Smu1 and RED play an important role for the activation of human spliceosomes

Dissertation

for the award of the degree

„Doctor rerum naturalium"

of the Georg-August-Universität Göttingen

within the doctoral program "Molecular Biology of Cells"

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

submitted by

Sandra Maria Keiper

from Berlin

Göttingen, 2018

Members of the Examination Board:

Prof. Dr. Reinhard Lührmann (1st Referee)

       Department of Cellular Biochemistry

       Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Heike Krebber (2nd Referee)

       Department of Molecular Genetics

       Georg-August-Universität Göttingen

Prof. Dr. Patrick Cramer

       Department of Molecular Biology

       Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Markus Zweckstetter

       Department for NMR-based Structural Biology

       Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Jörg Großhans

       Department of Developmental Biochemistry

       University Medical Center Göttingen

Prof. Dr. Ralph Kehlenbach

       Department of Molecular Biology

       University Medical Center Göttingen

Date of oral examination: September 27th 2018

For Oskar

Table of Contents

Abstract

In eukaryotes, pre-mRNA splicing is catalysed by the spliceosome, a highly complex and dynamic molecular machine, which assembles stepwise by the sequential recruitment of five small nuclear ribonucleoproteins (snRNPs) and numerous non-snRNP proteins. In humans, formation of the pre-catalytic, spliceosomal B complex with a stably associated tri-snRNP requires the action of the helicase Prp28, which displaces the U1 snRNP from the 5' splice site to be replaced by the U6 snRNA. At the same time, the so-called, B-specific proteins are recruited. Transformation of the B complex into an activated (B^{act}) spliceosome is initially triggered by the action of the helicase Brr2 that unwinds the U4/U6 RNA duplex. This results in the crucial release of the U4 snRNP, which allows the U6 snRNA to interact with the U2 snRNA forming essential components of the catalytic centre. Concomitantly, the B-specific proteins dissociate in the course of the activation, while the B^{act}-specific proteins, the Prp19/CDC5L complex and the intron binding complex (IBC) are stably integrated.

Smu1 and RED are two B-specific proteins that are conserved among higher eukaryotes but absent in *S. cerevisiae*. Currently, little is known about the precise function of these proteins. Their transient association with the B complex suggests that they may be involved in the B-to-B^{act} complex transition, but previous studies proposed that these proteins form a functional module that is involved in the regulation of alternative splicing. However, it is not clear how Smu1 and RED might contribute to this process. Furthermore, it has not been investigated whether these proteins participate in constitutive splicing and thus it remains possible that Smu1 and RED – like several other B-specific proteins – are in general important for intron excision in higher eukaryotes. To provide clarity on this subject, I analysed the function of Smu1 and RED in pre-mRNA splicing both *in vivo* and *in vitro*.

To address whether constitutive splicing is dependent on Smu1 and RED, an RNAseq analysis was performed with HeLa cells siRNA-depleted of Smu1 or RED. Knock-down of these two proteins resulted in profound changes in alternative splicing patterns and also led to the retention of constitutively spliced introns, suggesting that Smu1 and RED are important for splicing in general, and thus not only involved in the regulation of alternative splicing. A role for Smu1 and RED in constitutive splicing was also demonstrated *in vitro*, using HeLa nuclear extract that was immunodepleted of Smu1 and RED. By using a well-functioning Smu1-specific antibody, Smu1 was nearly quantitatively removed from the extract along with more than 90% of RED, suggesting that the majority of human Smu1 and RED exist as a dimer in HeLa cell extract. Splicing of MINX-120 was less efficient in the absence of Smu1/RED, with an apparent slow-down in the rate of mRNA production. Investigation of spliceosome assembly revealed a transient accumulation of spliceosomal B complexes in the Smu1/RED-depleted extract, while the subsequently formed B^{act} and C complexes were still formed but at a slower rate. These results indicate that defects in splicing triggered by the absence of Smu1 and RED were caused by impaired spliceosome activation.

To investigate whether Smu1 and RED need to interact to fulfil their function, spliceosome assembly in the presence of the individually expressed proteins was investigated. As the addition of single proteins to the Smu1/RED-depleted extract did not restore the B-to-B^{act} transition and splicing product formation, Smu1 and RED do not appear to function on their own. This was attributable to poor or less stable binding to the spliceosome of the individual

Abstract

proteins compared to the Smu1/RED dimer. Removal of the WD40 domain of Smu1 abolished binding of the dimer to the spliceosome, and spliceosome assembly remained blocked at B complex level, demonstrating that the WD40 domain is essential for proper interaction of the dimer with the spliceosome. Truncation of RED's N- and C-terminal regions, which contact U2 or tri-snRNP proteins, respectively, within the B complex, restored splicing and spliceosome activation partially or nearly fully, suggesting that the contacts RED establishes with U2 or U5 individually are not essential for the function of Smu1/RED. To determine whether Smu1 and RED function as a binding platform for other spliceosomal factors, I purified B complexes that accumulate in their absence and investigated their composition. As no additional proteins were missing, Smu1 and RED appear to play a direct role in splicing, as opposed to aiding the binding of other factors required for spliceosome activation. Thus, these results indicate that Smu1 and RED themselves are important for efficient conversion of the B complex into B^{act}.

To provide evidence that Smu1 and RED are also involved in spliceosome activation *in vivo*, I knocked-down these proteins in HeLa cells and investigated endogenous spliceosome assembly by immunoblotting, using antibodies that recognise phosphopeptides specifically associated with assembled B or B^{act} complexes. Knock-down of Smu1 or RED led to an increased B complex signal and a decreased B^{act} complex signal, indicating that the activation of the spliceosome is also impaired *in vivo* in the absence of Smu1 and RED.

While knock-down of Smu1 and RED affected introns of all sizes, the vast majority were shorter than 100 nt. Short introns constitute only a small fraction of introns in the human genome, and thus this result indicates that the splicing of very short introns *in vivo* is highly dependent on the presence of Smu1 and RED. *In vitro* splicing studies using truncated versions of the MINX-120 pre-mRNA and Smu1/RED-depleted extract, also demonstrated that splicing was more dependent on the presence of Smu1 and RED when intron length was shorter. Truncation of the intron to either 90 or 80 nt reduced the overall efficiency of splicing compared to MINX-120, but it also enhanced the inhibitory effect of Smu1/RED-depletion. While MINX-90 was spliced somewhat less efficiently than MINX-120, splicing of MINX-80 was nearly abolished in the absence of Smu1 and RED, and led to an apparent block at the B complex stage with little or no formation of B^{act} or catalytically-active C complexes. Thus, Smu1 and RED play a crucial role in the splicing of extremely short introns both *in vitro* and *in vivo.* To investigate whether the intron length *per se* or the distance between the 5'SS and the BS or between the BS and the 3'SS determines whether splicing is dependent on Smu1/RED or not, I compared spliceosome assembly on PM5 pre-mRNAs with a shortened 5'SS-BS distance or a shortened polypyrimidine (PY) tract (BS-3'SS distance).
While spliceosome activation was not affected by the truncation of the PY tract in the absence of Smu1 and RED, shortening of the 5'SS-BS distance to ~55nt (as found in MINX-80) blocked the assembly at the B complex level, indicating that the distance between the 5'SS and the BS is the decisive factor for a strong Smu1/RED-dependence.

The dependence of spliceosome activation on a minimal 5'SS-BS distance is probably due to physical limitations exerted by the intron. In this case, resolving this steric hindrance should allow spliceosome activation even without the support of Smu1 and RED. Indeed, when the MINX-80 pre-mRNA was cleaved into two RNAs, spliceosome activation was restored in the absence of Smu1/RED. This supports the idea that a short 5'SS-BS distance exerts a structural constraint, which can be relieved by physically separating the 5' and 3' regions of an intron.

2

Abstract

Based on my data and structural information obtained from the cryo-EM structure of the human B complex, a model for how 5'SS-BS distance might lead to the dependency of spliceosome activation on Smu1/RED was generated. In the human B complex, Smu1 and RED form a molecular bridge between the U2 snRNP protein SF3B3 and the RNA helicase Brr2. This interaction appears to be important for spliceosome activation, potentially either by tethering Brr2 in a position required to unwind the U4/U6 interaction during activation or by directly aiding in triggering Brr2 unwinding activity. In the B complex, the intron is base paired with the U6 ACAGAG box at the 5'SS and with the U2 snRNA at the BS region. In most cases the distance between the 5'SS and the BS of the intron is sufficiently long to grant the U2 domain enough flexibility to move towards Brr2 and to form a bridge directly, even when Smu1 and RED are absent, albeit it at a slower rate. However, when this distance is short, i.e. ~56 nt or less, the intron adopts a fully extended conformation that results in a structural constraint. This could potentially result in repositioning of the U2 domain away from Brr2 and/or hinder its ability to move towards Brr2, and in turn inhibit spliceosome activation. As Smu1 and RED extend the U2/Brr2 bridge, the negative effect of a short 5'SS to BS distance would be greatly enhanced in their absence. Taken together, my studies provide novel insights into the function of the B-specific proteins Smu1 and RED in splicing. They additionally elucidate how intron architecture impacts spliceosome assembly and splicing, and how spliceosomal proteins potentially help the splicing machinery to overcome the challenges created by short introns.

Introduction

1. Introduction

It has been more than 70 years since deoxyribonucleic acid (DNA), was identified as the molecule in which the genetic information of an organism is stored (Avery *et al.*, 1944; Watson & Crick, 1953). The conversion of this information into a functional protein is a laborious process and increases in complexity with increasing complexity of the organism. In simple prokaryotic organisms, a gene consists of a continuous DNA sequence that is transcribed in its entirety and is directly translated into a protein (Jacob & Monod, 1961). In more complex eukaryotic organisms, a gene that is transcribed in the nucleus by the RNA polymerase II needs to undergo several processing/maturation events before it can be exported to the cytoplasm, where it is translated into a protein by the ribosomal machinery (Clawson *et al.*, 1985).

The processing of the primary transcript, also called precursor messenger RNA (pre-mRNA), already begins during its transcription. Right after the appearance of its 5' end, a modified nucleotide, called N7-methylguanosine (m^7G), is attached to the 5' ribose by a rare 5'-to-5' triphosphate linkage (Shatkin, 1976; Chiu *et al.*, 2002). Towards the end of transcription, but before termination, the transcript is cleaved at the 3' end, and this is followed by the synthesis of a poly(A) tail comprising up to 200 adenosine residues. Both modifications are important for the transcript's stability and further promote its export to the cytoplasm and its productive translation (Shatkin & Manley, 2000; Rosonina *et al.*, 2003; Nagaike *et al.*, 2011).

In eukaryotic genes, especially of higher eukaryotes, protein-coding sequences termed exons are often disrupted by non-coding sequences termed introns, which are still included in the primary transcript. This demands an additional maturation step called splicing; in this step, intronic sequences are excised and the remaining exons are ligated together (Berk, 2016). The spliceosome, a highly complex and dynamic molecular machine, catalyses the splicing reaction and ensures its accuracy.

1.1 Basic structure of pre-mRNAs

The first demanding task of the spliceosome is the definition of the intron-exon boundaries. Since defects in splicing can have dramatic consequences for the cell and can ultimately cause severe diseases such as cancer or neuro-degenerative diseases in humans, the exact nucleotide identification of an intron's ends is a crucial process (Cooper *et al.*, 2009; Webb *et al.*, 2013; Scotti & Swanson, 2016). In eukaryotes, including yeast and humans, introns are flanked by short conserved sequences called splice sites, which define the boundaries of the intron and are important motifs for intron/exon recognition by the spliceosome (Fig. 1.1).

The 5' splice site (5'SS) sequence marks the boundary between an intron and its upstream exon. While this sequence (AG|GUAUGU, where '|' indicates the exon/intron boundary) is highly conserved in yeast (Lopez & Séraphin, 1999), it is much more variable in human cells (AG|GURAGU, where R = purine [G or A]) (M. Q. Zhang, 1998). The boundary between an intron and its downstream exon is determined by the 3' splice site (3'SS). In yeast and human, this consists of the characteristic sequence YAG (Y = pyrimidine [C or U]) (M. Q. Zhang, 1998; Lopez & Séraphin, 1999). Another essential element is the branch-point sequence (BPS) YUR<u>A</u>Y (human) (M. Q. Zhang,

1998) or UACUA<u>A</u>C (yeast) (Lopez & Séraphin, 1999; Spingola *et al.*, 1999). This motif is localized 18–40 nt or 10–40 nt respectively, upstream of the 3'SS. Most human introns further contain a polypyrimidine (PY) tract. This pyrimidine-rich region is located between the branch site (BS) and the 3'SS and helps to recruit and stabilize the interaction of spliceosomal factors with the pre-mRNA and is an important splicing motif in humans (Coolidge *et al.*, 1997).
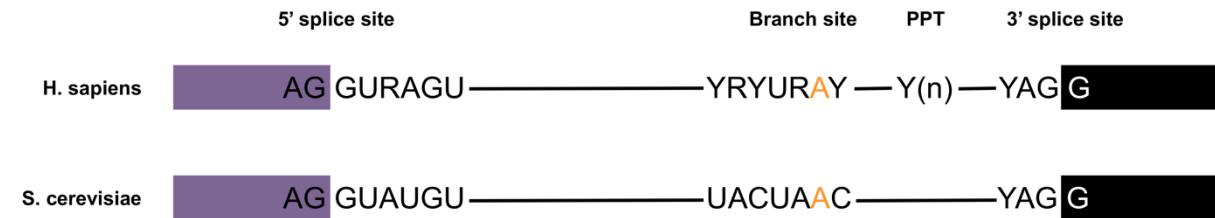
Figure 1.1. Conserved splicing sequences of pre-mRNA from *H. sapiens* and *S. cerevisiae*. Schematic representation of a pre-mRNA comprising one intron (black line) surrounded by two exons (purple and black boxes). The conserved intron defining consensus sequences are specified in black and white letters, while the branch-point adenosine is shown in orange. Y(n) indicates the polypyrimidine tract. (Y: pyrimidines (C or T), R: purines (A or G). (Will & Lührmann, 2011)

1.2 Pre-mRNA splicing mechanism

The excision of introns from a pre-mRNA includes two sequential S_N2-type transesterification reactions (Fig. 1.2). During the first reaction, the 2' hydroxyl group of the branch-point adenosine attacks nucleophilically the phosphodiester bond of the 5'SS. This results in the liberation of the 5' exon and formation of the 3' exon-intron lariat intermediate, in which the branch-point adenosine and the guanosine at the intron's 5' end are connected by a 2'–5' phosphodiester bond. The second splicing reaction involves the free 3' hydroxyl group of the 5' exon, which performs a nucleophilic attack on the phosphodiester bond at the 3'SS, resulting in exon ligation and the release of the intron lariat (Moore & Sharp, 1993).
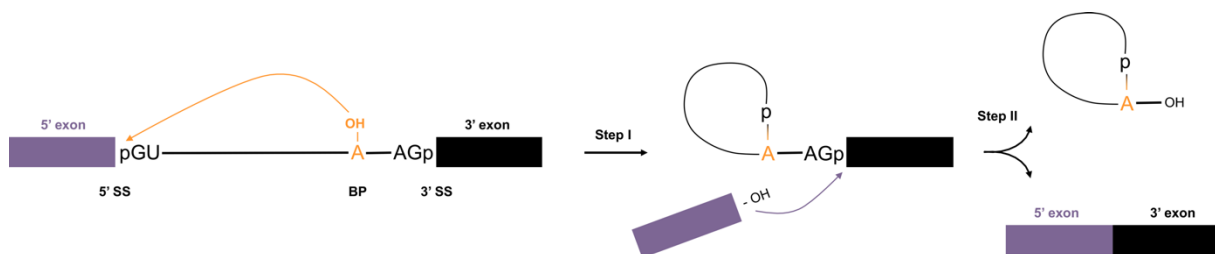
Figure 1.2. Chemistry of the splicing reaction. Schematic illustration of the two sequential transesterification reactions. The first step comprises nucleophilic attack by the 2' OH group of the branch-point adenosine on the phoshodiester bond at the 5' splice site, leading to the free 5' exon and intron-3' exon intermediates. During the second step the free 3' OH group of the 5' exon targets the phosphodiester bond at the 3'SS, resulting in the ligation of the two exons and the release of the intron lariat. Phosphate groups involved in the reactions are indicated by "p". (Will & Lührmann, 2011)

Introduction

The mechanism of pre-mRNA splicing resembles that of group II self-splicing introns. Both exhibit similar consensus sequences at the splice sites and branch point and form identical intermediate and end products in two transesterification reactions. Since pre-mRNA introns are unable to form catalytically active structures, as group II introns do, their removal depends on the spliceosomal machinery. Therefore, group II intron splicing is ATP-independent, while the spliceosome-dependent intron excision consumes energy during spliceosome assembly and the catalytically relevant rearrangements (Seetharaman *et al.*, 2006; Lambowitz & Zimmerly, 2011).

1.3 Trans-splicing

As described above, splicing commonly occurs within one RNA molecule and can thus be considered as cis-splicing. However, in some organisms, including trypanosomes, nematodes and plants, a rather unusual form of RNA splicing exists, called trans-splicing, in which splicing can occur across two distinct pre-mRNA transcripts (Fig. 1.3).
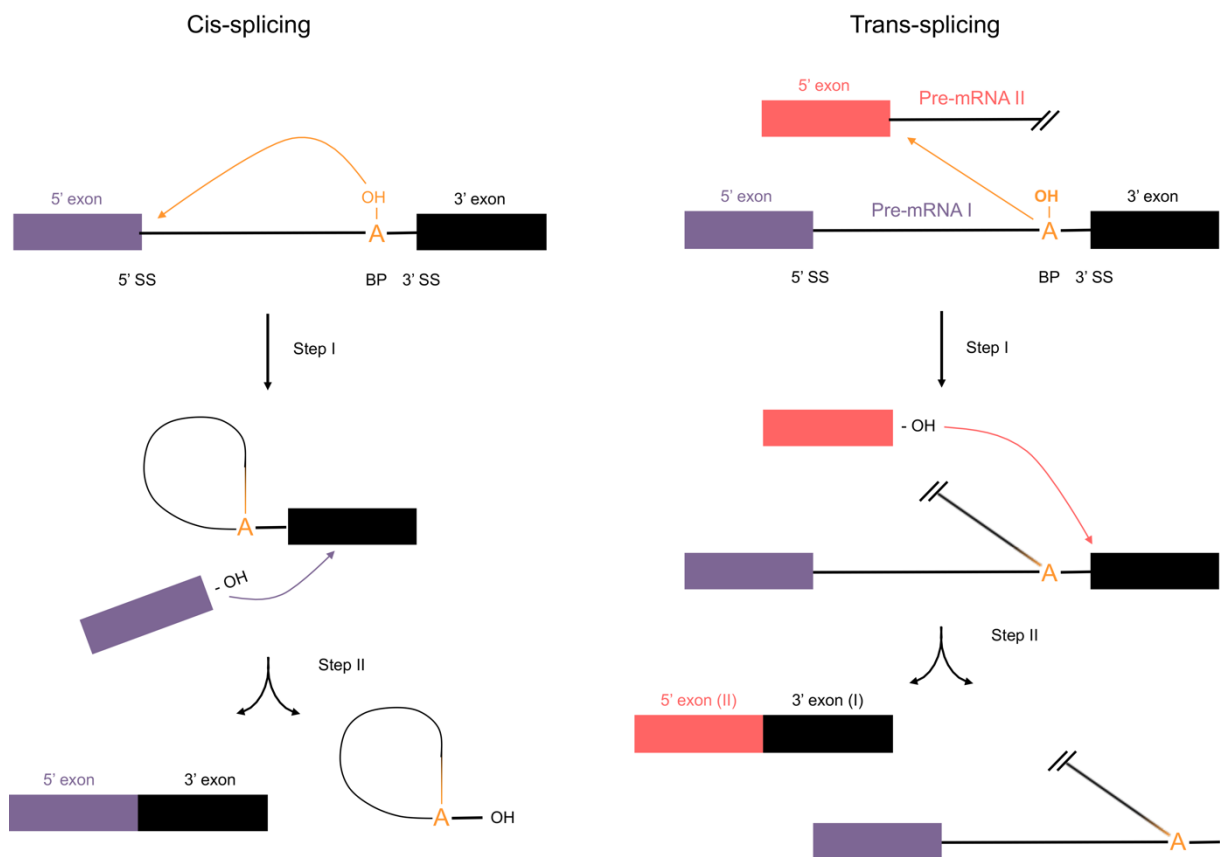
Figure 1.3. Cis-splicing *versus* trans-splicing. Schematic of the cis-splicing and trans-splicing reactions, both of which comprise two transesterification reactions. While cis-splicing involves only splicing elements from a single transcript, trans-splicing engages splicing signals from two distinct pre-mRNAs, each of which contributes an exon to the mature mRNA. The figure is adapted from (Garcia-Blanco, 2003).

During this process, the branch point adenosine of the first pre-mRNA (PI) attacks the 5'SS of another pre-mRNA (PII), which liberates the 5' exon of PII and ligates its downstream intron to PI. By attacking the 3'SS of PI, the free 5' exon (PII) is ligated to the 3' exon of PI, combining two exons derived from different transcripts.

Although trans-splicing has recently been described in humans, it appears to be rather seldom. It is unclear why cis-splicing is favored over trans-splicing, but it has been proposed that coupling of splicing to transcription might spatially restrict contacts between different transcripts. However, the mechanism underlying trans-splicing in mammalian cells is poorly understood. (Konarska *et al.*, 1985; Caudevilla *et al.*, 1998; Garcia-Blanco, 2003).

1.4 snRNPs – the major building blocks of the spliceosome

The spliceosome is a highly complex molecular machine with a remarkably dynamic structure and composition. The main components of the spliceosome are the small nuclear ribonucleoproteins (snRNPs) U1, U2 and U5 and a U4/U6 di-snRNP (Will & Lührmann, 2011).

1.4.1 The small nuclear RNAs

Each snRNP of the major spliceosome consists of a uridine-rich small nuclear RNA, called snRNA (in the case of the U4/U6 di-snRNP, two snRNAs), that are associated with a defined set of proteins (Lerner & Steitz, 1979; Bringmann *et al.*, 1984; Hashimoto & Steitz, 1984). Apart from the U6 snRNA, all snRNAs are transcribed by the RNA polymerase II and thus are equipped with a 5' terminal m^7G cap during transcription (Chandrasekharappa *et al.*, 1983; Kunkel *et al.*, 1986). The m^7G cap not only stabilizes the snRNAs (Furuichi *et al.*, 1977), but also acts as a nuclear export signal (Hamm & Mattaj, 1990). During maturation of the snRNA in the cytoplasm the survival of motor neuron protein (SMN) complex directs the assembly of the seven Sm proteins B/B', D1, D2, D3, E, F and G in a ring structure around each snRNA's highly conserved uridine-rich Sm-site, forming the so-called Sm-core (Raker *et al.*, 1996; Fischer *et al.*, 1997). This promotes processing of the 3' end of the snRNA and hypermethylation of the m^7G cap to a 2,2,7-trimethylguanosine (m_3G) cap, and the premature snRNP is reimported into the nucleus, where additional particle-specific proteins are recruited to form a complete snRNP (Mattaj, 1986; Fischer & Lührmann, 1990; Nesic *et al.*, 2004).

The U6 snRNA is transcribed by the RNA polymerase III and possesses a rare 5'-terminal γ-monomethylphosphate cap (Kunkel *et al.*, 1986). This cap increases the stability of the U6 snRNA but does not function as a localization signal (Shumyatsky *et al.*, 1993). As a result, the U6 snRNA remains in the nucleus during maturation (Spiller *et al.*, 2007). The U6-specific equivalent of Sm proteins are the LSm proteins (LSm2–8), which also form a heptameric ring around a uridine-rich sequence at the 3' end of the snRNA (Achsel *et al.*, 1999). After maturation, the U4 and U6 snRNAs form an extensive interaction network, which is followed by association of particle-specific proteins, forming the mature di-snRNP (Wersig *et al.*, 1992).

Introduction

The snRNAs form secondary structures, which are highly conserved among higher eukaryotes. Fig. 1.4 illustrates the anticipated structures of the human snRNAs within the respective snRNPs. In the di-snRNP U4 and U6 exhibit complementary regions, which are base-paired to one another (stems I and II) (Brow & Guthrie, 1988). During the dynamic assembly of the spliceosome some secondary structures, as well as interactions between snRNAs, are remodelled.
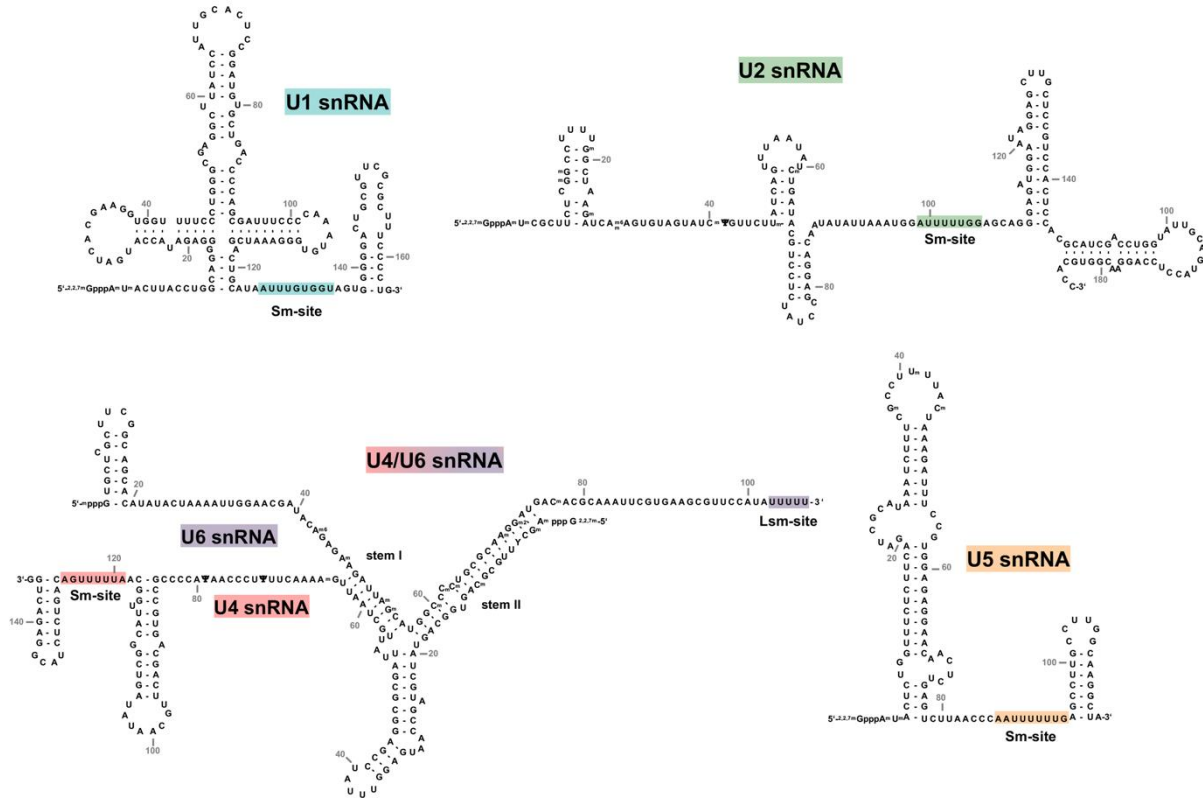
Figure 1.4. Sequences and conserved secondary structures of the human spliceosomal snRNAs. Illustration of the snRNA's structures as anticipated within the respective snRNPs. The basis for this figure was kindly provided by Berthold Kastner Department of Cellular Biochemistry, MPI-BPC.

1.4.2 The snRNP-specific proteins

In addition to the Sm or LSm proteins, the snRNAs further associate with a set of snRNP-specific proteins (Fig. 1.5). The 12S U1 snRNP comprises the particle-specific proteins U1-C, U1-A and U1-70K. Although the U1 snRNA recognizes the 5'SS by base-pairing, U1-C and U1-70K have an important function in stabilizing the interaction of the snRNP with the pre-mRNA (Heinrichs *et al.*, 1990; D. Zhang & Rosbash, 1999).

Introduction

The 17S U2 snRNP comprises the U2-A'/U2-B'' dimer and the two multi-protein sub-complexes SF3a and SF3b (Behrens *et al.*, 1993; Brosi *et al.*, 1993). The trimeric SF3a complex is composed of the proteins SF3a120, SF3a66 and SF3a60 and the heptameric SF3b complex includes SF3b155, SF3b145, SF3b130, SF3b49, SF3b14a/p14, SF3b14b and SF3b10 (Will *et al.*, 2002). The SF3a/SF3b proteins are important for A complex formation, as they support the interaction of the U2 snRNA with the BS region (Gozani *et al.*, 1996). Additionally, a number of proteins, called U2-related proteins loosely associate with the U2 snRNP (Will *et al.*, 2002).

The 20S U5 snRNP comprises six particle-specific proteins (hPrp8, hBrr2, hSnu114, hPrp6, hPrp28, hLin1, 40K and hDib1), some of which are important for functional spliceosome assembly (Bach *et al.*, 1989; Black & Pinto, 1989; Behrens & Lührmann, 1991). The two helicases Prp28 and Brr2 are essential for the formation of B and B^{act} complexes, respectively (Laggerbauer *et al.*, 1998; Staley & Guthrie, 1999; Boesler *et al.*, 2015). The largest U5 protein, Prp8, is located in the catalytic core of the spliceosome and has been shown to be in contact with the 5'SS, the 3'SS, the BPS as well as the U5 and U6 snRNAs (MacMillan *et al.*, 1994; Umen & Guthrie, 1995; Reyes *et al.*, 1996). Furthermore, Prp8 interacts with Snu114; both of these have been shown to regulate Brr2 activity (Achsel *et al.*, 1998; Bartels *et al.*, 2002; Small *et al.*, 2006; Maeder *et al.*, 2008; Mozaffari-Jovin *et al.*, 2012).
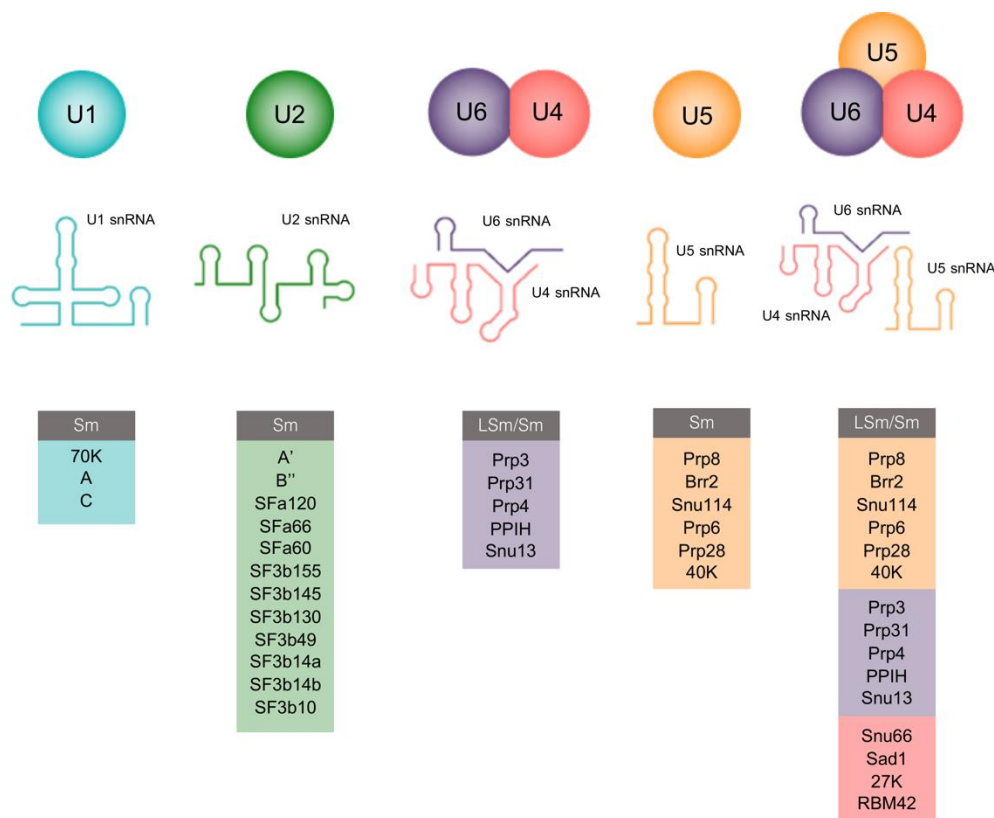
Figure 1.5. Composition of the human snRNPs. Each snRNP consists of one (or two) snRNA(s) and a set of complex-specific proteins. The secondary structure of the human U snRNAs is illustrated schematically. Sm/LSm proteins are shown in grey boxes and the particle-specific proteins of the respective snRNP are shown in the coloured boxes. (Will & Lührmann, 2011)

Introduction

The 13S U4/U6 di-snRNP includes five particle-specific proteins: hPrp3, hPrp31, hPrp4, PPIH and Snu13. The di-snRNP is further associated with U5, forming the 25S U4/U6.U5 tri-snRNP. This particle further contains four specific proteins hSnu66, hSad1, 27K and RBM42 (Behrens & Lührmann, 1991).

While the U4 and U6 snRNAs interact extensively in the di-snRNP, the association of U5 appears to be based rather on protein-protein interactions. The interaction between the U5 protein Prp6 and the di-snRNP protein Prp31 is a prerequisite for U4/U6.U5 tri-snRNP formation (Makarov *et al.*, 2000; Makarova *et al.*, 2002; Schaffert *et al.*, 2004).

1.5 Spliceosome assembly

A unique characteristic of the spliceosome is its sequential assembly *de novo* on each new pre-mRNA substrate (Fig.1.6). None of the snRNPs harbour a pre-formed active site; instead, this is created in a highly controlled manner during the assembly process. Its creation involves substantial remodelling of RNA/RNA and RNA-protein interactions. The driving forces behind these rearrangements are eight highly conserved DExD/H-box ATPases/helicases: UAP56, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22 and Prp43 (Staley & Guthrie, 1998; Wahl *et al.*, 2009; Will & Lührmann, 2011).

These helicases belong to the superfamily 2 (SF2) of helicases, which characteristically comprise two RecA-like domains and variable N- and/or C-terminal regions. The RecA-like domains comprise the highly conserved helicase motifs for binding and hydrolysis of ATP, for binding of RNA and for the coordination of ATP binding and RNA unwinding, respectively (Cordin & Beggs, 2013; Sloan & Bohnsack, 2018). During spliceosome assembly, spliceosome activation, splicing catalysis and spliceosome disassembly, these enzymes promote the specific removal of interaction partners (RNA or proteins) in an ATP-dependent manner, which leads to the formation of new base-pairing or protein-binding.

Furthermore, DExD/H-box helicases play a central role in proofreading, i.e., the rejection and discarding of suboptimal substrates, thereby ensuring the high fidelity of splicing (Semlow & Staley, 2012). While the majority of these enzymes associate only transiently with the spliceosome, the Brr2 helicase is recruited as part of the tri-snRNP and is released during spliceosome disassembly (Makarov *et al.*, 2002). Consequently, the extact regulation of Brr2's activity is a prerequisite for preventing premature unwinding of the U4/U6 interaction in the tri-snRNP or the B complex. This task appears to be mainly executed by several other spliceosomal components including the U5 protein Prp8 and probably several B-specific proteins.

Spliceosome assembly is initiated by the binding of the U1 snRNA to the 5'SS of the intron, which is supported by the U1-specific proteins (Zhuang & Weiner, 1986; Heinrichs *et al.*, 1990; D. Zhang & Rosbash, 1999). Additionally, SF1, U2AF65 and U2AF35 recognize the BPS, the PY tract and the 3'SS, respectively (Ruskin *et al.*, 1988; Berglund *et al.*, 1997; Merendino *et al.*, 1999).

Already during formation of this first assembly intermediate (the so-called E complex), all relevant splicing elements are recognized by the distinct splicing factors.

Although the U2 snRNP associates loosely with the E complex, its stable integration into the A complex requires the helicases UAP56 and Prp5 to perform ATP-dependent rearrangements that result in the removal of SF1 from the pre-mRNA and allow the U2 snRNP to interact with the branch-site region (Berglund *et al.*, 1997; Das *et al.*, 2000; M. Zhang & Green, 2001; Schwer, 2001; Will *et al.*, 2002), whereby the U2 snRNA base-pairs with the branch site (J. Wu & Manley, 1989; Zhuang & Weiner, 1989). In addition, numerous U2 SF3a/b proteins establish contacts to U2AF65 and to the branch-site adenosine itself, as well as to a region upstream of the branch site termed the "anchoring site" (Gozani *et al.*, 1996, 1998; Will *et al.*, 2001; Schellenberg *et al.*, 2006).

This comprehensive interaction network ensures that the splicing motif is identified with high precision. The binding of U2 further causes the branch-site adenosine to bulge out, which is important for the subsequent first step of splicing (Query *et al.*, 1994).
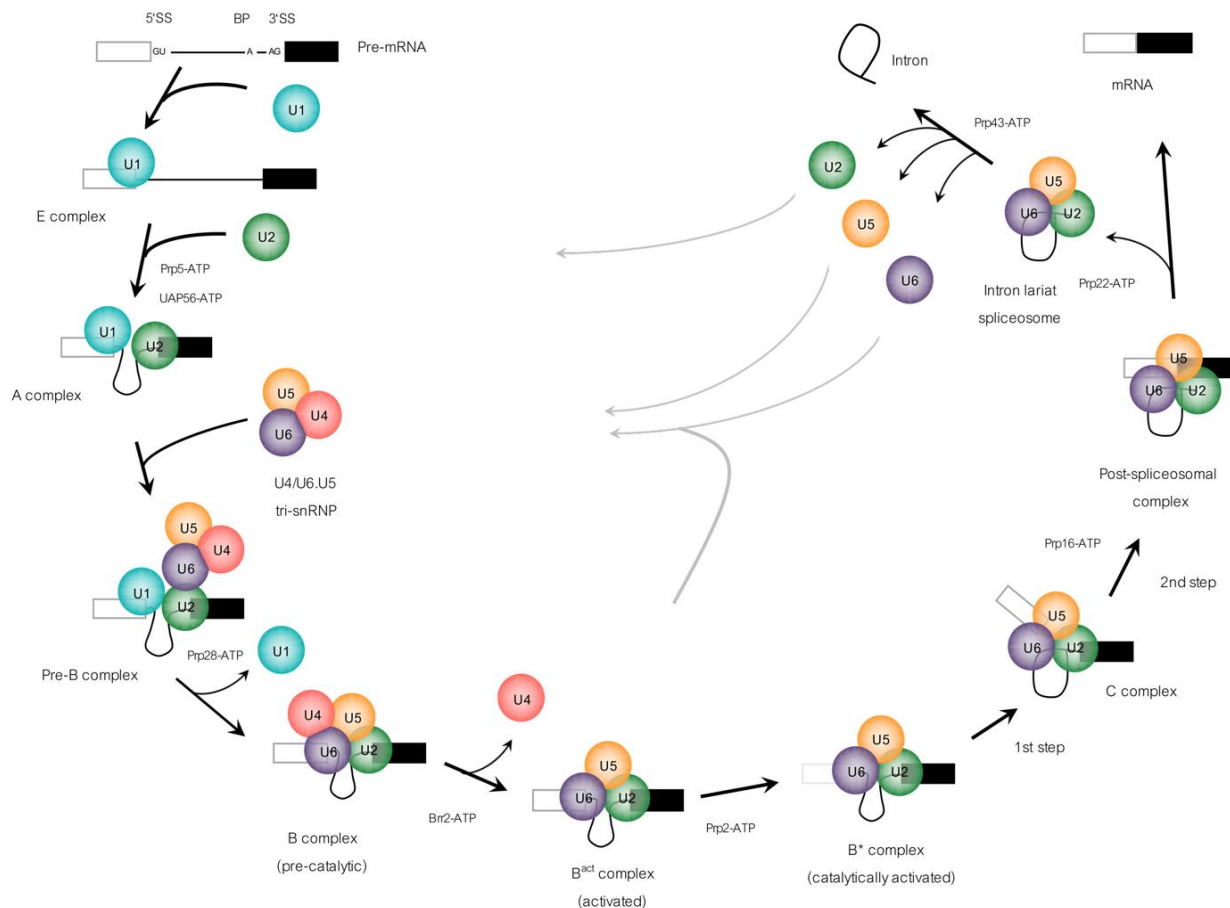
Figure 1.6. Spliceosomal splicing cycle. Schematic diagram of the sequential spliceosome assembly and disassembly during one splicing cycle. The assembly is initiated by the binding of U1 to the 5'SS (E complex). Stable binding of U2 to the BPS results in A complex formation. After loose association of the tri-snRNP (pre-B complex), its stable integration and the loss of U1 lead to the B complex. Structural rearrangements, including the release of U4 (B^{act} complex) as well as the subsequent action of Prp2 give rise to a catalytically active (B*) spliceosome. The first transesterification reaction results in the formation of complex C and the second splicing reaction in the mature mRNA and the intron lariat. The spliceosomal components dissociate and are regenerated for the next round of splicing (modified from Will and Lührmann, 2011 and Boesler *et al.*, 2015).

Introduction

Association of the pre-assembled U4/U6.U5 tri-snRNP with the A complex results in the relatively unstable 37S pre-B complex, which comprises all five snRNAs. In this complex the tri-snRNP is loosely docked to the spliceosome by interacting with U2 forming the U2/U6 helix II. To integrate the tri-snRNP stably into the spliceosome, the binding of the U1 snRNP to the 5'SS is destabilized by the DEAD-box helicase Prp28 and replaced by the ACAGAG box of the U6 snRNA, resulting in the formation of the pre-catalytic B complex (Staley & Guthrie, 1999; Boesler *et al.*, 2015). At this stage the U5 snRNA establishes contacts to the 5' exon close to the 5'SS, which are maintained throughout both steps of splicing (Sontheimer & Steitz, 1993; Newman, 2008).

Also, the U5 protein Prp8 is in contact with the pre-mRNA close to the 5'SS, an interaction which is thought to stabilize the U5 snRNA at the 5'SS and appears to be essential for tri-snRNP integration (Teigelkamp *et al.*, 1995; Boesler *et al.*, 2015). Formation of the B complex is further accompanied by phosphorylation of Prp31 (di-snRNP) and Prp6 (U5) through the Prp4 kinase, which is either crucial for or is a result of stable integration of the tri-snRNP (Schneider, Hsiao *et al.*, 2010; Boesler *et al.*, 2015).

To convert the B complex into an activated spliceosome (B^{act} complex), the B complex needs to undergo substantial conformational and compositional rearrangements. During activation, the base-pairing of the U4/U6 duplex is unwound by the Brr2 helicase, which results in the release of the U4 snRNA along with the U4/U6-associated proteins and factors that tether the U5 snRNP to the di-snRNP (Laggerbauer *et al.*, 1998; Bessonov *et al.*, 2010). The displacement of U4 allows the U6 snRNA to interact extensively with the U2 snRNA. The simultaneous interaction of U6 with the 5'SS and U2 brings the 5'SS and the BS into close proximity with one another, a conformation that allows the first splicing reaction to take place (Madhani & Guthrie, 1992; Sun & Manley, 1995).

However, the B^{act} complex is still catalytically inactive. As recently shown by cryo-electron microscopy, both the 5'SS and the U2/BS helix are shielded by proteins at this stage. Therefore, the action of the Prp2 helicase is required to promote the destabilization of the U2 SF3b proteins from the BS; it also appears to liberate the 5'SS for catalysis (Lardelli *et al.*, 2010; Ohrt *et al.*, 2012; Rauhut *et al.*, 2016; Haselbach *et al.*, 2018).

The resulting B* complex is now capable of catalysing the first transesterification reaction, where the branch-point adenosine attacks the 5'SS nucleophilically, thereby liberating the 5' exon and forming the 3' exon-lariat intermediate. This gives rise to the C complex, which is remodelled by the helicase Prp16 to prepare the spliceosome for the second step of splicing. This includes the displacement of the branch-site region from the catalytic centre and juxtaposition of the 5' exon's 3'-OH and the 3'SS (Schwer & Guthrie, 1991; Ohrt *et al.*, 2013; Fica *et al.*, 2017).

The resulting C* complex catalyses step two of splicing, during which the 3' hydroxyl group of the 5' exon attacks the 3'SS, resulting in exon ligation and the formation of the intron lariat. The Prp22 helicase then promotes the release of the mature mRNA from the post-catalytic spliceosome (Mayas *et al.*, 2006; Schwer, 2008) in form of a mRNP and is subsequently exported to the cytoplasm (Le Hir *et al.*, 2000).

The resulting intron lariat spliceosome (ILS) is then disassembled by the action of Prp43. As a result, the intron lariat is released, debranched and degraded, and the snRNPs are separated and recycled to enter another round of splicing (Arenas & Abelson, 1997; Gee *et al.*, 1997).

1.6 Gene architecture and pre-mRNA splicing

The architecture of pre-mRNAs can vary widely between different organisms. In the yeast genome, only ~4% of all genes include introns, and usually not more than one intron is found per gene. Furthermore, the length of yeast introns is relatively consistent; the vast majority of introns range in length between 50 and 500 bp, with none exceeding 1000 bp (Lopez & Séraphin, 1999; Spingola *et al.*, 1999).

In contrast, the human genome is much more complex. Around 90% of all genes include at least one intron, and the average human gene comprises 8.8 exons and 7.8 introns. While exons display a relatively uniform length of ~170 bp, the length of introns can vary immensely – from fewer than 100 up to more than one million bases. However, most introns appear to be in the range of 100–1000 bp, while introns with canonical ends are rarely shorter than 70 bases. This sophisticated human genome structure therefore requires many regulatory factors to ensure correct splice site selection and splicing. (Lander *et al.*, 2001; Sakharkar *et al.*, 2004; Abebrese *et al.*, 2017).

1.6.1 Alternative splicing

The importance of splicing increases with the complexity of an organism. Whereas splicing is completely absent in prokaryotes, simple eukaryotes, such as yeasts, predominantly make use of basic constitutive splicing, which means that a gene is always spliced in the same way and the product does not show any sequence variation. Complex eukaryotic organisms, such as humans, rely on a more sophisticated form of splicing, called alternative splicing. More than 90% of human genes containing more than one intron undergo alternative splicing (E. T. Wang *et al.*, 2008; Pan *et al.*, 2008).

Alternative splicing means that the exons of a pre-mRNA can be combined in different ways. In general, five kinds of alternative splicing events are known: (i) cassette exon skipping, (ii) alternative 5'SS or (iii) 3' SS selection, (iv) intron retention and (v) mutually exclusive introns (E. T. Wang *et al.*, 2008; Pan *et al.*, 2008) (Fig. 1.7). This flexibility in intron/exon definition makes possible the generation of several protein isoforms from a single gene, and thus results in a considerable expansion of the cell's protein diversity. By this means the relatively small human genome, which comprises 20,000–25,000 genes, yields an estimated 80,000–120,000 proteins (Liang *et al.*, 2000; Consortium, 2004; Yura *et al.*, 2006). As a result, constitutive and especially alternative splicing account for the greater part of the molecular and cellular complexity in higher eukaryotes (Blencowe, 2006; Y. Wang *et al.*, 2015).

