In Vitro Evidence for a Long Range Pseudoknot in the 5′-Untranslated and Matrix Coding Regions of HIV-1 Genomic RNA*

Jean-Christophe Paillart‡§, Eugene Skripkin‡§, Bernard Ehresmann‡, Chantal Ehresmann‡, and Roland Marquet‡

From the ‡UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, Strasbourg F-67084, France

The 5′-untranslated leader region of human immunodeficiency virus type 1 (HIV-1) RNA contains multiple signals that control distinct steps of the viral replication cycle such as transcription, reverse transcription, genomic RNA dimerization, splicing, and packaging. It is likely that fine tuned coordinated regulation of these functions is achieved through specific RNA-protein and RNA-RNA interactions. In a search for cis-acting elements important for the tertiary structure of the 5′-untranslated region of HIV-1 genomic RNA, we identified, by ladder selection experiments, a short stretch of nucleotides directly downstream of the poly(A) signal that interacts with a nucleotide sequence located in the matrix region. Confirmation of the sequence of the interacting sites was obtained by partial or complete inhibition of this interaction by antisense oligonucleotides and by nucleotide substitutions. In the wild type RNA, this long range interaction was intramolecular, since no intermolecular RNA association was detected by gel electrophoresis with an RNA mutated in the dimerization initiation site and containing both sequences involved in the tertiary interaction. Moreover, the functional importance of this interaction is supported by its conservation in all HIV-1 isolates as well as in HIV-2 and simian immunodeficiency virus. Our results raise the possibility that this long range RNA-RNA interaction might be involved in the full-length genomic RNA selection during packaging, repression of the 5′ polyadenylation signal, and/or splicing regulation.

The genomes of RNA viruses are multifunctional molecules. In retroviruses, including human immunodeficiency virus type 1 (HIV-1), the primary RNA transcript functions as pre-mRNA (splicing), mRNA (synthesis of Gag and Gag-Pol proteins), and genomic RNA for packaging into infectious particles. The 5′-untranslated leader region of the HIV-1 RNA genome contains cis-acting signals of recognition for proteins and RNAs responsible for regulating several crucial steps of the viral life cycle. This region includes a long terminal repeat consisting of the R (repeat) and U5 (unique at 5′) regions and the primer binding site (PBS), as well as exon 1 leader sequences downstream of U5 (Fig. 1A) (1).

The secondary structure of the 5′-untranslated region of HIV-1 genomic RNA has been extensively studied (2–5), and it is now well documented that almost all functional sites in the 5′-end fold into independent structural hairpin loop domains (Fig. 1B). It has been shown that disruption of either of these motifs is critical for several steps in the viral life cycle. The TAR hairpin is essential for Tat-mediated activation of viral transcription (6–8), reverse transcription (9, 10), and packaging (7, 11). The immediately adjacent poly(A) hairpin is critical for repression of the proximal poly(A) site (12–17) and encapsidation (7, 13). The PBS region forms a complex RNA structure that influences the binding of the tRNA5′ primer to the PBS and its subsequent elongation by reverse transcriptase (18–21). The region between the PBS and translation initiation site includes four distinct hairpin structures in different steps: initiation of genomic RNA dimerization (DIS) (22–26), splicing (SD) (27, 28), encapsidation (Psi) (29–31) and translation initiation (AUG) (2–4). In addition, several of these RNA motifs have been suggested to contribute to selective packaging of viral genomic RNA (32–36).

Despite numerous studies aimed to probe the structure of the 5′-untranslated region of HIV-1 genomic RNA, very little is known about its tertiary structure in vitro as well as in the virion. It is likely that specific RNA-protein and RNA-RNA interactions allow fine tuned coordinated regulation of the different functional sites in this region and permit the compaction of the genomic RNA in a 120-nm particle. Recent articles reported that the leader region of HIV-1 RNA can adopt a compactly folded structure (32) and that a conformational RNA switch could regulate various functions in the viral life cycle (5). In a search for cis-acting elements important for the tertiary structure of the 5′-untranslated region of HIV-1 genomic RNA, we used ladder selection experiments to identify a short stretch of nucleotides directly downstream of the poly(A) signal that interacts with a nucleotide sequence located in the matrix coding region. We report that site-directed mutagenesis disrupting either of these sequences inhibits the long distance interaction. Similarly, antisense oligonucleotides efficiently inhibit the interaction. The functional significance of this long range pseudoknot is further supported by phylogenetic sequence analysis that revealed conservation of this interaction in the genomes of all HIV-1, HIV-2, and SIV isolates.

**EXPERIMENTAL PROCEDURES**

Plasmid Construction for in Vitro Transcription—Site-directed mutagenesis using the QuikChange™ site-directed mutagenesis kit was

---

*This work was supported by grants from the “Agence Nationale de Recherches sur le SIDA” and a “Jeunes Equipes” grant from the Center National de la Recherche Scientifique (to R. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67094 Strasbourg cedex, France. Tel.: 33-3-88-41-70-91; Fax: 33-3-88-60-22-18; E-mail: JC.Paillart@ibmc.u-strasbg.fr.

§ Present address: Memorial Sloan-Kettering Cancer Center, Box 557, York Ave., New York, NY 10021.

