

**Regulation of early and late HIV-1 gene expression by  
alternative pre-mRNA splicing**

**Die Regulation der frühen und der späten HIV-1  
Genexpression durch alternatives prä-mRNA Spleißen**

Inaugural-Dissertation

zur

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

vorgelegt von

Corinna Asang

aus Düsseldorf

April 2010



aus dem Institut für Virologie  
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. H. Schaal

Koreferent: Prof. Dr. D. Willbold

Tag der mündlichen Prüfung: 27.10.2010



## Zusammenfassung

Das alternative Spleißen stellt einen bedeutenden Mechanismus für die zeitlich regulierte HIV-1 Genexpression dar, der für die virale Replikation unentbehrlich ist. Zahlreiche *cis*-wirkende Elemente beeinflussen die Prozessierung der viralen prä-mRNA zu unterschiedlichen mRNA-Isoformen, die die frühe, intermediäre und späte virale Genexpression kennzeichnen. Die vorliegende Arbeit untersuchte spleiß-regulatorische Mechanismen, die entweder die Bildung früher mRNAs oder die Expression genomischer RNA steuern.

Im ersten Teil der Arbeit wird gezeigt, dass ein zuvor identifizierter, mehrere Bindestellen umfassender, purin-reicher exonischer Spleiß-Enhancer (GAR ESE) im Exon 5 durch die bidirektionale Aktivierung der jeweiligen 3' ss und der 5' ss für den Einschluss der internen alternativen Exons 4c, 4a, 4b und 5 in frühe *rev*- und *nef*-mRNAs entscheidend ist. Es wurde erkannt, dass der GAR ESE eine doppelte spleiß-regulatorische Funktion ausübt, indem er (i) den internen Exon-Einschluss durch alle identifizierten Bindestellen für die SR-Proteine SF2/ASF und SRp40 synergistisch steigert und (ii) innerhalb der 3' ss-Gruppe stromaufwärts des Exons 5 die 3' ss A5 ausschließlich durch die proximalen SF2/ASF-Bindestellen spezifisch aktiviert. Die GAR ESE-vermittelte 3' ss Selektivität wird vermutlich durch die phosphorylierungs-abhängige Stabilisierung von U2AF<sup>65</sup> stromaufwärts der 3' ss A5 ausgeübt. Als weitere mögliche Faktoren für den GAR ESE-abhängigen Exon-Einschluss wurden die spleiß-regulatorischen Proteine hTra2- $\beta$ , hnRNP H und hnRNP Q identifiziert. Es wurde beobachtet, dass der interne Exon-Einschluss zusätzlich auf stromabwärts gelegene Elemente des Exons 5, der 5' ss D4 und der E42-Sequenz, angewiesen ist. Dies spricht für ein regulatorisches Netzwerk, das sich über das Exon 5 spannt. Diese Interaktionen sind auch für die Prozessierung intron-haltiger *vpu/env*-mRNAs essentiell. Daher reguliert der GAR ESE das Spleißen der viralen prä-mRNA sowohl in der frühen als auch in der intermediären Phase der HIV-1 Genexpression.

Im zweiten Teil dieser Arbeit offenbarte die Suche nach spleiß-regulatorischen Elementen, die für die Expression genomischer HIV-1 RNA verantwortlich sind, dass die mit der 5' ss D1 überlappenden Bindestellen für die SR-Proteine SC35 und SRp55, die Nutzung dieser Spleißstelle reduzieren. Im Gegensatz dazu verstärkte die Bindung der RS-Domäne von SC35 in einer nicht-überlappenden stromaufwärts gelegenen Position das Spleißen an D1. Dies deutet darauf hin, dass eine Konkurrenz zwischen SR-Proteinen und spleißosomalen Komponenten um die Bindung an den D1 besteht. Dieses Modell wurde durch die Beobachtung erhärtet, dass die weiter stromaufwärts des D1 liegenden Bindestellen für SR-Proteine und hnRNP H die Nutzung der 5' ss nicht beeinflussen. Darüber hinaus wurde beobachtet, dass D1-flankierende Sequenzen die interne 3' ss Auswahl und das Spleißen des distalen Introns beeinflussen. Jedoch reicht der regulatorische Einfluss der dem D1 benachbarten Sequenzen

vermutlich nicht aus, um den beachtlichen Anstieg der Menge vollständig ungespleißter RNA in der späten Phase der viralen Genexpression herbeizuführen. Ein wesentlicher spleiß-regulatorischer Effekt des viralen Proteins Gag auf das Spleißen der HIV-1 mRNA konnte nicht bestätigt werden. Hingegen wurde erkannt, dass das p17-ins Element stromabwärts des D1 durch die Wirkung als intronischer Spleiß-Enhancer das Spleißen an dieser Stelle kontrolliert. Zugleich ist dieses Element für die Rev-abhängige Expression vollständig ungespleißter RNA, aber nicht intermediärer *vpu/env*-mRNAs, entscheidend. Daher scheinen sich intron-haltige intermediäre und späte virale mRNAs in ihren Sequenzvoraussetzungen für die Rev-Abhängigkeit zu unterscheiden. Dies könnte die molekulare Ursache für die verzögerte Expression der vollständig ungespleißten genomischen HIV-1 mRNA sein.

## Summary

Alternative splicing is a major mechanism for temporally regulated HIV-1 gene expression essential for viral replication. Numerous *cis*-acting elements affect viral pre-mRNA processing into distinct mRNA isoforms characterising early, intermediate and late viral gene expression. This thesis investigated splicing regulatory mechanisms controlling either the generation of early mRNAs or the expression of genomic RNA.

In the first part of this work a previously identified purine-rich multisite exonic splicing enhancer (GAR ESE) located in exon 5 was shown to be crucial for inclusion of the alternative exons 4c, 4a, 4b and 5 into early *rev*- and *nef*-mRNAs by bidirectionally activating the respective 3' ss and the 5' ss. The GAR ESE was found to perform a dual splicing regulatory function (i) by synergistically enhancing internal exon inclusion through all identified binding sites for the SR proteins SF2/ASF and SRp40, and (ii) by specifically activating A5 of the 3' ss cluster upstream of exon 5 solely by the proximal SF2/ASF binding sites. GAR ESE-mediated 3' ss selectivity is likely exerted by phosphorylation-dependent stabilisation of U2AF<sup>65</sup> upstream of 3' ss A5. For GAR ESE-dependent exon inclusion the splicing regulatory proteins hTra2- $\beta$ , hnRNP H and hnRNP Q were identified as further candidate factors. Internal exon inclusion was observed to additionally depend on downstream elements of exon 5, i.e. 5' ss D4 and the E42 sequence indicating a regulatory network spanning across exon 5. These interactions are also essential for processing of intron-containing *vpu/env*-mRNAs. Therefore, the GAR ESE regulates viral pre-mRNA splicing in the early as well as in the intermediate phase of HIV-1 gene expression.

In the second part of this thesis, searching for splicing regulatory elements responsible for the expression of genomic HIV-1 RNA revealed that binding sites for the SR proteins SC35 and SRp55 partly overlapping D1 reduce 5' ss efficiency. In contrast, tethering the SC35 RS domain into a non-overlapping upstream position enhanced splicing at D1 indicating a competition of SR proteins and spliceosomal components for binding to D1. This model was substantiated by the finding that binding sites for SR proteins and hnRNP H further upstream of D1 did not affect 5' ss usage. In addition, sequences flanking D1 were found to affect internal 3' ss selection and distal intron removal. However, the regulatory impact of the sequences neighbouring D1 is likely not sufficient to mediate the striking increase in the level of completely unspliced RNA in the late phase of viral gene expression. A significant effect of the viral protein Gag on splicing of the HIV-1 mRNA could not be confirmed. In contrast, the p17-ins element downstream of D1 was found to control splicing at D1 by acting as intronic splicing enhancer. At the same time, this element is essential for Rev-dependent expression of completely unspliced mRNA, but not for intermediate *vpu/env*-mRNAs. Therefore, intron-containing intermediate and late viral mRNAs appear to differ in their sequence requirements for Rev-dependency. This might represent the molecular basis for the delayed expression of completely unspliced genomic HIV-1 mRNA.

## Table of contents

<b>Zusammenfassung</b> .....	<b>I</b>
<b>Summary</b> .....	<b>III</b>
<b>A. Introduction</b> .....	<b>1</b>
<b>A.1 The Human Immunodeficiency Virus Type 1 (HIV-1) – persisting a global health challenge</b> .....	<b>1</b>
<b>A.2 HIV-1 replication critically depends on RNA processing</b> .....	<b>2</b>
A.2.1 Gene expression of HIV-1 .....	3
A.2.2 Alternative splicing of the HIV-1 pre-mRNA ensures complete expression of the viral proteome .....	6
A.2.3 Exploiting and finally overriding the cellular splicing pathway regulates temporal HIV-1 gene expression .....	8
<b>A.3 Mechanism of pre-mRNA splicing</b> .....	<b>9</b>
A.3.1 Spliceosome assembly and catalysis .....	9
A.3.2 Splice site-bridging interactions .....	15
A.3.3 Splice site recognition – finding a needle in the haystack .....	17
A.3.3.1 5'ss recognition .....	18
A.3.3.2 3'ss recognition .....	20
A.3.4 Alternative splice site selection .....	27
A.3.5 <i>Cis</i> -acting splicing regulatory elements .....	31
A.3.6 Splicing regulatory proteins .....	37
A.3.6.1 SR proteins .....	37
A.3.6.2 Heterogeneous nuclear ribonucleoparticle (hnRNP) proteins .....	43
A.3.6.3 Combinatorial control of alternative splicing by splicing regulatory proteins .....	47
<b>A.4 Exploitation of cellular pre-mRNA splicing by HIV-1</b> .....	<b>48</b>
A.4.1 <i>Cis</i> -regulatory elements control splice site recognition in the HIV-1 pre-mRNA .....	48
A.4.2 Modulation of cellular splicing regulatory protein expression and activity after HIV-1 infection .....	51
<b>A.5 Expression of intron-containing HIV-1 mRNAs</b> .....	<b>52</b>
A.5.1 Rev-mediated expression of intron-containing HIV-1 mRNAs .....	53
A.5.2 Inhibitory elements interfere with the expression of intron-containing HIV-1 mRNAs .....	58
<b>A.6 Aim of this work</b> .....	<b>62</b>
<b>B. Materials and Methods</b> .....	<b>63</b>
<b>B.1 Materials</b> .....	<b>63</b>
B.1.1 Chemicals, culture media and solvents .....	63
B.1.2 Enzymes .....	65
B.1.3 Cells .....	65
B.1.3.1 Prokaryotic cells .....	65
B.1.3.2 Eukaryotic cells .....	66
B.1.4 Oligonucleotides .....	66
B.1.4.1 Cloning primer .....	66
B.1.4.2 Sequencing primer .....	68
B.1.4.3 RT-PCR primer .....	68
B.1.4.4 shRNA .....	70
B.1.4.5 <i>In vitro</i> transcription primer .....	70
B.1.5 Recombinant plasmids .....	72
B.1.5.1 HIV-1 derived minigene plasmids .....	72
B.1.5.2 Plasmids for protein expression .....	79
B.1.5.3 Plasmids for recombinant U1snRNA expression .....	80
B.1.5.4 Control plasmids .....	80
B.1.6 Antibodies .....	80
B.1.7 Algorithms and databases .....	82
B.1.7.1 Evaluation of splice site strength .....	82
B.1.7.2 Prediction of SR protein binding sites .....	82
B.1.7.3 2D protein separation and identification .....	82
<b>B.2 Methods</b> .....	<b>84</b>
B.2.1 Cloning .....	84
B.2.1.1 Polymerase-Chain-Reaction (PCR) .....	84

B.2.1.2	Ligation .....	84
B.2.1.3	Transformation .....	85
B.2.1.4	Analytical plasmid DNA isolation .....	85
B.2.1.5	Preparative plasmid DNA isolation .....	86
B.2.1.6	DNA sequencing .....	86
B.2.2	Expression and purification of recombinant T7 RNA polymerase .....	87
B.2.3	Eukaryotic cell culture .....	88
B.2.3.1	Maintenance of eukaryotic cells .....	88
B.2.3.2	Transfection of eukaryotic cells using FuGene 6 .....	88
B.2.3.3	Viral infection .....	89
B.2.3.4	shRNA-mediated knockdown of hnRNP H expression .....	89
B.2.3.5	Inhibition of SR protein phosphorylation .....	90
B.2.4	Immunofluorescence .....	91
B.2.5	Western Blot Analysis .....	92
B.2.5.1	Cell harvesting .....	92
B.2.5.2	Determination of protein concentration .....	92
B.2.5.3	Normalisation of transfection efficiency using a Luciferase assay .....	92
B.2.5.4	Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis (SDS-PAGE) .....	93
B.2.5.5	Western Blot .....	93
B.2.6	RNA isolation .....	93
B.2.6.1	Phenol/chloroform-based isolation of total RNA .....	94
B.2.6.2	Isolation of total RNA using anionic exchange columns .....	94
B.2.6.3	Fractionation of cytoplasmic and nuclear RNA .....	94
B.2.7	RT-PCR .....	95
B.2.7.1	Two-step RT-PCR .....	95
B.2.7.2	One-Step RT-PCR .....	96
B.2.7.3	Native gel electrophoresis followed by EtBr staining .....	96
B.2.7.4	Denaturing gel electrophoresis of fluorescently labelled cDNA .....	97
B.2.7.5	Isolation of cDNA from native polyacrylamide gels .....	97
B.2.8	RNA affinity chromatography .....	98
B.2.8.1	<i>In vitro</i> transcription .....	98
B.2.8.2	Protein isolation by RNA affinity chromatography .....	99
B.2.8.3	<i>In vitro</i> dephosphorylation of HeLa nuclear extract .....	99
B.2.8.4	2D gel electrophoresis .....	101
B.2.9	Protein sequencing by mass spectrometry .....	102
B.2.9.1	In gel digestion and sample preparation .....	102
B.2.9.2	Mass spectrometry .....	102
B.2.10	Microfluidics analyses using Lab-on-a-chip assays .....	103
B.2.10.1	Protein analyses .....	103
B.2.10.2	RNA analyses .....	103
<b>C.</b>	<b>Results .....</b>	<b>104</b>
<b>C.1</b>	<b>Splice site activation of the alternative exons 4c, 4a, 4b, and 5 during early and intermediate HIV-1 gene expression .....</b>	<b>106</b>
C.1.1	Characterisation of the GAR ESE within exon 5 .....	110
C.1.1.1	The GAR enhancer is essential for inclusion of exon 5 into early mRNA isoforms. ....	110
C.1.1.2	Individual SR protein binding sites of the GAR ESE synergistically increase recognition of the alternative exons 4a and 5. ....	114
C.1.1.3	The proximal SF2/ASF binding site specifically contributes to A5 usage. ....	116
C.1.1.4	All SR protein binding sites of the GAR ESE are required to efficiently activate A5 for the generation of intron-containing <i>vpu/env</i> -mRNAs. ....	117
C.1.1.5	A predicted third SF2/ASF binding site covering the intron/exon 5 border ensures preferential activation of A5. ....	118
C.1.1.6	The GAR ESE is essential for activation of each of the competing 3' ss clustering upstream of exon 5. ....	120
C.1.2	The GAR ESE – a key player in a network of regulatory elements across exon 5. ....	124
C.1.2.1	GAR ESE mediated 3' ss selectivity is augmented by an efficient downstream 5' ss ....	124
C.1.2.2	Binding of U1 snRNP at D4 is necessary to activate the 3' ss cluster in Rev-dependent, intron-containing mRNAs .....	128
C.1.2.3	The 3' region of exon 5 contains an additional splicing enhancer. ....	128
C.1.3	The mechanism of GAR ESE enhancer function .....	132
C.1.3.1	The GAR ESE stabilises binding of U2AF <sup>65</sup> at 3' ss A5. ....	132

## Table of contents

C.1.3.2	The GAR ESE does not stabilise spliceosomal components SF1/mBBP and SAP155..	135
C.1.3.3	The GAR ESE promotes binding of the splicing regulatory protein hTra2-β.....	136
C.1.3.4	Identification of additional proteins stabilised by the GAR ESE .....	140
C.1.3.5	U2AF heterodimer binding to the 3' ss ("complex formation") requires a phosphorylated factor. ....	145
C.1.3.6	The impact of splicing regulatory protein phosphorylation on the function of the GAR ESE complex <i>in vivo</i> .....	148
C.1.4	Summary: A functional network emanating from exon 5 and flanking splice sites regulates inclusion of exon 4c, a, b and 5.....	151
<b>C.2</b>	<b>Expression of unspliced RNA during late HIV-1 gene expression .....</b>	<b>153</b>
C.2.1	Identification of <i>cis</i> -acting splicing regulatory elements in the proximity of D1.....	154
C.2.1.1	Individual SR protein binding sites overlapping D1 differentially affect D1 efficiency and alternative 3' ss selection. ....	154
C.2.1.2	Tethering SC35 into an upstream exonic position enhances D1 usage. ....	162
C.2.1.3	hnRNP H binding at exon 1 does not influence splicing at D1.....	165
C.2.1.4	Identification of additional proteins binding to exon 1 and D1.....	168
C.2.1.5	Sequences surrounding D1 do not affect D1 activation, but modulate downstream 3' ss selection and distal intron removal.....	170
C.2.1.6	The viral Gag protein does not affect D1 activation. ....	172
C.2.2	The p17-inhibitory element in the viral pre-mRNA regulates temporal HIV-1 gene expression.....	178
C.2.2.1	The proximal 202 nt of the Gag-ORF are required for Rev-reactivity in the presence of the p17-inhibitory element. ....	178
C.2.2.2	The Rev binding site in stem loop 1 does not contribute to Rev-mediated increase of unspliced RNA.....	184
C.2.2.3	The first 202 nt of the p17-ins element contain an intronic splicing enhancer. ....	187
C.2.3	Summary: Splicing at D1 is oppositely regulated by SR protein binding sites overlapping the 5'ss and the downstream located intronic p17-ins element. ....	189
<b>D.</b>	<b>Discussion.....</b>	<b>192</b>
<b>D.1</b>	<b>The GAR ESE constitutes a main regulatory molecular hub for central splice site activation of the HIV-1 pre-mRNA. ....</b>	<b>192</b>
D.1.1	The GAR ESE is essential for exon recognition. ....	193
D.1.2	Exon recognition is synergistically activated by the SR protein binding sites of the GAR ESE. ....	194
D.1.3	The GAR ESE mediates 3' ss selectivity by its proximal SF2/ASF binding sites. ....	198
D.1.4	Mechanism of GAR ESE function .....	203
D.1.5	A functional network of exon-crossing interactions regulates splicing of the alternative exons 4c, 4a, 4b and 5.....	220
D.1.6	Future perspectives: Internal exon recognition mediated by the GAR ESE .....	224
<b>D.2</b>	<b>Expression of genomic HIV-1 RNA in the late phase of viral replication.....</b>	<b>225</b>
D.2.1	The p17-inhibitory element – a Janus regulator for the expression of genomic HIV-1 RNA .....	227
D.2.2	D1 usage is suppressed by SR proteins binding to the 5' ss.....	237
D.2.3	Sequences neighbouring D1 affect far distant splice site usage – implications for splicing kinetics. ....	239
D.2.4	Future perspectives: Regulation of D1 usage.....	243
<b>E.</b>	<b>Conclusions .....</b>	<b>245</b>
<b>F.</b>	<b>References .....</b>	<b>246</b>
<b>G.</b>	<b>Appendix .....</b>	<b>284</b>
<b>G.1</b>	<b>Experimental data .....</b>	<b>284</b>
<b>G.2</b>	<b>Abbreviations and units .....</b>	<b>288</b>
G.2.1	Abbreviations .....	288
G.2.2	Units .....	289
<b>G.3</b>	<b>Publications.....</b>	<b>290</b>
<b>G.4</b>	<b>Curriculum Vitae .....</b>	<b>291</b>
<b>G.5</b>	<b>Erklärung.....</b>	<b>292</b>
<b>Danksagung .....</b>	<b>293</b>	

## A. Introduction

### A.1 The Human Immunodeficiency Virus Type 1 (HIV-1) – persisting a global health challenge

In 1981, symptoms of immunological dysfunction like generalised lymphadenopathy, opportunistic infections and a variety of unusual cancers observed for a number of patients in the USA were first recognised to constitute a separate disease entity (211, 399, 558). The clinical picture summarised as Acquired Immunodeficiency Syndrome (AIDS) was associated with an infectious etiologic agent through isolation of a retrovirus from a lymph node of a patient with generalised lymphadenopathy often preceding the development of AIDS (33). Additional retroviral isolates displaying very similar biochemical, immunological, structural and cell culture characteristics derived from peripheral blood lymphocytes of patients with AIDS or conditions frequently progressing into AIDS substantiated this correlation (199). The newly identified retrovirus was taxonomically assigned to the *Lentivirus* genus by sequencing of proviral DNA isolated from infected human T-cells of either primary (629) or leukaemia cell line origin (486) and finally named human immunodeficiency virus (HIV) (123). Untreated, HIV-1 infection induces a progressive decline in the numbers and functionality of CD4<sup>+</sup> T cells [reviewed in (160, 213, 623)], resulting in the destruction of the host immune system and in AIDS development (175). In 1986, a second human retrovirus, HIV-2, was isolated from individuals living in West African countries, which is genetically related to HIV-1, but immunologically distant and less pathogenic (115).

The identification of HIV-1 as causative agent of AIDS initiated a plethora of basic and applied biomedical research aiming for therapeutic approaches to oppose viral replication. Despite joint efforts to prevent the spread of the infection, the latest UN report on the statistics of the HIV-1 pandemic estimates that, in 2008, 33 million people were infected with HIV-1 worldwide and that already 27 million deaths were caused by the infection till that time (616, 617). Among the people living with HIV-1, 67% reside in sub-Saharan Africa (616, 617). Considering the limited resources of the public health system in these strongly affected countries, treatments are required that are cost-effective and easy to use. The advent of highly active antiretroviral therapy (HAART) strongly reduced morbidity and mortality of HIV-1 infection (433, 460). Although antiretroviral therapy options have been largely broadened due to the optimisation of pharmaceutical compounds, the increasing percentage of HIV-1 infected people under

treatment elicited the rise of multidrug-resistant viral subspecies (359). Besides the growing problem of drug resistances, treatment with the HAART regime is cost-intensive and requires regular medication.

Generation of a vaccine preventing HIV-1 infection turned out to be complex, which was mainly attributed to a reservoir of chronically infected immune cells, the high diversity of viral structural proteins due to the error-prone reverse transcriptase activity and the viral targeting of the immune cells at the same time required for efficient HIV-specific adaptive immune response (568). Surprisingly, a recent phase III trial conducted in Thailand reported a moderate success in preventing HIV-1 infection (492). However, the statistical significance of the study outcome and the efficacy of the combination vaccine are controversially discussed. In parallel to vaccine research, viral replication has been tackled by a number of promising alternative approaches e.g. evolution of an HIV-1 targeting recombinase excising the proviral genome from the genome of infected HeLa cells (516) or generation of HIV-1 resistant T cells through transduction of hematopoietic precursor cells with a lentiviral vector expressing antiviral RNAs targeting viral regulatory proteins Tat and Rev and, in addition, CCR5, a cellular coreceptor for viral entry (26, 348). In the long term, HIV-1 resistant stem cells expressing therapeutic RNAs are envisioned to be retransplanted into patients after bone marrow irradiation (348). Although a proof of principle has been profoundly demonstrated for these innovative approaches, several challenging problems in the administration of these therapies remain to be solved, until they might be considered for clinical application. However, a particular case recently underlined the clinical importance of CCR5 as antiviral target by the observation that in a patient with HIV-1 infection allogenic transplantation of stem cells homozygous for the CCR5 $\Delta$ 32 mutation restricting coreceptor surface expression prevents viral rebound after discontinuing HAART (273). To continue the design of alternative therapeutic strategies, further progress in the knowledge on HIV-1 replication is required providing the molecular basis to identify potential antiviral targets.

## **A.2 HIV-1 replication critically depends on RNA processing**

The devastating effect of HIV-1 infection is predominantly caused by a decline and loss of function of CD4<sup>+</sup> immune cell populations, especially T lymphocytes thereby severely compromising host immune competence. Besides affecting programmed cell death

(PCD) in infected cells, secreted and surface-bound viral proteins also enhance PCD in non-infected bystander cells thus multiplying the loss of CD4<sup>+</sup> cells (623). The profound impact on immune cell function and lifespan is caused by synthesis of HIV-1 proteins. Besides transcriptional and translational mechanisms, alternative RNA processing crucially regulates lentiviral gene expression. Alternative splicing of the common HIV-1 pre-mRNA is exploited to fulfil two essential functions during viral gene expression by (i) facilitating the complete expression of the viral proteome and (ii) navigating the timely regulated expression of viral genes (312, 317, 413, 475). Disturbing the regulation of viral pre-RNA processing influences infectivity and pathogenesis of HIV-1 [(212, 646), reviewed in (587)], thus underscoring the importance of ordered gene expression for viral replication.

### **A.2.1 Gene expression of HIV-1**

Hallmarks of retroviral replication are reverse transcription of the viral RNA genome generating the proviral DNA and integration into the host cell genome. As a characteristic of lentiviruses, the pre-integration complex (PIC) containing the proviral genome actively invades the interphase nucleus allowing infection of non-dividing cells arrested in the G<sub>1</sub> phase of the cell cycle (Fig. I-1B). Although transcription from unintegrated proviral DNA was observed (585, 652), efficient retroviral gene expression depends on its integration committing HIV-1 gene expression to cellular pathways (532).

The HIV-1 provirus contains a single transcription unit, which is controlled by a promoter located within the 5'-Long Terminal Repeat (5'-LTR). Efficient LTR-mediated transcription by cellular RNA polymerase II (RNAP II) requires the viral protein Tat binding to a stem loop, termed trans-activation response (TAR) region, formed at the 5'-end of all nascent HIV-1 transcripts. Interaction of Tat with positive transcription elongation factor b (P-TEFb), composed of cyclinT1 and Cdk9, induces phosphorylation of the carboxy-terminal domain (CTD) of RNAP II, which stimulates its processivity and in turn facilitates transcription of full-length viral pre-mRNA. Differential expression and phosphorylation of P-TEFb components were suggested to contribute to the postentry block of HIV-1 replication observed in undifferentiated monocytes (158, 358). RNAP II-

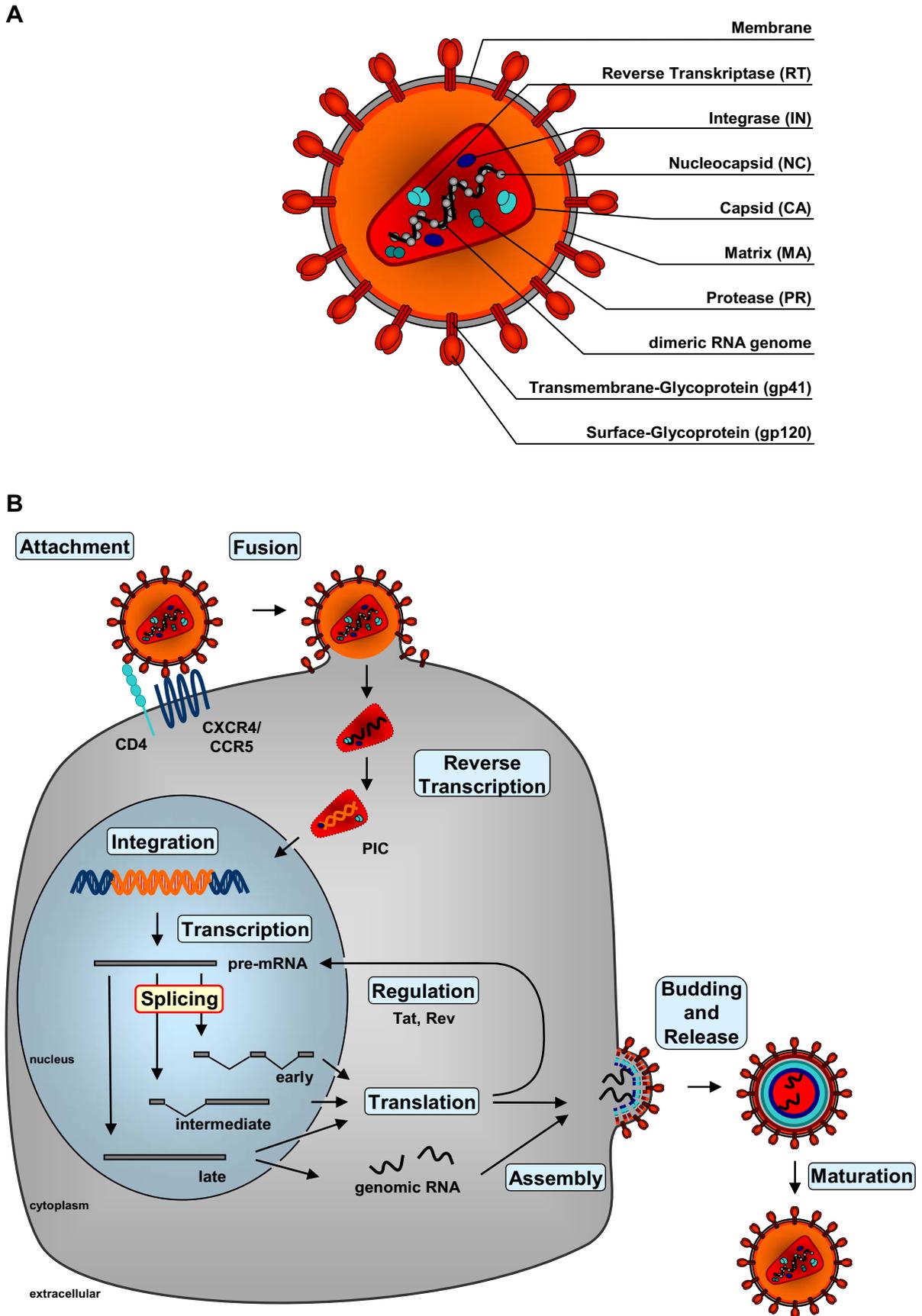


Fig. I-1: The HIV-1 genome integrates into host chromatin and exploits cellular gene expression pathways.

mediated transcription induces capping of the viral pre-mRNA and recruits the cellular polyadenylation machinery to the polyadenylation signal within the 3'-LTR of the full-length transcript. In addition to 5'- and 3'-end processing, full-length pre-mRNA is extensively spliced constituting the basis for temporally controlled expression of HIV-1 encoded enzymes, regulatory, accessory and structural proteins. In the early phase of viral gene expression small regulatory and accessory proteins Tat, Rev, Nef and Vpr are expressed acting on viral gene expression but also on cellular pathways [reviewed in (146, 473, 496)]. In the intermediate phase, Rev initiates the additional expression of the surface and transmembrane glycoprotein precursor gp160 (Env). Finally, core structural components and enzymes required for assembly and release of viral particles are translated as Gag and Gag/Pol precursor proteins from the unspliced mRNA in the late phase. To achieve early, intermediate and late protein expression, the translational capacity of HIV-1 mRNAs is altered by temporally regulated alternative splicing removing interfering start codons and open reading frames (ORFs) from the primary transcript.

---

**Fig. I-1: continued.**

---

**(A)** HIV-1 virion structure. The virus particle is enveloped by a host cell-derived membrane, in which the glycoprotein gp41 is embedded carrying the surface glycoprotein gp120. The Matrix protein is associated with the membrane through N-myristilation. The inner core of the virion is generated by the Capsid protein, which assembles into a cone-like structure. Within the capsid two copies of the (+)-RNA genome are bound by the Nucleocapsid protein. In addition, proteins required for virion maturation and infectivity are located within the capsid.

**(B)** HIV-1 virions productively infect target cells by attachment and fusion to the cell membrane through the initial interaction of the viral surface glycoprotein gp120 with the CD4 receptor (140, 315, 405). Conformational rearrangements mediated by the gp120-CD4 interaction expose the V3 domain of gp120 that interacts mainly with either chemokine receptor CCR5 (9, 152, 162) or CXCR4 (180). Coreceptor binding induces further structural transitions exposing the hydrophobic terminus of the viral transmembrane glycoprotein gp41, which triggers disorganisation of the host cell membrane and fusion with the viral membrane. After entry of the viral capsid into the host cell cytoplasm, the dimeric (+)-strand RNA genome (black) is reverse transcribed generating the proviral DNA (orange) embedded in a pre-integration complex (PIC), which actively invades the nucleus. After integration and transcription, alternative splicing of the viral pre-mRNA generates a heterogeneous population of mRNA isoforms, which is shifted towards intron-containing mRNA isoforms (intermediate/late) during the course of HIV-1 gene expression. Through the interaction of Gag with the viral full-length mRNA, a dimeric RNA genome is incorporated into the virion. After budding from the plasma membrane autocatalytic cleavage of the Gag and Gag-Pol precursors into structural proteins and enzymes results in structural rearrangements condensing the virion.

### A.2.2 Alternative splicing of the HIV-1 pre-mRNA ensures complete expression of the viral proteome

Cap-dependent ribosomal entry and subsequent scanning of the full-length HIV-1 mRNA result in exclusive translation of the Gag- and Gag-/Pol-ORFs located in the upstream half of the viral genome (Fig. I-2A). Therefore, viral genome organisation necessitates mechanisms to release the translational inhibition of the remaining seven genes encoded downstream of the cap-proximal ORFs, which are generally not reached by leaky scanning of the ribosomes. Internal ribosomal entry sites (IRES) identified in

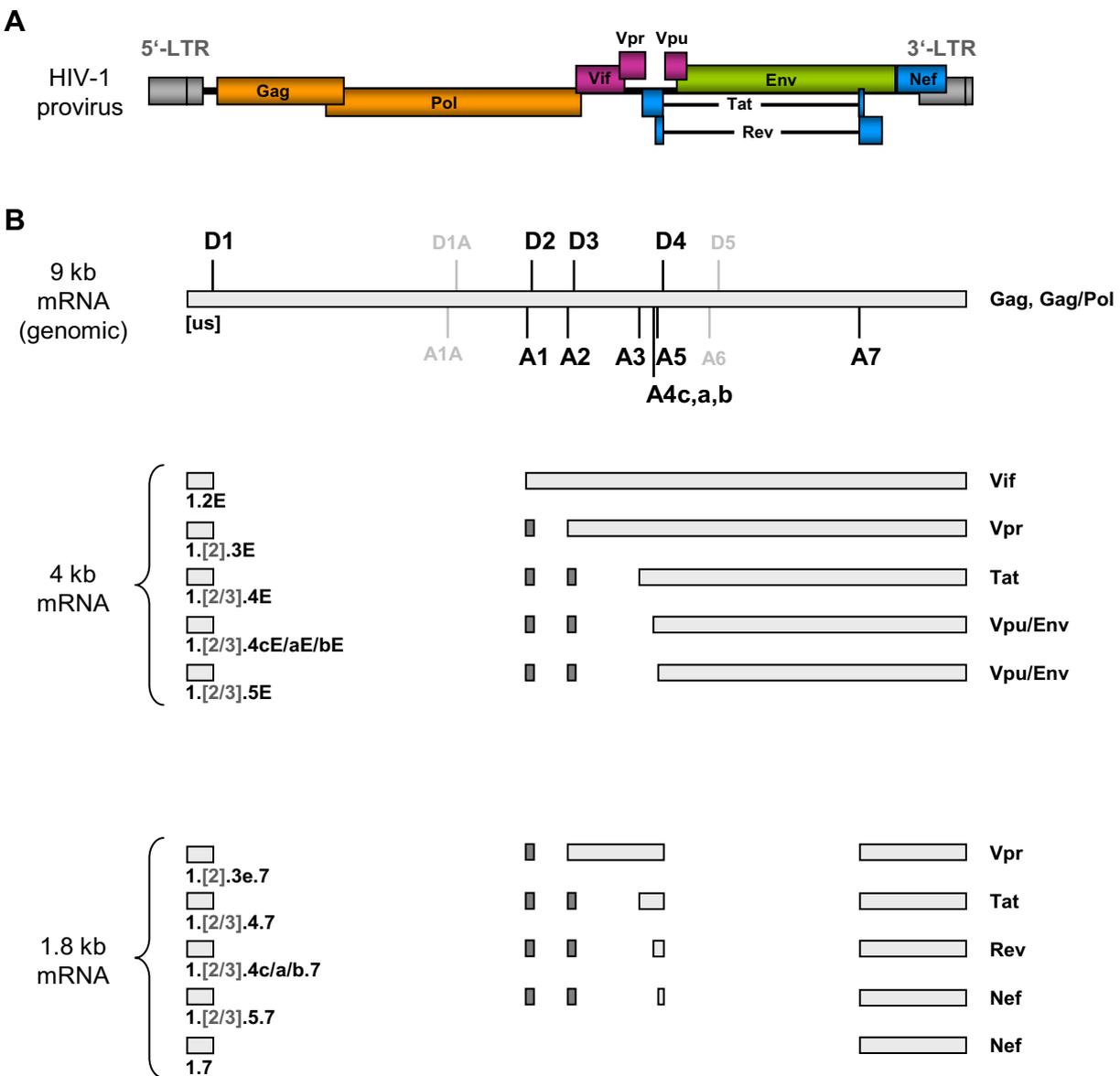


Fig. I-2: Alternative splicing of the HIV-1 pre-mRNA is required to utilise the complete coding potential of the complex retroviral genome.

the 5'-UTR (71) and within the Gag-ORF (77) only allow cap-independent expression of the full-length or a shortened Gag precursor protein after viral inhibition of cap-dependent translation (77) and, thus, are unable to facilitate the expression of downstream encoded viral genes. Although ribosomal frameshifting (281) and leaky scanning of bicistronic mRNAs (329, 528, 531) are additionally employed in the translation of downstream ORFs [reviewed in (65)], alternative splicing of the viral pre-mRNA emerged as the main regulatory mechanism to remove inhibitory translational start codons. Splicing removes intervening sequences (introns) from the pre-mRNA. By combinatorial removal of multiple introns alternative splicing of the HIV-1 pre-mRNA balances relative levels of viral mRNAs such that all structural, enzymatic, regulatory and accessory viral genes are appropriately expressed. To this end, extensive splicing of the viral pre-mRNA generates more than 40 viral mRNA isoforms throughout HIV-1 gene expression comprising the translation capacity for all viral proteins (197, 232, 481, 495, 527-529) (Fig. I-2B). Therefore, alternative splicing of the HIV-1 pre-mRNA essentially ensures complete expression of the viral proteome.

---

**Fig. I-2: continued.**

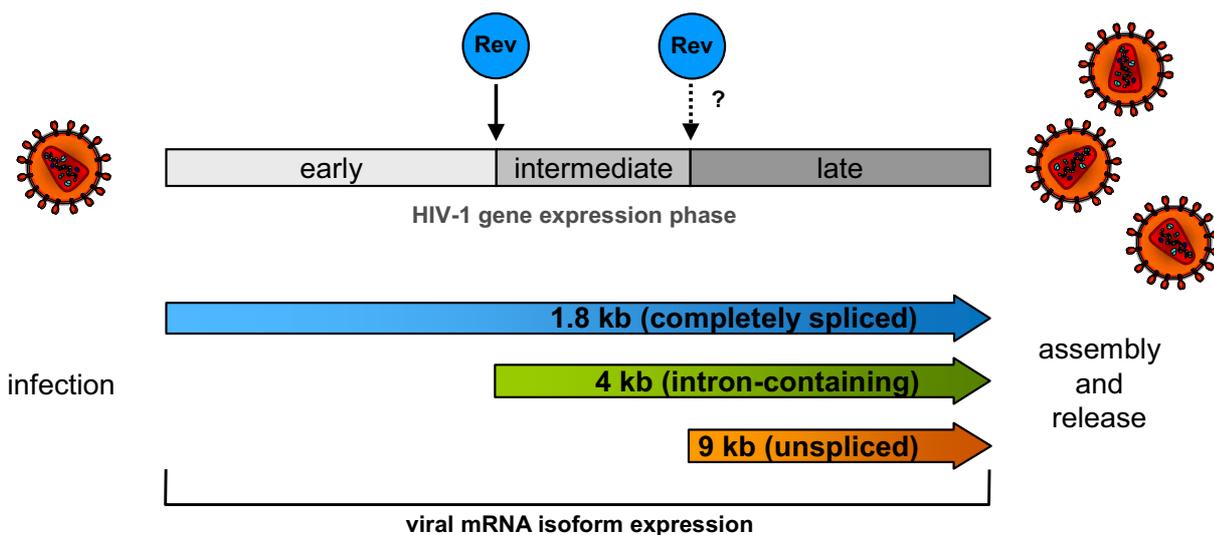
---

**(A)** Schematic depiction of the HIV-1 genome. A single transcription unit is expressed from the promoter located in the 5'-LTR. Nine ORFs (Gag, Gag/Pol, Env, Vif, Vpr, Vpu, Tat, Rev and Nef) encoding 15 proteins in case of the laboratory strain NL4-3 are organised in a highly condensed structure. With the Nef-ORF constituting the only exception, all other ORFs contained within the HIV-1 pre-mRNA strongly overlap or disperse each other. Cap-dependent as well as IRES-mediated ribosomal entry to the unspliced 9 kb pre-mRNA results solely in the expression of Gag and Gag-Pol protein precursors. Open reading frames are indicated by coloured boxes [blue: regulatory proteins (early expression), green: envelope glycoproteins (intermediate expression), magenta: accessory proteins, orange: structural proteins and enzymes (late expression)]. Flanking long terminal repeats (LTRs) are shown as grey boxes.

**(B)** Extensive alternative splicing of the full-length, 9 kb pre-mRNA is required to utilise the complete coding potential of the complex retroviral genome. Alternative usage of multiple 5' splice sites (5' ss: D1 to D5) and 3' splice sites (3' ss: A1 to A7) generates more than 40 mRNA isoforms from the HIV-1 pre-mRNA. Based on their resulting size spliced mRNA isoforms are classified as 4 kb and the 1.8 kb mRNAs. Whereas 4 kb mRNAs are essentially characterised by splicing from the major 5' ss D1 to one of the central 3' ss, 1.8 kb mRNAs are additionally spliced between D4 and A7 in the downstream half of the pre-mRNA. Alternative exons 2 and 3 (dark grey boxes) are alternatively included into mRNA isoforms. The splicing pattern of the respective mRNAs is denoted below the isoform scheme [e: exon 3 extended to D4, E: extended exons, us: unspliced].

### A.2.3 Exploiting and finally overriding the cellular splicing pathway regulates temporal HIV-1 gene expression

Alternative splicing of the HIV-1 pre-mRNA gives rise to three distinct mRNA classes that are classified according to their particular sizes of 1.8 kb, 4 kb and 9 kb. Analysing HIV-1 mRNA expression revealed that the respective mRNA classes are expressed in a defined chronological order (312, 317, 413, 475). The onset of viral gene expression is characterised by exclusive generation of the early 1.8 kb mRNA class through extensive splicing of the viral pre-mRNA (Fig. I-2B). Alternative splice site usage gives rise to at least 23 differentially spliced early mRNA isoforms encoding the accessory proteins Vpr and Nef and the regulatory proteins Tat and Rev (481). Generation of early mRNAs is maintained throughout HIV-1 gene expression (Fig. I-3). In the intermediate phase, Rev induces the additional expression of the intron-containing, 4 kb mRNAs class encoding



**Fig. I-3: Alternative splicing of the HIV-1 pre-mRNA temporally shifts viral gene expression from early to late phase by gradual intron retention.**

Following integration of the HIV-1 proviral DNA into the cell genome, early regulatory proteins are expressed from extensively spliced viral mRNA isoforms of the 1.8 kb class (blue). The expression of Rev profoundly shifts the splicing pattern of the viral pre-mRNA to intron-containing mRNA isoforms (148, 178) thereby initiating the intermediate phase of viral gene expression. Intermediate 4 kb mRNA isoforms (green) retain the *vpu-env*-coding sequence located in the 3'-region of the viral pre-mRNA. The expression of completely unspliced 9 kb genomic RNA (orange) encoding structural components and enzymes required for viral assembly and release is further delayed. Although the presence of Rev is a prerequisite for the expression of all intron-containing HIV-1 mRNAs, it remains to be demonstrated whether Rev decisively initiates the transition from intermediate to late viral gene expression.

the accessory proteins Vif and Vpr and the glycoprotein precursor gp160 (Env). Finally, in the late phase of viral gene expression an increasing fraction of the HIV-1 pre-mRNAs remains completely unspliced (312) encoding the precursor proteins Gag and Gag/Pol and also serving as viral genome during virion assembly. The progressive changes in the viral pre-mRNA splicing pattern from early to late HIV-1 gene expression shift the viral mRNA population towards isoforms with increasing intron content.

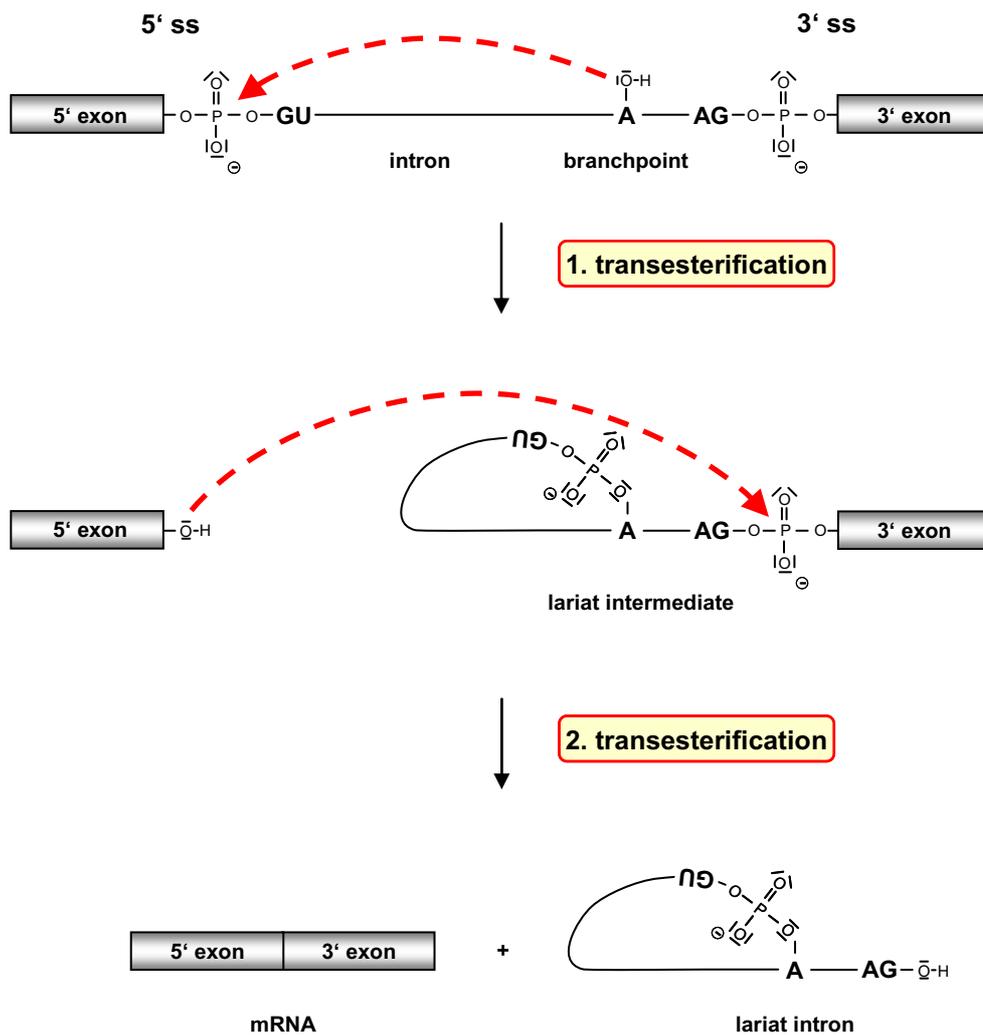
The main regulatory mechanism for the expression of intron-containing mRNAs identified so far is the interaction of the viral regulatory protein Rev with a structural RNA element within the *env*-coding sequence in the downstream half of the viral pre-mRNA termed Rev-responsive element (RRE) (141, 680) (120, 138, 250, 313, 389). The Rev-RRE-mediated nuclear mRNA export was the first mechanism described to circumvent the cellular requirement of intron removal prior to nucleocytoplasmic export (168, 179, 388). However, the underlying viral mechanisms interfering with cellular splicing known to date fail to explain the intermediate to late shift from intron-containing, but spliced 4 kb mRNA to completely unspliced genomic RNA.

### **A.3 Mechanism of pre-mRNA splicing**

Most metazoan genes are dispersed by non-coding sequences (introns), which are removed from the pre-mRNA by splicing, a process involving two sequential transesterification reactions. The splicing reaction is performed by the cellular spliceosome, a macromolecular ribonucleoprotein (RNP) machine, which identifies introns by interacting with sequence elements within the RNA, i.e. the 5' splice site (5' ss) defining the upstream and the 3' splice site (3' ss) defining the downstream border of the intron [reviewed in (649)].

#### **A.3.1 Spliceosome assembly and catalysis**

The spliceosome positions the intron-flanking splice sites for an RNA-based catalysis of the splicing reaction, which removes the intervening intron and ligates the adjacent sequences (exons) (Fig. I-4). Like their human equivalents, HIV-1-encoded splice sites are recognised by the major U2-dependent spliceosome, which catalyses splicing of



**Fig. I-4: The spliceosome catalyses two sequential transesterification reactions removing an intron from the pre-mRNA.**

In the first transesterification reaction, the oxygen of the 2'-hydroxyl group provided by the branch point adenosine (A) attacks the phosphorus atom that links the 5'-exon and the intron (top). In a nucleophilic substitution a lariat/3'-exon intermediate is generated and the bond to the 5'-exon displaced. In the second transesterification reaction, the oxygen atom of the nascent hydroxyl group of the 5'-exon forms a bond with the phosphorus atom that links the intron and the 3'-exon (middle) thereby substituting the bond between the lariat intron and the 3'-exon and ligating both exons to form the mRNA (bottom).

canonical GT-AG pre-mRNA introns constituting the vast majority of metazoan pre-mRNA introns. The major spliceosome consists of five uridine-rich small nuclear ribonucleoprotein particles (U snRNPs), i.e. U1, U2, U4/U6 and U5 snRNP, and a number of non-snRNP proteins [reviewed in (83, 289)]. Each U snRNP is composed of

one or two (U4/U6) U snRNAs complexed with seven common Sm proteins designated B, D1, D2, D3, E, F and G, or the homologous group of like-Sm (Lsm) proteins in case of U6 snRNP and several particle specific proteins [for a review see (649)]. Numerous *in vitro* analyses resulted in the model of ordered spliceosome assembly suggesting that the spliceosome is individually assembled for removal of each intron by sequential binding of preassembled U snRNPs and non-snRNP proteins [reviewed in (76)]. The isolation of preassembled penta-snRNPs from HeLa nuclear extracts suggested the existence of an alternative pathway for spliceosome assembly (381, 584). However, since the functionality of this complex in pre-mRNA splicing still remains elusive (289), ordered assembly currently represents the prevailing model for spliceosome formation.

Prior to catalysis of the transesterification reactions, spliceosomal components initially recognise, pair and position the splice sites, which is essential for correct splicing of the pre-mRNA substrate. The RNA elements that are identified by the spliceosome are defined by short consensus motifs with the upstream 5' ss containing a conserved GU-dinucleotide and the 3' ss consisting of three sequence elements, i.e. the branch point sequence (BPS), the polypyrimidine tract (PPT) and the AG-dinucleotide. In *in vitro* splicing systems, intermediates of ordered spliceosome assembly are defined by sequential association and release of U snRNPs and the state of the pre-mRNA substrate regarding the two-step chemical splicing reaction (Fig. I-5) [reviewed in (76, 649)]. Spliceosome assembly is a highly dynamic process employing numerous compositional and conformational rearrangements [reviewed in (580, 628)]. Spliceosome assembly starts with ATP-independent formation of the E complex (414), which contains U1 snRNP bound to the 5' ss through base pairing interactions of the 5' end of the U1 snRNA with the 5' ss and non-snRNP proteins SF1/mBBP (splicing factor 1/mammalian branch point binding protein) and U2AF (U2 snRNP auxiliary factor) recognising the BPS and the PPT of the 3' ss, respectively. Formation of this early splicing complex brings the 5' ss, the BPS and the 3' ss into proximity within a structured conformation (305, 306). The E' complex lacking U2AF can precede this complex (307). In the presence of ATP stable association of U2 snRNP with the BPS displaces SF1/mBBP converting the E complex into the A complex. U2 snRNA forms a duplex with the BPS, in which the bulging of an adenosine determines the branch-site acting as nucleophile in the first transesterification reaction (483). Subsequently, recruitment of the U4/U6\*U5 tri-snRNP links U1 and U2 snRNPs forming the B complex. Since U6 snRNA forms a RNA-RNA duplex with the 5' ss and at the same time with

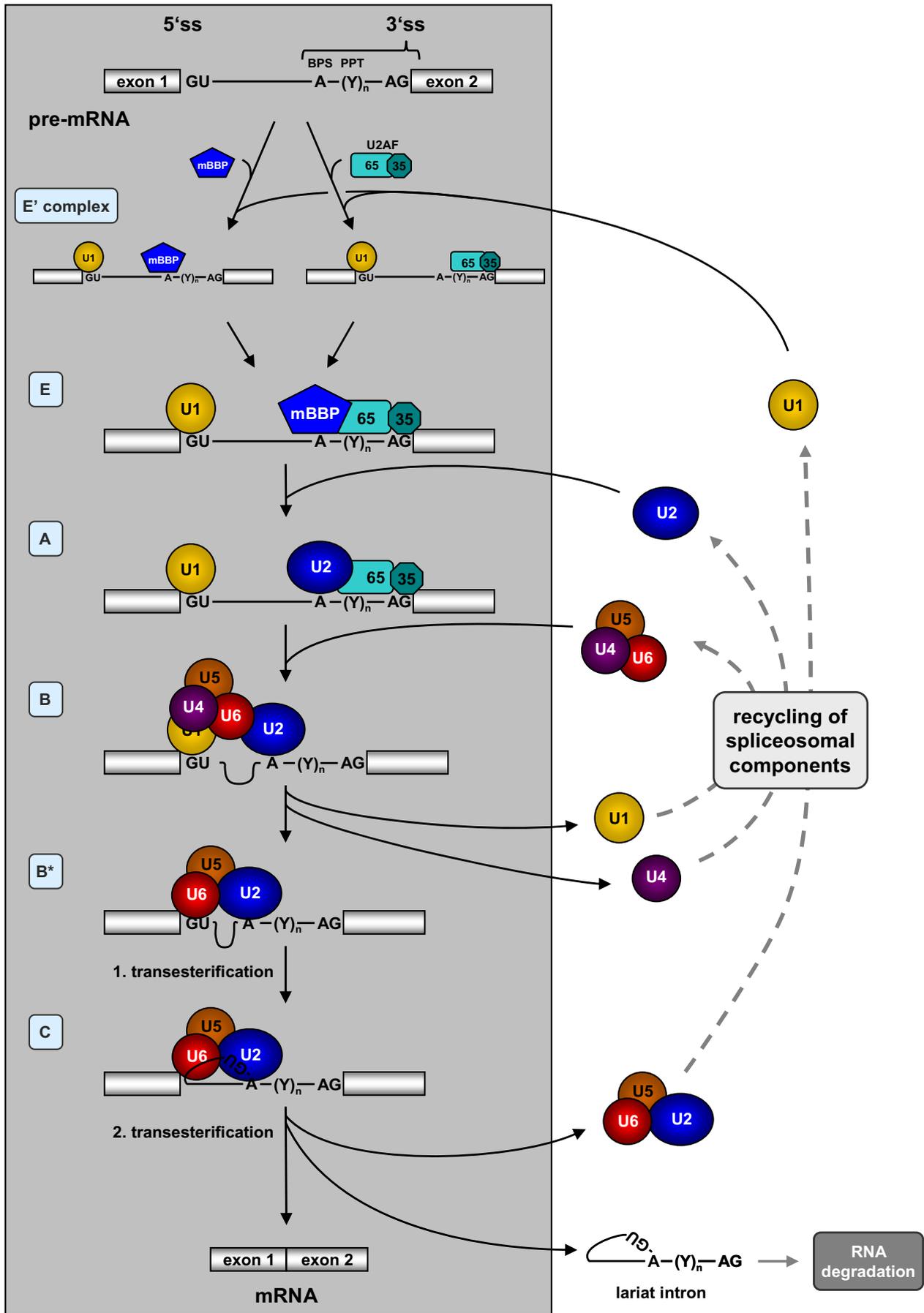


Fig. I-5: The spliceosome assembles in a stepwise manner on each intron of the pre-mRNA.

U2 snRNA attached to the BPS, U6 snRNP is proposed to juxtapose the intron-flanking splice sites. Cross-intron communication between the 5' ss and the 3' ss has been shown to be additionally promoted by non-snRNP splicing factors of the serine/arginine-rich (SR) protein superfamily (68, 193, 259, 655) and the SR-related splicing coactivator complex SRm160/300 (62, 167). Although the B complex includes all components required for splicing, a series of ATP-dependent rearrangements resulting in the dissociation of U1 and U4 snRNAs is required to catalytically activate the spliceosome (B\* complex) [reviewed in (76, 628)]. The activated spliceosome conducts the first transesterification reaction generating the C complex. Prior to the second catalytic step further rearrangements reposition the catalytic centre of the spliceosome (319). Coincident with the second step of the splicing reaction, factors involved in quality control, export and translation are recruited to the mRNA forming the exon junction complex (EJC), which is positioned 20-24 nt upstream of the generated exon-exon junction [(344, 688), reviewed in (342)]. After the second transesterification reaction the spliceosome dissociates releasing the mature mRNA complexed by RNA binding proteins.

---

**Fig. I-5: continued.**

---

Each intron is flanked by the 5' ss, an 11 nt sequence harbouring an invariant GU-dinucleotide at the beginning of the intron, and a tripartite 3' ss consisting of the branch point sequence (BPS), the downstream located polypyrimidine tract (PPT) and the invariant AG-dinucleotide marking the 3' end of the intron. Initially, the 5' ss and the 3' ss are recognised by spliceosomal U1 snRNP (U1) and a complex of the non-spliceosomal proteins SF1/mBBP (mBBP) and both subunits of U2AF (65 and 35), respectively, generating the E complex. Partial complexes lacking either U2AF (E' complex) or SF1/mBBP can precede the E complex. All further compositional and conformational changes in the complex require the presence of ATP or ATP as well as GTP. The E complex evolves into the A complex by recruitment of U2 snRNP (U2) to the BPS displacing SF1/mBBP. Subsequent entry of the U4/U6\*U5 tri-snRNP (U4/U5/U6) generates the B complex, which is activated (B\* complex) by structural rearrangements causing the release of U1 snRNP as well as U4 snRNP from the spliceosome. The first transesterification reaction generates a lariat intermediate, which is detached from the 5'-exon, but is held in proximity of the lariat intermediate by the C complex of the spliceosome. After the second transesterification reaction the spliceosome disassembles and releases the mature mRNA. The released lariat intron is degraded by debranching enzymes and exonucleases, whereas spliceosomal components are recycled to join a further splicing reaction. Spliceosomal complexes are named according to metazoan nomenclature. U snRNPs are simplified as ellipses [A: branch point adenosine, PPT: polypyrimidine tract ,(Y)<sub>n</sub>].

The initial recognition of 5' ss, BPS and PPT in the E complex occurs in the absence of ATP. Therefore, binding of U1 snRNP, SF/mBBP and U2AF to the particular splice site elements was expected to be reversible. However, it was found that after E complex assembly addition of excess pre-mRNA is unable to efficiently compete with already complexed pre-mRNA indicating that ATP-independent E complex formation commits the pre-mRNA to the general splicing pathway (414, 539). The observation that general commitment to the splicing pathway is already achieved in the early E' complex indicates that U2AF is dispensable for general splicing commitment (307). Additionally, the same study proposed that U1 snRNP-mediated commitment to the general splicing pathway might be performed prior to the base-pairing interaction of U1 snRNA with the pre-mRNA. In contrast to the general commitment to splicing, pairing of particular splice sites appears not to be definite prior to ATP-dependent formation of the prespliceosomal A complex (323, 355). Therefore, commitment to spliceosome assembly appears to be distinct from commitment to splice site pairing. Although the A complex determines the exons that are joined in the splicing reaction, the precise AG-dinucleotide participating in exon ligation is selected after the first catalytic step of the splicing reaction via a scanning mechanism (17, 18, 111, 563).

To avoid the generation of mRNAs encoding misfunctional or even deleterious proteins the spliceosome has to identify and remove pre-mRNA introns with high accuracy. Prerequisites for exact intron removal are the fidelity of splice site recognition, splice site pairing in the presence of multiple candidate splice sites and nucleotide selection in both transesterification reactions including repositioning of the catalytic centre. The fidelity of these processes is ensured by multiple mechanisms [reviewed in (489)]. Each of the fidelity mechanisms contains the potential to regulate splice site selection and splicing efficiency.

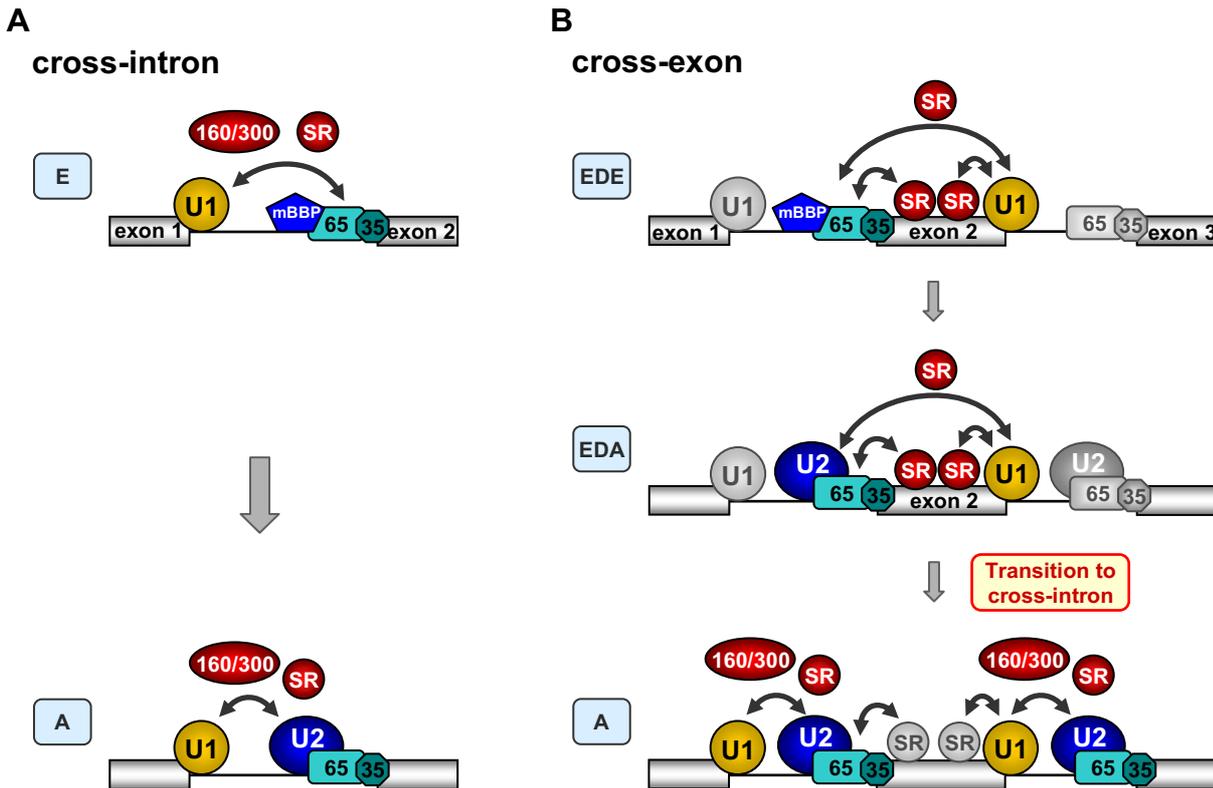
Metazoan splice sites are mainly recognised by the interplay of spliceosomal components binding to less conserved splice sites with auxiliary proteins interacting with splice site-proximal *cis*-acting splicing regulatory elements [reviewed in (564)]. The nuclear concentration and functionality of proteins interacting with splicing regulatory elements thus may strongly influence splice site recognition. In contrast to the nearby situated regions mediating splice site recognition, pairing of the correct splice sites in pre-mRNAs encoding multiple introns crucially depends on the exon/intron architecture in combination with the presence of numerous *cis-acting* splicing regulatory elements in a larger region of the pre-mRNA. After splice site pairing a number of ATP-dependent

rearrangements are required to generate the catalytic core of the spliceosome. The precision of nucleotide selection during the subsequent biochemical reactions is ensured by recurrent 5' ss and 3' ss identification through different spliceosomal components during spliceosome assembly and conformational rearrangements guiding both catalytic steps of the splicing reaction within the spliceosome [reviewed in (489)]. For the fidelity of the splicing reaction during dynamic remodelling of the spliceosome a kinetic-proofreading model was proposed (84, 85), in which a number of consecutive structural transitions mediated by ATP hydrolysis of several DExH/D-box ATPases require RNA substrates to fit into the catalytic core of the spliceosome [reviewed in (565)]. Suboptimal splicing substrates that do not meet the structural demands of the remodelled spliceosome conformation are discarded from the splicing pathway. Whereas ATP-dependent rearrangements of the spliceosome mainly ensure exclusion of suboptimal splice sites, additional spliceosomal proteins like SPF45 were implicated in proofreading 3' ss selection in the second transesterification reaction (338). In addition, hSlu7 has been proposed to contribute to the regulation of the second catalytic step of the splicing reaction by holding the detached upstream exon in the proper position for exon ligation (114).

Spliceosome assembly and catalysis have been shown to be strongly controlled by phosphorylation (574). The assembly of spliceosomal complexes is inhibited by phosphatases like PP1 (410) indicating that phosphorylated factors contribute to spliceosome assembly. In contrast, the requirement of the phosphatases PP1 and PP2A for the second step of the splicing reaction (552) suggests that dephosphorylation of components within the assembled spliceosome is needed for the second transesterification reaction. Together these results suggest that consecutive phosphorylation and dephosphorylation are required for pre-mRNA splicing.

### **A.3.2 Splice site-bridging interactions**

Spliceosome assembly *per se* requires an interaction across the respective intron (Fig. I-6A). However, direct juxtaposition of intron-flanking splice sites appears to be constrained by the exon/intron architecture of the pre-mRNA. In *Saccharomyces cerevisiae* only 3.8% of annotated genes contain introns (366) and less than 10 genes encode more than one intron (366, 576). In addition, the majority of yeast introns are smaller in size [ $<500$  nt (576)] than their flanking exons. For those yeast pre-mRNA



**Fig. I-6: Splice site-bridging interactions.**

**(A)** Cross-intron interactions. The splicing reaction requires interactions across the intron to position the splice sites into proximity. Prior to U4/U6\*U5 tri-snRNP entry into the spliceosome, regulatory factors like SR proteins (SR) as well as the splicing coactivator complex SRm300/160 (300/160) are able to mediate bridging of U1 snRNP and components recognising the 3' ss in spliceosomal complexes E and A [U1: U1 snRNP, U2: U2 snRNP, 35: U2AF<sup>35</sup>, 65: U2AF<sup>65</sup>].

**(B)** Cross-exon interactions. For short mammalian exons interactions across the exon precede those across the intron. Interactions are frequently mediated by SR proteins, which stabilise binding of spliceosomal components bound to the exon-flanking splice sites either by RNA-dependent or -independent mechanisms [EDE/EDA: exon-defined E/A complex].

substrates containing only short introns, direct assembly of the spliceosome across the intron had been observed. Because of the small size of most yeast introns, it was suggested that the majority of splice site pairs are brought into proximity by direct interactions across the intron (intron-definition) (42), which was proposed to apply for introns containing less than 250 nt [(188), reviewed in (256)].

Compared to *S. cerevisiae*, most metazoan pre-mRNAs exhibit a more complex exon/intron architecture harbouring multiple expanded introns typically ranging in size from  $10^3$ - $10^4$  nt, whereas more than 95% of internal vertebrate exons are smaller than 300 nt (42). Contrasting prevalent intron-definition initiating spliceosome assembly in yeast, pre-spliceosomal interactions across expanded metazoan introns are preceded by interactions across the flanking exons (exon-definition) (Fig. I-6B) [(494, 600), reviewed in (42)]. An A-like complex is proposed to establish across the exon, in which U2 snRNP and U1 snRNP recognising splice sites of distinct introns participate. In terminal 5' exons the cap binding complex (CBC) recognising the 7-methylguanosine triphosphate ( $m^7Gppp$ ) cap structure substitutes for U2 snRNP in defining the first exon (125, 347, 683). The exon definition model was substantiated by the finding that U1 snRNP bound to the 5' ss of a downstream located intron promotes 3' ss recognition of the upstream intron (260). A distance of more than 48 nt between the 3' ss and the downstream 5' ss is required for efficient U1 snRNP-mediated activation of splicing of the upstream intron (274). Exon length and splice site strength appear to act as additive factors in exon-definition (155, 156). Inspired by the increased size of introns compared to exons, it was assumed that the initial interaction between spliceosomal components recognising the splicing signals occurs via the shorter distance [(601), reviewed in (42)]. Therefore, in the case of long introns and short exons in vertebrates the first contact between splice sites likely takes place across the exon. To achieve splicing of the intron-flanking splice site pair, cross-exon interactions subsequently have to promote cross-intron interactions juxtaposing the neighbouring exons. However, the mechanistic details and the molecular forces driving the restructuring process from cross-exon to cross-intron defined pre-spliceosomal complexes have still to be resolved in more detail.

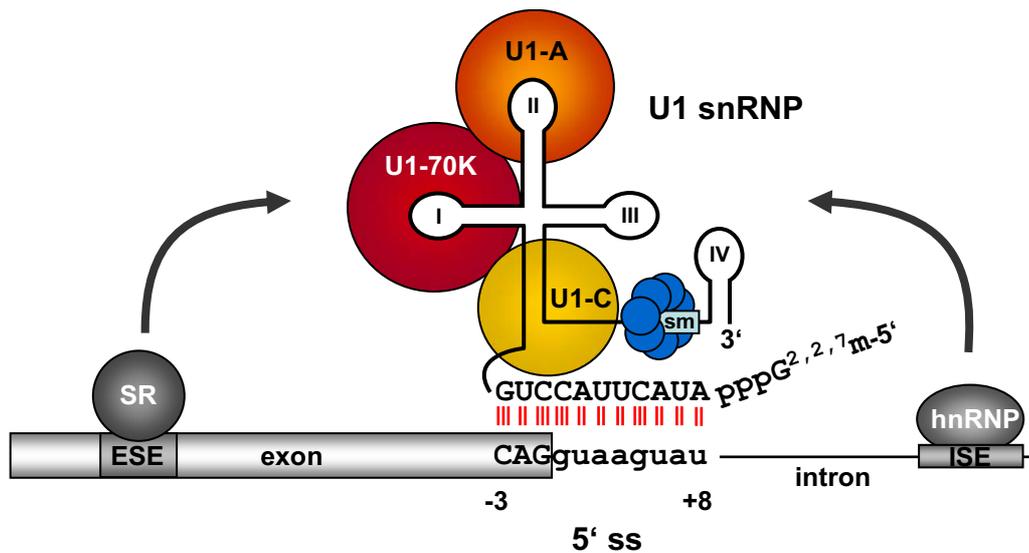
### **A.3.3 Splice site recognition – finding a needle in the haystack**

In contrast to the strongly conserved splice sites found in yeast, most metazoan splice sites are highly degenerated from the consensus motifs (83). In line with viral adaptation to the human host, HIV-1 splice sites are likewise found to cover a wide range of deviation from consensus splice site sequences resulting in considerable differences in intrinsic splice site strength (190, 296, 447). Low splice site conservation raises the problem to identify the borders of the large metazoan introns. Nevertheless, also metazoan splice sites are recognised with high accuracy (189). In a number of well-

studied examples the precision of splice site recognition is ensured by *cis*-acting splice site-proximal RNA sequences [reviewed in (637)]. Proteins binding to these regulatory sequences enhance splice site usage by stabilising the interaction of spliceosomal components with the splice sites. In contrast, splice site usage can also be suppressed by proteins interfering with splice site recognition by spliceosomal components. Overall a dynamic model of splice site recognition emerged, in which splice sites and splice site-proximal regulatory sequences cooperate to initiate spliceosome assembly. The importance of correct splice site identification is underlined by the finding that a number of diseases are associated with malfunctions of splice site recognition [reviewed in (29, 130, 606)].

### **A.3.3.1 5'ss recognition**

5' ss are composed of a single sequence element covering the exon/intron border. As one of the earliest steps in spliceosome assembly, the 5' ss is recognised by U1 snRNP, consisting of the U1 snRNA complexed by the common U snRNP core structure and the three U1 snRNP-specific proteins U1-70K, U1-A and U1-C (75, 440). Identification of the 5' ss occurs through formation of an RNA duplex between the 11 single stranded nucleotides at the 5' end of the U1 snRNA and the 5' ss sequence (694) (Fig. I-7). Determined by the position of the AC-dinucleotide of the U1 snRNA that base pairs with the invariant GU-dinucleotide at intronic position +1 and +2 of the 5' ss, the free 5'-end is capable of binding nucleotides from position -3 to +8 relative to the intron-exon border. Due to the recognition of the 5' ss via RNA duplex formation, the intrinsic 5' ss strength during early complex assembly depends its nucleotide complementarity to the free 5'-end of the U1 snRNA. Although a 5' ss consensus motif reveals a conservation of 8 nucleotides from position -2 to +6 relative to the exon/intron border, U1 snRNP-mediated minigene expression studies and RT-PCR analysis revealed that even six consecutive complementary nucleotides are already sufficient to mediate U1 snRNP binding and splicing (190, 191, 295). Furthermore, sequence analysis of 46,308 annotated human canonical 5' ss demonstrated that, despite the consensus motif generated by all 5' ss analysed, the 5' ss data set could be classified into three subgroups showing either exonic, centred or intronic complementarity to U1 snRNA (246). The observations converge into a model, in which all 11 nucleotides of the 5' ss can contribute to duplex formation with the U1 snRNA (295). Theoretically, to



**Fig. I-7: The complementarity to U1 snRNA defines the intrinsic 5' ss strength.**

U1 snRNP recognition of an intrinsically strong 5' ss displaying full complementarity to the 5' end of the U1 snRNA. The U1 snRNP core is constituted by the U1 snRNA shapening into a cloverleaf-like secondary structure with four stem loops (I-IV). Three U1 snRNP-specific proteins (U1-A, U1-70K and U1-C) and seven Sm proteins (blue circles) common to several U snRNPs bind to the U1 snRNA. The single-stranded 11 nt at the free 5'-end of the U1 snRNA are able to form hydrogen bonds (red vertical bars) with complementary bases of the 5' ss. The intrinsic strength of a 5' ss is determined by the number and continuity of bases complementary to U1 snRNA. In addition to RNA-RNA interaction, U1 snRNP can be further stabilised by protein-protein interactions between U1 snRNP-specific proteins and splicing regulatory proteins (SR, hnRNP) binding upstream or downstream of the 5' ss [ESE: exonic splicing enhancer, hnRNP: heterogeneous nuclear ribonucleoprotein, ISE: intronic splicing enhancer, SR: arginine-serine-rich protein].

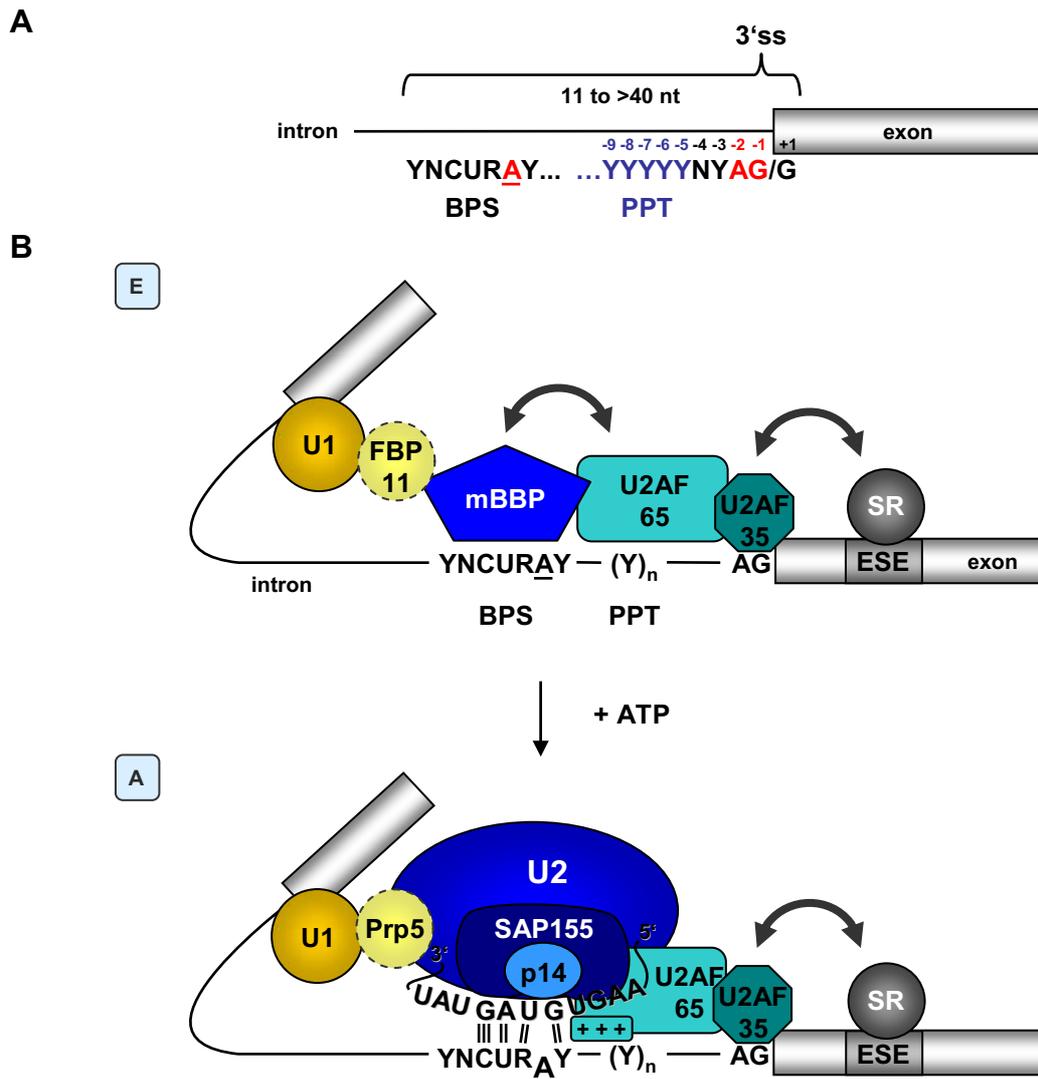
achieve U1 snRNP binding in the absence of flanking sequences the RNA-RNA interaction between the 5' ss and U1 snRNA has to reach a definite threshold of complementarity, which still remains to be defined.

Besides the RNA duplex between U1 snRNA and the 5' ss, also U1-C had been implicated to contribute to 5' ss recognition (110, 164, 253, 284, 650). Crystallographic analysis of the U1 snRNP structure suggested that a zinc-finger motif of U1-C interacts with the minor groove formed by an RNA duplex at the 5' ss and thus likely stabilises U1 snRNP binding (476). In particular pre-mRNA substrates, association of U1 snRNP

with the 5' ss requires the interaction of U1-70K (318) or U1-C (185) with splicing regulatory proteins of the SR protein family or the non-SR protein TIA-1 bound to splice site-proximal sequences. Based on U1 snRNP-mediated expression and splicing analyses, the concept evolved that neighbouring exonic and intronic sequences modulate the number of complementary bases required for U1 snRNP binding [F.-J. Grosseloh, diploma thesis (230)]. This model implicates that the overall strength of the 5' ss for initial splice site recognition is determined by the sequence of the 5' ss and the presence of splice site-proximal sequences stabilising or hindering U1 snRNP binding by recruiting *cis*-acting splicing regulatory factors. Subsequent to 5' ss recognition, U1 snRNP also functions to promote efficient association of U2 snRNP with the BPS during A complex formation (28). In addition to functions implicated in splicing, U1 snRNP binding also contributes to stabilisation of the pre-mRNA (295, 367). After recruitment of the U4/U6\*U5 tri-snRNP during transition from the A complex into the B complex, U1 snRNA is displaced by U6 snRNA and the 5' ss is additionally contacted by components of U5 snRNP.

### **A.3.3.2 3'ss recognition**

The downstream border of canonical introns is defined by the 3' ss consisting of the invariant AG-dinucleotide marking the intron/exon border, the upstream located branch point sequence (BPS) and the polypyrimidine tract (PPT) positioned in between (Fig. I-8A). The intrinsic strength of the 3' ss, i.e. the efficiency of splice site recognition by the spliceosomal machinery based on these elements, is assumed to largely depend on the complementarity of the BPS to the U2 snRNA and the quality of the PPT (129, 296), which differs in overall length and continuity of the pyrimidine-rich sequence. In addition, the immediately upstream and downstream flanking nucleotide of the AG-dinucleotide (positions -3 and +1 relative to exon/intron border) influence the efficiency of 3' ss recognition. Within the consensus sequences of the 3' ss a marked preference exists for a pyrimidine at position -3 (YAG/) (83), which coincides with a more efficient usage of the AG-dinucleotide for the second step of the splicing reaction (563) [L. Hartmann, diploma thesis (245)]. The preference for a pyrimidine at position -3 was proposed to be due to interactions between the 5' ss and the 3' ss mediated by U6 snRNA (124). Although the consensus sequence reveals a preference for guanosine in position +1, the first nucleotide in the exon appears to affect splice site recognition



**Fig. I-8: 3' ss recognition during early spliceosomal complex formation.**

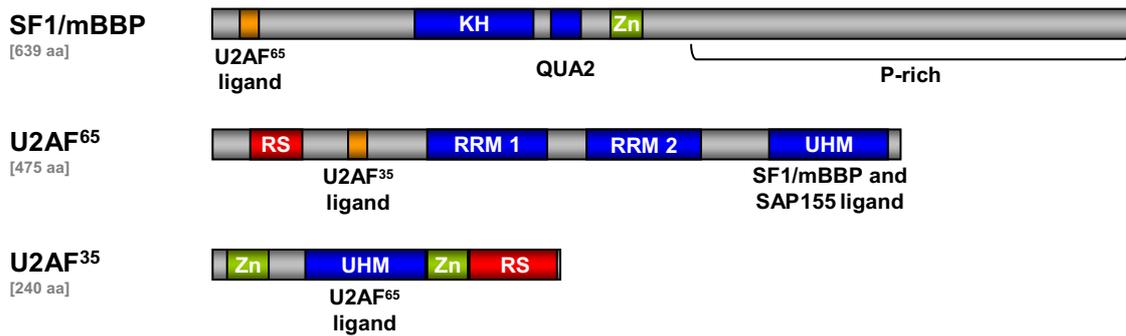
**(A)** Consensus sequences of mammalian 3' ss. The branch point adenosine and the AG-dinucleotide (red) together with the pyrimidine in position -3 represent the most conserved nucleotides. The PPT (blue) varies in length between individual 3' ss [A: branch adenosine, BPS: branch point sequence, N: any nucleotide, PPT: polypyrimidine tract, R: purine (A/G), Y: pyrimidine (C/U)] [modified from (83, 424)].

**(B)** Recognition of the BPS shifts from SF1/mBBP (mBBP) to U2 snRNP during the transition from E to A complex. Initially, the interaction between SF1/mBBP and the U2AF heterodimer stabilises the protein complex recognising the 3' ss in the E complex (top). Splice site-proximal splicing regulatory proteins can additionally contribute to complex stability. The proximity between both intron-flanking splice sites is thought to be caused by an interaction of FBP11 bridging U1 snRNP and SF1/mBBP. During ATP-dependent progression into the A complex (bottom), SF1/mBBP is displaced by U2 snRNP. Annealing of U2 snRNA and the pre-mRNA is enhanced by the RS domain of U2AF<sup>65</sup> (+++). For clarity only a section of the extensively structured U2 snRNA and the two U2 snRNP-specific proteins SAP155 and p14 especially contributing to BPS recognition are depicted. In the A complex, the proximity between the 5' ss and the 3' ss is likely mediated by Prp5 [ESE: exonic splicing enhancer, SR: serine-arginine-rich protein, U1: U1 snRNP, U2: U2 snRNP, (Y)<sub>n</sub>: polypyrimidine tract].

only in the presence of a suboptimal PPT with guanosine strongly promoting splicing compared to cytosine [L. Hartmann, diploma thesis (245)]. Furthermore, also appropriate spacing of the 3' ss elements contributes to the intrinsic strength (215, 563). The proximity of the PPT to the BPS increases the efficiency of the first step of the splicing reaction generating the intron lariat (487). In contrast, the AG-dinucleotide can be situated in a greater distance to BPS and PPT, because in case of an efficient recognition of BPS and PPT the AG-dinucleotide may not be required until the second transester reaction (487). Initial recognition of the 3' ss elements occurs by binding of the auxiliary proteins SF1/mBBP to the BPS (44) and U2AF to the PPT and the AG-dinucleotide (408, 509, 656, 677, 697) generating the E complex at the 3' ss.

The heterodimeric auxiliary splicing factor U2AF has been found to initially recognise the 3' ss and to recruit U2 snRNP to the BPS in most metazoan pre-mRNAs (326, 509, 677). U2AF consists of the large 65 kDa subunit (U2AF<sup>65</sup>) contacting the PPT (677, 679) and the small 35 kDa subunit (U2AF<sup>35</sup>) interacting with the downstream located AG-dinucleotide (408, 656, 697) (Fig. I-8B). Both U2AF subunits were found to be crucial for viability of various model organisms, e.g. *Drosophila melanogaster* (298, 507), *Caenorhabditis elegans* (698) and *Schizosaccharomyces pombe* (478, 644). However, the U2AF<sup>65</sup> homolog is dispensable for spliceosome assembly in the most intensively studied splicing model organism, *Saccharomyces cerevisiae*, (3). Furthermore, no homolog for U2AF<sup>35</sup> has been found in that organism. The functional redundancy of U2AF<sup>65</sup> in *S. cerevisiae* indicates that alternative pathways might exist to circumvent U2AF requirement in particular pre-mRNAs. This suggestion is consistent with the observation that the necessity for U2AF in mammalian splicing systems strongly depends on the pre-mRNA substrate (378).

In most metazoan pre-mRNAs, U2AF<sup>65</sup> is sufficient for U2 snRNP recruitment. U2AF<sup>65</sup> contains an N-terminal arginine/serine-rich (RS) domain consisting of RS-dipeptide repeats, an U2AF<sup>35</sup>-interaction domain [aa 85-112 (508, 685)] and three RRM-type RNA-binding domains (679) (Fig. I-9). Two of the RNA binding domains are canonical RRM motifs recognising the PPT in a single-stranded structure (557), whereas the third, the U2AF homology motif (UHM) named based on structural and sequence homologies with U2AF<sup>35</sup>, constitutes an atypical RRM specialised in protein-protein interactions [reviewed in (309)]. The UHM domain of U2AF<sup>65</sup> interacts with SF1/mBBP in the E complex (43, 484) and with SAP155 in the A complex (215). Analysis of U2AF<sup>65</sup>-directed RNase cleavage of the 3' ss elements (306) and X-ray analysis of a minimal



**Fig. I-9: Domain organisation of non-spliceosomal proteins recognising distinct 3' ss elements during early spliceosomal complex formation.**

[KH: hnRNP K homology domain, QUA2: Quaking homology 2 domain, RRM: RNA recognition motif, RS: arginine-serine-rich domain, UHM: U2AF homology motif, Zn: zinc finger]

U2AF<sup>65</sup> polypeptide containing RRM1 and RRM2 (557) revealed that binding of U2AF<sup>65</sup> to the PPT bends the RNA at the 3' ss, which was suggested to bring the 3' ss into the proximity of the BPS (306).

Binding of U2AF<sup>65</sup> to the PPT is supported by the small subunit of U2AF (700). Dimerisation with U2AF<sup>35</sup> strongly increases the RNA binding affinity of the heterodimer compared to the U2AF<sup>65</sup> monomer (506) by a sequence-dependent interaction with the AG-dinucleotide closely downstream of the PPT (408, 656, 697). Recognition of the AG-dinucleotide by U2AF<sup>35</sup> is controlled by DEK, a chromatin-associated protein, which interferes with binding of the U2AF heterodimer to PPTs lacking an AG-dinucleotide (567). In addition, U2AF<sup>35</sup> bridges U2AF<sup>65</sup> and non-spliceosomal regulatory proteins bound to the downstream exon in particular pre-mRNAs (700). The interaction of U2AF<sup>35</sup> with RS domains of other non-spliceosomal splicing factors of the SR protein family in enhancer-dependent splicing is mediated by its C-terminal RS domain (700), which is disrupted by 12 glycine residues (685). Although U2AF<sup>35</sup> contains a region with weak homology to an RRM-type RNA binding domain, the monomer exhibits only residual RNA binding affinity (408, 506). Instead, the atypical RRM was found to interact with U2AF<sup>65</sup> (685). The interface between both subunits is organised around two tryptophan residues, one contributed by the atypical RRM of U2AF<sup>35</sup> and the other one by a proline-rich region in the N-terminus of U2AF<sup>65</sup>, which are bound in hydrophobic pockets (310).

The impact of the interaction of the U2AF<sup>35</sup> subunit with the AG-dinucleotide appears to depend on the RNA substrate (234): AG-independent introns already perform the first transesterification reaction in splicing substrates lacking the AG-dinucleotide (487) and thus in the absence of the U2AF<sup>35</sup> subunit, whereas AG-dependent introns require the presence of the AG-dinucleotide and U2AF<sup>35</sup> for spliceosome assembly and the first transester reaction [(487, 656), reviewed in (424)]. AG-dependent introns were mostly found to contain short and discontinuous PPTs implying that – for these substrates – U2AF<sup>65</sup> binds only weakly to the PPT and has to be assisted by U2AF<sup>35</sup> for efficient recognition of the PPT. In line with this suggestion, interaction of U2AF<sup>35</sup> with the AG/G of the 3' ss stabilises U2AF<sup>65</sup> at weak PPTs (506, 656, 697). Besides stabilising U2AF<sup>65</sup> at suboptimal PPTs, U2AF<sup>35</sup> has been found to also promote splicing without increasing the amount of bound U2AF<sup>65</sup> (235). Therefore, it was concluded that U2AF<sup>35</sup> fulfils an additional role during spliceosome assembly, which may be rate-limiting for the splicing reaction of respective pre-mRNA substrates (235).

The interaction of U2AF<sup>35</sup> with the 3' ss during early spliceosome assembly has been suggested not necessarily to define the AG-dinucleotide that participates in exon ligation in the second step of the splicing reaction (424). This proposal was derived from the earlier observation that the AG-nucleotide used for spliceosome assembly can differ from that participating in exon ligation (469, 695). Additionally, the U2AF heterodimer frequently dissociates from the spliceosome prior to the first catalytic reaction (40), after which the AG-dinucleotide used for exon ligation is determined (18). Therefore, binding of U2AF<sup>35</sup> may be more important to identify suboptimal 3' ss deviating from consensus splice site sequences than for defining the exon/intron border.

SF1/mBBP, a 75 kDa protein initially purified from HeLa cells (325), binds to the BPS of the 3' ss (44) (Fig. I-8B). Although the function of SF1/mBBP is still a matter of debate, the protein was reported to bring the BPS into proximity of the 5' ss in the E' complex, which was hypothesised to be exerted by a potential interaction with FBP11 (4, 307), the human homolog of the yeast U1 snRNP protein component Prp40. Binding of SF1/mBBP to the pre-RNA is stabilised by U2AF. The proximity frequently found for BPS and PPT in pre-mRNA substrates facilitates SF1/mBBP and U2AF to cooperatively increase their binding at the 3' ss. Binding to U2AF<sup>65</sup> was found to enhance the binding affinity of SF1/mBBP to the BPS 20-fold, whereas SF1/mBBP increases the binding affinity of U2AF<sup>65</sup> to the PPT 5-fold (43). Protein-protein and protein-RNA interactions are performed by distinct structural domains of SF1/mBBP. The N-terminal region

contacting the atypical RNA recognition motif 3 (RRM3) of U2AF<sup>65</sup> (43, 484) is followed by the hnRNP K homology (KH) domain mediating the sequence-specific interaction of SF1/mBBP with the BPS (4, 45, 484) (Fig. I-9). NMR-derived structural data of SF1/mBBP complexed with a minimal RNA ligand containing the yeast BPS flanked by two nucleotides on each site revealed that, together with a Quaking homology 2 (QUA2) region located C-terminally, the KH domain recognises the complete BPS and deeply buries the branch point adenosine within the interaction interface (362). The adjacent zinc knuckle was reported to confer additional binding affinity (45). In contrast to the N-terminally located interaction domains, the following proline-rich region and the C-terminus are negligible for SF1/mBBP to function in spliceosomal assembly and for yeast viability (484).

Despite its well documented binding to the BPS and the interdependency with U2AF<sup>65</sup> during E complex assembly, SF1/mBBP exerts only a kinetic impact on particular pre-mRNA substrates carrying an optimal BPS matching the yeast consensus sequence (236, 510). However, SF1/mBBP was also demonstrated to be dispensable for the assembly of adenovirus- and IgM-derived pre-mRNAs into spliceosomal complexes (236). Moreover, RNAi-mediated depletion of SF1/mBBP did not affect splicing of several endogenous pre-mRNAs (602). Overall, these results indicate that the function of SF1/mBBP can be compensated by other splicing factors.

Recognition of BPS and PPT results in the formation of a ternary E complex at the 3' ss consisting of the pre-mRNA, SF1/mBBP and U2AF (4, 468). Binding of U2AF<sup>35</sup> is not necessarily required to generate the E complex, but might crucially assist U2AF<sup>65</sup> binding to RNA substrates containing suboptimal PPTs. The ternary complex is stabilised by cooperative interactions between SF1/mBBP and U2AF<sup>65</sup> increasing the affinity of both proteins to the RNA substrate (43). Two kinases, PKG-1 (cGMP-dependent protein kinase-1) (635) and KIS (kinase interacting with stathmin) (390), affect the interaction of SF1/mBBP with U2AF<sup>65</sup>. Binding to U2AF<sup>65</sup> is severely reduced by PKG-1-mediated phosphorylation of SF1/mBBP at the conserved Ser20 (635), which is located within a stretch of positively charged residues contributing to the molecular interface with U2AF<sup>65</sup> RRM3 (537). In contrast, phosphorylation of Ser80 and Ser82 by KIS appears to enhance the interaction of SF1/mBBP with U2AF<sup>65</sup> and thus ternary complex formation (390).

Although the E complex at the 3' ss is generally defined by the presence of SF1/mBBP as well as U2AF<sup>65</sup>, the order of assembly appears not to be absolute. Both proteins have been shown to bind particular RNA targets independently of each other generating the E' complex in the absence of U2AF (307) or an intermediate complex lacking SF1/mBBP (236). However, during further assembly both prespliceosomal complexes converge into the common E complex (Fig. I-5). Recognition of the 3' ss appears to represent a highly variable modular system regarding the sequence conservation of BPS and PPT and the requirement for the non-snRNP proteins that recognise the 3' ss in the E complex. Although SF1/mBBP and both U2AF subunits were found to contribute to 3' ss recognition, all of them have been found to be dispensable in particular pre-mRNAs: RNA substrates have been described, in which recognition of the BPS by SF1/mBBP is not required for spliceosome formation and splicing (236, 510). Despite its impact on 3' ss recognition and U2 snRNP recruitment, in the presence of excess SC35, a splicing regulatory factor of the SR protein family, U2AF<sup>65</sup> was found to be dispensable for splicing of the chimeric PIPβG pre-mRNA substrate *in vitro* (378). This suggests that although binding of U2AF<sup>65</sup> is essential for splicing in almost all cases, U2AF<sup>65</sup> binding appears not to be the rate-limiting step for every splicing substrate (247, 398). In addition, U2AF<sup>65</sup>-independent splicing of an IgM pre-mRNA had been observed in nuclear extracts of adenovirus-infected HeLa cells indicating that the requirement of U2AF<sup>65</sup> for splicing can be compensated by other factors (372). Also U2AF<sup>35</sup> has been shown to be dispensable for *in vitro* splicing (678, 679). In the case of a long and continuous PPT, U2AF<sup>35</sup> was found negligible for the splicing reaction (234, 656). Therefore, it appears that the quality of different 3' ss, i.e. sequence conservation and appropriate spacing between the BPS, PPT and AG-dinucleotide, may determine distinct steps in early spliceosome assembly to be rate-limiting for spliceosome formation. Nevertheless, 3' ss elements *in vivo* frequently deviate from the favoured nucleotide distances. It appears that optimal splice site elements reduce the necessity to involve all factors that are able to contribute to 3' ss recognition, whereas 3' ss displaying low sequence conservation and suboptimal distances between the 3' ss elements require the presence of SF1/mBBP as well as the U2AF heterodimer for recognition. Thereby, this variable system of proteins binding to the 3' ss elements may be an important prerequisite to guarantee flexible but highly accurate 3' ss recognition.

Although U2 snRNP does not interact with the BPS in the E complex, it is already loosely associated at this point (145, 414, 488). Moreover, its presence has been found

to be required for E complex formation (159). During transition into the prespliceosomal A complex the 17S U2 snRNP stably associates with the BPS in an ATP-dependent process mediated by two DEXH/D-box RNA unwindases, UAP56 (182) and the U2 snRNP-associated protein Prp5 (665). Binding of p14/SF3b14a, a subunit of the U2 snRNP-specific component SF3b, replaces SF1/mBBP from the BPS, which enables the UHM domain of U2AF<sup>65</sup> to bind to the SAP155 subunit of the SF3b component (98, 215, 609). These compositional and conformational changes at the 3'ss facilitate the formation of a bulged duplex between U2 snRNA and the BPS (483), which is promoted by the U2 snRNP-specific protein SAP155 binding to both sites of the BPS (215). In addition to the protein-protein interaction stabilising U2 snRNP at the RNA, U2AF<sup>65</sup> also promotes the annealing of U2 snRNA to the BPS by its N-terminal arginine-serine-rich (RS) domain (618). For this mechanism it was proposed that positively charged residues within the RS domain of U2AF<sup>65</sup> contact the BPS and stabilise the limited base pairing interaction between U2 snRNA and the BPS (618). The only requirement for the function of the RS domain appears to be a positive net charge that facilitates RNA annealing by compensating negative charges of the phosphate backbone. Binding of U2 snRNP at the pre-mRNA is further stabilised by RNA-specific but sequence-independent interactions of the U2 snRNP-specific, trimeric SF3a and heptameric SF3b components with a region upstream of the BPS (214).

#### **A.3.4 Alternative splice site selection**

The inclusion of different exons into an mRNA by alternative splice site usage affects regulatory aspects of RNA metabolism e.g. intracellular distribution and half-life, as well as the coding information resulting in mRNA isoforms encoding differing peptides with potentially altered chemical and biological activities. Since genome sequencing of diverse species revealed that the number of protein coding genes in an organism does not correlate with its overall cellular complexity (6, 341, 608, 642), it became evident that alternative splicing significantly contributes to proteome diversity and thus expands the coding capacity of the genome [reviewed in (53, 218, 392)]. Recently, high-throughput sequencing of human tissue-specific transcriptomes indicated that nearly all multi-exon genes composing 94% of human genes are alternatively spliced (631). The high frequency of alternatively spliced mRNA isoforms in the human transcriptome led to the proposal that splice sites might be paired less stringently by the spliceosome,

giving rise to numerous mRNA isoforms and thereby also to evolution (406, 407, 571). However, this idea was contrasted by experiments revealing that the spliceosome recognises and removes introns with high accuracy (189) indicating that alternative splice site selection is a highly regulated process. Regulation of alternative splicing provides the potential to modulate the spatio-temporal expression of a splice variant in different cell types, during development or in other biological processes induced in response to extra- and intracellular cell stimuli [reviewed in (553)], like e.g. apoptosis [reviewed in (533)]. Misregulation of alternative splicing may result in the failure to meet cell- and tissue-specific protein demands frequently leading to cellular dysfunction and disease [reviewed in (86, 453)].

Alternative splice site selection differentially alters the exon structure of the mRNA. Complete exons can either be included into the mRNA or skipped. In the case of mutually exclusive exons, only one of the competing exons is included into the mRNA. In contrast, alternate use of neighbouring competing splice sites modifies the length of the resulting exon. Furthermore, introns that are normally excised can be retained in the mRNA. The extensive transcript diversity generated by alternative splicing implies a significant flexibility of the spliceosome to identify and process splice site pairs within a given pre-mRNA. Experimental evidence regarding splice site activation converge into a model, in which combinatorial control of multiple parameters mediates splice site recognition (256). Alternative selection of splice sites results from differential factor recruitment to *cis*-acting splicing regulatory elements that either promote or suppress splice site usage and also from the transcription process itself. Although splice site sequences and exon/intron architecture constitute invariant pre-mRNA attributes, intrinsic splice site strength and genomic context compose the basis for general splice site recognition and thus decisively contribute to alternative splice site selection [reviewed in (80)].

The splice site itself contains insufficient information to initiate splicing by sole sequence recognition. The large subset of intronic pseudo splice sites matching the consensus sequence as well or even better than authentic splice sites evidences the necessity of additional parameters for splice site activation (106). Numerous studies employing site-directed mutagenesis of splicing reporter minigenes, conventional and functional SELEX (Systematic Evolution of Ligands by Exponential Enrichment) of RNA-binding splicing factors, cross-linking immunoprecipitation (CLIP) analyses (514) and computational approaches resulted in the identification of *cis*-acting splicing regulatory

---

elements consisting of short degenerated motifs that influence splice site activation [reviewed in (637)]. The best analysed examples of *cis*-acting splicing regulatory sequences act by recruiting regulatory proteins into the proximity of splice sites, which either enhance or repress spliceosome assembly on adjacent splice sites. In addition, also regulatory elements have been identified that affect splice site recognition by the RNA secondary structure they adopt. A higher level of regulatory complexity is achieved by frequent combination of multiple *cis*-regulatory sequences controlling splice site recognition [reviewed in (54, 564)]. The combinatorial interplay of positive and negative regulatory elements allows the gradual activation of splice site usage resulting in a variable ratio of mRNA isoforms.

Analysing alternative splice site preference in dependency of the intron length revealed that exons flanked by introns larger than 250 nt and thus recognised by exon definition have a higher probability of being excluded from the mRNA during alternative splicing than exons flanked by short introns, which are recognised by intron-definition (188). The observation that the pre-mRNA exon/intron architecture contributes to alternative splicing was further substantiated by the finding that minigene-based simulation of aberrant splicing patterns, frequently observed due to mutations in splice sites or splicing regulatory elements *in vivo*, strongly depends on the genomic context [E. Honisch, diploma thesis (263)] (30). It might be envisioned that splicing regulatory complexes established at individual exons communicate even across large intronic distances and thereby affect the overall outcome of splice site activation. Since a direct interaction between the spliceosomal and the transcriptional machinery was inferred from the tethering of an exon to the C-terminal domain (CTD) of an RNA polymerase II (RNAP II)-directed transcription complex (166), likely via the U1 snRNP-associated protein Prp40 (426), regulatory complexes might get already into proximity during transcription. An interaction of one or multiple U1 snRNP with the CTD of RNAP II would locate the 5' ss in the proximity of nascent alternative 3' ss and regulatory complexes thus increasing the relative local concentration of splice sites and splicing regulatory elements [reviewed in (207)].

Splicing of constitutive as well as alternative exons can occur already cotranscriptional (462), which has been initially deduced from electron micrographs of *Drosophila melanogaster* embryonic transcription units (48, 49, 454). Cotranscriptional splicing enables the transcription machinery to influence alternative splicing [reviewed in (322)]. Two transcription-related mechanisms have been demonstrated to modulate alternative

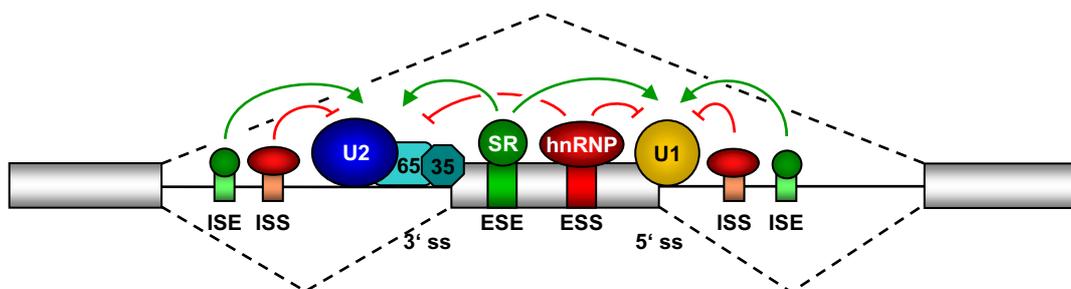
splice site selection: first, in the dynamic exon recognition hypothesis altered elongation rates of the RNAP II restrict the time available for splicing of a particular splice site pair until competing splice sites and regulatory elements are transcribed thus regulating splice site competition and also promoting the formation of different RNA secondary structures [kinetic model (86, 149, 321)]. The impact of transcriptional elongation on alternative splice site selection was first demonstrated by the predominant inclusion of the alternative EDI exon into the *fibronectin* mRNA following transcription in the presence of efficient activators of transcriptional elongation (291, 446). The transcriptional elongation rate has been found to be altered upon an environmental signal, ultraviolet (UV) irradiation, affecting alternative splicing of several pre-mRNAs, e.g. of the apoptotic regulator *bcl-x* (430). Second, depending on the promoter identity, splicing regulatory proteins are differentially recruited to the transcriptional start and also to nascent RNA resulting in an altered splicing pattern of the pre-mRNA [recruitment model (86, 321)]. The CTD of RNAP II appears to play a central role for loading of spliceosomal factors onto the pre-mRNA (184, 422), because splicing regulatory factors of the SR protein family (144, 422), PSF (169, 501), p54 (169) and the U1 snRNP-associated protein Prp40 (426) have been found to associate with the CTD. However, also interactions of additional transcription factors with e.g. the SF3a subunit of U2 snRNP (560) have been reported [reviewed in (461)]. In line with the coupling of transcription and splicing, promoter architecture has been found to influence alternative splicing e.g. in meiosis-specific splicing of fission yeast (423). Thus, variations in the assembly of the transcriptional complex and the rate of transcriptional elongation provide an additional regulatory impact on alternative splicing.

The splicing regulatory network represents a subnetwork of gene regulation controlling the generation of mRNA diversity by several layers of intercommunicating, *cis*-acting elements that distinguish exons from introns hence defining a splicing code [reviewed in (59, 632)]. The investigation of the network's regulatory circuits, which direct tissue-specific or signal-induced changes in alternative splicing patterns, has been highly stimulated by recent advances in high-throughput technologies used for microarray analyses and DNA sequencing [reviewed in (244)]. To assess the effect of changes in splice factor expression on global alternative splicing patterns, genome-wide RNA binding sites for splicing regulatory factors were identified by cross-linking and immunoprecipitation followed by sequencing thereby providing the basis for decrypting the general code underlying splicing regulation. RNA binding site annotation in

combination with alternative splicing information were used to generate large protein-specific RNA-protein interaction maps, e.g. for the splicing factors Nova (353), Fox-1 and Fox-2 (671, 682), revealing that the majority of genes regulated by a single splicing factor tend to be associated with similar functions. The potential discovery of novel mRNA isoforms by deep-sequencing of transcriptomes likely further extends these genome-wide maps of splicing regulatory networks, which compose the groundwork for predictions how mutations in splice factor and target genes affect mRNA splicing.

