

Identification and characterization of small  
molecule inhibitors of pre-mRNA splicing  
that block spliceosome assembly  
at novel stages

**Dissertation**

for the award of the degree

"Doctor rerum naturalium"

of the Georg-August University Göttingen

within the doctoral program Biology

of the Georg-August University School of Science (GAUSS)

submitted by

**Anzhalika Sidarovich**

from Soligorsk, Belarus

Göttingen 2015

### **Thesis committee**

Prof. Dr. Reinhard Lührmann  
Department of Cellular Biochemistry,  
Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Holger Stark  
Research group of 3D Electron Cryo-Microscopy,  
Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Heike Krebber  
Department of Molecular Genetics,  
Institute for Microbiology and Genetics, Georg August University Göttingen

### **Members of the Examination Board**

Reviewer:

Prof. Dr. Reinhard Lührmann  
Department of Cellular Biochemistry,  
Max Planck Institute for Biophysical Chemistry, Göttingen

Second Reviewer:

Prof. Dr. Holger Stark  
Research group of 3D Electron Cryo-Microscopy,  
Max Planck Institute for Biophysical Chemistry, Göttingen

Additional Reviewer:

Prof. Dr. Heike Krebber  
Department of Molecular Genetics,  
Institute for Microbiology and Genetics, Georg August University Göttingen

### **Further members of the Examination Board**

Prof. Dr. Markus Zweckstetter  
Department for NMR-based Structural Biology,  
Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Ralf Ficner  
Department for Molecular Structural Biology  
Institute for Microbiology and Genetics, Georg August University Göttingen

Prof. Dr. Kai Tittmann  
Department for Molecular Enzymology  
Georg August University Göttingen

Date of the oral examination: 17.04.2015

*"I wonder about turtles."*

*"What do you wonder about turtles?" Angela asked him.*

*"When they pull in their heads," he said, "do their spines buckle or contract?"*

Kurt Vonnegut, *Cat's Cradle*

*For my family with love*



# **Affidavit**

I hereby declare that my thesis entitled "Identification and characterization of small molecule inhibitor of pre-mRNA splicing that block spliceosome assembly at novel stages" has been written independently and with no other sources and aids than quoted. This thesis (wholly or in part) has not been submitted elsewhere for any academic award or qualification.

Anzhalika Sidarovich

February 2015

Göttingen



## Table of Contents

<b>Abstract</b> .....	<b>1</b>
<b>1 Introduction</b> .....	<b>3</b>
<b>1.1 Pre-mRNA splicing</b> .....	<b>3</b>
1.1.1 Structure of pre-mRNAs .....	3
1.1.2 Intron recognition sites .....	4
1.1.3 Mechanism of pre-mRNA splicing .....	4
1.1.4 Composition of the spliceosome .....	5
1.1.5 Spliceosome assembly.....	9
1.1.5.1 Exon definition, an alternative early assembly pathway .....	11
1.1.6 RNA-RNA interactions in the spliceosome.....	12
1.1.7 RNA-based catalysis of pre-mRNA splicing.....	14
1.1.8 Protein composition of the spliceosome.....	15
1.1.8.1 non-snRNP spliceosomal proteins.....	16
1.1.8.2 Dynamics of the spliceosome's protein composition .....	18
1.1.9 Posttranslational modification of proteins during splicing.....	19
1.1.10 Structure of the spliceosome and spliceosomal snRNPs .....	21
1.1.11 Small molecule inhibitors of pre-mRNA splicing.....	22
<b>1.2 Aim of this study</b> .....	<b>23</b>
<b>2 Materials and methods</b> .....	<b>25</b>
<b>2.1 Material</b> .....	<b>25</b>
2.1.1 Chemicals.....	25
2.1.2 Antisera, monoclonal and polyclonal antibodies .....	27
2.1.3 Enzymes and enzyme inhibitors, and MS2-MBP protein.....	28
2.1.4 Nucleotides.....	28
2.1.5 Oligonucleotides .....	29
2.1.6 Plasmids .....	29
2.1.7 Cell lines .....	29
2.1.8 Buffers.....	30
2.1.9 Commercial kits .....	33
2.1.10 Chromatography materials and consumables.....	33
2.1.11 Small Molecule Screening Apparati.....	34
2.1.12 Apparati .....	34
<b>2.2 Methods</b> .....	<b>36</b>

2.2.1 Molecular biology standard methods .....	36
2.2.1.1 <i>in vitro</i> transcription .....	36
2.2.1.2 Concentration determination of nucleic acids.....	37
2.2.1.3 Agarose gel electrophoresis of nucleic acids.....	37
2.2.1.4 Phenol-chloroform-isoamyl alcohol extraction .....	38
2.2.1.5 Proteinase K digestion .....	38
2.2.1.6 Denaturing Urea polyacrylamide gel electrophoresis .....	38
2.2.1.7 RNA silver staining .....	39
2.2.1.8 Radioactive labeling of the 5'-end of DNA oligonucleotides .....	39
2.2.1.9 Northern blotting .....	39
2.2.2 Standard protein biochemical methods.....	40
2.2.2.1 TCA precipitation of protein.....	40
2.2.2.2 Acetone precipitation of protein.....	40
2.2.2.3 Denaturing SDS polyacrylamide gel electrophoresis. ....	40
2.2.2.4 Coomassie staining of SDS-PAGE gels .....	41
2.2.2.5 Silver staining of SDS-PAGE gels.....	41
2.2.3 Immunological Methods.....	42
2.2.3.1 Immunoblotting (Western blotting).....	42
2.2.3.2 Immunoprecipitation.....	43
2.2.4 Special Methods .....	43
2.2.4.1 Preparation of HeLa cell nuclear extract.....	43
2.2.4.2 <i>In vitro</i> splicing reactions.....	43
2.2.4.3 <i>In vitro</i> high throughput-screening assay of small molecules .....	44
2.2.4.4 Analysis of splicing complexes by native agarose gel electrophoresis.....	45
2.2.4.5 Purification of spliceosomal complexes .....	45
2.2.4.6 Micrococcal nuclease treatment of HeLa nuclear extract.....	46
2.2.4.7 ATP depletion.....	46
2.2.4.8 Chasing of stalled complexes in the presence or absence of ATP .....	46
2.2.4.9 Purification of protein complexes with anti-FLAG-beads .....	47
2.2.5 RNA cross-linking and structure probing .....	47
2.2.5.1 Psoralen crosslinking.....	47
2.2.5.2 Chemical modification.....	47
2.2.5.3 Primer extension .....	48
2.2.6 Electron microscopy .....	49
2.2.6.1 Sample Preparation for Electron Microscopy.....	49
2.2.6.2 Negative staining .....	49
2.2.6.3 Preparation of unstained samples for electron cryo microscopy .....	49
2.2.6.4 Electron microscopy .....	50
2.2.6.5 Single-Particle Image Processing .....	50



2.2.7 Mass spectrometry.....	50
<b>3 Results.....</b>	<b>51</b>
<b>3.1 Identification of new small molecule inhibitors of pre-mRNA splicing.....</b>	<b>51</b>
3.1.1 Search for small molecule inhibitors of pre-mRNA splicing using a high throughput screening assay .....	51
3.1.2 Confirmation of splicing inhibition activity in an <i>in vitro</i> splicing assay.....	53
3.1.3 Determination of IC <sub>50</sub> values and the stage at which spliceosome assembly is inhibited .....	54
<b>3.2 Purification and characterization of spliceosomal complexes stalled with different splicing inhibitors.....</b>	<b>59</b>
3.2.1 Time course of splicing complex formation at the lowest concentration at which splicing is completely inhibited .....	59
3.2.2 Sedimentation behaviour of splicing complexes stalled with compounds 028, 120, 296 and 297 .....	60
3.2.3 Affinity-purification of spliceosomal complexes stalled in the presence of selected compounds.....	62
3.2.4 Isolation of splicing complexes stalled with compound 028 .....	64
3.2.5 Protein composition of affinity-purified spliceosomal A and B-like complexes formed in the presence of compound 028 .....	68
3.2.6 Immunoblotting confirms the absence/presence of selected proteins in affinity-purified spliceosomal complex B <sup>028</sup> .....	73
3.2.7 Attempts to isolate more homogeneous B <sup>028</sup> complexes.....	75
3.2.8 Analysis of RNA-RNA interactions in spliceosomes stalled by compound 028 .....	76
3.2.8.1 Identification of RNA-RNA interactions in B <sup>028</sup> complexes via psoralen crosslinking.....	78
3.2.8.2 Structure of U5 snRNA in B <sup>028</sup> complexes .....	79
3.2.8.3 Structure of U2 and U6 snRNAs in B <sup>028</sup> complexes .....	81
3.2.9 B <sup>028</sup> complexes can be chased into catalytically active spliceosomes. ....	86
<b>3.3 Electron microscopy studies of affinity-purified B<sup>028</sup> complexes .....</b>	<b>87</b>
3.3.1 Negative stain 2D electron microscopy of affinity-purified B <sup>028</sup> complexes....	88
3.3.2 3D reconstruction of the B <sup>028</sup> particle.....	91
<b>3.4 B<sup>028</sup> spliceosomes as a tool to investigate binding and release of proteins during activation.....</b>	<b>94</b>
<b>3.5 Mechanism of inhibitory action of compound 028.....</b>	<b>101</b>
3.5.1 U4/U6.U5 tri-snRNP is stable in the presence of compound 028 .....	101

3.5.2 Association of AD002 with the Prp19/CDC5L complex is affected by the presence of compound 028 .....	103
3.5.3 Gradient analysis of Prp19/CDC5L complex in the presence/absence of compound 028.....	104
<b>3.6 Structure-activity relationship (SAR) analysis of compound 028 .....</b>	<b>105</b>
3.6.1 Effect of the reduction of the linker between the two heterocyclic motifs on splicing inhibition <i>in vitro</i> .....	106
3.6.2 Effect of further modifications on splicing inhibition <i>in vitro</i> .....	108
<b>4 Discussion .....</b>	<b>116</b>
<b>4.1 Identification of new small molecule inhibitors of pre-mRNA splicing.....</b>	<b>116</b>
<b>4.2 Purification of a novel spliceosome assembly intermediate .....</b>	<b>117</b>
4.2.1 Composition of affinity-purified spliceosomal complexes stalled by compound 028 .....	117
4.2.2 Determination of the stage at which the spliceosome assembly is stalled in the presence of compound 028 by protein composition analysis .....	119
4.2.3 Characterization of the RNA-RNA network in spliceosomes stalled by compound 028.....	121
<b>4.3 Electron microscopy studies of affinity-purified B<sup>028</sup> complexes .....</b>	<b>124</b>
<b>4.4 Identification of chemical groups important for compound 028's splicing inhibition activity .....</b>	<b>128</b>
<b>4.5 B<sup>028</sup> spliceosomes can be used to investigate binding and release of proteins during activation.....</b>	<b>129</b>
<b>4.6 Perspectives.....</b>	<b>132</b>
<b>5 References .....</b>	<b>134</b>
<b>6 Appendix .....</b>	<b>146</b>
6.1 List of abbreviations.....	146
6.2 List of figures.....	149
6.3 List of tables.....	151
6.4 Acknowledgements.....	152
6.5 Curriculum vitae .....	154
6.6 Conferences .....	155

# Abstract

The spliceosome is a highly dynamic, megadalton ribonucleoprotein (RNP) molecular machine that catalyses the important step of the post-transcriptional processing of eukaryotic precursor messenger RNA (pre-mRNA), called splicing. During splicing the non-coding nucleotide intervening regions (introns) of newly transcribed pre-mRNA are excised and the protein-coding sequences (exons) are ligated *via* two consecutive transesterification reactions. The spliceosome is assembled *de novo* on each pre-mRNA intron by consecutive recruitment of five RNA-protein complexes (snRNPs) and numerous non-snRNP factors in a dynamic manner. Initially, the U1 and U2 snRNPs bind to the 5' splice site and branch point sequence, respectively, forming the A complex. In the next step, the preformed U4/U6.U5 tri-snRNP binds and complex B is assembled. Thereafter, complex B undergoes multiple rearrangements in terms of both its conformation and composition, including the dissociation of U1 and U4, resulting in the formation of the B<sup>act</sup> complex. After the formation of the catalytically active B\* complex, the first step of splicing takes place, generating the cleaved 5' exon and intron-3' exon lariat intermediates, and creating complex C, which performs the second step of splicing, resulting in the ligation of the 5' and 3' exons and release of an intron in the form of a lariat. Various assembly intermediates of the spliceosome have been experimentally defined and characterized. However, the dynamics of protein exchange, as well as RNA rearrangements during the spliceosome assembly remain in many aspects unclear.

The spliceosome assembly is highly regulated in multiple ways. Disruption or misregulation of alternative and constitutive splicing are the cause or severity modulator of many human diseases, including among others cancer, neurodegenerative and autoimmune diseases, making the spliceosome a highly attractive drug target. Small molecule inhibitors that block discrete steps of the extremely dynamic assembly and functional cycle of the spliceosome are not only of potential therapeutic value, but also highly useful for the detailed investigation of the structure and function of the spliceosome. However, only a limited number of small molecule inhibitors that specifically target the pre-mRNA splicing machinery have been identified at present. Using a robust, rapid and sensitive high throughput *in vitro* splicing assay, which monitors the formation of step I spliceosomes (i.e. the spliceosomal C complex) by measuring the association of a FLAG-tagged version of the DEAD box ATPase Abstrakt, a chemical library of ~172,000 small molecules was screened for splicing inhibition activity. All compounds showing at least a 50% decrease in the signal intensity of bound Abstrakt were subjected to a second test, and reproducible hits were finally tested in an *in vitro* splicing assay using <sup>32</sup>P-labeled adenovirus-derived MINX-MS2 pre-mRNA as substrate.

Ten compounds were confirmed to be inhibitors of pre-mRNA splicing *in vitro*, and exhibited IC<sub>50</sub> values ranging from 3 to 50 μM. To determine at which stage they inhibit splicing, I

performed a splicing time course and analysed the spliceosomal complexes formed by native agarose gel electrophoresis. Analysis of splicing complex formation revealed that at least one compound (hereafter designated as 028), led to an accumulation of A complexes and a complex denoted as B<sup>028</sup> that sediments slightly faster than the pre-catalytic B complex on glycerol gradients. Gradient centrifugation experiments indicated that U4/U6.U5 tri-snRNP stability is not affected significantly by 028. To characterize in more detail spliceosomal complexes formed in the presence of compound 028, I subjected the stalled splicing reactions to glycerol-gradient centrifugation and purified the complexes in a given peak by MS2-MBP affinity selection. Initial analysis of the RNA and protein composition of complexes affinity-purified from the “B-like” peak, suggested that compound 028 stalls splicing at a novel stage of the spliceosome activation step; i.e., inhibition appears to take place after release of the U4 snRNP, but prior to the release of the LSm and B-specific proteins and before the stable integration of the Prp19 complex, the recruitment of B<sup>act</sup> specific proteins and the phosphorylation of U2 protein SF3b155. Analysis of the RNA-RNA network within B<sup>028</sup> by RNA structure probing and psoralen crosslinking suggested, but did not definitively prove, that the catalytically important U6-ISL and U2/U6 helix Ia and Ib likely are formed in the spliceosomes stalled by compound 028. Purified B<sup>028</sup>-complexes could be chased into catalytically active spliceosomes upon addition of micrococcal nuclease-treated nuclear extract, demonstrating that they are functional complexes. We thus performed initial “chase” experiments with purified B<sup>028</sup> complexes to follow the recruitment and release of proteins during the activation step. Our results suggest that ATP is required for the recruitment of the Prp19/CDC5L complex proteins and B-specific proteins. Negative stain electron microscopy of B<sup>028</sup>-complexes and 3D reconstruction of the B<sup>028</sup>-particle revealed a unique morphology, which is different from the pre-catalytic B and activated B<sup>act</sup> spliceosomal complexes, but suggest common morphological features with both B and B<sup>act</sup>, consistent with them being stalled at an intermediate assembly stage. Finally, structure-activity relationship (SAR) studies, in which modifications of compound 028 were assayed for their effect on pre-mRNA splicing in vitro, revealed structural determinants that are important for the inhibition activity of compound 028. Thus, compound 028 allows us to obtain a novel snapshot of the spliceosome assembly pathway and to perform detailed structural and functional investigations, to improve our limited understanding of the dynamic rearrangement of spliceosomal components during spliceosome activation.

# 1 Introduction

The genetic information of a living organism is stored in its deoxyribonucleic acid (DNA). Whereas prokaryotic genes consist of continuous, protein coding DNA sequences, most genes of higher eukaryotes display a mosaic structure. Thereby protein-coding sequences (exons) are interrupted by long non-coding nucleotide intervening regions (introns). To allow correct translation of the protein-coding region, the introns of newly transcribed pre-mRNA (*precursor messenger RNA*) have to be removed precisely from the primary transcript, and the exons ligated together during the evolutionarily conserved process, called splicing. Additionally, the 5' end of the pre-mRNA is modified with a unique cap-structure, consisting of a N<sup>7</sup>-methylguanosine (m<sub>7</sub>G) covalently bound to the ribose via an unusual 5' to 5' triphosphate linkage (Shatkin, 1976; Shatkin and Manley, 2000). Furthermore, the 3' end of the pre-mRNA undergoes polyadenylation (post-transcriptionally) (Colgan and Manley, 1997). Capping and polyadenylation are essential for the stability of the pre-mRNA/mRNA, regulation of mRNA transport to the cytoplasm and efficient translation. Finally, mature mRNA is transported from the nucleus to the cytoplasm and used as a template in the protein translation process.

## 1.1 Pre-mRNA splicing

### 1.1.1 Structure of pre-mRNAs

Pre-mRNAs vary greatly in the number of introns they contain and their lengths. The yeast genome contains very few introns, namely, less than 250 of more than 6200 genes contain an intron (e.g. only 3.8%), and in most cases there is only one intron per gene (Ares et al., 1999; Spingola et al., 1999). In contrast, metazoan genes contain on average 8 introns each (Sakharkar et al., 2004).

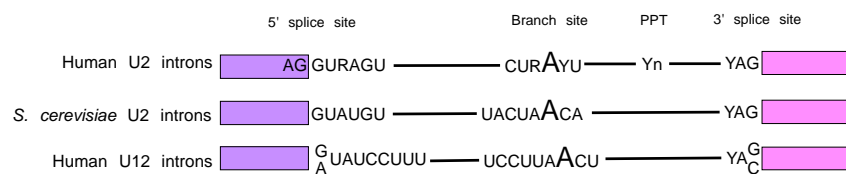
Some pre-mRNAs are spliced constitutively, that is all exons in a single pre-mRNA are spliced together in the same way, generating the same mRNA without any sequence variation (Will and Lührmann, 2011). When alternative splicing takes place, different sets of exons from a single pre-mRNA are spliced together, leading to variable mRNAs. In higher eukaryotes, alternative splicing leads to the expansion of the genome by vastly amplifying the palette of translated proteins (Black, 2003). It is estimated that 40–60% of human transcripts undergo alternative splicing (Fox-Walsh et al., 2005; Modrek and Lee, 2003). Malfunction/defects in the splicing machinery and its regulation, mutations in spliceosome components and *cis*-acting point mutations (e.g. within splice sites) can lead

to a large number of pathological diseases and/or alter their severity level, including cancer and neuro-degenerative diseases (Cooper et al., 2009; Webb et al., 2013).

### 1.1.2 Intron recognition sites

During splicing, different RNA-protein complexes (snRNPs) bind to the conserved sequences of the pre-mRNA, thereby assembling a spliceosome. The splice sites of the pre-mRNA are recognized with an astonishingly high level of fidelity. For the recognition of the introns, the short conserved intron consensus-sequences are of great importance. These consensus-sequences include the 5' splice site, 3' splice site and branch point sequence (BPS), which is located 18-40 nucleotides upstream from the 3' splice site (Burge et al., 1999) (Fig. 1.1). Generally, consensus sequences in higher eukaryotes are less evolutionarily conserved than those in *S. cerevisiae* (Berget, 1995). There are two types of introns in higher eukaryotes, so-called U2- and U12-introns, which are removed by two different spliceosomes. The U12-type introns, which are not very abundant (less than 1 % of introns in human), have more highly conserved 5' splice site and branch point sequences. U2-type introns contain additionally a pyrimidine-rich region (PPT), which is located between the BPS and 3' splice site (Burge et al., 1999). The consensus sequences are recognized during splicing by multiple snRNAs and proteins, ensuring the fidelity of splice site definition.

In addition, other *cis*-acting pre-mRNA elements also are of importance. These include exonic or intronic splicing enhancers (ESEs or ISEs) or silencers (ESIs or ISIs), which bind regulatory proteins, enabling the repression or stimulation of spliceosome assembly on the pre-mRNA substrate (Wang and Burge, 2008).

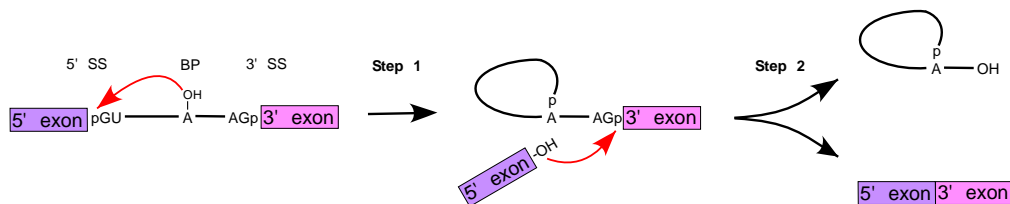


**Figure 1.1 Consensus sequences of pre-mRNA in yeast and human.** Here two exons (coloured boxes) are separated by an intron (black line). Consensus sequences are found at the 5' splice site, branch point sequence (BPS) and 3' splice site. A = adenine, C = cytosine, G = guanine, U = uracil, R = purine, Y = pyrimidine, pyrimidine-rich region Y<sub>(n)</sub>. N can be A, U, G or C.

### 1.1.3 Mechanism of pre-mRNA splicing

Pre-mRNA splicing is catalysed by a multimegadalton, ribonucleoprotein complex (RNP), called the spliceosome. It involves two sequential S<sub>N</sub>2-type transesterification

reactions (Moore et al., 1993). In the first step, a nucleophilic attack by the 2' hydroxyl group of the branchpoint adenosine takes place on the phosphodiester bond at the 5' splice site (5'SS). Via this reaction the pre-mRNA is cleaved, and the 5' end of the intron is connected to the BP adenosine via a 2'-5'-phosphodiester bond, leading to the formation of a branched RNA lariat structure (Fig. 1.2). The result is a "free" 5' exon with a 3' hydroxyl group and a lariat intermediate containing the intron and the downstream 3' exon. During the second step, a nucleophilic attack of the free 3' hydroxyl group of the 5' exon occurs on the phosphodiester bond at the 3' SS (Fig. 1.2). Thereby both exons are ligated together via a phosphodiester bond and the excised intron is released in the form of a lariat. The lariat intron is then debranched and degraded, whereas the spliced mRNA can be transported to the cytoplasm for protein biosynthesis.



**Figure 1.2 Schematic view of the two-step mechanism of pre-mRNA splicing.** In the first step a nucleophilic attack of the oxygen of the 2' OH group of the BP adenosine at the phosphodiester bond on the 5' splice site (5'SS) results in the formation of the free 5' exon and intron-3' exon intermediates. In the second step, a nucleophilic attack of the free 3' OH group of the 5' exon targets the phosphodiester bond at the 3' SS. Both exons are ligated together and the intron is released in the form of a lariat. The exons are represented as boxes in different colours, and the intron is shown as a solid line. Invariant nucleotides at the exon-intron boundaries and branch point adenosine are indicated. Phosphate groups involved in the reactions are indicated by (p).

### 1.1.4 Composition of the spliceosome

The precise excision of an intron from the pre-mRNA and the following ligation of its exons are catalysed by a dynamic macromolecular machine called the spliceosome. The spliceosome consists of five small nuclear ribonucleoprotein particles (snRNPs) that assemble on the pre-mRNA substrate in a stepwise manner. The small uridine-rich nuclear RNA components of these snRNPs (U snRNAs) build a complicated RNA-RNA network and participate in the catalysis of splicing. The proteins of snRNPs and many additional non-snRNP proteins assist in splice site recognition and in the stabilization of the RNA network.

There are two distinct types of spliceosomes: the *major* U2-dependent spliceosome, which is present in all eukaryotic cells and catalyses the excision of U2-type introns, and

the *minor* U12-dependent spliceosome, which is found only in a small set of eukaryotes and catalyses the excision of the less abundant U12-type introns (Patel and Steitz, 2003; Will and Lührmann, 2005). The *minor* spliceosome contains the snRNPs U11, U12, U4atac and U6atac, which are functional analogues of the U1, U2, U4 and U6 snRNPs which form the *major* U2-type spliceosomes. U5 snRNP is a component of both kinds of spliceosome.

#### 1.1.4.1 The snRNAs

The U1, U2, U4/U6 and U5 snRNPs are the basic integral modular components of the *major* spliceosome, which catalyses the excision of more than 99 % of human introns. The spliceosomal snRNPs contain a uridine-rich small nuclear RNA molecule (U1, U2 or U5 snRNAs) or two associated uridine-rich small nuclear RNA molecules (U4/U6 di-snRNA). The snRNAs bind to proteins specific to each snRNP (Fig. 1.3) and also seven common Sm proteins (Will and Lührmann, 2006).

The *major* human snRNAs are short (106-187 nts) and localized in the nucleus (Will and Lührmann, 2001). The snRNAs are folded in a characteristic manner, with stem loop structures continually interrupted by single-stranded regions (Fig. 1.3). The U4 and U6 snRNAs engage in stable base-pairing interactions via stem I and stem II, forming the U4/U6 di-snRNA (Guthrie and Patterson, 1988). The snRNAs contain highly conserved regions that typically are bound by proteins or interact with other RNAs (Will and Lührmann, 1997). A typical attribute of U1, U2, U4 and U5 snRNAs is the presence of a conserved Sm-site (Branlant et al., 1982), which is the binding site for the seven Sm-proteins that are present in all snRNPs. The Sm proteins (B/B', D1, D2, D3, E, F, G) are found in all *major* U snRNPs except for U6 snRNP, and are important for snRNP biogenesis (Kambach et al., 1999). The U6 snRNP contains seven Sm-like (Lsm) proteins instead, which bind the 3' end of U6 snRNA (Achsel et al., 1999).

U snRNAs contain posttranscriptional chemical modifications, including site-specific pseudouridine ( $\psi$ ) formation and 2'-O-methylation (Reddy and Busch, 1983). These are often found on extremely well-conserved nucleotides of the snRNAs, such as the branch site recognition region of vertebrate U2 snRNA (Massenet et al., 1998) and can be important for U snRNA function (Yu et al., 1998).

The spliceosomal U1, U2, U4 and U5 snRNAs are transcribed in the nucleus by RNA polymerase II and, after m7G capping, are transported to the cytoplasm, where the seven Sm proteins assemble onto the conserved Sm site of the snRNA, forming the so-called Sm core (Will and Lührmann, 2001). The Sm proteins build a heptameric ring structure,



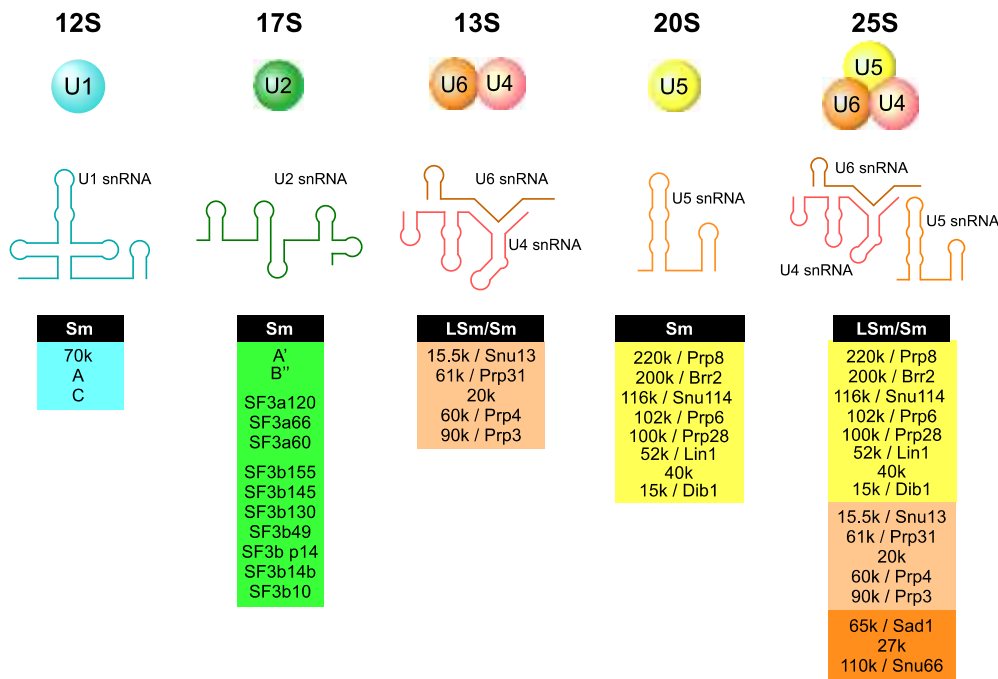


where particle-specific proteins bind to the snRNP. Finally, the association of particle-specific proteins is the last step in the maturation of snRNPs. In contrast, the U6 snRNA is transcribed by RNA Polymerase III and contains a  $\gamma$ -monomethyl triphosphate cap. The U6 snRNA does not contain an Sm site and associates instead with seven Sm-like proteins (LSm2-8), which bind to its 3' end and form a heptameric ring structure (Achsel et al., 1999; Seraphin, 1995). The U6 snRNP undergoes all of its maturation steps exclusively in the nucleus (Will and Lührmann, 2001).

#### **1.1.4.2 snRNP-specific proteins**

Each spliceosomal U snRNP also contains a set of particle-specific proteins (Fig. 1.4). The snRNP-specific proteins play essential roles during spliceosome assembly and the catalytic steps of splicing (Fig. 1.4). The 12S U1 snRNP contains three U1 snRNP-specific proteins (U1-C, U1-A and U1-70K) (Will and Lührmann, 2006), and U1-C and U1-70K aid the association of the U1 snRNA with the 5' splice site (Heinrichs et al., 1990). The functionally active 17S U2 snRNP contains at least twelve U2 snRNP-specific proteins. First the U2-A` and U2-B`` proteins bind to stem loop IV of U2 snRNA as a dimer and build up the 12S U2 snRNP. The heteromeric splicing factors SF3b and SF3a bind successively to the 12S U2 snRNP, creating the 17S U2 snRNP (Brosi et al., 1993a). U2-related proteins also associate with the U2 snRNP, but dissociate at salt concentrations above 250 mM (Will et al., 2002). The SF3a complex consists of 3 subunits with molecular weights of 120, 66 and 60 kDa (Brosi et al., 1993b), whereas the SF3b complex consists of seven subunits (SF3b155, SF3b145, SF3b130, SF3b49, p14, SF3b14b and SF3b10) (Will et al., 2002). SF3a and SF3b facilitate the interaction between U2 snRNP and the BPS, enabling the formation of spliceosomal complex A (Gozani et al., 1996; Kramer and Utans, 1991).

The 20S U5 snRNP consists of eight U5-specific proteins (Bach et al., 1989; Behrens and Lührmann, 1991). The largest, highly conserved U5 protein is 220K/hPrp8, which is present at the catalytic core of the spliceosome. It has been shown to interact with other U5 snRNP proteins, namely, the GTPase 116K/Snu114 and the DExD/H box helicase 200K/Brr2 (Achsel et al., 1998) and to cross-link with the pre-mRNA substrate at all catalytically important sites (5' SS, 3' SS, BP), as well as with U5 and U6 snRNAs (Galej et al., 2014; Grainger and Beggs, 2005). 200K/Brr2 is an integral U5 snRNP component, catalysing unwinding of the U4/U6 di-snRNA duplex during the spliceosome activation step. 100K/Prp28 activity displaces the U1 snRNP from the 5' splice site (Mohlmann et al., 2014; Strauss and Guthrie, 1994).



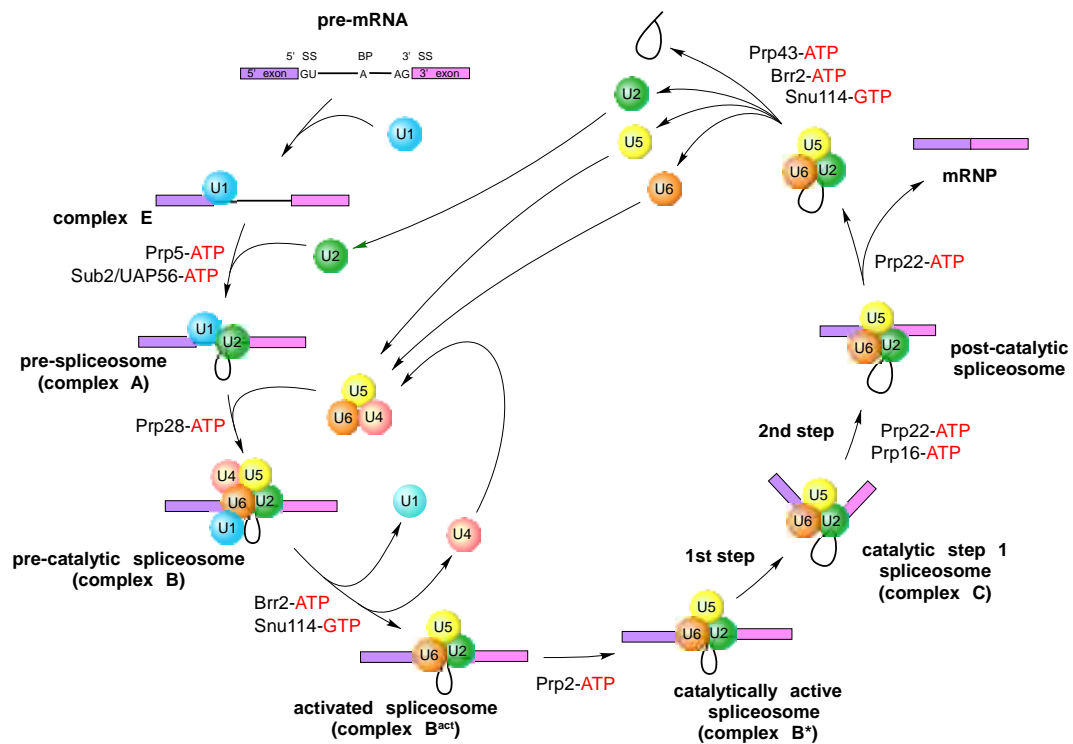
**Figure 1.4 Composition of the major spliceosomal U snRNPs.** The major spliceosome consists of U1, U2, U4/U6 and U5 snRNPs. Each snRNP consists of one (or two) snRNA molecules and different numbers of complex-specific proteins. The secondary structure of the human U snRNAs is shown schematically. The protein composition of the U snRNPs is shown in boxes of different colour. The U4/U6.U5 tri-snRNP is composed of U5 and U4/U6. Sm/Lsm proteins are shown in black boxes.

The 13S U4/U6 di-snRNP contains both LSm and Sm proteins, and 5 specific proteins: 90K/Prp3, 61K/Prp31, 60K/Prp4, 20K (also known as CypH) and 15.5K/Snu13 (Behrens and Lührmann, 1991; Lauber et al., 1997).

The 25S U4/U6.U5 tri-snRNP consists of the 13S U4/U6 di-snRNP and 20S U5 snRNP plus three additional proteins: 110K/Snu66, 65K/Sad1 and 27K (Behrens and Lührmann, 1991; Lauber et al., 1997). However, the U5-52K protein dissociates upon formation of tri-snRNP (Laggerbauer et al., 2005). The U4/U6.U5 tri-snRNP contains two sets of Sm proteins and one set of LSm proteins (Achsel et al., 1999; Salgado-Garrido et al., 1999). U4/U6-61K/hPrp31 and U5-102K/hPrp6 play a key role in the U4/U6.U5 tri-snRNP formation, interacting with each other and thus forming a bridge between the two particles (Makarova et al., 2002; Schaffert et al., 2004).

### 1.1.5 Spliceosome assembly

The spliceosome is a highly dynamic RNP machine (Wahl et al., 2009), and assembles on the pre-mRNA substrate in a stepwise manner. The spliceosome undergoes dynamic structural and compositional rearrangements, bringing the reactive sites of the pre-mRNA into close proximity to allow catalysis.



**Figure 1.5 The spliceosome cycle.** Spliceosome assembly is activated by the interaction of the U1 snRNP with the 5' splice site (complex E). Complex A formation occurs by stable binding of the U2 snRNP with the BPS. Incorporation of the pre-formed tri-snRNP (complex B) is followed by structural rearrangements, including U1 and U4 release, and the spliceosome becomes catalytically active after the action of Prp2. The first transesterification reaction results in the formation of complex C. After the second transesterification reaction, the mature mRNA and the lariat-intron are released. The spliceosomal components dissociate and are regenerated for the next round of splicing.

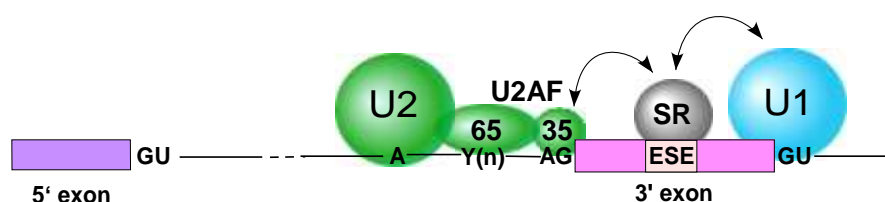
Native electrophoresis, affinity selection and glycerol gradient centrifugation revealed distinct spliceosome assembly stages that form via the sequential association and release of spliceosomal snRNPs (Will and Lührmann, 2006) (Fig. 1.5), as well as distinct complexes formed during the catalytic steps of splicing and during spliceosome dissociation. The initial step of spliceosome assembly occurs when the U1 snRNP binds to the 5' SS of an intron in an ATP-independent manner. The base-pairing interaction of the 5' end of the U1 snRNA with the 5' SS (Ruby and Abelson, 1988; Seraphin et al., 1988) is stabilized in higher eukaryotes by members of the SR protein family and U1 snRNP-associated proteins (Will and Lührmann, 1997). The SF1/BBP protein binds to the BPS and U2AF65 (U2 auxiliary factor) binds to the polypyrimidine sequence (PPT) (Reed, 1996). Furthermore, the AG-dinucleotide of the 3' SS is bound by the U2AF subunit U2AF35, which is tightly bound by U2AF65. These molecular interactions play a decisive role during the initial recognition of the 5' SS and 3' SS of an intron and lead to the formation of spliceosomal complex E (Das et al., 2000). In the next step, the U2

snRNA undergoes the ATP-dependent base pairing interaction with the BPS of the pre-mRNA, driven by the helicases UAP56 and hPrp5 (Zhang and Green, 2001), resulting in the formation of complex A. The branch site adenosine is bulged out, and thus prepared for its nucleophilic attack during the first catalytic step (Query et al., 1994). The U2/BP interaction is stabilized by the U2 snRNP-associated heteromeric protein complexes SF3a and SF3b, which contact the pre-mRNA at or close to the branch site (Gozani et al., 1996; Will and Lührmann, 2001). The formation of complex B occurs when the U4/U6 and U5 snRNPs, as a pre-formed U4/U6.U5 tri-snRNP, associate. Even though all snRNPs are present in complex B, this complex remains catalytically inactive and needs significant compositional and conformational rearrangements for its activation. Prior to catalytic activation, the DEAD-box helicase hPrp28 catalyses the release of U1 from the 5' SS of the pre-mRNA (Staley and Guthrie, 1999). Subsequently, the helicase Brr2 catalyses unwinding of the U4/U6 base pairing interaction (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998). This allows U6 snRNA to base pair with the 5' SS and the U2 snRNA, bringing the 5' SS and the BP sequence near to each other for the first step of splicing, and also to form an internal stem-loop (U6 ISL) which is important for splicing catalysis. In this way, the U1 and U4 snRNPs are destabilized, and the activated spliceosome B<sup>act</sup> is generated. Thereby, U5 snRNA contacts the nucleotides near the 5' SS. The action of the DEXH/D-box helicase Prp2 leads to the destabilization of U2 proteins from the BPS (Kim and Lin, 1996; Warkocki et al., 2009), resulting in the formation of the catalytically active complex B\*, which is able to perform the first step of splicing, thereby creating complex C. The spliceosome undergoes further conformational rearrangements catalysed by the helicase Prp16 (Schwer and Guthrie, 1992). After catalysing the second step of splicing, a post-spliceosomal complex is formed. The post-spliceosomal complex dissociates and the mRNA is released in the form of an mRNP, catalysed by the RNA helicase Prp22. Finally, U2, U5 and U6 snRNPs can be recycled to be used in the next round of splicing. The intron, in the form of a lariat, is also released by the action of the helicase Prp43 (Martin et al., 2002), debranched by Dbr1-catalysed hydrolysis of the 2'-5' phosphodiester bond (Chapman and Boeke, 1991), and is finally degraded. For each intron to be spliced, the spliceosome is assembled *de novo*.

#### **1.1.5.1 Exon definition, an alternative early assembly pathway**

The length of an intron determines if the initial assembly of the spliceosome occurs across an intron or across an exon. An intron length of 200 to 250 nucleotides ensures splice site recognition by spliceosome components initially across an intron (Fox-Walsh et al., 2005). When introns are longer (i.e. exceed 250 nts) splice sites are recognized

across an exon (Berget, 1995), which is the case for most pre-mRNAs in higher eukaryotes (Fox-Walsh et al., 2005). During exon definition, U1 snRNP binds the 5' SS downstream of an exon and U2 snRNP binds the BPS upstream of it, (first initiated by the association of U2AF with the PPT and the 3' SS). The proteins of the SR family bind ESE sequences, stabilizing the exon-defined complex (Hoffman and Grabowski, 1992; Reed, 2000). Subsequently, these cross-exon interactions are replaced by cross-intron interactions, a process that is not well understood. For some pre-mRNAs, this step is decisive in determining if an exon is included or skipped during the spliceosome cycle (Sharma et al., 2008). After the B complex stage, the cross-exon and cross-intron assembly pathways are the same.



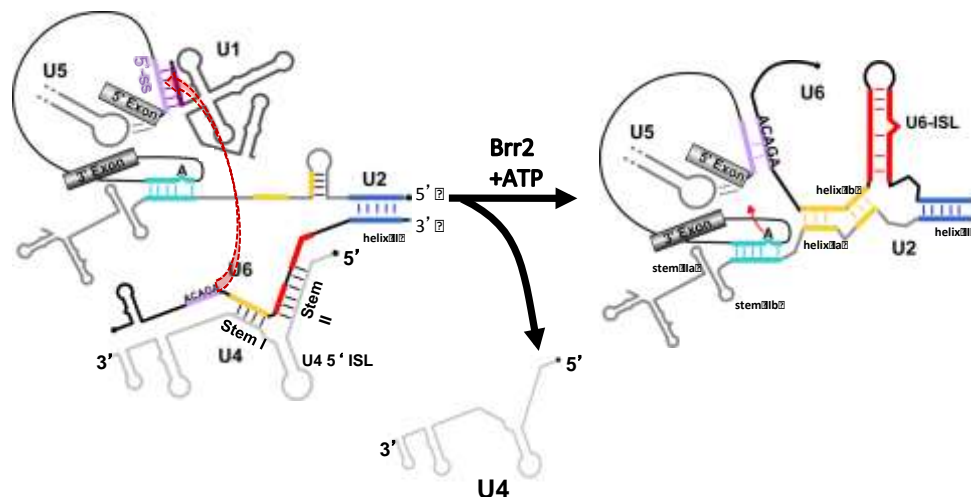
**Figure 1.6 Cross-exon definition.** U1 snRNP binds the 5' SS downstream of an exon and U2 snRNP binds the BPS upstream of it, initiated by the association of U2AF with the PPT and the 3' SS. SR proteins interact with an exonic splicing enhancer (ESE). Subsequently, the exon-defined complex is replaced by a complex with cross-intron interactions.

### 1.1.6 RNA-RNA interactions in the spliceosome

During spliceosome assembly and catalysis, a highly complicated, dynamic RNA-RNA network is formed (Nilsen, 1994; Will and Lührmann, 2011). At the initial stage of spliceosome assembly, the U1 snRNA undergoes base-pairing interactions with exon and intron nucleotides at the 5' SS of the pre-mRNA (5' SS-recognition) (Zhuang and Weiner, 1986) and the U2 snRNA base-pairs with the BPS (Parker et al., 1987; Wu and Manley, 1989) (Fig. 1.6). Thereby, a short U2 snRNA/pre-mRNA-duplex is formed, and as a result the BPS-adenosine is bulged out, freeing its nucleophilic 2' hydroxyl group for the first transesterification reaction at the 5' SS. After U4/U6.U5 tri-snRNP association, the 5' end of U2 snRNA base pairs with the 3' end of U6 snRNA, forming U2/U6 helix II (Madhani and Guthrie, 1992). The nucleotides of U6 snRNA form important regions of the spliceosome's active centre. However, at the beginning of the splicing cycle they remain inactive, forming within the U4/U6.U5 tri-snRNP and in complex B, two extended RNA helices with U4 snRNA, called U4/U6 stem I and stem II (Brow and

Guthrie, 1988; Rinke et al., 1985). Upon integration of the tri-snRNP, the stem loop I of U5 snRNA interacts with exon nucleotides near the 5' SS.

During activation of the spliceosome, U1 snRNA is destabilized from the 5' SS of the pre-mRNA by Prp28 (Staley and Guthrie, 1999) and subsequently replaced by the 5' end of U6 snRNA, which builds the base pair interactions with intron nucleotides at the 5' SS via the conserved ACAGAG box sequence (Sawa and Abelson, 1992). This interaction is important for 5' SS recognition during the first step of splicing (Fabrizio and Abelson, 1990; Lesser and Guthrie, 1993). The U4/U6 duplex is disrupted by Brr2 (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998; Staley and Guthrie, 1999), and after U4 snRNP release, short U2/U6-duplexes are formed, which are called helix Ia and helix Ib, which bring the 5' SS and BPS into close proximity for the first step of splicing (Fig. 1.7). The U2/U6 interaction can potentially form a structure corresponding to a 3-way junction. However, an alternative model for the U2/U6 interaction (a 4-way junction) was also proposed (Sun and Manley, 1995). In this model helix Ib is not formed, and instead the U6 nucleotides AGC are involved in the formation of a prolonged U6 ISL, whereas the U2 snRNA forms an additional stem loop 1 (SL1). Taking into account the highly dynamic nature of the RNA-RNA network, both conformations are possible, but recent studies showed that in human spliceosome a 3-way U2/U6 junction is formed (Anokhina et al., 2013).



**Figure 1.7 Dynamic RNA-RNA network in the spliceosome.** Exon sequences are highlighted as grey boxes and the intron sequence as a black line. Secondary structures of snRNA are shown schematically in grey or black, and the regions involved in base-pair interactions (short lines) are shown in different colours. The 5' end of an snRNA is labelled as a black dot. Solely loop I of the U5 snRNA is shown. During the transition of a pre-catalytic (left) to the catalytically activated spliceosome (right), U1 and U4 leave, and U6 and U2 are involved in new base-pair interactions. This figure was kindly provided by Dr. Bertold Kastner.

Release of U4 from U6 also leads to U6 snRNA re-folding such that the intra-molecular stem loop (U6-ISL) is formed. This ISL is extremely important for splicing catalysis, being involved in the positioning and mediating of metal ion binding (McManus et al., 2007; Yean et al., 2000).

After the first transesterification reaction, the spliceosome has to be further re-organized, to position the 3' SS and the 5' SS in close proximity for the second transesterification reaction (Smith and Konarska, 2008; Staley and Guthrie, 1998). The exact temporal coordination of these rearrangements and the conformation of RNA-RNA-interactions at this stage are still not well understood. Prior to the second transesterification reaction, U5 contacts exon nucleotides downstream of the 3' SS and binds the 5' exon to the spliceosome after the first transesterification reaction. Moreover, it brings both exons close to each other for the second catalytic step (Turner et al., 2004). Currently, information about the nature and dynamics of RNA tertiary interactions in the spliceosome is not clearly understood.

The yeast U2 snRNA toggles between two intra-molecular conformations, forming helix IIc, created from base pairing between loop nucleotides of U2 stem loop IIa and nucleotides downstream of stem loop IIb (Hilliker et al., 2007; Perriman and Ares, 2007). This stem IIc is relevant for the first step of splicing, and the refolding back to stem loop IIa is important for the second step of splicing. In humans, the formation of U2 helix IIc was thus far not detected (Anokhina et al., 2013).

### **1.1.7 RNA-based catalysis of pre-mRNA splicing**

Nuclear pre-mRNA introns and the spliceosomal snRNAs are unable to form a catalytically active structure in the absence of spliceosomal proteins. In spite of the necessity of numerous proteins for catalytic activation, much evidence indicates that the spliceosome is a ribozyme, i.e. the transesterification reactions are catalysed by snRNA components of the spliceosome, foremost U2 and U6.

Group II self-splicing introns share common features with nuclear pre-mRNAs such as consensus sequences at the 5' SS, 3' SS and branch point sequence, but some use water for the first nucleophilic attack. This similarity in mechanism and structure gave rise to the hypothesis that pre-mRNA introns and the spliceosomal snRNAs were evolutionary related to the group II self-splicing introns and therefore that pre-mRNA splicing is catalysed by RNA (Cech, 1986; Jacquier and Michel, 1990; Madhani, 2013; Sharp, 1985). Moreover, it was shown that both catalytic splicing reactions can be reversible under appropriate conditions, demonstrating further similarity between the splicing



mechanisms of these two types of introns (Tseng and Cheng, 2008). Furthermore, it was shown that U6 and U2 snRNAs are able to catalyse a splicing-related reaction with pre-mRNA-like substrates even in the absence of proteins (Valadkhan et al., 2007).

The crystal structure of a self-splicing group II intron from the alkaliphile *Oceanobacillus iheyensis* revealed a bulge-containing catalytic domain V in the ribozyme core, as well as a conserved catalytic triad coordinating two divalent metal ions. The structural similarities of the intramolecular stem loop of the U6 snRNA (U6 ISL) with the AGC catalytic triad were previously noted and evidence was provided that the U6 ISL also coordinates a metal ion (Yean et al., 2000). The divalent metal ions are bound by the U6 ISL in a manner similar to that of domain V of group II intron RNA, strongly suggesting evolutionary similarity between pre-mRNA splicing and the self-splicing of group II introns, including a two-metal ion catalytic mechanism (Toor et al., 2008).

Recently, the RNA-based catalysis of pre-mRNA splicing was verified, namely, the U6 snRNA was shown to drive both transesterification reactions by binding and positioning two divalent metal ions (Fica et al., 2013). These metal ions are thought to stabilize the leaving groups of both splicing reactions. The five catalytic metal ligands identified in the U6 snRNA structure correspond exactly to the ligands mediating catalytic divalent metal ions in crystal structures of a group II intron RNA. Thus, the group II introns and the spliceosome appear to use the identical RNA-driven catalytic mechanisms, strongly suggesting that they originate evolutionarily from a common ancestor (Fica et al., 2013). The crystal structure of yPrp8, which is a component of the U5 snRNP and is located in the catalytic centre of the spliceosome, revealed domains that were suggested to contribute to the formation of spliceosome's catalytic core, that resemble catalytic cofactors of group II introns (Galej et al., 2014). Thus, Prp8 may contribute in some way to the catalytic core of the spliceosome.

### **1.1.8 Protein composition of the spliceosome**

Proteins account for the largest share of the spliceosome's mass, and play important roles in the splicing process. The spliceosomal proteins enable and stabilize the dynamic RNA-RNA and RNA-protein interactions important for splice site recognition (Will and Lührmann, 2006). They enable all conformational and structural rearrangements needed for the catalytic activity of the spliceosome, mediating the organization of the catalytic centre for catalysis. More than 170 spliceosome-associated proteins were identified in affinity-purified, human *major* spliceosomal complexes by mass spectrometry analysis, whereas the individual complexes (i.e. B, B<sup>act</sup>, C) contain ~ 120 of the various proteins

(Wahl et al., 2009). The spliceosomal proteome contains U snRNP-associated proteins, as well as a large amount of non-snRNP proteins. Some of them interact with the spliceosome as individual proteins and the others are components of pre-formed subcomplexes.

#### **1.1.8.1 non-snRNP spliceosomal proteins**

Aside from snRNP proteins (see 1.1.4.2) a notable number of non-snRNP factors are associated with the spliceosome during the course of splicing (Wahl et al., 2009), including SR, hnRNP proteins and DExD/H-box RNA helicases. Members of the serine/arginine (SR)-rich protein family share common structural features, such as one or more amino-terminal RNA recognition motif(s) (RRMs) and a variable number of arginine/serine (RS) repetitive dipeptides in their carboxy-terminal domain (Birney et al., 1993; Graveley, 2000; Long and Caceres, 2009). Therefore, these proteins can interact at the same time with both RNA via their RRM domains and with other proteins via their RS domain (Graveley and Maniatis, 1998). For example, the recognition of exonic splicing enhancer (ESEs) regions of a pre-mRNA by SR proteins promotes spliceosome assembly at adjacent splice sites. SR proteins can also promote snRNP interactions with the pre-mRNA such as stabilisation of the U1/5'SS base pairing interaction at the initial stage of splicing assembly and thus modulate alternative splicing events (reviewed in Graveley, 2000). RS domains are also found in non-SR proteins, for instance, in U2AF-subunits (Zamore et al., 1992).

Most hnRNP proteins have RRM domains and glycine-rich domains (RG) and also regulate alternative splicing. They bind to exonic splicing silencer (ESS) elements of the pre-mRNA and impede/obstruct exon recognition by hindering the use of neighboring splice sites. HnRNP proteins counteract the positive regulators of splicing and/or recruit factors that block splicing, thereby initiating splicing repression/silencing (Smith and Valcarcel, 2000).

Another functionally important group of non-snRNP proteins is composed of the spliceosomal DExD/H-box RNA helicases (Unwindases/RNPases), which are the driving forces for all of the structural remodelling events taking place at different stages of spliceosome assembly or disassembly (Staley and Guthrie, 1998). These proteins include Sub2/UAP56, hPrp5, hPrp28, hBrr2, hPrp2, hPrp16, hPrp22 and hPrp43. They possess evolutionary conserved amino acid sequence motifs, which are essential for their functionality. These proteins modulate the structure and the function of specific RNAs and/or RNPs at different stages of spliceosome assembly (Rocak and Linder, 2004), by

catalysing the unwinding of RNA-RNA base pair interactions and/or RNA-proteins interactions, which allows the RNAs to partake in new interactions (Pyle, 2008; Staley and Guthrie, 1998). In this way, they enable structural rearrangements of the spliceosome, using the energy of ATP hydrolysis. They also facilitate the discard of non-productive splicing intermediates ensuring the fidelity of pre-mRNA splicing (Koodathingal and Staley, 2013). The aforementioned eight spliceosomal helicases are evolutionarily conserved between yeast and humans.

The DEAD-box helicases Prp5 and Sub2/UAP56 function during pre-spliceosome formation. Sub2/UAP56 catalyses the displacing U2AF and/or BPP from the BPS promoting the addition of U2 snRNP to the spliceosome (Kistler and Guthrie, 2001). Prp5 facilitates the stable binding of the U2 snRNP to the BPS (Xu and Query, 2007). The DEAD-box protein Prp28 and the DExD/H-box helicase Brr2 play key roles in the structural rearrangements of the spliceosome prior to/during activation of the spliceosome and are integral components of the U5 snRNP and the spliceosome. Prp28 catalyses the release of U1 snRNA from the 5' SS (Staley and Guthrie, 1999), apparently by dissociation of U1-C from the 5' SS (Chen et al., 2001). In humans, the phosphorylated form of Prp28 is necessary for its association with the tri-snRNP and the subsequent integration of the tri-snRNP into the spliceosome during B complex formation (Mathew et al., 2008). Brr2 catalyses the unwinding of the U4/U6 di-snRNA base pair interactions during spliceosome activation, and its activity is regulated by Prp8 and the GTPase Snu114 (Mozaffari-Jovin et al., 2013; Small et al., 2006).

