Recognition and masking of putative U1 snRNA binding sites

Erkennung und Maskierung von putativen U1 snRNA Bindesequenzen

Inaugural-Dissertation

zur Erlangung des Doktorgrades der

Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Anna-Lena Brillen

aus Oberhausen

Düsseldorf, März 2017

aus dem Institut für Virologie

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. H. Schaal

Koreferent: Prof. Dr. M. Feldbrügge

Tag der mündlichen Prüfung: 19.05.2017

Für meine Familie

Table of Contents

Abstract

Removal of intronic sequences during pre-mRNA processing requires accurate recognition of the exon-intron boundaries. Here, the initial step is the RNA duplex formation between the 5' splice site (5'ss) and the 5' end of U1 snRNA. Human as well as viral, e.g. HIV-1, 5'ss sequences however, do not depend on full complementarity to all 11 nucleotides of the 5' end of U1 snRNA. So-called *cis*-acting regulatory sequences (SREs), which can be bound by SR or hnRNP proteins, can greatly influence 5'ss recognition, with both positive and negative impacts. Algorithms like the hexamer-based 'HEXplorer', calculating enhancing or silencing properties of regions in the vicinity of splice sites, as well as algorithms (e.g., MaxEnt, HBond-Score (HBS)) calculating intrinsic splice site strength, aim at reliably describing splice site selection and uncovering splice variants that may result in different protein isoforms. As it is for endothelial cell function, suppressing or upregulating specific protein isoforms might lead to the development of new therapeutic strategies fighting inherited diseases that can even occur through silent point mutations. To take a step forward in understanding splice site regulation, we investigated splice donor usage and exon definition in two model systems. The human fibrinogen Bβ-chain gene (FGB) exon 7 and HIV-1 exon 2/2b both contain neighboring splice donor sites of comparable strength. We show that those donor sites are regulated via a multitude of SREs, acting in a strict position dependent manner. Additionally, we dissected the role of hnRNP D in viral pre-mRNA splicing and found hnRNP A1 and hnRNP D binding to ESSV at overlapping motifs. Thereby, levels in *vpr* mRNAs seem to be dynamically regulated. Disrupting only one of those elements leads to an altered splicing phenotype and might therefore represent valuable targets for LNA-mediated therapy.

Zusammenfassung

Während der mRNA Prozessierung erfordert die Entfernung von intronischen Sequenzen die genaue Erkennung von Exon-Intron Grenzen. Hierbei ist der erste Schritt die Bildung eines RNA-Duplex zwischen der 5‘ Spleißstelle (5’ss) und dem 5‘ Ende der U1 snRNA. Humane und virale, z.B. HIV-1, 5’ss Sequenzen benötigen jedoch keine völlige Komplementarität aller 11 Nukleotide zum 5‘ Ende der U1 snRNA. So genannte *cis*-agierende regulatorische Sequenzen (SREs), die von SR und hnRNP Proteinen gebunden werden können, können die 5’ss Erkennung positiv oder negativ beeinflussen. Algorithmen, wie der auf Hexamer-basierende ‚HEXplorer‘, der die fördernden und inhibierenden Eigenschaften der Sequenzen in der Nachbarschaft von Spleißstellen berechnet, und Algorithmen (z.B. MaxEnt, HBond-Score (HBS)), die die intrinsische Stärke von Spleißstellen kalkulieren, zielen darauf ab, die Spließstellenselektion zuverlässig zu beschreiben und Spleißvarianten aufzudecken, die zu unterschiedlichen Proteinisoformen führen. Wie bei Endothelzellen könnte eine Unterdrückung oder Hochregulation spezifischer Proteinisoformen zur Entwicklung von neuen Therapie-Strategien führen. Therapie-Strategien, die auch stille Mutationen einbeziehen, könnten auch Erbkrankheiten adressieren. Um dem Verständnis der Spleißstellenselektion einen Schritt näher zu kommen, haben wir die Benutzung von Spleißdonoren und die Exondefinition in zwei Modellsystemen untersucht. Exon 7 des humanen FGB-Gens und Exon 2/2b von HIV-1 beinhalten beide benachbarte Spleißdonoren vergleichbarer Stärke. Wir konnten zeigen, dass diese Donoren von einer Vielzahl positionsabhängiger SREs reguliert werden. Zusätzlich haben wir die Rolle von hnRNP D beim viralen Spleißen analysiert. Hier zeigte sich, dass hnRNP A1 und hnRNP D das ESSV Element an überlappende Motive binden. Dadurch werden *vpr* mRNA Level dynamsich reguliert. Die Inaktivierung von nur einem dieser Elemente

führt zu einem veränderten Spleiß-Phänotyp. Diese Zielsequenzen könnten deshalb als nützliche Angriffsziele einer LNA-basierenden Therapie dienen.

1. Introduction

1.1. Pre-mRNA splicing

Processing of messenger RNA is a characteristic feature of eukaryotic gene expression. The primary transcript is modified by at least three main processing steps before it leaves the nucleus and serves as template for protein synthesis [1]: the addition of a 7-methyl guanosine cap at its 5' end, the removal of intervening sequences, called introns, and the addition of a poly(A) tail at its 3'end. Introns are removed by joining the flanking sequences (exons) to generate the mature mRNA. The spliceosome, a large and highly dynamic protein-RNA complex, facilitates the two sequential transesterification reactions involved in splicing with single nucleotide precision. Besides splicing, binding of spliceosomal components to a splice site can also shield the mRNA from degradation [2], premature cleavage and polyadenylation [3]. Any interruption of this orchestrated process can lead to errors in gene expression.

1.1.1. Mechanisms of pre-mRNA splicing

The spliceosome precisely recognizes sequence elements within the precursor mRNA (pre-mRNA): the 5' splice site (5'ss or splice donor (SD)), the 3' splice site (3'ss or splice acceptor (SA)) and the branch point sequence (BPS). 99% of all 5'ss are characterized by an 11 nucleotide long sequence marked by a GU at the exon/intron border [4]. The largely degenerate consensus sequence CAG\GURAGUNN (R = purine, N = purine or pyrimidine, \ = exon-intron border) in metazoans permits the prevalence of alternative splice site usage, expanding proteomic diversity, which is not prominent in lower eukaryotes like *Saccharomyces cerevisiae* [5, 6]. The 3'ss is composed of a polypyrimidine tract (PPT, ~10-20 nucleotides long) and an invariant AG at the intron/exon border which is characterized by the CAG/G (/ = intron-exon border) consensus motif [5]. The BPS with its distinctive adenosine (YNYURAC (Y =

pyrimidine, R = purine, N = any nucleotide)) is located 15-50 nucleotides upstream of the 3'ss AG [7, 8].

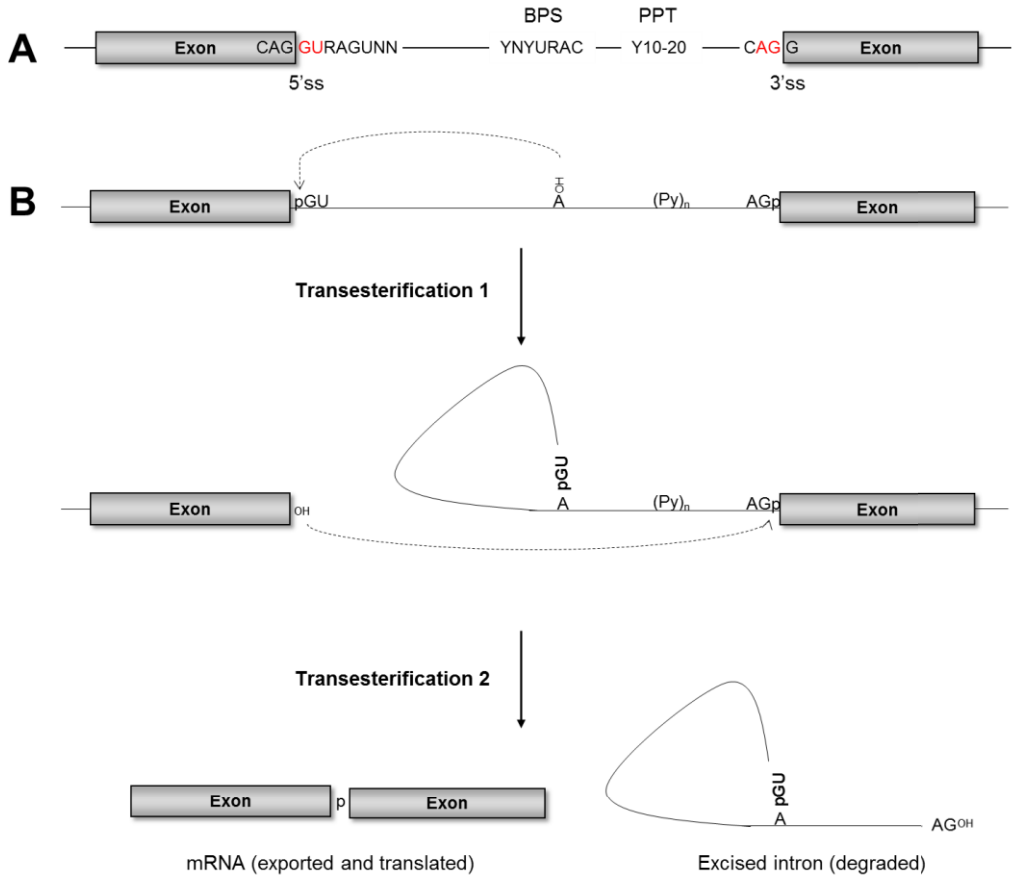


Figure 1 – The splicing reaction
(**A**) Shown are two exons (grey boxes) that are separated from each other by an intron (line). The exon/intron boundaries are marked by the 5'ss and 3'ss. The 5'ss is defined by an 11 nucleotide long sequence, including the canonical GU dinucleotide. The 3'ss is composed of a polypyrimidine tract (PPT), followed by the invariant AG dinucleotide. The BPS is situated 15-50 nucleotides further upstream. (**B**) During the first transesterification step, the phosphorus atom at the 5'ss is attacked by the 2' hydroxyl group of the BPS. In the second catalytic step, the oxygen atom of the hydroxyl group of the first exon forms a bond with the phosphorus atom at the intron/exon border of the second exon leading to exon ligation and an excised intron lariat. This illustration was adapted from [9].

The chemical mechanism of the splicing reaction entails two rather simple sequential transesterification steps (Figure 1). First, the 2' hydroxyl group of the BPS-adenosine attacks the phosphodiester bond that links the 5' exon and the intron, generating a 5' exon and a lariat intermediate. Subsequently, the phosphodiester bond at the 3'ss is attacked by the hydroxyl group of the 5' exon, leading to a replacement of the bond between the lariat intron and the ligated exons [10]. The splicing reaction is controlled by the spliceosome. The major spliceosome consists of five different subunits, U1, U2,

U4/U6 and U5 snRNPs (uridine-rich small nuclear RNPs) and a vast number of associated proteins [11]. Each U snRNP is composed of one (or two for U4/U6) U snRNAs, a ring of seven Sm (B/B', D1, D2, D3, E, F and G) or like-Sm proteins (for U6) and a variable number of snRNP particle-specific proteins [8, 12]. In addition to the "U2-type" spliceosome including U1 and U2 snRNPs, also a minor type (U12-type) coexists in a subset of eukaryotes in which U1 and U2 are substituted by U11 and U12, representing, however, only less than 0.5% of all introns present in human cells [13]. In contrast to the U2-type, the subunits of the U12-dependent spliceosome, are U11, U12, U5, and $U4_{atac}/U6_{atac}$ snRNPs [14]. The different U snRNPs assemble in a stepwise manner when they encounter a pre-mRNA substrate which involves broad dynamic changes within the architecture of the complex [8, 12, 15] (Figure 2). The first step is the ATP-independent formation of the E complex (E, early complex) in which the U1 snRNP is recruited to the 5'ss and interacts via base pairing with the mRNA through the 5' end of its RNA [16]. Additionally, non-snRNP factors such as SF1/mBBP (splicing factor 1/mammalian branch point binding protein) and both subunits of U2AF (U2 snRNP auxiliary factor) interact with the BPS, PPT and the AG of the 3'ss, respectively. In an ATP-dependent step, catalyzed by the DExD/H helicases Prp5 and Sub2, U2 snRNP displaces SF1/mBBP, progressing E into A complex [7, 11]. The U2 snRNP forms a duplex with the BPS and interacts with the U1 snRNP, resulting in the bulging of the adenosine, which serves as a nucleophile for the first transesterification step [7, 17].

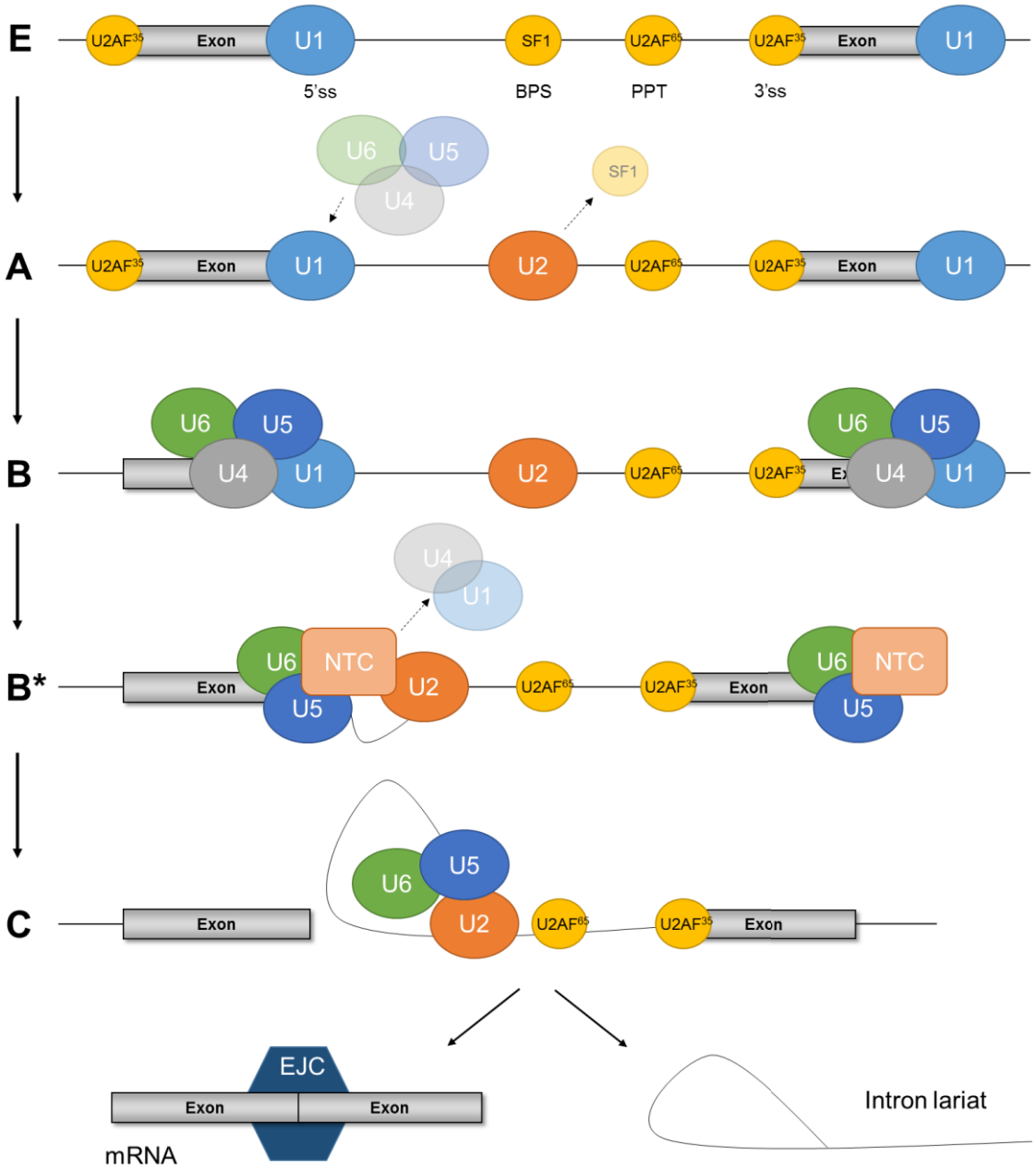


Figure 2 – Spliceosome assembly
The stepwise assembly of the spliceosome starts with the formation of the E complex via interactions of the U1 snRNP with the 5'ss and SF1, U2AF^{65}, U2AF^{35} to the BPS, PPT and the 3'ss, respectively. ATP-dependent binding of the U2 snRNP displaces SF1 from the BPS, generating the A complex. Subsequently, the pre-assembled tri-snRNP U4/U6*U5 joins, forming the B complex which is catalytically activated (B* complex) through conformational changes and binding of the NineTeen complex (NTC). This causes the release of U1 and U4 snRNPs from the spliceosome. The first transesterification reaction leads to an exon intermediate and lariat intron intermediate, forming complex C. After the second transesterification, spliceosomal components dissociate, the exons are ligated and the intron is degraded.

Subsequently, the preassembled U4/U6*U5 tri-snRNP is recruited to form the B complex which is catalyzed by another DExD/H helicase, Prp28. Furthermore, the NineTeen complex (NTC or Prp19/CDC5 complex) is involved in B complex formation,

enabling stable interactions of the U6 and U5 snRNPs with the pre-mRNA [18]. The catalytically inactive B complex is converted into a catalytically active form (B* complex) by ATP-dependent conformational and structural rearrangements: these include the displacement of U1 and U4 snRNAs and the interaction of U2 and U6 snRNAs, which is catalyzed by additional helicases (Brr2, Snu114 and Prp2) [11, 19, 20]. Thus, by interaction of U2-U6 snRNAs, the 5'ss and the BPS adenosine acquire close proximity, initiating the first transesterification reaction and converting B* into C complex [8]. Following further extensive structural rearrangements, which depend on Prp8, Prp16 and Slu7 [21], the second catalytic step takes place in which both exons are ligated and the lariat intron is degraded [22]. Concurrently with the second catalytic step, a protein complex is recruited 20-24nt upstream of the exon-exon junction (Exon Junction Complex; EJC) which is involved in mRNA export, translation and quality control [23]. Finally, the spliceosome complex falls apart and the single components are recycled for additional rounds of splicing.

1.1.2. Exon and Intron definition

The distance between splice sites have been shown to influence efficient spliceosomal assembly. Contrary to *Saccharomyces cerevisiae*, the human gene architecture is mainly determined by short exons (the average size of middle exons is 151nt [24] and long introns (hundreds to thousand base pairs long) [25]. Thus, the splicing machinery must precisely select those small exons within a plethora of considerably longer introns. Consequently, recognition of exons is very likely established through interactions spanning the exon, in a process called exon-definition [26, 27] (Figure 3). Here, binding of U1 snRNPs to a 5'ss result in a direct communication with the upstream located U2 auxiliary factor $U2AF^{65}$ and subsequently directs U2 snRNPs to the BPS [28]. Observations substantiated the model by the discovery that strengthening a 5'ss led to the cumulative usage of the upstream 3'ss, whereas

mutating a splice donor resulted in an increase in exon skipping events [29-31]. However, if an exon exceeds the length of 200-250 nucleotides, splice site recognition likely takes place across the intron via intron-definition complexes [32].

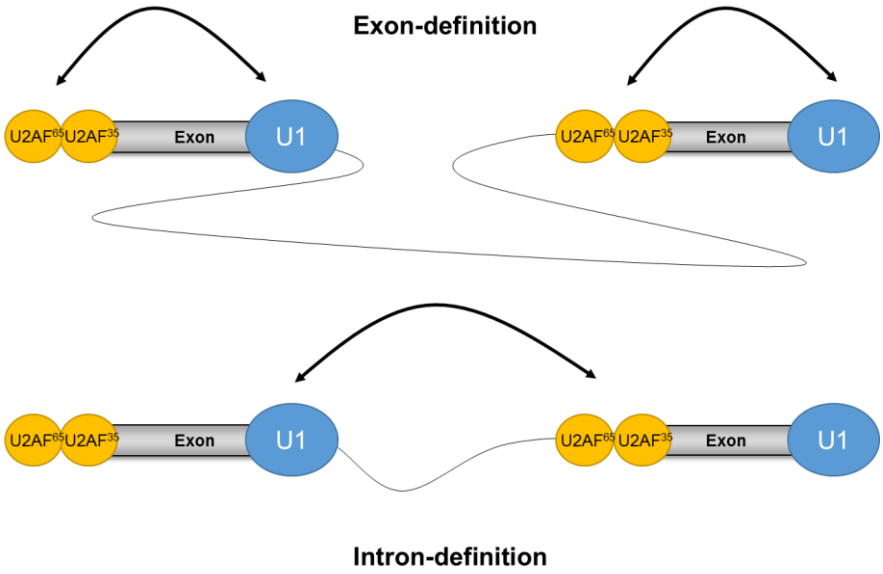


Figure 3 – Exon and intron definition
Human genes consist of short exons which are separated by long introns (top). Here, splice site recognition occurs mainly across exons in which the U1 snRNP interacts with the upstream located U2 associated factors $U2AF^{65}$ and $U2AF^{35}$. The lower panel shows the intron-definition model in which splice site pairing occurs across the intron.

But, since every splicing reaction occurs across the intron, each exon-definition state must be switched to an intron-definition complex in which the 5'ss interacts with the downstream located 3'ss. Until today, this switch is only poorly understood and may often involve support of splicing regulatory proteins [33, 34]. Another hypothesis is that the flip between exon- and intron-definition complexes occurs directly when an U4/U6*U5 tri-snRNP bound to the exon interacts with an upstream 5'ss [35]. However, it is important to bear in mind that the detection of the first and last exon, which each only has one flanking splice site, is facilitated by other mechanisms. On one hand, the 5' terminal exon requires the cap-binding complex that substitutes for U2 snRNP in communicating with the downstream located 5'ss [36]. On the other hand, the 3' terminal exon is recognized by U2AF, either interacting with the polyadenylation polymerase (PAP), the cleavage factor CF1m or the cleavage and polyadenylation

specificity factor (CPSF)[25, 37-39]. Moreover, transcription, capping, addition of a poly(A) tail and splicing do not proceed in a sequential order, as previously thought, but take place co-transcriptionally, allowing novel interplays between several regulatory elements [40]. Obvious indicators for co-transcriptional splicing are sequence analyses of RNA associated to chromatin which showed that more exons than introns were enriched in the samples [41], as well as the fact that 65% of intronic sequences have been spliced out after only five minutes [42] which coincides with the transcription rate of 1.8 to 4.0 kb per minute [40, 43-46]. Furthermore, elongation time can modify splice site decisions by influencing RNA binding of spliceosomal compartments and splicing regulatory proteins [40]. On the contrary, RNA binding proteins itself can also affect elongation time. The carboxy-terminal domain (CTD) of RNA polymerase II (Pol II) recruits and directly binds splicing factors. For SRSF3-mediated splicing regulation, the Pol II CTD is essential [47] and it has been shown that SRSF2 positively affects elongation by interactions with the positive transcription elongation factor (P-TEFb) [48, 49]. However, not all mRNAs are co-transcriptionally spliced but some are post-transcriptionally spliced which is thought to control e.g. the timing of gene expression [50, 51].

1.1.3. Splice site recognition

Many human splice sites are degenerate and only poorly match to the consensus sequence of the U1 snRNA. Only the GT at the 5'ss and the AG at the 3'ss are conserved to some extent [52]. However, alternative splicing does not occur because of inaccuracy of the spliceosome, but with surprisingly great precision [53]. It is estimated that 94-95% of all human genes generate at least two mRNA isoforms through alternative splicing, thereby including different sets of exons [54, 55]. Types of alternative splicing involve exon skipping, inclusion of mutually exclusive exons, alternative 5' or 3'ss selection or the retention of an intron (Figure 4). The degree to

which splice sites are used or not is regulated by both *cis*-regulatory sequences and *trans*-acting factors.

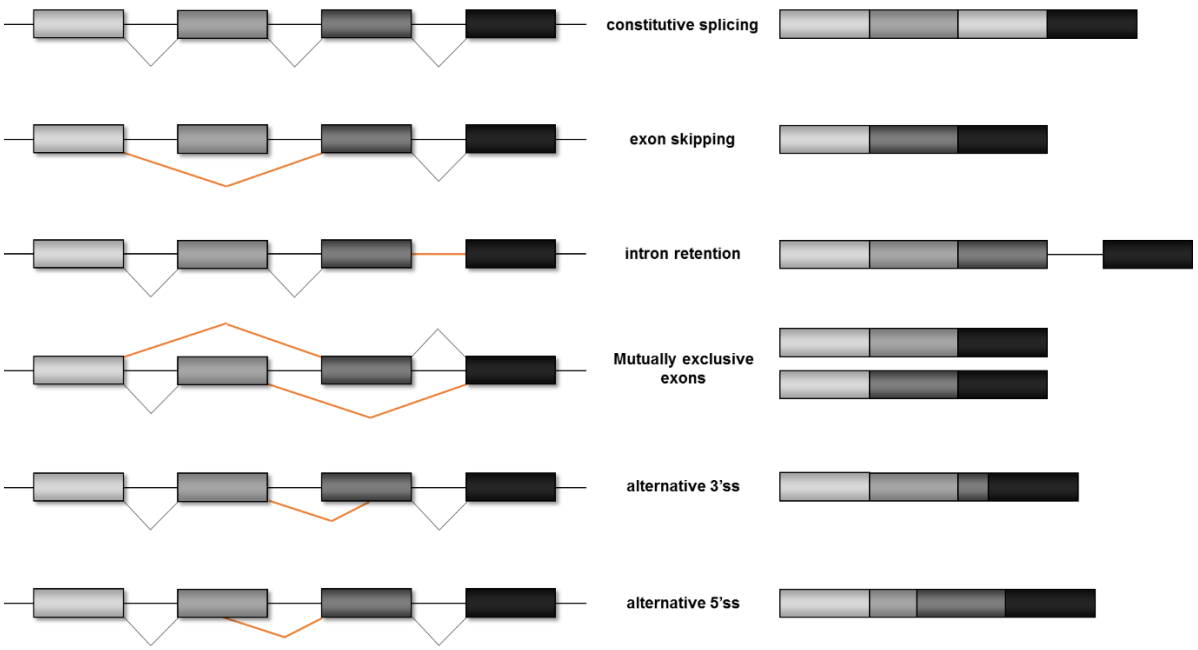


Figure 4 – Modes of alternative splicing
Model of the major forms of alternative splicing. Apart from constitutive splicing, alternative splicing allows the variable inclusion/exclusion of exons or the retention of an intron. Another possibility is that alternative 5'ss and 3'ss are used.

The intrinsic strength of a 5'ss is based on the ability to form hydrogen bonds with the single stranded 5' end of the U1 snRNA [2, 56]. Algorithms that most reliably calculate the intrinsic strength of a 5'ss are the HBond score (https://www2.hhu.de/rna/html/hbond_score.php; [56]), which is based on predicted hydrogen bond formation of the U1 snRNA and the 11 nucleotides of a 5'ss, and the MaxEnt score (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html; [57]), which calculates the distribution of nucleotides at specific positions within 9 nucleotides of a 5'ss.

Contrary to that, the strength of a 3'ss is assumed to critically dependent on the pyrimidine content of the PPT, the distance between the BPS and the 3'ss, as well as the complementarity between the BPS and the U2 snRNA [58-60]. Within a polypyrimidine tract, uridines are favored over cytidines and a stretch of 11 uridines is

most efficiently used for splicing [61, 62]. Furthermore, the two nucleotides surrounding the AG of the 3'ss have been implicated in affecting 3'ss recognition, in which a pyrimidine at position -3 (relative to exon/intron border) was shown to be relevant for U6 snRNA interactions, whereas a guanosine at position +1 seems to be only relevant for a specific subset of introns [63-65].

However, to determine the actual "functional strength" of a splice site, splicing regulatory elements (SREs) surrounding splice donor and acceptor sites have to be considered. Generally, those SREs are mainly bound by three classes of RNA-binding proteins: SR (Serine/arginine-rich), hnRNP (heterogeneous nuclear ribonucleoparticle) proteins and tissue-specific proteins such as Rbfox, PTB or Nova [66-73]. Depending on their position within the exon or the intron, SREs are classified as exonic/intronic splicing enhancers (ESE/ISE) or exonic/intronic splicing silencers (ESS/ISS) and act in a strict position-dependent manner, either positively or negatively influencing splice site recognition [74, 75] (Figure 5). Furthermore, SREs were shown to act cooperatively to combinatorically control splice site use [76, 77] and are not only essential for alternative but also for constitutive splicing [78-80].

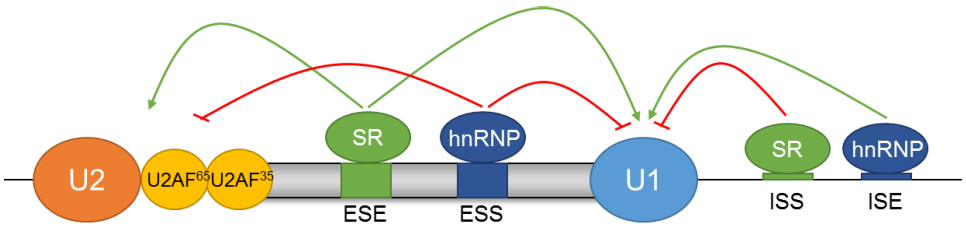


Figure 5 – Position-dependent regulation of splicing regulatory proteins
Binding of SR or hnRNP proteins to enhancer or silencer elements can promote or inhibit splice site recognition by spliceosomal components. SR proteins increase splice site usage if bound to the exon, whereas they repress splice site usage from intronic positions. hnRNP proteins show a mirror-inverted phenotype and promote splicing from the intron and inhibit splicing from the exon.

1.1.4. SR proteins

SR proteins represent a family of structurally related RNA binding proteins that possess repeats of serine (S) and arginine (R) dipeptide residues (RS) of variable length and

one or two N-terminal RNA recognition motifs (RRMs) which are interchangeable between all SR proteins [66]. In general, the RRM domain(s) facilitate binding to the pre-mRNA, whereas the RS domain triggers protein-protein or other protein-RNA interactions [81, 82]. However, it could be shown that the quality of RS domains within different SR proteins is directly linked to the number of included RS dipeptides and furthermore depends the phosphorylation state of serine residues within RS domains [83, 84]. So far, twelve SR proteins have been identified that range in size between 20-75kDa: SF2/ASF, SC35, SRp20, SRp75, SRp40, SRp55, 9G8, SRp46, SRp30c, SRp38, SRp54 and SRrp35, which were recently re-named in SRSF1-SRSF12 to clarify their relationship to each other [85] (Figure 6). Members of the SR protein family are not only found in all metazoan species and plants but also in some lower eukaryotes with exception of *Saccharomyces cerevisiae*, which only expresses three SR-like proteins [66, 86, 87].

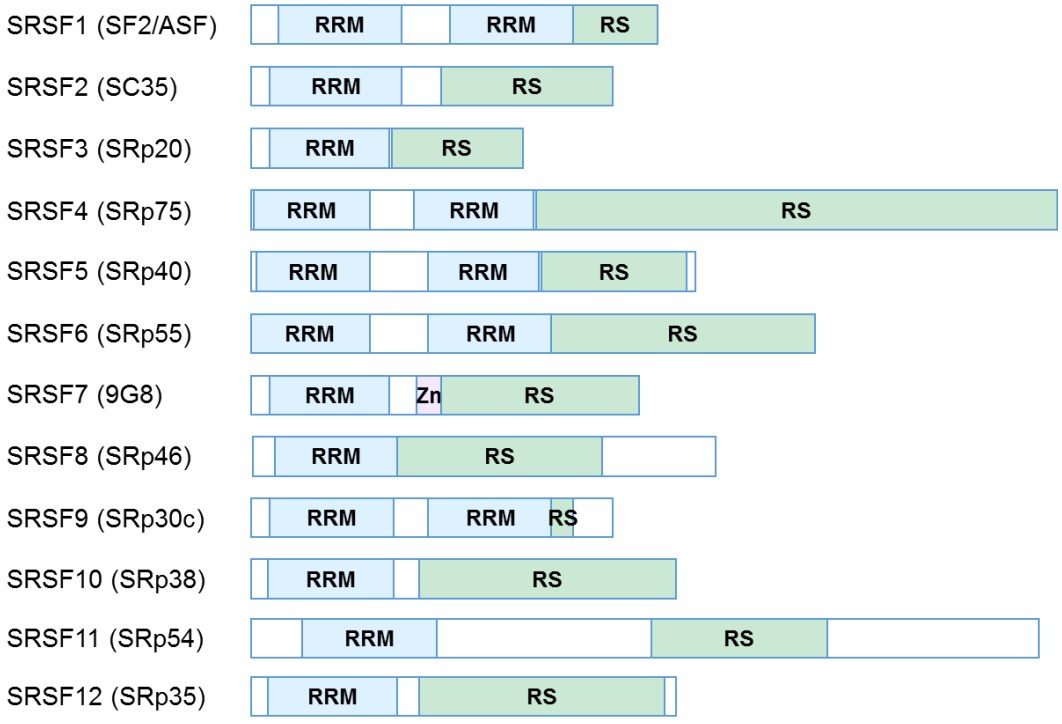


Figure 6 – Structural features of SR proteins
SR proteins share a modular structure, containing one or two RNA recognition motifs (RRM) at the N-terminus (blue) and an RS domain at the C-terminus (green). Only SRSF7 contains an additional element, a zinc-finger motif (Zn)(rose). This illustration was adapted from [88].

A multitude of high-affinity binding sites have been proposed for several SR proteins by using SELEX (selected evolution of ligands though exponential enrichment) or CLIP-seq techniques. However, even though SR proteins recognize distinct RNA recognition motifs, the motifs have been found to be rather degenerate and, in several cases, are bound redundantly by different SR proteins [89-91]. Moreover, the RNA binding capacity of SR proteins can be improved by other participating proteins like the SR-like protein Tra2 through cooperative binding with SRSF1 or SRSF7 [92-94] or by additive effects through the binding of many copies of the same enhancer [70]. SR proteins binding to ESE sequences were shown to promote 5'ss usage via interactions with the U1-specific protein U1-70K that itself contains two RS domains. While earlier studies suggested that the RS domain of SR proteins facilitate the recruitment of the U1 snRNP to the 5'ss [81, 82, 95], other studies indicated that it is rather the RRM that is responsible [96, 97]. Besides, it was shown that SR proteins interact with another U1-specific protein, U1-C [98], or directly with the U1 snRNP [99, 100] to enhance splice site activation. Additionally, SR proteins bound to ESE sequences influence the recognition of a 3'ss splice site by stabilizing the U2 snRNP auxiliary factors $U2AF^{35}$ and recruiting $U2AF^{65}$ to the AG dinucleotide and the PPT, respectively [66, 82, 101-104]. SR proteins are also able to enhance splicing without binding to ESE sequences. Here, SR proteins facilitate the incorporation of the U4/U6*U5 tri-snRNP during B complex formation [105, 106]. On the contrary, SR proteins have also been reported to bind to intronic sequences. According to their position-dependent effects, they were only associated with repression of splicing when bound to intronic positions [74, 107-110]. This mode of action is supposed to include the formation of a so-called "dead-end" complex. Here, the spliceosomal complex is trapped in A complex formation and cannot progress any further [33, 74, 111]. At present, there is, however, only little

understanding about the exact mechanism behind this SR-mediated inhibition of splicing.

1.1.5. hnRNP proteins

Another large group of splicing regulatory proteins represent the hnRNP proteins [71]. 20 major types of hnRNP proteins are known so far, named hnRNP A-U, which share some common features but are highly divergent regarding their composition and functional properties (Figure 7). Furthermore, related hnRNP-like proteins exist, like Nova2, hnRNP LL or hnRNPD-like JKTBP.

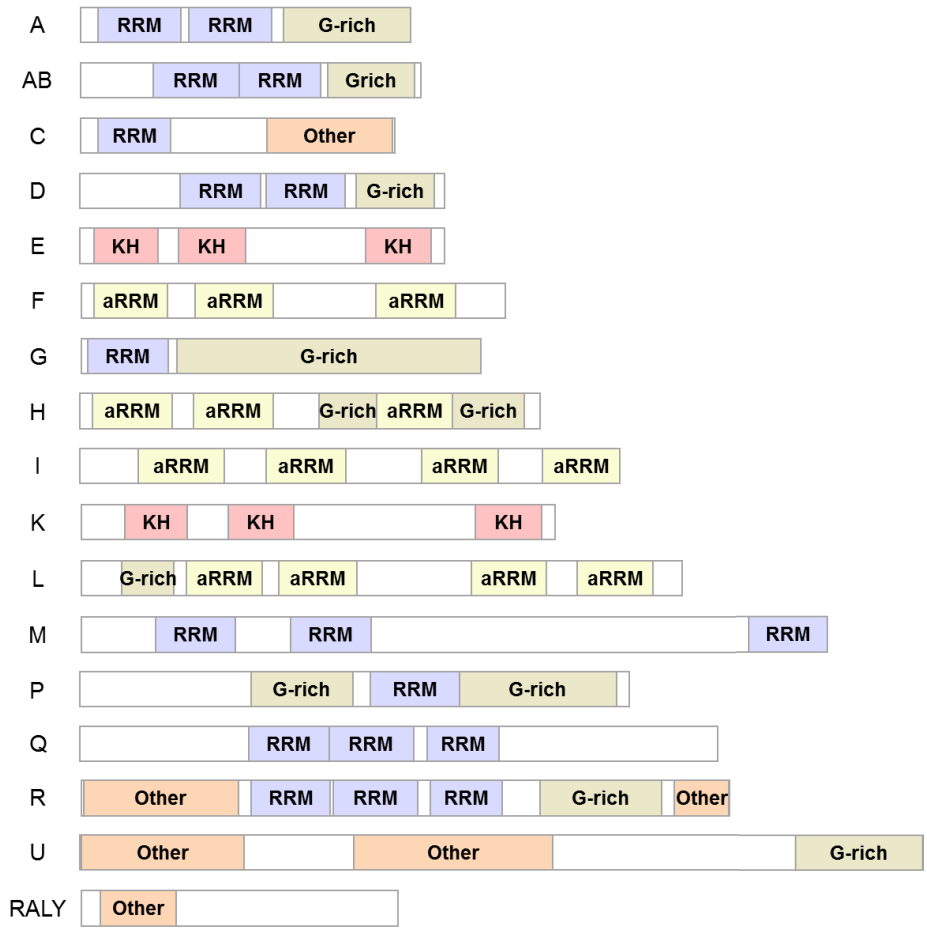


Figure 7 – Strucutral features of hnRNP proteins
The group of hnRNP proteins is highly diverse and consist of multiple different domains that are connected by linker regions.

The most common structural features of the hnRNPs are the RNA binding motif (RRM), the quasi (q)RRM- or hnRNP K-homology (KH)-type as well as auxiliary domains enriched in proline, glycine, tyrosine, arginine, glutamine or asparagine [71, 112, 113].

The variable structure of the different hnRNP proteins contribute to the great functional diversity of hnRNP proteins including telomere biogenesis, polyadenylation, translation, RNA editing, nuclear export and mRNA stabilization [114-116]. Furthermore, several members of the hnRNP protein family have been found being involved in alternative splicing. One of the best studied hnRNP subgroup with respect to alternative splicing is hnRNP A/B which includes isoforms A1, A2/B1, A3 and A0. They are known as splicing repressors, showing an inverse phenotype compared to SR proteins by inhibiting when located upstream, and promoting splicing when localized downstream of a 5'ss [74, 117]. Several scenarios are reported of how hnRNP proteins inhibit splice site usage. In the simplest one, binding of hnRNP A/B proteins restricts spliceosomal components or SR proteins in binding to a splice site or *cis*-regulatory element, respectively, via multimerization [118, 119]. Another possibility of negative regulation is the looping-out mechanism. Here, the intervening sequence is folded out, thereby bringing distant splice sites into close proximity, which finally leads to exon skipping [120, 121]. Also, hnRNP C has been implicated in splicing regulation and was shown to compete with $U2AF^{65}$ in binding to the polypyrimidine tract and prevents the recognition of cryptic 3' splice sites [122]. Furthermore, hnRNP C inhibits the binding of the hnRNP-like protein TIA-1 downstream of 5'ss, thereby counteracting its enhancing potential [122]. Unlike other hnRNP proteins, members of the hnRNP F/H subgroup (hnRNP H, H', F and 2H9) specifically recognize G runs which are defined by the consensus sequence DGGGD (where D is A, G, or U) [76, 123, 124]. hnRNP F/H are composed of three quasi-RRMs which were shown to interact with the RNA through three highly conserved loops [125]. Furthermore, they contain a G-rich domain with which they are able to self-interact, thereby looping out intervening introns as it has been shown for hnRNP A/B [126]. In general, hnRNP F/H proteins have been shown to inhibit splice donor usage from exonic positions [74, 127-

132], whereas they were also shown to promote splicing when bound to the intron [74, 133-138]. Herein, hnRNP F/H binding sites have been reported to accumulate downstream of alternatively spliced exons in order to guarantee 5'ss usage [76, 139, 140]. However, there are exceptions to the rule: hnRNP F/H was also found to activate splicing from intronic positions [124, 141]. Another hnRNP protein, hnRNP I, is better known as PTB (polypyrimidine tract-binding protein) due to its ability to inhibit 3'ss usage via binding to upstream polypyrimidine tracts [142]. It negatively acts on splicing by binding to CU-rich sequences [143, 144] and several mechanisms have been elucidated through which PTB inhibits splicing, including competition of binding sites with U2 snRNP [62, 145], looping-out of the skipped exon [146, 147] and preventing the pre-spliceosomal A complex to progress into a catalytic state ("dead-end" complex) [33]. Furthermore, it was reported that RRMs 1 and 2 of PTB directly interact with SLIV of the U1 snRNA, preventing further spliceosome assembly [111]. Despite its role as a negative regulator, several alternative exons have been characterized that are positively regulated by PTB [148]. Most recently, hnRNP D proteins associated with an HIV-1 silencer complex, ESSV, was shown to act through competing with hnRNP A1 binding to an overlapping motif [149]. All hnRNP D isoforms act as repressors of exon inclusion whose silencer activity positively correlates with the isoform-specific size of their C-terminal glycine-rich domains.

1.2. Splicing and disease

Until now, over 200 diseases could definitively be linked to mutations affecting splicing patterns [150] and computational predictions, comparative genomics and transcriptome profiling indicate that it is an even higher number that can lead to disease states [151]. Disease-causing mutations can range from mutations within 3'ss or 5'ss over mutations within silencer or enhancer motifs to mutations affecting splicing regulatory proteins or snRNPs itself [151]. The most abundant type, however, are

mutations that affect splice sites or regulatory elements [152] and depending on their location, lead to exon skipping, intron retention, cryptic splice site usage or affect the ratio of alternatively spliced mRNA isoforms. Among the first cases of human disease involving aberrant splicing were the β-globin thalassemia mutations and mutations affecting the SMN-2 gene, causing spinal muscular atrophy [73, 153-155]. For β-globin, a G>A mutation at position 1 of the intron, destroying the canonical GT of the 5'ss, was described, generating two types of abnormally spliced mRNAs, which do not encode for normal β-globin protein [153]. On the other hand, exon 7 of SMN-2 underlies a combinational control of SREs. The exon is bound by Tra2 which interacts with other SR proteins and thereby stimulates the inclusion of this exon. If a C>T transition at position +6 in exon 7 occurs, a new ISE is created which is bound by hnRNPA1, thereby inducing skipping of this exon, leading to a protein which is defective for self-association and SMN self-oligomerization [154, 155]. An example for mutations affecting a splicing regulatory protein itself is the degenerative muscle disease amyotrophic lateral sclerosis (ALS). It could be shown that ALS is caused by a mutation within the glycine-rich domain of hnRNPA1 protein, altering the dynamics of RNA granule assembly which is important for post-transcriptional regulation within neurons [156]. Besides those diseases, several cases of splicing errors leading to cancer and metastasis have been described which were linked to mutations in genes encoding for $U2AF^{35}$ or SRSF2 [73]. By now, several therapeutic approaches have been developed to counteract splicing errors in disease. Antisense-oligonucleotides can bind to specific target sequences on the pre-mRNA and block interactions of the spliceosome or splicing regulatory elements [151]. Clinical trials have been carried out e.g. for Duchenne muscular dystrophy (DMD) which is caused by genomic deletions leading to splicing of *out-of-frame* mRNAs. Antisense-oligonucleotides induce exon skipping of several exons, restoring the open reading frame [151]. Those oligonucleotides

specifically bind to complementary nucleotide regions on the pre-mRNA and sterically interfere with binding of SREs or components or the splicing machinery.

1.3. Prediction of localization of SREs

Identifying mutations affecting splicing and predicting their outcome is critical for diagnostics and treatment of patients suffering from genetic disorders. Recently, two algorithms have been developed (ΔHZ_{EI} and ΔtESRseq) that showed a high predictive power by indicating the direction and severity of the induced splicing defects [157, 158]. The HEXplorer score is based on hexamer weights calculated by a RESCUE-type approach (Relative Enhancer and Silencer Classification by Unanimous Enrichment) [159]. Since ESE sequences act in a strict position-dependent manner, the scaled difference of occurrence was calculated between hexamers in the exon versus intron near constitutively used splice sites (ΔEI) [159]. The HEXplorer score, as a further improvement, identifies the average score of all hexamers overlapping with the index nucleotide (ΔHZ_{EI}) [157]. The resulting HEXplorer plot illustrates splice enhancing and silencing properties of the splice site neighborhood (Figure 8). Mutational effects can be visualized and, additionally, the most effective mutations calculated. On the other hand, ΔtESRseq is based on relative splicing strength scores (ESRseq) which were assigned via transfection experiments of 4096 6-mers within a three-exon minigene. The total ESRseq score changes (ΔtESRseq), also using overlapping hexamers, were then calculated to measure the impact of exonic splicing mutations [158].

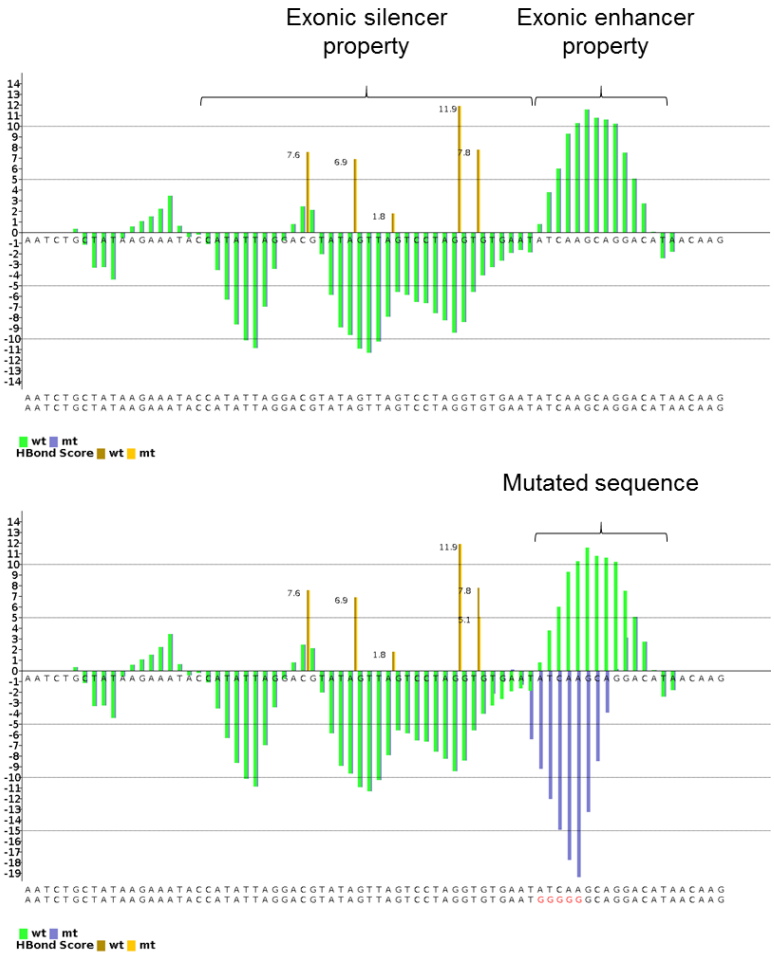


Figure 8 – HEXplorer plot
Depicted is an exemplary HEXplorer plot of HIV-1 exon 3. Regions showing HEXplorer scores below the horizontal axis represent sequences with exonic silencer properties whereas scores above the horizontal axis depict exonic enhancer properties. Additionally, HBond scores of all GTs are calculated (brownish). Point mutations that alter the HEXplorer plot are shown in purple. The HEXplorer can be accessed through the web page https://www2.hhu.de/rna/html/hexplorer_score.php.

1.4. The human immunodeficiency virus type 1 (HIV-1)

HIV-1 is the cause of the acquired immunodeficiency disease syndrome (AIDS). According to the WHO (World Health Organization), approximately 36.7 million people were living with HIV at the end of 2015 with 2.1 million new HIV infections and 1.1 million deaths due to AIDS in 2015. The disease is characterized by a severe loss of $CD4^+$ lymphocytes, leading to a progressive failure of the immune system and prevalently to death. After more than 30 years, research led to the development of effective antiretroviral drugs which efficiently suppress the virus and stop progression of the disease. However, so far these drugs do not kill dormant viruses and thus any

cure is not sterile yet. According to the WHO, only 46% of people living with HIV received lifelong antiretroviral treatment since it is cost-intensive, requires regular medication and is therefore almost only available in industrial countries. Therapeutic strategies face the problem of dormant virus reservoirs and the rise of multidrug-resistant viral subspecies due to the high genetic variability of HIV [160-162]. The reason for the high variability lies within the viral reverse transcriptase with an error rate of about 1.2×10^{-5} to 6.7×10^{-4} mutations per base per replication cycle [163]. In consequence, the increase of drug-resistant viral strains and the still very high prevalence of the disease raise the strong need for novel therapeutic strategies and thus, highlight the importance of an extended knowledge of the mechanisms of viral gene expression.

1.4.1. HIV-1 life cycle

HIV-1 belongs to the family of retroviruses and is an enveloped virus with two copies of a (+)-stranded RNA genome within an internal nucleocapsid (Figure 9). During infection, the HIV-1 viral surface glycoprotein gp120 interacts with cells expressing the CD4 receptor that is found on the surface of mononuclear cells such as T-cells, macrophages, and dendritic cells [164-166] (Figure 10). Furthermore, different HIV-1 isolates depend on one of the two different co-receptors for efficient entry *in vivo*: CCR5 or CXCR4 [165]. Here, the V3 loop of gp120 mainly determines the HIV-1 tropism [167].

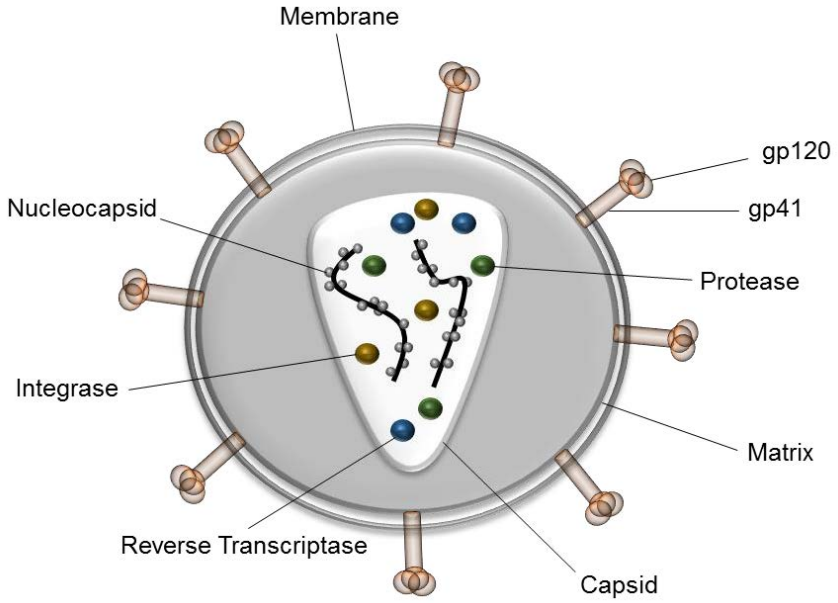


Figure 9 – HIV-1 virus particle structure
The HIV-1 particle is surrounded by a host cell-derived lipid membrane in which the glycoprotein is embedded which interacts with the matrix. The capsid contains two copies of the (+)-strand RNA genome that are associated with nucleocapsid proteins. Additionally, proteins required for viral infectivity are located within the capsid.

Subsequently, the virus fuses with the plasma membrane of the target cell and the capsid is released into the cytoplasm. The viral RNA is then reversely transcribed by the capsid-associated viral reverse transcriptase which uses a packaged host cell tRNA as primer to initiate plus-strand DNA synthesis [168]. Following completion of second strand synthesis, the DNA, which is associated with viral and cellular proteins, forming the pre-integration complex (PIC), is actively transported into the nucleus and integrated into the host genome as provirus [169-171]. The provirus is terminated by two long-terminal repeats (LTRs) at its 5' and 3' ends. HIV-1 gene expression is under the control of the viral promoter within the 5'-LTR which can be recognized by the host transcription machinery. However, efficient viral transcription by the cellular RNA polymerase II (Pol II) requires the viral protein Tat which binds to an U-rich bulge within the TAR element *(transactivation-responsive region)*, a stem-loop structure located at the 5'-end of each nascent HIV-1 transcript [172, 173]. Tat interacts with the positive transcription elongation factor b (P-TEFb) which hyperphosphorylates the C-terminal

domain (CTD) of Pol II, thereby promoting elongation [174]. However, only a single transcript is generated encoding the whole viral genome and alternative splicing enables the controlled expression of HIV-1 encoded enzymes, regulatory, accessory and structural proteins.

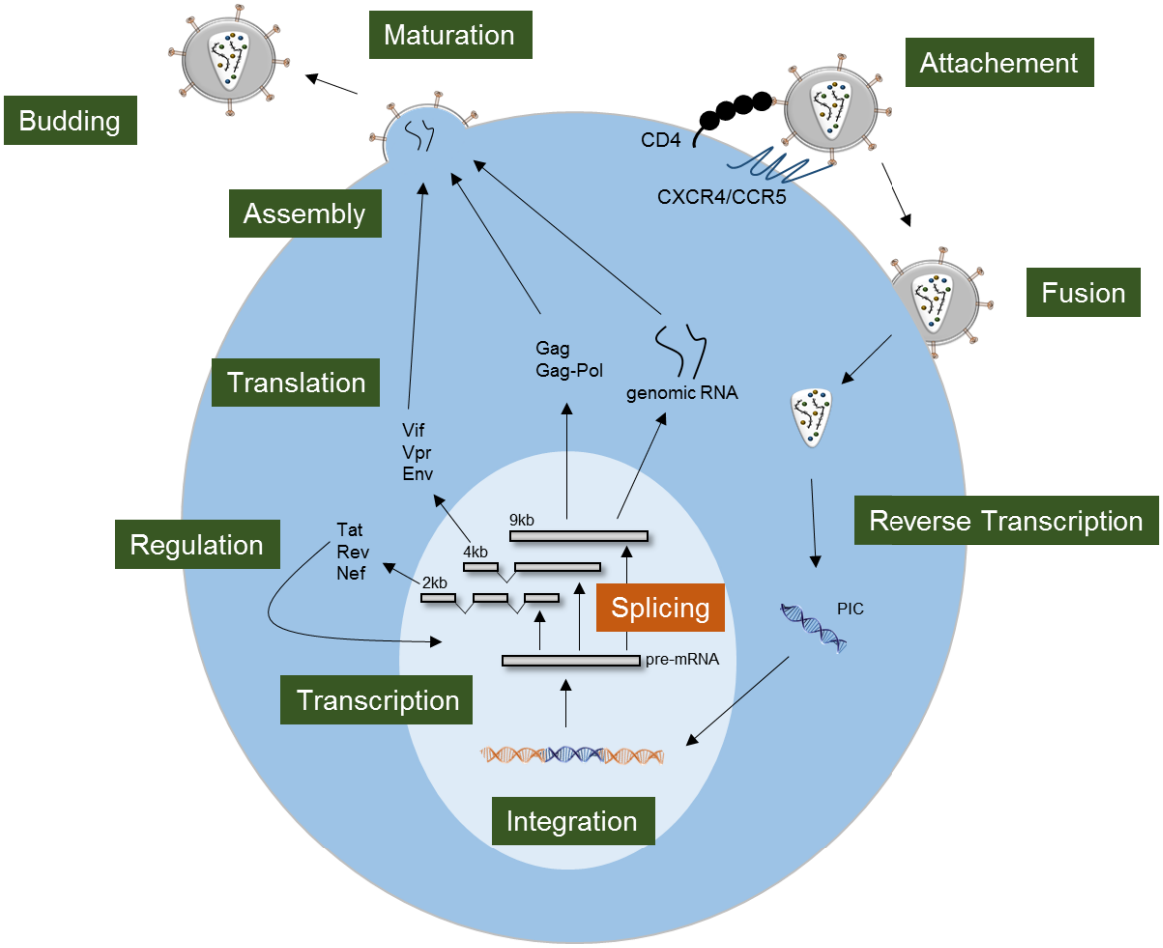


Figure 10 – HIV-1 replication cycle
After receptor-mediated entry into the host cell, the viral RNA is reversely transcribed, subsequently transported into the nucleus and integrated into the host's genome. After integration, the DNA is transcribed into the viral pre-mRNA, which is heavily spliced leading to the generation of a great variety of mRNA species which can be assigned to three different mRNA classes: intronless 2kb, intron-containing 4kb and unspliced 9kb mRNAs. 4kb and 9kb mRNAs are translated into structural and enzymatic proteins or serve as genomic RNA, which is incorporated into nascent virions. Viral particles are released and mature to new infectious virus particles.

1.4.2. HIV-1 splicing

Since translation starts at the 5' end of an mRNA by binding of the small ribosomal subunit to the most cap-proximal start codon, scanning of the 9kb HIV-1 primary transcript leads to sole translation of Gag and Gag/Pol open reading frames (ORF)

[175]. To circumvent this, HIV-1 uses alternative splicing to remove cap-proximal AUGs by utilization of the cellular splicing machinery, resulting in the generation of more than 50 viral mRNA isoforms [173, 176-178]. The emergence of most HIV-1 isoforms results from the variable usage of four different splice donor sites (D1, D2, D3 and D4) and eight different splice acceptor sites (A1, A2, A3, A4cab, A5 and A7), giving rise to three individual mRNA classes, intronless 2kb, intron-containing 4kb and unspliced 9kb mRNAs, that are expressed in a temporal order during the viral life cycle [179-181] (Figure 11).

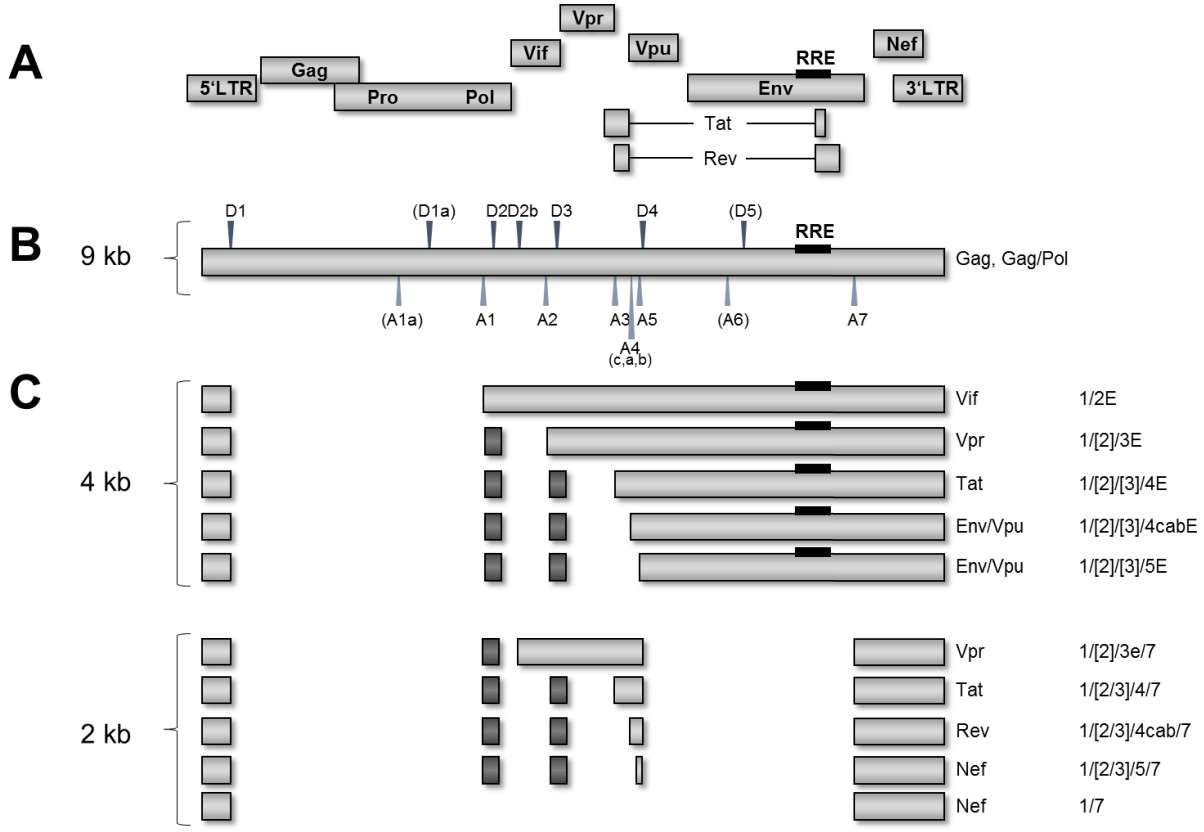


Figure 11 – The HIV-1 genome
(**A**) The eight open reading frames of HIV-1 (Gag/Pol, Env, Vif, Vpr, Vpu, Tat, Rev and Nef) are shown as grey rectangles. Furthermore, the Rev Response Element (RRE) localized within the Env ORF is depicted. (**B**) The unspliced 9kb message is coding for *gag* and *pol* and serves additionally as genomic RNA. Shown are the positions of predominantly used 5'ss and 3'ss. Splice sites shown in brackets only exist in some HIV strains. (**C**) 4kb and 2kb mRNAs include non-coding exon 1, whereas non-coding exons 2 and 3 (dark grey) are alternatively spliced. The 2kb mRNAs are additionally spliced from 5'ss D4 to 3'ss A7.

Shortly after successful infection, 2kb mRNAs encoding for Tat, Rev and Nef are generated though excessive splicing of the primary transcript. Tat and Rev shuttle back

into the nucleus to either promote transcription elongation, or serve as a nuclear export transactivator for intron-containing mRNAs by binding to the Rev Response Element (RRE). Coincidentally, intron-containing 4kb mRNAs with ORFs for Vif, Vpr, Vpu and Env are transcribed. Furthermore, unspliced 9kb mRNAs are transported into the cytoplasm and either translated or used as genomic RNA for newly formed viral particles. To allow the generation of intron-containing or unspliced mRNAs, HIV-1 splicing must be rather inefficient and indeed, many viral 5' and 3' splice sites show scarce conformity with the human consensus sequences (Table 1).

Table 1 – Intrinsic strengths of HIV-1 splice sites
The intrinsic strengths of HIV-1 5'ss were calculated with the HBond score, the 3'ss using the MaxEnt score.

Splice donor	Sequence	HBond score
D1	CtGGTGAGTAc	17.5
D2	aAGGTgAaggg	10.7
D2b	CAGGTgAtgAT	12.4
D3	aAGGTAgGatc	14.0
D4	gcaGTAAGTAg	15.7

Splice acceptor	Sequence	MaxEnt
A1	aattttcgggtttattacaggga	6.41
A2	ctattttgattgtttttcagaat	9.43
A3	ctgctgtttatccatttcagaat	9.76
A4c	gtgttgctttcattgccaagttt	3.74
A4a	agtttgtttcatgacaaaagcct	-1.75
A4b	tttcatgacaaaagccttaggca	-4.09
A5	ttaggcatctcctatggcaggaa	4.01
A7	attcaccattatcgtttcagacc	7.15

HIV-1 splice donor D1 and D4 are most efficiently used, which in turn is reflected by the highest similarity to the human U1 snRNA consensus sequence, whereas HIV-1 3'ss are rather intrinsically weak except for A2 and A3 [60, 182]. However, many *cis*-regulatory elements regulating HIV-1 splice site usage have been identified since then that determine splice site usage and the relative abundance of different viral proteins [183] (Figure 12).

