Deciphering principles of recognition of human immunodeficiency virus (HIV) splice sites

Inaugural-Dissertation

zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Jan Otto Peter

aus Düsseldorf

Düsseldorf, August 2015

aus dem Institut für Virologie

der Heinrich-Heine-Universität Düsseldorf

Für meine Eltern

Zusammenfassung

Ein Kennzeichen des Humanen Immundefizienz-Virus Typ 1 (HIV-1) ist die strikte Abhängigkeit der viralen Replikation vom alternativen Spleißen. Durch alternatives Spleißen entstehen aus der viralen prä-mRNA mehr als 40 verschiedene Spleißisoformen, die für wenigstens 18 virale Proteine kodieren. Um diese exprimieren zu können und eine effiziente virale Replikation zu ermöglichen, muss es eine fein ausgewogene Balance zwischen gespleißten und intronhaltigen mRNA-Isoformen geben. Die mutationsbasierte Inaktivierung des *exonic splicing silencer V* (ESSV) führt zu exzessivem Spleißen der HIV-1 prä-mRNA. Dadurch verschiebt sich die Balance zu gespleißten mRNA-Isoformen und verhindert somit die HIV-1 Replikation. Allerdings führte die Depletion des publizierten ESSV-Bindeproteins hnRNP A1, nicht wie angenommen, zu exzessivem Spleißen. Deshalb musste die Bindung eines anderen RNA-bindenden Proteins, welches die spleißinhibitorische Aktivität des ESSV aufrecht hält, angenommen werden. Im Rahmen dieser Doktorarbeit wurde gezeigt, dass hnRNP D, ein bekannter Regulator der mRNA-Stabilität und Erhaltung der Telomere, ebenfalls spezifisch an den ESSV binden kann. Weiterhin konnte nachgewiesen werden, dass die RNA-Bindung aller hnRNP D Isoformen die Erkennung von Spleißstellen beeinträchtigt und das Spleißen reprimiert. Zudem konnte durch Deletionsmutagenese demonstriert werden, dass in Abhängigkeit von der durch das Exon 7 kodierten Peptidsequenz hnRNP D auf zwei unterschiedliche Weisen das Spleißen der prä-mRNA inhibieren kann. Schließlich wurde gezeigt, dass eine exklusive hnRNP D Bindestelle anstelle des ESSV ebenfalls eine ausgewogene Balance zwischen gespleißter und intronhaltiger mRNA-Isoformen aufrechterhalten kann. Zusammengefasst konnte in dieser Doktorarbeit nachgewiesen werden, dass hnRNP D auch ein spleißregulatorisches Protein ist, das zusammen mit hnRNP A1 am ESSV das exzessive Spleißen der HIV-1 prä-mRNA unterbindet und somit zur effizienten HIV-1 Replikation beiträgt.

Summary

One hallmark of Human Immunodeficiency Virus Type 1 (HIV-1) is the strict dependence of its viral replication on alternative splicing. Driven by alternative splicing more than 40 different transcript isoforms arise from the viral pre-mRNA, which encode at least 18 viral proteins. In order to express all viral proteins allowing efficient viral replication, a fragile balance between the intronless and intron-containing mRNAs has to be kept. Mutational inactivation of the exonic splicing silencer V (ESSV) was shown to result in excessive splicing of the HIV-1 pre-mRNA, shifting the balance towards intronless mRNA isoforms and thus inhibiting HIV-1 replication. However, depletion of the published ESSV binding protein hnRNP A1 did not as anticipated result in the onset of excessive splicing. Thus, binding of another RNA-binding protein could be supposed which sustains the splice inhibitory activity of the ESSV. In this thesis it was shown that hnRNP D, a known regulator of mRNA stability and telomere maintenance, specifically binds to the ESSV. Further, binding of each hnRNP D isoform to the RNA could be observed to interfere with the recognition of splice sites and thus to inhibit splicing. Moreover, deletion mutagenesis demonstrated that dependent on the peptide sequence encoded by alternative exon 7 hnRNP D mediated inhibition of pre-mRNA splicing occurs by two different mechanisms. Finally, an exclusive hnRNP D binding site replacing the ESSV was shown to maintain a viable balance of intronless to intron-containing mRNA isoforms. Altogether, hnRNP D represents a splicing regulatory protein that binds along with hnRNP A1 to the ESSV and thereby contributes to the regulation of HIV-1 pre-mRNA splicing, the prevention of excessive splicing and thus allows efficient viral replication.

Index

Introduction

1.1 HIV-1 life cycle

The immunodeficiency virus type-1 (HIV-1) belongs to the family of *Retroviridae*. The hallmark of the *Retroviridae* is their reliance on reverse transcription of their single stranded RNA (ssRNA) genome into double stranded DNA (dsDNA), the active nuclear import and henceforth integration of the proviral DNA into the host genome. While with some genera of retroviruses infection is followed by the immediate onset of proviral gene expression, viruses belonging to the genus *Lentiviridae*, as HIV-1, characteristically show a long period of latency before progression to viral gene expression.
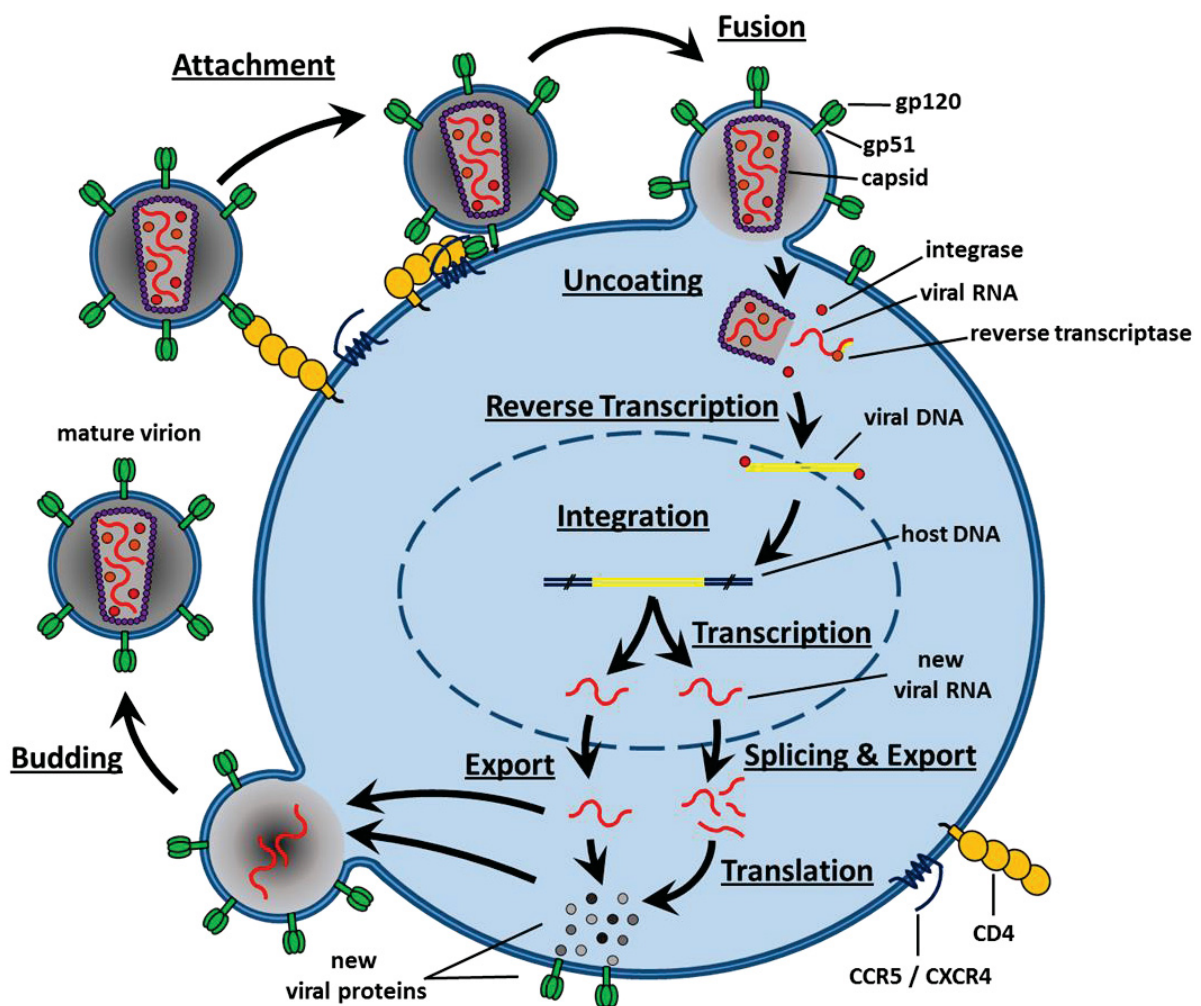
Introduction

Fig. 1.1: HIV-1 life cycle

The HIV-1 life cycle begins with binding of the viral surface glycoprotein Env gp120 to the cellular CD4 receptor. Receptor binding of Env gp120 results in conformational changes leading to the exposure of the coreceptor binding site. Thereafter, binding of Env gp120 to the coreceptor CXCR4 or CCR5 triggers Env gp41-mediated membrane fusion resulting in the release of the viral capsid into the cytoplasm. Following partial uncoating of the capsid, the viral ssRNA genome is reversed transcribed, actively transported into the nucleus and randomly integrated into the host's genome. After transcription of proviral DNA, the primary transcript is alternatively spliced leading to more than 40 different mRNA isoforms that encode at least 18 viral proteins. New virions consisting out of translated viral proteins and two copies of the unspliced genomic (+) ssRNA assemble at the cellular membrane and bud into the extracellular space thereby acquiring the viral envelope. Activation of the viral protease within the budded immature virions results in structural rearrangements and thus the maturation of the released virions.

The susceptibility of a given cell to HIV-1 infection relies on the presence of two distinct cellular surface glycoproteins functioning as viral receptors, the primary receptor CD4 and the coreceptor given by either the cellular chemokine receptor CCR5 or CXCR4. Attachment of HIV-1 to host cells is mediated by its viral glycoprotein Env binding to CD4 surface glycoprotein, which is expressed on mononuclear cells including a subset of T cells (CD4+), macrophages (Mφ) and dendritic cells (DCs) [41, 47, 193] (Fig. 1.1). The CD4-binding dependent conformational switch within the Env gp120 subunit results in proper positioning of the V-domain and exposure of the coreceptor binding site [47, 74]. Thereafter, specific coreceptor binding is mediated by the highly variable V3 loop of the Env gp120 subunit, whose exchange could be shown to alter the viral tropism by allowing specific binding of either CCR5, CXCR4 or both [74, 157, 170]. The interaction of the V3 loop of the Env gp120 subunit and the coreceptor triggers Env gp41 dependent membrane fusion of the viral envelope with the cellular membrane [38, 226]. The nucleocapsid released within the cytoplasm of the host cell undergoes rapid uncoating. Following the uncoating, the viral ssRNA (+) genome is reverse transcribed by the viral reverse transcriptase into a linear dsDNA molecule, the proviral DNA. After reverse transcription, the resulting proviral DNA is associated with viral and cellular proteins, forming the so called pre-integration complex (PIC). One unique feature of the genus of *Lentiviridae* is the capacity of the pre-integration complex to be actively imported into the nucleus, allowing HIV-1 also to infect non-dividing cells, such as macrophages. After nuclear import, the proviral DNA is randomly integrated into the host genome by the viral Integrase (IN). Viral gene expression is controlled by a

promoter located within the 5' long terminal repeat (5' LTR). Thereby, independent of the integration site RNA-polymerase II dependent transcription underlies epigenetic regulation [175, 213] and occurs shortly after integration at a rather low level [106]. Splicing of the first transcribed pre-mRNAs results in multiple spliced mRNA isoforms of which some encode the *trans*-activator of transcription (Tat). Once translated in the cytoplasm, Tat is actively imported into the nucleus and binds to the transactivation responsive region (TAR) forming on each nascent viral pre-mRNA during transcription. Subsequent recruitment of the positive transcription elongation factor b (pTEFb) results in hyperphosphorylation of the C-terminal domain of RNA-Polymerase II thereby dramatically increasing its transcriptional processivity [161, 167]. The positive feedback loop consisting of the expression of Tat resulting in enhancement of transcription at the 5' LTR promoter results in accumulation of multiple spliced mRNAs encoding the viral proteins Tat, Nef and Rev [173]. Following its translation, Rev is actively imported into the nucleus and binds to the Rev responsive element (RRE), which is located tat/rev intron flanked by its coding exons. Binding of Rev to the RRE mediates the export of all intron-containing mRNAs encoding the viral infectivity factor (Vif), the viral protein R (Vpr), the viral protein U (Vpu), and the envelope glycoprotein (Env). Additionally, the nuclear export of the completely unspliced viral mRNA, which serves as mRNA for the viral structural proteins Gag and Gag-Pol, but also represents the viral genomic RNA, is also mediated by Rev. Assembly of HIV-1 virions occurs at the plasma membrane. There the myristoylated HIV-1 Gag and Gag-Pol precursor protein mediate all necessary steps including binding to the plasma membrane, establishing protein-protein interactions necessary to generate a spherical shape, concentrating viral Env protein at the membrane and packaging of the viral genomic RNA [207]. Each HIV-1 virion incorporates two copies of the viral genomic RNA. Prior packaging both copies are non-covalently linked by the formation of an RNA-duplex, called kissing-loop structure, starting at the dimer initiation site (DIS) in the 5'UTR of both RNAs. The formed genomic RNA dimers are packaged within the assembling virions by interaction of the nucleocapsid protein incorporated within the Gag and Gag-Pol precursor protein and the packaging signal. Completely assembled virions bud from the cellular membrane and are released into the extracellular space, thereby acquiring their envelope harbouring the viral glycoproteins. All released virions are initially immature. Maturation is thought to start with the transient dimerization and subsequent auto-catalytic

cleavage of Gag-Pol precursor protein by the incorporated viral protease [210]. Thereafter, the auto-catalytically liberated protease subunits assemble into dimers that efficiently cleave more Gag and Gag-Pol precursor proteins resulting in structural rearrangements leading finally to the formation of mature and infectious virions.

1.2 The dependence of HIV-1 replication on alternative splicing

Initiation of eukaryotic translation depends on binding of the ribosomal 43S subunit to the 5'-cap structure of the mRNA. Subsequently the mRNA is scanned for a favorable start codon where translation is initiated. Defying that common rule of eukaryotic gene expression HIV-1 generates at least 40 different transcripts by alternative splicing [173] resulting in the expression at least 18 different viral proteins from a single pre-mRNA [101] (Fig. 1.2). Thereby, the inhibitory upstream located translational start codons are removed by alternative splicing and thus the juxtaposition of start codons to a cap-proximal position enables the translation of each viral open reading frame (ORF). The only exception to that principle is the glycoprotein expression from a bi-cistronic *vpu*/*env* mRNA. Here, a minimal upstream open reading frame (uORF) overlapping with the *vpu* translational start codon allows efficient translation of downstream located *env* ORF [5, 116]. During the early phase of viral gene expression only intronless mRNAs of the 2kb class are exported from the nucleus encoding the viral proteins Tat, Nef and Rev [112, 113, 151, 189]. HIV-1 Rev, then shuttles back into the nucleus, binds to the RRE and mediates the nuclear export of intron-containing viral mRNAs by the Crm1-dependent export pathway [169]. This enables the onset of late viral gene expression by mediating the export of not only intron-containing viral mRNAs belonging to the 4Kb class of viral mRNAs, but also of the unspliced 9Kb genomic RNA [112, 151, 190]. Finally, alternative splicing of the viral pre-mRNA, combined with Rev-mediated export, generates a balanced expression of all viral proteins resulting in efficient viral replication.
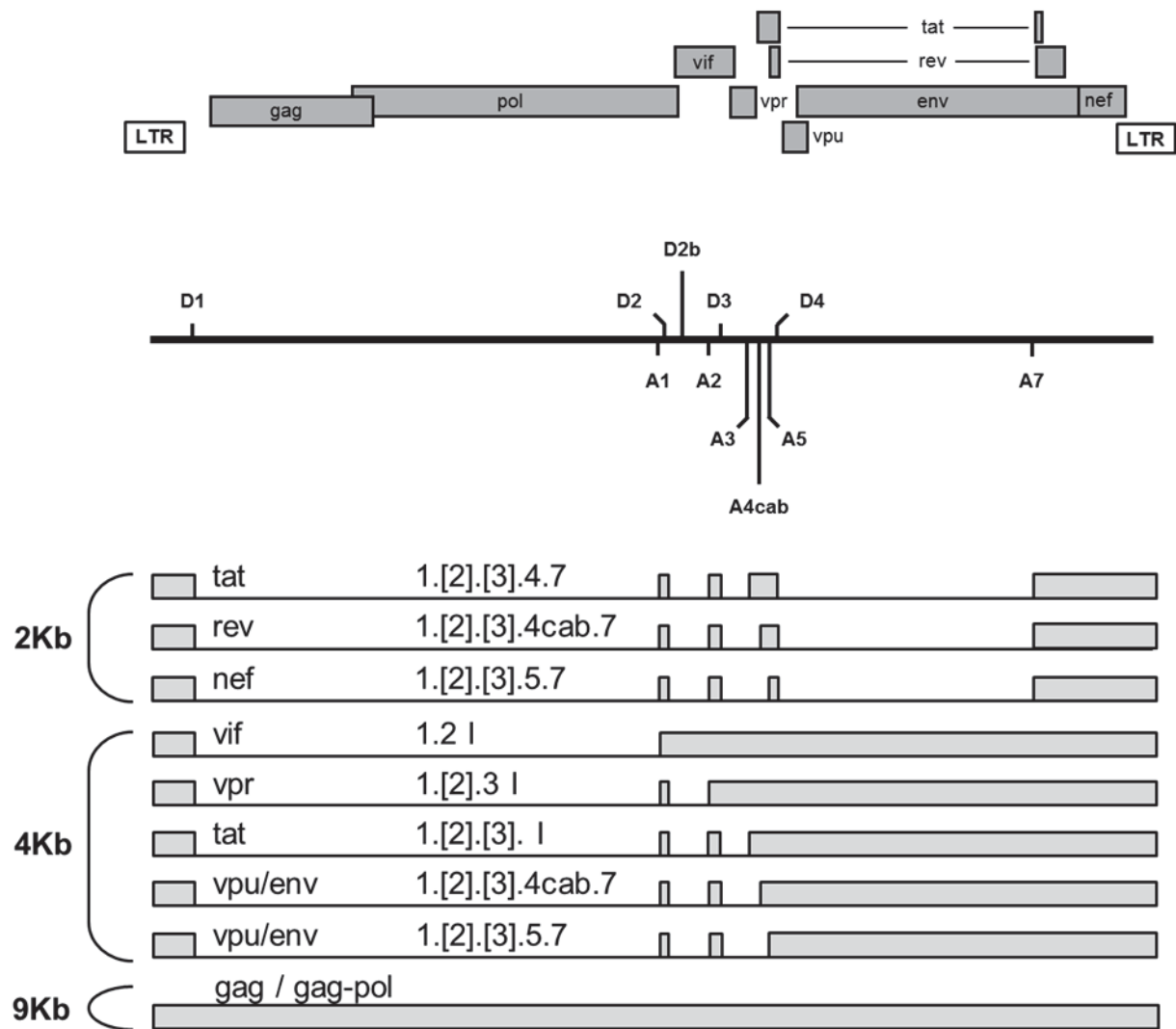
Fig. 1.2: Schematic diagram of viral mRNA isoforms generated by alternative splicing.

(Top panel) Organisation of the proviral genome. Open reading frames (ORFs) are indicated as grey boxes. The long terminal repeats (LTR) located at both ends of the HIV-1 genome are depicted as white boxes. (Center panel) HIV-1 pre-mRNA containing all viral 5' splice sites (5'ss) and 3' splice sites (3'ss). (Bottom panel) Illustration of viral mRNA isoforms belonging to the viral RNA classes 2Kb, 4Kb and 9Kb. Exons are depicted as boxes and introns are indicated by lines. Each depicted mRNA isoform is tagged according to its translated open-reading frame and the included exons. Extention of original exon is marked by an "I". Alternative non-coding leader exons of mRNA isoforms are marked by parenthesis within exon enumeration and the mRNA isoform containing all exons is depicted.

1.3   The spliceosome - catalysis and complex assembly

Splicing of the pre-mRNA is catalysed by the spliceosome resulting in the cleavage of the intron and joining of the adjacent exons by two transesterification reactions. The spliceosome contains five small nuclear RNAs (snRNAs), the U1, U2, U4, U5 and U6 snRNA, which form the corresponding five small nuclear ribonucleic particles (snRNPs) by association with several particle-specific proteins. The first step of spliceosome formation represents the recognition of RNA sequences defining the exon and intron boundaries, the 5' splice site (5'ss) characterised by the canonical GU-dinucleotide, the 3' splice site (3'ss) composed of the polypyrimidine tract (PPT) and the canonical AG-dinucleotide and the branch point sequence (BPS) featuring the branch point adenosine (Fig 1.3 I).
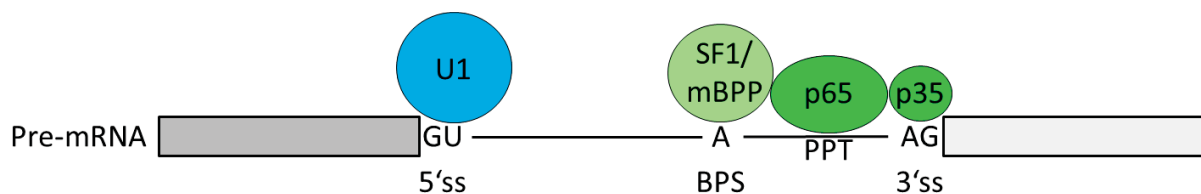
Fig. 1.3 I: Exon/intron bounderies and their recognition during the early complex.

Exon/intron bounderies are defined by three *cis*-acting elements, the 5' splice site (5'ss, indicated as GU), the 3' splice site (3'ss, indicated as PPT and AG) and the branch point sequence (BPS, indicated as A). During spliceosome-mediated removal of introns and the ligation of exons each *cis*-acting element is recognised by several specific factors of the spliceosome in a dynamic temporal order. In the early steps of spliceosome catalysis, termed the early complex (E complex), the 5'ss is recognised by the U1 snRNP, the 3'ss by the 65Kd (p65) and 35Kd (p35) U2 auxilliary factors (U2AF) and the BPS by splicing factor 1 / mammalian branch point binding protein (SF1/mBPP). Exons were depicted as boxes and introns indicated by lines. *Trans*-acting factors were represented as different coloured circles depicting the U1 snRNP in blue, U2AF p65 and p35 in dark green and SF1/mBPP in light green. (Diagram modified from [221])

The occurrence of splicing at such defined borders is mediated by the sequential assembly of spliceosomal complexes and dynamic changes of their structure and composition [206, 221, 228]. The first step of spliceosome-mediated catalysis, the formation of the early complex (E complex), is characterized by ATP-independent binding of the exon/intron border defining *cis*-acting elements by specific *trans*-acting factors of the spliceosome. The 5'ss is recognised by the U1 snRNP. Thereby, an RNA-duplex between the 5' end of the U1 snRNA and the 11nt of the 5'ss is formed [148]. The recognition of the 3'ss is mediated by binding of both subunits of U2 auxiliary factor (U2AF) to the PPT and the canonical AG-dinucleotide,

respectively. Finally, the BPS is bound by the non-snRNP proteins splicing factor 1 / mammalian branch point binding protein (SF1/mBPP). After formation of the E complex ATP-dependent association of the U2 snRNP with the BPS defines the progression into the A complex (Fig. 1.3 II). Notably, base paring between the U2 snRNA and the BPS results in displacement of the branch point adenosine out of the formed RNA duplex. This bulged out adenosine represents the necessary nucleophile for the first transesterification [174]. Progression into the B complex is defined by joining of the A complex by the preassembled tri-snRNP complex consisting out of U4, U5 and U6 snRNP. After joining of the tri-snRNP the so far splice inactive B complex is transformed into a splice competent state (B* complex) by extensive conformational and compositional rearrangements [25, 221]. While the U1 and U4 snRNP were released during that transformation, the interaction of the U6 snRNP with the 5'ss and the U2 snRNA bound to the BPS sufficiently brings both splice sites in close proximity. The occurrence of the first catalytic step results in conversion of the B* complex to the C complex. Thereby, the conducted transesterification results in the transfer of the phosphorester bond at the 3' hydroxyl group of the ribose belonging to the last ribonucleotide of the exon to the 2' hydroxyl group of the bulged out branch point adenosine belonging to the BPS.
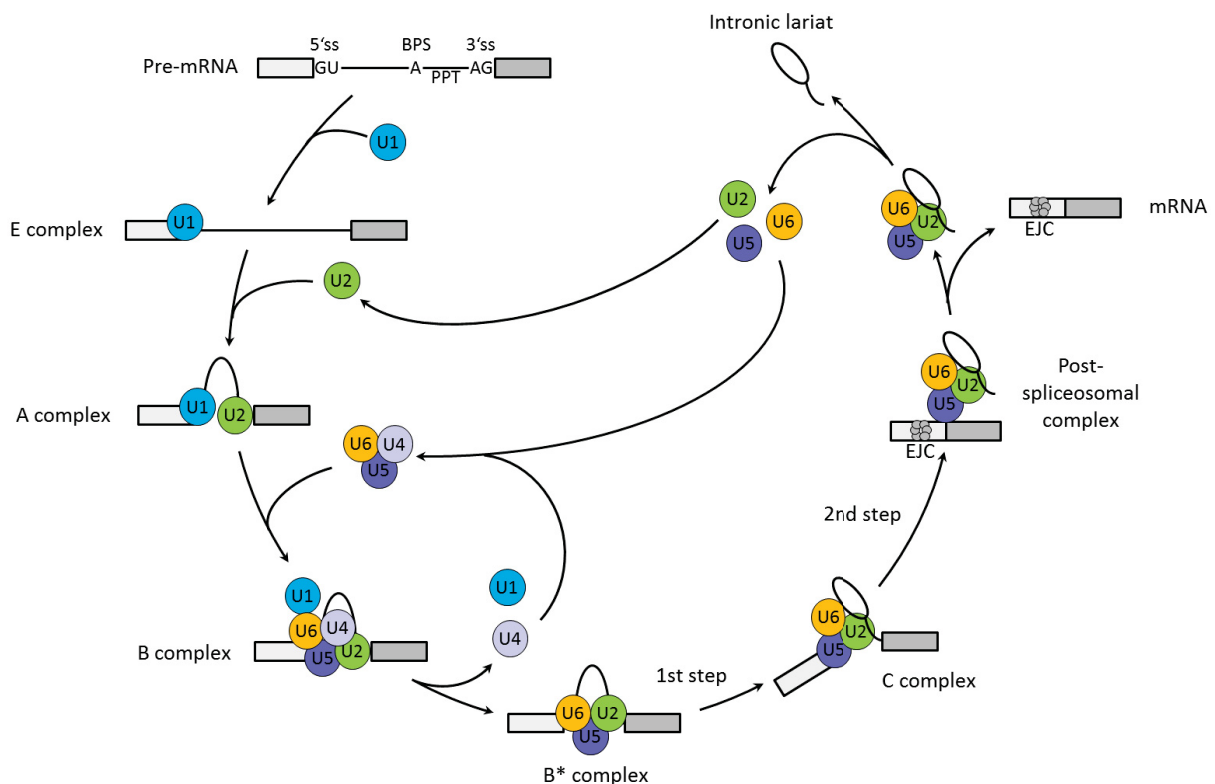
Introduction

Fig. 1.3 II: Assembly and disassembly cycle of the spliceosomal complexes during pre-mRNA splicing.

After RNA-polymerase II dependent transcription exons and introns of each pre-mRNA are defined by combinations of three *cis*-acting elements, the 5' splice site (5'ss, indicated as GU), the 3' splice site (3'ss, indicated as PPT and AG) and the branch point sequence (BPS, indicated as A). During spliceosome-mediated removal of introns and ligation of exons resulting in the intronless mRNA, the RNA bound in an dynamic temporal order of the mayor spliceosomal complexes the U snRNPs and a plethora of other associated proteins. During the early complex formation (E complex) the exon/intron borders are recognised especially by the U1 snRNP binding to the 5'ss. The A complex is formed by binding of the U2 snRNP to the BPS. The association with the tri-snRNP containing U4, U5 and U6 snRNP results in the pre-catalytic B complex. Conformational changes leading to the release of the U1 and U4 snRNP result in the catalytically active B* complex. The first transesterification (1st step) results in formation of the intronic lariat and defines the transition to the C complex. In the second transesterification (2nd step) the intronic lariat is cleaved and the flanking exons are ligated. During this step a protein complex is loaded upstream of the exon junction, termed exon junction complex (EJC), which resembles a seal of quality for the spliced mRNA. After disassembly of the formed post-spliceosomal complex and release of the spliced mRNA, the intronic lariat is degraded. All released spliceosomal components are eventually recycled to initiate a new round of spliceosome-mediated intron removal and exon ligation. Exons are depicted as boxes and introns indicated by lines. The 5'ss, 3'ss and BPS are indicated by the canonical GU, AG and A nucleotide respectively. Spliceosomal complexes are indicated as differently coloured circles, the U1 snRNP (blue), U2 snRNP (dark green), U4 snRNP (pale violett), U5 snRNP (dark violett) and the U6 snRNP (orange). Deposited proteins of the EJC are indicated by small grey circles. (The diagram is adapted from [221])

Notably, the so formed intronic lariat releases the 5' end of the upstream exon with another free 3' hydroxyl group representing a potent nucleophile for the next transesterification. Though, prior to the next step of catalysis the catalytic centre of the spliceosome needs to undergo a further extensive rearrangement. Thereby, the nucleophile at the end of the upstream exon is positioned in close proximity to the intron/exon junction of the downstream exon [115]. Noteworthy, at this step a protein complex is loaded at the upstream exon approximately 20-24nt apart from its end. This so called exon junction complex (EJC) marks productive intron removal conducted by the spliceosome and is of paramount importance in the latter life of the mRNA being exported from the nucleus, translated and assessed for its quality [19, 39, 122-124, 199, 246]. After completion of the structural rearrangements the phosphorester bond at intron/exon border of the downstream exon is transferred to the free 3' hydroxyl group of the upstream exon, thereby releasing the intronic lariat and joining both exons. Once the second transesterification is

completed the spliceosomal complexes dissociate from the spliced mRNA and participate in a new round of mRNA splicing.

1.4 Recognition of exon/intron borders

The first step of spliceosomal catalysis resulting in removal of the intronic sequence and joining of exons is the formation of the E complex. During E complex formation *cis*-acting elements marking the exon/intron borders were recognised by *trans*-acting factors belonging either directly to the spliceosomal U snRNPs or associated RNA-binding proteins. The exon/intron boundaries are marked by three types of *cis*-acting elements, the 5' splice site (5'ss), the 3'splice site (3'ss) and the branch point sequence (BPS) [221]. Interestingly, since being bound by invariant cellular *trans*-acting factors, the sequences characterising those *cis*-acting elements would be expected to be rather conserved. However, in contrast to lower eukaryotes as *Saccharomyces cerevisiae* showing high levels of conservation [233], the splice site sequences of higher eukaryotes appear to be more or less degenerate. Nevertheless, the end of each exon is defined by a 5'ss characterized by the consensus sequence CAG/GURAGUNN (R = purine, N = any nucleotide, / = exon/intron border) including the highly conserved GU-dinucleotide directly located at the 5' end of the following intron [34, 221] (Fig. 1.4 I). The start of each exon and thus the 3' end of each intron is defined by the 3'ss consisting out of the polypyrimidine tract (PPT) and the highly conserved AG-dinucleotide. While, the boundary of the intron to the exon is marked by the YAG/G (Y = pyrimidine, / = intron/exon border) consensus sequence including the highly conserved AG-dinucleotide, the PPT is located upstream within the intron possessing about 10 – 20nt of pyrimidines [34, 221] (Fig. 1.4 I). The BPS is usually located 15 to 50nt upstream of the 3'ss. In contrast to the 5'ss and 3'ss, the consensus sequence of BPS is less conserved and represented by the sequence YNYURAC containing the branch point adenosine (Y = pyrimidine, R = purine, N = any nucleotide) [34, 79, 221] (Fig. 1.4 I).
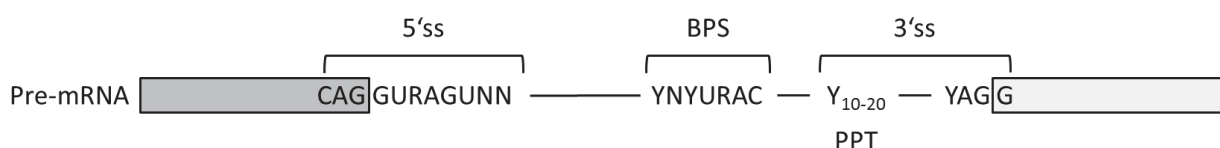
Fig. 1.4 I: Composition of *cis*-acting elements contributing to the recognition of exon/intron borders.

Exon/intron borders are defined by three *cis*-acting elements, the 5' splice site (5'ss), the 3' splice site (3'ss) and the branch point sequence (BPS). The 5'ss is characterized by the 11nt long sequence CAG/GURAGUNN (R = purine, N = any nucleotide, / = exon/intron border) which can form an RNA duplex with the 5' end of the U1 snRNA. The 3'ss consists out of the polypyrimidine tract (PPT) characterized by 10-20 pyrimidines and the consensus sequence YAG/G (Y = pyrimidine, / = intron/exon border) directly at the intron/exon border. The BPS is defined by the consensus sequence YNYURAC (Y = pyrimidine, N = any nucleotide) containing the canonical branch point adenosine. Exons were depicted as boxes and introns indicated by lines. (The diagram is modified from [221])

During E complex formation the 3' (3'ss) and 5' splice site (5'ss) are recognised by specific factors belonging to the spliceosome. A measure for the efficiency of that process is given by the intrinsic strength of a splice site. The intrinsic strength of a given 5'ss describes the efficiency of RNA duplex formation between the free 5' end of the U1 snRNA and the 11nt given by the 5'ss sequence [76, 103, 179, 248]. Current algorithms such as the HBond algorithm describe the quantity of hydrogen bonds, neighbour relations in between and interdependence of specific position within this RNA duplex [76, 87]. Additionally, however, the RNA duplex is also stabilised by the U1 snRNP associated protein U1C through extensive protein-RNA interactions [44, 63, 64, 90, 171]. Moreover, it could be shown that U1C could establish U1 snRNP binding also in absence of RNA duplex formation [63, 64]. Finally, RNA-duplex formation at the 5'ss was described to be essential for protection of the pre-mRNA against nuclear degradation [103]. In contrast to the 5'ss, the efficiency of the recognition of the 3'ss is dependent on the recognition of the three distinct sequence elements by their corresponding *trans*-acting factors, the PPT by U2AF p65, the YAG/G consensus sequence by U2AF p35 and the BPS by SF1/mBPP later replaced by the U2 snRNP during A complex formation. By that means an accurate estimation of the intrinsic strength of a 3'ss is more complex. However, a sufficient prediction about the intrinsic strength of a 3'ss could be made by determination of the pyrimidine content of the PPT and the complementarity of the BPS to the U2 snRNP [50]. In that line, it was shown that a PPT incorporating a continuous stretch of 11 pyrimidines sufficiently allows high efficient binding of U2AF p65. In contrast to the PPT and the BPS, the contribution of the YAG/G consensus sequence bound by U2AF p35 to the intrinsic strength of the 3'ss is appears only to be required, when the upstream PPT contains a low pyrimidine content [235]. Moreover, the contribution to the

recognition of the 3'ss was shown to be closely depend on the nucleotides present up- and downstream of the highly conserved AG-dinucleotide [49, 78, 133, 203].

However, besides the intrinsic strength the recognition of 5'ss and 3'ss could be shown to be enhanced or silenced by a network of additional *cis*-acting elements located either within the exon, called exonic splicing enhancer (ESE) or exonic splicing silencer (ESS), or within the intron, termed intronic splicing enhancer (ISE) or intronic splicing silencer (ISS), and their corresponding *trans*-acting factors frequently resembled by members of the serine/arginine-rich proteins (SR proteins) and heterogeneous nuclear ribonucleoprotein (hnRNP) families [15, 17, 141, 224] (Fig. 1.4 II).
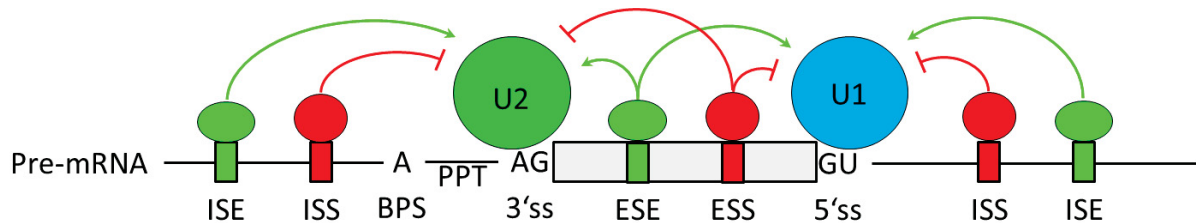
Fig. 1.4 II: Network of *cis*-acting elements and corresponding *trans*-acting factors regulating exon recognition.

Exon/intron borders are defined by three *cis*-acting elements, the 5' splice site (5'ss, indicated as GU), the 3' splice site (3'ss, indicated as PPT and AG) and the branch point sequence (BPS, indicated as A). During the early phases of the spliceosome assembly the 5'ss is bound by the U1 snRNP (blue circle, designated U1) and 3'ss by the U2 snRNP (dark green circle, designated U2) initiating spliceosome-mediated removal of introns and ligation of exons. However frequently, sequences from all three *cis*-acting elements defining the exon/intron borders deviate form their common consensus sequence lowering their capacity to be recognised by factors of the spliceosome, e.g. the U1 snRNP. In order to nevertheless recognise those degenerate sequences over a plethora of per chance similar sequences in the vincinity, an extensive network of *cis*-acting elements is employed. Those *cis*-acting elements are located either within exons or introns and bind *trans*-acting factors, commonly serine/arginine-rich proteins (SR proteins) or heterogeneous nuclear ribonucleoproteins (hnRNPs). Binding of the corresponding *trans*-acting factors either enhances (ESE/ISS) or silences (ESS/ISS) the recognition of splice sites and thus contributes to the defintion of the correct exon/intron borders. The exon is depicted as box and introns are indicated by lines. The 5'ss, 3'ss and BPS are indicated by the canonical GU, AG and A nucleotide respectively. Exonic and intronic splicing enhancer (ESE/ISE) are depicted as green boxes, their corresponding *trans*-acting factors as green circles. Exonic and intronic splicing silencer (ESS/ISS) are depicted as red boxes, their corresponding *trans*-acting factors as red circles. Green arrows with spiked head indicate an enhancing activity, whereas red arrows with flat head indicate a silencing activity. (The diagram is modified from [224])

1.5   Finding the proper matches – a shift from exon to intron definition

Approximately 95% of all genes belonging to higher order vertebrates are spliced. However, the recognition of intron/exon borders mediated within the formation of the E complex encounters some major obstacles. One of them is given by the need to detect the 5'ss, 3'ss and BPS defining the borders of relatively small exons (median size 124nt) within a sea of large introns (median size 1458nt) [187]. A solution to that problem was proposed by the exon definition model (Fig. 1.5 I).
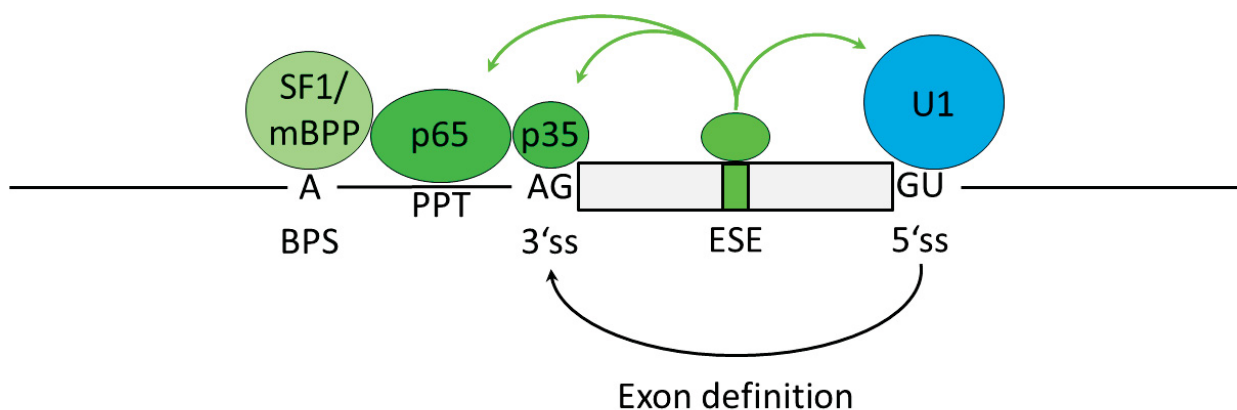
Fig. 1.5 I: Exon definition mediated by upstream, cross-exon interactions of the U1 snRNP

Exon/intron borders are defined by three *cis*-acting elements, the 5'ss (indicated as GU), the 3'ss (indicated as PPT and AG) and BPS (indicated as A). During spliceosome-mediated removal of introns and the ligation of exons each *cis*-acting element is recognised by several specific factors of the spliceosome in a dynamic temporal order. In the early steps of spliceosome catalysis, termed the early complex (E complex), 5'ss is recognised by the U1 snRNP, the 3'ss by the 65Kd (p65) and 35Kd (p35) U2 auxilliary factors (U2AF) and the BPS by splicing factor 1/mammalian branch point binding protein (SF1/mBPP). Recognition of the 5'ss by the U1 snRNP was found to increase the recognition of a directly upstream located 3'ss by U2AF and U2 subsequently [93]. Moreover, ESEs within the exon were described to stabilise and further promote the connection between both splice sites flanking the exon [21, 93]. Exons were depicted as boxes and introns indicated by lines. The *cis*-acting elements at the exon/intron border were indicated by their canonical nucleotides, the 5'ss (GU), the 3'ss (AG) and the BPS (A). The exonic splicing enhancer is depicted by a light green box. *Trans*-acting factors are represented as different coloured circles depicting the U1 snRNP in blue, U2AF p65 and p35 in dark green, SF1/mBPP in pale green and ESE binding factor in light green. Arrows with spiked heads indicate an enhancing activity.