Prp2 promotes a poorly understood rearrangement that converts the B<sup>act</sup> complex into the B\* complex prior to the first step of splicing (Warkocki et al., 2009). Prp16 facilitates conformational changes required for the second step of splicing, which have been poorly investigated up to now. Subsequently, Prp22 acts during the second step of splicing, and is needed for the release of the spliced mRNA (Company et al., 1991). Prp43 is necessary for the release of the excised lariat-intron from the post-splicing complex, which is accompanied by release of U2, U5 and U6, promoting the final disassembly of the spliceosome (Fourmann et al., 2013; Martin et al., 2002).

The human hPrp19/CDC5L complex (in yeast, NineTeen Complex, NTC) consists of the Prp19 protein and a number of other splicing factors that are stably recruited to the spliceosome during activation and remain associated during both catalytic steps (Bessonov et al., 2008; Makarov et al., 2002; Makarova et al., 2004). The NTC complex stabilizes the association of U5 and U6 with the spliceosome after release of U4 (Chan and Cheng, 2005; Chan et al., 2003). An additional group of hPrp19/CDC5L-related

proteins was shown to associate with Prp19/CDC5L complex (Deckert et al., 2006). hPrp19/CDC5L complex proteins interact with U5 snRNP during the activation step and remain associated during both catalytic steps. They are present in the RNP core of the C complex even after treatment with 1M NaCl (Bessonov et al., 2008). They have been detected in the 35S U5 snRNP that is proposed to be a dissociation product of a post-spliceosomal, intron-containing complex (Makarov et al., 2002).

A group of B<sup>act</sup>-specific proteins consisting of KIAA1604, Prp17, hPrp2, GPKOW, MGC20398, PPIL2, MGC23918, NY-CO-10, RNF113A and FRG1 as well as a group of proteins previously detected in B<sup>act</sup> complex associate with the spliceosome stably during B<sup>act</sup> complex formation, but they appear to be destabilized during the B<sup>act</sup> to C transition (Bessonov et al., 2010). These proteins appear to be required during the activation of the spliceosome, however, the role of these proteins in the splicing is not well understood.

The pre-mRNA retention and splicing (RES) complex consists of the Snu17, Bud13 and Pml1 proteins, and acts prior to the first step of splicing in yeast (Dziembowski et al., 2004). During splicing in yeast, it is essential for the retention of unspliced pre-mRNAs in the nucleus (Dziembowski et al., 2004).

#### **1.1.8.2 Dynamics of the spliceosome's protein composition**

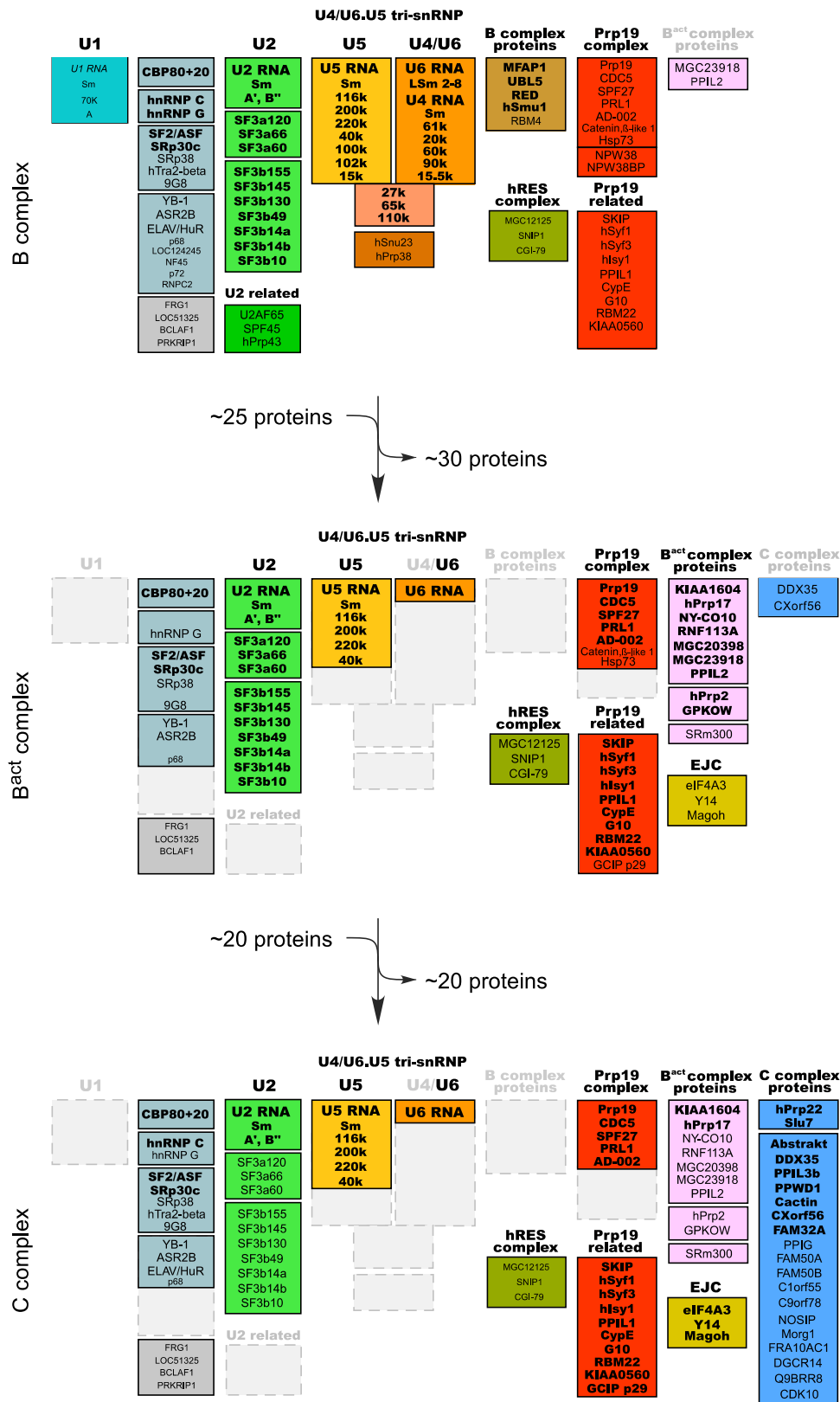
Mass spectrometry analysis of purified spliceosomes confirms that the protein composition of the spliceosome is highly dynamic, with extensive exchanges of proteins from one splicing step to another (Behzadnia et al., 2007; Bessonov et al., 2008; Deckert et al., 2006; Fabrizio et al., 2009). The human spliceosomal A complex consists of U1, U2 snRNPs and about 50 non-snRNP proteins (Behzadnia et al., 2007), several of which dissociate during the transition from the A to B complex. The human B complex contains in addition to U1 and U2 proteins, the U4/U6.U5 tri-snRNP proteins, B-specific proteins, and low amounts of the hPrp19/CDC5L complex and related proteins and proteins of the mRNA retention and splicing (RES) complex (Deckert et al., 2006). During the B to B<sup>act</sup> transition, about 30 proteins leave the spliceosome (Bessonov et al., 2010) (Fig. 1.8), including U1 and U4/U6 associated proteins, whereas about 25 other proteins incorporate, such as B<sup>act</sup>-specific proteins, including the hPrp2 helicase required for catalytic activation. The hPrp19/CDC5L complex proteins and related proteins are stabilized at this stage and present in stoichiometric amounts in B<sup>act</sup>, thereby the U5 snRNP undergoes remarkable remodelling generating a 35S form of the U5 snRNP (Makarov et al., 2002). During transition from the B<sup>act</sup> to C complex, SF3a and SF3b proteins are destabilized both in yeast and humans (Bessonov et al., 2008; Fabrizio et al., 2009; Warkocki et al.,

2009), suggesting these proteins are no longer required after the first step of splicing. B<sup>act</sup>-specific proteins are also underrepresented in the C complex. At the same time about 20 proteins are recruited to the spliceosome at this stage, including the group of C-complex specific proteins, and step II factors required prior to or during the second step of splicing, and the exon junction complex (EJC) (Bessonov et al., 2010; Bessonov et al., 2008) (Fig. 1.8). The exon junction complex (EJC) is deposited onto mRNA during splicing and is transported to the cytoplasm, regulating the cytoplasmic metabolism of spliced mRNAs. In particular, it modulates translational efficiency, detection of target mRNAs for the nonsense-mediated decay pathway, and assists in the localization of the spliced mRNA (Bono et al., 2006). Interestingly, C complexes treated with high salt reveal an RNP core consisting of only ~35 proteins, including Prp19/CDC5L proteins and Prp19-related factors, and U5 proteins such as Prp8 (Bessonov et al., 2008), indicating these proteins play important roles in maintaining the catalytically active RNA network of the spliceosome.

Fewer proteins were identified in yeast spliceosomal complexes compared to human complexes; however, nearly all of these proteins are evolutionarily conserved (Fabrizio et al., 2009). Many of the proteins identified in human but not found in the yeast spliceosome, are thought to be required for alternative splicing, which is absent in yeast. The large exchange of proteins observed during spliceosome assembly and activation is evolutionarily conserved in both human and yeast (Bessonov et al., 2008; Deckert et al., 2006; Fabrizio et al., 2009).

### **1.1.9 Posttranslational modification of proteins during splicing**

Some spliceosomal proteins undergo posttranslational modification such as reversible phosphorylation, which is involved in the regulation of spliceosome assembly. The phosphorylation of hPrp28 is essential for the stable integration of the U4/U6.U5 tri-snRNP during formation of the B complex, with hPrp28 phosphorylated by the kinase SRPK2 (Mathew et al., 2008). The tri-snRNP proteins Prp6 and Prp31 are phosphorylated by Prp4 kinase in human and this phosphorylation has been proposed to be necessary for stable tri-snRNP binding (Schneider et al., 2010). Dephosphorylation of the U1-70K protein (Tazi et al., 1993) and SR proteins (Cao et al., 1997) is required for the 1st step in mammals. The U2-associated SF3b155 protein is phosphorylated before or at the time of the first catalytic step (Wang et al., 1998) and dephosphorylated during the 2nd step of splicing, together with U5-116K/Snu114, by PP1/PP2A phosphatases (Shi et al., 2006).



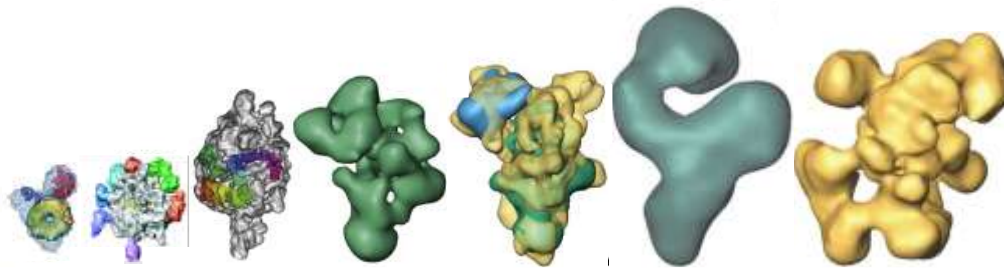
**Figure 1.8** Compositional dynamics of the human spliceosomal complexes B, B<sup>act</sup>, and C. Protein composition of human B, B<sup>act</sup>, and C spliceosomal complexes detected by mass spectrometry is shown. Proteins are grouped according to their function or association with an snRNP, protein complex or spliceosomal complex. The relative abundance of the proteins is indicated by light (substoichiometric amounts) or dark (stoichiometric amounts) lettering. This figure was kindly provided by Dr. S. Bessonov.

Additionally, some spliceosomal proteins undergo ubiquitination (Bellare et al., 2008), which appears to be required to prevent the premature unwinding of U4/U6 in the tri-snRNP (Bellare et al., 2006). The ubiquitination of Prp8 is proposed to facilitate the regulation of U4/U6 unwinding during catalytic activation, and U2/U6 unwinding during spliceosome disassembly (Bellare et al., 2008) via modulating Brr2 activity.

### **1.1.10 Structure of the spliceosome and spliceosomal snRNPs**

The structure of the spliceosome has been intensively investigated for many years. However, the determination of the structure of the spliceosome at the atomic level appears to be a very challenging task. The most obvious obstacles here are the highly dynamic nature and complex composition of the spliceosome. X-ray crystallography and NMR analyses have been restricted to particular snRNPs and spliceosomal proteins, alone or in complex with RNA or their protein interaction partner (Ritchie et al., 2009). These studies provide structural information with high resolution, which can be used to localize proteins or protein domains in the lower resolution structures obtained from EM studies. At present, structural features of the whole spliceosome can be visualized only by EM.

Electron microscopy is a powerful tool to study the architecture and morphology of spliceosomal snRNPs and their components. The resolution of the spliceosome is mainly limited by the heterogeneity of the spliceosomal complexes, their instability, and by the damage caused to the spliceosome particles during purification and preparation for cryo-EM. The GraFix method (Kastner et al., 2008) allowed the stabilization of the individual particles by mild chemical fixation during the centrifugation step prior to visualization resulted in the significant improvement of quality of samples destined for electron microscopy. Low resolution ( $\sim 12\text{--}40 \text{ \AA}$ ) 3D structures of the human spliceosomal snRNPs and spliceosomal complexes include the U1 snRNP (Stark et al., 2001), U2-associated SF3b subcomplex (Golas et al., 2003, 2005), U4/U6.U5 tri-snRNP and its components (Sander et al., 2006) and are shown in Figure 1.9. Additionally, 3D structures of endogenous yeast and human splicing complexes have been reported (Azubel et al., 2004; Ohi et al., 2007). Furthermore, 3D structures of the human complexes A, C and BAU1 have been obtained (Behzadnia et al., 2007; Boehringer et al., 2004; Golas et al., 2010). Recently, EM techniques have been greatly improved, including enhanced quality of the spliceosome preparation, chemical fixation and computational processing and refinement methods, allowing a much higher level of resolution. Single-particle cryo-electron microscopy is the most promising method to study the higher-order structure of the spliceosome.



**Figure 1.9** The currently available EM 3D structures of the spliceosome and spliceosomal components. From the left to the right: U1 snRNP, SF3b, U11/U12 di-snRNP, U5 snRNP, U4/U6.U5 tri-snRNP, BAU1 spliceosome, C complex (adapted from Lührmann and Stark, 2009).

Different labeling techniques can be applied to localize individual components in the whole spliceosome structure (Lührmann and Stark, 2009), extending the information obtained from EM structures with low to moderate resolution. These include labeling with antibodies and introducing tags into spliceosomal proteins. For example, genetic tagging was applied to map the position of key proteins in the yeast tri-snRNP (Hacker et al., 2008) and/or U6 snRNP.

2D EM views of various spliceosomal complexes purified under physiological conditions from human, *D. melanogaster* and *S. cerevisiae* have also been reported (Deckert et al., 2006; Fabrizio et al., 2009; Herold et al., 2009). Single-particle cryo-electron microscopy analysis of B, B<sup>act</sup>, and C complexes has confirmed remarkable morphological differences between these complexes (Fabrizio et al., 2009), which is consistent with the large compositional changes occurring at each stage of spliceosome assembly.

### 1.1.11 Small molecule inhibitors of pre-mRNA splicing

There is a growing interest in identifying small molecules that inhibit splicing. Small molecule inhibitors of pre-mRNA splicing can be used to stall spliceosomes at distinct and more defined stages of the spliceosome assembly, and thus potentially obtain more homogeneous/stabilised populations of spliceosomes for structural and functional studies of pre-mRNA splicing. Alternatively, they could potentially block splicing at a novel stage revealing new assembly intermediates of the spliceosome.

At present, only a limited number of inhibitors of pre-mRNA splicing have been identified. For example, the selective chelator of zinc, 1,10-phenanthroline causes a dose-dependent inhibition of the second step of pre-mRNA splicing *in vitro* (Shomron et al., 2002). Spliceostatin A (SSA), a methylated derivative of the natural chemical product FR901464, was reported to inhibit pre-mRNA splicing at submicromolar concentrations



by non-covalent interaction with the U2 snRNP associated SF3b protein complex (Kaida et al., 2007). During the A to B complex transition/conversion, SSA prevents the proper interaction of the SF3b155 protein with pre-mRNA, resulting in the unstable, non-productive base-pairing interaction between the U2 snRNA and sequences 5' of the BPS (Corrionero et al., 2011; Roybal and Jurica, 2010). Another structurally related compound, pladienolide D (Kotake et al., 2007; Yokoi et al., 2011), was also found to inhibit pre-mRNA splicing by binding to the SF3b protein complex. The synthetic derivative of pladienolide D, compound E7107, was shown to prevent stable binding of U2 snRNP to the pre-mRNA, abolishing the ATP-dependent remodelling event in U2 snRNP that exposes the branch point binding region (Folco et al., 2011). Recently, another synthetic analogue of FR901464 and derivative of SSA, sudemycin E, was shown to bind to SF3b155 and to induce U2 snRNP dissociation, and it altered alternative splicing events (Convertini et al., 2014). In addition, the natural product isoginkgetin was reported to inhibit splicing, presumably by preventing stable recruitment of the U4/U5/U6 tri-snRNP, resulting in accumulation of the pre-spliceosomal A complex (O'Brien et al., 2008). Moreover, cardiotonic steroids were identified to be alternative splicing modulators (Stoilov et al., 2008). Several small-molecule inhibitors of protein acetylation/deacetylation were found to block the splicing cycle *in vitro* at distinct stages; these include three small-molecule inhibitors of histone acetyltransferases (HATs), as well as three small-molecule inhibitors of histone deacetylases (HDACs). Two of these inhibitors, anacardic acid and garcinol, allow only the formation of A-like spliceosomes, while the other compounds inhibit spliceosome activation after incorporation of all U snRNPs into the spliceosome (Kuhn et al., 2009). A series of naphthalen-2-yl hydroxamate compounds was reported to inhibit splicing *in vitro* at the A complex stage (Patil et al., 2012). Four new compounds were confirmed to be pre-mRNA splicing inhibitors including lichen secondary metabolites, namely, psoromic and norstictic acids (Samatov et al., 2012). Psoromic acid was shown to lead to an accumulation of what appears to be mainly B<sup>act</sup> spliceosomal complexes (Samatov et al., 2012).

## 1.2 Aim of this study

The spliceosome is a large macromolecular complex that is assembled *de novo* for each splicing event on a substrate pre-mRNA in a stepwise and very dynamic manner. Its main task is to catalyse the splicing of non-coding sequences (introns) and subsequent ligation of the exons with astonishing fidelity. Large structural and compositional rearrangements occur during each step of spliceosome assembly, driven by the action of RNA helicases.

The most exciting and dramatic rearrangements occur during the activation of the B complex intermediate, which contains all snRNPs, but is still catalytically inactive, resulting in an activated spliceosome.

Small molecule inhibitors of the pre-mRNA splicing are potentially powerful tools to investigate splicing due to their potential ability to stall spliceosome assembly either at distinct known stages or at novel stages of the spliceosome assembly, leading to the isolation of highly pure and structurally more homogeneous populations of the spliceosome. At present there are only a limited number of potent inhibitors of pre-mRNA splicing that have been identified, and very little is known about the precise mechanisms by which these particular molecules inhibit splicing. Thus, I set out as a main goal of my thesis to identify novel compounds that inhibit pre-mRNA splicing *in vitro*, by screening different chemical libraries, employing a previously established screening assays. Of particular importance has been to determine the exact stage where the spliceosome is stalled in the presence of a particular compound. This has been achieved by analysing their RNA and protein composition, as well as the RNA-RNA network formed within the stalled spliceosomes. Based on this information, I have also tried to elucidate in more detail the molecular mechanism of inhibition by the identified splicing inhibitors. Furthermore, I have set out to elucidate whether the stalled complexes formed in the presence of a given compound are functionally active under splicing conditions and thus potentially can be used as a tool for functional studies such as examining the dynamic exchanges of the proteins after the inhibition step. Structural studies using different electron microscopy techniques were also carried out to see if the stalled complexes are good candidates for higher resolution EM studies. Identified compounds have been also subjected to SAR to identify structural determinants (motifs) of the molecule that confer splicing inhibition activity onto compound molecule, and that could potentially be used to increase its potency and/or be used for tagging without significant loss of its inhibition activity in order to further investigate possible targets of the compound under study in the splicing reaction. Taken together, this work should extend our knowledge about defined steps of spliceosome assembly, stalled by the presence of inhibiting compounds, and provide insights into the structure of this particular stage of the spliceosome and the potential function of proteins targeted by the inhibitor(s).

## 2 Materials and methods

### 2.1 Material

#### 2.1.1 Chemicals

Acetic acid (AcOH)	Merck, Germany
Acetone	Merck, Germany
Agarose electrophoresis grade (low melting point)	Invitrogen, UK
Ammonium peroxodisulphate (APS)	Merck, Germany
Boric acid	Sigma-Aldrich, Germany
Bovine serum albumin	Sigma-Aldrich, Germany
Bromophenol blue	Merck, Germany
Coomassie brilliant blue G-250	Serva, Germany
Coomassie brilliant blue R-250	Serva, Germany
Cacodylic acid	Sigma-Aldrich, Germany
Calcium chloride dihydrate	Merck, Germany
Chlorophorm	Roth, Germany
Creatine phosphate	Sigma-Aldrich, Germany
1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT)	Sigma-Aldrich, Germany
Dimethylsulphoxide (DMSO)	Roth, Germany
Dimethylsulfate (DMS)	FLUKA
DNA molecular weight marker	New England Biolabs, Germany
1,4-dithiothreitol (DTT)	Roth, Germany
Ethylenediamine-N;N;N',N'-tetraacetic acid, disodium salt dihydrate (EDTA)	Roth, Germany
Ethylene glycol bis(2-amoniethyl)-tetraacetic acid (EGTA)	Roth, Germany
Ethanol	Merck, Germany
Ethidium bromide (10 mg/ml) (3,8-Diamino-5-Ethyl-6-phenyl-phenanthridiniumbromid)	Roche, Germany
Fish sperm DNA (10 mg/ml)	Roche, Germany
Formaldehyde	Merck, Germany
Glycerol (87%)	Merck, Germany

Glycerol (100 %)	Merck, Germany
Glycogen	Roche, Germany
Glycine	Merck, Germany
Gluteraldehyde 25 % aqueous solution	EM Sciences
Heparin, sodium salt	Sigma-Aldrich, Germany
Hydrochloric acid (HCl)	Merck, Germany
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Roth, Germany
Igepal <sup>R</sup> CA-630 (NP-40)	Sigma-Aldrich, Germany
$\beta$ -ethoxy- $\alpha$ -ketobutyraldehyde (Kethoxal, KE)	Research Organics Inst.
Magnesium chloride hexahydrate	Merck, Germany
$\beta$ -Mercaptoethanol ( $\beta$ -ME)	Roth, Germany
Methanol	Merck, Germany
Milk powder, dry	Heirler, Germany
Phenol/Chloroform/Isoamyl alcohol (25:24:1)	Roth, Germany
Phenylmethylsulfonylfluoride (PMSF)	Roche, Germany
Ponceau S	Serva, Germany
Potassium chloride	Merck, Germany
Potassium hydroxide	Merck, Germany
2-Propanol	Merck, Germany
Protease inhibitor, EDTA free	Roche, Germany
Protein molecular weight marker (unstained, pre-stained)	Bio-Rad
Rotiphorese Gel 30 (30% acrylamide, 0.8% <i>N,N'</i> -methylene- <i>bis</i> -acrylamide)	Roth, Germany
Rotiphorese Gel 40 (38% acrylamide, 2% B <i>N,N'</i> -methylene- <i>bis</i> -acrylamide)	Roth, Germany
Roti-Phenol-Chloroform	Roth, Germany
Silver nitrate	Merck, Germany
Sodium acetate	Merck, Germany
Sodium carbonate	Merck, Germany
Sodium chloride	Merck, Germany
Sodium hydrogen carbonate	Merck, Germany
Sodium hydroxide	Merck, Germany
Sodium thiosulfate pentahydrate	Merck, Germany
Sodiumdodecylsulfate (SDS)	Serva, Germany

N, N, N', N'-Tetramethylethylenediamine (TEMED)	Sigma-Aldrich, Germany
Tricine	Sigma-Aldrich, Germany
Trichloroacetic acid (TCA)	Merck, Germany
Tris-(hydroxymethyl)-aminomethane (Tris)	VWR International, Germany
Triton X-100	Sigma-Aldrich, Germany
tRNA <i>E. coli</i>	Boehringer
Tween 20	Sigma-Aldrich, Germany
Urea	Merck, Germany
Xylene cyanol FF	Sigma-Aldrich, Germany

### 2.1.2 Antisera, monoclonal and polyclonal antibodies

Rabbit polyclonal antibodies were raised against peptides (indicated in brackets) derived from human proteins. All antibodies were affinity purified prior to use.

$\alpha$ -CDC5: C1 (aa 106-124)	Dept. Lührmann; Makarova <i>et al.</i> , 2004
$\alpha$ -hPrp17: 17.1 (aa 92-107)	Dept. Lührmann; Dr. E. Makarov
$\alpha$ -hPrp19: 19.2 (aa 176-191)	Dept. Lührmann; Makarova <i>et al.</i> , 2004
$\alpha$ -SF3b155: 155.1 (aa 99-113)	Dept. Lührmann; Will <i>et al.</i> , 2001
$\alpha$ -phospho-SF3b155: pT313 (aa 308-318)	Dept. Lührmann; Girard <i>et al.</i> , 2012
$\alpha$ -Lsm4: Supermario (aa 119-139)	Dept. Lührmann; Makarova <i>et al.</i> , 2004
$\alpha$ -MS2-MBP	Dept. Lührmann
$\alpha$ -60K: (aa 67-93)	Dept. Lührmann; Makarova <i>et al.</i> , 2004
$\alpha$ -Npw38 (aa 250-265)	Dept. Lührmann; Dr. E. Makarov
$\alpha$ -Npw38BP (aa 3-19)	Dept. Lührmann; Dr. E. Makarov
$\alpha$ -Prp38 (aa 1206-1223)	Dept. Lührmann
$\alpha$ -U5-116k: STAN (aa 1-109)	Dept. Lührmann; Fabrizio <i>et al.</i> , 1997
$\alpha$ -rabbit peroxidase	Dianova, Germany
$\alpha$ -Sm: Y12	Dept. Lührmann

### 2.1.3 Enzymes and enzyme inhibitors, and MS2-MBP protein

Calf intestine pyrophosphatase (CIP)	New England Biolabs, Germany
Complete protease inhibitor, EDTA free	Roche, Germany
<i>Pfu</i> DNA polymerase (10 U/ $\mu$ l)	Dept. Görlich, MPI-bpc
Klenow fragment of DNA polymerase I	New England Biolabs
Proteinase K	Fluka, Switzerland
Rnasin (RNase inhibitor, 40 U/ $\mu$ l)	Promega, Germany
RQ1 Dnase (1 U/ $\mu$ l)	Promega, Germany
Restriction enzymes: XbaI (20 U/ $\mu$ l), BamHI (20 U/ $\mu$ l), HincII (20 U/ $\mu$ l)	New England Biolabs, Germany
SP6 RNA polymerase	New England Biolabs, Germany
AMV Reverse Transcriptase	USB Products® Affymetrix, USA
Taq DNA polymerase	Promega, Germany
T4 DNA Ligase	New England Biolabs, Germany
T4 Polynucleotide kinase (PNK)	New England Biolabs, Germany
T7 RNA polymerase	Dept. Görlich, MPIbpc
MS2-MBP	Dept. Lührmann, MPIbpc

### 2.1.4 Nucleotides

Nucleoside-5'-triphosphate (100 mM): ATP, CTP, UTP, GTP	Promega, Germany
Deoxynucleoside-5'-triphosphate (100 mM): dATP, dCTP, dTTP, dGTP	Promega, Germany
Di-deoxynucleoside-5'-triphosphate (100 mM): dATP, dCTP, dTTP, dGTP	Promega, Germany
$m^7G(5')ppp(5')G$ cap (7-Monomethyl-diguanosine triphosphate)	Kedar, Poland
Radiolabeled nucleotides: $\alpha$ -[ $^{32}P$ ]-UTP (10 $\mu$ Ci/ $\mu$ l, 3000 Ci/mmol) $\gamma$ -[ $^{32}P$ ]-ATP (10 $\mu$ Ci/ $\mu$ l, 6000 Ci/mmol) $\alpha$ -[ $^{32}P$ ]-ATP (10 $\mu$ Ci/ $\mu$ l, 3000 Ci/mmol)	Perkin-Elmer, Germany

### 2.1.5 Oligonucleotides

DNA oligonucleotides for reverse transcription were purchased from Eurofins Genomics, Ebersberg, Germany.

Name	Sequence 5'→3'
B2	GCAAGGCTCAAAAAATTGGGT
K31	GTA CTGCAATACCAGGTCGAT
K34	TGGAACGCTTCACGAATTTGC
K91	CATTTAATATATTGTCCTCGG
M6	GGCGGTCTCGTC
M12	CTCGTCGGCAGC
acaga RT primer	TTGGTCTCCTTAAAGCTGTC

### 2.1.6 Plasmids

Name	Description	
pPM5	Plasmid encoding PM5 pre-mRNA	Bessonov <i>et al.</i> , 2008
pPM5-10	Plasmid encoding PM5-10 pre-mRNA	Bessonov <i>et al.</i> , 2010
pMINX	Plasmid encoding MINX pre-mRNA in pSP65-vector under control of SP6 promoter, Amp <sup>R</sup>	Zillman <i>et al.</i> , 1988
pMS2-MBP	Plasmid encoding MS2-MBP fusion protein in pMAL vector cloned between StuI (site destroyed) and HindIII, Amp <sup>R</sup> .	New England Biolabs, Germany

### 2.1.7 Cell lines

HeLa S3 cells (human cervical cancer cells)	GBF, Germany
HeLa S3 cells (FLAG-tagged AD002)	Grote <i>et al.</i> , 2010
HeLa S3 cells (FLAG-tagged Abstrakt)	Samatov <i>et al.</i> , 2012

### 2.1.8 Buffers

Commonly used media, buffers and solutions were prepared with Milli-Q deionized water and filter-sterilized (0.22  $\mu\text{m}$ ) or autoclaved according to Sambrook *et al.*, 1989. Solutions were autoclaved if necessary (121°C, 20 min, 1 bar). The pH was adjusted with (by the addition of) 37% HCl or 10 M NaOH if not stated otherwise.

10X TBE:	0.89 M Tris, pH 8.0 0.89 M Boric Acid 25 mM EDTA
SLAB4	50 mM Tris 380 mM glycine 0.1% (w/v) SDS
Blotting buffer (for western blot)	1.5 l SLAB4 0.6 l methanol 0.9 l deionized water
Blotting buffer (for northern blot)	25 mM NaPO <sub>4</sub> , pH 6.5
Blocking buffer	1×TBS 0.1% (v/v) Tween-20 5% (w/v) milk powder
Coomassie staining solution I	0.025% (w/v) Coomassie (R250) 50% (v/v) methanol 7.5% (v/v) acetic acid
Coomassie staining solution II	0.008 % (w/v) Coomassie (G250) 0.035 M HCl
2x DMS Stop buffer	400 mM Tris-HOAc pH 7.5 400 mM $\beta$ -mercaptoethanol 5 mM EDTA 600 mM NaOAc
Destaining solution (RNA gel)	10% (v/v) acetic acid
Destaining solution (protein gel)	40% (v/v) ethanol 10% (v/v) acetic acid
10X G150 buffer	200 mM HEPES-KOH, pH 7.9 1.5 M KCl 15 mM MgCl <sub>2</sub>
10x HY buffer:	0.5 M Tris-HCl, pH 8.4 0.6 M NaCl 0.1 M DTT



10X PBS	1.3 M NaCl 200 mM KH <sub>2</sub> PO <sub>4</sub> /K <sub>2</sub> HPO <sub>4</sub>
1X Roeder C buffer:	25 % (v/v) glycerol 20 mM HEPES-KOH, pH 7.9 420 mM NaCl 1.5 mM MgCl <sub>2</sub> 0.2 mM EDTA, pH 8.0 0.5 mM DTT 0.5 mM PMSF
1X Roeder D buffer:	10% glycerol 20 mM HEPES-KOH, pH 7.9 100 mM KCl 0.2 mM EDTA 1.5 mM MgCl <sub>2</sub> 0.5 mM PMSF 0.5 mM DTT
2X PK buffer:	200 mM Tris-HCl, pH 7.5 25 mM EDTA, pH 8.0 0.3 M NaCl or KCl 2 % (w/v) SDS
20X SSC:	3 M NaCl 0.3 M sodium citrate
2X SDS gel loading buffer:	20% glycerol 0.25 mM β-mercaptoethanol 100 mM Tris-HCl, pH 6.8 6.8% (w/v) SDS 0.05 % (w/v) Xylene cyanol 0.12 % (w/v) Bromphenol blue
TNES (RNA extraction buffer):	20 mM Tris-HCl pH 8.0 300 mM NaCl 5 mM EDTA, pH 8.0 0.1 % (w/v) SDS
4X agarose gel loading buffer:	1X TBE 30% (v/v) glycerol 0.05 % (w/v) Xylene cyanol 0.05% (w/v) Bromophenol blue

RNA loading dye:	80 % Formamide 1 mM EDTA, pH 8.0 0.05 % (w/v) Bromphenolblue 0.05 % (w/v) Xylene cyanol
10X TBS buffer:	200 mM Tris-HCl, pH 7.6 1.5 M NaCl 1 % Tween 20
Northern blotting pre-hybridisation and hybridisation buffer:	25 mM Na <sub>2</sub> HPO <sub>4</sub> , pH 6.5 6X SSC 0.5 % (w/v) SDS 50% deionized formamide 5X Denhardt's solution 100 µg/ml fish sperm DNA
5X SP6 transcription buffer:	600 mM HEPES-KOH, pH 7.5 160 mM MgCl <sub>2</sub> 10 mM spermidine 200 mM DTT
5X T7 transcription buffer:	1 M HEPES-KOH pH, 7.5 160 mM MgCl <sub>2</sub> 10 mM spermidine 200 mM DTT
1X Running buffer SDS-PAGE (Laemmli):	25 mM Tris-HCl, pH 6.8 1.92 mM glycine 1% (w/v) SDS
4X Separating gel buffer (Laemmli):	1.5 M Tris-HCl, pH 8.8 0.4% (w/v) SDS
4X Stacking gel buffer (Laemmli):	0.5 M Tris-HCl, pH 6.8 0.4% (w/v) SDS
1X Western blot transfer buffer:	20 mM Tris 150 mM NaCl 0.1 % (v/v) Tween 20
10x RT buffer:	0.5 M Tris-HCl, pH 8.4 0.6 M NaCl 0.1 M MgCl <sub>2</sub> 0.1 M DTT

### 2.1.9 Commercial kits

ECL Western Blot Detection Kit, Amersham	GE Healthcare, Germany
Prime It II Random Primer Labelling Kit	Stratagene, USA
QIAgen Plasmid Mini/Maxi Preparation Kit	Qiagen, USA
QIAquick Gel Extraction Kit	Qiagen, USA

### 2.1.10 Chromatography materials and consumables

Anti-FLAG M2 beads	Sigma, Steinheim
Amersham Hyperfilm ECL	GE Healthcare, Germany
Amersham™ Protran™ 0.2 µm NC membrane	GE Healthcare, Germany
Amylose resin	New England Biolabs, Germany
Autoradiography X-ray films BioMax MR	Kodak, USA
Chromatography columns (2.0, 10 ml)	Bio-Rad, Germany
Column MBPTrap™ HP (1 ml)	GE Healthcare, Germany
Column Illustra ProbeQuant™ G-25/G50	GE Healthcare, Germany
Column Zeba spin desalting (7K MWCO)	Thermo Fisher Scientific, USA
Dialysis membranes (MWCO 6000-8000 Da)	SpektraPor, USA
FLAG 1 x Peptide	Sigma-Aldrich, Germany
Fluotrac600 black plates (Cat. No 781077)	Greiner-Bio-One, Germany
Needles	Henke Sass Wolf, Germany
Needles BD Spinal 18GA 3.50 IN 1.2 x 90 mm	Becton Dickinson, Spain
Nylon membrane Biodyne B 0.45 µm	Pall Life Sciences, Mexico
Nitrocellulose membrane Protran BA 83	Whatman, Germany
Nesco-Parafilm	Roth, Germany
Pipettes	Gilson, USA
Pipettes	Eppendorf, Germany
Pipette tips (blue, yellow)	Roth, Germany
Probe Quant G-50 micro columns	GE Healthcare, Germany
Protein A Sepharose	GE Healthcare, Germany
Tubes (0.5 ml, 1.5 ml and 2.0 ml)	Eppendorf, Germany
Tubes (15 ml und 50 ml)	Sarstedt, Germany
Sterile filters (0.2 µm or 0.45 µm)	Sarstedt, Germany
Surgical blades	Martin, Tuttlingen
Syringes	Braun, Germany

Whatman Paper	Biometra, Germany
96-well-Plates	Greiner, Germany

### 2.1.11 Small Molecule Screening Apparati

Multidrop Combi	Thermo Fisher Scientific, USA
Robotic arm “FS”	Thermo CRS, USA
Elx405	BioTek, USA
Echo 520	Labcyte, USA
Paradigm	Molecular Devices, USA

### 2.1.12 Apparati

Autoclaves	H+P Labortechnik, Germany
Centrifuge Biofuge fresco	Heraeus, Germany
Centrifuge Biofuge pico	Heraeus, Germany
Centrifuge Megafuge 1.0R	Heraeus, Germany
Centrifuge Sorvall® Evolution RC Lynx 6000	Thermo Scientific, USA
Centrifuge Sorvall® Evolution RC	Kendro, USA
Centrifuge Sorvall® Discovery™ 90SE	Thermo Scientific, USA
Centrifuge Sorvall® WX ultra Series	Thermo Scientific, USA
Cryo electron microscope CM200 FEG	Philips, Netherlands
Cryo electron microscope Titan Krios	FEI, Netherlands
Ultracentrifuge Discovery M150	Thermo Scientific, USA
Fractionator BR-184X	Brandel, USA
Fractionator LKB-FRAC-100	Pharmacia, Sweden
Gel documentation unit	Bio-Rad, Germany
Gel electrophoresis apparatus	in-house
Gel dryer Model 583	Bio-Rad, Germany
Glassware	VWR International, Germany
Gradient Master™ Modell 106	BioComp Instruments, USA
Ice machine	Ziegra, UK
Incubators	Heraeus, Germany
“Head-over-tail” Rotor	Cole-Parmer, USA
Heating blocks	Eppendorf, Germany

Hybridization oven	Hybaid Biometra, Germany
4800 MALDI ToF/ToF	Applied Biosciences/MDS Sciex
Kodak X-Omat 2000 Processor	Kodak, USA
LTQ-Orbitrap Velos	Thermo Fisher Scientific, Germany
Magnetic Stirrer	IKA Staufen, Germany
Microfuge 5415D	Eppendorf, Germany
Microwave oven	Bosch
Milli-Q-water supply apparatus	Millipore
Orbitrap XL	Thermo Fisher Scientific, Germany
pH-meter, Orion 2-Star	Thermo Scientific, USA
pH-meter, Professional Meter PP-20	Sartorius, Germany
Phosphorimager screens	GE Healthcare, Germany
Phosphorimager Typhoon Trio+	GE Healthcare, Germany
PerfectBlue Semi-Dry Electro blotter	PeqLab
Phosphorimager Typhoon 8600	Amersham Pharmacia, Germany
Pipetman	Eppendorf, Germany
Power supply EPS 2A200	Hoefer
Power supply EPS 3501/XL	Amersham Pharmacia, Germany
Rocking platform MAX Q3000 und Q2000	Thermo Scientific, USA
S100AT4 rotor	Thermo Fisher Scientific, USA
F14-14x50cy rotor	Thermo Fisher Scientific, USA
Sorvall TST 41.14 rotor	Thermo Fisher Scientific, USA
Sorvall T660 rotor	Thermo Fisher Scientific, USA
Sorvall SureSpin rotor	Thermo Fisher Scientific, USA
Scale BP211D	Sartorius, Germany
Scale BP41OS	Sartorius, Germany
Scale XS4002S DeltaRange <sup>R</sup>	Mettler Toledo
Scale XS205 DualRange <sup>R</sup>	Mettler Toledo
Scintillation counter LS 1701/TRI-CARB 2100TR	Beckman/Packard, USA
Speed Vac Concentrator 5301	Eppendorf, Germany
Spectrophotometer Nanodrop ND-1000	Thermo Fisher Scientific, USA
Spectrophotometer Ultrospec 3000 pro	Amersham Pharmacia, Germany
Thermomixer	Eppendorf, Germany
UV-Stratalinker 2400	Stratagene, USA
Vitrobot Mark IV	FEI, Netherlands
Vortex	Scientific Industries, USA
Water bath	IKA Labortechnik, Germany

## 2.2 Methods

Unless otherwise stated, all molecular biological methods were performed as described in Sambrook, 1989. Commercial kits were used according to the instructions of the manufacturer.

### 2.2.1 Molecular biology standard methods

#### 2.2.1.1 *in vitro* transcription

[<sup>32</sup>P]-labeled, m<sup>7</sup>G(5')ppp(5')G-capped pre-mRNAs were synthesized *in vitro* by SP6 or T7 runoff transcription. A typical 100 µl transcription reaction (see below) was assembled at room temperature and subsequently incubated at 40°C (SP6) or 37°C (T7) for 2.5 h. The DNA template was hydrolysed by addition of 5 µl of 1 U/µl RQ1 DNase to the reaction mix and further incubation at 37°C for 10 min. In order to purify the RNA from free nucleotides and small fragments, the reaction mixture was mixed with an equal amount of formamide loading buffer and loaded onto a denaturing polyacrylamide gel.

Standard *in vitro* transcription reaction:

	Volume (µl)	Final concentration
5X T7 (SP6) transcription buffer	20	1X
0.1 M ATP	7.5	7.5 mM
0.1 M CTP	7.5	7.5 mM
0.1 M UTP	1.5	1.5 mM
0.1 M GTP	1.5	1.5 mM
0.151 M m <sup>7</sup> G(5')ppp(5')G-cap	3.312	5 mM
RNase inhibitor (40 U/µl)	3.75	1.5 U/□l
SP6 (T7) RNA polymerase (20 U/µl)	10	2.0 U/□l
DNA template (250 ng/µl) (PCR product or plasmid insert)	10	25 ng/□l
α-[ <sup>32</sup> P]-labelled UTP (10 □Ci/□l, 3000 Ci/mmol)	5	
H <sub>2</sub> O	29.938	
Total	100	

The RNA band was detected by autoradiography, in case of radioactive RNA or by UV-shadowing in case of unlabelled RNA, and cut out of the gel. RNA was eluted from the gel by incubating overnight in TNES buffer. RNA was Phenol / Chloroform / Isoamylalcohol (PCI) extracted and precipitated with 3 volumes of ethanol. The RNA precipitate was dissolved in sterile water. Transcribed pre-mRNA was purified from

unincorporated nucleotides by gel filtration using G50 micro columns. The radioactivity of 1  $\mu$ l of the solution was determined by counting in a scintillation counter according to the Cherenkov protocol and the specific activity of the transcript was calculated. Unlabelled RNA was transcribed according to the same protocol in the absence of  $\alpha$ -[<sup>32</sup>P]-labelled UTP, adjusting the final concentration of the cold UTP to 7.5 mM.

### **2.2.1.2 Concentration determination of nucleic acids**

To determine the concentration of nucleic acids, the absorption of an aqueous solution was measured at a wavelength of 260 nm against the reference (H<sub>2</sub>O). The nucleic acid concentration was determined using a NanoDrop spectrophotometer according to the manufacturer's instructions. The 260/280 nm ratio was used to determine the purity grade. A ratio lower than 1.8 indicates the presence of protein, phenol or other contaminants absorbing at 280 nm.

The following equations were used to determine concentrations:

$$1 \text{ OD}_{260} = 50 \text{ } \mu\text{g/ml double-stranded DNA}$$

$$1 \text{ OD}_{260} = 33 \text{ } \mu\text{g/ml single-stranded DNA}$$

$$1 \text{ OD}_{260} = 40 \text{ } \mu\text{g/ml single-stranded RNA}$$

Alternatively, the concentration of <sup>32</sup>P-labeled pre-mRNA was determined by the ratio of hot and cold UTP, the reference date of the  $\alpha$ -[<sup>32</sup>P]-UTP used and the absolute number of uridines in the transcript.

### **2.2.1.3 Agarose gel electrophoresis of nucleic acids**

Agarose gel electrophoresis was used for analysis and/or for purification of preparative amounts of nucleic acids, in particular, to separate DNA fragments after restriction enzyme digestions. Using agarose gel electrophoresis nucleic acids are first separated in an electric field, thereby the DNA fragments of different size migrate with a different speed due to the net negative charge of its sugar-phosphate backbone. Depending on the size of the DNA fragments, gels contain 0.8 – 2 % agarose (w/v) in 1x TBE buffer (0.8-1.0 % agarose for the purification of pUC-MINX-MS2 digestion products). The DNA was supplemented with 4x agarose gel loading dye and separated by electrophoresis at a constant voltage of 60-100 V in 1x TBE buffer. For detection of DNA an organic intercalating agent ethidium bromide at a concentration of 0.5  $\mu$ g/ml was added to the gel solution. DNA bands were visualized by illumination with UV-light at a wavelength of 365 nm. Commercial DNA ladders from Roche were used as a size marker.

To isolate the DNA fragments of interest from the gel, the corresponding DNA bands visualized by UV light at a wavelength of 365 nm were excised from the gel with a sterile razor blade and DNA was extracted using a QIAquick gel extraction kit (Qiagen) following the instructions of the manufacturer. Thereby, the agarose was melted, the DNA bound to the matrix, washed and eluted. The concentration of eluted DNA was determined by Nanodrop (as described in 2.2.1.2).

#### **2.2.1.4 Phenol-chloroform-isoamyl alcohol extraction**

Phenol-chloroform-isoamyl alcohol extraction (PCI-extraction) was used to separate nucleic acids from proteins. During PCI-extraction proteins are denatured by phenol and isoamyl alcohol, ensuring the complete inactivation of RNase activity. The solution was extracted with an equal volume of PCI-solution and mixed by vortexing. To separate the aqueous and organic phases, the emulsion was centrifuged for 10 min at room temperature at 13000 rpm in a microfuge. The upper aqueous phase containing nucleic acids was transferred to a new tube and nucleic acids were precipitated by the addition of 1/10 volume of 3M NaOAc (pH 5.2) and 3 volumes of ethanol in the presence of 1  $\mu$ l glycogen.

#### **2.2.1.5 Proteinase K digestion**

Proteinase K digestion was used to remove proteins from preparations of nucleic acid. In a standard proteinase K digestion reaction, 10  $\mu$ l of a splicing reaction aliquot was supplemented with 10  $\mu$ l 2x PK buffer and 2  $\mu$ l of 10 mg/ml proteinase K, and incubated for 30 min at 37°C. After the digestion, the volume was adjusted to 200  $\mu$ l with H<sub>2</sub>O, and RNA was isolated by PCI extraction (described in 2.2.1.4). The RNA pellet was dried in a Speed Vac Concentrator for 2 min, subsequently resuspended in RNA loading dye and analysed on a 0.5 mm 10-14 % denaturing polyacrylamide gel (PAGE) as described below.

#### **2.2.1.6 Denaturing Urea polyacrylamide gel electrophoresis**

Denaturing polyacrylamide gel electrophoresis in the presence of urea was used to separate RNA molecules. Urea is used as a denaturant to break the hydrogen bonds between the base pairs of the nucleic acid. Depending on the size of the RNA, the gels contained 5, 8, 10 or 14 % polyacrylamide, 8 M urea and 0.5x TBE. The gel was polymerized by addition of 30  $\mu$ l TEMED and 300  $\mu$ l 10 % (w/v) APS to 40 ml of the gel solution. The RNA samples were dissolved in formamide loading buffer, subsequently heated for 3 min at 96°C to promote denaturation of the RNA molecules prior to loading, and immediately chilled on



ice. The electrophoresis was performed at a constant voltage of 60-100 V in 0.5X TBE buffer. The RNA was visualized by silver staining, autoradiography or with a PhosphoImager.

#### **2.2.1.7 RNA silver staining**

Silver staining of RNA polyacrylamide gels was used for detecting nucleic acids and performed according to Caetano-Anollés and Gresshoff, 1994. For the fixation, the gel was incubated in a 10 % acetic acid solution for 30 min. Then, the gel was washed three times for 5 min with deionized water and incubated for 30 min in solution containing 1.5 g/L AgNO<sub>3</sub> and 0.056 % formaldehyde. The gel was then briefly rinsed twice with water and developed with solution containing 30 g/L Na<sub>2</sub>CO<sub>3</sub>, 0.056 % formaldehyde and 400 mg/L Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>. When the desired intensity was achieved, the development was stopped with 10 % acetic acid. The gel was scanned and dried on a vacuum gel dryer at 80°C for 45 min.

#### **2.2.1.8 Radioactive labeling of the 5'-end of DNA oligonucleotides**

To perform primer extension analysis, DNA oligonucleotides (primers) were radioactively labeled at their 5'-end in the presence of  $\gamma$ -[<sup>32</sup>P] ATP using T4-poly nucleotide kinase (PNK) prior to hybridization. To this end, a standard labeling reaction (10  $\mu$ l) containing 10 pmol of the DNA oligo, 1  $\mu$ l of 10x PNK buffer (New England Biolabs), 1  $\mu$ l of T4 PNK (10 U/ $\mu$ l) and 6  $\mu$ l of  $\gamma$ -[<sup>32</sup>P]-ATP (6000 Ci/mmol, 10 mCi/ml) was incubated for 45 – 60 min at 37°C. After incubation, the reaction volume was adjusted to 40  $\mu$ l with dH<sub>2</sub>O and the [<sup>32</sup>P]-labeled oligonucleotides were purified from unincorporated  $\gamma$ -[<sup>32</sup>P] ATP using ProbeQuant G-50 micro columns according to the manufacturer's instructions.

#### **2.2.1.9 Northern blotting**

Northern blot analysis was used for detecting RNA molecules in a sample via a [<sup>32</sup>P]-labeled DNA probe complementary to an RNA target sequence. First, RNA samples are separated by denaturing PAGE. Afterwards, RNAs were transferred to a nylon membrane (Biodyne B) by semi-dry blotting for 2 h at 2 mA/cm<sup>2</sup> in 0.5X TBE buffer. Alternatively, transfer occurred by blotting for 12 h at 1 mA in 25 mM NaPO<sub>4</sub>. After blotting, the nylon membrane was UV irradiated in a UV Stratalinker apparatus and pre-blocked with pre-hybridization buffer for 1 h at 42°C. Then the membrane was incubated for 12-24 h at 42°C with the [<sup>32</sup>P]-labeled DNA probe in hybridization buffer. To this end, the probe was incubated at 95°C for 5 min, chilled on ice and added to pre-hybridization buffer.

After hybridization, the membrane was washed thoroughly to avoid background signal and ensure to that the probe has bound specifically as follows: twice with wash buffer 1 (2x SSC, 0.5 % (w/v) SDS) for 10 min at room temperature, once with wash buffer 2 (2x SSC, 0.1 % (w/v) SDS) for 10 min at room temperature, and once more with wash buffer 2 for 30 min at 50°C. After washing, RNA was visualized by autoradiography.

## **2.2.2 Standard protein biochemical methods**

### **2.2.2.1 TCA precipitation of protein**

The protein samples were precipitated prior to gel electrophoresis to reduce the sample volume. For this purpose, to one volume of protein solution, 1/100 volume of 2% sodium deoxycholate (DOC) was added, which is an anionic detergent. The mix was vortexed and incubated for 30 min at 4°C. Trichloroacetic acid (TCA) was added to a final concentration of 10 %. The mix was vortexed, incubated at 4°C overnight and centrifuged 15 min at 4°C in a microfuge at maximum speed (13000 rpm). The supernatant was discarded and the pellet was washed at least twice with 1 ml of -20°C acetone, vortexed and centrifuged for 10 min at 13000 rpm in a microfuge. The pellet was dried in Speed Vac Concentrator for 2 min and resuspended in 10-20  $\mu$ l of 2x sample buffer.

### **2.2.2.2 Acetone precipitation of protein**

To one volume of protein solution 5 volumes of ice-cold acetone (-20°C) were added and vigorously mixed (vortex). The mixture was incubated for 10-12 h at -20°C and centrifuged for 15 min at 4°C in a microfuge at 13000 rpm. The pellet was washed twice with 80 % ethanol, dried in Speed Vac Concentrator for 2 min and resuspended in SDS loading buffer.

### **2.2.2.3 Denaturing SDS polyacrylamide gel electrophoresis.**

Denaturing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins according to their molecular mass. An anionic detergent, sodium dodecyl sulphate (SDS), was applied as a detergent. SDS-PAGE was adapted from Laemmli (Laemmli, 1970). Protein samples were dissolved in protein loading buffer and heated to 95°C for 5 min prior to loading on the gel. The gel consisted of a stacking gel and a separating gel. The stacking part of the gel, used to compress (stack) proteins into a thin layer before they reach the separating gel, contained 4 % polyacrylamide. The stacking gel was polymerized with 0.05 % (w/v) APS and 0.25 % (v/v) TEMED. The separating gel, which is used to separate polypeptides by size, contained usually 10 or 13 % polyacrylamide, and was polymerized with 0.033 % (w/v) APS and 0.33 % (v/v)

TEMED, and run vertically in SDS-PAGE running buffer. The gel was subsequently fixed and stained with Coomassie brilliant blue G-250 (2.2.2.4) or with silver (2.2.2.5). Protein molecular weight marker from Bio-Rad was used as a size standard.

Stacking gel:	1.66 ml Rotiphorese Gel 30 (30% acrylamide, 0.8% <i>N,N'</i> -methylene- <i>bis</i> -acrylamide) 2.5 ml 4x stacking buffer 5.84 ml dd H <sub>2</sub> O 50 µl 10 % (w/v) APS 25 µl TEMED
Separating gel (10 %):	10 ml Rotiphorese Gel 30 (30% acrylamide, 0.8% <i>N,N'</i> -methylene- <i>bis</i> -acrylamide) 7.5 ml 4x separating buffer 12.5 ml dd H <sub>2</sub> O 100 µl 10 % (w/v) APS 100 µl TEMED
Running buffer:	250 mM Tris-HCl, pH 8.8 0.1 % (w/v) SDS 0.192 M Glycine

#### 2.2.2.4 Coomassie staining of SDS-PAGE gels

Coomassie blue (triphenylmethane) dyes (R-250 and G-250) interact with proteins through interactions between dye sulfonic acid groups and protein amine groups, as well as through Van der Waals attractions. Gels are fixed in 40 % (v/v) methanol / 10 % (v/v) HoAc for 30 min and soaked in G-250 dye until the gel has a blue colour. Then the gel background is destained with 40 % (v/v) methanol and 10 % (v/v) acetic acid. The gel can be stored in 5 % acetic acid, or, alternatively, it can be scanned and dried on Whatmann paper with a vacuum gel dryer at 80°C for 45 min.

#### 2.2.2.5 Silver staining of SDS-PAGE gels

Silver staining of SDS-PAGE gels was performed as described in Blum *et al.*, 1987. During the procedure, all solutions were used at 10 times the gel volume and all steps were performed on a shaker. The protein gel was first fixed in a 50 % (v/v) methanol / 12 % (v/v) acetic acid solution for at least 30 min or overnight. After fixation, the gel was washed twice with 50 % (v/v) ethanol and once with 30 % (v/v) ethanol for 20 min each.

