Regulation of HIV-1 splicing
Müller, Nancy

Citation for published version (APA):

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 3

HIV-1 splicing is controlled by local RNA structure and binding of splicing regulatory proteins at the major 5’ splice site

Nancy Mueller, 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

Journal of General Virology, 2015, Volume 96, Pages 1906-1917
Abstract
The 5’ leader region of the HIV-1 RNA genome contains the major 5’ splice site (ss) that is used in the production of the many spliced viral RNAs. This splice-donor (SD) region can fold into a stable stem-loop structure and the thermodynamic stability of this RNA hairpin influences splicing efficiency. In addition, splicing may be modulated by binding of splicing-regulatory proteins, in particular SR proteins SRSF1 (SF2/ASF), SRSF2 (SC35), SRSF5 (SRp40) and SRSF6 (SRp55), to sequence elements in the SD region. The role of RNA structure and SR protein binding in splicing control was previously studied by functional analysis of mutant SD sequences. The interpretation of these studies was complicated by the fact that most mutations simultaneously affect both structure and sequence elements. We therefore tried to disentangle the contribution of these two variables by designing more precise SD region mutants with a single effect on either the sequence or the structure. The current analysis indicates that HIV-1 splicing at the major 5’ss is modulated by both the stability of the local RNA structure and the binding of splicing regulatory proteins.
Introduction

In eukaryotic cells, primary RNA transcripts are co-transcriptionally spliced in the spliceosome complex to remove intronic sequences that are located between splice donor and splice acceptor sites (5'ss and 3'ss, respectively). Splicing events between alternative 5'ss and 3'ss positions allow the formation of variant mRNAs and contribute to the proteome complexity. Formation of the spliceosome complex is initiated by binding of U1 snRNP to the 5'ss, which depends on sequence complementarity between the 5'ss and 5' end of the U1 snRNA component of U1 snRNP. In addition, splicing efficiency is controlled by splicing regulatory elements (SREs) to which regulatory proteins bind that either increase or decrease the recruitment of spliceosome components (1). It was initially described that SR proteins with a characteristic serine and arginine rich domain enhance splicing, whereas hnRNP would specifically inhibit splicing (1, 2). More recently, it was described that SR proteins and hnRNPs can both activate and repress splicing depending on their position relative to the 5'ss (3-8). The local RNA structure at the 5'ss can also influence binding of spliceosome components and modulate splice site usage (9-20).

HIV-1 is an attractive model to study the regulation of splicing because this virus produces a single primary transcript and numerous differentially spliced mRNAs. The unspliced transcript (~9 kb in size) is used as genomic RNA that is packaged in virus particles and as mRNA for the production of Gag and Pol proteins. This RNA encodes several 5’ss and 3’ss, and alternative splicing creates singly (~4 kb) and multiply (~2 kb) spliced mRNAs that are used for translation of the other 7 viral proteins. Splicing has to be strictly regulated to allow the balanced production of all viral RNAs and proteins. The HIV-1 transcript contains several SREs to enhance or inhibit splicing at nearby positioned splice sites (21-27). Furthermore, usage of HIV-1 splice sites can be influenced by local RNA structure (28-31).

The major 5'ss is used for the production of all spliced mRNAs and its control is thus key for splicing regulation. The 11-nucleotide (nt) 5'ss (5'CUGGUGAGUAC3') has high sequence complementarity to the 5' end of U1 snRNA (3'GUCCAUUCAUA5'), which will allow efficient U1 snRNA annealing and splicing. Both splicing regulatory proteins and local RNA structure may have a role in controlling splicing efficiency. Using ESEfinder, an algorithm based on SELEX-identified RNA binding sites for splicing regulatory proteins (32, 33), Asang et al. (34) recently predicted binding of the SR proteins SRSF1 (SF2/ASF), SRSF2 (SC35), SRSF5 (SRp40) and SRSF6 (SRp55) to the major 5'ss region (Fig. 1B). Mutations in the putative SRSF2 binding sites did indeed affect SRSF2 binding and increased the splicing efficiency, whereas mutations in the SRSF5 binding site reduced splicing. We recently described that splicing at the major 5'ss is also controlled by RNA structure. Sequences surrounding the major 5’ss can fold into a stem-loop structure, the SD hairpin, that partially occludes the 11-nt sequence to which U1 snRNA anneals (Fig. 1B). Different HIV-1 isolates and the related SIVmac93 can fold a similar 5'ss structure (30, 35). This evolutionary conservation supports the idea that this structure has an important role in viral replication. Destabilization of this hairpin structure increased splicing, which indicates that the wild-type (wt) structure suppresses splicing (29, 30). We demonstrated that such destabilization of the SD
hairpin reduced virus replication. Stabilization of the hairpin, which further reduced splicing, also inhibited virus replication. Apparently, HIV-1 splicing at the major 5'ss is fine-tuned by SD hairpin stability for optimal replication. The RNA structure likely modulates splicing by controlling accessibility for the splicing machinery, including the U1 snRNP and SR proteins. On the other hand, SR protein binding may also affect the SD hairpin stability. Thus, an intricate interplay between local RNA structure and RNA-binding proteins may control splicing at the major 5'ss.

