Human immunodeficiency virus type 1 splicing at the major splice donor site is controlled by highly conserved RNA sequence and structural elements

Nancy Mueller, Bep Klaver, Ben Berkhout and Atze T. Das

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

Human immunodeficiency virus type 1 (HIV-1) splicing has to be strictly controlled to ensure the balanced production of the unspliced and all differently spliced viral RNAs. Splicing at the major 5′ splice site (5′ss) that is used for the synthesis of all spliced RNAs is modulated by the local RNA structure and binding of regulatory SR proteins. Here, we demonstrate that the suboptimal sequence complementarity between this 5′ss and U1 small nuclear RNA (snRNA) also contributes to prevent excessive splicing. Analysis of a large set of HIV-1 sequences revealed that all three regulatory features of the 5′ss region (RNA structure, SR protein binding and sequence complementarity with U1 snRNA) are highly conserved amongst virus isolates, which supports their importance. Combined mutations that destabilize the local RNA structure, remove binding sites for inhibitory SR proteins and optimize the U1 snRNA complementarity resulted in almost complete splicing and accordingly reduced virus replication.

Two supplementary tables are available with the online Supplementary Material.

DOI 10.1099/jgv.0.000288

Received 11 June 2015
Accepted 16 September 2015

Human immunodeficiency virus type 1 (HIV-1) produces a single primary RNA transcript. The full-length transcript functions as RNA genome that is packaged into virions and as mRNA for translation of the Gag and Pol proteins. HIV-1 RNA contains several splice donor (SD) [5′ splice site (5′ss)] and splice acceptor [3′ splice site (3′ss)] sites. Differential usage of these sites results in the production of a variety of spliced RNAs that encode the other viral proteins. Splicing has to be strictly regulated for the balanced production of all viral RNAs and proteins.

Formation of the spliceosome complex is initiated by binding of U1 small nuclear ribonucleoprotein (snRNP) to the 5′ss, which is mediated by the complementarity between the 11 single-stranded nucleotides at the 5′ end of the U1 small nuclear RNA (snRNA) component and the 5′ss sequence (Freund et al., 2003). The major 5′ss located in the 5′ untranslated leader of HIV-1 transcripts is used for the production of all spliced mRNAs. As both spliced and unspliced transcripts are required for a productive replication cycle, splicing at this 5′ss has to be controlled to prevent complete processing. Previous studies indicated that both RNA structure and sequence elements are involved in this control. The 5′ss region can fold a stem-loop structure, the splice donor (SD) hairpin, that partially occludes the U1 annealing site and reduces splicing frequency (Abbink & Berkhout, 2008; Mueller et al., 2014). Splicing is also influenced by binding of several SR proteins to splice regulatory elements (SREs) in the 5′ss region (Asang et al., 2012; Mueller et al., 2015). The SD hairpin structure may also influence binding of these SR proteins. In addition, the extent of sequence complementarity between the 5′ss and U1 may influence snRNP binding and splicing efficiency.

Previous studies demonstrated that viral RNA structures and sequence motifs with an important role in viral replication have been highly conserved during HIV-1 evolution. Therefore, the major 5′ss region of different HIV-1 isolates was analysed for the capacity to form an SD hairpin stem-loop structure, presence of putative SR protein-binding sites and sequence complementarity with U1 snRNA. In total, 1184 sequences of HIV-1 group M isolates described in the Los Alamos HIV sequence database (Web alignment 2012; http://www.hiv.lanl.gov/) were analysed and compared with the subtype B LAI strain that is used as prototype in our studies (Table S1, available in the online Supplementary Material).

The ability to form an SD stem-loop structure was analysed by mfold RNA structure analysis (Zuker, 2003). The majority of the isolates (780 sequences) have an SD hairpin sequence (nt 282–300) identical to that of the LAI strain and can fold the same structure (Fig. 1a, structure A). Most isolates...
Fig. 1. Evolutionary conservation of the structure and sequence characteristics of the major 5’ss. (a) The capacity of the major 5’ss RNA region of 1184 HIV-1 group M isolates described in the Los Alamos sequence database (Web alignment 2012) to fold an SD stem–loop structure was analysed with mfold RNA structure analysis software. The SD hairpin structure.

