© 1993 Oxford University Press

Nucleic Acids Research, 1993, Vol. 21, No. 5 1171–1178

Secondary structure of the HIV-2 leader RNA comprising the tRNA-primer binding site

Ben Berkhout and Ilse Schoneveld
University of Amsterdam, Department of Virology, Academic Medical Center, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Received December 4, 1992; Revised and Accepted January 28, 1993

ABSTRACT

The initiation of reverse transcription of a retroviral RNA genome occurs by a tRNA primer bound near the 5′ end of the genomic RNA at a position called the primer-binding site (PBS). To understand the molecular basis for this RNA-RNA interaction, the secondary structure of the leader RNA of the human immunodeficiency virus type 2 (HIV-2) RNA was analyzed. *In vitro* synthesized HIV-2 RNA was probed with various structure-specific enzymes and chemicals. A computer program was then used to predict the secondary structure consistent with these data. In addition, the nucleotide sequences of different HIV-2 isolates were used to screen for the occurrence of co-variation among putative base pairs. The primary sequences have diverged rapidly in some HIV-2 isolates, however, some strikingly conserved secondary structure elements were identified. Most nucleotides in the leader region are involved in base pairing. An exception is the PBS sequence, of which 15 out of 18 nucleotides are exposed in an internal loop. These findings suggest that the overall structure of the HIV-2 genome has evolved to facilitate an optimal interaction with its tRNA primer.

INTRODUCTION

The 5′ leader region of retroviral RNA contains multiple sequences important for viral replication. These sequences do not code for proteins but are the *cis*-acting sites of recognition by proteins and RNAs responsible for mediating several phases of the viral life cycle. Reverse transcription of the retroviral genome, for example, is primed by a tRNA bound to an 18-nucleotide long, complementary region near the 5′ end of the genome. This region is termed the primer-binding site (PBS) (reviewed in 1–3). Furthermore, processes such as mRNA splicing, polyadenylation and translation are controlled by sequence elements in the leader transcript (reviewed in 2). Another leader region, termed Ψ, is necessary for the selective encapsidation of viral genomes into assembling virions (2). Although specific nucleotide sequences are involved in some of these processes, it may also be that the secondary or tertiary structure of the RNA leader is important. For instance, the human and simian immunodeficiency viruses (HIV and SIV) contain an additional *cis*-acting RNA element at the 5′ end of their RNA, the so-called TAR element (reviewed in 4). TAR forms the target for the transcriptional *trans*-activator protein Tat. Extensive mutational analyses indicated that both specific sequence and structure elements are critical for TAR RNA function (5–8).

Reverse transcription is mediated by the virion-associated enzyme reverse transcriptase (RT). Like all DNA polymerases, RT needs a primer with a free 3′ OH (hydroxyl). All retroviruses utilize a cellular tRNA molecule as primer, its identity depending on the particular virus species. In each instance 16–19 nucleotides at the 3′-CCA end of the tRNA are the exact complement of the corresponding viral PBS. Most mammalian retroviruses use proline tRNA, and avian retroviruses make use of tryptophan tRNA (1,2,9). Sequence analysis of the HIV-1 and HIV-2 viruses revealed a PBS corresponding to lysine tRNA (10), as is the case in other lentiviruses and the mouse mammary tumor virus. The temporal relationship of the association of the tRNA primer with the PBS relative to virus assembly, budding, genome dimerization, Gag protein maturation, *de novo* infection, and reverse transcription has not been rigorously characterized (2). Incorporation of the proper tRNA molecule into the virion was suggested to be the result of a specific interaction with the RT enzyme (11–14). However, these results were obtained with avian retroviruses and could not be reproduced with murine retroviruses (11,15). For the HIV-1 RT protein conflicting data on the specificity of $tRNA^{Lys,3}$ binding were published (16–18). To complicate matters even further, binding of the tRNA to the viral genome was suggested to be catalyzed by the nucleocapsid Gag protein (19). It is nevertheless reasonable to assume that, besides a potential role for the RT and Gag proteins, a critical role in tRNA selection is played by the PBS sequence itself.

There is some evidence for avian Rous sarcoma virus that the structure around the PBS plays a role in initiation of reverse transcription. A potential RNA structure for this region was proposed by Cobrinik *et al* (20,21, see also 22,23). Mutations outside the actual PBS sequence which disrupt this structure impair reverse transcription in infected cells, while mutations which alter the sequence but retain the stem structure have no effect. Studies with permeabilized viruses show that disruption

of the RNA structure does not affect the amount of tRNA primer bound to the viral RNA (20). However, a decrease in the incorporation of the first deoxynucleotides from the 3' OH terminus of the tRNA primer was measured. In an effort to understand the structural requirements of the PBS region in reverse transcription, we have biochemically analyzed the secondary structure of the RNA leader of human immunodeficiency virus type 2 (HIV-2). We previously reported on the structure of the TAR element at the extreme 5' end of the HIV-2 genome (24), here we probed a region extending 400 nucleotides into the leader transcript.

MATERIALS AND METHODS

Preparation of RNA

Two plasmids were constructed to facilitate transcription of HIV-2 RNA by the T7 RNA polymerase. Plasmid Blue-TAR2 contains a PstI-KpnI fragment of the HIV-2 molecular clone pROD10 (gift of Dr K.Peden) ligated downstream of the T7 promoter in the polylinker of Bluescript KS(+). This construct was previously described in detail (24) and encodes the HIV-2 leader up to position +420. Plasmid Blue-Gag2 contains the complete HIV-2 leader region (up to the PstI restriction site at position +886) in the correct orientation downstream of the T7 promoter in Bluescript KS(+). Both plasmids were linearized by restriction enzyme digestion in the plasmid sequences downstream of the HIV-2 insert (PvuII-digested Blue-TAR2 and EcoRI-digested Blue-Gag2). T7 transcripts were synthesized according to standard methods (25). Upon DNase treatment and deproteinization of the samples by phenol extraction, the RNA was recovered by ethanol precipitation, dissolved in renaturation buffer (10mM Tris−HCl pH7.5, 100mM NaCl) at approximately 1μg/μl and incubated at 72°C for 2 minutes, followed by slow cooling to 20°C.

