Regulation of HIV-1 splicing
Müller, Nancy

Citation for published version (APA):
Müller, N. (2016). Regulation of HIV-1 splicing

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 2

HIV-1 splicing at the major splice donor site is restricted by RNA structure

Nancy Mueller, Nikki van Bel, Ben Berkhout, Atze T. Das

Laboratory of Experimental Virology, Department of Medical Microbiology, Center for Infection and Immunity Amsterdam (CINIMA), Academic Medical Center, University of Amsterdam, The Netherlands

Virology, 2014, Volume 468-470, Pages 609-620
Abstract
The 5' leader region of the HIV-1 RNA contains the major 5' splice site (ss) that is used in the production of all spliced viral RNAs. This splice-donor (SD) region can fold a stem-loop structure. We demonstrate that whereas stabilization of this SD hairpin reduces splicing efficiency, destabilization increases splicing. Both stabilization and destabilization reduce viral fitness. These results demonstrate that the stability of the SD hairpin can modulate the level of splicing, most likely by controlling the accessibility of the 5'ss for the splicing machinery. The natural stability of the SD hairpin restricts splicing and this stability seems to be fine-tuned to reach the optimal balance between unspliced and spliced RNAs for efficient virus replication. The 5'ss region of different HIV-1 isolates and the related SIVmac239 can fold a similar structure. This evolutionary conservation supports the importance of this structure in viral replication.
Introduction

HIV-1 encodes a ~9 kb RNA transcript with at least 9 open reading frames and untranslated leader and trailer regions (Fig. 1A). This full-length RNA is packaged as genomic RNA in the virion and used as mRNA for translation of the Gag and Pol proteins. Additionally, a multitude of different singly and multiply spliced transcripts are produced that encode the other viral proteins (1-3). Splicing has to be strictly regulated to obtain the right balance of these RNAs for the synthesis of the appropriate level of all viral proteins. In the early phase of HIV-1 transcription, the primary transcript that is produced in the nucleus is fully spliced and the resulting multiply spliced RNAs (~2 kb) are transported to the cytoplasm where they are translated into the Tat, Rev and Nef proteins. The Rev protein that accumulates in the nucleus will target the Rev-responsive element (RRE) in the env region of the unspliced and singly spliced (~4 kb) RNAs, thus activating their export to the cytoplasm via the Crm1 pathway (4). The Vif, Vpr, Vpu and Env proteins will subsequently be produced from the singly spliced transcripts and the Gag and Gag-Pol proteins from the unspliced transcripts.

HIV-1 uses several splice donor (5'ss) and splice acceptor (3'ss) sites for the generation of the large variety of spliced mRNAs (3, 5). During splicing, the spliceosome complex consisting of several small nuclear ribonucleoproteins (snRNPs) is formed on the pre-mRNA. U1 snRNP binds to the 5'ss, which is mediated by the sequence complementarity with the 5' end of the U1 snRNA component. The U2 snRNP auxiliary factor (U2AF) interacts with the 3'ss (6, 7) followed by binding of U2 snRNP to the branch point sequence (7, 8). Subsequently, the U4/U6-U5 tri-snRNP particle will bind to complete the process of spliceosome assembly (7). Usage of the different 5'ss and 3'ss can be influenced by their sequence and by nearby positioned splicing enhancer or silencer motifs that bind cellular factors such as SR (serine and arginine rich) and hnRNP (heterogeneous nuclear ribonucleoprotein) proteins (5, 9). Most splice site sequences in HIV-1 differ from the consensus 5' and 3'ss sequences described for mammalian genes (5'ss: MAG/guragu; 3' ss: y(G/U)n[yag/G; with M= C or A; r= A or G; y= C or U; n= any nucleotide (nt); exonic nucleotides (nts) in capital letters; intronic nts in lower-case letters) (10, 11). The presence of suboptimal 5' and 3'ss elements has been proposed to allow the intricate alternative splicing of HIV-1 transcripts (2, 12-14). Furthermore, RNA structure may affect the accessibility of the splice signals and influence the splicing efficiency and pattern (15-19).

The major 5'ss that is used in the generation of all spliced HIV-1 RNAs is located upstream of the gag gene and embedded within the highly structured 5' leader region of the viral RNA (Fig. 1A-B). This splice donor (SD) region can fold into a stem-loop structure (the SD hairpin; Fig. 1C) and splicing at this 5'ss may therefore be influenced by both the primary sequence and the RNA secondary structure (20). Whereas an increased sequence complementarity between the 5'ss and the 5' end of the U1 snRNA will allow more efficient U1 snRNA binding and spliceosome assembly and increase splicing, a less optimal sequence complementarity will reduce complex formation and splicing. Similarly, occlusion of the 5'ss in a stable RNA structure will prevent U1 snRNA binding and splicing, whereas a better exposure of this site in an open or unstable structure may allow
efficient U1 snRNA binding and subsequent splicing. Additionally, it has recently been suggested that several SR proteins can bind to this SD region and reduce or activate splicing (21).

It was previously demonstrated that stabilization of the SD hairpin through the removal of the 1-nt bulge and 1-base pair (bp) extension of the stem (J7 mutant, Fig. 1B-C) decreases the splicing efficiency (18). Based on this observation, we hypothesized that -in the context of the wild-type (wt) virus - splicing at the major 5'ss may be suppressed by the SD hairpin structure to allow the regulation of alternative splicing. To test this hypothesis, HIV-1 leader RNA constructs and virus variants with a destabilized SD stem-loop structure were generated and analyzed for 5'ss usage.

Figure 1. The HIV-1 major 5'ss is positioned in a stable hairpin structure. (A) The HIV-1 provirus (DNA) is shown with its open reading frames and long terminal repeat regions (LTRs). The 5' and 3' LTR consist of U3, R and U5 regions. Transcription starts at the U3-R border (+1) of the 5' LTR and the transcripts are polyadenylated at the R-U5 border of the 3' LTR. Usage of different 5'ss and 3'ss sites results in the production of several singly (4-kb group) and multiply (2-kb group) spliced RNAs. The major 5'ss is used for the production of all spliced RNAs (the other 5'ss and 3'ss sites are shown simplified). (B) The sequence complementarity between the 5' end of the U1 snRNA and the major 5'ss (boxed in grey; ▲, cleavage site) is shown. Mutations (boxed in black; ∆, nt deletion) were introduced in the 5'ss region to inactivate the 5'ss (J1 mutant), to stabilize the SD hairpin (J7) or extend this hairpin (J2). (C) The structure and thermodynamic stability ($\Delta G$ in kcal/mol) of the wt, J2 and J7 SD hairpin as predicted with the Mfold RNA structure analysis software and confirmed by SHAPE-directed RNA structure determination (Supplementary Fig. 1) are shown.

