Investigation of the 3D structure of the human activated spliceosome by cryo-electron microscopy

Dissertation

for the award of the degree

„Doctor rerum naturalium“

of the Georg-August-Universität Göttingen

submitted by

Ilya Komarov

from Volgograd, Russia

Göttingen 2017

Members of the Examination Board:

Prof. Dr. Reinhard Lührmann (1st Referee)

Department of Cellular Biochemistry

Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Ralf Ficner (2nd Referee)

Department for Molecular Structural Biology

Georg-August-Universität, Göttingen

Prof. Dr. Claudia Höbartner

Research Group at the Institute for Organic and Biomolecular Chemistry

Georg-August-Universität, Göttingen

Prof. Dr. Holger Stark

Department of Structural Dynamics

Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Henning Urlaub

Research Group Bioanalytical Mass Spectrometry

Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Wolfgang Wintermeyer

Research Group Ribosome Dynamics

Max Planck Institute for Biophysical Chemistry, Göttingen

Date of oral examination: September, 15th

Table of contents

Abstract

The spliceosome is a highly dynamic megadalton ribonucleoprotein (RNP) complex that catalyses the removal of introns from eukaryotic precursor messenger RNA (pre-mRNA) in two consecutive transesterification reactions. The spliceosome is assembled *de novo* on each pre-mRNA intron by the sequential recruitment of five RNA-protein complexes (snRNPs) and numerous non-snRNP factors in a dynamic manner, driven by numerous DExH/D box ATPases or RNA helicases. After the U1 and U2 snRNPs bind to the 5' splice site (ss) and branch site (BS) of the pre-mRNA, respectively, the pre-formed U4/U6.U5 tri-snRNP is recruited to give spliceosomal complex B, which lacks an active site. Catalytic activation of the B complex occurs in a stepwise manner. Initially, Brr2 RNA helicase dissociates the U4/U6 snRNA duplex, allowing U6 snRNA to restructure and to form an intricate U2/U6 RNA network, which is at the heart of the catalytic centre of the spliceosome. During activation, all of the U4/U6 and several U5 proteins are dissociated, while more than 20 new proteins including the Prp19 complex proteins (NTC) are stably integrated into the spliceosome, yielding the B^{act} complex, which is still pre-catalytic. Catalytic activation requires the ATP-dependent action of the RNA helicase Prp2, which displaces several proteins and remodels the U2 SF3A and SF3B complex proteins, yielding the B* complex. B* then catalyses the first step of the splicing reaction, generating the C complex. Following an additional restructuring step, the resulting C* complex catalyses the 2^{nd} step of the splicing reaction to form the mRNA product.

During the last two years, high-resolution electron cryo-microscopy (cryo-EM) structures have been published for several assembly intermediates of the yeast spliceosome including the B, B^{act}, C, C* and intron-lariat complexes, providing completely new insight into the complex structure of the yeast spliceosome and its structural dynamics during the catalytic cycle. As of now, only the human C* complex has been investigated by cryo-EM. In this work, I have used cryo-EM to investigate the 3D structure of the human B^{act} complex. Human and yeast activated spliceosomes share a large number of conserved proteins but differ in their protein composition in several aspects. Human B^{act} contains numerous proteins that are absent in yeast, including numerous peptidyl-prolyl isomerases (PPIases) and the RNA helicase Aquarius (Aqr), which is required for catalytic activation of the human spliceosome in addition to Prp2. On the other hand, proteins conserved between yeast and human are missing from the human B^{act} spliceosome, raising the possibility that the 3D structure of the human and yeast B^{act} complexes may differ to some extent.

The B^{act} complex was assembled in HeLa nuclear extracts using a pre-mRNA construct, PM5-10, which contained the 5' exon and an intron that is truncated 10 nucleotides (nts) after the BS, and was affinity-purified for cryo-EM analysis. After exhaustive 3D multi-reference

refinement (3D classification) of the human B^{act} particles, two major forms, termed A and B, of the B^{act} structure were obtained at resolutions of 5.3 Å and 8.1 Å, respectively. While their overall structure is largely similar, forms A and B differ with respect to the presence/absence of several densities, as described below. The structure of the central domain of the human B^{act} complex, including the catalytic U2/U6 RNP core, is highly conserved between the human and yeast spliceosomes. At the catalytic U2/U6 RNA centre, density for catalytic metal ion 2 (M2) is present but not for M1, indicating that the catalytic centre in this state is not yet functionally active. The 5' end of the intron is engaged in base-pair interactions with the U6 ACAGA box and an additional ca. 10 nts of U6 snRNA, a distinctive feature not present in yeast spliceosomes. The 5'ss is positioned close to the catalytic centre but the 5'-terminal GU nts of the intron are engaged in protein interactions with Rnf113A (hCwc24) and probably also with U2 SF3A2. The BS forms an extended helix with U2 snRNA, which is clamped between the terminal HEAT repeats of the toroidal HEAT domain of SF3B1. The BS-adenosine (BS-A) is occluded in a protein pocket, comprised of C-terminal HEAT repeats and the PHF5A (hRds3) and is spatially separated from the catalytic centre by ca. 5 nm. The hPrp2 RNA helicase is bound to the convex side of SF3B1's HEAT domain, close to the site, where the 3' end of the intron exits the HEAT domain, but spatially separated from the U2/BS helix by ca. 7 nm. Thus, in a similar way to the scenario proposed for yeast B^{act} spliceosomes, the ATP-dependent Prp2-mediated remodelling may lead to conformational changes in SF3B1's HEAT domain that liberates the first-step reactants for catalysis.

In the human B^{act} structure, the U2/U6 helix II adopts a significantly different conformation, when compared with the yeast B^{act} structure and other spliceosomal assembly intermediates including human C*. Moreover, the conserved Syf2 protein, which binds to the base of U2/U6 helix II in yeast B^{act}, is absent from human B^{act}, probably because in the latter the U2/U6 helix II is sandwiched between two other proteins. This raises the interesting possibility that in humans the catalytic activation is more complex than in yeast and requires remodelling of U2/U6 helix II and concomitant integration of hSyf2 as a pre-requisite for B* complex formation. Another distinguishing feature of the human B^{act} structure is the existence of an intricate protein-protein interaction network that connects the complex of the Aqr helicase and protein Xab2 (hSyf1) to the main body of the spliceosome. Interestingly, all four PPIases present in human B^{act} are involved in protein-protein interactions, indicating that one of their functions appears to be that of serving as bridges between various protein modules of the spliceosome.

Finally, while forms A and B of the human B^{act} structure obtained by 3D classification share most of the structural features described above, they differ in respect of the presence/absence of densities for several protein domains. The most dramatic difference is the absence of density for the large elongated Prp19 helical bundle in form A, while its well defined density is present in form B of the

B^{act} structure. At the same time, density for Ppil1 appears in form B, forming a bridge between the centre of the helical bundle and the central body of B^{act}. Moreover, the position of the U5 40K WD40 domain also differs in form A and B. Experimental evidence further indicates that the absence of a protein density in one of the forms is not due to the physical absence of the respective proteins, but instead indicates conformational flexibility of the protein domains. Further evidence indicates that form A is a precursor of form B of the B^{act} structure. Thus, it was possible to capture by cryo-EM analysis two conformational states of the human B^{act} complex that differ in their degree of conformational maturation towards the catalytic activation step. As it is unlikely, that an ATP-requiring step is involved in the transition of the B^{act} complex from form A to B, this suggests that the observed conformational changes of the various protein domains are facilitated by the thermal energy of the system.

1 Introduction

Prokaryotes and eukaryotes store the blueprints for proteins and non-coding RNAs (e.g. tRNA and rRNA) as DNA-encoded information in their genomes. Transcription of genes into messenger RNA provides the templates for ribosomal translation into proteins. While prokaryotic transcripts are readily usable for protein synthesis, a eukaryotic transcript is initially transcribed in the nucleus as a precursor messenger RNA (pre-mRNA). Numerous maturation steps of the nuclear pre-mRNA are required before translation of the mature mRNA takes place on ribosomes in the cytoplasm. One of the most dramatic maturation steps that a pre-mRNA has to undergo is due to the fact that most eukaryotic genes, unlike prokaryotic genes, are "genes in pieces", i.e. the coding segments (called exons) of a gene are interspersed with non-coding segments (called introns) (Berget, Moore et al. 1977, Chow, Gelinas et al. 1977, Gilbert 1978, Tonegawa, Maxam et al. 1978). Since the primary transcript is a faithful copy of the DNA sequence containing the complete string of exons and introns as coded in the DNA, a maturation step is required to remove the introns and reconnect the exons. This process of precise excision and re-ligation is called splicing and takes place in the nucleus concomitantly with or right after transcription and is catalyzed by a highly dynamic macromolecular machine called the spliceosome. The mosaic structure of eukaryotic genes dramatically increases the coding capacity of the rather limited number of genes in eukaryotes. Unlike the prokaryotic "one gene – one protein" situation, one eukaryotic gene can code for a large group of related proteins. This variety is created during splicing when exons of a particular gene are either included or excluded from the mature mRNA ("alternative splicing"). Additional maturation steps modify the 5’ end with a 7-methylguanosine "cap" (m^7G) and the 3’ end with a polyadenylated "tail" (Colgan and Manley 1997, Shatkin and Manley 2000).

The spliceosome is a large RNA-protein complex comprised of five major RNA-protein subunits, the small nuclear ribonucleoprotein (snRNP) particles, and an additional group of non-snRNP *trans*-acting factors. Each of the snRNP subunits consists of one or two (in case of the U4/U6 di-snRNP) uridine-rich small nuclear RNAs (U1, U2, U4, U5, U6 snRNA), common proteins and a set of specific proteins (Wahl, Will et al. 2009). Within the spliceosome these subunits contribute to an intricate network of RNA-RNA, RNA-protein and protein-protein interactions.

1.1 Pre-mRNA structure

When various organisms are compared, pre-mRNAs are found to vary dramatically in number and length of introns. In the yeast *Saccharomyces cerevisiae* less than 5% of the approximately 6 000 genes are known to have introns. They are generally around 100 nucleotides (nts)

long (Spingola, Grate et al. 1999). In contrast, an average human gene (with ca. 20 000 protein-coding genes in total) contains about 6-9 introns, varying in length from less than 10 to more than 100 000 nts (Sakharkar, Chow et al. 2004).

The spliceosome assembles *de novo* on each individual intron to be spliced. For finding and recognizing the precise splicing position, the spliceosome assembly is guided and governed by recognition of evolutionarily conserved *cis*-elements within the pre-mRNA substrate. Recognizing these splice sites with high precision is required in order to avoid aberrant splicing products and subsequent aberrant proteins. In eukaryotes (e.g. *Homo sapiens*) these *cis*-elements are comprised of consensus sequences at the 5' splice-site (5'ss), the branch site (BS) and the 3'ss (Aebi, Hornig et al. 1986, Stephens and Schneider 1992) (Figure 1.1).