### A.3.5 *Cis*-acting splicing regulatory elements

*Cis*-regulatory sequences are frequently found in metazoan pre-mRNAs to either activate (enhance) or repress (silence) the use of adjacent splice sites by recruiting *trans*-acting sequence-specific RNA binding proteins or modulating RNA secondary structure [reviewed in (637)]. According to their location and activity these elements are referred to as intronic/exonic splicing enhancers (ESE/ISE) or exonic/intronic splicing silencers (ESS/ISS) (Fig. I-10). The effect of *cis*-regulatory sequences on overall splice site strength facilitates variable modulation of splice site usage leading to the



**Fig. I-10: *Cis*-acting elements regulate splice site recognition resulting in alternative splicing.**

*Trans*-acting factors (hnRNP and SR) recruited by *cis*-acting regulatory elements either enhance (green) or silence (red) the recognition of neighbouring splice sites by spliceosomal components. The influence of splicing regulatory proteins on splice site recognition frequently affects splicing of adjacent introns (lower dashed lines) and may also result in skipping of the internal exon (upper dashed line) [35: U2AF<sup>35</sup>, 65: U2AF<sup>65</sup>, ESE: exonic splicing enhancer, ESS: exonic splicing silencer, hnRNP: heterogeneous nuclear ribonucleoprotein particle protein, ISE: intronic splicing enhancer, ISS: intronic splicing silencer, SR: serine-arginine-rich protein, U1: U1 snRNP, U2: U2 snRNP] [adapted from (637)].

suggestion that *cis*-regulatory elements are an integral component of alternative splice site regulation. While early reports proposed that activation of constitutively used splice sites does not require splice site-proximal enhancer elements, more recent data indicated that also constitutively used splice sites might depend on the function of splicing enhancer elements (520, 687). The abundance of *cis*-acting splicing regulatory elements indicates that exonic as well as intronic mutations can impact alternative splicing. In former times exonic mutations were frequently considered only on protein level thereby causing amino acid changes (missense mutation) or truncated proteins (deletions or nonsense mutation). However, mutationally induced aberrant RNA splicing patterns displaying exon skipping or activation of cryptic splice sites have emerged as a frequent cause of disease [reviewed in (95, 459, 569)]. The identification of genomic mutations affecting *cis*-regulatory elements led to the generation of antisense-based therapeutic strategies targeting mutated *cis*-regulatory elements by either chemically modified nucleic acids silencing the use of cryptic splice sites or by hybrid protein nucleic acid (PNA)-peptides mimicking binding of a splicing regulator to correct pathologically altered splicing events [(96, 518, 561), reviewed in (607)].

Enhancer and silencer motifs identified so far tend to be short (typically 5-10 nt) and consist of relatively degenerate consensus sequences. General splicing regulatory motifs were determined by purely computational approaches comparing datasets expected to be enriched in splicing regulatory elements with one or more control sets [(174, 177, 670, 687), reviewed in (106)]. Sequence motifs that are substantially enriched are considered as putative binding sites for regulatory factors and, in general, group into families of highly similar sequences that together define a consensus binding motif. Nevertheless, computational approaches based on motif frequency are necessarily restricted to uncover binding sites for which motif overrepresentation can be observed. In addition to splicing regulatory motifs involved in general splicing, some computational approaches searched for motifs that are associated with alternative splicing by analysing datasets containing sequences flanking alternatively used splice sites (210, 636). Splicing regulatory motifs have also been determined by functional selection *in vivo*. Random sequences of defined length were substituted for essential enhancer or silencer motifs in reporter minigenes (132, 639). After transfection functional splicing regulatory motifs were identified by PCR amplification and sequencing. Following multiple rounds of cell-based selection by repeated

transfection/isolation rounds, resulting sequences were grouped and consensus sequences for splicing regulatory motifs deduced.

Identified splicing regulatory motifs differed between exonic and intronic input sequences, already indicating that distinct sets of splicing regulatory elements act dependent on their position. Experimental validation of computationally derived splicing regulatory elements confirmed that the activity of some splicing regulatory elements is strongly position-dependent (210). Consistent with this observation, an adenoviral regulatory element had been shown to inhibit splicing at a 3' ss when located upstream of the BPS, whereas an exonic position resulted in activation of the same 3' ss (300).

Most *cis*-regulatory elements recruit sequence-specific RNA binding proteins that either promote or inhibit activation of an adjacent splice site by affecting spliceosome assembly. Different spliceosomal complexes have been shown to be influenced by *cis*-regulatory proteins [reviewed in (266)]. Assembly of the E complex is a major control point for initial splice site recognition and pairing (415, 539) and, therefore, is thought to represent an important target for the regulation of alternative splicing (415). Numerous examples of splicing regulatory proteins acting via *cis*-regulatory elements have been described that affect splice site recognition in the E complex by stabilising or hindering the interaction of U1 snRNP with the 5' ss (89, 185, 284, 318, 655, 659) and U2AF with the PPT of 3' ss (223, 540, 604, 638, 700). Nevertheless, evidence provided more recently indicated that splicing can also be regulated during the successive transitions into the A or the B complex (66, 206, 265) or even between both catalytic reactions (112, 219, 338). In addition, also the rearrangement from an exon- to an intron-defined pre-spliceosomal A complex provides a regulatory check point for the splicing reaction (541). However, also proteins involved in ensuring the fidelity of the splicing reaction have been found to contribute to alternative splice site selection (463). The influence of *cis*-acting splicing regulatory elements on distinct complexes formed throughout spliceosome assembly suggests that alternative splice site usage is modulated by several distinct molecular interactions.

The best analysed examples of enhancer-mediated 5' ss activation are performed by RNA-dependent stabilisation of U1 snRNP by interactions of enhancer-bound splicing regulatory factors of the SR protein family and the splicing regulatory protein TIA-1 with U1 snRNP-specific proteins U1-70K (89, 284, 318, 655, 659) and U1-C (185, 284). From the observation that members of the SR protein family and the hnRNP family as

well as additional splicing regulatory proteins require a specific exonic or intronic position relative to the splice site, it appears that enhancer-mediated U1 snRNP stabilisation depends on the stereospecific arrangement of the regulatory proteins [S. Rosin, diploma thesis (500)]. Within the first exon of a pre-mRNA, U1 snRNP can also be stabilised by CBP80 bound to the cap structure (125, 347, 683). In addition to the interaction of regulatory proteins with U1 snRNP-specific proteins, binding of U1 snRNP to the 5' splice site was described to be stabilised also by direct interactions of the regulatory protein with U1 snRNA (544, 545).

The most intensively studied mechanism for 3' splice site activation is the recruitment of U2AF<sup>65</sup> to the PPT. In the U2AF<sup>65</sup> recruitment model, SR proteins bound to an ESE stabilise U2AF<sup>35</sup> at the AG-dinucleotide (638, 700), which serves as bridge recruiting U2AF<sup>65</sup> to the weak PPT of non-consensus 3' splice site [(700), reviewed in (58, 217)]. Although SR proteins have been found to directly interact with U2AF<sup>35</sup> consistent with the proposed bridging role for U2AF<sup>35</sup> (655), U2AF<sup>35</sup> is dispensable for enhancer-dependent splicing of some pre-mRNA substrates. In support of a model bypassing U2AF<sup>35</sup>-mediated stabilisation, it was found that U2AF<sup>65</sup> – but not U2AF<sup>35</sup> – interacts with the SR protein SRp54 (686). However, the U2AF<sup>65</sup> recruitment model was not confirmed as an universal mechanism of enhancer function by subsequent studies demonstrating that U2AF<sup>65</sup> binding was either unaffected by neighbouring enhancers or did not correlate with splicing activity of the pre-mRNA substrates (234, 297, 352). Recently, inactivating an ESE in exon 7 of the *SMN1/2* genes by providing an oligonucleotide complementary to this splicing regulatory element, reduced U2 snRNP recruitment to the 3' splice site of intron 6 without changing U2AF<sup>65</sup> recruitment suggesting that U2 snRNP binding can also be stabilised by splicing regulatory factors at a step following U2AF<sup>65</sup> binding (398). These results argue that U2AF<sup>65</sup> binding is not the rate-limiting step in every ESE-mediated splicing reaction and pointed to the involvement of other mechanisms in the regulation of enhancer-dependent 3' splice site activation. One of these alternative mechanisms might be the improved annealing of the U2 snRNA in the presence of the RS domain of an ESE-bound SR protein, which stabilises U2 snRNP binding to the BPS (545) like it has been described for the RS domain of U2AF<sup>65</sup> (618). Also ESE-dependent recruitment of the coactivator complex SRm160/300 during spliceosome assembly was found to enhance splice site usage (167).

Repression of splice site usage by splice site-proximal silencer elements has mainly been attributed to the antagonistic action of silencer-bound proteins on the effect of

positively acting splicing regulatory proteins. In the simplest case silencer-bound proteins share similar binding preferences to core splicing signals or regulatory elements and therefore compete with binding of spliceosome components or positively acting splicing regulatory proteins. The competition has been observed for polypyrimidine tract binding protein (PTB; also referred to as hnRNP I), which binds to the PPT and thus interferes with binding of U2AF. Silencer elements have been found to promote the formation of 'dead-end' complexes that resist further processing into catalytically active spliceosomes (131, 297, 333). However, the often cooperative nature of repressor binding extends the antagonistic mechanism by two additional scenarios: in the first, cooperative binding of negatively acting splicing regulatory proteins starting from high-affinity binding sites propagates across core splicing signals or enhancer elements hence suppressing splice site recognition by generating a 'zone of silencing' [(693), reviewed in (396, 400)]. In the second scenario, interactions between repressors bound to both sides of an alternatively spliced exon loop out the intervening RNA thereby making the core splicing signals inaccessible for spliceosome components and enhancer proteins and bringing otherwise distant exons in close proximity to each other.

Besides affecting splice site usage by recruiting splicing regulatory proteins, *cis*-regulatory elements have also been described to influence splice site selection by the RNA secondary structure they adopt [reviewed in (79)]. A bioinformatical analysis evaluating the calculated RNA stability of secondary structures containing core splicing signals of alternatively used splice sites revealed that up to 4% of alternative splicing events strongly correlate with the presence of stable RNA secondary structures and thus might be subject to RNA structure-mediated splice site regulation (549). The secondary structure of *cis*-regulatory elements profoundly modulates binding of regulatory proteins, which vary in the RNA pattern they recognise. Whereas the majority of RNA-binding domains in splicing regulatory proteins, e.g. the hnRNP K homology (KH) domain and RRM, is suggested to bind sequence-specific to single-stranded RNA motifs [reviewed in (24, 411)], also double-stranded hairpins exposing apical loops serve as binding site for splicing regulatory proteins as it had been had been demonstrated for HuR (365). Also binding of the splicing regulatory proteins Nova-1, SRp55 and its *Drosophila* homolog B52 were found to depend on a combination of RNA secondary structure and target nucleotide sequences (78, 435, 551). Therefore, changes in the RNA secondary structure potentially result in the association of a different set of splicing regulatory proteins with the pre-mRNA. *Vice versa* also protein

binding is able to influence the structure of the pre-mRNA. Protein-mediated RNA structuring has for example been revealed at the 3' ss, where binding of U2AF to the PPT was proposed to bend the RNA in a way that brings the AG-dinucleotide into proximity of the BPS (306).

Most reports revealing an impact of RNA secondary structure on splicing concern core splicing signals, which become inaccessible for recognition by spliceosomal components due to structural constraints. Consistently, sequestering 5' ss D1 of the HIV-1 pre-mRNA in mutant stem loops more stable than those formed in wild-type viral isolates was found to reduce viral replication and infectivity (2). Likewise, an extensive RNA secondary structure capturing the major 5' ss of the murine leukaemia virus (MLV) pre-mRNA interferes with efficient splice site usage thus providing genomic as well as low amounts of spliced viral RNA (327). Suboptimal activation of the MLV 5' ss depends on the limited complementarity between the U1 snRNA and the splice site (703). Structure-dependent sequestering has also been observed for splicing regulatory elements thus modulating their effect e.g. on the inclusion of the EDA exon into the *fibronectin* mRNA (82, 432). However, secondary structures can also promote splicing at correct splice sites by masking aberrant pseudo splice sites or by stabilising adjacent core splicing signals in a single-stranded conformation, which can be bound by spliceosomal components [reviewed in (79)]. RNA secondary structure can also shorten the distance between splice sites and thereby enhance the splicing reaction. Also the net outcome of splicing regulatory complexes can be modulated by bringing otherwise far-distant splicing regulatory elements into proximity, like it had been observed for splicing of the alternative exon IIIB in the *FGFR2* pre-mRNA (31). However, long-range RNA-RNA interactions can also loop out complete exons thus hindering the recognition by spliceosomal components, which might induce exon skipping. This 'loop out' mechanism essentially regulates alternative splicing of the *dscam* gene encoding an axon guidance receptor in *Drosophila* (221). The differential outcome of RNA secondary structure on splice site activation indicates that the effect of secondary structure strongly depends on the relative position of splice sites and regulatory sequences within the genomic context.

### A.3.6 Splicing regulatory proteins

The most extensively studied cellular effectors mediating *cis*-regulatory splicing enhancer and silencer functions belong to the serine-arginine-rich (SR) protein family and to the heterogeneous nuclear ribonucleoparticle (hnRNP) protein family. Analysing the mechanism of splicing regulation in metazoan and in viral pre-mRNAs brought up the general model that enhancer-bound SR proteins support splicing by stabilising spliceosomal components during splice site recognition, whereas hnRNP proteins bound to silencer elements interfere with spliceosome assembly, mainly by antagonising SR protein function (88, 404, 693). However, individual hnRNP proteins and SR proteins were found to regulate splicing both positively and negatively in different pre-mRNAs (271, 300, 505), indicating that the effect of a splicing regulatory protein may strongly vary depending on the targeted pre-mRNA substrate.

#### A.3.6.1 SR proteins

The SR proteins constitute a family of structurally related and phylogenetically conserved RNA binding proteins [reviewed in (69, 357, 363, 550)]. Members of the SR protein family are found in metazoans, plants and also in the fission yeast *Schizosaccharomyces pombe*, whereas *Saccharomyces cerevisiae* encodes only three SR-like proteins (52, 237, 651). Several SR proteins act as essential splicing factors, but also function in RNA export, non-sense-mediated decay (NMD) and translation. Recently, the requirement of SR proteins for genome stability (350, 351, 658) and their binding to histones even indicated roles ahead of RNA processing [(364), reviewed in (689)]. Due to the multitude of cellular pathways they are involved in, SR proteins were implicated as master regulators of gene expression (363).

The founding member of the SR protein family, SF2/ASF, was independently isolated by two groups due to its activity to complement splicing-deficient HeLa nuclear cell extract (splicing factor 2; SF2) (324) and to switch 5' splice site selection in an alternatively spliced SV40 early pre-mRNA (alternative splicing factor; ASF) (202). Subsequently, another protein (spliceosomal component 35; SC35) required for *in vitro* splicing in nuclear extract was identified (192). Based on the initial identification of SF2/ASF, SR proteins were originally defined by the ability to restore pre-mRNA splicing in splicing-deficient cytoplasmic HeLa S100 extracts and by their detection with the monoclonal antibody mAb104 recognising a conserved phospho-epitope (503, 504, 675). So far, seven SR

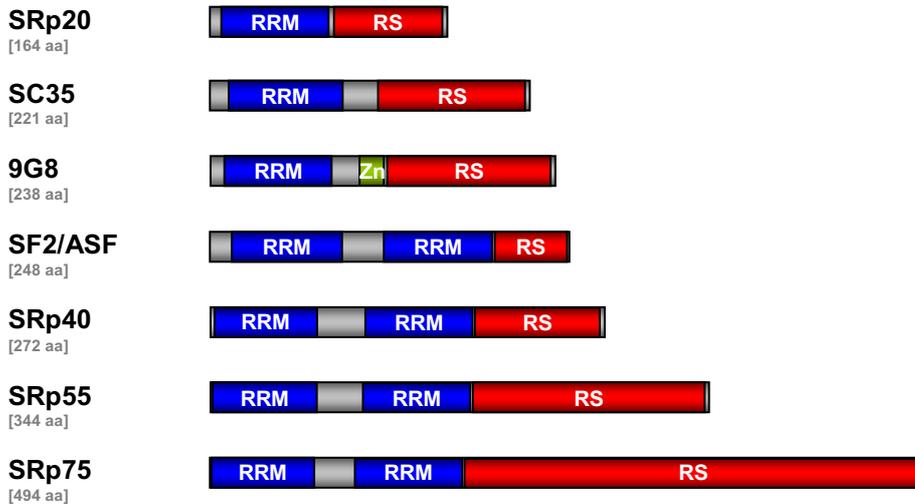
proteins (SF2/ASF, SC35, SRp20, SRp40, SRp55, SRp75, 9G8) are considered as classical members of the SR protein family due to their functions in constitutive as well as alternative splicing and due to their similar domain structure (357, 363). SR proteins have a modular structure characterised by one or two copies of an N-terminal RNA-recognition motif (RRM) followed by a C-terminal domain rich in arginine-serine dipeptide repeats (RS domain) (Fig. I-11A). In addition to classical SR proteins, a number of splicing factors contain an RS domain, e.g. U1-70K, both U2AF subunits and the splicing coactivators SRm160 and SRm300 (67) (Fig. I-11B). Due to their otherwise different domain structure, they are collectively referred to as SR-related proteins (SRr) [reviewed in (61)].

The identification of SF2/ASF by different experimental approaches reflects the dual functionality of SR proteins to serve as essential splicing factors in constitutive splicing but also as regulators of splice site selection in alternative splicing. For constitutive splicing SR proteins are assumed to bridge the 5' ss and the 3' ss flanking the intron by stabilising interactions within the spliceosome thereby contributing to intron-definition. This proposal is based on the observations that SC35 and SF2/ASF promote the interaction between spliceosomal components bound at the 5' ss and the 3' ss (193, 655) and that SR proteins are required to recruit the U4/U6\*U5 tri-snRNP into the spliceosome (498). Consistent with an exon-independent function in constitutive splicing, SR proteins were found to be required for lariat formation of an RNA substrate containing a single exonic nucleotide (259). Besides their contribution to intron-definition, SR proteins were also found to mediate exon-definition in constitutive splicing by promoting the communication of spliceosomal components bound to exon-flanking splice sites (68). In alternative splicing SR proteins are thought to act as molecular adaptors stabilising otherwise weak interactions between spliceosomal components and alternatively used splice sites by binding to splicing enhancer sequences [reviewed in (357)].

Enhancer-dependent functions of the SR proteins are mediated by one or two RNA recognition motifs (RRMs) containing two highly conserved, short submotifs, RNP-1 and RNP-2. X-ray crystallography of a canonical RRM obtained from U1-A revealed the existence of a four-strand antiparallel  $\beta$ -sheet, which is packed against two connecting  $\alpha$ -helices (434). The RNP motifs reside within the  $\beta$ 1 and  $\beta$ 3 strand and expose several aromatic residues into the solvent, which are thought to contact the RNA through stacking interactions and hydrogen bonds formed with single-stranded bases [(458),

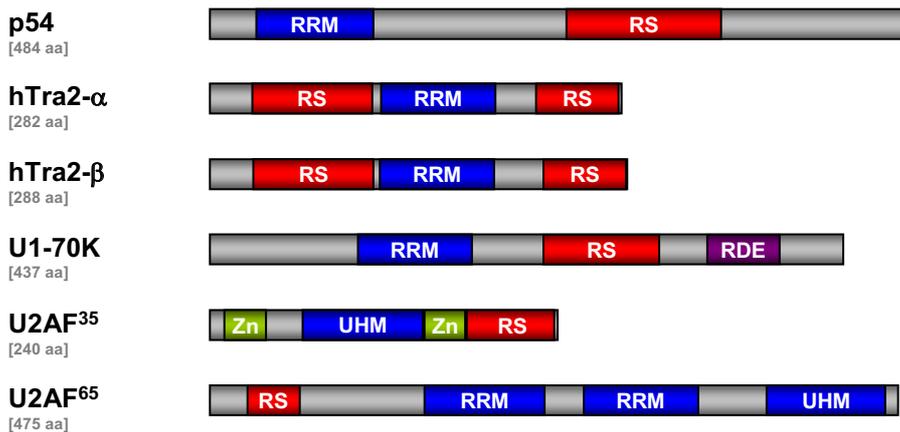
## A

## Classical SR proteins



## B

## Additional SR proteins



**Fig. I-11: Modular domain structure of the SR protein family.**

**(A)** Seven proteins constitute the subgroup of classical SR proteins characterised by one or two N-terminal RNA binding motifs (RRM) and a C-terminal domain rich in RS-dipeptides (RS domain). Classical SR proteins are characterised by structural similarity, dual functionality in constitutive and alternative splicing, recognition by the monoclonal antibody mAb104 recognising a common phospho-epitope and their precipitation in the presence of magnesium chloride. Respective domain positions were extracted from the Swiss-Prot database ([www.uniprot.org](http://www.uniprot.org)) [RRM: RNA recognition motif, RS: arginine-serine-rich domain, Zn: zinc finger motif].

**(B)** Examples of additional SR proteins. Several splicing regulatory proteins and spliceosomal components also contain an RS domain and thus are integrated into the subgroup of additional SR proteins. Although they share the RS domain, additional SR proteins lack the characteristics of the classical SR protein subgroup [RDE: domain rich in arginine, asparagine and glutamine, UHM: U2AF homology motif].

reviewed in (622)]. The second RRM found in some SR proteins (SF2/ASF, SRp75, SRp55 and SRp40) deviates from the canonical RRM structure. The atypical RRM is less conserved, but harbours an invariant signature heptapeptide, SWQDLKD (52), which contributes to RNA binding of SF2/ASF (614). Both RRMs of SF2/ASF were found to synergistically ensure optimal RNA binding (87, 701) and to determine RNA binding specificity (102, 596).

High-affinity RNA binding sites for individual SR proteins were experimentally determined by Systematic Evolution of Ligands by Exponential Enrichment (SELEX). Using this approach, the classical SR proteins SF2/ASF and SC35 (596), SRp40 (595), 9G8 and Srp20 (99) were shown to possess distinct RNA binding specificities mainly for purine-rich sequences generating consensus motifs of usually 6-10 nucleotides. Further analyses of SR protein target sites by functional SELEX approaches using activation of splicing as selection criteria (132, 610) surprisingly identified less well defined RNA binding sites for SF2/ASF, SC35, SRp40 and SRp55 than those obtained by conventional SELEX suggesting that also degenerated, low-affinity SR protein binding sites may function as splicing enhancers [(360), reviewed in (597)]. Meanwhile RNA binding motifs resulting from functional SELEX experiments for the SR proteins SF2/ASF, SC35, SRp55 and SRp40 have been implemented into a web-based algorithm enabling to search a given RNA target for sequences similar to these less defined consensus SR protein binding sites (97, 566).

The RNA binding specificity of splicing regulatory complexes appears to be modulated by the participating proteins as has been demonstrated for the sex determination factor transformer 2 (Tra2) in *Drosophila*. Tra2 is essential for female-specific activation of a multipartite ESE (dsxRE) in exon 4 of the *doublesex* (*dsx*) pre-mRNA (252, 277, 374, 611). Within the dsxRE, Tra2 binds to two distinct enhancer motifs by generating splicing regulatory protein complexes with the female-specific transformer (Tra) and an SR protein, either 9G8 or SF2/ASF (374, 375, 612). Since both enhancer motifs, a repeat element (RE) and a purine-rich element (PRE), differ significantly regarding their purine content, it was suggested that the specificity of the Tra2 RNA binding affinity is changed by cooperative interactions with different SR proteins thus allowing Tra2 binding to distinct enhancer sequences (374, 375, 598).

Two mammalian homologs of *Drosophila* Tra2, hTra2- $\alpha$  (147) and hTra2- $\beta$  (38, 401, 535), have been identified sharing 75% identity (598). Both proteins are structurally

related to classical SR proteins by containing a central RNA-binding domain flanked by two RS domains. Conventional SELEX studies revealed that both proteins preferentially bind to oligo-(A) or oligo-(GAA) sequences thereby overlapping with the RNA binding specificity of SF2/ASF (599). However, a subsequent study using *in vivo* splicing assays indicated that – in line with Tra2 binding to *dsx* enhancer elements containing a different purine content – hTra2- $\beta$  targets a highly degenerated RNA binding site (GHVVGANR; H=A/C/T, V=A/C/G, R=A/G, N=any nucleotide) (586). Binding of hTra2- $\beta$  was found to activate inclusion of different exons e.g. in the *SMN* and *hTra2- $\beta$*  pre-mRNA (261, 586). Due to their splicing activatory function, recent discussions suggested to consider hTra2- $\alpha$  and hTra2- $\beta$  as additional SR proteins (357, 363).

The distinctive feature of SR and SR-related proteins is represented by the RS domain composed of a number of consecutive arginine-serine dipeptides (675). The length and sequence of the RS domain is highly conserved for individual SR proteins (52). Since SF2/ASF was found to directly interact with the U1 snRNP-specific protein U1-70K in an RS domain-dependent manner (318, 659), it was suggested that SR proteins act in enhancer-dependent splicing by stabilising spliceosomal components at the splice sites via the RS domain. Consistently, RS dipeptide repeats were reported to be sufficient to constitute an enhancer domain (96, 222, 224, 472, 692). In addition to protein-protein interactions, the RS domain stabilises the interaction of the pre-mRNA-U snRNA duplex formed during splice site recognition by U1 and U2 snRNP and during spliceosome remodelling by U6 and U5 snRNA [(544-547), reviewed in (220)]. However, splice sites of several constitutively as well as alternatively spliced pre-mRNA substrates were found to be activated by SR proteins independent of the RS domain suggesting that the requirement of the RS domain for SR protein function depends on the pre-mRNA target (542, 604, 692).

The RS domain of SR proteins is extensively phosphorylated (503), which affects enhancer-dependent as well as general functions of SR proteins in splicing [(479, 595, 659, 660), reviewed in (574)]. SR proteins are phosphorylated by the SR protein kinase family (SRPK) (233, 331, 437, 633), the *cdc2*-like nuclear kinase family (Clk/Sty) (127, 439), DNA topoisomerase I (502), Akt (57), Akt2 (465), PKA C subunits (332) and glycogen synthase kinase-3 (GSK3) (254). Due to activation of the kinases by various signal transduction pathways, different intra- and extracellular stimuli can affect splicing patterns by modulating SR protein phosphorylation [reviewed in (581)]. In addition, modulating the extent of phosphorylation appears to be used to control the activity of

SR proteins e.g. during adenovirus infection (301) and during early development of the nematode *Ascaris lumbricoides* (513).

Phosphorylation of SR proteins has been reported to be essential to initiate spliceosome assembly and to recruit U4/U6\*U5 snRNP to the spliceosome in *in vitro* formation assays (90, 410, 498). However, transesterification reactions within the fully assembled spliceosome require the presence of dephosphorylated SR proteins (90, 410, 660), which is mediated by protein phosphatase-1 (PP1) and -2A (PP2A) (409, 410). Based on these results and the observation that hyper- as well as hypophosphorylation of the RS domain inhibits SR protein-dependent constitutive splicing (479), it was proposed that sequential phosphorylation and dephosphorylation is required for SR proteins to function as general splicing factors (90).

Besides its relevance for constitutive splicing, phosphorylation appears also to be involved in alternative splicing, because hyperphosphorylation induced by overexpression of Clk/Sty kinase but also dephosphorylation induces a shift in splice site selection (165, 301). The phosphorylation state of SR proteins influences alternative splice site selection by affecting RNA binding specificity and protein-protein interactions with spliceosomal components and regulatory proteins. Phosphorylation of mainly the RS domain, but also of the remaining protein, enhances the RNA-independent interaction of SF2/ASF with U1-70K (659). In addition, phosphorylation of the RS domain of SF2/ASF reduces self-interaction and binding to SRp40 and Tra2 $\alpha$ , whereas binding to U2AF<sup>35</sup> is not affected (660).

The RS domain has been suggested not to significantly determine RNA binding specificity, because it had been shown that SF2/ASF and SC35 deletion mutants bind sequence-specifically to RNA in the absence of the RS domain (596, 702). However, since SRp40 containing an unphosphorylated RS domain bound to RNA, but failed to select a specific sequence, whereas SRp40 carrying a phosphorylated RS domain efficiently identified a high affinity RNA binding site in SELEX assays (595), it appears that in the presence of the RS domain its phosphorylation is necessary for sequence specific RNA binding. Consistent with the results obtained for SRp40, the unphosphorylated RS domain of SF2/ASF also induces non-sequence-specific RNA binding, which is abrogated by phosphorylation of the RS domain (659). These results led to the suggestion that the highly positive charge of the unphosphorylated RS domain

resulted in an RNA-specific but sequence-independent binding of the SR protein prevailing over sequence-specific binding by the two RNA binding domains (595).

The similar domain structure and the ability to complement splicing-deficient cellular extracts indicated that SR proteins might fulfil redundant functions in general splicing. Consistent with this hypothesis, the global pattern of gene expression was not found to be dramatically altered upon tissue-specific depletion of SC35 in a mouse model (154) or SF2/ASF from the chicken DT-40 cell line (346). However, the RNA binding specificity of SR proteins and their distinct effects on alternative splicing implied that they might also exert non-redundant functions. The existence of fundamental, non-redundant functions of SR proteins were demonstrated *in vivo* by the finding that knockout mice for the SR proteins SRp20 (288), SC35 (154, 634) and SF2/ASF (664) showed an early embryonic lethal phenotype [reviewed in (425)]. From these experiments it appears that, although inactivation of individual SR proteins may not cause widespread defects, non-redundant functions of SR proteins crucially contribute to alternative splicing patterns mediating essential processes during early development.

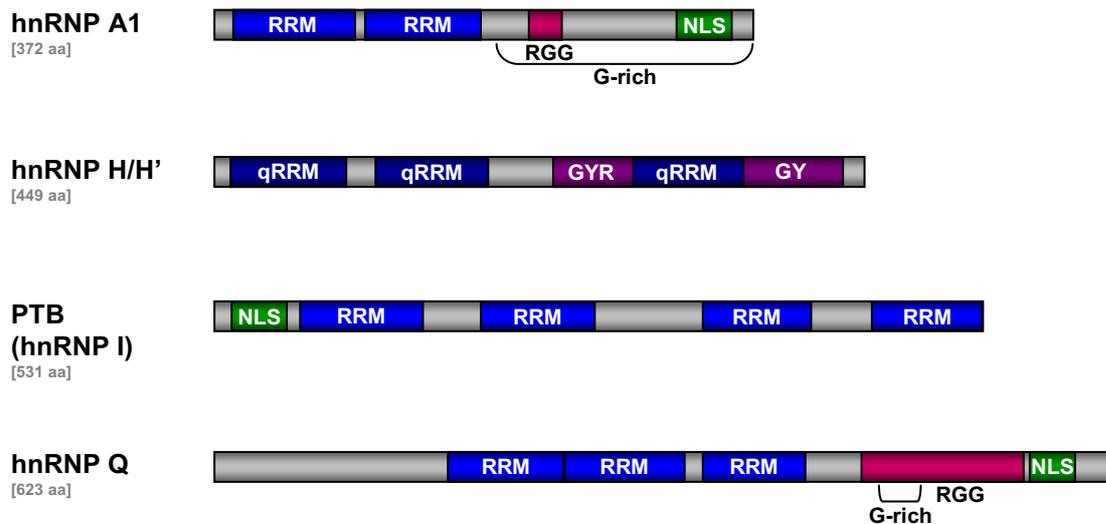
#### **A.3.6.2 Heterogeneous nuclear ribonucleoparticle (hnRNP) proteins**

Nascent pre-mRNA generated by RNAP II associates with various proteins and snRNPs. The term heterogeneous nuclear ribonucleoparticle (hnRNP) protein was initially applied to proteins associated with high molecular weight RNA (heterogeneous nuclear RNA) [reviewed in (163)]. The hnRNP protein family was further refined as group of proteins efficiently crosslinking by ultraviolet (UV) irradiation to hnRNA *in vivo* [reviewed in (396)]. Although hnRNP proteins are found generally associated with hnRNA, their non-randomly distribution bound to RNA suggested sequence-specific binding (403, 657). hnRNP proteins were implicated in various biological processes like telomere biogenesis, transcription, polyadenylation, RNA stability, nucleocytoplasmic RNA export, cytoplasmic RNA trafficking and translation [reviewed in (328)]. In addition, several members of the hnRNP protein family exert regulatory functions in alternative splicing [reviewed in (396)].

Due to their operational definition, hnRNP proteins constitute a structurally diverse family, in which currently more than 30 hnRNP proteins and derivative isoforms are grouped according to global protein identity (396). The most prominent structural features of the hnRNP protein family are at least one RNA binding motif of the canonical

RRM-, quasi (q)RRM- or hnRNP K-homology (KH)-type as well as auxiliary domains enriched in glycine, tyrosine, arginine, glutamine or asparagine (G-rich, GY-rich, GYR-rich, Q-rich, QN-rich or SRGY-rich regions) (Fig. I-12). Another striking characteristic is the number of protein isoforms generated by alternative splicing and posttranslational modification including phosphorylation (419, 620, 661), arginine methylation (255, 311, 361, 445, 455) and sumoylation (349, 624). Whereas phosphorylation of hnRNP proteins has been related to an indirect influence on alternative splice site selection (620), the exact role of methylation and modification by SUMO remains to be clarified.

Several members of the hnRNP protein family have been found to affect alternative splice site selection. Binding of hnRNP proteins was reported to predominantly repress splice site activation and exon recognition. Different models emerged for the mechanism of hnRNP-mediated changes in splice site selection. In the simplest case, binding of hnRNP proteins to splice sites or *cis*-regulatory enhancer sequences restricts the access of spliceosomal components or positively acting splicing regulatory proteins like it has been shown for the HIV-1 *cis*-regulatory element ESSV (157). The ability of hnRNP proteins to bind RNA cooperatively led to a propagation model, in which hnRNP proteins bind to one or more high-affinity binding sites and recruit additional hnRNP proteins to flanking low-affinity binding sites thereby creating a zone of local repression (693). Repression of splice site recognition frequently results in the activation of alternative splice sites (667). In addition to the inhibitory impact on early spliceosome assembly, polypyrimidine tract binding protein (PTB; hnRNP I) was reported to prevent cross-exon communication required for exon-definition [(279), reviewed in (575)] and, in another example, to block the transition from an exon-defining to an intron-defining spliceosome complex (541). A further mechanism of hnRNP-mediated splicing regulation is displayed by the interaction of hnRNP proteins bound at distant target sites, which was suggested to loop out the intervening sequence thereby bringing distant splice sites into proximity (56, 397, 448). Depending on whether or not an exon is located in the looped out sequence, interaction between distant hnRNP proteins can result in alternative splice site pairing or increased splicing of a constitutively spliced exon contrasting the prevalent inhibitory effect of hnRNP proteins on exon recognition. Additional evidence for a positive role of hnRNP proteins in splicing has been provided by reports demonstrating 5' splice site activation mediated by hnRNP protein binding in the neighbouring intron (200, 248, 271). However, so far no direct interaction between hnRNP proteins and U1 snRNP-specific proteins has been detected.



**Fig. I-12: Domain organisation of selected members of the hnRNP protein family.**

Members of the hnRNP protein family contain at least one RNA binding domain, but due to their operational definition display high diversity in protein interaction domains and overall domain structure. Domain positions were extracted from the Swiss-Prot database ([www.uniprot.org](http://www.uniprot.org)) [NLS: nuclear localisation signal; RGG: Arg-, Gly-rich sequence motif, RGG-box; RRM: RNA recognition motif; qRRM; quasi RRM].

The H subgroup of hnRNP proteins consisting of hnRNP H, H', F and 2H9 contains two (2H9) or three (hnRNP H, H' and F) copies of an RRM-related RNA binding domain (quasi-RRM; qRRM) and two auxiliary domains one rich in glycine, tyrosine and arginine (GYR-rich region) and the other glycine- and tyrosine-rich (GY-rich region) (Fig. I-12). Members of this subgroup share a high similarity (e.g. 96% similarity between hnRNP H and H' and 78% similarity between hnRNP H and F). All hnRNP H proteins recognise the consensus RNA motif DGGGD (where D is A, G, or U) (93, 521). Employing different modes of action, hnRNP H proteins appear to exert contrary effects on splicing regulation. Members of the hnRNP H group have been found to alter 5' ss selection, inhibit splicing and induce exon skipping (107, 466). For example, in the HIV-1 pre-mRNA activation of 3' ss A3 is inhibited by hnRNP H binding to ESS2p, which was suggested to result from an interference with U2AF<sup>35</sup> binding at the 3' ss (282). Additionally, exonic binding of hnRNP H proteins has been shown to interfere with splice site activation in several cellular pre-mRNAs (133, 497). In contrast, exonic as

well as intronic binding of hnRNP H has been reported to promote usage of neighbouring splice sites (94, 241). However, tethering of hnRNP H next to a 5' ss by the bacteriophage MS2 system activates 5' ss usage of an HIV-1-derived minigene only from the intronic position [S. Rosin, diploma thesis (500)]. Furthermore, it has been shown that hnRNP H proteins can improve splicing of distant splice sites without affecting U1 snRNP occupancy at the 5' ss, which was interpreted that hnRNP H proteins are able to contribute to the generation of splice site-activating intron loops (397). In tissue-specific splicing of the *c-src* pre-mRNA, hnRNP F and H were found to engage in an enhancer complex with neural-specific (n)PTB inducing inclusion of the N1 exon, whereas in non-neuronal cells complex formation with PTB generates a silencer complex causing exon skipping (113, 420). The assembly into regulatory complexes oppositely affecting *c-src* pre-mRNA splicing indicates that the effect of hnRNP H proteins on splicing may be directed by additional proteins within a splicing regulatory complex.

Polypyrimidine tract binding protein (PTB/hnRNP I) has been identified as splicing repressor in a number of alternatively spliced pre-mRNAs [reviewed in (575, 627)]. PTB consists of four RNA binding domains of the RRM-type connected by linker sequences and an N-terminal nuclear localisation signal (NLS) (Fig. I-12). Whereas RRM1 and RRM4 adopt the canonical RRM structure, RRM2 and RRM3 are extended by an additional  $\beta$ -strand [(128, 559), reviewed in (23)]. The RNA motif recognised by PTB consists of polypyrimidine-rich sequences with a 4-9 nt core motif of nearly alternating cytosines and uridines (13, 101, 470, 562). Consistent with the pyrimidine-rich RNA binding sites, PTB-mediated inhibition of 3' ss activation has been related to a competition with U2AF for binding to the PPT (562), e.g. in exon 3 splicing of the  $\alpha$ -*TM* pre-mRNA (356). However, splicing regulatory PTB-binding sites are not necessarily required to overlap 3' ss elements. Like hnRNP H and hnRNP A1, PTB has been suggested to loop out intervening introns by occupying distant binding sites (448), which might promote alternative splice site usage or enhance constitutive splicing of enlarged introns. Additionally, PTB has been described to inhibit splicing by interfering with the communication between spliceosomal components either across the exon e.g. in the *fas* pre-mRNA (279) or across the intron e.g. in the *c-src* pre-mRNA (541). In mammals, tissue-specific paralogues of PTB are expressed in the brain [neuron-specific (n)PTB] (395, 474) and in hematopoietic cells (ROD1) (666) facilitating differential splicing of PTB-regulated pre-mRNAs in these tissues.

In contrast to PTB and the hnRNP H subgroup, hnRNP Q is less well characterised. hnRNP Q is alternatively spliced into three mRNA isoforms (427). The longest isoform, hnRNP Q3 encodes three RRM-type RNA binding domains, a bipartite NLS and an RGG-box motif (Fig. I-12). In addition, an acidic region rich in glutamine and asparagine is located in the C-terminus. hnRNP Q has been reported to bind to U-rich and AU-rich sequences (55). The identification of hnRNP Q in purified spliceosomes indicated that it might contribute to splicing regulation (441). This suggestion was substantiated by the finding that depleting hnRNP Q and the structurally related hnRNP R from splicing extracts reduced splicing efficiency of adenovirus- and chicken *δ-crystallin*-derived pre-mRNA substrates (427). Recently, the homolog of hnRNP Q and R in *C. elegans*, HRP-2, was reported to regulate alternative exon inclusion in the *unc-52* mRNA by binding to intronic UCUAUC motifs (290). Contrasting its positive effect on splicing, hnRNP Q isoforms have also been implicated as negative factors contributing to skipping of exon 7 in the *SMN2* pre-mRNA (109).

#### **A.3.6.3 Combinatorial control of alternative splicing by splicing regulatory proteins**

Many alternative splicing events are regulated by a complex interplay of enhancers and silencers, in which the activatory function of SR proteins is frequently counteracted by members of the hnRNP family. Since both splicing regulatory families are ubiquitously expressed in metazoans, cell-specific splice site selection was suggested to be determined by the ratio of the members of both splicing regulatory protein families varying widely between different tissues and cell lines (242, 676). Consistent with this hypothesis, splice site selection is affected by the ratio of SR proteins and hnRNP A/B proteins bound to the RNA substrate (674). However, recently also cell-specific splicing regulatory factors have been identified (449, 641) indicating the possibility that the contribution of cell-specific splicing regulatory proteins has been underestimated so far.

In addition to the general antagonism observed for SR proteins and hnRNP proteins, counteracting effects have also been described between members of the SR protein family, e.g. for SRp20 and SF2/ASF regulating splicing of the *SRp20* pre-mRNA (287). Furthermore, the SR-related protein SRp86 was found to differentially modulate the activity of other SR proteins by enhancing the effect of SRp20, but interfering with the splicing regulatory function of SF2/ASF, SC35 and SRp55 (32), indicating that also SR-

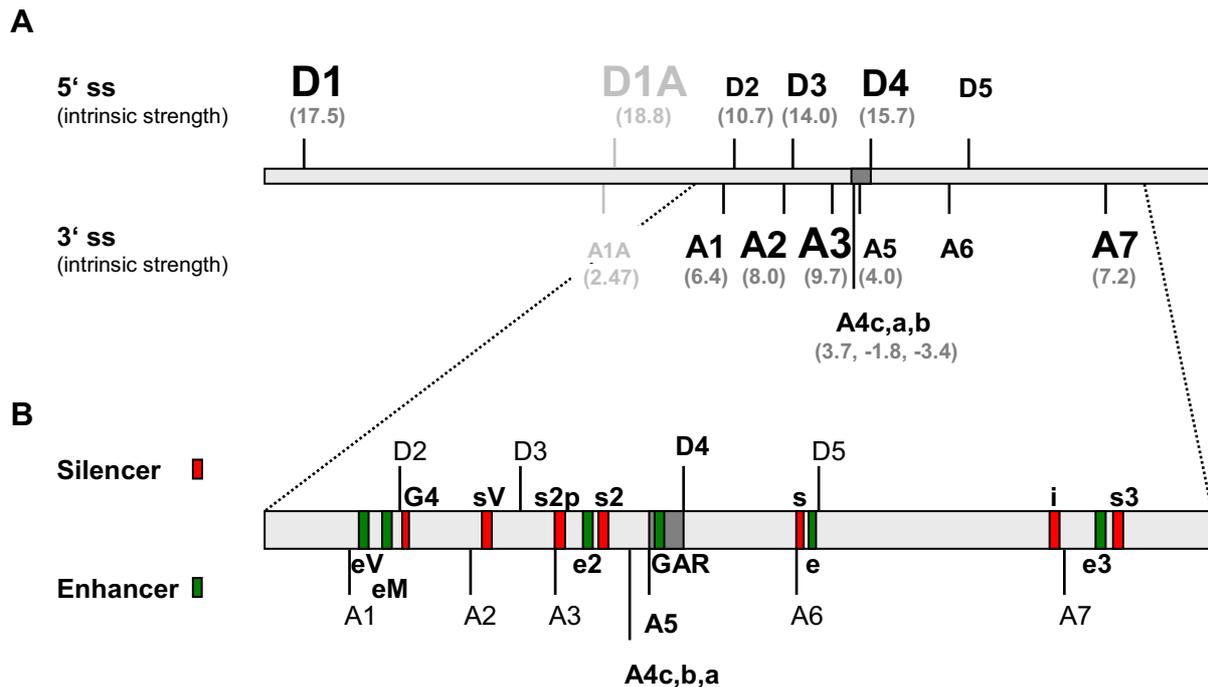
related proteins are able to antagonise the activity of classical SR proteins. Likewise, with the CELF proteins (CUG-BP and ETR3-like factors) (334) also antagonists other than SR proteins have been identified for hnRNP proteins. Changed ratios of both proteins have been described to oppositely regulate splicing of the *cTNT* and *α-actinin* pre-mRNA (104, 227). Therefore, although a large number of examples are documented, in which oppositely directed activities of SR proteins and hnRNP proteins control splicing, additional regulatory networks exist within splicing regulatory protein families and also with other regulatory factors.

#### **A.4 Exploitation of cellular pre-mRNA splicing by HIV-1**

Throughout viral gene expression, the cellular splicing machinery processes the HIV-1 pre-mRNA into mRNAs of the 1.8 kb class encoding regulatory and accessory proteins. However, in the intermediate and late phase the viral mRNA isoform pattern shifts towards intron-containing isoforms belonging to the 4 kb and 9 kb class indicating that efficient intron removal decreases in the course of HIV-1 gene expression. The interplay of splice sites and *cis*-acting splicing regulatory elements constitutes an essential prerequisite for the temporally regulated viral mRNA expression.

##### **A.4.1 *Cis*-regulatory elements control splice site recognition in the HIV-1 pre-mRNA**

Depending on the viral isolate, at least four 5' ss and eight 3' ss are used for alternative splicing of the HIV-1 pre-mRNA (Fig. I-13A). The high number of alternative mRNA isoforms was initially related to the low similarity of most HIV-1 splice sites to the consensus sequences. Early investigations using a subgenomic  $\beta$ -globin-derived minigene revealed that only 5' ss D1 and D4, displaying the highest similarity of the HIV-1 splice sites to the consensus sequences, are as efficiently used as the intrinsically strong  $\beta$ -globin control 5' ss, whereas D2 and D3 were used two to three times less efficiently (447). Analysing 3' ss strength, the same study inferred from inefficient usage of each HIV-1 3' ss in the  $\beta$ -globin hybrid minigene that all viral 3' ss are weak. However, a number of *cis*-regulatory sequences affecting HIV-1 splice site usage had been identified since then, which with a single exception are located in exons



**Fig. I-13: Intrinsic strength of HIV-1 splice sites and localisation of *cis-acting* splicing regulatory elements.**

**(A)** Distribution and intrinsic strength of 5' ss (D1-D5) and 3' ss (A1-A7) within the HIV-1 pre-mRNA. Depending on the viral isolate, the pre-mRNA contains at least 12 splice sites participating in alternative splicing. The nomenclature of the splice sites refers to (134, 373). The size of the label corresponds to the intrinsic strength of the respective splice site, which was assessed using the HBond Score algorithm ([www.uni-duesseldorf.de/rna](http://www.uni-duesseldorf.de/rna)) for 5' ss and the MaxEntScore algorithm ([http://genes.mit.edu/burgelab/maxent/Xmaxentscan\\_scoreseq\\_acc.html](http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html)) for 3' ss. The position of exon 5 and the upstream located alternative 3' ss cluster consisting of 3' ss 4c, 4a, 4b and 5 is indicated (dark grey box). A1A and D1A were identified in a cDNA screen of HIV-1 infected cells and appear to be preferentially involved in mRNA stabilisation (373). However, the relevance of these splice sites for alternative splicing during viral gene expression remains to be demonstrated. Splice site D5 has exclusively been reported for the HIV-1 isolate HXB2 (39).

**(B)** Location of *cis-acting* splicing regulatory elements within the HIV-1 pre-mRNA. The enlargement of the 3' half of the pre-mRNA depicts the location of the known enhancer (green) and silencer elements (red) regulating viral pre-mRNA splicing. Exon 5 is flanked by 5' ss D4 and 3' ss A5 and contains the GAR ESE in the 5'-half. The 3' ss A4c, A4b and A4a cluster directly upstream of exon 5 [e: ESE (646), e2: ESE2 (674), e3: ESE3 (579), eM: ESEM1/M2 (296), eV: ESEvif (172), G4: intronic guanosine-rich silencer (172), GAR: GAR ESE (91, 295), i: ISS (603), s: ESS (94, 646), s2: ESS2 (11, 555), s2p: ESS2p (282), s3: ESS3 (12, 556, 579), sV: ESSV (50, 380) (adapted from (588))].

(Fig. I-13B) [(11, 12, 50, 91, 172, 282, 295, 296, 380, 555, 556, 579, 603, 674), reviewed in (587)]. A re-evaluation of HIV-1 splice site strength in the absence of *cis*-regulatory elements revealed that A2 and A3 are intrinsically strong 3' ss, whereas A1, A4c, A4a, A4b, A5 and A7 are intrinsically weak (296). Incorporating the respective downstream flanking exonic sequences into the minigene reporter demonstrated that splicing at each HIV-1 3' ss is profoundly controlled by the neighbouring *cis*-regulatory elements (296). Based on these results, regulation of HIV-1 3' ss usage emerged to rely on the formation of distinct splicing regulatory complexes either enhancing intrinsically weak 3' ss or repressing intrinsically strong 3' ss. The conservation of the intrinsic splice site strength between HIV-1 isolates from different clades suggests that the particular splice site strength is essential for efficient virus replication (587).

The *cis*-regulatory sequences controlling HIV-1 alternative splicing appear to act via different mechanisms. All HIV-1 *cis*-regulatory sequences have been found to recruit splicing regulatory proteins belonging either to the SR protein family (SF2/ASF, SC35, SRp40, SRp75) or to the hnRNP protein family (hnRNP A/B, hnRNP H) (50, 91, 172, 282, 296, 603, 674). The impact of *cis*-regulatory elements on spliceosome assembly has been investigated for ISS, ESS2, ESS3 and ESSV, which were found to interfere with E or A complex formation (12, 157, 556, 603). The importance of regulated HIV-1 splice site usage was underlined by the observation that mutations within HIV-1 splicing silencer elements ESS2 and ESSV reduce the production of infectious virus particles [(380), reviewed in (587)]. Although the HIV-1 pre-mRNA is highly structured (46, 643, 648), only one *cis*-regulatory element located in the viral strain-specific exon 6D has been described to modulate HIV-1 alternative splice site activation via altered RNA secondary structure (280). Nevertheless, it has been proposed that mutations that stabilise RNA stem loops sequestering splice sites D1 or A3 restrict splice site accessibility and thus reduce splice site usage (2, 283).

Activation of the major 5' ss D1 is required for generation of all spliced HIV-1 mRNAs. Contrasting the general proximity of HIV-1 splice sites and splicing regulatory elements, it was suggested that neither upstream nor downstream sequences affect splicing at D1 (447). This proposal is consistent with the strong similarity of this 5' ss to the consensus sequence characterising D1 as intrinsically strong splice site. However, directly neighbouring exonic and intronic sequences were omitted from the initial analysis of overall D1 splice site strength (447). A recent study indicated that the downstream located region might support D1 usage in an *in vitro* splicing assay (521). Nevertheless,

in the same study splicing at D1 was not activated by the intron in the presence of the downstream located splicing silencer element ESS2. Therefore, it remains to be demonstrated whether the intron activates splicing at D1 during viral gene expression.

Within the central region of the HIV-1 pre-mRNA, the internal exon 5 is flanked by 5' ss D4 and 3' ss A5. Upon activation of both splice sites, inclusion of exon 5 results in the expression of *nef*-mRNA isoforms. Alternative usage of 3' ss A4c, A4a and A4b immediately preceding A5 results in the inclusion of enlarged internal exons thereby generating *rev*-mRNA isoforms. The unusual high frequency of 3' ss used to provide *rev*-mRNA isoforms underlines the importance of Rev for viral replication. Since Rev represents an essential export factor for intron-containing HIV-1 mRNAs (388) and Nef crucially downregulates the expression of cellular surface receptors and interferes with signal transduction pathways [reviewed in (187, 491)], balanced expression of both proteins is required for efficient HIV-1 replication. All 3' ss immediately upstream of exon 5 share a low similarity to the 3' ss consensus motifs for BPS and PPT and their usage depends on the presence of the downstream exon (296). Despite the similarity of D4 to the consensus sequence, recognition of the 5' ss by U1 snRNP and splicing depend on a guanosine-adenosine-rich (GAR) splicing enhancer located in the 5'-half of exon 5 (295). Although the GAR enhancer was described to also activate an upstream located heterologous 3' ss in *in vitro* splicing assays (91), it is not clear whether in the viral sequence context the GAR enhancer activates either the upstream or the downstream flanking splice site or both.

#### **A.4.2 Modulation of cellular splicing regulatory protein expression and activity after HIV-1 infection**

Considering the overall shift in the HIV-1 mRNA expression pattern to isoforms with increasing intron-content, it appears conceivable that splice site activation by splicing regulatory elements decreases during viral gene expression. Since most HIV-1 *cis*-regulatory sequences recruit splicing regulatory factors to modulate splice site usage, changes in the expression or functionality of splicing regulatory proteins might contribute to the shift in viral pre-mRNA alternative splicing. Consistent with this hypothesis, several investigations reported considerable infection-induced changes in mRNA and protein expression of splicing regulators of the SR protein family, like SF2/ASF and hTra2- $\beta$ , as well as the hnRNP protein family, like hnRNP A1 and

hnRNP H (1, 100, 161, 195, 383). Although the expression of SR proteins as well as hnRNP proteins is not uniformly up- or downregulated and appears to vary depending on the respective experimental system, virus-induced changes in the splicing factor expression profile likely participate in the temporal control of HIV-1 pre-mRNA splicing.

Besides altering the expression of splicing regulatory proteins, HIV-1 gene expression might also affect the functionality of splicing factors by influencing post-translational modifications required for effector function. Virus-induced dephosphorylation of SR proteins concomitant with changes in alternative splice site usage has been reported to occur during infection with adenovirus, vaccinia virus or herpes simplex virus (HSV) [(269, 301, 534), reviewed in (337)]. The HIV-1 regulatory protein Tat has been described to interact with protein phosphatase-1 (PP1) (14), which is also involved in SR protein dephosphorylation. Additionally, Tat was observed to redistribute SC35 from nuclear speckles into the proximity of the nucleolus (195). Since the localisation of SR proteins in nuclear speckles has previously been assigned to their phosphorylation state, the interaction of Tat with PP1 might be envisioned to affect SR protein functionality by modulating phosphorylation. Like Tat, also Nef was found to interact with cellular proteins involved in phosphorylation cascades. A number of kinases involved in different signal transduction pathways like e.g. Fyn, Lck and Src have been shown to bind to Nef [reviewed in (205, 491)]. Eventually, these interactions might result in the activation of the kinase Akt, which was shown to phosphorylate SR proteins SF2/ASF, SRp40 and 9G8 (57, 465). Therefore, signal transduction pathways activated by Nef might also result in the phosphorylation of SR proteins. Since phosphorylated as well as dephosphorylated SR proteins are essential for pre-mRNA splicing, increasing amounts of Tat and Nef might modify the composition of splicing regulatory complexes formed on the HIV-1 pre-mRNA thereby contributing to the switch from the early into the intermediate phase of viral gene expression.

## **A.5 Expression of intron-containing HIV-1 mRNAs**

The shift from early to intermediate and late HIV-1 gene expression is hallmarked by the appearance of intron-containing mRNAs of the 4 kb and 9 kb class encoding structural proteins and enzymes but also providing the viral genome in the form of the completely unspliced 9 kb mRNA. A prerequisite for the expression of intron-containing viral mRNA

is the inefficient recognition of the 3'ss in the HIV-1 pre-mRNA based on non-consensus splice site sequences and *cis*-regulatory elements (296, 447, 578), which was assumed to result in a slow splicing kinetic of the viral pre-mRNA thus increasing the half-life of intron-containing viral mRNAs within the nucleus compared to the presence of efficient splice sites (103, 296). However, to ensure temporally regulated HIV-1 gene expression, premature cytoplasmic appearance of intron-containing mRNA in the early phase of viral gene expression is inhibited by a number of negatively acting elements within the HIV-1 pre-mRNA referred to as instability (*ins*) elements or *cis*-repressive sequences (CRS) (74, 121, 239, 382, 438, 499, 524, 526, 530). The viral regulatory protein Rev expressed in the early phase binds to intron-containing viral mRNAs and counteracts the inhibitory effects of *ins*- and CRS-elements and also of cellular mRNA surveillance mechanisms, hence stabilising intron-containing HIV-1 mRNA (179, 386). Rev binding facilitates the nuclear export of 4 kb and unspliced viral mRNAs [(168, 179, 388), reviewed in (473)] by recruiting the cellular export factor Crm1 (chromosome region maintenance-1; exportin-1) to intron-containing viral mRNAs (186, 443, 654).

#### **A.5.1 Rev-mediated expression of intron-containing HIV-1 mRNAs**

The translocation of RNA into the cytoplasm is mediated by nuclear export receptors acting as molecular adaptors between the RNA cargo and nucleoporins lining the nuclear pore [reviewed in (37)]. The bulk of intron-containing metazoan mRNAs become export-competent by recruiting export factors via the exon junction complex (EJC), which is formed coincident with the second catalytic step of the splicing reaction 20-24 nt upstream of the generated exon-exon junction without RNA sequence specificity (302, 344, 345, 688). The EJC core components eIF4AIII, MLN51, Magoh and Y14 generate a dynamic platform for adaptor proteins functioning in RNA quality control, export and translation [(25, 343, 605), reviewed in (342)]. Although several nuclear export adaptors, e.g. Hpr1 and yeast Nab2, were implicated in mRNA export [reviewed in (276, 304)], the nuclear protein Aly represents the best analysed example of a nuclear export adaptor protein mediating EJC-dependent mRNA export. Aly is recruited to the mRNA by interacting with the transiently associated EJC component UAP56 (370, 590), an ATP-dependent RNA helicase also involved in spliceosome assembly [(182, 548), reviewed in (543)]. In turn mRNA-bound Aly recognises the heterodimeric nuclear

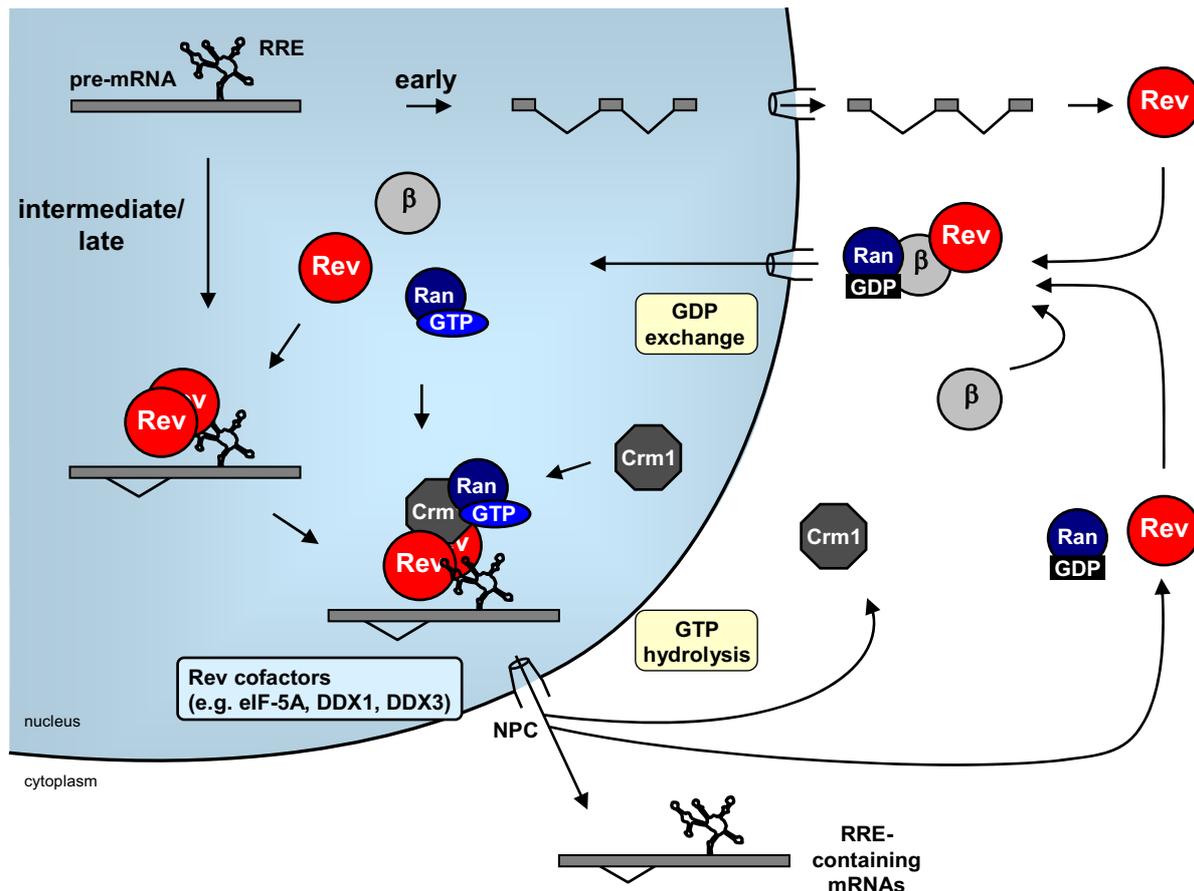
export factor TAP/Nxt1 (Nxf1/p15) (589, 592) thereby facilitating rapid and efficient nuclear export of spliced mRNAs [(343, 369, 691), reviewed in (490)].

In contrast to spliced mRNAs, export of unspliced RNAs lacking EJC necessitates the formation of an alternative platform to recruit nuclear export factors. Whereas simple retroviruses, e.g. the Mason-Pfizer monkey virus (MPMV), employ a constitutive transport element (CTE) (72, 171), an RNA sequence that is directly recognised by the cellular export receptor TAP (231), the complex retrovirus HIV-1 encodes an export adaptor, the regulatory protein Rev, in order to facilitate the nuclear export of intron-containing viral mRNA. Rev binds to the Rev-responsive element (RRE) (120, 138, 141, 250, 313, 389, 680), an RNA sequence located in the 3'-terminal intron of the viral pre-mRNA defined by D4 and A7 (148, 239, 388, 499). The Rev-RRE interaction has been shown to be essential for the expression of all intron-containing 4 kb and the genomic HIV-1 mRNA (148, 178).

Based on thermodynamic calculations and biochemical probing, it was suggested that the RRE forms a complex RNA secondary structure, which consists of four stem loops (designated II-V) protruding from the long central stem I (105, 313, 388, 393). The secondary structure of the RRE appears to be the major determinant for the interaction of Rev with the RRE (450). Rev binds to the RRE via a domain rich in basic residues located in the N-terminal half of the 116 aa protein (314, 681), which also serves as NLS and is covered by an extended domain required for multimerisation (264, 330, 384, 471, 681). Binding of Rev to the RRE was reported to start within a high-affinity binding site at the base of stem loop IIB (35, 251, 262, 389, 613). Further *in vitro* binding studies revealed that distinct stoichiometric Rev-RRE complexes are formed suggesting that starting from the high-affinity binding site multiple copies of Rev multimerise along the RNA through cooperative protein-protein and protein-RNA interactions (250, 251, 313, 385, 393). Recently, monitoring the dynamic of Rev multimerisation on an RRE target by single-molecule TIRF (Total Internal Reflection Fluorescence) microscopy revealed that up to four individual Rev monomers bind subsequently and cooperatively to an RRE target truncated within the central stem I (477). In fact, binding of a single Rev monomer to the RRE has been found to be unable to mediate RNA export indicating that at least dimerisation on the RRE target is essential for Rev function (384, 385).

In addition to the domains involved in multimerisation and RNA binding, the C-terminal half of Rev contains a leucine-rich activation domain acting as NES (181, 384, 412).

Due to the presence of both an NES and an NLS Rev constantly shuttles between the cytoplasm and the nucleus (293, 294, 412). The activation domain of Rev is bound by the karyopherin Crm1 (Exportin-1) (186, 443), a member of the importin- $\beta$  superfamily of nuclear transport receptors (577). Interaction of Rev with the RRE directs viral intron-containing mRNAs to the Crm1-mediated export pathway (186, 194, 443, 654) (Fig. I-14). Crm1 recognises leucine-rich NES of a variety of cellular proteins acting also



**Fig. I-14: Interaction of Rev with the RRE directs intron-containing HIV-1 mRNAs to the CRM1-mediated nuclear export pathway.**

In the early phase of HIV-1 gene expression Rev is translated from completely spliced mRNA of the 1.8 kb class. The nuclear localisation signal of Rev is recognised by importin- $\beta$  mediating Ran-GDP-dependent nuclear import. After conversion of Ran-GDP into Ran-GTP the Rev import complex dissociates. In the intermediate and late phase, Rev binds to an RNA secondary structure (RRE) located in the 3'-terminal intron of the viral pre-mRNA. After at least dimerisation of Rev, the RNA-protein complex is recognised by Crm1. Recruitment of Ran-GTP allows the translocation of the Rev-RNA export complex through the nuclear pore complex (NPC). Several cofactors have been found to be additionally involved in Rev-mediated RNA export. In the cytoplasm the cargo complex disassembles providing intron-containing viral RNA for translation and virion assembly. Rev and cellular transport proteins are re-imported into the nucleus thereby completing the Rev-mediated RNA export cycle [ $\beta$ : importin- $\beta$ , RRE: Rev-responsive element]

as adaptor proteins in RNA export, like e.g. APRIL and pp32 (73), and thereby translocates a subset of endogenous proteins and RNAs into the cytoplasm [(73, 525), reviewed in (135, 272)]. Via the HIV-1 adaptor protein Rev, Crm1 indirectly binds to the RRE and enables the translocation of viral intron-containing RNA through the NPC by interacting with nuclear pore proteins (NUPs) (20, 443). The association of Rev and Crm1 requires the presence of the GTP-bound form of Ran (186), a key GTPase regulating importin- $\beta$  function. Consistent with the only low affinity of the Rev NES to Crm1 (20), additional cofactors, e.g. eIF-5A and RNA helicases DDX1 and DDX3, have been found to be required for efficient translocation of the Rev-RRE complex into the cytoplasm [reviewed in (249, 593)]. Recently, it was found that the N-terminal aa 9-14 of Rev mediate the interaction with hnRNP A1, Q, K, R and U, which might allow Rev to be additionally involved in diverse cellular processes (238).

Besides recognising the RRE, *in vitro* assays indicated that Rev binds to the packaging signal (PSI) in the highly structured 5'-UTR of the HIV-1 pre-mRNA (198). The internal loop A of stem loop 1 (SL1) located upstream of D1 within the PSI contains a highly conserved purine-rich loop similar to the high-affinity Rev binding site of the RRE (225). Mutation of loop A decreased binding of Rev to SL1 (198) and caused a replication defect *in vivo* (226). So far, binding of Rev to SL1 in the 5'-UTR has been shown to stimulate the translation of reporter genes located downstream of the HIV-1 leader *in vitro* (228) underlining the role of Rev in enhancing cytoplasmic utilisation of viral mRNAs by associating them with the translation machinery [(136), reviewed in (229)].

In addition to Rev, the PSI is recognised by the viral protein precursor Gag via its nucleocapsid component (8, 150, 208, 209). A previous report suggested that binding of Gag to the unspliced mRNA might interfere with further processing of the RNA including splicing [C.K. Damgaard, PhD thesis (142)] and thus support Rev to facilitate the expression of completely unspliced HIV-1 RNA. However, the relevance of Gag binding to the PSI for the expression of intron-containing viral mRNA remains to be investigated.

Despite its molecular delineation, the subnuclear localisation of the Rev-RRE interaction is largely unknown. The predominant subnuclear accumulation of Rev in nucleoli and nuclear speckles characterised by the presence of the splicing factor SC35 (179, 384, 471) led to the suggestion that a specific localisation of Rev might be required for its function as nuclear export receptor. The major function of the nucleolus has traditionally been assigned to ribosomal RNA transcription and subunit assembly, but also several

non-ribosomal processes like e.g. U6 snRNA processing have been found to occur within the nucleolus [reviewed in (452, 467)]. However, controversial results were obtained regarding the relevance of the nucleolar accumulation of Rev for its function in RNA export. It has been shown that Rev multimerisation, which is required for Rev-mediated RNA export, predominantly occurs in the nucleoli (139). In addition, mutations of the nucleolar localisation signal (NoLS) in the Rev protein that resides within the NLS but extends into the N-terminus reduced the efficiency of Rev-mediated mRNA export (122, 626). Since also HIV-1 RNA localises to the nucleolus (416, 417), an interaction of Rev and intron-containing viral RNA might occur within the nucleoli. However, the finding that nucleolar degradation of viral RNA reduces viral replication (416) might represent an only indirect effect of depleting the pool of nucleoplasmic intron-containing viral mRNAs available for nuclear export and thus does not necessarily indicate an essential function of the nucleolus in viral RNA metabolism. The suggestion that nucleolar localisation of Rev might not be essential for its function in RNA export is underlined by the finding that redistributing Rev from the predominant nucleolar to a more nucleoplasmic localisation by overexpressing Nap1 (nucleosome assembly protein 1) enhanced Rev function (119). In view of these observations, the role of nucleolar localisation of Rev remains to be conclusively determined.

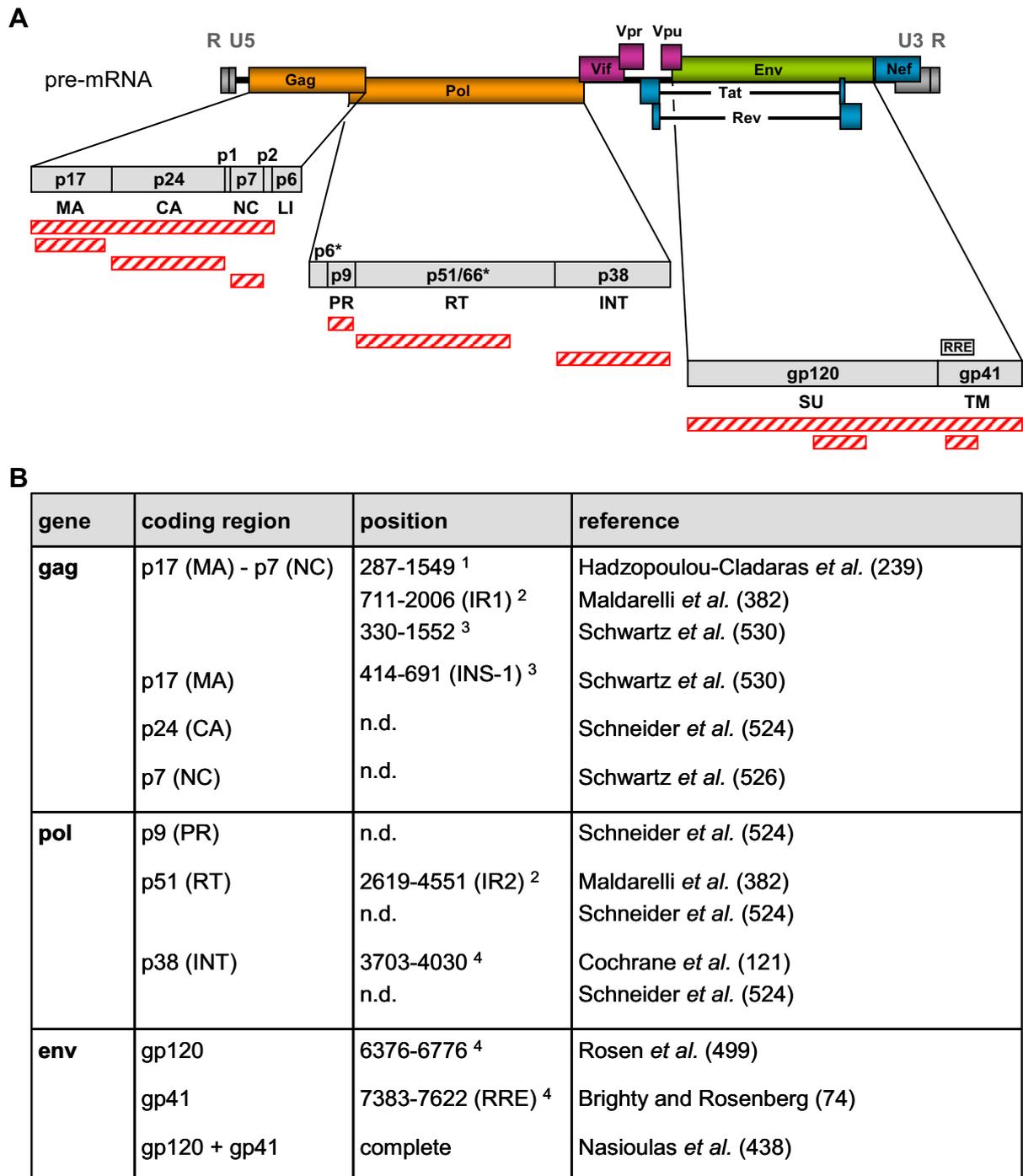
In addition to its accumulation within nucleoli, Rev was found to associate with SC35-containing nuclear speckles (137, 294). Nuclear speckles are thought to store splicing factors, which are supplied to nearby active sites of transcription thus facilitating cotranscriptional splicing [(421), reviewed in (340)]. Although nuclear speckles are primarily considered as storage compartment, a gradient of intron-containing pre-mRNA of several cellular genes from the periphery to the interior of nuclear speckles indicated that introns remaining after cotranscriptional pre-mRNA processing might be removed mainly at the periphery of but also within nuclear speckles (240, 285, 662, 663). Since it was proposed that the HIV-1 pre-mRNA is not spliced cotranscriptionally and in addition a partial colocalisation of a subgenomic *env*-mRNA with SC35-containing speckles was observed after minigene transfection (176, 684), it is conceivable that intron-containing viral pre-mRNAs reach nuclear speckles, where they might interact with Rev during the intermediate and late phase of viral gene expression. This idea was supported by the observation that Rev-deficiency of distinct transfected HIV-1 proviruses increases the amount of intron-containing viral RNA localised at nuclear speckles (47, 118). Accordingly, the interaction of Rev with the RRE might be envisioned to circumvent

subsequent splicing of the viral pre-mRNA at the periphery of nuclear speckles by targeting intron-containing viral mRNAs to nuclear export.

### **A.5.2 Inhibitory elements interfere with the expression of intron-containing HIV-1 mRNAs**

The chronological order of HIV-1 mRNA expression requires that premature expression of intron-containing viral transcripts is prevented in the early phase of viral gene expression. Several negatively acting *cis*-regulatory elements within the *gag/pol*- and *env*-coding sequences of the HIV-1 pre-mRNA referred to as inhibitory sequences (ins), inhibitory regions (IR) or *cis*-repressive sequences (CRS) have been found to interfere with the expression of structural proteins and enzymes from intron-containing mRNA in the absence of Rev (74, 121, 239, 382, 438, 499, 524, 526, 530) (Fig. I-15). HIV-1 inhibitory elements don't share a common sequence motif, which is in accordance with the observation that these elements affect different levels of gene expression. Whereas an inhibitory element in the *pol*-coding region apparently impedes efficient translation (121), *cis*-regulatory elements in the Gag-ORF (382) and within the RRE in the Env-ORF (74) were found to impair nuclear export of the respective RNA. As a mechanism for the reduced cytoplasmic appearance of reporter mRNA in the presence of inhibitory elements, it was suggested that these inhibitory elements harbour retention signals restricting intron-containing HIV-1 mRNA to the nucleus (74, 382). However, interacting proteins or structures retaining inhibitory element-containing mRNA within the nucleus remain to be identified. In addition to inhibitory elements influencing nucleocytoplasmic RNA distribution, the p17 (MA)-encoding sequence of the *gag* inhibitory region (p17-ins) (530) and the CRS element in the *integrase*-encoding region (121) were described to decrease the stability of reporter mRNAs. Consistently, in the absence of Rev intron-containing HIV-1 mRNAs exhibit a short nuclear half-life, whereas completely spliced viral mRNAs appear to be quite stable in the nucleus (386).

Further investigations regarding the stability of intron-containing HIV-1 mRNA concentrated on the destabilising activity of the p17-ins element, which was attributed to the high AU content existing throughout the Gag-/Pol-ORF (530). Substituting several nucleotides of the AU-rich clusters thereby decreasing the AU content (526) or optimising the Gag-ORF to the more GC-rich human codon usage (216) allowed constitutive expression of Gag from subgenomic HIV-1 minigenes in the absence of Rev



**Fig. I-15: Distribution of *cis*-regulatory elements interfering with the expression of structural proteins and enzymes from intron-containing HIV-1 mRNA isoforms.**

**(A)** Nearly all regions of the Gag-, Pol- and Env-ORFs encoded by the HIV-1 pre-mRNA harbour *cis*-acting elements (red striped boxes) counteracting the expression of the respective structural proteins and enzymes (upper panel) [n.d.: not determined, R: repeat; RRE: Rev responsive element, U5: unique 5' region; U3: unique 3' region; p6\*: extended linker protein p6 (LI) caused by ribosomal frameshift, p51/66\*: two isoforms of reverse transcriptase (RT) are proteolytically processed from the Gag-Pol precursor protein, which either contain (p66) or lack (p51) the carboxy-terminal RNase H activity].

**(B)** Localisation of negatively acting *cis*-regulatory elements. Nucleotide numbering <sup>1</sup> according to (645), <sup>2</sup> according to (5), <sup>3</sup> starts with the first nucleotide of R, <sup>4</sup> according to (486).

underlining the importance of AU-rich clusters for the inhibitory effect of the p17-ins element.

AU-rich elements (ARE) already emerged as common *cis*-acting motifs regulating the stability of a variety of cellular mRNAs [reviewed in (34, 308)]. AREs are frequently found in the 3'-UTR of early response genes (ERGs), e.g. interferons and cytokines, and are thought to function as binding sites for proteins cooperating or interfering with cellular RNA degradation mechanisms. It has been suggested that destabilising proteins binding to AREs in the 3'-UTR either deadenylate existing poly(A)-tails or interfere with efficient polyadenylation and thus allow subsequent 3'-5' exonucleolytic degradation by a large multi-protein complex, the exosome [(428), reviewed in (267, 523)]. Alternatively, mRNA decapping induces rapid degradation by the 5' exonucleases Rat1 and Xrn1 [reviewed in (464)] suggesting that ARE-containing mRNAs might be degraded in 3'-5' as well as in 5'-3' direction. Despite its high AU-content the p17-ins element lacks the characteristic pentanucleotide AUUUA found in classical AU-rich elements (108). Since the p17-ins element failed to cause efficient degradation of a subgenomic *β-globin* mRNA when compared to the ARE derived from the 3'-UTR of the *GM-CSF* mRNA (418), the p17-ins element presumably also functionally differs from classical AREs found in cellular mRNAs. This suggestion is supported by the finding that proteins identified to bind the p17-ins element, i.e. hnRNP A1 (436), all members of the subgroup of hnRNP H proteins (93, 521) and PABP1 (poly(A)-binding protein 1) (7), are typically not observed to induce regulated mRNA turnover. Additionally, an extended RNA target containing the p37-coding region of the Gag-ORF comprised of the p17 (MA)- and the p24 (CA)-encoding sequence is recognised by a heterodimer formed of p54nrb (nuclear RNA binding protein 54kDa) and PSF (PTB-associated splicing factor) (696). However, binding of none of these proteins to their target sequences in the Gag-ORF has conclusively been shown to account for the destabilisation mediated by the p17-ins element on HIV-1 RNA so far.

In contrast to typical cellular AREs, the impact of the p17-ins element on RNA stability strongly varies depending on the minigene used (382, 418, 530). Whereas a hybrid *tat/gag*-mRNA was strongly destabilised by insertion of the ins-sequence (530), the same element only marginally reduced RNA stability when situated downstream of an efficiently spliced *β-globin* minigene (418) or did not affect RNA steady state levels as part of the Gag-ORF inserted downstream of a *CAT* minigene (382). The reduced influence of the p17-ins element on RNA stability in heterologous minigenes indicates

---

that additional elements in the viral pre-mRNA might cooperate with the p17-ins element to interfere with the expression of intron-containing viral RNA. A synergistic cooperation of the p17-ins element has already been demonstrated with the RRE in a heterologous *β-globin* minigene (418, 436). This functional synergy has been shown to facilitate Rev-mediated expression of intron-containing reporter mRNAs, which otherwise fails using *β-globin*-derived minigenes harbouring the RRE in the intron (103) or the terminal exon (418). This finding was quite unexpected considering that the repressing effect of HIV-1 inhibitory elements on viral gene expression is rescued by the presence of Rev revealing that the Rev-RRE interaction is able to counteract nuclear RNA retention and degradation initiated by *cis*-regulatory elements within intron-containing HIV-1 mRNA. The synergistic activity of the RRE and the p17-ins sequence on the expression of intron-containing mRNA has been attributed to an inhibitory influence of hnRNP A1 and a non-specified 50 kDa protein bound to the p17-ins element on the removal of the upstream situated *β-globin* intron (436). Impeding the splicing reaction might allow Rev to interact with the RRE-containing mRNA thereby facilitating the expression of unspliced reporter mRNA in the cytoplasm. However, it remains to be demonstrated whether the p17-ins element cooperates with downstream located elements of the HIV-1 pre-mRNA within the viral sequence context.

## A.6 Aim of this work

Efficient replication of HIV-1 is based on the temporally regulated expression of viral genes. Tightly regulated gene expression is caused by a shift in the splicing pattern of the common HIV-1 pre-mRNA resulting in mRNA isoforms with increasing intron content. To achieve the chronological order of viral mRNA isoform expression the cellular splicing machinery is exploited during early HIV-1 gene expression generating extensively spliced viral mRNAs, whereas in the late phase the splicing machinery is bypassed by the viral pre-mRNA resulting in unspliced genomic RNA. To improve the knowledge on the temporal regulation of HIV-1 gene expression it is a long-term goal to identify the underlying *cis*-regulatory elements modulating alternative splice site usage in the HIV-1 pre-mRNA from early to late gene expression. In the thesis presented here one example for splicing regulation was analysed for early and late pre-RNA processing, respectively.

For the investigation of splicing regulatory events in the early phase of viral gene expression the work focused on the inclusion of the alternatively spliced internal exons 4c, 4a, 4b, and 5 into the 1.8 kb and 4 kb mRNA isoforms. Previously, a purine-rich exonic splicing enhancer in the centrally located exon 5 of the HIV-1 pre-mRNA was identified by our group (295). This work aims to determine whether the GAR ESE activates the upstream located 3'ss as well as the downstream located 5'ss in the presence of both flanking splice sites. In addition, the mechanism underlying splice site activation by the GAR ESE and the contribution of the neighbouring exonic E42 sequence to internal exon recognition is going to be elucidated.

The second part of this thesis concentrated on the basis of disuse of 5' ss D1 in the late phase of viral gene expression leading to the generation of genomic HIV-1 mRNA. It will be analysed whether splicing regulatory elements are located in the proximity of the major 5' ss D1. The contribution of the viral structural protein Gag to D1 efficiency will be evaluated. Finally, it will be investigated whether the p17-inhibitory element located downstream of D1 is involved in the regulation of splicing at this 5' ss and therefore participates in the expression of genomic RNA.

## B. Materials and Methods

### B.1 Materials

#### B.1.1 Chemicals, culture media and solvents

Unless otherwise noted, chemicals were supplied by Invitrogen, Merck, Riedel-de-Haen, Roth, Sigma and Serva. The formation of media and solvents is denoted in the respective experimental protocols or emanates from standard laboratory manuals (22, 512).

##### Cloning

Ampicillin	Roche	Cat.-No. 10835242001
BigDye® Terminator v1.1 Cycle Seq. Kit	Applied Biosystems	Cat.-No. 4336772
FastPlasmid™ Mini	Eppendorf	Cat. No. 0032007655
LB-Broth (Lennox)	Roth	Cat.-No. X964.2
LB Agar (Lennox L Agar)	Invitrogen	Cat.-No. 22700-025
Plasmid Midi Kit	Qiagen	Cat.-No. 12145
Plasmid Maxi Kit	Qiagen	Cat.-No. 12163
Pwo DNA Polymerase	Roche	Cat.-No. 1644955
T4-DNA Ligase	NEB	Cat.-No. M0202S

##### Gel electrophoresis

1 kb DNA Ladder	Invitrogen	Cat.-No. 15615-024
Amberlite IRN-120L	Pharmacia plusone	Cat.-No. 17-1326-01
LE Agarose	Biozym	Cat.-No. 840004
MetaPhor® Agarose	Biozym	Cat.-No. 50180E
PageRuler™ Prestained Protein Ladder Plus	Fermentas	Cat.-No. #SM1811
Prestained SDS-PAGE Standard, Low Range	Bio-Rad	Cat.-No. 161-0305
PS microtest plate, 72-well, with lid	Greiner	Cat.-No. 654102
Rotiphorese Gel 30 (37.5:1 acrylamide/bisacrylamide)	Roth	Cat.-No. 3029.1
Sieve GP Agarose (Genetic pure, low melting)	Biozym	Cat.-No. 850050

##### Protein purification

Ni-NTA agarose	Invitrogen	Cat.-No. R901-01
Lysozym	AppliChem	Cat.-No. A4972001
DNase I recombinant, RNase-free	Roche	Cat.-No. 04716728001
PMSF	Sigma	Cat.-No. P7626
Dialysis membrane tubing (6-8 kDa cut off)	Spectrapor	Cat.-No. 132655

## Materials and Methods

---

### Eukaryotic cell culture

DMEM	GIBCO	Cat.-No. 41966
DPBS –CaCl <sub>2</sub> –MgCl <sub>2</sub>	GIBCO	Cat.-No. 14190
FCS	GIBCO	Cat.-No. 10270-106
PenStrep (10.000 U/mL penicillin, 10.000 µg/mL streptomycin)	GIBCO	Cat.-No. 15140
0.05% Trypsin-EDTA	GIBCO	Cat.-No. 25300
Geneticin G-418 Sulphate	GIBCO	Cat.-No. 11811-031
Trypan Blue Stain 0.4%	GIBCO	Cat.-No. 15250-061

### Transfection

FuGENE® 6 Transfection Reagent	Roche	Cat.-No. 1815075
Lipofectamine™ 2000	Invitrogen	Cat.-No. 11668500

### Immunoblotting

CDP-Star	Roche	Cat.-No. 1685627
ECL™ Western Blotting Detection Reagents	Amersham	Cat.-No. RPN2106
Hyperfilm™ ECL	Amersham	Cat.-No. RPN3103K
Immobilon-P	Millipore	Cat.-No. IPVH00010
Luciferase Assay System	Promega	Cat.-No. E1501
Protein Assay	Bio-Rad	Cat.-No. 500-0006
Re-Blot Plus (Strong Antibody Stripping Solution)	Chemicon	Cat.-No. 2504

### RT-PCR

AmpliTaq® DNA Polymerase	Applied Biosystems	Cat.-No. N0808-0166
Aqua steril	Delta Select	Cat.-No. EPM062F605
DNase I recombinant, RNase-free	Roche	Cat.-No. 04716728001
GeneAmp® dNTPs (2.5 mM each)	Applied Biosystems	Cat.-No. N0808-0007
Primer p(dT) <sub>15</sub> [Oligo-(dT)]	Roche	Cat.-No. 10814270001
Platinum® Taq DNA Polymerase	Invitrogen	Cat.-No. 10966-018
Recombinant RNasin® Ribonuclease Inhibitor	Promega	Cat.-No. #N2511
SuperScript™ III Reverse Transcriptase	Invitrogen	Cat.-No. 18080-085
SuperScript™ III One-Step RT-PCR system	Invitrogen	Cat.-No. 12574-026

### RNA affinity chromatography

Calf Intestinal Alkaline Phosphatase	Promega	Cat.-No. M1821
Adipic acid dihydrazide-Agarose	Sigma	Cat.-No. A0802
Nuclear extracts of 5x10 <sup>9</sup> HeLa cells	Cilbiotech S.A	Cat.-No. CC-01-20-25

Phosphatase inhibitor Cocktail 1	Sigma	Cat.-No. P2850
Phosphatase inhibitor Cocktail 2	Sigma	Cat.-No. P5726
$\beta$ -Glycerophosphate disodium salt hydrate (dissolved in ddH <sub>2</sub> O at 0.5M, stored at -20°C)	Sigma	Cat.-No. G5422
ProteinLoBind tubes (1.5 mL)	Eppendorf	Cat.-No. 0030108.116
<u>Two-dimensional protein separation</u>		
Immobiline DryStrip pH 3-10, 7 cm	Amersham	Cat.-No. 17-6001-11
Immobiline DryStrip pH 3-11NL, 13 cm	Amersham	Cat.-No. 17-6003-75
<u>Sample preparation for mass spectrometry</u>		
Formic acid (1 M in H <sub>2</sub> O)	Fluka	Cat.-No. 06473
ProteinLoBind tubes (0.5 mL)	Eppendorf	Cat.-No. 0030108.094
Trypsin (proteomics grade)	Sigma	Cat.-No. T6567
ZipTip <sub>C18</sub> (standard bed)	Millipore	Cat.-No. ZTC 18S 008
<u>Lab-on-a-chip analyses</u>		
Protein 230 Kit	Agilent	Cat.-No. 5067-1517
RNA 6000 Nano Kit	Agilent	Cat.-No. 5067-1511

## B.1.2 Enzymes

Restriction enzymes were obtained from New England Biolabs (NEB), Fermentas (MBI), Invitrogen, Promega and Roche. They were used as specified by the manufacturer using the buffers provided. T7-RNA polymerase was expressed and purified in house (see B.2.2).

## B.1.3 Cells

### B.1.3.1 Prokaryotic cells

Recombinant plasmids were cloned in *Escherichia coli* (*E. coli*) strain DH5 $\alpha$ F'IQ (Invitrogen). Amplification of non-methylated plasmid DNA for cloning and restriction analyses employing methylation-sensitive restriction endonucleases were performed in *E. coli* strain DM1 (GIBCO), which lacks functional methylases. Recombinant T7-RNA polymerase was expressed and purified from the *E. coli* strain BL21(DE3) (Invitrogen) tolerating high concentrations of recombinant proteins.

Genotypes of *E. coli* strains used:

**DH5 $\alpha$ F'IQ:** F $\phi$ 80/lacZ $\Delta$ M15  $\Delta$ (lacZYA-argF) U169 *recA1 endA1 hsdR17* (*r<sub>k</sub><sup>-</sup>, m<sub>k</sub><sup>+</sup>*) *phoA supE44*  $\lambda^-$  *thi1 gyrA96 relA1/F' proAB+ lacIqZ $\Delta$ M15 zzz::Tn5 [KmR]*

**DM1:** F $\phi$  *dam<sup>-</sup>13::Tn9(Cm<sup>r</sup>) dcm mcrB hsdR<sup>-</sup>M<sup>+</sup> gal1 gal2 ara lac thr leu ton<sup>r</sup> tsx<sup>r</sup> Su<sup>o</sup>  $\lambda^-$*

**BL21(DE3):** *E. coli B, F<sup>-</sup>, dcm, ompT, hsdS<sub>B</sub> (r<sub>B</sub>m<sub>B</sub>), gal lon ( $\lambda$ DE3)*

### B.1.3.2 Eukaryotic cells

The human cervix carcinoma cell line HeLa and derivative cell lines HeLa-T4<sup>+</sup> and HeLa-T4R5 were used for transfection experiments. HeLa-T4<sup>+</sup> cells express the membrane bound, human CD4 receptor due to a stably integrated gene copy under control of the Moloney Murine Sarcoma Virus LTR promoter (379). Transgenic cells were positively selected by application of geneticin. HeLa-T4R5 cells contain stably integrated gene copies of the human CD4 and the CCR5 receptor (kindly provided by Dr. M. Schreiber, Bernhard-Nocht-Institut für Tropenmedizin, Hamburg) and were positively selected by supplementing geneticin and puromycin.

### B.1.4 Oligonucleotides

Oligonucleotides were initially synthesized in house using an automatic DNA-Synthesizer (Model 381 A, Applied Biosystems) by elimination of the terminal trityl group. Oligonucleotides were disconnected from the column substrate with NH<sub>4</sub>OH (3 x 30 min) and were incubated at 55°C ON to remove protective groups. Oligonucleotides were purified by size exclusion chromatography using NAP25 columns (Pharmacia) and eluted with TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA). DNA concentration and purity of collected fractions was determined by photometry at 260 nm and 280 nm (Ultrospec II-4050, Pharmacia). Fractions with the highest DNA concentration were pooled and the final DNA concentration determined.

Oligonucleotides numbered higher than #2289 and Cy5-labeled oligonucleotides for RT-PCR analysis using automated laser fluorescence (A.L.F) were synthesised by Metabion International AG (Martinsried) and purified by desalting. Lyophilised oligonucleotides were reconstituted with sterile H<sub>2</sub>O (Delta Select) to a stock concentration of 100  $\mu$ M. Reconstituted oligonucleotides were stored at -20°C.

#### B.1.4.1 Cloning primer

Endonuclease restriction sites are underlined:

#136(as): 5'- GTCTAGGATCTACTGGAGGCCTTTCTTGCTCTCCTCT

#640(as): 5'- CAATACTACTTCTTGTGGGTTGG

#865(as): 5'- TCTGGTACCTTCGAAGCGCCACTGCTAGAGATT  
#1493(s): 5'- ATCAGGCCTGAATTCGCGGTCGACATCGATAGAATGAGACGA  
#1494(as): 5'- ATCCTGCAGTCAGCAGTTCTTGAA  
#1509(as): 5'- ATCAGATCTTCCCTAGGAAGTTAGCCTGTCTCT  
#1510(as): 5'- ATCGGATCCTTACAAAACCTTTGCCTT  
#1511(as): 5'- ATCGGATCCTTAGTAATTTTGGCTGAC  
#1512(s): 5'- ATCGCGCGCACGGCAAGA  
#1513(s): 5'- ATCACCGGTTCTATAAAA  
#1544(s): 5'- CTTGAAAGCGAAAGTAAAGC  
#1690(as): 5'- ATCGAGCTCTTAAGCGAAGTCGCCGCTAATCCTGCCGTAGGAG  
#1721(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGCTTGATCTGGTGAGTACGCCAAAAATTTTGA  
#1722(as): 5'- ATCGAATTCAGGCCTCTCTCCTTCTAGCCTCCGCTAGTCAAATTTTGGCG  
#1737(s): 5'- ATCTCATGAAATGCAACCTATAAT  
#1749(s): 5'- ATCGCGCGCTCGGCAAGAGGCGAGGGGCTTGATCTGGTGAGTACGCCAAAAATTTTGA  
#1750(s): 5'- ATCGCGCGCTCTGCAAGAGGCGAGGGGCTTGATCTGGTGAGTACGCCAAAAATTTTGA  
#1779(as): 5'- ATCGAATTCAGGCCTCTCTCCTTCTAGCCTCCGCTAG  
#1807(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGCTTGGACTGGTGAGTACGCCAAAAATTTTGA  
#1808(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGCGGCAACTGGTGAGTACGCCAAAAATTTTGA  
#1809(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGAGGCGGCTGGTGAGTACGCCAAAAATTTTGA  
#1831(as): 5'- AGAGCTAGCTATCTGTTTTAAAG  
#1832(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGCGATAACTGGTGAGTACGCCAAAAATTTTGA  
#1833(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGCTTGGGCTGGTGAGTACGCCAAAAATTTTGA  
#1842(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGCGACATCTGGTGAGTACGCCAAAAATTTTGA  
#1861(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGATTACGCCAAAAATTTTGA  
#1864(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGCGATAACTGGTGAGTACGACAAAAATTTTGACTAGC  
GGAGGCTAGA  
#1867(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGGGGCTTGATCTGGTGAGTACCCGCAAATTTTACTAGC  
GGAGGCTAGA  
#1991(as): 5'- ATCGAGCTCTTCGTGCTGTCTCCGCTTCTTCCCGCCGTAGGAGATCCC  
#2031(s): 5'- GGTGCGAGAGCGTCA  
#2032(as): 5'- ATCGTCGACGTAATTTTGGCTGAC  
#2176(s): 5'- ATCGCGCGCTTAAGGAGTTTATATGAAACCCTTACGAATTCTTGACTAAGGAGTTTATATGGAA  
ACCCTTACGGGCTTGATCTGGTGAGTACGCCA  
#2177(s): 5'- ATCGCGCGCTTAAGGAGTTTATATGAAACCCTTACGAATTCTTGACTAAGGAGTTTATATGGAA  
ACCCTTACGGGCGGCGACTGGTGAGTACGCCA  
#2185(as): 5'- ATCGTCGACCGCGAATTCAGGCCTCTCTCCTTCTAGCCTCCGCTAGTCAAATTTTGGCGT  
ACTCACCAG  
#2249(s): 5'- GGC GCCAGGGCCAGC  
#2250(as): 5'- ATCGTCGACGTAGTTCTGGCTCAC  
#2289(s): 5'- ATCGCGCGCACGGCAAGAGGCGAGTGCCGGCGACTGGTGAGTACGCCAAAAATTTTGA  
#2295(s): 5'- AAGCTGAAGCACATCGTATGGGCAAGCAGG  
#2297(s): 5'- AGCTGCAGCCAGCCTTACAGACAGGATCAG  
#2299(s): 5'- CACCAAGGAGGCCCTAGACAAGATAGAGGA  
#2353(s): 5'- GCCCATGGCTGACTAATTTTTTT  
#2504(as): 5'- CAGCTCGCCGCCGCTTAATACTGACGCTCT  
#2505(as): 5'- CTTGCCGCCGGCCCTTAACCGAATTTTTTC

## Materials and Methods

---

- #2506(as): 5'- CCTGCTGGCCACACTATATGTTTTAATTT  
#2542(as): 5'- ATCGAGCTCTTCGTCGCTGTCTCCGCTTCTTCCCGCCGTAGGAGATCCCTAAGGTTTTTGTGCAT  
GAAACAA  
#2555(s): 5'- ATCGAGCTCATCAGAACAGTCAGACTCATCAAGCTTCTCTATCAAAGAGGTAACATCTACATGTA  
ATGCAAC  
#2556(s): 5'- ATCGAGCTCGCAGTAAGTAGTACA  
#2557(s): 5'- ATCGAGCTCGAGGTAACATCTACATGTAATGCAAC  
#2558(s): 5'- ATCGAGCTCGCAGTAAGACGTACATGTAATGCAAC  
#2559(s): 5'- ATCGAGCTCGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCGCAGTAAGTAGTACA  
#2560(s): 5'- ATCGAGCTCGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCGAGGTAACATCTACATGTA  
ATGCAAC  
#2561(s): 5'- ATCGAGCTCGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCGCAGTAAGACGTACATGTA  
ATGCAAC  
#2889(as): 5'- ATCGAGCTCTTCGTCGCTGTCTCCGCTTCTTCCCTGCCGTAGGAGATCCCTAAGGTTTTTGTGCAT  
GAAACAA  
#2890(as): 5'- ATCGAGCTCTTCGTCGCTGTCTCCGCTTCTTCCCTGCCGTAGGAGATCCCCAAGGCTTTTGTGCAT  
GAAACAA  
#2891(as): 5'- ATCGAGCTCTTCGTCGCTGTCTCCGCTTCTTCCCGCCGTAGGAGATCCCCAAGGCTTTTGTGCAT  
GAAACAA  
#2892(as): 5'- ATCGAGCTCTTCGTCGCTGTCTCCGCTTCTTCCCGCCGTAGGAGATCCCCAAGGTTTTTGTGCAT  
GAAACAA  
#2893(s): 5'- ATCATCGATTGGAAGGGCTAATTT  
#2907(s): 5'- ATCGCGCGCTTATTCATAATGATAGTAGGAGGCCTGGTGAGTACGCCA  
#2908(as): 5'- ATCGAATTCAGGCCTGCCTCCTACTATCATTATGAATAATTTTATATACCACAGTACTCACCAGT  
CGC  
#2909(as): 5'- ATCGAATTCAGGCCTCTCTATTCACTATAGAAAGTACAGCAAAAATTTCTTGTACTCACCAGT  
CGC  
#3104(as): 5'- ATCGTCGACGGGATGGTTGTAGCT  
#3105(as): 5'- ATCGTCGACGGCTGGGCTGCAGCT  
#3141(as): 5'- GATGCGCGCTTCAGCAAGCCGAGCCCCGCGTCGAGAGATCTC

### B.1.4.2 Sequencing primer

- #1158(as): 5'- GCCACTGTCTTCTGCTCTT  
#1513(s): 5'- ATCACCGGTTCTATAAAA  
#1542(as): 5'- CACCTTCTTCTTCTATTCCCTT  
#1543(as): 5'- CCCCCATCTCCACAA  
#1544(s): 5'- CTTGAAAGCGAAAGTAAAGC  
#2353(s): 5'- CCCCCATGGCTGACTAATTTTTTTT

### B.1.4.3 RT-PCR primer

Alternatively spliced RNA (ASR) of minigenes descending from SV leader SD1 SA5 env nef:

- #1544(s): 5'- CTTGAAAGCGAAAGTAAAGC  
#1543(as): 5'- CCCCCATCTCCACAA  
#1542(as): 5'- CACCTTCTTCTTCTATTCCCTT

Alternative primer pair for ASR published by Purcell and Martin (481):

**#2576(as)** [Odp.032]: 5'- CCGCAGATCGTCCCAGATAAG  
**#2577(as)** [Odp.084]: 5'- TCATTGCCACTGTCTTCTGCTCT  
**#2578(s)** [Odp.045]: 5'- CTGAGCCTGGGAGCTCTCTGGC

Total RNA (Total ASR) of the HIV-1 based minigene constructs (within *nef*-coding region):

**#224(s)**: 5'- TGGCCTGCTGTAAGGGAAAG  
**#225(as)**: 5'- CCAGTCCCCCCTTTTCTTTT

*hGH*-mRNA (amplicon covering exons 2-4):

**#1274(s)**: 5'- TTGACACCTACCAGGAGTTTGAAGAAG  
**#1273(as)**: 5'- GATGCGGAGCAGCTCTAGGTTGGATTT

*hGH*-mRNA (amplicon covering exons 3-5):

**#1225(s)**: 5'- CAACAGAAATCCAACCTAGAGCTGCT  
**#1224(as)**: 5'- TCTTCCAGCCTCCCATCAGCGTTTGG

Primer for generation of Cy5-labeled size marker for A.L.F:

**#1544-Cy5(s)**: 5'- CTTGAAAGCGAAAGTAAAGC  
**#2009(as)**: 5'- CGTACTCACCAGTCG (amplicon size: 100 bp)  
**#2010(as)**: 5'- CTTCTAGCCTCCGCT (amplicon size: 130 bp)  
**#2011(as)**: 5'- CGAATTCAGGCCTCT (amplicon size: 150 bp)  
**#2012(as)**: 5'- CCTTTCTTGCTCTCC (amplicon size: 200 bp)  
**#2013(as)**: 5'- AGTCTAGGATCTACT (amplicon size: 220 bp)  
**#2014(as)**: 5'- GGCTGACTTCCTGGA (amplicon size: 250 bp)  
**#2015(as)**: 5'- GGCAATGAAAGCAAC (amplicon size: 300 bp)  
**#2016(as)**: 5'- CTTCTTCCTGCCGTA (amplicon size: 350 bp)  
**#2017(as)**: 5'- GAAGCTTGATGAGTC (amplicon size: 400 bp)  
**#2019(as)**: 5'- TATGGACCACACAAC (amplicon size: 500 bp)  
**#2020(as)**: 5'- CTTCACTCTCATTGC (amplicon size: 600 bp)

Primer for generation of Cy5-labeled RT-PCR products for A.L.F:

**#224-Cy5(s)** [total ASR]: 5'- Cy5-TGGCCTGCTGTAAGGGAAAG  
**#1225-Cy5(s)** [hGH]: 5'- Cy5-CAACAGAAATCCAACCTAGAGCTGCT  
**#1544-Cy5(s)** [ASR]: 5'- Cy5-CTTGAAAGCGAAAGTAAAGC

Primer pairs for PCR amplification of *hnRNP H*- and *GAPDH*-cDNA were kindly provided by Prof. Dr. M. Caputi (Florida Atlantic University, USA).

*hnRNP H*

**MC20.54(s)**: 5'- CAGAGAAGGCAGACCAAGTG  
**MC20.55(as)**: 5'- TCCAACCCTGAGAAGAAGT

GAPDH

#3153: 5'- ACCACAGTCCATGCCATCAC

#3154: 5'- TCCACCACCCTGTTGCTGTA

**B.1.4.4 shRNA**

shRNAs targeting hnRNP proteins were kindly provided by Prof. Dr. M. Caputi (Florida Atlantic University, USA).

Oligonucleotide for *in vitro* transcription of shRNA targeting hnRNP H (antisense T7-promoter is doubly underlined):

**MC66.20:**

5'- AAAGGATTGGGTGTTGAAGCATACTTCGGTATGCTTCAACACCCAATCCTATAGTGAGTCGTATTA

**B.1.4.5 *In vitro* transcription primer**

*B.1.4.5.1 Annealing-Templates*

Sequences of 5'ss D1 and D4, respectively, are underlined. The T7-promoter sequence complementary to the T7sense oligo (#2058) is doubly underlined.

**#2058/#2503/#3070** (T7-promoter sense):

5'- TAATACGACTCACTATAGG

GAR ESE with downstream 5'ss D4

**#1916**(GAR ESE):

5'- TGTACTACTTACTGCGAGCTCTTCGTCGCTGTCTCCGCTTCTTCCTGCCTATAGTGAGTCGTATTA

**#1917**(HIV#18):

5'- TGTACTACTTACTGCGAGGCCTCCTACTATCATTATGAATAATTTGCCTATAGTGAGTCGTATTA

**#1918**(SRp40-):

5'- TGTACTACTTACTGCGAGCTCTTCGTCGCTAAATCCGCTTCTTCCTGCCTATAGTGAGTCGTATTA

**#1919**(SF2-):

5'- TGTACTACTTACTGCGAGCTCTTAAGCGCTGTCTCCGCTAATTCCTGCCTATAGTGAGTCGTATTA

**#1920**(SF2-, SRp40-):

5'- TGTACTACTTACTGCGAGCTCTTAAGCGAAGTCGCCGCTAATTCCTGCCTATAGTGAGTCGTATTA

GAR ESE with upstream intronic sequence

**#2581**(GAR ESE and 34 nt intron upstream):

5'- GAGCTCTTCGTCGCTGTCTCCGCTTCTTCCTGCCGTAGGAGATCCCTAAGGCTTTTGTGCATGACCTATAGTGA  
GTCGTATTA

**#2582**(GAR ESE and 23 nt intron upstream; extended)

5'- GAGCTCTTCGTCGCTGTCTCCGCTTCTTCCTGCCGTAGGAGATCCCTAAGGCCCTATAGTGAGTCGTATTA

**#2583**(GAR ESE and 17 nt intron upstream; short):

5'- GAGCTCTTCGTCGCTGTCTCCGCTTCTTCCTGCCGTAGGAGATCCCCCTATAGTGAGTCGTATTA

**#2596**(GAR ESE-; extended):

5'- GAGCTCTTAAGCGCTAAATCCGCTAATTCCTGCCGTAGGAGATCCCTAAGGCCCTATAGTGAGTCGTATTA

**#2597**(GAR ESE-; short):

5'- GAGCTCTTAAGCGCTAAATCCGCTAATTCCTGCCGTAGGAGATCCCCCTATAGTGAGTCGTATTA

**#2599**(2xSF2-; extended):

5'- GAGCTCTTAAGCGCTGTCTCCGCTAATTCCTGCCGTAGGAGATCCCTAAGGCCCTATAGTGAGTCGTATTA

**#2600**(2xSF2-; short):

5'- GAGCTCTTAAGCGCTGTCTCCGCTAATTCCTGCCGTAGGAGATCCCCCTATAGTGAGTCGTATTA

#### D1 with upstream exonic sequence

**#2247**(D1 and 29 nt upstream exon):

5'- TTGGCGTACTCACCAGTCGCCGCCCTCGCCTCTTGCCGTGCGCGCCTATAGTGAGTCGTATTA

**#2248**(5'SR-):

5'- TTGGCGTACTCACCAGATCAAGCCCCTCGCCTCTTGCCGTGCGCGCCTATAGTGAGTCGTATTA

**#2392**(-12T-10C):

5'- TTGGCGTACTCACCAGTCGCCGGCACTCGCCTCTTGCCGTGCGCGCCTATAGTGAGTCGTATTA

**#2393**(12T-10C and 5'SR-):

5'- TTGGCGTACTCACCAGATCAAGGCACTCGCCTCTTGCCGTGCGCGCCTATAGTGAGTCGTATTA

#### *B.1.4.5.2 Primer for PCR-derived templates*

The NL4-3 exon 5 sequence and mutants thereof were amplified including 17 nt or 23 nt intron and 12 nt downstream exon using different primer combinations.