¶ The abbreviations used are: HIV-1, human immunodeficiency virus type 1; PBS, primer binding site; DIS, dimerization initiation site; NC, nucleocapsid protein; Psi, packaging signal; poly(A), polyadenylation; SD site, splice donor site; SIV, simian immunodeficiency virus.
conducted according to the manufacturer (Stratagene), using plasmid DNAs pJCB and pJCB DIS(−) containing the 615 first nucleotides of HIV-1 Mal isolate, with a wild-type or mutated DIS sequence, respectively (24). Primers 5’-CTTAAGCCTCAATAAACGAAGCCTTGAGTG-(sense) and 5’-GGCGTACTCACAACTAGAGATCCC-3’ (antisense) were used to amplify plasmid pJCB, giving rise to plasmid pB-CG77/UU460 containing both substitutions was obtained by amplification of pB-CG77. Thermocycling was for 1 min at 95 °C, followed by 16 cycles of 30 s at 94 °C, 1 min at 55 °C, and 8 min at 68 °C. The same procedure was applied to plasmid pJCA (24) containing region 305–311, and 123–311 of HIV-1 genomic RNA, respectively. RNA-(305–311) and RNA-(1–123) were linearized with PvuII, XbaI, and AflII, and then the temperature was gradually increased by 1 °C steps. After a 5-min incubation at the appropriate temperature, an aliquot was loaded on a 1% (w/v) agarose gel after the addition of glycerol (20% final concentration) and run as previously described. Monomers and shifted RNA species were visualized after autoradiography or by using a BAS 2000 BIO-Imager (Fuji).

Thermal Stability of the Long Range Interaction—To determine the thermal stability of the long range interaction, samples were incubated 30 min at 37 °C, and then the temperature was gradually increased by 7 °C steps. After a 5-min incubation at the appropriate temperature, an aliquot was loaded on a 1% (w/v) agarose gel after the addition of ethidium bromide staining and quantified with the MacBas (Fuji) software. The melting temperature of the shifted RNA, T_m, was defined as the temperature at which the fraction of shifted RNA was reduced by 2-fold, as compared with its value at 37 °C.
Enzymatic RNA Probing with RNase T2—In a standard experiment, 400 nM 1–156 or 1–615 RNAs was dissolved in 8 μl of water, heated for 2 min at 90 °C, chilled on ice, and renatured for 30 min at 37 °C in 50 mM sodium cacodylate (pH 7.5), 5 mM MgCl₂, 300 mM KCl. After renaturation, the samples were cooled at room temperature for 10 min before treatment with RNase T2 (15 min at 37 °C; 0.002 units/μl). The positions of RNase hydrolysis were detected by primer extension with avian myeloblastosis reverse transcriptase as previously described (2).

Ladder Selection Experiments—5′- and 3′-end-labeled RNA molecules (400 nM) were submitted to limited alkaline hydrolysis in 50 mM NaCO₃ (pH 8.9) during 4 min at 90 °C. Alkaline ladders were neutralized with 300 mM sodium acetate (pH 5.6), ethanol-precipitated, and used in the mobility shift assay with the adequate nonhydrolyzed RNA partner (400 nM). Monomers and shifted RNA molecules were cut out from the 1% low melting agarose gel and extracted with phenol (v/v) for 15 min at 50 °C. After ethanol precipitation, RNA fragments were resuspended in 4 μl of formamide-containing loading buffer and separated by denaturing gel electrophoresis on an 8% acrylamide gel. RNase T1 (G-specific) and RNase U2 (A-specific) sequencing reactions were run in parallel to detect the borders of the selected population of RNA fragments. After autoradiography, films were scanned, and densitograms of the lanes corresponding to the initial alkali ladder and the selected RNA fragments were obtained with the program Quantity One (Bio-Rad).

Inhibition of the RNA-RNA Interaction by Oligodeoxyribonucleotides—Synthetic DNA oligodeoxyribonucleotides complementary to positions 313–334, 335–366, 367–394, 395–420, and 441–460 of the HIV-1 Mal sequence were used in the mobility shift assay. Briefly, the antisense oligodeoxyribonucleotide (400 nM or 1.6 μM) was first incubated with 400 nM of 3′-end-labeled RNA (305–615) in the binding buffer for 15 min at 37 °C. The RNA-oligonucleotide complexes were then incubated for 30 min at 37 °C with an equimolar amount of RNA (1–311) DIS(−) and analyzed by agarose gel electrophoresis. The fraction of monomer and shifted RNA molecules were quantified using a BAS 2000 Bio-Imager (Fuji) as previously described (24, 39).

RESULTS

RNA Sequences Upstream and Downstream of the SD Site Interact Together—It has previously been reported by us (25, 26, 39) and by others (22, 23) that the 5′-untranslated region of HIV-1 genomic RNA contains the major dimerization signal (DIS) but that other regions around the DIS could influence the scaffold of the RNA tertiary structure (24, 40). To better understand the HIV-1 RNA folding and to identify potential long distance contacts, we synthesized a set of HIV-1 RNAs differing by their length by in vitro transcription with T7 polymerase (Fig. 2A). We analyzed the shift induced by labeled RNA-(305–615) on this set of RNAs by native agarose gel electrophoresis (Fig. 2B).