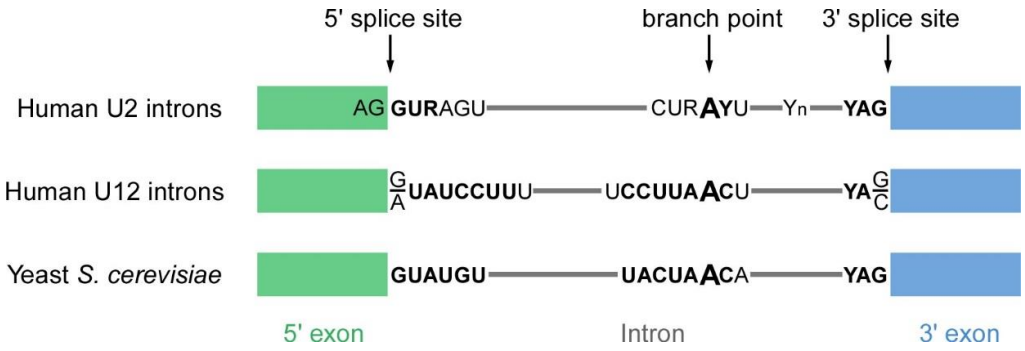


Figure 1.1: Conserved intronic consensus sequences in H. sapiens and *S. cerevisiae*

Schematic representation of conserved consensus sequences defining an intron. The branch point adenosine is underlined and the poly-pyrimidine tract is indicated by Yn. Y stands for pyrimidines and R for purines. The degree of conservation in higher eukaryotes is generally lower than in yeast (Burge, Tuschl et al. 1999).

Furthermore, regulatory RNA elements called exonic and intronic splicing enhancers (ESE and ISE) and silencers (ESS and ISS) are also recognized; these affect all splicing decisions by recruiting *trans*-acting protein factors that activate or repress spliceosome assembly at specific sites (Black 2003).

The 5'ss marks the exon/intron junction at the 5' end of the intron and includes the conserved GU dinucleotide. At the other end of the intron, the 3'ss region has the following conserved sequence elements: the BS with the highly conserved BS-adenosine, followed by a pyrimidine-rich region called the poly-pyrimidine tract (PPT), followed by the terminal AG dinucleotide (Reed 1989). The PPT plays an important regulatory role in the early stage of spliceosome formation, ensuring the fidelity of splice-site recognition (Coolidge, Seely et al. 1997). In human pre-mRNAs the branch point is usually located 18-40 nts upstream of the 3'ss (Zhang 1998), while in yeast the BS-A is usually 10-40 nts upstream of the 3'ss (Spingola, Grate et al. 1999).

There are two types of introns in higher eukaryotes, so-called U2- and U12-introns, which are removed by two different spliceosomes. The U12-type introns, which are not very abundant (less

than 1 % of introns in human), have more highly conserved 5'ss and BS (Figure 1.1). The consensus sequences are recognized during splicing by multiple snRNAs and proteins, ensuring the fidelity of splice-site definition.

1.2 The two-step mechanism of splicing

The excision of pre-mRNA introns requires two sequential transesterification reactions (Peebles, Perlman et al. 1986, Moore and Sharp 1993). The 2'-hydroxyl group of the BS-A performs a nucleophilic attack at the phosphodiester bond of the 5'ss (Figure 1.2).

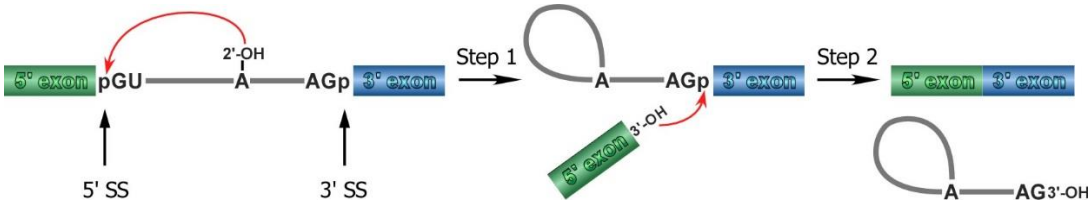


Figure 1.2: Chemistry of the splicing reaction

Two consecutive transesterification reactions lead to the excision of the intron with a lariat structure and the joining of the 5' and 3'exons.

This first reaction results in a free 5' exon and a lariat intermediate, the latter containing the intron and the downstream 3' exon. The BS-A and the G at the 5' end of the intron are connected by a 2'-5' phosphodiester bond. The second reaction occurs by the nucleophilic attack on the 3'ss by the free 3' hydroxyl of the 5' exon, leading to the spliced mRNA and intron lariat. Splicing of group II introns uses the same reaction mechanism producing the same intermediates and products.

The resulting free intron lariat is debranched and degraded, while the spliced mRNA is transported to the cytoplasm. Pre-mRNA splicing requires ATP hydrolysis during several steps of spliceosomal assembly and structural rearrangements.

1.3 U snRNPs, the major building blocks of the human spliceosome

The major (U2-type) spliceosome is formed through the interaction of five snRNPs (U1, U2, U5, U4/U6). Each snRNP contains a uridine-rich small nuclear RNA (U snRNA) (two in the case of the U4/U6 snRNP) and a variable number of proteins. A common set of proteins is made up by the seven Sm or Sm-like (LSm) proteins. U1, U2, U4 and U5 snRNA, but not U6, contain a binding site for the Sm proteins that is characterized by high uridine content. The seven Sm proteins bind to this single-stranded RNA stretch of the Sm site and form a ring-like protein structure (Guthrie and Patterson 1988, Will and Luhrmann 2001). SnRNAs with Sm-binding sites are transcribed by RNA polymerase II

as precursors containing a m^7G-cap and share a common maturation pathway undergoing multiple maturation steps including export to the cytoplasm (Matera, Terns et al. 2007, Matera and Wang 2014). The maturation of these snRNAs requires export to the cytoplasm, where the Sm-proteins B, D1, D2, D3, E, F and G are assembled on a uridine-rich stretch of the respective snRNA in a ring-like structure. Following the assembly of the Sm ring, the m^7G-cap is hypermethylated to a 2, 2, 7-trimethylguanosine cap and the 3' end of the snRNAs is trimmed. The Sm core and the hypermethylated cap then induce the re-import of the spliceosomal snRNAs into the nucleus for further maturation (Fischer, Sumpter et al. 1993). The newly imported snRNPs transiently accumulate in Cajal bodies, where they undergo site-specific pseudouridylation and 2'-O-methylation directed by scaRNAs (Kiss 2004).

The U6 snRNA, in contrast, does not bind Sm proteins, but rather LSm proteins. It is transcribed by RNA polymerase III, whose transcripts are characterized by a γ-monomethyl cap. The U6 snRNP does not assemble an Sm ring; instead, seven LSm (LSm 2-8) proteins assemble into a similar heptameric ring-like structure at the 3' end of U6 snRNA (Achsel, Brahms et al. 1999). The biogenesis and maturation of U6 snRNA are confined to the nucleus. During its maturation, the U6 snRNA also undergoes site-specific pseudouridylation and 2'-O methylation. However, in contrast to Sm snRNAs, internal modification of the U6 snRNA is mediated by small nucleolar RNAs (snoRNAs) (Kiss 2001, Kiss 2004).

While the characteristic secondary structures of the snRNAs are highly conserved in all eukaryotes, phylogenetic comparison of orthologous snRNAs reveals that U snRNAs from different species have high sequence diversity. Although the primary sequence is not conserved, the U4 and U6 snRNAs in metazoan as well as in yeast are always extensively base paired, forming a Y-shaped structure (Brow and Guthrie 1988, Will and Luhrmann 2006) (Figure 1.3). SnRNA regions involved in interactions with the pre-mRNA consensus sequences or other snRNAs, are highly conserved. The proposed secondary structures of the human snRNAs are shown in Figure 1.3. These structures are substantially rearranged during spliceosome assembly.

Although human and yeast snRNAs are remarkably similar in their critical secondary structure features, there are also some significant structural differences between them. Unlike humans, which possess one "long" U5 snRNA (Chanfreau, Elela et al. 1997), yeast has two distinct U5 snRNAs that differ only in their length at the 3' end (U5L/S for long and short U5 snRNA). Furthermore, U1 snRNA in yeast is significantly longer than its human counterpart (Kretzner, Krol et al. 1990). The most dramatic difference is observed in the yeast U2 snRNA. While the structural features of the human U2 are retained, yeast U2 displays a huge (non-essential) insertion of 945 nucleotides replacing the human stem loop III immediately downstream of the Sm site.

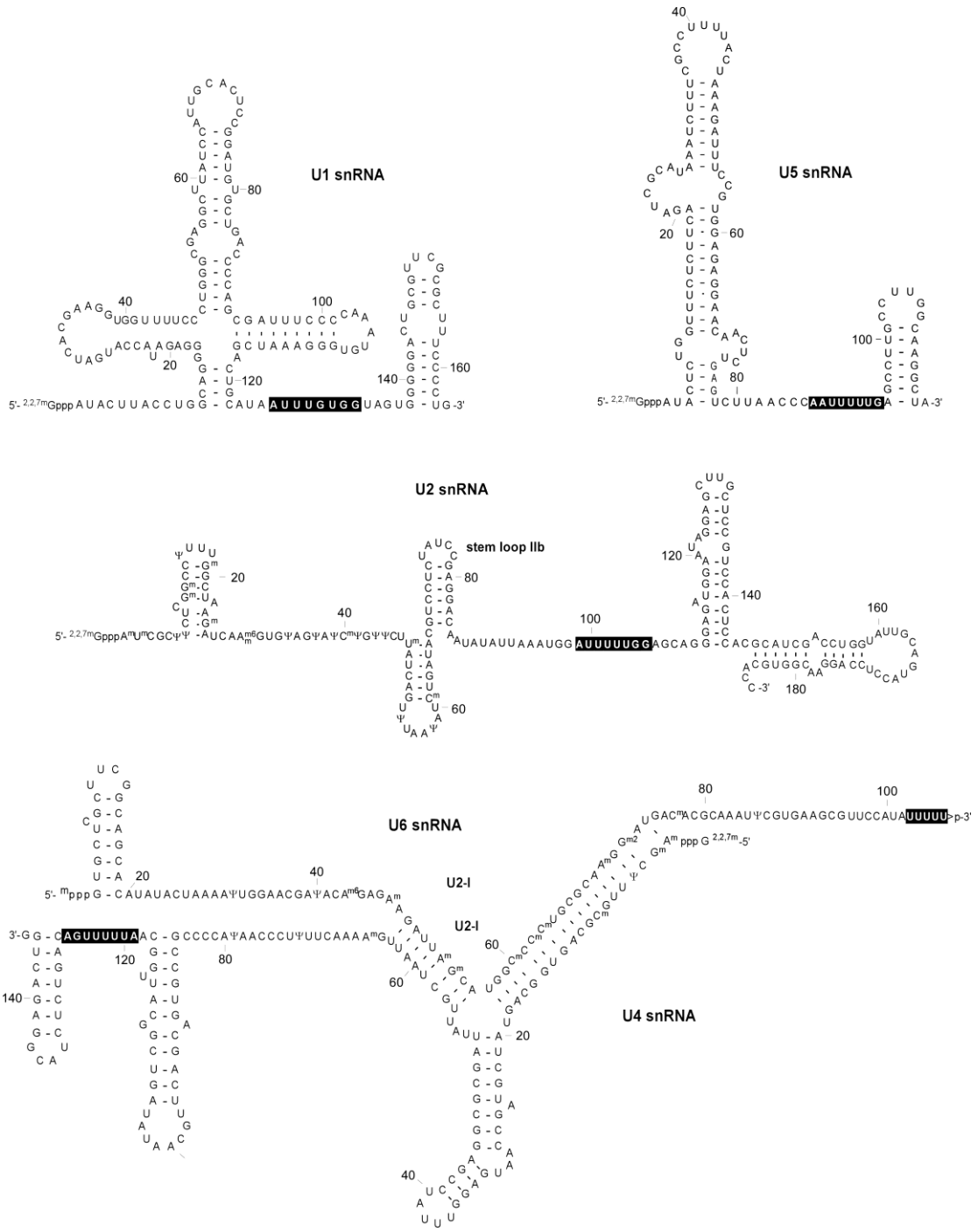


Figure 1.3: Sequences and predicted secondary structures of the human spliceosomal snRNAs

The proposed secondary structures of the human snRNAs. U4 and U6 snRNAs are extensively base paired, forming a Y-shaped structure. The Sm-sites are indicated as black boxes.