**#2825**(s) [short]:

5'- TAATACGACTCACTATAGGGGATCTCCTACGGC

**#2826**(s) [extended]:

5'- TAATACGACTCACTATAGGGCCTTAGGGATCTCC

**#2827**(s) [extended, A4b AG-]:

5'- TAATACGACTCACTATAGGGCCTTGGGGATCTCCTACGGCA

**#2828**(as) [full complementary 5'ss]:

5'- TGTAATACTTACCTGTTTGATAGAGAAGCT

**#2829**(as) [D4]:

5'- TGTACTACTTACTGC

**#2830**(as) [cs +1<sup>12</sup>]:

5'- TGTACGTCTTACTGCTTTGATAGAGAA

**#2831**(as) [3U]:

5'- TGTACTACTAACTGCTTTGATAGAG

**#2980**(as) [HIV#18 leader SD4]:

5'- TGTACTACTTACTGCGCCTCCTACTATCAT

### B.1.5 Recombinant plasmids

Sequences of all generated constructs were confirmed by DNA sequencing of substituted subgenomic or PCR-amplified fragments.

#### B.1.5.1 HIV-1 derived minigene plasmids

The 2-intron minigene SV leader SD1 SA5 env nef contains the 5'-region of the proviral genome starting from the repeat region (R) till the Gag translational start codon, which was joined by a short artificial linker sequence with the 3' part of the viral genome starting from the *Sal*I restriction site downstream of 3' ss A3 – thereby omitting the immediately flanking ESS – until the stop codon of the nef ORF. Translational start codons for Gag and the regulatory proteins Tat and Rev were inactivated by point mutations.

For construction of the 2-intron minigene **SV leader SD1 SA5 env nef** the *Sal*I/*Ce*II-fragment of ZK SV SD1 PSI env nef containing a restriction site linker was substituted with the *Sal*I/*Ce*II-fragment of SV E/X tat<sup>rev</sup> (295) introducing the *env*-coding region deviating from pNL4-3 (5). **ZK SV SD1 PSI env nef** was cloned by linking the SV40 pA site, the SV40 promoter and the pNL4-3 sequence of the *Pst*I/*Stu*I-fragment of SV1.4E-PSI-tat(-)rev(-)envCAT with a *Stu*I/*Pst*I-restricted PCR product introducing unique *Eco*R I, *Sal*I and *Cla*I restriction sites upstream of the *nef*-coding region by using the 5' primer #1493 and the 3' primer #1494 with pNLA1 (591), a derivative of pNL4-3 (5), as template.

#### Plasmid containing the HIV-1 LTR promoter

The plasmid **LTR leader SD1 SA5 env nef** was generated by substituting the 349 bp *Cla*I/*Afl*II-fragment of SV leader SD1 SA5 env nef containing the SV40 early promoter with the 521 bp *Cla*I/*Afl*II-restricted fragment of a PCR product generated by using primer pair #2893/#865 with pLTR636RL as template (kindly provided by Dr. J. Krummheuer). Positive clones were identified by restriction analysis revealing an additional *Ava*I restriction site, loss of an *Nco*I restriction site and an increased fragment size.

#### Plasmid carrying mutations in the hnRNP H binding site upstream of D1

**SV leader -12T-10C SD1 SA5 env nef** was mutated at positions -12 from G to T and -10 from G to C relative to the exon1/intron border thereby changing the consensus motif (AGGGG) for hnRNP H binding to AGtGc. The potential 5' ss generated by the -12G>T mutation was expected not to be activated due to insufficient U1 snRNA complementarity (HBond Score: 4.50 [ver. 4.1]). Since the resulting plasmid is unaltered regarding size and endonuclease restriction sites compared to SV leader SD1 SA5 env nef, the intermediate plasmid **ZK SV leader SD1mut SA5 env nef** was cloned by substituting the 85 bp *Bss*HII/*Eco*R I-fragment with the respective 387 bp fragment from SV1.3ΔSD3 ctat crev. Positive clones were identified by an additional

restriction site for each of the endonucleases *Nco* I and *Afl* II. To generate SV leader -12T-10C SD1 SA5 env nef a PCR amplicon was generated using the mutagenic sense primer #2289 and antisense primer #1722 without template. The PCR product was digested using endonucleases *Bss*H II and *Eco*R I and substituted for the 387 bp fragment of ZK SV leader SD1mut SA5 env nef. Positive clones were identified by loss of the *Nco* I restriction site of the intermediate plasmid and by size decrease compared to the parental construct.

#### Plasmids containing mutations in 5' ss D1

**SV leader 5U SA5 env nef** was cloned by substituting the 387 bp *Bss*H II/*Eco*RI-fragment of ZK SV leader SD1mut SA5 env nef with the 85 bp *Bss*H II/*Eco*R I-restricted PCR amplicon generated by using primer pair #1861/#1722 without template. Positive clones were identified by the decreased fragment size compared to the parental plasmid ZK SV leader SD1mut SA5 env nef.

**SV leader 5U ins(IIIB) SA5 env nef** and **SV leader 5U syn-ins(IIIB) SA5 env nef** were generated by substituting the 94 bp *Bss*H II/*Sa*I-fragment of SV leader SD1 SA5 env nef with the 475 bp *Bss*H II/*Sa*I-fragment of SV leader SD1 SA5 env nef with the 475 bp *Bss*H II/*Sa*I-fragment of SV leader SD1 SA5 env nef derived from using primer pair #1861/#136 with templates SV leader SD1 ins(IIIB) SA5 env nef and SV leader SD1 syn-ins(IIIB) SA5 env nef, respectively. Positive clones were selected by loss of an *Eco*R I restriction site in an *Eco*R I/*Sa*c I-restriction analysis.

**SV leader 5U ins(202) SA5 env nef** and **SV leader 5U syn-ins(202) SA5 env nef** were generated by substituting the 396 bp *Bss*H II/*Sa*I-fragment of ZK SV leader SD1mut SA5 env nef with the 281 bp *Bss*H II/*Sa*I-fragment of the PCR product generated with sense primer #1544 and either antisense primer #3104 using SV leader 5U ins(IIIB) SA5 env nef as template or antisense primer #3105 using SV leader 5U syn-ins(IIIB) SA5 env nef as template, respectively. Positive clones were identified by decreased fragment size and loss of the *Nco* I restriction site in the parental construct.

#### Plasmids containing substitutions in the upstream and downstream regions of 5' ss D1

**SV leader HIV#18 SD1 SA5 env nef**, **SV leader SD1 HIV#18up SA5 env nef**, and **SV leader SD1 HIV#18down env nef** were cloned by substituting the 387 bp *Bss*H II/*Eco*R I-fragment of the intermediate plasmid ZK SV leader SD1mut SA5 env nef with 85 bp *Bss*H II/*Eco*R I-digested PCR products amplified with SV leader SD1 SA5 env nef as template by using primer pair #2907/#136 for SV leader HIV#18 SD1 SA5 env nef, primer pair #1544/#2908 for SV leader SD1 HIV#18up SA5 env nef and primer pair #1544/#2909 for SV leader SD1 HIV#18down SA5 env nef. Positive clones were identified by loss of the *Nco* I restriction site of the intermediate plasmid and decreased fragment size.

### Plasmids containing bacteriophage-derived stem loop sequences

Two consecutive stem loops of the bacteriophage PP7 were introduced upstream of D1 by substituting the 94 bp *BssH II/SaI* I-fragment with the 143 bp *BssH II/SaI* I-fragment of a PCR product amplified without template encoding both stem loops in the presence of parental (primer pair #2177/#2185) or mutated SR protein binding sites at D1 (5'SR-) (primer pair #2176/#2185). PCR-derived fragments were inserted into SV leader SD1 SA5 env nef generating **SV leader 2xPP7 SD1 SA5 env nef** (#2177/#2185) and **SV leader 2xPP7 5'SR- SD1 SA5 env nef** (#2176/#2185). Replacing the *BssH II/SaI* I-fragment in the plasmid SV leader SD1 SA5 2xMS2 SD4 env nef yielded constructs **SV leader 2xPP7 SD1 SA5 2xMS2 SD4 env nef** (#2177/#2185) and **SV leader 2xPP7 5'SR- SD1 SA5 2xMS2 SD4 env nef** (#2176/#2185). Positive clones were identified by increased fragment size compared to the parental plasmids in an *EcoR I/Nco I* restriction analysis.

### Plasmids containing mutations in potential SR protein binding sites overlapping 5' ss D1

Mutations were introduced into predicted SR protein binding sites overlapping D1 by replacing the 387 bp *BssH II/EcoR I*-fragment of ZK SV leader SD1mut SA5 env nef with the respective 85 bp *BssH II/EcoR I*-fragment of PCR products generated without templates using primer pair #1807/#1722 for **SV leader SC35(1)- SD1 SA5 env nef**, #1808/#1722 for **SV leader SC35(2)- SD1 SA5 env nef**, #1809/#1722 for **SV leader SRp40- SD1 SA5 env nef**, #1832/#1722 for **SV leader SC35(1)- SC35(2)- SD1 SA5 env nef**, #1833/#1722 for **SV leader SC35(1)- SRp40- SD1 SA5 env nef**, #1842/#1722 for **SV leader SRp40- SC35(2)- SD1 SA5 env nef**, #1721/#1722 for **SV leader 5'SR- SD1 SA5 env nef**, #1864/#1779 for **SV leader SC35(1)- SRp40- SC35(2)- SRp55(2)- SC35(3)- SD1 SA5 env nef** and #1867/#1779 for **SV leader SC35(1)- SC35(2)- SC35(3)- SD1 SA5 env nef**. Positive clones were selected by loss of a restriction site for each of the endonucleases, *Nco I* and *Afl II*, and a decreased fragment size compared to the parental plasmid.

### Plasmids containing mutations of exonic SR protein binding sites far upstream of D1

SR protein binding sites predicted upstream of D1 were mutated by substituting the 387 bp *BssH II/EcoR I*-fragment of ZK SV leader SD1mut SA5 env nef with the 85 bp *BssH II/EcoR I*-fragment of the PCR amplicon derived with primer pair #1749/#1722 generating **SV leader (-28)SRp55- SD1 5'SR- SA5 env nef** or primer pair #1750/#1722 producing **SV leader (-30)SF2- (-28)SRp55- (-25)SF2- SD1 5'SR- SA5 env nef**. Positive clones were identified by loss of *Nco I* and *Afl II* restriction sites and by a decreased fragment size.

---

Plasmids containing regions of the p17-inhibitory element

Plasmids **SV leader SD1 ins(IIIB) SA5 env nef** and **SV leader SD1 syn-ins(IIIB) SA5 env nef** were cloned from pc wtgag(IIIB) and pc syngag(IIIB), respectively (generously provided by Prof. Dr. R. Wagner, Universität Regensburg) (216). The 13 bp *Stu* I/*Sal* I-fragment of the parental construct SV leader SD1 SA5 env nef was replaced by *Sal* I-digested 394 bp PCR products amplified using primer pair #2031/#2032 with pc wtgag(IIIB) as template for **SV leader SD1 ins(IIIB) SA5 env nef** and primer pair #2249/#2250 with pc syngag(IIIB) as template for **SV leader SD1 syn-ins(IIIB) SA5 env nef**. Positive clones were identified by loss of a *Stu* I restriction site and increased fragment size in a *Stu* I/*Afl* II restriction analysis.

Hybrid syn-ins-/ins-sequences were generated by overlap PCR. The first PCR was performed using a symmetrical 30 bp sense primer introducing the respective syn-ins-/ins-junction by containing 15 nt complementary to the syn-ins- and 15 nt complementary to the ins-sequence and the common antisense primer #136 with SV leader SD1 ins(IIIB) SA5 env nef as template. In a subsequent PCR reaction, PCR products from the first amplification were used as antisense primer in combination with the common sense primer #1544 using SV leader SD1 syn-ins(IIIB) SA5 env nef as template. Sense primers #2295, #2297 and #2299 were used in the first PCR to generate **SV leader SD1 syn-ins102(IIIB) SA5 env nef**, **SV leader SD1 syn-ins202(IIIB) SA5 env nef** and **SV leader SD1 syn-ins302(IIIB) SA5 env nef**, respectively. The 475 bp *BssH* II/*Sal* I-digested PCR products were substituted for the 94 bp *BssH* II/*Sal* I-fragment of SV leader SD1 SA5 env nef. Positive clones were identified in a *Hind* III/*Apa* L I restriction analyses by the gain of an additional *Hind* III restriction site and an increased plasmid size compared to the parental construct.

Applying a similar cloning strategy, hybrid ins-/syn-ins-sequences were generated by overlap PCRs. The first PCR was performed using the common sense primer #1544 and a symmetrical 30 bp antisense primer introducing the respective ins-/syn-ins junction by providing 15 bp complementary to each sequence, ins and syn-ins, with SV leader SD1 ins(IIIB) SA5 env nef as template. In the subsequent second PCR, products from the first PCR were used as primers in combination with the common antisense primer #136 with SV leader SD1 syn-ins(IIIB) SA5 env nef as template. In the first PCR reaction antisense primers #2504, #2505 and #2506, respectively, were used for generation of **SV leader SD1 ins(24)syn-ins SA5 env nef**, **SV leader SD1 ins(63)syn-ins SA5 env nef** and **SV leader SD1 ins(102)syn-ins SA5 env nef**. PCR products were digested with *BssH* II and *Sal* I and the resulting 475 bp fragments substituted for the 94 bp *BssH* II/*Sal* I-fragment of SV leader SD1 SA5 env nef. Positive clones were selected by loss of the *EcoR* I restriction site and increased fragment size compared to the parental construct in an *EcoR* I/*Sac* I restriction analysis.

To shorten the hybrid ins-/syn-ins-sequences to 202 nt, the 396 bp *BssH* II/*Sal* I-fragment of ZK SV leader SD1mut SA5 env nef was substituted with a 281 bp *BssH* II/*Sal* I-digested PCR

product generated using sense primer #1544 and antisense primer #3104 with SV leader SD1 ins(IIIB) SA5 env nef as template for **SV leader SD1 ins(202) SA5 env nef**, antisense primer #3105 with SV leader SD1 ins(102)syn-ins SA5 env nef as template for **SV leader SD1 ins(102)syn-ins(202) SA5 env nef**, antisense primer #3105 with SV leader SD1 ins(63)syn-ins SA5 env nef as template for **SV leader SD1 ins(63)syn-ins(202) SA5 env nef**, antisense primer #3105 with SV leader SD1 ins(24)syn-ins SA5 env nef as template for **SV leader SD1 ins(24)-syn-ins(202) SA5 env nef** and antisense primer #3105 with SV leader SD1 syn-ins(IIIB) SA5 env nef for **SV leader SD1 syn-ins(202) SA5 env nef**. Positive clones were identified by loss of an *EcoR* I restriction site and a decreased plasmid size compared to the parental construct in an *EcoR* I/*Sac* I restriction analysis.

### Plasmids carrying mutation in the SL1 loop

The preliminary plasmid **ZK SV mut-ups BssHII leader SD1 SA5 env nef** was cloned by substituting the 307 bp *Nco* I/*BssH* II-fragment of the 2-intron minigene SV leader SD1 SA5 env nef with a 684 bp fragment originating from plasmid pT-βbeta-c1-23,8 (kindly provided by Dipl. Biol. L. Hartmann) thereby inserting an additional *Pst* I restriction site.

The SL1 loop A in the leader upstream of D1 was mutated in the plasmid **SV leader GGGG SD1 SA5 env nef** by replacing the 684 bp *Nco* I/*BssH* II-fragment of ZK SV mut-ups BssHII leader SD1 SA5 env nef with a *BssH* II/*Nco* I-restricted PCR insert generated using primers #2353 and #3141 with template SV leader SD1 SA5 env nef.

The mutation of the SL1 loop A was combined with the most 5'-situated 202 nt of the p17-inhibitory element or the codon-optimised sequence in the plasmids **SV leader GGGG SD1 ins(202) SA5 env nef** and **SV leader GGGG SD1 syn-ins(202) SA5 env nef** by substituting the 709 bp *BssH* II/*Nde* I-fragment of SV leader GGGG SD1 SA5 env nef with the 896 bp *BssH* II/*Nde* I-fragment of SV leader SD1 ins(202) SA5 env nef and SV leader SD1 syn-ins(202) SA5 env nef, respectively. Positive clones were identified by increased fragment size in a *Sal* I/*Hind* III restriction analysis.

### Plasmids containing mutations in 5' ss D4

**SV leader SD1 3U env nef** contains a point mutation in the 5'ss D4 at position +3 from A to T. The plasmid was cloned by substituting the 1247 bp *Sac* I/*Nhe* I-fragment of SV leader SD1 SA5 env nef with the corresponding *Sac* I/*Nhe* I-fragment of SV E/X tat<sup>rev</sup> 3U (295) including the point mutation in the 5'ss D4.

Mutant construct **SV leader SD1 SA5 GTV env nef** was generated by substituting the 1247 bp *Sac* I/*Nhe* I-fragment of SV leader SD1 SA5 env nef with the corresponding *Sac* I/*Nhe* I-fragment of SV E/X tat<sup>rev</sup> GT V (295). Mutation of the cryptic 5' ss 13 nt downstream of D4 from GT to GA was achieved by substituting the *BspH* I/*Nde* I-fragment of the parental construct

SV leader SD1 SA5 env nef with the *Bsp*H I/*Nde* I-restricted PCR product amplified with primer pair #1737/#640 using SV leader SD1 SA5 env nef as template producing **SV leader SD1 SA5 GTV 15A env nef**.

The construct **SV leader SD1 SA5 cs+1<sup>12</sup> env nef** was generated by substituting the 1247 bp *Sac* I/*Nhe* I-fragment of SV leader SD1 SA5 env nef containing 5' ss D4 for the respective fragment of SV E/X tat<sup>-</sup> rev<sup>-</sup> cs+1<sup>12</sup>. The plasmid was cloned by I. Meyer. In the plasmid **SV leader SD1 SA5 cs-2<sup>14</sup> env nef** 5' ss D4 was substituted with the mutant 5' ss cs-2<sup>14</sup> by replacing the 397 bp *Sac* I/*Nde* I-fragment of ZK SV leader SD1mut SA5 env nef in a 2-insert ligation applying the additional endonuclease *Bss*H II with an *Sac* I/*Nde* I-restricted PCR product amplified with primer pair #2555/#1831 using SV leader SD1 SA5 env nef as substrate. Substitution of D4 with 5' ss cs-2<sup>14</sup> or cs+1<sup>12</sup>, respectively, was combined with deletion or substitution of the E42 sequence with the HIV#18 control sequence in constructs **SV leader SD1 SA5 H- cs-2<sup>14</sup> env nef**, **SV leader SD1 SA5 H- cs+1<sup>12</sup> env nef**, **SV leader SD1 SA5 HIV#18 cs-2<sup>14</sup> env nef** and **SV leader SD1 SA5 HIV#18 cs+1<sup>12</sup> env nef** by substituting the 397 bp *Sac* I/*Nde* I-fragment of ZK SV leader SD1mut SA5 env nef with PCR products amplified with mutagenic sense primers #2557, #2558, #2560 and #2561, respectively, and the common antisense primer #1831 using SV leader SD1 SA5 env nef as substrate. Constructs were cloned using a 2-insert cloning strategy including the additional endonuclease *Bss*H II. Positive clones were selected by fragment sizes applying an *Eco*R I/*Sac* I restriction analysis.

#### Plasmids containing mutations in 3' ss A4a, A4b or A5

The 3' ss clustering upstream of exon 5 were inactivated by point mutations within the respective AG-dinucleotides. PCR products carrying mutated 3' ss were generated by using the common sense primer #1544 in combination with the antisense primers #1991, #2890, #2889, #2891, #2542 and #2892 for generation of **SV leader SD1 SA5- env nef**, **SV leader SD1 SA4b- SA5 env nef**, **SV leader SD1 SA4a- SA5 env nef**, **SV leader SD1 SA4b- SA5- env nef**, **SV leader SD1 SA4a- SA5- env nef** and **SV leader SD1 SA4a- SA4b- SA5- env nef**, respectively. The 227 bp *Eco*R I/*Sac* I-fragment of each PCR product was substituted for the respective fragment of the parental construct SV leader SD1 SA5 env nef in a 2-insert cloning protocol using the additional endonuclease *Nhe* I. Positive clones were identified in an *Eco*R I/*Sac* I restriction analysis showing the identical fragment pattern as the parental construct.

#### Plasmids containing mutations in potential SR protein binding sites of the GAR ESE (exon 5)

For generation of mutant constructs **SV leader SD1 SA5 SF2(1)- SF2(2)- SRp40- env nef**, **SV leader SD1 SA5 SF2(1)- SRp40- SF2(2)- env nef**, **SV leader SD1 SA5 SRp40- SF2(2)- env nef**, **SV leader SD1 SA5 SF2(1)- SRp40- env nef**, **SV leader SD1 SA5 SF2(1)- SF2(2)- env**

nef, SV leader SD1 SA5 SRp40 15- env nef, SV leader SD1 SA5 SF2(2)- env nef, SV leader SD1 SA5 SF2(1)- env nef, SV leader SD1 SF2- SA5 env nef and SV leader SD1 SF2- SA5 ESE- env nef the 227 bp *EcoR I/Sac I*-fragment of SV leader SD1 SA5 env nef was substituted with *EcoR I/Sac I*-restricted PCR products amplified with the common sense primer #1544 and mutagenic antisense primers #1690, #2598, #2000, #2067, #1903, #1660, #1987, #1988, #2499 and #2501, respectively, using SV leader SD1 SA5 env nef as template. After a 2-insert cloning procedure using the additional endonucleases *Nhe I*, positive clones were identified in an *EcoR I/Sac I*-restriction analyses revealing identical fragment sizes as the parental construct.

#### Combinatorial mutations in the GAR ESE and in the 3' ss cluster upstream of exon 5

The initially used mutation of the GAR ESE [SF2(1)- SF2(2)- SRp40-] was combined with the inactivation of single or multiple 3' ss by substituting the 227 bp *EcoR I/Sac I*-fragment of SV leader SD1 SA5 env nef with the respective fragment of PCR products generated using the common sense primer #1544 in combination with mutagenic antisense primers on templates containing different mutations of the 3' ss (Tab. II-1). Applying the identical cloning strategy, an additional construct series was generated harbouring a triple mutation of the GAR ESE [SF1(1)- SRp40- SF2(2)-] differing in the nucleotide substitutions inactivating the SRp40 binding site. Constructs were cloned using a 2-insert protocol with the additional endonuclease *Nhe I*. Positive clones were identified in an *EcoR I/Sac I* restriction analysis.

**Tab. II-1: Mutagenic antisense primers and PCR templates used for generation of constructs carrying combinatorial mutations in the GAR ESE and in the 3' ss cluster upstream of exon 5.**

Cloned construct	Primer	PCR template
SV leader SD1 SA4b- SA5 SF2(1)- SF2(2)- SRp40- env nef	#1690	SV leader SD1 SA4b- SA5 env nef
SV leader SD1 SA4a- SA5- SF2(1)- SF2(2)- SRp40- env nef	#2543	SV leader SD1 SA5 env nef
SV leader SD1 SA5- SF2(1)- SF2(2)- SRp40- env nef	#2002	SV leader SD1 SA5 env nef
SV leader SD1 SA4a- SA4b- SA5- SF2(1)- SF2(2)- SRp40- env nef	#2002	SV leader SD1 SA4a- SA4b- SA5- env nef
SV leader SD1 SA4b- SA5 SF2(1)- SRp40- SF2(2)- env nef	#2598	SV leader SD1 SA4b- SA5 env nef
SV leader SD1 SA4a- SA5- SF2(1)- SRp40- SF2(2)- env nef	#3071	SV leader SD1 SA4a- SA5- env nef
SV leader SD1 SA5- SF2(1)- SRp40- SF2(2)- env nef	#3071	SV leader SD1 SA5- env nef
SV leader SD1 SA4a- SA4b- SA5- SF2(1)- SRp40- SF2(2)- env nef	#3071	SV leader SD1 SA4a- SA4b- SA5- env nef

Mutations in the E42 sequence

Substitution of the E42 sequence with the HIV#18 control in **SV leader SD1 SA5 ESE HIV#18 env nef** and deletion of the E42 sequence ( $\Delta$ E42) in the mutant construct **SV leader SD1 SA5 H- env nef** was performed by substituting the 397 bp *Sac I/Nde I*-fragment of ZK SV leader SD1mut SA5 env nef with *Sac I/Nde I*-restricted fragments of PCR products amplified with sense primers #2559 or #2556, respectively, and the common antisense primer #1831 using SV leader SD1 SA5 env nef as substrate. Constructs were cloned using a 2-insert cloning strategy including the additional endonuclease *BssH II*. Positive clones were selected by fragment sizes applying an *EcoR I/Sac I* restriction analysis.

**B.1.5.2 Plasmids for protein expression**Plasmids encoding the viral protein Gag

**3-CCCC** was generously provided by Prof. Dr. H.G. Kräusslich (Universitätsklinikum Heidelberg) (653). The plasmid encodes the complete Gag precursor protein (Pr55<sup>Gag</sup>) and the Protease subunit of the *pol* gene of the HIV-1 isolate BH10 (486). *Gag*-encoding transcripts are expressed from the cytomegalovirus (CMV) immediate early promoter and are polyadenylated using the polyadenylation signal of the bovine growth hormone (BGH). Efficient Rev-independent nucleocytoplasmic export of the mRNA is mediated by four copies of a constitutive RNA transport element (CTE) derived from Mason-Pfizer Monkey Virus (M-PMV) inserted downstream of the *gag/pol*-encoding region (653).

**CMV gag 3-CCCC** was cloned by substituting the 128 bp *Mfe I/Bgl II*-fragment of 3-CCCC with an *Mfe I/Bgl II*-digested PCR-amplicon derived by using sense primer #1513 and antisense primer #1509 introducing a mutation of the heptameric frame shift region using plasmid 3-CCCC as template.

**CMV MA/CA 3-CCCC** was generated by replacing the 1899 bp *BssH II/BamH I*-fragment of 3-CCCC with a 1171 bp *BssH II/BamH I*-restricted PCR product derived by using primer pair #1512/#1510 employing 3-CCCC as template.

**CMV MA 3-CCCC** was cloned by substituting the 1899 bp *BssH II/BamH I*-fragment of 3-CCCC with an 478 bp *BssH II/BamH I*-digested PCR amplicon generated using sense primer #1512 and antisense primer #1511 using 3-CCCC as template.

Plasmid encoding the viral protein Rev

**SVcrev** expresses the *rev*-cDNA derived from the HIV-1 isolate NL4-3 under control of the SV40 early promoter and was cloned by replacing the *EcoR I/Xho I*-fragment from pSVT7 with the *EcoR I/Xho I*-fragment from pUHcrev (519).

Expression plasmid encoding fusion proteins

**SV SD4/SA7 NLS-PP7 SC35pA** encodes a fusion protein of the PP7 envelope protein and the RS domain of SC35. The plasmid was cloned by Dr. C. Konermann [diploma thesis (320)].

Prokaryotic expression plasmid encoding recombinant T7 RNA polymerase

**pQE-80L-T7RNAP** was kindly provided by Prof. Dr. M. Caputi (Florida Atlantic University, USA). The gene encoding the RNA polymerase of the bacteriophage T7 was subcloned into the plasmid pQE-80L (Qiagen) generating plasmid pQE-80L-T7RNAP and is expressed with an N-terminal 6xHis-tag. The ampicillin resistance gene is used for selection.

**B.1.5.3 Plasmids for recombinant U1snRNA expression**

The U1 snRNA expression plasmids **pUCBU1 9T10G11C** (190), and **pUCBU1 6A** (295) harbouring mutations in the free 5'-end of the encoded U1 snRNA have already been described. pUCBU1 6Apm was kindly provided by Dr. M. Freund.

**B.1.5.4 Control plasmids**

**pSP73** (Promega) was used to adjust DNA amounts between samples during transfections for RNA analyses.

**pSVT7** (51) was employed to adjust DNA amounts in transfection experiments aiming on protein analyses.

**pGL3-control** (Promega) was used in transient transfection experiments to normalise samples regarding transfection efficiency. The plasmid encodes the firefly (*Photinus pyralis*) *luciferase* gene under control of the SV40 early promoter.

**pXGH5** (536) was cotransfected to monitor transfection efficiency in semi-quantitative RNA analyses. It expresses the human growth hormone (hGH) under the control of the mouse metallothionein-1 promoter (mMT1), which is otherwise only expressed in cells of the anterior pituitary.

**B.1.6 Antibodies**

Several antibodies targeting viral, spliceosome-associated or splicing-regulatory proteins were used in working dilutions described in Tab. II-2 and were detected applying appropriate secondary antibodies as listed in Tab. II-3.

Tab. II-2: Primary antibodies

Target subgroup	Protein	Provider	Cat.-No. (Ref.-No.)	Host	Clonality (clone)	Working Dilution
HIV-1	Gag	Prof. Dr. O. Adams	(#18928)	Hu	polyclonal	1:200
	CA	DuPont	NEA#9283	Ms	monoclonal	1:1.000 1:500 (IF)
	MA	Advanced Biotechnologies	13-103-100	Ms	monoclonal	1:5.000 1:500 (IF)
	Env	Dade-Behring	87-133/026	Ms	(237)	1:5000
	Nef	DuPont	NEA#9302	Ms	monoclonal	1:200
SR proteins	SR proteins (phospho-epitope)	Zymed	33-9400	Ms	(1H4)	1:1.000
	SF2/ASF	Zymed	32-4500	Ms	(96)	1:2.000
	SF2/ASF	Prof. Dr. A Krainer	-	Ms	(AK-96)	1:500
	SRp40	US Biological	S5554-58	Rb	polyclonal	1:500
	hTra2- $\beta$	Abcam	ab31353	Rb	polyclonal	1:300
	SC35	Prof. Dr. M. Caputi	mab532	Rb	polyclonal	1:1500
hnRNP proteins	hnRNP H	Prof. Dr. D. L. Black	AN113	Rb	polyclonal	1:1.000
	hnRNP A2/B1	Santa Cruz Biotechnology	sc-32316	Ms	(DP3B3)	1:200
	hnRNP A1	Santa Cruz Biotechnology	sc-10032	Go	polyclonal	1:200
	hnRNP A1	Santa Cruz Biotechnology	sc-56700	Ms	(9H10)	1:200
Spliceosome-associated proteins	U1-70K	Santa Cruz Biotechnology	sc-9571	Go	polyclonal	1:200
	U2AF35	ptglab	BC001923	Rb	polyclonal	1:1.000
	U2AF65	Prof. Dr. M. Hastings/ Prof Dr. A. Krainer	-	Ms	(MC3)	1:20
	SAP155	MBL Intl. Corporation	D221-3	Ms	monoclonal	1:5.000
	SF1/mBBP	Lifespan Biosciences	LS-C31720	Rb	polyclonal	1:1.000
control	$\beta$ -Actin	Sigma-Aldrich	A5316	Ms	(AC-74)	1:20.000

Unless otherwise noted working dilutions refer to immunoblotting [Rb: rabbit, Go: goat, Hu: human, IF: immunofluorescence, Ms: mouse, Ref.-No.: reference number].

**Tab. II-3: Secondary antibodies**

Epitope Ig species	Provider	Cat.-No.	Host	Conjugate	Working dilution
Rb	Sigma-Aldrich	A6154	Go	POD	1:2.500
Go	Dianova	705-035-147	Dn	POD	1:5.000
Hu	Promega	#S382B	Go	AP	1:2.000
Ms	Amersham	NA9310	Sh	POD	1:1.000
Ms	Jackson ImmunoResearch Laboratories	115-166-003	Go	Cy3	1:200 (IF)

Unless otherwise noted working dilutions refer to immunoblotting [AP: Alkaline Phosphatase, Dn: donkey, Go: goat, Hu: human, IF: immunofluorescence, Ms: mouse, POD: horseradish Peroxidase, Rb: rabbit, Sh: sheep].

## B.1.7 Algorithms and databases

### B.1.7.1 Evaluation of splice site strength

The intrinsic 5' ss strength was predicted using the splicefinder algorithm ([http://www.uni-duesseldorf.de/rna/html/hbond\\_score.php](http://www.uni-duesseldorf.de/rna/html/hbond_score.php)) and the MaxEntScore algorithm for 5' ss [[http://genes.mit.edu/burgelab/maxent/Xmaxentscan\\_scoreseq.html](http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html)] (669)]. The intrinsic strength of 3' ss was evaluated applying the MaxEntScore algorithm for 3' ss [[http://genes.mit.edu/burgelab/maxent/Xmaxentscan\\_scoreseq\\_acc.html](http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html)] (669)].

### B.1.7.2 Prediction of SR protein binding sites

Binding of the SR proteins SF2/ASF, SRp40, SC35 and SRp55 to RNA sequences was predicted using ESEfinder (ver. 2.0 and 3.0 as indicated) [[http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese\\_finder.cgi?process=home](http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese_finder.cgi?process=home)] (97, 566)]. The program scores potential SR protein binding sites based on their similarity to consensus binding sites generated by functional SELEX analyses. SR protein specific threshold values mark scores above which binding of the respective protein is suggested. For SR protein binding site predictions in this work default threshold values were applied.

### B.1.7.3 2D protein separation and identification

Isoelectric points (pI) of representative spliceosomal and splicing regulatory proteins in different phosphorylation states were calculated on [www.phosphosite.org](http://www.phosphosite.org) using the algorithm from

ExPASy's Compute pI/Mw program ([http://www.expasy.org/tools/pi\\_tool.html](http://www.expasy.org/tools/pi_tool.html)) to define the pH-range for isoelectric focussing in the first dimension of 2D gel electrophoresis. Spot identification and matching analyses for 2D-separation was performed using ImageMaster 2D Platinum 6.0 (GE Healthcare).

Recording of MS spectra was performed using Analyst® QS software (AB SCIEX). Peptides analysed by mass spectrometry were identified by comparison of the respective MS spectra with the MASCOT database (MASCOT MS/MS ions search; [www.matrixscience.com](http://www.matrixscience.com)).

## **B.2 Methods**

Standard methods of molecular biology e.g. DNA restriction analyses, DNA separation by agarose gel electrophoresis, DNA isolation from agarose gels and generation of transformation-competent bacteria were performed according to the laboratory handbooks 'Molecular Cloning' (512) and 'Current Protocols in Molecular Biology' (22).

### **B.2.1 Cloning**

#### **B.2.1.1 Polymerase-Chain-Reaction (PCR)**

DNA fragments used for cloning of recombinant plasmids were amplified in a volume of 100  $\mu$ L using 2.5 U Pwo DNA Polymerase (Roche) and 100 ng DNA template in a reaction containing 10 mM Tris-HCl, pH 8.85, 25 mM KCl, 5 mM  $(\text{NH}_4)_2\text{SO}_4$ , 2 mM  $\text{MgSO}_4$ , 200  $\mu$ M deoxy-nucleosidetriphosphates (dNTP) (Applied Biosystems), 200 nM sense and antisense primer, respectively. DNA was amplified in a Robocycler Gradient 96 (Stratagene) using following conditions: denaturation 1 x 94°C 3 min; amplification 30 x 94°C 30 sec, x°C 1 min (Annealing temperature ranged between 52°C and 56°C depending on templates and primers used for individual amplicons.), 72°C 1 min; final elongation 1 x 72°C 10 min. PCR products were purified from the reaction by adding 1 vol. phenol (Roth) and 1 vol. chloroform/isoamyl alcohol (24:1). After vortexing phases were separated by centrifugation (12.000 rpm, 5 min, Eppendorf microcentrifuge) and the supernatant again extracted with 1 vol. chloroform/isoamyl alcohol (24:1). After separation (12.000 rpm, 5 min, Eppendorf microcentrifuge), DNA in the aqueous phase was precipitated with 0.1 vol. 4M LiCl and 2.5 vol ethanol (96%) at -80°C for 20 min. After centrifugation (12.000 rpm, 15 min, Eppendorf microcentrifuge), DNA was washed with 200  $\mu$ L ethanol [70% (v/v)] (12.000 rpm, 10 min, Eppendorf microcentrifuge), air-dried and resuspended in 30  $\mu$ L ddH<sub>2</sub>O.

#### **B.2.1.2 Ligation**

Plasmids and PCR products were digested with respective restriction enzymes using the buffers provided by the manufacturer in a total reaction volume of 20  $\mu$ L. The fragment pattern was controlled in 1% LE agarose (Biozym) or 3% metaphor agarose gels for fragments smaller than 150 nt (Biozym) using TBE buffer (89 mM Tris borate, pH 8, 2 mM EDTA). Remaining DNA was separated on 0.8% low melting sieve GP agarose (Biozym) using TB<sup>1</sup>/<sub>10</sub>E buffer (89 mM Tris borate, pH 8, 0.2 mM EDTA). DNA fragments were visualised in "low melt" gels using long-wave UV light (320 nm) and cut from the gel. Agarose containing isolated DNA fragments were melt at 65°C. Ligation of DNA fragments was performed in 20  $\mu$ L reactions containing a maximum input of 3  $\mu$ L DNA fragments in melted agarose, 400 U T4 DNA ligase (NEB), 50 mM Tris-HCl,

pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 1 mM ATP and 25 µg/mL BSA. For 1-insert cloning 1 µL plasmid backbone fragment and 2 µL insert fragment were applied, whereas for 2-insert cloning 0.5 µL plasmid backbone fragment, 1 µL longer and 1.5 µL shorter insert fragment were used. DNA fragments were ligated at 37°C for 1 h. Usually, ligations were used for transformation of *E. coli* without further incubation. Nevertheless, it was noticed that additional incubation at RT ON increased ligation efficiency. Therefore, reactions ligating with low efficiency were subjected to *E. coli* transformation after prolonged incubation.

### **B.2.1.3 Transformation**

The *E. coli* strain DH5αF'IQ (Invitrogen) was used for plasmid amplification. For bacteria transformation with ligated plasmids 5 µL of the ligation reaction were applied to chemically competent bacteria. For transformation of bacteria with highly purified plasmids 0.5 ng DNA in ddH<sub>2</sub>O were used. Cells were incubated with DNA for 20 min on ice. After heat shock for 1.5 min at 42°C followed by a short cooling period on ice, cells were transferred into 400 µL SOC medium [2% (w/v) Bacto-Tryptone, 0.5% (w/v) yeast extract, 8.6 mM NaCl, 50 mM glucose and 10 mM MgCl<sub>2</sub>, pH 7] and incubated for 1 h at 37°C and 220 rpm. Subsequently, the complete bacteria suspension was plated on ampicillin-containing LB agar (100 µg/mL) and propagated ON at 37°C.

### **B.2.1.4 Analytical plasmid DNA isolation**

Single colonies were transferred into 5 mL LB medium containing ampicillin (100 µg/mL) and were grown at least 16 h ON at 37 °C and 220 rpm. Isolation of plasmid DNA was performed using a modified protocol of alkaline lysis. To this end, an 1 mL aliquot of the bacteria suspension was pelleted by centrifugation for 1 min at RT applying 14.000 rpm (Eppendorf microcentrifuge). The pellet was resuspended in 300 µL buffer 1 (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 400 µg/mL RNase A). Cells were lysed by addition of 300 µL buffer 2 [0.2 M NaOH, 1% (w/v) SDS] at RT for 5 min. The lysate was neutralised by addition of 300 µL buffer 3 (3 M KAc, pH 5.5). After centrifugation for 15 min using 14.000 rpm at 4°C DNA within the supernatant was precipitated with 0.7 vol. isopropanol. After centrifugation for 30 min at 11.000 rpm and RT the pellet was washed with 200 µL ethanol [75% (v/v)], air-dried and resuspended in TE buffer (pH 8). Positive clones were identified by restriction with appropriate DNA endonucleases, which was controlled by gel electrophoresis using 1% LE agarose (Biozym) or 3% metaphor agarose (Biozym) for DNA fragments smaller than 150 nt. For immediate sequencing plasmid DNA was purified using silica-columns (FastPlasmid™ Mini, Eppendorf).

### **B.2.1.5 Preparative plasmid DNA isolation**

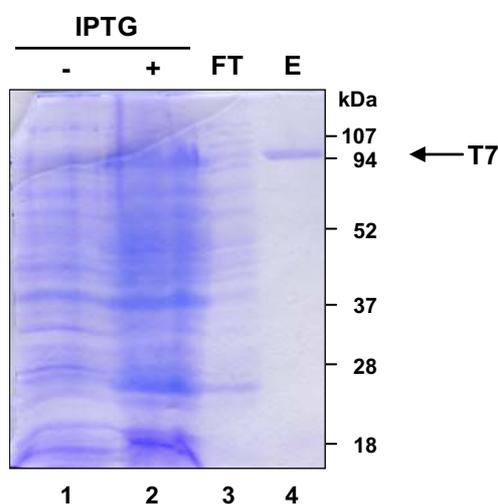
200 mL ampicillin-containing LB medium (100 µg/mL) were inoculated with a single colony derived from plated glycerol stocks (prepared from 700 µL overnight grown bacterial culture and 300 µL glycerol, stored at -80°C) or with 100 µL of a 5 mL preparatory culture and incubated ON at 37°C and 220 rpm. Bacteria were pelleted by centrifugation for 5 min at 4.000 rpm (Beckmann JS-21, JA14 Rotor) and resuspended in 10 mL buffer 1 (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 400 µg/mL RNase A). Cells were lysed by addition of 10 mL buffer 2 [0.2 M NaOH, 1% (w/v) SDS] and incubated for 5 min at RT. The lysate was neutralised by addition of 10 mL buffer 3 (3 M KAc, pH 5.5). After 15 min incubation on ice and centrifugation for 30 min at 10.000 rpm and 4°C (Beckmann JS-21, JA14 rotor) the supernatant was filtered through a folded filter (Schleicher & Schüll, 595½, Ø150 mm) and plasmid DNA purified using silica-based anion-exchange-columns (Plasmid DNA Maxi Kit, Qiagen) according to the manufacturer's instructions. After precipitation with 0.7 vol. isopropanol, followed by 30 min centrifugation at 10.000 rpm and a washing step with 75% ethanol, plasmid DNA was air-dried and resuspended in 50-200 µL TE buffer (pH 8). DNA concentration was measured by spectral photometry at 260 nm and 280 nm and adjusted to 1 µg/µL with TE buffer (pH 8). The quality of the preparation was confirmed by agarose gel electrophoresis.

### **B.2.1.6 DNA sequencing**

To confirm mutations introduced into plasmids or to identify RT-PCR products after gel electrophoretic separation, DNA was sequenced using the dideoxy method by Sanger et al. (515). Sequencing reactions were performed in a volume of 20 µL containing 5 pmol primer, 4 µL BigDye v1.1 RR-24 reaction mix (includes DNA-polymerase, labelled and unlabelled dNTPs and reaction buffer) and template-dependent DNA amounts (200-500 ng DNA for plasmid and 50 ng for PCR product sequencing). Sequencing PCR was performed by 26 cycles of denaturation (94°C, 0:30 min), annealing (55°C, 0:30 min) and elongation (60°C, 4:00 min). Sequencing reactions were purified by ethanol/sodium acetate precipitation [0.3 mM sodium acetate, pH 5.2, 78% (v/v) ethanol] for 15 min. After centrifugation at 14.000 rpm for 20 min the pellet was washed with 250 µL 75% (v/v) ethanol. After a second centrifugation at 14.000 rpm for 5 min the supernatant was removed, the pellet was air-dried and dissolved in 10 µL formamide. Sequencing reactions were protected from light and stored at 4°C till separation on an automated DNA sequencer (3130 Genetic Analyzer, Applied Biosystems). Additional sequencing reactions were performed by the Analytical Core Facility of the Biological-Medical Research Centre (BMFZ, HHUD).

## B.2.2 Expression and purification of recombinant T7 RNA polymerase

A glycerol stock of the *E. coli* strain BL21(DE3) transformed with the T7 RNA polymerase expression plasmid pQE-80-L-T7RNAP was streaked out on LB agar plates containing ampicillin (100 µg/mL). A single colony was transferred into 2 mL ampicillin-containing LB medium (100 µg/mL) and incubated for 2 h at 220 rpm and 37°C. The 2 mL preparatory culture was transferred into 50 mL LB medium containing ampicillin (100 µg/mL) and incubated ON at 220 rpm and 37°C. 10 mL of the overnight culture were transferred into 500 mL LB medium containing ampicillin (100 µg/mL) and propagated for 2-3 h at 220 rpm and 37°C until the optical density at 600nm reached 0.6. Recombinant protein expression was induced by addition of IPTG in a final concentration of 1 mM to the bacterial culture. 1 mL aliquots were taken prior to and after IPTG supplementation to control the induction of T7 RNA polymerase expression (Fig. II-1). After another 4 h cultivation bacteria were harvested by centrifugation (4.000 g, 4°C, 4 min) and resuspended in 5 mL chilled LB medium. After centrifugation (4.000 g, 4°C, 4 min) cells were resuspended in binding buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8, 300 mM NaCl, 10 mM imidazole) (2-5 mL/g bacteria pellet). After addition of PMSF (Sigma) to a final concentration of 2 mM, 2 mg/mL lysozym (AppliChem) and 1000 U DNase I, RNase-free (Roche) bacteria were incubated for 30 min on ice and subsequently sonicated for 4 x 10 sec. The bacteria suspension



**Fig. II-1: Control of T7 RNA polymerase expression and purification.**

Bacteria sedimented from 1 mL aliquots prior to and after IPTG induction were lysed in 2x protein sample buffer. Proteins were separated in a 10%-SDS-polyacrylamide gel and compared to the column flow-through and 10 µL purified T7 RNA polymerase (E) corresponding to the input amount of 1 mL bacterial culture. The gel was stained using a modified coomassie brilliant blue staining protocol (299) [FT: flow-through, E: eluate].

was cleared from cell debris by centrifugation at 35.000 rpm and 4°C for 30 min (Ultracentrifuge; Beckmann). Recombinant T7 RNA polymerase was purified from the supernatant by affinity chromatography of the His-tagged protein using Ni-NTA agarose (Invitrogen) in a C 26/40 chromatography column (GE Healthcare). Unspecifically bound proteins were removed from the affinity column by washing with 10 vol. binding buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8, 300 mM NaCl, 10 mM imidazole) followed by 8 vol. washing buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8, 300 mM NaCl, 20 mM imidazole). T7 RNA polymerase was eluted from the affinity column with 3 vol. elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8, 300 mM NaCl, 250 mM imidazole). Salt conditions of the eluted enzyme solution were adjusted by dialysing against dialysis buffer [20 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.7, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 50% (v/v) glycerol] for 12-22 h. The purified enzyme was stored at -20°C. An aliquot of purified T7 RNA polymerase, lysates of the bacteria culture prior to and after induction with IPTG and the flow-through of the purification procedure were analysed on 10% SDS-polyacrylamide gels to control expression of the T7 RNA polymerase and purity of the enzyme preparation (Fig. II-1).

### **B.2.3 Eukaryotic cell culture**

#### **B.2.3.1 Maintenance of eukaryotic cells**

Cells were cultivated at standard incubator conditions (humidified atmosphere, 5% CO<sub>2</sub>, 37°C). Cells were maintained in T75 cell culture flasks (TRP) using Dulbecco's modified Eagle's medium (DMEM, Invitrogen) containing 4mM L-Glutamine and additionally supplemented with 10% fetal calf serum (FCS, GIBCO), 100 U/mL penicillin and 100 µg/mL streptomycin (PenStrep, GIBCO). HeLa-T4<sup>+</sup> and HeLa-T4R5 cells were positively selected by application of 100 µg/mL Geneticin (GIBCO). In addition, HeLa-T4R5 cells were positively selected by application of 3 µg/mL Puromycin (Sigma-Aldrich). Subconfluent cell monolayers were detached for passaging or transfection after washing the cells twice with 5 mL PBS<sub>def</sub> (Dulbecco's Phosphate Buffered Saline deficient in Ca<sup>2+</sup> and Mg<sup>2+</sup>, DPBS, Invitrogen) each by trypsinisation with 1.3 mL Trypsin for 10 min at 37°C. The enzyme was inactivated by resuspending the cells in 10 mL DMEM for passaging or PBS<sub>def</sub> for cell counting.

#### **B.2.3.2 Transfection of eukaryotic cells using FuGene 6**

The number of living cells was determined by negative staining using trypan blue [0.4% (w/v), GIBCO] in a Neubauer counting chamber. For western blot analyses 3 x 10<sup>5</sup> cells and for RT-PCR and immunofluorescence analyses 2 x 10<sup>5</sup> cells were seeded in 6-well plates (TRP) in 2 mL DMEM supplemented with 10% FCS lacking antibiotics. For immunofluorescence assays cells were seeded on glass cover slips in 6-well plates. Transfection was performed after 24 h using FuGENE® 6 transfection reagent (Roche) according to the protocol provided by the

manufacturer. For Western Blot (B.2.5) and immunofluorescence analyses (B.2.4) the medium was exchanged 24 h after transfection and transfected cells were harvested after 48 h. For RT-PCR analyses (B.2.7) cells were harvested 30 h after transfection.

### B.2.3.3 Viral infection

Infection experiments were performed by Dr. I. Hauber (Heinrich-Pette-Institut, Hamburg). Cell cultures were maintained in RPMI 1640 medium containing 10% FCS (Pansystems GmbH) and antibiotics (penicillin and streptomycin). Viral stocks were prepared by transfecting 293T cells with the provirus expression vector pNL4-3 and expression confirmed by ELISA of culture supernatant detecting the capsid protein (p24) (Innotest HIV p24 Antigen mAB; Innogenetics N.V.). For HIV-1 infection  $5 \times 10^6$  PM1 cells (371), a CD4<sup>+</sup> Hut78-derived T cell clone, were resuspended in 500  $\mu$ L culture medium and incubated at 37°C for 3 h with 100 ng of the viral stock. After infection cells were washed twice with PBS<sub>def</sub> and further cultured in 5 mL medium for another 5 days. Subsequently, total RNA was prepared using Trizol reagent (Invitrogen) according to the protocol provided by the manufacturer.

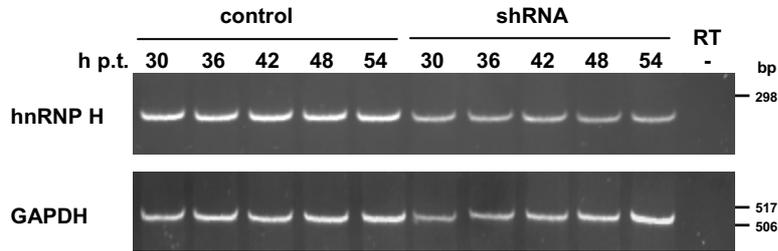
### B.2.3.4 shRNA-mediated knockdown of hnRNP H expression

To determine the influence of hnRNP H expression on the splicing pattern of the 2-intron construct,  $3 \times 10^5$  HeLa-T4<sup>+</sup> cells were transiently cotransfected with 1  $\mu$ g of the 2-intron minigene SV leader SD1 SA5 env nef, 1  $\mu$ g SVcrev or 1  $\mu$ g pSP73, and 4  $\mu$ g shRNA targeting *hnRNP H*-mRNA using Lipofectamine™ 2000 (Invitrogen). The transfection medium was changed against DMEM supplemented with 10% FCS 4 h post transfection. Cytoplasmic RNA was isolated 54 h post transfection (B.2.6.3), resuspended in 10  $\mu$ L DMDC-ddH<sub>2</sub>O and stored at -80°C until PCR analysis (B.2.7).

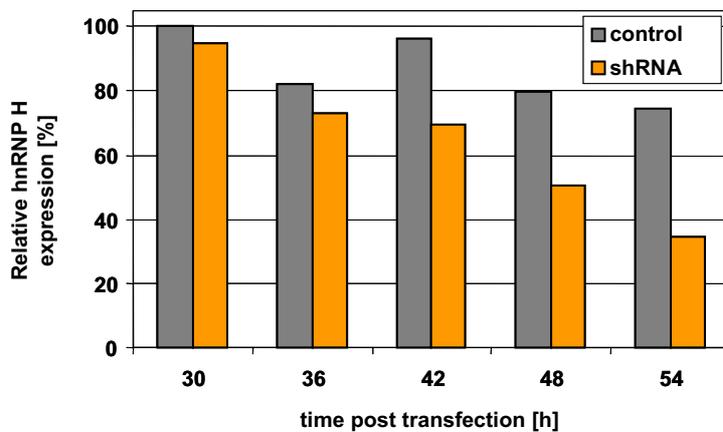
The prolonged cultivation time of transfected cells compared to general RT-PCR analysis applied to transient minigene transfection experiments resulted from a preliminary experiment monitoring the time course of hnRNP H knockdown after transfection of eukaryotic cells with shRNA. To this end,  $6 \times 10^5$  HeLa-T4<sup>+</sup> cells were seeded in 60 mm cell culture dishes (TRP) 24 h before transfection. Cells were transfected with 8  $\mu$ g shRNA (generously provided by Prof. Dr. M. Caputi) using Lipofectamine™ 2000 (Invitrogen) according to the protocol provided by the manufacturer. The transfection medium was changed 4 h post transfection against DMEM supplemented with 10% FCS. Total RNA was isolated 30, 36, 42, 48 and 54 h post transfection, respectively, using the 3.5-fold upscaled phenol-chloroform based protocol (see B.2.6.1), resuspended in 35  $\mu$ L DMDC-ddH<sub>2</sub>O and stored at -80°C. RT-PCR analysis of *hnRNP H*- and *GAPDH*-mRNA expression using primer pairs MC20.54/MC20.55 and #3153/#3154, respectively, revealed that knockdown of *hnRNP H*-mRNA expression is already

detectable 30 h post transfection, whereas the level of *GAPDH*-mRNA expression requires 54 h to achieve the expression level of control cells (Fig. II-2).

**A**



**B**



**Fig. II-2: Kinetic of *hnRNP H*-mRNA knockdown**

**(A)** RT-PCR analysis of HeLa-T4<sup>+</sup> cells transfected with an shRNA targeting *hnRNP H*-mRNA or mock-transfected. RT-PCR was performed applying a two-step protocol priming reverse transcription with oligo-(dT). PCR products were resolved in non-denaturing 6% polyacrylamide gels and stained with ethidiumbromide [h p.t.: hours post transfection, RT: reverse transcription].

**(B)** Quantification of *hnRNP H*-mRNA expression from the experiment shown in (A). Expression of *hnRNP H*-mRNA was normalised to *GAPDH*-mRNA expression. The amount of *hnRNP H*-mRNA expressed 30 h after shRNA transfection in mock-transfected control cells was set as 100%.

### B.2.3.5 Inhibition of SR protein phosphorylation

Two compounds were used for inhibition of SR protein phosphorylation. The benzothiazole compound TG003 [(Z)-1-(3-ethyl-5-methoxy-2,3-dihydro-benzothiazole-2-yliden)propan-2-on] (kindly provided by Dr. C. Ehlers and Prof. Dr. J. Hauber, Heinrich-Pette-Institut, Hamburg) was reported to abrogate phosphorylation of SF2/ASF through inhibition of the Clk1/Sty kinase activity (431). In a second set of experiments, the substance Drug29 (by courtesy of

Prof. J. Tazi, IGM Montpellier, France) was tested for the ability to influence the SR protein phosphorylation.

The respective SR protein kinase inhibitor was dissolved in DMSO. The compound TG003 was pre-warmed to 37°C and diluted in pre-warmed DMEM to avoid precipitation. The compounds were applied to the cell culture medium 3 h post transfection in final concentrations of 10 µM and 50 µM for TG003 and 1 µM and 2.5 µM for Drug29 thereby introducing DMSO in a final concentration of 0.5% (v/v). In addition, the medium of control cells was supplemented with only DMSO in a final concentration of 0.5% (v/v). Kinase inhibitors and DMSO remained in the medium until cells were harvested. 30 h post transfection phase contrast images were taken at 160-fold magnification using a Canon PowerShot A620 digital camera connected to a Nikon TS1000 microscope. Subsequently, RNA was isolated using a phenol/chloroform-based protocol (B.2.6.1) and analysed by RT-PCR (B.2.7.1).

#### **B.2.4 Immunofluorescence**

$2 \times 10^5$  HeLa-T4<sup>+</sup> cells were seeded on glass cover slips in 6-well plates (TRP) and after 24 h transfected with indicated Gag and MA expression plasmids. Protein expression from the transfected plasmids was analysed 30 h post transfection. All washing and incubation steps were performed in a vol. of 500 µL. Cells were washed in PBS<sub>def</sub> at RT for 5 min, fixed in 3% paraformaldehyde in PBS<sub>def</sub> for 20 min and subsequently washed in PBS<sub>def</sub> for 5 min. Cells were permeabilised by incubation in 0.02% saponin (Calbiochem-Merck) in PBS<sub>def</sub> for 15 min. Remaining binding sites were blocked with 2% normal goat serum (Dako) in permeabilisation solution diluted 1:10 with PBS<sub>def</sub> for 20 min. The Gag precursor protein was detected using a mouse monoclonal antibody targeting an epitope in the capsid subunit (#CA; DuPont). The matrix protein was detected with a monoclonal antibody targeting the matrix subunit (#p17; Advanced Biotechnologies). The respective primary antibody was diluted 1:500 in 1:10 PBS<sub>def</sub>-diluted blocking solution. After incubation with primary antibodies for 1 h cells were washed three times 5 min each with 1:10 PBS<sub>def</sub>-diluted permeabilisation solution. As secondary antibody Cy3-labelled goat anti-mouse IgG (H+L) (Jackson ImmunoResearch Laboratories) was applied in a 1:200 dilution with 1:10 PBS<sub>def</sub>-diluted blocking solution. After incubation in the dark for 45 min immunostained cells were washed twice 5 min each with PBS<sub>def</sub>. Cells were additionally incubated in 4,6-diamidino-2-phenylindole (DAPI; Invitrogen) for 3 min. After a final wash with PBS<sub>def</sub> for 5 min cover slips were mounted on microscope slides with FluoromountG (Southern Biotechnology Associates) and stored at 4°C ON. Immunofluorescence was visualised at 1.000-fold magnification using a Zeiss Axioscope equipped for epifluorescence. Images were acquired using a CCD camera (Micromax, Princeton Instruments) coupled to the Metaview Imaging System software (Universal Imaging Corp.).

## **B.2.5 Western Blot Analysis**

### **B.2.5.1 Cell harvesting**

Adherent cells were scraped into the culture medium 48 h after transfection, shortly centrifuged, washed with 1 mL PBS<sub>def</sub>, again centrifuged and resuspended in 500 µL PBS. From the cell suspensions 70 µL aliquots were collected to determine protein concentration (see B.2.5.2) and transfection efficiency (see B.2.5.3). Cell aliquots were solubilised by three cycles of freezing in methanol/dry ice and thawing at 37°C. Cellular debris was sedimented and the supernatant used for total protein and luciferase measurements.

Remaining cells were sedimented (1 min, 10.000 rpm, Eppendorf microcentrifuge) and cellular proteins denaturated by 200 µL treatment buffer [2% (w/v) SDS, 65 mM Tris-HCl, pH 6.8, 10% (v/v) glycerol, 10% (w/v) DTT], spiked with 0.05 vol. gel loading buffer [50% (v/v) glycerol, 0.25% (w/v) bromphenolblue, 0.25% (w/v) xylencyanol] and boiled for 10 min at 95°C. Aliquots of identical transfection efficiency were normalised with pSVT7-transfected cell extract („mock“) to identical protein amounts.

### **B.2.5.2 Determination of protein concentration**

Protein concentration of cellular lysates was determined according to Bradford (70). The method is based on the protein-mediated shift in the absorption maximum of an acid solution of Coomassie Brilliant Blue G250 from 465 nm to 595 nm. The protein concentration for each probe was measured in duplicates for two dilutions (1:10 and 1:20 in ddH<sub>2</sub>O). 50 µL of diluted protein samples were supplemented with 200 µL diluted Bradford reagent (Bio-Rad, Protein-Assay 1:5 dilution) in 96-well microtiter plates (Nunc). After incubating for 5 min absorption was measured at 595 nm in an ELISA-Reader (SLT Rainbow, SLT Labinstruments) using bovine serum albumin (BSA) as standard ranging from 0-100 µg/mL.

### **B.2.5.3 Normalisation of transfection efficiency using a Luciferase assay**

The cotransfected plasmid pGL3-Control (Promega) encodes the firefly (*Photinus pyralis*) luciferase gene under control of the SV40 early promoter. Luciferase catalyses the oxidation of luciferin (Luciferase Assay System, Promega) to oxyluciferin. Emitted bioluminescence was measured as relative light units (RLU) in a luminometer (Microlumat LB96P, EG&G Berthold). The luciferase-assay was performed in duplicates using 5 and 10 µL cell extract in black 96-well microtiter plates (Nunc) covered with 40 µg bovine serum albumin (BSA). Bioluminescence was measured 2 sec after substrate injection over a period of 15 sec. On the basis of RLU data as extent of transfection efficiency normalisation of western blot samples was performed.

#### **B.2.5.4 Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis (SDS-PAGE)**

Protein separation was performed under denaturing conditions (335) as vertical flat bed gel electrophoresis in discontinuous 0.1% SDS-10% polyacrylamide gels (Rotiphorese Gel 30, Roth). Mini-gels were operated in 1 x SDS running buffer [0.8% (w/v) SDS, 0.2 M Tris-Base, 1.9 M glycine] for 1 h applying a current of 20 mA per gel. To monitor protein size and subsequent blotting efficiency molecular weight markers (PageRuler™ Prestained Protein Ladder Plus, Fermentas; Prestained SDS-PAGE Standard, Low Range, Bio-Rad).

#### **B.2.5.5 Western Blot**

Proteins were transferred from SDS-polyacrylamide gels to PVDF membranes (Millipore, Immobilon-P) by electroblotting either in a tank blot system (Bio-Rad) in transfer buffer [200 mM glycine, 25 mM Tris-Base, 20% (v/v) methanol] for 1 h using 150 mA and additional cooling or in a semi-dry system (Biometra) for 1:30 h applying 0.8 mA/cm<sup>2</sup> membrane.

The membrane was blocked in TBS-T [20 mM Tris-HCl, pH7.5, 150 mM NaCl, 0.05% (v/v) Tween-20] containing 10% (w/v) dry milk for 1 h at RT or ON at 4°C. Binding of the primary antibody was performed for 1 h in TBS-T containing 5% dry milk. After washing the membrane three times for 10 min each in TBS-T, the membrane was incubated with appropriate secondary antibodies in TBS-T containing 5% dry milk for 30 min. The membrane was washed with TBS-T three times 10 min each and twice shortly with TBS.

Antibody binding was visualised using the ECL system (Amersham) for peroxidase-conjugated secondary antibodies, whereas the CDP Star system (Roche) was employed to detect alkaline phosphatase-conjugated secondary antibodies. Both detection assays were used according to the manufacturer's protocol. Chemiluminescence was measured by exposure to ECL hyperfilm (Amersham) or to the Lumi-Imager F1 operating a CCD camera (Roche).

For immunoblot reprobing antibodies were removed by incubating the membrane in Antibody Stripping Solution (Chemicon) for 15 min. Membranes were washed twice 5 min each in blocking solution [10% (w/v) dry milk in TBS-T] and either reprobbed immediately or stored at 4°C.

#### **B.2.6 RNA isolation**

All solutions prepared for RNA-based protocols were generated using dimethyl-dicarbonate (DMDC)-treated ddH<sub>2</sub>O.

### **B.2.6.1 Phenol/chloroform-based isolation of total RNA**

RNA from transfected cells was isolated by a phenol/chloroform-based protocol. Cells grown in 6-well culture plates were washed twice with 2 mL ice-cold PBS each and lysed by addition of 500  $\mu$ L 1 x Sol. D [4 M guanidinium-isothiocyanat, 25 mM sodium citrate, pH 7, 0.5% (v/v) sodium N-laurylsarcosine]. The lysate was detached from the well using a cell scraper (Nunc) and transferred to a 2 mL reaction tube. RNA was extracted by addition of 1.43% (v/v)  $\beta$ -mercapto-ethanol, 0.1 vol. 2 M NaAc (pH 4), 1 vol. phenol (pH 4, Roth) and 0.2 vol. chloroform/isoamyl alcohol (24:1). After vortexing for 15 sec, samples were cooled on ice for 15 min. After phase separation (10.000 rpm, 4°C, 20 min, Eppendorf microcentrifuge) the aqueous phase was transferred to a 1.5 mL reaction tube and RNA precipitated with 1 vol. isopropanol (4°C, ON). RNA was sedimented by centrifugation (10.000 rpm, 4°C, 20 min, Eppendorf microcentrifuge), washed twice with 200  $\mu$ L ethanol [70% (v/v)] each, air-dried and resolved in 10  $\mu$ L DMDC-ddH<sub>2</sub>O. Until further analyses RNA was stored at -80°C.

### **B.2.6.2 Isolation of total RNA using anionic exchange columns**

RNA of adherent cells was isolated using microspin columns containing a silica-matrix (GenElute™ Mammalian Total RNA Kit, Sigma). Cells were washed twice with 2 mL PBS each and lysed by addition of 250  $\mu$ L lysis buffer per 6-well. The lysate was centrifuged through filtration columns (14.000 rpm, 2 min, Eppendorf microcentrifuge). The flow-through was mixed with 1 vol. ethanol (70%) and loaded on RNA binding columns by centrifugation (14.000 rpm, 15 sec, Eppendorf microcentrifuge). Column-bound RNA was washed with 500  $\mu$ L wash buffer 1, the column transferred to another reaction tube and subsequently washed with 500  $\mu$ L wash buffer 2 (14.000 rpm, 15 sec, Eppendorf microcentrifuge). A second washing step with wash buffer 2 was performed for 2 min (14.000 rpm, Eppendorf microcentrifuge). RNA was eluted from the column with 50  $\mu$ L elution buffer. After determining the concentration by photometry at 260 nm and 280 nm RNA was stored at -80°C until further analyses.

### **B.2.6.3 Fractionation of cytoplasmic and nuclear RNA**

Adherent cells grown in 6-well culture plates were washed twice with 2 mL ice-cold PBS each and once with 2 mL hypotonic wash buffer (5 mM Tris-HCl pH 7.4, 2.5 mM MgCl<sub>2</sub>, 1.5 mM KCl). Cells were lysed by addition of 0.4 mL hypotonic lysis buffer [1% (v/v) Triton® X-100, 1% (w/v) sodium deoxycholate, 5 mM Tris-HCl pH 7.4, 2.5 mM MgCl<sub>2</sub>, 1.5 mM KCl]. After incubation on ice for 10 min the cytoplasmic cell lysate was removed and cleared from cellular debris (2.500 rpm, 4°C, 5 min, Eppendorf microcentrifuge). The supernatant was mixed with 1 vol. heated 2 x Sol. D [8 M guanidinium-isothiocyanat, 50 mM sodium citrate, pH 7, 1% (v/v) sodium N-laurylsarcosine]. Nuclei were detached from the culture plate wells into 0.4 mL hypotonic wash buffer using a cell scraper (Nunc), sedimented (2.500 rpm, 4°C, 5 min, Eppendorf

microcentrifuge) and lysed in 100  $\mu$ L 1 x Sol. D [4 M guanidinium-isothiocyanat, 25 mM sodium citrate, pH 7, 0.5% (v/v) sodium N-laurylsarcose]. RNA from both fractions was extracted applying the phenol/chloroform-based protocol described in B.2.6.1.

## **B.2.7 RT-PCR**

During this work the RT-PCR protocol was changed from a two-step assay performing reverse transcription and subsequent PCR in separated reactions to a one-step assay performing both reactions consecutively in the same reaction tube. The one-step assay increases the accuracy of the RT-PCR by reducing the amount of pipetting steps.

### **B.2.7.1 Two-step RT-PCR**

#### *B.2.7.1.1 Reverse Transcription*

Contaminating plasmid DNA in 2  $\mu$ L of the isolated RNA was digested using 10 U recombinant, RNase-free DNase I (Roche) in a 10  $\mu$ L reaction at 37°C for 30 min. The enzyme was heat-inactivated at 70°C for 5 min. 4.5  $\mu$ L of DNase I-treated RNA were incubated with 1  $\mu$ L Oligo-(dT)<sub>15</sub> (Roche), 4  $\mu$ L dNTPs (2.5 mM each, Applied Biosystems) and 3.5  $\mu$ L DMDC-ddH<sub>2</sub>O at 65°C for 5 min and subsequently cooled down on ice for at least 1 min. Reverse transcription was performed after addition of 4  $\mu$ L 5 x First Strand Buffer, 1  $\mu$ L 0.1 M DTT, 1  $\mu$ L (40 U) RNAsin (Promega) and 1  $\mu$ L (200 U) SuperScript™ III RNase H- Reverse Transcriptase (Invitrogen) at 50°C for 1 h. As control for plasmid DNA removal an identical reaction omitting only reverse transcriptase was prepared for each sample. The reaction was subsequently heat-inactivated at 70°C for 15 min. cDNA was stored at -20°C until PCR amplification.

#### *B.2.7.1.2 PCR*

cDNA from 2  $\mu$ L of the Reverse Transcription assay was amplified in a 50  $\mu$ L reaction containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.15 mM MgCl<sub>2</sub>, 0.001% (w/v) gelatin, 200  $\mu$ M dNTPs (Applied Biosystems), 200 nM of each respective primer and 1.25 U AmpliTaq DNA-Polymerase (Applied Biosystems) in 200  $\mu$ L PCR reaction tubes (Biozym) in a Robocycler Gradient 96 PCR cycler (Stratagene). To allow semi-quantitative PCR analysis, the exponential phase of the PCR reaction was determined for each primer pair in preliminary experiments (see Tab. II-4).

**Tab. II-4: Amplicon sizes and PCR conditions**

Target mRNA	Primer	amplicon length [bp]	Annealing temperature [°C]	amplification cycles
ASR	#1542-#1544	149-1011*	52	26-30
Total ASR (nef)	#224/#225	256	52	28
hGH (exon 2-4)	#1274/#1273	161	56	25-28
(exon 3-5)	#1225/#1224	187	56	25-28
hnRNP H	MC20.54/MC20.55	280	58	26
GAPDH	#3153/#3154	520	60	26

\*The amplicon length for mRNA expressed from 2-intron minigene depends on the splicing pattern of the pre-mRNA and the size of the analysed minigene (see Appendix, Fig. VII-1).

### B.2.7.2 One-Step RT-PCR

Remaining DNA in 2 µL of the RNA preparation was digested with 5 U recombinant, RNase-free DNase I (Roche) in a 10 µL reaction at 37°C for 20 min. The enzyme was heat-inactivated at 70°C for 5 min. The reaction was diluted 1:10 with DMDC-ddH<sub>2</sub>O. 0.02 µg DNase I-treated RNA was reverse transcribed (55°C, 30 min) using 0.2 µM of each antisense primer and 0.8 µL SuperScript™ III/Taq-DNA polymerase mixture (Invitrogen) and subsequently amplified according to the protocol provided by the manufacturer in a 20 µL reaction (initial denaturation: 94°C 3 min; amplicon-dependent number of amplification cycles: denaturation 94°C 30 sec, annealing 52°C 1 min, elongation 68°C 1 min; final elongation: 68°C 10 min). Control reactions were performed substituting the enzyme mixture with only PlatinumTaq DNA polymerase (Invitrogen) to control for DNA contamination.

### B.2.7.3 Native gel electrophoresis followed by EtBr staining

5-10 µL cDNA were separated in non-denaturing gels containing 6% polyacrylamide using 1 x TBE as running buffer. Gels were run with 200 V for 0:45 h up to 1:30 h. After gel electrophoretic separation DNA was stained with EtBr (4 µg/mL in 1 x TBE) for 10 min and visualised by UV light excitation in the LumiA3 (Roche).

#### **B.2.7.4 Denaturing gel electrophoresis of fluorescently labelled cDNA**

For generation of fluorescently labelled cDNA Cy5-labelled sense primers #1544 (ASR) and #1225 (hGH) were applied in the PCR. Cy5-labelled primers and PCR products were protected from light. PCR reactions were diluted in a range of 1:2 to 1:5 in ddH<sub>2</sub>O. 6 µL diluted cDNA were mixed with 6 µL of sample buffer in 72-well PS microtest plates (Greiner), covered with a lid and stored at 4°C in the dark. Prior to loading samples were heat-denatured in the microtest plate at 95°C for 5 min and subsequently chilled on ice. Cy5-labelled cDNA was separated in denaturing sequencing gels containing 7 M urea-4.5% polyacrylamide at a constant gel temperature of 40°C using 0.6 x TBE as running buffer in an Automated Laser Fluorescence DNA Sequencer (ALF, Pharmacia) controlled by ALFwin™ software (Pharmacia) measuring laser-induced fluorescence (19). Gels were operated with 38 mA using a maximum voltage of 1500 V for 3-8 h depending on the amplicon size. Cy5-labelled cDNA is excited by a fixed laser beam during electrophoresis. Emitted light is detected by photodetectors collecting Cy5-mediated cDNA signals every 2.5 sec thereby generating electrophoretic trace data. Raw data were analysed using P2 software (Pharmacia).

#### **B.2.7.5 Isolation of cDNA from native polyacrylamide gels**

cDNA bands identified in EtBr-stained polyacrylamide gels were cut from the gel under long wave UV light (320 nm) and reduced into small pieces in a 1.5 mL reaction tube. DNA was eluted by incubation in elution buffer [0.5 M NH<sub>4</sub>Ac, 0.1% (w/v) SDS, 1 mM EDTA] at 37°C ON. After short centrifugation (12.000 g, 1 min, 4°C) the supernatant was transferred in another 1.5 mL reaction tube. Gel pieces were again mixed with 0.5 vol. elution buffer, which after centrifugation were combined with the first supernatant. Remaining polyacrylamide gel pieces were removed from the supernatant by filtration through glass fibre filters (GF/C filter, Whatman). PCR products were precipitated by addition of 2 vol. ethanol (96%) on ice for 30 min. After centrifugation (12.000 g, 10 min, 4°C) the pellet was resuspended in 200 µL Tris-EDTA (pH 8) and 25 µL 3 M NaAc (pH 5.2). PCR products were again precipitated by addition of 2 vol. ethanol (96%) on ice for 30 min. After centrifugation (12.000 g, 10 min, 4°C) the supernatant was removed and the DNA pellet washed with 200 µL ethanol [70% (v/v)] (12.000 g, 10 min, 4°C), air-dried and resuspended in 5 µL ddH<sub>2</sub>O. After determining the concentration of the isolated cDNA by photometry, 50 ng cDNA were applied per 20 µL sequencing (B.2.1.6).

## **B.2.8 RNA affinity chromatography**

### **B.2.8.1 *In vitro* transcription**

#### *B.2.8.1.1 Template length below 100 nt*

Full-length oligonucleotides used for *in vitro*-transcription were purified by separating 100  $\mu$ L oligonucleotides (100  $\mu$ M) supplemented with 150  $\mu$ L 8M urea containing bromphenolblue in 15 % polyacrylamide gels (300 V, 2:30 h). DNA was detected by UV shadowing (320 nm) and full-length oligonucleotides cut from the gel. Gel pieces were further cut into smaller pieces and rotated in 600  $\mu$ L elution buffer [0.5 M  $\text{NH}_4\text{Ac}$ , 0.1% (w/v) SDS, 1 mM EDTA] at 4°C ON. Eluted DNA was purified by phenol-chloroform extraction. After addition of 0.1 vol 3 M NaAc (pH 5), 1 vol. phenol (pH 4) and 0.2 vol. chloroform/isoamyl alcohol (24:1) and centrifugation (13.000 rpm, 4°C, 5 min), DNA in the aqueous phase was precipitated with 1 mL ethanol (96%) at -20°C for 5 min. DNA was sedimented (13.000 rpm, 30 min) and air-dried. DNA was resolved in 52  $\mu$ L DMDC-ddH<sub>2</sub>O and the concentration determined by photometry.

For *in vitro*-transcription 500 pmol of the respective sequence-specific primer and the T7 primer were adjusted to a total volume of 500  $\mu$ L with DMDC-ddH<sub>2</sub>O. Primers were denatured at 90°C for 5 min and subsequently annealed by cooling down at RT for 5 min.

#### *B.2.8.1.2 Template length above 100 nt*

Templates exceeding 100 nt were amplified from plasmids by PCR using Pwo DNA polymerase (Roche) containing proofreading activity. The T7 promoter sequence was incorporated at the 5'-end of the sense primer. Using minigene specific primers the target sequence was amplified using a standard PCR protocol with 35 amplification cycles (B.2.1.1). PCR products were purified using a silica-based spin-column protocol (Gel Extraction Kit, Qiagen). 250 ng dsDNA template were used for *in vitro* transcription.

#### *B.2.8.1.3 In vitro transcription*

For *in vitro* transcription an 1 mL reaction containing 500 pmol pre-annealed oligonucleotides or 250 ng PCR products, 50 mM Tris-HCl, pH 7.5, 15 mM  $\text{MgCl}_2$ , 5 mM DTT, 5 mM NTPs (pH 8, Sigma), 2 mM spermidine and 60  $\mu$ L T7 RNA polymerase (B.2.2), aliquoted into 500  $\mu$ L and incubated at 37°C for 5 h. RNA was precipitated by addition of 1 mL ethanol (96%) to each aliquot and incubation at -80°C for 5 min. RNA was sedimented by centrifugation at 13.200 rpm for 7 min at 4°C. To purify full-length transcripts, the RNA pellet was resolved in 200  $\mu$ L 8 M urea containing bromphenolblue and separated in 15% polyacrylamide gels (300 V, 2-3h). RNA was detected by UV shadowing and the slowest migrating bands cut from the gel. Gel fragments of both 500  $\mu$ L aliquots were chopped into pieces, combined in a 15 mL falcon tube and eluted by rotating in 3 mL elution buffer [0.5 M  $\text{NH}_4\text{Ac}$ , 0.1% (w/v) SDS, 1 mM EDTA] at 4°C

ON. RNA was isolated by addition of 0.1 vol 3 M NaAc (pH 5), 1 vol. phenol (pH 4) and 0.2 vol. chloroform/isoamyl alcohol (24:1). After centrifugation (4.000 rpm, 4°C, 7 min, Eppendorf 5810 R), RNA in the aqueous phase was precipitated by addition of 6 mL ethanol (96%) at -80°C for 5 min and subsequently seeded by centrifugation (4.000 rpm, 4°C, 45 min, Eppendorf 5810 R). RNA pellets were air-dried, resolved in 52-102 µL DMDC-ddH<sub>2</sub>O depending on the pellet size and the RNA concentration photometrically determined. RNAs were stored at -80°C until RNA affinity chromatography.

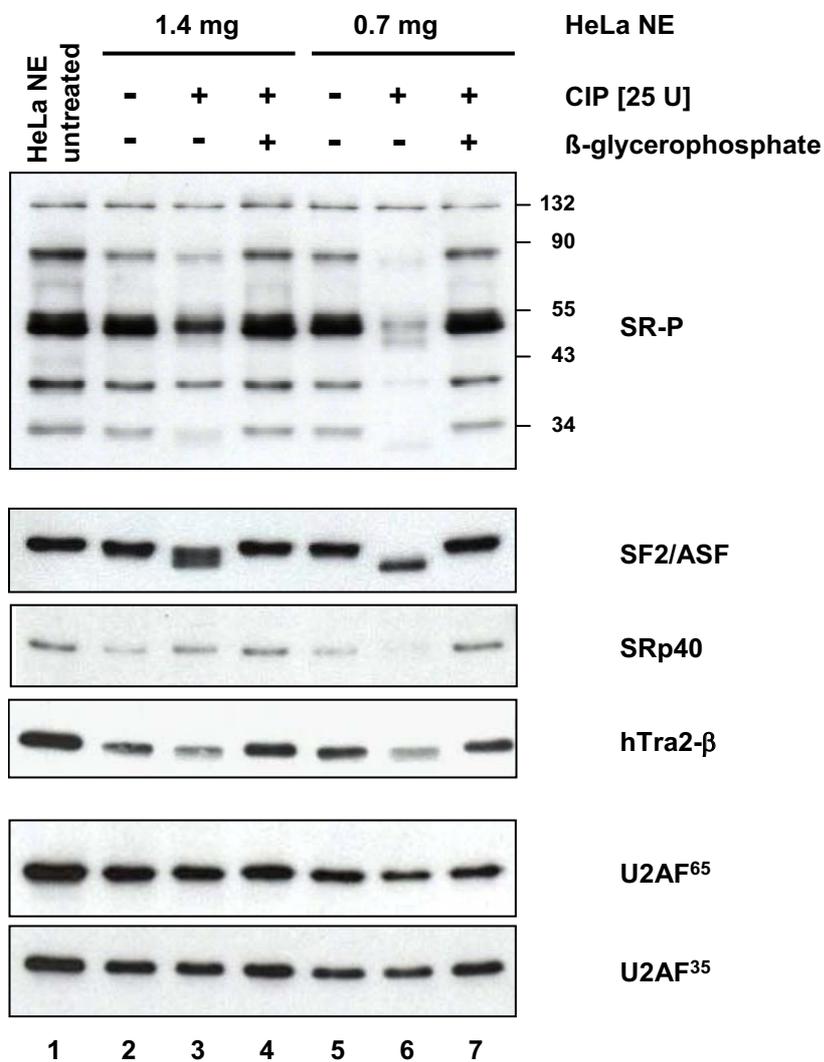
### **B.2.8.2 Protein isolation by RNA affinity chromatography**

RNA affinity chromatography was performed by modification of a published procedure (92). 900-2000 pmol of *in vitro* transcribed RNA were chemically activated in the dark in ProteinLoBind reaction tubes (Eppendorf) in a 400 µL reaction for 1 h (0.1 M NaAc, pH 5, 5 mM Na-m-JO<sub>4</sub>), precipitated with 0.2 vol. NaAc (1 M, pH 5) and 2.5 vol. ethanol (96%) at -80°C for exactly 5 min and sedimented (13.200rpm, 4°C, 30 min). For each sample 125 µL Adipic acid dihydrazide-Agarose suspension (Sigma) were washed four times with 0.1 M NaAc (pH 5) (300 rpm, 4°C, 3 min) and after the last washing step adjusted to 1 mL with 0.1 M NaAc (pH 5). Washed Adipic acid dihydrazide-Agarose beads were added given to the RNA precipitate and bound ON at 4°C. Unbound RNA was removed by two washing steps each with 1 mL 2 M NaCl (800 rpm, 2 min, Eppendorf microcentrifuge). Bound RNA was adjusted to the nuclear salt concentration by washing three times with 1 mL buffer D each [20 mM HEPES-KOH, pH 7.6, 5% (v/v) glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol (DTT)]. HeLa cell nuclear extract (Cilbiotech s.a., Belgium) was diluted with buffer D and rotated with the RNA-coupled agarose beads for 20-30 min at 30°C or, alternatively, at 37°C. The activity of phosphatases was inhibited by including Phosphatase inhibitor cocktails 1 and 2 (Sigma) in a concentration of 1 µL/500 µg nuclear protein. Unbound proteins were removed from the reaction by washing five times with 1 mL buffer D each containing 4mM MgCl<sub>2</sub> (800 rpm, 2 min, Eppendorf micro centrifuge). After final washing, 12.5-60 µL 2 x protein sample buffer [0.75 M Tris-HCl, pH 6.8, 20% (v/v) glycerol, 10% (v/v) β-mercapto-ethanol, 4% (w/v) SDS] were added to the bead pellet depending on the amount of input RNA. Proteins were dissociated from the RNA by incubating at 95°C for 10 min. Agarose beads were pelleted by centrifugation, the supernatant transferred to another ProteinLoBind reaction tube (Eppendorf) and stored at -20°C until protein analyses.

### **B.2.8.3 *In vitro* dephosphorylation of HeLa nuclear extract**

Proteins of HeLa nuclear extracts [Cilbiotech s.a., Belgium; prepared as described by Dignam et al. (153)] were dephosphorylated by Calf Intestinal Alkaline Phosphatase (CIP) treatment. For the nuclear extract a protein concentration of 7.3 µg/µL was quantified by Bradford-assay (B.2.5.2). To determine the CIP activity needed to efficiently dephosphorylate HeLa nuclear

extract proteins, dephosphorylation was performed in a preliminary experiment using either 18 U or 36 U CIP (Promega) per mg HeLa nuclear proteins in 250  $\mu$ L reactions containing buffer conditions as indicated by the phosphatase supplier. To assess the degree of dephosphorylation achieved by phosphatase treatment, Ser/Thr-phosphatases were inhibited in a control reaction by addition of  $\beta$ -glycerophosphate to the reaction in a final concentration of 20 mM. Reactions were incubated at 37°C for 30 min. Control reactions were incubated in parallel in the absence of CIP and  $\beta$ -glycerophosphate to confirm that dephosphorylation was performed by CIP. Dephosphorylation of nuclear proteins was analysed by immunoblotting and detection with antibodies against a common phospho-epitope of SR proteins (Zymed; clone 1H4), SF2/ASF (Zymed; clone 96) and SRp40 (U.S. Biological) (Fig. II-3). In addition, the effect



**Fig. II-3: Immunoblot analysis of HeLa nuclear proteins dephosphorylated *in vitro* applying different amounts of CIP activity.**

CIP- or mock-treated HeLa nuclear proteins were separated in 10% SDS-polyacrylamide gels and after blotting probed with indicated primary antibodies recognising phosphorylated (SR-P, SRp40 and hTra2- $\beta$ ) or non-phosphorylated epitopes (SF2/ASF, U2AF<sup>65</sup> and U2AF<sup>35</sup>).

of CIP-mediated dephosphorylation on U2AF<sup>65</sup> and U2AF<sup>35</sup> was analysed using a monoclonal antibody recognising U2AF<sup>65</sup> (generously provided by Prof. M. Hastings and Prof. A. Krainer, Cold Spring Harbor Laboratory, USA) and a polyclonal antibody recognising U2AF<sup>35</sup> (ptglab). Immunoblotting analysis revealed that dephosphorylation is incomplete using 18 U CIP/mg HeLa nuclear extract. Therefore, for RNA affinity chromatography HeLa nuclear extracts were dephosphorylated using 36 U CIP/mg HeLa nuclear proteins.

#### **B.2.8.4 2D gel electrophoresis**

For each target sequence triplicate reactions containing 2000 pmol RNA each were subjected to RNA affinity chromatography as described (B.2.8.2). Proteins were eluted from the RNA by rotating each agarose bead pellet after the final wash with 125  $\mu$ L 66% glacial acetic acid and 33 mM MgCl<sub>2</sub> for 30 min up to 1 h (173). Agarose beads were centrifuged at 800 rpm for 2 min and the supernatant transferred into another ProteinLoBind tube (Eppendorf). Bead pellets were again rotated with 125  $\mu$ L elution mix for 30 min up to 1 h and the supernatant after centrifugation (13.200 rpm, 2 min) combined with the first supernatant. Proteins contained in the supernatant were precipitated with 5 vol. acetone at -20°C ON. After centrifugation (12.000 rpm, 30 min, 4°C), proteins were washed five times with acetone:ddH<sub>2</sub>O (5:1) and air-dried. Protein pellets were stored at -20°C until isoelectric focusing.

For isoelectric focusing triplicate protein pellets were solubilised in 42  $\mu$ L thiourea/urea sample buffer [7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 2.5% (w/v) DTT, 0.5% (v/v) IPG buffer (pH 3-10), 0.002% (w/v) bromophenol blue] each by rocking at RT for 2 h. Triplicate samples were pooled after a short centrifugation resulting in a sample volume of 126  $\mu$ L required to rehydrate 7 cm IPG-strips. Proteins samples were passively loaded on immobilised pH gradients (IPG-strip pH 3-10, Amersham) by rehydration of the dried gel matrix with the protein solution ON. Proteins were focused in the pH gradients using an Ettan III IPGphor (Amersham) using a 250 V step for 1 h, a 4.000 V gradient for 1 h and a 4.000 V step for 7.5 h. Focused protein containing IPG strips were stored at -20°C at least ON.

Separation in the second dimension was performed by SDS-PAGE using a polyacrylamide concentration of 10%. SDS was removed from the gels by washing three times in dH<sub>2</sub>O. Proteins were stained using a modified coomassie brilliant blue staining solution ON (299). 2D-gels were digitalised using a Hewlett Packard scanner directed by LabScan software (Amersham). Protein spots were detected and analysed using ImageMaster 2D Platinum 6.0 software (GE Healthcare)

## **B.2.9 Protein sequencing by mass spectrometry**

### **B.2.9.1 In gel digestion and sample preparation**

Bands containing proteins to be identified were cut from the SDS polyacrylamide gels, cutted into approximately 1 mm<sup>3</sup> pieces and transferred into a 0.5 mL reaction tube (Protein-Low-Bind reaction tube, Eppendorf). To remove salts, which could interfere with peptide ionisation, gel pieces were agitated four times in 100 µL freshly prepared 25 mM ammonium hydrogen carbonate buffer/50% acetonitrile each, first for 10 min and then three times for 30 min at RT. Gel pieces were completely dehydrated by incubation in acetonitrile (100%) for 30 min and after removal of the acetonitrile dried in a vacuum centrifuge (DNA110 SpeedVac®, Thermo Scientific). Gel pieces were rehydrated in trypsin solution (0.1 µg/µL [Sigma] in 25 mM ammonium carbonate buffer, pH 8), excessive trypsin solution removed and overlaid with 25 mM ammonium carbonate buffer. Proteins were in gel digested for 12-16 h at 37°C.

The supernatant of the in gel digestion was collected in a second reaction tube (ProteinLoBind reaction tube, Eppendorf). Gel pieces were rocked in 2 vol. ddH<sub>2</sub>O for 5 min. After sonication for 5 min the supernatant was removed and combined with the supernatant extracted before. Afterwards gel pieces were three times agitated in 1 vol. elution buffer (50% acetonitrile, 5% formic acid) each for 30 min at RT and all supernatants pooled with the supernatants collected before. In the final elution gel pieces were agitated with 1 vol. acetonitrile (100%) for 30 min and the supernatant was also combined with the protein supernatant eluted before. Eluted proteins were lyophilised in a vacuum centrifuge (DNA110 SpeedVac®, Thermo Scientific) and stored at -20°C until mass spectrometry analyses.

### **B.2.9.2 Mass spectrometry**

Eluted proteins were dissolved in 5 µL 4% methanol/1% formic acid, desalted and concentrated by ZipTip<sub>C18</sub> reversed-phase purification (Millipore). The C18-resin of the ZipTip<sub>C18</sub> pipette tip was wetted three times with 60% methanol/1% formic acid and equilibrated three times with 4% methanol/1% formic acid. Proteins were loaded on the ZipTip resin by 10 x aspirating and dispensing the sample. Proteins bound to the C18-resin were washed four times with a total volume of 30 µL 4% methanol/1% formic acid and eluted in 5 µL 60% methanol/1% formic acid. Mass spectrometry was performed by Dr. W. Bouschen and Dr. N. Wiethölter using an ESI-Quadrupole-TOF (QSTAR XL; Applied Biosystems) at the Analytical Core Facility of the Biological-Medical Research Centre (BMFZ, HHUD).

---

## **B.2.10 Microfluidics analyses using Lab-on-a-chip assays**

### **B.2.10.1 Protein analyses**

Prior to preparative separation proteins isolated by RNA affinity chromatography (see B.2.8) were analysed by capillary gel electrophoresis using the Protein 230 Chip Assay (Agilent). The capillaries of the analysis chips were loaded with gel matrix and the protein ladder as described by the manufacturer. For each protein sample 4  $\mu\text{L}$  were denatured at 95°C for 5 min using 2  $\mu\text{L}$  of DTT-containing sample buffer. Denatured proteins were diluted in ddH<sub>2</sub>O to a final volume of 90  $\mu\text{L}$ . Triplicates of 6  $\mu\text{L}$  each were analysed for each protein sample on the Bioanalyzer 2100 system (Agilent). Data acquisition and analyses was performed using Bioanalyzer 2100 Expert Software (ver. B.02.05.SI360) applying standard running conditions provided for the Protein 230 chip assay.

### **B.2.10.2 RNA analyses**

Routinely, the quality of RNA samples used for RT-PCR analyses was analysed by native capillary gel electrophoresis using the RNA 6000 Nano Chip Assay (Agilent). Prior to chip loading 3  $\mu\text{L}$  of each column-purified RNA sample were denatured at 70°C for 2 min and immediately cooled down on ice to avoid secondary structure formation. 1  $\mu\text{L}$  of each RNA sample was loaded on the chip prepared according to the manufacturer's protocol and separated using the Bioanalyzer 2100 system (Agilent). Data acquisition was performed using Bioanalyzer 2100 Expert Software (ver. B.02.05.SI360) applying standard running conditions provided for the RNA 6000 Nano chip assay. An algorithm implemented into the Bioanalyzer Software classifies each sample based on the complete electrophoretic trace data by calculating an RNA integrity number (RIN) ranging from 1 to 10, with 1 representing the most degraded and 10 representing the most intact RNA. Samples yielding RIN-values of at least 8.5 were considered as high-quality RNA and used for RT-PCR analyses.

## C. Results

Efficient replication of HIV-1 depends on temporally changes in the expression of distinct viral protein subsets during the early and late phase of viral gene expression. Since the viral genome is transcribed as a single pre-mRNA, all viral proteins are encoded by the same pre-mRNA with partly overlapping open reading frames (ORFs). The chronological expression of distinct protein subsets originates from changes in the splicing pattern of this viral pre-mRNA.

During early gene expression extensive splicing of the viral pre-mRNA generates several transcript isoforms encoding the regulatory proteins Rev and Nef by including one of the alternative exons 4c, 4a, 4b or 5 into the mRNA. These exons differ in the 3' ss used for their generation but share usage of the downstream located 5' ss D4. Activation of the splice sites flanking these exons, however, requires splicing enhancer elements within its shortest isoform, i.e. exon 5 (295, 296). Research from our group analysing the splice site-dependent Rev-mediated HIV-1 glycoprotein expression from a subgenomic 1-intron-*env*-expression construct identified the GAR splicing enhancer (guanosine-adenosine-rich exonic splicing enhancer; GAR ESE) in exon 5 supporting U1 snRNA binding at the downstream 5' ss D4 (295). A follow-up study including results of this thesis revealed that the GAR ESE also activates an upstream located heterologous 3' ss in *in vitro* splicing assays (91). The first part of this thesis characterises the effect of the GAR enhancer on the inclusion of the internal HIV-1 exons in the presence of their authentic upstream 3' ss as well as their downstream 5' ss. Inclusion of the internal exons was investigated *in vivo* in a transient cell culture transfection system. To dissect the mechanism of the splicing enhancer function *in vitro* RNA-protein binding studies were employed.

During late viral gene expression the mRNA pool is shifted towards isoforms with increasing intron content due to progressive changes in the splicing pattern of the primary transcript. Finally, disuse of all viral pre-mRNA splice sites results in the expression of unspliced RNA, which serves both, as template for translation of structural proteins and enzymes and also as viral genome, which is encapsidated into budding virus particles. Since inactivating internal 3' ss is not sufficient for the expression of unspliced RNA – as revealed in the first part of this thesis – the hypothesis was derived that splicing of introns upstream of exon 5 might be regulated by the efficiency of 5' ss

---

D1. Therefore, in the second part of this thesis it was investigated whether splicing at D1 is also effected by regulatory elements surrounding D1.

This thesis has in part been published in peer reviewed journals (Freund et al. 2003; Caputi et al. 2004, Asang et al. 2008) and presented at symposia (s. G.3 Publications).

### C.1 Splice site activation of the alternative exons 4c, 4a, 4b, and 5 during early and intermediate HIV-1 gene expression

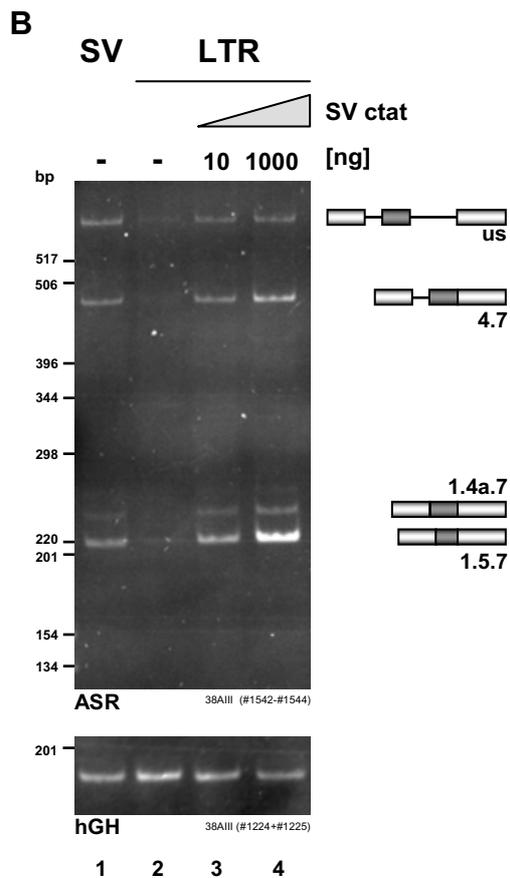
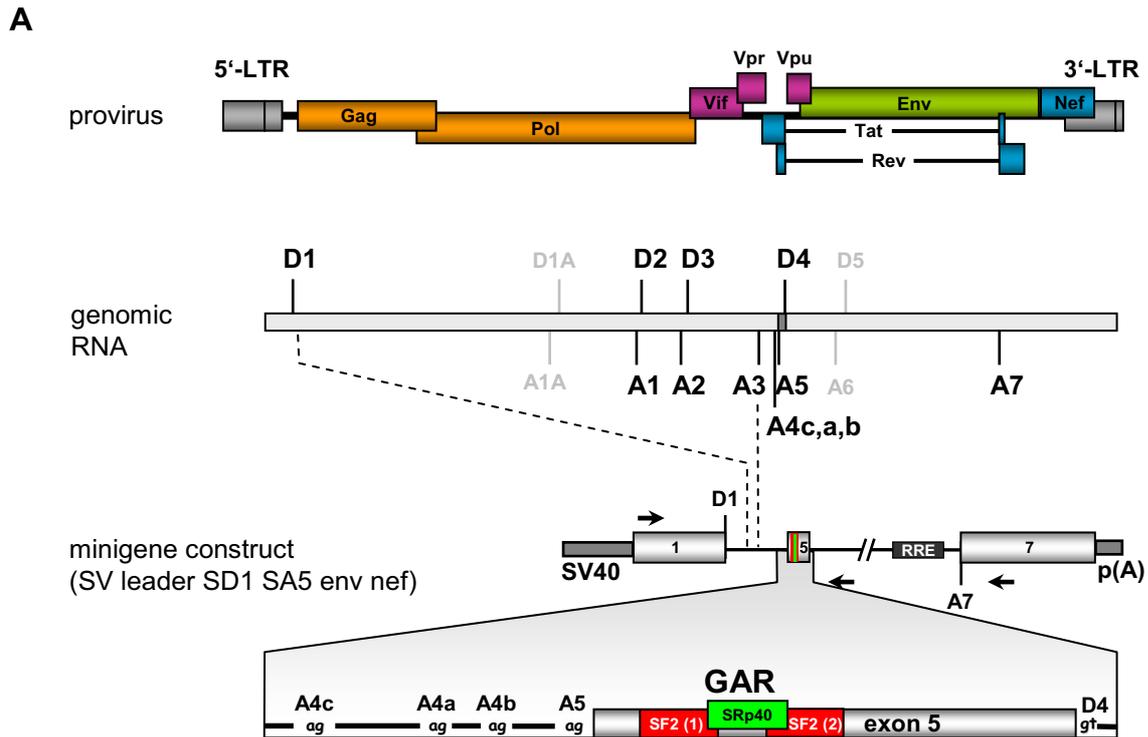
Previous research in our group identified a purine-rich exonic splicing enhancer (Guanosine-Adenosine-rich exonic splicing enhancer, GAR ESE) in the 5' region of HIV-1 exon 5 (295). Because at that time splice site activation by the enhancer had been tested only in the presence of either a single 3' ss or 5' ss (91), the first part of this thesis examined whether the GAR ESE could simultaneously activate both, the 3' ss and 5' ss flanking the GAR containing exons.

In order to analyse the impact of the GAR enhancer on splicing of the central alternatively used exons 4c, 4a, 4b and 5 of the HIV-1 pre-mRNA in a transient transfection model *in vivo*, a subgenomic minigene was constructed that contains the GAR ESE encoding exon 5 as internal exon and also the upstream located 3' ss A4c, A4a and A4b (SV leader SD1 SA5 env nef, Fig. III-1A).

To allow simultaneous detection of all differentially spliced mRNA isoforms in a multiplex RT-PCR assay, the 5' region of the proviral genome of the viral isolate NL4-3 was shortened by deletion of the 5'-half of the genome encoding the structural proteins and enzymes (Fig. III-1A). Thereby, 5' ss D2 and D3 as well as 3' ss A1, A2, and A3 were removed, so that the minigene with the resulting splice sites constitutes an HIV-1 3-exon-2-intron splicing reporter. To rule out any side-effects on transcription and splicing of the reporter construct due to the autoregulatory functions of the viral proteins

#### **Fig. III-1: Generation of a 3-exon-2-intron minigene construct to investigate alternative splicing regulation of the HIV-1 pre-mRNA.**

(A) The alternative splicing reporter minigene SV leader SD1 SA5 env nef was generated by deletion of *gag-pol*-, *vif*- and *vpr*-encoding sequences in the 5' region of the proviral genome of the laboratory strain NL4-3 (upper panel). The deletion simultaneously eliminates a number of splice sites (middle panel) thereby generating the 3-exon-2-intron architecture in the minigene (lower panel). The HIV-1 LTR promoter in the alternative splicing reporter minigene SV leader SD1 SA5 env nef was replaced by the Tat-independent SV40 early promoter (SV40). The previously identified GAR enhancer is situated in the centrally located exon 5 of the alternative splicing reporter construct. Positions of primers used for RT-PCR are indicated by arrows. The enlargement of the internal exon 5 illustrates the position of the GAR ESE in the 5'-half of the exon preceded by the four alternatively used 3' ss 4c, 4a, 4b and 5 [LTR: Long Terminal Repeat, SV40: Simian Virus 40, GAR: guanosine-adenosine-rich enhancer, RRE: Rev-responsive element].



**Fig. III-1: continued.**

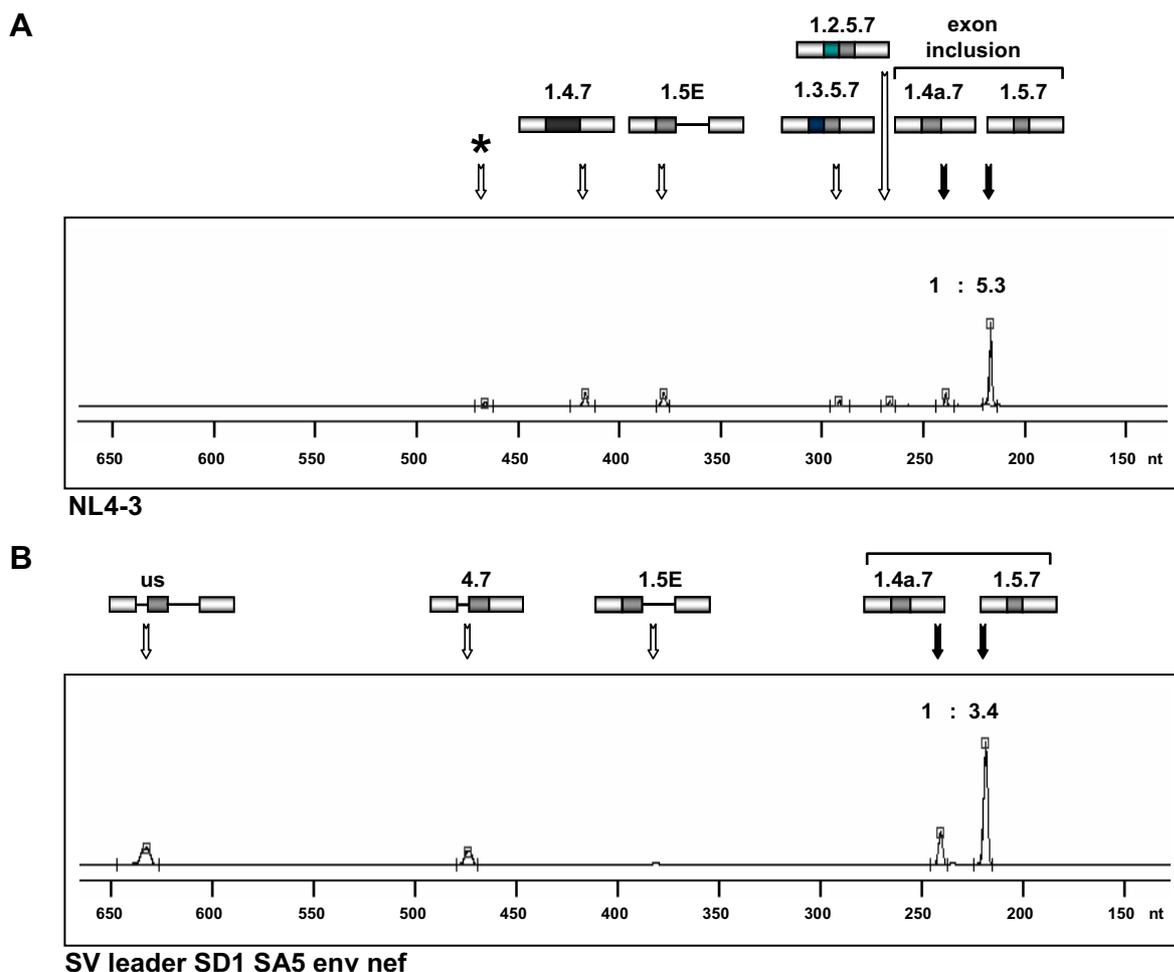
(B) RT-PCR analysis of HeLa-T4<sup>+</sup> cells transfected with the 2-intron minigene containing either the HIV-1 LTR or the SV40 early promoter. 2 x 10<sup>5</sup> cells were transiently transfected with 1 µg SV leader SD1 SA5 env nef (SV, lane 1) or 1 µg LTR SD1 SA5 env nef (LTR, lanes 2-4). Cells transfected with the LTR-minigene were cotransfected with 10 ng or 1000 ng SVctat expressing HIV-1 Tat (lanes 3 and 4). All cells were additionally cotransfected with 1 µg pXGH5 expressing the *hGH*-mRNA to monitor equal transfection efficiency. Transfected DNA amounts were equalised using pSP73. Total RNA was isolated 30 h post transfection. mRNA was reverse transcribed using oligo(dT) as primer. PCR was performed in the linear amplification range using primers #1542-#1544 for amplification of alternatively spliced reporter RNA and primers #1224 and #1225 for amplification of *hGH*-mRNA. PCR products were resolved on 6% polyacrylamide gels and stained with ethidium bromide. The exon/intron structure of alternatively spliced mRNA isoforms is depicted on the right [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, us: unspliced].