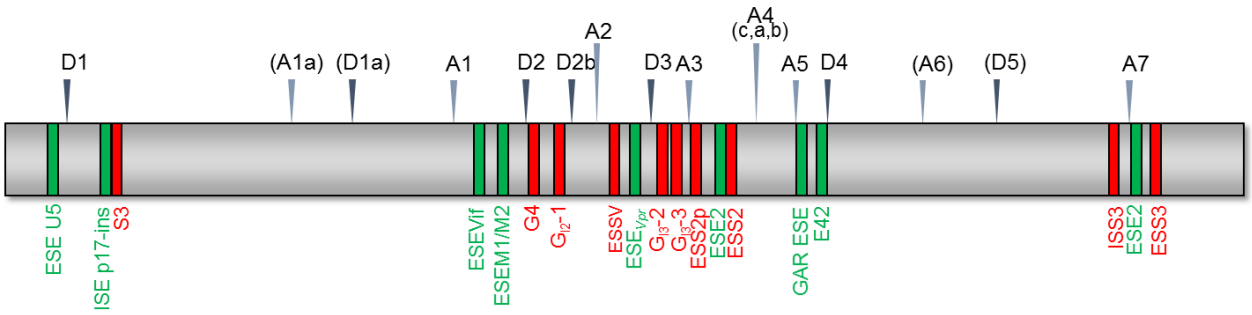


Figure 12 – HIV-1 splicing regulatory elements
Depicted are the locations of the known positively acting (green) and negatively acting (red) SREs regulating viral pre-mRNA splicing. This illustration was modified according to [173] and supplemented with more recent data.

1.4.3. HIV-1 proteins

i. Tat, Rev and Nef

Intronless, 2kb mRNAs encode for the regulatory proteins Tat, Rev and Nef.

Tat (Transactivator of Transcription) is a 14-16 kDa sized, low abundant protein that is essential for viral transcription by trans-activating efficient transcription from the LTR promotor [184, 185]. Tat is formed by splicing at 5'ss D1 to 3'ss A3 and D4 to A7 and can include the non-coding leader exon 2, exon 3 or both [176, 183]. Splice acceptor usage of A3 is repressed by two exonic silencer elements, termed ESS2 and ESS2p, which are bound by hnRNP A/B and hnRNP H, respectively, thereby significantly downregulating *tat* mRNA levels [132, 186-188]. Furthermore, an element positively influencing A3 usage was described, ESE2, which is bound by SRSF2 and competes with the bindingsite of hnRNP A1 to ESS2b [189, 190]. Besides that, splicing from D4 to A7 is a prerequisite for generating *tat*, *rev* and *nef* mRNAs, thereby eliminating the RRE-containing intron. Several SREs have been described to regulate A7 usage [183]. A negatively acting element located upstream of A7, ISS3, was shown to bind hnRNP A1 and competes with U2 snRNP binding to the branch point sequence [191]. Furthermore, two exonic splicing silencers downstream of A7 have been described: ESS3a and ESS3b, negatively acting on A7 splicing by cooperative binding of hnRNP A/B and masking the binding site of a further ESE element [118, 183, 187, 192-195].

Additionally, ESE2 was identified, which is bound by SRSF1 and thereby promotes A7 usage [183, 187, 193].

Rev (Regulator of Virion Protein Expression) is a 19 kDa sized protein that is essential for the export of intron-containing and unspliced viral mRNAs by binding to the RRE within the Env coding sequence [196]. Binding of Rev leads to the oligomerization of further Rev molecules that then interact with the cellular export receptor CRM1, exporting the intron-containing viral mRNA through the nuclear core complexes into the cytoplasm [197-200]. Subsequently, Rev shuttles back into the nucleus by binding to the nuclear import factor importin-β [201]. Rev is encoded by two exons and Rev-expressing mRNAs differ in including non-coding exons 2 and 3 and the usage of acceptors 4cab. Those acceptors are all intrinsically very weak and their usage depends on the strong guanosine–adenosine-rich ESE (GAR ESE), bound by SRSF1 and SRSF5, which does not only activate those 3' splice sites but also A5 and the downstream-lying donor D4 which is required for splicing of the *rev*, *nef* and *env* mRNAs [2, 202, 203]. In addition to that, the presence and strength of D4 is essential for upstream splice acceptor usage since U1 snRNPs bound to D4 and U2 snRNPs at 3'ss A4cab or A5 form cross-exon complexes and thereby facilitate exon definition [203].

Nef (Negative Factor) is a 27-35 kDa sized protein with a broad array of functions and required for full HIV-1 virulence. Among others, Nef down-regulates CD4, MHC class I and II to counteract apoptosis and evade from the immune system [204, 205]. Nef is encoded by one exon, and *nef* mRNA variants differ in the inclusion of exon 2 or 3 and exon 5. Differential splice acceptor usage of A5 is regulated as described above.

ii. Env, Vif, Vpr and Vpu

Messages of the 4kb class are spliced, but still intron-containing and therefore rely on Rev for nuclear export. They encode for Env, Vif, Vpr and Vpu.

Env (Envelope) forms the viral envelope. Translation occurs from the bicistronic env/vpu mRNA. The *env* gene codes for the gp160 protein which is proteolytically cleaved into the mature glycoproteins gp120 and gp41 [206]. As described above, gp120 binds to the CD4 receptor, CCR5 and CXCR4, thereby enabling the attachment to the host cell and membrane fusion which is mediated by gp41 [165, 207-209]. Env transcripts can consists of exons 2, 3, 4cab und 5, whose variable inclusion is mediated by ESS2, ESS2p, ESE2 and the GAR ESE.

Vif (Viral infectivity factor) is a 23 kDa sized protein and an antagonist of the host cell restriction factor APOBEC3G (A3G) [210]. A3G is a cytidine deaminase that is incorporated into newly assembling virions and after infection, triggers excessive deamination of deoxycytidine to deoxyuridine during reverse transcription, resulting in G-to-A hypermutations within the HIV-1 genome. Vif counteracts the incorporation of A3G into virions by provoking its ubiquitination and proteosomal degradation [211, 212]. *Vif* mRNAs are generated by splicing from D1 to SA1, the inclusion of exon 2 and the retention of the downstream intron, which includes the *vif* start codon. The definition of exon 2 is facilitated by cross-exon interactions between splice sites A1 and D2 and binding of U1 snRNPs to D2 without being actually spliced that is essential for the recognition of A1 und thus, proper *vif* mRNA production [127, 130, 213]. Inclusion of exon 2 is furthermore promoted by several splicing regulatory elements. Upstream of D2, ESEs M1 and M2, which are both bound by SRSF1, and the SRSF4-dependent element ESEVif, increase splice donor usage whereas a silencer element (G4 - GGGG) overlaps with D2 and exerts negative effects on exon inclusion, potentially by sterical hindrance of U1 snRNP binding [60, 130, 213]. Additionally, a second splice donor, D2b, is involved in the generation of *vif* mRNAs. This splice donor has an even higher intrinsic strength than D2 but is significantly less used which can be attributed to a G run (G_{I2}-1), bound by hnRNP F/H, inhibiting its usage [127].

Vpr (Virion protein R) has a variety of functions. The 14 kDa sized protein is i.a. involved in the import of the pre-integration complex and the cause of the cell cycle arrest during G2-phase, leading to a higher viral transcription rate [214-216]. In a comparable manner to *vif* mRNAs, *vpr* mRNAs are primarily formed by splicing from D1 to A2, the inclusion of exon 3 and the retention of the downstream intron, which includes the *vpr* start codon. Here, excessive usage of the intrinsically strong splice acceptor A2 is inhibited by a silencer element within exon 3, termed ESSV, which harbors a (Py/A)UAG binding motif and is bound by members of the hnRNP A/B protein family [217-219]. Binding of hnRNP A/B to ESSV interferes with the binding of $U2AF^{65}$ to the polypyrimidine tract of A2. Additionally, we could recently show that hnRNP D binds to the overlapping motif “UUAG” within ESSV [149]. Here we proposed that the counteractivity between hnRNP A/B and hnRNP D drives the relative formation of *vpr* mRNAs. Moreover, the inclusion of exon 3 is negatively influenced by a G run localized deeply within intron 3 which beyond this functions is involved in the mutually exon selection of *vif* and *vpr* mRNA formation [149]. On the contrary, since intron retention relies on cross-exon interactions of A2 and D3, exon 3 inclusion is positively influenced by Tra2α and Tra2β binding to the exonic enhancer ESE_{Vpr} [220].

Vpu (Virion protein U) is translated from the bicistronic mRNA that also encodes the *env* gene. Vpu counteracts cellular factors that inhibit the release of virions from infected cells by downregulating tetherin and CD4 [221, 222]. Like *env*, transcripts can consist of exons 2, 3, 4cab und 5, that are controlled by ESS2, ESS2p, ESE2 and GAR ESE as mentioned before.

iii. Gag and Pol

9kb mRNAs are not spliced and serve as genomic RNA for progeny virions and additionally, contain the open reading frames for the Gag-Pol precursor and Gag, encoding structural proteins and Pol, encoding viral enzymes. To translate both open

reading frames, a -1 ribosomal frameshift is necessary which is enabled by two elements: a “slippery site” and an RNA secondary structure [223-225]. Both genes are expressed as polyproteins, which are cleaved by the viral protease into p17 (Matrix), p24 (Capsid), p7 (Nucleocapsid), p6 and integrase, reverse transcriptase and protease, respectively.

1.5. Theses of this dissertation

Thesis I

Multiple splice-isoforms derived from a single primary transcript can contribute to proteomic diversity. The molecular machinery, however, that regulates proper splice site selection is prone to failure. Many diseases have been linked to aberrant splicing. Understanding signals that mediate alternative splicing will assist the development of algorithms as essential components for diagnostics and the development of therapeutic options. Endothelial dysfunction leading to cardiovascular diseases, the leading cause of death globally, has been associated with alterations within splicing patterns of numerous genes (**chapter 2**). During blood coagulation, fibrinogen is responsible for blood clot formation. Hereditary fibrinogen disorders can even occur through silent point mutations, if leading to activation of cryptic splice sites. A dense network of splicing regulatory elements (SREs) controls cryptic splice site selection (**chapter 3**). Activating cryptic splice sites in fibrinogen exon 7 might serve as a general model for splice site selection and exon end definition.

Thesis II

HIV-1 depends on the human splicing machinery for processing its primary transcript to provide mRNAs that can be efficiently translated by the scanning ribosome. Over 50 different mRNAs species arise by mainly using four splice donor and eight splice acceptor sites. Their selection is controlled by SREs that balance the emergence of all

viral transcripts dictating the viral protein levels. The levels of the viral protein Vif that counteracts the host restriction factor APOBEC3G is tightly regulated by SR and hnRNP proteins within exon 2/2b, exhibiting antagonistic functions (**chapter 4**). For hnRNP proteins, the G-rich domain plays a crucial role in silencing downstream donor usage as it is the case for hnRNP D binding to the ESSV element, regulating HIV-1 splicing within exon 3 and Vpr expression (**chapter 5**). Interfering with balanced splice site usage might represent a potential target for antiretroviral therapeutic approaches.

1.6. Thesen dieser Dissertation

<u>These I</u>

Die Vielzahl an Spleiß-Isoformen, die aus nur einem Primärtranskipt entstehen, können zur proteomischen Vielfalt beitragen. Die molekulare Maschinerie, die eine korrekte Spleißstellenselektion reguliert, ist jedoch fehleranfällig. Viele Krankheiten wurden mit aberrantem Spleißen in Verbindung gebracht. Aus diesem Grund ist das Verständnis über Signale, die das alternative Spleißen vermitteln, essentiell, um Algorithmen entwickeln zu können, die als wesentliche Komponenten in der Diagnostik und in der Entwicklung von Therapieoptionen eingesetzt werden könnten. Die endotheliale Dysfunktion führt zu Herz-Kreislauf-Erkrankungen, der häufigsten Haupttodesursache weltweit, und wurde mit Veränderungen im Spleißmuster vieler Gene in Zusammenhang gebracht (**Kapitel 2**). Fibrinogen ist verantwortlich für die Thrombenbildung während der Blutgerinnung. Erbkrankheiten, die die Bildung von Fibrinogen verändern, können auch durch stille Punktmutationen auftreten, wenn diese zur Aktivierung von kryptischen Spleißstellen führen. Ein dichtes Netzwerk von spleißregulatorischen Elementen (SREs) kontrolliert die kryptische Spleißstellenselektion (**Kapitel 3**). Die Aktivierung kryptischer Spleißstellen im

Fibrinogen Exon 7 könnte als allgemeingültiges Modell der Spleißstellenselektion und der Definition des Endes eines Exons dienen.

These II

Damit alle HIV-1 mRNAs vom Ribosom effizient translatiert werden können, ist das Virus für die Prozessierung seines Primärtranskripts auf die humane Spleißmaschinerie angewiesen. Über 50 verschiedene mRNA Spezies entstehen durch die hauptsächliche Nutzung von vier Spleißdonoren und acht Spleißakzeptoren. Die Spleißstellenselektion wird von SREs kontrolliert, die das Auftreten aller viralen Transkripte regulieren und somit die viralen Proteinlevel. Die Level des viralen Proteins Vif, das dem Wirtsrestriktionsfaktor APOBEC3G entgegenwirkt, werden von antagonistisch wirkenden SR und hnRNP Proteinen innerhalb von Exon2/2b streng reguliert (**Kapitel 4**). Bei hnRNP Proteinen spielt die G-reiche Domäne eine wichtige Rolle bei der Blockierung der stromabwärts liegenden Spleißdonoren. So wurde es für hnRNP D und dem ESSV Element gezeigt, welches das HIV-1 Spleißen innerhalb von Exon 3 und die Expression von Vpr reguliert (**Kapitel 5**). Eine Störung dieser ausbalancierten Spleißstellennutzung könnte als potentielles Target für antiretrovirale Therapien fungieren.

2. Critical Regulators of Endothelial Cell Functions: For a Change Being Alternative

The following review is published in Antioxid Redox Signal 22(14): 1212-1229. (doi: 10.1089/ars.2014.6023) by

Farrokh, S.,* ***Brillen, A. L.****, Haendeler, J., Altschmied, J., Schaal, H. (*shared first authors)*

Contribution

SF wrote parts of the manuscript dealing with endothelial cell function. ALB wrote parts of the manuscript dealing with splicing mechanisms. SF and ALB prepared the figures. JH, JA and HS participated in its design and coordination and helped to draft the manuscript.

Abstract

Significance: The endothelium regulates vessel dilation and constriction, balances hemostasis, and inhibits thrombosis. In addition, pro- and anti-angiogenic molecules orchestrate proliferation, survival, and migration of endothelial cells. Regulation of all these processes requires fine-tuning of signaling pathways, which can easily be tricked into running the opposite direction when exogenous or endogenous signals get out of hand. Surprisingly, some critical regulators of physiological endothelial functions can turn malicious by mere alternative splicing, leading to the expression of protein isoforms with opposite functions. Recent Advances: While reviewing the evidence of alternative splicing on cellular physiology, it became evident that expression of splice factors and their activities are regulated by externally triggered signaling cascades. Furthermore, genome-wide identification of RNA-binding sites of splicing regulatory

proteins now offer a glimpse into the splicing code responsible for alternative splicing of molecules regulating endothelial functions. Critical Issues: Due to the constantly growing number of transcript and protein isoforms, it will become more and more important to identify and characterize all transcripts and proteins regulating endothelial cell functions. One critical issue will be a non-ambiguous nomenclature to keep consistency throughout different laboratories. Future Directions: RNA-deep sequencing focusing on exon–exon junction needs to more reliably identify alternative splicing events combined with functional analyses that will uncover more splice variants contributing to or inhibiting proper endothelial functions. In addition, understanding the signals mediating alternative splicing and its regulation might allow us to derive new strategies to preserve endothelial function by suppressing or upregulating specific protein isoforms.

ANTIOXIDANTS & REDOX SIGNALING
Volume 22, Number 14, 2015

DOI: 10.1089/ars.2014.6023

Therapeutics

FORUM REVIEW ARTICLE

Critical Regulators of Endothelial Cell Functions: For a Change Being Alternative

Sabrina Farrokh,[1,*] Anna-Lena Brillen,[2,*] Judith Haendeler,[1,3] Joachim Altschmied,[1,#] and Heiner Schaal[2,#]

Abstract

Significance: The endothelium regulates vessel dilation and constriction, balances hemostasis, and inhibits thrombosis. In addition, pro- and anti-angiogenic molecules orchestrate proliferation, survival, and migration of endothelial cells. Regulation of all these processes requires fine-tuning of signaling pathways, which can easily be tricked into running the opposite direction when exogenous or endogenous signals get out of hand. Surprisingly, some critical regulators of physiological endothelial functions can turn malicious by mere alternative splicing, leading to the expression of protein isoforms with opposite functions. ***Recent Advances:*** While reviewing the evidence of alternative splicing on cellular physiology, it became evident that expression of splice factors and their activities are regulated by externally triggered signaling cascades. Furthermore, genome-wide identification of RNA-binding sites of splicing regulatory proteins now offer a glimpse into the splicing code responsible for alternative splicing of molecules regulating endothelial functions. ***Critical Issues:*** Due to the constantly growing number of transcript and protein isoforms, it will become more and more important to identify and characterize all transcripts and proteins regulating endothelial cell functions. One critical issue will be a non-ambiguous nomenclature to keep consistency throughout different laboratories. ***Future Directions:*** RNA-deep sequencing focusing on exon–exon junction needs to more reliably identify alternative splicing events combined with functional analyses that will uncover more splice variants contributing to or inhibiting proper endothelial functions. In addition, understanding the signals mediating alternative splicing and its regulation might allow us to derive new strategies to preserve endothelial function by suppressing or upregulating specific protein isoforms. *Antioxid. Redox Signal.* 22, 1212–1229.

Endothelium—Function and Dysfunction

THE CARDIOVASCULAR SYSTEM composed of heart and blood vessels transports blood through the organism to deliver oxygen and nutrients to all organs as well as carbon dioxide and metabolic end products to the lungs and excretory organs for exhalation and disposal. The border between the blood stream and the surrounding tissues is the vessel wall, which—in all larger vessels—principally consists of three layers (Fig. 1). The outer layer or adventitia is entirely composed of connective tissue containing elastic and collagen fibers synthesized by fibroblasts. It is followed by the media, which is made up of vascular smooth muscle cells required for vasoconstriction and -dilation; in veins, the media is much thinner than in arteries. Both these layers are not present in capillaries, the smallest of the body's blood vessels. The innermost layer, the intima, is separated from the media by an elastic membrane, the elastica interna. It consists of a monolayer of endothelial cells, which are directly at the interface with the blood stream. The endothelium plays a key role in many physiological processes such as regulation of vascular tone, hemostasis, thrombosis, and angiogenesis.

For a long time, it was assumed that the endothelium mainly represents an inert barrier, which simply separates

[1]Heisenberg-Group—Environmentally-Induced Cardiovascular Degeneration, IUF—Leibniz Research Institute for Environmental Medicine, Düsseldorf, Germany.
[2]Institute for Virology, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany.
[3]Central Institute of Clinical Chemistry and Laboratory Medicine, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany.
*These two authors contributed equally to the work.
#Shared senior authors.

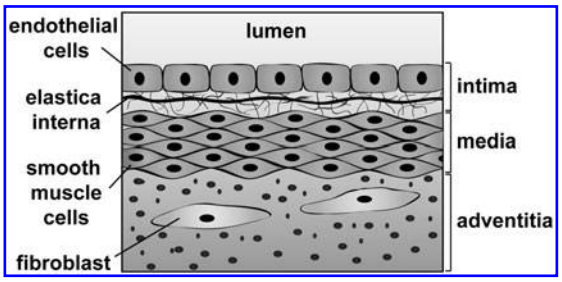


FIG. 1. Structure of the vessel wall. Schematic picture of a cross section through the wall of a large vessel showing the different layers and cell types.

circulating blood from surrounding tissues. In 1980, Furchgott and Zawadzki discussed for the first time an endothelial-dependent vasodilation as a response to acetylcholine (47). Later, it was discovered that nitric oxide (NO) is the key signal for vessel relaxation (104, 115). The major source of NO in the endothelium is the endothelial nitric oxide synthase (eNOS; the HGNC symbols of proteins relevant for the chapters on specific splice variants are listed in Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/ars). eNOS catalyzes the reaction from L-arginine to L-citrulline and NO. The most important physiological stimulus for eNOS activation, and by this the production of NO, is shear stress on the vessel wall (118). Furthermore, the activation of eNOS can be triggered by mediators such as acetylcholine or bradykinin binding to membrane receptors. NO released toward smooth muscle cells activates soluble guanylate cyclase in these contractile cells, which synthesizes the second messenger cyclic guanosine monophosphate. This signaling results in a decreased concentration of calcium in the cell, which, in turn, reduces the activity of the myosin light chain kinase. Since this kinase is responsible for initiation of vasoconstriction by phosphorylation of the motor protein myosin, reduced enzymatic activity results in relaxation of blood vessels (117).

Besides its role in dilation and constriction of blood vessels, the endothelium regulates hemostasis and inhibits thrombosis. Hemostasis is a physiological defense mechanism against bleeding due to vessel wall injury mediated by the coagulation system [for a review, see ref. (155)]. Thrombosis describes the pathophysiological process during which formation of thrombi, consisting of blood clots and aggregated platelets inside the vessel, obstructs the blood flow, which can lead to cardiac infarction or stroke. The appearance of thrombotic events is the consequence of an imbalance between pro- and anti-coagulative systems. Under physiological conditions, endothelial cells express anticoagulants and release prostacyclin and NO to suppress monocyte and platelet adhesion and aggregation of the latter [for a review, see refs. (54, 95)].

Endothelial cells not only have anti-coagulatory and antithrombotic properties, but also play a critical role in angiogenesis, the formation of new blood vessels from preexisting ones. This process requires a complex, well-balanced regulation of pro- and anti-angiogenic molecules and involves proliferation and migration of endothelial cells. Angiogenesis is critical during development; in the adult organism, it occurs mainly in wound healing and during the menstrual cycle. Enhanced or reduced angiogenesis will result in pathologic conditions, such as ischemia in diabetes or tumor vascularization; hence, it should be tightly controlled [for a review, see ref. (31)]. Angiogenesis is characterized by an initial vasodilation, followed by an increase in permeability and destabilization of the vessel wall, which then allows sprouting by proliferation and migration of endothelial cells. The most critical pro-angiogenic regulator is vascular endothelial growth factor A (VEGFA). VEGFA binding to VEGF receptor 2 (VEGFR2) on endothelial cells triggers a complex signaling cascade, which, on one hand, elevates eNOS activity, leading to vasodilation and, on the other hand, induces proliferation and migration of angiogenic endothelial cells (76). Besides VEGFA, a number of other pro-angiogenic molecules have been described, for example, several other growth factors such as transforming growth factor β1 (TGFβ1), platelet-derived growth factor BB, fibroblast growth factors, and cell surface molecules such as integrins. A further regulator of angiogenesis primarily expressed in endothelial cells is endoglin (ENG), an integral membrane glycoprotein and accessory coreceptor for TGFβ1. To prevent uncontrolled vessel outgrowth, there are negative feedback regulators such as angiostatin, endostatin, and thrombospondin. Most of these inhibitors are extrinsic to endothelial cells. Interestingly, endothelial cells also express anti-angiogenic molecules such as vasohibin 1 to induce a self-regulating, feedback inhibition response (160).

The response of endothelial cells to soluble factors and their receptors requires the conversion of signals, which in many cases is achieved by changes in transcription factor activities, leading to an altered cellular transcriptome. We have recently characterized the transcription factor grainyhead-like 3 (GRHL3) as a central regulator of endothelial cell migration and the maintenance of endothelial integrity by suppressing apoptosis (93).

The many functions of the endothelium explain why its dysfunction is a central event in many cardiovascular diseases. In general, the term endothelial dysfunction is used to describe a shift toward reduced vasodilation and impaired anti-inflammatory and anti-thrombotic capacity of the endothelium. It is associated with an increased production of reactive oxygen species (ROS), termed oxidative stress, and with a reduced NO-bioavailability. Oxidative stress occurs in cardiovascular diseases, for example, atherosclerosis or diabetes and other pathophysiological conditions such as ischemia/reperfusion or heart failure (38, 142) and during aging (39). oxidize cysteine residues and, thus, regulate the functions of many proteins, for example, kinases, phosphatases, and transcription factors (20, 93, 100). Moreover, ROS can directly damage cellular macromolecules such as proteins, lipids, and DNA. Decreased production of NO can have different reasons, for example, reduced expression and/or activation of eNOS (133, 164) or a lack of substrates and cofactors, which are required for NO production (120) and an early NO degradation as a result of high ROS levels (62).

As described earlier, numerous factors, from soluble proteins over enzymes producing vasoactive substances to transcription factors, determine the maintenance and proper function of the endothelium. Interestingly, the primary transcripts for many of these proteins are subject to alternative splicing and in several instances, the translation products of their alternatively spliced transcripts have divergent and sometimes opposing functions. Splice site choice is changed

under conditions where endothelial dysfunction is observed, for example, during aging and senescence (61, 98, 102). Furthermore, the expression of splicing factors is modulated by oxidative stress (23). This suggests that alterations in the splicing patterns can contribute to changes in endothelial function. Therefore, we will discuss pre-mRNA splicing and highlight the roles of selected examples for protein isoforms derived from alternative splicing in endothelial function.

Pre-mRNA Splicing

Pre-mRNA splicing is an essential feature of eukaryotic gene expression. The process in which introns are excised and exons are precisely joined together is called splicing (Fig. 2). Splicing is catalyzed by the spliceosome, a large intricate protein complex. The spliceosome recognizes two conserved sequence elements, the 5′splice site (5′ss) or splice donor (SD) and the 3′splice site (3′ss) or splice acceptor (SA). The canonical SD, which accounts for 99% of all SDs, is characterized by an 11-nucleotide-long sequence, containing a GU-dinucleotide at the beginning of the intron (14), whereas the SA is composed of the branch point sequence (BPS), the polypyrimidine tract (PPT), and an invariant AG at the intron/exon border. The spliceosome, comprising five different small nuclear ribonucleoprotein particles (U1, U2, U4, U5, or U6 snRNPs) and numerous associated non-snRNP proteins, assembles at the splice sites. In addition, most eukaryotes express a minor type of spliceosome, in which U1 and U2 are replaced by U11 and U12. These introns are classified as U12-type and are marked by AT and AC dinucleotides at their intron termini (131). During compositional and conformational changes of the spliceosome, two consecutive transesterification reactions are carried out. In the first step, the phosphorus atom at the SD is attacked by the oxygen of the 2′-hydroxyl group of the BPS-adenosine. Second, the oxygen atom at the SD forms a bond with the phosphorus atom at the SA leading to the ligation of both exons and an excised lariat intron.

However, the chemically simple splicing reaction faces a much more complicated organization of the spliceosome (156, 165). The stepwise spliceosome assembly begins with formation of the Early (E) complex. Here, the 5′ end of the U1 snRNA recognizes the 11 nucleotide long SD sequence and forms an RNA duplex. Besides splicing *per se*, base-pairing of the U1 snRNA to the SD can also protect the mRNA from degradation (73), as well as inhibit premature cleavage and polyadenylation (72). Subsequent to E complex formation,

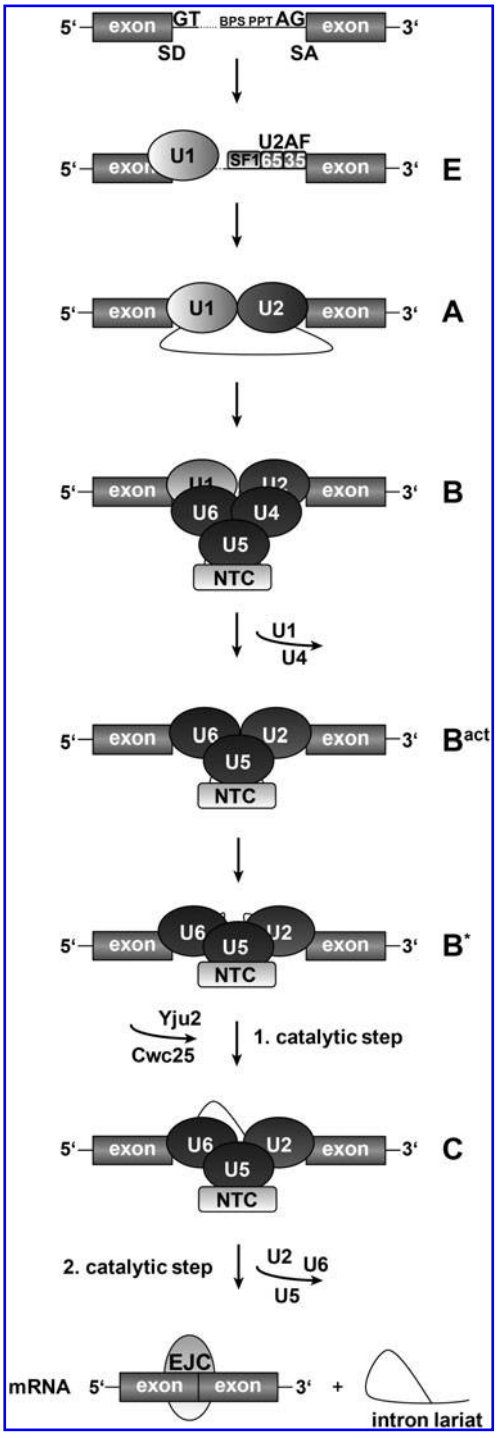


FIG. 2. Spliceosome assembly and splicing. The accurate recognition of exon/intron borders by the spliceosome is facilitated by conserved sequence elements: the splice donor (SD or 5′ss) and the splice acceptor (SA or 3′ss). Initially, the SD is recognized by the U1 snRNP, while the SA is bound by the non-spliceosomal proteins SF1 and both subunits of U2AF (65 and 35), generating the E complex. Particularly, the pyrimidine content of the PPT influences binding of U2AF65, and, thus, determines the intrinsic strength of the SA. This early (E) complex is then evolved in an ATP-dependent manner into A complex by subsequent binding of the U2 snRNP to the BPS, which releases SF1 from the pre-mRNA. Additional binding of the pre-assembled tri-snRNP (U4/U6*U5) and the NTC displacing U1 and U4 snRNPs forms the B complex, which is activated following structural and conformational changes (B^{act}). Additional ATPase remodeling forms the catalytically active B* complex, generating new binding sites for other proteins such as Yju2 and Cwc25. The first catalytic step progresses B* into C complex, which is followed by the second catalytic step. The remaining spliceosomal complex disassembles, releases the mature mRNA, a lariat intron, which will be degraded, and the remaining snRNPs and NTC proteins, which are recycled for the next splicing reaction. Coincidentally, a protein complex called EJC is loaded onto the mRNA, ~20–24 nucleotides upstream of the splice junction. 3′ss, 3′ splice site; 5′ss, 5′ splice site; BPS, branch point sequence; EJC, exon junction complex; NTC, NineTeen complex; PPT, polypyrimidine tract; SA, splice acceptor; SD, splice donor; snRNP, small nuclear ribonucleoprotein particle.

progression into the A complex requires the ATP-dependent association of U2 snRNP with the SA. Rearrangements follow in which the branch point adenosine is bulged out to serve as the nucleophile for the first catalytic step of splicing. After association of the U4/U6*U5 tri-snRNP, consisting of ~25 proteins, as well as 35 non-snRNP proteins join the spliceosome (156), forming the pre-catalytic B complex, which is activated *via* structural and conformational changes (B^{act} complex). Subsequently, several proteins are released from the spliceosome, whereas binding sites for the splicing factors Cwc25 and Yju2 are created, which are needed for the first step of splicing (B* complex) (111). Furthermore, during activation, the U6 snRNA simultaneously interacts with the SD and the U2 snRNA to bring the SD into close proximity to the branch point adenosine. However, the process by which the U1 snRNP is replaced with U6 snRNP seems to vary between higher eukaryotes (43) and *Saccharomyces cerevisiae* (137). Besides tri-snRNP binding, the essential NineTeen complex (NTC or Prp19/CDC5 complex) also participates in B complex formation and remains associated during both phases of transesterification, facilitating stable interactions of the U6 and U5 snRNPs with the pre-mRNA (18). The activated spliceosome conducts the first splicing step, which converts B* into C complex. Additional rearrangements of the spliceosomal complex follow before the second step of splicing in which both exons are then ligated and the lariat structure is degraded. After the second transesterification, the remaining spliceosomal components dissociate. Coincidentally, the exon junction complex (EJC), a protein complex involved in RNA quality control, export, and translation, is loaded onto the spliced mRNA. Another set of proteins associated with spliced mRNA are numerous serine/arginine-rich (SR) proteins, which have been reported to interact with the general export receptor TAP/NFX1, as well as the conserved mRNA export machinery TREX (transcription/export complex). Furthermore, the shuttling SR proteins SRSF1, SRSF3, and SRSF7 are tightly associated with the EJC (134).

However, additional information from *cis*-acting splicing regulatory elements is essential for proper splice site selection (63). In general, those elements are bound by sequence-specific RNA-binding proteins such as SR or heterogeneous nuclear ribonucleoparticle (hnRNP) proteins, which can either activate or repress splice site usage. SR proteins have classically been described as splicing enhancers, while hnRNP proteins have been considered repressors of splicing. However, genome-wide studies showing position-dependent RNA splicing maps and further systematic *in vitro* and *in vivo* analyses of different SR and hnRNP proteins with splicing enhancer-dependent reporter constructs could demonstrate that SR and hnRNP proteins can function as both, dependent on their position relative to the splice sites (37, 66, 90).

Approximately 94% of human genes produce at least two mRNA isoforms through alternative splicing, thereby including different combinations of exons into an mRNA from one primary transcript (159). Expression of different protein isoforms often results in profound alteration of their chemical and biological functions within the cell, and aberrant alternative splicing as well as mutations within the splicing machinery can result in numerous diseases (22). Cellular stress such as heat, starvation, or hypoxia can also affect the production of altered protein isoforms. Exon array experiments, examining hypoxia-related changes in alternative splicing in endothelial cells, revealed that altered splicing of several genes, which function in cytoskeletal remodeling and migration, leads to the expression of isoforms, positively contributing to the formation of new blood vessels (162). Another study showed that an extensive response of gene regulation exists to ensure a proper balance between cell survival and apoptosis under hypoxia in human umbilical vein endothelial cells, that is, the expression of different isoforms of interleukin 8 and hypoxia-inducible factor-1α (HIF-1α) (59). Furthermore, cellular stress can likewise lead to the modification of *trans*-acting factors such as splicing regulatory proteins. Changes in the homeostasis of ROS can lead to the alteration of the activity or concentration of splicing regulatory proteins. It could be shown that physiological concentrations of H_2O_2 result in a diminished RNA-binding affinity of the nuclear-restricted splicing regulatory protein hnRNP C to the pre-mRNA through phosphorylation of its C-terminal domain by protein kinase CK1α, which was assumed to regulate the post-transcriptional response to low H_2O_2, thereby releasing distinct mRNAs for nuclear export (74, 140). Furthermore, oxidative stress selectively decreases the expression of the splicing regulators phosphotyrosine binding (PTB) and hnRNP A2/B1 in human cancer cells, supposedly by an increase of proteasomal degradation. One consequence is the expression of an alternative splice variant of soluble guanylate cyclase with enhanced oxidation resistance as a potential adaptive response (23, 130).

In the next few chapters, we will describe several prominent examples of alternative splicing leading to the expression of protein isoforms, which affect endothelial functions in opposite directions.

Endothelial Nitric Oxide Synthase

The enzyme eNOS is the major NO source in the endothelium of mammalians. Functional eNOS is a homodimer that converts L-arginine to citrulline and NO with concomitant oxidation of $NADPH + H^+$ to nicotinamide adenine dinucleotide phosphate ($NADP^+$) (Fig. 3). Each subunit consists of an N-terminal oxygenase domain with binding sites for heme, L-arginine, and tetrahydrobiopterin (BH_4), a calmodulin (CaM)-binding region located in the central part of the protein and a C-terminal reductase domain with binding sites for the cofactor NADPH and the prosthetic groups FAD and FMN (141). Binding of the substrate L-arginine and the interaction with co-factors only occurs in the dimeric state of eNOS (25).

eNOS can be activated in a calcium-dependent and -independent way. The calcium-dependent activation of eNOS is triggered by the interaction with CaM (15). Endogenous agonists such as bradykinin and acetylcholine activate transmembrane receptors and induce phospholipase C signaling, which results in an increase of intracellular calcium levels. The calcium ions activate the regulatory protein CaM, which binds to the CaM-binding site of eNOS and facilitates an electron flux from the reductase to the oxygenase domain (128). In contrast, calcium-independent activation is mediated by shear stress and hormones such as estrogens and insulin (17, 80). Furthermore, the activity of eNOS is regulated by post-translational mechanisms, including translocation to the plasma membrane, phosphorylation, and interaction with other regulatory proteins and cofactors.

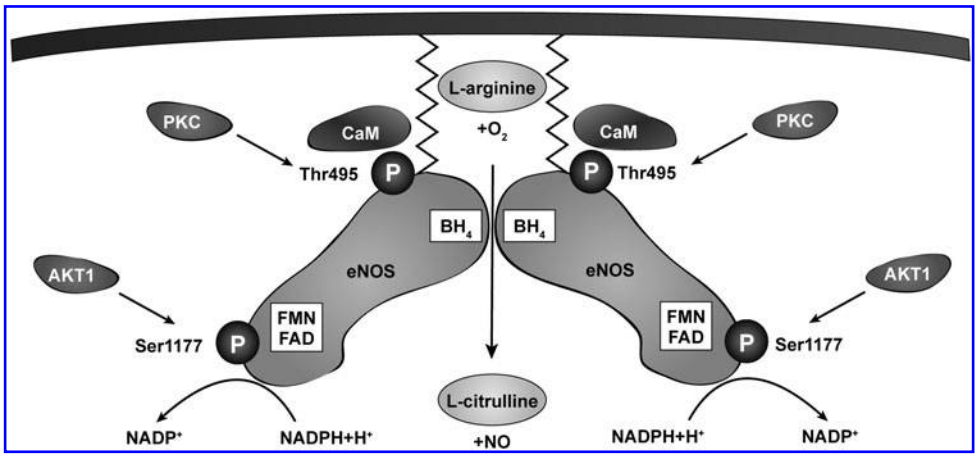


FIG. 3. eNOS structure, function, and regulation. Homodimeric eNOS catalyzes the formation of L-citrulline and NO from L-arginine and O_2 and the concomitant oxidation of $NADPH + H^+$ to $NADP^+$. This reaction depends on BH_4 and the prosthetic groups FMN and FAD, which are required for an electron flux from the reductase to the oxygenase domain. Calcium-dependent activation of eNOS depends on CaM binding. Furthermore, eNOS is activated by AKT1-dependent phosphorylation on Ser1177, whereas phosphorylation of Thr495 by PKC reduces eNOS activity by interfering with CaM binding. BH_4, tetrahydrobiopterin; CaM, calmodulin; eNOS, endothelial nitric oxide synthase; FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; $NADP^+$, nicotinamide adenine dinucleotide phosphate; NO, nitric oxide; PKC, protein kinase C.

The localization of eNOS at the plasma membrane plays an important role in its activity. In particular, the covalent attachment of a myristoyl moiety to a glycine at the extreme N-terminus of eNOS is absolutely required for the localization at the plasma membrane and maximum enzyme activity (125, 129). Another post-translational fatty acid modification influencing the cellular localization of eNOS is the reversible palmitoylation on cysteine 15 and 26 in the oxygenase domain. Palmitoylation of eNOS occurs in the Golgi and then directs the enzyme to the plasma membrane (136).

Besides myristoylation and palmitoylation, the interactions with the eNOS interacting protein (NOSIP) and the eNOS traffic inducer (NOSTRIN) also have an effect on the cellular localization of eNOS. NOSIP binds to the carboxy-terminal oxygenase domain and supports the translocation of eNOS from the plasma membrane to intracellular membranes, thereby inhibiting NO synthesis (28). Similar to NOSIP, NOSTRIN also binds to the oxygenase domain, but translocates eNOS from the plasma membrane to intracellular vesicles, leading to reduced eNOS activity (168).

However, eNOS activity is predominantly regulated by multi-site phosphorylation and dephosphorylation (41). So far, phosphorylation has been described for Tyr81 and Tyr657, Ser114, Ser615, Ser633, and Ser1177 as well as for Thr495 in the human protein. While phosphorylation of Ser615, Ser633, and Ser1177 activates eNOS, phosphorylation of Ser114 and Thr495 reduces eNOS activity (5). Phosphorylation of the tyrosine residues Tyr81 (44) and Tyr657 (40) is induced by shear stress, but the two modifications have opposite effects on eNOS activity. While Tyr81 phosphorylation by $pp60^{src}$ induces NO production, phosphorylation of Tyr657 by proline-rich tyrosine kinase 2 (PTK2) decreases eNOS activity (40, 46). This observation seems to provide a negative feedback insofar that phosphorylation of Tyr657 by PTK2 might limit the production of peroxynitrite from eNOS under conditions of increased shear stress (40). However, the reason for the simultaneous appearance of these two eNOS phosphorylations is not clear. The two most extensively investigated phosphorylation sites are Ser1177 and Thr495 (Fig. 3). Phosphorylation of Ser1177 is affected by a number of kinases, including protein kinase A, protein kinase B alpha (AKT1), and calcium/CaM-dependent protein kinase II (CaM kinase II) (45, 105). This phosphorylation increases the electron flux to the oxygenase domain and reduces the dissociation of the CaM-eNOS complex (99). In contrast, Thr495 is a target for protein kinase C and phosphorylation at this site prevents binding of CaM, thereby dampening eNOS activity (42). This suggests that the dual phosphorylation at these two residues is a major determinant of the net activity of eNOS. This regulation of eNOS by phosphorylation implicates that active dephosphorylation also plays a role in the control of enzyme activity. Indeed, dephosphorylation at Thr495 by protein phosphatase 1 (126) and at Ser116 by the Ca^{2+}/CaM-dependent phosphatase calcineurin (79) has been demonstrated.

An essential cofactor for eNOS activity is BH_4 (1). When BH_4 levels are low, eNOS becomes ''uncoupled'' and produces superoxide anions instead of NO (150, 166). Therefore, optimal levels of BH_4 are absolutely required for the production of NO. Under conditions of oxidative stress, BH_4 is rapidly oxidized, leading to reduced bioavailability of this cofactor and endothelial dysfunction (32).

Besides its dependence on cofactors, eNOS is also regulated by interactions with other proteins such as heat shock protein 90 (Hsp90) and the caveolar coat protein caveolin-1 (CAV1). Hsp90 is a chaperone responsible for protein trafficking and folding and is reversibly phosphorylated on tyrosine residues in response to various eNOS activating agonists (48, 153). Phosphorylated Hsp90 binds to eNOS and stimulates its catalytic activity (7). CAV1 is a structural protein located in specialized invaginations of the plasma membrane called caveolae (116). Binding of CAV1 requires

myristoylation and palmitoylation of eNOS and results in a decreased eNOS activity by preventing CaM binding (71, 129).

In addition to this multiplicity of modifications and interactions, alternative splicing was also described for eNOS in regulating vascular functions. The gene encoding for eNOS consists of more than 20 exons. Alternative splicing within intron 13 has been described in endothelial cells and gives rise to three alternative splice variants termed eNOS13A, eNOS13B, and eNOS13C (92). All of them contain a novel exon within this intron and use three different SAs, but a common polyadenylation signal leading to premature transcription termination (Fig. 4A). These alternatively spliced transcripts code for C-terminally truncated eNOS proteins lacking the reductase domain, which is important for binding of NADPH, FAD, and FMN. The splice variant eNOS13A was used as an example to further analyze the functions of the shortened eNOS isoforms. While eNOS13A on its own was catalytically inactive, co-expression with full-length eNOS suppressed the activity of the full-length protein. This was ascribed to the formation of heterodimers between the two proteins in which eNOS13A exerts a *trans*-dominant-negative effect (92) (Fig. 4C).

A mechanistic clue to these previously unknown splicing events of eNOS pre-mRNA was provided by the presence of a highly polymorphic $[CA]_n$ repeat in intron 13 (107). It was shown that a higher number of these CA repeats is associated with an increased risk to develop coronary artery disease (138). It was hypothesized that modulation of eNOS pre-mRNA splicing, regulated by the number of CA repeats, might be involved in promoting such diseases, as A/C-rich sequences have been shown to act as splicing enhancers both *in vivo* and *in vitro* (24). Consistent with this, it was shown that splicing efficiency in an eNOS minigene requires these CA repeats (68) and that the number of repeats influences eNOS isoform formation (92). The specific increase in eNOS13A mRNA observed with a higher number of CA repeats (Fig. 4C) led to the assumption that extended CA repeats induce a shift toward this *trans*-dominant-negative isoform, which might explain the enhanced risk for the development of coronary artery disease (92).

In an attempt to identify splicing regulatory proteins binding to the CA repeats by UV crosslinking experiments and mass spectrometry analysis, hnRNP L as well as the non-SR protein Y box-binding protein 1 (YB-1) was shown to specifically bind to this repetitive dinucleotide (68). Particularly for hnRNP L, it was demonstrated *via* depletion experiments that it functions as a splicing enhancer in the context of eNOS intron 13 (Fig. 4B). Later work indicated that this splicing enhancer functions *via* activation of the upstream SD (67). Furthermore, the CA repeats could also function as a splicing enhancer in a heterologous context as exemplified in four different human genes. Here, Systematic Evolution of Ligands by EXponential Enrichment (SELEX) analysis revealed that hnRNP L not only binds to continuous CA stretches but also binds to CA-rich clusters (67). Recently, these observations could be confirmed by genome-wide individual-nucleotide resolution Crosslinking and ImmunoPrecipitation (iCLIP) experiments demonstrating preferential binding of hnRNP L to introns (as in the case of eNOS) and 3′UTRs. Furthermore, it could be demonstrated that hnRNP L also has a splicing repressing activity when bound closely upstream of an SA, potentially *via* interference with recognition of the PPT (124). The second protein shown to interact with CA repeats, YB-1, has yet not been further investigated with regard to eNOS splicing regulation. However, it has at least been demonstrated that YB-1 binds to A/C rich exonic splicing enhancers, for example, in CD44, where it is seemingly required for inclusion of an alternative exon (139). In addition, it was recently shown that YB-1 is a spliceosome-associated protein supporting the recruitment of U2AF65 to the SA through direct protein–protein interactions (161).

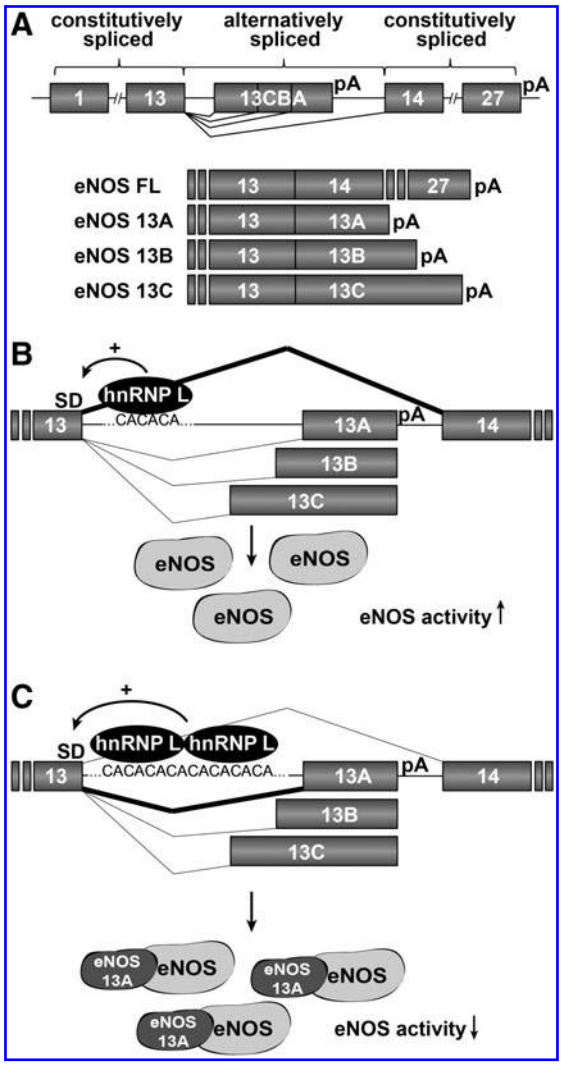


FIG. 4. Transcript architectures of eNOS splice variants and proposed splicing regulation. (A) Alternative recognition of splice sites between exons 13 and 14 of eNOS gives rise to three shortened splice variants (eNOS 13A, 13B, and 13C). The alternative polyadenylation signal within intron 13, spliced out in the full-length eNOS isoform (eNOS FL), leads to a premature transcription termination for the shortened variants. **(B)** Full-length eNOS splicing is regulated by polymorphic $[CA]_n$ repeats within intron 13, which are bound by hnRNP L. **(C)** An increased number of CA repeats correlates with an increase in eNOS13A mRNA formation, thereby leading to the formation of heterodimers between full-length eNOS and eNOS13A, which suppresses eNOS activity. hnRNP, heterogeneous nuclear ribonucleoparticle.

Besides these efforts to identify splicing regulatory proteins binding to *cis*-acting sequences within the eNOS primary transcript, also enzymes such as Cdc-like kinases and DNA topoisomerase I, known to phosphorylate SR proteins, were analyzed for their role in regulating alternative eNOS splicing. It was shown that stimulation of endothelial cells with tumor necrosis factor α (TNFα) increases the levels of eNOS isoforms 13A, B, and C, but not of full-length eNOS. This selective upregulation was abrogated by pharmacological inhibition and knockdown of DNA topoisomerase I, but not of Cdc2-like kinases, indicating that DNA topoisomerase I is necessary for alternative splicing regulation of the eNOS pre-mRNA (34). Furthermore, treatment with TNFα led to increased phosphorylation of SRSF4 and SRSF6, which was normalized by DNA topoisomerase I inhibition, indicating that both these SR proteins are involved in alternative splicing of eNOS and, thus, reduced NO production under pro-inflammatory conditions, possibly *via* the regulation of dynamic changes in SR protein phosphorylation.

Vascular Endothelial Growth Factor A

Angiogenesis is a well-balanced process regulated by endogenous activators and inhibitors [for a review, see ref. (31)]. In the mature organism, negative regulators predominate this balance and prevent uncontrolled vessel growth. Under physiological conditions, not only the action of positive regulators is increased, for example, in the case of wound healing and tissue repair, but also pathophysiological processes such as tumor growth can activate angiogenesis.

VEGFA is the most important stimulus for pro-angiogenic vessel growth. It belongs to the VEGF-family, including VEGFA, VEGFB, VEGFC, VEGFD, and placental growth factor [for a review, see ref. (64)]. VEGFA stimulates vascular growth by binding as a homodimer to the receptor tyrosine kinases VEGFR1 and VEGFR2 on endothelial cells [for a review, see ref. (108)]. Besides its role in angiogenesis, VEGFA is a survival factor in endothelial cells. On one hand, VEGFA protects cells against apoptosis induced, for example, by serum starvation through activation of the phosphatidylinositol PI3 kinase/AKT pathway (50); on the other hand, VEGFA stimulates the expression of the anti-apoptotic protein Bcl-2 and its relative A1 (49). In addition to its function in the cardiovascular system, VEGFA acts as a chemotactic agent in several non-endothelial cell types, for example, monocytes and granulocyte-macrophage progenitor cells (13, 21).

Both VEGFR1 and VEGFR2 are involved in angiogenesis regulated by VEGFA, with VEGFA demonstrating a higher affinity to VEGFR1 than to VEGFR2 (157). However, VEGFR2 is essential for pro-angiogenic signal transduction, which is based on a 10-fold higher tyrosine kinase activity of VEGFR2 in comparison to VEGFR1 (157). On binding of VEGFA, VEGFR2 dimerizes and becomes *trans*-autophosphorylated, thereby generating docking sites for modular Src homology 2 and PTB domains. Here, the two most important phosphorylation sites seem to be Tyr1175 and Tyr1214. While phosphorylation on Tyr1175 leads to phosphorylation and activation of phospholipase Cγ-1 and is involved in cell proliferation (57, 143), phosphorylation on Tyr1214 induces actin remodeling *via* activation of Cdc42 and the stress-activated protein kinase MAPK11, thereby triggering cell migration (83). This signaling results in the transcription and activation of cell survival genes and initiates vascular growth (Fig. 5). VEGFR1 has been shown to regulate the VEGFR2-mediated proliferative response of endothelial cells (16, 122). This is most likely due to formation of heterodimers between these two receptors (26). Besides the interaction with VEGFR1 and 2, VEGFA can also bind to neuropilin 1 (NRP1). It has been demonstrated that co-expression of NRP1 and VEGFR2 enhances the binding of VEGFA to the latter receptor, which led to the suggestion that NRP1 may modulate VEGFA-induced angiogenesis (135) (Fig. 5).

Gene expression of VEGFA is regulated on one hand by oxygen tension and on the other hand by signaling molecules such as growth factors, hormones, and oncogenes. Under hypoxic conditions, HIF-1α, which is constitutively transcribed and translated but promptly degraded in normoxia, becomes stabilized and—besides other target genes—activates the expression of VEGFA. Secretion of HIF-1α-induced VEGFA from hypoxic cells leads to the establishment of a VEGFA gradient, which initiates vessel sprouting to hypoxic tissue areas. In addition, many growth factors, hormones, and cytokines can induce the expression of VEGFA (114).

For a long time, VEGFA has exclusively been described as being pro-angiogenic by interacting with VEGFR2, leading to vascular growth as described earlier. Although multiple, alternatively spliced protein isoforms of VEGFA exist, which exhibit pro-angiogenic properties, other splice variants have been discovered encoding anti-angiogenic protein isoforms (60). Mechanistically, it was suggested that the anti-angiogenic isoforms studied bind to VEGFR2, but that phosphorylation only insufficiently takes places, leading to an attenuation of downstream signaling pathways (75) (Fig. 5).

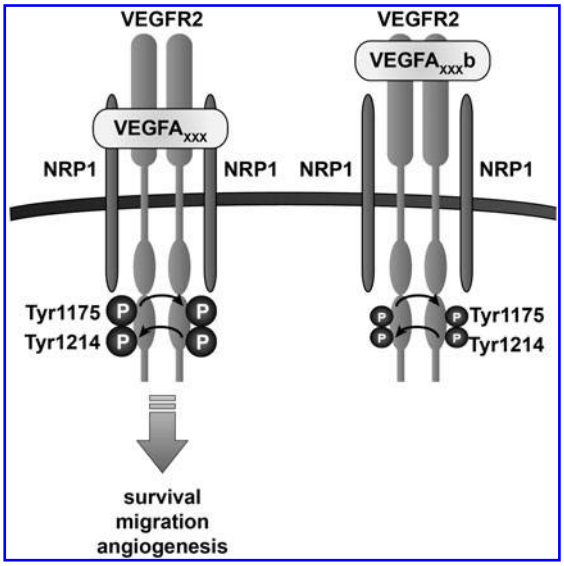


FIG. 5. VEGFA$_{xxx}$ and VEGFA$_{xxx}$b signaling. Both the pro-angiogenic and anti-angiogenic VEGFA isoforms interact with the homodimeric VEGFR2; the interaction with NRP1 enhances binding of VEGFA to VEGFR2. Compared with VEGFA$_{xxx}$b binding, the *trans*-autophosphorylation in the intracellular domain of VEGFR2 is more pronounced on an interaction with VEGFA$_{xxx}$ binding, thereby triggering survival, migration, and angiogenesis responses. NRP1, neuropilin 1; VEGFA, vascular endothelial growth factor A; VEGFR1 and VEGFR2, VEGF receptor 1 and 2.

Similar to the splicing pattern of many other alternatively spliced genes, VEGFA splicing is complex due to numerous alternative SD as well as SA sites. Currently, 13 isoforms of VEGFA have been described, which are generated in normal and pathological tissues by alternative splicing of exons 4, 5, 6, 7, and 8. These isoforms fall into two big classes: pro-angiogenic mediators named $VEGFA_{xxx}$ and anti-angiogenic molecules termed $VEGFA_{xxx}b$, with xxx representing the amino-acid number. The $VEGFA_{xxx}$ family consists of eight members ($VEGFA_{111}$, $VEGFA_{121}$, $VEGFA_{145}$, $VEGFA_{148}$, $VEGFA_{165}$, $VEGFA_{183}$, $VEGFA_{189}$, and $VEGFA_{206}$), the $VEGFA_{xxx}b$ family of five ($VEGFA_{121}b$, $VEGFA_{145}b$, $VEGFA_{165}b$, $VEGFA_{183}b$, and $VEGFA_{189}b$) (Fig. 6). As described earlier, the two VEGFA protein families have opposing functions. The pro-angiogenic $VEGFA_{xxx}$ isoforms are generated by using a proximal splice acceptor site (PSS) within exon 8, whereas the $VEGFA_{xxx}b$ family is formed by distal splice acceptor site (DSS) choice, thus resulting in two diverging exons, called exon 8a and 8b (Fig. 7). Both open reading frames code for six amino acids: exon 8a for CDKPRR and exon 8b for SLTRKD (4), leading to profound changes in structure and function, which cause either pro-angiogenic or anti-angiogenic signaling (Fig. 5). Furthermore, the usage of the PSS or DSS is combined with alternative splice events in exon 6 and 7. An example is exon 6a, which is included in splice variant $VEGFA_{189}$ and $VEGFA_{189}b$, but not in $VEGFA_{165}$ and its counterpart $VEGFA_{165}b$. Within a minigene, inclusion of exon 6a could be controlled by a putative 9-nucleotide-long splicing silencer sequence, which weakens its inclusion (158). Furthermore, the SR-related protein CAPER, also acting as a transcriptional co-activator for estrogen receptors and the transcription factor complex activator protein 1 (AP-1), was demonstrated to mediate splicing of VEGFA in cancer cells, especially in controlling the ratio of $VEGFA_{165}/VEGFA_{189}$ isoforms and, therefore, reducing vascular growth in tumor tissues. Since CAPER favors the expression of $VEGFA_{189}$, it was suggested that CAPER might target exon 6 by interacting with

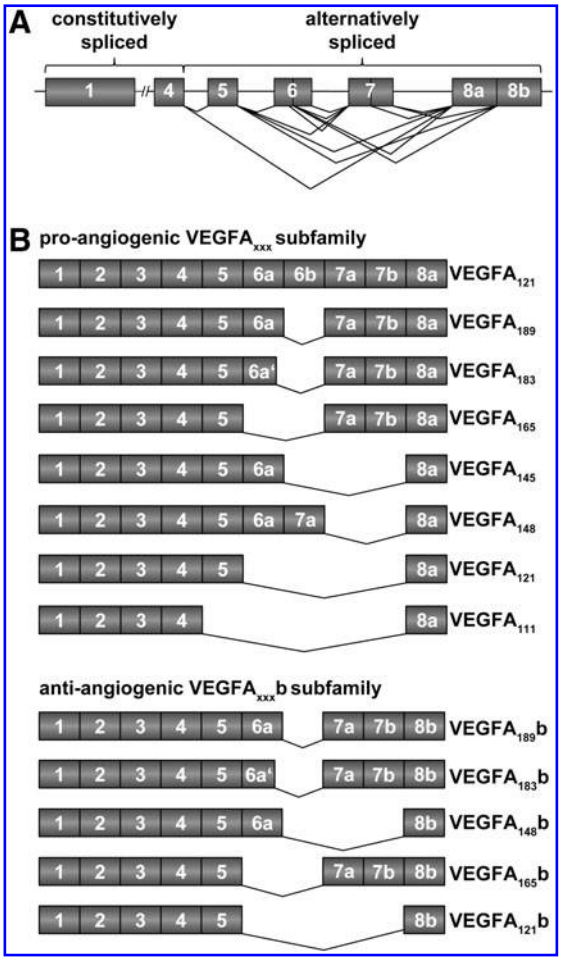


FIG. 6. Alternatively spliced VEGFA mRNA variants. **(A)** Alternative splicing at various SDs and SAs of VEGFA occurs in the region between exons 4 and 8a or the more downstream located exon 8b. **(B)** So far, a total of 13 transcript isoforms have been described as a result of alternative usage of SA8a or SA8b, resulting in a different terminal exon. The alternatively spliced VEGFA mRNA variants encode VEGFA isoforms harboring either pro- or anti-angiogenic properties. The pro-angiogenic subfamily ($VEGFA_{xxx}$) consists of eight family members, all containing the exon 8a encoding domain, while the five members of the anti-angiogenic subfamily ($VEGFA_{xxx}b$) contain the exon 8b encoding domain.

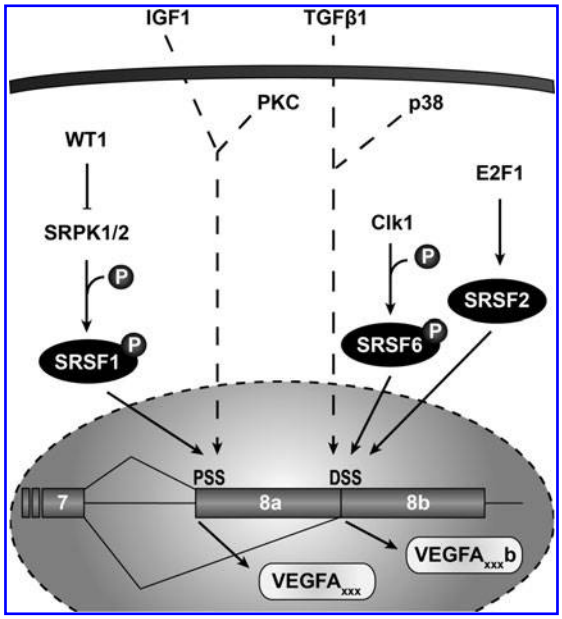


FIG. 7. VEGFA splicing regulation. Splice site regulation at the 3′-end of the VEGFA primary transcript leads to the formation of transcripts coding either for pro-angiogenic ($VEGFA_{xxx}$) protein isoforms or for anti-angiogenic ($VEGFA_{xxx}b$) isoforms. $VEGFA_{xxx}$ isoform encoding transcripts are generated by using the PSS whose usage is promoted by IGF1 and PKC as well as by SRPK1/2. Furthermore, phosphorylated SRSF1 was suggested to upregulate proximal splice site choice, which is inhibited by WT1. On the other hand, $VEGFA_{xxx}b$ encoding transcripts are generated by inclusion of exon 8b through DSS usage. Treatment with TGFβ1 favored DSS selection. p38 is also involved in DSS usage. Moreover, Clk1, which phosphorylates SRSF6, seems to be involved in DSS choice as well as E2F1, whose direct transcriptional target, SRSF2, activates DSS. DSS, distal splice acceptor site; IGF1, insulin-like growth factor; PKC, protein kinase C; PSS, proximal splice acceptor site; TGFβ1, transforming growth factor β1; WT1, Wilms' tumor suppressor 1 protein.

the spliceosome and, therefore, support exon 6 inclusion (65). A shortened version of exon 6a (exon 6a′) appears in isoforms $VEGFA_{183}$ and $VEGFA_{183}b$, resulting from usage of an SD site within exon 6a, which was shown to be highly conserved among animal species. Besides exon 6a, a truncated VEGFA splice variant, $VEGFA_{148}$, exists, which lacks exon 6 and the terminal parts of exon 7 and exon 8.