Chemical modification and RNase treatment

The renatured RNA (1 μg per reaction) was treated with increasing amounts of the ribonuclease T1 (0−0.001−0.005−0.025 Units), cobra venom nuclease CV (0−0.001−0.003−0.010 Units) or nuclease S1 (0−0.1−0.3−0.9 Units) as previously described (24). After an incubation for 5 min at 37°C, the RNA sample was diluted in 100 μl TE (10 mM Tris−HCl pH7.5, 1mM EDTA), phenol-extracted and recovered by ethanol precipitation. HIV-2 RNA (1 μg in 200 μl) was treated with the single-strand specific chemicals diethyl-pyrocarbonate (DEP; 0−1-2−4 μl), dimethylsulfate (DMS; 0−0.6−1.2−1.8 μl) or kethoxal (Ket; 0−4 μl). Upon the addition of 10 μg carrier tRNA, the modified transcript was recovered by ethanol precipitation.

Primer extension analysis

The following DNA oligonucleotides were used to map the modified RNA positions in a primer extension reaction (the complementary HIV-2 coordinates are indicated): HIV-2 U5 5' AGGAGAGATGGGAGCAC 3' (+183 to +199) HIV-2 PBS 5' CTGTTCAGGCGCCAACCT 3' (+299 to +316) HIV-2 psi 5' TCCGTCGTGGTTTGTTCCTGC 3' (+373 to +393) Gag-2 AUG 5' TCTCAAGACGGAGTTTCTCGCGCCCAT 3' (+546 to +572) Blue-T3 5' ATTAACCCTCACTAAAG 3' (complementary to Bluescript sequences in Blue-TAR2 transcripts). Primers were end-labeled with [γ -^{32}P]ATP and T4 polynucleotide kinase. The labeled oligonucleotide (approximately 0.2 ng) was mixed with 0.2 μg modified transcript in a total volume of 12 μl annealing buffer (83 mM Tris−HCl pH 7.5, 125 mM KCl) and incubated for 2 min at 85°C, 10 min at 65°C and slowly cooled down to 25°C. 1 μl (200 Units) of Mo-MuLV reverse transcriptase and 6.5 μl RT buffer (9 mM $MgCl_2$, 30 mM DTT, 150 μg/ml Actinomycin D and 1.5 mM of each dNTP) were added, followed by a 5 min incubation at 37°C and 5 min at 42°C. Upon ethanol precipitation, the samples were taken up in 5 μl formamide sample buffer, denatured at 90°C and analyzed in a 6% acrylamide- 8 M urea gel. The end-labeled primers were also used in dideoxy-sequencing of the appropriate plasmid.

RESULTS

Chemical and enzymatic probing of HIV-2 RNA structure

In vitro synthesized HIV-2 leader RNA was treated with structure-specific probes, followed by primer extension analysis to localize the sites of modification or cleavage. Nucleotides sensitive to RNase S1 or RNase T1, as well as nucleotides that were modified with kethoxal (Ket), dimethyl sulfate (DMS) or diethyl pyrocarbonate (DEP), are assumed to not be involved in base-pairing or base-stacking interactions. Bases that reacted with cobra venom nuclease (CV) were taken to be either in double-stranded regions or single-stranded, stacked regions. The sites of modification were identified by primer extension analysis using

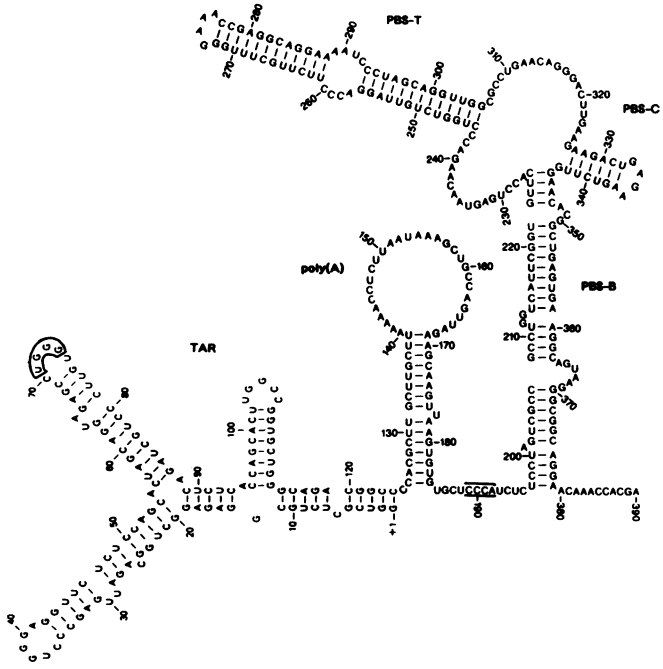


Figure 1. Model of the secondary structure of the HIV-2 leader RNA. Shown is the TAR element at the extreme 5'-end of the transcript (position +1 to +124), the hairpin containing the polyA signal AAUAAA (position +151 to +156) and the region around the PBS sequence (position +303 to +320), which is the perfect complement of the 3'-end of $tRNA^{Lys,3}$. A putative long-distance interaction between nucleotides in the second TAR loop (5'UGGG3', position +71/+74) and nucleotides in between the polyA and PBS structures (5'CCCA3', position +189/+192) is indicated by boxes. Consistent with this proposal, both regions were surprisingly insensitive towards the various single-strand reagents (Figure 2 and 7, reference 24). We like to note that while all leader sequences are present in the 5' end of HIV-2 RNA, only sequences up to position +172 are present in the repeat element at the 3' end.

several DNA primers positioned along the HIV-2 leader. The digestion data are summarized in Figures 4–7A, representative experiments are shown in Figures 2 and 3.

The result of all structure mapping experiments is the RNA structure model presented in Figure 1. This region encompasses the TAR element (position +1 to +124), of which the secondary structure was previously determined (24), the polyadenylation signal AAUAAA (position +151 to +156), and the 18 nucleotides of the PBS (position +303 to +320). For all regions of the leader, information on the secondary structure was obtained using at least two different primers and found to be comparable. In addition, we analyzed two HIV-2 transcripts that differ in their 3′ sequences. TAR-2 transcripts contain HIV-2 leader sequences from position +1 to +420, while Gag-2 transcripts run into the *gag* open reading frame (+1 to +886, including the *gag* startcodon at position +546). The results of the structure probing were comparable for these two RNA species. In the next sections, we will deal with the different structural domains depicted in Figure 1 in detail.

