Published online 14 July 2008 *Nucleic Acids Research, 2008, Vol. 36, No. 14* ***4653–4666***
doi:10.1093/nar/gkn447

RNA secondary structure of the feline immunodeficiency virus 5′UTR and Gag coding region

Laurie James and Bruno Sargueil*

CNRS UPR2167 – Centre de Génétique Moléculaire, Avenue de la Terrasse, 91190 Gif sur yvette, France

Received May 14, 2008; Revised June 26, 2008; Accepted June 27, 2008

ABSTRACT

The 5′ untranslated region (5′UTR) of lentiviral genomic RNA is highly structured, and is the site of multiple RNA–RNA and RNA–protein interactions throughout the viral life cycle. The 5′UTR plays a critical role during transcription, translational regulation, genome dimerization, reverse transcription priming and encapsidation. The 5′UTR structures of human lentiviruses have been extensively studied, yet the respective role and conformation of each domain is still controversial. To gain insight into the structure-function relationship of lentiviral 5′UTRs, we modelled the RNA structure of the feline immunodeficiency virus (FIV), a virus that is evolutionarily distant from the primate viruses. Through combined chemical and enzymatic structure probing and a thorough phylogenetic study, we establish a model for the secondary structure of the 5′UTR and Gag coding region. This work highlights properties common to all lentiviruses, like the primer binding site structure and the presence of a stable stem-loop at the 5′ extremity. We find that FIV has also evolved specific features, including a long stem loop overlapping the end of the 5′UTR and the beginning of the coding region. In addition, we observed footprints of Gag protein on each side of the initiation codon, this sheds light on the role of the sequences required for encapsidation.

INTRODUCTION

Feline immunodeficiency virus (FIV) is a lentivirus that infects the *felidae*, and is responsible for an immunodeficiency like virus in cats. Despite important differences in their genetic organisation FIV and HIV are clearly related viruses (1–3). Because of the low cost and the easy manageability of its host FIV has been used as a model to study lentivirus physiology (4) and to develop vaccine against AIDS (5,6). FIV encodes for Gag, Pol, Env and two accessory proteins: Vif, and a protein of unclear function called Orf-A (7–11), but it does not encode for any homologues of Tat, Rev, Vpr, Vpu and Nef (1–3). Despite thorough deletion studies defining the important determinants for viral packaging (12–17), very little is known about FIV 5′ untranslated region (5′UTR) RNA structure. The lentiviral genomic RNA 5′UTR, which is crucial for viral transcription (18), genomic RNA dimerization (19,20) and encapsidation (19). It is the site of many RNA–protein and RNA–RNA interactions throughout the viral life cycle. The viral RNA secondary structure plays a critical role in these interactions, and is regulated by the existence of potential alternative structures (21–24). We were interested in modelling the secondary structure of the FIV 5′UTR and the *gag* coding region, in order to compare it with those of other lentiviruses, and to gain insight into the structure–function relationship of the leader RNA and its interaction with the coding sequence.

In human immunodeficiency virus types 1 and 2 (HIV-1 and HIV-2), several stem loops have been clearly identified and found to fulfil at least one defined role. At the very 5′ terminus, the TAR structure stimulates viral genome transcription through the binding of the Tat protein (18). It has also been shown to be targeted by several other viral and cellular proteins, the functions of which are still not clearly identified (25–27). Another important secondary structure in HIV is the poly(A) stem loop (poly(A) SL), which occludes a poly(A) signal and thus prevents the premature termination of the genomic transcription (28). A few nucleotides downstream of the poly(A) SL is the primer binding site structure (PBS), which embeds a sequence complementary to the $tRNA^{lys3}$. This tRNA primes the reverse transcription of the genome prior to its integration into the host genome (29). Depending on the genome, a variable number of small stem loops are

*To whom correspondence should be addressed. Tel: +00 33 1 53 73 15 68; Fax: +00 33 1 53 73 99 25; Email: bruno.sargueil@univ-paris5.fr
Present address:
Laurie James and Bruno Sargueil, CNRS UMR8015, Laboratoire de cristallographie et RMN biologique, Faculté de Pharmacie 4 avenue de l'observatoire 75270, Cedex 06, Paris.

found 3′ of the PBS; all are bound by the nucleocapsid domain (NC) of the Gag polyprotein (30). One of them, termed 'DIS' for dimerization initiation site, contains a self-complementary loop sequence. Genomic RNA dimerization is suspected to be initiated by the formation of a kissing-loop complex in this loop (20), although recent studies suggest that the dimerization process might be more complex *in vivo* (31,32). Another small stem loop contains the major Splicing Donor consensus sequence (SD); this structure could regulate splicing efficiency by modulating the accessibility of the splicing site to the splicing machinery (33). In addition, two elements of tertiary structure have been identified: the first long distance interaction links U5 sequences to the initiation codon, the second associates the poly(A) loop with the coding region (22,34,35). It has been suggested that the former interaction is the switch for a structural rearrangement that regulates dimerization and encapsidation (22,23).

Gag polyprotein is known to bind the genomic RNA 5′UTR, and potentially the *gag* coding region (36). This interaction is required for genomic RNA packaging, but the exact binding site(s) remain unknown, potentially because they involve an as yet unidentified secondary or tertiary structure. We used a footprinting assay to better understand Gag interaction with the 5′UTR and the 500 first nucleotides of the coding sequence.

In this work, we show that FIV, HIV-1 and HIV-2 5′UTR structures share common features, although FIV exhibits some clear differences. Our results highlight the presence and conservation of a stable stem loop at the 5′ terminus of all lentiviruses, despite the absence of a Tat homologue in FIV. We hypothesize that this 5′ terminal stem-loop, SL1, is an important determinant for FIV genomic RNA dimerization. For several regions we propose several models compatible with our probing results. Notably, the region embedding the major splice site donor could adopt two different secondary structures that could be involved in splicing regulation. Finally, Gag footprints map within the 5′UTR and the Gag coding region itself, which is in agreement with previous results on FIV encapsidation, and suggests a model for preferential encapsidation of genomic over spliced RNA.

MATERIALS AND METHODS

Oligonucleotides sequences

The following oligonucleotides have been used in this study: FIV1: 5′ TAATACGACTCCTATAGATGG GG AATGGACAGGGGCGAGATTGG3′, FIV2: 5′ACTTC CTCACCTCCTAGTCCTTCTC3 ′, FIV3: 5′CCT ACTC CTACAGCAACATTA3′, FIV4: 5′CTTTCTTGTAAAT CGCAAAT3′, FIV5: 5′TGTCTAATCCCA TT TGAG AA3′, FIV6: 5′TAATACGACTCACTATAGGAGTCT CTTTGTTGAGGACT3 ′, FIV7: 5′ATCTCTCC TACC TTGTTGAC3′, FIV9: 5′TAGCTTCACTTCCTCAATC A3′, FIV10: 5′CACAGATACTCGACAGGG TT3′, FIV 11: 5′CCATCTGAAATTCCCTTCTCCA3′.

FIV6/FIV3, FIV1/FIV2 and FIV6/FIV2 were used to PCR amplify the 5′UTR, the Coding region and the full length constructs respectively. FIV4, FIV5 and FIV11 were used to prime reverse transcription within the coding region, while FIV7, FIV9 and FIV10 prime in the coding region.

In vitro Dimerization of FIV RNA

In vitro dimerization assays were conducted essentially as described (37,38). Five picomoles of RNA were denatured in water (volume final = 10 µl) at 90°C for 2 min and quenched-cooled on ice for 2 min. After the addition of 4 µl of 5 × monomer buffer [250 mM Tris–HCl (pH 7.5), 200 mM KCl and 0.5 mM $MgCl_2$] or 5 × dimer buffer [250 mM Tris–HCl (pH 7.5), 1.5 M KCl and 25 mM $MgCl_2$], dimerization was allowed to proceed for 60 min at 37°C. The samples were then cooled on ice to stabilize dimers formed during incubation. For experiments recapitulating the probing conditions, after denaturation, 2 µl of (500 mM Hepes (pH 7.5), 1 M KCl and 100 mM $MgCl_2$) were added and the mix was allowed to slow cool on the bench for 30 min, then for Gag containing samples, 2 µl of 3 µM HIV-1_{IIIB} p55Gag, 0.5 µl of 1 mM $ZnCl_2$ and 0.5 µl of 200 mM DTT were added, samples were incubated for 60 min at 37°C, then treated 15 min with 0.05% SDS and 0.5 µg proteinase K. Samples were then loaded on a 0.8% agarose gel with 2 µl of glycerol loading dye [6×: 40% glycerol, 44 mM Tris–borate (pH 8.3), and 0.25% bromophenol blue]. Electrophoresis was carried out for 2 h at 4°C and 5 V/cm in 44 mM Tris–borate (pH 8.3) and 0.1 mM $MgCl_2$ (TBM) or at room temperature (24°C) in 44 mM Tris–borate (pH 8.3) and 1 mM EDTA (TBE). After electrophoresis, RNA was stained with ethidium bromide, and the gel was quantified with a LAS (Fuji).

Chemical and enzymatic probing

The secondary structure of the FIV 5′UTR was probed using DiMethyl Sulfate (DMS), *N*-Cyclohexyl-*N*′-[*N*-Methylmorpholino)-ethyl]-Carbodiimide-4-Toluolsul fonate (CMCT) and V1 RNAse (Ambion) as described (39,40). Ten picomoles of RNA were resuspended in 36 µl of 50 mM Hepes pH 7.5 (or 50 mM borate potassium pH 8 for CMCT), denatured for 2 min at 80°C. Then, 4 µl of 1 M KCl and 100 mM $MgCl_2$ [and 1 µl of tRNA (10 µg/µl) for RNaseV1] was added and the mixture was incubated at room temperature for 10 min. DMS (25 mM final), CMCT (25 mM final) or RNAse V1 (0.05 or 0.025 Unit) was added, and the mixture was incubated for 5 min (10 min for RNAse V1). The modification reaction was stopped on ice by addition of 10 µg of total yeast tRNA (DMS and CMCT), or 1 µl of SDS 1% and 1 µl of proteinase K 10 mg/ml (RNAse V1). The reaction was then immediately ethanol precipitated on dry ice in presence of 0.5 M ammonium acetate. RNA was then re-suspended in 0.5 M ammonium acetate, ethanol precipitated, washed with 70% ethanol and re-suspended in 8 µl of H_2O. Modifications were revealed by reverse transcription (MMLV RNAse H^-) using ^{32}P-labelled primer according to the manufacturer instructions (Promega, Madison, WI, USA). Reverse transcription products were resolved on a 6% denaturing PAGE, the resulting gel was scanned on a Storm (GE Healthcare). The relative proportion of each product was determined drawing profiles with ImageQuant.