The minor U11, U12, U4atac and U6atac snRNAs, functional analogues of the major snRNAs, also show high level of similarity in secondary structure fold, despite only limited sequence homology and a lower degree of conservation of their sequence (Tarn and Steitz 1996).

In addition to the group of common proteins, the other important group of protein constituents of snRNPs is that of the particle-specific proteins. The human U1 snRNP contains only three particle-specific proteins, U1-70K, U1-A and U1-C (Hinterberger, Pettersson et al. 1983, Bringmann and

Luhrmann 1986). This composition results in an observed Svedberg (S) value of 12S. U1-C is involved in stabilization of the initial base pairing interactions between U1 snRNA and the 5'ss of the pre-mRNA (Heinrichs, Bach et al. 1990), whereas U1-70K interacts with SR proteins to stabilize the interaction of U1 snRNP with the pre-mRNA (Kohtz, Jamison et al. 1994).

The human 17S U2 snRNP is composed of proteins U2-A', U2-B'' and the heteromeric sub-complexes SF3A and SF3B. First the U2-A' and U2-B'' proteins bind to stem loop IV of U2 snRNA as a dimer, thus forming a 12S U2 snRNP. The SF3A and SF3B proteins help to stabilize the base pairing between the U2 snRNA and the BS by interactions with the pre-mRNA around the BS (Krämer and Utans 1991, Gozani, Feld et al. 1996, Valcárcel, Gaur et al. 1996), whereas SF3B6 (p14) can be crosslinked to the BS-A (Will and Luhrmann 2001). The SF3A complex consists of three subunits with molecular weights of 120, 66 and 60 kDa (hereinafter termed SF3A1, SF3A2and SF3A3, respectively) (Brosi, Hauri et al. 1993), whereas the SF3B complex consists of seven subunits (SF3B155, SF3B145, SF3B130, SF3B49, SF3B10, SF3Bp14 and SF3B14b) (hereinafter termed SF3B1 – B6 and PHF5A, respectively) (Will, Urlaub et al. 2002). Additionally, a set of protein factors has been identified as the U2-related proteins. This set of proteins, including hPrp43, Spf45, Spf30, Spf31, SR140, CHERP, PUF60, hPrp5 as well as the U2AF65/35 heterodimer, is loosely associated with the 17S U2 snRNP (Will, Urlaub et al. 2002).

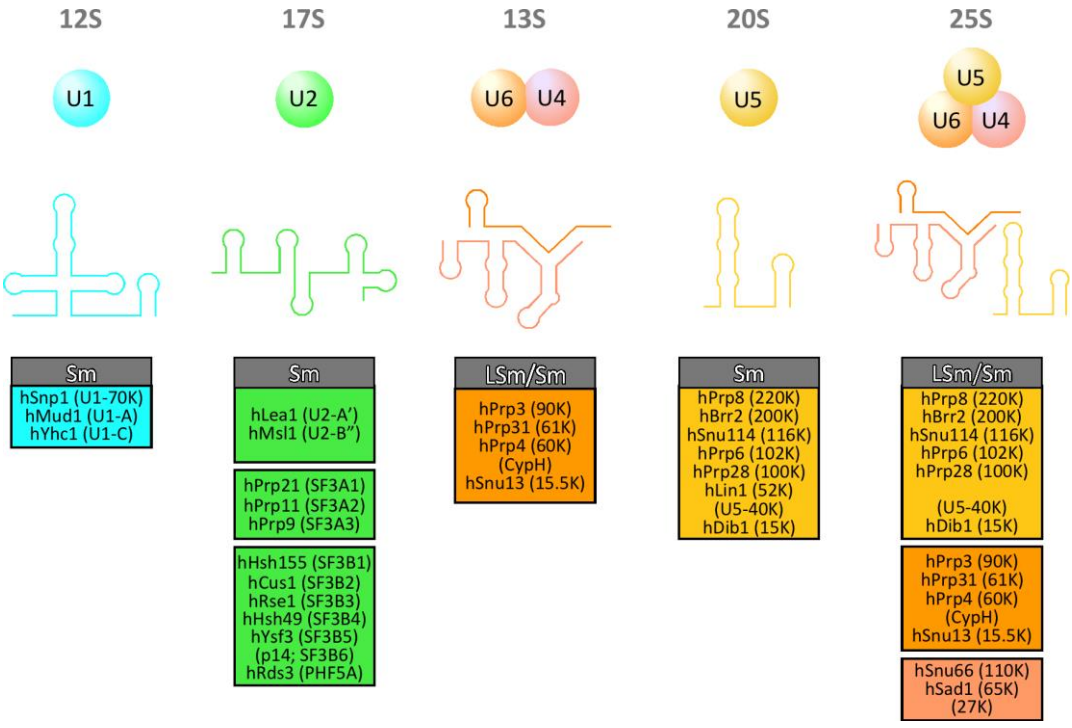


Figure 1.4: Composition of the spliceosomal U snRNPs

Each snRNP consists of one (or two) snRNA molecules and different numbers of complex-specific proteins. The secondary structure of the human U snRNAs is shown schematically. The protein composition of the U snRNPs is shown in boxes of different colors. The U4/U6.U5 tri-snRNP is composed of U5 and U4/U6. Sm/Lsm proteins are shown in gray boxes. Protein names are indicated according to yeast (human) nomenclature. "h" stands for "human".

The 20S U5 snRNP harbors eight particle-specific proteins, hPrp8, hBrr2, hSnu114, hPrp6, hPrp28, hLin1, 40K and hDib1 (Bach, Winkelmann et al. 1989). The hPrp8 protein is the largest protein in the spliceosome and was shown to be in contact with the 5'ss, 3'ss and the BS at different time points of the splicing cycle (Teigelkamp, Newman et al. 1995, Teigelkamp, Whittaker et al. 1995, Reyes, Kois et al. 1996, Reyes, Gustafson et al. 1999, McPheeters and Muhlenkamp 2003, Grainger and Beggs 2005). The U5-specific proteins hPrp28 and hBrr2 are RNA helicases were identified as crucial players during formation of the pre-catalytic spliceosome and its subsequent activation, respectively (Raghunathan and Guthrie 1998, Laggerbauer, Liu et al. 2005). hBrr2 unwinds the U4/U6 di-snRNP during spliceosome activation, hPrp28 instead displaces the U1 snRNP from the 5'ss (Strauss and Guthrie 1994, Möhlmann, Mathew et al. 2014). The U5-specific proteins hPrp8 and hSnu114 have been shown to regulate hBrr2 activity (Small, Leggett et al. 2006, Maeder, Kutach et al. 2009, Mozaffari-Jovin, Wandersleben et al. 2013, Mozaffari-Jovin, Wandersleben et al. 2014).

The U4/U6 di-snRNP is associated with the five specific proteins hPrp3, hPrp31, hPrp4, CypH and hSnu13, resulting in an S value of 13S (Behrens and Luhrmann 1991, Lauber, Plessel et al. 1997). Upon association of the 13S U4/U6 di-snRNP with the 20S U5 snRNP the 25S U4/U6.U5 tri-snRNP is formed, which additionally gains three tri-snRNP-specific proteins hSnu66, hSad1 and 27K (Black and Pinto 1989, Behrens and Luhrmann 1991). The U5-specific protein hLin1 is proposed to play a role during assembly of this 25S particle, but it is no longer present in the mature U4/U6.U5 tri-snRNP (Laggerbauer, Liu et al. 2005). The assembly of the U4/U6.U5 tri-snRNP appears to be mediated mainly through protein-protein interactions. The proteins hPrp6 and hPrp31 are important bridging factors, as removal of either protein abolishes U4/U6.U5 tri-snRNP formation *in vitro* and *in vivo* (Makarova, Makarov et al. 2002, Schaffert, Hossbach et al. 2004).

1.4 Dynamic assembly pathway of the spliceosome

Spliceosome assembly occurs by the ordered interaction of the five snRNPs and numerous splicing factors with the pre-mRNA. The spliceosome assembles *de novo* on each intron and follows a sequence of defined stages (Figure 1.5). Assembly and remodeling of the spliceosome during the full splicing cycle creates the catalytic center for intron excision and ligation of exons. The dynamic assembly and dismantling of the spliceosome requires several different driving forces and control mechanisms, which are provided by DExD/H-box ATPases or RNA helicases, a group of important non-snRNP proteins (Wahl, Will et al. 2009, Cordin, Hahn et al. 2012). These molecular motors use the energy of nucleoside triphosphate (NTP) hydrolysis for the rearrangement of RNA-RNA and RNA-protein interactions modulating the structure and function of specific RNAs and/or RNPs (Staley and Guthrie 1998, Pyle 2008). Eight DExD/H-box helicases (human/yeast: UAP56/Sub2, DDX46/Prp5, hBrr2/Brr2, DHX16/Prp2, DDX23/Prp28, DHX38/Prp16, DHX8/Prp22 and DHX15/Prp43), conserved

between humans and yeast, are sufficient to drive splicing in *S. cerevisiae*, whereas spliceosomes of higher eukaryotes possess four additional helicases: SF3B125, Aquarius (Aqr), Abstrakt, and DHX35 (De, Schmitzova et al. 2016).