To identify the detailed molecular mechanisms of VEGFA splice site choice, current research is aimed at uncovering the key players being responsible for switching between pro- and anti-angiogenic VEGFA splice variants. Understanding those molecular mechanisms of splice site choice would enable us to develop new targeted therapies, for example, to suppress tumor-angiogenesis. Nowak *et al.* (110) examined the influence of environmental stimuli such as growth factors and the potential role of SR proteins in alternative VEGFA splicing in retinal pigment epithelial cells and podocytes. They could show that treatment with the cytokines insulin-like growth factor 1 (IGF1) and TNFα led to a switch in splice site choice, favoring the pro-angiogenic $VEGFA_{xxx}$ isoforms. A mirror-inverted splicing phenotype toward the anti-angiogenic isoform could be observed after incubation with TGFβ1. Here, the level of the anti-angiogenic $VEGFA_{xxx}b$ family significantly increased (Fig. 7). Besides using an inhibitor of the p38 mitogen-activated protein kinase pathway, distal splice site selection could be suppressed by inhibition of the Cdc-like kinase family, which has been earlier shown to be involved in phosphorylation of SR proteins such as SRSF1, SRSF5, and SRSF6 (121), suggesting that at least one of these SR proteins mediates DSS choice. Indeed, overexpression experiments revealed that SRSF1 and SRSF5 enhanced PSS usage and, thus, increased $VEGFA_{xxx}$ isoforms, whereas SRSF6 favored DSS choice (Fig. 7). Moreover, pull-down experiments identified an SRSF6 binding site downstream of exon 8b, but due to the positional effects of SR proteins (37) this binding site can hardly explain a direct SRSF6-mediated enhancing effect on DSS choice. A later study (109) focused especially on exploring the link between the IGF1- and SRSF1-dependent switch in splice site choice toward pro-angiogenic VEGFA transcript isoforms. An SRSF1 binding site within the PPT upstream of the PSS could be mapped; however, these experiments do not necessarily prove that SRSF1 binds and upregulates PSS usage *in vivo* (109). Furthermore, PSS utilization could be efficiently suppressed by inhibition of protein kinase C and the SR protein kinases 1 and 2 (SRPK1, SRPK2). Since the latter two have been shown to phosphorylate and thereby activate SRSF1, this strongly suggests that SRSF1 is involved in splice site choice; the precise mechanism, however, remains elusive. Since the Wilms' tumor suppressor 1 protein (WT1) represses SRPK1 expression by binding to the promoter of its gene, WT1 indirectly is also involved in inhibiting SRSF1 and, thus, preventing angiogenesis (2) (Fig. 7). In addition, SRSF2 was described to play a role in the regulation of the relative amount of pro-angiogenic versus anti-angiogenic VEGFA isoforms in human cancer cells (101). Here, the transcription factor E2F1 was shown to increase $VEGFA_{165b}$, which could be prevented by knockdown of its direct transcriptional target SRSF2 (Fig. 7). On the other hand, overexpression of SRSF2 resulted in a boost of $VEGFA_{165b}$ and moreover, E2F1 and SRSF2 were able to reduce blood vessel formation in tumors *in vivo*.

Endoglin

Endoglin (ENG) plays an important role in vascular development and angiogenesis. It is a transmembrane glycoprotein with a large extracellular domain, a hydrophobic transmembrane domain, and a relatively short intracellular domain (51). It forms a disulfide-linked homodimer and can be phosphorylated in its serine/threonine-rich cytoplasmic tail (77, 85, 123). Its involvement in vessel development and angiogenesis was demonstrated by the embryonic lethality of *eng*-knockout mice, which exhibit defects in vascular development although vasculogenesis is not affected (89). This was underpinned by the observation that haploinsufficiency impairs angiogenesis in adult animals (70). Furthermore, such animals show reduced tumor angiogenesis, demonstrating a major role for ENG also in this process (33). Moreover, ENG has also been implicated in vasodilation by stabilizing eNOS and facilitating its interaction with Hsp90 (69, 147).

ENG on its own does not bind a ligand, but it is an accessory coreceptor of the transforming growth factor receptor system (19, 56). The TGFβ superfamily consists of a large number of proteins, including TGFβ1, 2, and 3, bone-morphogenetic proteins and activins, which have a multitude of different functions. They signal through heteromeric receptors composed of two transmembrane serine/threonine kinases termed type I and type II receptor. The TGFβs bind to a type II receptor, which already in the basal state is phosphorylated. The liganded type II receptor then recruits a type I receptor, which only in the complex is capable of ligand binding. In this complex, the type II receptor phosphorylates and, therefore, activates its heteromeric partner, which after a conformational change is then responsible for further downstream signaling *via* different SMAD proteins that transduce the extracellular signals to the nucleus to activate transcription. In addition to the ligand binding receptors, two auxiliary coreceptors have been identified, namely betaglycan and ENG [for a review of the TGFβ system, see refs. (96, 97)]. Betaglycan coexists with TGFβ receptors in a variety of cell types; however, in endothelial cells that have no or very little betaglycan (127), ENG is the major coreceptor (19). ENG interacts with both the TGFβ receptor 1 (TGFBR1) and 2 (TGFBR2) (56) and other members of this receptor family (3, 11). It binds TGFβ1 and 3 with a high affinity by associating with the constitutively active TGFBR2, but unlike the related betaglycan it does not bind TGFβ2 (3, 19). The function of ENG as a TGFβ coreceptor explains the similarity between phenotypes of mice deficient for ENG (89), TGFβ1 (30), TGFBR1 (84), and TGFBR2 (112). These knockout animals also underscore the importance of the TGFβ pathway in vascular development and remodeling.

Interestingly, TGFβ1 has different effects on endothelial cells, which seemingly depend on the dose and the downstream pathways activated. While low doses act pro-angiogenic and stimulate proliferation and migration, high doses have opposite effects. The pro-angiogenic function has been ascribed to signaling *via* the activin A receptor type II like 1 (ACVRL1), also called ALK1; the anti-angiogenic function is coupled to TGFBR1, also known as ALK5. The consequence of the activation of these different receptors is an activation of different SMAD proteins (52)—SMAD 1 and 5 downstream of ACVRL1 and SMAD 2 and 3 as effectors of TGFBR1 (Fig. 8)—and an upregulation of distinct target

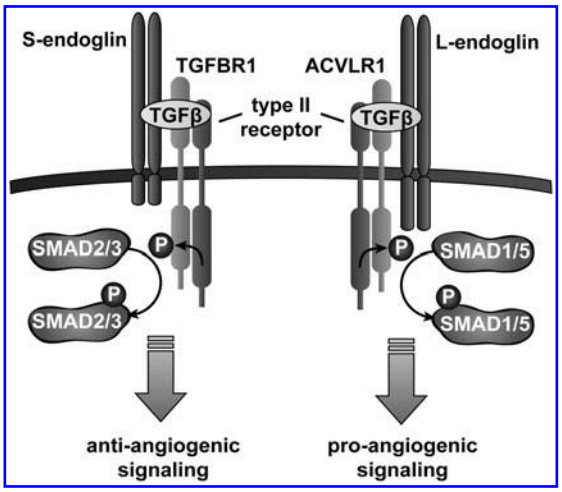


FIG. 8. Influence of endoglin isoforms on TGFβ1 signaling. S-endoglin promotes TGFβ1 signaling *via* TGFBR1, leading to phosphorylation of SMAD2 and 3, triggering anti-angiogenic gene expression programs. In contrast, L-endoglin supports signaling *via* ACVRL1, SMAD 1, and 5, thereby promoting angiogenesis. ACVRL1, activin A receptor type II like 1; TGFBR1, TGFβ receptor 1.

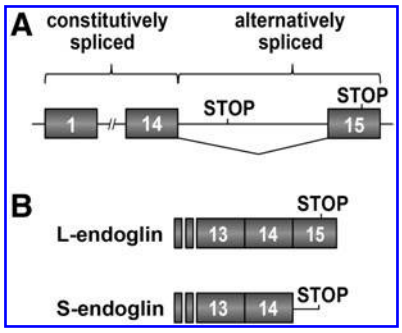


FIG. 9. Endoglin splice variants. (A) Alternative splicing of endoglin occurs at SD14. **(B)** Two alternatively spliced isoforms have been described for endoglin. Both isoforms share the first 13 exons. In the S-endoglin isoform, the SD of exon 14 is not recognized and, thus, the downstream intron between exons 14 and 15 is retained, leading to a premature stop codon and therefore a shorter protein product.

genes (113). Downregulation of endoglin (88) or the use of a neutralizing antibody (132) aggravated the TGFβ1-mediated inhibition of endothelial cell migration and proliferation, suggesting that endoglin is a negative regulator of the TGFBR1 pathway. This is in line with findings that endoglin promotes proliferation of normal endothelial cells (86), and elevated ENG levels correlate with proliferation of tumor endothelial cells (103). Taken together, these results suggest that ENG affects the outcome of TGFβ signals received by endothelial cells. Indeed, downregulation of endoglin suppresses pro-angiogenic ACVRL1 signal transduction. Interestingly, knockdown of TGFBR1 could rescue the proliferation defect observed after interference with endoglin expression, indicating that the ACVRL1 pathway indirectly suppresses TGFRB1 signaling (86).

Similar to VEGFA and eNOS mRNA processing, the endoglin pre-mRNA is also alternatively spliced, leading to two-protein coding and at least three additional non-protein coding transcript variants. The more abundant form of the protein coding transcripts, L (long)-endoglin consists of 15 exons and the less frequently expressed S (short)-endoglin of 14 exons (6). Both isoforms share the first 13 exons, which contain the extracellular and transmembrane domains. However, the last intron between exons 14 and 15 is retained in the shorter S-ENG isoform, leading to a premature stop codon within the thus created open reading frame extending into the intron (Fig. 9). As a consequence, the cytoplasmic tails of S- and L-endoglin differ in their length. While in L-ENG it is 47-amino-acids long, the corresponding region in S-ENG consists of 14 amino acids, the last 7 of which differ from the long form.

Studies focusing on the differential expression of those two isoforms revealed that the S-/L-endoglin ratio increases in senescent human endothelial cells and in vascularized tissues of old mice, both situations in which oxidative stress is increased (36, 53); furthermore, angiogenesis is impaired in aging (82). It was suggested that L- and S-endoglin differentially influence the two diverging TGFβ receptor pathways in endothelial cells. Although S-ENG can interact with ACVRL1 as well as with TGFBR1, the affinity to the latter is higher (10), whereas L-ENG shows a higher affinity to ACVRL1 (11) (Fig. 8). At the downstream end of the signaling cascades, that is, at the transcriptional level, it was shown that the shift toward S-endoglin in senescent cells is also reflected in the regulation of known target genes of the two TGFβ pathways. Moreover, reporter gene assays specifically measuring the output of the divergent signaling pathways showed that S-ENG can activate a TGFBR1-dependent reporter construct, which was repressed by L-ENG. Likewise, a reporter gene regulated by ACVRL1-dependent signals was not affected by S-endoglin, but strongly activated by the longer isoform (10). Another study assessing the isoform-specific functions after stable transfection of S- or L-ENG expression vectors into a cell line not expressing endoglin (87) demonstrated that the longer isoform promotes proliferation, which was suppressed by the shorter protein. In line with the work described earlier, L-ENG enhanced the ACVRL1 pathway, while S-ENG promoted TGFBR1 signaling (152) (Fig. 8). Notably, transgenic mice expressing S-endoglin in the endothelium (151) suffer from hypertension, decreased hypertensive responses to NO inhibition, a reduced vasodilatory response to TGFβ, and a decreased eNOS expression in some tissues, pointing toward a connection between endoglin and NO production (10).

With regard to regulation of alternative endoglin splicing, it could be demonstrated that SRSF1 plays a key role in intron retention of S-ENG (8). Overexpression of SRSF1 tilted the balance between the two isoforms toward S-ENG and *via* a computer-based analysis, two potential SRSF1 binding sites within the retained intron were predicted. Mutational analysis confirmed the critical role of one of these motifs for intron retention. Interestingly, the upregulation of S-ENG after replicative or stress-induced endothelial senescence was accompanied by a relocation of SRSF1 into the cytoplasm. Therefore, it was assumed that during senescence, SRSF1 interferes with binding of the minor spliceosome, which was suggested to be also active in the cytoplasm (78), and, thus, inhibits splicing at exons 14 and 15. However, whether the minor spliceosome is, indeed, located in the cytoplasm is still a subject of intense debate (149). Based on the involvement

of SRSF1 in the senescence-associated switch in endoglin isoforms and alternative splicing of other mediators of endothelial cell functions such as VEGFA and tissue factor, it was postulated that SRSF1 could be a marker of endothelial senescence (9).

Grainyhead-Like 3

The transcription factor grainyhead-like 3 (GRHL3) is a member of an ancient transcription factor family conserved throughout the animal kingdom (148, 154). The first member of this family to be described was *Drosophila* grainyhead (GRH). Its name is derived from flies carrying mutations in the *grainyhead* gene, leading to embryonic lethality and a cuticle phenotype, including a grainy and discontinuous head skeleton (12). The genomes of invertebrates such as *Drosophila melanogaster* and *Caenorhabditis elegans* contain a single *grh* gene (154), whereas three homologs, called *grhl1*, *grhl2*, and *grhl3* (81, 146, 163), exist in vertebrates. On the protein level all GRH-homologs are highly conserved in their activation-, DNA-binding, and dimerization domains and can form homo- and heterodimers with each other (146). GRHL3 was originally described as being essential for neural tube closure during development, skin barrier function, and wound repair (144, 145), but it also regulates urothelial differentiation (167). Interestingly, the function in barrier formation is not only evident in mammalians, but also in *D. melanogaster* (94) and *C. elegans* (154), although the cuticles in these animals are structured completely different, suggesting that this function arose very early in evolution (106).

A totally different role for *grhl3* was suggested by its identification in a gene-trap screen for anti-apoptotic genes (55). Later on, it was shown that in human endothelial cells, GRHL3 is required for basal and NO-induced migration. The pro-migratory effect of GRHL3 in these cells after overexpression was as strong as treatment with VEGFA, but seemed to be independent of it (55, 93). Furthermore, it was demonstrated that GRHL3 also serves an anti-apoptotic function in endothelial cells, which is in line with its identification in the original screen. This effect is dependent on NO production, as it could be abrogated by inhibition of eNOS. A mechanistic explanation was provided by the fact that overexpression of GRHL3 led to an activation of eNOS, which plays a crucial role in protection of endothelial cells, and its upstream regulator AKT1. Interestingly, treatment of endothelial cells with an NO donor upregulated the expression of GRHL3, indicating the existence of a regulatory feedback loop (93).

However, unlike the situation in mice, for which only a single GRHL3 transcript and the corresponding protein have been described, three isoforms of this transcription factor are derived from the human gene (146). This is due to an additional first, human-specific exon, transcribed from an alternative promoter, which gives rise to a second primary transcript. In addition, this transcript is alternatively spliced—including or skipping the second exon (Fig. 10). While the transcript including exon 2 codes for a protein

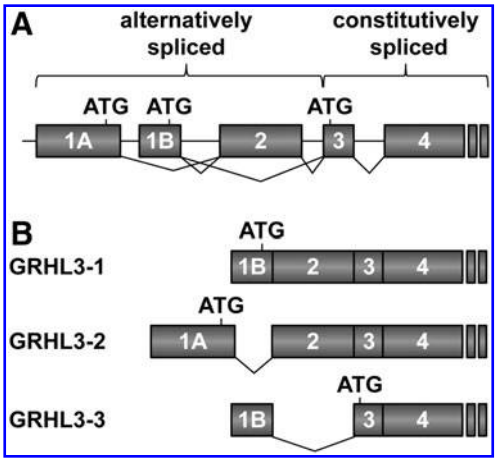


FIG. 10. GRHL3 splice variants. (A) The human *grhl3* gene contains two different first exons (1A and 1B) transcribed from separate promoters. In addition, the pre-mRNA beginning with exon 1B can be alternatively spliced, including or skipping exon 2. **(B)** The GRHL3 isoforms translated from the three mRNAs have distinct N-termini. While isoform1 (GRHL3-1) and 2 (GRHL3-2) differ only by a few amino acids, isoform 3 (GRHL3-3) represents an N-terminally truncated version of the other two proteins. (ATG—functional translation initiation codons). GRHL3, grainyhead-like 3.

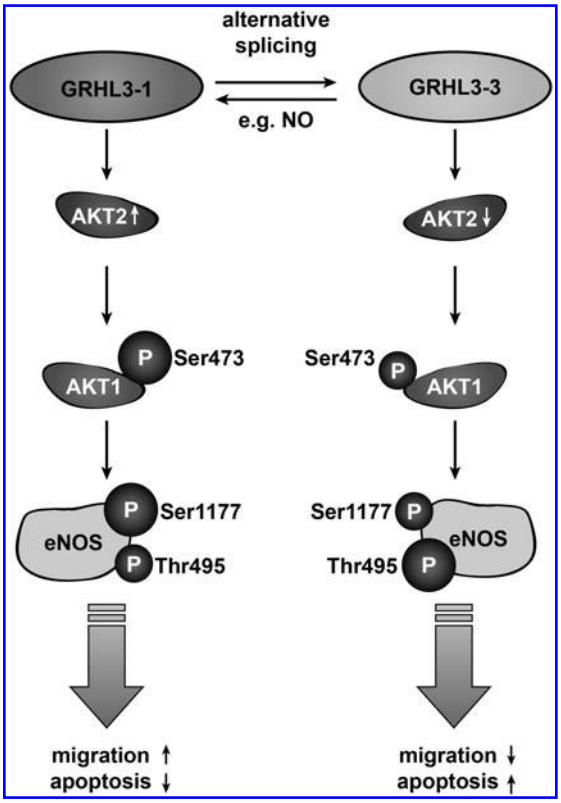


FIG. 11. Impact of GRHL3 splice variants on endothelial functions. GRHL3 isoforms 1 (GRHL3-1) and 3 (GRH3-3), which are derived from alternative splice variants of the same pre-mRNA, reciprocally regulate AKT2 expression. Since AKT2 is a master regulator of all AKT-isoforms, this could explain the opposite effects on the phosphorylation AKT1 and its target eNOS, finally leading to apoptosis inhibition and migration induction in the case of GRHL3-1 and the opposite in the case of GHRHL3-3. The balance between these GRHL3 isoforms can be regulated, for example, by NO.

(GRHL3-1), which is nearly identical to the mouse protein and human GRHL3-2, the open reading frame of the shorter variant GRHL3-3 starts with the translation initiation codon in exon 4, that is translated into an N-terminally truncated, but otherwise identical protein, which was previously described as a putative repressor (146). We could show that all three mRNAs are coexpressed in human endothelial cells and that physiological concentrations of NO induce a shift in the balance between the two alternative splice forms toward GRHL3-1. Intriguingly, the translation products of the two alternatively spliced mRNAs have opposing functions in these cells (58) (Fig. 11). The longer isoform 1 acts pro-migratory, whereas isoform 3 inhibits migration. A similar dichotomy was observed with regard to endothelial cell apoptosis. While GRHL3-1 suppresses apoptosis in an eNOS-dependent manner, GRHL3-3 acts pro-apoptotic. This was reflected in the reciprocal regulation of eNOS phosphorylation by these two proteins. In contrast to the report by Ting *et al.* (146), we demonstrated that not only isoform 1 is a transcriptional activator but also isoform 3 and that the two isoforms activate different sets of target genes. Validation of selected targets on the protein level showed that GRHL3-1 upregulates the expression of two proteins, which could serve to explain the anti-apoptotic and pro-migratory effects, the basic helix-loop-helix protein Max interactor 1 [MXI1], and protein kinase B β [AKT2] (58). While MXI1 could block the action of MYC, which has been implicated in endothelial cell apoptosis (91), AKT2 is a master regulator of all AKT isoforms (119) and might, therefore, be responsible for AKT1 and finally eNOS activation. The expression of both proteins was downregulated by GRHL3-3 (Fig. 11). The cumulative evidence suggests that fine-tuning the balance between the two GRHL3 isoforms determines the fate of endothelial cells. However, so far, nothing is known about the mechanisms regulating alternative splicing of GRHL3.

Conclusion

Endothelial dysfunction is not only the cause for many cardiovascular diseases, but also associated with aging and senescence. However, the simple analysis of gene expression changes in endothelial cells subject to physiological and pathophysiological conditions captures only a part of the picture, because the diversity of the human proteome cannot be explained by the number of protein coding genes in our genome. Alternative splicing is one major mechanism generating multiple protein isoforms from a single pre-mRNA. The examples described earlier illustrate that alternative splicing is not merely a whim of nature, but can have profound implications for endothelial function and, thus, vascular physiology. Interestingly, protein isoforms generated by alternative splicing can have different and often opposing functions, suggesting that this process might serve to finely tune cellular responses and outputs. Thus, for gene expression analyses, it seems imperative to comprehensively identify all splice variants as a prerequisite to uncover the functions of the corresponding proteins. The first attempts focusing on this aspect of in-depth transcriptome analyses have recently been made (35). This also requires a unanimous annotation and nomenclature of splice variants. A commendable example—at least for tracking the numerous spliceosome-associated proteins and snRNAs—is given by the Spliceosome Database (27). Moreover, mechanistic explanations of how alternative splicing of critical mediators of endothelial cell function is regulated, including the roles of specific splicing regulatory proteins involved in particular splice events, are needed. However, these underlying processes are only understood in very few cases and mostly not in full detail. Work in this direction in the cancer field has led to ideas about new therapeutic options (29) by interfering with specific protein isoforms or the splice process *per se*, and it is, thus, imaginable that similar options might be available in the future to treat endothelial dysfunction and, therefore, improve cardiovascular health.

Acknowledgment

The authors thank Nadine Dyballa-Rukes for critically reading this article.

References

1. Alp NJ and Channon KM. Regulation of endothelial nitric oxide synthase by tetrahydrobiopterin in vascular disease. *Arterioscler Thromb Vasc Biol* 24: 413–420, 2004.
2. Amin EM, Oltean S, Hua J, Gammons MV, Hamdollah-Zadeh M, Welsh GI, Cheung MK, Ni L, Kase S, Rennel ES, Symonds KE, Nowak DG, Royer-Pokora B, Saleem MA, Hagiwara M, Schumacher VA, Harper SJ, Hinton DR, Bates DO, and Ladomery MR. WT1 mutants reveal SRPK1 to be a downstream angiogenesis target by altering VEGF splicing. *Cancer Cell* 20: 768–780, 2011.
3. Barbara NP, Wrana JL, and Letarte M. Endoglin is an accessory protein that interacts with the signaling receptor complex of multiple members of the transforming growth factor-beta superfamily. *J Biol Chem* 274: 584–594, 1999.
4. Bates DO, Cui TG, Doughty JM, Winkler M, Sugiono M, Shields JD, Peat D, Gillatt D, and Harper SJ. VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, is down-regulated in renal cell carcinoma. *Cancer Res* 62: 4123–4131, 2002.
5. Bauer PM, Fulton D, Boo YC, Sorescu GP, Kemp BE, Jo H, and Sessa WC. Compensatory phosphorylation and protein-protein interactions revealed by loss of function and gain of function mutants of multiple serine phosphorylation sites in endothelial nitric-oxide synthase. *J Biol Chem* 278: 14841–14849, 2003.
6. Bellon T, Corbi A, Lastres P, Cales C, Cebrian M, Vera S, Cheifetz S, Massague J, Letarte M, and Bernabeu C. Identification and expression of two forms of the human transforming growth factor-beta-binding protein endoglin with distinct cytoplasmic regions. *Eur J Immunol* 23: 2340–2345, 1993.
7. Bender AT, Silverstein AM, Demady DR, Kanelakis KC, Noguchi S, Pratt WB, and Osawa Y. Neuronal nitric-oxide synthase is regulated by the Hsp90-based chaperone system *in vivo*. *J Biol Chem* 274: 1472–1478, 1999.
8. Blanco FJ and Bernabeu C. Alternative splicing factor or splicing factor-2 plays a key role in intron retention of the endoglin gene during endothelial senescence. *Aging Cell* 10: 896–907, 2011.
9. Blanco FJ and Bernabeu C. The splicing factor SRSF1 as a marker for endothelial senescence. *Front Physiol* 3: 54, 2012.
10. Blanco FJ, Grande MT, Langa C, Oujo B, Velasco S, Rodriguez-Barbero A, Perez-Gomez E, Quintanilla M, Lopez-Novoa JM, and Bernabeu C. S-endoglin expression

is induced in senescent endothelial cells and contributes to vascular pathology. *Circ Res* 103: 1383–1392, 2008.
11. Blanco FJ, Santibanez JF, Guerrero-Esteo M, Langa C, Vary CP, and Bernabeu C. Interaction and functional interplay between endoglin and ALK-1, two components of the endothelial transforming growth factor-beta receptor complex. *J Cell Physiol* 204: 574–584, 2005.
12. Bray SJ and Kafatos FC. Developmental function of Elf-1: an essential transcription factor during embryogenesis in *Drosophila*. *Genes Dev* 5: 1672–1683, 1991.
13. Broxmeyer HE, Cooper S, Li ZH, Lu L, Song HY, Kwon BS, Warren RE, and Donner DB. Myeloid progenitor cell regulatory effects of vascular endothelial cell growth factor. *Int J Hematol* 62: 203–215, 1995.
14. Burset M, Seledtsov IA, and Solovyev VV. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res* 28: 4364–4375, 2000.
15. Busse R and Mulsch A. Calcium-dependent nitric oxide synthesis in endothelial cytosol is mediated by calmodulin. *FEBS Lett* 265: 133–136, 1990.
16. Bussolati B, Dunk C, Grohman M, Kontos CD, Mason J, and Ahmed A. Vascular endothelial growth factor receptor-1 modulates vascular endothelial growth factor-mediated angiogenesis via nitric oxide. *Am J Pathol* 159: 993–1008, 2001.
17. Caulin-Glaser T, Garcia-Cardena G, Sarrel P, Sessa WC, and Bender JR. 17 beta-estradiol regulation of human endothelial cell basal nitric oxide release, independent of cytosolic Ca2+ mobilization. *Circ Res* 81: 885–892, 1997.
18. Chanarat S and Strasser K. Splicing and beyond: the many faces of the Prp19 complex. *Biochim Biophys Acta* 1833: 2126–2134, 2013.
19. Cheifetz S, Bellon T, Cales C, Vera S, Bernabeu C, Massague J, and Letarte M. Endoglin is a component of the transforming growth factor-beta receptor system in human endothelial cells. *J Biol Chem* 267: 19027–19030, 1992.
20. Chiarugi P. PTPs versus PTKs: the redox side of the coin. *Free Radic Res* 39: 353–364, 2005.
21. Clauss M, Gerlach M, Gerlach H, Brett J, Wang F, Familletti PC, Pan YC, Olander JV, Connolly DT, and Stern D. Vascular permeability factor: a tumor-derived polypeptide that induces endothelial cell and monocyte procoagulant activity, and promotes monocyte migration. *J Exp Med* 172: 1535–1545, 1990.
22. Cooper TA, Wan L, and Dreyfuss G. RNA and disease. *Cell* 136: 777–793, 2009.
23. Cote GJ, Zhu W, Thomas A, Martin E, Murad F, and Sharina IG. Hydrogen peroxide alters splicing of soluble guanylyl cyclase and selectively modulates expression of splicing regulators in human cancer cells. *PLoS One* 7: e41099, 2012.
24. Coulter LR, Landree MA, and Cooper TA. Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol Cell Biol* 17: 2143–2150, 1997.
25. Crane BR, Arvai AS, Ghosh DK, Wu C, Getzoff ED, Stuehr DJ, and Tainer JA. Structure of nitric oxide synthase oxygenase dimer with pterin and substrate. *Science* 279: 2121–2126, 1998.
26. Cudmore MJ, Hewett PW, Ahmad S, Wang KQ, Cai M, Al-Ani B, Fujisawa T, Ma B, Sissaoui S, Ramma W, Miller MR, Newby DE, Gu Y, Barleon B, Weich H, and Ahmed A. The role of heterodimerization between VEGFR-1 and VEGFR-2 in the regulation of endothelial cell homeostasis. *Nat Commun* 3: 972, 2012.
27. Cvitkovic I and Jurica MS. Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic Acids Res* 41: D132–D141, 2013.
28. Dedio J, Konig P, Wohlfart P, Schroeder C, Kummer W, and Muller-Esterl W. NOSIP, a novel modulator of endothelial nitric oxide synthase activity. *FASEB J* 15: 79–89, 2001.
29. Dehm SM. mRNA splicing variants: exploiting modularity to outwit cancer therapy. *Cancer Res* 73: 5309–5314, 2013.
30. Dickson MC, Martin JS, Cousins FM, Kulkarni AB, Karlsson S, and Akhurst RJ. Defective haematopoiesis and vasculogenesis in transforming growth factor-beta 1 knock out mice. *Development* 121: 1845–1854, 1995.
31. Distler O, Neidhart M, Gay RE, and Gay S. The molecular control of angiogenesis. *Int Rev Immunol* 21: 33–49, 2002.
32. Dumitrescu C, Biondi R, Xia Y, Cardounel AJ, Druhan LJ, Ambrosio G, and Zweier JL. Myocardial ischemia results in tetrahydrobiopterin (BH4) oxidation with impaired endothelial function ameliorated by BH4. *Proc Natl Acad Sci U S A* 104: 15081–15086, 2007.
33. Duwel A, Eleno N, Jerkic M, Arevalo M, Bolanos JP, Bernabeu C, and Lopez-Novoa JM. Reduced tumor growth and angiogenesis in endoglin-haploinsufficient mice. *Tumour Biol* 28: 1–8, 2007.
34. Eisenreich A, Boltzen U, Poller W, Schultheiss HP, and Rauch U. Effects of the Cdc2-like kinase-family and DNA topoisomerase I on the alternative splicing of eNOS in TNF-alpha-stimulated human endothelial cells. *Biol Chem* 389: 1333–1338, 2008.
35. Eksi R, Li HD, Menon R, Wen Y, Omenn GS, Kretzler M, and Guan Y. Systematically differentiating functions for alternatively spliced isoforms through integrating RNA-seq data. *PLoS Comput Biol* 9: e1003314, 2013.
36. El Assar M, Angulo J, and Rodriguez-Manas L. Oxidative stress and vascular inflammation in aging. *Free Radic Biol Med* 65: 380–401, 2013.
37. Erkelenz S, Mueller WF, Evans MS, Busch A, Schoneweis K, Hertel KJ, and Schaal H. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA* 19: 96–102, 2013.
38. Fearon IM and Faux SP. Oxidative stress and cardiovascular disease: novel tools give (free) radical insight. *J Mol Cell Cardiol* 47: 372–381, 2009.
39. Finkel T and Holbrook NJ. Oxidants, oxidative stress and the biology of ageing. *Nature* 408: 239–247, 2000.
40. Fisslthaler B, Loot AE, Mohamed A, Busse R, and Fleming I. Inhibition of endothelial nitric oxide synthase activity by proline-rich tyrosine kinase 2 in response to fluid shear stress and insulin. *Circ Res* 102: 1520–1528, 2008.
41. Fleming I and Busse R. Molecular mechanisms involved in the regulation of the endothelial nitric oxide synthase. *Am J Physiol Regul Integr Comp Physiol* 284: R1–R12, 2003.
42. Fleming I, Fisslthaler B, Dimmeler S, Kemp BE, and Busse R. Phosphorylation of Thr(495) regulates Ca(2+)/calmodulin-dependent endothelial nitric oxide synthase activity. *Circ Res* 88: E68–E75, 2001.
43. Freund M, Hicks MJ, Konermann C, Otte M, Hertel KJ, and Schaal H. Extended base pair complementarity between U1 snRNA and the 5′ splice site does not inhibit splicing in higher eukaryotes, but rather increases 5′ splice site recognition. *Nucleic Acids Res* 33: 5112–5119, 2005.

44. Fulton D, Church JE, Ruan L, Li C, Sood SG, Kemp BE, Jennings IG, and Venema RC. Src kinase activates endothelial nitric-oxide synthase by phosphorylating Tyr-83. *J Biol Chem* 280: 35943–35952, 2005.
45. Fulton D, Gratton JP, and Sessa WC. Post-translational control of endothelial nitric oxide synthase: why isn't calcium/calmodulin enough? *J Pharmacol Exp Ther* 299: 818–824, 2001.
46. Fulton D, Ruan L, Sood SG, Li C, Zhang Q, and Venema RC. Agonist-stimulated endothelial nitric oxide synthase activation and vascular relaxation. Role of eNOS phosphorylation at Tyr83. *Circ Res* 102: 497–504, 2008.
47. Furchgott RF and Zawadzki JV. The obligatory role of endothelial cells in the relaxation of arterial smooth muscle by acetylcholine. *Nature* 288: 373–376, 1980.
48. Garcia-Cardena G, Fan R, Shah V, Sorrentino R, Cirino G, Papapetropoulos A, and Sessa WC. Dynamic activation of endothelial nitric oxide synthase by Hsp90. *Nature* 392: 821–824, 1998.
49. Gerber HP, Dixit V, and Ferrara N. Vascular endothelial growth factor induces expression of the antiapoptotic proteins Bcl-2 and A1 in vascular endothelial cells. *J Biol Chem* 273: 13313–13316, 1998.
50. Gerber HP, McMurtrey A, Kowalski J, Yan M, Keyt BA, Dixit V, and Ferrara N. Vascular endothelial growth factor regulates endothelial cell survival through the phosphatidylinositol 3′-kinase/Akt signal transduction pathway. Requirement for Flk-1/KDR activation. *J Biol Chem* 273: 30336–30343, 1998.
51. Gougos A and Letarte M. Primary structure of endoglin, an RGD-containing glycoprotein of human endothelial cells. *J Biol Chem* 265: 8361–8364, 1990.
52. Goumans MJ, Valdimarsdottir G, Itoh S, Rosendahl A, Sideras P, and ten Dijke P. Balancing the activation state of the endothelium via two distinct TGF-beta type I receptors. *EMBO J* 21: 1743–1753, 2002.
53. Goy C, Czypiorski P, Altschmied J, Jakob S, Rabanter LL, Brewer AC, Ale-Agha N, Dyballa-Rukes N, Shah AM, and Haendeler J. The imbalanced redox status in senescent endothelial cells is due to dysregulated Thioredoxin-1 and NADPH oxidase 4. *Exp Gerontol* 56: 45–52, 2014.
54. Gross PL and Aird WC. The endothelium and thrombosis. *Semin Thromb Hemost* 26: 463–478, 2000.
55. Guardiola-Serrano F, Haendeler J, Lukosz M, Sturm K, Melchner H, and Altschmied J. Gene trapping identifies a putative tumor suppressor and a new inducer of cell migration. *Biochem Biophys Res Commun* 376: 748–752, 2008.
56. Guerrero-Esteo M, Sanchez-Elsner T, Letamendia A, and Bernabeu C. Extracellular and cytoplasmic domains of endoglin interact with the transforming growth factor-beta receptors I and II. *J Biol Chem* 277: 29197–29209, 2002.
57. Guo D, Jia Q, Song HY, Warren RS, and Donner DB. Vascular endothelial cell growth factor promotes tyrosine phosphorylation of mediators of signal transduction that contain SH2 domains. Association with endothelial cell proliferation. *J Biol Chem* 270: 6729–6733, 1995.
58. Haendeler J, Mlynek A, Buchner N, Lukosz M, Graf M, Guettler C, Jakob S, Farrokh S, Kunze K, Goy C, Guardiola-Serrano F, Schaal H, Cortese-Krott M, Deenen R, Kohrer K, Winkler C, and Altschmied J. Two isoforms of Sister-Of-Mammalian Grainyhead have opposing functions in endothelial cells and *in vivo*. *Arterioscler Thromb Vasc Biol* 33: 1639–1646, 2013.
59. Hang X, Li P, Li Z, Qu W, Yu Y, Li H, Shen Z, Zheng H, Gao Y, Wu Y, Deng M, Sun Z, and Zhang C. Transcription and splicing regulation in human umbilical vein endothelial cells under hypoxic stress conditions by exon array. *BMC Genomics* 10: 126, 2009.
60. Harper SJ and Bates DO. VEGF-A splicing: the key to anti-angiogenic therapeutics? *Nat Rev Cancer* 8: 880–887, 2008.
61. Harries LW, Hernandez D, Henley W, Wood AR, Holly AC, Bradley-Smith RM, Yaghootkar H, Dutta A, Murray A, Frayling TM, Guralnik JM, Bandinelli S, Singleton A, Ferrucci L, and Melzer D. Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell* 10: 868–878, 2011.
62. Harrison DG. Endothelial function and oxidant stress. *Clin Cardiol* 20: II-11–17, 1997.
63. Hartmann L, Theiss S, Niederacher D, and Schaal H. Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci* 13: 3252–3272, 2008.
64. Holmes DI and Zachary I. The vascular endothelial growth factor (VEGF) family: angiogenic factors in health and disease. *Genome Biol* 6: 209, 2005.
65. Huang G, Zhou Z, Wang H, and Kleinerman ES. CAPER-alpha alternative splicing regulates the expression of vascular endothelial growth factor(1)(6)(5) in Ewing sarcoma cells. *Cancer* 118: 2106–2116, 2012.
66. Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S, Ares M, Jr., and Yeo GW. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* 1: 167–178, 2012.
67. Hui J, Hung LH, Heiner M, Schreiner S, Neumuller N, Reither G, Haas SA, and Bindereif A. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J* 24: 1988–1998, 2005.
68. Hui J, Stangl K, Lane WS, and Bindereif A. HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat Struct Biol* 10: 33–37, 2003.
69. Jerkic M, Rivas-Elena JV, Prieto M, Carron R, Sanz-Rodriguez F, Perez-Barriocanal F, Rodriguez-Barbero A, Bernabeu C, and Lopez-Novoa JM. Endoglin regulates nitric oxide-dependent vasodilatation. *FASEB J* 18: 609–611, 2004.
70. Jerkic M, Rodriguez-Barbero A, Prieto M, Toporsian M, Pericacho M, Rivas-Elena JV, Obreo J, Wang A, Perez-Barriocanal F, Arevalo M, Bernabeu C, Letarte M, and Lopez-Novoa JM. Reduced angiogenic responses in adult Endoglin heterozygous mice. *Cardiovasc Res* 69: 845–854, 2006.
71. Ju H, Zou R, Venema VJ, and Venema RC. Direct interaction of endothelial nitric-oxide synthase and caveolin-1 inhibits synthase activity. *J Biol Chem* 272: 18522–18525, 1997.
72. Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, and Dreyfuss G. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468: 664–668, 2010.
73. Kammler S, Leurs C, Freund M, Krummheuer J, Seidel K, Tange TO, Lund MK, Kjems J, Scheid A, and Schaal H. The sequence complementarity between HIV-1 5′ splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. *RNA* 7: 421–434, 2001.

74. Kattapuram T, Yang S, Maki JL, and Stone JR. Protein kinase CK1alpha regulates mRNA binding by heterogeneous nuclear ribonucleoprotein C in response to physiologic levels of hydrogen peroxide. *J Biol Chem* 280: 15340–15347, 2005.
75. Kawamura H, Li X, Harper SJ, Bates DO, and Claesson-Welsh L. Vascular endothelial growth factor (VEGF)-A165b is a weak *in vitro* agonist for VEGF receptor-2 due to lack of coreceptor binding and deficient regulation of kinase activity. *Cancer Res* 68: 4683–4692, 2008.
76. Klagsbrun M and D'Amore PA. Vascular endothelial growth factor and its receptors. *Cytokine Growth Factor Rev* 7: 259–270, 1996.
77. Koleva RI, Conley BA, Romero D, Riley KS, Marto JA, Lux A, and Vary CP. Endoglin structure and function: determinants of endoglin phosphorylation by transforming growth factor-beta receptors. *J Biol Chem* 281: 25110–25123, 2006.
78. König H, Matter N, Bader R, Thiele W, and Müller F. Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. *Cell* 131: 718–729, 2007.
79. Kou R, Greif D, and Michel T. Dephosphorylation of endothelial nitric-oxide synthase by vascular endothelial growth factor. Implications for the vascular responses to cyclosporin A. *J Biol Chem* 277: 29669–29673, 2002.
80. Kuchan MJ and Frangos JA. Role of calcium and calmodulin in flow-induced nitric oxide production in endothelial cells. *Am J Physiol* 266: C628–C636, 1994.
81. Kudryavtseva EI, Sugihara TM, Wang N, Lasso RJ, Gudnason JF, Lipkin SM, and Andersen B. Identification and characterization of Grainyhead-like epithelial transactivator (GET-1), a novel mammalian Grainyhead-like factor. *Dev Dyn* 226: 604–617, 2003.
82. Lähteenvuo J and Rosenzweig A. Effects of aging on angiogenesis. *Circ Res* 110: 1252–1264, 2012.
83. Lamalice L, Houle F, Jourdan G, and Huot J. Phosphorylation of tyrosine 1214 on VEGFR2 is required for VEGF-induced activation of Cdc42 upstream of SAPK2/p38. *Oncogene* 23: 434–445, 2004.
84. Larsson J, Goumans MJ, Sjostrand LJ, van Rooijen MA, Ward D, Leveen P, Xu X, ten Dijke P, Mummery CL, and Karlsson S. Abnormal angiogenesis but intact hematopoietic potential in TGF-beta type I receptor-deficient mice. *EMBO J* 20: 1663–1673, 2001.
85. Lastres P, Martin-Perez J, Langa C, and Bernabeu C. Phosphorylation of the human-transforming-growth-factor-beta-binding protein endoglin. *Biochem J* 301 (Pt 3): 765–768, 1994.
86. Lebrin F, Goumans MJ, Jonker L, Carvalho RL, Valdimarsdottir G, Thorikay M, Mummery C, Arthur HM, and ten Dijke P. Endoglin promotes endothelial cell proliferation and TGF-beta/ALK1 signal transduction. *EMBO J* 23: 4018–4028, 2004.
87. Letamendia A, Lastres P, Botella LM, Raab U, Langa C, Velasco B, Attisano L, and Bernabeu C. Role of endoglin in cellular responses to transforming growth factor-beta. A comparative study with betaglycan. *J Biol Chem* 273: 33011–33019, 1998.
88. Li C, Hampson IN, Hampson L, Kumar P, Bernabeu C, and Kumar S. CD105 antagonizes the inhibitory signaling of transforming growth factor beta1 on human vascular endothelial cells. *FASEB J* 14: 55–64, 2000.
89. Li DY, Sorensen LK, Brooke BS, Urness LD, Davis EC, Taylor DG, Boak BB, and Wendel DP. Defective angiogenesis in mice lacking endoglin. *Science* 284: 1534–1537, 1999.
90. Llorian M, Schwartz S, Clark TA, Hollander D, Tan LY, Spellman R, Gordon A, Schweitzer AC, de la Grange P, Ast G, and Smith CW. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat Struct Mol Biol* 17: 1114–1123, 2010.
91. Lopez-Farre A, Sanchez de Miguel L, Caramelo C, Gomez-Macias J, Garcia R, Mosquera JR, de Frutos T, Millas I, Rivas F, Echezarreta G, and Casado S. Role of nitric oxide in autocrine control of growth and apoptosis of endothelial cells. *Am J Physiol* 272: H760–H768, 1997.
92. Lorenz M, Hewing B, Hui J, Zepp A, Baumann G, Bindereif A, Stangl V, and Stangl K. Alternative splicing in intron 13 of the human eNOS gene: a potential mechanism for regulating eNOS activity. *FASEB J* 21: 1556–1564, 2007.
93. Lukosz M, Mlynek A, Czypiorski P, Altschmied J, and Haendeler J. The transcription factor Grainyhead like 3 (GRHL3) affects endothelial cell apoptosis and migration in a NO-dependent manner. *Biochem Biophys Res Commun* 412: 648–653, 2011.
94. Mace KA, Pearson JC, and McGinnis W. An epidermal barrier wound repair pathway in *Drosophila* is mediated by grainy head. *Science* 308: 381–385, 2005.
95. Mackman N. New insights into the mechanisms of venous thrombosis. *J Clin Invest* 122: 2331–2336, 2012.
96. Massague J. TGF-beta signal transduction. *Annu Rev Biochem* 67: 753–791, 1998.
97. Massague J. How cells read TGF-beta signals. *Nat Rev Mol Cell Biol* 1: 169–178, 2000.
98. Mazin P, Xiong J, Liu X, Yan Z, Zhang X, Li M, He L, Somel M, Yuan Y, Phoebe Chen YP, Li N, Hu Y, Fu N, Ning Z, Zeng R, Yang H, Chen W, Gelfand M, and Khaitovich P. Widespread splicing changes in human brain development and aging. *Mol Syst Biol* 9: 633, 2013.
99. McCabe TJ, Fulton D, Roman LJ, and Sessa WC. Enhanced electron flux and reduced calmodulin dissociation may explain ''calcium-independent'' eNOS activation by phosphorylation. *J Biol Chem* 275: 6123–6128, 2000.
100. Meng TC, Fukada T, and Tonks NK. Reversible oxidation and inactivation of protein tyrosine phosphatases *in vivo*. *Mol Cell* 9: 387–399, 2002.
101. Merdzhanova G, Gout S, Keramidas M, Edmond V, Coll JL, Brambilla C, Brambilla E, Gazzeri S, and Eymin B. The transcription factor E2F1 and the SR protein SC35 control the ratio of pro-angiogenic versus antiangiogenic isoforms of vascular endothelial growth factor-A to inhibit neovascularization *in vivo*. *Oncogene* 29: 5392–5403, 2010.
102. Meshorer E and Soreq H. Pre-mRNA splicing modulations in senescence. *Aging Cell* 1: 10–16, 2002.
103. Miller DW, Graulich W, Karges B, Stahl S, Ernst M, Ramaswamy A, Sedlacek HH, Muller R, and Adamkiewicz J. Elevated expression of endoglin, a component of the TGF-beta-receptor complex, correlates with proliferation of tumor endothelial cells. *Int J Cancer* 81: 568–572, 1999.
104. Morley D, Maragos CM, Zhang XY, Boignon M, Wink DA, and Keefer LK. Mechanism of vascular relaxation induced by the nitric oxide (NO)/nucleophile complexes, a new class of NO-based vasodilators. *J Cardiovasc Pharmacol* 21: 670–676, 1993.

105. Mount PF, Kemp BE, and Power DA. Regulation of endothelial and myocardial NO synthesis by multi-site eNOS phosphorylation. *J Mol Cell Cardiol* 42: 271–279, 2007.
106. Moussian B and Uv AE. An ancient control of epithelial barrier formation and wound healing. *Bioessays* 27: 987–990, 2005.
107. Nadaud S, Bonnardeaux A, Lathrop M, and Soubrier F. Gene structure, polymorphism and mapping of the human endothelial nitric oxide synthase gene. *Biochem Biophys Res Commun* 198: 1027–1033, 1994.
108. Neufeld G, Cohen T, Gengrinovitch S, and Poltorak Z. Vascular endothelial growth factor (VEGF) and its receptors. *FASEB J* 13: 9–22, 1999.
109. Nowak DG, Amin EM, Rennel ES, Hoareau-Aveilla C, Gammons M, Damodoran G, Hagiwara M, Harper SJ, Woolard J, Ladomery MR, and Bates DO. Regulation of vascular endothelial growth factor (VEGF) splicing from pro-angiogenic to anti-angiogenic isoforms: a novel therapeutic strategy for angiogenesis. *J Biol Chem* 285: 5532–5540, 2010.
110. Nowak DG, Woolard J, Amin EM, Konopatskaya O, Saleem MA, Churchill AJ, Ladomery MR, Harper SJ, and Bates DO. Expression of pro- and anti-angiogenic isoforms of VEGF is differentially regulated by splicing and growth factors. *J Cell Sci* 121: 3487–3495, 2008.
111. Ohrt T, Prior M, Dannenberg J, Odenwalder P, Dybkov O, Rasche N, Schmitzova J, Gregor I, Fabrizio P, Enderlein J, and Luhrmann R. Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* 18: 1244–1256, 2012.
112. Oshima M, Oshima H, and Taketo MM. TGF-beta receptor type II deficiency results in defects of yolk sac hematopoiesis and vasculogenesis. *Dev Biol* 179: 297–302, 1996.
113. Ota T, Fujii M, Sugizaki T, Ishii M, Miyazawa K, Aburatani H, and Miyazono K. Targets of transcriptional regulation by two distinct type I receptors for transforming growth factor-beta in human umbilical vein endothelial cells. *J Cell Physiol* 193: 299–318, 2002.
114. Pages G and Pouyssegur J. Transcriptional regulation of the vascular endothelial growth factor gene—a concert of activating factors. *Cardiovasc Res* 65: 564–573, 2005.
115. Palmer RM, Ferrige AG, and Moncada S. Nitric oxide release accounts for the biological activity of endothelium-derived relaxing factor. *Nature* 327: 524–526, 1987.
116. Parton RG. Caveolae and caveolins. *Curr Opin Cell Biol* 8: 542–548, 1996.
117. Pfeifer A, Klatt P, Massberg S, Ny L, Sausbier M, Hirneiss C, Wang GX, Korth M, Aszodi A, Andersson KE, Krombach F, Mayerhofer A, Ruth P, Fassler R, and Hofmann F. Defective smooth muscle regulation in cGMP kinase I-deficient mice. *EMBO J* 17: 3045–3051, 1998.
118. Pohl U, Holtz J, Busse R, and Bassenge E. Crucial role of endothelium in the vasodilator response to increased flow *in vivo*. *Hypertension* 8: 37–44, 1986.
119. Polytarchou C, Iliopoulos D, Hatziapostolou M, Kottakis F, Maroulakou I, Struhl K, and Tsichlis PN. Akt2 regulates all Akt isoforms and promotes resistance to hypoxia through induction of miR-21 upon oxygen deprivation. *Cancer Res* 71: 4720–4731, 2011.
120. Pou S, Pou WS, Bredt DS, Snyder SH, and Rosen GM. Generation of superoxide by purified brain nitric oxide synthase. *J Biol Chem* 267: 24173–24176, 1992.
121. Prasad J, Colwill K, Pawson T, and Manley JL. The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Mol Cell Biol* 19: 6991–7000, 1999.
122. Rahimi N, Dayanir V, and Lashkari K. Receptor chimeras indicate that the vascular endothelial growth factor receptor-1 (VEGFR-1) modulates mitogenic activity of VEGFR-2 in endothelial cells. *J Biol Chem* 275: 16986–16992, 2000.
123. Ray BN, Lee NY, How T, and Blobe GC. ALK5 phosphorylation of the endoglin cytoplasmic domain regulates Smad1/5/8 signaling and endothelial cell migration. *Carcinogenesis* 31: 435–441, 2010.
124. Rossbach O, Hung LH, Khrameeva E, Schreiner S, Konig J, Curk T, Zupan B, Ule J, Gelfand MS, and Bindereif A. Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol* 11: 146–155, 2014.
125. Sakoda T, Hirata K, Kuroda R, Miki N, Suematsu M, Kawashima S, and Yokoyama M. Myristoylation of endothelial cell nitric oxide synthase is important for extracellular release of nitric oxide. *Mol Cell Biochem* 152: 143–148, 1995.
126. Schmitt CA, Heiss EH, Aristei Y, Severin T, and Dirsch VM. Norfuraneol dephosphorylates eNOS at threonine 495 and enhances eNOS activity in human endothelial cells. *Cardiovasc Res* 81: 750–757, 2009.
127. Segarini PR, Rosen DM, and Seyedin SM. Binding of transforming growth factor-beta to cell surface proteins varies with cell type. *Mol Endocrinol* 3: 261–272, 1989.
128. Sessa WC. eNOS at a glance. *J Cell Sci* 117: 2427–2429, 2004.
129. Sessa WC, Barber CM, and Lynch KR. Mutation of N-myristoylation site converts endothelial cell nitric oxide synthase from a membrane to a cytosolic protein. *Circ Res* 72: 921–924, 1993.
130. Sharina IG, Jelen F, Bogatenkova EP, Thomas A, Martin E, and Murad F. Alpha1 soluble guanylyl cyclase (sGC) splice forms as potential regulators of human sGC activity. *J Biol Chem* 283: 15104–15113, 2008.
131. Sharp PA and Burge CB. Classification of introns: U2-type or U12-type. *Cell* 91: 875–879, 1997.
132. She X, Matsuno F, Harada N, Tsai H, and Seon BK. Synergy between anti-endoglin (CD105) monoclonal antibodies and TGF-beta in suppression of growth of human endothelial cells. *Int J Cancer* 108: 251–257, 2004.
133. Shimokawa H, Flavahan NA, and Vanhoutte PM. Loss of endothelial pertussis toxin-sensitive G protein function in atherosclerotic porcine coronary arteries. *Circulation* 83: 652–660, 1991.
134. Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, and Moore MJ. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* 151: 750–764, 2012.
135. Soker S, Takashima S, Miao HQ, Neufeld G, and Klagsbrun M. Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. *Cell* 92: 735–745, 1998.
136. Sowa G, Liu J, Papapetropoulos A, Rex-Haffner M, Hughes TE, and Sessa WC. Trafficking of endothelial nitric-oxide synthase in living cells. Quantitative evidence supporting the role of palmitoylation as a kinetic trapping mechanism limiting membrane diffusion. *J Biol Chem* 274: 22524–22531, 1999.
137. Staley JP and Guthrie C. An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* 3: 55–64, 1999.

138. Stangl K, Cascorbi I, Laule M, Klein T, Stangl V, Rost S, Wernecke KD, Felix SB, Bindereif A, Baumann G, and Roots I. High CA repeat numbers in intron 13 of the endothelial nitric oxide synthase gene and increased risk of coronary artery disease. *Pharmacogenetics* 10: 133–140, 2000.
139. Stickeler E, Fraser SD, Honig A, Chen AL, Berget SM, and Cooper TA. The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *EMBO J* 20: 3821–3830, 2001.
140. Stone JR and Collins T. Rapid phosphorylation of heterogeneous nuclear ribonucleoprotein C1/C2 in response to physiologic levels of hydrogen peroxide in human endothelial cells. *J Biol Chem* 277: 15621–15628, 2002.
141. Stuehr DJ. Structure-function aspects in the nitric oxide synthases. *Annu Rev Pharmacol Toxicol* 37: 339–359, 1997.
142. Sugamura K and Keaney JF, Jr. Reactive oxygen species in cardiovascular disease. *Free Radic Biol Med* 51: 978–992, 2011.
143. Takahashi T, Yamaguchi S, Chida K, and Shibuya M. A single autophosphorylation site on KDR/Flk-1 is essential for VEGF-A-dependent activation of PLC-gamma and DNA synthesis in vascular endothelial cells. *EMBO J* 20: 2768–2778, 2001.
144. Ting SB, Caddy J, Hislop N, Wilanowski T, Auden A, Zhao LL, Ellis S, Kaur P, Uchida Y, Holleran WM, Elias PM, Cunningham JM, and Jane SM. A homolog of *Drosophila* grainy head is essential for epidermal integrity in mice. *Science* 308: 411–413, 2005.
145. Ting SB, Wilanowski T, Auden A, Hall M, Voss AK, Thomas T, Parekh V, Cunningham JM, and Jane SM. Inositol- and folate-resistant neural tube defects in mice lacking the epithelial-specific factor Grhl-3. *Nat Med* 9: 1513–1519, 2003.
146. Ting SB, Wilanowski T, Cerruti L, Zhao LL, Cunningham JM, and Jane SM. The identification and characterization of human Sister-of-Mammalian Grainyhead (SOM) expands the grainyhead-like family of developmental transcription factors. *Biochem J* 370: 953–962, 2003.
147. Toporsian M, Gros R, Kabir MG, Vera S, Govindaraju K, Eidelman DH, Husain M, and Letarte M. A role for endoglin in coupling eNOS activity and regulating vascular tone revealed in hereditary hemorrhagic telangiectasia. *Circ Res* 96: 684–692, 2005.
148. Traylor-Knowles N, Hansen U, Dubuc TQ, Martindale MQ, Kaufman L, and Finnerty JR. The evolutionary diversification of LSF and Grainyhead transcription factors preceded the radiation of basal animal lineages. *BMC Evol Biol* 10: 101, 2010.
149. Turunen JJ, Niemela EH, Verma B, and Frilander MJ. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA* 4: 61–76, 2013.
150. Vasquez-Vivar J, Kalyanaraman B, Martasek P, Hogg N, Masters BS, Karoui H, Tordo P, and Pritchard KA, Jr. Superoxide generation by endothelial nitric oxide synthase: the influence of cofactors. *Proc Natl Acad Sci U S A* 95: 9220–9225, 1998.
151. Velasco B, Ramirez JR, Relloso M, Li C, Kumar S, Lopez-Bote JP, Perez-Barriocanal F, Lopez-Novoa JM, Cowan PJ, d'Apice AJ, and Bernabeu C. Vascular gene transfer driven by endoglin and ICAM-2 endothelial-specific promoters. *Gene Ther* 8: 897–904, 2001.
152. Velasco S, Alvarez-Munoz P, Pericacho M, Dijke PT, Bernabeu C, Lopez-Novoa JM, and Rodriguez-Barbero A. L- and S-endoglin differentially modulate TGFbeta1 signaling mediated by ALK1 and ALK5 in L6E9 myoblasts. *J Cell Sci* 121: 913–919, 2008.
153. Venema VJ, Marrero MB, and Venema RC. Bradykinin-stimulated protein tyrosine phosphorylation promotes endothelial nitric oxide synthase translocation to the cytoskeleton. *Biochem Biophys Res Commun* 226: 703–710, 1996.
154. Venkatesan K, McManus HR, Mello CC, Smith TF, and Hansen U. Functional conservation between members of an ancient duplicated transcription factor family, LSF/Grainyhead. *Nucleic Acids Res* 31: 4304–4316, 2003.
155. Versteeg HH, Heemskerk JW, Levi M, and Reitsma PH. New fundamentals in hemostasis. *Physiol Rev* 93: 327–358, 2013.
156. Wahl MC, Will CL, and Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 136: 701–718, 2009.
157. Waltenberger J, Claesson-Welsh L, Siegbahn A, Shibuya M, and Heldin CH. Different signal transduction properties of KDR and Flt1, two receptors for vascular endothelial growth factor. *J Biol Chem* 269: 26988–26995, 1994.
158. Wang R, Crystal RG, and Hackett NR. Identification of an exonic splicing silencer in exon 6A of the human VEGF gene. *BMC Mol Biol* 10: 103, 2009.
159. Wang Z and Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14: 802–813, 2008.
160. Watanabe K, Hasegawa Y, Yamashita H, Shimizu K, Ding Y, Abe M, Ohta H, Imagawa K, Hojo K, Maki H, Sonoda H, and Sato Y. Vasohibin as an endothelium-derived negative feedback regulator of angiogenesis. *J Clin Invest* 114: 898–907, 2004.
161. Wei WJ, Mu SR, Heiner M, Fu X, Cao LJ, Gong XF, Bindereif A, and Hui J. YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. *Nucleic Acids Res* 40: 8622–8636, 2012.
162. Weigand JE, Boeckel JN, Gellert P, and Dimmeler S. Hypoxia-induced alternative splicing in endothelial cells. *PLoS One* 7: e42697, 2012.
163. Wilanowski T, Tuckfield A, Cerruti L, O'Connell S, Saint R, Parekh V, Tao J, Cunningham JM, and Jane SM. A highly conserved novel family of mammalian developmental transcription factors related to *Drosophila* grainyhead. *Mech Dev* 114: 37–50, 2002.
164. Wilcox JN, Subramanian RR, Sundell CL, Tracey WR, Pollock JS, Harrison DG, and Marsden PA. Expression of multiple isoforms of nitric oxide synthase in normal and atherosclerotic vessels. *Arterioscler Thromb Vasc Biol* 17: 2479–2488, 1997.
165. Will CL and Lührmann R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 3: a003707, 2011.
166. Xia Y, Tsai AL, Berka V, and Zweier JL. Superoxide generation from endothelial nitric-oxide synthase. A Ca2+/calmodulin-dependent and tetrahydrobiopterin regulatory process. *J Biol Chem* 273: 25804–25808, 1998.
167. Yu Z, Mannik J, Soto A, Lin KK, and Andersen B. The epidermal differentiation-associated Grainyhead gene Get1/Grhl3 also regulates urothelial differentiation. *EMBO J* 28: 1890–1903, 2009.
168. Zimmermann K, Opitz N, Dedio J, Renne C, Muller-Esterl W, and Oess S. NOSTRIN: a protein modulating nitric oxide release and subcellular distribution of endothelial nitric oxide synthase. *Proc Natl Acad Sci U S A* 99: 17167–17172, 2002.

Address correspondence to:
Dr. Joachim Altschmied
Heisenberg-Group—Environmentally-Induced Cardiovascular Degeneration
IUF—Leibniz Research Institute for Environmental Medicine
Auf'm Hennekamp 50
Düsseldorf 40225
Germany

E-mail: joachim.altschmied@uni-duesseldorf.de

Prof. Heiner Schaal
Institute for Virology
Heinrich-Heine-University Düsseldorf
Universitätsstrasse 1
Düsseldorf 40225
Germany

E-mail: schaal@uni-duesseldorf.de

Date of first submission to ARS Central, June 15, 2014; date of final revised submission, August 18, 2014; date of acceptance, September 5, 2014.

Abbreviations Used

3′ss = 3′ splice site
5′ss = 5′ splice site
ACVRL1 = activin A receptor type II like 1
BH_4 = tetrahydrobiopterin
BPS = branch point sequence
CaM = calmodulin
CAV1 = caveolin-1
DSS = distal splice acceptor site
EJC = exon junction complex
eNOS = endothelial nitric oxide synthase
FAD = flavin adenine dinucleotide
FMN = flavin mononucleotide
GRH = *Drosophila* grainyhead
GRHL3 = grainyhead-like 3
HIF-1α = hypoxia inducible factor-1α
hnRNP = heterogeneous nuclear ribonucleoparticle
Hsp90 = heat shock protein 90
IGF1 = insulin-like growth factor 1
MAPK = mitogen-activated protein kinase
$NADP^+$ = nicotinamide adenine dinucleotide phosphate
NO = nitric oxide
NOSIP = eNOS interacting protein
NOSTRIN = eNOS traffic inducer
NRP1 = neuropilin 1
NTC = NineTeen complex
PKC = protein kinase C
PPT = polypyrimidine tract
PSS = proximal splice acceptor site
PTB = phosphotyrosine binding
PTK2 = proline-rich tyrosine kinase 2
ROS = reactive oxygen species
SA = splice acceptor
SD = splice donor
snRNP = small nuclear ribonucleoprotein particle
SR proteins = serine/arginine-rich proteins
SRPK1 and 2 = SR protein kinases 1 and 2
TGFBR1 and 2 = TGFβ receptor 1 and 2
TGFβ1, 2 and 3 = transforming growth factor β1, 2 and 3
TNFα = tumor necrosis factor α
VEGFA = vascular endothelial growth factor A
VEGFR1 and 2 = VEGF receptor 1 and 2
WT1 = Wilms' tumor suppressor 1 protein
YB-1 = Y box-binding protein 1

This article has been cited by:

1. F. Wu, X. J. Dong, H. Q. Zhang, L. Li, Q. L. Xu, Z. F. Liu, Z. T. Gu, L. Su. 2016. Role of MnSOD in propofol protection of human umbilical vein endothelial cells injured by heat stress. *Journal of Anesthesia* **30**:3, 410-419. [CrossRef]
2. Eckers Anna, Haendeler Judith. 2015. Endothelial Cells in Health and Disease. *Antioxidants & Redox Signaling* **22**:14, 1209-1211. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]

3. Succession of splicing regulatory elements determines cryptic 5'ss functionality

The following data are published in Nucleic Acids Res. 2016 Dec 29. pii: gkw1317. (doi: 10.1093/nar/gkw1317) by

***Brillen, A.L.**, Schöneweis, K., Walotka, L., Hartmann, L., Müller, L., Ptok, J., Kaisers, W., Poschmann, G., Stühler, K., Buratti, E., Theiss, S., Schaal, H.*

Contribution

ALB and HS conceived the study and designed the experiments. ALB, KS, LW, LH and LM performed cloning, transfection experiments and (q)RT-PCR analyses. ST, ALB and HS performed HEXplorer analyses. ALB performed RNA-pull-down analyses. KS carried out FACS analysis. ST, JP, WK and HS carried out bioinformatical analyses. GP and KS performed MS analysis. EB helped to draft the manuscript. ALB, ST and HS wrote the manuscript.

Abstract

A critical step in exon definition is the recognition of a proper splice donor (5'ss) by the 5' end of U1 snRNA. In the selection of appropriate 5'ss, *cis*-acting splicing regulatory elements (SREs) are indispensable. As a model for 5'ss recognition, we investigated cryptic 5'ss selection within the human fibrinogen Bβ-chain gene (FGB) exon 7, where we identified several exonic SREs that simultaneously acted on up- and downstream cryptic 5'ss. In the FGB exon 7 model system, 5'ss selection iteratively proceeded along an alternating sequence of U1 snRNA binding sites and interleaved SREs which in principle supported different 3' exon ends. Like in a relay race, SREs either suppressed a potential 5'ss and passed the splicing baton on or splicing actually occurred. From RNA-Seq data, we systematically selected 19 genes containing exons

with silent U1 snRNA binding sites competing with nearby highly used 5'ss. Extensive SRE analysis by different algorithms found authentic 5'ss significantly more supported by SREs than silent U1 snRNA binding sites, indicating that our concept may permit generalization to a model for 5'ss selection and 3' exon end definition.