(b) 5’ss sequence | Structure | Isolates | SR protein-binding site
--- | --- | --- | ---
GGGCCGUGUguguacgccca | A | 604 | + + + + + + 17.5
GGGCCGUGUguguacgccca | C | 203 | + + + + + + 17.5
GGGCCGUGUguguacgccca | A | 79 | + + + + + + 17.5
GGGCCGUGUguguacgccca | A | 45 | + + + + + + 17.5
GGGCCGUGUguguacgccca | A | 35 | + + + + + + 17.5
GGGCCGUGUguguacgccca | F | 32 | + + + + + + 17.5
GGGCCGUGUguguacgccca | C | 31 | + + + + + + 17.5
GGGCCGUGUguguacgccca | P | 17 | + + + + + + 17.5
GGGCCGUGUguguacgccca | A | 11 | + + + + + + 17.5
GGGCCGUGUguguacgccca | K | 10 | – – – + + + + 17.3

(c) SRSF5 72.6% | SRSF6(1) 99.8%
SRSF2(1) 98.2% | SRSF6(1) 95.6%
SRSF2(1) 92.1% | SRSF6(1) 95.4%

(d) CAGUUGGCUGGUguguacgcccaaa
for the HIV-1 LAI strain is shown (structure A). The identical sequence and structure is observed for 780 isolates. The sequence and structure of different hairpin prototypes is shown (structures B–V) with nucleotides differing from the HIV-1 LAI sequence boxed and nucleotide insertions encircled. The frequency of every structure variant is indicated. Observed nucleotide differences in the prototypic sequences are indicated with the abbreviations referring to the sequence names shown in Table S1. The 5'ss nucleotides to which the U1 snRNA binds, are encircled in grey (cleavage site indicated by arrowhead). In two isolates with structure C (with an A→G variation at position 286) an additional base pair (not shown) can be formed in the lower stem region as a result of a nucleotide co-variation at positions 281 and 301 (CS5 isolate: A(281)U(301)) or a single nucleotide difference at position 281 (CS6 isolate: U(281)A(301)). (b, c) The major 5'ss region (nt 281–301 in LAI) was analysed with ESEfinder for the presence of putative SR protein-binding sites and with the HBond Score software to calculate the hydrogen bond score (HBS) of the 11 nt 5'ss sequence (highlighted in grey; exon and intron nucleotides indicated in upper and lower case, respectively). (b) The 10 most frequently observed sequences are shown. Nucleotides that differ from the most frequently observed sequence (corresponding to the HIV-1 LAI strain; upper row) are indicated in bold and underlined. The number of isolates with the corresponding sequence, the predicted structure (A–V, as shown in a), the presence (+) or absence (−) of putative SR protein-binding sites and the HBS value are indicated. The complete analysis of all sequences is shown in Table S2. (c) The position of the putative SR protein-binding sites in the HIV-1 LAI SD region and the percentage of virus isolates carrying each site are shown (5'ss nucleotides to which the U1 snRNA binds are boxed in grey). (d) Sequence variation in the 5'ss motif observed in a collection of 201 541 human 5'ss, as described by Roca et al. (2013). The height of each nucleotide reflects its conservation at the corresponding position.

Asang et al. (2012) previously used ESEfinder, an algorithm based on SELEX (systematic evolution of ligands by exponential enrichment)-identified RNA binding sites for splicing regulatory proteins (Cartegni et al., 2003; Smith et al., 2006), to identify potential binding sites for the SR proteins SRSF2 (SC35), SRSF5 (SRp40) and SRSF6 (SRp55) in the major 5'ss region. Binding of the SRSF2 protein was confirmed experimentally and shown to reduce splice site usage, but the other interactions remain putative. ESEfinder analysis of the HIV-1 group M sequences indicated that the previously described SR protein-binding sites are present in 776 of 1184 isolates (Fig. 1b, c, Table S2). The SRSF6(1) site is observed in nearly all isolates (99.8%), whereas the overlapping SRSF6(2) site is detected in 95.6% of the isolates. All but one of the 1184 isolates have at least one of these SRSF6 sites. Similarly, the overlapping SRSF2(1) and SRSF2(2) sites are detected in 92.1 and 98.2% of the isolates, respectively, with 98.9% of the isolates having at least one of these SRSF2 sites. The SRSF2(3) and SRSF5 sites are observed in 95.4 and 72.6% of the isolates, respectively. Thus, the presence of these putative SR protein-binding sites is highly conserved amongst group M isolates, with the possible exception of the SRSF5 site.