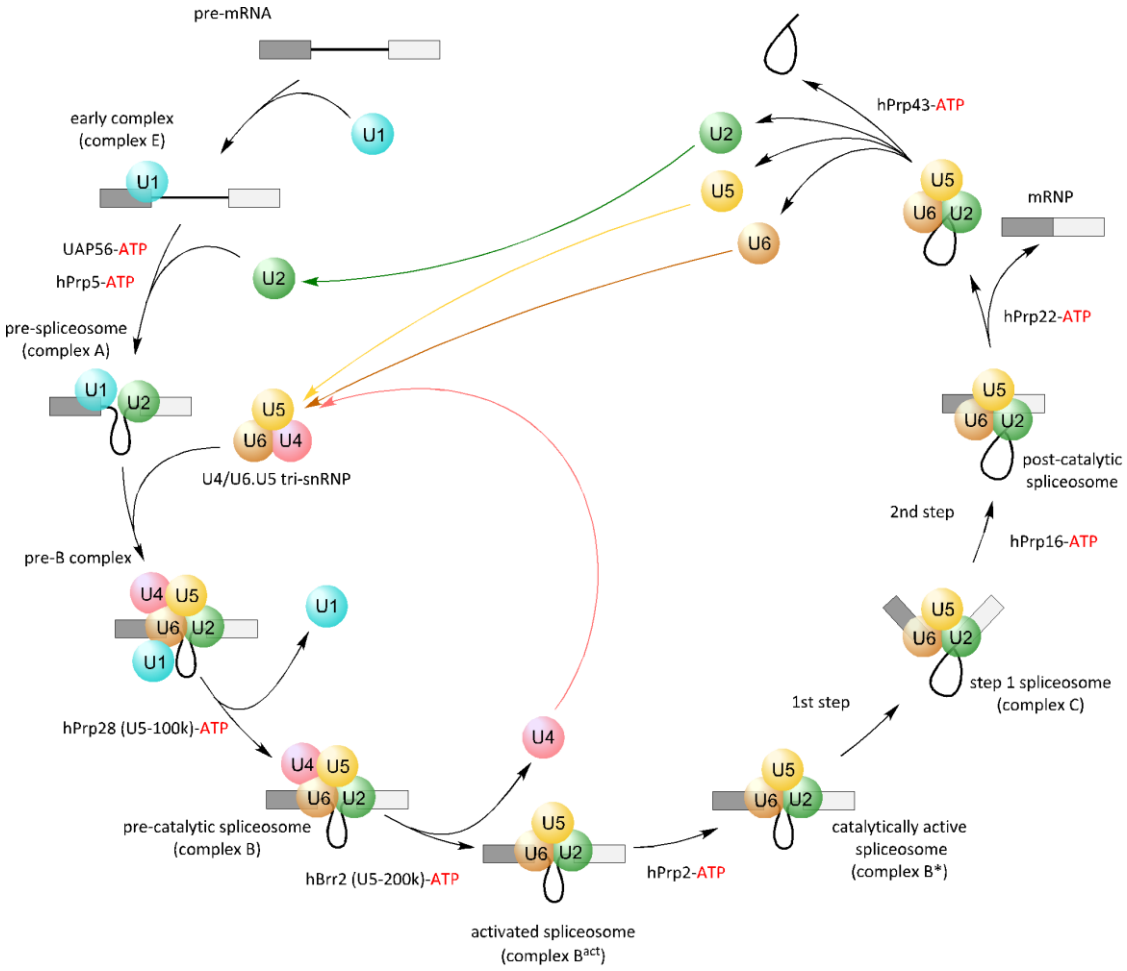


Figure 1.5: Stepwise assembly of the human spliceosome

Schematic representation of splicing cycle. For simplicity the U snRNPs are represented as circles labeled with its respective U snRNA. The spliceosomal complexes are named according to the metazoan nomenclature. The eight conserved DExD/H-box ATPases/helicases, which are required for important RNP remodeling events, are indicated. Exons and introns are indicated as colored boxes and bold lines, respectively [modified from Will and Lührmann, 2011].

Assembly of the spliceosome is initiated by recognition of the 5'ss by the U1 snRNP (Figure 1.5). At the same time, SF1 and U2AF65/35 bind to the BS and the PPT/3'ss, respectively (Ruskin, Zamore et al. 1988, Berglund, Chua et al. 1997). These associations lead to the formation of the so-called E complex (Reed 1990). The U2 snRNP is already present in the E complex, but an ATP-dependent rearrangement mediated by the RNA helicase Prp5 is required to establish a firm interaction of U2 snRNP with BS, which results in formation of the A complex (Dalbadie-McFarland and Abelson 1990, Fleckner, Zhang et al. 1997). These interactions involve base pairing of the U2 snRNA with the BS, which leads to the "bulging-out" of the BS-A and interactions of the U2-SF3A/B proteins near the BS. The BS-A is directly contacted by the SF3B6 protein (Query, Moore et al. 1994, Will and

Luhrmann 2001). Moreover, SF3B1 not only interacts with the pre-mRNA, but also with U2AF65 to stabilize the binding of U2 snRNP to the pre-mRNA (Gozani, Potashkin et al. 1998). Thus, the recognition of several splicing signals is a general mechanism of spliceosome assembly and splicing catalysis that ensures splicing accuracy.

Next, the U4/U6.U5 tri-snRNP is recruited to the spliceosome, forming a 37S pre-B complex, in which the tri-snRNP is only loosely bound (Boesler, Rigo et al. 2016). Stable integration of the U4/U6.U5 tri-snRNP into the spliceosome requires the action of Prp28, which destabilizes the binding of U1 snRNP and yields the B complex. During stabilization of the B complex the U4/U6.U5 tri-snRNP associated factors hPrp31 and hPrp6 are phosphorylated by the hPrp4 kinase, which was shown to be essential for pre-mRNA splicing *in vitro* (Schneider, Hsiao et al. 2010). The B complex is still in a catalytically inactive state. Activation of the spliceosome is initiated by the RNA helicase Brr2, which leads to the dissociation of U4 snRNP from the spliceosome and to the formation of the activated spliceosome or B^{act} complex (Laggerbauer, Achsel et al. 1998, Raghunathan and Guthrie 1998).

Subsequent action of the helicase Prp2 then transforms the B^{act} spliceosome into a catalytically activated state, the so-called B* complex, which carries out the first step of splicing, resulting in the C complex. The C complex contains the excised 5’ exon and the 3’ exon-intron lariat intermediate. Upon the action of the RNA helicase Prp16 the spliceosome undergoes another structural rearrangement, yielding the intermediate complex C*, which performs the second step of splicing (Umen and Guthrie 1995). During the second step, the lariat intron is excised and the 5’ and 3’ exons are ligated. Prp22 is needed for the release of the spliced mRNA from the intron-lariat spliceosome (Company, Arenas et al. 1991) and the mRNA is exported to the cytoplasm. Finally, Prp43 RNA helicase promotes the disassembly of the intron-lariat spliceosome, and the released snRNPs are reshuffled for a fresh round of splicing and the intron is degraded (Martin, Schneider et al. 2002, Fourmann, Schmitzova et al. 2013).

1.5 The dynamic network of spliceosomal RNA-RNA interactions and RNA-metal-mediated catalysis of pre-mRNA splicing

During spliceosome assembly an intricate network of RNA-RNA interactions involving snRNAs and the pre-mRNA substrate is established (Wahl, Will et al. 2009, Will and Luhrmann 2011). This network of RNAs undergoes significant rearrangements during the spliceosome's transition from one distinct type of complex to the next. Spliceosome assembly is initiated by the recognition of the 5’ ss through base-pairing interactions with the 5’ end of U1 snRNA (see Figure 1.6) (Zhuang and Weiner 1986) and the formation of a base-pairing interaction of the U2 snRNA with the BS (Parker, Siliciano et al. 1987, Wu and Manley 1989). In the short U2/BS helix, the BS-A is branched out. In the 37S

pre-B complex the 3' end of the U6 snRNA is base-paired to the 5' end of the U2 snRNA, forming the U2/U6 helix II (blue in Figure 1.6). In the pre-B complex, U1 snRNA is still base-paired to the 5’ ss and the tri-snRNP is not yet stably bound. Within the tri-snRNP the U4 and U6 snRNA are base-paired and form a Y-shaped structure composed of stems I and II and the U4 snRNA internal stem-loop. The ACAGA sequence directly upstream of stem I within the U6 snRNA (the so-called ACAGAG box) is highly conserved.

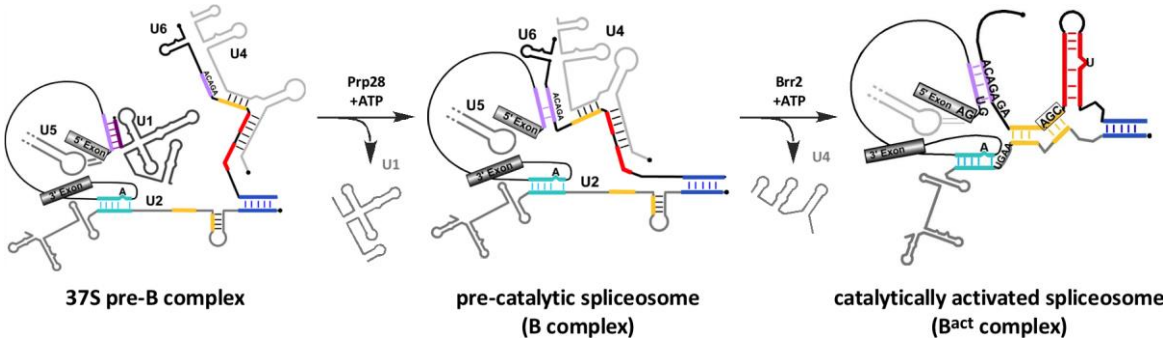


Figure 1.6 Rearrangements of the RNA-RNA network during formation of a catalytically activated spliceosome

The exons are shown as rectangular boxes. The RNA molecules are depicted as black and grey lines respectively. Important RNA sequences that are involved in dynamic base-pairing interactions are depicted in colour. See the text for the detailed explanation.

For stable integration of the tri-snRNP into the spliceosome, Prp28 removes the U1 snRNA from the 5’ss, allowing the latter to base-pair with ACAGA-box of the U6 snRNA (violet in Figure 1.6). In the B complex the stem-loop I of U5 snRNA interacts with the 3’ terminal nucleotides of the 5’ exon. The U5 loop I plays a role also later, in the positioning of both exons for both steps of splicing (Grainger and Beggs 2005). Activation of the pre-catalytic B complex by Brr2 helicase leads to the disruption of the U4/U6 base-pair interactions and the displacement of U4 snRNP from the spliceosome. U6 snRNA is now restructured, forming the intramolecular stem-loop (U6 ISL), and engages in new base-pair (bp) formation with U2 snRNA, yielding the U2/U6 helices Ia and Ib. (Figure 1.6) (Madhani and Guthrie 1992). The B^{act} complex is then catalytically activated by Prp2 RNA helicase, yielding the B* complex, which can catalyze the first step of the splicing reaction. Biochemical and genetic studies in yeast have provided strong evidence that within the B^{act} complex a catalytic U2/U6 RNA-RNA interaction network is assembled that is highly similar to the catalytic RNA network of group II self-splicing introns (Fica, Tuttle et al. 2013), and this was recently further confirmed by cryo-EM structures of yeast and human spliceosomes (Rauhut, Fabrizio et al. 2016, Yan, Wan et al. 2016). Thus, similar to the catalytic domain V of group II introns, the U2/U6 helix Ib and the U6 ISL adopt a secondary structure in which a conserved AGC triad is situated 5 bp away from a conserved bulged-out nt in U6 (U6-U80 in *S. cerevisiae*). Moreover, the U6 catalytic metal ligands, situated in the triad and the bulge, correspond one-to-one to the domain V ligands (Figure 1.7). Finally, similar to the

group II intron catalytic core, a group II-like triple helix also appears to form in U6 snRNA to juxtapose the two metal binding sites at the U6-U80 bulge loop and the AGC triad.

