Nucleic Acids Research, 1998, Vol. 26, No. 16 **3667–3676**

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

Christian Kroun Damgaard, Helle Dyhr-Mikkelsen and Jørgen Kjems*

Department of Molecular and Structural Biology, University of Aarhus, C.F. Møllers Allé, Building 130, DK-8000 Aarhus C, Denmark

Received May 20, 1998; Revised and Accepted July 6, 1998

ABSTRACT

Encapsidation of HIV-1 genomic RNA is mediated by specific interactions between the RNA packaging signal and the Gag protein. During maturation of the virion, the Gag protein is processed into smaller fragments, including the nucleocapsid (NC) domain which remains associated with the viral genomic RNA. We have investigated the binding of glutathione-*S*-transferase (GST) Gag and NC fusion proteins from HIV-1, to the entire HIV-1 and -2 leader RNA encompassing the packaging signal. We have mapped the binding sites at conditions where only about two complexes are formed and find that GST-Gag and GST-NC fusion proteins bind specifically to discrete sites within the leader. Analysis of the HIV-1 leader indicated that GST-Gag strongly associates with the PSI stem–loop and to a lesser extent with regions near the primer binding site. GST-NC binds the same regions but with reversed preferences. The HIV-1 proteins also interact specifically with the 5′-leader of HIV-2 and the major site of interaction mapped to a stem–loop, with homology to the HIV-1 PSI stem–loop structure. The different specificities of Gag and NC may reflect functionally distinct roles in the viral replication, and suggest that the RNA binding specificity of NC is modulated by its structural context.

INTRODUCTION

Packaging of retroviral genomic RNA into the virion is an essential step of the retroviral replication cycle. Two copies of the full-length genomic RNA are incorporated into the progeny virion while spliced viral RNAs and cellular RNAs are largely excluded. The specificity relies on an interaction between the nucleocapsid (NC) domain of the viral Gag protein and structural elements within the 5′-end of the viral genomic RNA, referred to as the packaging signal (PSI) (reviewed in 1 and 2). In HIV-1, the secondary structure of the viral leader encompassing the PSI has been investigated by computer modelling, phylogenetic comparison and enzymatic and chemical probing, and the results have led to different models (3–12). Most of the proposed models are compatible with the presence of seven stem–loop structures which have been named after their putative functions in the viral replication cycle (reviewed in 12): the TAR stem–loop which binds the viral transcriptional transactivator Tat; the Poly(A) stem–loop which harbors the inactive 5′-Poly(A) site; the PBS stem–loop containing the binding site for the tRNA primer for reverse transcription; the DIS stem–loop which is involved in dimerization through a kissing loop interaction; the SD stem–loop containing the major splice donor; the PSI stem–loop which has been mapped as a major determinant for specific packaging; and the AUG stem–loop encompassing the Gag initiation codon.

The PSI stem–loop has been shown to play a major role in packaging *in vivo* (10,11,13–20), but several regions have been shown to enhance this process. In particular the DIS stem–loop has, besides its possible role in the initiation of genomic dimerization (21), been shown to be important for RNA packaging (10,11,16,20,22–24). However, it is unclear whether the DIS stem–loop serves a direct role as a packaging signal or whether it is indirect in the sense that a DIS stem–loop mediated dimerization enhances packaging efficiency (20,22,23). In addition, sequences in the R and U5 region including the putative TAR and Poly(A) stem–loops and the 5′-end of the Gag open reading frame have also been shown to enhance packaging (11,19,25–27).

During virion maturation the Gag protein is processed into smaller fragments including the matrix (MA), capsid (CA) and NC proteins. After the excision from the Gag precursor it is estimated that ~2000 copies of the NC protein remain associated with the viral genome (28). Although the NC domain has a general non-specific binding activity towards nucleic acids (29,30), several lines of evidence point to the NC domain of Gag as being responsible for the specific RNA binding required for encapsidation of the viral genome. This has been demonstrated by introducing mutations in the conserved regions of NC and characterizing these mutants either by their viral RNA binding

*To whom correspondence should be addressed. Tel: +45 89 42 26 86; Fax: +45 86 19 65 00; Email: kjems@biobase.dk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

properties *in vitro* (29,31–37) or by their ability to specifically incorporate viral genomic RNA into assembling virions (14,34,38–41). In addition, studies using chimeric Gag proteins of HIV-1 and MLV show that the NC domain possesses the specificity required for the encapsidation of the cognate viral genome (42,43).

NC also seems to play an important role in other steps of the viral replication cycle. It has been reported to promote specific annealing of the tRNA primer to the primer binding site, minus strand strong stop cDNA synthesis, processivity of the reverse transcriptase, strand transfer during reverse transcription and annealing of complementary DNA oligonucleotides *in vitro* (31,44–50).

Binding of HIV-1 Gag and NC proteins to viral RNAs *in vitro* has been studied by UV-crosslinking (51), by north-western blotting (36,52), by filterbinding (9,35,37,53) and by electrophoretic mobility shift assays (6,32,33,37). Together, these results indicate that Gag and NC bind specifically to viral RNAs derived from the leader region and 5′-terminal Gag ORF sequences. The source of protein used in these studies ranges from synthetic non-tag peptides (6,35,37,51) to recombinant His-tagged (36,53) and glutathione-*S*-transferase (GST)-tagged (9,32,33,35) versions of NCp7, NCp15 and Gag. The GST tagged forms of Gag and NC have been shown to bind specifically and with high affinity to an RNA encompassing the DIS, SD, PSI and AUG stem–loops (9). Using smaller RNAs spanning combinations of one to three of the above stem–loops, an interaction was still detected, although of a lower affinity, suggesting a somewhat degenerate recognition mechanism (9,33). Moreover, transcripts spanning 5′-terminal Gag ORF sequences have been shown to enhance specific binding of GST-Gag and GST-NC (32,33). Thus, RNA sequences important for specific binding of Gag and NC appear to overlap with the regions important for *in vivo* packaging supporting a functional link between the two processes.

In this study, we have employed an RNA footprinting strategy using recombinant GST-tagged versions of Gag and NC from HIV-1 and RNAs encompassing the entire HIV-1 or HIV-2 leader RNA. Our results demonstrate that the GST-tagged NC binds to discrete positions within the HIV-1 viral leader, but with different preferences to structural elements of the RNA. Moreover, specific binding of HIV-1 Gag and NC to a PSI-like structure of HIV-2 provides a molecular model for efficient cross-packaging of the two viruses.

MATERIALS AND METHODS

Cloning, expression and purification of GST-tagged Gag and NC proteins

DNA fragments corresponding to full-length Gag and NC from the HIV-1 LAI strain were PCR amplified from the pGST-Gag plasmid (32) using the following primers: 5′-GAGGATCCATG-GGTGCGAGAGCGTCAGTA and 5′-GAGAATTCCTTGTG-ACGAGGGGTCGTTGC for Gag, and 5′-GAGGATCCATGC-AGAGAGGCAATTTTAGG and 5′-GAGAATTCCATTAGCC-TGTCTCTCAGTAC for NC. PCR products were digested with *Bam*HI and *Eco*RI and cloned into the *Bam*HI–*Eco*RI sites of the pGEX-GTH expression vector (54).

The proteins were expressed in *Escherichia coli* BL21 strain by induction with 1 mM IPTG after growth to an A_{600} of 1.0 and harvested after 3 h. Harvest, lysis and purification on glutathione agarose beads was performed essentially as described by the manufacturer. Cells were lysed by sonication on ice for 4 min (10 s burst, 20 s rest) in buffer L [20 mM Tris–HCl (pH 8.0), 100 mM NaCl, 0.5% Nonidet P-40, 0.5 mM EDTA, 50 mM DTT, 0.2 mM PMSF]. Bacterial debris was removed by centrifugation and the supernatant was incubated with glutathione–Sepharose 4B (Pharmacia) at 4°C for 3 h with gentle agitation. Beads were collected by centrifugation and washed three times in buffer L. At this step the fusion proteins were either stored, while bound to the affinity beads in buffer L containing 20% glycerol, or eluted at room temperature by gentle shaking in elution buffer [20 mM reduced glutathione, 100 mM Tris–HCl (pH 8.0), 120 mM NaCl]. The eluted proteins were stored at –80°C in elution buffer containing 20% glycerol. Protein concentration was estimated from Coomassie Blue stained SDS gels. Synthetic NC (NCp7) from HIV-1 LAV strain was generously provided by Jean-Luc Darlix (55).