Then the gel was incubated with 0.8 mM Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> for exactly 60 seconds, and immediately washed three times with ddH<sub>2</sub>O for 20 sec each to remove the thiosulfate from the gel surface. For staining, the gel was incubated in a 0.012 M AgNO<sub>3</sub> / 0.026 % (v/v) formaldehyde solution for 20 min. The stained gel was then washed three times with ddH<sub>2</sub>O to remove unbound silver nitrate. The gel was rinsed briefly with a solution containing 0.56 M Na<sub>2</sub>CO<sub>3</sub>, 0.0185 % (v/v) formaldehyde, and 16 μM Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> to remove excess silver from the surface of the gel. Then the solution was replaced with a solution containing 0.56 M Na<sub>2</sub>CO<sub>3</sub>, 0.0185 % (v/v) formaldehyde, and 16 μM Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>, and the gel was developed until the protein bands showed the desired staining intensity. The developing was stopped by incubating with a 50 % (v/v) methanol / 10 % (v/v) acetic acid solution. The stained gel was stored in 5 % acetic acid, or, alternatively, scanned and dried on Whatman paper with a vacuum gel dryer at 80°C for 45 min.

### **2.2.3 Immunological Methods**

#### **2.2.3.1 Immunoblotting (Western blotting)**

Western blotting was performed to detect specific proteins in a sample. First, the proteins were separated by SDS PAGE. A pre-stained MW protein marker from Bio-Rad was used as a size standard and to control for good transfer efficiency. Then the proteins were transferred from the gel to a nitrocellulose membrane. To this end, the polyacrylamide gel was placed on the blotting membrane, previously soaked in blotting buffer, avoiding air bubbles. It was subsequently covered by Whatman paper, placed between two sponges, assembled in a Trans-blot cell (BioRad). The transfer was carried out in blotting buffer for 2 h at 60 V, 4°C, followed by staining with Ponceau S dye to check the efficiency of the transfer. Subsequently, to prevent non-specific binding of antibodies to the membrane, the membrane was incubated in blocking buffer, containing 0.1 % (v/v) Tween-20 and 5 % (w/v) dry milk powder, overnight at 4°C or for 1 h at room temperature, and washed with TBS-Tween buffer (optional). To detect proteins of interest, the membrane was incubated with a primary antibody for 2 h at room temperature and washed one time for 15 min and four times for 5 min each with TBS-Tween buffer, containing 5 % (w/v) dry milk powder, to remove unbound antibodies. Subsequently, the membrane was incubated with horseradish peroxidase-conjugated anti-rabbit or anti-mouse antibodies for 1 h at room temperature (1:50000 dilution), followed by washing with TBS-Tween (1x 15 min, 4x 5 min). Next, the proteins were detected by enhanced chemiluminescence using an ECL kit according to the manufacturer's

instructions. The luminescence was detected by exposure to ECL Hyperfilm. The film was developed using a X-Omat 2000 developing machine.

### **2.2.3.2 Immunoprecipitation**

Protein A Sepharose was washed twice with three volumes of 1x PBS (pH 8.0) and subsequently incubated with antibodies in 1x PBS buffer at 4°C overnight with head-over-tail rotation. Next, sepharose with bound antibodies was washed four times with five volumes of 1x PBS and distributed in a defined number of tubes. The splicing reaction mixture was added to the Sepharose beads in IPP buffer containing 150-400 mM salt, and incubated at 4°C for 4 h with head-over-tail rotation. Then, the Sepharose beads with bound material were washed 5 times with IPP buffer containing the desired salt concentration. RNA was recovered by proteinase K digestion and subsequent PCI extraction, and precipitated with ethanol. Precipitated RNA was separated by denaturing, 10 % polyacrylamide gel electrophoresis and subjected to Northern gel analysis.

### **2.2.4 Special Methods**

#### **2.2.4.1 Preparation of HeLa cell nuclear extract**

Nuclear extracts were prepared from HeLa cells essentially as described by Dignam *et al.*, 1983, frozen in liquid nitrogen in Roeder C buffer and stored at -80°C. Prior to use in splicing reactions, nuclear extracts were thawed and dialyzed against 50 – 100 volumes of Roeder D buffer (100 mM KCl, 1.5 mM MgCl<sub>2</sub>, 20 mM HEPES-KOH, pH 7.9) and 10 % (v/v) glycerol, twice for 2.5 h. The extract was subsequently centrifuged at 8330 rpm at 4 °C for 10 min in F14-14x50cy rotor (centrifuge Sorvall® Evolution RC Lynx 6000). The precipitate was discarded. The HeLa nuclear extract was aliquoted, frozen in liquid nitrogen and stored at -80 °C.

#### **2.2.4.2 *In vitro* splicing reactions**

For *in vitro* splicing reactions, MINX and PM5 or PM5-10 pre-mRNA substrates were used. The 5'-end of the PM5 pre-mRNA and 3'-end of MINX pre-mRNA are preceded/followed by three MS2 coat protein binding sequences which are used for purification of spliceosomal complexes (Bessonov *et al.*, 2010). A standard splicing reaction contained 10 nM pre-mRNA and 40 % HeLa nuclear extract in buffer containing 3 mM MgCl<sub>2</sub>, 65 mM KCl, 20 mM HEPES-KOH (pH 7.9), 2 mM ATP and 20 mM creatine phosphate (final concentrations). Additionally, splicing reaction contained defined concentration of one of the compounds, and the control reaction contained DMSO, which is a solvent for the compounds. The reaction mixture was incubated at 30

°C for the times indicated. After splicing, reactions were chilled on ice. To analyse RNA products of splicing, reaction aliquots were first digested with proteinase K. RNA was subsequently recovered by PCI extraction, followed by ethanol or TCA precipitation, resolved on a denaturing polyacrylamide gel and visualized by autoradiography. The formation of spliceosomal complexes was analysed by native agarose gel electrophoresis (2.2.2.4).

### 2.2.4.3 *In vitro* high throughput-screening assay of small molecules

An *in vitro* high throughput-screening assay was used (Samatov et al., 2012) to screen large chemical libraries for splicing inhibition activity (Fig. 2.1). For screening, Fluotrac600 black plates were covered with  $\alpha$ -MS2-MBP antibodies at the concentration of 0.02 mg/ml (10  $\mu$ l/well) in 20 mM HEPES-KOH (pH 7.9) shaking at 4°C overnight using an automated setup. Next, the plates were washed three times with wash buffer, containing 1 mM HEPES-KOH (pH 7.9), 0.075 mM MgCl<sub>2</sub>, 7.5 mM NaCl, 0.005 % Tween 20 (100  $\mu$ l/well). Next, 20 nl compound were added to the wells at a concentration of 50  $\mu$ M followed by 10  $\mu$ l of an ice-cold splicing reaction containing HeLa nuclear extract prepared from cells expressing a FLAG-tagged version of the Abstract protein,

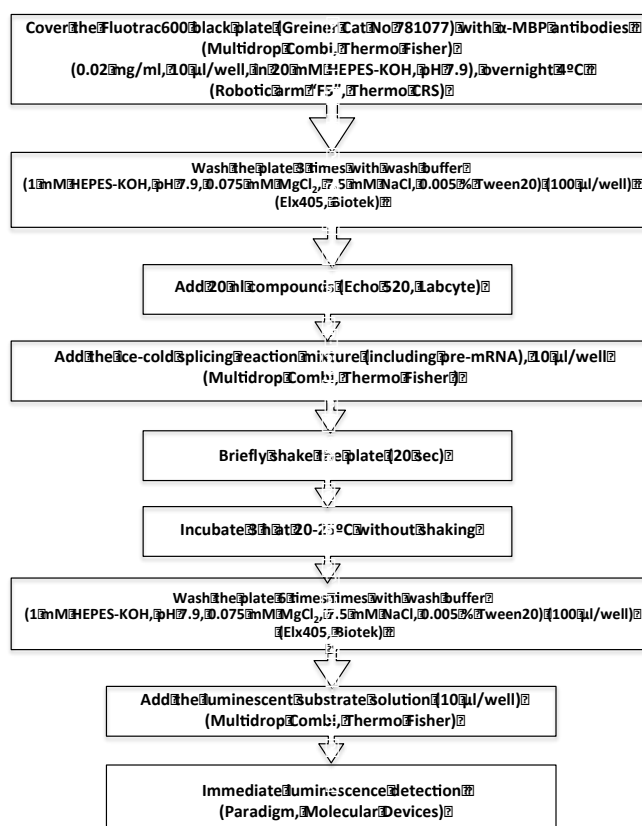


Figure 2.1 Schematic of high throughput *in vitro* screening procedure for inhibitors of pre-mRNA splicing.

which first associates with the spliceosome at the C complex stage. This FLAG-tagged Abstract protein is marked by adding peroxidase conjugated anti-FLAG antibody to the splicing reaction. After brief shaking of the plate (20 sec), the mixture was incubated for 3 h at 20-25°C without shaking. During this incubation time, the pre-mRNA substrate becomes immobilized in the wells of the microtiter plate by the bound antibodies that react with the MS2-MBP protein that binds the pre-mRNA via MS2 aptamers. Next, the plates were washed six times with wash buffer (100 µl/well) and then luminescent solution (100 µl/well) was added. The extent of C complex formation was then measured by a peroxidase luminescence reaction. The intensity of the luminescence detected reflects the amount of tagged protein present in the spliceosomal C complex, and therefore the efficiency of step I catalysis.

#### **2.2.4.4 Analysis of splicing complexes by native agarose gel electrophoresis**

Spliceosomal complexes A, B, B<sup>act</sup> and C are formed in a subsequent manner and show different mobility on a native agarose gel (Lamond *et al.*, 1987; Das and Reed, 1999; Bessonov *et al.*, 2010). Native agarose gel electrophoresis was used to resolve spliceosomal complexes for analytical purposes. Standard splicing reactions (2.2.4.1) were incubated at 30°C for the times indicated. Subsequently, heparin was added to the splicing reaction to a final concentration of 0.4 mg/ml and the mixtures were incubated for another 1 min at 30°C. Finally, 6 µl of 5x agarose loading dye was added and the samples were loaded onto a native agarose gel (26 cm x 15 cm) containing 1.5 % (w/v) low melting point agarose (Invitrogen) in 0.5 x TBE buffer. The electrophoresis was performed at a constant voltage of 40 V for 16 h in 0.5x TBE buffer. Gels were dried on a vacuum gel dryer at 60°C for 4 h and bands were visualized by exposing the gel to a phosphoimager screen.

#### **2.2.4.5 Purification of spliceosomal complexes**

Spliceosomal complexes were isolated using the MS2-MBP-affinity selection method (Deckert *et al.*, 2006). Pre-mRNA was incubated with 20 mM HEPES-KOH (pH 7.9) and a 20-fold molar excess of purified MS2-MBP fusion protein. Then, a 1 ml standard splicing reaction containing 3 mM MgCl<sub>2</sub>, 65 mM KCl, 2 mM ATP, 20 mM CP, 20 mM HEPES-KOH (pH 7.9) and 10 nM of [<sup>32</sup>P]-labeled pre-mRNA (MINX or PM5-10) was incubated for 4 (B complex), 22 (B<sup>028</sup> complex) or 170 min (B<sup>act</sup> complex) at 30°C. A 25-fold molar excess of DNA oligonucleotide complementary to nucleotides -6 to -18 and -12 to -24 relative to the 5' splice site (M6 and M12 oligos; PM5-10 complexes) was added to the reaction with B<sup>act</sup> complexes and incubated at 30°C for an additional 20 min. 1 ml of the splicing reaction was loaded onto 14 ml, linear 10-30 or 15-30 % (v/v) glycerol gradients

containing G-150 buffer. Gradients were centrifuged at 30000 rpm for 18 h (B and B<sup>028</sup> complexes) or at 28000 rpm for 16 h (B<sup>act</sup> complex) in a Sorvall TST 41.14 rotor and harvested manually in 500 µl fractions from the top. Peak fractions containing the complexes were pooled and loaded onto a column containing 250 µl packed volume of amylose beads equilibrated with G-150 buffer. The matrix was washed 6 times with 1 ml of G-150 buffer. Spliceosomal complexes were eluted dropwise with 500 µl of G-150 buffer containing 20 mM maltose. RNA was recovered from the complexes, analysed on an 8 M urea-14 % polyacrylamide gel, and visualized by silver staining and autoradiography. To analyse the protein composition of the eluted complexes, proteins were precipitated, separated by 8-10 % SDS-PAGE and subjected to mass spectrometry. Alternatively, proteins were separated by 8-10 % SDS-PAGE and analysed by Western blotting.

#### **2.2.4.6 Micrococcal nuclease treatment of HeLa nuclear extract**

To deplete splicing active HeLa nuclear extract of all endogenous snRNPs, CaCl<sub>2</sub> was added to HeLa nuclear extract in Roeder D buffer to a final concentration of 1.5 mM followed by addition of micrococcal nuclease (10 U/µl, Amersham Pharmacia) to a final concentration of 0.5 U/µl. The mixture was incubated for 15 min at 30°C and immediately transferred to ice. To quench the nuclease activity, EGTA-KOH, pH 8.0 was added to a final concentration of 4.5 mM.

#### **2.2.4.7 ATP depletion**

To deplete ATP from nuclear extract, 1 unit hexokinase and 6 mM glucose was added to 1 ml nuclear extract and incubated for 10-15 min at 30°C. Then, the nuclear extract was chilled in ice.

#### **2.2.4.8 Chasing of stalled complexes in the presence or absence of ATP**

1.0 pmole of affinity purified stalled complexes (2.2.4.5) still bound to the matrix were washed 6 times with 1 ml of G-150 buffer and subsequently incubated under *in vitro* splicing conditions either with 400 µl buffer alone or with nuclear extract treated with micrococcal nuclease under standard splicing conditions in the presence or in the absence of ATP at 30°C to 1 h. To ensure the absence of ATP in the reaction mixture, the HeLa nuclear extract was ATP depleted prior to the splicing reaction. Spliceosomal complexes were eluted dropwise using a BioRad column with 500 µl of G-150 buffer containing 20 mM maltose. RNA was recovered from the complexes, analysed on a 8 M urea-14 % polyacrylamide gel, and visualized by silver staining and autoradiography. To analyse the



protein composition of the eluted complexes, proteins were precipitated, separated by SDS-PAGE and subjected to mass spectrometry.

#### **2.2.4.9 Purification of protein complexes with anti-FLAG-beads**

To purify hPrp19/CDC5L protein complexes, splicing reactions were performed under standard conditions in the presence of HeLa nuclear extract from cells expressing FLAG-tagged AD002/Cwc15 protein. Subsequently, the salt concentration was adjusted to 250 mM. FLAG-tagged protein and/or complexes containing FLAG-tagged protein were bound to anti-FLAG-M2 agarose beads by incubating at 4°C for 4 h with “head-over-tail” rotation. Then, the agarose beads with bound complexes were extensively washed with G250 buffer, and complexes were eluted by incubating with 0.2 mg/ml FLAG peptide for 10-15 min. The eluted proteins were precipitated with TCA and analysed by SDS-PAGE.

#### **2.2.5 RNA cross-linking and structure probing**

##### **2.2.5.1 Psoralen crosslinking**

Psoralen crosslinking was used for the identification of double-stranded (base paired) regions and higher-order structures within RNA in affinity-purified spliceosomes. Psoralen is an aromatic tricyclic furocumarin, which intercalates between RNA base pairs and photoreacts with double-stranded regions via formation of cycloadducts with pyrimidine bases. This reaction is initiated by the long wavelength UV irradiation of nucleic acids at 364 nm in the presence of psoralen upon the absorption of the photon to form monoadducts, and the absorption of another photon results in the formation of a crosslink between the strands. The crosslinked RNAs were identified by Northern blot analysis by sequential probing, or by direct autoradiography of the blot to detect the position of <sup>32</sup>P-labelled, crosslinked pre-mRNA. Briefly, MS2-affinity purified spliceosomes (0.5-1.0 pmole) formed on PM5-10 pre-mRNA were subjected to long-wave UV irradiation in the presence of psoralen, and RNA was separated by gel electrophoresis. Northern blotting was performed by sequential hybridization of <sup>32</sup>P-labelled probes against U1, U2, U4, U5 and U6 prepared with a Prime It II Random Primer Labelling Kit (Stratagene). The migration of different bands detected with each probe can be compared and correlated with each other to identify the cross-linked RNA.

##### **2.2.5.2 Chemical modification**

To study RNA structure in stalled complexes, spliceosomal complexes were formed on PM5-10 and/or MINX pre-mRNAs under splicing conditions *in vitro*, and isolated via MS2 selection. Chemical modifications were then performed with 100 µl of purified

complexes (3-5 pmoles) in the presence of *E. coli* tRNA (10ng/ml) using three chemical reagents: 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT) which modifies Watson-Crick interactions mainly at positions N1-G and N3-U, dimethylsulfate (DMS) which reacts with N1-A and N3-C, and  $\beta$ -ethoxy- $\alpha$ -ketobutyraldehyde (Kethoxal) which modifies specifically N1-G and N2-G. DMS-containing reactions (0.5  $\mu$ l DMS per 100  $\mu$ l reaction volume) were incubated at 20°C for 8 min in IPP buffer, the reaction was stopped with 100  $\mu$ l 2x DMS Stop buffer and precipitated immediately with five volumes of ethanol in the presence of 1/10 volume of 3M NaOAc (pH 5.2). CMCT-containing reactions (final concentration, 70 mg/ml) were incubated at 20°C for 30 min in IPP buffer, and precipitated immediately with five volumes of ethanol in the presence of NaOAc and 1  $\mu$ l glycogen (20 mg/ml). Kethoxal-containing reactions (2  $\mu$ l Kethoxal per 100  $\mu$ l reaction volume) were incubated on ice for 1 hr in IPP buffer, and precipitated immediately with five volumes ethanol in the presence of 1/10 volume of 3M NaOAc (pH 5.2) and 1  $\mu$ l glycogen (20 mg/ml). After stopping the reactions and precipitating with ethanol, all samples were dissolved in proteinase K buffer (for Kethoxal-containing reactions potassium borate (pH 7.0) was added to 50 mM) and digested with proteinase K. RNA was isolated by phenol/chloroform extraction and finally precipitated with ethanol.

### **2.2.5.3 Primer extension**

To perform primer extension, oligonucleotides complementary to chosen sequences in the U snRNAs or pre-mRNA were radioactively labelled on their 5' end using T4 polynucleotide kinase prior to hybridization and used as primers for reverse transcriptase. Primer was added to isolated RNA and allowed to first hybridize by heating and then cooling down. Primer extension was performed (in the direction 3' to 5' end) by adding reverse transcriptase and the four dNTPs (0.1 mM each) to RNA samples, incubating them for 45-60 min at 37°C. If chemical modification takes place on the RNA template, the primer extension product will be stopped exactly one nucleotide before the chemical modification point, resulting in truncated cDNA fragments instead of fully-elongated chains. To obtain a sequencing ladder, *in vitro* transcribed snRNAs were treated in an identical manner, but in the presence of 0.5 mM dideoxy-NTPs. Reverse transcripts were separated on denaturing, 9.6 % polyacrylamide sequencing gels (8.3 M urea) and visualized by autoradiography.

### **2.2.6 Electron microscopy**

The structure of isolated spliceosomal B<sup>028</sup> complex was analysed by electron microscopy (EM). All EM experiments were performed in collaboration with Dr. Berthold Kastner, and 3D Electron Cryo-Microscopy with Prof. Dr. Holger Stark.

#### **2.2.6.1 Sample Preparation for Electron Microscopy**

For electron microscopy (EM) studies, B<sup>028</sup> spliceosomes were isolated via the MS2 affinity selection method, and the B<sup>028</sup> complexes eluted from the amylose column were subjected to a second glycerol gradient centrifugation step, using the GraFix procedure (Kastner et al., 2008), which stabilizes the individual particles by mild chemical fixation during the centrifugation step. Briefly, 15 pmol of spliceosomes were loaded onto a glycerol gradient (10-30 %) in G150 buffer supplemented with a fixation agent, namely an increasing concentration of glutaraldehyde (0 to 0.1 %). The gradients were centrifuged for 4 h 42 min at 50000 rpm in a TH660 rotor at 4°C. Fractions of 175 µl were collected from the bottom using a Pharmacia fractionator. Peak fractions were identified by scintillation counting and used for EM studies.

#### **2.2.6.2 Negative staining**

For negative staining, the particles were adsorbed onto a carbon film floated on the surface of sample solution for approximately 2-60 min, depending on the concentration of particles in the sample, using the single carbon method. Subsequently, the carbon film was transferred to the staining solution (~2 % uranyl formate in ddH<sub>2</sub>O). The carbon film was extracted from the staining solution using a copper EM grid covered with a perforated carbon film. Excess liquid was removed by blotting using Whatman filter paper. For room temperature negative staining, the EM grid was dried and stored in a dry place.

#### **2.2.6.3 Preparation of unstained samples for electron cryo microscopy**

For an unstained cryo EM preparation, the high percentage of glycerol in the sample solution was removed by buffer exchange of the sample with G-150 buffer, performed with a Zeba spin desalting column (Thermo Fischer Scientific Inc.). Sample loss during the buffer-exchange procedure was judged by scintillation counting and by negative stain electron microscopy, which also allowed a determination of the appropriate adsorption time. The particles were adsorbed onto a carbon film floated on the surface of sample solution for a defined period of time (dependent on the concentration of particles in a sample). Then, the carbon film was picked out of the particle solution by using a copper EM grid, and several microliters of ddH<sub>2</sub>O were applied to the grid. This EM grid was

subsequently mounted in a Vitrobot Mark IV. Careful blotting using Whatman filter paper was used to remove excess liquid desirable for the vitrification. Then the grid was plunged into liquid ethane cooled down by liquid nitrogen. Finally, the cryo grids were transferred and stored in liquid nitrogen until imaging.

#### **2.2.6.4 Electron microscopy**

Negative stained EM specimens were visualized with a Philips CM200 FEG Transmission Electron Microscope (TEM) at a room temperature at an acceleration voltage of 160 kV.

#### **2.2.6.5 Single-Particle Image Processing**

Individual particle images were selected semi-automatically (Sander et al., 2003a, b). For each data set, several thousands of individual single-particle images were collected. Iterative rounds of image processing were performed with IMAGIC-5 and CowEyes program. The two-dimensional (2D) image processing was performed until the result was stable, including alignment using resampling to polar coordinates (Sander et al., 2003b) and multivariate statistical classification (van Heel, 1984). The three-dimensional (3D) structure was determined by Prof. Holger Stark using the angular reconstitution method (Van Heel, 1987; van Heel et al., 2000).

#### **2.2.7 Mass spectrometry**

For MS analysis, proteins from purified complexes were separated on 4-12 % NuPAGE gradients gels (Invitrogen) and stained with Coomassie staining solution. For protein identification, lanes were cut into 23 slices and proteins were in-gel digested with trypsin and extracted as previously described (Shevchenko *et al.*, 1996). Peptides were analysed in an Orbitrap (LTQ Velos Orbitrap or Orbitrap XL) mass spectrometer under standard conditions. Proteins were identified by searching fragment spectra against the NCBI non-redundant (nr) database using Mascot as search engine. Peptides were constrained to be tryptic with a maximum of two missed cleavage sites. Carbamidomethylation of cysteines and oxidation of methionine residues were considered as variable modifications. The highest scoring peptide from each protein, as well as single hit peptides were manually inspected to eliminate any false positives in the data set. The overall numbers of sequenced peptides were counted and only fragment spectra with an ion score above 30 (Mascot) were included.

## 3 Results

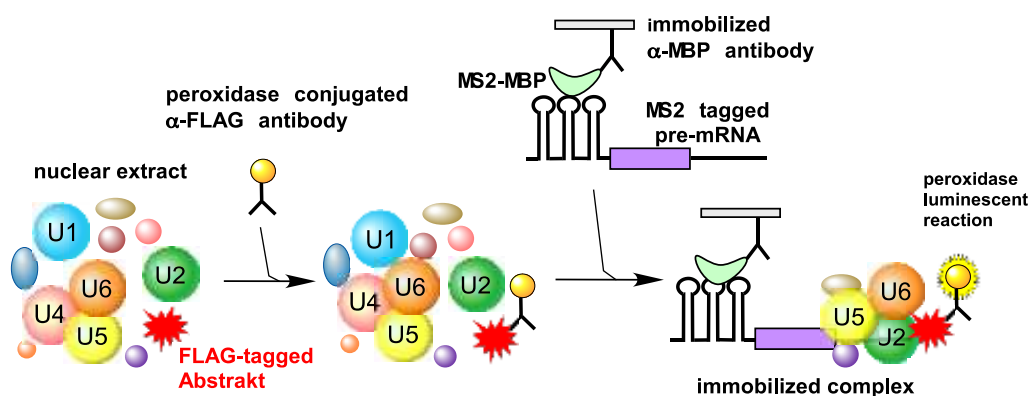
### 3.1 Identification of new small molecule inhibitors of pre-mRNA splicing

Small molecule inhibitors that block or stall splicing at a specific stage can potentially be used for isolating homogeneous/stabilised populations of spliceosomes for structural and functional studies of pre-mRNA splicing. Furthermore, they could potentially block splicing at a novel stage and thus help to isolate new assembly intermediates of the spliceosome. Finally, misregulation of alternative splicing can cause or modify the severity of many human diseases, such as autoimmune and neurodegenerative diseases and cancer. Thus, small molecule inhibitors could potentially be of therapeutic value if they also affect alternative splicing outcomes. We thus set out to identify novel compounds that inhibit pre-mRNA splicing *in vitro*.

#### 3.1.1 Search for small molecule inhibitors of pre-mRNA splicing using a high throughput screening assay

In order to screen large chemical libraries we used an *in vitro* high throughput-screening assay that was developed in our laboratory (Samatov et al, 2002) (Fig. 3.1). This assay measures the interaction of one of the spliceosomal proteins, the DEAD box ATPase Abstract, with the spliceosome and can be used to search for inhibitors of pre-mRNA splicing. This protein is first recruited to the spliceosome when the C complex is formed, before the second catalytic step of splicing, and thus can be used as a marker for catalysis of the first step of splicing. In this assay, the pre-mRNA substrate is immobilized in the wells of a microtiter plate by antibodies that react with the MS2-MBP protein that binds the pre-mRNA via MS2 loops (three aptamers). This pre-mRNA, called PM5, contains a 5'exon, 5'splice site, an intron containing a branch point sequence (BPS) and a 60 nts long polypyrimidine tract (PPT), but lacks a 3'splice site and 3'exon. Under splicing conditions, this truncated RNA substrate undergoes 5'splice site cleavage, but cannot proceed to step II due to the lack of both the 3'splice site sequence and 3'exon. For this reason splicing is stalled at the C complex step after splicing intermediates have formed, but before the second step of the splicing reaction. Splicing is performed with nuclear extract obtained from HeLa cells stably-expressing a FLAG-tagged version of the Abstract protein. This protein is marked by adding peroxidase conjugated anti-FLAG antibody to the splicing reaction. If C complexes form, the Abstract protein will be

incorporated and also immobilized on the plate even after extensive washing. The extent of C complex formation is then measured by a peroxidase luminescence reaction. If splicing is inhibited by a compound at any stage prior to step II, a loss of this signal, compared to the positive control, will be observed. The intensity of the luminescence detected reflects the amount of tagged protein present in the spliceosomal C complex, and therefore the efficiency of step I catalysis.



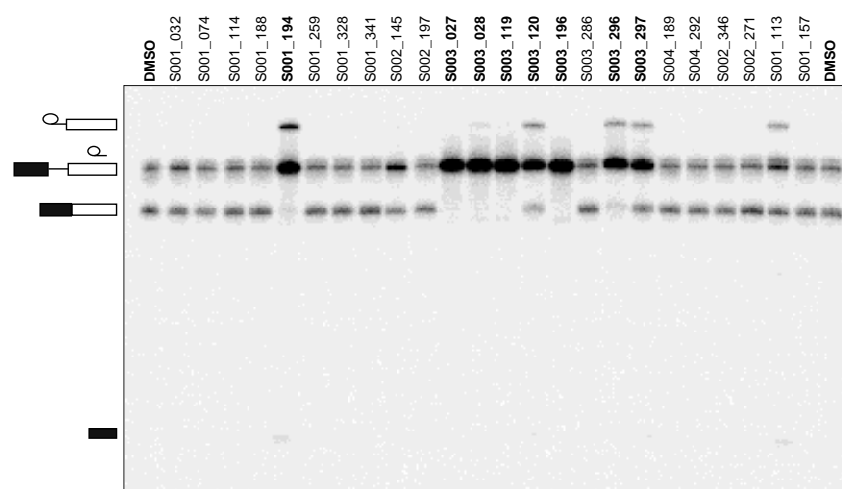
**Figure 3.1 Schematic of the high throughput *in vitro* screening assay.** A pre-mRNA substrate lacking a 3' splice site and 3' exon is shown bound by the MS2-MBP protein via three MS2 protein aptamers introduced at its 5' end. It is immobilized in the well of a microtiter plate by anti-MBP antibodies that react with the MS2-MBP protein. Splicing is performed with nuclear extract containing a FLAG-tagged Abstract protein, and this protein is recognized by peroxidase conjugated anti-FLAG antibody. The extent of C complex formation is then measured by a peroxidase luminescence reaction. The figure is adapted from Samatov et al, 2002.

To screen large chemical libraries, an automated setup using pipetting robots for most steps was established by Timur Samatov and Sonja Sievers at the COMAS (Compound Management and Screening centre) facility in Dortmund. Using this automated setup, they initially screened a chemical library of about 160,000 compounds at the concentration of 50  $\mu$ M for splicing inhibition activity (primary screen). Approximately 2000 primary hits (~1.5%) showing at least a 50 % decrease in signal intensity were subjected to the second test (retested in plates in duplicate) and only 30 reproducible hits with at least a 50% decrease in signal intensity were selected for further test in an *in vitro* splicing assay in order to eliminate any false positives. In a second round of screening, I tested an additional 12,000 compounds from the COMAS small molecule libraries, yielding 48 positive hits. I also performed a third screening with ca. 2,000 compounds from the LDC library (commercial and proprietary project-related compounds), including a sub-library of kinase inhibitors. From this screen 44 reproducible hits were selected for further testing in an *in vitro* splicing assay.

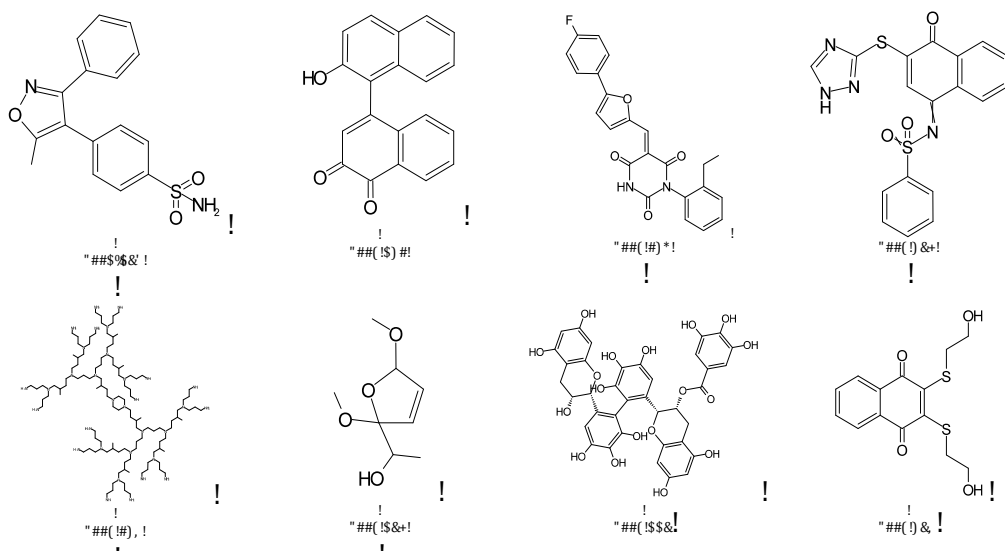
### 3.1.2 Confirmation of splicing inhibition activity in an *in vitro* splicing assay

Reproducible hit compounds from the first screen were tested in an *in vitro* splicing assay using radioactively-labelled, adenovirus-derived MINX-MS2 pre-mRNA as substrate. We tested their effect on pre-mRNA splicing using a wildtype HeLa nuclear extract. We could observe a dramatic decrease (compared to the DMSO control) in the formation of the splicing products, with an increase in the unspliced pre-mRNA and/or intermediates, on denaturing gels upon addition of some of these compounds at a concentration of 50  $\mu$ M (Fig. 3.2A). In this way, eight compounds were initially confirmed to be inhibitors of pre-mRNA splicing *in vitro*. The structures of these compounds are shown in Fig. 3.2B.

A.



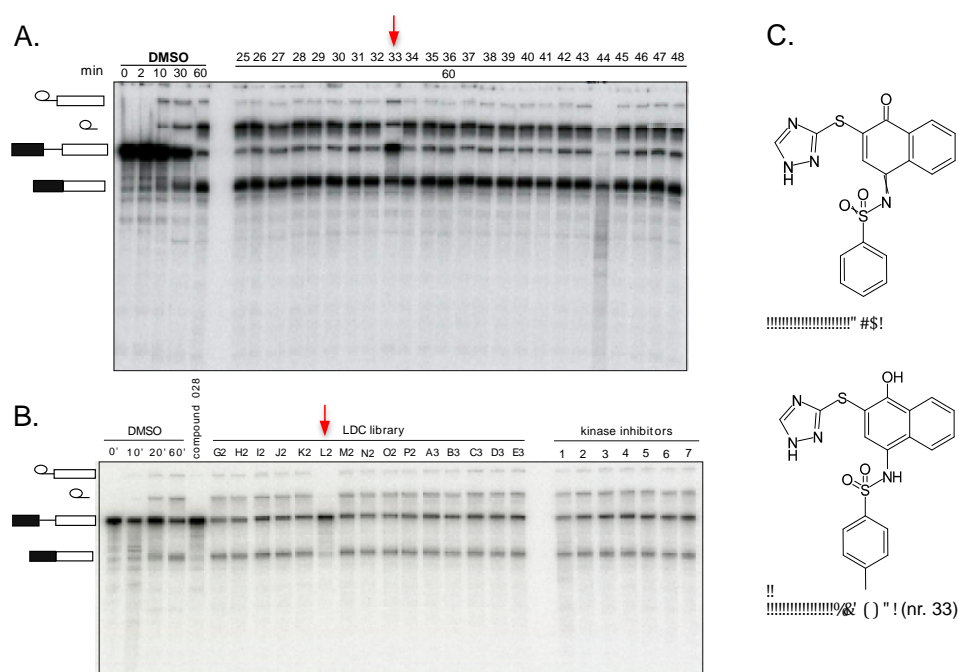
B.



**Figure 3.2 Analysis of *in vitro* splicing performed in the presence of different compounds at a concentration of 50  $\mu$ M.** A. Splicing of  $^{32}$ P-labelled MINX-MS2 pre-mRNA in HeLa nuclear extract for 60 min. in the presence of the indicated compounds. RNA was analysed by denaturing PAGE and visualized by autoradiography (DMSO: control reaction with solvent). The positions of lariat-intron/3'-exon, excised lariat intron, pre-mRNA, spliced mRNA, and 5'-exon (top to bottom) are indicated on the left. Performed by Dr. T. Samatov. B. Chemical structures of the confirmed inhibitors of pre-mRNA splicing *in vitro*.

Next, 48 primary hits from screen Nr. 2 were tested in a similar manner (Fig. 3.3A and data not shown). One of these compounds, designed 187042 (nr. 33 in Fig. 3.3A) was confirmed to inhibit pre-mRNA splicing *in vitro* at a concentration of 50  $\mu$ M after 60 min. (Fig. 3.3A). This compound is structurally related to compound 296, which is one of the confirmed hits from screen Nr. 1 (Fig. 3.3C).

Finally the 44 positive hits from screen Nr. 3 were tested in an *in vitro* splicing assay (Fig. 3.3B and data not shown). At least one of them was shown to inhibit pre-mRNA splicing at the concentration of 50  $\mu$ M. All together, 10 novel compounds were confirmed to be small molecule inhibitors of pre-mRNA splicing *in vitro*.



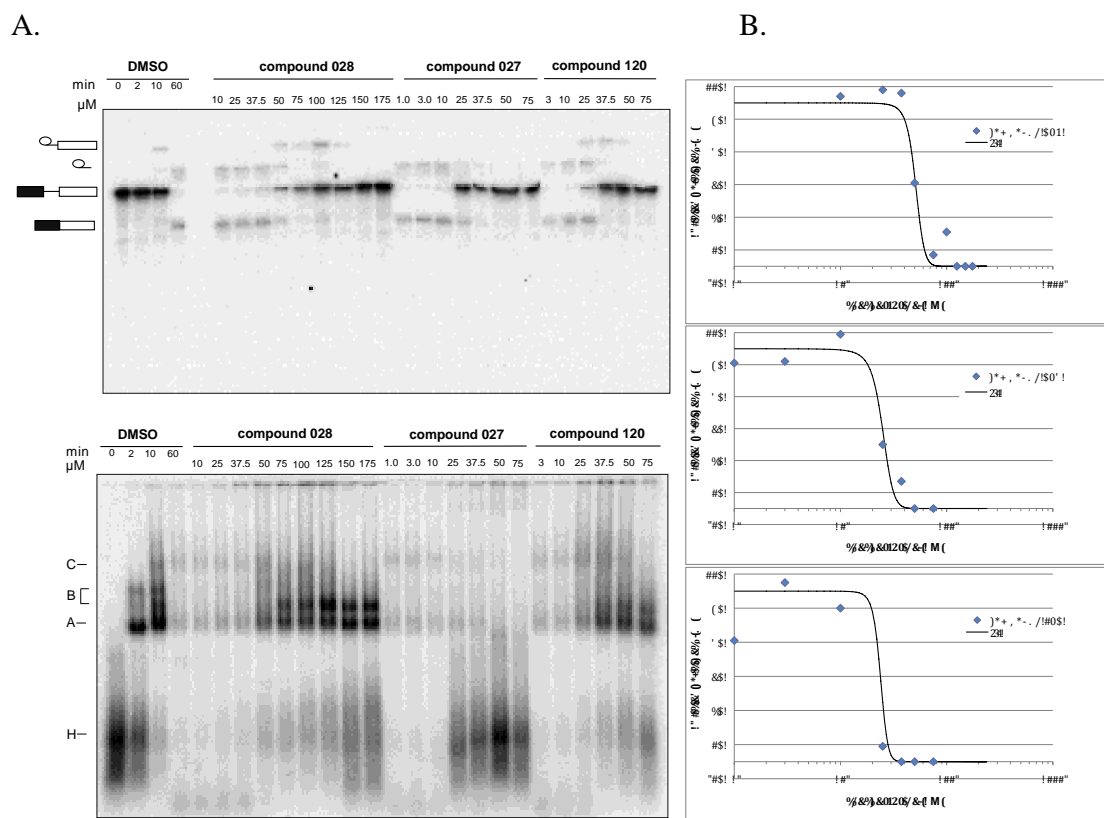
**Figure 3.3 Analysis of *in vitro* splicing performed in the presence of a subset of hit compounds from screen nr. 2 and 3 of the COMAS library at the concentration of 50  $\mu$ M.** A. Splicing of <sup>32</sup>P-labeled MINX-MS2 pre-mRNA in HeLa nuclear extract for 60 min. in the presence of compounds from screen nr. 2 (Nr. 25-48). B. Splicing of <sup>32</sup>P-labeled MINX-MS2 pre-mRNA in HeLa nuclear extract after 60 min. in the presence of compounds from screen nr. 3 (LDC library, including set of kinase inhibitors). RNA was analysed by denaturing PAGE and visualized by autoradiography (DMSO: control reaction with solvent). The positions of lariat-intron/3'-exon, excised lariat-intron, pre-mRNA, and spliced mRNA (top to bottom) are indicated on the left. C. Chemical structure of the confirmed hit 187042 (nr. 33) and structurally-related compound S003 296.

### 3.1.3 Determination of IC<sub>50</sub> values and the stage at which spliceosome assembly is inhibited

We next investigated pre-mRNA splicing inhibition in more detail. First, we performed *in vitro* splicing assays, varying the concentration of the various hit compounds. We measured the formation of the splicing products and intermediates on denaturing gels. In



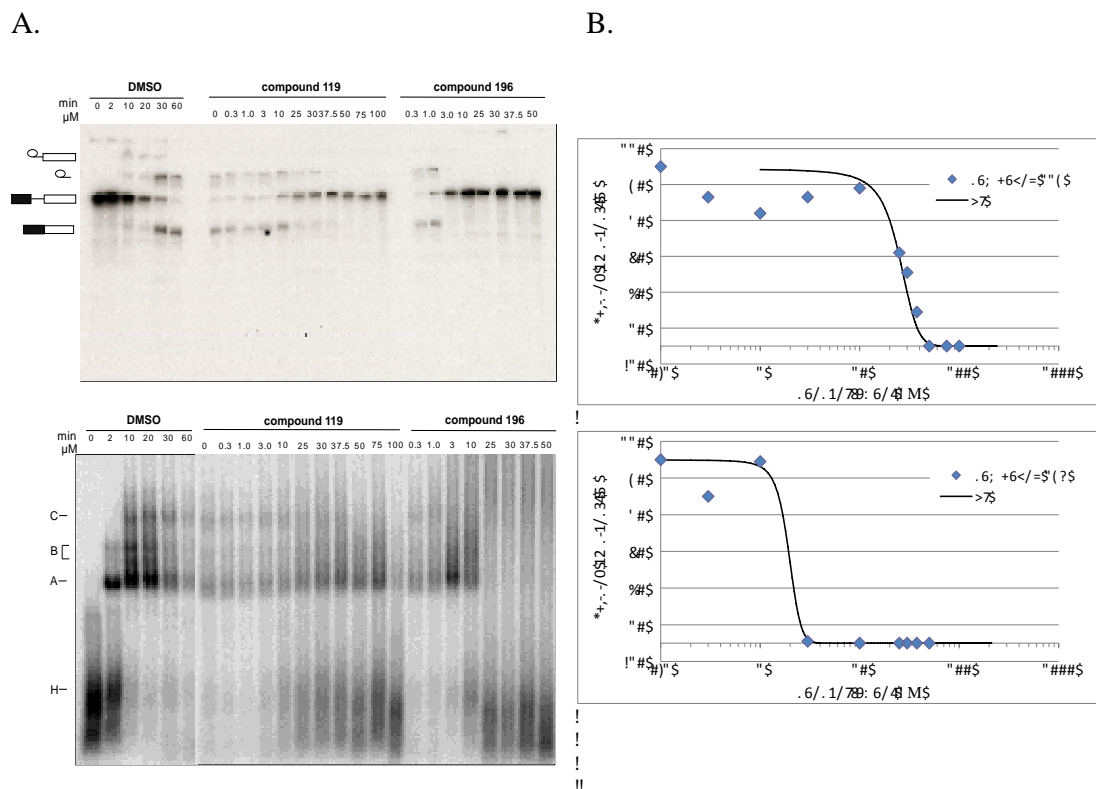
this way, IC<sub>50</sub> values (i.e. the concentration at which the second step of splicing is inhibited by 50 %) for each compound could be determined enabling the estimation of the potency of each compound. Additionally we investigated splicing complex formation on native agarose gels in order to determine at which stage each compound blocks splicing.



**Figure 3.4 Effects of increasing compound concentrations on the splicing of <sup>32</sup>P-labeled MINX-MS2 pre-mRNA *in vitro*.** A. Splicing was performed in the presence of various concentrations of three different compounds in HeLa nuclear extract for 1 h. RNA was analysed by denaturing PAGE (upper panel) and spliceosomal complexes were analysed on an agarose gel (lower panel). The positions of lariat-intron/3'-exon, lariat-intron, pre-mRNA, spliced mRNA and the positions of spliceosomal complexes C, B, A, H are indicated on the left. Bands were visualized by autoradiography. DMSO, control reaction with the solvent. B. Quantification of splicing efficiency and determination of IC<sub>50</sub> values. The splicing efficiency values for different concentrations *c* were determined from the gels shown in panel A using the IMAGE J program (circles in plots). These values were fitted using the logistic function  $f(c) = \frac{100}{1 + \exp[(c - IC_{50})/slope]}$  (solid curve). The fitting parameters, including the IC<sub>50</sub>, were determined by the nonlinear least-square fit method implemented in the Microsoft Office Excel program.

With the DMSO control (which is the solvent for the inhibitors) after 10 minutes of splicing, the lariat-3'exon intermediate (a product of the first step of splicing) could be seen on a denaturing polyacrylamide gel, and the intensity of the substrate (pre-mRNA) band was reduced to some extent to this time point (Fig. 3.4A). Products of the second

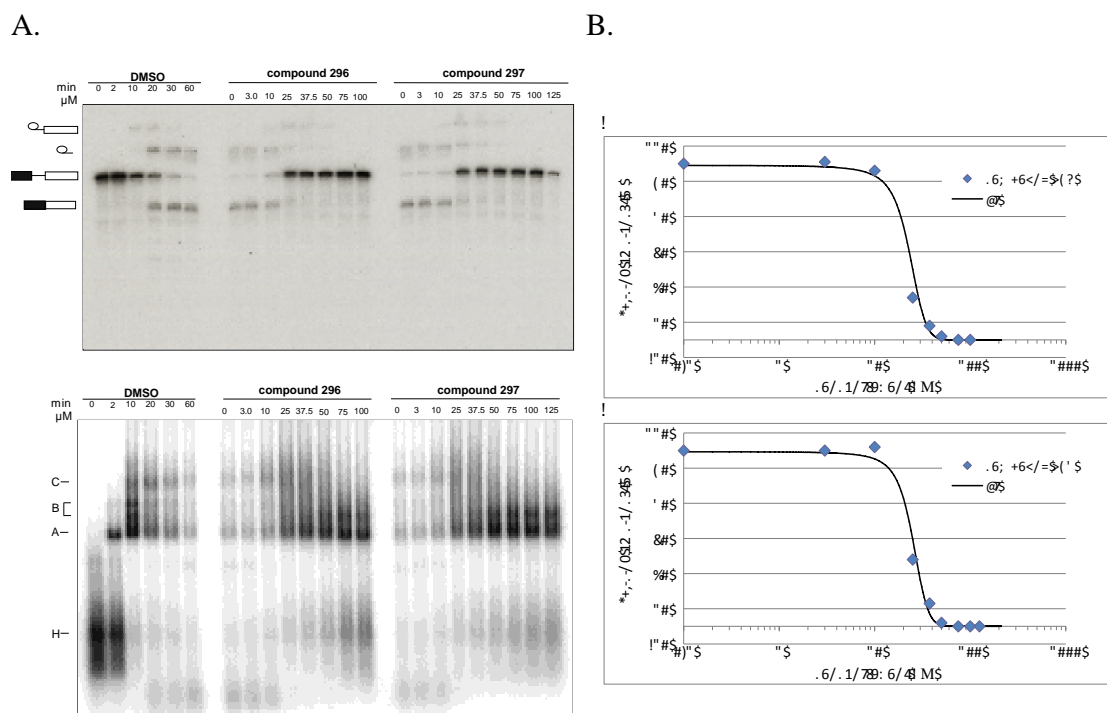
step of splicing (spliced mRNA and lariat-intron) were clearly visible after 20 min, and after 60 min predominantly spliced mRNA and excised lariat-intron were visible on the gels, with only traces of the pre-mRNA and lariat-3' exon intermediate.



**Figure 3.5 Effects of increasing compound concentrations on the splicing of  $^{32}\text{P}$ -labeled MINX-MS2 pre-mRNA *in vitro*.** A Splicing was performed in the presence of various concentrations of the indicated two compounds in HeLa nuclear extract for 1 h. RNA was analysed by denaturing PAGE (upper panel) and spliceosomal complexes were analysed on an agarose gel (lower panel). The positions of lariat-intron/3'-exon, lariat-intron, pre-mRNA, spliced mRNA and the positions of spliceosomal complexes C, B, A, H are indicated on the left. Bands were visualized by autoradiography. DMSO, control reaction with the solvent. B. Quantification of splicing efficiency and determination of  $\text{IC}_{50}$  values.  $\text{IC}_{50}$  values were determined as described in the legend to Fig. 3.4.

When increasing amounts of inhibitor 028 were added to the *in vitro* splicing reaction, unspliced pre-mRNA and/or lariat-3' exon intermediate started to accumulate at 50  $\mu\text{M}$  relative to the DMSO control, after 60 min. of splicing (Fig. 3.4A). We could observe strong inhibition of mRNA formation at 75  $\mu\text{M}$ . An accumulation of the A and a B-like complex was observed at concentrations above 75  $\mu\text{M}$  on a native agarose gel after 60 min of splicing. With the DMSO control only C complexes and trace amounts of B and A were visible after 60 min. At a concentration of 150-175  $\mu\text{M}$  splicing was inhibited completely. The  $\text{IC}_{50}$  value for compound 028 was determined to be 51  $\mu\text{M}$  (Fig. 3.4B; Table 3.1). Compound 027 inhibited mRNA formation completely at 50  $\mu\text{M}$ , with an  $\text{IC}_{50}$  equal to 25  $\mu\text{M}$ . However, only non-spliceosomal complex H (hnRNP) could be detected

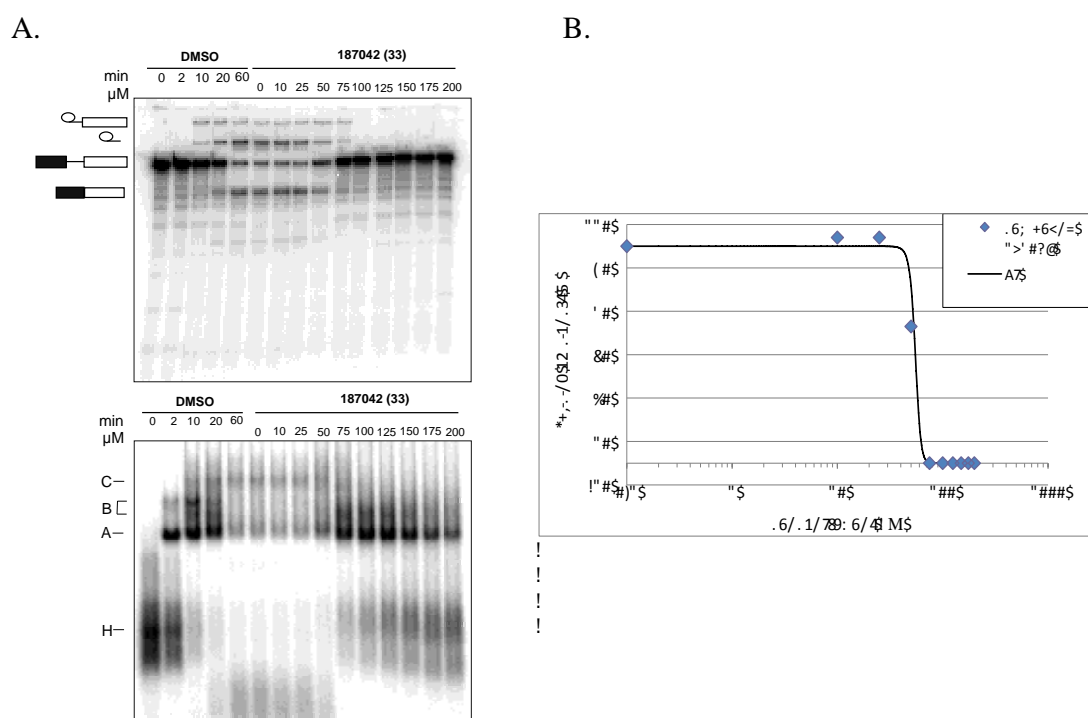
on a native agarose gel, suggesting that it probably inhibits splicing before complex A formation. Compound 120 inhibited splicing completely at 75  $\mu\text{M}$ , with an  $\text{IC}_{50}$  value of 24  $\mu\text{M}$  (Fig. 3.4B; Table 3.1), and the native gel analysis revealed an accumulation of spliceosomal complexes A and B, but to a lesser extent as observed with compound 028.



**Figure 3.6 Effects of increasing compound concentrations on the splicing of  $^{32}\text{P}$ -labeled MINX-MS2 pre-mRNA *in vitro*.** A and B. Splicing was performed in the presence of various concentrations of the indicated two compounds in HeLa nuclear extract for 1 h. RNA was analysed by denaturing PAGE (upper panel) and spliceosomal complexes were analysed on an agarose gel (lower panel). The positions of lariat-intron/3'-exon, lariat-intron, pre-mRNA, spliced mRNA and the positions of spliceosomal complexes C, B, A, H are indicated on the left. Bands were visualized by autoradiography. DMSO, control reaction with the solvent. B. Quantification of splicing efficiency and determination of  $\text{IC}_{50}$  values.  $\text{IC}_{50}$  values were determined as described in the legend to Fig. 3.4.

Compound 119 inhibited pre-mRNA splicing fully at the concentration of 75  $\mu\text{M}$  with an  $\text{IC}_{50}$  of 25  $\mu\text{M}$  (Fig. 3.5; Table 3.1), and native gel analysis revealed the accumulation of spliceosomal complex A, but at low amounts compared to the 2 min. DMSO control (Fig. 3.5). Compound 196 was found to inhibit splicing with high potency, blocking completely at the concentration of 10  $\mu\text{M}$  and exhibiting an  $\text{IC}_{50}$  of 2  $\mu\text{M}$  (Fig. 3.5; Table 3.1). However, mainly A complexes accumulated at 3  $\mu\text{M}$ , and at higher concentrations no splicing complex formation was observed.

Compounds 296 and 297 showed similar inhibition patterns, with IC<sub>50</sub> values of 23 and 25 μM, respectively (Table 3.1), and both A and B spliceosomal complexes accumulating at a concentration of 75 μM or higher (Fig. 3.6). Compound 187042, which is structurally related to compound 296 (Fig. 3.3B), inhibited splicing with an IC<sub>50</sub> value of 55 μM (Fig. 3.7A), indicating that it is approximately two times less potent. In contrast to compound 296, mainly complex A accumulated in the presence of 75 μM or higher concentrations of compound 187042. Finally, compound L2 from LDC library inhibited pre-mRNA splicing *in vitro* with an IC<sub>50</sub> value of 40 μM, blocking splicing mainly at the A complex stage (data not shown).



**Figure 3.7 Effects of increasing compound concentrations on the splicing of <sup>32</sup>P-labeled MINX-MS2 pre-mRNA *in vitro*.** A. Splicing was performed in the presence of various concentrations of the indicated compound in HeLa nuclear extract for 1 h. RNA was analysed by denaturing PAGE and spliceosomal complexes were analysed on an agarose gel. The positions of lariat-intron/3'-exon, lariat-intron, pre-mRNA, and spliced mRNA, and the positions of spliceosomal complexes C, B, A, H are indicated on the left. Bands were visualized by autoradiography. DMSO, control reaction with the solvent. B. Quantification of splicing efficiency and determination of IC<sub>50</sub> values (see the caption of Fig. 3.4 for details).

Several of these compounds were not analysed further due to either the lack of accumulation of splicing complexes in their presence (e.g. S003\_027 in Fig. 3.2) or the inactivity of the subsequent batches of the compounds (e.g. S001\_194 in Fig. 3.2). Or in some cases, the structures of the compounds suggested that they might be non-specific inhibitors of splicing (e.g. S003\_027 in Fig. 3.2).

**Table 3.1 Summary of IC<sub>50</sub> values determined for small molecules confirmed to be inhibitors of pre-mRNA splicing *in vitro* and the stage at which spliceosome assembly is inhibited.** The splicing efficiency values for different concentrations  $c$  were determined from the gels using the IMAGE J program. These values were fitted using the logistic function  $f(c) = \frac{100}{1+\exp[(c-IC_{50})/slope]}$ . The IC<sub>50</sub>, was determined by the nonlinear least-square fit method implemented in the Microsoft Office Excel program.