The 11 nt 5'ss sequence to which U1 snRNA binds varies only slightly amongst different HIV-1 isolates (Fig. 1b, Table S2). The most frequently observed sequence (C–UGGUGAGUAC+8) deviates only at position −2 from the consensus sequence for eukaryotic 5'ss elements (Fig. 1d). Whereas most HIV-1 isolates have a U at position −2 that cannot pair with U1 snRNA (Fig. 1c), >50% of eukaryotic 5'ss elements have an A that can interact with U1 (Fig. 1d). The hydrogen bond score (HBS) (Freund et al., 2003) reflects the base-pairing potential between the U1 snRNA and the three exonic (nt −3 to −1) and eight intronic nucleotides (nt +1 to +8) at the 5'ss. The HBS can vary from 1.8 for a 5'ss that contains only the essential G+U+2 nucleotides to 23.8 for a 5'ss with optimal complementarity to U1. To compare the complementarity between U1 and the major 5'ss of the different HIV-1 isolates, we calculated the HBS for each variant with the HBond Score Web-Interface (http://www.uni-duesseldorf.de/rna/html/hbond_score.php). The most frequently observed 5'ss sequence has an HBS value of 17.5 (Fig. 1b, Table S2), which reflects a high, but incomplete sequence complementarity with U1. Due to the small sequence variation in the 5'ss sequence, the HBS score varied slightly for other HIV-1 isolates, and a value between 17.3 and 17.7 was calculated for 98.9% of the M group isolates (1171 of 1184). These results demonstrate that the sequence complementarity between U1 snRNA and the 5'ss is also highly conserved during HIV-1 evolution.

Our in silico analysis demonstrated that the SRSF2-binding sites are highly conserved amongst HIV-1 isolates, which indicates a role for these SREs in splicing control, although in a previous study the deletion of the SRSF2(1) site did not
**Fig. 2.** Mutation of the SD hairpin region affects 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 aa of Gag, Flag-tag peptide and firefly luciferase. The luciferase transcript contains a cryptic 3’ss. The unspliced RNA encodes an active luciferase protein, whereas the spliced RNA does not. The sequences of the WT and mutant 5’ss regions are shown. The 5’ss nucleotides to which the U1 snRNA anneals are boxed in grey (cleavage site indicated by arrowhead). Nucleotide substitutions are boxed in black. (b) The structure of the WT and mutant SD hairpins as predicted with the mfold RNA structure analysis software are shown. The 5’ss nucleotides are encircled in grey. (c) The WT and mutant 5’ss regions were analyzed with ESEfinder. The presence (+) or absence (−) of the putative SR protein-binding sites is shown. In addition, the RNA structure stability (ΔG; structures shown in b), HBS...
The observed luciferase levels (Fig. 2d) were used to estimate results in activity. Previously, we demonstrated that transcription was established by basal LTR promoter infected without Tat plasmid, when a lower level of expression (Fig. 2d). Mutants A3a, A3b and A4 showed a splicing of the WT transcript results in detectable luciferase analysed by measuring the luciferase activity in transfected LTR promoter. The effect of the mutations on splicing was Tat-expressing plasmid to activate transcription from the (Mueller et al., 2014) to analyse the effect of SD mutations on 5′ss usage. (d) Splicing of the WT transcript results in detectable luciferase expression (Fig. 2d). Mutants A3a, A3b and A4 showed a decrease in luciferase activity, which reflects increased splicing. Similar effects were observed when cells were transfected without Tat plasmid, when a lower level of transcription was established by basal LTR promoter activity. Previously, we demonstrated that ~85% of the transcripts containing a WT SD hairpin are spliced, which results in ~15% unspliced RNA (Mueller et al., 2014). The observed luciferase levels (Fig. 2d) were used to estimate the unspliced and spliced RNA level of the mutant constructs (unspliced RNA mutant = RLU mutant/RLU WT × 15%; spliced RNA mutant = 100% − unspliced RNA mutant %). Splicing of the A3a mutant was slightly increased to ~90% (~10% unspliced RNA), whereas the A3b and A4 mutants showed a further increase to ~94–95% splicing (~5–6% unspliced RNA; Fig. 2c). These latter mutants, lacking both repressive SRSF2-binding sites upstream of the 5′ss, thus show the highest level of splicing. As the thermodynamic stability of all mutant RNA structures is predicted to be the same, it seems likely that removal of both splicing-inhibitory SRSF2 sites was responsible for the increased splicing frequency of the A3b and A4 mutants. These results confirm that 5′ss splicing is inhibited by binding of SRSF2 proteins in the SD region. Removal of both binding sites did not result in complete splicing, which indicates that other limiting factors are present.

The suboptimal sequence complementarity between the 5′ss and U1 snRNA may be one of the factors that cause incomplete splicing. To investigate this possibility, we constructed two SD mutants with increased U1-binding capacity. In the SDa mutant, U2 C was replaced by A to increase the HBS to 21.1 (Fig. 2). This mutation slightly destabilized the SD hairpin structure (ΔG = −7.9 kcal mol⁻¹; Fig. 2c). In the SDaa mutant, in addition the C3 was replaced by A, because an A is frequently observed at this position in eukaryotic 5′ss (Fig. 1d). This double mutation resulted in an HBS of 19.4 and further destabilized the hairpin structure (ΔG = −5.5 kcal mol⁻¹). ESEfinder analysis indicated that the SDas and SDaas mutations inactivated both putative SRSF2-binding sites upstream of the cleavage site (Fig. 2c).