RNA transcription

All transcripts are numbered with respect to the 5′-end of genomic viral RNA. They were synthesized *in vitro* by T7 RNA polymerase run-off transcription. The HIV-1 RNA was expressed from the pT7HIV1–627 plasmid containing nt 1–627 of the HIV-1 LAI strain cloned into a *Stu*1 site downstream of a T7 promotor (unpublished data). Templates linearized with *Hae*III were used for the generation of transcripts corresponding to positions 1–401 containing wild-type 5′- and 3′-termini. The plasmid pUC8-HIV-2, encoding the RNA leader of the HIV-2 ROD strain genome was kindly provided by Benjamin Berkhout (56). The plasmid was linearized with *Bst*YI to generate a template corresponding to positions 1–738 of the viral genome. Radiolabelled RNA was generated with the T7-MEGAshort-script™ kit (Ambion). Each reaction contains 7.5 mM ATP, 7.5 mM CTP, 7.5 mM GTP, 1.5 mM UTP, 25 µCi [α-^{32}P]UTP (Amersham, 400 Ci/mmol), 3 µg DNA template and buffer supplied with the kit in a reaction volume of 20 µl. Reactions were incubated for 2 h at 37°C and the RNA was gel purified and quantified by the amount of ^{32}P incorporation. RNA containing the IIB helix from Rev response element was prepared according to Kjems *et al.* (57).

Electrophoretic mobility shift assay

RNA was renatured at a concentration of ~0.1–0.2 pmol/µl by incubation at 80°C for 5 min in renaturation buffer [10 mM HEPES–KOH (pH 7.6), 100 mM KCl] followed by slow cooling from 65 to 25°C and left on ice. Complexes were formed by incubating 0.35 pmol of renatured RNA with 0–4.6 µg rRNA as indicated and 0–66 pmol GST-Gag or GST-NC (diluted in 3 µl elution buffer) in 10 µl binding buffer [50 mM Tris–HCl (pH 8.0), 50 mM NaCl, 100 µM $ZnCl_2$, 10 U RNasin™ (Promega), 5 mM DTT]. After incubation for 15 min at room temperature, 2 µl 30% glycerol were added and reactions were loaded on a pre-run non-denaturing acrylamide gel [100 mM Tris–borate (pH 8.3) and 1 mM EDTA]. The gel was run at 12 V/cm for 2.5 h at 4°C and bands were visualized by autoradiography.

RNase probing and primer extension

Four picomoles of HIV-1 or -2 leader RNA were renaturated in 20 µl as described above and mixed with 50 µg (for HIV-1) or 97 µg (for HIV-2) rRNA in a final volume of 210 µl binding

buffer. This mixture was split into three aliquots of 70 µl, which were incubated with 10 µl of either storage buffer, or 65 or 130 pmol of protein (GST-Gag or GST-NC). After incubation for 15 min at room temperature, each reaction was further split into four 20 µl aliquots to which either 2 µl ddH_2O, 2 µl RNase T_1 (12 U/ml), 2 µl RNase T_2 (200 U/ml) or 2 µl RNase V_1 (350 U/ml) were added on ice. After incubation for 20 min on ice, the reactions were terminated by addition of 180 µl 300 mM NaAc (pH 6.0), and the RNA was recovered by phenol extraction followed by precipitation. The RNA pellets were either stored at –80°C or redissolved in 4.8 µl ddH_2O. For the primer extension reaction, 2.4 µl of the RNA were mixed with 0.2 pmol of 5′-end-labelled primer in a volume of 6 µl of annealing buffer [10 mM Tris–HCl (pH 6.9), 40 mM KCl, 0.5 mM EDTA]. After a 1 min incubation at 95°C, the samples were transferred to 50°C for 10 min and stored on ice. To each sample, 4 µl of an extension mixture containing 1 µl 10× RT buffer [500 mM Tris–HCl (pH 8.4), 100 mM $MgCl_2$, 20 mM DTT], 1.75 mM dNTPs and 2 U of AMV reverse transcriptase (Amersham) were added, giving a final concentration of 0.7 mM of each dNTP, and the incubation was continued at 46°C for 30 min before the addition of 40 µl of 300 mM NaAc (pH 6.0) and precipitation. The cDNA was resolved on 6 or 8% denaturing polyacrylamide gels [8 M urea, 100 mM Tris–borate (pH 8.3) and 1 mM EDTA]. A marker sequence was obtained from four reactions with untreated RNA where individual dideoxynucleotides were included at a final concentration of 0.1 mM of ddATP or ddTTP or 0.05 mM ddCTP or ddGTP. The gels were exposed and quantified on a PhosporImager. Primers used for reverse transcription were: HIV-1, 5′-CCTTAACCGAATTTTTTCCC (positions 382–401) or 5′-CTCGCACCCATCTCTCTCC (positions 328–346); HIV-2, 5′-CTGATTCTTTCTAATTCATCTGC (positions 581–603). 5′-end labelling and RNase protection mapping of end-labelled HIV-1 leader RNA was done according to Kjems *et al.* (57).

RESULTS

Gag and NC bind specifically to an HIV-1 RNA spanning the viral leader

Specific binding of Gag and NC protein to HIV-1 RNA has previously been established (6,9,32,33,35,36,53). We have examined the binding of Gag and NC to a viral RNA (1–401) corresponding to the entire 5′-UTR and 66 nt of the Gag open reading frame in more detail. Initially, synthetic NCp7 peptide was tested in band shift analysis, but only very large complexes were formed which remained in the gel slot (data not shown). In a search for conditions under which a single or only a few proteins bind the RNA, we found that discrete complexes were formed in band shift assays when using recombinant GST-tagged versions of Gag and NC (Fig. 1A and B). Both proteins bound the viral RNA with ~100-fold higher affinity than unrelated RNA, whereas GST alone did not bind to the RNA (data not shown). To analyse the specificity of the complexes, titrations of protein and competitor rRNA concentrations were made (Fig. 1A and B). The number of complexes formed between GST-NC and HIV-1 leader RNA ranged from one to five depending on the protein concentration (20–100-fold molar excess to viral RNA). In contrast, only one complex was observed when incubating a 39 nt long RNA fragment corresponding to helix IIB of the Rev response element (57) in the presence of 100–1000 molar excess

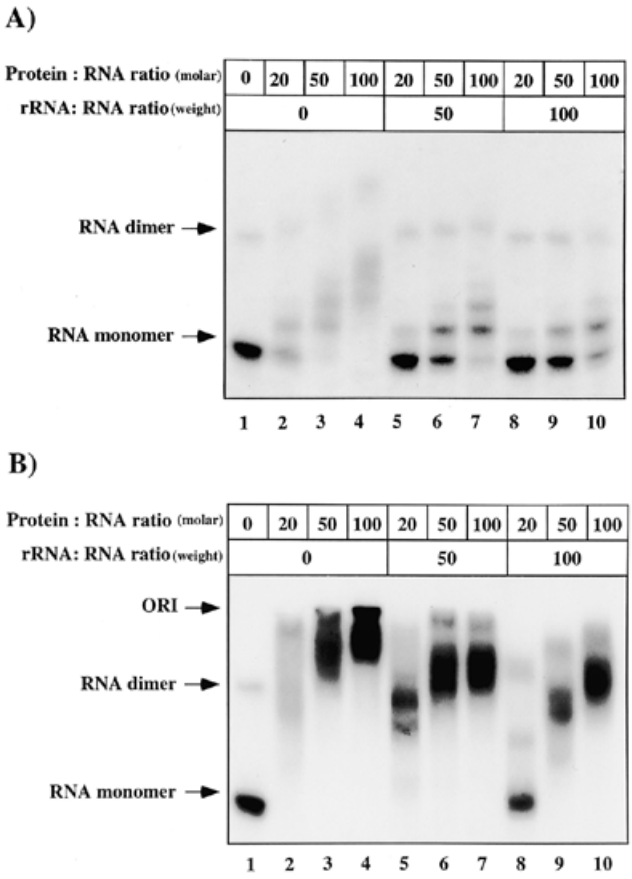

Figure 1. Gel mobility shift assay of complex formation between HIV-1 RNA (1–401) and GST-NC (**A**) or HIV-1 GST-Gag (**B**). Forty-five nanograms (0.35 pmol) of renatured HIV-1 RNA were mixed with no (lanes 1–4), 2.2 µg (lanes 5–7) or 4.5 µg (lanes 8–10) of rRNA and increasing amounts of GST-NC or GST-Gag protein (0, 7, 17.5 or 35 pmol) corresponding to the indicated protein:HIV-1 RNA ratio. After incubation for 15 min at room temperature the resulting complexes were separated on a 5% native acrylamide gel. Position of the gel top (ORI) and the bands corresponding to monomeric and dimeric RNA alone are indicated.