Compound	IC <sub>50</sub> value	Splicing complexes accumulating
028	51 μM	A and B-like complexes
027	25 μM	Non-spliceosomal complex H
120	24 μM	A and B-like complexes
119	25 μM	A complexes
196	2 μM	A complexes
296	23 μM	A and B-like complexes
297	25 μM	A and B-like complexes
187042	55 μM	Mainly A and B-like complexes

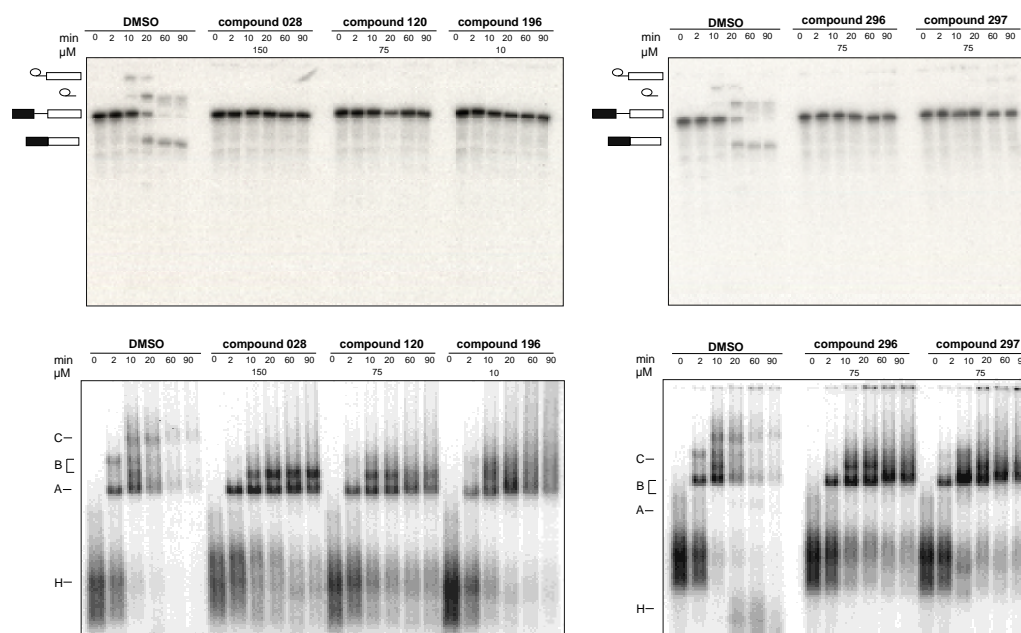
## 3.2 Purification and characterization of spliceosomal complexes stalled with different splicing inhibitors

### 3.2.1 Time course of splicing complex formation at the lowest concentration at which splicing is completely inhibited

Several hit compounds inhibited pre-mRNA splicing *in vitro* and led to the accumulation of spliceosome assembly intermediates. With the goal of affinity purifying splicing complexes stalled by these inhibitors, we next analysed splicing complex formation as a function of time. For these studies we used a concentration where the first and second step of splicing were completely inhibited. These data allowed us to determine the time point when one or more complexes had accumulated to the highest amount.

After 2 min of splicing, mainly A complexes were seen on the native agarose gels with the DMSO control, and after 10 min of splicing predominantly B complexes and lower levels of C complexes (Fig. 3.8). After 20 min, A and B complexes were converted into C complexes and after 60 min only traces of the C complexes are still visible. In the presence of 150 μM of compound 028, the kinetics of splicing complex assembly differs dramatically. Here, after 10 min, A complexes and complexes migrating faster than B are mainly observed. These complexes appear to accumulate to the highest amount after 20 min of splicing. Similar results were obtained for compounds 120, 296 and 297, which

were tested at 75  $\mu\text{M}$ , except that at later time points mainly complex A was observed with compounds 296 and 297. At a concentration of 10  $\mu\text{M}$ , an accumulation of mainly spliceosomal complex A, which peaked at the 20 min time point, was seen with compound 196. Based on these results, we could choose the splicing incubation time, where the most spliceosomal complexes accumulated for each compound.

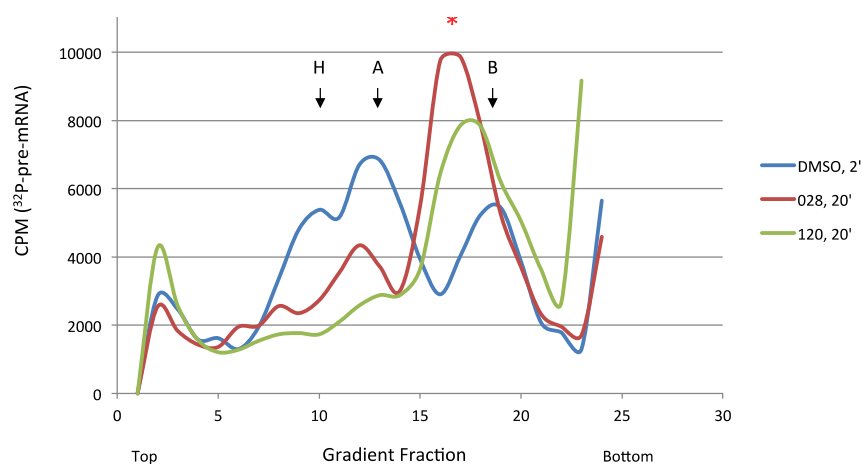


**Figure 3.8 Kinetics of splicing complex formation with  $^{32}\text{P}$ -labeled MINX-MS2 pre-mRNA in the presence of the indicated concentrations of various compounds.** Splicing was performed for 0 to 90 min in the presence of an inhibitor at the concentration resulting in full inhibition of pre-mRNA splicing in HeLa nuclear extract. Reactions were stopped at the indicated time points. RNA was analysed by denaturing PAGE (upper gels) and splicing complexes were analysed on an agarose native gel (lower gels), and visualized by phosphorimaging (DMSO: control reaction with the solvent). The positions of the lariat-intron/3'-exon, lariat-intron, pre-mRNA, and spliced mRNA (top to bottom), and the positions of spliceosomal complexes C, B, A, H are indicated on the left.

### 3.2.2 Sedimentation behaviour of splicing complexes stalled with compounds 028, 120, 296 and 297

We next checked the sedimentation behaviour of complexes stalled in the presence of various small molecule inhibitors on glycerol gradients on an analytical scale. The control splicing reaction was performed in the presence of DMSO and incubated for 2 min, during which mainly A and B complexes were formed. Splicing mixtures were loaded on to glycerol gradients. Splicing complexes formed on  $^{32}\text{P}$ -labelled MINX-MS2 pre-mRNA were then separated by glycerol gradient centrifugation and fractionated from the top. The cpm of each gradient fraction was determined and plotted. The gradient profile of the

DMSO control revealed two peaks in fractions 12-14 and 18-20 that sediment with S-values of approximately 30 and 40-45S, respectively, which correspond to the S-values of spliceosomal complexes A and B, respectively (Fig. 3.9). When the splicing reaction was performed in the presence of compound 120 (75  $\mu$ M) for 20 min, a main peak in fractions 16-19 was detected. In contrast, when splicing was performed for 20 min. in the presence of compound 028 (150  $\mu$ M), complexes sedimenting similar to the A complex were observed plus complexes that sediment slightly slower than the B complex (Fig. 3.9). This shift was a first hint that the B-like complex accumulating in the presence of compound 028 is different from wildtype B/B<sup>act</sup> complexes that have higher S-values. When splicing was performed for 20 min in the presence of 75  $\mu$ M compound 296, a main peak was seen in fractions 18-20 and thus complexes migrating like B/B<sup>act</sup> complexes in the DMSO control were formed (Fig. 3.11A). Interestingly, spliceosomal complexes formed in the presence of 75  $\mu$ M compound 297 after 10 min appeared to have a higher S-value than the B complexes, peaking ~ 2 fractions closer to the bottom of the gradient (Fig. 3.11A). Native gel analysis of splicing complex formation in the presence of these compounds revealed predominantly A complexes (Fig. 3.8), which seems to contradict the lack of an A complex peak when analysed on glycerol gradients. However, the B-like complexes observed on gradients after incubating with compound 296 or 297 may be unstable and thus dissociate during native gel analysis which is performed in the presence of heparin which can destabilize loosely associated components.



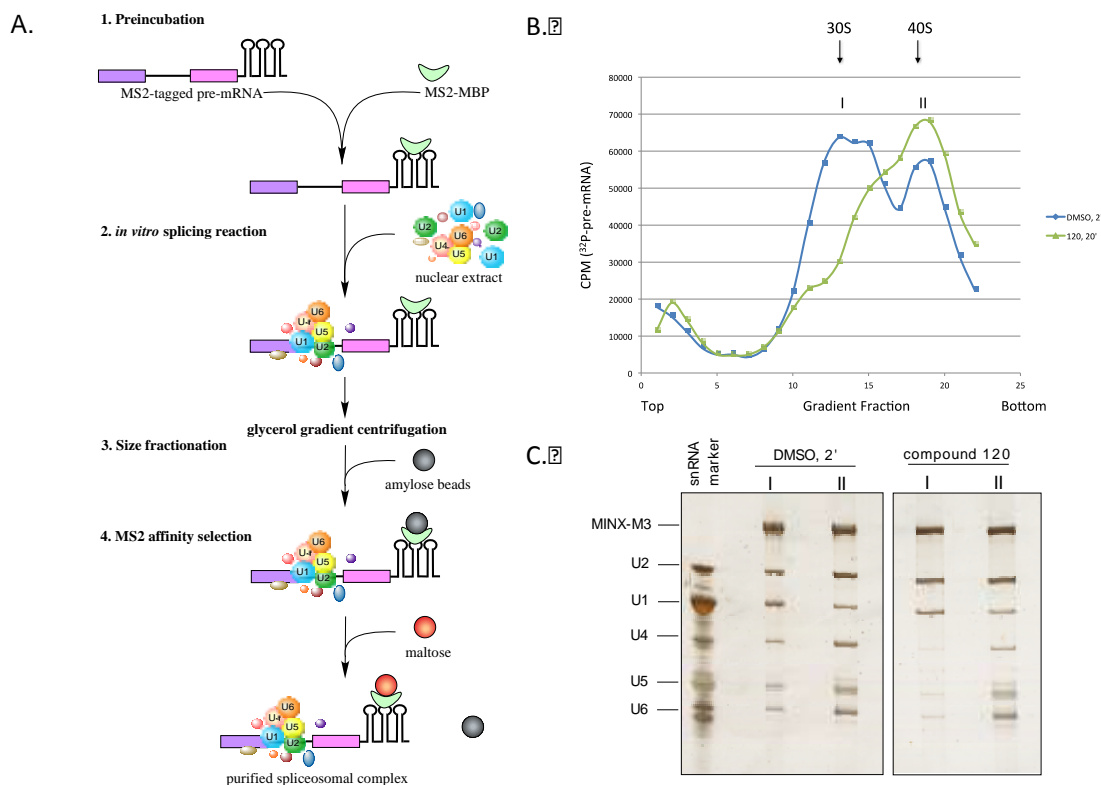
**Figure 3.9 Glycerol gradient sedimentation profiles of splicing complexes formed in the presence of compounds 028 (150  $\mu$ M) and 120 (75  $\mu$ M).** In vitro splicing was performed under standard splicing conditions. Spliceosomal complexes formed on radioactively-labelled MINX pre-mRNA were separated on a linear 10-30% glycerol gradient containing 150 mM salt, and fractionated from the top. The radioactivity of spliceosomal complexes detected in each gradient fraction is plotted. Peaks corresponding to the H, A and B spliceosomal complexes are indicated with arrows. \* = “B-like” complexes formed in the presence of compound 028.

### 3.2.3 Affinity-purification of spliceosomal complexes stalled in the presence of selected compounds

To characterize the complexes stalled with various compounds under study in more detail, we purified preparative amounts of them via the MS2 affinity selection method (shown schematically in Fig. 3.10A). To this end, *in vitro* splicing reactions were carried out using a radioactively-labelled, adenovirus-derived MINX pre-mRNA substrate containing three MS2 bacteriophage coat protein-binding aptamers at its 3' end (MINX-MS2). This MS2-tagged MINX pre-mRNA substrate was first bound by MBP-MS2 fusion protein, consisting of MS2 bacteriophage coat protein and the maltose-binding protein MBP. *In vitro* splicing was then performed under standard splicing conditions in the presence of HeLa nuclear extract. Spliceosomal complexes that form were then separated by glycerol gradient centrifugation and the distribution of <sup>32</sup>P-pre-mRNA determined by Cherenkov counting. Complexes from individual gradient peaks (30S region for complex A and ~40S region for complex B) were then incubated with amylose beads, which are bound by MBP, and after washing, complexes were eluted with maltose, which competes with amylose for binding to MBP. The RNA and protein composition of the eluted complexes was then analysed by denaturing PAGE (RNA) and LC-MS/MS mass spectrometry (protein).

We first purified complexes formed in the presence of 75  $\mu$ M compound 120. The glycerol gradient profile of splicing complexes formed in the presence of inhibitor 120 revealed two gradient peaks: peak I in a ~30S region of the gradient, and peak II in a ~40S region of the gradient (Fig. 3.10B). Subsequently, complexes were affinity purified from peak I and peak II, and the RNA composition from each peak was analysed by denaturing PAGE and subsequent silver staining (Fig. 3.10C). RNA analysis of peak I revealed mainly the presence of the pre-mRNA, U1 and U2 snRNAs, indicating that they are A complexes. The A complex peak of the DMSO control was found to be contaminated by B complexes, based on the presence of low levels of U4, U5 and U6 snRNAs and tri-snRNP proteins as determined by mass spectrometry (Table 3.2). This hindered a comparison of the protein composition of A<sup>DMSO</sup> with A<sup>120</sup> complexes. Affinity-purified complexes from the 40-45S region (peak II) of the DMSO control and compound 120 contained pre-mRNA, U1, U2, U4, U5 and U6 snRNA, indicating they are mainly B complexes. Mass spectrometry of 40S complexes stalled by compound 120 confirmed that they are B complexes, because the protein composition of B<sup>120</sup> and B complexes formed with DMSO, were nearly identical (Table 3.2).

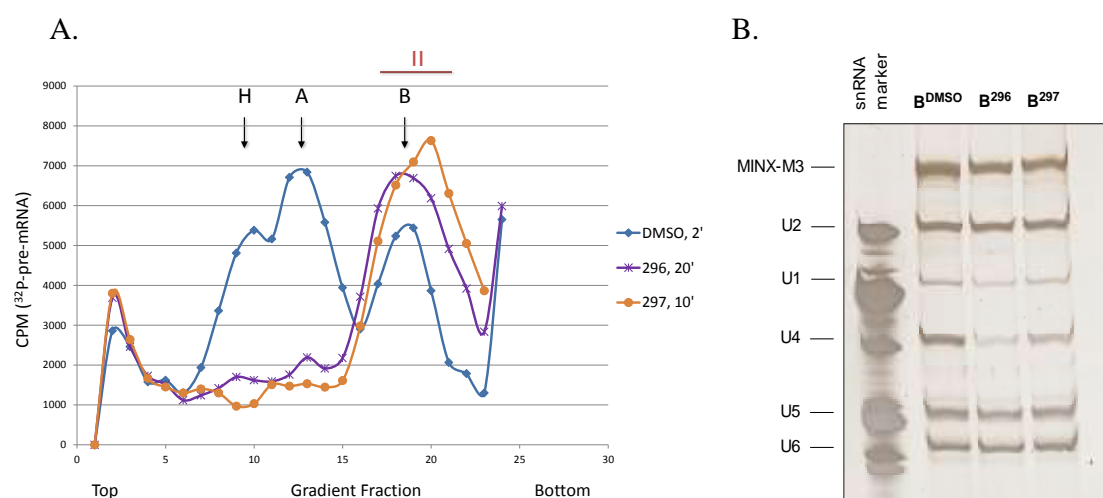




**Figure 3.10 Purification of complexes stalled with compound 120.** A. Schematic view of MS2-affinity purification of human spliceosomal complexes. MS2-tagged pre-mRNA substrate was pre-incubated with MS2-MBP fusion protein. Subsequently, splicing was performed in the presence of 40 % HeLa nuclear extract under standard splicing conditions. Spliceosomal complexes formed were then separated by size on a linear 10 to 30 % glycerol gradient and detected via their  $^{32}\text{P}$ -labelled pre-mRNA. Spliceosomal complexes from each peak were affinity-selected using amylose beads and eluted with maltose. B. Glycerol gradient sedimentation profiles of the spliceosomal complexes formed in control (DMSO) or compound 120 containing splicing reactions. C. RNA composition of MS2-affinity purified splicing complexes formed in the presence of DMSO or compound 120. RNA recovered from splicing complexes affinity-purified from peaks I and II was analysed by denaturing PAGE and detected by silver staining. RNA identities are indicated at the left.

We next affinity-purified complexes stalled by compounds 296 and 297 (Fig. 3.11). RNA analysis of spliceosomal complexes formed in the presence of compounds 296 and 297 recovered from peak II revealed equimolar amounts of pre-mRNA and U2, U5, U6 snRNAs, with U1 and U4 snRNA underrepresented to some extent, suggesting that these “B-like” complexes are different from the B complex formed in the presence of DMSO ( $\text{B}^{\text{DMSO}}$ ), but their snRNA composition is more similar to the  $\text{B}^{\text{act}}$  complex (Fig. 3.11B). The protein composition analysis of these two complexes was determined by LC-MS/MS mass spectrometry (Table 3.2). Complexes stalled in the presence of compounds 296 and 297 contain similar amounts of U2 and U5 snRNP proteins as complex B, with U4/U6 and tri-snRNP proteins somewhat underrepresented based on the number of peptides detected by MS. Proteins of the Prp19/CDC5L complex and the Prp19-related proteins

Skip and RBM22 appear to be incorporated reasonably well, but there is only little recruitment of typical B<sup>act</sup>-specific proteins, such as RNF113 and Prp17. This is a first indication that compounds 296 and 297 may stall spliceosomes during activation, specifically after stable incorporation of Prp19/CDC5L complex proteins, but before recruitment of B<sup>act</sup>-specific proteins. However, B-specific proteins and even second step factors (in the case of 296) were also present in these stalled complexes. This suggests that these purified complexes contain a mixture of B, B<sup>act</sup> and later complexes. Thus, further optimization of the purification of these complexes is necessary to characterize them in more detail.



**Figure 3.11 RNA composition of MS2-affinity purified B/B-like splicing complexes formed in the presence of compounds 296 and 297.** A. Glycerol gradient sedimentation profiles of splicing complexes formed in the presence of compounds 296 and 297. *In vitro* splicing was performed under standard conditions. Spliceosomal complexes formed on radioactively-labelled MINX-MS2 pre-mRNA were separated on a linear 10-30% glycerol gradient containing 150 mM salt and fractionated from the top. The gradient profiles were determined by counting the <sup>32</sup>P-MINX radioactivity in each fraction. Peaks corresponding to the H, A and B of spliceosomal complexes are indicated with arrows. B. RNA recovered from splicing complexes affinity-purified from peak II was analysed by denaturing PAGE and detected by silver staining. RNA identities are indicated at the left.

### 3.2.4 Isolation of splicing complexes stalled with compound 028

Finally, we affinity-purified complexes formed in the presence of compound 028 and investigated their RNA and protein composition. To get a better separation of complexes formed in the presence of compound 028 (B<sup>028</sup>) from other complexes in the glycerol gradient, we first optimized the gradient centrifugation conditions, namely, we changed the glycerol gradient from 10-30% to 15-30%, and optimized sedimentation parameters so, that the B<sup>028</sup> complexes sediment closer to the bottom (Fig. 3.12A). We then isolated

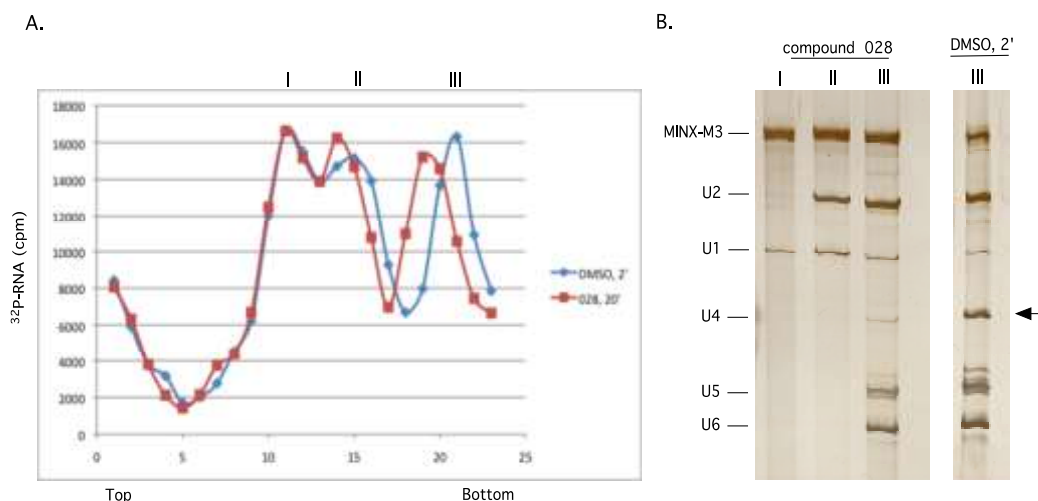
**Table 3.2 Proteins identified by LC-MS/MS in human spliceosomal complexes A, B and complexes stalled in the presence of compounds 120 (B<sup>120</sup>), 296 (B<sup>296</sup>) and 297 (B<sup>297</sup>).** The number shown is the total number of peptides sequenced for the indicated protein. Proteins are grouped according to function or association. Proteins widely considered as common contaminants, such as ribosomal proteins and non-reproducibly detected proteins are not shown.

	GenBank accession no.	kDa	A	A 120	B	B 120	B	B 296	B 297
fmol			250	250	250	250	1,000	1,000	1,000
<b>Sm proteins</b>									
B	gi 4507125	24.6	1	7	18	12	108	127	88
D1	gi 5902102	13.3	1	3	7	5	11	12	9
D2	gi 29294624	13.5	7	3	13	11	87	82	87
D3	gi 4759160	13.9	5	6	10	5	79	67	67
E	gi 4507129	10.8	6	5	13	6	54	48	42
F	gi 4507131	9.7	4	5	8	6	32	30	18
G	gi 4507133	8.5	3	2	5	1	24	15	19
<b>U1 snRNP</b>									
U1-70K	gi 29568103	51.6	10	6	8	3	5	10	0
U1-A	gi 4759156	31.3	1	8	7	6	21	32	22
U1-C	gi 119624201	17.4	2	2	2	2	4	3	2
<b>U1 snRNP associated</b>									
FBP11	gi 151301228	105.8					1	30	27
S164 (fSAP94)	gi 219520438	100.1					0	20	10
<b>17S U2 snRNP</b>									
U2A'	gi 50593002	28.4	2	13	26	24	87	142	128
U2B''	gi 4507123	25.4	1	9	11	9	33	67	30
SF3a120	gi 5032087	88.9	37	34	53	56	430	387	350
SF3a66	gi 21361376	49.3	8	7	17	13	48	52	55
SF3a60	gi 5803167	58.5	18	17	30	24	140	167	133
SF3b155	gi 54112117	145.8	77	65	110	78	666	489	545
SF3b145	gi 55749531	100.2	23	19	57	46	359	329	371
SF3b130	gi 54112121	135.5	70	53	84	78	626	662	648
SF3b49	gi 5032069	44.4	3	3	5	4	17	26	24
SF3b14a/p14	gi 7706326	14.6	10	9	11	9	61	75	81
SF3b14b	gi 14249398	12.4	5	7	6	7	25	39	50
SF3b10	gi 13775200	10.1	1	1	6	4	24	29	16
<b>17S U2 related</b>									
hPRP43	gi 68509926	90.9	24	7	35	22	137	118	94
SPF45	gi 14249678	45	5	5	9	7	23	16	12
SPF30	gi 37087924	26.7					26	21	15
U2AF65	gi 6005926	53.5	3	7	10	8	29	29	23
U2AF35	gi 5803207	27.9		2	6	2	9	16	9
SPF31	gi 3599415	31		1		0	0	3	2
hPRP5	gi 664806097	117.4					1	0	0
SR140 (fSAPa)	gi 1122937227	118.2					51	45	30
CHERP	gi 119226260	100	11	4	5	4	41	32	23
PUF60	gi 74761960	59.9	4	21	1	15	3	27	23
SF3b125	gi 45446747	103		2	0	5	0	1	1
<b>A proteins</b>									
THRAP3 (TRAP150)	gi 167234419	108.5	13		8	0	1	10	6
BUB3	gi 56550081	37.2					8	42	37
FLJ10839, ICCAR1	gi 46852388	132.8					1	88	70
RBM10	gi 12644371	103.5	9	8	12	10	30	13	23
SF1	gi 42544130	68.3					2	14	11
RBM5/LUCA15	gi 5032031	92.1	6	0	7	2	18	14	27
SF4 (F23858)	gi 33469964	72.5					0	19	13
<b>U5 snRNP</b>									
220K (Prp8)	gi 3661610	273.7	90	16	185	122	836	799	798
200K (Brr2)	gi 45861372	244.5	100	17	189	129	934	859	821
116K (Snu114)	gi 41152056	109.4	60	17	82	69	468	503	445
40K	gi 4758560	39.3	1	4	33	31	130	138	160
102K (Prp6)	gi 40807485	106.9	29		59	36	424	222	262
15K (Dib1)	gi 5729802	16.8					32	12	8
100K (Prp28)	gi 41327771	95.6	7		23	3	20	24	24
52K (Lin1)	gi 5174409	37.6					25	11	6
<b>LSm proteins</b>									
LSm2	gi 10863977	10.8	4	2	8	5	35	22	20
LSm3	gi 7657315	11.8					12	11	11
LSm4	gi 6912486	15.4	1		10	5	39	28	23
LSm5	gi 10720081	9.9					2	2	4
LSm6	gi 5919153	9.1	1		2	2	24	18	14
LSm7	gi 7706423	11.6	0		3	2	14	10	11
LSm8	gi 7706425	10.4					24	19	14

	GenBank accession no.	kDa	A	A 120	B	B 120	B	B 296	B 297
fmol			250	250	250	250	1,000	1,000	1,000
<b>U4/U6 snRNP</b>									
90K	gi 4758556	77.6	19	2	48	32	246	118	165
60K	gi 45861374	58.4	16	5	41	27	226	80	131
20K	gi 5454154	20	9	2	16	7	82	31	55
61K	gi 40254869	55.4	13	1	27	13	126	38	76
15.5K	gi 4826860	14.2					20	12	13
<b>U4/U6.U5 snRNP</b>									
110K	gi 13926068	90.2	19	1	62	41	431	223	225
65K	gi 56550051	65.4			10		12	17	5
27K	gi 24307918	18.9			2				
<b>B proteins</b>									
hSnu23		23.6					20	43	24
hPRP38	gi 24762236	37.5			1	17	52	66	41
hSmu-1 (fSAP57)	gi 8922679	57.5	8	8	33	34	275	270	289
MFAP1	gi 50726968	51.9	1	4	12	15	73	101	90
RED	gi 10835234	65.6	1	0	37	37	246	230	230
FBP21	gi 67461858	42.5			10	7	42	10	18
<b>hPRP19/CDC5L complex</b>									
hPRP19	gi 7657381	55.2	3	1	17	22	32	216	218
CDC5L	gi 11067747	92.2	0		30	36	41	218	237
SPF27	gi 5031653	21.5			10	6	8	61	60
PRL1	gi 4505895	57.2					20	124	94
CCAP1 (hsp73)	gi 5729877	70.4	7	9	11	11	24	17	6
CCAP2 (AD-002)	gi 7705475	26.6			0	4	3	18	24
CTNBL1	gi 18644734	65.1			1	6	2	30	39
Npw38BP	gi 7706501	70		0	11	23	43	109	89
Npw38	gi 5031957	30.5					11	44	37
<b>hPRP19/CDC5L related</b>									
RBM122 (fSAP47)	gi 8922328	46.9			2	9	15	59	63
hSYF1 (XAB2)	gi 55770906	100	2		8	9	4	147	128
CRNKL1 (hSYF3)	gi 30795220	100.6					13	133	81
hIsy1 (fSAP133)	gi 20149304	33			3	7	5	29	26
SKIP	gi 6912676	51.1			8	20	30	160	161
Cyp-E	gi 5174637	33.4					3	48	23
PPIase-like (PPL1)	gi 7706339	18.2					11	39	39
KIAA0560 (fSAP164)	gi 38788372	171.3	1		11	12	11	219	149
G10 (fSAP17)	gi 32171175	17			2	5	8	38	26
<b>A/B proteins</b>									
TCERG1 (CA150)	gi 21327715	123.9					0	65	44
ELAV (HuR)	gi 38201714	36.1					12	39	29
p72/DDX17	gi 3122595	80.5		3	2	7			
DDX9, DDX9	gi 4503297	142	30	4	15	1	42	19	2
NFAR	gi 62512150	95.4	5	1	1		22	8	1
NF45	gi 24234747	43					5	4	0
LOC124245	gi 31377595	104	6	0	14	17	15	64	52
YB-1	gi 34098946	35.9	3	6	9	10	33	51	15
ASR2B	gi 33383233	100	22	7	18	23	4	103	128
p68 (DDX5)	gi 4758138	69.2	2	3	8	9	20	35	30
<b>Recruited prior to Bact</b>									
hsp27	gi 4504517	22.8					0	2	0
UBL5	gi 13236510	8.5					14	19	26
hPRP4-Kinase	gi 89276756	117.1	1		6		0	6	11
HsKin17	gi 13124883	45.4					0	47	52
BCLAF1	gi 7661958	106					1	11	3
SKIV2L2 (KIAA0052)	gi 193211480	117.8	18	1	11	4	22	6	7
<b>Abundant first in Bact</b>									
KIAA1604 (fSAPb)	gi 55749769	105.5					0	45	8
hPRP17	gi 7706657	65.5			8	7	17	76	40
hPRP2	gi 4503293	119.2					0	77	24
GPKOW	gi 15811782	52.1					0	11	0
NY-CO-10	gi 64276486	53.8					0	15	7
RNF113A	gi 5902158	38.8					0	9	5
MGC20398	gi 49472814	42					0	23	28
PPL2 (Cyp-60)	gi 7657473	59.5					0	53	55
MGC23918	gi 21389497	19.2					4	41	28

	GenBank accession no.	kDa	A	A 120	B	B 120	B	B 296	B 297
fmol			250	250	250	250	1,000	1,000	1,000
<b>2nd step factors</b>									
hPRP22	gi 4826690	139.3					0	31	0
hPRP16	gi 17999539	140.5					0	9	0
hSLU7	gi 27477111	68.4					0	15	0
Abstrakt GCIP29 (fSAP29)	gi 21071032	69.8						17	
	gi 46371998	28.7					1	25	1
<b>detected in complex</b>									
eIF3b6	gi 7705433	66.6					17	8	0
RBM4	gi 93277122	40.2		2	4	3	7	7	8
matrin3	gi 21626466	94.6	7		5		9	13	4
<b>EJC/mRNP</b>									
eIF4A3	gi 7661920	46.9					3	16	6
Magoh	gi 4505087	17.2					0	3	0
Y14	gi 4826972	19.9					0	3	0
DDX3	gi 87196351	73.3	1	6	6	13	10	41	38
<b>mRNA binding proteins</b>									
PABP1	gi 46367787	70.5	2		23				
<b>cap binding complex</b>									
CBP20	gi 110349727	18	4	2	5	7	13	17	12
CBP80	gi 4505343	91.8	26	12	36	29	91	225	150
<b>RES complex</b>									
SNIP1	gi 21314720	45.8					0	28	26
MG13125 (fSAP71)	gi 14249338	70.5			4	8	1	53	66
CGI-79	gi 4929627	39.7					0	23	27
<b>SR proteins</b>									
SF2/ASF	gi 5902076	27.8					14	39	41
9G8	gi 72534660	27.4					8	22	26
SRp20	gi 4506901	19.4					0	21	11
SRp30c	gi 4506903	25.5					9	10	1
SRp38	gi 5730079	31.3					0	3	3
SRp55	gi 20127499	39.6					4	10	37
hTra-2b (beta) (FRS10)	gi 4759098	33.7					4	3	3
<b>SR related proteins</b>									
SRm300	gi 4759098	300					0	32	21
<b>hnRNP</b>									
hnRNP A0	gi 8134660	30.9					10	6	4
hnRNP A1	gi 4504445	38.7					27	28	26
hnRNP A3	gi 34740329	39.6					24	24	17
hnRNP A2/B1	gi 14043072	37.4					11	4	4
hnRNP C	gi 4758544	33.3					17	11	10
hnRNP D	gi 14110420						2	2	1
hnRNP E	gi 148470406	45.7					5	16	19
hnRNP G	gi 56699409	42.4					14	12	12
hnRNP H1	gi 5031753	49.1					11	26	21
hnRNP H2	gi 2500576	49.3					0	18	0
hnRNP H3	gi 23503095	36.9					0	0	0
hnRNP K	gi 14165435	51					0	22	17
hnRNP L	gi 5031755	70.9					8	10	4
hnRNP U	gi 14141161	90.6					9	10	8
hnRNP U-L2	gi 118601081	85					7	2	3
PCBP1	gi 5453854	37.5					14	15	36
E1B-AP5	gi 8051631	95.7					4	5	11
<b>Miscellaneous proteins</b>									
RBM7	gi 4503293	30.5					5	2	0
dermcidin	gi 20141302	11.3					50	28	36
RACK1 (GNB2L1)	gi 54037168	35.1					34	17	12
desmoglein 1	gi 292495005	114					36	26	21
desmoplakin	gi 115502381	332					37	22	26
S100a18 (Hornerin)	gi 45476906	282					42	24	52
IGF2BP3 (Koc1)	gi 30795212	63.6					13	16	8
DHX36	gi 313104099	115					15	7	1
filaggrin 2	gi 74096732	248					14	31	26
GAPDH	gi 120649	36					8	9	2
human serum albumin	gi 4502027	66					24	19	41
beta-actin	gi 46397333	41					10	7	6
tubulin (beta-5)	gi 56757569	50					55	33	45
RBM39	gi 124297482	57					5	9	4
FUBP2	gi 313104306	77					0	24	9

splicing complexes from peaks I, II and III using MS2 affinity selection. With compound 028, complexes in peak II contained equimolar amounts of unspliced pre-mRNA and U2, U5 and U6, but had much less U1 snRNA, and nearly no U4 snRNA, suggesting they are mainly activated spliceosomes. However, with compound 028, complexes in peak II were shifted 1-2 fractions toward the top; thus, they have a lower S-value than B<sup>act</sup> and B complexes, suggesting they might be a novel spliceosome assembly intermediate. Peak I corresponds to H complexes, while peak II contained A complexes, with only U1, U2 and pre-mRNA.



**Figure 3.12 Purification of splicing complexes stalled with compound 028.** A. Glycerol gradient sedimentation profiles of the separated spliceosomal complexes. B. RNA composition of MS2-affinity purified splicing complexes. RNA recovered from complexes affinity-purified from peaks I, II and III was analysed by denaturing PAGE and detected by silver staining. RNA identities are indicated at the left.

### 3.2.5 Protein composition of affinity-purified spliceosomal A and B-like complexes formed in the presence of compound 028

The protein composition of MS2 affinity-purified B<sup>028</sup>, B and B<sup>act</sup> spliceosomal complexes, as well as A<sup>028</sup> and A complexes was determined by LC-MS/MS mass spectrometry in several independent experiments. The MS analysis was performed in collaboration with the group of Prof. Dr. H. Urlaub. The number of peptides sequenced for spliceosomal proteins in the purified A or B/B<sup>act</sup> complexes is shown in Table 3.2. To isolate B<sup>act</sup> complexes we used the PM5-10 substrate, which is a truncated version of the PM5 pre-mRNA, containing a 5' exon, an intron with a branch site, a 5 nt stretch between the BPS and PPT, and a polypyrimidine tract shortened to 10 nt (compared to 60 nt in the full-length version), but no terminal AG dinucleotide and no 3' exon. Additionally, it

contains triple MS2 aptamers at its 5' end, enabling affinity selection of spliceosomal complexes. Spliceosome assembly utilizing this pre-mRNA is stalled prior to the catalytic activation step (i.e. B\* formation), and thus allows activated spliceosomes (B<sup>act</sup>) to be isolated.

Mass spectrometry showed that the A complexes formed in the presence of compound 028 contain similar amounts of U1 and U2 proteins, compared to A complexes, formed in the presence of DMSO (Table 3.3). However there were some differences in the amounts of some U2-related proteins, such as SPF45, SPF30, CHERP and also the A or A/B proteins DDX9 and RBM5 (based on total peptides identified for each protein), which appear to be less abundant in A<sup>028</sup> complexes. The SPF30 protein is known to associate with both the U2 snRNP and the tri-snRNP (Rappsilber et al., 2001; Meister et al., 2001), suggesting the accumulation of A complexes in the presence of compound 028 could potentially be caused by the absence of SPF30, blocking the transition from A complex to B complex. Purified A complexes from the DMSO control apparently also contain low amounts of contaminating B complexes, based on the presence of U5, U4/U6 and tri-snRNP proteins.

Based on the number of detected peptides, many proteins appear to be present in equal/similar amounts in B<sup>028</sup> and B complexes formed in the presence of DMSO (Table 3.3). These include nearly all proteins associated with U1 snRNP, U2 snRNP, and also U5 snRNP, including the DExD/H-box protein Brr2, which catalyses the unwinding of U4/U6, and Prp8 and the GTPase Snu114, which among others regulate/activate Brr2 activity during the activation step (Small et al., 2006; Maeder et al. 2009; Mozaffari-Jovin et al., 2013). However, several U5-specific proteins appear to be underrepresented in B<sup>028</sup> such as 102K/Prp6, 15K/Dib1 and the DEAD-box phosphoprotein 100K/hPrp28, which catalyses the release of U1 from the 5'SS of pre-mRNA during spliceosome activation, so that U6 can interact with the 5' splice site (Staley and Guthrie 1999, Chen et al. 2001).

Distinct from B complexes, B-like complexes formed in the presence of compound 028 contain highly underrepresented amounts of U4/U6 snRNP proteins, which is characteristic for B<sup>act</sup> spliceosomal complexes (Table 3.3). Indeed, only few peptides were sequenced for U4/U6 snRNP proteins in purified B<sup>act</sup> complexes. Lsm proteins, which dissociate during activation, are detected in B complexes. They appear to be still abundant in B<sup>028</sup>; by comparison only low levels of them are detected in activated spliceosomes (B<sup>act</sup>), suggesting the inhibition of spliceosome assembly by 028 occurs before Lsm destabilisation.

**Table 3.3** Proteins identified by LC-MS/MS in human spliceosomal complexes B, B<sup>act</sup> and complexes stalled in the presence of compound 028 (B<sup>028</sup>). The number shown is the total number of peptides sequenced for the indicated protein. Proteins are grouped according to function or association. Proteins considered as common contaminants, such as ribosomal proteins and non-reproducibly detected proteins are not shown.

	GenBank accession no.	kDa	A	A 028	A	A 028	B	B 028	B	B 028	Bact
famol			250	250	400	500	250	250	800	800	800
<b>MS2-MBP</b>									196	276	316
<b>Sm proteins</b>											
B	gi 4507125	24.6	1	7	39	15	18	12	55	55	57
D1	gi 5902102	13.3	1	3	3	20	7	4	8	13	14
D2	gi 29294624	13.5	7	7	40	17	13	7	33	47	64
D3	gi 4759160	13.9	5	6	30	24	10	6	37	39	47
E	gi 4507129	10.8	6	6	16		13	7	19	19	27
F	gi 4507131	9.7	4	4	5		8	6	18	23	29
G	gi 4507133	8.5	3	3	5	1	5	4	6	9	14
<b>U1 snRNP</b>											
U1-70K	gi 29568103	51.6	10	7	36	58	8	3			
U1-A	gi 4759156	31.3	1	7	34	21	7	3	7	15	9
U1-C	gi 119624201	17.4	2	3			2	2		4	2
<b>U1 snRNP associated</b>											
FBP11	gi 151301228	105.8								9	9
S164 (fSAP94)	gi 219520438	100.1	5	1				6	1	9	6
<b>17S U2 snRNP</b>											
U2A'	gi 50593002	28.4	2	13	89	60	26	15	47	91	87
U2B''	gi 4507123	25.4	1	6	37	40	11	9	23	30	47
SF3a120	gi 5032087	88.9	37	41	322	355	53	50	205	210	194
SF3a66	gi 21361376	49.3	8	13	50	43	17	13	24	33	34
SF3a60	gi 5803167	58.5	18	22	135	138	30	25	62	95	75
SF3b155	gi 54112117	145.8	77	70	671	652	110	91	337	404	327
SF3b145	gi 55749531	100.2	23	15	219	313	57	37	185	238	228
SF3b130	gi 54112121	135.5	70	60	833	723	84	86	365	481	496
SF3b49	gi 5032069	44.4	3	3	43	15	5	3	10	15	18
SF3b14a/p14	gi 7706326	14.6	10	10	38	24	11	11	14	25	28
SF3b14b	gi 14249398	12.4	5	6	32	29	6	8	13	22	25
SF3b10	gi 13775200	10.1	1	2	9		6	4	10	12	11
<b>17S U2 r related</b>											
hPRP43	gi 68509926	90.9	24	6	115	84	35	16	41	39	56
SPF45	gi 14249678	45	5	3	56	18	9	3	2	7	5
SPF30	gi 37087924	26.7			44	9					
U2AF65	gi 6005926	53.5	3	5	5	12	10	5	7	13	
U2AF35	gi 5803207	27.9		2	2	4	6	6	2	7	1
SPF31	gi 3599415	31		2	0	6				4	
hPRP5	gi 664806097	117.4									
SR140 (fSAPa)	gi 122937227	118.2			10	40			5	12	16
CHERP	gi 119226260	100	11	1	87	19	5	3	2	4	6
PUF60	gi 74761960	59.9	4	8			1	6		1	
SF3b125	gi 45446747	103					0	1			
<b>A proteins</b>											
THRAP3 (TRAP150)	gi 167234419	108.5	13				8	0			
BUB3	gi 56550081	37.2			0	2				6	4
FLJ10839, CCAR1	gi 46852388	132.8			3	7					
RBM10	gi 12644371	103.5	9	10	84	30	12	7	4	13	7
SF1	gi 42544130	68.3			2	1				14	1
RBM5/LUCA15	gi 5032031	92.1	6		48	1	7	1			
ZNF207	gi 6137311	50.8									
SF4 (F23858)	gi 33469964	72.5			3						40
<b>U5 snRNP</b>											
220K / Prp8	gi 3661610	273.7	90	7	39		185	141	505	644	748
200K / Brr2	gi 45861372	244.5	100	12	44		189	160	563	702	820
116K / Snu114	gi 41152056	109.4	60	9	65	0	82	79	256	338	407
40K	gi 4758560	39.3	1	1	13		33	29	59	85	123
102K / Prp6	gi 40807485	106.9	29	1	8		59	26	181	77	24
15K / Dib1	gi 5729802	16.8					15	3	9	5	1
100K / Prp28	gi 41327771	95.6	7				23		31	21	17
52K / Lin1	gi 5174409	37.6					6		8	3	4
<b>LSm proteins</b>											
LSm2	gi 10863977	10.8	4	1	1		8	6	11	18	9
LSm3	gi 7657315	11.8							2	6	2
LSm4	gi 6912486	15.4	1	0			10	7	10	18	5
LSm5	gi 10720081	9.9							2	3	2
LSm6	gi 5919153	9.1	1				2	2	7	11	4
LSm7	gi 7706423	11.6	0				3	2	7	6	1
LSm8	gi 7706425	10.4					7	5	10	9	6



	GenBank accession no.	kDa	A	A 028	A	A 028	B	B 028	B	B 028	Bact
fmlol			250	250	400	500	250	250	800	800	800
<b>U4/U6 di-snRNP</b>											
90K / Prp3	gi 4758556	77.6	19	1	26	0	48	19	128	47	9
60K / Prp4	gi 45861374	58.4	16	1	22		41	12	90	35	11
20K	gi 5454154	20	9	2	4		16	9	37	27	4
61K / Prp31	gi 40254869	55.4	13	0	3		27	12	52	16	5
15.5K / Snu13	gi 4826860	14.2							10	4	1
<b>U4/U6.U5 tri-snRNP</b>											
110K / Snu66	gi 13926068	90.2	19		5		62	31	238	108	21
65K / Sad1	gi 56550051	65.4					10		6	2	3
27K	gi 24307918	18.9					2				
<b>B proteins</b>											
hSnu23	gi 52783532	23.6					6	6	13	19	12
hPRP38	gi 24762236	37.5					1	7	61	81	35
hSnu-1 (fSAP57)	gi 8922679	57.5	8	5			33	34	131	191	78
M FAP1	gi 50726968	51.9	1	2			12	11	63	73	46
RED	gi 10835234	65.6	1	0			37	28	91	147	60
FBP21	gi 67461858	42.5					10	3	21	7	
<b>hPRP19/CDC5L complex</b>											
hPRP19	gi 7657381	55.2	3		3		17	8	77	33	296
CDC5L	gi 11067747	92.2	0				30	5	89	35	249
SPF27	gi 5031653	21.5			0		10	2	17	9	56
PRL1	gi 4505895	57.2					4	0	38	23	193
CCAP1 (hsp73)	gi 5729877	70.4	7	10			11	8	21	26	55
CCAP2 (AD-002)	gi 7705475	26.6					0	0	8	2	24
CTNNBL1	gi 18644734	65.1					1	1	10	10	34
Npw38BP	gi 7706501	70		0			11	24	43	115	44
Npw38	gi 5031957	30.5					6	7	4	27	9
<b>hPRP19/CDC5L related</b>											
PRCC	gi 40807447	52.4					0	0	9	1	15
RBM 22 (fSAP47)	gi 8922328	46.9					2	0	29	6	73
hSYF1 (XAB2)	gi 55770906	100	2				8	2	45	7	290
CRNKL1 / hSYF3	gi 30795220	100.6					4		60	17	200
hlsy1 (fSAP133)	gi 20149304	33					3	1	16	6	86
SKI P	gi 6912676	51.1					8	5	60	24	143
Cyp-E	gi 5174637	33.4					3		12	3	66
PPIase-like 1 (PPIL1)	gi 7706339	18.2					3	1	8	13	30
K1AA0560 (fSAP164)	gi 38788372	171.3	1				11	4	60	13	432
G10 (fSAP17)	gi 32171175	17					2	3	13	26	30
<b>A/B proteins</b>											
TCERG1 (CA150)	gi 21327715	123.9									
ELAV (HuR)	gi 38201714	36.1									
p72/DDX17	gi 3122595	80.5					2				
DDX9, DHX9	gi 4503297	142	30	5	77	32	15	5			
NFAR	gi 62512150	95.4	5	3			1		3	8	
NF45	gi 24234747	43			5	8					
LOC124245	gi 31377595	104	6		0	1	14	3			
YB-1	gi 34098946	35.9	3	10	29	24	9	8			
ASR2B	gi 33383233	100	22	1	4	9	18	23			
p68 (DDX5)	gi 4758138	69.2	2	1		2	8	4	5	16	15
<b>Recruited prior to Bact</b>											
hsp27	gi 4504517	22.8				4			2	2	4
UBL5	gi 13236510	8.5							12	14	14
hPRP4-Kinase	gi 89276756	117.1	1				6		2	1	
HskIn17	gi 13124883	45.4							16	7	18
BCLAF1	gi 7661958	106									2
SKI V2L2 (K1AA0052)	gi 193211480	117.8	18	1			11	2	10	1	5
<b>Abundant first in Bact</b>											
K1AA1604 (fSAPb)	gi 55749769	105.5							8		183
hPRP17	gi 7706657	65.5					8	1	42	7	171
hPRP2	gi 4503293	119.2							18	7	184
GPLOW	gi 15811782	52.1							5	5	104
NY-CO-10	gi 64276486	53.8							6		54
RNF113A	gi 5902158	38.8							1		41
M GC20398	gi 49472814	42							11		40
PPIL2 / Cyp-60	gi 7657473	59.5							20	3	111
M GC23918	gi 21389497	19.2							11	1	36
FRG1	gi 4758404	29.2									9
<b>Detected in Bact</b>											
TRIM24	gi 47419911	116.7									9
LENG1	gi 24308289	30.4									62
DDX50	gi 13129006	82.4									4
FLJ38348	gi 167900474	30.1									11
CRIPT, HSPC139	gi 7661798	11.1								1	
GTL3 (fSAP23)	gi 8392875	22.6								1	22
M OV10	gi 14211540	113.5								9	2
FUBP3	gi 100816392	61.5				3				4	9

	GenBank accession no.	kDa	A	A 028	A	A 028	B	B 028	B	B 028	Bact
fmol			250	250	400	500	250	250	800	800	800
<b>2nd step factors</b>											
hPRP22	gj4826690	139.3									89
hPRP16	gj1799539	140.5									
hPRP18	gj4506123	39.9									
hSLU7	gj2747711	68.4									31
Abstrakt	gj21071032	69.8									39
HSP70	gj5123454	70			10	34					
GClp29(fSAP29)	gj46371998	28.7									20
DDX35	gj20544129	78.9									35
TFIP11	gj8393259	96.8									
<b>detected in complex</b>											
eIF356	gj7705433	66.6									
WDR70	gj8922301	73.1									
RBM4	gj93277122	40.2					4	0			
BCAS1(AIBC1, NABC1)	gj191251777	61.2									
DBPA	gj20070160	40.1			25	16					
matrin3	gj21626466	94.6	7	1			5	6			
<b>EJC/mRNP</b>											
eIF4A3	gj7661920	46.9							0	0	50
Magoh	gj4505087	17.2									
Y14	gj4826972	19.9									5
Pinin	gj33356174	81.6									8
RNPS1	gj6857826	34.2									3
Acinus(fSAP152)	gj7662238	151.8									
UAP56	gj18375623	49.1									
ELG	gj8923771	38.9							2		14
DDX3	gj87196351	73.3	1				6	3		8	6
<b>mRNA binding proteins</b>											
PABP1	gj46367787	70.5	2				23	3			
PABPC4, PABP, APP1	gj4504715	70.6									
PABPN1, PAB2	gj4758876	32.6							1		1
<b>cap binding complex</b>											
CBP20	gj110349727	18	4	4	4	3	5	6	7	20	18
CBP80	gj4505343	91.8	26	13	30	57	36	32	80	124	213
<b>TREX</b>											
THOC1(HPR1)	gj154448890	75.6								1	2
THOC2	gj125656165	169.6							1		3
THOC3	gj14150171	38.8							1		2
THOC5(KIAA0983, fSAP79)	gj50959110	78.5									
THOC6(WDR58, MGC2655)	gj31543164	37.5								1	5
THOC7(FUJ23445, fSAP24)	gj13376623	23.7									1
<b>RES complex</b>											
SNIP1	gj21314720	45.8							13	3	64
MGC13125(fSAP71)	gj14249338	70.5					4	0	24	15	115
CGI-79	gj4929627	39.7									
<b>SR proteins</b>											
SF2/ASF	gj5902076	27.8	0	7	5	14	12	12			
9G8	gj72534660	27.4							1	2	21
SRp20	gj4506901	19.4								1	10
SRp30c	gj4506903	25.5		2			2	3	1	4	12
SRp38	gj5730079	31.3									35
SRp46	gj14141216										3
SRp55	gj20127499	39.6	9	2			11	5		2	22
SRp75	gj21361282	56.8	3				2				
hTra-2alpha	gj9558733	32.7									9
hTra-2beta(fSFRS10)	gj4759098	33.7				5	10	4	1		19
<b>SR related proteins</b>											
SRm160	gj42542379	102.5									
SRm300	gj4759098	300							23	10	196
<b>hnRNP</b>											
hnRNP A0	gj8134660	30.9		3	9	1	3	1	6	8	
hnRNP A1	gj4504445	38.7		3	1	1	11	5	15	12	3
hnRNP A3	gj34740329	39.6		2			4	4	8	6	5
hnRNP A2/B1	gj14043072	37.4		4	1		8	4	1	3	1
hnRNP C	gj4758544	33.3		5	21	10	7	3			
hnRNP D	gj14110420	38.4									
hnRNP E	gj148470406	45.7		2		2	2	5		11	7
hnRNP G	gj56699409	42.4		1	4	0	6	2	5	4	23
hnRNP H1	gj5031753	49.1	0	8		4	4	10	3	21	12
hnRNP H2	gj2500576	49.3						1			11
hnRNP H3	gj23503095	36.9					1	2			4
hnRNP K	gj14165435	51	2	6	1	5	5	6	2	14	1
hnRNP L	gj215274006	64.1									1
hnRNP M	gj14141152	77.5	6		1		11	3			
hnRNP R	gj5031755	70.9				1			10	2	8
hnRNP U	gj14141161	90.6	7		3	0	3	0			
hnRNP U-L2	gj118601081	85	10	2	6	1	5	2	1	4	
PCBP1	gj5453854	37.5	1	6			10	8			7
E1B-AP5	gj8051631	95.7									

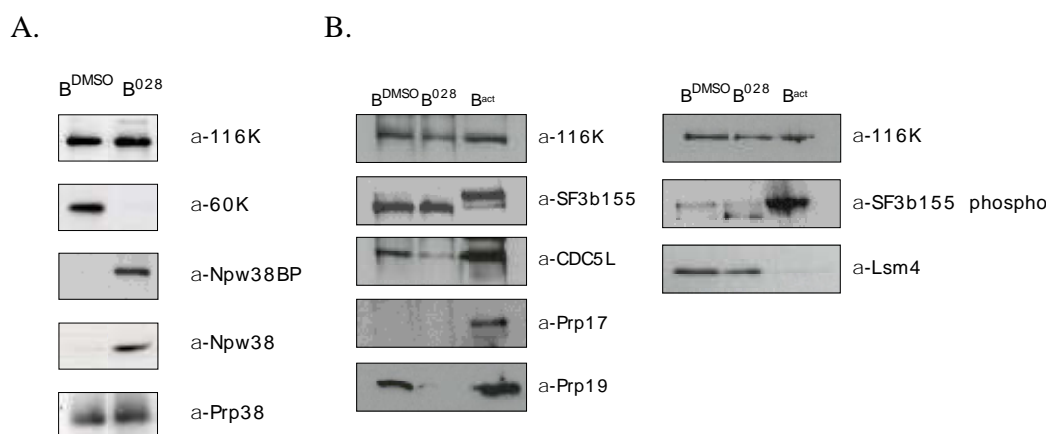
fmol	GenBank accession no.	kDa	A		A 028		B		B 028		Bact	
			250	250	400	500	250	250	800	800	800	
<b>Miscellaneous proteins</b>												
RBM 7	gi 4503293	30.5			8							
dermcidin	gi 20141302	11.3	5	7	47	79	4	4				
S100 A9	gi 115444	13			4	10						
S100 A8	gi 115442	11	3	4	10	19	3	7	15	10	11	
desmoglein 1	gi 292495005	114	18	19	4	20	15	20	57	71	59	
desmoplakin	gi 115502381	332	25	13	8	47	15	54	17	14	35	
S100a18, Hornerin	gi 45476906	282	23	10	35	84	13	16				
FABP5	gi 232081	15	4	2		2	3	3				
IGF2BP3 (Koc1)	gi 30795212	63.6	6	5	4	2	5	0				
filaggrin 2	gi 74096732	248	7	6	21	32	5	19	30	34	47	
PDCD6	gi 29791717	25	7	1			7	5				
GAPDH	gi 120649	36	5	2	3	20	1	5				
human serum albumin	gi 4502027	66	9	5	109	18	10	10	7	7	6	
alpha tubulin	gi 37492	46	8	12	25	13	19	21				
beta tubulin	gi 338695								30	26	74	
beta-actin	gi 46397333	41	3	2			3	5				
tubulin beta-5	gi 56757569	50	12	17	13	7	26	19	11	7	5	
JUP, PDGB, gamma-Catenin	gi 15080189	82	15	9	0	17	12	23	15	8	14	
RBM 39	gi 124297482	57				11						
FUBP2	gi 313104306	77				3						

MS analysis also revealed that most B-specific proteins are present in B and B<sup>028</sup> complexes in similar amounts. One exception was the FBP21 protein, which appears to be underrepresented in B<sup>028</sup> complexes. In contrast to B<sup>act</sup> complexes, B<sup>028</sup> complexes have little or no B<sup>act</sup>-specific spliceosomal proteins that are recruited during activation. Thus, the inhibition by 028 occurs prior to B<sup>act</sup> complex formation. Based on the detection of low amounts of B<sup>act</sup>-specific proteins, B<sup>DMSO</sup> complexes also appear to contain a low level of B<sup>act</sup> complexes. Interestingly, B<sup>028</sup> complexes also have much fewer Prp19/CDC5L complex proteins as well as Prp19/CDC5L related factors, relative to B<sup>act</sup> complexes. These proteins are usually detected in B complexes in low amounts, but they are stabilized during activation, and thus found enriched in B<sup>act</sup> and subsequently formed C complexes (Bessonov et al, 2008). On the other hand, there are higher amounts of Npw38BP and Npw38 in the B<sup>028</sup> complex compared to B and B<sup>act</sup> complexes. These proteins normally dissociate during the activation step and are not detected in B<sup>act</sup>, suggesting that spliceosome assembly is stalled in the presence of compound 028 prior to the stable integration of the Prp19/CDC5L complex and prior to Npw38 and Npw38BP dissociation. Taken together, the protein composition of B<sup>028</sup> complexes suggests that compound 028 stalls splicing at an early stage of the spliceosome activation step.