Based on the finding that the addition of a 5'ss at the end of a 2 exon splice substrate appeared to increase the recognition of the upstream 3'ss, the definition of the downstream exon was deduced to increased splicing of the upstream intron [12, 178]. Further evidence was provided by studies showing that mutations inactivating either the 3'ss or the 5'ss resulted rather in skipping of the entire exon than sole interference with splicing at the inactivated splice site. Finally it was shown that directly after binding of the U1 snRNP to the 5'ss a molecular bridge was established reaching across the exon, resulting in enhanced assembly of U2AF and subsequently U2 snRNP at the 3'ss [93]. Noteworthy, ESEs located within an exon appear to stabilise and further promote the connection between both splice sites flanking the exon [21, 93] (Fig. 1.5 I).However, after accurate definition of each exon the intronic sequences need to be defined, whose cleavage results in joining of the flanking exons [18, 95, 192]. This switch from exon to intron definition was shown to occur at the border from A to B complex. In vitro purification experiments showed that the tri-snRNP consisting out U4, U5 and U6 snRNP was first bound to the exon definition complex spanned between two splice sites. Thereafter, intron definition was mediated by engagement of the tri-snRNP with the U1 snRNP bound to another RNA substrate containing a 5'ss [188] (Fig. 1.5 II).
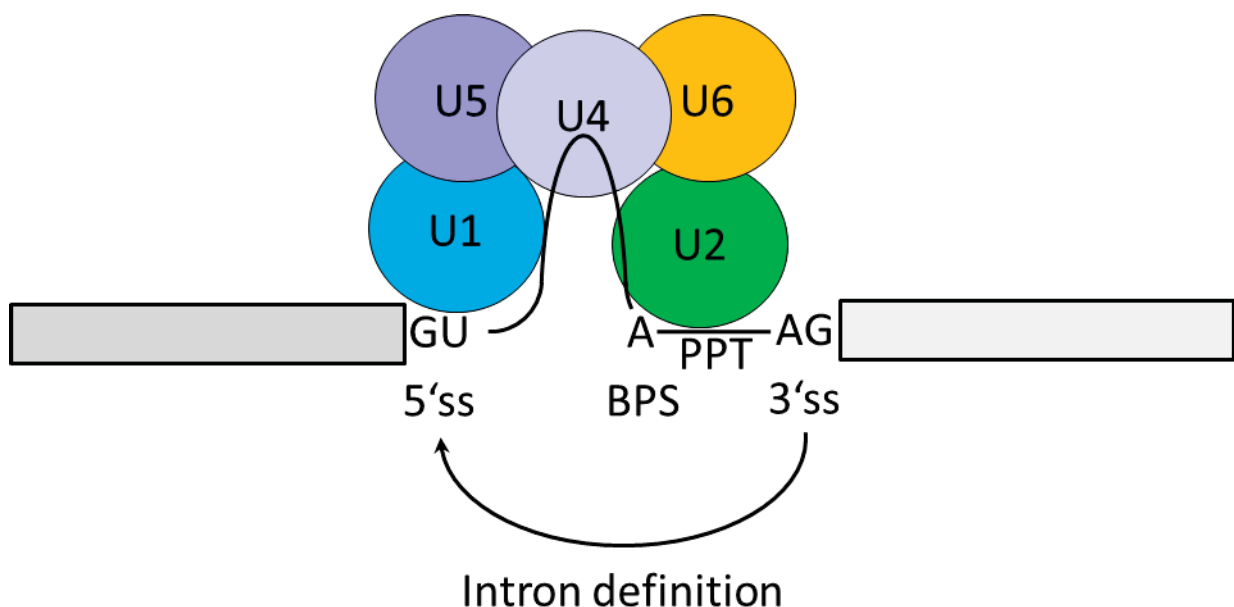
Fig. 1.5 II: Intron definition mediated by cross-intron interaction of the tri-snRNP

During spliceosome-mediated removal of introns and ligation of exons each *cis*-acting element on the pre-mRNA is recognised by several specific factors of the spliceosome in an dynamic temporal order. With formation of the spliceosomal A complex the 5'ss (indicated as GU) is still bound by the U1 snRNP and the BPS (inidcated as A) by the U2 snRNP. The association of the A complex with the tri-snRNP containing the U4, U5 and U6 snRNP initiates the formation of the B complex. After initial binding of the tri-snRNP to the exon definition complex of the downstream exon, in a process termed intron defintion the association of the tri-snRNP with the U1 snRNP bound to the 5'ss of the upstream exon defines the intron to be removed by the spliceosome. Exons are depicted as boxes and introns indicated by lines. The *cis*-acting elements at the exon/intron border are indicated by their canonical nucleotides, the 5'ss (indicated as GU), the 3'ss (indicated as PPT and AG) and the BPS (indicated as A). The binding U snRNPs are depicted as coloured circles, the U1 (blue), U2 (dark green), U4 (pale purple), U5 (dark purple) and U6 (orange). (Diagram modified from [95])

1.6 Alternative splicing

As said about 95% of all genes belonging to higher vertebrates are spliced. However, besides generating only one mRNA by joining of the direct adjacent exons, known as constitutive splicing, also multiple mRNAs are generated from one pre-mRNA by selection of different subsets of exons, known as alternative splicing. Defying the original 'one gene one protein hypothesis' alternative splicing results in expansion of the human proteomic diversity above genomic restriction to about 20,000 genes by about five fold [137, 156, 163]. A lot of knowledge about the mechanisms underlying alternative splicing had been gained by mutations resulting in aberrant splicing [30, 34, 60, 160, 223]. Hitherto, three major regulatory determinants of alternative splicing have been identified, the intrinsic strength of the splice sites, the surrounding network of splicing enhancers and splicing silencers and the local RNA structure regulating the overall accessibility to the involved *cis*-acting elements [10, 194, 224]. However, it had been shown that most human splice sites representing only a poor match to the consensus sequence were chosen over cryptic splice sites often found to be the even better matches. Thus apparently, the choice of one splice site over the other only by means of their intrinsic strength is clearly insufficient [40]. Moreover, not all SREs, which contribute to the recognition of splice sites during constitutive splicing, also contribute to the discrimination of alternative splice sites [185, 245]. Nevertheless, the differential composition of some SREs could be shown to allow the discrimination between alternative splice sites by binding of the corresponding *trans*-acting factors [15, 85, 204].

Additionally, tissue specific differences including the expression of those *trans*-acting factors, as for example Nova and Fox1/2, had been observed to result in tissue specific expression of alternative transcripts [215, 238, 242].

1.7   The SR protein family

Hitherto the members of the SR protein family were commonly described to contribute to the regulation of constitutive and alternative pre-mRNA splicing. Though, after contributing to splicing regulation SR proteins often remain associated with spliced mRNAs and thus contribute to later steps of eukaryotic gene expression as nuclear mRNA export, quality control by regulation of nonsense-mediated decay (NMD) and translation [82, 129, 195]. Generally, the expression of SR proteins was shown to occur within all metazoan species and some lower eukaryotes with the exception of budding yeast *Saccharomyces cerevisiae* [83, 132, 240]. Since prone to extensive phosphorylation SR proteins could be identified by the monoclonal antibody mAB104 detecting phosphorylated SR proteins [181]. Further, their contribution to the stimulatory effect of S100 extracts on pre-mRNA splicing [240] and their overall conversation among vertebrates and invertebrates defines a member of the SR protein family. Fitting to all three of those requirements so far twelve human SR proteins have been identified and named as SR splicing factors (SRSF), SRSF1 (SF2/ASF), SRSF2 (SC35), SRSF3 (SRp20), SRSF4 (SRp75), SRSF5 (SRp40), SRSF6 (SRp55), SRSF7 (9G8), SRSF8 (SRp46), SRSF9 (SRp30c), SRSF10 (SRp38), SRSF11 (SRp54) and SRSF12 (SRrp35) [138]. Additionally, to those twelve human SR proteins genome-wide surveys identified several proteins, termed SR-related or SR-like proteins, that as a matter of fact contain an RS-domain, but fail either to rescue pre-mRNA splicing of a splice-deficient HeLa S100 extract or to be stained by mAB104 [20]. SR protein structure is characterized by one or two N-terminal RNA recognition motives (RRMs) and a C-terminal arginine/serine-rich domain (RS domain). While the RRM mediates sequence specific RNA binding, the RS domain enables contacts to components of the splicing machinery employed in constitutive and alternative splicing [82, 114, 234]. Moreover, besides facilitating the regulation of pre-mRNA splicing, the RS domain contains a nuclear localization signal (NLS) targeting nuclear import using a SR protein specific receptor transportin-SR [31, 110, 121]. A long time SR proteins have been commonly described as enhancer of pre-mRNA splicing by recruiting the U1 snRNP to the 5'ss via their

RS domain and thus increasing exon definition [77, 128]. However, since recent literature provided evidence that SR proteins bound to intronic SREs can interfere with exon recognition [26, 96, 105], a more sophisticated model of position dependency of splicing regulation by SREs was proposed [67].

1.8 The hnRNP family

The term heterogeneous nuclear RNA (hnRNA) summarises all transcribed nuclear, protein coding RNAs despite of their state of processing. These hnRNAs were commonly found to be associated with various proteins, termed heterogeneous nuclear ribonucleoproteins (hnRNPs), forming protein-RNA complexes [61]. Isolation of the hnRNPs and subsequent molecular characterization let to the differentiation of 20 hnRNP families, termed hnRNP A to U [86, 168]. However, despite their common origin hnRNPs exhibit a high diversity regarding their structural composition and functional properties. One widespread structural feature of hnRNPs lies within the manner of RNA binding using at least one RNA recognition motif (RRM). Notably, the typical RRM fold is defined by the typical βαββαβ sequence of secondary structures and the two RNP consensus sequences (RNP-1 and RNP-2) [14, 94, 131, 162]. Characteristically the RRM fold displays two α-helices packed against four strands of an antiparallel β-sheet providing the main protein-RNA interaction surface. While the conserved RNP motives within the central β-strands mainly expose aromatic side chains allowing unspecific RNA binding by hydrophobic stacking interactions with the RNA nucleotides, the composition of the exposed residues by the β-sheet outside the RNP motifs generates binding specificity [140]. Noteworthy, besides the canonical RRM also other functional equivalent RNA binding domains (RBDs), such as the quasi-RRM (qRRM) and the K-homology motif (KH), have been identified to mediate RNA binding of hnRNPs [9, 57, 58, 202]. Further, apart from one RBD mediating RNA binding, hnRNPs often contain auxiliary domains rich of certain amino acids, as glycine, proline or acidic amino acids in general. Furthermore, adding to the structural diversity most pre-mRNAs encoding hnRNPs are alternatively spliced and the translated proteins extensively posttranslational modified including phosphorylation [150, 217, 236], sumoylation [126, 218] and arginine methylation [45, 92, 155, 164]. Although the structural diverse family hnRNPs also affects a plethora of other cellular processes, the majority contributes to the regulation of alternative splicing

[86]. While binding of hnRNPs to exonic SREs had been well established to interfere with pre-mRNA splicing [42, 100, 182], also enhancement had been shown from intronic located SREs [186, 222]. Thus, similar to SR-proteins, though functionally inverted regulation of pre-mRNA splicing by hnRNP proteins was shown to be position-dependent [67].

1.9   HnRNP A1

Hitherto, the hnRNP A/B family of proteins represent the currently best known negative regulators of splicing regulatory proteins shown to interfere with mRNA splicing by binding to a wide variety of either exonic [51, 59, 80, 107, 144, 145, 180] or intronic [37, 108, 211] *cis*-acting elements. All proteins of the hnRNP A/B family contain a common structure consisting of two N-terminal located RNA-binding domains (RBDs) and a glycine-rich C-terminus, termed 2xRBD-Gly structure [27, 62, 142]. While the binding specificity of most prominent member of the family hnRNP A1 is determined by their two RRM domains [1, 28, 29, 55, 91, 177, 191, 237], two regions within the glycine-rich C-terminus were identified, the RGG-box presumably contributing to RNA-binding and mediating binding cooperativity [36, 48, 118, 145, 152] and the M9 sequence necessary for nuclear shuttling [201, 225]. Further, the glycine-rich C-terminus was shown to mediate homo- and heteromeric protein-protein interactions to hnRNP A1 and other members of the hnRNP A/B protein family [35]. Hitherto, several mechanisms of hnRNP A1-mediated regulation of alternative splicing have been published. In this line, RNA-binding of hnRNP A1 and subsequent steric hindrance was described to displace positive splicing regulatory factors from overlapping sites [56, 84]. Further, self-association of hnRNP A/B proteins binding at exon flanking SREs was shown to result in looping-out of alternative exons and thus in intron-definition [16, 154]. Finally, cooperativity-induced spreading of hnRNP A1 binding to low affinity binding sites was described to displace positive splicing regulatory factors from their binding sites [159, 247]. Moreover, characterization of the latter mechanism, termed cooperative-spreading, showed that spreading occurred bidirectional in 3'-5' and 5'-3' direction, though in 5'-3' direction to a weaker extent.

1.10 HnRNP D

The heterogeneous nuclear ribonucleoprotein D (hnRNP D), also known as AU-rich element RNA-binding protein 1 (AUF1), was initially identified as cytosolic factor accelerating the decay of *c-myc* mRNA *in vitro* [23, 24]. Since then, hnRNP D was shown to regulate the decay kinetics of a plethora of transcripts, which encode cell cycle regulators, early response genes, inflammatory mediators and cytokines, G-protein coupled receptors and oncoproteins [81, 251]. Purification and identification of the factor revealed four similar proteins originating out of one alternative spliced pre-mRNA (Fig. 1.10), which assembled into complexes efficiently bind to a variety of RNA substrates containing adenine/uridine-rich elements (AREs) [52, 54, 220].
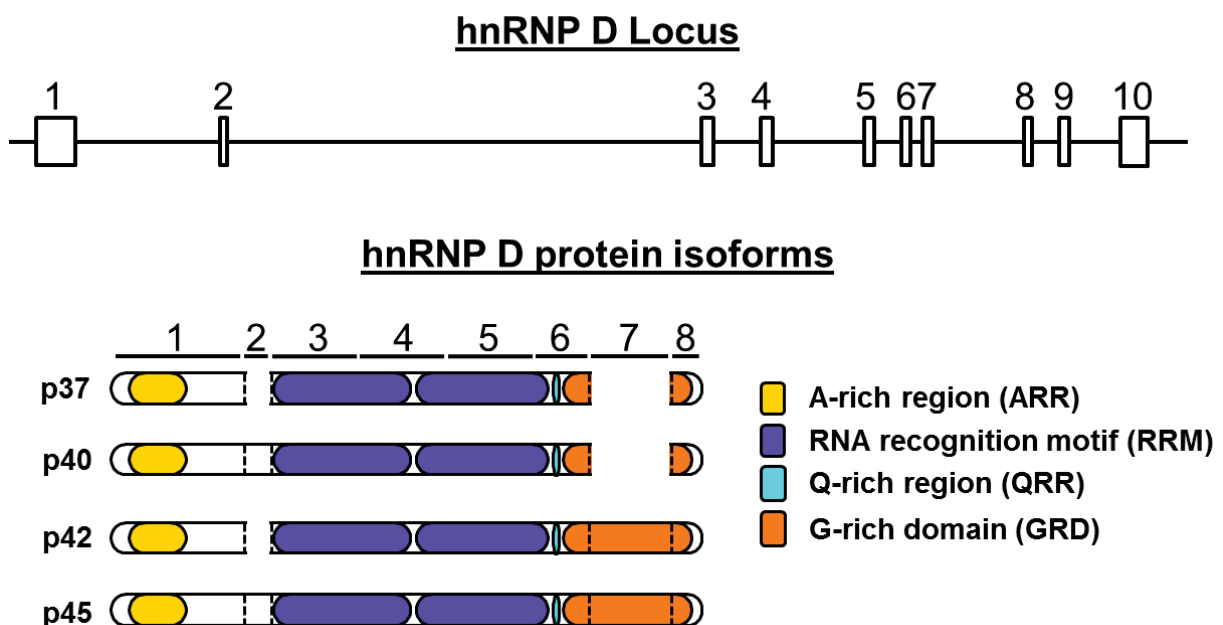
Fig. 1.10: Genome organisation and protein isoforms of hnRNP D.

The hnRNP D locus (upper panel) comprises 10 exons, of which the first eight encode the four hnRNP D isoforms (bottom panel). Splicing of the alternative exons 2 and 7 generates four hnRNP D isoforms, named according to their molecular weight p37, p40, p42 and p45. All hnRNP D isoforms contain two non-identical RNA-recognition motifs (RRMs), an N-terminal alanine-rich region (ARR), C-terminal glutamine-rich region (QRR) and a C-terminal glycine-rich domain (GRD).

Differing by their cellular localisation hnRNP D isoforms p42 and p45 commonly show a nuclear localization, while isoforms p37 and p40 are located within the cytoplasm and nucleus [6, 97, 230, 243]. Mapping of nuclear localization sequence (NLS) of hnRNP D so far indicated that a 19 amino acid long domain within the C-terminus common to all isoforms

might mediate nuclear shuttling by binding to transportin 1 [208]. However another model stated that the inclusion of alternative exon 7 with hnRNP D isoforms p42 and p45 inhibits NLS-mediated nuclear import and thus a co-transport with alternative nuclear cargos might be required [184]. Moreover, complex formation with nuclear scaffold attachment factor-β or cytosolic 14-3-3σ [6, 89] might serve as reservoir to enrich specific hnRNP D isoforms in either in the cytoplasm or the nucleus. Similar to the members of the hnRNP A/B family all hnRNP D isoforms contain the common 2xRBD-Gly structure, containing two RNA-binding domains (RBDs) and C-terminal glycine-rich domain (GRD) [102]. The two RBDs are provided by two in tandem arranged, non-identical RRM domains [220, 243]. Both RRM domains are required, but not sufficient for high-affinity RNA binding [53]. Besides the RRMs the N-terminal alanine-rich region (ARR) and C-terminal glutamine-rich region (QRR) have been shown to be necessary for high affinity binding [53]. Moreover, inclusion of alternative exon 2 resulting in the insertion of 19 amino acids directly adjacent to the N-terminal to RRM1 was shown to lower RNA-binding affinity of hnRNP D isoforms p40 and p42 by 3 to 5 fold [220, 249]. All hnRNP D isoforms form stable dimers in solution presumably mediated by the alanine-region (ARR) [53]. After initial binding of an hnRNP D dimer to an RNA substrate, a tetrameric complex is formed on extended RNA substrates by association with another hnRNP D dimer [231, 232, 249]. RNA-binding kinetics of the second dimer showed that inclusion of 49 amino acids encoded by alternative exon 7 results in a higher stability of the formed tetrameric complex on extended RNA substrates by hnRNP D isoforms p42 and p45 [231, 249]. All hnRNP D protein isoforms bind to ARE containing RNA substrates with mid to low nanomolar affinity [53, 249]. Sequence specificity of binding seems to be relaxed, since hnRNP D was published to efficiently bind to canonical ARE sequence motives, the r(AUUUA) pentamers [219, 243], but also to other uridine-rich sequences with a similar affinity [231, 232]. Moreover, recently hnRNP D was isolated from HeLa cell nuclear extracts binding to d(TTAGGG) sequence repeats, the human telomeric repeat, and even more tight to its RNA counterpart r(UUAGGG) [98]. Further, both RRMs could independently be shown to mediate sequence specific binding to the telomeric repeat sequence [65, 102, 109, 153]. At the chromosome ends the single stranded telomere repeats were shown to fold into a G-quartet structure [229] resulting in interference with the replication of telomeres by telomerase [73, 241]. Since hnRNP D binding could be shown to destabilise G-quartet structures and an

interaction of hnRNP D with the telomerase was described [71], an additional role of hnRNP D in telomere maintenance was proposed.

1.11 Splicing regulatory elements identified within the HIV-1 genomic RNA

By alternative splicing the viral pre-mRNA HIV-1 generates more than 40 different mRNA isoforms needed to express at least 18 viral proteins. Thereby, alternative splicing of HIV-1 pre-mRNA is mediated by the combinatorial recognition of specific subsets of the viral splice sites including eight 3'ss incorporating the canonical AG and five 5'ss characterized by the canonical GU [173]. All viral 3'ss and 5'ss were recently assessed for their intrinsic strength (Fig. 1.11). While the 5'ss D1 and D4 were shown to be relatively strong, in line with their low complementary to the 5' end of the U1 snRNA the 5'ss D2, D2b and D3 displayed to be rather weak [75, 158, 227]. Notably, apart from the other viral splice donors the actual recognition of 5'ss D2b was described to be very low due to the inhibition by an intronic G-run G_{I2-1} (Fig. 1.11) [173, 227]. Experiments determining the intrinsic strength of the viral 3'ss discovered that besides 3'ss A2 and A3 featuring an intermediate intrinsic strength all other viral 3'ss, the 3'ss A1, A4c, A4b, A4d, A5 and A7, are intrinsically weak [104]. However, apparently deviating from the given limit of their intrinsic strength the actual recognition of 3'ss A1, A4cab, A5 and A7 could be shown to be increased, while that of 3'ss A2 and A3 displayed to be decreased, when being measured in presence of their respective exonic sequences [104]. That apparent gap in regulation was shown to be filled by a network of SREs located within the respective exonic and intronic sequences, which predominantly determines the actual recognition of each viral 3'ss (Fig. 1.11). Therefore, some ORFs, such as *vpu*/*env* and *nef*, are expressed quite efficiently, while other ORFs, as such encoding Vpr, Vif and Tat proteins depict low expression levels [173]. However, due to the high complexity the employed SRE network providing differential expression of each viral ORF and enabling viable HIV-1 replication has been shown to tolerate only low margins of error. In that line, recently a mutant of viral SRE was found to induce excessive splicing shown to be detrimental to virus particle production [13, 59, 134, 136]. Thus, efficient HIV-1 replication depends on one hand on alternative splicing, in order to differentially express its viral proteins. However, on the other hand HIV-1 pre-mRNA splicing needs to be inefficient, in

order to retain sufficient amounts of intron-containing viral RNA enabling the expression of late viral genes and keep sufficient amounts of the viral genomic RNA.
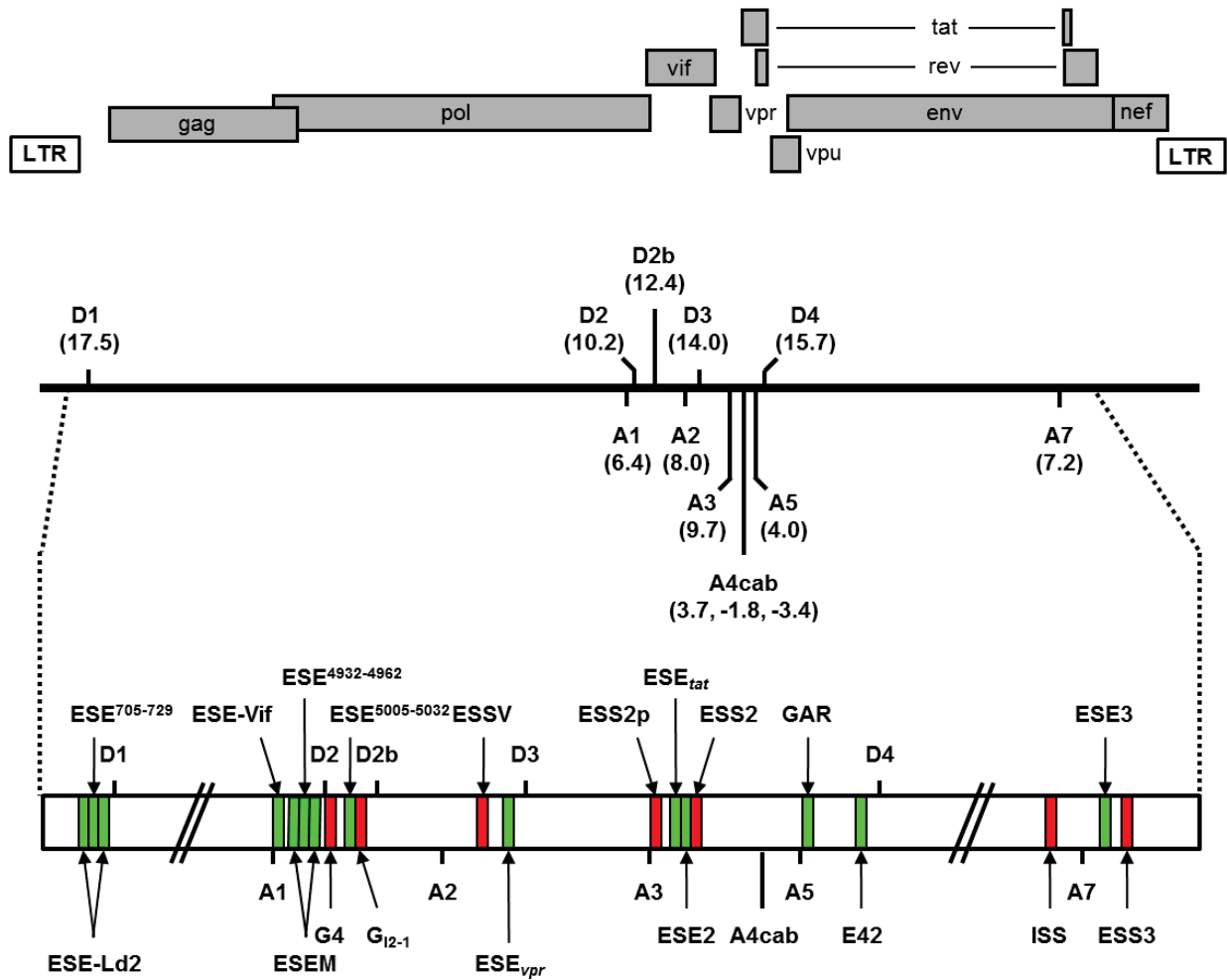
Fig. 1.11: Schematic diagram of known splicing regulatory elements (SREs) within the HIV-1 genome. (Top panel) Organisation of the proviral genome. Open reading frames (ORFs) were indicated as grey boxes. The long terminal repeats (LTR) at both ends of the HIV-1 genome were depicted as white boxes. (Center panel) HIV-1 pre-mRNA containing the viral 5' splice sites (5'ss) and 3' splice sites (3'ss). The intrinsic strength of the viral 5'ss is was determined by a web-based algorithm (HBond Score, http://www.uni-duesseldorf.de/RNA). The intrinsic strength of the viral 3'ss was determined by a web-based algorithm (MaxEntScore, http://genes.edu/burgelab/maxentscan_scoreseq_acc.html). (Bottom panel) Relative positions of SREs within the HIV-1 pre-mRNA: ESE-Ld2 [7], ESE[705-729] [69], ESEVif [72]; ESEM [104], ESE[4932-4962] [69], G4 [72], ESE[5005-5032] [69], G[i2-1] [227], ESSV [13, 59, 134], ESE[vpr] [68, 69], ESS2p [100], ESE[tat] [66, 69], ESE2 [84, 239], ESS2 [3, 33, 197], GAR [8, 32, 103], E42 [8], ISS [211], ESE3 [205] and ESS3 [4, 198, 205]. (Modification of existing diagram [66])

1.12 Aim of this work

In order allow efficient viral replication, HIV-1 essentially relies upon the maintenance of a fragile balance between spliced mRNA isoforms and the unspliced viral mRNA. While, alternative splicing of the viral pre-mRNA generates more than 40 different mRNA isoforms resulting in the expression of at least 18 viral proteins, a certain amount of the pre-mRNA has to be left unspliced. This unspliced pre-mRNA represents the genome of viral progeny and also the mRNA encoding the viral structural proteins Gag and Gag-Pol. Any shift of the balance to excessive splicing combined with a decrease in unspliced 9Kb RNA, and thus decrease in Gag expression had been described to inhibit HIV-1 replication [135]. Further, the observed low levels of Gag protein led to a processing defect [136], additionally contributing to the loss in viral replication. This thin line of splicing regulation has been shown for the recognition of HIV-1 exon 3 [13, 59, 134]. Primary the recognition of HIV-1 exon 3, but at the same time retention of its downstream intron by exclusive splicing at 3'ss mediates expression of the *vpr* ORF. Moreover, besides mediating *vpr* expression, the recognition of HIV-1 exon 3 and splicing of both adjacent introns results in incorporation of the exon within the leader of several viral transcripts. Both outcomes mediated by recognition of HIV-1 exon 3 rely on three determinants, the complementarity of the 5'ss D3 to the U1 snRNA, the exonic splicing silencer ESSV and the exonic splicing enhancer ESE$_{vpr}$. Recognition of exon 3 was mediated by the complementarity of 5'ss D3 to the 5' end of the U1 snRNA and the presence of ESE$_{vpr}$ synergistically resulting in recruitment of U1 snRNP to 5'ss D3, thus enabling exon definition [68]. However, recognition of HIV-1 exon 3 needs to be restricted by the dominant splicing repressive regulatory element ESSV located within this exon. Binding of hnRNP A1 to ESSV had been shown to lead to displacement U2AF p65 [59], thereby overriding exon definition emanating from the 5'ss. Moreover, disruption of the ESSV by site-directed mutagenesis resulting in excessive splicing demonstrated the importance of ESSV-mediated interference with exon 3 recognition for HIV-1 replication [13, 134]. However, exclusive contribution hnRNP A1 to ESSV-mediated interference was drawn into question, since it was shown that depletion of cellular hnRNP A1 by siRNA-mediated knock down failed to induce excessive splicing of HIV-1 exon 3 [99]. This finding gave rise to the notion, that another splicing regulatory protein along with hnRNP A1 might bind to the ESSV and thereby close the observed gap in interference.

Interestingly, a first hint to answer this question was given by a recent study proposing the high AU-rich content of exon 3 to be an mRNA instability element [117]. Following this proposal putative binding sites of a known mRNA stability regulator, hnRNP D, were found to overlap with the already published binding sites of hnRNP A1. However, while the proposed influence on mRNA stability of the putative hnRNP D binding sites could be ruled out [134], the idea manifested that hnRNP D might represent the other splicing regulatory protein beside hnRNP A1 contributing to ESSV-mediated interference with exon recognition. Thus, the primary goal of this thesis was to investigate the possibility of hnRNP D binding to the ESSV and the possible outcome of that event on the recognition of HIV-1 exon 3. In particular, along with hnRNP A1, hnRNP D binding to the ESSV might contribute to the interference with the recognition of HIV-1 exon 3, thereby holding excessive splicing at bay and thus represent essential host factors required for HIV-1 replication.

2 Material and Methods

2.1 Material

2.1.1 Eukaryotic cell lines

HeLa cells (ATCC-No. CCL-2):

Established out of a tissue sample of the cervix carcinoma patient Henrietta Lacks the so called HeLa cells represent the first known cell line. Immortalized by infection of human papillomavirus type 2 (HPV-2) this cell line was used for transient transfection experiments.

HEK-293T cells (ATCC-No. CRL-11268):

The parental HEK-293 cell line was derived by transformation out of human embryonic kidney tissue by incorporation of human adenovirus 5 DNA (ATCC-No. CRL-1573). This cell line was further manipulated by the insertion of the SV40 virus T-antigen, resulting in the HEK-293T cell line. The production of the SV40 large T-antigen allows the episomal replication of plasmids containing an SV40 origin. Another hallmark of HEK-293T cells is their high expression capacity. Therefore this cell line was used for transient transfection experiments such as overexpression experiments or viral particle production.

CB3 / CB3 A1 cells

The CB3 cells were derived from spleen of a BALB/c mouse infected at birth with F-MuLV [196]. In contrast to other F-MuLV-induced erythroleukemic cell clones derived from retroviral insertion events and further mutagenesis, CB3 cells are homozygous containing a retroviral insertion at the Fli-2 site immediately downstream of the murine hnRNP A1 gene locus. The insertion could be shown to diminish hnRNP A1 expression [11]. The CB3 A1 cell line was created by retroviral transduction of a murine hnRNP A1 expression cassette, restoring hnRNP A1 expression [119].

2.1.2 Bacterial strains

***E. coli* K12 DH5α F'IQ:**   F-φlacZΔM15   Δ(*lac*ZYA-*arg*F)   U169   *rec*A1   *end*A1   *hsd*R17

(rk-, mk+) *pho*A *sup*E44 λ- *thi*-1 *gyr*A96 *rel*A1/F´ *pro*AB+

*lac*IqZΔM15 zzf::Tn5 [KmR]

2.1.3 DNA-Oligonucleotides

All listed oligonucleotides are depicted in 5'-3' orientation.

2.1.3.1 Oligonucleotides for RT-PCR

#640: 3'-primer for RT-PCR intron 5 downstream SD5

CAATACTACTTCTTGTGGGTTGG

#1224: 3'-primer for RT-PCR exon junction 4/5 of hgh

TCTTCCAGCCTCCCATCAGCGTTTGG

#1225: 5'-primer for exon junction 3/4 of hgh

CAACAGAAATCCAACCTAGAGCTGCT

#1544: 5'-primer for RT-PCR exon1 upstream SD1

CTTGAAAGCGAAAGTAAAGC

#2588: 3´-primer for RT-PCR downstream exon 4 within the bacterial chloramphenicol acetyltransferase (CAT) ORF

CTTTACGATGCCATTGGGA

#3392: 3'-primer for RT-PCR downstream SA7

CGTCCCAGATAAGTGCTAAGG

#3632: 3'-primer for RT-PCR downstream SA3

TGGATGCTTCCAGGGCTC

#3387: 5'-primer HIV-1 exon 7

TTGCTCAATGCCACAGCCAT

#3388: 3'-primer for HIV-1 exon 7

TTTGACCACTTGCCACCCAT

2.1.3.2 Oligonucleotides for molecular cloning

#3442: 3´-primer for cloning hnRNP D for ala aa8-41 - fragment 1 over NcoI (90nt, 72.2°C TM)

CGCCACCATGGCTCCCTCCTGCTCGCCCGCCGAGCCGCCTACCGCCGCCGTTGCCGCTGCCG
CCGCCCCGTCCCCGCCGAACTGCTCCTC

#3443: 5´-primer for cloning hnRNP D for ala aa8-41 - fragment 2 over NcoI (60nt, 76.5°C TM)

GGGAGCCATGGTGGCGGCGACACAGGGGGCAGCGGCGGCGGCGGGAAGCGGAGCCGGGAC

#3444: 3´-primer for cloning hnRNP D for ala aa8-41 - fragment 2 (21nt, 66.5°C TM)

GCCAAAACCCCTTGATCGCCC

#3641: 5´-primer for mutagenesis ESSV 15-16/28-29/34-35 over AlwNI (77nt, 72.8°C TM)

TTTCAGAATCTGCTATAAGAAATACCATATTCTGACGTATAGTTCTTCCTCTGTGTGAATAT
CAAGCAGGACATAAC

#4290: 5´-primer for cloning hnRNP D Q/G-rich regions (31nt, 62.3°C TM)

CCCCCCGGATCCCAGCAACAGCAACAGTGGG

#4291: 3´-primer for cloning hnRNP D Q/G-rich regions (34nt, 62.3°C TM)

CCCCCCCTCGAGTTACCCATAACCACTCTGCTGG

#4293: 5´-primer for mutagenesis ESSV>Neutral over AlwNI (84nt, 63.1°C TM)

CCCCCCCAGAATCTGCTATAAGAAATACCCCAAACAACCAAACAACCAAACAACCAAGTGAA
TATCAAGCAGGACATAACAAGG

#4294: 5´-primer for mutagenesis ESSV>ARE over AlwNI (84nt, 63.1°C TM)

CCCCCCCAGAATCTGCTATAAGAAATACCTATTTATTTATTTATTTATTTATTTATTGTGAA
TATCAAGCAGGACATAACAAGG

#4295: 5´-primer for mutagenesis ESSV>TR over AlwNI (84nt, 63.1°C TM)

CCCCCCCAGAATCTGCTATAAGAAATACCTTAGGGTTAGGGTTAGGGTTAGGGTTAGGTGAA
TATCAAGCAGGACATAACAAGG

2.1.3.3 Oligonucleotides for in vitro transcription

#4324: 5´-primer RNA affinity chromatography T7 (19nt)

TAATACGACTCACTATAGG

#4325: 3´-primer RNA affinity chromatography ESSV NL4-3 (70nt)

ACACCTAGGACTAACTATACGTCCTAATATGGACATGGGTGATCCTCATGTCCTATAGTGAG
TCGTATTA

#4328: 3´-primer RNA affinity chromatography ESSV 15-16/28-29/34-35 (70nt)

ACACAGAGGAAGAACTATACGTCAGAATATGGACATGGGTGATCCTCATGTCCTATAGTGAG
TCGTATTA

#4330: 3´-primer RNA affinity chromatography ESSV>Neutral (70nt)

ACTTGGTTGTTTGGTTGTTTGGTTGTTTGGGGACATGGGTGATCCTCATGTCCTATAGTGAG
TCGTATTA

#4331: 3´-primer RNA affinity chromatography ESSV>ARE (70nt)

ACAATAAATAAATAAATAAATAAATAAATAGGACATGGGTGATCCTCATGTCCTATAGTGAG
TCGTATTA

#4332: 3´-primer RNA affinity chromatography ESSV>TR (70nt)

ACCTAACCCTAACCCTAACCCTAACCCTAAGGACATGGGTGATCCTCATGTCCTATAGTGAG
TCGTATTA

2.1.4 Recombinant plasmids

2.1.4.1 HIV-1 4 exon minigene

Template	PCR-Fragment	Donor vector	Product
LTR 4 Exon	#3641/#2588 AlwNI-SpeI	LTR 4 Exon	LTR 4 Exon ESSV$^-$
LTR 4 Exon	#4393/#2588 AlwNI-SpeI	LTR 4 Exon	LTR 4 Exon Neutral
LTR 4 Exon	#4294/#2588 AlwNI-SpeI	LTR 4 Exon	LTR 4 Exon ARE
LTR 4 Exon	#4295/#2588 AlwNI-SpeI	LTR 4 Exon	LTR 4 Exon TR

2.1.4.2 Proviral DNA

Template	Fragment	Donor vector	Product
pNL4-3	PflMI-EcoRI	LTR 4 exon ESSV$^-$	pNL4-3 ESSV$^-$
pNL4-3	PflMI-EcoRI	LTR 4 exon ESSV Neutral	pNL4-3 Neutral
pNL4-3	PflMI-EcoRI	LTR 4 exon ESSV TR	pNL4-3 TR
pNL4-3	PflMI-EcoRI	LTR 4 exon int1ext ESSV ARE	pNL4-3 ARE

2.1.4.3 MS2-fusion protein overexpression vector

Template	Fragment	Donor vector	Product
SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p37) del Ala-rich (aa 8-41)pA	#2850/#3442 NheI-NcoI #3443/#3444 NcoI-BglII	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p37) del Ala-rich (aa 8-41)	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p37)pA
SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p40) del Ala-rich (aa 8-41)pA	#2850/#3442 NheI-NcoI #3443/#3444 NcoI-BglII	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p45) del Ala-rich (aa 8-41)pA	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p40)pA
SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p45)pA	BamHI – StuI	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p37)pA	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p42)pA
SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p45) del Ala-rich (aa 8-41)pA	#2850/#3442 NheI-NcoI #3443/#3444 NcoI-BglII	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p45) del Ala-rich (aa 8-41)pA	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p45)pA
SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p37)pA	BamHI – StuI	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p45) (N)pA	SVSD4/SA7 scNLS-MS2 delFG hnRNP D p37/p42 ΔCTpA
SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p45)pA	BamHI – StuI	SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p45) (N)pA	SVSD4/SA7 scNLS-MS2 delFG hnRNP D p40/p45 ΔCTpA
SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p37)pA	#4290/#2851 BamHI-XhoI	SVSD4/SA7 scNLS-MS2 delFG hnRNP C1/C2pA	SVSD4/SA7 scNLS-MS2 delFG hnRNP D p37/p40 CTpA
SVSD4/SA7 scNLS-MS2 delFG hnRNP D (p45)pA	#4290/#2851 BamHI-XhoI	SVSD4/SA7 scNLS-MS2 delFG hnRNP C1/C2pA	SVSD4/SA7 scNLS-MS2 delFG hnRNP D p42/p45 CTpA

2.1.5 Antibodies

2.1.5.1   Primary antibodies

Target group	Protein	Provider	Cat.-No	Host	Clone	Working dilution
hnRNPs	hnRNP A1	Santa Cruz Biotechnology	sc-56700	Ms	9H10	1:500
	AUF1 (hnRNP D)	Millipore	07-260	Rb	poly	1:500
	hnRNP K/J	Santa Cruz	Sc-32307	Ms	3C2	1:1000
TAGs	FLAG	Sigma-Aldrich	F1804	Ms	M2	1:2500
Controls	β-Actin	Sigma-Aldrich	A5316	Ms	AC-74	1:100000
	MS2	Tetracore	TC-7004	Rb	poly	1:2500
HIV-1	p24	Aalto	D7320	Sh	poly	1:2500
	vpr	NIH	2221	Rb	poly	1:500

2.1.5.2   Secondary antibodies

Species	Provider	Cat.-No	Host	Conjugate	Working dilution
Mouse	GE Healthcare	NXA931	Sh	HRP	1:2500
Rabbit	GE Healthcare	NA943V	Dk	HRP	1:2500
Sheep	Dianova	713-035-003	Dk	HRP	1:2500

2.1.6 Kits

DIG Northern Starter Kit [Roche]   Cat.-No. 12039672910

Plasmid Midi Kit [Qiagen]   Cat.-No. 12145

Gel Extraction Kit [Qiagen]   Cat.-No. 28704

Ribomax^{TM} Large Scale RNA Production kit   [Promega]   Cat.-No. P1300

2.1.7 Reagents and chemicals

If not explicitly mentioned otherwise, chemicals of analytical chemistry grade were used from Merck, Riedel-de-Haen, Roth, Sigma and Serva. Preparation of growth media and solvents is described in the respective experimental protocols or derives from standard laboratory manuals.