As expected from previous studies (24), RNA-(305–615) is unable to dimerize (Fig. 2B, lane 2). Unexpectedly, we observed a significant mobility shift when labeled RNA-(305–615) was incubated with unlabeled RNA-(1–311) (Fig. 2B, lane 5). Indeed, several shifted species could be observed, probably because this region of the HIV-1 leader RNA contains the DIS hairpin structure (nucleotides 265–287). Indeed, the number and position of the shifted bands indicated that they most likely corresponded (from the bottom to the top of the gel) to one RNA-(305–615) molecule bound to monomeric RNA-(1–311), and one or two RNA-(305–615) molecules bound to dimeric RNA-(1–311). In keeping with this interpretation, only one shifted band was observed when labeled RNA-(306–615) was incubated with an unlabeled RNA-(1–311) bearing point mutations in the DIS loop that prevented DIS-mediated RNA dimerization (Fig. 2B, lane 6).

When labeled RNA-(305–615) was incubated with unlabeled RNA-(1–615) or RNA-(1–615) DIS(−), only very weak shifted bands were detected (Fig. 2B, lanes 3 and 4). Our interpretation of this result is that the interaction between the sequences located upstream and downstream of the SD site was intramolecular in RNA-(1–615) (and RNA-(1–615) DIS(−)). Thus, the downstream site of the long RNA efficiently competed with the homologous site of the truncated RNA-(305–615) for binding to the upstream site, reducing the level of the intermolecular interaction with RNA-(305–615). Taken together, these results suggest that the long distance interaction is independent of the RNA dimerization process and that this interaction requires two elements apart from the SD site that interact with each other through intramolecular base pairing in RNA-(1–615).

Characterization of the Sequences Involved in the Long Distance Interaction—To further characterize the long distance interaction, we next analyzed RNA mobility shift by using a
FIG. 3. Determination of RNA sequences involved in the long distance interaction by ladder selection. A, schematic diagram of the strategy used to define the 5’ border of the 5’ interaction domain using 5’-end labeled RNA-(1-311). Similar strategies were used to define the 3’ border of the 5’ interaction domain and the 5’ and 3’ borders of the 3’ interaction domain. After statistical alkali hydrolysis of 32P-labeled RNA-(1-311), the ladder was incubated with RNA-(305-615). Shifted molecules were visualized on native agarose gel and purified, and the
truncated version of the 1–615 RNA. To avoid dealing with multimeric complexes on agarose gels (Fig. 2B, lane 5), we used dimerization-deficient RNAs (DIS (−); see “Experimental Procedures”). Both RNAs starting at positions 100 and 123 were unable to shift RNA-(305–615) (Fig. 2B, lanes 10 and 11). A similar result was obtained with RNA-(1–126) (Fig. 2B, lane 7). On the contrary, RNA-(1–102) and RNA-(1–152) gave a shift with a yield comparable with the one obtained with RNA-(1–311) DIS (–) (Fig. 2B, lanes 8 and 9). These results indicated that one of the sequences required for the interaction was located between nucleotides 62 and 102, corresponding to the poly(A) hairpin loop (Fig. 1B). Similar experiments were conducted using unlabeled 3′-truncated RNAs starting at position 123 or 305 with labeled RNA-(1–311) DIS (–). They allowed us to delimit the 3′ interacting domain downstream of nucleotide 415 (data not shown).

To test the number of base pairs involved in the long range interaction, we performed thermal denaturation experiments of the complex formed by RNA-(1–311) DIS (−) and RNA-(305–615) (Fig. 2C). The T_m of this complex is 43–45 °C. This value is fully comparable with the T_m obtained with the DIS hairpin of HIV-1, where the two RNA monomers are able to interact through six Watson-Crick base pairs (24).

**Ladder Selection Experiments—**In an attempt to map more precisely the 5′ and 3′ borders of the RNA domains that form the long distance interaction, we performed ladder selection experiments. In these experiments, the RNA subfragments, obtained from a pool of RNAs generated by mild alkaline hydrolysis of either 5′- or 3′-end-labeled RNA-(1–311) or RNA-(305–615), were selected for their capability to retain binding to their unlabeled RNA partner (Fig. 3A). After extraction of the bound RNA fragments from the agarose gel (see “Experimental Procedures”), selected molecules were analyzed by electrophoresis on a denaturing polyacrylamide gel (Fig. 3A), and the selection boundaries were determined from the densitograms of the initial and selected RNA species. The boundaries were identified as the starting points of a strong continuous selection. Peaks corresponding to isolated selected fragments, such as those observed at position 134 in Fig. 3B and positions 434, 436, and 447 in Fig. 3C, were not taken into account, since they most likely reflect artifactual selections due to aberrant folding of these RNA fragments.