The polyA hairpin (position +125 to +184)

Treatment with several single-strand reagents (DEP,DMS,S1,T1) produced strong bands in the +140/+170 region, but not in the immediate upstream and downstream sequences (Figure 2). The results are summarized in Figure 4A, in which the locations of the nucleotides that were modified under native conditions are superimposed on the RNA secondary structure model. A stem-loop hairpin structure consistent with the digestion pattern can be drawn that presents the polyA signal (position +151/+156) in the single-stranded loop. We will refer to this structure as the polyA hairpin. Cobra venom nuclease reactivity occurs at multiple positions on both sides of the proposed single-stranded region, suggesting that these regions may be stacked on top of the base-paired stem. Two A-rich stretches are present in the loop; AAAAA at position +140/+144 and the polyA signal AAUAAA at position +151/+156. Interestingly, both stretches are highly reactive towards the chemicals DMS and DEP, while only the latter is susceptible to RNase S1 treatment. Because extensive base-stacking was observed in the AAAAA region, but not in the AAUAAA region, it is possible that base stacking or steric hindrance blocks access of the S1 enzyme.

Secondary structure predictions based on the primary sequence of the polyA region were made using several computer algorithms (26,27). Both programs predict the polyA hairpin, even without the inclusion of the modification data of Figure 4 as a restrictive parameter. The only difference is additional base-pairing between the 5′ and 3′ side of the loop (5′AAC3′ paired to 5′GUU3′), which is not consistent with the experimental data (Figure 4A).

Sequence comparison of RNA from different HIV-SIV isolates has been very instructive in solving the TAR RNA structure (24,28). We performed a similar phylogenetic analysis for the polyA region using nucleotide sequences from thirteen retroviruses (10). Besides seven HIV-2 isolates (Figure 4B, numbers 1–7), we included sequences from the closely related simian immunodeficiency viruses isolated from macaques and sooty mangabeys (SIV-MAC and SIV-SMM, numbers 8–11 and 12–13, respectively). The RNA structure of the polyA region is generally supported by the phylogenetic analysis (Figure 4B). The majority of the nucleotide substitutions observed were in the single-stranded loop and bulge regions and therefore do not affect

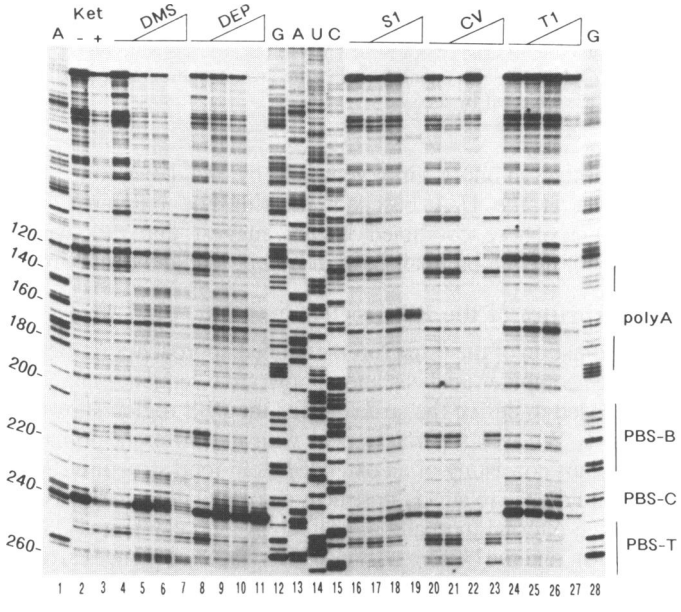


Figure 2. Nuclease digestion and chemical modification of TAR-2 RNA under native conditions. *In vitro* transcribed HIV-2 leader RNA was treated with limiting amounts of several single-strand specific reagents (Ket=kethoxal (G-specific), DMS=dimethyl sulfate (A,C-specific), DEP (diethyl pyrocarbonate (A-specific), S1=RNase S1, T1=RNase T1 (G-specific)) and the double-strand-specific RNase CV (Cobra Venom). Each treatment (indicated above the lanes) was performed with increasing amounts of the chemical-RNase (see materials and methods for details, left lanes represent mock incubations). Cleavage-modification sites were detected using primer-extension analysis with the HIV-2 psi primer. Electrophoresis was on 6% polyacrylamide/7M urea gels. For reference, the psi primer was used in a DNA sequence reaction (lanes 12–15; GAUC and lane 1 (A) and 28 (G). Numbers on the left represent HIV-2 coordinates (+1 being the transcription start site). The position of double-stranded stem regions is schematically indicated on the right.

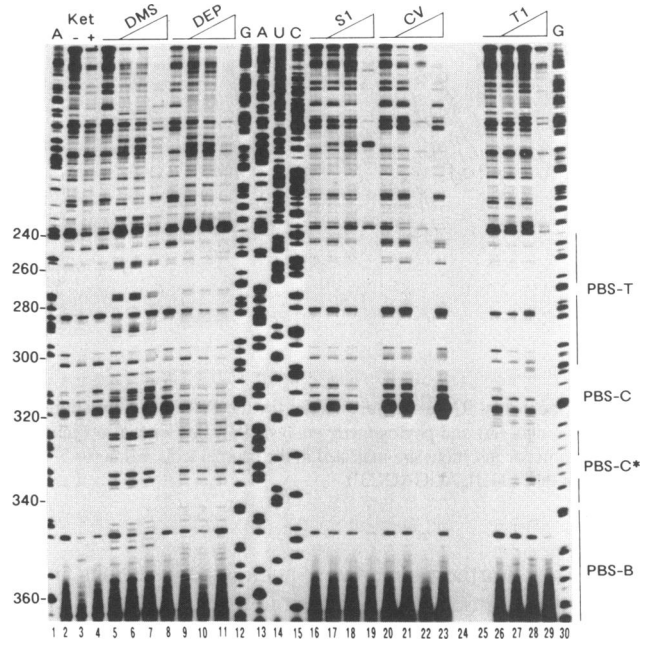


Figure 3. Nuclease digestion and chemical modification of TAR-2 RNA under native conditions. See legend to Figure 2 for details. The position of the different leader regions is schematically indicated on the right. The small hairpin in the central PBS domain is indicated by PBS-C*.