Adipic acid dihydrazide-Agarose [Sigma-Aldrich]   Cat.-No. A0802

Ampicillin [Roche]   Cat.-No. 10835242001

AmpliTaq® DNA polymerase [Life Technologies].   Cat.-No. N0808-0166

Complete Protease Inhibitors [Roche]   Cat.-No. 1674498

Dithiothreitol (DTT) [Serva]   Cat.-No. 20710

DNase I recombinant, RNAse-free [Roche]   Cat.-No. 04716728001

dNTP Mix [Qiagen]   Cat.-No. 201901

ECL^{TM} Western Blotting Detection Reagent [GE healthcare]   Cat.-No. RPN2106

FCS [PAN Biotech]   Cat.-No. 3302

Gibco® DMEM [Life Technologies]   Cat.-No. 41966

Gibco® DPBS -CaCl$_2$ -MgCl$_2$ [Life Technologies]   Cat.-No. 14190

Gibco® RPMI 1640 GlutaMAX^{TM} [Life Technologies]   Cat.-No. 61870-010

Gibco® Trypan Blue Stain 0.4% [Life Technologies]   Cat.-No. 15250-061

Gibco® Trypsin-EDTA 0.05% [Life Technologies]   Cat.-No. 25300

HeLa nuclear extract of $5x10^9$ cells [Cilbiotech]   Cat.-No. CC-01-20-25

High Fidelity Polymerase [Roche]   Cat.-No. 1173265001

Hyperfilm™ ECL™ [GE healthcare]   Cat.-No. RPN3103K

LB Agar (Lennox L Agar) [Life Technologies]   Cat.-No. 22900-025

LB-Broth (Lennox) [Roth]   Cat.-No. X964.2

LE Agarose [Biozym]   Cat.-No. 840004

NP40/Igepal [Sigma-Aldrich]   Cat.-No. I3021

Nylon Membranes, positively charged [Roche]   Cat.-No. 11417240001

Oligo (dT)$_{15}$ primer [Roche]   Cat.-No. 10814290001

Pen Strep [Gibco, Life Technologies]   Cat.-No. 15140

Polyethylenimine, branched [Sigma-Aldrich]   Cat.-No. 408727

Protein-Assay [Biorad]   Cat.-No. 500-0006

Protran® Nitrocellulose Membranes [GE healthcare]   Cat.-No. 10401196

ReBlot Strong Plus [Millipore]   Cat.-No. 2509

Recombinant RNasin® Ribonuclease Inhibitor [Promega]   Cat.-No. N2511

Rotiphorese Gel 30 (37.5:1 acrylamide/bisacrylamide) [Roth] Cat.-No. 3029.1

Sieve GP Agarose (Genetic pure, low melting) [Biozym]   Cat.-No. 850050

SuperScript® III Reverse Transcriptase [Life Technologies]   Cat.-No 18080-085

T4-DNA Ligase [NEB]   Cat.-No. M0202S

TransIT®-LT1 [Mirus Bio LLC]   Cat.-No. 731-0028

Whatman™ paper [GE Healthcare]   Cat.-No. 3001-017

2.2 Methods

2.2.1 Cell culture

2.2.1.1   Maintenance of adherent eukaryotic cell lines

All eukaryotic cell lines were maintained according to the recommendations by ATCC. HeLa and HEK-293T cells were grown in Dulbecco's modified Eagle's Medium (DMEM, Gibco, Life Technologies) supplemented to contain 10% $^{vol}/_{vol}$ fetal calf serum (PAN) and 1% $^{vol}/_{vol}$ Pen/Strep (Gibco, Life Technologies). Adherent cell lines were split twice a week. Briefly, old media was aspirated, cells were rinsed twice with DPBS (Gibco, Life Technologies) and detached from their surface using 0.05% Trypsin-EDTA solution (Gibco, Life Technologies). Trypsin treatment was stopped using FCS containing media and an appropriate amount of cells were seeded according to their confluence.

2.2.1.2   Maintenance of eukaryotic cell lines growing in suspension

CB3 A1 cells were grown using MEM α (Gibco, Life Technologies) supplemented to contain 10% $^{vol}/_{vol}$ fetal calf serum (PAN) and 1% $^{vol}/_{vol}$ Pen/Strep. CB3 A1 cells were selected to contain the retroviral hnRNP A1 expression cassette by addition of G418 sulfate to a final concentration of 400mg/ml. The old media of suspension cells was changed twice a week. At confluence populations were split.

2.2.1.3   Determination of live cell count of a cell suspension

Cell titres were determined using a Neubauer chamber. Cells were diluted with an equal volume of trypan blue solution (Gibco, Life Technologies) and applied to the Neubauer chamber. All non-stained cells were counted in all four quarters and live cell count was calculated as mean cell number per quarter multiplied with the dilution factor (here 2) and volume factor (10,000, transition from 0.1 µl to ml) and expressed as cells/ml.

2.2.1.4   Establishment of long-term cryo-cultures

In order to establish cryo-cultures, cells were counted, spun down for 5min at 500g and resuspended in freezing medium containing 50% $^{vol}/_{vol}$ FCS, 40% $^{vol}/_{vol}$ DMEM or RPMI, respectively, and 10% $^{vol}/_{vol}$ dimethyl sulfoxide (DMSO, Roth). $1 \bullet 10^6$ cells were supplied per

cryo-tube (Nunc) and cells were carefully frozen overnight at -80°C using a freeze apparatus (Nunc). On the following day cells were permanently store in liquid nitrogen until use.

2.2.1.5   Transient transfection of eukaryotic cells

Eukaryotic cells were transfected either by using *Trans*IT®-LT1 reagent (Mirus) or polyethylenimine (PEI, Roth). Polyethylenimine was diluted to a final concentration of 1mg/ml before use. 2 to 3µl of both reagents were applied per µg of total DNA. Reagents and DNA were diluted separately in the appropriate FCS-free media and incubated for 5min at RT. Both fractions were mixed and incubated for 20min at RT. The transfection mix was applied to the cells close to the surface in a spiral movement.

2.2.2 Molecular cloning

2.2.2.1   Preparation of chemo-competent E. *coli* DH5α F'IQ

A 5ml LB pre-culture containing 25µg/ml kanamycin was inoculated with E. *coli* DH5α F'IQ (Invitrogen) and incubated overnight at 37°C on a shaker. The following day 200ml LB culture containing kanamycin 25µg/ml was inoculated with the total pre-culture, grown to an OD_{600} of 0.5 and harvested by centrifugation at 4,000rpm and 4°C for 5 min with centrifuge 5810R (Eppendorf). Pellets were resuspended into 20ml TFB 1 buffer (100mM RbCl; 50mM $MnCl_2$; 30mM KAc; 10mM $CaCl_2$; 15% $^{vol}/_{vol}$ glycerine] and incubated on ice for 90min. Cells were spun down at 4,000rpm and 4°C for 5min and pellets were resuspended in TFB 2 buffer [10mM Mops; 10mM RbCl; 75mM $CaCl_2$; 15% $^{vol}/_{vol}$ glycerine]. Chemo-competent cells were split in aliquots, immediately shock frozen using liquid nitrogen and stored at -80°C until use.

2.2.2.2   Heat shock transformation of chemo-competent E. *coli* DH5α F'IQ

An aliquot of 50µl chemo-competent E. *coli* DH5α F'IQ was subjected to 10µl of the ligation reaction and incubated for at least 20min on ice. DNA/cell mix was heat shocked for 90s at 42°C and immediately chilled on ice for 2min. 800µl LB medium was added and the culture was left for phenotypic expression at 37°C for 1h on a shaker. 400µl of the culture were plated on LB-agar plates containing 100µl/ml ampicillin and incubated overnight at 37°C.

2.2.2.3 Preparative PCR

Preparative PCR were performed using Expand High Fidelity System (Roche Applied Science) and Robo Cycler Gradient 96 (Stratagene). PCR reactions were performed according to the manufactures recommendations and in dependence of the melting points of the used primer set and amplicon size. A standard program is listed below.

Initial denaturation	94°C	3min	
Denaturation	94°C	30s	
Annealing	53°C	1min	35 cycles
Elongation	72°C	1min	
Final extension	72°C	10min	

2.2.2.4 Purification of DNA fragments by phenol-chloroform extraction

Prior extraction, ddH$_2$O was added to the DNA sample until a final volume of 200µl. 200µl phenol-chloroform-isoamyl alcohol mix (125:24:1, Roti Phenol/C/I, Roth) was added and the mixture was vortexed about 5s until white staining. Phases were separated by centrifugation at 11,000rpm for 15min using centrifuge 5417R (Eppendorf). The aqueous phase was transferred into a new tube and 200µl of chloroform-isoamyl alcohol mix (24:1) was added. After 5s on the vortex mixer phases were separated again by centrifugation. The aqueous phase was transferred into a new tube and 0.1 volumes of 5M LiCl and 2.5 volumes of absolute, pre-cooled ethanol were added. DNA was precipitated on dry ice for 20min and spun down at 14,000rpm 4°C for 30min. The pellet was washed with 200µl 70% $^{vol}/_{vol}$ ethanol and centrifuged again for 10min. Afterwards the pellet was air-dried and resuspended with 15µl ddH2O.

2.2.2.5 Photometric determination of nucleic acid concentrations

The concentration of nucleic acids within aqueous solutions was determined using a Nano Drop 1500 (Peqlabs). Measurement was performed according the manufactures recommendations.

2.2.2.6   Mini preparation of plasmids by alkaline lysis

Single colonies were transferred into a 5ml LB culture containing 100µg/ml ampicillin and incubated overnight at 37°C on a shaker. 2ml cell suspension spun down at 14,000rpm for 5min using centrifuge 5417R (Eppendorf). Bacterial pellets were resuspended with 300µl buffer 1 [50mM Tris-HCl pH8; 10mM EDTA; 100µg/ml RNAse A]. Cell lysis was performed by addition of 300µl buffer 2 [0.2M NaOH; 1% $^w/_{vol}$ SDS] and incubation for 5min at RT. Lysis was stopped by addition of 300µl of buffer 3 [3M KAc pH5.5]. Cell debris was spun down 15min at 14,000rpm. Supernatant was transferred into a new tube and 600µl isopropanol was added. DNA was precipitated 30min at 14,000rpm and 4°C. The pellet was washed with 200µl 70% $^{vol}/_{vol}$ ethanol and 10min spun down again. After air-drying the pellet was resuspended in 20µl ddH$_2$O.

2.2.2.7   Midi preparation of plasmids by using the Qiagen Midi Prep Kit

Preparative isolation of plasmid DNA was done using Plasmid Midi Kit (Qiagen). For Midi preparations 150ml LB culture containing 100µg/ml ampicillin was inoculated and incubated overnight at 37°C on a shaker. Bacteria were spun down for 15min at 5,000rpm and 4°C using Beckman Coulter centrifuge Avanti J-E and Rotor JA10. The pellet was resolved in 4ml buffer 1 [50mM Tris-HCl pH 7.5; 10mM EDTA pH 8.0; 400µg/µl RNase A]. Alkaline lysis was performed by addition of 4ml buffer 2 [0.2M NaOH, 1% $^w/_{vol}$ SDS] and 5min incubation at RT. Lysis was stopped by addition of 4ml buffer 3 [3M KAc pH5.5]. Cellular RNAs were removed by incubation for 15min on ice. Cell debris was removed by 10min centrifugation at 5,000rpm and 4°C using Beckman Coulter centrifuge Avanti J-E and Rotor JA10. Columns were equilibrated by addition and flow-through of 4ml QBT buffer [750mM NaCl; 50mM MOPS pH 7.0; 15% $^{vol}/_{vol}$ isopropanol; 0.15% $^{vol}/_{vol}$ Triton X-100]. Plasmid containing supernatant was filtrated through Whatman® prepleated qualitative filter paper (Whatman) in the equilibrated columns and flow through. Columns were washed twice by addition and flow-through of 10ml QC buffer [1M NaCl; 50mM MOPS pH 7; 15% $^{vol}/_{vol}$ isopropanol]. Bound plasmids were eluted from the columns by addition and flow-through of 5ml QF buffer [1.25M NaCl; 50mM Tris-HCl pH 8.5; 15% $^{vol}/_{vol}$ isopropanol] in corex centrifugation tubes containing 3.5ml isopropanol. Plasmid DNA was precipitated by centrifugation for 30min at 10,000rpm and 4°C using Beckman Coulter centrifuge Avanti J-E and Rotor JS 13.1. Pellet was

washed by addition of 2ml 70% $^{vol}/_{vol}$ ethanol and centrifugation for 10min at 10,000rpm and 4°C. Pellets were air-dried and resuspended in 50µl ddH$_2$O.

2.2.2.8 Preparation of bacterial cryo-cultures

Bacterial cryo-cultures were prepared by addition of 300µl glycerol to 700µl cell suspension and stored at -80°C.

2.2.2.9 Preparative restriction endonuclease digest

For all restriction endonucleases, buffers (New England Biolabs) were used according to the manufacturers' recommendations. Preparative digest were performed within a total volume of 20µl using approximately 3µg template and 10 units of each enzyme for 2h at 37°C.

2.2.2.10 Analytical restriction endonuclease digest of mini preparations

All restriction endonucleases and buffers (New England Biolabs) were used according to the manufacturers' recommendations. Analytical digests of mini preparations were performed using 10µl sample and 3 units of each enzyme for 1h at 37°C.

2.2.2.11 Recovery of DNA fragments from low-melting-temperature agarose gels

The gel was casted using 0.8% $^{w}/_{vol}$ low-melting-temperature agarose (Biozym) resolved in 1xTB^1/$_{10}$E buffer [89mM Tris; 89mM boric acid; 0.2mM EDTA]. The DNA-fragments were sliced out using a sharp razorblade and solubilized at 65°C.

2.2.2.12 Recovery of DNA fragments from standard agarose gels

DNA fragments were sliced out using a sharp razorblade and weighted. DNA was extracted using the Gel Extraction Kit (Qiagen) following the manufactures' recommendations.

2.2.2.13 Ligation of DNA fragments

All ligation reactions were performed using the T4 ligase [400U/µl] and the respective buffer from New England Biolabs in a total volume of 20µl. DNA fragments were ligated with an insert/backbone ratio of about 3/1 using 400 cohesive end units ligase per reaction and incubated overnight at 16°C.

2.2.2.14 Agarose gel electrophoresis

Typically 1% $^w/_{vol}$ agarose (Biozym) gels were casted using 1x TBE buffer [89mM Tris; 89mM boric acid; 2mM EDTA]. Electrophoresis was performed at 80mA. Agarose gels were stained with ethidium bromide diluted in 1x TBE buffer. Following staining, gels were exposed to UV irradiation and documented using camera optics provided by Intas.

2.2.2.15 DNA sequencing

DNA sequencing analysis was performed by the "Biomedizinisches Foschungszentrum" using a template concentration of 250ng/µl for plasmids, 25ng/µl for linear DNA-fragments and primer concentration of 10pmol/µl.

2.2.3 RNA methods

2.2.3.1 Total RNA isolation from eukaryotic cells

The protocol for RNA isolation from eukaryotic cells represents a modified method, published by Chomczynski and Sacchi [46]. Briefly, every step was done on ice or at 4°C, respectively. The cells were washed twice with PBS and 500µl solution D [4M guanidinium thiocyanate, 25mM sodium citrate, 0.5% $^w/_{vol}$ N-lauroylsarcosine, 0.1M β-mercaptoethanol] was added. Cells were scraped off using a cell scraper and transferred into a new tube. In order to extract the RNA, 1vol acidic phenol (Roti Aqua-Phenol, Roth), 0.1vol 2M sodium acetate pH 4, 7.2µl β-mercaptoethanol and 0.2vol chloroform isoamyl alcohol (24:1) was added. After vigorously shaking for 15s, samples were chilled for 15min on ice and phases separated by centrifugation for 20min at 11,000rpm using centrifuge 5417R (Eppendorf). In a total of 400µl the aqueous phase was mixed with 400µl isopropanol and RNA was precipitated overnight at -20°C. The following day RNA was pelleted by centrifugation for 30min at 14,000rpm. Pellets were washed twice by addition of 200µl 70% $^{vol}/_{vol}$ ethanol diluted with DEPC-H$_2$O [0.1% $^{vol}/_{vol}$ diethyl pyrocarbonate] and subsequent centrifugation for 10min at 14,000rpm. The pellet was air-dried, resolved in 10µl DEPC-H$_2$O [0.1% $^{vol}/_{vol}$ diethyl pyrocarbonate] and stored until use at -80°C.

2.2.3.2   DNase I digest and cDNA synthesis

Prior to cDNA synthesis a DNase I digest was performed, in order to remove residual plasmid contaminations. In a total volume of 10µl containing 3µg RNA and 1µl RNAse-free DNase I [10U/µl] (Roche Applied Science) DNase I digest was allowed for 20min at 37°C and subsequent 40min at RT. Following DNase I heat inactivation for 5min at 70°C 4.5µl were subjected to subsequent cDNA synthesis. 6.5µl DEPC-H_2O [0.1% $^{vol}/_{vol}$ diethyl pyrocarbonate], 1µl 10mM dNTP mix (Qiagen), 1µl 7.5µM oligo-dT_{15} primer (Roche Applied Science) were added, following 5min of incubation at 65°C, to dissolve RNA secondary structures. 4µl 4x first strand buffer (Life Technologies), 1µl 0.1M DTT, 1µl RNasin [40U/µl] (Promega) and 1µl superscript III RNase H-deficient reverse transcriptase [200U/µl] (Life Technologies) were added, oligo-dT_{15} primer was annealed for 60min at 50°C and cDNA was synthesized for 15min at 72°C.

2.2.3.3   Semi-quantitative PCR

Semi-quantitative PCR was done using AmpliTaq^R PCR system (Life Technologies) and Robo Cycler Gradient 96 (Stratagene). Briefly, to 5µl cDNA as template, 5µl 10x PCR reaction buffer [100mM Tris-HCl, pH 8.3, 500mM KCl, 15mM $MgCl_2$, 0.01% $^{w}/_{vol}$ gelatine] (Life Technologies), 1µl 10mM dNTP mix (Qiagen), 0.25µl AmpliTaq^R polymerase [5U/µl] (Life Technologies), 1µl each of the respective 10mM primer and ddH_2O to a final volume of 50µl were added. PCR was performed using a standard program listed below.

Initial denaturation	94°C	3min	
Denaturation	94°C	30s	
Annealing	53°C	1min	x cycles
Elongation	72°C	1min	
Final extension	72°C	10min	

Typically, semi-quantitative PCRs were conducted using 35 cycles. However, in order to detect cDNA derived from human growth hormone transcripts, the primer set #1224/#1225 and 25 cycles were used.

2.2.3.4 Native polyacrylamide gel electrophoresis

Typically native 10% $^{vol}/_{vol}$ polyacrylamide (Rotiphorese gel 30, Roth) gels were casted using 1x TBE buffer [89mM Tris; 89mM NaBO$_4$; 2mM EDTA]. Electrophoresis was performed at 30mA. Native gels were stained within 1x TBE buffer containing 0.04% $^{vol}/_{vol}$ ethidium stock solution [10mg/ml ethidium bromide]. Following staining, gels were exposed to UV irradiation and documented using camera optics provided by Intas.

2.2.3.5 Recovery of DNA-fragments from polyacrylamide gels

DNA-fragments were sliced out using a sharp razorblade and transferred into a 1.5ml reaction tube (Eppendorf). 200µl elution buffer [0.5M ammonium acetate, 10mM magnesium acetate, 1mM EDTA pH 8, 0,1% $^{w}/_{vol}$ SDS] were added and the suspension was incubated overnight at 37°C on a shaker. Eluted DNA was further isolated using the gel extraction kit (Qiagen) following the manufactures' recommendations assuming a 200mg agarose fragment.

2.2.3.6 Production of digoxigenin-labelled probes for northern blot analysis

Unlabelled probes for northern blot analysis were produced by preparative PCR using the primer pair #3387/#3388 spanning 153nt of HIV-1 exon 7 and pNL4-3. Unlabelled probes were analysed by agarose gel electrophoresis, recovered from the gel and purified using Gel Extraction Kit (Qiagen) following the manufactures' recommendations. Dig-labelled HIV-1 exon 7 probes were produced by preparative PCR using primer pair #3387/#3388 and 10x PCR DIG probe synthesis mix [2mM dATP, 2mM dCTP, 2mM dGTP, 1.3mM dTTP, 0.7mM DIG-dUTP] (Roche). Deviating from the standard protocol (see 2.2.2.3) preparative PCR was conducted in a total volume of 40µl. Successful Dig-labelling was verified by agarose gel electrophoresis of the labelled and unlabelled probes. 20µl Dig-labelled HIV-1 exon 7 probe were added to 10ml Dig Easy Hyb Granules (DIG Northern Starter Kit, Roche Applied Science) and stored at -20°C until use.

2.2.3.7 Northern blot

Prior to northern blotting a 1% $^w/_{vol}$ agarose gel was casted containing 1xMEN buffer [20mM MOPS, 5mM sodium acetate, 1mM ETDA] and 2% $^{vol}/_{vol}$ formaldehyde. The samples were prepared by dilution of 3µg RNA per sample with DEPC-H$_2$O [0.1% $^{vol}/_{vol}$ diethyl pyrocarbonate] to a final volume of 10µl and addition of 2.5µl RNA loading dye [1mM EDTA, 50% $^{vol}/_{vol}$ glycerine, 0.25% $^w/_{vol}$ bromophenol blue, 0.25% $^w/_{vol}$ xylencyanol, 40µg/ml ethidium bromide]. Subsequently, RNAs were denatured for 10min at 70°C and loaded onto the gel. Electrophoresis was done in 1xMEN buffer at 60V. RNA isolation and equal sample loading indicated by the 28S and 18S rRNA bands was documented by UV light exposure of the gel. Gel and the positively charged nylon membrane were washed briefly with ddH$_2$O and incubated in 20xSSC [3M NaCl, 0.3M sodium citrate, pH 7] until use. The blot was assembled from bottom to top using an appropriate amount absorptive paper, three layers of dry Whatman^{TM} paper (GE healthcare), the positive charged nylon membrane (Roche), the gel, three layer of Whatman^{TM} paper soaked in 20xSSC and a Whatman^{TM} paper based bridge linking blot and a 20xSSC reservoir. Capillary blotting was done overnight at RT. The blot was UV cross-linked using CL-1000 Ultraviolet Crosslinker (UVP) [λ=254nm, 120,000 µJ/cm²]. After washing the membrane twice with ddH$_2$O, the membrane was pre-hybridized by incubation with DIG Easy Hyb Granules (DIG Northern Starter Kit, Roche Applied Science) for 2h at 55°C. 10ml DIG Easy Hyb Granules containing the HIV-1 exon 7 probe was incubated for 10min at 100°C, in order to denature the dsDNA of the probe. After a subsequent cold shock on ice preventing renaturation, the HIV-1 exon 7 probe was applied to the membrane. Hybridization was carried out overnight at 55°C. In order to reduce unspecific binding of the probe, affinity washes were carried out. Briefly, the membrane was washed once briefly with ddH$_2$O, twice for 5min at RT with stringent wash buffer I [2xSSC, 0.1 % $^w/_{vol}$ SDS], twice briefly with ddH$_2$O and twice for each 20min at 68°C with stringent wash buffer II [0.2xSSC, 0.1 % $^w/_{vol}$ SDS]. After affinity washes the membrane was washed twice in ddH$_2$O and then incubated in maleic acid buffer [0.1M maleic acid, 0.15M NaCl, pH 7.5] at RT. The 1x blocking solution, was prepared out of 10x blocking solution (DIG Northern Starter Kit, Roche Applied Science) diluted to 1x concentration with maleic acid buffer, applied to the membrane and incubated 30min at RT. After blocking, polyclonal anti-digoxigenin Fab-fragments conjugated with an alkaline phosphatase (DIG Northern Starter

Kit, Roche Applied Science) diluted 1:20,000 in 1x blocking solution was applied and incubated for 30min RT. Following three washes with maleic acid buffer 10min each at RT, the membrane was dried using Whatman™ paper and incubated 5min RT with CDP Star (DIG Northern Starter Kit, Roche Applied Science). The northern blot was exposed between 10min to 1h to an Amersham Hyperfilm™ ECL™ (GE Healthcare). The exposed film was developed using Cawomat 2000IR (Cawo GmbH), Developer G153 (Agfa), Rapid Fixer G354 (Agfa) and ddH$_2$O. Alternatively, fluorescence emitted by peroxidase catalysed reaction was detected and quantified using Gel Chemo Cam Imager system (Intas Science Imaging).

2.2.3.8 *In vitro* transcription of RNA oligonucleotides

In vitro transcription of RNA oligonucleotides was performed using RiboMAX™ Large Scale RNA Production System (Promega) following the manufactures' recommendations. The templates for in vitro transcription were produced by annealing of an anti-sense DNA oligonucleotide containing in 3' to 5' direction the T7 polymerase binding site, MS2 phage loop sample sequence with a sense DNA oligonucleotide containing only the T7 binding site. Briefly, 2µg of the anti-sense oligonucleotide and an equimolar amount of the sense oligonucleotide were combined within a 1.5ml reaction tube (Eppendorf), incubated for 10min at 100°C in a water bath and left stirring, gradually cooling down overnight at 4°C. *In vitro* transcription reaction was set up by mixing the annealed template with 4µl 5x T7 transcription buffer (400mM HEPES-KOH pH 7.5, 120mM MgCl$_2$, 10mM spermidine, 200mM DTT), 6µl rNTPs (25mM each) and 2µl T7 enzyme mix and adjusted to a final reaction volume of 20µl with nuclease-free ddH$_2$O. After 2-4h incubation at 37°C residual DNA-template was removed by DNase I digestion. Briefly, 1U RQ1 RNase-free DNase I [1U/µl] per µg template was added and digestion performed for 15min at 37°C. After DNase I digest residual protein was removed by phenol-chloroform extraction. Prior extraction ddH$_2$O was added to the DNA sample until a final volume of 200µl. 100µl acidic phenol (Roti Aqua-Phenol, Roth) and 100µl chloroform-isoamyl alcohol mix (24:1) were added and the mixture was vortexed for at least 1min until white staining. Phases were separated by centrifugation at 11,000rpm for 15min using centrifuge 5417R (Eppendorf). The aqueous phase was transferred into a new tube and 200µl of chloroform-isoamyl alcohol mix (24:1) was added. After 1min on the vortex mixer the phases were again separated by centrifugation. The aqueous phase was

transferred into a 1.5ml reaction tube (Eppendorf) and 0.1 volumes of 3M sodium acetate pH 5.2 and 2.5 volumes of absolute, precooled ethanol were added. RNA was precipitated on ice for 5min and spun down at 14,000rpm 4°C for 30min. The pellet was washed with 200µl 70% $^{vol}/_{vol}$ ethanol diluted with DEPC-H$_2$O [0.1% $^{vol}/_{vol}$ diethyl pyrocarbonate] and centrifuged again for 10 min. Afterwards the pellet was air-dried and resuspended with 15µl DEPC-H$_2$O [0.1% $^{vol}/_{vol}$ diethyl pyrocarbonate]. Integrity of RNA oligonucleotides was checked on an agarose gel.

2.2.3.9 RNA pull-down analysis

RNA pull-down analysis was done using RNA oligonucleotides immobilized on agarose beads. In order to couple RNA oligonucleotides to adipic acid dihydrazine-agarose beads, first the RNA oligonucleotides need to be oxidized at the 3'-end ribose. 1,000pmol of *in vitro* transcribed RNA oligonucleotides were diluted in DEPC-H$_2$O [0.1% $^{vol}/_{vol}$ diethyl pyrocarbonate] to a final volume of 340µl. After addition of 40µl 1M sodium acetate pH 5 diluted in DEPC-H$_2$0 and 20µl of a saturated solution sodium meta-periodate diluted in DEPC-H$_2$0, the reaction mix was incubated for 1h RT in the darkness. Subsequently 80µl 1M sodium acetate pH 5 diluted in DEPC-H$_2$0 and 2.5vol absolute, precooled ethanol was added. After precipitation for 5min at -80°C, RNAs were pelleted by centrifugation for 30min at 14,000rpm and 4°C using centrifuge 5417R (Eppendorf) and air-dried. Meanwhile for each sample 150µl adipic acid dihydrazine-agarose beads were washed four times with 0.1M sodium acetate pH 5 diluted in DEPC-H$_2$0 using centrifuge 5810R (Eppendorf) each time for 5min at 4,000rpm and 4°C. After washing, the bead suspension was adjusted to a final volume of 1ml per sample with 0.1M sodium acetate pH 5 diluted in DEPC-H$_2$0 and added to the RNAs. Coupling of the activated RNA oligonucleotides to the adipic acid dihydrazine-agarose beads was performed overnight at 4°C on a shaker. In order to remove uncoupled RNA oligonucleotides from immobilized RNAs, the beads were washed twice by addition of 1ml 2M NaCl diluted in DEPC-H$_2$0, following vigorous shaking and subsequent centrifugation for 1min at 14,000rpm using centrifuge 5417R (Eppendorf). In order to adjust samples to nucleus-like conditions, samples were washed thrice with 1ml buffer D [20mM HEPES-KOH pH 7.9, 5% $^{vol}/_{vol}$ glycerol, 0.1M KCL, 0.2mM EDTA, 0.5mM DTT] and subsequent centrifugation. In a total volume of 200µl HeLa nuclear extract was diluted to concentration

of 50% with buffer D and supplemented with 1µg MS2 coat protein per sample. 200µl of the diluted nuclear extract were added to the immobilized RNAs and protein binding was allowed for 30min at 30°C. Five sequential affinity washes were carried out at elevated salt conditions by addition of 1ml buffer D_{Mg} [20mM HEPES-KOH pH 7.9, 5% $^{vol}/_{vol}$ glycerol, 0.1M KCL, 0.2mM EDTA, 0.5mM DTT, 4mM $MgCl_2$] containing 100mM, 250mM, 500mM, 250mM or 100mM KCl and subsequent centrifugation. After washing, an equal volume protein sample buffer [0.75M Tris-HCl, pH 6.8, 20% $^{vol}/_{vol}$ glycerol, 10% $^{vol}/_{vol}$ β-mercaptoethanol, 4% $^{w}/_{vol}$ SDS] was added to the bead pellet and the bound proteins were eluted from the immobilized RNAs by incubation for 10 min at 95°C. Beads were spun down for 15min at 14,000rpm, the elution fraction was transferred into a new 1.5ml reaction tube (Eppendorf) and stored until use at -20°C.

2.2.4 Protein methods

2.2.4.1 Protein isolation of eukaryotic cells

Cells were washed twice with PBS, scraped form cell culture dishes and transferred into a new 1.5ml reaction tube (Eppendorf). Cells were pelleted by centrifugation for 5min at 1,500rpm and the pellet was resolved in 50µl RIPA buffer [25mM Tris-HCl pH 7.6, 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS]. Subsequently total cell lysates were established by freeze and thaw method. Therefore samples were frozen for 10min at -20°C and thawed on ice. Cell lysates were stored at -20°C until use.

2.2.4.2 Determination of protein concentration by Bradford assay

Protein concentrations were determined using the protein assay (Bio-Rad) based on a default Bradford assay [22]. Dye solution was prepared by 1:5 dilution of the stock solution. 1ml of dye solution was transferred into a new 1.5ml reaction tube (Eppendorf), 1 – 10 µl of the sample was added and after vigorous shaking incubated for at least 5min RT. OD_{595} was measured using UV Mini 1240 UV-Vis spectrometer (Shimadzu) and protein concentration calculated using standard curve data out of serial dilutions of BSA.

2.2.4.3   Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Proteins samples were separated under denaturing conditions as published by Laemmli [120] in discontinuous sodium dodecyl sulfate (SDS) polyacrylamide gels (Rotiphorese Gel 30, Roth) using Mini-Protean vertical gel electrophoresis system (Bio-Rad). Gels were run in 1x Laemmli running buffer [1% $^w/_{vol}$ SDS, 0.25M Tris-Base, 1.9M glycine] for 1-2h at 30mA and 4°C. Discontinuous polyacrylamide gels were casted in the following compositions (Tab. 1)

Table 1: Discontinuous SDS-PAGE compositions

Separation gel			Stacking gel	
Ingredients:	12%	15%	Ingredients:	5%
30% PAA	14.4ml	18.0 ml	30% PAA	4.5ml
2M Tris-HCl pH 8	7.6ml	7.6ml	0.5M Tris-HCl pH 6.8	3.6ml
20% $^w/_{vol}$ SDS	180µl	180µl	20% $^w/_{vol}$ SDS	135µl
ddH$_2$O	13.5	9.9ml	60% $^w/_{vol}$ Sucrose	6.3ml
10% $^w/_{vol}$ APS	432µl	432µl	ddH$_2$O	12.6ml
TEMED	72µl	72µl	10% $^w/_{vol}$ APS	360µl
			TEMED	36µl

2.2.4.4   Immunoblotting

Separated proteins by SDS-PAGE were immobilized on Whatman^{TM} Protran® nitrocellulose membranes (GE healthcare) by electroblotting. Prior blot assembly the PVDF membrane and Whatman^{TM} paper were soaked in 1x blotting buffer. The blot was assembled from cathode to the anode using, three layers of Whatman^{TM} paper (GE healthcare), the PVDF membrane, the gel, three layer of Whatman^{TM} paper. Electroblotting was performed using a semi-dry blotting chamber (Transblot® SD Semi-Dry, Bio-Rad) and 1x blotting buffer [192mM glycine, 25mM Tris-Base, 20% $^{vol}/_{vol}$ methanol] for 70min at constant 15V. Further a tank blotting chamber (Mini Trans-Blot® Cell, Bio-Rad) and 1x trans blotting buffer [192mM glycine, 25mM Tris-Base, 0.1% $^w/_{vol}$ SDS, 20% $^{vol}/_{vol}$ methanol] were used for electroblotting using the following parameters, 15min at 150mA constant, 20min at 300mA constant. After blotting the protein transfer to the PVDF membrane was controlled by Ponceau S staining [2% $^w/_{vol}$

Ponceau-S, 30% $^{vol}/_{vol}$ trichloroacetic acid, 30% $^{w}/_{vol}$ 5-sulfosalicylic acid]. Subsequently the membrane was destained using 1x TBS-T buffer [20mM Tris-HCl pH 7.5, 150mM NaCl, 0.1% $^{vol}/_{vol}$ Tween 20] and blocked for 1h at RT using 10% $^{w}/_{vol}$ non-fat dry milk powder diluted in 1x TBS-T buffer. Following blocking membrane was probed with the primary antibody diluted in 1x TBS-T buffer containing 5% $^{w}/_{vol}$ non-fat dry milk powder for 2h at RT or overnight at 4°C. Thereafter the membrane was washed three times for 10min at RT with 1x TBS-T buffer and secondary antibody conjugated to horseradish peroxidase (HRP) diluted in 1x TBS-T buffer containing 5% $^{w}/_{vol}$ non-fat dry milk powder was applied to the membrane for 1h at RT. After at least three washing steps using 1x TBS-T buffer immunodetection was performed using ECL Western Blotting System (GE Healthcare) following the manufacturers' recommendations. The membrane was exposed between 1s to 1h to an Amersham Hyperfilm™ ECL™ (GE Healthcare). The film was developed using Cawomat 2000IR (Cawo GmbH), Developer G153 (Agfa), Rapid Fixer G354 (Agfa) and ddH$_2$O. In alteration to X-ray film based evaluation of immunoblotting, fluorescence emitted by peroxidase catalysed reaction was detected and quantified using Gel Chemo Cam Imager system (Intas Science Imaging). For reprobing of the membranes, bound antibodies were removed through incubation in stripping solution (ReBlot Plus Strong, Millipore) for 5 to 20min at RT.

2.2.4.5 Mass spectrometric-compatible silver staining of protein gels

Separated proteins by SDS-PAGE were subjected to silver staining based on the published method by Heukeshoven [176]. Briefly, the gel was fixated overnight in buffer A [50% $^{vol}/_{vol}$ ethanol, 10% $^{vol}/_{vol}$ acetic acid]. After fixation the gel was sensitized by incubation in buffer B [30% $^{vol}/_{vol}$ ethanol, 500mM sodium acetate, 8mM NaS$_2$O$_3$] for 30min and subsequently stained with buffer C [6mM AgNO$_4$] for 30min. After staining and a short wash with buffer D [236mM Na$_2$CO$_3$], the gel was developed with buffer E [236mM Na$_2$CO$_3$, 0.01% $^{vol}/_{vol}$ formaldehyde] for a period of 1 to 10min. Development was stopped by incubation in buffer F [50mM EDTA] for 20min. Gels were washed afterwards with ddH$_2$0 and stored until use submerged at 4°C.

2.2.4.6 Mass spectrometric analysis of silver stained protein bands

Sample preparation and MS-analysis were performed by the Molecular Proteomics Laboratory (MPL) of the BMFZ, Heinrich-Heine-University Düsseldorf. Briefly, in order to

prepare silver stained protein bands for analysis by mass spectrometry, first the bound silver has to be removed, second the proteins need to be fragmented and third the proteins need to be extracted from the gel. Destaining of the gel slices was achieved by addition of 10µl destaining solution [15mM $Na_2S_2O_3$, 50mM $K_3FeIII(CN)_6$] and incubation for 1min. Following destaining, the slices were washed three times for 10min with 30µl of solution A [10mM NH_4HCO_3] and solution B [5mM NH_4HCO_3, 50% $^{vol}/_{vol}$ acetonitrile] in a consecutive manner (ABABAB). After drying the slices using a SpeedVac, in gel digestion of the proteins was performed overnight by addition of 5µl trypsin solution [0.033µg/µl] and incubation at 37°C. Digested proteins were extracted twice by addition of 30µl extraction solution [50% $^{vol}/_{vol}$ acetonitrile, 0.05% $^{vol}/_{vol}$ trifluoroacetic acid], following 10min sonification in a sonifier bath. Protein containing supernatants were pooled and dried out using a SpeedVac. Protein containing pellets were resuspended in 17µl 0.1% $^{vol}/_{vol}$ trifluoroacetic acid. 15µl sample were subjected to liquid chromatography using RSLCnano system (Thermo Fischer Scientific). Separated protein fragments were directly subjected to ESI-MS spectrometer containing a 3D ion-trap (HCTultra PTM Discoverer System, Bruker Daltonics).

2.2.4.7   Immunoprecipitation

7.5 10^6 HEK293T cells were seeded in 100mm tissue culture dishes. Adherent cells were transfected twice with 20µg expression plasmid and incubated overnight. The cells were washed off the dish, spun down with 5,000rpm for 1min using centrifuge 5810R (Eppendorf), and washed twice with 500µl pre-cooled PBS. Cell lysis was carried out by resuspending the cell pellet in 600µl IP lysis buffer [140mM NaCl, 5mM $MgCl_2$, 20mM Tris-HCl pH 7.6, 1% $^{vol}/_{vol}$ NP-40, cOmplete protease inhibitor (Roche)] and incubation on ice for 30min. Lysates were pre-cleared by centrifugation at 14,000rpm and 4°C for 30min using centrifuge 5417R (Eppendorf). 35µl of the lysate were subjected to an equal volume of 2x protein sample buffer [0.75M Tris-HCl, pH 6.8, 20% $^{vol}/_{vol}$ glycerol, 10% $^{vol}/_{vol}$ β-mercaptoethanol, 4% $^{w}/_{vol}$ SDS], incubated at 95°C for 10min and stored until use at -20°C. 4µl monoclonal ANTI-FLAG® M2 antibody (Sigma) were added to the residual lysate and shaken overnight using a circulatory shaker at 4°C. Protein G Sepharose 4 Fast Flow (GE healthcare) was washed with buffer B [150mM NaCl, 10mM Tris-HCl pH 7.6, 2mM EDTA, 1% $^{vol}/_{vol}$ NP-40, cOmplete protease inhibitor (Roche)] and refilled with buffer B up to the starting volume. Two drops

out of a cropped 1,000 μl tip (approx. 100μl) were added to the lysates and the mixture was left on the circulatory shaker at 4°C for 1h. Affinity washing steps were carried out consecutively by centrifugation at 14,000rpm and 4°C for 1min using centrifuge 5417R (Eppendorf), following resuspension twice in buffer B, once in buffer C [500mM NaCl, 10mM Tris-HCl pH 7.6, 2mM EDTA, 1% $^{vol}/_{vol}$ NP-40, cOmplete protease inhibitor (Roche)] and once in buffer D [10mM Tris-HCl pH 6,8, cOmplete protease inhibitor (Roche)]. After pelleting sepharose by centrifugation at 14,000rpm and 4°C for 1min using centrifuge 5417R (Eppendorf), an equal volume of 2x protein sample buffer was added and the bound proteins eluted at 95°C for 10min. Samples were separated from the sepharose by centrifugation at 14,000rpm and 4°C for 1min using centrifuge 5417R (Eppendorf) and stored at -20°C until use.

