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 7

Summary and Discussion
Summary of the thesis

RNA splicing is a critical step in the HIV-1 replication cycle because it controls the expression of the full viral proteome. The major 5'ss that is positioned in the untranslated leader of the HIV-1 RNA transcript is of particular interest because it is used for the production of all differentially spliced mRNAs. This thesis describes the role of the major 5'ss in HIV-1 splicing and the possibility that it is regulated by a repressive local RNA structure, the splice donor (SD) hairpin, that masks the 11 nucleotides (nts) of the 5'ss signal for recognition by U1 snRNA of the spliceosome machinery. We also set up a screen to identify cellular transcripts that are similarly splicing-suppressed by local 5'ss RNA structure. The situation at the HIV-1 5'ss is much more complex because the SD region encodes multiple sequence elements to which enhancer and silencer proteins bind. We tried to dissect the influence of all these factors. In addition, we describe a new splicing role for the HIV-1 Tat protein in usage of this major 5'ss.

It was previously reported that stabilization of the hairpin that surrounds the HIV-1 major 5'ss decreases the splicing efficiency and consequently causes a severe delay in virus replication (1). We now assessed whether 5'ss splicing is suppressed by local RNA structure in the wild-type (wt) HIV-1 transcript (chapter 2). To do so, mutants with a stabilized or destabilized structure were analyzed for the impact on virus replication and splicing. We could confirm that further stabilization of the wt hairpin structure reduces 5'ss usage. Importantly, we observed that destabilization of the wt hairpin does increase the level of HIV-1 splicing. This result indicates that HIV-1 RNA splicing is indeed at least partially suppressed by the SD stem-loop structure that encompasses the 5'ss. Both stabilization and destabilization of the SD hairpin affects virus replication, suggesting that HIV-1 requires a sub-optimal level of splicing in order to express all viral transcripts and proteins in the amount required for optimal virus replication.

In addition to local RNA structure, sequence elements in the SD region may bind SR proteins that negatively or positively regulate HIV-1 splicing. The initial mutations used to affect the SD hairpin structure also altered these putative SR protein binding sites. We tried to disentangle these two effects with newly designed mutants (chapter 3). This analysis indicates that there is an intricate interplay of positive and negative proteins and repressive RNA structure in the regulation of HIV-1 splicing at the major 5'ss.

U1 snRNA binding to 11 nts of the major 5'ss is likely a determining factor in splicing regulation. We tried to design 5'ss mutants with optimized splicing activity by opening the local RNA structure, disrupting negative SR protein binding sites and optimization of the U1 snRNA binding site, but without disrupting the positive SR sites (chapter 4). We indeed obtained nearly complete splicing, which coincided with a loss of virus replication capacity, thus confirming that HIV-1 5'ss splicing is deliberately suppressed to allow the expression of the full viral transcriptome and proteome. We also demonstrate that these regulatory features (RNA structure, SR protein and U1 snRNA binding sites) are well-conserved among HIV-1 isolates of
The HIV-1 Tat protein is well-known for its essential stimulatory role in viral transcription, but other functions have also been described (2-6). For instance, a report indicated that Tat influences 3'ss selection (2). In chapter 5 we boost the Tat-splicing connection by showing that this small viral protein enhances splicing at the major HIV-1 5'ss. As for the mechanism of transcriptional activation, the Tat splicing effect requires elements encoded within the HIV-1 LTR promoter and the TAR hairpin that is present at the 5'end of all viral transcripts. These results suggest that Tat-activated transcription and splicing regulation are coupled processes.

We provide multiple lines of evidence in chapters 2, 3 and 4 that the local RNA structure suppresses HIV-1 5'ss splicing. There are other reports of splicing regulation by RNA structure (7-18). We performed a screen of ~50,000 human transcripts to identify cellular 5'ss motifs that are suppressed by local RNA structure. We identified many transcripts with a candidate 5'ss stem-loop structure and selected five elements for subsequent splicing tests in reporter-based assays (chapter 6). The RNA structure of 5'ss\(^{16}\) in the primary FAM73B transcript was shown to inhibit splicing in a reporter construct. In addition, an alternative 5'ss was found embedded within the hairpin of 5'ss\(^{48}\) that was activated upon inactivation of 5'ss\(^{48}\). These combined results warrant a more intensive search for transcripts of which splicing is controlled by RNA structure.