Introduction

The regulation of alternative splicing appears to be quite complex and is significantly modulated by trans-acting factors, such as the SR proteins and the hnRNPs. The family of SR proteins is characterized by the so-called RS domain, which comprises long stretches that are rich in the amino acids serine (S) and arginine (R) and promotes the interaction with other protein components. SR proteins further include at least one RNA recognition motif (RRM), by which they bind to pre-mRNAs (Shepard & Hertel, 2009).
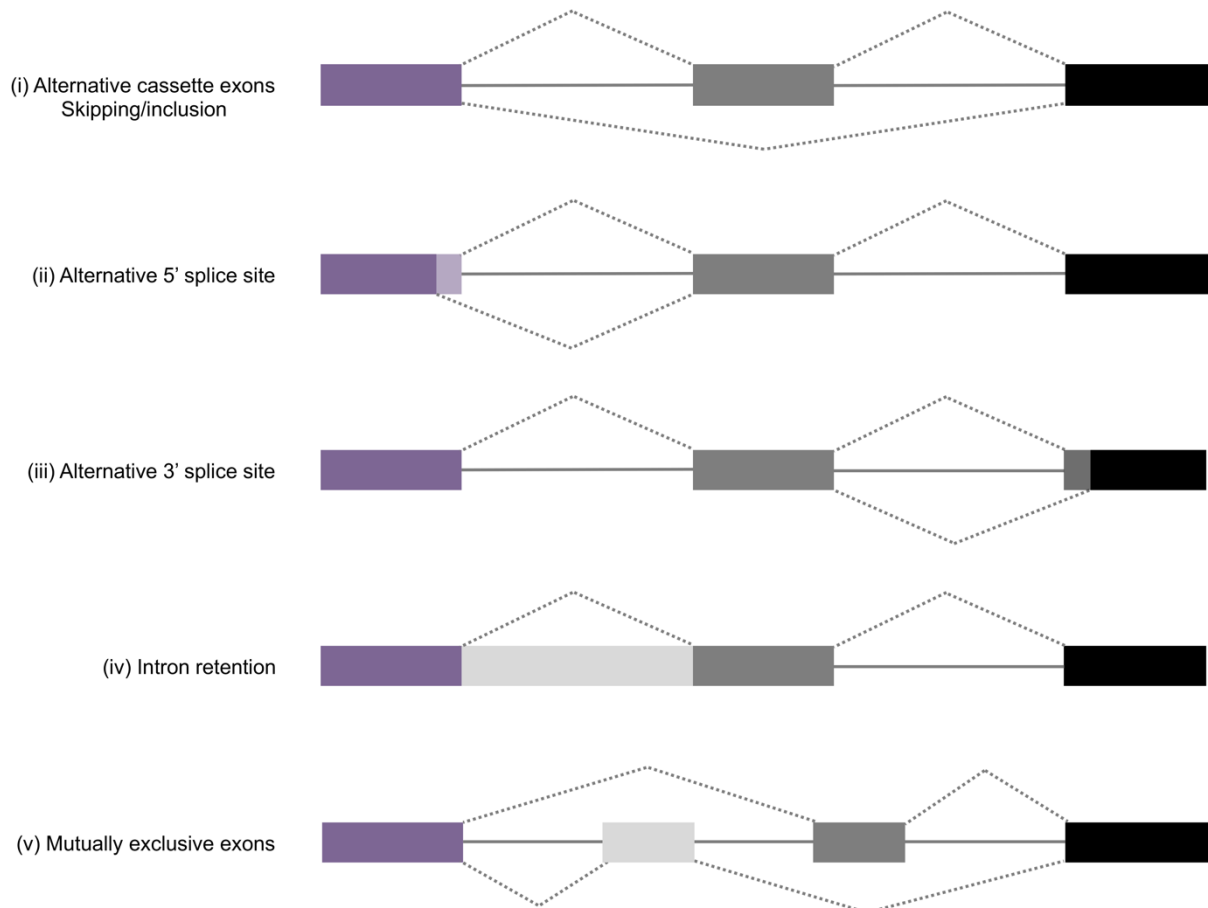
Figure 1.7. Types of alternative splicing. Schematic representation of the different types of alternative splicing. (i) Alternative cassette exons are either included or skipped (omitted). During skipping, the exon and both surrounding introns are excised from the pre-mRNA. Skipping of alternative cassette exons represents the majority of alternative splicing events. (ii & iii) Selection of an alternative 5' or 3' splice site results in the exclusion of partial exon segments from, or their inclusion in, the mature mRNA. (iv) Intron retention describes the inclusion of an entire intron into the spliced product. (v) When two exons are mutually exclusive, only one of them will be included in the mature transcript. Constitutively spliced exons are indicated as purple and black boxes and introns are shown as solid lines. Alternatively spliced regions are represented by grey boxes and splicing choices are marked by dotted lines. The figure is adapted from Wang *et al.*, 2015.

The hnRNPs (heterogeneous nuclear RNPs) comprise a more heterogeneous set of nuclear proteins, which specifically associate with primary transcripts of the RNA polymerase II (Geuens *et al.*, 2016). These proteins are structurally more diverse than the SR proteins; typically, they contain several structural domains, of which the RRM domain is the most common motif (Han *et al.*, 2010). Additionally, hnRNPs tend to include auxiliary domains, such as the common RGG repeat, which functions as a protein-protein interaction module to allow interaction with other hnRNPs (Birney *et al.*, 1993).

These splicing factors interact with specific cis-elements that can be located within the intron or the exon. While SR proteins preferentially interact with exonic or intronic splicing enhancers (ESE & ISE), splicing silencers (ESS & ISS) are preferentially bound by hnRNPs. The interplay of these elements and splicing factors, which play opposing roles in splicing, are the decisive factors for splice-site choices under given conditions (Black, 2003; Bradley *et al.*, 2014).

1.6.2 Co-transcriptional splicing

Although splicing functions independently of transcription, most splicing events occur co-transcriptionally (Beyer & Osheim, 1988; Goldstrohm *et al.*, 2001; Pandya-Jones & Black, 2009; Girard *et al.*, 2012).

Pandya-Jones and Black established a cell-fractionation procedure to separate chromatin-associated RNA from soluble nucleoplasmic RNA. Subsequent quantification by quantitative RT-PCR of the nascent RNA fraction and the RNA faction already released showed that the majority of constitutive introns of the c-Src and fibronectin pre-mRNAs were excised while the pre-mRNA was still being transcribed by the RNA Polymerase II. Also introns abutting alternative exons were often removed co-transcriptionally, but with varying excision efficiency among distinct, differentially regulated, cell lines.

This cell-fractionation approach was also applied by Girard *et al.* to detect co-transcriptional splicing in HeLa cells with an independent readout. They analysed the nucleoplasmic and chromatin fractions by western blotting using antibodies specific for the phosphorylated U2 protein SF3b155 (P-SF3b155). As phosphorylation of SF3b155 occurs only in activated spliceosomes, this antibody thus allows one to detect active spliceosomes in HeLa cells. Quantification of the P-SF3b155 signal revealed that ~80% of activated spliceosomes were associated with the chromatin, suggesting that only 20% of pre-mRNA splicing occurs post-transcriptionally.

Coupling of pre-mRNA splicing to transcription not only stimulates both processes (Hirose *et al.*, 1999; Fong & Zhou, 2001; Millhouse & Manley, 2005; David *et al.*, 2011), but also improves splicing fidelity and makes possible the coordinated regulation of the three processing events (Neugebauer, 2002; Bentley, 2014).

Co-transcriptional splicing also appears to be an important factor for the regulation of alternative splicing.

Two models have been proposed that may explain how the RNA polymerase II exerts influence on splice site selection; these models are not necessarily mutually exclusive.

Introduction

The **recruitment model** suggests that the C-terminal domain (CTD) of the RNA polymerase II recruits specific splicing factors, including SR proteins and other early spliceosomal factors such as the U1 snRNP, to the nascent transcript, thereby influencing splice-site selection (Misteli & Spector, 1999; Morris & Greenleaf, 2000; de la Mata & Kornblihtt, 2006; David *et al.*, 2011). The **kinetic model** concerns the elongation rate of the RNA polymerase II, which again is determined by certain histone marks. While rapid elongation favours strong splice sites, slow elongation can facilitate spliceosome assembly on introns with weak splice sites or allow binding of splicing enhancers/inhibitors (de la Mata *et al.*, 2003; Dujardin *et al.*, 2014). Since various further factors have been shown to influence alternative splicing it will be a major challenge to understand fully its complex regulation.

1.6.3 Exon-definition pathway

In the human genome, intron size can vary from less than 100 nt up to several hundred thousand nt. The spliceosome, however, can only assemble across an intron (intron definition) when the length of the intron does not exceed ~300 nt (Sterner *et al.*, 1996). Therefore, the more frequent longer introns require an alternative to the traditional recognition pathway that allows splice-site identification across the exon (exon definition) (Robberson *et al.*, 1990). In this case, the U1 snRNP interacts with the exon's downstream 5'SS, while the U2 snRNP and U2AF65/35 bind to the BPS or the PY tract/ 3'SS upstream of the exon (Fig. 1.8). SR proteins that bind to ESE sequences support the assembly and stability of the exon-defined A-like complexes (Lam & Hertel, 2002). Through their RS domains SR proteins can interact with other splicing factors that also include a RS domain, as is the case for U2AF35 and U1-70K. Thus, SR proteins could form an important bridge between the U1 and U2 snRNPs in a cross-exon A-like complex (J. Y. Wu & Maniatis, 1993; Ram & Ast, 2006).
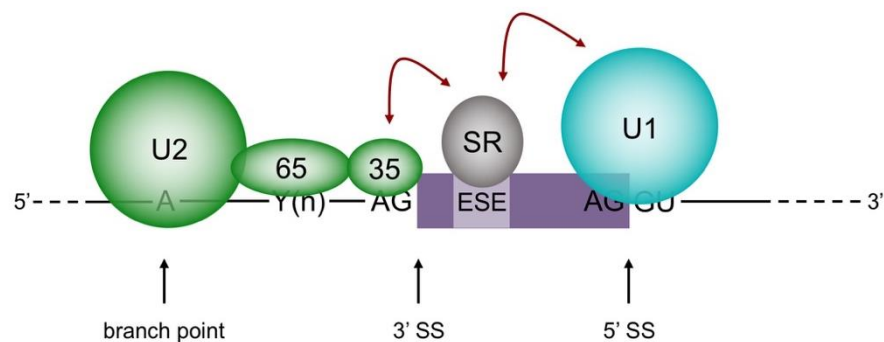
Figure 1.8. Molecular interactions during spliceosome assembly across an exon. In the exon-defined, A-like complex, U2 snRNP is bound to an upstream BPS and U1 snRNP to the downstream 5'SS. The U2AF65/35 heterodimer interacts with the upstream PY tract/3'SS. SR proteins, which bind to exonic splicing enhancers (ESEs), bridge interactions with components of the general splicing machinery (indicated by red arrows). (Ram & Ast, 2006)

The splicing reaction, however, can only occur across the intron, and this necessitates the conversion of the exon-defined complex to an intron-defined spliceosome. Little is known about the mechanism of this conversion, but, although the commitment to splice-site pairing has been attributed to the formation of the A complex (Lim & Hertel, 2004), recent studies suggest that exon-defined A-like complexes do not need to undergo a transition to an intron-defined A complex. As the U4/U6.U5 tri-snRNP can interact directly with the cross-exon A-like complex, the proximity of an adequate 5'SS allows the subsequent formation of a B-like complex (Fig. 1.9). This complex resembles the common intron-defined B complex and proceeds in the same manner with its catalytic activation and splicing catalysis (Schneider, Will *et al.*, 2010; Boesler *et al.*, 2016).

Numerous studies indicate that the transition to an intron-defined spliceosome is an important checkpoint for determining whether an exon is included or skipped during alternative splice site selection. If an exon cannot be recognized, or the transition to an intron-defined organization is prevented, the exon will inevitably be skipped (Izquierdo *et al.*, 2005; House & Lynch, 2006; Bonnal *et al.*, 2008; Sharma *et al.*, 2008).
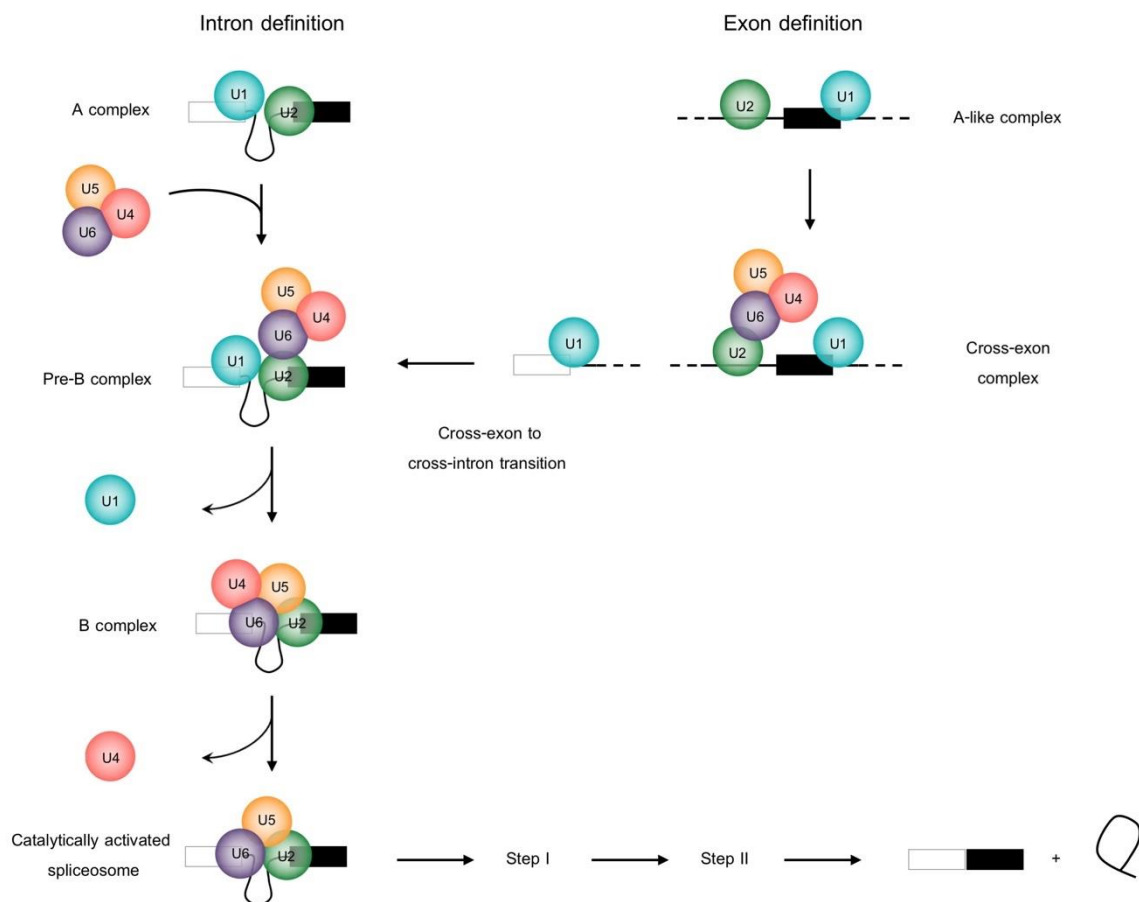
Figure 1.9. Intron- *versus* exon-definition pathway. Spliceosome assembly during intron and exon definition. In the intron-defined pathway, U1 and U2 snRNPs bind to the 5'SS and BPS of the same intron. Upon association and integration of the U4/U6.U5 tri-snRNP, the B complex is formed. During exon definition, U2 snRNP interacts with an upstream BPS and the U1 snRNP with the downstream 5'SS of the exon. The recruitment of the U4/U6.U5 tri-snRNP results in the cross-exon complex, which makes the transition to an intron-defined organization. Subsequently, spliceosome assembly proceeds along the canonical splicing pathway. The figure is adapted from Schneider, Will *et al.*, 2010.

Introduction

1.6.4 Spliceosome assembly requires a minimum intron length

Intron length represents an important factor in pre-mRNA splicing. While the alternative exon-definition pathway enables the spliceosome to deal with introns that exceed the acceptable size for intron definition, a minimum intron length appears to be a prerequisite for the assembly of functional spliceosomes. Investigation of short introns in human cells showed that canonical introns, which are thought to be excised by the spliceosome, are rarely shorter than 70 nt (Abebrese *et al.*, 2017). Consistently, in the 1980s several studies showed that a minimum of 60–80 nt must separate the 5'SS and the 3'SS to allow proper splicing (Wieringa *et al.*, 1984; Ulfendahl *et al.*, 1985). Soon after this discovery, other studies defined the decisive factor more precisely as the distance between the 5'SS and the branch site, as elongating the intron by PY-tract extension could not restore splicing (Fu *et al.*, 1988; Köhrer & Domdey, 1988; Smith & Nadal-Ginard, 1989). A distance between the 5'SS and the BS of 45–52 nt was shown to result in the selection of an upstream cryptic 5'SS, which increases the distance to the branch site. In the absence of an alternative splice site, the intron was shown to be retained in the mature mRNA (Krainer *et al.*, 1984; Köhrer & Domdey, 1988).

Steric hindrance has been suggested as a possible reason for a minimum length requirement. The U1 and (especially) the U2 snRNP are large complexes, whose simultaneous binding to the 5'SS and to the BS may require a certain amount of space and flexibility (Wieringa *et al.*, 1984; Fu *et al.*, 1988). This idea has been supported by native gel and sedimentation analysis, which showed impaired spliceosome assembly on short non-functional introns (Köhrer & Domdey, 1988; Smith & Nadal-Ginard, 1989; Himmelspach *et al.*, 1991). Thus, intron size seems to be a relevant determinant for spliceosome assembly and splicing. However, the exact mechanism of intron-length-dependent spliceosome assembly is only poorly understood.

1.7 Dynamic RNA/RNA interaction network

The formation of a functional spliceosome is associated with extensive remodelling of the RNA/RNA interaction network (Fig. 1.10) (Wahl *et al.*, 2009; Will & Lührmann, 2011). The initial recognition of the 5'SS occurs through base-pairing with the 5' end of the U1 snRNA (Zhuang & Weiner, 1986). The second interaction with the pre-mRNA is formed by the U2 snRNA, resulting in a short U2/BPS-duplex, where the branch-point adenosine is bulged out (J. Wu & Manley, 1989; Zhuang & Weiner, 1989; Query *et al.*, 1994). This exposes the adenosine's 2' hydroxyl group for nucleophilic attack at the 5'SS.

Upon recruitment of the U4/U6.U5 tri-snRNP, the 5' end of the U2 snRNA and the 3' end of the U6 snRNA form a short helix (U2/U6 helix II). Stable integration of the tri-snRNP, however, involves the displacement of U1 from the 5'SS and also base-pair formation between the 5'SS and the U6 ACAGAG box. In the B complex U6 is thus base-paired to U2, to the pre-mRNA and extensively to U4. (Staley & Guthrie, 1999; Boesler *et al.*, 2015).

The U6 snRNA includes important components of the catalytic centre, which are kept in an inactive state by hybridization with the U4 snRNA in the tri-snRNP and the B complex to prevent premature catalysis. Consequently, the catalytic activation of the spliceosome requires the disruption of the U4/U6 snRNA duplex (U4/U6 stem I and stem II), which is followed by release of the U4 snRNP (Laggerbauer *et al.*, 1998; Raghunathan & Guthrie, 1998). This allows U6 to form interactions with the U2 snRNA (U2/U6 helix Ia and I), thereby juxtaposing the 5'SS and the BPS for the first splicing reaction. The release of U4 further allows U6 to form an internal stem-loop (U6 ISL) that structurally resembles domain V (D5) of group II self-splicing introns and, likewise, coordinates catalytically important metal ions (Madhani & Guthrie, 1992; Anokhina *et al.*, 2013).

Cryo-electron microscopy of the first *S. pombe* spliceosome illustrated the major structural similarities between the U6 ISL and D5 (Hang *et al.*, 2015). In the spliceosome, three conserved motifs are involved in metal-binding. These include (i) the so-called catalytic triad AGC, which is located within the U2/U6 helix Ib at the stem of the U6 ISL, (ii) a bulged-out uridine located five nucleotides from this triad and (iii) the U6 ACAGAG box. These three structures formed by U2 and U6 are important for the RNA-metal-mediated catalysis of both pre-mRNA splicing steps (Keating *et al.*, 2010; Fica *et al.*, 2013). Thereby, the catalytic AGC triad forms a triple helix with the terminal GA of the ACAGA<u>GA</u> sequence and the bulged-out uridine of the U6 ISL. This so-called catalytic triplex coordinates two magnesium ions at a distance of about 4 Å; these mediate the catalysis of both steps of pre-mRNA splicing. In the same manner, magnesium ions are coordinated by group II self-splicing introns, suggesting that group II introns and the spliceosome share an evolutionary heritage.
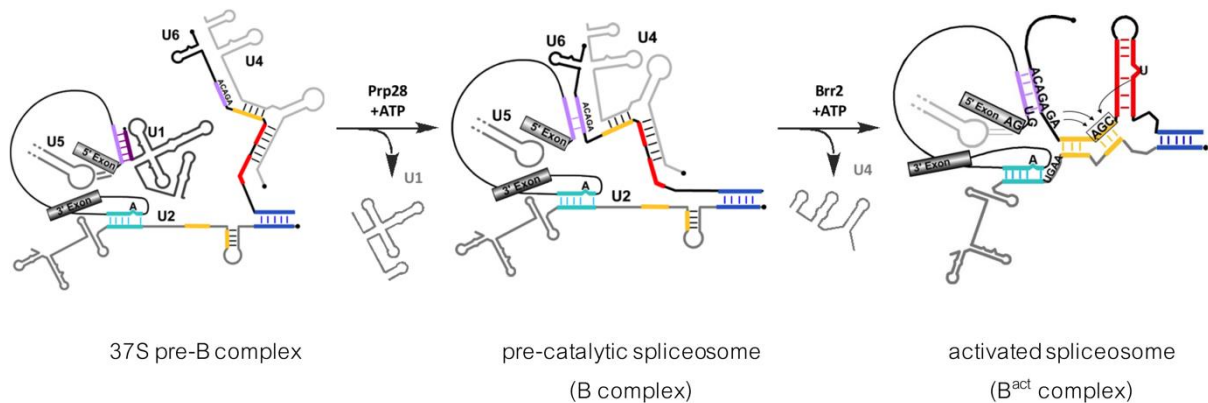
Figure 1.10. Dynamic RNA-RNA network in the spliceosome. Schematic illustration of the rearrangements of RNA-RNA interactions during formation of a catalytically activated spliceosome. In the pre-B complex, the U1 snRNA is base-paired with the 5'SS (aubergine), and the U2 snRNA is bound to BPS (turquoise). Initial association of the U4/U6.U5 tri-snRNP occurs between the U6 and U2 snRNAs (blue). Stable integration of the U4/U6.U5 tri-snRNP during B complex formation involves the disruption of base-pairing between the U1 snRNA and the 5'SS, which allows the 5'SS to interact with the conserved ACAGAG motif of the U6 snRNA (purple). During activation, the base-pairing between U4 and U6 snRNAs is disrupted, resulting in extensive base-pairing between the U2 and U6 snRNAs via helix (yellow). The U6 snRNA further forms an essential internal-stem loop (ILS) (red), while the U2 snRNA remains base-paired with the BPS. U5 snRNA contacts nucleotides of the exon. Exons are represented as grey boxes, while introns are shown as a black line. Secondary structures of snRNAs are shown schematically in grey or black. Of the U5 snRNA, only loop I is shown. The regions involved in base-pairing (short lines) are shown in different colours. This figure was kindly provided by Berthold Kastner, Department of Cellular Biochemistry, MPI-BPC.

Introduction

The U5 snRNA comprises a stem loop (U5 stem loop I), which interacts with the 5' exon before and after the first step of splicing and further contacts the 3' exon after the first catalytic step has occurred. Thus, the U5 snRNA plays an important role in properly positioning the two exons for the second splicing reaction (Sontheimer & Steitz, 1993; Newman, 2008).

1.8 Dynamic protein composition of the spliceosome

The spliceosome is a particularly protein-rich molecular RNP machine. Around 170 proteins have been identified as spliceosomal components in humans, while individual spliceosomal complexes can contain ~110 proteins (Fig. 1.11). The yeast spliceosome comprises fewer (~90) protein factors (Fabrizio *et al.*, 2009), most of which have homologues in higher eukaryotes. This indicates that the yeast spliceosome represents the basic, evolutionarily conserved core spliceosome. Consistently with this idea, it appears that many proteins that are not conserved from yeast to human have regulatory roles, such as the regulation of alternative splicing.

In addition to the ~45 snRNP-associated factors, a number of non-snRNP proteins accompany the spliceosome. Their attachment is often transient and limited to the stages when their presence or action is required. (Wahl *et al.*, 2009; Will & Lührmann, 2011).

Mass-spectrometric analysis of spliceosomal complexes assembled *in vitro* have allowed the determination of the protein composition of the spliceosome's major assembly stages, thereby revealing the extent of protein exchange between one intermediate and the next. This in turn allowed the identification of groups of proteins that are exclusively present at each particular stage, thereby giving a unique proteomic signature to distinct intermediate spliceosomal complexes (Deckert *et al.*, 2006; Behzadnia *et al.*, 2007; Bessonov *et al.*, 2010; Agafonov *et al.*, 2011).

During B complex formation a number of designated A complex proteins and U1 snRNP proteins are released, while about 25 proteins join the spliceosome as part of the U4/U6.U5 tri-snRNP together with a number of non-snRNP protein factors. These include a group of proteins, previously identified as B-specific proteins, as well as the pre-assembled Prp19/CDC5L complex, a set of Prp19/CDC5L-related proteins and the RES complex. Owing to their sub-stoichiometric abundance in the B complex, the latter spliceosomal factors were thought be only loosely associated with the B complex, with their stable integration following during the activation. However, recent evidence suggests that these proteins are recruited only after B^{act} complex formation.

This protein composition was based on the analysis of so-called kinetic B complexes, which were obtained by incubation of the pre-mRNA under splicing conditions for 5–8 minutes before the assembly was stopped. During this time no first-step products are formed, indicating that the majority of spliceosomes are in a pre-catalytic state. Nevertheless, when B complexes are kinetically stalled the possibility exists that a small fraction of the population undergoes activation, which can give a false impression of the actual situation.
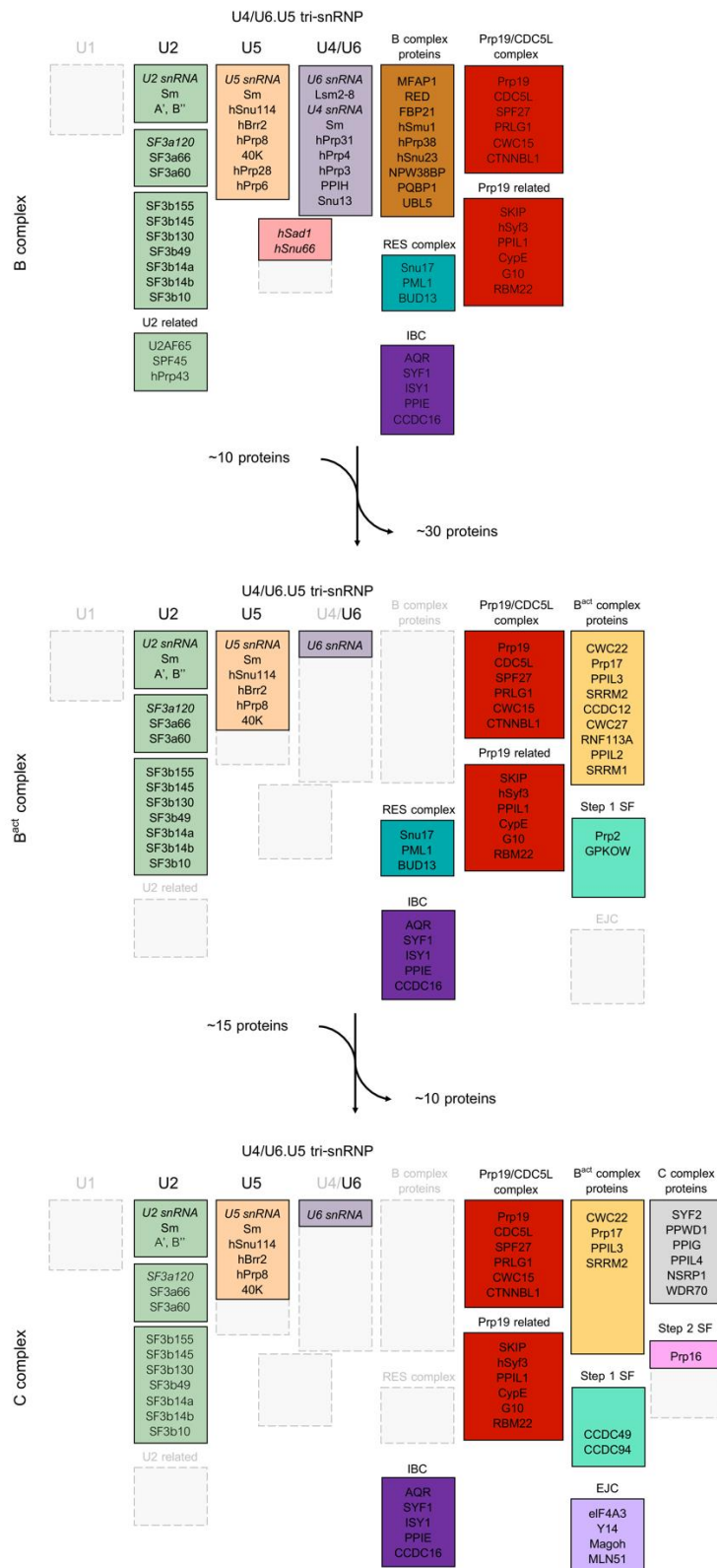
Figure 1.11. Dynamic protein composition of the human spliceosome. The protein composition of the human B, B^act and C complexes as determined by 2D gel-electrophoresis and mass spectrometry. The C complex was stalled before the Prp16-driven rearrangements by using a dominant negative Prp16 mutant. The relative abundance of proteins is indicated by bold (stoichiometric amounts) or light (substoichiometric amounts) lettering. Proteins are grouped according to their snRNP association, function, presence in a stable heteromeric complex or association with a particular spliceosomal complex, as indicated. SR and hnRNP proteins, as well as those present in very low amounts, are not shown. This figure is adapted from results kindly obtained by Leyla El Ayoubi and Dmitry Agafonov, Department of Cellular Biochemistry, MPI-BPC.

Introduction

Recently, two studies showed that when B complexes were stalled with the ATP analogue ATPγS ($B^{ATPγS}$) (Agafonov, Van Santen *et al.*, 2016) or by lowering the concentration of $MgCl_2$ ($B^{DeltaMgCl2}$) (Bertram, Agafonov, Dybkov *et al.*, 2017) the Prp19/ CDC5L complex, the Prp19-related proteins as well as the RES complex were essentially absent; only RBM22 and SKIP were present in sub-stoichiometric amounts. In contrast, the B-specific proteins were stoichiometrically present. This suggests that in the purified the kinetic B complexes the PRP19/CDC5L and Prp19-related proteins may be derived from contaminating B^{act} complexes, whereas the stalled B complexes represent a more homogeneous B complex population.

The exclusive association of the B-specific proteins with the B complex led to the initial assumption of a role for them in B complex formation. However, recent studies by Boesler *et al.* have suggested that these proteins are not necessarily required for stable B complex formation *in vitro* (Boesler *et al.*, 2015, 2016).

The subsequent activation of the spliceosome is also highly dynamic at the level of protein composition. Proteins associated specifically with the U4/U6 di-snRNP and the U4/U6.U5 tri-snRNP, as well as the B-specific proteins, are released (~30 proteins), while ~10 proteins - including B^{act}-specific factors, the Prp19/CDC5L complex, the inron-binding complex (IBC) and the RES complex - are stably integrated and are present in stoichiometric amounts in the B^{act} complex. While the Prp19/CDC5L complex was proposed to be essential for spliceosome activation and/or the first step of splicing (Makarova *et al.*, 2004), little is known about the function of the RES complex in splicing. A recent study suggests that in yeast this complex is involved in the B-to-B^{act} transition (Bao *et al.*, 2017). However, while these proteins are already associated with the yeast B complex, they appear to be recruited to the human spliceosome after its activation.

During C complex formation, only moderate structural changes take place, but ~10 proteins, the presence of which is no longer required, are released while the C complex-specific proteins, the first-step splicing factors and members of the exon junction complex (EJC) are recruited (~15 proteins). The EJC is placed onto the pre-mRNA upstream of the splice junction and remains associated with the mature mRNA to modulate its downstream metabolism in the cell (Le Hir & Séraphin, 2008).

Finally, the second-step factors and a group of putative C*-specific proteins join the spliceosome after the first step of splicing, while some C complex proteins are released.

1.9 Post-translational phosphorylation of proteins during splicing

Besides the dramatic compositional and conformational changes that the spliceosome undergoes during splicing, alterations in the phosphorylation state of distinct spliceosomal proteins have also been recorded.

Reversible protein phosphorylation is the most common post-translational modification, and the phosphorylation and dephosphorylation of certain proteins has been shown to be essential for spliceosome assembly and splicing catalysis. Interestingly, the majority of spliceosome-related phosphorylation events identified so far take place on serine/threonine residues. At the beginning of the splicing cycle, for example, SR proteins are intensively

phosphorylated (Xiao & Manley, 1997) and several snRNP-associated proteins undergo phosphorylation in early stages of the splicing cycle. This includes extensive phosphorylation of the U1 snRNP protein 70K during E complex formation; this phosphorylation step has been shown to be essential for spliceosome assembly (Woppmann *et al.*, 1993; Hernández *et al.*, 2009).

Additionally, B complex formation is accompanied by phosphorylation events, such as that of Prp28 (by SRPK2), Prp31 (by Prp4-kinase; Fig. 1.12) and Prp6 (also by Prp4-kinase). These proteins were shown to be phosphorylated during stable integration of the tri-snRNP (Schneider, Hsiao *et al.*, 2010).

Furthermore, the U2-associated protein SF3b155 is phosphorylated at specific serine/threonine residues in the B[act] complex before the first catalytic step and is subsequently dephosphorylated in the C complex before the second step of splicing (Agafonov *et al.*, 2011; Girard *et al.*, 2012).

The two redundant phosphatase families PP1 and PP2A were identified as mediating splicing catalysis by dephosphorylating their main substrates, phosphorylated U2-SF3b155 and U5-116K, respectively, which is relevant for initiating the second catalytic splicing step. Similarly, dephosphorylation of 70K, which already occurs during A complex formation, is a prerequisite for the first splicing step to occur (Shi *et al.*, 2006; Wahl *et al.*, 2009).

In general, it can be said that phosphorylation occurs predominantly during spliceosome assembly and activation, whereas catalysis and disassembly are rather characterized by dephosphorylation events. At present it is unclear how the phosphorylation/dephosphorylation of specific proteins controls the progression of the splicing cycle. However, it is believed that phosphorylation events are likely to exert control over protein–protein interactions.

1.10 The structure of the human B complex and the organization of the B-specific proteins

While the composition of the spliceosome has been investigated extensively over many years, it is only recently that detailed structural information could be provided. The size of the spliceosome and its flexibility render comprehensive structural analysis challenging. In the past years, however, several high-resolution structures of distinct spliceosomal complexes derived from yeast and human have been resolved by cryo-electron microscopy (cryo-EM). These include the tri-snRNP (Nguyen *et al.*, 2015, 2016; Agafonov, Kastner *et al.*, 2016), the B complex (Bertram, Agafonov, Dybkov *et al.*, 2017; Plaschka *et al.*, 2017; Bai *et al.*, 2018), the B[act] complex (Rauhut *et al.*, 2016; Yan *et al.*, 2016; Haselbach *et al.*, 2018; X. Zhang *et al.*, 2018), the C complex (Galej *et al.*, 2016; Wan *et al.*, 2016; Zhan *et al.*, 2018), the C* complex (Bertram, Agafonov, Liu *et al.*, 2017; Fica *et al.*, 2017; Yan *et al.*, 2017), the P complex (Bai *et al.*, 2017; Liu *et al.*, 2017; Wilkinson *et al.*, 2017) and the ILS (Wan *et al.*, 2017). Together, these structures have provided a wealth of information about the location of RNAs and proteins in the various spliceosomal "snapshots" and their dynamics during the transition from one complex to the next.

Introduction

Bertram, Agafonov, Dybkov, *et al.* (2017) were able to resolve the structure of the human B complex (Fig. 1.12A), which also revealed the locations of several domains of the B-specific proteins. In the human B complex the U2-containing globular head domain is found on top of a triangular body; the latter comprises the tri-snRNP and the B-specific proteins. The two domains are connected by three main bridges. Bridge B1 comprises the U2/U6 helix II, B2 includes spliceosomal proteins and B3 most likely contains the protein-bound intron, which interacts with both the U2 and U6 snRNAs.

As observed for the human B^{act} (Haselbach *et al.*, 2018; X. Zhang *et al.*, 2018) and C* complexes (Bertram, Agafonov, Liu *et al.*, 2017), the intron forms extended helices with the U6 ACAGAG box at the 5'SS and U2 at the BS. These helices extend beyond the short 5'SS or BS regions and are thought to stabilize the interaction of the snRNAs with the pre-mRNA in the spliceosome, which is further supported by the binding of protein factors.

The U4 Sm core and helix I of the U4/U6 duplex are connected by a single-stranded region of U4, which serves as a docking site for the N-terminal helicase cassette of Brr2 and is essential for the helicase's U4/U6 duplex-unwinding activity *in vitro* (Mozaffari-Jovin *et al.*, 2012). In the B complex, the Brr2 helicase is loaded onto this single-stranded region of the U4 snRNA, but it is not yet in contact with the U4/U6 duplex region. As this isolated B complex remains stable in the presence of ATP, the helicase activity of Brr2 is likely to be negatively regulated in this complex. It appears that at least some of the B-specific proteins are involved in the regulation of Brr2 in the spliceosome when it is primed for activation, which makes their position in the B complex particularly interesting (see below).

During formation of the B complex, the Prp28 helicase is destabilized, which allows the three B-specific proteins Prp38, MFAP1 and Snu23 to bind as a trimeric protein complex close to the extended U6 ACAGAG/5' SS helix. Although it has been suggested that Prp38 and MFAP1 are important for a functional activation step, the underlying mechanism is poorly understood (Lybarger *et al.*, 1999; Schütze *et al.*, 2016; Ulrich & Wahl, 2017). The structure of the human B complex shows that the NTD of Prp38 as well as known α- helices of Snu23 and MFAP1 are in direct contact with the U6 ACAGAG/5'SS helix region (Fig. 1.12B). Therefore, involvement in repositioning the 5'SS during activation has been proposed for these proteins.

FBP21 is another member of the set of B-specific proteins. It is conserved among the higher eukaryotes but does not have any known counterpart in yeast. In the human B complex FBP21 appears to bind through its zinc finger domain to the minor groove of the U6 ACAGAG/5'SS helix and also to contact Brr2 (Fig. 1.12B), suggesting that FBP21 might be important for maintaining Brr2 in an inactivate state in the B complex. Consistently with this, a recent study suggested that the helicase activity of Brr2 is inhibited in the presence of FBP21 (Henning *et al.*, 2017). Interestingly, in spliceosomal complexes stalled with the small-molecule inhibitor cp028 (designated as the B028 complex), FBP21 is the only B-specific protein that is absent (Sidarovich *et al.*, 2017). This compound inhibits spliceosome maturation at an intermediate state after the release of the U4 snRNP, but before the release of the LSm and B-specific proteins and the recruitment of the Prp19/CDC5L complex, the B^{act}-specific proteins etc. It appears that with the release of FBP21 a physical barrier is removed, which makes possible the Brr2-driven activation of the spliceosome.

Taken together, these results suggest that FBP21 prevents premature unwinding of the U4/U6 interaction by inhibiting Brr2 activity in early B complexes. This essential role of FBP21 in the regulation of spliceosome assembly underlines both the importance of the precise control of Brr2 and the possibility that splicing factors specific for higher eukaryotes may be indispensable for basic splicing.
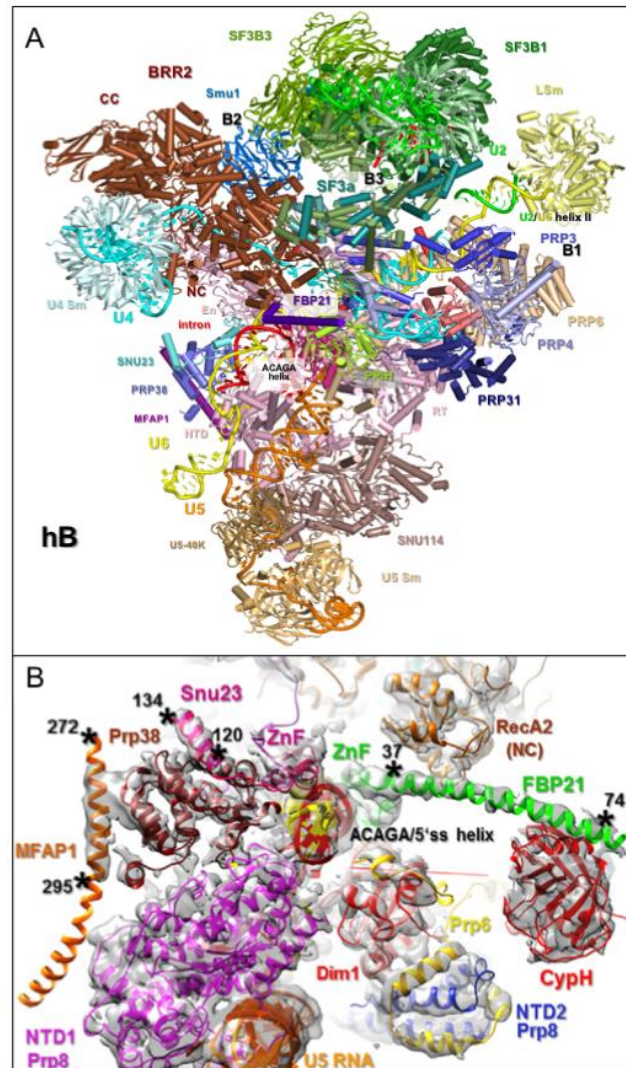
Figure 1.12. Structure of the human B complex. (A) Cryo-eletron microscopy structure of the human B complex. Green proteins and snRNA belong to the U2 head domain. Proteins and snRNAs with other colours are part of the tri-snRNP or belong to the B-specifc proteins. The three bridges are indicated by black letters. The proteins are labelled in the corresponding colours. (B) Locations of the B-specific proteins Prp38, Snu23, MFAP1 and FBP21 within the human B complex as determined by cryo-electron microscopy. This part of the figure was taken from Bertram, Agafonov, Dybkov *et al.*, 2017

Introduction

Like FBP21, Smu1 and RED are found in higher eukaryotes while being absent in yeast. These two proteins have been the targets of several studies, yet their function in splicing is poorly understood. Owing to their absence in simple organisms, it has been proposed that they may act as regulatory factors rather than having an essential core function. This is supported by a recent study that showed that siRNA-mediated depletion of Smu1 or RED changes the pattern of alternative splicing in human cells (Papasaikas *et al.*, 2015). Interestingly, alternative splicing appeared to be affected in a very similar manner by siRNA-mediated knock-down of Smu1 or RED, suggesting that these two proteins function as a single unit. Consistently with this, several other studies have provided evidence for a physical and functional connection between Smu1 and RED. For instance, a Smu1-RED complex was shown to be recruited by the influenza virus RNA polymerase whose transcripts appeared to be spliced differently upon depletion of RED/Smu1 (Fournier *et al.*, 2014). Furthermore, the siRNA-induced depletion of either Smu1 or RED in cells has been shown to result in co-depletion of the other, suggesting that the interaction between these two proteins improves the stability of both of them (Spartz *et al.*, 2004).
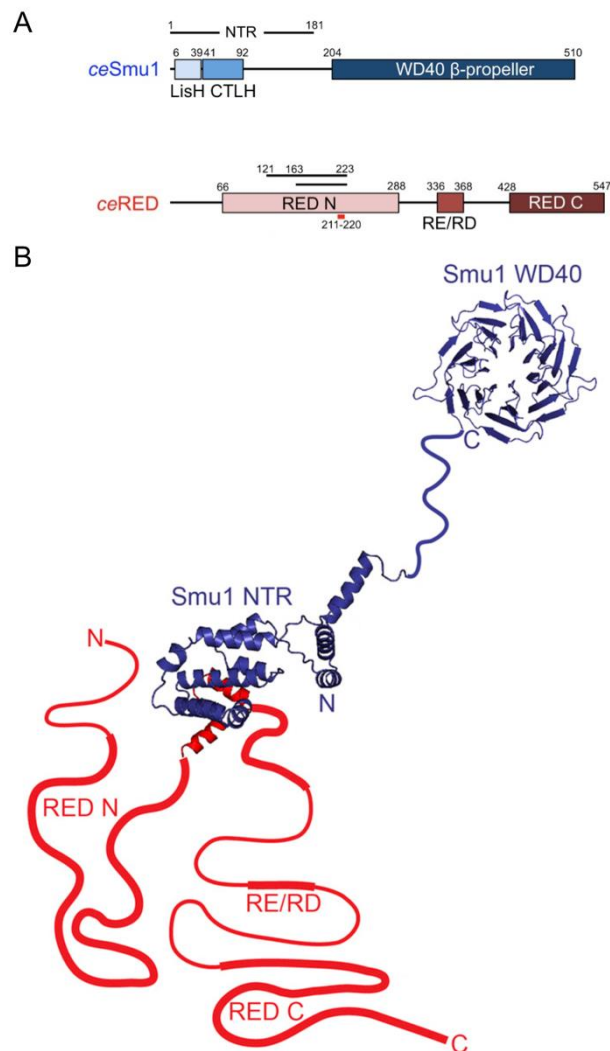
Figure 1.13. Structure of *C. elegans* Smu1 and RED. (A) Schematic illustration of Smu1 and RED from *C. elegans*. **(B)** The structures of the NTR of Smu1 and the short helices of RED that interact with Smu1 were determined by crystallography. The intrinsically disordered protein RED is illustrated schematically and a common folding of the Smu1-WD40 is shown. (Ulrich *et al.*, 2016)

Yeast two hybrid and structural studies further proposed that Smu1 and RED form a functional module that recruits further spliceosomal factors to the spliceosome (Hegele *et al.*, 2012). Apart of a LisH-CTLH homo-dimerization motif and a globular core domain, which is involved in hetero-dimer formation with RED, Smu1 further comprises a C-terminal WD40 domain. This domain is known to mediate protein-protein interactions and was thus suggested to serve as a binding platform for other proteins. RED, which is intrinsically disordered, was shown to form a short helical structure only upon interaction with Smu1 and was proposed to contact additional spliceosomal proteins in a similar manner (Fig. 1.13) (Ulrich *et al.*, 2016).

The cryo-EM structure of the human B complex has indicated a new possible role for Smu1 and RED - that is, in forming a bridge between U2 and the tri-snRNP in the B complex (Bertram, Agafonov, Dybkov *et al.*, 2017). Specifically, Smu1 directly contacts the WD40 domain 2 of U2-SF3B3 and the C-terminal cassette of Brr2 through its WD40 domain (Fig. 1.12A). It is possible that this interaction stabilizes the position of Brr2 relative to its U4/U6 substrate and might be important during the activation. The precise location of RED in the B complex could not be determined since this intrinsically disordered protein lacks clear structural domains. However, chemical crosslinking experiments suggest that RED makes contact with several U2 proteins and with Smu1 as well as with Prp8. This shows that RED and Smu1 engage in several contacts within the spliceosome and may thus jointly stabilize the interaction between the U2 snRNP and the tri-snRNP in the human B complex.

1.11 Aims

The B-specific proteins interact transiently with the spliceosome at the B complex level. While Prp38, MFAP1 and Snu23 are highly conserved and have been proposed to be essential for the activation of the spliceosome, FBP21, Smu1 and RED are only shared among higher eukaryotes and little is known about their function. Recently, two independent studies came to the conclusion that FBP21 might also help to keep the Brr2 helicase inactive at the critical stage of preparing the spliceosome for its activation (Bertram, Agafonov, Dybkov *et al.*, 2017; Henning *et al.*, 2017). All in all, it appears that the B-specific proteins, conserved or not, are important for supporting or regulating the Brr2-driven rearrangements.

The role of Smu1 and RED, however, remains poorly understood. Although these two proteins have recently been the objects of several studies, each study led to different proposed functions in splicing, so that no clear picture has yet emerged. While Smu1 and RED were first thought to be regulators of alternative splicing (Papasaikas *et al.*, 2015) or mediators of protein–protein interactions (Hegele *et al.*, 2012), more recent considerations rather attribute to them a role in spliceosome stability (Bertram, Agafonov, Dybkov *et al.*, 2017). The one point on which structural and functional studies unanimously agree is that the two proteins undergo a mutual physical interaction that results in a combined, joint function.

Introduction

Therefore, my thesis addresses the question of the role that Smu1 and RED actually play during splicing. Currently, it is widely accepted that Smu1 and RED are auxiliary factors that control alternative splicing. However, their transient association with the B complex might indicate an involvement in the formation or stabilization of this assembly intermediate. To reveal whether Smu1 and RED function as regulatory or general splicing factors, I set out to seek convincing evidence by investigating splicing and spliceosome assembly *in vivo* and *in vitro*. It has been proposed several times that Smu1 and RED act as a unit during splicing. To understand the importance and nature of this interaction, I further investigated the individual contribution of each protein to their joint function in splicing. A final elucidation of the role of Smu1 and RED within the spliceosome will help us to understand why higher eukaryotes need a more complex spliceosomal protein composition, and how these additional factors enable the spliceosome to fulfil its tasks in more complicated organisms.

