contrast, endogenous viruses were not included because they would not necessarily have a functional packaging signal. Thirty-three ASLV sequences encompassing the PBS and μY regions were found in GenBank. Among them, 32 contain the wild-type PBS that is complementary to the 3′-end of tRNA\(^{\text{Trp}}\). The PBS sequence of ALV-J displays one single base change compared to the wild-type PBS. This base change may be due to a sequencing mistake, and does not alter the CCCGA sequence that can interact with the UCGGG sequence of O3SLa. ALVs that use tRNA\(^{\text{Trp}}\) have a growth advantage over ALVs that use other tRNAs (41). Therefore, it is not surprising that all ASLV strains contain a PBS that is complementary to the 3′-end of tRNA\(^{\text{Trp}}\). The 33 ASLV sequences were aligned according to the consensus described for ASLV strains (6). Eight types of sequences were found in the region forming the terminal loop of O3SLa (Fig. 7B). All of the sequences, with the exception of MC29, conserve the potential to form four base pairs with the CCCG sequence located in the internal loop of PBS. The conservation of the CGGG sequence suggests that it is functionally important. The UCGGG sequence, complementary to the CCCGA sequence of PBS, is conserved in 25 strains. Except in the MAV strains, the PBS/O3SLa interaction involving five base pairs can occur in all replication competent viruses (for example, ALV and RSV strains).

Figure 5. Chemical probing of the U5/PBS region with DMS. Modifications of A(N-1) and C(N-3) with DMS were carried out under conditions described in Materials and Methods. Lanes C are controls without DMS treatment. RNAs 1–626SA266, 1–626wt and 1–626PBS1 were incubated with DMS for 4 min (lanes t1) or 8 min (lanes t2). Sequence lanes (A, C, G and U) were run in parallel. Arrows point out the modified bases.
DISCUSSION

The present study demonstrates that the R–U5–PBS region can interact with the minimal packaging signal in ALV. Indeed, we found that RNA transcripts representing the R–U5–PBS and mY regions formed loose RNA complexes. Using enzymatic probing, antisense oligonucleotides and site-directed mutagenesis, we clearly showed that the subunits of loose RNA complexes are linked by a base pairing interaction involving the internal loop of PBS and the terminal loop of O3SLa. The base pairing possibilities and the electrophoretic properties of loose RNA complexes strongly suggest that the base pairing interaction is restricted to five base pairs (Fig. 7A). Despite the presence of Mg²⁺ during electrophoresis on a native agarose gel at 4°C, we found that the level of loose RNA complexes was low (Fig. 3). This is not surprising since the dimerization yield of HIV-1 RNAs varied from 12 to 80% depending on the sequence of the 6 bp duplex forming the linkage structure (42,43). One explanation is that the short duplex composed of 5 bp is partially stable during electrophoresis in TBM buffer. In other words, Mg²⁺ partially prevents the dissociation of loose RNA complexes during electrophoresis.

To confirm that the PBS/O3SLa interaction exists in the complete 5’ leader of ALV genomic RNA, we analyzed by chemical probes RNA transcripts containing this region. For RSV, the internal loop of PBS in the context of the 5’ leader has been investigated by enzymatic probing (35). The residue G₁₁₂ was a moderate T1 target. We found that this residue was unreactive to kethoxal (data not shown). We cannot easily explain this minor discrepancy between our results and those reported by Morris and Leis (35). However, consistent with our results, the 109–113 sequence of RSV displayed very limited accessibility to the single-strand-specific RNase A. Using three separate experimental approaches (absence of Mg²⁺, binding of oligonucleotide dPBS to the PBS sequence and mutations in the internal loop of PBS), we found that the guanine residues located in the terminal loop of O3SLa became strongly reactive to kethoxal under conditions that were expected to disrupt the PBS/O3SLa interaction. We used the mfold program of Zuker (40) and several chemical probes to analyze the secondary structure adopted by the 1–400 region of ALV genomic RNA (unpublished results). From this analysis, we constructed a secondary structure model for the 1–400 region that was the most consistent with probing data. This model contains the PBS/O3SLa interaction. Taken together, our results strongly suggest that the base pairing interaction between the internal loop of PBS and the terminal loop of O3SLa occurs in the 5’-end of ALV genomic RNA.