Tat and Rev, which are encoded within the 3' region of the viral genome and thus within this splicing reporter, their translational start codons were inactivated by point mutations. Efficient transcription by the HIV-1 promoter, however, requires *trans*-activation by Tat. Since it was intended to develop a reporter construct, whose expression is independent of cotransfected expression plasmids, it was evaluated whether the HIV-1 LTR promoter could be substituted with a Tat-independent promoter without changing the minigene splicing pattern. Therefore, two minigene constructs were generated, which are controlled either by the Tat-dependent HIV-1 LTR promoter or by the Tat-independent promoter of the early transcriptional unit of Simian Virus 40 (SV40 early promoter). In addition, the polyadenylation signal in the non-coding region of the 3'-LTR was substituted with the early polyadenylation signal originating from SV40. The minigene construct and mutants thereof were used for transient transfections of eukaryotic cells under non-replicative conditions.

Since it has been described that promoter strength can affect splice site selection [(292), reviewed in (322)], the splicing pattern of the minigene construct under control of the SV40 promoter was compared to that of the minigene construct driven by the LTR promoter. As expected, in the absence of Tat only marginal expression of LTR-driven minigene transcripts was observed (Fig. III-1B, lane 2). Cotransfection with 10 ng Tat expression vector, however, induced expression of the LTR-controlled minigene construct (Fig. III-1B, lane 3). Besides a low amount of unspliced RNA (us) transcript isoforms with only the proximal intron spliced (4.7) and the proximal and distal intron spliced (1.4a.7/1.5.7) were observed. For inclusion of the internal exons 3' ss 4a and 5 were used. RNA expression from the minigene construct could further be increased by raising the amount of Tat expression vector to 1000 ng (Fig. III-1B, cf. lane 3 and 4). The splicing pattern of the minigene under transcriptional control of the SV40 promoter was identical to that of the LTR-controlled minigene in the presence of 10 ng Tat expression vector (Fig. III-1B, lane 1). Therefore, the minigene construct under control of the SV40 promoter constitutes a reliable reporter to analyse processing events of HIV-1 primary transcripts.

It was reported that from 3' ss A4c, A4a, A4b and A5, clustered upstream of the exonic GAR ESE, A5 is most frequently used for generation of viral mRNAs (232, 481, 495). To examine, if the reporter construct resembles the unbalanced ratio of 3' ss activation of full-length HIV-1 pre-mRNA during infection, the splicing pattern of the transfected minigene construct was compared to that of NL4-3 infected PM1 cells by RT-PCR.

Since additional splice sites in the 5' region of the viral genome, lacking in the minigene construct, caused a more complex splicing pattern of the pre-mRNA than in the minigene construct, RT-PCR products were labelled with the fluorochrome Cy5, resolved under denaturing conditions and detected by automated laser fluorescence (ALF). Analysing the splicing pattern of viral mRNA isoforms in infected PM1 cells confirmed that activation of A5 is preferred for inclusion of the internal exons into mRNA, whereas A4a is less frequently used (Fig. III-2A).



**Fig. III-2: The 3-exon-2-intron minigene includes the internal exons 4a and 5 in transient transfection experiments like viral pre-mRNA during infection.**

RT-PCR analysis of PM1 cells infected with the viral isolate NL4-3 **(A)** compared to the mRNA expression pattern of HeLa-T4<sup>+</sup> cells transiently transfected with the 2-intron minigene construct SV leader SD1 SA5 env nef **(B)**. Cy5-labeled RT-PCR products were separated on denaturing urea-polyacrylamide gels and detected by Automated Laser Fluorescence (ALF). RT-PCR products are shown as processed fluorescence curve data of the electrophoretic separation. Exon junctions of spliced mRNAs are indicated by numbering above the respective lanes. The asterisk marks a low amount RNA signal, which was identified as unspecific signal by sequencing. The ratio of exon 5 to 4a inclusion in mRNAs with single internal exon recognition is indicated above the respective peaks.

Usage of 3' splice sites A4c and A4b was not observed under these experimental settings. In addition to RNA isoforms containing only one internal exon, RNA isoforms additionally including either the non-coding leader exon 2 or 3 as second exon were observed. Since both exons had been deleted in the minigene construct, these RNA isoforms were absent in the splicing pattern of cells transfected with the minigene (Fig. III-2B). Due to the position of the RT-PCR primers unspliced RNA in virus infected cells generates a considerably longer RT-PCR product than unspliced minigene RNA. As a result of its length it is discriminated in the RT-PCR amplification and therefore not displayed in this experimental setting. Nevertheless, it was ascertained, that the preferred usage of 3' splice site A5 for inclusion of the internal exons compared to usage of A4a matched between transcripts expressed from the minigene or the viral genome. Thus, the minigene reflects viral splice site usage and depicts a model for splice site regulation of the internal exons 4c, 4a, 4b and 5.

### **C.1.1 Characterisation of the GAR ESE within exon 5**

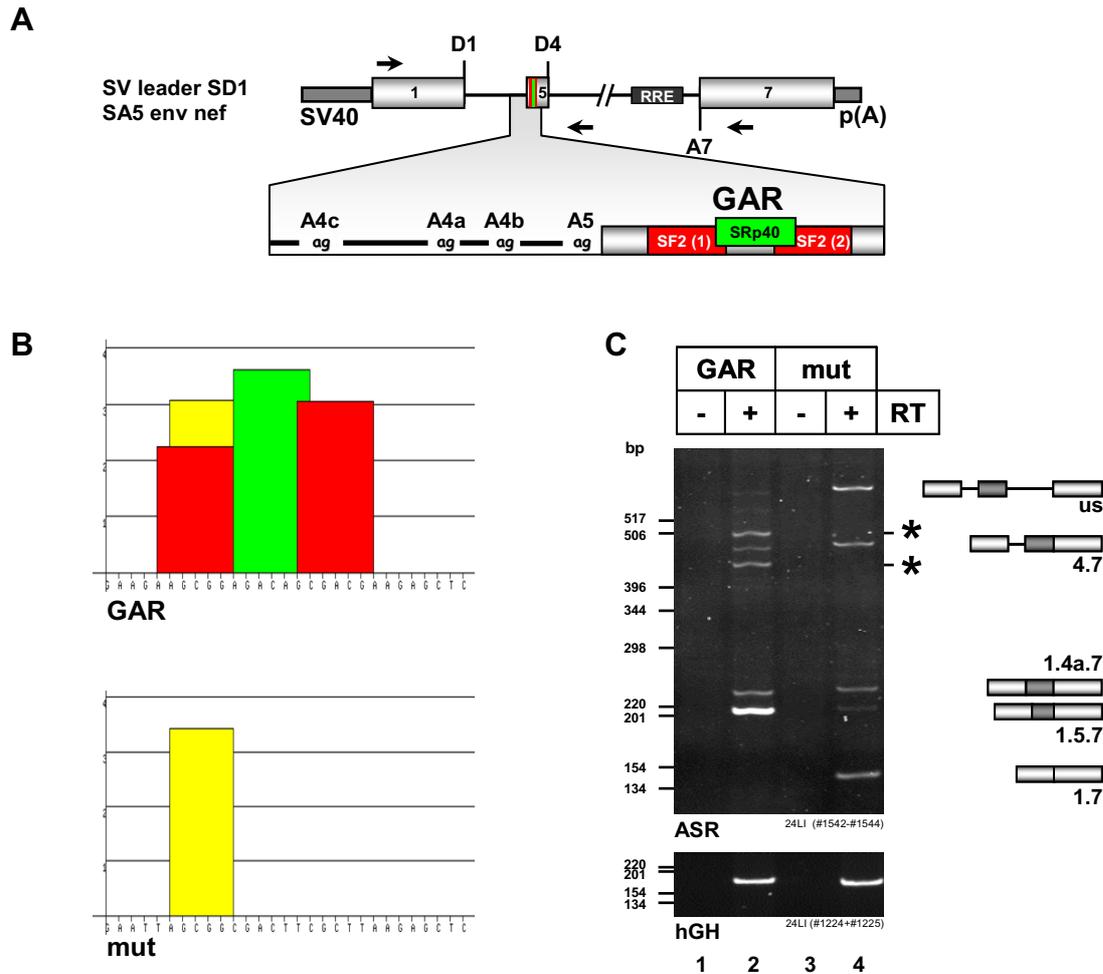
#### **C.1.1.1 The GAR enhancer is essential for inclusion of exon 5 into early mRNA isoforms.**

In the GAR ESE sequence two binding sites for the SR proteins SF2/ASF and one for SRp40 and SRp55, respectively, were predicted using an algorithm based on experimental SELEX data (ESEfinder, ver. 2.0; [<http://rulai.cshl.edu/tools/ESE2/> (97)]) (Fig. III-3A and B). Purified SF2/ASF2 as well as SRp40 were able to rescue *in vitro* splicing of a GAR ESE-dependent heterologous *dsx*-substrate in splicing-deficient HeLa S100 extract, whereas the enhancer did not respond to SRp55 (91). To investigate, which function the GAR ESE exerts in the recognition of the internal exon of the 3-exon-2-intron minigene, a combination of point mutations was inserted into the GAR ESE of

**Fig. III-3: The GAR ESE is essential for inclusion of the central exon 5 into the mRNA expressed from the 3-exon-2-intron minigene construct.**

---

(A) Scheme of the alternative splicing reporter SV leader SD1 SA5 env nef. The internal exon 5 is enlarged to illustrate the location of the predicted binding sites for the SR proteins SF2/ASF and SRp40. A cluster of four 3' splice sites precedes the internal exon allowing the alternative generation of upstream extended exon isoforms 4c, 4a, and 4b. Positions of primers used in RT-PCR are indicated by arrows.



**Fig. III-3: continued.**

**(B)** The ESEfinder algorithm predicts two SF2/ASF, an SRp40, and an SRp55 binding site in the 5'-half of exon 5 (GAR ESE, upper panel). Since SRp55 could not rescue splicing in an *in vitro* splicing assay, only binding sites for SF2/ASF and SRp40 were analysed regarding their influence on splicing of the alternative splicing reporter. Inactivation of the SR proteins binding sites generates the ESE- mutant (mut, lower panel).

**(C)** RT-PCR assay of total RNA isolated from HeLa-T4<sup>+</sup> cells transiently transfected with 1  $\mu$ g of the parental reporter construct SV leader SD1 SA5 env nef (GAR; lanes 1-2) or the mutant SV leader SD1 SA5 SF2(1)- SF2(2)- SRp40- env nef (mut; lanes 3-4). Cells were cotransfected with 1  $\mu$ g pXGH5 to control equal transfection efficiency (hGH) and RNA isolated 30 h post transfection. DNase I-treated RNA was subjected to 2-step RT-PCR using two antisense primers to allow amplification of cDNA from unspliced and spliced RNAs in a multiplex PCR assay. The exponential phase of amplification was determined in a pioneer experiment. The exon/intron structure of alternative mRNA isoforms is indicated at the right. Asterisks mark heteroduplex PCR products as revealed by DNA sequencing [ASR: alternatively spliced reporter RNA; hGH: human Growth Hormone, RT: reverse transcription, us: unspliced].

the reporter construct predicted to specifically inactivate the putative binding sites for SF2/ASF and SRp40. HeLa-T4<sup>+</sup> cells were transiently transfected with the respective minigene constructs and the splicing pattern was analysed by multiplex RT-PCR of total RNA.

Inactivation of the predicted binding sites for SF2/ASF and SRp40 strongly reduced the recognition of the internal exon 5 (1.5.7) (Fig. III-3C, lane 4). In addition, an mRNA isoform was observed, in which the internal exon was skipped (1.7). Concomitantly with the changes in exon recognition expression of completely unspliced RNA (us) and RNA spliced only at the distal intron (4.7) increased.

Since mutating the GAR ESE decreases the activity of the flanking splice sites thereby increasing the expression of alternatively spliced mRNAs as well as unspliced mRNA, the enhancer is essential for efficient recognition of the centrally located exon 5.

To confirm that SF2/ASF binding to the GAR ESE was indeed reduced due to the mutation introduced into the minigene construct, binding of SF2/ASF to the NL4-3 GAR ESE and mutant enhancer sequences was analysed. *In vitro* transcribed RNA was used to capture GAR ESE bound SF2/ASF, which was subsequently detected by immunoblotting. SF2/ASF bound strongly to the GAR ESE RNA sequence, whereas it bound in a considerably lower amount to a control sequence derived from the viral genome (HIV#18) (Fig. III-4B, cf. lane 1 and 2), which has previously been shown to be unable to activate the downstream located 5' ss D4 (91). Mutation of the GAR ESE's SRp40 binding site reduced SF2/ASF binding only slightly, whereas mutation of both SF2/ASF binding sites strongly decreased SF2/ASF binding. Nevertheless, only combining SF2/ASF and SRp40 binding site mutations reduced SF2/ASF binding to the level of the control sequence. The presence of U1-70K, a protein component of the U1 snRNP, was only detected using the GAR ESE sequence and was unaffected by mutation of the SRp40 binding site, whereas it was nearly abrogated by mutation of the two SF2/ASF or all three analysed SR protein binding sites (Fig. III-4B, middle panel, cf. lanes 1 and 3 with lanes 4 and 5). Both RNA sequences that efficiently bound U1-70K showed the strongest SF2/ASF binding suggesting a threshold level of SF2/ASF to be necessary for U1 snRNP binding to the 5' ss.

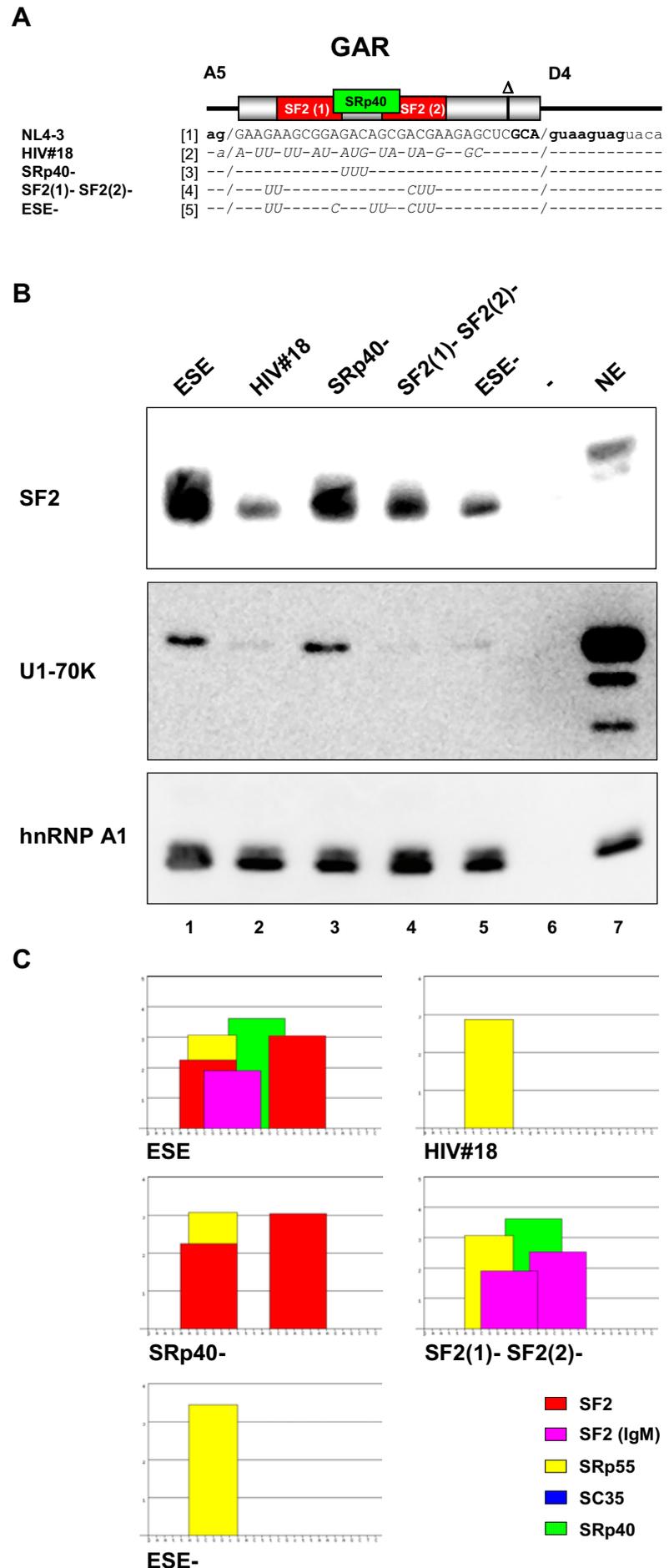
Although binding of SF2/ASF to the GAR ESE was reduced after mutations of the enhancer sequence, the relative amounts of SF2/ASF binding *in vitro* did not exactly

**Fig. III-4: Mutations of the GAR ESE reduce binding of SF2/ASF and U1-70K.**

**(A)** RNA sequences employed in RNA affinity chromatography. *In vitro* transcribed RNAs contained the AG-dinucleotide of the upstream intron (lower case letters) followed by the exon 5 sequence (capital letters) harbouring the GAR ESE. The 3' half of exon 5 was deleted up to 5' ss D4 ( $\Delta$ ). The HIV#18 template served as control sequence, which had been shown not to act as splicing enhancer (91). Either the predicted SRp40 binding site (SRp40-), the SF2/ASF binding sites (2x SF2-) or binding sites for both proteins (ESE-) were mutated in the RNA target sequences (italic letters).

**(B)** RNA affinity chromatography of RNA targets depicted in (A). 1000 pmol of respective RNA targets were immobilised and incubated in HeLa nuclear extract. Bound proteins were analysed by immunoblotting using antibodies against SF2 (AK-96, kindly provided by Prof. M. Caputi), U1-70K (sc-9571) and hnRNP A1 (9H10). A control binding reaction was performed lacking target RNA (-). Protein input was controlled by analysing HeLa nuclear extract without further treatment (NE).

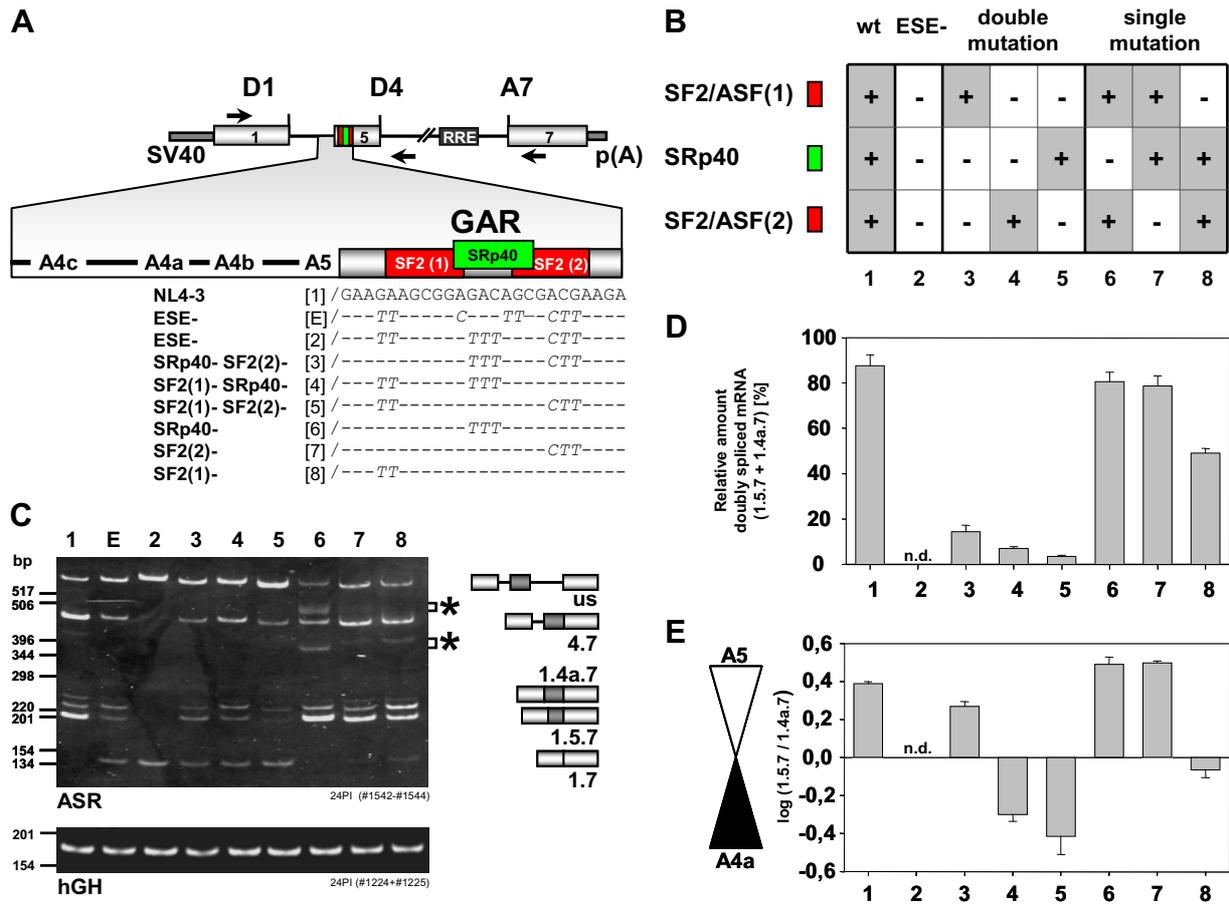
**(C)** Binding sites for SR proteins predicted by the ESEfinder algorithm in the GAR ESE sequence and mutations thereof.



correspond with the predictions derived from the ESEfinder algorithm. Therefore, the GAR ESE sequence and mutations used to examine SF2/ASF binding were reanalysed regarding their predicted SR protein binding sites by using a more recent version of the ESEfinder algorithm containing an additional score matrix (IgM-BRCA1) for the prediction of SF2/ASF binding sites [ESEfinder, ver. 3.0 (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home>) (97, 566)]. Analysis of the GAR ESE with the improved ESEfinder algorithm predicted an additional SF2/ASF binding site overlapping the proximal SF2/ASF binding site [SF2(1)] (Fig. III-4C). Assuming that overlapping binding sites can not be occupied at the same time, but increase the probability of SR protein binding at the respective sequence, experimentally assessed SF2/ASF binding to the GAR ESE and mutations thereof revealed a higher correlation to the predictions derived from the updated version of the ESEfinder.

#### **C.1.1.2 Individual SR protein binding sites of the GAR ESE synergistically increase recognition of the alternative exons 4a and 5.**

Although purified SF2/ASF2 as well as SRp40 rescued splicing of a GAR ESE-dependent heterologous 3' ss in splicing-deficient HeLa S100 extract *in vitro*, mutation of the putative SRp40 binding site in previous *in vitro* as well as *in vivo* experiments reduced the activation of 5' ss D4 only slightly (91). However, mutating both SF2/ASF binding sites strongly reduced activation of D4 as well as the heterologous 3' ss. Therefore, the contribution of the individual binding sites for the SR proteins SF2/ASF and SRp40 to the enhancer function in the 2-intron minigene was further examined by single and combinatorial mutations of the SR protein binding sites. The point mutations that were used in previous experiments to inactivate the SRp40 binding site in combination with mutations of both SF2/ASF binding sites of the GAR ESE could not be used for the single inactivation of the SRp40 binding site, since this mutation also inactivated both neighbouring predicted SF2/ASF binding sites. Therefore, another mutation specifically inactivating the SRp40 binding site was analysed. None of the mutations in the SR protein binding sites of the GAR ESE in the 2-intron minigene construct were predicted to generate additional SR protein binding sites according to the ESEfinder algorithm initially (ver. 2.0) used. The splicing pattern of the minigene constructs was again analysed after transient transfection of HeLa-T4<sup>+</sup> cells by multiplex RT-PCR (Fig. III-5).



**Fig. III-5: Binding sites of the GAR ESE for the SR proteins SF2/ASF and SRp40 synergistically activate exon recognition.**

**(A)** Schematic representation of GAR ESE mutations introduced into the HIV-1 2-intron minigene. Positions of RT-PCR primers are indicated by arrows. The enlargement illustrates the organisation of the alternative 3' ss cluster upstream of exon 5 and the mutations introduced in the GAR enhancer [RRE: Rev-responsive element, SV40: SV40 early promoter, p(A): SV40 polyadenylation signal].

**(B)** Summary of SR protein binding site mutations in the GAR ESE used in this assay.

**(C)** RT-PCR analysis of the parental minigene and GAR ESE mutants. HeLa-T4<sup>+</sup> cells were transiently transfected with the construct SV leader SD1 SA5 env nef or mutant derivatives [numbering according to (B)] and pXGH5. Total RNA was subjected to RT-PCR with the primers indicated in (A) (ASR) or with primers specific for *hGH*-mRNA as control for transfection efficiency (hGH). Asterisks mark heteroduplex PCR products as revealed by sequencing emerging due to non-denaturing separation conditions [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, us: unspliced].

**(D)** Quantification of internal exon recognition after SR protein binding site mutations. PCR products were Cy5-labelled, separated on denaturing gels and detected by ALF. Exon recognition was quantified as percentage of doubly spliced mRNA using either A4a or A5 relative to total mRNA (sum of spliced and unspliced mRNA isoforms) in each sample. Error bars indicate standard deviation from the mean of three independent experiments.

**(E)** Quantification of 3' ss usage in doubly spliced mRNA. Raw data resulting from the experiments shown in (D) were analysed regarding 3' ss selectivity depicted as ratio of 1.5.7- to 1.4a.7-mRNA.

Simultaneous mutation of all three predicted SR protein binding sites completely abrogated internal exon 4a and 5 recognition (Fig. III-5C, cf. lanes 2 and E). This result demonstrated, that the function of the GAR ESE extends beyond the sole backup of exon 5 recognition and thereby is likewise important for the inclusion of the internal exon 4a. The strongest effect after mutation of a single SR protein binding site was exerted by loss of the proximal SF2/ASF binding site [SF2(1)], reducing exon inclusion from 88% to 49% and concomitantly causing exon skipping (Fig. III-5C and D, lane 8). Mutating either the distal SF2/ASF [SF2(2)] or the SRp40 binding site only led to a slight decrease in exon recognition of about 10% (Fig. III-5C and D, lane 6 and 7). Combinatorial mutations of two of the SR proteins, however, dramatically reduced exon recognition to inclusion rates below 14% independent of which site had been mutated (Fig. III-5C and D, lanes 3-5). Together these results show that each of the GAR ESE SR protein binding sites analysed here contributes to exon recognition with the proximal SF2/ASF binding site having the highest impact on exon recognition. Furthermore, it can be concluded that any two of the SR protein binding sites synergistically mediate exon recognition (Fig. III-5C and D, cf. 3-5 with 6-8), whereas the presence of the third leads to an only additive effect on exon recognition (Fig. III-5C and D, cf. 6-8 with 1).

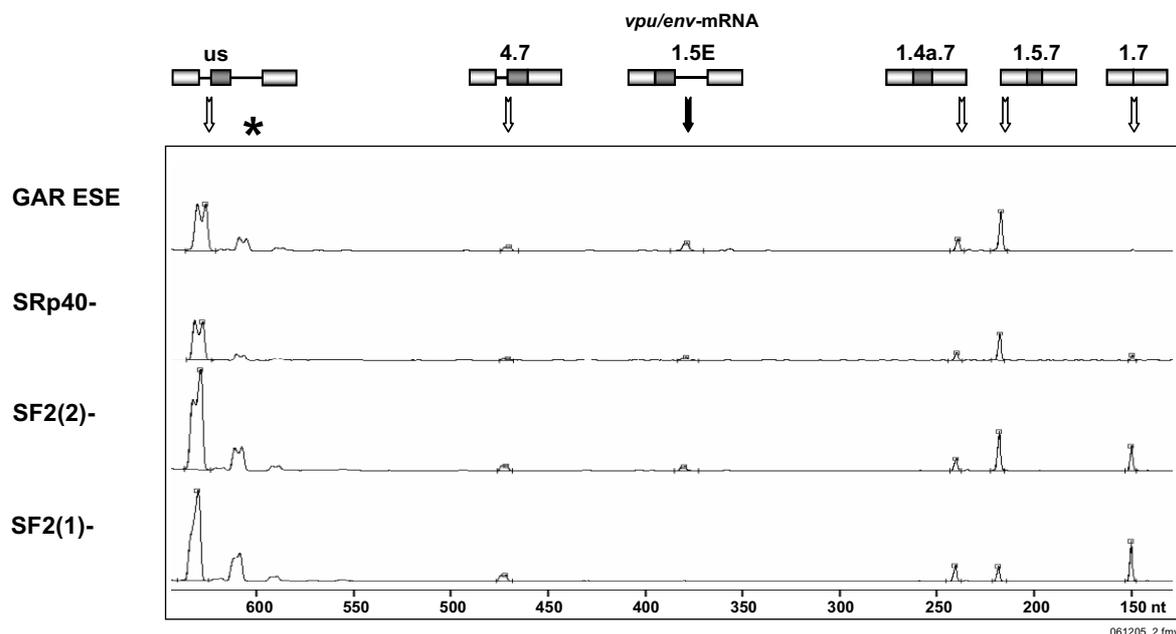
#### **C.1.1.3 The proximal SF2/ASF binding site specifically contributes to A5 usage.**

Analysing the ratio of exon 5 to exon 4a inclusion, a dominant influence of the GAR ESE's proximal SF2/ASF binding site [SF(1)] on the efficiency of exon 5 recognition was observed. Irrespective of the overall exon inclusion efficiency, mutation of SF2(1) led to a decrease in exon 5 and a concomitant increase in exon 4a recognition (Fig. III-5C and E, cf. lane 1 and 8). In contrast, mutation of the distal SF2/ASF or SRp40 binding site not only maintained the dominant usage of A5 but even slightly increased A5 over A4a usage suggesting that the distal SR protein binding sites might preferentially activate A4a (Fig. III-5C and E, cf. lane 1, 6 and 7). However, combining the mutations of the proximal SF2/ASF with one distal SR protein binding site augmented the preferential activation of A5 rather than causing an intermediate ratio of 3' ss activation (Fig. III-5C and E, cf. lane 4, 5 and 8). This indicates that besides its general enhancing role in internal exon recognition only the proximal SF2/ASF binding site confers selective activation of 3' ss A5. From these results it is concluded that the HIV-1 exon 5 GAR ESE fulfils a dual role in the regulation of alternative splicing by (i) ensuring exon

recognition of the alternative exons 5 and 4a and by (ii) determining the ratio of internal exon recognition in favour of exon 5.

#### C.1.1.4 All SR protein binding sites of the GAR ESE are required to efficiently activate A5 for the generation of intron-containing *vpu/env*-mRNAs.

While the early phase of HIV-1 gene expression is marked by complete intron removal from the viral pre-mRNA, in the intermediate phase splicing at 5' ss D4 is constrained by the interaction of the viral regulatory protein Rev with the RRE. This regulation results in the group of 4 kb mRNA isoforms differentially spliced in their 5' region but unspliced at the distal intron between 5' ss D4 and 3' ss A7. For these intron-containing mRNA isoforms an even more frequent usage of A5 compared to 4a was observed than in completely spliced 2 kb mRNAs (481). To investigate whether the GAR ESE also influences the ratio of internal 3' ss activation in the 4 kb mRNA class the splicing pattern of minigene constructs harbouring various mutations of the GAR ESE was analysed in the presence of Rev (Fig. III-6). Simultaneous mutation of all three SR



**Fig. III-6: All three SR protein binding sites of the GAR ESE are required to activate the 3' ss cluster for generation of intron-containing *vpu/env*-mRNA.**

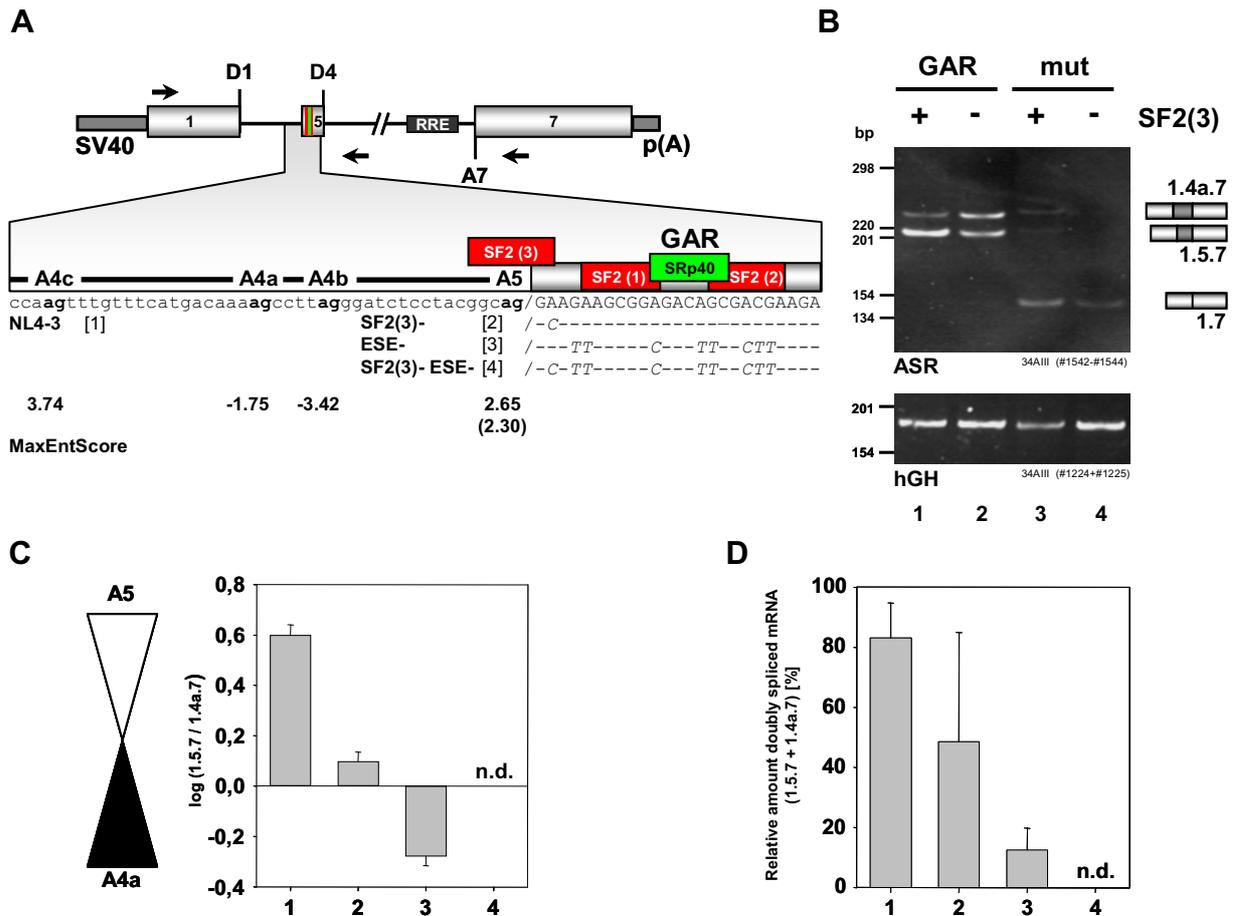
RT-PCR analysis of total RNA from HeLa-T4<sup>+</sup> cells transiently transfected with the parental 2-intron minigene or derivatives thereof carrying mutations of single SR protein binding sites (cf. Fig. 5A). Cells were cotransfected with SVcrev expressing the viral regulatory protein Rev and pXGH5. Cy5-labeled PCR products were separated and analysed by ALF as described in Fig. III-2. PCR products are shown as processed fluorescence curve data. The distinct isoforms of alternatively spliced mRNA are indicated at the top. The asterisk marks an unspecific signal identified by sequencing [us: unspliced].

protein binding sites of the GAR enhancer abolished expression of *vpu/env*-mRNA (data not shown). Since individual SR protein binding sites of the GAR ESE were found to contribute differentially to its enhancer function in completely spliced mRNA, this dependency was also examined in *vpu/env*-mRNAs with constructs carrying mutations in either one of the SR protein binding sites. Mutation of the SF2(2) or the SRp40 binding site reduced *vpu/env*-mRNA expression about two-fold [Fig. III-6, SRp40- and SF2(2)-], whereas mutating SF2(1) was already sufficient to eliminate its expression [Fig. III-6, SF2(1)-]. In contrast, the amount of unspliced mRNA remained unchanged for mutation of the SRp40 binding site and even increased after mutation of any SF2/ASF binding site like it had been observed after mutation of the GAR ESE in the absence of Rev (cf. C.1.1.1).

These results indicate that the individual SR protein binding sites of the GAR ESE maintain their relative impact on 3' ss activation also for processing of intron-containing *vpu/env*-mRNA. Moreover, GAR ESE dependency may be stronger than in completely spliced mRNA, since any single SR protein binding site mutation clearly reduced *vpu/env*-mRNA expression.

#### **C.1.1.5 A predicted third SF2/ASF binding site covering the intron/exon 5 border ensures preferential activation of A5.**

Extending the bioinformatical analyses of SR protein binding sites [ESEfinder, ver. 3.0 (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home>) (97, 566)] into the intron upstream of exon 5 predicted a third SF2/ASF binding site overlapping 3' ss A5 [SF2(3)] (Fig. III-7A). To investigate whether this putative site additionally contributes to 3' ss selection, SF2(3) was mutated at position +2 of exon 5 (A to C). Within the context of the wild-type GAR ESE this mutation abrogated the preferential selection of A5 as evidenced by an equal activation of A5 and A4a (Fig. III-7B and C, lane 2). Although this mutation was introduced at position +2 of the exon, which does not constitute a known functional element of the 3' ss, changes in that position are suggested to effect the intrinsic strength of the 3' ss A5 (MaxEntScore [[http://genes.mit.edu/burgelab/maxent/Xmaxentscan\\_scoreseq\\_acc.html](http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html)] (669))). To rule out that the decreased A5 usage is due to the slightly reduced intrinsic strength of this 3' ss (MaxEntScore decreases from 2.65 to 2.30), the influence of SF2(3) on the splicing pattern was analysed in presence of the GAR ESE mutation (ESE-). To this end, the originally tested triple GAR ESE



**Fig. III-7: An additional SF2/ASF binding site covering the intron/exon 5 border specifically activates A5.**

**(A)** An additional SF2/ASF binding site is predicted upstream of the GAR ESE by the ESEfinder algorithm (ver. 3.0) (97, 566).

**(B)** RT-PCR analysis of total RNA isolated from HeLa-T4<sup>+</sup> cells transfected with the parental 2-intron minigene or derivatives carrying SR protein binding site mutations. 2 x 10<sup>5</sup> cells were transiently transfected with 1 µg SV leader SD1 SA5 env nef (lane 1), SV leader SD1 SF2- SA5 env nef (lane 2), SV leader SD1 SA5 SF2(1)- SF2(2)- SRp40- env nef (lane 3) or SV leader SD1 SF2- SA5 ESE- env nef (lane 4) and 1 µg pXGH5 to monitor transfection efficiency. RNA isolation and RT-PCR was performed as described in Fig. III-1B. The exon/intron structure of alternatively spliced RNA isoforms is depicted at the right [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, us: unspliced].

**(C)** Quantification of 3' ss usage in doubly spliced mRNA. 3' ss selectivity is shown as ratio of doubly spliced 1.5.7- to 1.4a.7-mRNA. Error bars indicate standard deviation from the mean of three independent experiments [n.d.: not detected].

**(D)** Quantification of exon inclusion. Raw data resulting from the transfection experiments shown in (C) were analysed regarding the amount of doubly spliced mRNA relative to total reporter mRNA isoforms in each sample. Error bars represent standard deviation.

mutation was used that still facilitated residual exon recognition and led to an increased ratio of A4a to A5 usage (Fig. III-3, and Fig. III-7B, lane 3) allowing the detection of increasing as well as decreasing amounts of exon inclusion. If the reduced exon 5 inclusion was caused by the decreased intrinsic strength of A5, the additional mutation of SF2(3) had been expected to lead to an increased 3' ss 4a use. However, mutating SF2(3) in this context completely abolished exon recognition (Fig. III-7B and D, lane 4) indicating that SF2(3) is also part of the GAR ESE complex regulating splicing at the 3' ss cluster preceding exon 5 and fulfils a non-redundant function to the proximal SF2/ASF binding site of the GAR ESE.

#### **C.1.1.6 The GAR ESE is essential for activation of each of the competing 3' ss clustering upstream of exon 5.**

So far efficient activation of 3' ss A4c and A4b could not be detected in the 2-intron minigene, although their usage had already been described by others (232, 481, 493). Since it had been reported that inactivation of individual 3' ss of the cluster can lead to compensatory increases in the use of alternative 3' ss (481, 493, 594), it was rather unexpected that neither A4c nor A4b was activated in the minigene after reduced A4a and A5 usage due to mutation of the GAR ESE. This raised the question whether 3' ss A4c and A4b are also activated by the GAR ESE. If this were the case mutating the enhancer would inactivate these 3' ss as well. Alternatively, 3' ss A4c and A4b could be too weak to be used under this experimental settings and might require a more specific activation. To test whether in the minigene 3' ss A4c and A4b can be basically activated, individual or combinatorial mutations of the AG-dinucleotides of 3' ss A4a, A4b, and A5 were introduced into the minigene (Fig. III-8A). RT-PCR analysis of transiently transfected HeLa-T4<sup>+</sup> cells revealed that inactivation of 3' ss A5, which is predominantly used in the wild-type construct, led to a compensatory increase in 3' ss A4a use (Fig. III-8B, lane 2). Interestingly, inactivating A4b, although not used in this experimental setting, completely abolished activation of A4a, resulting in the exclusive activation of A5 (Fig. III-8B, lane 3), which had already been described (493, 594). This suggested the downstream region of 3' ss A4a to be crucial for activation of this 3' ss. Mutation of A4a left the strong activation of A5 unchanged, but considerably activated A4b (Fig. III-8B, lane 4). Efficient activation of 3' ss A4c could only be initiated through concomitant inactivation of at least 3' ss A5 and A4b (Fig. III-8B, lane 5), whereas



exclusive use of 3' ss A4c could only be achieved following additional mutation of 3' ss 4a (Fig. III-8B, lane 7). Sole activation of 3' ss A4b could only be detected following inactivation of 3' ss A5 and A4a (Fig. III-8B, lane 6).

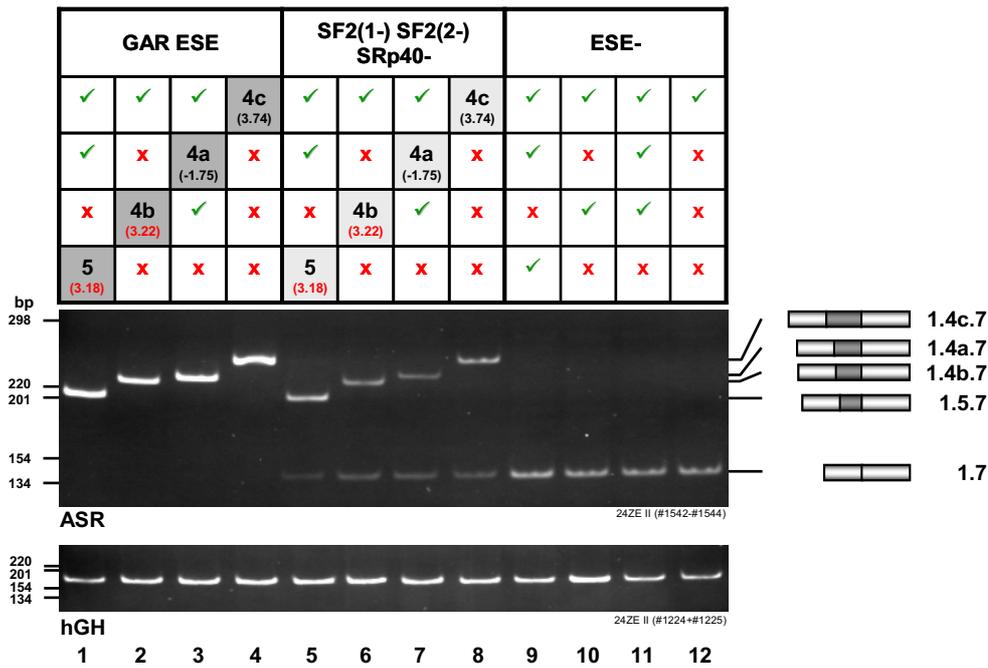
These results demonstrate that basically all 3' ss of the cluster are fully functional in the minigene construct, but obviously, strongly compete with each other. Therefore, inactivation of the 3' ss predominantly used in the wild-type construct did not lead to a more equal activation of the non-mutated splice sites, but resulted in one predominantly used alternative 3' ss.

To determine if all 3' ss of the cluster depend on the GAR ESE the splicing pattern of the minigenes with prevalent activation of an individual 3' ss was compared using two different mutations of the GAR ESE. First, the effect of the triple mutation of the GAR ESE on activation of the respective 3' ss was analysed, which reduced but still allowed residual exon recognition using A4a and A5 (cf. Fig. III-3C, lane 4). RT-PCR analysis of transiently transfected HeLa-T4<sup>+</sup> cells showed that each respective 3' ss of the cluster responded to this mutation of the GAR ESE with reduced exon recognition and concomitant exon skipping (Fig. III-9A, lane 5-8). Nevertheless, the relative amount of mRNA isoforms including the respective internal exon decreased more pronounced for the isoform using A4a, whereas those using A4b and A4c responded to an intermediate degree. A5 showed the lowest reduction in exon recognition after mutation of the GAR ESE suggesting that activation of A4a depends on the GAR ESE the most compared to the other 3' ss of the cluster (Fig. III-9C, cf. 7 with 5, 6, and 8).

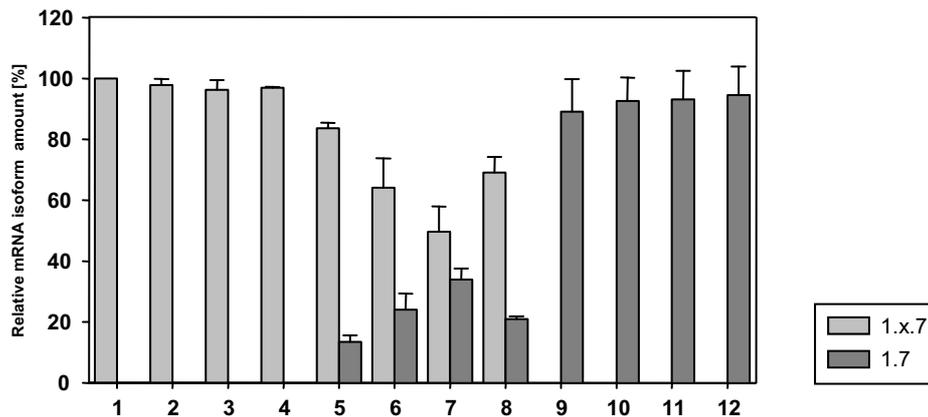
To investigate whether all 3' ss critically depend on the GAR ESE for their activation the GAR ESE mutation, which completely abrogated exon recognition in the presence of all 3' ss of the cluster (Fig. III-5C, lane 2), was introduced into the mutant minigenes using mainly one 3' ss. Inactivation of the GAR ESE completely abolished exon recognition mediated by each individual 3' ss of the cluster and increased the amount of skipped mRNA (Fig. III-9B, lanes 9-12).

These findings indicate, that all four 3' ss strongly compete with each other and that all critically depend on the GAR ESE for their activation albeit to a different degree. The extent, however, appears not to correlate with the predicted intrinsic strength of the 3' ss.

A



B



**Fig. III-9: All 3' ss of the cluster are activated by the GAR ESE.**

(A) RT-PCR analysis of total RNA from HeLa-T4<sup>+</sup> cells transiently transfected with 2-intron minigenes carrying mutations in the 3' ss in combination with the GAR ESE (lanes 1-4) or mutated enhancer sequences (lanes 5-12). Efficient internal exon inclusion is marked by dark grey highlighting, less efficient exon inclusion by light grey highlighting. Exclusive exon skipping is denoted without colouring.

(B) Quantification of alternatively spliced mRNA isoforms. Internal exon inclusion (1.x.7) is denoted in light grey, exon skipping (1.7) is represented in dark grey bars. The amount of the respective mRNA isoform was calculated relative to the total mRNA amount in each sample. Error bars represent standard deviation derived from the mean of three independent experiments.

### **C.1.2 The GAR ESE – a key player in a network of regulatory elements across exon 5.**

Activation of alternatively used splice sites is frequently regulated by interactions of multiple elements [reviewed in (564)]. To elucidate if the GAR ESE is embedded in a regulatory network it was analysed whether the intrinsic strength of the downstream 5' ss and additional sequences in exon 5 contribute to internal exon recognition.

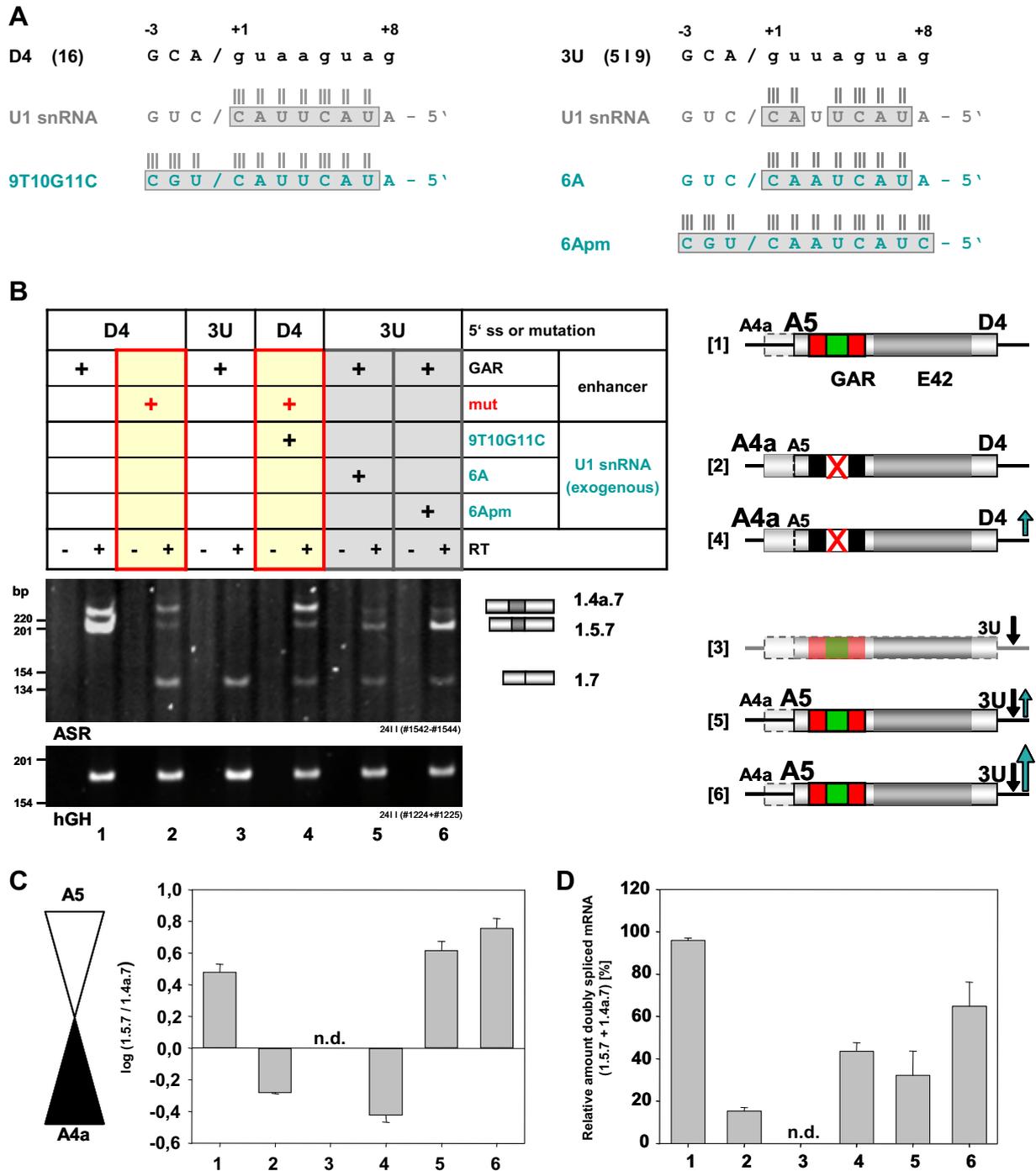
#### **C.1.2.1 GAR ESE mediated 3' ss selectivity is augmented by an efficient downstream 5' ss**

It has previously been shown that besides activating splicing at the upstream 3' ss the GAR ESE also recruits U1 snRNP to the downstream 5' ss D4 (91). Since U1 snRNP has been shown to assist exon recognition in general (260), the role of cross-exon interactions between U1 snRNP bound at D4 and the 3' ss cluster in the preferential activation of A5 was studied. To this end, it was investigated, if increasing U1 snRNP binding is able to shift back the mutationally induced preferential activation of A4a towards the dominant A5 usage observed in the wild-type minigene construct. Preferential activation of A4a was again achieved by introducing the triple SR protein binding site mutation into the GAR ESE, which still facilitated residual exon recognition (Fig. III-3C, lane 4). Cotransfection of an expression plasmid coding for an U1 snRNA with full complementarity to D4 solely increased recognition of the internal exon 4a but failed to restore preferential 3' ss A5 usage (Fig. III-10B, C, and D, cf. 2 and 4). This

#### **Fig. III-10: U1 snRNP binding at 5' ss D4 augments exon recognition in general thereby promoting GAR ESE-mediated 3' ss selectivity.**

**(A)** Complementarity of the parental 5' ss D4 (left) and the 3U mutant (right) to the free 5'-end of the endogenous U1 snRNA (grey) or recombinant U1 snRNAs optimised for higher complementarity (blue). Cotransfection of the U1 snRNA expression plasmid pUCBU1 6A (6A) harbouring a compensatory mutation reconstitutes the base pairing of the mutated 5' ss 3U, whereas plasmids pUCBU1 9T10G11C (9T 10G 11C) or pUCBU1 6Apm (6Apm) extend the base pairing to neighbouring. Hydrogen bonds are represented by vertical bars. Exonic nucleotides of the 5' ss are denoted in capital, intronic nucleotides in lower case letters.

**(B)** RT-PCR analysis of total RNA isolated from HeLa-T4<sup>+</sup> cells transiently cotransfected with the 2-intron minigenes SV leader SD1 SA5 env nef (lane 1, D4 GAR), SV leader SD1 SA5 SF2(1)<sup>-</sup> SF2(2)<sup>-</sup> SRp40<sup>-</sup> env nef (lane 2 and 4, D4 mut) or SV leader SD1 SA5 3U env nef (lane 3, 5 and 6, 3U GAR) (left), U1 snRNA expression plasmids as indicated and pXGH5 to monitor transfection efficiency. A schematic



**Fig. III-10: continued.**

overview of the RT-PCR results (right) demonstrates that increasing U1 snRNA complementarity solely augments GAR ESE-mediated 3' ss selectivity. RNA isolation and RT-PCR were performed as described in Fig. III-1B. The exon/intron structure of the alternatively spliced mRNAs (ASR) is indicated at the right.

**(C)** Quantification of 3' ss selectivity in doubly spliced mRNA. 3' ss selectivity was calculated as ratio of 1.5.7- to 1.4a.7-mRNA. Error bars represent standard deviation from the mean of three independent experiments.

**(D)** Quantification of exon inclusion. Raw data from the experiments in (C) were analysed regarding the percentage of doubly spliced mRNA relative to total mRNA isoforms per sample. Error bars indicate standard deviation.

indicates that increased RNA duplex formation between U1 snRNA and D4 only advanced overall exon recognition, whereas preferential A5 selection was ensured by the GAR ESE. As a control the effect of increasing U1 snRNP binding stability on the ratio of A5 to A4a usage was also analysed in the presence of the wild-type GAR ESE. Because exon recognition already predominates in the parental minigene construct containing the GAR ESE and D4 and cannot be increased by providing higher complementarity of the 5' ss to the U1 snRNA [C. Konermann, diploma thesis (320)], D4 was first inactivated by introducing a point mutation at position +3 from A to U (Fig. III-10A, 3U). This mutation has been shown to lower the complementarity of D4 to U1 snRNA thereby abrogating U1 snRNP binding, and can be rescued upon cotransfection of a corresponding U1 snRNA expression plasmid carrying a compensatory mutation (295, 373). Inactivating D4 (+3A>U) eliminated internal exon recognition in the 2-intron minigene construct (Fig. III-10B and D, lane 3). As expected, cotransfecting plasmids expressing compensatory (high or full 5' ss complementarity) U1 snRNAs rescued internal exon recognition (Fig. III-10B and D, 5 and 6). However, as seen before with the mutated GAR ESE, increasing U1 snRNP binding solely augmented activation of the 3' ss selected by the SR protein binding sites of the GAR ESE (Fig. III-10B, cf. 1 with 5 and 6). This confirms that 3' ss selection is exclusively determined by the GAR enhancer, whereas U1 snRNP binding at the downstream 5' ss supports the overall level of 3' ss selectivity through cross-exon interactions.

Although increased U1 snRNP binding did not account for 3' ss selectivity, it generally supported GAR ESE-mediated 3' ss activation (Fig. III-10B and D, cf. 2 with 4 and 5 with 6). Substituting D4 with a weaker 5' ss of intermediate complementarity to endogenous U1 snRNA showed no effect on 3' ss selection and general exon recognition (Fig. III-11B, lane 2). Surprisingly, introducing a low complementarity 5' ss activated the previously described cryptic 5' ss 13 nt downstream of D4 (438) (Fig. 11A), leading to inclusion of an additional intronic sequence into exons 4a and 5 (Fig. III-11B, lane 3). Introduction of this weak 5' ss resulted in a moderate decrease in 3' ss selectivity (Fig. III-11C, 3). This demonstrates that the degree of GAR enhancer-mediated 3' ss selectivity depends on the stability of U1 snRNP binding at the downstream 5' ss.

Taken together these results show that specific 3' ss selection is conferred by the GAR ESE, rather than by U1 snRNP binding stability. Nevertheless, U1 snRNP binding at

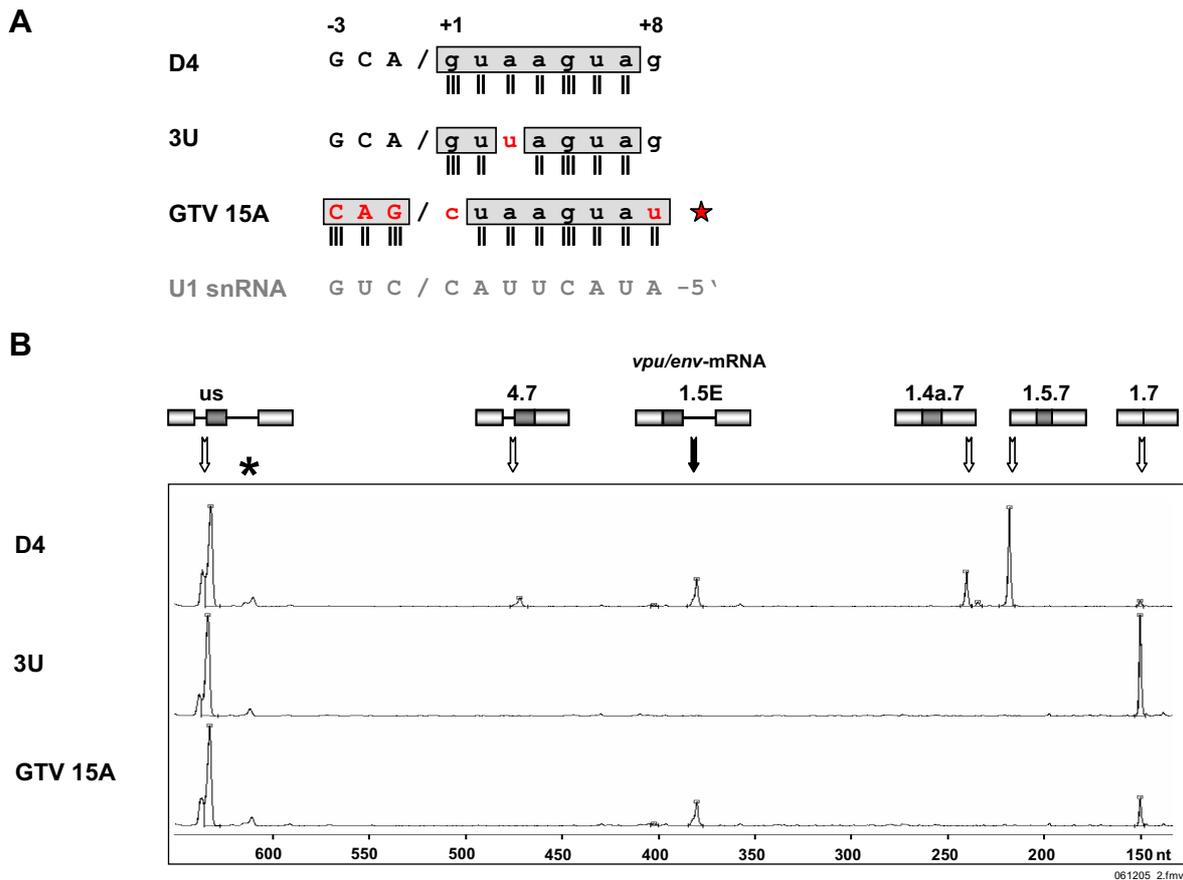


### **C.1.2.2 Binding of U1 snRNP at D4 is necessary to activate the 3' ss cluster in Rev-dependent, intron-containing mRNAs**

To examine the role of U1 snRNP binding at D4 on activation of the 3' ss cluster in the intermediate and late phase of viral gene expression 5' ss D4 was compared with two mutations in absence of its splicing function: 3U abolishing U1 snRNP binding and GTV 15A restitution of U1 snRNP binding but suppressing splicing at D4 due to a mutation of the canonical GT-dinucleotide (Fig. III-12A). The 3U-mutation led to complete loss of Rev-dependent *vpu/env*-mRNA, whereas the amount of unspliced mRNA remained unchanged, demonstrating that none of the 3' ss in the cluster upstream of exon 5 had been activated (Fig. III-12B, cf. 3U vs. D4). To confirm that lack of detectability of the *vpu/env*-mRNA was caused by loss of U1 snRNP binding at D4, U1 snRNP binding was reconstituted by substituting D4 with a splicing inactive U1 snRNP binding site (GTV) (295). The GTV sequence possesses full complementarity to the 5' end of the U1 snRNA except for position +1, which was mutated from G to C to prevent splicing. After introducing the GTV sequence, again splicing from the cryptic 5' ss 13 nt downstream of the inactivated 5' ss (438) was observed (data not shown). To exclude splicing of the downstream intron, this cryptic 5' ss was additionally inactivated (Fig. III-12A, GTV 15A). Restitution of U1 snRNP binding rescued the expression of *vpu/env*-mRNA to the same extent as observed for D4 (Fig. III-12B, cf. GTV 15A vs. D4). From this experiment it was concluded that U1 snRNP binding at D4 is needed for 3' ss activation for generation of Rev-dependent *vpu/env*-mRNAs in the intermediate and late phase of viral gene expression.

### **C.1.2.3 The 3' region of exon 5 contains an additional splicing enhancer.**

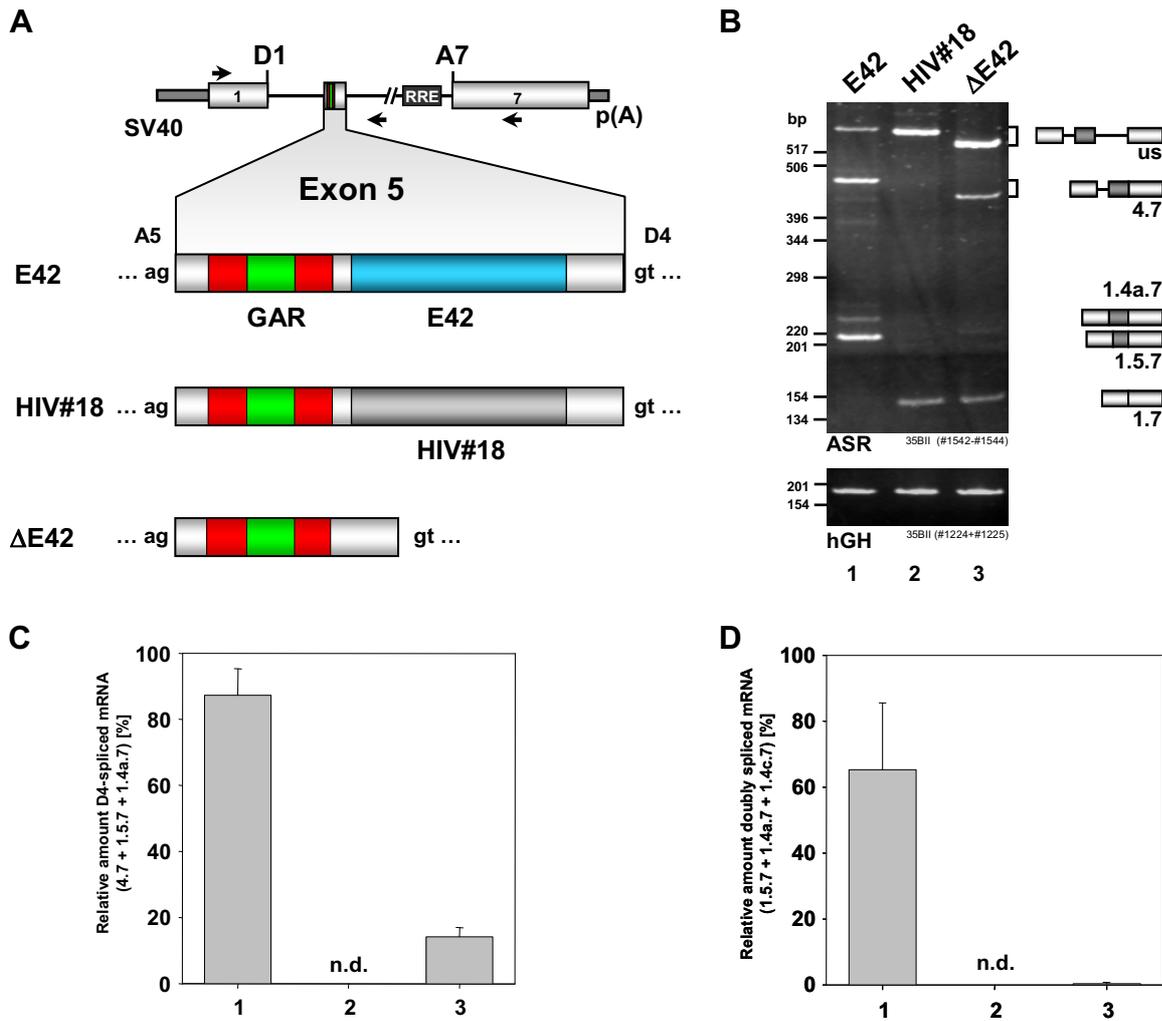
The finding that U1 snRNP bound to the downstream 5' ss of exon 5 takes part in the GAR ESE-mediated 3' ss activation raised the question of whether the exonic sequence located between the enhancer and D4, termed E42, might also be involved in the formation of the cross-exon interactions. To investigate the impact of the E42 sequence on exon recognition the splicing pattern of the 2-intron minigene construct was analysed after either substituting the E42 sequence with an HIV-1 control sequence of identical length, which did not support U1 snRNP binding at D4 (HIV#18) (91), or deleting E42 ( $\Delta$ E42) (Fig. III-13A). RT-PCR analysis of transiently transfected HeLa-T4<sup>+</sup> cells revealed that substituting the E42 sequence completely abolished internal exon



**Fig. III-12: U1 snRNP binding at D4 is essential for activation of the 3' ss cluster generating intron-containing *vpu/env*-mRNAs.**

**(A)** Complementarity of parental 5' ss D4 and mutants thereof to endogenous U1 snRNA included in the U1 snRNP. Exonic nucleotides are depicted in capital letters; the exon/intron border is depicted by a slash. Nucleotides complementary to U1 snRNA are highlighted in grey. Hydrogen bonds are illustrated by vertical bars. Nucleotide substitutions in mutated 5' ss are shown in red. The star marks the additional mutation of the cryptic 5' ss 13 nt downstream of D4 (438) (+15 T>A, numbering relative to position +1 in D4).

**(B)** RT-PCR analysis of total RNA from HeLa-T4<sup>+</sup> cells transfected with the 2-intron reporter construct carrying wild-type D4 (D4), a mutant unable to stably bind endogenous U1 snRNA (3U) or a splicing-deficient U1 snRNP binding site with the additional mutation of the cryptic 5' ss downstream (GTV 15A). Cells were cotransfected with 1  $\mu$ g SVcrev, an expression plasmid for the viral regulatory protein Rev, allowing the export of intron-containing reporter mRNA into the cytoplasm, and with 1  $\mu$ g pXGH5. RT-PCR was performed using Cy5-labelled sense primer #1544, separated on denaturing gels and detected by Automated Laser Fluorescence (ALF) as described in Fig. III-2. RT-PCR products are shown as processed fluorescence curve data of the electrophoretic separation. The exon/intron structure of alternatively spliced mRNA isoforms is depicted above the lanes. The asterisk marks an RT-PCR signal, which was identified as unspecific signal by DNA sequencing.



**Fig. III-13: The E2F2 sequence exerts an additional enhancer function ensuring recognition of the internal exons 4a and 5.**

**(A)** Composition of splicing regulatory elements in the central exon 5 of the 2-intron minigene reporter and mutant constructs. The E2F2 fragment in 3' region of exon 5 has either been substituted for a control HIV-1 sequence (HIV#18) or deleted ( $\Delta$ E2F2).

**(B)** RT-PCR analysis of total RNA isolated from HeLa-T4<sup>+</sup> cells transfected with the parental 2-intron minigene or E2F2-mutants as illustrated in (A).  $2 \times 10^5$  cells were transiently transfected with 1  $\mu$ g SV leader SD1 SA5 env nef (E2F2, lane 1), SV leader SD1 SA5 HIV#18 SD4 env nef (HIV#18, lane 2) or SV leader SD1 SA5 H- SD4 env nef ( $\Delta$ E2F2) and 1  $\mu$ g pXGH5 to control equal transfection efficiency. RNA isolation and RT-PCR was performed as described in Fig. III-1B. [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, us: unspliced].

**(C)** Quantification of D4 activation. The sum of doubly and singly spliced RNA using D4 (4.7 + 1.4a.7 + 1.5.7) was calculated relative to total mRNA per sample. Error bars indicate standard deviation from the mean of three independent experiments [n.d.: not detected].

**(D)** Quantification of exon inclusion. Raw data originating from the experiments in (C) were analysed regarding the percentage of doubly spliced mRNA using A5, A4a or A4c relative to total mRNA isoforms per sample. Error bars represent standard deviation.

---

recognition (mRNAs isoforms 1.4a.7 and 1.5.7) and, unexpectedly, also activation of D4 resulting in the loss of singly spliced RNA (4.7) (Fig. III-13B, C and D, cf. lanes 1 and 2). Additionally, the lack of internal exon recognition – as previously observed (cf. 1.1.1) – not only led to appearance of the 1.7 mRNA but also increased the amount of unspliced RNA. This demonstrates that although E42 is insufficient to activate D4 by itself (91), this sequence is essential to connect the activating effect of the GAR ESE to the 5' ss. On the other hand D4 was still activated in the absence of E42, while exon recognition was nearly abolished resulting in the identification of mostly unspliced and singly spliced RNA using either splice site pair D1/A7 (1.7) or D4/A7 (4.7) (Fig. III-13B, C and D, lane 3). This shows that the GAR ESE in proximity to the 5' ss exclusively restored the enhancing function on D4, whereas internal exon recognition was only slightly recovered by use of A4c, the most upstream 3' ss of the cluster. Therefore, the GAR ESE when positioned directly upstream of D4 enhances splicing at this 5' ss but failed to efficiently activate any of the 3' ss simultaneously. Although inhibitory elements in the substituted sequence or exon length constraints cannot be ruled out formally, it was reasoned from these experiments that in addition to the GAR enhancer the E42 sequence is essential for establishing cross-exon interactions necessary for 5' ss activation and efficient exon recognition in completely spliced mRNAs.

### C.1.3 The mechanism of GAR ESE enhancer function

To elucidate the molecular mechanism of GAR ESE-mediated 3' ss cluster activation, it was investigated whether the enhancer promotes 3' ss selectivity and exon recognition by stabilising spliceosomal proteins. To this end, the GAR ESE-dependent binding of nuclear proteins to *in vitro* transcribed RNA consisting of the sequence upstream of 3' ss A5 as well as the enhancer sequence located downstream was examined.

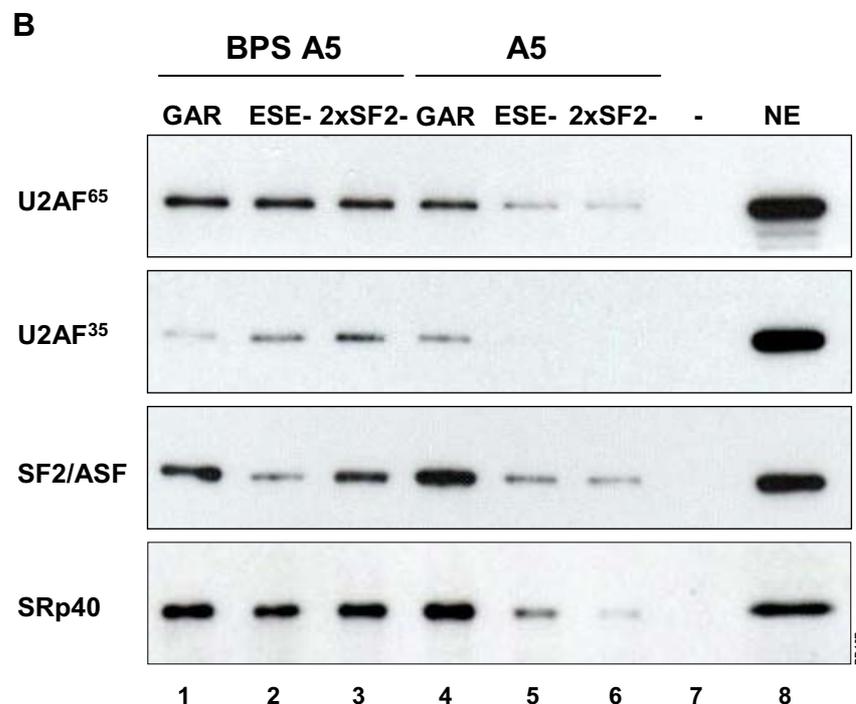
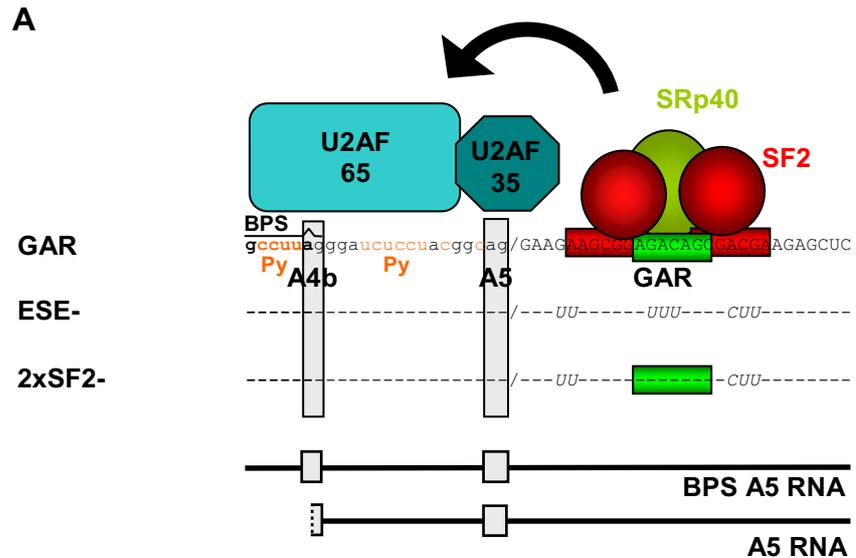
#### C.1.3.1 The GAR ESE stabilises binding of U2AF<sup>65</sup> at 3' ss A5.

In one of the earliest complexes committed to the splicing pathway, the ATP-independent E complex, the 65 kDa subunit of U2AF (U2AF<sup>65</sup>) recognises the polypyrimidine tract (PPT) (677, 679), whereas the 35 kDa subunit (U2AF<sup>35</sup>) binds to the downstream located AG-dinucleotide (408, 656, 697). A number of splicing enhancers have been shown to increase splicing efficiency by stabilising the U2AF heterodimer or one of its subunits at the 3' ss (223, 604, 638, 700). However, since other studies had reported contradictory results (234, 297, 352), enhancers might be classified by the mechanistic interactions they exert. To examine whether the GAR ESE promotes 3' ss activation by stabilising the U2AF heterodimer at 3' ss A5, which was predominantly used in the *in vivo* splicing assays (cf. Fig. III-2), binding of both subunits, U2AF<sup>65</sup> and U2AF<sup>35</sup>, to 3' ss A5 was compared *in vitro* in the presence of the GAR ESE and mutations thereof using RNA affinity chromatography (Fig. III-14A). For this purpose, mutations were selected, which either eliminated exon inclusion (ESE-) or severely diminished exon inclusion concomitantly shifting the ratio of 3' ss selection from A5 to A4a (2xSF2-) in the *in vivo* splicing assay (cf. Fig. III-5).

Immunoblot analyses of U2AF binding to *in vitro* transcribed RNA containing the GAR ESE flanked by the AG-dinucleotide of 3' ss A5 and the preceding PPT up to but excluding the most proximal BPS (594) revealed that binding of U2AF<sup>65</sup> decreased after inactivation of the GAR ESE (Fig. III-14B, cf. lane 4 with lane 5) and also after mutation of two SF2/ASF binding sites of the GAR ESE (Fig. III-14B, lane 6). In addition, binding of U2AF<sup>35</sup> was abrogated by each of the mutations (Fig. III-14B, cf. lane 4 with lanes 5 and 6). These findings demonstrated that the GAR ESE stabilises binding of the U2AF heterodimer to the 3' ss in the presence of a minimal PPT.

**Fig. III-14: The GAR ESE stabilises binding of U2AF<sup>65</sup> to RNAs containing 3' ss A5.**

**(A)** RNA sequences employed in RNA affinity chromatography. Intronic nucleotides are denoted in lower case letters, exonic nucleotides in capital letters. The binding sites of the GAR ESE for the SR proteins SF2/ASF and SRp40 are marked within the sequence by red and green boxes. Hypothesised positions of the spliceosomal proteins U2AF<sup>65</sup> and U2AF<sup>35</sup> recognising distinct functional elements of 3' ss A5 are indicated. Pyrimidines (Py) constituting the polypyrimidine tract (PPT) are shown in orange. The questioned stabilisation of the U2AF heterodimer by enhancer-bound proteins is illustrated by an arrow. Below the sequences short and extended target RNAs used in



(B) are depicted containing the intronic sequence either excluding the most proximal BPS of A5 and the A of the AG-dinucleotide of A4b (A5 RNA) or including the BPS of A5 (BPS A5 RNA) [BPS: branch point sequence].

**(B)** Immunoblot analyses of proteins isolated from HeLa nuclear extract by RNA affinity chromatography using RNA targets described in (A). For each sample 2000 pmol RNA were immobilised and incubated with 60% HeLa nuclear extract at 30°C for 20 min in the presence of phosphatase inhibitors. Proteins were eluted from the bead matrix by heating to 95°C in SDS-sample buffer. 4% of the isolated proteins were separated in SDS-polyacrylamide gels and subjected to immunoblotting using antibodies against U2AF<sup>65</sup> (generously provided by Prof. Dr. M. Hastings and Prof. Dr. A. Krainer, Cold Spring Harbor Laboratory, USA), U2AF<sup>35</sup> (ptglab), SF2/ASF (Zymed), and SRp40 (US Biological). A negative control of the binding reaction was performed lacking RNA thereby confirming that protein binding in the samples occurs at the target RNA. HeLa nuclear extract (6 µg) was used as positive control (NE).

Since the PPT of 3' ss A5 is dispersed by a number of purines, it was tested whether additional pyrimidines upstream of the AG-dinucleotide of 3' ss A4b, which at the same time represent the most proximal BPS of 3' ss A5, contribute to U2AF heterodimer binding. Extending the RNA target at its 5' end by only six nucleotides thereby introducing four additional pyrimidines, revealed that the amount of U2AF<sup>65</sup> bound to the extended RNA targets was unaffected by inactivation as well as mutation of only two SF2/ASF binding sites of the GAR ESE (Fig. III-14B, lanes 1-3) indicating that the additional pyrimidines strengthen the interaction between U2AF<sup>65</sup> and the RNA target. However, binding of U2AF<sup>35</sup> unexpectedly increased after mutation of the GAR ESE (Fig. III-14B, cf. lane 1 with lanes 2 and 3). Since the increase in U2AF<sup>35</sup> binding to extended RNA targets after mutations of the GAR ESE contrasted the decrease of U2AF<sup>35</sup> bound to RNA targets carrying the same GAR ESE mutations but lacking the PPT extension, it is likely that the GAR ESE does not directly influence U2AF<sup>35</sup> binding. Instead, increasing amounts of U2AF<sup>35</sup> at constant U2AF<sup>65</sup> binding might hint to an increase of U2AF heterodimers recognising the 3' ss A5 of the RNA target. This result suggested that the GAR ESE stabilises monomeric U2AF<sup>65</sup> at sufficient long PPTs and that mutations of the GAR ESE are compensated by increasing the proportion of U2AF heterodimers bound to the RNA target. This hypothesis is in agreement with previous findings that recombinant U2AF<sup>35</sup> binds with only low intrinsic binding affinity to an AG-dinucleotide containing RNA substrate (408). In contrast, U2AF<sup>65</sup> is able to efficiently bind to polypyrimidine-rich RNA in the absence of U2AF<sup>35</sup>. However, dimerisation of both subunits strongly increases the binding affinity of the U2AF heterodimer to RNA (506). Therefore, this result suggested that the GAR ESE stabilises U2AF<sup>65</sup> at the PPT upstream of 3' ss A5 in the absence of U2AF<sup>35</sup>.

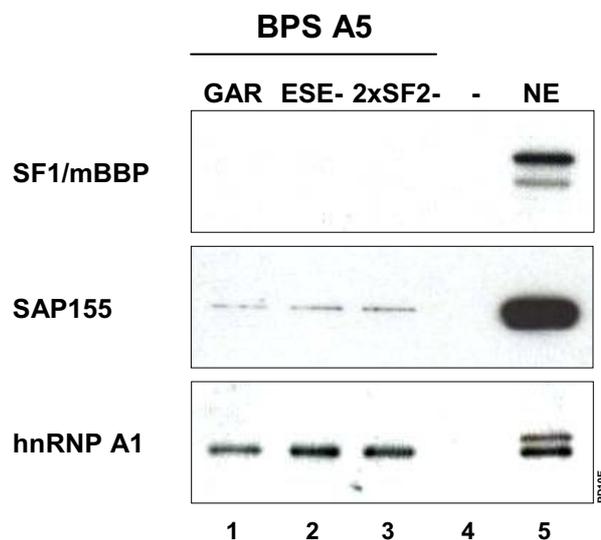
Since the GAR ESE-mediated changes in the binding of both U2AF subunits observed *in vitro* are not correlating with the respective levels of exon recognition in the *in vivo* splicing assays, the GAR ESE-mediated stabilisation of U2AF<sup>65</sup> monomers might be involved in 3' ss selection rather than exon recognition. Therefore, this result indicated that the GAR ESE promotes general 3' ss activation by a mechanism distinct from U2AF stabilisation.

### C.1.3.2 The GAR ESE does not stabilise spliceosomal components SF1/mBBP and SAP155.

Since U2AF<sup>65</sup> bound equally strong to RNA targets containing the BPS and the GAR ESE or mutations thereof, it was questioned whether addition of the BPS switches the protein stabilised by the GAR ESE from U2AF<sup>65</sup> in short RNA targets to SF1/mBBP, which recognises the BPS in the early E' (307) and E complex (4, 44) (Fig. I-8B) or even to U2 snRNP, which is already loosely associated in the E complex (145, 414, 488). To elucidate whether assembly of the early BPS-dependent E complex at A5 is promoted by the GAR ESE in the extended RNA targets, binding of SF1/mBBP and SAP155, a component of U2 snRNP, to BPS-containing RNA targets was monitored by immunoblotting of proteins isolated by RNA affinity chromatography (Fig. III-15). Both proteins, however, bound only very inefficiently to the target RNAs compared to the control protein hnRNP A1 (Fig. III-15, lower panel). Although other 3' ss had been reported to assemble into the E complex in the absence of the upstream 5' ss, from this result it was concluded that under the experimental conditions used here the GAR ESE does neither promote binding of SF1/mBBP nor U2 snRNP to RNA targets containing the most proximal BPS of 3' ss A5.

**Fig. III-15: Binding of SF1/mBBP and SAP155 to RNA substrates containing the most proximal BPS of 3' ss A5 is unaffected by the GAR ESE.**

Immunoblot analysis of proteins isolated by RNA affinity chromatography. RNA targets containing the BPS upstream of 3' ss A5 and the GAR ESE or mutants thereof (RNA sequences cf. Fig. III-14A) were subjected to RNA affinity chromatography as described in Fig. III-14B. Immunoblots were probed using antibodies against SF1/mBBP (Lifespan), SAP155 (MBL), and hnRNP A1 (Y-15, SCBT). Equal amounts of HeLa nuclear extract (6 µg) were loaded as positive control. Two SF1/mBBP isoforms of 67 kDa and 75 kDa are expressed in HeLa cells, which are both able to bind to U2AF<sup>65</sup> (484).

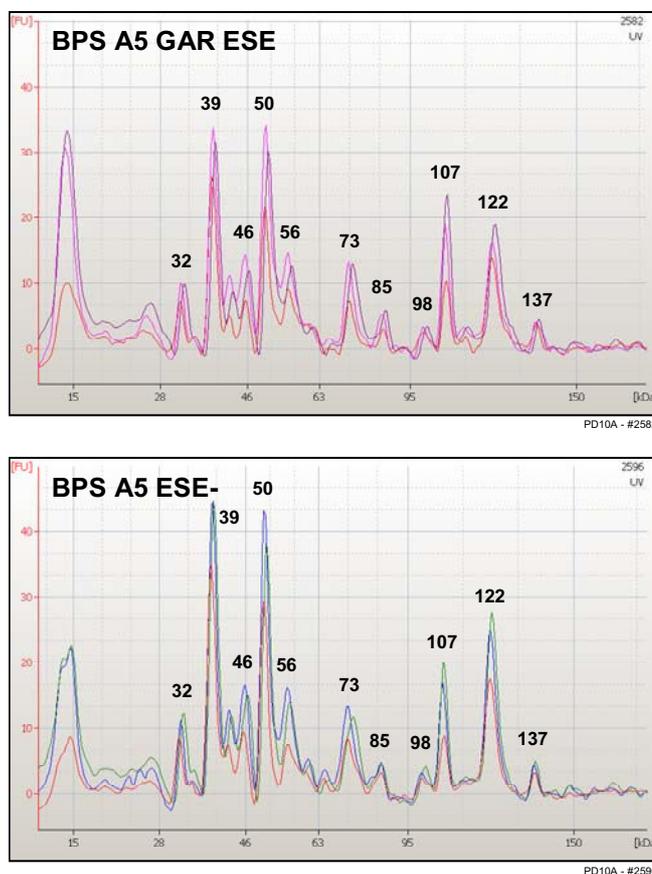


Therefore, the GAR ESE does not switch from U2AF stabilisation in RNA targets containing a minimal PPT to stabilisation of SF1/mBBP or U2 snRNP in the BPS-containing RNA model substrates. However, this result does not rule out a GAR ESE-mediated stabilisation of SF1/mBBP or U2 snRNP binding to additional alternative BPS upstream of the one most proximal to 3' ss A5 in the 2-intron minigene.

### **C.1.3.3 The GAR ESE promotes binding of the splicing regulatory protein hTra2- $\beta$ .**

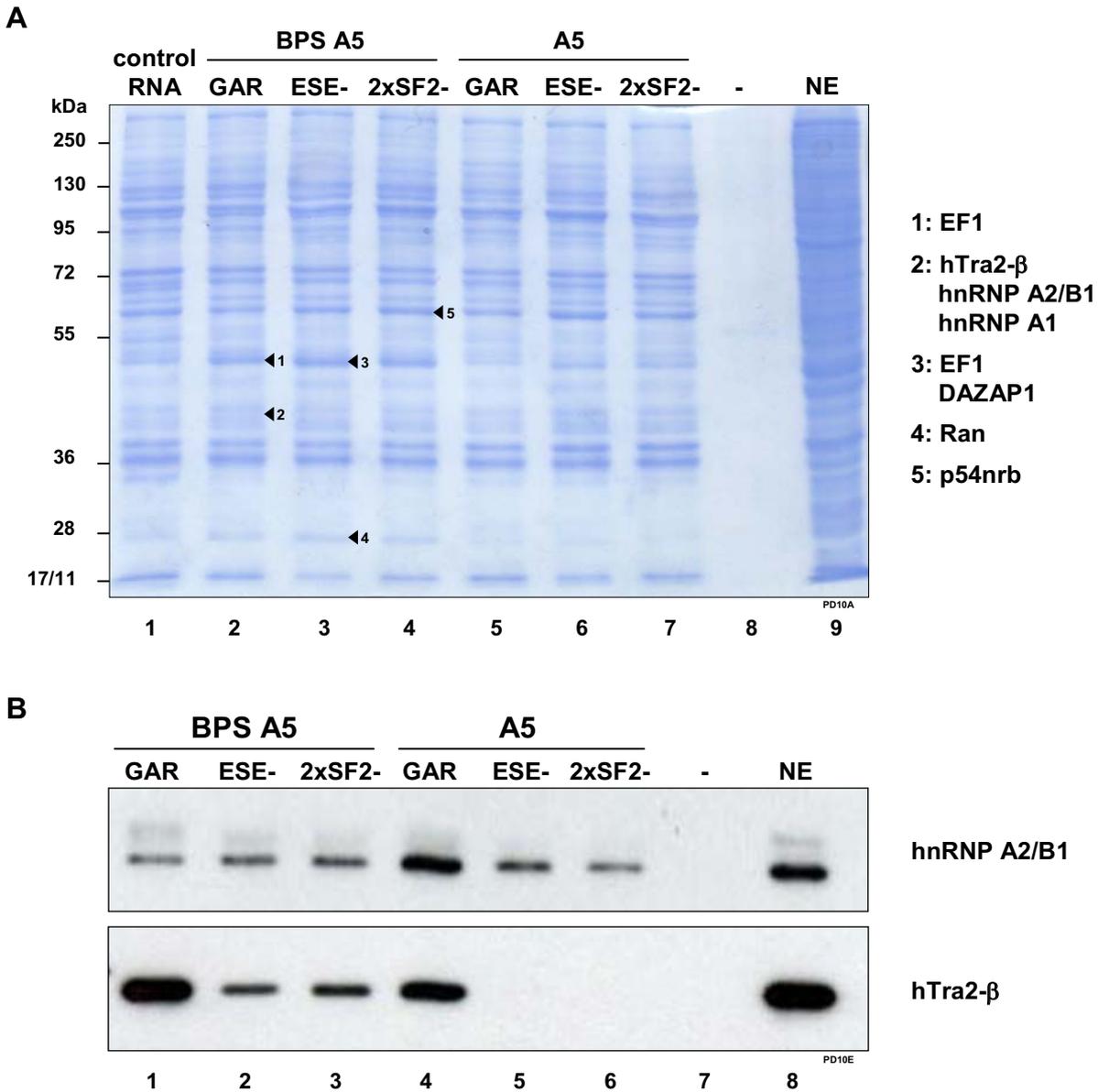
Since GAR ESE-mediated exon recognition observed in the minigene splicing assays *in vivo* could not be attributed to the stabilisation of early spliceosomal components U2AF, SF1/mBBP and SAP155, a more general approach was applied to screen for candidate proteins mediating GAR ESE-dependent exon recognition. To first obtain an overview of proteins bound to RNA targets used for analyses of GAR ESE-mediated stabilisation of early spliceosomal components, proteins isolated by RNA affinity chromatography were subjected to capillary gel electrophoresis and detected by fluorescence (Fig. III-16). The electropherograms of proteins bound to the RNA target containing the BPS and PPT upstream and the GAR ESE located downstream of 3' ss A5 revealed that a large number of proteins bound to the RNA with the most prominent protein signals detected at sizes of 39 kDa, 50 kDa, 107 kDa, and 122 kDa, respectively (Fig. III-16, upper panel, BPS A5 GAR ESE). Although the dense protein pattern in the separation range from 14-230 kDa and the variability of the individual electropherograms hindered the detection of individual protein peaks, slight differences were observed in response to inactivation of the GAR ESE e.g. in the separation range from 55-65 kDa (Fig. III-16, lower panel, BPS A5 ESE-). However, the fluorescent detection of the protein samples containing a variety of high abundant proteins might be too sensitive to analyse differences in the binding pattern of low abundant proteins eventually resulting in a biased identification of high abundant proteins in capillary gel electrophoresis.

In order to identify also low abundant proteins which, dependent on the GAR ESE, bind more strongly to the RNA, proteins isolated by RNA affinity chromatography were separated by SDS-PAGE (Fig. III-17A). As seen before in the fluorescence-detected capillary gel electrophoresis, coomassie stained-gels revealed only little differences in the protein binding pattern after mutation of the GAR ESE (Fig. III-17A, cf. lane 2 with



**Fig. III-16: Analysis of GAR ESE-dependent protein binding to 3' ss A5-containing RNA targets by capillary gel electrophoresis.**

Proteins isolated from HeLa nuclear extracts by RNA affinity chromatography as described in Fig. III-14B were separated by denaturing capillary gel electrophoresis and detected by fluorescence using a microfluid lab-on-a-chip system (Bioanalyzer 2100, Agilent). 4  $\mu$ L of the isolated proteins were denatured with 2  $\mu$ L DTT-containing sample buffer at 95°C, diluted in ddH<sub>2</sub>O to a final volume of 90  $\mu$ L, and analysed in triplicates of 6  $\mu$ L on a protein chip covering a separation range from 14-230 kDa (Protein 230 Kit, Agilent). Fluorescent curve data of triplicate measurements were aligned for each probe.



**Fig. III-17: Identification of proteins stabilised at the 3' ss cluster by the GAR ESE.**

**(A)** Coomassie staining of proteins isolated by RNA affinity chromatography using RNA targets described in Fig. III-14A. An RNA target containing the HIV-1 5' ss D1 and the neighbouring exonic and intronic sequence was used as control (RNA control, lane 1). Half the amount of proteins isolated using 2000 pmol RNA was separated in a 10% SDS-polyacrylamide gel. Numbered bands were isolated from the gel and proteins were digested by trypsin. Resulting peptides were analysed using mass spectrometry (Dr. W. Bouschen, BMFZ, HHUD).

**(B)** Immunoblot analyses of protein samples described in (A) using antibodies against hnRNP A2/B1 (SCBT) (upper panel) and hTra2-β (Abcam) (lower panel).

lanes 3 and 4). Nevertheless, besides minor differences in band size and differences in band intensity between short and BPS-containing RNA targets, one faint coomassie gel band was observed to reduce intensity after mutation of the GAR ESE (Fig. III-17A, band 2, cf. lanes 2 and 3) and was therefore suggested to contain proteins that might be involved in GAR ESE-mediated regulation of 3' ss activation. This band might have been hidden in the capillary gel electrophoresis by surrounding high abundant proteins. Sequencing of proteins purified from the coomassie gel band (Fig. III-17A, band 2, lane 2) by mass spectrometry identified the splicing regulatory proteins hnRNP A2/B1 and hTra2- $\beta$ . Verification of protein binding by immunoblotting revealed that hnRNP A2/B1 binding to the RNA was only stabilised by the GAR ESE in short RNA targets, whereas addition of the BPS and extended PPT, respectively, abrogates GAR ESE-mediated stabilisation of hnRNP A2/B1 (Fig. III-17B, upper panel, cf. lanes 1 and 4) suggesting that hnRNP A2/B1 is not involved in GAR ESE-mediated 3' ss activation *in vivo*. In contrast, enhancer-dependent binding of hTra2- $\beta$  to short but also to extended RNA targets was confirmed by immunoblotting. hTra2- $\beta$  bound strongly to GAR ESE-containing RNA target sequences in the presence of the minimal PPT (Fig. III-17B, lower panel, lane 4). Inactivation and also mutation of only two SF2/ASF binding sites of the GAR ESE in RNA target sequences lacking the BPS completely abolished hTra2- $\beta$  binding (Fig. III-17B, lower panel, lanes 5 and 6). Since the purine-rich consensus sequence found in a number of hTra2- $\beta$ -regulated mRNAs (GAARGARR) (586) largely coincides with the GAR ESE sequence and also the intron/exon 5 border, the immunoblot results pointed to a direct binding of hTra2- $\beta$  to the GAR ESE-containing 5'-half of exon 5. Nevertheless, binding of hTra2- $\beta$  still increased in the presence of the additional pyrimidines in the extended RNA target (BPS A5) (Fig. III-17B, cf. lanes 1 and 4). This Increase might be caused either by an additional hTra2- $\beta$  binding site generated by addition of the BPS or by stabilisation of exon 5-bound hTra2- $\beta$  via an yet unidentified factor binding to the BPS. Mutating the GAR ESE in BPS-containing RNA targets did not abolish but severely reduced hTra2- $\beta$  binding (Fig. III-17B, lower panel, lanes 2 and 3) underlining the potential binding of hTra2- $\beta$  to the GAR ESE and the stabilisation of hTra2- $\beta$  to BPS-containing RNA targets. Since binding of hTra2- $\beta$  to BPS-containing RNA targets in the presence of the wild-type GAR ESE or mutations thereof correlated with the GAR ESE-mediated exon recognition *in vivo*, it was suggested from these results that, in addition to SF2/ASF and SRp40, hTra2- $\beta$  is

another splicing regulatory factor involved in GAR ESE-mediated exon recognition. The more prominent reduction of hTra2- $\beta$  binding in response to inactivation of the GAR ESE (ESE-) compared to mutation of only two SF2/ASF binding sites (2xSF2-) might reflect the loss of exon inclusion caused by this mutation in the *in vivo*-splicing assays (cf. Fig. III-5). This result points to a threshold level of hTra2- $\beta$  binding to be necessary for internal exon inclusion.

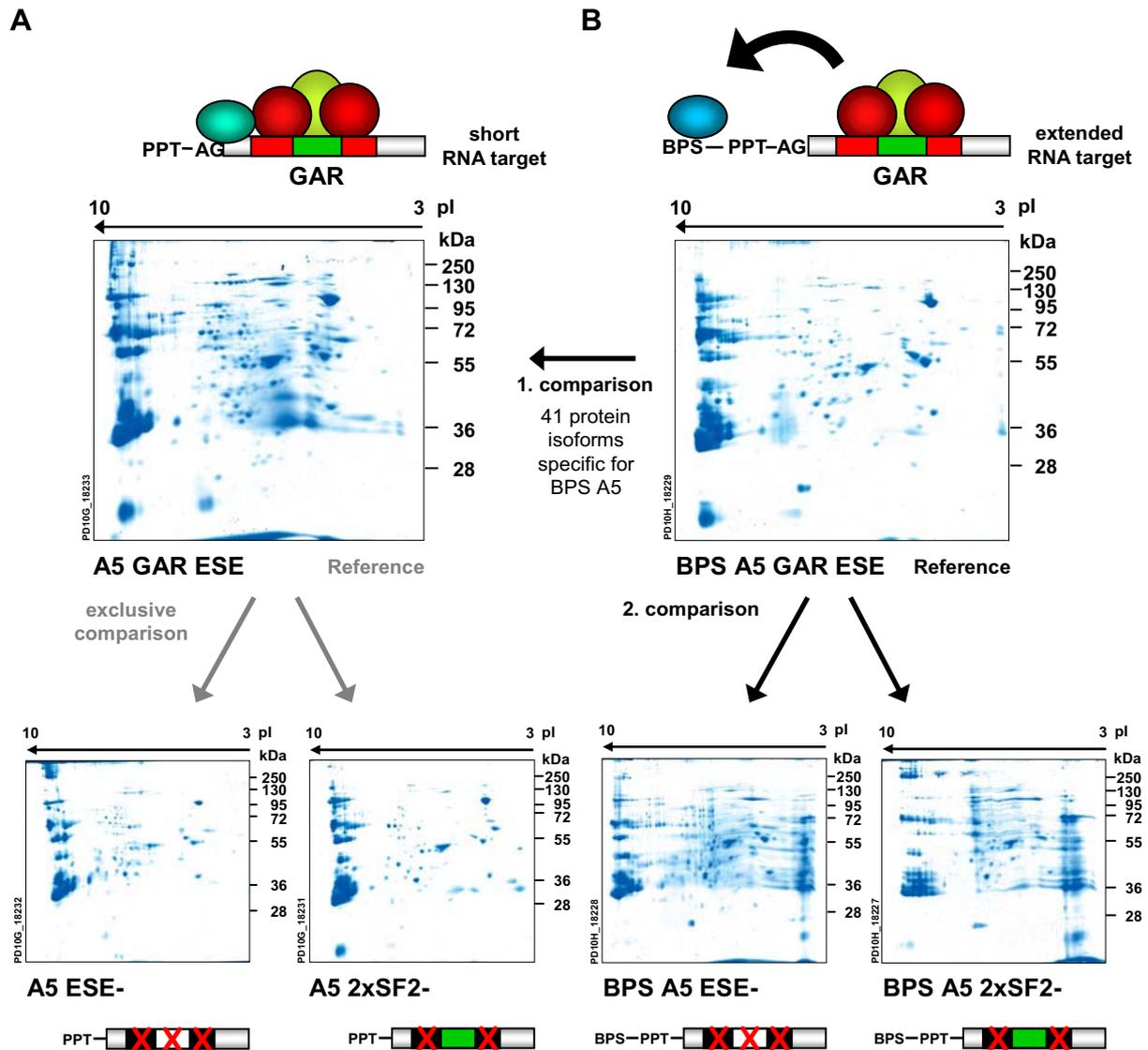
#### **C.1.3.4 Identification of additional proteins stabilised by the GAR ESE**

The binding pattern of hTra2- $\beta$  to intron/exon 5 RNA targets indicated that besides SF2/ASF and SRp40 additional proteins might contribute to the GAR enhancer function and also revealed enhanced binding of hTra2- $\beta$  in the presence of the PPT/BPS extension. Therefore, it was aimed (i) to identify additional proteins binding to the RNA dependent on the GAR ESE sequence and (ii) to analyse whether proteins other than U2AF<sup>65</sup> and SF1/mBBP bind to the PPT/BPS sequence in extended RNA targets, which might further stabilise hTra2- $\beta$  binding to the RNA and thereby promote exon recognition. To this end, proteins isolated by RNA affinity chromatography as performed before were separated by two-dimensional gel electrophoresis to detect differences hidden till this point in the vast numbers of proteins bound to the respective RNA sequences (Fig. III-16 and Fig. III-17A).

To identify proteins binding GAR ESE-dependent to the RNA, proteins isolated by RNA affinity chromatography using short RNA targets containing the minimal PPT upstream and the GAR ESE downstream of the 3' ss A5 (Fig. III-18A, A5 GAR ESE) were compared to the proteins isolated using RNA targets of identical length carrying

#### **Fig. III-18: Analyses of RNA affinity purified proteins by 2D-gel electrophoresis.**

Triplicates of RNA affinity chromatography reactions including phosphatase inhibitors were performed for RNA targets excluding (A) or including the BPS upstream of 3' ss A5 (B) as depicted in Fig. III-14A. Entire protein preparations were passively loaded onto IPG strips pH 3-10 (Amersham) and after isoelectric focusing (Ettan IPGphor III) mounted on 10% SDS-polyacrylamide minigels for separation in the second dimension. Proteins were visualised by sensitive coomassie blue staining using a protocol appropriate for subsequent mass spectrometry analyses (299). As control for RNA specific protein binding affinity chromatography reactions were also performed in the absence of RNA (see Appendix, Fig. VII-2).



**Fig. III-18: continued.**

**(A)** Identification of proteins binding GAR ESE-dependent to 3'ss A5 and the downstream enhancer sequence (short RNA targets). The sketch illustrates SF2/ASF and SRp40 binding to the GAR ESE and the questioned enhancer-mediated binding of potential splicing regulatory proteins. Protein spots derived from the RNA target containing the parental GAR ESE (A5 GAR ESE, reference) were compared to the protein spots resulting from RNA affinity chromatography of the respective RNA targets containing either mutations in the SF2/ASF binding sites and in the SRp40 binding site (A5 ESE-) or only in two SF2/ASF binding sites (A5 2xSF2-).

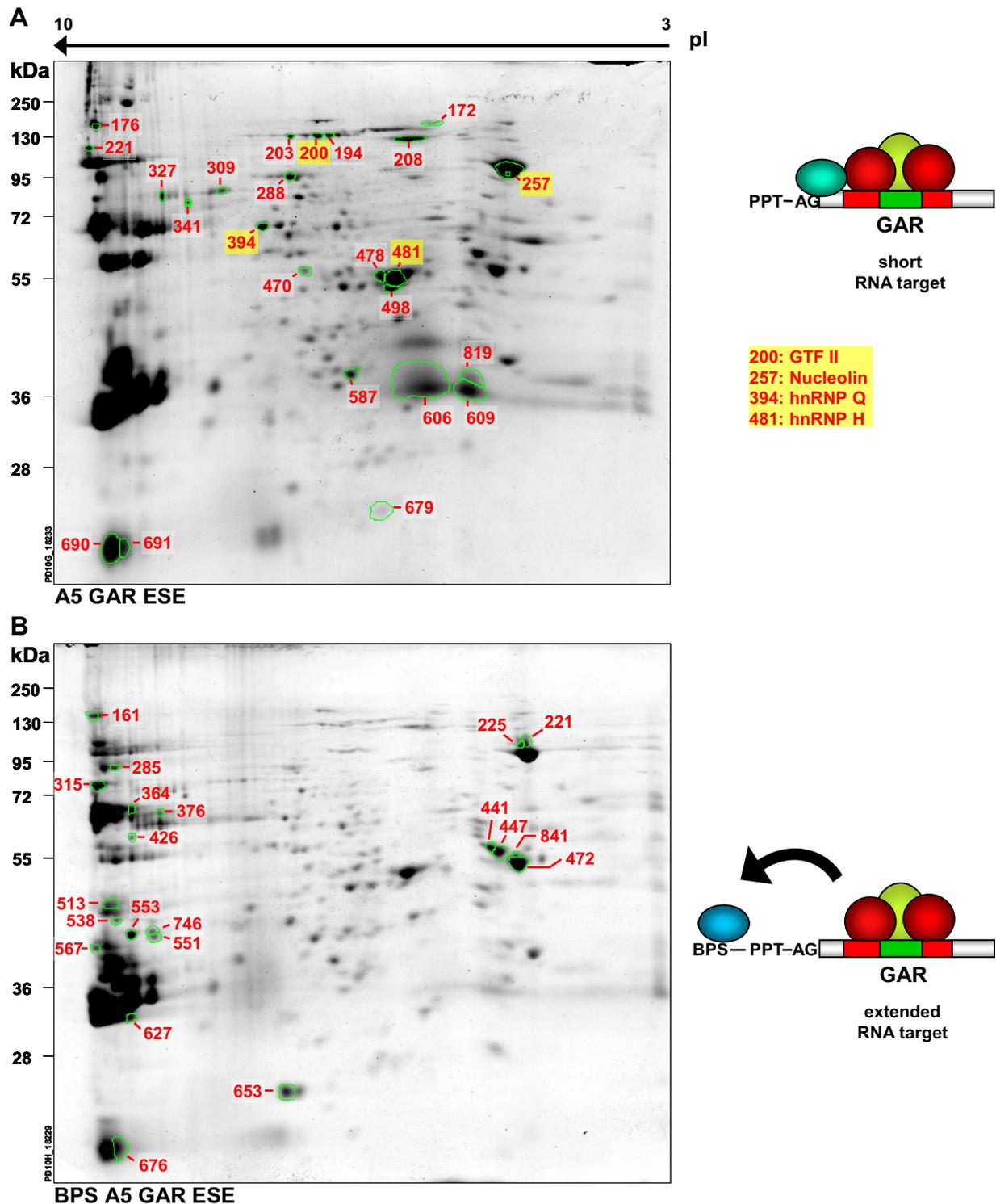
**(B)** Analysis of proteins binding GAR ESE-mediated to RNA targets carrying an extended PPT (extended RNA targets). Protein spots identified after RNA affinity chromatography of RNA targets additionally containing a PPT extension and the most proximal BPS A5 (BPS A5 GAR ESE, reference) were first compared to protein spots derived for the corresponding RNA target lacking the 5'-extension (Fig. III-18A, A5 GAR ESE) yielding 41 protein isoforms binding to the RNA target dependent on the 5'-extension. To determine, which of these proteins bind GAR ESE-mediated to the 5'-extension, protein spots were in a second analysis compared to their respective positional counterparts in 2D-gels derived from RNA affinity chromatography of extended RNA targets carrying mutations in the GAR ESE.

mutations in the GAR ESE (Fig. III-18A, A5 ESE-). To further refine the subgroup of proteins contributing to exon recognition, the simultaneous mutation of two SF2/ASF binding sites of the GAR ESE was used as control to eliminate proteins which might be involved in 3' ss selection, but which still allow low levels of exon recognition and therefore interfere with the detection of proteins essentially mediating exon recognition (Fig. III-18A, A5 2xSF2-). Proteins were first separated according to their isoelectric point (pI) using an immobilised pH gradient and in the second dimension according to their size (Fig. III-18A). Since calculated pI-values ([www.phosphosite.org](http://www.phosphosite.org)) for representative splicing regulatory proteins were found to extend into the alkaline region, a broad pH-gradient ranging from pH 3-10 was chosen for isoelectric focusing in the first dimension. Nevertheless, binding of vast numbers of protein isoforms with alkaline pI-values to the RNA targets impaired their differentiation into individual proteins spots resulting in an only suboptimal separation resolution in the alkaline area.