Results

Stabilization of the SD hairpin structure affects virus replication

Before assessing whether SD hairpin destabilization increases the splicing efficiency, the effect of the previously tested J7 mutation (18) and another stabilizing mutation
(J2) was analyzed in the virus context and in reporter constructs (Fig. 1B-C). In J2, the AAAA sequence downstream of the SD hairpin (nt +302 to +305) was substituted by a UCCU sequence to extend the stem with 4 bp. For comparison, we included the J1 mutant in which the U1 snRNA binding sequence was mutated. These J1 mutations prevent splicing ((18) and Fig. 3B). The corresponding virus did not replicate at all and attempts to obtain revertant virus variants with improved replication capacity by long-term culturing were unsuccessful (18).

Mfold RNA structure analysis predicted that J2 stabilized the SD hairpin to a similar extent as J7, with the thermodynamic stability changed from ΔG= -9.0 to -12.7 (J2) or -13.6 (J7) kcal/mol. To confirm that the mutated RNAs did fold the intended SD hairpin structure, we determined the structure of wt, J2 and J7 leader RNA molecules (nts +1 to +368, with +1 being the HIV-1 transcription start site) by SHAPE-directed RNA structure analysis. This analysis demonstrated that the wt RNA did fold the predicted leader RNA structure with the DIS (dimer-initiation signal) and Ψ (packaging) hairpins upstream and downstream of the SD hairpin, respectively (Supplementary Fig. 1). Analysis of the J2 and J7 RNAs demonstrated that these RNAs did fold the stabilized SD hairpin that was predicted by MFold, and wt-like DIS and Ψ hairpins (Supplementary Fig. 1). Furthermore, the more upstream and downstream RNA regions showed a similar reactivity pattern in the SHAPE analysis, demonstrating that the introduced mutations do not cause any unexpected structural rearrangement in the 5’ leader region (data not shown).

The effect of the J2 and J7 mutations on virus replication was analyzed in the SupT1 T cell line that supports HIV-1 replication. The wt virus replicated fast, which resulted in a rapid increase in the extracellular CA-p24 level and the appearance of syncytia. The J2 mutant showed a slight delay in virus replication, whereas the J7 mutant replicated poorly (Fig. 2A). The same ranking in replication efficiency was scored in peripheral blood mononuclear cells (PBMCs; data not shown). To confirm that J2 replicated less efficiently than the wt virus, competition experiments were performed. Virus cultures were started by transfection of an equimolar mixture of the wt and J2 molecular clones into SupT1 cells. The viruses were passaged at the peak of infection when massive syncytia were apparent. At the second passage, integrated proviral DNA was isolated from infected cells and the leader RNA region was sequenced. Sequencing of the input virus resulted in a mixed sequence at the positions at which the wt and mutant virus differ, in agreement with the presence of both viruses at the start of the experiment (input samples in Fig. 2B-C). Sequencing of the virus population present at the second passage revealed the presence of only the wt nt at the discriminatory positions (output samples in Fig. 2B-C), which demonstrates that the relative level of the wt virus had increased, while the level of the mutant virus had dropped significantly. These results demonstrate that the wt virus replicated more efficiently and outcompeted both the J2 mutant (Fig. 2B) and the J7 mutant (Fig. 2C), which shows that the stabilizing mutations in the SD hairpin reduce virus replication.
Figure 2. Stabilization of the SD hairpin reduces virus replication. (A) SupT1 T cells were infected with equal amounts of 293T-produced wt and mutant virus. Viral replication was monitored by measuring the CA-p24 level in the culture supernatant. The replication curves shown are representative of data obtained in three experiments. (B) and (C) Competition experiments. SupT1 T cells were transfected with an equimolar mixture of wt and mutant virus-encoding plasmids. The virus was cultured for several weeks and passaged onto fresh cells when virus-induced syncytia were massively present. The viral leader RNA region of the input DNA and integrated proviral DNA isolated from infected cells at the second passage (day 17-21; output) was sequenced (population sequencing) with a reverse primer annealing downstream of the mutations. The nts mutated in J2 and J7 indicated are in red. The 1-nt deletion in J7 (marked with \( \Delta \)) causes a misalignment of the upstream J7 and wt sequences.

Stabilization of the SD stem-loop structure reduces splicing
To analyze the effect of the stabilizing mutations on 5'ss usage, LTR-driven leader RNA-luciferase reporter constructs (Fig. 3A) with a wt or mutated SD element were transfected into 293T cells. These cells were co-transfected with a Tat-expressing plasmid to activate transcription from the LTR promoter. The production of unspliced and spliced RNA was analyzed 48 h after transfection. For this, RNA was isolated from the transfected cells and reverse transcribed. The cDNA was PCR-amplified with a primer combination that detects both unspliced and spliced transcripts (710-bp and 429-bp products, respectively; Fig. 3A) and analyzed by agarose gel electrophoresis. With the wt construct, both RNA products were detected, but the spliced RNA signal was stronger than the unspliced RNA signal. As previously observed, the J1 mutation blocked 5'ss usage and this variant showed only the unspliced product, while the J7 mutation strongly reduced splicing, resulting in an increase in the amount of unspliced transcript and a concomitant decrease in the spliced transcript level (Fig. 3B). The J2 variant showed a similar pattern as the wt construct, which indicates that the J2 mutation did not significantly affect 5'ss usage.