**Discussion: Alternatives for the SD hairpin structure**

The HIV-1 untranslated leader RNA region can fold several stem-loop structures that have important regulatory functions, including the TAR, polyA, DIS, SD and psi hairpins (Fig. 1A) (19, 20). This thesis mainly focuses on the SD hairpin (nts 282-300; boxed in blue in Fig. 1A) and its regulatory role in HIV-1 RNA splicing. However, two recent reports suggested a slightly different multiple-hairpin conformation of the leader RNA in which the SD region folds differently (21, 22). We will therefore review in detail what is known about this SD hairpin and the alternative conformations.

**The SD hairpin**

The presence of the SD hairpin was first suggested in 1993 based on the unusual migration of a 44 nt-leader RNA molecule encoding the SD and psi sequences during native gel electrophoresis (23). The SD hairpin structure was subsequently confirmed in multiple studies that used a variety of techniques, including Mfold RNA structure prediction, chemical probing and mutational analysis (24-26). In addition, a base pair (bp) co-variation was described in the lower stem region (A\(^{282}\),U\(^{300}\) instead of G\(^{282}\),C\(^{300}\)) of diverse HIV-1 isolates, thus providing phylogenetic support for the proposed base pairing scheme (25). An initial NMR structure analysis of a short leader RNA fragment (nt 282-300) also supported the SD hairpin, but bp A\(^{286}\),U\(^{295}\) was suggested to form a novel base-triple platform with the bulged nucleotide (nt) A\(^{296}\) (27). SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension)-based RNA structure probing of the complete HIV-1 RNA genome also confirmed folding of the SD hairpin (28).
Figure 1. Proposed structures for the SD region in the HIV-1 leader RNA. (A) The multiple hairpin structure in which the RNA folds the SD hairpin (▲, 5'ss cleavage site) and several other stem-loop structures that encode important replication signals (TAR, transacting responsive element; polyA, polyadenylation signal; PAS, primer activation signal; PBS, primer binding site; DIS, dimerization initiation signal; Ψ, packaging signal). (B) The alternative SDa hairpin structure. (C) The 3WJ structure in which the SD region interacts with upstream nts (SD-3WJ interaction). (A-C) The HIV-1 LAI prototype sequence is shown with the nts that form the U5-AUG long-distance interaction boxed in red (*, AUG start codon) and the nts that form the 2-bp U5-AUG extension in the 3WJ structure indicated with red filled boxes. In C, the 17 unpaired or weakly-paired guanosines that have been suggested to serve as potential NC binding sites are encircled in green.
The analysis of mutant and revertant viruses provides an additional means to score the importance of specific RNA structures. For instance, we demonstrated that opening of the TAR hairpin is detrimental to HIV-1 replication. Revertant viruses that were selected upon long-term culturing of the TAR-mutated virus repaired this structure – not necessarily the sequence - by fixation of spontaneous mutations (29). This forced evolution approach mimics the search for bp co-variations in a phylogenetic survey to identify biologically relevant RNA structures (30). HIV-1 mutation studies indicated that the thermodynamic stability of the SD stem-loop structure is important for virus replication (1, 31, 32). Stabilization of the hairpin caused a severe virus replication defect and decreased the splicing efficiency. Upon long-term culturing of these mutants, revertant viruses could be selected that acquired additional mutations that opened the stabilized hairpin, thus demonstrating that a too stable SD hairpin hampers HIV-1 replication (1). Perhaps more important for demonstrating the existence of the SD hairpin, we showed that destabilization of this structure increased HIV-1 RNA splicing and reduced virus replication. We did however not observe evolutionary repair of the destabilized structure in long-term culture experiments, which may be due to the fact that the mutations did only cause a mild replication defect and evolutionary repair may not sufficiently improve virus replication to outcompete the original mutant in the time frame of the experiment (31).