To study the regulatory role of RNA structure in HIV-1 splicing, we previously modified the SD hairpin stability by mutation, followed by measurement of the splicing efficiency. The introduced mutations were carefully chosen not to affect the U1 snRNA binding site but may have affected other sequence elements, in particular the putative SR-protein binding sites (34). We therefore designed specific sets of SD region mutants that allowed us to unravel the role of RNA structure and sequence elements in splicing regulation.

Results

Mutations in the SD hairpin sequence affect splicing efficiency

In our previous study that aimed to analyze the role of SD hairpin stability in splicing control, we introduced nt substitutions in the left hand side of the lower stem region (mutants A<sub>L</sub>, B<sub>L</sub> and AB<sub>L</sub>; Fig. 1A) (30). These mutations destabilized the RNA hairpin but also changed the accessibility of 5'ss nts for interaction with U1 snRNA (Fig. 1B). In particular, the cleavage site (indicated with ▲ in Fig. 1B) that was presented in the loop of the wt hairpin was occluded in the stem region of the mutants. We therefore created new SD hairpin variants in which nts in the right-hand side of the stem were substituted (A<sub>R</sub>, B<sub>R</sub>, AB<sub>R</sub>). These mutations also reduced hairpin stability but did not (A<sub>R</sub>) or only slightly (B<sub>R</sub>, AB<sub>R</sub>) modify the presentation of the U1 snRNA binding site. We also combined the left- and right-hand side mutations (A<sub>LR</sub>, B<sub>LR</sub> and AB<sub>LR</sub>; Fig. 1A). These double mutants fold a wt-like hairpin with similar presentation of the 5'ss, but its stability was slightly less than the wt hairpin because C-G base pairs (bps) were replaced by A-U bps.

Quantitative analysis of splicing at the 5'ss in the complete HIV-1 context is complicated because of pleiotropic effects. For example, the regulatory Tat and Rev proteins are produced from fully spliced transcripts. Tat enhances transcription from the viral LTR promoter (36 and references therein) and influences splice site selection (37). Rev stimulates nuclear export of the unspliced and singly spliced viral RNAs. Moreover, Rev may influence other replication processes, like translation and encapsidation of viral RNAs (reviewed in (38)). Thus, both the level of the viral RNAs and the ratio between the spliced and unspliced RNAs will be influenced indirectly and in an uncontrolled way by changes in the initial splicing efficiency. Therefore, we used a simplified splicing reporter: an LTR-driven leader RNA-luciferase construct with the major 5'ss in the HIV-1 leader region and a 3'ss in the luciferase gene (Fig. 2A). We previously demonstrated that mutations in the SD hairpin that increase or decrease splicing of this reporter RNA also affect virus...
Figure 1. Mutation of the SD hairpin region. (A) The structure and thermodynamic stability (ΔG in kcal/mol) of the wt and mutant SD hairpin structures as predicted with the MFold RNA structure analysis software are shown. The nts of the major 5'ss, to which the 5' end of the U1 snRNA binds, are encircled in grey (▲, cleavage site). Nt substitutions are indicated in the wt structure (with arrows) and mutant structures (black boxes). (B) Alignment of the wt and mutant SD region sequences. The position of the SR protein binding sites is indicated. The base-paired nts are boxed in grey. Nt substitutions are boxed in black. The 5'ss nts are in bold (▲, cleavage site).
replication when introduced in the complete HIV-1 genome (29, 30). Furthermore, these SD mutations did not affect other gene expression processes, like RNA production, polyadenylation or translation, or the effects were relatively small compared to the observed splicing effects (30).