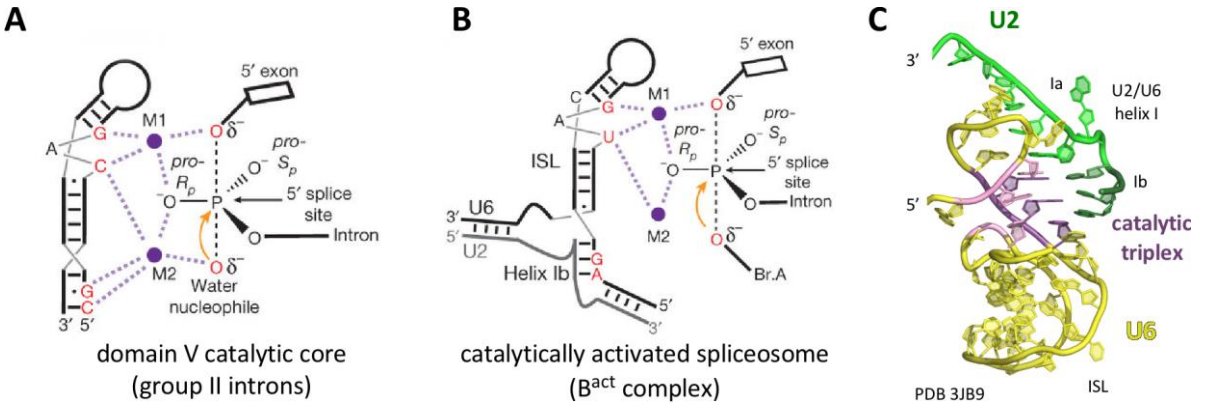


Figure 1.7 Model for catalytic metal interactions

Catalytic metal interactions during domain V hydrolysis **(A)** and pre-mRNA splicing **(B)**. The reactive oxygens are colored red, the pre-mRNA scissile phosphate is depicted in a transition state, and interactions between specific ligands and the reactive oxygens mediated by M1 and M2 are shown as light magenta dashed lines. Modified from Fica *et. al.,* 2014. (C) 3D model of the catalytic triad, which comprises the ISL of U6 snRNA and helix I of the U2/U6 duplex (PDB 3JB9, Hang *et al*., 2015)

In yeast, this spliceosomal triplex comprises Hoogsteen interactions of the terminal G52 and A53 of the conserved U6 ACAGA box with nucleotides G60 and A59, and of U6-U80 with C61 of the AGC triad. Cryo-EM 3D structures of the *S. cerevisiae* B^{act} complex demonstrated that this intricate catalytic U2-U6 RNA network is already in place in the activated spliceosome (Rauhut, Fabrizio et al. 2016, Yan, Wan et al. 2016). Moreover, the cryo-EM 3D structures of the yeast C complex (Galej, Wilkinson et al. 2016, Wan, Yan et al. 2016), the yeast and human C* complexes, which are activated for catalysis of step II of splicing (Bertram, Agafonov et al. 2017, Fica, Oubridge et al. 2017, Yan, Wan et al. 2017) and of the *S. pombe* intron-lariat spliceosome (ILS) (Hang, Wan et al. 2015, Yan, Hang et al. 2015), revealed that the catalytic U2-U6 RNA core exhibits a similar overall 3D structure during the two catalytic phases of the spliceosome.

1.6 Non-snRNP proteins of the spliceosome

The isolation of spliceosomal complexes representing distinct snapshots of the full spliceosomal cycle, combined with mass-spectrometric analysis, identified a large number of so-called non-snRNP proteins (Will and Luhrmann 2006). These include the groups of early splicing proteins and the spliceosomal RNA helicases discussed above. Altogether, more than 170 proteins have been identified in the human spliceosome, while only about 90 proteins have been detected in yeast spliceosomal complexes (including snRNP proteins). Interestingly, the vast majority of the yeast spliceosomal proteins have counterparts in the human spliceosome, indicating that the yeast spliceosome is the evolutionarily conserved core design of the splicing machinery. In the following,

emphasis is placed on the description of the non-snRNP proteins of the activated spliceosome, as these play an important role in the generation of a functionally active U2/U6 catalytic RNA core and will be discussed in detail in the cryo-EM structure of the human B^{act} complex in the results section below. Moreover, several of the non-snRNP proteins of the B^{act} complex are pre-organized into stable multimeric protein complexes.

In yeast, a major protein complex is represented by the so-called NineTeen Complex (NTC). All but two of the NTC proteins are evolutionarily conserved in humans; however, they are pre-organized in different sub-complexes. For example, the human NTC proteins Cdc5, Prl1, AD002, Spf27 (yeast analogues: Cef1, Prp46, Cwc15 and Snt309, respectively) and hPrp19 are present in the human hPrp19/Cdc5L complex, which in addition contains the CTNNBL1 and Hsp73 proteins. The NTC proteins Isy1 and Syf1 are part of the so-called intron binding complex (IBC) which also contains the RNA helicase Aqr, the cyclophilin E (CypE) and CCDC16 (De, Bessonov et al. 2015) (Figure 1.8).

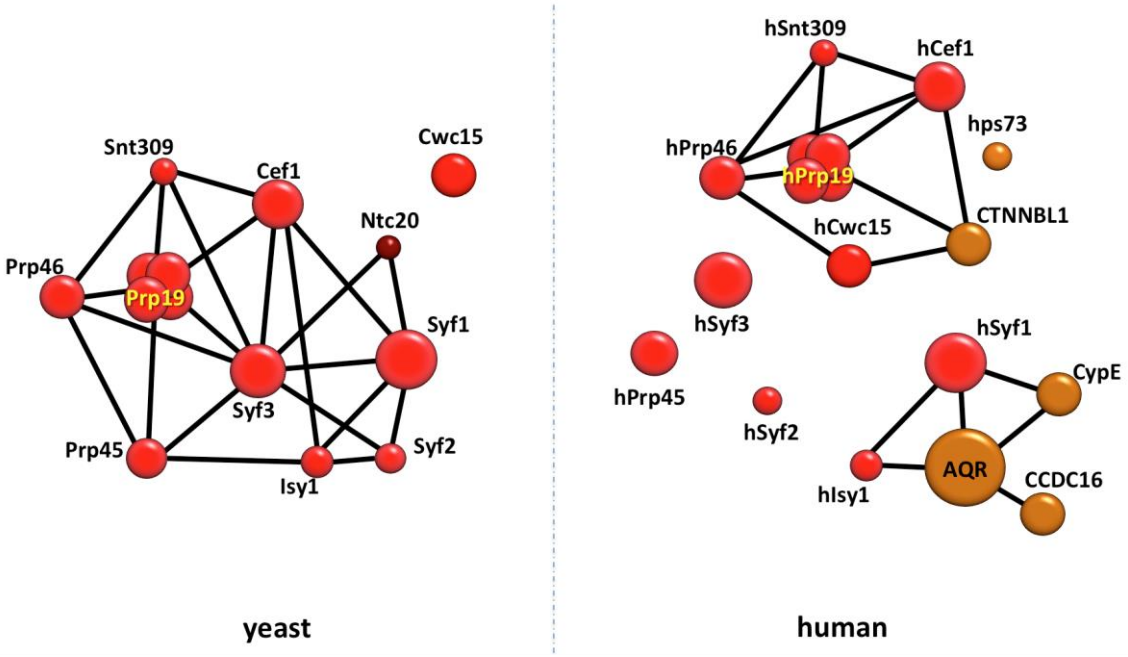


Figure 1.8 Organisation of yeast NTC proteins and their human analogues

In yeast, 10 proteins are recruited as a pre-formed complex (left). Ntc20 is depicted in dark red, it has no counterparts in human. Proteins depicted in orange are human-specific. Human proteins are pre-organised into two major building blocks.

The human homologues of the *S. cerevisiae* NTC complex proteins, Skip (hPrp45), Crnkl1 (hSyf3) and hSyf2, are not part of the human Prp19/Cdc5L or IBC complexes (Figure 1.8). Except for hSyf2, which is recruited first to the human C complex, all other proteins are recruited to the B^{act} spliceosome and play an important role in the formation of the catalytic U2/U6 RNP core (see below).

The heterotrimeric RES (for REtention and Splicing) complex is a conserved spliceosome-associated protein module that was shown to enhance the splicing of a subset of transcripts and to promote the nuclear retention of unspliced pre-mRNAs in yeast (Gottschalk, Bartels et al. 2001, Dziembowski,

Ventura et al. 2004). The human homologues of the yeast RES complex subunits, hBud13, Snip1 (hPml1) and RbmX2 (hSnu17) associate stably with the spliceosome during its activation phase (Deckert, Hartmuth et al. 2006, Bessonov, Anokhina et al. 2008, Bessonov, Anokhina et al. 2010). Additional proteins that associate with the human activated spliceosome are KIAA1604 (hCWC22), Rnf113A (hCwc24), hPrp17, CCDC12, Rbm22, G10 (hBud31), PRCC, the cyclophilins PPIL1, PPIL2, and NY-CO-10 (hCwc27), and the exon junction complex (EJC) proteins eIF4A3, Y14 and Magoh (Bessonov, Anokhina et al. 2010) (Figure 1.9).

Several of the proteins that are recruited to the B^{act} complex are later displaced from the spliceosome. Also, other defined sets of non-snRNP proteins are transiently recruited to other assembly intermediates of the spliceosome, as will be discussed below.

Finally, another important group of non-snRNP splicing factors regulating splicing activity and modulating alternative splicing, consist of two families of proteins. The first is represented by serine/arginine-rich (SR) proteins. SR proteins share common structural features, such as one or more amino-terminal RNA recognition motif(s) (RRMs) and a variable number of arginine/serine (RS) repetitive dipeptides in their carboxy-terminal domain (Birney, Kumar et al. 1993, Graveley 2000, Long and Caceres 2009). Therefore, these proteins can interact at the same time with both RNA through their RRM domains and with other proteins through their RS domain (Graveley and Maniatis 1998). For example, the recognition of ESE regions of a pre-mRNA by SR proteins promotes spliceosome assembly at adjacent splice sites. SR proteins can also promote snRNP interactions with the pre-mRNA, such as stabilization of the U1/5'ss base-pairing interaction at the initial stage of splicing assembly, and thus modulate alternative splicing events (reviewed in Graveley, 2000). RS domains are also found in non-SR proteins, for instance, in U2AF subunits (Zamore, Patton et al. 1992).

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

1.7 The dynamics of the protein composition of the human spliceosome

Proteins account for the largest share of the spliceosome's mass and play important roles throughout the entire splicing cycle. They enable all conformational and structural rearrangements needed for the catalytic activity of the spliceosome. A total of more than 170 different spliceosome-associated

proteins were identified in affinity-purified human spliceosomal complexes, with each complex containing approximately 120 different proteins (Wahl, Will et al. 2009). Mass spectrometry analysis of purified spliceosomes confirms that the protein composition of the spliceosome is highly dynamic, with extensive changes of the protein inventory from one splicing step to the next (Deckert, Hartmuth et al. 2006, Behzadnia, Golas et al. 2007, Bessonov, Anokhina et al. 2008, Fabrizio, Dannenberg et al. 2009). Validation of quantitative changes of proteins was furthermore helped by 2D gel electrophoresis (Agafonov, Deckert et al. 2011).