The results obtained with oligonucleotide dPBS and mutations within the internal loop of PBS (Fig. 4) demonstrated that the PBS/O3SLa interaction is not required for dimerization of ALV RNA. Consistent with these results, it has been reported for ALV that the PBS sequence does not play a critical role in the stability of dimer RNA in vivo (44). Conversely, the absence of reactivity of 184GGG₁₈₆ to kethoxal in monomeric RNA 1–626SA266 and dimeric RNA 1–626wt showed that the accessibility of the O3SLa terminal loop does not depend on ALV genomic RNA dimerization. On the other hand, the reactivity of 184GGG₁₈₆ strongly increased under conditions that disrupted the PBS/O3SLa interaction in monomeric RNA 1–626SA266 (Fig. 6). These findings strongly suggest that the internal loop of PBS interacts with the terminal loop of O3SLa through intramolecular base pairing in monomeric and dimeric RNAs 1–626. However, the dimerization yield was 10% lower with RNA 1–626PBS1 than with RNA 1–626wt (average of three independent experiments). A likely interpretation is that the tertiary structure of RNA 1–626wt is slightly more appropriate for dimerization than that of RNA 1–626PBS1. We favor this
interpretation because the PBS/O3SLa interaction probably belongs to the class of structural elements called pseudoknots, which are involved in RNA three-dimensional folding (45–50). Classically, the formation of pseudoknots involves Watson–Crick base pairing between a hairpin loop and a complementary sequence outside that loop. It should be
emphasized that pseudoknots are generally stabilized by Mg$^{2+}$ (51,52). Pseudoknots are known to have functional roles. For example, pseudoknots resulting only from three base pairs formation are important to ribosome function (53,54). The PBS/O3SLa interaction may be formed in vivo since it occurs under in vitro conditions (temperature, salt concentrations) that are close to physiological conditions. The PBS/O3SLa interaction may be important for efficient replication of ASLV since the sequences forming this structural element are highly conserved (Fig. 7B). However, naturally occurring compensatory sequence (covariant) changes were not found. This is not surprising because there is a strong selective pressure to maintain the PBS complementary to tRNA$^{\text{Trp}}$ (41,55). The CCGGA sequence of PBS has only the potential to base pair with the UCGGG and UUGGG sequences. The stability of the PBS/O3SLa interaction should be reduced with the UUGGG sequence. The UUGGG sequence was not found in the O3SLa terminal loop of ASLV strains (Fig. 7B). In contrast, except in the MC29 strain, all ASLV strains conserve the CGGG sequence. These observations suggest that the UGGG sequence cannot base pair with the PBS sequence, or the PBS/O3SLa interaction formed with the UGGG sequence does not provide a growth advantage to the virus. However, we cannot exclude the possibility that the conservation of the CGGG sequence is not correlated with the PBS/O3SLa interaction.

This is the first in vitro study that shows that two sequences located in the 5′ leader of retroviral RNA can interact to form probably a pseudoknot. Until now, only retroviral pseudoknots that involve at least one region of gag, pro or pol genes have been described (2,56). The fact that the PBS/O3SLa interaction involves the PBS and the minimal packaging signal suggests that this specific interaction could regulate the processes of tRNA$^{\text{Trp}}$ primer annealing and RNA packaging. It is possible that the PBS/O3SLa interaction prevents primer annealing before packaging of genomic RNA into the virion. However, an alternative hypothesis is that the interaction promotes primer annealing after RNA packaging. Indeed, it is conceivable that the PBS/O3SLa interaction destabilizes the base pairing interactions involving the PBS and U5 sequences (stems formed by nucleotides 69–72/114–117 and 75–78/103–106, see Fig. 1A). Whitcomb et al. (41) produced a series of mutations that changed the PBS of ALV so that it was homologous to the 3′ ends of six different chicken tRNAs. These mutations could prevent the PBS/O3SLa interaction since several of them were located in the internal loop of PBS. The PBS mutants bound the alternate tRNA primers and packaged approximately equal amounts of RNA per virion. These results suggest that the PBS/O3SLa interaction is not absolutely required for tRNA primer placement and RNA packaging. However, it is not completely excluded that the interaction was maintained in the mutant viruses. Indeed, the PBS sequences of mutants conserved the ability to form three to five base pairs with the terminal loop of O3SLa.