Reporter constructs with the wt or SD-mutated HIV-1 leader region were transfected into 293T cells. 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 at 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. 2A) and the PCR products were analyzed by agarose gel electrophoresis. With the wt and mutant constructs, both RNA products were detected. The 429-bp spliced RNA product was more abundantly present than the 710-bp unspliced RNA product, indicating that splicing at this 5’ss is very efficient, which is in agreement with previous results (30). The level of unspliced RNA seemed to be slightly decreased for the mutants when compared with wt, while the level of the spliced product was slightly increased (Fig. 2B). These results indicate that all tested mutations increase the splicing efficiency.

Because the RT-PCR analysis provides only semi-quantitative results, the effect of the mutations on splicing was also quantitatively analyzed by measuring the luciferase activity in transfected cells. Whereas the unspliced RNA encodes an active luciferase protein, the spliced RNA does not (Fig. 2A). All mutants showed reduced luciferase activity compared to wt, which is in agreement with the observed increase in splicing. Similar effects were observed when the cells were transfected without Tat plasmid and low-level transcription is driven by the basal LTR promoter activity (Fig. 2C). In our previous study (30), we demonstrated that the luciferase construct with a wt SD hairpin is efficiently spliced (~85% splicing). Based on the observed luciferase levels (Fig. 2C), we calculated that splicing of the A\textsuperscript{L}, B\textsuperscript{L} and AB\textsuperscript{L} mutants is increased to ~92-93% (Fig. 2D), which is in agreement with our previous observations (30). The A\textsuperscript{R}, B\textsuperscript{R} and AB\textsuperscript{R} mutants, which were similarly destabilized as the A\textsuperscript{L}, B\textsuperscript{L} and AB\textsuperscript{L} set, but differ in the presentation of the 5’ss nts, exhibited a similar increase in splicing efficiency. These results demonstrate that destabilization of the SD RNA structure increased splicing and that changes in the presentation of the U1 snRNA binding nucleotides (nts) had no or only a minor effect, which confirms that splicing is restricted by the wt SD hairpin stability. The A\textsuperscript{LR} and B\textsuperscript{LR} mutants in which the left- and right-hand side A and B mutations were combined demonstrated a similar increase in splicing as the single-side mutants, while the AB\textsuperscript{LR} demonstrated a further increase in splicing (~96%) compared to the single-side AB mutants. These results were surprising because we expected that partial restoration of the hairpin stability would at least partially restore the inhibitory effect of RNA structure on splicing. Possibly, underlying sequence elements that influence splicing were affected by the introduced mutations such that the double mutants were similarly or even more severely affected than the single mutants.
Asang et al. (34) predicted binding of several SR proteins to sequences overlapping or surrounding the 5’ss of the HIV-1 NL4-3 strain. Our constructs are based on the related HIV-1 LAI isolate. The 5’ss regions of NL4-3 and LAI differ only at the position immediately upstream of the SD hairpin sequence (position 281 in Fig. 1A; A in LAI, C in NL4-3) and the isolates can fold an identical SD hairpin structure and encode the same putative SR protein binding sites (Fig. 1B). Mutational analysis experiments by Asang et al. indicated that binding of the SR protein SRSF2 reduces splicing, whereas SRSF5 and SRSF6 stimulate splicing (34). To evaluate the effect of the introduced A, B and AB mutations on binding of these splice factors, we analyzed the wt and mutant SD sequences with the ESEfinder software (Table 1). The A\textsuperscript{L} mutation did not affect any of the predicted binding sites present in wt, but all other mutations did affect one or two of these sites. The AB\textsuperscript{LR} mutations resulted in the loss of two SRSF2
binding sites, which in combination with the relative low hairpin stability may explain
the very high splicing efficiency observed for this mutant. Thus, except for the A\textsuperscript{i} mutant, both the stability of the local RNA structure and the presence of SR protein
binding sites may have influenced splicing of the SD hairpin variants.

Table 1. SR protein binding sites, stability and splicing efficiency of wt and mutant SD hairpins.