The RNA produced by wt and SD-mutated LTR-luciferase reporter constructs was also analyzed by Northern blotting (Fig. 3C). Analysis of the wt construct revealed the presence of the anticipated unspliced (~2.3 kb) and spliced (~2.0 kb) luciferase RNAs. An additional smaller RNA (~1.0 kb) was detected, which most likely resulted from splicing to an alternative 3'ss in the luciferase coding region. Both J1 and J7 showed an increased level of unspliced RNA. The J1 construct yielded no spliced RNA, whereas J7 produced a very low level of the spliced RNAs. These observations are in agreement with the complete inhibition of splicing by the J1 mutation and strong inhibition of splicing by the J7 mutation, as observed in the RT-PCR analysis (Fig. 3B). The level of the spliced and unspliced RNAs observed for the J2 mutant did
Figure 3. Stabilization of the SD stem-loop structure reduces splicing. (A) In the pLTR-gag-flag-luc construct, the HIV 5’ LTR and leader RNA sequences including the major 5’ss (with +1 indicating the transcription start site) are coupled to the gene encoding a fusion protein consisting of the N-terminal 25 amino acids of Gag, the Flag tag peptide and the firefly luciferase protein (18). The luciferase transcript contains a cryptic 3’ss (18). The primers used for RT-PCR analysis are indicated. This RT-PCR will result in a product of 710 bp for the unspliced RNA and 429 bp for the spliced RNA. (B-C) 293T cells were transfected with the wt or mutated pLTR-gag-flag-luc constructs and a Tat-expressing plasmid. Intracellular RNA was isolated after 48 h and analyzed by RT-PCR (panel B) and northern blotting (panel C). m, DNA size marker. The position of the 18S rRNA is indicated. (D) 293T cells were transfected with one of the pLTR-gag-flag-luc constructs with or without a Tat-expressing plasmid. The intracellular luciferase production was measured after 48 h. The mean and SD of 4 measurements are shown. Statistical analysis was used to compare the wt and mutant values (*, p < 0.05). Mock, control cells transfected with the empty pBluescript plasmid instead of the pLTR-gag-flag-luc construct. (E) The luciferase level measured upon cotransfection with Tat (panel C) was used to calculate the splicing efficiency. The J1 mutant did not show any splicing (panel B) and the luciferase level measured with this constructs reflects 100% unspliced luciferase-encoding RNA. In the absence of unspliced RNA (0%; mock transfected cell) the luciferase level was close to zero. These 0% and 100% levels were used to interpolate the splicing frequency of the other constructs ([observed luciferase level / J1 luciferase level] x 100%).
not noticeably differ from the wt level, which is also in agreement with the RT-PCR results.

Because the semi-quantitative RT-PCR and the Northern blot analyses may not reveal small differences in splicing efficiency, we also analyzed the effect of the mutations on splicing by measuring the luciferase activity in the transfected cells, which provides a more quantitative measure. Whereas the unspliced RNA encodes an active luciferase protein, the spliced RNA does not (Fig. 3A). The wt construct generated a low luciferase level, which correlates with efficient splicing. The luciferase level of the J1 mutant was significantly higher, which is in agreement with the increased unspliced transcript level observed for this mutant in the RT-PCR analysis. J7 also produced a high luciferase level, which is in agreement with the very low splicing activity observed in the RT-PCR assay. J2 showed a 1.6-fold increase in luciferase expression, which reflects a slight decrease in splicing efficiency (Fig. 3D). Apparently, this decrease was too small to be detected in the RT-PCR analysis, but correlates with the minor virus replication deficit. A similar pattern in luciferase activity was observed when the cells were transfected without Tat plasmid and transcription is driven by the low basal LTR-promoter activity (Fig. 3D). The splicing defect of the J1 mutant allowed us to estimate the splicing efficiency of the other variants, as the luciferase level observed with J1 is generated in the absence of any splicing (0%), whereas complete splicing (100%) will result in undetectable luciferase activity. Interpolation of the luciferase levels observed for the other constructs, indicated that the wt SD hairpin resulted in ~85% splicing of the transcripts, whereas the J2 and J7 variants exhibited ~74% and ~13% splicing, respectively (Fig. 3E). Thus, both the J2 and J7 mutated constructs demonstrate that stabilization of the SD hairpin decreases 5'ss splicing. As observed in the virus replication analysis, the effect of the J7 mutation was much more pronounced than that of the J2 mutation. Although both mutations have a similar effect on the overall hairpin stability, J7 stabilizes the top part where splicing occurs whereas J2 extends the bottom part. The more severe effect of the J7 mutation may indicate that local accessibility of the 5'ss is important.

**Destabilization of the SD hairpin reduces virus replication**

It seems possible that the wt hairpin stability partially restricts 5'ss usage to prevent excessive splicing. To test this hypothesis, the effect of SD hairpin destabilization on splicing was analyzed. To prevent effects on known and unknown sequence elements in the SD region, we introduced nt substitutions that had been observed in previous evolution experiments with the J7-mutated HIV-1 variant (18). These nt substitutions were all positioned outside the 9-nt U1-snRNA binding site (A, B and C mutants, Fig. 4A-B). In addition, we combined these nt changes to further destabilize the structure (AB, BC, AC and ABC mutants). Mfold RNA structure analysis predicted that the SD hairpin was destabilized in all mutants, with the thermodynamic stability varying from \(\Delta G = -5.5\) (C) to \(-3.7\) (AC, BC and ABC) kcal/mol compared to the wt value of -9.0. SHAPE RNA structure analysis of corresponding leader RNA molecules confirmed that the mutant RNAs did fold the modified SD hairpin structure without affecting other domains of the leader RNA structure (Supplementary Fig. 1 and data not shown).
The replication capacity of the virus variants was tested in the SupT1 T cell line. All mutants replicated, but slightly delayed when compared with the wt virus (Fig. 5). A similar reduction of virus replication was observed upon infection of PBMCs (data not shown). To confirm that the mutants replicate less efficiently than the wt virus, competition experiments were started by transfection of an equimolar mixture of wt and mutant constructs (either A, B, C, AB, BC, AC or ABC). The input DNA and the integrated proviral DNA isolated from infected cells after passaging of the viruses was sequenced. Sequencing of the input virus resulted in a mixed sequence at the mutated positions, confirming the presence of both viruses at the start of the experiment (input samples in Fig. 6). For all cultures, sequencing of the virus population after passaging revealed the exclusive presence of the wt nt at the discriminatory positions, which demonstrated that the wt virus always outcompeted the mutant virus (output samples in Fig. 6). These results confirm that the SD hairpin mutations affect the virus replication capacity.
Destabilization of the SD hairpin increases splicing