Nucleic Acids Research, 2016 **1–15**
doi: 10.1093/nar/gkw1317

Succession of splicing regulatory elements determines cryptic 5'ss functionality

Anna-Lena Brillen[1], Katrin Schöneweis[2], Lara Walotka[1], Linda Hartmann[1], Lisa Müller[1], Johannes Ptok[1], Wolfgang Kaisers[3], Gereon Poschmann[4], Kai Stühler[4,5], Emanuele Buratti[6], Stephan Theiss[7,*] and Heiner Schaal[1,*]

[1]Institute for Virology, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany, [2]Department of Infectious Diseases, Molecular Virology, University Hospital Heidelberg, 69120 Heidelberg, Germany, [3]Department of Anesthesiology, University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany, [4]Molecular Proteomics Laboratory, BMFZ, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany, [5]Institute for Molecular Medicine, University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany, [6]International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park, 34149 Trieste, Italy and [7]Institute of Clinical Neuroscience and Medical Psychology, Heinrich-Heine-University Düsseldorf, 40225

Received July 26, 2016; Revised December 7, 2016; Editorial Decision December 18, 2016; Accepted December 19, 2016

ABSTRACT

A critical step in exon definition is the recognition of a proper splice donor (5'ss) by the 5' end of U1 snRNA. In the selection of appropriate 5'ss, *cis*-acting splicing regulatory elements (SREs) are indispensable. As a model for 5'ss recognition, we investigated cryptic 5'ss selection within the human fibrinogen Bβ-chain gene (FGB) exon 7, where we identified several exonic SREs that simultaneously acted on up- and downstream cryptic 5'ss. In the FGB exon 7 model system, 5'ss selection iteratively proceeded along an alternating sequence of U1 snRNA binding sites and interleaved SREs which in principle supported different 3' exon ends. Like in a relay race, SREs either suppressed a potential 5'ss and passed the splicing baton on or splicing actually occurred. From RNA-Seq data, we systematically selected 19 genes containing exons with silent U1 snRNA binding sites competing with nearby highly used 5'ss. Extensive SRE analysis by different algorithms found authentic 5'ss significantly more supported by SREs than silent U1 snRNA binding sites, indicating that our concept may permit generalization to a model for 5'ss selection and 3' exon end definition.

INTRODUCTION

Alternative 5′ splice site selection is a highly regulated process involving degenerate sequence elements that are recognized by a large intricate protein complex, the spliceosome, which is composed of five small nuclear ribonucleoprotein particles (snRNPs). Spliceosome assembly starts with the interaction of the U1 snRNP and the 5'ss at the exon–intron border. Since within vertebrates, relatively small exons are separated by much longer introns, splice site pairing is supposed to first occur across the exon through subsequent binding of the U2 snRNP to the branch point sequence of the upstream 3′ splice site (3′ss or splice acceptor) (1,2). Both snRNPs interact with each other forming the 'exon-definition' complex (2), which is later converted into 'intron-definition' complexes (3,4), connecting U1 and U2 snRNPs across the intron and triggering the splicing reaction.

Splice donor choice, however, is not only directed by the spliceosome itself, recognizing the 11 nt long sequence of the 5′ss to form an RNA duplex with the 5′end of U1 snRNA (3,4), but also critically depends on RNA binding proteins that bind to splicing regulatory elements (SREs) in the vicinity of splice sites, like SR (serine-arginine-rich) or hnRNP (heterogeneous nuclear ribonucleoparticle) proteins (5). SR proteins are composed of one or two RNA binding domains (RRM) and an arginine-serine (RS-)rich domain that participate directly in the interaction with other proteins or with RNA itself. Both domains have been shown to be capable of participating in U1 snRNP recruitment to the 5'ss via the U1-specific protein U1-70K (6–9). Dependent on their binding position to the exon or intron, SR proteins can generally act in a position-dependent manner, either activating or silencing splice donor usage (10,11). Generally, SR proteins enhance 5'ss use, when they bind to the upstream exon, while they repress splicing from

*To whom correspondence should be addressed. Tel: +49 211 81 12393; Fax: +49 211 81 10856; Email: schaal@uni-duesseldorf.de
Correspondence may also be addressed to Stephan Theiss. Email: theiss@uni-duesseldorf.de

the downstream intron. During repression, proteins bound to inhibitory SREs interfere with further progression into late spliceosomal complexes and form so-called 'dead-end' complexes (10,12,13).

In human genetics, the computational identification of aberrant splice donor usage due to nucleotide exchanges is vitally needed for diagnostics, and evaluating a mutation's biological relevance for clinical treatment of patients with hereditary disorders is indispensable (14,15).

By now, variations in *cis*- or *trans*-acting elements within protein coding genes have been associated with altering splicing patterns and thereby inducing genetic defects that cause human diseases (16). Estimates of the fraction of human inherited disease mutations that affect splicing range from 10% for mutations located directly within splice sites (17) and can even reach 22–25% if mutations within SREs were considered (18,19). Thus, roughly 1/3 of all nucleotide mutations leading to human disease result in exon skipping, use of cryptic splice sites or intron retention, leaving aside SREs that have not been discovered yet. However, since cryptic sites are splicing inactive as long as the authentic 5′ss is functional, it seems that splice site choice simply follows a 'winner-takes-it-all' rule. If the authentic 5′ss is weakened, however, it is generally unclear whether exon skipping or cryptic splicing occurs.

Although highly desirable, there is no single *in silico* tool available yet, providing reliable predictions of splice site usage. Algorithms like MaxEnt (20) and HBond (3,4) for 5'ss scoring, as well as e.g. ΔtESRseq (15) or HEXplorer-based (21) approaches calculating enhancing or silencing properties of regions in the vicinity of splice sites, greatly assist in this daunting task.

In this work we show that a tight cluster of alternating multiple SREs and U1 snRNA binding sites controls cryptic splice donor usage throughout the human fibrinogen Bβ-chain gene (FGB) exon 7. Based on HEXplorer profiles, we predicted several SREs that we confirmed by mutational analyses. Motifs identified in these *cis*-acting SREs exhibited some degeneracy with respect to the binding splicing regulatory proteins SRSF1 and Tra2β, indicating a possible redundancy. Splicing regulatory proteins bound to these SREs acted in a strictly position-dependent manner, each functioning as a gateway that either terminated the exon or passed on an 'exon end signal' to the next U1 snRNA binding site.

MATERIALS AND METHODS

Single-intron splicing constructs

Constructs SV guanosine-adenosine-rich (GAR) SD4 Δvpu env eGFP D36G, SV GAR¯ SD4 Δvpu env eGFP D36G, SV GAR¯ESE¯ SD4 Δvpu env eGFP D36G are based on the HIV-1 glycoprotein/eGFP expression plasmid and have been described before (3,22). Inserting the neutral sequence CCAAACAA (23) was carried out by replacing GAR with a polymerase chain reaction (PCR) product obtained with primer pair #3378/#3379. All FGB exon 7-dervied fragments were inserted into SV GAR SD4 Δvpu env eGFP D36G, replacing the GAR element with DNA fragments obtained with primer pairs #3168/#3169 (FGB7-A), #3166/#3167 (FGB7-B), #3170/#3171 (FGB7-C), #3172/#3173 (FGB7-D), #3326/#3327 (FGB7-E) and #3174/#3175 (FGB7-F), respectively. SV FGB7-D(8A) SD4 Δvpu env eGFP D36G, SV FGB7-D(5C) SD4 Δvpu env eGFP D36G and SV FGB7-D(5C+8A) SD4 Δvpu env eGFP D36G were constructed by replacing GAR with PCR products resulting from primer pair #3479/#3480, #3481/#3482 and #3483/#3484, respectively.

Fibrinogen Bβ minigenes

The plasmids pT-Bβ-WT and pT-Bβ-IVS7+1G>T were previously described (24,25). pT-Bβ-IVS7+1G>A and pT-Bβ-IVS7+2T>A were cloned via mutagenesis PCR of pT-Bβ-WT using the primer pairs #5659/#5660, and #5661/#5660, respectively. pT-Bβ-IVS7+1G>T-mt-c1 was cloned via a mutagenesis PCR of pT-Bβ-IVS7+1G>T with primers #2619/#2622 and #2620/#2621; pT-Bβ-IVS7+1G>T-mt-c1/c2* #2619/#2647 and #2620/#2646; pT-Bβ-IVS7+1G>T-mt-c1/c2*/c3 with primers #2619/#2624 and #2620/#2623; pT-Bβ-WT-c1-15.8 was cloned via a mutagenesis PCR of pT-Bβ-WT using primers #2619/#2765 and #2620/#2764. pT-Bβ-WT-c1-18.8 was cloned via a mutagenesis PCR of pT-Bβ-WT-c1-15.8 using primers #2619/#2872 and #2620/#2871; pT-Bβ-WT-c1-20.8 using primers #2619/#2874 and #2620/#2873. pT-Bβ-WT-c3-15.8 was cloned via a mutagenesis PCR of pT-Bβ-WT with primers #2619/#2925 and #2620/#2924; pT-Bβ-WT-c3-18.8 with primers #2619/#2927 and #2620/#2926; pT-Bβ-WT-c3-20.8 with primers #2619/#2929 and #2620/#2928. HEXplorer-guided mutations of fragments B-D were inserted via mutagenesis PCR of pT-Bβ-WT or –IVS, respectively, with primers #5568/#2620 and #2619/#5569 (B), #5566/#2620 and #2619/#5567 (C), #3548/#2620 and #3549/#2619 (D), #5571/#2620 and 2619/#5569 (B/C), #5568/#2620 and #2619/#5569 (B/D), #5570/#2620 and #2619/#5567 (C/D), #5571/#2620 and #2619/#5569 (B/C/D). Exon 7 was replaced with only splicing neutral sequences (25) by using a customized synthetic gene from Invitrogen and inserted into pT-Bβ-IVS7+1G>T via EcoNI/Bpu10I. FGB7-derived fragments were inserted with PCR products resulting from primer pairs #4835/2620 (B), #5179/2620 (B MUT), #5581/#2620 (C), #5585/#2620 (C MUT), #4703/#2620 (D) and #4791/#2620 (D MUT). Fragments derived from the E1α PDH gene were inserted with PCR products resulting from primer pairs #5497/#5498 (WT) and #5499/5500 (MUT) and fragments derived from the SNAPC4 gene (ENSG00000165684) with primer pairs #5498/#5490 (WT) and #5491/#5492 (MUT).

Expression plasmids

pXGH5 (26) was cotransfected to monitor transfection efficiency.

Oligonucleotides

All oligonucleotides used were obtained from Metabion GmbH (Planegg, Germany) (see Supplementary Table S1).

Cell culture and RT-PCR analysis

HeLa cells were cultivated in Dulbecco's high-glucose modified Eagle's medium (Invitrogen) supplemented with 10% fetal calf serum and 50 μg/ml penicillin and streptomycin each (Invitrogen). Transient-transfection experiments were performed with six-well plates at 2.5×10^5 cells per well by using TransIT®-LT1 transfection reagent (Mirus Bio LLC US) according to the manufacturer's instructions. Total RNA was isolated 24 h post-transfection by using acid guanidinium thiocyanate-phenol-chloroform as described previously (27). For (q)RT-PCR analyses, RNA was reversely transcribed by using Superscript III Reverse Transcriptase (Invitrogen) and Oligo(dT) primer (Invitrogen). For the analyses of the single-intron splicing constructs primer pair #3210/#3211 was used; for the analyses of the Fibrinogen Bβ minigenes, primer pair #2648/#2649 was used. Quantitative RT-PCR analysis was performed by using the qPCR MasterMix (PrimerDesign Ltd) and LightCycler 1.5 (Roche). For normalization, primers #1224/#1225 were used and the level of hGH present in each sample was monitored.

FACS analysis

Fluorescence-activated cell sorting (FACS) analysis for the measurement of quantitative eGFP expression was carried out using FACS Canto2 (BD Biosciences). First, cells were washed with PBS and incubated with trypsin for 5 min. After several washing steps (PBS + 3% FCS), samples were acquired to the cytometer. Next, data was edited using the FlowJo analysis software (Tree Star, Inc.).

Inhibition of translation by cycloheximide

In order to detect nonsense-mediated decay (NMD)-sensitive transcripts, cells were incubated with 50 μg/ml of the translational inhibitor cycloheximide (CHX) 6 h prior to harvesting. As control for CHX treatment, we amplified RNA encoding SRSF3 (SRp20) with specific primers binding within exon 1 and 5 (#4003/#4004). WT SRSF3-messages exclude the poison cassette exon 4, while transcripts including exon 4 contain a pre-mature stop codon and get degraded by NMD (28).

Protein isolation by RNA affinity chromatography

Three thousand picomoles of short RNA oligonucleotides for either wild-type (WT) or mutant version of FGB7-derived fragments B (#5170; #5173), C (#5572; #5573) and D (#5167; #5168) or the neutral sequence (#5169), respectively, were covalently coupled to adipic acid dihydrazide-agarose beads (Sigma). 60% of HeLa nuclear extract (Cilbiotech) was added to the immobilized RNAs. After stringent washing with buffer D containing different concentrations of KCl (20mM HEPES-KOH [pH 7.9], 5%[vol/vol] glycerol, 0.1-0.5 M KCl, 0.2 M ethylenediaminetetraacetic acid, 0.5 mM dithiothreitol, 0.4M $MgCl_2$), precipitated proteins were eluted in protein sample buffer. Samples were heated up to 95°C for 10 min and loaded onto sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) for western blot analysis. Samples were transferred to a nitrocellulose membrane probed with primary and secondary antibodies (SRSF1 (Invitrogen 32–4500), Tra2β (abcam ab31353), MS2 (Tetracore TC-7004-002)) and developed with ECL chemiluminescence reagent (GE Healthcare).

HEXplorer score calculation

HEXplorer score profiles of pairs of WT and mutant sequences were calculated using the web resource https://www.hhu.de/rna/html/hexplorer_score.php (21).

Mass spectrometric analysis

Protein samples were shorty separated over about 4 mm running distance in a 4–12% polyacrylamide gel. After silver staining, protein containing bands were excised and prepared for mass spectrometric analysis as described (29). Briefly, samples were destained, reduced with dithiothreitol, alkylated with iodoacetamide and digested with trypsin. Resulting peptides were extracted from the gel piece and finally resuspended in 0.1% trifluoroacetic acid.

Initially, peptides were separated by liquid chromatography on an Ultimate 3000 Rapid Separation Liquid Chromatography system (RSLC, Thermo Scientific, Dreieich, Germany). A trap column (Acclaim PepMap100, 3 μm C18 particle size, 100 Å pore size, 75 μm inner diameter, 2 cm length, Thermo Scientific, Dreieich, Germany) was used for peptide pre-concentration at a flow rate of 6 μl/min for ten minutes using 0.1% trifluoroacetic acid as mobile phase. Subsequently, peptides were separated on a 25 cm length analytical column (Acclaim PepMapRSLC, 2 μm C18 particle size, 100 Å pore size, 75 μm inner diameter, Thermo Scientific, Dreieich, Germany) at a flow rate of 300 nl/min at 60°C using a 2 h gradient from 4 to 40% solvent B (0.1% (v/v) formic acid, 84% (v/v) acetonitrile in water) in solvent A (0.1% (v/v) formic acid in water). Peptides were injected into the mass spectrometer by distal coated Silica Tip emitters (New Objective, Woburn, MA, USA) via a nano electrospray ionization source using a spray voltage of 1.4 kV.

Tandem mass spectra were recorded in a data dependent setting with an Orbitrap Elite (Thermo Scientific, Dreieich, Germany) hybrid mass spectrometer in positive mode. Full scans (resolution 60 000) were recorded over a scan range of 350–1700 m/z with a maximal ion time of 200 ms and the target value for automatic gain control set to 1000 000 in profile mode in the orbitrap part of the instrument. Subsequently, up to twenty precursors at charge states two and three were isolated (isolation window 2 m/z), fragmented by collision induced dissociation and analyzed with a maximal ion time of 50 ms and the target value for automatic gain control set to 3000 (available mass range 50–2000 m/z, resolution 5400) in the linear ion trap part of the instrument. Already analyzed precursors were excluded from further isolation and fragmentation for 45 s.

Data analysis within the MaxQuant environment (version 1.5.5.1, Max Planck Institute of Biochemistry, Planegg, Germany) was performed independently for the two replicate sample batches with standard parameters if not otherwise stated. Spectra were searched against 70 615 entries from the UniProt KB homo sapiens proteome

UP000005640 (downloaded on 16 June 2016) with label-free quantification enabled as well as the 'match between runs' option. Tryptic cleavage specificity was chosen, as well as carbamidomethyl at cysteines as fixed and methionine oxidation and acetylation at protein n-termini as variable modifications. For precursor masses, the mass tolerances were set to 20 ppm (first search) and 4.5 ppm (second search after recalibration) and for fragment masses to 0.5 Da. Peptides and proteins were accepted at a false discovery rate of 1% and only proteins considered showing two or more identified different peptides.

RESULTS

Cryptic splice site activation is mediated by SREs acting in a strictly position-dependent manner

To exemplify the complexity of aberrant splicing and the difficulty of predicting the splicing outcome caused by human pathogenic mutations, we revisited cryptic splice site usage embedded in a splicing-regulatory network of the human FGB. Here, the FGB c.1244+1G>T (aka IVS7+1G>T) minigene analysis revealed that beside exon 7 skipping this mutation caused the activation of three cryptic splice donors localized in the upstream exon: two canonical at 106 nt (c1) and 24 nt (c3), and one non-canonical 40 nt (c2*, indicated by the asterisk) upstream of the physiological 5'ss, leading to a loss of functional fibrinogen (24,25). Additionally, we observed activation of the intron localized cryptic site p1 158 nt downstream of +1G>T.

Calculating the HBond scores (HBS) of all 11 nt long GT sequences within exon 7 and the downstream intron ((6); https://www.hhu.de/rna/html/hbond_score.php) confirmed that the physiological 5′ss had the highest HBS, indicative of the highest complementarity to U1 snRNA, followed by activated cryptic splice sites in the neighborhood of the physiological 5′ss (Figure 1).

To analyze whether activation of the three exonic cryptic splice sites caused by the mutant physiological 5'ss is solely mediated by the previously identified naturally silent SRSF1 (aka SF2/ASF) binding site (25), we mutated the exonic cryptic sites one after the other. The impact of these mutations on the splicing pattern was analyzed by RT-PCR following transient transfection assays of WT or +1G>T in a three exon minigene (25). As shown in Figure 2A, following +1G>T mutation c1, c2*, c3 and p1 were activated. Surprisingly, individual inactivation of c1, however, seemed to neither effectively shift the overall splice site use toward the other cryptic sites nor change the level of exon skipping (Figure 2A, cf. lanes 2 and 3). This obvious lack of competition between these cryptic splice sites suggested that the other cryptic sites might be regulated independently of c1. In line with this, inactivation of both c1 and c2* (Figure 2A, lane 4), or all three exonic cryptic splice sites (Figure 2A, lane 5) caused much less exon skipping as expected, strengthening our hypothesis that at least one additional SRE might be located downstream of c1.

Next, we investigated whether the canonical cryptic sites c1 or c3 could outcompete the physiological WT 5'ss, if they were modified to have a similar U1 snRNA complementarity as the physiological 5'ss, i.e. similar HBS (HBS 15.0). Adapting c1 from HBS 12.2 to 15.8 (Figure 2B) did not change splice site usage (Figure 2C, cf. lanes 1 and 2), supporting our hypothesis that at least another SRE was localized within exon 7, between c1 and WT 5'ss. At the same time, such an SRE would repress c1 usage and enhance any downstream splice site (10). Thus, likewise adapting c3 from HBS 10.8 to 15.8 can be expected to switch splice site use from the physiological 5'ss to the 24 nt more proximal cryptic site c3. Indeed, increasing c3 from HBS 10.8 to 15.8 fully activated this cryptic splice site even in the presence of the physiological 5'ss (Figure 2C, cf. lanes 1 and 3), thereby shortening the exon by 24 nt. Interestingly, exclusive use of c1 was not observed even when it was increased to HBS 18.8 or even 20.8 (Figure 2C, lanes 4 and 6), speaking again for at least a second SRE downstream of c1 which would simultaneously repress the nearest upstream splice site and enhance the nearest downstream splice site, thereby extending the exon.

Multiple exonic splicing enhancers are located within FGB exon 7

To examine the validity of this concept, we analyzed splice site recognition in more detail and searched for additional functional SREs. First, we experimentally determined the impact of the putative exonic splicing enhancers on splice site recognition. For this, we used our well-characterized enhancer-dependent single intron eGFP splicing reporter (3,30), permitting to measure eGFP fluorescence intensity proportional to U1 snRNP binding to the 5'ss (21). In this way, splice site recognition can not only be measured via (q)RT-PCR, but also be quantified in an independent experimental setup by flow cytometry. Furthermore, the leader sequence of this enhancer reporter can be substituted with any putative SRE. As reference for grading downstream enhancer impact we used the HIV-1 GAR splicing enhancer, containing two SRSF1- and one SRSF5-binding sites and its mutations GAR^{-} and GAR^{-}ESE^{-} (3,22,30) (Supplementary Figure S1A). To confirm the GAR inactivating mutations, we additionally substituted the inactive GAR^{-}ESE^{-} with a CCAAACAA repeat previously shown to be splicing neutral (23) and determined their impact on 5'ss recognition. In fact, using both qRT-PCR and flow cytometry, we could measure an up to 230-fold increase in splice site recognition mediated by the GAR element, an up to 7-fold increase mediated by the GAR^{-}, but none for GAR^{-}ESE^{-} (0.5-fold activation) (Supplementary Figure S1B). In summary, the enhancer reporter allows to functionally rank strong, intermediate and not enhancing properties of SREs in comparison to the splicing neutral reference sequence.

Next, we examined six FGB exon 7-derived fragments (named FGB7-A to FGB7-F), where FGB7-B corresponded to the naturally silent SRSF1 binding site (25), and inserted each fragment into this enhancer reporter. Here, B, C and D showed an increase in splice site recognition of more than 100 to even 1000 times compared to GAR^{-}ESE^{-} (Figure 3A). Interestingly, fragment D, but not B showed the highest splicing enhancing activity and was further subjected to mutational analyses to identify the splicing regulatory protein binding to it. We used the HEXplorer algorithm (21) to predict the most promising

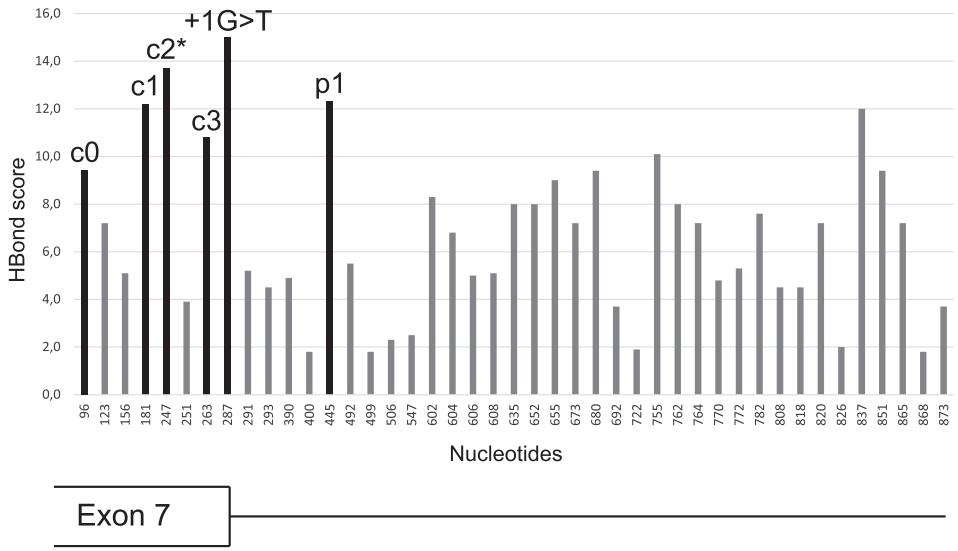
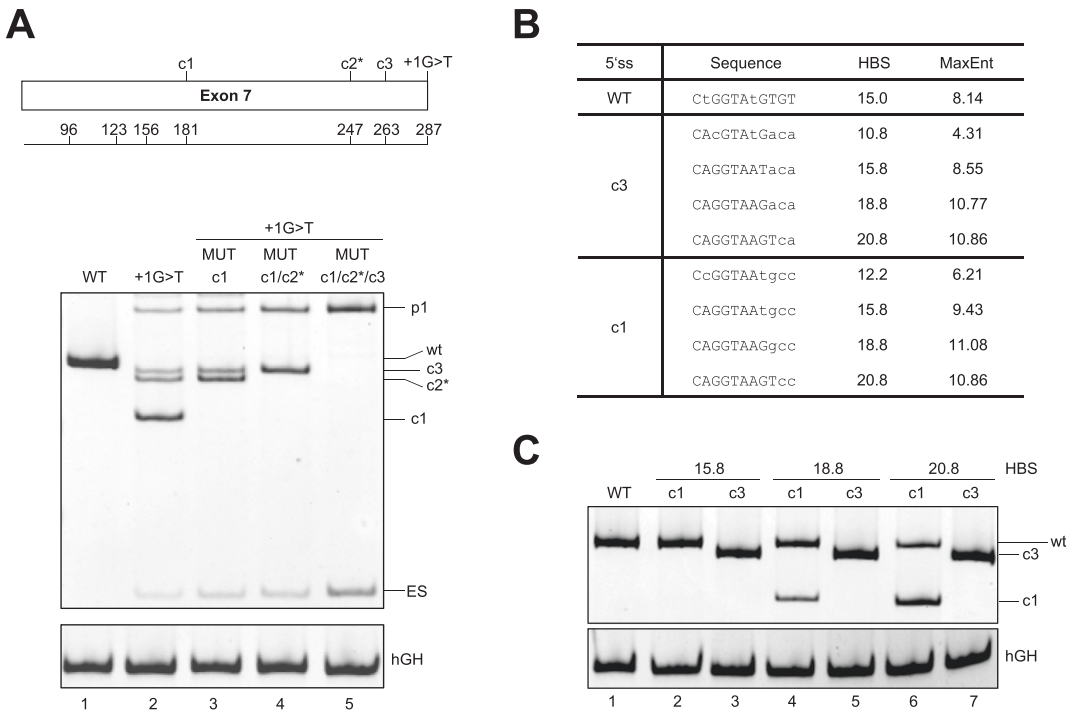


Figure 1. HBond score (HBS) profile of all 11-nt long GT sequences within the human fibrinogen Bβ-chain gene exon 7 and its downstream intron. Additionally, the GC-splice site c2* (with a substituted GT for HBS calculation) is considered and indicated by an asterisk. Numbers on the x-axis describe positions starting from the beginning of exon 7.



5'ss	Sequence	HBS	MaxEnt
WT	CtGGTAtGTGT	15.0	8.14
c3	CAcGTAtGaca	10.8	4.31
	CAGGTAATaca	15.8	8.55
	CAGGTAAGaca	18.8	10.77
	CAGGTAAGTca	20.8	10.86
c1	CcGGTAAtgcc	12.2	6.21
	CAGGTAAtgcc	15.8	9.43
	CAGGTAAGgcc	18.8	11.08
	CAGGTAAGTcc	20.8	10.86

Figure 2. Cryptic splice donor activation. (**A**) Schematic drawing of the 287-nt long human fibrinogen Bβ-chain gene exon 7 and its cryptic splice sites (top). 2.5×10^5 HeLa cells were transiently transfected with 1 μg of each construct together with 1 μg of pXGH5 (hGH) to monitor transfection efficiency. Twenty-four hours after transfection, RNA was isolated and subjected to RT-PCR analysis using primer pairs #2648/#2649 and #1224/#1225 (hGH). PCR products were separated by 10% non-denaturing polyacrylamide gel electrophoresis and stained with ethidium bromide (bottom). Exonic (c1–c3) and intronic (p1) cryptic splice donor sites as well as the skipped exon (ES) are depicted on the right hand side. (**B**) Sequences of the wild-type 5'ss (WT), c1 and c3 variants including their different HBond/MaxEnt scores. (**C**) RT-PCR analysis of splicing patterns of cryptic splice sites mutated to have higher HBS according to (B).

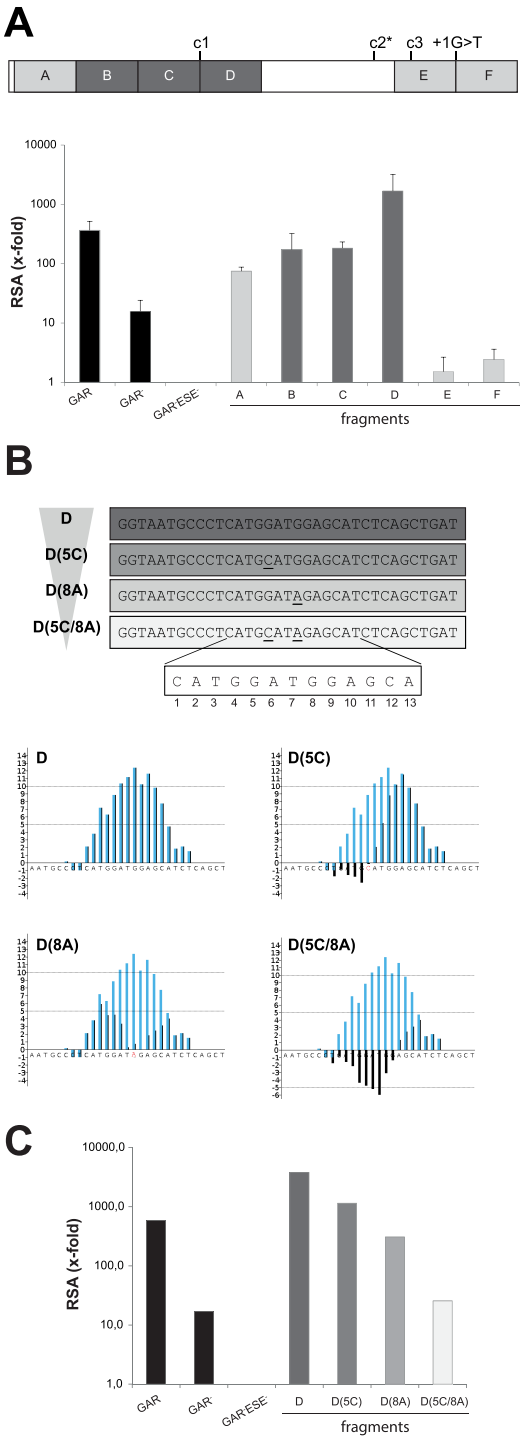
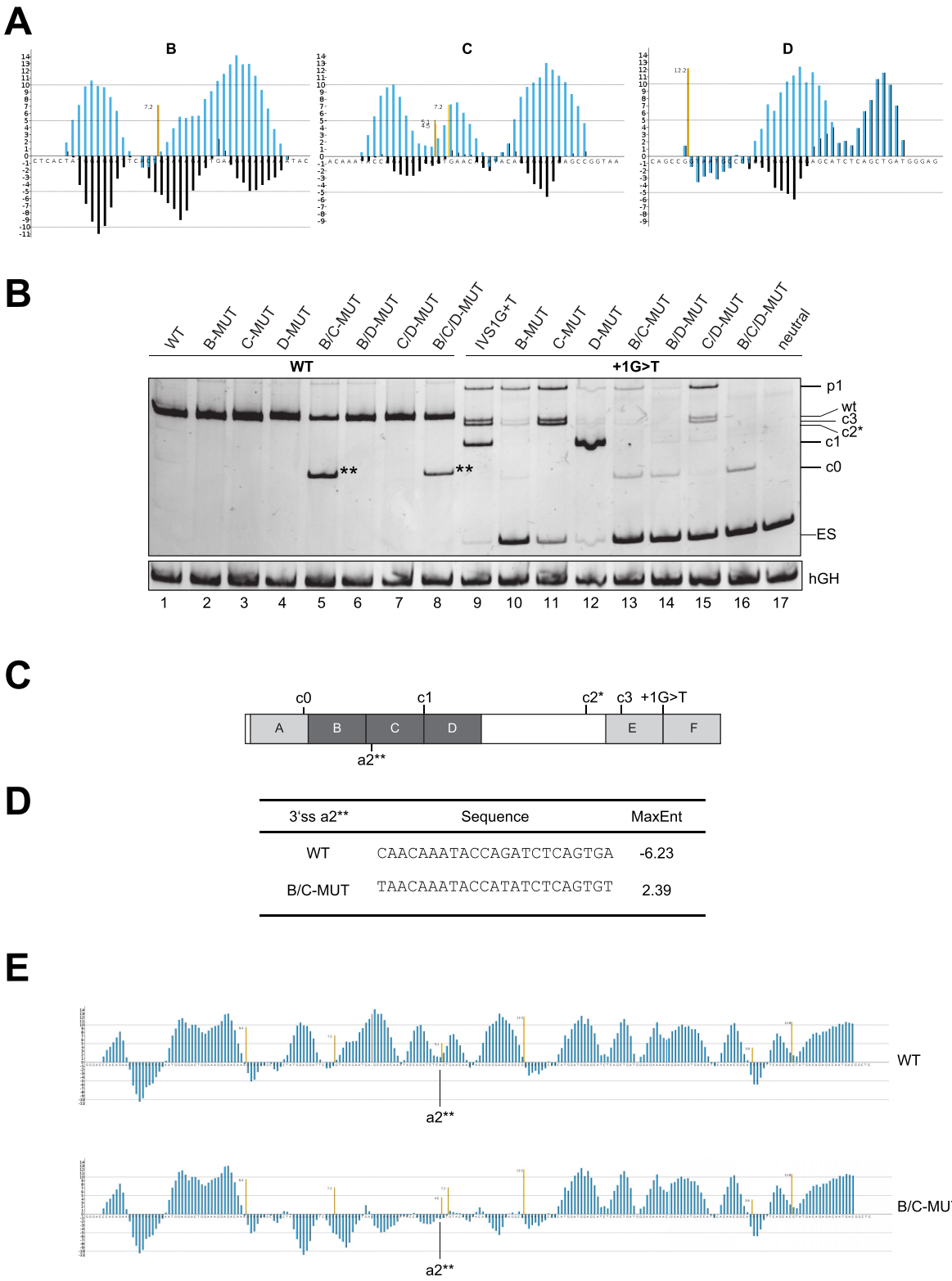


Figure 3. SREs within FGB exon 7. (**A**) Schematic of the localization of FGB exon 7 fragments used in SRE analysis. 2.5×10^5 HeLa cells were transiently transfected with 1 μg of each construct and 1 μg of pXGH5. At 24 h after transfection, total-RNA samples were collected and used for qRT-PCR with primer pair #3210/#3211 and normalized to hGH (#1224/#1225). Relative splicing activity (RSA). (**B**) Sequences (top) and HEXplorer profiles (bottom) of fragment D and its mutations. The WT profile is shown in blue and mutant profiles in black. (**C**) Real-time PCR of transcripts expressed from the enhancer reporter. cDNA samples were prepared as described for panel A and used in real-time PCR assays to specifically quantitate the relative abundances of spliced mRNA. Relative splicing activity (RSA).

inactivating mutations for fragment D. The HEXplorer is based on a RESCUE-type approach (31), calculating the different distributions of hexamer frequencies within introns versus exons. The profiles of genomic regions depict exonic enhancing and silencing properties, while HEXplorer score (HZ_{EI}) differences can assess mutational effects within SREs. Here, the sequence CATGGATGGAGCA was shown to have the longest contiguous HZ_{EI}-positive stretch, reflecting splicing enhancing properties (Figure 3B). In both the proximal and the distal parts of fragment D, we selected point mutations (5G>C, 8G>A) strongly decreasing the HZ_{EI}-positive area. The double mutation 5G>C/8G>A was predicted to maximally neutralize the enhancing properties of FGB7-D (Figure 3B). To examine this prediction, the mutations were tested within the eGFP enhancer reporter and inserted into FGB7-D upstream of the reporter 5'ss, and HeLa cells were transfected to monitor splice site activity using semi-quantitative RT-PCR. Indeed, the predicted mutations turned out to drastically impair the enhancing functionality of this fragment confirming its activity as another SRE within FGB exon 7 (Figure 3C).

Cryptic splice donor selection is highly dependent on each single SRE

On the basis of the above findings, we extended our analyses to determine whether the newly identified SRE was essential for splice site selection in the physiological exonic context. First, to test if the c.1244+1G>T mutation not only disrupted U1 snRNP binding but by itself might have created an SRE, we analyzed two additional splicing inactivating mutations within the three-exon minigene (c.1244+1G>A, c.1244+2T>A). However, only marginal differences in the splicing pattern could be observed, so that creation of a new SRE can be ruled out as the main cause for the observed splicing pattern (Supplementary Figure S2A). The slight increase in p1 usage for c.1244+1G>A and c.1244+2T>A was compatible with a formation of a moderate putative SRE located directly upstream of the exon/intron boundary (Supplementary Figure S2B). We therefore used the WT as well as the pathogenic FGB c.1244+1G>T three-exon minigenes for further analyses. To complete the picture, we performed HEXplorer-based mutational analyses for fragment C, but also B in order to compare HEXplorer-based inactivation to deletion of the naturally silent SRSF1 binding site (25) (Figure 4A). In agreement with Spena *et al.* (24), we did not observe any effect on cryptic 5'ss activation for an individual mutation as long as the physiological 5'ss was present (Figure 4B, lanes 1–4). Combining, however, either mutations within B and C (Figure 4B, lane 5) or all three parts at the same time (Figure 4B, lane 8), but not B and D (Figure 4B, lane 6) or C and D (Figure 4B, lane 7) resulted in activation of a cryptic 3'ss (Figure 4B, lane 5 and 8; Figure 4C, a2 (**)). This, however, could simply be explained by the accidental upregulation of this cryptic 3'ss (MaxEnt score from −6.23 to 2.39) located within C, and therefore also be present in the combined fragments B and C (Figure 4D). Aside from this, this cryptic 3′ss usage might also be supported by the changed sequence profile after HEXplorer-guided mutagenesis (Figure 4E). Indeed, the sequence environment preceding the AG is composed of a



3'ss a2**	Sequence	MaxEnt
WT	CAACAAATACCAGATCTCAGTGA	-6.23
B/C-MUT	TAACAAATACCATATCTCAGTGT	2.39

Figure 4. Splicing pattern of the FGB minigenes. (**A**) HEXplorer profiles of WT fragments B, C and D (blue) and mutant profiles (black). (**B**) RT-PCR analysis of splicing patterns of WT and c.1244+1G>T minigenes. Neutral sequence is CCAAACAA-repeat. 2.5×10^5 HeLa cells were transiently transfected with 1 μg of each construct and 1 μg of pXGH5. Twenty-four hours after transfection RNA was isolated and subjected to RT-PCR analysis using primer pairs #2648/#2649 and #1224/#1225 (hGH). PCR products were separated by 10% non-denaturing polyacrylamide gel electrophoresis and stained with ethidium bromide. (**C**) Positions of newly identified cryptic splice donor c0 and acceptor site a2** within FGB exon 7. (**D**) Sequences of the cryptic WT 3'ss a2** and the cryptic 3'ss generated upon mutation B/C-MUT, together with their MaxEnt scores. (**E**) HEXplorer profiles of FGB exon 7 of WT and B/C-MUT.

HZ_{EI}-negative stretch of hexamers reflecting intronic rather than exonic sequences (21).

As seen before, as soon as the physiological canonical 5'ss was rendered non-canonical (c.1244+1G>T), all cryptic splice sites c1, c2*, c3 and p1 were activated but still almost no exon skipping could be observed (Figure 4B, lane 9).

As expected, fragments B and C seemed to activate their proximal downstream splice donor c1. Strikingly, even mutating only one of these fragments completely abolished c1 donor usage and concomitantly enhanced exon skipping (Figure 4B, lanes 10 and 11), demonstrating that both fragments had to act in concert to activate c1. However, they did not differentially affect activation of c2* and c3, indicating that these two sites are independently regulated by another SRE upstream of both c2* and c3.

In agreement with the individual fragments' splicing regulatory activity (Figure 3A), changing the enhancing properties of D had the strongest effect on splice site selection, leading to an almost exclusive c1 donor usage and very little exon skipping, thereby shortening the exon (Figure 4B, lane 12). Further mutation of any combination of fragments drastically reduced exon 7 recognition (Figure 4B, lanes 13–16), and also activated the fourth exonic cryptic 5′ss c0 with an HBS of 9.4 (Figure 4B, lanes 13–16; Figure 4C). Since fragment A increased splice donor recognition 75-fold within the enhancer reporter (Figure 3A), it is likely that c0 was activated when there was no concurrent position-dependent inhibition by B or C.

Eventually, we inserted HEXplorer-guided point mutations into B instead of deleting B (25) to maintain constant exon length. Inactivating B by point mutations resulted in complete loss of c1 usage and an increase in exon skipping, whereas deleting fragment B only moderately impacted the splicing pattern (Supplementary Figure S3). This apparent discrepancy might be explained by the circumstances that the deletion brings fragments A and C in juxtaposition with each other, increasing the overall enhancing properties of this area.

We also treated WT and c.1244+1G>T mutant minigenes with the protein synthesis inhibitor CHX to examine if the observed mutation-induced splicing pattern also depended on NMD. However, as no difference in the splicing patterns could be observed, we exclude NMD as being responsible for the pattern of mutation-induced transcript isoforms (Supplementary Figure S4).

In summary, all four fragments (A–D) regulated both exon recognition and splice site selection by inhibiting upstream splice donor usage and simultaneously stimulating downstream splice donor usage. They were required to repress weak 5′ss along the way to the physiological 3′ exon end.

Variation of 5′ss complementarity systematically controls FGB exon 7 inclusion in the presence of various SREs

To examine the impact on exon recognition and splice site selection of a single SRE and the 5′ss it supports, we investigated splice site activation of fragments B, C and D individually. To this end, the FGB-7 sequence was fully substituted with neutral sequences maintaining only c1, c3 and c.1244+1G>T. Each fragment was then individually replaced back into this simplified splicing neutral exon at its physiological position either upstream of c1 or c3 (Figure 5A). Additionally, the HBS of c1 or c3 were stepwise increased to examine the interaction between the splice site proper and surrounding SREs.

Gradually increasing the HBS either for c1 from 12.2 or for c3 from 10.8 up to 20.8 led to a strong increase in splice site recognition in an otherwise fully splicing-neutral environment (Figure 5B–D, cf. lanes 1–4). In this particular neutral context, an HBS threshold of 18.8 was required for splice site recognition (Figure 5B–D, lane 3). Exon skipping, however, could not be totally eliminated even by increasing the HBS up to 20.8. As expected, inserting either B, C or D into this neutral exon substantially increased exon recognition and combined with an increase in complementarity of the supported splice site fully restored exon recognition (Figure 5B–D, cf. lanes 5–8). In the same way, the mutant versions of all individual fragments supported exon recognition much less, confirming the splicing enhancing activity of fragments B, C, D (Figure 5B–D, cf. lanes 5–8 with lanes 9–12). Furthermore, this experimental setting also allowed to estimate that e.g. fragment C contributed equally to exon recognition as an increase in splice site complementarity from HBS 15.8 to 20.8 (Figure 5C, cf. lanes 4 and 6).

Multiple SR proteins bind to FGB exon 7

To identify splicing regulatory proteins binding to RNA fragments B, C or D we performed RNA affinity purification assays, extending the work of Spena *et al*. on fragment B binding SRSF1 (25). We therefore incubated short WT or mutant sequence RNA oligonucleotides with HeLa nuclear extract (32). After several washing steps, the remaining specifically bound proteins were eluted, separated by SDS-PAGE and analyzed via mass spectrometry analysis. For B and C, 13 out of 14 SR protein abundance ratios 'mutant/WT' were below 1, indicating that a significant number of SR proteins showed higher affinities for WT sequences. For fragment B, 5 out of 6 SR protein intensities were lower in the mutant sequence, covering a range of 0.36—0.82 ($p = 0.04$, Fisher's exact test). For fragment C, all 8 SR protein intensities were lower in the mutant sequence, covering a range of 0.11–0.93 ($p = 0.0022$, Fisher's exact test) (Supplementary Table S2). For fragment D, a diverse picture emerged: 4 out of 8 SR protein intensities—including SRSF1—were lower in the mutant sequences (range 0.54–0.94, n.s.), although specifically Tra2β was found at slightly elevated levels of 1.14. In particular, SRSF1 showed the highest counts of unique peptides of all SR proteins and was decreased in all three mutant fragments.

For SRSF1 and Tra2β, we additionally carried out western blot analyses. As expected, not only control RNA B, but also both RNA oligo C and D were bound by SRSF1, and the respective mutations clearly impaired binding, confirming the mass spectrometry results. Tra2β was also bound to all individual WT oligos B–D, and the reduction in binding to mutant sequences was even more pronounced than for SRSF1 (Figure 6A). Analyzing the sequence composition of the individual fragments revealed

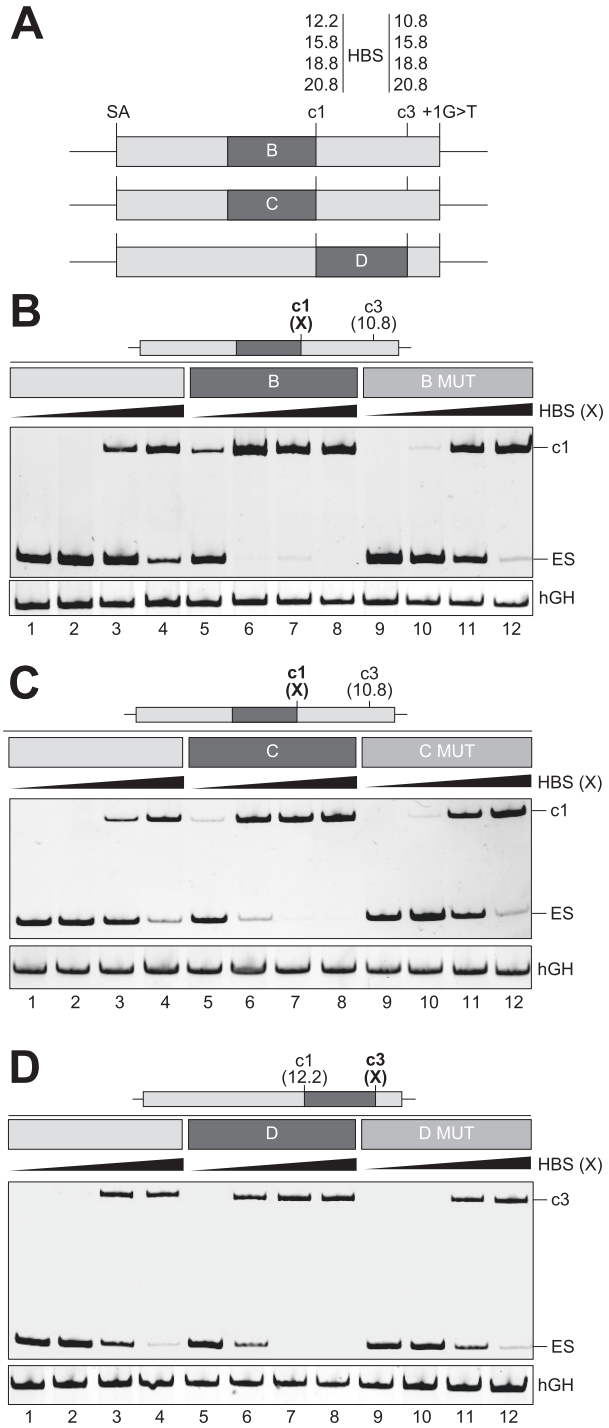


Figure 5. Impact of SREs on exon 7 recognition. (**A**) Schematic overview of the modified c.1244+1G>T minigene containing only neutral sequences (CCAAACAA-repeats, light gray boxes), c1, c3, c.1244+1G>T and either fragment B, C or D, respectively. Additionally, the HBS of c1 and c3 were stepwise increased to the values depicted above. (**B**–**D**) RT PCR analyses of the splicing pattern of the minigenes as shown in (A). HBS (X) at the right hand side of the black wedges above lanes 1–4, 5–8, 9–12 indicates for either cryptic site c1 or c3 (marked as bold and with X) increasing HBS values given in (A). 2.5×10^5 HeLa cells were transiently transfected with 1 μg of each construct and 1 μg of pXGH5. Cells were subjected to RT-PCRs using primer pairs #2648/#2649 and #1224/#1225 (hGH). PCR amplicons were separated on a non-denaturing 10% polyacrylamide gel and stained with ethidium bromide. Exon skipping (ES).

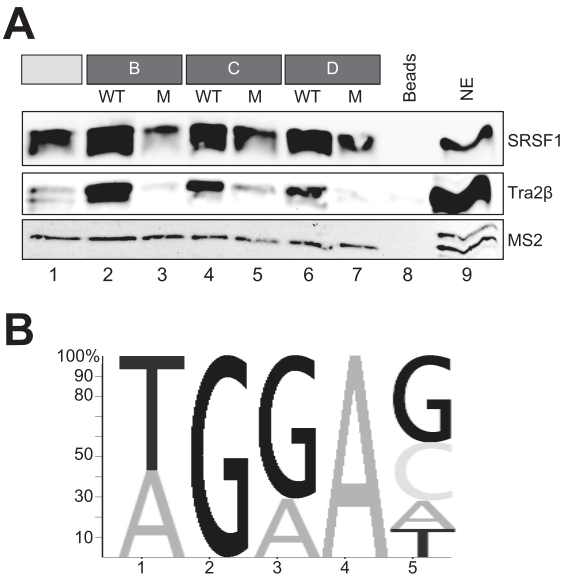


Figure 6. Western blot of SRSF1 and Tra2β binding to each fragment but not to the mutant variants. (**A**) RNAs including MS2 loops were immobilized using agarose beads, and analyzed for proteins binding by western blot. After the precipitated proteins have been resolved by SDS-PAGE (12%), specific antibodies directed against SRSF1 and Tra2β were used. MS2 coat protein added to the nuclear extract served as a loading control. (**B**) Sequence logo generated from a sequence motif generated by manual alignment of fragments A, B, C and D. The size of the letters reflects the relative frequency of the nucleotides at the position in the alignment.

that all were enriched in common purine-rich sequence motifs ((A/T)GGA; TGAA) (Figure 6B), previously shown to be bound by SRSF1 (33–35) or Tra2β (36–39).

Based on these results, FGB exon 7 seems to be regulated by multiple SR protein binding sites, for the most part by SRSF1, which regulates cryptic splice donor usage.

HEXplorer-guided mutations beyond FGB exon 7 induce cryptic splice site activation

In order to examine whether the splice site selection concept depending on both 5'ss complementarity and position dependent activity of up- and downstream SREs can be extended beyond FGB exon 7, we tested two examples outside the FGB gene. In particular, in both selected examples the competing U1 snRNA binding site and the WT splice site have similar U1 snRNA complementarity.

First, we revisited the well-documented pathogenic intronic SRE mutation (G to A substitution at +26, termed '759+26G>A') downstream of E1α PDH exon 7, which was found in a patient suffering from encephalopathy and lactic acidosis. This mutation has been shown to create a *de novo* SRSF2 binding site leading to activation of a cryptic splice site located within E1α PDH intron 7 (40,41). This cryptic splice site has even higher U1 snRNA complementarity (HBS 13.7) than the weak physiological 5'ss (HBS 12.2).

From E1α PDH, we derived the physiological 5'ss, the cryptic 5'ss and the intronic region in between containing the SRSF2 binding site, and inserted these into our FGB splicing neutral three-exon minigene. We furthermore inserted two copies of SRSF7 binding sites upstream of the

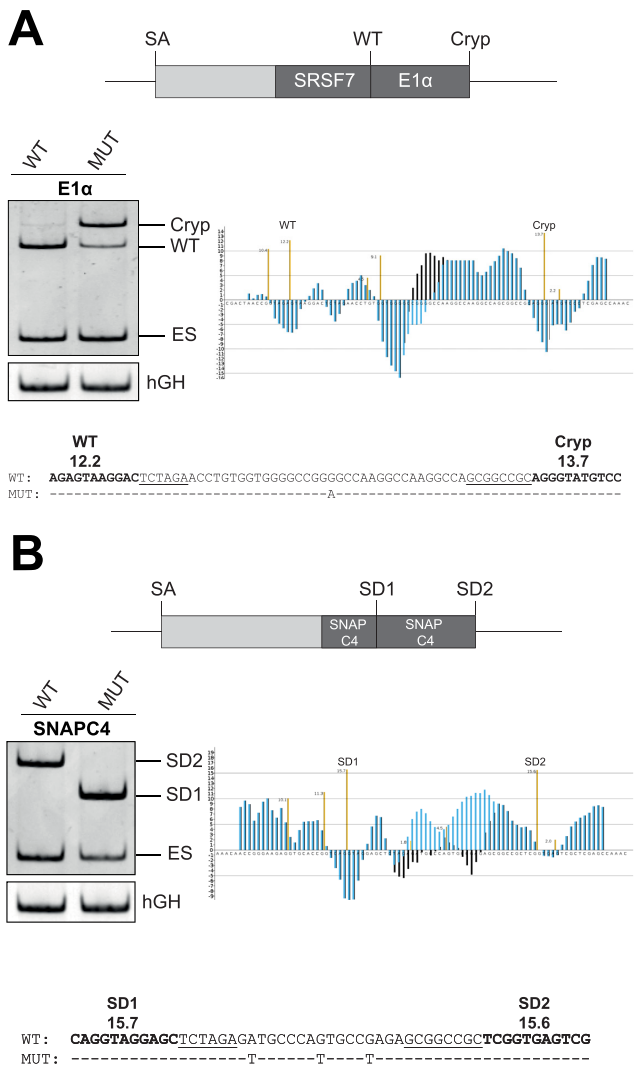


Figure 7. Extension of splice site selection concept from FGB exon 7 three-exon minigene to two other genes. (**A**) The middle exon (top) contains a fragment derived from the E1α PDH gene including the two corresponding splice donor sites (WT, Cryp) at their authentic positions, and else neutral sequences (CCAAACAA-repeats, light gray boxes) and two SRSF7 binding sites (61). HEXplorer profile of mutant sequence (black) shows stronger splice enhancing properties than WT (blue). 2.5×10^5 HeLa cells were transiently transfected with 1 μg of the construct and 1 μg of pXGH5. At 24 h after transfection, total-RNA samples were collected and used for RT-PCR with primer pair #2648/#2649 and normalized to hGH (#1224/#1225). (**B**) Same minigene containing two fragments derived from the SNAPC4 gene including the two corresponding splice donor sites at their authentic positions in between restriction sites (top). HEXplorer profiles of mutant sequence (black) shows weaker splice enhancing property than WT (blue). RT-PCR of transcripts amplified from the enhancer reporter (left hand side). Samples were prepared as described for panel A.

physiological 5'ss, since it has been shown that recognition of this rather weak E1α PDH splice donor is dependent on a strong SRE (41) (Figure 7A, top). In agreement with previous results, the 759+26G>A mutation led to a switch in 5'ss usage (wt to crypt, Figure 7A, left). Furthermore, HEXplorer analysis of the sequence between wt and crypt confirmed this observed phenotype: the pathogenic 759+26G>A mutation positively shifted the HEXplorer profile ($\Delta HZ_{EI} = 72$) indicative of increased downstream SRE activity (Figure 7A, right).

Second, we randomly selected an exon with an unused upstream U1 snRNA binding site of comparable complementarity (HBS 15.7) as the physiological 5'ss (HBS 15.6) from our fibroblast transcriptome dataset (see below): exon 12 in the SNAPC4 transcript (ENST00000298532.2). We inserted the following four segments into our FGB splicing neutral three-exon minigene: (i) upstream of the unused U1 snRNA binding site, to account for possible natural SRE context, (ii) exonic U1 snRNA binding site, (iii) physiological 5'ss, (iv) region between these sites (Figure 7B, top).

Following insertion of these SNAPC4 segments into the splicing reporter, we transfected HeLa cells and analyzed the splicing pattern confirming exclusive usage of the physiological 5'ss (wt; Figure 7B, left). HEXplorer-guided mutagenesis decreased the enhancing properties in the region between the U1 snRNA binding site and the 5'ss ($\Delta HZ_{EI} = -162$; Figure 7B, right), and splicing completely switched from the physiological 5'ss to the further upstream located U1 snRNA binding site (crypt; Figure 7B, left). Again, splice site usage seemed to be regulated by promoting downstream splice donor usage and simultaneously repressing upstream splice donor usage.

Taken together, we have confirmed our SRE dependent splice site selection concept in two examples beyond FGB exon 7: each had a pair of physiological 5'ss and U1 snRNA binding site with similar complementarity—one exonic and one intronic.

Can SREs explain 5'ss selection between GT sites of similar U1 snRNA complementarity?

We independently tested our 5'ss selection concept on individual pairs of a 5'ss and a nearby rarely used exonic U1 snRNA binding site with even higher complementarity, systematically selected from a dataset of 54 human RNA-Seq samples. These samples were derived from short term cultivated *in vivo* aged human dermal fibroblasts, collected from 30 healthy control subjects. Alignment with STAR (Ensembl 82) identified 2,050,307 multiply covered exon-exon junctions (E-MTAB-4652; Kaisers *et al*. PLoS One, in revision). From these, we selected exons with highly used canonical 5'ss (>10 000 exon junction reads). In this subset, we additionally selected only exons containing a U1 snRNA binding site within 35 nucleotides upstream that (i) had high complementarity (HBS > 14), (ii) had higher complementarity than the authentic 5'ss, but (iii) was silent (# reads < 1.3% of authentic 5'ss; median # reads 3). To allow for a putative SRE hexamer between the GT sites but not overlapping with either one, we furthermore required at least 17 (8+6+3) nucleotides in between. Application of these strict selection criteria left only 19 such exons from 19 different genes.

In each of these 19 hits, we scanned three separate regions of equal size for SREs: (i) upstream the silent GT-site, (ii) between silent GT-site and real 5'ss, (iii) downstream the authentic 5'ss. Regions (i) and (ii) each included the full 11 nucleotides of the silent GT-site and 5'ss. The size of these re-

gions was individually determined by the distance between the pair of silent GT-site and 5'ss.

We assessed these three regions for SREs using web resources for the common algorithms ESEfinder 3.0, RESCUE-ESE, FAS-ESS-hex3, PESX, ESRsearch and HEXplorer (31,42–47), applying respective default settings for SRE detection.

In order to find a semi-quantitative measure for the 'overall enhancer effect' in a given upstream exonic neighborhood of a GT-site, we first assigned a weight of +1 or −1 for each enhancer or silencer motif in this region predicted by any of the algorithms. To account for the direction dependence of enhancer action, we performed the same calculation for a mirror region of equal size downstream the GT-site and defined the *splice site enhancer weight* as the difference between sum of upstream weights and sum of downstream weights (Figure 8A). In this way, we treated exonic splicing enhancer (ESEs) occurring downstream of a GT-site as exonic splicing silencers with the same weight. This splice site enhancer weight was designed to capture both enhancing and silencing properties of equally sized regions up- and downstream of any GT-site, and its construction is analogous to the 'exonic splicing motif difference' defined by Ke *et al*. (48). In the same way (21), we calculated splice site weights using the ESRseq and HEXplorer algorithms as the average ESRseq and HEXplorer difference between the up- and downstream regions of any GT-site.

In the 19 genes containing exons with highly used 5'ss and nearby silent GT-sites with higher complementarity, all three splice site enhancer weights were significantly higher for authentic 5'ss than for silent GT-sites (enhancer/silencer weights $p = 0.009$, ESRseq $p = 0.006$, HEXplorer $p = 0.0002$; Figure 8B–D). Indeed, 17 out of 19 genes contained predicted SREs expected to repress U1 snRNA binding sites with even higher complementarity in favor of the authentic 5'ss, which is consistent with our 5'ss selection concept derived from the FGB exon 7 model system. These results suggest that this concept may not be limited to FGB exon 7.

DISCUSSION

In this study, we provide evidence for multiple SREs within the human FGB exon 7. Predictions obtained by both HEXplorer (21) and HBS (3,4), combined with the position-dependence of SREs (10), allow us a glimpse at understanding specific splicing outcomes of human pathogenic 5'ss mutations, based on the model exon FGB7. Here, we propose that, starting from the 3'ss, 5' splice site selection iteratively proceeds along an alternating sequence of U1 snRNA binding sites and interleaved SREs which can in principle support different 3' exon ends. Like in a relay race, SREs can either suppress a potential 5'ss and pass the splicing baton on or splicing actually occurs. This picture may permit generalization to a model for 5'ss selection and 3' exon end definition.

Binding of the 5' end of U1 snRNA to 5'ss initiates spliceosome formation. A higher base pair complementarity to U1 snRNA thereby supports splice site recognition to a higher degree (4,49). In many cases it has been shown that cryptic splice sites are significantly weaker than their authentic 5'ss counterparts (50). In line with this, assessment of the hydrogen bonding patterns of potential U1 snRNA binding sites (http://www.uni-duesseldorf.de/rna/html/hbond_score.php) within FGB exon 7 revealed that the authentic 5'ss had a higher HBS than the cryptic sites and that in turn the cryptic sites were stronger than the remaining potential U1 snRNA binding sites. Incidentally, we found the splice site p1 within the downstream intron, which was used in the presence of the c.1244+1G>T mutation and which seemed to be below detection limit in previous experiments but had an HBS comparable to c1 (25). However, even slightly different expression levels of splicing regulatory proteins under these different experimental conditions might have caused this difference.

Generally, mutations targeting an SRE may not only activate cryptic splice sites but can also lead to complete loss of exon recognition (17). We showed that either inserting SREs or strengthening the cryptic splice sites c1 and c3 within our simplified splicing neutral exon led to a gradual increase in exon recognition, which depended both on cryptic 5'ss HBS and support by splicing regulatory proteins.

By using an enhancer reporter, we identified multiple elements (A–D) within exon 7, which are each able to promote downstream splice donor usage. This is not surprising, since at least 3/4 of all nucleotides within a normal exon have been shown to be involved in splicing regulation (51). Furthermore, it could be shown that multiple enhancer elements increase the overall rate of splicing (10,52) which was attributed to the fact that more SREs improve the chance of an enhancer element promoting U1 snRNP binding to a 5'ss.

An important part in selecting a 5'ss in the presence of multiple simultaneously acting SREs is played by the strict position-dependency of splicing regulatory proteins. Plenty of work has shown that the same splicing regulatory proteins can activate splice donor usage from upstream positions as well as inhibit from downstream positions (18,53–55). Minigene analyses could show that the position-dependency seems to be a common mechanism of several SR and hnRNP proteins (10,11). This is in line with our findings showing that mutating fragment B and C upstream of c1 led to an impaired c1 donor usage, whereas mutating D upstream of c3 reduced c3 usage. However, mutating e.g. fragment D at the same time led to an upregulation of the upstream located splice donor site c1.