of NC, implying that NC oligomerization is negligible under these conditions (data not shown). Addition of rRNA as competitor to the binding reaction (50–100-fold excess to viral RNA) reduced the amount of the slowest migrating complexes, but two of the complexes were relatively resistant to competition (Fig. 1A, compare lanes 3 and 4 with lanes 6, 7, 9 and 10). Assuming that the ladder of complexes corresponds to addition of single GST-NC molecules, this indicates that about two GST-NC molecules bind strongly and specifically to the HIV-1 RNA. GST-NC also shifts the band corresponding to the RNA dimer (Fig. 1A, lanes 1–4) but this shift disappears in the presence of 50–100-fold excess of rRNA, suggesting that the GST-NC does not interact specifically with the dimer RNA.

Binding of GST-Gag to the HIV-1 RNA (1–401) gave a similar result (Fig. 1B). In the absence of rRNA, increasing concentrations of GST-Gag gave rise to a smear of complexes of decreasing mobility (Fig. 1B, lanes 1–4) and it was difficult to resolve the complexes into distinct bands. As seen with GST-NC, the lower complexes became more distinct in the presence of a large

excess of rRNA, and remained resistant to competition (Fig. 1B, lanes 5–10). This indicates a specific interaction of several GST-Gag molecules with the viral RNA, but from these band shift gels it is difficult to assess whether GST-Gag also bound specifically to the RNA dimer.

GST-Gag and GST-NC bind to structural motifs within the viral leader with different affinity and specificity

In order to map primary binding sites of the GST-tagged versions of Gag and NC within the HIV-1 RNA leader, we employed an RNA footprinting assay. The RNA was digested in the absence and in the presence of GST-NC or GST-Gag with one of three different ribonucleases (T_1, T_2 and V_1) and the extent of cleavage was subsequently detected by primer extension (Fig. 2A and B) or by using end-labelled RNA (Fig. 2C). RNase T_1 cleaves 3′ to guanosines in single stranded RNA, RNase T_2 cleaves preferentially 3′ to adenosines in single stranded RNA and RNase V_1 cleaves non-specifically within double stranded regions. The cleavage pattern obtained in the absence of protein yielded information about the RNA structure and the results from several experiments are summarized in Figure 3. As estimated from Figure 1, lane 1, ~95% of the RNA is monomeric in the absence of protein and structural data originating from dimeric RNA molecules is therefore negligible. Our data are largely consistent with the structure depicted in Figure 3. Nearly all single strand specific cleavages were seen in the regions expected to form loops, while double strand specific cleavages were primarily found in the stem region of the stem–loop structures (see Discussion). The RNase cleavage reactions in the presence of either GST-Gag or GST-NC were performed under the same conditions as used for the band shift analyses in lanes 9 and 10 of Figure 1A and B where only one or two molecules of protein are expected to bind per molecule of RNA. Two different primers complementary to the transcript were used to resolve the whole RNA. The footprinting was repeated at least three times for both Gag and NC and essentially the same result was obtained each time (Fig. 2A and B; summarized in Fig. 4A and B). Several positions within the RNA became specifically protected against RNase cleavage by GST-Gag and GST-NC, and other cleavages were specifically enhanced. For GST-Gag, the major site of protection was found in the PSI loop (positions G317–G320) where a >3-fold protection was observed using both T_2 and T_1 RNases (Fig. 2A, lanes 1–2 and 4–6). In the preceding SD stem–loop structure, the binding of Gag resulted in enhanced single strand specific RNase T_1 cleavages in the loop (G289–G290), whereas the double strand specific RNase V_1 cleavage in the stem (C287–U288) became strongly reduced (Fig. 2A, lanes 11–13 and 4–6). Strong protection against RNase T_1 cleavage was also observed in the region between the DIS and SD stems (G272–G273) and between the Poly(A) and PBS stems (G106 and G108) as well as a minor protection at positions just downstream of the Gag start codon (G346 and G348) (Fig. 2A).

Incubation of the same HIV-1 RNA with GST-NC gave rise to a protection pattern which showed both similarities to and differences from the pattern seen with GST-Gag (compare Fig. 4A and B). A somewhat weaker protection (1.5–2-fold) in the PSI loop was apparent while the characteristic enhancements and protections in the SD stem–loop structure were unchanged. The major difference, as compared to the Gag footprint, was the strong protection to T_1 at a region downstream from the primer binding site (G212–G214) and the increased protection against T_1 and T_2 of nucleotides within the primer binding site (G195–G197, G181–G184) and (A191, A192 and A194) respectively (Fig. 4A and B). In addition, GST-NC also gave minor protections upstream from the primer binding site (U153–U156 and A168–A170). Protections at A203–A205 were seen for both proteins. The observed protections against cleavages in the primer binding site region were confirmed by reverse transcription of the same RNase treated RNA with primers closer to the PBS (data not shown).

In the TAR and the Poly(A) stem–loop regions, the reverse transcription gave a relatively high background of non-specific terminations in the control reactions where no RNase was added. We therefore employed a different strategy to map this particular region. Instead of using the reverse transcription method to map the cleavage sites, we used 5′-end labelled HIV-1 RNA which could be visualized directly on a sequencing gel (Fig. 2C). The obtained cleavage pattern was compatible with the existence of the two stem–loops, TAR and Poly(A), but showed neither protections nor enhancements of RNase cleavage upon incubation with Gag and NC, indicating that neither Gag nor NC bind within this region under the conditions used. This is consistent with a band shift analysis using an RNA corresponding to positions 1–120 of the HIV-1 genome as substrate. No band shift was observed upon incubation with up to a 200-fold molar excess of neither GST-Gag nor GST-NC in the presence of competitor rRNA (data not shown).

HIV-1 Gag and NC interact with a loop in the HIV-2 PSI 3 loop

HIV-2 and SIV genomes are efficiently packaged *in vivo* by HIV-1 proteins (58,59). We therefore wanted to test the binding of GST-Gag and GST-NC from HIV-1 to the HIV-2 RNA leader *in vitro*. Band shift assays were performed with a 738 nt RNA spanning the entire 3′ UTR region and 83 nt of the Gag ORF and under the same conditions as used for the HIV-1 RNA. The HIV-1 Gag and NC proteins readily formed complexes with the HIV-2 RNA and based on the band shifts, binding affinities appeared similar to the ones seen with the HIV-1 RNA (data not shown). As for the HIV-1 RNA, about one to two shifted complexes were resistant to the presence of a 100-fold competitor rRNA.