2.2.4.8   Recombinant expression and purification of MBP-MS2

A 5ml LB pre-culture containing 100μg/ml ampicillin was inoculated with E.coli K12 containing lac-repressor controlled expression cassette of the MBP-MS2 fusion protein (kindly provided by Klemens Hertel). After incubation overnight at 37°C on a shaker, a 500ml LB culture containing 100μg/ml ampicillin was inoculated with the total pre-culture and grown for 2-3h at 37°C on a shaker until an OD_{600} of 0.40 to 0.45 was reached. Bacterial expression of MBP-MS2 was induced by addition of IPTG to a final concentration of 0.2mM and maintained for 4h at 37°C on a shaker. Bacteria were spun down for 15min at 4,000rpm and 4°C using centrifuge Avanti J-E (Beckman Coulter) and Rotor JA10 (Beckman Coulter). The pellet was resuspended with 10ml lysis buffer [20mM Tris-HCl pH 7.5, 200mM NaCl, cOmplete protease inhibitor (Roche)] and spun down again for 10min at 2,000rpm and 4°C. Supernatant was discarded, the pellet resuspended again in 15ml lysis buffer and distributed equally in three 50ml Falcons. Cell lysis was performed for using Branson Sonifier 450 with a cup horn pre-cooled to 4°C. Sonification was carried out three times for 10min using output 5.5 and a duty cycle of 50%. From this point on samples were handled at 4°C or on ice. Lysates were pre-cleared by centrifugation for 10min at 9,000rpm using centrifuge 3K30 (Sigma) and 19776 rotor (Sigma). MBP affinity chromatography was prepared by addition of 550μl agarose resin (NEB) into a disposable 10ml polypropylene column (Thermo Scientific) and washed with a column volume of buffer D [20mM HEPES, 100mM KCl, 1mM MgCl$_2$,

0,2mM EDTA, 5mM CaCl$_2$, cOmplete protease inhibitor (Roche)]. Pre-cleared lysate was supplemented with 1M CaCl$_2$ until final concentration of 5mM. Chromatin digestion was started by addition of 5µl RNAse-free DNase I [10U/µl] (Roche). Lysates were applied onto the column and binding was allowed overnight on a circulatory shaker. Loaded resin was subjected to three affinity washes by the consecutive application of a column volume of lysis buffer, 10min incubation on the circulatory shaker and flow through. Column was equilibrated with a column volume phosphate buffer [5mM Na$_3$PO$_4$ pH 7]. During equilibration a new column was prepared with 550µl heparin-agarose resin (Gentauer) and washed with a column volume of phosphate buffer. After equilibration, bound proteins were eluted from the first column direct into the second. Therefore four times 500µl MBP elution buffer [5mM Na$_3$PO$_4$ pH 7, 15mM maltose, cOmplete protease inhibitor (Roche)] was applied and the column was incubated for 15min on the circulatory shaker. Following elution binding of the eluted proteins to the heparin column was allowed overnight on the circulatory shaker. After a single wash of the heparin column with one column volume phosphate buffer, bound proteins were eluted. Elution from the heparin column was done by three times application of 500µl elution buffer [20mM HEPES, 100mM KCl, 1mM MgCl$_2$, 0,2mM EDTA, 5mM CaCl$_2$, 10% $^{vol}/_{vol}$ glycerol, cOmplete protease inhibitor (Roche)] and incubation for 15min on the circulatory shaker. Protein concentration was measured by Bradford assay. MS2 protein was stored as 20µl aliquots at -80°C.

3   Results

3.1   ESSV of HIV-1 is not exclusively controlled by hnRNP A1

Since siRNA-mediated knock-down of hnRNP A1 failed to induce excessive HIV-1 exon 3 splicing [99], another splicing regulatory protein was assumed to additionally or alternatively bind to the ESSV. It was further proposed that the AU-rich sequences within HIV-1 exon 3 might contribute to the regulation of mRNA stability [117]. Consistently with that proposal, putative binding sites within the ESSV were identified that were presumably bound by the known mRNA stability regulatory protein hnRNP D. However, when mRNA stability regulation by exon 3 sequence could be ruled out [134], the picture emerged that hnRNP D binding to the putative binding sites might represent the presumed splicing regulatory protein additionally contributing to ESSV activity. Though, prior addressing that notion any further, the underlying evidence had to be validated. Therefore an HIV-1 derived splicing reporter system was established, which is able to detect changes in the splicing pattern induced by the absence of hnRNP A1. However, since HIV-1 harbours at least 4 hnRNP A1-dependent splice silencer elements throughout its genome, the ESSV, ESS2, ESS3 and ISS, the splicing reporter was restricted to the first 4 exons of the HIV-1 genome resulting in an HIV-1 derived minigene with a truncated intron 1, termed LTR 4 exon ESSV [227]. This minigene was derived from the minigene LTR ex2 ex3 [200] by insertion of the EcoRI/NdeI fragment of HIV-1 NL4-3 (Fig. 3.1 I A). Due to the exonic splicing silencer activity of the ESSV it was expected that this splicing reporter reflects poor recognisability of the internal exons but on the other hand allows the detection of the so-called excessive splicing phenotype following inactivation of the ESSV. Thus, to be able to analyse the excessive splicing phenotype, the ESSV inactivating mutation ESSV$_x$ [13] was inserted as control (LTR 4 exon ESSV$^-$). As expected, the processed transcripts derived from LTR 4 exon ESSV (ESSV) predominantly showed exon 2 and exon 3 exclusion (tat1) reflecting the splicing repressive phenotype of this subgenomic splicing reporter (Fig. 3.1 I B lane 1). Furthermore, albeit at low level, inclusion of exon 2 (tat2) appeared to be dominant over inclusion of exon 3 (tat3) and the intron-containing *vpr*-mRNA (vpr) (Fig. 3.1 I B lane 1).
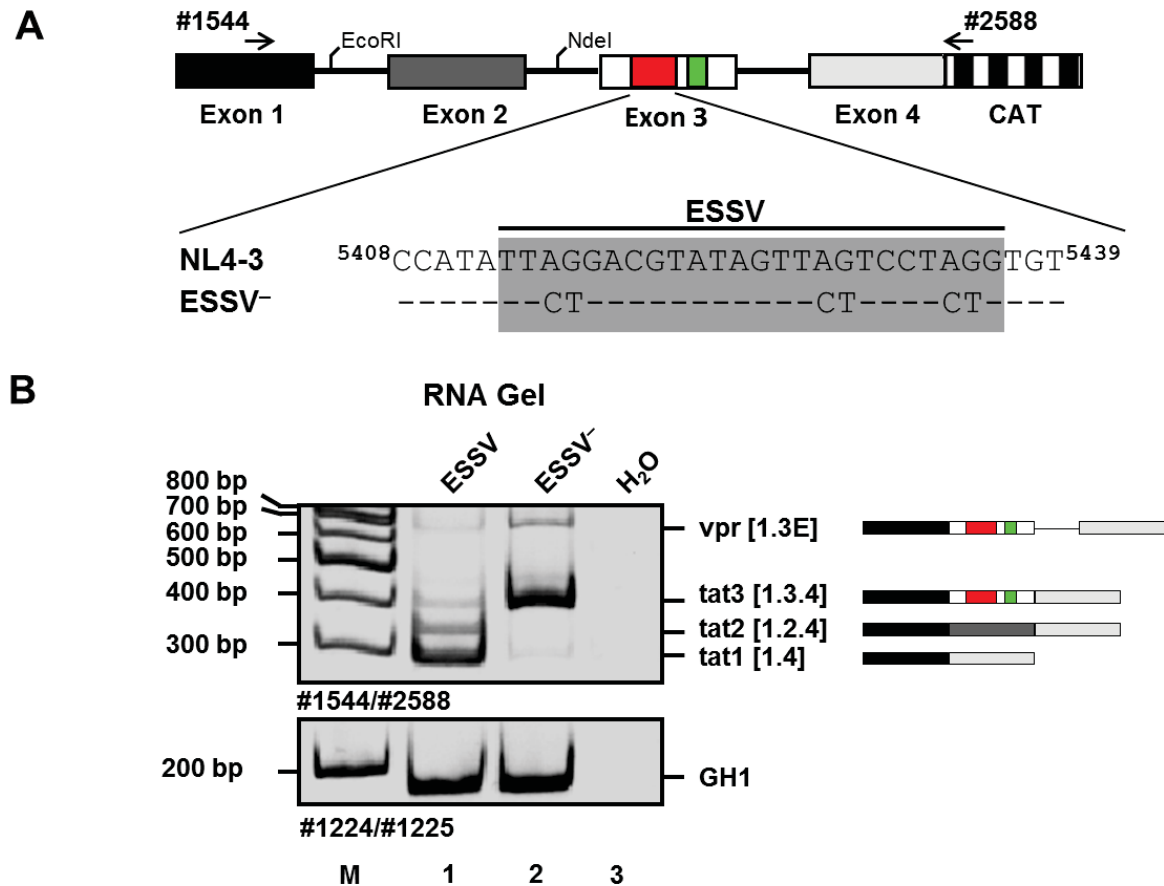
Fig 3.1 I: Splice patterns derived from LTR 4 exon minigenes in dependence on the presence of the ESSV.

(A) Schematic drawing of HIV-1 derived minigene LTR 4 exon ESSV containing the first four exons of the HIV-1 NL4-3 genome with a truncated, but extended intron 1 [227]. The SREs within HIV-1 exon 3, the ESSV and ESE$_{vpr}$ are coloured in red and light green. The reporter splicing pattern was assessed by semi-quantitative RT-PCR using primer set #1544/#2588 (arrows). ESSV$^-$ mutation equals the ESSV$_x$ mutation published by Bilodeau et al. [13], targeting all three (Py/A)-UAG motifs of the ESSV. Nucleotide positions were designated in reference to HIV-1 Hxb2 strain. (B) Transfection of minigenes. 2.5 x 10^5 HeLa cells were seeded in 6 well plates and transfected with 1μg of the designated LTR 4 exon splicing reporter, 1μg pxGH5 and 0.2μg SVctat. 24h post transfection total RNA was isolated, DNase I digested and reverse transcribed. The resulting cDNA template was amplified by semi-quantitative PCR using the reporter specific primer set #1544/#2588. In order to control the transfection efficiency, the expression of the human growth hormone (GH1) mRNA by pxGH5 was monitored using the exon-junction specific primer set #1224/#1225 resulting in PCR-fragment of 186nt. Reporter specific splice isoforms were termed according to Purcell and Martin [173] and schematically illustrated on the right side of the panel. The PCR-products of the designated viral splice isoforms obtained by using the designated primer sets were depicted within the supplementary material (Fig. 6.2 I). Lanes containing DNA-marker were highlighted (M).

However in contrast to recent observations with infectious viruses, the used HIV-1 derived minigene lacks the preference for the intron-containing *vpr*-mRNA (vpr) over the inclusion of exon 3 (tat3) [227]. In line with recent literature, however, substitution of the ESSV with ESSV⁻ led to almost exclusive exon 3 recognition displayed by high levels of the exon 3 containing mRNA isoforms tat3 and also somewhat higher levels of *vpr*-mRNA (vpr) (Fig. 3.1 I B lane 2). Notably, the previously observed preference of exon 3 inclusion (tat3) over the intron-containing *vpr*-mRNA (vpr) was maintained. In summary, the LTR 4 exon ESSV minigene showed that the ESSV-mediated interference with the recognition of HIV-1 exon 3 could be set aside by substitution of ESSV with the ESSV⁻ mutant sequence allowing efficient inclusion of HIV-1 exon 3 (Fig. 3.1 I B cf. lanes 1 with 2). Thus, a reliable reporter for excessive HIV-1 exon 3 splicing was constructed.

In order to determine, if the observed excessive splicing phenotype exhibited by LTR 4 exon ESSV⁻ could exclusively attributed to the loss of hnRNP A1 binding, the splicing pattern was analysed in the hnRNP A1 deficient CB3 cell line. CB3 cells represent a murine erythroleukemic cell line induced by infection with murine friend leukemia virus (F-MuLV) [196]. Retroviral insertion and further mutagenesis, was shown to completely abolish hnRNP A1 expression [11]. Since hnRNP A1 represents a highly abundant cellular protein and siRNA knock down experiments commonly fail to result in a 100% knock out of the target gene, the CB3 cell line was preliminary chosen for further analysis even though it is not an HIV-1 host cell. The CB3 A1 cell line was used as an internal control. This cell line was retrovirally transduced with a murine hnRNP A1 expression cassette restoring hnRNP A1 expression [119]. First, in order to validate the functionality of the knock out system, total protein lysates of CB3 and CB3 A1 cells were subjected to immunoblot analysis (Fig. 3.1 II A). As expected, only CB3 A1 cells exhibited detectable amounts of hnRNP A1 whereas both cell lines showed comparable levels of β-actin. The latter was used as an internal control (Fig. 3.1 II A cf. lane 1 with lane 2). Following verification of the hnRNP A1 knock out, the abundance of hnRNP D within both cell lines was compared (Fig. 3.1 II A). As anticipated, the immunoblot analysis showed comparable levels of hnRNP D expression for both cell lines (Fig. 3.1 II A cf. lane 1 with lane 2). Thus, the CB3/CB3 A1 knock out system represents a suitable system to verify the previous knock down data [99] and the hypothesis that hnRNP D might also contribute to the ESSV activity.

First, in order to control previous results, HeLa cells were transfected with the HIV-1 derived minigenes LTR exon 4 ESSV and ESSV⁻ containing either the functional or inactivated ESSV (Fig. 3.1 II B lane 1-2). As anticipated, transfection of HeLa cells with LTR 4 exon ESSV (ESSV) dominantly led to exclusion of HIV-1 exon 2 and 3 (tat1), whereas LTR 4 exon ESSV⁻ (ESSV⁻) mainly displayed HIV-1 exon 3 including transcripts, tat3 and vpr (cf. Fig. 3.1 I B lanes 1-2 with Fig. 3.1 II B lanes 1-2). Transfection of CB3 A1 cells resulted in a comparable splicing pattern as transfection of HeLa cells excluding obvious cell type specific HIV-1 exon 3 splicing differences (Fig. 3.1 II B cf. lanes 1-2 with lanes 5-6). Consistent with recent literature [99], in absence of hnRNP A1 expression (CB3 knock out cells) the presence of a considerable amount of tat1 mRNA indicated that the recognition of exon 3 was still clearly repressed (Fig. 3.1 II A lane 3). Thus, the lack of hnRNP A1 expression does only partly offset the exon 3 splicing repressing effect and another splicing regulatory factor must be involved in exon 3 repression. Moreover, since the splicing pattern of LTR 4 exon ESSV⁻ in CB3 cells was comparable to the other two hnRNP A1 expressing cells (Fig. 3.1 II A cf. lanes 2, 4, 6), these data suggest that the ESSV inactivating mutations also affect binding of the assumed factor, possibly hnRNP D.
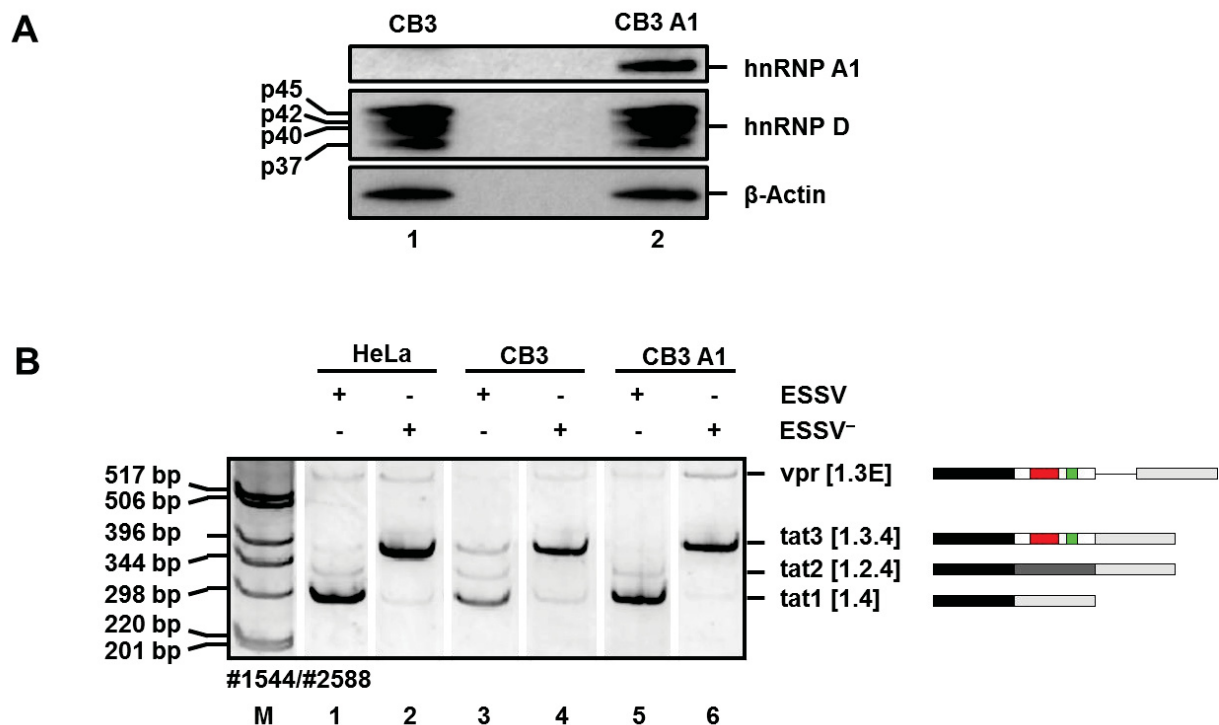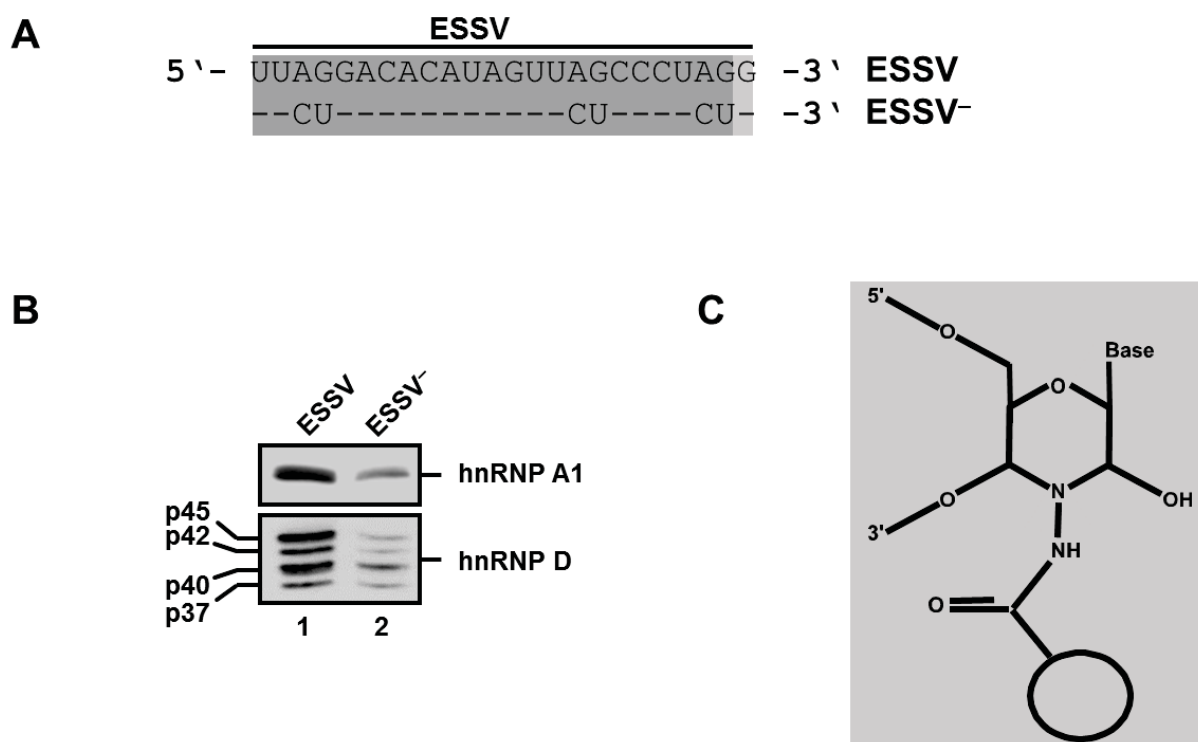
Fig. 3.1 II: ESSV function is only partially dependent on hnRNP A1

(A) Immunoblot analysis of total cellular proteins of CB3 and CB3 A1 cells. 5 x 10[6] cells of each were collected and subjected to protein isolation. After determination of protein concentration using Bradford assay, equal amounts of protein lysates were separated using 10% SDS-PAGE and electro-blotted. Immunoblot analysis was done using anti-hnRNP A1, anti-hnRNP D and anti-β-actin antibodies. (B) Transfection of LTR 4 exon minigenes. 2.5 x 10[5] HeLa, CB3 or CB3 A1 cells were seeded in 6 well plates and transfected with 1µg of the designated LTR 4 exon splicing reporter, 1µg pxGH5 and 0.2µg SVctat. In order to enable HIV-1 Tat activity in murine cells, additionally CB3 and CB3 A1 cells were co-transfected with human cyclin T1 and CDK9 expression vectors, to reconstitute human pTEFb. In this line, CB3 and CB3 A1 cells were co-transfected with 1µg each of pcDNA cyclin T1 and pcDNA CDK9 expression vector, whereas HeLa cells were co-transfected with 2µg of non-coding pcDNA3.1(+). 24h post transfection total RNA was isolated, DNase I digested and reverse transcribed. The resulting cDNA template was amplified by semi-quantitative PCR using the reporter specific primer set #1544/#2588. Reporter specific splice isoforms were termed according to Purcell and Martin [173] and schematically illustrated on the right side of the panels. The PCR-products of the designated viral splice isoforms obtained by using the designated primer sets were depicted within the supplementary material (Fig. 6.2 I). For reasons of comprehensibility, lanes of the same gel were rearranged and the resulting transitions were represented by gaps. Lanes containing DNA-marker were highlighted (M).

In order to test the hypothesis that hnRNP D can also bind to the ESSV sequence and that its binding capacity is equally impaired by the ESSV⁻ sequence RNA-affinity chromatography was performed (Fig. 3.1 III).

A

```
                      ESSV
5'-  UUAGGACACAUAGUUAGCCCUAGG  -3'  ESSV
     --CU-----------CU----CU-  -3'  ESSV⁻
```

B

p45
p42
p40
p37

hnRNP A1

hnRNP D

C

Fig. 3.1 III: HnRNP D binds to the ESSV.

(A) Design of RNA affinity chromatography probes. RNA oligonucleotides were produced by in vitro transcription using commercial available RiboMAX™ Kit (Promega). The resulting RNA oligonucleotides encompass the 24nt sequence of HIV-1 NLA1 ESSV (grey). Only mutations within ESSV⁻ sequence were depicted, identical sequences were indicated by a hyphen. B) RNA affinity chromatography approach including immunoblot analysis. The 3'-end standing ribose of each RNA oligonucleotide was oxidized using periodate and linked to adipic acid dihydrazide coated agarose beads. The immobilized RNA oligonucleotides were incubated in the presence of HeLa nuclear extract. Bound proteins were subjected to stringent affinity washes and eluted. Following separation of the eluted proteins by SDS-PAGE and western blotting, immunoblot analysis was carried out using antibodies against hnRNP A1 and hnRNP D. (C) Schematic representation of the immobilized RNA bound to the beads by covalent linkage of its 3'-standing ribonucleotide.

As expected, immunoblot analysis of precipitated proteins revealed comparable binding of hnRNP A1 and hnRNP D to the ESSV sequence (Fig. 3.1 III B lane 1), thereby confirming hnRNP D as putative candidate protein contributing to ESSV-mediated interference with the recognition of HIV-1 exon 3. Further, the ESSV⁻ mutant described by Bilodeau and co-workers [13] not only impaired binding of hnRNP A1, but comparably also binding of hnRNP D (Fig 3.1 III B lane 2). Thus, the observed partial offset of ESSV activity, when hnRNP A1 expression was lost, together with the verification of hnRNP D binding to the ESSV and the decrease in binding of hnRNP D along with hnRNP A1 leading to the onset of excessive splicing, strongly indicates a contribution of hnRNP D to the interference with HIV-1 exon 3 recognition mediated by the ESSV.

3.2   hnRNP D – a splicing regulatory protein

Since the preceding experiments strongly indicated that hnRNP D binding to the ESSV might contribute to its interference with HIV-1 exon 3 recognition, an MS2-tethering assay was employed to assess the effect of hnRNP D binding on exon recognition. The MS2-tethering system allows guiding of the protein of interest to a pre-defined spot on a chosen RNA, where the effect of the protein can be studied independently of its natural RNA binding site. The system itself is based on high specific binding of the bacteriophage derived RNA-binding protein, the MS2 phage coat protein, to its viral target RNA structure, the MS2-loop [166, 172, 216] (Fig. 3.2 I B). In order to apply this system to be able to assess the effect of a specific protein on mRNA splicing, each protein of interest had to be expressed as an MS2-

fusion protein where ideally the domain of interest is covalently linked to a single chain version of the MS2 dimer, forming a single "dimeric" MS2-fusion protein (Fig. 3.2 I A). The expressed MS2-fusion protein in turn binds highly specific to the MS2-loops exposed on a modified pre-mRNA co-expressed by a splicing reporter construct (Fig. 3.2 I B). By that means, the effect of any MS2-fusion protein on pre-mRNA splicing of the splicing reporter could be analysed by comparing the obtained splicing patterns. However, employing the MS2-tethering assay faces two mayor obstacles to be tackled. First, in order to allow engagement of the MS2-fusion protein with the MS2-loop exposed on the splicing reporter pre-mRNA, each MS2-fusion protein is required to be transported into the nucleus. Therefore nuclear import was mediated by incorporation of the NLS of HIV-1 Tat protein within the N-terminus of each MS2-fusion protein, thus enabling its nuclear import in an importin-β dependent manner [214]. Second, since the genuine MS2 phage coat protein was observed to exhibit unwanted inter-dimer interactions mediated by the so called FG-loop structure, the FG-loops of the two MS2 phage coat protein subunits expressed within each MS2-fusion protein were deleted [125]. The modified MS2-fusion proteins used here were expressed by the established SV scNLS-MS2 ΔFG vector system [200] (Fig. 3.2 I A). In the splicing assay the HIV-1 derived splicing reporter LTR ex2 ex3 (2xMS2) SD3$_{Down}$ [200] was used, which encompasses the first four exons of pNLA1 with a truncated version of intron 1. Within this splicing reporter the ESSV sequence within HIV-1 exon 3 was replaced by two MS2-loops (Fig. 3.2 I C). In order to maintain in absence of the ESSV balanced exon 3 inclusion levels, the complementarity of the 5'ss D3 to the U1 snRNA was artificially decreased (HBS 14.0 > 12.3), termed SD3$_{Down}$. Altogether, using both vector systems the effect of each MS2-fusion protein could be assessed 24h post transfection by RNA isolation, DNase I digest and semi-quantitative RT-PCR.
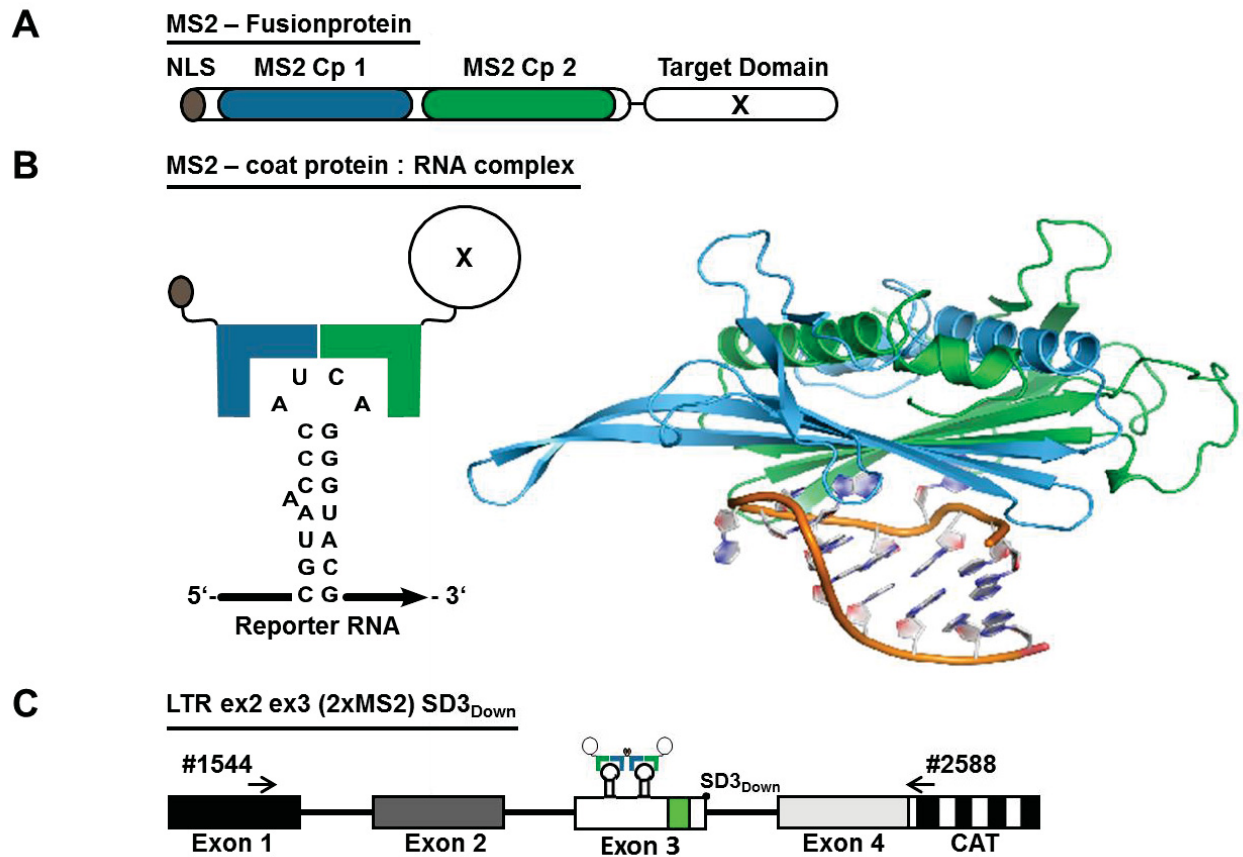
Fig. 3.2 I: The underlying principles of the MS2-tethering assay.

(A) Schematic drawing of the MS2-fusion protein expressed by SV scNLS-MS2 ΔFG vector system containing the NLS of HIV-1 protein Tat (grey), two MS2 phage coat proteins (dark blue/dark green), each modified by deletion of their FG-loop, and the target peptide. (B) Illustration of the MS2-fusion protein binding to the MS2-loop structure within the reporter RNA (right panel). Further detail of the MS2:Loop-interaction is provided by the ribbon diagram of an MS2 dimer binding to the MS2-loop. This diagram originates from van den Worm and co-workers [216] (PDB ID: 1AQ3) and was illustrated and kindly provided by Dr. Neil Ranson (Leeds University, UK). The combined β-sheet of the two MS2 coat subunits represents the protein-RNA interaction surface with the MS2-loop (left panel). (C) Schematic drawing of the HIV-1 derived minigene LTR ex2 ex3 (2xMS2) SD3$_{Down}$ containing the first four exons of pNLA1 with a truncated version of intron 1 and modified by replacement of the ESSV sequence within exon 3 by two MS2-loops [200]. The adjacent element ESE$_{vpr}$ [68] (light green) was left unaltered. Nevertheless, in order to allow a balanced inclusion pattern of exon 3, the complementarity of 5'ss D3 to the U1 snRNA was decreased (SD3$_{Down}$, HBS 14.0 > 12.3) [200]. The splicing regulatory effect of each MS2-fusion protein on splicing of the reporter pre-mRNA could be assessed by semi-quantitative RT-PCR using the primer set #1544/#2588.

However, prior to assessing the effect of hnRNP D on exon recognition, the suitability of the constructed splicing reporter and the effect of the MS2-coat protein *per se* needed to be assessed. Therefore two controls were included within the MS2-tethering assay (Fig. 3.2 II lanes 1-2). First, the splicing reporter was transfected without a MS2-fusion protein expression vector (Fig. 3.2 II lane 1). Second, the splicing reporter was transfected with an expression vector, expressing the modified MS2-coat protein dimer with its N-terminal NLS domain, but lacking a target domain (Fig. 3.2 II lane 2). Transfection of the reporter only, led to the expression of 3 distinct splice isoforms, termed in reference to the original HIV-1 derived mRNA isoforms tat1, tat3 and vpr (Fig. 3.2 II lane 1). As anticipated, when compared to LTR 4 exon ESSV, the substitution of the ESSV by the MS2-loops within the LTR ex2 ex3 (2xMS2) SD3$_{Down}$ splicing reporter led to a high degree of HIV-1 exon 3 recognition resulting in exon 3 including splice isoforms vpr and tat3 (cf. Fig. 3.1 I B lane 1 with Fig. 3.2 II lane 1). Though notably, instead of featuring almost exclusive exon 3 inclusion as it was observed with the LTR 4 exon ESSV$^-$ splicing reporter, the decrease in complementarity of SD3 (SD3$_{Down}$, HBS 14.0 > 12.3) to the U1 snRNA within LTR ex2 ex3 (2xMS2) SD3$_{Down}$ led to a decreased recognition of exon 3 resulting in levels of tat1 mRNA comparable to that of tat3 (cf. Fig. 3.1 I B lane 2 with Fig. 3.2 II lane 1). In summary, the LTR ex2 ex3 (2xMS2) SD3$_{Down}$ minigene represents a suitable splicing reporter for the MS2-tethering assay enabling a wide detectable range of either splicing silencing or splicing enhancement. Moreover, coexpression of the MS2-coat protein dimer binding to the MS2-loop did not change the observed splicing pattern (Fig. 3.2 II cf. lane 1 with lane 2) indicating that binding of the MS2-coat protein dimer to the MS2-loops not affects exon recognition. Thus, any deviating effects on exon recognition observed, when an MS2-fusion protein is tethered to the MS2-loops, could be assigned to the protein domain coupled to the MS2-coat dimer. In order to assess the effect of hnRNP D on exon recognition, all four hnRNP D isoforms were coexpressed as MS2-fusion proteins and their effect on the splicing pattern of the splicing reporter was assessed by RT-PCR (Fig. 3.2 II lanes 3-6). In line with the hypothesis that hnRNP D binding to the ESSV interferes with HIV-1 exon 3 recognition, tethering of all four hnRNP D isoforms similarly resulted in decreased exon recognition (Fig. 3.2 II cf. lanes 1-2 with lanes 3-6). In particular, the observed decrease in tat3 and intron-containing *vpr*-mRNA (vpr) was accompanied by an increase in tat1 mRNA lacking

both, exon 2 and 3. Noteworthy, due to the introduction of the HIV-1 Tat derived NLS, nuclear import of all MS2-fusion proteins is mediated independently of their endogenous NLS. Thus, the intensity of splice inhibition observed here for each hnRNP D isoform might slightly vary.
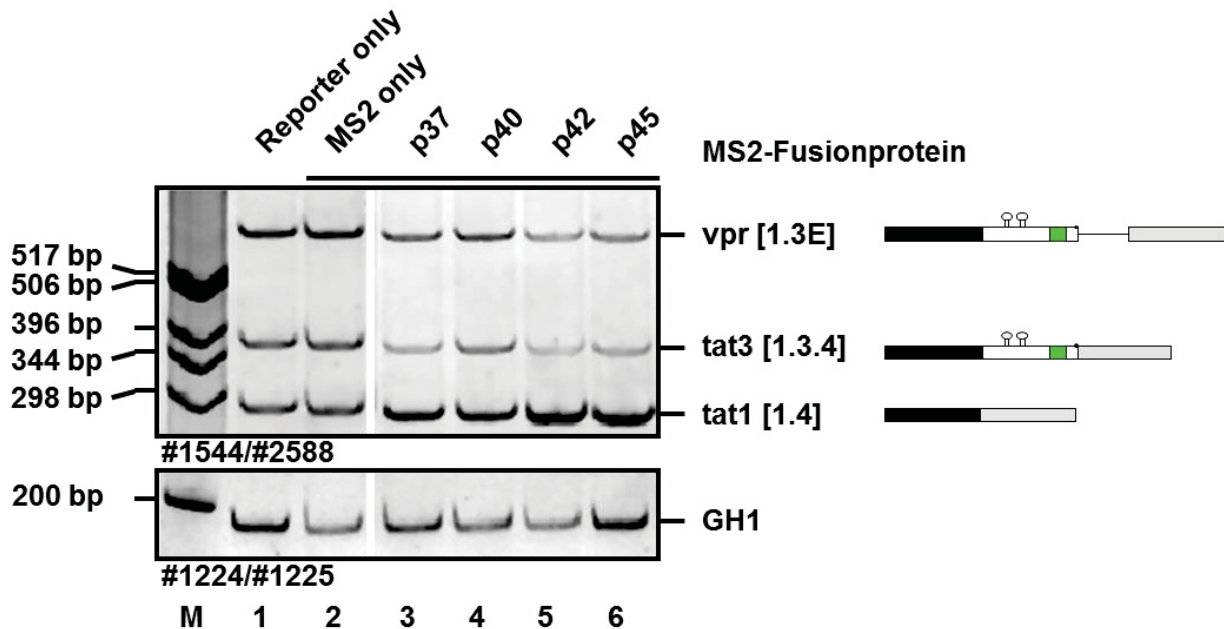

Fig. 3.2 II: MS2-tethering of all hnRNP D isoforms results in decreased exon recognition.

MS2-tethering assay. 2.5×10^5 HeLa cells were seeded in 6 well plates and transfected with 1µg of the designated SV scNLS-MS2 ΔFG vector and 1µg of the splicing reporter LTR ex2 ex3 (2xMS2) SD3$_{Down}$, 1µg pxGH5 and 0.2µg SVctat. 24h post transfection total RNA was isolated, DNase I digested and reverse transcribed. The resulting cDNA template was amplified by semi-quantitative PCR using the reporter specific primer set #1544/#2588. In order to monitor the transfection efficiency, the expression of the human growth hormone (GH1) mRNA by pxGH5 was assessed using exon-junction specific primer set #1224/#1225 resulting in a PCR-fragment of 186nt. Reporter specific splice isoforms were termed in reference to Purcell and Martin [173] and schematically illustrated on the right side of this panel. The PCR-products of the designated viral splice isoforms obtained by using the designated primer sets were depicted within the supplementary material (Fig. 6.2 I). For reasons of comprehensibility, lanes of the same gel were rearranged and the resulting transitions were represented by gaps. Lanes containing DNA-marker were highlighted (M).

In order to identify a specific domain that might contribute the observed interference, deletion mutagenesis was carried out, to generate diverse hnRNP D truncation mutants (Fig. 3.2 III).
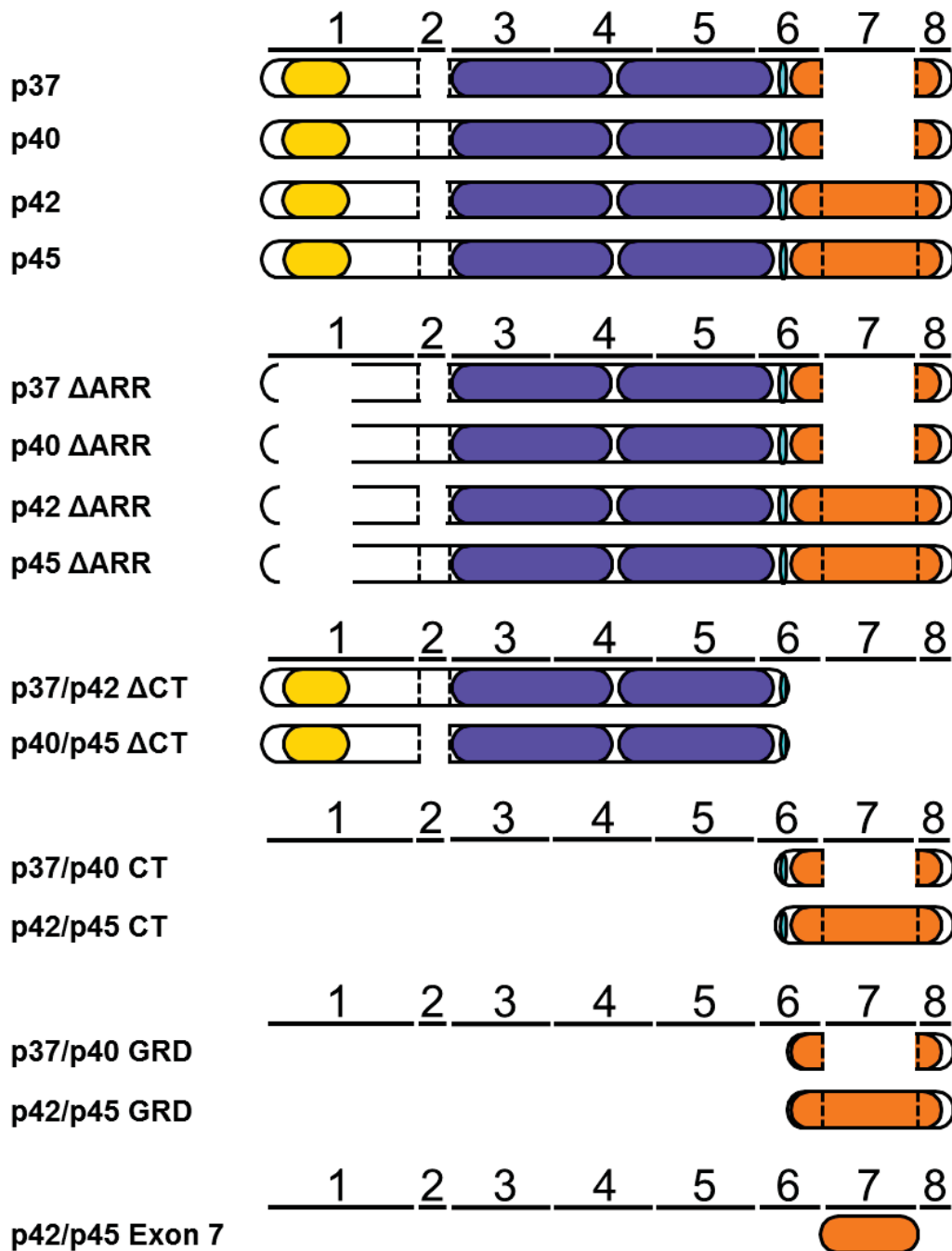
Fig. 3.2 III: Generation of hnRNP D deletion mutants.