### 3.2.6 Immunoblotting confirms the absence/presence of selected proteins in affinity-purified spliceosomal complex B<sup>028</sup>

To confirm the presence or absence of selected proteins in B<sup>028</sup> complexes we performed immunoblotting with affinity-purified B, B<sup>028</sup> and B<sup>act</sup> complexes. The signal intensity of U5 116K/Snu114 protein was used as a loading control. As shown in figure 3.13, the U4/U6 60K/Prp4 protein, also a component of the tri-snRNP, was detected in complex B, but not in B<sup>028</sup> complexes with anti-60K antibodies. The B-specific protein Prp38 was

detected both in B and B<sup>028</sup> complexes in approximately equal amounts. In contrast, Npw38 and Npw38BP, which belong to the group of Prp19/CDC5L-related proteins, were not detected by immunoblotting in the B complex, but were present in the stalled B<sup>028</sup> complexes. Spliceosomes formed in the presence of compound 028 contained underrepresented amounts of Prp19/CDC5L complex proteins (i.e. Prp19 and CDC5L) (Fig. 3.13B), which are normally first abundant in the B<sup>act</sup> complex (i.e. after activation). Furthermore, the amount of these proteins is even less than in the affinity purified B complexes, suggesting compound 028 might affect the association of the Prp19/CDC5L complex in a negative way. Additionally B<sup>028</sup> complexes lack the B<sup>act</sup> specific protein Prp17 that is recruited during activation. However they still contain similar amounts of the Lsm4 protein as the B complex.



**Figure 3.13 Western blot analyses of the purified B, B<sup>028</sup> and B<sup>act</sup> complexes.** Affinity-purified B and B<sup>028</sup> (A) or B, B<sup>028</sup> and B<sup>act</sup> complexes (B) were subjected to the western blot analysis. Purified complexes were separated on an 10% SDS-PAGE, transferred to a nitrocellulose membrane and blotted with the indicated antibodies. To ensure that equal amounts of the both complexes were analysed, antibodies against the U5-116K protein, a stable core component of the spliceosome, were used.

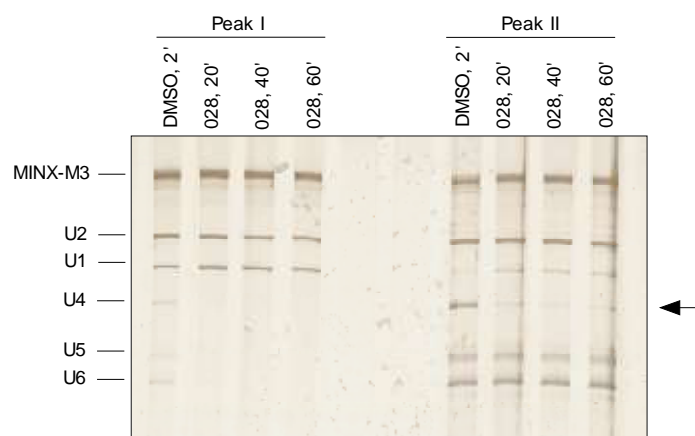
A component of the U2 snRNP, the SF3b155 protein, is known to be phosphorylated during/after the activation step but prior the first catalytic step of splicing (Wang et al., 1998). If detected by anti-SF3b155 antibodies, both states of this protein – phosphorylated and not phosphorylated – can be identified, but they exhibit different migration behaviour during SDS-PAGE. To check whether the SF3b155 in the B<sup>028</sup> complexes is phosphorylated, we performed immunoblotting with anti-SF3b155 antibodies. Indeed, we could observe that the SF3b155 band in B and also in B<sup>028</sup> complexes migrates faster than the SF3b155 signal detected by the same antibodies in B<sup>act</sup> complexes, suggesting that SF3b155 is phosphorylated in B<sup>act</sup> complexes, but not in B or B<sup>028</sup>. An antibody against phosphorylated threonine 313 of SF3b155 is able to detect only the phosphorylated form of SF3b155 (Girard et al., 2012). Western blotting with these

antibodies revealed a slower migrating SF3b155 band solely in B<sup>act</sup> complexes, but not in the B or B<sup>028</sup> complexes consistent with our results with  $\alpha$ -SF3b155 antibodies. Taken together, these results show that B<sup>028</sup> complexes are stalled before phosphorylation of the SF3b155 protein.

Thus, our western blot analyses generally correlate with the mass spectrometry data. Taken together, our results demonstrate that the spliceosomal B-like complex stalled with compound 028 represents a novel intermediate between B and B<sup>act</sup>, in which U4/U6 proteins have dissociated but not most B-specific proteins nor the U6 Lsm proteins. On the other hand, Prp19/CDC5L proteins are clearly underrepresented, if present at all, and B<sup>act</sup> proteins have not yet associated. Thus, structural and functional investigation of B<sup>028</sup> may provide insight into the dynamic RNP rearrangements occurring during the activation step.

### 3.2.7 Attempts to isolate more homogeneous B<sup>028</sup> complexes

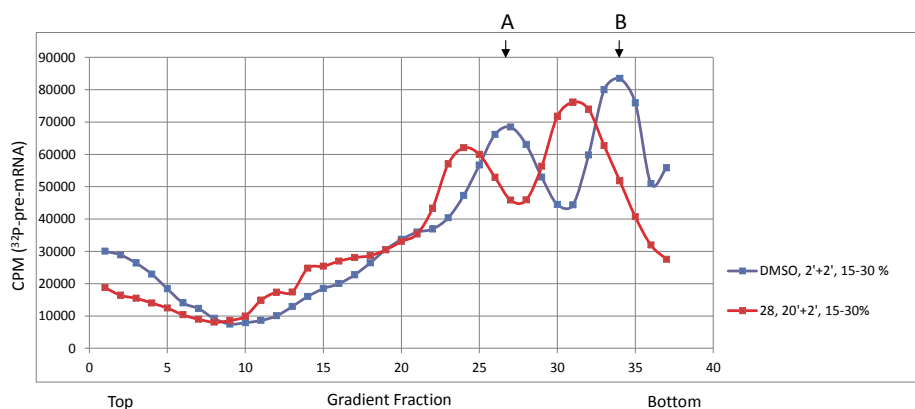
The RNA composition of B<sup>028</sup> spliceosomes suggests there is a mixture of complexes lacking U1 and U4 with those still containing them. We thus checked whether more pronounced changes in the RNA composition of B<sup>028</sup> complexes, namely, complete loss of U1 and U4 occur at later time points. We thus performed the splicing reaction for 20, 40 or 60 min and isolated the spliceosomal complexes using MS2 affinity selection. The RNA composition of the A and B complexes formed in the presence of compound 028 appeared to be identical even at later (40 and 60 min) time points (Fig. 3.14). Also, no significant changes in the protein composition of B<sup>028</sup> were observed by mass spectrometry at later incubation times (data not shown).



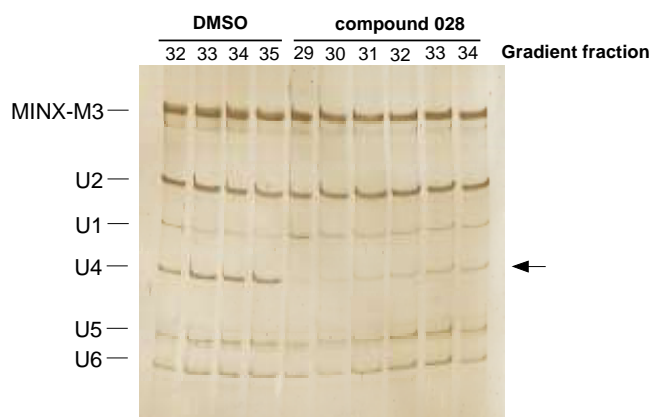
**Figure 3.14 RNA composition of MS2-affinity purified B<sup>028</sup> complexes formed in the presence of compound 028 at different time points of the splicing reaction.** RNA recovered from complexes affinity-purified from peaks I and II was analysed by denaturing PAGE and detected by silver staining. RNA identities are indicated at the left.

We also purified complexes from every fraction of the B<sup>028</sup>-containing peak and analysed their RNA composition (Fig. 3.15). In the fractions located closer to the top (# 29-30) higher amounts of U1 snRNA were detected, probably due to contamination from the neighbouring peak containing complex A. On the other hand, the fractions located closer to the bottom (# 33-34) displayed slightly increased amounts of U4 snRNA, suggesting the presence of B spliceosomal complexes. Thus, highly purified B<sup>028</sup> complexes subsequently used for structural and functional studies were taken only from the middle of the peak.

A.



B.

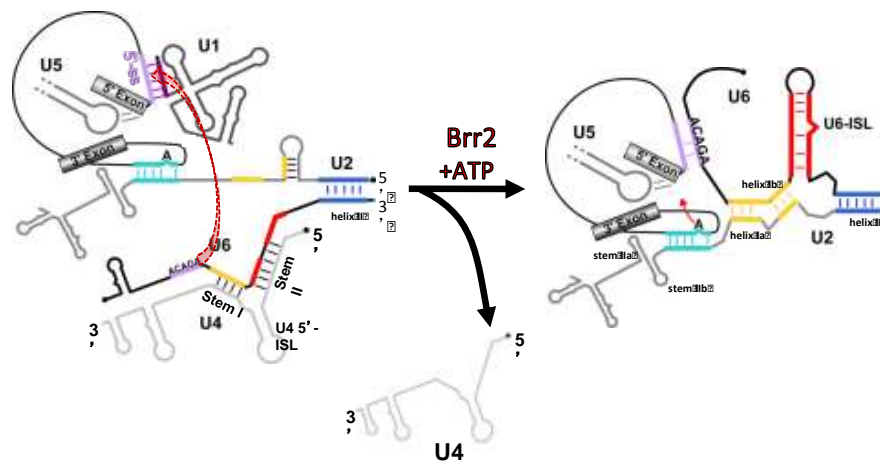


**Figure 3.15 Purification of stalled B<sup>028</sup> complexes and characterization of different fractions of the B-like peak.** A. Glycerol gradient sedimentation profile of the spliceosomal complexes formed under splicing conditions in the presence/absence of compound 028. B. RNA composition of MS2-affinity purified splicing complexes formed in the presence of compound 028. RNA recovered from the indicated fractions of each gradient was analysed by denaturing PAGE and detected by silver staining. RNA identities are indicated at the left.

### 3.2.8 Analysis of RNA-RNA interactions in spliceosomes stalled by compound 028

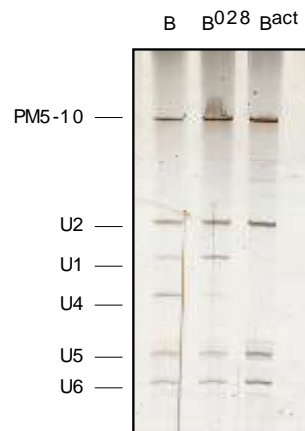
During activation, the RNA-RNA network of the spliceosome is rearranged. Unwinding of the U4/U6 base pairing by Brr2 enables U6 to form new base pairs with U2 (U2/U6 helix Ia and Ib) and the pre-RNA's 5' splice site. Also a catalytically important internal

stem loop (U6 ISL) is formed (Fig. 3.16). A major question is whether U6 has rearranged in the  $B^{028}$  complex, and what the RNA-RNA network looks like in this complex.



**Figure 3.16 Dynamic network of RNA-RNA interactions in the spliceosome.** Exon sequences are highlighted as grey boxes and the intron sequence as black line. Secondary structures of snRNA are shown schematically in grey or black, and the regions involved in base-pair interactions (short lines) are shown in different colours. The 5' end of an snRNA is labelled as a black dot. Solely loop I of the U5 snRNA is shown. During the transition of a pre-catalytic (left scheme) to the catalytically-activated spliceosome (right scheme) U1 and U4 leave, and U6 and U2 are involved in new base-pair interactions.

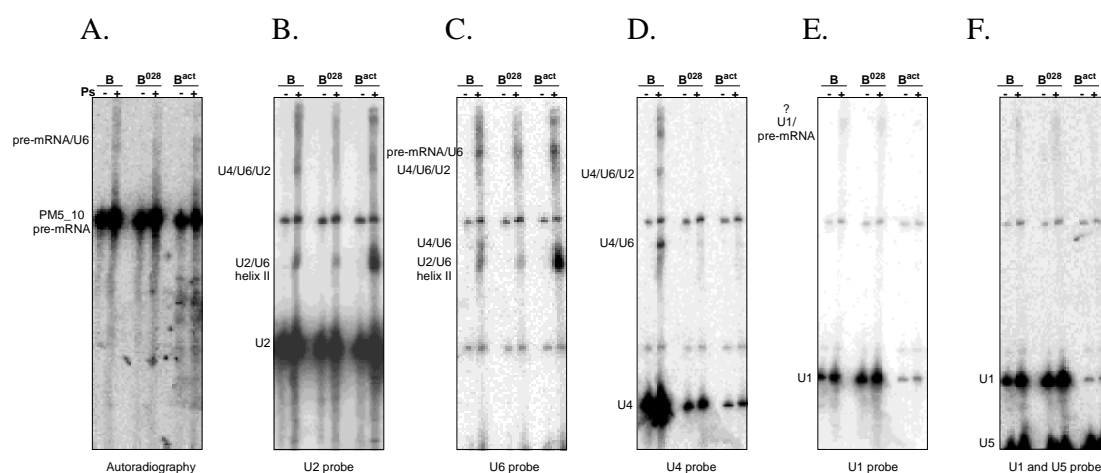
To study RNA-RNA interactions in  $B^{028}$ , complexes were formed under splicing conditions *in vitro*, and purified by MS2 affinity selection, using PM5-10 and/or MINX pre-mRNAs as substrates. B and  $B^{act}$  complexes were isolated in parallel. RNA denaturing gel analysis demonstrated that complexes formed on PM5-10 in the presence of 028, have a similar RNA composition relative to those formed on MINX (Fig. 3.17).



**Figure 3.17 RNA composition of MS2 affinity-purified splicing complexes B,  $B^{028}$  and  $B^{act}$  formed on PM5-10 and used for structure probing studies.** RNA recovered from the complexes was analysed by denaturing PAGE and detected by silver staining. RNA identities are indicated at the left.

### 3.2.8.1 Identification of RNA-RNA interactions in B<sup>028</sup> complexes via psoralen crosslinking

RNA-RNA interactions in B<sup>028</sup> were first monitored using UV light-induced psoralen-mediated crosslinking in collaboration with Dr. M. Anokhina. Psoralen crosslinking is an efficient method that allows the identification the double-stranded (base paired) regions and higher-order structure within the RNA molecule. Psoralen is an aromatic tricyclic furocoumarin, which intercalates between RNA base pairs and photoreacts with double-stranded regions via formation of cycloadducts with pyrimidine bases. This reaction is initiated by the long wavelength UV irradiation of nucleic acids at 364 nm in the presence of psoralen upon the absorption of the photon to form monoadducts, and the absorption of other photon results in the formation of a crosslink between the strands. The crosslinked regions of the RNA(s) can be followed by northern blot analysis by sequential probing, or by direct autoradiography of blot to detect the position of <sup>32</sup>P-labelled crosslinked pre-mRNA. The migration of different bands detected with each probe can be compared and correlated with each other to identify the cross-linked RNA. If a band is detected by two different probes we can conclude it represents a crosslink between those two snRNAs.



**Figure 3.18 Identification of RNA-RNA interactions via psoralen crosslinking in purified complexes B, B<sup>028</sup> and B<sup>act</sup>.** RNAs were visualized by Northern blot analysis with <sup>32</sup>P-probes against U2, U6, U4, U1 and U5 and/or the <sup>32</sup>P-labelled pre-RNA was visualized by autoradiography. Ps: psoralen. Crosslinks and the PM5-10 pre-mRNA are indicated on the left of the panel.

Purified B, B<sup>028</sup> and B<sup>act</sup> spliceosomes formed on PM5-10 pre-mRNA were subjected to long-wave UV irradiation in the presence of psoralen, and RNA was separated by gel electrophoresis. Northern blotting was performed with <sup>32</sup>P-labelled probes against U1, U2, U4, U5 and U6 sequentially. First, autoradiography revealed the position of



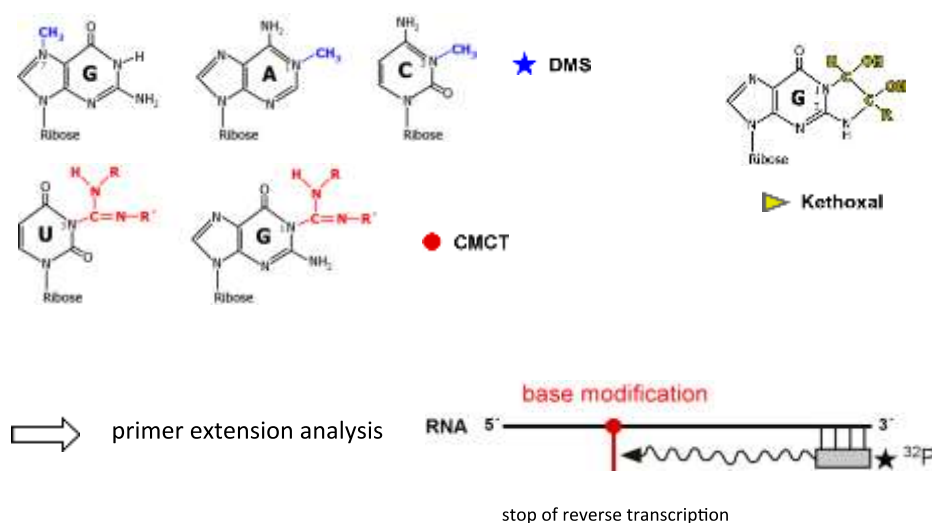
crosslinks containing radioactively labelled PM5-10 pre-mRNA (Fig. 3.18A). Hybridization with  $^{32}\text{P}$ -probes against U2 and U6 revealed a U2/U6 crosslink, based on the signal from both U2 and U6 probes and its migration behaviour indicates it is U2/U6 helix II crosslink (Anokhina et al., 2013). U2/U6 helix II was detected mainly in  $\text{B}^{\text{act}}$  complex and to a lesser extent in B and  $\text{B}^{028}$  complexes, suggesting it was shielded in  $\text{B}/\text{B}^{028}$  by proteins and therefore less accessible to psoralen (Fig. 3.18, B and C). In addition, U4/U6 crosslinks were also detected with both U4 and U6 probes in complex B, but not in  $\text{B}^{028}$  and  $\text{B}^{\text{act}}$ , due to release of U4 snRNA from  $\text{B}^{028}$  and  $\text{B}^{\text{act}}$  complexes (Fig. 3.18, C and D). Similarly, a U2/U4/U6 triple crosslink was detected with U2, U4 and U6 probes in complex B, but not in  $\text{B}^{028}$  and  $\text{B}^{\text{act}}$ , confirming the absence of this crosslink, due to the absence of U4 snRNA at later stages of spliceosomal assembly. We didn't observe any strong crosslinks with U1 probes in all three complexes, except a very weak slow migrating signal was detected in both B and  $\text{B}^{028}$  complexes, due to the presence small amounts of U1 snRNA (Fig. 3.18E), suggesting it might be the crosslink of U1 base paired with the 5' splice site. Additionally, a crosslink between U6 snRNA and the pre-mRNA was detected in all three complexes, providing evidence for the base pairing interaction of the U6 ACAGAG box intron nucleotides near the 5'ss. No crosslinks were detected with a U5 probe. However, lack of a crosslink doesn't mean there were no interactions.

### 3.2.8.2 Structure of U5 snRNA in $\text{B}^{028}$ complexes

To determine at which stage of activation compound 028 blocks splicing we investigated RNA-RNA interactions in  $\text{B}^{028}$  complex by performing RNA structure probing. Structure probing allows us to test the reactivity of each nucleotide towards specific chemical agents and to determine the secondary and tertiary structure of RNA molecules, and also whether they are engaged in interactions with proteins, via detection of the modified nucleotides. Bases in the RNA molecule are modified if they do not interact with other RNAs, i.e. it shows if a particular nucleotide in an RNA molecule is base paired or not. To study at which stage of RNA rearrangement compound 028 blocks spliceosome assembly, structure probing was performed with B,  $\text{B}^{028}$  and  $\text{B}^{\text{act}}$  complexes. All structure probing experiments were performed in collaboration with Dr. Maria Anokhina from our laboratory.

Chemical modifications were performed with three chemical reagents: 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT), which modifies bases at positions normally involved in Watson-Crick interactions, mainly at positions

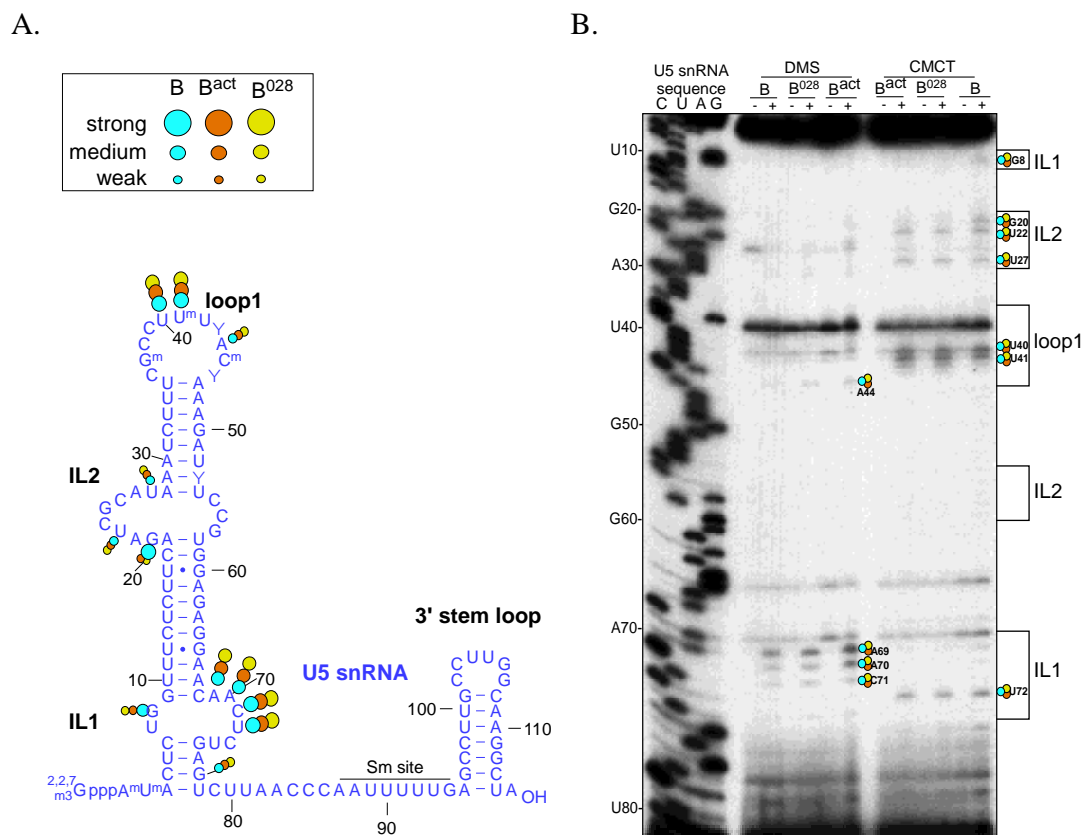
N1-G and N3-U, dimethylsulfate (DMS) which reacts with N1-A and N3-C, and  $\beta$ -ethoxy- $\alpha$ -ketobutyraldehyde (Kethoxal), which modifies specifically N1-G and N2-G (Fig. 3.19; described in Material and Methods). After incubating spliceosomes with CMCT, DMS or Kethoxal, RNA was isolated by proteinase K digestion followed by phenol/chloroform extraction. Primer extension assays were then performed with 5'-end labelled oligonucleotides complementary to chosen sequences in the U snRNAs. Primer extension occurs (in the direction 3' to 5' end) by adding reverse transcriptase (RT). If chemical modification takes place on the RNA template, the elongation of the primer will be stopped exactly one nucleotide before the chemical modification point, resulting in truncated cDNA fragments instead of fully-elongated chains. To obtain a sequencing ladder, *in vitro* transcribed snRNAs were treated in an identical manner, but in the presence of 0.5 mM dideoxy-NTPs. Reverse transcripts were separated on denaturing polyacrylamide sequencing gels and visualized by autoradiography. Accessibility to a given chemical was indicated as strong, medium or weak depending the intensity of the RT-stop (minus the background which is seen without DMS, CMCT or Kethoxal).



**Figure 3.19** Positions of chemical modifications by chemical reagents DMS, CMCT and Kethoxal and schematic view of primer extension using modified RNA as a template.

U5 snRNA undergoes only minor changes during the B to B<sup>act</sup> transition, with IL2 and loop 1 more accessible in complex B versus B<sup>act</sup> (Anokhina et al, 2012) (Fig. 3.20), because the latter contacts the 5' exon before step I of splicing (Wyatt et al., 1992; Sontheimer and Steitz, 1993). Analysis of U5 modification in B, B<sup>028</sup> and B<sup>act</sup> complexes revealed that all regions involved in known internal base pair interactions were fully-

protected in all three complexes. The RNA protection pattern observed within the complex B<sup>028</sup> is identical to that seen with complex B<sup>act</sup> (Fig. 3.20). In particular, the internal loop 2 nt G20 and the internal loop 1 nt G8 are more protected in B<sup>028</sup> and B<sup>act</sup>, compared to B, which indicates less accessibility in these regions. Nucleotides U42 and Ψ43 were found to be unreactive against chemical modification in all three tested complexes, consistent with U5 interacting with the 5' exon of the pre-mRNA via these nucleotides. Nucleotides involved in formation of the internal loops 1 and 2 (IL1 and IL2, respectively) have a nearly identical modification pattern in B, B<sup>028</sup> and B<sup>act</sup>, confirming that the structure of this region of U5 is not affected by the presence of compound 028.



**Figure 3.20 U5 snRNA protection pattern in spliceosomal complexes is identical in B<sup>028</sup> and B<sup>act</sup> complexes.** A. Schematic summary of U5 snRNA modification pattern in different spliceosomal complexes. The different sizes and colours of the dots indicate different modification accessibility for different spliceosomal complexes. B. Chemical modification pattern of B, B<sup>act</sup> and B<sup>028</sup> complexes formed on PM5-10 treated with the indicated reagents. (-) = no reagent (reveals background stops). Primer extension was performed with oligonucleotides complementary to U5 nts 83-103.

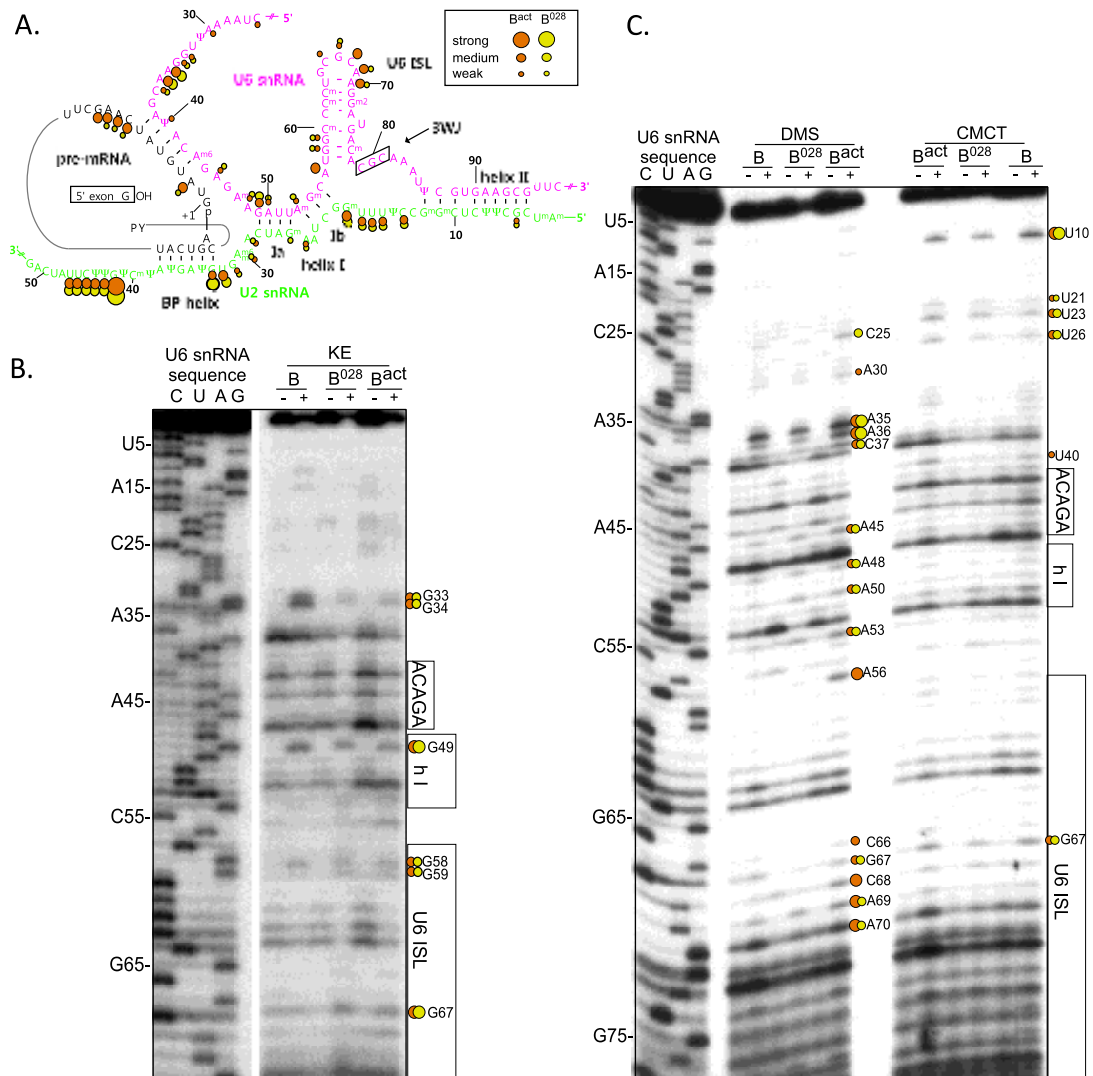
### 3.2.8.3 Structure of U2 and U6 snRNAs in B<sup>028</sup> complexes

To clarify to which extent the interactions between U2 and U6 are formed in B<sup>028</sup> complexes and whether U2/U6 helices, which are typical for B<sup>act</sup>, are present in B<sup>028</sup>, we

investigated the U6 and U2 snRNA protection patterns in B, B<sup>028</sup> and B<sup>act</sup>. In the B complex, U4 and U6 snRNAs are engaged in base pair interactions building two long helices, stem I and stem II. This U4/U6 structure does not exist in B<sup>act</sup>, because U4 is released. In the activated spliceosome, U6 snRNA undergoes remarkable structural rearrangements after Brr2-catalysed U4/U6 unwinding and U4 release. At this time U6 snRNA base pairs with the 5' splice site and the U2 snRNA, and forms an internal stem-loop (U6 ISL) (Fig. 3.16, 3.21A). This U6 ISL is extremely important for catalysis and is present in B<sup>act</sup>. We first focused on the accessibility of nucleotides forming the U6 ISL. Nucleotides involved in the loop (A70-G67) of the U6 ISL are highly accessible in B<sup>act</sup> and there is a very low modification of nucleotides involved in the formation of the stem of the ISL (C61-C63, G71-A73) (Fig. 3.21). Nucleotides involved in the formation of the U6 ISL loop are more protected in B<sup>028</sup> compared to those in B<sup>act</sup> (A69, C68, C66). However nucleotides G67, A69 and A70 are accessible to a low extent. This enhanced protection of loop nucleotides of the putative ISL sequence in B<sup>028</sup> could potentially be explained by shielding due to bound proteins during modification. The other U6 nucleotides involved in the ISL formation are protected against modification in B<sup>028</sup>, such as nts A56-G65, suggesting that the stem of the U6 ISL is formed. Because they are close to the 3' end of U6, it was not possible to analyse the modification of nts on the other side of the stem. Moreover, U6 nucleotide A56 that is located close to the ISL sequence is well-protected in B<sup>028</sup>, compared to reproducible accessibility of this particular nucleotide in B<sup>act</sup>, confirming that this nucleotide is either base-paired or shielded by proteins in B<sup>028</sup>. These results suggest that the RNP structure in this region is clearly different between B<sup>028</sup> and B<sup>act</sup>. However, because the U6 ISL loop nucleotides are only poorly accessible, we cannot conclude with certainty that U6 ISL has formed in B<sup>028</sup>.

In B complex, the U6 nucleotides A50-C55 are base paired with nucleotides from U4 snRNA under the formation of stem I. After U4 release, these U6 nucleotides interact with 5' end of U2 snRNA to form U2/U6 helix I. U2/U6 helix I together with U6 ISL represent the catalytic centre of the spliceosome, and thus the RNA structure in these regions is of particular interest. The U6 snRNA modification pattern analysis revealed that U6 nucleotides A50-U52 are relatively well-protected in both B<sup>028</sup> and B<sup>act</sup> with only weak modification of A50 (G49 is accessible for chemical modification in both, B<sup>act</sup> and B<sup>028</sup>) (Fig. 3.21), suggesting U2/U6 helix Ia is formed in the B<sup>028</sup> complex. Complementary U2 nucleotides involved in helix Ia formation (G25-C28) are also protected against modification (Fig. 3.21A, 3.22). U6 nucleotides involved in the formation of U2/U6 helix Ib, i.e. nucleotides A53-C55, appear to be also protected

against chemical modification in both,  $B^{028}$  and  $B^{act}$  complexes (Fig. 3.21). In addition, U2 nucleotides (G20-U22) are also protected (Fig. 3.21A and 3.22), suggesting helix Ib is also formed in  $B^{028}$ .

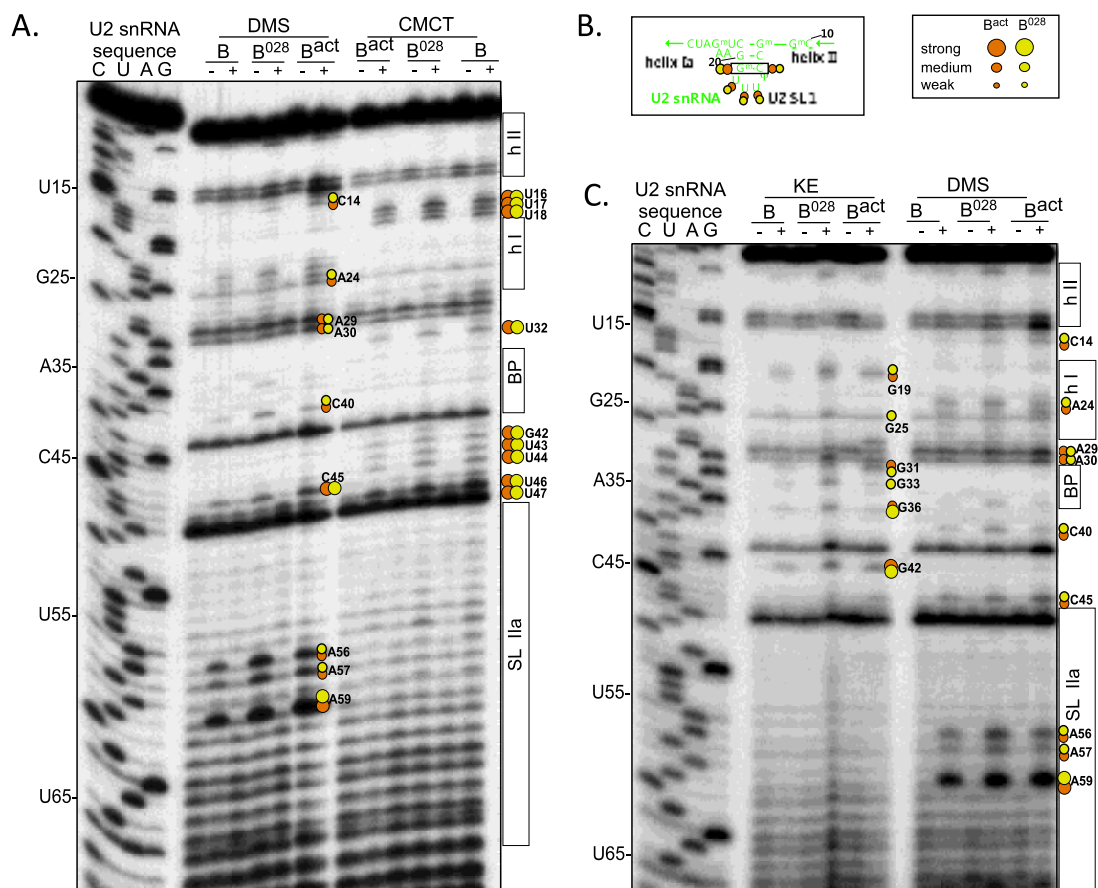


**Figure 3.21 Primer extension analysis of U6 snRNAs in chemically modified B,  $B^{028}$  and  $B^{act}$  complexes.** A. Schematic summary of modification pattern for U2 and U6 snRNA interactions in spliceosomal complexes  $B^{028}$  and  $B^{act}$ . Different size and colour of the dots indicate different degrees of modification accessibility in the different spliceosomal complexes. B and C. Primer extension analysis of U6 snRNA from CMCT, DMS and KE modified complexes. Chemical modification pattern of B,  $B^{028}$  and  $B^{act}$  complexes formed on PM5-10 treated with reagents indicated. Primer extension was performed with oligonucleotides complementary to U6 nts 80-100.

U2/U6 helix II, which involves base pairing interactions between the 5' end of the U2 snRNA and 3' region of the U6 snRNA (Madhani and Guthrie, 1992) appears to be formed in all three tested complexes based on the chemical inaccessibility of U2

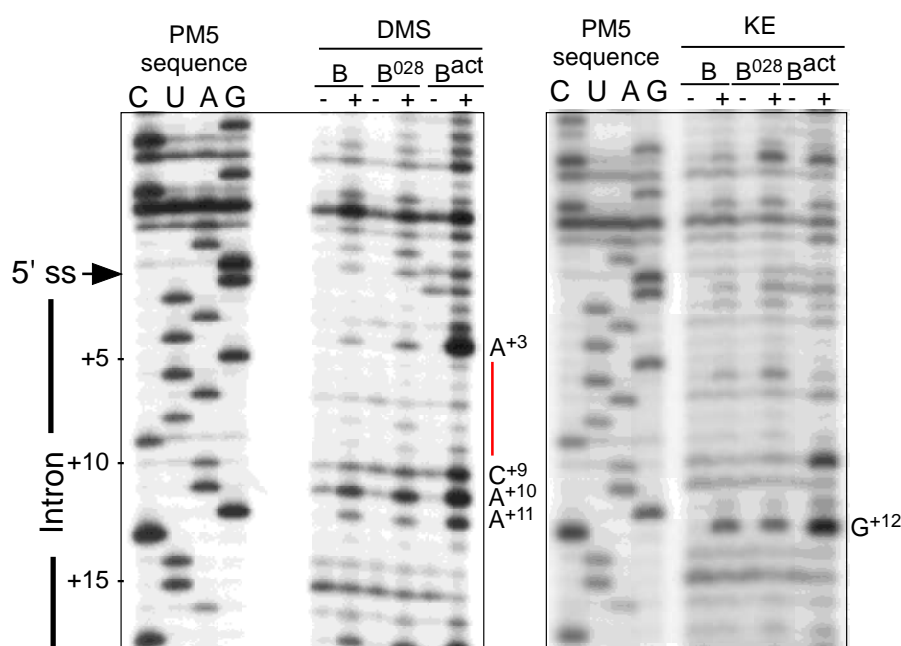
nucleotides 3-11 of this helix (Fig. 3.21A, 3.22 A and C). The accessibility of this RNA region in U6 snRNA wasn't analysed, because its location on the RNA prevents primer extension analysis (it is too close to the 3' end of U6).

Previous work from our lab showed that U2/U6 interaction forms a 3-way junction in  $B^{act}$  as shown in Fig. 3.21 (Anokhina et al., 2013). An alternative conformation of U2/U6 (a 4-way junction) was also previously proposed (Sun and Manley, 1995). In the proposed four-way junction model of U6/U2 interaction, but not in the 3-way junction model, the formation of U2 stem loop 1 (SL1) should be observed. Our results do not support the existence of this loop in both  $B^{028}$  and  $B^{act}$  complexes, based on accessibility of U2 nts C14 and G19, which should be base paired if there is formation of the short U2 stem in this model (Fig. 3.22). However, additional analyses after the extension of the 3' end of U6 snRNA are necessary to rule out the 4W junction structure with certainty.



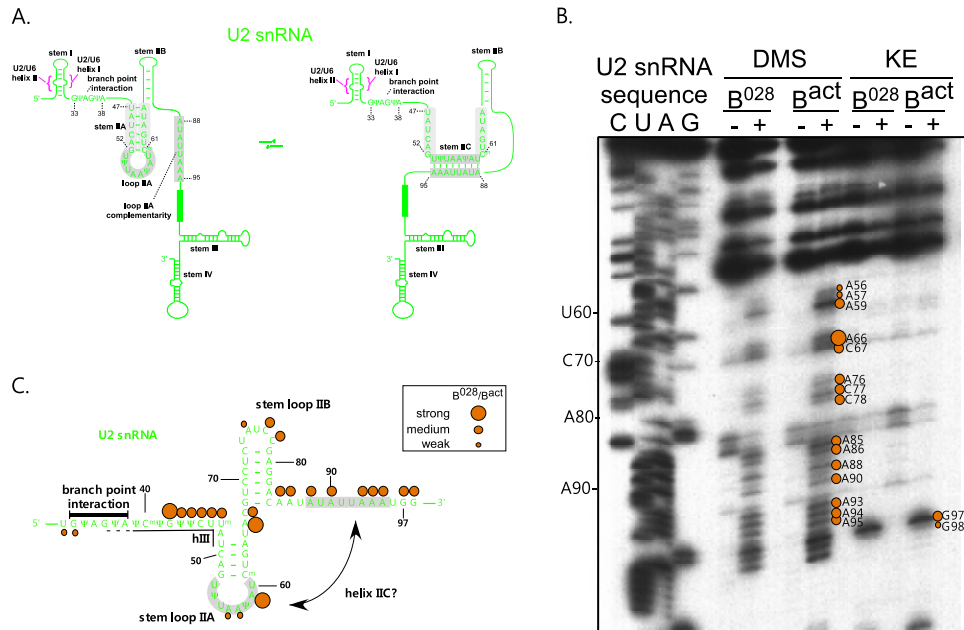
**Figure 3.22 U2 snRNA chemical modification pattern in B, B<sup>act</sup> and B<sup>028</sup> complexes.** A and C. Primer extension analysis of U2 snRNA from DMS, CMCT and KE treated complexes. Primer extension was performed with oligonucleotides complementary to U2 nts 75-95 and 146-163. B. Scheme of modification pattern for U2 snRNA in the four-way junction (4WJ) model.

The nucleotides from the U6 ACAGA box (A40-A44 nts) were found to be protected against chemical modification in both B<sup>act</sup> and B<sup>028</sup> complexes (Fig. 3.20). Likewise, intron nts +4 to +8 were also protected from modification in B, B<sup>028</sup> and B<sup>act</sup> complexes (Fig. 3.23). Thus, the interactions of the U6 ACAGA box (A41-A45) with the pre-mRNA intron near the 5' ss appears to be formed in B<sup>028</sup> complexes. Chemical modification also supported the formation of the U2/BPS interaction; all U2 nucleotides involved (G33-A38) with the exception of G33 were not accessible in B<sup>028</sup> or B<sup>act</sup> (Fig. 3.21A, 3.22).



**Figure 3.23 Pre-mRNA chemical modification pattern in B, B<sup>act</sup> and B<sup>028</sup> complexes.** Primer extension analysis of the pre-mRNA from DMS and KE treated complexes. Primer extension was performed with an oligonucleotide complementary to PM5 pre-mRNA nts 281-262.

Previously, it was reported that yeast U2 snRNA toggles between different intra-molecular conformations (Fig. 3.24A). This includes the formation of so called helix IIc, which involves base pairing between loop nucleotides (53-60) of U2 stem loop IIa and nucleotides downstream of stem loop IIb (88-96) (Hilliker et al., 2007; Perriman et al., 2007). As shown in figure 3.24, U2 nucleotides A88-A95 and U53-U60, which form base pairs if helix IIc is formed, are accessible to all reagents tested in B<sup>act</sup> and B<sup>028</sup> complexes. Thus, U2 helix IIc is not formed in the B<sup>028</sup> complex; nor in all other human spliceosomal complexes tested thus far (Fig. 3.24).



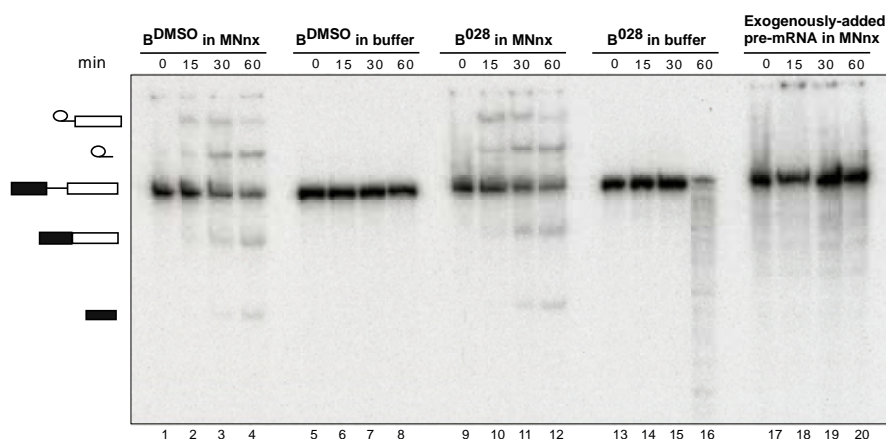
**Figure 3.24 U2 snRNA protection pattern in spliceosomal complexes reveals that U2 helix IIc is not formed in the B<sup>028</sup> complex.** A. Structure of U2 stem IIc and stem-loop IIa. Nucleotides involved in stem IIc and stem loop IIa are shadowed. The nucleotides of U2 that bind the intronic branch site consensus sequence are shown. B. Chemical modification of B<sup>act</sup> and B<sup>028</sup> complexes with the reagents indicated. (-) = background. Primer extension was performed with oligonucleotides complementary to U2 nts 75-95 and 146-163 and analyzed by denaturing PAGE. C. Schematic of U2 snRNA modification pattern in spliceosomal complexes B<sup>028</sup> and B<sup>act</sup>. Different size and colour of the dots indicate different degrees of modification accessibility for different spliceosomal complexes.

### 3.2.9 B<sup>028</sup> complexes can be chased into catalytically active spliceosomes.

We next tried to elucidate whether the complex formed in the presence of compound 028 is functionally active or a dead-end complex. To this end, affinity-purified splicing complexes were incubated under *in vitro* splicing conditions either with buffer alone or with nuclear extract treated with micrococcal nuclease (MN). When incubated with micrococcal nuclease, the snRNAs are digested leading to disruption of all snRNPs. An excess of unlabelled pre-mRNA was also added together with the MN-treated extract to rule out a disassembly of B<sup>028</sup> complexes with subsequent reassembly of spliceosomes as a possible cause of the observed splicing activity. When incubated in the presence of MN-treated HeLa nuclear extract, the B<sup>028</sup> complexes can be chased into catalytically active spliceosomes with the same efficiency of the first and second steps of splicing (Fig. 3.25, lanes 9-12) as B complexes chased in parallel as a control (Fig. 3.25, lanes 1-4). In contrast, when B<sup>028</sup> or B complexes were incubated in the absence of nuclear extract, no splicing was observed in both cases (Fig. 3.25, lanes 5-8 and 13-16). This shows that protein factors in nuclear extract necessary for subsequent spliceosome activation and



catalytic activity can bind to B or B<sup>028</sup>. Exogenously-added, <sup>32</sup>P-labelled pre-mRNA incubated with MN-digested nuclear extract under splicing conditions is not spliced, indicating that the treatment of nuclear extract with MN was efficient (Fig. 3.25, lanes 17-20). We conclude that the affinity-purified stalled B<sup>028</sup> complexes are not dead-end complexes that are irreversibly blocked at a step between B and B<sup>act</sup> complexes, but instead are a functional assembly intermediate.



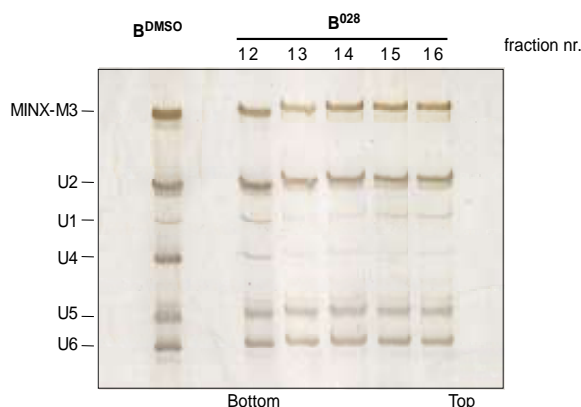
**Figure 3.25 B<sup>028</sup> complexes can be chased into catalytically active spliceosomes in the presence HeLa nuclear extract treated with micrococcal nuclease.** The affinity-purified splicing complexes were incubated under *in vitro* splicing conditions either with buffer alone (lanes 5-8 and 13-16) or with nuclear extract treated with micrococcal nuclease plus an excess of unlabelled pre-mRNA (lanes 1-4 and 9-12) for the indicated time points. Lanes 17-20: [<sup>32</sup>P]-labeled pre-mRNA was incubated in the presence of micrococcal nuclease-treated HeLa nuclear extract. RNA was isolated, analyzed by denaturing PAGE and visualized by autoradiography. Position of splicing products, intermediates or unspliced MINX pre-mRNA is indicated at the left.

### 3.3 Electron microscopy studies of affinity-purified B<sup>028</sup> complexes

The spliceosome undergoes major rearrangements concerning its RNA and protein composition during the activation step. These remarkable changes are reflected in different morphologies of the spliceosomal complex B and subsequently formed activated spliceosomes (B<sup>act</sup>) that have been visualized by electron microscopy (EM) (Bessonov et al., 2010). A better understanding of this highly dynamic step is of particular importance, since this decisive step prepares the spliceosome to fulfil its main task, namely, the two subsequently performed transesterification reactions. Using electron microscopy, we endeavoured to determine the structure of complex B<sup>028</sup>, which represents an intermediate step during the transition of spliceosomal complex B to an activated spliceosome, to provide insight in the structural dynamics of this particular step. EM was performed in collaboration with Dr. Berthold Kastner, Dr. Prakash Dube and Prof. Dr. Holger Stark.

### 3.3.1 Negative stain 2D electron microscopy of affinity-purified B<sup>028</sup> complexes

To obtain the first images of the stalled B<sup>028</sup> complex, we performed two-dimensional negative stain EM. We thus first isolated B<sup>028</sup> spliceosomes via the MS2 affinity selection method (Fig. 3.10A). We then subjected the eluted B<sup>028</sup> complexes to a second glycerol gradient centrifugation step, using the GraFix procedure (Kastner et al., 2008). The GraFix method stabilizes the individual particles by mild chemical fixation during the centrifugation step, and protects them against aggregation. For this purpose, the glycerol gradient is supplemented with a fixation agent, namely an increasing concentration of glutaraldehyde (0 to 0.1 %). During this second centrifugation step, the particles of interest can also be separated from aggregates and defective particles, reducing the heterogeneity. RNA analysis of the purified B<sup>028</sup>-containing fractions of the second gradient revealed an RNA pattern typical for B<sup>028</sup> (Fig. 3.26). The B<sup>028</sup>-containing peak fractions were then subjected to negative staining EM performed by Dr. Berthold Kastner and Dr. Prakash Dube. Thereby, particles were absorbed onto a carbon film, using the single carbon film method, and negatively-contrasted with 2 % uranyl formate.

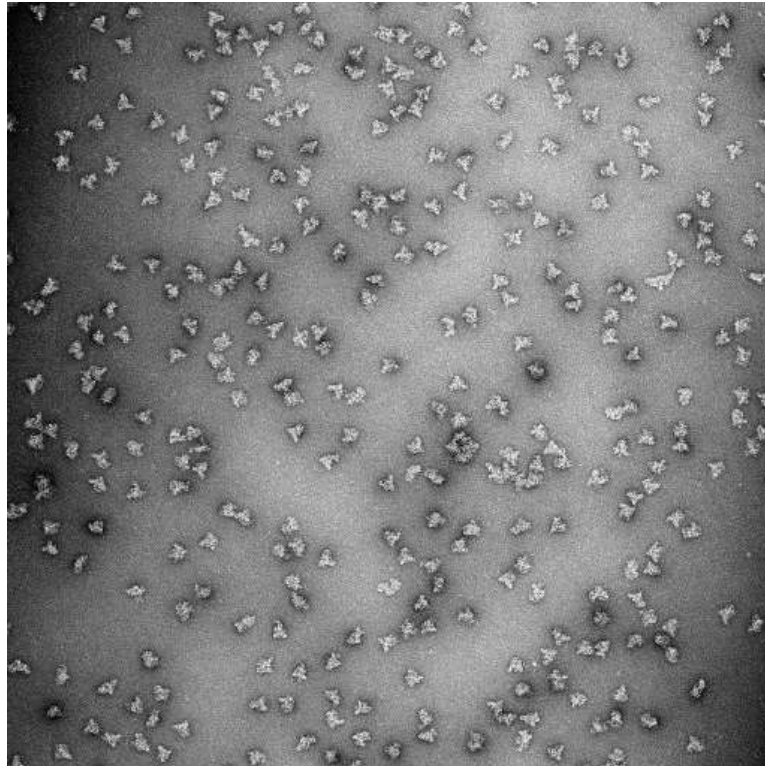


**Figure 3.26 RNA composition of the purified B<sup>028</sup>-containing fractions compared to B complexes formed in the presence of DMSO.** RNA recovered from splicing complexes from each fraction was analysed by denaturing PAGE and detected by silver staining. RNA identities are indicated at the left.