RNA footprinting experiments with HIV-1 GST-Gag and GST-NC on the HIV-2 RNA were performed (Fig. 5A and B, respectively) and the results are summarized in Figure 6. The structural information from the cleavage pattern obtained in the absence of protein is in good agreement with the previously suggested structural model shown in Figure 6A (12). Both Gag and NC gave a similar protection pattern including a very strong protection in the PSI 3 loop at position U522-A525 (Fig. 5A and B, lanes 1–6; Fig. 6B and C) as well as at positions C515–U516 at the base of this stem. Less strong protections were seen in the PSI 1 stem–loop (A452–G454 and G448) for both Gag and NC. Both proteins protected positions U466–G467 in the SD stem but varied in their degree of protection of the SD loop. Moreover, only NC enhanced the cleavages in the PSI 2 loop (positions A500–A501). Thus, GST-Gag and GST-NC appear to share a

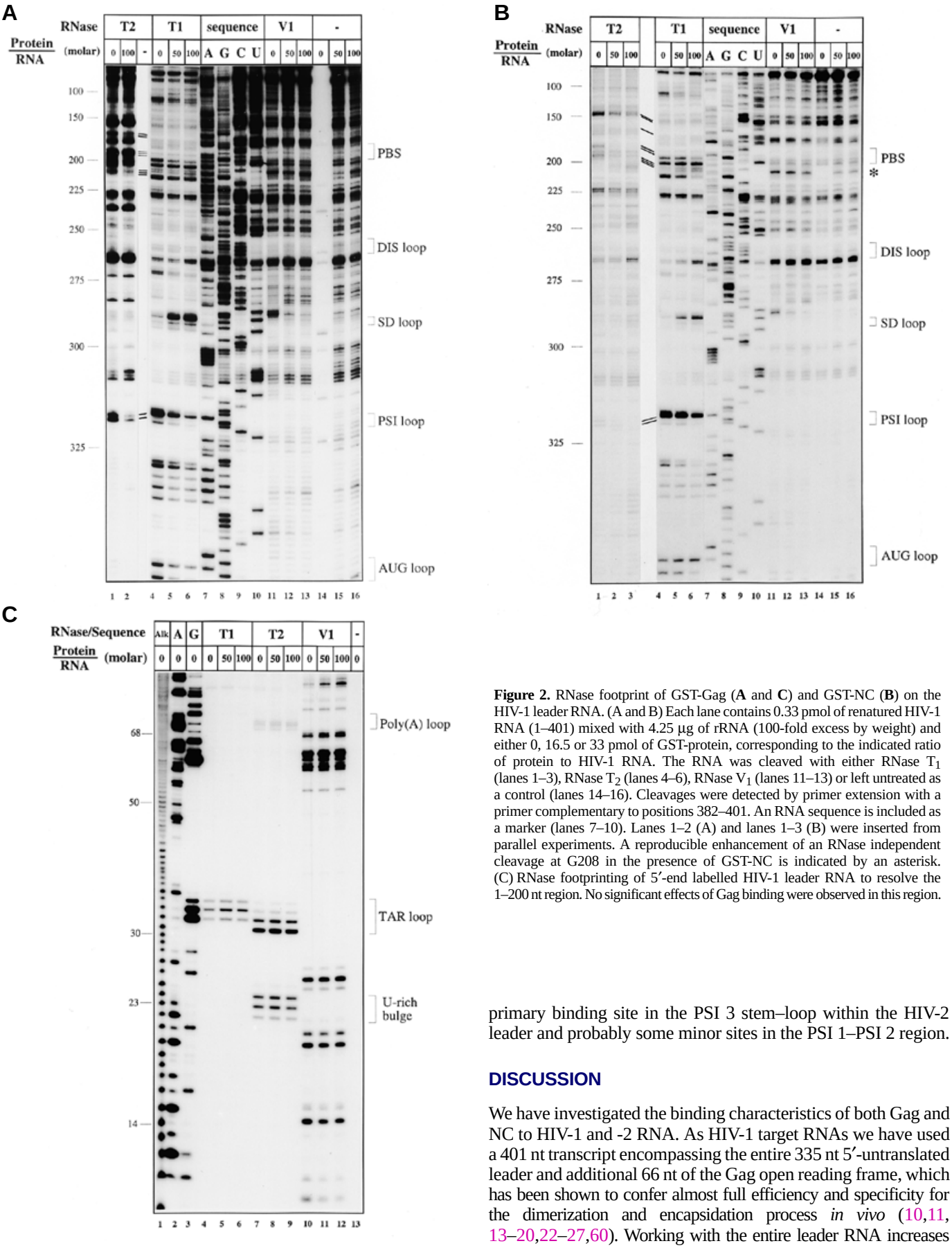

Figure 2. RNase footprint of GST-Gag (**A** and **C**) and GST-NC (**B**) on the HIV-1 leader RNA. (A and B) Each lane contains 0.33 pmol of renatured HIV-1 RNA (1–401) mixed with 4.25 µg of rRNA (100-fold excess by weight) and either 0, 16.5 or 33 pmol of GST-protein, corresponding to the indicated ratio of protein to HIV-1 RNA. The RNA was cleaved with either RNase T_1 (lanes 1–3), RNase T_2 (lanes 4–6), RNase V_1 (lanes 11–13) or left untreated as a control (lanes 14–16). Cleavages were detected by primer extension with a primer complementary to positions 382–401. An RNA sequence is included as a marker (lanes 7–10). Lanes 1–2 (A) and lanes 1–3 (B) were inserted from parallel experiments. A reproducible enhancement of an RNase independent cleavage at G208 in the presence of GST-NC is indicated by an asterisk. (C) RNase footprinting of 5′-end labelled HIV-1 leader RNA to resolve the 1–200 nt region. No significant effects of Gag binding were observed in this region.

primary binding site in the PSI 3 stem–loop within the HIV-2 leader and probably some minor sites in the PSI 1–PSI 2 region.

DISCUSSION

We have investigated the binding characteristics of both Gag and NC to HIV-1 and -2 RNA. As HIV-1 target RNAs we have used a 401 nt transcript encompassing the entire 335 nt 5′-untranslated leader and additional 66 nt of the Gag open reading frame, which has been shown to confer almost full efficiency and specificity for the dimerization and encapsidation process *in vivo* (10,11, 13–20,22–27,60). Working with the entire leader RNA increases