There are reports that support the involvement of the PBS/O3SLa interaction in the ASLV RNA packaging process. Indeed, it has been shown that heterologous RNAs containing only the minimal packaging signal (nucleotides 156–237) are packaged 2.6-fold less efficiently than genomic RSV RNA (6,57). This finding suggests that sequence(s) and/or structure(s) outside of μΨ are required for optimal RNA packaging. The PBS/O3SLa interaction could be the additional RNA structure that promotes RNA packaging. RNA folding constraints are not the same in genomic RNA and heterologous RNAs. Therefore, the PBS/O3SLa interaction may be more important for RNA packaging in natural context than in a heterologous context. Interestingly, it has been demonstrated that an extension of the stem upstream of the PBS sequence caused a significant decrease in the amount of genomic RNA packaged into viral particles (12). In addition, this extension increased the RNase A sensitivity for nucleotides located in the PBS internal loop (35). In contrast, the RNA packaging defect and increase in RNase A sensitivity were not produced by nucleotide insertions in the terminal loop upstream of the PBS sequence (12,35). These data suggest that an extension of the stem composed of nucleotides 82–88 and 94–100 (see Fig. 1A) prevents the PBS/O3SLa interaction by altering the overall topology of the 5′ leader region of the genomic RNA. Taken together, the data of previous studies are consistent with the hypothesis that the PBS/O3SLa interaction increases the RNA packaging efficiency although it is dispensable for this process. We favor a model in which the PBS/O3SLa interaction indirectly promotes the interaction of the Gag protein with specific sites within μΨ. The PBS/O3SLa interaction may prevent the O3SLa stem–loop from interfering with access of Gag to its binding sites in μΨ.

Since the 5′ leader of ASLV genomic RNA contains stable secondary structures and three ORFs upstream of the Gag start codon, it is possible that the RNA folding constraints induced by the PBS/O3SLa interaction regulate the initiation of Gag translation. Furthermore, the PBS sequence is located within uORF2 that is poorly translated (17). This may indicate that the PBS/O3SLa interaction prevents ribosome access to the uORF2 initiation codon. Translational initiation can be driven by RNA structural elements referred to as internal ribosome entry sites (IRESs). The 5′ leader of ASLV genomic RNA contains a bipartite IRES that lies from the PBS sequence to the Gag start codon (18). Therefore, the PBS/O3SLa interaction is probably involved in IRES three-dimensional folding and IRES activity. It has been shown that a pseudoknot can be essential for IRES activity (58).

Finally, an attractive hypothesis is that the PBS/O3SLa interaction has been selected by evolution to allow that primer tRNA annealing, RNA packaging and initiation of Gag translation occur with the optimal efficiency. An analysis of sequences from a variety of retroviruses reveals that each PBS sequence has up to three possibilities to form a pseudoknot with sequences located in the terminal loops of stem–loops (Fig. 7C). These stem–loops contribute to the function of packaging signals (4,5,59,60). Interestingly, the sequences involved in the putative pseudoknots are poorly or not reactive against chemical and enzymatic probes that interact with unpaired nucleotides (61–63). Thus, it is tempting to speculate that formation of a pseudoknot acting as a cis-regulatory element in the leader is common among retroviruses.

ACKNOWLEDGMENTS

We thank Marylène Mougel for critical reading of the manuscript. This work was supported by grant 9398 from the Association pour la Recherche sur le Cancer (to P.F.). I.K., N.V. and H.D.-O. were the recipients of postdoctoral