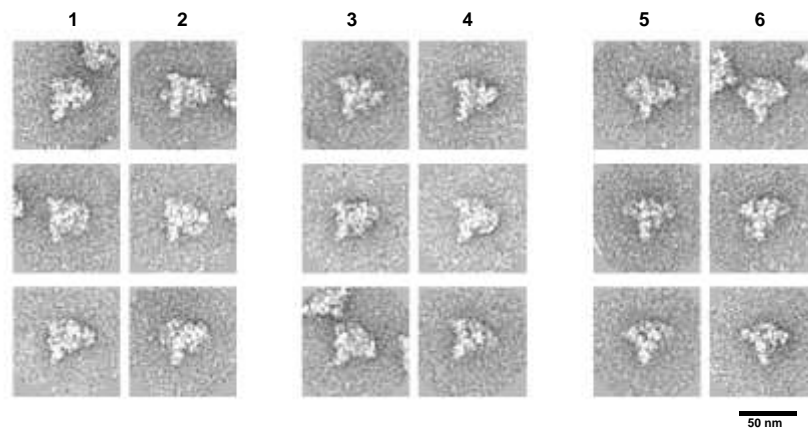
An overview of the negatively-stained raw images revealed a homogeneous population of monodisperse particles of the same basic morphology (Fig. 3.27A) with a maximum dimension of about approximately 40 nm. The majority of the B<sup>028</sup> images exhibit a roughly triangular shape. One corner of the triangle appears mostly as a stick-like protrusion. In the gallery of typical B<sup>028</sup> images (Fig. 3.27B) this protrusion is oriented to the bottom, thereby making it the foot of the particle. From this foot a ca 10 nm wide density runs up to the top in most images. This density axis is positioned in most cases directly at the left edge of the triangular shaped images (columns 1-4). The remaining

part of the images has a more globular shape. It is positioned at the upper 2/3 of the elongated axis and makes up the right corner of the triangle. In more rare cases the B<sup>028</sup> images have a mushroom-like shape (columns 5-6). Here, the elongated density appears to be bent or is not traceable at all. Some images show features of both the triangular and mushroom shape. An example is shown in the last row of column 4 and 5, where one and the same image is oriented according to the triangular view (column 4) and the mushroom view (column 5). This suggests that both types of views can be generated by one and the same 3D structure.

A.

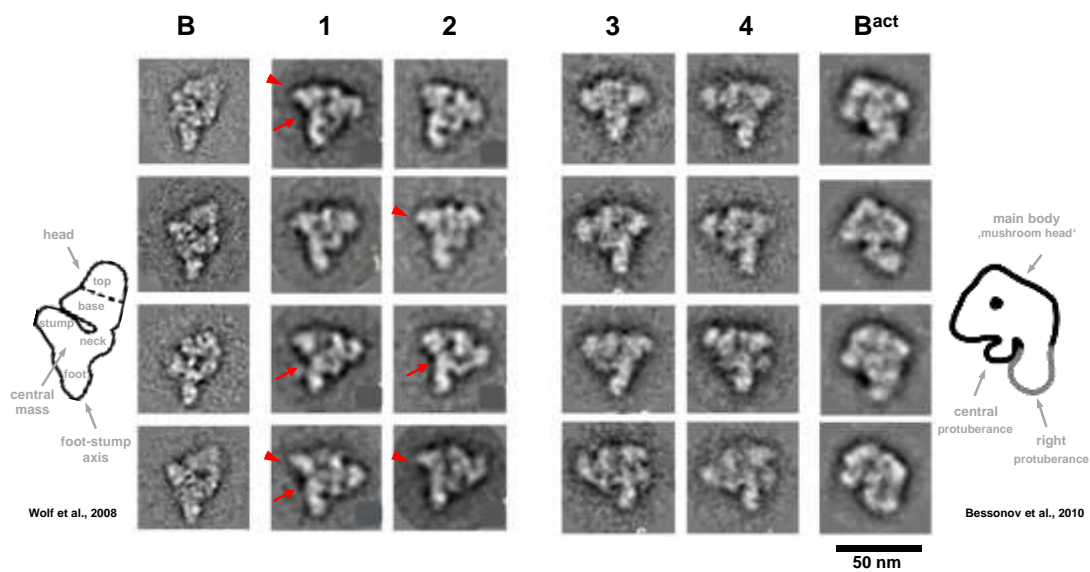


B.



**Figure 3.27 Electron microscopy of negatively-stained B<sup>028</sup> particles assembled on the MINX pre-mRNA substrate.** (A) Overview of raw images of the B<sup>028</sup> complex obtained by negative stain EM. (B) Gallery of raw images of stalled B<sup>028</sup> complex with the most frequently found views. Scale bars correspond to 50 nm.

The single particle images were then subjected to image processing. After several rounds of iterative translational and rotational image alignment and image classification, the data set was partitioned into 200 image classes and the class members averaged. With the class averages (Fig. 3.28, columns 1-4) more structural features of the  $B^{028}$  particle are now recognizable. Dominant are classes showing the typical triangular shapes with the long density axis on the left (columns 1-2). At the upper one-third, a dividing line of stain (see arrow) can often be seen at the 40 nm long and on average 10-12 nm wide axis. Typically, the delimited upper one-third density part frequently sticks out to the side like a beak (see arrowheads). The globular density element shows a lot of fine structure in the class averages. The complex distribution of stain accumulated in cavities and holes in the particle indicates an intricate network of structural elements. The classes with the typical mushroom-like shape (column 3-4, first three rows) were observed less frequently, similar to the classes showing both mushroom and triangular shape features (columns 3-4, last row).



**Figure 3.28  $B^{028}$  particles exhibit a morphology between  $B$  and  $B^{act}$ .** Selected class averages of the negatively-stained, stalled  $B^{028}$  complex (columns 1-4) compared to published spliceosomal complexes  $B$  (left) and activated spliceosomes ( $B^{act}$ ) (right). The columns 1-2 versus columns 3-4 show the two major types of orientation of  $B^{028}$  particles on the carbon film which appear to be mirror images. Scale bars correspond to 50 nm.

The views of the  $B^{028}$  particle show in their basic morphology similarities to the  $B$  complex (Fig. 3.28, left column), when the triangular views are considered, and similarities to the  $B^{act}$  complex (right column), when the mushroom views are compared.

The B complex shows dominant views with a rhombic and triangular shape (Deckert et al., 2006). In these views an elongated axis running along one edge also prevails (“foot-stump axis”). There is a protrusion at the lower end of the B complex axis (“foot”) that shares similarities with the foot region of the B<sup>028</sup> complex. The upper end of this B complex main axis, called the “stump” - structural designations are according to Boehringer et al. (2004), Deckert et al. (2006) and Wolf et al. (2009) - appears a bit different as the protruding beak-like feature is missing. Beside the axis-like element, the B complex images also have one additional domain that is attached to the upper part of the elongated structure (“head”) via short a “neck” region. But here it is in an upper right position pointing more upwards and even exceeds the height of the long axis, thereby creating rhombic-shaped views. This different arrangement makes up the most obvious global morphological difference compared to the B<sup>028</sup> triangular views. The globular density is oriented in the B<sup>028</sup> particle more toward the central region of the long density axis, creating the shape more like an isosceles triangle. The globular domain in the B complex classes shows much less fine structure than seen at the B<sup>028</sup> complex. This domain might be more compact in the B complex.

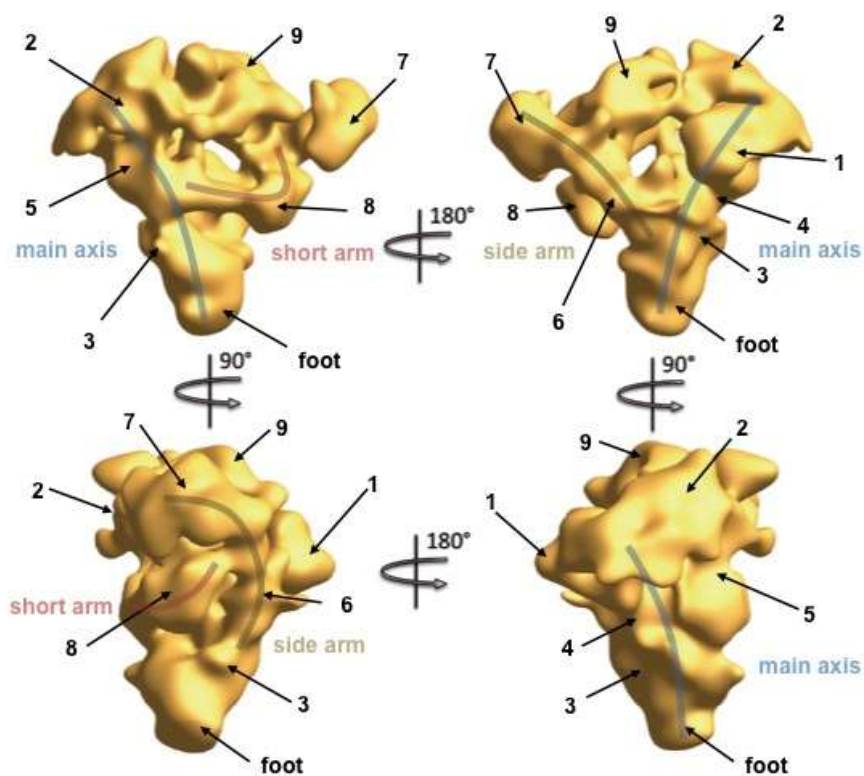
Mushroom-like views are not typical for the B complex but rather dominate in images of the B<sup>act</sup> complex. The dominant B<sup>act</sup> views show typically a centrally elevated and upward-oriented main body with a central stain accumulation and a more complex foot domain, consisting of two protuberances, which are situated in the center and at the right in the dominant classes (Fig. 3.28, right). The appearance of the main body seems to be mostly invariant, whereas the shape and position of the protuberances differ significantly in the various classes. In general the structural similarities between the B<sup>act</sup> (right column) and B<sup>028</sup> views are quite limited and the structural details seen in B<sup>028</sup> are not recognizable in B<sup>act</sup>. However, this is likely due to differences in their adsorption to the carbon film which leads to different views of B<sup>028</sup> versus B<sup>act</sup>. Nonetheless, the overall morphology of B<sup>028</sup> appears to be between that of B and B<sup>act</sup>.

### **3.3.2 3D reconstruction of the B<sup>028</sup> particle**

Because the B<sup>028</sup> was very homogeneous in the 2D EM analysis, it appeared to be a good candidate for 3D EM structure analysis. To prepare B<sup>028</sup> complexes for 3D reconstruction, splicing was performed under standard conditions, and the splicing reaction mixture was applied to the first glycerol gradient. Then B<sup>028</sup>-containing fractions were loaded onto the MS2 column. The eluate was then subjected to a second glycerol gradient this time containing two cross-linking agents, namely, an increasing concentration (0 to 0.1%) of

glutaraldehyde and a decreasing concentration (0.2 to 0 %) of N-(p-Maleimidophenyl) isocyanate (PMPI). Negatively-stained EM grids were first made from the B<sup>028</sup> complex-containing gradient fractions. EM imaging and image processing was performed by Dr. Holger Stark. 11360 images of single particles were extracted and processed with the CowEyes program (Prof. Dr. H. Stark). The aligned and classified data set was partitioned into image classes and the class members averaged.

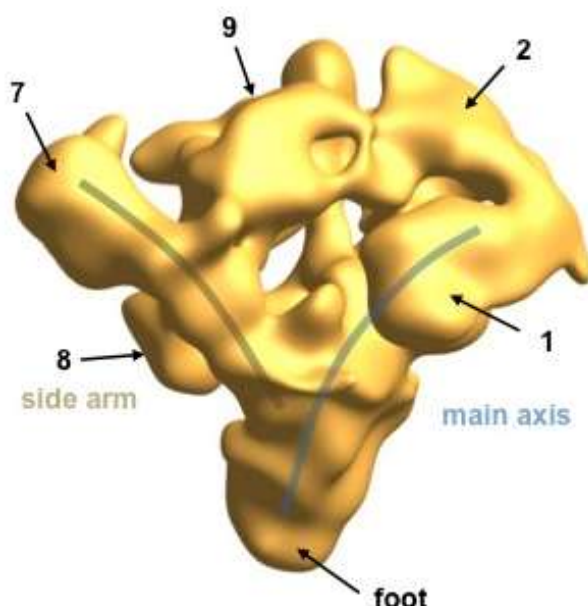
Raw images of B<sup>028</sup> particles revealed again a homogeneous population and the class averages demonstrated the characteristic two-dimensional averages of the B<sup>028</sup> complex. For the single-particle 3D reconstitution of the B<sup>028</sup> particle, the angular reconstitution method was performed (van Heel, 1987; van Heel et al., 2000). The orientation of each particle in three-dimensional space can be described by Eulerian angles, representing the rotation degree/event. Determination of the Eulerian angles for each particle differently oriented using the angular reconstitution method allows us to assign the projection directions of the class averages. The 3D structure reconstitution was finally calculated computationally from images of ~ 3000 particles and resulted in a 3D structure/architecture of the B<sup>028</sup> particle determined with a resolution of ca. 25 Å, shown in Fig. 3.29.



**Figure 3.29** The morphology of the B<sup>028</sup> complex assembled on the MINX pre-mRNA. (C) 3D structure (surface representation) of the B<sup>028</sup> complex reconstructed using 11360 negatively-stained EM images.

The B<sup>028</sup> 3D reconstruction has an approximately 20 nm thick isosceles triangle-like structure with a side length of 38 to 40 nm. Only the foot narrows down to a thickness of 8 nm, while a region in the upper part (characterized by the domains labelled 1 and 2 in the figure) is up to 28 nm thick. The prominent elongated axis already seen in the images turned out to be indeed a continuous, elongated density element running from the foot via the central region (3) to the thick upper corner (1, 2). This upper corner consists of two structural elements, one with a more globular structure containing a central cavity (1) and one with a long, flat structure (2), which are connected to the central region (3) by two bridges (4, 5). The two upper corner structures are connected to each other. The flat one (2) forms the beak-like density seen already in the 2D images.

3D reconstruction reveals a network of elongated structures in the globular domain. One slightly bent, that in total is nearly 30 nm long and has a 6 to 10 nm thick side arm (6), runs from a connecting point at the lower central position of the main axis directly to the outer corner of the globular domain. At this corner the side arm ends with a 6 x 8 x 10 nm sized globular structure (7). About half way down, the side arm is connected to another shorter arm (8), which originates also centrally from the elongated axis, but on the other side. Finally, at the top of the 3D reconstruction there is a complex density (9) that joins the flat structure of the upper region (2) of the long axis with both arms, the side arm (5) and the short arm (8).



**Figure 3.30** The mushroom-like view of the B<sup>028</sup> complex tilted forward on its horizontal axis. 3D structure of the B<sup>028</sup> complex reconstructed using negatively stain EM 11360 images, surface representation.

When the B<sup>028</sup> 3D structure is oriented to show a mushroom-like view, the long axis runs also along one lateral side but it is now tilted toward the plain of the paper (Fig. 3.30). The flat upper structure (2) is now seen well above the globular structure (1). In the mushroom-view oriented B<sup>028</sup> 3D reconstruction, holes going straight through the center of the particle are present. The intense accumulation of stain seen in class averages of B<sup>028</sup> mushroom-like images are easily explained by the presence of these holes.

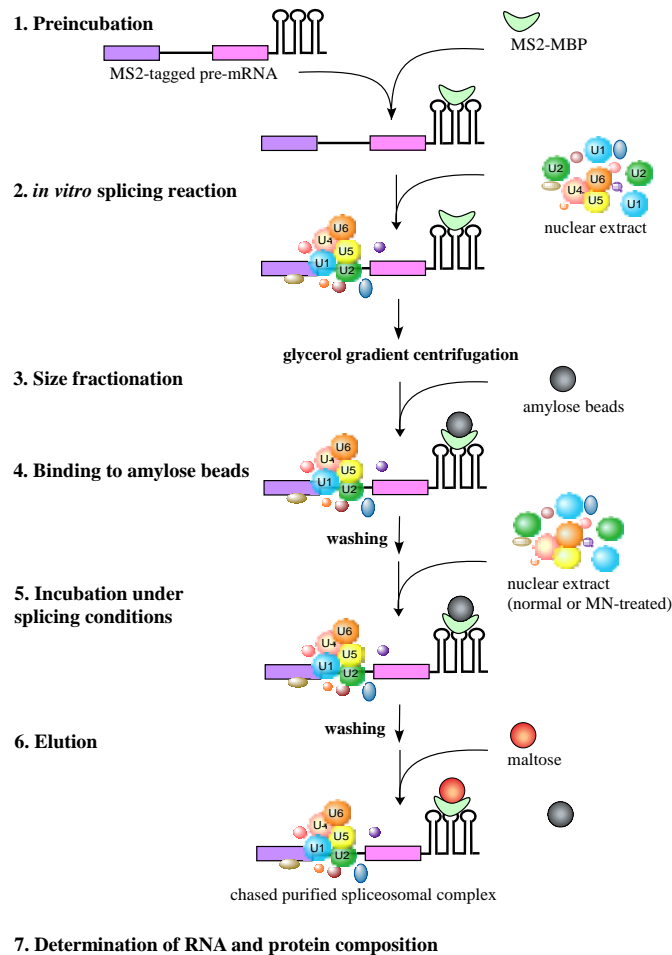
At present, the B<sup>028</sup> complexes are being analysed by cryo-EM (in progress). Cryo-EM enables images of hydrated biological particles at low temperatures to be obtained and therefore protects the particles against radiation damage. To avoid poor vitrification quality due to the presence of glycerol in the gradient fractions, the purified B<sup>028</sup> complexes were subjected to buffer exchange, so that the glycerol-containing environment was substituted with glycerol-free buffer. Then particles were transferred onto a copper EM grid, vitrified by liquid nitrogen and stored in liquid nitrogen until imaging.

### **3.4 B<sup>028</sup> spliceosomes as a tool to investigate binding and release of proteins during activation.**

Remarkable rearrangements occur in the spliceosome during its activation step concerning both its conformation and composition, including an exchange of many proteins. The B<sup>028</sup> complex is stalled apparently after U4 snRNA has been displaced by the helicase Brr2, but before release of B-specific (except FBP21) and Lsm proteins, and before B<sup>act</sup>-specific and Prp19/CDC5L-associated proteins have interacted. Because B<sup>028</sup> complexes can be chased into catalytically active complexes they could potentially be used to study the recruitment and release of these proteins during activation.

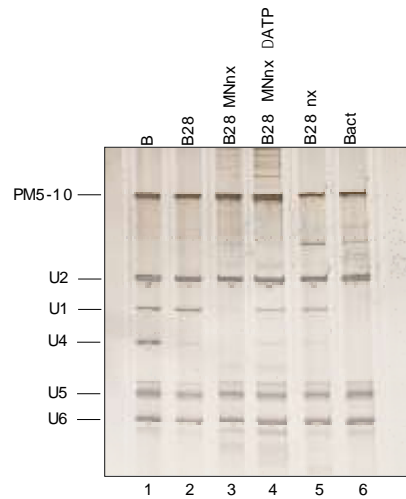
To this end, I performed splicing *in vitro* and separated the complexes, which were formed in the presence of compound 028, on a glycerol gradient. Complexes in the B/B<sup>act</sup> peak fractions were bound to amylose beads and washed (Fig. 3.31). Then B<sup>028</sup> complexes still bound to the matrix were incubated either with buffer alone or with normal or MN-treated nuclear extract under splicing conditions in the presence or in the absence of ATP. To ensure the absence of ATP in the reaction mixture, the HeLa nuclear extract was treated with hexokinase and glucose. Then the extensively washed complexes were eluted with maltose (Fig. 3.31). The RNA composition was analysed by denaturing PAGE and the protein composition via mass spectrometry.





**Figure 3.31 Schematic view of MS2-affinity purification of human spliceosomal B<sup>028</sup> complexes with subsequent chasing of isolated complexes with normal or MN-treated nuclear extract under splicing conditions.** MS2-tagged pre-mRNA substrate was pre-incubated with MS2-MBP fusion protein. Subsequently, splicing *in vitro* was performed in the presence of HeLa nuclear extract under standard splicing conditions. Spliceosomal complexes formed were then separated by size on a linear 10 to 30 % glycerol gradient and detected via  $^{32}\text{P}$ -labelled pre-mRNA. Spliceosomal complexes from each peak were affinity-selected using amylose beads. After washing, the isolated complexes were incubated under splicing conditions in the presence of normal or MN-treated nuclear and, after extensive washing, eluted with maltose. The recruitment or loss of proteins was determined by mass spectrometry.

The RNA analysis revealed that B<sup>028</sup> complexes (Fig. 3.32, lane 2) remain intact under these conditions. Interestingly, B<sup>028</sup> complexes lose their U1 snRNA when incubated under splicing conditions (+ATP) in the presence of MN-treated extract (3.32, lane 3), showing an snRNA composition identical to that of the activated spliceosome (3.32, lane 6). In the absence of ATP, U1 snRNA stays associated with the spliceosome (Fig. 3.32, lane 4). However, B<sup>028</sup> chased in the presence of nuclear extract not digested with MN also contains U1 snRNA (Fig. 3.32, lane 5), but can be chased into activated spliceosomes very efficiently, based on the protein composition analysis (see below). Thus, the presence/absence of low amounts of U1 is not a good indicator for activation.



**Figure 3.32 RNA composition of B<sup>028</sup> complexes after chasing with normal or MN-treated nuclear extract.** RNA composition of MS2-affinity purified B, B<sup>028</sup> and B<sup>act</sup> splicing complexes and B<sup>028</sup> complexes incubated with MN-treated or normal nuclear extract (nx). RNA recovered was analysed by denaturing PAGE and detected by silver staining. RNA identities are indicated at the left.

We first tried chasing B<sup>028</sup> complexes formed on the MINX substrate and could detect the recruitment of B<sup>act</sup>-specific and Prp19/CDC5L complex proteins (data not shown). However, as spliceosome assembly continues past the activation step with this substrate, B<sup>act</sup>-specific proteins are lost in the course of splicing, impeding the interpretation of whether or not they bind. Thus, we switched to the PM5-10 pre-mRNA substrate, to stall the spliceosome assembly prior to the catalytic activation step (i.e. B\* formation) at the stage of B<sup>act</sup> formation in order to study only the activation stage.

The number of peptides sequenced for each protein in the various complexes is shown in Table 3.4. By comparing these numbers with those seen with purified B<sup>act</sup> complexes, the amount of B<sup>028</sup> converted to B<sup>act</sup> can be estimated. There were little or no changes in the amounts of the Sm, U2, U5 or CBP20/80 proteins between B<sup>028</sup>, the chased complexes and B<sup>act</sup>, which are not expected to be lost during activation.

The human hPrp19/CDC5L complex proteins are stably recruited to the spliceosome during activation (Makarov et al. 2002; Makarova et al., 2004; Wahl et al. 2009), and remain associated throughout both catalytic steps, stabilizing the association of U5 and U6 with the spliceosome after U4 is released (Chan et al., 2003, 2005). Mass spectrometry analysis shows that B<sup>028</sup> contains only low amounts of Prp19/CDC5L complex proteins, such as Prp19, CDC5L, SPF27, PRL1, as well as Prp19/CDC5L-related proteins such as RBM22, Syf1, CRNKL1, Isy1, SKIP and KIAA0560 in comparison to both B and B<sup>act</sup> (Table 3.4). However, after chasing with MN-treated nuclear extract or untreated nuclear extract under splicing conditions, there is an increase

in the peptides sequenced for Prp19/CDC5L proteins and related proteins. Notably, in the absence of ATP these proteins appear to be less efficiently integrated into the spliceosome (i.e. fewer peptides were identified by mass spectrometry), demonstrating the requirement of ATP for the association of this group of proteins with the spliceosome during the activation step. Based on the number of peptides identified by MS, there are fewer Prp19/CDC5L and related proteins bound in the chased complexes compared to the  $B^{\text{act}}$  complex. This is an indication that not all  $B^{028}$  complexes can be chased to  $B^{\text{act}}$ . Interestingly, Npw38 and Npw38BP, which normally are released during the activation step, were found in high amounts in  $B^{028}$  and  $B^{028}$  complexes chased with MN-treated extract (+ATP), but lower amounts were found in the complex chased in the absence of ATP, or chased with normal nuclear extract.

$B^{\text{act}}$ -specific proteins, which integrate into the spliceosome during activation, are strongly underrepresented in B and  $B^{028}$  complexes. Remarkably, they are also recruited to the spliceosome after incubating with MN-treated extract or normal nuclear extract, but only in the presence of ATP. This suggests that their association is ATP-dependent or ATP is required at a stage prior to their recruitment.

B-specific proteins, which leave the spliceosome during activation, are present in  $B^{028}$  complexes as well as in B complexes, and they appear to be partially lost in chased complexes in an ATP-independent manner. Also, Lsm proteins, which should dissociate during activation, are partially lost when  $B^{028}$  complexes are incubated with normal or MN-treated extract. The 2<sup>nd</sup> step factor Prp22 and RES complex proteins, which are recruited during activation, are underrepresented in  $B^{028}$  complexes: upon incubation with MN-treated extract or normal extract they are recruited efficiently to the spliceosome, but only in the presence of ATP. Interestingly, some spliceosomal RNA helicases, including Prp5 and Prp43, bind to the spliceosome in large amounts in the absence of ATP, suggesting their dissociation is ATP-dependent. Generally, the chasing of the  $B^{028}$  complexes was more efficient in the presence of normal nuclear extract relative to that in the presence of MN-treated extract. However, with non-treated nuclear extract, it cannot be ruled out that  $B^{028}$  complexes dissociate and new complexes form during the chase.

Taken together, these results show that our  $B^{028}$  complexes can be chased into activated spliceosomes efficiently. Notably, the absence of ATP during chasing experiment prevents binding of an important set of proteins such as Prp19/CDC5L complex proteins,  $B^{\text{act}}$ -specific and RES complex proteins. The results demonstrate that ATP is required for progression of spliceosome assembly after U4 snRNA release to form an activated spliceosome.

**Table 3.4 Proteins identified by LC-MS/MS in human spliceosomal complexes B, B<sup>act</sup>, B<sup>028</sup> and in B<sup>028</sup> chased in the presence of MN-treated HeLa extract (MNnx) with or without ATP or normal extract plus ATP. The number shown is the total number of peptides sequenced for the indicated protein. Proteins are grouped according to function or association. Proteins widely considered as common contaminants, such as ribosomal proteins and non-reproducibly detected proteins are not shown.**

	GenBank accession no.	kDa	B	B28	B28 in MNnx	B28 in MNnx ΔATP	B28 in nx	Bact
<b>pmol</b>			0.5	0.5	0.5	0.5	0.5	0.5
<b>MS2-MBP</b>		89	170	406	482	374	586	347
<b>Sm proteins</b>								
<b>B</b>	gi 4507125	24.6	36	61	65	69	49	69
<b>D1</b>	gi 5902102	13.3	10	14	11	13	15	19
<b>D2</b>	gi 29294624	13.5	29	65	47	39	52	65
<b>D3</b>	gi 4759160	13.9	36	52	46	48	47	64
<b>E</b>	gi 4507129	10.8	19	31	23	27	27	45
<b>F</b>	gi 4507131	9.7	24	24	25	20	27	27
<b>G</b>	gi 4507133	8.5	7	9	13	12	8	8
<b>U1 snRNP</b>								
<b>U1-70K</b>	gi 29568103	51.6						
<b>U1-A</b>	gi 4759156	31.3	12	18	16	19	12	7
<b>U1 snRNP associated</b>								
<b>FBP11</b>	gi 151301228	105.8		16	24	27	12	8
<b>S164 (fSAP94)</b>	gi 219520438	100.1	0	18	39	37	17	10
<b>17S U2 snRNP</b>								
<b>U2A'</b>	gi 50593002	28.4	32	66	66	59	79	100
<b>U2B''</b>	gi 4507123	25.4	21	37	34	30	35	46
<b>SF3a120</b>	gi 5032087	88.9	125	205	216	187	150	223
<b>SF3a66</b>	gi 21361376	49.3	20	16	24	26	15	20
<b>SF3a60</b>	gi 5803167	58.5	46	93	83	85	74	92
<b>SF3b155</b>	gi 54112117	145.8	213	410	370	377	263	309
<b>SF3b145</b>	gi 55749531	100.2	140	248	196	173	133	195
<b>SF3b130</b>	gi 54112121	135.5	305	557	503	507	443	529
<b>SF3b49</b>	gi 5032069	44.4	12	17	29	31	19	26
<b>SF3b14a/ p14</b>	gi 7706326	14.6	28	43	43	48	48	53
<b>SF3b14b</b>	gi 14249398	12.4	5	16	14	13	16	17
<b>SF3b10</b>	gi 13775200	10.1	15	20	12	12	7	15
<b>17S U2 related</b>								
<b>hPRP43</b>	gi 68509926	90.9	90	109	27	104	32	32
<b>SPF45</b>	gi 14249678	45	25	14	8	13	3	2
<b>U2AF65</b>	gi 6005926	53.5	1	16	30	36	9	
<b>U2AF35</b>	gi 5803207	27.9		6	6	7	2	
<b>hPRP5/ DDX46</b>	gi 664806097	117.4		2	33	113	8	
<b>SR140 (fSAPa)</b>	gi 122937227	118.2	44	35	24	36	13	3
<b>CHERP</b>	gi 119226260	100	25	10	3	11	2	
<b>PUF60</b>	gi 74761960	59.9	1	12	51	97	9	
<b>A proteins</b>								
<b>THRAP3</b>	gi 167234419	108.5	5		27	9	26	7
<b>BUB3</b>	gi 56550081	37.2		13	21	27	13	4
<b>SF1</b>	gi 42544130	68.3		19	8	9	8	
<b>SF4 (F23858)</b>	gi 33469964	72.5	27	10	7	12	16	9
<b>U5 snRNP</b>								
<b>220K / Prp8</b>	gi 3661610	273.7	367	554	506	447	521	669
<b>200K / Brr2</b>	gi 45861372	244.5	379	604	592	472	620	764
<b>116K / Snu114</b>	gi 41152056	109.4	186	326	341	321	361	403
<b>40K</b>	gi 4758560	39.3	53	88	98	79	90	115
<b>102K / Prp6</b>	gi 40807485	106.9	118	33	39	19	47	23
<b>15K / Dib1</b>	gi 5729802	16.8	11	3	2	4	3	
<b>100K / Prp28</b>	gi 41327771	95.6	38	11	23	14	17	19
<b>52K / Lin1</b>	gi 5174409	37.6	7	2	2	3	2	3
<b>LSm proteins</b>								
<b>LSm2</b>	gi 10863977	10.8	11	21	15	13	9	6
<b>LSm3</b>	gi 7657315	11.8	1	4	6	3	1	1
<b>LSm4</b>	gi 6912486	15.4	3	16	10	11	11	3
<b>LSm5</b>	gi 10720081	9.9		1	1	2	1	1
<b>LSm6</b>	gi 5919153	9.1	4	9	11	6	4	2
<b>LSm7</b>	gi 7706423	11.6	1	3	3	2	0	
<b>LSm8</b>	gi 7706425	10.4	9	18	12	11	10	6
<b>U4/ U6 di-snRNP</b>								
<b>90K / Prp3</b>	gi 4758556	77.6	75	19	13	18	17	4
<b>60K / Prp4</b>	gi 45861374	58.4	64	19	13	17	16	5
<b>20K</b>	gi 5454154	20	20	5	8	5	10	5
<b>61K / Prp31</b>	gi 40254869	55.4	34	11	12	10	9	1
<b>15.5K / Snu13</b>	gi 4826860	14.2	6	2	2	3	3	1

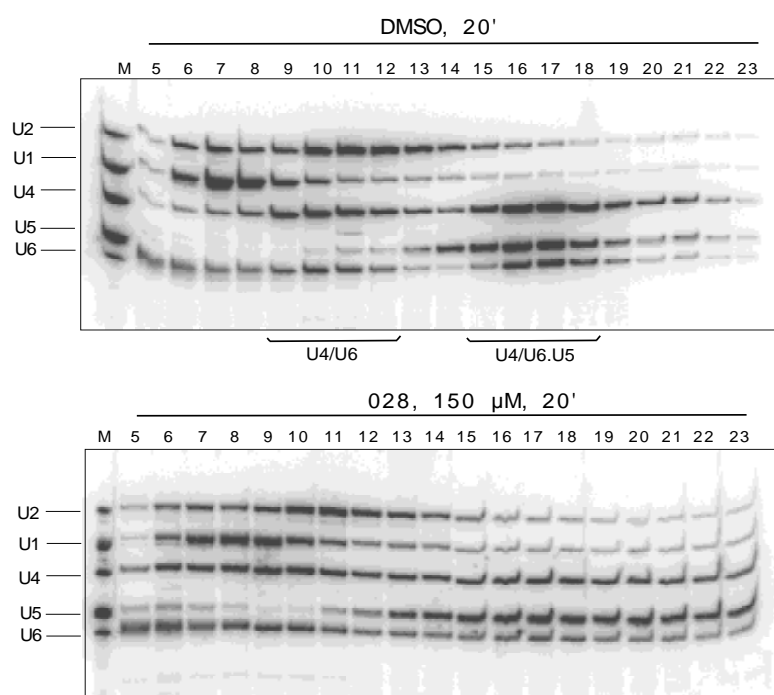
	GenBank accession no.	kDa	B	B28	B28 in MNnx	B28 in MNnx ΔATP	B28 in nx	Bact
<b>pmol</b>			0.5	0.5	0.5	0.5	0.5	0.5
<b>U4/U6.U5 tri-snRNP</b>								
110K / Snu66	gj 13926068	90.2	109	74	36	25	30	14
65K / Sad1	gj 56550051	65.4	5	2	9	8	11	9
27K	gj 24307918	18.9				1	1	
<b>B proteins</b>								
hSnu23	gj 52783532	23.6	6	19	17	12	11	8
hPRP38	gj 24762236	37.5	22	58	41	23	17	20
hSmu-1 (fSAP57)	gj 8922679	57.5	99	304	205	189	99	76
MFAP1	gj 50726968	51.9	42	64	48	39	25	41
RED	gj 10835234	65.6	46	111	107	91	44	44
FBP21	gj 67461858	42.5	7	2		1	4	
<b>hPRP19/ CDC5L</b>								
hPRP19	gj 7657381	55.2	62	19	114	40	204	309
CDC5L	gj 11067747	92.2	77	32	183	68	200	291
SPF27	gj 5031653	21.5	18	8	28	14	44	54
PRL1	gj 4505895	57.2	41	17	90	26	109	156
CCAP1 (hsp73)	gj 5729877	70.4	4	9	8	34	54	36
CCAP2 (hspc148, AD-002)	gj 7705475	26.6	9		15	4	20	19
catenin, b-like 1 (CTNNB1)	gj 18644734	65.1	13		29	9	34	32
Npw 38BP	gj 7706501	70	27	102	141	84	43	31
Npw 38	gj 5031957	30.5	3	21	24	12	12	5
<b>hPRP19/ CDC5L related</b>								
PRCC	gj 40807447	52.4	6		13		10	9
RBM22 (fSAP47)	gj 8922328	46.9	14	2	37	17	46	59
hSYF1 (XAB2)	gj 55770906	100	43	17	131	49	220	301
CRNKL1/ hSYF3	gj 30795220	100.6	44	9	56	22	134	279
hI sy1 (fSAP133)	gj 20149304	33	12	15	97	76	22	18
SKI P	gj 6912676	51.1	45	13	115	51	80	117
Cyp-E	gj 5174637	33.4	9	3	21	6	47	84
PPIase-like 1 (PPIL1)	gj 7706339	18.2	10	3	18	9	23	32
KI AA0560 (fSAP164)	gj 38788372	171.3	77	17	174	74	306	455
G10 (fSAP17)	gj 32171175	17	7	10	19	15	23	22
<b>A/ B proteins</b>								
TCERG1 (CA150)	gj 21327715	123.9	3	23	150	108	76	12
RNPC2 (CC1.3)	gj 28201880	58.5		12	32	31	9	10
p72/ DDX17	gj 3122595	80.5	4	13	67	65	33	19
NFAR	gj 62512150	95.4		6	15	2	9	7
NF45	gj 24234747	43			10	2	1	17
LOC124245	gj 24234748	104	10	22	66	57	59	10
YB-1	gj 24234749	35.9	31	37	78	78	37	57
p68 (DDX5)	gj 24234751	69.2	7	27	108	87	27	24
<b>Recruited prior to Bact</b>								
hsp27	gj 4504517	22.8		2	6	12	15	4
UBL5	gj 13236510	8.5	6	12	12	10	7	4
hPRP4-Kinase	gj 89276756	117.1	4	4	40	8	33	2
HsKin17	gj 13124883	45.4	12	2	14		6	11
BCLAF1	gj 7661958	106	2		8	4	11	5
SKI V2L2 (KI AA0052)	gj 193211480	117.8	5	12	13	35	20	12
<b>Abundant first in Bact</b>								
KI AA1604 (fSAPb)	gj 55749769	105.5	15	1	36	6	106	161
hPRP17	gj 7706657	65.5	23	1	32	8	105	181
hPRP2	gj 4503293	119.2	17	4	56	8	104	175
GPLOW (T54, GPATC5)	gj 15811782	52.1	12	1	12	1	57	87
NY-CO-10	gj 64276486	53.8	9		12	4	34	72
RNF113A	gj 5902158	38.8	4		9	1	13	41
MGC20398	gj 49472814	42	9		21	7	23	41
PPIL2/ Cyp-60	gj 7657473	59.5	17		24	6	69	121
MGC23918	gj 21389497	19.2	7		12	3	16	37
FRG1	gj 4758404	29.2			3		6	11
<b>Detected in Bact</b>								
p30 DBC	gj 9937501	102.8				2	3	6
FAM58A	gj 156630447	26			1		2	3
NCOR1	gj 189441644	270.1						2
TRI M24 (TIF1,LOC51591)	gj 47419911	116.7						
LENG1	gj 24308289	30.4					6	10
DDX50	gj 13129006	82.4		5	2			
CRI PT	gj 7661798	11.1					4	3
GTL3 (fSAP23)	gj 8392875	22.6	3	2	3	11	2	
MOV10	gj 14211540	113.5		0	9	4	4	14
FUBP3	gj 100816392	61.5	6	9	31	14	17	6
<b>2nd step factors</b>								
hPRP22	gj 4826690	139.3	7	2	21	6	91	71
hSLU7	gj 27477111	68.4			5		29	21

	GenBank accession no.	kDa	B	B28	B28 in MNnx	B28 in MNnx ΔATP	B28 in nx	Bact
<b>pmol</b>			0.5	0.5	0.5	0.5	0.5	0.5
<b>abundant first in C complex</b>								
Abstrakt	gi 21071032	69.8			8	9	35	22
GClP p29 (fSAP29)	gi 46371998	28.7	1		8	2	22	26
DDX35	gi 20544129	78.9			9	5	15	30
<b>detected in C complex</b>								
matrin 3	gi 21626466	94.6		9	41	34	14	13
<b>EJC/ mRNP</b>								
el F4A3	gi 7661920	46.9	5	6	33	25	27	40
Magoh	gi 4505087	17.2			4	4	14	15
Pinin	gi 33356174	81.6		1	29	21	10	15
RNPS1	gi 6857826	34.2		1	8	7	3	3
DDX3	gi 87196351	73.3	5	12	76	50	20	9
<b>mRNA binding proteins</b>								
PABP1	gi 46367787	70.5						
PABPC4	gi 4504715	70.6			32	26	9	9
PABPN1	gi 4758876	32.6		1			5	1
<b>cap binding complex</b>								
CBP20	gi 110349727	18	6	10	12	14	17	21
CBP80	gi 4505343	91.8	101	199	167	158	176	221
TREX								
THOC1 (HPR1)	gi 154448890	75.6			15	16	5	9
THOC2	gi 125656165	169.6			26	10	6	14
THOC3	gi 14150171	38.8			4	3		4
THOC6 (WDR58, MGC2655)	gi 31543164	37.5			11	11	5	8
THOC7 (FLJ23445, fSAP24)	gi 13376623	23.7			6	6	3	5
<b>RES complex</b>								
SNI P1	gi 21314720	45.8	17	1	23	1	37	58
MGC13125 (fSAP71)	gi 14249338	70.5						
CGI-79	gi 4929627	39.7	10	2	10	2	27	49
<b>SR proteins</b>								
SF2/ ASF	gi 5902076	27.8	14	9	17	76	26	35
9G8	gi 72534660	27.4	8	6	22	42	20	23
SRp20	gi 4506901	19.4				37	12	
SRp30c	gi 4506903	25.5	7	9	42	46	14	32
SRp38	gi 5730079	31.3	9	13	45	32	29	41
SRp46	gi 14141216	31	1	1	1	2	2	3
SRp55	gi 20127499	39.6			2	24	10	11
SRp75	gi 21361282	56.8				9		
hTra-2 alpha	gi 9558733	32.7	3	1	2	8	4	7
hTra-2 beta/ SFRS10	gi 4759098	33.7	2	4	16	22	16	15
<b>SR related proteins</b>								
SRm160	gi 42542379	102.5	4	4	15	19	26	37
SRm300	gi 4759098	300	40	22	46	43	77	148
<b>hnRNP</b>								
hnRNP A0	gi 8134660	30.9			4		5	1
hnRNP A1	gi 4504445	38.7		15	110	76	28	6
hnRNP A3	gi 34740329	39.6		9	45	27	10	5
hnRNP A2/ B1	gi 14043072	37.4		8	98	58	31	3
hnRNP D	gi 14110420	38.4		3	11	8	7	2
hnRNP F	gi 148470406	45.7	2	27	41	40	11	12
hnRNP G	gi 56699409	42.4	12	15	97	76	22	18
hnRNP H1	gi 5031753	49.1	4	24	93	70	21	14
hnRNP H3	gi 23503095	36.9		11	7	5	2	
hnRNP K	gi 14165435	51		19	172	145	30	6
hnRNP L	gi 215274006	64.1		12	21	7	10	2
hnRNP M	gi 14141152	77.5	7	13	62	68	21	22
hnRNP Q	gi 15809590	69.6		4	39	15	16	5
hnRNP R	gi 5031755	70.9		8	110	47	26	13
hnRNP U	gi 14141161	90.6		10	51	36	8	0
hnRNP U-L2	gi 118601081	85		13	50	14	7	5
<b>Miscellaneous proteins</b>								
LOC51325	gi 22035565	104.8	18	42	12	20	17	13
BAT1	gi 114306812	51	1	6	14	9	18	11
HSPA5	gi 14916999	72	1	7	11	28	42	22
FNBP4	gi 313104235			3	24	12	12	2
WTAP	gi 47117889	43		1	10	10	1	3
ZNF 318	gi 166215018	24	6	19	17	12	11	8
CCDC130	gi 134037158						3	14
FUBP2	gi 313104306	77		59	62	49	26	

## 3.5 Mechanism of inhibitory action of compound 028

### 3.5.1 U4/U6.U5 tri-snRNP is stable in the presence of compound 028

We next were interested in elucidating the mechanism how compound 028 blocks spliceosome assembly. Both A and B-like complexes accumulate when splicing is performed in the presence of compound 028. We thus checked if tri-snRNP formation/stability is affected by the presence of compound 028. To answer this question, HeLa nuclear extract was incubated under splicing conditions (except no pre-mRNA was added) in the presence of DMSO or 028, and the splicing mix was subjected to glycerol gradient centrifugation (Fig. 3.33). RNAs were isolated from gradient fractions, separated by denaturing PAGE, transferred to a nylon membrane and hybridized with <sup>32</sup>P-labelled DNA probes specific for U1, U2, U4, U5 and U6 snRNAs.

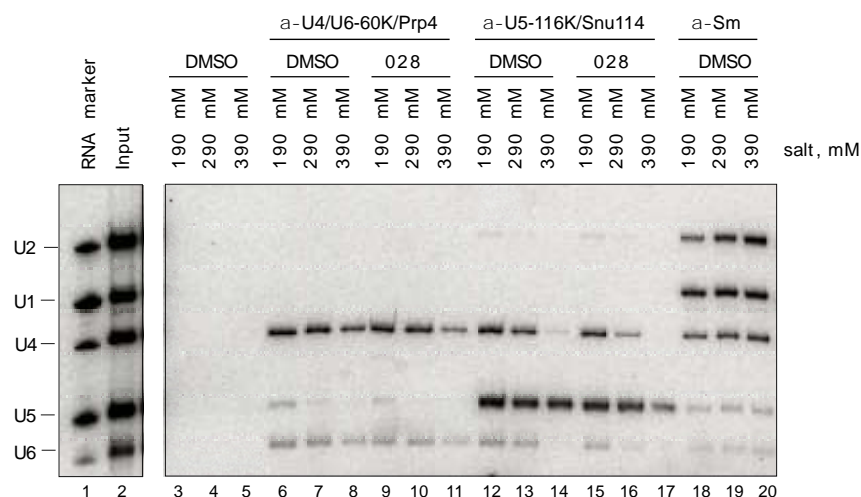


**Figure 3.33 The U4/U6.U5 tri-snRNP is stable in the presence of compound 028.** Control reaction in the absence of compound 028 (DMSO) was performed in parallel. HeLa nuclear extract plus DMSO (upper panel) or compound 028 (lower panel) was incubated under splicing conditions at 30° for 20 min and subjected subsequently to 10-30 % glycerol gradient centrifugation. RNAs were isolated from gradient fractions, separated by denaturing PAGE, and analysed by northern blotting with radioactively labelled DNA probes against the U2, U1, U4, U5 and U6 snRNAs. Hybridized DNA probes were detected by autoradiography. RNA identities are indicated at the left.

Northern blot analysis revealed that tri-snRNPs formed in the presence of DMSO peak in fractions 15 to 18, based on the presence of U4, U5 and U6 snRNAs. In the presence of

028, tri-snRNPs sedimented as a broader peak in fractions 14 to 23, suggesting compound 028 leads to some aggregation of the tri-snRNP such that it sediments closer to the bottom of the gradient. However, there was little increase in free U4/U6 snRNPs, which peak in fractions 9-12 in the DMSO control. Thus, the tri-snRNP remains to the most part intact. The sedimentation of U1 and U2 snRNPs was also largely unaffected by compound 028.

To check if the stability of the U4/U6.U5 tri-snRNPs is affected by 028, we performed immunoprecipitations with splicing reactions and antibodies directed against the U4/U6 60K/Prp4 protein and looked for coprecipitation of U5 (indicating tri-snRNP formation) at different salt concentrations. Likewise we used antibodies against the U5 116K/Snu114 protein to check for coprecipitation of U4/U6. In the presence of 028 we observed a slightly decreased signal intensity of U5 coprecipitated with anti-U4/U6-60K antibodies, especially at 290 and 390 mM salt, in comparison to the control (Fig. 3.34, lanes 6-8 versus 9-11). Likewise, less U4/U6 snRNP was coprecipitated with anti-U5-116K antibodies when compound 028 was added (Fig. 3.34, lanes 12-14 versus 15-17). In general, we conclude that compound 028 affects tri-snRNP stability to only a low extent, and that the majority of tri-snRNP is still intact under splicing conditions.

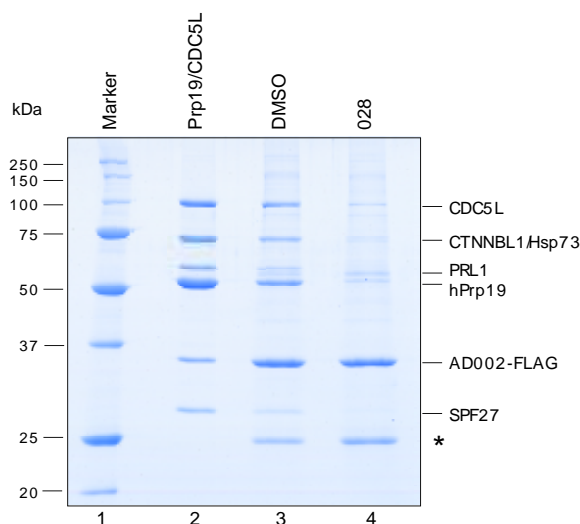


**Figure 3.34 U4/U6.U5 tri-snRNPs appear to be slightly less salt stable in the presence of compound 028.** Tri-snRNP particles were immunoprecipitated with antibodies directed against U4/U6 snRNP protein 60K/Prp4 or the U5 snRNP protein 116K/Snu114 or Sm proteins (Y12) as a control, at three different salt concentrations. Initially, HeLa nuclear extract was incubated under splicing conditions at 30° for 20 min in the presence or absence of compound 028, and then added to the PAS-bound antibodies, incubated for 4 h with end-over-end rotation and extensively washed in buffer containing 190, 290, 390 mM salt. Co-precipitated RNAs were isolated, separated by denaturing PAGE and analysed by northern blotting. Hybridized <sup>32</sup>P-labelled DNA probes were detected by autoradiography. RNA identities are indicated at the left.



### 3.5.2 Association of AD002 with the Prp19/CDC5L complex is affected by the presence of compound 028

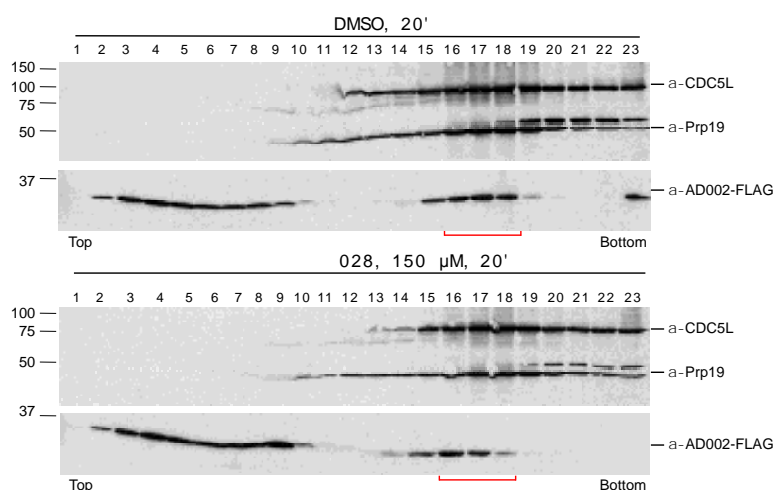
The hPrp19/CDC5L complex is required for spliceosome activation and for stabilization of interactions of U5 and U6 with the pre-mRNA in B<sup>act</sup> (Chan et al., 2003, 2005). Based on mass spectrometry analysis, it was shown that the hPrp19/CDC5L complex is clearly underrepresented in our stalled B<sup>028</sup> complexes (Table 3.3). Thus, the block during the activation step could be due to an effect of compound 028 on the Prp19/CDC5L complex. To test this, we next isolated hPrp19/CDC5L complexes from splicing reactions in the presence/absence of compound 028. For this purpose, splicing reactions containing HeLa nuclear extract expressing FLAG-tagged AD002/Cwc15 protein were performed under standard conditions in the presence of DMSO or compound 028. Then the salt concentration was adjusted to 250 mM. Complexes containing AD002-FLAG were bound to anti-FLAG-M2 agarose beads and, after extensive washing, eluted with FLAG peptide. SDS-PAGE analysis revealed that intact Prp19/CDC5L complexes containing CDC5L, CTNNBL1, HSP73, PRL1, Prp19 and AD002, were successfully isolated in the absence of compound 028 (Fig. 3.35). In the presence of compound 028, an equal amount of AD002 was precipitated, but only low levels of CDC5L, CTNNBL1, Hsp73, PRL1 and Prp19, compared to the control reaction, indicating that mainly free AD002-FLAG had been isolated. Taken together, these results suggest that the association of AD002 with the Prp19/CDC5L complex is affected by the presence of compound 028.



**Figure 3.35 The Prp19/CDC5L complex is affected by the presence of compound 028.** Splicing reactions in the absence of pre-mRNA were performed in the presence or absence of compound 028 using 40 % HeLa nuclear extract expressing FLAG-tagged AD002 protein. FLAG-tagged protein with interacting partners was bound to anti-FLAG-M2 agarose in buffer containing 250 mM salt, washed and eluted with 0.2 mg/ml FLAG peptide. Eluted complexes were precipitated with TCA, separated by SDS-PAGE and stained with coomassie. The Prp19/CDC5L complex proteins were identified by mass spectrometry and are indicated on the right. \* = IgG light chain.

### 3.5.3 Gradient analysis of Prp19/CDC5L complex in the presence/absence of compound 028

To check the stability of the Prp19/CDC5L complex and the association of the AD002/Cwc15 protein with Prp19/CDC5L complexes in the presence of compound 028, we performed splicing under standard conditions (except no pre-mRNA was added) with HeLa nuclear extract expressing a FLAG-tagged version of AD002 protein in the presence or absence of compound 028. The splicing reaction mixture was subjected to glycerol gradient centrifugation (150 mM salt) and gradients were fractionated from the top. The proteins from each fraction were separated by SDS-PAGE, and selected proteins were then detected by western blotting with antibodies directed against CDC5L, Prp19 and the FLAG tag. Western blot analysis revealed similar amounts of AD002-FLAG protein migrating alone (not in complex with Prp19 or CDC5L) in fractions 3 to 9 in both the DMSO control and when compound 028 was present (Fig. 3.36). Prp19 and CDC5L proteins sedimented together in fractions 16 to 18 in both preparations, demonstrating that the Prp19/CDC5L complexes do not dissociate once we add compound 028. In contrast, FLAG-tagged AD002 protein co-migrated in the control reaction, but showed a slight shift if incubated in the presence of compound 028, suggesting that the adjusting the salt concentration from 150 to 250 mM could lead to the dissociation of AD002 protein component from Prp19/CDC5L complex. Thus, additional experiments are needed to clarify the effect of compound 028 on the Prp19/CDC5L complex.



**Figure 3.36** The sedimentation of the Prp19/CDC5L complex is not affected by compound 028. HeLa nuclear extract expressing FLAG-tagged AD002 protein was incubated under splicing conditions but no exogenous pre-mRNA was added. Complexes were separated using glycerol gradient centrifugation and subsequently fractionated from the top. The proteins from each fraction were separated on SDS-PAGE and subjected to western blotting procedure. Western blotting was performed using antibodies directed against selected Prp19/CDC5L complex proteins and against the FLAG tag. Molecular weight is indicated at the left, and the antibodies used are indicated at the right.

### 3.6 Structure-activity relationship (SAR) analysis of compound 028

Our results showed that compound 028 completely inhibits pre-mRNA splicing *in vitro* at a concentration of 150  $\mu\text{M}$ , with an  $\text{IC}_{50}$  value of 51  $\mu\text{M}$ , leading to accumulation of A complexes as well as complexes stalled after the formation of B complex, but before  $\text{B}^{\text{act}}$  complexes emerge, i.e. during the activation step. However, the exact relationship between 028 compound structure and splicing inhibition is not clear. A structure–activity relationship analysis (SAR) of this compound can reveal molecule features essential for splicing inhibition. On the basis of a SAR study we hoped to enhance the inhibition property of compound 028 by modifying its original structure. Furthermore, we intended to identify parts of the molecule that could potentially be used for tagging without significant loss of its inhibition activity. A tagged version of 028 is of particular importance for further investigation of possible target(s) of compound 028 in the splicing reaction.