SR proteins are composed of one or two RRMs and one RS domain that participate directly in the interaction with other proteins or with the RNA itself. Until now, there is controversial data about the exact mechanism by which SR proteins promote splice site recognition. It has been shown that SRSF1 targets the U1-specific protein U1-70K to facilitate recruitment of the spliceosome to a splice donor site via RS-RS domain interactions (6,7). However, it also has been shown that the RS domain is not responsible for the interaction with U1 snRNP but rather the RRM (8,9). Notwithstanding, for the SR-related protein Tra2 it has been shown that splice site repression and activation occur via different effector domains (56). Therefore, it is tempting to speculate that SRSF1 acts in the same fashion; whether it is the RRM or the RS domain that is involved in activation or repression needs to be subject of closer examination. Fur-

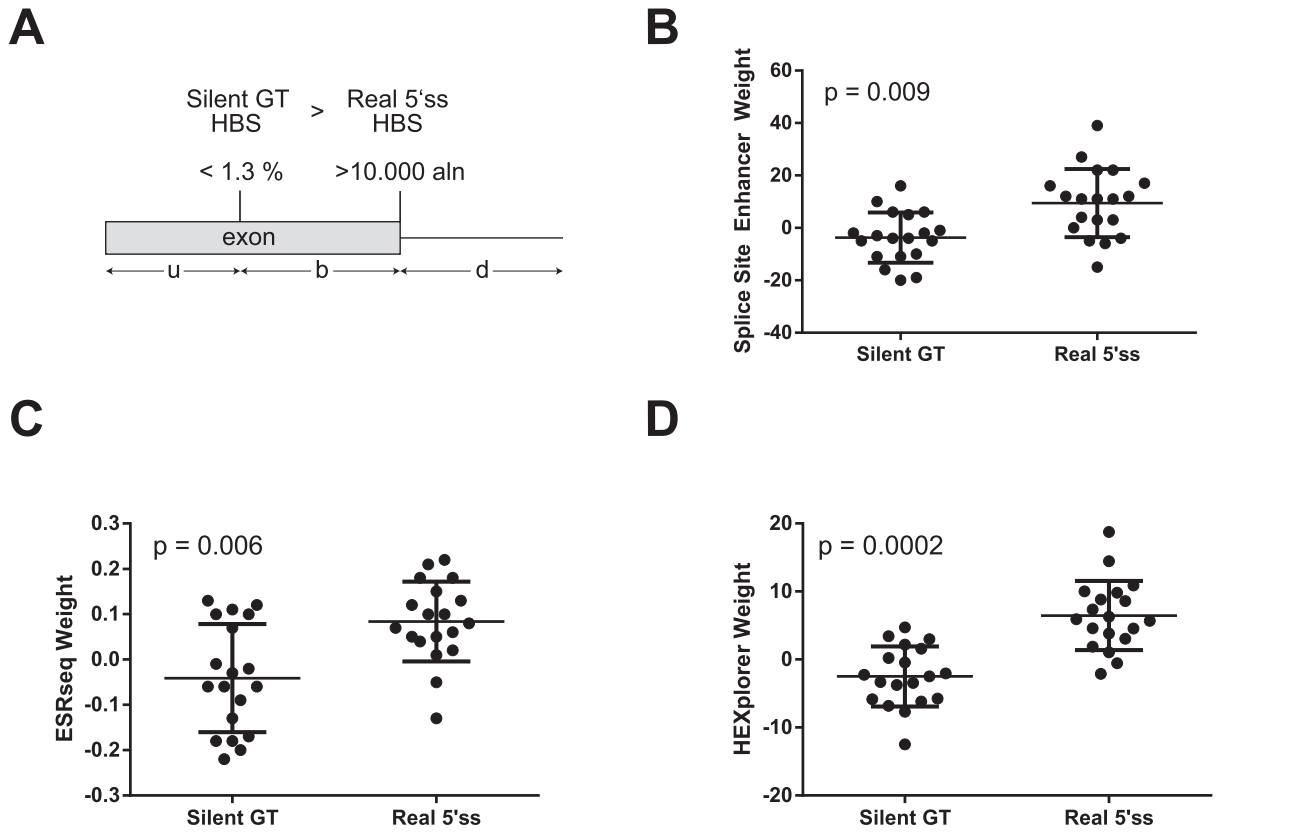


Figure 8. SREs support highly used 5'ss more than nearby silent GT-sites with similar or higher U1 snRNA complementarity. We screened 19 exons with pairs of a real 5'ss (>10 000 exon junction reads, *aln*) and a nearby silent exonic U1 snRNA binding site (HBS > 14, > 5'ss, <1.3% 5'ss *aln*) for splicing regulatory elements. (**A**) For each of these 19 exons, we scanned three separate regions of equal size: upstream the silent GT-site (u), between silent GT-site and real 5'ss (b), downstream the real 5'ss (d). (**B**) *Splice site enhancer weights*, (**C**) ESRseq weights and (**D**) HEXplorer weights, calculated as differences between sum of upstream weights and sum of downstream weights (A; 'u-b', 'b-d') were significantly higher for real 5'ss than for silent GT-sites (Wilcoxon signed rank test).

thermore, both functions might act simultaneously to inhibit the use of upstream cryptic splice sites. This perfectly matches with our results revealing that each SRE activated downstream donor usage but at the same time inhibited upstream located 5'ss.

Furthermore, it was assumed that SREs can be recognized by more than one SR protein (34) and that a purine-rich enhancer element can enhance splicing if bound by a protein complex (57). This is in accordance with our mass spectrometric analysis of proteins bound to SREs where we found a couple of SR proteins showing a selective higher abundance in wt sequence based affinity purifications. Pandit *et al*. (34) suggested that different kinds of SR proteins exist: some SR proteins like SRSF1 bind rather loosely to exonic positions, while others bind to more distinct binding motifs (58). It was additionally shown that SR proteins can directly interact with each other, like Tra2 and SRSF1 (59) and that Tra2 recruits other splicing factors to ESE sequences (60). From our data we cannot decide whether SR proteins bind directly to SREs or indirectly in complex with other SR proteins. Therefore, we can extend our hypothesis that the simultaneous inhibition or promotion of splice donor usage by SREs is facilitated by a set of splicing regulatory proteins which only together ensure specific binding as it has been shown for FUS and hnRNP H (11).

Aberrant splicing is one major cause of human genetic disease, and does not only involve mutations within splice sites but also within SREs, which makes computational mutation assessment within SREs highly important. All investigated elements within FGB exon 7 identified using the HEXplorer algorithm (21) could be experimentally validated as enhancer elements. Only recently, the HEXplorer has been proven a powerful tool to predict not only mutational effects within SREs but also their severity (15). Therefore, we propose that for a 'functional 5'ss usage prediction' both the splice site complementarity to U1 snRNA and in addition the sequence environment must be considered (21). The data presented in this work suggest a model in which SREs surrounding cryptic splice sites act in a strictly position-dependent manner, possibly supporting or antagonizing each other. This effect, however, is invisible in the presence of the strong authentic splice site (Figure 9, WT). As soon as the authentic splice site is disrupted, the SRE effects within FGB exon 7 become visible, leading to exon shortening. Even though c1 is supported by the SREs in B and C, fragment D lowers the enhancing potential for this cryptic splice site, letting the SREs in B and C appear

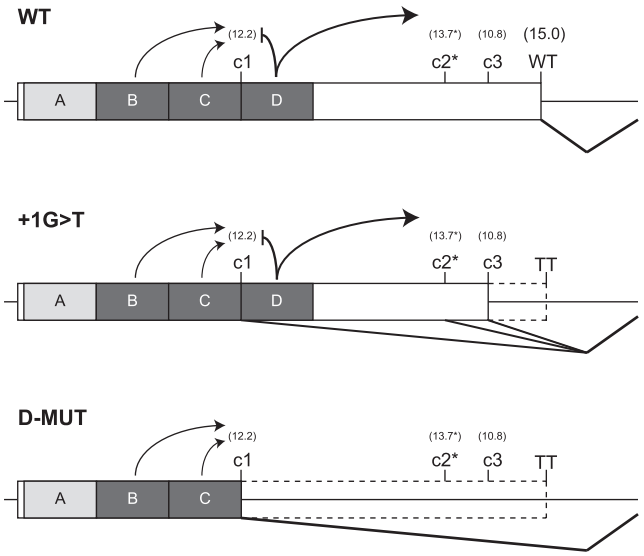


Figure 9. Model for FGB exon 7 recognition. Authentic splice donor usage is facilitated by the WT 5'ss with an HBS of 15.0 exceeding all cryptic splice sites and by the bi-directional properties of SREs that antagonize each other (B, C versus D) (WT). Inactivation of the WT 5'ss (GT>TT) leads to the usage of cryptic splice sites within exon 7, thereby shortening the exon (c.1244+1G>T) to either c1, c2* or c3. Mutation of fragment D shortens the exon exclusively to c1, as it is the only remaining SRE-supported splice site which is no longer repressed by the downstream localized SRE (D-MUT).

'naturally silent' and leading to activation of c1, c2* and c3 in comparable amounts (Figure 9, c.1244+1G>T). This is supported by the fact that mutating fragment D leads to a drastic increase in c1 and loss of c2* and c3 usage, further shortening the exon (Figure 9, D-MUT).

Moreover, we could substantiate our findings of SRE-dependent splice site selection by evaluating two further examples beyond FGB exon 7 in the same reporter system. Within both, E1α and SNAPC4, competing U1 snRNA binding sites exist that are regulated through position-dependent SREs. Mutational analyses could switch splicing in either direction, matching the calculated HEXplorer scores by interrupting or creating a sequence with exonic enhancer properties, respectively (Figure 7). Extending these examples of competing U1 snRNA binding sites, we specifically selected exons from 19 genes with highly used 5'ss downstream of nearby silent GT-sites with similar complementarity (average HBS difference 1.2) from our fibroblast RNA-Seq transcriptome dataset. This choice of exons permitted focusing on predicted SRE effects on 5'ss selection rather than 5'ss strength. Simultaneously taking potential SREs in regions both up- and downstream of a GT-site into account, as suggested by our extensive FGB exon 7 analyses, we found authentic, highly used 5'ss significantly more supported by SREs than silent GT-sites. This result was consistently found in analyses using eight different SRE identifying algorithms, indicating that the proposed concept may be more generally valid beyond our FGB exon7 model system.

This concept of potentially iterated position-dependent SRE action may highlight the important role of SREs not only in alternative splice site selection, but also as key for constitutive splicing in exons containing internal U1 snRNA binding sites that must be ignored to obtain an appropriate exon end.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Imke Meyer and Björn Wefers for excellent technical assistance and Philipp Peter for implementing the HEXplorer algorithm on our RNA website.

FUNDING

Deutsche Forschungsgemeinschaft (DFG) [SCHA 909/4-1]; Jürgen Manchot Stiftung (to A.L.B., L.W., L.H., H.S.); Stiftung für AIDS-Forschung, Düsseldorf (to H.S.). Funding for open access charge: Heinrich-Heine University.
Conflict of interest statement. None declared.

REFERENCES

1. Will,C.L. and Luhrmann,R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.*, **3**, a003707.
2. Schneider,M., Will,C.L., Anokhina,M., Tazi,J., Urlaub,H. and Luhrmann,R. (2010) Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol. Cell*, **38**, 223–235.
3. Kammler,S., Leurs,C., Freund,M., Krummheuer,J., Seidel,K., Tange,T.O., Lund,M.K., Kjems,J., Scheid,A. and Schaal,H. (2001) The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. *RNA*, **7**, 421–434.
4. Freund,M., Asang,C., Kammler,S., Konermann,C., Krummheuer,J., Hipp,M., Meyer,I., Gierling,W., Theiss,S., Preuss,T. *et al.* (2003) A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res.*, **31**, 6963–6975.
5. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
6. Wu,J.Y. and Maniatis,T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75**, 1061–1070.
7. Kohtz,J.D., Jamison,S.F., Will,C.L., Zuo,P., Luhrmann,R., Garcia-Blanco,M.A. and Manley,J.L. (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, **368**, 119–124.
8. Xiao,S.H. and Manley,J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.*, **11**, 334–344.
9. Cho,S., Hoang,A., Sinha,R., Zhong,X.Y., Fu,X.D., Krainer,A.R. and Ghosh,G. (2011) Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 8233–8238.
10. Erkelenz,S., Mueller,W.F., Evans,M.S., Busch,A., Schoneweis,K., Hertel,K.J. and Schaal,H. (2013) Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA*, **19**, 96–102.
11. Reber,S., Stettler,J., Filosa,G., Colombo,M., Jutzi,D., Lenzken,S.C., Schweingruber,C., Bruggmann,R., Bachi,A., Barabino,S.M. *et al.* (2016) Minor intron splicing is regulated by FUS and affected by ALS-associated FUS mutants. *EMBO J.*, **35**, 1504–1521.
12. Domsic,J.K., Wang,Y., Mayeda,A., Krainer,A.R. and Stoltzfus,C.M. (2003) Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts. *Mol. Cell. Biol.*, **23**, 8762–8772.

13. Sharma,S., Kohlstaedt,L.A., Damianov,A., Rio,D.C. and Black,D.L. (2008) Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat. Struct. Mol. Biol.*, **15**, 183–191.
14. Hartmann,L., Theiss,S., Niederacher,D. and Schaal,H. (2008) Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front. Biosci.*, **13**, 3252–3272.
15. Soukarieh,O., Gaildrat,P., Hamieh,M., Drouet,A., Baert-Desurmont,S., Frebourg,T., Tosi,M. and Martins,A. (2016) Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools. *PLoS Genet.*, **12**, e1005756.
16. Daguenet,E., Dujardin,G. and Valcarcel,J. (2015) The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO Rep.*, **16**, 1640–1655.
17. Krawczak,M., Thomas,N.S., Hundrieser,B., Mort,M., Wittig,M., Hampe,J. and Cooper,D.N. (2007) Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.*, **28**, 150–158.
18. Lim,K.H., Ferraris,L., Filloux,M.E., Raphael,B.J. and Fairbrother,W.G. (2011) Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11093–11098.
19. Sterne-Weiler,T., Howard,J., Mort,M., Cooper,D.N. and Sanford,J.R. (2011) Loss of exon identity is a common mechanism of human inherited disease. *Genome Res.*, **21**, 1563–1571.
20. Yeo,G. and Burge,C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
21. Erkelenz,S., Theiss,S., Otte,M., Widera,M., Peter,J.O. and Schaal,H. (2014) Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. *Nucleic Acids Res.*, **42**, 10681–10697.
22. Asang,C., Hauber,I. and Schaal,H. (2008) Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. *Nucleic Acids Res.*, **36**, 1450–1463.
23. Zhang,X.H., Arias,M.A., Ke,S. and Chasin,L.A. (2009) Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *RNA*, **15**, 367–376.
24. Spena,S., Duga,S., Asselta,R., Malcovati,M., Peyvandi,F. and Tenchini,M.L. (2002) Congenital afibrinogenemia: first identification of splicing mutations in the fibrinogen Bbeta-chain gene causing activation of cryptic splice sites. *Blood*, **100**, 4478–4484.
25. Spena,S., Tenchini,M.L. and Buratti,E. (2006) Cryptic splice site usage in exon 7 of the human fibrinogen B beta-chain gene is regulated by a naturally silent SF2/ASF binding site within this exon. *RNA*, **12**, 948–958.
26. Selden,R.F., Howie,K.B., Rowe,M.E., Goodman,H.M. and Moore,D.D. (1986) Human growth hormone as a reporter gene in regulation studies employing transient gene expression. *Mol. Cell. Biol.*, **6**, 3173–3179.
27. Chomczynski,P. and Sacchi,N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.
28. Lareau,L.F., Inada,M., Green,R.E., Wengrod,J.C. and Brenner,S.E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, **446**, 926–929.
29. Poschmann,G., Seyfarth,K., Besong Agbo,D., Klafki,H.W., Rozman,J., Wurst,W., Wiltfang,J., Meyer,H.E., Klingenspor,M. and Stuhler,K. (2014) High-fat diet induced isoform changes of the Parkinson's disease protein DJ-1. *J. Proteome Res.*, **13**, 2339–2351.
30. Caputi,M., Freund,M., Kammler,S., Asang,C. and Schaal,H. (2004) A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. *J. Virol.*, **78**, 6517–6526.
31. Fairbrother,W.G., Yeh,R.F., Sharp,P.A. and Burge,C.B. (2002) Predictive identification of exonic splicing enhancers in human genes. *Science*, **297**, 1007–1013.
32. Erkelenz,S., Hillebrand,F., Widera,M., Theiss,S., Fayyaz,A., Degrandi,D., Pfeffer,K. and Schaal,H. (2015) Balanced splicing at the Tat-specific HIV-1 3'ss A3 is critical for HIV-1 replication. *Retrovirology*, **12**, 29.
33. Anczukow,O., Akerman,M., Clery,A., Wu,J., Shen,C., Shirole,N.H., Raimer,A., Sun,S., Jensen,M.A., Hua,Y. *et al.* (2015) SRSF1-regulated alternative splicing in breast cancer. *Mol. Cell*, **60**, 105–117.
34. Pandit,S., Zhou,Y., Shiue,L., Coutinho-Mansfield,G., Li,H., Qiu,J., Huang,J., Yeo,G.W., Ares,M. Jr and Fu,X.D. (2013) Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol. Cell*, **50**, 223–235.
35. Ray,D., Kazan,H., Chan,E.T., Pena Castillo,L., Chaudhry,S., Talukder,S., Blencowe,B.J., Morris,Q. and Hughes,T.R. (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat. Biotechnol.*, **27**, 667–670.
36. Tacke,R., Tohyama,M., Ogawa,S. and Manley,J.L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, **93**, 139–148.
37. Grellscheid,S., Dalgliesh,C., Storbeck,M., Best,A., Liu,Y.L., Jakubik,M., Mende,Y., Ehrmann,I., Curk,T., Rossbach,K. *et al.* (2011) Identification of evolutionarily conserved exons as regulated targets for the splicing activator Tra2 beta in development. *PLoS Genet.*, **7**, e1002390.
38. Tsuda,K., Someya,T., Kuwasako,K., Takahashi,M., He,F., Unzai,S., Inoue,M., Harada,T., Watanabe,S., Terada,T. *et al.* (2011) Structural basis for the dual RNA-recognition modes of human Tra2-beta RRM. *Nucleic Acids Res.*, **39**, 1538–1553.
39. Erkelenz,S., Poschmann,G., Theiss,S., Stefanski,A., Hillebrand,F., Otte,M., Stuhler,K. and Schaal,H. (2013) Tra2-mediated recognition of HIV-1 5' splice site D3 as a key factor in the processing of vpr mRNA. *J. Virol.*, **87**, 2721–2734.
40. Mine,M., Brivet,M., Touati,G., Grabowski,P., Abitbol,M. and Marsac,C. (2003) Splicing error in E1alpha pyruvate dehydrogenase mRNA caused by novel intronic mutation responsible for lactic acidosis and mental retardation. *J. Biol. Chem.*, **278**, 11768–11772.
41. Gabut,M., Mine,M., Marsac,C., Brivet,M., Tazi,J. and Soret,J. (2005) The SR protein SC35 is responsible for aberrant splicing of the E1alpha pyruvate dehydrogenase mRNA in a case of mental retardation with lactic acidosis. *Mol. Cell. Biol.*, **25**, 3286–3294.
42. Cartegni,L., Wang,J., Zhu,Z., Zhang,M.Q. and Krainer,A.R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.*, **31**, 3568–3571.
43. Zhang,X.H. and Chasin,L.A. (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.*, **18**, 1241–1250.
44. Smith,P.J., Zhang,C., Wang,J., Chew,S.L., Zhang,M.Q. and Krainer,A.R. (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum. Mol. Genet.*, **15**, 2490–2508.
45. Wang,Z., Rolish,M.E., Yeo,G., Tung,V., Mawson,M. and Burge,C.B. (2004) Systematic identification and analysis of exonic splicing silencers. *Cell*, **119**, 831–845.
46. Goren,A., Ram,O., Amit,M., Keren,H., Lev-Maor,G., Vig,I., Pupko,T. and Ast,G. (2006) Comparative analysis identifies exonic splicing regulatory sequences–The complex definition of enhancers and silencers. *Mol. Cell*, **22**, 769–781.
47. Ke,S., Shang,S., Kalachikov,S.M., Morozova,I., Yu,L., Russo,J.J., Ju,J. and Chasin,L.A. (2011) Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res.*, **21**, 1360–1374.
48. Ke,S., Zhang,X.H. and Chasin,L.A. (2008) Positive selection acting on splicing motifs reflects compensatory evolution. *Genome Res.*, **18**, 533–543.
49. Freund,M., Hicks,M.J., Konermann,C., Otte,M., Hertel,K.J. and Schaal,H. (2005) Extended base pair complementarity between U1 snRNA and the 5' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5' splice site recognition. *Nucleic Acids Res.*, **33**, 5112–5119.
50. Roca,X., Sachidanandam,R. and Krainer,A.R. (2003) Intrinsic differences between authentic and cryptic 5 ' splice sites. *Nucleic Acids Res.*, **31**, 6321–6333.
51. Chasin,L.A. (2007) Searching for splicing motifs. *Adv. Exp. Med. Biol.*, **623**, 85–106.
52. Hertel,K.J. and Maniatis,T. (1998) The function of multisite splicing enhancers. *Mol. Cell*, **1**, 449–455.
53. Cereda,M., Pozzoli,U., Rot,G., Juvan,P., Schweitzer,A., Clark,T. and Ule,J. (2014) RNAmotifs: prediction of multivalent RNA motifs that control alternative splicing. *Genome Biol.*, **15**, R20.

54. Ule,J., Stefani,G., Mele,A., Ruggiu,M., Wang,X., Taneri,B., Gaasterland,T., Blencowe,B.J. and Darnell,R.B. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**, 580–586.
55. Llorian,M., Schwartz,S., Clark,T.A., Hollander,D., Tan,L.Y., Spellman,R., Gordon,A., Schweitzer,A.C., de la Grange,P., Ast,G. *et al.* (2010) Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat. Struct. Mol. Biol.*, **17**, 1114–1123.
56. Shen,M. and Mattox,W. (2012) Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. *Nucleic Acids Res.*, **40**, 428–437.
57. Yeakley,J.M., Morfin,J.P., Rosenfeld,M.G. and Fu,X.D. (1996) A complex of nuclear proteins mediates SR protein binding to a purine-rich splicing enhancer. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 7582–7587.
58. Anko,M.L., Muller-McNicoll,M., Brandl,H., Curk,T., Gorup,C., Henry,I., Ule,J. and Neugebauer,K.M. (2012) The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.*, **13**, R17.
59. Amrein,H., Hedley,M.L. and Maniatis,T. (1994) The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. *Cell*, **76**, 735–746.
60. Lynch,K.W. and Maniatis,T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes Dev.*, **10**, 2089–2101.
61. Cavaloc,Y., Bourgeois,C.F., Kister,L. and Stevenin,J. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA*, **5**, 468–483.

4. Analysis of competing HIV-1 splice donor sites uncovers a tight cluster of splicing regulatory elements within exon 2/2b

The following data are submitted for publication to Journal of Virology (MS ID: JVI00389-17) by

***Brillen A.L.**, Walotka L., Hillebrand F., Müller L., Widera M., Theiss S., Schaal H.*

Contribution

ALB and HS conceived the study and designed the experiments. ALB, LW, LM and MW performed cloning, transfection experiments and (q)RT-PCR analyses. ALB performed RNA-pull-down analyses. FH performed LNA-related experiments. ST, ALB and HS performed HEXplorer analyses. ST provided statistical analyses. ALB, ST and HS wrote the manuscript.

Abstract

Abstract: The HIV-1 accessory protein Vif is essential for viral replication by counteracting the host restriction factor APOBEC3G (A3G), and balanced levels of both proteins are required for efficient viral replication. Non-coding exons 2/2b contain the Vif start codon between their alternatively used 5'ss D2 and D2b. For the vif mRNA, intron 1 must be removed, while intron 2 must be retained. Thus, 3'ss A1 must be activated by U1 snRNP binding to either D2 or D2b, while splicing at D2 or D2b must be prevented. Here, we unravel the complex interaction between previously known and novel constituents of the splicing regulatory network regulating HIV-1 exon 2/2b inclusion into viral mRNAs. In particular, using RNA pulldown experiments and MS analysis we found members of the hnRNP A/B family binding to the novel SRE ESS2b, and Tra2/SRSF10 binding to nearby ESE2b. Splice site selection between D2 and D2b

was examined in a minigene reporter by HEXplorer guided mutational analysis of the identified SREs. Furthermore, the impact of these SREs on the viral splicing pattern and protein expression was exhaustively analyzed in viral particle production and replication experiments. Masking of these protein binding sites by usage of locked nucleic acids (LNAs) impaired Vif expression and viral replication. Importance: Based on our results, we propose a model in which a dense network of SREs regulates vif mRNA and protein expression, crucial to maintain viral replication within host cells with varying A3G levels and at different stages of infection. This regulation is maintained by several SR and hnRNP proteins binding to those elements. Targeting this cluster of SREs with LNAs may lead to the development of novel effective therapeutic compounds.

Analysis of competing HIV-1 splice donor sites uncovers a tight cluster of splicing regulatory elements within exon 2/2b

Anna-Lena Brillen[1], Lara Walotka[1], Frank Hillebrand[1], Lisa Müller[1], Marek Widera[1,2], Stephan Theiss[3], and Heiner Schaal[1,#]

[1]Heinrich-Heine-University Düsseldorf, Institute of Virology, Düsseldorf, Germany

[2]University Hospital, University of Duisburg-Essen, Institute of Virology, Essen, Germany

[3]Heinrich-Heine-University Düsseldorf, Institute of Clinical Neuroscience and Medical Psychology, Düsseldorf, Germany

Corresponding address: schaal@uni-duesseldorf.de

Word count complete abstract: 205

Word count text: 5754

Abstract

Abstract The HIV-1 accessory protein Vif is essential for viral replication by counteracting the host restriction factor APOBEC3G (A3G), and balanced levels of both proteins are required for efficient viral replication. Non-coding exons 2/2b contain the Vif start codon between their alternatively used 5'ss D2 and D2b. For the *vif* mRNA, intron 1 must be removed, while intron 2 must be retained. Thus, 3'ss A1 must be activated by U1 snRNP binding to either D2 or D2b, while splicing at D2 or D2b must be prevented.

Here, we unravel the complex interactions between previously known and novel components of the splicing regulatory network regulating HIV-1 exon 2/2b inclusion into viral mRNAs. In particular, using RNA pulldown experiments and MS analysis we found members of the hnRNP A/B family binding to the novel SRE ESS2b, and Tra2/SRSF10 binding to nearby ESE2b. Using a minigene reporter, we performed HEXplorer guided mutational analysis to narrow down SRE motifs affecting splice site selection between D2 and D2b. Eventually, the impact of these SREs on the viral splicing pattern and protein expression was exhaustively analyzed in viral particle production and replication experiments. Masking of these protein binding sites by usage of locked nucleic acids (LNAs) impaired Vif expression and viral replication.

Importance Based on our results, we propose a model in which a dense network of SREs regulates *vif* mRNA and protein expression, crucial to maintain viral replication within host cells with varying A3G levels and at different stages of infection. This regulation is maintained by several SR and hnRNP proteins binding to those elements. Targeting this cluster of SREs with LNAs may lead to the development of novel effective therapeutic strategies.

Introduction

During LTR (long terminal repeat)-driven transcription, over 50 mRNA isoforms emerge by alternative splicing of the HIV-1 precursor mRNA (1, 2). According to their distinct sizes, mRNA isoforms can be divided into three different classes: 2kb mRNAs (intronless), encoding for Tat, Rev and Nef, intron-containing 4kb mRNAs, encoding for Vif, Vpr, Vpu and Env, and 9kb unspliced mRNAs, encoding Gag and Gag-Pol (3). Viral gene expression underlies a strict chronological order (4-6). In the early phase, only

intronless mRNAs are transported out of the nucleus and translated, whereas intron-containing 4kb and 9kb mRNAs depend on the accumulation of Rev protein, which facilitates their export into the cytoplasm in the later phase.

Primarily responsible for the vast amount of mRNA isoforms are four splice donor sites (D1—D4, 5'ss), eight splice acceptor sites (A1—A7, 3'ss) and several only rarely used sites like splice donor 2b (D2b, (1-3)). Their recognition depends on intrinsic strength as well as *cis*-acting splicing regulatory elements (SREs) bound by e.g. SR or hnRNP proteins (7).

Splicing itself is a highly regulated process, controlled by several components of the spliceosomal complex. It starts with U1 snRNP binding to the 5'ss, followed by U2 snRNP binding to the branch point sequence of the upstream 3'ss (8). U1 and U2 snRNPs pair in a process named "exon-definition" (9), which is later transformed into an "intron-definition" process (10, 11), in which U1 and U2 snRNPs couple across the intron and thereby initiate the splicing reaction. SR or hnRNP proteins can support U1 snRNP binding to a splice donor, depending on their exonic or intronic position (12).

Up to this time, many SREs have been identified within the pre-mRNA of HIV-1 (Fig. 1A). Only recently, five novel SREs could be identified using the HEXplorer algorithm (13). This algorithm reflects potential enhancing and silencing properties of hexamers in the neighbourhood of a 5'ss. Any disruption of a splice site or an SRE can lead to a profound weakening of viral replication (14). Exclusively within HIV-1 exon 2 and 2b, six different SREs have already been described (Fig. 1B). Within exon 2, the SRSF1-dependent exonic splicing enhancers (ESEs) M1 and M2 (15) as well as the SRSF4-dependent ESE-Vif (16) have been shown to activate D2, whereas two G runs suppress exon 2/2b inclusion (16, 17). Furthermore, a novel HEXplorer-identified SRE within exon

2b (ESE$^{5005\text{-}5032}$, from now on called ESE2b), was shown to activate downstream splice donor usage within minigene analysis (13).

In addition to 3'ss A1 recognition and removal of the most 5'-proximal intron, use of the downstream 5´ss sites must be prevented to result in the formation of *vif* mRNAs. Downstream splice donor sites D2 and D2b, however, have to be recognized by U1 snRNP to activate A1, but rendered splicing incompetent to maintain the *vif* open reading frame (ORF) whose start codon lies within the downstream intron of D2 (17).

Vif is a low abundant, 23 kDa small protein that is incorporated into newly assembling virions. Vif counteracts the host restriction factor APOBEC3G (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G; A3G) (18), that is also encapsidated into virus particles and primarily triggers G-to-A hypermutations in the viral genome during reverse transcription. Vif binds to A3G to provoke ubiquitination and proteosomal degradation. Although Vif is absolutely essential for an efficient HIV-1 replication in A3G-expressing cells, excessive Vif is equally deleterious, since massive levels of Vif inhibit proteolytic Gag processing (19).

In the present study, we focused on the functional importance of splicing regulatory elements within exon 2/2b. On the basis of our results, we provide evidence that multiple SREs within exon 2/2b tightly regulate proper *vif* mRNA production. We could underline the functional importance of ESE2b, bound by Tra2 and SRSF10, and the newly discovered ESS2b, bound by hnRNP A/B proteins, for splice donor usage and exon recognition. Point mutations within those SREs predicted via the HEXplorer algorithm as well as LNA masking altered both viral *vif* mRNA and Vif protein amounts by regulating exon 2/2b inclusion, and led to a drop in viral particle production.

Results

Tra2 and SRSF10 act via ESE2b to activate the downstream-located 5'ss D2b

To understand splice site selection critical for HIV-1 *vif* mRNA formation, we focused on the exonic 2b region downstream of splice site D2 (Fig. 1). The *vif* start codon is localized upstream of an alternative splice site, termed D2b, which defines the 3' end of exon 2b but needs to be repressed to retain the downstream intronic sequence coding for Vif. Previously, we have shown that D2b is repressed by a conserved immediately upstream located G run (G_{I2}-1) which is bound by hnRNP F/H (17). As inactivating G_{I2}-1 led to an upregulation of this intrinsically rather weak splice donor D2b, we hypothesized that G_{I2}-1 not only represses D2b but might additionally shield an upstream bound SR protein from activating D2b (17). This assumption was further supported by the observation that, in the presence of multiple exonic SREs, the SRE closest to the 5'ss likely dominates splicing decisions (12). Therefore, we tested the region between D2 and D2b for splice site enhancing properties, and split the region into four overlapping segments indicated in Fig. 1B. To test the segments for 5'ss enhancing properties, we used an HIV-1 subgenomic reporter, which allows monitoring of SRE-mediated U1 snRNP binding to 5'ss SD4, forming an eGFP encoding mRNA by splicing to 3'ss A7 (20-22) (Fig. 1C, top). Following transient transfection, fluorescence microscopy allowed a first rough estimation of enhancing properties in the four exonic 2b segments. We used the sequence CCAAACAA (23) as a splicing neutral reference and the very strongly enhancing SRE HIV-1 GAR H fragment as a positive control (20, 22). As expected, fragment IV covering G_{I2}-1 did not support downstream 5'ss usage, while ESE2b ($ESE^{5005\text{-}5032}$ (13)) contained in fragment III enhanced D2b usage. Neither fragment I nor II led to an increased eGFP expression (Fig. 1C, bottom), demonstrating

that ESE2b was the only SRE in the 3' part of exon 2b capable of supporting downstream 5'ss usage.

To identify splicing regulatory proteins binding to ESE2b, we made use of the previously published inactivating nucleotide substitutions predicted by the HEXplorer algorithm ($ESE2b^{MUT}$ (ΔHZ_{EI} -267), termed "5015A>T; 5025A>T (dm)" in (13)). We performed RNA affinity purification assays with RNA oligonucleotides containing either the ESE2b or the $ESE2b^{MUT}$ sequence. After coupling to agarose beads, the oligonucleotides were incubated with HeLa nuclear extract. After washing and elution, bound proteins were analyzed via mass-spectrometry (MS). Besides weak binding to several members of the SR protein family, we found a significant loss of the proteins Tra2α, Tra2β and SRSF10 in the mutant ESE2b sequence and no significant change in the level of any hnRNP protein (Table 1, P = 0.05, t-test).

SREs between D2 and ESE2b are necessary for maintaining splicing at D2

To test the impact of ESE2b on D2/D2b splice donor selection, we used a heterologous three-exon minigene splicing reporter (Fig. 2A), previously shown to be suitable to dissect the role of *cis*-acting SREs in splice site decisions even in complex splicing networks (24). Within this splicing reporter, the artificial internal exon was not recognized at all when completely composed of splicing neutral sequences (23), but could be exonized upon replacing neutral sequences by *cis*-acting SREs or increasing the U1 snRNA complementarity of its 5'ss (HBS > 15.8; (24)). When we inserted both viral 5' splice sites, D2 (HBS 10.7) and D2b (12.4), into this context of neutral sequences, this exon was not recognized even though it is bordered by an intrinsically strong SA (MaxEnt score 10.25, Fig. 2B, lane 1). To recapitulate HIV-1 exon 2 splice site recognition, all known exon 2 localized SREs (herein for simplicity collectively referred to

as ESE2: ESE-Vif (16), M1 and M2 (15), the GGGG motif (16), as well as ESE2b and G_{I2}-1 were inserted either individually or in combination into this exon at their authentic positions either upstream or downstream of D2 (Fig. 2A).

Replacing corresponding neutral sequences with ESE2 alone comparably activated D2 and D2b (Fig. 2B, cf. lanes 1, 2). Additionally substituting neutral sequences with ESE2b and G_{I2}-1 switched splice site selection to almost exclusive D2b rather than D2 usage (Fig. 2B, lane 3), indicating that ESE2b did not only strongly support D2b selection, overriding the repressive G_{I2}-1 activity, but at the same time blocked the upstream localized D2. Even though D2b has a higher complementarity to U1 snRNA than D2 (HBS 12.4 vs. 10.7), in the viral context it is used rarely: 0.2% relative to 5.3% D2 usage (17). To examine the impact of ESE2b variants on 5'ss selection, we tested two ESE2b mutations that reduced its splice enhancing activity (WT > ΔHZ_{EI}-94 > ΔHZ_{EI}-267) (Fig. 2C).

As shown in (Fig. 2B), a stepwise switch towards D2 usage occurred when we reduced the ESE2b HEXplorer score by 2-nt-mutations, thus weakening its splice enhancing activity (Fig. 2B, cf. lanes 3-5). This D2b-to-D2 transition occurred both with intact and inactivated G_{I2}-1 (Fig. 2B, cf. lanes 6-8), but in the latter case a larger reduction in ESE2b splice enhancing activity was required to switch to D2 selection. Thus, splicing occurred at the weaker upstream 5'ss D2, if the combined splice enhancing property of ESE2b and G_{I2}-1 did not suffice to move splice site selection to downstream located D2b.

So far, however, HIV-1 D2 usage as in the viral context could not yet be mimicked with this minigene indicating that there may be an additional *cis*-acting element in the viral sequence. Such an SRE, localized between D2 and ESE2b might act like an "insulator"

166 separating the ESE2 from ESE2b activities. Therefore, we profiled exon 2b for further
167 enhancing and silencing properties of splice site neighborhoods. A region showing
168 predominantly upstream enhancing and downstream silencing properties was located
169 directly downstream of splice donor D2 (Fig. 3A, top panel). This region includes four
170 subsequent peaks (A, B, C, D) of the HEXplorer profile. We then substituted either the
171 whole fragment (A—D) or individual fragments (A, B, C, D) for neutral sequences of the
172 same length in the minigene reporter. After RT-PCR analysis, it became obvious that
173 indeed the region from A through D reversed splice site selection from D2b back to the
174 native HIV-1 splice site D2, which is more frequently used in the viral context (Fig. 3B,
175 cf. lanes 1 and 2). Further analyses of the individual fragments demonstrated that
176 fragments C and D rather than fragment A or B affected splice site choice (Fig. 3B, lanes
177 3-6). However, as neither fragment C nor D on its own was sufficient to fully induce the
178 splice site switch, we concluded that the potential SRE spanned both fragments, and
179 termed it ESS2b. To examine our hypothesis, we specifically changed ESS2b by
180 HEXplorer-guided mutagenesis in fragments C, D or both in the context of A—D (Fig.
181 3A, bottom 3 panels). Analysis of the splicing pattern (Fig. 3C) revealed that mutating
182 either C or D led to a partial splice site switch, whereas simultaneously mutating C and
183 D showed the same splicing phenotype as the neutral sequence (Fig. 3C, cf. lanes 1
184 and 5). These results demonstrate that ESS2b spans C and D, enhances upstream D2
185 and represses downstream D2b recognition even in the presence of downstream
186 ESE2b. Next, to identify splicing regulatory proteins binding to ESS2b, we again
187 performed RNA affinity purification with WT and mutant sequences as described above.
188 Subsequent MS analysis revealed that besides hnRNP DL binding, especially members
189 of the hnRNP A/B family (hnRNP AB family includes isoforms A1, A2/B1, A3 and A0)

were markedly enriched in the WT compared to the mutant sample, whereas, in contrast to that, no SR protein was significantly enriched (Table 2, P = 0.05, t-test).

Taken together, multiple SREs within exon 2/2b balance splice site selection in a strictly position-dependent manner.

ESS2b and ESE2b regulate balanced splice donor usage in provirus-transfected cells

To analyze the impact of ESS2b and ESE2b on viral pre-mRNA splicing, we inserted both most promising inactivating mutations ESE2b^{MUT} (ΔHZ_{EI} = -267) and ESS2b^{MUT} (C+D^{MUT}) either individually or in combination into proviral plasmid DNA pNL4-3 (GenBank accession no. 19921, (25)), with and without the inactivating G_{I2}-1 mutation (17). Following transfection of HEK293T CD4^{+} cells, RNA was isolated 48 hrs later, subjected to Northern blot analysis, and detected with an exon 7 probe hybridizing to all viral mRNAs. Mutating ESE2b showed no shift in viral mRNA levels compared to the wild-type proviral clone (Fig. 4A, cf. lanes 1 and 2), whereas, in contrast, inactivating ESS2b caused a strong increase in 2kb and 4kb *vif* mRNAs which was accompanied by a reduction of 9kb mRNAs (Fig. 4A, cf. lanes 1 and 3). Interestingly, inactivating mutations of both SREs seem to nearly compensate each other (Fig. 4A, cf. lanes 1 and 4), suggesting that even though there seems to be no obvious effect for mutating ESE2b in viral mRNA distribution at first glance, both these SREs together critically regulate the balance of HIV-1 RNA classes. In agreement with our previous results (17), mutating G_{I2}-1 caused an increased amount of 2kb and, particularly, of 4kb *vif* mRNAs, which was comparable to inactivating ESS2b (Fig. 4A, cf. lanes 3 and 5). Inactivation of ESS2b and G_{I2}-1 resulted in an even stronger effect (Fig. 4A, cf. lanes 1, 5, 7).

214 Next, to quantitatively measure individual HIV-1 transcript ratios, RT-PCRs were set up
215 with different primer pairs, each normalized to the total amount of all viral mRNAs
216 measured with primers detecting exon 7 (#3387/#3388, Fig 4B). Since exon-2- and
217 exon-3-recognition underlie inverse regulation (26-29), we used exon-junction primer
218 pairs specifically detecting *vif* and *vpr* or [1.2.5] and [1.3.5] *nef* mRNAs as two distinct
219 targets for exon 2 vs. exon 3 inclusion into viral mRNAs (#3395/#3396 (*vif*);
220 #3397/#3398 (*vpr*); #3395/#4843 ([1.2.5 *nef*]) and #3397/#3636 ([1.3.5 *nef*]), Fig 4B). As
221 expected, inactivation of ESE2b showed no significant change in *vif*, *vpr* and [1.2.5],
222 [1.3.5] *nef* mRNA levels (1-way ANOVA with Dunnett's post-hoc test), whereas
223 disruption of ESS2b induced a huge upregulation of *vif* and [1.2.5] *nef* ($P<0.001$,
224 Dunnett's post-hoc test) and a reduction of *vpr* and [1.3.5] *nef* mRNAs (Fig. 4C a, b, cf.
225 bars 2 and 3). Inactivation of both SREs resulted in mRNA levels comparable to wild-
226 type (Fig. 4C a, b, bar 4). Likewise, inactivation of G_{I2}-1 led to comparable effects with
227 an overall higher level of *vif* mRNAs (Fig. 4C a, bars 5-8, *vif*). Furthermore, we
228 measured the levels of unspliced and multiply spliced mRNAs with both intact and
229 inactivated G_{I2}-1 (#3389/#3390 (unspliced); #3391/#3392 (multiply spliced), Fig 4B).
230 There was no significant difference to the wild-type after disruption of ESE2b (Dunnett's
231 post-hoc test), but a clear decrease in unspliced mRNAs for inactivating ESS2b (Fig. 4C
232 c, cf. bars 2 and 3; 6 and 7), which could again be compensated by additionally mutating
233 ESE2b (Fig. 4C c, lanes 4 and 8).
234 To break down which impact both mutations had on distinct mRNA species, we also
235 performed semi-quantitative RT-PCR. In line with minigene analyses and position-
236 dependent effects, inactivating ESE2b revealed a complete loss of D2b usage (Fig. 4D,
237 D2b splicing, lane 2, e.g. Tat2b), whereas there was an elevated level of D2b usage

after inactivating ESS2b (Fig. 4D, D2b splicing, lane 3, e.g. Nef3b) and an upregulation of otherwise low abundant mRNA species (Fig. 4D, 2 kb species, lane 3, e.g. Gp41b [1.2b.5.7] (17)). Moreover, inactivating ESE2b led to a slight decrease of exon-2-containing transcripts like *vif2* or *tat2* (Fig. 4D, Ex1-4 splicing, lane 2). A mirror-inverted phenotype occurred after inactivation of ESS2b, where an increased degree of exon 2 inclusion could be observed (*vif2*), entailing a drop of exon 3 inclusion and *vpr* messages (*vpr3*), thereby sustaining their mutually regulated role in HIV-1 splicing as it has been shown by qPCR analysis in Fig. 4C (Fig. 4D, Ex1-4 splicing and 4kb, lane 3). Comparing overall 2kb and 4kb mRNA species in general, only marginal differences to wild-type pNL4-3 could be detected for ESE2b (Fig. 4D, 2kb and 4kb, cf. lanes 1 and 2), compatible with Northern blot analysis. As expected, for ESS2b, elevated levels of exon 2 inclusion with a concomitant reduction in exon 3 including mRNAs could be observed (Fig. 4D, 2kb and 4kb, cf. lanes 1 and 3). Again, for all detected mRNA species, a splicing pattern comparable to wild-type pNL4-3 was observed, if both SREs had been mutated (Fig. 4D, cf. lanes 1 and 4). As shown before, inactivation of G_{I2}-1 resulted in an enhanced exon 2b inclusion, followed by an increased amount of exon 2 containing transcripts, supporting the exon-bridging function of A1 with respect to D2 and D2b (Fig. 4D, cf. lanes 1 and 5). Additionally mutating ESE2b or ESS2b had none or only minor effects on the splicing patterns (Fig. 4D, cf. lanes 2 and 6; 3 and 7). In summary, RT-PCR analyses of RNA expressed from proviral clone pNL4-3 confirmed the results of the minigene analyses, revealing ESE2b and ESS2b as essential SREs regulating splice donor usage within exon 2/2b and thus *vif* mRNA processing.

ESE2b and ESS2b are essential for viral infectivity

261 To test to what extent changes in exon 2/2b inclusion reflect viral protein expression we
262 performed immunoblot analysis. No obviously different phenotype for the investigated
263 proteins was observed after inactivating ESE2b (Fig. 5A, cf. lanes 1 and 2). In
264 agreement with the data obtained from (q)RT-PCR analysis, a strong increase in Vif
265 protein level could be observed after inactivating ESS2b (Fig. 5A, cf. lanes 1 and 3). As
266 expected, mutating both SREs brought Vif protein level back to wild-type pNL4-3 level
267 (Fig. 5A, cf. lanes 1 and 4). Additionally interrupting G_{I2}-1 enhanced the effect of ESS2b
268 and further increased Vif protein expression (Fig. 5A, lane 7). Moreover, a drop in
269 intracellular p24 Gag levels as well as in viral capsid within the supernatant could be
270 observed for the ESS2b mutant with intact or inactivated G_{I2}-1 (Fig. 5A, lanes 3 and 7).
271 Furthermore, we used an antibody directed against the C-terminal domain of Gp41
272 (Chessie 8, (30)) to examine the presence of the previously described Gp41b isoform
273 (17). In agreement with RT-PCR analyses, also Gp41b protein was enriched after
274 ESS2b mutation (Fig. 5A, cf. lanes 3 and 7).
275 Eventually, we tested whether viral particles within the supernatants harboring either
276 individual mutations or both were still infectious. For this, we used GHOST cells that
277 stably expressed the CD4 receptor and contained an LTR-dependent gene of the
278 enhanced green fluorescence protein (eGFP). Thus, after successful infection and Tat-
279 mediated transactivation of the LTR promoter, eGFP expression can be easily monitored
280 via fluorescence microscopy. 48 hours post infection, a strong eGFP expression was
281 observed for wild-type pNL4-3, and it was clearly reduced in the ESE2b mutant infected
282 cells (Fig. 5B, cf. panels 1 and 2). Furthermore, infection with ESS2b mutant viral
283 particles led to a complete loss of eGFP expression, which was partially restored in cells
284 infected with viral particles harboring both mutations (Fig. 5B, cf. panels 3 and 4). p24

levels within the supernatant reflected the observed eGFP expression (Fig. 5C). In summary, the severely altered phenotype of the ESS2b-mutant already observed during (q)RT-PCR, Northern and Western blot analysis, led to a complete failure of infectiousness. Surprisingly, mutating ESE2b already showed a clear drop in eGFP expression, which was not indicated by the transfection experiments alone. Thus, an already slight imbalance in viral exon 2 splicing could lead to an impairment of proper viral particle production. In viral particles containing both mutations, balance could be restored at least to some extent.

Masking of ESE2b and ESS2b restricts viral particle production

As it was shown before (26, 27), the usage of locked-nucleic acids (LNAs) can mimic the mutational analysis of SREs within the provirus. Those modified antisense oligonucleotides are able to mask any specific sequence, in particular SREs, and thereby inhibit the binding of SR or hnRNP proteins. We used LNAs targeting either ESE2b or ESS2b, and co-transfected them with pNL4-3 (Fig. 6A). Scrambled LNAs not targeting any viral sequence were used as a control. 48 hrs post transfection, RNA and protein were isolated and analyzed for mRNA levels and protein expression. Northern blot analysis revealed a similar distribution of viral mRNA classes when the two SREs were masked by LNAs, as was obtained by SRE mutation (cf. Fig 6B with Fig. 4A). Here, LNAs targeting ESE2b showed a slight reduction of 4kb mRNAs, whereas LNAs targeting ESS2b showed a strong increase in 4kb *vif* mRNA and a decrease in unspliced 9kb mRNA (Fig. 6B). Furthermore, we examined the levels of both intracellular Gag protein and virus particles released into the supernatant (Fig. 6C). In agreement with the p24 levels detected after virus infection (Fig. 5C), we observed significantly less p24 Gag both within cells and supernatant for both LNAs. Additionally, RT-PCR analysis

showed a dramatic loss of exon 2/2b inclusion for LNAs targeting ESE2b (Fig. 6D, e.g. *vif2* and *tat2b*, cf. lanes 1 and 3), followed by an increase in exon 3 inclusion (Fig. 6D, e.g. *vpr3*, cf. lanes 1 and 3). Conversely, splicing shifted towards exon 2 inclusion when LNAs against ESS2b were applied (Fig. 6D, e.g. *vif2* and *tat2*, cf. lanes 1 and 4), while exon 3 inclusion was reduced at the same time (Fig. 6D, e.g. *vpr3*, cf. lanes 1 and 4). Taken together, masking ESE2b or ESS2b with LNAs showed a phenotype very similar to infection experiments and was able to inhibit proper virus particle production.

In summary, data obtained in these experiments highlight the existence of a tight cluster of splicing regulatory elements within exon 2/2b that balances viral mRNA and protein production. Inhibiting protein binding to those elements disrupts viral particle production and infectivity.

Multiple SRE sequence variations between HIV-1 subtypes

Aligning the HIV-1 consensus sequences A1 to AE of HIV-1 exon 2/2b using the RIP 3.0 software (https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html) showed that sequence variations between viral strains occurred strikingly more often within the regions containing the splicing regulatory elements ESS2b, ESE2b, and G_{I2}-1, while the flanking sequences were mainly conserved (Fig. 7A). The impact of these natural nucleotide variations on splice enhancing properties was reflected in their HEXplorer profiles: Indeed, HIV-strains showed a wide range of ΔHZ_{EI} scores. In order to examine one exemplary naturally occurring variation, we substituted in the minigene reporter the subtype K sequence exhibiting both high ΔHZ_{EI} and an additional deletion of five nucleotides within ESS2b. In fact, subtype K experimentally showed a splicing phenotype similar to A—D with a slight tendency towards D2 usage (Fig. 7B, left, cf.

lanes 1 and 3). The HEXplorer profile of subtype K (Fig. 7B, right, black bars) showed only a minor effect on ESS2b compared to pNL4-3 (blue bars), and a weaker ESE2b. Both effects tend to shift splice site selection further towards D2, which is barely visible since D2 already dominates splicing in pNL4-3. The high SRE sequence variability between HIV-1 subtypes may suggest an equally wide range of splicing regulatory properties that permits adjusting Vif levels to A3G levels in a variety of cellular host environments.

Discussion

The data presented in this work show that a splicing regulatory network (Fig. 8) regulates HIV-1 exon 2/2b inclusion into viral mRNAs, thus optimizing viral replication via competing actions of several SREs located close to D2 and D2b. In particular, we identified the Tra2/SRSF10-binding site ESE2b and the hnRNP A/B-binding site ESS2b, that could be specifically masked by LNAs. Both SREs contribute to regulating 5'ss D2/D2b and 3'ss A1, as well as *vif* mRNA and protein production.

During alternative splicing, recognition of splice sites is most often not only facilitated by conserved sequence elements like the 5'ss and 3'ss, but also by RNA secondary structure (2, 31-33) and a multitude of splicing regulatory elements. While splicing patterns of various HIV-1 subtypes are mostly conserved, frequency of splice site usage can depend on temperature (2) and presence of splicing regulatory proteins (14).

Within the non-coding exon 2/2b, already six different SREs have been described. Three elements exist that enhance recognition of splice donor D2 and thereby inclusion of exon 2 into viral mRNAs: ESEs M1 and M2 (bound by SRSF1) (15) and ESE-Vif (bound by SRSF4) (16). Furthermore, an inhibitory GGGG-motif, overlapping with the already

intrinsically weak D2, inhibits its usage and exon 2 inclusion (16), potentially through sterical hindrance of the U1 snRNP. We have previously reported that a G run located downstream of exon 2 inhibits the further downstream lying splice donor D2b by binding of hnRNP F/H (G_{I2}-1) (17). Inactivation of this G_{I2}-1 motif led to a strong increase in usage of the otherwise only little-used donor D2b. This was attributed to the fact that binding of hnRNP F/H leads to the formation of a “dead-end” complex, meaning that the U1 snRNP binds to the 5’ss without actually splicing at this position (12, 34, 35).

Upregulation of D2b usage following G_{I2}-1 inactivation indicated that an SR protein binding site could be located within exon 2b. We have previously found an enhancing element located downstream of D2 within a HEXplorer-based screen of total HIV-1 mRNA (13). Continuing analysis of this element here showed that the enhancer ESE2b strongly activates D2b and simultaneously inhibits D2, which is facilitated by binding of SRSF10 and Tra2. Tra2β was previously shown to bind to GA-rich sequence elements (36-39), similar to the sequence of ESE2b. Cloning this element into the minigene indeed led to an excessive splicing phenotype at D2b, which however, was not observed in a physiological HIV-1 splicing context. During infection, we could confirm by RNA deep-sequencing that D2b is only marginally used (0.2%) compared to D2 (5.3%) (17). Here, we resolve this apparent discrepancy between splicing patterns of minigene and infection experiments by identifying a novel SRE located within exon 2b, ESS2b, which counteracts the strong ESE2b effects. By using MS analysis, we show that ESS2b is bound by members of the hnRNP A/B-family, which fits earlier studies showing that those proteins bind to sequences that include a “TAG” motif (40, 41).

It might be surprising that such a multitude of SREs should regulate 5’ss selection in an even non-coding exon. However, in order to obtain Vif, 3’ss A1 must be used, and A1

itself seems to require activation by an exon definition complex (42, 43) in which U1 snRNP binding to either D2 or D2b promotes the recognition of the upstream-located splice acceptor A1 by U2 snRNPs. On the other hand, splicing at D2 or D2b prevents Vif expression, which relies on intron 2 retention. This is similar to *env* mRNA processing where U1 snRNP binding to an even splicing incompetent D4 was needed for 3'ss A5 activation (44). Thus, the commonly observed higher amounts of both intron-retaining—leading to Vif expression—and exon 2 including mRNAs can be due to increased U1 snRNA complementarity or mutations of neighboring SREs (16, 17, 29, 45).

Only balanced levels of Vif expression contribute to maximal viral replication, while excessive Vif expression is detrimental to viral replication due to perturbation of proteolytic Gag processing (19). On the other hand, excessive splicing at D2 leads to a decrease of unspliced mRNAs, and consequently, a reduction of Gag/Gag-Pol expression levels and a defect in virion production. This effect was also termed "oversplicing" and is in line with our observation, revealing that excessive Vif expression after mutating or masking of ESS2b leads to a reduction of overall unspliced mRNAs and an impairment of cellular Gag and viral particles within supernatant. Yet, not only excessive Vif levels, but also insufficient amounts are deleterious to viral replication. Vif is essential for counteracting the host cell restriction factor A3G, and an imbalance of Vif to A3G ratio strongly affects viral replication. It was shown that if restriction pressure is low, lower Vif levels are sufficient to counteract A3G, whereas excessive Vif impedes viral replication ability (19, 46). However, on the contrary, HIV-1 only replicates in cells with high restriction pressure, if sufficient Vif is present (17, 46).

Nomaguchi et al. identified natural single-nucleotide variations within different HIV-1 isolates proximal to HIV-1 SA1 (SA1prox), which could be shown to regulate *vif* mRNA

and Vif protein expression and were linked to the fact that an optimal Vif to A3G ratio is decisive for proper viral replication (17, 46). Here, we find nucleotide variations predominantly within splicing regulatory elements in exon 2/2b.

Therefore, it is tempting to speculate that the vast number of SREs within exon 2/2b ensures viral replication in cells with different A3G or splicing regulatory protein concentrations, e.g. by a mechanism like mutual evolution (46).

Materials and Methods

Single-intron splicing constructs

All eGFP single-intron splicing reporters are based on the well-established HIV-1 glycoprotein/eGFP expression plasmid (20). Insertion of exon 2b Parts I-IV was carried out by replacing GAR H of SV GAR H SD4 Δvpu env eGFP D36G (22) with a PCR product obtained with primer pairs #4200/#4201 (Part I), #4202/#4203 (Part II), #4204/#4205 (Part III) and #4206/#4207 (Part IV), respectively. The neutral sequence (23) was inserted 3.5 times as described above with primer pair #4213/#4214.

Three-exon minigenes

The three-exon minigenes are derived from the Fibrinogen Bβ minigene pT-Bβ-IVS7+1G>T (47, 48). The middle exon was replaced with only splicing neutral sequences (23) by using a customized synthetic gene from Invitrogen and inserted into pT-Bβ-IVS7+1G>T via EcoNI/Bpu10I. HIV-1-derived splice donors D2 and D2b were inserted with PCR products resulting from primer pair #4793/#4794. ESE-Vif, -M1, -M2 were inserted by PCR with primer pair #4853/#2620. Fragments of HIV-1 exon 2/2b were added at their authentic positions relative to D2 or D2b, respectively, by using primer pairs #4795/#2620 (ESE2b and G_{I2}-1), #4798/#2620 (ΔHZ_{EI}-94 and G_{I2}-1),

#5318/#2620 (ΔHZ_{EI}-267 and G_{I2}-1), #4796/#2620 (ESE2b and G_{I2}-1^{MUT}), #5319/#2620 (ΔHZ_{EI}-94 and G_{I2}-1^{MUT}), #5317/#2620 (ΔHZ_{EI}-267 and G_{I2}-1^{MUT}), #5251/#2620 (ESS2b A-D), #5337/#2620 (ESS2b part A), #5339/#2620 (ESS2b part B), #5341/#2620 (ESS2b part C) and #5343/#2620 (ESS2b part D), respectively. The fragment of HIV-1 subtype K was added at its authentic position flanked by D2 and D2b using primer pair #5712/#5713. HEXplorer guided mutations of ESS2b were inserted via PCR products resulting from primer pairs #5392/#2620 (C^{MUT}), #5393/#2620 (D^{MUT}) and #5394/#2620 (C^{MUT} D^{MUT}), respectively.

Proviral plasmids

Proviral DNA pNL4-3 ESE2b^{MUT} was generated by overlapping PCR technique using primers #5549/#4773 and #5553/#5550; pNL4-3 ESS2b^{MUT} using primer pairs #5547/#4773 and #5553/#5548; pNL4-3 ESE2b^{MUT} ESS2b^{MUT} using primer pairs #5551/#4773 and #5553/#5552. pNL4-3 G_{I2}-1^{MUT} has been described previously (17) and used as a template instead of pNL4-3 using primer pairs depicted above to generate double or triple mutations, respectively.

Expression plasmids

pXGH5 (49) was cotransfected to monitor transfection efficiency.

pCL-dTOM was cotransfected to detect transfection efficiency of each sample in fluorescence microscopy analysis. The plasmid expresses the fluorescent protein Tomato and was kindly provided by Dr. H. Hanenberg.

Oligonucleotides

All oligonucleotides used were obtained from Metabion GmbH (Planegg, Germany) (see Table 3). RNase-Free HPLC purified LNAs were purchased from Exiqon (Denmark).

Cell culture and transfection

450 HeLa, HEK293T ($CD4^+$) or GHOST (3) CXCR4+ cells (50) were cultured in Dulbecco's
451 high-glucose modified Eagle's medium (Invitrogen) supplemented with 10% fetal calf
452 serum and 50 µg/ml penicillin/streptomycin (Invitrogen). For transient transfection, $2x10^5$
453 cells per six-well were used. Transient-transfection experiments were performed by
454 using TransIT®-LT1 transfection reagent (Mirus Bio LLC US) according to the
455 manufacturer's instructions. LNA transfection was performed as described in (26).

456 **RNA isolation and RT-PCR**

457 24 hrs or 48 hrs post transfection, total RNA was isolated by using acid guanidinium
458 thiocyanate-phenol-chloroform (51). For semi-quantitative and quantitative RT-PCR
459 analyses, RNA was reversely transcribed by using Superscript III Reverse Transcriptase
460 (Invitrogen) and Oligo(dT) primers (Invitrogen), and amplified using primer pairs depicted
461 in Fig. 4B.

462 **Northern blotting**

463 3 µg of total RNA isolated by using acid guanidinium thiocyanate-phenol-chloroform (51)
464 was separated on denaturing 1% agarose gel and then capillary blotted onto positively
465 charged nylon membrane. Hybridization was carried out using a digoxigenin (DIG)-
466 labeled HIV-1 exon 7 PCR-amplicon (#3387/#3388) as previously described (17).

467 **Protein isolation and western blotting**

468 Proteins samples were heated up to 95°C for 10 min and loaded onto sodium dodecyl
469 sulphate-polyacrylamide gel (SDS PAGE) for Western blot analysis. Samples were
470 transferred to a nitrocellulose membrane probed with primary and secondary antibodies
471 (Sheep antibody against HIV-1 p24 CA from Aalto; mouse monoclonal antibodies
472 specific for HIV-1 Vif (ab66643) from Abcam; mouse anti-gp41 (Chessie 8, (30)) and

mouse anti β-actin monoclonal antibody (A5316) from Sigma-Aldrich) and developed with ECL chemiluminescence reagent (GE Healthcare).

RNA affinity purification assay

3000 picomoles RNA oligonucleotides for either wild-type (WT) or mutant version of ESE2b and ESS2b, respectively, were covalently coupled to adipic acid dihydrazide agarose beads (Sigma). 60% of HeLa nuclear extract (Cilbiotech) was added to the immobilized RNAs. After five stringent washing steps with buffer D containing different concentrations of KCl (0.1, 0.25, 0.5, 0.25, 0.1 M KCl, together with 20 mM HEPES-KOH [pH 7.9], 5% [vol/vol] glycerol, 0.2 M ethylenediaminetetraacetic acid, 0.5 mM dithiothreitol, 0.4 M $MgCl_2$), precipitated proteins were eluted in protein sample buffer. Samples were sent to the Molecular Proteomics Laboratory, BMFZ, Heinrich Heine University, Düsseldorf for MS analysis.

HEXplorer score calculation

HEXplorer score profiles of wild-type and mutant sequences were calculated using the web interface (https://www2.hhu.de/rna/html/hexplorer_score.php; (13, 24)).

qPCR statistics

In qPCR experiments, expression levels relative to WT were calculated as *exp(-Δct)* ratios. Bar graphs show mean and standard deviation of three replicates. Statistical significance was determined separately for each sample (*vif*, *vpr*, exon 2, exon 3, unspliced, multiply spliced) by 1-way ANOVA followed by Dunnett's post-hoc test correcting for multiple comparisons.

Acknowledgements

We thank Björn Wefers for excellent technical assistance. The following reagents were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: Chessie 8 from Dr. George Lewis, GHOST (3) CXCR4+ cells from Dr. Vineet N. KewalRamani and Dr. Dan R. Littman.

Funding was provided by Deutsche Forschungsgemeinschaft (DFG) [SCHA 909/8-1]; Jürgen Manchot Stiftung (to A.L.B., L.W., H.S.); Stiftung für AIDS-Forschung, Düsseldorf (to H.S.).

Author details

ALB and HS conceived the study and designed the experiments. ALB, LW, LM and MW performed cloning, transfection experiments and (q)RT-PCR analyses. ALB performed RNA-pull-down analyses. FH performed LNA-related experiments. ST, ALB and HS performed HEXplorer analyses. ST provided statistical analyses. ALB, ST and HS wrote the manuscript. All authors read and approved the final manuscript.

References

1. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res 40:10345-10355.

2. Emery A, Zhou S, Pollom E, Swanstrom R. 2017. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. J Virol Feb 28;91(6). pii: e02515-16.

3. Purcell DF, Martin MA. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67:6365-6378.

4. Kim SY, Byrn R, Groopman J, Baltimore D. 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. J Virol 63:3708-3713.

5. Klotman ME, Kim S, Buchbinder A, DeRossi A, Baltimore D, Wong Staal F. 1991. Kinetics of expression of multiply spliced RNA in early human immunodeficiency virus type 1 infection of lymphocytes and monocytes [published erratum appears in Proc. Natl. Acad. Sci. U. S. A. 1992 89:1148]. Proc Natl Acad Sci U S A 88:5011-5015.

6. Mohammadi P, Desfarges S, Bartha I, Joos B, Zangger N, Munoz M, Gunthard HF, Beerenwinkel N, Telenti A, Ciuffi A. 2013. 24 hours in the life of HIV-1 in a T cell line. PLoS Pathog 9:e1003161.

7. Karn J, Stoltzfus CM. 2012. Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression, p a006916. In Frederic D. Bushman GJN, and Ronald Swanstrom (ed), Cold Spring Harb Perspect Med, 2012/02/23 ed, vol 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

8. Will CL, Luhrmann R. 2011. Spliceosome structure and function. Cold Spring Harb Perspect Biol 3.

9. Berget SM. 1995. Exon recognition in vertebrate splicing. J Biol Chem 270:2411-2414.

10. Romfo CM, Alvarez CJ, van Heeckeren WJ, Webb CJ, Wise JA. 2000. Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. Mol Cell Biol 20:7955-7970.

11. Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A 102:16176-16181.

12. Erkelenz S, Mueller WF, Evans MS, Busch A, Schoneweis K, Hertel KJ, Schaal H. 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA 19:96-102.

13. Erkelenz S, Theiss S, Otte M, Widera M, Peter JO, Schaal H. 2014. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic Acids Res 42:10681-10697.

14. Stoltzfus CM. 2009. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. Adv Virus Res 74:1-40.

15. Kammler S, Otte M, Hauber I, Kjems J, Hauber J, Schaal H. 2006. The strength of the HIV-1 3' splice sites affects Rev function. Retrovirology 3:89.

16. Exline CM, Feng Z, Stoltzfus CM. 2008. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol 82:3921-3931.

17. Widera M, Erkelenz S, Hillebrand F, Krikoni A, Widera D, Kaisers W, Deenen R, Gombert M, Dellen R, Pfeiffer T, Kaltschmidt B, Munk C, Bosch V, Kohrer K, Schaal H.