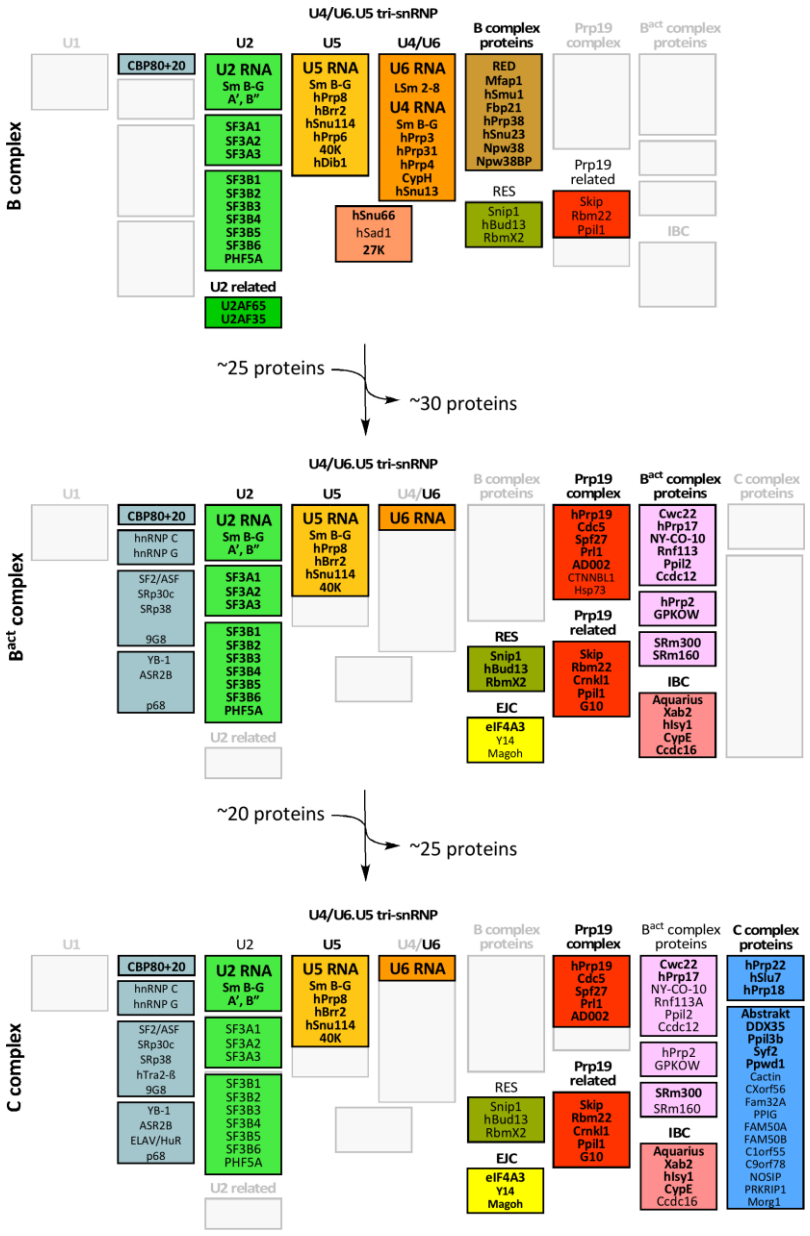


Figure 1.9: Compositional dynamics of the human spliceosome

The protein composition of the human B, B^{act}, and C complexes based on 2D gel-electrophoresis and mass spectrometry analysis is shown. The relative abundance of proteins is indicated by bold (stoichiometric amounts) or light (substoichiometric amounts) lettering. Proteins are grouped according to snRNP association, function and presence in a stable heteromeric complex or association with a particular spliceosomal complex, as indicated.

In the following paragraphs, I will summarize the dynamic behavior only of the most abundant proteins that are found in the purified human B, B^{act} and C complexes (Figure 1.9). Stable B complexes, which have been isolated in the presence of ATP-γ-S (Agafonov, van Santen et al. 2016), contain all the U2 and tri-snRNP proteins, except that Prp28, Rbm42 and Sad1 are already largely absent. On the other hand, a set of 8 proteins, termed B-specific proteins, RED, Smu1, Mfap1, Fbp21, hPrp38, hSnu23, Npw38 and Npw38BP have been recruited to the stable B complex, and these are not present in the 37S pre-B complex. Moreover, small amounts of the hRES complex proteins are also present (Agafonov, Deckert et al. 2011, Boesler, Rigo et al. 2016).

The most dramatic exchange of proteins occurs during the transition from the B complex to the activated spliceosome. On the one hand, all of the U4/U6 specific proteins, the tri-snRNP proteins and the U5 proteins hPrp6 and hDib1, most of the B-specific proteins, as well as some A complex proteins are largely absent from the B^{act} complex. On the other hand, the hPrp19/CDC5L and IBC complex proteins and all other B^{act} proteins mentioned above, as well as hPrp2 and its co-activator GPKOW, are stably recruited to the B^{act} complex. Moreover, the RES proteins and the EJC proteins eIF4A3, Y14 and Magoh are also present in significant amounts in the B^{act} complex (Figure 1.9).

In contrast to yeast, the catalytic activation of the human B^{act} complex to the B* complex has not yet been investigated, nor has the hB* complex been isolated. In yeast, Prp2 and its co-activator are sufficient to transform the B^{act} into the B* complex. This is accompanied by major RNP-remodeling events. For example, the yeast B^{act} proteins Cwc24 and Cwc27, as well as part of the RES proteins are displaced from the spliceosome (Warkocki, Odenwälder et al. 2009, Ohrt, Prior et al. 2012). Moreover, the U2 SF3A/B proteins are destabilized and their association with the spliceosome becomes salt-sensitive (Warkocki, Odenwälder et al. 2009, Lardelli, Thompson et al. 2010, Ohrt, Prior et al. 2012). As in the human C complex Rnf113A and NY-CO-10 are also largely underrepresented and the U2 SF3A/B proteins are likewise destabilized, it is reasonable to assume that these remodeling events also occur in the human spliceosome, already during the catalytic activation process (Bessonov, Anokhina et al. 2010). However, there is evidence that the catalytic activation process of human B^{act} complexes is more complex than in yeast. For example, the ATPase activity of the RNA helicase Aqr, which is absent in *S. cerevisiae*, is also required for generating a human B* like complex (De, Sessonov et al. 2015). Moreover, at least two DExH/D box RNA helicases are abundant in the purified human C complex, and it is not known whether one or both of these proteins may also play a role in B* formation.

The protein composition of the purified human C complex differs significantly from that of the B^{act} complex. In addition to Rnf113A and NY-CO-10 also Ccdc12, Ccdc16, Ppil2 and hPrp2 and its coactivator GPKOW are largely absent in the C complex. On the other hand numerous proteins,

including the RNA helicases Abstrakt and DDX35, the cyclophilins Ppil3b and and Ppwd1, the step 2 factors hSlu7, hPrp18 and hPrp22, and the proteins Cactin, Cxorf56 and Fam32A are first recruited to the C complex. While these represent the more abundant C complex-specific proteins, several others are also present in C complexes but are clearly sub-stoichiometric (Bessonov, Anokhina et al. 2008, Agafonov, Deckert et al. 2011).

1.8 3D structures of snRNPs and spliceosomes

Until very recently, the molecular architecture and three-dimensional (3D) organization of the spliceosome remained only poorly understood. A number of high resolution 3D structures of several larger protein domains or protein-protein and protein-RNA complexes were obtained primarily by X-ray crystallography. These include, among others, the helicase domain of Brr2 (Santos, Jovin et al. 2012, Nguyen, Li et al. 2013), parts of Prp8 (Pena, Liu et al. 2007, Pena, Rozov et al. 2008, Ritchie, Schellenberg et al. 2008, Yang, Zhang et al. 2008, Galej, Oubridge et al. 2013), various U4 (Li, Leung et al. 2016) and U6 RNA-protein complexes (Montemayor, Curran et al. 2014) and even recombinantly produced (Kondo, Oubridge et al. 2015) or native U1 snRNPs (Pomeranz Krummel, Oubridge et al. 2009). However, for a large ensemble such as the spliceosome, which contains many flexible parts, crystallisation appears to be a major challenge if not unreachable. Therefore, cryo-EM is the method of choice for studying the structures of large RNP complexes.

Low to medium-resolution (<12–40 Å) 3D cryo-EM structures of the human pre-spliceosomal A, spliceosomal B and C complexes (Boehringer, Makarov et al. 2004, Jurica, Sousa et al. 2004, Behzadnia, Golas et al. 2007), as well as some of their building blocks such as the U1 snRNP (Stark, Dube et al. 2001), the U2-associated SF3B protein complex (Golas, Sander et al. 2003, Golas, Sander et al. 2005), or the U4/U6.U5 tri-snRNP were initially reported. Moreover, immuno-labeling of spliceosomal components was used to assign major building blocks in the low resolution structures of the spliceosome (Stark and Lührmann 2006).

Recently, mainly owing to the invention of direct detectors and powerful software packages, a revolution has taken place in the EM field, allowing the near-atomic resolution of large molecular ensembles (Kühlbrandt 2014). A first breakthrough was achieved by the group of Yigong Shi, who determined the structure of the major part of the endogenous intron-lariat spliceosome from *S. pombe* (Hang, Wan et al. 2015, Yan, Hang et al. 2015). This provided, for the first time, molecular insight into the organisation of the catalytic RNP core of the spliceosome. First of all, it proved that the catalytic U2/U6 RNA network adopts a three-dimensional structure very similar to that of the catalytic RNA core of group II self-splicing introns, including the existence of the catalytic triplex (Fica, Mefford et al. 2014). Moreover, the two catalytic Mg^{2+} ions were coordinated by nucleotides of the

catalytic triplex, which is consistent with biochemical results obtained by phosphorothioate substitutions (Fica, Tuttle et al. 2013). Secondly, the cryo-EM structure not only revealed the atomic structure of the entire Prp8 protein, but also demonstrated that Prp8 forms an active site, made up by the N-terminal domain and part of its central domain, into which the catalytic U2/U6 RNA network is docked. In addition, about 10 proteins or major parts thereof, many of them belonging to the NTC, interact with the catalytic RNA network, providing clear evidence that a functional catalytic U2/U6 RNA structure can only exist as part of this RNP core. Another interesting structural feature of the ILS complex is the existence of a helical bundle comprising the coiled-coil domain of the four Cwf8 (Prp19) copies, a long C-terminal alpha helix of Cdc5 and a long alpha-helical part of Cwf7 (Spf27), which was connected to the main body of the spliceosome by small bridges (Yan, Hang et al. 2015) (Figure 1.10).

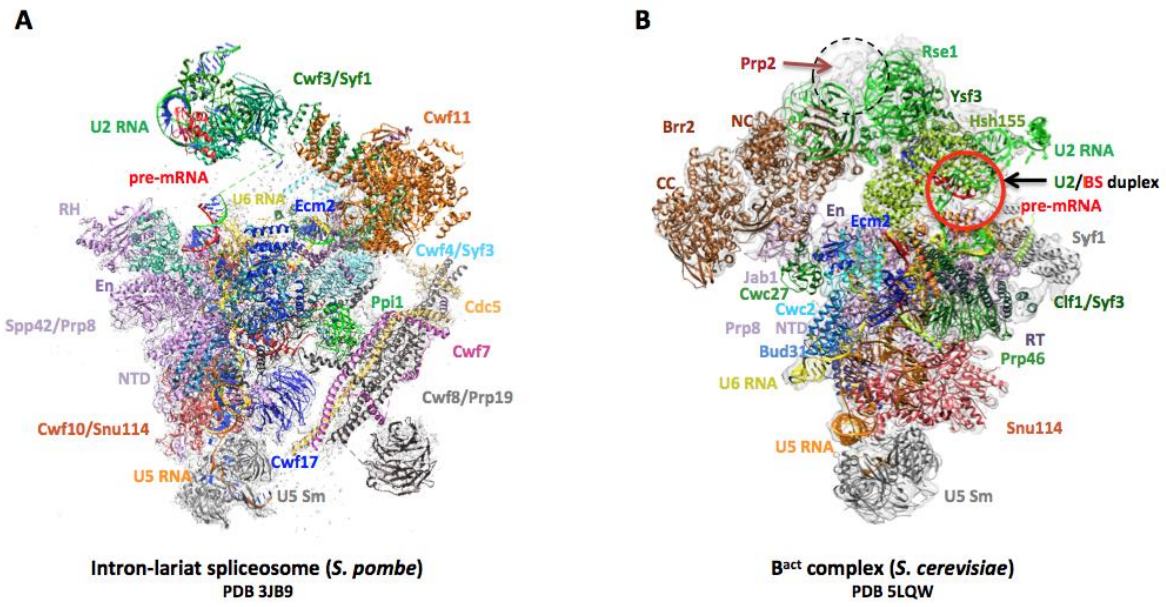


Figure 1.10: 3D cryo-EM structures of the intron-lariate spliceosome from S. pombe and B^{act} complex from S. cerevisiae

EM structures of *S. pombe* intron-lariat **(A)** and *S. cerevisiae* activated spliceosomes **(B)**, modified from Yan *et al.,* 2015, and Rauhut *et al.,* 2016, respectively. Position of Prp19 helical bundle in the ILS structure is shown on the left **(A)**. The spatial separation of the Prp2 RNA helicase (black, dashed circle) from the U2/BS RNA duplex (red circle) in the B^{act} spliceosome is indicated on the right **(B)**. Proteins and RNA components are colour-coded.