When the 1–311 ladder was used in the mobility shift assay with unlabeled RNA-(305–615), the size of the retained RNA subfragments varied from full-length down to position U 84 for 615), the size of the retained RNA folding of these RNA fragments. They most likely reflect artifactual selections due to aberrant, were not taken into account, since 434, 436, and 447 in Fig. 3. Tm of 32P-labeled RNA fragments 1–615 RNA. To avoid dealing with

Fig. 4. Enzymatic probing experiments with Ribonuclease T2. RNA-(1–156) (lane 1, 400 nM) was incubated with increasing amounts of RNA-(123–615) (lane 2, 140 nM; lane 3, 400 nM; lane 4, 1200 nM) and treated with RNase T2 (0.002 units/μl) for 15 min at 37 °C. Lane 5, RNA-(1–615) (400 nM) treated with RNase T2. Sites of RNase hydrolysis were identified by primer extension with avian myeloblastosis reverse transcriptase as previously described (2). A sequencing reaction of the polyadenylation region was performed in parallel.

selected molecules were separated on a 8% denaturing polyacrylamide gel. B and C, ladder selection experiments with 5′-end and 3′-end 32P-labeled RNA fragments 1–311 (B) and 305–615 (C). Lanes 1, RNAs submitted to RNase T1 digestion; lanes 2, RNAs submitted to RNase U2 digestion; lanes 3, RNAs statistically hydrolyzed with alkali; lanes 4, RNAs statistically hydrolyzed with alkali and that have been selected by the corresponding RNA partner. The borders corresponding to the start of strong selection are indicated in boldface type on the left of the gels. The densitograms of lanes 3 (gray) and 4 (black) are shown beside the autoradiographs.
RNase T2. These data reinforces the hypothesis that nucleotides 77–83 constitute the 5’ interaction site. It is worthy of note that the RNase T2 accessibility of the poly(A) hairpin loop was independent of the integrity of the DIS loop (data not shown). Assays to test the downstream sequence were unsuccessful, probably due to the structural versatility of this domain (results not shown).

Inhibition of the Long Distance Interaction by Antisense Oligonucleotides—In a second step, to validate the implication of region 457–463 in the long distance interaction, we wondered whether the interaction could be inhibited by antisense DNA oligonucleotides directed against this region (Fig. 5). We were particularly interested to understand the behavior of the matrix region, since no convincing results were obtained from probing experiments. RNA-(305–615) and the different antisense DNA oligonucleotides were heat-annealed as described under “Experimental Procedures” and were further used in the gel mobility shift assay. As shown in Fig. 5, some antisense DNA molecules had no effect on the shift of RNA-(305–615) (Fig. 5, AS313–334 and AS367–394). An additional band was observed with AS-(313–334) that could be explained by the induction of a conformational switch of the RNA by DNA annealing, as previously observed (32). Annealing of AS-(335–366) and AS-(395–420) partially inhibited the RNA shift, and AS-(441–460) almost completely prevented RNA-(305–615) from shifting (Fig. 5). Taken together, those results confirm our ladder selection data (Fig. 3) but raise the possibility that multiple domains in the matrix coding sequence might directly or indirectly affect the long range interaction. This observation which correlates with the absence of clear probing information, suggests that the tertiary interaction depends on particular features of the global versatile structure of the Matrix coding region.

Inhibition of the Long Distance Interaction by Site-directed Mutagenesis—To test the putative base pairing interaction (Fig. 6A) between the poly(A) and the matrix coding regions, mutants were constructed in which the predicted interaction was disrupted. Fig. 6B shows the three types of mutations that were introduced in the different size RNA fragments (1–311 DIS(−), 305–615, or 1–615 DIS(−)). Note that as previously explained, all mutant RNAs tested in this study were mutated in the DIS hairpin loop, so that only one shifted species can be formed. Mutant RNA CG77 contains a four-nucleotide substitution in the poly(A) hairpin loop (5′CGAA3′ instead of GCCUU); RNA UU460 contains 460UUCG463 instead of AAGC in the matrix region, and the compensatory mutant RNA CG77/ UU460 contains both substitutions. Those mutations have been designed with the final aim of studying their effect in a proviral context; thus, they do not change the sequence of the Gag protein. Indeed, mutations around amino acids 35–37 of the matrix protein have been shown to have deleterious effects on viral replication and assembly (for a review, see Ref. 41).

We next analyzed the capacity of the RNA mutants to give a shift on agarose gel with different RNA partners (Fig. 6, C–E). In the first set of experiments (Fig. 6C), we showed that the substitution of the poly(A) hairpin loop in the 1–311 context almost completely suppressed the interaction with RNA-(305–615) (Fig. 6C, compare lanes 2 and 4). Similar results were obtained with RNA-(123–615) (data not shown). On the other hand, the mutation in the matrix coding region only partially inhibited the interaction with DIS(−) RNA-(1–311) (Fig. 6C, lanes 4 and 5). Similarly, the base pairing capacity was not completely restored when using a pair of RNA mutants containing complementary sequences (Fig. 6C, lanes 2 and 3). The residual interaction between RNA-(1–311) and RNA-(305–615) or RNA-(305–615) and RNA-(1–311) might be due to the complementarity between the 7′AAGC7′ sequence in the poly(A) hairpin loop and 459′GCUU462′ in the mutated region. The inefficient trans-complementation between the two mutated sequences suggests the existence of an important structural element that would be disrupted in the UU460 mutant.

To test this possibility, we investigated the capability of wild-type or mutant RNA-(1–311), (305–615), and (1–615) to shift RNA-(1–615) bearing either the poly(A) or the matrix mutation (Fig. 6, D and E). Remember that wild type RNA-(1– 311) and (305–615) only weakly interacted with wild type RNA-(1–615), due to the long range intramolecular pseudoknot in the latter RNA (Fig. 2 and data not shown). If the mutation...
The long range pseudoknot in the 5′-end of HIV-1 genomic RNA has been shown to interact with the poly(A) hairpin loop and the matrix domain. A model of the long distance interaction between the poly(A) hairpin loop and the matrix domain is shown in Fig. 6A. B. Localization of the substitutions in the poly(A) and matrix regions. C-F. RNA mobility shift assays. The 32P-labeled RNA fragment is marked with an asterisk and incubated with the corresponding RNA as identified at the tops of the lanes. The position of the monomers and shifted RNA fragments (300/300, 300/600, or 600/600) is marked by an asterisk.