557 2013. An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif
558 mRNA. J Virol 87:2707-2720.

559 18. Sheehy AM, Gaddis NC, Choi JD, Malim MH. 2002. Isolation of a human gene
560 that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418:646-
561 650.

562 19. Akari H, Fujita M, Kao S, Khan MA, Shehu-Xhilaga M, Adachi A, Strebel K. 2004.
563 High level expression of human immunodeficiency virus type-1 Vif inhibits viral infectivity
564 by modulating proteolytic processing of the Gag precursor at the p2/nucleocapsid
565 processing site. J Biol Chem 279:12355-12362.

566 20. Kammler S, Leurs C, Freund M, Krummheuer J, Seidel K, Tange TO, Lund MK,
567 Kjems J, Scheid A, Schaal H. 2001. The sequence complementarity between HIV-1 5'
568 splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-
569 mRNA. RNA 7:421-434.

570 21. Freund M, Asang C, Kammler S, Konermann C, Krummheuer J, Hipp M, Meyer I,
571 Gierling W, Theiss S, Preuss T, Schindler D, Kjems J, Schaal H. 2003. A novel
572 approach to describe a U1 snRNA binding site. Nucleic Acids Res 31:6963-6975.

573 22. Caputi M, Freund M, Kammler S, Asang C, Schaal H. 2004. A bidirectional
574 SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency
575 virus type 1 rev, env, vpu, and nef gene expression. J Virol 78:6517-6526.

576 23. Zhang XH, Arias MA, Ke S, Chasin LA. 2009. Splicing of designer exons reveals
577 unexpected complexity in pre-mRNA splicing. RNA 15:367-376.

578 24. Brillen AL, Schoneweis K, Walotka L, Hartmann L, Muller L, Ptok J, Kaisers W,
579 Poschmann G, Stuhler K, Buratti E, Theiss S, Schaal H. 2016. Succession of splicing
580 regulatory elements determines cryptic 5'ss functionality. Nucleic Acids Res
581 doi:10.1093/nar/gkw1317.

582 25. Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, Martin MA.
583 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in
584 human and nonhuman cells transfected with an infectious molecular clone. The Journal
585 of Virology 59:284-291.

586 26. Widera M, Hillebrand F, Erkelenz S, Vasudevan AA, Munk C, Schaal H. 2014. A
587 functional conserved intronic G run in HIV-1 intron 3 is critical to counteract
588 APOBEC3G-mediated host restriction. Retrovirology 11:72.

589 27. Erkelenz S, Hillebrand F, Widera M, Theiss S, Fayyaz A, Degrandi D, Pfeffer K,
590 Schaal H. 2015. Balanced splicing at the Tat-specific HIV-1 3'ss A3 is critical for HIV-1
591 replication. Retrovirology 12:29.

592 28. Madsen JM, Stoltzfus CM. 2005. An exonic splicing silencer downstream of the 3
593 ' splice site A2 is required for efficient human immunodeficiency virus type 1 replication.
594 J Virol 79:10478-10486.

595 29. Mandal D, Feng Z, Stoltzfus CM. 2010. Excessive RNA splicing and inhibition of
596 HIV-1 replication induced by modified U1 small nuclear RNAs. J Virol 84:12790-12800.

597 30. Abacioglu YH, Fouts TR, Laman JD, Claassen E, Pincus SH, Moore JP, Roby
598 CA, Kamin Lewis R, Lewis GK. 1994. Epitope mapping and topology of baculovirus-

599 expressed HIV-1 gp160 determined with a panel of murine monoclonal antibodies. AIDS
600 Res Hum Retroviruses 10:371-381.

601 31. Abbink TE, Berkhout B. 2008. RNA structure modulates splicing efficiency at the
602 human immunodeficiency virus type 1 major splice donor. J Virol 82:3090-3098.

603 32. Zychlinski D, Erkelenz S, Melhorn V, Baum C, Schaal H, Bohne J. 2009. Limited
604 complementarity between U1 snRNA and a retroviral 5' splice site permits its attenuation
605 via RNA secondary structure. Nucleic Acids Res 37:7429-7440.

606 33. Mueller N, van Bel N, Berkhout B, Das AT. 2014. HIV-1 splicing at the major
607 splice donor site is restricted by RNA structure. Virology 468-470C:609-620.

608 34. Domsic JK, Wang Y, Mayeda A, Krainer AR, Stoltzfus CM. 2003. Human
609 immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV
610 antagonizes binding of U2AF65 to viral polypyrimidine tracts. Mol Cell Biol 23:8762-
611 8772.

612 35. Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL. 2008. Polypyrimidine
613 tract binding protein controls the transition from exon definition to an intron defined
614 spliceosome. Nat Struct Mol Biol 15:183-191.

615 36. Tacke R, Tohyama M, Ogawa S, Manley JL. 1998. Human Tra2 proteins are
616 sequence-specific activators of pre-mRNA splicing. Cell 93:139-148.

617 37. Grellscheid S, Dalgliesh C, Storbeck M, Best A, Liu Y, Jakubik M, Mende Y,
618 Ehrmann I, Curk T, Rossbach K, Bourgeois CF, Stevenin J, Grellscheid D, Jackson MS,

619 Wirth B, Elliott DJ. 2011. Identification of evolutionarily conserved exons as regulated
620 targets for the splicing activator tra2beta in development. PLoS Genet 7:e1002390.

621 38. Tsuda K, Someya T, Kuwasako K, Takahashi M, He F, Unzai S, Inoue M, Harada
622 T, Watanabe S, Terada T, Kobayashi N, Shirouzu M, Kigawa T, Tanaka A, Sugano S,
623 Guntert P, Yokoyama S, Muto Y. 2011. Structural basis for the dual RNA-recognition
624 modes of human Tra2-beta RRM. Nucleic Acids Res 39:1538-1553.

625 39. Erkelenz S, Poschmann G, Theiss S, Stefanski A, Hillebrand F, Otte M, Stuhler
626 K, Schaal H. 2013. Tra2-mediated recognition of HIV-1 5' splice site D3 as a key factor
627 in the processing of vpr mRNA. J Virol 87:2721-2734.

628 40. Burd CG, Dreyfuss G. 1994. RNA binding specificity of hnRNP A1: significance of
629 hnRNP A1 high-affinity binding sites in pre-mRNA splicing. EMBO J 13:1197-1204.

630 41. Kajita Y, Nakayama J, Aizawa M, Ishikawa F. 1995. The UUAG-specific RNA
631 binding protein, heterogeneous nuclear ribonucleoprotein D0. Common modular
632 structure and binding properties of the 2xRBD-Gly family. J Biol Chem 270:22167-
633 22175.

634 42. Robberson BL, Cote GJ, Berget SM. 1990. Exon Definition May Facilitate Splice
635 Site Selection in Rnas with Multiple Exons. Mol Cell Biol 10:84-94.

636 43. Hoffman BE, Grabowski PJ. 1992. U1 snRNP targets an essential splicing factor,
637 U2AF65, to the 3' splice site by a network of interactions spanning the exon. Genes Dev
638 6:2554-2568.

639 44. Asang C, Hauber I, Schaal H. 2008. Insights into the selective activation of
640 alternatively used splice acceptors by the human immunodeficiency virus type-1
641 bidirectional splicing enhancer. Nucleic Acids Res 36:1450-1463.

642 45. Madsen JM, Stoltzfus CM. 2006. A suboptimal 5 ' splice site downstream of HIV-1
643 splice site A1 is required for unspliced viral mRNA accumulation and efficient virus
644 replication. Retrovirology 3:10.

645 46. Nomaguchi M, Doi N, Sakai Y, Ode H, Iwatani Y, Ueno T, Matsumoto Y, Miyazaki
646 Y, Masuda T, Adachi A. 2016. Natural Single-Nucleotide Variations in the HIV-1
647 Genomic SA1prox Region Can Alter Viral Replication Ability by Regulating Vif
648 Expression Levels. J Virol 90:4563-4578.

649 47. Spena S, Duga S, Asselta R, Malcovati M, Peyvandi F, Tenchini ML. 2002.
650 Congenital afibrinogenemia: first identification of splicing mutations in the fibrinogen
651 Bbeta-chain gene causing activation of cryptic splice sites. Blood 100:4478-4484.

652 48. Spena S, Tenchini ML, Buratti E. 2006. Cryptic splice site usage in exon 7 of the
653 human fibrinogen B beta-chain gene is regulated by a naturally silent SF2/ASF binding
654 site within this exon. RNA 12:948-958.

655 49. Selden RF, Howie KB, Rowe ME, Goodman HM, Moore DD. 1986. Human
656 growth hormone as a reporter gene in regulation studies employing transient gene
657 expression. Mol Cell Biol 6:3173-3179.

658 50. Morner A, Bjorndal A, Albert J, Kewalramani VN, Littman DR, Inoue R,
659 Thorstensson R, Fenyo EM, Bjorling E. 1999. Primary human immunodeficiency virus

660 type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in
661 coreceptor usage. J Virol 73:2343-2349.

662 51. Chomczynski P, Sacchi N. 1987. Single-step method of RNA isolation by acid
663 guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156-159.

664 Figure Legends

665 **Figure 1.** Analysis of splicing regulatory elements (SREs) in HIV-1 exon 2/2b. (**A**) Black
666 (silencer) and grey (enhancer) bars indicate published SREs. Splice donor sites (D1—
667 D4), splice acceptor sites (A1—A7) and protein ORFs are shown. (**B**) Known SREs
668 within exon 2/2b and schematic of exon 2b parts I—IV. (**C**) Fluorescence microscopic
669 analysis of fragments I—IV. Top: Schematic overview of the single-intron eGFP splicing
670 reporter. Bottom: HeLa cells were transiently transfected with 1 µg of each construct
671 together with 1 µg of pCL-dTOM to monitor transfection efficiency. 24 h after
672 transfection, fluorescence microscopy was carried out.

673 **Figure 2.** Impact of ESE2b on D2b recognition. (**A**) Schematic of the three-exon
674 minigene. The middle exon is composed of only neutral CCAAACAA repeats (23) except
675 for D2, D2b and the depicted SREs. (**B**) RT-PCR analyses of the minigene (A) splicing
676 pattern. HeLa cells were transiently transfected with 1 µg of each construct and 1 µg of
677 pXGH5. RNA isolated from cells were subjected to RT-PCRs using primer pairs
678 #2648/#2649 and #1224/#1225 (hGH). PCR amplicons were separated on a non-
679 denaturing 10% polyacrylamide gel and stained with ethidium bromide. (**C**) HZ_{EI} plots of
680 ESE2b, ΔHZ_{EI} = -94 and ΔHZ_{EI}= -267 (black: mutated sequence; blue: wild-type
681 reference).

682 **Figure 3.** ESS2b located between D2 and ESE2b is bound by members of hnRNP A/B
683 family and counteracts ESE2b. (**A**) HEXplorer score profiles of sequence A—D
684 (indicated by black bars) and mutations of fragments C, D, or both (black: mutated
685 sequence; blue: wild-type reference) composing ESS2b. (**B-C**) Mutational analysis of
686 ESS2b. HeLa cells were transiently transfected with 1 µg of each construct and 1 µg of
687 pXGH5. Twenty-four hours after transfection, RNA was isolated from the cells and
688 subjected to RT-PCR analysis using primer pairs #2648/#2649 and #1224/#1225 (hGH).
689 **Figure 4.** ESE2b and ESS2b cause alterations in proviral pre-mRNA processing. (**A**)
690 Northern blot analysis of total RNA isolated from HEK293T $CD4^+$ cells transfected with
691 wild-type or mutant pNL4-3. A hybridization probe was used specifically detecting HIV-1
692 exon 7. (**B**) Binding sites of (q)RT-PCR primers. (**C**) qRT-PCR of total RNA isolated from
693 the same RNA preparation as in (A) to specifically quantitate the levels of (a) *vif* vs. *vpr*,
694 (b) [1.2.5] vs. [1.3.5.] and (c) multiply spliced vs. unspliced mRNA species, displaying
695 *exp(-Δct)* ratios normalized to wild-type splicing pattern. Bar graphs show mean and
696 standard deviation of three replicates. Primer pair #3387/#3388 specifically detecting
697 exon 7 was used for normalization. Following primer pairs were used: *vif*: #3395/#3396;
698 *vpr*: #3397/#3398; [1.2.5]: #3395/#4843; [1.3.5]: #3397/#3636; multiply spliced:
699 #3391/#3392; unspliced: #3389/#3390. (**D**) RT-PCR analysis of RNA used in (A) and
700 (C). Primer pairs: #1544/#3632 (Ex1-4 splicing), #2710/#3392 (D2b splicing),
701 #1544/#3392 (2kb species), #1544/#640 (4kb species). HIV-1 mRNA species are
702 indicated on the left hand side of each gel image according to (3). Exon numbers are
703 indicated in square brackets; those including an E read through D4.
704 **Figure 5**. Impairment of proper viral particle production. (**A**) Immunoblot analysis of
705 proteins of pelleted virions from the supernatant (SN) of transfected cells described in

706 Fig. 4. (**B-C**) 2.5×10^5 HEK293T cells were transfected with pNL4-3 and mutant
707 proviruses. 48h post transfection, the supernatant was collected for infection of GHOST
708 $CD4^+$ cells, an indicator cell line which expresses eGFP after HIV-1 infection. Infection
709 and viral replication was analyzed 48h post infection, both by fluorescence microscopy
710 (B) and by p24-gag Western blot analysis (C) of supernatants of the infected GHOST
711 $CD4^+$ cells.

712 **Figure 6.** LNA-directed masking of ESE2b and ESS2b mimics mutational phenotype.
713 (**A**) Schematic of the LNA binding sites. (**B**) Northern blot analysis of total RNA. HeLa
714 cells were co-transfected with pNL4-3 and either LNAs masking ESE2b or ESS2b or the
715 scrambled LNA. Total RNA was isolated 24 h post transfection and subjected to
716 Northern blot analysis using an HIV-1 exon 7 probe. (**C**) Western blot analysis of cellular
717 (Cell) and supernatant (SN) Gag of co-transfected cells from (B). (**D**) RT-PCR analysis
718 of different viral mRNA species. Following primer pairs were used: #1544/#3632 (Ex1-4
719 splicing), #2710/#3392 (D2b splicing), #1544/#3392 (2kb species), #1544/#640 (4kb
720 species). HIV-1 mRNA species are indicated on the left hand side of each gel image
721 according to (3). Exon numbers are indicated in square brackets; those including an E
722 read through D4.

723 **Figure 7.** Analysis of SREs within exon 2/2b of different HIV-1 subtypes. (**A**) pNL4-3-
724 derived HIV-1 exon 2/2b consensus sequences from A1 to AE of the different HIV-1
725 subtypes, together with their HEXplorer score differences ΔHZ_{EI}. Conserved sequences
726 are represented by "–" and differences by letters. Regions with SREs are shown with red
727 or green background. The subtype sequences were analyzed with the RIP 3.0 software
728 (http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html). (**B**) Left: Splicing patterns of
729 the splicing reporter carrying SRE regions of subtype K (lane 1) and pNL4-3 (lane 3).

For reference, lanes 2 and 4 corresponding to the neutral sequence and to D^{MUT} are also shown. HeLa cells were transiently transfected with 1 µg of each construct and 1 µg of pXGH5. 24 h after transfection, RNA was isolated from the cells and subjected to RT-PCR analysis using primer pairs #2648/#2649 and #1224/#1225 (hGH). Right: HEXplorer profiles of pNL4-3 and exemplary subtype K containing a 5 nt deletion (between vertical red lines) and several single nt variations. Blue bars depict HEXplorer profile for pNL4-3 and black bars for subtype K.

Figure 8. Model for exon2/2b recognition. Exon 2/2b inclusion and splice donor usage is regulated by a complex network of SREs. (**A**) SR proteins binding to both ESE2 and ESE2b support U1 snRNP binding at the downstream located splice donors D2 and D2b. Exon definition leads to the concomitant upregulation of splice acceptor A1, and to higher *vif* mRNA expression (left-pointing arrows below exon 2/2b). (**B**) Lower levels of SR proteins as well as hnRNP binding to sites ESS2b and G_{I2}-1 reduce U1 snRNP binding to D2 and D2b.

Table 1. Mass spectrometry analysis of ESE2b. Log_2 differences and unique peptides of SR and hnRNP proteins enriched after RNA affinity purification are shown. Ratios of normalized protein intensities from purifications of wild type divided by mutated sequence samples are calculated.

	Unique Peptides	Log_2 Difference
TRA2A	10.5	3.97
TRA2B	13.5	3.20
SRSF10	11.5	3.16
SRSF3	4.5	1.39
SRSF7	7	1.32
SRSF6	4	0.99
SRSF4	4.5	0.68

SRSF9	12	0.45
SRSF1	25	0.45
SRSF2;SRSF8	4	-0.06
SRSF11	6	-0.32
HNRNPUL2-BSCL2;HNRNPUL2	12.5	1.48956667
HNRNPR	12	1.38463333
HNRNPL	14	0.5345
HNRNPDL	6	0.50626667
HNRNPU	26	0.45483333
HNRNPA2B1	20	0.3558
HNRNPH3	10.5	0.3429
HNRNPM	26	0.32243333
HNRNPA3	17.5	0.2572
HNRNPA1;HNRNPA1L2	21	0.24806667
HNRNPH2	9	0.19285
HNRNPLL;HNRPLL	17	0.14186667
HNRNPK	27	0.1075
HNRNPH1	6.5	0.06496667
HNRNPUL1	17	0.0539
HNRNPF	16	-0.0073
HNRNPC	5	-0.22633333
HNRNPD	6	-0.26476667
HNRNPA0	9.5	-1.01033333

748

749 **Table 2.** Mass spectrometry analysis of ESS2b. Log_2 differences and unique peptides of
750 SR and hnRNP proteins enriched after RNA affinity purification are shown. Ratios of
751 normalized protein intensities from purifications of wild type divided by mutated
752 sequence samples are calculated.

	Unique Peptides	**Log_2 Difference**
HNRNPDL	6	2.11
HNRNPA1;HNRNPA1L2	21	1.34
HNRNPA2B1	20	1.13
HNRNPA3	17.5	1.11
HNRNPA0	9.5	0.94
HNRNPUL2	12.5	0.53
HNRNPH3	10.5	0.41

HNRNPF	16	0.25
HNRNPH2	9	0.23
HNRNPH1	6.5	0.19
HNRNPR	12	0.16
HNRNPK	27	0.14
HNRNPL	14	0.13
HNRNPD	6	0.12
HNRNPU	26	0.10
HNRNPLL;HNRPLL	17	0.05
HNRNPUL1	17	-0.01
HNRNPC	5	-0.08
HNRNPM	26	-0.13
HNRNPH1;HNRNPH2	3	-0.31
SRSF2;SRSF8	2	0.38435
SRSF10	11.5	0.37666667
TRA2A	10.5	0.25796667
TRA2B	13.5	0.2346
SRSF1	20.5	0.1378
SRSF4	4.5	0.0756
SRSF9	6	0.05275
SRSF11	6	0.0258
SRSF6	4	0.0229
SRSF7	7	-0.06006667
SRSF3	4.5	-0.31765

753

754 **Table 3.** Primers used for cloning, (q)RT-PCR analyses and sequences of LNAs.

	CLONING
Primer	**Primer sequence**
#2620	GATCCCGGGAAAGATTTGTTGTCACATACAGAAG
#4200	AATTCGCAGTAGTAATACAAGATAATAGTGACATAGAGCT
#4201	CTATGTCACTATTATCTTGTATTACTACTGCG
#4202	AATTCGATAATAGTGACATAAAAGTAGTGCCAAGAGAGCT
#4203	CTCTTGGCACTACTTTTATGTCACTATTATCG
#4204	AATTCTAGTGCCAAGAAGAAAAGCAAAGATCATCAGAGCT
#4205	CTGATGATCTTTGCTTTTCTTCTTGGCACTAG

#4206	AATTCTCATCAGGGATTATGGAAAACAGATGGCAGGAGCT
#4207	CCTGCCATCTGTTTTCCATAATCCCTGATGAG
#4213	AATTCCCAAACAACCAAACAACCAAACAACCAAACGAGCT
#4214	CGTTTGGTTGTTTGGTTGTTTGGTTGTTTGGG
#4773	TGGATGCTTCCAGGGCTC
#4793	AACAAACCGGTAAGGTGAAGGGTCTAGACCAAACAACCAAACAAC
#4794	AACAGCGTACGTTGTTTGGTTGTTTGGTTGTTTGGTTGTTTGGTTGTTTGGGTCGACATCATC ACCTGGCGGCCGCTTGTTTG
#4795	AAGGGGCTAGCCCAAGAAGAAAAGCAAAGATCATCAGGGATTATGGAAAACAGGCGGCCG CCAGGT
#4796	AAGGGGCTAGCCCAAGAAGAAAAGCAAAGATCATCCGCGATTATGGAAAACAGGCGGCCG CCAGGT
#4798	AAGGGGCTAGCCCAAGATGAAAAGCAATGATCATCAGGGATTATGGAAAACAGGCGGCCG CCAGGT
#4853	AACAACCTTAGGGGACAGCAGAGATCCAGTTTGGAAAGGACCAGCAAAGCTCCTCTGGAAA GGGGACCCAAGGTGAAG
#5251	AAGGGGCTAGCGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAA AGCAAAGATCATCA
#5317	AAGGGGCTAGCCCAAGTAGAAAAGCATAGATCATCCGCGATTATGGAAAACAGGCGGC
#5318	AAGGGGCTAGCCCAAGTAGAAAAGCATAGATCATCAGGGATTATGGAAAACA
#5319	AAGGGGCTAGCCCAAGATGAAAAGCAAGATCATCCGCGATTATGGAAAACAGGCGG
#5337	AAGGGGCTAGCGCAGTAGTAATACACCAAACAACCAAACAACCAAACAACCAAGAAGAAAA GCAAAGATCATCA
#5339	AAGGGGCTAGCCAACCAAACAAAATACAAGATAACCAAACAACCAAACAACCAAGAAGAAA AGCAAAGATCATCA
#5341	AAGGGGCTAGCCCAAACAACCAAACAAAGATAATAGTGACCCAAACAACCAAGAAGAAAAG CAAAGATCATCA
#5343	AAGGGGCTAGCCCAAACAACCAAACAACCAAACAAGTGACATAAAAGTAGTGCCAAGAA
#5392	AAGGGGCTAGCGCAGTAGTAATACAAGATACTCGTGACATAAAAGTAGTGCCAAGAA

#5393	AAGGGGCTAGCGCAGTAGTAATACAAGATAATAGTGACATACAAGTACTGCCAAGAAGAAAAGCAAAGATCAT
#5547	CTCGTGACATACAAGTACTGCCAAGAAGAAAAGCAAAGATCAT
#5548	GTACTTGTATGTCACGAGTATCTTGTATTACTACTGCCCCTT
#5549	TAGAAAAGCATAGATCATCAGGGATTATGGAAAAC
#5550	ATGCTTTTCTACTTGGCACTACTTTTATGTCACT
#5551	CTCGTGACATACAAGTACTGCCAAGTAGAAAAGCATAGATCATCAGGGATTATGGAAAAC
#5552	ATGCTTTTCTACTTGGCAGTACTTGTATGTCACGAGTATCTTGTATTACTACTGCCCCTT
#5553	CTGGCAGAAAACAGGGAGATT
#5712	GGGCTAGCGCAGTAGTAATACAATAGTGAGATAAAGGTAGTACCAAGAAGAAAAGCAAAGAT
#5713	CCTGGCGGCCGCCCATCTGTTTTCCATAATCCCTAATAATCTTTGCTTTTCTTCTTGG
	(q)RT-PCR
Primer	**Primer sequence**
#640	CAATACTACTTCTTGTGGGTTGG
#1224	TCTTCCAGCCTCCCATCAGCGTTTGG
#1225	CAACAGAAATCCAACCTAGAGCTGCT
#1544	CTTGAAAGCGAAAGTAAAGC
#2648	AGTGATTCAGAACCGTCAAG
#2649	TCCACCACCGTCTTCTTTAG
#2710	GGGGGGATCGATAATTAAGGAGTTTATATGGAAACCCTTAAAGGTAAAGGGGCAGTAGTAATACAA
#3387	TTGCTCAATGCCACAGCCAT
#3388	TTTGACCACTTGCCACCCAT
#3389	TTCTTCAGAGCAGACCAGAGC
#3390	GCTGCCAAAGAGTGATCTGA
#3391	TCTATCAAAGCAACCCACCTC

#3392	CGTCCCAGATAAGTGCTAAGG
#3395	GGCGACTGGGACAGCA
#3396	CCTGTCTACTTGCCACAC
#3397	CGGCGACTGAATCTGCTAT
#3398	CCTAACACTAGGCAAAGGTG
#3632	TGGATGCTTCCAGGGCTC
#3636	CCGCTTCTTCCTTGTTATGTC
#4843	CCGCTTCTTCCTTTCCAGAGG
	LNAs
LNA	**LNA sequence**
Scrambled	GACGCGTCCTTACGCG
ESE2b	TCTTTGCTTTTCTTCT
ESS2b	CTACTTTTATGTCACTAT

755

756

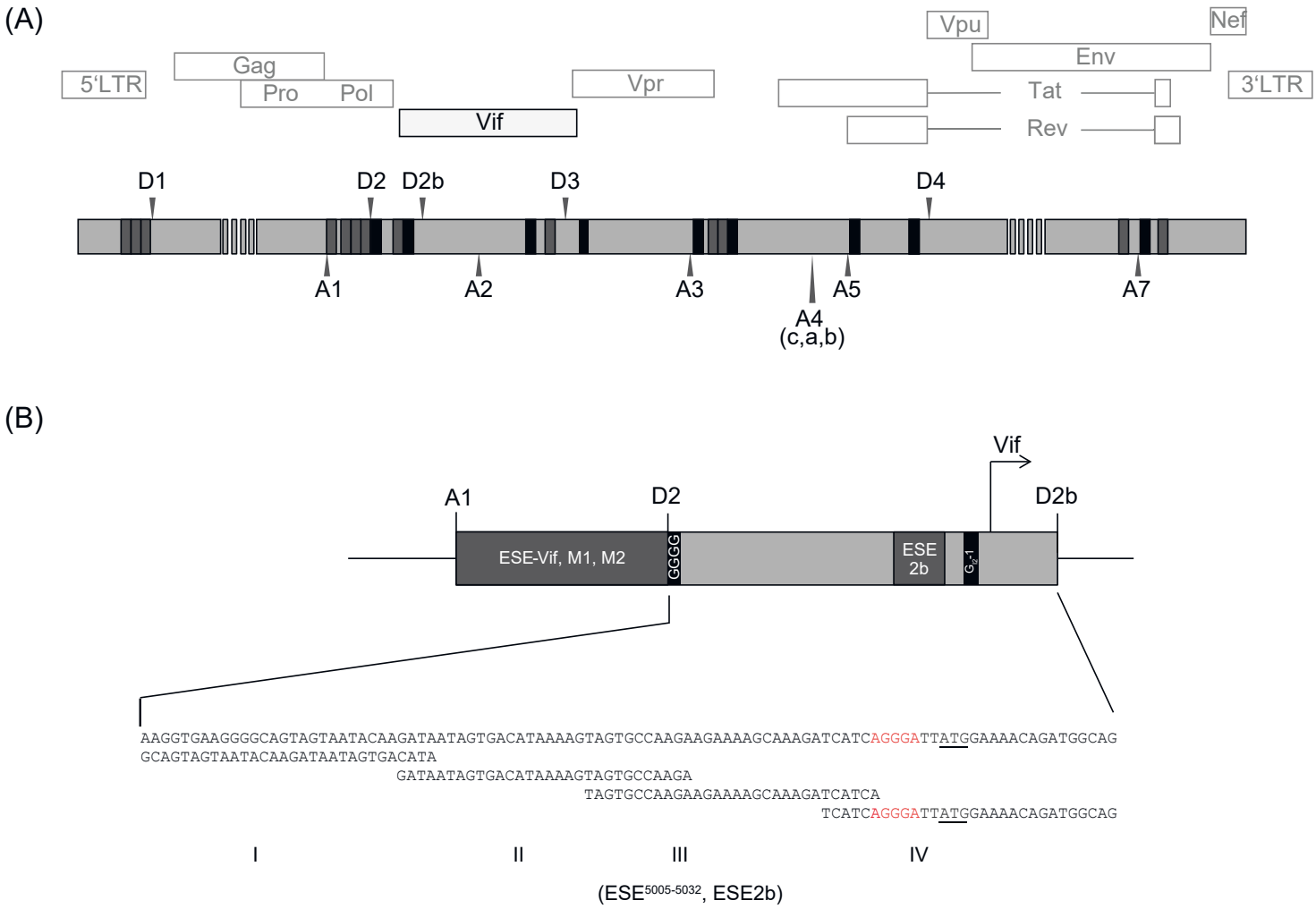


Figure 1

Analysis of splicing regulatory elements (SREs) in HIV-1 exon 2/2b. (A) Black (silencer) and grey (enhancer) bars indicate published SREs. Splice donor sites (D1—D4), splice acceptor sites (A1—A7) and protein ORFs are shown. (B) Known SREs within exon 2/2b and schematic of exon 2b parts I—IV. (C) Fluorescence microscopic analysis of fragments I—IV. Top: Schematic overview of the single-intron eGFP splicing reporter. Bottom: HeLa cells were transiently transfected with 1 µg of each construct together with 1 µg of pCL-dTOM to monitor transfection efficiency. 24 h after transfection, fluorescence microscopy was carried out.

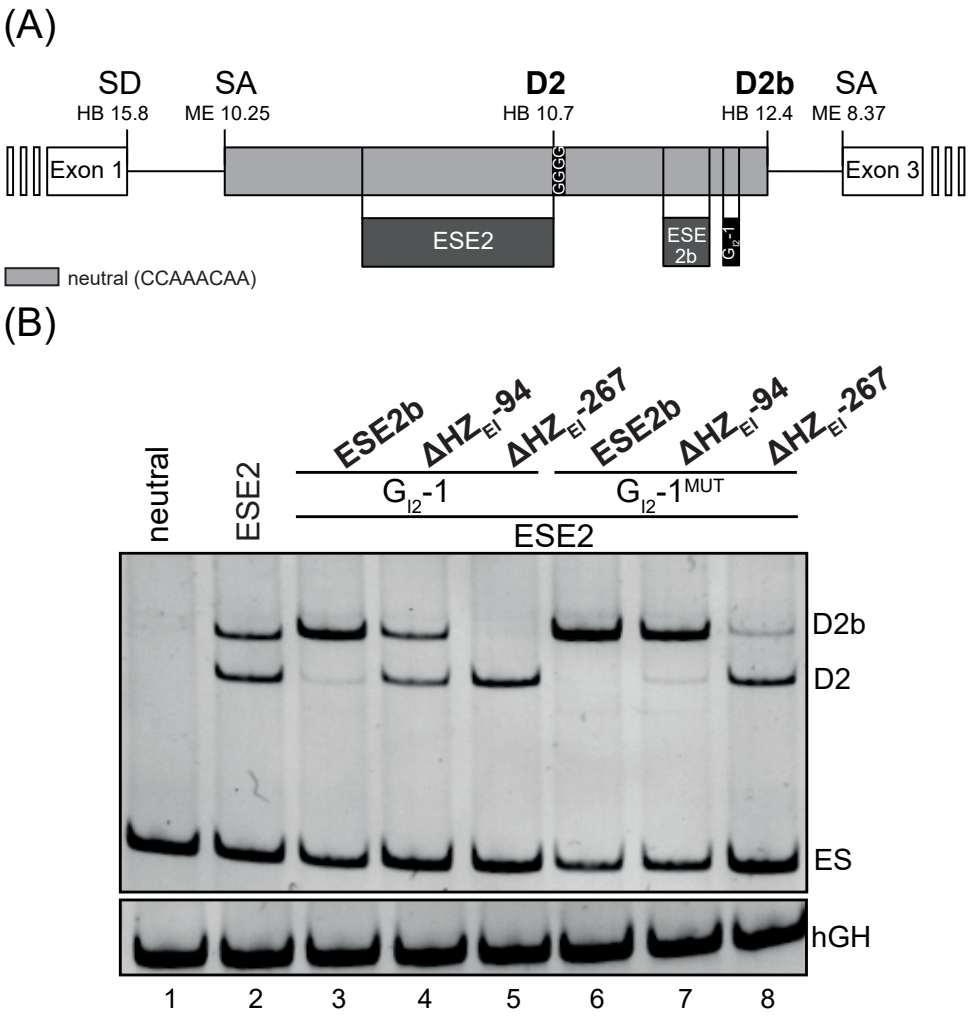


Figure 2

Impact of ESE2b on D2b recognition. (A) Schematic of the three-exon minigene. The middle exon is composed of only neutral CCAAACAA repeats (23) except for D2, D2b and the depicted SREs. (B) RT-PCR analyses of the minigene (A) splicing pattern. HeLa cells were transiently transfected with 1 µg of each construct and 1 µg of pXGH5. RNA isolated from cells were subjected to RT-PCRs using primer pairs #2648/#2649 and #1224/#1225 (hGH). PCR amplicons were separated on a non-denaturing 10% polyacrylamide gel and stained with ethidium bromide. (C) HZEI plots of ESE2b, ΔHZEI = -94 and ΔHZEI= -267 (black: mutated sequence; blue: wild-type reference).

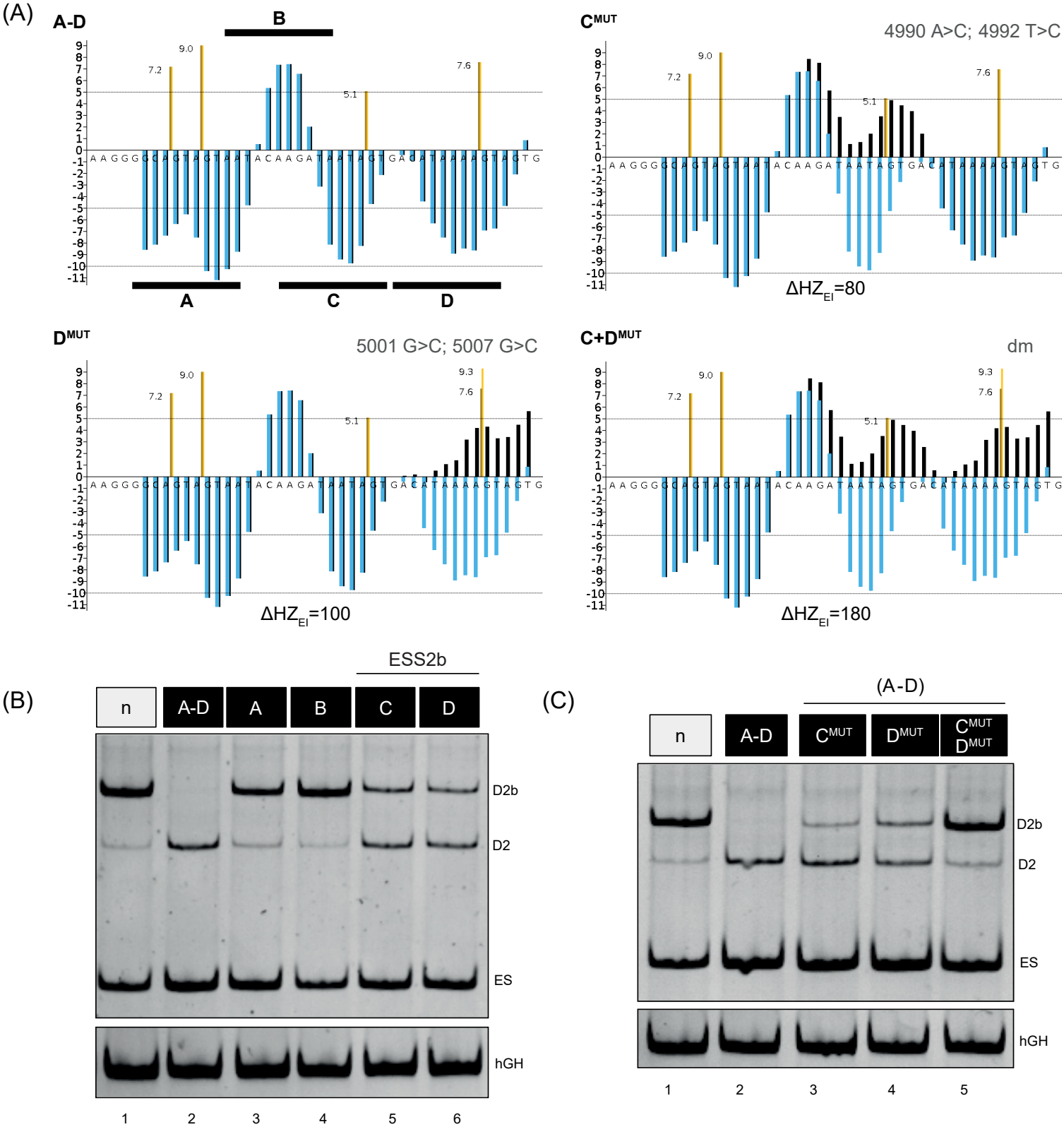


Figure 3

ESS2b located between D2 and ESE2b is bound by members of hnRNP A/B family and counteracts ESE2b. (A) HEXplorer score profiles of sequence A—D (indicated by black bars) and mutations of fragments C, D, or both (black: mutated sequence; blue: wild-type reference) composing ESS2b. (B-C) Mutational analysis of ESS2b. HeLa cells were transiently transfected with 1 μg of each construct and 1 μg of pXGH5. Twenty-four hours after transfection, RNA was isolated from the cells and subjected to RT-PCR analysis using primer pairs #2648/#2649 and #1224/#1225 (hGH).

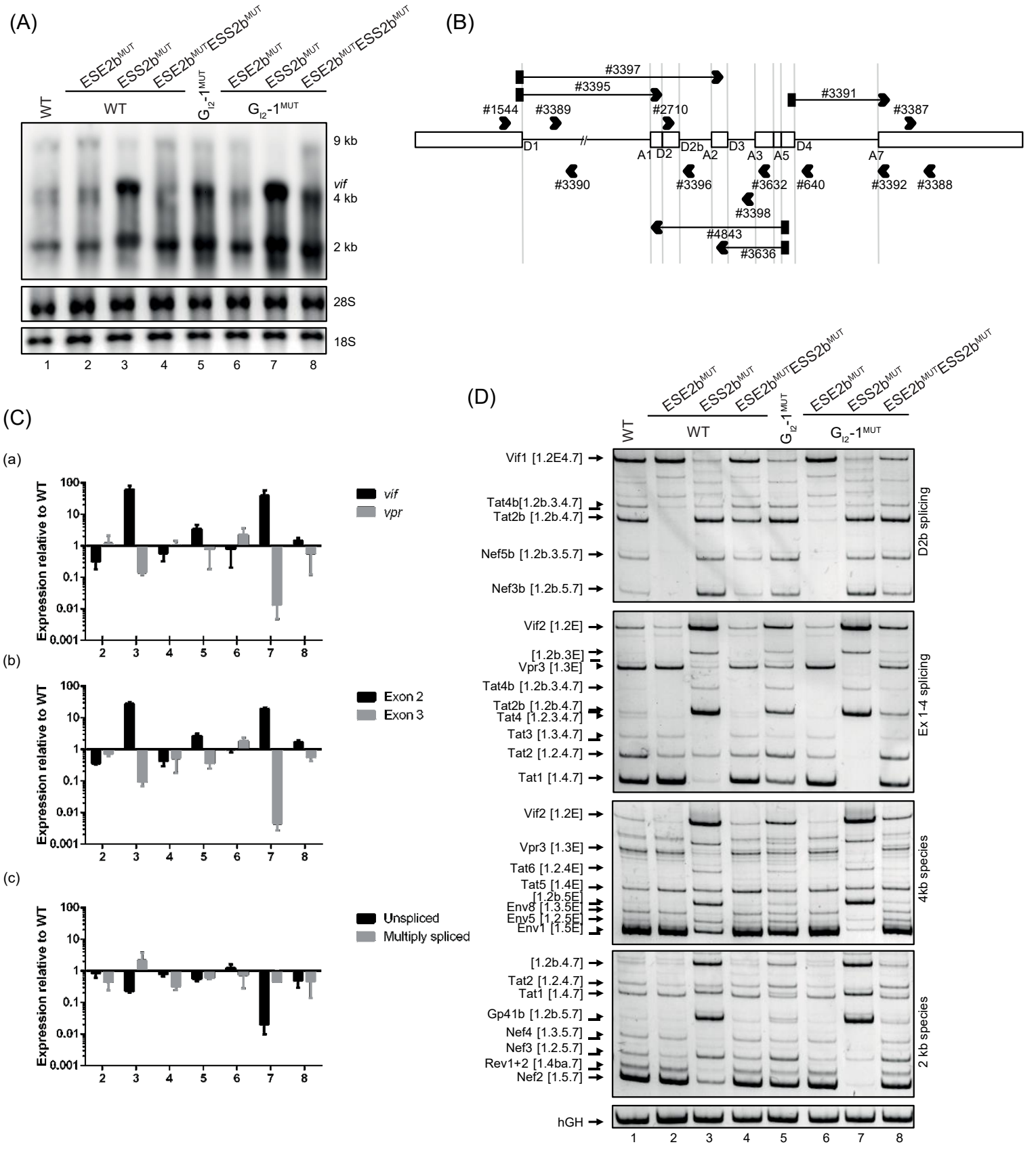


Figure 4

ESE2b and ESS2b cause alterations in proviral pre-mRNA processing. (A) Northern blot analysis of total RNA isolated from HEK293T CD4+ cells transfected with wild-type or mutant pNL4-3. A hybridization probe was used specifically detecting HIV-1 exon 7. (B) Binding sites of (q)RT-PCR primers. (C) qRT-PCR of total RNA isolated from the same RNA preparation as in (A) to specifically quantitate the levels of (a) vif vs. vpr, (b) [1.2.5] vs. [1.3.5.] and (c) multiply spliced vs. unspliced mRNA species, displaying exp(-Δct) ratios normalized to wild-type splicing pattern. Bar graphs show mean and standard deviation of three replicates. Primer pair #3387/#3388 specifically detecting exon 7 was used for normalization. Following primer pairs were used: vif: #3395/#3396; vpr: #3397/#3398; [1.2.5]: #3395/#4843; [1.3.5]: #3397/#3636; multiply spliced: #3391/#3392; unspliced: #3389/#3390. (D) RT-PCR analysis of RNA used in (A) and (C). Primer pairs: #1544/#3632 (Ex1-4 splicing), #2710/#3392 (D2b splicing), #1544/#3392 (2kb species), #1544/#640 (4kb species). HIV-1 mRNA species are indicated on the left hand side of each gel image according to (3). Exon numbers are indicated in square brackets; those including an E read through D4.

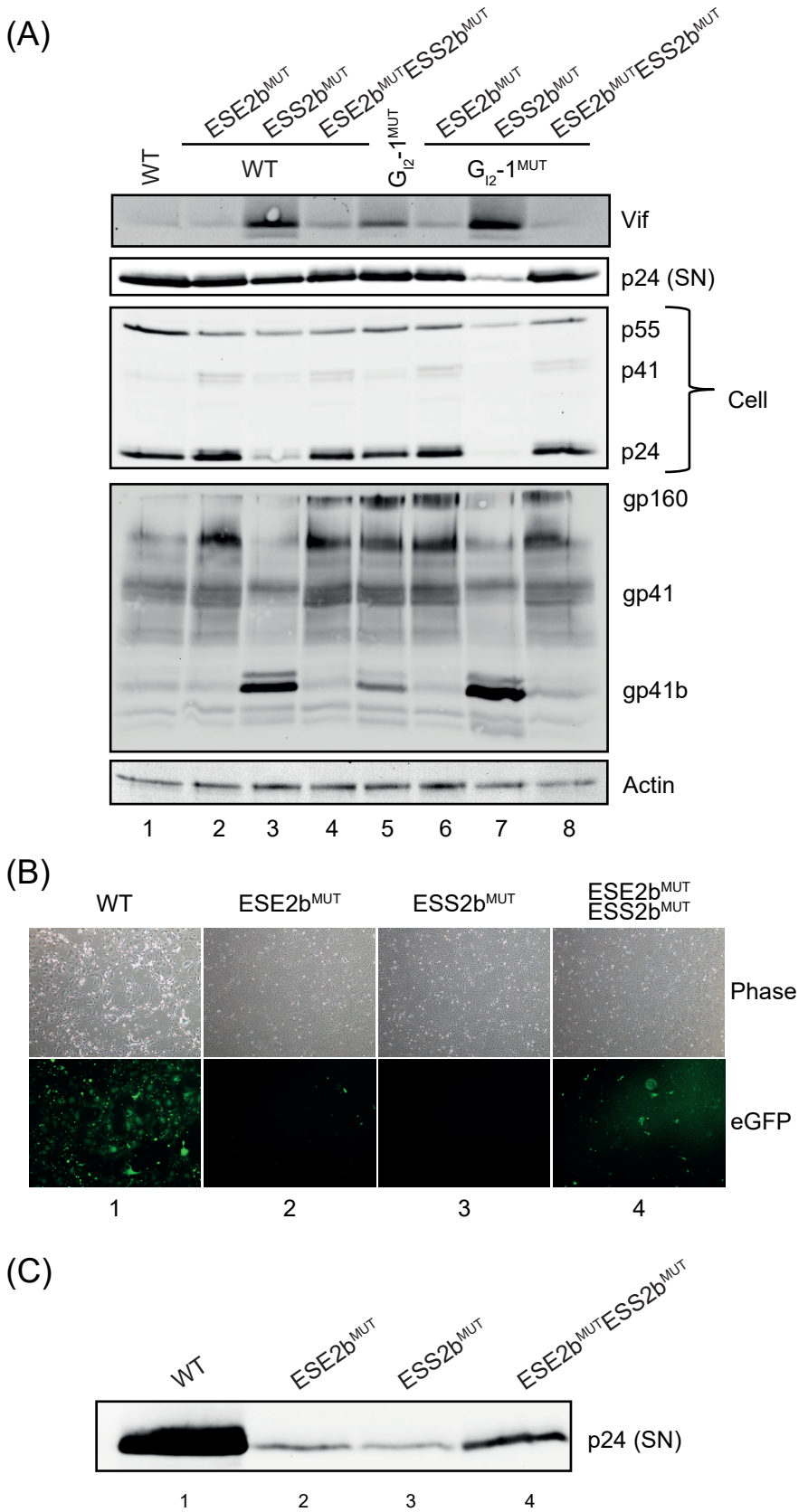


Figure 5

Impairment of proper viral particle production. (A) Immunoblot analysis of proteins of pelleted virions from the supernatant (SN) of transfected cells described in Fig. 4. (B-C) 2.5 × 105 HEK293T cells were transfected with pNL4-3 and mutant proviruses. 48h post transfection, the supernatant was collected for infection of GHOST CD4+ cells, an indicator cell line which expresses eGFP after HIV-1 infection. Infection and viral replication was analyzed 48h post infection, both by fluorescence microscopy (B) and by p24-gag Western blot analysis (C) of supernatants of the infected GHOST CD4+ cells.

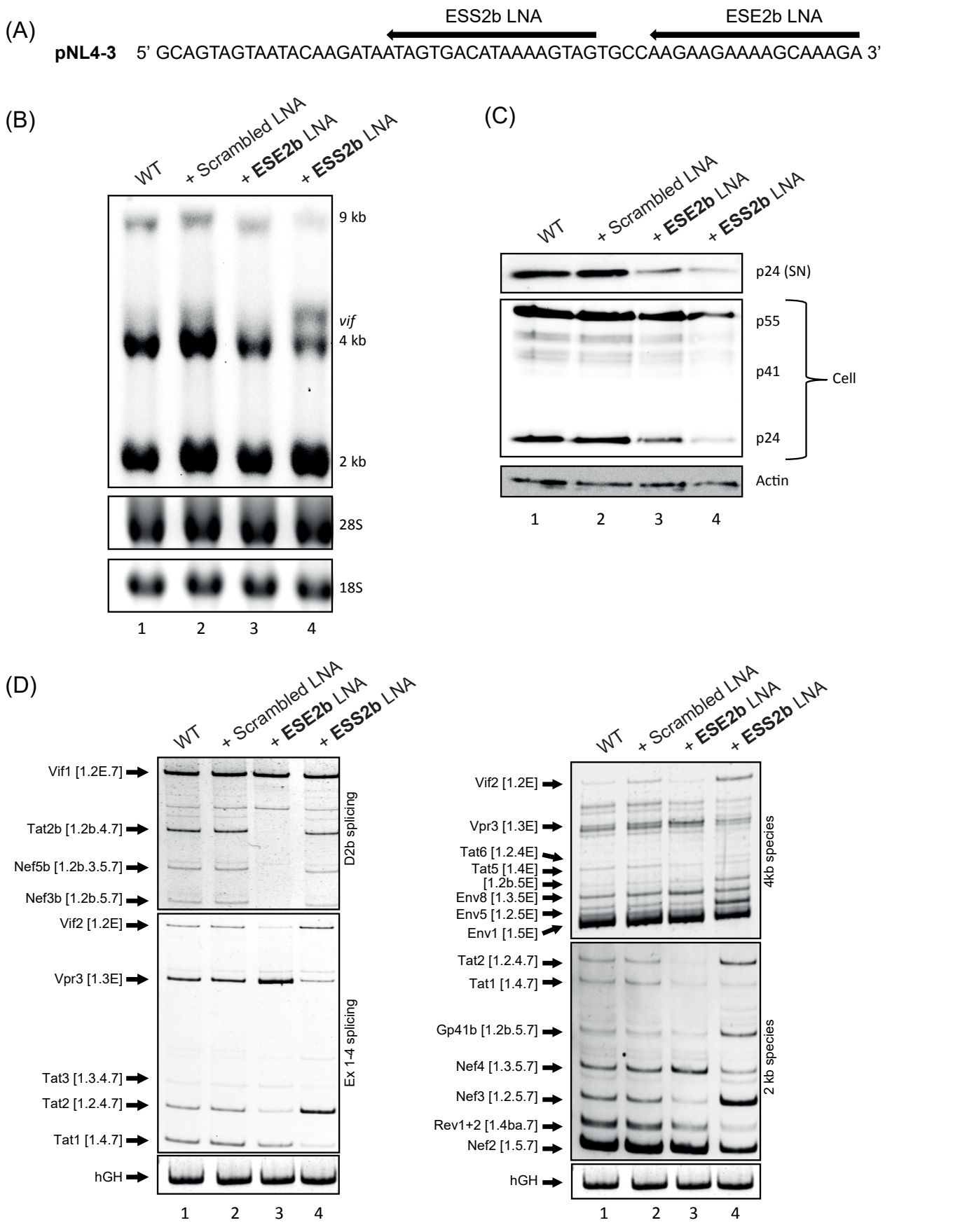


Figure 6

LNA-directed masking of ESE2b and ESS2b mimics mutational phenotype. (A) Schematic of the LNA binding sites. (B) Northern blot analysis of total RNA. HeLa cells were co-transfected with pNL4-3 and either LNAs masking ESE2b or ESS2b or the scrambled LNA. Total RNA was isolated 24 h post transfection and subjected to Northern blot analysis using an HIV-1 exon 7 probe. (C) Western blot analysis of cellular (Cell) and supernatant (SN) Gag of co-transfected cells from (B). (D) RT-PCR analysis of different viral mRNA species. Following primer pairs were used: #1544/#3632 (Ex1-4 splicing), #2710/#3392 (D2b splicing), #1544/#3392 (2kb species), #1544/#640 (4kb species). HIV-1 mRNA species are indicated on the left hand side of each gel image according to (3). Exon numbers are indicated in square brackets; those including an E read through D4.

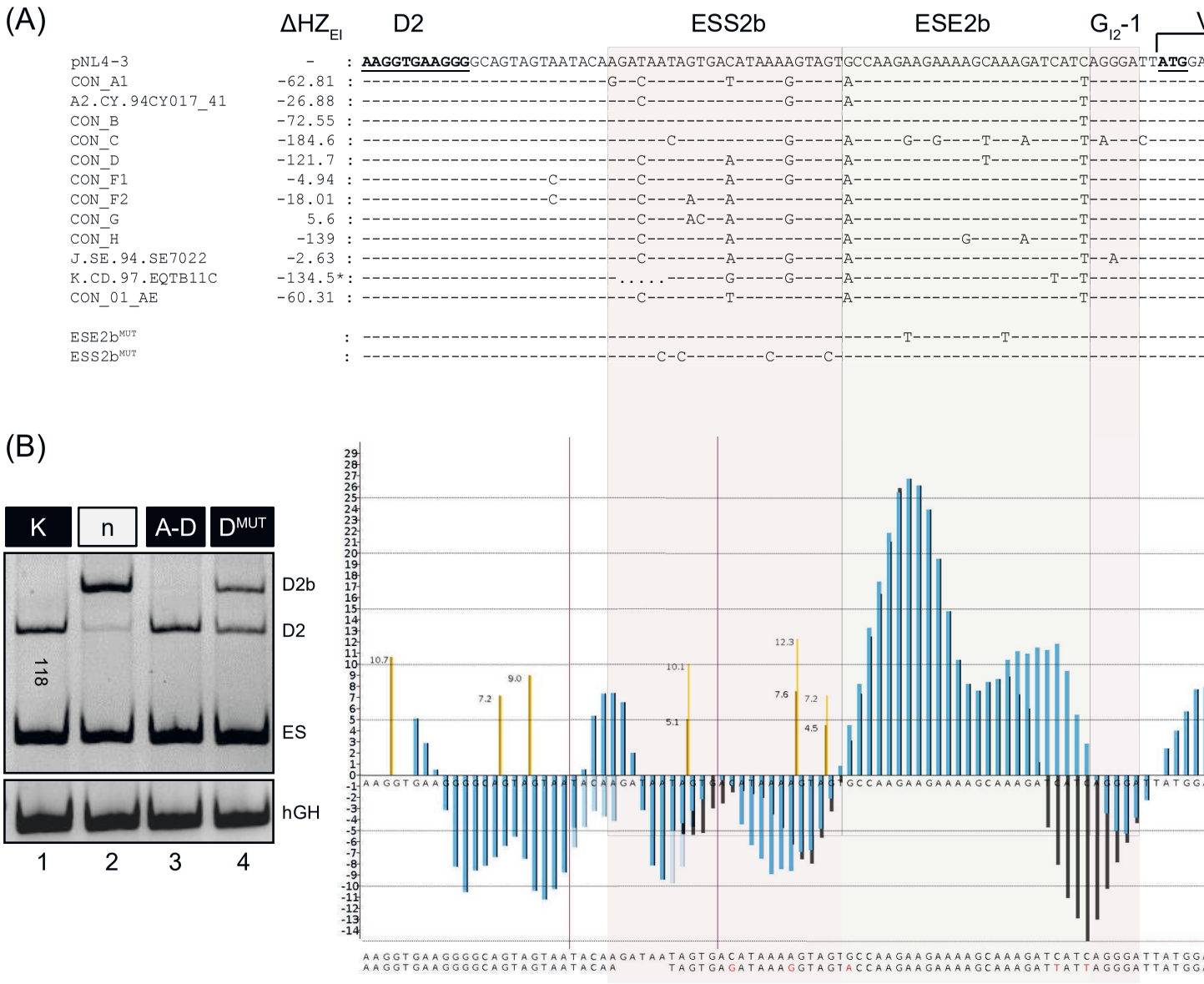


Figure 7

Analysis of SREs within exon 2/2b of different HIV-1 subtypes. (A) pNL4-3-derived HIV-1 exon 2/2b consensus to AE of the different HIV-1 subtypes, together with their HEXplorer score differences ΔHZEI. Conserved sequen by "–" and differences by letters. Regions with SREs are shown with red or green background. The subtype seq analyzed with the RIP 3.0 software (http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html). (B) Left: Splicing pa reporter carrying SRE regions of subtype K (lane 1) and pNL4-3 (lane 3). For reference, lanes 2 and 4 correspo sequence and to DMUT are also shown. HeLa cells were transiently transfected with 1 µg of each construct and h after transfection, RNA was isolated from the cells and subjected to RT-PCR analysis using primer pairs #264 #1224/#1225 (hGH). Right: HEXplorer profiles of pNL4-3 and exemplary subtype K containing a 5 nt deletion (b lines) and several single nt variations. Blue bars depict HEXplorer profile for pNL4-3 and black bars for subtype

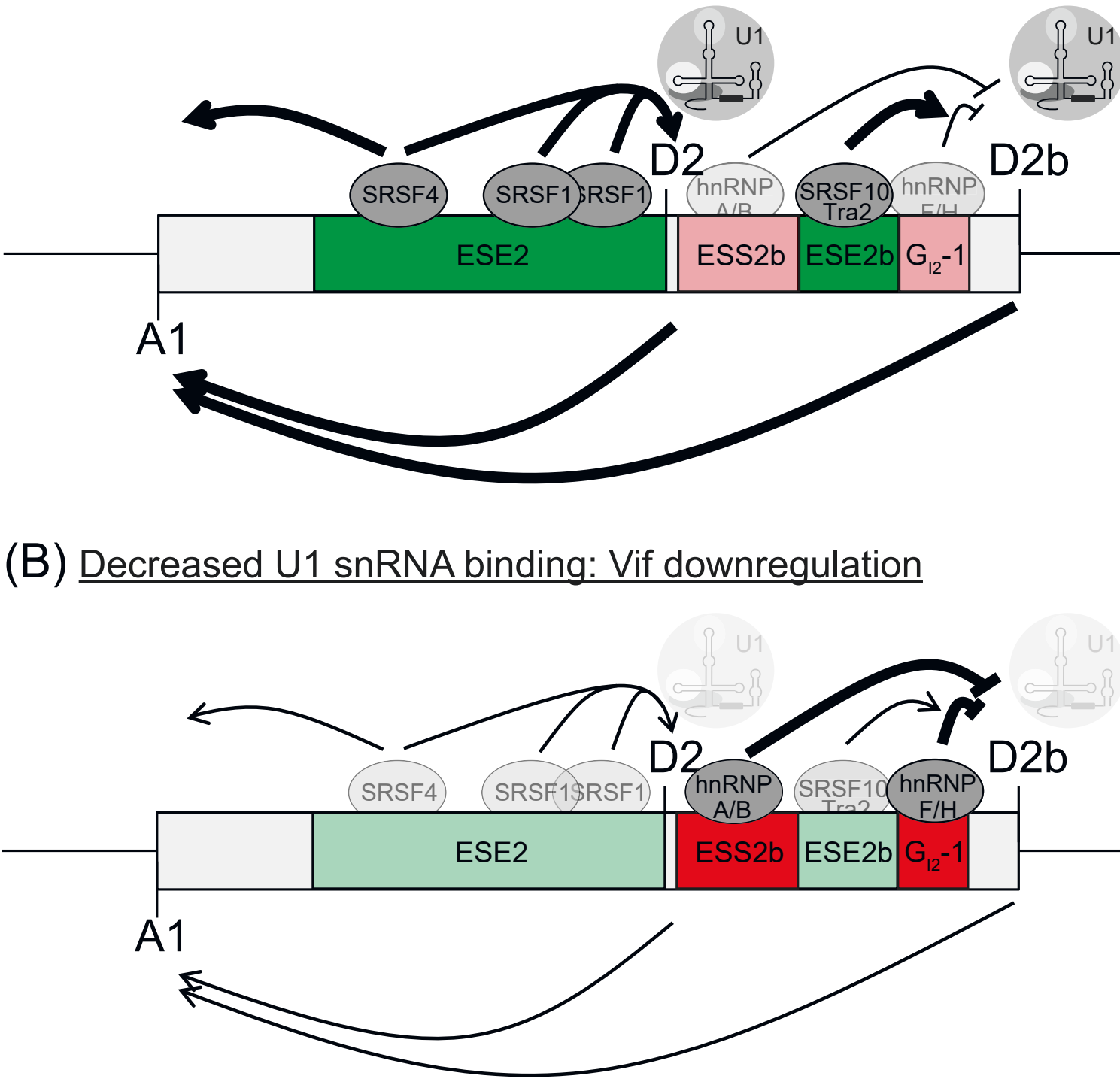


Figure 8

Model for exon2/2b recognition. Exon 2/2b inclusion and splice donor usage is regulated by a complex network of SREs. (A) SR proteins binding to both ESE2 and ESE2b support U1 snRNP binding at the downstream located splice donors D2 and D2b. Exon definition leads to the concomitant upregulation of splice acceptor A1, and to higher vif mRNA expression (left-pointing arrows below exon 2/2b). (B) Lower levels of SR proteins as well as hnRNP binding to sites ESS2b and GI2-1 reduce U1 snRNP binding to D2 and D2b.

5. Differential hnRNP D isoform incorporation may confer plasticity to the ESSV-mediated repressive state across HIV-1 exon 3

The following data are published in Biochim Biophys Acta. 2017 Feb;1860(2):205-217. (doi: 10.1016/j.bbagrm.2016.12.001) by

Hillebrand, F, Peter, J.O., ***Brillen, A.L.****, Otte, M., Schaal, H., Erkelenz, S.*

Contribution

FH, JP and SE conceived and designed MS2 tethering experiments, HIV-related transfection and readout experiments, designed and performed expression and RNA-pull-down analysis. ALB designed and performed cloning of the MS2/SR fusion protein expressing plasmids and related MS2 tethering experiments. HS and SE conceived the study, supervised its design and its coordination, and wrote the manuscript.

Abstract

Even though splicing repression by hnRNP complexes bound to exonic sequences is well-documented, the responsible effector domains of hnRNP proteins have been described for only a select number of hnRNP constituents. Thus, there is only limited information available for possible varying silencer activities amongst different hnRNP proteins and composition changes within possible hnRNP complex assemblies. In this study, we identified the glycine-rich domain (GRD) of hnRNP proteins as a unifying feature in splice site repression. We also show that all four hnRNP D isoforms can act as genuine splicing repressors when bound to exonic positions. The presence of an extended GRD, however, seemed to potentiate the hnRNP D silencer activity of isoforms p42 and p45. Moreover, we demonstrate that hnRNP D proteins associate with the HIV-1 ESSV silencer complex, probably through direct recognition of "UUAG"

sequences overlapping with the previously described "UAGG" motifs bound by hnRNP A1. Consequently, this spatial proximity seems to cause mutual interference between hnRNP A1 and hnRNP D. This interplay between hnRNP A1 and D facilitates a dynamic regulation of the repressive state of HIV-1 exon 3 which manifests as fluctuating relative levels of spliced vpr- and unspliced gag/pol-mRNAs.

Biochimica et Biophysica Acta 1860 (2017) 205–217

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Differential hnRNP D isoform incorporation may confer plasticity to the ESSV-mediated repressive state across HIV-1 exon 3

Frank Hillebrand [a,1], Jan Otto Peter [a,1], Anna-Lena Brillen [a], Marianne Otte [c], Heiner Schaal [a,*], Steffen Erkelenz [a,b,**]

[a] Institute of Virology, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany
[b] Institute for Genetics and Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany
[c] Institute of Evolutionary Genetics, Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany

ARTICLE INFO

Article history:
Received 27 July 2016
Received in revised form 22 November 2016
Accepted 1 December 2016
Available online 3 December 2016

Keywords:
HIV-1
Splicing
hnRNP proteins
hnRNP D

ABSTRACT

Even though splicing repression by hnRNP complexes bound to exonic sequences is well-documented, the responsible effector domains of hnRNP proteins have been described for only a select number of hnRNP constituents. Thus, there is only limited information available for possible varying silencer activities amongst different hnRNP proteins and composition changes within possible hnRNP complex assemblies. In this study, we identified the glycine-rich domain (GRD) of hnRNP proteins as a unifying feature in splice site repression. We also show that all four hnRNP D isoforms can act as genuine splicing repressors when bound to exonic positions. The presence of an extended GRD, however, seemed to potentiate the hnRNP D silencer activity of isoforms p42 and p45. Moreover, we demonstrate that hnRNP D proteins associate with the HIV-1 ESSV silencer complex, probably through direct recognition of "*UUAG*" sequences overlapping with the previously described "*UAGG*" motifs bound by hnRNP A1. Consequently, this spatial proximity seems to cause mutual interference between hnRNP A1 and hnRNP D. This interplay between hnRNP A1 and D facilitates a dynamic regulation of the repressive state of HIV-1 exon 3 which manifests as fluctuating relative levels of spliced *vpr*- and unspliced *gag/pol*-mRNAs.