Each isoform was cloned by PCR using specific primers and BamHI – XhoI restriction sites into the SV scNLS-MS2 ΔFG vector system. Depicted are all full length and truncated isoforms of hnRNP D used in the MS2-tethering assay. Each protein isoform is depicted in reference to the coding exons of the hnRNP D locus (lines above). The colour code unifies similar elements: A-rich region (yellow), G-rich domain (orange), Q-rich region (light blue) and RRMs (purple). Boxes with dotted lines represent protein sequences encoded by the alternative exons 2 (small box) and 7 (large box) of the hnRNP D locus.

In order to assess the activity of the established hnRNP D deletion mutants on exon recognition, they were expressed as MS2-fusion proteins and thus evaluated by the MS2-tethering assay using the four full-length protein isoforms as positive control (Fig. 3.2 IV). Consistent with previous results transfection of the LTR ex2 ex3 (2xMS2) SD3$_{Down}$ splicing reporter without and with coexpression of the MS2-coat protein dimer without a target domain resulted in the detection of the three distinct transcript isoforms, corresponding to tat1, tat3 and vpr (cf. Fig. 3.2 II lanes 1-2 with Fig. 3.2 IV lanes 1-2). Moreover, all tethered full-length isoforms of hnRNP D in comparison to the controls resulted in decreased exon 3 recognition (Fig. 3.2 IV cf. lanes 1-2 with lanes 3-6). Consistent with the previous MS2-tethering assay, the induced shift from exon 3 including mRNA isoforms (tat3, vpr) to tat1 mRNA excluding exon 2 and exon 3 further indicates the capacity of all hnRNP D isoforms to interfere with exon recognition (cf. Fig. 3.2 II lanes 3-6 with Fig. 3.2 IV lanes 3-6). Since deletion of the N-terminal A-rich region (ΔARR) failed to result in any increase of exon 3 including mRNA isoforms, tat3 and vpr, the A-rich region appears to be negligible for the interference with exon recognition (Fig. 3.2 IV cf. lanes 7-10 with lanes 3-6). The deletion of the C terminus resulted in two truncated, hnRNP D isoforms (ΔCT) deviating from each other only by the presence of the alternative exon 2 peptide (Fig. 3.2 III). Interestingly, tethering both C-terminal truncated hnRNP D isoforms restored exon 3 recognition to control level indicating a contribution to splicing inhibition (Fig. 3.2 IV cf. lanes 11-12 with lanes 1-2). Noteworthy, an alternative hypothesis to the observed loss of splicing inhibition might be that the deletion of the C terminus either results in aberrant folding or cellular mis-localisation of the truncated protein. However, since it was shown, that accurate folding of both RRMs of hnRNP D occurs completely independent of other parts of the protein [53], aberrant folding seems to be less likely albeit it cannot formally ruled out. Moreover, since the HIV-1 Tat NLS was incorporated into all MS2-fusion proteins, a deficit in nuclear localisation is highly unlikely. Thus, the C termini of all hnRNP D isoforms contribute to the interference with exon recognition. In order to test the hypothesis if independently from other parts of the protein the C terminus of hnRNP D was sufficient to exhibit interference with exon recognition, the N-terminal part of hnRNP D was deleted resulting in two truncated C termini (CT) deviating from each other only by the presence of the domain encoded by alternative exon 7 (Fig. 3.2 III).
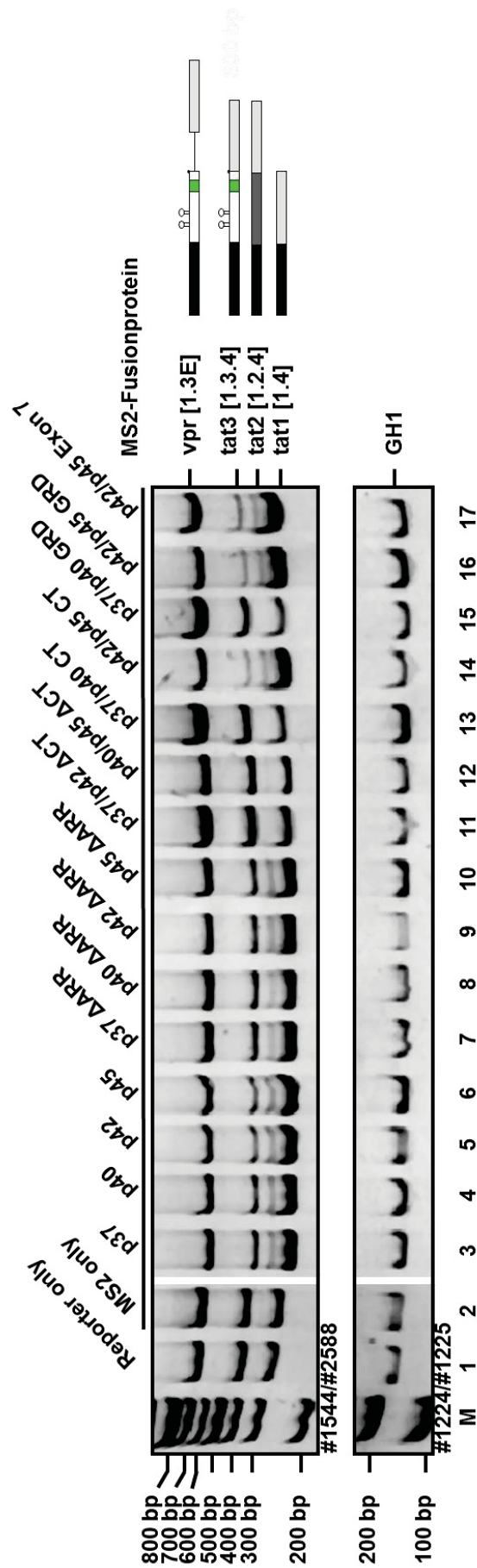
Fig. 3.2 IV: The hnRNP D domain encoded by the alternative exon 7 interferes with exon recognition.

MS2-tethering assay. 2.5×10^5 HeLa cells were seeded in 6 well plates and transfected with 1µg of the designated SV scNLS-MS2 ΔFG vector and 1µg of the splicing reporter LTR ex2 ex3 (2xMS2) SD3$_{Down}$, 1µg pxGH5 and 0.2µg SVctat. 24h post transfection total RNA was isolated, DNase I digested and reverse transcribed. The resulting cDNA template was amplified by semi-quantitative PCR using the reporter specific primer set #1544/#2588. Reassembling the transfection control, the expression of the human growth hormone (GH1) mRNA by pxGH5 was monitored using exon-junction specific primer set #1224/#1225 resulting in PCR-fragment of 186nt. Reporter specific splice isoforms were termed in reference to Purcell and Martin [173] and schematically illustrated on the right side of this panel. The PCR-products of the designated viral splice isoforms obtained by using the designated primer sets were depicted within the supplementary material (Fig. 6.2 I). For reasons of comprehensibility, lanes of the same gel were rearranged and the resulting transitions were represented by gaps. Lanes containing DNA-marker were highlighted (M).

As anticipated, tethering of the C terminus of hnRNP D isoforms p42 and p45 resulted in interference with exon 3 recognition mapping the effector domain to the C terminus (Fig. 3.2 IV cf. lanes 14 with lanes 5-6). Interestingly, tethering of the C terminus of isoforms p37 and p40 exhibited a splicing pattern similar to the controls and thus appeared not to affect exon recognition (Fig. 3.2 IV cf. lane 13 with lanes 1-2). However, in contrast to the C-terminal truncation mutants tethering of full-length hnRNP D p37 and p40 showed interference with exon recognition (Fig. 3.2 IV cf. lanes 3-4 with lane 11). Thus, in contrast to hnRNP D isoforms p42 and p45 the interference with exon recognition by hnRNP D isoforms p37 and p40 appears not only to be provided by the C terminus indicating the contribution of another yet unidentified region. However, since the C terminus of hnRNP D isoforms p42 and p45 appeared to independently exhibit interference with exon recognition, mapping was continued by tethering of the corresponding GRD (GRD) and the GRD of hnRNP D isoforms p37 and p40 as control (Fig. 3.2 IV lane 15-16). As anticipated, similar to the corresponding C terminus the GRD of hnRNP D isoforms p37 and p40 failed to interfere with exon recognition (Fig. 3.2 IV cf. lane 15 with lanes 1-2), while the GRD of hnRNP D isoforms p42 and p45 comparable to the corresponding C terminus exhibited interference with exon recognition (Fig. 3.2 IV cf. lane 16 with lane 14). However, since both GRDs deviate only by the presence of the domain encoded by the alternative exon 7, the contribution of the exon 7 encoded domain was assessed (Fig. 3.2 IV lane 17). Interestingly, tethering of the exon 7 derived domain independently of other parts of the GRD equally reconstituted the shift from exon 3 including splice isoforms (tat3, vpr) to tat1 mRNA excluding exon 3 and exon 2

indicating interference with the recognition of exon 3. Moreover, the observed interference of the domain encoded by the alternative exon 7 was comparable to the interference exhibited by the corresponding full-length isoforms of hnRNP D p42 and p45 (Fig. 3.2 IV cf. lane 17 with lanes 5-6). Thus, the domain encoded by alternative exon 7 appears to be the effector domain of hnRNP D isoforms p42 and p45, which establishes the interference with exon recognition exhibited by those hnRNP D isoforms. In summary, both MS2-tethering experiments showed that all isoforms of hnRNP D interfere with exon recognition. Moreover, the domain encoded by alternative exon 7 apparently enables hnRNP D isoforms p42 and p45 to interfere with exon recognition. Whereas, interference with exon recognition exhibited by hnRNP D isoforms p37 and p40 appeared to be partially dependent on the C-terminal part of the protein and on another yet unidentified domain.

3.3 Overexpression of hnRNP D suggests additional hnRNP D binding sites

As it could be reproducibly demonstrated by the MS2-tethering assays, all hnRNP D isoforms interfere with exon recognition, when tethered to the MS2-loops placed within exon 3. In order to verify that binding of hnRNP D to the ESSV apart from MS2-tethering results in interference with exon 3 recognition, the distinct hnRNP D isoforms were ectopically overexpressed in presence of the HIV-1 derived minigene LTR 4 exon ESSV (Fig. 3.3). Ectopic overexpression was achieved using a pcDNA3.1 derived vector set, expressing the proteins of interest with an N-terminal extension containing a Flag-tag linked by four glycine residues to the NLS of HIV-1 Tat protein. Transfection of HEK-293T cells with the corresponding expression vectors on two consecutive days resulted in a high expression of all overexpressed proteins when compared with mock expression using non-coding pcDNA3.1 vector (Fig. 3.3 C cf. lanes 2-7 with lane 1). As anticipated, RT-PCR based analysis of LTR 4 exon ESSV pre-mRNA splicing showed no effect of mock expression on the splicing pattern of the HIV-1 derived minigene, when compared with the previously obtained data (cf. Fig. 3.1 I B lane 1 with Fig. 3.3 B lane 1). In line with MS2-tethering experiments, ectopic overexpression of hnRNP D isoforms lacking their C-termini (ΔCT) comparable to mock expression failed to exhibit any effect on the splicing pattern of LTR 4 exon ESSV splicing-reporter (Fig. 3.3 B cf. lanes 6-7 with lane 1). Notably, since the lack of the C-terminal located

NLS of hnRNP D [43, 184, 208] was compensated by the N-terminal incorporation of the Tat NLS, a defect in nuclear localisation could be excluded to be responsible for the observed lack of interference. Thus, when taken together with finding that the C terminus was shown not to the affect RNA-binding affinity of hnRNP D [53], the observed lack of interference further confirms the necessity of the C terminus for hnRNP D-mediated interference with exon recognition. As expected, when compared to C-terminal deleted isoforms, ectopic overexpression of the full-length hnRNP D isoforms dramatically changed the splicing pattern obtained with LTR 4 exon ESSV minigene (Fig. 3.3 B cf. lanes 6-7 with lanes 2-5). However, instead of the anticipated shift from the exon 3 including mRNA-isoforms (tat3, vpr) to tat1 mRNA excluding exon 3 and exon 2, tat1 mRNA and the intron-containing vpr-mRNA (vpr) appeared to be decreased, while increased tat2, tat3 and tat4 mRNA isoforms seemed to be increased (Fig. 3.3 B lanes 2-5).
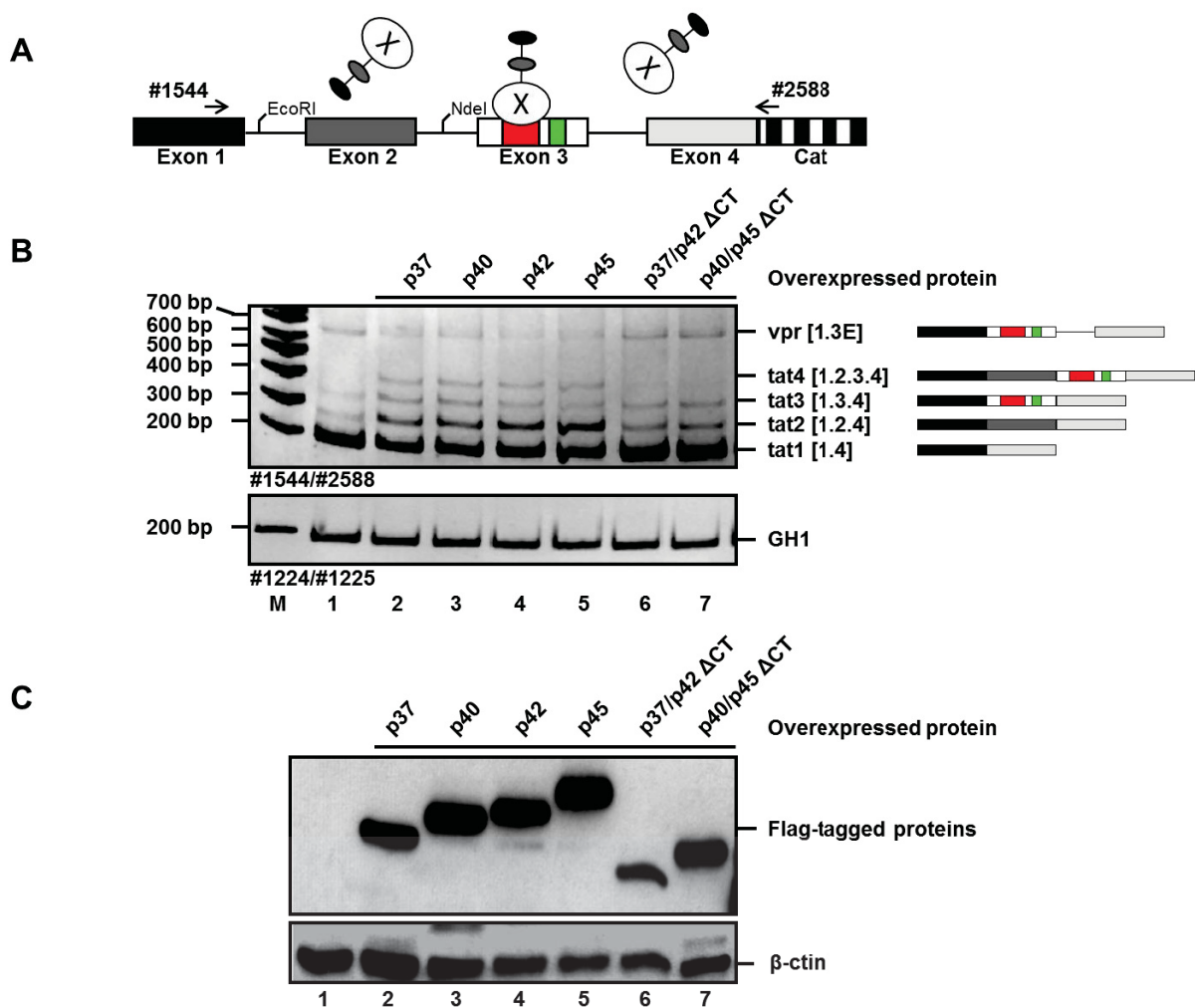
Fig. 3.3: Ectopic overexpression of hnRNP D isoforms suggests at least one additional hnRNP D binding site.

(A) Schematic diagram of the splicing reporter LTR 4 exon bound by ectopic overexpressed target proteins. Overexpressed proteins were depicted containing N-terminal Flag-tag (black), followed by the NLS of HIV-1 Tat protein (grey) and target protein ("X"). PCR primers were depicted as arrows. (B,C) Ectopic overexpression experiment. 2.5×10^5 HEK-293T cells were seeded in 6 well plates. The following day cells were transfected with 2µg of the designated pcDNA3.1 Flag-NLS vector or pcDNA3.1 respectively, to establish pre-expression. At the second day 2 µg of the designated pcDNA3.1 Flag-NLS vector or pcDNA3.1, 1µg of the splicing reporter LTR 4 exon ESSV, 1µg pxGH5 and 0.2µg SVctat were transfected. At the third day cells were washed off their wells, washed with PBS and transferred into two equal aliquots. (B) Using one aliquot total RNA was isolated, DNase I digested and reverse transcribed. The resulting cDNA template was amplified by semi-quantitative PCR using the reporter specific primer set #1544/#2588. Reassembling the transfection control, the expression of the human growth hormone (GH1) mRNA by pxGH5 was monitored using exon-junction specific primer set #1224/#1225 resulting in PCR-fragment of 186nt. Reporter specific splice isoforms were termed in reference to Purcell and Martin [173] and schematically illustrated on the right side of this panel. The PCR-products of the designated viral splice isoforms obtained by using the designated primer sets were depicted within the supplementary material (Fig. 6.2 I). (C) Total protein lysates originating out of the second cell aliquot were subjected to discontinuous SDS-PAGE and western blotting. Immunoblot analysis was carried out using anti-Flag and anti-β-actin antibodies.

As anticipated, the decrease in intron-containing vpr mRNA indicates overexpression induced enhancement of hnRNP D binding to the ESSV subsequently resulting in increased interference with exon 3 recognition. However, simultaneously overexpression of hnRNP D appeared to increase exon 2 and exon 3 containing tat2, tat3 and tat4 mRNA, while tat1 mRNA excluding exon 2 and exon 3 appeared to be decreased. Thus, besides the observed ESSV-mediated interference with exon 3 recognition, triggered by the overexpression another hnRNP D dependent, but yet unknown SRE might somehow result in preferentially increased inclusion of HIV-1 exon 2 and exon 3. In summary, ectopic overexpression of C-terminal truncated hnRNP D isoforms confirmed the contribution of the C-terminus to hnRNP D-mediated regulation of pre-mRNA splicing. Moreover, the observed complex splicing pattern indicates the presence of another, yet unknown hnRNP D dependent SRE within LTR 4 exon ESSV minigene.

3.4 An hnRNP D high affinity binding site could functionally replace ESSV

Since the overexpression experiment failed to ultimately prove the contribution of hnRNP D binding to ESSV-mediated interference with the recognition of HIV-1 exon 3, sequence analysis was employed, to identify mutations that specifically diminish either hnRNP A1 or hnRNP D binding (Fig 3.4 I). In this line, analysis of the ESSV sequence revealed the four prominent TAG motifs known to be bound by hnRNP A1 [13, 59, 134]. Moreover, RNA-affinity chromatography experiments showed that mutations inactivating the TAG motifs (ESSV⁻) resulted in diminished binding of both hnRNPs indicating overlapping binding sites of hnRNP A1 and hnRNP D. Additionally, the consensus motif resulting from the alignment of the four TAG motifs displayed a striking similarity to the vertebrate telomere repeat, characterised by TTAGGG sequence repeats published to be bound by hnRNP D [65, 98, 102]. However, so far every attempt to specifically diminish binding of either hnRNP A1 or hnRNP D failed (data not shown).
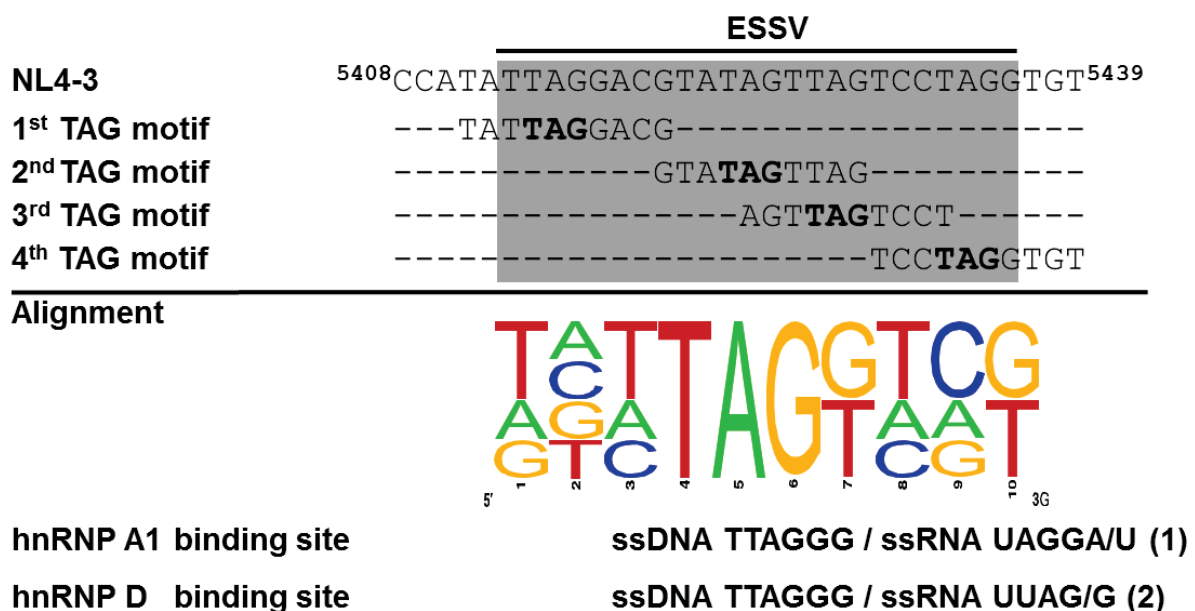
Fig 3.4 I: Sequence analysis of putative hnRNP D binding sites within the ESSV sequence.

In order to identify a consensus motif out of all TAG-containing sequence elements within the ESSV, 10nt of the sequence environment embedding each motif was collected and aligned. Below the alignment binding sites for hnRNP A1 and hnRNP D were listed: (1) indicating hnRNP A1 binding site published by MacKay and Cooke (1992) [147], Ishikawa et al. (1993) [98], Burd and Dreyfuss (1994) [28], Abdul-Manan and Williams (1996) [2], Abdul-Manan et al. (1996) [1], Erlitzki and Fry (1997) [70], Mayeda et al. (1998) [146] and Ding et al. (1999) [55]; (2) indicating hnRNP D binding sites published by Ishikawa et al. (1993) [98], Nagata et al. (1999) [153], Katahira et al. (2001) [109] and Enokizono et al. (2005) [65].

Since discrimination of hnRNP A1 and D binding by site directed mutagenesis of the ESSV sequence failed, literature was screened for already published hnRNP A1 and D binding sites. In this line of thinking, binding sites were requested that resemble the genuine ESSV bound by hnRNP A1 and hnRNP D and a binding site bound exclusively by hnRNP D (Fig. 3.4 II).
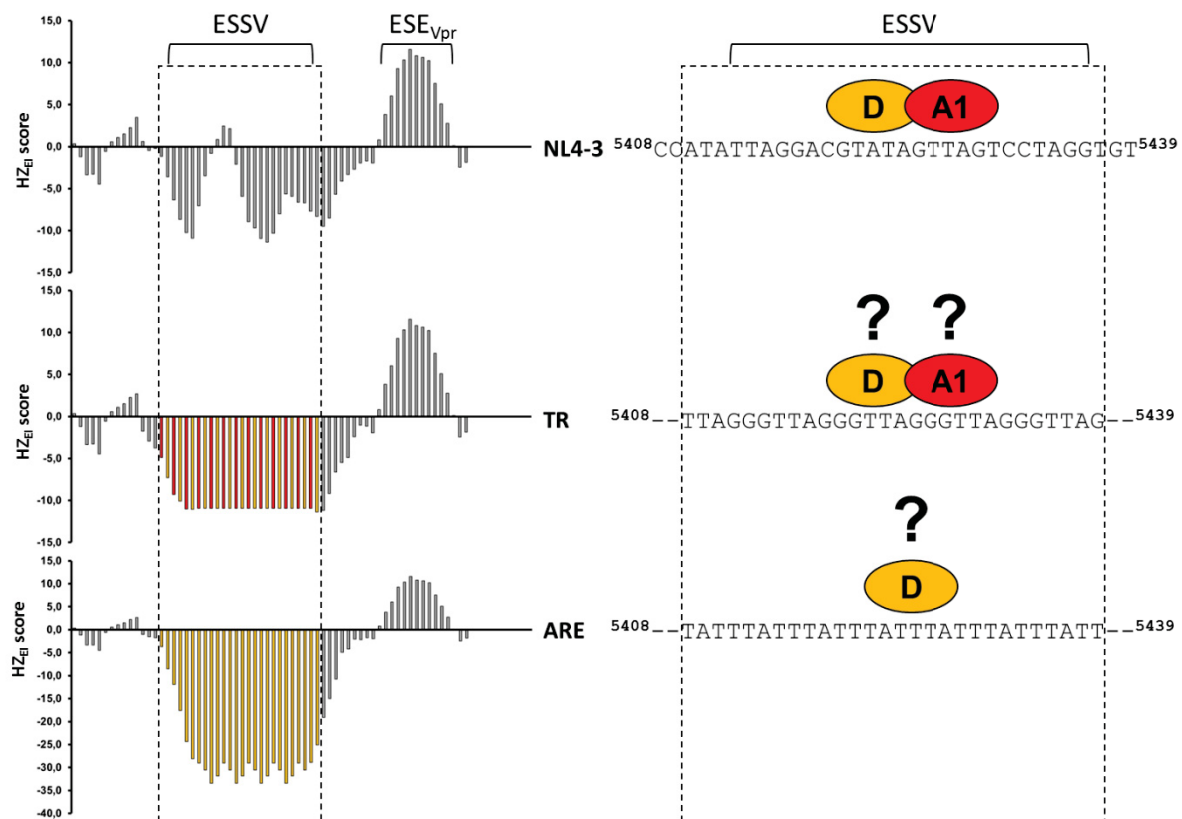
Fig. 3.4 II: Hexplorer analysis of TR and ARE sequences generated to replace the ESSV.

The ESSV together with both sequences generated to replace it were depicted on the left side of the panel encompassing the base sequence 5408 – 5439 in reference to Hxb2 provirus derived base numeration. The relative position of the 16nt ESSV is indicated as continuous lined bracket, whereas the actually replaced sequence is highlighted by a box comprising dotted frame lines. By that manner the sequences derived from the minigenes LTR 4 exon ESSV (ESSV), LTR 4 exon TR (TR) and LTR 4 exon ARE (ARE) were displayed with their published *trans*-acting factors hnRNP A1 (red circle) and hnRNP D (orange circle). The complete HIV-1 exon 3 sequence of the designated LTR 4 exon minigenes was subjected to analysis using Hexplorer algorithm [69] displayed on the right site of the panel. The Hexplorer algorithm based analysis assigns to each nucleotide of the analysed sequence a specific score value, the Hexplorer Z_{EI} (HZ_{EI}) score. The obtained HZ_{EI} score resembles the probability by which the occurrence of the six individual hexamers overlapping with the index nucleotide happens to be overrepresented either within exons (positive scores) or introns (negative scores). Since the definition of exon and introns apart of the 3'ss, the 5'ss and the branch point sequence also relies on the contribution of specific *trans*-acting factors bound to those overrepresented *cis*-acting elements, the individual algebraic sign and the size of HZ_{EI} score interval encompassing a specific nucleotide sequence allows a prediction upon the capacity of that sequence either to enhance or silence mRNA splicing. Since Hexplorer based analysis of TR and ARE sequence displays both a negative HZ_{EI} score interval similar to that of the ESSV, both sequences can presumed to silence mRNA splicing.

Fitting to the requirements, recent literature described specific binding of hnRNP A1 and hnRNP D to the "TTAGGG" sequence known as the telomere repeat (TR) [98] and exclusive binding of hnRNP D to "AUUUA" sequence commonly found in AREs (ARE) [243]. Repeats of both sequences were designed to replace the ESSV, while maintaining the genuine length of HIV-1 exon 3 and the precise location of the SRE within the exon. In order to gain a first idea of the splicing regulatory capacity of these sequences, Hexplorer analysis [69] was employed to predict the effect of TR and ARE sequences replacing the ESSV within HIV-1 exon 3 (Fig. 3.4 II). Thereby, the algorithm used by the Hexplorer assigns to each nucleotide of the analysed sequence a specific score value, the Hexplorer Z_{EI} (HZ_{EI}) score. The individual score value resembles a measure for the probability by which the six individual hexamers overlapping with the index nucleotide were overrepresented either within exons (positive scores) or introns (negative scores) belonging to original dataset used to develop the algorithm. Since the definition of exon and introns apart of the 3'ss, the 5'ss and the branch point sequence also rely on the contribution of specific *trans*-acting factors bound to those overrepresented *cis*-acting elements, the individual algebraic sign and the size of HZ_{EI} score interval encompassing a specific nucleotide sequence allows a prediction upon the capacity

of that sequence either to enhance or silence mRNA splicing. As expected a Hexplorer based analysis assigned a negative score interval to ESSV sequence known to interfere with exon recognition and a positive score interval to the ESE_{vpr} shown to increase the recognition of the exon. Interestingly, similar to the ESSV Hexplorer based sequence analysis assigned a negative score interval not only to TR, presumably bound by hnRNP A1 and D, but also to ARE sequence published to exclusively bind hnRNP D.

In order to complete the experimental set up a replacement for $ESSV^-$ was needed that in contrast to TR sequence is bound only by low levels of hnRNP A1 and hnRNP D. Therefore, similar in length to the TR and ARE sequence, a sequence containing "CCAAACAA" repeats was designed (Neutral), which was published by Zhang and co-workers [244] to exhibit no effect on mRNA splicing and thus presumably induce excessive splicing. Though, prior to assessing the effects of the inserted sequences on mRNA splicing, the assumed binding properties of the generated sequences had to be validated by RNA affinity chromatography based analysis (Fig. 3.4 III).
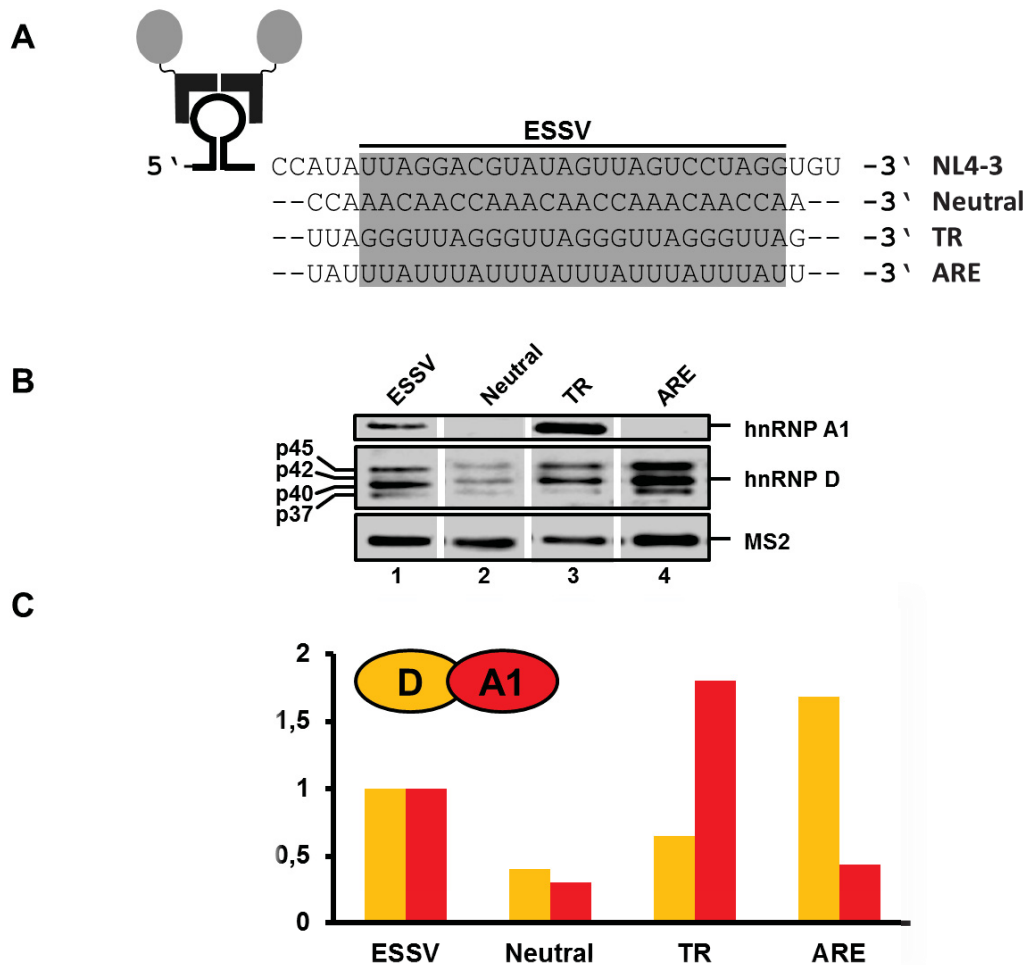
Results

Fig. 3.4 III: Binding profiles of Neutral, TR and ARE sequence in comparison to the ESSV

(A) Design of RNA affinity chromatography probes. RNA oligonucleotides were reverse-transcribed using commercial available RiboMAX™ Kit (Promega). All oligonucleotides were designed to contain an MS2-loop at their 5'-end, allowing binding to MS2-MBP fusion protein serving as loading control. Downstream of the MS2-loop the target sequence encompassing the ESSV (grey) was placed. (B) RNA affinity chromatography approach with *post hoc* immunoblot analysis. 3'-end standing ribose of RNA oligonucleotides were oxidized using periodate and linked to adipic acid dihydrazide coated agarose beads. Immobilized RNA oligonucleotides were incubated in the presence of HeLa nuclear extract supplemented with recombinant MS2-MBP fusion protein. Bound proteins were subjected to stringent affinity washes and eluted. Following separation of the eluted proteins by SDS-PAGE and western blotting, immunoblot analysis was carried out using antibodies against hnRNP A1, hnRNP D. For reasons of comprehensibility, lanes of the same gel were rearranged and the resulting transitions were represented by gaps. (C) Band Intensity readout. As an indicator of hnRNP binding the relative intensity of all bands was measured using ImageJ software. The measured intensities of all samples were normalized in reference to the MS2 loading control. Normalized band intensities were depicted relative to the ESSV sequence.

As demonstrated before, ESSV sequence could be shown to be bound by both hnRNPs, hnRNP A1 and hnRNP D (cf. Fig. 3.1 IV B lane 1 with Fig. 3.4 III BC lane 1). As anticipated, in comparison to the ESSV Neutral sequence featuring the CCAAACAA repeats exhibited only low binding of hnRNP A1 and D (Fig. 3.4 III BC cf. lane 2 with lane 1). Further, similar to the ESSV TR sequence appeared to be bound by both hnRNPs, albeit more abundantly by hnRNP A1 (Fig. 3.4 III BC cf. lane 3 with lane 1). Finally, ARE sequence appeared to be bound more abundantly by hnRNP D than the ESSV (Fig. 3.4 III BC cf. lane 4 with lane 1). Whereas, comparable to Neutral sequence binding of hnRNP A1 to ARE sequence appeared to be low (Fig. 3.4 III BC cf. lane 4 with lane 2) indicating almost exclusive binding of hnRNP D. In summary, since the replacement of the ESSV by the TR, ARE and Neutral sequence yielded in sequences bound by either hnRNP A1 and D, only hnRNP D or none of both hnRNPs, the experimental set up to evaluate contribution of hnRNP D on the recognition of HIV-1 exon 3 was complete. Since the binding properties of the generated sequences were confirmed by RNA affinity chromatography analysis, all designed sequences were transferred into the LTR 4 exon minigene yielding in the minigenes LTR 4 exon TR (TR), LTR 4 exon ARE (ARE) and LTR 4 exon Neutral (Neutral). In order to assess the splicing regulatory effect of the inserted sequences, HeLa cells were transfected with LTR 4 exon ESSV functioning as control and the new generated minigenes (Fig.3.4 IV A). In line with previous results, transfection of LTR 4

exon ESSV minigene (ESSV) showed predominantly exclusion of exon2 and exon 3 (tat1 mRNA) indicating the contribution of hnRNP A1 and D to ESSV-mediated interference (cf. Fig. 3.1 I B lane 1 with Fig. 3.4 IV lane 1). As anticipated by the observed low binding of hnRNP A1 and hnRNP D to Neutral sequence (Fig. 3.4 III BC lane 2), transfection of LTR 4 exon Neutral minigene (Neutral) showed predominant inclusion of HIV-1 exon 3 (tat3, vpr) (Fig. 3.4 IV B lane 2) indicating excessive recognition of HIV-1 exon 3. In contrast to the LTR 4 exon Neutral minigene demonstrating the excessive splicing phenotype, transfection of LTR 4 exon TR minigene showed an ESSV-like splicing pattern (Fig. 3.4 IV B cf. lane 3 to lane 1). Though notably, in contrast to ESSV more abundant binding of hnRNP A1 to TR sequence appeared to result in an almost complete loss of tat3 mRNA reflecting the in comparisons to the ESSV more abundant binding of hnRNP A1 (Fig. 3.4 III BC lane 3). Interestingly, transfection of LTR 4 exon ARE minigene resulted in an ESSV-like splicing pattern comparable to that of LTR 4 exon TR (Fig. 3.4 IV B cf. lane 4 with lane 3).
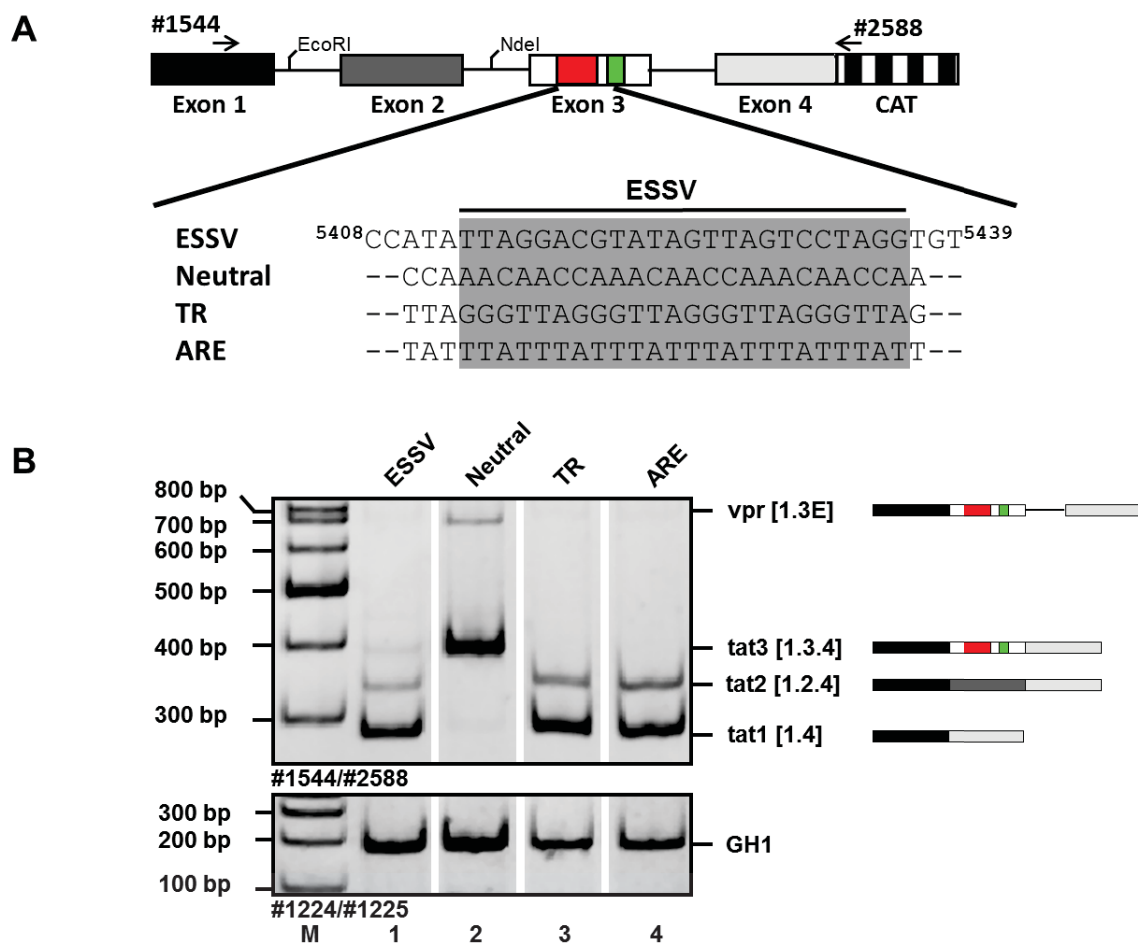
Fig. 3.4 IV: An hnRNP D exclusive binding site is sufficient to replace ESSV.

