Biochemical and NMR Study on the Competition between Proteins SC35, SRp40, and Heterogeneous Nuclear Ribonucleoprotein A1 at the HIV-1 Tat Exon 2 Splicing Site

Houda Hallay1, Nicolas Locker1,2, Lilia Ayadi1, Delphine Ropers1, Eric Guittet1, and Christiane Branlant1,3

From the 1Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-Université Henri Poincaré-Nancy I, Boulevard des Aiguillettes, BP239, 54506 Vandoeuvre-lès-Nancy Cedex and 2Laboratoire de Chimie et Biologie Structurales, ICSN-CNRS, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette, France

The human immunodeficiency virus, type 1, Tat protein plays a key role in virus multiplication. Because of its apoptotic property, its production is highly controlled. It depends upon the A3 splicing site utilization. A key control of site A3 activity is the ESS2 splicing silencer, which is located within the long stem-loop structure 3 (SLS3), far downstream from site A3. Here, by enzymatic footprints, we demonstrate the presence of several heterogeneous nuclear ribonucleoprotein (hnRNP) A1-binding sites on SLS3 and show the importance of the C-terminal Gly domain of hnRNP A1 in the formation of stable complexes containing several hnRNP A1 molecules bound on SLS3. Mutations in each of the UAG triplets in ESS2 strongly reduce the overall hnRNP A1 binding, showing the central role of ESS2 in hnRNP A1 assembly on SLS2-SLS3. Using NMR spectroscopy, we demonstrate the direct interaction of ESS2 with the RNA recognition motifs domains of hnRNP A1. This interaction has limited effect on the RNA two-dimensional structure. The SR proteins SC35 and SRp40 were found previously to be strong activators of site A3 utilization. By enzymatic and chemical footprints, we delineate their respective binding sites on SLS2 and SLS3 and find a strong similarity between the hnRNP A1-, SC35-, and SRp40-binding sites. The strongest SC35-binding site only has a modest contribution to site A3 activation. Hence, the main role of SR proteins at site A3 is to counteract hnRNP A1 binding on ESS2 and ESE2. Indeed, we found that ESE2 has inhibitory properties because of its ability to bind hnRNP A1.

Retrovirus RNAs are transcribed from an integrated proviral genome. Part of the transcripts has to be transported to the cytoplasm in an unspliced form and is used as mRNAs for the production of the viral, Gag, Gag-Pol, and Env protein precursors and as genomic RNAs for new virion assembly. Another part of the transcripts has to be spliced for production of the Tat, Rev, Nef, Vif, and Vpr proteins. Because of the presence of four splicing donor sites (5’ ss) and eight splicing acceptor sites (3’ ss) in HIV-1 RNA, the splicing machinery of infected cells generates at least 40 distinct mRNAs from a unique RNA primary transcript. The relative abundance of these mRNAs depends greatly on the relative efficiencies of the 3’ ss, which are suboptimal (2–11). During the early phase of cell infection, the five 3’ ss (A3, A4c, A4a, A4b, and A5) located in a small central part of the viral RNA are used for production of the tat, rev, and nef mRNAs (1). All the tat mRNAs are spliced at site A3. The rev mRNAs are spliced at sites A4a, A4b, or A4c, and most of the nef mRNAs are spliced at site A5 (1,12).

The Tat protein plays a key role in virus multiplication as it is needed for production of full-length HIV-1 transcripts (13). However, because of the apoptotic activity of this protein on both the infected cells and the neighboring cells (13–15), virus HIV-1 strongly controls its production. In both lymphoid and nonlymphoid infected cells, the steady-state level of the doubly spliced tat mRNAs is considerably lower than levels of doubly spliced rev mRNAs and singly spliced env/vpu mRNAs (1). This seems to be due to the poor efficiency of the A3 splicing site compared with the other downstream 3’ ss (2, 3, 9, 16). Surprisingly, despite its low efficiency, site A3 has the most optimized polypyrrimidine tracts compared with the competitor sites (8,9). One explanation for this apparent discrepancy is the presence of both an upstream (ESS2p) (17) and a downstream (ESS2) exon-splicing silencer acting on site A3 (9). The proximal ESS2p element binds protein hnRNPH generating a steric hindrance at site A3 (17). In contrast, ESS2 is located far downstream from site A3 (69 nucleotides) (2,9). It inhibits an early step of spliceosome assembly (3), and binding of proteins hnRNP A/B is responsible for the inhibition (18). However, the
hnRNP A1 and SC35 Binding on HIV-1 Tat Exon 2

mechanism through which the binding of protein hnRNP A1 on ESS2 can limit the utilization of site A3, which is located 69 nucleotides upstream, remains unknown. This long distance activity may be linked to the peculiar secondary structure of the HIV-1 RNA region containing ESS2. Indeed, ESS2 is located in a long irregular stem-loop structure (SLS3) highly conserved in HIV-1 isolates (8). As protein hnRNP A1 contains a Gly domain allowing its oligomerization (19), the presence of several hnRNP A1-binding sites in SLS3 may favor its oligomerization along the RNA. Such findings would explain why, when inserted in a foreign context, at a distance from the 3’ ss similar to the one in the authentic HIV-1 context, the ESS2-dependent inhibition was reduced as compared with the authentic one (9).

The ESS2 context effect in HIV-1 RNA might also be due to the presence of neighboring SR-binding sites, which are required for the activity of site A3. They may be buried upon protein hnRNP A/B binding on ESS2 (20). Indeed, an overexpression of the SR proteins SC35 or SRp40 in HeLa cells and in infected T lymphocytes strongly increases site A3 utilization at the expense of site A1, A2, A4a–c, and A5 utilizations (21, 22). A steric hindrance hypothesis limiting the access of SR protein is as much realistic, as an SC35-binding site (ESE2) was recently identified in the vicinity of ESS2 (20). Enzymatic footprints were used to try to identify the protein hnRNP A1 and SC35-binding sites in a short HIV-1 RNA fragment containing the ESS2 element (20). However, as the hnRNP A1 protein used in these assays was a GST-hnRNP A1 fusion protein, the hnRNP A1-binding sites were not precisely delineated. In addition, the short RNA fragment used in the experiment could not form the SLS3 structure. Hence, the importance of the RNA secondary structure on the binding of proteins hnRNP A1 and SC35 was not investigated. Finally, binding of the two proteins was only studied on a short portion of the site A3 downstream region.

Therefore, for a better understanding of how proteins hnRNP A1, SC35, and SRp40 modulate site A3 utilization, it was of high importance to identify how these proteins interact with the two stem-loop structures SLS2 and SLS3 of HIV-1 RNA. The SLS2 stem-loop contains the A3 splicing site and the ESS2p inhibitor (17); the downstream SLS3 stem-loop contains the ESS2 inhibitor and the ESE2 activator (2, 20). It was also interesting to compare the binding property on SLS3 of the entire hnRNP A1 protein to that of its N-terminal domain (protein UP1). Indeed, both the Gly and UP1 domains of protein hnRNP A1 were proposed to contribute to the RNA-binding property (23). However, the UP1 domain, containing the two RNA recognition motifs (RRM), is expected to be responsible for the RNA binding specificity (23). The Gly domain would account for both protein multimerization and a reinforcement of the RNA-protein interaction (19, 23).