2. Materials and methods

2.1 Materials

2.1.1 Chemicals

Chemical	**Supplier**
Acetic acid	Merck, Germany
Agarose (low melting point)	Thermo Fisher Scientific, USA
Agarose	Serva, Germany
Ammonium peroxidisulphate (APS)	Serva, Germany
Ampicillin	Roth, Germany
Boric acid	Merck, Germany
Bromphenol blue	Serva, Germany
Calcium chloride dehydrate ($CaCl_2 \cdot 2H_2O$)	Merck, Germany
Coomassie brilliant blue G-250	Serva, Germany
Creatine phosphate	Sigma-Aldrich, Germany
Dimethyl pimelimidate (DMP)	Thermo Fisher Scientific, USA
Dimethylsulphoxide (DMSO)	Roth, Germany
Dithiobis succinimidyl propionate (DSP)	Thermo Fisher Scientific, USA
Dithiothreitol (DTT)	Roth, Germany
DNA molecular weight marker (2log, 1kb)	Roth, Germany
DNA molecular weight marker (5kb, 100bp)	New England Biolabs, Germany
EDTA (disodium salt dihydrate)	Roth, Germany
Ethanol	Merck, Germany
Ethidium bromide	Serva, Germany
Ficoll 400	Sigma-Aldrich, Germany
Fish sperm DNA (10mg/mL)	Roche, Germany
Formamide	Merck, Germany
Gentamycin	Roth, Germany
Glycerol	Merck, Germany
Glycine	Merck, Germany
GlycoBlue™ Coprecipitant	Thermo Fisher Scientific, USA
Glycogen (20 mg/mL)	Sigma-Aldrich, Germany
Heparin (sodium salt)	Roth, Germany
HEPES (N-2-Hydroxyethylpiperazin-N-2-ethansulfonic acid)	Sigma-Aldrich, Germany
Hydrochloric acid fuming 37%	Merck, Germany
Imidazole	Serva, Germany

Materials and Methods

IPTG	Roth, Germany
Isopropanol	Merck, Germany
Kanamycin	Roth, Germany
Maltose	Merck, Germany
Magnesium chloride hexahydrate ($MgCl_2 \cdot 6H_2O$)	Merck, Germany
Methanol	Merck, Germany
β-Mercaptoethanol	Roth, Germany
Milk powder	Roth, Germany
MOPS	Roth, Germany
Nonidet P-40	
Penicillin/Streptomycin (10 mg/mL)	Thermo Fisher Scientific, USA
PMSF (Phenylmethylsulfonylfluoride)	Roche, Germany
Poly (ethylene glycol)-block-poly(propylene glycol)-block- poly(ethylene glycol) (PEG-PPG) average Mn ~ 14000	Sigma-Aldrich, Germany
Polyvinylpyrrolidone average Mn 10000	Sigma-Aldrich, Germany
Ponceau S	Serva, Germany
Potassium chloride (KCl)	Merck, Germany
Potassium dihydrogen phosphate (KH_2PO_4)	Merck, Germany
Potassium hydroxide (KOH)	Merck, Germany
Pre-stained protein-molecular weight marker	Bio-Rad, Germany
Roti phenol-chloroform-isoamyl alcohol (PCI)	Roth, Germany
Rotiphorese Gel 30 (37.5:1) Polyacrylamide	Roth, Germany
Rotiphorese Gel 40 (19:1) Polyacrylamide	Roth, Germany
Sodium acetate	Merck, Germany
Sodium azide	Merck, Germany
Sodium chloride (NaCl)	Merck, Germany
Sodium dodecyl sulfate (SDS)	Serva, Germany
Sodium hydroxide (NaOH)	Merck, Germany
Spermidine	Sigma-Aldrich, Germany
SYBR® Gold Nucleic Acid Gel Stain	Thermo Fisher Scientific, USA
TEMED (N, N, N', N'-Tetramethylethylendiamine)	Sigma-Aldrich, Germany
Tetracyclin	Roth, Germany
Triethanolamine	Sigma-Aldrich, Germany
Tris [Tris-(hydroxymethyl)aminomethane]	VWR International, Germany
Triton X-100	Sigma-Aldrich, Germany
tRNA from *E. coli*	Roche, Germany
Tween-20	Roth, Germany
Urea	Merck, Germany

X-gal

Xylene cyanol FF   Sigma-Aldrich, Germany

2.1.2 Consumables

Consumable	Supplier
6-well plates	Greiner-Bio-One, Germany
Biodyne B Membrane	Pall Corporation, Germany
Cell scraper	Sarstedt, Germany
Centrifugal filters, Amicon Ultra	Millipore, USA
Centrifuge tubes (TH660)	Seton scientific, USA
Centrifuge tubes (TST41.14)	Beranek Laborgeräte, Germany
Dialysis membranes (MWCO 6000-8000 Da)	SpektraPor, USA
Electroporation cuvettes	Bio-Rad, Germany
NuPAGE™ gels (1.5 mm, 4–12%)	Thermo Fisher Scientific, USA
Nesco-/Parafilm	Roth, Germany
Petri dishes (for bacteria)	Sarstedt, Germany
Greiner CELLSTAR® dish (for adhesive cell lines)	Greiner, Germany
Greiner CELLSTAR® serological pipette	Greiner, Germany
Greiner culture tubes (14 mL)	Greiner, Germany
Protran BA 83 nitrocellulose	GE Healthcare, UK
ReadyLyzer 3, MWCO 6–8 kDa	Serva, Germany
Safe-lock tubes (0.5, 1.5 and 2 mL)	Eppendorf, Germany
Standard Polystyrene Semi-micro Cuvettes	Sarstedt, Germany
Sterile filters (0.2 µm or 0.45 µm)	Sarstedt, Germany
Surgical blades	Martin, Germany
Tips, disposable	StarLab, UK
Tubes (15 mL and 50 mL)	Sarstedt, Germany
Whatman 3MM paper	Whatman, UK

Materials and Methods

2.1.3 Chromatographic resins and columns

Resin / column	Supplier
Amylose resin	New England Biolabs, Germany
Chromatography columns	Bio-Rad, Germany
Dynabeads™ Protein A	Thermo Fisher Scientific, USA
Roti®garose-His/Ni Beads	Roth, Germany
ProbeQuant™ G-50 micro columns	GE Healthcare, UK

2.1.4 Commerical kits and buffers

Kit	Supplier
BCA™ protein assay kit	Thermo Fisher Scientific, USA
ECL Western Blot Detection Kit	GE Healthcare, USA
High Pure Plasmid Isolation Kit	Roche, Germany
Lipofectamine RNAiMAX Transfection Reagent	Thermo Fisher Scientific, USA
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel, Germany
Prime It II random-primer labeling kit	Agilent Technologies, USA
PureLink™ HiPure Plasmid Filter Maxiprep Kit	Thermo Fisher Scientific, USA
X-tremeGENE™ 9 DNA Transfection Reagent	Sigma-Aldrich, Germany

Buffer	Supplier
Gel Loading Dye, Purple (6X), no SDS	New England Biolabs, Germany
Phusion® HF Buffer Pack (5x)	New England Biolabs, Germany

2.1.5 Machines

Machine	Supplier
Autoclaves	H+P Labortechnik, Germany
Axiovert 25 Inverted Microscope	Zeiss, Germany
Biological safety Cabinets Hera Safe	Thermo Fisher Scientific, USA
Biofuge fresco	Heraeus, Germany
Biofuge pico	Heraeus, Germany
Bioruptor Sonicator	Diagenode, Belgium
Casy Cell Counter + Analyzer System Model TT	Innovatis, Germany
Centrifuge 5424/5424 R	Eppendorf, Germany
CO_2 incubator BBD 6220	Heraeus, Germany

EG&G, and radiation hand-foot monitor	Berthold, Germany
Electroporation system MicroPulser	Bio-Rad, Germany
Fiberlite™ F14-14 x 50cy Fixed-Angle Rotor	Thermo Fisher Scientific, USA
Gel documentation unit	Bio-Rad, Germany
Gel electrophoresis apparatus	Bio-Rad, Germany
Gel electrophoresis apparatus	In-house
Gel dryer model 583	Bio-Rad, Germany
Glass cell culture flasks	Duran Group, Germany
Glass-ware	VWR International, Germany
Gradient master model 106	BioComp Instruments, Canada
Head-over-tail rotor	Cole-Parmer, USA
Heating blocks HB-130	Unitek, USA
Hybridization oven	Hybaid Biometra, UK
Ice machine	Ziegra, UK
IKA-Vibrax-VXR table-top shaker	IKA-Werke, Germany
KB incubator	Binder, Germany
LB122, contamination monitor	Berthold, Germany
MaxQ 3000 Benchtop Shaker	Thermo Fisher Scientific, USA
Megafuge 1.0R	Kendro, USA
Milli-Q-water supply apparatus	Millipore, USA
Mini-Monitor Series 900 Radiation Monitor	Mini Instruments, UK
Mini Protean 3 Electrophoresis Cell	Bio-Rad, USA
Mini Trans-Blot Electrophoretic Transfer Cell	Bio-Rad, USA
Multifuge 1S-R	Heraeus, Germany
Multitron incubator shaker	INFORS HT, Swiss
Orion 3 Star pH meter	Thermo Fisher Scientific, USA
Phosphorimager Typhoon 8600	Amersham Pharmacia, Germany
Power supply EPS 2A200	Hoefer Pharmacia Biotech, USA
Power supply EPS 3501/XL	Amersham Pharmacia, Germany
Q Exactive™ HF Hybrid Quadrupol-Orbitrap™ Mass Spectrometer	Thermo Fisher Scientific, USA
Rocking Platform WT15	Biometra, Germany
Scintillation counter LS 1701	Packard, USA
Sonifier W-250D	Heinemann, Germany
Sorvall™ LYNX™ Superspeed Centrifuge	Thermo Fisher Scientific, USA
Sorvall TH660 rotor	Kendro, USA
Sorvall TST41.14 rotor	Kendro, USA
Sorvall Ultracentrifuge Discovery 90 SE	Sorvall/Kendro, USA
Sorvall WX Ultracentrifuge	Sorvall/Kendro, USA

Materials and Methods

Speed Vac Konzentrator 5301	Eppendorf, Germany
Spectrophotometer Nanodrop ND-1000	Thermo Fisher Scientific, USA
Spectrophotometer Ultrospec 3000 pro	GE Healthcare, USA
Test tube rotator	Snijders, The Netherlands
T3 Thermocycler	Biometra, Germany
Thermomixer comfort	Eppendorf, Germany
Trans-Blot Electrophoretic Transfer Cell	Bio-Rad, Germany
UV-Stratalinker 2400	Stratagene, USA
Vortex	Scientific industries, USA
X-ray film developer X-Omat 2000	Kodak, USA

2.1.6 Nucleotides

Nucleotide	Supplier
Nucleoside-5'-triphosphate (100 mM): ATP, CTP, UTP, GTP	Pharmacia, Germany
Deoxynucleoside-5'-triphsophate (100 mM): dATP, dCTP, dTTP, dGTP	Pharmacia, Germany
7-monomethyl-diguanosine triphosphate (m^{7}G(5')ppp(5')G-cap)	Kedar, Poland

2.1.7 Radiolabeled nucleotides

Nucleotide	Supplier
α-^{32}P-UTP [10 μCI/μl, 3000 Ci/mmol]	Perkin-Elmer, Germany
α-^{32}P-dATP [10 μCI/μl, 3000 Ci/mmol]	Perkin-Elmer, Germany

2.1.8 Oligonucleotides

Name	Sequence (5' → 3')
Smu1-F	CAATCACTCGACGAAGAC
Smu1-R	CTTCCGTGTTTCAGTTAGC

Strep-D-WD40-R	ATTATGCATTCATTATTTTTCGAACTGAGGG
	TGGGACCACTGTCCGAACTTGATGTGCC
Smu1-His-F	ATACGAATTCCATATGTCTATCGAAATCGAGTCC
Smu1-His-R	ATACCTGCAGGATTTAAGGCTTCCACAGCTTCAGCA
RED-F	ATGCGAATTCAGCATGGGACCTGAAAGG
RED-R	ATGCCTGCAGGATGCTCATTAGTACTTAG
D-N-RED #2-F	ATACGAATTCCATATGAACTACAGGGCCGTCGGACCTACCG
D-C-RED #1-R-Stop	ATACCTGCAGGATCTAGGAGTAGTCCACTTCTTCGTCA
Primer 550 (F)	CCATCTCGCAAATAAATAAG
Primer 551 (R)	CATTTTATGTTTCAGGTTC
Primer 552 (F)	TCATAGCGCGGGTTCCTTCC
Primer 553 (R)	CTTTAATTCAACCCAACACA
MINX-70 Primer A	TCACACAGGAAACAGCTATGAC
MINX-70 Primer B	TCTTACCGTTCGGAGG
MINX-70 Primer C	CCTCCGAACGGTAAGAGGGCGCAGTAGTCCAG
MINX-70 Primer D	GTAAAACGACGGCCAGTG
MINX-80 Primer A	TCACACAGGAAACAGCTATGAC
MINX-80 Primer B	AGTTCTACATGCTAGGCTCTTACC
MINX-80 Primer C	GGTAAGAGCCTAGCATGTAGAACT GTAGTCCAGGGTTTCCTT
MINX-80 Primer D	GTAAAACGACGGCCAGTG
MINX-90 Primer A	TCACACAGGAAACAGCTATGAC
MINX-90 Primer B	AGTTCTACATGCTAGGCTCTTACC
MINX-90 Primer C	GGTAAGAGCCTAGCATGTAGAACT CTAGGGCGCAGTAGTCCAG
MINX-90 Primer D	GTAAAACGACGGCCAGTG
MINX-150 Primer A	TCACACAGGAAACAGCTATGAC
MINX-150 Primer B	AGACTGAGACTGAGACTGAGGGCTCTTACCGTTCGGAG
MINX-150 Primer C	CTCAGTCTCAGTCTCAGTCTCAGTCTGCCTgtagaactggttacctgcagcc
MINX-150 Primer D	GTAAAACGACGGCCAGTG
MINX-80-cleaved Primer A	TCACACAGGAAACAGCTATGAC
MINX-80-cleaved Primer B	TGAGACTGAGACTGAGGGCTC
MINX-80-cleaved Primer C	<u>TAATACGACTCACTATAGGG</u>TCTCAGGGTTTCCTTGATG
MINX-80-cleaved Primer D	GTAAAACGACGGCCAGTG
PM5$_{56}$ and PM5-10$_{56}$ Primer A	TCACACAGGAAACAGCTATGAC
PM5$_{56}$ and PM5-10$_{56}$ Primer B	GTAAGCTTGATACATACCTTGGC
PM5$_{56}$ and PM5-10$_{56}$ Primer C	GCCAAGGTATGTATCAAGCTTACGTGACTGATAGAACACTACCTG
PM5$_{56}$ and PM5-10$_{56}$ Primer D	GTAAAACGACGGCCAGTG

Materials and Methods

U1 snRNA forward K107	GATACTTACCTGGCAGGGGAG
U1 snRNA reverse K50	CGCGGATCCAGGGGAAAGCGCGAACGCAGTC
U2 snRNA forward K78	CCTAATACTCACTATAGATCGCTTCTCGGCCTTTTGCG
U2 snRNA reverse K79	GGGTGCACCGTTCCTGGAGGTAC
U4 snRNA forward K47	GGGAATTCCTAATACGACTCACTA
U4 snRNA reverse K48	CGCGAATCCAGTCTCCGTAGAGAC
U5 snRNA forward K115	GATACTCTGGTTTCTCTTCAG
U5 snRNA reverse KJ7	CCCAAGCTTTAGCCTTGCCAAGGCAAGG
U6 snRNA forward K54	CCTAATACGACTCACTATAGGTGCTCGCTTCGGCAGC
U6 snRNA reverse K55	AAAAATATGGAACGCTTCACG

2.1.9 siRNAs

siRNA	Target Sequence	Supplier
Prp38A-1	CTGCTGTTATTCAAAGATCAA	Laboratory of R. Lührmann, MPI-bpc
Prp38A-2	TCCACGGACATCGATAATCAA	Laboratory of R. Lührmann, MPI-bpc
MFAP1-1	AAGGACCGAGTGACAGTTCAA	Laboratory of R. Lührmann, MPI-bpc
MFAP1-2	TACGGCGTTTACAGAACCGTA	Laboratory of R. Lührmann, MPI-bpc
RED-1	AACCCGTAACAAGAAGCTTAA	Laboratory of R. Lührmann, MPI-bpc
RED-2	CAGCGAGTATATGAACAACAA	Laboratory of R. Lührmann, MPI-bpc
Smu1-1	CACAACTGAGCAGGCATATTA	Laboratory of R. Lührmann, MPI-bpc
Smu1-2	TACGGCTATGTCGATCGAAAT	Laboratory of R. Lührmann, MPI-bpc
Smu1-3	GTAAAGATCTGGAATATGA	Laboratory of R. Lührmann, MPI-bpc
AllStars Negative Control siRNA		QIAGEN

2.1.10 Plasmids

Name	Description	Reference
hSmu1_pMK-RQ	Synthetic gene of the full-length Smu1; Kan^R	Geneart, Germany
RED-IK_pMK-RQ	Synthetic gene of the full-length RED; Kan^R	Geneart, Germany
pFL	Multibac transfer vector (acceptor) for bacmid preparation; Gen^R and Amp^R	Dr. Imre Berger, EMBL, Grenoble
pFL-RED-His$_{10}$	N-His$_{10}$, RED full-length gene inserted via EcoRI/PstI; Gen^R and Amp^R	This study
pFL-ΔN-RED- His$_{10}$	N-His$_{10}$, N-terminally truncated RED gene inserted via EcoRI/PstI; Gen^R and Amp^R	This study

36

pFL-RED-ΔC- His$_{10}$	N-His$_{10}$, C-terminally truncated RED gene inserted via EcoRI/PstI; Gen^R and Amp^R	This study
pFL-ΔΔ-RED- His$_{10}$	N-His$_{10}$, N- and C- terminally truncated RED gene inserted via EcoRI/PstI; Gen^R and Amp^R	This study
pFL-Smu1-StrepII	C-StrepII, Smu1 full-length gene inserted via XhoI/NsiI; Gen^R and Amp^R	This study
pFL-Smu1-ΔWD40-StrepII	C-StrepII, Smu1 WD40 domain deletion mutant gene inserted via XhoI/NsiI; Gen^R and Amp^R	This study
pFL-Smu1- His$_{10}$	N-His$_{10}$, Smu1 full-length gene inserted via EcoRI/PstI; Gen^R and Amp^R	This study
pMINX	Plasmid encoding MINX pre-mRNA; Amp^R	(Zillmann *et al.*, 1988)
pPM5	Plasmid encoding PM5 pre-mRNA; Amp^R	(Bessonov *et al.*, 2008)
pPM5-10	Plasmid encoding PM5-10; Amp^R	(Bessonov *et al.*, 2010)

2.1.11 Antibodies

Antibody	**Supplier**
Anti-hSnu66	Laboratory of R. Lührmann, MPI-bpc (Makarova *et al.*, 2001)
Anti-hPrp31, phospho-specific	Laboratory of R. Lührmann, MPI-bpc (Schneider, Hsiao *et al.*, 2010)
Anti-P-SF3b155 (T313)	Laboratory of R. Lührmann, MPI-bpc (Girard *et al.*, 2012)
Anti-Prp31	Laboratory of R. Lührmann, MPI-bpc (Makarova *et al.*, 2002)
Anti-SF3b155	Laboratory of R. Lührmann, MPI-bpc (Will *et al.*, 2001)
Anti-RED	Laboratory of R. Lührmann, MPI-bpc
Anti-MFAP1	Laboratory of R. Lührmann, MPI-bpc
Anti-FBP21	Santa Cruz Biotechnology, USA
Anti-hSmu-1	Santa Cruz Biotechnology, USA
Anti-hPrp38	Laboratory of R. Lührmann, MPI-bpc
Anti-hSnu23	Laboratory of R. Lührmann, MPI-bpc
Anti-H3 (Histone 3)	Abcam, UK
Anti-Snu114	Laboratory of R. Lührmann, MPI-bpc (Fabrizio et al.,1997, Sidarovich et al., 2017)
Anti-SF3A2	Laboratory of R. Lührmann, MPI-bpc (Will et al., 2002)
Goat-anti-rabbit (horseradish peroxidase coupled)	Jackson Immunoresearch, USA
Goat-anti-mouse (horseradish peroxidase coupled)	Jackson Immunoresearch, USA

Materials and Methods

2.1.12 Proteins, enzyme inhibitors and enzymes and

Protein	Supplier
Bovine serum albumin (BSA), acetylated	Sigma-Aldrich, Germany
MS2-MBP	Laboratory of R. Lührmann, MPI-bpc

Enzyme inhibitor	Supplier
Complete protease inhibitor, EDTA-free	Roche, Germany
PhosSTOP™	Roche, Germany
RNase Inhibitor [40 U/µl]	Molox, Germany

Enzyme	Supplier
Accutase® solution	Sigma-Aldrich, Germany
Pfu Tripple DNA Polymerase	Laboratory of D. Görlich, MPI-bpc
Proteinase K (10 mg/mL)	Bioron, Germany
Restriction endonucleases	New England Biolabs, Germany
RQ1 DNase [1 U/µl]	Promega, USA
SP6 RNA Polymerase	New England Biolabs, Germany
T4 DNA Ligase	Laboratory of D. Görlich, MPI-bpc
T7 RNA Polymerase [20 U/µl]	Laboratory of D. Görlich, MPI-bpc
Yeast inorganic pyrophosphatase (YIPP) [0.1 U/µl]	New England Biolabs, Germany

2.1.13 Bacteria strains

Strain	Supplier
Escherichia coli XL-10 Gold	Agilent Technologies, USA
Escherichia coli DH10MultiBacY	Dr. Imre Berger, EMBL, Grenoble

2.1.14 Cell lines

HeLa S3 cells were used for the preparation of nuclear extract. HeLa SS6 cells were used for RNAi experiments.

The *Spodoptera frugiperda* cell line Sf9 was used for recombinant baculovirus propagation and isolation.

The *Trichoplusia ni* cell line High Five was used for recombinant protein expression.

Cell line	Supplier
HeLa S3 cells (human cervical cancer cells)	GBF, Germany
HeLa SS6 cells (human cervical cancer cells)	ATCC, USA
High Five™ cells	Invitrogen, Germany
Sf9 cells	Invitrogen, Germany

2.1.15 Commercial media

Medium	Supplier
Dulbecco's Modified Eagle Medium (DMEM)	Thermo Fisher Scientific, USA
ESF 921 Insect Cell Culture Medium, Protein Free	Expression Systems, USA
Gibco™ Sf-900™ III SFM	Fisher Scientific, USA
LB Agar Medium (Powder)	MP Biomedicals, USA
LB Medium (Capsules)	MP Biomedicals, USA
Opti-MEM - Reduced Serum Media	Thermo Fisher Scientific, USA

2.1.16 Buffers, media and solutions

Media, buffers and solutions were prepared with deionized water (Millipore) and autoclaved if necessary (121 °C, 20 min/15 min 1 bar). Solutions with heat-labile components were filter-sterilized (0.22 µm).
The pH was adjusted with 37% HCl, 5 M NaOH or 5–10 M KCl.

Coomassie staining solution

100 µM	Coomassie brilliant blue G-250
40 mM	HCl

Denhardt's Solution (100x)

2%	BSA
2%	SDS
2%	Polyvinylpyrrolidone

10x G-75 buffer

200 mM	HEPES
750 mM	KCl
15 mM	$MgCl_2$

pH adjusted to 7.9

Materials and Methods

10x G-150 buffer

200 mM	HEPES
1.5 M	KCl
15 mM	$MgCl_2$

pH adjusted to 7.9

Gel fixing solution

10% (v/v)	Acetic acid
40%	Ethanol or Methanol

HeLa Cell Lysis Buffer

2 mM	EDTA, pH 8.0
150 mM	NaCl
30 mM	Tris-HCl, pH 7.5
1%	Triton X-100
2x	PhosSTOP™
2x	Complete protease inhibitor
1 mM	$MgCl_2$

HeLa IP Dilution Bufer

235 mM	NaCl
30 mM	Tris-HCl, pH 7.5
2%	Triton X-100
2x	PhosSTOP™
2x	Complete protease inhibitor
3 mM	$MgCl_2$
0.2 U/µL	RNasin

HeLa IP Wash Bufer

200 mM	NaCl
30 mM	Tris-HCl, pH 7.5
1%	Triton X-100
1.5 mM	$MgCl_2$

HeLa Sonication Buffer

2 mM	EDTA, pH 8.0
60 mM	Tris-HCl, pH 7.5
2%	PEG-PPG
2x	PhosSTOP™
2x	Complete protease inhibitor

HeLa IP Dilution Bufer

2 mM	EDTA, pH 8.0
30 mM	Tris-HCl, pH 7.5
1%	Triton X-100
2x	PhosSTOP™
2x	Complete protease inhibitor
0.3 U/µL	RNasin

High Five Lysis Buffer

50 mM	HEPES-KOH, pH 7.9
500 mM	NaCl
15 mM	Imidazole, pH 8.0
0.1%	NP-40
15%	Glycerol
2 mM	DTT

High Five Dilution Buffer

50 mM	HEPES-KOH, pH 7.9
50 mM	NaCl
15%	Glycerol
2 mM	DTT

High Five Wash Buffer

50 mM	HEPES-KOH, pH 7.9
750 mM	NaCl
20 mM	Imidazole
15%	Glycerol
2 mM	DTT

Materials and Methods

High Five Elution Buffer

20 mM	HEPES-KOH, pH 7.9
200 mM	NaCl
300 mM	Imidazole
15%	Glycerol
2 mM	DTT

20x MOPS buffer

1 M	Tris
1 M	MOPS
2%	SDS
20 mM	EDTA

5x Native gel loading dye

90 mM	Tris
90 mM	Boric acid
2.5 mM	EDTA, pH 8.0
30% (v/v)	Glycerol
0.05% (w/v)	Bromphenol blue

Northern blot transfer buffer

176 mM	NaH_2PO_4
90 mM	Na_2HPO_4

Northern blot hybridization solution

50% (v/v)	Formamide
10x	SSC
0.5%	SDS
30 mM	Tris-HCl, pH 7.5
5x	Denhardt's Solution

Northern blot wash buffer

2x	SSC
0.1%	SDS

10x PBS (-T)

1.3 M	NaCl
27 mM	KCl
80 mM	Na_2HPO_4
15 mM	KH_2PO_4
0.05%	Tween-20 (optional)

PK Buffer

50 mM	Tris-HCl, pH 7.5
70 mM	NaCl
0.2 %	SDS
1 mM	$CaCl_2$
0.1 U/µL	RNase inhibitor

2x RNA loading dye

80%	Formamide
1 mM	EDTA, pH 8.0
0.05% (w/v)	Bromphenol blue
0.05% (w/v)	Xylene cyanol

Roeder D buffer

10% (v/v)	Glycerol
20 mM	HEPES-KOH, pH 7.9
100 mM	KCl
1.5 mM	$MgCl_2$
0.2 mM	EDTA, pH 8.0
0.25 mM	DTT
0.25 mM	PMSF

6x SDS Loading Dye

244 mM	Tris-HCl, pH 7.5
2.4%	SDS
250 mM	DTT
21.2% (w/v)	Glycerol
0.12% (w/v)	Bromphenol blue

Materials and Methods

SDS-PAGE running buffer

25 mM	Tris-HCl, pH 6.8
192 mM	Glycine
1% (w/v)	SDS
21.2% (w/v)	Glycerol
0.12% (w/v)	Bromphenol blue

20x SSC buffer

3 M	NaCl
300 mM	Trisodium citrate

pH adjusted to 7.9

10x T4 DNA ligase buffer

500 mM	Tris-HCl, pH 7.5
100 mM	$MgCl_2$
100 mM	DTT
10 mM	ATP
250 µg/mL	BSA

10x TBE

0.89 M	Tris-HCl, pH 7.5
0.89 M	Boric acid
25 mM	EDTA, pH 8.0

10x TBS-T

200 mM	Tris-HCl, pH7.5
1.5 M	NaCl
1%	Tween-20

Triethanolamine solution

1x	PBS
200 mM	Triethanolamine

pH adjusted to 8.6

5x TRO buffer

1 M	HEPES-KOH, pH 7.5
200 mM	DTT
160 mM	$MgCl_2$
10 mM	Spermidine

Western blot blocking buffer

1x	TBS-T
5% (w/v)	Milk powder

Western blot stripping solution

62.5 mM	Tris-HCl, pH 6.7
2% (w/v)	SDS
85 mM	β-Mercaptoethanol

10x Western blot transfer buffer

500 mM	Tris-HCl, pH7.5
400 mM	Glycine
0.1% (w/v)	SDS

2.2 Methods

2.2.1 Molecular biology standard methods

2.2.1.1 PCR amplification

DNA fragments were amplified in a volume of 50 µL by PCR as illustrated in Tables 2.1 and 2.2.

Table 2.1: Composition of a standard PCR

Component	Stock conc.	Final conc.
HF Buffer	5x	1x
dNTP mix	10mM	200 µM
Primer 1 (Fw)	10 µM	200 nM
Primer 2 (Rev)	10 µM	200 nM
Template (plasmid)		1 ng/µL
Pfu Tripple DNA Pol.	100 ng/µL	1 ng/µL

Materials and Methods

Table 2.2: Program of a standard PCR

Step	Temperature	Time	
Initial denaturation	95°C	3 min	
Denaturation	95°C	45 sec	
Annealing	Primer specific	40 sec	31 cycles
Extension	72°C	1-2 min/kb	
Final extension	72°C	15 min	

2.2.1.2 Restriction digest of DNA

Restriction digests of PCR products and plasmids were performed in a reaction volume of 50 µL at 37 °C overnight (Table 2.3). When possible, PCR products were digested directly in PCR reaction buffer. For digestions with enzymes incompatible with this buffer, PCR products were purified by using the NucleoSpin® Gel and PCR Clean-up kit from Macherey-Nagel before digestion.

Table 2.3: Composition of a standard restriction digest

Component	Stock conc.	Final conc.
Cut Smart Buffer	10x	1x
Enzyme 1		10 U/µg DNA
Enzyme 2		10 U/µg DNA
DNA		1–5 µg

Digested PCR products were purified directly by using the NucleoSpin® Gel and PCR Clean-up kit, while plasmids were resolved on a preparative agarose gel and the band of interest was used for gel extraction with the NucleoSpin® Gel and PCR Clean-up kit.

2.2.1.3 Ligation of digested insert and vector

The composition of a standard ligation reaction is illustrated in Table 2.4. The ligation was performed in 20 µL reaction volume at 16 °C overnight and immediately transformed into chemically competent *E. coli* cells.

Table 2.4: Composition of a standard ligation reaction

Component	Stock conc.	Final conc.
T4 DNA ligase buffer	10x	1x
T4 DNA ligase	1U/µL	0.1 U/µL
Vector		50-100 ng
Insert		300 ng -1.5 µg

2.2.1.4 Nucleic acid quantification

To determine the nucleic acid concentration of an aqueous solution, its absorption at wavelengths of 260 nm and 280 nm was measured by using a NanoDrop spectrophotometer according to the manufacturer's instructions. The respective buffer without any nucleic acids served as reference for the measurement.

The ratio of OD260/OD280 indicates the purity of the solution. Thereby, a ratio of 2.0 is considered to correspond to pure RNA or oligonucleotides and 1.8 to pure DNA. Ratios below 1.8 indicate the presence of impurities like protein, phenol or other contaminants absorbing at 280 nm.

The following equations were used to determine the concentrations:

1 OD260: 50 µg/ml double-stranded DNA

1 OD260: 33 µg/ml single-stranded DNA

1 OD260: 40 µg/ml single-stranded RNA

2.2.1.5 PCI extraction

To separate nucleic acids from proteins, a phenol-chloroform-isoamyl alcohol (PCI) extraction was performed. During this procedure, the proteins are denatured by the phenol/chloroform mixture and retained in the organic phase, whereas the nucleic acids remain in the aqueous phase. The reactions' volume was adjusted to 200 µl and an isovolume of PCI was added; the samples were thoroughly mixed on a table-top shaker for 10 min. The suspension was centrifuged for 10 min at room temperature at 13000 rpm in a microfuge to separate the aqueous from the organic phase. The aqueous phase was collected and supplemented with 0.1 volumes 3 M NaOAC pH 5, 3 volumes absolute ethanol and if required 1 µL of GlycoBlue™ Coprecipitant. Precipitation was performed at -20 °C for at least an hour, usually overnight. The nucleic acids were pelleted by centrifugation at 13 000 rpm and 4 °C for 45 min in a microfuge and the pellet was washed once with 75% (v/v) ethanol. The pellet was then air-dried and resuspended in water or 1x RNA Loading dye.

2.2.1.6 Proteinase K digestion

To prune RNAs of associated proteins, a proteinase K digest was performed before the PCI extraction.
PK buffer and 0.5 µg/µl proteinase K was added to the RNA sample and incubated at 37 °C for 30 min.
The RNA was subsequently isolated by PCI extraction as described in (2.2.1.5).

Materials and Methods

2.2.1.7 Generation of truncated MINX, PM5 and PM5-10 constructs

The truncated MINX constructs MINX-70, -80 and -90 were generated by fusion PCR. The MINX pre-mRNA, cloned in PUC18 plasmid, served as template. In brief, a set of four primers was designed. Two primers were placed upstream and downstream of the MINX construct (Fig. 2.1 primers A and D), while the two other primers flanked the sequence to be deleted (primers B and C). In the first step, the fragments upstream (fragment 1) and downstream (fragment 2) of the region to be deleted were amplified separately. The 5' end of the forward primer (primer C), used to amplify the fragment 2 was equipped with a 20-nucleotide stretch complementary to the 3' extremity of the fragment 1. The gel-purified fragments served as template in a second PCR reaction. Here, the two fragments were first annealed, and this was followed by synthesis of the sequences complementary to the 5' overhangs by the Pfu DNA polymerase. The resulting constructs, containing the deletion, were subsequently amplified in a third PCR using the primers upstream and downstream of the MINX sequence (primers A and D). The final PCR product was gel-purified and used as template for *in vitro* transcription. $PM5_{56}$ and $PM5\text{-}10_{56}$ were generated by following a similar strategy.
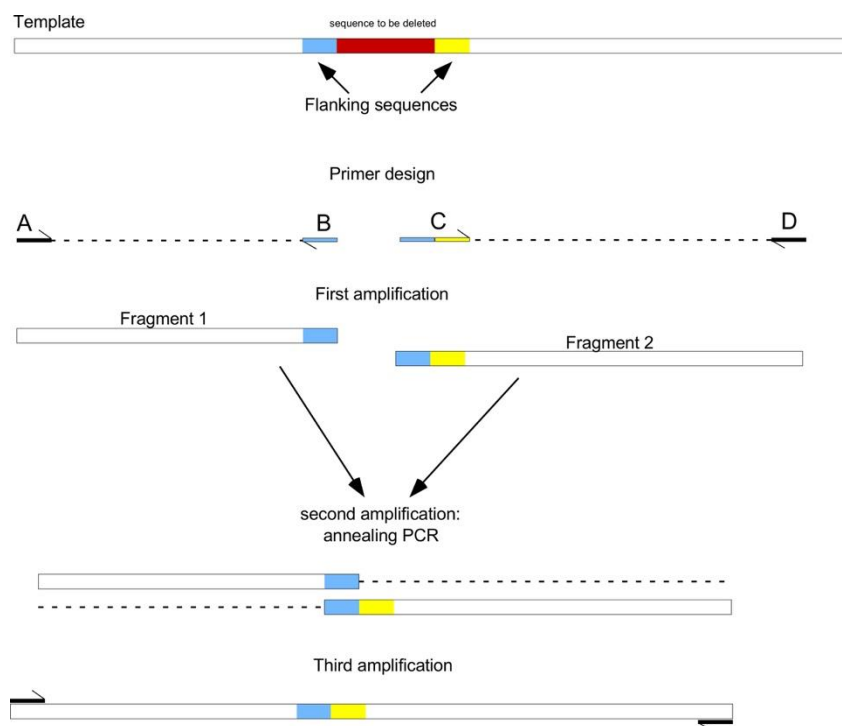
Figure 2.1. Schematic illustration of a fusion PCR. A fusion PCR consists of three PCR reactions. In the first reaction, the two fragments of interest are individually amplified using primer A and B or C and D, respectively. Primer C includes an overhang complementary to the 3' sequence of Fragment 1, allowing the annealing of the two fragments in a second PCR. The hybridized region serves as a priming sequence for the DNA polymerase, which synthesizes the missing strands. In a third PCR the final product is amplified to obtain sufficient quantities.

The same principle has also been applied to generate the MINX-150 construct. Here, the primers B and C were positioned next to one another flanking the zone where the additional sequence was to be inserted. The primer B and primer C contained overhangs with the sequence to be inserted, which in part overlapped.

Table 2.5: Composition of reaction mixtures for the first and third amplifications

Component	Stock conc.	Final conc.
HF Buffer	5x	1x
dNTP mix	10mM	200 µM
Primer 1 (Fw)	10 µM	200 nM
Primer 2 (Rev)	10 µM	200 nM
Template		
(plasmid)/		0.4 ng/µL
(PCR product)		0.04 ng/µL
Pfu Tripple DNA Pol.	100 ng/µL	1 ng/µL

Table 2.6: Program of the first and third amplifications

Step	Temperature	Time	
Initial denaturation	95 °C	3 min	
Denaturation	95 °C	30 sec	
Annealing	56–62 °C	30 sec	29 cycles
Extension	72 °C	1min	
Final extension	72 °C	5 min	

Table 2.7: Composition of reaction mixtures for the second amplification (annealing)

Component	Stock conc.	Final conc.
HF Buffer	5x	1x
dNTP mix	10mM	200 µM
Primer 1 (Fw)	10 µM	200 nM
Primer 2 (Rev)	10 µM	200 nM
Templates		
(Fragment 1) +		0.4 ng/µL
(Fragment 2)		0.4 ng/µL
Pfu Tripple DNA Pol.	100 ng/µL	1 ng/µL

Table 2.8: Program of the second amplification (annealing)

Step	Temperature	Time	
Initial denaturation	95 °C	3 min	
Denaturation	95 °C	30 sec	
Annealing	55 °C	45 sec	10 cycles
Extension	72 °C	1 min	
Final extension	72 °C	5 min	

Materials and Methods

2.2.1.8 Generation of MINX-80-cleaved

The MINX-80-cleaved construct was generated by PCR as two separate fragments.

Amplification of the 5' fragment (MINX-80-cleaved Primer A and B) comprised exon 1 and 25 nucleotides downstream of the 5'SS.

The 3' fragment (MINX-80-cleaved Primer C and D) comprised the remaining intronic sequence including the branch site and polypyrimidine tract as well as the exon 2.

To allow *in vitro* transcription Primer C contained the T7 promoter sequence at its 5' extremity (underlined).

Both fragments were gel purified and used as template in *in vitro* transcription reactions.

2.2.1.9 In vitro transcription

For *in vitro* transcription of radioactively labelled or non-labelled pre-mRNAs, linearized plasmids or PCR products were used as DNA templates (see Section 2.2.1.7 and 2.2.1.8). To improve the transcript's stability and to allow affinity purification, all pre-mRNAs were equipped with a $m^7G(5')ppp(5')G$-cap and a MS2-tag, respectively. Only part B of trans-MINX-80 was transcribed without cap. MINX pre-mRNA constructs were synthesized by the T7 RNA polymerase and PM5 or PM5-10 constructs by the SP6 RNA polymerase.

Table 2.9 illustrates the composition of a standard *in vitro* transcription reaction. For transcriptions of trans-MINX-80 part B, the $m^7G(5')ppp(5')G$-cap was excluded and the final concentration of GTP was raised to 7.5 mM. For transcription of non-radioactive pre-mRNAs α-^{32}P-UTP was omitted and the final concentration of UTP increased to 7.5 mM.

Table 2.9: Composition of a standard *in vitro* transcription reaction

Component	Stock conc.	Radioactive	Without cap	Non-radioactive
TRO buffer	5x	1x	1x	1x
ATP	100 mM	7.5 mM	7.5 mM	7.5 mM
CTP	100 mM	7.5 mM	7.5 mM	7.5 mM
UTP	100 mM	1.5 mM	1.5 mM	7.5 mM
GTP	100 mM	1.5 mM	7.5 mM	1.5 mM
m^7Gcap	114 mM	5 mM	-	5 mM
Rnasin	40 U/µL	1.5 U/µL	1.5 U/µL	1.5 U/µL
YIPP	100 U/µL	1.25 U/µL	1.25 U/µL	1.25 U/µL
T7/SP6 RNA Pol.	20 U/µL	20 U/µL	20 U/µL	20 U/µL
Template		10 ng/µL PCR 50 ng|µL plasmid	10 ng/µL PCR 50 ng|µL plasmid	10 ng/µL PCR 50 ng|µL plasmid
α-^{32}P-UTP	3000 Ci/mmol 10 µCi/µL	750 Ci/mmol	750 Ci/mmol	-

The transcription was performed at 37 °C for 3 h. The DNA template was digested by addition of 0.5 µl RQ1 DNase (1U/µl, NEB) followed by incubation at 37 °C for 30 min. To eliminate unincorporated nucleotides and proteins, the transcription reaction was purified on a ProbeQuant G-50 column (GE Healthcare) according to the manufacturer's protocol and subsequently subjected to PCI extraction (Section 2.2.1.5).

2.2.1.10 Denaturing polyacrylamide gel electrophoresis

Pre-mRNA splicing reactions or snRNA samples were resolved by denaturing polyacrylamide gel electrophoresis (PAGE) in the presence of 8 M urea. Denaturing gels were composed of 1x TBE, 8 M urea and 6–14% (v/v) polyacrylamide (40%, 19:1), depending on the size of the RNAs to be analysed. Polymerization was initiated by addition of 0.06% (w/w) APS and 0.06% TEMED. RNA samples were dissolved in RNA loading dye, denatured for 10 min at 80 °C and subsequently chilled on ice. The electrophoresis was performed in 1x TBE buffer at 20–35 mA. The RNA was visualized by autoradiography, with the Phosphorimager Typhoon 8600 or with SYBR® Gold Nucleic Acid Gel Stain.

2.2.1.11 SYBR® Gold Nucleic Acid Gel Staining

To stain snRNA resolved by denaturing PAGE (Section 2.2.1.10), gels were fixed in RNA gel-fixing solution for 30 min at room temperature. Next, the gels were soaked in water for 3 times 10 min each and equilibrated in 0.5x TBE buffer for 2 times 10 min. Staining was performed with 1x SYBR® Gold Nucleic Acid Gel Stain in 0.5x TBE for 15 min. Since the stain is light-sensitive, the gels were protected from light. After a short rinse with 0.5x TBE buffer, RNAs were visualized using the Fluorescence function of the Phosphorimager Typhoon 8600.

2.2.1.12 Agarose gel electrophoresis of nucleic acids

Analysis or purification of DNA samples, such as PCR products or digested plasmids, was performed by agarose-gel electrophoresis: Agarose gels contained 1x TBE and 0.8–2% agarose depending on the size of the DNA fragments. For DNA visualization gels were stained with ethidium bromide, an organic intercalating agent, at a concentration of 0.5 µg/ml. Samples were mixed with Purple Gel Loading Dye (6x, NEB) and the DNA content was separated by electrophoresis at 60–120 V in 1x TBE buffer. In order to evaluate the size of the DNA fragments a commercial DNA ladder (Roth, NEB) was loaded next to the samples. To recover specific DNA fragments from the gel, the DNA was visualized with UV light at a wavelength of 365 nm and the band of interest was excised with a sterile razor blade.

Materials and Methods

The DNA was extracted from the gel by using the NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) as described in the user's manual. This includes melting of the gel piece, binding of the DNA to the provided matrix and the subsequent elution of the purified fragment. The concentration of the eluted DNA was determined by Nanodrop (as described in Section 2.2.1.4).

2.2.2 Protein-biochemistry standard methods

2.2.2.1 Protein quantification

To measure the concentration of protein samples, the BCA™ protein assay kit (Thermo Fisher Scientific) was used according to the manufacturer's protocol. This assay is based on bicinchoninic acid (BCA) for the colorimetric detection. A BSA standard was prepared for calibration and the quantification of total protein was analysed at an absorption maximum of 562 nm. Since the assay only tolerates an imidazole concentration up to 50 mM, samples with higher concentrations (proteins purified from insect cells) could not be quantified by this method. These were estimated by quantification of gel bands, which were related to protein bands of similar size with known concentration.

2.2.2.2 Denaturing SDS polyacrylamide gel-electrophoresis (SDS-PAGE)

Proteins were separated by denaturing SDS-PAGE according to Laemmli (Laemmli, 1970) Protein samples were supplemented with 6x SDS loading dye and denatured for 5 min at 95 °C before loading. The separating gel included 10% polyacrylamide (30%, 37.5:1) and the stacking gel comprised 4% polyacrylamide. The gel was run in 1x SDS running buffer at 60–120 V. Proteins were visualized by Coomassie staining or transferred to a membrane for immunoblotting.

Table 2.10: Compositions of SDS gels

Component	Stock conc.	Stacking gel	Separating gel
Acrylamide	30%	4%	10%
Tris-HCl, pH 6.8	1M	125 mM	-
Tris-HCl, pH 8.8	1M	-	375 mM
SDS	10% (w/v)	0.1%	0.1%
APS	20% (w/w)	0.1%	0.14%
TEMED	100%	0.1%	0.07%

In some cases, for instance for mass spectrometry, proteins were separated on 4–12% NuPAGE™ gradient gels (Thermo Fisher Scientific) and run in 1x MOPS buffer at 200 V.

2.2.2.3 Coomassie staining

Coomassie staining of proteins was performed according to Sambrook *et al*. (Sambrook *et al.*, 1989). The protein gel was fixed with gel-fixing solution for 30 min, soaked in water for 3x 10 min and then stained with hot Coomassie staining solution for 30 min. The gel was then de-stained in water until clear bands were visible.

2.2.2.4 Western blot

For western blot analysis, the denaturing SDS gel was transferred to a nitrocellulose membrane (Protan BA83 nitrocellulose, 0.2 µm) using the wet-tank transfer procedure. For the assembly of the transfer sandwich 6 Whatman papers and the membrane were saturated with 1x western blot transfer + 20% methanol and assembled as follows: 3 Whatman papers, gel, membrane and again 3 Whatman papers. The electro transfer was conducted at 60 V (400 mA max.) with the Bio-Rad Mini Trans-Blot Electrophoretic Transfer Cell system for 2 h at 4 °C. The membrane was subsequently blocked for 1 h with 5% milk in 1x TBS-T, which also served as antibody solution. Immunoblotting was performed for at least 1 h (or overnight) at room temperature at 4 °C by using the primary antibodies listed in Section 2.1.9. The blot was washed 3 times for 10 min each with 1x TBS-T and subsequently incubated with the secondary antibody for 1 h at room temperature. As secondary antibody horseradish-peroxidase-conjugated anti-rabbit (1:50,000) or anti-mouse (1:10,000) antibodies were used. After 3 washes with 1x TBST the proteins were detected by enhanced chemiluminescence using the ECL detection kit (Amersham) as described in the user's manual.

In order to re-blot a membrane with other antibodies, previously blotted antibodies were stripped by incubation in western-blot stripping solution for 20–30 min at 50 °C. The membrane was then rinsed twice with water and washed 3 times with 1x TBS-T for 10 min at room temperature. Afterwards, the membrane was blocked again and subsequently incubated with new antibodies.

2.2.2.5 Northern blot

Northern probe preparation

The RNA was detected with [α^{32}P]-dATP (3000 Ci/mmol) radioactively labelled probes. For probe generation the Prime It II Random Primer Labelling Kit (Stratagene) was used according to the manufacturers' instructions. PCR synthesized DNA-oligonucleotides hybridizing with one of the five snRNAs served as templates.