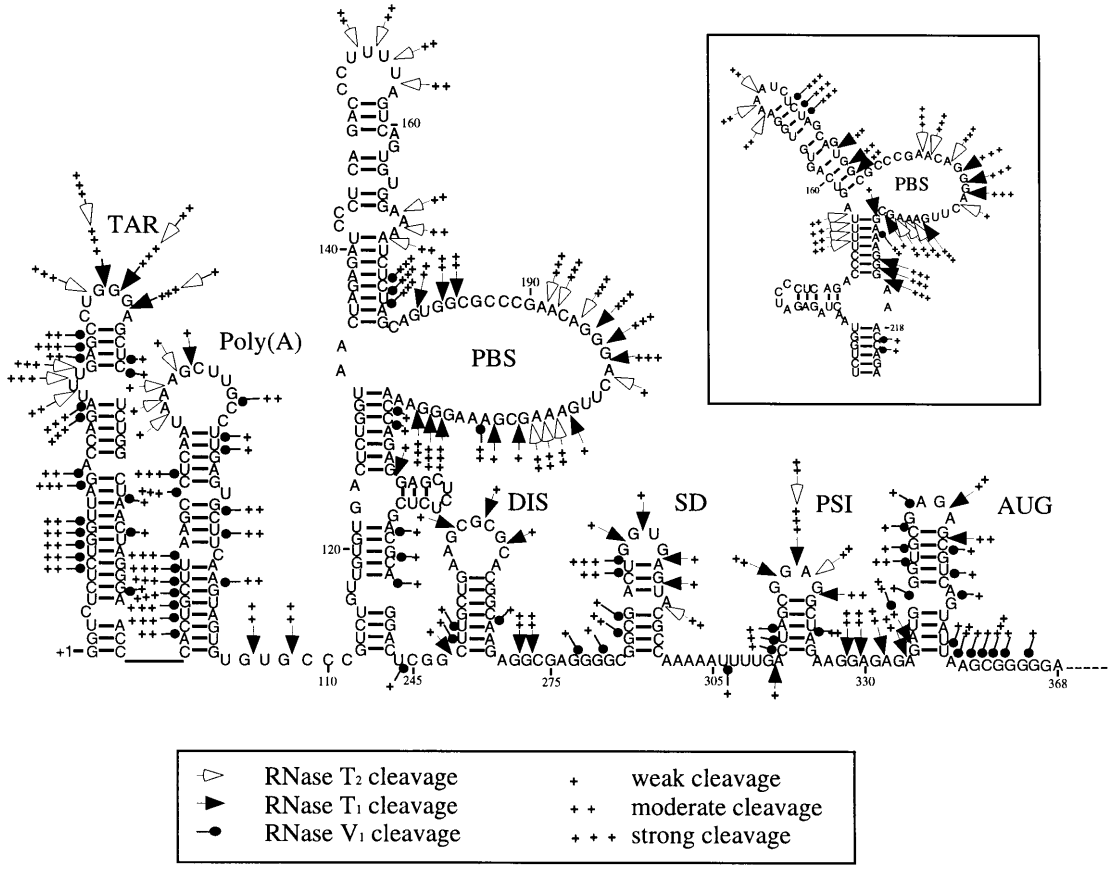

Figure 3. Summary of all cleavage data obtained from several experiments with HIV-1 RNA (1–401) in the absence of protein. The data in the TAR and Poly(A) stem–loops was obtained with end labelled RNA while data in the downstream region was obtained by use of end labelled primers complementary to positions 328–346 and 382–401. The secondary structure is drawn according to Berkhout (12) with a modification of the PBS stem–loop structure. The structure of the PBS region proposed by Berkhout is shown as an insert. Numbers are given relative to the start of transcription and the stems are named as follows: TAR, binding site for the transcriptional activator Tat; Poly(A), polyadenylation site in the 3′LTR; PBS, primer binding site; DIS, dimer initiation site; SD, major splice donor; PSI, packaging signal; AUG, Gag start codon.

the chances of correct folding of secondary structural elements and potential long distance tertiary interactions in the RNA.

The structure of the HIV-1 leader RNA was mapped using single and double strand specific RNases, and the result was largely consistent with the secondary structure model drawn in Figure 3 which is similar to the previously proposed model by Berkhout (12) but modified in the PBS region according to Rizvi *et al.* (7). Berkhout's model of the PBS stem–loop structure (shown as an insert in Fig. 3) is based on genetic data (61), but it is incompatible with our footprinting data. In particular, the strong T_1 and T_2 cleavages of G181–G184, G212–G214 and U153–U156 suggest that these regions are unpaired in the naked RNA. However, it is possible that NC induces a conformational change in this region which could account for the discrepancies of the structural models. The lower part of the PBS stem–loop structure is more speculative and lacks phylogenetic support. In the absence of sufficient data for the 5′-side of this stem we cannot fully evaluate the validity of the proximal part of the PBS stem in this study.

Another intriguing observation is the general low accessibility to single strand specific RNases, but not double strand specific RNases, of the inter-helical regions between the PBS, DIS, SD and PSI stems and downstream from the AUG stem. It indicates that some of the helical segments are more extended than drawn in the model (Fig. 3) and/or that co-helical stacking between adjacent helices may induce a tight overall structure. One possibility is that the DIS stem is more extended as proposed by Harrison and Lever (4). This view is reinforced by the observation that a DNA oligonucleotide complementary to the DIS-SD inter-helical region, which was intended as a primer for the reverse transcription in this study, annealed inefficiently to the viral RNA (data not shown). Potential base pairing partners to the G-rich region downstream of the AUG region include the inaccessible U-stretch between SD and PSI. The distal loops of the DIS and SD hairpins are also relatively inaccessible to enzymes. This is surprising considering that the DIS loop has been implicated in the formation of a kissing loop interaction in the dimer (20–24,62) and that the 5′-splice site within the SD loop is highly active in splicing (63). It is possible that the loops are buried in a higher order structure in the monomer and in the absence of splicing factors.

The formation of nucleoprotein complexes was studied by electrophoretic mobility shift and RNA footprinting assays using three different sources of protein: synthetic non-tag NC (data not shown), GST-tagged NC and GST-tagged Gag, and the results showed characteristic differences. Both GST-tagged versions of NC and Gag bound specifically to the RNA, forming multiple complexes, similar to what has been observed previously (32,33). The higher order complexes are not simply a result of RNA-independent oligomerization based on the observation that GST-NC

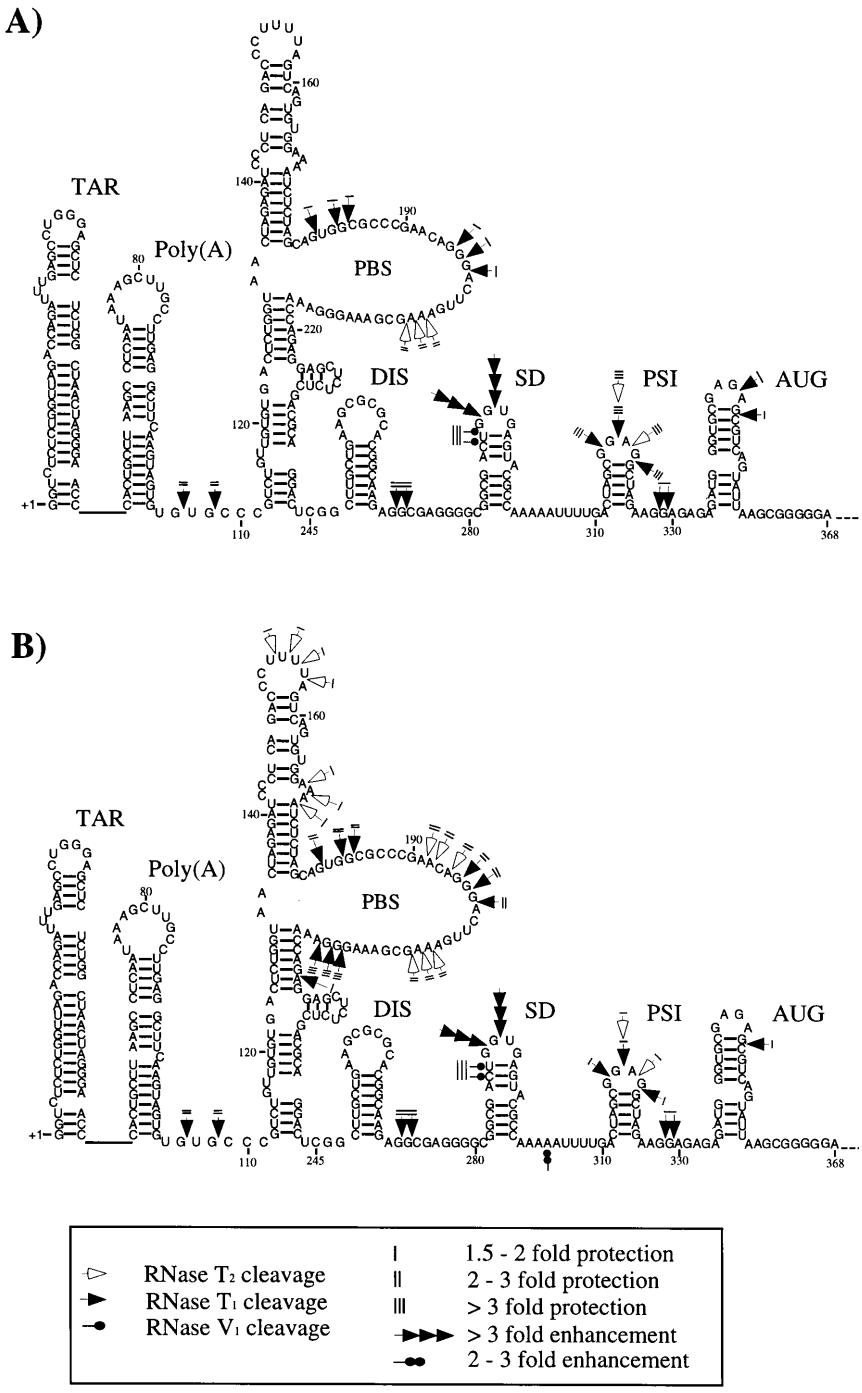

Figure 4. Summary of the protections and enhancements obtained for HIV-1 RNA (1–401) in the presence of GST-Gag (**A**) or GST-NC (**B**). The data are derived from at least three independent experiments for each primer.

binding at higher molar concentrations to short unrelated RNAs primarily forms single complexes (data not shown). Addition of a >100-fold excess of non-specific competitor reduced the amount of larger complexes and gave rise to 1–2 distinct bands that were relatively unaffected by the competitor, implying the existence of two high affinity binding sites. In contrast, non-tag synthetic NC formed only high molecular weight complexes that did not enter the gel (data not shown). The different binding characteristics observed between non-tag NC and GST-NC imply that the GST-tag does not interfere with the binding specificity of NC but merely inhibits the nucleation of the RNA into high molecular weight complexes. GST-NC did not increase the amount of the RNA dimer band as it has been demonstrated with non-tag NC protein by others (6,31,51), implying that the annealing capacity of GST-tagged NC is inefficient. It is possible that the GST-tag distorts the folding of the cluster of basic amino acids in the N-terminal of the NC, which has been shown to be the primary determinant for the annealing and dimerization activity of NC (31,45).