Most recently, the cryo-EM structures of earlier assembly intermediates of the catalytic cycle of the *S.cerevisiae* spliceosome – such as the B, C and C* complexes, as well as the human C* complex – have also been solved (Galej, Wilkinson et al. 2016, Rauhut, Fabrizio et al. 2016, Yan, Wan et al. 2016, Bertram, Agafonov et al. 2017, Fica, Oubridge et al. 2017, Plaschka, Lin et al. 2017, Yan, Wan et al. 2017, Zhang, Yan et al. 2017). Collectively, these structures revealed that the catalytic RNP core not only exhibits a very similar structure throughout the catalytic cycle of the yeast spliceosome, but that

its structure also reveals many similarities when the yeast and human C* complexes are compared. In sharp contrast to the largely stable organisation of U5 snRNA and the U2/U6 catalytic RNP core, the 3' part of U2 snRNA and its Sm core domain is highly mobile and undergoes large-scale movement during the catalytic cycle of the spliceosome (Bertram, Agafonov et al. 2017, Plaschka, Lin et al. 2017).

Furthermore, the cryo-EM structure of the yeast-activated spliceosome has shown that the catalytic U2/U6 RNP core domain is already largely established in the B^{act} complex. Moreover, the B^{act} cryo-EM model has provided the structural basis for the earlier biochemical finding that the activated spliceosome is still in a pre-catalytic state. That is, the first-step reactants, i.e. the 5'ss and the BS-A, were not only occluded by proteins, but in addition were spatially separated by about 5 nm, preventing premature first step catalysis. Surprisingly, the Prp2 RNA helicase was found to be positioned spatially separated from the U2/BS duplex by more than 7 nm, suggesting that Prp2 would act from a distance to liberate the first step reactants from their protein clamps, as a pre-requisite for catalytic activation of the B^{act} spliceosome (Rauhut, Fabrizio et al. 2016) (Figure 1.10).

.

2 Aims

During the last two years, high-resolution cryo-EM structures have been published for several assembly intermediates of the yeast spliceosome including the B, B^{act}, C, C* and intron-lariat complexes, providing completely new insight into the complex structure of the yeast spliceosome and its structural dynamics during the catalytic cycle. However, as of now, only the human C* complex has been investigated by cryo-EM. In collaboration with Dr. David Haselbach from the Department of Structural Dynamics, MPIbpc, I have used cryo-EM to investigate the 3D structure of the human B^{act} complex.

Human and yeast activated spliceosomes share a large number of conserved proteins but differ in their protein composition in several aspects. Thus, human B^{act} contains numerous proteins, which are conserved in higher eukaryotes but are absent in yeast, including numerous peptidyl-prolyl isomerases (PPILs) and the RNA helicase Aquarius (Aqr), which is required for catalytic activation of the human spliceosome in addition to Prp2. On the other hand, proteins conserved between yeast and human, such as Yju2 and Syf2, are missing from the human B^{act} spliceosome. Furthermore, conserved proteins of the Prp19/Cdc5L complex are differentially pre-organised in distinct sub-complexes in yeast and human cells, suggesting that the order of protein recruitment during spliceosome activation differs between the two organisms, which together raises the possibility that the 3D structure of the human and yeast B^{act} complexes may differ to some extent.

The B^{act} complex was assembled in HeLa nuclear extracts using a pre-mRNA construct, PM5-10, which contained the 5’ exon and an intron that is truncated 10 nts after the BS. In the first part of my work, I have used affinity-purified B^{act} complexes for the screening of buffer conditions in a high-throughput system termed ProteoPlex, to find biochemical conditions, which would stabilize the structure of the B^{act} spliceosome

After exhaustive 3D multi reference refinement (3D classification) of the human B^{act} particles, two major forms, termed A and B, of the B^{act} structure were obtained at resolutions of 5.3 Å and 8.1 Å, respectively. While their overall structure is largely similar, forms A and B differ with respect to the presence/absence of several densities.

The structure of the central domain of the human B^{act} complex including the catalytic U2/U6 RNP core is highly conserved between the human and yeast spliceosomes. This is also true for the architecture of the 5’ss and the U2/BS helix and their occlusion by homologous proteins to prevent pre-mature step 1 catalysis of the splicing reaction. However, significant structural differences were observed within the peripheral domains of the human B^{act} structure, including a unique conformation

of the U2/U6 helix II. The data are consistent with the possibility that the U2/U6 helix II may have to rearrange prior to or during catalytic activation to allow the binding of the conserved Syf2 protein close to the rearranged RNA helix. Thus, the catalytic activation process of the human B^{act} spliceosome may be more complex and more regulated than in yeast.

Finally, while the main forms of the hB^{act} structure obtained by 3D classification share most of the structural features, they differ in respect of the presence/absence of densities for several protein domains. Experimental evidence further indicates that the absence of a protein density in one of the forms is not due to the physical absence of the respective proteins. Instead, the differences indicate conformational flexibility of the protein domains. Further evidence indicates that form A is a precursor of form B of the B^{act} structure. Thus, it was possible to capture by cryo-EM analysis two conformational states of the human B^{act} complex that differ in their degree of conformational maturation towards the catalytic activation step.

3 Materials and methods

3.1 Materials

3.1.1 Chemicals

Acetic acid	Merck, Germany
Acetobutyrate cellulose in ethyl acetate (Triafol)	Sigma-Aldrich, Germany
Adenosine triphosphate (ATP)	Amersham Biosciences
Alkylbenzyldimethylammonium chloride (Osvan solution)	Sigma-Aldrich, Germany
Agarose (low melting point)	Invitrogen, Netherlands
Agarose	Invitrogen, Netherlands
Ammonium peroxidisulphate (APS)	Merck, Germany
Ampicillin	Sigma-Aldrich, Germany
Bicine (N, N-Bis(2-hydroxyethyl)glycine)	Sigma-Aldrich, Germany
BisTris (Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane)	Sigma-Aldrich, Germany
Boric acid	Merck, Germany
Bromphenol blue	Merck, Germany
BS3 (Bis[sulfosuccinimidyl]suberate)	Thermo Fisher Scientific, USA
Coomassie brilliant blue G-250	Serva, Germany
Creatine phosphate	Sigma-Aldrich, Germany
Dimethylsulphoxide (DMSO)	Roth, Germany
di-Potassium hydrogen phosphate	Merck, Germany
di-Sodium hydrogen phosphate	Merck, Germany
Dithiothreitol (DTT)	Roth, Germany
DNA molecular weight marker	Gibco, New Zealand
EDTA (Disodium salt dihydrate)	Roth, Germany
Ethanol	Merck, Germany
Ethidium bromide	Roche, Germany
Formamide	Merck, Germany
Glutaraldehyde	Electron Microscopy Sciences, USA
Glycerol	Merck, Germany
Glycine	Merck, Germany

Glycoblue	Ambion, USA
Heparin (sodium salt)	Roth, Germany
HEPES (N-2-Hydroxyethylpiperazin-N-2-ethanesulfonic acid)	Calbiochem, USA
Hydrochloric acid (HCl)	Merck, Germany
Isopropanol	Merck, Germany
IPTG	Merck, Germany
LB-Agar	Q-Bio-gene, USA
LB-liquid media	Q-Bio-gene, USA
Maltose	Merck, Germany
Magnesium acetate	Merck, Germany
Magnesium chloride	Merck, Germany
MES (2-(n-morpholino) ethanesulfonic acid)	Roth, Germany
Methanol	Merck, Germany
β-Mercaptoethanol	Roth, Germany
MOPS	Invitrogen, Netherlands
PMSF (Phenylmethylsulfonyl fluoride)	Roche, Germany
Potassium acetate	Merck, Germany
Potassium chloride	Merck, Germany
Potassium dihydrogen phosphate	Merck, Germany
Pre-stained protein-molecular weight marker	Bio-Rad, Germany
Protease inhibitor tablet, EDTA free	Roche, Germany
Phenol-chloroform-isoamyl (25:24:1) alcohol (PCI)	Roth, Germany
Rotiphorese gel 30 solution	Roth, Germany
Rotiphorese gel 40 solution	Roth, Germany
Sodium acetate	Merck, Germany
Sodium chloride	Merck, Germany
Sodium dodecyl sulfate (SDS)	Serva, Germany
snRNA marker	In-house
Sucrose	Merck, Germany
Spermidine	Sigma-Aldrich, Germany
SYBR Gold Nucleic Acid Gel Stain	Thermo Fisher Scientific, Germany
SYPRO orange	Life Technologies
TERGITOL (Type NP-40) detergent	Sigma-Aldrich, Germany

TEMED (N, N, N', N'-Tetramethylethylendiamine)	Sigma-Aldrich, Germany
Tris [Tris-(hydroxymethyl)aminomethane]	VWR International, Germany
Tween 20	Sigma-Aldrich, Germany
Uranyl formate	In-house
Urea	Merck, Germany
Xylene cyanol	Fluka, Switzerland

3.1.2 Laboratory materials and consumables

96 well Plates	Bio-Rad, Germany
Amicon Ultra 0.5 ml, 15 ml (MWCO 50 and 100 kD)	Merck, Germany
Amylose resin	New England Biolabs, Germany
Carbon rods, highest grade	Ringsdorff Werke GmbH, Bonn
Centrifuge tubes	Beranek, Germany
Copper EM grids, 200 mesh square fine bar	Science Services, Munich, Germany
Copper EM grids Quantifoil R2/2 and R3.5/1, Cu, 200 mesh	Quantifoil Micro Tools GmbH, Jena
MBPTrap HP Column (1 and 5 ml)	GE Healthcare, Germany
HiTrap Heparin HP Column (1ml)	GE Healthcare, Germany
Custom made teflon blocks	in-house
Dialysis membranes (MWCO 6000-8000 Da)	SpektraPor, USA
Ethane (liquid)	Messer, Sulzbach, Germany
Filter paper, 90 mm	Whatman (GE Healthcare)
Mica, 75x25 mm	Plano, Wetzlar, Germany
Needles BD Spinal 18GA 3.50 IN 1.2 x 90 mm	Becton Dickinson, Spain
Nitrocellulose, 0.2 µm	Whatman (GE Healthcare)
Nitrogen (liquid)	Air Liquide, Paris, France
NuPAGE™ gels (1.5 mm, 4-12%)	Invitrogen, Netherlands
Parafilm	Roth, Germany
Pipettes	Eppendorf, Germany
Pipette tips (1.0, 0.2, 0.01 ml)	Roth, Germany
Poly-Prep columns	Bio-Rad, USA
ProbeQuant™ G-25/ G-50 micro columns	GE Healthcare, UK
Slide-A-Lyzer dialysis units (MWCO 6 kDa)	Pierce, USA
Sterile filters (0.2 µm or 0.45 µm)	Sarstedt, Germany

Microfuge tubes (0.5 ml, 1.5 ml and 2.0 ml)	Eppendorf, Germany
Microfuge tubes (15 ml und 50 ml)	Sarstedt, Germany

3.1.3 Commerical kits

BCA protein assay kit	Pierce, USA
Bradford Assay	Bio-Rad, Germany
NuPAGE™ gels (1.5 mm, 4-12%)	Invitrogen, Netherlands
pHClear Screen	Qiagen, Germany
QIAGEN Plasmid Mini / Maxi Kits	Qiagen, Germany