<table>
<thead>
<tr>
<th></th>
<th>Splicing silencers*</th>
<th>Splicing enhancers*</th>
<th>(\Delta G) (kcal/mole)</th>
<th>splicing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>+ + + + + + +</td>
<td>+ + + + + +</td>
<td>-9.0</td>
<td>85</td>
</tr>
<tr>
<td>A\textsuperscript{i}</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>-5.2</td>
<td>92</td>
</tr>
<tr>
<td>A\textsuperscript{IK}</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>-6.3</td>
<td>89</td>
</tr>
<tr>
<td>A\textsuperscript{IK}</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>-6.9</td>
<td>93</td>
</tr>
<tr>
<td>B\textsuperscript{i}</td>
<td>- + + + + +</td>
<td>+ + + + + +</td>
<td>-5.2</td>
<td>93</td>
</tr>
<tr>
<td>B\textsuperscript{IK}</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>-5.5</td>
<td>93</td>
</tr>
<tr>
<td>B\textsuperscript{IK}</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>-7.7</td>
<td>92</td>
</tr>
<tr>
<td>AB\textsuperscript{i}</td>
<td>- + + + + +</td>
<td>+ + + + + +</td>
<td>-7.8</td>
<td>93</td>
</tr>
<tr>
<td>AB\textsuperscript{IK}</td>
<td>+ + + + + +</td>
<td>+ + + + + +</td>
<td>-5.3</td>
<td>96</td>
</tr>
</tbody>
</table>

\* SR protein binding sites as predicted with ESEfinder and described by Asang et al (34)

The stability of the SD hairpin structure influences splicing efficiency
We designed a new set of SD hairpin mutants in which exclusively the RNA stability
and not the predicted protein binding sites were affected. The 11-nt U1 snRNA
binding site was also not modified. To identify suitable mutants, the SR protein
binding capacity of SD hairpin sequences with all possible 1-nt substitutions was
analyzed with ESEfinder (Table 2). Most nt substitutions affected one or several of
the previously described SR protein binding sites or created a new binding site.

Apart from the A\textsuperscript{i} mutation, two 1-nt mutations were possible that did not affect
any putative SR protein binding site (T2 and T3 mutants in Table 2). As shown in
Fig. 3, these mutations destabilized the SD hairpin structure to a different extent
(\(\Delta G\textsubscript{T2} = -5.2\) and \(\Delta G\textsubscript{T3} = -7.1\) kcal/mol). Unfortunately, the T3 mutation resulted in
considerable complementarity to the 5’ end of U1 snRNA (\(^\text{5’}GGAGUGACUGG\text{3’};\) nts
complementary to U1 snRNA underlined), including a GU dinucleotide at position
4-5, which is an invariable characteristic of 5’ss sequences. To estimate the strength
of the new 5’ss, we calculated the hydrogen bond score (HBS), which reflects the base
pairing potential with U1 snRNA (39, 40). Whereas the wt 5’ss has an HBS value of
17.5, the new site has an HBS of 10.7, which indicates that this site could function
as a weak alternative 5’ss. The T2 mutation also created a GU dinucleotide, but
with modest complementarity to U1 snRNA (\(^\text{5’}GGAGUCGACUG\text{3’};\) HBS= 1.8), which
makes it unlikely that this sequence will function as 5’ss. Although the presence of
an alternative 5’ss may complicate the analysis, we included the T3 mutant in our
experiments because of the limited possibilities to destabilize the SD hairpin without
affecting putative SR protein binding sites.

To analyze the effect of these destabilizing mutations on 5’ss usage, T2 and
T3 mutated leader RNA-luciferase reporter constructs were generated. 293T
cells were transfected with the wt and mutated constructs and luciferase
Figure 3. Destabilization of the SD hairpin structure without affecting SR protein binding sites increases splicing. (A) Alignment of the wt and mutant SD hairpin sequences (see Fig. 1B for details). (B) The structure and thermodynamic stability (ΔG in kcal/mol) of the wt and mutant SD hairpin structures are shown (see Fig. 1A for details). (C) Luciferase production in cells transfected with the wt or mutant pLTR-gag-flag-luc plasmids with or without a Tat plasmid was analyzed as described in Fig. 2C. The mean and SD of 3 independent experiments are shown. (D) The splicing frequency of the wt and mutant constructs was calculated as described in Fig. 2D. (E) The thermodynamic stability (ΔG) of the wt, A, T2 and T3 SD hairpins was plotted against their splicing efficiency.
Table 2. Effect of single-nt substitutions on SR protein binding sites.