Footprinting

For footprinting, RNAs were treated exactly as described above for probing, but they were incubated beforehand for 15 min at room temperature with 4 µl of 3 µM HIV-1_{IIIB} p55Gag (NIH AIDS research and reagent program) (reactions were 300 nM final), and 0.5 µl of $ZnCl_2$ 1 mM (10 µM final) and 0.5 µl of 200 mM DTT (2 mM final). The RNA protein mix was then probed with DMS and V1 nuclease as described above.

Filter-binding assay for Gag-RNA interactions

Filter-binding assays were performed in triplicate and consisted of a 20 µl reaction mixture. Labelled FIV RNA (5000 c.p.m., <5 nM) was heated in water for 5 min at 80°C. After addition of microlitre buffer 1 [30 mM of *N*-2-hydroxyethylpiperazine-*N*′-2ethanesulfonic acid (Hepes; pH 7.5), 100mM of KCl, 10 µM of $ZnCl_2$, 2 mM dithiotreitol (DTT), 5 µg of tRNA and 10 mM of $MgCl_2$ (final concentrations), the reaction mixture was slowly renatured at room temperature for 10 min The indicated amounts of HIV-1_{IIIB} p55Gag was added to the mixture. After 30 min at room temperature (24°C), reaction mixtures were placed on wet nitrocellulose disks (0.45 µM-pore-size HAWP type, Millipore), filtered through and rinsed with 0.3 ml of buffer 1. Radioactivity retained on filter was determined by Cerenkov counting. All data were corrected for background binding of probe to filters in the absence of protein (typically <3% of total radioactivity) and are expressed as a percentage of counts bound relative to the maximal binding achieved with a given RNA target within each experiment.

Native gel analysis

RNAs were analysed on a 6% polyacrylamide native gel containing 66 mM Hepes, 34 mM Tris and 5 mM $MgCl_2$. RNA were heat denatured for 2 min at 80°C, and slowly cooled to room temperature in 20 mM Tris pH 8, 0.5 mM $MgCl_2$ and 75 mM KCl for 10 min 10% sucrose was added to the sample before loading them on the gel which was run for 5 h at 8 W.

RNA preparation

All RNAs were transcribed with T7 polymerase from PCR products as described (40). After transcription, full-length products were purified on a 6% denaturing PAGE. They were then extracted from the gel using a passive diffusion method in a buffer containing 0.5 M NH_4 acetate, 0.1 mM EDTA and 0.1% SDS. Recovered RNA was desalted on a G50 column (GE Healthcare) and quantified by absorption at 260 nM (1 OD equivalent to 37 µg/ml).

RESULTS

Three different constructs were used to probe and model the secondary structure

To detect potential interactions between the 5′UTR and the coding region, we conducted all our probing and modelling experiments in parallel on three different *in vitro* transcribed RNA constructs: the 5′UTR (from nucleotide +1 to 412, where +1 is the transcription start), the *gag* coding region (from 412 to 920) and the full-length 5′UTR-gag construct (from +1 to 920). Our experiments were conducted on a molecular clone from a domestic cat virus (clone 34TF10, accession # NC_001482) (1).

Mg^{2+}-dependent dimerization of 5′UTR and 5′UTR-gag FIV RNAs

Lentivirus leaders contain the sequences which are responsible for initiating viral genome dimerization and have been shown to dimerize *in vitro* (41). Therefore, the conformational state of the RNAs were evaluated by monitoring their electrophoretic mobility under several non-denaturing conditions.

First, RNAs were run on a non-denaturing polyacrylamide gel in presence of Mg^{2+}. All three constructs migrate as a single tight band, indicating a lack of broadly divergent alternative structure and aggregation (data not shown). We then used agarose gel electrophoresis in native ('TBM conditions', see 'Materials and methods' section) or in semi denaturing conditions ('TBE conditions': no magnesium at room temperature see 'Materials and methods' section). In the case of HIV-1 and HIV-2, this method was able to detect two different species of dimers (38,42). The first, only stable in presence of magnesium, are suspected to be mediated by kissing hairpins and have been termed 'loose dimers' (38,42). On the other hand, RNA dimers that are stable in semi denaturing conditions are referred to as 'tight' dimers (38,42). Henceforth, we will use the terms 'loose' and 'tight' dimers to refer to these different complexes.

We monitored the dimerization state of the three constructs when treated and incubated in buffer conditions identical to those used in our probing assay. In addition, RNAs were also incubated in the additional presence of 300 nM of Gag protein, which could influence the proportion of dimer in solution. As a control, the three constructs were folded under conditions known to yield essentially monomer (0.1 mM Mg^{2+}) or dimer (5 mM Mg^{2+}) when used with HIV-1 or HIV-2 leader (37,38,42). As can be seen on Figure 1, the 5′UTR and the full-length constructs run as two bands under native TBM conditions while they run essentially as a single band corresponding to the monomer under semi denaturing conditions. The coding region runs essentially as a single band corresponding to the monomeric form under all conditions. This indicates that under our native probing conditions, 90% of the 5′UTR and 40% of the full length RNA exist as 'loose' dimers, but no stable 'tight' dimers are formed. The presence of the Gag protein does not seem to influence the dimerization equilibrium. The secondary structure of the 'loose' dimer is thought to be similar to the monomer structure, except for an intermolecular contact in the form of kissing complexes between one or several loops. Thus, we reason that the probing pattern of the 'loose' dimer should differ from the monomer's only by the absence of single-strand-specific hits in loops harbouring self-complementary sequences. The structure probing was therefore conducted in absence or presence of Mg^{2+}.

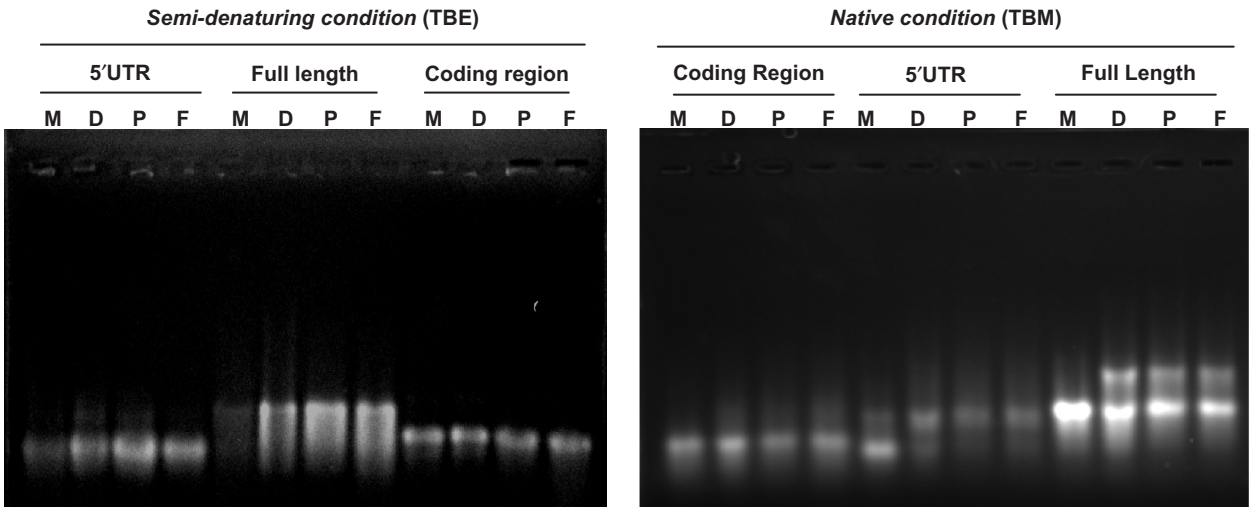

Figure 1. FIV 5′UTR forms loose dimers *in vitro*. Three RNA constructs, the 5′UTR, the coding region and the full-length, were renatured and incubated under different conditions: M (Monomer: 0.1 magnesium), D (Dimer: 5 mM magnesium), P (Probing: 10 mM magnesium) or F (Footprint: Probing + 300 nM p55Gag), and then run under native (TBM) or semi denaturing (TBE) conditions (see 'Materials and methods' section). For each construct, the fast migrating band represents monomer, while the dimer runs above.

However, caution should be taken, as for HIV-1 and HIV-2 the equilibrium between the monomer and dimer is regulated by an alternative structure that relies on a long distance interaction (21,22,43). Such a long distance interaction has not been detected in FIV (44), and although we cannot rule out the existence of a widely divergent secondary structure, we found only one model fully consistent with the probing and phylogenetic data. This conformation is reminiscent of the 'Branched structure with Multiple Hairpins' (BMH) modelled for HIV-1 and HIV-2 (22,43).