Here, we used enzymatic and chemical footprints to delineate the binding sites of proteins UP1, hnRNP A1, SC35, and SRp40 on SLS2 and SLS3. Gel-shift assay experiments were performed to compare the affinities of the identified protein-binding sites, and their influences on splicing efficiency were tested in vitro after site-directed mutagenesis. In addition, NMR approaches were used to identify the nucleotides of ESS2 contacting the UP1 domain. This is the first NMR study of a UP1-RNA complex; only the three-dimensional structure of a UP1-DNA complex had been studied previously (24). Altogether, the data presented demonstrate the presence of several binding sites for proteins hnRNP A1, SC35, and SRp40 on SLS3 and the SLS2-SLS3 linking region and the existence of strong overlapping between the binding sites of these three proteins. The 5’ strand of the internal loop II in SLS3 is characterized as the strongest SC35-binding site. However, despite the close distance between loop II and site A3, binding of protein SC35 to this loop is found to have only a modest activation effect on splicing at site A3. Our data demonstrate a crucial role of ESS2 in the stable association of several hnRNP A1 molecules on SLS3. Both the UP1 and Gly domains of hnRNP A1 play a role in the formation of this multimolecular complex, and formation of this complex can explain the strong inhibitory property of ESS2 compared with the other hnRNP A1-binding sites detected in SLS3. Initiation by ESS2 of the assembly of a stable complex containing several hnRNP A1 molecules also explains the high importance of the SC35- and SRp40-binding sites that overlap ESS2, in the mechanism of activation of site A3. Based on the present data, a model explaining the long distance effect of ESS2 and the specific activation effects of proteins SC35 and SRp40 is proposed.

EXPERIMENTAL PROCEDURES

Plasmid Construction—The previously described pLD-L3-U1 plasmid (17) was used for WT L3-U1 RNA production. This RNA contains the two HIV-1 BRU regions extending from positions 1 to 384 and 5172 to 5637 fused together. A first series of pLD-L3-U1 plasmid derivatives carrying one or two deletions (ΔESS2p, Δ5’loopII, ΔESS2, or Δ3’loopII) in the SLS2 and/or SLS3 coding region, and a second series of pLD-L3-U1 plasmid derivatives with dinucleotide substitutions in the sequence coding the RNA segment from position 5420 to 5447 were produced by site-directed mutagenesis, using the PCR approach (see the sequences of the oligonucleotides used in supplemental Table 1). PCR-amplified DNA fragments were obtained in conditions described previously (17), using the WT pLD-L3-U1 plasmid or one of its derivatives as the template and one of the oligonucleotide pairs described in supplemental Table 1 as primers.

RNA Production—For enzymatic and chemical footprinting assays, the A3M WT or variant RNAs were produced by SP6 RNA polymerase transcription using PCR-amplified DNA fragments from a WT or variant pLD-L3-U1 plasmid as the matrix and the O-Sp6 and O-548 oligonucleotides as the primers (supplemental Table 1). SP6 in vitro transcriptions were performed as described previously (8). For gel-shift experiments, radioactively labeled RNA transcripts corresponding to the WT or variant HIV-1 RNA regions from position 5339 to 5515 or from position 5367 to 5466, respectively, were transcribed in vitro with the SP6 RNA polymerase, in conditions described previously (8). PCR-amplified DNAs (primers pairs O-SP6 and 548 or 1989 and 2658, see supplemental Table 1) were used as the matrices. For splicing assays, uniformly labeled capped transcripts were synthesized by T7 RNA polymerase in conditions described previously (17). WT and variant pLD-L3-U1 plasmids (Δ5’loopII, ΔESS2, ΔESS2p, ΔESS2pΔ5’loopII, ΔESS2Δ5’loopII, and the 48–74 series) were cleaved by the Pst I nuclease prior to transcription. For NMR analysis, RNAs...
(NLN1, NLN2, NLN3, NLN4, NLN7, NLN8, or NLN9) were transcribed in vitro with the T7 RNA polymerase, using a DNA template containing a 2’ OMe-G at position 2, as described previously (25). After polyacrylamide gel purification, electrodialysis, and ethanol precipitation, the RNA was dialyzed against the NMR buffer (10 mM sodium phosphate buffer, pH 6.5) and was refolded by a 2-min heating at 95°C, followed by a slow cooling at room temperature.

Proteins Production and Purification—hnRNP A1 protein without any additional tag sequence was produced and purified as described previously (10), by using plasmid pET11a-hnRNP A1. A gene coding the UP1 domain was generated by insertion of three successive stop codons (positions 674, 677, and 680) in the hnRNP A1 open reading frame of a recombinant pGEX plasmid encoding a GST-hnRNP A1 fusion protein (generous gift of A. Mereau). Escherichia coli BL21 cells were then used for production of the recombinant UP1 protein. Details on its production and purification are given in the Supplemental Material.

For SR protein assays, we used recombinant SR proteins SC35 and SRp40 expressed in baculovirus, as described previously (26) (generous gift of Dr J. Stevenin, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg, France).

RNA and RNP Enzymatic and Chemical Footprinting Assays—The A3M WT or variants RNA transcribed by the SP6 RNA polymerase were refolded by a 5-min incubation at 65°C, followed by a slow cooling at room temperature. The RNP complexes were formed in 3.6 μl of buffer D (0.2 mM EDTA, 0.1 mM KCl, 0.5 mM dithiothreitol, 0.25 mM phenylmethylsulfonyl fluoride, 20% glycerol, 20 mM HEPES, pH 7.9) containing 2.5 mM magnesium chloride, as described previously (10). The [protein]/[RNA] molar ratios used ranged from 2 to 100 for assays with hnRNP A1, UP1, SC35, and SRp40. About 3.4 pmol of the A3M WT or mutated RNA were used for each enzymatic assay, and the RNA concentration was fixed to 230 ng/μl by addition of a yeast tRNA mixture. T1 (Roche Applied Science) and T2 (Invitrogen) RNase digestions were performed with 2.5 units/μg of RNA. V1 RNase (Kemotex) digestions were made with 2.5 × 10⁻³ units/μg of RNA. Incubations were carried out for 10 min at 30°C and stopped as described previously (8). Kethoxal modifications were made as described previously (8). For comparison, the free RNA was digested (T1, T2, and V1 RNases) or modified (Kethoxal) in the same conditions. Positions of enzymatic cleavages and chemical modifications were identified by extension of the 5’-end-labeled O-548 primer (supplemental Table 1), by using the avian myeloblastosis virus-reverse transcriptase, as described previously (8).

NMR Spectroscopy—For NMR experiments in H₂O, RNAs were dried and dissolved in 5% D₂O, 95% H₂O. Experiments involving nonexchangeable resonances were performed on RNAs dissolved in 100% D₂O. RNA sample concentrations were kept below 1 mM to avoid significant duplex formation. Formation of unique species for each transcript was assayed by native electrophoresis. ¹H-¹H and ¹H-¹⁵N NMR spectra were recorded on a Bruker DRX-600 with a z-gradient TXI cryoprobe or on a Bruker DRX-500 with an xyz-gradient TXI probe. The one- and two-dimensional spectra in 5% D₂O, 95% H₂O were recorded at 277 K for assignment. The two-dimensional NOESY spectra were recorded with 50- or 300-ms mixing times. A combination of Watergate and Jump and Return sequences was used for water suppression (27). RNAs were submitted to temperature titration from 277 to 308 K, and NOESY experiments used for assignment were recorded at 277 K. ¹⁵N HSQC spectra were acquired on ¹⁵N-labeled NLN8 samples at 293 K in NMR buffer to confirm the assignments and in buffer D lacking glycerol to map the interactions with UP1 as described later. To confirm the assignment and directly monitor Watson-Crick base pair formation, HNN-COSY experiments were recorded at 288 K on ¹⁵N-labeled NLN8 (28). ¹⁵N HSQC recorded on ¹⁵N-labeled UP1 in buffer D lacking glycerol were recorded at 293 K and titrations with NLN8 RNAs were performed as described later. Proton and nitrogen chemical shifts were calibrated against trimethylsilylpropionic acid. The two-dimensional spectra were collected in a phase-sensitive mode via the STATES-TPPI method and processed accordingly using the GIFA software (29).

RNA NMR Assignment and Analysis—Standard methods were used to deduce the secondary structure and global folding from the assignment of the exchangeable protons (30). Initial assignments were based on standard NOE walk analysis. The base-paired NH resonances from the predicted Watson-Crick pairs and those corresponding to unpaired but protected residues were assigned by analyzing NOE spectra recorded in H₂O. The well-established patterns of NOE and chemical shifts predicted for Watson-Crick pairs and the strong NOEs involving the exchangeable resonances provided unambiguous assignments for all slowly exchanging NH resonances. Formation of Watson-Crick base pair was further evidenced from HNN-COSY experiments yielding to observation of a direct coupling across the hydrogen bonding. Based on NOEY correlations between imino protons and between amino and imino protons, which are showing whether a nucleotide is base-paired or not, the folding of RNAs was deduced from NMR experiments in H₂O and confirmed by preliminary assignment of aromatic to anomic sequential NOEs when needed.