Materials and Methods

Northern blotting

For northern blot analysis, samples were separated by denaturing polyacrylamide-gel electrophoresis and transferred to a nylon membrane by wet transfer (Trans-Blot Electrophoretic Transfer Cell). The transfer sandwich was assembled from 2 sheets of Whatman paper, gel, membrane and again 2 sheets of Whatman paper. Whatman papers and membrane had previously been soaked in northern Transfer Buffer. The transfer was carried out overnight at 20 V in a cold room. Subsequently, RNAs were crosslinked to the membrane by UV irradiation in a UV-Stratalinker 2400 (Auto Crosslink) with an energy of 1200 J/cm^2. Pre-hybridization was performed for 4 h at 42 °C in 20 mL northern blot hybridization solution supplemented with 0.2 mg/mL denatured fish sperm DNA. The labelled probes were boiled at 95 °C for 5 min and together with fresh denatured fish-sperm DNA added to 20 mL northern blot hybridization solution. Hybridization was performed over night at 42 °C. After collection of the Northern blot hybridization solution the membrane was washed 3 times with 20–25 mL pre-warmed northern blot wash buffer for 30 min. The membrane was subsequently sealed in plastic foil. Signal visualization was performed by autoradiography.

2.2.3 Special methods

2.2.3.1 Preparation of splicing active HeLa nuclear extract

Nuclear extracts were prepared from HeLa cells essentially as described by Dignam *et al*. (Dignamr *et al.*, 1983). The extracts (in Roeder C buffer) were subsequently frozen in liquid nitrogen and stored at –80 °C. For use in splicing reactions, the nuclear extracts were thawed and dialyzed against 50–100 volumes of Roeder D buffer for 2x 2.5 h. To eliminate precipitate, the extracts were centrifuged at 8330 rpm at 4 °C for 10 min in F14-14x50cy rotor (centrifuge Sorvall® Evolution RC Lynx 6000) and aliquots of the supernatants were frozen in liquid nitrogen to be stored at –80 °C.

2.2.3.2 Transient transfection with siRNA and cell fractionation

HeLa SS6 cells were grown in 10 cm dishes containing Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 100 µg/mL penicillin/streptomycin. Cell cultures were kept at at 37 °C in 5% CO$_2$. For cell-culture maintainance, the cells were washed twice with 1x PBS and dissociated from the dish by addition of Accutase® solution. The cells were then resuspended in fresh DMEM.
For transient transfection with siRNAs, DMEM supplemented with 10% FBS but without antibiotics was used. About 3 x 10^5 cells were cultured in 6 wells one day before transfection. For transfection the Lipofectamine® RNAiMAX Transfection Reagent (Thermo Fisher Scientific) was used as described in the user's manual. After a transfection time of 60 hours, cells were washed twice with ice-cold 1x PBS and then resuspended to ~10×10^6 cells/mL into HeLa Cells Lysis Buffer using a sterile cell scraper. After brief vortexing, the samples were incubated on ice for 15 min. The nucleoplasmic fraction was separated from the chromatin faction by centrifugation at 4 °C and 1100 x g

for 10 min. The soluble nucleoplasmic fraction was collected and the chromatin faction resuspended in a volume of HeLa Cell Lysis Buffer equivalent to the nucleoplasmic fraction. In a 1.5 mL reaction tube, the chromatin was solubilized by sonication in a Bioruptor for 3 min (30 sec on, 30 sec off) at maximum intensity in a water bath at 2° C.

The protein concentration was determined using the BCA Protein Assay Kit (Thermo Fisher Scientific) according to the user's manual. The protein composition was analysed by SDS-PAGE followed by western blotting.

2.2.3.3 RNA seq and data-processing

For RNA seq analysis, HeLa cells were transfected with siRNAs specific for Smu1, RED or MFAP1 as described in Section 2.2.3.2. As control, an siRNA without target in the human transcriptome was transfected. Cells were harvested 50 hours after transfection and their total RNA was extracted. The RNA was further processed by the laboratory of Juan Valcárcel; this comprised reverse transcription of the RNA, Illumina sequencing and data analysis.

The reads obtained by Hiseq-Illumina Paired End sequencing (100bp reads post adapter/barcode removal) were mapped on the human hg19 genome and the UCSC transcriptome by using TopHat2 (v2.0.8). Differential transcript splicing was quantified by using the Cuffdiff program of the Cufflinks suite (v2.1.1) and alternative splicing events, including usage of alternatively spliced exons, alternative 5'SS, alternative 3'SS and retained introns were quantified by using the MISO package (v0.5.2).

2.2.3.4 Insect-cell expression of recombinant proteins

Smu1 and RED were expressed separately or jointly in SF-9 or High Five cells using synthetic genes optimized for expression (GeneArt, Life Technologies). For Baculovirus generation, the Smu1 and RED full-length (FL) or truncation (Δ) constructs were amplified by PCR using a primer comprising the required restriction sites. It was subsequently inserted into the pFL vector. To co-purify the dimers, RED was tagged with an N-terminal poly-His$_{10}$ tag under control of the polyhedrin (polh) promoter (EcoRI/PstI) while Smu1 was cloned with a C-terminal Strep-II tag (from provider) under control of the p10 promoter (XhoI/NsiI). Standard digest and ligation reaction are illustrated in Tables 2.3 and 2.4. Since Strep-II tag purification proved to be inefficient, Smu1-FL was additionally cloned with an N-terminal His$_{10}$ tag (EcoRI/PstI) for single purification. For amplification, obtained vectors were transfected into chemically competent *E. coli* XL-10 Gold cells. The entire ligation was incubated with 100 µL cells on ice for 30 min. Heat shock was performed at 42 °C for 1 min, followed by a 5 min incubation on ice. For recovery 1 mL LB was added to the cells, which were subsequently incubated in a thermomixer at 37 °C and 300 rpm for 1h. Cells were then plated on LB plates containing 100 µg/mL ampicillin.

Materials and Methods

Plasmids were isolated using the High Pure Plasmid Isolation Kit from Roche or the PureLink™ HiPure Plasmid Filter Maxiprep Kit from Thermo Fisher Scientific as described in the manufacturers protocol. The sequence was verified by Sanger sequencing (Microsynth Seqlab).

Errorless vectors were transformed into DH10MultiBacY cells for inclusion into baculoviral DNA. For that purpose, 600 ng of vector was added to 100 µL of electro-competent DH10MultiBacY cells and incubated on ice for 30 min. The cells were transferred to an electroporation cuvette and a single 2000 V pulse was applied. Immediately, 1 mL of LB medium was added to the cells, which were transferred to a Greiner culture tube and incubated in a Multitron incubator shaker (Infors HT) at 37 °C and 180 rpm overnight. The transformation products were plated onto LB-Agar containing 100 µg/mL ampicillin, 10 µg/mL gentamycin, 50 µg/mL kanamycin, 12 µg/mL tetracycline, 1 mM IPTG and 150 µg/mL X-gal. Overnight cultures in LB Medium, containing only antibiotics, were prepared from positive (white) colonies. Viral plasmids were isolated using the High Pure Plasmid Isolation Kit from Roche according to the user's manual until step 4. This means that after removal of the precipitate the supernatant was not transferred to the High Pure Filter Tube, but supplemented with 700 µL isopropanol. The plasmid was precipitated at –20 °C for at least 30 min and the DNA was collected by centrifugation at 13,000 rpm and 4 °C for 15 min in a microfuge. After one wash with 75% ethanol the pellet was dissolved in 20 µL deionized and autoclaved water. 6 well plates, each well containing 3 mL of 0.3×10^6 Sf9 cells/mL in Gibco® Sf-900™ III SFM medium, were prepared and transfected with the isolated plasmid using X-tremeGENE™ 9 DNA Transfection Reagent according to the manufacturers protocol. 60–72 h after transfection at 27 °C (KB incubator), the virus containing supernatant was collected (V_0). Fresh Sf9 cells, maintained in suspension in Gibco® Sf-900™ III SFM medium at 0.7×10^6 cells/mL were transfected to generate the V_1 virus and incubated at 25 °C and 90–100 rpm on a MaxQ 3000 Benchtop Shaker. For protein expression, the optimized High Five™ cell line was used. The cells were maintained in suspension in ESF 921 Insect Cell Culture Medium at 0.7×10^6 cells/mL, 25 °C and 90–100 pm. The cells were harvested 60–72 h after transfection by centrifugation at 4 °C and 3000 rpm for 10 min using a Multifuge 1S-R and either immediately subjected to protein purification or frozen in liquid nitrogen and stored at –80 °C.

2.2.3.5 Purification of recombinant proteins from insect cells

The High Five™ cell pellet was resuspended in 10 mL/g cells of pre-chilled High Five Lysis Buffer and disrupted by ultrasonic probe sonication at 30% amplitude for 15–30 min (30 msec on, 30 msec off) until the lysate was clear. After centrifugation for 1 h at 4 °C and 10,000 rpm in a F14–14 x 50cy rotor (Sorvall LYNX 6000 Superspeed Centrifuge) the soluble cell fraction was diluted 1:1 with High Five Dilution Buffer and incubated with previously equilibrated Roti®garose-His/Ni Beads (Roth) at 4 °C for 3 h with head-over-tail rotation. The beads were collected and washed three times with High Five Wash Buffer for 10 minutes at 4 °C with head-over-tail rotation. Elution was performed with 1.5–3 bead volumes High Five Elution Buffer on ice for 15 minutes. Eluates were frozen in liquid nitrogen and stored at –80 °C. Analysis of the purified proteins was performed by SDS polyacrylamide gel-electrophoresis, followed by staining with Coomassie blue.

2.2.3.6 Crosslinking of antibodies to Dynabeads™ Protein A

In order to obtain antibodies covalently associated with beads, antibodies were crosslinked to the beads by using dimethyl pimelimidate (DMP). The crosslinking procedure is based on the protocol provided by Abcam.

600 µL beads were washed 3 times with 1x PBS head-over-tail at 4 °C for at least 15 min. 50 mg of anti-Smu1 antibody in 800 µL PBS was added to the beads and incubated overnight with head-over-tail rotation at 4 °C. As a control (mock) another set of beads was incubated with buffer without any antibody. The following steps were performed with head-over-tail rotation at room temperature. The beads were washed 3 times with 1x PBS for 10 min and subsequently equilibrated with triethanolamine solution for 2 times 10 min. The DMP was dissolved in a triethanolamine solution to a concentration of 25 mM. Crosslinking was performed with 1 mL 25 mM DMP for 1 h and another 30 min with fresh DMP. The crosslinker was quenched by incubation with 2x 50 mM Tris, pH 7.5 in 1x PBS for 10 min.

After two washes with PBS, the excess of antibody was removed by incubation in 200 mM glycine, pH 2.5 for 2 min. The beads were subsequently washed three times with 1x PBS-T for 10 min. The beads were stored in 1x PBS-T containing additionally 0.09% sodium azide.

2.2.3.7 Immunodepletion of HeLa nuclear extract

Immunodepletion of Smu1/RED from HeLa nuclear extract (Section 2.2.3.1) was performed using anti-peptide antibodies specific for Smu1. The crosslinked beads were washed 3 times with 1x PBS and then blocked overnight in 1x PBS supplemented with 0.5 mg/mL BSA, 0.05 mg/mL $tRNA_{E.c.}$ and 0.05 mg/mL Glycogen.

To minimize unspecific binding to the beads, the salt concentration of the nuclear extract was increased to 750 mM KCl (SE-750). After three washes with Roeder D buffer containing an increased salt concentration of 750 mM KCl (RD-750 buffer), the beads were divided into two equal portions and the SE-750 was incubated successively with each portion for 2 hr at 4 °C with head-over-tail rotation. The extract was subsequently dialyzed against RD-100 buffer for 5 h at 4 °C. Mock-depleted extract was treated in a similar manner while the antibody was omitted.

2.2.3.8 In vitro splicing reaction

A typical splicing reaction was carried out in the presence of 10 nM MINX, PM5 or PM5-10 pre-mRNA and 40 % (v/v) HeLa nuclear extract or 50% depleted extract. The composition of a standard splicing reaction is illustrated in Table 2.11.

The splicing reaction mixtures were incubated at 30 °C for the times indicated in the respective figure (Section 3, Results).

Materials and Methods

Table 2.11: Composition of a standard *in vitro* splicing reaction

Component	Stock conc.	Final conc.
HEPES-KOH, pH 7.9	1 M	20 mM
MgCl$_2$	500 mM	3 mM
KCl	3000 mM	60 mM
ATP	100 mM	2 mM
Creatine phosphate	500 mM	20 mM
Nuclear extract		40–50%
(^{32}P-) pre-mRNA		5–10 nM

For *in vitro* trans-splicing, 10 nM of the 5' fragment and 50 nM of the 3' fragment were assembled in a standard splicing reaction.

To assemble spliceosomes in the presence of recombinant Smu1 and/or RED proteins, the nuclear extract was first supplemented with a 30-fold excess of recombinant protein and pre-incubated for 15 min at 30 °C before addition to the splicing reaction.

For chase experiments, the splicing reaction was incubated for the time indicated and the recombinant protein was then added along with a 10-fold excess of unlabelled pre-mRNA. To exclude the risk that an effect might have been due to the increased salt concentration or the presence of imidazole, a control was performed using buffer without protein.

2.2.3.9 Analysis of in vitro splicing by Denaturing PAGE

In order to analyse the products of *in vitro* pre-mRNA splicing, the splicing reaction was subjected to PK digestion followed by PCI extraction, and the RNA isolated was resolved on a denaturing polyacrylamide gel. To visualize pre-mRNA, splicing intermediates and products, the gel was dried at 70 °C for 2 h and visualised with a phosphorimager screen.

2.2.3.10 Analysis of splicing complexes by native agarose gel-electrophoresis

The assembly of the spliceosome *in vitro* can be analysed by using a mobility-retardation assay, since the A, B B^{act} and C complexes exhibit a decreased mobility on native agarose gels (Lamond *et al.*, 1987; Kent & MacMillan, 2002). To that end, splicing reaction mixtures were supplemented with 0.5 µg/µL heparin and 5x native gel-loading dye before loading the sample onto a native gel (2% w/v low-melting-point agarose in 1x TBE buffer). Electrophoresis was performed at 50–60 mA for 18 h at room temperature. The gels were subsequently dried at 60 °C for 3–4 h and visualised with a phosphorimager screen.

2.2.3.11 MS2 affinity-selection of splicing complexes

The isolation of spliceosomal complexes was performed by MS2 affinity-selection as described previously (Bessonov *et al.*, 2010). To allow purification, the pre-mRNA was pre-incubated with a 20-fold molar excess of purified MS2-MBP fusion protein in the presence of 20 mM HEPES-KOH, pH 7.9 for 30 min at 4 °C.

Subsequently, the pre-mRNA was added to a 1 mL standard splicing reaction. Kinetic and ΔSmu1/RED B complexes were incubated for 8 min at 30 °C. Spliceosomes were assembled on MINX-70 for 15 minutes.

The splicing reaction was loaded onto a 14 ml linear 10–30 % (v/v) glycerol gradient containing G-150 buffer. Ultracentrifugation of the gradients was performed at 22,000 rpm for 15 h 20 min at 4 °C in a Sorvall TST 41.14 rotor, and the 500 µl gradient fractions were harvested manually from top to bottom. The distribution of ^{32}P-labeled pre-mRNA was analysed by Cherenkov counting.

For affinity selection, the peak fractions containing the respective spliceosomal complexes were pooled and loaded onto a pre-equilibrated column containing 200 µl of packed amylose beads (NEB) and incubated for 15 min. The matrix was then washed with 50 column volumes of G-75 buffer. Elution of the spliceosomal complexes was performed by addition of 400 µl of G-75 buffer containing 25 mM maltose. The snRNA composition was analysed by denaturing polyacrylamide-gel electrophoresis after PK digestion and PCI extraction or by submission to mass spectrometry.

2.2.3.12 Mass spectrometry

The protein composition of spliceosomal complexes was analysed by mass spectrometry. To this end, isolated spliceosomal complexes were concentrated by ethanol precipitation (addition of 0.1 volume 3 M NaOAC pH 5 and 5 volumes absolute ethanol) at –80 °C for 48 h. The RNP pellet obtained was dissolved in 1x NuPAGE™ LDS Sample Buffer, supplemented with NuPAGE™ Sample Reducing Agent and denatured at 70 °C for 10 min. The proteins were separated on 4–12% NuPAGE^{TM} gradient gels (Invitrogen) and stained with Coomassie blue.

Each lane was divided into 23 slices and the proteins of each slice were digested in-gel with trypsin and extracted as described by Shevchenko et al. (Shevchenko *et al.*, 1996). The extracted peptides were subsequently analyzed in a Q Exactive™ HF Hybrid Quadrupol-Orbitrap™ Mass Spectrometer (Thermo Fisher Scientific) under standard conditions. Proteins were identified by searching fragment spectra against the UniProt data base. Sample preparation and mass spectrometric analysis was performed by the laboratory of Henning Urlaub.

Materials and Methods

2.2.3.13 Purification of endogenous spliceosomes from the chromatin fraction

Crosslinking and sonication

For the purification of spliceosomal complexes from the chromatin fraction, cells were lysed as described in Section 2.2.3.2. The HeLa Cell Lysis Buffer, however, contained 30 mM HEPES-KOH pH 7.5 instead of Tris-HCl and the chromatin was pelleted by centrifugation at 700 x g at 4 °C for 10 min. The chromatin pellet was resuspended in a volume of HeLa W/O Salt Buffer equivalent to the lysis volume. The chromatin was then passed through a 26G needle until homogenized. For crosslinking dithiobis succinimidyl propionate (DSP) was dissolved in dimethyl sulfoxide (DMSO) and added to a final concentration of 1.5 mM. Crosslinking was performed on ice, at 4 °C for 1 h. An equal volume of HeLa Sonication Buffer was added and 250 µL per 1.5 mL reaction tube was sonicated in a Bioruptor at maximum intensity (30 sec on, 30 sec off) in a 0.5 °C cold water bath. Sonication was performed for 2 times 60 min with an incubation on ice for 30 min in between.

Glycerol-gradient sedimentation

The cross-linked and sonicated samples were layered on a 10–40% glycerol gradient containing 65 mM NaCl, 30 mM Tris-HCl pH 7.5 and 1 M EDTA. Centrifugation was performed at 4 °C and 60,000 rpm in a Th660 (Thermo scientific) rotor for 1h (Sorvall WX Ultracentrifuge). The gradient was fractionated into 175 µL fractions, which were subsequently analyzed by western blotting. For crosslink reversal 3x loading dye containing 180 mM Tris, 6% SDS, 300 mM DTT and bromophenol blue was mixed with the samples and the mixture was kept at 95 °C for 7 min. From the fractions of interest P-Prp31- and P-SF3b155-associated spliceosomes were purified by immunoprecipitation.

Immunoprecipitation

For immunoprecipitation the peak fractions were pooled and diluted 1:1 with HeLa IP Dilution Buffer.
Unless otherwise stated, incubations were performed at 4 °C under rotation. ~5 µg of the purified antibodies were directly added to the diluted fractions, incubated at 4 °C for 2 h and then with ~30 µL protein A Sepharose beads. After precipitation the beads were sedimented by centrifugation at 3,000 rpm for 3 min in a Biofuge fresco (Heraeus) and the supernatant was collected for further investigation. The beads were washed 4x with HeLa IP Wash Buffer followed by incubation on ice for 2 min. The precipitate's RNA content was analysed by northern blotting.

3. Results

The spliceosome is characterised by its highly complex and dynamic structure and composition. While lower eukaryotes get by with the core components necessary for intron excision, higher eukaryotes require additional factors to cope with the increased challenges of a more complex gene architecture. Smu1 and RED are two non-conserved spliceosomal factors whose function in splicing is currently poorly understood. In order to shed some light on this issue, this study targets the role of Smu1 and RED in splicing with the help of a combination of genome wide, *in vivo* and *in vitro* approaches.

3.1 Alternative and constitutive splicing are affected by knockdown of Smu1 and RED

Several studies targeting Smu1 and RED have revealed that their knock-down/gene mutation affects alternative splicing in plants, worms and humans (Spartz *et al.*, 2004; Chung *et al.*, 2009; Papasaikas *et al.*, 2015). However, as these studies focused on known alternative splicing events, the question of whether constitutive splicing is altered in the absence of Smu1 and RED has not been investigated. To address this issue, I examined how their depletion generally effects splicing in HeLa cells by RNA-seq analysis. To be able to differentiate between Smu1/RED-specific effects on splicing and effects shared with other B-specific proteins, the knock-down of MFAP1 was additionally investigated. For that purpose, HeLa cells were transfected with siRNAs specific for Smu1, RED or MFAP1, respectively. As a control, an siRNA without target in the human transcriptome was transfected. 50 h after transfection the cells were harvested and the total RNA was extracted for Illumina sequencing. The RNAseq analysis was performed by Panagiotis Papasaikas in the laboratory from Juan Juan Valcárcel (CRG, Bercelona).

Hiseq-Illumina Paired End sequencing (100bp reads post adapter/barcode removal) generated 2 x 135,376,977 reads for the control, 2 x 144,695,392 reads in ΔRED cells, 2 x 129,201,923 reads in ΔSmu1 cells and 2 x 131,781,731 reads in ΔMFAP1 cells. About 89%, 88%, 86% and 85% of the total reads for the control, ΔRED, ΔSmu1 and ΔMFAP1 cells, respectively, were mapped on the human hg19 genome and the UCSC transcriptome by using TopHat2 (v2.0.8). Differential transcript splicing was quantified by using the Cuffdiff program of the Cufflinks suite (v2.1.1) and alternative splicing events, including usage of alternatively spliced exons, alternative 5'SS (Alt5'SS), alternative 3'SS (Alt3'SS) and retained introns were quantified by using the MISO package (v0.5.2). As shown in Fig. 3.1A, the alternative splicing of cassette exons and intron retention represented the majority of changes in splicing events upon depletion of Smu1, RED and MFAP1. A substantial majority of cassette exons (Cex) were skipped (Fig. 3.1B). Usage of alternative 5'SS and 3'SS detected in Smu1 or RED depleted cells were largely overlapping (~56%) while the overlap level dropped to 35–40% between Smu1/RED and MFAP1 depleted cells (Fig. 3.1C and D).

Results

Similarly, a high level of similarity (72% overlap) between Smu1- and RED-depleted cells was observed for the alternative splicing of cassette exons. This observation is consistent with the idea that the two factors function as a dimer. In comparison, the extent of overlap observed for cells depleted of Smu1 or RED with ΔMFAP1 cells was reduced to 51% (Fig. 3.1E). Nevertheless, a similarity of 51% is not a small matter and is indicative of factors that associate concomitantly with the spliceosome.
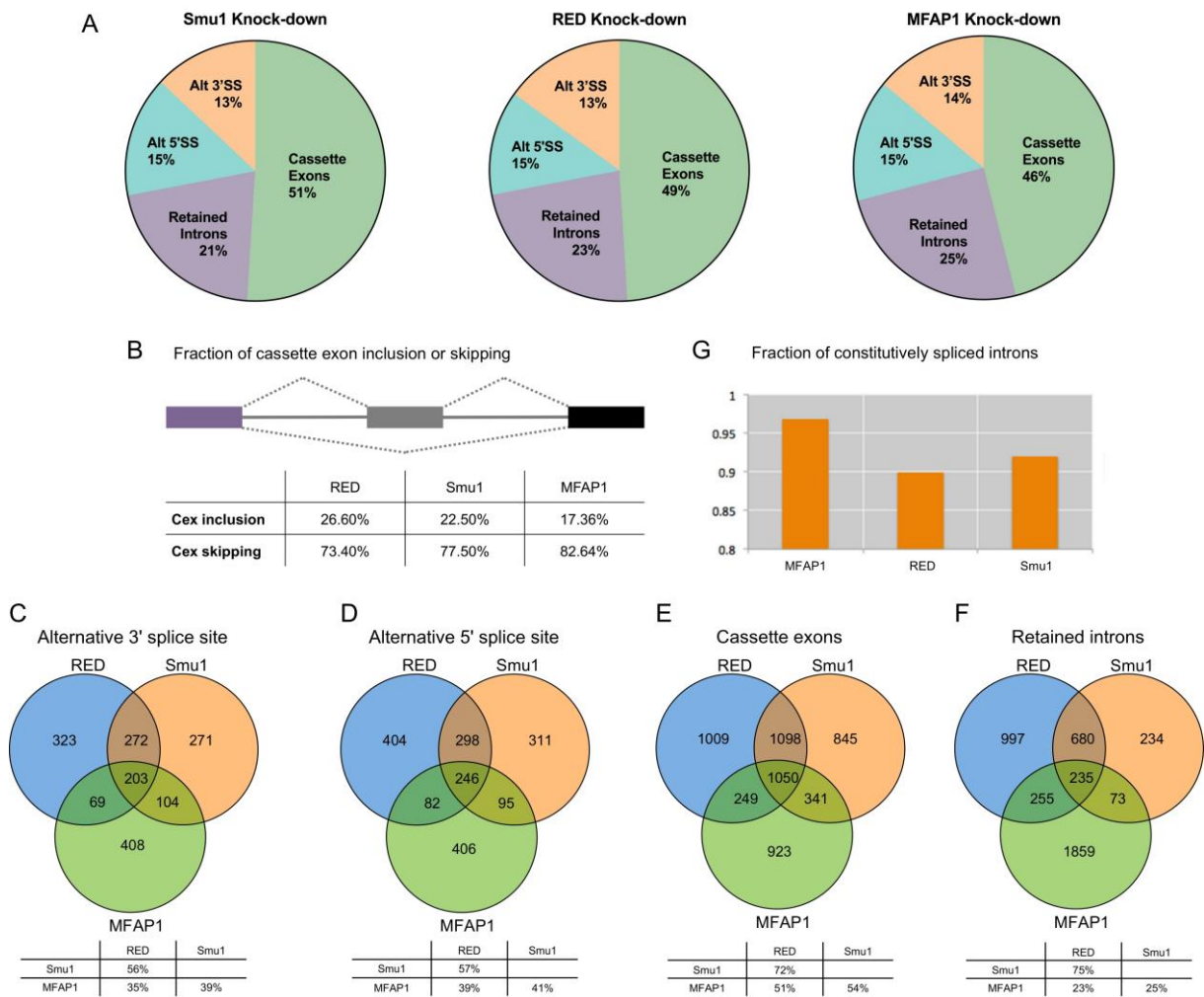
Figure 3.1. Splicing events effected by siRNA-mediated depletion of Smu1, RED or MFAP1. (A) Pie charts showing class distribution of all alternative splicing events detected: Usage of alternatively spliced cassette exons, alternative 5'SS (Alt5'SS), alternative 3'SS (Alt3'SS) and retained introns in siRNA-mediated knock-down cells for Smu1, RED and MFAP1. (B) Relative fraction of cassette exon (Cex) inclusion or skipping among the population of cassette exons whose alternative splicing is affected in ΔSmu1, ΔRED and ΔMFAP1 HeLa cells. (C-F) Venn diagrams showing the extent of the AS events overlap between each class of AS, i.e. alternative 3' splice sites (C), alternative 5' splice sites (D), cassette exon (E) and retained introns (F) in the three different cell types i.e. ΔSmu1, ΔRED and ΔMFAP1 HeLa cells. Numbers represented in the circles represent the absolute number of the corresponding AS events detected in each depletion experiment. The table shows the relative overlap in %. (G) Chart representing the fraction of constitutively spliced introns among the introns retained (unspliced) in ΔSmu1, ΔRED and ΔMFAP1 cells.

As observed for cassette exons, introns retained upon Smu1 and RED depletion also showed a 75% overlap. Strikingly, the overlap of ΔSmu1/RED cells with MFAP1-depleted cells was (at 25%) dramatically reduced, indicating that the majority of introns retained upon MFAP1 depletion were normally processed in Smu1/RED-depleted cells and *vice versa* (Fig. 3.1F). Introns retained in ΔRED, ΔSmu1 and ΔMFAP1 cells appeared to represent in their vast majority (>90%) introns that are normally constitutively spliced (Fig. 3.1G) by the spliceosome. This suggests that Smu1 and RED are important for splicing in general, and are not only involved in the regulation of alternative splicing.

3.2 Short introns are predominantly retained upon knock-down of Smu1 and RED
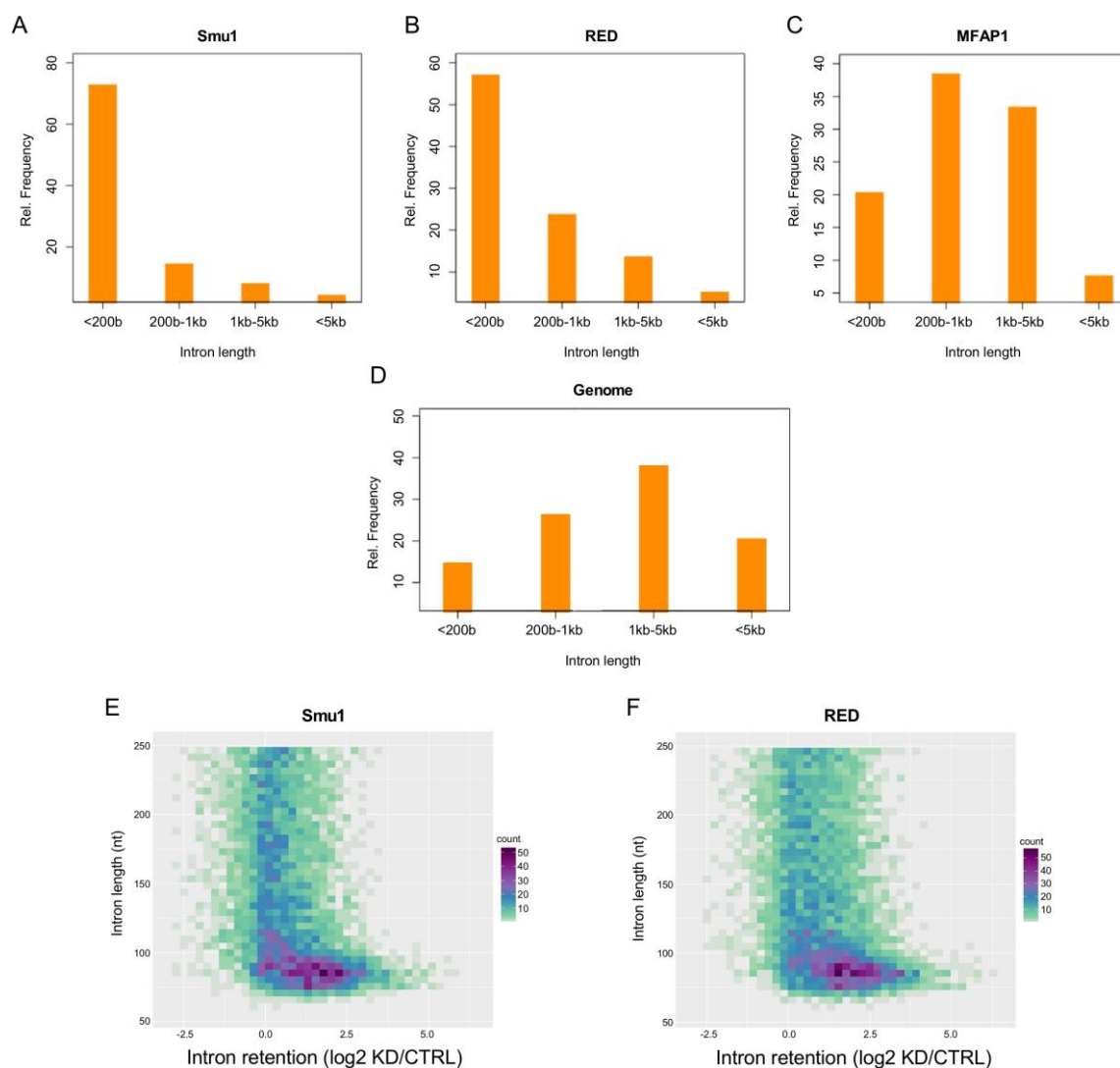
Figure 3.2. siRNA-mediated depletion of Smu1 or RED results in retention of short constitutively spliced introns. (A-C) Introns were sorted into 4 classes according to their length i.e. <200 bases, 200 to 1Kb, 1Kb to 5Kb and >5Kb. Chart showing the relative distribution among those 4 length classes of the introns retained in ΔSmu1(A), ΔRED (B) and ΔMFAP1 (C) cells. (D) Chart showing the relative distribution of intron length (whole genome) among these 4 classes. (E & F) Heat maps showing the level of intron retention (x-axis) relative to the intron length (y-axis) upon depletion of Smu1 (E) or RED (F). The colour code ranges from pale green for low abundance introns to deep red for highly abundant introns (referred to as "count").

Results

All attempts to identify common features of the introns retained after Smu1 and RED knock-down, failed to reveal any mutual motives or sequences. However, surprisingly, I found that introns shorter than 200 nt were preferentially retained. As shown in Fig. 3.2A and B, ~60% of the introns retained in ΔSmu1 and ΔRED cells exhibited a length under 200 nucleotides, even though such short introns only have a prevalence in the whole human genome of about 15% (Fig. 3.2D). In comparison, MFAP1-depletion did not show any preference for short introns, as they represented only 20% of the introns retained (Fig. 3.2 C). Thus, defects in constitutive splicing showed a strong bias toward short introns in Smu1- and RED-depleted cells. Closer examination indicated that introns presenting the highest level of retention in ΔSmu1 and ΔRED cells had a strikingly homogenous length of around 75–100 nt (Fig. 3.2E and F). This distribution supports the idea that intron length may be a co-determinant of the Smu1/RED requirement for efficient intron splicing.

3.3 A Smu1-specific antibody effectively co-depletes Smu1 and RED from HeLa nuclear extract

The RNAseq analysis of Smu1- or RED-depleted cells suggested that these two proteins are required for splicing of introns shorter than 100 nt. To recapitulate the effect seen *in vivo* and to dissect the roles of Smu1 and RED in pre-mRNA splicing, I set up a suitable *in vitro* system. As a first measure, I removed Smu1 from splicing-competent nuclear extract by immunodepletion with a Smu1-specific antibody. To exclude the possibility that an effect is due to the procedure (e.g. to the increase in salt concentration during the depletion, or to the exposure to beads) mock-depleted extract was prepared in exactly the same manner, but with omission of the antibody. This extract served as a control extract in the later splicing experiments. To further assess unspecific binding to the beads, SE750 (splicing extract with 750 mM KCl) was also investigated; in this control the extract had a higher salt concentration but was not exposed to beads.

Analysis of the depleted extract by immunoblotting revealed that Smu1 was nearly completely depleted from the extract, along with around 90% of RED (Fig. 3.3A and B). Apparently, the physical interaction between these two proteins resulted in an efficient co-depletion of RED with Smu1, even at 1 M KCl, underlining the stability of the Smu1/RED dimer. In contrast, other B-specific proteins (Prp38, MFAP1 and FBP21) and also core spliceosomal factors (U5-Snu114K, tri-snRNP protein Snu66 and U2-SF3a66) were not affected by the depletion or by contact with the beads, indicating that the Smu1-specific antibody indeed selectively depleted the Smu1/RED dimer from the extract.
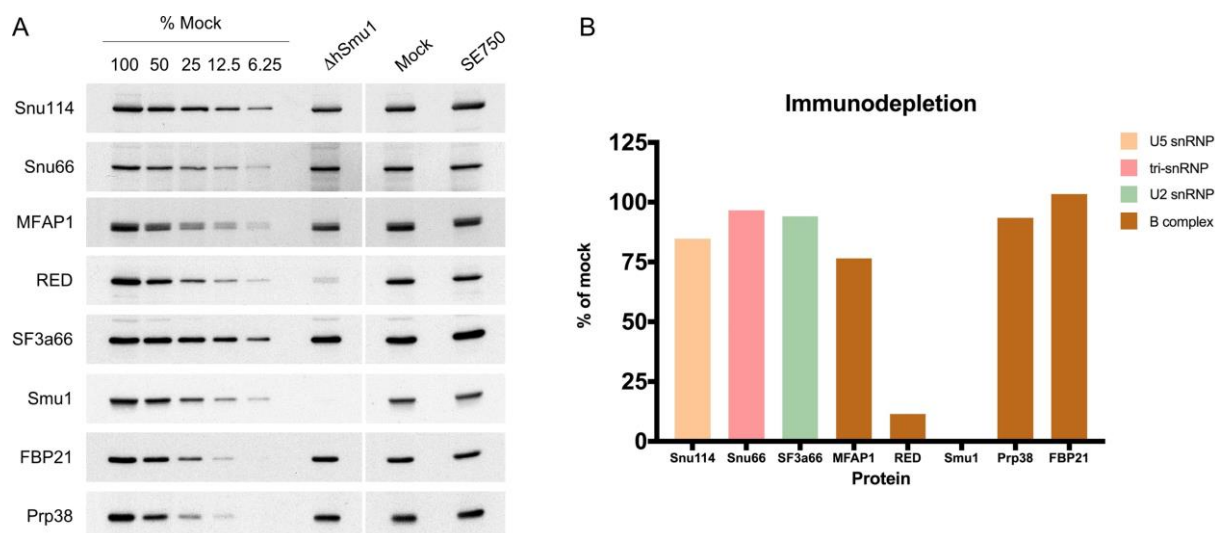
Figure 3.3. Smu1 and RED levels are efficiently and specifically decreased by immunodepletion. (A) Western blot analysis of immunodepleted extract. For immunodepletion, the salt concentration of splicing-competent nuclear extract was increased to 750 mM KCl and incubated with two sets of beads for 2 h each, followed by dialysis against RD100 buffer. A Smu1-specific antibody (ΔSmu1) or no antibody (Mock) was coupled to the beads beforehand. To monitor bead-specific effects, a parallel extract was analysed that was not exposed to beads, but only to an increased salt concentration (SE750). Equal protein concentrations from each extract were loaded, as well as a ladder of the mock-depleted extract, to assess depletion efficiency. The extracts were separated by 4–12% SDS-PAGE, transferred to a nitrocellulose membrane and analysed by immunoblotting using antibodies against Smu1 and RED as well as core splicing factors (116K, 110K and SF3a66) or other B-specific proteins (MFAP1, FBP21 and hPrp38) as controls. (B) Quantification of protein levels in the depleted extract relative to the mock extract.

3.4 Spliceosome activation is slowed down in ΔSmu1/RED extracts

In order to assess the role of Smu1/RED, I investigated the kinetics of *in vitro* splicing of the MINX-MS2 pre-mRNA in the absence or presence of Smu1/RED. This pre-mRNA comprises a 120-nt-long intron, flanked by two exons; it is therefore able to perform both splicing reactions. The co-transcribed MS2 tag of the pre-mRNA binds efficiently to an MS2-MBP fusion protein, which in turn binds to amylose beads. However, the affinity of the MBP (maltose binding protein) for amylose is lower than that for maltose. This makes possible the affinity purification of spliceosomes assembled *in vitro* for compositional analysis by gel electrophoresis or mass spectrometry. Analysis of MINX pre-mRNA splicing by denaturing gel electrophoresis revealed that both splicing steps occurred in the absence of Smu1/RED but less efficiently; this was manifested by the delayed appearance of both intermediates and products upon depletion (Fig. 3.4A).

After 60 minutes of splicing ~85% of the MINX pre-mRNA had been converted into mature mRNA in mock depleted extract, while only ~40% had been spliced in ΔSmu1/RED extract (Fig. 3.4C). Investigation of the spliceosome assembly on a native gel showed an increased B complex signal relative to the mock at all incubation times when Smu1/RED had been depleted (Fig. 3.4B). However, the signal decreased over time and the subsequently formed B^{act} and C complexes could be detected.

Results

This suggests that the accumulation of B complexes was only transient and that the activation was impaired, but not blocked. In contrast no accumulation of A complexes was observed. Taken together, these results suggest that Smu1 and RED are not essential to form functional spliceosomes on the MINX pre-mRNA, but that their activation is delayed in the absence of the two proteins. Thus, Smu1 and RED appear to facilitate/accelerate the transition from the B to the B^{act} complex.
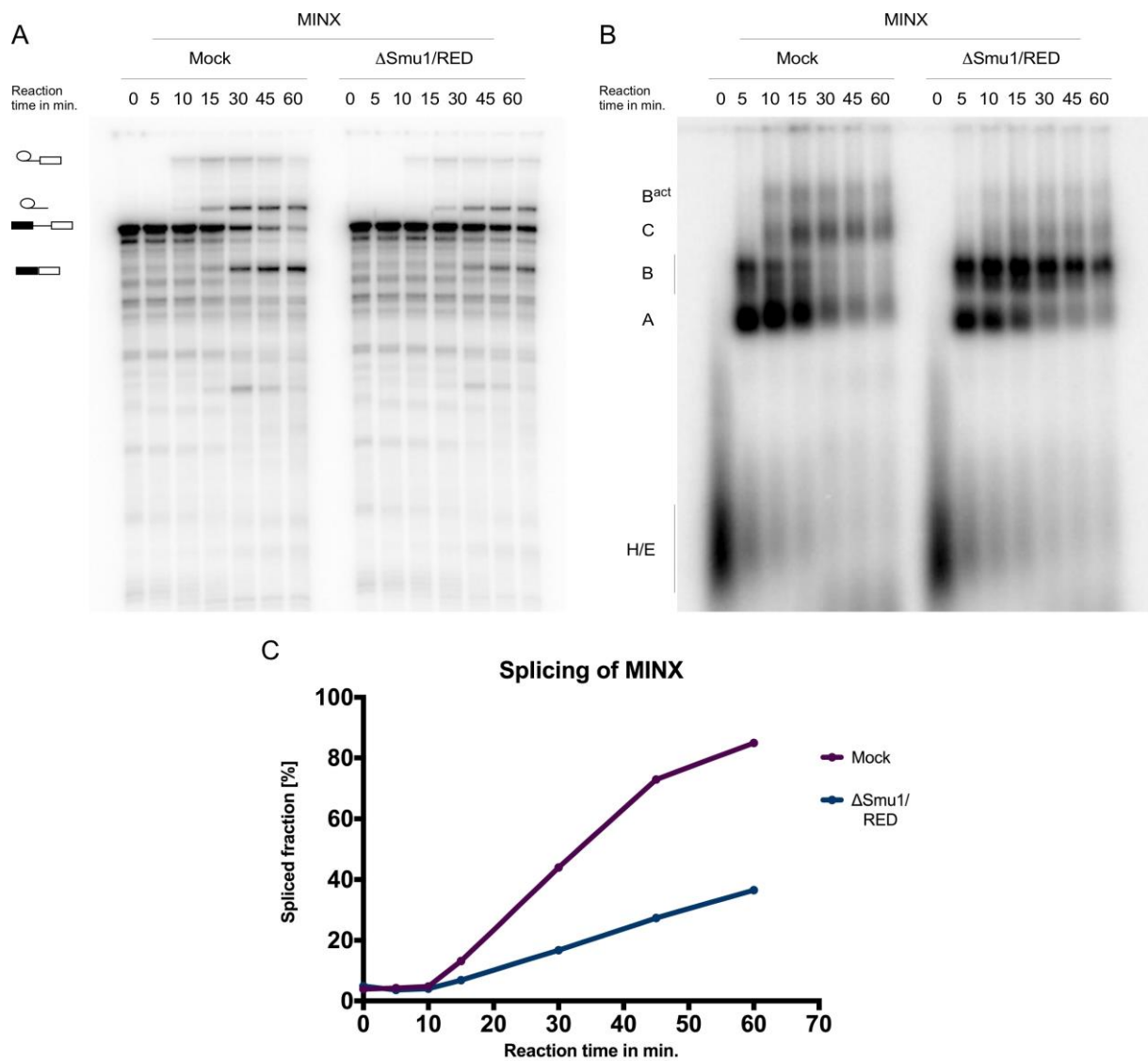
Figure 3.4. The B-to-B^{act} transition is slowed down in ΔSmu1/RED extract. Kinetics of *in vitro* splicing (A) and spliceosome assembly (B) of MINX-120 pre-mRNA in the presence or absence of Smu1/RED. ^{32}P-labelled pre-mRNA was incubated under splicing conditions in the presence of depleted or mock-depleted HeLa nuclear extract for the times indicated. (A) RNA was analysed by denaturing PAGE and visualized by autoradiography. The pre-mRNA, splicing intermediates and producs are indicated on the left. (B) Spliceosomal complex formation was analysed on a native agarose gel and visualized by autoradiography. The positions of the H/E, A, B, C, and B^{act} complexes are indicated on the left. (C) Quantification of the kinetics of pre-mRNA to mRNA conversion over the reaction time.

3.5 Splicing efficiency is intron length dependent

The RNAseq data indicated that the introns retained in cells upon Smu1/RED depletion had a pre-dominant length of 75 - 100 nt, while the splicing of longer introns was much less strongly affected. Therefore, I designed MINX-MS2 constructs containing introns with 70, 80, 90 or 150 nucleotides (MINX-70, -80, -90 and -150) by removing or adding sequences between the 5'SS and the branch site (BS) (Fig. 3.5A). Thereby, the 5'SS - BS distance was extended/reduced from 96 nt to 126, 66, 56 and 46 nt, respectively, while the distance between the BS and the 3'SS remained unchanged. Importantly, the essential splicing motifs - including both splice sites, BS and polypyrimidine tract (PY tract) - remained unaltered.

In untreated nuclear extract, splicing of the MINX-150 pre-mRNA was as efficient as that of MINX-120. Both exhibited similar kinetics and most of both pre-mRNAs (~75%) had been spliced after 45 minutes (Fig. 3.5B and C). This indicates that elongation of the MINX intron by 30 nucleotides did not have a significant effect on splicing or spliceosome formation. When the intron was shortened to 90 nt, splicing-product formation occurred quite efficiently, although with somewhat altered kinetics. At each time point less of the splicing products had been formed, and 60 minutes were required to splice ~75% of the pre-mRNA. Further shortening of the intron to 80 nt revealed an even greater delay in splicing; after 60 minutes only ~55% had been spliced. When the intron was truncated to 70 nt barely any spliced product could be detected. The differences in splicing efficiency suggest that the productivity of the spliceosome depends on the intron length when this length falls below a certain critical value (<120 nt).

Investigation of the spliceosome assembly on native gels showed that spliceosomal complexes were formed with similar efficiency on MINX-120 and MINX-150 (Fig. 3.5D). When spliceosomes were assembled on MINX-90 the B^{act} complex signal at later time points (30–60 min) appeared slightly stronger, while on MINX-80 the B complex was clearly increased after 10 minutes. This is consistent with the moderately reduced splicing efficiencies of the shortened pre-mRNAs which appears to be caused by an impaired spliceosome assembly.

With MINX-70 a strong accumulation of A complexes was obtained at all time points. While moderate signals for the B complex could be observed, no B^{act} or C complexes were visible, indicating the B complex formation is challenging on such a short intron. Taken together, the sequential shortening of the MINX intron resulted in increasingly impaired spliceosome assembly and splicing, suggesting that both depend on the intron length and can be completely inhibited when it falls below a critical size of ~70 nt.