Probing strategy and modelling pathway

The secondary structure of the three constructs was probed using chemical and enzymatic reagents. CMCT and DMS were used to detect single stranded regions, while RNAse V1 reveals stacked or paired nucleotides. DMS specifically methylates unpaired adenines at N-1 and cytosines at N-3; CMCT modifies unpaired guanosines and uracils at N-1 and N-3, respectively. Those modifications cause reverse transcriptase to stop at the nucleotide immediately 3′ to the solvent accessible position. RNAse V1 cleavage leaves a 3′OH terminus; thus primer extension on those products stops at the position hit (45). The reverse transcription pattern of the modified RNA is compared to the profile obtained with a non-modified RNA. A typical example of our results is shown in Figure 2. Modifications were classified into two categories according to the relative intensity of the stop they induced by comparison with the control reaction. They were qualified as 'weak' when inducing a 2- to 3-fold increase in intensity of the RT stop, and as 'strong' for higher intensity (up to 9-fold has been observed). For DMS and CMCT, a weak modification reveals single stranded regions but can also be localized within poorly stable stems, or even at the terminus of a stable stem due to fraying. Nucleotides hit by the V1 nuclease can be part of a helix, or be stacked for other reasons. Moreover, one should keep in mind that not all nucleotides involved in base pairs are substrates of the V1 nuclease. For all those reasons, in a first round of secondary structure modelling, we took into account only the strong hits observed with DMS or CMCT. In this first step, those constraints were used as input constraints for the Mfold secondary structure free energy minimization software (46). The different models issued from Mfold were further evaluated in regards to the V1 and the weak DMS and CMCT hits observed, and then compared to phylogenetic data (a set of 72 different FIV sequences were aligned—see Supplementary Data). Each of the secondary structure models generated by Mfold were assessed individually in the light of the experimental and phylogenetic data. Validated pairings were then used as additional constraints to generate new models with the help of Mfold. This process was repeated until the model generated agreed with all the experimental and phylogenetic data. The probing data obtained with the isolated 5′UTR or the coding region were essentially similar to those observed with the full length construct but for one notable exception (detailed below). Therefore, the probing data were pooled and are presented on the secondary model established for the full-length construct (Figure 3). The data shown on this figure is the result of at least three independent experiments including different RNA preparations, and represents the compilation of all those experiments. The individual secondary structure elements, and the data that support these structures, are described below. In discussing the phylogenetic data, we will use the term 'conservative' changes for transition on one side of the helix that does not disturb the pairing (as U–G to C–G or A–U to G–U), while the term 'compensatory' is restricted to the observation of a double mutation preserving the pairing (as A–U to G–C).

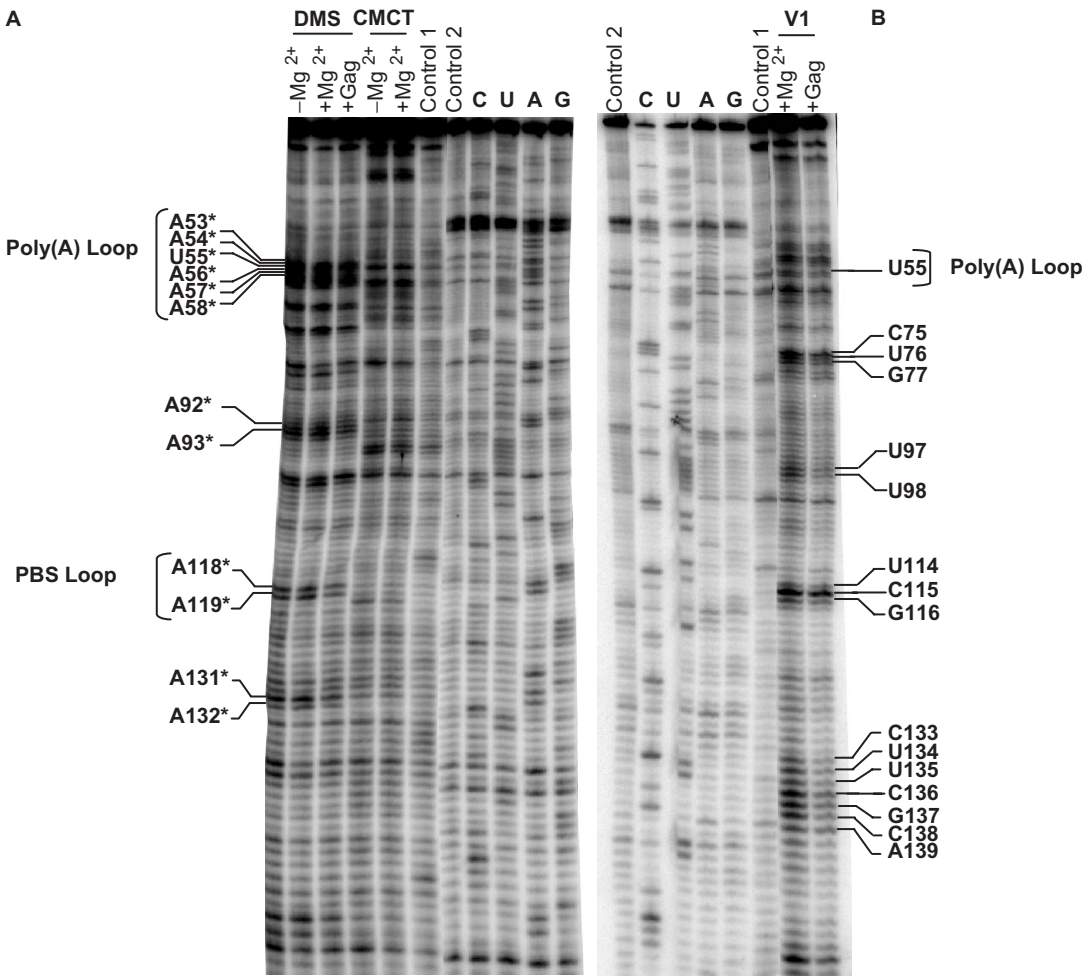

Figure 2. Chemical and enzymatic probing of the PBS and polyA region in the 5′UTR. (**A**) Full length RNA was renatured and treated with DMS and CMCT as described in 'Materials and methods' section. The probing was conducted in presence or in absence of magnesium. For control 1, RNA was treated exactly like the probed RNA in presence of magnesium, except that the chemical reagent was omitted (elongation of an unmodified RNA with the MMLV RNAse H^-). For DMS, Gag footprint on the RNA was established by conducting the probing in presence of 300 nM of protein (see 'Materials and methods' section). Reverse transcription products were separated on a 6% gel, along with a control and four sequencing lanes. Control 2 is a control for the sequencing reaction, it consist in a reverse transcription of an unmodified RNA with the AMV reverse transcriptase. Some interesting positions are indicated on the side of the autoradiogram. Note that chemical reagents induce a premature reverse transcriptase one nucleotide before the nucleotide hit. Therefore the chemically induced stops run one faster than the corresponding sequence and are therefore indicated with an asterisk. (**B**) Same that in A, except that RNA was incubated with RNAse V1. Only the condition with magnesium has been performed because divalent ions are required for the nuclease activity. The reverse transcriptase stop induced by RNAse cleavage runs at the same position that its corresponding sequence.

Stem-Loop 1

This 45-nt structure begins at the very first transcribed nucleotide. The probing observed is fully consistent with the modelled stem/internal bulge/stem loop and this structure is strictly conserved among the 44 sequences available (see Figures 3 and 4; Supplementary Data). The sequence is highly conserved, few conservative and compensatory changes are observed within the upper loop, and there are few variable nucleotides in the bulge (U_9 is a C in all but three sequences, U_{10} is a A or a G in two isolates, G_{11} is an A in one instance), while the apical loop is a UUUG in four instances, and a CUCG in one isolate. In 39 out of the 44 isolates, the sequence of the predicted capping loop is UUCG. This is another evidence for the existence of SL1, as UUCG is a sequence overrepresented in natural RNA loops (47) and forms unusually stable tetraloops (48). Furthermore, several conservative and compensatory changes (A_{18}–U_{27} to G_{18}–C_{27} and C_{19}–G_{26} to A_{19}–U_{26}) are observed in the apical stem. Interestingly, the 4 nt of the tetraloops are part of self-complementary sequence in all the isolates (YUYGAG where Y stands for a pyrimidine).

SL2, the poly(A) stem-loop

A canonical polyadenylation signal (AAUAAA) is present in all isolates, and a stem-loop can be formed with the adjacent nucleotides on each side (Figure 3). A stem loop containing the poly(A) signal is predicted in all 44 isolates available; nevertheless its sequence and position varies. In some instances, the poly(A) signal is not in the loop but partially buried in the stem, in others it appears to be only marginally stable (see for example isolate #D31939 in

Supplementary Data). The formation of a poly(A) stem loop is fully substantiated by the probing we have conducted on the 34FT10 isolate: Positions C_{52} to A_{58} are strongly hit by DMS and CMCT, and we observed strong and weak RNAse V1 cuts on the 5′ and the 3′ side of the stem respectively (Figure 3). Some nucleotides on the 3′ side of the stem are also susceptible to single-strand-specific probes, this probably reflects its poor stability (calculated $\Delta G = -0.6$ kcal/mol).

SL3, the PBS structure

This structure encompasses the PBS sequence that is complementary to the $tRNA^{lys3}$ which is used to prime reverse transcription. The 18-nt PBS sequence is strictly conserved among the FIV isolates and is also present in HIV-1, HIV-2 and SIV_{Mac}. The PBS sequence has been modelled as a predominately bulged loop between the two major helices SL3-A and SL3-B; part of the tRNA-binding site can be folded as a small hairpin which stems contains three G-C base pairs (SL3-C) (Figure 3). The formation of the two main stems (SL3-A and SL3-B) of this structure is fully substantiated by the probing data since most of the nucleotides that composed them are hit by the V1 nuclease (Figure 2). The only exceptions are nucleotides that are involved in potentially unstable base pairs located at the terminus of the helix or close to bulges. The 2 nt U_{166} and G_{167} have been modelled as a bulge, but are substrates of the V1 nuclease. This is not completely unexpected since single or double nucleotide bulges are often stacked within the helix, even if dynamically so. We have modelled two alternative structures of similar stability for the lower part