NMR Titration Experiments—For UP1 protein titration by NLN8 RNA, 100 μl of ¹⁵N UP1 at a 0.075 mM concentration in buffer D lacking glycerol was used. Two molar equivalents of RNA were lyophilized and dissolved in 100 μl of the same buffer. For titration, aliquots of this solution were added stepwise to the labeled UP1 protein, and ¹⁵N HSQCs were recorded at 293 K. The [RNA]/[UP1] ratios used were of 0:1, 1:1, 1.5:1, and 2:1. For NLN8 RNA titration by UP1 protein, the same strategy was used except that in addition to ¹⁵N HSQC, HNN-COSY experiments were recorded at 288 K to assay directly the base pairing of NLN8 upon UP1 binding.

Electrophoretic Mobility Shift Assays—Electrophoretic mobility shift assays were performed on 3.4 pmol of cold RNA mixed with 25 fmol of ³²P uniformly labeled transcript or on only 25 fmol of ³²P uniformly labeled transcript. RNP complexes were formed in buffer D by 15 min of incubation on ice (8). The [protein]/[RNA] molar ratios used ranged from 0 to 640. For competition experiments with cold RNAs, the 15 min of incubation of the labeled RNA with protein hnRNP A1 or SC35 was followed by a 15-min incubation with cold competitor RNA (0–80 pmol). The free RNA and RNP complexes were fractionated by electrophoresis on 6% polyacrylamide gels (38:2).
Several hnRNP A1-binding Sites in SLS3—Protein hnRNP A1 frequently binds cooperatively to several close binding sites (10, 31–33). For a better understanding of how it modulates site A3 utilization, we first delineated its binding sites on the previously identified (8) SLS2 and SLS3 stem-loop structures (Fig. 1). To this end, enzymatic footprinting experiments were performed on complexes formed between protein hnRNP A1 and the A3M RNA containing these two stem-loops (Fig. 1). According to gel-shift assays, a protein hnRNP A1 molar excess of 5 (protein concentration of 0.8 μM) compared with A3M RNA was sufficient for total conversion of the free A3M RNA into bound RNA (supplemental Fig. S1A). Upon increasing the [protein]/[RNA] molar ratio above 5, several hnRNP A1 molecules were bound to the RNA, as evidenced by the continuous step by step decrease of the electrophoretic mobility of the RNP complex (supplemental Fig. S1A). Thus, for evaluation of the relative affinities of protein hnRNP A1 for its various binding sites, the enzymatic probing of the A3M RNA accessibility was performed on complexes formed at [protein]/[RNA] molar ratios ranging from 2 to 10. T1 and T2 RNases were used in conditions defined for a specific cleavage of single-stranded accessible regions (8). The V1 RNase, which cleaves specifically the double-stranded and stacked RNA regions, was also used. At [protein]/[RNA] ratios ranging from 2 to 5 (Fig. 2, A, C, and D), protections were observed in ESS2, the terminal loop of the B motif, both strands of the internal loop II, and the 3′ strand of stem c. At a [protein]/[RNA] ratio of 10, the level of protection was strongly increased, and a new protection site was detected in the single-stranded segment linking SLS3 to SLS2 (Fig. 2, A and D).

FIGURE 1. Organization of the HIV-1 genome and the HIV-1 RNA fragments used in this study. The 5′ ss(D1 to D4) and the 3′ ss(A1 to A7) are indicated. Boxes represent the open reading frames. The location within the HIV-1 genome of the two HIV-1 segments present in the L3-U1 WT RNA as well as the HIV-1 RNA regions corresponding to the A3M and A3S RNAs are shown. Positions of the splicing regulatory elements (ESS2p, ESE2, and ESS2) (9, 17, 20) and the SLS2 and SLS3 stem-loop structures (8) are indicated.
SLS2-SLS3 HIV-1 RNA region. (ii) The two RRM s of protein hnRNP A1 can bind ESS2. Thus, we decided to perform a detailed study of the UP1-ESS2 interaction by NMR spectroscopy. Indeed, hnRNP A1 was not the appropriate molecule for an NMR spectroscopy analysis because of its large size and its numerous binding sites on SLS3 protein. The SLS3 stem-loop structure was also not suitable for NMR structural analysis because of its length of 130 nucleotides. Therefore, we first had to define a shorter RNA containing a correctly folded ESS2 element.

A Partial Structuration of ESS2 in SLS3 RNAs Missing Stems a−d—To focus the study on the UP1-ESS2 interaction, we had to eliminate the internal loop II, which contained another UP1-binding site. Thus, stems a−d and loops I−III were deleted in four RNA variants (NLN1, NLN2, NLN3, and NLN4; see Fig. 3A). They only differed one from the other by the number of additional G and C residues in the terminal stem. The effect of the deletion on the RNA structure was studied by one- and two-dimensional $^1$H-$^1$H NOESY analyses (Fig. 3, B and E). The short NLN7 RNA was used to assign the imino resonance of residues in the B motif (Fig. 3, C and F). The information obtained was then used for assignment of imino resonances of residues in larger RNAs. In the one- and two-dimensional $^1$H-$^1$H NOESY spectra of RNA NLN7, the formation of stem g is evidenced by the five sharp resonances with NOE connectivities. They reveal a continuous stretch of seven Watson-Crick base pairs involving segments G$^{5438}$-G$^{5444}$ and C$^{5451}$-G$^{5457}$ and an additional more loose base pair interaction between residues A$^{5445}$ and U$^{5450}$ (broad resonance at 14.2 ppm). The additional resonance at 10.7 ppm was assigned to the loop residue G$^{5447}$ (Fig. 3, C and F). Hence, RNA NLN7 consisted of an 8-bp stem capped by an AGCA tetraloop (Fig. 3D).

Similar assignment of imino resonances in the NOESY spectra of RNA NLN1 also showed the formation of the B motif (Fig. 3, B and E, dashed lines). The broader imino resonance corresponding to residue U$^{5450}$ suggests that the U$^{5450}$-A$^{5455}$ pair is loosely closed in RNA NLN1, and addition of MgCl$_2$ up to 10 mM did not increase its stability (data not shown). Most of the remaining resonances were broad showing rapid exchange with the water signal and thus a low stability of the corresponding helices. According to the NOE differences and two-dimensional NOE-NOESY (Fig. 3, B and E), three of the seven base pairs of helix f were formed. The four terminal base pairs of stem e, including the G$^{5472}$-U$^{5422}$ pair with characteristic chemical shifts, were present. Hence, RNA NLN1 had a more open structure than SLS3 (Fig. 3H). NMR analysis revealed very similar structures for RNAs NLN2, NLN3, and NLN4 (data not shown). In particular, despite the presence of additional C-G pairs in stem e, no extension and stabilization of stem f was detected. Therefore, we concluded that in the absence of stems a−d, ESS2 adopts a loose structure.