Results
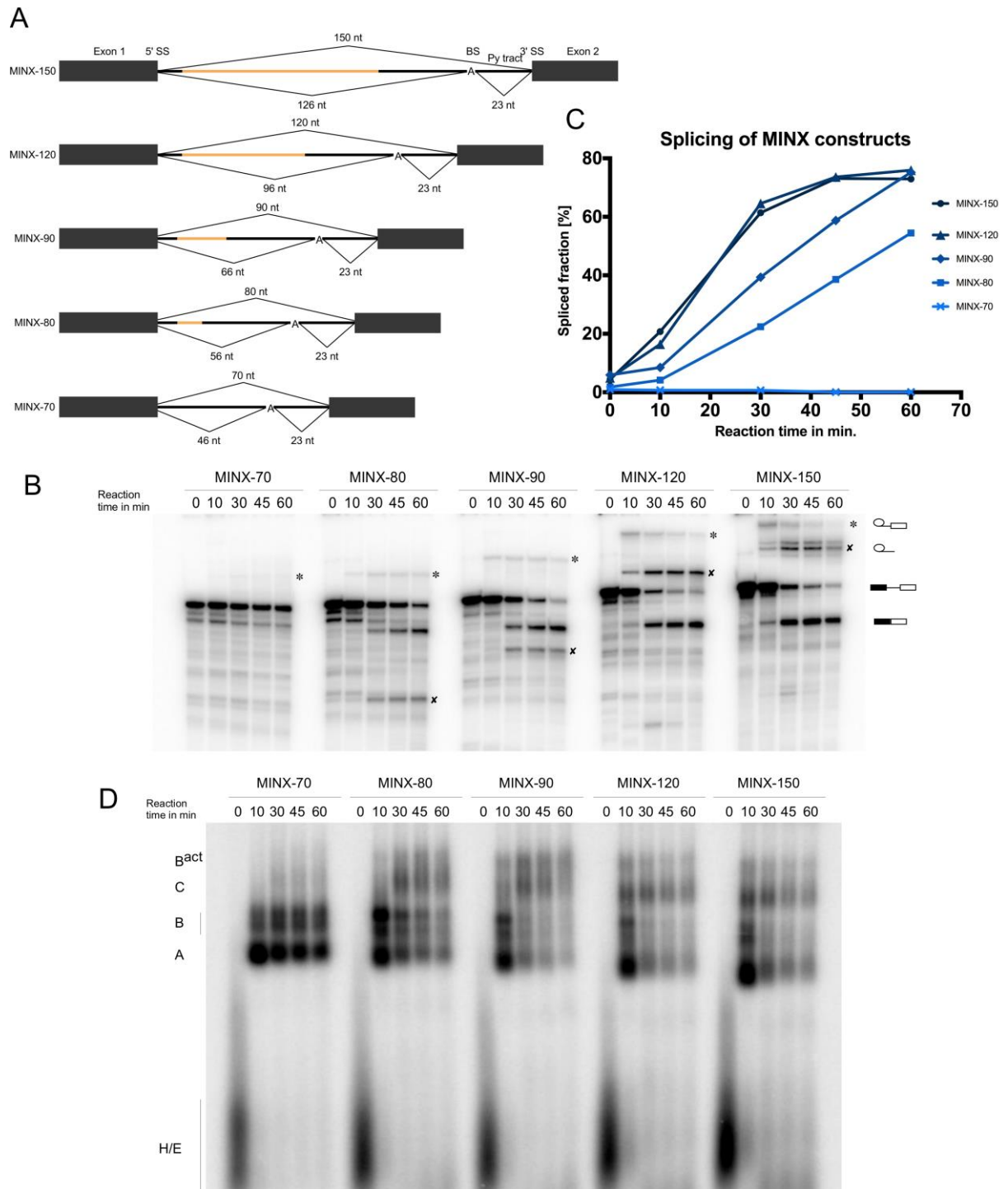
Figure 3.5. Splicing efficiency and spliceosome assembly are length-dependent. (A) Schematic of MINX-MS2 constructs with varying intron lengths. The MINX-120 intron (120 nt) was extended by 30 nt (MINX-150) or truncated by 30 nt (MINX-90), 40 nt (MINX-80) or 50 nt (MINX-70). Reactive sites (5'SS, BP, 3'SS) and the polypyrimidine (PY) tract remained unchanged. Kinetics of *in vitro* splicing (B) and spliceosome assembly (D) of the distinct MINX pre-mRNA constructs. ^{32}P-labelled pre-mRNAs were incubated under splicing conditions in the presence of regular HeLa extract for the indicated times. (B) RNA was analysed by denaturing PAGE and visualized by autoradiography. The pre-mRNA and splicing intermediates and products are indicated on the right. "∗" and "x" mark the intron-3'exon intermediate or spliced-out intron of the distinct pre-mRNA constructs, which migrate differently owing to their different sizes. (C) Quantification of the kinetics of the conversion of pre-mRNA to mRNA over the reaction time. (D) Spliceosomal complex formation was analysed on a native agarose gel and visualized by autoradiography. The positions of the H/E, A, B, C, and B[act] complexes are indicated on the left.

68

3.6 Smu1 and RED are essential for *in vitro* splicing and spliceosome assembly when introns are short
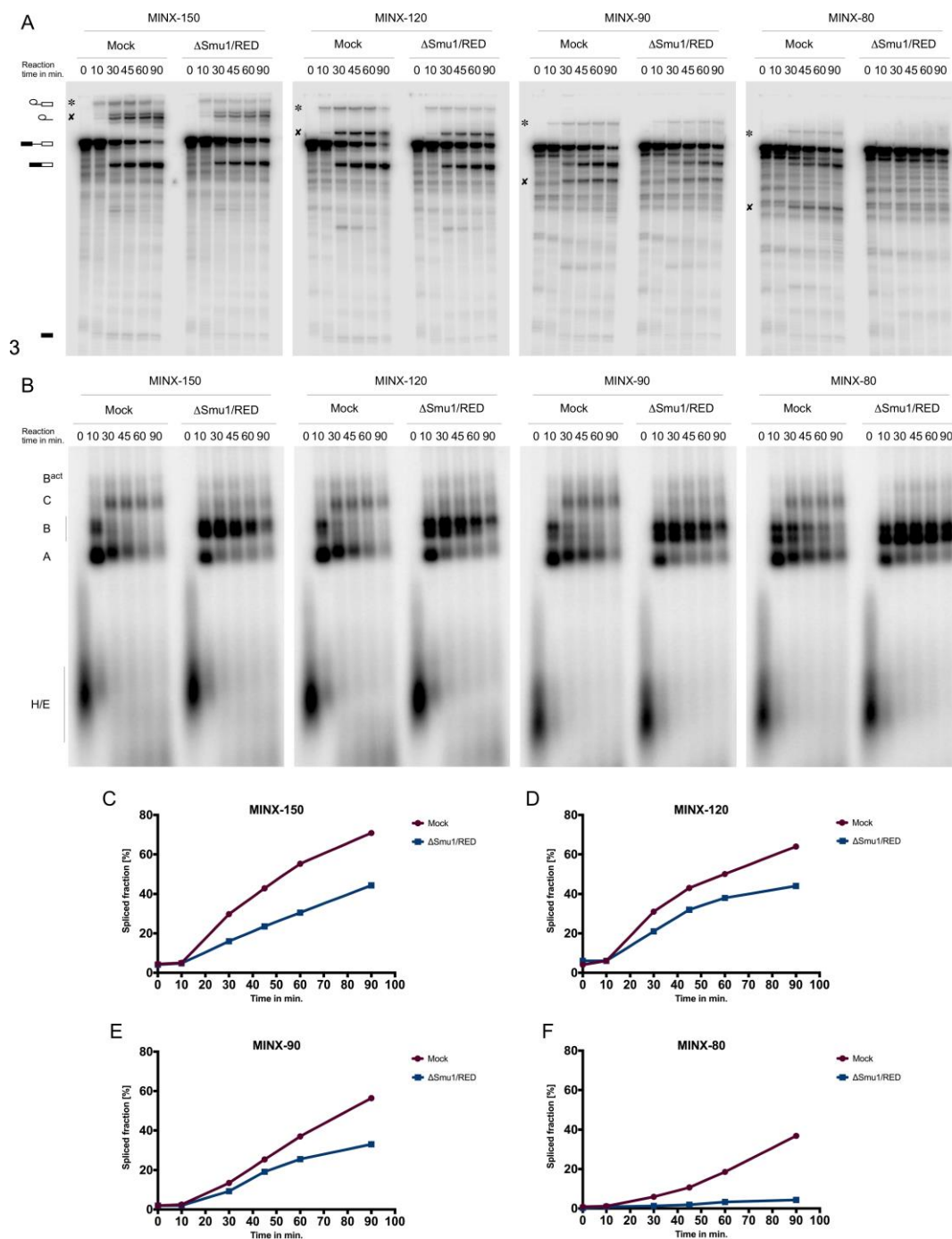
Figure 3.6. Splicing and spliceosome assembly are slowed down/inhibited by the Smu1/RED depletion. Kinetics of splicing (A) and spliceosome assembly (B) *in vitro* of MINX pre-mRNA constructs in the presence or absence of Smu1/RED. ^{32}P-labelled pre-mRNA was incubated under splicing conditions in the presence of depleted or mock-depleted HeLa nuclear extract for the times indicated. (A) RNA was analysed by denaturing PAGE and visualized by autoradiography. The pre-mRNA and splicing intermediates and products are indicated on the right. "*" and "x" mark respectively the intron-3'exon intermediate and the intron of the distinct pre-mRNA constructs, which migrate differently owing to their different sizes. (B) Spliceosomal complex formation was analysed on a native agarose gel and visualized by autoradiography. The positions of the H/E, A, B, C, and B^{act} complexes are indicated on the left. (C-F) Quantification of pre-mRNA to mRNA conversion in ΔSmu1/RED extract relative to the mock-depleted sample.

Results

As a next step, I investigated the effect of the Smu1/RED depletion on the splicing of the four functional MINX constructs (MINX-80, -90, -120 and -150) in the Smu1/RED-depleted extract. To that end, each pre-mRNA was incubated with Smu1/RED-depleted or mock-depleted extract and spliceosome assembly or splicing-product formation was analysed on native or denaturing gels, respectively. In the absence of Smu1/RED, MINX-150 and MINX-120 were spliced ~30% less efficiently (Fig. 3.6A, C and D) than in the mock-depleted extract, and in contrast to the mock control there was a strong accumulation of B complexes at all time points with B complexes still detectable after 90 minutes (Fig. 3.6B). Splicing of MINX-90 in ∆Smu1/RED extract was 50% less efficient than in the mock-depleted extract (Fig. 3.6E) and slightly more B complexes accumulated. Thus, MINX-90 appears to be moderately more affected by the depletion than MINX-150 and MINX-120. Splicing of MINX-80, however, was drastically reduced in the absence of Smu1/RED, with the pre-mRNA-to-mRNA conversion reduced by 90% (Fig. 3.6F). Consistent with this, a significantly increased B complex signal relative to the mock extract could be detected on the native gel, while no signals corresponding to later B^{act} or C spliceosomal complexes were visible, suggesting a severe impairment of the B-to-B^{act} transition. Taken together, these results indicate that Smu1 and RED function to facilitate spliceosome assembly on longer introns and become essential factors for the activation of spliceosomes assembling on short introns with a length of ~80 nt.

3.7 Splicing of MINX-80 can be restored by addition of Smu1/RED

In vitro splicing experiments with Smu1/RED-depleted extract suggested that these two proteins are involved in the activation of spliceosomal B complexes. To exclude the possibility that the observed effect was caused by co-depletion of additional splicing factors, I performed a complementation experiment with human Smu1 and RED proteins, obtained by co-expression in insect cells. For that purpose, RED was equipped with an N-terminal His tag and Smu1 with a Strep-II tag, and for both proteins baculoviral DNA was generated in *E. coli* DH10MultiBacY cells. The viral DNA was transfected into Sf9 insect cells for reconstitution of infectious virus particles. For protein expression, optimised High Five insect cells were transfected with the virus particles, and after 60–72 hours, the dimer could be purified from the cells by nickel-column chromatography (Fig 3.7A). The purified protein was incubated with HeLa nuclear extract before initiating the splicing reaction, to allow the formation of potentially important interactions. Since splicing of MINX-80 was most strongly affected by the Smu1/RED-depletion, it was used as substrate in the following experiments, except where stated otherwise.

As shown in Fig. 3.7B, complementation of the Smu1/RED-depleted reaction with the purified dimer completely resolved the assembly blockade and allowed the formation of B^{act} and C complexes. Furthermore, splicing of MINX-80 was restored to levels comparable to that observed with mock-depleted extract (Fig. 3.7C and D). This indicates that indeed no splicing factors other than Smu1/RED were responsible for stalling the spliceosome assembly.
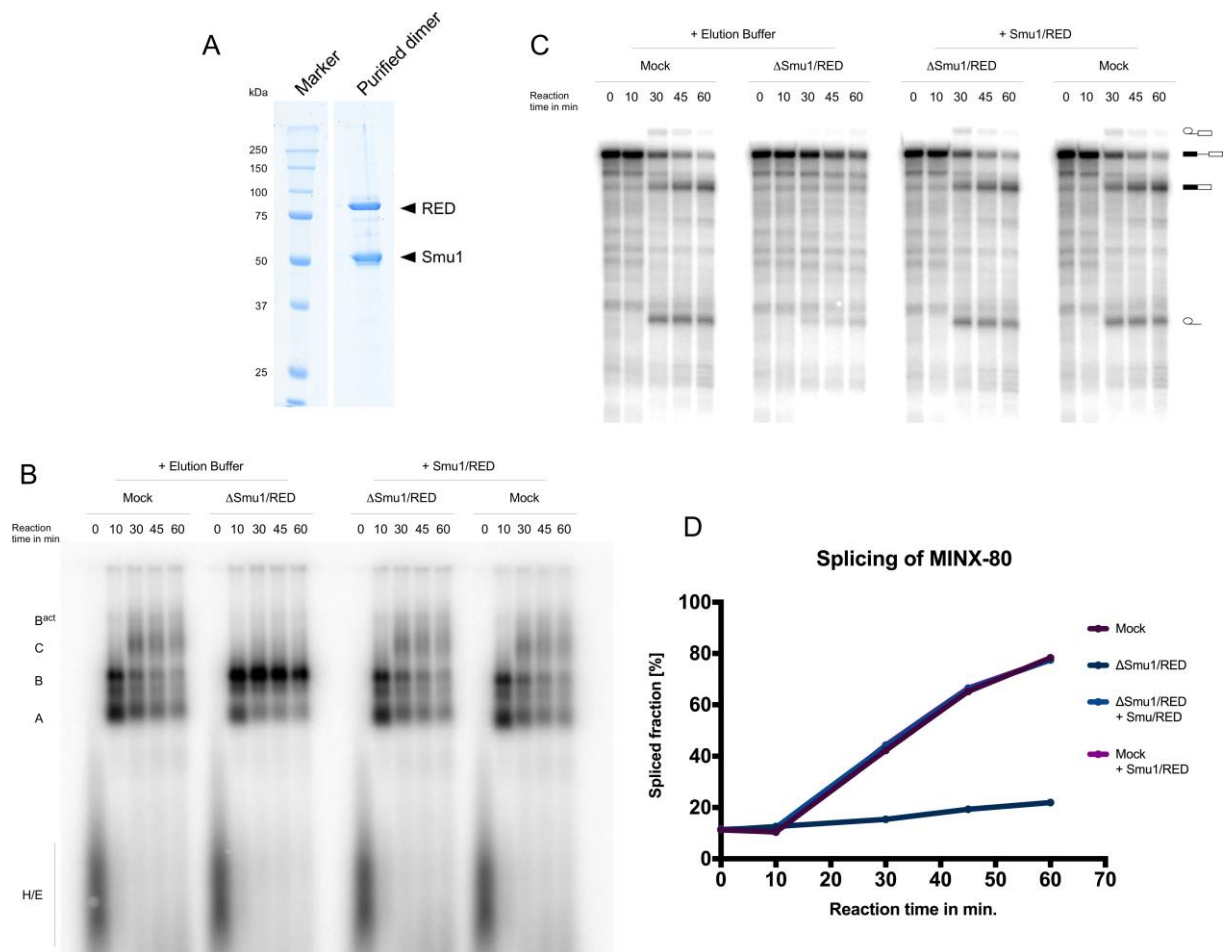
Figure 3.7. Purified recombinant Smu1/RED is sufficient to restore spliceosome activation and splicing. (A) Human Smu1 and RED, expressed and purified from High Five insect cells were separated by SDS-PAGE and stained with Coomassie. Kinetics of spliceosome assembly (B) and *in vitro* splicing (C) after complementation with the Smu1/RED dimer obtained from co-expression in insect cells. The purified proteins, or only High Five Elution buffer, were pre-incubated with depleted or mock-depleted extract and then with the ^{32}P-labelled MINX-80 pre-mRNA under splicing conditions for the times indicated. (B) Spliceosomal complex formation was analysed on a native agarose gel and visualized by autoradiography. The positions of the H/E, A, B, C, and B^{act} complexes are indicated on the left. (C) RNA was analysed by denaturing PAGE and visualized by autoradiography. The pre-mRNA and splicing intermediates and products are indicated on the right. (D) Quantification of the kinetics of pre-mRNA-to-mRNA conversion over the reaction time.

3.8 Smu1/RED immunodepletion does not affect the recruitment of additional spliceosomal factors to the B complex

Complementation of depleted extract with the purified Smu1/RED dimer restored splicing of the MINX-80 pre-mRNA to mock-depletion levels, indicating that the observed defects in splicing and spliceosome assembly were not due to co-depletion of other spliceosomal factors required for the activation. Previously, it has been proposed that Smu1 and RED might serve as a binding platform (Ulrich *et al.*, 2016) for other spliceosomal proteins.

Results

To investigate whether Smu1 and RED recruit spliceosomal factors that are required for a functional activation, or promote the activation themselves, I assembled mock (i.e wildtype) and ΔSmu1/RED-stalled B complexes on MINX-80, separated the particles by glycerol-gradient ultracentrifugation, performed MS2 affinity selection and compared the snRNA and protein composition of the affinity-purified complexes. The mass spectrometry analysis was performed by the laboratory of Henning Urlaub (MPI-bpc, Göttingen).

The gradient profile is illustrated in Fig. 3.8A. It shows that the B complexes assembled in mock-depleted extract after 8 min of *in vitro* splicing, peaked in the fractions 14–15, while ΔSmu1/RED B complexes (after incubation for 8 min) peaked in fraction 12–14, and thus migrated one or two fractions slower. This distinct migration might suggest that the stalled complexes are smaller than the B complexes formed after 8 min in the mock extract, or that they may have adopted a more relaxed conformation, or possibly both. However, the snRNA composition of ΔSmu1/RED B complexes (Fig. 3.8B) was identical to that of the mock B complexes, comprising the characteristic U2, U4, U5 and U6 snRNAs. The absence of the U1 snRNA suggests that the Prp28-mediated destabilization of the U1 snRNP was not affected by the depletion of Smu1/RED, while the presence of the U4 snRNA indicates that the Brr2-mediated unwinding of the U4/U6 helix had not yet occurred. Thus, depletion of Smu1/RED stalled spliceosome assembly after the U1 snRNP release and before the unwinding of the U4/U6 duplex by Brr2.
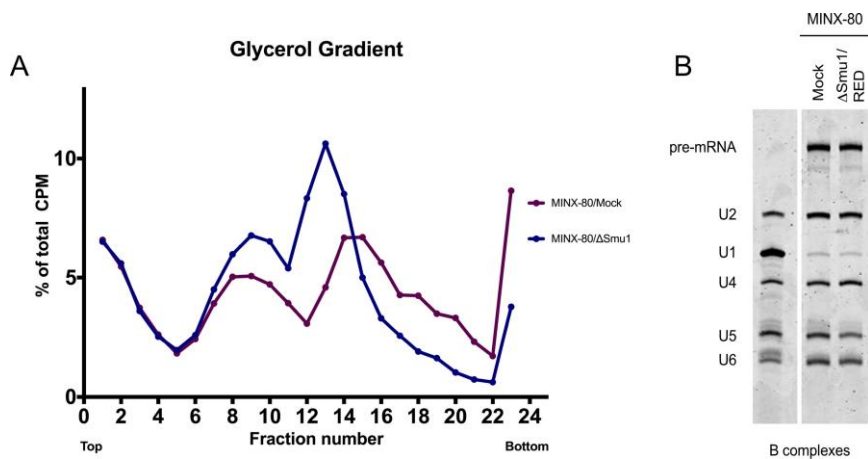
Figure 3.8. Depletion of Smu1/RED stalls spliceosome assembly at the B complex level. (A) The glycerol-gradient sedimentation profile of B complexes assembled on MINX-80 in the presence or absence of Smu1/RED. B complexes were assembled for 8 min under splicing conditions in the presence of depleted or mock-depleted HeLa nuclear extract. The migration of the ^{32}P-pre-mRNA was determined by Cherenkov counting. (B) snRNA composition of the spliceosomal complexes. The spliceosomal complexes in the peak fractions (ΔSmu1/RED B complex, fractions 12–14; mock B complex, fractions 14–16) were subjected to MS2 affinity purification and the RNA extracted from the purified complexes was analysed by denaturing PAGE followed by staining with SYBR Gold.

Table 3.1: Protein compositions of the ΔSmu1/RED and mock (kinetically stalled) B complexes. Proteins of the affinity-purified complexes indicated were identified by LC-MS/MS after separation by 4–12 % SDS-PAGE. 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.

Protein	hGene Name	kDa	MINX-80 Mock	MINX-80 ΔSmu1/RED
Sm proteins				
B	SNRPB	24	45	48
D1	SNRPD1	13	19	20
D2	SNRPD2	14	30	41
D3	SNRPD3	14	46	62
E	SNRPE	11	26	26
F	SNRPF	10	5	8
G	SNRPG	8	7	10
U1 snRNP				
U1-70K	SNRNP70	52	11	22
U1-A	SNRPA	31	11	13
U1-C	SNRPC	17	1	2
17S U2 snRNP				
U2A'	SNRNPA1	28	75	73
U2B"	SNRPB2	25	20	26
SF3a120	SF3A1	89	159	161
SF3a66	SF3A2	51	25	26
SF3a60	SF3A3	59		
SF3b155	SF3B1	146	295	323
SF3b145	SF3B2	100	144	181
SF3b130	SF3B3	136	337	389
SF3b49	SF3B4	44		
SF3b14a/p14	SF3B14	15	18	22
SF3b10	SF3b5	10	5	7
17S U2 related				
hPRP43	DHX15	91	45	36
U2AF65	U2AF2	54	28	48
SPF45	RBM17	45	16	5
SR140 (fSAPa)	U2SURP	118	23	51
CHERP	CHERP	104	12	15
SF3b125	DDX42	103	8	2
U2AF35	U2AF1	28	11	19
SPF31	DNAJC8	30	1	7
hPRP5	DDX46	117	5	25
PUF60	PUF60	60	16	27
SPF30	SMNDC1	27	11	21
U5				
220K	PRPF8	274	658	647
200K	SNRNP200	245	803	754
116K	EFTUD2	109	343	344
40K	WDR57	39	95	80
102K	PRPF6	107	271	254
15K	TXNL4A	17	13	17
100K	DDX23	96	127	116
52K	CD2BP2	38	31	10
LSm proteins				
LSm2	LSM2	11	15	15
LSm3	LSM3	12	3	7
LSm4	LSM4	15	5	9
LSm5	LSm5	10		
LSm6	LSM6	9	3	7
LSm7	LSM7	12	4	8
LSm8	LSM8	10	13	14
U4/U6				
90K	PRPF3	78	111	158
60K	PRPF4	58	108	124
20K	PPIH	19	19	32
61K	PRPF31	55	75	88
15.5K	NHP2L1	14	7	14
U4/U6.U5				
110K	SART1	90	278	350
65K	USP39	65	59	46
27K		27		
RBM42	RBM42	48		

Results

Protein	hGene Name	kDa	MINX-80	
			Mock	ΔSmu1/RED
A complex				
THRAP3	THRAP3	109	23	22
CCAR1	CCAR1	133	2	6
RBM5	RBM5	92	6	7
S164 (fSAP94)	RBM25	100	6	15
RBM10	RBM10	104		
SF1	SF1	68		
FBP11	PRPF40A	109	4	9
SF4	SUGP1	72	2	12
Pre-B complex				
hPRP4-Kinase	PRPF4B	115	72	62
B complex				
hPRP38	PRPF38A	37	45	45
hSnu23	ZMAT2	24	12	16
MFAP1	MFAP1	52	30	35
RED	IK	66	83	4
hSmu-1 (fSAP57)	SMU1	58	87	3
FBP21	WBP4	43	13	29
Npw38BP	WBP11	70	7	0
Npw38	PQBP1	30	2	1
UBL5	UBL5	9	5	4
hPRP19/CDC5L				
hPRP19	PRPF19	55	85	40
CDC5L		92	95	43
SPF27	BCAS2	26	13	2
PRL1	PLRG1	57	46	21
CCAP1	HSPA8	71	156	139
CCAP2	CWC15	27	6	2
catenin, b-like 1	CTNNBL1	62	20	5
CRNKL1 (hSYF3)	CRNKL1	99	76	31
SKIP	SNW1	61	60	22
RBM22	RBM22	47	15	2
PPIase-like 1 (PPIL1)	PPIL1	18	6	2
G10 (fSAP17)	BUD31	17	11	4
IBC				
KIAA0560 (fSAP164)	AQR	171	72	35
hSyf1	XAB2	100	51	27
MGC20398	ZNF830	42	12	0
hIsy1 (fSAP133)	ISY1	33	13	7
Cyp-E	PPIE	33	18	5
B^{act} complex				
KIAA1604 (fSAPb)	CWC22	106	35	1
hPRP17	PRPF17/CDC4	66	27	12
NY-CO-10	CWC27	54	10	2
RNF113A	RNF113A	39		
PPIL2 (Cyp-60)	PPIL2	60	28	5
PRCC	PRCC	52	12	2
MGC23918	CCDC12	21	8	5
SRm160	SRRM1	103	16	3
SRm300	SRRM2	300	71	48
PPIase-like 3b	PPIL3	19		
1st step factors				
hPRP2	DHX16	119	19	7
GPKOW	GPKOW	52	4	1
CCDC49	CWC25	50		
C complex				
GCIP p29 (fSAP29)	SYF2	29	3	4
C9orf78 (HSPC220)	C9orf78	34		
WDR70	WDR70	33		
NOSIP	NOSIP	33		
PPWD1 (KIAA0073, CyP64)	PPWD1	74		
PPIL4	PPIL4	35		
CDK10	CDK10	35		
PPIG (SRcyp)	PPIG	89	2	0
2nd step factors				
hPRP22	DHX8	139	8	2
hPRP18	PRPF18	40		
hSlu7	SLU7	68	0	2
hPRP16	DHX38	140	26	14

Protein	hGene Name	kDa	MINX-80	
			Mock	ΔSmu1/RED
Putative C* proteins				
DDX35	DHX35	79	7	4
Abstrakt	DDX41	70	4	8
C19orf29 (NY-REN-24, cactin)	CACTIN	89		
CXorf56 (FLJ22965)	CXorf56	25,6		
FAM50A (HXC-26, XAP5)	FAM50A	40		
SR proteins				
SF2/ASF	SRSF1	28	37	42
SC35	SRSF2	25		
SRp20	SRSF3	19	10	12
SRp75	SRSF4	57	12	12
SRp40	SRSF5	31		
SRp55	SRSF6	40	16	16
9G8	SRSF7	27	22	22
SRp46	SRSF8	32		
SRp30c	SRSF9	26	8	8
SRp38	SRSF10	31	7	3
SRSF11	SRSF11	54	0	4
hTra-2 alpha	TRA2A	33		
hTra-2 beta (SFRS10)	TRA2B	34		
hnRNP				
hnRNP A0	HNRNPA0	31	10	12
hnRNP A1	HNRNPA1	39	44	67
hnRNP A3	HNRNPA3	40	17	21
hnRNP AB	HNRNPAB	35	11	15
hnRNP A2/B1	HNRNPA2B1	37	53	80
hnRNP C	HNRNPC	33	17	17
hnRNP D	HNRNPD0	24		
hnRNP DL	HNRNPDL	46	8	6
hnRNP F	HNRNPF	46	6	2
hnRNP H1	HNRNPH1	49	9	2
hnRNP H3	HNRNPH3	37	10	6
hnRNP K	HNRNPK	51	25	29
hnRNP L	HNRNPL	64	12	10
hnRNP M	HNRNPM	78	14	8
hnRNPQ	HNRNPQ	70	22	44
hnRNP R	HNRNPR	71	29	43
hnRNP U	HNRNPU	91	27	39
hnRNP UL-1	HNRNPUL1	72	2	2
hnRNP UL-2	HNRNPUL2	91	9	5
PCBP1	PCBP1	38	16	16
PCBP2	PCBP2	32	6	8
RALY	RALY	33		
EJC/mRNP				
eIF4A3	EIF4A3	47	9	0
magoh	MAGOH (B)	17	3	2
Y14	RBM8A	20		
Pinin	PNN	82	6	4
Acinus (fSAP152)	ACIN1	147		
SAP18	SAP18	18	1	3
UAP56	DDX39B	49	15	6
RNPS1	RNPS1	34	4	1
RES complex				
SNIP1	SNIP1	46	20	2
MGC13125 (fSAP71)	BUD13	71	17	7
CGI-79	RBMX2	40		
TREX				
THOC1	THOC1	76	14	6
THOC2	THOC2	183	41	11
THOC3	THOC3	39	3	2
Aly/REF (THOC4)	ALYREF	28	4	5
THOC5 (KIAA0983, fSAP79)	THOC5	79	13	8
THOC6 (WDR58, MGC2655)	THOC6	38	6	6
THOC7 (FLJ23445, fSAP24)	THOC7	24	5	2
mRNA binding proteins				
PABP1	PABPC1	71	21	21
PABP2	PABP2	33		
PABPC4 (IPABP, APP1)	PABPC4	17	16	11
YB-1	YBX1	36	42	35
YBX3	YBX3	40	43	46
NF45	ILF2	43	14	21
ELAV	ELAVL1	39	2	6
ASR2B	SRRT	100	40	49
KIN17	KIN17	45	18	0
RBM15	RBM15	106	11	9

Results

Protein	hGene Name	kDa	MINX-80	
			Mock	ΔSmu1/RED
cap binding complex				
CBP80	NCBP1	**92**	50	51
CBP20	NCBP2	**18**	9	10
NCBP3	NCBP3	**71**		
DExD/H-box RNA Helicases				
p68 (DDX5)	DDX5	**69**	27	42
DHX36	DHX36	**115**	31	23
DHX38	DHX38	**140**	26	14
DDX3X	DDX3X	**73**	17	19
DDX21	DDX21	**87**	11	0
p72 (DDX17)	DDX17	**81**	17	28
DDX9	DHX9	**142**	59	49
Transcription				
TCERG1	TCERG1	**124**	12	2
Miscellaneous				
YWHAZ	YWHAZ	**28**	88	88
JUP	JUP	**66**	76	56
YWHAE	YWHAE	**29**	66	77
ENO2	ENO2	**47**	64	56
DPYSL2	DPYSL2	**62**	61	45
HSPA1A	HSPA1A	**70**	51	46
HSPA5	HSPA5	**72**	51	43
YWHAH	YWHAH	**28**	49	54
ATP1A3	ATP1A3	**112**	47	41
TFIP11	TFIP11	**97**	41	15
NF110b	NF110b	**96**	40	43
LDHB	LDHB	**37**	39	40
PAXBP1	PAXBP1	**105**	37	24
PPIA	PPIA	**18**	35	31
CALM3	CALM3	**17**	34	21
STXBP1	STXBP1	**68**	30	18
GAPDH	GAPDH	**36**	29	22
YWHAB	YWHAB	**28**	28	39
SKIV2L2	SKIV2L2	**118**	28	35
ATP1A1	ATP1A1	**113**	27	0
SFN	SFN	**28**	25	35
TPI1	TPI1	**31**	25	30
MDH1	MDH1	**36**	24	22
PHF5A	PHF5A	**12**	24	22
Nucleolin	NCL	**77**	23	16
BCLAF1	BCLAF1	**106**	22	15
ALDOA	ALDOA	**39**	21	19
SON	SON	**264**	20	24
PGK1	PGK1	**41**	19	14
PRDX1	PRDX1	**22**	18	21
PIP	PIP	**17**	18	15
NPM1	NPM1	**33**	18	5
PRRC2C	PRRC2C	**317**	17	9
ATP5A1	ATP5A1	**60**	15	9
HRNR	HRNR	**282**	15	7
FNBP4	FNBP4	**110**	15	0
NF45	NF45	**43**	14	21
ANXA2	ANXA2	**39**	13	14
PLP1	PLP1	**30**	13	14
IGF2BP3	IGF2BP3	**64**	13	12
LYZ	LYZ	**17**	12	15
AHNAK	AHNAK	**629**	12	3
DNA topoisomerase 1	TOP1	**91**	12	2
PRCC	PRCC	**49**	12	2
ZCCHC8	ZCCHC8	**79**	11	15
GOT1	GOT1	**46**	11	10
AZGP1	AZGP1	**34**	11	8
EXOSC10	EXOSC10	**101**	11	7
ZC3H11A	ZC3H11A	**89**	11	5
SSB	SSB	**47**	11	3
Matrin-3	MATR3	**100**	10	11

Mass-spectrometric analysis (Table 3.1) further showed that the U2, U5 and U4/U6 di-snRNP proteins, as well as the tri-snRNP specific proteins were equally represented in ΔSmu1/RED B complexes and kinetically-stalled B complexes. Only few peptides corresponding to Smu1 and RED were detected in ΔSmu1/RED B complexes, which is consistent with efficient depletion from the extract. Concurrently, levels of the other B-specific proteins remained essentially unchanged, suggesting that the absence of RED and Smu1 did not affect the recruitment of the other B-specific factors. This indicates that the B-specific proteins do not form a pre-assembled complex as is the case for the Prp19/CDC5L complex or the IBC. Instead, they appear to bind the spliceosome as individual proteins or as small subunits like the Smu1/RED dimer.

In contrast, factors known to associate with the spliceosome during and after the activation, such as proteins from the Prp19/CDC5L complex or the B^{act}-specific factors, were ~50% less abundant in ΔSmu1/RED B complexes as compared with kinetically-stalled B spliceosomes. The presence of these proteins indicates that a small fraction of the kinetically-stalled spliceosomes have already undergone activation. Their clear reduction in the ΔSmu1/RED B complexes is consistent with stalling of the spliceosome assembly process before the B-to-B^{act} transition. Taken together, these results show that B complexes assembled in the absence of Smu1 and RED have an snRNA and protein composition highly similar to that of B complexes assembled in mock-depleted extract, suggesting that Smu1 and RED do not exert their function in spliceosome activation indirectly by altering the recruitment of the other B-specific proteins required for activation.

3.9 The distance between the 5'SS and the branch site, rather than total intron length per se, is the decisive factor for Smu1/RED dependence

The results accrued so far suggested uniformly that Smu1 and RED are required spliceosome activation essentially when introns reach a critical length of ~80 nt. As an intron can be divided into two main segments - the sequence separating the 5'SS and BS and the region between the BS and 3'SS - the question arises whether the intron length *per se* or alternatively the length of one of these sections is the decisive factor for the strong dependence on Smu1 and RED. To address this question, I truncated the introns of the PM5$_{150}$ and PM5-10$_{150}$ pre-mRNAs. PM5$_{150}$ is a pre-mRNA that contains a 5' exon and a 236 nt long intron, with an extended PY tract of 60 nt (Fig. 3.9A). PM5-10$_{150}$ is a derivative of PM5$_{150}$, the PY tract of which has been truncated to 10 nt (Fig. 3.9B). Consequently, the PY tract and intron of PM5-10$_{150}$ are shorter than those in PM5$_{150}$ (where the intron length is 166 nt), while the distance between 5'SS and BS is 150 nt long in both. Comparison of the two pre-mRNAs allows one to determine effects that are based on the distance between BS and 5'SS.

To be further able to assess the importance of the 5'SS-BS distance, the segment of this intron was additionally truncated from 150 to 56 nt in PM5-10$_{56}$, resulting in an overall intron length of 73 nt. The short PY tract of PM5-10 prevents the first splicing step from taking place and results in B^{act} complex stalling. As no splicing products or intermediates can be formed with the PM5-10 constructs, only spliceosome assembly was investigated.

Results

Native gels showed that Smu1/RED depletion resulted in an increased B complex signal and somewhat diminished B^{act} complex signal with PM5-10$_{150}$, indicating that the activation was impaired (Fig. 3.9C). Nevertheless, a strong signal corresponding to the B^{act} complex was still obtained, suggesting that the activation was less efficient but not abolished.
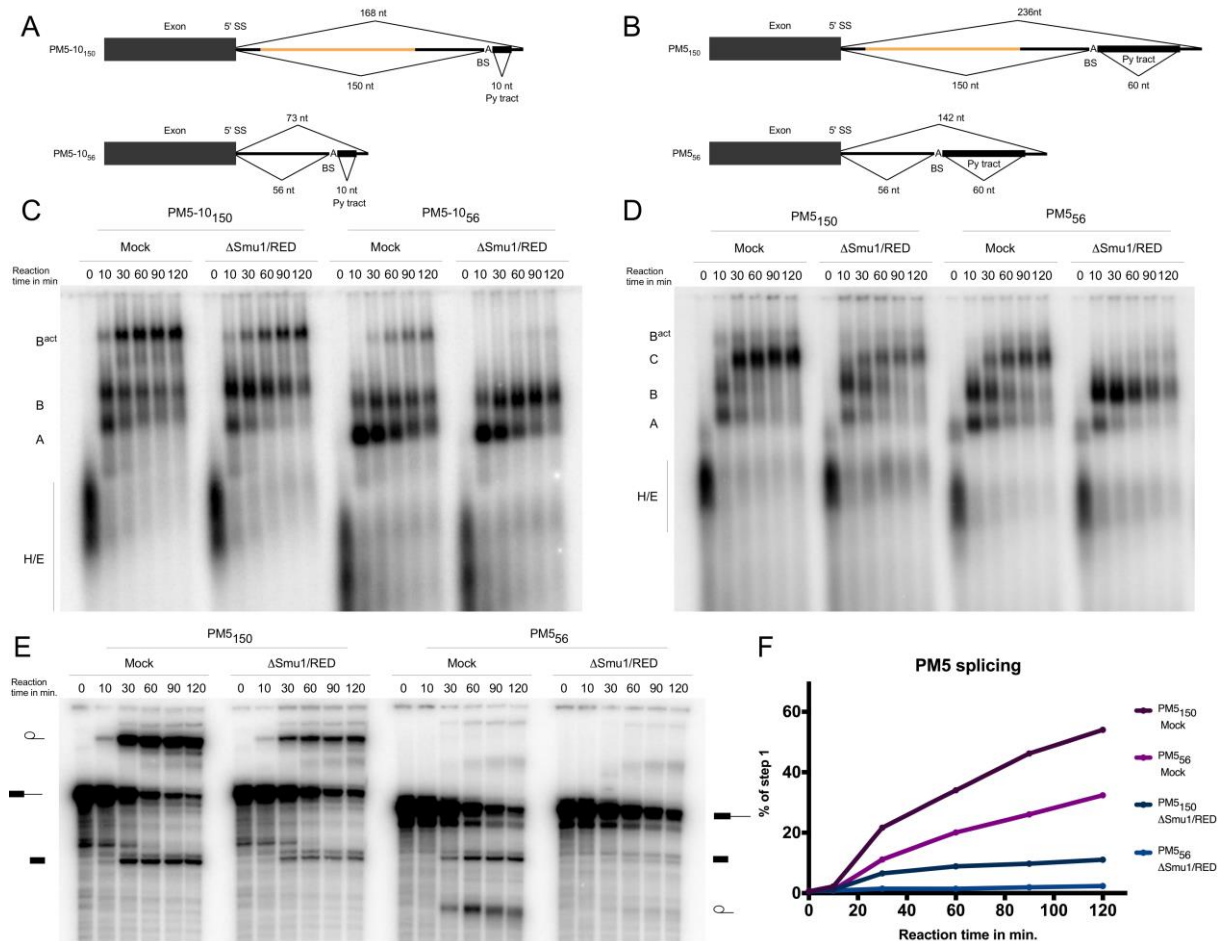

Figure 3.9. PY tract extension does not compensate for a short 5'SS-BS distance. (A) Schematic of PM5-10 constructs. The 5'SS-BS distance from PM5-10$_{150}$ was shortened to 56 nt in PM5-10$_{56}$. Reactive sites (5'SS, BP, 3'SS) and the PY tract remained unchanged. In PM5-10 constructs, truncation of the PY tract to 10 nt blocks spliceosome assembly at the B^{act} complex level. (B) Schematic of PM5 constructs. The 5'SS-BS distance from PM5$_{150}$ was shortened to 56 nt in PM5$_{56}$. Reactive sites (5'SS, BP, 3'SS) and PY tract remained unchanged. Owing to a lack of a 3' exon, stalls spliceosome assembly on PM5 constructs stalls at the C complex stage. (C) and (D) Kinetics of spliceosome assembly on PM5-10 or PM5 constructs in the presence or absence of Smu1/RED. ^{32}P-labelled pre-mRNAs were incubated under splicing conditions in the presence of Smu1/RED-depleted or mock-depleted HeLa nuclear extract for the times indicated. Spliceosomal complex formation was analysed on a native agarose gel and visualized by autoradiography. The positions of the H/E, A, B, C, and B^{act} complexes are indicated on the left. (E) Kinetics of *in vitro* splicing. The ^{32}P-labelled PM5 pre-mRNA was incubated with Smu1/RED-depleted or mock-depleted extract under splicing conditions for the times indicated, separated by denaturing PAGE and visualized by autoradiography. The pre-mRNA and splicing intermediates are indicated on the left and right. (F) Quantification of the kinetics of the formation of splicing intermediates over the reaction time.

Reduction of the distance between 5'SS and BS in PM5-10 from 150 to 56 nt appeared generally to slow down the transition from one complex to the next under conditions of mock-depletion. In ΔSmu1/RED extract, however, only traces of activated spliceosomes were detected, while the B complex signal was stably increased even after 120 min, indicating that the B-to-B^{act} transition was completely blocked.

Truncation of the 5'SS-BS distance in PM5 from 150 to 56 nt resulted in an overall intron length of 142 nt, which is still above the determined threshold of 100 nt. While the extended PY tract of PM5 constructs allows the formation of activated spliceosomes, the assembly is blocked at the C complex level, owing to the lack of a 3' exon. As the first splicing step can nonetheless occur, splicing was additionally investigated by denaturing PAGE. Spliceosome assembly on PM5$_{150}$ appeared less responsive to the Smu1/RED depletion, as a strong increase in B complex signal was only observed after 10 minutes, and it subsequently quickly decreased (Fig. 3.9D). However, the C complex signal in ΔSmu1/RED extract was significantly decreased compared with mock-depleted extract, suggesting that the assembly was still restricted by the Smu1/RED-depletion. Consistent with this, the first splicing step of PM5 was significantly less efficient without Smu1/RED (Fig. 3.9E); while 55% of PM5$_{150}$ had undergone the first splicing step after 120 minutes in mock-depleted extract, this was only the case for 11% of PM5-10$_{150}$ in extract depleted of Smu1/RED. (Fig. 3.9F). Interestingly, shortening of the PM5 intron between 5'SS and BS decreased the efficiency of spliceosome assembly in mock-depleted extract, with ~45% less of the splicing intermediates formed after 120 minutes (~31%). Remarkably, in the absence of Smu1/RED, spliceosome assembly on PM5$_{56}$ was clearly stalled at the B complex level and only traces of C complexes were detected. At the same time, only a very small fraction of the pre-mRNA had been converted to splicing intermediates.

As the PM5 and PM5-10 constructs were similarly affected by the depletion of Smu1 and RED, the distance between the BS and the 3'SS does not appear to be a relevant factor. In contrast, shortening of the 5'SS-BS distance blocked the activation of spliceosomes formed on both PM5$_{56}$ and PM5-10$_{56}$ indicating that this distance is the decisive factor for the Smu1/RED-dependence of short-intron splicing. Consequently, it is not the intron length *per se* that determines whether Smu1 and RED are essential, but the distance between the 5'SS and the BS.

3.10 Cleaving the MINX-80 substrate into two parts restores activation on short introns independently of Smu1/RED

My results indicate that Smu1 and RED jointly support the activation of the spliceosome, and that this support becomes essential when the distance between 5'SS and BS reaches a critical length of ~56 nt or shorter. A reasonable explanation for this length-dependence is that short introns might exercise a physical constraint upon the spliceosome, thereby preventing movements required for the activation. Consequently, releasing the physical constraint exerted through the pre-mRNA on the stalled ΔSmu1/RED spliceosome (by splitting the short MINX-80 intron between the 5'SS and BS) should restore the flexibility that the spliceosome needs in order to adopt the conformation required for its activation and for splicing.

Results

To test this, I physically separated the intron of the MINX-80 pre-mRNA by transcribing it in two independent parts (MINX-80-cleaved; Fig. 3.10A). The first part comprised Exon1 and 25 nt of the intron downstream of the 5'SS and the second part included the 31 remaining nucleotides upstream of the BS, the BS itself, the PY tract and Exon2.
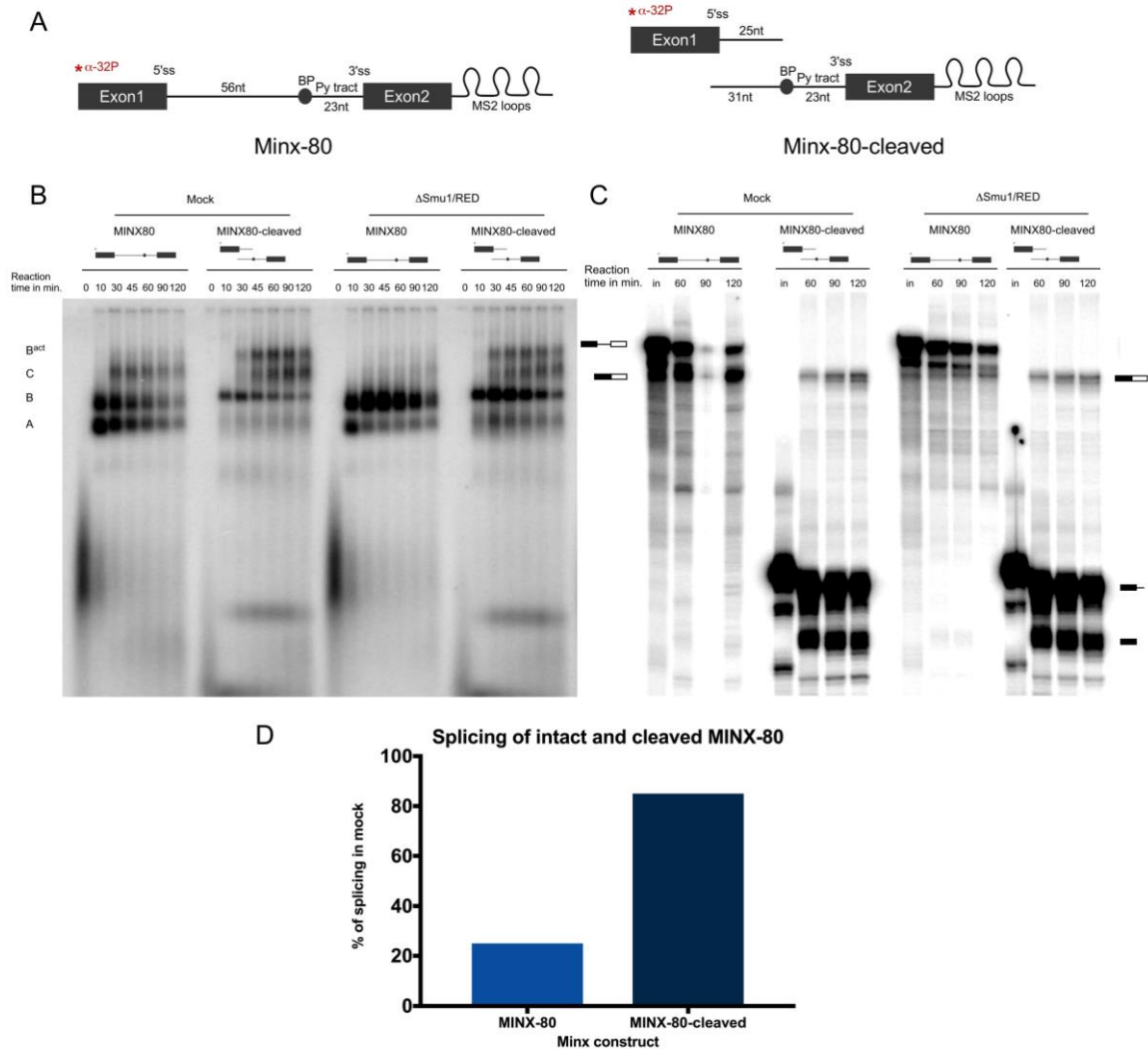

Figure 3.10. MINX-80-cleaved allows spliceosome activation and splicing in the absence of Smu1 and RED. (A) Schematic of MINX-80 and MINX-80-cleaved. To obtain the MINX-80-cleaved construct, MINX-80 pre-mRNA was transcribed in two parts. The first contained Exon 1, the 5'SS and 25 nt of the intron. The second part comprised the rest of the intron, including the BP, the PY tract and the 3'SS, as well as Exon 2. The first part was radiolabelled for detection by autoradiography. The two constructs were identical regarding sequence and intron length. Kinetics of spliceosome assembly (B) and *in vitro* splicing (C) of the intact and cleaved MINX-80 constructs in the presence or absence of Smu1/RED. Pre-mRNAs were incubated under splicing conditions in depleted or mock-depleted extract for the times indicated. (B) Spliceosomal complex formation was analysed on a native agarose gel and visualized by autoradiography. The positions of the H/E, A, B, C, and B^{act} complexes are indicated. (C) RNA was analysed by denaturing PAGE and visualized by autoradiography. The pre-mRNA and splicing intermediates and products are indicated on the right or left. (D) Quantification of pre-mRNA to mRNA conversion in ΔSmu1/RED extract after 120 min relative to the mock sample.