The effect of SD hairpin destabilization on splicing was investigated by transfection of 293T cells with wt and mutant LTR-leader-luciferase constructs and subsequent analysis of the intracellular spliced and unspliced transcripts by RT-PCR. The mutants A, B and AB demonstrated a slightly reduced level of unspliced RNA compared to the wt construct, which indicates that these destabilizing mutations slightly increase splicing (Fig. 7A). In contrast, the C mutant and the BC, AC and ABC mutants did not show such a decrease in unspliced RNA. Northern blot analysis of the RNA produced by the A, B and C mutated constructs confirmed that the unspliced RNA level was slightly decreased for the A and B mutants, but not for the C mutant (Fig. 3C). Because these effects are difficult to observe in the RT-PCR and Northern-blot assays, we used the more quantitative luciferase assay that specifically measures the protein product of the unspliced RNA (Fig. 7B). The mutants A, B and AB showed a ~2-fold decrease in luciferase activity which reflects an increase in splicing compared to the wt. This result is in agreement with our hypothesis that the stability of the wt hairpin restricts splicing. Surprisingly, the C mutant showed a slight increase in luciferase activity (~1.3-fold compared to wt; p=0.084), which indicates a small decrease in splicing. The mutants AC, BC and ABC showed a luciferase level similar to the wt construct, which indicates that the splicing enhancing effect of the A, B and AB mutations is neutralized by the splicing inhibitory effect of the C mutation. Similar effects were observed when the cells were transfected without Tat plasmid and a lower level of transcription was reached by the basal LTR-promoter activity (Fig. 7B). The observed luciferase levels were used to estimate the splicing efficiency of the destabilized mutants, as described above for the stabilized mutants. Compared to the ~85% splicing efficiency of the wt transcript, the A, B and AB mutants showed an increased splicing level of ~93%. The splicing frequency was reduced to ~80% for the C mutant. The BC, AC and ABC combination mutants demonstrated a splicing frequency similar to that of the wt (~85-88%; Fig. 7C).

Figure 5. Destabilization of the SD hairpin reduces virus replication. SupT1 T cells were infected with equal amounts of wt and mutant virus. Virus replication was monitored by measuring the CA-p24 level in the culture supernatant. The replication curves shown are representative of data obtained in three experiments.
Figure 6. SD hairpin destabilization reduces virus replication. Virus-competition cultures were started by transfection of SupT1 T cells with an equimolar mixture of wt and mutant virus-encoding plasmids. The viral leader RNA region of the input DNA and integrated proviral DNA isolated from infected cells after culturing was sequenced as described in Fig. 2B-C.
Figure 7. Destabilization of the SD stem-loop structure increases splicing. (A) 293T cells were transfected with a wt or mutated pLTR-gag-flag-luc plasmid and a Tat plasmid. Intracellular RNA was isolated after 48 h and analyzed by RT-PCR as described in Fig. 3B. (B) Luciferase production in cells transfected with the wt or mutated pLTR-gag-flag-luc plasmid with or without a Tat plasmid was analyzed as described in Fig. 3D. The mean and SD of 4 independent experiments are shown. Statistical analysis was used to compare the wt and mutant values (*, p < 0.05). (C) The splicing frequency of the wt and mutant constructs was calculated as described in Fig. 3E.

Communication between the 5’ss and the 5’ polyadenylation signal
One complication with the analysis of the LTR-driven leader-RNA luciferase constructs could be that modulation of the 5’ss activity may have an impact on the polyadenylation signal that is present in the upstream HIV-1 sequences. In fact, the polyadenylation signal is part of the repeat (R) region that is present at both the 5’ and 3’ end of the HIV-1 RNA genome. Binding of U1 snRNA to the 5’ss has been demonstrated to suppress the 5’ polyadenylation signal (22, 23), which is part of the regulatory mechanism to control selective usage of the 3’ polyadenylation signal. Thus, improved recognition of the 5’ss (as in the A and B mutants) could result in further suppression of 5’ polyadenylation and thus indirectly affect the luciferase mRNA and protein levels. To avoid this indirect effect, luciferase constructs with a 1-nt substitution in the 5’ polyadenylation signal (AAUAAA to AACAAA) were tested. This mutation completely inactivates the 5’ polyadenylation signal, which prevents any premature polyadenylation of the luciferase transcripts (24). The activity profiles of the new constructs (Fig. 8) were similar to that of the original constructs (Fig. 3D and Fig. 7B). Also with the new constructs, the SD-hairpin stabilizing mutations reduced splicing slightly (J2) or significantly (J7), which resulted in an increased luciferase level (Fig. 8A). The destabilizing A and B mutations again increased splicing and reduced the luciferase level, whereas the C mutation had an opposite effect (Fig. 8B). Therefore, we conclude that the differential behavior of the 5’ss mutants is not due to indirect polyadenylation effects.
Figure 8. The stability of the SD hairpin influences splicing. 293T cells were transfected with pLTR-gag-flag-luc constructs in which the SD hairpin was either stabilized (A) or destabilized (B) and in which the HIV-1 5' polyadenylation signal (encoded by the R region in the 5' LTR) was inactivated through a 1-nt mutation (AAUAAA to AACAAA). When indicated, cells were co-transfected with a Tat-expressing plasmid (plus Tat). The intracellular luciferase production was measured after 48 h. The mean and SD of 3 independent experiments are shown. Mock, control cells transfected with the empty pBluescript plasmid instead of the pLTR-gag-flag-luc construct. Statistical analysis was used to compare the wt and mutant values (*, p < 0.05).

Discussion
The major 5' ss in the HIV-1 leader RNA is involved in the production of all spliced viral RNAs, yet should be ignored to produce the unspliced RNA genome. The sequence surrounding this 5' ss can fold into a stem-loop structure and our results indicate that this SD hairpin structure modulates the splicing efficiency, most likely by influencing U1 snRNA annealing that initiates the splicing process. We demonstrated that destabilization of the SD hairpin increases 5' ss usage (A, B and AB mutants), which demonstrates that the natural (wt) stability inhibits splicing. By occlusion of the 5' ss in a hairpin structure, HIV-1 may prevent excessive splicing of the primary transcript and allow the regulation of alternative splicing. Further stabilization of the SD hairpin further reduces 5' ss usage (J2 and J7 mutants). Both increased and reduced levels of splicing had a negative impact on virus replication. These results suggest that HIV-1 splicing is delicately balanced for the production of the optimal amount of the different viral RNAs and proteins. Even slight deviations from this balance seem to influence virus replication.