Replacement of the Internal Loop IV by a Stable Stem Stabilizes the ESS2 Base Pair Interaction—We hypothesized that replacement of the internal loop IV by a stable stem (RNAs NLN8 and NLN9, see Fig. 3A) may stabilize the ESS2 base pairing in truncated versions of SLS3. The folding of RNA NLN8 was assayed using one- and two-dimensional $^1$H-$^1$H NOESY (Fig. 3, D and G) and $^{15}$N HSQC experiments, using an $^{15}$N-labeled RNA (supplemental Fig. S2). The $^1$H low field region of the NLN8 spectra displayed 13 paired imino protons. According to the $^{15}$N chemical shifts in the HSQC spectra, three of them correspond to U-A pairs and seven to G-C pairs. Identification of the sequential walk between the imino resonances and assignment of NOE connectivity established the formation of stem f. The broad resonance at 13.01 ppm was assigned to the 5’-terminal G residue. A stop was observed in the sequential walk from the G(C)$\Rightarrow$G(C)$\Rightarrow$G(C) stretch assigned to the terminal G$^{5425}$-G$^{5427}$ sequence to the U(A)$\Rightarrow$U(A)$\Rightarrow$G stretch assigned to the first CUAGA motif. This stop should correspond to the unpaired residues A$^{5457}$ and C$^{5458}$, as proposed by probing experiments (8) (Fig. 2D). Identification of sequential anomic/aromatic NOE in the two-dimensional $^1$H-$^1$H NOESY connectivities, as well as in the NOE difference experiments revealed interaction between residues G$^{5462}$ and U$^{5434}$, and residues U$^{5434}$ and G$^{5436}$, indicating the formation of the C$^{5433}$-G$^{5462}$, U$^{5434}$-A$^{5461}$, and G$^{5436}$-C$^{5460}$ pairs, and the stacking of residues U$^{5459}$ and A$^{5437}$ on top of them (data not shown). Altogether, the data clearly established that a helix, including the second CUAGA motif, is formed in RNA NLN8. The absence of detection of the imino protons from residue U$^{5434}$ in the one- and two-dimensional $^1$H-$^1$H NOESY experiments is an indication for lability of the U$^{5434}$-A$^{5461}$. Resonances belonging to the terminal B motif were easily identified in NOESY spectra by comparison with the NLN7 spectra (Fig. 3, F and G, dashed lines). Thus, in agreement with our hypothesis, ESS2 was well folded in RNA NLN8, with the formation of seven Watson-Crick base pairs (Fig. 3F). However, the internal loop V was still slightly larger than in SLS3, because the G$^{5455}$-C$^{5460}$ pair was not formed, and the U$^{5459}$-A$^{5437}$ pair was loosely closed. Here again, addition of MgCl$_2$ up to 10 mM did not stabilize these 2 bp. A very similar folding was found for RNA NLN9 with an A$^{5467}$ to G substitution inducing the formation of an additional C$^{5467}$-C$^{5428}$ pair (Fig. 3A).

Protein UP1 Interacts with ESS2 in NMR Conditions—As our NMR analysis revealed that ESS2 has almost the same structure in NLN8 RNA as in SLS3, RNA NLN8 was used for NMR footprinting experiments with protein UP1. First, to assess the folding of the recombinant UP1 protein, $^1$H-$^1$5N-HSQC of the free UP1 protein was recorded at 600 MHz and 293 K (in black in supplemental Fig. S3). It revealed a highly folded structure, as evidenced by the well dispersed resonances observed. The ability of protein UP1 to interact with RNA NLN8 was then tested by gel-shift assays (RNA concentration of 2 $\mu$M and protein concentration from 0 to 100 $\mu$M, in buffer D), an apparent $K_D$ of 15 $\mu$M was determined from this experiment (supplemental Fig. S4A). To delineate the protected RNA region, we also performed enzymatic footprinting assays at high RNA and protein concentrations (RNA concentration of 17 $\mu$M and protein concentration of 34 $\mu$M) (supplemental Fig. S4B). In agreement with the above foot-printing data on the UP1-A3M RNA complex, protections were detected in the ESS2 and the B motif. Residue G$^{5431}$ was protected against T1 RNase digestion, whereas residue G$^{5458}$ in loop V was cleaved at the same level with or without protein UP1. A more limited protection was observed for residues 5436 and 5438. All the V1 RNase cleavages in ESS2 and the B motif were markedly decreased, whereas
hnRNP A1 and SC35 Binding on HIV-1 Tat Exon 2
those in the segment 5459–5466, complementary to ESS2, had an almost unmodified level in the presence of protein UP1. Taking all these data into account, an NMR titration experiment was performed on a sample of 15N-labeled UP1 protein at a 75 μM concentration in buffer D (seven times more than the estimated Kd) by stepwise addition of unlabeled RNA NLN8, until reaching a final 150 μM RNA concentration. The amide signals of protein UP1 in the presence of increasing concentrations of RNA NLN8 were monitored using 1H–15N HSQC. As shown in the HSQC spectra (supplemental Fig. S3) of the free protein and after addition of the RNA target at a final [RNA]/[protein] molar ratio of 2, some of the resonances (circled in black in the supplemental Fig. S3) disappeared, which was in agreement with an interaction between protein UP1 and RNA NLN8, under the high RNA and protein concentrations needed for NMR analysis.

Protein UP1 Interacts with Residue G5431 in the First UAG of ESS2 and Residue G5444 in Stem g—To identify the nucleotides of RNA NLN8, which are in direct contact with protein UP1 in the UP1–NLN8 complex, the exchangeable imino resonances of an 15N-labeled RNA NLN8 were monitored by 1H–15N HSQC upon addition of increasing amounts of unlabeled UP1 protein (Fig. 4A). The RNA concentration used was of 300 μM. At a 1:1.5 [RNA]/[protein] molar ratio, the resonance of residues G5431 and G5444 were markedly shifted, and those of residues G5436 and G5447 were modified to a lesser extent. No other chemical shift differences were observed upon further addition of UP1. The marked shift of the imino proton of residue G5431 in ESS2 was in perfect agreement with its protections in the enzymatic footprint (supplemental Fig. S4B). Residue G5444 belongs to the B motif. As it was not cleaved by T1 RNase digestion in the enzymatic footprint both with or without protein, no information on this residue was obtained from the footprint experiment.

To confirm that RNA NLN8 remained base paired upon association with protein UP1, we performed HNN-COSY experiments (Fig. 4B) in the presence or the absence of protein UP1. This allowed detection of scalar coupling through hydrogen bonding between nucleotides. The HNN-COSY spectrum of the free RNA reflects the folding established previously. Detection of HNN-COSY peaks for all nucleotides assigned in the HSQC confirmed their involvement in base pair interactions. Upon addition of protein UP1, the same shifts occurred as detected in the HSQC, showing that these nucleotides are indeed involved in the binding of protein UP1. Moreover, nucleotides U5429, G5431, and G5436 in ESS2 and residues G5462 and U5465 in the complementary strand, residues G5444, U5442, G5444, and G5454 in the B motif and residues G5427, G5469 in the upper stem remained base paired as shown by the detection of an HNN–COSY signal for each of them. Therefore, upon binding of protein UP1, RNA NLN8 seems to retain the folding established in its free form (Fig. 4C). Hence, the reduction of V1 RNase cleavages in ESS2 and the B motif found in the footprinting experiment is more likely because of steric hinderance than to two-dimensional structure disruption. Based on the enzymatic and NMR footprints, we conclude for two UP1 anchoring sites on RNA NLN8, ESS2 and the B motif.

Proteins SC35, SRp40, and hnRNP A1 Bind Similar Segments of SLS3—One possible explanation for the strong activation of site A3 (21) by the two SR proteins SC35 and SRp40 might be their capability to displace protein hnRNP A1 from its numerous binding sites in the SLS2-SLS3 region. To test this hypothesis, enzymatic footprinting experiments similar to the ones performed on hnRNP A1-A3M complexes were repeated on complexes formed by proteins SC35 or SRp40 and the A3M RNA. In addition, as we could not produce protein SC35 in the large scale needed for NMR analysis, the data obtained by enzymatic probing of the SC35-A3M RNA complexes were completed by chemical probing, using kethoxal as the probe. This reagent modifies single-stranded accessible G residues, and the modified residues can be detected by primer extension analysis (10). Because of the small size of kethoxal, unmodified residues are expected to be in direct contact with the protein. Preliminary enzymatic footprinting assays performed at different [protein]/[RNA] molar ratios revealed a satisfactory protection for ratios between 2 and 4 (protein concentrations ranging from 0.33 to 0.45 μM). SC35 and SRp40 proteins both protected the two strands of the internal loop II, ESS2, and the terminal loop of the B motif. Some protection of residue G5379 in the SLS2-SLS3 linking segment was also detected (Fig. 5, A, B, C, F, and G), as well as a protection of residue G5454 in the internal loop V (Fig. 5, A, B, C, F, and G). Probing of the SC35 complexes by the use of kethoxal confirmed the interaction of protein SC35 with the 5′ strand of loop II, ESS2, and the terminal loop of the B motif (Fig. 5, B and F). However, protections mainly concerned the second UAGA motif of ESS2 and the neighboring loop V. For both proteins SC35 and SRp40, strong protections were also detected in the 3′ strand of SLS3 (Fig. 5, D, E, F, and G). Altogether, the enzymatic and chemical probing revealed the occurrence of several SC35 and SRp40–binding sites in SLS3 (Fig. 5, F and G), and in agreement with the strong activation properties of proteins SC35 and SRp40 on site A3 (21), several...
hnRNP A1 and SC35 Binding on HIV-1 Tat Exon 2
**hnRNP A1 and SC35 Binding on HIV-1 Tat Exon 2**