<table>
<thead>
<tr>
<th></th>
<th>Splicing silencers</th>
<th>Splicing enhancers</th>
<th>New sites</th>
<th>ΔG (kcal/mol)</th>
<th>Splicing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>SRSF2$^{(1)}$</td>
<td>SRSF2$^{(2)}$</td>
<td>SRSF2$^{(3)}$</td>
<td>SRSF5</td>
<td>SRSF6$^{(1)}$</td>
</tr>
<tr>
<td>AGGGACTGTGATACGCA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$-C-$------</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$-A-$------</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$-T-$------</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$-C-$------</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>A$^L$</td>
<td>$-A-$------</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>T2</td>
<td>$-T-$------</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>B$^L$</td>
<td>$-A-$------</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>T3</td>
<td>$-T-$------</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$-C-$------</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$-A-$------</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$-A-$------</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$-T-$------</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$-G-$------</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

SRSF1$^1$ as predicted using the original SRSF1 (IgM) matrix (32)
SRSF1$^2$ as predicted using the improved SRSF1 (IgM-BRCA1) matrix (33)
production was measured after 48 h to estimate the splicing efficiency. Both mutants showed a decrease in luciferase activity, which reflects an increase in splicing compared to the wt (Fig. 3C). The T2 mutated SD hairpin has a similar thermodynamic stability as the A1 mutant (ΔG = -5.2 kcal/mol) and the same increase in splicing was measured (92%) (Fig. 3D). The results obtained with the T2 mutant thus confirm that the stability of the SD hairpin influences splicing. The more moderately destabilized T3 variant (ΔG = -7.1 kcal/mol) demonstrated a smaller increase in splicing efficiency (90%), which is in agreement with the inverse correlation between RNA structure stability and the level of splicing (Fig. 3E). We did not investigate whether alternative splicing is induced for the T3 mutant. The observed correlation between RNA structure stability and splicing efficiency does suggest that if splicing occurred at the new 5’ss, this did not significantly affect the outcome of the experiment, possibly because both 5’ss sites are similarly suppressed by the same RNA structure.

**SR protein binding sites in the SD hairpin sequence affect splicing efficiency**

To critically test the effect of SR protein binding on splicing we generated another set of SD hairpin variants in which we tried to modify the predicted SR protein binding sites without RNA structure consequences. For this, base pair (bp) co-variations were introduced in the left- and right-hand side of the lower hairpin stem. In the X, Y, Z variants (Fig. 4A, B), a G-C bp was replaced by C-G or vice versa. Two or three of these co-variations were combined in the XY, YZ, XZ and XYZ variants. The thermodynamic stability of the mutated hairpins was similar to that of the wt hairpin (ΔG = -9.0 kcal/mol) or differed only slightly (ranging from -7.8 to -9.8 kcal/mol). ESEfinder predicted that the mutations removed at least one SRSF2 or SRSF6 binding site, except for the YZ mutant (Table 3). Some of the mutations created a new candidate SR protein binding site.

![Image](image.png)

**Table 3. SR protein binding sites, stability and splicing efficiency of SD hairpin mutants with wt-like stability.**

<table>
<thead>
<tr>
<th>Splicing silencers</th>
<th>Splicing enhancers</th>
<th>new sites</th>
<th>ΔG (kcal/mol)</th>
<th>% splicing</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRSF2(1)</td>
<td>SRSF2(2)</td>
<td>SRSF2(3)</td>
<td>SRSF5</td>
<td>SRSF6</td>
</tr>
<tr>
<td>wt</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>X</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Y</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Z</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>XY</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>YZ</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>XZ</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>XYZ</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Upon transfection of cells with the wt and mutated leader RNA-luciferase reporter constructs, the luciferase level was measured (Fig. 4C) and used to determine the splicing efficiency (Fig. 4D). Complex effects were scored that are indicative of both binding site and RNA structure effects. The Y and YZ hairpins have the same ΔG as wt and these mutants showed a slightly increased level of splicing (~88%), which can be explained by deletion of predicted binding sites for the splicing-inhibiting SRSF2 protein or creation of a binding site for the splicing-enhancing SRSF1 protein.
The XY, XZ and XYZ mutations, which all slightly destabilized the hairpin to the same level ($\Delta G = -8.6$ kcal/mol), increased splicing to different levels. Splicing was almost complete for the XY mutant (97%), which can be explained by deletion of two SRSF2 sites and creation of a single SRSF1 site. Splicing of the XZ and XYZ mutants was only moderately increased (90% and 92%, respectively). The XZ mutant holds the same putative binding sites as the XY mutant, except that one SRSF6 binding site was lost. This site was also absent for the XYZ mutant. SRSF6 binding stimulates splicing (41) and loss of this site may explain the relatively small increase in splicing efficiency of the XZ and XYZ mutants.