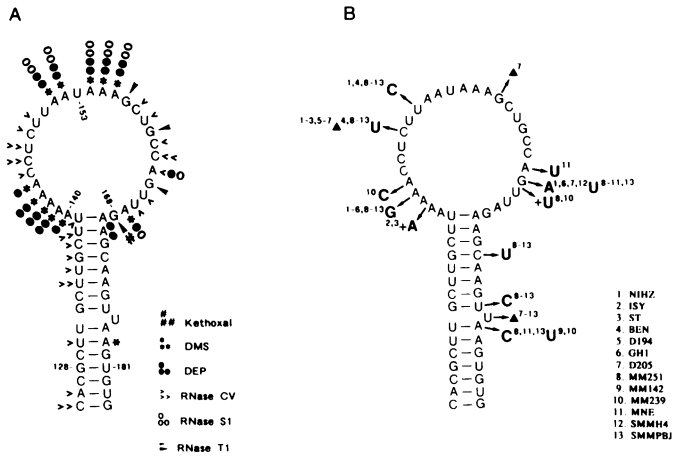


Figure 4. Secondary RNA structure model for the polyA region; summary of the digestion data **(A)** and phylogenetic analysis **(B)**. Sensitivities of the RNA to the various reagents are indicated by symbols (see insert). The sequence/structure of the HIV-2 ROD isolate is shown as prototype and compared to 13 different isolates (1−7=HIV-2; 8−11=SIV-MAC; 12,13=SIV-SMM). Nucleotide changes occurring in these isolates are indicated in bold with the number of the particular isolate in superscript. Deletions are shown as ▲, insertions are indicated by a + sign.

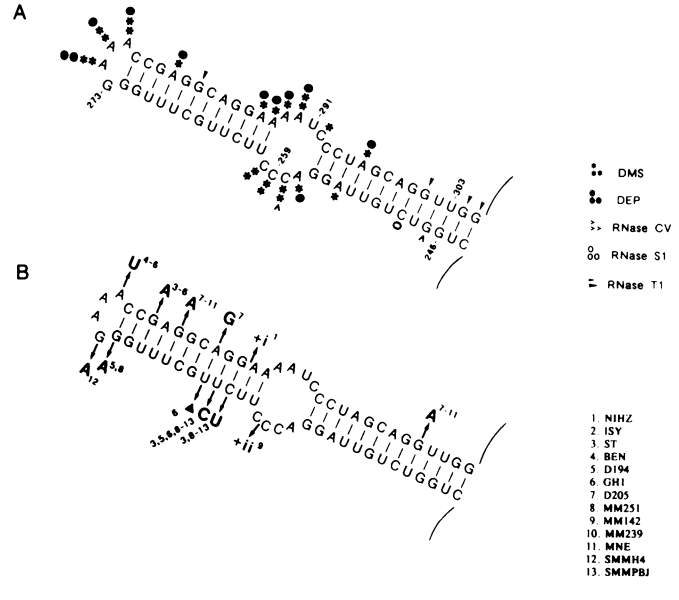


Figure 5. Secondary RNA structure model for the PBS-T region; summary of the digestion data **(A)** and phylogenetic analysis **(B)**. See legend to Figure 4 for details. Two major insertions are indicated in the internal loop region (i=5′CGG3′, ii=5′CUGGUCUGUUAGGACC3′).

base-pairing. Furthermore, two 1-nucleotide insertions and two 1-nucleotide deletions were present in the loop. Most sequence changes in the stem area do not disturb the basepairing scheme either. Interestingly, six isolates (8−13) mutate a G-U basepair into the more stable G-C pair, and at the same time change a G-C into the less stable G-U pair. Conservation of stem stability can also explain the alterations seen for many isolates in the bulge area. A deletion of the 1-nucleotide bulge (isolates 7−13) is

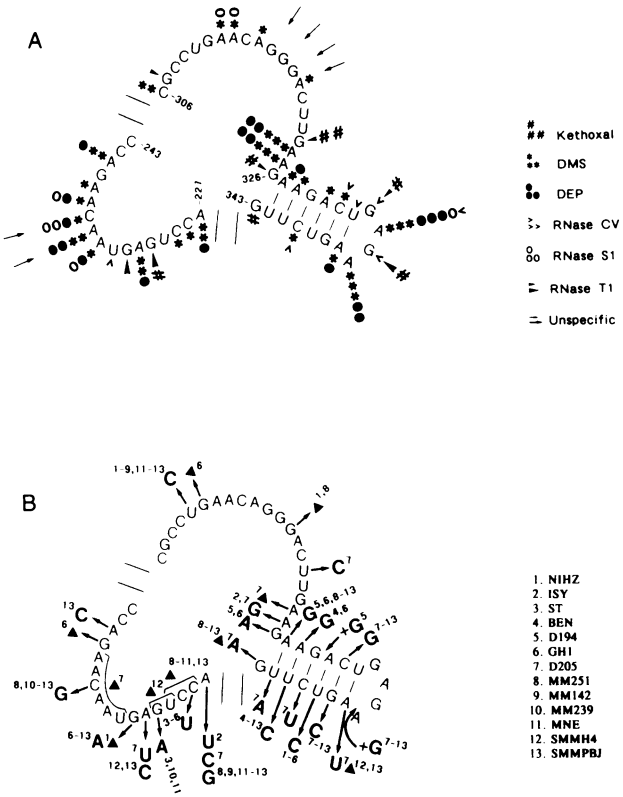


Figure 6. Secondary RNA structure model for the PBS-C region; summary of the digestion data **(A)** and phylogenetic analysis **(B)**. See legend to Figure 4 for details. Positions at which strong, non-specific reverse transcriptase stops were observed are indicated by arrows.

compensated in most viral genomes by the generation of a mismatch at the U-A basepair flanking the bulge (isolates 8−11,13). These combined results suggest that the overall stability of this 15-basepair hairpin is kept within narrow limits.