(A) Schematic drawing of HIV-1 derived minigene LTR 4 exon containing the first four exons of HIV-1 NL4-3 with a truncated, but functional version of intron 1 [227]. The SREs within HIV-1 exon 3, the ESSV and ESE$_{vpr}$, are coloured in red and light green. The splicing pattern of the reporter was assessed by semi-quantitative RT-PCR using the primer set #1544/#2588 (arrows). Nucleotide positions were designated in reference to HIV-1 Hxb2 strain. (B) Transfection of minigenes. $2.5 \cdot 10^5$ HeLa cells were seeded in 6 well plates and transfected with 1µg of the designated splice reporter LTR 4 exon, 1µg pxGH5 and 0.2µg SVctat. 24h post transfection total RNA was isolated, DNase I digested and reverse transcribed. The resulting cDNA template was amplified by semi-quantitative PCR using the reporter specific primer set #1544/#2588. Reassembling the transfection control, the expression of the human growth hormone (GH1) mRNA by pxGH5 was monitored using exon-junction specific primer set #1224/#1225 resulting in PCR-fragment of 186nt. Reporter specific splice isoforms were termed in reference to Purcell and Martin [173] and schematically illustrated on the right side of this panel. The PCR-products of the designated viral splice isoforms obtained by using the designated primer sets were depicted within the supplementary material (Fig. 6.2 I). For reasons of comprehensibility, lanes of the same gel were rearranged and the resulting transitions were represented by gaps. Lanes containing DNA-marker were highlighted (M).

However importantly, in contrast to TR sequence the ARE sequence was shown almost exclusively to be bound by hnRNP D (Fig. 3.4 III BC cf. lane 4 with lane 3) providing the evidence that the observed interference with exon 3 recognition could only be provided by hnRNP D. Thus finally, as previously indicated by the hnRNP A1 knock out experiment and the MS2-tethering assays, binding of hnRNP D along with hnRNP A1 to the ESSV indeed contributes to the interference with HIV-1 exon 3 recognition.

3.5 HnRNP D binding sequence permits efficient HIV-1 particle production

Hitherto within this thesis, hnRNP D could be shown to resemble a splicing regulatory protein, which when bound to an exonic position results in interference with exon recognition. In that line, using HIV-1 derived LTR 4 exon minigene hnRNP D together with hnRNP A1 could be shown to bind to the ESSV and thus contribute to its interference with the recognition of HIV-1 exon 3. In order to verify that hnRNP D can contribute to HIV-1 virus particle production by binding to the ESSV, the established binding site variants were transferred from the LTR 4 exon minigenes to pNL4-3 proviral vectors correspondingly termed pNL4-3 ESSV (ESSV), pNL4-3 TR (TR), pNL4-3 ARE (ARE) and pNL4-3 Neutral (Neutral).

Results

After generation of the required proviral vectors, HIV-1 particle production was assessed by transfection of HEK-293T producer cell line with the proviral vectors. 48 hours post transfection splicing of HIV-1 pre-mRNA was analysed using specific primer sets aiming to cover a wide range of HIV-1 transcripts (Fig. 3.5 I). Three primers sets were chosen to amplify *tat* and *vpr* mRNAs (#1544/#3632), mRNA isoforms of the 2kb class (#1544/3932) and 4Kb class (#1544/#640). Consistently with the results obtained from transfection of the LTR 4 exon ESSV minigene, transfection of the corresponding proviral vector pNL4.3 ESSV resulted primarily in exon 2 and exon 3 excluding mRNA isoforms (tat1, nef2, rev1+2, env1) and low levels of exon 3 including transcript variants (tat3, vpr, nef4, env5) (Fig. 3.5 I B lane 1). Though notably, in comparison to the LTR 4 exon ESSV minigene the recognition of HIV-1 exon 3 resulted more frequently in splicing at 3'ss A2, but retention of intron 3 (vpr mRNA) than in inclusion of HIV-1 exon 3 by splicing at 3'ss A2 and 5'ss D3 (tat3 mRNA). (cf. Fig. 3.4 IV B lane 1 with Fig. 3.5 I B lane 1).
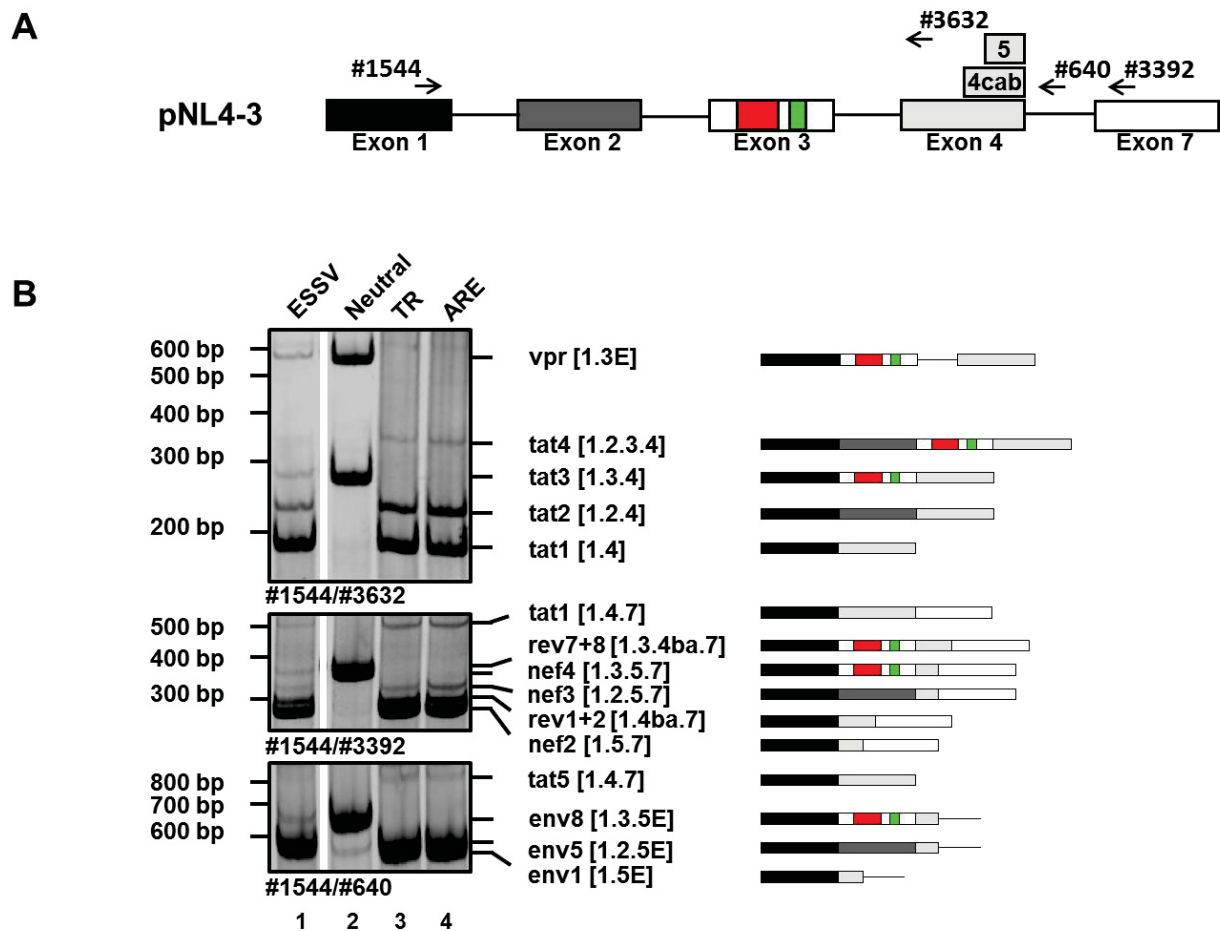
Fig. 3.5 I: HIV-1 NL4-3 ESSV can functionally be replaced by the exclusive hnRNP D binding sequence.

(A) Schematic drawing of HIV-1 NL4-3 proviral DNA. The SREs within HIV-1 exon 3, the ESSV and ESE$_{vpr}$ are coloured in red and light green. The used primer sets to assess the splicing pattern of the viral primary transcript are depicted as arrows. (B) Proviral transfection experiment. 2.5 x 10^5 HEK-293T cells were seeded in 6 well plates and transfected with 3µg of the designated pNL4-3 vector. 48 h post transfection total RNA was isolated, DNase I digested and reverse transcribed. The resulting cDNA template was amplified by semi-quantitative PCR using HIV-1 specific primer sets covering different mRNA subclasses (see A). Viral splice isoforms were termed in reference to Purcell and Martin [173] and schematically illustrated on the right side of this panel. The PCR-products of the designated viral splice isoforms obtained by using the designated primer sets were depicted within the supplementary material (Fig. 6.2 II). For reasons of comprehensibility, lanes of the same gel were rearranged and the resulting transitions were represented by gaps.

As anticipated, similar to transfection of LTR 4 exon Neutral minigene transfection of pNL4-3 Neutral led predominantly to exon 3 including mRNA isoforms (tat3, vpr, nef4, rev7+8, env5) (cf. Fig. 3.4 IV B lane 2 with Fig. 3.5 I B lane 2) indicating excessive recognition of HIV-1 exon 3 due to low hnRNP D and hnRNP A1 binding levels. Comparable to the corresponding minigenes, in comparison to pNL4-3 Neutral transfections of pNL4-3 TR and pNL4-3 ARE resulted in a shift from predominantly exon 3 including mRNA isoforms (tat3, vpr, nef4, rev7+8, env5) to dominant exclusion of exon 3 (tat1, nef2, rev1+2, env1) (Fig. 3.5 I B cf. lanes 3-4 with lane 2). Thus, apparently binding of hnRNP D and hnRNP A1 to TR sequence as well as exclusive binding of hnRNP D to ARE sequence resulted in interference with exon 3 recognition providing further evidence for the contribution of hnRNP D to ESSV-mediated regulation of HIV-1 splicing. Moreover, the in contrast to pNL4-3 ESSV observed complete loss of exon 3 including mRNA isoforms (tat3, vpr3, nef4, env8) with pNL4-3 TR and pNL4-3 ARE indicated a slightly higher capacity of TR and ARE sequence to interfere with the recognition of HIV-1 exon 3 (Fig. 3.5 I B cf. lanes 3-4 with lane 1). Altogether, in confirmation with the preceding results obtained by transfection of the LTR 4 exon minigenes, transfection of pNL4-3 proviral vectors could verify that exclusive binding of hnRNP D to ARE sequence sufficiently sustains a splicing pattern comparable to that mediated by the ESSV bound by hnRNP A1 and hnRNP D.
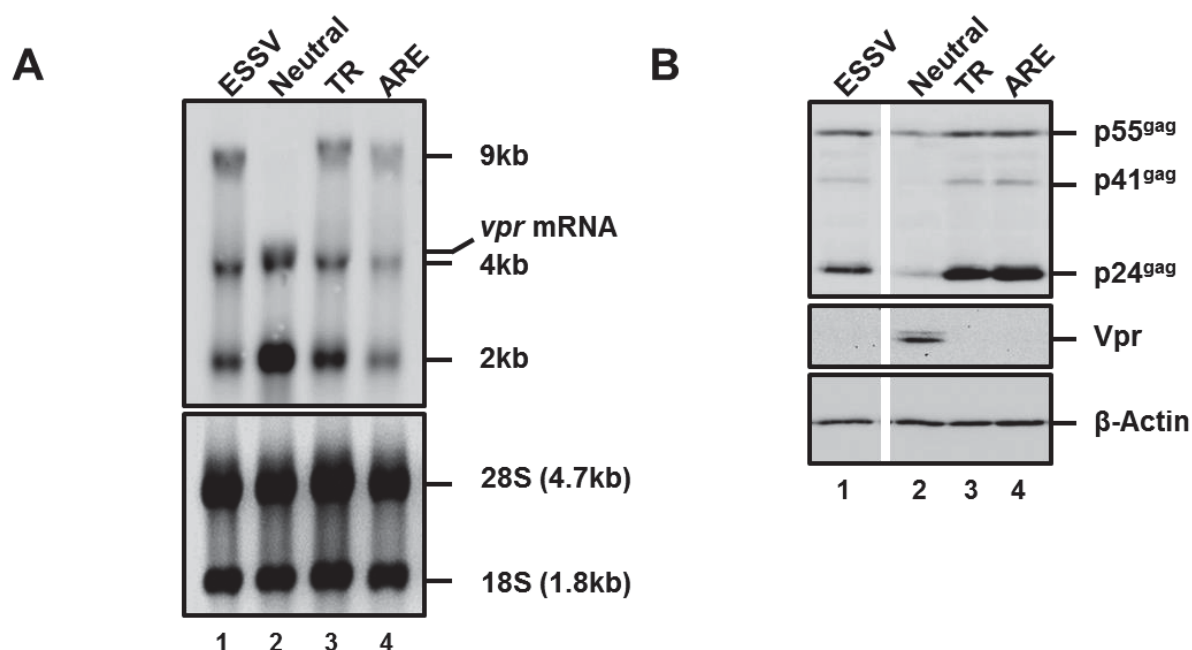
Fig. 3.5 II: HnRNP D binding within HIV-1 exon 3 sustains HIV-1 particle production.

(A) Proviral transfection experiment. 2.5 x 10^5 HEK-293T cells were seeded in 6 well plates and transfected with 3µg of the designated proviral vector pNL4-3 X. 48 h post transfection total RNA was isolated and subjected to agarose gel electrophoresis. Equal RNA loading was verified by ethidium bromide staining of ribosomal RNAs. After capillary blotting, the northern blot was probed with a DNA probe specific for HIV-1 exon 7. (B) Proviral transfection experiment. 2.5·10^5 HEK-293T cells were seeded in 6 well plates and transfected with 3µg of the designated proviral vector pNL4-3 X. 48 h post transfection cell-lysis was carried out. After lysis, total protein lysates were subjected to discontinuous SDS-PAGE and western blotting. Immunoblot analysis was carried out using anti-HIV-1 p24, anti-HIV-1 Vpr and anti-β-actin antibodies. For reasons of comprehensibility, lanes of the same gel were rearranged and the resulting transitions were represented by gaps.

Since exclusive binding of hnRNP D to the ARE sequence sufficiently maintained an ESSV-like splice pattern, the efficiency of HIV-1 particle production was probed by assessing the abundance of viral RNA classes by northern blotting (Fig.3.5 II A) and the HIV-1 Gag protein by western blotting (Fig. 3.5 II B). Transfection of pNL4-3 ESSV provirus displayed the three viral RNA classes, the unspliced 9Kb class, the single spliced 4Kb class and the multiple spliced 2Kb class (Fig. 3.5 II A lane 1). Analysis of HIV-1 Gag protein by immunoblotting using HIV-1 p24^{gag} specific antibodies showed the expected p24^{gag} containing proteins, the gag-precursor p55^{gag} and p41^{gag} and the completely processed p24^{gag} (Fig. 3.5 II B lane 1). Moreover, immunoblot analysis showed a higher abundance of completely processed p24^{gag} than that of both precursor proteins indicating efficient posttranslational processing of HIV-1

Gag protein. Consistent with previous experiments showing excessive recognition of HIV-1 exon 3 leading to predominant inclusion of exon 3 within HIV-1 transcripts, transfection of pNL4-3 Neutral resulted in an increase in exon 3 containing *vpr* mRNA isoforms migrating above the 4Kb class (Fig. 3.5 II A cf. lane 2 with lane 1) and correspondingly detectable levels of Vpr protein (Fig. 3.5 II B cf. lane 2 with lane 1). However importantly, excessive recognition of exon 3 also resulted in a total loss of the 9Kb RNA, while the abundance of the 2Kb RNA was dramatically increased (Fig. 3.5 II cf. lane 2 with lane 1). Consistent with the detected loss of 9Kb RNA, immunoblot analysis showed a strong decrease in p24^{gag} protein (Fig. 3.5 II B cf. lane 2 with lane 1). Further, the abundance of p24^{gag} containing precursor proteins could be shown to be higher than that of completely processed p24^{gag} indicating a defect in procession of the precursor proteins. Altogether, the loss of hnRNP D and A1 binding to Neutral sequence resulting in excessive recognition of HIV-1 exon 3 led to the loss of viral 9Kb RNA, decreased HIV-1 Gag expression and a procession defect of the Gag precursor protein, thus indicating a severe impairment of HIV-1 particle production. In contrast to pNL4-3 Neutral, transfection of pNL4-3 TR or pNL4-3 ARE in line with the previously detected splicing patterns resulted in expression levels of the three viral RNA classes comparable to pNL4-3 ESSV (Fig. 3.5 II cf. lanes 3-4 with lane 1). Although, transfection of pNL4-3 TR, as well as pNL4-3 ARE appeared to result in a slight increase of 2Kb class (Fig. 3.5 II cf. lanes 3-4 with lane 1) an overall increased gag expression could be detected (Fig. 3.5 II B cf. lanes 3-4 with lane 1) indicating efficient HIV-1 particle production. In summary, the phenotype observed with the Neutral sequence aligns perfectly with the published excessive splicing phenotype detrimental to HIV-1 replication [135, 136]. Moreover, since not only replacement of the Neutral sequence by the TR sequence bound by hnRNP A1 and hnRNP D, but also by ARE sequence exclusively bound by hnRNP D could functionally replace the ESSV in maintenance of efficient HIV-1 particle production, hnRNP D resembles an essential splicing regulatory protein regulating HIV-1 splicing and thus sustaining HIV-1 replication.

3.6   Complex formation of hnRNP D with hnRNP K

So far the experimental data presented within this thesis provided sufficient evidence that hnRNP D represents a splicing regulatory protein that when bound to an exonic position

interferes with exon recognition. Further, the obtained data by the MS2-tethering assay displayed that the interference mediated by all hnRNP D isoforms is strictly dependent on the presence of their C termini. Moreover, since in contrast to full-length hnRNP D isoforms p37 and p40 tethering of the shared C terminus alone failed to induce interference with exon recognition, the contribution of another, yet unidentified domain to the interference mediated by hnRNP D isoforms p37 and p40 could be supposed. However, the interference with exon recognition mediated by hnRNP D isoforms p42 and p45 could be mapped to the domain encoded by alternative exon 7 by using the MS2-tethering approach. In this line, recently the domain encoded by alternative exon 7 could be mapped to mediate the protein-protein interaction of hnRNP D isoform p42 and p45 with tristetraprolin (TTP) [111]. Thus, in order to pursue the notion that a protein-protein interaction mediated by the domain encoded by alternative exon 7 might contribute to interference with exon recognition, Flag-tagged hnRNP D p45 was overexpressed and immunoprecipitated to identify possible interaction partners (Fig. 3.6).
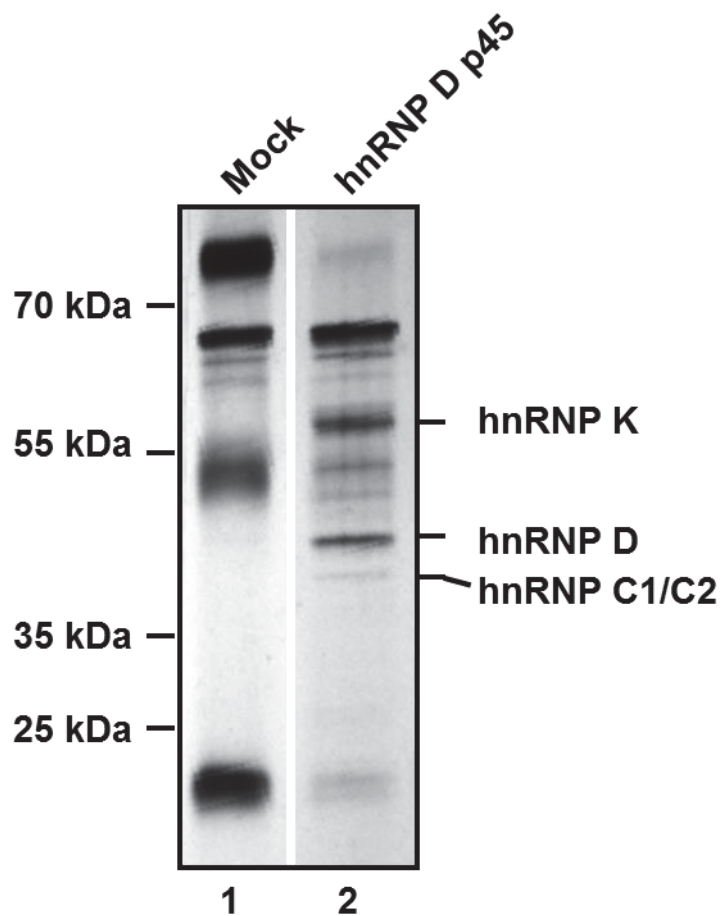
Fig. 3.6: HnRNP D forms a complex with hnRNP K.

7.5×10^6 HEK293T cells were seeded in 100mm tissue culture dishes and transfected on two consecutive days with 20μg of the designated pcDNA3.1 Flag-NLS expression vector or pcDNA3.1 (Mock) each and incubated overnight. Total protein lysates were subjected to primary immunoprecipitation using ANTI-FLAG® M2 antibody (Sigma) subsequently immobilized on Protein G Sepharose 4 Fast Flow. Precipitated proteins were eluted and separated by SDS-Page. SDS-PAGE gel was silver stained based on the published method by Heukeshoven [176]. Bands of interest were sliced out and send to the Molecular Proteomics Laboratory (MPL) of the BMFZ, Heinrich-Heine-University Düsseldorf. Samples were prepared by trypsin fragmentation and subsequently subjected to mass spectrometric analysis using ESI-MS spectrometer containing a 3D ion-trap (HCTultra PTM Discoverer System, Bruker Daltonics). Mascot score, sequence coverage and the sequences of the identified peptides were depicted within the supplementary material (Fig. 6.2 III). For reasons of comprehensibility, lanes of the same gel were rearranged and the resulting transitions were represented by gaps.

Ectopic overexpression of hnRNP D p45 was achieved by transfection of pcDNA3.1 Flag-NLS vector system on two consecutive days. The overexpressed hnRNP D isoform p45 containing N-terminal Flag-tag was immunoprecipitated using anti-flag antibody immobilized on protein g sepharose. Immunoprecipitated proteins were eluted, subjected to separation by SDS-PAGE and analysed subsequently by silver staining of the gel (Fig. 3.6). Protein identities were determined by MS-based analysis of the extracted and trypsin-treated protein fragments (Fig. 6.2 III). Notably, primary immunoprecipitation led to precipitation of three distinct proteins, hnRNP K, hnRNP D and hnRNP C (Fig. 3.6 cf. lane 2 with lane 1). Precipitation of both, hnRNP D and hnRNP C, was anticipated, since on one hand ectopically overexpressed hnRNP D isoform p45 was used for precipitation and on the other hand the cellular RNA-protein complexes were left intact during precipitation, thus facilitating precipitation of the major mRNA binding protein hnRNP C. However interestingly, besides hnRNP D and hnRNP C also large amounts of hnRNP K could be precipitated raising the question, if hnRNP K plays a role in hnRNP D-mediated interference with exon recognition.

3.7 Protein complex at ESSV contained hnRNP A1, hnRNP D and hnRNP K

Since recent literature indicated involvement of hnRNP K in HIV-1 mRNA splicing concerning HIV-1 tat exon 3 [139] and the primary immunoprecipitation of hnRNP D resulted in high amounts of precipitated hnRNP K, the question was raised, if hnRNP K is also present at the

ESSV. In order to address this question and to verify binding of hnRNP K, together with hnRNP A1 and D to HIV-1 ESSV sequence, RNA affinity chromatography was carried out (Fig. 3.7).

A

B
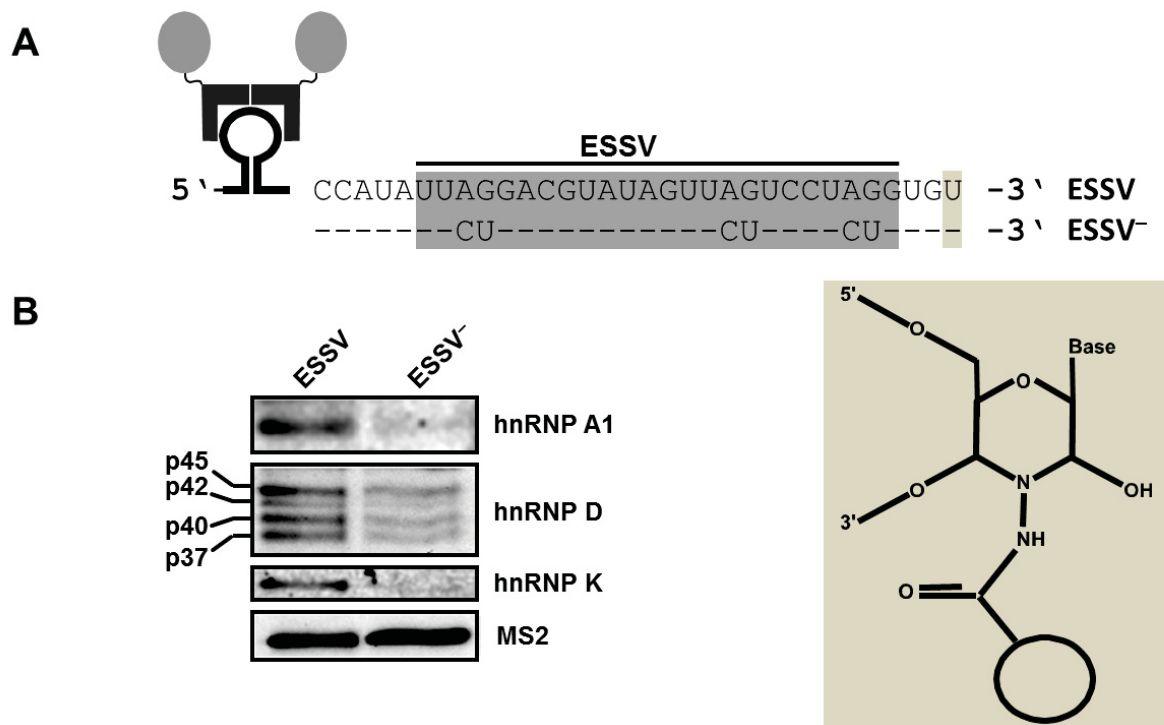
Fig. 3.7: hnRNP K bind only binding sequence is sufficient to replace ESSV.

(A) Design of RNA affinity chromatography probes. RNA oligonucleotides were reverse-transcribed using commercial available RiboMAX™ Kit (Promega). All oligonucleotides were designed to contain a MS2 loop at their 5'-end, allowing binding to MS2-MBP fusion protein serving as loading control. Downstream of the MS2-loop the target sequence encompassing the ESSV (grey) was placed. (B) RNA affinity chromatography approach with *post hoc* immunoblot analysis. 3'-end standing ribose of RNA oligonucleotides were oxidized using periodate and *post hoc* linked to adipic acid dihydrazide coated agarose beads. Immobilized RNA oligonucleotides were incubated in presence of HeLa nuclear extract supplemented with recombinant MS2-MBP fusion protein. Bound proteins were subjected to stringent affinity washes and eluted. Following separation of the eluted proteins by SDS-PAGE and *post hoc* western blotting, immunoblot analysis was carried out using antibodies against hnRNP A1, hnRNP D and hnRNP K.

As observed previously, hnRNP A1 and hnRNP D could be shown to bind to ESSV sequence and binding of both was lost with ESSV$^-$ sequence (Fig. 3.7 cf. lane 1 with lane 2). Interestingly, as indicated by the previous experiment showing the immunoprecipitation of hnRNP K by hnRNP D isoform p45 (Fig. 3.6 lane 2), hnRNP K could be pulled down by ESSV sequence along with hnRNP A1 and hnRNP D (Fig. 3.7 lane 1). Moreover, along with hnRNP

A1 and hnRNP D also hnRNP K binding showed to be decreased with ESSV$^-$ sequence (Fig 3.7 cf. lane 2 with lane 1). However, if hnRNP K binding beside hnRNP A1 and hnRNP D contributes to interference with the recognition of HIV-1 exon 3 remains to be addressed by future experiments.

4   Discussion

4.1   ESSV-mediated regulation of HIV-1 splicing is dependent on hnRNP D

Recognition of HIV-1 exon 3 and its 3'ss A2 rely on three determinants, the complementary of the 5'ss D3 to the U1 snRNA, and both splicing regulatory sequences, ESSV and ESE$_{vpr}$ [13, 59, 68]. The complementarity of the 5'ss D3 to U1 snRNA was shown to contribute to the recognition of exon 3 by regulating U1 snRNP binding that in turn supports exon definition [68]. Moreover, exon 3 recognition was shown to be dependent on the binding of Tra2 proteins to ESE$_{vpr}$ synergistically increasing with the 5'ss complementarity the recruitment of the U1 snRNP to 5'ss D3 [68]. However, the ESSV was shown to override to a large extent the established exon recognition resulting in low levels of transcripts including exon 3 and intron-containing *vpr* mRNA. Recent publications showed that binding of hnRNP A1 to the ESSV interferes with U2AF p65 binding to the PPT of 3'ss A2 [13, 59]. Moreover, the mutational inactivation of the ESSV was shown to result in aberrantly high levels of exon recognition, termed excessive splicing [13, 59, 134]. Thereby, the aberrantly high levels of exon recognition were described to induce a shift from intron-containing to intronless transcripts, which was shown to result in the complete loss of the unspliced viral 9Kb RNA, expression of HIV-1 Gag and Gag-Pol precursor protein and finally HIV-1 replication [68, 134, 135]. However, a recent study published by Jablonski and co-workers showed that hnRNP A1 knock out failed to induce excessive splicing [99], arguing for another splicing regulatory protein contributing together with hnRNP A1 to ESSV-mediated regulation of HIV-1 splicing. Analogously presented within this thesis, hnRNP A1 deficiency of CB3 cell line failed to induce excessive splicing of the HIV-1 minigene pre-mRNA. RNA affinity chromatography using ESSV containing RNA substrates led to the initial identification of hnRNP D as additional ESSV binding protein. Whereas, the inability to precipitate either hnRNP A1 or hnRNP D by RNA substrates containing a mutational inactivated ESSV indicated a role of hnRNP D in ESSV-mediated regulation of HIV-1 splicing. Additional evidence to that hypothesis was provided by MS2-tethering assay showing that tethering of all hnRNP D

isoforms displayed interference with exon recognition. Finally, substitution of the ESSV by ARE sequence, which consistently to recent publications was shown to be exclusively bound by hnRNP D [243], resulted in efficient HIV-1 particle production. Thus, the contribution of hnRNP D to ESSV-mediated regulation of HIV-1 mRNA splicing by hnRNP D could be assumed.

Further evidence to the role of hnRNP D in ESSV-mediated regulation of HIV-1 splicing was presented by a study of Lund and co-workers [130]. When using a HeLa cell line transduced with a replication inactive HIV-1 genome derivate, data of a simultaneous knock down of all hnRNP D isoforms revealed a slight increase of nef4 and env8 transcripts that both contain HIV-1 exon 3. Moreover, the hnRNP D knock down appeared to induce a decrease in the abundance of unspliced 9Kb RNA matched by a decrease in Gag protein levels. Notably, the authors attributed these findings completely to the loss of hnRNP D isoforms p42 and p45, which they supposed to contribute to the export of the 9Kb RNA. However, an alternative interpretation of these findings might be that siRNA-induced loss of hnRNP D binding to the ESSV results in increased, but due to the presence of hnRNP A1 not excessive recognition of HIV-1 exon 3, which in turn led to a decrease in 9Kb RNA and Gag protein levels. Finally, the importance of hnRNP D binding along with hnRNP A1 to the ESSV for HIV-1 replication was further underlined by data derived from two proteomic approaches [88, 165]. Therein, HIV-1 infection of monocyte derived macrophages was shown to induce the up regulation of hnRNP D expression along with the expression of hnRNP A1 further indicating the necessity for both proteins in regulation of HIV-1 mRNA splicing.

4.2 The mechanisms behind ESSV-mediated regulation of HIV-1 splicing

The recognition of HIV-1 exon 3 was recently described to be dependent on the efficiency of RNA-duplex formation of the U1 snRNA with 5'ss D3 and tra2 proteins bound to ESE$_{vpr}$, which synergistically enhance U1 snRNP binding [68] (Fig. 4.2 A). Recently, hnRNP A1 binding to the ESSV was shown to interfere with U2AF p65 binding to the PPT of 3'ss A2 resulting in interference with exon 3 recognition [13, 59]. Recent publications showed that binding of hnRNP A1 to HIV-1 ESS3 similarly to the ESSV resulted in interference with U2AF p65 binding

to the PPT of 3'ss A7 [159, 247]. Thereby the GRD of hnRNP A1 bound to a high affinity binding site was shown to enable cooperative binding of other hnRNP A1 proteins to adjacent low affinity binding sites, termed cooperative spreading. Thus, cooperative spreading of hnRNP A1 proteins in 3'-5' direction was proposed to result in displacement of U2AF p65 from the PPT of 3'ss A7. In that line, cooperative spreading of hnRNP A1 emanating from the ESSV in 3'-5' direction might also lead to displacement of U2AF p65 from the PPT of 3'ss A2 (Fig. 4.2 B). Interestingly, ESSV containing splicing substrates subjected to HeLa nuclear extracts were shown to form E-like complexes containing the U1 snRNP bound to its 5'ss [59]. Thus, while cooperative spreading of hnRNP A1 emanating from the ESSV in 3'-5' direction might lead to displacement of U2AF p65, spreading in 5'-3' direction might inhibit Tra2 binding to ESE$_{vpr}$, but not U1 snRNP binding to 5'ss D3. Additionally, analysis of ESSV binding proteins led to the identification of hnRNP K. However, since hnRNP K binding sites in contrast to the ESSV were described to be C-rich [143, 209, 212], hnRNP K might rather be recruited to the ESSV by protein-protein-interactions.
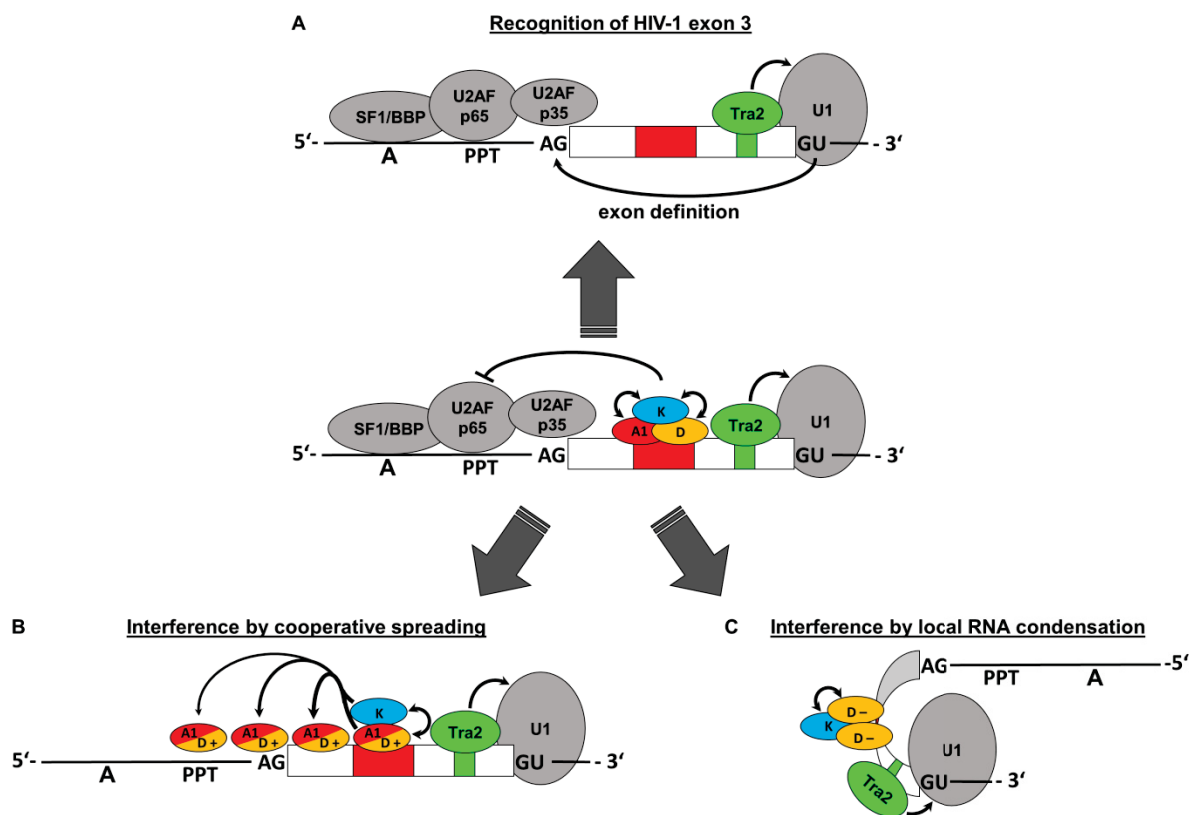
Discussion

Fig. 4.2: Model of HIV-1 exon 3 recognition.

Schematic drawing of the working model concerning the *cis-* and *trans-*acting factors that contribute to the regulation of HIV-1 exon 3 recognition. The middle panel depicts the putative complex containing all *cis-*acting elements bound by the corresponding *trans-*acting factors. The drawing represents a section of the viral pre-mRNA in 5'-3' direction encompassing HIV-1 exon 3 sequence (white box) and the flanking intronic sequence (black line). The depicted pre-mRNA contains the following *cis-*acting elements, the 3'ss containing the polypyrimidine tract (PPT) and the canonical AG-dinucleotide (AG), the 5'ss containing the canonical GU-dinucleotide (GU), and the branch-point site (A). The 5'ss D3 is bound by the U1 snRNP ("U1", grey sphere). The canonical AG-dinucleotide and the PPT are bound by U2-auxilliary factor p35 ("U2AF p35", grey sphere) and U2-auxilliary factor p65 ("U2AF p65", grey sphere). The BPS is bound by the splicing factor 1/mammalian branch point binding protein ("SF1/mBPP" grey sphere). The SREs located within HIV-1 exon 3, the exonic splicing silencer V (ESSV) and exonic splicing enhancer vpr (ESE$_{vpr}$), were depicted as red and green rectangles. The ESE$_{vpr}$ is bound by Tra2 proteins ("tra2", green sphere) which enhance U1 recruitment to the 5'ss (pointed arrow). The ESSV is bound to PPT by a complex containing hnRNP A1 ("A1", red sphere), hnRNP D ("D", orange sphere) and hnRNP K ("K", blue sphere). The bound complex interferes with U2AF p65 binding to the PPT (flat arrow). The interaction of hnRNP K with either hnRNP A1 or hnRNP D results in mutual enhanced recruitment of each factor to the complex (double pointed arrow). (A) Recognition of HIV-1 exon 3. Binding of Tra2 proteins to ESE$_{vpr}$ with the intrinsic strength of the 5'ss synergistically result in enhanced U1 snRNP recruitment (pointed arrow). Bridging interactions established by the U1 snRNP binding bound to the 5'ss lead to enhanced recruitment of U2AF p65 to the PPT, resulting in exon definition (pointed arrow). (B) Interference with recognition of HIV-1 exon 3 by cooperative spreading. Cooperative spreading is initiated either by hnRNP A1 or the exon 7 peptide containing hnRNP D isoforms p42 or p45 (D+) bound to the ESSV. Thereby the G-rich domain (GRD) of the bound hnRNP enables cooperative binding resulting in binding of another hnRNP protein to an adjacent low affinity binding site. Cooperative spreading in 5'-3' direction results in displacement of U2AF proteins from the 3'ss leading skipping of HIV-1 exon 3. Both, hnRNP A1 as well as hnRNP D isoforms p42 and p45, might recruit hnRNP K to the complex. Association of hnRNP K might result in mutually enhanced recruitment (double pointed arrow), and thus enhanced cooperative spreading. (C) Interference with recognition of HIV-1 exon 3 by local condensation of the RNA structure. The ESSV is bound by hnRNP D isoforms p37 or p40 (D-), which both contain a GRD lacking the peptide encoded by exon 7. Binding of multiple hnRNP D proteins to the ESSV might result in local condensation of the RNA structure. Recruitment of hnRNP K might result in mutually enhanced recruitment (double pointed arrow) leading to a stabilisation of hnRNP D binding and thus enhanced condensation of the RNA structure. The induced local condensation of the RNA structure might displace Tra2 proteins binding to ESE$_{vpr}$, but not binding of the U1 snRNP to the 5'ss. Thus, the observed interference with exon recognition by hnRNP D isoforms p37 and p40 might be rather due to prevention of the recognition of 3'ss A2. In this line, the induced local condensation of RNA structure might either interfere with the formation of the exon bridging complex linking 5'ss D3 with 3'ss A2 or directly displace U2AF p65 from the PPT.