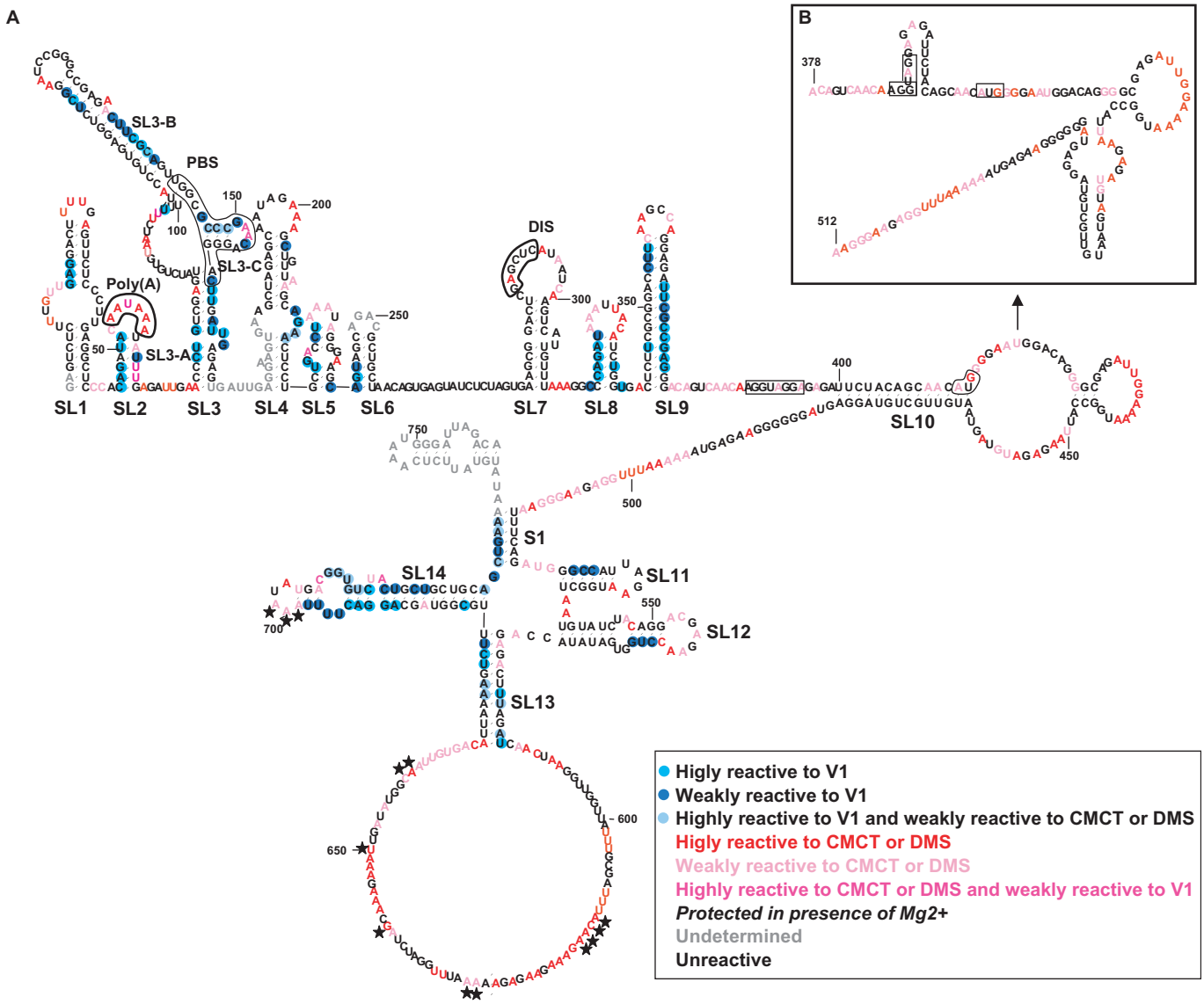

Figure 3. Secondary structure model of FIV 5′UTR and *gag* coding region. Colours reflect the accessibility of the nucleotides to the different probes as indicated in the boxed area. 'Weakly reactive' refers to reverse transcriptase stops enhanced 2- to 3-fold compared to the untreated RNA, 'Strongly reactive' position are enhanced over 4-fold. The star denotes position protected by at least 2-fold in presence of magnesium. (**A**) Model of the full-length RNA. The polyA signal, the primer binding site and the self complementary sequence in the DIS loop are boxed. (**B**) Alternative secondary structure model for the region surrounding the initiation codon. The splice site donor sequence and the initiation codon are boxed.

of the PBS structure and the junction region (Figure 5A). The three structures are in agreement with our probing data and our phylogenetic observations (on 16 sequences). The phylogenetic study also confirms the overall structure, since the apical stem (SL3-B) is conserved in all 70 sequences available, although the sequences are variable. Interestingly, in addition to the tRNA-binding site, there are two other conserved sequences: the apical loop from U_{120} to G_{125}, and the 9 nt preceding the primer binding site (5′CYUCGCAGU3′). In 53 isolates, the apical loop sequence is 5′U_{120}CCGGG$_{125}$3′, a self-complementary sequence, while the 18 remaining isolates harbour only a 4-nt self-complementary sequence, because C_{121} is missing.

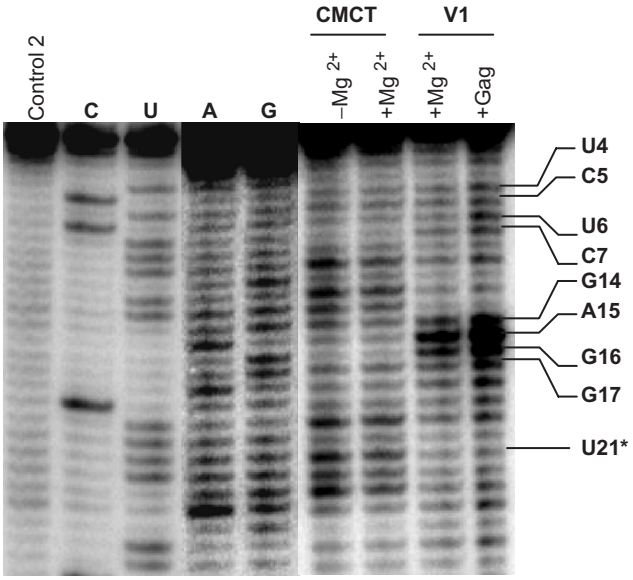

Figure 4. Chemical and enzymatic probing of the SL1 region—Gag footprint on SL1. For comments see Figure 2. Note that chemical reagents induce a premature reverse transcriptase one nucleotide before the nucleotide hit. Therefore the chemically induced stops run one nucleotide further than the corresponding sequence and are therefore indicated with an asterisk.

SL4 to SL9; six hairpin stem-loops of unknown function

The region 3′ to the PBS structure can be folded as a succession of six stem-loops. The existence of SL4 is supported by the DMS hits observed in the apical loop and on the bulged adenosine on the upper stem, and by one V1 cut in the stem. Interestingly, the 3 nt in the 3′ side of the central purine rich bulge (5′GAA3′/3′AGA5′) are also cut by V1. This could be because this bulge is structured, or simply because purines have a tendency to form stacking interactions even when not paired. This loop contains a potential tandem AG/GA sheared base pairs, which is a recurrent motif often involved in tertiary or protein contact (49–54), and more importantly is a known determinant for HIV-1 Gag binding (55–57).

The SL4 structure and sequence are conserved in 41 available sequences, and only very rare conservative changes are observed. In three isolates (AY600517, AF474246 and AY220074), an SL4 like structure in which the bulge is larger (7 nt on 5′ side and 8 nt on the 3′ side), and the apical loop shorter (4 nt) can be modelled.

The SL5 stem loop is entirely consistent with the probing signatures observed (Figure 3), the sequence of the stem is conserved in all the 41 available sequences, but the identity of the nucleotides in the apical loop is not conserved.

The modelling of SL6 relies mostly on the phylogenetic analysis. The sequence of this stem-loop is conserved in the 40 sequences we have aligned, with the exception of two conservative changes (C_{247}–U_{247} in one isolate, C_{256}–U_{256} in two isolates), and one mutation (A_{245}–U_{245}) that disrupt 1 bp in eight isolates. However, in those cases the stem can systematically be elongated, to the detriment

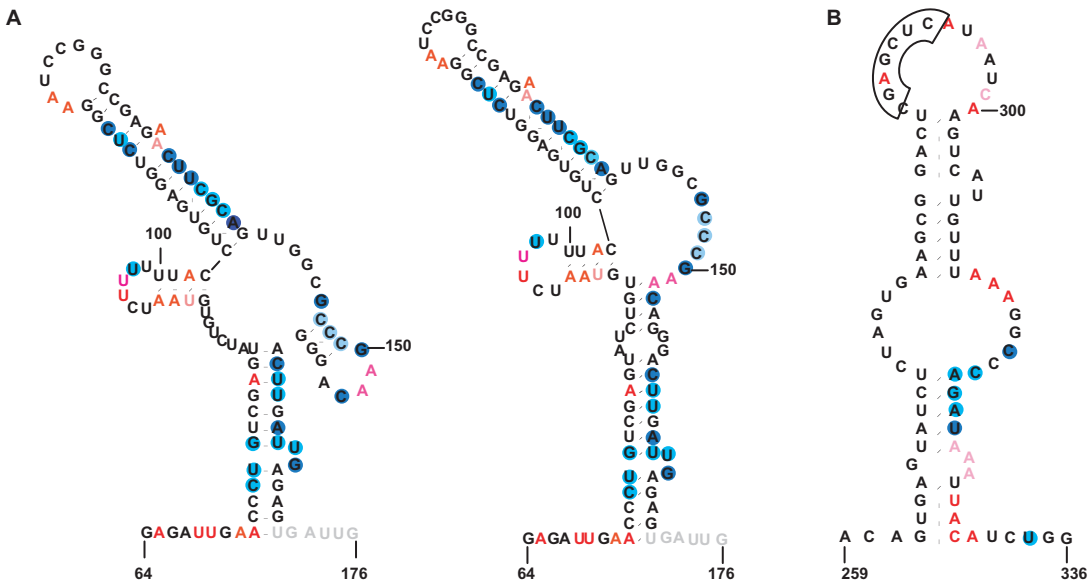

Figure 5. Alternative models compatible with the probing data. (**A**) Alternative models for the PBS. (**B**) Alternative model for the SL7/SL8 region. SL7 can be elongated at its base, this results in the disruption of SL8. Legend and colour code are the same as that in Figure 3.

of SL5, by 2 bp C_{240}–G_{259} and A_{241}–U_{258} at its base (see for example isolate #AY369381.1 in the Supplementary Data). The probing data confirm the phylogenetic analysis: we observed three V1 nuclease cuts on the 5′ of the helix, and no single-strand-specific hits in the stem (Figure 3).

SL7 has been modelled according to the DMS hits observed in the loop. This hairpin can be modelled in all 40 sequences gathered, although its sequence can be quite variable. Interestingly, the apical loop contains a self complementary sequence (5′GAGCUC3′) which is conserved in 30 out of 40 sequences. In three of them, no self-complementary sequence can be identified, in five others conservative or compensatory changes are observed (5′GGGCUC3′, 5′GAAUUC3′ or 5′GAGCUU3′) and finally, two potential base pairs are disrupted in two others (5′AAGCUC3′). In the majority of the isolates (33 out of 35 self-complementary sequences), the first 2 nt of this sequence are embedded in the stem. Interestingly, this stem loop can be elongated at its base, to the detriment of SL8, by pairing 5′G_{262}–U_{271}3′ to 3′C_{330}–A_{320}5′ (Figure 5B). Such an alternative structure can be modelled in HIV-2 and has been shown to regulate dimerization/encapsidation (42,58,59). However, the nucleotides at the 3′ base of this stem are strongly hit by DMS, this argues that this stem loop is not be stable *in vitro*.