The structure of the PBS region (position +197 to +379)

A combination of the three forementioned approaches (structure probing, computer analysis and comparative sequence analysis) was needed to derive the structure for the PBS region shown in Figure 1. We will separately deal with the different domains; the top part consisting of an extended stem-loop structure (PBS-T, Figure 5), the relatively unstructured central domain consisting of a large internal loop and a small stem-loop element (PBS-C, Figure 6), and the bottom part in which leader sequences around position +210 are annealed to sequences around +360 (PBS-B, Figure 8).

The PBS-T region

The structure shown was predicted by the algorithm for RNA folding and is strongly supported by the experimental structure analysis (Figure 2 and 3, results are summarized in Figure 5A). Most informative are the DMS and DEP treatments that show highly susceptible nucleotides in the loop (position +274/+276) and internal loop (+257/+260 and +287/+292) domains. Some RNase T1 reactivity was observed on the 3′-side of the lower stem region (positions +301, +304 and +305). Interestingly, this region contains four consecutive G-U basepairs that might destabilize the helix (see discussion).

For a rigorous phylogenetic analysis, there should be dual base changes such that the alteration of the primary sequence at two residues in a stem will be compensatory to retain structure (e.g. G-C to A-U). Among the thirteen HIV-SIV isolates that were analyzed, no such dual base changes were observed for the PBS-T region (Figure 5B). However, many of the substitutions do conserve basepairing (e.g. U-G to U-A at position +279 and +281). For seven isolates (3, 8–13), a stem-destabilizing change at position +263 (C-G to U-G) was compensated for by a structure-stabilizing mutation in the next basepair (U-G to C-G, position +264). Major differences do occur in the sequences of isolate 9 (a 16-nucleotide sequence duplication at position +259) and isolate 1 (a 3-nucleotide insertion at position +286). Because of their location in or close to the internal loop area, these changes are not expected to disrupt the overall stem-structure.

The PBS-C region

The central domain of the PBS structure (PBS-C, Figure 6) consists of a 6-basepair stem-loop element (position +327/+343) and a relatively large internal loop (17 nucleotides at the 5′ side [position +227/+243] and 21 nucleotides at the 3′ side [position +306/+326]). The 5′ side of the internal loop is characterized by high susceptibility to all single-strand specific reagents (Figure 6A) and a dramatic variability in sequence among the different HIV-SIV isolates (Figure 6B). In fact, all nucleotides except two are mutated or deleted in one or more viral isolates. This invariant dinucleotide is the CC sequence adjacent to the PBS-T stem (position +242/+243). Given these results, combined with the inability to pair the +227/+243 region to any other sequence in the HIV-2 leader (either by computer analysis or by hand), we believe this region to not be involved in hydrogen bonding.

The analysis of the 3′ side of the internal loop was different from that of the 5′ side in that less reactivity was observed (Figure 6A), and that only a few base substitutions-deletions were detected (Figure 6B). The absence of major nucleotide changes is easily explained by the important function of this region as a primary sequence element; at least 15 nucleotides (position +306/+320) constitute the PBS, which is complementary to the $tRNA^{Lys,3}$ primer. Analysis of the complete HIV-2 leader sequences did not reveal any possibility for basepairing of this region. In addition, the reactivity pattern of this region was independent from the actual 3′ end of the HIV-2 sequences (either +420 or +886). The combined results suggest that this region is not highly structured. In the absence of phylogenetic data, we have therefore drawn this region as single-stranded. At this moment, however, we cannot vigorously exclude local basepairing (e.g. CCU310-GGG318), interactions within the internal loop (e.g. CCCUG246-GGGAC314), or tertiary interactions that limit the accessibility of this region.

The structure of the small hairpin in the central domain was suggested by the digestion experiments (Figure 3, PBS-C* area, results are summarized in Figure 6A) that showed major reactivity in the tetranucleotide loop area (position +333/+337). However, the presence of this hairpin was initially corroborated neither by the computer analysis, nor by the comparative sequence analysis (Figure 6B). At first glance, the existence of the helical stem is supported by only one compensatory base change (C-G to G-C for isolates 7–13), while multiple mutations are expected to disrupt the stem at most other basepair positions. A closer look at the individual sequences, however, does reveal that all isolates can adopt a similar stem-loop structure through

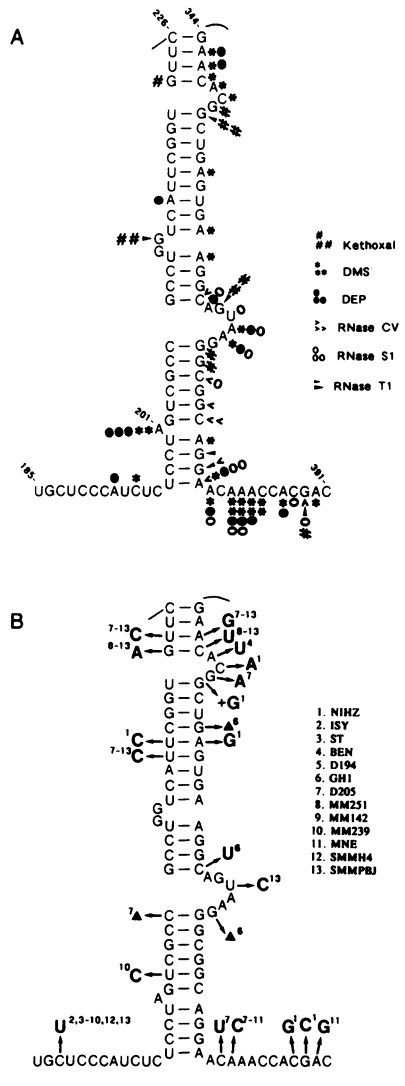


Figure 7. Alternative stem-loop structures in the region 3′ of the PBS. The dramatic nucleotide variability seen in this region for most HIV-2 isolates do affect the folding pattern as proposed for the prototype HIV-2 ROD (Figure 6). Alternative stem-loop elements can be formed through realignment of the mutated sequences. Bases that differ from the prototype ROD sequence are circled. The position of deletions (▲) and insertions (+) are indicated. Two alternative folding patterns are shown for both the D194 and the H4/PBJ isolates. Putative hydrogen bonding between A and C is indicated by ●. The possibility of non-Watson–Crick C-A basepairing has been described for tRNA (29), rRNA (30) and phage RNA (31).

realignment of the primary sequences (Figure 8). These results suggest that RNA structure is of primary importance for this region just downstream of the PBS sequence. Interestingly, for all alternative hairpin structures, the conserved pentanucleotide sequence UGAGA is present at the apex of the stem.