A model of the long distance interaction between the poly(A) hairpin loop and the matrix domain. B. Localization of the substitutions in the poly(A) and matrix regions. C-F. RNA mobility shift assays. The 32P-labeled RNA fragment is marked with an asterisk and incubated with the corresponding RNA as identified at the tops of the lanes. The position of the monomers and shifted RNA fragments (300/300, 300/600, or 600/600) is indicated. E has been visualized after ethidium bromide staining. All RNAs used in these experiments had their DIS mutated to prevent DIS-mediated dimerization, except wild type RNA-(1–615) (E). d, dimers of wild type RNA-(1–615).

Introduced in the 5′ or 3′ interaction site of RNA-(1–615) does disrupt the intramolecular long range interaction, then the wild-type remaining 5′ or 3′ site should be able to interact with wild type RNA-(305–615) or -(1–311), respectively. On the contrary, no shift should be observed between these RNAs if the mutated sequences are not involved in the long range interaction. Indeed, RNA-(1–615) CG77, unlike wild type RNA-(1–615), strongly interacted with labeled RNA-(1–311) (Fig. 6D, lanes 1 and 2). Moreover, no interaction was observed between RNA-(1–615) CG77 and RNA-(1–311) CG77 (Fig. 6D, lane 4), indicating that the interaction between RNA-(1–615) CG77 and wild type RNA-(1–311) was mediated by the poly(A) hairpin loop of the latter RNA. Similarly, labeled RNA-(305–615) interacted with RNA-(1–615) UU460 but not with RNA-(1–615) CG77 (Fig. 6D, lanes 6 and 7), again supporting the involvement of the poly(A) hairpin loop in the long range interaction. In addition, RNA-(1–615) UU460 did not interact with RNA-(305–615) UU460 (Fig. 6D, lane 9), as expected if the mutated region were directly involved in the long range interaction.

Finally, we tested the effects of the CG77 and UU460 mutations using various combinations of RNAs 615 nucleotides in length (Fig. 6E). With the exception of the control RNA used in lane 1, all RNAs were mutated in the DIS. Thus, RNA-(1–615) CG77 (lane 2) and RNA-(1–615) UU460 (lane 3) migrated as the expected monomeric species. Furthermore, we already showed that in RNA-(1–615), the long range interaction is intramolecular (Fig. 2). Therefore, an RNA mutated in the DIS and bearing wild-type sequences in the poly(A) hairpin loop and matrix coding region also migrated as a monomer (lane 2). Similarly, the migration as a monomer of the RNA bearing the compensatory mutations CG77 and UU460 reflects either an intramolecular long range interaction or the absence of such an interaction (lane 5). To address this question, we used a variety of RNA-(1–615) combinations (Fig. 6E, lanes 6–10). Remarkably, only combination of RNA-(1–615) UU460 and RNA-(1–615) CG77 was able to form a stable complex (lane 10). This trans-complementation strongly suggests that regions CG77 and UU460 are indeed those involved in the long range interaction.

The fact that trans-complementation only occurs between large fragments and not between short fragments suggests that the negative effect of the UU460 mutation is more pronounced in the truncated RNA than in the intact RNA. This can be correlated with the versatility of the matrix domain that is more sensible to its context than stable regions. Combined together, these results strongly suggest that the sequences we mutated in the poly(A) hairpin loop and in the matrix coding region constitute the 5′ and 3′ interaction site of the long range pseudoknot.

Conservation of a Long Distance Pseudoknot in HIV-1, HIV-2, and SIV—To test the biological significance of the long range interaction we identified, we performed an extensive sequence comparison of the 5′-polyadenylation signal region and the matrix coding sequence around amino acids 35–37 in human and simian lentiviruses (available on the World Wide Web at hiv-web.lanl.gov). We took all sequences of the nucleotide alignments of HIV-1/HIV-2/SIV complete genomes (42)
Results

Interestingly, more variation was found among the poly(A) hairpin loop and the matrix coding sequences of HIV-2 and SIV isolates, but the alignment revealed the maintenance of the potential base pairing between these regions (Fig. 7A). Indeed, despite the lack of strong sequence homology within these two groups, we observed a high degree of conservative changes (GC changed to AU or vice versa). Regarding the HIV-2 group, the strongest co-variations were observed for the HIV-2_B_D205 and HIV-2_B_EHO isolates, where base deletions and substitutions in the poly(A) hairpin were compensated by base substitutions in the matrix sequence. All other HIV-2 and SIV isolates also demonstrated the same base pairing conservation, despite the lack of strong sequence homology within these two groups, we observed a high degree of conservative changes (GC changed to AU or vice versa).

Discusion

The genome of HIV-1 is composed of two homologous RNA molecules about 9200 nucleotides long, which are packaged in a 120-nm diameter particle. The virus has developed specific mechanisms to package its genome, involving either RNA-RNA (DIS) or RNA-protein (NC-DIS, NC-Psi) interactions. However, other RNA-RNA interactions must be present all along the genome to allow its compaction and its correct folding.