**FIGURE 3. Two-dimensional structure analysis of SLS3 truncated RNAs by NMR spectroscopy.** A, the truncated SLS3 RNAs subjected to NMR analysis are drawn. Nonviral nucleotides are boxed. Nucleotides belonging to the ESS2 sequence are displayed in gray characters. Residues are numbered as in SLS3. Helices and internal loop numbering is the one used for SLS3. Only the sequence differences in RNAs NLN2, NLN3, NLN4 as compared with RNA NLN1 are indicated. The same kind of representation is used for RNA NLN9, which is closely related to NLN8. Numbering was designed so as to keep the same number for common viral nucleotides in all constructs. B–J, secondary structure analyses of RNAs NLN1, NLN7, and NLN8 by NMR spectroscopy. The NMR spectra of exchangeable protons of RNAs NLN1 (B) and NLN7 (C) at 277 and 298 K and RNA NLN8 (D) at 277 and 303 K are shown, with the imino proton assignments denoted (only the last two numbers of the HIV sequence numbering are displayed). The resonance assignment of imino protons of NLN1 (E), NLN7 (F), and NLN8 (G) are shown on a two-dimensional NOESY spectrum recorded in H2O on 600 MHz (4°C, mixing time 300 ms). Black lines sketch the sequential walks through imino proton resonances. Dotted lines correspond to the B motif residues. Additional connectivities observed at mixing time of 50 ms are boxed. The secondary structure of RNAs NLN1 (H), NLN7 (I), and NLN8 (J) as deduced from the NMR data are illustrated. Helices, loops, and stems numbering is according to that used for SLS3 (Fig. 2D).

**FIGURE 4. NMR mapping of the UP1/NLN8 interactions.** A, 1H-15N HSQC spectrum of the uniformly 15N-labeled NLN8 RNA free (black) and complexed with protein UP1 (green). Imino resonances affected upon binding to UP1 are circled. B, HNN-COSY spectra of 1H-uniformly labeled NLN8 RNA free in solution (black) or bound to the protein UP1 (green), imino resonance affected by the binding to UP1 are circled. Dashed lines sketched hydrogen bonding across Watson Crick base pair and direct coupling between N1(G)/N3(C) and N3(U)/N1(A). Assignment of imino resonances is denoted. C, nucleotides for which NMR signals are affected upon UP1 binding are circled in green on the proposed secondary structure of RNA NLN8 (pale dark green for limited and marked shifts, respectively).

of the binding sites identified for these two proteins overlap the hnRNP A1-binding sites.

**hnRNP A1 Has a Stronger Affinity for ESS2, whereas SC35 Preferentially binds the 5′ Strand of Loop II**—To test the relative affinities of proteins hnRNP A1 and SC35 for the two strong binding sites that we identified in SLS3 (the ESS2 region and the 5′ strand of loop II), we performed competition assays, using the radiolabeled A3M WT RNA for complex formation, and cold A3M WT or variant RNA for the competition. Complexes were formed with protein SC35 (2 μM concentration) or hnRNP A1 (1.5 μM concentration). Two variant A3M RNAs were used in these competition assays, the A3M Δ5′loop II RNA, which was missing the GAGGAG sequence from position 5398 to 5403 (Fig. 6A), and the A3M ΔESS2 RNA with a deletion of the sequence extending from position 5429 to 5443 (Fig. 6A). The A3M Δ5′loop II RNA had a limited destabilization effect on the complexes formed by protein SC35 and the WT RNA (Fig. 6B). This was in contrast to the strong effects observed for the A3M WT or ΔESS2 RNA. Conversely, compared with the A3M ΔESS2 RNA, A3M Δ5′loop II RNA had a stronger destabilization effect on complexes formed by protein hnRNP A1 and the WT RNA (Fig. 6C). These data were a strong indication for the following: (i) a higher affinity of protein SC35 for the 5′ strand of loop II than for ESS2, and (ii) a stronger affinity of protein hnRNP A1 for the segment deleted in the A3M ΔESS2 RNA than for the 5′ strand of loop II. To verify that the results obtained were not dependent on the identity of the HIV-1 RNA region used in the assays, the experiments were repeated using complexes formed between protein SC35 and the A3S WT, Δ5′loopII, or ΔESS2 RNA (supplemental Fig. S5). The A3S RNA did not contain SLS2 and was missing the SLS3 sequence from position 5467 to 5515 (Fig. 1). The results obtained were very similar to the ones obtained for the long versions of the RNAs. The stronger affinity of protein SC35 for the HIV-1 BRU segment 5398–5403 was verified by gel-shift assays performed with the A3S WT, Δ5′loopII, or ΔESS2 RNA (supplemental Fig. S5A) or by competition with the A3S WT or ΔESS2 RNA (supplemental Fig. S5B). According to published data (20), ESE2 was expected to contain an SC35-binding site. For evaluation of the SC35 affinity for ESE2, gel-shift assays with and without competitor RNA were repeated using an A3M or A3S RNA containing the 2 base substitutions (G to A and A to C, positions 5421 and 5423), which Zahler et al. (20) generated in ESE2 to probe SC35 interaction (our mutated RNAs are designated A3M or A3S mut ESE2 RNAs) (Figs. 6D and S5C). As shown in Fig. 6D, an RNA carrying these 2 base substitutions had the same capability as the A3M WT RNA to destabilize the complexes formed by protein SC35 and the A3M WT RNA. Similar results were obtained when the experiment was repeated with the short versions of the RNAs (supplemental Fig. S5B), and these data were confirmed by gel-shift assays without competitor RNA (supplemental Fig. S5A). Therefore, pro-
tein SC35 was not found to have a stronger affinity for the 5’ strand of loop II than for ESE2, as well in a fully folded RNA (A3M RNA), as in a partially folded RNA (A3S RNA). Accordingly, the sequence of the 5’ strand of loop II fits with one of the consensus sequences established for protein SC35 by use of the SELEX approach (34).
**Despite Its Strong Affinity for Protein SC35, the 5′ Strand of Loop II Has Limited Activation Properties**—As protein SC35 had a high affinity for the 5′ strand of loop II, we tested whether this loop could be involved in the activation of site A3 by protein SC35. To this end, we produced a series of L3-U1 RNA variants and compared their in vitro splicing efficiencies. RNA L3-U1 contains the D1, A3, A4a–c, and A5 splicing sites of HIV-1 RNA (Fig. 1) and was used previously to study site A3 activation by SR proteins in vitro (21). In the L3-U1 variant RNAs that we produced, the 5′ or the 3′ strand of loop II (RNA L3-U1 Δ5′loopII or L3-U1 Δ3′loopII, respectively), the ESS2p (the proximal inhibitor) (RNA L3-U1 ΔESS2p), or ESS2 (RNA L3-U1 ΔESS2) sequences were deleted (Fig. 1). One RNA with a combined deletion of the 5′ strand of loop II and ESS2 was also produced (RNA L3-U1 Δ5′loop II−ΔESS2), as well as an RNA missing ESS2p and the 5′ strand of loop II (RNA L3-U1 Δ5′loop II−ΔESS2p). Splicing assays were performed in the nuclear extracts of HeLa cells, in the presence (300 ng) or the absence of protein SC35 (Fig. 7A). As found previously (21) for the WT L3-U1 RNA, the A3 and A5, and to a lesser extent the A4a and A4b, acceptor sites were used in the absence of protein SC35 addition, and splicing almost only occurred at site A3 upon addition of protein SC35. Without addition of protein SC35 in the extract, deletion of the ESS2p element only had a limited effect on the splicing efficiency at site A3 (MA3/P increase by a factor of 1.5) (Fig. 7A), and the level of activation upon addition of protein SC35 was very similar to the one found for the WT RNA (Fig. 7B). As expected, upon deletion of ESS2, a strong increase of site A3 utilization was observed (factor of 15), and the activation by protein SC35 was only a factor of 3.5. The deletion of the 5′ strand of loop II did not modify significantly the splicing efficiency in nuclear extract. However, the activation by protein SC35 was decreased by a factor of about 1.5. Interestingly, in the absence of SC35 addition, the highest yield of utilization of site A3 was obtained when both the ESS2 region and the 5′ strand of loop II were deleted, which was an indication for an inhibitory property of the 5′ strand of loop II. This inhibitory property may be explained by the ability of the 5′ strand of loop II to bind protein hnRNP A1. A small activation by protein SC35 was still observed for the L3-U1 Δ5′loop II−ΔESS2 (factor of 2). Deletion of the 3′ strand of loop II only had a limited effect on splicing at site A3. However, activation by protein SC35 was increased by a factor of 2 (supplemental Fig. S6). Altogether, the data revealed that the strong SC35-binding site located in the 5′ strand of loop II only has a limited contribution to the activation of site A3 by protein SC35 (25% of the overall activation). Therefore, the most important parameter of site A3 activation was expected to be the displacement of protein hnRNP A1 from ESS2 by protein SC35.