SL8 is a small hairpin, the stem of which is cut in many instances by the double-strand-specific nuclease, while its apical loop is hit by both DMS and CMCT (Figure 3). The adenosine of the last U-A base pair is also hit by DMS, and this can be explained by the poor stability of this terminal pair. C_{316} and U_{336} on each side of the stem are cut by the V1 nuclease, those 2 nt are conserved, they could participate in a yet unidentified interaction. This hairpin can be modelled for all 40 sequences, although its sequence can be very divergent (compare, for instance, NC_001482.1 with AY600517.1 in the Supplementary Data).

It is then followed by a 16-nt-long stem in which we have observed numerous V1 nuclease cuts, while the nucleotides in the loop are hit by DMS. The sequence of this stem loop (SL9) is essentially conserved in the 41 isolates, sporadic conservative mutations are observed all along the stem, and in some isolates 1 bp is disrupted.

The last 14 nt of the 5′UTR are paired with coding sequences

The probing data between the 3′ end of SL9, to the AUG reported in Figure 3 was derived exclusively from experiments performed on the full-length construct, since this region is too close to the most 3′ terminal primer binding site of the 5′UTR construct. After a detailed phylogenetic analysis, we modelled a splicing donor stem-loop similar to what is observed in HIV-1 and HIV-2. A 6-nt-long stem embedding the intronic portion of a potential splicing donor site (AG/GUAGGA) could be modelled. This stem loop and its sequence are highly conserved among the 41 sequences we have aligned. Some changes are observed in the loop of nine isolates, and only one conservative change and two disruptive mutations can be observed in the stem of only one isolate for each of them. However, the probing observed on the full length construct was not completely consistent with the formation of this helix (Figure 3). In addition, we noticed that the probing signatures observed on the first 80 nt of the coding region depend dramatically on the construct used. With the full length construct, we observed single-strand-specific hit localised in three regions (from G_{414} to G_{420}; from A_{434} to A_{442} and from U_{450} to A_{460}), while with the 'coding region' construct, we mainly observed V1 cuts in between U_{443} and A_{459}. This discrepancy, which is the only exception, suggests that some interactions between the coding region and the 5′UTR might exist, and led us to consider an alternative folding for the end of the 5′UTR. Actually, we found that a 14 nt long stem linking the very end of the 5′UTR to the coding region would be in agreement with the probing data on the full-length construct. This long stem is followed by a large purine rich bulge, closed by a short hairpin, the loop of which is heavily hit by single-strand-specific probes. The sequence of the basal stem is very conserved in all FIV isolates, and we observed two conservative changes (A_{404}–G_{404}, and A_{477}–G_{477}), and two mutations (U_{400}–G_{400} and U_{403}–A_{403}) that disrupt 1 bp out of 14, each in only one isolate. The three G–C base pairs that close the bulge are conserved in 39 of the 41 sequences available. Three A–U base pairs are weakly hit by DMS, and this could reflect some breathing at the 3′ end of this helix which is A–U rich. Alternatively, this may simply reflect the fact that we are probing a mix of the two structures described above.

The *gag* coding region is structured

A global look at the probing data indicates that, except for two long single stranded regions (G_{478}–A_{512} and C_{586}–C_{668}), the coding region is structured. We have modelled it as four stem-loops (SL11–SL14) and a long range stem S1 that brings SL11 and SL14 in close proximity.

S1 was modelled after the V1 cuts observed on the 3′ side of the stem (Figure 3). Those two sequences appear to be highly conserved, except for C_{516} which is a U_{516} (conservative change in 23 sequences out of 40) and for U_{729}–C_{729} that disrupt 1 bp in two isolates. Nevertheless, one has to keep in mind that analysing the nucleotide conservation within a coding region is not necessarily fully relevant for the structure since the sequence is constrained by its ability to be translated into a protein.

SL11 and SL12 are fully consistent with the probing (Figures 3 and 7), SL11 is strictly conserved except for the first and the last nucleotides of the loop (see Supplementary Data). SL12 sequence is strictly conserved but for U_{540} or A_{543} the mutation of which disrupts 1 bp. It is interesting to note that those two changes are never observed concomitantly.

The SL13 stem is mostly a substrate for V1 cutting, although A_{575} and A_{577} are also weakly reactive to DMS, and this is not surprising since SL13 is mostly constituted of A–U and G–U base pairs. The sequence of this segment is well conserved, we observed few conservative and compensatory changes; and the disruption of a base pair on scarce occasions. Several stretches of nucleotides within the loop are protected from modification upon

Mg^{2+} addition (Figure 3). Interestingly, most of them are embedded in 4–6 nt self complementary sequences. Each of those sequences are independently conserved in 75% of the isolates, and conservative changes are observed in some cases (see Supplementary Data).

SL14 is fully substantiated by the probing data, three A–U base pairs are weakly hit by DMS, this could be expected for at least two of them which are located at both extremities of a 3 bp helix. Another peculiarity of the probing signature are the V1 hits in the internal bulge, but one can note that this bulge could be structured involving formation of at least two G–U base pairs. The phylogenetic analysis shows a good conservation of the sequences in the stem, but some variation is observed in the pairs adjacent to the internal bulge. Overall, SL14 can be formed in all isolates and some conservative and compensatory changes are observed. It is noteworthy that the sequence of the loop is very conserved and protected from DMS in presence of Mg^{2+}. The three adenosines protected lie within a palindromic sequence $_{695}$UUUAAA$_{701}$ which is conserved in 29 out of the 41 sequences available. In all isolates where this sequence is mutated, another self-complementary sequence is found $_{691}$GGAAUUCU$_{698}$ in which conservative changes are observed (see Supplementary Data). The protection observed upon Mg^{2+} addition, and the RNAse V1 cleavage of U_{696}–U_{698} could reflect intermolecular contacts.

Gag binds the 5′UTR, and the coding region

It has been shown that HIV-1 Gag is able to promote FIV packaging (60) and therefore could binds FIV genomic RNA.

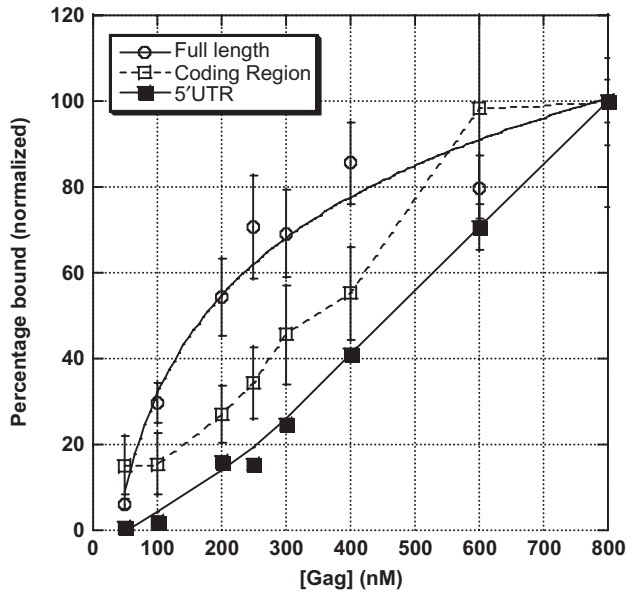

Figure 6. HIV-1 Gag binds to FIV leader and coding region. Filter binding assays of HIV-1 Gag polyprotein binding to FIV full length (hollow circles), FIV 5′UTR (filled squares) or the coding region (hollow squares). To facilitate comparison, results have been normalized to the value of the maximum that corresponds to a plateau except for the 5′UTR. The maximum bindings observed were 57% for the full length, 62% for the coding region and 40% for the 5′UTR. Error bars represent the SD from the mean from the data of three independent experiments.

We performed filter binding assays using HIV-1 Gag and radiolabelled RNAs corresponding to the three constructs: the 5′UTR, the coding region and the 'full-length' RNA. Gag binds to all three RNAs, albeit with different affinities (Figure 6). None of the binding curves fit a theoretical curve describing a bimolecular interaction. This curve may reflect multiple superimposed phenomena: multiple binding sites are likely to be present, and Gag binding may be cooperative in some cases and not in others. For those reasons we are reluctant to derive an observed K_D from those experiments. Nevertheless, the value of half-binding (150 nM for full length, 350 nM for the coding region and 450 nM for the 5′UTR) indicate that the affinity for any FIV derived RNA is somewhat weaker than what is observed with HIV-1 leader (K_D = 50 nM). However, the binding data reflect specific binding, since Gag does not bind to all fragments of the 5′UTR, even at concentrations as high as 1 μM (see below). These results suggest that Gag binding sites are within both the coding region and the 5′UTR, and that cooperation between the two structures is required for optimal binding. To map the binding sites and/or the structural rearrangement induced by Gag binding, we conducted DMS and V1 nuclease probing in presence of the protein. RNA (250 nM) was incubated in presence of 300 nM of Gag, a protein concentration does not promote additional alternative structure of any of the constructs (Figure 1).