1. Introduction

Following integration into the host genome, HIV-1 transcription starts at the 5′-LTR promoter. Multiple spliced viral mRNAs with a size of approximately 1.8 kb are the first HIV mRNAs detectable within the cytoplasm of the infected cell and are translated into the HIV-1 regulatory proteins Tat, Rev and Nef (Fig. 1A; [1–3]). Following relocalization into the nucleus, Tat induces processive transcription, whereas Rev mediates the nuclear export of intron-containing (~4 kb) and unspliced (~9 kb) viral mRNAs, which normally would be retained within the nucleus (Fig.1A; [4], for a recent review see [5]). This in turn allows the expression of structural (Gag, Env), enzymatic (Pol) and accessory viral proteins (Vpr, Vif, Vpu). Furthermore, as the unspliced RNA accumulates, the genomic RNA can be packaged into newly formed virus particles. Viral protein expression is CAP-dependent, i.e. the ribosomal 40S subunit enters the RNA at its 5′-end RNA and scans along the transcript until it encounters an appropriate translational start codon. Selection of alternative viral 3′'ss upstream of each HIV-1 open reading frame (ORF) determines which of the viral AUGs is pieced together with the 5′-end of respective viral RNAs and thus which viral protein can be efficiently synthesized (Fig. 1A). Viral splice site selection is controlled by positive and negative splicing regulatory elements (SREs), which are distributed across the HIV-1 genome and are mostly found to be positioned in the direct vicinity of splice sites (see Supplementary Fig. 1; for a recent review see [6]). Nuclear RNA binding proteins, which belong to the protein inventory of an infected cell, act through these SREs to enhance or repress nearby splice sites (for a review see [7,8]). The two major splicing factor families are the serine-arginine (SR) proteins and the heterogenous ribonucleoproteins (hnRNPs). While hnRNPs repress splicing when bound to an exon and activate splicing following relocation to the opposite intron, SR proteins show a reversed position-dependence [9]. SR and hnRNP proteins possess a modular domain organization, which includes the presence of at least one RNA recognition motif (RRM) for sequence-specific SRE binding. In addition, SR proteins carry a C-terminal arginine-serine (RS) rich domain of variable size which serves as an effector domain to interact with general splicing components during splice site activation [10–12]. Although several studies revealed that the glycine-rich domain (GRD) of individual hnRNP proteins functions as an analogous effector domain for the establishment of splice site repression (e.g. [13–15]), no systematic analyses

* Correspondence to: H. Schaal, Institute of Virology, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany.
** Correspondence to: S. Erkelenz. Institute for Genetics and Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany.
E-mail address: steffen.erkelenz@gmx.de (S. Erkelenz).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bbagrm.2016.12.001
1874-9399/

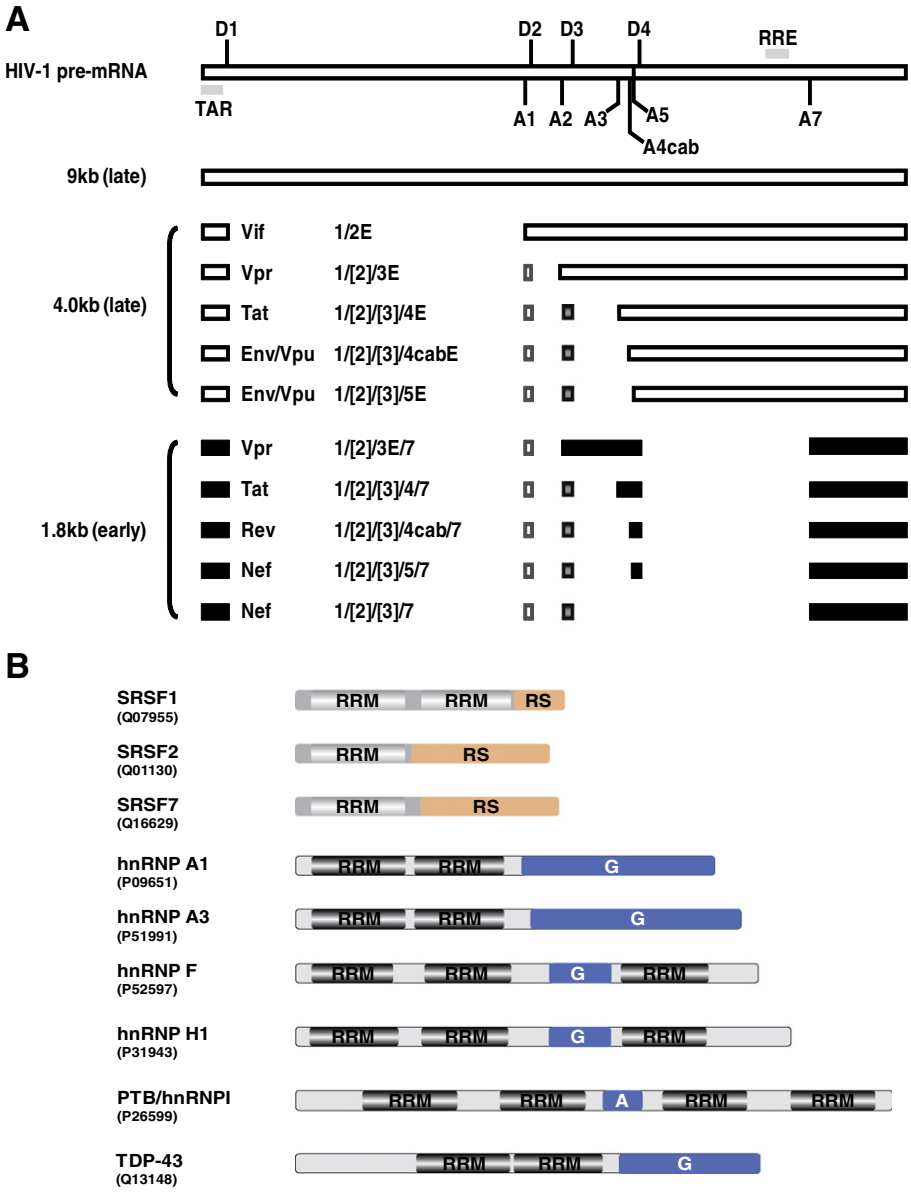


Fig. 1. Alternatively spliced HIV-1 mRNAs and sketch of SR and hnRNP proteins analyzed in this study. (A) Schematic drawing of the HIV-1 pre-mRNA with splice donor (D) and acceptor (A) sites as well as the *trans*-activating response (TAR) and the Rev-responsive elements (RRE). The exon/intron compositions of the viral unspliced 9 kb, the intron-containing 4 kb and the intron-less 1.8 kb mRNA classes are shown. E: Extended exon (B) Domain structure of splicing regulating proteins. UniProtKB accession numbers are provided below each protein. Candidate SR and hnRNP protein effector domains are highlighted in orange (RS: arginine-serine-rich domain) or blue (G: glycine-rich domain, A: alanine-rich domain). RRM: RNA recognition motif.

has been carried out so far to compare the repressive properties of different hnRNP-derived GRDs with one another. HIV-1 critically depends on coordinated interactions of cellular SR and hnRNP proteins with the viral pre-mRNA in order to guarantee the emergence of an intact viral transcriptome within the infected cell encompassing >40 different viral mRNAs. For example, since the translational Vpr start codon is localized within the downstream intron of exon 3, expression of Vpr-encoding mRNAs critically relies on activation of 3′ss A2, but silencing of 5′ss D3 (both flanking the noncoding HIV-1 leader exon 3). Splice site usage is negatively regulated by an exonic splicing silencer (ESSV), which is embedded within the center of exon 3 [16–18]. Recently, two independent transcriptome-wide studies confirmed that "*UAG*" motifs serve as cellular binding sites for hnRNP A1 [19,20]. The viral ESSV element contains three (pyrimidine) "*UAG*" motifs which have been previously shown to capture hnRNP A1 proteins, inhibiting splicing at the exon 3 splice sites [16]. Inactivation of ESSV causes a dramatic increase in the levels of exon 3-containing and *vpr*-mRNA species [18,21]. Excessive exon 3 splice site activation seriously perturbs the normal balance between spliced and unspliced viral mRNAs, leading to a paucity in the formation of unspliced RNAs and a severely impaired ability of the virus to replicate [18,21]. For this reason, silencing of exon 3 safeguards the accumulation of sufficient amounts of unspliced viral mRNA within the infected cell. However, it also needs to be leaky to occasionally permit the generation of singly spliced *vpr*-encoding mRNAs.

In this study, we show that the GRDs of different hnRNP proteins are all capable to inhibit splicing when recruited to an HIV-1 exon 3-containing splicing reporter. Furthermore, we show that hnRNP D proteins also possess a silencing activity and that the extent of splicing repression imposed by hnRNP D proteins positively correlates with the isoform-specific size of their C-terminal GRD. Finally, we provide evidence for specific hnRNP D binding to the ESSV within HIV-1 exon 3. In summary, we propose a model in which alternative hnRNP A/D

Table 1
Primers used for cloning.

Cloned construct	Primer	Sequence
Proviral HIV-1 plasmids		
Forward		
pNL4-3 "AD dm"	#3641	5′ TTT CAG AAT CTG CTA TAA GAA ATA CCA TAT TCT GAC GTA TAG TTC TTC CTC TGT GTG AAT ATC AAG CAG GAC ATA AC 3′
pNL4-3 "TR"	#4295	5′ CCC CCC CAG AAT CTG CTA TAA GAA ATA CCT TAG GGT TAG GGT TAG GGT TAG GGT TAG GTG AAT ATC AAG CAG GAC ATA ACA AGG 3′
pNL4-3 "ARE"	#4294	5′ CCC CCC CAG AAT CTG CTA TAA GAA ATA CCT ATT TAT TTA TTT ATT TAT TTA TTT ATT GTG AAT ATC AAG CAG GAC ATA ACA AGG 3′
pNL4-3 "neutral"	#4293	5′ CCC CCC CAG AAT CTG CTA TAA GAA ATA CCC CAA ACA ACC AAA CAA CCA AAC AAC CAA GTG AAT ATC AAG CAG GAC ATA ACA AGG 3′
pNL4-3 "A down"	#4296	5′ CCC CCC CAG AAT CTG CTA TAA GAA ATA CCA TAT TAG CAC GTA TAG TTA GTC CTA GCT GTG AAT ATC AAG CAG GAC ATA ACA AGG 3′
pNL4-3 "D up"	#4297	5′ CCC CCC CAG AAT CTG CTA TAA GAA ATA CCA TAC TAG GAC GTA TAG CTA GTC CTA GGT GTG AAT ATC AAG CAG GAC ATA ACA AGG 3′
Reverse	#2588	5′ CTT TAC GAT GCC ATT GGG 3′
MS2 fusion proteins expression plasmids		
Forward		
SV SD4/SA7 scNLS-MS2 ΔFG SRSF1 (FL)	#4229	5′ GTG GTG GAT CCT CGG GAG GTG GTG TGA TT 3′
	#4230	5′ CTA GAC TCG AGT TAT GTA CGA GAG CGA GAT CTG 3′
SV SD4/SA7 scNLS-MS2 ΔFG SRSF1 (ΔRS)	#4229	5′ GTG GTG GAT CCT CGG GAG GTG GTG TGA TT 3′
	#4231	5′ CTA GAC TCG AGT TAG GGC CCA TCA ACT TTA ACC 3
SV SD4/SA7 scNLS-MS2 ΔFG SRSF7 (ΔRS)	#4238	5′ GTG GTG GAT CCT CGC GTT ACG GGC GGT A 3′
	#4239	5′ CTA GAC TCG AGT TAA TGA CAA TCA TAA GCA TAA TGT CCC TTTTCG 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP A1 (ΔGRD)	#2754	5′ GGG GGG GGA TCC TCT AAG TCA GAG TCT CCT AAA GAG CCC 3′
	#2935	5′ CCC CCT CGA GTT ATT AAC TAG TTC TTT GGC TGG ATG AAG CAC 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP A1 (GRD)	#2881	5′ CCC CCG GAT CCG GTC GAA GTG GTT CTG GAA A 3′
	#2755	5′ GGG GGG CTC GAG TTA AAA TCT TCT GCC ACT GCC ATA GCT AC 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP A3	#2839	5′ GGT GGA TCC GAG GTA AAA CCG CCG C 3′
	#2840	5′ CCC CCT CGA GTT AGA ACC TTC TGC TAC CAT ATC 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP A3 (ΔGRD)	#2839	5′ GGT GGA TCC GAG GTA AAA CCG CCG C 3′
	#2938	5′ CCC CCT CGA GTT ATT AAC TAG TAG CAG ACT GCA TTT CTT GTT TAG AA 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP A3 (GRD)	#2883	5′ CCC CCG GAT CCG GAT CAC AGA GAG GTC GT 3′
	#2840	5′ CCC CCT CGA GTT AGA ACC TTC TGC TAC CAT ATC 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP H1 (ΔGRD)	#3449	5′ GTC ATA AGG ACC TGG CCG CTG 3′
	#3450	5′ GCT GCT AGG TCC TTC TAC TTT CCA GAG CAC AAC AGG 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP H1 (GRD)	#3451	5′ GCT GCT GGA TCC GGT CCT TAT GAC AGA CCT GGG 3′
	#3452	5′ AGC AGC CTC GAG GCC ACC ATC CCC GTA TCT G 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP F (ΔGRD)	#3445	5′ AGC AGC CAG TAG CTG AGT GGC CGC TGC ACG GAC ATG 3′
	#3446	5′ GCT GCT CAG CTA CTG TTT CAC AGT GCA GAG CAC CAC 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP F (GRD)	#3447	5′ GCT GCT GGA TCC GGG CCC TAT GAC CGG C 3′
	#3448	5′ AGC AGC CTC GAG CTC ACT GTC GCC GTA TCT GTG 3′
SV SD4/SA7 scNLS-MS2 ΔFG PTB (ΔARD)	#3830	5′ CCA GCA ATA CAG AAT TTC CGC TCT GCC CGG CCA TC 3′
	#3831	5′ GGA AAT TCT GTA TTG CTG GTC AGC AAC CTC AAC CCA GAG AGA GTC 3′
SV SD4/SA7 scNLS-MS2 ΔFG PTB (ARD)	#3653	5′ TTG GTG GAT CCC CCG TGC TCA GGA TCA TC 3′
	#3654	5′ TTG GTC TCG AGT GCC CCC GCC AGG CC 3′
SV SD4/SA7 scNLS-MS2 ΔFG TDP-43	#3176	5′ TGG TGG ATC CTC TGA ATA TAT TCG GGT AAC CGA A 3′
	#3177	5′ ATT ACT CGA GCT ACA TTC CCC AGC CAG AAG A 3′
SV SD4/SA7 scNLS-MS2 ΔFG TDP-43 (ΔGRD)	#3176	5′ TGG TGG ATC CTC TGA ATA TAT TCG GGT AAC CGA A 3′
	#3179	5′ ATT ACT CGA GCT AAC TTC TTT CTA ACT GTC TAT TGC TAT TG 3′
SV SD4/SA7 scNLS-MS2 ΔFG TDP-43 (GRD)	#3178	5′ TGG TGG ATC CGG AAG ATT TGG TGG TAA TCC AG 3′
	#3177	5′ ATT ACT CGA GCT ACA TTC CCC AGC CAG AAG A 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP D p37 – p45	#2465	5′ GGT GGA TCC TCG GAG GAG CAG TTC GGC 3′
	#2466	5′ AGA CTC GAG TTA GTA TGG TTT GTA GCT ATT TTG 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP D p37/p42 (ΔC)	#2465	5′ GGT GGA TCC TCG GAG GAG CAG TTC GGC 3′
	#2939	5′ CCC CCT CGA GTT ATT AAC TAG TCC ACT GTT GCT GTT GCT GAT 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP D p40/p45 (ΔC)	#2465	5′ GGT GGA TCC TCG GAG GAG CAG TTC GGC 3′
	#2939	5′ CCC CCT CGA GTT ATT AAC TAG TCC ACT GTT GCT GTT GCT GAT 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP D p37/p40 (C)	#2941	5′ CCC CGG ATC CGG ATC TAG AGG AGG ATT TGC 3′
	#2466	5′ AGA CTC GAG TTA GTA TGG TTT GTA GCT ATT TTG 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP D p42/p45 (C)	#2941	5′ CCC CGG ATC CGG ATC TAG AGG AGG ATT TGC 3′
	#2466	5′ AGA CTC GAG TTA GTA TGG TTT GTA GCT ATT TTG 3′
SV SD4/SA7 scNLS-MS2 ΔFG hnRNP D Exon 7	#3048	5′ GGT GGA TCC GGT GGC CCC AGT CAA AA 3′
	#3049	5′ CCC CCT CGA GTT ACT GCT GGT TGC TAT AAT CAC C 3′

silencer compositions might confer variations to the repressive state of exon 3 and thereby determine the relative formation of *vpr*-encoding mRNAs.

2. Material and methods

2.1. Oligonucleotides

Oligonucleotides were synthesized at Metabion GmbH (Martinsried, Germany).

Primers used for site-directed mutagenesis (**see** Tables 1 **in separate file**).

Primers used for semi-quantitative and quantitative RT-PCR analyses (**see** Table 2 **in separate file**).

Table 2
Primers used for semi-quantitative and quantitative RT-PCR.

Target RNA	Primer	Sequence
Viral mRNA classes	#1544 (E1)	5′ CTT GAA AGC GAA AGT AAA GC 3′
	#3392 (E7)	5′ CGT CCC AGA TAA GTG CTA AGG 3′
	#640 (I4)	5′ CAA TAC TAC TTC TTG TGG GTT GG 3′
	#3632 (E4)	5′ TGG ATG CTT CCA GGG CTC 3′
All viral RNAs	#3387	5′ TTG CTC AAT GCC ACA GCC AT 3′
	#3388	5′ TTT GAC CAC TTG CCA CCC AT 3v
LTR ex 2 ex3 (reporter RNAs)	#1544	5′ CTT GAA AGC GAA AGT AAA GC 3′
	#2588	5′ CTT TAC GAT GCC ATT GGG 3′
GH1	#1224	5′ CCA CTC CTC CAC CTT TGA 3′
	#1225	5′ ACC CTG TTG CTG TAG CCA 3′

Table 3
Primers used for RNA *in vitro* binding assays.

RNA substrate	Primer	Sequence
forward (T7)	#4324	5′ TAA TAC GAC TCA CTA TAG G 3′
Reverse		
ESSV	#4325	5′ ACA CCT AGG ACT AAC TAT ACG TCC TAA TAT GGA CAT GGG TGA TCC TCA TGT CCT ATA GTG AGT CGT ATT A 3′
“D up”	#4326	5′ACA CCT AGG ACT AGC TAT ACG TCC TAG TAT GGA CAT GGG TGA TCC TCA TGT CCT ATA GTG AGT CGT ATT A 3′
“AD dm”	#4328	5′ACA CAG AGG AAG AAC TAT ACG TCA GAA TAT GGA CAT GGG TGA TCC TCA TGT CCT ATA GTG AGT CGT ATT A 3′
“ESS2”	#4329	5′ ACG GTT GTT TGG TCT AGT CTA GTT GTT TGG GGA CAT GGG TGA TCC TCA TGT CCT ATA GTG AGT CGT ATT A 3′
“neutral”	#4330	5′ACT TGG TTG TTT GGT TGT TTG GTT GTT TGG GGA CAT GGG TGA TCC TCA TGT CCT ATA GTG AGT CGT ATT A 3′
“ARE”	#4331	5′ ACA ATA AAT AAA TAA ATA AAT AAA TAA ATA GGA CAT GGG TGA TCC TCA TGT CCT ATA GTG AGT CGT ATT A 3′
“TR”	#4332	5′ ACC TAA CCC TAA CCC TAA CCC TAA CCC TAA GGA CAT GGG TGA TCC TCA TGT CCT ATA GTG AGT CGT ATT A 3′
“A down”	#4333	5′ ACA CGT AGG AGT AAC TAT ACG TCG TAA TAT GGA CAT GGG TGA TCC TCA TGT CCT ATA GTG AGT CGT ATT A 3′

Primers used for RNA *in vitro* binding assays (see Table 3 in separate file).

2.2. Plasmids

MS2 fusion protein expressing plasmids were cloned by replacing the *Bam*HI, *Xho*I fragment of SV scNLS-MS2 ΔFG-ΔRS HA [22] with PCR products using appropriate forward and reverse primers (see Table 1 and [19,20]). Plasmids for the expression of Flag-tagged hnRNP D isoforms were generated by insertion of the respective *Bam*HI/*Xho*I fragments from SV scNLS-MS2 ΔFG hnRNP D p37-p45 into pcDNA3.1-Flag. Construction of the HIV-1-based subgenomic splicing reporter construct LTR ex2 ex3 (2xMS2) SD3down has been described previously [22]. Proviral HIV-1 exon 3 ESSV mutants were obtained using PCR mutagenesis. Therefore, the *Alw*NI/*Eco*RI fragment of LTR ex2 ex3 was replaced by PCR products using appropriate forward PCR primer (see Table 1) and #3632 as a reverse PCR primer containing *Alw*NI and *Eco*RI restriction sites. Finally, proviral HIV-1 variants were cloned by replacing the *Nde*I/*Eco*RI fragment of proviral plasmid pNL4-3 (GenBank Accession No. M19921). All cloned PCR amplicons were controlled by sequencing.

2.2.1. Culturing of cells and transient transfections

HEK293T were cultured in DMEM (Invitrogen) supplemented with 10% fetal calf serum (FCS) and 50 μg/ml of each penicillin and streptomycin (P/S) (Invitrogen). Plasmid transfections were performed using six-well plates with 2.5×10^5 HEK293T per plate and TransIT®-LT1 reagent (Mirus Bio LLC) following the manufacturer's instructions.

2.2.2. RNA extraction, RT-PCR and northern blot analyses

Total cellular RNA was harvested 48 h post transfection. For RT-PCR analyses RNAs were reversely transcribed with Superscript III Reverse Transcriptase (Invitrogen) and Oligo(dT) primer (Invitrogen). For analyses of LTR ex2 ex3 (2xMS2) SD3down-derived reporter mRNA splicing pattern PCR was carried out with primers #1544 and #2588. For viral *tat*-mRNAs and *vpr*-mRNA splicing, cDNA was used in a PCR reaction with primers #1544 (E1) and #3632 (E4) (see Table 2). For the analysis of intronless 2 kb HIV-1 mRNAs, a PCR reaction was carried out with forward primer #1544 (E1) and reverse primer #3392 (E7). Intron-containing 4.0 kb HIV-1 mRNAs were detected with primers #1544 (E1) and #640 (I4). All primer sequences used for semi-quantitative RT-PCR analyses are listed in Table 2. PCR products were separated on 8% non-denaturing polyacrylamide gels and stained with ethidium bromide for visualization. For Northern blot analysis of viral mRNAs, total RNA was separated on denaturating 1% agarose gels, transferred onto positively charged nylon membranes and probed with an digoxigenin (DIG)-labeled HIV-1 exon 7 PCR product (#3387/#3388).

2.3. Antibodies

The primary antibodies were used upon immunoblot studies as follows: Mouse antibody against hnRNP A1 (9H10) was obtained from Santa-Cruz Biotechnology. Rabbit antibody against hnRNP D (AUF1; 07–260) was purchased from Merck Millipore. Rabbit antibody against MS2 was provided by Tetracore (TC7004). Mouse antibody against α-actin (A2228) was obtained from Sigma-Aldrich. Sheep antibody against HIV-1 $p24^{Gag}$ was provided by Biochrom AG. Rabbit antiserum against Vpr was kindly provided through the NIH AIDS Research and Reference Reagent Program from Jeffrey Kopp. For detection, we used a horseradish peroxidase (HRP)-conjugated anti-mouse antibody (NA931) from GE Healthcare, a HRP-conjugated anti-rabbit antibody (A6154) from Sigma-Aldrich and a HRP-conjugated anti-sheep antibody from Jackson Immunoresearch Laboratories Inc.

2.4. Protein analysis

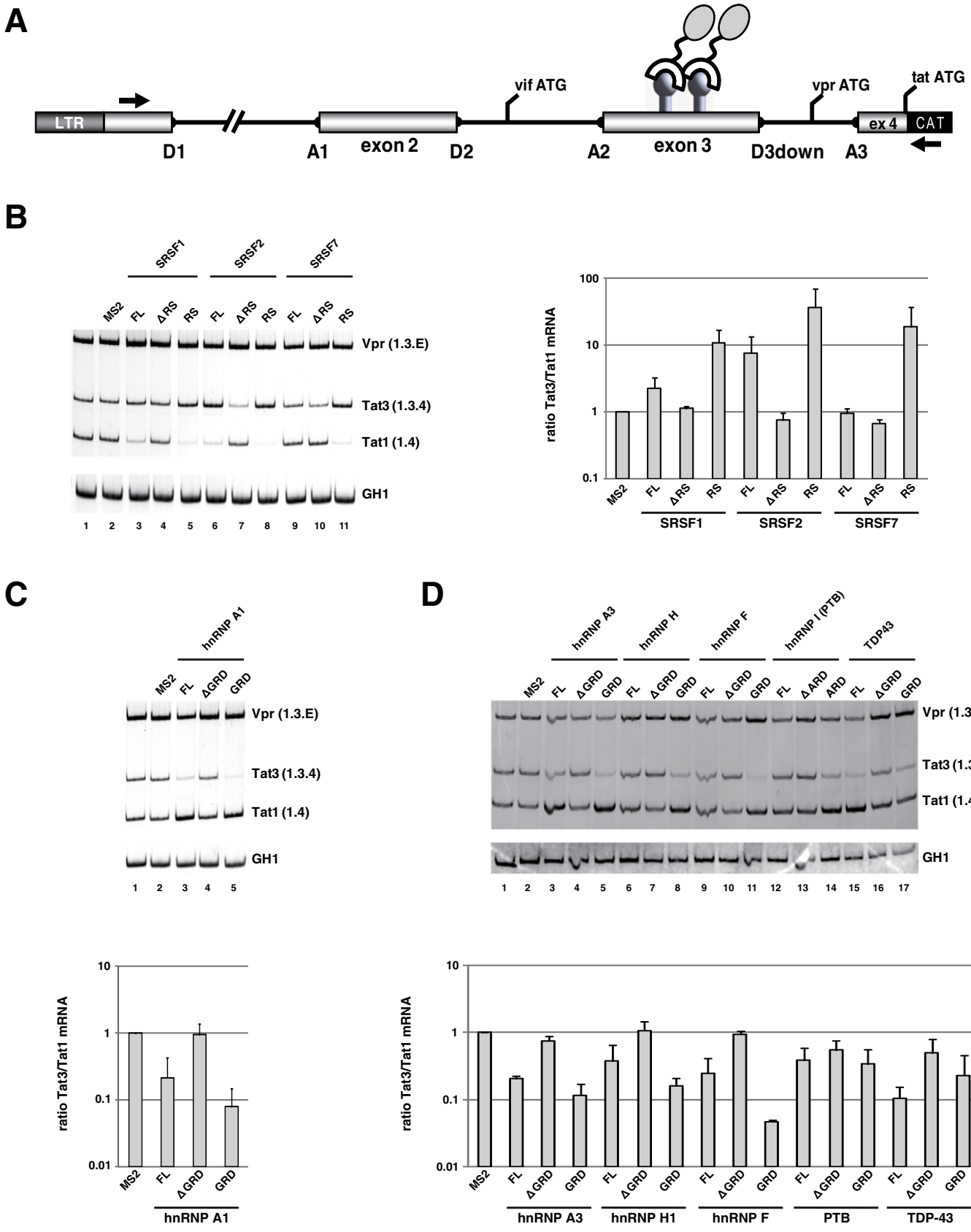
Cells were lysed in RIPA buffer (25 mM Tris·HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, protease inhibitor cocktail (Roche)). Subsequently, purified proteins were resolved using SDS polyacrylamide gel electrophoresis, blotted onto a nitrocellulose membrane, probed with specific antibodies and developed with ECL chemiluminescence reagents (GE Healthcare).

2.4.1. Covalent coupling of in vitro transcribed RNAs to agarose beads and RNA affinity chromatography (RAC) assays

For *in vitro* transcription of substrate RNAs, templates were generated by annealing of a sense T7 DNA primer with an anti-sense DNA oligonucleotide containing in 3′ to 5′ direction the T7 polymerase binding site, an MS2 RNA binding site and the wild-type or mutant ESSV, or the ESS2, TR, ARE or neutral sequence (see Table 3). RNA was synthesized using the RiboMax™ large scale RNA production system (P1300, Promega) according to the manufacturer's recommendations. Substrate RNAs were immobilized on adipidic acid dihydrazide-Agarose beads as previously described [23,24]. Subsequently, coupled RNAs were incubated in 15% HeLa cell nuclear extract (Cilbiotech)/buffer D (20 mM

Fig. 2. The glycine-rich domains (GRD) of hnRNP proteins mediate splicing repression. (A) Sketch of the HIV-1 based splicing reporter used for MS2 tethering experiments. The ESSV splicing silencer sequence [16] within HIV-1 exon 3 has been substituted for tandem RNA binding sites for the bacteriophage MS2 coat protein. In order to screen for both, negative and positive effects on exon 3 splicing, viral 5′ss D3 has been mutated to decrease its intrinsic strength (HBs 14.0 > 12.3, “D3 down”) and obtain almost equal levels of exon 3 inclusion versus exclusion in absence of MS2 fusion protein co-expression. Positions of primers #1544 (exon 1) and #2588 (CAT) used in RT-PCR analysis are indicated. LTR: Long Terminal Repeat; CAT: Chloramphenicol Acetyltransferase; E: Extended exon (B) RT-PCR analysis of spliced reporter mRNAs after co-expression of MS2/SR fusion proteins. Full-length (FL), deletion of the RS domain (ΔRS) or RS domain only (RS). 2.5×10^5 HEK293T cells were transiently transfected with 1 μg of each: the HIV-1 based splicing reporter, the respective MS2 fusion protein expressing plasmid and pXGH5 (expressing human growth hormone 1; GH1) to monitor transfection efficiency. Furthermore, viral Tat protein was co-expressed within the cells (SVctat) to drive efficient transcription from the viral LTR promoter. Subsequently, RNA was collected 48 h post transfection and subjected to RT-PCR analyses as described in the “Material and Methods” section. The bar graph represents the mean of the ratio between exon 3 including and skipping transcripts (ratio Tat3/Tat1 mRNA) and the standard error of the mean (SEM) relative to the “MS2” control (bar 1), which was set to 1. (C–D) RT-PCR analysis of spliced reporter mRNAs after co-expression of MS2/hnRNP protein fusions. HEK293T cells were co-transfected and RNA preparations were performed as described in (B). Data show the mean ± SEM of three (C, $n = 3$) or two (D, $n = 2$) independent experiments.

HEPES-KOH [pH 7.9], 5% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT) for a maximum of 20 min at 30 °C. Recombinant MS2 coat protein was added to nuclear extract dilutions to control for equal precipitation efficiencies. After washing off unspecifically bound proteins, the remaining fraction on the RNAs was eluted by addition of an equal volume of 2× protein sample buffer and heating at 95 °C for 10 min. Samples were



then separated by SDS-PAGE and transferred to nitrocellulose membranes for probing with specific antibodies.

3. Results

3.1. The glycine-rich domains (GRDs) of hnRNP proteins act as splicing effector domains eliciting splice site repression

Although hnRNP proteins are well-described repressors of exon inclusion, so far no systematic analysis has been carried out addressing whether hnRNP proteins exploit a common effector domain to repress splice site use. By contrast, arginine/serine (RS)-rich domains are in generally accepted to be responsible for splicing activation elicited by SR proteins bound to an exon [9,10]. However, a few studies indicated that the glycine-rich domains (GRDs) of hnRNP proteins are necessary for splicing repression [14,15,25,26]. To comparatively evaluate splicing activation and repression by RS and GRD domains derived from a selection of prominent SR and hnRNP proteins (Fig. 1B), we carried out MS2 tethering assays as described previously [22]. MS2 coat fusions were recruited to a tandem MS2 coat protein binding site, replacing the ESSV sequence within HIV-1 exon 3-containing reporter mRNAs (Fig. 2A) [16,22]. Substitution of the ESSV led to a nearly exclusive detection of transcripts including exon 3. Therefore, we reduced the intrinsic strength of viral 3′ss D3 (HBS 14.0 > 12.3, "D3 down"; www.uni-duesseldorf.de/rna [27]) and hence, could detect basal exon 3 splicing activity in absence of ESSV (Fig. 2B, lanes 1 and 2). This enabled us to measure up and down effects on splice site use following MS2 fusion protein binding.

In our screens, we tested full-length proteins (FL) for their influence on splice site selection as well as deletion variants lacking their (putative) effector domain (ΔRS or ΔGRD). Additionally, we assayed the effector domains alone (RS and GRD). As expected, all RS domains alone (SRSF1, SRSF2 and SRSF7) stimulated exon 3 inclusion (Fig. 2B, cf. lane 2 with lanes 5, 8 or 11), while SRSF fusion proteins lacking their RS domain failed to promote splice site usage (Fig. 2B, cf. lane 2 with lanes 4, 7 and 10). However, full-length SR proteins showed a somewhat less pronounced enhancement of exon 3 inclusion (Fig. 2B, cf. lanes 3 and 5; 6 and 8; 9 and 11). This might be explained by intramolecular sequestration of the RS domain by the unbound RRM [28]. For hnRNP A1, it has already been reported that the GRD essentially contributes to splice site repression [25,26], although a more recent study indicated that a subregion of the domain is dispensable for certain modes of splice site repression [29]. Consistent with our expectations, MS2-hnRNP A1 fusion proteins efficiently suppressed splicing when recruited to exon 3. This was evident by a shift from exon3-containing to exon3-skipped reporter mRNAs (Fig. 2C, cf. lanes 1–3). Deletion of the GRD domain resulted in the entire loss of repression as indicated by an exon inclusion/exclusion ratio which was comparable to the controls (Fig. 2C, cf. lanes 1 and 2 with 4). However, when the GRD was tested alone, we could barely detect exon 3 inclusion (Fig. 2C, cf. lanes 2 and 5, Tat3) consistent with previous data [25,26] and a role of the GRD as repressor domain.

Next we wished to test whether the GRDs of other hnRNP proteins (or the biochemically similar alanine-rich domain (ARD) of PTB) are equally capable of repressing exon 3 splicing. Indeed, we found that all of the tested GRDs were able to efficiently repress exon 3 inclusion (Fig. 2D, cf. lane 2 with 5, 8, 11, 14 or 17). Moreover, hnRNP proteins lacking their GRD domain either completely lost their inhibitory property (Fig. 2D, cf. lane 2 with lanes 4, 7, 10 or 13) or were at least less repressive (Fig. 2D, cf. lanes 2 and 16).

From these findings we concluded that all tested GRDs are able to potently repress splicing when tethered to an exonic position. Therefore, each GRD can be regarded as a general splicing repressor domain that represents the functional counterpart to the arginine-serine (RS)-rich domains of SR proteins.

3.2. hnRNP D acts as a repressor of exon inclusion

Previous studies indicated that hnRNP D controls splice site usage through binding of AU-enriched or "*UAGG*" motif-containing RNA target sequences [30,31]. However, it still remains open whether all or only individual hnRNP D isoforms are capable of regulating splice site selection. Alternative inclusion of exons 2 and 7 generates four different hnRNP D isoforms, p37, p40, p42 and p45 (according to their molecular weight) [32]. Interestingly, all four hnRNP D isoforms contain a C-terminal GRD, although of different lengths (Fig. 3A). To answer the question of whether the repressing activity of these domains correlate with their length, we performed MS2 tethering assays as described above. All four hnRNP D isoforms inhibited splicing when recruited to exon 3 (Fig. 3B, cf. lane 2 with lanes 3 to 6), which was in agreement with their proposed role as splicing repressors. However, the hnRNP D variants p42 and p45 showed a slightly higher potency to repress splicing (Fig. 3B, lanes 5 and 6, relative splicing efficiency: 0.18 and 0.23) when compared to their lower weight counterparts p37 and p40 (Fig. 3B, lanes 3 and 4, relative splicing efficiency: 0.33 and 0.55). We therefore wished to determine whether the exon 7-encoded extended GRD (Fig. 3A) might be responsible for the increased repressive activity of the longer isoforms p42 and p45. As it was seen for the other hnRNP proteins, the GRD seemed to be indispensable for splice site repression, since deletion mutants either lacking the short (p37 and p40) or the extended GRD (p42 and p45) failed to repress exon 3 inclusion (Fig. 3B, cf. lane 2 with lanes 7 and 8). Of note, the deletion mutants lacking the GRD were properly expressed (Supplementary Fig. 2). Interestingly, only the extended GRD present in hnRNP D p42 and p45 was independently able to efficiently repress exon 3 inclusion when tested alone (Fig. 3B, cf. lanes 9 and 10), suggesting different GRD-dependent repressor properties for the hnRNP D isoforms.

Repression by the extended GRD appeared to be entirely dependent on the presence of the exon 7-encoded amino acids. Consistently, the exon 7 peptide alone was sufficient to potently inhibit splicing in our MS2 tethering screen (Supplementary Fig. 3). Altogether, these findings support the notion that the repressor strength of the four hnRNP D isoforms depends on the presence of an extended GRD within the C-terminus of hnRNP D.

3.3. hnRNP D can repress splicing in the context of native target sequences

Based on an obvious overlap between the hnRNP D and hnRNP A1 RNA binding sites (hnRNP D: "*UUAGG/G*"; hnRNP A1: "*UAGGA/U*") and the common ability of both proteins to repress exon inclusion, we investigated whether hnRNP D might also be involved in ESSV-dependent HIV-1 exon 3 repression. The ESSV contains three "*UAG*" binding motifs, which were previously shown to be bound by hnRNP A1 proteins [16]. However, two of these motifs also match the consensus motif for hnRNP D proteins ("*UUAG*") and therefore might also or alternatively be bound by hnRNP D proteins (Fig. 4A). In addition, we found that similar to hnRNP A1, hnRNP D p45 could comparably repress exon 3 inclusion in hnRNP A1-negative cells (Supplementary Fig. 4). Based on these findings, it is likely that hnRNP D proteins contribute to exon 3 splicing regulation by at least partially competing with hnRNP A1 for binding to the ESSV. Either hnRNP D could functionally replace hnRNP A1 at the ESSV or co-assemble into the ESSV repressor complex to fine-tune exon 3 splice site selection.

In a first approach, we carried out RNA affinity chromatography (RAC) assays with RNA substrates containing the wild-type ESSV or a previously described hnRNP A1 binding mutant (ESSV^{-}), specifically screening for the presence of hnRNP D and hnRNP A1. It was previously demonstrated that the ESSV mutant dramatically increases exon 3 splicing in the context of the infectious clone pNL4-3 [18,21]. Exon 3 oversplicing was accompanied by a strong reduction in the levels of unspliced viral mRNAs, which severely impaired viral Gag expression and the formation of new virus particles [18,21]. RNA oligonucleotides

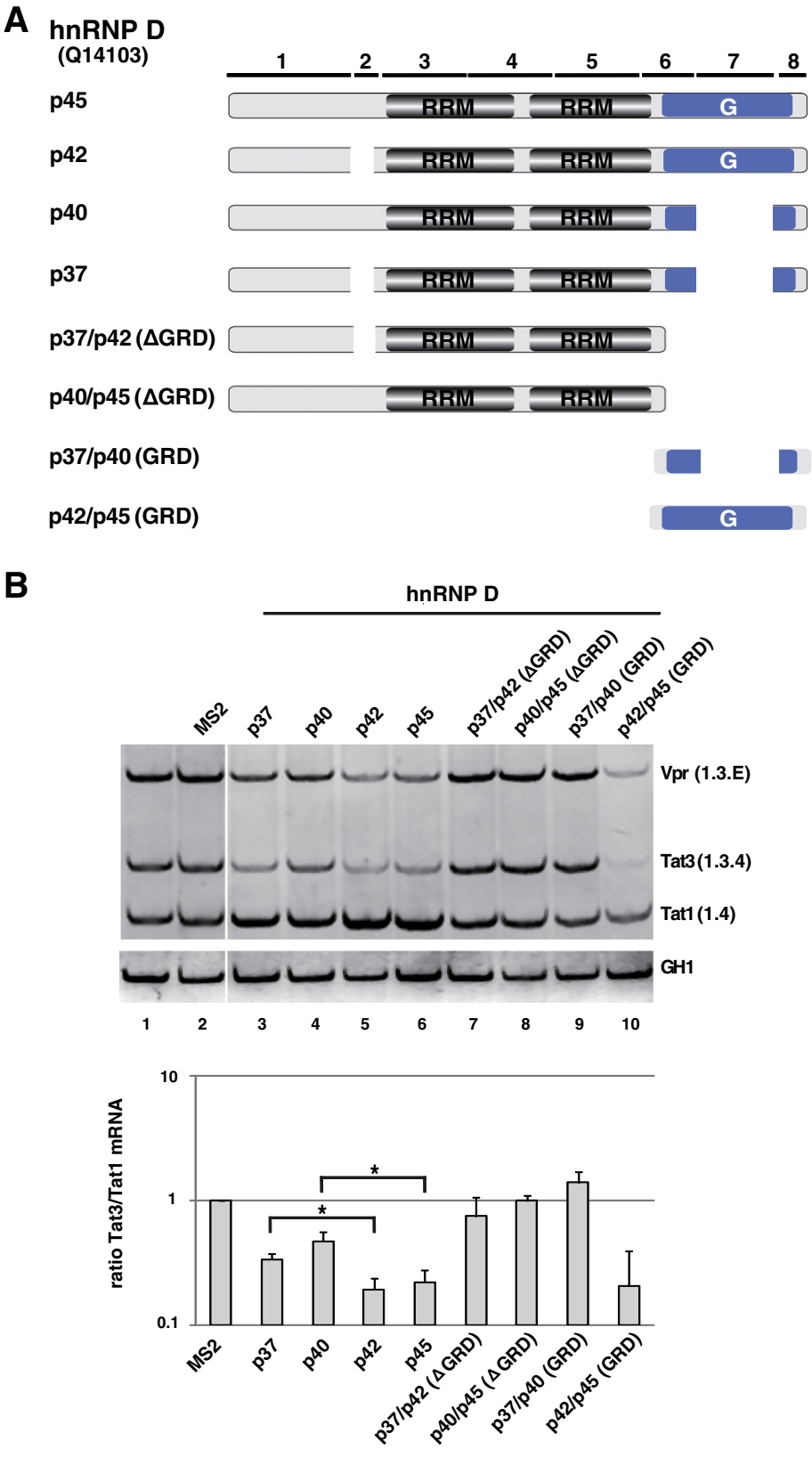


Fig. 3. hnRNP D proteins can efficiently inhibit splicing when tethered to a reporter exon. (A) Scheme of the hnRNP D isoforms and derived variants used in MS2 tethering experiments. (B) Top: RT-PCR analysis of spliced reporter RNAs following MS2/hnRNP D fusion protein co-expression. Experiments were performed as described in Fig. 2. Cropped gel image: all samples were run on the same gel. Bottom: Quantification of the RT-PCRs. The bar graph shows the mean ± SEM from three independent transfections experiments ($n = 3$; * $p < 0.05$).

were *in vitro* transcribed, immobilized on agarose beads and incubated in HeLa nuclear extracts as previously described [23,24]. Each RNA substrate was equipped with a single copy of an MS2 coat protein binding site. MS2 coat protein was added as recombinant protein to the nuclear extracts to control for precipitation efficiencies. All four hnRNP D isoforms could be precipitated with wild-type mRNA. Precipitation

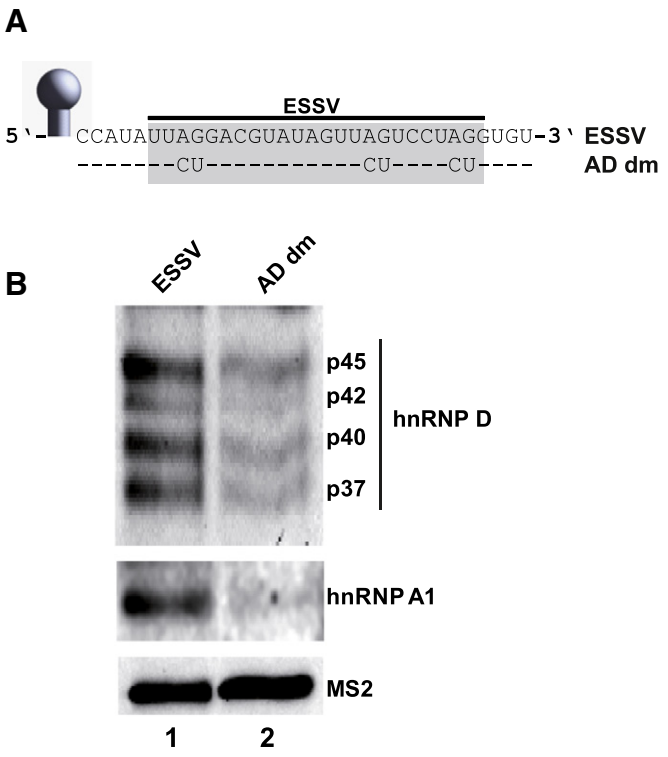


Fig. 4. hnRNP D protein isoforms specifically associating with ESSV. (A) Wild-type and mutant ESSV *in vitro* RNA substrates used in RNA affinity chromatography (RAC) experiments. Mutated nucleotides are indicated below the wild-type reference (pNL4-3) at corresponding positions and "-" denotes wild type nucleotide. (B) RAC assay using RNAs containing wild-type (ESSV) or mutant ESSV (AD dm) sequence. *In vitro* transcribed RNAs were immobilized on agarose beads and incubated with HeLa nuclear extract containing recombinant MS2 coat protein. After washing, specifically bound proteins were resolved by SDS PAGE and transferred onto nitrocellulose membranes for probing with specific antibodies against hnRNP A1, hnRNP D and MS2.

efficiency was greatly reduced when mutated RNA substrates were used (Fig. 4B, cf. lanes 1 and 2), confirming an interaction of hnRNP D with the ESSV sequence. We also confirmed hnRNP A1 binding on wild-type but not on the ESSV^{-} substrate RNA (Fig. 4B, lanes 1 and 2). Therefore, the ESSV mutant sequence will from now on be referred to as AD double mutant ("AD dm").

Although we show that all hnRNP D isoforms can bind the ESSV sequence, this does not necessarily indicate a functional significance for the hnRNP D•ESSV interactions for HIV-1 exon 3 splicing. Additionally, it remains open whether hnRNP D could also independently repress splicing when associated with other known target sequences. To separately study the individual properties of hnRNP D and hnRNP A1 proteins to repress exon 3 inclusion, we first measured their respective precipitation efficiencies together with different but known RNA target sequences supposed to either predominantly bind hnRNP A1 (Telomere Repeat: "TR"; [33]) or hnRNP D (AU-rich element: "ARE" [34]; Fig. 5A). A splicing neutral sequence ("neutral" [35]) neither bound by hnRNP A1 nor hnRNP D as well as a second HIV-1 repressor sequence ("ESS2" [36–39]) previously described to associate with hnRNP A1 were included as controls.

As expected, hnRNP D was most efficiently precipitated with the ARE sequence, followed by both ESSV and TR. It was inefficiently precipitated by ESS2 and the splicing neutral sequence (Fig. 5B, hnRNP D). Conversely, hnRNP A1 was efficiently precipitated with the TR substrate, followed by its ESSV target but not with any of the other three substrates (Fig. 5B, hnRNP A1). Our failure to detect hnRNP A1 on the ESS2 sequence (Fig. 5B, lane 3) is in contrast to a preceding study [39], but might simply be due to critical upstream sequences for hnRNP A1 binding not present in our RNA substrates [39].

After having confirmed the different binding specificities of these sequences we next analyzed whether the ARE sequence is able to functionally replace the ESSV sequence in the context of the infectious clone pNL4-3. Following transfection we performed RT-PCR analysis to measure the levels of exon 3 inclusion into viral mRNA species (Fig. 5C). As expected, substitution of the ESSV by the splicing neutral sequence led to complete de-repression of exon 3 and consequently, an accumulation of mRNA species including exon 3 (Fig. 5C, left panel, cf. lanes 1 and 2, Tat3). Replacement of ESSV by either TR or ARE led to exclusive detection of exon 3-less viral mRNA species (Fig. 5C, left panel, cf. lanes 1 and 2 with 3 and 4, Tat3; middle and right panel, cf. lanes 1 and 2 with 3 and 4, Nef4; Env8), demonstrating that either one of both sequences could not only functionally substitute for the ESSV repressive activity but was an even more efficient repressor of exon 3 splicing. Remarkably, ESSV substitution also led to moderate activation of the normally dormant 5′ss D2b within intron 2 (Fig. 5C, left panel, cf. lanes 1 and 2 with 3 and 4, Tat2b; [40]).

In line with these findings, northern and western blot analyses revealed normal levels of unspliced RNA and Gag protein following substitution of the ESSV sequence with either the TR or the ARE sequence (Fig. 5D and E). Collectively, these results demonstrate that hnRNP D can fully repress HIV-1 exon 3 splicing depending on the RNA target sequence. Furthermore, due to the fact that neither the TR sequence bound by hnRNP A1 nor the ARE sequence bound by hnRNP D allowed for any exon 3 inclusion at all suggests that the spatial proximity of the hnRNP binding sites of ESSV causes mutual interference between hnRNP A1 and hnRNP D allowing to dynamically regulate the repressive state of HIV-1 exon 3. Such dynamic hnRNP A1/D binding at ESSV might be necessary to allow residual exon 3 recognition adjusting relative levels of spliced *vpr*- and unspliced *gag/pol*-mRNAs.

3.4. hnRNP D binding to ESSV may fine-tune exon 3 splice site selection

To interfere with hnRNP A1/D binding at ESSV we designed different sets of point mutations with the aim of selectively changing binding of either hnRNP A1 ("A down") or hnRNP D ("D up") (Fig. 6A). RAC assays were performed to analyze the hnRNP · ESSV interaction profiles (Fig. 6B). Substituting the three core AG dinucleotides with CU dinucleotides impaired precipitation efficiencies for both p37 and slightly hnRNP A1 (Fig. 6B, cf. lanes 1 and 2). RNAs containing the "A down" mutations, however, precipitated lower levels of hnRNP A1 proteins, whereas binding of hnRNP D p37 was selectively increased (Fig. 6B, cf. lane 1 and 3). By contrast, bound fractions of "D up" mutant RNA substrates showed almost unaffected levels of hnRNP A1 proteins, whereas the levels of hnRNP D p40 and p45 were increased (Fig. 6B, cf. Lames 1 and 4). Of note, from the overall protein precipitation efficiencies in this assay we inferred that the relative weak precipitation differences were probably due to partial saturated levels of hnRNPA1 and hnRNP D isoforms precipitating together with "AD dm" mutant RNAs (Fig. 6B, lane 2).

All mutations, however, were subsequently tested for their effects on exon 3 and *vpr*-mRNA splicing in the context of infectious pNL4-3. RT-PCR analyses revealed that the levels of exon 3 inclusion and *vpr*-mRNA appeared to be dependent on the relative binding efficiencies of hnRNPA1 and hnRNP D to the mutant ESSV sequences (Fig. 6C). Accordingly, the wild-type ESSV bound by almost equal amounts of both hnRNP A1 and hnRNP D efficiently repressed exon 3 splice site selection (Fig. 6C, cf. lane 1, Tat1, Nef2, Env1: exon 3 skipping). However, the overall loss of hnRNP binding to "AD dm" mutant RNAs led to almost entire inclusion of exon 3 into viral mRNA species (Fig. 6C, cf. lanes 1 and 2, Tat3, Nef4, Env8: exon 3 inclusion), which suggests failure of formation of functional repressor complexes. Repression of exon 3 skipping, however, could also not be readjusted by increased binding of hnRNP D p37 to the "A down" mutation (Fig. 6C, cf. lanes 1 and 3), indicating that additional incorporation of hnRNP D p37 into the repressor complex cannot compensate for a lack of hnRNP A1. The precipitation levels of

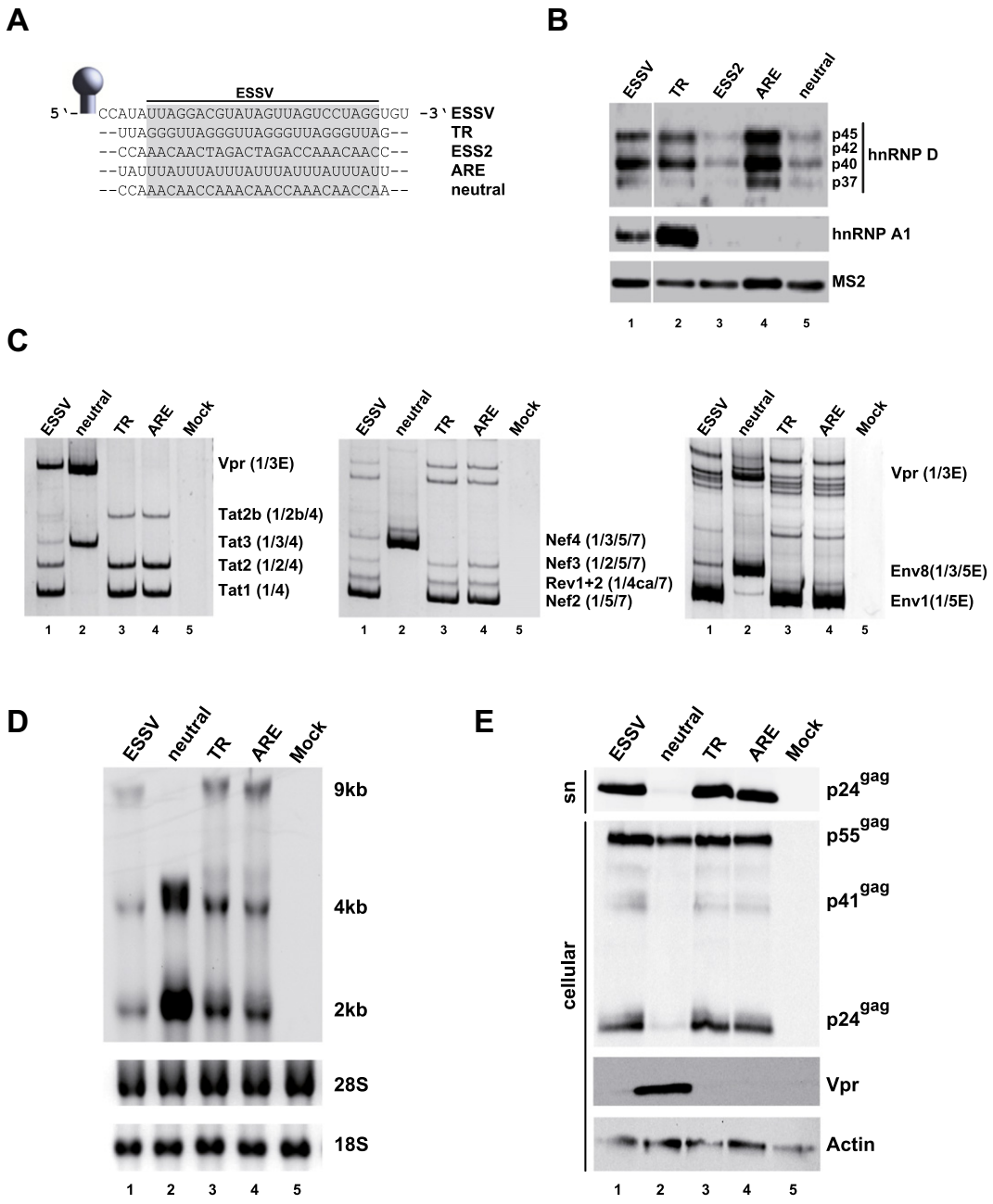


Fig. 5. A natural hnRNP D binding sequence is capable to substitute for ESSV and repress HIV-1 exon 3 inclusion. (A) Panel of *in vitro* transcribed RNA substrates used in RNA affinity chromatography (RAC) experiments. (B) RAC assays were performed with RNAs from (A) as described in Fig. 4B. Samples were probed with primary antibodies specific for hnRNP A1 and hnRNP D. Recombinant MS2 coat protein was added to HeLa cell nuclear extracts and served as a control for equal precipitation efficiencies. Cropped gel image: All samples were run on the same gel. (C) RT-PCR analysis of total RNA isolated from 2.5×10^5 HEK293T cells, which were either transfected with 1 μg of wild-type or mutant proviral pNL4-3 plasmid. The different sets of primer pairs used in RT-PCR analysis are provided in Supplementary Fig. 1. HIV-1 mRNA species are indicated to the right of the gel images according to the nomenclature published previously [3]. (D) Northern blot analysis of total RNA taken from the same RNA preparation as in (C). For detection a hybridization probe was used specifically directed against HIV-1 exon 7. **(E)** Western Blot analysis of viral Gag and Vpr expressed by wild-type or mutant proviral pNL4-3 plasmid. Supernatants (sn) and cellular lysates (cellular) were probed with a specific antibody against HIV-1 $p24^{gag}$ or HIV-1 Vpr. Equal levels of cell lysates were monitored by detection of α-actin.

hnRNP A1 seemed to be unaffected by the "D up" mutations, whereas the relative binding efficiencies of hnRNP D p40 and p45 appeared to be selectively increased (Fig. 6B, cf. lanes 1 and 4). Interestingly, this shift towards hnRNP D was only connected to a partial loss of repression of exon 3, resulting in nearly equivalent amounts of exon 3-including versus exon 3-less mRNA species and levels of *vpr*-mRNA in between those seen for the wild-type ESSV (Fig. 6C, cf. lanes 1, 3 and 4). Northern (Fig. 6D) and Western blot analyses (Fig. 6E) correlated well with the respective exon 3 splicing profiles obtained by RT-PCR. While "AD dm" and "A down" exhibited a striking defect in the expression of unspliced RNA (Fig. 6D, cf. lanes 1 to 3) and Gag protein (Fig. 6E, cf. lanes 1 to 3) as it was expected from their exon 3 oversplicing phenotypes, the increase in exon 3 splicing by "D up" appeared to be still compatible with normal levels of both (Fig. 6D, E; cf. lanes 1 and 4). Surprisingly, Vpr expression seemed to be even higher for the "D up" mutant, which might be explained by hnRNP D-mediated facilitation of *vpr*-mRNA export [41]. From these results, we concluded that increased hnRNP D binding to the ESSV alters the silencer activity in an isoform-dependent manner.

In support of this finding, we found that higher levels of hnRNP D were accompanied by an increased inclusion of exon 3 into wild-type ESSV carrying reporter mRNAs [19] (Fig. 7A). Interestingly, this

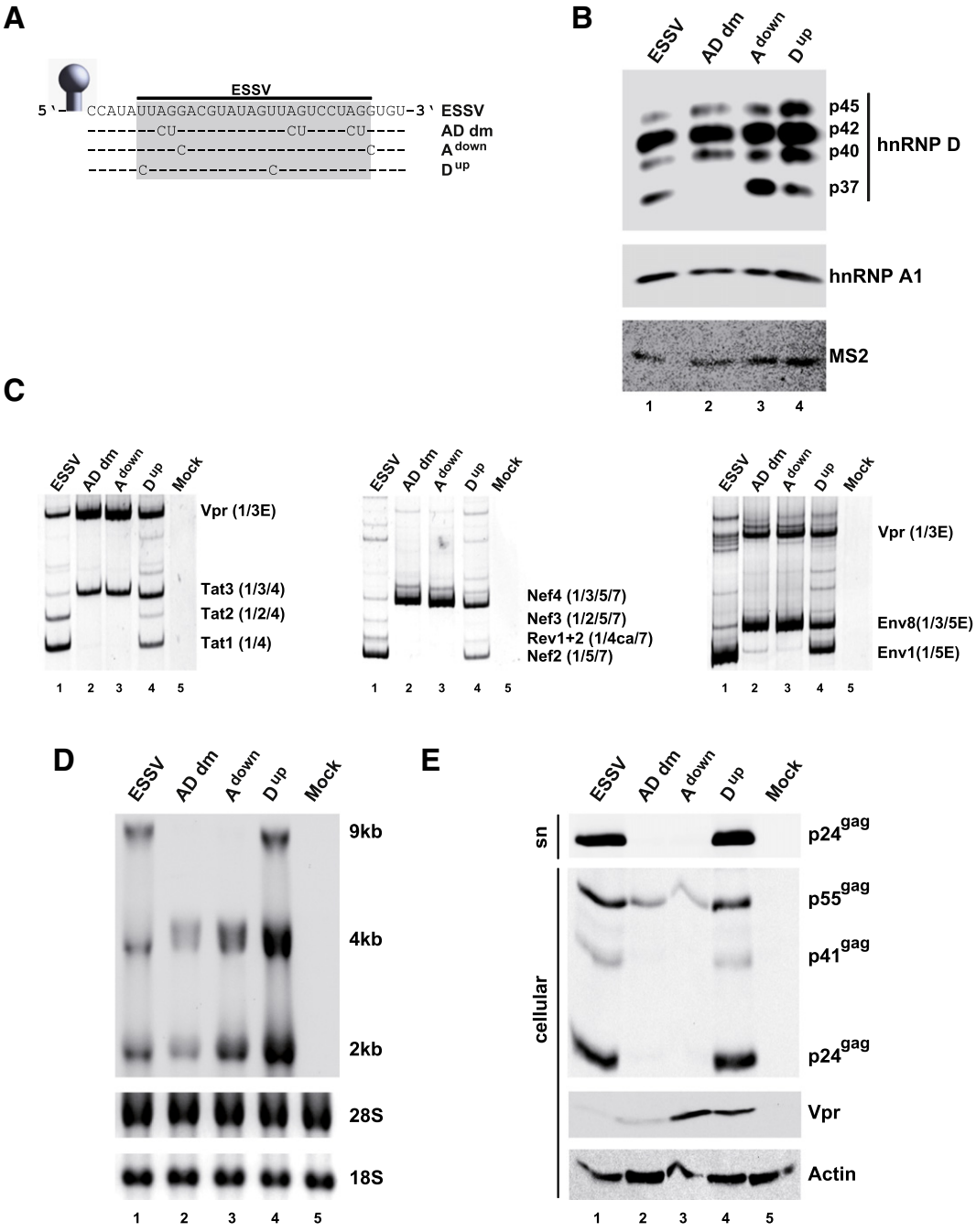


Fig. 6. Relative hnRNP A1 versus hnRNP D isoform compositions of the ESSV silencer complex determine the extent of exon 3 silencing. (A) *In vitro* transcribed RNA substrates used for the RAC assays. Mutated nucleotides are indicated below the wild-type ESSV reference sequence of pNL4-3 at corresponding positions. "-" denotes wild type nucleotide. (B) RNAs were coupled to agarose beads and analyzed for specific binding of hnRNP A1 and hnRNP D. Detection of recombinant MS2 coat protein was used as a control for equal precipitation efficiencies. (C) 2.5×10^5 HEK293T cells were transfected with pNL4-3 and mutant provirus. 48 h after transfection RNAs were analyzed by RT-PCR or (D) northern blot. (E) Supernatants (sn) and cell lysates (cellular) from transfected HEK293T cells were analyzed for viral Gag and Vpr expression as described in Fig. 5.

phenotype appeared to be more pronounced upon co-expression of the isoform p40 (Fig. 7B, cf. lanes 1–5, Tat3 and Tat4), either suggesting a higher efficiency to displace hnRNP A1 from the ESSV or alternatively, a lower ability to repress splicing when bound to exon 3. Interestingly, exon 2 splicing was also upregulated independently of the hnRNP D isoform which has been used in the co-expression experiment (Fig. 7B, cf. lanes 1–5, Tat2 and Tat4). This indicated the presence of other binding sites within the reporter RNA responding to hnRNP D overexpression or an indirect effect resulting from higher levels of hnRNP D within the cells. To briefly sum up, we propose that variations in the composition of hnRNP aggregates or mutual interference of hnRNPs might alter their splicing repressive activity. This might be a consequence of separate, non-redundant strategies of repression exploited by different hnRNP proteins.