The PBS-B region

The structure of the PBS-B region was predicted by computer analysis and confirmed by the digestion experiments (Figure 3, summarized in Figure 7A) and the phylogenetic analysis (Figure 7B). All single-stranded areas were positively identified. For instance, the bulged-out A at position +201 was highly accessible to the chemicals DEP and DMS, but no reactivity of the adjacent nucleotides was scored (Figure 7A). Relatively little

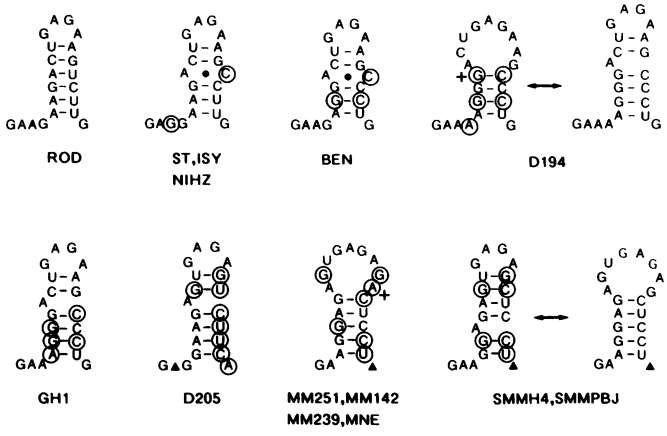


Figure 8. Secondary RNA structure model for the PBS-B region; summary of the digestion data (**A**) and phylogenetic analysis (**B**). See legend to Figure 4 for details.

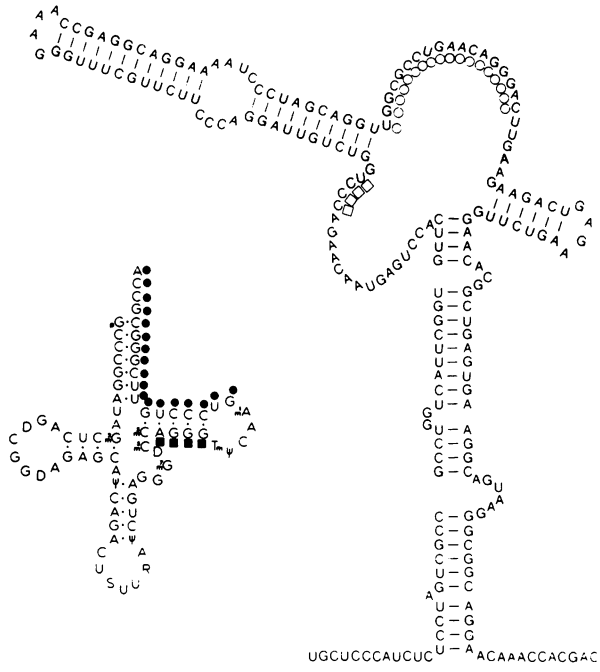
reactivity was observed for the region that connects the polyA and PBS structures (position +185/+196). A potential tertiary interaction of these sequences with upstream sequences in the TAR element may be responsible for this shielding effect (see Figure 1).

Three base changes on one side of the PBS-B stem are compensated by base substitutions in the opposite strand (Figure 7B: the 3rd, 4th and 9th basepair from the top). Thus, the basepairing is conserved even though the sequence is not. Furthermore, the overall stability of the upper part of the stem seemed to be conserved among isolates; the U-A to C-G change for isolates 7−13 was accompanied by a G-C to A-U change at the next basepair.

DISCUSSION

Thusfar, limited experimental data are available on RNA secondary structure of retroviral leader sequences (5,20,22−24,32,33). Here, we present a structural model for HIV-2 RNA leader based on the susceptibility of various regions of the molecule to cleavage by nucleases and on phylogenetic comparison of sequences from other HIV-2 viruses (Figure 1). The RNA chain folds into a short polyA stem-loop structure and a much larger, irregular PBS structure. The structure of the first 400 nucleotides of the HIV-2 leader consists of many short helix segments, which are interrupted by a variety of structural irregularities (bulges, internal loops etc.). In general, the stability of the individual helices is maintained relatively constant even though multiple base substitutions do occur among viral isolates. The tendency not to produce perfect stems may reflect the biological role of the RNA molecule as template in reverse transcription and translation. Progression of the RT enzyme and ribosomes, respectively, is best served by the presence of weak helices.

The existence of most stem regions is supported by compensatory base substitutions. The phylogenetic data can be summarized as follows. For basepaired nucleotides, a G to A change is expected to occur more frequently in the context of G-U pairs (G-U to A-U, conservation of basepairing) than G-C pairs (G-C to A C, loss of pairing). Indeed, out of 5 G to A changes in the HIV-2 helixes, 3 were in the context of G-U pairs




```
Standard tRNA-PBS interaction (●○)

Loss of bp in tRNA:              -12
Loss of bp in PBS-RNA:           - 3
Gain of bp in tRNA-PBS complex:  +18
                                 ----
                                 + 3

Extended tRNA-PBS interaction (●○ +■□)

Loss of bp in tRNA:              -12
Loss of bp in PBS-RNA:           - 3
Gain of bp in tRNA-PBS complex:  +22
                                 ----
                                 + 7
```

Figure 9. Secondary RNA structures of the $tRNA^{Lys}$ primer and the HIV-2 PBS region. Compared to Figure 1, part of the PBS-T stem is melted to expose the complete PBS sequence in order to allow an 18-basepair interaction with the 3′ end of the tRNA. The initial contact sites are marked by circles (tRNA; filled-in circles, HIV-2 PBS; open circles). The number of basepairs lost and formed during this 'standard' interaction is indicated. Subsequent stabilization of this complex is possible through an additional 8 basepairs between the regions marked by squares (tRNA; filled-in squares, HIV-2 PBS; open squares). The total number of basepairs lost and gained in this 'extended' annealing reaction is shown.

and only 1 G-C pair was altered. The 5th mutation was a co-variation (G-C to A-U) that conserved basepairing. This analysis becomes even more striking when one takes into account the total number of G-U and G-C pairs (14 and 33, respectively). A similar analysis can be performed for U to C changes, that are expected to predominate in U-G over U-A pairs. Out of a total of 14 U-G pairs, 4 were found to change into C-G. In contrast, no individual U to C change was observed for the 26 U-C basepairs. Additionally, 4 co-variations (U-C to C-G) were scored.