In this study, we provide strong evidence that a long distance pseudoknot exists in the 5′-end of HIV-1 genomic RNA. By using RNA fragments of different lengths and ladder selection on the first 600 nucleotides of HIV-1 genomic RNA, we identified a heptanucleotide sequence located immediately downstream of the polyadenylation signal (77GCUUGCC83) that interacts with a complementary sequence located in the matrix coding region, about 400 nucleotides downstream of the polyadenylation signal (77GCUUGCC83) that interacts with a complementary sequence located in the matrix coding region, about 400 nucleotides downstream of the polyadenylation signal (77GCUUGCC83). The overall architecture of the poly(A) hairpin is very well conserved among lentiviruses but that the poly(A) signal in the SIVSYK173 isolate is also very remarkable. Indeed, it was also possible to draw a putative long distance interaction between the poly(A) hairpin loop and the matrix region when one of the two interacting sequence is mutated in the poly(A) region of isolate SIVSYK173 is located upstream of the poly(A) signal (see “Results”).

The high pressure of selection to conserve the long distance interaction between the poly(A) hairpin loop and a region located in the matrix coding region in HIV-1, HIV-2, and SIV supports the functional importance of this novel tertiary interaction in HIV-1 replication.

Alignment of the sequences surrounding the poly(A) signal and the matrix region was not very informative as far as we considered only the HIV-1 isolates. Indeed, these two sequences were absolutely conserved in all HIV-1 clades. This conservation suggests that the two aligned sequences are functionally important, but it did not give us any information about their interaction.

Interestingly, more variation was found among the poly(A) hairpin loop and the matrix coding sequences of HIV-2 and SIV isolates, but the alignment revealed the maintenance of the potential base pairing between these regions (Fig. 7A). Indeed, despite the lack of strong sequence homology within these two groups, we observed a high degree of conservative changes (GC changed to AU or vice versa). Regarding the HIV-2 group, the strongest co-variations were observed for the HIV-2_B_D205 and HIV-2_B_EHO isolates, where base deletions and substitutions in the poly(A) hairpin were compensated by base substitutions in the matrix sequence. All other HIV-2 and SIV isolates also demonstrated the same base pairing conservation, ranging from 6 (SIVMM251) to 10 (SIVVER963) base pairs for the putative interaction (Fig. 7A). For those having an increased number of base pairs (8–10), it was interesting to note the appearance of one (HIV-2_B_EHO, SIVVER963, SIVHJOEST, and SIVAGM677) or two (SIVGRF677) GU base pairs, suggesting that the overall stability of the long-range pseudoknot must be maintained within narrow limits.

The SIVSYK173 isolate is also very remarkable. Indeed, it was also possible to draw a putative long distance interaction with the matrix domain, but in this case, the complementary sequence was located upstream of the poly(A) signal (Fig. 7A). Interestingly, it has been shown by phylogenetic analysis that the overall architecture of the poly(A) hairpin is very well conserved among lentiviruses but that the poly(A) signal in the SIVSYK173 isolate is shifted toward the 3′ side of the loop due to a nine-nucleotide duplication (12) (Fig. 7B). Thus, the sequence complementary to the matrix coding region is part of the poly(A) hairpin loop, and therefore, it is available to interact with its putative partner.

The identification of the present long distance RNA-RNA interaction raises a number of questions about its biological...
of the SD site to prevent encapsidation of spliced RNAs. Indeed, the packaging signal of HIV-1 RNA is multipartite (35, 44), and regions outside Psi (TAR, poly(A), DIS, Gag) contribute to optimal packaging (7, 11, 34, 36, 45–47). The long distance interaction between the poly(A) hairpin and the matrix coding region could facilitate the discrimination between unspliced and spliced RNAs by the retroviral Gag proteins. This hypothesis is consistent with a recent study showing that substitution of the sequence directly downstream of the poly(A) signal reduced the ratio of genomic/spliced RNAs in virions (7) but does not affect the dimerization efficiency of the RNA genome (48).

Poladenylation—Due to the duplication of the poly(A) signal at both ends of the HIV-1 genome, a fine tuned mechanism must exist to restrict the proximal poly(A) signal used. Two main inhibition mechanisms have been proposed: proximity of the transcription initiation site (49) and occlusion by Tat (50) or by the SD site via its interaction with U1 snRNP (51, 52). It should be noted that an RNA encompassing the hairpin loop interacts with a sequence further downstream in the matrix coding region, one can easily imagine that the access of poladenylation factors would be impeded by such a mechanism. It should be noted that an RNA encompassing the interaction site in the matrix coding domain was not able to interact in vitro with a 600-nucleotide-long RNA corresponding to the 3′-end of the HIV-1 genome (data not shown).

Translation—Likewise, it is conceivable that the long range pseudoknot might regulate translation of the gag gene. It has recently been suggested that the HIV-1 leader region could form alternative structures in vitro and that a conformational RNA “switch,” from a rodlike structure to a branched structure, would regulate key steps in the replication cycle, such as translation to dimerization, packaging, and reverse transcription (5). Interestingly, the long range pseudoknot can only take place in the branched structure, since the poly(A) hairpin does not exist in the rodlike structure. Although additional experiments are required to test this possible NC-mediated RNA “switch,” it has been previously shown ex vivo that HIV-1 unspliced RNA constitutes a single pool that can function interchangeably as mRNA and as genomic RNA (53–56). This mechanism differs from the one described for HIV-2, which packages RNA predominantly in cis (i.e. the newly synthesized Gag preferentially encapsidates the RNA from which it was produced) (54, 55).