Thus, MINX-80 and MINX-80-cleaved have identical sequences but MINX-80-cleaved is split between its 5'SS and BS. The MINX-80 and MINX-80-cleaved pre-mRNAs were tested in *in vitro* splicing reactions performed with Smu1/RED-depleted or mock-depleted extract, and the splicing outcome and spliceosome assembly were investigated by denaturing and native gel analysis, respectively

As shown in Fig. 3.10B, C and D, B, B^{act} and C complexes were assembled on both MINX-80 and MINX-80-cleaved in mock-depleted extract, and the spliced products could readily be detected, indicating that both constructs are splicing-competent. However, there appeared to be an accumulation of B^{act}/C complexes with the cleaved versus uncleaved MINX-80 pre-mRNA. Also, less A complexes were detected, likely because they do not withstand the native gel conditions when U1 and U2 bind to physically separated pieces of the pre-mRNA. As described above, spliceosome assembly on MINX-80 pre-mRNA in Smu1/RED-depleted extract was stalled at the B complex level, as indicated by a strong accumulation of B complexes at all times, while the B^{act} and C complex signals were reduced as compared with the mock-depleted sample (Fig. 3.10B). Consistent with this, splicing was ~80% less efficient in the absence of Smu1 and RED. However, spliceosomes assembled on MINX-80-cleaved could be activated, as strong signals corresponding to B^{act} and C complexes were detected. Also, the spliced product could readily be detected, although its formation was 20% less efficient than that observed under mock conditions (Fig. 3.10D).

Thus, the release of the physical constraint exerted by the pre-mRNA on the spliceosome in MINX-80-cleaved reactions restored the B-to-B^{act} transition on short introns in the absence of Smu1/RED, but also the formation of splicing product. This strongly supports the idea that the Smu1/RED complex is essential to promote the adoption of a conformation required for spliceosome activation when short introns exert a physical constraint.

3.11 Stalled ∆Smu1/RED B complexes can be activated by addition of the purified dimer

To obtain a better understanding of why spliceosome assembly is stalled in the absence of Smu1/RED, I tested whether ∆Smu1/RED B complexes are functional assembly intermediates or dead-end complexes whose splicing activity cannot be restored. Therefore, I assembled stalled B spliceosomes in ∆Smu1/RED extract for 30 minutes and subsequently added the purified dimer. Concomitantly, a 10-fold excess of unlabelled MINX-80 pre-mRNA was added to sequester newly assembled spliceosomes, and thus ensure that any subsequently formed splicing complexes (i.e., the C complex) could only be derived from the B complexes formed before addition of Smu1/RED. Spliceosome assembly was analysed by native gel electrophoresis (Fig. 3.11A).

The signal of the ∆Smu1/RED-stalled B complexes decreased rather slowly upon addition of High Five Elution Buffer and no B^{act}/C complex formation was detected. In contrast 15 minutes after addition of Smu1/RED a signal corresponding to B^{act} and C complexes could be detected while the signal of the stalled B complexes was significantly reduced after 30 mins. Concomitantly, no clear formation of spliced mRNA or excised intron could be detected at any time after addition of elution buffer to the stalled B complexes (Fig. 3.11B and C). However, when

Results

the purified dimer was added, both mRNA and spliced intron were detected. After 120 minutes ca. 40% of the pre-mRNA had been spliced, which is approximately 50% of the mock control. This indicates that the stalled B complexes could be chased into active spliceosomes that catalyse splicing. Consequently, the B complexes that accumulate in the absence of Smu1/RED are functional assembly intermediates, which are merely stalled prior to activation.
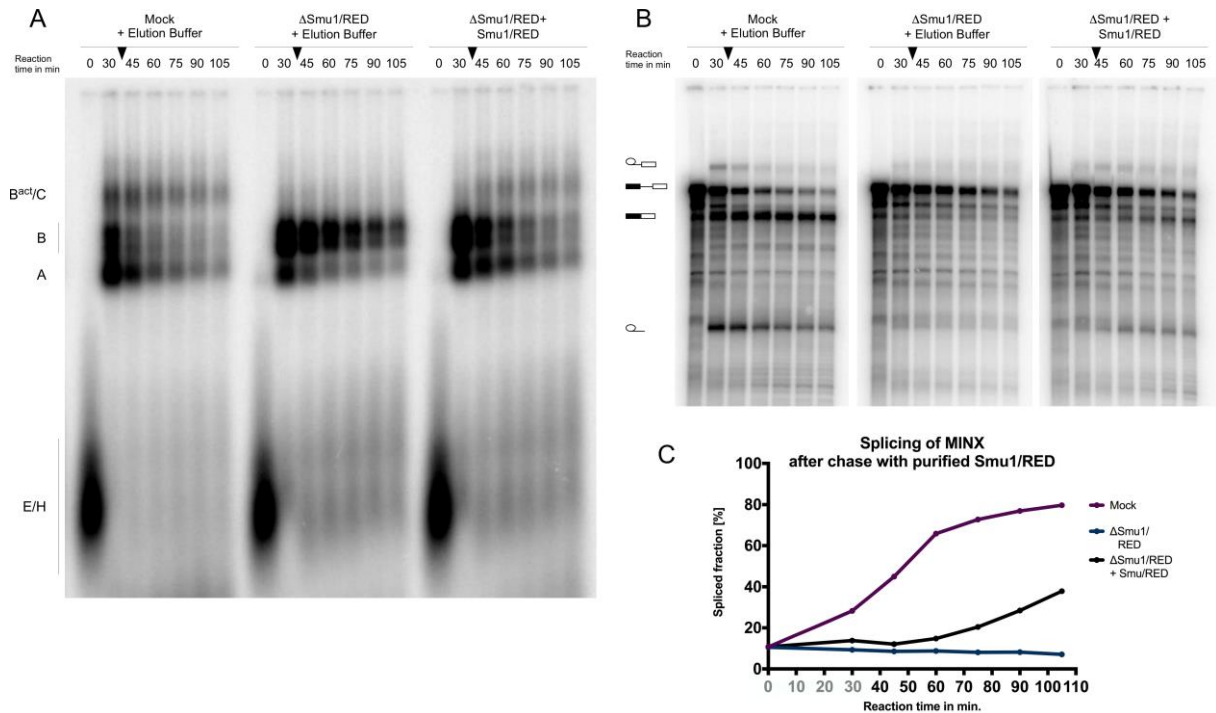

Figure 3.11. ΔSmu1/RED B complexes can be chased into activated spliceosomes. Kinetics of spliceosome assembly (A) and *in vitro* splicing (B) after chase with the Smu1/RED dimer, co-expressed in insect cells. ΔSmu1/RED B complexes were assembled on ^{32}P-labelled MINX-80 pre-mRNA under splicing conditions for 30 minutes. Subsequently, purified Smu1/RED dimer or High Five Elution Buffer was added (black arrow) along with a 10-fold excess of cold pre-mRNA. (A) Spliceosomal complex formation was analysed on a native agarose gel and visualized by autoradiography. The positions of the H/E, A, B, C, and B^{act} complexes are indicated on the left. (B) RNA was analysed by denaturing PAGE and visualized by autoradiography. The pre-mRNA and splicing intermediates and products are indicated on the left. (C) Quantification of the kinetics of the conversion of pre-mRNA to mRNA. Measurements indicated in grey were taken before addition of the elution buffer or purified Smu1/RED dimer, and those indicated in black after this.

3.12 The interaction between Smu1 and RED is essential for their function in splicing

Smu1 and RED form a stable module, which was shown to improve the stability of both of these proteins (Spartz *et al.*, 2004). It is likely that the interaction is also important for their function. To investigate how each protein contributes to the dimer's function, I performed a series of complementation experiments using the MINX-80 pre-mRNA and Smu1/RED depleted extract. The tight interaction of Smu1 and RED had made separate investigations *in vivo* and *in vitro* impossible.

Therefore, to examine to which extent each protein contributes to in splicing, I separately expressed and purified recombinant His-Smu1 and His-RED proteins from insect cells. For more detailed functional studies, I additionally generated truncated Smu1 and RED proteins and investigated the effect on spliceosome activation.

3.12.1 Smu1 and RED can only support the activation co-operatively

The cryo-EM structure of the human B complex revealed that the interaction of Smu1 and RED is maintained in the spliceosome, suggesting that this interaction is important for their function. Both Smu1 and RED contact U2 and tri-snRNP proteins, suggesting that they stabilize the interaction of the U2 and the tri-snRNP within the B complex, which appears to be important during the spliceosome activation step (Bertram, Agafonov, Dybkov *et al.*, 2017). However, it is unclear whether both proteins are required to provide sufficient stability or if one would be sufficient for this task. To that end, I investigated whether Smu1 or RED alone is able to restore splicing activity, by complementing ΔSmu1/RED nuclear extract with the individually purified proteins (Fig. 3.12A) and analysing the formation of spliceosome assembly intermediates by native gel electrophoresis (Fig. 3.12B). A strong accumulation of B complexes was detected in Smu1/RED-depleted extract and addition of buffer alone did not relieve this accumulation. Interestingly, when Smu1 or RED were added separately little or no effect on B complex accumulation was observed and spliceosomes did not undergo activation. However, complementation with both single proteins together promoted the B-to-B^{act} transition as efficiently as the co-expressed dimer did. Concurrently, no splicing product was formed when the single proteins were added, while splicing was very efficient in the presence of both proteins (3.12C) This shows that the individually purified proteins are functional but cannot fulfil their function separately. Furthermore, it shows that Smu1 and RED do not need to be co-expressed to form a functional heterodimer.

While the interaction of Smu1 and RED appears to be crucial for their functionality, it is unclear whether dimer formation is required for their association with the spliceosome. To answer this question, I purified the B complexes assembled in the presence of RED, Smu1 or both proteins and analysed their association with the spliceosome by immunoblotting. Interestingly, addition of each single protein resulted in spliceosomes that migrated more slowly in the glycerol gradient (fractions 14–15) than complexes containing both proteins (fractions 15–16); this resembles the sedimentation behaviour of ΔSmu1/RED and kinetically-stalled B complexes, respectively (Fig. 3.13A).

Western blot analysis of the purified B complexes showed that a strong signal for both proteins was seen when they were added together (Fig. 3.13B). However, when only one protein was added only low levels of RED were detected and Smu1 was nearly absent. This suggests that the single proteins - especially Smu1 - do not bind efficiently to the spliceosome on their own and, thus, that their interaction is important for efficient association of both proteins.
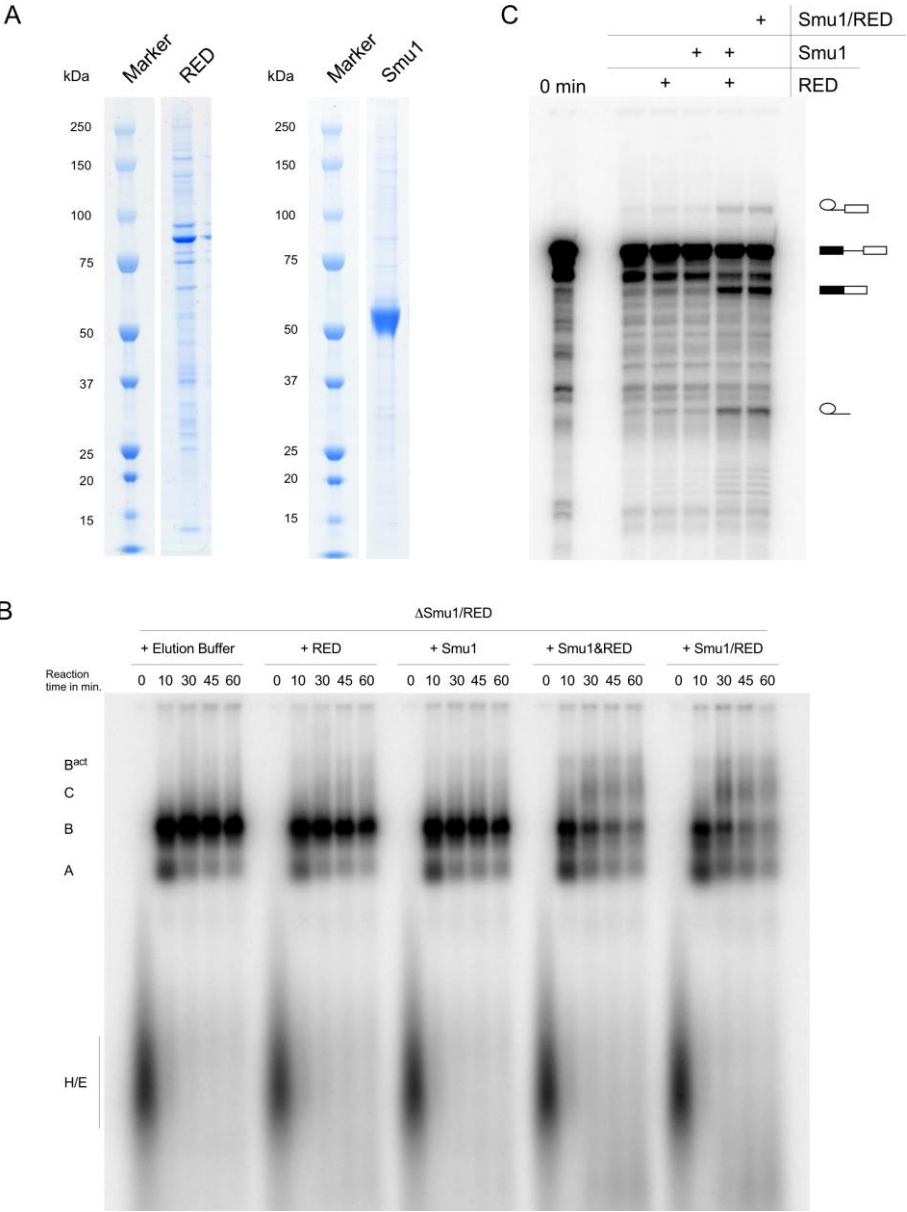
Results

Figure 3.12. The interaction between Smu1 and RED appears to be important for their function. (A) Smu1 and RED, individually expressed and purified from High Five insect cells were separated by SDS-PAGE and stained with Coomassie. (B) Single proteins were added separately or together to Smu1/RED-depleted splicing extract. After 15 minutes incubation under splicing conditions, ^{32}P-labelled MINX-80 was added and incubated for the times indicated. Spliceosome assembly was monitored by native gel electrophoresis. As controls, High Five Elution Buffer without protein or the functional co-expressed (Smu1/RED) dimer was added. (C) Splicing of MINX-80 after 60 minutes in the presence of the single proteins or both Smu1 and RED, as indicated in the table above. RNA was analysed by denaturing PAGE and visualized by autoradiography. The pre-mRNA and splicing intermediates and products are indicated on the right.
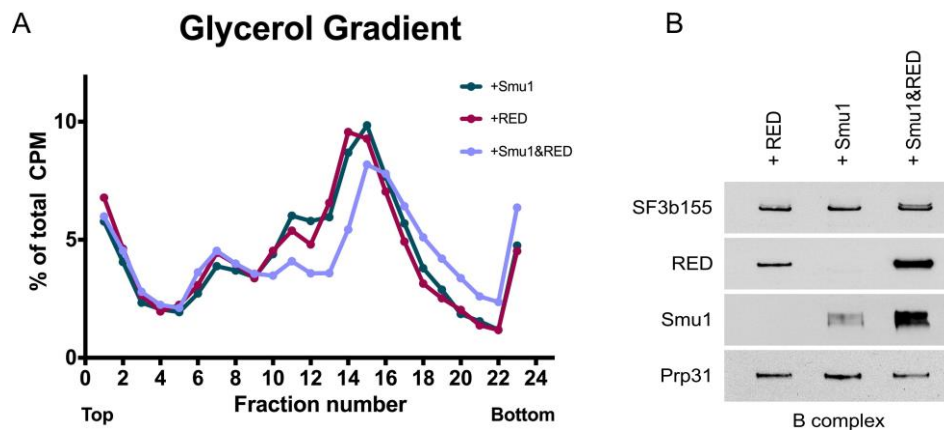
Figure 3.13. The interaction between Smu1 and RED enhances their association with the spliceosome. (A) Purified B complexes assembled in the presence of Smu1, RED or both proteins were seperated by glycerol-gradient ultracentrifugation and their migration was determined by Cherenkov counting. (B) The spliceosomal complexes in the peak fractions (+ Smu1 and + RED, fractions 14 & 15; + Smu1 & RED, fractions 15 & 16) were subjected to MS2 affinity purification and analysed by SDS-PAGE followed by immunoblotting using antibodies against SF3b155, RED, Smu1, and Prp31.

3.12.2 The WD40 domain of Smu1 is required for association of Smu1/RED with the spliceosome

The structure of the human B complex showed that the WD40 domain of Smu1 is located directly at the interface of the U2 protein SF3B3 and the U5-associated helicase Brr2 (Bertram, Agafonov, Dybkov *et al.*, 2017), which catalyses the unwinding of the U4/U6 duplex during activation. It is possible that the Smu1-WD40 domain stabilizes the positioning of Brr2 relative to its pre-mRNA substrate, thereby regulating the activity of the helicase. Consequently, the Smu1-WD40 domain might be crucial for the function of the Smu1/RED dimer.

To assess the importance of this region for the B-to-B^{act} transition, I expressed a truncation mutant of Smu1, lacking the C-terminal WD40 domain (Fig. 3.14A). The short remaining fragment still included the RED interaction surface and could be co-purified with full-length (FL) RED (Fig. 3.14B), implying that hetero-dimerization was not inhibited.

As shown in Fig. 3.14C, complementation with Smu1-ΔWD40/RED did not restore spliceosome activation, as the B complexes remained stable over the entire incubation time, comparable to the negative control (with elution buffer), while barely any later complexes could be detected. Consistently with this, barely any splicing product had been formed in the presence of the Smu1-ΔWD40/RED dimer (Fig. 3.14D). This suggests that the Smu1-WD40 domain is indeed important for the function of the Smu1/RED dimer. Closer examination of the purified complexes by immunoblotting revealed that the Smu1-ΔWD40/RED dimer could not associate with the spliceosome (Fig. 3.14E).

Results

Whereas Smu1/RED was readily detectable in the control complexes, Smu1-ΔWD40/RED could barely be detected in the purified B complexes. As the association of the dimer appears to be highly dependent on the WD40 domain of Smu1, an investigation of its direct role during the activation of the spliceosome is not possible.
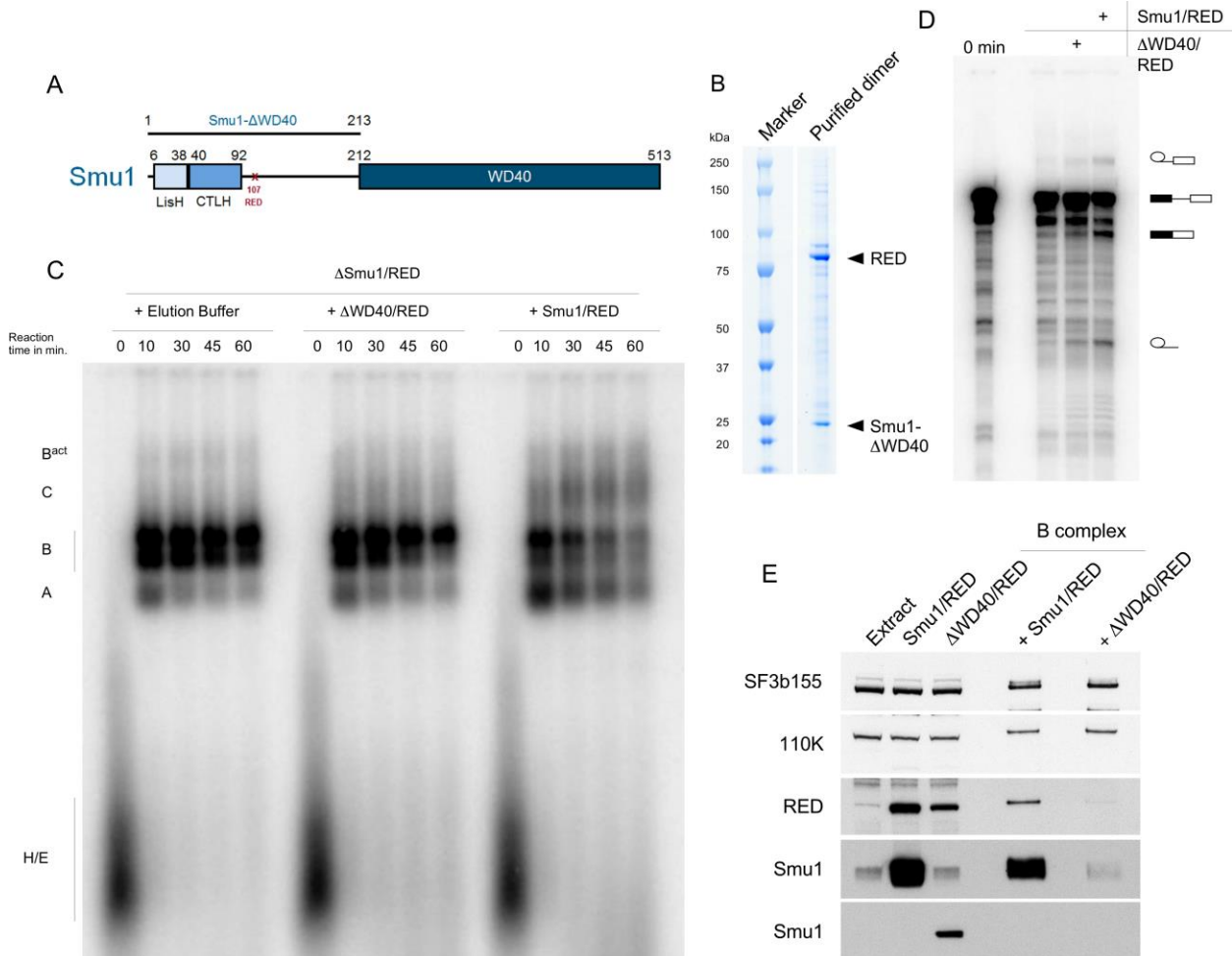

Figure 3.14. The WD40 domain of Smu1 is important for interaction of the dimer with the spliceosome and for its function. (A) Schematic of Smu1. Structural domains are indicated by blue boxes. The amino acid that was crosslinked to RED in the human B complex (Bertram, Agafonov, Dybkov et al., 2017) is marked by a red "x" and indicates the RED interaction region. The part of Smu1 that is included in the Smu1-ΔWD40 mutant is indicated above. To investigate the function of the WD40 domain of Smu1, a truncated version of the protein (Smu1-ΔWD40) was co-expressed with RED in insect cells. (B) Smu1-ΔWD40 and RED, co-expressed and purified from High Five insect cells were separated by SDS-PAGE and stained with Coomassie. (C) Kinetics of spliceosome assembly on MINX-80 in the absence/ presence of Smu1/RED or Smu1-ΔWD40/RED. The co-expressed Smu1-ΔWD40/RED dimer was incubated with Smu1/RED-depleted extract under splicing conditions and subsequently with ^{32}P-labelled MINX-80 pre-mRNA for the indicated times. The complexes formed were analysed on a native gel. As controls, the buffer without protein or the functional co-expressed dimer was added. (D) Splicing of MINX-80 after 60 minutes in the presence of Smu1/RED or Smu1-ΔWD40/RED. RNA was analysed by denaturing PAGE and visualized by autoradiography. The pre-mRNA and splicing intermediates and products are indicated on the right. (E) The B complexes assembled in the presence of Smu1/RED or Smu1-ΔWD40/RED were resolved by glycerol-gradient ultracentrifugation and affinity-purified. Their protein composition was investigated by immunoblotting with the antibodies indicated at the left. Smu1/RED-depleted extract alone or complemented with Smu1/RED or Smu1-ΔWD40/RED, was loaded a reference for protein migration.
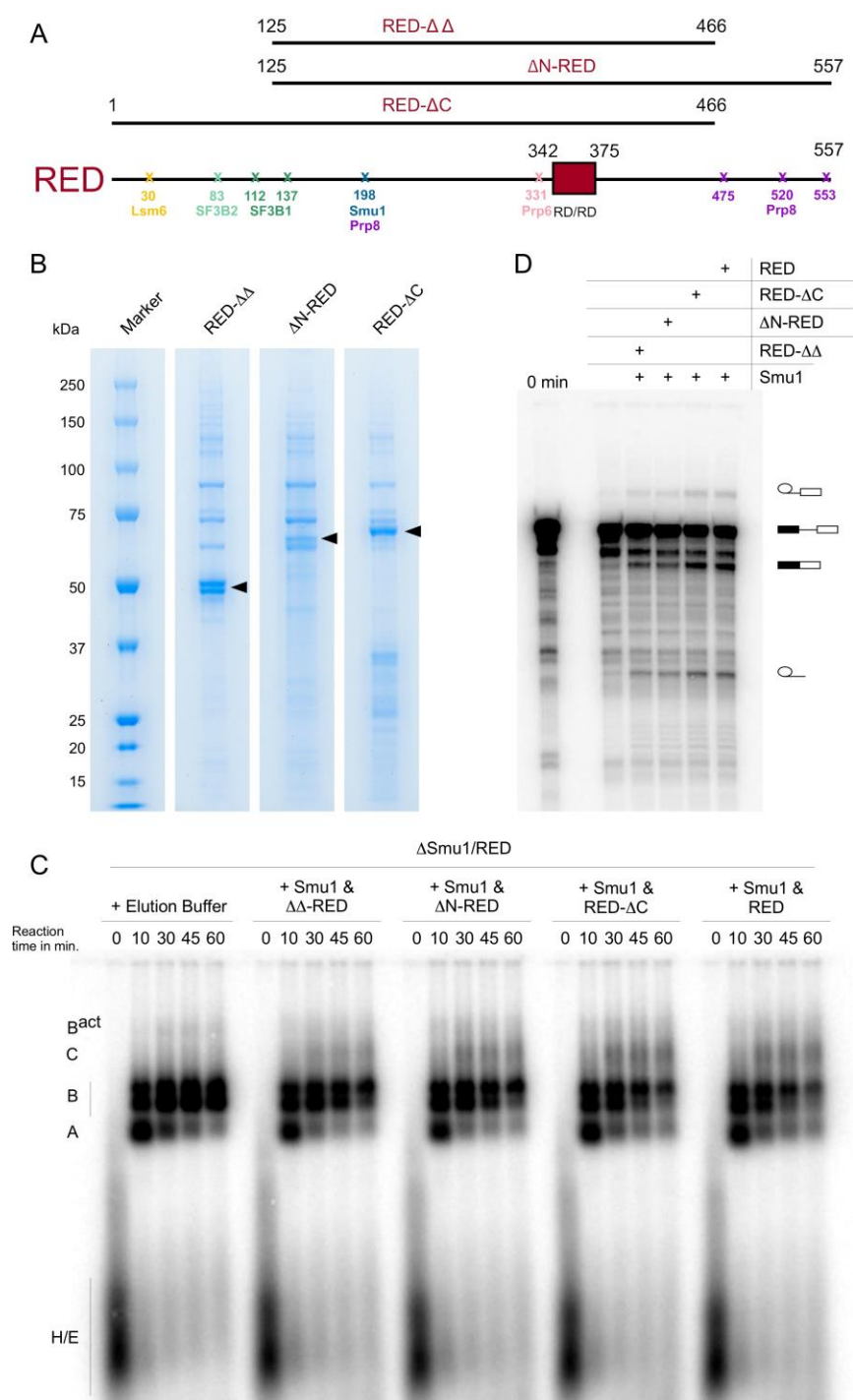
3.12.3 Structural requirements of RED for supporting B^act complex formation

Figure 3.15. RED contributes to the function of the Smu1/RED dimer. (A) Schematic of RED. The amino acids that were crosslinked to other proteins in the human B complex (Bertram, Agafonov, Dybkov et al., 2017) are shown by a coloured "x" indicating the interaction areas. The parts included in the RED-ΔΔ, ΔN-RED or RED-ΔC mutants are shown above. (B) RED-ΔΔ, ΔN-RED or RED-ΔC mutants, expressed and purified from High Five insect cells were separated by SDS-PAGE and stained with Coomassie. (C) Kinetics of spliceosome assembly on MINX-80 in the absence/presence of Smu1/RED or the indicated Smu1/RED-mutant. Together with Smu1-FL, the mutants were added to Smu1/RED-depleted splicing reactions and the complexes formed were analysed on a native gel. As controls, the buffer without protein or RED-FL was added. (D) Splicing of MINX-80 after 60 minutes in the presence of Smu1 and the different RED mutants as indicated in the table above. RNA was analysed by denaturing PAGE and visualized by autoradiography. The pre-mRNA and splicing intermediates and products are indicated on the right.

Results

To determine the role of RED in splicing, the protein was truncated N- and/or C-terminally (Fig. 3.15A) and added to the splicing reaction together with full-length Smu1. Investigation of the purified RED mutant proteins by SDS-PAGE (Fig. 3.15B), revealed bands of the expected sizes for all of them. As could be observed for individually expressed and purified RED, some extra bands could be detected. Interestingly, co-expression with Smu1, did not improve the yield or purity of the protein.

Investigation of spliceosome formation on native gels (Fig. 3.15C) revealed that when the double mutant RED-ΔΔ was added, B complexes remained stable for 60 minutes and minor amounts of C complexes were detectable. When ΔN-RED or RED-ΔC were added, the B complex signal decreased over time, while a clear C complex signal appeared after 30 minutes. This indicates that truncation of RED either N- or C-terminally alone does not abolish the function of the dimer, but that a full-length RED protein leads to more efficient dimer function. Interestingly, investigation of the splicing of MINX-80 revealed that minor amounts of splicing product were formed when RED-ΔΔ or ΔN-RED was added to the splicing mixture, while splicing in the presence of RED-ΔC was similarly efficient as that observed with the full-length protein (Fig.3.15D). This suggests that the N-terminus of RED plays a more important role than its C-terminus.

3.13 MINX-70 stalls spliceosome assembly after tri-snRNP association but before efficient U1 displacement

Splicing with the MINX-70 pre-mRNA was completely abolished, consistent with previous studies, indicating that splicing requires a minimal distance between the 5'SS and the BS of 50 nt (Smith & Nadal-Ginard, 1989; Himmelspach *et al.*, 1991). However, both B and A complexes appeared to form (Fig. 3.5), in contrast to previous results.
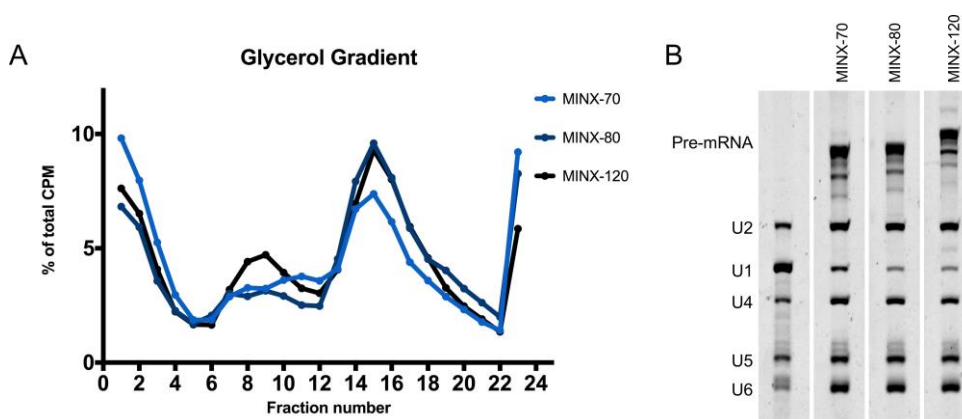

Figure 3.16. Spliceosomes formed on MINX-70 contain higher amounts of U1. (A) Glycerol-gradient sedimentation profile of complexes assembled on MINX-70, MINX-80 and MINX-120. MINX-70 complexes were assembled for 12 min and kinetic B complexes were assembled on MINX-80 or MINX-120 for 8 min under splicing conditions. The migration of the ^{32}P-pre-mRNA was determined by Cherenkov counting. (B) snRNA composition of the spliceosomal complexes. The spliceosomal complexes in the peak fractions (fractions 14–16) were subjected to MS2 affinity purification and the RNA extracted from the purified complexes was analysed by denaturing PAGE followed by staining with SYBR Gold.

Table 3.2: Protein compositions of spliceosomal complexes assembled on MINX-70, -80 and -120. Proteins of the affinity-purified indicated complexes were identified by LC-MS/MS after separation by 4–12 % SDS-PAGE. The number shown is the total number of peptides sequenced for each protein indicated. 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.

Protein	hGene Name	kDa	MINX70	MINX80	MINX120
Sm proteins					
B	SNRPB	24	84	68	72
D1	SNRPD1	13	56	36	44
D2	SNRPD2	14	91	71	88
D3	SNRPD3	14	75	62	78
E	SNRPE	11	37	28	22
F	SNRPF	10	21	17	11
G	SNRPG	8	21	13	12
U1 snRNP					
U1-70K	SNRNP70	52	30	20	23
U1-A	SNRPA	31	8	4	6
U1-C	SNRPC	17			
17S U2 snRNP					
U2A'	SNRNPA1	28	107	93	86
U2B"	SNRPB2	25	29	21	24
SF3a120	SF3A1	89	252	192	230
SF3a66	SF3A2	51	31	28	34
SF3a60	SF3A3	59	153	136	166
SF3b155	SF3B1	146	444	438	521
SF3b145	SF3B2	100	203	209	237
SF3b130	SF3B3	136	529	473	491
SF3b49	SF3B4	44	22	33	29
SF3b14a/p14	SF3B14	15	36	36	41
SF3b10	SF3b5	10	8	15	9
17S U2 related					
hPRP43	DHX15	91	103	69	81
U2AF65	U2AF2	54	45	37	28
SPF45	RBM17	45	20	12	12
SR140 (fSAPa)	U2SURP	118	39	23	34
CHERP	CHERP	104	26	15	28
SF3b125	DDX42	103	9	3	7
U2AF35	U2AF1	28	11	14	11
SPF31	DNAJC8	30			
hPRP5	DDX46	117	3	1	5
PUF60	PUF60	60	8	4	9
SPF30	SMNDC1	27	26	20	6
U5					
220K	PRPF8	274	876	786	739
200K	SNRNP200	245	976	942	976
116K	EFTUD2	109	529	475	507
40K	WDR57	39	117	108	112
102K	PRPF6	107	400	349	323
15K	TXNL4A	17	17	20	20
100K	DDX23	96	203	108	111
52K	CD2BP2	38	18	24	21
LSm proteins					
LSm2	LSM2	11	26	28	19
LSm3	LSM3	12	8	9	8
LSm4	LSM4	15	18	14	10
LSm5	LSm5	10	2	1	1
LSm6	LSM6	9	10	5	7
LSm7	LSM7	12	9	10	11
LSm8	LSM8	10	26	20	19
U4/U6					
90K	PRPF3	78	206	201	190
60K	PRPF4	58	194	185	189
20K	PPIH	19	35	29	24
61K	PRPF31	55	88	88	71
15.5K	NHP2L1	14	13	10	19
U4/U6.U5					
110K	SART1	90	382	274	340
65K	USP39	65	100	63	58
27K		27			
RBM42	RBM42	48	7	7	9

Results

Protein	hGene Name	kDa	MINX70	MINX80	MINX120
A complex					
THRAP3	THRAP3	109	20	24	18
CCAR1	CCAR1	133	9	3	3
RBM5	RBM5	92	22	2	15
S164 (fSAP94)	RBM25	100	7	5	8
RBM10	RBM10	104			
SF1	SF1	68	4	3	4
FBP11	PRPF40A	109	10	7	5
SF4	SUGP1	72	2	0	1
Pre-B complex					
hPRP4-Kinase	PRPF4B	115	132	101	89
B complex					
hPRP38	PRPF38A	37	79	68	84
hSnu23	ZMAT2	24	12	14	15
MFAP1	MFAP1	52	43	41	53
RED	IK	66	120	131	161
hSmu-1 (fSAP57)	SMU1	58	132	151	161
FBP21	WBP4	43	15	20	19
Npw38BP	WBP11	70	26	36	57
Npw38	PQBP1	30	6	3	11
UBL5	UBL5	9	6	6	6
hPRP19/CDC5L					
hPRP19	PRPF19	55	107	116	175
CDC5L		92	128	122	283
SPF27	BCAS2	26	28	33	52
PRL1	PLRG1	57	37	76	136
CCAP1	HSPA8	71	32	37	39
CCAP2	CWC15	27	7	4	11
catenin, b-like 1	CTNNBL1	62	23	35	58
CRNKL1 (hSYF3)	CRNKL1	99			
SKIP	SNW1	61	61	69	154
RBM22	RBM22	47	16	23	45
PPIase-like 1 (PPIL1)	PPIL1	18	4	4	17
G10 (fSAP17)	BUD31	17	8	14	23
IBC					
KIAA0560 (fSAP164)	AQR	171	112	150	266
hSyf1	XAB2	100	76	94	209
MGC20398	ZNF830	42	11	17	22
hIsy1 (fSAP133)	ISY1	33	16	27	47
Cyp-E	PPIE	33	6	16	29
B^{act} complex					
KIAA1604 (fSAPb)	CWC22	106	25	30	71
hPRP17	PRPF17/CDC40	66	27	25	53
NY-CO-10	CWC27	54	9	12	22
RNF113A	RNF113A	39	6	5	14
PPIL2 (Cyp-60)	PPIL2	60	29	37	60
PRCC	PRCC	52	9	13	23
MGC23918	CCDC12	21	12	13	20
SRm160	SRRM1	103	10	8	25
SRm300	SRRM2	300	45	58	173
PPIase-like 3b	PPIL3	19			
1st step factors					
hPRP2	DHX16	119	41	44	115
GPKOW	GPKOW	52	9	15	23
CCDC49	CWC25	50	7	10	32
C complex					
GCIP p29 (fSAP29)	SYF2	29	4	6	17
C9orf78 (HSPC220)	C9orf78	34	0	0	6
WDR70	WDR70	33	4	2	29
NOSIP	NOSIP	33			
PPWD1 (KIAA0073, CyP64)	PPWD1	74	12	6	54
PPIL4	PPIL4	35	23	24	44
CDK10	CDK10	35			
PPIG (SRcyp)	PPIG	89	4	2	13
2nd step factors					
hPRP22	DHX8	139	17	7	63
hPRP18	PRPF18	40			
hSlu7	SLU7	68	3	0	16
hPRP16	DHX38	140	14	24	43

Protein	hGene Name	kDa	MINX70	MINX80	MINX120
Putative C* proteins					
DDX35	DHX35	**79**	8	10	23
Abstrakt	DDX41	**70**	11	6	31
C19orf29 (NY-REN-24, cactin)	CACTIN	**89**	3		25
CXorf56 (FLJ22965)	CXorf56	**26**	2	0	7
FAM50A (XAP5)	FAM50A	**40**	1		8
SR proteins					
SF2/ASF	SRSF1	**28**	71	63	74
SC35	SRSF2	**25**			
SRp20	SRSF3	**19**	12	13	15
SRp75	SRSF4	**57**	25	24	23
SRp40	SRSF5	**31**	36	35	32
SRp55	SRSF6	**40**	39	34	36
9G8	SRSF7	**27**	38	33	41
SRp46	SRSF8	**32**			
SRp30c	SRSF9	**26**	23	14	20
SRp38	SRSF10	**31**	22	16	14
SRSF11	SRSF11	**54**			
hTra-2 alpha	TRA2A	**33**			
hTra-2 beta (SFRS10)	TRA2B	**34**			
hnRNP					
hnRNP A0	HNRNPA0	**31**	9	14	22
hnRNP A1	HNRNPA1	**39**	74	111	132
hnRNP A3	HNRNPA3	**40**	14	18	21
hnRNP AB	HNRNPAB	**35**	5	2	2
hnRNP A2/B1	HNRNPA2B1	**37**	48	53	62
hnRNP C	HNRNPC	**33**	39	22	24
hnRNP D	HNRNPD0	**24**			
hnRNP DL	HNRNPDL	**46**	4	2	4
hnRNP F	HNRNPF	**46**	8	5	7
hnRNP H1	HNRNPH1	**49**	16	6	10
hnRNP H3	HNRNPH3	**37**	4	1	3
hnRNP K	HNRNPK	**51**	10	9	10
hnRNP L	HNRNPL	**64**	7	5	4
hnRNP M	HNRNPM	**78**	17	15	27
hnRNPQ	HNRNPQ	**70**	37	23	28
hnRNP R	HNRNPR	**71**	45	31	44
hnRNP U	HNRNPU	**91**	20	12	22
hnRNP UL-1	HNRNPUL1	**72**	12	7	6
hnRNP UL-2	HNRNPUL2	**91**	6	1	7
PCBP1	PCBP1	**38**	12	14	8
PCBP2	PCBP2	**32**			
RALY	RALY	**33**	5	6	3
EJC/mRNP					
eIF4A3	EIF4A3	**47**	7	8	16
magoh	MAGOH (B)	**17**	2	0	7
Y14	RBM8A	**20**	2	1	4
Pinin	PNN	**82**	2	1	2
Acinus (fSAP152)	ACIN1	**147**	22	27	40
SAP18	SAP18	**18**	2	1	2
UAP56	DDX39B	**49**	6	8	12
RNPS1	RNPS1	**34**	3	4	2
RES complex					
SNIP1	SNIP1	**46**	19	15	23
MGC13125 (fSAP71)	BUD13	**71**			
CGI-79	RBMX2	**40**			
TREX					
THOC1	THOC1	**76**	14	13	19
THOC2	THOC2	**183**	37	21	28
THOC3	THOC3	**39**	5	2	0
Aly/REF (THOC4)	ALYREF	**28**	3	2	5
THOC5 (KIAA0983, fSAP79)	THOC5	**79**	16	12	12
THOC6 (WDR58, MGC2655)	THOC6	**38**	6	5	3
THOC7 (FLJ23445, fSAP24)	THOC7	**24**	4	2	2
mRNA binding proteins					
PABP1	PABPC1	**71**	21	20	16
PABP2	PABP2	**33**	6	6	4
PABPC4 (IPABP, APP1)	PABPC4	**17**	17	10	18
PABPN1	PABPN1	**33**	6	6	4
YB-1	YBX1	**36**	58	60	63
YBX3	YBX3	**40**	53	54	64
ELAV	ELAVL1	**39**	13	9	11
ASR2B	SRRT	**100**	165	89	91
KIN17	KIN17	**45**	15	31	34
RBM4	RBM4	**40**	14	10	12
RBMX	RBMX	**42**	15	10	11
RBM14	RBM14	**69**	16	11	9
RBM15	RBM15	**106**	8	4	11

Results

Protein	hGene Name	kDa	MINX70	MINX80	MINX120
cap binding complex					
CBP80	NCBP1	**92**	49	38	39
CBP20	NCBP2	**18**	27	19	21
NCBP3	NCBP3	**71**			
DExD/H-box RNA Helicases					
p68 (DDX5)	DDX5	**69**	22	17	35
DHX36	DHX36	**115**	11	9	10
DDX3X	DDX3X	**73**	13	14	17
p72 (DDX17)	DDX17	**81**	19	14	19
DDX9	DHX9	**142**	35	37	59
Transcription					
TCERG1	TCERG1	**124**	26	46	82
Miscellaneous					
SPEN	SPEN	**402**	71	103	134
DKFZp762M013		**87**	61	83	173
Rnpc2	Rnpc2	**57**	60	58	52
TFIP11	TFIP11	**97**	56	36	51
PAXBP1	PAXBP1	**105**	53	55	63
ZC3HAV1	ZC3HAV1	**101**	39	25	29
SKIV2L2	SKIV2L2	**118**	37	36	35
NF110b		**96**	29	26	45
SSB	SSB	**47**	28	29	22
DKFZp686E2459		**110**	27	15	20
SON	SON	**264**	24	20	33
PHF5A	PHF5A	**12**	24	20	26
Matrin-3	MATR3	**100**	23	9	14
RBBP6	RBBP6	**202**	20	22	16
FLJ76180		**71**	18	41	71
SETD2	SETD2	**288**	18	9	7
ZCCHC8	ZCCHC8	**79**	17	19	16
LARP7	LARP7	**67**	17	10	14
BCLAF1	BCLAF1	**106**	16	20	26
Li 3a		**35**	16	13	15
PRRC2C	PRRC2C	**317**	16	11	4
SFSWAP	SFSWAP	**105**	16	1	3
GAPDH		**36**	15	17	12
GPANK1	GPANK1	**39**	15	17	12
Peptidyl-prolyl cis-trans isomerase		**18**	14	16	11
NF45		**43**	14	13	17
EXOSC10	EXOSC10	**101**	14	6	13
PRRC2A	PRRC2A	**229**	14	9	4
FUS interacting protein		**31**	13	8	15
IGF2BP3	IGF2BP3	**64**	11	7	13
JUP	JUP	**82**	11	7	13
RSRC1	RSRC1	**39**	10	6	11
NPM1	NPM1	**33**	10	4	4
IGF2BP1	IGF2BP1	**63**	10	6	9
STRBP	STRBP	**74**	9	8	15
ZFR	ZFR	**117**	9	8	13
FLJ94609		**66**	9	7	12
C17orf85		**71**	8	6	12
RNA-binding protein		**32**	7	25	22
Heat shock 70kDa protein 1A		**77**	7	8	10
BCDIN3		**74**	5	12	11
SCAF11	SCAF11	**165**	4	10	7
HSPA5	HSPA5	**72**	4	11	11
MSI2	MSI2	**35**	3	18	9
PPP1CA	PPP1CA	**38**	3	13	11
WDR83	WDR83	**34**	3	2	11
SLC4A1AP	SLC4A1AP	**89**	2	10	24
PPP1R8	PPP1R8	**38**	2	5	12
TTC14	TTC14	**88**	1		14

Interestingly, B complex formation on MINX-70 featured a migration behaviour similar to that of kinetic B complexes assembled on MINX-80 and MINX-120, with the main peak in fractions 14–16 (Fig. 3.16A). Complexes affinity-purified from the peak fractions 14–16 contained stochiometric amounts of U2, U4, U5 and U6 for all of the pre-mRNAs (Fig. 3.16B), while U1 was enriched in MINX-70 spliceosomes, as compared with complexes assembled on MINX-80 and -120. This suggests that MINX-70 indeed allows association of the tri-snRNP with the A complex, but the release of U1 is impaired. Interestingly, the protein composition of MINX-70-associated spliceosomes was barely distinguishable from kinetic B complexes assembled on MINX-80 (Table 3.2). Furthermore, the MINX-70 and MINX-80 complexes differed from B spliceosomes assembled on MINX-120 only in reduced levels of later splicing factors such as the Prp19/CDC5L and intron binding complex as well as B^{act}, first-step, C and second-step factors.

The presence of U4, U5 and U6 as well as tri-snRNP proteins indicates that the tri-snRNP is able to associate with A complexes assembled on MINX-70. The enrichment of the U1 snRNA compared with kinetic B complexes is reminiscent of the pre-B complex. This recently identified intermediate complex includes a loosely associated tri-snRNP before its stable integration by Prp28-driven removal of U1 from the 5'SS, resulting in the formation of a stable 5'SS/U6 ACAGAG helix. However, the B-specific proteins, which have been shown to be absent in pre-B complexes were highly abundant in spliceosomes assembled on MINX-70. It is possible that the short intron found in MINX-70 impairs the formation of a stable 5'SS/U6 ACAGAG helix, which might be formed only in equilibrium and still promote the binding of the B-specific proteins. Thus, MINX-70 might allow the formation of an intermediate spliceosome between the pre-B and B complexes.

3.14 Smu1 and RED are required for spliceosome activation *in vivo*

In vitro experiments pointed to a connection between Smu1/RED-induced defects in splicing and a delayed or stalled B-to B^{act} transition, which was characterized by an accumulation of the B complex. To determine whether the changes in splicing observed upon Smu1 and RED depletion were also caused by impaired activation of the spliceosome *in vivo*, I depleted each protein from HeLa cells by RNAi and investigated the effect on spliceosome assembly. For each protein, two siRNAs with different target sequences were tested, which gave comparable results in all cases. Thus, only one siRNA was used in the subsequent experiments.