One striking feature of the structure model is the predominant single-strand character of the PBS sequences. Only 3 basepairs of the PBS-T stem need to open up in order to expose the complete 18-nucleotide PBS (Figure 9, the open circles). Remarkably, this region of the PBS-T stem contains four

consecutive G-U basepairs and some susceptibility to single-strand reagents was detected (Figure 5A). Thus, melting of this part of the PBS-T stem, in order to facilitate binding of the tRNA 3′-terminal CCA, is not expected to be a major hurdle. In fact, since all tRNA species share the same 3′-terminal CCA sequence, the initial inaccessibility of the GGU complement might contribute to the process of tRNA selection.

Interaction of the primer $tRNA^{Lys,3}$ with the HIV-2 PBS sequence of the retroviral genome involves 18 basepairs. For this intermolecular basepairing to occur, however, the tRNA sequences should not be involved in stable intramolecular basepairing. This implies that part of the cloverleaf structure of tRNA has to be unfolded (Figure 9). A total of 12 basepairs in the tRNA molecule must be sacrificed to allow a complete 18-basepair interaction with the HIV-2 target sequences (Figure 9, the circles in tRNA and PBS molecules). In addition, some tertiary interactions within the DHU loop (T_m and $_mA$) and between nucleotides in the DHU and TΨC loops (ΨC and GG) are likely to be lost at this point. It is possible that binding of the RT enzyme, which was reported to recognize the anticodon domain of $tRNA^{Lys,3}$ (16), does induce a conformational change in the tRNA molecule that enables the acceptor and TΨC stems to open up. The RT protein may also actively promote helix unwinding, as was described for the avian retroviral enzyme on RNA−DNA and DNA−DNA duplexes, but not on RNA−RNA duplexes (34). Alternatively, a role for the Gag nucleocapsid protein in the unwinding of the tRNA structure has been reported (19).

Based on sequence analysis, it has been suggested that other regions of the tRNA molecule can form additional contacts with the retroviral genome (reviewed in 2). Haseltine *et al.* (35,36) reported that 7 bases of the TΨC loop of the $tRNA^{Trp}$ primer could anneal to upstream sequences of the Rous sarcoma virus RNA. Consistent with this idea, Cordell *et al.* (37) showed that a 3′-terminal, 18-nucleotide tRNA fragment lacking the TΨC loop could not efficiently prime the reverse transcription reaction. Extension of this fragment with an additional 9 nucleotides from the TΨC loop did produce an active primer species. Recently, elegant work by Aiyar *et al.* (38) provided experimental evidence for the importance of this putative second interaction site. It was demonstrated that the additional interaction is necessary for efficient initiation of reverse transcription *in vitro* as well as in permeabilized virions. Furthermore, their experiments showed that besides primary sequences also certain structure elements are important for PBS function (20,21,38).

In our structure model of HIV-2 RNA, 8 additional basepairs can be formed between the TΨC loop of $tRNA^{Lys,3}$ (Figure 9, filled-in squares) and nucleotides on the 5′ side of the internal PBS loop (open squares). Sequence analysis of this second binding site in a number of HIV-2 and SIV isolates (Figure 5B,6B) indicated that this sequence motif is highly conserved, which is consistent with its proposed role in tRNA binding. Interestingly, the extended tRNA−PBS interaction is possible without the opening up of additional basepairs in the tRNA molecule and only one G-U pair of the PBS-T stem needs to be melted. It is the initial interaction between tRNA and PBS through the 'standard' basepairs (circles in Figure 9) that forces most nucleotides needed for the 'extended' interaction (squares in Figure 9) into the single-stranded form. These sequences (in the TΨC stem of the tRNA molecule and the 5′-side of the HIV-2 PBS-T stem) can subsequently anneal to further stabilize the complex. In this view, interaction between the retroviral genome and its tRNA primer occurs in two steps and the structure of the PBS region is suggested to play an important role in the process of reverse transcription. The secondary structure of HIV-2 RNA brings two regions of the genome into juxtaposition (the 'standard' PBS [circles, position 310] and the 'extended' PBS [squares, position 240]) that together facilitate stable binding of the tRNA primer. Since similar structures and interactions can be found for several retroviral RNAs (38), it is likely that these viruses utilize a common mechanism to initiate reverse transcription.

It should be stressed that we do not believe genomic RNA to be the only viral component needed for tRNA selection and binding. For avian viruses, a specific interaction between the RT enzyme and the tRNA primer was reported (11,12). This idea is supported by the apparent absence of primer from Rous sarcoma virus particles that lack polymerase (13,39) and the inclusion of primer in murine leukemia virus particles that lack viral RNA (40). However, the situation appears to be different in the murine retroviruses where no specific RT−tRNA binding was found (15). This was most conclusively shown by the genetic data of Colicelli and Goff (9). They describe a revertant of the Moloney murine leukemia virus with a PBS no longer specific for the usual $tRNA^{Pro}$, but for $tRNA^{Glu}$. This result suggests that the murine RT protein can utilize different tRNAs and that tRNA selection may be governed by the PBS sequences. Interestingly, the reversion-mutation was not restricted to the 18-nucleotide PBS region. Instead, a much larger genome segment around the PBS was acquired. Given the importance of the secondary, upstream tRNA binding site and the observation that both PBS sites are part of a much larger, specific RNA structure, it seems likely that the $tRNA^{Glu}$-PBS can function only in the appropriate sequence context. For instance, in order for the HIV-2 PBS sequence to fold into the correct secondary structure, a large segment of the sequences flanking the PBS need to be present (Figure 1, position +200 to +380). Further studies on the minimal requirements of the PBS region for primer binding (20,41,42) may reveal more details of this RNA−RNA interaction.

ACKNOWLEDGEMENTS

We thank Jeroen van Wamel and Maarten Jebbink for oligonucleotide synthesis, Wim van Est for photography work and Ms J.H.van Horssen-Medema for excellent artwork. We thank Rob Benne and Charles Boucher for critically reading of the manuscript and members of the Berkhout laboratory for helpful discussions. This work was sponsored by the Netherlands Organization for Scientific Research (NWO).