After dividing the protein population binding to the RNA target containing the minimal PPT into 468 protein isoforms represented by individual protein spots, comparison with protein isoforms identified to bind to the RNA target carrying a triple mutation in the GAR ESE (ESE-) revealed 39 proteins lacking a matching protein spot after inactivation of the GAR ESE (Fig. III-18A). Comparison of these proteins with 35 protein isoforms, which remained unmatched after mutation of two SF2/ASF binding sites (2xSF2-), showed that only 17 of the 39 proteins absent after inactivation of the GAR ESE were specific for this mutation (Tab. III-1). These proteins were considered as candidate

Change		Protein isoforms	
		A5	BPS A5
Absent ESE-/2xSF2-		19	13
Absent ESE-	Unchanged 2xSF2-	17	7
	Reduced 2xSF2-	3	2
Absent 2xSF2-	Unchanged ESE-	13	8
	Reduced ESE-	3	1
Reduced ESE-	Unchanged 2xSF2-	22	-
Reduced 2xSF2-	Unchanged ESE-	7	-
Reduced ESE-/2xSF2-		1	1

**Tab. III-1: Report of protein isoforms binding GAR ESE-mediated to RNA targets carrying the minimal PPT (A5) or to the BPS of extended RNA targets (BPS A5).**



**Fig. III-19: Depiction of protein isoforms hypothesised to regulate exon recognition bound to BPS-lacking and BPS-containing RNA targets isolated for identification by mass spectrometry.**

Protein spots reduced in volume or absent in 2D-SDS-polyacrylamide gels derived from mutant enhancer sequences (cf. Fig. III-18) were isolated from reference gels derived from parental GAR ESE-containing RNA targets in the absence [(A), A5 GAR ESE] or presence of the most proximal BPS [(B), BPS A5 GAR ESE] and analysed by ESI-Quad-TOF mass spectrometry. Schemes of the respective RNA targets are depicted on the right. Spots isolated for MS analyses are encircled in green. Spot numbers (red) were assigned by ImageMaster 2D Platinum 6.0 software (GE Healthcare) during gel analyses.

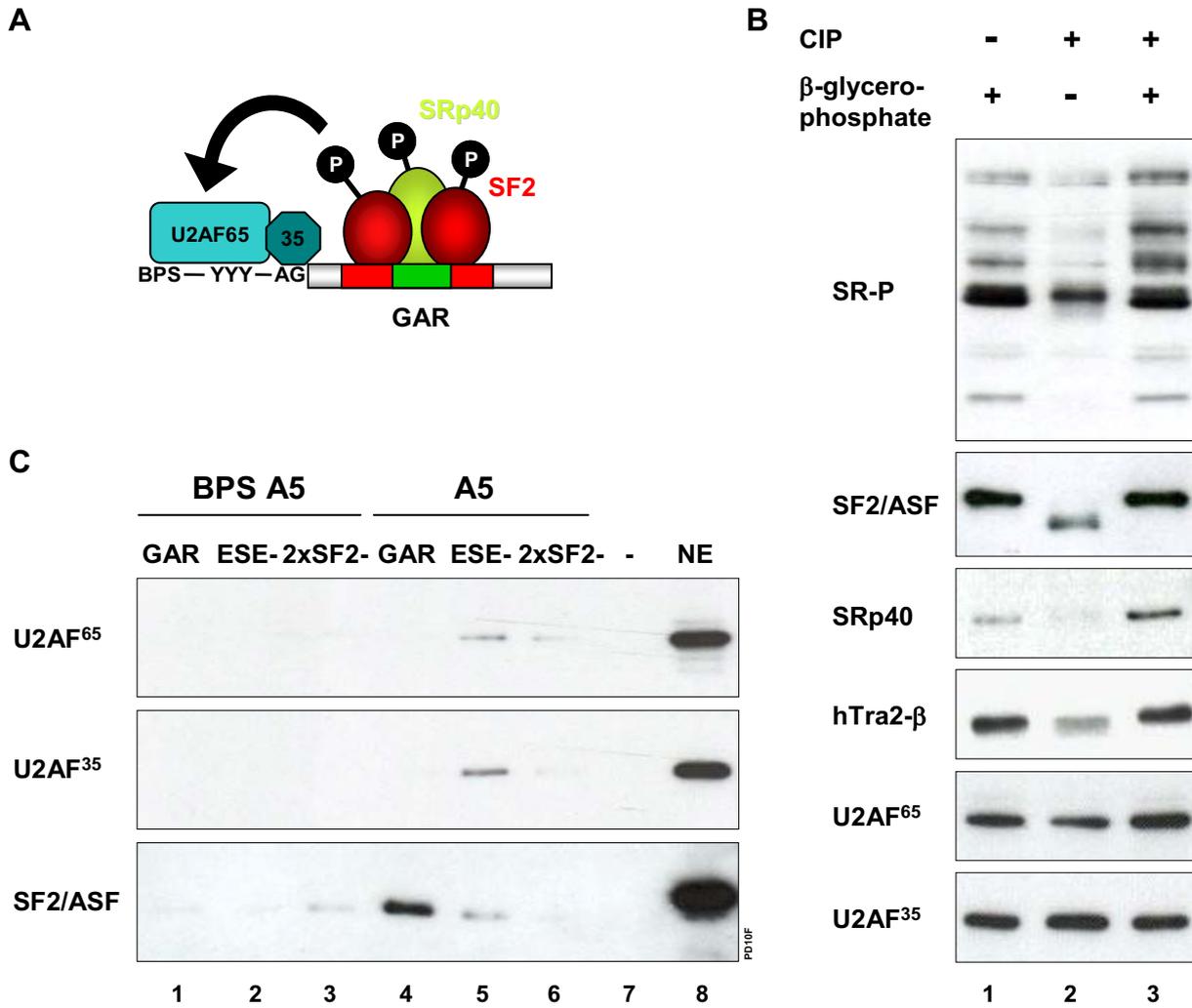
proteins contributing to GAR ESE-mediated exon recognition. Besides the complete loss of protein binding also reduced binding of regulatory proteins might decrease exon recognition. After inactivation of the GAR ESE using the triple mutation (ESE-), 22 protein isoforms binding at least 2.5-fold less to the RNA target were found specific for the inactivation of the GAR enhancer. Proteins characterised by specifically abrogated or diminished binding to the RNA after inactivating the GAR ESE (ESE-) were isolated (Fig. III-18C, MS, and Fig. III-19A) and are currently subjected to mass spectrometry analyses. The first proteins identified from that screen are Nucleolin, GTF II, hnRNP Q and hnRNP H (Fig. III-19A).

In a second experiment, proteins binding to the set of extended RNA targets were analysed by 2D-gel electrophoresis to identify proteins binding GAR ESE-dependent to the 5'-extension. To this end, the two dimensional separation of proteins bound to the BPS-containing RNA target in the presence of the GAR ESE (Fig. III-18B, BPS A5 GAR ESE) was compared to proteins bound to the RNA target lacking the BPS (Fig. III-18A, A5 GAR ESE). In this first comparison, 215 of 544 protein spots identified for the BPS-containing GAR ESE RNA target were found to bind specifically to the extended RNA target (Fig. III-18B, BPS A5 GAR ESE). These proteins were further classified by application of a threshold of 0.04 for the absolute protein volume of the protein spot. The threshold was defined as the lowest absolute protein volume allowing visual identification of the protein spots in coomassie-stained gels thereby enabling their isolation and subsequent sequencing. Application of the threshold narrowed the number of identifiable protein spots down to 41 of the 215 protein spots initially found to bind BPS-mediated to the RNA target. Comparison of the 41 protein spots with their positional counterparts in case the GAR ESE had been mutated yielded 32 putative GAR ESE-dependent protein isoforms binding BPS-mediated to the RNA target. From this group 9 proteins, which were absent exclusively after inactivation of the GAR ESE, might contribute to exon inclusion. The remaining 23 protein spots, which were absent or reduced after mutation of the SF2/ASF binding sites alone or after inactivation of the GAR ESE as well as mutation of the SF2/ASF binding sites, might play only a minor role in exon inclusion, but are likely to contribute to 3' ss selection. The most promising candidates showing differences between intronic sequences followed by wild-type GAR ESE or mutated enhancer sequences (Fig. III-18D, MS and Fig. III-19B) were isolated and are currently analysed by mass spectrometry (Dr. W. Bouschen, BMFZ, HHUD).

### **C.1.3.5 U2AF heterodimer binding to the 3' ss (“complex formation”) requires a phosphorylated factor.**

Phosphorylation of SR proteins has been shown to affect their intracellular localisation (270, 336, 668) and splicing regulatory function [(479, 595, 659, 660), reviewed in (574)]. To investigate whether phosphorylation of SF2/ASF and SRp40 is necessary for their splicing enhancer function in the GAR ESE complex, it was analysed whether the GAR ESE stabilises U2AF<sup>65</sup> binding to the intronic region of 3' ss A5 lacking the BPS also in the presence of dephosphorylated nuclear proteins. If phosphorylation is necessary for the stabilising effect of the GAR ESE, *in vitro* dephosphorylated HeLa nuclear extracts should reduce binding of U2AF<sup>65</sup> to GAR ESE-containing RNA targets. Although U2AF<sup>65</sup> as well as U2AF<sup>35</sup> have been reported to be phosphorylated on four and three serine residues, respectively (204, 402, 451), phosphorylated residues of both U2AF subunits have been found outside of the RRM and with one exception also outside of the regions that constitute the interface for U2AF heterodimerisation (508, 685). Likewise, the binding site of U2AF<sup>65</sup> for SF1/mBBP harbours none of the phosphorylated serines (43, 484). Since in addition a minimal U2AF heterodimer generated using recombinant proteins was shown to bind to RNA (310), it was assumed that dephosphorylation would neither affect U2AF dimerisation nor RNA binding affinity of the U2AF heterodimer in the *in vitro* assay.

HeLa nuclear extracts were dephosphorylated *in vitro* using calf intestinal phosphatase (CIP) and subsequently employed in RNA affinity chromatography of wild-type and mutant GAR ESE sequences initially used to assess the GAR ESE-mediated stabilisation of spliceosomal components (cf. Fig. III-14A). Dephosphorylation of CIP-treated HeLa nuclear extract was confirmed by immunoblotting using a non-phospho-epitope-specific antibody against SF2/ASF, which demonstrated SF2/ASF dephosphorylation by a shift of the protein from a hyperphosphorylated to a low phosphorylated and a non-phosphorylated form (Fig. III-20B, lane 2).



**Fig. III-20: Dephosphorylation abrogates the GAR ESE-mediated stabilisation of U2AF at 3' ss A5.**

(A) Schematic representation of proteins bound to the GAR ESE in exon 5 and the upstream intron during early spliceosomal assembly. Phosphoresidues (P) on RS domains are represented by filled black circles. The questioned interaction of phosphorylated regulatory proteins with the U2AF heterodimer is indicated (black arrow).

(B) HeLa nuclear extract proteins were dephosphorylated *in vitro* using calf intestinal phosphatase (CIP) for 30 min at 37°C and dephosphorylation confirmed in immunoblot analyses using antibodies recognising a common SR protein phospho-epitope (1H4, Zymed) or epitopes specific for SF2/ASF (Zymed), SRp40 (US Biological), hTra2- $\beta$  (Abcam), U2AF<sup>65</sup> (generously provided by Prof. Dr. M. Hastings and Prof. Dr. A. Krainer, Cold Spring Harbor Laboratory, USA) and U2AF<sup>35</sup> (ptglab).

(C) RNA affinity chromatography of RNA targets described in Fig. III-14A using dephosphorylated HeLa nuclear extract (B).

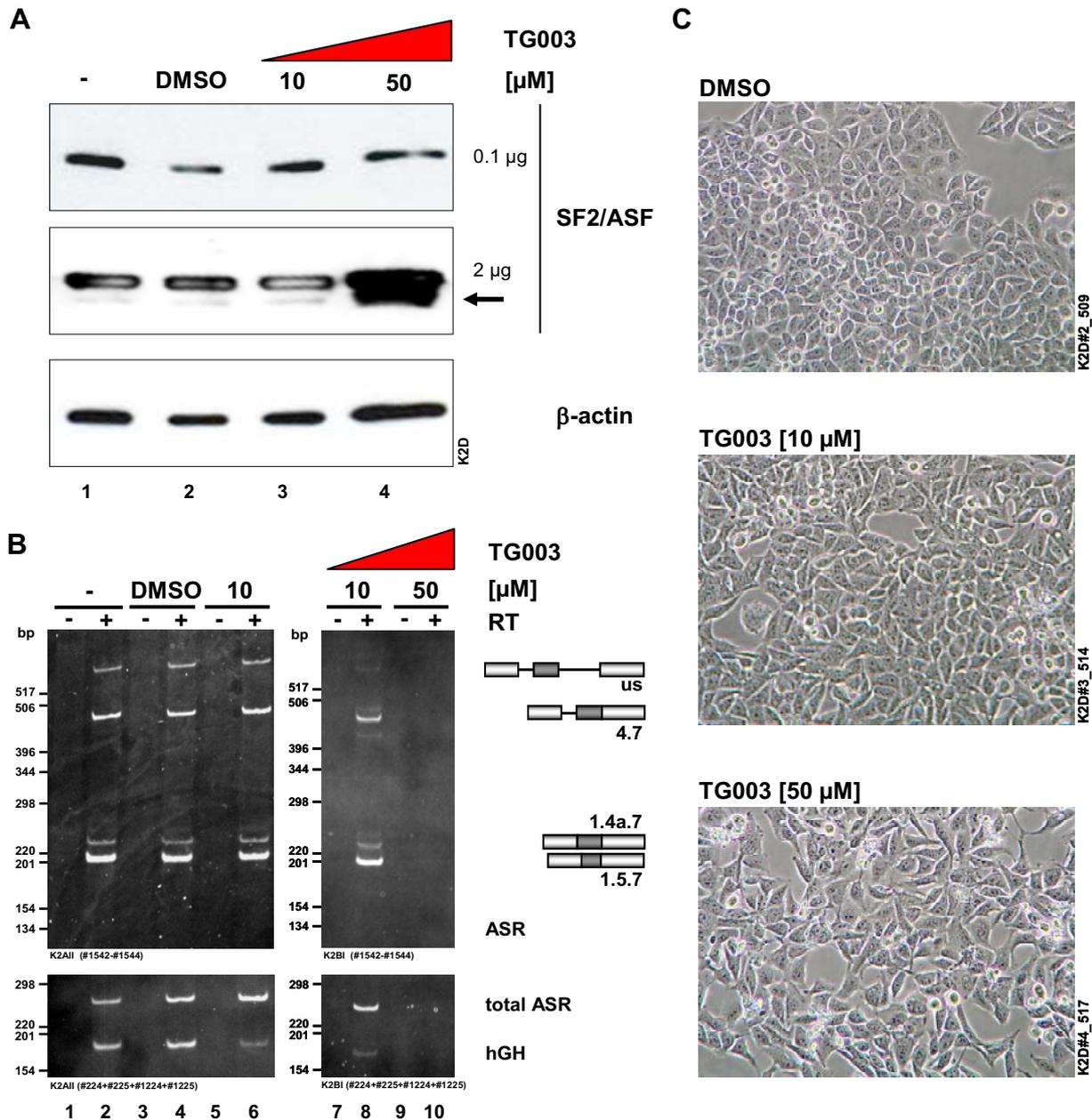
Application of a phospho-epitope-specific antibody also validated SRp40 dephosphorylation. In addition, dephosphorylation of SR-proteins SRp75, SRp55, SRp40, SF2/ASF and SRp20 was analysed with an antibody recognising a phospho-epitope common to several SR proteins revealing that only residual amounts of SR-proteins remain phosphorylated after CIP treatment (Fig. III-20B, SR-P). Analyses of the protein input treated with CIP and the phosphatase inhibitor  $\beta$ -glycerophosphate confirmed that changes in size and signal intensity were caused by CIP-mediated dephosphorylation (Fig. III-20B, lane 3). Identical amounts of both U2AF subunits in mock- and CIP-treated HeLa nuclear extracts were controlled by using U2AF-specific phospho-independent antibodies (Fig. III-20B, U2AF<sup>65</sup> and U2AF<sup>35</sup>).

RNA affinity chromatography using dephosphorylated HeLa nuclear extract demonstrated that binding of SF2/ASF was strongly diminished (Fig. III-20C, SF2/ASF), but still detectable in BPS-lacking RNA targets followed by the wild-type GAR ESE although to a much lower degree compared to its binding to the same RNA target in untreated HeLa nuclear extract (cf. Fig. III-14B, identical amount of HeLa NE loaded as positive control). Therefore, overall binding of this splicing regulatory protein to the RNA targets is diminished after dephosphorylation. However, SF2/ASF binding to extended RNA targets appears to depend more on phosphorylation (cf. Fig. III-14B with Fig. III-20B, lanes 1-3). One explanation for stronger decrease in SF2/ASF binding to extended RNA targets might be a competition between SF2/ASF and hTra2- $\beta$ , which was stabilised on extended RNA targets using untreated HeLa nuclear extracts. Surprisingly, using dephosphorylated HeLa nuclear extracts binding of U2AF<sup>65</sup> as well as U2AF<sup>35</sup> was abolished for all RNA targets containing an extended PPT and also for the RNA target lacking the extension but containing the parental GAR ESE (Fig. III-20C, lanes 1-4), which all showed the highest amount of U2AF binding in untreated HeLa nuclear extract (Fig. III-14B, lanes 1-4). In contrast, binding of U2AF<sup>65</sup> to short RNA targets carrying mutations in the GAR ESE remained unchanged, raising the question whether the low level of U2AF<sup>65</sup> bound to short RNA targets carrying mutations in the GAR ESE using untreated HeLa nuclear extract might have been caused by generation of an artificial polypyrimidine-rich region serving as suboptimal low-affinity U2AF<sup>65</sup> binding site due to a number of nucleotide substitutions to polypyrimidines (Fig. III-14A, ESE- and 2xSF2-). The corresponding increase in U2AF<sup>35</sup> binding to this short RNA target (Fig. III-20C, lane 5) might indicate that the U2AF heterodimer is needed to occupy this low-affinity binding site. In summary, these results demonstrated that efficient binding of

both U2AF subunits to short as well as extended RNA targets depends on phosphorylated proteins in HeLa nuclear extract. Therefore, it is suggested that the GAR ESE-mediated stabilisation of U2AF<sup>65</sup> is executed by phosphorylated domains of SF2/ASF, SRp40 and hTra2- $\beta$ .

#### **C.1.3.6 The impact of splicing regulatory protein phosphorylation on the function of the GAR ESE complex *in vivo***

To elucidate *in vivo* the relevance of SR protein phosphorylation on GAR ESE function, it was aimed to analyse the splicing pattern of the minigene splicing reporter in the presence of dephosphorylated SR proteins. The benzothiazole compound TG003 was reported to abrogate phosphorylation of recombinant SF2/ASF through inhibition of Clk1/Sty kinase activity (431). To examine whether GAR ESE-mediated exon inclusion or 3' ss selection depend on SF2/ASF phosphorylation, the splicing pattern of the 2-intron minigene was analysed in HeLa-T4<sup>+</sup> cells treated with TG003 (kindly provided by Dr. C. Ehlers and Prof. J. Hauber, Heinrich-Pette-Institut, Hamburg). The membrane-permeable substance was applied into the culture media in the recommended concentration of 10  $\mu$ M and in the five-fold increased concentration of 50  $\mu$ M, respectively, 3 h post transfection with the minigene. In parallel assays either proteins or total RNA were prepared 30 h post transfection. Immunoblotting 0.1  $\mu$ g total cellular proteins with an SF2/ASF-specific antibody failed to detect any changes in the phosphorylation of SF2/ASF in the presence of 10  $\mu$ M and 50  $\mu$ M TG003 (Fig. III-21A, upper panel), which would have resulted in a faster migrating protein isoform (cf. Fig. III-20B). The immunoblot was also probed against  $\beta$ -actin confirming equal protein loading (Fig. III-21A, lower panel). Increasing the protein amount for the immunoblot analysis to 2  $\mu$ g hampered the linear detection of the hyperphosphorylated SF2/ASF protein signal, because the strong bands immediately exhausted the peroxidase substrate giving rise to "burn out" ghost bands (Fig. III-21A, middle panel, lanes 1-3). Analysing the increased protein input confirmed the unchanged phosphorylation status of SF2/ASF in the presence of 10  $\mu$ M TG003 by constant low amounts of less phosphorylated SF2/ASF. The immunoblot analysis contrasts the observation that this concentration had been reported to inhibit Clk1/Sty activity *in vitro*, to rescue the redistribution of phosphorylated SR proteins into the nucleoplasm, which is induced by overexpression of recombinant Clk1/Sty in HeLa cells and to reduce SR protein



**Fig. III-21: TG003, an inhibitor of Clk1/Sty activity, only weakly affects SF2/ASF phosphorylation and fails to change the splicing pattern of the 2-intron minigene.**

**(A)** Immunoblot analyses of SF2/ASF phosphorylation using 0.1 μg (upper panel) and 2 μg total cellular protein (middle panel) applying a phospho-epitope-independent monoclonal antibody (Zymed). The faster migrating hypophosphorylated SF2/ASF protein is indicated by an arrow. Equal protein loading was controlled by detection of β-actin (Sigma-Aldrich) (lower panel).

**(B)** RT-PCR analysis of total RNA expressed from the 2-intron minigene in the presence of TG003. HeLa-T4<sup>+</sup> cells were transiently cotransfected with SV leader SD1 SA5 env nef and pXGH5 to control transfection efficiency. Cells were treated with 10 μM or 50 μM TG003 dissolved in DMSO or DMSO only as solvent control. Untreated cells served as negative control (-). The splicing pattern in the presence of 50 μM TG003 was detected beyond the linear amplification range using 30 PCR cycles (lanes 7 to 10) to verify the loss of plasmid-derived RNA signals.

**(C)** Phase contrast images of TG003- or control-treated HeLa-T4<sup>+</sup> cells (magnification x 160).

phosphorylation in COS-7 cells (431). Raising the concentration of TG003 to 50  $\mu\text{M}$  increased the level of dephosphorylated SF2/ASF and since the hyperphosphorylated SF2/ASF isoform was not able to exhaust the peroxidase substrate, it has to be assumed that the level of hyperphosphorylated SF2/ASF is at best slightly reduced (Fig. III-21A, middle panel, lane 4). However, even at this high compound concentration the level of phosphorylated SF2/ASF still outweighed the amount of the less phosphorylated protein isoform. Nevertheless, the increased concentration of TG003 strongly influenced cell morphology (Fig. III-21C, TG003 [50 $\mu\text{M}$ ]) and increased cell death suggesting that TG003 treatment at this concentration severely affects cell physiology. Consistent with the SF2/ASF phosphorylation pattern in the immunoblot analyses, splicing of the 2-intron minigene did not change in the presence of 10  $\mu\text{M}$  TG003 (Fig. III-21B, cf. lane 6 with lane 4). Increasing the concentration of TG003 to 50  $\mu\text{M}$  completely abolished expression of both transfected minigenes, i.e. the 2-intron alternative splicing reporter (ASR) as well as the reporter used for transfection control (hGH) (Fig. III-21B, lane 10).

Due to the fact that the compound TG003 did not sufficiently reduce phosphorylation of SF2/ASF in a non-toxic concentration range, a second potential inhibitor of SF2/ASF phosphorylation, the indole derivative Drug29 (personal communication and by courtesy of Prof. J. Tazi, IGM Montpellier, France), was tested for its ability to reduce the steady-state phosphorylation level of SF2/ASF *in vivo*. However, this substance completely failed to effect SF2/ASF phosphorylation as tested in immunoblot assays even at higher concentrations than recommended (data not shown).

Since the question whether phosphorylation of SF2/ASF is critical for GAR ESE function *in vivo* could not conclusively solved by application of kinase inhibitors, other experimental assays have to be established to achieve levels of dephosphorylated SF2/ASF sufficient to provide significant results.

#### **C.1.4 Summary: A functional network emanating from exon 5 and flanking splice sites regulates inclusion of exon 4c, a, b and 5.**

In exon 5 of the HIV-1 pre-mRNA our research group identified a guanosine-adenosine-rich exonic splicing enhancer (GAR ESE). The GAR ESE activated either an enhancer-dependent 5' ss or a 3' ss in the presence of the SR proteins SF2/ASF or SRp40 in respective subgenomic 1-intron constructs. In this thesis, the GAR ESE was characterised regarding its function for internal exon recognition in the subgenomic sequence context of the internal exon 5, which is flanked by a cluster of alternatively used 3' ss and by 5' ss D4.

It was found that the GAR ESE is essential for activation of each 3' ss of the internal cluster. In this process the GAR enhancer fulfils a dual splicing regulatory function by (i) synergistically activating exon recognition by the individual SR protein binding sites and (ii) by specifically conferring 3' ss selectivity within the alternatively used 3' ss cluster through the proximal SF2/ASF binding sites. 3' ss selectivity mediated by the GAR ESE depends on the stability of U1 snRNP binding at 5' ss D4 suggesting that interactions across the internal exon activate the 3' ss cluster. The E42 sequence located between the GAR enhancer and 5' ss D4 appears to be essential for formation of these cross-exon interactions, since E42 deletion almost completely abrogates splicing at the 3' ss cluster. Activation of the 3' ss cluster, which is mediated by the GAR ESE and binding of U1 snRNP at the 5' ss D4, remains essential for processing of the intron-containing *vpu/env*-mRNA in the intermediate and late phase of viral gene expression.

Elucidating the GAR ESE mechanism, it was found that the enhancer stabilises binding of U2AF<sup>65</sup> upstream of 3' ss A5 thereby likely mediating 3' ss selectivity. Although phosphorylation of splicing factors appears to promote binding of the U2AF heterodimer to the 3' ss *in vitro*, the relevance of SR protein phosphorylation on GAR ESE function *in vivo* remains to be validated. Searching for factors mediating GAR ESE-dependent exon recognition, stabilisation of SF1/mBBP, a component of the E complex, was not confirmed. Application of a more general approach to detect proteins mediating exon recognition identified hTra2- $\beta$  as another splicing regulatory protein binding to the GAR ESE. Therefore, it was further investigated whether the number of proteins involved in GAR ESE function had been underestimated. However, hTra2- $\beta$  binding still increases upon extension of the PPT pointing to an additional binding site within extension or to the stabilisation of hTra2- $\beta$  via an additional protein binding to the

extended PPT. By the use of two-dimensional gel electrophoresis, additional proteins (i) binding GAR ESE-mediated to the RNA or (ii) binding to the extended PPT in dependency of the GAR ESE, thereby possibly acting as adaptor protein, were isolated and are currently analysed by mass spectrometry. Among the first proteins identified in this screen are hnRNP H and hnRNP Q, two members of the hnRNP family, whose contribution to internal exon recognition will be evaluated in further investigations.

Since GAR ESE-mediated activation of the downstream located 5' ss and the upstream located 3' ss cluster is essential to generate almost all spliced mRNA isoforms during viral gene expression, the GAR ESE substantially contributes to the regulation of viral replication.

## C.2 Expression of unspliced RNA during late HIV-1 gene expression

During late viral gene expression disuse of all splice sites of the HIV-1 pre-mRNA allows Rev-mediated nucleo-cytoplasmic export of unspliced genomic RNA, which serves as template for the translation of the Gag-Pol precursor protein and also as genomic RNA that is packaged into progeny virus. Based on the insights into the activation of splice sites located centrally in the HIV-1 pre-mRNA the second part of this thesis investigated mechanisms circumventing viral pre-mRNA splicing during the late phase of viral gene expression.

The rationale to investigate regulatory mechanisms allowing the expression of genomic RNA originates from the analyses of HIV-1 mRNA expression kinetics revealing that genomic viral RNA is expressed with a delay of approximately 12 hours compared to *env*-mRNAs (312). Although expression of both of these intron-containing RNA isoforms had been shown to depend on the presence of the viral regulatory protein Rev, their differential expression kinetics pointed to an additional regulatory mechanism besides the Rev-mediated nucleo-cytoplasmic export, which is a prerequisite for the expression of all intron-containing HIV-1 RNA isoforms (168, 179, 388).

In this part of the thesis, two parameters of the viral pre-mRNA metabolism, i.e. 5' splice site efficiency and the impact of the RNA inhibitory element in the 5'-region of the *gag*-gene, were examined, which are conceivable to contribute to the regulated expression of completely unspliced viral RNA during the course of host cell infection and which are not mutually exclusive.

In the first case, it was assumed that the amount of completely unspliced RNA increases in the late phase, because splicing efficiency at 5' splice site D1 declines. Decreased splicing efficiency at D1 might be mediated either by a splicing enhancer activating D1 in the early phase or by a splicing silencer inhibiting D1 in the late phase of viral gene expression. Therefore, in transient transfection experiments 5' splice site D1 and the neighbouring exonic and intronic sequences were analysed regarding the presence of *cis*-acting splicing regulatory elements. The efficiency of D1 might be decreased not only by host cell splicing regulatory proteins but also by the interaction of viral proteins with flanking RNA sequences. In particular, it was reported that the viral structural protein Gag binds to the RNA leader via its nucleocapsid component (8, 150, 208, 209) and that the presence of Gag might increase the amount of unspliced RNA [C.K. Damgaard, PhD thesis (142)]. Therefore, it was also investigated whether Gag

regulates D1 splicing. To this end, the influence of the precursor Gag and its subunits, respectively, on the amount of unspliced RNA was assessed in transient transfection assays.

As a second alternative mechanism for the expression of viral genomic RNA, it was postulated in this thesis that unspliced RNA might be constantly generated during viral gene expression, but is degraded during the early phase by destabilising elements in the first intron. The mechanism, however, preventing degradation of the viral genomic RNA in the late phase of viral gene expression is still not conclusively resolved. Degradation of unspliced RNA might be counteracted by cellular proteins induced by the virus in the late phase.

### **C.2.1 Identification of *cis*-acting splicing regulatory elements in the proximity of D1**

The 5' ss D1 is used in each of the more than 40 alternatively spliced HIV-1 mRNAs. In accordance with its frequent usage, D1 is classified as an intrinsically strong 5' ss by different algorithms [splicefinder (<http://www.uni-duesseldorf.de/rna/>), MaxEntScore ([http://genes.mit.edu/burgelab/maxent/Xmaxentscan\\_scoreseq.html](http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html)) (669)]. Previous work from our group had demonstrated that in the absence of an upstream located splicing enhancer U1 snRNP does not efficiently bind to D1 in a subgenomic 1-intron *env*-expression transcript. Therefore, it was hypothesised that splicing at D1 in its authentic position in the viral pre-mRNA is also supported by a splicing enhancer.

#### **C.2.1.1 Individual SR protein binding sites overlapping D1 differentially affect D1 efficiency and alternative 3' ss selection.**

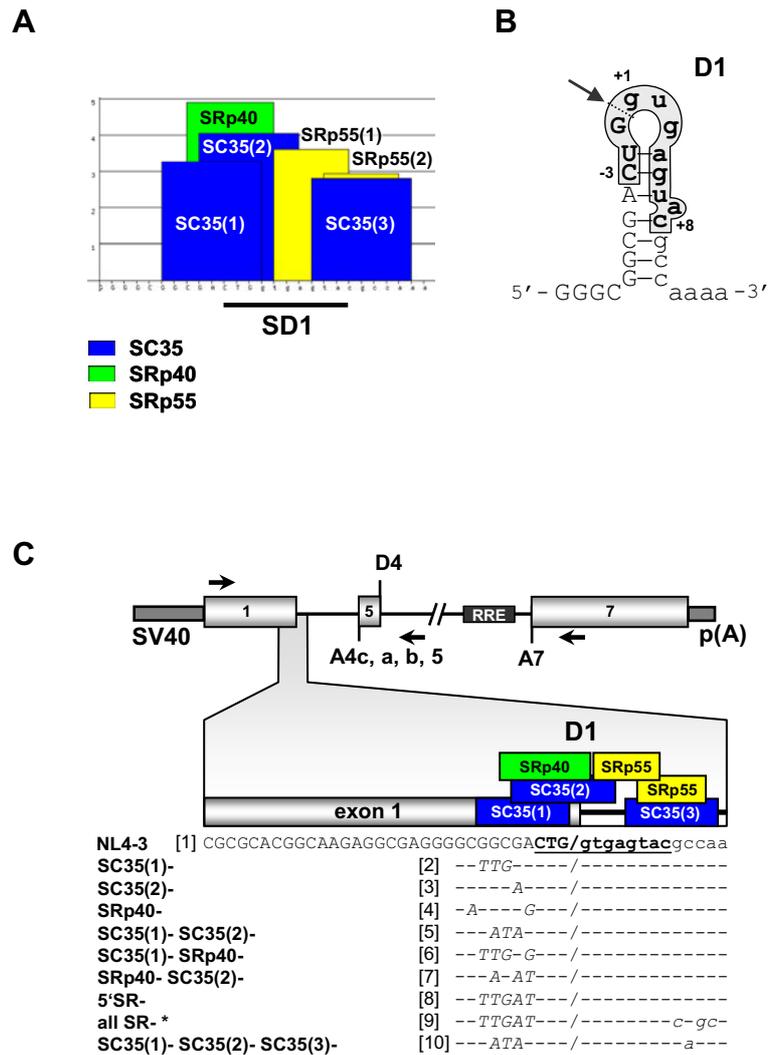
To elucidate, if splicing regulatory elements are located in the immediate proximity of D1, this region was analysed *in silico* for the presence of putative SR protein binding sites (ESEfinder, ver. 3.0 [<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home>] (97, 566)). Six SR protein binding sites were predicted to partially or completely overlap the D1 sequence (Fig. III-22A). Three of these binding sites for SC35 and SRp40 partially cover D1 in the upstream flanking region, whereas an SRp55 and an additional SC35 binding site overlap D1 in the flanking downstream region. A second SRp55 binding site is entirely located within the D1 sequence [SRp55(1)]. To analyse

**Fig. III-22: Localisation of potential SR protein binding sites overlapping D1.**

**(A)** Using the default values of the ESEfinder algorithm (ver. 3.0) (97, 566), six binding sites for the SR proteins SC35, SRp40 and SRp55 were predicted in and overlapping the D1 sequence denoted at the bottom. The y-axis describes the score value, which was appointed by the ESEfinder algorithm, reflecting the match of a predicted SR protein binding site with the respective experimentally determined consensus binding motif (97, 566).

**(B)** In the highly structured leader of the HIV-1 pre-mRNA D1 and its neighbouring sequences form a stem loop structure, in which the exon/intron border is located in the apical loop (arrow, dashed line) (10, 116, 511). The nucleotides from position -3 in the exon (capital letters) to +8 in the intron (lower case letters) are able to contribute to U1 snRNA binding (highlighted in grey).

**(C)** Individual and combinatorial mutations of the SR protein binding sites predicted in (A) were introduced into the 2-intron minigene SV leader SD1 SA5 env nef. The exon 1/intron-border containing D1 is enlarged. Positions of SR protein binding sites predicted by the ESEfinder algorithm are indicated. Exonic sequences are denoted in capital letters, whereas intronic sequences are indicated in lower case letters. Nucleotide exchanges (italic letters) for each mutated minigene are given below the parental sequence. In the "all SR-" minigene (\*) all SR protein binding sites except for the central SRp55 binding site [SRp55(1)] were mutated, because mutation of this site would have altered the intrinsic strength of D1. The numbering of the individual mutants corresponds to the lanes in Fig. III-23 [SV40: SV40 early promoter, p(A): SV40 early polyadenylation site, RRE: Rev-responsive element].



whether the potential SR protein binding sites indeed contribute to 5' ss activation, all five SR protein binding sites only partially overlapping D1 were mutated in the 2-intron minigene (Fig. III-22C, all SR-\* [9]). By introducing point mutations outside of D1 its intrinsic strength remained unaltered. Due to its position, the SRp55 binding site located within the D1 sequence [SRp55(1)] could not be mutated without affecting the intrinsic strength of D1. Therefore, the SRp55(1) binding site remained intact in the following experiments.

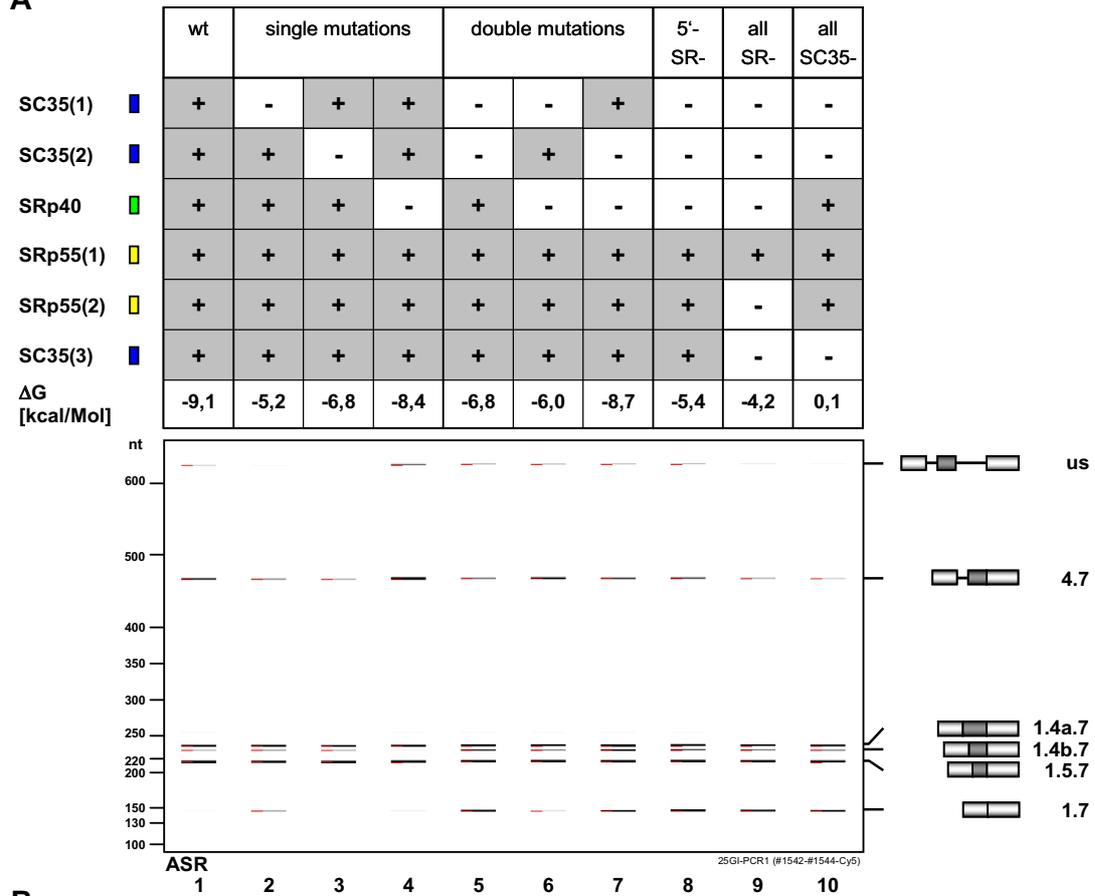
After transient transfection of HeLa-T4<sup>+</sup> cells with the parental 2-intron minigene, RT-PCR analysis revealed that the majority of mRNAs were spliced at D1 including one of the alternative internal exons 4a, 4b or 5 (1.4a.7, 1.4b.7, and 1.5.7), whereas only 8% of the mRNAs remained unspliced at D1 (us and 4.7) (Fig. III-23A and B, lane 1). Mutating all SR protein binding sites overlapping D1 decreased the expression of mRNA isoforms that remained unspliced at D1 to 1.4% (Fig. III-23A and B, cf. lanes 1 and 9). This decrease correlated with an increased expression of mRNA isoforms

**Fig. III-23: Individual SR protein binding sites overlapping D1 differentially contribute to suppression of D1 usage and 3' ss activation.**

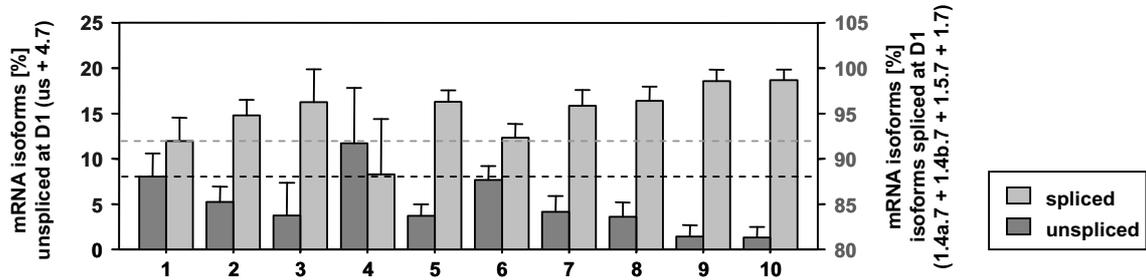
**(A)** RT-PCR analysis of HeLa-T4<sup>+</sup> cells transiently transfected with the parental and mutant minigenes depicted in Fig. III-22C. Individual and combinatorial SR protein binding site mutations were introduced into the 2-intron minigene SV leader SD1 SA5 env nef resulting in stem loops with different thermodynamic stabilities ( $\Delta G$ , upper panel).  $2 \times 10^5$  HeLa-T4<sup>+</sup> cells were transiently transfected with 1  $\mu$ g of the parental 2-intron minigene or mutant derivatives thereof. Total RNA was isolated 30 h after transfection. mRNA was reverse transcribed using oligo(dT) as primer. PCR products were labelled using Cy5-modified primers #1544 for the amplification of alternatively spliced reporter RNA (ASR) and #1225 for RT-PCR of *hGH*-mRNA. PCR products were separated on denaturing urea-polyacrylamide gels and detected by ALF. Fluorescence data were processed and a virtual gel electrophoresis image was generated by using P2 software (Pharmacia) (lower panel). The exon/intron structure of the respective mRNA isoforms is depicted at the right. Integrated fluorescence signals of bands labelled with red bars exceed the threshold (set as 100), which distinguishes PCR product signals from background noise.

**(B-D)** Relative quantification of mRNA isoforms unspliced or spliced at D1 (B), 1.4b.7-mRNA (C) and 1.7-mRNA (D) after SR protein binding site mutations at D1. Relative mRNA amounts were calculated as percentage of the respective mRNA isoforms relative to total mRNA isoforms (sum of spliced and unspliced mRNA isoforms) in each sample. Dashed lines mark the expression of the respective mRNA isoform from the parental 2-intron minigene. The mean values of three independent experiments are shown with error bars representing standard deviation.

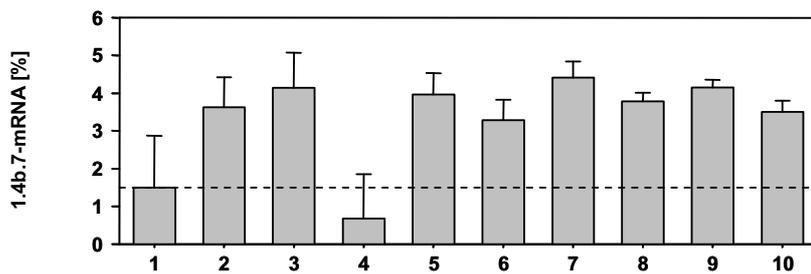
A



B



C



D

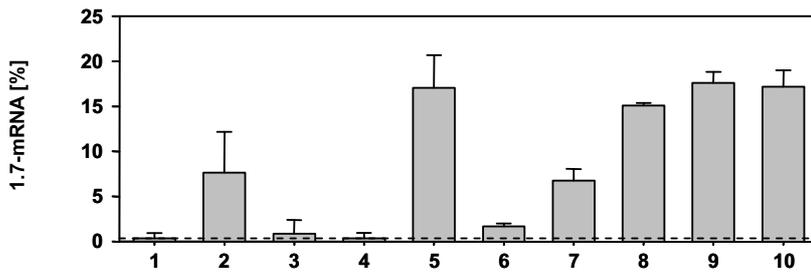


Fig. III-23: continued.

spliced at D1, which was unexpectedly caused by mRNAs either including the internal exon 4b (1.4b.7) or excluding any internal exon (1.7) (Fig. III-23A, C and D, cf. lanes 1 and 9), whereas the expression of mRNA isoforms including either the internal exon 4a or 5 was unaffected (see Appendix, Fig. VII-3). These results suggested that the SR protein binding sites overlapping D1 rather reduce than enhance the efficiency of this 5' ss thereby supporting the expression of mRNAs unspliced at D1. This unexpected finding contrasts the more prevalently observed splicing activating function of SR proteins. However, suppression of 5' ss usage by the SR protein binding sites analysed might be conceivable to emerge from competition of SR proteins with the U1snRNA for binding at D1.

To investigate whether the five SR protein binding sites overlapping D1 support expression of mRNA isoforms unspliced at D1 to a different extent, the effect of these binding sites was analysed by individual and combinatorial SR protein binding site mutations in the 2-intron minigene (Fig. III-22C and Fig. III-23A, lanes 2-8 and lane 10). Transient transfections of HeLa-T4<sup>+</sup> cells with the parental or mutant minigene constructs followed by RT-PCR of total RNA showed that mutations of the SR protein binding sites differentially reduced the level of unspliced RNA (Fig. III-23B). Individual mutation of the SR protein binding sites overlapping D1 from the exon revealed that the SC35 binding site located proximally to D1 [SC35(2)] inhibited D1 efficiency stronger than the more upstream located SC35 binding site [SC35(1)] – supporting the competition model –, whereas the SRp40 binding site appeared to even support D1 activation (Fig. III-23B, lanes 2-4). Introducing simultaneous mutations in two SR protein binding sites confirmed the stronger impact of SC35(2) on 5' ss usage, because both minigenes carrying mutations in the SC35(2) binding site expressed less RNA unspliced at D1 than the mutant carrying mutations in the more upstream located SC35(1) and the SRp40 binding site (Fig. III-23B, lanes 5-7). However, simultaneous mutation of two SR protein binding sites overlapping D1 from the exon seemed to act non-additive on D1 efficiency. Even mutating all three exonic SR protein binding sites overlapping D1 did not further decrease the amount of RNA unspliced at D1 observed after the individual mutation of the SC35(2) binding site (Fig. III-23B, cf. lanes 3 and 8). Only the additional mutation of SR protein binding sites overlapping D1 from the intron further reduced the amount of RNA unspliced at D1. The additional mutation of the third SC35 binding site covering D1 was sufficient for this reduction and resulted in the same level of RNA unspliced at D1 as the mutation of all five SR protein binding sites partially overlapping

with D1 (Fig. III-23B, cf. lanes 9 and 10). From these results it was concluded that the SR protein binding sites partially covering D1 reduce 5' ss efficiency to a different extent. The proximally located upstream SC35(2) binding site reduces 5' ss activation the most, whereas the second exonic SC35 binding site exerts a lower impact on D1 efficiency. The SRp40 binding site appeared to even support D1 usage. However, an additive effect of SR protein binding sites overlapping D1 was only observed for simultaneous mutations of exonic and intronic SR protein binding sites.

Unexpectedly, mutations of the SR protein binding sites differentially affected the expression of the 1.4b.7- and 1.7-mRNA isoforms. The relative amount of 1.4b.7-mRNA moderately corresponded with the increase in mRNAs spliced at D1 (Fig. III-23, cf. B and C). In contrast, expression of 1.7-mRNA did not correlate with changes in the expression of mRNA spliced at D1. For generation of 1.7-mRNA the most upstream located SC35 binding site [SC35(1)] was the only SR protein binding site, whose individual mutation induced the expression of mRNA excluding the internal exon (Fig. III-23A and D, cf. lane 2 with 3 and 4). Additional mutation of the second exonic SC35 binding site [SC35(2)] appeared to synergistically increase the exclusion of the internal exons (Fig. III-23A and D, lane 5). Combining the mutation of one of the SC35 binding sites in exon 1 with the mutation of the potential binding site for SRp40 showed a contrary impact of the SRp40 binding site on the amount of RNA excluding any internal exon. In combination with the mutation of the most upstream SC35 binding site [SC35(1)] mutation of the SRp40 binding site reduced exon exclusion compared to the sole mutation of SC35(1) (Fig. III-23A and D, cf. lane 2 with 6). However, in combination with the mutation of the proximal binding site for SC35 in exon 1 [SC35(2)] mutation of the SRp40 binding site further pronounced exon exclusion compared to sole mutation of SC35(2) (Fig. III-23A and D, cf. lane 3 with 7). Although simultaneous mutation of the two exonic SC35 binding sites appeared to synergistically activate splicing from D1 to A7, additional mutation of the SRp40 binding site did not further enhance the amount of 1.7-mRNA (Fig. III-23D, lane 8). Likewise, additional mutations of SR protein binding sites overlapping D1 from the intron did not increase the amount of 1.7-mRNA (Fig. III-23D, lane 9 and 10). Overall, both SC35 binding sites covering D1 from the exon exert the most repressive influence on the generation of 1.7-mRNA. To this repressive effect on internal exon exclusion the most upstream located SC35 binding site contributed stronger than the more proximally located SC35 binding site.

In summary, increasing amounts of 1.4b.7-mRNA correlated with an increase in the expression of mRNA isoforms unspliced at D1 suggesting that both expression changes are linked to the same SR protein binding sites overlapping D1. In contrast, the appearance of 1.7-mRNA did not correlate with the expression changes of mRNAs unspliced at D1. Since the impact of individual SR protein binding sites differed between the generation of 1.7- and 1.4b.7-mRNA, this result suggested that in the parental sequence generation of 1.7-mRNA excluding any internal exon is suppressed by mechanisms distinct from those reducing 5' ss D1 efficiency.

Structure analyses employing chemical and enzymatic probing (36, 116, 243, 511) and NMR spectroscopy (10) revealed that the D1 sequence folds into a stem loop (Fig. III-22B). A recent study reported that the stability of the D1 RNA stem loop structure impairs the functionality of this 5' ss (2). In the mutational analysis of the SR protein binding sites presented here neither changes in the expression level of mRNAs unspliced at D1 nor the generation of 1.7-mRNA excluding any internal exon correlated with the calculated stability of the resulting D1 stem loop structures (Fig. III-23A,  $\Delta G$ ). Exclusion of the internal exons was already induced with stem loops predicted to fold with stabilities similar to the wild-type stem loop (Fig. III-23A and D, cf. lane 1 with 7). In addition, mutations yielding similar hairpin stabilities induced considerably different exon exclusion ratios (Fig. III-23A and D, cf. lane 3 with 5) and comparable high exon exclusion ratios were caused by sequences, which possessed almost the highest difference in the predicted stability of the hairpin structure (Fig. III-23A and D, cf. lane 5 with 10). Thus, thermodynamic stability of the RNA hairpin structure is not likely to cause the overall increased splicing at D1 and the generation of 1.7-mRNA in the mutated minigene context.

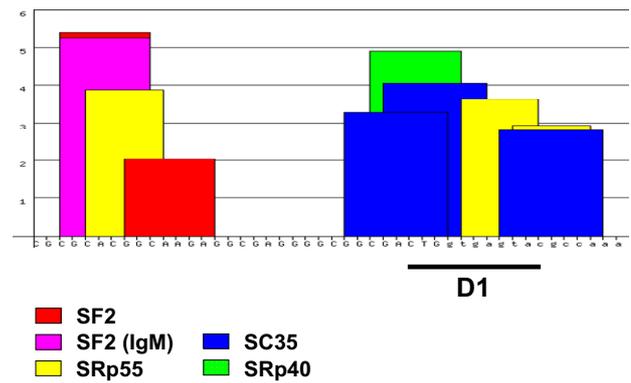
To investigate whether the examined SR protein binding sites overlapping D1 cooperate with the more upstream flanking exonic region in supporting internal exon recognition and D1 suppression, the sequence upstream of D1 was additionally analysed with respect to predicted SR protein binding sites. Using the ESEfinder algorithm, additional two SF2/ASF binding sites and one SRp55 binding site were predicted (Fig. III-24A). To examine whether these SR protein binding sites further reduce D1 efficiency and support internal exon recognition, they were mutated in the 2-intron minigene carrying mutations in the three exonic SR protein binding sites overlapping D1 (Fig. III-24B). The effect of SR protein binding site mutations on the splicing pattern of the minigene was analysed in transiently transfected HeLa-T4<sup>+</sup> cells. However, RT-PCR analysis revealed

**Fig. III-24: SR protein binding sites A**  
**in the leader sequence upstream of D1 do not contribute to the suppression of D1 usage.**

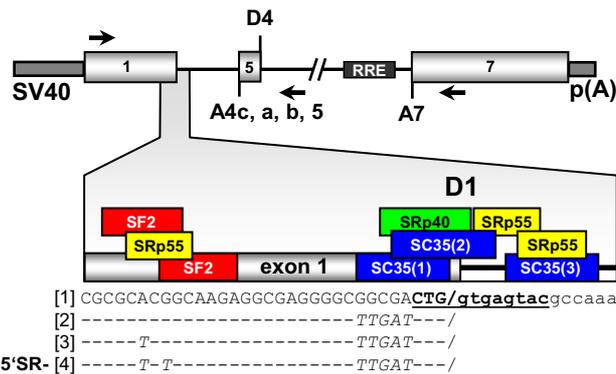
**(A)** Depiction of binding sites for SR proteins SF2/ASF and SRp55 predicted upstream of D1 using the ESEfinder algorithm (ver. 3.0) (97, 566). The most upstream SF2/ASF binding sites is predicted by two different score matrices applied by the ESEfinder (red, magenta). The D1 sequence is underlined.

**(B)** Mutations introduced upstream of D1 inactivating either the SRp55 or all three predicted SR protein binding sites far upstream of D1 in addition to the 5'SR- mutation next to D1 in the 2-intron minigene. Respective mutations are denoted in italic letters [RRE: Rev-responsive element].

**(C)** RT-PCR analysis of the parental 2-intron minigene and mutants derivatives depicted in (B). HeLa-T4<sup>+</sup> cells were transiently transfected with SV leader SD1 SA5 env nef (lanes 1 and 2), SV leader SD1 5'SR- SA5 env nef (lanes 3 and 4), SV leader (-28)SRp55- SD1 5' SR- SA5 env nef (lanes 5 and 6) or SV leader (-30)SF2- (-28)SRp55- (-25)SF2- SD1 5' SR- SA5 env nef (lanes 7 and 8) and pXGH5 to control equal transfection efficiency. Total RNA was isolated 30 h post transfection and mRNA reverse transcribed using oligo(dT). PCR was performed using primers #1542-#1544 for alternatively spliced reporter mRNA (ASR) and #1224/#1225 for hGH-mRNA (hGH). The exon/intron structure of the alternatively spliced mRNA isoforms is indicated at the right. Asterisks mark heteroduplex PCR products as verified by sequencing.

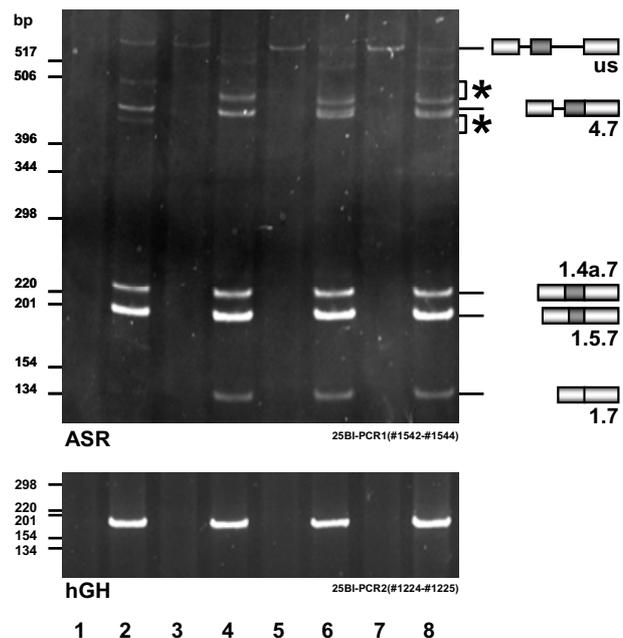


**B**



**C**

wt		5' SR-				splice site	
-		-		SRp55-	2xSF2-SRp55-	upstream exon	
-	+	-	+	-	+	-	+
							RT



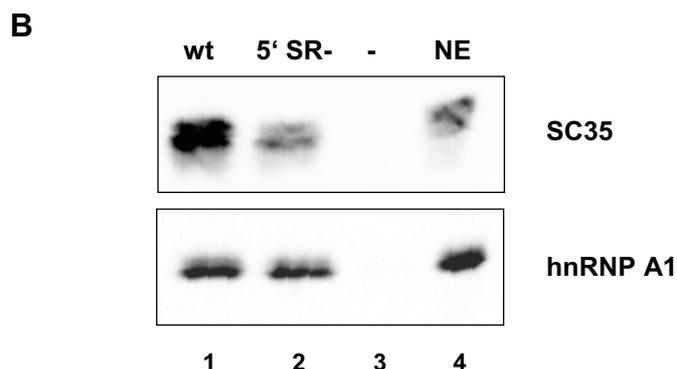
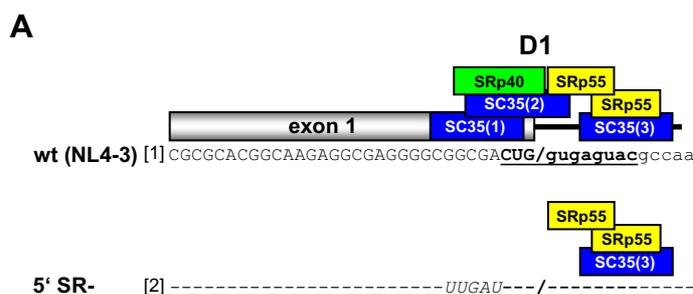
that the additional mutations of the upstream exonic SR protein binding sites did neither decrease the amount of mRNA isoforms unspliced at D1 nor affect the expression of 1.7-mRNA (Fig. III-24C, cf. lane 2 with 3 and 4). Therefore, additional binding sites for SF2/ASF and SRp55 located in the upstream region of D1 are not involved in the regulation of D1 efficiency.

#### **C.2.1.2 Tethering SC35 into an upstream exonic position enhances D1 usage.**

In the mutational analysis of SR protein binding sites overlapping D1 a strong decrease in the expression of mRNA isoforms unspliced at D1 and the strongest exclusion of the internal exons was already caused by simultaneous mutation of the predicted exonic SC35 binding sites [SC35(1) and SC35(2)] and could only be slightly intensified by additional mutations in the 2-intron minigene. To test whether decreased exon recognition correlates with reduced SC35 binding at D1, binding of SC35 to the parental RNA sequence derived from the viral laboratory strain NL4-3 was compared to binding to the RNA sequence carrying mutations in the three exonic SC35 and SRp40 binding sites upstream of D1 *in vitro* (Fig. III-25, 5'SR-). This combination of SR protein binding site mutations induced one of the most prominent exon exclusion ratios in the 2-intron minigene *in vivo* (Fig. III-23, 5'SR- [8]). Immunoblot analysis of proteins isolated using RNA affinity chromatography confirmed binding of SC35 to exon 1 (Fig. III-25B, upper panel, lane 1). Compared to the NL4-3 sequence the mutated sequence bound significantly less SC35 (Fig. III-25B, upper panel, lane 2) and thereby validated the position of the predicted SC35 binding sites at D1. Remaining SC35 bound to the RNA after mutation of the SR protein binding sites suggested that the most downstream predicted SC35 binding site overlapping D1 [SC35(3)] might also be functional. Binding of hnRNP A1, a protein that binds to RNA of sufficient length, served as loading control (Fig. III-25B, lower panel).

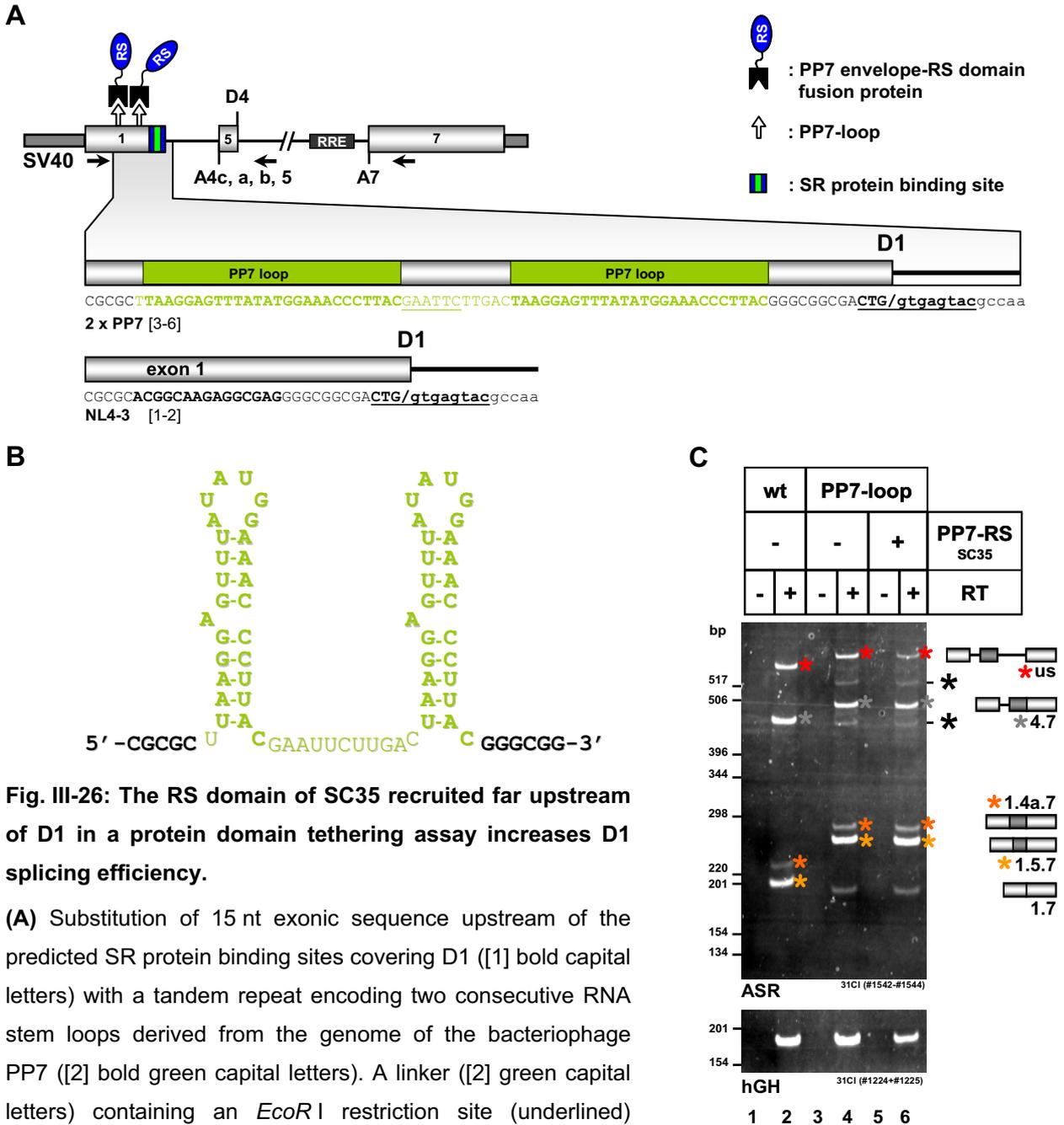
**Fig. III-25: SC35 binds to the predicted binding sites in the 5'-half of D1.**

**(A)** RNA sequences used as targets for affinity chromatography of HeLa nuclear proteins. RNAs containing 32 nt upstream and 13 nt downstream of the exon1/intron-border containing the parental exonic sequence derived from the laboratory strain NL4-3 [1] or the mutant sequence [2] were transcribed *in vitro*. Mutation of five nucleotides (italic letters) upstream of the D1 sequence (bold, underlined) abrogates the prediction of the three SR protein binding sites for SC35 and SRp40 overlapping D1 from the exon.



**(B)** Immunoblot analysis of proteins isolated by RNA affinity chromatography using RNA targets described in (A). 2700 pmol of parental or mutant RNA were immobilised and incubated with 60% HeLa nuclear extract. A negative control was performed lacking RNA (-). Proteins were eluted from the RNA in 100  $\mu$ L 2 x protein sample buffer. 28% of the isolated proteins were separated on 10% SDS-polyacrylamide gels and blotted onto PVDF membrane. HeLa nuclear extract served as positive control (NE). Proteins were detected using antibodies recognising SC35 (by courtesy of Prof. M. Caputi, Florida Atlantic University, USA) (upper panel) or hnRNP A1 (Santa Cruz Biotechnology) (lower panel).

To investigate whether SC35 is able to promote splicing at D1 when positioned more distantly from the 5' ss and thereby cannot interfere with U1 snRNA binding, the splicing pattern of the 2-intron minigene was analysed in the presence of the splicing regulatory RS domain of SC35 tethered further upstream of D1. To this end, the leader region of D1 was substituted with a tandem repeat of a hairpin structure of the bacteriophage PP7 RNA (Fig. III-26A and B). During phage assembly the hairpin structure is selectively bound by the PP7 coat protein. This RNA-protein interaction was employed to position the RS domain of SC35 upstream of D1 by generating a fusion protein of the PP7 coat protein and the SC35 RS domain. HeLa-T4<sup>+</sup> cells were transiently cotransfected with the 2-intron minigene construct encoding two PP7 hairpin loops in exon 1 and a plasmid expressing the PP7 coat-SC35 RS domain fusion protein. RT-PCR analysis of the 2-intron minigene carrying the PP7 loops instead of the D1 leader



**Fig. III-26: The RS domain of SC35 recruited far upstream of D1 in a protein domain tethering assay increases D1 splicing efficiency.**

(A) Substitution of 15 nt exonic sequence upstream of the predicted SR protein binding sites covering D1 ([1] bold capital letters) with a tandem repeat encoding two consecutive RNA stem loops derived from the genome of the bacteriophage PP7 ([2] bold green capital letters). A linker ([2] green capital letters) containing an *EcoRI* restriction site (underlined) connects both PP7 loops [RRE: Rev-responsive element].

(B) Secondary structure model of the genetically fused tandem PP7 stem loops. An adenosine bulges out from each stem. A translational start codon is located in each loop [adapted from (354)].

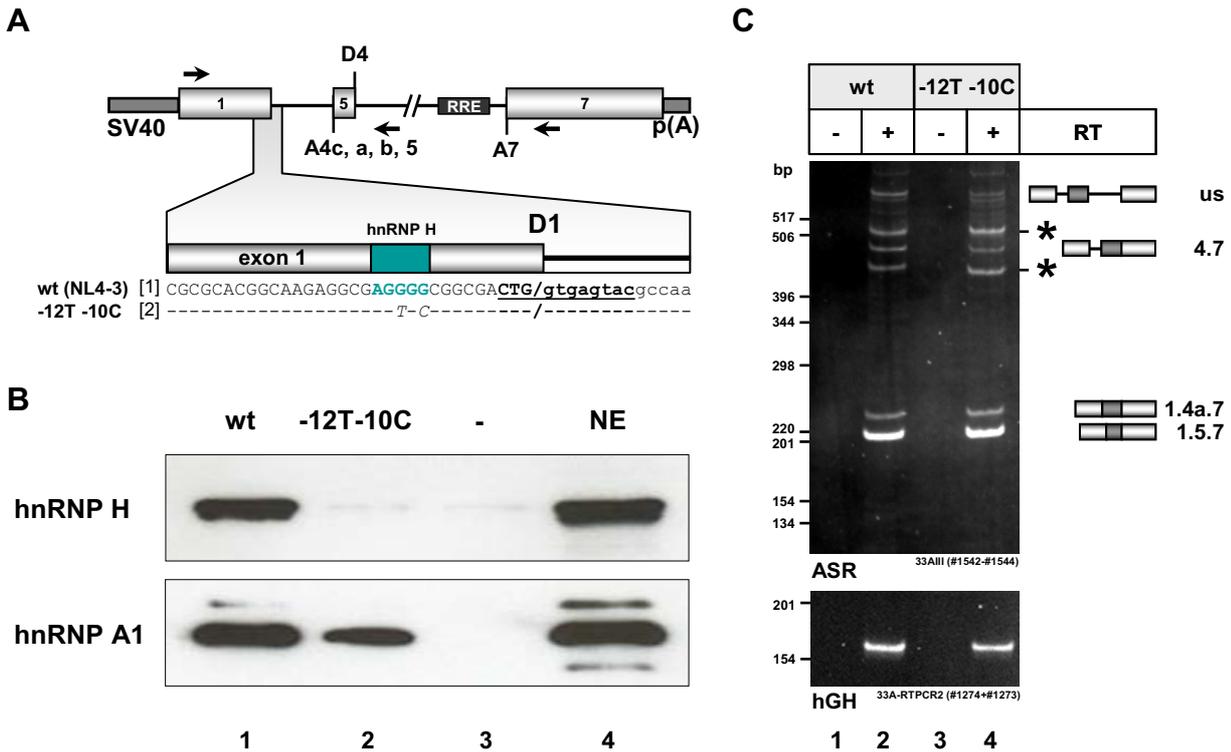
(C) RT-PCR analysis of HeLa-T4<sup>+</sup> cells transiently transfected with the parental 2-intron minigene SV leader SD1 SA5 env nef (lane 1 and 2) or the mutant minigene SV leader 2xPP7 SD1 SA5 env nef containing PP7 loops as illustrated in (A) (lanes 3-6) and pXGH5 as transfection control. Cells were additionally cotransfected with 1 µg SV SD4/SA7 NLS-PP7 SC35 expressing a PP7 coat-SC35 RS domain fusion protein (PP7-RSSC35, lane 5 and 6). The exon/intron structure of alternatively spliced mRNA isoforms is indicated at the right. Insertion of the PP7 loops expands the size of exon 1, thus increasing the size of each mRNA isoform by 49 bp. Identically spliced mRNA isoforms are indicated by coloured asterisks. Black asterisks mark heteroduplex PCR products [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, RT: reverse transcription, us: unspliced].

---

sequence revealed that the loop insertion induced the expression of the mRNA isoform excluding the internal exons (1.7) (Fig. III-26B, cf. lane 2 with 4). Cotransfection of the PP7-RS domain-encoding plasmid did not change the expression of the 1.7-mRNA isoform, but reduced the amount of completely unspliced RNA (us) (Fig. III-26B, lane 6). Thus, it was concluded that the RS domain of SC35, although not sufficient to abrogate the expression of 1.7-mRNA excluding the internal exons, in a more distant exonic position increased D1 splicing efficiency as monitored by the reduced amount of completely unspliced RNA. Since SC35 influenced D1 efficiency in opposite directions depending on the distance of the regulatory domain from the 5' ss, these findings further support the model that SR protein binding sites covering D1 reduce D1 efficiency by competing with U1 snRNA for binding at D1.

### **C.2.1.3 hnRNP H binding at exon 1 does not influence splicing at D1.**

The hnRNP proteins constitute a family of splicing regulatory proteins predominantly reported to inhibit pre-mRNA splicing [reviewed in (396)]. The sequence upstream of D1 contains a motif (AGGGG) matching the consensus binding site of the hnRNP H protein [DGGGD, where D is A, G, or U (93, 521)], which had already been implicated in splicing regulation (81, 94, 113, 133, 241, 248, 282, 394, 497, 521). To investigate whether hnRNP H regulates splicing at D1, the potential binding site in the leader sequence, which had not been affected by the mutations introduced so far, was mutated by two nucleotide substitutions at position -12 and -10 with respect to the exon 1/intron-border (Fig. III-27A).



**Fig. III-27: hnRNP H binding in exon 1 does not influence splicing at D1.**

**(A)** Mutation of the hnRNP H binding site upstream of D1 in the 2-intron minigene. The sequence AGGGG at the position -14 to -10 (depicted in blue) with respect to the exon/intron border matches the consensus motif identified to bind proteins of the hnRNP H family (93, 521). The predicted hnRNP H binding site was inactivated by two point mutations disrupting the stretch of four guanosines (-12T -10C [2]). Mutated nucleotides are denoted in italic letters [RRE: Rev-responsive element].

**(B)** RNA affinity chromatography of D1 and flanking sequences. RNAs encoding wild-type [1] or mutated sequences [2] enlarged in (A) were in vitro transcribed and 900 pmol of each RNA were covalently bound to adipic acid agarose beads. A negative control was performed omitting RNA in the binding reaction (-). After incubation in 60% HeLa nuclear extract for 30 min at 37°C in the presence of phosphatase inhibitors, beads were extensively washed and proteins eluted in 12.5 µL 2 x protein sample buffer. 8% of the eluted proteins were separated and immunoblotted as described in Fig. III-25B. Proteins were detected using antibodies recognising hnRNP H (generously provided by Prof. D.L. Black, University of California, Los Angeles, USA) and hnRNP A1 (clone 9H10, Santa Cruz Biotechnology). HeLa nuclear extract (NE, 5.6 µg) served as positive control.

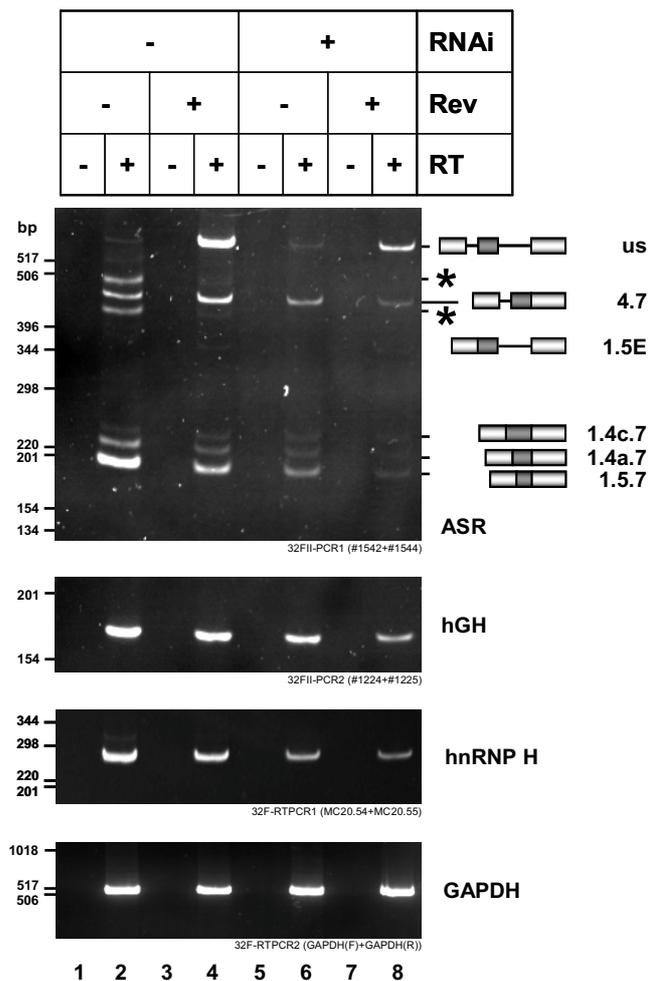
**(C)** RT-PCR analysis of the 2-intron minigene carrying the mutated hnRNP H binding site compared with the parental 2-intron minigene. HeLa-T4<sup>+</sup> cells were transiently transfected with 1 µg SV leader SD1 SA5 env nef (lane 1 and 2) or 1 µg SV leader -12T-10C SD1 SA5 env nef (lane 3 and 4) and 1 µg pXGH5 serving as transfection control. RT-PCR was essentially performed as described in Fig III-24C except that for amplification of hGH-mRNA primer pair #1273/#1274 was used. The exon/intron structure of PCR products is illustrated at the right. Asterisks mark heteroduplex PCR products [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, RT: reverse transcription, us: unspliced].

To confirm that mutations introduced into the 2-intron minigene affect hnRNP H binding, protein binding to the sequence upstream of D1 was examined by RNA affinity chromatography *in vitro*. Immunoblot analysis of proteins isolated using the same parental RNA target region, which was already applied for the analyses of SC35 binding (Fig. III-25A), revealed that hnRNP H strongly binds to the NL4-3 sequence and is reduced to the level of the negative control after introducing the point mutations at position -12 and -10 (Fig. III-27B, upper panel, cf. lane 1 with 2). However, RT-PCR analysis of total RNA isolated from HeLa-T4<sup>+</sup> cells transiently transfected with the parental or mutant 2-intron minigene revealed no difference in the splicing pattern of the 2-intron minigene after mutation of the hnRNP H binding site (Fig. III-27C, cf. lane 2 with 4). Therefore, binding of hnRNP H to exon 1 does not influence splicing at D1.

To rule out that binding of hnRNP H to an yet unidentified site obscures changes in the splicing pattern, the amount of *trans*-acting cellular hnRNP H was reduced by using RNA interference (RNAi). To this end, cellular hnRNP H expression was impaired by transient transfection of a short hairpin RNA (shRNA) complementary to the *hnRNP H*-mRNA (kindly provided by Prof. M. Caputi, Florida Atlantic University, USA). RT-PCR analysis of total RNA isolated 54 hours after simultaneous transfection of HeLa-T4<sup>+</sup> cells with the hnRNP H-targeting shRNA and the 2-intron minigene revealed an overall lower expression level of the transfected plasmids, the 2-intron splicing reporter as well as the minigene expressing *hGH*-mRNA, serving as transfection control (Fig. III-28). Nevertheless, taking into account the reduced transfection efficiency, expression levels of mRNA isoforms containing the proximal intron (us and 4.7) remained unchanged after reduction of cellular hnRNP H expression in the absence as well as in the presence of the viral regulatory protein Rev (Fig. III-28, cf. lane 2 with 6 and lane 4 with 8). However, after hnRNP H knockdown a strong decrease in the expression of mRNA isoforms including the internal exons 4a and 5 was observed (1.4a.7/1.5.7), whereas inclusion of exon 4c appeared to be activated (Fig. III-28, cf. lane 2 with 6). These results provided further evidence that hnRNP H does not regulate splicing at D1 in the 2-intron minigene but clearly contributes to inclusion of the internal exons 4a and 5 into the mRNA.

**Fig. III-28: Knockdown of hnRNP H expression does not affect D1 usage in the 2-intron minigene, but reduces internal exon inclusion.**

RT-PCR analysis of HeLa-T4<sup>+</sup> cells cotransfected with the 2-intron reporter and an shRNA targeting *hnRNP H*-mRNA. Cells were transiently transfected with the 2-intron minigene SV leader SD1 SA5 env nef (lanes 1-8) and cotransfected with SVcrev where indicated (Rev, lane 3-4 and lane 7-8) or a control plasmid. Cells were simultaneously transfected with 4 µg shRNA targeting *hnRNP H*-mRNA where indicated (RNAi, lanes 3-8). In a preliminary experiment the time course of *hnRNP H*-mRNA expression after transfection of HeLa-T4<sup>+</sup> cells with shRNA was examined (see B.2.3.4). Since expression of *GAPDH*-mRNA requires 54 h to recover from shRNA transfection, cytoplasmic RNA for RT-PCR analyses of shRNA-treated cells was harvested after the prolonged incubation of 54 h. RNA was reverse transcribed using oligo(dT) and cDNA amplified using primers #1542-#1544 for alternatively spliced reporter RNA (ASR), primer pair #1224/#1225 for *hGH*-mRNA (hGH), primer pair MC20.54/MC20.55 for *hnRNP H*-mRNA (hnRNP H) and primer pair GAPDH(F)/GAPDH(R) for *GAPDH*-mRNA (GAPDH). PCR products were resolved on 6% polyacrylamide gels and stained using ethidium bromide. The exon/intron structure of alternatively spliced mRNA isoforms of the 2-intron minigene is shown at the right. Asterisks mark heteroduplex PCR products [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, RT: reverse transcription, us: unspliced].



amplified using primers #1542-#1544 for alternatively spliced reporter RNA (ASR), primer pair #1224/#1225 for *hGH*-mRNA (hGH), primer pair MC20.54/MC20.55 for *hnRNP H*-mRNA (hnRNP H) and primer pair GAPDH(F)/GAPDH(R) for *GAPDH*-mRNA (GAPDH). PCR products were resolved on 6% polyacrylamide gels and stained using ethidium bromide. The exon/intron structure of alternatively spliced mRNA isoforms of the 2-intron minigene is shown at the right. Asterisks mark heteroduplex PCR products [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, RT: reverse transcription, us: unspliced].

### C.2.1.4 Identification of additional proteins binding to exon 1 and D1

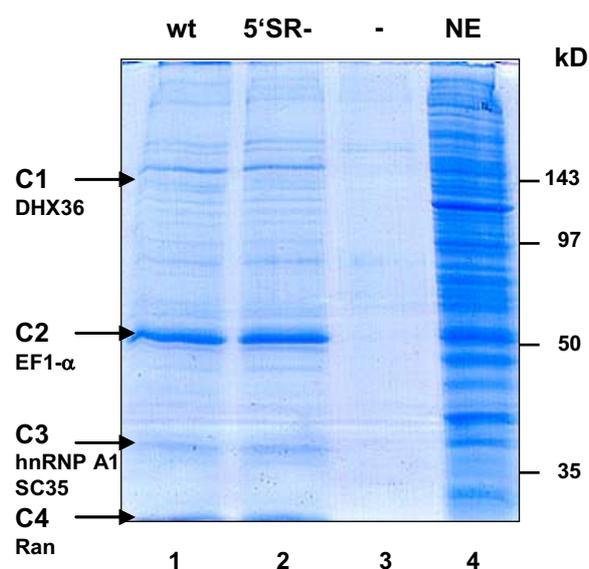
Since inhibition of 1.7-mRNA expression could neither be attributed to the RS domain of SC35 nor to hnRNP H, additional proteins binding to exon 1 and D1 were isolated by RNA affinity chromatography and subsequently identified by mass spectrometry. The D1 sequence and additional 5 nt downstream and 26 nt upstream were used as RNA target (Fig. III-25A, NL4-3). Proteins binding to the NL4-3 RNA sequence were

compared to those binding to an RNA target carrying mutations in the three exonic SR protein binding sites overlapping D1 (Fig. III-25A, 5'SR-). After gel electrophoretic separation and coomassie staining the mutated exon 1 sequence showed only a minor difference in the protein bands compared to the NL4-3 sequence (Fig. III-29). Mass spectrometry of conspicuous protein bands binding to the wild-type RNA identified the proteins DHX36, EF1- $\alpha$  and Ran. The number of proteins binding to the NL4-3 RNA was further extended by Nucleolin, which was identified in an additional separation (data not shown). Additionally peptides were identified corresponding to SC35 and hnRNP A1 thereby further confirming binding of SC35 to the D1 RNA target as it was already demonstrated in this thesis (Fig. III-25B, upper panel). Although hnRNP A1 has been reported to act as splicing regulatory protein, the previous experiment revealed equal binding to the parental and the mutant RNA sequence (Fig. III-25B, lower panel). The remaining proteins identified by mass spectrometry are implicated in RNA-related processes like transcription [DHX36 (278)], RNA degradation [DHX36 (615)] as well as RNA stabilisation [Nucleolin (456, 538)], translation [EF-1 (303)] and transport [Ran (368)]. However, none of these proteins could be linked to the splicing machinery and its regulators. Therefore, confirmation of their binding by western blot was not pursued.

**Fig. III-29: Identification of additional proteins binding to exon 1.**

Affinity chromatography of RNA targets containing 5' ss D1 and the flanking sequence derived from the viral laboratory strain NL4-3 (wt) or containing mutations in the SR protein binding sites overlapping D1 from the exon (5'SR-). RNA targets as depicted in Fig. III-25A were transcribed *in vitro* and 1000 pmol covalently bound to adipic acid agarose beads. A negative control was performed by omitting RNA in the reaction (-). Immobilised RNA targets were incubated with 60% HeLa nuclear extract for 20 min at 30°C. Proteins were eluted from the RNA in 60  $\mu$ L 2x protein sample buffer.

25% of the protein eluate were separated in a 10% SDS-polyacrylamide gel and stained with coomassie brilliant blue G250 using a modified protocol allowing subsequent mass spectrometry of isolated bands (299). Indicated bands were isolated from the gel, digested using trypsin and after peptide purification analysed using an ESI-Quad-TOF (Dr. N. Wiethölter, BMFZ, HHUD).



### C.2.1.5 Sequences surrounding D1 do not affect D1 activation, but modulate downstream 3' ss selection and distal intron removal.

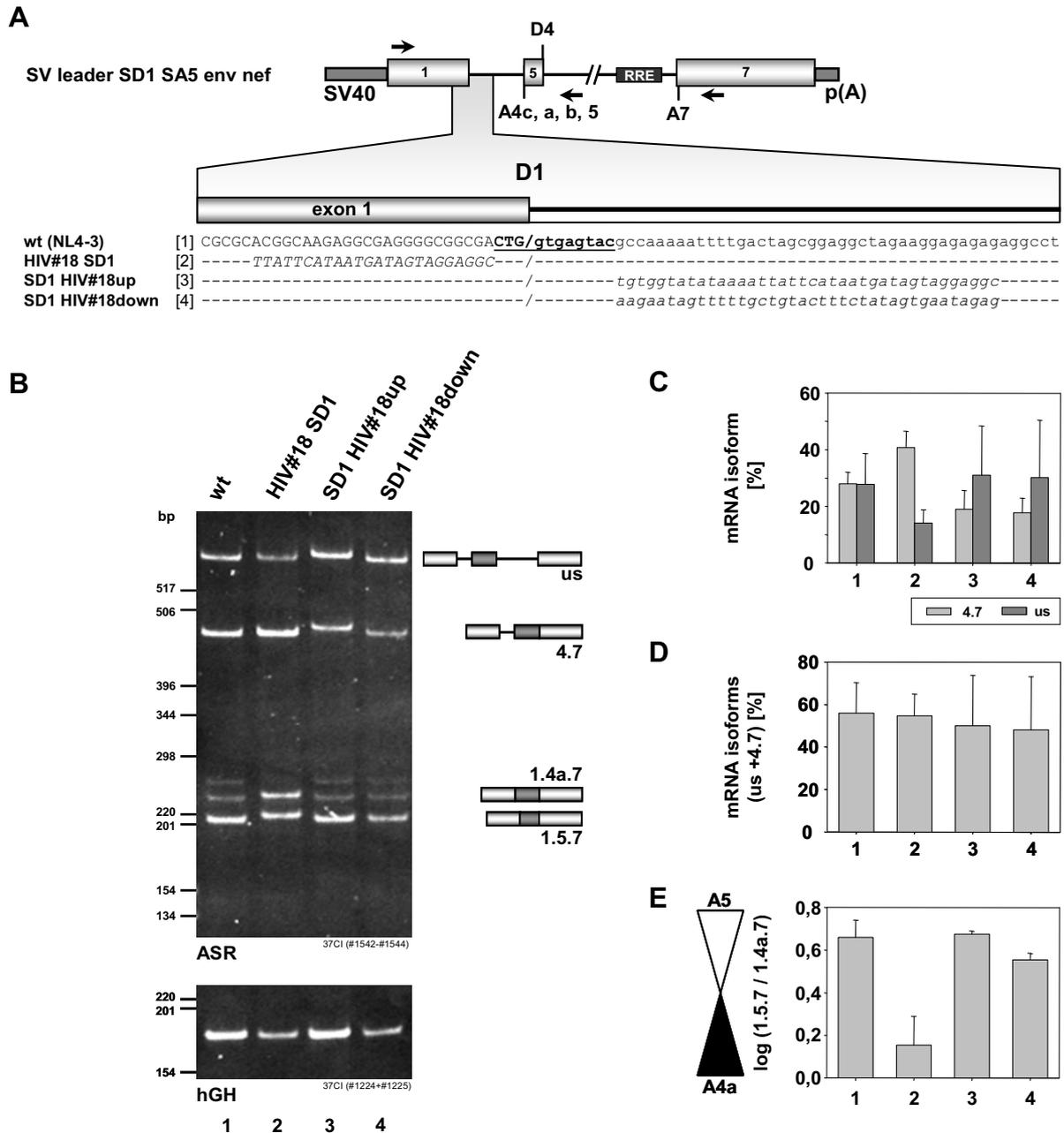
In an alternative, more general approach it was investigated whether the sequences flanking D1, which had not been affected by the mutations analysed so far, contain splicing regulatory elements modulating 5' ss efficiency. To this end, the flanking exonic 24 nt and intronic 38 nt, respectively, were substituted in the 2-intron minigene with an HIV-1 control sequence (HIV#18), which had previously been shown not to contain a splicing enhancer activity (91) (Fig. III-30A). As had been seen in the previous experiments, the parental 2-intron minigene already expresses a low level of intron-containing and completely unspliced RNA, therefore allowing the identification of enhancing as well as silencing splicing regulatory elements surrounding D1. RT-PCR analysis of transiently transfected HeLa-T4<sup>+</sup> cells revealed that substitution of the exonic

#### **Fig. III-30: D1 neighbouring sequences affect internal 3' ss selection and distal intron removal.**

**(A)** Illustration of sequence substitutions upstream and downstream of D1 in the 2-intron minigene. The 24 nt located upstream of D1 were substituted with a control sequence of identical length preceding a non-functional HIV-1 splice site (HIV#18 [2]) (91). The 38 nt of the 5'-UTR downstream of D1 were substituted with an extended version of the same control sequence (D1 HIV#18up [3]) or a sequence originating from downstream of the HIV#18 site (D1 HIV#18down [4]). Mutated nucleotides are shown in italic letters. The 11 nt of 5' ss D1 are denoted in bold letters and underlined [RRE: Rev-responsive element].

**(B)** RT-PCR analysis of HeLa-T4<sup>+</sup> cells transfected with the parental 2-intron minigene or mutants carrying substitutions of the D1 flanking sequences as shown in (A).  $2 \times 10^5$  cells were transiently transfected with 1  $\mu$ g SV leader SD1 SA5 env nef (lane 1), 1  $\mu$ g SV leader HIV#18 SD1 SA5 env nef (lane 2), SV leader SD1 HIV#18up SA5 env nef (lane 3) or SV leader SD1 HIV#18down (lane 4) and additionally cotransfected with 1  $\mu$ g pXGH5 to control transfection efficiency. RNA isolation and RT-PCR were performed as described in Fig. III-24C. Although some minor differences in the length of the PCR products are visible in the native gel electrophoresis, RT-PCR analysis under denaturing conditions performed for quantification (C-E) confirmed the identity of the PCR products (data not shown) [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, us: unspliced].

**(C)** Relative quantification of individual mRNA isoforms unspliced at D1. Transient transfection experiments were performed as described in (B). RT-PCR products were labelled using Cy5-labelled primers #1544 for alternatively spliced reporter RNA and #1225 for amplification of *hGH*-mRNA. Denaturing gel electrophoresis of PCR products was performed as described in Fig. III-23A. Expression of both mRNA isoforms unspliced at D1 (4.7 and us) were calculated relative to total mRNA expression from the respective minigene. Error bars represent the standard deviation from the mean of three independent experiments.



**Fig. III-30: continued.**

**(D)** Relative quantification of total mRNA isoforms unspliced at D1. Raw data collected in (C) were analysed to show the percentage of all mRNA isoforms unspliced at D1 relative to total alternatively spliced mRNA isoforms expressed in each sample. Errors bars represent standard deviation from the mean of three independent experiments.

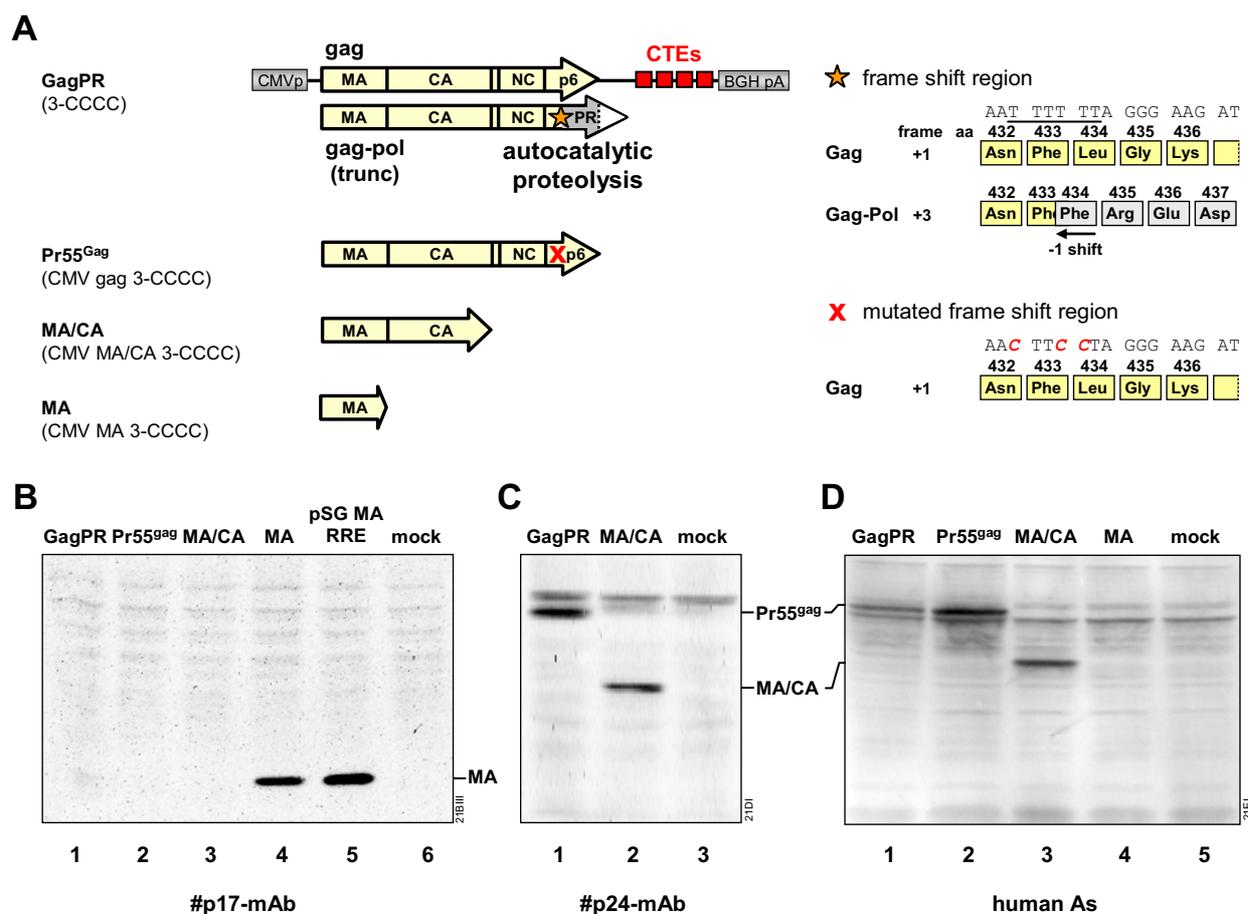
**(E)** Quantification of 3' ss selectivity in doubly spliced mRNA. Raw data resulting from the transfection experiments shown in (C) were analysed regarding selectivity of the internal 3' ss, which was calculated as ratio of doubly spliced 1.5.7- to 1.4a.7-mRNA. Error bars represent standard deviation.

region upstream of D1 increased the amount of RNA unspliced at D1 (4.7) from 28% to 41%, but at the same time decreased the amount of completely unspliced RNA (us) from 28% to 14% (Fig. III-30B and C, HIV#18 SD1, lane 2). In contrast, substitution of the intronic sequence with the same control sequence slightly decreased the amount of RNA unspliced at D1 (4.7) to 19% and increased the amount of completely unspliced RNA (us) to 31% (Fig. III-30B and C, SD1 HIV#18up, lane 3). Substitution of the downstream flanking region with a second control sequence confirmed these results (Fig. III-30B and C, SD1 HIV#18down, lane 4). Although the upstream as well as the downstream flanking region of D1 modulated expression of both mRNA isoforms unspliced at D1, the overall amount of mRNAs unspliced at D1 remained unaffected (Fig. III-30D). Unexpectedly, substitution of the 24 nt upstream of D1 also equally activated 3' ss A4a and A5 for inclusion of the internal exon (Fig. III-30E, lane 2). This was not observed for any substitution of the sequence downstream of D1. Therefore, substitution of the flanking exonic and intronic regions affected the splicing pattern of the 2-intron reporter construct, however, alterations in the splicing pattern were not due to changes in D1 activation, but in the 3' ss chosen by D1 and the splicing efficiency of the downstream located intron defined by 5' ss D4 and 3' ss A7. These effects on 3' ss selectivity and splicing of the downstream intron might hint to an involvement of D1 neighbouring sequences in splicing kinetics.

#### **C.2.1.6 The viral Gag protein does not affect D1 activation.**

Experiments analysing the requirements for packaging of HIV-1 genomic RNA into virions revealed that the viral protein Gag binds to the intronic region downstream of D1 (8, 150, 208, 209). A continuative investigation analysing the binding properties of the Gag polyprotein precursor and its nucleocapsid subunit to the packaging signal in the viral pre-mRNA reported a Gag precursor protein-mediated increase in the expression of a reporter gene located in an intronic position [C.K. Damgaard, PhD thesis (142)]. This finding suggested a role for the Gag-RNA interaction in viral mRNA processing. Based on these observations it was examined whether Gag increases the amount of RNA unspliced at D1 during the late phase of viral replication.

Gag expression plasmids encoding the full-length Gag open reading frame (ORF) contain the -1 frame shift region in the 3'-region of the ORF inducing a frame shift in 5% of the translating ribosomes, which therefore switch into the Protease-ORF (Fig. III-31,



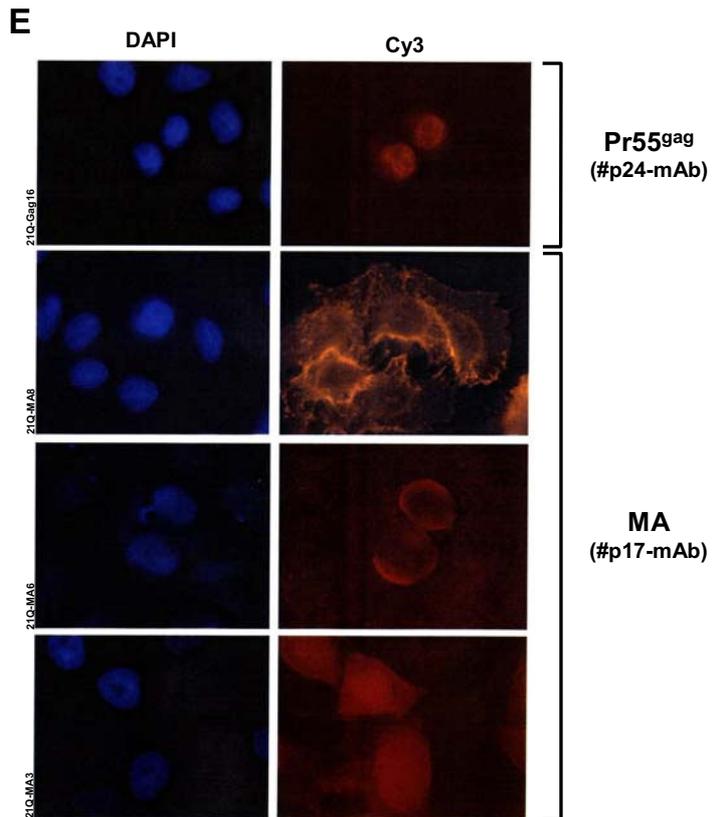
**Fig. III-31: Expression and subcellular localisation of full-length Gag and mutant derivatives.**

**(A)** Illustration of expression vectors encoding full-length or mutant Gag proteins derived from the HIV-1 isolate BH10 (486). Translating the mRNA expressed from the vector encoding the full-length Gag protein (left panel, GagPR) [(653), generously provided by Prof. H.G. Kräusslich (Universitätsklinikum Heidelberg)], 5% of the translating ribosomes shift into the Protease-ORF (right panel). Since the resulting protein was reported to be proteolytically active and to cleave the Gag precursor protein into its subunits, the frame shift region was mutated by a point mutation resulting in sole translation of the gag precursor protein (Pr55<sup>Gag</sup>). The full-length Gag expression plasmid was deleted in the 3' half to allow the expression of the fused matrix- and capsid-subunits (MA/CA) or the matrix subunit only (MA) [CMVp: Cytomegalovirus promoter, CTE: Constitutive Transport Element, BGH pA: Bovine Growth Hormone polyadenylation site].

**(B-D)** Immunoblot analyses of proteins expressed from respective *gag*-encoding plasmids. HeLa-T4<sup>+</sup> cells were transiently transfected with 1  $\mu$ g 3-CCCC (B-D, lane 1, GagPR), CMV gag 3-CCCC (B and D, lane 2), CMV MA/CA 3-CCCC (B-D, lane 3), CMV MA 3-CCCC (B and D, lane 4), pSVT7 serving as negative control (B, lane 6; C, lane 3; D, lane 5) or pSG MA RRE serving as positive control for MA expression, which was induced by cotransfection with 1  $\mu$ g of the Rev expression plasmid SVcrev (B, lane 5). Cells were harvested 48 h after transfection. Protein samples were normalised to equal transfection efficiency, adjusted to equal total protein amounts using mock-transfected cell lysates and separated in 10% SDS-polyacrylamide gels. Immunoblots were probed with antibodies targeting (B) MA (#p17, Advanced Biotechnologies), (C) CA (#p24, DuPont) or (D) human antiserum (As) recognising various HIV-1 proteins (by courtesy of Prof. Dr. O. Adams, Universitätsklinikum Düsseldorf).

**Fig. III-31: Expression and subcellular localisation of full-length Gag and mutant derivatives (continued).**

(E) Immunofluorescence analysis of HeLa-T4<sup>+</sup> cells transiently transfected with Gag expression plasmids.  $3 \times 10^5$  cells were grown on glass cover slips in 6-well plates and transiently transfected with 1  $\mu$ g CMV gag 3-CCCC (Pr55<sup>Gag</sup>) or CMV MA 3-CCCC (MA) as indicated. Cells were fixed 30 h after transfection and permeabilised using 0.02% saponin. After blocking monoclonal antibodies recognising either MA (#p17, Advanced Biotechnologies) or CA (#p24, DuPont) were applied. After incubation with a Cy3-labelled secondary antibody, cover slips were analysed using a Zeiss Axioscope (1000 x magnification).



GagPR). Albeit the protease is not completely encoded by the full-length expression vector used here, the C-terminally truncated Protease has been reported to be active and proteolytically cleave the Gag precursor protein (Pr55<sup>Gag</sup>) into its subunits MA, CA, p2, NC, p1 and p6. To distinguish effects of the Gag precursor protein from those mediated by individual Gag cleavage products, a full-length Gag expression plasmid encoding Pr55<sup>Gag</sup> was generated expressing exclusively the Gag precursor protein due to a mutation in the frame shift region (Fig. III-31A, Pr55<sup>Gag</sup>). In addition, two plasmids were generated exclusively expressing either the MA/CA fusion protein or the MA subunit due to differential deletions of the 3'-coding region (Fig. III-31A, MA/CA and MA). Mutant Gag protein expression in transiently transfected cells was confirmed by immunoblotting (Fig. III-31B). Detection of all generated Gag derivatives by a monoclonal antibody recognising an epitope in the MA subunit failed, but only confirmed the expression of the sole MA subunit and an MA control (Fig. III-31B, lanes 4 and 5). Therefore, protein expression of the full length Gag precursor and the MA/CA subunit was controlled using a monoclonal antibody recognising an epitope in the CA subunit (Fig. III-31C, lanes 1 and 2). Unexpectedly, proteolytic cleavage products of the full length Gag precursor protein were hardly detected. Therefore, the expression of all

generated Gag derivatives was also analysed using a human antiserum recognising epitopes in various HIV-1 proteins (kindly provided by Prof. Dr. O. Adams, Universitätsklinikum Düsseldorf). However, application of the antiserum only confirmed the absence of proteolytic Gag cleavage products in amounts sufficient for immunoblot detection. Nevertheless, overall expression of the Gag precursor protein was enhanced after mutation of the frame shift site (Fig. III-31D, cf. lane 1 with 2). To clarify whether the Gag precursor protein or a subunit of the precursor enters the nucleus and thus is able to interact with pre-mRNA prior to splicing, localisation of Pr55<sup>gag</sup> and MA after transient transfection was examined by immunofluorescence. Microscopy data revealed that the Gag precursor protein was localised mainly nuclear (Fig. III-31E, Pr55<sup>gag</sup>). In contrast, the MA subunit was found in different subcellular compartments of individual cells. In some cells MA was detected mainly cytoplasmic and at the cell membrane (Fig. III-31E, MA, top and middle panel). However, in other cells MA located in both, the cytoplasm as well as the nucleus (Fig. III-31E, MA, bottom panel) confirming the previously described nuclear localisation signal (NLS) in the MA subunit. Therefore, MA as well as the Gag precursor protein localised at least transiently in the nucleus allowing an effect of MA as well as the Gag precursor protein on the splicing pattern at D1.