The probing experiments were repeated at least two times, and we observed three different classes of footprints upon addition of Gag: some positions are protected from DMS (see for example Figure 7, A_{621}–A_{622} or Figure 2, A_{118}–A_{119}), others are protected from V1 nuclease cleavage (see for example Figure 2, C_{133}–A_{139} and U_{98}–U_{99}), and surprisingly, in some instances the V1 cleavage is enhanced upon gag binding (see for example Figure 4, G_{14}–G_{17} and Figure 8, C_{523}–G_{525}). Finally, some modifications remained unchanged (see for example Figure 2,

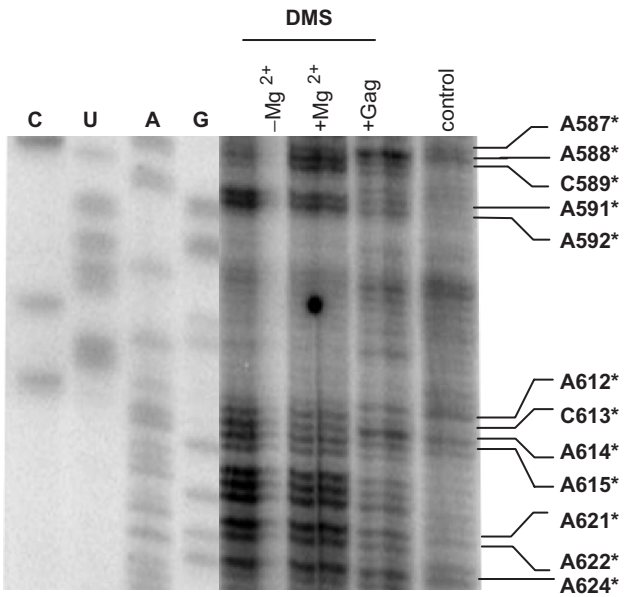

Figure 7. Chemical of SL13 in the coding region—Gag footprint on SL13. For comments see Figure 2.

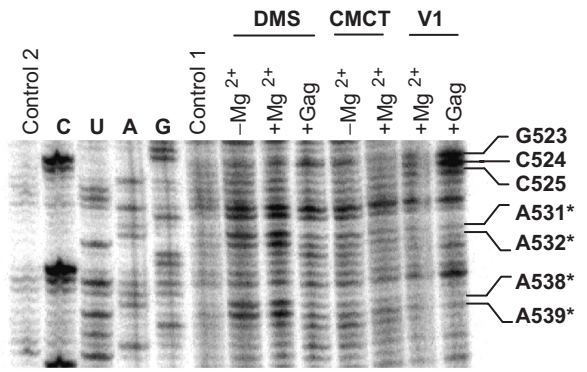

Figure 8. Chemical and enzymatic probing of SL11 in the coding region—Gag footprint on SL11. For comments see Figure 2. Note that chemical reagents induce a premature reverse transcriptase one nucleotide before the nucleotide hit. Therefore the chemically induced stops run one nucleotide further than the corresponding sequence and are therefore indicated with an asterisk.

A_{53}–A_{58}, Figure 7, A_{587}), which shows the specificity of the binding and the effect observed. Moreover, the same experiments, performed in presence of BSA at equivalent concentration in the same buffer, yield results identical to those obtained in absence of protein (data not shown). The observed protection patterns could be due to direct protein binding, or to an indirect effect upon a structural rearrangement induced by Gag. The experiments conducted do not allow us to formally distinguish between those two possibilities.

The changes of the modification pattern observed within the 5′UTR and in the coding region are reported in Figure 9. In four stems, SL1, SL11, SL12 and SL14, the presence of Gag enhances or even induces nuclease V1 cleavage (Figures 4 and 8). Although this is unexpected, such a footprint has already been observed with Gag (30,61), and it may originate from several phenomena that are discussed in the discussion section. In the 5′UTR, we observed a clear protection from V1 cleavage in the upper part of the PBS structure, and at the base of stems 2, 5 and 6. We observed protection from DMS in the loops of SL2, SL8, SL9, SL11, SL12 and the PBS. Interestingly, we observed many protections from DMS modification on nucleotides that are within or in the vicinity of SL10, the structure that spans from the 5′UTR to the coding region. Protections from DMS were also observed in the large loop located within the coding region (Figures 7 and 9).

Finally, to define the binding regions, we monitored Gag binding to short synthetic fragments by filter binding assays. We were unable to detect any binding to the 5′UTR fragments including SL4, SL5, SL6 (A_{158}–G_{275}), or including SL8 and SL9 (from G_{335} to U_{400}). Conversely, Gag binds to a fragment containing SL9 and SL10 (from G_{335} to U_{520}), although with a poor affinity (up to 30% of binding at 800 nM Gag). Altogether, our results suggest that the SL10 region is one of the important Gag-binding sites, although cooperation between different RNA regions is obviously required for efficient binding.

DISCUSSION

FIV 5′UTR is highly structured and shows some analogy with HIV-1 and HIV-2 leader

We used chemical and enzymatic probing and phylogenetic analysis to derive a secondary structure model for the Feline Immunodeficiency Virus leader and the *gag* coding sequence. FIV 5′UTR begins with a stem loop starting at the first transcribed nucleotide. In HIV-1 and HIV-2, such a structure, known as the trans-activation RNA element (TAR), is the target of the translational trans-activator protein (Tat). This protein, through its binding to both TAR and the cyclin T1, stimulates HIV transcription (18). FIV orfA (also designated as orf2) encodes for a small protein that stimulates gene expression driven from FIV LTR (11). However, this appears to be independent from the presence of 5′UTR sequences and thus orfA does not act through binding to SL1 (7–9). Other proteins have been shown to bind to HIV-1 TAR; for example, the translation initiation factor eif2 (25), and the viral accessory protein Vif (27). However, no clear homology other than its location could be made between HIV-1 TAR and FIV SL1 and thus we cannot extrapolate those results to FIV. Nevertheless, this structure and its sequence are conserved in all FIV viruses, which is undoubtedly a hallmark of a physiologically important element.

In HIV type 1 and 2 genomes, the major splice Site Donor is sequestered in a stem-loop, which is thought to regulate splicing efficiency (33). Such a hairpin could be formed, and is conserved, in FIV. Nevertheless, in agreement with our probing/footprinting data, and *in vivo* packaging experiments (see below), we modelled a structure incompatible with the SD stem loop. Nevertheless, conservation of a potential splice site donor stem loop indicates that this structure has a functional role. It is interesting to speculate that the two different structures could exist and regulate splicing and/or packaging (see below).

Elements involved in FIV dimerization

Lentiviruses genomic RNA are packaged in the virion particle as non-covalent dimers. The dimerization process is thought to be initiated by the annealing of a self-complementary sequence present in the loop of the DIS stem-loop. This leads to the formation of a kissing complex, which holds together the two genomic RNA molecules in the form of loose dimers. The latter are supposed to be converted into tight dimers by isomerization and extension of the pairing into the stem sequences (20). However, extended duplexes were not observed *in vivo* (62), where the dimerization process appears more complex than anticipated. It could rely on other structural elements such as TAR (32,63), or palindromic sequences located immediately 5′ of DIS (31), although this is not confirm by *in vivo* structure probing (62). The different elements could be redundant or act synergistically. Although we have not carried out a thorough study of FIV dimerization, our work indicates that FIV genome efficiently forms 'loose' dimers *in vitro*.

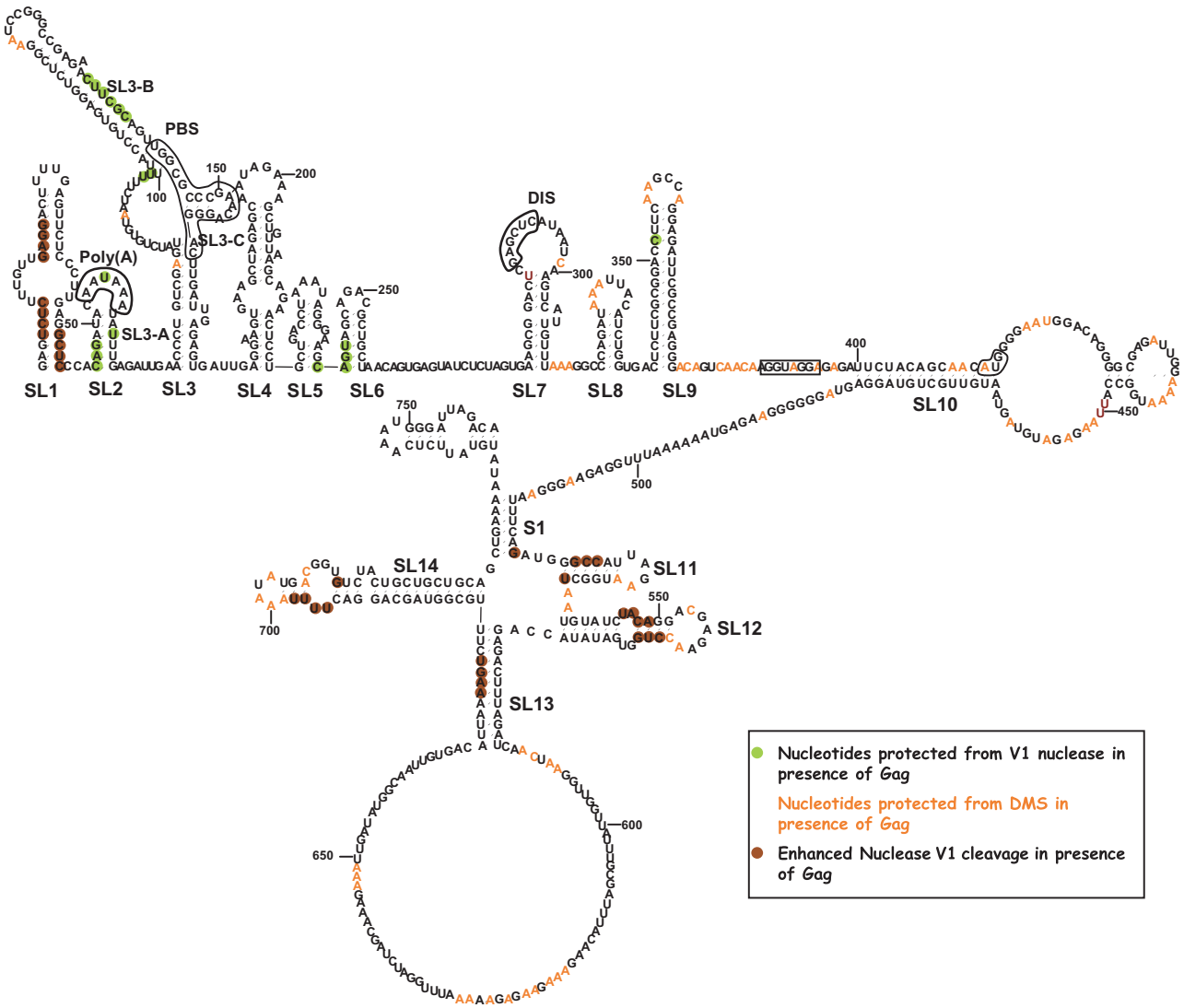

Figure 9. Footprint of the Gag protein on the 5′UTR and *gag* coding region. Colours reflect the modification of the accessibility of each nucleotide induced by Gag binding. Protection or enhancement was at least 2-fold.