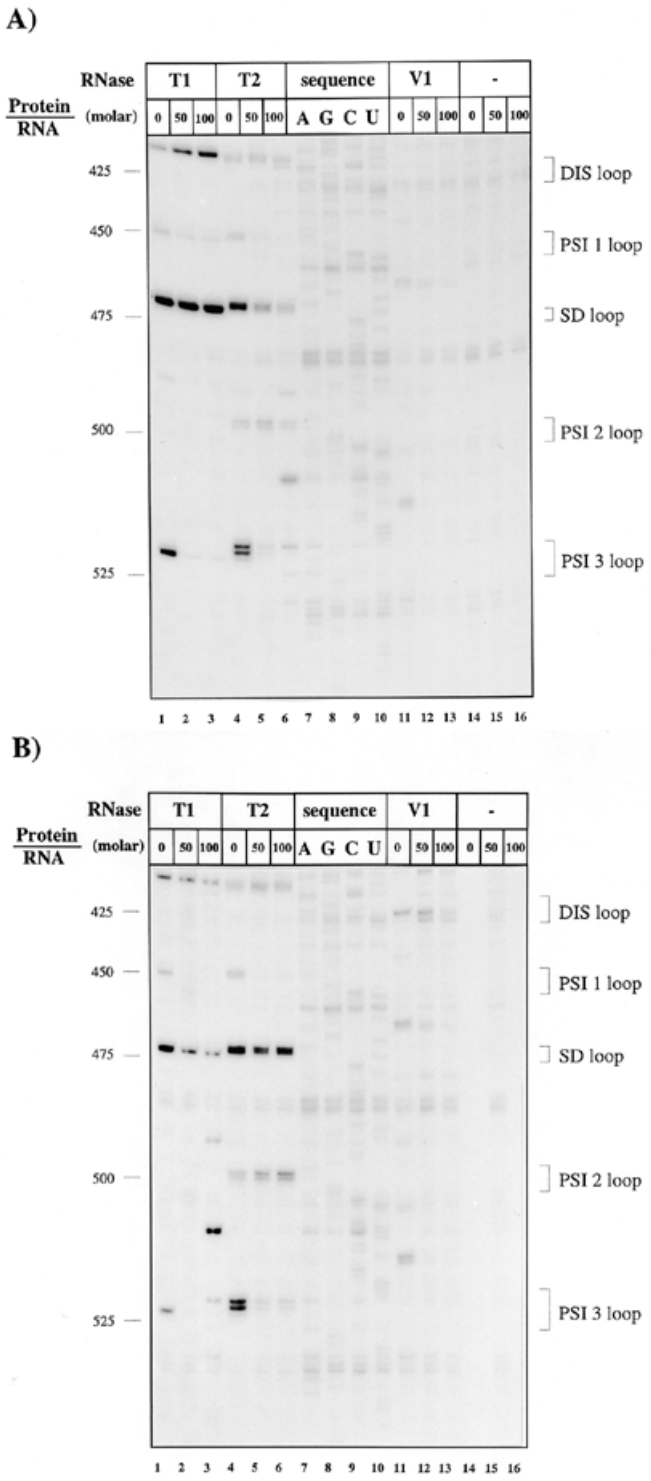

Figure 5. RNase footprint of GST-Gag (**A**) and GST-NC (**B**) on HIV-2 leader RNA (1–738). See legend to Figure 2 for experimentals and labels.

Mapping the discrete binding sites of GST-Gag and GST-NC

The complex formation between GST-Gag or GST-NC and viral RNA was investigated by RNA footprinting analysis under conditions where only the 1–2 high affinity complexes are detected in mobility shift analysis. The strongest effect by Gag was observed in the PSI loop, suggesting that this stem–loop represents a primary binding site for Gag. This is in good agreement with studies *in vivo*, where the PSI stem–loop has been shown to be important for efficient and specific packaging (10,11,13,14,18–20) and *in vitro*, where the NMR structure of the PSI stem–loop has been solved in complex with with the NC protein (64). Moreover, binding of Gag strongly increased the enzymatic cleavage of the SD loop accompanied by a protection against RNase V_1 cleavage of the adjacent base pairs. It is possible that Gag binding induces a conformational change exposing the SD loop to the solvent or disrupting the helical structure in this region.

The inhibition against T_1 cleavages is also observed in the spacer regions between the Poly(A) and PBS stems and between the DIS and SD stems. If the proposed structure in this region is correct, these sites are in close proximity and may form a potential binding site for Gag.

Binding of GST-NC to the RNA gave both similarities and pronounced differences to GST-Gag. The most striking difference was a reduced protection of the PSI loop accompanied by an increased protection in and near the PBS loop that harbours the PBS sequence to which the viral tRNA primer anneals to initiate reverse transcription. The lack of protections in the TAR and Poly(A) regions implies that these stem–loops do not harbour any specific Gag or NC binding sites. This is in agreement with a previous finding showing that an RNA spanning these two stem–loops bound to Gag with low affinity (9). However, the TAR and Poly(A) stem–loops have recently been reported to be important for efficient packaging (60,65). Taken together, these findings suggest that the TAR and the Poly(A) stem–loops might be important to other events which affects packaging efficiency. The DIS stem–loop, which is unaffected by the complex formation in regard to the RNase accessibility, has previously been shown to interact with GST-NCp15 (9). Due to the relatively low accessibility of the DIS stem–loop to RNase digestion, we cannot rule out that Gag and/or NC also bind this structure in our assay.

A general concern when performing RNA footprinting on a mixed population of monomer and dimer RNAs is that the protein may influence the level of dimerization. However, since the assay is performed under conditions at which we observe only 5–10% dimerization, and since this level is largely unaffected by GST-Gag and GST-NC, inter-molecular contacts are not likely to contribute significantly to the observed protections and enhancements.

HIV-1 GST-Gag and GST-NC bind to a homologous structure within the HIV-2 leader RNA

The HIV-2 leader also contains sequences important for genomic encapsidation (66,67) and HIV-2 and SIV genomic RNA is specifically packaged by the HIV-1 Gag protein (58,59). To characterize the interaction of HIV-1 Gag and NC proteins with HIV-2 RNA, we performed gel mobility shift analysis and protein footprinting using GST-Gag and GST-NC and an RNA spanning 738 nt of the HIV-2 leader. As was observed with the HIV-1 RNA, 2 distinct complexes were relatively resistant to a 100-fold excess of rRNA. Moreover, a high affinity binding site observed by RNA footprinting in the HIV-2 leader mapped to the stem–loop PSI-3 which has homology to the PSI stem–loop of the HIV-1 leader. In particular the conservation of nucleotides at the base of the stems ($_{312}$CUA–//–UAGAAG$_{329}$ in HIV-1 and $_{515}$CUA–//–UAGAAG$_{534}$ in HIV-2) suggest that these stem–loops are functionally equivalent. Previously published *in vitro* data on HIV-2 packaging