The same binding sites are predicted for the Y mutant as for the XY mutant (deletion of two SRSF2 sites and creation of a SRSF1 site). These mutants differ only slightly in hairpin stability ($\Delta G_y = -9.0; \Delta G_{xy} = -8.6$), but demonstrated a large difference in splicing efficiency (88% and 97%, respectively). These results indicate that slight destabilization of the hairpin can significantly increase splicing when the predicted SR protein binding sites are modified at the same time. The X mutation significantly increased splicing (95%) which may be due to reduced hairpin stability ($\Delta G = -7.8$ kcal/mol) in combination with the loss of two SRSF2 sites. The Z mutation slightly stabilized the SD hairpin ($\Delta G = -9.8$ kcal/mol), which in combination with the loss of a SRSF6 site, may have caused the observed reduction in splicing, despite the loss of an SRSF2 site. Taken together, these results indicate an intricate interplay of positive and negative protein binding sites and repressive RNA structure in the regulation of HIV-1 splicing.

Discussion

The SD hairpin in the HIV-1 leader RNA encompasses the major 5′ss that is used in the production of all spliced viral RNAs. We demonstrate that both the stability of the SD hairpin and underlying sequence elements to which SR proteins can bind control the level of splicing. Whereas previous reports described either RNA structure (30) or sequence effects (34), we designed 3 sets of mutants to disentangle the contribution of sequence and structure components.

In the first set, we compared hairpin variants with nt substitutions on the left, right, and both sides of the base-paired SD stem. The single mutants with a destabilized hairpin increased the splicing frequency as expected. Surprisingly, the double mutants that (partially) restored hairpin stability did not restore the wt phenotype, but increased splicing to the same or higher level as the single-side mutations. These results indicated that the mutations did not only exert a structure but also a sequence effect. In silico analysis of the mutated SD sequences for the presence of SR protein binding sites indicated that - except for the single-side A mutation - all mutations may indeed affect SR protein binding which obviously complicated the analysis.

In a second approach, we set out to design mutations in the lower SD stem that changed the RNA structure, but that did not remove existing or create new putative
Figure 4. Mutation of the SR protein binding sites without affecting the SD hairpin structure affects splicing. (A) Alignment of the wt and mutant SD hairpin sequences (see Fig. 1B for details). (B) The structure and thermodynamic stability (ΔG in kcal/mol) of the wt and mutant SD hairpin structures are shown (see Fig. 1A for details). (C) Luciferase production in cells transfected with the wt or mutant pLTR-gag-flag-luc plasmids with or without a Tat plasmid was analyzed as described in Fig. 2C. The mean and SD of 3 independent experiments are shown. (D) The splicing frequency of the wt and mutant constructs was calculated as described in Fig. 2D.
SR protein binding sites. This resulted in two 1-nt mutations (T2 and T3) that - like A1 - destabilized the hairpin structure without affecting SR protein sites. Both mutations increased the splicing efficiency and an inverse correlation was observed between the thermodynamic stability of the hairpin and splicing efficiency (Fig. 3E). These results confirmed the role of RNA structure in the modulation of HIV-1 splicing.

To analyze the role of the SR protein binding in splicing control, we designed a third set of hairpin mutants with a minimally affected RNA structure, but with mutations in the predicted SR protein binding sites. Comparison of the activity of 3 constructs with a wt hairpin stability (wt, Y and YZ; ΔG = -9.0 kcal/mol) and different configuration of putative SR protein binding sites indicated that splicing is most likely affected by the binding of SR proteins. Analysis of 3 variants with a slightly destabilized hairpin (XY, XZ, XYZ; ΔG = -8.6 kcal/mol) and differences in the SR binding-site configuration similarly implicate the role of SR protein binding in splicing control. These results are in agreement with previous observations by Asang et al. (34).

The Y and XY mutations similarly changed the predicted SR binding site configuration (loss of two SRSF2 splicing-inhibiting sites and creation of a candidate SRSF1 splicing enhancer). Whereas the corresponding SD hairpins differed only slightly in thermodynamic stability (ΔG_Y = -9.0; ΔG_XY = -8.6), a notable large difference in splicing efficiency was observed (88% and 97%, respectively). Thus, a relatively small destabilization (ΔΔG = 0.4 kcal/mol) resulted in nearly complete splicing (9% increase). In contrast, comparison of the hairpin stability and splicing activity of the wt (ΔG = -9.0; 85% splicing) and T3 variant (ΔG = -7.1; 90%), which both have a wt binding site configuration, reveals a smaller increase in splicing (only 5%) upon a more pronounced hairpin destabilization (ΔΔG = -1.9). Thus, a change in RNA stability translates into a smaller splicing effect in a wt SR protein binding site configuration than in a splicing-activated context. These results demonstrate the interplay between RNA structure and SR protein binding in the control of splicing and indicate that both regulatory components act synergistically on the major 5’ss.