Long distance interactions have been involved in the regulation of RNA synthesis and/or gene expression in a variety of (+)strand RNA viruses (57, 58) and eukaryotic mRNAs (59). Our structural data demonstrate the existence of a long range pseudoknot in the 5′-end of the HIV-1 genome, although additional experiments are necessary to understand its exact role in the replication cycle. However, the demonstration of the functional role of the proposed interaction may prove rather difficult. First, the involved poly(A) region is duplicated at both 5′- and 3′-ends of the genome. Thus, if one mutates the 5′ site only, one can only study a single replication cycle because the 3′ site will be mutated during reverse transcription. If both sites are mutated, it will be difficult to distinguish between effects due to mutation of the functional 3′ poladenylation site and disruption of the long range interaction. Preliminary results of single replication cycle experiments did not allow us to detect significant differences between wild-type and mutant viruses regarding Gag expression or viral particle release, thus suggesting that the long distance interaction is not involved in Gag translation, assembly, and protein maturation. This negative result does not rule out the existence of this long range pseudoknot in vitro and its functional role. Indeed, we are presently developing chemical probing of HIV RNA in infected cells and inside the viral particles, and our preliminary data are consistent with the proposed interaction. In addition, HIV and SIV have had thousands of years of evolution to select for features that only slightly increase the viral fitness. Of course, such features cannot be tested in a single replication cycle. For instance, mutations of the dimerization initiation site of the HIV-1 RNA produce a packaging defect that can only be observed in multiple cycle infections, although the genomic RNA of all retroviruses is dimeric (36, 46, 60). Furthermore, our phylogenetic analysis strongly supports the functional importance of this tertiary interaction.

REFERENCES

1. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997) Retroviruses, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
2. Baudin, F., Marquet, R., Isel, C., Darlix, J.-L., Ehresmann, B., and Ehresmann, C. (1983) J. Mol. Biol. 229, 382–397
3. Berkhou, B. (1996) Prog. Nucleic Acids Res. Mol. Biol. 54, 1–34
4. Harrison, G. P., and Lever, A. M. L. (1992) J. Virol. 66, 4144–4153
5. Huthoff, H., and Berkhou, B. (2000) RNA T4, 143–157
6. Aboul-ela, F., Karr, J., and Varani, G. (1995) J. Mol. Biol. 253, 313–322
7. Clever, J. L., Eckstein, D. A., and Parslow, T. G. (1999) J. Virol. 73, 101–109
8. Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H., and Jones, K. A. (1998) Cell 92, 451–462
9. Harrich, D., Ulch, C., and Gaynor, R. B. (1996) J. Virol. 70, 4017–4027
10. Ohi, Y., and Clever, J. L. (2000) J. Virol. 74, 8324–8334
11. Helga-Maria, C., Hammarskjold, M. L., and Rekosh, D. (1999) J. Virol. 73, 4127–4135
12. Berkhou, B., Klaver, B., and Das, A. T. (1995) Virology 207, 276–281
13. Das, A., Klaver, B., Klasens, B., van Wamel, J., and Berkhou, B. (1997) J. Virol. 71, 2346–2356
14. Das, A. T., Klaver, B., and Berkhou, B. (1999) J. Virol. 73, 81–91
15. Gravely, B. R., Fleming, E. S., and Gilmartin, G. M. (1996) J. Biol. Chem. 271, 33654–33663
16. Klasens, B. I., Das, A. T., and Berkhou, B. (1998) Nucleic Acids Res. 26, 1870–1876
17. Klasens, B. I., Thiessen, M., Virtanen, A., and Berkhou, B. (1999) Nucleic Acids Res. 27, 446–454
18. Isel, C., Ehresmann, C., Keigh, G., Ehresmann, B., and Marquet, R. (1995) J. Mol. Biol. 247, 236–250
19. Isel, C., Lanchy, J. M., Le Grice, S. F., Ehresmann, C., Ehresmann, B., and Marquet, R. (1996) EMBO J. 15, 917–924
20. Isel, C., Westhof, E., Massire, C., Le Grice, S. F., Ehresmann, B., Ehresmann, C., and Marquet, R. (1999) EMBO J. 18, 1038–1048
21. Lanchy, J. M., Isel, C., Ehresmann, C., Marquet, R., and Ehresmann, B. (1996) Biochimie (Paris) 78, 1087–1096
22. Laughaer, M., and Jette, L. (1994) Biochemistry 33, 1346–1347
23. Muriaux, D., Girard, P. M., Bonnet, M. B., and Paolotti, J. (1985) J. Biol. Chem. 270, 8209–8216
24. Paillart, J.-C., Marquet, R., Skripkin, E., Ehresmann, B., and Ehresmann, C. (1994) J. Biol. Chem. 269, 27486–27493
25. Paillart, J.-C., Skripkin, E., Ehresmann, B., Ehresmann, C., and Marquet, R. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 5572–5577
26. Skripkin, E., Paillart, J.-C., Marquet, R., Ehresmann, B., and Ehresmann, C. (1996) Proc. Natl. Acad. Sci. U. S. A. 91, 4945–4949
27. Amarasighe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000) J. Mol. Biol. 301, 491–511
28. Amarasighe, G. K., De Guzman, R. N., Turner, R. B., and Summers, M. F. (2000) J. Mol. Biol. 296, 153–156
29. De Guzman, R. N., Wu, Z. R., Stalling, C. L., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) Science 279, 384–388
30. Pappalardo, L., Kerwood, D. J., Pelzer, I., and Borer, P. N. (1998) J. Mol. Biol. 282, 801–818
31. Zeffman, A., Hassard, S., Varani, G., and Lever, A. (2000) J. Mol. Biol. 297, 877–893
32. Berkhou, B., and van Wamel, J. L. (2000) RNA 6, 282–295