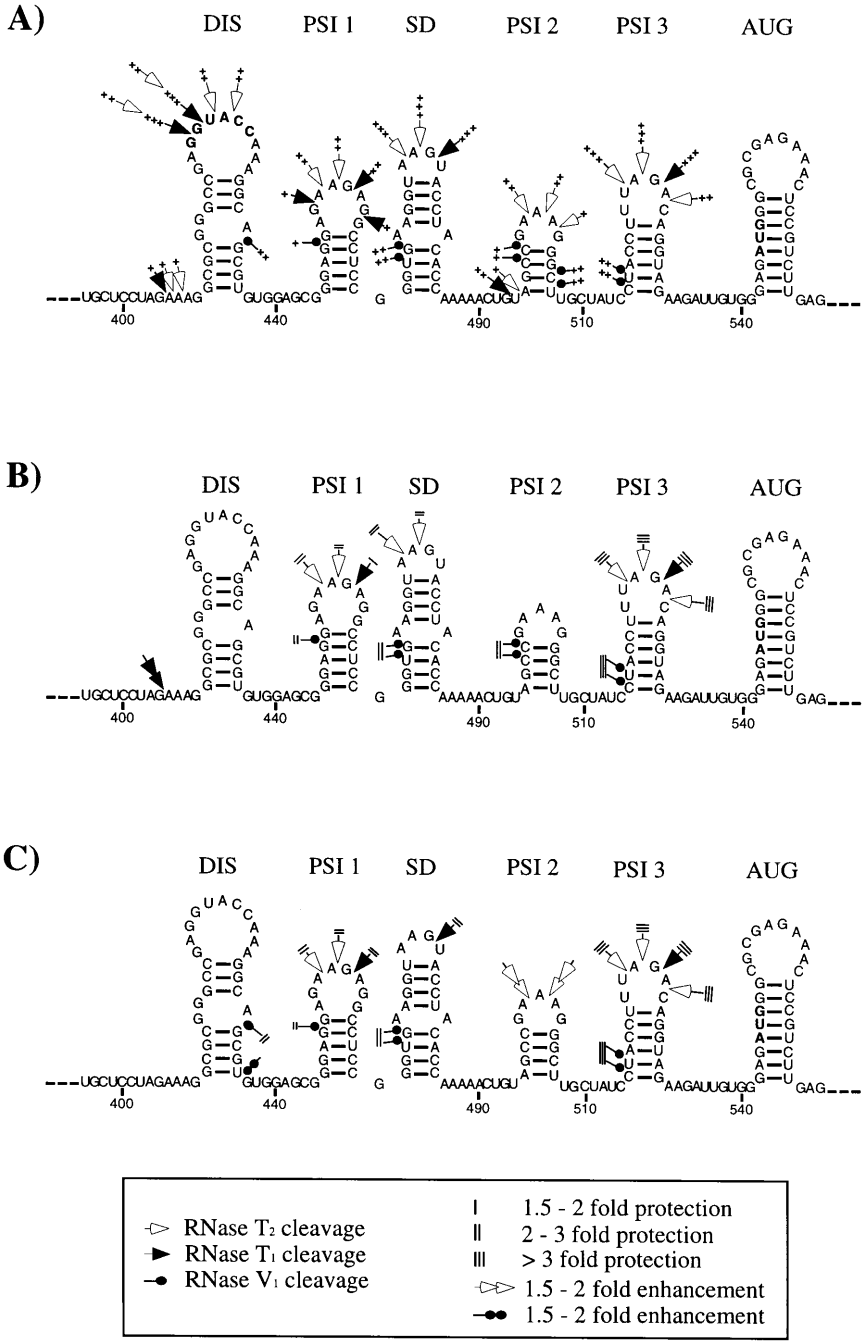

Figure 6. Summary of all cleavage data obtained for HIV-2 RNA (1–738) in the absence of protein (**A**) and of the protections and enhancements in the presence of GST-Gag (**B**) or GST-NC (**C**). The secondary structure is drawn according to Berkhout (12) and numbers are given relative to the start of transcription. Only the sequence from position U396 to G570 is shown. Abbreviations are as in Figure 2.

signals are relatively sparse. However, an *in vitro* binding study using UV-crosslinking of an HIV-2 RNA spanning this region to its cognate NC protein (NCp8) also supports the importance of the PSI-3 stem–loop (68).

Insights into the Gag/NC recognition mechanism

The strong protection by Gag and NC of G317–G320 in the PSI loop against enzymatic cleavages points to these residues as being major binding determinants. This finding was recently substantiated by an NMR analysis of HIV-1 NC bound to a PSI stem–loop RNA. It revealed that the N- and C-terminal CCHC-type zinc knuckles of the NC protein formed hydrogen bonds with the H1 and the O6 atoms in the heterocyclic rings of G318 and G320, respectively, and a conserved arginine in the basic domain between the zinc-knuckles interacts with A319 (64). The N-terminal residues L3–R10 adopt an α-helical structure which is positioned within a widened major groove of the stem primarily interacting non-specifically with the phosphate backbone of the RNA. The binding of multiple NC molecules to other parts of the

leader suggests that the RNA binding motif is rather flexible, probably partly because of the non-specific interactions between the basic residues of the α-helix and the RNA backbone. The NMR model also provides a possible explanation for the differences in binding characteristics of GST-NC and non-tag NC that we observe. It is possible that the interaction between the N-terminal α-helix of NC and the RNA is less favourable when situated in the context of an N-terminal GST-fusion, and that the inactivation of such a 'non-specific clamp' for RNA binding causes the discrete binding characteristic of GST-NC as compared to non-tag NC.

The differences in the binding pattern of NC and Gag that we observe may resonate distinct functional roles at different steps of the viral life cycle: the strong binding of the Gag precursor to the PSI may promote the selection of unspliced viral RNA in the packaging step, whereas the preference of NC for the primer binding region may reflect its proposed role in the annealing of the primer tRNA to the primer binding site (31,45,46).

ACKNOWLEDGEMENTS

We thank B.Berkhout for pUC8-HIV-2, S.P.Goff for GST-NC and GST-Gag plasmids and J.-L.Darlix for kindly sharing some synthetic NCp7 peptide with us. R.Rosendahl is thanked for excellent technical assistance. The work was supported in part by grants from the Danish Cancer Society, the Danish Research Council of Health, the Danish National Research Council of Science, the EU Biomed 2 program and the Karen Elise Jensen Foundation.