The SD-hairpin mutations introduced in the LTR-leader RNA-luciferase reporter constructs (Fig. 3 and 7) could also influence other gene expression processes, like RNA production, polyadenylation and translation. Significant effects of the mutations on translation are unlikely because these sequence changes did not create a novel AUG start codon that could usurp scanning ribosomes (25, 26). Alternatively, the stabilized RNA structures could block the scanning ribosomes and reduce luciferase production, but we observed an increased luciferase level for the stabilized J2 and J7 mutants. Similarly, the destabilizing mutations could allow a more progressive movement of the ribosomes along the RNA and increase luciferase production, but we observed a decreased luciferase level for the A and B mutants. Because binding of U1 snRNA to the 5' ss has been demonstrated to suppress premature polyadenylation at the polyadenylation signal present in the 5' leader RNA (22, 23), we included LTR-leader luciferase constructs in which the 5' polyadenylation signal
was inactivated by mutation. The activity profile of these constructs (Fig. 8) was similar to that of the original constructs (Fig. 3D and Fig. 7B), which demonstrated that the differential behavior of the SD-hairpin mutated constructs is not due to indirect polyadenylation effects. It was previously also shown that a 5′ss, when positioned in close proximity to the promoter region, can stimulate transcription initiation, possibly through the recruitment of transcription factors by U1 snRNA (27-30). The mutations that we introduced did not affect the 5′ss sequence, but their RNA structure effect could indirectly influence transcription by affecting the binding of U1 snRNA to the 5′ss. However, Northern blot analysis of the RNAs produced by the LTR-leader luciferase constructs did not show large differences in the total RNA signal for the wt and mutant variants (Fig. 3C). Moreover, the stabilizing J2 and J7 mutations would reduce U1 snRNA binding and HIV-1 transcription in this scenario, whereas we observed an increase in luciferase production (Fig. 3D). Similarly, the destabilizing A and B mutations would increase U1 snRNA binding and transcription, whereas we observed reduced luciferase levels (Fig. 7B). Notably, the J1 mutation that did affect the 5′ss sequence, prevented splicing completely but did not block transcription (Fig. 3C) and resulted in an increased luciferase level (Fig. 3D). These results suggest that the SD hairpin mutations did either not affect transcription or that such an effect was relatively small compared to the splicing effect.

The sequence around the major 5′ss is highly conserved in different HIV-1 group M isolates (31), but some sequence variation is present. Importantly, nearly all alternative nts are compatible with the SD hairpin structure (Fig. 9A). For example, a G-to-A change is frequently observed at position 282, but in nearly all cases a simultaneous C-to-U change is observed at position 300. As a result of this bp co-variation, the G-C bp is substituted by an A-U bp. The frequently observed A-to-G nt change at position 286 is also well tolerated in the SD hairpin structure because an A-U bp is replaced by a G-U bp. Some isolates have multiple nt changes in the SD region that at first glance are not compatible with the hairpin structure. However, Mfold structure prediction indicated that these isolates can fold a slightly different hairpin (Fig. 9B-C). The strong conservation of a structure around the 5′ss supports an important role in HIV-1 replication. As shown by SHAPE-directed RNA structure determination, the closely related SIV mac239 isolate (32) can also fold a similar stem-loop structure that embeds the major 5′ss. Modulation of 5′ss usage by RNA structure may therefore be a general characteristic of primate lentiviruses.

In a previous study, we showed that stabilization of the SD hairpin through deletion of a 1-nt bulge and extension of the lower stem region with 1 bp almost completely blocked 5′ss-mediated splicing and virus replication (J7 mutant). We now observed that stabilization of the hairpin by extending the lower stem (J2 mutant) only modestly reduced splicing and replication. These results may indicate that the base pairing in the upper part of the stem-loop structure, which is the region to which the U1 snRNA anneals, is more critical for 5′ss restriction than base pairing in the lower part of the stem. It has previously been reported that annealing of other RNA molecules, e.g. siRNA inhibitors (33), can be restricted by the occlusion of the target sequence in a
stable secondary structure. Also in these studies, masking of the target region that is directly involved in the initial annealing step demonstrated the largest impact. An alternative explanation for the difference between the J7 and J2 phenotypes may come from the suggestion that U1 snRNA binds with 11 nt - instead of 9 nt - to a 5' ss (34-36). This will extend the binding site with 2 nt on the 3' side. Accordingly, the 1-nt bulge (A-296) that was deleted in the stabilized J7 mutant would be part of the U1 snRNA binding site. This binding site modification may have contributed to the severe splicing defect of this mutant compared to the equally stabilized J2 mutant. The extended 5'ss sequence would also include the C-297 nt that was mutated in the C mutant, which may explain why this destabilizing mutation surprisingly decreased splicing efficiency, whereas the other destabilizing mutations increased splicing. Alternatively, this C mutation may have affected another splice signal, although we carefully designed our mutants to avoid the alteration of important sequence elements. For example, this mutation may have affected the binding site for an SR protein involved in 5'ss usage (21).

Figure 9. The SD stem-loop structure is conserved among HIV-1 isolates. The major 5'ss region of 111 HIV-1 group M isolates (as aligned in the HIV Sequence Compendium 2013) (31) was analyzed for the capacity to fold the SD hairpin. (A) The sequence at the 5'ss region of 103 isolates is identical or very similar to that of the subtype B LAI strain used as prototype in this study and thus folds the same or a very similar structure. The sequence variation observed in the different subtypes and recombinants is shown. The labeling refers to the subtype (cpx, complex) and the accession number of the isolate (B4, JX140658; BG1, FJ670522; BF2, DQ085876; A6, AF286238; B12, JF320019; B15, AF086817; C1, JX140663; C3, EU884500; C5, EF469243; C8, AB254141; F4, JX140671; F7, GQ290462; G4, GU362882; cpx3, AF460972; cpx5, AM851091; BC3, AY727527; D1, K03454; D2, JX140670; D4, DQ054367; D5, AB485648; D7, EF633445; DF1, AF193253; A2D, AF286239). Some sequence changes occur simultaneously and isolates with bp co-variations are indicated in bold. (B) The subtype H isolates H1 (AF190127), H2 (AF190128) and H4 (FJ711703) have a 2-nt insertion (boxed nts) and fold a slightly different hairpin structure that includes an additional bp at the stem. (C) The subtype D isolates D3 (AY322219) and D6 (JX236672), and the subtype CD recombinant CD1 (AF289548) differ at 5 nt positions (boxed nts) from the LAI sequence, but can fold a similarly stable hairpin structure.
The SD stabilizing J2 mutation and the destabilizing A, B and AB mutations reduced viral replication fitness. However, virus replication was certainly not abrogated and we needed competition experiments to confirm the small effect on viral fitness. In contrast, the J7 mutation severely reduced virus replication. Estimation of the splicing frequency in the different leader RNA-luciferase constructs showed that splicing at the wt SD hairpin was fairly efficient, causing splicing in ~85% of the transcripts (Fig. 3E). Splicing of the J2 and J7 variants was reduced to ~74% and ~13%, respectively, whereas splicing of the A, B and AB mutants was increased to ~93% (Fig. 7C). Thus, the effect on the spliced:unspliced ratio is much larger for the J7 mutant (13:87) than for J2 (74:26) and the set of A, B and AB mutants (93:7) when compared to wt (85:15), which can explain the more dramatic effect of the J7 mutation on virus replication.