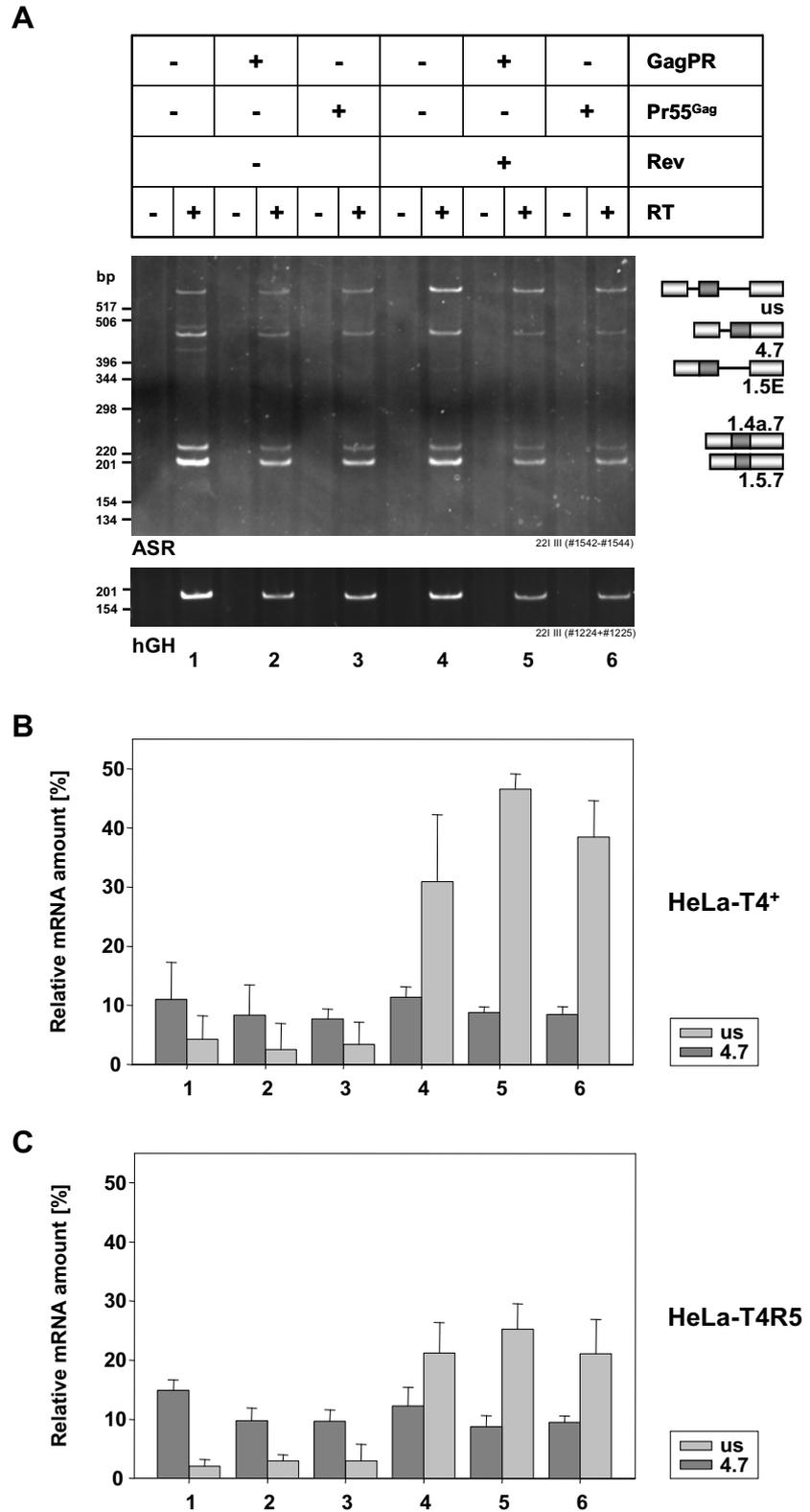
To elucidate the impact of Gag on the amount of RNA unspliced at D1, the splicing pattern of the reporter minigene was analysed by RT-PCR of total RNA derived from HeLa-T4<sup>+</sup> cells transiently cotransfected with the 2-intron minigene and the respective Gag expression vectors (Fig. III-32A). Quantification of mRNA isoforms unspliced at D1 (us and 4.7) revealed that expression of both isoforms is unaffected by coexpression of the proteolytically active Gag precursor protein or the Gag mutant expressing only the Gag precursor protein (Fig. III-32B, cf. 1 with 2 and 3). Additional coexpression of Rev increases the expression of completely unspliced RNA from the 2-intron minigene. Although coexpression of Gag proteins appeared to slightly increase the amount of completely unspliced RNA in the presence of Rev, this difference was not statistically significant ( $p=0.11$ ) (Fig. III-32B, cf. 4 with 5 and 6). Consistently, the expression of 4.7-mRNA, which remained unspliced at D1 but is spliced at the distal intron defined by D4 and A7, was also not influenced by coexpression of the Gag proteins in the presence of Rev. Surprisingly, expression of 4.7-mRNA in the presence of Rev, which was expected to be reduced due to interference of Rev with splicing of the distal intron, did not decrease, but remained unchanged (Fig. III-32B, cf. 1-3 with 4-6).

**Fig. III-32: Expression of Gag subunits does not affect splicing at D1.**

**(A)** RT-PCR analysis of HeLa-T4<sup>+</sup> cells transiently transfected with the 2-intron minigene and Gag expression plasmids. 2 x 10<sup>5</sup> cells were transiently transfected with 1 µg SV leader SD1 SA5 env nef, 1 µg 3-CCCC (GagPR) or CMV gag 3-CCCC (Pr55<sup>Gag</sup>), 1 µg SVcrev (Rev) where indicated or a control plasmid, and 1 µg pXGH5 to monitor transfection efficiency. RNA isolation and RT-PCR were performed as described in Fig. III-24C. Negative controls omitting reverse transcription were performed to rule out DNA contamination (RT-) [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, RT: reverse transcription, us: unspliced].

**(B-C)** Relative quantification of mRNA isoforms unspliced at D1 in HeLa-T4<sup>+</sup> (B) and HeLa-T4R5 cells (C). Transient transfection experiments were performed as described in (A). RT-PCR products were fluorescently labelled using Cy5-marked primers as described in Fig. III-23A. Raw

data were analysed regarding the percentage of unspliced and 4.7-mRNA relative to total mRNA isoforms expressed from the respective 2-intron minigene. Error bars indicate standard deviation from the mean of three independent experiments. Slight differences observed in (B) in the presence of Rev are not statistically significant as evaluated using the paired student's t-test (p=0.11).



---

To test whether a statistically significant difference in the expression of completely unspliced mRNA could be induced by increasing the ratio of processed Gag subunits, the transfection experiment was performed using HeLa-T4R5 cells, additionally expressing the CCR5 coreceptor, which had been reported to allow a more efficient cleavage of the Gag precursor protein (653). However, the apparent effect of the proteolytically active Gag precursor protein (GagPR) was not confirmed in HeLa-T4R5 cells (Fig. III-32B and C, cf. lanes 4 and 5).

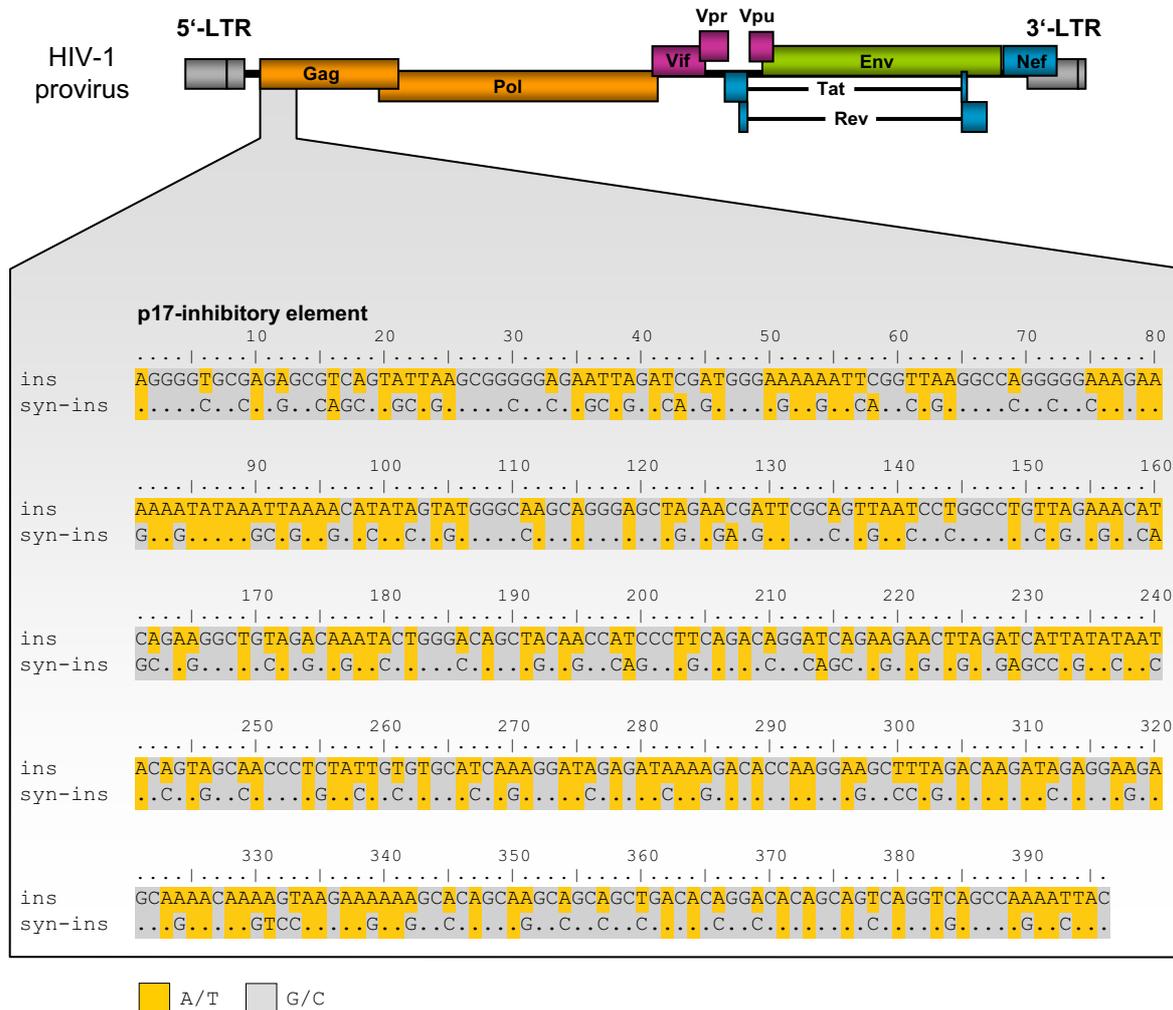
These results demonstrated that the viral Gag protein does not influence splicing at D1 in the minigene and therefore the interaction of Gag with the viral mRNA is not likely to act as a regulatory mechanism necessary for the tremendous increase in the expression of the completely unspliced RNA observed 24 h after viral infection of the cell.

### **C.2.2 The p17-inhibitory element in the viral pre-mRNA regulates temporal HIV-1 gene expression.**

Analysing the splicing pattern of the 2-intron minigene, it was an unexpected finding that a substantial amount of mRNAs remained unspliced in the proximal intron (4.7). Therefore, the question arose whether D1 might be much less efficiently used than it was anticipated from its intrinsic strength. The hypothesis that D1 is only inefficiently used was substantiated by the suppressive effect of the SR protein binding sites overlapping D1 on 5' ss usage. However, neither mutation of predicted SR protein binding sites overlapping D1 nor substitution of flanking sequences could completely activate D1. Therefore, the absence of unspliced mRNA early in viral gene expression might be caused by multiple inhibitory sequences (ins) dispersed throughout the *gag-pol*-coding region within the genomic context, which were reported to interfere with the expression of ins-containing HIV-1 mRNAs (121, 530), but which were lacking in the initially analysed 2-intron minigene.

#### **C.2.2.1 The proximal 202 nt of the Gag-ORF are required for Rev-reactivity in the presence of the p17-inhibitory element.**

To analyse whether the expression of completely unspliced RNA might be modulated by the sequence downstream of the Gag translational start codon, the first 396 nt of the Gag-ORF encoding the MA subunit (Fig. III-33) were reinserted at their authentic position within the 2-intron minigene (ins) (Fig. III-34A). To prevent translation of the Gag-ORF the translational start codon (ATG) was inactivated by introducing a point mutation from T to G at nucleotide +2 of the Gag-ORF. The inserted MA-encoding sequence contained the inhibitory elements M1-M4, which had been reported to confer instability to the *gag*-mRNA (526, 530). As a control, which should not decrease the stability of the intron-containing RNA, the p17-ins element was substituted with the corresponding region of a codon-optimised form, syn-gag [generously provided by Prof. Dr. R. Wagner (Universität Regensburg) (216)] (Fig. III-33 and Fig. III-34A, syn-ins), in which the codons of the Gag-ORF had been substituted with those most often used in highly expressed human genes ("human codon usage"). Adaptation to human codon usage allows Rev-independent Gag expression from otherwise unstable *gag*-mRNA expressed from intronless plasmids (216).



**Fig. III-33: Comparison of the MA-encoding sequence of the Gag-ORF containing the p17-inhibitory element and the respective codon-optimised syn-gag sequence.**

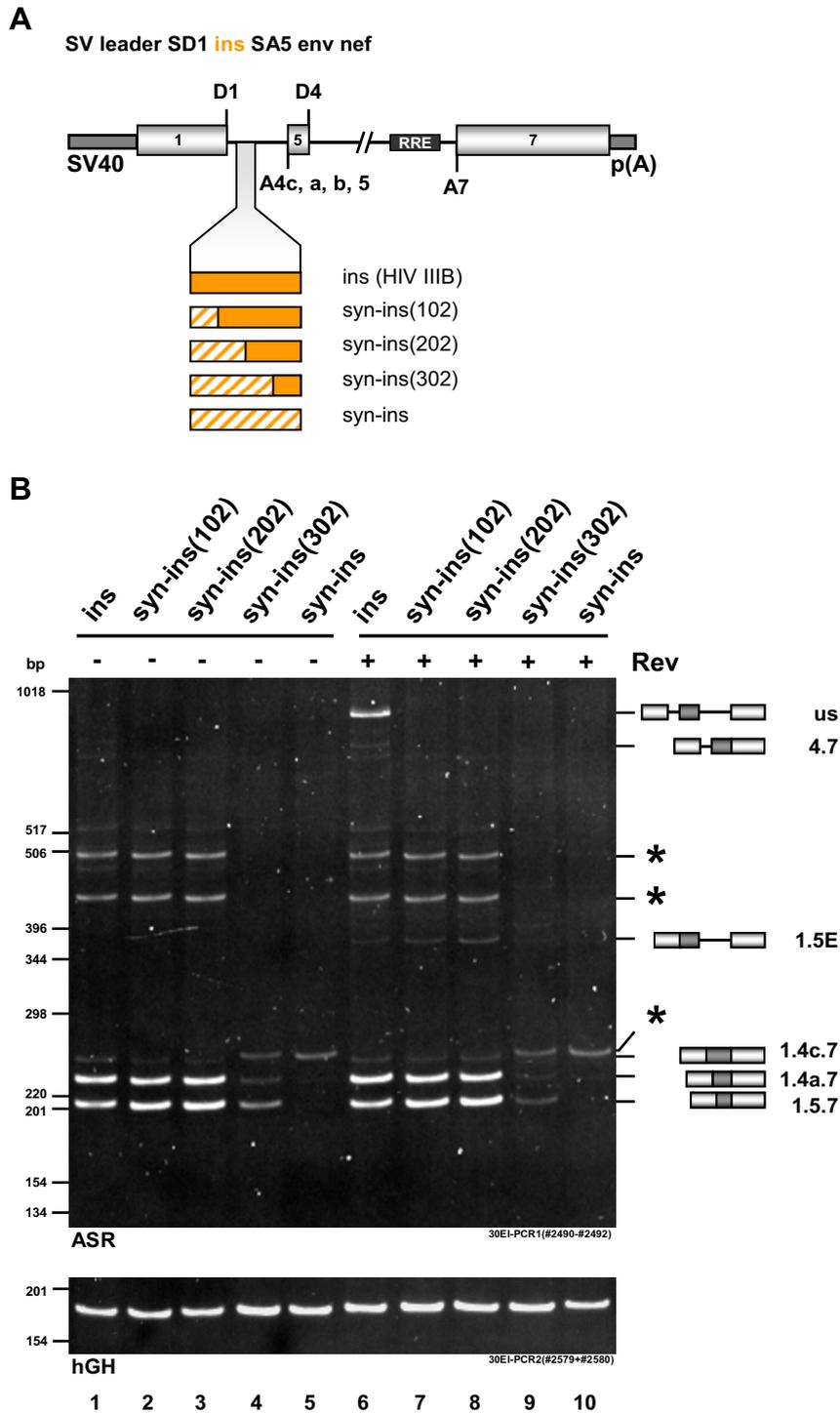
Illustration of parental and mutant p17-ins sequences inserted into the 2-intron minigene. The most upstream 396 nt of the Gag-ORF encoding the matrix subunit (MA) harbour the p17-inhibitory sequence (p17-ins), which had been reported to induce RNA degradation (530). To abrogate the inhibitory effect of the p17-ins element on Gag expression, a codon-optimised form of the Gag-ORF was generated by adapting the codons to the codon usage of frequently expressed human genes by introducing 115 translationally silent nucleotide substitutions (216). The resulting syn-gag sequence contains a decreased AT-content compared to the parental gag-encoding sequence. To avoid translation of the p17-ins sequence after insertion into the 2-intron minigene the translational start codon in the ins- as well as in the syn-ins sequence was inactivated by a point mutation from T to G at +2 of the Gag-ORF. Adenosines and thymidines are highlighted in orange, whereas guanosines and cytosines are highlighted in grey.

The splicing pattern of the 2-intron minigenes carrying the insertion of the ins- or syn-ins element, respectively, downstream of the inactivated Gag translational start codon was analysed after transient transfection of HeLa-T4<sup>+</sup> cells by RT-PCR of total RNA. Insertion of the p17-ins element clearly reduced the amount of RNA unspliced at D1 in the absence of Rev (cf. Fig. III-34B, lane 1 with Fig. III-30B, lane 1). However, the insertion did not completely abrogate the expression of RNA unspliced at D1 in the context of the subgenomic 2-intron minigene. Furthermore, 3' ss A4a and A5 were equally used for inclusion of the internal exons (1.4a.7 and 1.5.7). Surprisingly, substitution of the inhibitory element with the codon-optimised version abrogated the detectability of any 2-intron minigene expressed mRNA. Sequencing of the only visible band revealed a non-specific amplification of human *cofilin*-mRNA (Fig. III-34B, lane 5). These contrary effects of the p17-ins element and its codon-optimised version might hint to an additional function of the inhibitory element other than sole degradation. However, it should be noted that lack of mRNA detectability under these experimental conditions does not necessarily imply degradation of the RNA. Alternatively, an upstream cryptic splice site might be activated due to the inserted sequence leading to loss of the primer binding sites used in this RT-PCR assay.

**Fig. III-34: The proximal 102 nucleotides of the p17-inhibitory element are required for Rev-mediated expression of completely unspliced RNA.**

(A) Schematic diagram illustrating the insertion of the ins-sequence or mutant sequences into the 2-intron minigene. The proximal 396 nt of the p17-inhibitory element (ins, orange) or the respective region of the codon-optimised syn-gag sequence (syn-ins, orange fasciated) were inserted at their authentic position immediately downstream of the inactivated translational start codon of the Gag-ORF. Hybrids of ins- and syn-ins-containing minigenes were generated by successively replacing the downstream region of the ins- with the syn-in-sequence [syn-ins(102), syn-ins(202), syn-ins(302)] [p(A): SV40 polyadenylation site, RRE: Rev-responsive element, SV40: SV40 early promoter].

(B) RT-PCR analysis of HeLa-T4<sup>+</sup> cells transfected with the 2-intron minigene containing the proximal 396 nt of the ins-sequence or mutations thereof.  $2 \times 10^5$  cells were transiently transfected with 1  $\mu$ g SV leader SD1 ins(IIIB) SA5 env nef [ins, lane 1 and 6], SV leader SD1 syn-ins102(IIIB) SA5 env nef [syn-ins(102), lane 2 and 7], SV leader SD1 syn-ins202(IIIB) SA5 env nef [syn-ins(202), lane 3 and 8], SV leader SD1 syn-ins302(IIIB) SA5 env nef [syn-ins(302), lane 4 and 9] or SV leader SD1 syn-ins(IIIB) SA5 env nef [syn-ins, lane 5 and 10]. Cells were either cotransfected with 1  $\mu$ g SVcrev, a plasmid expressing the viral regulatory protein Rev, or a control plasmid. Additionally all cells were transfected with 1  $\mu$ g pXGH5 to control equal transfection efficiency. RNA isolation and RT-PCR were performed as described



**Fig. III-34: continued.**

in Fig. III-24C. The exon/intron structure of alternatively spliced mRNA isoforms is depicted at the right. mRNA isoform 1.4c.7 and heteroduplex or unspecific PCR products (marked by asterisks) were confirmed by sequencing of the respective bands purified from the gel [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, us: unspliced].

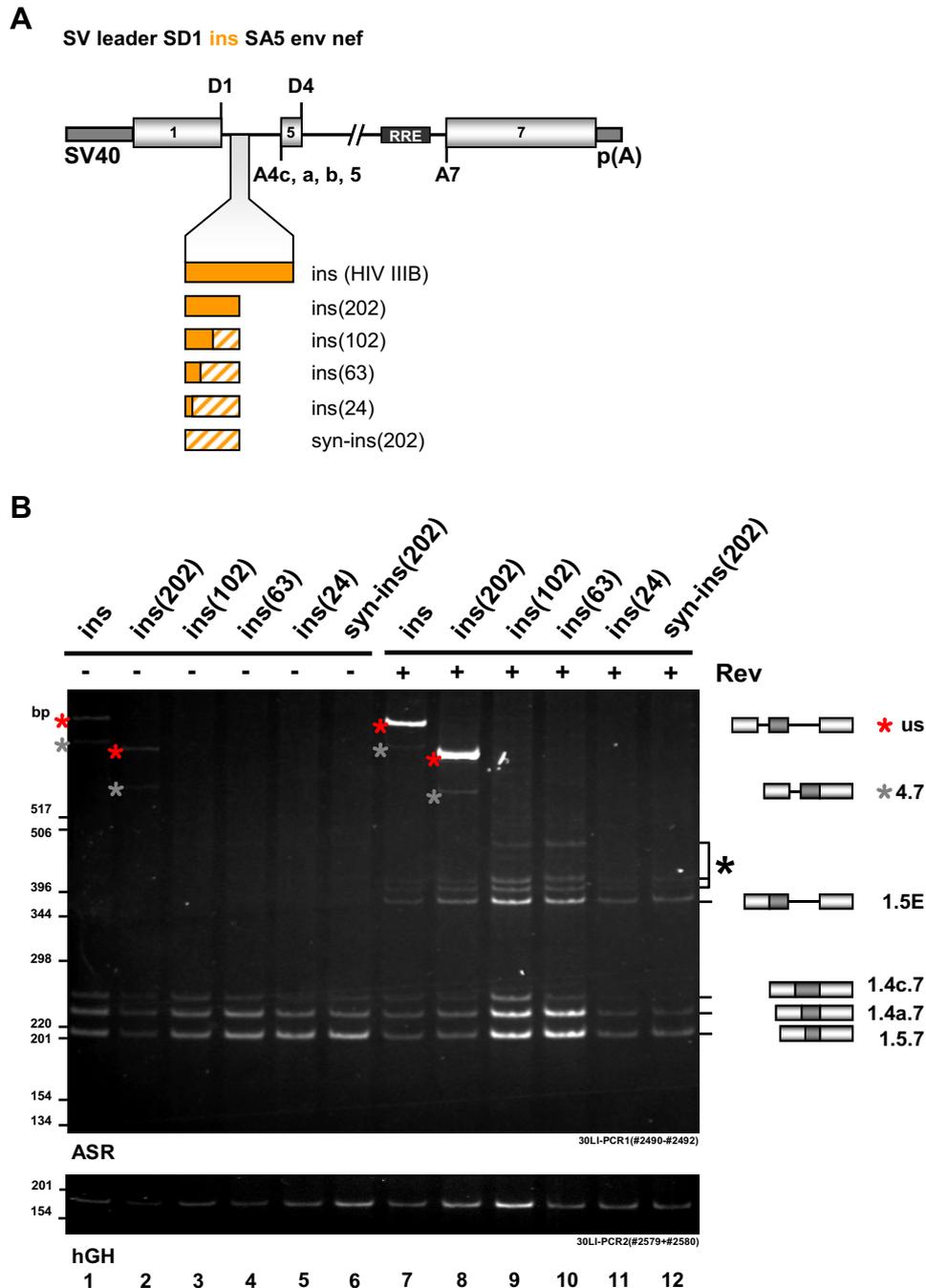
To identify the region within the codon-optimised syn-ins sequence responsible for RT-PCR signal loss, the syn-ins sequence was subdivided into four regions of nearly identical length and successively replaced by the p17-ins element (Fig. III-34A). RT-PCR analysis of transiently transfected HeLa-T4<sup>+</sup> cells indicated that nucleotides 202-301 are essential for loss of the RT-PCR signal in this 2-intron minigene context (Fig. III-34B, lane 4). Interestingly, strong Rev-mediated expression of completely unspliced RNA was only observed in the presence of the p17-ins element, but was abrogated after partial substitution with the codon-optimised version (Fig. III-34B, cf. lane 6 with lanes 7-8). However, Rev-mediated expression of the mRNA isoform containing the distal intron defined by D4 and A7 (1.5E) was observed for all minigenes carrying substitutions in the 5'-half of the p17-ins element. From these results it was concluded that Rev-mediated expression of RNA unspliced at D1 depends on elements distinct from those interfering with the removal of the distal intron defined by splice sites D4 and A7.

In the previous experiment, it was striking that substitution of the 5'-located 102 nt of the inserted p17-ins element abrogated Rev-mediated expression of unspliced RNA (Fig. III-34B, lane 7). To examine whether the proximal 102 nt of the p17-ins element are sufficient to confer Rev-reactivity regarding D1 to a syn-ins-containing 2-intron minigene, the proximal 102 nt of the syn-ins sequence were replaced with the proximal 102 nt of the p17-ins sequence [Fig. III-35A, ins(102)]. In order to narrow down the

**Fig. III-35: The downstream flanking region (nt 103-202) of the p17-ins element is additionally needed to mediate Rev-dependent expression of completely unspliced RNA.**

**(A)** Illustration of ins-/syn-ins-hybrid sequences generated in the context of the 2-intron minigene. To avoid loss of RT-PCR product detectability the inserted p17-ins sequence was shortened to 202 nt [ins(202), orange]. As control a syn-ins-containing minigene was generated containing only the proximal 202 nt of the codon-optimised syn-ins sequence [syn-ins(202), orange fascinated]. Hybrid minigenes were generated containing proximal proportions of the ins-sequence and distal proportions of the syn-ins sequence [ins(102), ins(63) and ins(24)] [p(A): SV40 polyadenylation site, RRE: Rev-responsive element, SV40: SV40 early promoter].

**(B)** RT-PCR analysis of HeLa-T4<sup>+</sup> cells transfected with the 2-intron minigene containing the shortened p17-ins sequence or mutants thereof. 2 x 10<sup>5</sup> HeLa-T4<sup>+</sup> cells were transiently transfected with 1 µg SV leader SD1 ins(IIIB) SA5 env nef [ins, lane 1 and 7], SV leader SD1 ins(202) SA5 env nef [ins(202), lane 2 and 8], SV leader SD1 ins(102)-syn-ins(202) SA5 env nef [ins(102), lane 3 and 9], SV leader SD1 ins(63)-syn-ins(202) SA5 env nef [ins(63), lane 4 and 10], SV leader SD1 ins(24)-syn-ins(202) SA5 env



**Fig. III-35: continued.**

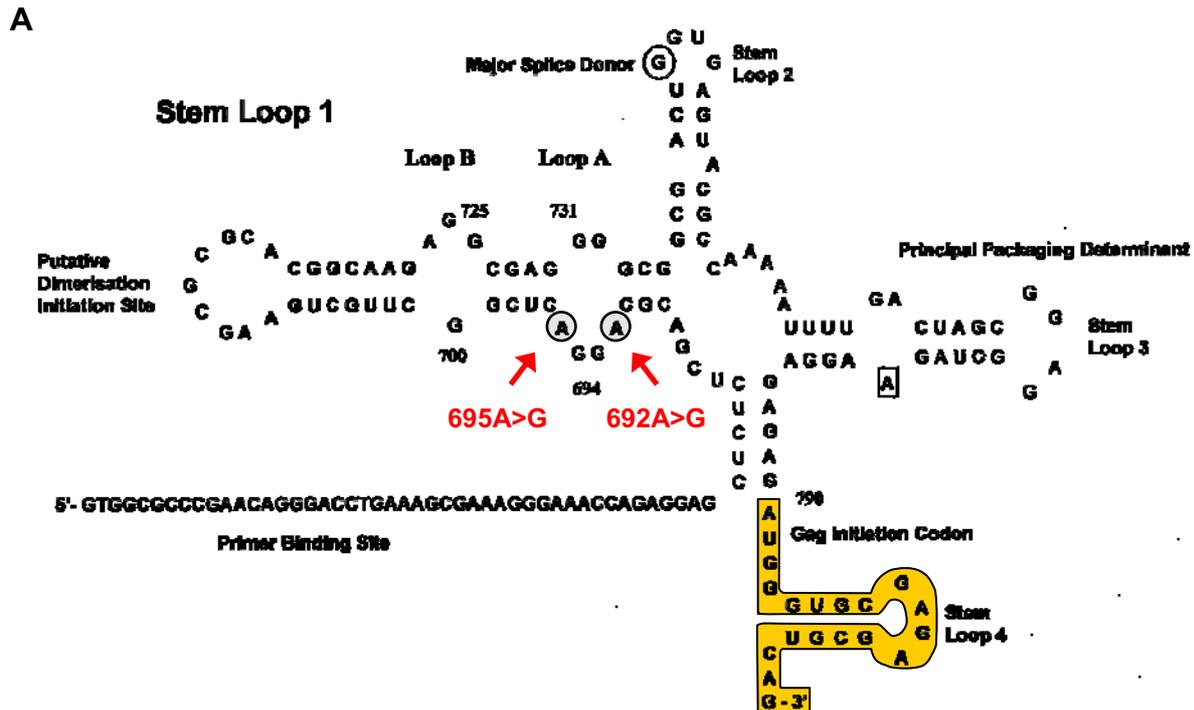
nef [ins(24), lane 5 and 11] or SV leader SD1 syn-ins(202) SA5 env nef [ins(202), lane 6 and 12] and cotransfected with either 1  $\mu$ g SVcrev, where indicated (Rev), or a control plasmid. Additionally, all cells were transfected with 1  $\mu$ g pXGH5 to control equal transfection efficiency. RNA isolation and RT-PCR were performed as described in Fig. III-24C. The exon/intron structure of differentially spliced mRNA isoforms is shown at the right. Due to the shortened ins-insertion, RT-PCR products derived from mRNAs retaining the proximal intron (4.7 and us) differ in size for the respective minigenes. RT-PCR products originating from mRNAs with identical exon/intron structure are indicated by coloured asterisks. The black asterisk marks heteroduplex PCR products [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, us: unspliced].

sequence required for Rev-reactivity additional plasmids were generated containing different proportions of the p17-ins element deduced from RNA structural elements predicted to form in this region, i.e. the proximal 24 nt (SL4) or 63 nt (SL4 and ORF1) [C.K. Damgaard, PhD thesis (142)]. To ensure RT-PCR detection, the inserted inhibitory elements were restricted to the proximal 202 nt. RT-PCR of transiently transfected HeLa-T4<sup>+</sup> cells verified that the Rev-dependent expression of unspliced RNA had been preserved in the 2-intron minigene containing only 202 nt of the p17-ins element, whereas it was totally abrogated in the syn-ins-containing control minigene (Fig. III-35B, lanes 8 and 12). Unexpectedly, insertion of the first 102 nt were not sufficient to transfer Rev-reactivity of D1 to a minigene containing nucleotides 103-202 of the codon-optimised syn-gag sequence (Fig. III-35B, lane 9). Consistently, both shorter proportions of the p17-ins element failed to express completely unspliced RNA in the presence of Rev (Fig. III-35B, lanes 10 and 11). Although completely unspliced RNA was not expressed from syn-ins-containing minigenes, all ins-/syn-ins-hybrid minigenes examined here expressed Rev-dependent 1.5E-mRNA demonstrating that all minigene derived pre-mRNAs remain Rev-reactive regarding the distal intron (Fig. III-35B, lanes 7-12). In addition, the syn-ins sequence affected the ratio of the 3' ss used for inclusion of the internal exon, so that the ratio of 3' ss selection is similar to that of the 2-intron minigene lacking the p17-ins element (Fig. III-35B, lane 6).

These results revealed that in the presence of syn-gag sequences besides the proximal 102 nt additional nucleotides of the region 103-202 of the p17-ins element are needed to confer Rev-reactivity of D1 for the generation of completely unspliced RNA to the control minigene.

### **C.2.2.2 The Rev binding site in stem loop 1 does not contribute to Rev-mediated increase of unspliced RNA.**

Initially, Rev was described to interact with the HIV-1 pre-mRNA at an RNA secondary structure termed Rev-responsive element (RRE) located in the intron flanked by D4 and A7. In a recent study, a second Rev binding site was identified in stem loop 1 (SL1) located upstream of D1 in a RNA secondary structure (198) forming within the 5'-untranslated region (5'-UTR) (Fig. III-36A). The p17-ins sequence might be involved in the formation of the RNA secondary structure of the 5'-UTR generating the binding site for Rev. To analyse whether the Rev binding site contributes to the Rev-reactivity of D1



**Fig. III-36: Mutation of the potential Rev binding site in stem loop 1 of the 5'-UTR does not change the splicing pattern of the 2-intron minigene.**

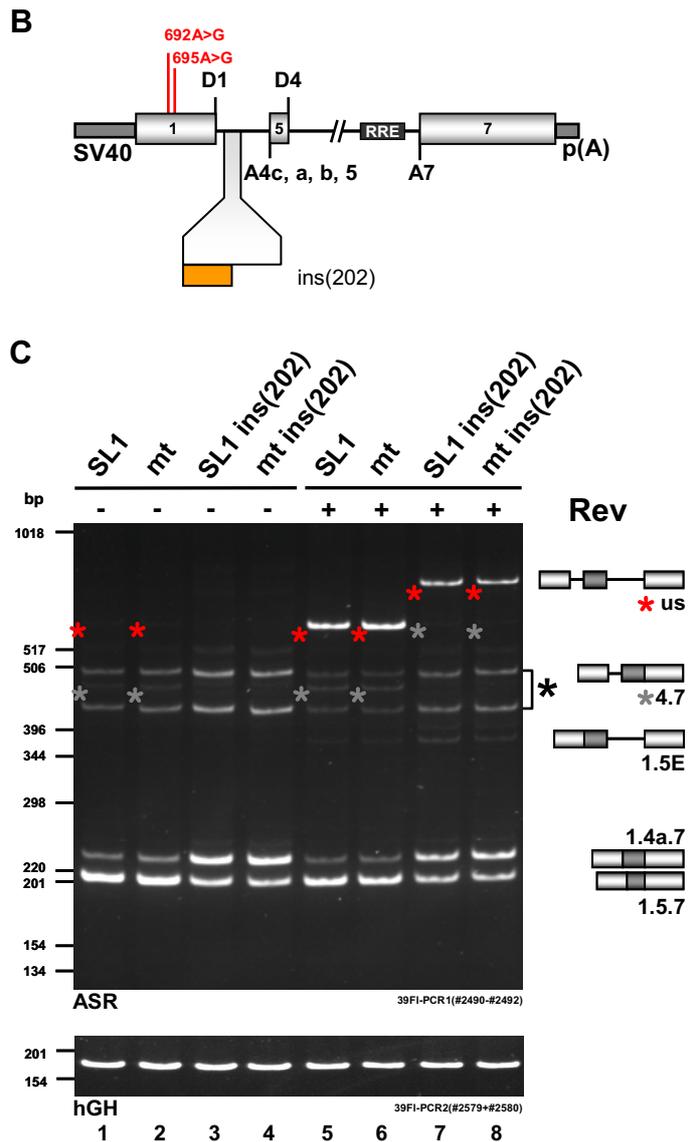
**(A)** The 5'-UTR harbouring D1 folds into a large secondary structure [adapted from (198)], in which four stem loops have been described (SL1-SL4). Recently, loop A of SL1 was reported to act as additional binding site for the viral regulatory protein Rev (198). Two point mutations from A to G at position 692 and 695 in the 5'-UTR had been reported to reduce binding of Rev to SL1 (198). Although quite distant, the p17-ins sequence (most proximal nucleotides are highlighted in orange) might assist the generation of SL1 through stabilisation of a central loop at the basis of SL1, SL2 and SL3.

in ins-containing minigenes, the Rev binding site upstream of D1 was mutated by introducing two point mutations from A to G in SL1 (Fig. III-36A) reported to reduce Rev binding *in vitro* (198) and to result in a viral replication defect (226). Analysis of the splicing pattern of the 2-intron minigene lacking the p17-ins element revealed that mutation of the Rev binding site in the leader upstream of D1 did not at all affect the splicing pattern of the 2-intron minigene neither in the absence nor in the presence of Rev (Fig. III-36B, cf. lane 1 with 2 and lane 3 with 4).

**Fig. III-36: Mutation of the potential Rev binding site in stem loop 1 of the 5'-UTR does not change the splicing pattern of the 2-intron minigene (continued).**

**(B)** Illustration of point mutations corresponding to nucleotides 692 and 695 in SL1 introduced into the 2-intron minigene in the absence or the presence of the proximal 202 nt of the p17-ins element. The constructs contain the shortened version of the p17-ins element to ensure the expression of detectable mRNA isoforms.

**(C)** RT-PCR analysis of HeLa-T4<sup>+</sup> cells transfected with 2-intron minigene derivatives carrying mutations in SL1. Cells were transiently transfected with SV leader SD1 SA5 env nef [SL1, lane 1 and 5], SV leader GGGG SD1SA5 env nef [mt, lane 2 and 6], SV leader SD1 ins(202) SA5 env nef [SL1 ins(202), lane 3 and 7] or SV leader GGGG SD1 ins(202) SA5 env nef [mt ins(202), lane 4 and 8], cotransfected with SVcrev, where indicated (Rev), and pXGH5 to monitor equal transfection efficiency. RT-PCR was performed as described in Fig. III-24C. The exon/intron structure of alternatively spliced mRNA isoforms is depicted at the right. Due to the insertion of the ins sequence, RT-PCR products derived from unspliced and 4.7-mRNA differ in size for the respective minigenes. RT-PCR products originating from mRNAs with identical exon/intron structure are indicated by coloured asterisks. The black asterisk marks heteroduplex PCR products [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, us: unspliced].



Also, introducing the SL1 mutation in the 2-intron minigene containing the proximal 202 nt of the ins-sequence did not influence the splicing pattern of the 2-intron minigene (Fig. III-36B, lanes 7 and 8). These unexpected results suggested that the Rev binding site in SL1 is dispensable for Rev-reactivity in minigenes containing the proximal 202 nt of the p17-ins element.

### C.2.2.3 The first 202 nt of the p17-ins element contain an intronic splicing enhancer.

As a control that the insertion of the p17-ins element caused degradation of RNA unspliced at D1, the ins element was combined with a point mutation of 5' ss D1 at position +5 from G to T (Fig. III-37A and B, 5U). This mutation lowers the complementarity of the 5' ss to the U1 snRNA and was expected to abrogate splicing at D1 caused by loss of U1 snRNA binding. RT-PCR analysis of transiently transfected HeLa-T4<sup>+</sup> cells revealed that inactivation of D1 resulted predominantly in mRNA, which was only spliced between D4 and A7 (4.7), whereas the amount of unspliced RNA (us) increased only marginally (Fig. III-37C, lane 4).

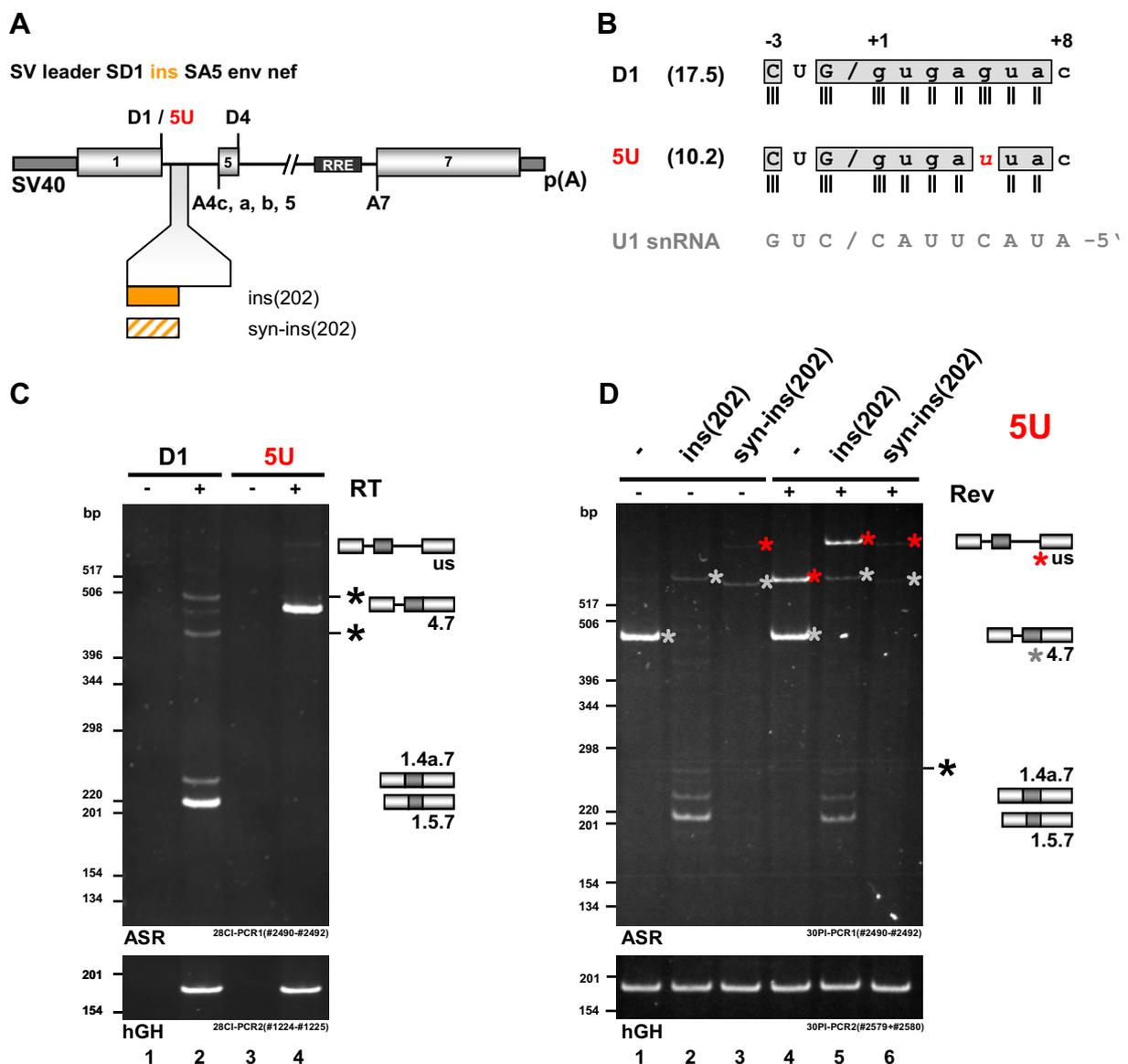


Fig. III-37: Splicing of a suboptimal D1 mutant is promoted by an intronic splicing enhancer in the p17-ins element.

**Fig. III-37: continued.**

**(A)** Scheme of 2-intron minigenes carrying the proximal 202 nt of the p17-ins element or the codon-optimised syn-ins control sequence in combination with either the parental 5' ss D1 or the mutant 5' ss 5U [p(A): SV40 polyadenylation site, RRE: Rev-responsive element, SV40: SV40 early promoter].

**(B)** Sequence comparison between the parental 5' ss D1 and the mutant 5' ss 5U (+5 G to T; shown in red). Due to the high complementarity to the free 5' end of the U1 snRNA (grey letters) D1 is classified as intrinsically strong splice site as monitored by high score values (denoted in brackets) assigned by the splice finder algorithm [splicefinder (ver. 4.1) ([www.uni-duesseldorf.de/rna](http://www.uni-duesseldorf.de/rna))]. Complementary base pairs are highlighted in grey. Hydrogen bonds, which can be formed between 5' ss nucleotides and the U1 snRNA, are depicted as vertical bars.

**(C)** RT-PCR analysis of HeLa-T4<sup>+</sup> cells transfected with the parental 2-intron minigene or the minigene carrying the mutant 5' ss 5U in the absence of the p17-ins element. Cells were transiently transfected with SV leader SD1 SA5 env nef (D1, lane 1 and 2) or SV leader 5U SA5 env nef (5U, lane 3 and 4) and pXGH5 to control equal transfection efficiency. RT-PCR was performed as described in Fig. III-24C. PCR reactions were controlled for plasmid DNA contamination by omitting reverse transcription (RT -, lane 1 and 3). The exon/intron structure of alternatively spliced mRNA isoforms is depicted at the right. The asterisks mark heteroduplex PCR products [ASR: alternatively spliced reporter RNA, hGH: human Growth Hormone, us: unspliced].

**(D)** RT-PCR analysis of HeLa-T4<sup>+</sup> cells transfected with the 2-intron minigene carrying the 5U mutation in the presence of the proximal 202 nt of the p17-ins element or the syn-ins control sequence. HeLa-T4<sup>+</sup> cells were transiently transfected with SV leader 5U SA5 env nef [lane 1 and 4], SV leader 5U ins(202) SA5 env nef [ins(202), lane 2 and 5] or SV leader 5U syn-ins(202) SA5 env nef [syn-ins(202), lane 3 and 6], SVcrev, where indicated, or a control plasmid and pXGH5 to control equal transfection efficiency. RT-PCR was performed as described in Fig. III-24C. The exon/intron structure of alternatively spliced mRNA isoforms is shown at the right. Due to the insertion of the ins-sequence or the control sequence, RT-PCR products retaining the proximal intron are enlarged compared to the parental 2-intron minigene. Coloured asterisks mark RT-PCR products resulting from mRNA isoforms with identical exon/intron structure. RT-PCR products corresponding to 1.4a.7- and 1.5.7-mRNA were confirmed by sequencing.

Insertion of the proximal 202 nt of the p17-ins element downstream of the 5U mutation of D1 surprisingly showed that splicing from this mutant 5' ss was reactivated (Fig. III-37D, lane 2). In contrast, insertion of the 202 nt syn-ins sequence did not activate splicing at the 5U mutation (Fig. III-37D, lane 3). Regarding the stability of ins- and syn-ins-containing mRNAs, it was observed that the mRNA isoform, which retained the proximal intron and thereby the respective ins- or syn-ins-sequence (4.7), was expressed to a similar extent from the ins- as well as from the syn-ins-containing

minigene (Fig. III-37D, cf. lanes 2 and 3) suggesting that the proximal 202 nt of both sequences did not differentially alter mRNA turnover. However, in combination with the 5U mutation at D1 only the syn-ins- and not the ins-containing minigene expressed completely unspliced RNA (us) in the absence of Rev. Nevertheless, Rev-reactivity at the 5U mutation of D1 was still only observed for the ins-containing pre-mRNA, whereas mRNA isoforms unspliced at D1 were at best slightly increased in the syn-ins-containing minigene by coexpression of Rev (Fig. III-37D, cf. lane 2 with 5 and lane 3 with 6).

These results suggested that the proximal 202 nt of the p17-ins sequence harbour an intronic splicing enhancer ensuring activation of D1 even after mutation of the 5' ss causing limited complementarity to U1 snRNA.

### **C.2.3 Summary: Splicing at D1 is oppositely regulated by SR protein binding sites overlapping the 5'ss and the downstream located intronic p17-ins element.**

The second part of this thesis aimed to identify sequence requirements necessary for the expression of genomic HIV-1 RNA. To this end, it was investigated whether splicing regulatory elements control 5' ss D1 efficiency and whether the p17-ins element located downstream of D1 influences the usage of this 5' ss.

To evaluate whether D1 efficiency is downregulated in the late phase of viral gene expression due to the presence of splicing regulatory elements, the 5' ss and flanking sequences were analysed with respect to binding sites for two splicing regulatory protein families, the hnRNPs and the SR proteins. Mutational analyses of the D1 sequence itself revealed that binding sites for the SR proteins SC35 and SRp55 overlapping D1 suppress 5' ss efficiency. Binding of SC35 to the exonic sequence of D1 was confirmed *in vitro*. These results might indicate a competition of SR proteins with U1 snRNA for binding to D1. The competition model was substantiated by the two further findings that suppression of D1 usage is not supported by additional SR protein binding sites in the region upstream of D1 and that also tethering the SC35 RS domain into a more upstream exonic position, as expected, activated splicing rather than inhibiting splicing at D1. In contrast to the SR protein binding sites, an hnRNP H binding site upstream of D1 does not regulate splicing at D1, which was shown by analysis of the mutant minigene as well as shRNA-mediated knockdown of cellular hnRNP H. However, hnRNP H is needed for efficient inclusion of the internal exon into the

minigene mRNA suggesting a role of hnRNP H in the activation of splice sites flanking the internal exons 4c, 4a, 4b, and 5.

Searching for additional splicing regulatory factors binding to D1, the proteins Nucleolin, DHX36, EF1- $\alpha$  and RAN were identified, which are involved in RNA metabolism, but could not be linked to splicing regulation. In a more general approach, substitution of sequences directly neighbouring D1 in the exon or the intron influenced the splicing pattern of the 2-intron minigene regarding internal 3' splice selection and distal intron removal. However, overall activation of D1 was not affected. Therefore, the D1 sequence and the flanking regions appear to be only marginally involved in regulation of D1 efficiency, however, the changes in the generation of mRNA isoforms using distant splice sites pointed to the involvement of the SR protein binding sites and the flanking sequences in splicing kinetics.

A major contribution to regulation of splicing at D1 by the viral protein Gag could not be confirmed by the experiments presented in this work. Although the expression of Gag subunits appeared to slightly increase the expression of mRNA isoforms unspliced at D1, this increase turned out to be not statistically significant and much less pronounced than expected in light of the previous findings.

Introducing the p17-inhibitory element (p17-ins) into the 2-intron minigene revealed that Rev-dependent expression of completely unspliced mRNA from ins-containing constructs requires the presence of the proximal 202 nt of the p17-ins element. The Rev binding site in stem loop 1 in the 5'-UTR is dispensable for Rev-reactivity of D1 in these 2-intron minigenes. Examining the effects of the p17-ins sequence insertion on steady-state expression levels revealed that mRNA isoforms containing the p17-ins or the control sequence in the proximal intron are expressed at the same amount. However, this experiment demonstrated for the first time that the proximal 202 nt of the p17-ins element act as an intronic splicing enhancer (ISE) *in vivo*.

The results presented in the second part of this thesis revealed that the sequence of D1 itself as well as neighbouring sequences affect the usage of this 5' splice site, internal 3' splice selection, and distal intron removal. However, the regulatory impact of these sequences is likely not sufficient to mediate the strong increase in the expression of completely unspliced RNA in the late phase of viral gene expression. A more robust effect on the expression of completely unspliced RNA was observed by the insertion of the p17-ins element into the 2-intron minigene revealing that the proximal 202 nt of the ins

---

sequence are essential for Rev-reactivity at D1, but not needed for Rev-reactivity of the distal intron. Therefore, both classes of Rev-dependent mRNA isoforms, i.e. completely unspliced mRNA and mRNAs spliced only in the distal intron, appear to differ in the sequence requirements necessary for Rev-dependency: whereas expression of mRNAs unspliced only in the distal intron depends solely on the presence of the RRE in the distal intron, completely unspliced mRNA additionally requires the p17-ins element in the proximal intron.

## D. Discussion

In this thesis, two mechanisms regulating alternative splicing of the HIV-1 pre-mRNA in the early and late phase of viral gene expression were investigated.

At the beginning of this work, it was not clear whether the GAR ESE located in the central exon 5 supports exon recognition by concomitant activation of both, the upstream located 3' ss as well as the downstream flanking 5' ss D4. Results obtained in this thesis demonstrated that the GAR ESE is essential for bidirectional simultaneous activation of the 3' ss cluster preceding exon 5 and of the downstream D4. Further investigations in this work evidenced that the GAR ESE is embedded in an intricate network of interactions emerging from the E42 sequence and the splice sites flanking exon 5 that regulates the inclusion of exons 4c, 4a, 4b and 5 into the mRNA.

The second part of this thesis questioned whether splicing regulatory elements located in the proximity of the major 5' ss D1 modulate the efficiency of this splice site. The results revealed that the p17-inhibitory sequence (p17-ins) located downstream of D1 in the Gag-ORF harbours an intronic splicing enhancer (ISE) increasing D1 usage *in vivo*. In addition, the first 202 nt of the p17-ins element were found to be crucial to confer Rev-reactivity on D1 resulting in the expression of unspliced mRNA from a subgenomic 2-intron minigene. Contrary to the enormous impact of the p17-ins element, a minor interference with splicing at D1 was observed for SR protein binding sites overlapping this 5' ss.

### D.1 The GAR ESE constitutes a main regulatory molecular hub for central splice site activation of the HIV-1 pre-mRNA.

The purine-rich multi-site GAR ESE in HIV-1 exon 5 activates both an upstream located 3' ss as well as a downstream 5' ss (91). Using a 2-intron minigene, whose splicing pattern is comparable to that of HIV-1 NL4-3-infected PM1 cells (Fig. III-2), it was shown in this thesis that the GAR ESE performs bidirectional 3' ss and 5' ss activation. The GAR ESE-mediated bidirectional splice site activation was demonstrated to additionally depend on the interplay with downstream regulatory elements of exon 5. It was discovered that the GAR ESE fulfils a dual splicing regulatory function (i) by synergistically enhancing inclusion of the internal exons 4a and 5 through all identified SR protein binding sites and (ii) by specifically activating A5 of the 3' ss cluster solely

through the two proximal SF2/ASF binding sites. Analysing the interaction of the GAR ESE with U1 snRNP at the downstream 5' ss D4 and the E42 sequence located in between revealed that the enhancer was not only essential for general but also for selective activation of the alternative 3' ss cluster, which in turn is necessary for generation of *rev-*, *nef-* and *vpu-/env-*mRNAs in the early and late phase of HIV-1 gene expression.

#### **D.1.1 The GAR ESE is essential for exon recognition.**

Experiments presented in this thesis demonstrate that the GAR ESE is crucial for activation of the splice sites flanking the internal exon 5, which led to the inclusion of one of the alternative exons 4c, 4a, 4b or 5 into the mRNA (Fig. III-5C). Since mRNA isoforms including one of these exons encode for the viral regulatory proteins Rev or Nef, which both fulfil important functions in viral replication [reviewed in (146, 229)], the GAR enhancer thus is a prerequisite for the efficient onset of viral replication. Furthermore, besides mediating exon recognition in the early phase of viral gene expression, the enhancer is also pivotal for activation of the 3' ss cluster generating intron-containing *vpu-/env-*mRNAs (Fig. III-6). Therefore, the GAR ESE influences splicing of nearly all mRNA isoforms except unspliced RNA, and thereby constitutes an essential *cis*-regulatory element during early and intermediate HIV-1 gene expression.

It has been reported that exonic sequences even in the absence of their splice sites can suppress joining of flanking exons when inserted into an intron (275). From these data it was concluded that exonic enhancer sequences may ensure exon recognition by acting as barriers preventing the exon removal (275). This, however, contrasts the results observed in this thesis for HIV-1 exon 5, because the GAR ESE was unable to prevent exon skipping caused by mutation of the downstream flanking 5' ss D4 (Fig. III-10B, lane 3). Since inactivating the GAR ESE also induced exon skipping in the 2-intron minigene (Fig. III-5C), the GAR ESE is necessary but not sufficient to obviate internal exon skipping. The simultaneous requirement for the GAR ESE and a functional 5' ss (Fig. III-12B) suggests that splicing-competent complexes assembling at the borders of exon 5 outcompete splicing from D1 to A7 and thus interfere with exon skipping rather by their functionality than by their sole presence. Therefore, *trans*-acting factors binding to the GAR ESE do not serve as a barrier to prevent splicing of the flanking exons as it had been proposed for other exonic enhancers (275). These dissimilar effects of HIV-1

exon 5 and exonic sequences derived from the *β-globin*-mRNA analysed by Maniatis and colleagues (275) indicate that inserted enhancer sequences might differ in their ability to suppress exon joining. These differences might relate to the exon/intron architecture of the model pre-mRNAs or the regulatory proteins recognising the enhancer elements and the mechanisms they elicit.

In addition to exon skipping, mutations of the GAR ESE or D4 consistently increased the expression of completely unspliced RNA (Fig. III-5C and Fig. III-13B). Unexpectedly, mRNA isoforms, which remained unspliced at D1, i.e. completely unspliced mRNA and mRNA retaining only the proximal intron (4.7-mRNA), were already detected after transfection of the parental 2-intron minigene in the absence of Rev, albeit at a lower level. The detectability of mRNA isoforms unspliced at D1 suggests that D1 was only suboptimally activated. This might be caused by a splicing silencer element located upstream of A4c (ESS2) (11), which is conceivable to reduce the efficiency of proximal intron removal in the 2-intron minigene. In pre-mRNAs unspliced in the proximal intron both, D1 and D4, are able to compete for usage of the most downstream 3' ss, A7. In the parental 2-intron minigene only D4 – but not D1 – was used generating 4.7-mRNA (Fig. III-5C, lane 1). However, decreasing the efficiency of D4 usage in the 2-intron minigene by mutations of the GAR ESE or the 5' ss itself, relieved the preferential usage of D4 and nearly equals usage of both 5' ss. After inactivating the GAR ESE, D1 usage (1.7-mRNA) even outcompeted D4 usage (Fig. III-5C, lane 2). However, the only suboptimal activation observed for D1 in the parental 2-intron minigene likely applies also to the removal of the proximal intron in mutant minigenes. Therefore, suboptimal activation of D1 is suggested to account for the increased expression of completely unspliced RNA observed after mutating the central exon.

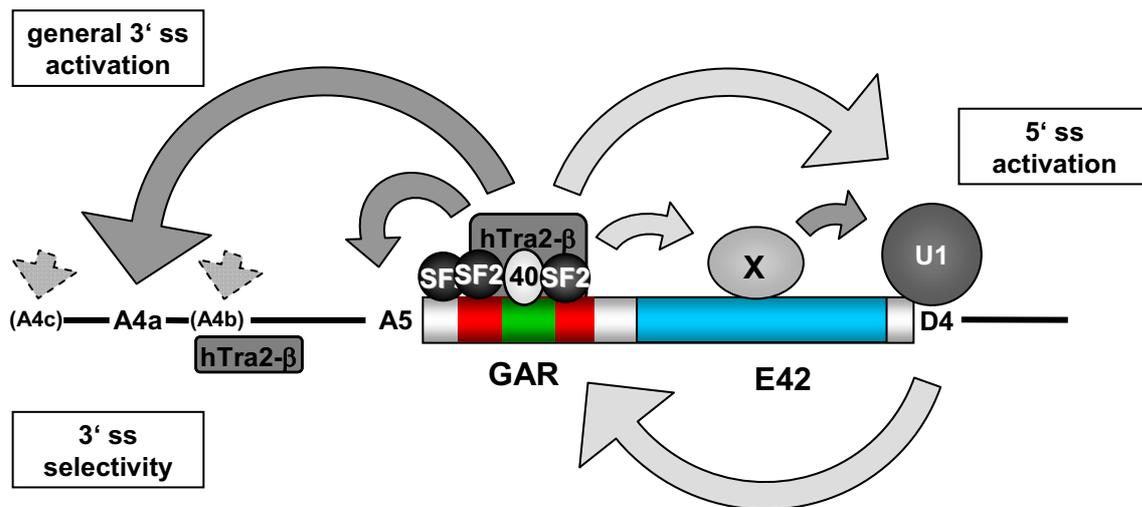
#### **D.1.2 Exon recognition is synergistically activated by the SR protein binding sites of the GAR ESE.**

Whereas most of the HIV-1 pre-mRNA sequence is removed by extensive splicing in the early phase of viral gene expression, the internal exons 4c, 4a, 4b and 5 have been described to be frequently included into the 1.8 kb mRNA class due to activation of the respective alternative 3' ss and 5' ss D4 (232, 317, 429, 481, 493, 495, 527). The results of this thesis evidence that the three binding sites for the SR proteins SF2/ASF and SRp40 initially described to form the GAR ESE (91) individually contribute to the

recognition of the internal exons 4a and 5 in the 2-intron minigene (Fig. III-5). In doing so, the individual SR protein binding sites differed in their extent of activating exon recognition [SF2(1) > SF2(2) > SRp40], albeit at an overall rather low level.

Inclusion of the internal exons 4a and 5 into mature mRNA isoforms was synergistically mediated by the presence of two SR protein binding sites of the GAR ESE (Fig. III-5D). Although already one SR protein binding site induced a marginal level of exon recognition, the combination of any two SR protein binding sites in the GAR enhancer complex synergistically activated exon recognition way beyond the sum of the individual contributions. Two SF2/ASF binding sites mediated the inclusion of the internal exons 4a and 5 to a similar extent as an SRp40 and an SF2/ASF binding site, which is in line with the finding that the RS domains are interchangeable between different SR proteins for some enhancer functions (102, 597, 621). However, the presence of a third binding site only slightly increased exon recognition.

Synergy in regulatory processes, like e.g. alternative splice site activation, has been proposed to be mediated either by cooperative binding of regulatory factors to enhancer sequences or by enhancement of two or more inefficient steps of the regulated process by direct interactions (257). Analysing the function of the GAR ESE, any single SR protein binding site was found to be already sufficient for sole activation of 5' ss D4 (Fig. III-5), but not for efficient exon inclusion, suggesting downstream directed cross-exon interactions of the GAR enhancer. Since any synergistic enhancement of splicing at D4 was observed in the presence of two or more SR protein binding sites of the GAR ESE, it seems highly unlikely that its synergistic effect on exon recognition is mediated through cooperative binding of SR proteins, as has been found in the *Drosophila doublesex* repeat element (258, 374). Previous work of our group also could not confirm a synergistic effect of the individual SR protein binding sites of the downstream GAR ESE on activation of a heterologous 3' ss (91). Therefore, it is hypothesised in this work that the synergy found in exon recognition (Fig. III-5) stems from bidirectional interactions emanating from the enhancer into opposite directions, the 3' ss and 5' ss, thereby promoting different steps in the spliceosome assembly at the respective upstream and downstream intron (Fig. IV-1).



**Fig. IV-1: Bidirectional cross-exon interactions emanating from the GAR ESE ensure internal exon recognition.**

SR protein binding sites of the GAR ESE ensure internal exon inclusion into the HIV-1 mRNA presumably by generating downstream as well as upstream directed interactions enhancing the recognition of the flanking splice sites. Although only 3' ss A4a and 5 were activated in the parental 2-intron minigene, the enhancer is essential for activation of all 3' ss within the cluster preceding exon 5. The enhancer was likewise essential for activation of 5' ss D4. However, it remains to be determined whether GAR ESE-bound proteins directly promote D4 recognition or whether activation is transferred by an yet unidentified factor (x) binding to the E42 fragment. In any case, the E42 sequence was found to be crucial for exon recognition and thus an interaction between proteins binding to the E42 sequence and the U1 snRNP recognising D4 is assumed. U1 snRNP binding at D4 additionally supports 3' ss selectivity mediated by the GAR ESE suggesting an also upstream directed interaction of U1 snRNP. Positioning of hTra2- $\beta$  binding at the GAR ESE and at the intronic purine-rich region proximal to A4b is based on the consensus binding motif of its RRM (586, 599) [40: SRp40, SF2: SF2/ASF, U1: U1 snRNP].

A similar synergistic action has been reported for three purine-rich elements regulating exon v3 inclusion into the *CD44*-mRNA (625), but has not automatically to be associated with every bidirectional enhancer (339). Therefore, synergistic splicing enhancer action may hint to different regulatory mechanisms than present in cases of additive enhancer action.

Synergy in complex assembly has been implicated as a molecular switch for regulatory processes (257). The synergistic activation of internal exon inclusion by the three distinct binding sites for the SR proteins SF2/ASF and SRp40 of the GAR ESE might allow the virus to switch from the shortest *nef1*-mRNA isoform, consisting only of the terminal exons 1 and 7 of the HIV-1 pre-mRNA, to internal exon-containing mRNAs coding for the regulatory proteins Rev or Nef. Additionally, cell type-dependent differences in the expression of the HIV-1 mRNA isoforms had been reported. In particular, it has been shown that exons 4c, 4a, 4b and 5 are less frequently included into HIV-1 mRNAs during infection of monocyte-derived macrophages (MDM) compared to CD4<sup>+</sup> lymphocytes (495, 570). Based on experimental data showing that mutations of the GAR ESE lower the inclusion of the internal exons (Fig. III-5), reduced exon inclusion in macrophages might be caused by impaired GAR enhancer activity. Although SF2/ASF expression in MDM was recently confirmed (161), the amount of this splicing regulatory protein might be too low to ensure efficient GAR enhancer activity.

In contrast to the synergistic activation of exon recognition by two SR protein binding sites, the presence of a third SR protein binding site of the GAR ESE exerted an only slightly additive effect on the inclusion of the internal exons 4a and 5 (Fig. III-5C). However, all three SR protein binding sites of the enhancer became essential for efficient activation of the 3' ss cluster generating intron-containing *vpu/env*-mRNA (Fig. III-6) in the intermediate and late phase of viral gene expression. Expression of intron-containing *vpu/env*-mRNAs is initiated by the viral protein Rev connecting mRNA isoforms unspliced in the distal intron defined by D4 and A7 to the Crm1-dependent export pathway (186, 194, 443, 654). Generation of *vpu/env*-mRNAs might require an improved splicing kinetic to efficiently remove the proximal intron upstream of the alternative internal exons prior to Rev-mediated nucleocytoplasmic export. Therefore, the enhancer might have to act more rapidly on 3' ss activation to mediate splicing of the proximal intron before the RNA is exported. This accelerated splicing kinetic might be facilitated by three SR protein binding sites more efficiently stabilising spliceosomal proteins at the 3' ss cluster.

### **D.1.3 The GAR ESE mediates 3' ss selectivity by its proximal SF2/ASF binding sites.**

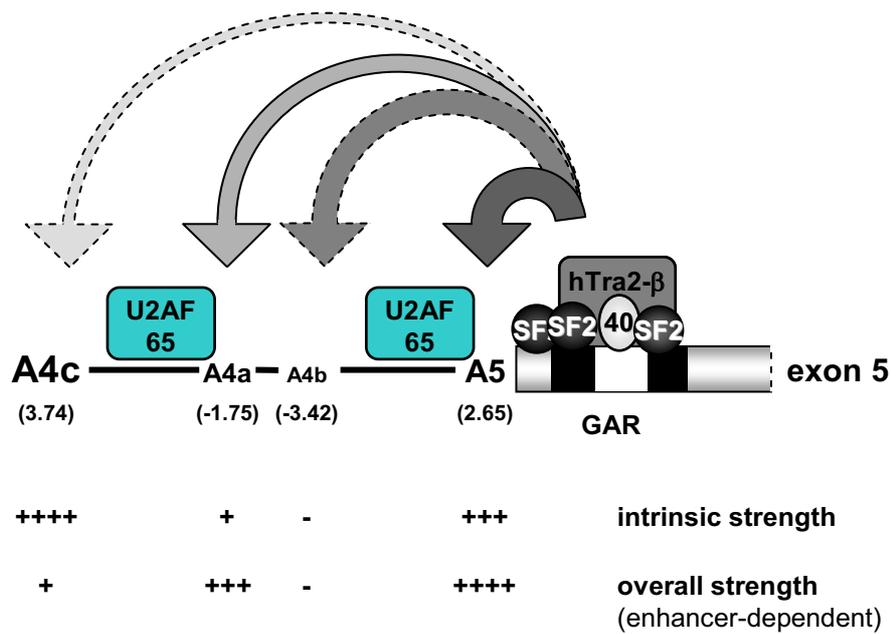
The internal exon 5 is preceded by four alternative 3' ss, from which A5 is located most downstream and flanks exon 5. Usage of one of the three 3' ss upstream of A5 results in mRNA isoforms encoding the viral regulatory protein Rev, whereas usage of A5 leads to the generation of an mRNA isoform encoding Nef. In this work, the preferential activation of A5 was identified as an additional function of the GAR ESE resulting in the predominant inclusion of the alternative exon 5 compared to exon 4a in the 2-intron minigene (Fig. III-5E). Unbalanced activation of the competitive 3' ss A4a and A5 has also been observed during viral gene expression and revealed a considerably higher expression of mRNAs using A5 than any other 3' ss of this cluster (232, 481, 495), leading to Nef-encoding mRNAs required in the early phase of viral gene expression. Nef fulfils a number of regulatory functions during viral replication, e.g. cellular surface receptor down-modulation (146), which are executed through mostly weak interactions of distinct Nef domains with a variety of cellular proteins (205). An elevated Nef concentration in the infected cell might thus contribute to efficient viral replication. Increased inclusion of the alternative internal exon 5 compared to exon 4a, as detected in the experiments presented here, may thus lend a further functionality to the GAR enhancer for viral replication.

From a kinetic point of view, it might be considered that alternative splice sites are chosen according to the spliceosomal assembly rate at the respective 3' ss. Recognition of the 3' ss might at first be mediated by the sequence elements reported to directly interact with spliceosomal components and thereby constitute the intrinsic splice site strength, i.e. the BPS, the PPT and the downstream flanking nucleotides of the AG-dinucleotide. In the 2-intron minigene all 3' ss clustering upstream of exon 5 are considered to exhibit a low intrinsic strength due to non-consensus BPS and discontinuous PPTs. This suggestion was supported by an early work reporting that the 3' ss cluster was only inefficiently used in a heterologous 1-intron minigene derived from the  *$\beta$ -globin* gene (447). However, due to an incomplete knowledge on *cis*-acting splicing regulatory elements at that time, the upstream located splicing silencer element ESS2 (11) was also included in the heterologous minigene. More recently, the weak intrinsic strength of all 3' ss within the cluster was experimentally confirmed by our group using a subgenomic 1-intron minigene lacking the downstream exon 5 and the upstream ESS2 (296). Nevertheless, the 3' ss within the cluster can still be ranked

---

according to their intrinsic strength using the MaxEntScore algorithm [[http://genes.mit.edu/burgelab/maxent/Xmaxentscan\\_scoreseq\\_acc.html](http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html) (669)] resulting in A4c (3.74) > A5 (2.65) >> A4a (-1.75) > A4b (-3.42) (Fig. III-7A).

From the intrinsic strength one would expect that A4c is predominantly activated due to its characterisation as strongest 3' ss within this cluster (MaxEntScore 3.74). However, activation of A4c was not observed in the parental 2-intron minigene (Fig. III-8B, cf. lane 1 with lanes 5 and 7). Instead A5 was predominantly used for internal exon inclusion, although this 3' ss is classified to exhibit a weaker intrinsic strength (MaxEntScore 2.65). These results argue against a strict control of the 3' ss cluster by splicing kinetics based only on their intrinsic strength. This conclusion is further substantiated by the finding that even activation of A4a – featuring a considerable weaker intrinsic strength (MaxEntScore -1.75) than A4c – was clearly favoured over A4c usage. Remarkably, the 3' ss located in close proximity to the GAR ESE, i.e. A5 and A4a, were more frequently used than expected from their intrinsic strength. This disproportionate activation of the 3' ss cluster indicates that the GAR ESE contributes to the overall strength of the 3' ss with an impact correlating to the proximity of the enhancer and 3' ss. The SR protein-mediated enhancement of complex assembly on the closely neighbouring splice sites thereby not only equalises differences in the intrinsic strength of the 3' ss, but even results in a stronger overall strength for 3' ss A5 and A4a. A similar concept was suggested by Krainer and colleagues proposing that SR protein binding sites might enhance recognition of neighbouring 5' ss by U1 snRNP thereby rendering both splice sites equally competitive (170). From this a model is evolved, in which the GAR ESE supports the recognition of the 3' ss located in close proximity by increasing recognition by spliceosomal components (Fig. IV-2).



**Fig. IV-2: Model of GAR ESE-mediated stabilisation of spliceosomal components depending on the proximity of the respective 3' ss within the cluster preceding HIV-1 exon 5.**

Despite considerable differences in the intrinsic splice site strength, each 3' ss of the cluster was efficiently used for exon inclusion in the absence of competing 3' ss. Therefore, their usage in the HIV-1 pre-mRNA depends on the combination of the intrinsic strength and the position to the GAR ESE relative to other 3' ss within the cluster. The GAR enhancer presumably stabilises spliceosomal components at the 3' ss most effectively at the nearest upstream 3' ss A5. Although A4a is located upstream of A5, its usage is also activated by the GAR ESE. However, both 3' ss interfere with the effect of the enhancer on the more upstream located 3' ss A4c. Disuse of A4b might relate to its low intrinsic strength or, alternatively, to its specific function for A4a activation. The intrinsic splice site strength (given in brackets) for the 2-intron minigene was classified using the MaxEntScore algorithm and is illustrated by the size of the splice site numbering, with A4c representing the strongest 3' ss within the cluster.

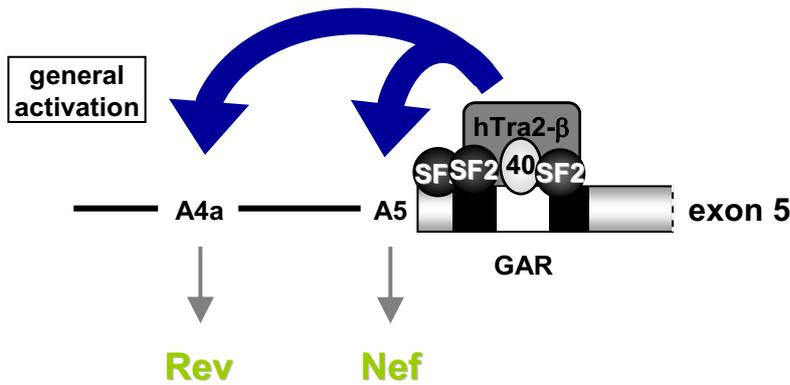
The model that the GAR ESE supports 3' ss recognition dependent on the distance from the respective 3' ss is only opposed by the disuse of A4b located between both activated 3' ss. However, based on the lowest intrinsic strength, the overall strength of A4b even enhanced by the GAR ESE might be too weak to compete with the neighbouring 3' ss. Alternatively, disuse of A4b might originate from its specific function within the 3'ss cluster, which was deduced from the observation that inactivating A4b also abrogates splicing at A4a (Fig. III-8B, lane 3). Therefore, the AG-dinucleotide of

A4b appears to essentially facilitate A4a usage. This result obtained in the minigene experiment is consistent with reports showing that A4b mutation also abrogates splicing at 3' ss A4a *in vitro* (594) and during viral infection *in vivo* (493). A similar dependency of two alternative 3' ss had been observed for the small tumour antigen mRNA (StAg) of the polyoma virus, where activation of the StAg 3' ss requires an additional 3' ss located 14 nt downstream (203). Another example was subsequently characterised within the regulation of the *Drosophila sexlethal (sxl)* exon 3 (469), in which activation of the proximal 3' ss is assisted by a distal 3' ss recruiting U2AF<sup>35</sup>. Due to the suggested function of A4b for activation of the upstream 3'ss, it is conceivable that A4b is occupied by splicing-related proteins necessary for A4a activation, which either block the enhancing effect of the GAR ESE on A4b or, alternatively, even transfer the enhancing effect onto A4a.

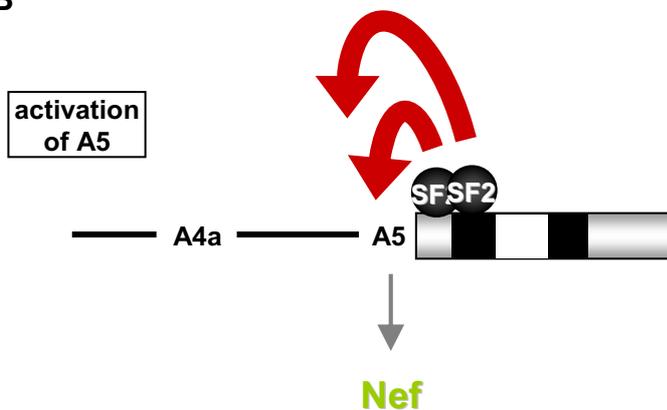
Loss of efficient 3' ss activation for internal exon inclusion after E42 deletion indicates that in the absence of the 3' half of exon 5 all SR protein binding sites of the GAR ESE are needed to stabilise binding of spliceosomal proteins at D4 (Fig. III-13B, lane 3). In this context, activation of the 3' ss cluster depends on the intrinsic strength as evidenced by the sole activation of A4c characterised by the highest intrinsic strength within the competing 3' ss cluster. This observation further substantiates the model that the GAR ESE confers 3' ss selectivity by stabilising spliceosomal components at the 3' ss and thereby overrides their intrinsic splice site strength.

Predominant activation of A5 within the 3' ss cluster was found to be exerted by the two proximal SF2/ASF binding sites of the GAR ESE (Fig. III-7B). Unexpectedly, both, the third SF2/ASF binding site located most downstream of the 3' ss cluster [SF2(2)] and also the SRp40 binding site, are not involved in 3' ss selectivity but solely contribute to overall exon recognition (Fig. IV-3).

**A**



**B**



**Fig. IV-3: SR protein binding sites of the GAR ESE fulfil a dual role in 3' ss activation for internal exon recognition.**

**(A)** All SR protein binding sites of the GAR ESE generally activate the 3' ss cluster upstream of exon 5. Nevertheless, due to differences in the position of the enhancer relative to the respective 3' ss within the cluster and their intrinsic strength A5 and A4a are exclusively but non-uniformly activated generating mainly *nef*- but also *rev*-mRNAs.

**(B)** The proximal SF2/ASF binding sites of the GAR ESE specifically activate A5 leading to the predominant expression of *nef*-mRNA. The exclusive ability of the most upstream SF2/ASF binding sites to specifically activate the most proximal splice site supports the suggestion that molecular interactions specifically activating A5 might differ from those involved in general 3' ss activation within the cluster.

The differential effects of the proximal and distal SR protein binding sites of the GAR ESE on 3' ss selectivity could emerge either from their distinct positions relative to the 3' ss cluster or from the proteins that bind to the binding sites. Distinct consensus binding sites for SR proteins have been identified using different SELEX methods (Systematic Evolution of Ligands by Exponential Enrichment). However, the RRM of

SR proteins bind to highly degenerated sequence motifs in splicing regulatory elements. Although binding of SF2/ASF to the GAR ESE sequence was confirmed *in vitro* (Fig. III-4B, cf. lane 1 with 4), the contribution of the respective SF2/ASF binding sites to overall SF2/ASF binding of the enhancer was not evaluated. Therefore, the distal SR protein binding site of the GAR ESE might also be recognised by other splicing regulatory proteins that do not support A5 activation.

As alternative explanation it might be proposed that the predominant activation of A5 by the proximal SF2/ASF binding sites depends on the proximity of the bound proteins to the 3' ss. This would imply that the GAR ESE-mediated activation of A5 requires a stereospecific confirmation of the enhancer complex assembling at the 3' ss that allows an SF2/ASF domain to activate A5. So far, it is an open question which protein domain of SF2/ASF mediates the pronounced activation of A5. SR proteins are likely to mediate the function of the GAR ESE through their RS domain, which is known to interact with proteins (318, 655) and has more recently been shown to also stabilise double stranded RNAs in pre-mRNA-U snRNA intermediates during spliceosomal assembly (544, 545). Both protein-protein and RNA-protein interactions might be involved in the 3' ss selectivity performed by the GAR ESE. An interaction of an RS domain with double stranded RNA might depend on their stereo-specific arrangement, and might therefore be restricted to the proximal SF2/ASF binding sites. In contrast, protein-protein interactions between the RS domains and proteins of the splicing machinery might be less reliant on stereo-specific requirements, and thereby might act more uniformly on the whole 3' ss cluster.

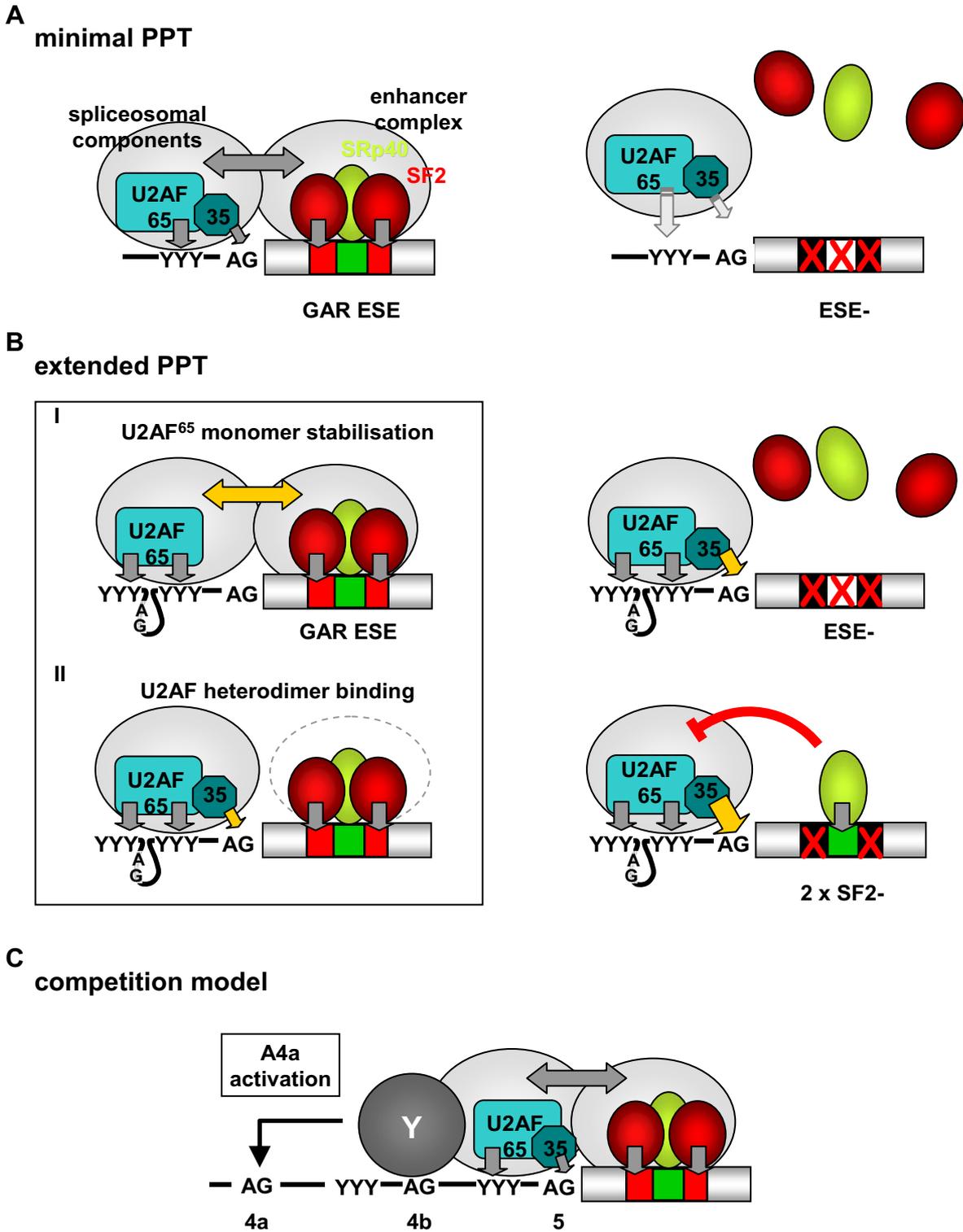
#### **D.1.4 Mechanism of GAR ESE function**

The GAR enhancer complex exerts a dual splicing regulatory function by (i) generally activating all 3' ss of the cluster and (ii) by specifically activating A5 by its proximal SF2/ASF binding sites (Fig. III-5 and Fig. III-7). Preferred activation of A5 is caused solely by the two proximal SR protein binding sites of the GAR enhancer complex contrasting the participation of all examined SR protein binding sites in general 3' ss activation (Fig. IV-3). The distinct SR protein binding site requirements for the respective enhancer function hint to distinct mechanisms applied by the GAR ESE to activate specifically A5 or generally all 3' ss within the cluster. Although many splicing regulators were found to affect early E complex assembly (223, 408, 517, 619, 638,

640, 690, 700), the picture arising from more recent studies implicates that splicing can be regulated throughout spliceosomal assembly and catalysis. For instance, it has been shown that regulation of alternative splicing can occur during the transition from A to B complex (265, 541) or between both transester reactions catalysed in the C complex (338). Therefore, general 3' ss activation might be enhanced by the GAR ESE during early complex formation or alternatively in subsequent assembly steps. In contrast, 3' ss selectivity exerted by the proximal SF2/ASF binding sites of the GAR ESE is most likely performed in a spliceosomal assembly step prior or up to A complex formation, since formation of this complex commits a splice site pair to one spliceosome (323, 355).

Regulation of E complex assembly via modulation of U2AF heterodimer binding has emerged as a common mechanism employed by numerous splicing enhancer (223, 638, 700) and silencer elements (408, 517, 619, 640, 690). Modulation of U2AF<sup>65</sup> binding had already been reported for two splicing regulatory elements of the HIV-1 pre-mRNA. The presence of an exonic splicing silencer element (ESS V) decreases U2AF<sup>65</sup> binding at the upstream 3' ss A2 (157), whereas stabilisation of U2AF<sup>65</sup> was observed by an exonic splicing enhancer located downstream of 3' ss A7 (ESE3) (604). Analysis of splicing regulation in two different HIV-1 *tat* pre-mRNA substrates revealed that the RS domain of SF2/ASF appears to be dispensable for activation of 3' ss A7 by ESE3 (604, 692). This observation led to the suggestion that SF2/ASF bound at ESE3 might enhance splicing mainly by counteracting the neighbouring ESS3 element (604, 692). Alternatively, it was hypothesised that binding of recombinant SF2/ASF lacking the RS domain might recruit other positively acting splicing regulatory factors (604). In this thesis, no experimental evidence was obtained supporting the existence of a splicing silencer within exon 5. Therefore, it was hypothesised that activation of A5 is likely performed by a direct positive mechanism of the proteins bound to the GAR ESE presumably through stabilisation of spliceosomal components or regulatory proteins at the flanking splice sites. Investigating the mechanism underlying splicing regulation of the internal exon 5, results of this thesis revealed that the GAR ESE stabilises binding of both U2AF subunits at 3' ss A5 in the presence of a minimal PPT (Fig. III-14B, lanes 4-6). Since binding of U2AF<sup>65</sup> to A5 RNA targets was equally diminished in the presence of mutations that caused either a shift in 3' ss selectivity or a decreased overall exon inclusion *in vivo*, stabilisation of U2AF<sup>65</sup> by the GAR ESE is suggested to be involved in specific activation of A5 rather than in general exon inclusion.

Unexpectedly, extending the RNA targets by six additional intronic nucleotides rendered U2AF<sup>65</sup> binding insensitive to GAR ESE mutations (Fig. III-14B, lanes 1-3). Additionally, binding of U2AF<sup>35</sup> even increased after introducing the GAR ESE mutations thereby opposing its binding pattern to minimal PPT-containing RNAs. This indicates that the amount of U2AF<sup>35</sup> bound to A5 in the presence of the GAR ESE in the extended as well as in the short RNA target was below saturation and points to considerably frequent binding of U2AF<sup>65</sup> monomers in the absence of U2AF<sup>35</sup>. Increased U2AF<sup>35</sup> binding to the RNA targets upon mutation of the GAR ESE might be caused by reduced sterical hindrance between U2AF<sup>35</sup> and enhancer-bound SR proteins, which might otherwise compete with U2AF<sup>35</sup> for RNA binding. However, this view is difficult to reconcile with the finding that mutating three SR protein binding sites of the GAR ESE resulted in an only intermediate increase in U2AF<sup>35</sup> binding, whereas mutation of only two SR protein binding sites allowed a stronger binding of U2AF<sup>35</sup>. Since the U2AF heterodimer was reported to possess a higher RNA binding affinity than the respective monomeric subunits (408, 506, 656), the GAR ESE might alternatively stabilise binding of U2AF<sup>65</sup> to the PPT as a monomer (Fig. IV-4). Therefore, mutations of the GAR ESE would select for higher levels of U2AF heterodimers at 3' ss A5. Since the presence of the SRp40 binding site of the GAR enhancer induced a stronger proportion of U2AF heterodimer binding to the RNA target than observed after complete inactivation of the GAR ESE, it seems that SRp40 binding at the GAR enhancer interferes with efficient binding of U2AF<sup>65</sup> monomers at the PPT of A5. This suggestion is in line with the predominant activation of A4a in the 2-intron minigene after mutation of two SF2/ASF binding sites (Fig. III-5C and E, lane 5). It should be noted that the mutated sequences inducing stronger binding of U2AF<sup>35</sup> reduced exon recognition and 3' ss selectivity *in vivo*, therefore increased U2AF<sup>35</sup> binding appears not to be involved in the mechanism underlying GAR ESE function.



**Fig. IV-4: Model of GAR ESE-mediated stabilisation of U2AF<sup>65</sup> monomers upstream of A5.**

(A) The presence of a minimal PPT presumably facilitates U2AF<sup>65</sup> binding only via a single RRM. The GAR ESE is assumed to stabilise the interaction of the U2AF heterodimer by interacting with the U2AF<sup>65</sup> subunit (left). After inactivation of the GAR ESE the U2AF heterodimer lacks stabilisation by the enhancer and can only inefficiently bind to the PPT upstream of A5 (right).

The interpretation that the GAR ESE enhances binding of the U2AF<sup>65</sup> monomer to A5 in the absence of U2AF<sup>35</sup> is in line with early results showing that in the presence of a sufficient long PPT the first step of the splicing reaction can be carried out in the absence of the 3' ss AG-dinucleotide (487). Reconsidering this correlation in the light of more recent findings, the necessity of a long PPT adjacent to the BPS for the first step of the splicing reaction might reflect the requirement for efficient binding of U2AF<sup>65</sup> to initiate spliceosome assembly. The PPT upstream of A5 provided in the *in vitro* binding assays consists of 11 nt and 20 nt, respectively, in short and extended RNA substrates, which are disrupted by purines. Since a PPT of 14 nt was shown to be unable to conduct the first step of the splicing reaction in the absence of the AG-dinucleotide, whereas a PPT containing 28 nt led to the formation of branched lariat intermediates [(487), reviewed in (424)], the PPT directly upstream of exon 5 is presumably shorter than the PPT of introns that exerted the first step of the splicing reaction in the absence of the AG-dinucleotide. However, the requirement for a high-quality, i.e. a long and continuous, PPT might be bypassed in HIV-1 exon 5 splicing regulation by the stabilising effect of the GAR ESE on U2AF<sup>65</sup> binding. Therefore, mutations of the enhancer might reduce U2AF<sup>65</sup> monomer binding resulting in a relative accumulation of U2AF heterodimers at A5 that have been reported to exert a higher RNA-binding affinity than the individual monomers (408, 506, 656).

**Fig. IV-4: continued.**

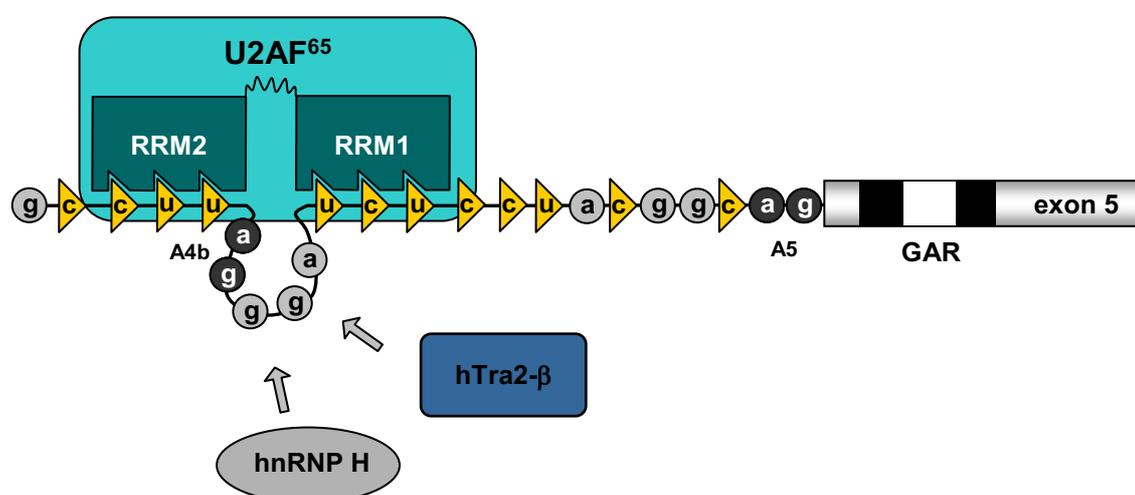
**(B)** Extending the PPT upstream of A5 by six additional nucleotides renders U2AF heterodimer binding GAR ESE-independent. Additional pyrimidines upstream of the AG-dinucleotide of A4b might allow U2AF<sup>65</sup> to recognise the PPT via two RRM. The GAR ESE is able to efficiently stabilise U2AF<sup>65</sup> monomers at the PPT in the absence of U2AF<sup>35</sup> (I). However, also GAR ESE-independent binding of the U2AF heterodimer to the extended PPT might be feasible due to the increased RNA affinity of the U2AF heterodimer compared to the respective subunits (II). After inactivation of the GAR ESE, U2AF heterodimers are still able to bind to the RNA target due to the extended PPT (right, ESE-). However, the proportion of U2AF<sup>35</sup> binding to the RNA target increases, because RNA binding solely via two RRMs only insufficiently attaches U2AF<sup>65</sup> monomers to the PPT. The SRp40 binding site of the GAR ESE was found to increase the amount of U2AF<sup>35</sup> at the RNA target. Since this binding site led to predominant activation of A4a, it appears that preferential A4a activation is achieved by hindering efficient U2AF<sup>65</sup> binding. The interference with U2AF<sup>65</sup> binding is assumed to result in a higher proportion of U2AF<sup>35</sup> binding than after inactivation of the complete GAR ESE. Essential interactions for U2AF<sup>65</sup> binding at the RNA are indicated by yellow arrows.

**(C)** *In vivo* the GAR ESE might essentially activate A5 by stabilising the U2AF heterodimer in the presence of a competing factor binding at 3' ss A4b and inducing A4a usage.

Enhanced binding of U2AF<sup>65</sup> mediated by the GAR ESE likely favours A5 usage by improving the first step of the splicing reaction. Since all 3' ss of the cluster upstream of exon 5 use distinct BPS, accelerating the first transester reaction choosing one of the BPS would thereby commit the spliceosome to the respective AG-dinucleotide. GAR ESE-mediated stabilisation of U2AF<sup>65</sup> at the PPT of A5 is only able to enhance activation of A5, but is unlikely to regulate usage of the upstream located 3' ss within the cluster. Nevertheless, the GAR ESE also ensures exon recognition via 3' ss A4a. GAR ESE-mediated activation of A4a – and of the remaining 3' ss in the cluster – might require a specific mechanism for activation as it had been observed for A5 in this thesis. It would be feasible that U2AF binding at A4a is likewise increased by a stabilising interaction with the enhancer. However, since general exon recognition and thus activation of A4a is mediated also by the distal SR protein binding sites of the enhancer, it appears unlikely that SR protein binding sites located more distant from the 3' ss cluster should exert the same stabilising interaction on U2AF<sup>65</sup> at A4a positioned upstream of A5. Alternatively, it might be hypothesised that for general 3' ss activation in the presence of the GAR ESE a second mechanism, distinct from that specifically activating A5, promotes an additional step in spliceosomal assembly that is common to all alternatively used 3' ss upstream of exon 5.

In summary, the differential changes in the binding of U2AF<sup>65</sup> and U2AF<sup>35</sup> to extended RNA targets upon mutation of the GAR ESE support two conclusions: first, GAR ESE-dependent binding of U2AF<sup>35</sup> to short RNA substrates containing a minimal PPT is proposed to occur via U2AF<sup>65</sup> and not by direct interactions with enhancer bound proteins. Therefore, in short RNA targets harbouring a suboptimal PPT the GAR ESE stabilises mainly U2AF<sup>65</sup>. This result contrast previous results showing that SF2/ASF directly interacts with U2AF<sup>35</sup> (655, 660, 668). However, a stabilising effect of SF2/ASF on U2AF<sup>35</sup> might depend on the respective stereospecific arrangement of the enhancer complex. Therefore, not every SF2/ASF-dependent enhancer necessarily has to act via stabilisation of U2AF<sup>35</sup> (398). Second, the nucleotide extension upstream of the PPT relieves the enhancer-dependency of U2AF<sup>65</sup> binding to the analysed RNA targets *in vitro*, presumably due to additional pyrimidines included in the added sequence (Fig. IV-4). However, the high U2AF<sup>65</sup> concentration in the HeLa nuclear extract might be crucial to allow this efficient enhancer-independent binding. Since high levels of spliceosomal components are able to compensate the necessity for otherwise essential splicing regulatory factors (700), limiting the amounts of spliceosomal proteins in the

binding reaction might restore a stabilising effect of the GAR ESE on U2AF<sup>65</sup> binding in the presence of additional pyrimidines. Nevertheless, extending the PPT upstream of A5 might strengthen U2AF<sup>65</sup> binding by providing a larger interaction interface between U2AF<sup>65</sup> and the RNA. Although a discontinuous PPT extension was used in this work, it is conceivable that interfering purines are looped out and thereby provide a nearly continuous linear stretch of pyrimidines (27) improving the affinity of the U2AF<sup>65</sup> binding site (Fig. IV-5).



**Fig. IV-5: Loop out model of interfering purines in the PPT upstream of A5.**

Two RNA binding motifs (RRM) of U2AF<sup>65</sup> might bring distant pyrimidines (yellow triangles) into proximity thereby generating a loop of interfering purines (grey circles) [adapted from (27)]. The purines, which in this model are not occupied by U2AF<sup>65</sup>, may be accessible for additional proteins like e.g. hTra2- $\beta$  or hnRNP H identified to bind to the intron/exon 5 RNA *in vitro*. The invariant AG-dinucleotides of A4b and A5 are denoted in white letters.

Although the GAR ESE appeared to lose its stabilising function on U2AF<sup>65</sup> binding with increasing intron length in the *in vitro* assay, the relevance of the GAR ESE's potential to stabilise U2AF<sup>65</sup> binding to a suboptimal PPT might become apparent in the light of the *in vivo* situation. Without taking into account the presence of alternative 3' ss upstream of A5, it might be suggested that U2AF<sup>65</sup> binds to the 3' ss without being

assisted by an enhancer element. However, considering the close proximity of the four alternatively used 3' ss upstream of exon 5, a more competitive view of 3' ss recognition is more likely and is also consistent with the experimental data obtained after inactivation of the individual 3' ss (Fig. III-8B). The intronic sequence added in the extended RNA targets constitutes the BPS of A5, but at the same time contains the AG-dinucleotide of A4b, which was demonstrated to be essential for A4a activation in this work and also by others (Fig. III-8B) (493, 594). Additionally, mutation of A4b also resulted in a compensatory increase of A5 usage (Fig. III-8B, cf. lanes 1 and 3) (481, 594) leading to the suggestion that splicing regulatory factors bound at A4b might sterically interfere with BPS recognition for A5 usage (588, 594). Therefore, *in vivo* binding of U2AF<sup>65</sup> to the nucleotide extension containing the AG-dinucleotide of A4b as well as additional pyrimidines might be competed by proteins mediating A4a activation and thus might depend on the GAR enhancer for efficient binding (Fig. IV-4C). The fact that this proposed competition was not detected in the *in vitro* binding studies presented in this thesis might be due to the terminal position of the short nucleotide extension eventually hampering recognition of the AG-dinucleotide of A4b by spliceosomal components. In conclusion, the GAR ESE might be crucial to mediate U2AF<sup>65</sup> binding to the PPT thereby enhancing E complex formation under competitive conditions *in vivo*. This appealing hypothesis will be further investigated using RNA targets incorporating additional intronic sequences up to A4c.

Whereas a low level of U2 snRNP associates with to the parental as well as mutant sequences as monitored by SAP155 binding, SF1/mBBP did not at all bind to the RNA targets containing extended PPTs (Fig. III-15B). This was an unexpected finding, since binding of SF1/mBBP is thought to guide binding of U2 snRNP to the BPS, which concomitantly replaces SF1/mBBP. It could be argued that the replacement of SF1/mBBP with U2 snRNP at the BPS proceeds immediately after recognition of the BPS resulting in an only transient binding of SF1/mBBP that could not be detected. However, a stronger binding of SAP155 would be expected to substantiate this scenario. Therefore, it is proposed that the low levels of SAP155 bind without assistance of SF1/mBBP *in vitro* and display U2 snRNP binding to the intron/exon 5 target either via protein-protein interactions with U2AF<sup>65</sup> or by direct recognition of the BPS. SF1/mBBP might have failed to bind to the RNA target, because the BPS used deviates from the consensus sequence. However, since the BPS constitutes the 5'-end of the RNA target in the assay used here, alternatively additional nucleotides upstream

might be required to allow efficient protein binding at the BPS *in vitro*. Nevertheless, SF1/mBBP appears not to be required for 3' ss recognition in general, because depletion of SF1/mBBP did not affect splicing of several endogenous and reporter mRNAs *in vitro* (510) and *in vivo* (602) and only slightly improves the kinetic of A complex assembly on an RNA substrate containing a strong, consensus BPS (236, 510). Therefore, the lack of SF1/mBBP in the RNA-protein complex assembled on extended RNA targets might hint to an SF1/mBBP-independent BPS recognition at HIV-1 A5.

The characteristic RS domain of the SR protein family is extensively phosphorylated [(376, 377, 503, 504), reviewed in (581)]. It has been reported that phosphorylation of SRp40 is essential for specific binding to its target sequence (595). Phosphorylation of the RS domain of SF2/ASF was also observed to affect the recruitment to its RNA target sequence and to reduce non-specific RNA binding (659). However, it was hypothesised that the RS domain itself does not constitute an RNA-binding motif but instead interferes in a non-phosphorylated state with sequence-specific binding of the RRM to the RNA due to its highly-charged residues, which are compensated by phosphorylation thus changing the binding characteristic from a sequence-independent to a sequence-specific RNA-protein interaction (595). With regard to the GAR ESE in HIV-1 exon 5, this thesis found that dephosphorylation of SF2/ASF dramatically reduces its binding to RNA targets containing the 5'-half of exon 5 and the upstream intronic region (cf. Fig. III-20C with Fig. III-14B). Only the parental GAR ESE in the presence of the minimal PPT was significantly bound by dephosphorylated SF2/ASF (Fig. III-20C, lane 4). However, this sequence already showed the strongest binding of phosphorylated SF2/ASF (Fig. III-14B, lane 4). Since phosphorylation of SF2/ASF was reported to enhance target specificity (659), residual binding of dephosphorylated SF2/ASF to the RNA target in the *in vitro* binding assay here might indicate a sequence-independent binding of SF2/ASF to the RNA target. However, this view is difficult to reconcile with the finding that sequence-independent binding was not observed on the other RNA targets carrying similar exonic sequences. Thus, the observed binding of dephosphorylated SF2/ASF to the GAR enhancer in RNA targets carrying the minimal PPT is more likely to represent reduced binding of SF2/ASF to the binding sites of the GAR ESE. These results imply that phosphorylation of SF2/ASF is essential for efficient binding of the protein to the GAR ESE.

Analysing the requirement of phosphorylation for the splicing regulatory function of SR proteins, it was reported that using non-phosphorylated SF2/ASF in a spliceosome formation assay neither A nor B or C complexes assemble on a pre-mRNA substrate and instead the amounts of the early H complex increased (90). Additionally, it was reported that phosphorylation of SF2/ASF is required for direct interaction with the splicing factor U1snRNP via the specific protein component U1-70K and for enhancer-dependent splicing of an HIV-1 *tat*-pre-mRNA substrate *in vitro* (659). Investigating the mechanism of 3' ss activation by the GAR ESE demonstrated that dephosphorylation of HeLa nuclear extract abrogated the GAR ESE-mediated stabilisation of the U2AF heterodimer at the RNA containing a minimal PPT that was observed in the presence of phosphorylated SR proteins (cf. Fig. III-14B with Fig. III-20C). This loss of U2AF stabilisation albeit low levels of dephosphorylated SF2/ASF bound (Fig. III-20C, lane 4) indicates that dephosphorylation not only impairs binding of SF2/ASF to the GAR ESE, but likely also affects direct or indirect interactions with U2AF causing stabilisation in untreated nuclear extracts. It should be noted that U2AF<sup>65</sup> as well as U2AF<sup>35</sup> have been reported to be phosphorylated (204, 402, 451). However, phosphorylated residues of U2AF<sup>65</sup> were found outside of the RRM and also outside of the regions that are involved in the interaction with U2AF<sup>35</sup> [aa 85-112 (508, 685)] and SF1/mBBP [aa 334-475 (43, 484)], indicating that its binding affinity to RNA is likely unchanged. Similar to U2AF<sup>65</sup>, phosphorylation of U2AF<sup>35</sup> was also reported outside of the RRM and with one exception outside of the domain necessary for the interaction with U2AF<sup>65</sup> [aa 47-172 (685)]. Additionally, recombinant peptides constituting the minimal U2AF regions required for interaction form stable heterodimers, which also bind to RNA (310). Although it cannot completely ruled out that dephosphorylation also affects the intrinsic binding ability of both U2AF subunits to the RNA, low levels of U2AF<sup>65</sup> binding to the RNA targets carrying mutations in the GAR ESE argue at least for an unaffected U2AF<sup>65</sup>-RNA interaction. Therefore, these results suggest that U2AF<sup>65</sup> binding to the RNA in dephosphorylated nuclear extracts is still feasible, whereas GAR ESE-mediated stabilisation of U2AF<sup>65</sup> fails. Although early interactions between non-phosphorylated as well as phosphorylated SF2/ASF, U1 snRNP and a 5' ss of a pre-mRNA substrate were observed (90, 284, 318), this does not necessarily hint to the formation of a respective complex at the 3' ss. These results contrast the finding that SF2/ASF bound at HIV-1 splicing enhancer ESE3 downstream of 3' ss A7 stabilises U2AF<sup>65</sup> binding independent of the RS domain, which indicated that phosphorylation is not required for SF2/ASF to

mediate splicing via U2AF<sup>65</sup> stabilisation (604). This difference in the requirement of phosphorylation for SF2/ASF-mediated splicing activation might be based on the respective 3' ss encoded by the RNA substrates used in this thesis and in the previous reports. The combination of different 3' ss with various splicing regulatory elements results in the assembly of highly diverse RNA-protein complexes [reviewed in (564)], in which interactions of the same protein, e.g. SF2/ASF, might involve different protein domains depending on the enhancer complex architecture.