REFERENCES

1 Rein,A. (1994) *Arch. Virol.* Suppl., **9**, 513–522.
2 Berkowitz,R.D., Fischer,J. and Goff,S.P. (1996) *Curr. Top. Microbiol. Immunol.*, **214**, 177–218.
3 Muesing,M.A., Smith,D.H. and Capon,D.J. (1987) *Cell*, **48**, 691–701.
4 Harrison,G.P. and Lever,A.M. (1992) *J. Virol.*, **66**, 4144–4153.
5 Baudin,F., Marquet,R., Isel,C., Darlix,J.L., Ehresmann,B. and Ehresmann,C. (1993) *J. Mol. Biol.*, **229**, 382–397.
6 Sakaguchi,K., Zambrano,N., Baldwin,E.T., Shapiro,B.A., Erickson,J.W., Omichinski,J.G., Clore,G.M., Gronenborn,A.M. and Appella,E. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 5219–5223.
7 Rizvi,T.A. and Panganiban,A.T. (1993) *J. Virol.*, **67**, 2681–2688.
8 Isel,C., Ehresmann,C., Keith,G., Ehresmann,B. and Marquet,R. (1995) *J. Mol. Biol.*, **247**, 236–250.
9 Clever,J., Sassetti,C. and Parslow,T.G. (1995) *J. Virol.*, **69**, 2101–2109.
10 McBride,M.S. and Panganiban,A.T. (1996) *J. Virol.*, **70**, 2963–2973.
11 McBride,M.S., Schwartz,M.D. and Panganiban,A.T. (1997) *J. Virol.*, **71**, 4544–4554.
12 Berkhout,B. (1996) *Prog. Nucleic Acid Res. Mol. Biol.*, **54**, 1–34.
13 Lever,A., Gottlinger,H., Haseltine,W. and Sodroski,J. (1989) *J. Virol.*, **63**, 4085–4087.
14 Aldovini,A. and Young,R.A. (1990) *J. Virol.*, **64**, 1920–1926.
15 Clavel,F. and Orenstein,J.M. (1990) *J. Virol.*, **64**, 5230–5234.
16 Kim,H.J., Lee,K. and O'Rear,J.J. (1994) *Virology*, **198**, 336–340.
17 Kaye,J.F., Richardson,J.H. and Lever,A.M. (1995) *J. Virol.*, **69**, 6588–6592.
18 Hayashi,T., Shioda,T., Iwakura,Y. and Shibuta,H. (1992) *Virology*, **188**, 590–599.
19 Luban,J. and Goff,S.P. (1994) *J. Virol.*, **68**, 3784–3793.
20 Clever,J.L. and Parslow,T.G. (1997) *J. Virol.*, **71**, 3407–3414.
21 Skripkin,E., Paillart,J.C., Marquet,R., Ehresmann,B. and Ehresmann,C. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 4945–4949.
22 Laughrea,M., Jette,L., Mak,J., Kleiman,L., Liang,C. and Wainberg,M.A. (1997) *J. Virol.*, **71**, 3397–3406.
23 Berkhout,B. and van Wamel,J.L. (1996) *J. Virol.*, **70**, 6723–6732.
24 Paillart,J.C., Skripkin,E., Ehresmann,B., Ehresmann,C. and Marquet,R. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 5572–5577.
25 Buchschacher,G.J. and Panganiban,A.T. (1992) *J. Virol.*, **66**, 2731–2739.
26 Parolin,C., Dorfman,T., Palu,G., Gottlinger,H. and Sodroski,J. (1994) *J. Virol.*, **68**, 3888–3895.
27 Vicenzi,E., Dimitrov,D.S., Engelman,A., Migone,T.S., Purcell,D.F., Leonard,J., Englund,G. and Martin,M.A. (1994) *J. Virol.*, **68**, 7879–7890.
28 Darlix,J.L., Lapadat Tapolsky,M., de Rocquigny,H. and Roques,B.P. (1995) *J. Mol. Biol.*, **254**, 523–537.
29 Lapadat,T.M., De,R.H., Van,G.D., Roques,B., Plasterk,R. and Darlix,J.L. (1993) *Nucleic Acids Res.*, **21**, 831–839.
30 Khan,R. and Giedroc,D.P. (1994) *J. Biol. Chem.*, **269**, 22538–22546.
31 De Rocquigny,H., Gabus,C., Vincent,A., Fournie Zaluski,M.C., Roques,B. and Darlix,J.L. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6472–6476.
32 Berkowitz,R.D., Luban,J. and Goff,S.P. (1993) *J. Virol.*, **67**, 7190–7200.
33 Berkowitz,R.D. and Goff,S.P. (1994) *Virology*, **202**, 233–246.
34 Dorfman,T., Luban,J., Goff,S.P., Haseltine,W.A. and Gottlinger,H.G. (1993) *J. Virol.*, **67**, 6159–6169.
35 Dannull,J., Surovoy,A., Jung,G. and Moelling,K. (1994) *EMBO J.*, **13**, 1525–1533.
36 Schmalzbauer,E., Strack,B., Dannull,J., Guehmann,S. and Moelling,K. (1996) *J. Virol.*, **70**, 771–777.
37 Surovoy,A., Dannull,J., Moelling,K. and Jung,G. (1993) *J. Mol. Biol.*, **229**, 94–104.
38 Gorelick,R.J., Nigida,S.J., Bess,J.J., Arthur,L.O., Henderson,L.E. and Rein,A. (1990) *J. Virol.*, **64**, 3207–3211.
39 Gorelick,R.J., Chabot,D.J., Rein,A., Henderson,L.E. and Arthur,L.O. (1993) *J. Virol.*, **67**, 4027–4036.
40 Kaye,J.F. and Lever,A.M. (1996) *J. Virol.*, **70**, 880–886.
41 Poon,D.T.K., Wu,J. and Aldovini,A. (1996) *J. Virol.*, **70**, 6607–6616.
42 Berkowitz,R.D., Ohagen,A., Hoglund,S. and Goff,S.P. (1995) *J. Virol.*, **69**, 6445–6456.
43 Zhang,Y. and Barklis,E. (1995) *J. Virol.*, **69**, 5716–5722.
44 Tsuchihashi,Z. and Brown,P.O. (1994) *J. Virol.*, **68**, 5863–5870.
45 Lapadat Tapolsky,M., Pernelle,C., Borie,C. and Darlix,J.L. (1995) *Nucleic Acids Res.*, **23**, 2434–2441.
46 Barat,C., Lullien,V., Schatz,O., Keith,G., Nugeyre,M.T., Gruninger,L.F., Barre,S.F., LeGrice,S.F. and Darlix,J.L. (1989) *EMBO J.*, **8**, 3279–3285.
47 Weiss,S., Konig,B., Morikawa,Y. and Jones,I. (1992) *Gene*, **121**, 203–212.
48 Peliska,J.A., Balasubramanian,S., Giedroc,D.P. and Benkovic,S.J. (1994) *Biochemistry*, **33**, 13817–13823.
49 Ji,X., Klarmann,G.J. and Preston,B.D. (1996) *Biochemistry*, **35**, 132–143.
50 Li,X., Quan,Y., Arts,E.J., Zhuo,L., Preston,B.D., De Rocquigny,H., Roques,B.P., Darlix,J.-L., Kleiman,L., Parniak,M. and Wainberg,M.A. (1996) *J. Virol.*, **70**, 4996–5004.
51 Darlix,J.L., Gabus,C., Nugeyre,M.T., Clavel,F. and Barre,S.F. (1990) *J. Mol. Biol.*, **216**, 689–699.
52 Luban,J. and Goff,S.P. (1991) *J. Virol.*, **65**, 3203–3212.
53 Berglund,J.A., Charpentier,B. and Rosbash,M. (1997) *Nucleic Acids Res.*, **25**, 1042–1049.
54 Jensen,T.H., Jensen,A. and Kjems,J. (1995) *Gene*, **162**, 235–237.
55 De,R.H., Gabus,C., Vincent,A., Fournie,Z.M., Roques,B. and Darlix,J.L. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6472–6476.
56 Oude Essink,B.B., Das,A.T. and Berkhout,B. (1996) *J. Mol. Biol.*, **264**, 243–254.
57 Kjems,J., Calnan,B.J., Frankel,A.D. and Sharp,P.A. (1992) *EMBO J.*, **11**, 1119–1129.
58 Kaye,J.F. and Lever,A.M.L. (1998) *J. Virol.*, **72**, 5877–5885.
59 Rizvi,T.A. and Panganiban,A.T. (1993) *J. Virol.*, **67**, 2681–2688.
60 McBride,M.S. and Panganiban,A.T. (1997) *J. Virol.*, **71**, 2050–2058.
61 Berkhout,B. (1997) *Nucleic Acids Res.*, **25**, 4013–4017.
62 Clever,J.L., Wong,M.L. and Parslow,T.G. (1996) *J. Virol.*, **70**, 5902–5908.
63 O'Reilly,M.M., McNally,M. and Beemon,K.L. (1995) *Virology*, **213**, 373–385.
64 De Guzman,R.N., Wu,Z.R., Stalling,C.C., Pappalardo,L., Borer,P.N. and Summers,M.F. (1998) *Science*, **279**, 384–388.
65 Das,A.T., Klaver,B., Klasens,B.I.F., Wamel,J.L.B.V. and Berkhout,B. (1998) *J. Virol.*, **71**, 2346–2356.
66 Garzino Demo,A., Gallo,R.C. and Arya,S.K. (1995) *Hum. Gene Ther.*, **6**, 177–184.
67 McCann,E.M. and Lever,A.M. (1997) *J. Virol.*, **71**, 4133–4137.
68 Tsukahara,T., Komatsu,H., Kubo,M., Obata,F. and Tozawa,H. (1996) *Biochem. Mol. Biol. Int.*, **40**, 33–42.