SR protein binding analysis of mutant sequences was based on the ESEfinder software. Asang et al. (34) previously used ESEfinder to identify the potential SR protein binding sites in the wt SD region. SRSF2 protein binding was confirmed experimentally, but other interactions remain putative. It should be noted that a high ESE score does not guarantee SR protein binding (32). Furthermore, we only scored the presence or absence of a binding site, but the affinity of the SR protein may vary for a slightly different sequence and such effects may have contributed to the observed differences in splicing efficiency. In fact, some of the binding sites overlap and only the protein with the highest affinity may bind. Moreover, ESEfinder only predicts SR protein binding sites and the mutations could also inactivate or create binding sites for other regulatory proteins like hnRNPs. Despite these complexities, our analysis demonstrates that splicing at the HIV-1 major 5’ss does not only depend on the SD hairpin stability but also on the sequence of this region. The introduced mutations did not affect the 11-nt 5’ss sequence to which the U1 snRNA anneals and it seems likely that the observed sequence effects are due to altered binding of splicing regulatory
proteins. In fact, it seems possible that the SD hairpin stability will also affect binding of these regulatory proteins by determining the accessibility of their binding sites.

HIV-1 replication requires both the unspliced and the many spliced RNAs, and the ratio between these RNAs varies during the replication cycle. For example, at the onset of transcription the virus needs fully spliced transcripts for the production of the regulatory proteins (Tat and Rev) and at a later stage the virus needs singly spliced and unspliced RNAs for the production of the virion components (Gag, Pol and Env proteins and genomic RNA). The HIV-1 5’ss is potentially very strong because of the high sequence complementarity with the 5’ end of U1 snRNA. It seems likely that HIV-1 cannot select a weaker 5’ss with reduced U1 sequence complementarity to limit splicing, because this would hamper efficient splicing at all stages. Therefore, HIV-1 needs regulatory mechanisms to control the level of splicing. We here demonstrate that both local RNA structure and local SR protein binding sites are involved in this control. To prevent excessive splicing, the major 5’ss is positioned in a hairpin structure that partially occludes the U1 snRNA binding site. The level of splicing can subsequently be fine-tuned by the binding of the regulatory SR proteins. At a later stage, the HIV-1 Rev protein will also influence splicing by stimulating the nuclear export of unspliced and singly spliced transcripts that encode the Rev-responsive element (RRE). These complex mechanisms allow HIV-1 to balance the production of spliced and unspliced RNAs in a temporal manner.

We recently demonstrated that even small changes in the splicing level inhibit HIV-1 replication (30). Targeting the balanced production of spliced and unspliced HIV-1 mRNAs is therefore an interesting, yet relatively unexplored target for antiviral therapy. For example, splicing can be increased by using modified U1 snRNAs that bind more efficiently to the 5’ss (42). Another possibility is the use of antisense oligonucleotides that modify the SD structure (43), block the 5’ss (44) or target other splicing regulatory elements (45-49). Splice regulatory proteins could be targeted by small molecules. For example, indole derivate compounds (IDC) have been identified that selectively inhibit SR protein family members (50). Certain IDC potently inhibited HIV-1 replication in cell culture experiments (50, 51). Other IDC inhibit replication of murine leukemia virus (MLV) and protect mice from MLV-induced pathogenesis (52). This differential inhibition of HIV and MLV by different IDC most likely reflects distinct viral requirements for cellular splicing factors (52). Thus, drugs that modulate the splicing process can be used for the development of a novel HIV-1 therapy that should be fully active against multidrug-resistant virus variants. Understanding how HIV-1 splicing is regulated at the molecular level and the identification of all cellular and viral factors involved may help in the design of new antiviral strategies.