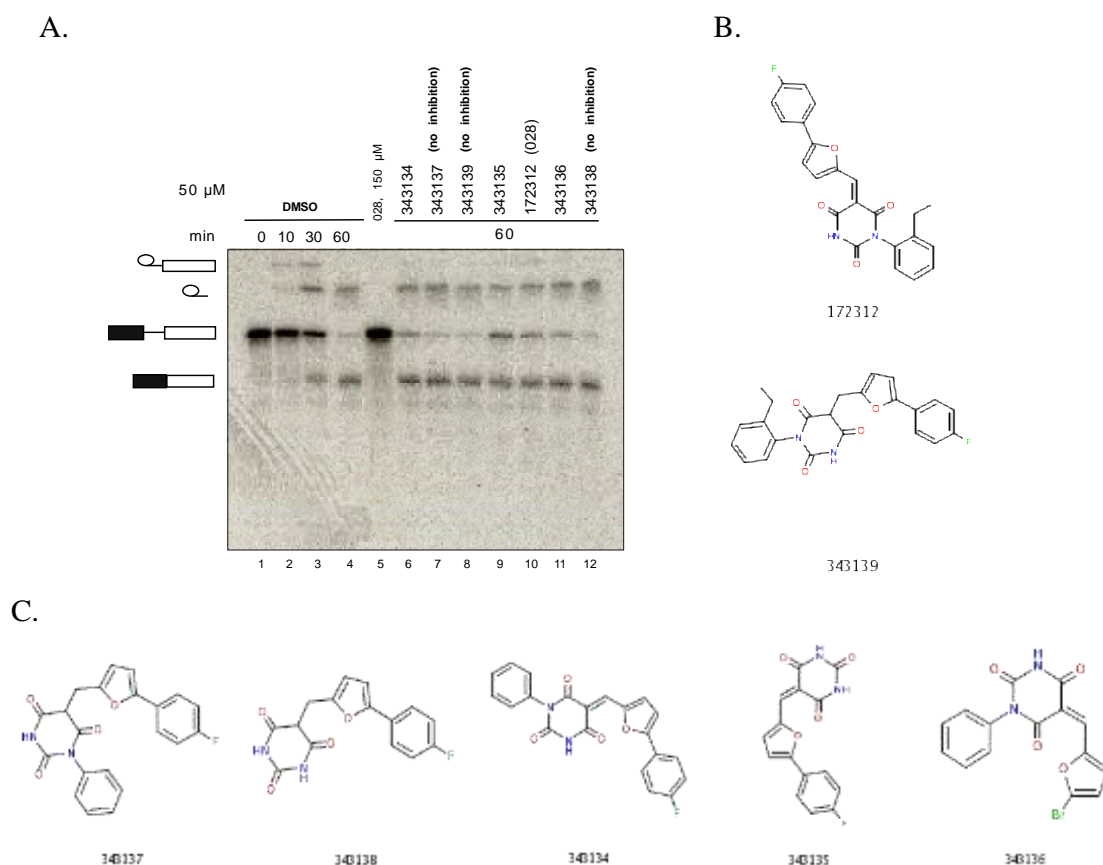
To elucidate structural determinants (motifs) that confer splicing inhibition activity on compound 028, a number of structural modifications at specific sites of the original molecule were carried out by Javier Ceballos in Prof. Dr. Waldmann's group (MPI for Molecular Physiology, Dortmund). The synthesis of 028 analogues was performed with a wide diversity of modifications introduced separately or combined at different positions. These modifications include reduction of the linker between the two heterocyclic motifs of the original molecule, modifications in the *o*-ethylphenyl group and/or in *p*-fluorophenyl group (with the goal of changing electron properties on the rings and/or changing the size of the substitution), modifications in the furan group located between the *p*-fluorophenyl group and heterocyclic barbituric acid-related group, and modifications in the heterocyclic barbituric acid-related group based on the pyrimidine heterocyclic skeleton. Derivatives were synthesized and tested in an *in vitro* splicing assay. All together, I tested 31 derivatives, whose structures are shown in Fig. 3.37B and C and in Table 3.4.

With the first set of analogues, we examined the importance of the double bond in the short unsaturated linker between the two heterocyclic motifs. The original compound is expected to behave reactively towards nucleophiles due to its Michael acceptor character. Therefore, this feature might be relevant for the splicing inhibition activity of the original compound. To check this, we tested the reduced derivative of compound 028 alone (343139) or in combination with other modifications like removing the ethyl group from the *o*-ethylphenyl group (343137) or removing the entire *o*-ethylphenyl group (343138).

For comparison, splicing was assayed with a version of compound 028 that was also synthesized by the Waldmann group whose structure is identical to the commercially available compound 028, called here 172312.

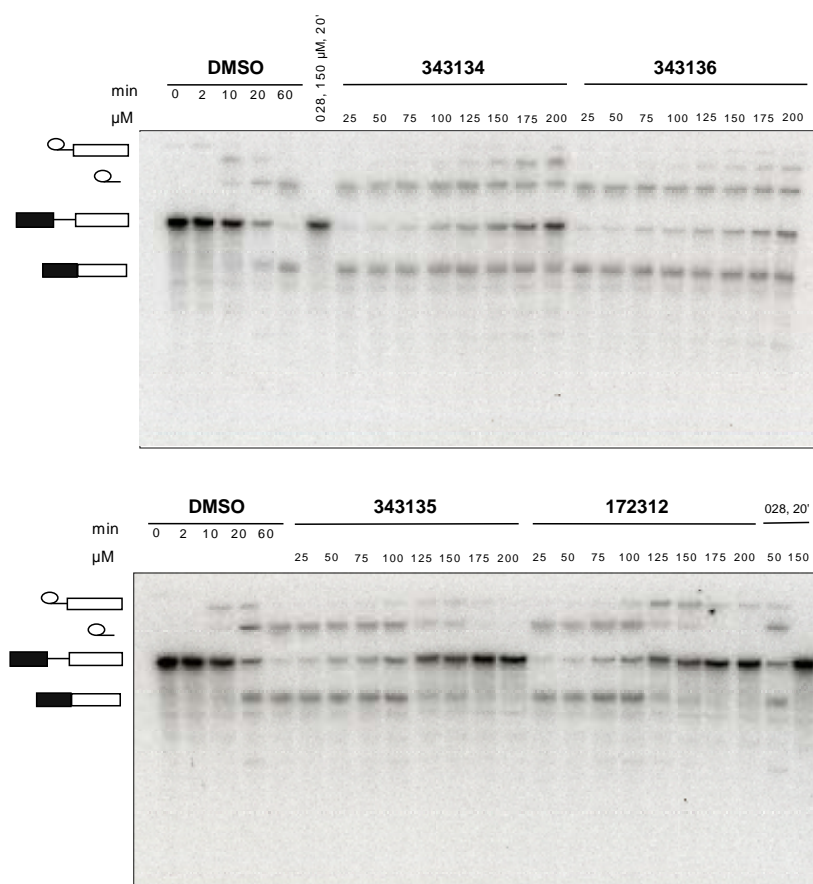
### 3.6.1 Effect of the reduction of the linker between the two heterocyclic motifs on splicing inhibition *in vitro*

To assay for enhanced inhibition we chose the concentration 50  $\mu\text{M}$ , where only slight inhibition of splicing is observed by the synthesized compound 172312 (Fig. 3.37, lane 10). At 50  $\mu\text{M}$  all of the modified analogues (343139, 343137, 343138) (Fig. 3.37, lanes 8, 7, 12, Table 3.6) were as active as the DMSO control and less potent than 172312, displaying a slight increase in inhibition activity relative to 172312. At the same time, removing the ethyl group alone from the non-reduced molecule (343134) did not change



**Figure 3.37** Analysis of *in vitro* splicing performed in the presence of various derivatives of compound 028 at a concentration of 50  $\mu\text{M}$ . Splicing was performed with radioactively-labelled adenovirus-derived, MINX-MS2 pre-mRNA as a substrate. A. Effects of different analogues on the splicing of  $^{32}\text{P}$ -labelled MINX-MS2 pre-mRNA in HeLa nuclear extract after 60 min. RNA was analysed by denaturing PAGE and visualized by autoradiography (DMSO: control reaction with solvent). The positions of the lariat-intron/3'-exon, pre-mRNA, spliced mRNA, (top to bottom) are indicated on the left. B. Chemical structure of the original molecule synthesized for this study. C. Chemical structures of the synthesized analogues of compound 028.

the inhibition activity compared to 172312 (Fig. 3.37, lane 6, Table 3.6). However, we could observe slightly enhanced inhibition activity (accumulation of pre-mRNA), when the entire o-ethylphenyl group was removed without reduction of the double bond of the linker (343135) (Fig. 3.37, lane 9) when compared to the synthesized original compound 172312. In addition, we tested a molecule with the p-fluorophenyl group substituted by bromine in combination with the removal of the ethyl residue from the o-ethylphenyl group (343136) (Fig. 3.37). The bromine was added as a substitute for reactive fluorine. This compound was less effective in inhibiting splicing *in vitro* in comparison to the original hit compound (Fig. 3.37 lane 11, 3.38). Taken together, our initial results suggest that the Michael acceptor character of the original compound is necessary for splicing inhibition. However, to be able to make a more precise conclusion, we decided to test the activity of the more promising analogues at higher concentrations.



**Figure 3.38 Effect of increasing concentrations of selected analogues of compound 028 on the splicing of  $^{32}\text{P}$ -labeled MINX-MS2 pre-mRNA.** Splicing was performed in the presence of various concentrations (25-200  $\mu\text{M}$ ) of three analogues in HeLa nuclear extract for 1 h. Formation of the splicing products and intermediates was analysed by denaturing PAGE and visualized by autoradiography. The positions of lariat-intron/3'-exon, lariat-intron, pre-mRNA, spliced mRNA are indicated on the left. DMSO, control reaction (solvent).

To analyse the splicing inhibition efficiency of 343134, 343135 and 343136 in more detail, I carried out a titration of these compounds (Fig. 3.38). From this experiment it is clear that compounds 343134 and 343136 are less potent than 172312. Thus, deletion of the ethyl group (343134) or substitution of the p-fluorophenyl group with bromine in combination with the elimination of ethyl group from o-ethylphenyl (343136) reduced somewhat splicing inhibition activity (Fig. 3.38, upper panel). In contrast, removal/elimination of the entire o-ethylphenyl group did not significantly alter the potency of 343135 compared to 172312 (Fig. 3.38, lower panel), suggesting it is not required for inhibition activity. Interestingly, the lariat-intron 3'exon intermediate is more abundant at a concentration of 125-150  $\mu$ M of 172312 compared to 343135. This suggests that the mechanism of inhibition might differ somewhat if the o-ethylphenyl group is removed.

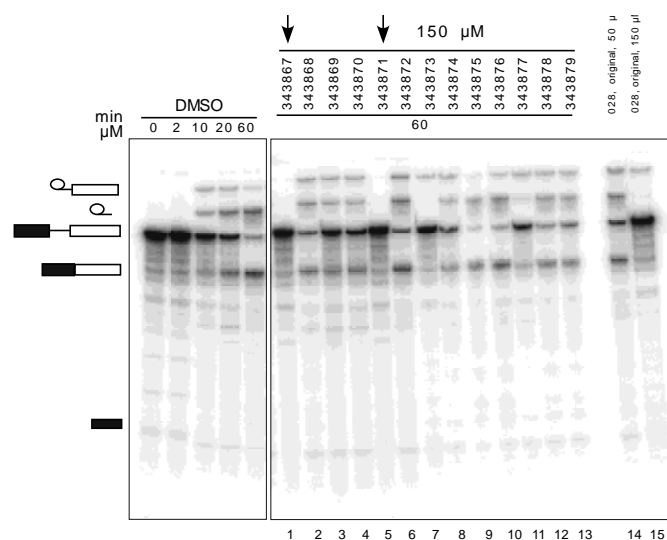
### **3.6.2 Effect of further modifications on splicing inhibition *in vitro***

We next analysed the effect of a wide variety of modifications on the splicing inhibition activity of compound 028 (172312). Changing the position of fluorine from para- to ortho-position abolished inhibition activity completely (343876) as well as replacing the p-fluorophenyl group with heterocycle piperidine (non aromatic ring) (344353). Two neighbouring halogens in the meta- and para-positions (343872) reduced the inhibitory activity dramatically as well as the replacement of fluorine on the p-fluorophenyl group by a methyl group (electron donator) (343875) or bulky phenyl group (343868).

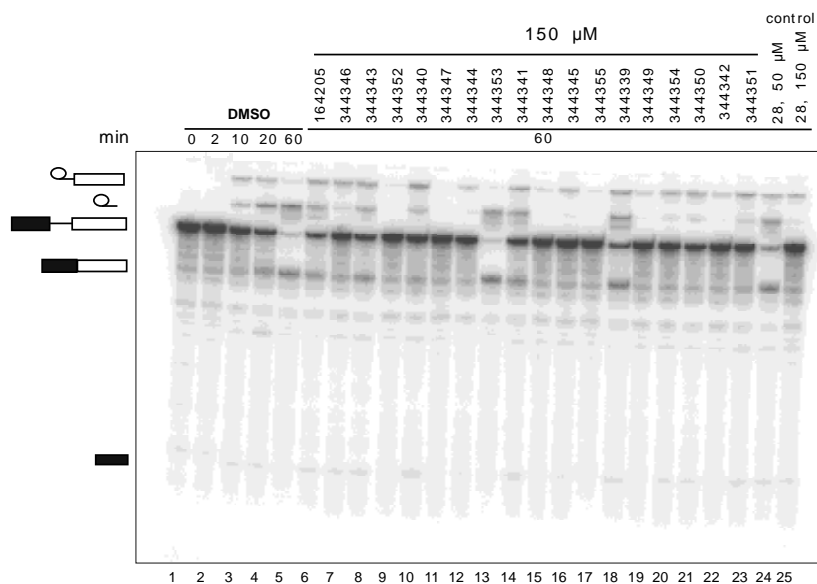
Eliminating the furan group from the heterocyclic barbituric acid-related group, which leads to the phenyl group being directly connected via unsaturated short linker with a halogen (chlorine or bromine) substituting the fluorine in the para-position (343878, 343879) also clearly weakens inhibition activity in comparison to the original hit. Thus, our study revealed that the p-fluorophenyl group together with reactive fluorine, as well as the furan group are required for the inhibition activity/function.

Minor modifications in the p-fluorophenyl group such as changing the position of fluorine from the para- to meta-position (343867) or substituting fluorine with chlorine in the meta-position (343871) clearly enhanced the inhibition effect of the original compound, suggesting once again that the p-fluorophenyl group is extremely important for the function of the compound.

A.

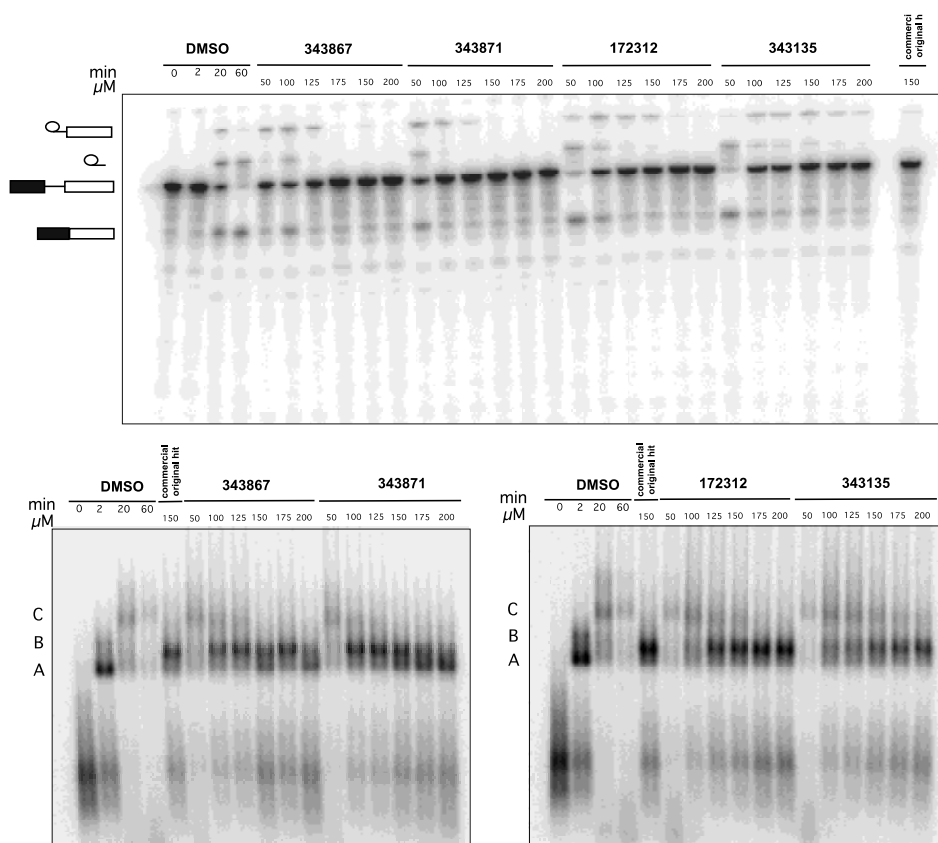


B.



**Figure 3.39 Effects of various analogues of compound 028 on the splicing of <sup>32</sup>P-labeled MINX-MS2 pre-mRNA in HeLa nuclear extract after 60 min.** RNA was analysed by denaturing PAGE and visualized by autoradiography (DMSO: control reaction with solvent). Positions of lariatintron/3'-exon, lariatintron, pre-mRNA, spliced mRNA, and 5'-exon (top to bottom) are indicated on the left.

The other modifications in the p-fluorophenyl group, such as two halogens in the meta-positions (343873), dimethylamino group (electron donating group) attached to the phenyl group (343869) or an acetyl group (electron withdrawing group) in para-position (343870) didn't change the potency of the molecule significantly. Substituting fluorine with a trifluoromethyl ether group (344355) appears to show rather strong inhibition. Introducing a methoxy group in the ortho-position leads to reduced splicing inhibition (343874), whereas two methoxy groups in the para- and meta-position (343877) restore the initial potency of the molecule.



**Figure 3.40** Effect of increasing concentrations of selected synthesized analogues of compound 028 on the splicing of  $^{32}\text{P}$ -labeled MINX-MS2 pre-mRNA. Splicing was performed in the presence of various concentrations of four analogues in HeLa nuclear extract for 1 h. Formation of the splicing products and intermediates was analysed by denaturing PAGE (upper panel), and spliceosomal complexes were analysed on an agarose gel (lower panel). The positions of lariatintron/3'-exon, lariatintron, pre-mRNA, spliced mRNA and the positions of spliceosomal complexes C, B, A, H are indicated on the left. Gels were visualized by autoradiography. DMSO, control reaction (solvent).

The modification on the o-ethylphenyl group generally also didn't have a strong effect on the activity of compound. The elimination of the ethyl group from the o-ethylphenyl group and introduction of fluorine, chlorine or bromine in para- or meta-position (Table 3.5 and 3.6, 344342, 344345, 344347 and 344350) had a minimal effect on the splicing inhibition potency (Fig. 3.39B). Interestingly, the introduction of fluorine in a meta-position (344350) weakened the inhibition potency of the molecule in comparison to chlorine, which appears to slightly enhance inhibition potency. Changing the position or substitution of the ethyl group with methyl (344349, 344346) didn't affect inhibition activity significantly, as well as methyl groups in the meta-/para- or meta-/ortho-positions (344348, 344352). No strong effect was also observed by the substitution of the ethyl group with an ethoxy group attached to phenyl at the meta-position (344343) or in the para-position (344344). Also a methylenedioxy- group (very strong electron donating



group) attached to the phenyl composing a methylenedioxyphenyl functional group, which is widely found in natural products, including multiple drugs (344354) didn't significantly affect the inhibition activity of the original inhibitor.

Eliminating the p-fluorophenyl group alone or accompanied by the introduction of additional groups like a phenyl to the nitrogen in the heterocyclic barbituric acid-related group (elimination of N-H moiety) weakened inhibition (344339, 164205), while the addition of methoxy groups to both phenyl groups in the para- or ortho-positions reduced the potency to a lesser extent, especially when the methoxy groups are in o-positions (344340, 344341). Substituting the oxygen with a sulphur atom in the barbituric acid-related group (344351) had a negative impact on the inhibitor potency. Taken together, these modifications had only minimal or no effect on the inhibition activity. Thus, their positions in the original molecule can potentially be used for tagging.

We failed in our efforts to find an analogue that was significantly more active than the original molecule. The most promising two active analogues found were 343867 and 343871, which possessed only minor modifications in the o-fluorophenyl group. Their inhibition potency was slightly higher than that of 028.

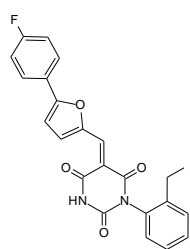
From the results obtained, we can conclude that the presence of the unsaturated linker between two heterocyclic motifs in the original compound appears to be required for inhibiting the splicing of pre-mRNA *in vitro*. Our studies also revealed that a reactive fluorine on the p-fluorophenyl group is necessary for the inhibition activity. Changing the position of fluorine from the para- to ortho-position or the presence of two neighbouring halogens in the meta- and para-positions abolishes the inhibition effect. The highest inhibitory potencies were observed using highly reactive halogens like fluorine or chlorine at the meta-position compared to fluorine at the para-position of the original molecule. Substituting the p-fluorophenyl group with heterocycle piperidine (344353) abolishes the inhibition potency. We also showed that analogues with eliminated or substituted ethyl on the o-ethylphenyl group have a similar or slightly decreased inhibition activity compared to the original product. These observations indicate that the primary ethyl group is not essential for the inhibition of pre-mRNA splicing *in vitro* and can potentially be used for tagging of the original compound 172312.

**Table 3.5 Summary of modifications to compound 028 and their effect on the inhibition of pre-mRNA splicing *in vitro* at 150  $\mu$ M.**

	<b>ID</b>	<b>Modification type</b>	<b>Potency</b>
<b>I</b>	343139	Reduced analogue	less active at 50 $\mu$ M
	343137	Reduced analogue and removed ethyl group from o-ethylphenyl group	less active at 50 $\mu$ M
	343138	Reduced analogue with removed o-ethylphenyl group	less active at 50 $\mu$ M
	343134	Removed ethyl group from o-ethylphenyl group	less active at 50 $\mu$ M
	343135	Removed o-ethylphenyl group	same as 028
	343136	Substitution of p-fluorophenyl group with bromine as a substitute for the entire reactive group in combination with eliminated ethyl group from o-ethylphenyl	less active at 50 $\mu$ M
<b>II</b>	343876	Changing the position of fluorine from para- to ortho-position	inactive
	344353	Replacing the p-fluorophenyl group with heterocycle piperidine (non aromatic ring)	inactive
	343872	Two neighbouring halogens in meta- and para-position	almost inactive
	343875	Replacement of fluorine on the p-fluorophenyl group by methyl group (an electron donor group)	inactive
	343868	Replacement of fluorine on the p-fluorophenyl group by phenyl group	clearly reduced
	343878	Eliminating the furan group with heterocyclic barbituric acid-related group connected to phenyl group with chlorine in para-position	clearly reduced
	343879	Eliminating the furan group with heterocyclic barbituric acid-related group connected to phenyl group with bromine in para-position	clearly reduced
<b>III</b>	343867	Changing position of fluorine from para- to meta-position	<b>stronger inhibition</b>
	343871	Substituting fluorine with chlorine in meta-position	<b>stronger inhibition</b>
<b>IV</b>	344342	Elimination of ethyl group from the o-ethylphenyl group and introduction of bromine in para-position	same as 028
	344345	Elimination of ethyl group from the o-ethylphenyl group and introduction of bromine in meta-position	same as 028
	344347	Elimination of ethyl group from the o-ethylphenyl group and introduction of chlorine in meta-position	slightly more active
	344350	Elimination of ethyl group from the o-ethylphenyl group and introduction of fluorine in meta-position	slightly less active
<b>V</b>	344349	Changing the position of the ethyl group from ortho- to para-position (same electron properties of the ring)	same as 028
	344346	Substitution of the ethyl group with methyl (same electron properties)	same as 028
	344348	Di-methylation in meta-/para-positions	same as 028
	344352	Di-methylation in meta-/ortho-positions	same as 028
<b>VI</b>	344343	Substitution of the ethyl group with an ethoxy group attached to phenyl in para-position (more electron donating properties)	same as 028
	344344	Substitution of the ethyl group with an ethoxy group attached to phenyl in meta-position	slightly less active
	344354	Methylenedioxy group attached to phenyl composing methylenedioxyphenyl functional group (very strong electron donating group)	slightly less active
	344351	Substituting the oxygen with sulphur atom in barbituric acid-related group	less active

<b>VII</b>	343873	Two halogens in meta-positions	same as 028
	344355	Substituting fluorine with trifluoromethyl ether group	slightly more active
<b>VIII</b>	343869	Dimethylamino group attached to a phenyl group (very strong donating)	slightly less active
	343870	Acetyl in para-position (electron withdrawing group)	slightly less active
	343874	Substituting the fluorine with methoxy group in ortho-position	less active
	343877	Two methoxy groups in para- and meta-position	same as 028
<b>IX</b>	344339	Eliminating the p-fluorophenyl group	less active
	164205	Eliminating the p-fluorophenyl group, introduction of phenyl group to the nitrogen in heterocyclic barbituric acid-related group (Elimination of N-H moiety)	less active
	344340	Adding the methoxy groups to both phenyl groups in para- position	less active
	344341	Adding the methoxy groups to both phenyl groups in ortho-position	less active

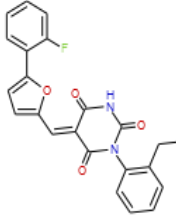
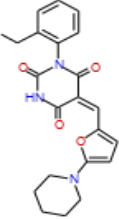
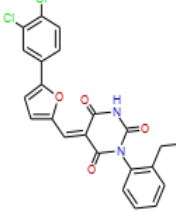
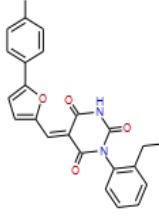
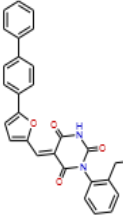
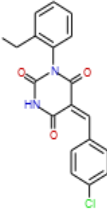
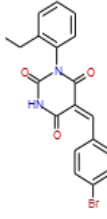
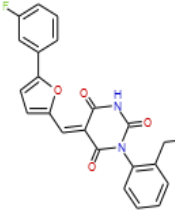
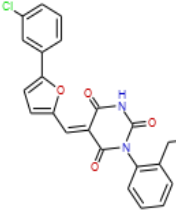
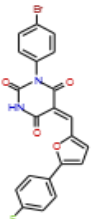
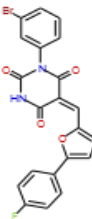
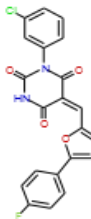
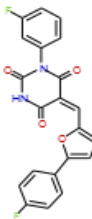
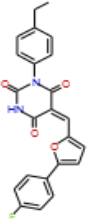
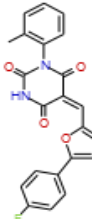
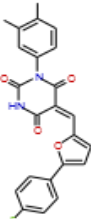
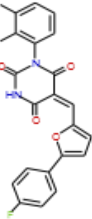
**Table 3.6 Chemical structures of synthesized analogues of compound 028.**

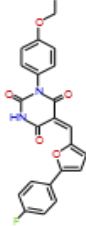
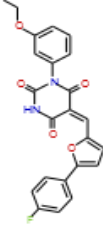
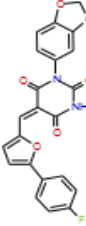
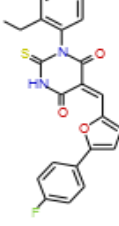
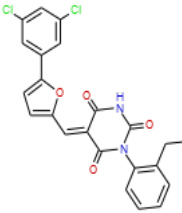
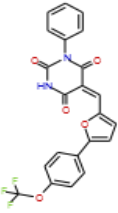
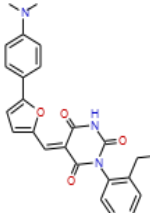
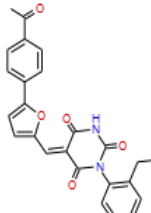
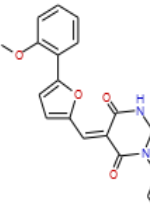
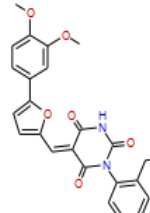
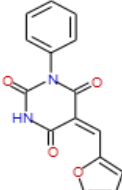
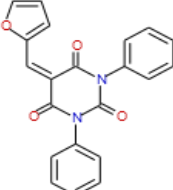
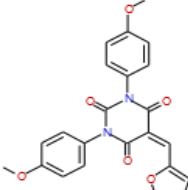
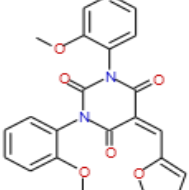


compound 028 (172312)

(-) no inhibition  
 (+) weak inhibition  
 (++) unchanged inhibition  
 (+++) enhanced inhibition

<b>I</b>	 343139	 343137	 343138
Inhibition activity	+	+	+
	 343134	 343135	 343136
Inhibition activity	+	++	+

II	 343876	 344353	 343872	 343875
Inhibition activity	-	-	-	-
	 343868	 343878	 343879	
Inhibition activity	+	+	+	
III	 343867		 343871	
Inhibition activity	+++		+++	
IV	 344342	 344345	 344347	 344350
Inhibition activity	++	++	++	++
V	 344349	 344346	 344348	 344352
Inhibition activity	++	++	++	++

VI	 344343	 344344	 344354	 344351
Inhibition activity	++	++	++	+
VII	 343873		 344355	
Inhibition activity	++		++	
VIII	 343869	 343870	 343874	 343877
Inhibition activity	++	++	+	++
IX	 344339	 164205	 344340	 344341
Inhibition activity	+	+	+	+

## 4 Discussion

Pre-mRNA splicing is a highly intricate process due to the highly complex composition of the spliceosome and its continuous structural rearrangements during splicing. The dynamic nature of the spliceosome is a serious impediment for the investigation of its structure. Stalling spliceosome assembly at distinct stages is a powerful tool for understanding the compositional and structural dynamics of the spliceosome, because it potentially allows us to isolate homogeneous/stabilised populations of spliceosomes for structural and functional studies. The aim of this work was to identify new small molecule inhibitors of pre-mRNA splicing that stall spliceosome assembly and enable the isolation of more homogeneous complexes or complexes stalled at a new intermediate stage of splicing. In the presented work, I not only managed to identify several inhibitors of pre-mRNA splicing *in vitro*, which block pre-mRNA splicing at different stages of spliceosome assembly, but also succeeded with one compound (028) to stall the spliceosome at a novel stage of the splicing cycle – i.e. during its catalytic activation. This new assembly intermediate was subsequently isolated and characterized in terms of its composition, its RNA-RNA network, and also its overall structure by EM. The stalled complex was functionally active, as it could be chased into a catalytically active spliceosome in the presence of HeLa nuclear extract treated with micrococcal nuclease. Furthermore, we carried out an initial structure-activity relationship study that revealed the structural motifs of compound 028 essential for its inhibition activity.

### 4.1 Identification of new small molecule inhibitors of pre-mRNA splicing

We first screened large chemical libraries to search for small molecule inhibitors of pre-mRNA splicing. Several approaches to screen for inhibitors of pre-mRNA splicing were previously reported. Some of them take advantage of reporter constructs, allowing the screening for an increase in a fluorescent reporter protein that is expressed from pre-mRNA designed such that only the unspliced version generates active protein (O'Brien et al., 2008). Alternative splicing modulators were identified using a two-color fluorescent reporter system in cellular assays that allows detection of the inclusion or exclusion of a particular exon (Stoilov et al., 2008). High throughput splicing assays were also developed using other techniques, such as reverse transcription followed by quantitative PCR readout (Effenberger et al., 2013). For my research I used an *in vitro*, high throughput screening assay that was developed in our lab (Samatov et al., 2012). This

assay was shown to be an effective method to screen large chemical libraries to identify compounds that inhibit pre-mRNA splicing prior to the formation of the spliceosomal complex C (Pawellek et al., 2014; Samatov et al., 2012). It measures the interaction of the spliceosomal C-specific DEAD box ATPase Abstract with the spliceosome and therefore is used as a marker for catalysis of the first step of splicing (described in detail in section 3.1.1). Using this assay ca. 172000 compounds were screened, and 10 novel compounds were initially confirmed to be inhibitors of pre-mRNA splicing *in vitro*. None of these compounds were structurally related, excluding compounds designed 296 and 187042. The detected compounds inhibit splicing in a dose-dependent manner with IC<sub>50</sub> values ranging from 2 to 55 μM. This is significantly less potent compared to some known splicing inhibitors, such as spliceostatin A that inhibits splicing at nanomolar concentration (Kaida et al., 2007), but similar to others, such as psoromic acid with an IC<sub>50</sub> value of 56 μM or norstictic acid with an IC<sub>50</sub> value of 28 μM (Samatov et al., 2012). These compounds inhibited spliceosome assembly mainly at the A and/or B/B<sup>act</sup> complex stage. The most extensive accumulation of these complexes was seen with the compound denoted as 028. Thus, the effect of this particular compound on spliceosome assembly was studied in more detail.

## **4.2 Purification of a novel spliceosome assembly intermediate**

### **4.2.1 Composition of affinity-purified spliceosomal complexes stalled by compound 028**

To characterize the complexes accumulating in the presence of the identified compounds in detail, I purified them via the MS2 selection method and analysed their RNA and protein composition via denaturing gel analysis and/or northern blotting (RNA) or mass spectrometry and/or western blotting (protein). Previously, various human spliceosomal complexes were purified and their RNA and protein compositions were characterized (Bessonov et al., 2010; Bessonov et al., 2008). These known intermediates (B or B<sup>act</sup> complexes) were isolated in parallel and used for comparison with the complexes stalled by small molecule inhibitors.

The isolated B-like complexes stalled by compound 120 were nearly identical to B complexes formed with DMSO, whereas complexes formed in the presence of compounds 296 and 297 appeared to be a mixture of B, B<sup>act</sup> and later complexes. Thus, these complexes were not studied further (120) or will be studied in more detail in the future (296 and 297) after optimization of their isolation conditions.

Both A and B-like complexes accumulated when splicing was carried out in the presence of compound 028. Interestingly, the A complexes formed in the presence of compound 028 ( $A^{028}$ ) peaked on glycerol gradients one fraction closer to the top compared to the A complexes formed under the inhibitor-free conditions, suggesting that their composition or structure may be different. However, the RNA composition of these complexes was unchanged and their protein composition was affected only to a small extent, containing lesser amounts of several U2-related proteins (SPF45, SPF30, CHERP), as well as the A or A/B proteins DDX9 and RBM5 compared to A complexes formed in the presence of DMSO. Recently, functional studies showed that SPF30 is an essential splicing factor that associates with both U4/U6.U5 tri-snRNP and U2 snRNP components (Rappsilber et al., 2001). Its immunodepletion from nuclear extract was shown to inhibit the first step of pre-mRNA splicing by preventing the formation of complex B (Meister et al., 2001; Rappsilber et al., 2001). Taking into account the necessity of the SPF30 protein for the formation of complex B, we suggest that the accumulation of A complexes in the presence of compound 028 could potentially be caused by the absence of SPF30, blocking the transition from the A to B complex. Compound 028 could alternatively hinder other proteins needed for B complex formation, such as SRPK2 (Mathew et al., 2008) or the tri-snRNP proteins 65K and 110K (Makarova et al., 2001).

Since A, and also B-like complexes, accumulate during the inhibition of spliceosome assembly by compound 028, we tested whether tri-snRNP formation/stability might be affected by the presence of compound 028. The sedimentation behaviour on glycerol gradients of tri-snRNPs formed in the presence of compound 028 suggests that compound 028 leads to some aggregation of the tri-snRNP (Fig. 3.30). However, the tri-snRNP, and also U1 and U2 snRNP, appeared to be for the most part intact and unaffected by compound 028. The stability of the U4/U6.U5 tri-snRNPs in the presence of compound 028 was assayed by immunoprecipitation studies performed at different salt concentrations with antibodies directed against selected tri-snRNP proteins (U4/U6 60K/Prp4 and U5 116K/Snu114) to check for coprecipitated U5 and U4/U6, respectively, which demonstrates tri-snRNP formation (Fig. 3.31). The stability of the U4/U6.U5 tri-snRNPs was shown to be only slightly affected in the presence of compound 028. Thus, alterations in tri-snRNP formation do not appear to contribute to the accumulation of A or B complexes in the presence of compound 028.

The sedimentation behaviour and composition of affinity-purified B-like complexes formed in the presence of compound 028 ( $B^{028}$ ) demonstrated clear and intriguing differences compared to B or  $B^{act}$  complexes. First,  $B^{028}$  complex sedimented on glycerol



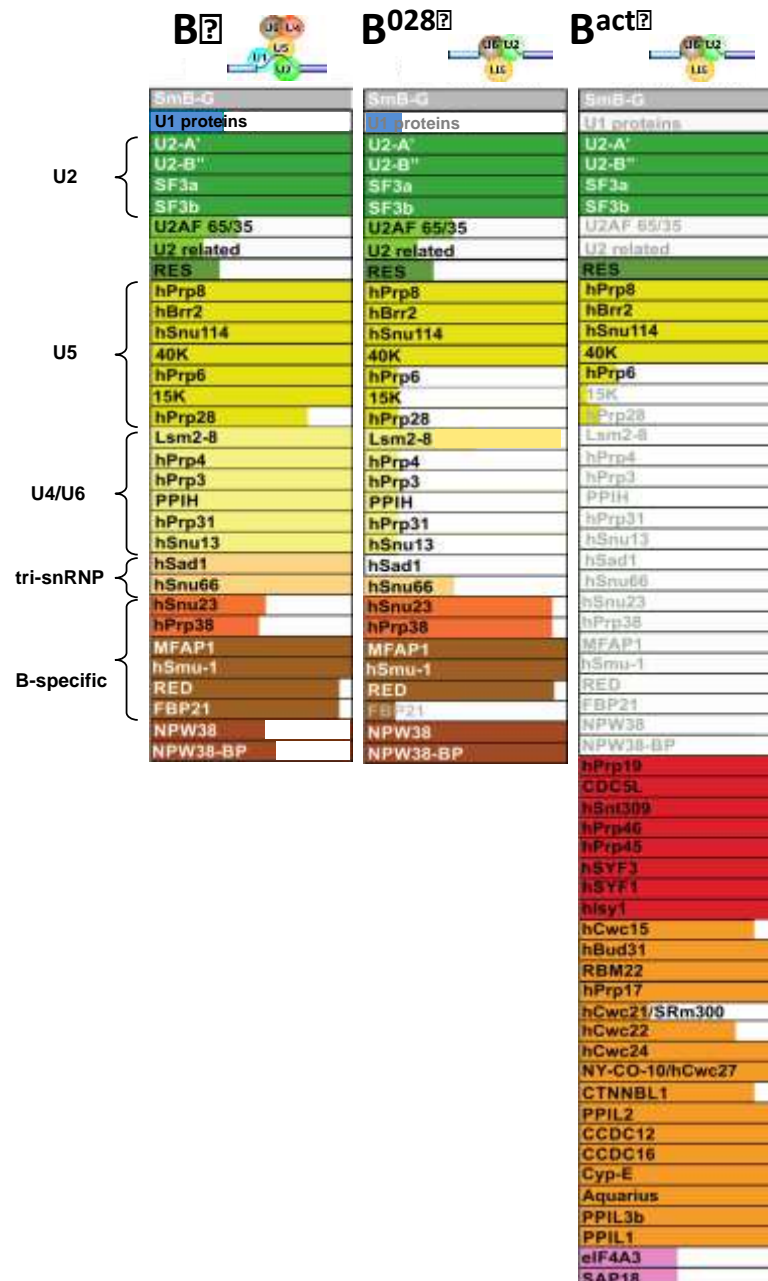
gradients with a lower S-value compared to that of complexes B<sup>act</sup> and B formed under inhibitor-free conditions, suggesting that it might be a novel spliceosome assembly intermediate. The RNA composition of the stalled B<sup>028</sup> complexes revealed that they contain equimolar amounts of unspliced pre-mRNA and U2, U5 and U6, but a lesser amount of U1 snRNA, and almost no U4 snRNA. This RNA composition suggests that splicing inhibition occurs during or after the activation stage of spliceosome assembly.

#### **4.2.2 Determination of the stage at which the spliceosome assembly is stalled in the presence of compound 028 by protein composition analysis**

Dramatic protein exchanges occur during the transition from B to B<sup>act</sup> complexes, as evidenced by mass spectrometry analysis (Bessonov et al., 2010). Many proteins (about 30) dissociate from the spliceosome during this stage of spliceosome assembly, including the B-specific proteins and U6-associated Lsm proteins. The latter is thought to enable the base-pairing interaction of the U6 snRNA with intron nucleotides near the 5' SS (Chan et al., 2003). Our research shows that the inhibition of spliceosome assembly by compound 028 occurs before destabilisation of Lsm proteins because they are clearly present in B and B<sup>028</sup> complexes, but not in B<sup>act</sup> complexes (Fig. 4.1). Similarly, B-specific proteins, which are released during activation, are still present in B and B<sup>028</sup> complexes in similar amounts. However, the B-specific FBP21 protein, which was shown to bind to several splicing factors, such as the core splicing proteins SmB/B' and SF3b49, and to be important for splicing (Huang et al., 2009; Klippel et al., 2011), is underrepresented in B<sup>028</sup> complexes, suggesting its association with the spliceosome is affected by compound 028 or the inhibition takes place after FBP21 release during activation. The role of the FBP21 protein in the observed inhibition by compound cannot be estimated, since the function of this particular protein in pre-mRNA splicing is still poorly understood.

Several groups of proteins were detected in B<sup>028</sup> and B complexes formed in the presence of DMSO in equal/similar amounts (Fig. 4.1). These include all proteins associated with U1 snRNP, U2 snRNP, and most associated with U5 snRNP, including the DExD/H-box protein hBrr2, hPrp8 and the GTPase hSnu114, which are involved in Brr2-catalysed unwinding of U4/U6 base-pair interactions. The inhibition by compound 028 takes place after the Brr2-catalysed unwinding of U4/U6 base-pair interactions as shown by the release of U4 snRNA in inhibitor-treated reactions, which is accompanied by dissociation of U4/U6 snRNP proteins. The release of U4 snRNP makes B<sup>028</sup> similar to B<sup>act</sup>. Thus, the inhibited complexes appear to proceed further during the activation step, compared to B

complexes. Moreover, several U5-specific proteins, which dissociate during activation, are underrepresented in  $B^{028}$  such as 102K/Prp6, 15K/Dib1 and the DEAD-box phosphoprotein 100K/hPrp28, which catalyses the release of U1 from the 5' SS of pre-mRNA during spliceosome activation, so that U6 can interact with the 5' SS (Chen et al., 2001; Staley and Guthrie, 1999). This suggests that the U5 snRNP is at least partially remodeled in  $B^{028}$ .



**Figure 4.10** Summary of selected proteins identified by mass spectrometry in the affinity-purified spliceosomal complexes  $B$ ,  $B^{028}$  and  $B^{act}$ . The distribution of color in each protein box reflects schematically the abundance of each identified protein. This figure was kindly provided by Dr. Berthold Kastner.

About 30 proteins are stably recruited to the spliceosome upon activation. These include B<sup>act</sup>-specific spliceosomal proteins and non-snRNP Prp19/CDC5L complex proteins, as well as Prp19/CDC5L-related factors (Bessonov et al., 2010; Chan and Cheng, 2005; Chan et al., 2003). Our results demonstrate that the inhibition by compound 028 occurs during the B to B<sup>act</sup> transition based on the absence of B<sup>act</sup>-specific spliceosomal proteins in the affinity-purified B<sup>028</sup> complexes. Interestingly, Prp19/CDC5L complex proteins and related factors, which stabilize the interactions of U5 and U6 with the pre-mRNA in B<sup>act</sup> (Chan and Cheng, 2005; Chan et al., 2003), are also underrepresented in B<sup>028</sup> complexes compared to B<sup>act</sup> complexes, suggesting that the inhibition could potentially be caused by an effect of compound 028 on the association or stable binding of the Prp19/CDC5L complex with the spliceosome. Strikingly, our results show that the Prp19/CDC5L complexes remain largely intact upon addition of compound 028 to the splicing reaction, but the association of the FLAG-tagged Prp19/CDC5L component AD002 with the Prp19/CDC5L complex appeared to be affected by the presence of compound 028 (Fig. 3.32). Surprisingly, the Npw38BP and Npw38 proteins, which dissociate during activation and are not detected in B<sup>act</sup>, are highly abundant in the B<sup>028</sup> complex (based on peptide counts and immunoblotting), suggesting that the inhibition of spliceosome assembly in the presence of compound 028 occurs prior to Npw38 and Npw38BP dissociation. Thus, proteins in the Prp19/CDC5L complex appear to be potential targets of compound 028 and interference with their activity could lead to the inhibition of pre-mRNA splicing *in vitro* by compound 028.

The U2-associated SF3b155 protein, which is known to be phosphorylated during/after the activation step but prior to the first catalytic step (Wang et al., 1998), was in a non-phosphorylated state in B and B<sup>028</sup>, but phosphorylated in B<sup>act</sup>, as shown by immunoblotting with antibodies specific for phosphorylated SF3b155 (Fig. 3.13). This indicates that the inhibition takes place before phosphorylation of SF3b155.

#### **4.2.3 Characterization of the RNA-RNA network in spliceosomes stalled by compound 028**

Monitoring the RNA-RNA network in B<sup>028</sup> complexes by UV light-induced psoralen crosslinking allowed the identification of double-stranded regions between RNA molecules of the stalled complexes. To determine what the RNA-RNA network looks like in B<sup>028</sup> and thus at which precise stage it is blocked, we also performed RNA structure probing with B, B<sup>028</sup> and B<sup>act</sup> complexes, via chemical modification of snRNA nucleotides in these three complexes.

B complex formation is accompanied by the base pairing interactions between the 5' end of U2 snRNA and 3' end of U6 snRNA, building a U2/U6 helix II (Madhani and Guthrie, 1992). Our results show that U2/U6 helix II is formed in B<sup>028</sup> and B<sup>act</sup> complexes based on the chemical inaccessibility of the corresponding U2 region. Psoralen-mediated crosslinking confirmed the formation of U2/U6 helix II in B<sup>028</sup>. However, there was a weaker crosslink signal in both B and B<sup>028</sup> complexes compared to B<sup>act</sup>. This could be due to shielding of this RNA region by proteins at early stages of spliceosome activation. To further confirm the presence of U2/U6 helix II in B<sup>028</sup> complexes, inaccessibility of the 3' end of U6 snRNA has to be shown. To look at protection in this region, the 3' end of U6 snRNA must be extended by an RNA "adaptor" fragment to enable the primer extension analysis of the 3' end of U6. However, taken together our results indicate that the U2/U6 helix II remains unaffected in stalled B<sup>028</sup> spliceosomes.

In complex B, the U6 snRNA is base-paired with the U4 snRNA, forming U4/U6 stem I and stem II. During activation, the helicase Brr2 catalyses the disruption of U4/U6 base pair interactions and U4 release occurs (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998; Staley and Guthrie, 1999), enabling further RNA rearrangements necessary for the formation of the activated spliceosome, such as the formation of an U6 snRNA internal stem-loop (U6 ISL), base pairing between U6 snRNA and the 5' splice site, and additional base pairing with U2 snRNA. It follows from its RNA and protein composition that the B<sup>028</sup> complex is inhibited after the Brr2-catalysed disruption of U4/U6 base-pair interactions. Indeed, psoralen-mediated crosslinking revealed U4/U6 and U2/U4/U6 crosslinks in complex B exclusively, but not in B<sup>028</sup> or B<sup>act</sup>, which can be attributed to the release of U4 snRNA in B<sup>028</sup> and B<sup>act</sup> complexes. Our results also show the presence of a weak crosslink of U1 snRNA with the pre-mRNA in B and B<sup>028</sup> complexes, consistent with the low levels of U1 in these spliceosomes.

The most important events occurring during activation are the formation of an U6 snRNA internal stem-loop (U6 ISL) and base pairing interactions of U6 snRNA with the 5' SS of pre-mRNA and the U2 snRNA, leading to the formation of short U2/U6-duplexes (helix Ia and helix Ib), which brings the 5' SS and BS in close proximity to each other for the first step of splicing. A psoralen-mediated crosslink between U6 snRNA and the pre-mRNA was detected in all three complexes, consistent with a base pairing interaction of the U6 ACAGAG box with intron nucleotides near the 5' SS. Structure probing revealed that the nucleotides from the U6 ACAGA box (nts A40-A44), as well as intron nts +4 to +8 are protected against chemical modification in both B<sup>act</sup> and B<sup>028</sup> complexes, confirming the interaction of the U6 ACAGA box (A41-A45) with the pre-mRNA intron

near the 5' SS in B<sup>028</sup> complexes. However, the presence of Lsm proteins and absence of Prp19/CDC5L complex proteins in B<sup>028</sup> complexes suggest that this interaction is still not stabilised (Chan and Cheng, 2005), and the structural rearrangements typical for B<sup>act</sup> in this region might not be fully-formed. The formation of the U2/BPS interaction also appears to be formed in B<sup>028</sup> complexes, based on the inaccessibility of the U2 nucleotides involved (G33-A38) for chemical modification.

Whether the U6 ISL, which plays a key role in splicing catalysis, is formed in B<sup>028</sup> spliceosomes is not clear based on our current results. Based on the accessibility of U6 snRNA nucleotides involved, the formation of the U6 ISL loop appears to be at least partially formed in B<sup>028</sup>. But some loop nucleotides are less reactive against chemical modification compared to those in B<sup>act</sup>. However, this could be due to the presence of proteins that shield these nucleotides. The stem of the U6 ISL appears to be at least partially formed in B<sup>028</sup> based on the RNA protection pattern of U6 nts G65-A56 (Fig. 3.20). However, analysis of the U6 snRNA region C79-Ψ86 still needs to be performed to clarify this question. This will require the addition of a 3' end adaptor to allow annealing of a primer before the 3' end of U6.

The formation of the U2/U6 helices Ia and Ib, through interactions of U6 nucleotides with the 5' end of U2 snRNA after U4 release, appears to occur in B<sup>028</sup>. This is supported by the inaccessibility of the involved U6 nucleotides (A50-U52 and A53-C55) and complementary U2 nucleotides (G25-C28 and G20-U22, respectively) revealed by the U6 and U2 snRNAs modification pattern analysis (Fig. 3.20, 3.21).

Previously it was shown that U2 and U6 snRNAs form a structure called a 3-way junction in activated human spliceosomes (Anokhina et al., 2013). An alternative four-way junction model for the U6/U2 interaction was also proposed (Sun and Manley, 1995). The four-way junction model of U6/U2 interaction includes the formation of U2 stem loop 1 (SL1), and was not observed in either B<sup>028</sup> and B<sup>act</sup> complexes. This conclusion is based on the accessibility of U2 nts C14 and G19, which are basepaired in the U2 stem in SL1 (Fig. 3.21). Thus, U2 and U6 don't appear to form a four-way junction in the B<sup>028</sup> complex. However, the detailed investigation of the secondary structure of the 3' end of U6 snRNA is required to decisively exclude the existence of a 4 way junction.

The U5 snRNA undergoes only minor changes during the B to B<sup>act</sup> transition, with IL2 and loop 1 more accessible in complex B versus B<sup>act</sup> (Anokhina et al., 2013). This is consistent with the fact that loop 1 contacts the 5' exon before step I of splicing (Sontheimer and Steitz, 1993; Wyatt et al., 1992). Our analysis revealed that the RNA

protection pattern within the B<sup>028</sup> and B<sup>act</sup> complexes are identical, confirming that the structure of this region of U5 is not affected by the presence of compound 028. The loop 1 nucleotides U42 and Ψ43 might interact with the 5' exon of the pre-mRNA in B<sup>028</sup>, as well as in B and B<sup>act</sup> due to their inaccessibility to chemical modification.

Finally, the toggling of yeast U2 snRNA between two different intra-molecular conformations was reported previously (Hilliker et al., 2007; Perriman and Ares, 2007). This toggling involves the formation of helix IIc, which is formed by base pairing between loop nucleotides (53-60) of U2 stem loop IIa and nucleotides downstream of stem loop IIb (88-96). Analysis of the RNA protection pattern of nucleotides involved in U2 helix IIc (nucleotides A88-A95 and U53-U60) in both B<sup>028</sup> and B<sup>act</sup> complexes demonstrated that all nucleotides involved in helix IIc are accessible to chemical modification. Thus, the existence of U2 helix IIc was not detected in the B<sup>028</sup> complex.

### **4.3 Electron microscopy studies of affinity-purified B<sup>028</sup> complexes**

We also succeeded to visualize the stalled B<sup>028</sup> complex by two-dimensional negative stain electron microscopy and compared its 2D overall structure with the previously reported 2D structures of human B and B<sup>act</sup> complexes (Bessonov et al., 2010). The typical B<sup>028</sup> complex images have a maximum dimension of ca. 40 nm and show shapes ranging from isosceles triangles to mushroom-like forms. The triangular views are characterized by an up to 40 nm long density axis present on one side of the B<sup>028</sup> particle (left edge in Fig. 3.28 columns 1-2), and a more globular density on the opposite side of the particle. A mushroom-like projection form of the B<sup>028</sup> complex was observed less frequently. This B<sup>028</sup> complex form typically has a relatively flat mushroom head and a mushroom foot that is displaced to one side. Sometimes a density axis running from the foot to one corner of the head can also be seen in the mushroom-like views. The presence of such intermediate views indicates that the triangular and mushroom-like views represent structurally similar particles bound to the carbon carrier film via different particle faces.

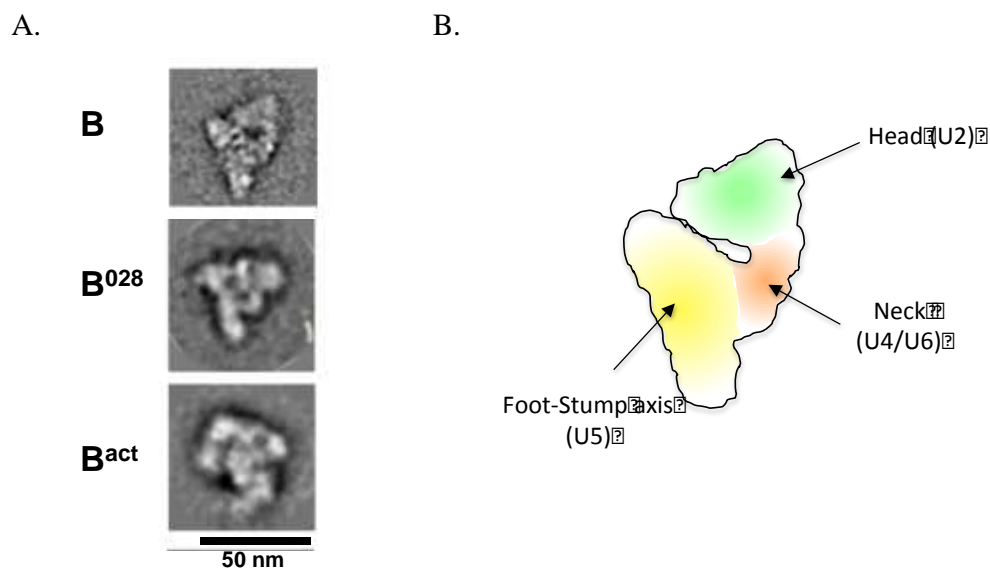
The B<sup>028</sup> complex exhibits roughly the same size as the B and B<sup>act</sup> complexes and also shares some structural similarities with both complexes, but it also exhibits obvious differences. Typical images of the B complex (Deckert et al., 2006) also show an elongated density axis on one edge (Fig. 4.1). The B complex also contains a globular domain (designated the “head”). However, compared to B<sup>028</sup>, it exhibits a different position relative to the density axis, being oriented more upwards and thereby giving rise to the typical rhombic-shaped views of the B complex. In the B<sup>028</sup> complex the globular

domain is shifted more toward the center of the elongated density axis, so that triangular views dominate (Fig. 4.1; 3.28, columns 1-4). In the B complex images, the globular domain has very different shapes with hardly any fine structure visible and thus it appears in class averages typically as a quite diffuse density area. The long density axis, called “foot-stump axis” in the B complex (Wolf et al., 2008), as well as the “arm” element linking the globular domain to the center of the axis, show a great deal of fine structure in the class averages. The intricate differences between B and B<sup>028</sup> images demonstrate that the B<sup>028</sup> complex is not just a B complex that has lost one structural element, i.e., the U4 snRNP. Instead, there appears to be extensive rearrangements in the globular domain.

Like the B<sup>028</sup> complex, the human B<sup>act</sup> complex also exhibits mushroom-like views. Mushroom-like views appear to be typical for B<sup>act</sup> complexes, as they are the dominant views displayed by the yeast B<sup>act</sup> particle (Fabrizio et al., 2009). The similarities between the B<sup>028</sup> images and those of the human B<sup>act</sup> images are more or less limited to the presences of views with mushroom-like outlines (Fig 3.28, 4.1, column 3-4, first three rows). As such mushroom-like views are not seen in B complex images, the appearance of mushroom-like views in the B<sup>028</sup> particle can be taken as an indication that the conversion of the B to the B<sup>028</sup> complex is morphologically a step towards the B<sup>act</sup> configuration. A likely reason why human B<sup>act</sup> views are so different is the presence of a large numbers of additional proteins in the B<sup>act</sup> complex, in particular the large Prp19/CDC5L complex. Because the B<sup>028</sup> and B<sup>act</sup> images do not just differ by the presence or absence of additional domains, the proteins additionally present in the B<sup>act</sup> particle likely create new preferred carbon film binding sites on the particle so that different views are seen in the EM images.