The alternative SD* hairpin
Although the original SHAPE-probing study supported the SD hairpin (28), a more recent SHAPE analysis by the same group of Kevin Weeks suggested an alternative hairpin structure (22). This SD* hairpin (Fig. 1B) presents the 5'ss cleavage site in the base-paired stem and not in the single-stranded loop as in the prototype SD hairpin. In fact, the SD* hairpin was proposed in early phylogenetic studies and also supported by chemical probing and RNA structure prediction software (33, 34). However, the SD hairpin was subsequently believed to represent the more likely RNA fold (24). Interestingly, we also observed this alternative SD* structure in SHAPE-probing experiments, but only for mutants in which the SD hairpin was destabilized (31). Mfold analysis confirmed SD* folding for these destabilized mutants, unlike the wt sequence that adopts the SD hairpin (31, 35).

The 3WJ structure
A recent NMR study on a larger fragment of the HIV-1 leader RNA (nt 105-345) of the NL4-3 strain by the Michael Summers team suggested a rather different folding of the SD RNA region (21). A tandem three-way junction (3WJ) structure was proposed for this RNA fragment (Fig. 1C), in which the SD nts interact with upstream leader sequences (interaction indicated with SD-3WJ interaction), in particular with nts that form the 5' and 3' sides of the PBS stem in the earlier-proposed multiple-hairpin conformation (Fig. 1A) (20). The 3WJ structure contains the U5-AUG duplex that was previously described for the multiple-hairpin structure (boxed in red in Fig. 1A and C), but this interaction is now extended with two base pairs (bps) compared to the original structure (filled red boxes). The original U5-AUG interaction is supported by HIV-1 mutational studies, phylogenetic (co-variation) analyses and biochemical
assays (chemical probing and gel electrophoresis) (20, 36). The initial report (20) also demonstrated conservation of this U5-AUG interaction in different HIV-1, HIV-2 and SIV isolates despite significant sequence variation.

The new NMR study proposes that the 3WJ structure is intrinsically involved in the regulation of different leader RNA processes, such as mRNA translation and packaging of the dimeric RNA genome in virions, but only the latter was tested experimentally. Analysis of NC protein binding and RNA packaging for wt and mutant RNAs suggested the importance of unpaired and weakly-paired “junction” guanosines that are exposed by the 3WJ scaffold (encircled in green in Fig. 1C). The suggested translation regulatory function remains speculative, however. In fact, a role in translation regulation was previously proposed for the U5-AUG interaction as it masks the Gag start codon, but a subsequent mutational survey failed to provide support for such a mechanism (37).

In search for phylogenetic support for SD, SDa or 3WJ
The recent controversy about the actual folding of the SD domain triggered us to perform an in depth phylogenetic analysis. This unbiased method was key for solving the structure of ribosomal RNA and many other complex RNA molecules (38-40). Such an in silico analysis seems particularly valuable because different experimental HIV-1 RNA probing methods seem to yield different answers. A brief phylogenetic survey was performed in the recent NMR study by Keane et al. (21) and it was concluded that “the tandem three-way junction structure is highly conserved, and the rare variations that disrupt base pairing are due to transient polymorphisms”. We set out to perform a more detailed sequence variation analysis for the bps that constitute the SD, SDa and 3WJ structures.

SD phylogeny
We analyzed the sequence of all HIV-1 group M isolates described in the Los Alamos HIV sequence database (www.hiv.lanl.gov; Web alignment of complete sequences 2012) but excluded ambiguous nts, nt insertions and deletions and nt variations that were observed only in a single virus isolate, as these may represent reverse transcription, PCR or sequencing errors. We scored 459 variations in the SD region (nt 282-300) when 1128 isolates were compared to the prototype subtype B LAI strain (Fig. 2A). A large fraction of the variations affect base-paired nts (indicated with bp 1-6 in Fig. 2A). In 32 isolates, a bp co-variation is observed at bp 1 (A-U instead of G-C bp) that strongly supports this interaction. The majority of the other variations also allow formation of the same bps. For example, the A286G change that is observed in 263 isolates causes a G-U bp instead of an A-U at bp 4. Other variations that open bps and that at first glance do not support the SD hairpin, result in a small realignment in the stem region and in the formation of a comparable SD hairpin structure. For example, the frequently observed double variation C287A/U288C opens two bps (bp 5 and 6: C287-G294 and U288-A295) but results in the formation of two alternative bps (A287-U285 and C288-G294). Some virus isolates hold multiple nt differences in this leader RNA domain, but as previously shown (Chapter 4) most isolates can fold a similar SD hairpin structure. Only a minority of the variations does
not allow Watson-Crick bp formation. For each bp at which nt variation was observed, we calculated the percentage of co-variations, structure conserving variations (all other variations that do allow bp formation in an SD hairpin structure) and structure disrupting variations (not allowing bp formation in an SD hairpin structure) (Fig. 2B), which revealed that in particular the variation observed at bp 1, 3, 4 and 5 is compatible with the SD hairpin structure. Taken together, we observed 8% co-variations, 81% structure conserving variations, and only 11% disrupting variations in the SD region (Fig. 3), which gives phylogenetic support for the existence of the SD hairpin.