**Mutation of the ESS2 G5431 Residue Found to Contact Protein UP1 Reduces ESS2 Activity**—An important point to address was why among the various hnRNP A1-binding sites present in SLS3, ESS2 has such an inhibitory effect on site A3 utilization. For a better delineation of its mechanism of action, we then produced another large series of L3-U1 RNA variants. Starting from position 5420 to position 5447, we substituted each successive pairs of residues by their complementary sequence (variant L3-U1 RNAs 48–74) (Fig. 8A). The efficiency of site A3 utilization in these variant RNAs was tested by gel-shift assays on 6% polyacrylamide gels in TBE buffer (as indicated above each lane, 0–80 pmol of cold labeled RNA and protein SC35 at a 2 μM concentration, in conditions described under “Experimental Procedures.”) Its stability in the presence of increasing amounts of cold WT, ΔESS2p, or ESS2 RNAs, and the base substitutions in A3M mut ESE2 RNA are shown. B, the SC35-A3M WT RNA complex was formed with 25 fmol of radio-labeled RNA and protein SC35 at a 2 μM concentration, in conditions described under “Experimental Procedures.” Its stability in the presence of increasing amounts of cold WT, ΔESS2p, or ESS2 A3M RNAs was tested by gel-shift assays on 6% polyacrylamide gels in TBE buffer (as indicated above each lane, 0–80 pmol of cold competitor RNA were used). For each assay, the amount of bound RNA is expressed as a percentage of the total RNA. C, the same kind of competition experiment was performed, using the hnRNP A1-A3M WT RNA complex formed with protein hnRNP A1 at a 1.5 μM concentration, D, the destabilization effect of the addition of cold WT or mut ESE2 RNA on the SC35-A3M WT RNA complex was tested. D, same as in B except that the destabilization effect of the addition of cold WT or mut ESE2 RNA on the SC35-A3M WT RNA complex was tested.

**FIGURE 6.** Protein SC35 preferentially binds the 5′ strand of loop II, whereas protein hnRNP A1 has a stronger affinity for ESS2. A, the deletions made in A3M Δ5′loopII, A3M ΔESS2 RNAs, and the base substitutions in A3M mut ESE2 RNA are shown. B, the SC35-A3M WT RNA complex was formed with 25 fmol of radiolabeled RNA and protein SC35 at a 2 μM concentration, in conditions described under “Experimental Procedures.” Its stability in the presence of increasing amounts of cold WT, Δ5′loopII, or ΔESS2 A3M RNAs was tested by gel-shift assays on 6% polyacrylamide gels in TBE buffer (as indicated above each lane, 0–80 pmol of cold competitor RNA were used). For each assay, the amount of bound RNA is expressed as a percentage of the total RNA. C, the same kind of competition experiment was performed, using the hnRNP A1-A3M WT RNA complex formed with protein hnRNP A1 at a 1.5 μM concentration, D, the destabilization effect of the addition of cold WT or mut ESE2 RNA on the SC35-A3M WT RNA complex was tested. D, same as in B except that the destabilization effect of the addition of cold WT or mut ESE2 RNA on the SC35-A3M WT RNA complex was tested.
dependent upon the identity of the substituted residues. In agreement with previous results (9), the mutations in ESS2, which had the strongest positive effects on splicing, were the AG (5430–5431) to UC (variant 58) and GA (5436–5437) to CU (variant 64) substitutions in each of the UAG motifs of ESS2, respectively (Fig. 8, B and C). Interestingly, the A5430G5431 dinucleotides contains residue G5431, which was found to contact protein UP1 by NMR titration (Fig. 4 A). The data also revealed that two dinucleotide substitutions, generated in ESE2, also increased splicing efficiency (variants 50 and 52) (Fig. 8). It should be noticed that they both affected a third UAG motif located upstream the ESS2 UAGs. Hence, the previously described ESE2 enhancer (20) also has inhibitory properties.

**Mutations of the U5422A5423, A5430G5431, and G5436A5437 Dinucleotides Strongly Reduce hnRNP A1 Affinity for ESS2—**

We tested whether the negative effects of the A5430G5431 to UC (variant 58), G5436A5437 to CU (variant 64) and U5422A5423 to AU (variant 50) substitutions on the inhibitory activity of ESS2 were because of a lower affinity of protein hnRNP A1 for the entire SLS3 region. To answer this question, A3M RNAs with the 58 (A5430G5431 to UC), 64 (G5436A5437 to CU), or 50 (U5422A5423 to AU) dinucleotide substitutions were generated, and footprinting experiments were performed on the complexes that they formed with protein hnRNP A1. As illustrated in Fig. 9, even at an [hnRNP A1]/[A3M RNA] molar ratio of 10, no protection of the ESS2 and the B motif was detected for the A3M variant RNAs 58 and 64, and the protection in the 5 strand of loop II also disappeared. Protection of the 3 strand of loop II was still detected but to a lower yield. There-
fore, each of the AG (residues 5430–5431) and GA (residues 5436–5437) dinucleotides in ESS2 is required for binding of protein hnRNP A1 on ESS2, the B motif, and the 3’ strand of loop II and for efficient protection of the 5’ strand of loop II. Protection of ESS2 and the B motif was also decreased upon mutation of the UAG motif in ESE2 (residues 5422–5424), showing an interrelation of the three UAG motifs for hnRNP A1 binding in the ESE2-ESS2 region (supplemental Fig. S7). Thus, we concluded that the UAG motif in ESE2 contributes to reinforce the binding of protein hnRNP A1 on SLS3, which probably explains its splicing inhibitory property.

FIGURE 9. Mutations in each UAG motif of ESS2 limit protein hnRNP A1 binding on the entire SLS2-SLS3 HIV-1 RNA region. Primer extension analysis of the A3M WT and 58 and 64 variant RNAs digested with T2 RNase in the absence (−) or the presence of protein hnRNP A1. The [protein]/[RNA] ratios ([A1]/[A3M]) used for complex formation were 2 or 10, as indicated above the lanes. The enzymatic reactions were performed in buffer D. Control extensions were performed in the absence of RNase T2 (−). Lanes U, G, C, and A correspond to the sequencing ladders. Numbering of the nucleotides in the HIV-1/BRU RNA is indicated on the left side of the panel. Positions of the helices, loops, and of the splicing regulatory elements ESS2p, ESS2, and ESE2 are shown on the right side of the panel.