As the investigation of endogenous spliceosomes requires tools that allow their detection in their cellular environment, I took advantage of the fact that Prp31 and SF3b155 are specifically phosphorylated in assembled B complexes or activated B^{act} spliceosomal complexes, respectively (Fig. 3.17). The specificity of the P-SF3b155-specific antibody for B^{act} complexes has already been verified *in vitro* and *in vivo* (Girard et al., 2012).
In addition, *in vitro* studies showed that Prp31 is specifically phosphorylated in assembled B spliceosomes (Schneider et al., 2010) and dissociates from the spliceosome during activation (Agafonov et al., 2011).

Results

Therefore, antibodies generated against these two phosphopeptides are excellent markers for assembled B and B[act] spliceosomes *in vivo*.

As a negative control (CTRL), an siRNA without a target in mammalian cells was transfected into HeLa cells. Furthermore, Prp38 and MFAP1 were depleted for comparison. Prp38 has been shown in yeast and human (Xie *et al.*, 1998; Schütze *et al.*, 2016) to be required for spliceosome activation, while Spp381, the yeast homologue of MFAP1 has been proposed to interact with Prp38 and to contribute to its function in splicing (Lybarger *et al.*, 1999; Ulrich & Wahl, 2017). It is likely that MFAP1 also supports the function of Prp38 in humans. Thus, removal of these two proteins served as a positive control for defects in spliceosome activation.

After lysis, the transfected cells were fractionated into a soluble nucleoplasmic and an insoluble chromatin fraction by centrifugation. As splicing occurs predominantly co-transcriptionally, the majority of spliceosomes are associated with the chromatin through the nascent transcript and the RNA polymerase II and are thus associated with the chromatin fraction (Pandya-Jones & Black, 2009; Girard *et al.*, 2012). After solubilisation of the chromatin by sonication, both fractions were separated by SDS-PAGE and transferred to a nitrocellulose membrane for immunoblotting.
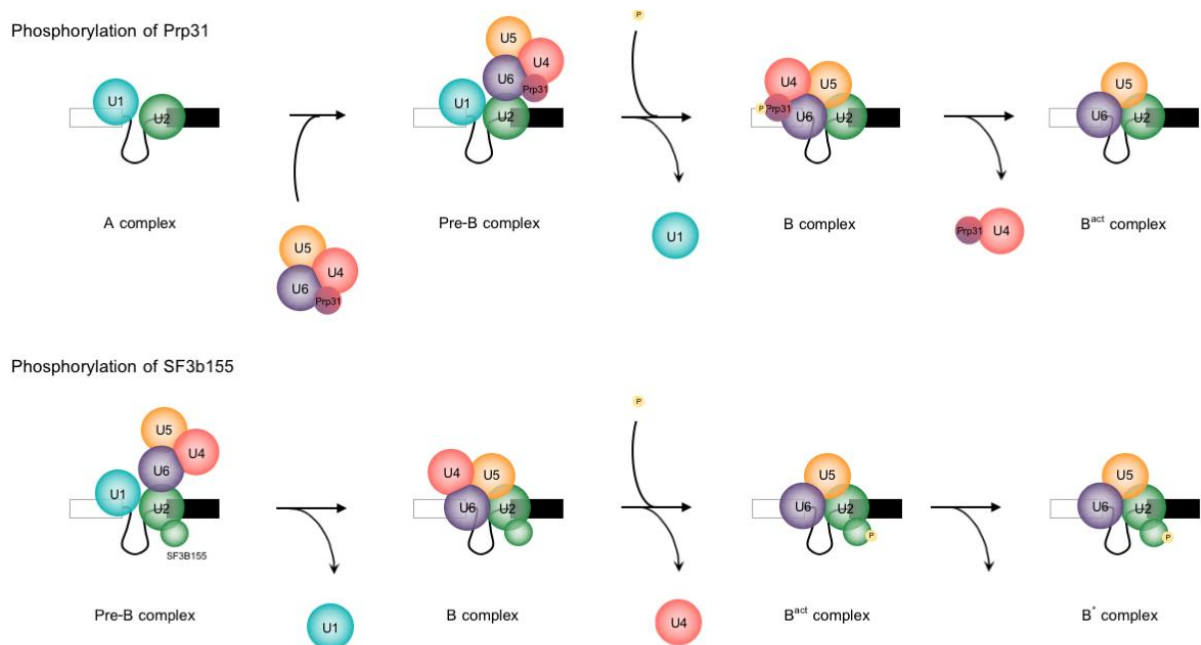
Figure 3.17. Phosphorylation of Prp31 and SF3b155. B complex formation depends on phosphorylation of Prp31, which is subsequently released with U4 during spliceosome activation. Only assembled B complexes harbour the phosphoprotein. SF3b155 phosphorylation occurs during the activation and its dephosphorlyation is essential for the second splicing reaction. Thus, the phosphorylated protein is specifically associated with catalytically activated spliceosomes.
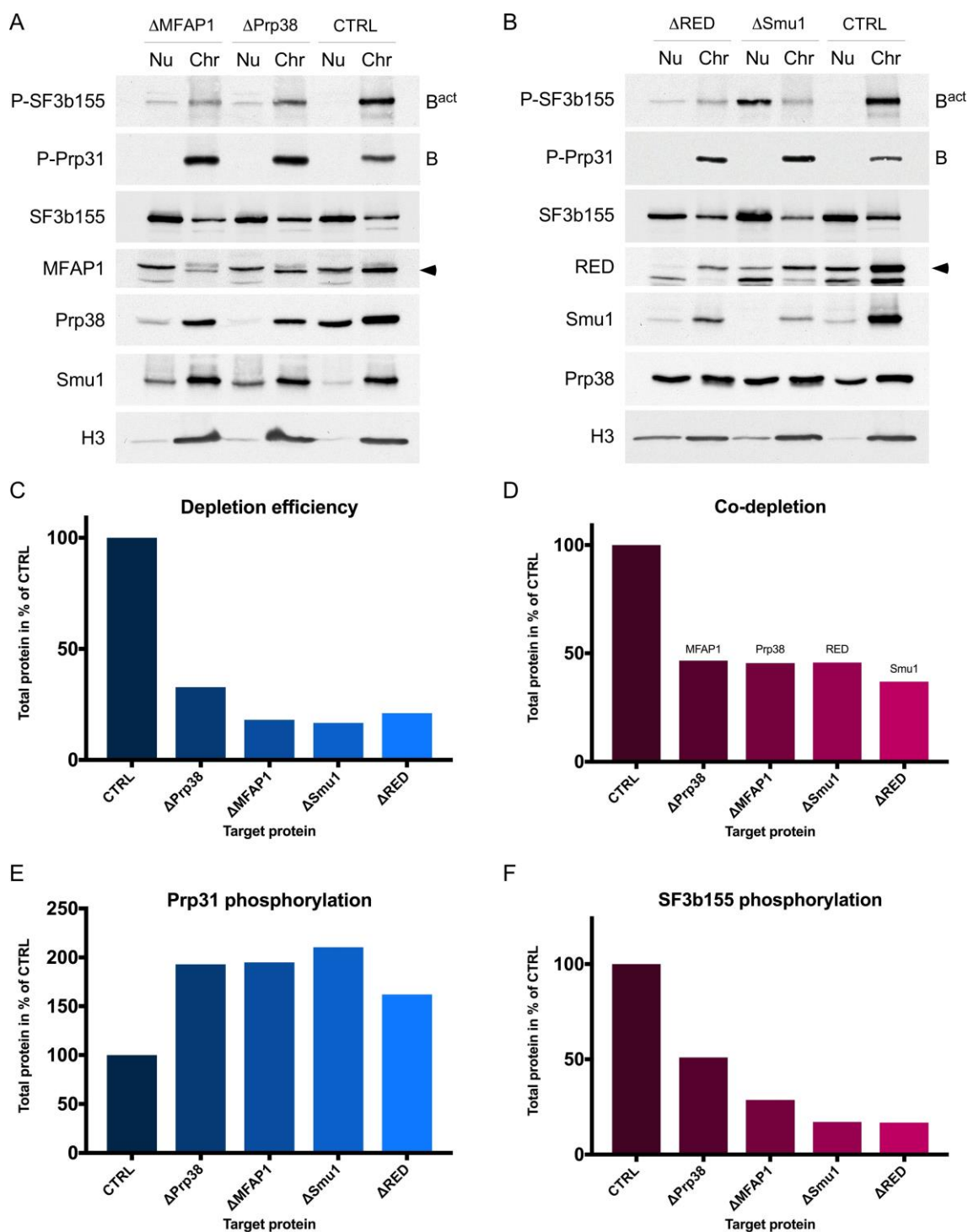
Figure 3.18. siRNA-mediated depletion of Prp38, MFAP1, Smu1 and RED results in altered phosphorylation of Prp31and SF3b155. Western blot analysis of MFAP1- or Prp38-depleted (A) or RED- or Smu1-depleted (B) cells. HeLa cells were lysed and fractionated into a soluble nulceoplasmic (Nu) fraction and an insoluble chromatin (Chr) fraction. Proteins were separated by SDS-PAGE (10–15% step gel) and blotted onto a nitrocellulose membrane. Immunodetection was performed with antibodies specific for the proteins indicated. (C) Quantification of the total protein level after depletion of the respective protein. The protein level was related to its level in CTRL cells. (D) Depletion of MFAP1/Prp38 or RED/Smu1 resulted in co-depletion of the respective partner. The total protein levels of the distinct proteins after depletion of their respective binding partner were quantified and related to their levels in CTRL cells. (E) Quantification of chromatin-associated phosphorylation signals of SF3b155. (F) Quantification of chromatin-associated phosphorylation signals of Prp31.

Results

As shown in Fig. 3.18A, Prp38 and MFAP1 are found in both the nucleoplasmic and chromatin factions under control conditions, with the majority being chromatin associated. siRNA treatment reduced significantly the total protein levels of Prp38 and MFAP1, by 70–80% (Fig. 3.18C) and appeared to result in recruitment of the remaining protein to the chromatin, consistent with involvement in co-transcriptional splicing. Interestingly, depletion of one of these proteins caused a ~55% co-depletion of the other (Fig. 3.18D). The two proteins were shown to interact within the B complex and probably also interact before association with the spliceosome (Bertram, Agafonov, Dybkov *et al.*, 2017).

Prp38 and MFAP1 depletion further resulted in a ~2-fold increase of P-Prp31 signal concomitant with a 50–75% decrease of P-SF3b155 signal (Fig. 3.18E and F). This indicates impaired formation of the B^{act} complex, resulting in accumulation of its precursor, the B complex. As a function in spliceosome activation has been suggested for the two proteins, this result is consistent with previous findings and provides supportive evidence that the phosphospecific antibodies monitor the correct spliceosomal intermediates *in vivo*.

Depletion of Smu1 and RED gave similar results (Fig. 3.18B). Upon depletion total protein levels of Smu1 and RED were reduced by 80–85% (Fig. 3.18C), each concomitant with a 55–65% co-depletion of the other protein (Fig. 3.18D). This is consistent with the previous observation that Smu1 and RED form a heterodimer. (Spartz *et al.*, 2004; Ulrich *et al.*, 2016; Bertram, Agafonov, Dybkov *et al.*, 2017).

Importantly, Prp31 and SF3b155 phosphorylation patterns were similar to those in ΔPrp38 and ΔMFAP1 cells. While the chromatin-associated P-SF3b155 signal was decreased by ~80% in Smu1- and RED- depleted cells, the P-Prp31 signal was increased by 75–100% (Fig. 3.18E and F). However, Prp38 levels were not affected by knock-down of Smu1 or RED. This suggests that Smu1 and RED, similar to several other B-specific proteins, are involved in spliceosome activation also *in vivo*.

To make sure that the increased P-Prp31 signal corresponded to assembled B complexes, I immunopurified these spliceosomes from the chromatin fraction of the transfected cells. The used procedure was adapted from Girard *et al.*, 2012, where endogenous B^{act} complexes were investigated. As shown in Fig. 3.19, this procedure is based on the principles used in chromatin immunoprecipitation (ChIP) and uses the DMS protein-protein crosslinking of the chromatin fraction to keep the spliceosomes intact during the subsequent sonication step. Sonication results in fragmentation of the chromatin, which singularizes the spliceosomal particles. The sheared chromatin fraction was subjected to glycerol-gradient centrifugation and the migration of P-Prp31 (or P-SF3b155) was monitored by western blotting (Data not shown). The RNA was extracted from the peak fractions and analysed by Northern blotting by using a mixture of the radioactively-labelled snRNAs as probes.
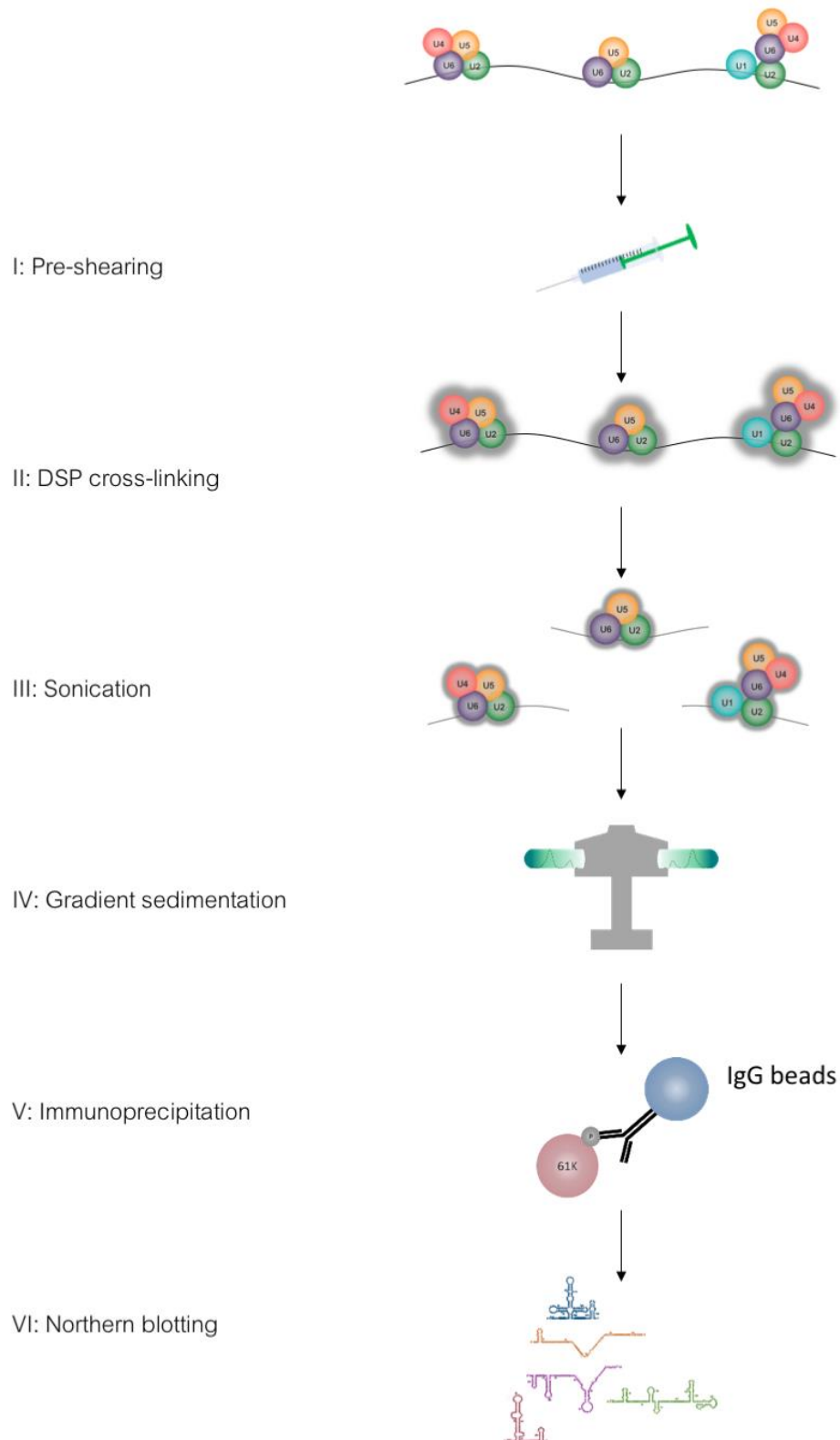
Figure 3.19. Procedure used for the purification of endogenous spliceosomes. The purification procedure included the isolation of the chromatin fraction from HeLa cells followed by rough pre-shearing via passing through a needle in order to disentangle the chromatin. Subsequent DSP crosslinking ensured the maintenance of the spliceosomes integrity during the sonication treatment, which served to separate the spliceosomes through chromatin fragmentation. The spectrum of different complexes was then resolved by glycerol gradient sedimentation. Via immunoprecipitation, the particles of interest were purified from the corresponding fractions and the precipitate's RNA composition was analysed by northern blotting.

Results

As shown in Fig. 3.20A, the P-Prp31-specific antibody purified complexes containing all snRNAs from untreated HeLa cells. In contrast, in spliceosomes purified with the P-SF3b155-specific antibody, U1 and U4 were underrepresented, which is consistent with the snRNA composition of activated spliceosomes. Although U1 is not found in *in vitro* assembled B spliceosomes, it is likely that the P-Prp31-associated complexes are indeed endogenous B complexes and that the presence of U1 is an artefact of the purification procedure.

Investigation of the snRNA composition of P-Prp31-associated spliceosomal particles purified from cells depleted of Prp38, MFAP1, Smu1 or RED revealed the presence of all five snRNAs, as it was the case for the control cells (Fig. 3.20B). This supports the idea that the enriched P-Prp31 signal detected in the chromatin fraction of Prp38-, MFAP1-, Smu1- and RED-depleted cells derives from assembled B spliceosomal complexes and is consistent with my *in vitro* results indicating that Smu1 and RED promote the B-to-B^{act}-transition during spliceosome assembly.
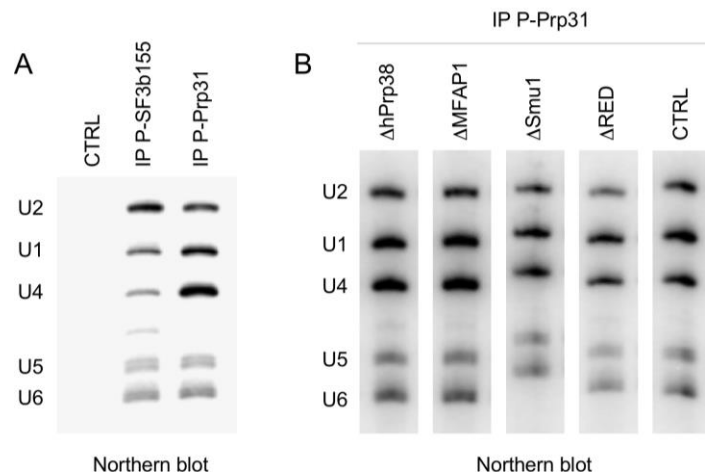
Figure 3.20. RNA composition of endogenous spliceosomes precipitated with anti-P-Prp31 antibodies. (A) Northern blot analysis of the snRNA composition of endogenous spliceosomes. Spliceosomal complexes were purified from HeLa cells by using antibodies specific for the phosphopeptides P-SF3b155 (B^{act}) or P-Prp31 (B). For the control IP (CTRL) the antibody was omitted during the purification. Northern blotting was performed by using probes against the snRNAs, which have been radioactively labelled at the 3' end. (B) Northern blot analysis of the snRNA composition of endogenous spliceosomes stalled by siRNA-mediated depletion of Prp38, MFAP1, Smu1 or RED.

4. Discussion

The spliceosome is a complex molecular machine that catalyses the removal of intronic sequences from a pre-mRNA. While yeast spliceosomes are only comprised of the basic core spliceosomal proteins, spliceosomes in higher eukaryotes are equipped with numerous additional factors, likely to meet the requirements of a much more complex gene architecture and the resulting prevalence of alternative splicing. Two examples of proteins found in human but not yeast spliceosomes are Smu1 and RED. These so-called B-specific proteins transiently associate with the spliceosome at the B complex stage and are released/destabilized during spliceosome activation. While several other members of the B-specific proteins contribute to the activation of spliceosomal B complexes, the role of Smu1 and RED during pre-mRNA splicing is poorly understood. Previous studies have proposed roles for Smu1 and RED as mediators of spliceosomal protein interactions (Hegele *et al.*, 2012), as alternative splicing regulators (Papasaikas *et al.*, 2015), as well as factors that support the stability of the B complex (Bertram, Agafonov, Dybkov *et al.*, 2017).

During my thesis I have investigated the function of Smu1 and RED in pre-mRNA splicing via a combination of genome wide, *in vivo* and *in vitro* approaches. My data revealed that Smu1 and RED are general splicing factors that enhance the B-to-B^{act} transition, and are essential for spliceosome activation when introns exhibit a critical distance between the 5'SS and the BS of ~56 nt. The involvement of these proteins in an essential spliceosome assembly step indicates that Smu1 and RED are not only important for alternative splicing, but also function as general splicing factors that are required for productive and efficient activation of the human spliceosome.

4.1 Efficient splicing is dependent on Smu1 and RED *in vivo* and *in vitro*

Previous studies showed that knock-down of Smu1 or RED in human cells significantly changes alternative splicing, resulting in their classification as alternative splicing regulators (Spartz *et al.*, 2004; Chung *et al.*, 2009; Papasaikas *et al.*, 2015). However, these studies did not address whether also constitutive splicing is affected by depletion of Smu1 and RED. To address this question, HeLa cells were siRNA-depleted of Smu1 or RED and splicing was analysed via RNAseq analysis (Fig. 3.1). Consistent with previous results, depletion of Smu1 or RED changed alternative splicing patterns significantly. Thereby, half of the altered splicing events were alternatively spliced cassette exons - which were preferentially skipped, while 30% involved alternative splice site selection (5'SS or 3'SS). Interestingly, more than 20% of the splicing events that were altered in Smu1 or RED-depleted cells represented retained introns. Importantly, these introns were predominantly constitutively spliced, suggesting that Smu1 and RED are important for splicing in general, and are not only involved in the regulation of alternative splicing. Importantly, cells depleted of Smu1 and RED showed striking overlaps in altered constitutive and alternative splicing events, but many of the altered splicing events were not affected by depletion of the B-specific protein MFAP1. This is consistent with the idea that Smu1 and RED form a functional module that acts independently of the other B-specific proteins.

Discussion

A role for Smu1 and RED in constitutive splicing was also demonstrated *in vitro* for the MINX-120 pre-mRNA substrate, using HeLa nuclear extract that was immunodepleted of Smu1 and RED (Fig. 3.4). The efficiency of splicing was clearly less in the ΔSmu1/RED extract compared to the mock-depleted extract but splicing was not completely inhibited. Instead there appeared to be a slow-down in the rate of mRNA production when Smu1 and RED were missing. Investigation of spliceosome assembly on native agarose gels revealed a transient accumulation of spliceosomal B, but not A complexes, in the Smu1/RED-depleted extract with MINX-120, while the subsequently assembled B^{act} and C complexes were still formed but at a slower rate. These results indicate that defects in splicing triggered by the absence of Smu1 and RED were specifically caused by impaired spliceosome activation, while the association and integration of the tri-snRNP were not affected. The slower/less efficient activation thus appears to be the basis for the slower pre-mRNA-to-mRNA conversion. The effect of the Smu1/RED-depletion on *in vitro* splicing was not substrate-specific as other pre-mRNAs, with a very different architecture such as PM5, were also spliced less efficiently in the absence of Smu1/RED, while B complexes temporarily accumulated (Fig. 3.9).

An effect on the kinetics of spliceosome activation could also explain how, at least in some cases, Smu1 and RED knock-down affect alternative splicing *in vivo*. Similar to the elongation rate of the RNA polymerase II, which is thought to be an important factor for the regulation of alternative splicing (de la Mata *et al.*, 2003; Dujardin *et al.*, 2014), a delay in spliceosome activation might be rate determining during splicing and consequently favour the selection of weaker splice sites. Thus, the changes in alternative splicing observed by RNA-seq analyses in Smu1- and RED-depleted cells may be due to a generally slowed-down spliceosome assembly.

4.2 Evidence that knock-down of Smu1 and RED also hinders spliceosome activation *in vivo*

To provide evidence that Smu1 and RED are also involved in spliceosome activation *in vivo*, I utilised phosphopeptide-specific antibodies against SF3b155 and Prp31 (Fig. 3.17). B complex formation depends on phosphorylation of Prp31 (Schneider, Hsiao *et al.*, 2010), which is subsequently released with U4 during spliceosome activation. Thus, only assembled B complexes harbour the Prp31 phosphoprotein. SF3b155 is hyperphosphorylated first during the activation step and its dephosphorlyation is essential for the second catalytic step of splicing (Shi *et al.*, 2006; Girard *et al.*, 2012). Thus, phosphorylated SF3b155 is specifically associated with catalytically activated spliceosomes. As a positive control to confirm that these antibodies can be used to detect changes in B and B^{act} complex formation *in vivo*, the B-specific proteins Prp38 and MFAP1 were knocked-down in HeLa cells (Fig. 3.18). The depleted cells were fractionated into a nucleoplasmic and a chromatin fraction and analysed by western blotting. Prp38 has been shown in yeast and human (Xie *et al.*, 1998; Schütze *et al.*, 2016) to be involved in spliceosome activation and Spp381, the yeast counterpart of MFAP1, appears to interact with yPrp38 thereby contributing to its function during the activation (Lybarger *et al.*, 1999; Ulrich & Wahl, 2017). Most likely, MFAP1 likewise supports Prp38 function in humans.

Discussion

Indeed, Prp38- and MFAP1-depletion in HeLa cells led to an increased signal of P-Prp31 in the chromatin fraction. As splicing occurs predominantly co-transcriptionally in the cell, active spliceosomes are mainly associated with this fraction (Pandya-Jones & Black, 2009; Girard *et al.*, 2012). Concomitant with the increased P-Prp31 signal, chromatin-associated P-SF3b155 appeared to be reduced. This indicates that knock-down of Prp38 or MFAP1, results in accumulation of B complexes on the chromatin at the expense of B^{act} complex formation. As these changes are consistent with impaired spliceosome activation, this result supports the idea that P-Prp31 and P-SF3b155 are good markers for the formation of B and B^{act} spliceosomes in the cell. As the investigation of spliceosome assembly *in vivo* is challenging, the availability of tools like these phospho-specific antibodies is extremely helpful and can potentially be extended to additional assembly intermediates to allow a better overall picture of spliceosome assembly *in vivo*.

To further demonstrate the specificity of P-Prp31 and P-SF3b155 for assembled B and B^{act} complexes, respectively, I purified the spliceosomal particles that contain the respective phoshopeptide using a method similar to chromatin-immunoprecipitation (Fig. 3.19). This procedure includes protein-protein crosslinking to maintain the integrity of the complexes during the subsequent sonication step, which fragments the nucleic acids resulting in the separation of individual spliceosomal complexes. While the P-Prp31-specific antibody precipitated complexes containing all five snRNAs, U1 and U4 were significantly reduced in spliceosomes containing P-SF3b155 (Fig. 3.20A). Although the formation of a B complex essentially requires the destabilization/release of the U1 snRNP, P-Prp31 containing spliceosomes are nonetheless likely pre-catalytic B complexes and the strong abundance of U1 appears to be an artefact of the purification method. U1 snRNPs might remain loosely associated with the spliceosome after its removal from the 5'SS and be covalently attached due to the cross-linking step. Alternatively, as U1 is the most abundant snRNP in the nucleus of mammalian cells (Baserga & Steitz, 1993) and the first to be recruited to the pre-mRNA, it is possible that U1 is bound to 5' splice sites upstream or downstream of the intron, thereby increasing the apparent abundance of the U1 snRNA in the P-Prp31 containing spliceosomes. Nevertheless, P-Prp31 and P-SF3b155 are clearly associated with assembled spliceosomes, which exhibit snRNA profiles consistent with them being B and B^{act} complexes, respectively.

To examine whether the changes in constitutive splicing in ΔSmu1 and ΔRED cells were likewise likely caused by defective spliceosome activation, I performed RNAi-mediated knockdowns of Smu1 and RED in Hela cells and followed the effects on spliceosome activation using the anti-P-Prp31 and P-SF3b155 antibodies. Interestingly, knockdown of Smu1 or RED led to a strong co-depletion of the respective other protein. This supports the idea that the stability of each protein is improved when they are able to interact with one another (Spartz *et al.*, 2004). It further indicates that the effects observed in previous *in vivo* studies reflect the combined effect of the depletion of both proteins. Investigation of endogenous spliceosome assembly by western blotting revealed an increase in the P-Prp31 signal and a decrease in the P-SF3b155 signal in ΔSmu1 and ΔRED cells, similar to what was observed for Prp38- and MFAP1-depleted cells. This indicates that the activation of the spliceosome is also impaired *in vivo* in the absence of Smu1 and RED.

Discussion

4.3 The majority of Smu1 and RED exist as a unit in HeLa cells

Immunodepletion of Smu1 from splicing-competent nuclear extract, using a well-functioning Smu1-specific antibody, led to the near quantitative removal of Smu1 and additionally more than 90% of RED, indicating that the two proteins physically interact in extract. The Smu1-RED interaction appears to be highly stable, as it resisted salt concentrations up to 1M KCl (data not shown). Smu1 and RED have previously been anticipated to form a functional module based on functional similarities between these two proteins (Papasaikas *et al.*, 2015; Ulrich *et al.*, 2016). This conclusion was also supported by biochemical techniques to identify protein–protein interactions, such as the yeast two-hybrid system (Y2H), which suggested that Smu1 and RED are able to physically interact (Hegele *et al.*, 2012). Binding studies with Smu1 and RED from *Chaenorhabditis elegans*, which are highly homologous to the human Smu1 and RED proteins, also demonstrated that these proteins interact and regions of the proteins minimally required for stable interaction were defined (Ulrich *et al.*, 2016). The highly stable and specific co-depletion of the vast majority of RED with Smu1 provides additional convincing evidence that the vast majority of human Smu1 and RED naturally exist as a unit in HeLa cells (Fig. 3.3). Recent cryo-EM studies of the human B complex revealed that Smu1 and RED interact within the spliceosome (Bertram, Agafonov, Dybkov *et al.*, 2017). The C-terminal WD40 domain and the N-terminal region (NTR) of Smu1 could be localized in the B complex together with a short alpha helix of RED (aa 209–222) that was shown previously to interact with the NTR of Smu1 in *C. elegans* (Ulrich *et al.*, 2016), indicating that Smu1 and RED retain their dimer status even after they are integrated into the spliceosome (Fig. 4.1).
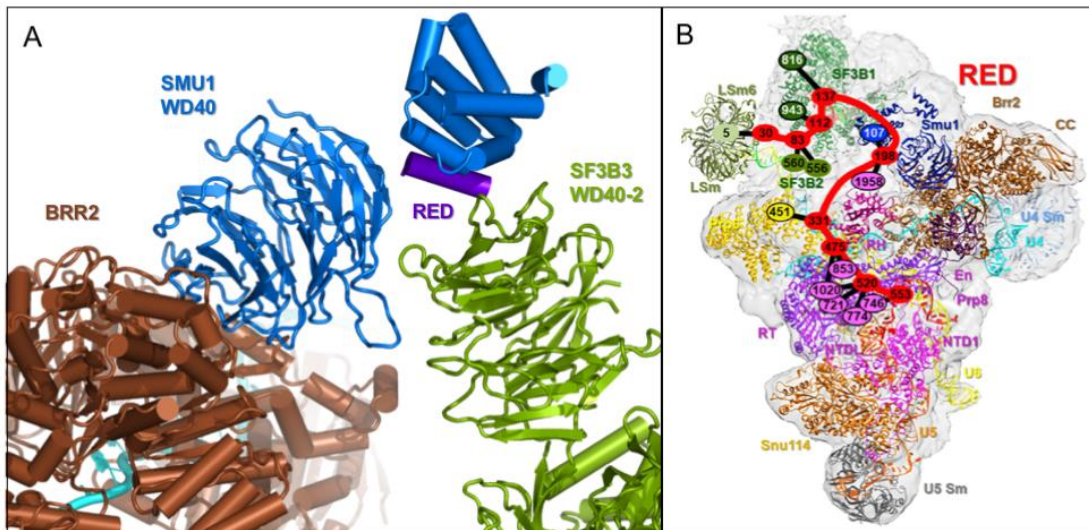
Figure 4.1. Location of Smu1 and RED within the human B complex. (A) Locations of the C-terminal WD40 domain and the N-terminal region (NTR) of Smu1 at the interface between the U5-associated helicase Brr2 and U2 protein SF3B3. The short alpha helix of RED (aa 209–222) that interacts with the NTR of Smu1 is shown in purple. (B) Overview of the likely location of RED based on cross-linking experiments. The anticipated path of RED is indicated as a red line, while the colored ovals with numbers indicate the intermolecular crosslinks between RED and other spliceosomal factors. The illustration was taken from Bertram, Agafonov, Dybkov *et al.*, 2017.

4.4 The interaction between Smu1 and RED is essential for their spliceosomal association and function

Depletion of Smu1 and RED *in vitro* suggested that nearly the entire populations of these proteins exist as a functional unit in the cell. According to the cryo-EM structure of the human B complex, this interaction is maintained after association of Smu1 and RED with the spliceosome, suggesting that it might be important for their mutual function (Bertram, Agafonov, Dybkov *et al.*, 2017). To investigate whether Smu1 and RED need to interact to support the activation of the spliceosome or whether one of these proteins could fulfil this function on their own, Smu1 and RED were individually expressed in insect cells and added to the ΔSmu1/RED extract (Fig. 3.12). Addition of the single proteins did not substantially restore the B-to-B^{act} transition or splicing product formation. However, when the two proteins were added together, B complexes no longer accumulated, and B^{act} and C complexes were formed. This shows that Smu1 and RED do not need to be co-expressed to form a functional module, and indicates that Smu1 and RED do not function on their own.

As Smu1 and RED interact in HeLa nuclear extract before their association with the spliceosome, it is possible that their interaction not only directly affects their function, but may be required for their efficient incorporation into the spliceosome. Consistent with this idea, B complexes formed after addition of Smu1 or RED to ΔSmu1/RED splicing extract contained only low levels of each protein, whereas the association of each protein was strongly enhanced when both proteins were added together (Fig. 3.13). This shows that Smu1 and RED bind poorly or less stably to the spliceosome as individual proteins. Their interaction might introduce important structural changes in Smu1 and RED that enable their proper/stable interaction with other B complex-associated spliceosomal components. Interestingly, the NTR domain of Smu1 and the short helix of RED that are required for dimerization are localized in close proximity to the Smu1-WD40 domain at the SF3B3-Brr2 interface (Fig.4.1A). It is possible that Smu1 requires RED to adopt this conformation, which is a prerequisite for it to simultaneously contact SF3B3 and Brr2, and thus to be stably incorporated into the B complex. Indeed, crystallography studies of Smu1 and RED from *C. elegans*, showed that Smu1 undergoes global conformational changes upon RED binding (Ulrich *et al.*, 2016). Taken together these results show that Smu1 and RED do not bind efficiently or productively to the spliceosome on their own nor function independently of one another.

4.5 Dissection of regions of Smu1 and RED required for their function in splicing

Smu1 contains a C-terminal WD40 domain, which is known to mediate protein-protein interactions. Its N-terminal alpha-helical domain interacts with RED, which is intrinsically disordered, but forms two short helices in the region where it interacts with Smu1. The position of the Smu1-WD40 domain at the U2/tri-snRNP interface in the B complex (Fig. 4.1A) suggests an important role for it in mediating the association of Smu1 and RED with the spliceosome. When full-length RED was co-expressed with a Smu1-WD40 truncation mutant, the dimer could be purified suggesting that the Smu1/RED interaction was not impaired by the absence of the WD40 domain.

Discussion

Addition of the Smu1-ΔWD40/RED dimer to Smu1/RED-depleted extract was not able to restore spliceosome activation, suggesting that this domain is important for the function of the entire dimer (Fig. 3.14) However, analysis of B complexes assembled in the presence of the Smu1-ΔWD40/RED dimer revealed that the dimer was not associated with these complexes, indicating that the WD40 domain is essential for proper interaction of the dimer with the spliceosome. As spliceosome binding was impaired, the role of the WD40 domain in spliceosome activation and splicing catalysis could not be investigated.

To assess the structural requirements of RED for the function of the Smu1/RED dimer, RED was N- or C-terminally truncated, but its region required for stable interaction with Smu1 was retained. Based on crosslinking experiments with purified human B complexes (Bertram, Agafonov, Dybkov *et al.*, 2017), RED's N-terminal region interacts with the U2 proteins SF3B1 (SF3b155) and SF3B3, while its C-terminus contacts Prp8 at several positions (Fig. 4.1B). Consequently, these truncations could potentially reduce the stability of the interaction of RED with the spliceosome. Addition of the ΔN-RED or RED-ΔC mutants together with full-length Smu1 to ΔSmu1/RED extract allowed spliceosome activation, but the latter was strongly reduced when both termini were removed (RED-ΔΔ) (Fig. 3.15). Thus, RED is – to a certain extent – still able to fulfil its function as long as only one interaction surface was removed. This suggests that the contacts RED establishes with U2 or U5 individually are not essential for the function of Smu1/RED, at least not for the splicing of the MINX-80 substrate.

4.6 Smu1 and RED do not indirectly affect activation by acting as a binding platform for other B complex proteins

I could show that Smu1 and RED contribute to efficient activation of the spliceosome and thus the question arises whether these proteins affect the activation directly or indirectly and by what mechanism. Investigation of the protein-protein interaction network in the spliceosome by Y2H experiments suggested that Smu1 and RED directly or indirectly contact numerous spliceosomal proteins (Hegele *et al.*, 2012). These interactions included MFAP1, as well as single components of the U5 snRNP, the U4/U6 di-snRNP, the RES complex, the IBC and the Prp19-related proteins. Protein-protein crosslinking with purified human B complexes also indicated that RED meanders through the spliceosome and contacts multiple spliceosomal proteins (Fig. 4.1B). Therefore, both proteins were proposed to act as binding platforms for additional spliceosomal proteins. Depletion of Smu1 and RED could thus potentially alter the binding of other B complex proteins and this could in turn compromise the efficiency of the activation process. Interestingly, B complexes assembled in Smu1/RED-depleted extract sedimented differently in glycerol-gradients compared to B complexes assembled in wildtype extracts, peaking one or two fractions closer to the top of the gradient. This slower migration can be explained by a lower molecular weight (e.g. if the complex contains fewer proteins) or by a less compact conformation (Fig. 3.8A).

Discussion

With the exception of the absence of Smu1 and RED, the composition of the ΔSmu1/RED B complexes was essentially identical to B complexes assembled in mock-depleted extract. Both contained stoichiometric amounts of U2, U4, U5 and U6, while U1 was nearly absent (Fig. 3.8B). This is consistent with the composition of stable B complexes formed after the action of Prp28, which leads to U1 displacement, but before Brr2-driven unwinding of the U4/U6 duplex.

All known B complex proteins were present including U2 and tri-snRNP proteins, as well as the remaining set of B-specific proteins (Table 3.1). Importantly, factors known to be required for the activation of B complexes, such as Prp38 or FBP21, were clearly present. This demonstrates that their recruitment is independent of Smu1/RED and that the stalling at the B complex level is not caused by the absence of these functionally important proteins.

Some differences in the protein composition of the wildtype versus ΔSmu1/RED B complexes could be observed for proteins known to be recruited during/after activation such as the Prp19/CDC5L complex, the IBC or B[act] and C specific proteins, which were essentially absent in the ΔSmu1/RED B complexes. These proteins are often found to a minor extent in kinetically stalled B complexes but are absent from B complexes stalled with ATPγS or by a low MgCl$_2$ concentration (Agafonov, Van Santen *et al.*, 2016; Bertram, Agafonov, Dybkov *et al.*, 2017). This is likely due to the more homogenous composition of the latter B complex preparations, while kinetically stalled B complex preparations include low levels of activated spliceosomes. Taken together the snRNA and protein compositions of the B complexes stalled by the absence of Smu1/RED are consistent with assembly being stalled after Prp28-promoted release of U1 but before Brr2-catalyzed unwinding of U4 and U6. As no factors known to be required for spliceosome activation were missing, it appears that Smu1 and RED play a direct role in the activation process.

However, the aberrant sedimentation behaviour of the ΔSmu1/RED B complexes suggests that the absence of these proteins leads to structural changes in the spliceosome, in particular a less compact conformation. Thus, it is still conceivable that the absence of Smu1 and RED may still indirectly affect activation by changing the interactions within the spliceosome involving other proteins such as Prp38 or FBP21. Smu1 and RED form a bridge between the U2 SF3B3 protein and Brr2 in the B complex (Fig. 4.1). First structural investigations of ΔSmu1/RED B complexes by negative staining electron microscopy, revealed a structure highly similar to that of DeltaMgCl$_2$ B complexes and no extended gap between U2 and Brr2 was visible in the absence of Smu1/RED. Cautiously speaking, U2 and especially Brr2 appeared slightly fuzzier in B complexes lacking Smu1 and RED, but the low resolution of this method prohibits drawing solid conclusions. Future cryo-EM of ΔSmu1/RED B complexes may provide sufficient resolution to allow insight into any structural changes due to the absence of these proteins.

Discussion

4.7 Smu 1 and RED might directly affect Brr2 helicase activity during the earliest step of activation

Spliceosome activation is a multi-step process, which is initiated by unwinding of the U4/U6 duplex by the RNA helicase Brr2, which is a stable U5 snRNP component. This allows U6 to rearrange and also to interact with U2, which forms the catalytically active U2/U6 RNA structures in the spliceosome. This also leads to loss of U4 and U4/U6 proteins and also the B-specific proteins. Further, the Prp19/CDC5L complex and related proteins, IBC and B[act]-specific proteins are recruited or stably interact. Brr2 is present before activation and also afterwards and its activity thus must be tightly regulated to prevent premature activation. Several mechanisms repress Brr2 before and after activation and its RNA unwinding activity must be triggered in a currently unknown way to start the activation process (Absmeier *et al.*, 2016). Smu1 and RED bridge the U2 SF3B3 protein and Brr2 in the B complex (Fig. 4.1), with Smu1 directly contacting the N-terminal and C-terminal helicase cassettes of Brr2. Thus, these proteins could contribute to positioning Brr2 relative to its substrate in a way that promotes U4/U6 unwinding. Consistent with this idea, Brr2 is very flexible and undergoes a large-scale movement from its position in the tri-snRNP to its position close to U4/U6 in the B complex (Bertram, Agafonov, Dybkov *et al.*, 2017), and Smu1 and RED could help to anchor Brr2 in this position. Alternatively or additionally, Smu1 could help in triggering Brr2 helicase activity by its direct interaction with the helicase cassettes via a mechanism that is currently not clear.

4.8 Smu1 and RED are essential for the activation of spliceosomes formed on very short introns

Although the majority of introns found in human cells are much longer than 200 nt, introns retained after knock-down of Smu1 or RED were predominantly shorter, with most between 75 and 100 nt long (Fig. 3.2). As this preference was not observed after knockdown of MFAP1, Smu1 and RED appear to be specifically required for the splicing of very short introns. Nevertheless, a substantial fraction (30-40%) of the retained introns were significantly longer, consistent with a general requirement for Smu1 and RED for efficient splicing also of longer introns. *In vitro* splicing studies using truncated versions of the MINX-120 pre-mRNA, also demonstrated that splicing is more dependent on the presence of Smu1 and RED when intron length was shorter (Fig. 3.6). That is, I performed splicing in mock-depleted or Smu1/RED immunodepleted HeLa nuclear extract and compared the splicing of the standard MINX pre-mRNA substrate - which contains a 120 nt long intron - with two truncated versions and one elongated variant. In these the intron was shortened or extended between the 5'SS and the BS, changing the overall intron length from 120 nt to 90 or 80 nt, or to 150 nt, respectively. Splicing of MINX-120 and MINX-150 was less efficient in the absence of Smu1/RED compared to the mock-depleted extract, but no increase in splicing efficiency was detected with MINX-150. As extending the intron did not have a significant effect on splicing, 120 nt appears to be above the critical length needed to observe enhanced dependency on Smu1/RED.

Importantly, spliceosome activation was not only delayed but completely blocked when introns were very short. Truncation of the intron to either 90 or 80 nt reduced the overall efficiency of splicing compared to MINX-120, but it also enhanced the inhibitory effect of Smu1/RED depletion. While MINX-90 was spliced somewhat less efficiently than MINX-120, splicing of MINX-80 was nearly abolished in the absence of Smu1 and RED. Analysis of splicing complex formation on MINX-80 showed a strong accumulation of B complexes with little or no B^{act} or catalytically-active C complexes formed when Smu1 and RED were absent. Splicing and spliceosome activation of MINX-80 could be restored to mock levels by addition of the purified dimer to the depleted extract, demonstrating that the observed inhibitory effect was solely due to the absence of Smu1 and RED and not due to the co-depletion of any other essential splicing factors. Thus, both *in vitro* and *in vivo* Smu1 and RED play a crucial role in the splicing of extremely short introns.

4.9 The distance between the 5'SS and the BS is critical for splicing and determines whether Smu1 and RED are essential

In MINX-90, -80 and -70, the distance from the BS-A to the 5'SS was shortened from 96 nt, to 66, 56 and 46 nts, respectively, and there was a progessive loss of splicing activity in the mock-depleted extract, with splicing completely blocked with MINX-70. This is consistent with previous studies showing that a minimum 5'SS-BS distance is critical for splicing (Ruskin *et al.*, 1985; Fu *et al.*, 1988; Köhrer & Domdey, 1988; Smith & Nadal-Ginard, 1989; Himmelspach *et al.*, 1991). With MINX-80, impaired spliceosome activation was seen under standard conditions, but the B-to-B^{act} transition was completely prevented in the absence of Smu1 and RED. Interestingly, spliceosome activation and splicing were also entirely blocked when the 5'SS-BS distance was truncated from 150 to 56 nt in the PM5$_{56}$ and PM5-10$_{56}$ pre-mRNAs (Fig. 3.9). While the intron of the PM5-10$_{56}$ pre-mRNA was with 73 nt shorter than the MINX-80 intron, the PM5$_{56}$ intron was with 142 nt considerably longer. As spliceosome assembly was similarly impaired with both pre-mRNAs, the 5'SS-BS distance rather than the overall intron length appears to be the decisive factor for the Smu1/RED-dependence. Additionally, spliceosome assembly on both PM5$_{56}$ and PM5-10$_{56}$ was equally blocked at B complex level. As they only differ in the length of their PY tracts, the distance between BS and 3'SS does not appear to have an impact on whether or not Smu1 and RED are essential. Taken together these results indicate that the distance between the 5'SS and the BS rather than the BS-3'SS distance or intron length *per se* is the decisive factor for the Smu1/RED-dependency during short intron splicing.

In the human B complex, a 17 nt long extended helix is formed between U6 (i.e. the U6 ACAGAG box and adjacent nucleotides) and nucleotides within the intron near the 5'SS (Fig. 4.2A). Similarly, the BS forms a 14 nt long helical duplex with the U2 snRNA (Fig. 4.2B). (Bertram, Agafonov, Liu *et al.*, 2017; Haselbach *et al.*, 2018; X. Zhang *et al.*, 2018). In B complexes formed on MINX-120 pre-mRNA, these two extended helices are separated by 15 nm (Bertram, Agafonov, Dybkov *et al.*, 2017), which corresponds to ~21 nt of RNA in an extended conformation (Fig. 4.2C).