Discussion

In this line recently a study showed a possible interaction of hnRNP K with hnRNP A1 large scale primary immune precipitation combined with MS identification [149]. Moreover, hnRNP K was shown to be associated with the hnRNP A1 binding sites ESE3/ESS3 and to a lesser extent the ISS [139]. Interestingly, association of hnRNP K instead of displacing hnRNP A1 was shown to further increase hnRNP A1 binding, which subsequently increased recruitment of hnRNP K to the complex. Thus, the association of hnRNP K with the ESSV might result in a positive feedback loop leading to an increased binding capacity of hnRNP A1 and by that way further enhance cooperative-spreading emanating from the ESSV (Fig. 4.2 B).

Apart from hnRNP A1, the experimental data presented within this thesis unequivocally provides a role of hnRNP D as a splicing regulatory protein in ESSV-mediated regulation of HIV-1 splicing. So far, MS2-tethering showed interference of all hnRNP D isoforms with exon recognition. Moreover, since deletion of the A-rich region failed to significantly diminish interference of all hnRNP D isoforms, dimerization mediated by that region [53] appears not to be necessary for hnRNP D-mediated interference. However, the mechanism, by which hnRNP D binding results in interference with exon recognition, remains speculative. MS2-tethering based domain mapping indicated that depending on the presence of the domain encoded by the alternative exon 7 hnRNP D binding might result in interference with exon recognition by two different mechanisms. In this line, the domain encoded by alternative exon 7 present in hnRNP D isoforms p42 and p45 could be mapped to exhibit interference with exon recognition. Thereby, the finding that the mapped domain encoded by alternative exon 7 was shown to facilitate cooperative binding of hnRNP D [249] indicated that hnRNP D might similar to hnRNP A1 induce cooperative spreading. Thus, the observed interference with exon recognition by hnRNP D isoforms p42 and p45 might be dependent on cooperative spreading in 3'-5' direction leading to the displacement of U2AF p65 bound to the PPT of 3'ss A2 (Fig. 4.2 B). Moreover, since ESSV containing splice substrates were described to form E-like complexes containing the U1 snRNP, but not U2AF p65 [59], similar to hnRNP A1 cooperative spreading by hnRNP D isoforms p42 and p45 might displace Tra2 proteins binding to ESE$_{vpr}$, but not the U1 snRNP binding to 5'ss D3. Additionally, since overexpression and primary immunoprecipitation of Flag-tagged hnRNP D isoform p45 resulted in precipitation of hnRNP K, a possible protein-protein interaction might result in recruitment

of hnRNP K. In accordance with this finding the same study identifying a possible interaction of hnRNP A1 with hnRNP K using large scale primary immune precipitation combined with MS identification, detected also a putative interaction of hnRNP K with hnRNP D [149]. Thus similar to hnRNP A1, the interaction of hnRNP D isoforms p42 and p45 with hnRNP K might result in recruitment of hnRNP K to the ESSV. Moreover, the ESSV-associated hnRNP K might enhance recruitment of further hnRNP D isoforms and thus further enhance hnRNP D-mediated cooperative spreading.

Besides hnRNP D isoform p42 and p45 containing the domain encoded by alternative exon 7 within their GRD, also hnRNP D isoforms p37 and p40 could be shown to interfere with exon recognition by MS2-tethering assay. A likely explanation for the interference observed with hnRNP D isoform p37 and p40 not containing the domain previously mapped to exhibit interference might be the recruitment of an hnRNP D isoform containing it. However, since the deletion of the A-rich region published to facilitate dimerization of hnRNP D isoforms failed to diminish interference exhibited by hnRNP D isoforms p37 and p40, the rescue of interference by dimerization with either hnRNP isoform p42 or p45 could be excluded. Thus, interference with exon recognition by hnRNP D isoforms p37 and p40 might be independent of the presence of the exon 7 encoded domain indicating another mechanism. Nevertheless, also deletion of the C terminus containing the GRD of hnRNP isoforms p37 and p40 was shown to diminish interference indicating at least a contribution of the GRD to the interference mechanism. However, since tethering of the GRD of hnRNP D isoforms p37 and p40 alone showed no interference with exon recognition, the interference mechanism appears to require another yet unidentified domain of hnRNP D isoforms p37 and p40. A possible link to the mechanism might be the finding that the formation of a tetrameric complex containing either hnRNP D isoform p37 or p40 resulted in local condensation of RNA structure in an RRM-independent manner around the consensus binding site [249, 250]. In this line, binding of hnRNP D p37 and p40 to the ESSV might induce local RNA condensation resulting in interference with exon recognition (Fig. 4.2 C). Noteworthy, since deletion of the A-rich region showed that hnRNP D-mediated interference might be independent of dimerization, the interference observed with hnRNP D isoforms p37 and p40 might rather result from simultaneous binding of multiple hnRNP D isoforms to the ESSV than formation of a tetrameric complex. Thereby, the resulting local RNA condensation

might block Tra2 protein binding to ESE_{vpr}, but not U1 snRNP binding to 5'ss D3, since RNA substrates containing the ESSV were shown to form E-like complexes featuring the U1 snRNP bound to the 5'ss, but lacking U2AF p65 binding to the PPT [59]. Thus, apart from displacement of the U1 snRNP from the 5'ss and thereby interfere with exon definition, the local condensation of the RNA structure might rather interfere with the recognition of 3'ss A2. In this line, the condensation of local RNA structure might either interfere with exon 3 recognition by inhibiting the formation of the exon bridging complex linking 5'ss D3 and 3'ss A2 or by direct displacement of U2AF p65 from the PPT of 3'ss A2. Finally, the interaction of hnRNP D isoforms p37 and p40 with hnRNP K resulting in mutually enhanced binding to the ESSV might further stabilise hnRNP D binding and thus result in enhanced local RNA condensation.

In line with the hypothesis of two mechanisms of interference dependent on the presence or absence of the exon 7 encoded peptide, a study by Li and co-workers described differential effects of hnRNP D isoforms binding to an SRE on the downstream located HPV16 5'ss SD_{3632} [127]. Interestingly, overexpression of hnRNP D isoforms p37 and p40 was shown to decrease the usage of 5'ss SD_{3632} resulting in low levels of L1 mRNA and the encoded protein. Whereas, overexpression of hnRNP D isoforms p42 or p45, respectively, led to increased usage of SD_{3632}, high levels of L1 mRNA and the encoded protein. Consistent with the working model, simultaneous binding of the overexpressed hnRNP D isoforms p37 or p40 to the two UAG-containing motifs within the published SRE might induce local condensation of the RNA structure. Since in contrast to HIV-1 exon 3, the SRE is located directly upstream of the 5'ss SD_{3632} the induced local condensation of the RNA structure might reduce the accessibility of the 5'ss SD_{3632} for U1 snRNP binding and thus result in decreased SD_{3632} usage. Interestingly, apart from decreasing the usage of SD_{3632} the overexpression of hnRNP D isoforms p37 and p40 was not reported to affect the recognition of the upstream located 3'ss SA_{3358}. This finding might be explained by the long distance of over 200nt between the SRE and 3'ss SD_{3358}, which compared to about 20nt between the ESSV and the 3'ss A2, might negated the effect of the induced local RNA condensation on exon recognition. Moreover, the observed reduction in 5'ss SD_{3632} usage indicating decreased U1 snRNP binding and thus decreased exon definition, might be compensated by

the recently discovered E4 ESE, which was shown to be indispensable for the recognition of the upstream located 3'ss SA_{3358} [183].

Notably, the overexpression of hnRNP D isoforms p42 and p45 was described to increase SD_{3632} usage. According to the working model, binding of the overexpressed hnRNP D isoforms p42 or p45 might result in cooperative spreading. Similar to HIV-1 exon 3 cooperative spreading from the SRE in 5'-3' direction might fail to displace the U1 snRNP and thus fail to decrease the usage of 5'ss SD_{3632}. Moreover, the observed increase in SD_{3632} usage might be due to the increased abundance of hnRNP D isoforms p42 or p45 interfering with binding of the endogenous hnRNP D isoforms p37 and p40 to the SRE and thus leading to increased usage of SD_{3632}. Interestingly, in contrast to HIV-1 exon 3 cooperative spreading emanating in 3'-5' direction from the SRE appeared not to affect the recognition of the upstream 3'ss SA_{3358}. Besides the long range of over 200nt to be covered by cooperative spreading, this discrepancy might be explained by the presence of a kind of roadblock represented by the *trans*-acting factor bound to E4 ESE located downstream of SA_{3358} [183]. In this line, Zhu and co-workers showed that an upstream SRE binding SRSF1 efficiently stops cooperative spreading of hnRNP A1 in 3'-5' direction [247]. Thus, E4 ESE binding protein might sterically block further procession of cooperative spreading of hnRNP D isoforms p42 or p45 and thereby .

Altogether, ESSV-mediated regulation of HIV-1 splicing is strictly dependent on hnRNP D and hnRNP A1. While hnRNP A1 binding interferes with exon recognition by cooperative spreading, the mechanism of hnRNP D-mediated interference might be dependent on the presence of the domain encoded by alternative exon 7. In the presence of exon 7 encoded domain the induction of cooperative binding might result in cooperative spreading by hnRNP D isoforms p42 and p45. Thereby, cooperative spreading might result in displacement of U2AF p65 from the upstream located PPT belonging to 3'ss A2 and thus interfere with exon 3 recognition. In contrast to hnRNP D isoforms p42 and p45, interference exhibited by hnRNP D isoforms p37 and p40 could be shown to be dependent on the C-terminus and another yet unidentified domain. Thereby, simultaneous RNA-binding of multiple hnRNP D isoforms might induce local condensation of RNA structure resulting in interference with exon recognition either by direct displacement of U2AF p65 from the PPT or by inhibiting the establishment of an exon bridging complexes. Finally, mutually enhanced recruitment of

hnRNP K by either hnRNP D or hnRNP A1 might further enhance ESSV-mediated interference.

4.3   Future prospects

Future experiments will focus on the verification and further elucidation of the bipartite interference mechanism exhibited by hnRNP D. Moreover, since ectopic overexpression of hnRNP D indicated the presence of other hnRNP D-dependent SREs within the HIV-1 genome. Therefore all other UAG-containing SREs of HIV-1 published to be bound by hnRNP A1, the ESS2, ESS3 and the ISS, should be examined regarding their hnRNP D dependency. Finally, the role of hnRNP K at the ESSV should be further elucidated, especially if binding of hnRNP D and hnRNP K is mutually facilitated.

5 References

1. **Abdul-Manan, N., S. M. O'Malley, and K. R. Williams.** 1996. Origins of binding specificity of the A1 heterogeneous nuclear ribonucleoprotein. Biochemistry **35:**3545-3554.

2. **Abdul-Manan, N., and K. R. Williams.** 1996. hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base-specific binding. Nucleic acids research **24:**4063-4070.

3. **Amendt, B. A., D. Hesslein, L. J. Chang, and C. M. Stoltzfus.** 1994. Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. Molecular and cellular biology **14:**3960-3970.

4. **Amendt, B. A., Z. H. Si, and C. M. Stoltzfus.** 1995. 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. Molecular and cellular biology **15:**6480.

5. **Anderson, J. L., A. T. Johnson, J. L. Howard, and D. F. Purcell.** 2007. Both linear and discontinuous ribosome scanning are used for translation initiation from bicistronic human immunodeficiency virus type 1 env mRNAs. Journal of virology **81:**4664-4676.

6. **Arao, Y., R. Kuriyama, F. Kayama, and S. Kato.** 2000. A nuclear matrix-associated factor, SAF-B, interacts with specific isoforms of AUF1/hnRNP D. Archives of biochemistry and biophysics **380:**228-236.

7. **Asang, C., S. Erkelenz, and H. Schaal.** 2012. The HIV-1 major splice donor D1 is activated by splicing enhancer elements within the leader region and the p17-inhibitory sequence. Virology **432:**133-145.

8. **Asang, C., I. Hauber, and H. Schaal.** 2008. Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. Nucleic acids research **36:**1450-1463.

9. **Baber, J. L., D. Libutti, D. Levens, and N. Tjandra.** 1999. High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. Journal of molecular biology **289:**949-962.

10. **Barash, Y., J. A. Calarco, W. Gao, Q. Pan, X. Wang, O. Shai, B. J. Blencowe, and B. J. Frey.** 2010. Deciphering the splicing code. Nature **465:**53-59.

11. **Ben-David, Y., M. R. Bani, B. Chabot, A. De Koven, and A. Bernstein.** 1992. Retroviral insertions downstream of the heterogeneous nuclear ribonucleoprotein A1 gene in erythroleukemia cells: evidence that A1 is not essential for cell growth. Molecular and cellular biology **12:**4449-4455.

12. **Berget, S. M.** 1995. Exon recognition in vertebrate splicing. The Journal of biological chemistry **270:**2411-2414.

References

13.   **Bilodeau, P. S., J. K. Domsic, A. Mayeda, A. R. Krainer, and C. M. Stoltzfus.** 2001. 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. Journal of virology **75:**8487-8497.

14.   **Birney, E., S. Kumar, and A. R. Krainer.** 1993. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic acids research **21:**5803-5816.

15.   **Black, D. L.** 2003. Mechanisms of alternative pre-messenger RNA splicing. Annual review of biochemistry **72:**291-336.

16.   **Blanchette, M., and B. Chabot.** 1999. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. The EMBO journal **18:**1939-1952.

17.   **Blencowe, B. J.** 2006. Alternative splicing: new insights from global analyses. Cell **126:**37-47.

18.   **Bonnal, S., C. Martinez, P. Forch, A. Bachi, M. Wilm, and J. Valcarcel.** 2008. RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. Molecular cell **32:**81-95.

19.   **Bono, F., and N. H. Gehring.** 2011. Assembly, disassembly and recycling: the dynamics of exon junction complexes. RNA biology **8:**24-29.

20.   **Boucher, L., C. A. Ouzounis, A. J. Enright, and B. J. Blencowe.** 2001. A genome-wide survey of RS domain proteins. RNA **7:**1693-1701.

21.   **Boukis, L. A., N. Liu, S. Furuyama, and J. P. Bruzik.** 2004. Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles. The Journal of biological chemistry **279:**29647-29653.

22.   **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry **72:**248-254.

23.   **Brewer, G.** 1991. An A + U-rich element RNA-binding factor regulates c-myc mRNA stability in vitro. Molecular and cellular biology **11:**2460-2466.

24.   **Brewer, G., and J. Ross.** 1989. Regulation of c-myc mRNA stability in vitro by a labile destabilizer with an essential nucleic acid component. Molecular and cellular biology **9:**1996-2006.

25.   **Brow, D. A.** 2002. Allosteric cascade of spliceosome activation. Annual review of genetics **36:**333-360.

26.   **Buratti, E., C. Stuani, G. De Prato, and F. E. Baralle.** 2007. SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer. Nucleic acids research **35:**4359-4368.

27.   **Burd, C. G., and G. Dreyfuss.** 1994. Conserved structures and diversity of functions of RNA-binding proteins. Science **265:**615-621.

28.   **Burd, C. G., and G. Dreyfuss.** 1994. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. The EMBO journal **13:**1197-1204.

References

29. **Buvoli, M., F. Cobianchi, G. Biamonti, and S. Riva.** 1990. Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. Nucleic acids research **18:**6595-6600.

30. **Caceres, J. F., and A. R. Kornblihtt.** 2002. Alternative splicing: multiple control mechanisms and involvement in human disease. Trends in genetics : TIG **18:**186-193.

31. **Caceres, J. F., T. Misteli, G. R. Screaton, D. L. Spector, and A. R. Krainer.** 1997. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. The Journal of cell biology **138:**225-238.

32. **Caputi, M., M. Freund, S. Kammler, C. Asang, and H. Schaal.** 2004. A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. Journal of virology **78:**6517-6526.

33. **Caputi, M., A. Mayeda, A. R. Krainer, and A. M. Zahler.** 1999. hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. The EMBO journal **18:**4060-4067.

34. **Cartegni, L., S. L. Chew, and A. R. Krainer.** 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nature reviews. Genetics **3:**285-298.

35. **Cartegni, L., M. Maconi, E. Morandi, F. Cobianchi, S. Riva, and G. Biamonti.** 1996. hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. Journal of molecular biology **259:**337-348.

36. **Casas-Finet, J. R., J. D. Smith, Jr., A. Kumar, J. G. Kim, S. H. Wilson, and R. L. Karpel.** 1993. Mammalian heterogeneous ribonucleoprotein A1 and its constituent domains. Nucleic acid interaction, structural stability and self-association. Journal of molecular biology **229:**873-889.

37. **Chabot, B., M. Blanchette, I. Lapierre, and H. La Branche.** 1997. An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. Molecular and cellular biology **17:**1776-1786.

38. **Chan, D. C., D. Fass, J. M. Berger, and P. S. Kim.** 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell **89:**263-273.

39. **Chang, Y. F., J. S. Imam, and M. F. Wilkinson.** 2007. The nonsense-mediated decay RNA surveillance pathway. Annual review of biochemistry **76:**51-74.

40. **Chasin, L. A.** 2007. Searching for splicing motifs. Advances in experimental medicine and biology **623:**85-106.

41. **Chatterjee, A., A. Rathore, S. Vidyant, K. Kakkar, and T. N. Dhole.** 2012. Chemokines and chemokine receptors in susceptibility to HIV-1 infection and progression to AIDS. Disease markers **32:**143-151.

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

References

43. **Chen, C. Y., N. Xu, W. Zhu, and A. B. Shyu.** 2004. Functional dissection of hnRNP D suggests that nuclear import is required before hnRNP D can modulate mRNA turnover in the cytoplasm. RNA **10:**669-680.

44. **Chen, J. Y., L. Stands, J. P. Staley, R. R. Jackups, Jr., L. J. Latus, and T. H. Chang.** 2001. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Molecular cell **7:**227-232.

45. **Chen, Y., X. Zhou, N. Liu, C. Wang, L. Zhang, W. Mo, and G. Hu.** 2008. Arginine methylation of hnRNP K enhances p53 transcriptional activity. FEBS letters **582:**1761-1765.

46. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical biochemistry **162:**156-159.

47. **Clapham, P. R., and A. McKnight.** 2001. HIV-1 receptors and cell tropism. British medical bulletin **58:**43-59.

48. **Cobianchi, F., R. L. Karpel, K. R. Williams, V. Notario, and S. H. Wilson.** 1988. Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. Large-scale overproduction in Escherichia coli and cooperative binding to single-stranded nucleic acids. The Journal of biological chemistry **263:**1063-1071.

49. **Collins, C. A., and C. Guthrie.** 2001. Genetic interactions between the 5' and 3' splice site consensus sequences and U6 snRNA during the second catalytic step of pre-mRNA splicing. RNA **7:**1845-1854.

50. **Coolidge, C. J., R. J. Seely, and J. G. Patton.** 1997. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. Nucleic acids research **25:**888-896.

51. **Del Gatto-Konczak, F., M. Olive, M. C. Gesnel, and R. Breathnach.** 1999. hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. Molecular and cellular biology **19:**251-260.

52. **DeMaria, C. T., and G. Brewer.** 1996. AUF1 binding affinity to A+U-rich elements correlates with rapid mRNA degradation. The Journal of biological chemistry **271:**12179-12184.

53. **DeMaria, C. T., Y. Sun, L. Long, B. J. Wagner, and G. Brewer.** 1997. Structural determinants in AUF1 required for high affinity binding to A + U-rich elements. The Journal of biological chemistry **272:**27635-27643.

54. **Dempsey, L. A., M. J. Li, A. DePace, P. Bray-Ward, and N. Maizels.** 1998. The human HNRPD locus maps to 4q21 and encodes a highly conserved protein. Genomics **49:**378-384.

55. **Ding, J., M. K. Hayashi, Y. Zhang, L. Manche, A. R. Krainer, and R. M. Xu.** 1999. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes & development **13:**1102-1115.

56. **Doktor, T. K., L. D. Schroeder, A. Vested, J. Palmfeldt, H. S. Andersen, N. Gregersen, and B. S. Andresen.** 2011. SMN2 exon 7 splicing is inhibited by binding of hnRNP A1 to a common ESS motif that spans the 3' splice site. Human mutation **32:**220-230.

References

57. **Dominguez, C., and F. H. Allain.** 2006. NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. Nucleic acids research **34:**3634-3645.

58. **Dominguez, C., J. F. Fisette, B. Chabot, and F. H. Allain.** 2010. Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. Nature structural & molecular biology **17:**853-861.

59. **Domsic, J. K., Y. Wang, A. Mayeda, A. R. Krainer, and C. M. Stoltzfus.** 2003. Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts. Molecular and cellular biology **23:**8762-8772.

60. **Douglas, A. G., and M. J. Wood.** 2011. RNA splicing: disease and therapy. Briefings in functional genomics **10:**151-164.

61. **Dreyfuss, G., V. N. Kim, and N. Kataoka.** 2002. Messenger-RNA-binding proteins and the messages they carry. Nature reviews. Molecular cell biology **3:**195-205.

62. **Dreyfuss, G., M. J. Matunis, S. Pinol-Roma, and C. G. Burd.** 1993. hnRNP proteins and the biogenesis of mRNA. Annual review of biochemistry **62:**289-321.

63. **Du, H., and M. Rosbash.** 2002. The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. Nature **419:**86-90.

64. **Du, H., and M. Rosbash.** 2001. Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing. RNA **7:**133-142.

65. **Enokizono, Y., Y. Konishi, K. Nagata, K. Ouhashi, S. Uesugi, F. Ishikawa, and M. Katahira.** 2005. Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. The Journal of biological chemistry **280:**18862-18870.

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

67. **Erkelenz, S., W. F. Mueller, M. S. Evans, A. Busch, K. Schoneweis, K. J. Hertel, and H. Schaal.** 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA **19:**96-102.

68. **Erkelenz, S., G. Poschmann, S. Theiss, A. Stefanski, F. Hillebrand, M. Otte, K. Stuhler, and H. Schaal.** 2013. Tra2-Mediated Recognition of HIV-1 5' Splice Site D3 as a Key Factor in the Processing of vpr mRNA. Journal of virology **87:**2721-2734.

69. **Erkelenz, S., S. Theiss, M. Otte, M. Widera, J. O. Peter, and H. Schaal.** 2014. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic acids research **42:**10681-10697.

70. **Erlitzki, R., and M. Fry.** 1997. Sequence-specific binding protein of single-stranded and unimolecular quadruplex telomeric DNA from rat hepatocytes. The Journal of biological chemistry **272:**15881-15890.

71. **Eversole, A., and N. Maizels.** 2000. In vitro properties of the conserved mammalian protein hnRNP D suggest a role in telomere maintenance. Molecular and cellular biology **20:**5425-5432.

72. **Exline, C. M., Z. Feng, and C. M. Stoltzfus.** 2008. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. Journal of virology **82:**3921-3931.

73. **Fletcher, T. M., D. Sun, M. Salazar, and L. H. Hurley.** 1998. Effect of DNA secondary structure on human telomerase activity. Biochemistry **37:**5536-5541.

74. **Freed, E. O.** 2001. HIV-1 replication. Somatic cell and molecular genetics **26:**13-33.

75. **Freund, M.** Heinrich-Heine-Universität Düsseldorf, 2004, PhD thesis. Die Funktion des U1 snRNPs in der HIV-1 env-Expression.

76. **Freund, M., C. Asang, S. Kammler, C. Konermann, J. Krummheuer, M. Hipp, I. Meyer, W. Gierling, S. Theiss, T. Preuss, D. Schindler, J. Kjems, and H. Schaal.** 2003. A novel approach to describe a U1 snRNA binding site. Nucleic acids research **31:**6963-6975.

77. **Fu, X. D., and T. Maniatis.** 1992. The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. Proceedings of the National Academy of Sciences of the United States of America **89:**1725-1729.

78. **Fu, Y., A. Masuda, M. Ito, J. Shinmi, and K. Ohno.** 2011. AG-dependent 3'-splice sites are predisposed to aberrant splicing due to a mutation at the first nucleotide of an exon. Nucleic acids research **39:**4396-4404.

79. **Gao, K., A. Masuda, T. Matsuura, and K. Ohno.** 2008. Human branch point consensus sequence is yUnAy. Nucleic acids research **36:**2257-2267.

80. **Goina, E., N. Skoko, and F. Pagani.** 2008. Binding of DAZAP1 and hnRNPA1/A2 to an exonic splicing silencer in a natural BRCA1 exon 18 mutant. Molecular and cellular biology **28:**3850-3860.

81. **Gratacos, F. M., and G. Brewer.** 2010. The role of AUF1 in regulated mRNA decay. Wiley interdisciplinary reviews. RNA **1:**457-473.

82. **Graveley, B. R.** 2000. Sorting out the complexity of SR protein functions. RNA **6:**1197-1211.

83. **Gross, T., K. Richert, C. Mierke, M. Lutzelberger, and N. F. Kaufer.** 1998. 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 research **26:**505-511.

84. **Hallay, H., N. Locker, L. Ayadi, D. Ropers, E. Guittet, and C. Branlant.** 2006. 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. The Journal of biological chemistry **281:**37159-37174.

85. **Han, K., G. Yeo, P. An, C. B. Burge, and P. J. Grabowski.** 2005. A combinatorial code for splicing silencing: UAGG and GGGG motifs. PLoS biology **3:**e158.

86. **Han, S. P., Y. H. Tang, and R. Smith.** 2010. Functional diversity of the hnRNPs: past, present and perspectives. The Biochemical journal **430:**379-392.

87. **Hartmann, L., S. Theiss, D. Niederacher, and H. Schaal.** 2008. Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? Frontiers in bioscience : a journal and virtual library **13:**3252-3272.

88. **Haverland, N. A., H. S. Fox, and P. Ciborowski.** 2014. Quantitative proteomics by SWATH-MS reveals altered expression of nucleic acid binding and regulatory proteins in HIV-1-infected macrophages. Journal of proteome research **13:**2109-2119.

89. **He, C., and R. Schneider.** 2006. 14-3-3sigma is a p37 AUF1-binding protein that facilitates AUF1 transport and AU-rich mRNA decay. The EMBO journal **25:**3823-3831.

90. **Heinrichs, V., M. Bach, and R. Luhrmann.** 1990. U1-specific protein C is required for efficient complex formation of U1 snRNP with a 5' splice site. Molecular biology reports **14:**165.

91. **Herrick, G., and B. Alberts.** 1976. Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. The Journal of biological chemistry **251:**2124-2132.

92. **Herrmann, F., M. Bossert, A. Schwander, E. Akgun, and F. O. Fackelmayer.** 2004. Arginine methylation of scaffold attachment factor A by heterogeneous nuclear ribonucleoprotein particle-associated PRMT1. The Journal of biological chemistry **279:**48774-48779.

93. **Hoffman, B. E., and P. J. Grabowski.** 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. Genes & development **6:**2554-2568.

94. **Hoffman, D. W., C. C. Query, B. L. Golden, S. W. White, and J. D. Keene.** 1991. RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. Proceedings of the National Academy of Sciences of the United States of America **88:**2495-2499.

95. **House, A. E., and K. W. Lynch.** 2006. An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. Nature structural & molecular biology **13:**937-944.

96. **Ibrahim, E. C., T. D. Schaal, K. J. Hertel, R. Reed, and T. Maniatis.** 2005. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. Proceedings of the National Academy of Sciences of the United States of America **102:**5002-5007.

97. **Inoue, A., Y. Arao, A. Omori, S. Ichinose, K. Nishio, N. Yamamoto, Y. Kinoshita, and S. Mita.** 2003. Identification of S1 proteins B2, C1 and D1 as AUF1 isoforms and their major role as heterogeneous nuclear ribonucleoprotein proteins. The Biochemical journal **372:**775-785.

98. **Ishikawa, F., M. J. Matunis, G. Dreyfuss, and T. R. Cech.** 1993. Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. Molecular and cellular biology **13:**4301-4310.

99. **Jablonski, J. A., and M. Caputi.** 2009. Role of cellular RNA processing factors in human immunodeficiency virus type 1 mRNA metabolism, replication, and infectivity. Journal of virology **83:**981-992.

References

100. **Jacquenet, S., A. Mereau, P. S. Bilodeau, L. Damier, C. M. Stoltzfus, and C. Branlant.** 2001. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. The Journal of biological chemistry **276:**40464-40475.

101. **Jager, S., P. Cimermancic, N. Gulbahce, J. R. Johnson, K. E. McGovern, S. C. Clarke, M. Shales, G. Mercenne, L. Pache, K. Li, H. Hernandez, G. M. Jang, S. L. Roth, E. Akiva, J. Marlett, M. Stephens, I. D'Orso, J. Fernandes, M. Fahey, C. Mahon, A. J. O'Donoghue, A. Todorovic, J. H. Morris, D. A. Maltby, T. Alber, G. Cagney, F. D. Bushman, J. A. Young, S. K. Chanda, W. I. Sundquist, T. Kortemme, R. D. Hernandez, C. S. Craik, A. Burlingame, A. Sali, A. D. Frankel, and N. J. Krogan.** 2012. Global landscape of HIV-human protein complexes. Nature **481:**365-370.

102. **Kajita, Y., J. Nakayama, M. Aizawa, and F. Ishikawa.** 1995. The UUAG-specific RNA binding protein, heterogeneous nuclear ribonucleoprotein D0. Common modular structure and binding properties of the 2xRBD-Gly family. The Journal of biological chemistry **270:**22167-22175.

103. **Kammler, S., C. Leurs, M. Freund, J. Krummheuer, K. Seidel, T. O. Tange, M. K. Lund, J. Kjems, A. Scheid, and H. Schaal.** 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.

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

105. **Kanopka, A., O. Muhlemann, and G. Akusjarvi.** 1996. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. Nature **381:**535-538.

106. **Karn, J., and C. M. Stoltzfus.** 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harbor perspectives in medicine **2:**a006916.

107. **Kashima, T., and J. L. Manley.** 2003. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. Nature genetics **34:**460-463.

108. **Kashima, T., N. Rao, and J. L. Manley.** 2007. An intronic element contributes to splicing repression in spinal muscular atrophy. Proceedings of the National Academy of Sciences of the United States of America **104:**3426-3431.

109. **Katahira, M., Y. Miyanoiri, Y. Enokizono, G. Matsuda, T. Nagata, F. Ishikawa, and S. Uesugi.** 2001. Structure of the C-terminal RNA-binding domain of hnRNP D0 (AUF1), its interactions with RNA and DNA, and change in backbone dynamics upon complex formation with DNA. Journal of molecular biology **311:**973-988.

110. **Kataoka, N., J. L. Bachorik, and G. Dreyfuss.** 1999. Transportin-SR, a nuclear import receptor for SR proteins. The Journal of cell biology **145:**1145-1152.

111. **Kedar, V. P., B. E. Zucconi, G. M. Wilson, and P. J. Blackshear.** 2012. Direct binding of specific AUF1 isoforms to tandem zinc finger domains of tristetraprolin (TTP) family proteins. The Journal of biological chemistry **287:**5459-5471.

112. **Kim, S. Y., R. Byrn, J. Groopman, and D. Baltimore.** 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. Journal of virology **63:**3708-3713.

References

113. **Klotman, M. E., S. Kim, A. Buchbinder, A. DeRossi, D. Baltimore, and F. Wong-Staal.** 1991. Kinetics of expression of multiply spliced RNA in early human immunodeficiency virus type 1 infection of lymphocytes and monocytes. Proceedings of the National Academy of Sciences of the United States of America **88:**5011-5015.

114. **Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, and J. L. Manley.** 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature **368:**119-124.

115. **Konarska, M. M., J. Vilardell, and C. C. Query.** 2006. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. Molecular cell **21:**543-553.

116. **Krummheuer, J., A. T. Johnson, I. Hauber, S. Kammler, J. L. Anderson, J. Hauber, D. F. Purcell, and H. Schaal.** 2007. A minimal uORF within the HIV-1 vpu leader allows efficient translation initiation at the downstream env AUG. Virology **363:**261-271.

117. **Krummheuer, J., C. Lenz, S. Kammler, A. Scheid, and H. Schaal.** 2001. Influence of the small leader exons 2 and 3 on human immunodeficiency virus type 1 gene expression. Virology **286:**276-289.

118. **Kumar, A., J. R. Casas-Finet, C. J. Luneau, R. L. Karpel, B. M. Merrill, K. R. Williams, and S. H. Wilson.** 1990. Mammalian heterogeneous nuclear ribonucleoprotein A1. Nucleic acid binding properties of the COOH-terminal domain. The Journal of biological chemistry **265:**17094-17100.

119. **LaBranche, H., S. Dupuis, Y. Ben-David, M. R. Bani, R. J. Wellinger, and B. Chabot.** 1998. Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. Nature genetics **19:**199-202.

120. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.

121. **Lai, M. C., R. I. Lin, S. Y. Huang, C. W. Tsai, and W. Y. Tarn.** 2000. A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. The Journal of biological chemistry **275:**7950-7957.

122. **Le Hir, H., and G. R. Andersen.** 2008. Structural insights into the exon junction complex. Current opinion in structural biology **18:**112-119.

123. **Le Hir, H., E. Izaurralde, L. E. Maquat, and M. J. Moore.** 2000. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. The EMBO journal **19:**6860-6869.

124. **Le Hir, H., and B. Seraphin.** 2008. EJCs at the heart of translational control. Cell **133:**213-216.

125. **LeCuyer, K. A., L. S. Behlen, and O. C. Uhlenbeck.** 1995. Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. Biochemistry **34:**10600-10606.

126. **Li, T., E. Evdokimov, R. F. Shen, C. C. Chao, E. Tekle, T. Wang, E. R. Stadtman, D. C. Yang, and P. B. Chock.** 2004. Sumoylation of heterogeneous nuclear ribonucleoproteins, zinc finger proteins, and nuclear pore complex proteins: a proteomic analysis. Proceedings of the National Academy of Sciences of the United States of America **101:**8551-8556.

References

127. **Li, X., C. Johansson, J. Glahder, A. K. Mossberg, and S. Schwartz.** 2013. 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 research.

128. **Lin, S., and X. D. Fu.** 2007. SR proteins and related factors in alternative splicing. Advances in experimental medicine and biology **623:**107-122.

129. **Long, J. C., and J. F. Caceres.** 2009. The SR protein family of splicing factors: master regulators of gene expression. The Biochemical journal **417:**15-27.

130. **Lund, N., M. P. Milev, R. Wong, T. Sanmuganantham, K. Woolaway, B. Chabot, S. Abou Elela, A. J. Mouland, and A. Cochrane.** 2012. Differential effects of hnRNP D/AUF1 isoforms on HIV-1 gene expression. Nucleic acids research **40:**3663-3675.

131. **Lunde, B. M., C. Moore, and G. Varani.** 2007. RNA-binding proteins: modular design for efficient function. Nature reviews. Molecular cell biology **8:**479-490.

132. **Lutzelberger, M., T. Gross, and N. F. Kaufer.** 1999. Srp2, an SR protein family member of fission yeast: in vivo characterization of its modular domains. Nucleic acids research **27:**2618-2626.

133. **Luukkonen, B. G., and B. Seraphin.** 1997. The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in Saccharomyces cerevisiae. The EMBO journal **16:**779-792.

134. **Madsen, J. M., and C. M. Stoltzfus.** 2005. An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. Journal of virology **79:**10478-10486.

135. **Mandal, D., Z. Feng, and C. M. Stoltzfus.** 2010. Excessive RNA splicing and inhibition of HIV-1 replication induced by modified U1 small nuclear RNAs. Journal of virology **84:**12790-12800.

136. **Mandal, D., Z. Feng, and C. M. Stoltzfus.** 2008. Gag-processing defect of human immunodeficiency virus type 1 integrase E246 and G247 mutants is caused by activation of an overlapping 5' splice site. Journal of virology **82:**1600-1604.

137. **Maniatis, T., and B. Tasic.** 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature **418:**236-243.

138. **Manley, J. L., and A. R. Krainer.** 2010. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). Genes & development **24:**1073-1074.

139. **Marchand, V., M. Santerre, C. Aigueperse, L. Fouillen, J. M. Saliou, A. Van Dorsselaer, S. Sanglier-Cianferani, C. Branlant, and Y. Motorin.** 2011. Identification of protein partners of the human immunodeficiency virus 1 tat/rev exon 3 leads to the discovery of a new HIV-1 splicing regulator, protein hnRNP K. RNA biology **8:**325-342.

140. **Maris, C., C. Dominguez, and F. H. Allain.** 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. The FEBS journal **272:**2118-2131.

141. **Matera, A. G., and Z. Wang.** 2014. A day in the life of the spliceosome. Nature reviews. Molecular cell biology **15:**108-121.

142. **Matunis, E. L., M. J. Matunis, and G. Dreyfuss.** 1992. Characterization of the major hnRNP proteins from Drosophila melanogaster. The Journal of cell biology **116:**257-269.

143. **Matunis, M. J., W. M. Michael, and G. Dreyfuss.** 1992. Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. Molecular and cellular biology **12:**164-171.

144. **Mayeda, A., and A. R. Krainer.** 1992. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. Cell **68:**365-375.

145. **Mayeda, A., S. H. Munroe, J. F. Caceres, and A. R. Krainer.** 1994. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. The EMBO journal **13:**5483-5495.

146. **Mayeda, A., S. H. Munroe, R. M. Xu, and A. R. Krainer.** 1998. Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. RNA **4:**1111-1123.

147. **McKay, S. J., and H. Cooke.** 1992. hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGGn. Nucleic acids research **20:**6461-6464.

148. **Michaud, S., and R. Reed.** 1991. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes & development **5:**2534-2546.

149. **Mikula, M., A. Dzwonek, J. Karczmarski, T. Rubel, M. Dadlez, L. S. Wyrwicz, K. Bomsztyk, and J. Ostrowski.** 2006. Landscape of the hnRNP K protein-protein interactome. Proteomics **6:**2395-2406.

150. **Mikula, M., J. Karczmarski, A. Dzwonek, T. Rubel, E. Hennig, M. Dadlez, J. M. Bujnicki, K. Bomsztyk, and J. Ostrowski.** 2006. Casein kinases phosphorylate multiple residues spanning the entire hnRNP K length. Biochimica et biophysica acta **1764:**299-306.

151. **Munis, J. R., R. S. Kornbluth, J. C. Guatelli, and D. D. Richman.** 1992. Ordered appearance of human immunodeficiency virus type 1 nucleic acids following high multiplicity infection of macrophages. The Journal of general virology **73 (Pt 8):**1899-1906.

152. **Nadler, S. G., B. M. Merrill, W. J. Roberts, K. M. Keating, M. J. Lisbin, S. F. Barnett, S. H. Wilson, and K. R. Williams.** 1991. Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative, UP1, with RNA and DNA: evidence for multiple RNA binding domains and salt-dependent binding mode transitions. Biochemistry **30:**2968-2976.

153. **Nagata, T., Y. Kurihara, G. Matsuda, J. Saeki, T. Kohno, Y. Yanagida, F. Ishikawa, S. Uesugi, and M. Katahira.** 1999. Structure and interactions with RNA of the N-terminal UUAG-specific RNA-binding domain of hnRNP D0. Journal of molecular biology **287:**221-237.

154. **Nasim, F. U., S. Hutchison, M. Cordeau, and B. Chabot.** 2002. 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:**1078-1089.

References

155.   **Nichols, R. C., X. W. Wang, J. Tang, B. J. Hamilton, F. A. High, H. R. Herschman, and W. F. Rigby.** 2000. The RGG domain in hnRNP A2 affects subcellular localization. Experimental cell research **256:**522-532.

156.   **Nilsen, T. W., and B. R. Graveley.** 2010. Expansion of the eukaryotic proteome by alternative splicing. Nature **463:**457-463.

157.   **O'Brien, W. A., Y. Koyanagi, A. Namazie, J. Q. Zhao, A. Diagne, K. Idler, J. A. Zack, and I. S. Chen.** 1990. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. Nature **348:**69-73.

158.   **O'Reilly, M. M., M. T. McNally, and K. L. Beemon.** 1995. Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. Virology **213:**373-385.

159.   **Okunola, H. L., and A. R. Krainer.** 2009. Cooperative-binding and splicing-repressive properties of hnRNP A1. Molecular and cellular biology **29:**5620-5631.

160.   **Orengo, J. P., and T. A. Cooper.** 2007. Alternative splicing in disease. Advances in experimental medicine and biology **623:**212-223.

161.   **Ott, M., M. Geyer, and Q. Zhou.** 2011. The control of HIV transcription: keeping RNA polymerase II on track. Cell host & microbe **10:**426-435.

162.   **Oubridge, C., N. Ito, P. R. Evans, C. H. Teo, and K. Nagai.** 1994. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature **372:**432-438.

163.   **Pan, Q., O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe.** 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature genetics **40:**1413-1415.

164.   **Park, S. H., G. H. Park, H. Gu, W. I. Hwang, I. K. Lim, W. K. Paik, and S. Kim.** 1997. Heterogeneous nuclear RNP protein A1-arginine methylation during HCT-48 cell cycle. Biochemistry and molecular biology international **42:**657-666.

165.   **Pathak, S., G. A. De Souza, T. Salte, H. G. Wiker, and B. Asjo.** 2009. HIV induces both a down-regulation of IRAK-4 that impairs TLR signalling and an up-regulation of the antibiotic peptide dermcidin in monocytic cells. Scandinavian journal of immunology **70:**264-276.

166.   **Peabody, D. S.** 1993. The RNA binding site of bacteriophage MS2 coat protein. The EMBO journal **12:**595-600.

167.   **Peterlin, B. M., and D. H. Price.** 2006. Controlling the elongation phase of transcription with P-TEFb. Molecular cell **23:**297-305.