In murine leukemia virus (MLV), a simple retrovirus, splicing at the major 5’ss is also controlled by local RNA structure (37). This 5’ss is part of a complex RNA structure that restricts 5’ss accessibility and U1 snRNA binding. Modulation of the activity of the major 5’ss by means of RNA structure may thus be a general retrovirus property. Also in cellular transcripts, RNA structure has been shown to affect splicing efficiency and to have an important role in the regulation of alternative splicing (16, 17, 19, 38-43). For example, alternative splicing of the mRNAs for the microtubule-associated protein Tau is controlled by a stem-loop structure that includes the 5’ss of exon 10. Mutations that disrupt this stem-loop structure do increase 5’ss usage and have been associated with frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17). In these patients, exon 10 is more frequently incorporated in the tau mRNAs, which distorts the delicate balance between different Tau isoforms (44-46).

RNA structure also influences 3’ss usage in HIV-1. The 3’ss that is used in the production of tat mRNAs is located in a stem-loop structure that is highly conserved in different HIV-1 and SIVcpz strains (47). Destabilization of this structure increased 3’ss usage. Further stabilization reduced splicing, but this effect was lost when the sequence of the 3’ss was optimized, which confirms that the combination of sequence and structure elements determines the strength of a particular splice site (47). To add complexity, the activity of splicing silencer and enhancer elements can also be influenced by RNA structure (48). For example, it has been demonstrated that the binding of hnRNP A1 to a splicing silencer motif downstream of the 3’ss used in the production of the completely spliced tat, rev and nef mRNAs was reduced by the folding of this RNA region in a stable secondary structure (49-51). Interestingly, Jablonski et al. demonstrated that an RNA structure located in the HIV-1 env region regulates splicing of a cryptic exon (52). A single, naturally occurring nt substitution in this region affected the RNA structure, which resulted in the inclusion of this cryptic exon and reduced viral infectivity. It remains unclear whether this altered RNA structure affects the binding of splicing factors or whether it disrupts another mechanism that is important for splicing (52).

Taken together, our results indicate that splicing in HIV-1 RNA is partially suppressed by the SD stem-loop structure that encompasses the major 5’ss. All virus variants with a stabilized or destabilized SD hairpin replicated less efficiently than the wt
virus, which demonstrates that the stability of the SD hairpin is fine-tuned to reach an optimal balance between spliced and unspliced mRNAs for efficient virus replication. Other mechanisms, like suboptimal splice signal sequences, may contribute to this splicing control. Modifying the balance between spliced and unspliced RNAs could be an interesting therapeutic approach to inhibit HIV-1 replication. Targeting the HIV-1 leader RNA region with small molecules, for example antisense oligonucleotides, that modify the SD hairpin structure will interfere with HIV-1 splicing and reduce viral fitness (53). Splicing could also be influenced by antisense oligonucleotides that block the 5’ss (54) or target splicing cis-elements (55-59) or by modified U1 snRNAs that bind more efficiently to the 5’ss site and increase splicing (60). Alternatively, cellular splicing factors could be targeted with small molecule drugs. Such cellular factors could be proteins (e.g. SR proteins) that bind to splicing enhancer elements and regulate the splicing efficiency. Selective binding of these drugs to the proteins could prevent early steps in splicing and block HIV-1 replication (61, 62).

Conclusion
Regulation of splicing in HIV-1 is essential for the balanced production of all viral RNAs and proteins. We show that splicing in HIV-1 RNA is partially suppressed by the SD stem-loop structure that encompasses the major 5’ss. HIV-1 uses RNA structure to modulate the level of splicing and the stability of the SD hairpin structure is fine-tuned for optimal virus replication.

Acknowledgements
We thank Stephan Heynen for performing CA-p24 ELISA. This research was sponsored by the Netherlands Organisation for Scientific Research (Chemical Sciences Division; NWO-CW; TOP grant).

Material and Methods
DNA constructs
The pLAI plasmid encodes the full-length proviral genome of the HIV-1 LAI strain (63). The pLTR-gag-flag-luc plasmid contains the HIV-1 5’ LTR promoter region, the leader RNA and adjacent Gag sequences (position +336 to +410 with +1 as the transcription start site) encoding the 25 N-terminal amino acids, an in-frame Flag-tag coding sequence and the firefly luciferase open reading frame (ORF) (18). The J2 and J7 variants of pLAI and pLTR-gag-flag-luc were previously described (64).

Luciferase constructs with a mutated SD hairpin sequence were created by mutagenesis PCR (65) with pLTR-gag-flag-luc as template. A first PCR (PCR1) was performed with primers TA014 (64) and Q (wt fragment) or C (mutant C) (Table 1). PCR2 was performed with the primers TA033 (26) and R (wt), A, B or AB (mutants A, B or AB respectively). The PCR1 and PCR2 products were purified and combined in a third PCR with the outer primers TA033 and TA014. For the A, B and AB mutant, the wt PCR1 fragment was combined with the A, B or AB PCR2 fragment. For the C, AC, BC and ABC mutants, the C PCR1 fragment was combined with the wt, A, B or AB PCR2 fragment, respectively. The resulting PCR3 products were HindIII and NcoI digested and ligated into the corresponding sites of the pLTR-gag-flag-luc vector.

To mutate the HIV-1 polyadenylation signal in pLTR-gag-flag-luc, this plasmid was used as template for mutagenesis PCR with primers TA081 and polyAmut (Table 1). The PCR product was digested with BglII.
and HindIII and ligated into the corresponding sites of pLTR-qag-flag-luc, which resulted in pLTR-pA\textsuperscript{wt}-qag-flag-luc. To combine the SD and polyA signal mutations, the HindIII-NcoI fragment of the SD-mutated pLTR-qag-flag-luc constructs was used to replace the corresponding fragment of pLTR-pA\textsuperscript{wt}-qag-flag-luc. To introduce the SD mutations into pLAI, the mutated region of the pLTR-qag-flag-luc constructs was PCR amplified with the primers TATA-1 and TA115 (Table 1). The resulting PCR fragments were digested with BshIII and Clai and ligated into the corresponding sites of pLAI.