Discussion

Consequently, to span the distance between the U6 ACAGAG/5'SS helix and the U2/BS helix, without the need for a substantial conformational change in the spliceosome, a minimum of ~52 nt would be necessary between the 5'SS and the BS (Fig. 4.2C). This is approximately the distance found in MINX-80, PM5[56] and PM5-10[56] suggesting that the intron nucleotides in this region are likely in a fully extended conformation in the B complex. This could potentially limit the flexibility of the spliceosome during its remodelling events, thereby impeding or – in the absence of Smu1/RED - completely preventing its activation.
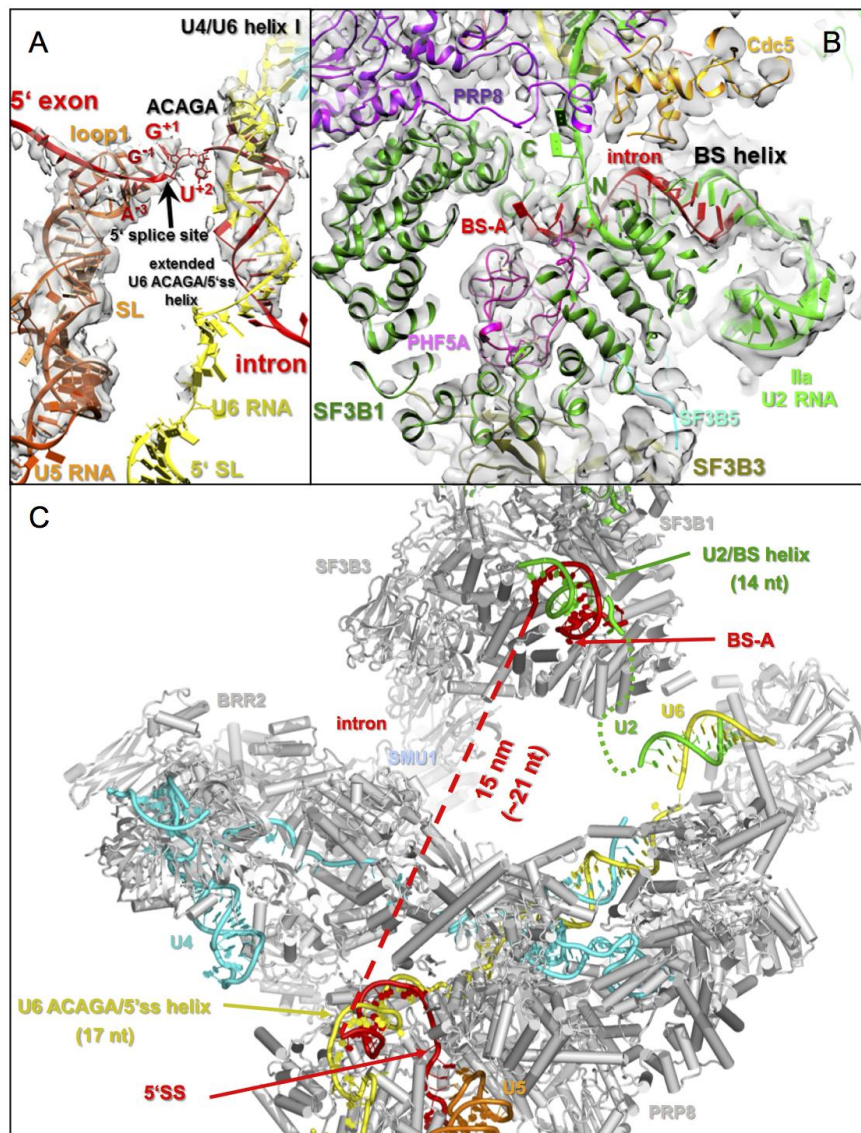
Figure 4.2. Location and interactions of the pre-mRNA within the human B complex. The intron forms extended helices with the U6 ACAGAG box and adjacent U6 nucleotides at the 5'SS (A) and the U2 snRNA at the BS (B). The two helices are separated by 15 nm (~21) in the human B complex. This would accommodate an estimated minimum of 54 nt between the 5'SS and the BS, without altering the structure of the spliceosome. These structures were kindly provided my Berthold Kastner, Department of Cellular Biochemistry, MPI-BPC.

4.10 A short 5'SS-BS distance represents a steric hindrance for the spliceosome

Previously, the requirement for a minimal 5'SS-BS distance was explained by steric hindrances that prevent the binding of the large snRNPs. The dependence of spliceosome activation on a minimal 5'SS-BS distance is probably due to physical limitations exerted by the intron (Fig. 4.2). In this case, resolving this steric hindrance should allow spliceosome activation even without the support of Smu1 and RED. Indeed, when the MINX-80 pre-mRNA was cleaved into two RNAs, spliceosome activation was restored both in the presence or absence of Smu1/RED (Fig 3.10). Specifically, while the intact MINX-80 resulted in a strong accumulation of B complexes, the cleaved substrate allowed the formation of activated spliceosomes and C complexes indicating that the activation was no longer inhibited. Thus, physically separating the 5'SS and the BS to release any length limitations restores the conversion of B to B^{act} complexes.

4.11 Model for how the 5'SS-BS distance affects the dependency of spliceosome activation on Smu1/RED

Taking into consideration my data and utilizing the structural information obtained from the cryo-EM structure of the human B complex, a model for how 5'SS-BS distance might lead to the dependency of spliceosome activation on Smu1/RED was generated (Fig. 4.3). In the human B complex Smu1 and RED form a molecular bridge between the U2 and tri-snRNPs (Bertram, Agafonov, Dybkov *et al.*, 2017). In the yeast spliceosome where Smu1 and RED do not exist, this bridge is formed directly by the Brr2 and SF3B3 homologs (Plaschka *et al.*, 2017). Here, a WD40 domain of Rse1, the yeast counterpart of SF3B3, is positioned between the two cassettes of Brr2, which is replaced by the WD40 domain of Smu1 in the human B complex. This interaction likely stabilizes the position of Brr2 relative to its U4/U6 substrate or might even directly stimulate Brr2 activity thereby promoting activation (see above). Plaschka *et al.* described the Rse1-Brr2 interaction in yeast as transient and suggested that U2 and Brr2 might need to be able to move independently from each other during spliceosome activation. In contrast, my data suggest that the stability of this bridge is important. It is possible that the apparent transience of this bridge in the yeast spliceosome rather represents two distinct states; B complexes before and after establishment of the Rse1-Brr2 interaction. Alternatively, this bridge could be less important in yeast.

In the human spliceosome, Smu1 and RED could ensure that the contact between U2 and Brr2 is established faster and remains stable during activation. As longer introns allow spliceosome activation in the absence of Smu1 and RED, but with reduced efficiency, it appears that a sufficiently long intron grants the U2 head domain enough flexibility to move towards Brr2, coincidentally forming the apparently essential SF3B3-Brr2 bridge, only at a slower pace (Fig. 4.3A). In contrast, an intron with a critical 5'SS-BS distance of ~56 nt would be completely extended in between those nucleotides forming a duplex with U2 or U6, and therefore physically restrict the mobility of U2. This would prevent the movements necessary for U2-SF3B3 to contact Brr2 (Fig. 4.3B) and this would be most severe when Smu1 and RED are absent. In this case, physically separating the 5'SS from the BS would release this

Discussion

physical constraint. Consequently, the flexibility of U2 that allows it to move towards Brr2 would be restored, allowing spliceosome activation even without the support of Smu1 and RED.
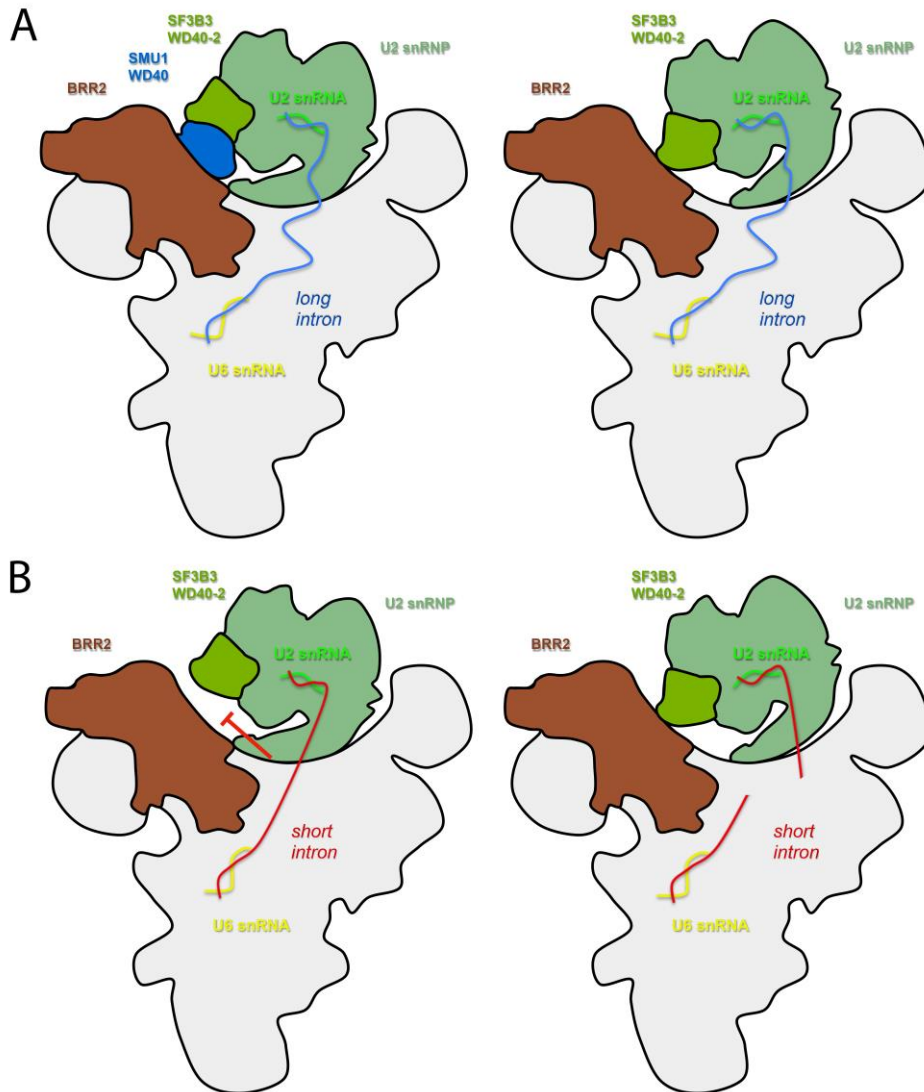
Figure 4.3. Model for how the 5'SS-BS distance affects the dependency of spliceosome activation on Smu1/RED. (A) On long introns Smu1 supports/stabilizes the formation of the Brr2-U2 bridge, which is likely important for spliceosome activation. In the absence of Smu1/RED, this bridge can still be formed, albeit at a slower rate, as long as the intron's distance between the 5'ss and BS is sufficiently long to allow movement of the U2 domain towards Brr2, resulting in a direct Brr2-U2 interaction. (B) On very short introns, with a short distance between the 5'SS and the BS, the U2 snRNP may be repositioned and/or its ability to move towards Brr2 may be inhibited due to the structural constraint of the intron. Therefore, no direct contact between U2 and Brr2 can be established, and Smu1 and RED become essential for activation. Splitting the MINX-80 pre-mRNA into two pieces may relieve this constraint, allowing the B-to-B[act] transition even in the absence of Smu1 and RED.

4.12 The tri-snRNP still binds the MINX-70 pre-mRNA

Previous studies indicated that the binding of snRNPs to the pre-mRNA was abolished when the 5'SS-BS distance was too short (Ruskin *et al.*, 1985; Fu *et al.*, 1988; Köhrer & Domdey, 1988; Smith & Nadal-Ginard, 1989; Himmelspach *et al.*, 1991). For example, binding of U2 was prevented when the distance between the 5'SS and the BS was extremely short - e.g. less than 35 nt in the case of an adenovirus pre-mRNA - and stable tri-snRNP binding (i.e., B complex formation) was abolished when this distance was less than 50 nt (Himmelspach *et al.*, 1991). Recent cryo-EM studies of the human B complex provide a structural basis for this distance dependency. As mentioned above, the structure of a functional B complex formed on MINX-120 would likely not require substantial rearrangement/changes if the 5'SS/BS distance is above ca. 50 nt, but substantial changes, that are likely inhibitory, would occur if this distance is much shorter (Fig. 4.2C).

In contrast to previous studies, my results show that tri-snRNP binding can still occur if this distance is less than 50 nt, and further show that the subsequent remodelling of the spliceosme during activation is affected when the distance between the 5'SS and the BS is short. That is, investigation of the spliceosome assembly on MINX-70 by native gel electrophoresis indicated a strong accumulation of A complexes, but also indicated that B complexes were still substantially formed (Fig. 3.5). A more detailed characterization of the spliceosomes formed on MINX-70 showed that their sedimentation behaviour on glycerol gradients was identical to that of B complexes formed on MINX-120 and MINX-80 pre-mRNA (Fig. 3.16). Their protein compositions were also highly similar, with the presence of B-specific proteins in all cases, consistent with B complexes being assembled on MINX-70 (Table 3.2).

A notable difference was the presence of substantially more U1 snRNA in the MINX-70 B complexes, compared to MINX-120 or MINX-80 B complexes (Fig. 3.16). During spliceosome assembly the tri-snRNP first associates loosely with the A complex to form the pre-B complex (Boesler *et al.*, 2016). In pre-B complexes U1 is stoichiometrically present and base pairs with the 5'SS. Stable tri-snRNP binding is then observed after U1 is displaced from the 5'SS and the U6 ACAGA box sequence base pairs with the 5'SS. This replacement of U1 by U6 is catalysed by the Prp28 (U5-100K) RNA helicase (Staley & Guthrie, 1999). Thus, stable B complexes normally contain very low levels of U1, which probably are not, base-paired with the 5'SS but instead are still associated via protein-protein contacts with the B complex. The enrichment of the U1 snRNA in MINX-70 B complexes is reminiscent of the situation with pre-B complexes. However, it has been shown that the B-specific proteins, which are present in the particles assembled on MINX-70, are not normally recruited before the action of Prp28. It is thus possible that MINX-70 allows the formation of an intermediate spliceosome between the pre-B and B complexes, where an aberrant structure caused by the short 5'SS to BS distance impairs Prp28 action but still allows relatively stable binding of the tri-snRNP. Interestingly, Prp28 (U5-100K) was clearly enriched in the MINX-70 spliceosomes, which suggests that it may not have acted and removed U1 from the 5'SS.

Discussion

Future RNA-RNA crosslinking studies should clarify if U1 still base pairs with the 5'SS in the MINX-70 B-like complexes or if the U6 ACAGAG box duplex is formed. An alternative explanation would be that a different structural organization of the MINX-70 complexes results in stabler U1 binding via protein-protein contacts alone. Future cyro-EM of spliceosomes formed on MINX-70 might reveal changes in their structure due to the very short intron.

4.13 Perspectives

The aim of this thesis was to clarify the role of the non-conserved, B-specific proteins Smu1 and RED. I could show that Smu1 and RED are involved in the Brr2-driven activation of the spliceosome *in vitro* and *in vivo*. In the majority of cases these two proteins facilitate/accelerate the B-to-B^{act} transition, but when the distance between the 5'SS and the BS reaches a critical range of around 56 nt Smu1 and RED are essential. The cryo-EM structure of the human B complex revealed that Smu1, via its WD40 domain, bridges the U2-tri-snRNP interaction by simultaneously binding to SF3B3 and Brr2 (Bertram, Agafonov, Dybkov *et al.*, 2017). Possibly, this interaction, which is formed directly in the yeast spliceosome (Plaschka *et al.*, 2017), positions Brr2 properly relative to its U4/U6 substrate, but it might also directly stimulate the helicase activity of Brr2. This could be tested in the future by using a helicase assay, where the unwinding efficiency of the U4/U6 duplex is monitored in the presence or absence of Smu1 and RED. Similar experiments conducted with Brr2 and fragments of FBP21 indicated that FBP21 decreases the unwinding efficiency of the helicase *in vitro* (Henning *et al.*, 2017) and could thus help to prevent pre-mature activation in spliceosomes primed for activation. For Smu1 and RED this type of experiment would provide information about the mechanism whereby they promote spliceosome activation and clarify whether it is strictly structural/conformational or if the activity of Brr2 is directly affected by the presence of these proteins.

At this moment it is not entirely clear how Smu1 and RED mediate the interaction between SF3B3 and Brr2, especially when the 5'SS-BS distance is short. One possibility is that Smu1 and RED hold SF3B3 and Brr2 close together, and that this interaction is disrupted when the intron or 5'SS-BS distance is short. Alternatively, if Smu1 and RED do not actively counteract the physical hindrance that a short 5'SS-BS distance exerts, Smu1 might simply span the gap between U2 and the tri-snRNP, when U2 is not flexible enough to move towards Brr2. In that case, subtle differences in the structure of the spliceosome might not be observable by negative stain electron microscopy. Therefore, examination of these complexes by cryo-electron microscopy might provide more detailed structural information about the conformational state of B complexes lacking Smu1 and RED, resulting in a better understanding of the function of these proteins. Likewise, cryo-EM may in the future reveal structural differences between human B complexes formed on very short versus long introns.

5. References

Abebrese, E.L., Ali, S.H., Arnold, Z.R., Andrews, V.M., Armstrong, K., Burns, L., Crowder, H.R., Day, R.T., Hsu, D.G., Jarrell, K., Lee, G., Luo, Y., Mugayo, D., Raza, Z., and Friend, K. (2017). Identification of human short introns. *PLoS ONE* 12, 1–19.

Absmeier, E., Santos, K.F., and Wahl, M.C. (2016). Functions and regulation of the Brr2 RNA helicase during splicing. *Cell Cycle* 15, 3362–3377.

Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Lührmann, 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. *EMBO Journal* 18, 5789–5802.

Agafonov, D.E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C.L., Urlaub, H., Luhrmann, R., Odenwälder, P., Bessonov, S., Will, C.L., Urlaub, H., and Lührmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Molecular and Cellular Biology* 31, 2667–82.

Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R. V., Liu, W.T., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* 351.

Agafonov, D.E., Van Santen, M., Kastner, B., Dube, P., Will, C.L., Urlaub, H., and Luhrmann, R. (2016). ATPγS stalls splicing after B complex formation but prior to spliceosome activation. *Rna* 22, 1329–1337.

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. *EMBO Journal* 32, 2804–2818.

Arenas, J.E., and Abelson, J.N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proceedings of the National Academy of Sciences of the United States of America* 94, 11798–11802.

Avery, O.T., Macleod, C.M., McCarty, M., and Peltier, L.F. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from Pheumococcus type III. *Clinical Orthopaedics and Related Research*.

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, 6038–42.

Bai, R., Wan, R., Yan, C., Lei, J., and Shi, Y. (2018). Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. *Science* 0325, 1–11.

Bai, R., Yan, C., Wan, R., Lei, J., and Shi, Y. (2017). Structure of the Post-catalytic Spliceosome from Saccharomyces cerevisiae. *Cell* 171, 1589–1598.e8.

Bao, P., Will, C.L., Urlaub, H., Boon, K.L., and Lührmann, R. (2017). The RES complex is required for efficient transformation of the precatalytic B spliceosome into an activated Bactcomplex. *Genes and Development* 31, 2416–2429.

Bartels, C., Klatt, C., Lührmann, R., and Fabrizio, P. (2002). The ribosomal translocase homologue Snu114p is involved in unwinding U4/U6 RNA during activation of the spliceosome. *EMBO Reports* 3, 875–880.

References

Baserga, S.J., and Steitz, J. a. (1993). The Diverse World of Small Ribonucleoproteins. *The RNA World* 359–381.

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

Behrens, S.E., Tyc, K., Kastner, B., Reichelt, J., and Lührmann, R. (1993). Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. *Molecular and Cellular Biology*.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *The EMBO Journal* 26, 1737–48.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews. Genetics* 15, 163–75.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89, 781–787.

Berk, A.J. (2016). Discovery of RNA splicing and genes in pieces. *Proceedings of the National Academy of Sciences* 113, 801–805.

Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* 170, 701–713.e11.

Bertram, K., Agafonov, D.E., Liu, W.T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* 542, 318–323.

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 (New York, N.Y.)* 16, 2384–403.

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, 846–50.

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes & Development* 2, 754–765.

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, 5803–16.

Black, D.L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry* 72, 291–336.

Black, D.L., and Pinto, A.L. (1989). U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6. *Molecular and Cellular Biology* 9, 3350–9.

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

Boesler, C., Rigo, N., Agafonov, D.E., Kastner, B., Urlaub, H., Will, C.L., and Lühmann, R. (2015). Stable tri-snRNP integration is accompanied by a major structural rearrangement of the spliceosome that is dependent on Prp8 interaction with the 5′ splice site. *Rna* 21, 1993–2005.

Boesler, C., Rigo, N., Anokhina, M.M., Tauchert, M.J., Agafonov, D.E., Kastner, B., Urlaub, H., Ficner, R., Will, C.L., and Lührmann, R. (2016). A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nature Communications* 7.

Bonnal, S., Martínez, C., Förch, P., Bachi, A., Wilm, M., and Valcárcel, J. (2008). RBM5/Luca-15/H37 Regulates Fas Alternative Splice Site Pairing after Exon Definition. *Molecular Cell* 32, 81–95.

Bradley, T., Cook, M.E., and Blanchette, M. (2014). SR proteins control a complex network of RNA-processing events. *RNA (New York, N.Y.)* 21, 75–92.

References

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Lührmann, R. (1984). Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *The EMBO Journal* 3, 1357–63.

Brosi, R., Hauri, H.P., and Kramer, A. (1993). Separation of splicing factor SF3 into two components and purification of SF3a activity. *Journal of Biological Chemistry*.

Brow, D.A., and Guthrie, C. (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334, 213–218.

Caudevilla, C., Serra, D., Miliar, A., Codony, C., Asins, G., Bach, M., and Hegardt, F.G. (1998). Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proceedings of the National Academy of Sciences* 95, 12185–12190.

Chandrasekharappa, S.C., Smith, J.H., and Eliceiri, G.L. (1983). Biosynthesis of Small Nuclear RNAs in Human Cells. *Journal of Cellular Physiology* 117, 169–174.

Chiu, Y.L., Ho, C.K., Saha, N., Schwer, B., Shuman, S., and Rana, T.M. (2002). Tat stimulates cotranscriptional capping of HIV mRNA. *Molecular Cell* 10, 585–597.

Chung, T., Wang, D., Kim, C.-S., Yadegari, R., and Larkins, B.A. (2009). Plant SMU-1 and SMU-2 Homologues Regulate Pre-mRNA Splicing and Multiple Aspects of Development. *Plant Physiology* 151, 1498–1512.

Clawson, G.A., Feldherr, C.M., and Smuckler, E.A. (1985). Nucleocytoplasmic RNA transport. *Molecular and Cellular Biochemistry* 67, 87–100.

Consortium, I.H.G.S. (2004). Finishing the euchromatic sequence of the human genome. *Nature* 50, 162–8.

Coolidge, C.J., Seely, R.J., and Patton, J.G. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Research* 25, 888–896.

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

Cordin, O., and Beggs, J.D. (2013). RNA helicases in splicing. *RNA Biology* 6286, 83–95.

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

David, C.J., Boyne, A.R., Millhouse, S.R., and Manley, J.L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes and Development* 25, 972–982.

de la Mata, M., Alonso, C.R., Fededa, J.P., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. *Malecular Cell* 12, 525–532.

de la Mata, M., and Kornblihtt, A.R. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nature Structural & Molecular Biology* 13, 973–980.

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, 5528–43.

Dignamr, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research* 36, 1475–1489.

Dujardin, G., Lafaille, C., de la Mata, M., Marasco, L.E., Muñoz, M.J., Le Jossic-Corcos, C., Corcos, L., and Kornblihtt, A.R. (2014). How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. *Molecular Cell* 54, 683–690.

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., Oubridge, C., Galej, W.P., Wilkinson, M.E., Bai, X.C., Newman, A.J., and Nagai, K. (2017). Structure of a spliceosome remodelled for exon ligation. *Nature* 542, 377–380.

References

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 catalyzes nuclear pre-mRNA splicing. *Nature* 25, 368–379.

Fischer, U., Liu, Q., and Dreyfuss, G. (1997). The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell* 90, 1023–1029.

Fischer, U., and Lührmann, R. (1990). An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. *Science* 249, 786–790.

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature* 414, 929–933.

Fournier, G., Chiang, C., Munier, S., Tomoiu, A., Demeret, C., Vidalain, P.O., Jacob, Y., and Naffakh, N. (2014). Recruitment of RED-SMU1 Complex by Influenza A Virus RNA Polymerase to Control Ciral mRNA Splicing. *PLoS Pathogens* 10.

Fu, X.Y., Colgan, J.D., and Manley, J.L. (1988). Multiple cis-acting sequence elements are required for efficient splicing of simian virus 40 small-t antigen pre-mRNA. *Molecular and Cellular Biology* 8, 3582–3590.

Furuichi, Y., LaFiandra, A., and Shaktin, A.J. (1977). 5'-Terminal structure and mRNA stability. *Nature* 269, 235–239.

Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. *Nature* 537, 197–201.

Garcia-Blanco, M.A. (2003). Messenger RNA reprogramming by spliceosome-mediated RNA trans-splicing. *Journal of Clinical Investigation* 112, 474–480.

Gee, S., Krauss, S.W., Miller, E., Aoyagi, K., Arenas, J., and Conboy, J.G. (1997). Cloning of mDEAH9, a putative RNA helicase and mammalian homologue of Saccharomyces cerevisiae splicing factor Prp43. *Proceedings of the National Academy of Sciences of the United States of America* 94, 11803–11807.

Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. *Human Genetics* 135, 851–867.

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., Luhrmann, R., and Lührmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nature Communications* 3, 994.

Goldstrohm, A.C., Greenleaf, A.L., and Garcia-Blanco, M.A. (2001). Co-transcriptional splicing of pre-messenger RNAs: Considerations for the mechanism of alternative splicing. *Gene* 277, 31–47.

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 and Development* 10, 233–243.

Gozani, O., Potashkin, J., and Reed, R. (1998). A Potential Role for U2AF-SAP 155 Interactions in Recruiting U2 snRNP to the Branch Site. *Molecular and Cellular Biology* 18, 4752–4760.

Hamm, J., and Mattaj, I.W. (1990). Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* 63, 109–118.

Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. *Biochemical Journal* 430, 379–392.

Hang, J., Wan, R., Yan, C., and Shi, Y. (2015). Structural basis of pre-mRNA splicing. *Science* 349, 1191–1198.

Haselbach, D., Komarov, I., Agafonov, D.E., Hartmuth, K., Graf, B., Dybkov, O., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2018). Structure and Conformational Dynamics of the Human Spliceosomal BactComplex. *Cell* 172, 454–464.e11.

Hashimoto, C., and Steitz, J.A. (1984). U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Research* 12, 3283–3293.

References

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic Protein-Protein Interaction Wiring of the Human Spliceosome. *Molecular Cell* 45, 567–580.

Heinrichs, V., Bach, M., Winkelmann, G., and Luhrmann, R. (1990). U1-specific C protein needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science* 247, 69–72.

Henning, L.M., Santos, K.F., Sticht, J., Jehle, S., Lee, C.T., Wittwer, M., Urlaub, H., Stelzl, U., Wahl, M.C., and Freund, C. (2017). A new role for FBP21 as regulator of Brr2 helicase activity. *Nucleic Acids Research* 45, 7922–7937.

Hernández, H., Makarova, O. V., Makarov, E.M., Morgner, N., Muto, Y., Krummel, D.P., and Robinson, C. V. (2009). Isoforms of U1-70k control subunit dynamics in the human spliceosomal U1 snRNP. *PLoS ONE* 4.

Himmelspach, M., Gattoni, R., Gerst, C., Chebli, K., and Stévenin, J. (1991). Differential block of U small nuclear ribonucleoprotein particle interactions during in vitro splicing of adenovirus E1A transcripts containing abnormally short introns. *Molecular and Cellular Biology* 11, 1258–1269.

Hirose, Y., Tacke, R., and Manley, J.L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes & Development*.

House, A.E., and Lynch, K.W. (2006). An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nature Structural and Molecular Biology* 13, 937–944.

Izquierdo, J.M., Majós, N., Bonnal, S., Martínez, C., Castelo, R., Guigó, R., Bilbao, D., and Valcárcel, J. (2005). Regulation of fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Molecular Cell* 19, 475–484.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3, 318–356.

Keating, K.S., Toor, N., Perlman, P.S., and Pyle, A.M. (2010). A structural analysis of the group II intron active site and implications for the spliceosome. *RNA (New York, N.Y.)* 1–9.

Kent, O.A., and MacMillan, A.M. (2002). Early organization of pre-mRNA during spliceosome assembly. *Nature Structural Biology* 9, 576–581.

Köhrer, K., and Domdey, H. (1988). Splicing and spliceosome formation of the yeast MATaI transcript require a miniimum distance from the 5' splice site to the internal branch acceptor site. *Nucleic Acids Research* 16, 9457–9475.

Konarska, M.M., Padgett, R.A., and Sharp, P.A. (1985). Trans-splicing of mRNA precursors in vitro . *Cell* 42, 165–171.

Krainer, A.R., Maniatis, T., Ruskin, B., and Green, M.R. (1984). Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell* 36, 993–1005.

Kunkel, G.R., Maser, R.L., Calvet, J.P., and Pederson, T. (1986). U6 small nuclear RNA is transcribed by RNA polymerase III. *Proceedings of the National Academy of Sciences of the United States of America* 83, 8575–8579.

Laemmli, U.K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 227, 680.

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

Lam, B.J., and Hertel, K.J. (2002). A general role for splicing enhancers in exon definition. *RNA (New York, N.Y.)* 8, 1233–1241.

Lambowitz, A.M., and Zimmerly, S. (2011). Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harbor Perspectives in Biology* 3, 1–19.

References

Lamond, A.I., Konarska, M.M., and Sharp, P.A. (1987). A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. *Genes & Development* 1, 532–543.

Lander, E.S., Linton, L.M., and Birren, B. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Lardelli, R.M., Thompson, J.X., Yates, J.R., and Stevens, S.W. (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA (New York, N.Y.)*.

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

Le Hir, H., and Séraphin, B. (2008). EJCs at the Heart of Translational Control. *Cell* 133, 213–216.

Lerner, M.R., and Steitz, J.A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proceedings of the National Academy of Sciences of the United States of America* 76, 5495–5499.

Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S.L., and Quackenbush, J. (2000). Gene index analysis of the human genome estimates approximately 120,000 genes. *Nature Genetics* 25, 239–240.

Lim, S.R., and Hertel, K.J. (2004). Commitment to splice site pairing coincides with a complex formation. *Molecular Cell* 15, 477–483.

Liu, S., Li, X., Zhang, L., Jiang, J., Hill, R.C., Cui, Y., Hansen, K.C., Zhou, Z.H., and Zhao, R. (2017). Structure of the yeast spliceosomal postcatalytic P complex. *Science* 358, 1278–1283.

Lopez, P.J., and Séraphin, B. (1999). Genomic-scale quantitative analysis of yeast pre-mRNA splicing : implications for splice-site recognition. *RNA (New York, N.Y.)* 5, 1135–1137.

Lybarger, S., Beickman, K., Brown, V., Dembla-Rajpal, N., Morey, K., Seipelt, R., and Rymond, B.C. (1999). Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. *Molecular and Cellular Biology* 19, 577–584.

MacMillan, A.M., Query, C.C., Allerson, C.R., Chen, S., Verdine, G.L., and Sharp, P.A. (1994). Dynamic association of proteins with the pre-mRNA branch region. *Genes and Development* 8, 3008–3020.

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.

Maeder, C., Kutach, A.K., and Guthrie, C. (2008). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nature Structural and Molecular Biology* 16, 42–48.

Makarov, E.M., Makarova, O. V., Achsel, T., and Lührmann, R. (2000). The human homologue of the yeast splicing factor Prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions. *Journal of Molecular Biology* 298, 567–575.

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. *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. *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. *EMBO Journal* 23, 2381–2391.

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

Mayas, R.M., Maita, H., and Staley, J.P. (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nature Structural and Molecular Biology* 13, 482–490.

References

Merendino, L., Guth, S., Bilbao, D., Concepcion, M., and Juan, V. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* 402, 838–841.

Millhouse, S., and Manley, J.L. (2005). The C-Terminal Domain of RNA Polymerase II Functions as a Phosphorylation-Dependent Splicing Activator in a Heterologous Protein. *Molecular and Cellular Biology* 25, 533–544.

Misteli, T., and Spector, D.L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Molecular Cell* 3, 697–705.

Moore, M.J., and Sharp, P.A. (1993). Evidence for 2 active sites in the spliceosome provided by stereochemistry of pre-messenger RNA splicing. *Nature* 365, 364–368.

Morris, D.P., and Greenleaf, A.L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA Polymerase II. *Journal of Biological Chemistry* 275, 39935–39943.

Mozaffari-Jovin, S., Santos, K.F., Hsiao, H.H., Will, C.L., Urlaub, H., Wahl, M.C., and Lührmann, R. (2012). The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes and Development* 26, 2422–2434.

Nagaike, T., Logan, C., Hotta, I., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. (2011). Transcriptional Activators Enhance Polyadenylation of mRNA Precursors. *Molecular Cell* 41, 409–418.

Nesic, D., Tabackovic, G., and Krämer, A. (2004). A role for Cajal bodies in the final steps of U2 snRNP biogenesis. *Journal of Cell Science* 117, 4423–4433.

Neugebauer, K.M. (2002). On the importance of being co-transcriptional. *Journal of Cell Science* 115, 3865–3871.

Newman, A. (2008). RNA Interactions in mRNA Splicing. *Encyclopedia of Life Sciences* 1–7.

Nguyen, T.H.D., Galej, W.P., Bai, X., Oubridge, C., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2016). CryoEM structure of the yeast U4 / U6 . U5 tri-snRNP at 3 . 7 Å resolution. *Nature* 530, 298–302.

Nguyen, T.H.D., Galej, W.P., Bai, X.C., Savva, C.G., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* 523, 47–52.

Ohrt, T., Odenwälder, P., Dannenberg, J., Prior, M., Warkocki, Z., Schmitzová, J., Karaduman, R., Gregor, I., Enderlein, J., Fabrizio, P., and Lührmann, R. (2013). Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA (New York, N.Y.)* 19, 1–14.

Ohrt, T., Prior, M., Dannenberg, J., Odenwälder, P., Dybkov, O., Rasche, N., Schmitzová, J., Gregor, I., Fabrizio, P., Enderlein, J., and Lührmann, R. (2012). Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA (New York, N.Y.)* 18, 1244–1256.

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

Pandya-Jones, A.M.Y., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. *RNA (New York, N.Y.)* 15, 1896–1908.

Papasaikas, P., Tejedor, J.R., Vigevani, L., and Valcárcel, J. (2015). Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Molecular Cell* 57, 7–22.

Plaschka, C., Lin, P.C., and Nagai, K. (2017). Structure of a pre-catalytic spliceosome. *Nature* 546, 617–621.

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 and 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* 8, 847–855.

Raker, V.A., Plessel, G., and Lührmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *The EMBO Journal* 15, 2256–69.

Ram, O., and Ast, G. (2006). SR proteins: a foot on the exon before the transition from intron to exon definition. *Trends in Genetics* 23, 5–7.

References

Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C., and Urlaub, H. (2016). Molecular architechture of the Saccharomyces cerevisiae activated spliceosome. *Science (New York, N.Y.)* 353, 1399–1405.

Reyes, J.L., Kois, P., Konforti, B.B., and Konarska, M.M. (1996). The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA (New York, N.Y.).*

Robberson, B.L., Cote, G.J., and Berget, S.M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular and Cellular Biology* 10, 84–94.

Rosonina, E., Bakowski, M.A., McCracken, S., and Blencowe, B.J. (2003). Transcriptional Activators Control Splicing and 3'-End Cleavage Levels. *Journal of Biological Chemistry* 278, 43034–43040.

Ruskin, B., Greene, J.M., and Green, M.R. (1985). Cryptic branch point activation allows accurate in vitro splicing of human β-globin intron mutants. *Cell* 41, 833–844.

Ruskin, B., Zamore, P.D., and Green, M.R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* 52, 207–219.

Sakharkar, M.K., Chow, V.T.K., and Kangueane, P. (2004). Distributions of exons and introns in the human genome. *In Silico Biology* 4, 387–93.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2nd ed. *Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.*

Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T., and L??hrmann, R. (2004). RNAi knockdown of hPrp31 to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO Journal* 23, 3000–3009.

Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Luhrmann, R., Glover, J.N.M., and MacMillan, A.M. (2006). Crystal structure of a core spliceosomal protein interface. *Proceedings of the National Academy of Sciences* 103, 1266–1271.

Schneider, M., Hsiao, H.H., Will, C.L., Giet, R., Urlaub, H., and Lührmann, R. (2010). Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. *Nature Structural and Molecular Biology* 17, 216–221.

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.

Schütze, T., Ulrich, A.K.C., Apelt, L., Will, C.L., Bartlick, N., Seeger, M., Weber, G., Lührmann, R., Stelzl, U., and Wahl, M.C. (2016). Multiple protein-protein interactions converging on the Prp38 protein during activation of the human spliceosome. *RNA (New York, N.Y.)* 22, 265–277.

Schwer, B. (2001). A new twist on RNA helicases : DExH / D box proteins as RNPases. *Nature* 8, 113–116.

Schwer, B. (2008). A Conformational Rearrangement in the Spliceosome Sets the Stage for Prp22-Dependent mRNA Release. *Molecular Cell* 30, 743–754.

Schwer, B., and Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* 349, 494–499.

Scotti, M.M., and Swanson, M.S. (2016). RNA mis-splicing in disease. *Nature Reviews Genetics* 17, 19–32.

Seetharaman, M., Eldho, N. V., Padgett, R.A., and Dayie, K.T. (2006). Structure of a self-splicing group II intron catalytic effector domain 5: Parallels with spliceosomal U6 RNA. *RNA (New York, N.Y.)* 12, 235–247.

Semlow, D.R., and Staley, J.P. (2012). Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem Sciences* 37, 263–273.

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 and Molecular Biology* 15, 183–191.

Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. *Cell* 9, 645–653.

References

Shatkin, A.J., and Manley, J.L. (2000). The ends of the affair: Capping and polyadenylation. *Nature Structural Biology* 7, 838–842.

Shepard, P.J., and Hertel, K.J. (2009). The SR protein family. *Genome Biology* 10, 242.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical Chemistry* 68, 850–858.

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, 819–29.

Shumyatsky, G., Wright, D., and Reddy, R. (1993). Methylphosphate cap structure increases the stability of 7SK, B2 and U6 small RNAs in Xenopus oocytes. *Nucleic Acids Research* 21, 4756–4761.

Sidarovich, A., Will, C.L., Anokhina, M.M., Ceballos, J., Sievers, S., Agafonov, D.E., Samatov, T., Bao, P., Kastner, B., Urlaub, H., Waldmann, H., and Lührmann, R. (2017). Identification of a small molecule inhibitor that stalls splicing at an early step of spliceosome activation. *ELife* 6, 1–32.

Sloan, K.E., and Bohnsack, M.T. (2018). Unravelling the Mechanisms of RNA Helicase Regulation. *Trends in Biochemical Sciences* 43, 237–250.

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.J., and Nadal-Ginard, B. (1989). Mutually exclusive splicing of α-tropomyosin exons enforced by an unusual lariat branch point location: Implications for constitutive splicing. *Cell* 56, 749–758.

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.

Spartz, A.K., Herman, R.K., and Shaw, J.E. (2004). SMU-2 and SMU-1, Caenorhabditis elegans Homologs of Mammalian Spliceosome-Associated Proteins RED and fSAP57, Work Together To Affect Splice Site Choice. *Molecular and Cellular Biology* 24, 6811–6823.

Spiller, M.P., Boon, K.L., Reijns, M.A.M., and Beggs, J.D. (2007). The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. *Nucleic Acids Research* 35, 923–929.

Spingola, M., Grate, L., Haussler, D., and Manuel, A. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA (New York, N.Y.)* 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.

Sterner, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. *Proceedings of the National Academy of Sciences of the United States of America* 93, 15081–15085.

Sun, J.S., and Manley, J.L. (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes and Development* 9, 843–854.

Teigelkamp, S., Newman, a J., and Beggs, J.D. (1995). Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO Journal* 14, 2602–2612.

Ulfendahl, P.J., Pettersson, U., and Akusjarvi, G. (1985). Splicing of the adenovirus-2 EIA 13S mRNA requires a minimal intron length and specific intron signals. *Nucleic Acids Research* 36, 6299–6315.

Ulrich, A.K.C., Schulz, J.F.F., Kamprad, A., Schütze, T., and Wahl, M.C.C. (2016). Structural Basis for the Functional Coupling of the Alternative Splicing Factors Smu1 and RED. *Structure* 24, 762–773.

Ulrich, A.K.C., and Wahl, M.C. (2017). Human MFAP1 is a cryptic ortholog of the Saccharomyces cerevisiae Spp381 splicing factor. *BMC Evolutionary Biology* 17.

Umen, J.G., and Guthrie, C. (1995). A Novel Role for a U5 snRNP Protein in 3'-Splice-Site Selection. *Genes & Development* 9, 855–868.

References

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718.

Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016). Structure ofa yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* 353, 895–904.

Wan, R., Yan, C., Bai, R., Lei, J., and Shi, Y. (2017). Structure of an Intron Lariat Spliceosome from Saccharomyces cerevisiae. *Cell* 171, 120–132.e12.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.

Wang, Y., Liu, J., Huang, B., Xu, Y., Li, J., Huang, L., Lin, J., Zhang, J., Min, Q., Yang, W., and Wang, X. (2015). Mechanism of alternative splicing and its regulation (Review). *Biomedical Reports* 3, 152–158.

Watson, J., and Crick, F. (1953). Molecular structure of nucleic acids. *Nature.* 171, 737–738.

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.

Wersig, C., Guddat, U., Pieler, T., and Bindereif, A. (1992). Assembly and nuclear transport of the U4 and U4 U6 snRNPs. *Experimental Cell Research* 199, 373–377.

Wieringa, B., Hofer, E., and Weissmann, C. (1984). A minimal intron length but no specific internal sequence is required for splicing the large rabbit β-globin intron. *Cell* 37, 915–925.

Wilkinson, M.E., Fica, S.M., Galej, W.P., Norman, C.M., Newman, A.J., and Nagai, K. (2017). Postcatalytic spliceosome structure reveals mechanism of 3 '-splice site selection. *Science* 358, 1283–1287.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology* 3, 1–23.

Will, C.L., Schneider, C., Macmillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Lührmann, R., and Query, C.C. (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO Journal* 20, 4536–4546.

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. *EMBO Journal* 21, 4978–4988.

Woppmann, A., Will, C.L., Kornstädt, U., Zuo, P., Manley, J.L., and Lührmann, R. (1993). Identification of an snrnp-associated kinase activity that phosphorylates arginine/serine rich domains typical of splicing factors. *Nucleic Acids Research* 21, 2815–2822.

Wu, J., and Manley, J.L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes & Development* 3, 1553–1561.

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

Xiao, S.H., and Manley, J.L. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes & Development* 11, 334–344.

Xie, J., Beickman, K., Otte, E., and Rymond, B.C. (1998). Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. *The EMBO Journal* 17, 2938–2946.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2016). Structure ofa yeast activated spliceosome at 3.5 Å resolution. *Science* 353, 904–911.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2017). Structure of a yeast step II catalytically activated spliceosome. *Science* 355, 149–155.

Yura, K., Shionyu, M., Hagino, K., Hijikata, A., Hirashima, Y., Nakahara, T., Eguchi, T., Shinoda, K., Yamaguchi, A., Takahashi, K.-I., Itoh, T., Imanishi, T., Gojobori, T., and Go, M. (2006). Alternative splicing in human transcriptome: functional and structural influence on proteins. *Gene* 380, 63–71.

Zhan, X., Yan, C., Zhang, X., Lei, J., and Shi, Y. (2018). Structure of a human catalytic step I spliceosome. *Science*

359, 537–545.

Zhang, D., and Rosbash, M. (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes and Development* 13, 581–592.

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 and Development* 15, 30–35.

Zhang, M.Q. (1998). Statistical features of human exons and their flanking regions. *Human Molecular Genetics* 7, 919–932.

Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., and Shi, Y. (2018). Structure of the human activated spliceosome in three conformational states. *Cell Research* 28, 307–322.

Zhuang, Y., and Weiner, A.M. (1986). Compensatory Base Change in Ul snRNA suppresses a 5 ' Splice Site Mutation. *Cell* 46, 827–835.

Zhuang, Y., and Weiner, A.M. (1989). A compensatory base change in human U2 snRNP can supress a branch site mutation. *Genes and Development* 3, 1545–1552.

Zillmann, M., Zapp, M.L., and Berget, S.M. (1988). Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Molecular and Cellular Biology* 8, 814–21.

Appendix

6. Appendix

6.1 Abbreviations

2'O-me	2'O-ribose methylated
3'SS	3' splice site
5'SS	5' splice site
Δ	Depletion / mutant
Å	Ångström unit
A	Adenosine / ampere
aa	Amino acid
Alt3'SS	Alternative 3'SS
Alt5'SS	Alternative 5'SS
Amp	Ampicillin
APS	Ammonium peroxodisulfate
ATP	Adenosine triphosphate
B1-3	Bridge 1-3
BCA	Bicinchoninic acid
bp	Base pair
BP	Branch point
BPS	Branch point sequence
BS	Branch site
BSA	Bovine serum albumin
°C	Degree celsius
C	Cytosine / Carboxy
C. elegans	*Caenorhabditis elegans*
Cex	Cassette exon
Ci	Curie
cm	Centimeter
cpm	Counts per minute
CTD	C-terminal domain
CTP	Cytosine triphosphate
CTRL	Control
d	desoxy
D5	Domain V of group II introns
dd	Double distilled/ didesoxy
Da	Dalton

DExD/H	Consensus sequence of helicases
DMP	dimethyl pimelimidate
DMSO	Dimethyl sulfoxide
DNA	Desoxyribonucleic acid
DSP	dithiobis(succinimidyl propionate)
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
E. coli	*Escherichia coli*
EDTA	Ethylendiamintetraacetate
EJC	Exon junction complex
EM	Electron microscopy
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
et al.	*Et alii*
f	Femto
FBS	Fetal bovine serum
FL	Full length
Fw	forward
G	Guanosine
g	Gram / centrifugal force
GTP	Guanosine triphosphate
h	Hour
h. sapiens	*Homo sapiens*
H_2O	Water
HCl	Hydrochloric acid
HeLa	Henrietta Lacks
HEPES	N-2-Hydroxyethylpiperazin-N-2-ethansulfonic acid
hn	Heterogenous nuclear
IBC	Intron binding complex
ILS	Intron lariat spliceosome
IPTG	Isopropyl-β-D-thiogalactopyranosid
ISE	Intronic splicing enhancer
ISL	Internal-stem loop
ISS	Intronic splicing silencer
J	Joule
k	Kilo
kb	Kilo bases
KCL	Potassium chloride

Appendix

kDa	Kilodalton
L	Liter
LB	Luria Bertani
LC	Liquid chromatography
Lsm	Like-Sm
m_3G	2,2,7-trimethylguanosine
m^7G	N7-methylguanosine
M	Molar
m	Milli / meter
μ	Micro
MBP	Maltose-binding protein
min	Minutes
$MgCl_2$	Magnesium chloride
mM	Millimolar
mRNA	Messenger RNA
MS	Mass spectrometry
MW	Molecular weight
N	Amino
n	Nano
NTD	N-terminal domain
NTP	Nucleoside triphosphate
nt	Nucleotides
OD	Optical density
-OH	Hydroxyl group
PI	Pre-mRNA I
PII	Pre-mRNA II
P	Phosphate
p	Pico
PAGE	Polyacrylamide gel-electrophoresis
PCI	Phenol-chloroform-isoamyl alcohol
PCR	Polymerase chain reaction
pH	Preponderance of hydrogen ions
PMSF	Phenylmethylsulfonylfluoride
PY	Polypyrimidine
Pre-mRNA	Precursor-mRNA
%	Percent
R	Purine base
RES	Retention and splicing

Rev	reverse
RGG	Arginine glycine glycine repeats
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
RNP	Ribonucleoprotein
rpm	Revolutions per minute
RRM	RNA recognition motif
RT	Room temperature/ Reverse transcription
S	Svedberg unit
s	Second
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Saccharomyces pombe*
SDS	Sodium dodecylsulfate
SE	Splicing extract
siRNA	Small interfering RNA
Sm	'Smith', patient in which Sm proteins were first discovered
sn	Small nuclear
snRNA	Small nuclear ribonucleic acid
snRNP	Small nuclear ribonucleoparticles
SR	Serine arginine rich
T	Thymin
TEMED	N, N, N', N'-Tetramethylethylendiamine
Tris	Tris-(hydroxymethyl)-aminomethane
tRNA	Transfer RNA
U	Uridine / unit
U snRNA	Uridine rich small nuclear RNA
U snRNP	Uridine rich small nuclear ribonucleoparticles
UTP	Uridine triphosphate
UV	Ultraviolet
V	Volt
Vol	Volume
v/v	Volume per volume
W	Watts
w/v	Weight per volume
y	yeast
Y	Pyrimidine base

Appendix

6.2 List of figures

Appendix

6.3 List of tables

6.4 Acknowledgements

Appendix