REFERENCES

1. Taylor, J.M. (1977) *Biochim. Biophys. Acta* **473**, 57−71.
2. Weiss, R., N. Teich, H. Varmus and J. Coffin (ed.). 1985. *Molecular biology of RNA tumor viruses*, 2nd ed. Cold Spring Laboratory, Cold Spring Harbor, N.Y.
3. Panganiban, A.T. (1990) *Seminars in Virology* **1**, 187−194.
4. Jeang, K-T., Y. Chang, B. Berkhout, M-L. Hammarskjöld and D. Rekosh (1991) *AIDS* **5** (suppl 2) 3−14.
5. Muesing, M.A., D.H. Smith and D.J. Capon (1987) *Cell* **48**, 691−701.
6. Feng, S. and E.C. Holland (1988) *Nature* **334**, 165−167.
7. Berkhout, B. and K-T. Jeang (1989) *J. Virol.* **63**, 5501−5504.
8. Berkhout, B., R.H. Silverman and K-T. Jeang (1989) *Cell* **59**, 273−282.
9. Colicelli, J. and S.P. Goff (1986) *J. Virol.* **57**, 37−45.

10. Myers, G., J.A. Berzofsky, A. Rabson, T.F. Smith and F. Wong-Staal (1990) *In Human Retroviruses and AIDS* (Los Alamos Natl. Lab., Los Alamos, NM).
11. Haseltine, W.A., A. Panet, D. Smoler, D. Baltimore, G. Petus, F. Harada and J.E. Dahlberg (1977) *Biochemistry* **16**, 3625–3632.
12. Araya, A., L. Sarih and S. Litvak (1979) *Nucl. Acids Res.* **6**, 3831–3843.
13. Peters, G.G. and J. Hu (1980) *J. Virol.* **36**, 692–700.
14. Litvak, S. and A. Araya (1982) *Trends Biochem. Sci.* **7**, 361–364.
15. Panet, A. and Berliner, H. (1978) *J. Virol.* **26**, 692–700.
16. Barat, C., V. Lullien, O. Schatz, G. Keith, M.T. Nugeyre, F. Gruninger-Leitch, F. Barre-Sinoussi, S.F.J. LeGrice and J.L. Darlix (1989) *EMBO J.* **8**, 3279–3285.
17. Sallafranque-Andreola, M.L., D. Robert, P.J. Barr, M. Fournier, S. Litvak, L. Sarih-Cottin and L. Tarrago-Litval (1989) *Eur. J. Biochem.* **184**, 367–374.
18. Sobol, R.W., R.J. Suhadolnik, A. Kumar, B.J. Lee, D.L. Hatfield and S.H. Wilson (1991) *Biochemistry* **30**, 10623–10631.
19. Prats, A.C., L. Sarih, C. Gabus, S. Litvak, G. Keith and J.L. Darlix (1988) *EMBO J.* **7**, 1777–1783.
20. Cobrinik, D., L. Soskey and J. Leis (1988) *J. Virol.* **62**, 3622–3630.
21. Cobrinik, D., A. Aiyar, Z. Ge, M. Katzman, H. Huang and J. Leis (1991) *J. Virol.* **65**, 3864–3872.
22. Darlix, J.L., M. Zuker and P.F. Spahr (1982) *Nucl. Acids Res.* **10**, 5183–5196.
23. Hackett, P.B., M.W. Dalton, D.P. Johnson and R.B. Petersen (1991) *Nucl. Acids Res.* **19**, 6929–6934.
24. Berkhout, B (1992) *Nucl. Acids Res.* **20**, 27–31.
25. Gurerick, V.V., I.D. Pokrovskaya, T.A. Obukhova and S.A. Zozulya (1991) *Anal. Biochem.* **195**, 207–213.
26. Zuker, M. and P. Stiegler (1981) *Nucl. Acids Res.* 9, 133–148.
27. Abrahams, J.P., M. van den Berg, E. van Batenburg and C. Pleij (1990) *Nucl. Acids Res.* **18**, 3035–3044.
28. Berkhout, B., A. Gatignol, J. Silver and K-T. Jeang (1990) *Nucl. Acids Res.* **18**, 1839–1846.
29. de Bruijn, H.M.L. and A. Klug (1983) *EMBO J.* **2**, 1309–1321.
30. Gutell, R.R., B. Weiser, C. Woese and H.F. Noller (1985) *Progr. Nucl. Acids Res. Biol.* **32**, 155–216.
31. Skripkin, E.A., M.R. Adhin, M.H. de Smit and J. van Duin (1990) *J. Mol. Biol.* **211**, 447–463.
32. Alford, R.L., S. Honda, C.B. Lawrence and J.W. Belmont (1991) *Virology* **183**, 611–619.
33. Harrison, G.P. and A.M.L. Lever (1992) *J. Virol.* **66**, 4144–4153.
34. Collett, M.C., J.P. Leis, M.S. Smith and A.J. Faras (1978) *J. Virol.* **26**, 498–509.
35. Haseltine, W.A., D.G. Kleid, A. Panet, E. Rothenberg and D. Baltimore (1976) *J. Mol. Biol.* **106**, 109–131.
36. Haseltine, W.A., A.M. Maxam and W. Gilbert (1977) *Proc. Natl. Acad. Sci. USA* **74**, 989–993.
37. Cordell, B., R. Swanstrom, H.M. Goodman and J.M. Bishop (1979) *J. Biol. Chem.* **254**, 1866–1874.
38. Aiyar, A., D. Cobrinik, Z. Ge, H-J. Kung and J. Leis (1992) *J. Virol.* **66**, 2464–2472.
39. Sawyer, R.C. and H. Hanafusa (1979) *J. Virol.* **29**, 863–871.
40. Levin, J.G. and J.G. Seidman (1979) *J. Virol.* **29**, 328–335.
41. Nagashunmugam, T., A. Velpaldi, C.S. Goldsmith, S.R. Zaki, V.S. Kalyanaraman and A. Srinivasan (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4114–4118.
42. Rhim, H., J. Park and C.D. Morrow (1991) *J. Virol.* **65**, 4555–4564.