2 J. C. Paillart, unpublished results.
3 J. C. Paillart and M. Dettenhofer, manuscript in preparation.
Long Range Pseudoknot in the 5’-End of HIV-1 Genomic RNA

33. Clever, J. L., Wong, M. L., and Parslow, T. G. (1996) J. Virol. 70, 5902–5908
34. Clever, J. L., and Parslow, T. G. (1997) J. Virol. 71, 3407–3414
35. McBride, M. S., and Panganiban, A. T. (1996) J. Virol. 70, 2963–2973
36. Paillart, J.-C., Berthoux, L., Ottmann, M., Darlix, J.-L., Marquet, R., Ehresmann, B., and Ehresmann, C. (1996) J. Virol. 70, 8348–8354
37. Marquet, R., Baudin, F., Gabus, C., Darlix, J. L., Mougel, M., Ehresmann, C., and Ehresmann, B. (1991) Nucleic Acids Res. 19, 2349–2357
38. Jossinet, F., Paillart, J.-C., Westhof, E., Hermann, T., Skripkin, E., Lodmell, J. S., Ehresmann, C., Ehresmann, B., and Marquet, R. (1999) RNA 5, 1222–1234
39. Marquet, R., Paillart, J.-C., Skripkin, E., Ehresmann, C., and Ehresmann, B. (1994) Nucleic Acids Res. 22, 145–151
40. Laughrea, M., and Jette, L. (1997) Biochemistry 36, 9501–9508
41. Freed, E. O. (1998) Virology 251, 1–15
42. (1999) in Human Retroviruses and AIDS (Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Mullins, J., Wolinsky, S., and Korber, B., eds) pp. 201–296, Theoretical Biology and Biophysics, Los Alamos, NM
43. Paillart, J.-C., Marquet, R., Skripkin, E., Ehresmann, C., and Ehresmann, B. (1996) Biochimie (Paris) 78, 639–653
44. Berkowitz, R., Fisher, J., and Goff, S. P. (1996) Curr. Top. Microbiol. Immunol. 214, 177–218
45. Das, A. T., Klaver, B., and Berkhout, B. (1998) J. Virol. 72, 9217–9223
46. Laughrea, M., Jette, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M. (1997) J. Virol. 71, 3397–3406
47. Luban, J., and Goff, S. P. (1994) J. Virol. 68, 3784–3793
48. Laughrea, M., Shen, N., Jette, L., Darlix, J.-L., Kleiman, L., and Wainberg, M. A. (2001) Virology 281, 109–116
49. Weichs an der Glon, C., Monks, J., and Proudfoot, N. J. (1991) Genes Dev. 5, 244–253
50. Weichs an der Glon, C., Ashe, M., Eggermont, J., and Proudfoot, N. J. (1993) EMBO J. 12, 2119–2128
51. Ashe, M. P., Pearson, L. H., and Proudfoot, N. J. (1997) EMBO J. 16, 5752–5763
52. Ashe, M. P., Furger, A., and Proudfoot, N. J. (2000) RNA 6, 170–177
53. Butsch, M., and Boris-Lawrie, K. (2000) J. Virol. 74, 11531–11537
54. Dorman, N., and Lever, A. (2000) J. Virol. 74, 11413–11417
55. Kaye, J. F., and Lever, A. M. (1999) J. Virol. 73, 3023–3031
56. McBride, M. S., and Panganiban, A. T. (1997) J. Virol. 71, 2050–2058
57. Honda, M., Rinbrand, R., Abell, G., Kim, D., and Lemon, S. M. (1999) J. Virol. 73, 4941–4951
58. Zhang, G., Slowinski, V., and White, K. A. (1999) RNA 5, 550–561
59. Scherpe, W., Meinema, D., Holthuizen, P. E., and Sussenbach, J. S. (1995) Mol. Cell. Biol. 15, 235–245
60. Berkhout, B., and van Wamel, J. L. (1996) J. Virol. 70, 6723–6732
In Vitro Evidence for a Long Range Pseudoknot in the 5′-Untranslated and Matrix Coding Regions of HIV-1 Genomic RNA

Jean-Christophe Paillart, Eugene Skripkin, Bernard Ehresmann, Chantal Ehresmann and Roland Marquet

J. Biol. Chem. 2002, 277:5995-6004. doi: 10.1074/jbc.M108972200 originally published online December 13, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M108972200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 58 references, 35 of which can be accessed free at http://www.jbc.org/content/277/8/5995.full.html#ref-list-1