Cells were transfected with WT or SD-mutated luciferase constructs and the Tat plasmid, and luciferase production was measured after 48 h. Both mutants showed around fivefold decrease in luciferase activity, which reflects a significant increase in splicing compared with WT (Fig. 2d). Similar effects were observed when the analysis was performed at basal transcription levels without Tat. Based on these luciferase levels, we estimate that splicing of the SDs and SDaas mutants is increased to ~97%, yielding only ~3% unspliced RNA (Fig. 2c). This high splicing level can be explained by the combinatorial effects: the improved 5′ss–U1 snRNA complementarity, removal of both SRSF2(1) and SRSF2(2) sites, and a destabilized RNA structure that results in a better 5′ss exposure. Splicing of the SDa mutant with a slightly destabilized SD hairpin (ΔG = −7.9; HBS 21.1; 97% splicing) was higher than that of A3b and A4 that lack the same SRSF2 sites and have a more dramatically destabilized RNA structure (ΔG = −5.2; HBS 17.5; 94–95% splicing). This difference in splicing frequency suggests an important role of the 5′ss–U1 snRNA complementarity (reflected by the HBS) in splicing efficiency. Splicing of the SDaas mutant (ΔG = −5.5; HBS 19.4; 97% splicing) was also higher than that of mutants A3b and A4, whilst the RNA
structures varied only slightly in stability, which confirms the important role of 5'ss–U1 snRNA complementarity. Taken together, these results indicate that HIV-1 splicing at the major 5'ss is restricted by local RNA structure, binding of SRSF2 and a suboptimal 5'ss sequence. By removing these restrictions, we generated a very efficient splice site mediator near the context of a doxycycline (Dox)-dependent HIV-1 variant. In this HIV-rtTA variant, the viral Tat-TAR transcription activation mechanism was inactivated and functionally replaced by the components of the Dox-inducible Tet-Off gene expression system (Das et al., 2004a) (Fig. 3a). HIV-rtTA replication can be modulated by varying the Dox concentration, lacking replication in the absence of Dox and very efficient replication at 1000 ng Dox ml⁻¹ (Das et al., 2004b). To detect small replication differences, we used a suboptimal Dox level. SupT1 cells were therefore infected with the WT and SD-mutated HIV-rtTA variants and cultured at 25 ng Dox ml⁻¹. The J2 mutation that stabilizes the SD hairpin and reduces splicing (Mueller et al., 2014) was included as control. When compared with the HIV-rtTA with a WT SD region, infection with the J2 variant resulted in a ~40% reduced CA-p24 level in the culture supernatant at 4 and 5 days after infection (Fig. 3b), which is in agreement with the previously observed replication defect of the corresponding HIV-1 mutant (Mueller et al., 2014). Replication of the A4, A3b, SDa and SDaa mutated HIV-rtTAs resulted in a 36–65% lower CA-p24 level compared with the WT HIV-rtTA, which shows that these SD mutations (that increase the splicing frequency) decrease virus replication.

Taken together, the present and previous studies demonstrate that splicing at the major 5'ss is regulated by the stability of the local RNA structure, the presence of SREs to which SR proteins can bind, and the base-pairing potential between the 5'ss and the U1 snRNA. The analysis of a large set of HIV-1 sequences revealed that all three regulatory features of the 5'ss region are highly conserved in different group M viruses, which supports an important function during virus replication. Modulation of the splicing frequency reduces HIV-1 replication and, as previously discussed (Mueller et al., 2015), the HIV-1 splicing process is an interesting target for antiviral therapy.

![Fig. 3. Mutations that alter the splicing frequency reduce virus replication.](image)

(a) Proviral genome of the Dox-controlled HIV-rtTA variant. In HIV-rtTA, the Tat-TAR transcriptional activation mechanism was inactivated by mutations in Tat and TAR (TARm) and functionally replaced by the Dox-inducible transcription mechanism by introducing tetO-binding sites in the U3 promoter region and by replacing the nef gene with the rtTA gene. As a consequence, transcription and replication of this virus are dependent on Dox administration. (b) SupT1 T-cells were infected with equal amounts of WT and mutant virus (CA-p24 input: 3 ng ml⁻¹) and cultured with 25 ng Dox ml⁻¹ to activate viral replication. Virus replication was monitored by measuring the CA-p24 level in the culture supernatant at 4 and 5 days after infection. Data represent mean ± SEM of four experiments.

References