However, under conditions in which HIV-1 and HIV-2 genomes yield at least 50% of stable dimers, FIV does not [Figure 1, data not shown; (22,38,64)]. Similarly to what is observed for HIV-1 and HIV-2, we have identified a hairpin in the 5′UTR, SL7, which loop harbours a 6-nt self-complementary sequence. However, it should be noted that not every self-complementary sequence exposed in a loop is able to form kissing complexes (20). Moreover, such a sequence is absent in three isolates which may have evolved an alternative solution. It seems therefore likely that, as shown for primate lentiviruses, other elements are involved in the dimerization process (32,63,65). First, SL1 harbours a self-complementary sequence that could be a dimerization site. Second, the PBS structure could be a good candidate as a primary or secondary dimerization site because its apical loop exposes a 4- to 6-nt self-complementary sequence. Furthermore, a sequence that has been implicated in HIV-2 dimerization (65) is also present in FIV PBS. Finally, our probing data indicate that dimerization contacts may occur in the coding region, but in any case, they could be only secondary dimerization site, since the coding region on its own does not form dimers (Figure 1). All the above hypotheses should be taken cautiously, and need a mutational analysis to be established. Several lines of evidences suggest that HIV-1 and HIV-2 dimerization process could be regulated by a conformational switch influenced by a long distance interaction between a sequence in U5 which is upstream from the primer-binding site, and bases surrounding the initiation codon (22–24,32). We were unable to detect such an interaction, or an analogous conformation involving such an extensive alternative structure in the FIV genomic RNA.

Gag-binding sites and genome packaging

HIV-1 Gag is a 55 kDa polyprotein which is cleaved by the viral protease into three main structural proteins, the Matrix, the Capsid and the Nucleocapsid.

However, the immature polyprotein is able to package the viral genome yielding immature viral particles (66). It mainly binds to small hairpins termed SL1–SL4, or psi sites, but also to the PBS (30). Gag-binding sites appear to be fairly degenerate, and it is difficult to predict a Gag-binding site from its sequence or structure (66,67). This could be explained if the Gag-binding site relies on a yet unidentified tertiary structure, although it should be noted that Gag binds with a reduced affinity to small fragments of its cognate 5′UTR (56,57,61,68,69). Using the nuclease V1 and DMS, we observed the HIV-1 Gag footprint on the FIV 5′UTR and *gag* coding region. Interestingly, we observed that in some instan ces (30,61), the presence of the protein induces ribonuclease V1 cleavage. We see several possible explanations for this. First, the presence of Gag prevents the formation of alternative structure of those stems, or promotes their cooperative folding, thus enhancing their stability and their susceptibility to nuclease V1. Second, Gag binding could disrupt elements of the tertiary structure leading to helix packing, thus uncovering those stems. Finally, Gag could induce new inter or intramolecular contacts. In summary, within the 5′UTR, the polyprotein alters the modification pattern within the poly(A) stem loop, the PBS structure, to SL6, SL8 and SL9, and to a single-strand region in between SL9 and the major splice site donor. SL6 had already been modelled as a Gag-binding hairpin because it is reminiscent of HIV-1 SL3, and lies within in a region important for packaging (16). Gag also induces footprints in a structure encompassing the end of the 5′UTR and the beginning of Gag-coding region, SL10, and to SL11, SL12 and SL13 within the open reading frame. Deletion experiments *ex vivo* have suggested that the important determinants for FIV packaging are the PBS region, the sequences downstream and the first 310 nt of the *gag* coding region (12–17). In the coding region, the first 90 nt downstream to the AUG appear to be critical for encapsidation, and addition of up to 230 nt further increases the packaging efficiency (15,17). This is particularly interesting because the first 90 nt enable the formation of SL10, and the 140 next nucleotides contain SL11, SL12 and at least the purine rich sequence in SL13 loop in which Gag induces a clear footprint as indicated by many protection patterns. Overall, our results are in good agreement with those obtained previously on packaging efficiency. However, the sequences required for the formation of SL10 are in between the splice site donor and AUG start codon, and have been shown to be dispensable for packaging (14,16). Nevertheless, this observation could be reconciled with our results: a careful analysis of the sequences of the mutants used for the packaging studies revealed that only 4 bp at the bottom of SL10 were disrupted (16,17). Thus, those constructs probably still harbour a shortened, but stable SL10. The involvement of the poly(A) in the encapsidation process is controversial (12,16). Our experiments suggest that Gag binds the poly(A) SL, this argues in favour of a role of this structure in FIV encapsidation. The involvement of the structures on both sides of the initiation codon suggests a discrimination pathway between the genomic and the spliced RNAs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by the CNRS and a grant from the 'Agence Nationale pour la Recherche sur le S.I.D.A' (ANRS). The following reagent was obtained through the NIH AIDS research and reference reagent program, division of AIDS, NIAID, NIH: HIV-1_{IIIB} p55Gag (cat. #3276) from Bio-Molecular Technology, Inc. Authors would like to thank Drs T. Ohlmann, N. Locker, F.Michel for critical reading of the manuscript, Drs T. Ohlmann and V. Camerini for sharing unpublished results. Special thanks go to Dr S. Butcher for thorough proofreading of our manuscript, and invaluable comments on our data. Research in our laboratory is funded by ANRS (Agence Nationale de la Recherche contre le S.I.D.A), ANR (Agence Nationale pour la Recherche) and CNRS. L.J is a recipient of a fellowship from ANRS. Funding to pay the Open Access publication charges for this article was provided by CNRS and ANRS.

Conflict of interest statement. None declared.

REFERENCES

1. Talbott,R.L., Sparger,E.E., Lovelace,K.M., Fitch,W.M., Pedersen,N.C., Luciw,P.A. and Elder,J.H. (1989) Nucleotide sequence and genomic organization of feline immunodeficiency virus. *Proc. Natl Acad. Sci. USA*, **86**, 5743–5747.
2. Foley,B.T. (2000) In Kuiken,C., Foley,B., Freed,E., Hann,B., Korber,B., Marx,P., McCutchan,F., Melors,J., Mullins,J. and Sodorski,J. (eds), *HIV Sequence Compendium 2000*. Los Alamos Natl Lab, Los Alamos, NM, pp. 35–43.
3. Miyazawa,T., Tomonaga,K., Kawaguchi,Y. and Mikami,T. (1994) The genome of feline immunodeficiency virus. *Arch. Virol.*, **134**, 221–234.
4. Burkhard,M.J. and Dean,G.A. (2003) Transmission and immunopathogenesis of FIV in cats as a model for HIV. *Curr. HIV Res.*, **1**, 15–29.
5. Pistello,M. (2008) Should accessory proteins be structural components of lentiviral vaccines? Lessons learned from the accessory ORF-A protein of FIV. *Vet. Immunol. Immunopathol.*, **123**, 144–149.
6. Dunham,S.P. (2006) Lessons from the cat: development of vaccines against lentiviruses. *Vet. Immunol. Immunopathol.*, **112**, 67–77.
7. Sundstrom,M., Chatterji,U., Schaffer,L., de Rozieres,S. and Elder,J.H. (2007) Feline immunodeficiency virus OrfA alters gene expression of splicing factors and proteasome-ubiquitination proteins. *Virology* **371**, 394–404.
8. Chatterji,U., de Parseval,A. and Elder,J.H. (2002) Feline immunodeficiency virus OrfA is distinct from other lentivirus transactivators. *J. Virol.*, **76**, 9624–9634.
9. Gemeniano,M.C., Sawai,E.T. and Sparger,E.E. (2004) Feline immunodeficiency virus Orf-A localizes to the nucleus and induces cell cycle arrest. *Virology*, **325**, 167–174.
10. Gemeniano,M.C., Sawai,E.T., Leutenegger,C.M. and Sparger,E.E. (2003) Feline immunodeficiency virus ORF-Ais required for virus particle formation and virus infectivity. *J. Virol.*, **77**, 8819–8830.
11. de Parseval,A. and Elder,J.H. (1999) Demonstration that orf2 encodes the feline immunodeficiency virus transactivating (Tat) protein and characterization of a unique gene product with partial rev activity. *J. Virol.*, **73**, 608–617.
12. Ghazawi,A., Mustafa,F., Phillip,P.S., Jayanth,P., Ali,J. and Rizvi,T.A. (2006) Both the 5′ and 3′ LTRs of FIV contain minor RNA encapsidation determinants compared to the two core