The pRL-CMV plasmid contains the Renilla luciferase reporter gene under the control of the cytomegalovirus immediate early (CMV-IE) promoter (Promega). The plasmid pTat-exon expresses the HIV-1 LAI Tat protein under the control of the CMV-IE promoter (66).

Table 1. Oligonucleotides used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence\textsuperscript{a}</th>
<th>Position\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>GGTGAGTACGCCAATTTTGTACACTAGCG</td>
<td>+289 to +317</td>
</tr>
<tr>
<td>R</td>
<td>ACTCACGACGTGCGCTCCCCTCGCC</td>
<td>+295 to +272</td>
</tr>
<tr>
<td>A</td>
<td>ACTACACAGCTGCTCCCCTCGCC</td>
<td>+295 to +272</td>
</tr>
<tr>
<td>B</td>
<td>ACTACACAGCTGCTCCCCTCGCC</td>
<td>+295 to +272</td>
</tr>
<tr>
<td>AB</td>
<td>ACTACACAGCTGCTCCCCTCGCC</td>
<td>+295 to +272</td>
</tr>
<tr>
<td>C</td>
<td>GTGGAATAGCAGAAAAATTTTGTACAGCG</td>
<td>+289 to +317</td>
</tr>
<tr>
<td>TATA-1</td>
<td>CAGATGCTGCATATAAGCAGCT</td>
<td>-39 to -18</td>
</tr>
<tr>
<td>TA081</td>
<td>ATGATTACGCAAAGCTCG</td>
<td>-708 to -691</td>
</tr>
<tr>
<td>TA115</td>
<td>AGATCCAGAGGAGTTCTAT</td>
<td>+1043 to +1026</td>
</tr>
<tr>
<td>LAI5Xba</td>
<td>TGGAAAGGCTAATTCACTCC</td>
<td>-454 to -434</td>
</tr>
<tr>
<td>polyAmut</td>
<td>AGGCAAGCTTGTGAGGCCTA</td>
<td>+86 to +65</td>
</tr>
<tr>
<td>A6838</td>
<td>CACCCGCTGACATGATAATG</td>
<td>-32 to +21</td>
</tr>
<tr>
<td>MO13</td>
<td>TCCCCGCTTAATACCTGA</td>
<td>+368 to +351</td>
</tr>
<tr>
<td>AH-AUG368</td>
<td>TCCCCGCTTAATACCTGACG</td>
<td>+368 to 348</td>
</tr>
</tbody>
</table>

\textsuperscript{a} nt numbers refer to the position in the HIV-1 LTR region with +1 as the transcription start site, except for the primer polyAmut that anneals to plasmid sequences upstream of the LTR region.

\textsuperscript{b} nt numbers refer to the position in the wt sequence.

Cell and virus cultures
293T cells were cultured at 37°C and 5% CO\textsubscript{2} in DMEM medium containing 10% fetal bovine serum (FBS), nonessential amino acids, 40 units/ml penicillin and 40 µg/ml streptomycin (67). Cells were transfected by calcium phosphate precipitation as previously described (67). For the production of virus stocks, 293T cells (~1.5 x 10\textsuperscript{5} cells per 2 cm\textsuperscript{2}, 60% confluency) were transfected with 1 µg of the wt or mutant pLAI constructs and the culture supernatant was harvested after 48 h. Virus production was measured by CA-p24 ELISA (68).

SupT1 T cells were grown at 37°C and 5% CO\textsubscript{2} in RPMI1640 advanced medium containing 1% FBS, 15 units/ml penicillin and 15 µg/ml streptomycin. SupT1 cells (5 x 10\textsuperscript{6} cells in 5 ml medium) were transfected with 1 µg wt or mutant pLAI plasmid by electroporation, as previously described (67), or infected with an equal amount of 293T-produced virus (corresponding to 0.25 ng CA-p24). Cells were split into fresh medium twice a week. Competition experiments (69) were started by transfection of an equal amount of the virus constructs. At the peak of infection, when massive syncytia were present, the virus was passaged onto fresh cells. The integrated proviral DNA was isolated from cells at the 2\textsuperscript{nd} passage, as described previously (70). The LTR region was PCR amplified with primers AH-GAG-Kpnl (66) and LAI 5’Xba and sequenced with the primer AH-GAG-Kpnl.

Luciferase assay
293T cells (~1.5 x 10\textsuperscript{5} cells per 2 cm\textsuperscript{2}) were transfected with 20 ng wt or mutant pLTR-qag-flag-luc, 0.5 ng pRL-CMV (as internal control), 5 ng pTat-exon or the empty pcDNA3 vector and 485 ng pBluescript (as carrier DNA). The cells were cultured for 48 h and subsequently lysed in passive lysis buffer (Promega). Firefly and Renilla luciferase activities were determined with the dual-luciferase assay (Promega) and the firefly luciferase activity was normalized to the Renilla luciferase activity. Data were corrected for between-session variation with the factor correction software (71).
Splice site analysis
293T cells (~1.5 x 10^5 cells per 2 cm^2) were transfected with 1 µg wt or mutant pLTR-gag-flag-luc construct and 50 ng pTat-exon or pcDNA3. After 48 h, cells were washed with phosphate-buffered saline (PBS), total cellular RNA was isolated with the RNeasy kit (QIAGEN) and contaminating DNA was removed with RNase-free DNase during isolation (QIAGEN). The RNA was used as template for cDNA synthesis with oligo dT and random hexamers as primer (ThermoScript RT-PCR system; Invitrogen). The cDNA product was amplified by PCR with primers TA033 (26) and TA113 (18) as previously described (18). The PCR products were analyzed by agarose gel electrophoresis.