Material and Methods

DNA constructs

The wt plTR-gag-flag-luc plasmid and the variants with an A1, B1 and A1B1 mutated SD hairpin sequence were described previously (30). New SD variants were created by mutagenesis PCR (53) with plTR-gag-flag-luc as template. PCR1 was performed with primers TA109 (Table 4) and TATA-1(30). PCR2 was performed with the primers TA016 luc and one of the mutagenic primers shown in Table 4. PCR1 and PCR2 products were
purified and combined in a third PCR with outer primers TATA-1 and TA016luc. The resulting PCR3 product was HindIII and Ncol digested and ligated into the corresponding sites of the pLTR-gag-flag-luc vector. The prL-CMV plasmid contains the Renilla luciferase reporter gene under control of the cytomegalovirus immediate early (CMV-IE) promoter (Promega). The plasmid pTat-exon expresses the HIV-1 LAI Tat protein under control of the CMV-IE promoter (36).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA016luc</td>
<td>GGTCCATTTTCCAGCGG</td>
<td>+325 to +315</td>
</tr>
<tr>
<td>A^K</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+503 to +486</td>
</tr>
<tr>
<td>A^R</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +314</td>
</tr>
<tr>
<td>B^R</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +314</td>
</tr>
<tr>
<td>B^K</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +314</td>
</tr>
<tr>
<td>A^B</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +311</td>
</tr>
<tr>
<td>A^B^R</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +311</td>
</tr>
<tr>
<td>T2</td>
<td>GAACCCGCACAGGCAGAGCGAGGGAGCGCTGTTGATACGTCAAAATTTTGAC</td>
<td>+254 to +312</td>
</tr>
<tr>
<td>T3</td>
<td>GAACCCGCACAGGCAGAGCGAGGGAGCGCTGTTGATACGTCAAAATTTTGAC</td>
<td>+254 to +312</td>
</tr>
<tr>
<td>X</td>
<td>GAACCCGCACAGGCAGAGCGAGGGAGCGCTGTTGATACGTCAAAATTTTGAC</td>
<td>+254 to +312</td>
</tr>
<tr>
<td>Y</td>
<td>GAACCCGCACAGGCAGAGCGAGGGAGCGCTGTTGATACGTCAAAATTTTGAC</td>
<td>+254 to +312</td>
</tr>
<tr>
<td>Z</td>
<td>GAACCCGCACAGGCAGAGCGAGGGAGCGCTGTTGATACGTCAAAATTTTGAC</td>
<td>+254 to +312</td>
</tr>
<tr>
<td>XY</td>
<td>GAACCCGCACAGGCAGAGCGAGGGAGCGCTGTTGATACGTCAAAATTTTGAC</td>
<td>+254 to +312</td>
</tr>
<tr>
<td>YZ</td>
<td>GAACCCGCACAGGCAGAGCGAGGGAGCGCTGTTGATACGTCAAAATTTTGAC</td>
<td>+254 to +312</td>
</tr>
<tr>
<td>XZ</td>
<td>GAACCCGCACAGGCAGAGCGAGGGAGCGCTGTTGATACGTCAAAATTTTGAC</td>
<td>+254 to +312</td>
</tr>
</tbody>
</table>

* nts that differ from the wt sequence are underlined

Table 4. Oligonucleotides used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA016luc</td>
<td>GGTCCATTTTCCAGCGG</td>
<td>+325 to +315</td>
</tr>
<tr>
<td>A^K</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+503 to +486</td>
</tr>
<tr>
<td>A^R</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +314</td>
</tr>
<tr>
<td>B^R</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +314</td>
</tr>
<tr>
<td>B^K</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +314</td>
</tr>
<tr>
<td>A^B</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +311</td>
</tr>
<tr>
<td>A^B^R</td>
<td>GCTGAGCGCGCACACAGAAGCAGGGGACCTGTTGATACGTCAAAATTTTGACTA</td>
<td>+251 to +311</td>
</tr>
</tbody>
</table>

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 (54). Cells were transfected by calcium phosphate precipitation as previously described (54).

Luciferase assay
293T cells (~1.5 x 10\textsuperscript{5} cells per 2 cm\textsuperscript{2}, 60% confluency) were transfected with 20 ng wt or mutant pLTR-gag-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 (55). Data were subjected to independent-samples T test analysis (IBM SPSS) when indicated. The differences were considered significant when p<0.05.

Splice site analysis
293T cells (~1.5 x 10\textsuperscript{5} cells per 2 cm\textsuperscript{2}, 60% confluency) 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 (56) and TA113 (29) as described (29). PCR products were analyzed by agarose gel electrophoresis.

In silico analyses
The RNA secondary structure of the SD hairpin region was analyzed with the Mfold algorithm (version 3.4) offered by the MBCMR Mfold server (http://mfold.burnet.edu.au) using standard settings (57). SR protein binding sites in the major 5'ss region were identified with ESEfinder (version 3.0) using the standard threshold values (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home; (32, 33)).
Acknowledgements
This research was sponsored by the Netherlands Organisation for Scientific Research (Chemical Sciences Division; NWO-CW; TOP grant).

References