packaging determinants within the 5′ untranslated region. *Microbes Infect*., **8**, 767–778.
13. Browning,M.T., Mustafa,F., Schmidt,R.D., Lew,K.A. and Rizvi,T.A. (2003) Sequences within the gag gene of feline immunodeficiency virus (FIV) are important for efficient RNA encapsidation delineation of sequences important for efficient packaging of feline immunodeficiency virus RNA. *Virus Res*., **93**, 199–209.
14. Mustafa,F., Ghazawi,A., Jayanth,P., Phillip,P.S., Ali,J. and Rizvi,T.A. (2005) Sequences intervening between the core packaging determinants are dispensable for maintaining the packaging potential and propagation of feline immunodeficiency virus transfer vector RNAs. *J. Virol*., **79**, 13817–13821.
15. Browning,M.T., Mustafa,F., Schmidt,R.D., Lew,K.A. and Rizvi,T.A. (2003) Delineation of sequences important for efficient packaging of feline immunodeficiency virus RNA. *J. Gen. Virol*., **84**, 621–627.
16. Kemler,I., Barraza,R. and Poeschla,E.M. (2002) Mapping the encapsidation determinants of feline immunodeficiency virus. *J. Virol*., **76**, 11889–11903.
17. Kemler,I., Azmi,I. and Poeschla,E.M. (2004) The critical role of proximal gag sequences in feline immunodeficiency virus genome encapsidation. *Virology*, **327**, 111–120.
18. Gatignol,A. (2007) Transcription of HIV: Tat and cellular chromatin. *Adv. Pharmacol*., **55**, 137–159.
19. Lever,A.M. (2007) HIV-1 RNA packaging. *Adv. Pharmacol*., **55**, 1–32.
20. Paillart,J.C., Shehu-Xhilaga,M., Marquet,R. and Mak,J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol*., **2**, 461–472.
21. Huthoff,H. and Berkhout,B. (2002) Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry*, **41**, 10439–10445.
22. Dirac,A.M., Huthoff,H., Kjems,J. and Berkhout,B. (2002) Regulated HIV-2 RNA dimerization by means of alternative RNA conformations. *Nucleic Acids Res*., **30**, 2647–2655.
23. Abbink,T.E., Ooms,M., Haasnoot,P.C. and Berkhout,B. (2005) The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry*, **44**, 9058–9066.
24. Ooms,M., Huthoff,H., Russell,R., Liang,C. and Berkhout,B. (2004) A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J. Virol*., **78**, 10814–10819.
25. Ben-Asouli,Y., Banai,Y., Hauser,H. and Kaempfer,R. (2000) Recognition of 5'-terminal TAR structure in human immunodeficiency virus-1 mRNA by eukaryotic translation initiation factor 2. *Nucleic Acids Res*., **28**, 1011–1018.
26. Haase,A.D., Jaskiewicz,L., Zhang,H., Laine,S., Sack,R., Gatignol,A. and Filipowicz,W. (2005) TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep*., **6**, 961–967.
27. Henriet,S., Richer,D., Bernacchi,S., Decroly,E., Vigne,R., Ehresmann,B., Ehresmann,C., Paillart,J.C. and Marquet,R. (2005) Cooperative and specific binding of Vif to the 5′ region of HIV-1 genomic RNA. *J. Mol. Biol*., **354**, 55–72.
28. Das,A.T., Klaver,B. and Berkhout,B. (1999) A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection. *J. Virol*., **73**, 81–91.
29. Abbink,T.E. and Berkhout,B. (2007) HIV-1 reverse transcription: close encounters between the viral genome and a cellular tRNA. *Adv. Pharmacol*., **55**, 99–135.
30. Damgaard,C.K., Dyhr-Mikkelsen,H. and Kjems,J. (1998) Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res*., **26**, 3667–3676.
31. L'Hernault,A., Greatorex,J.S., Crowther,R.A. and Lever,A.M. (2007) Dimerisation of HIV-2 genomic RNA is linked to efficient RNA packaging, normal particle maturation and viral infectivity. *Retrovirology*, **4**, 90.
32. Song,R., Kafaie,J. and Laughrea,M. (2008) Role of the 5′ TAR Stem-Loop and the U5-AUG duplex in dimerization of HIV-1 Genomic RNA. *Biochemistry*, **47**, 3283–3293.
33. Abbink,T.E. and Berkhout,B. (2008) RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J. Virol*., **82**, 3090–3098.
34. Abbink,T.E. and Berkhout,B. (2003) A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J. Biol. Chem*., **278**, 11601–11611.
35. Paillart,J.C., Skripkin,E., Ehresmann,B., Ehresmann,C. and Marquet,R. (2002) In vitro evidence for a long range pseudoknot in the 5′-untranslated and matrix coding regions of HIV-1 genomic RNA. *J. Biol. Chem*., **277**, 5995–6004.
36. D'Souza,V. and Summers,M.F. (2005) How retroviruses select their genomes. *Nat. Rev. Microbiol*., **3**, 643–655.
37. Lanchy,J.M., Rentz,C.A., Ivanovitch,J.D. and Lodmell,J.S. (2003) Elements located upstream and downstream of the major splice donor site influence the ability of HIV-2 leader RNA to dimerize in vitro. *Biochemistry*, **42**, 2634–2642.
38. Laughrea,M. and Jette,L. (1997) HIV-1 genome dimerization: kissing-loop hairpin dictates whether nucleotides downstream of the 5′ splice junction contribute to loose and tight dimerization of human immunodeficiency virus RNA. *Biochemistry*, **36**, 9501–9508.
39. Butcher,S.E. and Burke,J.M. (1994) Structure-mapping of the hairpin ribozyme. Magnesium-dependent folding and evidence for tertiary interactions within the ribozyme-substrate complex. *J. Mol. Biol*., **244**, 52–63.
40. Weill,L., Louis,D. and Sargueil,B. (2004) Selection and evolution of NTP-specific aptamers. *Nucleic Acids Res*., **32**, 5045–5058.
41. Paillart,J.C., Dettenhofer,M., Yu,X.F., Ehresmann,C., Ehresmann,B. and Marquet,R. (2004) First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem*., **279**, 48397–48403.
42. Lanchy,J.M., Ivanovitch,J.D. and Lodmell,J.S. (2003) A structural linkage between the dimerization and encapsidation signals in HIV-2 leader RNA. *RNA*, **9**, 1007–1018.
43. Huthoff,H. and Berkhout,B. (2001) Two alternating structures of the HIV-1 leader RNA. *RNA*, **7**, 143–157.
44. Damgaard,C.K., Andersen,E.S., Knudsen,B., Gorodkin,J. and Kjems,J. (2004) RNA interactions in the 5′ region of the HIV-1 genome. *J. Mol. Biol*., **336**, 369–379.
45. Brunel,C. and Romby,P. (2000) Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol*., **318**, 3–21.
46. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*., **31**, 3406–3415.
47. Woese,C.R., Winker,S. and Gutell,R.R. (1990) Architecture of ribosomal RNA: constraints on the sequence of ''tetra-loops''. *Proc. Natl Acad. Sci. USA*, **87**, 8467–8471.
48. Molinaro,M. and Tinoco,I. Jr. (1995) Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic Acids Res*., **23**, 3056–3063.
49. Leontis,N.B., Lescoute,A. and Westhof,E. (2006) The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol*., **16**, 279–287.
50. Lescoute,A., Leontis,N.B., Massire,C. and Westhof,E. (2005) Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res*., **33**, 2395–2409.
51. Lescoute,A. and Westhof,E. (2006) The interaction networks of structured RNAs. *Nucleic Acids Res*., **34**, 6587–6604
52. Schuwirth,B.S., Borovinskaya,M.A., Hau,C.W., Zhang,W., Vila-Sanjurjo,A., Holton,J.M. and Cate,J.H. (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science*, **310**, 827–834.
53. Shankar,N., Kennedy,S.D., Chen,G., Krugh,T.R. and Turner,D.H. (2006) The NMR structure of an internal loop from 23S ribosomal RNA differs from its structure in crystals of 50s ribosomal subunits. *Biochemistry*, **45**, 11776–11789.
54. Tolbert,B.S., Kennedy,S.D., Schroeder,S.J., Krugh,T.R. and Turner,D.H. (2007) NMR structures of (rGCUGAGGCU)2 and (rGCGGAUGCU)2: probing the structural features that shape the thermodynamic stability of GA pairs. *Biochemistry*, **46**, 1511–1522.
55. Hagan,N.A. and Fabris,D. (2007) Dissecting the protein-RNA and RNA-RNA interactions in the nucleocapsid-mediated dimerization and isomerization of HIV-1 stemloop 1. *J. Mol. Biol*., **365**, 396–410.
56. Yuan,Y., Kerwood,D.J., Paoletti,A.C., Shubsda,M.F. and Borer,P.N. (2003) Stem of SL1 RNA in HIV-1: structure and

nucleocapsid protein binding for a 1 x 3 internal loop. *Biochemistry*, **42**, 5259–5269.
57. Zeffman,A., Hassard,S., Varani,G. and Lever,A. (2000) The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein. *J. Mol. Biol.*, **297**, 877–893.
58. Baig,T.T., Lanchy,J.M. and Lodmell,J.S. (2007) HIV-2 RNA dimerization is regulated by intramolecular interactions in vitro. *RNA*, **13**, 1341–1354.
59. Lanchy,J.M. and Lodmell,J.S. (2007) An extended stem-loop 1 is necessary for human immunodeficiency virus type 2 replication and affects genomic RNA encapsidation. *J. Virol.*, **81**, 3285–3292.
60. Browning,M.T., Schmidt,R.D., Lew,K.A. and Rizvi,T.A. (2001) Primate and feline lentivirus vector RNA packaging and propagation by heterologous lentivirus virions. *J. Virol.*, **75**, 5129–5140.
61. Maddison,B., Marya,P. and Heaphy,S. (1998) HIV-1 Gag binds specifically to RNA stem-loops in the 5' leader sequence. *Biochim. Biophys. Acta*, **1398**, 305–314.
62. Wilkinson,K.A., Gorelick,R.J., Vasa,S.M., Guex,N., Rein,A., Mathews,D.H., Giddings,M.C. and Weeks,K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.*, **6**, e96.
63. Andersen,E.S., Contera,S.A., Knudsen,B., Damgaard,C.K., Besenbacher,F. and Kjems,J. (2004) Role of the trans-activation response element in dimerization of HIV-1 RNA. *J. Biol. Chem.*, **279**, 22243–22249.
64. Lanchy,J.M. and Lodmell,J.S. (2002) Alternate usage of two dimerization initiation sites in HIV-2 viral RNA in vitro. *J. Mol. Biol.*, **319**, 637–648.
65. Jossinet,F., Lodmell,J.S., Ehresmann,C., Ehresmann,B. and Marquet,R. (2001) Identification of the in vitro HIV-2/SIV RNA dimerization site reveals striking differences with HIV-1. *J. Biol. Chem.*, **276**, 5598–5604.
66. Scarlata,S. and Carter,C. (2003) Role of HIV-1 Gag domains in viral assembly. *Biochim. Biophys. Acta*, **1614**, 62–72.
67. Cruceanu,M., Urbaneja,M.A., Hixson,C.V., Johnson,D.G., Datta,S.A., Fivash,M.J., Stephen,A.G., Fisher,R.J., Gorelick,R.J., Casas-Finet,J.R. *et al.* (2006) Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res.*, **34**, 593–605.
68. Amarasinghe,G.K., De Guzman,R.N., Turner,R.B., Chancellor,K.J., Wu,Z.R. and Summers,M.F. (2000) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.*, **301**, 491–511.
69. De Guzman,R.N., Wu,Z.R., Stalling,C.C., Pappalardo,L., Borer,P.N. and Summers,M.F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*, **279**, 384–388.