168.   **Pinol-Roma, S., Y. D. Choi, M. J. Matunis, and G. Dreyfuss.** 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. Genes & development **2:**215-227.

169.   **Pollard, V. W., and M. H. Malim.** 1998. The HIV-1 Rev protein. Annual review of microbiology **52:**491-532.

170.   **Polzer, S., M. T. Dittmar, H. Schmitz, and M. Schreiber.** 2002. The N-linked glycan g15 within the V3 loop of the HIV-1 external glycoprotein gp120 affects coreceptor usage, cellular tropism, and neutralization. Virology **304:**70-80.

171. **Pomeranz Krummel, D. A., C. Oubridge, A. K. Leung, J. Li, and K. Nagai.** 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature **458:**475-480.

172. **Powell, A. J., and D. S. Peabody.** 2001. Asymmetric interactions in the adenosine-binding pockets of the MS2 coat protein dimer. BMC molecular biology **2:**6.

173. **Purcell, D. F., and M. A. Martin.** 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. Journal of virology **67:**6365-6378.

174. **Query, C. C., M. J. Moore, and P. A. Sharp.** 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes & development **8:**587-597.

175. **Quivy, V., S. De Walque, and C. Van Lint.** 2007. Chromatin-associated regulation of HIV-1 transcription: implications for the development of therapeutic strategies. Sub-cellular biochemistry **41:**371-396.

176. **R., H. J. D.** 1985. Simplified method for silver staining of proteins in polyacrylamide gels and the mechanism of silver staining. Electrophoresis **6:**103-112.

177. **Riva, S., C. Morandi, P. Tsoulfas, M. Pandolfo, G. Biamonti, B. Merrill, K. R. Williams, G. Multhaup, K. Beyreuther, H. Werr, and et al.** 1986. Mammalian single-stranded DNA binding protein UP I is derived from the hnRNP core protein A1. The EMBO journal **5:**2267-2273.

178. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Molecular and cellular biology **10:**84-94.

179. **Roca, X., A. R. Krainer, and I. C. Eperon.** 2013. Pick one, but be quick: 5' splice sites and the problems of too many choices. Genes & development **27:**129-144.

180. **Rooke, N., V. Markovtsov, E. Cagavi, and D. L. Black.** 2003. Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. Molecular and cellular biology **23:**1874-1884.

181. **Roth, M. B., A. M. Zahler, and J. A. Stolk.** 1991. A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. The Journal of cell biology **115:**587-596.

182. **Rothrock, C. R., A. E. House, and K. W. Lynch.** 2005. HnRNP L represses exon splicing via a regulated exonic splicing silencer. The EMBO journal **24:**2792-2802.

183. **Rush, M., X. Zhao, and S. Schwartz.** 2005. A splicing enhancer in the E4 coding region of human papillomavirus type 16 is required for early mRNA splicing and polyadenylation as well as inhibition of premature late gene expression. Journal of virology **79:**12002-12015.

184. **Sarkar, B., J. Y. Lu, and R. J. Schneider.** 2003. Nuclear import and export functions in the different isoforms of the AUF1/heterogeneous nuclear ribonucleoprotein protein family. The Journal of biological chemistry **278:**20700-20707.

185. **Schaal, T. D., and T. Maniatis.** 1999. Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Molecular and cellular biology **19:**261-273.

References

186. **Schaub, M. C., S. R. Lopez, and M. Caputi.** 2007. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. The Journal of biological chemistry **282:**13617-13626.

187. **Scherer, S.** 2008. A Short Guide to the Human Genome. Cold Spring Harbor, NY. Cold Spring Harbor Laboratory Press.

188. **Schneider, M., C. L. Will, M. Anokhina, J. Tazi, H. Urlaub, and R. Luhrmann.** 2010. Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. Molecular cell **38:**223-235.

189. **Schwartz, S., B. K. Felber, D. M. Benko, E. M. Fenyo, and G. N. Pavlakis.** 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. Journal of virology **64:**2519-2529.

190. **Schwartz, S., B. K. Felber, and G. N. Pavlakis.** 1991. Expression of human immunodeficiency virus type 1 vif and vpr mRNAs is Rev-dependent and regulated by splicing. Virology **183:**677-686.

191. **Shamoo, Y., U. Krueger, L. M. Rice, K. R. Williams, and T. A. Steitz.** 1997. Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 A resolution. Nature structural biology **4:**215-222.

192. **Sharma, S., L. A. Kohlstaedt, A. Damianov, D. C. Rio, and D. L. Black.** 2008. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nature structural & molecular biology **15:**183-191.

193. **Shen, R., H. E. Richter, and P. D. Smith.** 2011. Early HIV-1 target cells in human vaginal and ectocervical mucosa. American journal of reproductive immunology : AJRI : official journal of the American Society for the Immunology of Reproduction and the International Coordination Committee for Immunology of Reproduction **65:**261-267.

194. **Shepard, P. J., and K. J. Hertel.** 2008. Conserved RNA secondary structures promote alternative splicing. RNA **14:**1463-1469.

195. **Shepard, P. J., and K. J. Hertel.** 2009. The SR protein family. Genome biology **10:**242.

196. **Shibuya, T., and T. W. Mak.** 1983. Isolation and induction of erythroleukemic cell lines with properties of erythroid progenitor burst-forming cell (BFU-E) and erythroid precursor cell (CFU-E). Proceedings of the National Academy of Sciences of the United States of America **80:**3721-3725.

197. **Si, Z., B. A. Amendt, and C. M. Stoltzfus.** 1997. 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 research **25:**861-867.

198. **Si, Z. H., D. Rauch, and C. M. Stoltzfus.** 1998. The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. Molecular and cellular biology **18:**5404-5413.

199. **Silva, A. L., and L. Romao.** 2009. The mammalian nonsense-mediated mRNA decay pathway: to decay or not to decay! Which players make the decision? FEBS letters **583:**499-505.

References

200. **Singh, K. K., S. Erkelenz, S. Rattay, A. K. Dehof, A. Hildebrandt, K. Schulze-Osthoff, H. Schaal, and C. Schwerk.** 2010. Human SAP18 mediates assembly of a splicing regulatory multiprotein complex via its ubiquitin-like fold. RNA **16:**2442-2454.

201. **Siomi, H., and G. Dreyfuss.** 1995. A nuclear localization domain in the hnRNP A1 protein. The Journal of cell biology **129:**551-560.

202. **Siomi, H., M. J. Matunis, W. M. Michael, and G. Dreyfuss.** 1993. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. Nucleic acids research **21:**1193-1198.

203. **Smith, C. W., T. T. Chu, and B. Nadal-Ginard.** 1993. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. Molecular and cellular biology **13:**4939-4952.

204. **Smith, C. W., and J. Valcarcel.** 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. Trends in biochemical sciences **25:**381-388.

205. **Staffa, A., and A. Cochrane.** 1995. Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. Molecular and cellular biology **15:**4597-4605.

206. **Staley, J. P., and C. Guthrie.** 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell **92:**315-326.

207. **Sundquist, W. I., and H. G. Krausslich.** 2012. HIV-1 assembly, budding, and maturation. Cold Spring Harbor perspectives in medicine **2:**a006924.

208. **Suzuki, M., M. Iijima, A. Nishimura, Y. Tomozoe, D. Kamei, and M. Yamada.** 2005. Two separate regions essential for nuclear import of the hnRNP D nucleocytoplasmic shuttling sequence. The FEBS journal **272:**3975-3987.

209. **Swanson, M. S., and G. Dreyfuss.** 1988. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. Molecular and cellular biology **8:**2237-2241.

210. **Tang, C., J. M. Louis, A. Aniana, J. Y. Suh, and G. M. Clore.** 2008. Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. Nature **455:**693-696.

211. **Tange, T. O., C. K. Damgaard, S. Guth, J. Valcarcel, and J. Kjems.** 2001. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. The EMBO journal **20:**5748-5758.

212. **Thisted, T., D. L. Lyakhov, and S. A. Liebhaber.** 2001. Optimized RNA targets of two closely related triple KH domain proteins, heterogeneous nuclear ribonucleoprotein K and alphaCP-2KL, suggest Distinct modes of RNA recognition. The Journal of biological chemistry **276:**17484-17496.

213. **Tripathy, M. K., W. Abbas, and G. Herbein.** 2011. Epigenetic regulation of HIV-1 transcription. Epigenomics **3:**487-502.

214. **Truant, R., and B. R. Cullen.** 1999. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. Molecular and cellular biology **19:**1210-1217.

References

215. **Ule, J., G. Stefani, A. Mele, M. Ruggiu, X. Wang, B. Taneri, T. Gaasterland, B. J. Blencowe, and R. B. Darnell.** 2006. An RNA map predicting Nova-dependent splicing regulation. Nature **444:**580-586.

216. **van den Worm, S. H., N. J. Stonehouse, K. Valegard, J. B. Murray, C. Walton, K. Fridborg, P. G. Stockley, and L. Liljas.** 1998. Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. Nucleic acids research **26:**1345-1351.

217. **van der Houven van Oordt, W., M. T. Diaz-Meco, J. Lozano, A. R. Krainer, J. Moscat, and J. F. Caceres.** 2000. The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. The Journal of cell biology **149:**307-316.

218. **Vassileva, M. T., and M. J. Matunis.** 2004. SUMO modification of heterogeneous nuclear ribonucleoproteins. Molecular and cellular biology **24:**3623-3632.

219. **Vriz, S., and M. Mechali.** 1989. Analysis of 3'-untranslated regions of seven c-myc genes reveals conserved elements prevalent in post-transcriptionally regulated genes. FEBS letters **251:**201-206.

220. **Wagner, B. J., C. T. DeMaria, Y. Sun, G. M. Wilson, and G. Brewer.** 1998. Structure and genomic organization of the human AUF1 gene: alternative pre-mRNA splicing generates four protein isoforms. Genomics **48:**195-202.

221. **Wahl, M. C., C. L. Will, and R. Luhrmann.** 2009. The spliceosome: design principles of a dynamic RNP machine. Cell **136:**701-718.

222. **Wang, E., W. F. Mueller, K. J. Hertel, and F. Cambi.** 2011. 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. The Journal of biological chemistry **286:**4059-4071.

223. **Wang, G. S., and T. A. Cooper.** 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. Nature reviews. Genetics **8:**749-761.

224. **Wang, Z., and C. B. Burge.** 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA **14:**802-813.

225. **Weighardt, F., G. Biamonti, and S. Riva.** 1995. Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. Journal of cell science **108 (Pt 2):**545-555.

226. **Weissenhorn, W., A. Dessen, S. C. Harrison, J. J. Skehel, and D. C. Wiley.** 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature **387:**426-430.

227. **Widera, M., S. Erkelenz, F. Hillebrand, A. Krikoni, D. Widera, W. Kaisers, R. Deenen, M. Gombert, R. Dellen, T. Pfeiffer, B. Kaltschmidt, C. Munk, V. Bosch, K. Kohrer, and H. Schaal.** 2013. An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA. Journal of virology **87:**2707-2720.

228. **Will, C. L., and R. Luhrmann.** 2011. Spliceosome structure and function. Cold Spring Harbor perspectives in biology **3**.

229. **Williamson, J. R.** 1994. G-quartet structures in telomeric DNA. Annual review of biophysics and biomolecular structure **23:**703-730.

References

230. **Wilson, G. M., J. Lu, K. Sutphen, Y. Sun, Y. Huynh, and G. Brewer.** 2003. Regulation of A + U-rich element-directed mRNA turnover involving reversible phosphorylation of AUF1. The Journal of biological chemistry **278:**33029-33038.

231. **Wilson, G. M., Y. Sun, H. Lu, and G. Brewer.** 1999. Assembly of AUF1 oligomers on U-rich RNA targets by sequential dimer association. The Journal of biological chemistry **274:**33374-33381.

232. **Wilson, G. M., K. Sutphen, K. Chuang, and G. Brewer.** 2001. Folding of A+U-rich RNA elements modulates AUF1 binding. Potential roles in regulation of mRNA turnover. The Journal of biological chemistry **276:**8695-8704.

233. **Woolford, J. L., Jr.** 1989. Nuclear pre-mRNA splicing in yeast. Yeast **5:**439-457.

234. **Wu, J. Y., and T. Maniatis.** 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell **75:**1061-1070.

235. **Wu, S., C. M. Romfo, T. W. Nilsen, and M. R. Green.** 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature **402:**832-835.

236. **Xie, J., J. A. Lee, T. L. Kress, K. L. Mowry, and D. L. Black.** 2003. Protein kinase A phosphorylation modulates transport of the polypyrimidine tract-binding protein. Proceedings of the National Academy of Sciences of the United States of America **100:**8776-8781.

237. **Xu, R. M., L. Jokhan, X. Cheng, A. Mayeda, and A. R. Krainer.** 1997. Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. Structure **5:**559-570.

238. **Yeo, G. W., N. G. Coufal, T. Y. Liang, G. E. Peng, X. D. Fu, and F. H. Gage.** 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nature structural & molecular biology **16:**130-137.

239. **Zahler, A. M., C. K. Damgaard, J. Kjems, and M. Caputi.** 2004. 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. The Journal of biological chemistry **279:**10077-10084.

240. **Zahler, A. M., W. S. Lane, J. A. Stolk, and M. B. Roth.** 1992. SR proteins: a conserved family of pre-mRNA splicing factors. Genes & development **6:**837-847.

241. **Zahler, A. M., J. R. Williamson, T. R. Cech, and D. M. Prescott.** 1991. Inhibition of telomerase by G-quartet DNA structures. Nature **350:**718-720.

242. **Zhang, C., Z. Zhang, J. Castle, S. Sun, J. Johnson, A. R. Krainer, and M. Q. Zhang.** 2008. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. Genes & development **22:**2550-2563.

243. **Zhang, W., B. J. Wagner, K. Ehrenman, A. W. Schaefer, C. T. DeMaria, D. Crater, K. DeHaven, L. Long, and G. Brewer.** 1993. Purification, characterization, and cDNA cloning of an AU-rich element RNA-binding protein, AUF1. Molecular and cellular biology **13:**7652-7665.

244. **Zhang, X. H., M. A. Arias, S. Ke, and L. A. Chasin.** 2009. Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. RNA **15:**367-376.

245. **Zhang, X. H., and L. A. Chasin.** 2004. Computational definition of sequence motifs governing constitutive exon splicing. Genes & development **18:**1241-1250.

246. **Zhang, Z., and A. R. Krainer.** 2007. Splicing remodels messenger ribonucleoprotein architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. Proceedings of the National Academy of Sciences of the United States of America **104:**11574-11579.

247. **Zhu, J., A. Mayeda, and A. R. Krainer.** 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Molecular cell **8:**1351-1361.

248. **Zhuang, Y., and A. M. Weiner.** 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell **46:**827-835.

249. **Zucconi, B. E., J. D. Ballin, B. Y. Brewer, C. R. Ross, J. Huang, E. A. Toth, and G. M. Wilson.** 2010. Alternatively expressed domains of AU-rich element RNA-binding protein 1 (AUF1) regulate RNA-binding affinity, RNA-induced protein oligomerization, and the local conformation of bound RNA ligands. The Journal of biological chemistry **285:**39127-39139.

250. **Zucconi, B. E., and G. M. Wilson.** 2013. Assembly of Functional Ribonucleoprotein Complexes by AU-rich Element RNA-binding Protein 1 (AUF1) Requires Base-dependent and -independent RNA Contacts. The Journal of biological chemistry.

251. **Zucconi, B. E., and G. M. Wilson.** 2011. Modulation of neoplastic gene regulatory pathways by the RNA-binding factor AUF1. Frontiers in bioscience **16:**2307-2325.

6 Appendix

6.1 Abbreviations

3'ss   3'-splice site

5'ss   5'-splice site

AIDS   acquired immunodeficiency syndrome

ARE   Adenine/uridine-rich element

ARR   alanine-rich region

BALB   Bagg albino

BPS   branch point sequence

BSA   bovine serum albumin

CAT   chloramphenicol acetyltransferase

CCR   C-C chemokine receptor

CD   cluster of differentiation

cf.   lat. "confer" synonymously for compare

CT   C-terminus

CXCR   C-X-C chemokine receptor

DC   dendritic cell

ddH$_2$O   deionised and distilled water

DEPC   diethylpyrocarbonate

DMEM   Dulbecco's modified Eagle's medium

DNA   deoxyribonucleic acid

DNase   desoxyribonuclease

dsDNA	double stranded deoxyribonucleic acid
DTT	dithiothreitol
E.coli	*Escherichia coli*
EDTA	ethylendiaminetetraacetic acid
EJC	exon junction complex
env	gene for the viral membrane protein (envelope)
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
EtBr	ethidium bromide (3,8-Diamino-6-ethyl-5-phenylphenatridiumbromide)
FCS	fetal calf serum
F-MuLV	friend murine leukemia virus
gag	gene for viral structural proteins (group specific antigen)
GAR	guanine-arginine-rich
gp	glycoprotein
GRD	glycine-rich domain
HEK	human embryonic kidney
HeLa	cell-line named after Henrietta Lacks
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hGH	human growth hormone
HIV	human immunodeficiency virus
hnRNA	heterogeneous nuclear RNA
hnRNP	heterogeneous nuclear ribonucleoprotein
HPV	human papilloma virus
ISE	intronic splicing enhancer
ISS	intronic splicing silencer

Appendix

Kd	kilodalton
LB	Luria Broth base
LTR	long terminal repeat
KH	K-homology motif
mAB	monoclonal antibody
MEM	minimum essential medium
mRNA	messenger ribonucleic acid
Mφ	macrophage
NLS	nuclear localisation sequence
NES	nuclear export sequence
nef	negative factor
NMD	nonsense mediated decay
nt	nucleotide
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
Pen/Strep	penicillin / streptomycin
PIC	pre-integration complex
pol	gene for the viral enzymes (polymerase)
PPT	polypyrimidine tract
py	pyrimidine
RBD	RNA binding domain
rev	gene for the viral protein Rev (regulator of viral protein expression)
RNA	ribonucleic acid
RNAse	ribonuclease

RNP	ribonucleoprotein
RRE	rev responsive element
RRM	RNA recognition motif
RS	arginine/serine-rich
RT-PCR	reverse transcription polymerase chain reaction
qRRM	quasi RNA recognition motif
SA	splice acceptor
SDS	sodium dodecyl sulfate
SD	splice donor
siRNA	small interfering ribonucleic acid
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleic protein
SR	serine/arginine-rich
SRE	splicing regulatory element
SRp	serine-arginine-rich protein
SRSF	serine/arginine-rich splicing factor
ssDNA	single stranded deoxyribonucleic acid
ssRNA	single stranded deoxyribonucleic acid
SV40	simian virus 40
tat	gene for viral protein Tat (trans-activator of transcription)
tar	transactivation responsive region
TBE	tris-borate-EDTA buffer
TE	tris-EDTA buffer
Tris	tris-(hydoxymethyl)-aminomethane
U2AF	U2 auxiliary protein

uORF   upstream open reading frame

U snRNA   uridine-rich small nuclear ribonucleic acid

U snRNP   uridine-rich small nuclear ribonucleic protein

UTR   untranslated region

UV   ultraviolet

vif   gene for viral protein Vif (viral infectivity factor)

$^{vol}/_{vol}$ volume per volume

vpr   gene for viral protein Vpr (viral protein R)

vpu   gene for viral protein Vpu (viral protein U)

$^{w}/_{vol}$ weight per volume

6.2 Supplementary material

A

B

C

LTR 4 exon		LTR ex2 ex3 (2xMS2) SD3$_{Down}$	
Viral isoform	**#1544 / #2588**	**Viral isoform**	**#1544 / #2588**
tat1 [1.4]	267	tat1 [1.4]	267
tat2 [1.2.4]	317	tat2 [1.2.4]	317
tat3 [1.3.4]	341	tat3 [1.3.4]	372
tat4 [1.2.3.4]	391	tat4 [1.2.3.4]	422
vpr [1.3E]	654	vpr [1.3E]	685
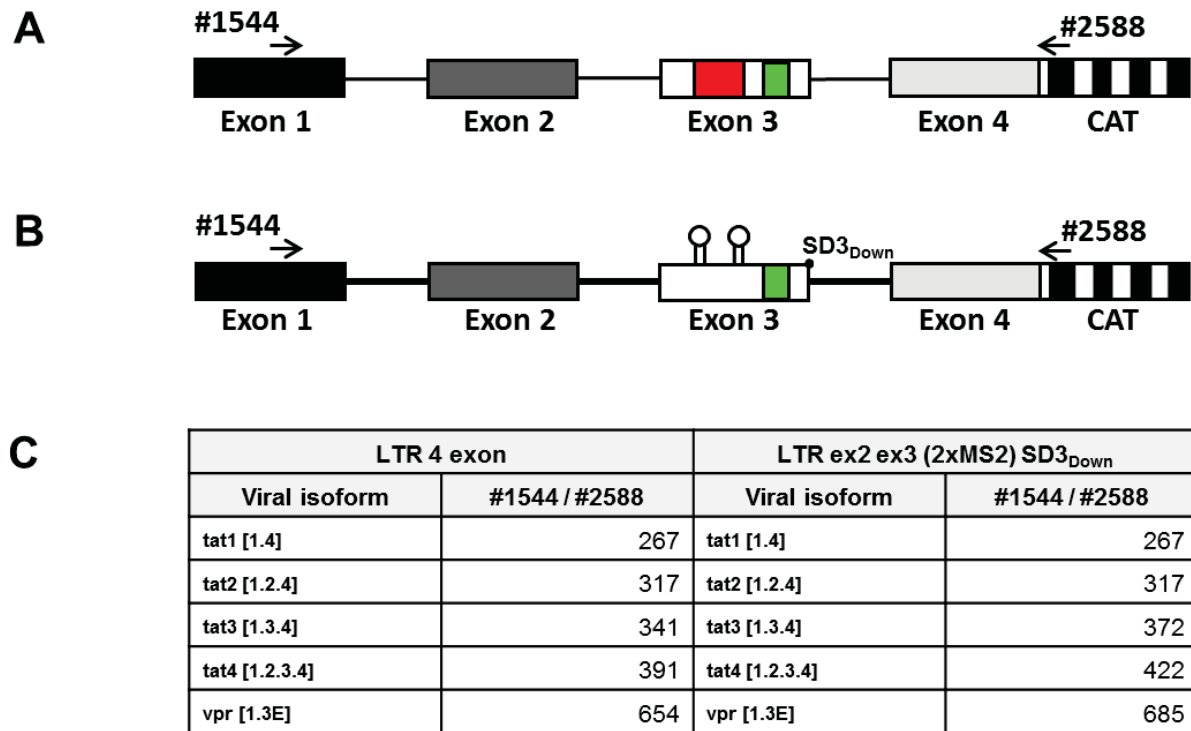
Fig. 6.2 I: Detection of HIV-1 minigene derived mRNA isoforms.

(A) Schematic diagram of HIV-1 derived minigene LTR 4 exon encompassing the first four exons. Exons were depicted as boxes, while the introns sequences were displayed as lines. Splice regulatory elements of HIV-1 exon 3 were highlighted as follows, the ESSV in red and ESE$_{vpr}$ in green. Primer positions were indicated as arrows combined with the corresponding identification number. (B) Schematic diagram of HIV-1 derived minigene LTR ex2 ex3 (2xMS2) SD3$_{Down}$ encompassing the first four exons, but being modified by replacement of the ESSV by two MS2-loops and a decreased complementarity of 5'ss SD3 to the U1 snRNA, termed SD3$_{Down}$. Exons were depicted as boxes, while the introns sequences were displayed as lines. Splice regulatory elements of HIV-1 exon 3 ESE$_{vpr}$ was highlighted in green. Primer positions were indicated as arrows combined with the corresponding identification number (C) PCR-Product sizes obtained by semi-quantitative RT-PCR analysis using the different reporter sets were displayed for each viral splice isoform.

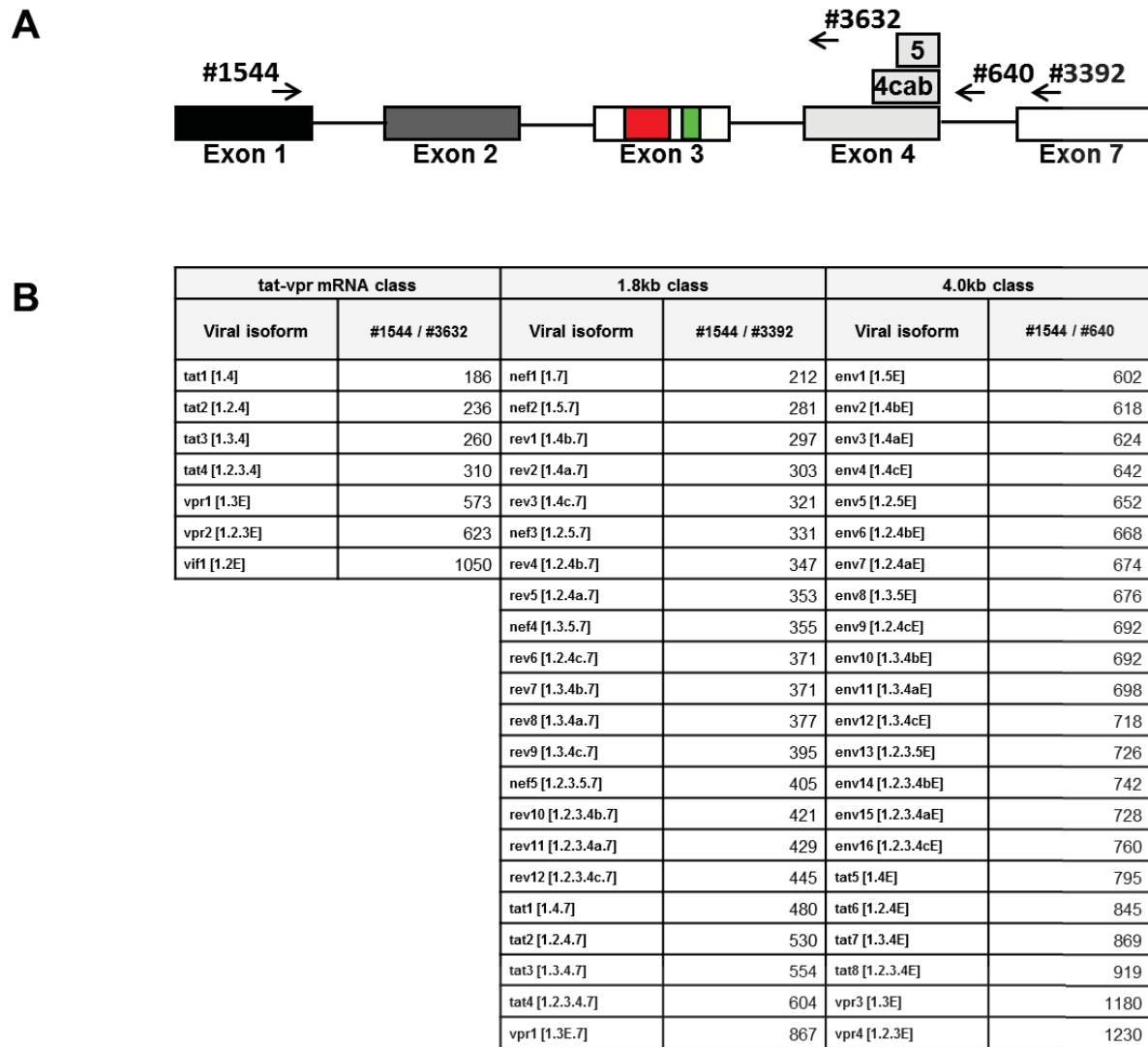| tat-vpr mRNA class		1.8kb class		4.0kb class	
Viral isoform	#1544 / #3632	Viral isoform	#1544 / #3392	Viral isoform	#1544 / #640
tat1 [1.4]	186	nef1 [1.7]	212	env1 [1.5E]	602
tat2 [1.2.4]	236	nef2 [1.5.7]	281	env2 [1.4bE]	618
tat3 [1.3.4]	260	rev1 [1.4b.7]	297	env3 [1.4aE]	624
tat4 [1.2.3.4]	310	rev2 [1.4a.7]	303	env4 [1.4cE]	642
vpr1 [1.3E]	573	rev3 [1.4c.7]	321	env5 [1.2.5E]	652
vpr2 [1.2.3E]	623	nef3 [1.2.5.7]	331	env6 [1.2.4bE]	668
vif1 [1.2E]	1050	rev4 [1.2.4b.7]	347	env7 [1.2.4aE]	674
		rev5 [1.2.4a.7]	353	env8 [1.3.5E]	676
		nef4 [1.3.5.7]	355	env9 [1.2.4cE]	692
		rev6 [1.2.4c.7]	371	env10 [1.3.4bE]	692
		rev7 [1.3.4b.7]	371	env11 [1.3.4aE]	698
		rev8 [1.3.4a.7]	377	env12 [1.3.4cE]	718
		rev9 [1.3.4c.7]	395	env13 [1.2.3.5E]	726
		nef5 [1.2.3.5.7]	405	env14 [1.2.3.4bE]	742
		rev10 [1.2.3.4b.7]	421	env15 [1.2.3.4aE]	728
		rev11 [1.2.3.4a.7]	429	env16 [1.2.3.4cE]	760
		rev12 [1.2.3.4c.7]	445	tat5 [1.4E]	795
		tat1 [1.4.7]	480	tat6 [1.2.4E]	845
		tat2 [1.2.4.7]	530	tat7 [1.3.4E]	869
		tat3 [1.3.4.7]	554	tat8 [1.2.3.4E]	919
		tat4 [1.2.3.4.7]	604	vpr3 [1.3E]	1180
		vpr1 [1.3E.7]	867	vpr4 [1.2.3E]	1230

Fig. 6.2 II: Detection of infectious virus derived mRNA isoforms.

(A) Schematic diagram of HIV-1 genome. Exons were depicted as boxes, while the introns sequences were displayed as lines. Splice regulatory elements of HIV-1 exon 3 were highlighted as follows, the ESSV in red and ESE$_{vpr}$ in green. Primer positions were indicated as arrows combined with the corresponding identification number. (B) PCR-Product sizes obtained by semi-quantitative RT-PCR analysis using the designated primer sets to assess the following subsets of viral transcript isoforms, tat-vpr mRNA class, 1.8kb class and 4kb class.

hnRNP K	Human heterogeneous nuclear ribonucleoprotein K
Score	559
Sequence Coverage	27%
Sequence (identified peptides shown in red)	``` 1 METEQPEETF PNTETNGEFG KRPAEDMEEE QAFKRSRNTD EMVELRILLQ SKNAGAVIGK 61 GGKNIKALRT DYNASVSVPD SSGPERILSI SADIETIGEI LKKIIPTLEE GLQLPSPTAT 121 SQLPLESDAV ECLNYQHYKG SDFDCELRLL IHQSLAGGII GVKGAKIKEL RENTQTTIKL 181 FQECCPHSTD RVVLIGGKPD RVVECIKIIL DLISESPIKG RAQPYDPNFY DETYDYGGFT 241 MMFDDRRGRP VGFPMRGRGG FDRMPPGRGG RPMPPSRRDY DDMSPRRGPP PPPPGRGGRG 301 GSRARNLPLP PPPPPRGGDL MAYDRRGRPG DRYDGMVGFS ADETWDSAID TWSPSEWQMA 361 YEPQGGSGYD YSYAGGRGSY GDLGGPIITT QVTIPKDLAG SIIGKGGQRI KQIRHESGAS 421 IKIDEPLEGS EDRIITITGT QDQIQNAQYL LQNSVKQYSG KFF```

hnRNP D	Human heterogeneous nuclear ribonucleoprotein D
Score	413
Sequence Coverage	26%
Sequence (identified peptides shown in red)	``` 1 MSEEQFGGDG AAAAATAAVG GSAGEQEGAM VAATQGAAAA AGSGAGTGGG TASGGTEGGS 61 AESEGAKIDA SKNEEDEGHS NSSPRHSEAA TAQREEWKMF IGGLSWDTTK KDLKDYFSKF 121 GEVVDCTLKL DPITGRSRGF GFVLFKESES VDKVMDQKEH KLNGKVIDPK RAKAMKTKEP 181 VKKIFVGGLS PDTPEEKIRE YFGGFGEVES IELPMDNKTN KRRGFCFITF KEEEPVKKIM 241 EKKYHNVGLS KCEIKVAMSK EQYQQQQQWG SRGGFAGRAR GRGGGPSQNW NQGYSNYWNQ 301 GYGNYGYNSQ GYGGYGGYDY TGYNNYYGYG DYSNQQSGYG KVSRRGGHQN SYKPY```

hnRNP C	Human heterogeneous nuclear ribonucleoprotein C1/C2
Score	999
Sequence Coverage	27%
Sequence (identified peptides shown in red)	``` 1 MASNVTNKTD PRSMNSRVFI GNLNTLVVKK SDVEAIFSKY GKIVGCSVHK GFAFVQYVNE 61 RNARAAVAGE DGRMIAGQVL DINLAAEPKV NRGKAGVKRS AAEMYGSSFD LDYDFQRDYY 121 DRMYSYPARV PPPPPIARAV VPSKRQRVSG NTSRRGKSGF NSKSGQRGSS KSGKLKGDDL 181 QAIKKELTQI KQKVDSLLEN LEKIEKEQSK QAVEMKNDKS EEEQSSSSVK KDETNVKMES 241 EGGADDSAEE GDLLDDDDNE DRGDDQLELI KDDEKEAEEG EDDRDSANGE DDS```

Fig. 6.2 III: MS-based analysis of proteins derived from primary immune precipitation.

Protein bands derived by primary immune precipitation and *post hoc* SDS-Paige were sliced out, incorporated proteins extracted, fragmented by trypsin and by means of identification subjected to mass spectrometric analysis using ESI-MS spectrometer containing a 3D ion-trap (HCTultra PTM Discoverer System, Bruker Daltonics). For each identified protein the obtained mascot score and sequence coverage were depicted. Identified peptides were displayed within the designated protein sequences and highlighted in red.

6.3   List of publications

Erkelenz, S., Theiss, S., Otte, M., Widera, M., Peter J.O., Schaal H. 2014. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic Acids Res, 42(16):10681-97.

Kaus, A., Widera, D., Kassmer, S., Peter, J., Zaenker, K., Kaltschmidt, C., Kaltschmidt, B. 2010. Neural stem cells adopt tumorigenic properties by constitutively activated NF-kappaB and subsequent VEGF up-regulation. Stem Cells Dev, 19(7): p. 999-1015.

Vizoso Pinto, M.G., Villegas, J. M., Peter, J., Haase, R., Haas, J., Lotz, A. S., Muntau, A. C., Baiker, A. 2009. LuMPIS--a modified luminescence-based mammalian interactome mapping pull-down assay for the investigation of protein-protein interactions encoded by GC-low ORFs. Proteomics, 9(23): p. 5303-8.

6.4   Curriculum vitae

Jan Otto Peter

Address	Neusser Weg 56
	40474 Düsseldorf
Contact information	mobile: 0176 - 29988215
	Email: jan.otto.peter@gmail.com
Date of Birth	02.05.1982
Place of Birth	Düsseldorf
Marital	single

Graduation history

10/09 – 11/15   **PhD thesis**, Institute of Virology, Heinrich-Heine-University, Düsseldorf, „Deciphering principles of recogniton of human immunodeficiency virus (HIV) splice sites"

10/06 – 03/09   **Master of Science** biochemistry and molecular cell biology, Witten/Herdecke University, Title of master thesis, „Analyzation of varicella-zoster virus (VZV) tegumentproteins interactions protein-protein interactions by two newly developed methods", Average: 1,0 (with distinction)

10/03 – 10-06   **Bachelor of Science** biochemistry, Heinrich-Heine-University, Düsseldorf, „Development of an algorithm supporting the diagnosis of hereditary diseases with splice site mutations", Average: 1.9 (very good)

08/99-02/06   **A-levels**, Georg-Büchner-Gymnasium, Düsseldorf, Average: 1.5 (very good)

6.5   Eidesstattliche 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.

Jan Otto Peter

Düsseldorf, den

6.6   Danksagung

Diese Arbeit wäre ohne die Unterstützung vieler Menschen nicht möglich gewesen. Diesen wunderbaren Menschen möchte ich hiermit den letzten Teil dieser Arbeit widmen.

Bereits zu Beginn der Oberstufe am Gymnasium wusste ich, dass ich nach dem Abitur ein naturwissenschaftliches Studium absolvieren würde. Dies führte mich zum Studiengang Bachelor Biochemie an der Heinrich-Heine-Universität. Dort lernte ich bereits im Rahmen meiner Bachelorarbeit die für meinen weiteren Werdegang wohl wichtigste Person kennen, **Prof. Dr. Heiner Schaal**. Schon damals tendierte die Geduld und Unterstützung die mir durch ihn zuteilwurde gegen die Unendlichkeit. Nicht zuletzt deshalb wurde diese sehr schöne als auch produktive Zeit mit dem meinem Bachelorabschluss gekrönt. Schließlich führte mich meine Doktorarbeit zurück an das Institut für Virologie der HHU und damit zur AG Schaal. Jedoch nach einem anfänglich guten Start ließen mich die teilweise katastrophalen privaten Probleme immer mehr vom Weg abkommen. In dieser Zeit hatte **Prof. Dr. Heiner Schaal** immer ein offenes Ohr, aufmunternde Worte und konstruktive Ratschläge für mich parat. Und obwohl auf Grund der Probleme die Doktorarbeit oft in weite Ferne rückte, glaubte er weiterhin fest an mich und spornte mich an mich wiederaufzurappeln und weiterzumachen. Allein deshalb darf ich heute diese letzten Absätze meiner Doktorarbeit schreiben. Allein deshalb geht mein herzlicher Dank an **Prof. Dr. Heiner Schaal**, ohne den ich nie soweit gekommen wäre.

An dieser Stelle möchte ich ganz besonders meinen Kollegen **Anna-Lena Brillen**, **Nora Diehl**, **Dr. Steffen Erkelenz, Isabel Heide**, **Linda Hartmann**, **Hanna Heinrichs**, **Dr. Frank Hillebrand**, **Dr. Claus Lenski**, **Lisa Müller**, **Tomek Ochman**, **Katrin Schöneweis**, **Lara Walotka**, **Björn Wefers** und **Dr. Marek Widera** für ihre Hilfe und Unterstützung danken. Allein die wunderbare Arbeitsatmosphäre in der AG Schaal lies mich stets mit Freude ans Institut für Virologie kommen. Und wenn auch der Tag nicht immer weiterhin so zufrieden stellen verlief wie er begonnen hatte, so waren es gerade die vielen, vielen Gespräche (und das ein oder andere gemeinsame Bier), die das Gewesene einfach wieder verblassen lies und Platz machte für neuen Optimismus. Hierfür möchte ich insbesondere **Dr. Steffen Erkelenz**, **Dr. Frank Hillebrand, Tomek Ochman** und **Dr. Marek Widera** danken, die mich oft durch den gemeinsam verzapften Unsinn, aber auch durch Ihre stete uneingeschränkte Unterstützung nicht an meiner Doktorarbeit verzweifeln ließen.

Weiterhin möchte ich **Marcel Andrée**, **Jens Bauer, Dr. Daniel Digrandi, Dr. Marco Maiwald, Dr. Albert Zimmermann** für Ihre Unterstützung auch von außerhalb der AG Schaal danken. Insbesondere möchte ich **Dr. Albert Zimmermann** für die stets faszinierenden, wenn auch manchmal unappetitlichen Diskussionen beim Essen danken. Leider blieb dabei nach wie vor die Frage unbeantwortet, ob es Spleißen tatsächlich gibt oder nicht doch das Protein vorher da war.

Insbesondere möchte ich hiermit auch **Christiane Cramer**, **Prof. Dr. Ingo Drexler**, **Dr. Inge Krümpelbeck**, **Juliane Noffke** und **Markus Thiel** danken. Denn Sie haben mir von Zeit zu Zeit den Kopf wieder gerade gerückt und mich so wieder auf den richtigen Kurs gebracht.

Danken möchte ich ferner allen Mitarbeiterinnen und Mitarbeitern des Institutes für Virologie und des Institutes für Mikrobiologie für die angenehme Arbeitsatmosphäre. Insbesondere möchte ich dabei den Institutsleitern **Prof. Dr. Hartmut Hengel** und **Prof. Dr. Jörg Timm** für die Möglichkeit danken im Institut für Virologie meine Doktorarbeit fertigzustellen.

Dank gebührt auch außerhalb des Instituts **Axel Auer, Alexander Grunwald, Pfarrerin Ute Kaufmann** und **Pfarrer Daniel Kaufmann**. Ihr habt mir gezeigt, dass es für jedes noch so unlösbar erscheinende Probleme eine passende Lösung gibt. Danke für das Wegbereiten. Danke für das Entzünden und Bewahren der Hoffnung in der Dunkelheit.

Zuletzt, aber vor allem, möchte ich an dieser Stelle meiner Familie danken.

Meine stete Liebe und unendliche Dankbarkeit gebührt dabei **meinen Eltern**, die mich bis über Ihren Tod hinaus noch auf meinem Weg mit ihrer unendlichen Liebe unterstützen und beschützen. Ohne Euch beide liebe **Tine**, lieber **Jobst**, wäre ich nie soweit gekommen.

Danke auch an meine Tanten **Heide Pane**, **Monika Peter** und **Caren Stahl,** meinen Onkel **Rolf Pane**, **Ralf Hertel**, meine Cousine **Kerstin Pane** und meinen Cousin **Sven Pane**. Ohne Euren bedingungslosen Rückhalt und Eure grenzenlose Unterstützung hätten mich die zurückliegenden Ereignisse umgerissen und entwurzelt. Ich verdanke Euch nicht mehr oder weniger als meine Zukunft.