FIGURE 10. Patterns of activation by protein SC35 of the A3 splicing sites contained in WT and variant L3-U1 RNAs. A, the WT and variant (50, 54, 58, 64, mut ESE2) L3-U1 RNAs were spliced for 2 h in the nuclear extract of HeLa cells, without or with increasing amounts of protein SC35 (0, 25, 50, 100, 200, and 400 ng) (supplemental Fig. S8). The splicing products were fractionated by gel electrophoresis, and the MA3/p values were calculated based on the amount of radioactivity measured in the bands of gel. For each RNA studied, the graph represents the evolution of the MA3/p value as a function of the amount of protein SC35 added in the extract. B, for each RNA, the evolution of the fold of activation of site A3 is represented as a function of the amount of protein SC35 added in the extract.

**Influence of hnRNP A1 and SC35 Binding on Site A3 Activation**—To complete the characterization of the ESE2 and ESS2 properties, we then compared the activation properties of protein SC35 on the L3-U1 variant RNAs 50, 54, 58, and 64. Variant 54 was used as a control as its two base substitutions were not found to modify the splicing activity at site A3 (Fig. 8). The L3-U1 mut ESE2 RNA, containing the two base substitutions that Zahler et al. (20) generated in ESE2, was also introduced for comparison. For each RNA, the effect on splicing at site A3 of the addition of increasing amounts of protein SC35 (25, 50, 100, 200, and 400 ng, respectively) in the nuclear extract was tested (supplemental Fig. S8). In the absence of SC35 addition, all the tested mutations, except the control mutation 54,
hnRNP A1 and SC35 Binding on HIV-1 Tat Exon 2

increased splicing efficiency at site A3 (Fig. 10A). This confirmed our finding of an inhibitory property of the ESE2 sequence. Upon addition of protein SC35, site A3 utilization was also increased in all the tested RNAs (Fig. 10B). However, the dose effect of protein SC35 was very different for the WT RNA and the variant RNA L3-U1 54, as compared with the four other variant RNAs (Fig. 10). The data obtained confirmed the negative effect on splicing efficiency of the three successive UAG sequences in the ESE2-ESS2 region. A larger amount of protein SC35 was required to get an optimum utilization of site A3 in the WT L3-U1 RNA and the variant L3-U1 54 RNA. The maximum splicing efficiency obtained for the mut ESE2 L3-U1 variant RNAs was slightly lower compared with that obtained for the 50 and 58 variant RNAs, which may be due to an alteration of the SC35-binding site in ESE2.

DISCUSSION

This study reveals how protein hnRNP A1 and the SR proteins SC35 and SRp40 develop several interactions with the SLS2-SLS3 region of HIV-1 RNA that contains the A3 splicing site and its regulatory elements. The data obtained provide important information toward the understanding of modulation of site A3 utilization by these proteins.

The RRM Domains of Protein hnRNP A1 Can Interact with Both ESS2 and the B Motif—Gel-shift experiments and enzymatic and NMR footprints demonstrate the physical interaction of the RRM domains of protein hnRNP A1 with ESS2 and the B motif. This is the first demonstration of a direct interaction between protein hnRNP A1 and ESS2. As in the 1H-15N HSQC of RNA NLN8 obtained upon progressive addition of protein UP1, the imino resonances of residues were maintained either at an identical position or with a shifted position, and the base-paired interactions present in the naked RNA were expected to be preserved in the RNP complex, and this was confirmed by the HNN-COSY experiment. Thus, a complete opening of the RNA double-stranded structure can be ruled out. The UP1-ESS2 interaction can likely occur because of the presence of five single-stranded residues in the peculiar two-dimensional structure of ESS2 in RNA NLN8 (G5428, A5431, U5434, A5435, and A5437). In addition, several of the base-paired residues are purine residues (G5427, A5430, G5431, and G5436), and they may form hydrogen bonds through the N-7 position of their bases. Furthermore, they may be involved in stacking interaction with phenylalanine residues, as found in the UP1-telomeric DNA interaction (24), where four Phe residues are stacked on the bases of the AG dinucleotides of the two UAG motifs. However, in contrast to the UP1-telomeric DNA interaction, the UP1-RNA interaction may involve ribose 2'-OH groups. Another difference between ESS2 and telomeric DNA is the different size of the segment linking the two (U or T)AGPu sequences (one residue in ESS2 against two residues in telomeric DNA). Finally, the additional interaction of protein UP1 with residue G5444 in the stem of the motif B likely also contributes to the specificity. The overall UP1-NLN8 interaction is not very strong, as the estimated $K_d$ is of 15 μM. However, taking all the above points into consideration, it is conceivable that an RNA segment, which is only partially base-paired with its complementary strand, as is the case for ESS2, can interact specifically with the two RRM's of protein UP1. The marked chemical shift observed for residue G5431 may reflect the interaction with one of the RRM's, and the more limited chemical shift found for residue G5436 may result from the interaction with the second RRM. The strong NMR footprinting detected for residue G5444 may correspond to an interaction through the C-terminal helix, as found in the U1A/RNA (35). Therefore, the shifts observed in the HSQC analysis may be explained by the interaction of a single UP1 molecule with RNA NLN8.

The hnRNP A1 Interaction with ESS2 and the B Motif Is Coupled with That in the 3' Strand of Loop II and ESE2—We found concerted appearances and concerted disappearances of the protections in the ESS2, the B motif, and the 3' strand of loop II. Their simultaneous appearance was detected in enzymatic footprints upon the increase of the hnRNP A1 or UP1 protein concentration. Their simultaneous disappearance was observed upon mutation of the A5430G5431 or G5436A5437 dinucleotides (ESS2 motif). A decrease of the ESS2 protection by protein hnRNP A1 was also observed upon mutation of the U5422A5423 dinucleotide in the ESE2 motif (supplemental Fig. S7). Taken together, these data strongly suggest a cooperative binding of two UP1 or two hnRNP A1 molecules on SLS3. The two RRM domains of one molecule likely interact with ESS2 and the B motif, whereas the two RRM domains of the second one would interact with the UAGA sequence in ESE2 and the 5' strand of loop II, which is an A-rich segment. Accordingly, protein hnRNP A1 was found previously to bind poly(A) stretched (36). Based on our site-directed mutagenesis experiments, the A5430G5431 and G5436A5437 dinucleotides in ESS2 play a crucial role in the assembly of this multiple hnRNP A1 complex on SLS3. The UAGA motif in ESE2 could not be identified as an hnRNP A1-binding site in our enzymatic footprints, because none of the enzymes used cleaved this segment in the naked RNA. Nevertheless, its mutation revealed its importance in hnRNP A1 recruitment, and thus in silencing of site A3. In contrast, despite the observed strong protection of the B motif and the 3' strand of loop II by protein hnRNP A1, these two sequences are not essential for the interaction. This is evidenced by protein UP1 interaction with RNA NLN8, which is missing the 5' strand of loop II and by enzymatic footprints on the A3M 70, 72, and 74 and Δ3'loopII RNAs variants (data not shown). The UAGA motifs in ESE2 and ESS2 are likely the key determinants for initiation of the hnRNP A1-SLS3 interaction. Then, the interaction is probably extended by contacts established with the B motif and the 3' strand of loop II (Fig. 11). Therefore, in addition to be a splicing enhancer (20), ESE2 is a splicing silencer. It is a multifunctional sequence, because it also belongs to one of the hnRNP A2 protein-target site (37). Binding of protein hnRNP A2 to this site was proposed to be involved in HIV-1 RNA transport (38).