Northern Blot analysis
5 µg RNA was mixed with denaturing loading dye (final concentration: 40 mM MOPS pH 7.0, 10 mM Na-acetate, 5% formaldehyde, 0.05 mg/ml ethidium bromide, 0.5 mg/ml orange G, 7 g/ml sucrose) followed by electrophoresis in a 1% agarose gel in MOPS buffer (40 mM MOPS, 10 mM Na-acetate pH 7.0) with 7% formaldehyde at 100 V for 4 h. The RNA was transferred onto a positively charged nylon membrane (Roche) using 20 x SSC (3.0 M NaCl, 0.3 M Na-citrate pH 7.0) overnight by capillary force. The RNA was cross-linked to the membrane by an UV crosslinker (Stratagene). The membrane was incubated for 1 h in ULTRAhyb (Ambion) at 55°C. The probe, consisting of the NcoI and XbaI digested luciferase fragment of pLTR-gag-flag-luc, was denatured for 10 min at 96°C and labeled with [α-^32P]dCTP by using the High Prime DNA labeling kit (Roche). The probe was added to the prehybridized membrane and subsequently hybridized for 16 h at 55°C. The membrane was washed twice for 15 min at room temperature in low-stringency buffer (2x SSC, 0.2% SDS) and twice for 30 min at 60°C in high-stringency buffer (0.1×SSC, 0.2% SDS). Images were obtained using the Storm 860 phosphorimager (Amersham Biosciences). The size of the luciferase transcripts was estimated using the rRNAs as marker. The amount of 18S and 28S rRNA was used as a loading control and was similar for all samples.

RNA secondary structure prediction
The RNA secondary structure of the SD hairpin region was analyzed using the Mfold algorithm (version 3.4) offered by the MBCMR Mfold server (http://mfold.burnet.edu.au) using standard settings (72).

SHAPE-directed RNA structure determination
For the production of wt and mutant leader RNA fragments (nts +1 to +368, with +1 being the HIV-1 transcription start site), pLAI plasmids containing the wt or mutated SD hairpin sequence were used as a template in a PCR with primers A6838 and MO13. The PCR product was purified using the Gel and PCR Clean-up Kit (Macherey-Nagel), ethanol-precipitated and dissolved in RNase-free water. The PCR products were in vitro transcribed using the MEGashortscript T7 Kit (Ambion) for 16 h at 37°C after which the DNA template was removed by DNase treatment. The RNA was purified from a denaturing 4% polyacrylamide gel and the full-length RNA transcript was excised and eluted from gel for 16 h at 30°C in 0.5 M ammonium acetate, 10 mM EDTA and 0.5% SDS. The RNA was extracted once with phenol-chloroform-isoamylalcohol (25:24:1 v/v; Invitrogen), ethanol-precipitated and dissolved in RNase-free water. RNA folding was performed according to Wilkinson et al. (73). 6 pmol RNA was dissolved in 9 µl RNase-free water and 9 µl 2x dimer-buffer (100 mM HEPES pH8.0, 400 mM KAc pH8.0, 10 mM MgCl₂) was added. RNA was denatured at 90°C for 3 min, snap-cooled on ice for 1 min and incubated at 37°C for 1 h to allow folding of the RNA. Chemical modification and primer extension of the RNA were performed essentially as described by Mortimer and Weeks (74). In short, 1 µl 400 mM BzCN (Sigma-Aldrich) in DMSO or 1 µl DMSO was added to 9 µl folded RNA-sample and mixed vigorously by pipetting at room temperature for 3 s. BzCN is rapidly hydrolyzed by water (half-life of 0.25 s at 37°C), which makes a quenching step unnecessary (75). RNA was mixed with 90 µl water, 5 µl 4 M NaCl, 0.5 µl (7.5 µg) GlycoBlue (Ambion) and 400 µl ethanol and precipitated by centrifugation at 21,000 x g and 4°C for 20 min. The RNA pellet was washed twice with 70% ethanol and resuspended in 8 µl water. BzCN and DMSO-treated RNA was mixed with 1 µl (1 pmol) FAM-labeled AH-AUG368 primer (Eurogentec) and 1 µl 10 mM dNTPs (dATP, dCTP, dGTP and dTTP), heated at 85°C for 2 min and 65°C for 5 min to anneal the primer, and subsequently snap-cooled on ice for 1 min. 1 µl 100 mM DTT, 4 µl 5x First-Strand buffer (Invitrogen) and 1 µl (200 U/µl) SuperScript III RT (Invitrogen) were added and the mixture was incubated for 1 min at 45°C and 60 min at 55°C for primer extension, and for 15 min at 70°C to inactivate RT. The RNA template was degraded by addition of 1 µl (2 U) RNaseH for 20 min at 37°C, after which 4 µl 50 mM EDTA pH8.0 was added. For the sequencing ladders, 3 pmol RNA in 10 µl water was mixed with 1 µl (0.3 pmol) NED-labeled AH-AUG368.
primer (Eurogentec), 1 µl 10 mM dNTPs, 1 µl 10 mM ddCTP (Trilink Biotechnologies), 1 µl 100 mM DTT, 4 µl 5x First-Strand buffer and 1 µl (200 U/µl) SuperScript III RT. Primer extension and subsequent RNA template degradation were performed as described for the BzCN/DMSO samples. Each BzCN or DMSO-treated sample was mixed with a sequence ladder and the cDNA was precipitated with 100 µl ethanol in the presence of 0.5 µl GlycoBlue. The pellet was washed twice with 70% ethanol and dissolved in 20 µl deionized formamide with 0.5 µl GeneScan™ 500 ROX™ Size Standard (Applied Biosystems). Samples were analyzed by capillary electrophoresis (Applied Biosystems AB3130 instrument). QuSHAPE software was used to align the sequences, calculate the relative peak areas and determine the reactivity of individual nts as described (76). RNA structures were predicted using the RNAstructure v5.6 software (77) in which the obtained SHAPE reactivity data were used as pseudo-free energy constraints. RNA structures were drawn using XRNa software (http://rna.ucsc.edu/rnacenter/xrna/xrna.html).

**Statistical analyses**
Data were subjected to independent-samples T-test analysis (IBM SPSS version 21) where indicated in the figure legends. The obtained p-values were considered significant when p<0.05.

**References**


23. Ashe MP, Pearson LH, Proudfoot NJ. 1997. The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. The EMBO journal 16:5752-5763.


30. Alexander MR, Wheatley AK, Center RJ, Purcell DF. 2010. Efficient transcription through an intron requires the binding of an Sm-type U1 snRNP with intact stem loop II to the splice donor. Nucleic acids research 38:3041-3053.


Supplementary Figure 1. SHAPE-directed RNA structure probing of wt and SD-mutated leader RNAs. The SHAPE reactivity of leader RNA transcripts (+1 to +368) was determined in 2-5 independent experiments. The structure of the region that folds the DIS, SD and ψ hairpins (+243 to +331) is shown. The color of the nt indicates its reactivity (black: low reactivity, 0 - 0.3; green: medium reactivity, 0.3 - 0.8; red: high reactivity, > 0.8). The DIS loop nts GCGCGC are not reactive because they are involved in the RNA dimer interaction.