**SDa phylogeny**

In the RNA region that folds the alternative SDa hairpin (nt 288-301), 270 nt variations were observed in 1135 isolates that affect one of the bps in the stem (bp 1-3 in Fig. 2C and D). At bp 1, a co-variation is observed in 14 isolates (U-A to G-C), which represents ~7% of the observed variation at this position (Fig. 2D). Approximately 43% of the variation at this position is structure conserving and 51% is structure disrupting (Fig. 2D). Co-variation is also observed in 2 isolates at bp 2 (G-C to A-U), while most of the sequence variation at this position is structure conserving. At bp 3, only structure disrupting variation is observed. In total, 6% of the variations in the SDa region represent bp co-variations and 49% are structure conserving (Fig. 3). A large fraction (45%) of the variations disrupts the SDa structure, while this level was much smaller for the SD hairpin structure (only 8%; Fig. 3).

**3WJ phylogeny**

We similarly analyzed the interaction between the SD sequence and the upstream nts (SD-3WJ domain) and the extended U5-AUG interaction in the 3WJ structure. Analysis of the SD-3WJ nts yielded 642 variations in 708 isolates that affected 10 of the 12 putative bps in this region (bp 7-16 in Fig. 2E and F). No bp co-variations were observed. For 7 of the 12 bps, all observed variation is structure disrupting. Only at bp 11 and 14, all sequence variation is structure conserving. However, the frequently observed A286G change that is responsible for the variation at bp 14 also supports the SD hairpin structure (Fig. 2A, bp 4). In total, 69% of the SD-3WJ variations are structure disrupting and only 31% are structure conserving (Fig. 3).

We observed 59 variations in 462 extended U5-AUG sequences that affected only 6 of the 13 putative bps (bp 1-6 in Fig. 2E and F). Also in this domain no co-variations were observed. The variation observed at bp 3 and 4 in the original duplex and at bp 6 in the U5-AUG extension is structure conserving. Disrupting variation is only seen at bp 1 and 2 in the original duplex and at bp 5 in the extended structure. In total, only 24% of the variation in the original U5-AUG duplex (bp 1-4) is structure disrupting, whereas this percentage is 73% for the duplex extension (bp 5-6) (Fig3). Taken together, our phylogenetic analysis does not support the interaction between the SD nts and upstream sequences (SD-3WJ) or the U5-AUG duplex extension (bp 8-9) that are characteristic for the 3WJ structure (20).
Figure 2. Phylogenetic analysis of the SD, SDa and 3WJ structures. The sequence of the SD hairpin (A-B), SDa hairpin (C-D) and 3WJ structure (U5-AUG and SD-3WJ interactions; E-F) in HIV-1 isolates described in the 2012 web alignment of the Los Alamos HIV-1 sequence database were compared to the corresponding sequence in the HIV-1 LAI strain. Sequences with ambiguous nts, nt insertions and deletions and nt variations that were observed in a single virus isolate were excluded from the analysis. (A, C, E) Variations observed in base paired nts (numbered bps) and their frequency in 1128 SD, 1135 SDa, 708 3WJ-SD and 462 U5-AUG sequences are shown. (B, D, F) Nt variations were scored as co-variation (grey bar), as structure conserving (all other variations that do allow Watson-Crick bp formation in the same or similar structure; white bar) and structure disrupting variations (not allowing bp formation in the same or similar structure; black bar). Co-variations were counted as single events.
Figure 3. Phylogenetic analysis supports the SD hairpin structure. All nt variations observed in the base paired nts of the SD and SD\textsuperscript{a} hairpins, and the SD-3WJ interaction, U5-AUG duplex (bp 1-4) and U5-AUG duplex extension (bp 5-6) of the 3WJ structure were scored as co-variation (grey bar), as structure conserving (white bar) or as structure disrupting (black bar). Co-variations and double variations were counted as single events. The total frequency of each category is shown.