4. Discussion

In this study, we could show that regions enriched in small non-polar amino acids (glycine- or alanine-rich domains; GRD or ARDs) derived from a representative set of prominent hnRNP proteins share a common ability to repress splicing from exonic sites, thereby supporting their long-proposed role as a common splicing repressor

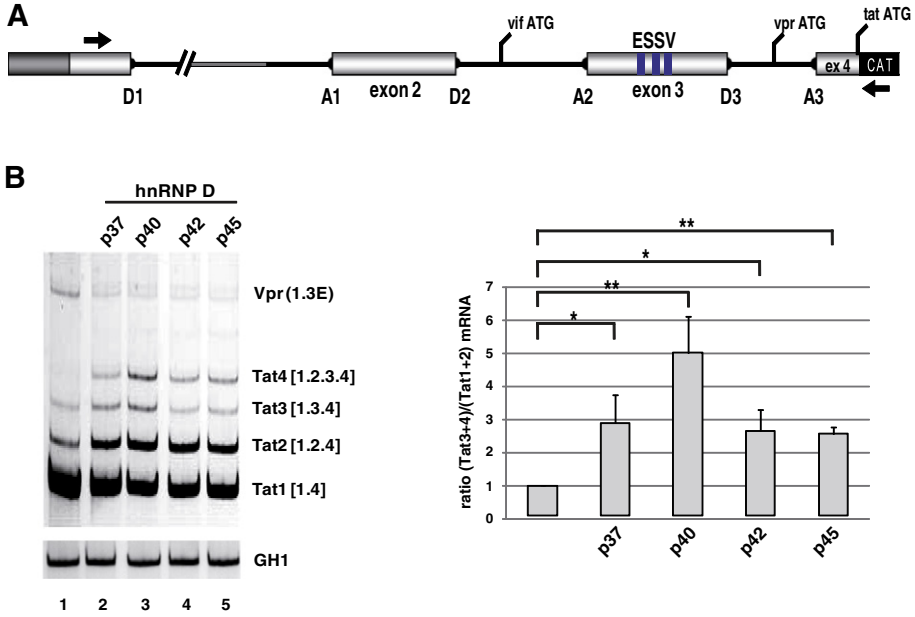


Fig. 7. Effects of hnRNP D isoform co-expression on HIV-1 exon 3 splicing. (A) Diagram of the HIV-1 based splicing reporter "LTR (intron 1) ex2 ex3" [24] containing the wild-type ESSV sequence. (B) 2.5×10^5 HEK293T cells were transiently co-transfected with 1 μg LTR (intron 1) ex2 ex3, 0.1 μg SVctat and 1 μg pcDNA3.1(+) or the respective hnRNP D isoform expressing plasmid. RNAs were analyzed by RT-PCR. Data show the mean ± SEM from three independently performed experiments (* $p < 0.05$, ** $p < 0.01$).

domain of hnRNP proteins. Activation by RS domains has previously been shown to depend on their size and the total number of arginine-serine repeats [9,10], which raised the question of whether the size of GRDs might also determine the repressive strength of hnRNP proteins. However, a general correlation between exon splicing efficiency and the size of the effector domain could not be confirmed for the GRDs tested. The sole exceptions were the different GRDs derived from the four hnRNP D isoforms. Herein, extension of the C-terminal GRD seemed to be linked to a relatively higher repression by the longer hnRNP D isoforms p42 and p45. Collectively, however, our data argue against a simplified model of size-dependent repression by GRDs. It rather suggests that the composition or the occurrence of specific motifs inside of the domains determines their relative strength, which is also supported by a recent study [15]. Noteworthy, in the MS2 tethering assay we failed to demonstrate a repressive activity for synthetic domains composed of only four consecutive recurring repeats ("GRGG", "GGYGG", "GYGG") (*data not shown*) contradicting results obtained with arginine/serine (RS) dipeptides of SR proteins [42].

Importantly, hnRNP D proteins were not only capable of repressing splicing after tethering to the exon, but also when bound to a natural high affinity RNA binding site (AU-rich element, ARE) in absence of hnRNP A1. Based on these results we propose that hnRNP D and its isoforms, aside from their known molecular functions in mRNA decay (for a recent review see [43]), are also regulators of pre-mRNA splicing. This is in line with two preceding studies showing that hnRNP D plays a role during human pappilomavirus-16 (HPV-16) splice site silencing [31] and that as a consequence of perturbations in the normal hnRNP D protein expression profile wide-scale alterations in the cellular splicing pattern could be observed [30]. On the other hand, analyzing the effects of hnRNP D depletion on HIV-1 gene expression, Lund et al. came to the conclusion that hnRNP D would mainly act on HIV-1 gene expression through facilitating the nuclear export of viral intron-containing RNAs to the cytoplasm [41]. They observed few to no effects on splice site selection after hnRNP D downregulation [41]. However, splicing profiles were not captured after overexpression of individual hnRNP D isoforms [41], favouring displacement of hnRNP A1 from the ESSV sequence. Unfortunately, our approach to generate an ESSV-based mutant that exclusively binds hnRNP A1, but lacks hnRNP D isoform binding, to screen for potential changes in exon 3 inclusion and to unequivocally pin down a role of hnRNP D for exon 3 splicing regulation failed. However, as mentioned above, we could at least show that the ARE sequence solely bound by hnRNP D proteins could effectively repress exon 3 inclusion when substituted for ESSV in the viral context. Furthermore, plasmid-driven overexpression of p45 in hnRNP A1-deficient CB3 cells was able to efficiently rescue exon 3 repression. Remarkably, overexpression of hnRNP D had the opposite effect on exon 3 splicing in presence of normal levels of hnRNP A1, leading to weak de-repression of the splice sites in an isoform-dependent manner. It might be plausible that higher concentrations of hnRNP D resulted in outcompeting hnRNP A1 at the two "*UUAG*" motifs within ESSV, thereby rearranging the silencer complex. In line with these results, we could observe that the relative hnRNPA1/D stoichiometries of the silencer complex appeared to play a role for the final splicing outcome. Sequences either predominantly interacting with hnRNP A1 (TR) or exclusively interacting with hnRNP D (ARE) turned out to be the most potent repressors of exon 3 splicing. By contrast, comparable amounts of hnRNP A1 and hnRNP D (ESSV) slightly relieved silencing and allowed residual exon 3 recognition. Finally, further de-repression could be achieved by an additional shift of mixed A1/D complexes towards hnRNP D isoforms ("D up"). Despite the fact that hnRNP D themselves can act as potent repressors of splicing, all or just individual isoforms, p37 or p40 might be less suitable cooperation partners for hnRNP A1 upon splice site repression. As a result, variations in the relative levels of hnRNP A1/D bound to the ESSV could modify the GRD-dependent silencer activity, thereby finely adjusting the relative levels of spliced *vpr* and unspliced *gag/pol*-mRNAs. Therefore, it is tempting to speculate that hnRNP proteins may exploit alternative, maybe mutually exclusive, silencing pathways to interfere with splice site recognition, although they all can act as splicing repressors. A variety of mechanisms had been described how hnRNP protein · RNA interactions can lead to inhibited splice site recognition, including (i) sterical hindrance that impedes binding of SR proteins or general splicing components to the RNA (e.g. [17,36,44]), also as a consequence of hnRNP oligomerisation (referred to as "zone of silencing") (e.g. [13]), (ii) self-interaction of hnRNP proteins engaging two distal binding sites to loop out and bury internal splice sites (e.g. [26,45,46]) and (iii) formation of dead-end splicing complexes, which are stalled

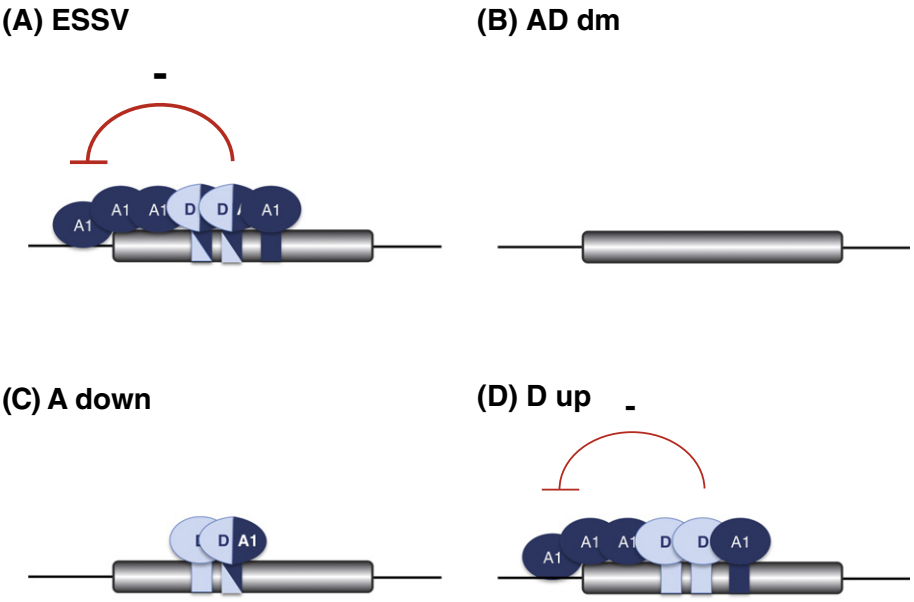


Fig. 8. Model of HIV-1 exon 3 splicing repression by compositionally different hnRNP A1/D assemblies. (A) hnRNP A1 binds to the wild-type ESSV sequence and efficiently spreads along the upstream 3′-sequence via self-interactions between hnRNP A1 proteins. However, hnRNP D binding to the first "*UUAG*" motifs might interfere with "*zone of silencing*" establishment, thereby occasionally de-repressing 3′ss A2 for *vpr*-mRNA formation. (B–C) Absence of sufficient hnRNP A1 binding to exon 3 entirely relieves repression from exon 3. (D) Increased hnRNP D binding to the ESSV further decreases spreading efficiency and concomitantly, exon 3 splice site repression.

for progression into a functional, catalytic active spliceosome [9,47–49]. Although the GRDs of hnRNP A1 and hnRNP H were documented to be functionally interchangeable for "*looping out*" of internal RNA sequences throughout splicing regulation [26] and hnRNP proteins have been documented to cooperate during splice site repression [47,50,51], certain combinations of hnRNP proteins may also get in each other's way [52], as it is speculated here for hnRNP A1 and D binding to the ESSV sequence. From this, it still remains unclear whether individual hnRNP proteins are capable to carry out all or just some of the modes of repression and to which extent different hnRNPs can cooperate during splice site repression. In one possible scenario, replacement of hnRNP A1 by p40 and p45 isoforms at the upstream "*UUAG*" motif (as seen for the "D up" mutant) might decrease hnRNP oligomerisation towards the 3′ ss A2 [13] (Fig. 8). Herein, particularly p40 (due to its shorter GRD) could be a less active collaborator for interaction with hnRNP A1 or p45 throughout formation of a "*zone of silencing*", thereby partially relieving exon 3 silencing (Fig. 7). Alternatively, replacement of hnRNP A1 by hnRNP D proteins could lead to a switch in the repression pathways, for example from a "*zone of silencing*" towards a "*looping out*" mode (or vice versa). The idea that some hnRNP proteins cooperate throughout splicing regulation, while others counteract each other was recently confirmed using a genome-wide approach by Huelga et *al.* [19]. In this study, alterations in splicing after depletion of six hnRNP proteins, A1, A2/B1, H1, F, M and U were complemented by global maps of RNA binding sites to identify cooperative as well as antagonistic actions of hnRNP proteins on exon recognition. For instance, it was found that hnRNP A1 showed a general tendency to oppose the effects of other hnRNP proteins on co-regulated exons, supporting the hypothesis that dynamic hnRNP A1/D arrangements at the ESSV may fine-tune the relative HIV-1 exon 3 inclusion levels. For upcoming studies, it will be of interest to expand mutational analyses and individually inactivate binding of hnRNP A1, hnRNP D and both at each of the three "*UAG*" motifs to determine their relative contribution to exon 3 splice site repression. This may reveal a functional hierarchy between the binding motifs in the control of exon 3 inclusion as it was found for other hnRNPA1-dependent cassette exons [20].

Revisiting the study of Lund et al. [41], the relative hnRNP A1/D isoform compositions might not only determine the silencing but also the viral mRNA export competence of the ESSV-bound complexes. It was found that unspliced viral RNA and Gag protein expression was increased in the presence of higher levels of p42 and p45, whereas overexpression of p37 and p40 showed an opposite effect [41]. Unfortunately, effects on viral RNA splice site selection were not controlled upon co-expression experiments. The longer hnRNP D isoforms at the ESSV might couple partial maintenance of splice site silencing to a stimulated export of the resultant intron-containing RNAs which would also explain higher Gag and Vpr protein levels for "D up" than expected from their intermediate exon 3 splicing pattern. By contrast, the shorter hnRNP D isoforms might further enhance exon 3 inclusion, leading to an additional decrease in the accumulation of *vpr*-encoding and unspliced viral RNAs. The loss of Gag expression could then be the result of a stronger impairment of the ESSV activity rather than to be due to a defect in viral RNA export stimulation. This is supported by the detection of increased amounts of Vpr protein for the "A down" mutant when compared to the "AD dm" variant. Both mutants showed depleted levels of hnRNP A1, but "A down" was selectively increased in p37 binding, indicating that even p37 would be capable to stimulate intron-containing RNA export and thus, Vpr protein expression. Accordingly, it might also be plausible that simultaneous binding of hnRNP A1 and D proteins to the ESSV functionally couples splice site regulation to stimulation of viral mRNA export into the cytoplasm. However, whether hnRNP D really interconnects splicing of intron-containing viral RNAs and their subsequent export into the cytoplasm, awaits further studies.

Authors' contributions

FH, JP and SE conceived and designed MS2 tethering experiments, HIV-related transfection and readout experiments, designed and performed expression and RNA-pull-down analysis. FH conceived and designed HIV-related transfection and readout experiments, designed and performed expression analysis. AB designed and performed cloning of the MS2/SR fusion protein expressing plasmids and related MS2 tethering experiments. HS and SE conceived the study, supervised its design and its coordination, and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgements

We are grateful to Björn Wefers for excellent technical assistance. These studies were funded by the DFG (SCHA 909/3-1), the Stiftung für AIDS-Forschung, Düsseldorf (H.S.) and Jürgen-Manchot-Stiftung (J.P, H.S.). We thank John Conboy, Massimo Caputi, Chris Smith, Alain Cochrane and David Peabody for providing plasmids. We thank Colin Donohoe for critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagrm.2016.12.001.

References

[1] M.E. Klotman, et al., Kinetics of expression of multiply spliced RNA in early human immunodeficiency virus type 1 infection of lymphocytes and monocytes, Proc. Natl. Acad. Sci. U. S. A. 88 (11) (1991) 5011–5015.
[2] S.Y. Kim, et al., Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression, J. Virol. 63 (9) (1989) 3708–3713.
[3] D.F. Purcell, M.A. Martin, Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity, J. Virol. 67 (11) (1993) 6365–6378.
[4] B.R. Cullen, Regulation of human immunodeficiency virus replication, Annu. Rev. Microbiol. 45 (1991) 219–250.
[5] J. Karn, C.M. Stoltzfus, Transcriptional and posttranscriptional regulation of HIV-1 gene expression, Cold Spring Harb. Perspect. Med. 2 (2) (2012) a006916.
[6] C.M. Stoltzfus, Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication, Adv. Virus Res. 74 (2009) 1–40.
[7] M. Chen, J.L. Manley, Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches, Nat. Rev. Mol. Cell Biol. 10 (11) (2009) 741–754.
[8] X.D. Fu, M. Ares Jr., Context-dependent control of alternative splicing by RNA-binding proteins, Nat. Rev. Genet. 15 (10) (2014) 689–701.
[9] S. Erkelenz, et al., Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms, RNA 19 (1) (2013) 96–102.
[10] B.R. Graveley, T. Maniatis, Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing, Mol. Cell 1 (5) (1998) 765–771.
[11] H. Shen, M.R. Green, RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans, Genes Dev. 20 (13) (2006) 1755–1765.
[12] H. Shen, M.R. Green, A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly, Mol. Cell 16 (3) (2004) 363–373.
[13] H.L. Okunola, A.R. Krainer, Cooperative-binding and splicing-repressive properties of hnRNP A1, Mol. Cell. Biol. 29 (20) (2009) 5620–5631.
[14] G. Shankarling, K.W. Lynch, Minimal functional domains of paralogues hnRNP L and hnRNP LL exhibit mechanistic differences in exonic splicing repression, Biochem. J. 453 (2) (2013) 271–279.
[15] N.M. Keppetipola, et al., Multiple determinants of splicing repression activity in the polypyrimidine tract binding proteins, PTBP1 and PTBP2, RNA (2016).
[16] P.S. Bilodeau, et al., RNA splicing at human immunodeficiency virus type 1 3′ splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element, J. Virol. 75 (18) (2001) 8487–8497.
[17] J.K. Domsic, et al., Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts, Mol. Cell. Biol. 23 (23) (2003) 8762–8772.
[18] J.M. Madsen, C.M. Stoltzfus, An exonic splicing silencer downstream of the 3′ splice site A2 is required for efficient human immunodeficiency virus type 1 replication, J. Virol. 79 (16) (2005) 10478–10486.
[19] S.C. Huelga, et al., Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins, Cell Rep. 1 (2) (2012) 167–178.
[20] G.H. Bruun, et al., Global identification of hnRNP A1 binding sites for SSO-based splicing modulation, BMC Biol. 14 (2016) 54.
[21] S. Erkelenz, et al., Tra2-mediated recognition of HIV-1 5′ splice site D3 as a key factor in the processing of vpr mRNA, J. Virol. 87 (5) (2013) 2721–2734.
[22] K.K. Singh, et al., Human SAP18 mediates assembly of a splicing regulatory multiprotein complex via its ubiquitin-like fold, RNA 16 (12) (2010) 2442–2454.
[23] S. Erkelenz, et al., Balanced splicing at the Tat-specific HIV-1 3′ss A3 is critical for HIV-1 replication, Retrovirology 12 (2015) 29.
[24] M. Widera, et al., A functional conserved intronic G run in HIV-1 intron 3 is critical to counteract APOBEC3G-mediated host restriction, Retrovirology 11 (2014) 72.
[25] F. Del Gatto-Konczak, et al., hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer, Mol. Cell. Biol. 19 (1) (1999) 251–260.
[26] J.F. Fisette, et al., hnRNP A1 and hnRNP H can collaborate to modulate 5′ splice site selection, RNA 16 (1) (2010) 228–238.
[27] M. Freund, et al., A novel approach to describe a U1 snRNA binding site, Nucleic Acids Res. 31 (23) (2003) 6963–6975.
[28] S. Cho, et al., Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70 K snRNP protein determines early spliceosome assembly, Proc. Natl. Acad. Sci. U. S. A. 108 (20) (2011) 8233–8238.
[29] J. Jean-Philippe, et al., A truncated hnRNP A1 isoform, lacking the RGG-box RNA binding domain, can efficiently regulate HIV-1 splicing and replication, Biochim. Biophys. Acta 1839 (4) (2014) 251–258.
[30] J.H. Yoon, et al., PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity, Nat. Commun. 5 (2014) 5248.
[31] X. Li, et al., Suppression of HPV-16 late L1 5′-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs, Nucleic Acids Res. 41 (22) (2013) 10488–10508.
[32] B.J. Wagner, et al., Structure and genomic organization of the human AUF1 gene: alternative pre-mRNA splicing generates four protein isoforms, Genomics 48 (2) (1998) 195–202.
[33] F. Ishikawa, et al., Nuclear proteins that bind the pre-mRNA 3′ splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n, Mol. Cell. Biol. 13 (7) (1993) 4301–4310.
[34] W. Zhang, et al., Purification, characterization, and cDNA cloning of an AU-rich element RNA-binding protein, AUF1, Mol. Cell. Biol. 13 (12) (1993) 7652–7665.
[35] X.H. Zhang, et al., Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing, RNA 15 (3) (2009) 367–376.
[36] H. Hallay, et al., Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site, J. Biol. Chem. 281 (48) (2006) 37159–37174.
[37] B.A. Amendt, et al., Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1, Mol. Cell. Biol. 14 (6) (1994) 3960–3970.
[38] B.A. Amendt, Z.H. Si, C.M. Stoltzfus, Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors, Mol. Cell. Biol. 15 (8) (1995) 4606–4615.
[39] A.M. Zahler, et al., SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing, J. Biol. Chem. 279 (11) (2004) 10077–10084.
[40] M. Widera, et al., An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA, J. Virol. 87 (5) (2013) 2707–2720.
[41] N. Lund, et al., Differential effects of hnRNP D/AUF1 isoforms on HIV-1 gene expression, Nucleic Acids Res. 40 (8) (2012) 3663–3675.
[42] D. Philipps, et al., Arginine/serine repeats are sufficient to constitute a splicing activation domain, Nucleic Acids Res. 31 (22) (2003) 6502–6508.
[43] F.M. Gratacos, G. Brewer, The role of AUF1 in regulated mRNA decay, Wiley Interdiscip. Rev. RNA 1 (3) (2010) 457–473.
[44] T.K. Doktor, et al., SMN2 exon 7 splicing is inhibited by binding of hnRNP A1 to a common ESS motif that spans the 3′ splice site, Hum. Mutat. 32 (2) (2011) 220–230.
[45] F.U. Nasim, et al., High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5′ splice site selection in support of a common looping out and repression mechanism, RNA 8 (8) (2002) 1078–1089.
[46] J. Konig, et al., iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution, Nat. Struct. Mol. Biol. 17 (7) (2010) 909–915.
[47] N.T. Chiou, G. Shankarling, K.W. Lynch, hnRNP L and hnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly, Mol. Cell 49 (5) (2013) 972–982.
[48] S. Sharma, et al., Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome, Nat. Struct. Mol. Biol. 15 (2) (2008) 183–191.
[49] S. Sharma, et al., U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression, Mol. Cell 41 (5) (2011) 579–588.
[50] F. Nasrin, et al., HnRNP C, YB-1 and hnRNP L coordinately enhance skipping of human MUSK exon 10 to generate a Wnt-insensitive MuSK isoform, Sci. Rep. 4 (2014) 6841.
[51] A. Damianov, et al., Rbfox proteins regulate splicing as part of a Large Multiprotein Complex LASR, Cell 165 (3) (2016) 606–619.
[52] M.A. Rahman, et al., HnRNP L and hnRNP LL antagonistically modulate PTB-mediated splicing suppression of CHRNA1 pre-mRNA, Sci. Rep. 3 (2013) 2931.

6. Summary and conclusion

Accurate precursor mRNA splicing is regulated by the combinatorial control of multiple splicing regulatory elements either promoting or repressing splice site usage. In this thesis, the interplay of competing or cooperating SREs near splice sites have been analyzed. For this, biochemical, virological experimental methods as well as bioinformatics tools have been used to dissect the role of SR and hnRNP protein networks that control splice site use within human and viral genes. In **chapter 1**, a general overview of the mechanism of pre-mRNA splicing and its role in disease and HIV-1 replication was presented. Splice site choice is based on the intrinsic strengths of 5'ss and 3'ss to interact with components of the splicing machinery as well as on the sequence neighborhood. Here, SR and hnRNP proteins binding to the RNA guide splicing into both directions in a position-dependent manner. However, predicting the location of a putative SRE to identify distinct splicing outcomes still remains a challenging task. An easy identification of aberrant splicing events, however, is essential for diagnostics and treatment of patients suffering from many genetic disorders. As an example, changes in splicing patterns of factors mediating blood vessel regulation can contribute to changes in endothelial function. In **chapter 2**, alternative splicing as a potential cause for cardiovascular disease was reviewed. The cardiovascular system is responsible for transporting blood through the organism, thereby delivering oxygen, hormones and essential nutrients to cells and organs. The innermost layer of each blood vessel is composed of endothelial cells that are responsible for dilation and constriction of blood vessels, as well as for regulating hemostasis and angiogenesis. One key player for vessel relaxation is the endothelial nitric oxide synthase (eNOS) which catalyzes the reaction from L-arginine to L-citrulline and NO. Alternative recognition of splice sites of eNOS exons 13 and 14 is controlled by hnRNP L and regulates vascular functions by promoting the formation of negatively

acting heterodimers. Furthermore, changes in splicing pattern of Vascular Endothelial Growth Factor A (VEGFA), Endoglin and Grainyhead-Like 3 (GRHL3) result in serious consequences for vascular physiology. Consequently, identifying all existing splice variants in genes mediating endothelial cell function and resolving the mechanistic principles behind it might lead to ideas about new therapeutic approaches. In **chapter 3** we focused on 5'ss recognition within the human fibrinogen Bβ-chain gene (FGB) exon 7. Here, a mutation within its authentic splice donor sequence (c.1244+1G>T) results in the activation of four cryptic splice sites (c1-c3, p1). So far, only a single binding site for SRSF1 had been described, activating c1. Extending these findings, we identified a whole cluster of alternating multiple SREs and U1 snRNA binding sites that controls cryptic splice donor usage. In minigene analysis, additional elements (B (= published SRSF1 binding site), C, D) were found as key players in regulating FGB splicing and deeply analyzed for their influence in the splicing outcome. With aid of the HEXplorer algorithm, mutational analysis confirmed their positive effect in activating downstream and simultaneously inhibiting upstream located splice donor sites. Additionally, pull-down analysis showed that multiple SR proteins are able to bind to these sequences, predominantly SRSF1 and Tra2β. We furthermore could expand this concept of splice site regulation to other genes and validated by statistical analyses, that for competing 5'ss, highly used 5'ss are significantly more supported by SREs than silent GT-sites. Hence, we proposed that this mode of splicing regulation seems not to be restricted to FGB splice donor selection but rather to be a general concept to define exon length. In **chapter 4** we further analyzed complex splicing regulatory networks. HIV-1 uses the human splicing machinery to generate more than 50 mRNA isoforms that encode for fifteen viral proteins. Especially within exon 2/2b, several SREs have been found to regulate proper *vif* mRNA production which is crucial to counteract the host restriction factor A3G. Studies using a minigene showed that besides known

SREs, an additional silencer element (ESS2b) is needed for viral splicing. We performed MS analysis after RNA affinity purification with WT and mutant sequences and showed that ESS2b is bound by hnRNP proteins of the A/B family. Furthermore, we revealed the functional importance of ESS2b and ESE2b within the proviral clone pNL4-3 and demonstrate that this cluster of splicing regulatory elements tightly regulates exon 2/2b inclusion and D2/D2b usage. Mutating either ESS2b or ESE2b led to aberrant splicing, a shift in the distribution of viral mRNA classes, a defect in Vif protein expression and, finally, to a loss of infectiousness. Furthermore, masking ESS2b and ESE2b with locked nucleic acids (LNAs) resulted in restricted viral particle production. By aligning HIV-1 subtype consensus sequences we found that sequence variations occurred far more often within regions containing ESS2b and ESE2b than outside and speculated that this cluster of SREs guarantees proper viral replication in cells with different A3G levels or splicing regulatory protein concentrations. Thus, in this chapter we propose targeting those SREs may lead to the development of novel effective therapies against HIV-1. In **chapter 5** the G-rich domains (GRDs) of hnRNP proteins was approved as general splicing repressors. Four different hnRNP D isoforms exist that all contain a C-terminal GRD of varying length. All can act as repressor of exon inclusion, however, we showed that the repressive state correlates with the presence of an extended GRD. Moreover, we revealed that hnRNP A1 and hnRNP D both bind to the HIV-1 silencer element within exon 3, ESSV. However, due to their overlapping binding motifs within ESSV, hnRNP A1 and hnRNP D interfere with each other. Thus, we proposed that hnRNP A1 and hnRNP D isoforms facilitate a dynamic regulation of spliced *vpr*- and unspliced *gag*/*pol*-mRNAs.

7. Curriculum Vitae

Personal Information

Anna-Lena Brillen, M.Sc.

Education

Since 04/2013	PhD Thesis, Heinrich-Heine-Universität Düsseldorf “Recognition and masking of putative U1 snRNA binding sites“
10/2010 - 12/2012	Master Biology, Heinrich-Heine-Universität Düsseldorf “Funktionelle Kartierung von Effektordomänen spleißregulatorischer Proteine“, Grade: 1,0
10/2007 - 10/2010	Bachelor Biology, Heinrich-Heine-Universität Düsseldorf „Stammzell-modulierte Neuroregeneration: Mechanismen und neuroethische Komplikationen“, Grade: 1,7

Publications

1. S Farrokh*, **AL Brillen***, J Haendeler, J Altschmied, H Schaal. Critical regulators of endothelial cell functions: for a change being alternative. Antioxid Redox Signal. 2015 May 10;22(14):1212-29. *shared first author
2. **AL Brillen**, K Schöneweis, L Walotka, L Hartmann, L Müller, J Ptok, W Kaisers, G Poschmann, K Stühler, E Buratti, S Theiss, H Schaal. Succession of splicing regulatory elements determines cryptic 5'ss functionality. Nucleic Acids Res. 2016 Dec 29. pii: gkw1317.
3. **AL Brillen**, L Walotka, F Hillebrand, M Widera, S Theiss, H Schaal. Analysis of competing HIV-1 splice donor sites uncovers a tight cluster of splicing regulatory elements within exon 2/2b. Journal of Virology (*Submitted - JVI00389-17*).
4. F Hillebrand*, JO Peter*, **AL Brillen**, M Otte, H Schaal, S Erkelenz. Differential hnRNP D isoform incorporation may confer plasticity to the ESSV-mediated repressive state across HIV-1 exon 3. Biochim Biophys Acta. 2017 Feb;1860(2):205-217.

Talks

AL Brillen, L Walotka, F Hillebrand, M Widera, S Theiss, H Schaal. Analysis of competing HIV-1 splice donor sites uncovers a tight cluster of splicing regulatory elements within exon 2/2b. 26th Annual Meeting of the Society for Virology, 06.04.-09.03.2016, Münster

Poster

AL Brillen, H Schaal. Processing of HIV-1 intron-containing mRNAs. 25th Annual Meeting of the Society for Virology, 18.3-21.03.2015, Bochum

AL Brillen, K Schöneweis, S Theis, H Schaal. Position-dependent effects of multiple splicing regulatory elements determine exon length and are critical for cryptic splice site activation. Meeting on Eukaryotic mRNA processing, August 18 - 22, 2015, Cold Spring Harbor, New York

8. Acknowledgements

Danken möchte ich an dieser Stelle:

Professor Dr. Heiner Schaal für die steten Hilfestellungen und Geduld, die wissenschaftlichen Freiräume, die Motivation und die vielen langen Sitzungen gemeinsam mit Stephan damit ich meine Arbeit kumulativ beenden kann.

Prof. Dr. Michael Feldbrügge für die bereitwillige Übernahme des Co-Referats.

Dr. Stephan Theiss für konstruktive Diskussionen und Anregungen in den mittlerweile unzähligen Sitzungen bei Heiner im Büro.

Der Jürgen Manchot Stiftung und der Graduiertenschule „Moleküle der Infektion II (MOI II)“.

Des Weiteren möchte ich mich bei Prof. Dr. Jörg Timm und Prof Dr. Hartmut Hengel dafür bedanken, dass ich diese Arbeit am Institut für Virologie fertigstellen konnte.

Ganz besonders danken möchte ich natürlich meinen Kollegen Nora Diehl, Steffen Erkelenz, Isabel Heide, Hanna Heinrichs, Frank Hillebrand, Claus Lenski, Lisa Müller, Jan Otto Peter, Aljoscha Tersteegen, Johannes Ptok, Lara Walotka, Björn Wefers und Marek Widera für die einfach großartige Zeit im Labor!

Mein unermesslicher Dank gilt Alex und meiner Familie.

9. Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Anna-Lena Brillen

Düsseldorf, den

10. References

1. Proudfoot, N.J., A. Furger, and M.J. Dye, *Integrating mRNA processing with transcription.* Cell, 2002. **108**(4): p. 501-12.
2. Kammler, S., et al., *The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA.* RNA, 2001. **7**(3): p. 421-34.
3. Kaida, D., et al., *U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation.* Nature, 2010. **468**(7324): p. 664-8.
4. Burset, M., I.A. Seledtsov, and V.V. Solovyev, *Analysis of canonical and non-canonical splice sites in mammalian genomes.* Nucleic Acids Res, 2000. **28**(21): p. 4364-75.
5. Cartegni, L., S.L. Chew, and A.R. Krainer, *Listening to silence and understanding nonsense: exonic mutations that affect splicing.* Nat Rev Genet, 2002. **3**(4): p. 285-98.
6. Sheth, N., et al., *Comprehensive splice-site analysis using comparative genomics.* Nucleic Acids Res, 2006. **34**(14): p. 3955-67.
7. Matera, A.G. and Z. Wang, *A day in the life of the spliceosome.* Nat Rev Mol Cell Biol, 2014. **15**(2): p. 108-21.
8. Will, C.L. and R. Luhrmann, *Spliceosome structure and function.* Cold Spring Harb Perspect Biol, 2011. **3**(7).
9. Rymond, B., *Targeting the spliceosome.* Nat Chem Biol, 2007. **3**(9): p. 533-5.
10. Moore, M.J. and P.A. Sharp, *Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing.* Nature, 1993. **365**(6444): p. 364-8.
11. Wahl, M.C. and R. Luhrmann, *SnapShot: Spliceosome Dynamics I.* Cell, 2015. **161**(6): p. 1474-e1.

12. Wahl, M.C., C.L. Will, and R. Luhrmann, *The spliceosome: design principles of a dynamic RNP machine.* Cell, 2009. **136**(4): p. 701-18.

13. Levine, A. and R. Durbin, *A computational scan for U12-dependent introns in the human genome sequence.* Nucleic Acids Res, 2001. **29**(19): p. 4006-13.

14. Will, C.L. and R. Luhrmann, *Splicing of a rare class of introns by the U12-dependent spliceosome.* Biol Chem, 2005. **386**(8): p. 713-24.

15. Huranova, M., et al., *The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells.* J Cell Biol, 2010. **191**(1): p. 75-86.

16. Michaud, S. and R. Reed, *An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway.* Genes Dev, 1991. **5**(12B): p. 2534-46.

17. Query, C.C., M.J. Moore, and P.A. Sharp, *Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model.* Genes Dev, 1994. **8**(5): p. 587-97.

18. Chanarat, S. and K. Strasser, *Splicing and beyond: the many faces of the Prp19 complex.* Biochim Biophys Acta, 2013. **1833**(10): p. 2126-34.

19. Raghunathan, P.L. and C. Guthrie, *RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2.* Curr Biol, 1998. **8**(15): p. 847-55.

20. Sun, J.S. and J.L. Manley, *A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing.* Genes Dev, 1995. **9**(7): p. 843-54.

21. Konarska, M.M., J. Vilardell, and C.C. Query, *Repositioning of the reaction intermediate within the catalytic center of the spliceosome.* Mol Cell, 2006. **21**(4): p. 543-53.

22. Smith, D.J., C.C. Query, and M.M. Konarska, *"Nought may endure but mutability": spliceosome dynamics and the regulation of splicing.* Mol Cell, 2008. **30**(6): p. 657-66.

23. Le Hir, H. and G.R. Andersen, *Structural insights into the exon junction complex.* Curr Opin Struct Biol, 2008. **18**(1): p. 112-9.

24. Scherer, S., *A short guide to the human genome*. 2008, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. xiv, 173 p.

25. De Conti, L., M. Baralle, and E. Buratti, *Exon and intron definition in pre-mRNA splicing.* Wiley Interdiscip Rev RNA, 2013. **4**(1): p. 49-60.

26. Robberson, B.L., G.J. Cote, and S.M. Berget, *Exon definition may facilitate splice site selection in RNAs with multiple exons.* Mol Cell Biol, 1990. **10**(1): p. 84-94.

27. Berget, S.M., *Exon recognition in vertebrate splicing.* J Biol Chem, 1995. **270**(6): p. 2411-4.

28. Hoffman, B.E. and P.J. Grabowski, *U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon.* Genes Dev, 1992. **6**(12B): p. 2554-68.

29. Talerico, M. and S.M. Berget, *Effect of 5' splice site mutations on splicing of the preceding intron.* Mol Cell Biol, 1990. **10**(12): p. 6299-305.

30. Grabowski, P.J., et al., *Combinatorial splicing of exon pairs by two-site binding of U1 small nuclear ribonucleoprotein particle.* Mol Cell Biol, 1991. **11**(12): p. 5919-28.

31. Kuo, H.C., F.H. Nasim, and P.J. Grabowski, *Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle.* Science, 1991. **251**(4997): p. 1045-50.

32. Fox-Walsh, K.L., et al., *The architecture of pre-mRNAs affects mechanisms of splice-site pairing.* Proc Natl Acad Sci U S A, 2005. **102**(45): p. 16176-81.

33. Sharma, S., et al., *Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome.* Nat Struct Mol Biol, 2008. **15**(2): p. 183-91.

34. Ram, O. and G. Ast, *SR proteins: a foot on the exon before the transition from intron to exon definition.* Trends Genet, 2007. **23**(1): p. 5-7.

35. Schneider, M., et al., *Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes.* Mol Cell, 2010. **38**(2): p. 223-35.

36. Izaurralde, E., et al., *A nuclear cap binding protein complex involved in pre-mRNA splicing.* Cell, 1994. **78**(4): p. 657-68.

37. Vagner, S., C. Vagner, and I.W. Mattaj, *The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing.* Genes Dev, 2000. **14**(4): p. 403-13.

38. Millevoi, S., et al., *An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries.* EMBO J, 2006. **25**(20): p. 4854-64.

39. Kyburz, A., et al., *Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing.* Mol Cell, 2006. **23**(2): p. 195-205.

40. Bentley, D.L., *Coupling mRNA processing with transcription in time and space.* Nat Rev Genet, 2014. **15**(3): p. 163-75.

41. Tilgner, H., et al., *Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs.* Genome Res, 2012. **22**(9): p. 1616-25.

42. Windhager, L., et al., *Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution.* Genome Res, 2012. **22**(10): p. 2031-42.

43. Singh, J. and R.A. Padgett, *Rates of in situ transcription and splicing in large human genes.* Nat Struct Mol Biol, 2009. **16**(11): p. 1128-33.

44. Danko, C.G., et al., *Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells.* Mol Cell, 2013. **50**(2): p. 212-22.

45. Darzacq, X., et al., *In vivo dynamics of RNA polymerase II transcription.* Nat Struct Mol Biol, 2007. **14**(9): p. 796-806.

46. Boireau, S., et al., *The transcriptional cycle of HIV-1 in real-time and live cells.* J Cell Biol, 2007. **179**(2): p. 291-304.

47. de la Mata, M. and A.R. Kornblihtt, *RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20.* Nat Struct Mol Biol, 2006. **13**(11): p. 973-80.

48. Ji, X., et al., *SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase.* Cell, 2013. **153**(4): p. 855-68.

49. Lin, S., et al., *The splicing factor SC35 has an active role in transcriptional elongation.* Nat Struct Mol Biol, 2008. **15**(8): p. 819-26.

50. Hao, S. and D. Baltimore, *RNA splicing regulates the temporal order of TNF-induced gene expression.* Proc Natl Acad Sci U S A, 2013. **110**(29): p. 11934-9.

51. Takashima, Y., et al., *Intronic delay is essential for oscillatory expression in the segmentation clock.* Proc Natl Acad Sci U S A, 2011. **108**(8): p. 3300-5.

52. Gesteland, R.F., T. Cech, and J.F. Atkins, *The RNA world : the nature of modern RNA suggests a prebiotic RNA*. 2nd ed. Cold Spring Harbor monograph series,.

1999, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. xxv, 709 p.

53. Fox-Walsh, K.L. and K.J. Hertel, *Splice-site pairing is an intrinsically high fidelity process.* Proc Natl Acad Sci U S A, 2009. **106**(6): p. 1766-71.

54. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nat Genet, 2008. **40**(12): p. 1413-5.

55. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-6.

56. Freund, M., et al., *A novel approach to describe a U1 snRNA binding site.* Nucleic Acids Res, 2003. **31**(23): p. 6963-75.

57. Yeo, G. and C.B. Burge, *Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals.* J Comput Biol, 2004. **11**(2-3): p. 377-94.

58. Coolidge, C.J., R.J. Seely, and J.G. Patton, *Functional analysis of the polypyrimidine tract in pre-mRNA splicing.* Nucleic Acids Res, 1997. **25**(4): p. 888-96.

59. Luukkonen, B.G. and B. Seraphin, *The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in Saccharomyces cerevisiae.* EMBO J, 1997. **16**(4): p. 779-92.

60. Kammler, S., et al., *The strength of the HIV-1 3' splice sites affects Rev function.* Retrovirology, 2006. **3**: p. 89.

61. Mullen, M.P., et al., *Alpha-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice.* Genes Dev, 1991. **5**(4): p. 642-55.

62. Singh, R., J. Valcarcel, and M.R. Green, *Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins.* Science, 1995. **268**(5214): p. 1173-6.

63. Collins, C.A. and C. Guthrie, *Genetic interactions between the 5' and 3' splice site consensus sequences and U6 snRNA during the second catalytic step of pre-mRNA splicing.* RNA, 2001. **7**(12): p. 1845-54.

64. Smith, C.W., T.T. Chu, and B. Nadal-Ginard, *Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns.* Mol Cell Biol, 1993. **13**(8): p. 4939-52.

65. Fu, Y., et al., *AG-dependent 3'-splice sites are predisposed to aberrant splicing due to a mutation at the first nucleotide of an exon.* Nucleic Acids Res, 2011. **39**(10): p. 4396-404.

66. Graveley, B.R., *Sorting out the complexity of SR protein functions.* RNA, 2000. **6**(9): p. 1197-211.

67. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing.* Annu Rev Biochem, 2003. **72**: p. 291-336.

68. Blencowe, B.J., *Alternative splicing: new insights from global analyses.* Cell, 2006. **126**(1): p. 37-47.

69. Matlin, A.J., F. Clark, and C.W. Smith, *Understanding alternative splicing: towards a cellular code.* Nat Rev Mol Cell Biol, 2005. **6**(5): p. 386-98.

70. Wang, Z. and C.B. Burge, *Splicing regulation: from a parts list of regulatory elements to an integrated splicing code.* RNA, 2008. **14**(5): p. 802-13.

71. Geuens, T., D. Bouhy, and V. Timmerman, *The hnRNP family: insights into their role in health and disease.* Hum Genet, 2016. **135**(8): p. 851-67.

72. Sahebi, M., et al., *Towards understanding pre-mRNA splicing mechanisms and the role of SR proteins.* Gene, 2016. **587**(2): p. 107-19.

73. Lee, Y. and D.C. Rio, *Mechanisms and Regulation of Alternative Pre-mRNA Splicing.* Annu Rev Biochem, 2015. **84**: p. 291-323.

74. Erkelenz, S., et al., *Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms.* RNA, 2013. **19**(1): p. 96-102.

75. Reber, S., et al., *Minor intron splicing is regulated by FUS and affected by ALS-associated FUS mutants.* EMBO J, 2016. **35**(14): p. 1504-21.

76. Han, K., et al., *A combinatorial code for splicing silencing: UAGG and GGGG motifs.* PLoS Biol, 2005. **3**(5): p. e158.

77. Smith, C.W. and J. Valcarcel, *Alternative pre-mRNA splicing: the logic of combinatorial control.* Trends Biochem Sci, 2000. **25**(8): p. 381-8.

78. Zhang, X.H. and L.A. Chasin, *Computational definition of sequence motifs governing constitutive exon splicing.* Genes Dev, 2004. **18**(11): p. 1241-50.

79. Maatz, H., et al., *RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing.* J Clin Invest, 2014. **124**(8): p. 3419-30.

80. Yang, J., et al., *RBM24 is a major regulator of muscle-specific alternative splicing.* Dev Cell, 2014. **31**(1): p. 87-99.

81. Kohtz, J.D., et al., *Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors.* Nature, 1994. **368**(6467): p. 119-24.

82. Wu, J.Y. and T. Maniatis, *Specific interactions between proteins implicated in splice site selection and regulated alternative splicing.* Cell, 1993. **75**(6): p. 1061-70.

83. Graveley, B.R., K.J. Hertel, and T. Maniatis, *A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers.* EMBO J, 1998. **17**(22): p. 6747-56.

84. Soret, J. and J. Tazi, *Phosphorylation-dependent control of the pre-mRNA splicing machinery.* Prog Mol Subcell Biol, 2003. **31**: p. 89-126.

85. Manley, J.L. and A.R. Krainer, *A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins).* Genes Dev, 2010. **24**(11): p. 1073-4.

86. Birney, E., S. Kumar, and A.R. Krainer, *Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors.* Nucleic Acids Res, 1993. **21**(25): p. 5803-16.

87. Gross, T., et al., *Identification and characterization of srp1, a gene of fission yeast encoding a RNA binding domain and a RS domain typical of SR splicing factors.* Nucleic Acids Res, 1998. **26**(2): p. 505-11.

88. Twyffels, L., C. Gueydan, and V. Kruys, *Shuttling SR proteins: more than splicing factors.* FEBS J, 2011. **278**(18): p. 3246-55.

89. Tacke, R. and J.L. Manley, *Determinants of SR protein specificity.* Curr Opin Cell Biol, 1999. **11**(3): p. 358-62.

90. Liu, H.X., M. Zhang, and A.R. Krainer, *Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins.* Genes Dev, 1998. **12**(13): p. 1998-2012.

91. Cavaloc, Y., et al., *The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers.* RNA, 1999. **5**(3): p. 468-83.

92. Lynch, K.W. and T. Maniatis, *Synergistic interactions between two distinct elements of a regulated splicing enhancer.* Genes Dev, 1995. **9**(3): p. 284-93.

93. Lynch, K.W. and T. Maniatis, *Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer.* Genes Dev, 1996. **10**(16): p. 2089-101.

94. Tacke, R., et al., *Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing.* Cell, 1998. **93**(1): p. 139-48.

95. Cao, W. and M.A. Garcia-Blanco, *A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding.* J Biol Chem, 1998. **273**(32): p. 20629-35.

96. Cho, S., et al., *Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8233-8.

97. Xiao, S.H. and J.L. Manley, *Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing.* Genes Dev, 1997. **11**(3): p. 334-44.

98. Jamison, S.F., et al., *U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements.* Nucleic Acids Res, 1995. **23**(16): p. 3260-7.

99. Shen, H. and M.R. Green, *A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly.* Mol Cell, 2004. **16**(3): p. 363-73.

100. Shen, H. and M.R. Green, *RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans.* Genes Dev, 2006. **20**(13): p. 1755-65.

101. Lavigueur, A., et al., *A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding.* Genes Dev, 1993. **7**(12A): p. 2405-17.

102. Tian, M. and T. Maniatis, *A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA.* Cell, 1993. **74**(1): p. 105-14.

103. Wang, Z., H.M. Hoffmann, and P.J. Grabowski, *Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity.* RNA, 1995. **1**(1): p. 21-35.

104. Zuo, P. and T. Maniatis, *The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing.* Genes Dev, 1996. **10**(11): p. 1356-68.

105. Makarova, O.V., E.M. Makarov, and R. Luhrmann, *The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes.* EMBO J, 2001. **20**(10): p. 2553-63.

106. Roscigno, R.F. and M.A. Garcia-Blanco, *SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome.* RNA, 1995. **1**(7): p. 692-706.

107. Buratti, E., et al., *SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer.* Nucleic Acids Res, 2007. **35**(13): p. 4359-68.

108. Dauksaite, V. and G. Akusjarvi, *Human splicing factor ASF/SF2 encodes for a repressor domain required for its inhibitory activity on pre-mRNA splicing.* J Biol Chem, 2002. **277**(15): p. 12579-86.

109. Kanopka, A., O. Muhlemann, and G. Akusjarvi, *Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA.* Nature, 1996. **381**(6582): p. 535-8.

110. Shen, M. and W. Mattox, *Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position.* Nucleic Acids Res, 2012. **40**(1): p. 428-37.

111. Sharma, S., et al., *U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression.* Mol Cell, 2011. **41**(5): p. 579-88.

112. Hoffman, D.W., et al., *RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins.* Proc Natl Acad Sci U S A, 1991. **88**(6): p. 2495-9.

113. Han, S.P., Y.H. Tang, and R. Smith, *Functional diversity of the hnRNPs: past, present and perspectives.* Biochem J, 2010. **430**(3): p. 379-92.

114. Pinol-Roma, S., *HnRNP proteins and the nuclear export of mRNA.* Semin Cell Dev Biol, 1997. **8**(1): p. 57-63.

115. Krecic, A.M. and M.S. Swanson, *hnRNP complexes: composition, structure, and function.* Curr Opin Cell Biol, 1999. **11**(3): p. 363-71.

116. Jean-Philippe, J., S. Paz, and M. Caputi, *hnRNP A1: the Swiss army knife of gene expression.* Int J Mol Sci, 2013. **14**(9): p. 18999-9024.

117. Wang, Y., et al., *A complex network of factors with overlapping affinities represses splicing through intronic elements.* Nat Struct Mol Biol, 2013. **20**(1): p. 36-45.

118. Zhu, J., A. Mayeda, and A.R. Krainer, *Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins.* Mol Cell, 2001. **8**(6): p. 1351-61.

119. Okunola, H.L. and A.R. Krainer, *Cooperative-binding and splicing-repressive properties of hnRNP A1.* Mol Cell Biol, 2009. **29**(20): p. 5620-31.

120. Blanchette, M. and B. Chabot, *Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization.* EMBO J, 1999. **18**(7): p. 1939-52.

121. Nasim, F.U., et al., *High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism.* RNA, 2002. **8**(8): p. 1078-89.

122. Zarnack, K., et al., *Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements.* Cell, 2013. **152**(3): p. 453-66.

123. Caputi, M. and A.M. Zahler, *Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family.* J Biol Chem, 2001. **276**(47): p. 43850-9.

124. Schaub, M.C., S.R. Lopez, and M. Caputi, *Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes.* J Biol Chem, 2007. **282**(18): p. 13617-26.

125. Dominguez, C., et al., *Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs.* Nat Struct Mol Biol, 2010. **17**(7): p. 853-61.

126. Martinez-Contreras, R., et al., *Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing.* PLoS Biol, 2006. **4**(2): p. e21.

127. Widera, M., et al., *An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA.* J Virol, 2013. **87**(5): p. 2707-20.

128. Chen, C.D., R. Kobayashi, and D.M. Helfman, *Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene.* Genes Dev, 1999. **13**(5): p. 593-606.

129. Buratti, E., et al., *hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes.* Nucleic Acids Res, 2004. **32**(14): p. 4224-36.

130. Exline, C.M., Z. Feng, and C.M. Stoltzfus, *Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression.* J Virol, 2008. **82**(8): p. 3921-31.

131. Fogel, B.L. and M.T. McNally, *A cellular protein, hnRNP H, binds to the negative regulator of splicing element from Rous sarcoma virus.* J Biol Chem, 2000. **275**(41): p. 32371-8.

132. Jacquenet, S., et al., *A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H.* J Biol Chem, 2001. **276**(44): p. 40464-75.

133. Chou, M.Y., et al., *hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells.* Mol Cell Biol, 1999. **19**(1): p. 69-77.

134. Garneau, D., et al., *Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x.* J Biol Chem, 2005. **280**(24): p. 22641-50.

135. Hastings, M.L., C.M. Wilson, and S.H. Munroe, *A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA.* RNA, 2001. **7**(6): p. 859-74.

136. McCullough, A.J. and S.M. Berget, *G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection.* Mol Cell Biol, 1997. **17**(8): p. 4562-71.

137. Min, H., R.C. Chan, and D.L. Black, *The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event.* Genes Dev, 1995. **9**(21): p. 2659-71.

138. Yeo, G., et al., *Variation in sequence and organization of splicing regulatory elements in vertebrate genes.* Proc Natl Acad Sci U S A, 2004. **101**(44): p. 15700-5.

139. Majewski, J. and J. Ott, *Distribution and characterization of regulatory elements in the human genome.* Genome Res, 2002. **12**(12): p. 1827-36.

140. Xiao, X., et al., *Splice site strength-dependent activity and genetic buffering by poly-G runs.* Nat Struct Mol Biol, 2009. **16**(10): p. 1094-100.

141. Wang, E., et al., *G Run-mediated recognition of proteolipid protein and DM20 5' splice sites by U1 small nuclear RNA is regulated by context and proximity to the splice site.* J Biol Chem, 2011. **286**(6): p. 4059-71.

142. Patton, J.G., et al., *Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing.* Genes Dev, 1991. **5**(7): p. 1237-51.

143. Chan, R.C. and D.L. Black, *Conserved intron elements repress splicing of a neuron-specific c-src exon in vitro.* Mol Cell Biol, 1995. **15**(11): p. 6377-85.

144. Perez, I., et al., *Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo.* RNA, 1997. **3**(7): p. 764-78.

145. Ashiya, M. and P.J. Grabowski, *A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart.* RNA, 1997. **3**(9): p. 996-1015.

146. Amir-Ahmady, B., et al., *Exon repression by polypyrimidine tract binding protein.* RNA, 2005. **11**(5): p. 699-716.

147. Spellman, R. and C.W. Smith, *Novel modes of splicing repression by PTB.* Trends Biochem Sci, 2006. **31**(2): p. 73-6.

148. Llorian, M., et al., *Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB.* Nat Struct Mol Biol, 2010. **17**(9): p. 1114-23.

149. Hillebrand, F., et al., *Differential hnRNP D isoform incorporation may confer plasticity to the ESSV-mediated repressive state across HIV-1 exon 3.* Biochim Biophys Acta, 2017. **1860**(2): p. 205-217.

150. Wang, J., et al., *SpliceDisease database: linking RNA splicing and disease.* Nucleic Acids Res, 2012. **40**(Database issue): p. D1055-9.

151. Singh, R.K. and T.A. Cooper, *Pre-mRNA splicing in disease and therapeutics.* Trends Mol Med, 2012. **18**(8): p. 472-82.

152. Scotti, M.M. and M.S. Swanson, *RNA mis-splicing in disease.* Nat Rev Genet, 2016. **17**(1): p. 19-32.

153. Treisman, R., et al., *A single-base change at a splice site in a beta 0-thalassemic gene causes abnormal RNA splicing.* Cell, 1982. **29**(3): p. 903-11.

154. Lorson, C.L., et al., *A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy.* Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6307-11.

155. Kashima, T. and J.L. Manley, *A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy.* Nat Genet, 2003. **34**(4): p. 460-3.

156. Kim, H.J., et al., *Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS.* Nature, 2013. **495**(7442): p. 467-73.

157. Erkelenz, S., et al., *Genomic HEXploring allows landscaping of novel potential splicing regulatory elements.* Nucleic Acids Res, 2014. **42**(16): p. 10681-97.

158. Soukarieh, O., et al., *Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools.* PLoS Genet, 2016. **12**(1): p. e1005756.

159. Fairbrother, W.G., et al., *Predictive identification of exonic splicing enhancers in human genes.* Science, 2002. **297**(5583): p. 1007-13.

160. Little, S.J., et al., *Antiretroviral-drug resistance among patients recently infected with HIV.* N Engl J Med, 2002. **347**(6): p. 385-94.

161. Belmonte, L., et al., *Reservoirs of HIV replication after successful combined antiretroviral treatment.* Curr Med Chem, 2003. **10**(4): p. 303-12.

162. Eisele, E. and R.F. Siliciano, *Redefining the viral reservoirs that prevent HIV-1 eradication.* Immunity, 2012. **37**(3): p. 377-88.

163. Alvarez, M. and L. Menendez-Arias, *Temperature effects on the fidelity of a thermostable HIV-1 reverse transcriptase.* FEBS J, 2014. **281**(1): p. 342-51.

164. Maddon, P.J., et al., *The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain.* Cell, 1986. **47**(3): p. 333-48.

165. Clapham, P.R. and A. McKnight, *HIV-1 receptors and cell tropism.* Br Med Bull, 2001. **58**: p. 43-59.

166. Chan, D.C. and P.S. Kim, *HIV entry and its inhibition.* Cell, 1998. **93**(5): p. 681-4.

167. Berger, E.A., P.M. Murphy, and J.M. Farber, *Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease.* Annu Rev Immunol, 1999. **17**: p. 657-700.

168. Saadatmand, J. and L. Kleiman, *Aspects of HIV-1 assembly that promote primer tRNA(Lys3) annealing to viral RNA.* Virus Res, 2012. **169**(2): p. 340-8.

169. Bukrinsky, M.I., et al., *Active nuclear import of human immunodeficiency virus type 1 preintegration complexes.* Proc Natl Acad Sci U S A, 1992. **89**(14): p. 6580-4.

170. Popov, S., et al., *Viral protein R regulates docking of the HIV-1 preintegration complex to the nuclear pore complex.* J Biol Chem, 1998. **273**(21): p. 13347-52.

171. Craigie, R. and F.D. Bushman, *HIV DNA integration.* Cold Spring Harb Perspect Med, 2012. **2**(7): p. a006890.

172. Leblanc, J., J. Weil, and K. Beemon, *Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA.* Wiley Interdiscip Rev RNA, 2013. **4**(5): p. 567-80.

173. Karn, J. and C.M. Stoltzfus, *Transcriptional and posttranscriptional regulation of HIV-1 gene expression.* Cold Spring Harb Perspect Med, 2012. **2**(2): p. a006916.

174. Ott, M., M. Geyer, and Q. Zhou, *The control of HIV transcription: keeping RNA polymerase II on track.* Cell Host Microbe, 2011. **10**(5): p. 426-35.

175. Kozak, M., *Pushing the limits of the scanning mechanism for initiation of translation.* Gene, 2002. **299**(1-2): p. 1-34.

176. Purcell, D.F. and M.A. Martin, *Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity.* J Virol, 1993. **67**(11): p. 6365-78.

177. Ocwieja, K.E., et al., *Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing.* Nucleic Acids Res, 2012. **40**(20): p. 10345-55.

178. Emery, A., et al., *Characterizing HIV-1 Splicing by Using Next-Generation Sequencing.* J Virol, 2017. **91**(6).

179. Kim, S.Y., et al., *Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression.* J Virol, 1989. **63**(9): p. 3708-13.

180. Klotman, M.E., et al., *Kinetics of expression of multiply spliced RNA in early human immunodeficiency virus type 1 infection of lymphocytes and monocytes.* Proc Natl Acad Sci U S A, 1991. **88**(11): p. 5011-5.

181. Michael, N.L., et al., *Induction of human immunodeficiency virus type 1 expression in chronically infected cells is associated primarily with a shift in RNA splicing patterns.* J Virol, 1991. **65**(3): p. 1291-303.

182. O'Reilly, M.M., M.T. McNally, and K.L. Beemon, *Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA.* Virology, 1995. **213**(2): p. 373-85.

183. Stoltzfus, C.M., *Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication.* Adv Virus Res, 2009. **74**: p. 1-40.

184. Dayton, A.I., et al., *The trans-activator gene of the human T cell lymphotropic virus type III is required for replication.* Cell, 1986. **44**(6): p. 941-7.

185. Fisher, A.G., et al., *The trans-activator gene of HTLV-III is essential for virus replication.* Nature, 1986. **320**(6060): p. 367-71.

186. Amendt, B.A., et al., *Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1.* Mol Cell Biol, 1994. **14**(6): p. 3960-70.

187. Amendt, B.A., Z.H. Si, and C.M. Stoltzfus, *Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors.* Mol Cell Biol, 1995. **15**(11): p. 6480.

188. Si, Z., B.A. Amendt, and C.M. Stoltzfus, *Splicing efficiency of human immunodeficiency virus type 1 tat RNA is determined by both a suboptimal 3' splice site and a 10 nucleotide exon splicing silencer element located within tat exon 2.* Nucleic Acids Res, 1997. **25**(4): p. 861-7.

189. Hallay, H., et al., *Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site.* J Biol Chem, 2006. **281**(48): p. 37159-74.

190. Zahler, A.M., et al., *SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing.* J Biol Chem, 2004. **279**(11): p. 10077-84.

191. Tange, T.O., et al., *The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element.* EMBO J, 2001. **20**(20): p. 5748-58.

192. Si, Z.H., D. Rauch, and C.M. Stoltzfus, *The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly.* Mol Cell Biol, 1998. **18**(9): p. 5404-13.

193. Staffa, A. and A. Cochrane, *Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1.* Mol Cell Biol, 1995. **15**(8): p. 4597-605.

194. Damgaard, C.K., T.O. Tange, and J. Kjems, *hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure.* RNA, 2002. **8**(11): p. 1401-15.

195. Marchand, V., et al., *A Janus splicing regulatory element modulates HIV-1 tat and rev mRNA production by coordination of hnRNP A1 cooperative binding.* J Mol Biol, 2002. **323**(4): p. 629-52.

196. Malim, M.H., et al., *The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA.* Nature, 1989. **338**(6212): p. 254-7.

197. Zapp, M.L., et al., *Oligomerization and RNA binding domains of the type 1 human immunodeficiency virus Rev protein: a dual function for an arginine-rich binding motif.* Proc Natl Acad Sci U S A, 1991. **88**(17): p. 7734-8.

198. Malim, M.H. and B.R. Cullen, *HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency.* Cell, 1991. **65**(2): p. 241-8.

199. Fischer, U., et al., *The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs.* Cell, 1995. **82**(3): p. 475-83.

200. Fornerod, M., et al., *CRM1 is an export receptor for leucine-rich nuclear export signals.* Cell, 1997. **90**(6): p. 1051-60.

201. Henderson, B.R. and P. Percipalle, *Interactions between HIV Rev and nuclear import and export factors: the Rev nuclear localisation signal mediates specific binding to human importin-beta.* J Mol Biol, 1997. **274**(5): p. 693-707.

202. Caputi, M., et al., *A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression.* J Virol, 2004. **78**(12): p. 6517-26.

203. Asang, C., I. Hauber, and H. Schaal, *Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer.* Nucleic Acids Res, 2008. **36**(5): p. 1450-63.

204. Kirchhoff, F., et al., *Role of Nef in primate lentiviral immunopathogenesis.* Cell Mol Life Sci, 2008. **65**(17): p. 2621-36.

205. Greenway, A.L., et al., *HIV-1 Nef control of cell signalling molecules: multiple strategies to promote virus replication.* J Biosci, 2003. **28**(3): p. 323-35.

206. Stein, B.S. and E.G. Engleman, *Intracellular processing of the gp160 HIV-1 envelope precursor. Endoproteolytic cleavage occurs in a cis or medial compartment of the Golgi complex.* J Biol Chem, 1990. **265**(5): p. 2640-9.

207. Klatzmann, D., et al., *Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes.* Science, 1984. **225**(4657): p. 59-63.

208. Choe, H., et al., *The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates.* Cell, 1996. **85**(7): p. 1135-48.

209. Schaal, H., et al., *Requirement of N-terminal amino acid residues of gp41 for human immunodeficiency virus type 1-mediated cell fusion.* J Virol, 1995. **69**(6): p. 3308-14.

210. Sheehy, A.M., et al., *Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein.* Nature, 2002. **418**(6898): p. 646-50.

211. Harris, R.S., J.F. Hultquist, and D.T. Evans, *The restriction factors of human immunodeficiency virus.* J Biol Chem, 2012. **287**(49): p. 40875-83.

212. Wissing, S., N.L. Galloway, and W.C. Greene, *HIV-1 Vif versus the APOBEC3 cytidine deaminases: an intracellular duel between pathogen and host restriction factors.* Mol Aspects Med, 2010. **31**(5): p. 383-97.

213. Mandal, D., et al., *Regulation of Vif mRNA splicing by human immunodeficiency virus type 1 requires 5' splice site D2 and an exonic splicing enhancer to counteract cellular restriction factor APOBEC3G.* J Virol, 2009. **83**(12): p. 6067-78.

214. Goh, W.C., et al., *HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo.* Nat Med, 1998. **4**(1): p. 65-71.

215. Heinzinger, N.K., et al., *The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells.* Proc Natl Acad Sci U S A, 1994. **91**(15): p. 7311-5.

216. Jowett, J.B., et al., *The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle.* J Virol, 1995. **69**(10): p. 6304-13.

217. Bilodeau, P.S., et al., *RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element.* J Virol, 2001. **75**(18): p. 8487-97.

218. Domsic, J.K., et al., *Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts.* Mol Cell Biol, 2003. **23**(23): p. 8762-72.

219. Madsen, J.M. and C.M. Stoltzfus, *An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication.* J Virol, 2005. **79**(16): p. 10478-86.

220. Erkelenz, S., et al., *Tra2-mediated recognition of HIV-1 5' splice site D3 as a key factor in the processing of vpr mRNA.* J Virol, 2013. **87**(5): p. 2721-34.

221. Gonzalez, M.E., *Vpu Protein: The Viroporin Encoded by HIV-1.* Viruses, 2015. **7**(8): p. 4352-68.

222. Guatelli, J.C., *Interactions of viral protein U (Vpu) with cellular factors.* Curr Top Microbiol Immunol, 2009. **339**: p. 27-45.

223. Jacks, T., et al., *Characterization of ribosomal frameshifting in HIV-1 gag-pol expression.* Nature, 1988. **331**(6153): p. 280-3.

224. Wilson, W., et al., *HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems.* Cell, 1988. **55**(6): p. 1159-69.

225. Gaudin, C., et al., *Structure of the RNA signal essential for translational frameshifting in HIV-1.* J Mol Biol, 2005. **349**(5): p. 1024-35.