The strong phosphorylation dependency of GAR ESE-mediated U2AF<sup>65</sup> stabilisation encouraged the evaluation of the influence of SF2/ASF phosphorylation on the splicing pattern of the HIV-1 pre-mRNA *in vivo*. To reduce the steady-state level of phosphorylated SF2/ASF, inhibitors of the cellular kinase Clk1/Sty reported to phosphorylate SF2/ASF (126, 127, 479) were applied to HeLa-T4<sup>+</sup> cells transiently transfected with the 2-intron minigene. However, reduced levels of phosphorylated SF2/ASF were not achieved using two alternative compounds (Fig. III-20A and data not shown) described to inhibit the activity of the Clk1/Sty kinase [(431) and personal communication by Prof. Dr. J. Tazi (IGM Montpellier, France)]. It could be assumed that the inhibitors failed to elicit a physiological response in the cell culture assay. However, changed cell morphology and reduced viability of treated cells argue against this possibility (Fig. III-20C). An alternative explanation for the at best modestly reduced phosphorylation of SF2/ASF might be that only very few residues of SF2/ASF are phosphorylated by Clk1/Sty. Therefore, specific inhibition of this kinase might induce only a minor shift in the apparent molecular weight during immunoblot analysis and thus could not be monitored. However, application of these chemical compounds to the cell culture medium did not change the splicing pattern of the 2-intron minigene *in vivo* (Fig. III-21 and data not shown). Assuming successful inhibition of Clk1/Sty-mediated phosphorylation of SF2/ASF, these results would suggest that the amino acids of SF2/ASF specifically phosphorylated by Clk1/Sty are not involved in the splicing regulatory function exerted by the GAR ESE on HIV-1 exon 5. However, since numerous kinases have been described to phosphorylate SR proteins (581), a third appealing hypothesis would be that impaired Clk1/Sty kinase activity might be compensated by other non-restricted cellular kinases like SRPK1 (233), SRPK2 (331, 633) and other members of cdc-like kinases, i.e. Clk2, Clk3 and Clk4, which are not affected by the kinase inhibitors used in this work. Especially, an interplay of SRPK1 and Clk/Sty has been reported to orchestrate SF2/ASF phosphorylation (444). This

hypothesis is supported by the recent finding that also in mouse P19 cells overall phosphorylation of SF2/ASF remains unaffected after treatment with TG003, one of the inhibitors used in this thesis (672). This interpretation might imply that inhibitors specifically targeting single kinases are presumably unable to reduce the amount of phosphorylated SR proteins sufficiently to validate the necessity for SF2/ASF phosphorylation for GAR ESE-mediated exon recognition and 3' ss selectivity. Moreover, in the light of these results it might be questioned whether the therapeutic promise of targeting spliceosomal assembly in missplicing-caused diseases can be achieved by targeting single kinases in the case of SR proteins phosphorylated by multiple kinases [(572), reviewed in (573)]. Whether phosphorylation of SF2/ASF is essential for exon recognition and 3' ss selectivity mediated by the GAR ESE *in vivo* remains to be elucidated using alternative approaches.

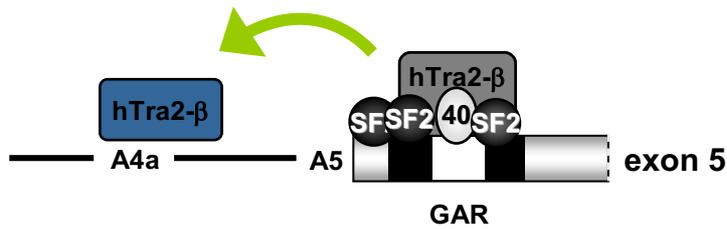
Investigating the mechanism of GAR ESE-mediated general 3' ss activation, binding of the splicing regulatory protein hTra2- $\beta$  was determined to correlate with the level of exon recognition observed in the minigene construct (cf. Fig. III-17B and Fig. III-5). Although not a classical SR protein, hTra2- $\beta$  was recently classified as an additional SR protein (357, 363). hTra2- $\beta$  was reported to display an RNA-binding specificity to purine-rich sequences similar to that of SF2/ASF (596, 598, 599). Cross-linking assays in nuclear extracts indicated that, perhaps as a result of the overlapping RNA-binding specificities, at least some purine-rich enhancers might bind multiple proteins, including SR and likely Tra2 proteins (485, 596). However, the binding pattern of hTra2- $\beta$  and SF2/ASF to the identical intron/exon 5 RNA targets differed in the *in vitro* binding analysis (cf. Fig. III-17B and Fig. III-14B). Whereas SF2/ASF still bound – albeit with a reduced amount - to RNA targets lacking the BPS after mutation of the SR protein binding sites of the GAR ESE, binding of hTra2- $\beta$  was completely abrogated. Since additional binding sites for SF2/ASF were predicted further upstream of A5 by the ESEfinder algorithm (data not shown), residual binding of SF2/ASF might indicate that these binding sites are also occupied by SF2/ASF. In contrast, loss of hTra2- $\beta$  binding to the RNA targets after mutation of two SF2/ASF binding sites of the enhancer implies that no binding sites for hTra2- $\beta$  are located outside of the GAR ESE in the RNA targets lacking the BPS. However, extending the RNA target at the 5'-end with the BPS increased hTra2- $\beta$  binding in the presence of the GAR ESE. This again contrasts the binding affinity of SF2/ASF, which was reduced after extending the RNA target. It was

generally observed in this thesis that the BPS increased hTra2- $\beta$  binding to the RNA targets, whereas binding of SF2/ASF was decreased in the presence of the BPS. From these results it can be concluded that although SF2/ASF and hTra2- $\beta$  presumably share binding sites within the GAR ESE sequence, they might participate in different interactions during 3' ss recognition and splicing at the 3' ss cluster.

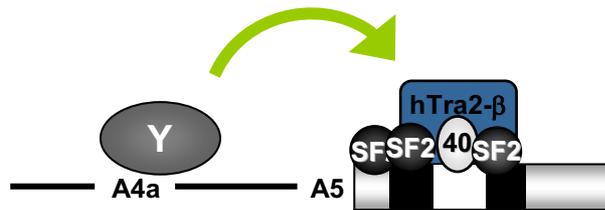
The hTra2- $\beta$  homolog in *Drosophila*, Tra2, constitutes one of the regulatory proteins essential for splicing activation mediated by the *Drosophila* doublesex enhancer in the *dsx* pre-mRNA (252, 277, 374, 611). In HeLa nuclear extracts Tra2 recruits the SR protein 9G8 to the *dsx*RE elements and is also able to mediate binding of SF2/ASF to the PRE element of the enhancer (375). Since neither Tra2 nor 9G8 alone bound significantly to the single repeat elements (252, 375), cooperative interactions between Tra2 and different SR proteins were assumed to generate highly specific RNA-protein complexes on the pre-mRNA (374, 375, 612). Consistent with the necessity of Tra2 for SR protein recruitment, Tra2 was found to directly interact with SR proteins like SF2/ASF (15, 655, 660). These results suggested that the human homolog, hTra2- $\beta$ , might also be able to recruit SR proteins, like SF2/ASF, into multi-protein enhancer complexes. However, this thesis provides evidence that binding of SF2/ASF to the HIV-1 exon 5 GAR ESE does not depend on simultaneous binding of hTra2- $\beta$  to the RNA as observed using RNA targets excluding the BPS (cf. Fig. III-14B with Fig. III-17B). Nevertheless, SF2/ASF, SRp40 and hTra2- $\beta$  might cooperatively regulate internal exon recognition by improving distinct steps in spliceosomal assembly.

The increased binding of hTra2- $\beta$  to BPS-containing RNA targets, compared to RNA targets lacking the BPS, suggests that the nucleotide extension either supports hTra2- $\beta$  binding at the GAR ESE through a stabilising interaction of a BPS-bound factor or generated an additional hTra2- $\beta$  binding site at the 5'-end of the RNA target (Fig. IV-6). The nucleotide extension (GCCUUA) does not match the classical purine-rich sequence that was reported to be recognised by hTra2- $\beta$  (586). Recently, an additional RNA-binding motif for the *Drosophila* homolog, Tra2, was identified (CAAGR) that tolerates a pyrimidine (482), but, however, also this motif is neither found in the intronic sequence nor in the nucleotide extension. Therefore, it is unlikely that the added sequence itself is bound by hTra2- $\beta$ . Nevertheless, in line with the proposal that the nucleotide extension contributes to a high-affinity U2AF<sup>65</sup> binding site by looping out interfering purines, hTra2- $\beta$  might recognise the assumed purine-rich loop (Fig. IV-5). Although likewise

A



B



**Fig. IV-6: Model of hTra2- $\beta$  stabilisation by ternary complex formation on BPS-containing intron/exon 5 RNA substrates.**

hTra2- $\beta$  may bind to two distinct sequences on the RNA substrate, which correspond to its reported purine-rich consensus binding motive, located either in the intron or within the GAR ESE itself. Since mutations of the GAR ESE completely abrogated hTra2- $\beta$  binding to short RNA targets, it is most likely that hTra2- $\beta$  binds to the GAR ESE itself. However, because of the increased, GAR ESE-dependent binding of hTra2- $\beta$  to extended RNA targets, an additional binding site in the nucleotide extension might be occupied by hTra2- $\beta$ , which is stabilised by GAR ESE-bound proteins (**A**). Alternatively, the BPS-extension might be recognised by an yet unidentified factor stabilising hTra2- $\beta$  at its binding site within the GAR ESE (**B**).

present in short RNA targets, the purine-rich sequence might be masked there by U2AF<sup>65</sup>. However, only five purines interrupt the PPT and might generate a loop after extension of the RNA target. Although the core RNA-binding motif identified in the *Drosophila* M1 intron of the *tra2* pre-mRNA consists of only 5 nt, several repeats of this core sequence appeared to be required for efficient Tra2-dependent splicing regulation in the *Drosophila* RNA substrate (482). Furthermore, earlier studies reported an 8 nt consensus sequence to constitute an RNA-binding site for hTra2- $\beta$  (586). Therefore, it

---

remains to be conclusively demonstrated whether hTra2- $\beta$  binds to the short purine stretch located within the PPT upstream of A5.

As an alternative to the generation of an additional binding site, binding of hTra2- $\beta$  to RNA targets containing the PPT might be increased by stabilisation at its binding sites within the GAR ESE. Since U2AF<sup>65</sup> binding did not increase in BPS-containing RNA targets compared to the short RNA targets in the presence of the GAR ESE, it appears unlikely that U2AF<sup>65</sup> participates in the increased binding of hTra2- $\beta$  in the presence of the BPS. However, hTra2- $\beta$  binding might be increased due to an interaction with other factors recognising the BPS. A candidate factor for this interaction might be U2 snRNP, which already loosely associates with the extended RNA targets under ATP-deficient conditions (Fig. III-15B). This proposal is supported by early findings demonstrating that SR protein-dependent exonic enhancers may directly stabilise U2 snRNP binding (544, 547) and also by the more recent finding that an hTra2- $\beta$ -dependent enhancer in SMN1 exon 7 recruits U2 snRNP without affecting U2AF<sup>65</sup> binding (398). Although binding of hTra2- $\beta$  was equally abrogated by two different mutations of the GAR ESE in RNA substrates lacking the BPS, analysis of BPS-containing RNA targets revealed that mutating the GAR ESE using the triple binding site mutation (ESE-) affects hTra2- $\beta$  binding stronger than mutation of only two SF2 binding sites (2xSF2-) (Fig. III-17B, cf. lanes 4-6 with lanes 1-3). This indicates that hTra2- $\beta$  might also bind to the sequence for which SRp40 binding was predicted. In RNA targets containing the BPS, hTra2- $\beta$  might bind to even suboptimal binding sites within the GAR ESE due to an interaction with U2 snRNP bound to the BPS. Therefore, a BPS-associated factor, like e.g. U2 snRNP, might compensate GAR ESE mutations resulting in low (ESE-) and intermediate hTra2- $\beta$  binding (2xSF2-).

It emerged as one general concept of enhancer-mediated splicing that binding of splicing activatory proteins enhance the recognition of flanking splice sites by stabilising spliceosomal components. However, assuming sole binding of hTra2- $\beta$  to the GAR ESE, increased binding of hTra2- $\beta$  in the presence of the BPS extends this concept by indicating that also factors recognising the BPS are *vice versa* able to stabilise a downstream located splicing regulatory protein. These results are consistent with the interpretation that a ternary complex forms on 3' ss A5 regulating general and/or specific 3' ss activation. Within a proposed ternary complex consisting of hTra2- $\beta$ , the RNA target and a yet non-confirmed factor associating with the BPS, three possible

interactions might contribute to binding of this regulatory complex on the RNA (Fig. IV-6). Decreased binding stability of hTra2- $\beta$  might be compensated by the interaction with other BPS-binding factors, e.g. U2 snRNP. However, since binding of U2 snRNP remains unchanged in the presence of the GAR ESE mutations, hTra2- $\beta$  appears not to stabilise U2 snRNP in early complex assembly. However, hTra2- $\beta$  binding might affect the ATP-dependent transition from E to A complex, during which U2 snRNP becomes stably associated with the BPS.

Besides other SR proteins also Tra2 has been reported to interact with U2AF<sup>35</sup> (655). Although binding of U2AF<sup>35</sup> did not correlate with the binding of hTra2- $\beta$  in the *in vitro* RNA binding assays representing early complexes assembly (cf. Fig. III-14B with Fig. III-17B), this result does not exclude the possibility that hTra2- $\beta$  interacts with U2AF<sup>35</sup> in subsequent steps of the spliceosome assembly. Since it was hypothesised in this thesis that in the presence of the GAR ESE the first catalytic step of the splicing reaction at 3' ss A5 might depend less on U2AF<sup>35</sup>, an interaction of hTra2- $\beta$  with U2AF<sup>35</sup> might improve recognition of the AG-dinucleotide during the second catalytic step and thereby promote overall splicing efficiency.

Besides hTra2- $\beta$ , SF2/ASF and SRp40, a number of additional protein isoforms were found to bind GAR ESE-mediated to the upstream region of exon 5 and the preceding intronic sequence (Fig. III-19A), which might contribute to general 3' ss activation by the GAR ESE. Mass spectrometry-based characterisation of proteins differentially bound to RNA targets carrying the parental GAR ESE or mutations thereof identified hnRNP H and hnRNP Q, two members of the hnRNP protein family. hnRNP proteins have been extensively analysed regarding their role in splicing regulation [reviewed in (396)]. In general, hnRNP proteins are considered to negatively regulate alternative splicing, however, also positive splicing regulation by hnRNPs was described (200, 241, 248). Since the proteins analysed by mass spectrometry (Fig. II-19A) were selected by comparison to mutant control sequences which resulted in reduced exon recognition in the minigene *in vivo*, hnRNP H and hnRNP Q, if they indeed contribute to splicing regulation of the internal HIV-1 exon 5, are likely to enhance usage of the 3' ss cluster upstream of exon 5.

The members of the hnRNP H subgroup consisting of hnRNP H, H', F and 2H9 predominantly bind to the consensus motifs GGGA, whereas hnRNP H and hnRNP H' additionally recognise the sequence motif GGGGCG (93, 521). Unexpectedly, these

consensus RNA-binding motifs of the hnRNP H family are not found within the 5'-half of exon 5 included in the RNA targets used for the binding experiments in this work. However, the GGGA motif is found at the 5' end of the RNA target in the intron sequence overlapping the AG-dinucleotide of A4b (Fig. III-14A). Although in the HIV-1 pre-mRNA this potential hnRNP H binding site is positioned in between both stretches of pyrimidines thought to interact with U2AF<sup>65</sup> from the protein binding studies, the "loop out" model might allow the GGGA motif to be bound by hnRNP H (Fig. IV-5).

Binding of hnRNP H has most often been reported to repress neighbouring splice sites. For example, it was found that hnRNP H antagonises exon recognition mediated by 9G8 thereby repressing inclusion of exon 2 of the *α-tropomyosin* mRNA (133). However, also an activating effect of hnRNP H on recognition of adjacent splice sites was observed, e.g. in the regulation of the N1 exon of c-src (113) and of the apoptotic regulator Bcl-x<sub>s</sub> (200). hnRNP H was also found to enhance or silence, respectively, splicing of the HIV-1 pre-mRNA via interactions with distinct splicing regulatory sequences. The activating function of hnRNP H is involved in the HIV-1 isolate-specific inclusion of exon 6D (94). In contrast, binding of hnRNP H to the exonic splicing silencer ESS2p in the second tat exon of the HIV-1 pre-mRNA reduces usage of A3 and was suggested to interfere with U2AF<sup>35</sup> binding (282). An hnRNP H-mediated decrease in U2AF<sup>35</sup> is consistent with the reduced levels of U2AF<sup>35</sup> detected at A5 in the presence of the GAR ESE (Fig. III-14B). Although highly speculative at this point, it might be imagined that hnRNP H supports a more equal activation of the 3' ss within the cluster by interfering with U2AF<sup>35</sup> binding at the AG-dinucleotide of 3' ss A5 thereby slowing down the second step of splicing catalysis.

The importance of hnRNP H binding to the HIV-1 RNA containing the 5'-half of exon 5 and an upstream intronic region observed *in vitro* was further substantiated by the finding that reducing cellular hnRNP H expression by RNAi decreased internal exon inclusion in the 2-intron minigene (Fig. III-28, ASR, cf. lane 2 with 6). However, the impact of hnRNP H on the regulation of exon 5 alternative splicing has to be further elucidated and will also to be extended to hnRNP Q.

Hints to the involvement of hnRNP Q in splicing regulation resulted from its identification in purified spliceosomes (441) and from its requirement for efficient *in vitro* splicing of adenovirus- and chicken  $\delta$ -crystallin-derived pre-mRNA substrates (427). Additionally, it was reported to bind the SMN protein (427) and to influence alternative splicing of SMN

exon 7 (109). Although hnRNP Q was reported to recognise U-rich and AU-rich sequences [(55), reviewed in (396)], binding of hnRNP Q to an *apoB* target RNA is in addition to poly(U) RNA also competed by poly(G) RNA (55). Addition of adenosine in poly(AU) RNAs did not abrogate the competing effect of poly(U) competitor RNAs. Since the effect of poly(GA) competitor RNA on the binding of hnRNP Q to the *apoB* pre-mRNA was not investigated by Davidson and colleagues (55), the GA-rich sequence of the GAR ESE in exon 5 might also serve as binding site for hnRNP Q.

#### **D.1.5 A functional network of exon-crossing interactions regulates splicing of the alternative exons 4c, 4a, 4b and 5.**

pre-mRNAs containing only one intron can be efficiently spliced *in vivo*. This could be reproduced *in vitro* using different single-intron pre-mRNA substrates (427, 596, 611, 659). Nevertheless, the presence of a downstream 5' ss was reported to promote splicing of an upstream located intron (68) in some cases by stabilising U2AF at the preceding 3' ss (260). These observations are consistent with the exon definition model, in which internal exons, i.e. exons flanked by a 3' ss and a 5' ss, are recognised via interactions across the exon (42). Experiments in this thesis demonstrated that recognition of the internal exons 4a and 5 is enhanced by strengthening U1 snRNP binding at the downstream flanking 5' ss D4 (Fig. III-10B and D). It was shown that U1 snRNP binding at D4 is required for efficient activation of the upstream 3' ss cluster (Fig. III-12). This result proposes that HIV-1 exon 5 is spanned by cross-exon interactions between U1 snRNP bound at D4 and the spliceosomal proteins recognising the 3' ss. Furthermore, strengthening the interaction of U1 snRNP with the 5' ss D4 was shown to increase the extent of 3' ss selectivity mediated by the GAR enhancer (Fig. III-10B and C). This additionally supports the model that U1 snRNP binding participates in 3' ss activation through cross-exon interactions, presumably by stabilising either spliceosomal components at the 3' ss or a specific 3' ss selecting SR protein configuration at the GAR enhancer.

Although D4 activates the upstream 3' ss cluster by U1 snRNP binding, usage of D4 itself likewise depends on activation by an enhancer element. Results in this thesis reveal that the GAR ESE, in addition to activating the upstream 3' ss cluster, is also essential for activation of the downstream 5' ss D4 (Fig. III-5C, 4.7-mRNA). Therefore, the GAR ESE is suggested to function as bidirectional enhancer activating the 3' ss

cluster as well as the 5' ss presumably by stabilising spliceosomal proteins at the respective splice site. A direct interaction of GAR ESE-bound proteins with spliceosomal components at the 3' ss is quite appealing due to the enhancer's spatial proximity to the upstream splice sites. This is also consistent with the GAR ESE-mediated stabilisation of U2AF<sup>65</sup> (Fig. III-14B). However, experiments in this thesis evidence that, even in close proximity to both, upstream and downstream, flanking splice sites, the GAR ESE is not sufficient to simultaneously activate both splice sites in the absence of the E42 sequence. Therefore, simultaneous activation might require an additional direct stabilisation of the U1 snRNP at D4 by the E42 sequence. Since mutational analyses of the GAR ESE revealed that two SR protein binding sites are needed for efficient internal exon recognition, it might be speculated that at least one SR protein binding site forms interactions to the 3' ss and one to the 5' ss thereby inducing a synergistic increase in exon recognition by promoting two distinct steps in the two splicing reactions resulting in exon inclusion (Fig. IV-1).

The model of splice site activation through oppositely directed interactions originating from the GAR ESE is supported by the additional requirement of the E42 sequence, located between the GAR enhancer and D4, for internal exon recognition in the subgenomic 2-intron minigene construct (Fig. III-13). As E42 failed to support U1 snRNP binding at the 5' ss in the absence of the GAR enhancer (91), it is assumed that it recruits an yet unidentified factor required for GAR ESE-dependent activation of D4. This factor might either act indirectly by facilitating cross-exon interactions between GAR ESE-bound SR proteins and U1 snRNP or establish an independent additional weak interaction with the U1 snRNP providing the stabilisation necessary to activate D4 (Fig. IV-1). The importance of this element for recognition of the short HIV-1 exon 5, flanked by long introns, supports the model that recognition of short exons is predominantly promoted by interactions of proteins across the exon (42, 188, 494).

Despite being essential for internal exon recognition in the HIV-1 pre-mRNA, the E42 sequence is noticeably less conserved between different HIV-1 strains than the GAR ESE (91). The only exception is a central heptamer motif (TCATCAA) that was found to be highly conserved. Nevertheless, a subsequent project based on results obtained in this thesis revealed that short sequences flanking the conserved heptamer contribute considerably more to internal exon recognition than the heptamer motif itself [A. Flößer, diploma thesis (183)]. In an initial screen the splicing regulatory proteins PTB, hnRNP Q1, and p54 were identified to specifically bind to the E42 sequence and

therefore are candidate proteins that might ensure internal exon recognition by E42 [A. Flößner, diploma thesis (183)]. However, since in particular PTB is commonly thought to exert a suppressive impact on exon recognition (279, 517, 540, 541, 554), the functional relevance of these proteins in the network regulating internal exon recognition in HIV-1 has to be further evaluated.

Analysing the splicing pattern of the 2-intron minigene, three essential elements were identified that are necessary to ensure the recognition of the internal exons 4c, 4a, 4b and 5: the GAR ESE, the E42 sequence and the 5' ss D4. Although each element is required for efficient exon recognition, none of them is able to ensure exon recognition independently of the other. Based on the experiments of this thesis, it is proposed that a network of interactions is spanned across exon 5, in which the most distant functional interaction is observed by the activation of the 3' ss cluster by the downstream D4 (Fig. IV-1). The regulatory network across HIV-1 exon 5 described in this thesis underscores the complexity of the functional interplay between *cis*-acting regulatory elements controlling alternative splice site selection [reviewed in (256, 564)].

Despite 5' ss D4 being unaltered, inactivating the GAR ESE by mutating three of the SR protein binding sites abrogated internal exon recognition in the 2-intron minigene (Fig. III-5C and D, lane 2). Likewise, substitution of the E42 sequence suppressed exon recognition (Fig. III-13B, lane 2). This is consistent with the interpretation that proteins binding to the GAR ESE and the E42 sequence are required to communicate the activation induced by U1 snRNP bound at D4 to the 3' ss upstream of exon 5. However, in light of the finding that most vertebrate exons are between 50 nt and 300 nt in length (42), exon 5 consisting of only 69 nt constitutes a short exon. Considering the proximity between flanking 3' ss and 5' ss, it is quite surprising that the presence of the GAR ESE and the E42 sequence might be needed to transfer the activating effect of U1 snRNP binding at D4 to the upstream 3' ss cluster. In this regard, it has been reported that in a distance of approx. 100 nt of the 5' ss U1 snRNP-mediated recruitment of U2 snRNP depends on the presence of SR proteins (68), even though U1 snRNP binding was observed in the absence of SR proteins. However, this earlier work did not differentiate whether SR proteins act via binding to an enhancer sequence or enhancer-independent to communicate the U1 snRNP-mediated increase in 3' ss recognition. For cross-intron communication between the 5' ss and the downstream 3' ss both, SR proteins and RS domain-containing splicing coactivators, were described to participate in an enhancer-independent fashion (60, 62, 193, 259, 498, 655). Similar to their function in general

splicing activation, it might thus be imagined that SR proteins also establish cross-exon interactions without the necessity for RNA binding. However, the requirement of the GAR ESE and the E42 sequence for exon recognition reveals that RNA-independent cross-exon interactions alone are insufficient to mediate efficient recognition of internal HIV-1 exon 5.

Based on the experimental evidence presented in this thesis showing that GAR ESE-mediated U2AF stabilisation was observed also in the absence of D4 *in vitro*, it might be argued that 3' ss A5 could in general be recognised also in the absence of downstream U1 snRNP binding. So why is downstream U1 snRNP binding required for exon recognition *in vivo*? Since particular interactions can still be observed albeit omitting *cis*- or *trans*-acting interaction partners emerging to be essential *in vivo*, it might be speculated that not all essential interactions are required at the same time. Taking into consideration the highly dynamic process of spliceosome assembly [for a recent review see (628)], interactions like e.g. between U2AF and GAR ESE-bound SR proteins might only be beneficial for early spliceosome assembly, whereas stabilising interactions with other spliceosomal components will become more important in later assembly steps. Complete loss of exon recognition in the absence of one of these essential factors, either the GAR ESE, the E42 sequence or downstream 5' ss D4 might be interpreted such that one of these interactions always has to be present to stabilise the respective spliceosomal complex and to ensure assembly progress. Therefore, it might be an attractive proposal that exon recognition *in vivo* depends on a number of different interactions, which subsequently substitute each other thereby providing a constant stabilisation of the spliceosomal assembly process.

Besides its function in the intronless 1.8 kb mRNA class, this work presents evidence that U1 snRNP binding at D4 is absolutely required also for activation of the 3' ss cluster in Rev-dependent 4 kb *vpu/env*-mRNAs, which are not spliced at D4. Based on earlier experiments using 1-intron minigenes, binding of U1 snRNP at D4 has been attributed to stabilise *vpu/env*-mRNAs, which remain unspliced at D4 in the presence of the regulatory protein Rev, (295, 367). However, in this thesis usage of 2-intron minigenes revealed that U1 snRNP-mediated stabilisation of the subgenomic HIV-1 pre-mRNA, if needed at all, can also be sufficiently performed by the most upstream 5' ss, D1. Since equal amounts of unspliced RNA were detected for several mutations of D4 (Fig. III-12B), it is assumed in this thesis that in the context of the additional upstream 5' ss D1, U1 snRNP binding at D4 is needed for 3' ss activation rather than for transcript stability

(295, 367). Nevertheless, this does not exclude the possibility that in the context of the genomic HIV-1 pre-mRNA additional inhibitory sequences might redirect a stabilising effect of D1 to the upstream half of the RNA. Therefore, D4 might be more important for stabilisation of the full-length RNA.

In the distinct isoforms of *vpu/env*-mRNAs the 3' ss activation ratio of A5 over A4a exceeded that of the 1.8 kb mRNAs, as has been described for viral infection (481). Since U1 snRNP is not displaced from D4 in the 4 kb mRNA class, its upstream directed cross-exon interactions are likely to persist longer, mediating the pronounced 3' ss selectivity of the GAR ESE. Thereby, the GAR ESE might contribute to the preferential generation of the *vpu/env*-mRNA isoforms efficiently expressing Vpu (16), which has been shown to support viral replication.

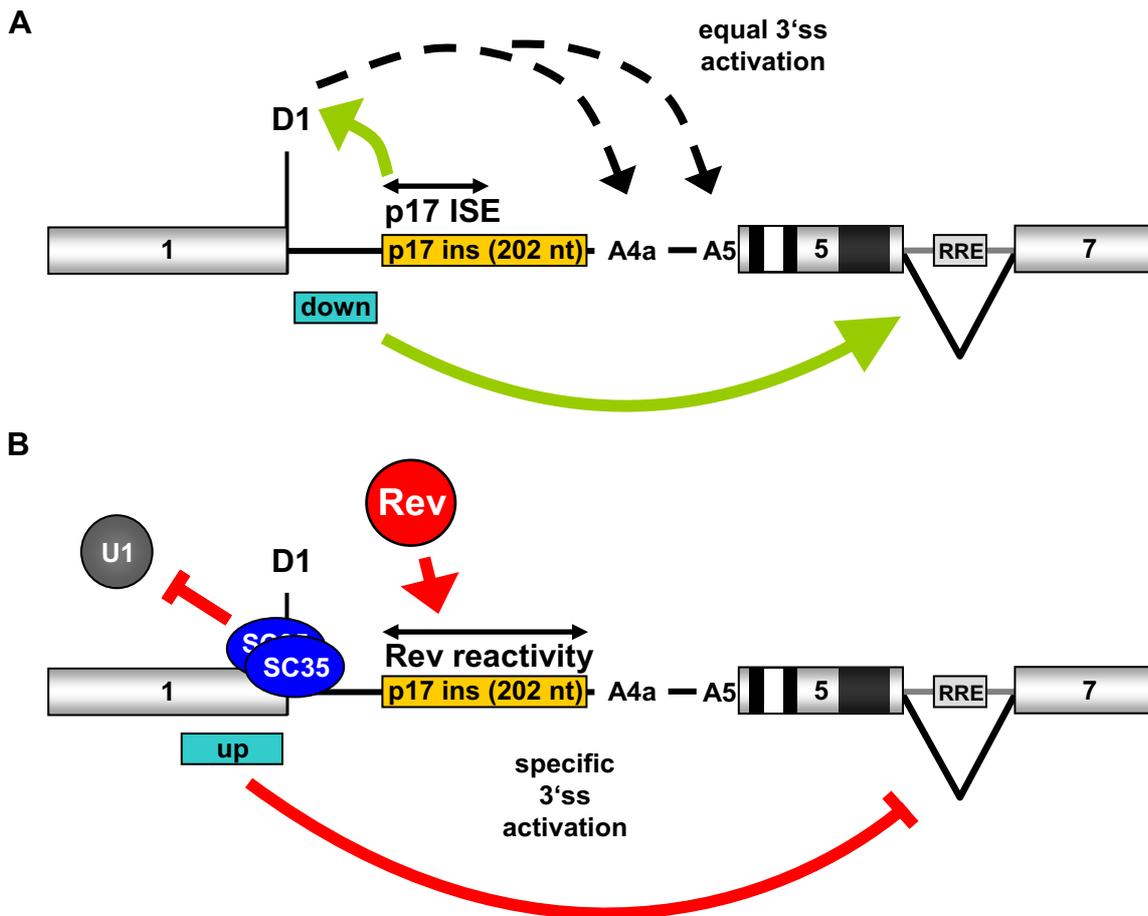
#### **D.1.6 Future perspectives: Internal exon recognition mediated by the GAR ESE**

This thesis provides evidence that the GAR ESE in exon 5 is a key regulatory element for splicing regulation of early and intermediately expressed HIV-1 mRNAs. The analysis of viral replication after mutating the GAR ESE might be a valuable approach to substantiate this finding. Additionally, the analysis of possibly emerging revertants might give further insight into the regulatory network across exon 5. The participation of the splicing regulatory proteins hTra2- $\beta$ , hnRNP H and hnRNP Q, identified for the first time in this work, requires further characterisation using e.g. protein overexpression or knockdown *in vivo* or *in vitro* systems. With regard to the RNA affinity chromatography used for protein binding studies *in vitro*, additional extensions of the intronic region and the analysis of later stages during spliceosomal assembly might provide further mechanistic details of the GAR ESE-mediated 3' ss activation.

---

## D.2 Expression of genomic HIV-1 RNA in the late phase of viral replication

Unspliced genomic HIV-1 RNA is essential for generation of replication-competent viruses. The expression of unspliced 9 kb genomic RNA and also intron-containing *vpu-env*-mRNA constituting the 4 kb mRNA class requires the presence of the viral regulatory protein Rev (148, 178, 179, 201, 239, 387, 389, 529). Nevertheless, delayed expression of genomic RNA compared to intron-containing *vpu-env*-mRNA (312) suggested that the expression of genomic RNA depends on an additional regulatory mechanism beside the Rev-RRE interaction. The second part of this thesis investigated whether *cis*-acting elements within 5' ss D1 and the neighbouring regions contribute to the regulated expression of HIV-1 genomic RNA. This thesis reveals that in the presence of the p17-inhibitory sequence (ins) within the 2-intron minigene the first 202 nt of this element become essential for Rev-dependent expression of completely unspliced, genomic RNA (Fig. IV-7).



**Fig. IV-7: Splicing at D1 is regulated by several exonic and intronic sequence elements.**

(A) Splicing activatory elements downstream of D1. The p17-ins element contains an ISE that promotes usage of D1 likely by an upstream directed interaction. In the presence of the p17-ins element A4a and A5 within the internal 3' ss cluster upstream of exon 5 were observed to be equally used. In addition, the sequence downstream of D1 (down) was found to activate splicing of the distal intron defined by D4 and A7 [RRE: Rev-responsive element].

(B) Splicing inhibitory elements up- and downstream of D1. The p17-ins element is required for Rev-dependent expression of completely unspliced RNA. Binding sites for the SR protein SC35 overlapping D1 decrease usage of the 5' ss likely by competing with U1 snRNP (U1) for binding. The sequence immediately upstream of D1 (up) was found to reduce splicing at the distal intron and to ensure specific activation of A5.

Additionally, it was found that these 202 nt also act as an intronic splicing enhancer (ISE) *in vivo*. Besides the major impact of the p17-ins element on splicing regulation of the viral pre-mRNA, a minor reduction of D1 usage was observed by SR protein binding sites overlapping D1. The decreased D1 usage caused by SR protein binding sites

suggests a direct competition of SR proteins and U1 snRNP for binding at D1. Furthermore, although not affecting the overall amount of RNA unspliced at D1, sequences flanking D1 modulated internal 3' ss selection and distal intron removal indicating that these sequences exert an effect on the downstream located splice sites maybe due to kinetic implications. Therefore, the p17-ins element, D1 flanking sequences and the D1 sequence itself were found to modulate the splicing pattern of a viral subgenomic model RNA both, positively as well as negatively. In contrast, an effect of the viral Gag protein on the splicing pattern of the 2-intron minigene could not be confirmed.

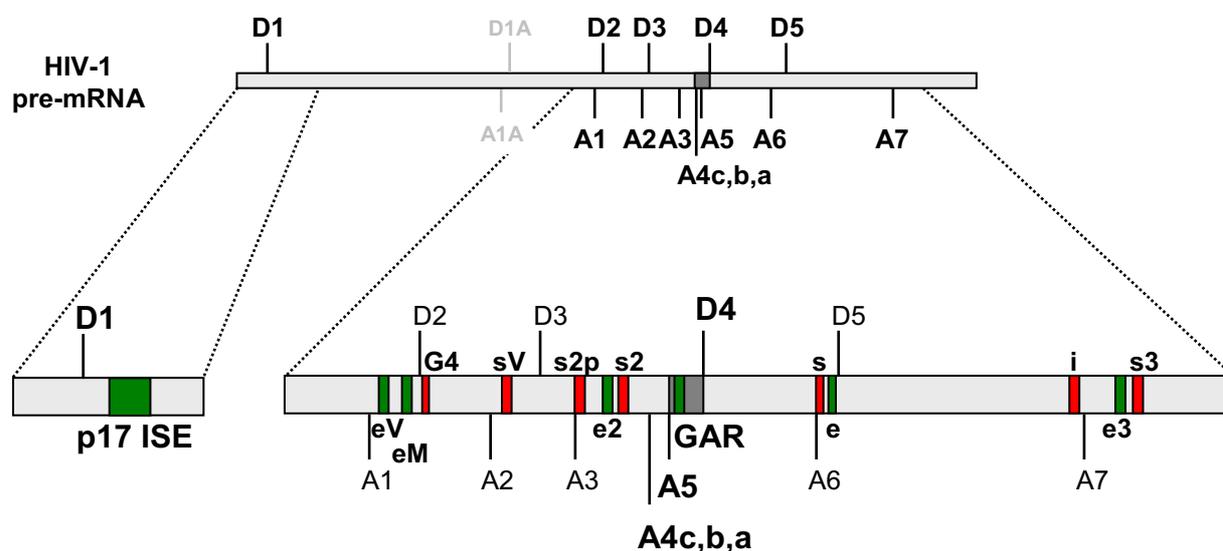
### **D.2.1 The p17-inhibitory element – a Janus regulator for the expression of genomic HIV-1 RNA**

The p17-ins element was initially identified by its inhibitory effect on the expression of the regulatory protein Tat, when placed in an artificial position downstream of 3' ss A7 in the HIV-1 *tat*-mRNA. Reduced protein expression was caused by a decrease in RNA stability as had been shown by measuring the RNA half-life in the presence of actinomycin D (530). Investigating the impact of the p17-ins element on HIV-1 splicing regulation in its authentic position downstream of D1, this thesis indicates that the p17 inhibitory sequence fulfils two splicing regulatory functions *in vivo* (i) by facilitating Rev-mediated expression of completely unspliced RNA and (ii) by enhancing splicing at D1. These splicing regulatory functions exerted by the p17-ins element might be missed in previous studies, because the characteristics of the HIV-1 inhibitory sequences were frequently investigated by Northern Blot analysis restricted to unspliced RNA (216, 524, 526, 530). Additionally, analysed reporter constructs often lacked D1 or a downstream 3' ss due to the application of subgenomic deletion mutants or cDNA constructs thereby displacing the p17-ins element out of its alternative intronic position (382, 418, 436, 524, 526, 530). Both functions of the p17-ins element promote the generation of different mRNA isoforms, i.e. RNA spliced at D1 and RNA unspliced at D1, and thus at first appear contradictory. However, the diverse abilities of the p17-ins element to accelerate RNA-degradation (530), to enhance splicing of the pre-mRNA (Fig. III-37D) and to mediate Rev-reactivity (Fig. III-34B) do not necessarily oppose each other, but instead might complement each other to timely coordinate the expression of unspliced and spliced viral mRNAs. In the early phase of viral gene expression degradation of mRNAs

that remain unspliced at D1 might be crucial to prevent the untimely expression of *gag*- and *gag/pol*-mRNAs. For expression of completely unspliced RNA in the late phase of viral gene expression, a cooperation of Rev with the p17-ins element might be required in addition to the Rev-RRE interaction in the downstream intron defined by D4 and A7. However, to maintain sufficient expression levels of the regulatory proteins Tat and Rev and the structural protein Env encoded by completely spliced and intron-containing mRNAs, respectively, the ISE might contribute essentially to D1 activation in the late phase of viral gene expression. Therefore, it might be proposed that multiple elements in the p17-ins sequence coordinate the expression of genomic HIV-1 RNA. This underscores the importance of the intronic sequence downstream of D1, which in its authentic position has previously mainly been implicated in decreased steady state levels of the *gag*- and *gag/pol*-mRNA, for regulated expression of genomic HIV-1 RNA (216, 524, 526).

A number of inhibitory sequences (ins) located in the *gag*-, *pol*- and *env*-coding regions of the HIV-1 pre-mRNA have been proposed to interfere with the expression of the respective genes (74, 121, 168, 239, 382, 438, 499, 524, 526, 530). In case of the p17-ins element, inhibition of gene expression has been interpreted to result from faster degradation of the RNA due to the high AU-content of the ins elements (530), whereas another investigation could not confirm a correlation between ins-mediated RNA degradation and the AU-content of the particular ins-element (S. Steck, PhD thesis (583)]. In contrast to those previous reports, experiments in this thesis indicated that mRNAs containing the first 202 nt of the p17-ins element (4.7-mRNA) appear as stable as a codon-optimised control sequence (Fig. III-37D, cf. lane 2 with 3). However, these experiments unexpectedly revealed that the first 202 nt of the p17-ins element activate splicing of a suboptimal D1 mutant in the 2-intron minigene (Fig. III-37D, lane 2; Fig. IV-7). First hints to the participation of the p17-ins element in HIV-1 splicing regulation came from recent *in vitro* splicing assays demonstrating that mutation of a guanosine repeat located 28 nt downstream of the Gag translational start codon abrogated splicing of an HIV-1 derived subgenomic RNA substrate (521). However, the exonic splicing silencer ESS2 located upstream of A4c had to be inactivated to allow splicing of the RNA substrate *in vitro*. Therefore, the importance of the guanosine repeat for splicing in the presence of additional splicing regulatory elements remained elusive. This thesis extends the *in vitro* results of Schaub and coworkers by demonstrating that also *in vivo* the first 202 nt of the p17-ins element, including the guanosine repeat,

activate D1 usage in the 2-intron minigene (Fig. III-37D, lane 2). Additionally, data presented in this work reveal that *in vivo* the ISE is also able to act in the presence of ESS2 and thus represents the first splicing regulatory element identified to act on D1 usage (Fig. IV-8).



**Fig. IV-8: The p17-ins element harbours an ISE activating D1 usage.**

So far, 13 *cis*-acting splicing regulatory sequences (lower panel) had been identified that either enhance or silence neighbouring splice sites of the HIV-1 pre-mRNA (upper panel). The newly identified p17 ISE is located downstream of D1 [e: ESE (646), e2: ESE2 (674), e3: ESE3 (579), eM: ESEM1/M2 (296), eV: ESEvif (172), G4: intronic guanosine-rich silencer (172), GAR: GAR ESE (91, 295), i: ISS (603), s: ESS (94, 646), s2: ESS2 (11, 555), s2p: ESS2p (282), s3: ESS3 (12, 556, 579), sV: ESSV (50, 380) (adapted from (588))].

Since D1 shares a high complementarity to the free 5'-end of the U1 snRNA, this 5' ss is assumed to constitute an efficiently used splice site. Consistently, D1 was used for splicing even in the absence of the ISE in the 2-intron minigene (Fig. III-36C). However, since a considerable level of 4.7-mRNA retaining the upstream intron is evident (Fig. III-1B), splicing of the upstream intron in the 2-intron minigene appears to function with only intermediate efficiency. Incomplete splicing of the upstream intron is somewhat surprising, because in the 2-intron minigene the intron is substantially shorter than in the viral pre-mRNA (Fig. III-1A). The identification of an intronic splicing

enhancer, which supports D1 activation, might explain the only intermediate splicing efficiency of the proximal intron in the 2-intron minigene in the absence of the p17-ins element. Nevertheless, it still might be argued that the p17 ISE is not essential for D1 activation. However, this element might become more important for D1 activation in splicing of the full-length HIV-1 pre-mRNA, in which the intron between D1 and the most proximal 3' splice active in splicing, A1, is considerably longer than in the 2-intron minigene (Fig. III-1A) and the RNA substrate used for the *in vitro* splicing assays (521). Furthermore, the p17 ISE might essentially contribute to the expression of completely spliced RNA in the late phase of viral gene expression thereby providing sufficient amounts of *tat*- and *rev*-mRNAs. Finally, during viral replication the p17 ISE might additionally ensure D1 usage after potential mutation of this splice site facilitated by the error rate of reverse transcription.

In the absence of the viral regulatory protein Rev, intron-containing HIV-1 mRNAs, like *vpu-env*-mRNA and genomic RNA, are intrinsically unstable as revealed by their short half-lives in the presence of actinomycin D (179, 386, 530). In light of the ISE within the p17-ins element, it might be questioned whether the reduced level of 4.7-mRNA, which remains unspliced at D1, in the presence of the p17-ins element is caused by the ISE-mediated increase in splicing at D1 or whether these intron-containing transcripts are degraded due to the destabilising effect of the p17-ins element. Although the HIV-1 ins elements were identified due to their inhibitory influence on gene expression, insertion of the p17-ins sequence into heterologous reporter RNAs induced variable effects on RNA stability. The first report identifying the p17-ins element showed that the half-life of a hybrid HIV-1 *tat/gag*-mRNA was strongly reduced by insertion of the p17-ins sequence (530). However, other investigations indicated that RNA stability was unaffected after insertion of the p17-ins element into heterologous RNAs derived from the *CAT* or  *$\beta$ -globin* genes (382, 418, 436). Since the ins elements share an unusual high AU-content and based on the observation that they destabilised originally selected reporter RNAs, the ins elements were attributed to the class of AU-rich elements (530). AU-rich elements have been shown to determine the extremely short half-life of mRNAs encoding regulatory proteins, e.g. involved in inflammation or signal transduction, that are only transiently expressed [reviewed in (308)]. Nevertheless, since the HIV-1 p17-ins element did not match the consensus motif AUUUA essential for the two main classes of AU-rich elements described, it can only be assigned into the additional third class lacking a defined consensus motif [reviewed in (34, 108)]. This thesis provides

experimental evidence that mRNAs containing the first 202 nt of the p17-ins or a control sequence are equally expressed (Fig. III-37D, lanes 2 and 3), although they largely differ in their AU-content (Fig. III-33). This finding implies that the AU-rich regions in the first 202 nt of the p17-ins element do not accelerate RNA degradation compared to the syn-ins control sequence in the 2-intron minigene. Since the most upstream 147 nt of the p17-ins element were reported to considerably reduce the steady state level of a hybrid HIV-1 *tat-/gag*-mRNA (530), the destabilising function of the p17-ins element appears to depend strongly on the sequence context of the respective reporter RNA. These controversial observations might indicate that the p17-ins element by itself is insufficient to reduce RNA expression, but depends on its interplay with other regulatory elements.

The p17-ins element is bound by all members of the hnRNP H family (93, 521). Binding of hnRNP H at an intronic position has been reported to activate a nearby 5' splice site in several pre-mRNAs (113, 200, 241, 248). In light of this finding, it was unexpected that only one out of five potential hnRNP H binding sites within the first 120 nt of the p17-ins element, the guanosine repeat 28 nt downstream of the Gag translational start codon, activated splicing of a subgenomic HIV-1 RNA substrate *in vitro* (521). Mutation of the guanosine repeat reduced binding of all proteins of the hnRNP H family at this site (521). Surprisingly, the remaining four intronic binding sites, one of them located even more closely to the 5' splice site, and another exonic potential binding site for the hnRNP H family did not contribute to splicing of the reporter RNA (521). This observation raised the question whether definite spatial requirements exist for hnRNP H either to bind to the RNA or to enhance splicing. In addition, the accessibility of the hnRNP H binding site might depend on distinct RNA secondary structures or alternatively might require an interaction with other regulatory elements to function as splicing enhancer. It might be conceivable that the *in vivo* function of the p17 ISE observed in this thesis is mediated by the hnRNP H binding site 28 nt downstream of the Gag translational start codon. However, the functional relevance of this hnRNP H binding site remains to be demonstrated within the 2-intron minigene used in the *in vivo* splicing assay.

The p17 ISE was identified in this thesis by activation of a suboptimal mutant 5' splice site, 5U, replacing the most proximal D1 in the 2-intron minigene (Fig. III-37D). The reduced complementarity of the mutant 5' splice site to the 5'-end of the U1 snRNA was found to be insufficient to facilitate splicing at D1 presumably due to a less efficient recruitment of U1 snRNP. The p17 ISE might activate the upstream located suboptimal 5' splice site 5U in the

2-intron minigene by stabilising U1 snRNP-binding at the 5' ss like it has been proposed for an hnRNP H/F-dependent splicing enhancer in the *DM20*-mRNA (630). However, *in vitro* analyses of the spliceosomal assembly on an HIV-1 derived subgenomic RNA substrate revealed that the E complex, which includes U1 snRNP bound to the 5' ss, was assembled with nearly equal efficiency after mutation of the guanosine repeat 28 nt downstream of the Gag translational start codon or depletion of the hnRNP H family (521). However, add-back experiments using recombinant proteins showed that the combination of hnRNP H, hnRNP F and hnRNP 2H9 supported only assembly of A and B complexes, but failed to promote C complex assembly (521) suggesting that another factor, either individually or in addition to the analysed hnRNPs, was required for splicing catalysis in those experiments. GRSF-1 might be a plausible candidate to allow C complex assembly, because the protein was depleted from the nuclear extracts, but afterwards not included in the mixture of hnRNPs that were added to the splicing reaction. Therefore, if the p17 ISE acts via the guanosine repeat 28 nt downstream of the Gag translational start codon, the enhancer presumably activates splicing by a mechanism distinct from U1 snRNP stabilisation.

Although binding of members of the hnRNP H family to the p17-ins element appeared to contribute to splicing regulation, it remains to be investigated whether their binding site at nt 28 (521) is the basis of the splicing enhancing effect of the p17-ins element. Further experiments will be aimed at localising the ISE within the first 202 nt of the p17-ins element in more detail. However, insertion of the ISE resulted in an at least equal usage of 3' ss A4a and A5 for internal exon inclusion compared to the parental 2-intron minigene lacking the p17-ins element (Fig. III-35B, lanes 2 and 8). Therefore, it might be deduced already from these experiments that the ISE likely localises within the first 63 nt of the p17-ins element, which is the shortest sequence of the p17-ins element that induced equal activation of A4a and A5 in the 2-intron minigene (Fig. III-35B, lanes 4 and 10). Additionally, this impact of splicing enhancer elements in the intron between D1 and the 3' ss cluster might hint to the presence of a second intronic splicing enhancer located between nt 202-302. Substitution of this region of the p17-ins element restored the unequal activation of the internal 3' ss cluster in the presence of the ISE located in the first 202 nt (Fig. III-34B, lanes 4 and 9) and thus points to less efficient spliceosomal assembly.

The expression of genomic RNA and *vpu/env*-mRNAs, which remain unspliced between D4 and A7, requires the interaction of Rev with the RRE element and an intrinsically

weak downstream 3' ss, A7 (103, 296). In line with these findings, not only *vpu/env*-mRNAs but also completely unspliced transcripts were expressed from the 2-intron minigene in the presence of Rev. However, after extending the upstream intron in the 2-intron minigene, the expression of completely unspliced RNA became dependent on the presence of the first 202 nt of the p17-ins element, whereas *vpu/env*-mRNA was expressed also from constructs carrying the mutated version of the p17-ins element, syn-ins (Fig. III-34B, cf. lane 8 with 12). Therefore, the results obtained in this thesis support the hypothesis that Rev-mediated retention of the upstream and the downstream intron, respectively, is regulated by mechanisms employing distinguishable sequence elements. Hints to the differential regulation of *vpu/env*-mRNA and genomic RNA already originated from previous investigations of the mRNA expression kinetics after viral infection of H9 cells demonstrating that completely unspliced RNA is expressed with a delay of approx. 12 h compared to *vpu/env*-mRNAs (312). However, the effect of the p17-ins element on HIV-1 gene expression was previously investigated using subgenomic 1-intron constructs or even in the absence of splice sites (216, 382, 524, 526, 530). The 2-intron minigene used in this work extends the results obtained for the p17-ins element in earlier studies, because it allows the differentiation of the splicing inhibitory function of the Rev-RRE interaction in the distal intron defined by D4 and A7 from the mechanisms necessary to generate completely unspliced RNA, which is additionally unspliced in the most upstream intron flanked by D1.

Expression of completely unspliced HIV-1 RNA was reported to depend on the viral regulatory protein Rev (178, 179, 239, 387, 389). Consistent with this early observation, expression of completely unspliced RNA from the parental 2-intron minigene lacking the p17-ins element was only detected in the presence of Rev (Fig. III-36B, lane 5). Nevertheless, in the parental 2-intron minigene this might mainly reflect the impact of Rev on splicing of the downstream intron defined by D4 and A7. Since the upstream intron flanking D1 was spliced with only intermediate efficiency in the parental 2-intron minigene lacking the p17-ins element, Rev might equally interfere with splicing of the downstream intron between D4 and A7 in pre-mRNAs, which are either spliced or still unspliced in the upstream intron, resulting in a mixture of completely unspliced and 1.5E-mRNA. A similar effect of the Rev-RRE interaction on an heterologous upstream intron has already been described (418, 436). However, after insertion of the first 396 bp of the p17-ins element into the upstream intron of the 2-intron minigene Rev-mediated expression of completely unspliced RNA additionally depended on the presence of the

first 202 nt of the p17-ins element (Fig. III-34B and Fig. III-35B). From this it might be concluded that the first 202 nt of the p17-ins element contain an additional Rev binding site. In the parental 2-intron minigene the ability of Rev to interfere with splicing via its interaction with the RRE might function only in an assumed maximum range of approx. 2 kb. Extending the upstream intron with the 396 bp of the p17-ins element might exceed this maximum range and thus renders the resulting intron too long to be controlled by the downstream Rev-RRE interaction. Therefore, the extension of the upstream intron might necessitate an additional Rev binding site for the generation of completely unspliced RNA. Alternatively, insertion of the first 202 nt of the p17-ins element might result in more efficient splicing of the upstream intron due to the presence of the ISE. Improved splicing of the upstream intron might hinder the retention of the upstream intron by Rev, because Rev-mediated expression of unspliced reporter RNAs has been described to depend on the presence of suboptimal splice sites (103, 296). Therefore, an additional Rev binding site within the upstream intron might be required to cope with the improved splicing of the upstream intron.

It has been shown that Rev fails to induce the expression of unspliced RNA from a heterologous  $\beta$ -globin 1-intron minigene with the HIV-1 RRE located within the intron (103). Further investigations using  $\beta$ -globin-derived 2-intron minigenes carrying the RRE in the terminal exon confirmed that observation (418, 436). However, the proximal 107 nt of the p17-ins element complemented Rev activity facilitating the expression of unspliced RNA in the cytoplasm (436). This effect was attributed to the binding of hnRNP A1 and an additional 50 kDa protein to this region of the p17-ins element. The synergistic activity of the RRE and the p17-ins sequence on the expression of unspliced RNA was suggested to be due to an inhibitory influence of hnRNP A1 bound to the ins element on the removal of the upstream situated  $\beta$ -globin intron (436). This might reduce the splicing efficiency of the  $\beta$ -globin 3' ss thus allowing Rev to interact with the RRE-containing mRNA. This interaction detaches the RNA from the splicing pathway finally leading to the appearance of unspliced RNA in the cytoplasm. However, in the heterologous  $\beta$ -globin minigene with the ins element as well as the RRE in artificial exonic positions downstream of the terminal heterologous 3' ss and also with both elements positioned side by side (418, 436), it might be questioned whether the observed effects represent the mechanisms underlying HIV-1 pre-mRNA processing.

Based on the synergistic action of the p17-ins element and the RRE described for a heterologous  $\beta$ -globin minigene, the identification of an additional Rev binding site upstream of D1 (198, 225) led to the proposal in this thesis that a synergistic interaction of Rev and the p17-ins element might also exist in the context of the authentic HIV-1 pre-mRNA. However, it was found that mutations in loop A of SL1 which were shown to reduce binding of Rev to this stem loop structure (198) did not affect the expression of completely unspliced RNA in the presence of the p17-ins element (Fig. III-36C). This result indicates that the Rev binding site within SL1 is presumably not involved in the regulation of splicing at D1 and thus does not cooperate with the p17-ins element to generate completely unspliced RNA. Therefore, another binding site for the Rev protein has to be assumed to be present in the proximity of the p17-ins element, if the inhibitory element complements Rev-reactivity by direct interactions. An alternative hypothesis would be that the synergy observed for the p17-ins element and the RRE for the generation of completely unspliced HIV-1 RNA emerges from the two individual functions conducted by each element, which are both required for expression of completely unspliced RNA under these experimental conditions.

In this thesis two newly identified splicing regulatory functions, enhancement of splicing at D1 and participation in Rev-mediated expression of completely unspliced RNA, were attributed to the first 202 nt of the p17-ins element. From that the question might arise whether a single regulatory element within this region accounts for both observed effects. However, since alternative usage of the internal 3' ss cluster indicated that the ISE might locate within the first 63 nt, whereas the first 102 nt of the p17-ins element are not sufficient to mediate Rev-reactivity for the expression of completely unspliced RNA in the 2-intron minigene, it is likely that two different regions of the p17-ins element mediate each particular function.

Unexpectedly, it was found that insertion of nt 203-402 of the codon-optimised form of the p17-ins element (syn-ins) abrogated the detectability of the 2-intron minigene-derived RT-PCR signals (Fig. III-34B). It might be hypothesised that CpG motifs created by codon-optimisation (151) reduce the expression of the syn-ins-containing mRNA, because CpG motifs have been implicated to interfere with gene expression [reviewed in (117, 286, 316)]. However, since mutation of the CpG motifs did not increase protein expression from a full-length syngag minigene (151) and codon-optimisation of the 396 nt p17-ins element analysed in this thesis only replaced a unique CpG motif within the 5'-region of the gag-ORF with another one, this possibility appears highly unlikely.

Furthermore, a region downstream of 3' ss A7, which is common to all HIV-1 mRNAs and also to those derived from the 2-intron minigene, could still be detected by RT-PCR (data not shown). This finding strongly argues against the degradation of syn-ins-containing mRNAs. However, the detection of minigene-derived RT-PCR products in the 3'-region of the pre-mRNA is consistent with the interpretation that codon-optimisation of the p17-ins element might alter the splicing pattern of the 2-intron minigene resulting in the removal of the binding site for the 5'-primer which is used for the amplification of all PCR products derived from alternatively spliced reporter mRNAs. It seems conceivable that due to the numerous mutations throughout the p17-ins element alternative splice sites have been generated in the syn-ins sequence or that RNA binding sites for splicing regulatory proteins had been destroyed or generated.

Besides a potential impact of the 396 nt syn-ins sequence on the activation of a cryptic splice site in the 2-intron minigene, the syn-ins sequence was also described to use an RNA export pathway distinct from that employed by the Rev-RRE system (216). The viral regulatory protein Rev, which is essential for the expression of intron-containing HIV-1 mRNA in the cytoplasm, was found to contain a nucleolar localisation signal (NoLS) that is necessary for efficient expression of a Rev-dependent reporter gene (122). Additionally, it has been deduced from nucleolar ribozyme-mediated RNA targeting that viral RNA at least transiently localises in the nucleolus (416). Although these observations could not be reproduced using subgenomic minigenes (684), it might be proposed that the Rev-RNA complex traffics via the nucleolus. Subsequent translocation of intron-containing HIV-1 mRNAs through the NPC by Rev had been reported to employ the nuclear export receptor Crm1 and is inhibited by leptomycin B treatment (186, 194, 443, 654). In contrast to HIV-1 *gag*-mRNA, the codon-optimised *syngag*-mRNA, which carries a number of mutations throughout the *gag*-encoding region, was found to be exported from the nucleus via a leptomycin B-independent pathway (216). Since the syn-ins sequence appears to induce an export pathway distinct from that applied by the p17-ins element, the RNA might - even in the presence of Rev - bypass nucleolar trafficking. By omitting nucleolar localisation an essential step for mediated expression of unspliced RNA might thus be circumvented. Therefore, it might be questioned whether the export pathway contributes to the splicing regulation of the subgenomic HIV-1 pre-mRNA.

### D.2.2 D1 usage is suppressed by SR proteins binding to the 5' ss.

The best characterised examples of alternative splice site usage in eukaryotes are regulated by members of two protein families, the hnRNP family and the SR proteins [reviewed in (357, 363, 396)]. Investigating the sequence requirements for efficient activation of the major 5' ss, D1, analysis of the D1 sequence itself revealed a striking number of binding sites for the SR proteins SC35, SRp55 and SRp40 covering the 5' ss (Fig. III-22). Since SR proteins have mainly been shown to activate neighbouring splice sites (357, 363), it was quite unexpected that mutating the SC35 binding sites overlapping D1 increased splicing at this 5' ss (Fig. III-23A and B). From these results it is suggested that SR protein binding sites overlapping the 5' ss prevent efficient activation of D1. It is hypothesised in this work that suppression of D1 usage is caused by competition of SR proteins bound at D1 with the U1 snRNP, which recognises the 5' ss in the early spliceosomal E complex.

Suppression of 5' ss usage by restricting the accessibility of U1 snRNP to the 5' ss had been described for a member of the mostly negatively acting hnRNP family, hnRNP H, after disease-causing 5' ss mutations in *NF-1* exon 3 and *TSH $\beta$*  exon 2 (81). However, the influence of hnRNP H binding on U1 snRNP binding was difficult to evaluate in the previous work, because improved binding of hnRNP H might reduce U1 snRNP binding, but on the other hand the reduced complementarity of the U1 snRNA to the 5' ss might *vice versa* facilitate increased binding of hnRNP H within the 5' ss. In this thesis, SR protein binding sites in the proximity of D1 were mutated outside of the 11 nt sequence of D1, thus leaving its intrinsic strength unaltered (Fig. III-22C). Therefore, increased levels of spliced mRNAs expressed from the 2-intron construct after mutation of the SC35 binding sites were not due to an increased intrinsic strength of D1, but were caused by the mutations of the neighbouring nucleotides (Fig. III-23B).

Contrasting the splicing inhibitory effect of the SC35 binding sites, the SRp40 binding site appears to slightly activate splicing at D1 in the 2-intron minigene (Fig. III-23). Since binding sites for SC35 and SRp40 oppositely influence D1 usage, D1 suppression might be not a general consequence of SR proteins bound at D1. However, whereas binding of SC35 to both most upstream binding sites was confirmed (Fig. III-25), binding of SRp40 was not investigated and thus the predicted binding site might not be occupied by SRp40. Alternatively, mutations introduced into the proximity of the D1 sequence might be assumed to alter the stability of the stem loop that is generated by the 5' ss

and surrounding nucleotides (Fig. III-22B) (10, 673). The stability of the stem loop has been reported to affect D1 usage (2). However, analysing the mutations introduced into the D1 proximity regarding the stability of the resulting stem loops demonstrated that the alterations in the splicing pattern of the 2-intron minigene did not correlate with the stem loop stability in the experimental system used in this thesis (Fig. III23A, B and C).

It might be envisioned that a competition of SC35 and U1 snRNP for binding to D1 assists the Rev-dependent generation of completely unspliced RNA in the late phase of viral gene expression. This hypothesis implies that during the course of viral replication SC35 levels increase or, alternatively, U1 snRNP levels decrease in target cells to achieve a shift in the binding ratio of both factors. Although not assayed during single rounds of infection, recent data, however, suggested that in general hnRNPs are upregulated, whereas the level of SC35 appears to be downregulated during long-term infected cultures of monocyte-derived macrophages (MDM) (161). Nevertheless, a high variability in the expression of splicing regulatory proteins was observed between different PBMC donors in these experiments. This finding contrasts an earlier report demonstrating that SC35 levels of the H9 T cell line increase three days after infection (383). These oppositely directed changes in the expression of SC35 might be interpreted that either SC35 expression after HIV-1 infection of the cell is biphasically modulated with an early increase and a subsequent decrease or that increased SC35 expression is restricted to infected T cells. Anyway, these data support the possibility that at least in T cells increasing SC35 levels may contribute to the expression of completely unspliced RNA in the late phase of viral gene expression.

It was shown that the 5'-UTR of the HIV-1 pre-mRNA containing D1 is essential for generation of *gag*-mRNAs, unspliced at D1 (216). This is consistent with the concept that U1 snRNP binding at a 5' ss is able to stabilise mRNAs which are vulnerable to degradation (295, 367, 373). This stabilisation might be functionally associated with the recruitment of transcriptional initiation factors by promoter-proximal 5' ss, which has been uncovered in the last years (63, 143, 196). The assumed interference of SR protein binding sites with U1 snRNA binding would diminish U1 snRNP-mediated stabilisation of the viral pre-mRNA. Therefore, it might be questioned whether stabilisation of the viral pre-mRNA is needed at D1 at all or whether stabilisation can also be performed by SR proteins bound at D1. Since SR proteins have already been shown to generate splicing enhancer complexes across exons, it might be conceivable that SC35 bound at D1 might also form a complex with the 5' cap structure of the pre-

mRNA. This complex might render the RNA-protein complex less susceptible for degradation.

A previous study analysing the binding properties of the Gag polyprotein precursor and its nucleocapsid (NC) subunit to the viral packaging signal encoded downstream of D1 revealed that the presence of each, the Gag precursor protein and the NC subunit, increased the expression of a reporter gene localised in an intronic position [C.K. Damgaard, PhD thesis (142)]. This finding suggested a role for the Gag-RNA interaction in viral mRNA processing. Enhanced stability of the reporter RNA and reduced splicing efficiency were considered as possible causes for increased protein expression. Although it appeared an attractive hypothesis that Gag increases the level of unspliced RNA in an autoregulatory loop during late viral replication, this hypothesis could not be confirmed in this thesis. Data presented here revealed that the presence of Gag did not affect the splicing pattern of the 2-intron minigene containing the viral packaging signal at its authentic position downstream of D1 (Fig. III-32). However, since in this thesis HeLa-T4<sup>+</sup> and HeLa-T4R5 cells were used to assess the influence of Gag on the HIV-1 pre-mRNA splicing pattern, it cannot completely ruled out that a Gag-dependent increase in the expression of the intronic reporter gene in the analysis performed by Damgaard might represent a distinctive feature of HEK 293 cells. Nevertheless, since the splicing pattern of HIV-1 mRNAs in PBMCs was shown to be reproduced in HeLa-T4<sup>+</sup> cells (481), results of this thesis suggest that the effect of Gag in target cells of HIV-1 is likely restricted to functions other than splicing, e.g. RNA stabilisation or enhanced translation.

### **D.2.3 Sequences neighbouring D1 affect far distant splice site usage – implications for splicing kinetics.**

As outlined above, SR protein binding sites overlapping D1 and the p17-ins element were found to regulate D1 usage. However, since substitution of the sequences flanking D1 did not alter the splicing efficiency at D1 in the 2-intron minigene (Fig. III-30D), the hypothesis that additional splicing regulatory elements reside in the immediate proximity of D1 could not be confirmed. Nevertheless, these experiments unexpectedly revealed that substitution of the sequences neighbouring D1 - although leaving the amount of RNA unspliced at D1 unchanged - altered the respective amount of 4.7-mRNA and completely unspliced RNA (Fig. III-30C). Whereas the amount of 4.7-mRNA, lacking the

distal intron defined by D4 and A7, increased after substitution of the upstream sequence with the control sequence, the level of unspliced RNA decreased correspondingly. Substitution of the sequence downstream of D1 with the control sequence inversely affected the expression of both mRNA isoforms. The changes in the splicing pattern of the 2-intron minigene were confirmed by inserting a different control sequence downstream of D1 (Fig. III-30B and C, cf. lane 3 with 4). Because of the distance between the sequences flanking D1 and the splice sites of the distal intron and the position of the internal 3' ss cluster in-between D1 and the distal intron, it appears unlikely that the sequence substitutions in the proximity of D1 directly affect spliceosome assembly at the distal intron. However, since in the last years a strong coupling of the different gene expression pathways, i.e. transcription, 5'-capping, splicing, 3'-processing, transcript release and RNA export became evident [reviewed in (41, 391, 442, 461, 480)], it might be speculated that the sequences neighbouring D1 indirectly affect splicing efficiency of the distal intron by influencing one of these steps, e.g. transcription rate or transcript release from active sites of transcription.

The characterisation of cotranscriptional RNA processing uncovered an additional level of splicing regulation [(49, 454), reviewed in (207, 322)]. Since it has been demonstrated that the rate of transcriptional elongation can dictate splice site selection *in vivo* (149, 268, 292, 446), it might be assumed that the sequences in the proximity of D1 in the HIV-1-derived reporter RNA influence splicing of the distal intron by affecting the speed of transcription by RNA polymerase II. Changes in the transcriptional elongation rate would have been expected to affect also splicing of the proximal intron between D1 and the internal 3' ss. However, splicing at D1 remained unaltered after substitution of the sequences neighbouring D1. Therefore, it appears unlikely that the sequences flanking D1 affect splicing of the distal intron by modulation of transcriptional elongation. This conclusion is substantiated by the finding that splicing of a subgenomic HIV-1 *env*-transcript could not be observed at active sites of transcription in a fluorescence-based microscopy approach (684). This led to the proposal that splicing of the HIV-1 pre-mRNA – in contrast to a number of cellular pre-mRNA introns – does not necessarily occur cotranscriptionally probably due to its considerable number of functionally weak splice sites (684).

Based on the observation that the frequency of D1 usage remains unaffected after substitution of its neighbouring sequences, it is hypothesised that these sequences are involved in a processing step subsequent to transcriptional elongation. It might be

speculated that D1-proximal sequences modulate splicing of the distal intron defined by D4 and A7 by altering the release of the nascent transcript from the DNA and its subsequent movement from the transcriptionally active perichromatin to the nuclear pore complex (NPC). Since the transcriptionally active sites are assumed to be supplied with splicing factors from nuclear speckles (340), they might contain a higher concentration of splicing factors. Assuming that splicing of the 2-intron minigene as well does not occur cotranscriptionally, the reporter pre-mRNA would be released from the transgene still containing at least one intron. Therefore, releasing the pre-mRNA from active sites of transcription might reduce the splicing efficiency of the remaining introns due to a starvation of splicing factors on the transition of the intron-containing pre-mRNA to the NPC. Accelerating the release of the pre-mRNA from the active sites of transcription would likely decrease the splicing efficiency of introns which are spliced with a slow kinetic and, in addition, might stronger affect introns which are located in the 3' region of the gene. This might explain why splicing of the distal intron defined by D4 and A7 responds to substitutions of the sequences neighbouring D1, whereas the proximal intron defined by D1 and the internal 3' ss cluster remains unaffected by these substitutions. Therefore, an impact of the sequences in the proximity of D1 on gene expression pathways beyond transcriptional elongation might explain the differences in the expression of 4.7-mRNA and completely unspliced RNA from the 2-intron minigene.

Cotranscriptional splicing implicates that introns located in the 5' region of the pre-mRNA are spliced prior to the transcription of introns in the 3' region. From this it might be deduced that most pre-mRNAs containing multiple introns are overall spliced in a 5'-to 3'-order, although in a microenvironment of at least two neighbouring introns the downstream intron might be spliced prior to the upstream one (647). Even though it was proposed that splicing of the HIV-1 pre-mRNA does not occur cotranscriptionally (684), more recent work postulated that, nevertheless, all introns are removed in 5'-to 3'-order (64) [M. Otte, PhD thesis (457)]. These ideas do not necessarily contradict each other, but together imply that the order of intron removal might be imprinted on the pre-mRNA during transcription. From the ordered intron removal of the HIV-1 pre-mRNA it was concluded that inhibition of splicing at D1 is sufficient to restrict usage of all downstream splice sites thereby generating completely unspliced RNA (64). This implicated that splicing of the distal intron between D4 and A7 depends on splicing of the upstream intron. However, data presented in this thesis demonstrate that usage of the downstream splice sites D4 and A7 is not prevented by the inhibition of splicing at D1

(Fig. III-37C, lane 4) and therefore, usage of the 3' ss cluster upstream of exon 5 is no prerequisite for splicing of the downstream intron in the 2-intron minigene. Thus, generation of completely unspliced RNA besides the inhibition of D1 usage also depends on the additional Rev-RRE interaction in the 3'-half of the viral pre-mRNA.

In addition to its influence on splicing of the distal intron, the sequence upstream of D1 was demonstrated to contribute to the unbalanced internal 3' ss activation in the 2-intron minigene as evidenced by the almost equal usage of the internal 3' ss cluster in the presence of the control sequence upstream of D1 (Fig. III-30E). Because of the exon/intron architecture with D1 located between the upstream sequence and the downstream 3' ss, it appears unlikely that proteins binding to the sequence upstream of D1 directly affect spliceosomal assembly at the respective 3' ss. Although examples have been described, in which an RNA duplex formation results in looping out a complete series of introns and exons (221), bioinformatic analysis of the HIV-1 reporter RNA used in this thesis did not support the formation of an RNA duplex between the sequence upstream of D1 and the sequences neighbouring the internal 3' ss cluster (data not shown) [Mfold (<http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi>) (699)]. Since insertion of the ISE within the p17-ins element downstream of D1 indicated that the efficiency of the upstream 5' ss may influence internal 3' ss selection, it might be imagined that the sequence upstream of D1 influences the kinetic of the splicing reaction. Therefore, these findings might support a strong impact of splicing kinetics on internal 3' ss competition in the HIV-1 pre-mRNA.

Improving splicing kinetics of the proximal intron is conceivable to be achieved by promoting recognition of D1 itself. However, the experiments revealed that overall splicing at D1 is not affected by substitution of the upstream sequence arguing against a direct impact of the sequence upstream of D1 on recognition of this 5' ss (Fig. III-30D, lane 2). Alternatively, the upstream sequence might slow down the kinetic of the splicing reaction by interfering with the transition from the exon-defined early spliceosomal complex at exon 1 to subsequent intron-defined spliceosomal complexes (522), in which splice site of both exons are paired in a common spliceosomal complex. Restructuring spliceosomal components from an exon-defined to an intron-defined complex has recently been shown to regulate the neuron-specific inclusion of the N1 exon into the *c-src*-mRNA (541). However, in the *c-src*-mRNA an intronic element was shown to bind an hnRNP family member, PTB, which interferes with the formation of an intron-defined spliceosomal complex. If an impaired switch from exon- to intron-definition is likewise

involved in the regulation of splicing kinetics at D1, the sequence upstream of the 5' ss due to its exonic position likely acts by recruiting an alternative splicing regulatory factor. Analyses of predicted binding sites for SR proteins SF2/ASF and SRp55 within the sequence upstream of D1 revealed that they are not involved in the unbalanced 3' ss activation and the decreased splicing of the distal intron mediated by the 24 nt upstream of D1 (Fig. III-24C). Likewise, abrogating hnRNP H binding within the sequence upstream of D1 did not resemble the effect of the complete substitution of the 24 nt upstream of D1 (Fig. III-27). The recent finding that the hnRNP H binding site upstream of D1 is not essential for *in vitro* splicing at D1 (521) was extended in this thesis showing that also *in vivo* another mutation, which was confirmed to abrogate hnRNP H binding, did not affect splicing at D1 (Fig. III-27). In summary, since the impact of the sequence upstream of D1 on splicing of the reporter RNA could neither be attributed to binding sites for the SR proteins SF2/ASF and SRp55 nor to hnRNP H binding, this work did not identify *cis*-acting factors mediating unbalanced 3' ss activation and reduced splicing of the distal intron. However, since many proteins bound to an RNA target containing the 24 nt upstream of D1 (Fig. III-29, lane 1), it cannot be ruled out that binding sites for other splicing regulatory proteins exist within the sequence upstream of D1, which possibly affect internal 3' ss selectivity and splicing efficiency of the distal intron.

The hypothesis that the splicing kinetics of the upstream intron is influenced by the sequence upstream of D1 was nevertheless further supported by the finding that a second substitution of this sequence encoding two RNA hairpin structures appeared to promote 3' ss selectivity and induced partial skipping of the internal exon (Fig. III-26C, lane 4). From this result it might be deduced that increasing the secondary structure immediately upstream of D1 might reduce the efficiency of D1 usage and favour the usage of the internal 3' ss with the higher overall strength.

#### **D.2.4 Future perspectives: Regulation of D1 usage**

The major regulatory effect on the expression of genomic unspliced HIV-1 RNA identified in this work was exerted by the p17-ins element. The analysis of 2-intron minigenes carrying additional ins/syn-ins hybrid sequences will likely identify the particular sequences within the p17-ins element required for its splicing enhancer

function and for Rev-mediated expression of completely unspliced RNA. Mutation of the hnRNP H binding site at nt 28 of the Gag-ORF (521) in the 2-intron minigene might reveal whether this site essentially contributes to the function of the p17 ISE. In addition, analysing the splicing pattern of the 2-intron minigene harbouring the p17-ins element in cells expressing decreased levels of hnRNP H might evaluate whether hnRNP H act as *trans*-acting factor regulating splicing of the HIV-1 pre-mRNA at D1. Finally, analysis of cryptic splice site activation in the presence of the 396 nt syn-ins sequence in the 2-intron minigene by 5' RACE (Rapid Amplification of cDNA Ends) might uncover splicing regulatory mechanisms underlying splicing of the HIV-1 pre-mRNA that remained unnoticed so far.

## E. Conclusions

The viral pre-mRNA employs a number of *cis*-acting regulatory sequences to ensure the timely coordinated expression of viral genes. This thesis reveals that almost the complete sequence of exon 5 is essential for inclusion of the alternative exons 4c, 4a, 4b and 5 into early mRNAs encoding the regulatory proteins Rev and Nef by generating a functional network of cross-exon interactions. The GAR ESE within exon 5 was found to constitute a complex substantial splicing regulatory element throughout HIV-1 gene expression. Although less conserved between different HIV-1 isolates than the GAR ESE, the E42 sequence located in the 3' half of exon 5 is also crucial for internal exon inclusion. However, the low level of sequence conservation might indicate that the mechanism that underlies the contribution of the E42 sequence to internal exon inclusion is more variable than that of the GAR ESE and might be fulfilled by different proteins in various HIV-1 isolates or by a factor requiring a considerably less stringent binding sequence. Since GAR ESE-mediated activation of the downstream located 5' splice site and the upstream located 3' splice site cluster is essential to generate almost all spliced mRNA isoforms during viral gene expression, the functional network emanating from the GAR ESE, the E42 element, and the flanking splice sites substantially contributes to the regulation of viral replication. The GAR ESE has recently been successfully targeted to inhibit HIV-1 multiplication (21), and further insights regarding the mechanism of GAR ESE-mediated splice site activation and selectivity might create new strategies to specifically inhibit generation of essential regulatory HIV-1 mRNAs.

In the second part of this thesis the sequences flanking D1 and the p17-ins element in the downstream intron were found to regulate splicing of the HIV-1 pre-mRNA. Slight effects were observed for the sequences directly neighbouring D1 on splicing of the downstream intron defined by D4 and A7 indicating that sequences upstream in the viral pre-mRNA may also affect distant downstream splicing events. However, the most substantial effect on the expression of genomic RNA appeared to be exerted by the p17-ins element within the Gag-ORF, which is essential for the expression of completely unspliced RNA but not for *vpu/env*-mRNAs. This might suggest that Rev applies different mechanisms to inhibit splicing of the downstream and the upstream intron of the HIV-1 pre-mRNA resulting in the expression of *vpu/env*-mRNAs and completely unspliced RNA. Understanding these virus specific mechanisms interfering with the cellular splicing machinery might provide additional options to restrict HIV-1 replication.

## F. References

1. 't Wout,A.B., Lehrman,G.K., Mikheeva,S.A. *et al.* (2003) Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4(+)-T-cell lines. *J.Virol.* **77**, 1392-1402.
2. **Abbink,T.E. and Berkhout,B.** (2008) RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J.Virol.* **82**, 3090-3098.
3. **Abovich,N., Liao,X.C. and Rosbash,M.** (1994) The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes Dev.* **8**, 843-854.
4. **Abovich,N. and Rosbash,M.** (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403-412.
5. **Adachi,A., Gendelman,H.E., Koenig,S. et al.** (1986) Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J.Virol.* **59**, 284-291.
6. **Adams,M.D., Celniker,S.E., Holt,R.A. et al.** (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-2195.
7. **Afonina,E., Neumann,M. and Pavlakis,G.N.** (1997) Preferential binding of poly(A)-binding protein 1 to an inhibitory RNA element in the human immunodeficiency virus type 1 gag mRNA. *J.Biol.Chem.* **272**, 2307-2311.
8. **Aldovini,A. and Young,R.A.** (1990) Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J.Virol.* **64**, 1920-1926.
9. **Alkhatib,G., Combadiere,C., Broder,C.C. et al.** (1996) CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* **272**, 1955-1958.
10. **Amarasinghe,G.K., De Guzman,R.N., Turner,R.B. et al.** (2000) NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform. *J.Mol.Biol.* **299**, 145-156.
11. **Amendt,B.A., Hesslein,D., Chang,L.J. et al.** (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. *Mol.Cell Biol.* **14**, 3960-3970.
12. **Amendt,B.A., Si,Z.H. and Stoltzfus,C.M.** (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. *Mol.Cell Biol.* **15**, 4606-4615.
13. **Amir-Ahmady,B., Boutz,P.L., Markovtsov,V. et al.** (2005) Exon repression by polypyrimidine tract binding protein. *RNA.* **11**, 699-716.
14. **Ammosova,T., Jerebtsova,M., Beullens,M. et al.** (2005) Nuclear targeting of protein phosphatase-1 by HIV-1 Tat protein. *J.Biol.Chem.* **280**, 36364-36371.
15. **Amrein,H., Hedley,M.L. and Maniatis,T.** (1994) The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. *Cell* **76**, 735-746.
16. **Anderson,J.L., Johnson,A.T., Howard,J.L. et al.** (2007) Both Linear and Discontinuous Ribosome Scanning Are Used for Translation Initiation from Bicistronic Human Immunodeficiency Virus Type 1 env mRNAs. *J.Virol.* **81**, 4664-4676.

17. **Anderson,K. and Moore,M.J.** (1997) Bimolecular exon ligation by the human spliceosome. *Science* **276**, 1712-1716.
18. **Anderson,K. and Moore,M.J.** (2000) Bimolecular exon ligation by the human spliceosome bypasses early 3' splice site AG recognition and requires NTP hydrolysis. *RNA*. **6**, 16-25.
19. **Ansorge,W., Sproat,B.S., Stegemann,J. et al.** (1986) A non-radioactive automated method for DNA sequence determination. *J.Biochem.Biophys.Methods* **13**, 315-323.
20. **Askjaer,P., Bachi,A., Wilm,M. et al.** (1999) RanGTP-regulated interactions of CRM1 with nucleoporins and a shuttling DEAD-box helicase. *Mol.Cell Biol.* **19**, 6276-6285.
21. **Asparuhova,M.B., Marti,G., Liu,S. et al.** (2007) Inhibition of HIV-1 multiplication by a modified U7 snRNA inducing Tat and Rev exon skipping. *J.Gene Med.* **9**, 323-334.
22. **Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A., Struhl,K. and eds.** (1991) *Current protocols in molecular biology*. Greene Publishing Associates and Wiley-Interscience, New York.
23. **Auweter,S.D. and Allain,F.H.** (2008) Structure-function relationships of the polypyrimidine tract binding protein. *Cell Mol.Life Sci.* **65**, 516-527.
24. **Auweter,S.D., Oberstrass,F.C. and Allain,F.H.** (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* **34**, 4943-4959.
25. **Ballut,L., Marchadier,B., Baguet,A. et al.** (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat.Struct.Mol.Biol.* **12**, 861-869.
26. **Banerjea,A., Li,M.J., Bauer,G. et al.** (2003) Inhibition of HIV-1 by lentiviral vector-transduced siRNAs in T lymphocytes differentiated in SCID-hu mice and CD34+ progenitor cell-derived macrophages. *Mol.Ther.* **8**, 62-71.
27. **Banerjee,H., Rahn,A., Davis,W. et al.** (2003) Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. *RNA*. **9**, 88-99.
28. **Barabino,S.M., Blencowe,B.J., Ryder,U. et al.** (1990) Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. *Cell* **63**, 293-302.
29. **Baralle,D., Lucassen,A. and Buratti,E.** (2009) Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep.* **10**, 810-816.
30. **Baralle,M., Skoko,N., Knezevich,A. et al.** (2006) NF1 mRNA biogenesis: effect of the genomic milieu in splicing regulation of the NF1 exon 37 region. *FEBS Lett.* **580**, 4449-4456.
31. **Baraniak,A.P., Lasda,E.L., Wagner,E.J. et al.** (2003) A stem structure in fibroblast growth factor receptor 2 transcripts mediates cell-type-specific splicing by approximating intronic control elements. *Mol.Cell Biol.* **23**, 9327-9337.
32. **Barnard,D.C. and Patton,J.G.** (2000) Identification and characterization of a novel serine-arginine-rich splicing regulatory protein. *Mol.Cell Biol.* **20**, 3049-3057.
33. **Barre-Sinoussi,F., Chermann,J.C., Rey,F. et al.** (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**, 868-871.
34. **Barreau,C., Paillard,L. and Osborne,H.B.** (2005) AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.* **33**, 7138-7150.
35. **Bartel,D.P., Zapp,M.L., Green,M.R. et al.** (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell* **67**, 529-536.

## References

---

36. **Baudin,F., Marquet,R., Isel,C. et al.** (1993) Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J.Mol.Biol.* **229**, 382-397.
37. **Bednenko,J., Cingolani,G. and Gerace,L.** (2003) Nucleocytoplasmic transport: navigating the channel. *Traffic.* **4**, 127-135.
38. **Beil,B., Screaton,G. and Stamm,S.** (1997) Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing. *DNA Cell Biol.* **16**, 679-690.
39. **Benko,D.M., Schwartz,S., Pavlakis,G.N. et al.** (1990) A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. *J.Virol.* **64**, 2505-2518.
40. **Bennett,M., Michaud,S., Kingston,J. et al.** (1992) Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes Dev.* **6**, 1986-2000.
41. **Bentley,D.** (2002) The mRNA assembly line: transcription and processing machines in the same factory. *Curr.Opin.Cell Biol.* **14**, 336-342.
42. **Berget,S.M.** (1995) Exon recognition in vertebrate splicing. *J.Biol.Chem.* **270**, 2411-2414.
43. **Berglund,J.A., Abovich,N. and Rosbash,M.** (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**, 858-867.
44. **Berglund,J.A., Chua,K., Abovich,N. et al.** (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**, 781-787.
45. **Berglund,J.A., Fleming,M.L. and Rosbash,M.** (1998) The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA.* **4**, 998-1006.
46. **Berkhout,B. and van Wamel,J.L.** (2000) The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA.* **6**, 282-295.
47. **Berthold,E. and Maldarelli,F.** (1996) cis-acting elements in human immunodeficiency virus type 1 RNAs direct viral transcripts to distinct intranuclear locations. *J.Virol.* **70**, 4667-4682.
48. **Beyer,A.L., Bouton,A.H. and Miller,O.L., Jr.** (1981) Correlation of hnRNP structure and nascent transcript cleavage. *Cell* **26**, 155-165.
49. **Beyer,A.L. and Osheim,Y.N.** (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* **2**, 754-765.
50. **Bilodeau,P.S., Domsic,J.K., Mayeda,A. et al.** (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. *J.Virol.* **75**, 8487-8497.
51. **Bird,P., Gething,M.J. and Sambrook,J.** (1987) Translocation in yeast and mammalian cells: not all signal sequences are functionally equivalent. *J.Cell Biol.* **105**, 2905-2914.
52. **Birney,E., Kumar,S. and Krainer,A.R.** (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**, 5803-5816.
53. **Black,D.L.** (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* **103**, 367-370.
54. **Black,D.L.** (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu.Rev.Biochem.* **72**, 291-336.

55. **Blanc,V., Navaratnam,N., Henderson,J.O. et al.** (2001) Identification of GRY-RBP as an apolipoprotein B RNA-binding protein that interacts with both apobec-1 and apobec-1 complementation factor to modulate C to U editing. *J.Biol.Chem.* **276**, 10272-10283.
56. **Blanchette,M. and Chabot,B.** (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* **18**, 1939-1952.
57. **Blaustein,M., Pelisch,F., Tanos,T. et al.** (2005) Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat.Struct.Mol.Biol.* **12**, 1037-1044.
58. **Blencowe,B.J.** (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem.Sci.* **25**, 106-110.
59. **Blencowe,B.J.** (2006) Alternative splicing: new insights from global analyses. *Cell* **126**, 37-47.
60. **Blencowe,B.J., Bauren,G., Eldridge,A.G. et al.** (2000) The SRm160/300 splicing coactivator subunits. *RNA.* **6**, 111-120.
61. **Blencowe,B.J., Bowman,J.A., McCracken,S. et al.** (1999) SR-related proteins and the processing of messenger RNA precursors. *Biochem.Cell Biol.* **77**, 277-291.
62. **Blencowe,B.J., Issner,R., Nickerson,J.A. et al.** (1998) A coactivator of pre-mRNA splicing. *Genes Dev.* **12**, 996-1009.
63. **Bohne,J. and Krausslich,H.G.** (2004) Mutation of the major 5' splice site renders a CMV-driven HIV-1 proviral clone Tat-dependent: connections between transcription and splicing. *FEBS Lett.* **563**, 113-118.
64. **Bohne,J., Wodrich,H. and Krausslich,H.G.** (2005) Splicing of human immunodeficiency virus RNA is position-dependent suggesting sequential removal of introns from the 5' end. *Nucleic Acids Res.* **33**, 825-837.
65. **Bolinger,C. and Boris-Lawrie,K.** (2009) Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. *Retrovirology.* **6**, 8.
66. **Bonnal,S., Martinez,C., Forch,P. et al.** (2008) RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Mol.Cell* **32**, 81-95.
67. **Boucher,L., Ouzounis,C.A., Enright,A.J. et al.** (2001) A genome-wide survey of RS domain proteins. *RNA.* **7**, 1693-1701.
68. **Boukis,L.A., Liu,N., Furuyama,S. et al.** (2004) Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles. *J.Biol.Chem.* **279**, 29647-29653.
69. **Bourgeois,C.F., Lejeune,F. and Stevenin,J.** (2004) Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog.Nucleic Acid Res.Mol.Biol.* **78**, 37-88.
70. **Bradford,M.M.** (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal.Biochem.* **72**, 248-254.
71. **Brasey,A., Lopez-Lastra,M., Ohlmann,T. et al.** (2003) The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. *J.Virol.* **77**, 3939-3949.
72. **Bray,M., Prasad,S., Dubay,J.W. et al.** (1994) A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proc.Natl.Acad.Sci.U.S.A* **91**, 1256-1260.
73. **Brennan,C.M., Gallouzi,I.E. and Steitz,J.A.** (2000) Protein ligands to HuR modulate its interaction with target mRNAs in vivo. *J.Cell Biol.* **151**, 1-14.

74. **Brighty,D.W. and Rosenberg,M.** (1994) A cis-acting repressive sequence that overlaps the Rev-responsive element of human immunodeficiency virus type 1 regulates nuclear retention of env mRNAs independently of known splice signals. *Proc.Natl.Acad.Sci.U.S.A* **91**, 8314-8318.
75. **Bringmann,P. and Luhrmann,R.** (1986) Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* **5**, 3509-3516.
76. **Brow,D.A.** (2002) Allosteric cascade of spliceosome activation. *Annu.Rev.Genet.* **36**, 333-360.
77. **Buck,C.B., Shen,X., Egan,M.A. et al.** (2001) The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. *J.Virol.* **75**, 181-191.
78. **Buckanovich,R.J. and Darnell,R.B.** (1997) The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol.Cell Biol.* **17**, 3194-3201.
79. **Buratti,E. and Baralle,F.E.** (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol.Cell Biol.* **24**, 10505-10514.
80. **Buratti,E., Baralle,M. and Baralle,F.E.** (2006) Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res.* **34**, 3494-3510.
81. **Buratti,E., Baralle,M., De Conti,L. et al.** (2004) hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Res.* **32**, 4224-4236.
82. **Buratti,E., Muro,A.F., Giombi,M. et al.** (2004) RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol.Cell Biol.* **24**, 1387-1400.
83. **Burge,C.B., Tuschl,T. and Sharp,P.A.** (1999) Splicing of Precursors to mRNAs by the Spliceosomes. In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds.), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 525-560.
84. **Burgess,S., Couto,J.R. and Guthrie,C.** (1990) A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell* **60**, 705-717.
85. **Burgess,S.M. and Guthrie,C.** (1993) Beat the clock: paradigms for NTPases in the maintenance of biological fidelity. *Trends Biochem.Sci.* **18**, 381-384.
86. **Caceres,J.F. and Kornblihtt,A.R.** (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* **18**, 186-193.
87. **Caceres,J.F. and Krainer,A.R.** (1993) Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J.* **12**, 4715-4726.
88. **Caceres,J.F., Stamm,S., Helfman,D.M. et al.** (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* **265**, 1706-1709.
89. **Cao,W. and Garcia-Blanco,M.A.** (1998) A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding. *J.Biol.Chem.* **273**, 20629-20635.
90. **Cao,W., Jamison,S.F. and Garcia-Blanco,M.A.** (1997) Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. *RNA.* **3**, 1456-1467.
91. **Caputi,M., Freund,M., Kammler,S. et al.** (2004) A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. *J.Virol.* **78**, 6517-6526.
92. **Caputi,M., Mayeda,A., Krainer,A.R. et al.** (1999) hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO J.* **18**, 4060-4067.

93. **Caputi,M. and Zahler,A.M.** (2001) Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J.Biol.Chem.* **276**, 43850-43859.
94. **Caputi,M. and Zahler,A.M.** (2002) SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *EMBO J.* **21**, 845-855.
95. **Cartegni,L., Chew,S.L. and Krainer,A.R.** (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat.Rev.Genet.* **3**, 285-298.
96. **Cartegni,L. and Krainer,A.R.** (2003) Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nat.Struct.Biol.* **10**, 120-125.
97. **Cartegni,L., Wang,J., Zhu,Z. et al.** (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31**, 3568-3571.
98. **Cass,D.M. and Berglund,J.A.** (2006) The SF3b155 N-terminal domain is a scaffold important for splicing. *Biochemistry* **45**, 10092-10101.
99. **Cavaloc,Y., Bourgeois,C.F., Kister,L. et al.** (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA.* **5**, 468-483.
100. **Chan,E.Y., Sutton,J.N., Jacobs,J.M. et al.** (2009) Dynamic host energetics and cytoskeletal proteomes in human immunodeficiency virus type 1-infected human primary CD4 cells: analysis by multiplexed label-free mass spectrometry. *J.Virol.* **83**, 9283-9295.
101. **Chan,R.C. and Black,D.L.** (1997) The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. *Mol.Cell Biol.* **17**, 4667-4676.
102. **Chandler,S.D., Mayeda,A., Yeakley,J.M. et al.** (1997) RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proc.Natl.Acad.Sci.U.S.A* **94**, 3596-3601.
103. **Chang,D.D. and Sharp,P.A.** (1989) Regulation by HIV Rev depends upon recognition of splice sites. *Cell* **59**, 789-795.
104. **Charlet,B., Logan,P., Singh,G. et al.** (2002) Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol.Cell* **9**, 649-658.
105. **Charpentier,B., Stutz,F. and Rosbash,M.** (1997) A dynamic in vivo view of the HIV-I Rev-RRE interaction. *J.Mol.Biol.* **266**, 950-962.
106. **Chasin,L.A.** (2007) Searching for splicing motifs. *Adv.Exp.Med.Biol.* **623**, 85-106.
107. **Chen,C.D., Kobayashi,R. and Helfman,D.M.** (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 Dev.* **13**, 593-606.
108. **Chen,C.Y. and Shyu,A.B.** (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem.Sci.* **20**, 465-470.
109. **Chen,H.H., Chang,J.G., Lu,R.M. et al.** (2008) The RNA binding protein hnRNP Q modulates the utilization of exon 7 in the survival motor neuron 2 (SMN2) gene. *Mol.Cell Biol.* **28**, 6929-6938.
110. **Chen,J.Y., Stands,L., Staley,J.P. et al.** (2001) Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol.Cell* **7**, 227-232.
111. **Chen,S., Anderson,K. and Moore,M.J.** (2000) Evidence for a linear search in bimolecular 3' splice site AG selection. *Proc.Natl.Acad.Sci.U.S.A* **97**, 593-598.

## References

---

112. **Chew,S.L., Liu,H.X., Mayeda,A. et al.** (1999) Evidence for the function of an exonic splicing enhancer after the first catalytic step of pre-mRNA splicing. *Proc.Natl.Acad.Sci.U.S.A* **96**, 10655-10660.
113. **Chou,M.Y., Rooke,N., Turck,C.W. et al.** (1999) hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol.Cell Biol.* **19**, 69-77.
114. **Chua,K. and Reed,R.** (1999) The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. *Nature* **402**, 207-210.
115. **Clavel,F., Guetard,D., Brun-Vezinet,F. et al.** (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233**, 343-346.
116. **Clever,J., Sasseti,C. and Parslow,T.G.** (1995) RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J.Virol.* **69**, 2101-2109.
117. **Clouaire,T. and Stancheva,I.** (2008) Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell Mol.Life Sci.* **65**, 1509-1522.
118. **Cmarko,D., Boe,S.O., Scassellati,C. et al.** (2002) Rev inhibition strongly affects intracellular distribution of human immunodeficiency virus type 1 RNAs. *J.Virol.* **76**, 10473-10484.
119. **Cochrane,A., Murley,L.L., Gao,M. et al.** (2009) Stable complex formation between HIV Rev and the nucleosome assembly protein, NAP1, affects Rev function. *Virology* **388**, 103-111.
120. **Cochrane,A.W., Chen,C.H. and Rosen,C.A.** (1990) Specific interaction of the human immunodeficiency virus Rev protein with a structured region in the env mRNA. *Proc.Natl.Acad.Sci.U.S.A* **87**, 1198-1202.
121. **Cochrane,A.W., Jones,K.S., Beidas,S. et al.** (1991) Identification and characterization of intragenic sequences which repress human immunodeficiency virus structural gene expression. *J.Virol.* **65**, 5305-5313.
122. **Cochrane,A.W., Perkins,A. and Rosen,C.A.** (1990) Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function. *J.Virol.* **64**, 881-885.
123. **Coffin,J., Haase,A., Levy,J.A. et al.** (1986) What to call the AIDS virus? *Nature* **321**, 10.
124. **Collins,C.A. and Guthrie,C.** (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.
125. **Colot,H.V., Stutz,F. and Rosbash,M.** (1996) The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes Dev.* **10**, 1699-1708.
126. **Colwill,K., Feng,L.L., Yeakley,J.M. et al.** (1996) SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. *J.Biol.Chem.* **271**, 24569-24575.
127. **Colwill,K., Pawson,T., Andrews,B. et al.** (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J.* **15**, 265-275.
128. **Conte,M.R., Grune,T., Ghuman,J. et al.** (2000) Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J.* **19**, 3132-3141.
129. **Coolidge,C.J., Seely,R.J. and Patton,J.G.** (1997) Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* **25**, 888-896.

130. **Cooper,T.A., Wan,L. and Dreyfuss,G.** (2009) RNA and disease. *Cell* **136**, 777-793.
131. **Cote,J., Dupuis,S., Jiang,Z. et al.** (2001) Caspase-2 pre-mRNA alternative splicing: Identification of an intronic element containing a decoy 3' acceptor site. *Proc.Natl.Acad.Sci.U.S.A* **98**, 938-943.
132. **Coulter,L.R., Landree,M.A. and Cooper,T.A.** (1997) Identification of a new class of exonic splicing enhancers by in vivo selection. *Mol.Cell Biol.* **17**, 2143-2150.
133. **Crawford,J.B. and Patton,J.G.** (2006) Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol.Cell Biol.* **26**, 8791-8802.
134. **Cullen,B.R.** (1991) Regulation of human immunodeficiency virus replication. *Annu.Rev.Microbiol.* **45**, 219-250.
135. **Cullen,B.R.** (2003) Nuclear RNA export. *J.Cell Sci.* **116**, 587-597.
136. **D'Agostino,D.M., Felber,B.K., Harrison,J.E. et al.** (1992) The Rev protein of human immunodeficiency virus type 1 promotes polysomal association and translation of gag/pol and vpu/env mRNAs. *Mol.Cell Biol.* **12**, 1375-1386.
137. **D'Agostino,D.M., Ferro,T., Zotti,L. et al.** (2000) Identification of a domain in human immunodeficiency virus type 1 rev that is required for functional activity and modulates association with subnuclear compartments containing splicing factor SC35. *J.Virol.* **74**, 11899-11910.
138. **Daefler,S., Klotman,M.E. and Wong-Staal,F.** (1990) Trans-activating rev protein of the human immunodeficiency virus 1 interacts directly and specifically with its target RNA. *Proc.Natl.Acad.Sci.U.S.A* **87**, 4571-4575.
139. **Daelemans,D., Costes,S.V., Cho,E.H. et al.** (2004) In vivo HIV-1 Rev multimerization in the nucleolus and cytoplasm identified by fluorescence resonance energy transfer. *J.Biol.Chem.* **279**, 50167-50175.
140. **Dalgleish,A.G., Beverley,P.C., Clapham,P.R. et al.** (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**, 763-767.
141. **Daly,T.J., Cook,K.S., Gray,G.S. et al.** (1989) Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. *Nature* **342**, 816-819.
142. **Damgaard, C. K.** University of Aarhus (2001), PhD thesis, Macromolecular interactions regulating genomic splicing, dimerisation and packaging in HIV-1. 1-143.
143. **Damgaard,C.K., Kahns,S., Lykke-Andersen,S. et al.** (2008) A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol.Cell* **29**, 271-278.
144. **Das,R., Yu,J., Zhang,Z. et al.** (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol.Cell* **26**, 867-881.
145. **Das,R., Zhou,Z. and Reed,R.** (2000) Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol.Cell* **5**, 779-787.
146. **Das,S.R. and Jameel,S.** (2005) Biology of the HIV Nef protein. *Indian J.Med.Res.* **121**, 315-332.
147. **Dauwalder,B., Amaya-Manzanares,F. and Mattox,W.** (1996) A human homologue of the Drosophila sex determination factor transformer-2 has conserved splicing regulatory functions. *Proc.Natl.Acad.Sci.U.S.A* **93**, 9004-9009.
148. **Dayton,A.I., Terwilliger,E.F., Potz,J. et al.** (1988) Cis-acting sequences responsive to the rev gene product of the human immunodeficiency virus. *J.Acquir.Immune.Defic.Syindr.* **1**, 441-452.

## References

---

149. **de la Mata,M., Alonso,C.R., Kadener,S. et al.** (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol.Cell* **12**, 525-532.
150. **De Rocquigny,H., Gabus,C., Vincent,A. et al.** (1992) Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc.Natl.Acad.Sci.U.S.A* **89**, 6472-6476.
151. **Demi,L., Bojak,A., Steck,S. et al.** (2001) Multiple effects of codon usage optimization on expression and immunogenicity of DNA candidate vaccines encoding the human immunodeficiency virus type 1 Gag protein. *J.Virol.* **75**, 10991-11001.
152. **Deng,H., Liu,R., Ellmeier,W. et al.** (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661-666.
153. **Dignam,J.D., Lebovitz,R.M. and Roeder,R.G.** (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475-1489.
154. **Ding,J.H., Xu,X., Yang,D. et al.** (2004) Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J.* **23**, 885-896.
155. **Dominski,Z. and Kole,R.** (1991) Selection of splice sites in pre-mRNAs with short internal exons. *Mol.Cell Biol.* **11**, 6075-6083.
156. **Dominski,Z. and Kole,R.** (1992) Cooperation of pre-mRNA sequence elements in splice site selection. *Mol.Cell Biol.* **12**, 2108-2114.
157. **Domsic,J.K., Wang,Y., Mayeda,A. et al.** (2003) Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts. *Mol.Cell Biol.* **23**, 8762-8772.
158. **Dong,C., Kwas,C. and Wu,L.** (2009) Transcriptional restriction of human immunodeficiency virus type 1 gene expression in undifferentiated primary monocytes. *J.Virol.* **83**, 3518-3527.
159. **Donmez,G., Hartmuth,K. and Luhrmann,R.** (2004) Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA.* **10**, 1925-1933.
160. **Douek,D.C., Picker,L.J. and Koup,R.A.** (2003) T cell dynamics in HIV-1 infection. *Annu.Rev.Immunol.* **21**, 265-304.
161. **Dowling,D., Nasr-Esfahani,S., Tan,C.H. et al.** (2008) HIV-1 infection induces changes in expression of cellular splicing factors that regulate alternative viral splicing and virus production in macrophages. *Retrovirology.* **5**, 18.
162. **Dragic,T., Litwin,V., Allaway,G.P. et al.** (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**, 667-673.
163. **Dreyfuss,G., Kim,V.N. and Kataoka,N.** (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat.Rev.Mol.Cell Biol.* **3**, 195-205.
164. **Du,H. and Rosbash,M.** (2002) The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419**, 86-90.
165. **Duncan,P.I., Stojdl,D.F., Marius,R.M. et al.** (1997) In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. *Mol.Cell Biol.* **17**, 5996-6001.
166. **Dye,M.J., Gromak,N. and Proudfoot,N.J.** (2006) Exon tethering in transcription by RNA polymerase II. *Mol.Cell* **21**, 849-859.
167. **Eldridge,A.G., Li,Y., Sharp,P.A. et al.** (1999) The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proc.Natl.Acad.Sci.U.S.A* **96**, 6125-6130.

168. **Emerman,M., Vazeux,R. and Peden,K.** (1989) The rev gene product of the human immunodeficiency virus affects envelope-specific RNA localization. *Cell* **57**, 1155-1165.
169. **Emili,A., Shales,M., McCracken,S. et al.** (2002) Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. *RNA*. **8**, 1102-1111.
170. **Eperon,I.C., Makarova,O.V., Mayeda,A. et al.** (2000) Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol.Cell Biol.* **20**, 8303-8318.
171. **Ernst,R.K., Bray,M., Rekosh,D. et al.** (1997) A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. *Mol.Cell Biol.* **17**, 135-144.
172. **Exline,C.M., Feng,Z. and Stoltzfus,C.M.** (2008) Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. *J.Virol.* **82**, 3921-3931.
173. **Expert-Bezancon,A., Le Caer,J.P. and Marie,J.** (2002) Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6A from chicken beta-tropomyosin pre-mRNA. *J.Biol.Chem.* **277**, 16614-16623.
174. **Fairbrother,W.G., Yeh,R.F., Sharp,P.A. et al.** (2002) Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007-1013.
175. **Fauci,A.S.** (1988) The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. *Science* **239**, 617-622.
176. **Favaro,J.P., Borg,K.T., Arrigo,S.J. et al.** (1998) Effect of Rev on the intranuclear localization of HIV-1 unspliced RNA. *Virology* **249**, 286-296.
177. **Fedorov,A., Saxonov,S., Fedorova,L. et al.** (2001) Comparison of intron-containing and intron-lacking human genes elucidates putative exonic splicing enhancers. *Nucleic Acids Res.* **29**, 1464-1469.
178. **Feinberg,M.B., Jarrett,R.F., Aldovini,A. et al.** (1986) HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell* **46**, 807-817.
179. **Felber,B.K., Hadzopoulou-Cladaras,M., Cladaras,C. et al.** (1989) rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc.Natl.Acad.Sci.U.S.A* **86**, 1495-1499.
180. **Feng,Y., Broder,C.C., Kennedy,P.E. et al.** (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872-877.
181. **Fischer,U., Huber,J., Boelens,W.C. et al.** (1995) The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82**, 475-483.
182. **Fleckner,J., Zhang,M., Valcarcel,J. et al.** (1997) U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* **11**, 1864-1872.
183. **Flößer, A.** Heinrich-Heine-Universität Düsseldorf (2008), diploma thesis, Charakterisierung cis-wirkender Elemente für die HIV-1 Exon 5 Erkennung. 1-125.
184. **Fong,N. and Bentley,D.L.** (2001) Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev.* **15**, 1783-1795.
185. **Forch,P., Puig,O., Martinez,C. et al.** (2002) The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *EMBO J.* **21**, 6882-6892.