Little evidence for SD\textsuperscript{a} and 3WJ structures
Our broad phylogenetic analysis of the three proposed structures for the SD region indicates that the observed sequence variation is highly compatible with the SD hairpin structure, less with the SD\textsuperscript{a} structure and only poorly with the 3WJ folding (Fig. 3). In particular, no bp co-variations were observed in the proposed SD-upstream nts interaction in the 3WJ structure. Possibly, such sequence variation may occur less frequently in tertiary RNA structure motifs like 3WJ than in relatively simple hairpin structures. However, bp co-variations have been described for other tertiary RNA folds like pseudoknots (41-44). In fact, the proposed 3WJ structure is supported exclusively by NMR data (21). Even though it was mentioned by Keane et al. that the new 3WJ structure is consistent with available SHAPE structure probing data, this provides only non-exclusive support as the same data were previously used to support a different RNA structure (45). Although the NMR study represents a tour de force concerning the size of the RNA fragment studied (155 nts), a caveat of the study is the use of multiple truncations (deletion of the TAR and polyA hairpins and the apical PBS domain). This mono-disciplinary data set seems too restricted to recall existing RNA structure models for the HIV-1 leader RNA.

Some of the confusion about the folding of this leader RNA domain may come from the fact that all of the proposed foldings are semi-stable. For example, MFold RNA structure analysis predicts a thermodynamic stability of -7.1 and -3.6 kcal/mol for the SD and SD\textsuperscript{a} hairpins, respectively. These hairpins are much less stable than some of the other leader RNA hairpins, like TAR (-24.8 kcal/mol) and polyA (-15.3 kcal/mol). We cannot formally disproof the SD\textsuperscript{a} and 3WJ structures and it remains of interest to
test their candidate functions, possibly as part of a riboswitch, as multiple structures may co-exist or distinct structures may exist under different conditions. Such a riboswitch has been proposed for other leader RNA domains (19, 46, 47). Distinct structures may exist under different conditions to play a specific role in one of the many leader RNA functions in the HIV-1 replication cycle. Accordingly, different SD domain structures may be adopted at different stages of the HIV-1 replication cycle, e.g. as nuclear versus cytoplasmic RNA or upon packaging of the viral RNA genome in virion particles, possibly induced by coating with the viral Nucleocapsid (NC) protein. Functional tests by means of a mutational analysis are seriously hampered by the exquisite sequence and structural requirements of these leader RNA domains (48).

**Limited phylogenetic support for the SD structure**

Although the SD sequence analysis was informative, we previously obtained much stronger phylogenetic support for other HIV-1 RNA structures. For instance, we documented multiple co-variations and even clustered co-variations in a single duplex for other leader domains, like the TAR and polyA hairpins, thus providing very strong phylogenetic support for the existence of these structures (49-54). We did not see such prominent co-variations in the SD region of the HIV-1 genome. This may indicate that the SD hairpin is less important for virus replication than the TAR and polyA hairpins. This may also indicate a stronger sequence requirement in SD than TAR. Mutation-reversion analysis of the lower stem of the TAR hairpin indeed indicated that maintenance of the structure is more critical than the sequence requirement (29). Conservation of the SD sequence is likely related to the presence of the U1snRNA and SR protein binding sites (32, 35, 55, 56). These sequence requirements may have prevented the evolution of more sequence variation.

The SD hairpin structure is not only supported by our phylogenetic analysis, but also by previous studies using different techniques, including virus mutant/revertant analysis, RNA structure probing, gel electrophoresis and NMR analysis (23-28). The mutational analysis and SHAPE structure probing presented in this thesis also support the SD hairpin structure. Moreover, we demonstrate that the SD hairpin has a regulatory function in the control of HIV-1 splicing and that both the structure and sequence of the hairpin are critical for this function.

**References**


35. Mueller N, Berkhout B, Das AT. 2015. HIV-1 splicing is controlled by local RNA structure and binding of splicing regulatory proteins at the major 5’ splice site. The Journal of general virology.


47. Huthoff H, Berkhout B. 2001. Mutations in the TAR hairpin affect the equilibrium between alternative conformations of the HIV-1 leader RNA. Nucleic acids research 29:2594-2600.