The globular head domain in the human B complex was reported to contain U2 snRNP components (U2-associated SF3b155 protein), as well as the intron and both exons of the pre-mRNA (Wolf et al., 2009). More recent immuno-EM labeling experiments in our lab with yeast B complexes, which are morphologically very similar to human B complexes, indicate that U5 proteins like Prp8 and Brr2 are located in the long density axis of the B complex, while the neck contains U4/U6 proteins (Dr. Norbert Rigo and Dr. Berthold Kastner, personal communication) (Fig. 4.2B). Based on their morphological similarities the gross organisation of the components of the long axis in the B complex seem to be maintained to a large degree in the B<sup>028</sup> complex. In contrast, the globular head domain of the B complex, containing the pre-mRNA and U2 snRNP proteins, appears to have undergone major structural changes during the transition to the B<sup>028</sup> complex, moving downwards toward the central part of the long axis. The U5 snRNP protein Prp8 is of

central importance for formation and maintaining the spliceosome's catalytic center. The apparent movement of the pre-mRNA and the U2 proteins present in the globular domain closer to the center of the long axis that harbors Prp8, suggests that there has been initiation of catalytic center formation in the  $B^{028}$  complex.



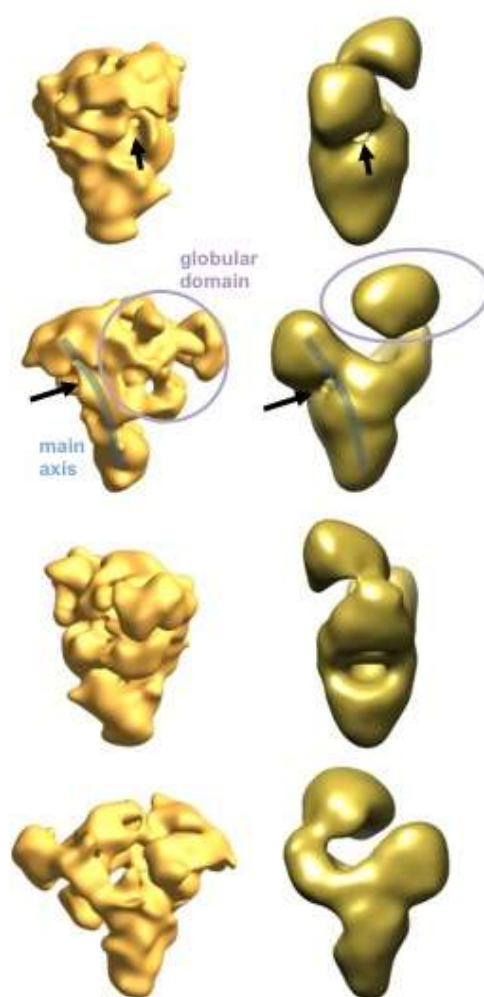
**Figure 4.2 Comparison of 2D structures of spliceosomal complexes B, B<sup>act</sup> and intermediate B<sup>028</sup>.** A. The typical negative stained 2D structures are shown. Scale bar corresponds to 50 nm. B. The position of the U2, U4/U6 and U5 snRNPs within the B complex is indicated.

The quality of the isolated  $B^{028}$  complexes in terms of their structural homogeneity allowed us to perform 3D EM structure analysis and to generate a 3D reconstruction of the intact  $B^{028}$  spliceosome (performed by Prof. Dr. Holger Stark) (Fig. 3.29, 3.30) with a resolution of ca. 25 Å. The reconstructed  $B^{028}$  structure demonstrated a unique 3D shape, distinct from all known spliceosomal complexes. It exhibited an almost triangular structure formed from an elongated, relatively massive main axis running from the foot to the upper corner, which is composed of a globular structure with a central cavity and a flatter structure with a beak-like protuberance (Fig. 3.29, 1, 2). Attached to the main axis is a more globular domain (Fig. 3.29, 7). The main axis and globular domain, are connected to each other by a complicated network of structural elements (described in detail in section 3.3.2). Notably, by slightly tilting the 3D structure (Fig. 3.30) a mushroom-like shape as observed in the 2D class averages can be envisaged.

A 3D structure of a human B-type spliceosomal complex was reconstructed from EM images more than ten years ago (Boehringer et al., 2004). These complexes were treated



with heparin and were not stabilized by GraFix, and the 3D structure could only be determined to about 4 nm because of substantial structural heterogeneity in its globular head domain. The fairly flat 3D structure shows the foot-stump axis as a compact mass with quite similar dimensions as the long density axis of the B<sup>028</sup> complex 3D structure (Fig. 4.2). Even a hole between the upper and lower portion of in the main density axis, as seen in B028, is visible in the BΔU1 3D structure (marked with an arrow in Fig. 4.2). Due to structural heterogeneity, the head domain appears as a smaller globular density in the BΔU1 3D structure and is connected via the U4-containing neck structure to the centre of the foot-stump main axis. As discussed in the original paper (Boehringer et al., 2004), the reduced head domain must be enlarged drastically to better reflect the B complex structure. Nevertheless, comparison of the BΔU1 3D structure with that of the B<sup>028</sup> complex (Fig. 4.3) gives an initial hint as to how the globular domain of the B complex might be bent downward toward the main axis during formation of the B<sup>028</sup> complex.



**Figure 4.3 Comparison of the 3D structures of B<sup>028</sup> and human BΔU1.**

The future EM investigation of the spliceosome, especially generating a 3D structure of the human B<sup>act</sup> complex and a B complex structure at a higher resolution, will allow a better, more detailed comparison with the B<sup>028</sup> complex 3D structure. This could provide deeper insight into the structural dynamics during activation and help localize spliceosomal components, e.g. those present in B or B<sup>act</sup> but missing in B<sup>028</sup>. Moreover, the cryo-EM grids already successfully produced should allow a higher resolved cryo-EM 3D reconstitution of the B<sup>028</sup> complex that should give a more detailed insight into its architecture in the near future.

#### **4.4 Identification of chemical groups important for compound 028's splicing inhibition activity**

A structure–activity relationship study (SAR) was performed to reveal structural motifs that are important for the splicing inhibition activity of compound 028. To this end a wide spectrum of structural modifications were introduced at specific sites of the original molecule and the effect of these modifications on pre-mRNA splicing was tested in an *in vitro* splicing assay using radioactively-labeled MINX pre-mRNA.

Analysis of compound 028 derivatives with a reduced linker between the two heterocyclic motifs of the original molecule indicated the importance of the Michael acceptor character of compound 028 for its splicing inhibition activity, suggesting the reactivity of compound 028 towards nucleophiles. In the next set of experiments, we investigated the effect of modifications in the o-ethylphenyl group and/or in p-fluorophenyl group, with the goal of changing the electron properties on the rings (electron donor or electron withdrawing) and/or changing the size of the substitution. Our study revealed that the p-fluorophenyl group, including its particular components like a singular reactive fluorine at the para- or meta-positions, as well as an aromatic phenyl group and the furan group located between the heterocyclic barbituric acid-related group and the p-fluorophenyl group, are required for splicing inhibition activity, as evidenced by the negative effect of modifications introduced at these positions. Changing the fluorine position from the para- to meta-position (343867) or substituting fluorine with chlorine in the meta-position (343871) clearly enhanced the splicing inhibition effect of the original compound, confirming that the p-fluorophenyl group is extremely important for the activity of compound 028, and suggesting that it might play a direct role in the inhibition reaction or reaction cascade (i.e. in the targeting of compound 028 to its inhibition partner). Interestingly, the replacement of fluorine by a methyl group (electron donator) (343875) or bulky phenyl group (343868) abolished the inhibition activity.

However, some of the modifications introduced into the p-fluorophenyl group, such as two halogens in the meta-positions (343873), a dimethylamino group (electron donating group) attached to the phenyl group (343869) or an acetyl group (electron withdrawing group) in the para-position (343870) didn't have any significant effect on the inhibition activity, whereas substituting a fluorine with a trifluoromethyl ether group (344355) appears to enhance inhibition. Likewise, nearly all modifications in the o-ethylphenyl group didn't alter the inhibition activity of compound 028 significantly, suggesting it is not required for the observed splicing inhibition.

Taken together, these results revealed the importance of the unsaturated linker between two heterocyclic motifs, the reactive fluorine on the p-fluorophenyl group and the entire aromatic p-fluorophenyl group for inhibiting the splicing of pre-mRNA *in vitro*. Unfortunately, we succeeded only in minimally enhancing the inhibition activity of compound 028 by minor modification in its original structure. Due to the rather high IC<sub>50</sub> value for compound 028 (i.e. 51 μM) enhancing its potency remains an important goal in the future. Moreover, we identified parts of the molecule, which appear not to contribute to the inhibition activity, based on our results that the modifications located in these parts have only minimal or no effect on the splicing inhibition activity of compound 028. Thus, these particular sites can potentially be used to introduce a tag into compound 028, without significant loss of its inhibition activity. We suggest the primary ethyl group on the o-ethylphenyl group is a very promising site for tagging. A tagged version of 028 could be used for further investigation of possible target(s) of compound 028 in the splicing reaction. To this end, covalently attaching biotin to the compound might be highly useful due to its efficiency, specificity and small size. To identify the binding partners of compound 028, a biotinylated version could be added to the splicing reaction and affinity selection with streptavidin beads, followed by extensive washing and elution of bound molecules could be carried out. The co-purified proteins can be finally identified by mass spectrometry and/or immunoblotting. Alternatively, we could use a fluorescence- or photoaffinity-tagged version of compound 028 for subsequent cross-linking studies, to search for the protein(s) that specifically bind to compound 028.

#### **4.5 B<sup>028</sup> spliceosomes can be used to investigate binding and release of proteins during activation.**

The investigation of the composition of the B<sup>028</sup> complex formed in the presence of compound 028 revealed a novel intermediate between B and B<sup>act</sup> not detected previously. Furthermore, we demonstrate that a large fraction of the affinity-purified B<sup>028</sup> complexes

can be chased into catalytically active spliceosomes that catalyse both steps of splicing when incubated with MN-treated HeLa nuclear extract. Thus, the block by compound 028 can be overcome by incubating with the pool of spliceosomal proteins in nuclear extract. This means that compound 028 might prevent certain proteins from binding (this is overcome by adding NE without the inhibitor) or there is a displacement of a protein bound by 028 in the  $B^{028}$  complex with a protein not bound by the inhibitor, which is now active in splicing and activation can proceed. The affinity-purified stalled  $B^{028}$  complex is a functionally active assembly intermediate, which is stalled after Brr2-catalysed U4 snRNA release, but before release of B-specific and Lsm proteins, and before the recruitment of the  $B^{act}$ -specific and Prp19/CDC5L-associated proteins to the spliceosome. We thus used stalled  $B^{028}$  complexes to investigate the dynamics of the subsequent recruitment/release of proteins during the activation step. This was done by incubating purified  $B^{028}$  complexes that were bound to amylose beads either with buffer alone or with normal or MN-treated nuclear extract in the presence or in the absence of ATP. The binding or loss of proteins was determined by mass spectrometry. To stall the spliceosome assembly after activation, i.e. at the stage of  $B^{act}$  formation, we used the PM5-10 pre-mRNA substrate. This was important to follow whether  $B^{act}$ -specific proteins were recruited, since they are released again already when the C complex forms.

The RNA analysis of the “chased”  $B^{028}$  complexes showed that the complexes stay intact during the 60 min chase. Indeed, mass spectrometry analysis showed that nearly all proteins that are not expected to be lost during activation were present in the chased  $B^{028}$  complexes in similar or same amounts compared to the unchased  $B^{028}$ , including the U2 and U5 snRNP proteins. This also shows that the complexes do not dissociate under chase conditions. In contrast, several groups of proteins were incorporated and some lost, showing the suitability of the  $B^{028}$  complexes for examining the exchange of proteins during activation.

The human hPrp19/CDC5L complex proteins incorporate into the spliceosome stably during activation (Makarov et al., 2002; Makarova et al., 2004; Wahl et al., 2009), and remain associated throughout both transesterification reactions, stabilizing the association of U5 and U6 with the spliceosome after U4 release (Chan et al., 2003, 2005). The mass spectrometry analysis shows that  $B^{028}$ , in contrast to both B and  $B^{act}$ , contains underrepresented amounts of Prp19/CDC5L complex proteins, as well as many Prp19/CDC5L-related proteins. The incubation of  $B^{028}$  with MN-treated nuclear extract or untreated nuclear extract under splicing conditions leads to a substantial increase in the amounts of bound Prp19/CDC5L complex proteins and related proteins. However, the

efficient recruitment of this group of proteins requires ATP, based on the result that only small amounts of Prp19/CDC5L proteins are detected in the B<sup>028</sup> spliceosomes chased under ATP-free conditions. However, even in the presence of ATP, the number of peptides identified for the Prp19/CDC5L and related proteins (as well as B<sup>act</sup>-specific proteins) in the chased complexes is significantly below that identified in the purified B<sup>act</sup> complex. Thus, not all B<sup>028</sup> complexes can be chased into B<sup>act</sup> under these conditions. Interestingly, Npw38 and Npw38BP, which are normally released during activation but accumulate in the stalled B<sup>028</sup> spliceosomes, were not released from B<sup>028</sup> complexes chased with MN-treated extract (+ATP) (i.e. peptide counts went up), but were found in lower amounts only when the complexes were chased with MN-treated extract in the absence of ATP, or chased with normal nuclear extract (+ATP). This particular difference between complexes chased in the presence of MN-treated extract and those chased in the presence of normal extract, suggests that there is a negative effect of MN treatment on the composition and/or splicing activity of nuclear extract components, but only in the presence of ATP. Generally, the efficiency of chasing the B<sup>028</sup> complexes into activated complexes (based on protein composition) was better in the presence of normal nuclear extract compared to that in the presence of the MN-treated extract. However, in case of the non-treated nuclear extract, it cannot be entirely ruled out that the observed splicing activity is due to the disassembly of B<sup>028</sup> complexes and subsequent reassembly of new spliceosomes.

Another group of spliceosomal proteins integrated into the spliceosome during activation are the B<sup>act</sup>-specific proteins. We demonstrate that these proteins are also recruited to the B<sup>028</sup> spliceosome after incubating with the MN-treated extract or with the normal nuclear extract in an ATP-dependent manner. This could be a direct requirement (at the time when the B<sup>act</sup>-specific proteins bind) or ATP could be needed for the binding or release of other spliceosomal proteins, which is a prerequisite for the binding of the B<sup>act</sup>-specific proteins. Also as mentioned below, the requirement for ATP could also reflect the need for a protein phosphorylation event.

B-specific and Lsm proteins are present in B complexes, but dissociate from the spliceosome during activation. Stalled B<sup>028</sup> complexes were shown to contain both groups of proteins, indicating that the inhibition takes place before their dissociation. Our chase experiments suggest that they leave the B<sup>028</sup> spliceosome in an ATP-independent manner, when incubated with normal or MN-treated extract. However, mass spectrometry analysis shows significant amounts of these groups of proteins remaining in our B<sup>act</sup> complexes, impeding the estimation of the efficiency of their release during activation. In contrast,

the 2<sup>nd</sup> step factor Prp22 and the RES complex proteins are recruited to the spliceosome upon incubation with the MN-treated extract or the normal extract under the splicing conditions, but only in the presence of ATP. Moreover, the spliceosomal RNA helicases Prp5 and Prp43 were identified in the chased spliceosomes in large amounts in the absence of ATP, suggesting their dissociation is ATP-dependent.

Thus, B<sup>028</sup> complexes are functionally active and they can proceed to activated spliceosomes under splicing conditions when supplemented with spliceosomal proteins in nuclear extract. The initial chase experiments indicate a requirement of ATP for the recruitment of key proteins such as Prp19/CDC5L complex proteins, B<sup>act</sup>-specific and RES complex proteins and/or to enable the release of other proteins. Why might ATP be required after U4 snRNA release to form an activated spliceosome? A simple reason might be that ATP is needed to “free up” a protein needed for activation that is in an inactive complex in the extract. Another is that enzymatic activity (hydrolysis ATP) is required, such as kinase activity, for the recruitment or release of proteins during activation. Thus, phosphorylation of proteins might play a role here. This in turn suggests that compound 028 might possibly block the activity of a kinase catalysing the proposed phosphorylation event that is needed for activation to proceed. It will be of interest to see if ATP hydrolysis is really required by using a non-hydrolysable form of ATP in future experiments. Also, a kinetic analysis (shorter incubation time points) might provide information about the order of binding – for example, if loss of Lsm proteins must occur before the release of B-specific proteins.

## 4.6 Perspectives

Additional studies with B<sup>028</sup> complexes should be performed in the future to elucidate the mechanism by which spliceosome assembly is stalled by compound 028, and to extend our understanding of structural and compositional dynamic rearrangements occurring during the activation step. Further EM studies with the goal of improving resolution could be done to learn more about the structural organization of stalled B<sup>028</sup> complexes (in collaboration with Prof. Dr. Holger Stark). A higher resolution structure would provide the opportunity to model proteins or snRNPs into the density map and thus localize them in the B<sup>028</sup> spliceosomes. To elucidate the RNA-RNA network in B<sup>028</sup> in more detail, the analysis of the RNA protection pattern at the 3' end of U6 snRNA in B<sup>028</sup> complexes will be undertaken. This will require that the 3' end of U6 snRNA is extended by an RNA “adaptor” fragment and analysed using the structure probing procedure to clarify whether the U6 ISL is formed in B<sup>028</sup> complex. The tagging of the compound 028 on the basis of

the obtained SAR data might be helpful to identify its target(s), with a goal to uncover the mechanism of the inhibition of pre-mRNA splicing by compound 028. In addition, the investigation of the inhibition activity of compound 028 in yeast (*S. cerevisiae*) pre-mRNA splicing system should elucidate whether the inhibition target is common to both yeast and human spliceosomes. Further chasing experiments are of particular importance with the goal of understanding the dynamics of protein recruitment by the analysis of protein exchange after shorter kinetic steps. Additionally, chasing experiments could be performed with non-hydrolysable ATP analogues to determine whether ATP hydrolysis is needed for the binding of B<sup>act</sup>-specific and Prp19/CDC5L complex proteins. Moreover, it will be interesting to clarify the phosphorylation status of different proteins during the chasing of B<sup>028</sup> complexes in the presence/absence of ATP. Finally, the inhibition of spliceosome assembly by other identified compounds (296/297) will be investigated in the future in more detail, to obtain hopefully new information about the dynamic protein recruitment/release and structural rearrangements during the activation step. Taken together, these studies will deepen our understanding of the structural and compositional rearrangements during the spliceosome activation step.

## 5 References

Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Luhrmann, R. (1998). The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Molecular and cellular biology* 18, 6756-6766.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *The EMBO journal* 18, 5789-5802.

Anokhina, M., Bessonov, S., Miao, Z., Westhof, E., Hartmuth, K., and Lührmann, R. (2013). RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *The EMBO journal* 32, 2804-2818.

Ares, M., Jr., Grate, L., and Pauling, M.H. (1999). A handful of intron-containing genes produces the lion's share of yeast mRNA. *Rna* 5, 1138-1139.

Azubel, M., Wolf, S.G., Sperling, J., and Sperling, R. (2004). Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Molecular cell* 15, 833-839.

Bach, M., Winkelmann, G., and Lührmann, R. (1989). 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proceedings of the National Academy of Sciences of the United States of America* 86, 6038-6042.

Behrens, S.E., and Lührmann, R. (1991). Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes & development* 5, 1439-1452.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., *et al.* (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *The EMBO journal* 26, 1737-1748.

Bellare, P., Kutach, A.K., Rines, A.K., Guthrie, C., and Sontheimer, E.J. (2006). Ubiquitin binding by a variant Jab1/MPN domain in the essential pre-mRNA splicing factor Prp8p. *Rna* 12, 292-302.

Bellare, P., Small, E.C., Huang, X., Wohlschlegel, J.A., Staley, J.P., and Sontheimer, E.J. (2008). A role for ubiquitin in the spliceosome assembly pathway. *Nature structural & molecular biology* 15, 444-451.

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

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *Rna* 16, 2384-2403.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452, 846-850.

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 research* 21, 5803-5816.



- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry* 72, 291-336.
- Boehringer, D., Makarov, E.M., Sander, B., Makarova, O.V., Kastner, B., Lührmann, R., and Stark, H. (2004). Three-dimensional structure of a pre-catalytic human spliceosomal complex. *Nature structural & molecular biology* 11, 463-468.
- Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126, 713-725.
- Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Haendler, B., and Jacob, M. (1982). U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *The EMBO journal* 1, 1259-1265.
- Brosi, R., Groning, K., Behrens, S.E., Lührmann, R., and Kramer, A. (1993a). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science* 262, 102-105.
- Brosi, R., Hauri, H.P., and Kramer, A. (1993b). Separation of splicing factor SF3 into two components and purification of SF3a activity. *The Journal of biological chemistry* 268, 17640-17646.
- Brow, D.A., and Guthrie, C. (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334, 213-218.
- Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of Precursors to mRNAs by the Spliceosomes. In *The RNA world : the nature of modern RNA suggests a prebiotic RNA*, R.F. Gesteland, T. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press), pp. 525-560.
- Cao, G., Garcia, C.K., Wyne, K.L., Schultz, R.A., Parker, K.L., and Hobbs, H.H. (1997). Structure and localization of the human gene encoding SR-BI/CLA-1. Evidence for transcriptional control by steroidogenic factor 1. *The Journal of biological chemistry* 272, 33068-33076.
- Cech, T.R. (1986). The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* 44, 207-210.
- Chan, S.P., and Cheng, S.C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *The Journal of biological chemistry* 280, 31190-31199.
- Chan, S.P., Kao, D.I., Tsai, W.Y., and Cheng, S.C. (2003). The Prp19p-associated complex in spliceosome activation. *Science* 302, 279-282.
- Chapman, K.B., and Boeke, J.D. (1991). Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell* 65, 483-492.
- Chen, J.Y., Stands, L., Staley, J.P., Jackups, R.R., Jr., Latus, L.J., and Chang, T.H. (2001). Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Molecular cell* 7, 227-232.
- Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes & development* 11, 2755-2766.
- Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* 349, 487-493.

Convertini, P., Shen, M., Potter, P.M., Palacios, G., Lagisetti, C., de la Grange, P., Horbinski, C., Fondufe-Mittendorf, Y.N., Webb, T.R., and Stamm, S. (2014). Sudemycin E influences alternative splicing and changes chromatin modifications. *Nucleic acids research* *42*, 4947-4961.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. *Cell* *136*, 777-793.

Corrionero, A., Minana, B., and Valcarcel, J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes & development* *25*, 445-459.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Molecular cell* *5*, 779-787.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Molecular and cellular biology* *26*, 5528-5543.

Dziembowski, A., Ventura, A.P., Rutz, B., Caspary, F., Faux, C., Halgand, F., Laprevote, O., and Seraphin, B. (2004). Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *The EMBO journal* *23*, 4847-4856.

Effenberger, K.A., Perriman, R.J., Bray, W.M., Lokey, R.S., Ares, M., Jr., and Jurica, M.S. (2013). A high-throughput splicing assay identifies new classes of inhibitors of human and yeast spliceosomes. *Journal of biomolecular screening* *18*, 1110-1120.

Fabrizio, P., and Abelson, J. (1990). Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science* *250*, 404-409.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Molecular cell* *36*, 593-608.

Fica, S.M., Tuttle, N., Novak, T., Li, N.S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature* *503*, 229-234.

Fischer, U., Sumpter, V., Sekine, M., Satoh, T., and Lührmann, R. (1993). Nucleocytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *The EMBO journal* *12*, 573-583.

Folco, E.G., Coil, K.E., and Reed, R. (2011). The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. *Genes & development* *25*, 440-444.

Fourmann, J.B., Schmitzova, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.L., Fabrizio, P., and Lührmann, R. (2013). Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes & development* *27*, 413-428.

Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 16176-16181.

- Galej, W.P., Nguyen, T.H., Newman, A.J., and Nagai, K. (2014). Structural studies of the spliceosome: zooming into the heart of the machine. *Current opinion in structural biology* 25, 57-66.
- Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Lührmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nature communications* 3, 994.
- Golas, M.M., Sander, B., Bessonov, S., Grote, M., Wolf, E., Kastner, B., Stark, H., and Luhrmann, R. (2010). 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. *Molecular cell* 40, 927-938.
- Golas, M.M., Sander, B., Will, C.L., Lührmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. *Science* 300, 980-984.
- Golas, M.M., Sander, B., Will, C.L., Lührmann, R., and Stark, H. (2005). Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Molecular cell* 17, 869-883.
- 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 & development* 10, 233-243.
- Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. *Rna* 11, 533-557.
- Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. *Rna* 6, 1197-1211.
- Graveley, B.R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Molecular cell* 1, 765-771.
- Guthrie, C., and Patterson, B. (1988). Spliceosomal snRNAs. *Annual review of genetics* 22, 387-419.
- Hacker, I., Sander, B., Golas, M.M., Wolf, E., Karagoz, E., Kastner, B., Stark, H., Fabrizio, P., and Lührmann, R. (2008). Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nature structural & molecular biology* 15, 1206-1212.
- Heinrichs, V., Bach, M., Winkelmann, G., and Lührmann, R. (1990). U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science* 247, 69-72.
- Herold, N., Will, C.L., Wolf, E., Kastner, B., Urlaub, H., and Lührmann, R. (2009). Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes. *Molecular and cellular biology* 29, 281-301.
- Hilliker, A.K., Mefford, M.A., and Staley, J.P. (2007). U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes & development* 21, 821-834.
- 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 & development* 6, 2554-2568.

Huang, X., Beullens, M., Zhang, J., Zhou, Y., Nicolaescu, E., Lesage, B., Hu, Q., Wu, J., Bollen, M., and Shi, Y. (2009). Structure and function of the two tandem WW domains of the pre-mRNA splicing factor FBP21 (formin-binding protein 21). *The Journal of biological chemistry* 284, 25375-25387.

Jacquier, A., and Michel, F. (1990). Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. *Journal of molecular biology* 213, 437-447.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., *et al.* (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nature chemical biology* 3, 576-583.

Kambach, C., Walke, S., and Nagai, K. (1999). Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Current opinion in structural biology* 9, 222-230.

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., *et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature methods* 5, 53-55.

Kim, S.H., and Lin, R.J. (1996). Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Molecular and cellular biology* 16, 6810-6819.

Kistler, A.L., and Guthrie, C. (2001). Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes & development* 15, 42-49.

Klippel, S., Wieczorek, M., Schumann, M., Krause, E., Marg, B., Seidel, T., Meyer, T., Knapp, E.W., and Freund, C. (2011). Multivalent binding of formin-binding protein 21 (FBP21)-tandem-WW domains fosters protein recognition in the pre-spliceosome. *The Journal of biological chemistry* 286, 38478-38487.

Koodathingal, P., and Staley, J.P. (2013). Splicing fidelity: DEAD/H-box ATPases as molecular clocks. *RNA biology* 10, 1073-1079.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nature chemical biology* 3, 570-575.

Kramer, A., and Utans, U. (1991). Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. *The EMBO journal* 10, 1503-1509.

Kuhn, A.N., van Santen, M.A., Schwienhorst, A., Urlaub, H., and Lührmann, R. (2009). Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. *Rna* 15, 153-175.

Laggerbauer, B., Achsel, T., and Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 95, 4188-4192.

Laggerbauer, B., Liu, S., Makarov, E., Vornlocher, H.P., Makarova, O., Ingelfinger, D., Achsel, T., and Lührmann, R. (2005). The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation. *Rna* 11, 598-608.

Lauber, J., Plessel, G., Prehn, S., Will, C.L., Fabrizio, P., Groning, K., Lane, W.S., and Lührmann, R. (1997). The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *Rna* 3, 926-941.

Lesser, C.F., and Guthrie, C. (1993). Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* 262, 1982-1988.

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

Lührmann, R., and Stark, H. (2009). Structural mapping of spliceosomes by electron microscopy. *Current opinion in structural biology* 19, 96-102.

Madhani, H.D. (2013). snRNA catalysts in the spliceosome's ancient core. *Cell* 155, 1213-1215.

Madhani, H.D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71, 803-817.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Lührmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298, 2205-2208.

Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P., and Lührmann, R. (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6\*U5 tri-snRNP formation and pre-mRNA splicing. *The EMBO journal* 21, 1148-1157.

Makarova, O.V., Makarov, E.M., and Lührmann, R. (2001). The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *The EMBO journal* 20, 2553-2563.

Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *The EMBO journal* 23, 2381-2391.

Martin, A., Schneider, S., and Schwer, B. (2002). Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *The Journal of biological chemistry* 277, 17743-17750.

Massenet, S., Mougín, A., and Branlant, C. (1998). Posttranscriptional modifications in the U small nuclear RNAs. In *The Modification and Editing of RNA* H. Grosjean, and R. Benne, eds. (ASM Press, Washington, DC), pp. 201-228.

Mathew, R., Hartmuth, K., Mohlmann, S., Urlaub, H., Ficner, R., and Lührmann, R. (2008). Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nature structural & molecular biology* 15, 435-443.

Mattaj, I.W. (1986). Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* 46, 905-911.

McManus, C.J., Schwartz, M.L., Butcher, S.E., and Brow, D.A. (2007). A dynamic bulge in the U6 RNA internal stem-loop functions in spliceosome assembly and activation. *Rna* 13, 2252-2265.

Meister, G., Hannus, S., Plattner, O., Baars, T., Hartmann, E., Fakan, S., Laggerbauer, B., and Fischer, U. (2001). SMNrp is an essential pre-mRNA splicing factor required for the formation of the mature spliceosome. *The EMBO journal* *20*, 2304-2314.

Modrek, B., and Lee, C.J. (2003). Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nature genetics* *34*, 177-180.

Mohlmann, S., Mathew, R., Neumann, P., Schmitt, A., Lührmann, R., and Ficner, R. (2014). Structural and functional analysis of the human spliceosomal DEAD-box helicase Prp28. *Acta crystallographica Section D, Biological crystallography* *70*, 1622-1630.

Moore, M.J., Query, C.C., and Sharp, P.A. (1993). Splicing of precursors to mRNA by the spliceosome. In *The RNA world : the nature of modern RNA suggests a prebiotic RNA world*, R.F. Gesteland, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 303-357.

Mozaffari-Jovin, S., Wandersleben, T., Santos, K.F., Will, C.L., Luhrmann, R., and Wahl, M.C. (2013). Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* *341*, 80-84.

Nelissen, R.L., Will, C.L., van Venrooij, W.J., and Lührmann, R. (1994). The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *The EMBO journal* *13*, 4113-4125.

Nilsen, T.W. (1994). RNA-RNA interactions in the spliceosome: unraveling the ties that bind. *Cell* *78*, 1-4.

O'Brien, K., Matlin, A.J., Lowell, A.M., and Moore, M.J. (2008). The biflavonoid isoginkgetin is a general inhibitor of Pre-mRNA splicing. *The Journal of biological chemistry* *283*, 33147-33154.

Ohi, M.D., Ren, L., Wall, J.S., Gould, K.L., and Walz, T. (2007). Structural characterization of the fission yeast U5.U2/U6 spliceosome complex. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 3195-3200.

Parker, R., Siliciano, P.G., and Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* *49*, 229-239.

Patel, A.A., and Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. *Nature reviews Molecular cell biology* *4*, 960-970.

Patil, V., Canzoneri, J.C., Samatov, T.R., Lührmann, R., and Oyelere, A.K. (2012). Molecular architecture of zinc chelating small molecules that inhibit spliceosome assembly at an early stage. *Rna* *18*, 1605-1611.

Pawellek, A., McElroy, S., Samatov, T., Mitchell, L., Woodland, A., Ryder, U., Gray, D., Lührmann, R., and Lamond, A.I. (2014). Identification of Small Molecule Inhibitors of Pre-mRNA Splicing. *The Journal of biological chemistry* *289*, 34683-34698.

Perriman, R.J., and Ares, M., Jr. (2007). Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes & development* *21*, 811-820.

Pyle, A.M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annual review of biophysics* *37*, 317-336.

- 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 & development* 8, 587-597.
- Raghunathan, P.L., and Guthrie, C. (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Current biology* : CB 8, 847-855.
- Rappsilber, J., Ajuh, P., Lamond, A.I., and Mann, M. (2001). SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome. *The Journal of biological chemistry* 276, 31142-31150.
- Reddy, R., and Busch, H. (1983). Small nuclear RNAs and RNA processing. *Progress in nucleic acid research and molecular biology* 30, 127-162.
- Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. *Current opinion in genetics & development* 6, 215-220.
- Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. *Current opinion in cell biology* 12, 340-345.
- Rinke, J., Appel, B., Digweed, M., and Lührmann, R. (1985). Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking. *Journal of molecular biology* 185, 721-731.
- Ritchie, D.B., Schellenberg, M.J., and MacMillan, A.M. (2009). Spliceosome structure: piece by piece. *Biochimica et biophysica acta* 1789, 624-633.
- Rocak, S., and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. *Nature reviews Molecular cell biology* 5, 232-241.
- Roybal, G.A., and Jurica, M.S. (2010). Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. *Nucleic acids research* 38, 6664-6672.
- Ruby, S.W., and Abelson, J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science* 242, 1028-1035.
- Sakharkar, M.K., Chow, V.T., and Kanguane, P. (2004). Distributions of exons and introns in the human genome. *In silico biology* 4, 387-393.
- Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S., and Seraphin, B. (1999). Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *The EMBO journal* 18, 3451-3462.
- Samatov, T.R., Wolf, A., Odenwalder, P., Bessonov, S., Deraeve, C., Bon, R.S., Waldmann, H., and Lührmann, R. (2012). Psoromic acid derivatives: a new family of small-molecule pre-mRNA splicing inhibitors discovered by a stage-specific high-throughput in vitro splicing assay. *Chembiochem : a European journal of chemical biology* 13, 640-644.
- Sander, B., Golas, M.M., Makarov, E.M., Brahms, H., Kastner, B., Lührmann, R., and Stark, H. (2006). Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Molecular cell* 24, 267-278.
- Sander, B., Golas, M.M., and Stark, H. (2003a). Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra. *J Struct Biol* 142, 392-401.

- Sander, B., Golas, M.M., and Stark, H. (2003b). Corrim-based alignment for improved speed in single-particle image processing. *Journal of structural biology* 143, 219-228.
- Sawa, H., and Abelson, J. (1992). Evidence for a base-pairing interaction between U6 small nuclear RNA and 5' splice site during the splicing reaction in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 89, 11269-11273.
- Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T., and Lührmann, R. (2004). RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *The EMBO journal* 23, 3000-3009.
- Schneider, M., Will, C.L., Anokhina, M., Tazi, J., Urlaub, H., and Lührmann, R. (2010). Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Molecular cell* 38, 223-235.
- Schwer, B., and Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *The EMBO journal* 11, 5033-5039.
- Seipelt, R.L., Zheng, B., Asuru, A., and Rymond, B.C. (1999). U1 snRNA is cleaved by RNase III and processed through an Sm site-dependent pathway. *Nucleic acids research* 27, 587-595.
- Seraphin, B. (1995). Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *The EMBO journal* 14, 2089-2098.
- Seraphin, B., Kretzner, L., and Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *The EMBO journal* 7, 2533-2538.
- Sharma, S., Kohlstaedt, L.A., Damianov, A., Rio, D.C., and Black, D.L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nature structural & molecular biology* 15, 183-191.
- Sharp, P.A. (1985). On the origin of RNA splicing and introns. *Cell* 42, 397-400.
- Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. *Cell* 9, 645-653.
- Shatkin, A.J., and Manley, J.L. (2000). The ends of the affair: capping and polyadenylation. *Nature structural biology* 7, 838-842.
- 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. *Molecular cell* 23, 819-829.
- Shomron, N., Malca, H., Vig, I., and Ast, G. (2002). Reversible inhibition of the second step of splicing suggests a possible role of zinc in the second step of splicing. *Nucleic acids research* 30, 4127-4137.
- Small, E.C., Leggett, S.R., Winans, A.A., and Staley, J.P. (2006). The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Molecular cell* 23, 389-399.
- Smith, C.W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in biochemical sciences* 25, 381-388.



- Smith, D.J., and Konarska, M.M. (2008). Mechanistic insights from reversible splicing catalysis. *Rna* 14, 1975-1978.
- Sontheimer, E.J., and Steitz, J.A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262, 1989-1996.
- Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *Rna* 5, 221-234.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315-326.
- Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Molecular cell* 3, 55-64.
- Stark, H., Dube, P., Lührmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* 409, 539-542.
- Stoilov, P., Lin, C.H., Damoiseaux, R., Nikolic, J., and Black, D.L. (2008). A high-throughput screening strategy identifies cardiotoxic steroids as alternative splicing modulators. *Proceedings of the National Academy of Sciences of the United States of America* 105, 11218-11223.
- Strauss, E.J., and Guthrie, C. (1994). PRP28, a 'DEAD-box' protein, is required for the first step of mRNA splicing in vitro. *Nucleic acids research* 22, 3187-3193.
- Sun, J.S., and Manley, J.L. (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes & development* 9, 843-854.
- Tazi, J., Kornstadt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C., and Lührmann, R. (1993). Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature* 363, 283-286.
- Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008). Crystal structure of a self-spliced group II intron. *Science* 320, 77-82.
- Tseng, C.K., and Cheng, S.C. (2008). Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science* 320, 1782-1784.
- Turner, I.A., Norman, C.M., Churcher, M.J., and Newman, A.J. (2004). Roles of the U5 snRNP in spliceosome dynamics and catalysis. *Biochemical Society transactions* 32, 928-931.
- Valadkhan, S., Mohammadi, A., Wachtel, C., and Manley, J.L. (2007). Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *Rna* 13, 2300-2311.
- van Heel, M. (1984). Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy* 13, 165-183.
- Van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* 21, 111-123.
- van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., *et al.* (2000). Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly reviews of biophysics* 33, 307-369.

- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* *136*, 701-718.
- Wang, C., Chua, K., Seghezzi, W., Lees, E., Gozani, O., and Reed, R. (1998). Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes & development* *12*, 1409-1414.
- 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.
- Warkocki, Z., Odenwalder, P., Schmitzova, J., Platzmann, F., Stark, H., Urlaub, H., Ficner, R., Fabrizio, P., and Lührmann, R. (2009). Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nature structural & molecular biology* *16*, 1237-1243.
- Webb, T.R., Joyner, A.S., and Potter, P.M. (2013). The development and application of small molecule modulators of SF3b as therapeutic agents for cancer. *Drug discovery today* *18*, 43-49.
- Will, C.L., and Lührmann, R. (1997). Protein functions in pre-mRNA splicing. *Current opinion in cell biology* *9*, 320-328.
- Will, C.L., and Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Current opinion in cell biology* *13*, 290-301.
- Will, C.L., and Lührmann, R. (2005). Splicing of a rare class of introns by the U12-dependent spliceosome. *Biological chemistry* *386*, 713-724.
- Will, C.L., and Lührmann, R. (2006). Spliceosome structure and function. In *The RNA world : the nature of modern RNA suggests a prebiotic RNA world*, R.F. Gesteland, T. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press), pp. pp. 369-400.
- Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* *3*.
- Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *The EMBO journal* *21*, 4978-4988.
- Wolf, E., Kastner, B., Deckert, J., Merz, C., Stark, H., and Lührmann, R. (2009). Exon, intron and splice site locations in the spliceosomal B complex. *The EMBO journal* *28*, 2283-2292.
- Wu, J., and Manley, J.L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes & development* *3*, 1553-1561.
- Wyatt, J.R., Sontheimer, E.J., and Steitz, J.A. (1992). Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes & development* *6*, 2542-2553.
- Xu, Y.Z., and Query, C.C. (2007). Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Molecular cell* *28*, 838-849.
- Yean, S.L., Wuenschell, G., Termini, J., and Lin, R.J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* *408*, 881-884.

Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., Sugi, N.H., Sagane, K., Hamaguchi, M., Iwata, M., *et al.* (2011). Biological validation that SF3b is a target of the antitumor macrolide pladienolide. *The FEBS journal* 278, 4870-4880.

Yu, Y.T., Shu, M.D., and Steitz, J.A. (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *The EMBO journal* 17, 5783-5795.

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.

Zhang, M., and Green, M.R. (2001). Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes & development* 15, 30-35.

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

# 6 Appendix

## 6.1 List of abbreviations

---

<b>Abbreviation</b>	
3'SS	3' splice site
5'SS	5' splice site
°C	degree Celsius
$\alpha$	anti
$\beta$ -ME	2-mercaptoethanol
A	adenosine
APS	ammonium peroxodisulfate
ATP	adenosine triphosphate
bp	base pair
BP	branch point
BPS	branch point sequence
BSA	bovine serum albumin
CTP	cytosine triphosphate
ca.	circa
Ci	Curie
cpm	counts per minute
cDNA	complementary DNA
d	desoxy
Da	Dalton
DExD/H	consensus sequence of helicases
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	1,4-Dithiothreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
EJC	exon junction complex
EM	electron microscopy
ESE	exonic splicing enhancer
ESI	exonic splicing inhibitor
EtOH	ethanol
et al.	et alii
f	femto

---

**Abbreviation**

---

GTP	guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hn	heterogenous nuclear
i.e.	<i>id est</i>
ISE	intronic splicing enhancer
ISI	intronic splicing inhibitor
ISL	intramolecular stem-loop
k	kilo
kDa	kilodalton
LSm	Sm-like protein
M	molar
m	milli
μ	micro
m <sub>3</sub> <sup>2,2,7</sup> G	2,2,7-trimethylguanosine
m <sub>7</sub> G	7-methylguanosine
min	minutes
mM	millimolar
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein
n	nano
nt	nucleotides
NTC	NineTeen Complex
NTP	nucleotide triphosphate
ORF	open reading frame
p	pico
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCI	phenol chloroform isoamylalkohol
PCR	polymerase chain reaction
pH	power of hydrogen
PMSF	phenylmethylsulfonyl fluoride
PNK	polynucleotide kinase
PPIase	peptidyl-prolyl isomerase
PPT	polypyrimidine tract
pre-mRNA	precursor mRNA
Prp	pre-mRNA processing

---

**Abbreviation**

---

RES complex	pre-mRNA retention and splicing complex
RNA	ribonucleic acid
RRM	RNA recognition motif
RNase	ribonuclease
RNasin	ribonuclease-inhibitor
RNP	ribonucleoprotein
rpm	revolutions per minute
RRM	RNA recognition motif
RS domain	arginine/serine rich domain
RT	room temperature
S	Svedberg unit ( $10^{-13}$ s)
s	second
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	sodium dodecylsulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SL	stem loop
sn	small nuclear
SR	serine-arginine rich
SS	splice site
TBE buffer	TRIS-borate-EDTA buffer
TBS	TRIS-buffered saline
TEMED	N,N,N',N'-tetraethylenediamide
TRIS	Tris-(hydroxymethyl)-aminomethane
tRNA	transfer RNA
U	unit
U snRNA	uridine-rich small nuclear RNA
U2AF	U2 auxiliary factor
UTP	uridine triphosphate
UTR	untranslated region
UV	ultraviolet
V	volt
v/v	volume per volume
w/v	weight per volume
Y	pyrimidine base

## 6.2 List of figures

Figure	Title
Figure 1.1	Consensus sequences of pre-mRNA in yeast and human.
Figure 1.2	Schematic view of the two-step mechanism of pre-mRNA splicing.
Figure 1.3	Sequences and predicted secondary structures of the human spliceosomal snRNAs.
Figure 1.4	Composition of the major spliceosomal U snRNPs.
Figure 1.5	The spliceosome cycle.
Figure 1.6	Cross-exon definition.
Figure 1.7	Dynamic RNA-RNA network in the spliceosome.
Figure 1.8	Compositional dynamics of the human spliceosomal complexes B, B <sup>act</sup> , and C.
Figure 1.9	The currently available EM 3D structures of the spliceosome and spliceosomal components.
Figure 2.1	Schematic of high throughput <i>in vitro</i> screening procedure for inhibitors of pre-mRNA splicing.
Figure 3.1	Schematic of the high throughput <i>in vitro</i> screening assay.
Figure 3.2	Analysis of <i>in vitro</i> splicing performed in the presence of different compounds at a concentration of 50 $\mu$ M.
Figure 3.3	Analysis of <i>in vitro</i> splicing performed in the presence of a subset of hit compounds from screen nr. 2 and 3 of the COMAS library at the concentration of 50 $\mu$ M.
Figure 3.4	Effects of increasing compound concentrations on the splicing of <sup>32</sup> P-labeled MINX-MS2 pre-mRNA <i>in vitro</i> .
Figure 3.5	Effects of increasing compound concentrations on the splicing of <sup>32</sup> P-labeled MINX-MS2 pre-mRNA <i>in vitro</i> .
Figure 3.6	Effects of increasing compound concentrations on the splicing of <sup>32</sup> P-labeled MINX-MS2 pre-mRNA <i>in vitro</i> .
Figure 3.7	Effects of increasing compound concentrations on the splicing of <sup>32</sup> P-labeled MINX-MS2 pre-mRNA <i>in vitro</i> .
Figure 3.8	Kinetics of splicing complex formation with <sup>32</sup> P-labeled MINX-MS2 pre-mRNA in the presence of the indicated concentrations of various compounds.
Figure 3.9	Glycerol gradient sedimentation profiles of splicing complexes formed in the presence of compounds 028 (150 $\mu$ M) and 120 (75 $\mu$ M).
Figure 3.10	Purification of complexes stalled with compound 120.
Figure 3.11	RNA composition of MS2-affinity purified B/B-like splicing complexes formed in the presence of compounds 296 and 297.
Figure 3.12	Purification of splicing complexes stalled with compound 028.
Figure 3.13	Western blot analyses of the purified B, B <sup>028</sup> and B <sup>act</sup> complexes.
Figure 3.14	RNA composition of MS2-affinity purified B <sup>028</sup> complexes formed in the presence of compound 028 at different time points of the splicing reaction.

<b>Figure</b>	<b>Title</b>
Figure 3.15	Purification of stalled B <sup>028</sup> complexes and characterization of different fractions of the B-like peak.
Figure 3.16	Dynamic network of RNA-RNA interactions in the spliceosome.
Figure 3.17	RNA composition of MS2 affinity-purified splicing complexes B, B <sup>028</sup> and B <sup>act</sup> formed on PM5-10 and used for structure probing studies.
Figure 3.18	Identification of RNA-RNA interactions via psoralen crosslinking in purified complexes B, B <sup>028</sup> and B <sup>act</sup> .
Figure 3.19	Positions of chemical modifications by chemical reagents DMS, CMCT and Kethoxal and schematic view of primer extension using modified RNA as a template.
Figure 3.20	U5 snRNA protection pattern in spliceosomal complexes is identical in B <sup>028</sup> and B <sup>act</sup> complexes.
Figure 3.21	Primer extension analysis of U6 snRNAs in chemically modified B, B <sup>028</sup> and B <sup>act</sup> complexes.
Figure 3.22	U2 snRNA chemical modification pattern in B, B <sup>act</sup> and B <sup>028</sup> complexes.
Figure 3.23	Pre-mRNA chemical modification pattern in B, B <sup>act</sup> and B <sup>028</sup> complexes.
Figure 3.24	U2 snRNA protection pattern in spliceosomal complexes reveals that U2 helix IIc is not formed in the B <sup>028</sup> complex.
Figure 3.25	B <sup>028</sup> complexes can be chased into catalytically active spliceosomes in the presence HeLa nuclear extract treated with micrococcal nuclease.
Figure 3.26	RNA composition of the purified B <sup>028</sup> -containing fractions compared to B complexes formed in the presence of DMSO.
Figure 3.27	Electron microscopy of negatively-stained B <sup>028</sup> particles assembled on the MINX pre-mRNA substrate.
Figure 3.28	B <sup>028</sup> particles exhibit a morphology between B and B <sup>act</sup> .
Figure 3.29	The morphology of the B <sup>028</sup> complex assembled on the MINX pre-mRNA.
Figure 3.30	The mushroom-like view of the B <sup>028</sup> complex tilted forward on its horizontal axis.
Figure 3.31	Schematic view of MS2-affinity purification of human spliceosomal B <sup>028</sup> complexes with subsequent chasing of isolated complexes with normal or MN-treated nuclear extract under splicing conditions.
Figure 3.32	RNA composition of B <sup>028</sup> complexes after chasing with normal or MN-treated nuclear extract.
Figure 3.33	The U4/U6.U5 tri-snRNP is stable in the presence of compound 028.
Figure 3.34	U4/U6.U5 tri-snRNPs appear to be slightly less salt stable in the presence of compound 028.
Figure 3.35	The Prp19/CDC5L complex is affected by the presence of compound 028.
Figure 3.36	The sedimentation of the Prp19/CDC5L complex is not affected by compound 028.
Figure 3.37	Analysis of <i>in vitro</i> splicing performed in the presence of various derivatives of compound 028 at a concentration of 50 $\mu$ M.



Figure	Title
Figure 3.38	Effect of increasing concentrations of selected analogues of compound 028 on the splicing of <sup>32</sup> P-labeled MINX-MS2 pre-mRNA.
Figure 3.39	Effects of various analogues of compound 028 on the splicing of <sup>32</sup> P-labeled MINX-MS2 pre-mRNA in HeLa nuclear extract after 60 min.
Figure 3.40	Effect of increasing concentrations of selected synthesized analogues of compound 028 on the splicing of <sup>32</sup> P-labeled MINX-MS2 pre-mRNA.
Figure 4.1	Summary of selected proteins identified by the mass spectrometry in the affinity-purified spliceosomal complexes B, B <sup>028</sup> and B <sup>act</sup> .
Figure 4.2	Comparison of 2D structures of spliceosomal complexes B, B <sup>act</sup> and intermediate B <sup>028</sup> .
Figure 4.3	Comparison of the 3D structures of B <sup>028</sup> and human BAU1.

### 6.3 List of tables

Table	Title
Table 3.1	Summary of IC <sub>50</sub> values determined for small molecules confirmed to be inhibitors of pre-mRNA splicing <i>in vitro</i> and the stage at which spliceosome assembly is inhibited.
Table 3.2	Proteins identified by LC-MS/MS in human spliceosomal complexes A, B and complexes stalled in the presence of compounds 120 (B <sup>120</sup> ), 296 (B <sup>296</sup> ) and 297 (B <sup>297</sup> ).
Table 3.3	Proteins identified by LC-MS/MS in human spliceosomal complexes B, B <sup>act</sup> and complexes stalled in the presence of compound 028 (B <sup>028</sup> ).
Table 3.4	Proteins identified by LC-MS/MS in human spliceosomal complexes B, B <sup>act</sup> , B <sup>028</sup> and in B <sup>028</sup> chased in the presence of MN-treated HeLa extract (MNnx) with or without ATP or normal extract plus ATP.
Table 3.5	Summary of modifications to compound 028 and their effect on the inhibition of pre-mRNA splicing <i>in vitro</i> at 150 μM.
Table 3.6	Chemical structures of synthesized analogues of compound 028.

## 6.4 Acknowledgements

“All you need is love...”

John Lennon

I would like to express my deep gratitude to Prof. Dr. Reinhard Lührmann for providing an opportunity to work on this interesting project as well as for his support, valuable suggestions and fruitful and stimulating discussions.

I am very grateful to Dr. Cindy Will for being a great supervisor and sharing her scientific experience with me, for her continuous support and encouragement, numerous ideas, helpful comments concerning my thesis.

I would like to thank the members of my thesis committee, Prof. Dr. Holger Stark and Prof. Dr. Heike Krebber for their helpful suggestions and interest in my results.

I'm also grateful to Holger Stark for the contribution in 3D EM studies. Many thanks go to Dr. Berthold Kastner and Dr. Prakash Dube for their help with the electron microscopy analysis. I also thank Berthold Kastner for critical reading of EM part of my thesis.

I would like to thank Prof. Dr. Waldmann, Dr. Sonja Sievers and Javier Ceballos from MPI for molecular physiology in Dortmund for the very productive collaboration, impressive assistance with screening procedure and synthesis of all necessary chemicals.

I would like to thank Dr. Henning Urlaub, Monika Raabe and Annika Kühn for performing mass spectrometric analysis so necessary for my project.

I would like to express my thanks to Dr. Maria Anokhina for the excellent and very pleasant cooperation within the RNA studies and Dr. Sergey Bessonov for the advices as to splicing assays.

I express my gratitude to Dr. Partizia Fabrizio and Dr. Reinhard Rauhut for their kind support and helpful advices.

I'm grateful to Dr. Kum-Loong Boon for providing antibodies necessary for my investigations and Dr. Cyrille Girard for the interesting experiment *in vivo*.

I am thankful to Dr. Klaus Hartmuth for the opportunity to study different methodological aspects of RNA investigation at his practical courses, as well as for very interesting conversations on multiple subjects.

I express my deep gratitude to Gabi Heyne for all the welcoming support, care, attention and continuous help in daily lab issues.

I appreciate the help of Juliane Moses for solving a multitude of everyday problems. I am also thankful to Irene Öchsner, Thomas Conrad and Hossein Kohansal for the excellent technical support.

My appreciations go to all members of the Department of Cellular Biochemistry for the amazing cooperation and interesting communication. In particular, I would like to thank the members of the lab no. 118 Gabi, Cindy, Masha, Olex and Elke for all friendly and always exciting scientific and non-scientific discussions.

And last but not least: My sincere gratitude goes to my husband Sascha for his support, love and patience during my PhD time; to my loving parents Tadeush and Irina, my charming sister Inessa, and of course, my adorable godchildren Viktoriya and Daniel for their delightful support, comfort, powerful encouragement, splendid disturbances and endless love.

## 6.5 Curriculum vitae

<b>Personal information</b>	
Name:	Anzhalika Sidarovich
Date of birth:	26.07.1975
Place of birth:	Soligorsk, Belarus
Citizenship:	Belarus
<b>Education and Employment</b>	
1982 – 1992	Secondary school no. 3, Soligorsk, Belarus
1992 – 1997	Diploma in Education Science and Psychology and practical Psychology, State Pedagogical University, Minsk, Belarus
1997 – 2000	Psychologist in preschool place in Soligorsk, Belarus
2001 – 2003	Editor in Publishing House „Narodny Sud“, Minsk, Belarus
2003 - 2004	Study German language in VHS, Göttingen
2004 - 2010	<p>Diploma in biology, Georg-August-University Göttingen</p> <p>Major subject: Biochemistry</p> <ol style="list-style-type: none"> <li>1. Minor subject: Microbiology</li> <li>2. Minor subject: Organic Chemistry</li> </ol> <p>Oral examination: Juli 2009</p> <p>Diploma thesis: “Overexpression of SF3a and SF3b protein complexes for functional and structural studies” at the Department “Cellular Biochemistry”, Max-Plank-Institute for biophysical chemistry, Göttingen</p> <p>Supervisor: Prof. Dr. Reinhard Lührmann</p>
2010 - 2015	<p>PhD project at the department “Cellular Biochemistry”, Max-Plank-Institute for biophysical chemistry, Göttingen, Germany</p> <p>Title: “Identification and characterization of small molecule inhibitor of pre-mRNA splicing that block spliceosome assembly at novel stages”</p> <p>Supervisor: Prof. Dr. Reinhard Lührmann</p>

## 6.6 Conferences

8<sup>th</sup> International PhD Student Symposium “Horizons in Molecular Biology”, 2011, Göttingen, Germany

9<sup>th</sup> International PhD Student Symposium “Horizons in Molecular Biology”, 2012, Göttingen, Germany

18<sup>th</sup> Annual Meeting of the RNA Society, 2013, Davos, Switzerland

Poster presented: “Identification of small molecule pre-mRNA splicing inhibitors using a stage-specific high-throughput in vitro splicing assay”

10<sup>th</sup> International PhD Student Symposium “Horizons in Molecular Biology”, 2013, Göttingen, Germany

Poster presented: “Stalling of pre-mRNA splicing at the spliceosome activation stage by a newly identified small molecule inhibitor”

EMBO/EMBL Symposium: Molecular Machines – Lessons from Integrating Structure, Biophysics and Chemistry, 2014, Heidelberg, Germany

Poster presented: “Stalling of pre-mRNA splicing at the spliceosome activation stage by a newly identified small molecule inhibitor”