## References

---

186. **Fornerod,M., Ohno,M., Yoshida,M. et al.** (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051-1060.
187. **Foster,J.L. and Garcia,J.V.** (2008) HIV-1 Nef: at the crossroads. *Retrovirology*. **5**, 84.
188. **Fox-Walsh,K.L., Dou,Y., Lam,B.J. et al.** (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc.Natl.Acad.Sci.U.S.A* **102**, 16176-16181.
189. **Fox-Walsh,K.L. and Hertel,K.J.** (2009) Splice-site pairing is an intrinsically high fidelity process. *Proc.Natl.Acad.Sci.U.S.A* **106**, 1766-1771.
190. **Freund,M., Asang,C., Kammler,S. et al.** (2003) A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res.* **31**, 6963-6975.
191. **Freund,M., Hicks,M.J., Konermann,C. et al.** (2005) Extended base pair complementarity between U1 snRNA and the 5' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5' splice site recognition. *Nucleic Acids Res.* **33**, 5112-5119.
192. **Fu,X.D. and Maniatis,T.** (1990) Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**, 437-441.
193. **Fu,X.D. and Maniatis,T.** (1992) The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proc.Natl.Acad.Sci.U.S.A* **89**, 1725-1729.
194. **Fukuda,M., Asano,S., Nakamura,T. et al.** (1997) CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390**, 308-311.
195. **Fukuhara,T., Hosoya,T., Shimizu,S. et al.** (2006) Utilization of host SR protein kinases and RNA-splicing machinery during viral replication. *Proc.Natl.Acad.Sci.U.S.A* **103**, 11329-11333.
196. **Furger,A., O'Sullivan,J.M., Binnie,A. et al.** (2002) Promoter proximal splice sites enhance transcription. *Genes Dev.* **16**, 2792-2799.
197. **Furtado,M.R., Balachandran,R., Gupta,P. et al.** (1991) Analysis of alternatively spliced human immunodeficiency virus type-1 mRNA species, one of which encodes a novel tat-env fusion protein. *Virology* **185**, 258-270.
198. **Gallego,J., Groatorex,J., Zhang,H. et al.** (2003) Rev binds specifically to a purine loop in the SL1 region of the HIV-1 leader RNA. *J.Biol.Chem.* **278**, 40385-40391.
199. **Gallo,R.C., Salahuddin,S.Z., Popovic,M. et al.** (1984) Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**, 500-503.
200. **Garneau,D., Revil,T., Fiset,J.F. et al.** (2005) Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J.Biol.Chem.* **280**, 22641-22650.
201. **Garrett,E.D., Tiley,L.S. and Cullen,B.R.** (1991) Rev activates expression of the human immunodeficiency virus type 1 vif and vpr gene products. *J.Virol.* **65**, 1653-1657.
202. **Ge,H. and Manley,J.L.** (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* **62**, 25-34.
203. **Ge,H., Noble,J., Colgan,J. et al.** (1990) Polyoma virus small tumor antigen pre-mRNA splicing requires cooperation between two 3' splice sites. *Proc.Natl.Acad.Sci.U.S.A* **87**, 3338-3342.
204. **Gevaert,K., Staes,A., Van Damme,J. et al.** (2005) Global phosphoproteome analysis on human HepG2 hepatocytes using reversed-phase diagonal LC. *Proteomics.* **5**, 3589-3599.

205. **Geyer,M., Fackler,O.T. and Peterlin,B.M.** (2001) Structure--function relationships in HIV-1 Nef. *EMBO Rep.* **2**, 580-585.
206. **Giles,K.E. and Beemon,K.L.** (2005) Retroviral splicing suppressor sequesters a 3' splice site in a 50S aberrant splicing complex. *Mol.Cell Biol.* **25**, 4397-4405.
207. **Goldstrohm,A.C., Greenleaf,A.L. and Garcia-Blanco,M.A.** (2001) Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene* **277**, 31-47.
208. **Gorelick,R.J., Chabot,D.J., Rein,A. et al.** (1993) The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *J.Virol.* **67**, 4027-4036.
209. **Gorelick,R.J., Nigida,S.M., Jr., Bess,J.W., Jr. et al.** (1990) Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J.Virol.* **64**, 3207-3211.
210. **Goren,A., Ram,O., Amit,M. et al.** (2006) Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol.Cell* **22**, 769-781.
211. **Gottlieb,M.S., Schroff,R., Schanker,H.M. et al.** (1981) Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N.Engl.J.Med.* **305**, 1425-1431.
212. **Gottlinger,H.G., Dorfman,T., Cohen,E.A. et al.** (1992) The role of the tnv protein and tnv RNA splicing signals in replication of HIV-1 IIIB isolates. *Virology* **189**, 618-628.
213. **Gougeon,M.L. and Piacentini,M.** (2009) New insights on the role of apoptosis and autophagy in HIV pathogenesis. *Apoptosis.* **14**, 501-508.
214. **Gozani,O., Feld,R. and Reed,R.** (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.* **10**, 233-243.
215. **Gozani,O., Potashkin,J. and Reed,R.** (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol.Cell Biol.* **18**, 4752-4760.
216. **Graf,M., Bojak,A., Deml,L. et al.** (2000) Concerted action of multiple cis-acting sequences is required for Rev dependence of late human immunodeficiency virus type 1 gene expression. *J.Virol.* **74**, 10822-10826.
217. **Graveley,B.R.** (2000) Sorting out the complexity of SR protein functions. *RNA.* **6**, 1197-1211.
218. **Graveley,B.R.** (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**, 100-107.
219. **Graveley,B.R.** (2002) Sex, AGility, and the regulation of alternative splicing. *Cell* **109**, 409-412.
220. **Graveley,B.R.** (2004) A protein interaction domain contacts RNA in the prespliceosome. *Mol.Cell* **13**, 302-304.
221. **Graveley,B.R.** (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**, 65-73.
222. **Graveley,B.R., Hertel,K.J. and Maniatis,T.** (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J.* **17**, 6747-6756.
223. **Graveley,B.R., Hertel,K.J. and Maniatis,T.** (2001) The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA.* **7**, 806-818.
224. **Graveley,B.R. and Maniatis,T.** (1998) Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol.Cell* **1**, 765-771.

## References

---

225. **Greatorex,J., Gallego,J., Varani,G. et al.** (2002) Structure and stability of wild-type and mutant RNA internal loops from the SL-1 domain of the HIV-1 packaging signal. *J.Mol.Biol.* **322**, 543-557.
226. **Greatorex,J.S., Palmer,E.A., Pomerantz,R.J. et al.** (2006) Mutation of the Rev-binding loop in the human immunodeficiency virus 1 leader causes a replication defect characterized by altered RNA trafficking and packaging. *J.Gen.Virol.* **87**, 3039-3044.
227. **Gromak,N., Matlin,A.J., Cooper,T.A. et al.** (2003) Antagonistic regulation of alpha-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein. *RNA.* **9**, 443-456.
228. **Groom,H.C., Anderson,E.C., Dangerfield,J.A. et al.** (2009) Rev regulates translation of human immunodeficiency virus type 1 RNAs. *J.Gen.Virol.* **90**, 1141-1147.
229. **Groom,H.C., Anderson,E.C. and Lever,A.M.** (2009) Rev: beyond nuclear export. *J.Gen.Virol.* **90**, 1303-1318.
230. **Grosseloh, F.-J.** Heinrich-Heine-Universität Düsseldorf (2006), diploma thesis, Beschreibung der Funktionalität einer 5'-Spleißstelle. 1-54.
231. **Gruter,P., Tabernero,C., von Kobbe,C. et al.** (1998) TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol.Cell* **1**, 649-659.
232. **Guatelli,J.C., Gingeras,T.R. and Richman,D.D.** (1990) Alternative splice acceptor utilization during human immunodeficiency virus type 1 infection of cultured cells. *J.Virol.* **64**, 4093-4098.
233. **Gui,J.F., Lane,W.S. and Fu,X.D.** (1994) A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature* **369**, 678-682.
234. **Guth,S., Martinez,C., Gaur,R.K. et al.** (1999) Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). *Mol.Cell Biol.* **19**, 8263-8271.
235. **Guth,S., Tange,T.O., Kellenberger,E. et al.** (2001) Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol.Cell Biol.* **21**, 7673-7681.
236. **Guth,S. and Valcarcel,J.** (2000) Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF. *J.Biol.Chem.* **275**, 38059-38066.
237. **Hacker,S. and Krebber,H.** (2004) Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. *J.Biol.Chem.* **279**, 5049-5052.
238. **Hadian,K., Vincendeau,M., Mausbacher,N. et al.** (2009) Identification of a heterogeneous nuclear ribonucleoprotein-recognition region in the HIV Rev protein. *J.Biol.Chem.* **284**, 33384-33391.
239. **Hadzopoulou-Cladaras,M., Felber,B.K., Cladaras,C. et al.** (1989) The rev (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a cis-acting sequence in the env region. *J.Virol.* **63**, 1265-1274.
240. **Hall,L.L., Smith,K.P., Byron,M. et al.** (2006) Molecular anatomy of a speckle. *Anat.Rec.A Discov.Mol.Cell Evol.Biol.* **288**, 664-675.
241. **Han,K., Yeo,G., An,P. et al.** (2005) A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS.Biol.* **3**, e158.
242. **Hanamura,A., Caceres,J.F., Mayeda,A. et al.** (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA.* **4**, 430-444.

- 
243. **Harrison,G.P. and Lever,A.M.** (1992) The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J.Virol.* **66**, 4144-4153.
244. **Hartmann,B. and Valcarcel,J.** (2009) Decrypting the genome's alternative messages. *Curr.Opin.Cell Biol.* **21**, 377-386.
245. **Hartmann, L.** Heinrich-Heine-Universität Düsseldorf (2005), diploma thesis, Charakterisierung funktionaler Sequenzelemente der HIV-1 Spleißakzeptoren . 1-79.
246. **Hartmann,L., Theiss,S., Niederacher,D. et al.** (2008) Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci.* **13**, 3252-3272.
247. **Hastings,M.L., Allemand,E., Duelli,D.M. et al.** (2007) Control of Pre-mRNA Splicing by the General Splicing Factors PUF60 and U2AF. *PLoS.ONE.* **2**, e538.
248. **Hastings,M.L., Wilson,C.M. and Munroe,S.H.** (2001) A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA. *RNA.* **7**, 859-874.
249. **Hauber,J.** (2001) Nuclear Export Mediated by the Rev/Rex Class of Retroviral Trans-activator Proteins. In Hauber,J. and Vogt,P.K. (eds.), *Nuclear Export of Viral RNAs*. Springer-Verlag, Vol. 259, pp. 55-76.
250. **Heaphy,S., Dingwall,C., Ernberg,I. et al.** (1990) HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. *Cell* **60**, 685-693.
251. **Heaphy,S., Finch,J.T., Gait,M.J. et al.** (1991) Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. *Proc.Natl.Acad.Sci.U.S.A* **88**, 7366-7370.
252. **Hedley,M.L. and Maniatis,T.** (1991) Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. *Cell* **65**, 579-586.
253. **Heinrichs,V., Bach,M., Winkelmann,G. et al.** (1990) U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science* **247**, 69-72.
254. **Hernandez,F., Perez,M., Lucas,J.J. et al.** (2004) Glycogen synthase kinase-3 plays a crucial role in tau exon 10 splicing and intranuclear distribution of SC35. Implications for Alzheimer's disease. *J.Biol.Chem.* **279**, 3801-3806.
255. **Herrmann,F., Bossert,M., Schwander,A. et al.** (2004) Arginine methylation of scaffold attachment factor A by heterogeneous nuclear ribonucleoprotein particle-associated PRMT1. *J.Biol.Chem.* **279**, 48774-48779.
256. **Hertel,K.J.** (2008) Combinatorial control of exon recognition. *J.Biol.Chem.* **283**, 1211-1215.
257. **Hertel,K.J., Lynch,K.W. and Maniatis,T.** (1997) Common themes in the function of transcription and splicing enhancers. *Curr.Opin.Cell Biol.* **9**, 350-357.
258. **Hertel,K.J. and Maniatis,T.** (1998) The function of multisite splicing enhancers. *Mol.Cell* **1**, 449-455.
259. **Hertel,K.J. and Maniatis,T.** (1999) Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. *Proc.Natl.Acad.Sci.U.S.A* **96**, 2651-2655.
260. **Hoffman,B.E. and Grabowski,P.J.** (1992) U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev.* **6**, 2554-2568.

## References

---

261. **Hofmann,Y., Lorson,C.L., Stamm,S. et al.** (2000) Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). *Proc.Natl.Acad.Sci.U.S.A* **97**, 9618-9623.
262. **Holland,S.M., Chavez,M., Gerstberger,S. et al.** (1992) A specific sequence with a bulged guanosine residue(s) in a stem-bulge-stem structure of Rev-responsive element RNA is required for trans activation by human immunodeficiency virus type 1 Rev. *J.Virol.* **66**, 3699-3706.
263. **Honisch, E.** Heinrich-Heine-Universität Düsseldorf (2005), diploma thesis, Analyse von alternativen Spleißvarianten tumorassoziierter Gene. 1-91.
264. **Hope,T.J., McDonald,D., Huang,X.J. et al.** (1990) Mutational analysis of the human immunodeficiency virus type 1 Rev transactivator: essential residues near the amino terminus. *J.Virol.* **64**, 5360-5366.
265. **House,A.E. and Lynch,K.W.** (2006) An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nat.Struct.Mol.Biol.* **13**, 937-944.
266. **House,A.E. and Lynch,K.W.** (2008) Regulation of alternative splicing: more than just the ABCs. *J.Biol.Chem.* **283**, 1217-1221.
267. **Houseley,J., LaCava,J. and Tollervey,D.** (2006) RNA-quality control by the exosome. *Nat.Rev.Mol.Cell Biol.* **7**, 529-539.
268. **Howe,K.J., Kane,C.M. and Ares,M., Jr.** (2003) Perturbation of transcription elongation influences the fidelity of internal exon inclusion in *Saccharomyces cerevisiae*. *RNA.* **9**, 993-1006.
269. **Huang,T.S., Nilsson,C.E., Punga,T. et al.** (2002) Functional inactivation of the SR family of splicing factors during a vaccinia virus infection. *EMBO Rep.* **3**, 1088-1093.
270. **Huang,Y., Yario,T.A. and Steitz,J.A.** (2004) A molecular link between SR protein dephosphorylation and mRNA export. *Proc.Natl.Acad.Sci.U.S.A* **101**, 9666-9670.
271. **Hui,J., Stangl,K., Lane,W.S. et al.** (2003) HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat.Struct.Biol.* **10**, 33-37.
272. **Hutten,S. and Kehlenbach,R.H.** (2007) CRM1-mediated nuclear export: to the pore and beyond. *Trends Cell Biol.* **17**, 193-201.
273. **Hutter,G., Nowak,D., Mossner,M. et al.** (2009) Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N.Engl.J.Med.* **360**, 692-698.
274. **Hwang,D.Y. and Cohen,J.B.** (1997) U1 small nuclear RNA-promoted exon selection requires a minimal distance between the position of U1 binding and the 3' splice site across the exon. *Mol.Cell Biol.* **17**, 7099-7107.
275. **Ibrahim,e.C., Schaal,T.D., Hertel,K.J. et al.** (2005) Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc.Natl.Acad.Sci.U.S.A* **102**, 5002-5007.
276. **Iglesias,N. and Stutz,F.** (2008) Regulation of mRNP dynamics along the export pathway. *FEBS Lett.* **582**, 1987-1996.
277. **Inoue,K., Hoshijima,K., Higuchi,I. et al.** (1992) Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of doublesex primary transcript for sex-specific RNA processing. *Proc.Natl.Acad.Sci.U.S.A* **89**, 8092-8096.
278. **Iwamoto,F., Stadler,M., Chalupnikova,K. et al.** (2008) Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU. *Exp.Cell Res.* **314**, 1378-1391.

- 
279. **Izquierdo,J.M., Majos,N., Bonnal,S. et al.** (2005) Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol.Cell* **19**, 475-484.
280. **Jablonski,J.A., Buratti,E., Stuani,C. et al.** (2008) The secondary structure of the human immunodeficiency virus type 1 transcript modulates viral splicing and infectivity. *J.Virol.* **82**, 8038-8050.
281. **Jacks,T., Power,M.D., Masiarz,F.R. et al.** (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331**, 280-283.
282. **Jacquetet,S., Mereau,A., Bilodeau,P.S. et al.** (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. *J.Biol.Chem.* **276**, 40464-40475.
283. **Jacquetet,S., Ropers,D., Bilodeau,P.S. et al.** (2001) Conserved stem-loop structures in the HIV-1 RNA region containing the A3 3' splice site and its cis-regulatory element: possible involvement in RNA splicing. *Nucleic Acids Res.* **29**, 464-478.
284. **Jamison,S.F., Pasman,Z., Wang,J. et al.** (1995) U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. *Nucleic Acids Res.* **23**, 3260-3267.
285. **Johnson,C., Primorac,D., McKinstry,M. et al.** (2000) Tracking COL1A1 RNA in osteogenesis imperfecta. splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain. *J.Cell Biol.* **150**, 417-432.
286. **Jones,P.A. and Takai,D.** (2001) The role of DNA methylation in mammalian epigenetics. *Science* **293**, 1068-1070.
287. **Jumaa,H. and Nielsen,P.J.** (1997) The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.* **16**, 5077-5085.
288. **Jumaa,H., Wei,G. and Nielsen,P.J.** (1999) Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. *Curr.Biol.* **9**, 899-902.
289. **Jurica,M.S. and Moore,M.J.** (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol.Cell* **12**, 5-14.
290. **Kabat,J.L., Barberan-Soler,S. and Zahler,A.M.** (2009) HRP-2, the *Caenorhabditis elegans* homolog of mammalian heterogeneous nuclear ribonucleoproteins Q and R, is an alternative splicing factor that binds to UCUAUC splicing regulatory elements. *J.Biol.Chem.* **284**, 28490-28497.
291. **Kadener,S., Cramer,P., Nogues,G. et al.** (2001) Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *EMBO J.* **20**, 5759-5768.
292. **Kadener,S., Fededa,J.P., Rosbash,M. et al.** (2002) Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proc.Natl.Acad.Sci.U.S.A* **99**, 8185-8190.
293. **Kalland,K.H., Szilvay,A.M., Brokstad,K.A. et al.** (1994) The human immunodeficiency virus type 1 Rev protein shuttles between the cytoplasm and nuclear compartments. *Mol.Cell Biol.* **14**, 7436-7444.
294. **Kalland,K.H., Szilvay,A.M., Langhoff,E. et al.** (1994) Subcellular distribution of human immunodeficiency virus type 1 Rev and colocalization of Rev with RNA splicing factors in a speckled pattern in the nucleoplasm. *J.Virol.* **68**, 1475-1485.
295. **Kammler,S., Leurs,C., Freund,M. et al.** (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.

## References

---

296. **Kammler,S., Otte,M., Hauber,I. et al.** (2006) The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology*. **3**, 89.
297. **Kan,J.L. and Green,M.R.** (1999) Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. *Genes Dev.* **13**, 462-471.
298. **Kanaar,R., Roche,S.E., Beall,E.L. et al.** (1993) The conserved pre-mRNA splicing factor U2AF from Drosophila: requirement for viability. *Science* **262**, 569-573.
299. **Kang,D., Gho,Y.S., Suh,M. et al.** (2002) Highly sensitive and fast protein detection with coomassie brilliant blue in sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Bull.Korean Chem.Soc.* **23**, 1511-1512.
300. **Kanopka,A., Muhlemann,O. and Akusjarvi,G.** (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**, 535-538.
301. **Kanopka,A., Muhlemann,O., Petersen-Mahrt,S. et al.** (1998) Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* **393**, 185-187.
302. **Kataoka,N., Yong,J., Kim,V.N. et al.** (2000) Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol.Cell* **6**, 673-682.
303. **Kaziyo,Y.** (1978) The role of guanosine 5'-triphosphate in polypeptide chain elongation. *Biochim.Biophys.Acta* **505**, 95-127.
304. **Kelly,S.M. and Corbett,A.H.** (2009) Messenger RNA Export from the Nucleus: A Series of Molecular Wardrobe Changes. *Traffic*.
305. **Kent,O.A. and MacMillan,A.M.** (2002) Early organization of pre-mRNA during spliceosome assembly. *Nat.Struct.Biol.* **9**, 576-581.
306. **Kent,O.A., Reayi,A., Foong,L. et al.** (2003) Structuring of the 3' splice site by U2AF65. *J.Biol.Chem.* **278**, 50572-50577.
307. **Kent,O.A., Ritchie,D.B. and MacMillan,A.M.** (2005) Characterization of a U2AF-independent commitment complex (E') in the mammalian spliceosome assembly pathway. *Mol.Cell Biol.* **25**, 233-240.
308. **Khabar,K.S.** (2005) The AU-rich transcriptome: more than interferons and cytokines, and its role in disease. *J.Interferon Cytokine Res.* **25**, 1-10.
309. **Kielkopf,C.L., Lucke,S. and Green,M.R.** (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes Dev.* **18**, 1513-1526.
310. **Kielkopf,C.L., Rodionova,N.A., Green,M.R. et al.** (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **106**, 595-605.
311. **Kim,S., Merrill,B.M., Rajpurohit,R. et al.** (1997) Identification of N(G)-methylarginine residues in human heterogeneous RNP protein A1: Phe/Gly-Gly-Gly-Arg-Gly-Gly-Gly/Phe is a preferred recognition motif. *Biochemistry* **36**, 5185-5192.
312. **Kim,S.Y., Byrn,R., Groopman,J. et al.** (1989) Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J.Virol.* **63**, 3708-3713.
313. **Kjems,J., Brown,M., Chang,D.D. et al.** (1991) Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. *Proc.Natl.Acad.Sci.U.S.A* **88**, 683-687.
314. **Kjems,J., Calnan,B.J., Frankel,A.D. et al.** (1992) Specific binding of a basic peptide from HIV-1 Rev. *EMBO J.* **11**, 1119-1129.

315. **Klatzmann,D., Champagne,E., Chamaret,S. et al.** (1984) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**, 767-768.
316. **Klose,R.J. and Bird,A.P.** (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem.Sci.* **31**, 89-97.
317. **Klotman,M.E., Kim,S., Buchbinder,A. et al.** (1991) Kinetics of expression of multiply spliced RNA in early human immunodeficiency virus type 1 infection of lymphocytes and monocytes. *Proc.Natl.Acad.Sci.U.S.A* **88**, 5011-5015.
318. **Kohtz,J.D., Jamison,S.F., Will,C.L. et al.** (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**, 119-124.
319. **Konarska,M.M., Vilardell,J. and Query,C.C.** (2006) Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol.Cell* **21**, 543-553.
320. **Konermann, C.** Heinrich-Heine-Universität Düsseldorf (2004), diploma thesis, Der Einfluss *cis*-wirkender Sequenzen auf die Regulation des alternativen Spleißens der HIV-1 Exons mit der 5' Spleißstelle #4. 1-80.
321. **Kornblihtt,A.R.** (2005) Promoter usage and alternative splicing. *Curr.Opin.Cell Biol.* **17**, 262-268.
322. **Kornblihtt,A.R.** (2007) Coupling transcription and alternative splicing. *Adv.Exp.Med.Biol.* **623**, 175-189.
323. **Kotlajich,M.V., Crabb,T.L. and Hertel,K.J.** (2009) Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex. *Mol.Cell Biol.* **29**, 1072-1082.
324. **Krainer,A.R., Conway,G.C. and Kozak,D.** (1990) Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev.* **4**, 1158-1171.
325. **Kramer,A.** (1992) Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. *Mol.Cell Biol.* **12**, 4545-4552.
326. **Kramer,A. and Utans,U.** (1991) Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. *EMBO J.* **10**, 1503-1509.
327. **Kraunus,J., Zychlinski,D., Heise,T. et al.** (2006) Murine leukemia virus regulates alternative splicing through sequences upstream of the 5' splice site. *J.Biol.Chem.* **281**, 37381-37390.
328. **Krecic,A.M. and Swanson,M.S.** (1999) hnRNP complexes: composition, structure, and function. *Curr.Opin.Cell Biol.* **11**, 363-371.
329. **Krummheuer,J., Johnson,A.T., Hauber,I. et al.** (2007) A minimal uORF within the HIV-1 vpu leader allows efficient translation initiation at the downstream env AUG. *Virology* **363**, 261-271.
330. **Kubota,S., Siomi,H., Satoh,T. et al.** (1989) Functional similarity of HIV-I rev and HTLV-I rex proteins: identification of a new nucleolar-targeting signal in rev protein. *Biochem.Biophys.Res.Commun.* **162**, 963-970.
331. **Kuroyanagi,N., Onogi,H., Wakabayashi,T. et al.** (1998) Novel SR-protein-specific kinase, SRPK2, disassembles nuclear speckles. *Biochem.Biophys.Res.Commun.* **242**, 357-364.
332. **Kvissel,A.K., Orstavik,S., Eikvar,S. et al.** (2007) Involvement of the catalytic subunit of protein kinase A and of HA95 in pre-mRNA splicing. *Exp.Cell Res.* **313**, 2795-2809.
333. **Labourier,E., Adams,M.D. and Rio,D.C.** (2001) Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol.Cell* **8**, 363-373.

## References

---

334. **Ladd,A.N., Charlet,N. and Cooper,T.A.** (2001) The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol.Cell Biol.* **21**, 1285-1296.
335. **Laemmli,U.K.** (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
336. **Lai,M.C., Lin,R.I., Huang,S.Y. et al.** (2000) A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J.Biol.Chem.* **275**, 7950-7957.
337. **Lai,M.C., Peng,T.Y. and Tarn,W.Y.** (2009) Functional interplay between viral and cellular SR proteins in control of post-transcriptional gene regulation. *FEBS J.* **276**, 1517-1526.
338. **Lallena,M.J., Chalmers,K.J., Llamazares,S. et al.** (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* **109**, 285-296.
339. **Lam,B.J. and Hertel,K.J.** (2002) A general role for splicing enhancers in exon definition. *RNA.* **8**, 1233-1241.
340. **Lamond,A.I. and Spector,D.L.** (2003) Nuclear speckles: a model for nuclear organelles. *Nat.Rev.Mol.Cell Biol.* **4**, 605-612.
341. **Lander,E.S., Linton,L.M., Birren,B. et al.** (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
342. **Le Hir,H. and Andersen,G.R.** (2008) Structural insights into the exon junction complex. *Curr.Opin.Struct.Biol.* **18**, 112-119.
343. **Le Hir,H., Gatfield,D., Izaurralde,E. et al.** (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**, 4987-4997.
344. **Le Hir,H., Izaurralde,E., Maquat,L.E. et al.** (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* **19**, 6860-6869.
345. **Le Hir,H., Moore,M.J. and Maquat,L.E.** (2000) Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.* **14**, 1098-1108.
346. **Lemaire,R., Prasad,J., Kashima,T. et al.** (2002) Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes Dev.* **16**, 594-607.
347. **Lewis,J.D., Izaurralde,E., Jarmolowski,A. et al.** (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev.* **10**, 1683-1698.
348. **Li,M.J., Kim,J., Li,S. et al.** (2005) Long-term inhibition of HIV-1 infection in primary hematopoietic cells by lentiviral vector delivery of a triple combination of anti-HIV shRNA, anti-CCR5 ribozyme, and a nucleolar-localizing TAR decoy. *Mol.Ther.* **12**, 900-909.
349. **Li,T., Evdokimov,E., Shen,R.F. et al.** (2004) Sumoylation of heterogeneous nuclear ribonucleoproteins, zinc finger proteins, and nuclear pore complex proteins: a proteomic analysis. *Proc.Natl.Acad.Sci.U.S.A* **101**, 8551-8556.
350. **Li,X. and Manley,J.L.** (2005) Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365-378.
351. **Li,X., Niu,T. and Manley,J.L.** (2007) The RNA binding protein RNPS1 alleviates ASF/SF2 depletion-induced genomic instability. *RNA.* **13**, 2108-2115.

- 
352. **Li,Y. and Blencowe,B.J.** (1999) Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. *J.Biol.Chem.* **274**, 35074-35079.
353. **Licatalosi,D.D., Mele,A., Fak,J.J. et al.** (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-469.
354. **Lim,F. and Peabody,D.S.** (2002) RNA recognition site of PP7 coat protein. *Nucleic Acids Res.* **30**, 4138-4144.
355. **Lim,S.R. and Hertel,K.J.** (2004) Commitment to splice site pairing coincides with A complex formation. *Mol.Cell* **15**, 477-483.
356. **Lin,C.H. and Patton,J.G.** (1995) Regulation of alternative 3' splice site selection by constitutive splicing factors. *RNA*. **1**, 234-245.
357. **Lin,S. and Fu,X.D.** (2007) SR proteins and related factors in alternative splicing. *Adv.Exp.Med.Biol.* **623**, 107-122.
358. **Liou,L.Y., Herrmann,C.H. and Rice,A.P.** (2002) Transient induction of cyclin T1 during human macrophage differentiation regulates human immunodeficiency virus type 1 Tat transactivation function. *J.Virol.* **76**, 10579-10587.
359. **Little,S.J., Holte,S., Routy,J.P. et al.** (2002) Antiretroviral-drug resistance among patients recently infected with HIV. *N.Engl.J.Med.* **347**, 385-394.
360. **Liu,H.X., Zhang,M. and Krainer,A.R.** (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* **12**, 1998-2012.
361. **Liu,Q. and Dreyfuss,G.** (1995) In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol.Cell Biol.* **15**, 2800-2808.
362. **Liu,Z., Luyten,I., Bottomley,M.J. et al.** (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098-1102.
363. **Long,J.C. and Caceres,J.F.** (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem.J.* **417**, 15-27.
364. **Loomis,R.J., Naoe,Y., Parker,J.B. et al.** (2009) Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol.Cell* **33**, 450-461.
365. **Lopez,d.S., I, Zhan,M., Lal,A. et al.** (2004) Identification of a target RNA motif for RNA-binding protein HuR. *Proc.Natl.Acad.Sci.U.S.A* **101**, 2987-2992.
366. **Lopez,P.J. and Seraphin,B.** (1999) Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition. *RNA*. **5**, 1135-1137.
367. **Lu,X.B., Heimer,J., Rekosh,D. et al.** (1990) U1 small nuclear RNA plays a direct role in the formation of a rev-regulated human immunodeficiency virus env mRNA that remains unspliced. *Proc.Natl.Acad.Sci.U.S.A* **87**, 7598-7602.
368. **Lui,K. and Huang,Y.** (2009) RanGTPase: A Key Regulator of Nucleocytoplasmic Trafficking. *Mol.Cell Pharmacol.* **1**, 148-156.
369. **Luo,M.J. and Reed,R.** (1999) Splicing is required for rapid and efficient mRNA export in metazoans. *Proc.Natl.Acad.Sci.U.S.A* **96**, 14937-14942.
370. **Luo,M.L., Zhou,Z., Magni,K. et al.** (2001) Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413**, 644-647.

## References

---

371. **Lusso,P., Cocchi,F., Balotta,C. et al.** (1995) Growth of macrophage-tropic and primary human immunodeficiency virus type 1 (HIV-1) isolates in a unique CD4+ T-cell clone (PM1): failure to downregulate CD4 and to interfere with cell-line-tropic HIV-1. *J.Virol.* **69**, 3712-3720.
372. **Lutzelberger,M., Backstrom,E. and Akusjarvi,G.** (2005) Substrate-dependent differences in U2AF requirement for splicing in adenovirus-infected cell extracts. *J.Biol.Chem.* **280**, 25478-25484.
373. **Lutzelberger,M., Reinert,L.S., Das,A.T. et al.** (2006) A novel splice donor site in the gag-pol gene is required for HIV-1 RNA stability. *J.Biol.Chem.* **281**, 18644-18651.
374. **Lynch,K.W. and Maniatis,T.** (1995) Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev.* **9**, 284-293.
375. **Lynch,K.W. and Maniatis,T.** (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes Dev.* **10**, 2089-2101.
376. **Ma,C.T., Hagopian,J.C., Ghosh,G. et al.** (2009) Regiospecific phosphorylation control of the SR protein ASF/SF2 by SRPK1. *J.Mol.Biol.* **390**, 618-634.
377. **Ma,C.T., Velazquez-Dones,A., Hagopian,J.C. et al.** (2008) Ordered multi-site phosphorylation of the splicing factor ASF/SF2 by SRPK1. *J.Mol.Biol.* **376**, 55-68.
378. **MacMillan,A.M., McCaw,P.S., Crispino,J.D. et al.** (1997) SC35-mediated reconstitution of splicing in U2AF-depleted nuclear extract. *Proc.Natl.Acad.Sci.U.S.A* **94**, 133-136.
379. **Maddon,P.J., Dalgleish,A.G., McDougal,J.S. et al.** (1986) The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**, 333-348.
380. **Madsen,J.M. and Stoltzfus,C.M.** (2005) An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. *J.Virol.* **79**, 10478-10486.
381. **Malca,H., Shomron,N. and Ast,G.** (2003) The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. *Mol.Cell Biol.* **23**, 3442-3455.
382. **Maldarelli,F., Martin,M.A. and Strebel,K.** (1991) Identification of posttranscriptionally active inhibitory sequences in human immunodeficiency virus type 1 RNA: novel level of gene regulation. *J.Virol.* **65**, 5732-5743.
383. **Maldarelli,F., Xiang,C., Chamoun,G. et al.** (1998) The expression of the essential nuclear splicing factor SC35 is altered by human immunodeficiency virus infection. *Virus Res.* **53**, 39-51.
384. **Malim,M.H., Bohnlein,S., Hauber,J. et al.** (1989) Functional dissection of the HIV-1 Rev trans-activator--derivation of a trans-dominant repressor of Rev function. *Cell* **58**, 205-214.
385. **Malim,M.H. and Cullen,B.R.** (1991) HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* **65**, 241-248.
386. **Malim,M.H. and Cullen,B.R.** (1993) Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. *Mol.Cell Biol.* **13**, 6180-6189.
387. **Malim,M.H., Hauber,J., Fenrick,R. et al.** (1988) Immunodeficiency virus rev trans-activator modulates the expression of the viral regulatory genes. *Nature* **335**, 181-183.
388. **Malim,M.H., Hauber,J., Le,S.Y. et al.** (1989) The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254-257.
389. **Malim,M.H., Tiley,L.S., McCarn,D.F. et al.** (1990) HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* **60**, 675-683.

- 
390. **Manceau,V., Swenson,M., Le Caer,J.P. et al.** (2006) Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65. *FEBS J.* **273**, 577-587.
391. **Maniatis,T. and Reed,R.** (2002) An extensive network of coupling among gene expression machines. *Nature* **416**, 499-506.
392. **Maniatis,T. and Tasic,B.** (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236-243.
393. **Mann,D.A., Mikaelian,I., Zimmel,R.W. et al.** (1994) A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J.Mol.Biol.* **241**, 193-207.
394. **Marcucci,R., Baralle,F.E. and Romano,M.** (2007) Complex splicing control of the human Thrombopoietin gene by intronic G runs. *Nucleic Acids Res.* **35**, 132-142.
395. **Markovtsov,V., Nikolic,J.M., Goldman,J.A. et al.** (2000) Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol.Cell Biol.* **20**, 7463-7479.
396. **Martinez-Contreras,R., Cloutier,P., Shkreta,L. et al.** (2007) hnRNP proteins and splicing control. *Adv.Exp.Med.Biol.* **623**, 123-147.
397. **Martinez-Contreras,R., Fisette,J.F., Nasim,F.U. et al.** (2006) Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS.Biol.* **4**, e21.
398. **Martins,d.A., Bonnal,S., Hastings,M.L. et al.** (2009) Differential 3' splice site recognition of SMN1 and SMN2 transcripts by U2AF and U2 snRNP. *RNA.* **15**, 515-523.
399. **Masur,H., Michelis,M.A., Greene,J.B. et al.** (1981) An outbreak of community-acquired *Pneumocystis carinii* pneumonia: initial manifestation of cellular immune dysfunction. *N.Engl.J.Med.* **305**, 1431-1438.
400. **Matlin,A.J., Clark,F. and Smith,C.W.** (2005) Understanding alternative splicing: towards a cellular code. *Nat.Rev.Mol.Cell Biol.* **6**, 386-398.
401. **Matsuo,N., Ogawa,S., Imai,Y. et al.** (1995) Cloning of a novel RNA binding polypeptide (RA301) induced by hypoxia/reoxygenation. *J.Biol.Chem.* **270**, 28216-28222.
402. **Matsuoka,S., Ballif,B.A., Smogorzewska,A. et al.** (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160-1166.
403. **Matunis,E.L., Matunis,M.J. and Dreyfuss,G.** (1993) Association of individual hnRNP proteins and snRNPs with nascent transcripts. *J.Cell Biol.* **121**, 219-228.
404. **Mayeda,A. and Krainer,A.R.** (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**, 365-375.
405. **McDougal,J.S., Kennedy,M.S., Sligh,J.M. et al.** (1986) Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **231**, 382-385.
406. **Melamud,E. and Moul,J.** (2009) Stochastic noise in splicing machinery. *Nucleic Acids Res.* **37**, 4873-4886.
407. **Melamud,E. and Moul,J.** (2009) Structural implication of splicing stochasticity. *Nucleic Acids Res.* **37**, 4862-4872.
408. **Merendino,L., Guth,S., Bilbao,D. et al.** (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**, 838-841.

## References

---

409. **Mermoud,J.E., Cohen,P. and Lamond,A.I.** (1992) Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res.* **20**, 5263-5269.
410. **Mermoud,J.E., Cohen,P.T. and Lamond,A.I.** (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J.* **13**, 5679-5688.
411. **Messias,A.C. and Sattler,M.** (2004) Structural basis of single-stranded RNA recognition. *Acc.Chem.Res.* **37**, 279-287.
412. **Meyer,B.E. and Malim,M.H.** (1994) The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes Dev.* **8**, 1538-1547.
413. **Michael,N.L., Morrow,P., Mosca,J. et al.** (1991) Induction of human immunodeficiency virus type 1 expression in chronically infected cells is associated primarily with a shift in RNA splicing patterns. *J.Virol.* **65**, 1291-1303.
414. **Michaud,S. and Reed,R.** (1991) An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.* **5**, 2534-2546.
415. **Michaud,S. and Reed,R.** (1993) A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev.* **7**, 1008-1020.
416. **Michienzi,A., Cagnon,L., Bahner,I. et al.** (2000) Ribozyme-mediated inhibition of HIV 1 suggests nucleolar trafficking of HIV-1 RNA. *Proc.Natl.Acad.Sci.U.S.A* **97**, 8955-8960.
417. **Michienzi,A., De Angelis,F.G., Bozzoni,I. et al.** (2006) A nucleolar localizing Rev binding element inhibits HIV replication. *AIDS Res.Ther.* **3**, 13.
418. **Mikaelian,I., Krieg,M., Gait,M.J. et al.** (1996) Interactions of INS (CRS) elements and the splicing machinery regulate the production of Rev-responsive mRNAs. *J.Mol.Biol.* **257**, 246-264.
419. **Mikula,M., Karczmariski,J., Dzwonek,A. et al.** (2006) Casein kinases phosphorylate multiple residues spanning the entire hnRNP K length. *Biochim.Biophys.Acta* **1764**, 299-306.
420. **Min,H., Chan,R.C. and Black,D.L.** (1995) The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes Dev.* **9**, 2659-2671.
421. **Misteli,T., Caceres,J.F. and Spector,D.L.** (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**, 523-527.
422. **Misteli,T. and Spector,D.L.** (1999) RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol.Cell* **3**, 697-705.
423. **Moldon,A., Malapeira,J., Gabrielli,N. et al.** (2008) Promoter-driven splicing regulation in fission yeast. *Nature* **455**, 997-1000.
424. **Moore,M.J.** (2000) Intron recognition comes of AGE. *Nat.Struct.Biol.* **7**, 14-16.
425. **Moroy,T. and Heyd,F.** (2007) The impact of alternative splicing in vivo: mouse models show the way. *RNA.* **13**, 1155-1171.
426. **Morris,D.P. and Greenleaf,A.L.** (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J.Biol.Chem.* **275**, 39935-39943.
427. **Mourelatos,Z., Abel,L., Yong,J. et al.** (2001) SMN interacts with a novel family of hnRNP and spliceosomal proteins. *EMBO J.* **20**, 5443-5452.
428. **Mukherjee,D., Gao,M., O'Connor,J.P. et al.** (2002) The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J.* **21**, 165-174.

429. **Munis,J.R., Kornbluth,R.S., Guatelli,J.C. et al.** (1992) Ordered appearance of human immunodeficiency virus type 1 nucleic acids following high multiplicity infection of macrophages. *J.Gen.Virol.* **73 ( Pt 8)**, 1899-1906.
430. **Munoz,M.J., Perez Santangelo,M.S., Paronetto,M.P. et al.** (2009) DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* **137**, 708-720.
431. **Muraki,M., Ohkawara,B., Hosoya,T. et al.** (2004) Manipulation of alternative splicing by a newly developed inhibitor of Clks. *J.Biol.Chem.* **279**, 24246-24254.
432. **Muro,A.F., Caputi,M., Pariyarath,R. et al.** (1999) Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol.Cell Biol.* **19**, 2657-2671.
433. **Murphy,E.L., Collier,A.C., Kalish,L.A. et al.** (2001) Highly active antiretroviral therapy decreases mortality and morbidity in patients with advanced HIV disease. *Ann.Intern.Med.* **135**, 17-26.
434. **Nagai,K., Oubridge,C., Jessen,T.H. et al.** (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**, 515-520.
435. **Nagel,R.J., Lancaster,A.M. and Zahler,A.M.** (1998) Specific binding of an exonic splicing enhancer by the pre-mRNA splicing factor SRp55. *RNA.* **4**, 11-23.
436. **Najera,I., Krieg,M. and Karn,J.** (1999) Synergistic stimulation of HIV-1 rev-dependent export of unspliced mRNA to the cytoplasm by hnRNP A1. *J.Mol.Biol.* **285**, 1951-1964.
437. **Nakagawa,O., Arnold,M., Nakagawa,M. et al.** (2005) Centronuclear myopathy in mice lacking a novel muscle-specific protein kinase transcriptionally regulated by MEF2. *Genes Dev.* **19**, 2066-2077.
438. **Nasioulas,G., Zolotukhin,A.S., Taberner,C. et al.** (1994) Elements distinct from human immunodeficiency virus type 1 splice sites are responsible for the Rev dependence of env mRNA. *J.Virol.* **68**, 2986-2993.
439. **Nayler,O., Stamm,S. and Ullrich,A.** (1997) Characterization and comparison of four serine- and arginine-rich (SR) protein kinases. *Biochem.J.* **326 ( Pt 3)**, 693-700.
440. **Neubauer,G., Gottschalk,A., Fabrizio,P. et al.** (1997) Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc.Natl.Acad.Sci.U.S.A* **94**, 385-390.
441. **Neubauer,G., King,A., Rappsilber,J. et al.** (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat.Genet.* **20**, 46-50.
442. **Neugebauer,K.M.** (2002) On the importance of being co-transcriptional. *J.Cell Sci.* **115**, 3865-3871.
443. **Neville,M., Stutz,F., Lee,L. et al.** (1997) The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr.Biol.* **7**, 767-775.
444. **Ngo,J.C., Chakrabarti,S., Ding,J.H. et al.** (2005) Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Mol.Cell* **20**, 77-89.
445. **Nichols,R.C., Wang,X.W., Tang,J. et al.** (2000) The RGG domain in hnRNP A2 affects subcellular localization. *Exp.Cell Res.* **256**, 522-532.

## References

---

446. **Nogues,G., Kadener,S., Cramer,P. et al.** (2002) Transcriptional activators differ in their abilities to control alternative splicing. *J.Biol.Chem.* **277**, 43110-43114.
447. **O'Reilly,M.M., McNally,M.T. and Beemon,K.L.** (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.
448. **Oberstrass,F.C., Auweter,S.D., Erat,M. et al.** (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054-2057.
449. **Ohno,G., Hagiwara,M. and Kuroyanagi,H.** (2008) STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo. *Genes Dev.* **22**, 360-374.
450. **Olsen,H.S., Nelbock,P., Cochrane,A.W. et al.** (1990) Secondary structure is the major determinant for interaction of HIV rev protein with RNA. *Science* **247**, 845-848.
451. **Olsen,J.V., Blagoev,B., Gnad,F. et al.** (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648.
452. **Olson,M.O. and Dundr,M.** (2005) The moving parts of the nucleolus. *Histochem.Cell Biol.* **123**, 203-216.
453. **Orengo,J.P. and Cooper,T.A.** (2007) Alternative splicing in disease. *Adv.Exp.Med.Biol.* **623**, 212-223.
454. **Osheim,Y.N., Miller,O.L., Jr. and Beyer,A.L.** (1985) RNP particles at splice junction sequences on Drosophila chorion transcripts. *Cell* **43**, 143-151.
455. **Ostareck-Lederer,A., Ostareck,D.H., Rucknagel,K.P. et al.** (2006) Asymmetric arginine dimethylation of heterogeneous nuclear ribonucleoprotein K by protein-arginine methyltransferase 1 inhibits its interaction with c-Src. *J.Biol.Chem.* **281**, 11115-11125.
456. **Otake,Y., Soundararajan,S., Sengupta,T.K. et al.** (2007) Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA. *Blood* **109**, 3069-3075.
457. **Otte, M.** Heinrich-Heine-Universität Düsseldorf (2006), PhD thesis, Identifizierung von *cis*-wirkenden Sequenzen in den alternativen HIV-1 Leaderexons und ihre funktionelle Bedeutung für die Spleißregulation. 1-178.
458. **Oubridge,C., Ito,N., Evans,P.R. et al.** (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-438.
459. **Pagani,F. and Baralle,F.E.** (2004) Genomic variants in exons and introns: identifying the splicing spoilers. *Nat.Rev.Genet.* **5**, 389-396.
460. **Palella,F.J., Jr., Delaney,K.M., Moorman,A.C. et al.** (1998) Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N.Engl.J.Med.* **338**, 853-860.
461. **Pandit,S., Wang,D. and Fu,X.D.** (2008) Functional integration of transcriptional and RNA processing machineries. *Curr.Opin.Cell Biol.* **20**, 260-265.
462. **Pandya-Jones,A. and Black,D.L.** (2009) Co-transcriptional splicing of constitutive and alternative exons. *RNA.* **15**, 1896-1908.
463. **Park,J.W., Parisky,K., Celotto,A.M. et al.** (2004) Identification of alternative splicing regulators by RNA interference in Drosophila. *Proc.Natl.Acad.Sci.U.S.A* **101**, 15974-15979.

- 
464. **Parker,R. and Song,H.** (2004) The enzymes and control of eukaryotic mRNA turnover. *Nat.Struct.Mol.Biol.* **11**, 121-127.
465. **Patel,N.A., Kaneko,S., Apostolatos,H.S. et al.** (2005) Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase Cbeta11 alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40. *J.Biol.Chem.* **280**, 14302-14309.
466. **Paul,S., Dansithong,W., Kim,D. et al.** (2006) Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. *EMBO J.* **25**, 4271-4283.
467. **Pederson,T.** (1998) The plurifunctional nucleolus. *Nucleic Acids Res.* **26**, 3871-3876.
468. **Peled-Zehavi,H., Berglund,J.A., Rosbash,M. et al.** (2001) Recognition of RNA branch point sequences by the KH domain of splicing factor 1 (mammalian branch point binding protein) in a splicing factor complex. *Mol.Cell Biol.* **21**, 5232-5241.
469. **Penalva,L.O., Lallena,M.J. and Valcarcel,J.** (2001) Switch in 3' splice site recognition between exon definition and splicing catalysis is important for sex-lethal autoregulation. *Mol.Cell Biol.* **21**, 1986-1996.
470. **Perez,I., Lin,C.H., McAfee,J.G. et al.** (1997) Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. *RNA.* **3**, 764-778.
471. **Perkins,A., Cochrane,A.W., Ruben,S.M. et al.** (1989) Structural and functional characterization of the human immunodeficiency virus rev protein. *J.Acquir.Immune.Defic.Syindr.* **2**, 256-263.
472. **Philipps,D., Celotto,A.M., Wang,Q.Q. et al.** (2003) Arginine/serine repeats are sufficient to constitute a splicing activation domain. *Nucleic Acids Res.* **31**, 6502-6508.
473. **Pollard,V.W. and Malim,M.H.** (1998) The HIV-1 Rev protein. *Annu.Rev.Microbiol.* **52**, 491-532.
474. **Polydorides,A.D., Okano,H.J., Yang,Y.Y. et al.** (2000) A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc.Natl.Acad.Sci.U.S.A* **97**, 6350-6355.
475. **Pomerantz,R.J., Trono,D., Feinberg,M.B. et al.** (1990) Cells nonproductively infected with HIV-1 exhibit an aberrant pattern of viral RNA expression: a molecular model for latency. *Cell* **61**, 1271-1276.
476. **Pomeranz Krummel,D.A., Oubridge,C., Leung,A.K. et al.** (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475-480.
477. **Pond,S.J., Ridgeway,W.K., Robertson,R. et al.** (2009) HIV-1 Rev protein assembles on viral RNA one molecule at a time. *Proc.Natl.Acad.Sci.U.S.A* **106**, 1404-1408.
478. **Potashkin,J., Naik,K. and Wentz-Hunter,K.** (1993) U2AF homolog required for splicing in vivo. *Science* **262**, 573-575.
479. **Prasad,J., Colwill,K., Pawson,T. et al.** (1999) The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Mol.Cell Biol.* **19**, 6991-7000.
480. **Proudfoot,N.J., Furger,A. and Dye,M.J.** (2002) Integrating mRNA processing with transcription. *Cell* **108**, 501-512.
481. **Purcell,D.F. and Martin,M.A.** (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J.Virol.* **67**, 6365-6378.
482. **Qi,J., Su,S. and Mattox,W.** (2007) The doublesex splicing enhancer components Tra2 and Rbp1 also repress splicing through an intronic silencer. *Mol.Cell Biol.* **27**, 699-708.

## References

---

483. **Query,C.C., Moore,M.J. and Sharp,P.A.** (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**, 587-597.
484. **Rain,J.C., Rafi,Z., Rhani,Z. et al.** (1998) Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA.* **4**, 551-565.
485. **Ramchatesingh,J., Zahler,A.M., Neugebauer,K.M. et al.** (1995) A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol.Cell Biol.* **15**, 4898-4907.
486. **Ratner,L., Haseltine,W., Patarca,R. et al.** (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **313**, 277-284.
487. **Reed,R.** (1989) The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**, 2113-2123.
488. **Reed,R.** (1990) Protein composition of mammalian spliceosomes assembled in vitro. *Proc.Natl.Acad.Sci.U.S.A* **87**, 8031-8035.
489. **Reed,R.** (2000) Mechanisms of fidelity in pre-mRNA splicing. *Curr.Opin.Cell Biol.* **12**, 340-345.
490. **Reed,R. and Hurt,E.** (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108**, 523-531.
491. **Renkema,G.H. and Saksela,K.** (2000) Interactions of HIV-1 NEF with cellular signal transducing proteins. *Front Biosci.* **5**, D268-D283.
492. **Reks-Ngarm,S., Pitisuttithum,P., Nitayaphan,S. et al.** (2009) Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. *N.Engl.J.Med.*
493. **Riggs,N.L., Little,S.J., Richman,D.D. et al.** (1994) Biological importance and cooperativity of HIV-1 regulatory gene splice acceptors. *Virology* **202**, 264-271.
494. **Robberson,B.L., Cote,G.J. and Berget,S.M.** (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol.Cell Biol.* **10**, 84-94.
495. **Robert-Guroff,M., Popovic,M., Gartner,S. et al.** (1990) Structure and expression of tat-, rev-, and nef-specific transcripts of human immunodeficiency virus type 1 in infected lymphocytes and macrophages. *J.Virol.* **64**, 3391-3398.
496. **Romani,B., Engelbrecht,S. and Glashoff,R.H.** (2010) Functions of Tat: the versatile protein of human immunodeficiency virus type 1. *J.Gen.Virol.* **91**, 1-12.
497. **Romano,M., Marcucci,R., Buratti,E. et al.** (2002) Regulation of 3' splice site selection in the 844ins68 polymorphism of the cystathionine Beta -synthase gene. *J.Biol.Chem.* **277**, 43821-43829.
498. **Roscigno,R.F. and Garcia-Blanco,M.A.** (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA.* **1**, 692-706.
499. **Rosen,C.A., Terwilliger,E., Dayton,A. et al.** (1988) Intragenic cis-acting art gene-responsive sequences of the human immunodeficiency virus. *Proc.Natl.Acad.Sci.U.S.A* **85**, 2071-2075.
500. **Rosin, S.** Heinrich-Heine-Universität Düsseldorf (2008), diploma thesis, Die Rekrutierung spleißregulatorischer Proteindomänen in die Nähe von HIV-1 Spleißstellen und ihre Wirkung auf die Spleißstellennutzung. 1-88.
501. **Rosonina,E., Ip,J.Y., Calarco,J.A. et al.** (2005) Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. *Mol.Cell Biol.* **25**, 6734-6746.

- 
502. **Rossi,F., Labourier,E., Forne,T. et al.** (1996) Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature* **381**, 80-82.
503. **Roth,M.B., Murphy,C. and Gall,J.G.** (1990) A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J.Cell Biol.* **111**, 2217-2223.
504. **Roth,M.B., Zahler,A.M. and Stolk,J.A.** (1991) A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J.Cell Biol.* **115**, 587-596.
505. **Rothrock,C.R., House,A.E. and Lynch,K.W.** (2005) HnRNP L represses exon splicing via a regulated exonic splicing silencer. *EMBO J.* **24**, 2792-2802.
506. **Rudner,D.Z., Breger,K.S., Kanaar,R. et al.** (1998) RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high-affinity binding. *Mol.Cell Biol.* **18**, 4004-4011.
507. **Rudner,D.Z., Kanaar,R., Breger,K.S. et al.** (1996) Mutations in the small subunit of the Drosophila U2AF splicing factor cause lethality and developmental defects. *Proc.Natl.Acad.Sci.U.S.A* **93**, 10333-10337.
508. **Rudner,D.Z., Kanaar,R., Breger,K.S. et al.** (1998) Interaction between subunits of heterodimeric splicing factor U2AF is essential in vivo. *Mol.Cell Biol.* **18**, 1765-1773.
509. **Ruskin,B., Zamore,P.D. and Green,M.R.** (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**, 207-219.
510. **Rutz,B. and Seraphin,B.** (1999) Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA.* **5**, 819-831.
511. **Sakaguchi,K., Zambrano,N., Baldwin,E.T. et al.** (1993) Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc.Natl.Acad.Sci.U.S.A* **90**, 5219-5223.
512. **Sambrook,J., Fritsch,E.F. and Maniatis,T.** (1989) *Molecular cloning. A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
513. **Sanford,J.R. and Bruzik,J.P.** (1999) Developmental regulation of SR protein phosphorylation and activity. *Genes Dev.* **13**, 1513-1518.
514. **Sanford,J.R., Coutinho,P., Hackett,J.A. et al.** (2008) Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS.ONE.* **3**, e3369.
515. **Sanger,F., Nicklen,S. and Coulson,A.R.** (1977) DNA sequencing with chain-terminating inhibitors. *Proc.Natl.Acad.Sci.U.S.A* **74**, 5463-5467.
516. **Sarkar,I., Hauber,I., Hauber,J. et al.** (2007) HIV-1 proviral DNA excision using an evolved recombinase. *Science* **316**, 1912-1915.
517. **Sauliere,J., Sureau,A., Expert-Bezancon,A. et al.** (2006) The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol.Cell Biol.* **26**, 8755-8769.
518. **Scaffidi,P. and Misteli,T.** (2005) Reversal of the cellular phenotype in the premature aging disease Hutchinson-Gilford progeria syndrome. *Nat.Med.* **11**, 440-445.
519. **Schaal,H., Pfeiffer,P., Klein,M. et al.** (1993) Use of DNA end joining activity of a *Xenopus laevis* egg extract for construction of deletions and expression vectors for HIV-1 Tat and Rev proteins. *Gene* **124**, 275-280.

## References

---

520. **Schaal,T.D. and Maniatis,T.** (1999) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol.Cell Biol.* **19**, 261-273.
521. **Schaub,M.C., Lopez,S.R. and Caputi,M.** (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. *J.Biol.Chem.* **282**, 13617-13626.
522. **Schellenberg,M.J., Ritchie,D.B. and MacMillan,A.M.** (2008) Pre-mRNA splicing: a complex picture in higher definition. *Trends Biochem.Sci.* **33**, 243-246.
523. **Schmid,M. and Jensen,T.H.** (2008) The exosome: a multipurpose RNA-decay machine. *Trends Biochem.Sci.* **33**, 501-510.
524. **Schneider,R., Campbell,M., Nasioulas,G. et al.** (1997) Inactivation of the human immunodeficiency virus type 1 inhibitory elements allows Rev-independent expression of Gag and Gag/protease and particle formation. *J.Virol.* **71**, 4892-4903.
525. **Schutz,S., Chemnitz,J., Spillner,C. et al.** (2006) Stimulated expression of mRNAs in activated T cells depends on a functional CRM1 nuclear export pathway. *J.Mol.Biol.* **358**, 997-1009.
526. **Schwartz,S., Campbell,M., Nasioulas,G. et al.** (1992) Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. *J.Virol.* **66**, 7176-7182.
527. **Schwartz,S., Felber,B.K., Benko,D.M. et al.** (1990) Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J.Virol.* **64**, 2519-2529.
528. **Schwartz,S., Felber,B.K., Fenyo,E.M. et al.** (1990) Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J.Virol.* **64**, 5448-5456.
529. **Schwartz,S., Felber,B.K. and Pavlakis,G.N.** (1991) Expression of human immunodeficiency virus type 1 vif and vpr mRNAs is Rev-dependent and regulated by splicing. *Virology* **183**, 677-686.
530. **Schwartz,S., Felber,B.K. and Pavlakis,G.N.** (1992) Distinct RNA sequences in the gag region of human immunodeficiency virus type 1 decrease RNA stability and inhibit expression in the absence of Rev protein. *J.Virol.* **66**, 150-159.
531. **Schwartz,S., Felber,B.K. and Pavlakis,G.N.** (1992) Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. *Mol.Cell Biol.* **12**, 207-219.
532. **Schwartzberg,P., Colicelli,J. and Goff,S.P.** (1984) Construction and analysis of deletion mutations in the pol gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell* **37**, 1043-1052.
533. **Schwerk,C. and Schulze-Osthoff,K.** (2005) Regulation of apoptosis by alternative pre-mRNA splicing. *Mol.Cell* **19**, 1-13.
534. **Sciabica,K.S., Dai,Q.J. and Sandri-Goldin,R.M.** (2003) ICP27 interacts with SRPK1 to mediate HSV splicing inhibition by altering SR protein phosphorylation. *EMBO J.* **22**, 1608-1619.
535. **Segade,F., Hurlle,B., Claudio,E. et al.** (1996) Molecular cloning of a mouse homologue for the Drosophila splicing regulator Tra2. *FEBS Lett.* **387**, 152-156.
536. **Selden,R.F., Howie,K.B., Rowe,M.E. et al.** (1986) Human growth hormone as a reporter gene in regulation studies employing transient gene expression. *Mol.Cell Biol.* **6**, 3173-3179.
537. **Selenko,P., Gregorovic,G., Sprangers,R. et al.** (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol.Cell* **11**, 965-976.

- 
538. **Sengupta,T.K., Bandyopadhyay,S., Fernandes,D.J. et al.** (2004) Identification of nucleolin as an AU-rich element binding protein involved in bcl-2 mRNA stabilization. *J.Biol.Chem.* **279**, 10855-10863.
539. **Seraphin,B. and Rosbash,M.** (1989) Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**, 349-358.
540. **Sharma,S., Falick,A.M. and Black,D.L.** (2005) Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol.Cell* **19**, 485-496.
541. **Sharma,S., Kohlstaedt,L.A., Damianov,A. et al.** (2008) Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat.Struct.Mol.Biol.* **15**, 183-191.
542. **Shaw,S.D., Chakrabarti,S., Ghosh,G. et al.** (2007) Deletion of the N-terminus of SF2/ASF permits RS-domain-independent pre-mRNA splicing. *PLoS.ONE.* **2**, e854.
543. **Shen,H.** (2009) UAP56- a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export. *BMB.Rep.* **42**, 185-188.
544. **Shen,H. and Green,M.R.** (2004) A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol.Cell* **16**, 363-373.
545. **Shen,H. and Green,M.R.** (2006) RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev.* **20**, 1755-1765.
546. **Shen,H. and Green,M.R.** (2007) RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nat.Struct.Mol.Biol.* **14**, 597-603.
547. **Shen,H., Kan,J.L. and Green,M.R.** (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol.Cell* **13**, 367-376.
548. **Shen,H., Zheng,X., Shen,J. et al.** (2008) Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. *Genes Dev.* **22**, 1796-1803.
549. **Shepard,P.J. and Hertel,K.J.** (2008) Conserved RNA secondary structures promote alternative splicing. *RNA.* **14**, 1463-1469.
550. **Shepard,P.J. and Hertel,K.J.** (2009) The SR protein family. *Genome Biol.* **10**, 242.
551. **Shi,H., Hoffman,B.E. and Lis,J.T.** (1997) A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52. *Mol.Cell Biol.* **17**, 2649-2657.
552. **Shi,Y., Reddy,B. and Manley,J.L.** (2006) PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. *Mol.Cell* **23**, 819-829.
553. **Shin,C. and Manley,J.L.** (2004) Cell signalling and the control of pre-mRNA splicing. *Nat.Rev.Mol.Cell Biol.* **5**, 727-738.
554. **Shukla,S., Gatto-Konczak,F., Breathnach,R. et al.** (2005) Competition of PTB with TIA proteins for binding to a U-rich cis-element determines tissue-specific splicing of the myosin phosphatase targeting subunit 1. *RNA.* **11**, 1725-1736.
555. **Si,Z., Amendt,B.A. and Stoltzfus,C.M.** (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 Res.* **25**, 861-867.
556. **Si,Z.H., Rauch,D. and Stoltzfus,C.M.** (1998) The exon splicing silencer in human immunodeficiency virus type 1 Tat exon 3 is bipartite and acts early in spliceosome assembly. *Mol.Cell Biol.* **18**, 5404-5413.

## References

---

557. **Sickmier,E.A., Frato,K.E., Shen,H. et al.** (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol.Cell* **23**, 49-59.
558. **Siegal,F.P., Lopez,C., Hammer,G.S. et al.** (1981) Severe acquired immunodeficiency in male homosexuals, manifested by chronic perianal ulcerative herpes simplex lesions. *N.Engl.J.Med.* **305**, 1439-1444.
559. **Simpson,P.J., Monie,T.P., Szendroi,A. et al.** (2004) Structure and RNA interactions of the N-terminal RRM domains of PTB. *Structure.* **12**, 1631-1643.
560. **Sims,R.J., III, Millhouse,S., Chen,C.F. et al.** (2007) Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol.Cell* **28**, 665-676.
561. **Singh,N.N., Shishimorova,M., Cao,L.C. et al.** (2009) A short antisense oligonucleotide masking a unique intronic motif prevents skipping of a critical exon in spinal muscular atrophy. *RNA.Biol.* **6**, 341-350.
562. **Singh,R., Valcarcel,J. and Green,M.R.** (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173-1176.
563. **Smith,C.W., Chu,T.T. and Nadal-Ginard,B.** (1993) Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol.Cell Biol.* **13**, 4939-4952.
564. **Smith,C.W. and Valcarcel,J.** (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem.Sci.* **25**, 381-388.
565. **Smith,D.J., Query,C.C. and Konarska,M.M.** (2008) "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol.Cell* **30**, 657-666.
566. **Smith,P.J., Zhang,C., Wang,J. et al.** (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum.Mol.Genet.* **15**, 2490-2508.
567. **Soares,L.M., Zanier,K., Mackereth,C. et al.** (2006) Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. *Science* **312**, 1961-1965.
568. **Sodora,D.L., Allan,J.S., Apetrei,C. et al.** (2009) Toward an AIDS vaccine: lessons from natural simian immunodeficiency virus infections of African nonhuman primate hosts. *Nat.Med.* **15**, 861-865.
569. **Solis,A.S., Shariat,N. and Patton,J.G.** (2008) Splicing fidelity, enhancers, and disease. *Front Biosci.* **13**, 1926-1942.
570. **Sonza,S., Mutimer,H.P., O'Brien,K. et al.** (2002) Selectively reduced tat mRNA heralds the decline in productive human immunodeficiency virus type 1 infection in monocyte-derived macrophages. *J.Virol.* **76**, 12611-12621.
571. **Sorek,R., Shamir,R. and Ast,G.** (2004) How prevalent is functional alternative splicing in the human genome? *Trends Genet.* **20**, 68-71.
572. **Soret,J., Bakkour,N., Maire,S. et al.** (2005) Selective modification of alternative splicing by indole derivatives that target serine-arginine-rich protein splicing factors. *Proc.Natl.Acad.Sci.U.S.A* **102**, 8764-8769.
573. **Soret,J., Gabut,M. and Tazi,J.** (2006) SR proteins as potential targets for therapy. *Prog.Mol.Subcell.Biol.* **44**, 65-87.
574. **Soret,J. and Tazi,J.** (2003) Phosphorylation-dependent control of the pre-mRNA splicing machinery. *Prog.Mol.Subcell.Biol.* **31**, 89-126.

- 
575. **Spellman,R. and Smith,C.W.** (2006) Novel modes of splicing repression by PTB. *Trends Biochem.Sci.* **31**, 73-76.
576. **Spingola,M., Grate,L., Haussler,D. et al.** (1999) Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA.* **5**, 221-234.
577. **Stade,K., Ford,C.S., Guthrie,C. et al.** (1997) Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**, 1041-1050.
578. **Staffa,A. and Cochrane,A.** (1994) The tat/rev intron of human immunodeficiency virus type 1 is inefficiently spliced because of suboptimal signals in the 3' splice site. *J.Virol.* **68**, 3071-3079.
579. **Staffa,A. and Cochrane,A.** (1995) Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. *Mol.Cell Biol.* **15**, 4597-4605.
580. **Staley,J.P. and Guthrie,C.** (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-326.
581. **Stamm,S.** (2008) Regulation of alternative splicing by reversible protein phosphorylation. *J.Biol.Chem.* **283**, 1223-1227.
582. **Stark,H., Dube,P., Luhrmann,R. et al.** (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539-542.
583. **Steck, S.** Universität Regensburg (2003), PhD thesis, Bedeutung des Kodongebrauchs für die zeitlich regulierte Genexpression im Rahmen der HIV-Replikation: Ansatzpunkte für die Entwicklung attenuierter Lebendimpfstoffe? 1-144.
584. **Stevens,S.W., Ryan,D.E., Ge,H.Y. et al.** (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol.Cell* **9**, 31-44.
585. **Stevenson,M., Haggerty,S., Lamonica,C.A. et al.** (1990) Integration is not necessary for expression of human immunodeficiency virus type 1 protein products. *J.Virol.* **64**, 2421-2425.
586. **Stoilov,P., Daoud,R., Nayler,O. et al.** (2004) Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum.Mol.Genet.* **13**, 509-524.
587. **Stoltzfus,C.M.** (2009) Regulation of HIV-1 Alternative RNA Splicing and Its Role in Virus Replication. In Maramorosch,K., Shatkin,A.J. and Murphy,F.A. (eds.), *Advances in Virus Research.* Academic Press, Burlington, Vol. 74, pp. 1-40.
588. **Stoltzfus,C.M. and Madsen,J.M.** (2006) Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. *Curr.HIV.Res.* **4**, 43-55.
589. **Strasser,K. and Hurt,E.** (2000) Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J.* **19**, 410-420.
590. **Strasser,K. and Hurt,E.** (2001) Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* **413**, 648-652.
591. **Strebel,K., Daugherty,D., Clouse,K. et al.** (1987) The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* **328**, 728-730.
592. **Stutz,F., Bachi,A., Doerks,T. et al.** (2000) REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA.* **6**, 638-650.
593. **Suhasini,M. and Reddy,T.R.** (2009) Cellular proteins and HIV-1 Rev function. *Curr.HIV.Res.* **7**, 91-100.

## References

---

594. **Swanson,A.K. and Stoltzfus,C.M.** (1998) Overlapping cis sites used for splicing of HIV-1 env/nef and rev mRNAs. *J.Biol.Chem.* **273**, 34551-34557.
595. **Tacke,R., Chen,Y. and Manley,J.L.** (1997) Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc.Natl.Acad.Sci.U.S.A* **94**, 1148-1153.
596. **Tacke,R. and Manley,J.L.** (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* **14**, 3540-3551.
597. **Tacke,R. and Manley,J.L.** (1999) Determinants of SR protein specificity. *Curr.Opin.Cell Biol.* **11**, 358-362.
598. **Tacke,R. and Manley,J.L.** (1999) Functions of SR and Tra2 proteins in pre-mRNA splicing regulation. *Proc.Soc.Exp.Biol.Med.* **220**, 59-63.
599. **Tacke,R., Tohyama,M., Ogawa,S. et al.** (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* **93**, 139-148.
600. **Talerico,M. and Berget,S.M.** (1990) Effect of 5' splice site mutations on splicing of the preceding intron. *Mol.Cell Biol.* **10**, 6299-6305.
601. **Talerico,M. and Berget,S.M.** (1994) Intron definition in splicing of small Drosophila introns. *Mol.Cell Biol.* **14**, 3434-3445.
602. **Tanackovic,G. and Kramer,A.** (2005) Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo. *Mol.Biol.Cell* **16**, 1366-1377.
603. **Tange,T.O., Damgaard,C.K., Guth,S. et al.** (2001) The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J.* **20**, 5748-5758.
604. **Tange,T.O. and Kjems,J.** (2001) SF2/ASF binds to a splicing enhancer in the third HIV-1 tat exon and stimulates U2AF binding independently of the RS domain. *J.Mol.Biol.* **312**, 649-662.
605. **Tange,T.O., Shibuya,T., Jurica,M.S. et al.** (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA.* **11**, 1869-1883.
606. **Tazi,J., Bakkour,N. and Stamm,S.** (2009) Alternative splicing and disease. *Biochim.Biophys.Acta* **1792**, 14-26.
607. **Tazi,J., Durand,S. and Jeanteur,P.** (2005) The spliceosome: a novel multi-faceted target for therapy. *Trends Biochem.Sci.* **30**, 469-478.
608. **The C.elegans Sequencing Consortium** (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012-2018.
609. **Thickman,K.R., Swenson,M.C., Kabogo,J.M. et al.** (2006) Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. *J.Mol.Biol.* **356**, 664-683.
610. **Tian,H. and Kole,R.** (1995) Selection of novel exon recognition elements from a pool of random sequences. *Mol.Cell Biol.* **15**, 6291-6298.
611. **Tian,M. and Maniatis,T.** (1992) Positive control of pre-mRNA splicing in vitro. *Science* **256**, 237-240.
612. **Tian,M. and Maniatis,T.** (1993) A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell* **74**, 105-114.

613. **Tiley,L.S., Malim,M.H., Tewary,H.K. et al.** (1992) Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein. *Proc.Natl.Acad.Sci.U.S.A* **89**, 758-762.
614. **Tintaru,A.M., Hautbergue,G.M., Hounslow,A.M. et al.** (2007) Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep.* **8**, 756-762.
615. **Tran,H., Schilling,M., Wirbelauer,C. et al.** (2004) Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol.Cell* **13**, 101-111.
616. **UNAIDS.** 2008 Report on the global AIDS epidemic. [http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/2008/2008\\_Global\\_report.asp](http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/2008/2008_Global_report.asp), 1-360. 2008.
617. **UNAIDS.** AIDS Epidemic Update 2009. <http://www.unaids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2009/default.asp>, 1-100. 2009.
618. **Valcarcel,J., Gaur,R.K., Singh,R. et al.** (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* **273**, 1706-1709.
619. **Valcarcel,J., Singh,R., Zamore,P.D. et al.** (1993) The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* **362**, 171-175.
620. **van Der Houven Van Oordt, Diaz-Meco,M.T., Lozano,J. et al.** (2000) The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J.Cell Biol.* **149**, 307-316.
621. **van Der Houven Van Oordt, Newton,K., Screaton,G.R. et al.** (2000) Role of SR protein modular domains in alternative splicing specificity in vivo. *Nucleic Acids Res.* **28**, 4822-4831.
622. **Varani,G. and Nagai,K.** (1998) RNA recognition by RNP proteins during RNA processing. *Annu.Rev.Biophys.Biomol.Struct.* **27**, 407-445.
623. **Varbanov,M., Espert,L. and Biard-Piechaczyk,M.** (2006) Mechanisms of CD4 T-cell depletion triggered by HIV-1 viral proteins. *AIDS Rev.* **8**, 221-236.
624. **Vassileva,M.T. and Matunis,M.J.** (2004) SUMO modification of heterogeneous nuclear ribonucleoproteins. *Mol.Cell Biol.* **24**, 3623-3632.
625. **Vela,E., Hilari,J.M., Roca,X. et al.** (2007) Multisite and bidirectional exonic splicing enhancer in CD44 alternative exon v3. *RNA.* **13**, 2312-2323.
626. **Venkatesh,L.K., Mohammed,S. and Chinnadurai,G.** (1990) Functional domains of the HIV-1 rev gene required for trans-regulation and subcellular localization. *Virology* **176**, 39-47.
627. **Wagner,E.J. and Garcia-Blanco,M.A.** (2001) Polypyrimidine tract binding protein antagonizes exon definition. *Mol.Cell Biol.* **21**, 3281-3288.
628. **Wahl,M.C., Will,C.L. and Luhrmann,R.** (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701-718.
629. **Wain-Hobson,S., Sonigo,P., Danos,O. et al.** (1985) Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**, 9-17.
630. **Wang,E. and Cambi,F.** (2009) Heterogeneous nuclear ribonucleoproteins H and F regulate the proteolipid protein/DM20 ratio by recruiting U1 small nuclear ribonucleoprotein through a complex array of G runs. *J.Biol.Chem.* **284**, 11194-11204.
631. **Wang,E.T., Sandberg,R., Luo,S. et al.** (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476.

## References

---

632. **Wang,G.S. and Cooper,T.A.** (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat.Rev.Genet.* **8**, 749-761.
633. **Wang,H.Y., Lin,W., Dyck,J.A. et al.** (1998) SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J.Cell Biol.* **140**, 737-750.
634. **Wang,H.Y., Xu,X., Ding,J.H. et al.** (2001) SC35 plays a role in T cell development and alternative splicing of CD45. *Mol.Cell* **7**, 331-342.
635. **Wang,X., Bruderer,S., Rafi,Z. et al.** (1999) Phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly. *EMBO J.* **18**, 4549-4559.
636. **Wang,X., Wang,K., Radovich,M. et al.** (2009) Genome-wide prediction of cis-acting RNA elements regulating tissue-specific pre-mRNA alternative splicing. *BMC.Genomics* **10 Suppl 1**, S4.
637. **Wang,Z. and Burge,C.B.** (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA.* **14**, 802-813.
638. **Wang,Z., Hoffmann,H.M. and Grabowski,P.J.** (1995) Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. *RNA.* **1**, 21-35.
639. **Wang,Z., Rolish,M.E., Yeo,G. et al.** (2004) Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831-845.
640. **Warf,M.B., Diegel,J.V., von Hippel,P.H. et al.** (2009) The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc.Natl.Acad.Sci.U.S.A.*
641. **Warzecha,C.C., Sato,T.K., Nabet,B. et al.** (2009) ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol.Cell* **33**, 591-601.
642. **Waterston,R.H., Lindblad-Toh,K., Birney,E. et al.** (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562.
643. **Watts,J.M., Dang,K.K., Gorelick,R.J. et al.** (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716.
644. **Webb,C.J. and Wise,J.A.** (2004) The splicing factor U2AF small subunit is functionally conserved between fission yeast and humans. *Mol.Cell Biol.* **24**, 4229-4240.
645. **Weiss,R., Teich,N. and Varmus,H.** (1985) *RNA tumor viruses. Molecular biology of tumor viruses.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
646. **Wentz,M.P., Moore,B.E., Cloyd,M.W. et al.** (1997) A naturally arising mutation of a potential silencer of exon splicing in human immunodeficiency virus type 1 induces dominant aberrant splicing and arrests virus production. *J.Virol.* **71**, 8542-8551.
647. **Wetterberg,I., Bauren,G. and Wieslander,L.** (1996) The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. *RNA.* **2**, 641-651.
648. **Wilkinson,K.A., Gorelick,R.J., Vasa,S.M. et al.** (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS.Biol.* **6**, e96.
649. **Will,C.L. and Luhrmann,R.** (2006) Spliceosome Structure and Function. In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds.), *The RNA World.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, Vol. 43, pp. 369-400.

650. **Will,C.L., Rumpfer,S., Klein,G.J. et al.** (1996) In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.* **24**, 4614-4623.
651. **Windgassen,M. and Krebber,H.** (2003) Identification of Gbp2 as a novel poly(A)<sup>+</sup> RNA-binding protein involved in the cytoplasmic delivery of messenger RNAs in yeast. *EMBO Rep.* **4**, 278-283.
652. **Wiskerchen,M. and Muesing,M.A.** (1995) Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. *J.Virol.* **69**, 376-386.
653. **Wodrich,H., Schambach,A. and Krausslich,H.G.** (2000) Multiple copies of the Mason-Pfizer monkey virus constitutive RNA transport element lead to enhanced HIV-1 Gag expression in a context-dependent manner. *Nucleic Acids Res.* **28**, 901-910.
654. **Wolff,B., Sanglier,J.J. and Wang,Y.** (1997) Leptomycin B is an inhibitor of nuclear export: inhibition of nucleo-cytoplasmic translocation of the human immunodeficiency virus type 1 (HIV-1) Rev protein and Rev-dependent mRNA. *Chem.Biol.* **4**, 139-147.
655. **Wu,J.Y. and Maniatis,T.** (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-1070.
656. **Wu,S., Romfo,C.M., Nilsen,T.W. et al.** (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**, 832-835.
657. **Wurtz,T., Kiseleva,E., Nacheva,G. et al.** (1996) Identification of two RNA-binding proteins in Balbiani ring premessenger ribonucleoprotein granules and presence of these proteins in specific subsets of heterogeneous nuclear ribonucleoprotein particles. *Mol.Cell Biol.* **16**, 1425-1435.
658. **Xiao,R., Sun,Y., Ding,J.H. et al.** (2007) Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol.Cell Biol.* **27**, 5393-5402.
659. **Xiao,S.H. and Manley,J.L.** (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.* **11**, 334-344.
660. **Xiao,S.H. and Manley,J.L.** (1998) Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *EMBO J.* **17**, 6359-6367.
661. **Xie,J., Lee,J.A., Kress,T.L. et al.** (2003) Protein kinase A phosphorylation modulates transport of the polypyrimidine tract-binding protein. *Proc.Natl.Acad.Sci.U.S.A* **100**, 8776-8781.
662. **Xing,Y., Johnson,C.V., Dobner,P.R. et al.** (1993) Higher level organization of individual gene transcription and RNA splicing. *Science* **259**, 1326-1330.
663. **Xing,Y., Johnson,C.V., Moen,P.T., Jr. et al.** (1995) Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J.Cell Biol.* **131**, 1635-1647.
664. **Xu,X., Yang,D., Ding,J.H. et al.** (2005) ASF/SF2-regulated CaMKII $\delta$  alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell* **120**, 59-72.
665. **Xu,Y.Z., Newnham,C.M., Kameoka,S. et al.** (2004) Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *EMBO J.* **23**, 376-385.
666. **Yamamoto,H., Tsukahara,K., Kanaoka,Y. et al.** (1999) Isolation of a mammalian homologue of a fission yeast differentiation regulator. *Mol.Cell Biol.* **19**, 3829-3841.
667. **Yang,X., Bani,M.R., Lu,S.J. et al.** (1994) The A1 and A1B proteins of heterogeneous nuclear ribonucleoproteins modulate 5' splice site selection in vivo. *Proc.Natl.Acad.Sci.U.S.A* **91**, 6924-6928.

## References

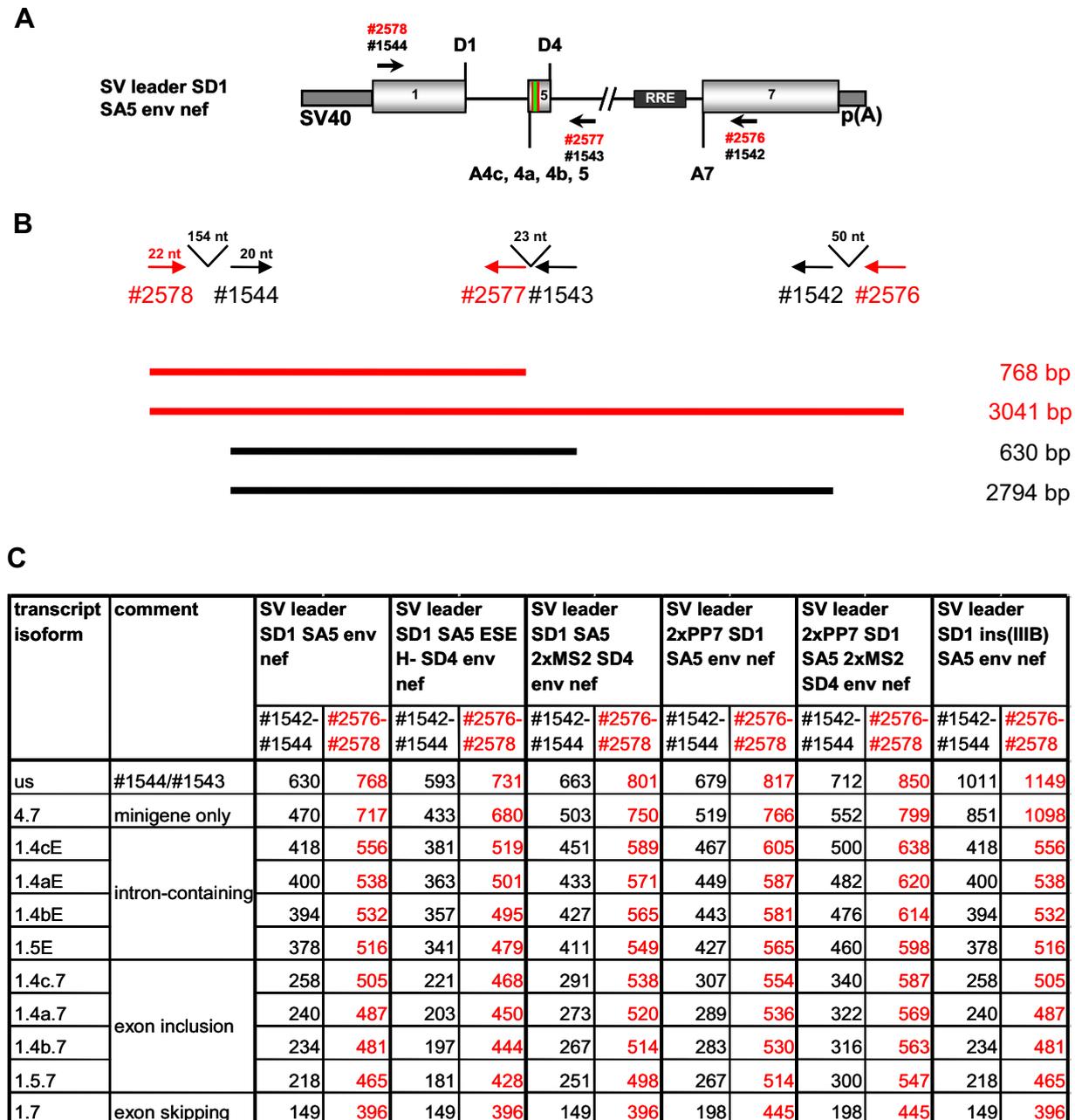
---

668. **Yeakley,J.M., Tronchere,H., Olesen,J. et al.** (1999) Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. *J.Cell Biol.* **145**, 447-455.
669. **Yeo,G. and Burge,C.B.** (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J.Comput.Biol.* **11**, 377-394.
670. **Yeo,G., Hoon,S., Venkatesh,B. et al.** (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc.Natl.Acad.Sci.U.S.A* **101**, 15700-15705.
671. **Yeo,G.W., Coufal,N.G., Liang,T.Y. et al.** (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat.Struct.Mol.Biol.* **16**, 130-137.
672. **Yomoda,J., Muraki,M., Kataoka,N. et al.** (2008) Combination of Clk family kinase and SRp75 modulates alternative splicing of Adenovirus E1A. *Genes Cells* **13**, 233-244.
673. **Yu,E.T., Hawkins,A., Eaton,J. et al.** (2008) MS3D structural elucidation of the HIV-1 packaging signal. *Proc.Natl.Acad.Sci.U.S.A* **105**, 12248-12253.
674. **Zahler,A.M., Damgaard,C.K., Kjems,J. et al.** (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. *J.Biol.Chem.* **279**, 10077-10084.
675. **Zahler,A.M., Lane,W.S., Stolk,J.A. et al.** (1992) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* **6**, 837-847.
676. **Zahler,A.M., Neugebauer,K.M., Lane,W.S. et al.** (1993) Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* **260**, 219-222.
677. **Zamore,P.D. and Green,M.R.** (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc.Natl.Acad.Sci.U.S.A* **86**, 9243-9247.
678. **Zamore,P.D. and Green,M.R.** (1991) Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J.* **10**, 207-214.
679. **Zamore,P.D., Patton,J.G. and Green,M.R.** (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-614.
680. **Zapp,M.L. and Green,M.R.** (1989) Sequence-specific RNA binding by the HIV-1 Rev protein. *Nature* **342**, 714-716.
681. **Zapp,M.L., Hope,T.J., Parslow,T.G. et al.** (1991) Oligomerization and RNA binding domains of the type 1 human immunodeficiency virus Rev protein: a dual function for an arginine-rich binding motif. *Proc.Natl.Acad.Sci.U.S.A* **88**, 7734-7738.
682. **Zhang,C., Zhang,Z., Castle,J. et al.** (2008) Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22**, 2550-2563.
683. **Zhang,D. and Rosbash,M.** (1999) Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev.* **13**, 581-592.
684. **Zhang,G., Zapp,M.L., Yan,G. et al.** (1996) Localization of HIV-1 RNA in mammalian nuclei. *J.Cell Biol.* **135**, 9-18.
685. **Zhang,M., Zamore,P.D., Carmo-Fonseca,M. et al.** (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc.Natl.Acad.Sci.U.S.A* **89**, 8769-8773.
686. **Zhang,W.J. and Wu,J.Y.** (1996) Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol.Cell Biol.* **16**, 5400-5408.

- 
687. **Zhang,X.H. and Chasin,L.A.** (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* **18**, 1241-1250.
688. **Zhang,Z. and Krainer,A.R.** (2007) Splicing remodels messenger ribonucleoprotein architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. *Proc.Natl.Acad.Sci.U.S.A* **104**, 11574-11579.
689. **Zhong,X.Y., Wang,P., Han,J. et al.** (2009) SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol.Cell* **35**, 1-10.
690. **Zhou,H.L. and Lou,H.** (2008) Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. *Mol.Cell Biol.* **28**, 5507-5516.
691. **Zhou,Z., Luo,M.J., Straesser,K. et al.** (2000) The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**, 401-405.
692. **Zhu,J. and Krainer,A.R.** (2000) Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes Dev.* **14**, 3166-3178.
693. **Zhu,J., Mayeda,A. and Krainer,A.R.** (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol.Cell* **8**, 1351-1361.
694. **Zhuang,Y. and Weiner,A.M.** (1986) A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**, 827-835.
695. **Zhuang,Y. and Weiner,A.M.** (1990) The conserved dinucleotide AG of the 3' splice site may be recognized twice during in vitro splicing of mammalian mRNA precursors. *Gene* **90**, 263-269.
696. **Zolotukhin,A.S., Michalowski,D., Bear,J. et al.** (2003) PSF acts through the human immunodeficiency virus type 1 mRNA instability elements to regulate virus expression. *Mol.Cell Biol.* **23**, 6618-6630.
697. **Zorio,D.A. and Blumenthal,T.** (1999) Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* **402**, 835-838.
698. **Zorio,D.A. and Blumenthal,T.** (1999) U2AF35 is encoded by an essential gene clustered in an operon with RRM/cyclophilin in *Caenorhabditis elegans*. *RNA.* **5**, 487-494.
699. **Zuker,M.** (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415.
700. **Zuo,P. and Maniatis,T.** (1996) The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* **10**, 1356-1368.
701. **Zuo,P. and Manley,J.L.** (1993) Functional domains of the human splicing factor ASF/SF2. *EMBO J.* **12**, 4727-4737.
702. **Zuo,P. and Manley,J.L.** (1994) The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. *Proc.Natl.Acad.Sci.U.S.A* **91**, 3363-3367.
703. **Zychlinski,D., Erkelenz,S., Melhorn,V. et al.** (2009) Limited complementarity between U1 snRNA and a retroviral 5' splice site permits its attenuation via RNA secondary structure. *Nucleic Acids Res.* **37**, 7429-7440.

## G. Appendix

### G.1 Experimental data

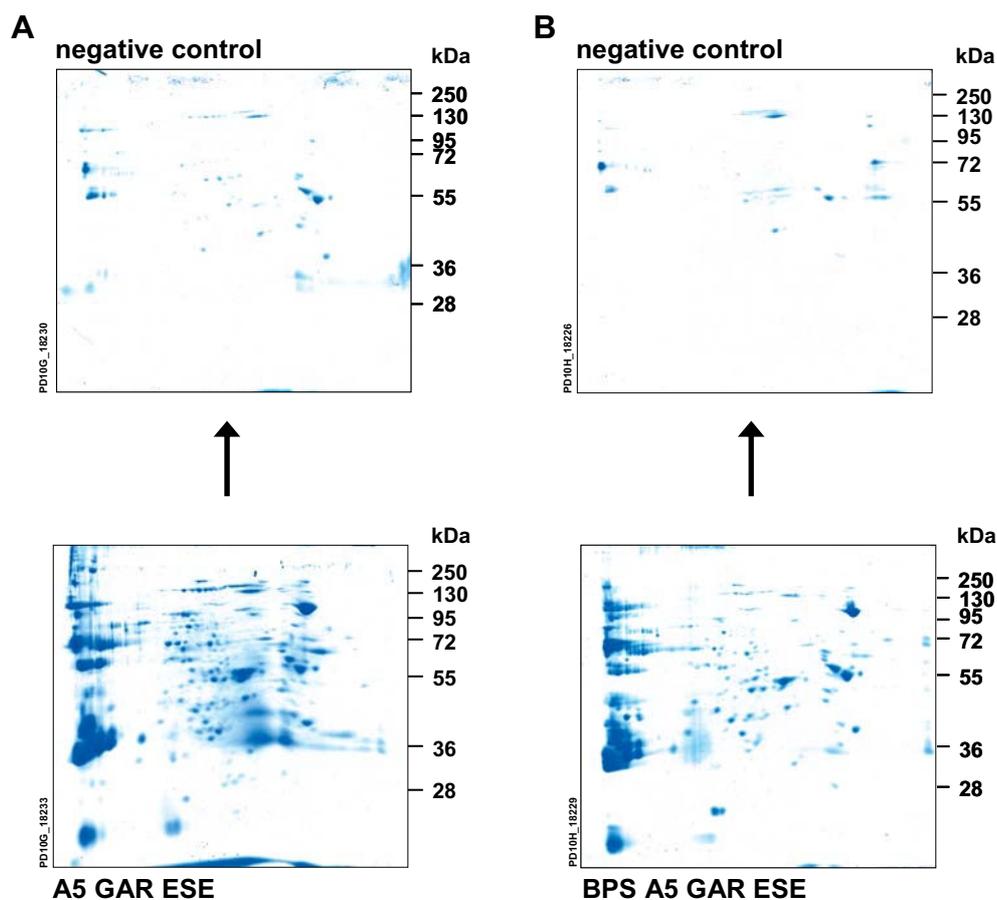


**Fig. VII-1: Primer position and PCR product sizes for 2-intron minigenes with varying exon and intron sizes.**

(A) Primer positions for amplification of alternatively spliced reporter RNA (ASR) expressed from parental and mutant 2-intron minigenes.

(B) PCR product length after amplification of unspliced RNA from 2-intron minigenes using primers #1542-#1544 or #2576-#2578.

(C) PCR product sizes for alternatively spliced reporter RNA applying primers depicted in (B).

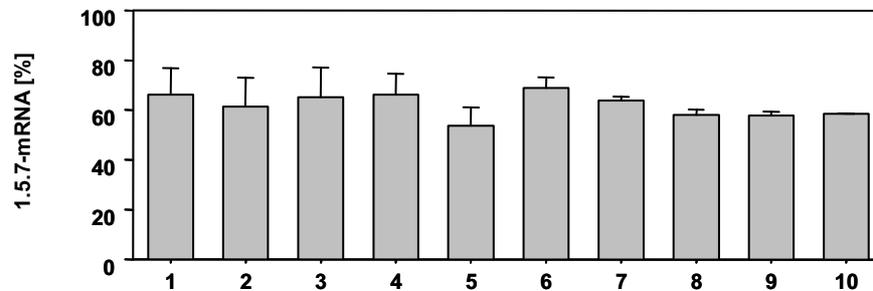
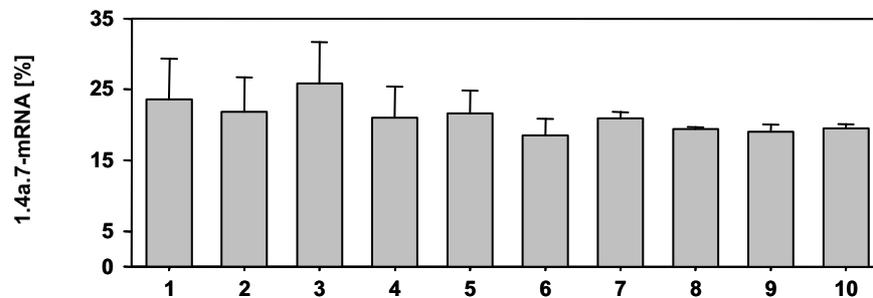


**Fig. VII-2: Negative controls for 2D-gel electrophoresis of proteins isolated from GAR ESE-containing RNA targets.**

During RNA affinity chromatography, control reactions were processed in the absence of short (**A**) or extended (**B**) target RNA displaying several proteins binding to the immobilisation bead matrix. Protein isoforms detected in negative controls were excluded from further analyses.

**A**

	wt	single mutations			double mutations			5'-SR-	all SR-	all SC35-
SC35(1)	+	-	+	+	-	-	+	-	-	-
SC35(2)	+	+	-	+	-	+	-	-	-	-
SRp40	+	+	+	-	+	-	-	-	-	+
SRp55(1)	+	+	+	+	+	+	+	+	-	+
SRp55(2)	+	+	+	+	+	+	+	+	+	+
SC35(3)	+	+	+	+	+	+	+	+	-	-
$\Delta G$ [kcal/Mol]	-9,1	-5,2	-6,8	-8,4	-6,8	-6,0	-8,7	-5,4	-4,2	0,1

**B****C**

**Fig. VII-3: Mutation of SR protein binding sites overlapping D1 does not affect the expression of 1.4a.7- and 1.5.7-mRNA.**

Experimental data showing the splicing pattern of the parental 2-intron minigene and derivate constructs carrying mutations in predicted SR protein sites overlapping D1 as indicated in (A) were analysed regarding the expression of doubly spliced 1.4a.7- (B) and 1.5.7-mRNA isoforms (C).

Tab. VII-1: Published constructs

Thesis	Publication	Reference
SV leader SD1 SA5 env nef	pSV-1-env	Asang et al. 2008
SV leader SD1 SA5 env nef	SV-1-env	Caputi et al. 2004
SV leader SD1 SA5 SF2 1- SF2 2- SRp40- env nef	pSV-1-env SF2- SRp40-	Asang et al. 2008
SV leader SD1 SA5 SF2 1- SF2 2- SRp40- env nef	SV-1-env SF2- SRp40-	Caputi et al. 2004
SV leader SD1 SA5 SF2(1)- SRp40- SF2(2)- env nef	pSV-1-env ESE-	Asang et al. 2008
SV leader SD1 SA5 SRp40- SF2(2)- env nef	pSV-1-env SRp40- SF2(2)-	Asang et al. 2008
SV leader SD1 SA5 SF2(1)- SRp40- env nef	pSV-1-env SF2(1)- SRp40-	Asang et al. 2008
SV leader SD1 SA5 SF2(1)- SF2(2)- env nef	pSV-1-env SF2(1)- SF2(2)-	Asang et al. 2008
SV leader SD1 SA5 SRp40 15- env nef	pSV-1-env SRp40-	Asang et al. 2008
SV leader SD1 SA5 SF2(2)- env nef	pSV-1-env SF2(2)-	Asang et al. 2008
SV leader SD1 SA5 SF2(1)- env nef	pSV-1-env SF2(1)-	Asang et al. 2008
SV leader SD1 SF2- SA5 env nef	pSV-1-env SF2(3)-	Asang et al. 2008
SV leader SD1 SF2- SA5 ESE-	pSV-1-env SF2(3)- GARmut	Asang et al. 2008
SV leader SD1 SA5 3U env nef	pSV-1-env 3U	Asang et al. 2008
SV leader SD1 SA5 3U env nef	SV-1-env 3U	Caputi et al. 2004
SV leader SD1 SA5 cs+1 <sup>12</sup> env nef	pSV-1-env cs+1 <sup>12</sup>	Asang et al. 2008
SV leader SD1 SA5 cs-2 <sup>14</sup> env nef	pSV-1-env cs-2 <sup>14</sup>	Asang et al. 2008
SV leader SD1 SA5 H- env nef	pSV-1-env $\Delta$ E42	Asang et al. 2008
SV leader SD1 SA5 HIV#18 env nef	pSV-1-env HIV#18	Asang et al. 2008
SV leader SD1 SA5 GTV env nef	pSV-1-env GTV	Asang et al. 2008
SV leader SD1 SA5 GTV 15A env nef	pSV-1-env GTV 15A	Asang et al. 2008
SV leader SD1 SA5 SD2 env nef	pSV-1-env SD2	Freund et al. 2003
SV leader SD1 SA5 -1G3U env nef	pSV-1-env -1G3U	Freund et el. 2003
SV leader SD1 SA5 4G7C env nef	pSV-1-env 4G7C	Freund et al. 2003
SV leader SD1 SA5 4C8U env nef	pSV-1-env 4C8U	Freund et al. 2003

For clarity published 2-intron minigenes were renamed in respective manuscripts.

## G.2 Abbreviations and units

### G.2.1 Abbreviations

3' ss	3' splice site
5' ss	5' splice site
Ac	Acetate
AIDS	Acquired Immunodeficiency Syndrome
Amp	Ampicillin
AP	Alkaline Phosphatase
as	antisense
ATP	adenosine-5'-triphosphat
BSA	bovine serum albumin
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CTP	cytidine-5'-triphosphate
DAPI	4',6-diamidino-2-phenylindole
ddH <sub>2</sub> O	deionised and distilled water
DMDC	dimethyl-dicarbonate
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
<i>env</i>	gene for the viral membrane protein (envelope)
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
EtBr	ethidium bromide (3,8-Diamino-6-ethyl-5-phenylphenatriumbromid)
FCS	fetal calf serum
<i>gag</i>	gene for the viral structural proteins (group specific antigen)
gp	glycoprotein
GTP	guanosine-5'-triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hGH	human growth hormone
HIV-1	Human Immunodeficiency Virus Type 1
hnRNP	heterogeneous nuclear ribonucleoprotein particle
LB	Luria Broth base
LTR	long terminal repeat
mRNA	messenger ribonucleic acid
ORF	open reading frame
ori	origin of replication
pA	polyadenylation signal
PBS	phosphate buffered saline
PBS <sub>def</sub>	Dulbecco's phosphate buffered saline deficient in Ca <sup>2+</sup> and Mg <sup>2+</sup>
PCR	polymerase chain reaction
PMSF	phenylmethane-sulfonyl-fluoride
POD	peroxidase
<i>pol</i>	gene for the viral enzymes (polymerase)
poly(A) <sup>+</sup>	polyadenylated
<i>rev</i>	gene for the viral protein Rev (regulator of viral protein expression)

---

RNA	ribonucleic acid
RNase	ribonuclease
RRE	Rev-responsive element
RS	arginine/serine-rich
s	sense
SA	splice acceptor
SD	splice donor
SELEX	Systematic Evolution of Ligands by Exponential enrichment
SDS	sodium dodecyl sulfate
SR	serine/arginine-rich
SSC	standard saline citrate
SU	viral surface envelope protein
SV40	Simian Virus 40
TE	Tris-EDTA buffer
TM	viral transmembrane envelope protein
Tris	Tris-(hydroxymethyl)-aminomethane
TTP	thymidine-5'-triphosphate
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume

### G.2.2 Units

bp	base pairs	m	meter
°C	degree Celsius	min	minutes
M	molar		
g	gram	n	nano ( $10^{-9}$ )
h	hour	nt	nucleotide
kb	kilobase	RLU	relative light units
kDa	kilodalton	rpm	rotations per minute
l	liter	sec	second
μ	micro ( $10^{-6}$ )	U	unit
m	milli ( $10^{-3}$ )	V	volt

### G.3 Publications

The results of the work presented here were in part published in scientific journals or presented at international conferences.

#### Original articles

**Asang C., Hauber I., Schaal H.** 2008. Insights into the selective activation of alternatively used splice acceptors by the Human Immunodeficiency Virus Type-1 bidirectional splice enhancer; *Nucleic Acids Res.* **36** (5): p. 1450-1463

**Caputi M., Freund M., Kammler S., Asang C., Schaal H.** 2004. A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates Human Immunodeficiency Virus Type 1 *rev*, *env*, *vpu* and *nef* gene expression; *J. Virol.* **78** (12): p. 6517-6526

**Freund M., Asang C., Kammler S., Konermann C., Krummheuer J., Hipp M., Meyer I., Gierling W., Theiss S., Preuss T., Schindler D., Kjems J., Schaal H.** 2003. A novel approach to describe a U1 snRNA binding site; *Nucleic Acids Res.* **31** (23): p. 6963-75

#### Poster presentations

**Asang C., Schaal H.** Expression of HIV-1 genomic RNA is regulated by the p17 gag instability element and the viral protein Gag.  
13th Annual Meeting of the RNA Society, Berlin, 28/07-03/08/2008

**Asang C., Flößer A., Bouschen W., Metzger S., Schaal H.** Elucidating the splicing regulatory network across HIV-1 exon 5  
Retroviruses, Cold Spring Harbor Laboratory, NY, 19-24/05/2008

**Asang C., Flößer A., Bouschen W., Metzger S., Schaal H.** A bidirectional SR protein-dependent exonic splicing enhancer regulates Rev-mediated HIV-1 *env* gene expression.  
Jahrestagung der Gesellschaft für Virologie, Heidelberg, 05-08/03/2008

**Asang C., Schaal H.** Selective activation of the alternatively used splice acceptors by the HIV-1 GAR ESE.  
Retroviruses, Cold Spring Harbor Laboratory, NY, 22-27/05/2007

**Asang C., Schaal H.** Dual ESE functions determine the alternative recognition of the HIV-1 *rev/nef* exon.  
Symposium on alternate transcript diversity II – Biology and therapeutics, EMBL Heidelberg, 21-23/03/2006

**Asang C., Schaal H.** Die SR-Proteine SC35 und SF2/ASF regulieren die Spleißstellen-Selektion der HIV-1 prä-mRNA.  
Tag des wissenschaftlichen Nachwuchts, Heinrich-Heine-Universität Düsseldorf, 24-25/06/2005

**Asang C., Schaal H.** Gag polyprotein precursor promotes genomic HIV-1 RNA expression.  
Jahrestagung der Gesellschaft für Virologie, Hannover, 16-19/03/2005

**Neveling K., Kammler S., Konermann C., Asang C., Freund M., Hipp M., Meyer I., Peabody D., Schaal H.** Splice site networking: A dynamic model of exon recognition.  
9th Annual Meeting of the RNA Society, Madison, WI, 01-06/06/2004

**Schaal H., Freund M., Kammler S., Asang C., Caputi M.** A bidirectional SR protein-dependent exonic splicing enhancer regulates Rev-mediated HIV-1 *env* gene expression.  
8th Annual Meeting of the RNA Society, Wien, 01-06/07/2003

---

## G.4 Curriculum Vitae

**Corinna Asang**

---

Dienstanschrift: Universitätsklinikum Düsseldorf  
 Institut für Virologie  
 Universitätsstr.1  
 40225 Düsseldorf  
 phone: +49/211-81-12781  
 asang@uni-duesseldorf.de

Privatanschrift: Itterstr. 158  
 40589 Düsseldorf  
 asang@alice-dsl.net

### **Persönliche Daten:**

---

**Familienstand:** ledig  
**Geburtsdatum:** 22.03.1974  
**Geburtsort:** Düsseldorf-Benrath

### **Ausbildung:**

---

- |                      |  |
|----------------------|--|
| <b>10/94 – 03/02</b> | <b>Studium der Biologie an der Heinrich-Heine-Universität Düsseldorf</b><br>Vordiplom 09/07/1997<br>Diplom 20/03/2002<br>Diplomarbeit:<br>Der Einfluss der 3'-kodierenden Region des env-Gens auf die Env-Expression;<br>ausgeführt am Institut für Medizinische Mikrobiologie und Virologie;<br>AG PD Prof. Dr. H. Schaal       |
| <b>07/02 – 04/10</b> | <b>Promotionsstudium der Biologie an der Heinrich-Heine-Universität Düsseldorf</b><br>Promotionsthema:<br>Die Regulation der frühen und späten Phase der HIV-1 Genexpression durch<br>alternatives prä-mRNA-Spleißen;<br>ausgeführt am Institut für Virologie des Universitätsklinikums Düsseldorf;<br>AG PD Prof. Dr. H. Schaal |

### **Auslandsaufenthalt:**

---

- |                      |   |
|----------------------|---|
| <b>11/05 – 12/05</b> | <b>Gastaufenthalt an der Florida Atlantic University, Boca Raton, USA</b><br><b>Assistant Prof. Dr. M. Caputi</b><br>Charles E. Schmidt College of Science, Department of Biomedical Science<br>777 Glades Road, P.O. Box 3091, Boca Raton, FI 33431, USA |
|----------------------|---|

### **Stipendium:**

---

- |                      |   |
|----------------------|---|
| <b>11/05 – 12/05</b> | <b>Reisestipendium des Boehringer Ingelheim Fonds</b> |
|----------------------|---|

### **Weiterbildung:**

---

- |                   |  |
|-------------------|--|
| <b>seit 01/09</b> | <b>Heinrich-Heine-Universität Düsseldorf</b><br>Mentee des Selma-Meyer-Mentoring-Programms |
|-------------------|--|

## **G.5 Erklärung**

Hiermit erkläre ich, dass ich die vorliegende Dissertation eigenständig und ohne unerlaubte Hilfe angefertigt habe und diese in der vorgeliegenden oder in einer ähnlichen Form noch bei keiner anderen Institution eingereicht habe.

(Corinna Asang)

Düsseldorf, den 28. April 2010

## Danksagung

Viele haben mich auf diesem Weg unterstützt. Mein besonderer Dank gilt:

**Prof. Dr. Heiner Schaal** für die Bereitstellung des wissenschaftlichen Freiraums zur weitgehend selbstständigen Bearbeitung des Themas, zahlreiche interessante Diskussionen und seine Hilfsbereitschaft bei jeglichen Problemen.

**Prof. Dr. Dieter Willbold** für die fakultätsübergreifende Betreuung dieser Arbeit.

**Prof. Dr. Hartmut Hengel** für die Möglichkeit die Arbeit am Institut für Virologie fertigzustellen.

**Prof. Dr. Ortwin Adams, Steffen Erkelenz, Dr. Marcel Freund, Linda Hartmann, Dr. Marianne Hipp, Dr. Susanne Kammler, Dr. Carolin Konermann, Dr. Jörg Krummheuer, Imke Meyer, Stephan Theiss** und **Björn Wefers** für die konstruktive Zusammenarbeit.

**Dr. Werner Bouschen, Nadine Dyballa, Dr. Sabine Metzger** und **Dr. Nicola Wiethölter** für ihre Unterstützung bei der MS-basierten Protein-Identifizierung.

**Dr. Beate Betz** und **Dr. Dieter Niederacher** für die Bereitstellung des DNA-Sequenziergeräts und die Einweisung in die automatisierte Gelelektrophorese.

allen nicht namentlich genannten **Kolleginnen und Kollegen des Instituts für Virologie, des Instituts für Medizinische Mikrobiologie und Krankenhaushygiene sowie des Zentrums** für ihre Hilfsbereitschaft und die angenehme Arbeitsatmosphäre.

**Prof. Dr. Massimo Caputi** und **Michael Schaub** für ihre Gastfreundschaft während meines Aufenthalts an der Florida Atlantic University.

**meinen Freunden**, die nicht locker ließen und mir keine Begründung ersparten. Ganz besonders möchte ich mich bei **Heike Auras** für die Bereitstellung des Westflügels und die sportlichen Höchstleistungen bedanken.

**Josefine Asang** für ihre ansteckende Lebensfreude.

**meinen Eltern** für ihr Vertrauen und ihre liebevolle Unterstützung.