The Gly Domain of Protein hnRNP A1 Increases the Affinity for SLS3—in agreement with previous results (23, 39), gel-shift and footprinting experiments reveal a much stronger affinity of protein hnRNP A1 for SLS3, as compared with protein UP1. Hence, the Gly domain of protein hnRNP A1 participates in some way to the interaction with SLS3. The reinforced binding may be due both to the RNA binding capacity of the Gly
hnRNP A1 and SC35 Binding on HIV-1 Tat Exon 2

explains the specific activation properties of these two SR proteins at site A3. Indeed, protein ASF/SF2, which is not an enhancer of site A3 utilization, shows a different pattern of interaction. Interestingly, although located in the close vicinity of site A3, the strongest SC35-binding site only has a modest effect on site A3 activation. This demonstrates that the most important role of protein SC35 in the activation mechanism is not to bring an RS domain in the close vicinity of the 3’ ss but to displace protein hnRNP A1 from its binding sites in the ESE2-ESS2 region. According to the present data, protein SC35 only has a low affinity for the previously described ESE2 enhancer (20). Its affinity for the downstream sequence containing ESS2 is higher. Furthermore, in a HeLa cell nuclear extract, the inhibitory property of ESE2 is greater than its enhancer activity, because mutations in ESE2 increase site A3 utilization. Hence, as previously found for the ESS3 regulatory element that modulates site A7 utilization in HIV-1 RNA (10), the ESE2-ESS2 stretch in SLS3 has both splicing inhibitory and enhancer properties, depending on the protein partner. Altogether, the complexity of the network of hnRNP A1 interactions on SLS3 is very similar to that found at the A7 splicing site (10). However, a main difference is the implication of protein ASF/SF2 in activation at site A7 and of proteins SC35 and SRp40 in activation at site A3. Interestingly, as the 3’ strand of loop II contains the branch site sequences for the splicing site A4c, which is used for rev mRNA production (41), ESS2 may control both site A3 and site A4c utilizations through hnRNP A1 recruitment.

Acknowledgments—We thank N. Birlirakis for the strong support in NMR data acquisition and D. Thomas and J. Ugolini for their technical assistance. We thank D. Fourmy for the helpful discussions in NMR data interpretation. We also thank Dr. J. Stevenin and R. Gattoni for their generous gift of proteins SC35 and SRp40.

REFERENCES

1. Purcell, D. F., and Martin, M. A. (1993) J. Virol. 67, 6365–6378
2. Amendt, B. A., Hesslein, D., Chang, L. J., and Stoltzfus, C. M. (1994) Mol. Cell. Biol. 14, 3960–3970
3. Amendt, B. A., Si, Z. H., and Stoltzfus, C. M. (1995) Mol. Cell. Biol. 15, 4606–4615
4. O’Reilly, M. M., McNally, M. T., and Beemon, K. L. (1995) Virology 213, 375–385
5. Staffa, A., and Cochrane, A. (1995) Mol. Cell. Biol. 15, 4597–4605
6. Staffa, A., and Cochrane, A. (1994) J. Virol. 68, 3071–3107
7. Damier, L., Domencjoud, L., and Branlant, C. (1997) Biochem. Biophys. Res. Commun. 237, 182–187
8. Jacquesnet, S., Ropers, D., Billouet, P. S., Damier, L., Moguin, A., Stoltzfus, C. M., and Branlant, C. (2001) Nucleic Acids Res. 29, 464–478
9. Si, Z., Amendt, B. A., and Stoltzfus, C. M. (1997) Nucleic Acids Res. 25, 861–867
10. Marchand, V., Mereau, A., Jacquesnet, S., Thomas, D., Moguin, A., Gattoni, R., Stevenin, J., and Branlant, C. (2002) J. Mol. Biol. 323, 629–652
11. Dyhr-Mikkelsen, H., and Kjems, J. (1995) J. Biol. Chem. 270, 24060–24066
12. Stoltzfus, C. M., and Madsen, J. M. (2006) Curr. HIV Res. 4, 43–55
13. Karn, J. (1999) J. Mol. Biol. 293, 235–254
14. Bonavida, R., Bajetto, A., Barbero, S., Albin, A., Noonan, D. M., and Schettini, G. (2001) Biochem. Biophys. Res. Commun. 288, 301–308

5 D. Ropers, L. Ayadi, D. Thomas, and C. Branlant, unpublished data.
hnRNP A1 and SC35 Binding on HIV-1 Tat Exon 2

15. Li, C. J., Friedman, D. J., Wang, C., Metelev, V., and Pardee, A. B. (1995) Science 268, 429–431
16. Bilodeau, P. S., Domsic, J. K., and Stoltzfus, C. M. (1999) J. Virol. 73, 9764–9772
17. Jacquenet, S., Mereau, A., Bilodeau, P. S., Damier, L., Stoltzfus, C. M., and Branlant, C. (2001) J. Biol. Chem. 276, 40464–40475
18. Caputi, M., Mayeda, A., Krainer, A. R., and Zahler, A. M. (1999) EMBO J. 18, 4060–4067
19. Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S., and Bianmonti, G. (1996) J. Mol. Biol. 259, 337–348
20. Zahler, A. M., Damgaard, C. K., Kjems, J., and Caputi, M. (2004) J. Biol. Chem. 279, 10077–10084
21. Ropers, D., Ayadi, L., Gattoni, R., Jacquenet, S., Damier, L., Branlant, C., and Stevenin, J. (2004) J. Biol. Chem. 279, 29963–29973
22. Jacquenet, S., Decimo, D., Muriaux, D., and Darlix, J. L. (2005) Retrovirology 2, 33–46
23. Mayeda, A., Munroe, S. H., Caceres, J. F., and Krainer, A. R. (1994) EMBO J. 13, 5483–5495
24. Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R. M. (1999) Genes Dev. 13, 1102–1115
25. Kao, C., Zheng, M., and Rudisser, S. (1999) RNA (Cold Spring Harbor) 5, 1268–1272
26. Cavalo, Y., Bourgeois, C. F., Kister, L., and Stevenin, J. (1999) RNA (Cold Spring Harbor) 5, 468–483
27. Collin, D., van Heijenoort, C., Boiziau, C., Toulme, J. J., and Guittet, E. (2000) Nucleic Acids Res. 28, 3386–3391
28. Dингley, A. J., and Grzesiek, S. (1998) J. Am. Chem. Soc. 120, 8293–8297
29. Pons, J. L., Malliavin, T. E., and Delsuc, M. A. (1996) J. Biomol. NMR 8, 445–452
30. Varani, G., Aboul-ela, F., and Allain, F. H. (1996) Prog. NMR Spectroscopy 29, 51–127
31. Bilodeau, P. S., Domsic, J. K., Mayeda, A., Krainer, A. R., and Stoltzfus, C. M. (2001) J. Virol. 75, 8487–8497
32. Damgaard, C. K., Tange, T. O., and Kjems, J. (2002) RNA (Cold Spring Harbor) 8, 1401–1415
33. Eperon, I. C., Makarova, O. V., Mayeda, A., Munroe, S. H., Caceres, J. F., Hayward, D. G., and Krainer, A. R. (2000) Mol. Cell. Biol. 20, 8303–8318
34. Tacke, R., and Manley, J. L. (1995) EMBO J. 14, 3540–3551
35. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Nature 372, 432–438
36. Pinol-Roma, S., and Dreyfuss, G. (1992) Nature 355, 730–732
37. Mouland, A. J., Xu, H., Cui, H., Krueger, W., Munro, T. P., Prasol, M., Mercier, J., Rekosh, D., Smith, R., Barbaresco, E., Cohen, E. A., and Carson, J. H. (2001) Mol. Cell. Biol. 21, 2133–2143
38. Beriault, V., Clement, J. F., Levesque, K., Lebel, C., Yong, X., Chabot, B., Cohen, E. A., Cochrane, A. W., Rigby, W. F., and Mouland, A. J. (2004) J. Biol. Chem. 279, 44141–44153
39. Mayeda, A., Munroe, S. H., Xu, R. M., and Krainer, A. R. (1998) RNA (N. Y.) 4, 1111–1123
40. Casas-Finet, J. R., Smith, J. D., Jr., Kumar, A., Kim, J. G., Wilson, S. H., and Karpel, R. L. (1993) J. Mol. Biol. 229, 873–889
41. Swanson, A. K., and Stoltzfus, C. M. (1998) J. Biol. Chem. 273, 34551–34557