3.1.4 Machines

Autoclaves	H+P Labortechnik, Germany
Balances	Sartorius, Germany
Centrifuge Biofuge fresco	Heraeus, Germany
Centrifuge Biofuge pico	Heraeus, Germany
Centrifuge Megafuge 1.0R	Heraeus, Germany
Cryo electron microscope CM200 FEG	Philips, Netherlands
Cryo electron microscope Titan Krios	FEI, Eindhoven, Netherlands
Falcon 3EC Direct Electron Detector	Thermo Fisher Scientific, Germany
FiberLite Fixed Angle Rotor F14-14 × 50	Thermo Fisher Scientific, Germany
FiberLite Fixed Angle Rotor F14-6 × 250	Thermo Fisher Scientific, Germany
Fractionator LKB-FRAC-100	Pharmacia Biotech, Sweden
Fujifilm FLA-7000 laser scanner	FUJIFILM Life Science, USA
Gel documentation unit	Bio-Rad, Germany
Gel electrophoresis apparatus	In-house
Gel dryer model 583	Bio-Rad, Germany
Glassware	VWR International, Germany
Gradient master model 106	BioComp Instruments, Canada
Head-over-tail rotor	Cole-Parmer, USA
Heating blocks	Eppendorf, Germany
Ice machine	Ziegra, UK
Incubators	Heraeus, Germany
LTQ-Orbitrap Velos	Thermo Fisher Scientific, Germany
Magnetic Stirrer	IKA Staufen, Germany

Microwave oven	Bosch, USA
Milli-Q-water supply apparatus	Millipore, USA
Orbitrap Fusion Lumos Tribrid Mass Spectrometer	Thermo Fisher Scientific, Germany
pH-Meter	Thermo Fisher Scientific, Germany
Peristaltic Pump P-1	Pharmacia Biotech, Sweden
Phosphorimager screens	GE Healthcare, Germany
Phosphorimager Typhoon Trio+	GE Healthcare, Germany
Pipetting Robot Microlab Star LET	Hamilton, USA
Power supply EPS 2A200	Hoefer Pharmacia Biotech, USA
Power supply EPS 3501/XL	Amersham Pharmacia, Germany
Rocking platform MAX Q3000 and Q2000	Thermo Fisher Scientific, Germany
Room temperature EM specimen holder	Philips, Eindhoven, Netherlands
Q Exactive HF Hybrid Quadrupol-Orbitrap Mass Spectrometer	Thermo Fisher Scientific, Germany
Scintillation counter LS 1701	Packard, USA
Sorvall HB-6 rotor	Kendro, USA
Sorvall SA800 AT4 rotor	Kendro, USA
Sorvall SS-34 rotor	Kendro, USA
Sorvall TH660 rotor	Kendro, USA
Sorvall TST41.14 rotor	Kendro, USA
Speed Vac Konzentrator 5301	Eppendorf, Germany
Spectrophotometer Nanodrop ND-1000	Thermo Fisher Scientific, Germany
Spectrophotometer Ultrospec 3000 pro	Amersham Pharmacia, Germany
SureSpin630 Rotor	Thermo Scientific, Germany
Tabletop centrifuge 5415D	Eppendorf, Germany
TemCam F415 (slow scan 4K x 4K CCD camera)	Tietz Video Systems, Germany
Thermomixer	Eppendorf, Germany
Ultracentrifuge Discovery 90 SE	Sorvall/Kendro, USA
Ultracentrifuge Discovery M150	Sorvall/Kendro, USA
Ultracentrifuge Sorvall Evolution RC Lynx 6000	Thermo Scientific, USA
Ultracentrifuge Sorvall WX ultra 80	Thermo Scientific, USA
Vitrobot Mark IV	FEI, Eindhoven, Netherlands
Vortex	Janke & Kunkel, Germany

3.1.5 Nucleotides

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^7G(5')ppp(5')G-cap)	Kedar, Poland

3.1.6 Radiolabeled nucleotides

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

3.1.7 DNA oligonucleotides

DNA-oligonucleotides RNase H digestions were purchased from IBA, Göttingen, Germany

Name	Sequence 5' → 3'
M6	GGCGGTCTCGTC
M12	CTCGTCGGCAGC

3.1.8 Enzymes

Complete protease inhibitor, EDTA-free	Roche, Germany
Proteinase K	Fluka, Switzerland
Restriction endonucleases	New England Biolabs, Germany
RNasin (RNase inhibitor) [40 U/µl]	Promega, USA
RQ1 DNase [1 U/µl]	Promega, USA
SP6 RNA polymerase	New England Biolabs, Germany
T7 RNA polymerase [20 U/µl]	New England Biolabs, Germany
Sequencing grade trypsin, porcine	Promega, Germany
Yeast inorganic pyrophosphatase (YIPP) [0.1 U/µl]	New England Biolabs, Germany

3.1.9 Bacteria strains

E. coli DH5α	Novagen, Germany

3.1.10 Cell line

HeLa S3 cells	GBF, Germany
(Human adenocarcinoma cell line)	

3.1.11 Plasmids

pMINX	U2-dependent pre-mRNA construct (MINX) recloned from pSP65-vector under control of SP6 promoter (Zillmann, Zapp et al. 1988) in pUC18-vector under control of T7 promoter, Amp^R.
pMS2-MBP	MS2-MBP fusion protein in pMAL vector (NEB), Amp^R.
pPM5-10	U2-dependent pre-mRNA construct (pPM5-10) in pUC18-vector under control of SP6 promoter, Amp^R (Bessonov, Anokhina et al. 2010).

3.1.12 Buffers

Commonly used media, buffers and solutions were prepared with deionized water (Millipore) and autoclaved if necessary (121 °C, 20 min, 1 bar). Solutions with heat-labile components were filter-sterilized (0.22 µm).

Coomassie staining solution:	100 µM Coomassie G-250
	0.13 mM (v/v) HCl
6x DNA loading dye:	60 % (v/v) glycerol
	10 mM Tris-HCl, pH 7.5
	60 mM EDTA, pH 8.0
	0.05% (w/v) bromphenol blue
	0.05% (w/v) xylene cyanol
10x G-150 buffer:	200 mM HEPES
	1.5 M KCl
	15 mM $MgCl_2$
	Adjust to pH 7.9

Fixation buffer:	40% (v/v) Methanol 10% (v/v) Acetic acid
1x MBP150 buffer	20 mM HEPES-KOH, pH 7.9 150 mM NaCl 0.05% (v/v) NP-40
1x MC buffer:	10 mM HEPES-KOH, pH 7.6 10 mM KOAc 0.5 mM MgOAc 5 mM DTT
5x Native gel loading dye:	1x TBE 30% (v/v) glycerol 0.05% (w/v) bromphenol blue
10X PBS (pH 7.4):	1.3 M NaCl 160 mM K_2HPO_4 40 mM KH_2PO_4
10X PBS (pH 8.0):	1.3 M NaCl 188 mM K_2HPO_4 12 mM KH_2PO_4
2x PK buffer:	200 mM Tris-HCl, pH 7.5 25 mM EDTA, pH 8.0 2% (w/v) SDS
2x RNA loading dye:	80% (v/v) formamide 1 mM EDTA, pH 8.0 0.05% (w/v) bromphenol blue 0.05% (w/v) xylene cyanol

1x Roeder C buffer:	25% (v/v) glycerol
	20 mM HEPES-KOH, pH 7.9
	420 mM NaCl
	1.5 mM $MgCl_2$
	0.2 mM EDTA, pH 8.0
	0.5 mM DTT
	0.5 mM PMSF
1x 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
	0.5 mM DTT
	0.5 mM PMSF
4x SDS loading dye:	40% glycerol
	400 mM β-mercaptoethanol
	200 mM Tris-HCl, pH 6.8
	4% (w/v) SDS
	0.05% (w/v) Xylene cyanol
	0.12% (w/v) Bromphenol blue
5X SP6 / T7 transcription buffer:	600 mM HEPES-KOH, pH 7.5
	160 mM MgCl2
	10 mM spermidine
	200 mM DTT
10x TBE:	0.89 M Tris
	0.89 M boric acid
	25 mM EDTA, pH 8.0

10X TBS buffer:	200 mM Tris-HCl, pH 7.6
	1.5 M NaCl
	1% Tween 20

3.1.13 Software programs

Program	Source
Coot	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/ (Emsley, Lohkamp et al. 2010)
coweyes	Custom made (Department of Structural Dynamics, MPI-BPC)
Eman 2	http://blake.bcm.edu/emanwiki/EMAN2 (Tang, Peng et al. 2007)
Gautomatch	http://www.mrc-lmb.cam.ac.uk/kzhang/
Gctf	http://www.mrc-lmb.cam.ac.uk/kzhang/ (Zhang 2016)
ImageJ	http://imagej.nih.gov/ij/ (Schneider, Rasband et al. 2012)
johnHenry	Custom made (Department of Structural Dynamics, MPI-BPC)
Relion 1.4	http://www2.mrc-lmb.cam.ac.uk/relion (Scheres 2012)
ResMap	http://resmap.sourceforge.net/ (Kucukelbir, Sigworth et al. 2014)
UCSF Chimera	http://www.cgl.ucsf.edu/chimera/ (Pettersen, Goddard et al. 2004)
unblur software suite	http://grigoriefflab.janelia.org/unblur (Campbell, Cheng et al. 2012)

3.2 Methods

3.2.1 Molecular biology standard methods

3.2.1.1 Transformation of DH5α cells

For preparation of competent cells according to the $CaCl_2$ method, a pre-culture of DH5α was inoculated in 5 ml Luria Bertani (LB) medium and incubated at 37°C overnight. The pre-culture was used to inoculate 50 ml LB. The culture was incubated at 37°C until it reached 0.3-0.5 OD/ml, transferred into a 50 ml Falcon tube, and centrifuged for 10 min at 2000 rpm in Megafuge 1.0R (Heraeus). Cells were resuspended in 25 ml cold 50 mM $CaCl_2$ and centrifuged again. After resuspension in 3 ml of cold 50mM $CaCl_2$, sterile glycerol was added to 10%, and the cells were kept at -80°C in 300 µl aliquots. Transformation was performed as described (Sambrook, Fritsch et al. 1989): 10-50 ng of plasmid DNA were mixed with 50 µl of competent cells and incubated on ice for 30 min. Heat-shock was performed for 1 min at 42°C and cells were immediately chilled on ice for 1 min. Subsequently, 800 µl of LB- medium were added and incubated for 1 h at 37°C. Cells were centrifuged briefly in a microfuge and the cell pellet was resuspended in 50 µl of 10 mM Tris/HCl, pH 7.5. Cells were plated on the selection media and incubated overnight at 37°C. The isolation of the plasmid DNA from the transformed strain was carried out using QIAGEN Plasmid Mini and Maxi Kits according to the manufacturer's instructions.

3.2.1.2 Generation of templates for run-off in vitro transcriptions

DNA plasmids encoding the MINX or PM5-10 pre-mRNA were extracted from *E. coli* DH5α cells using the QIAGEN Plasmid Maxi Kit (Qiagen) according to the manufacturer's protocol. Plasmid DNA was eluted in mQ H_2O and linearized by restriction digestion to generate templates for run-off *in vitro* transcriptions. The reactions were performed with the respective restriction endonuclease (NEB) according to the instructions of the manufacturer. Briefly, 10 units of enzyme were used to cut 1 µg of plasmid DNA by incubating for 1 h at 37 °C.

Agarose gel-electrophoresis was then performed to ensure linearization of the DNA after restriction. Samples were supplemented with 6x DNA loading dye. Gels were prepared using 1% (w/v) agarose and 0.4 µg/ml ethidium bromide in 1x TBE buffer. Gels were run at 100 V for approximately 1 h.

3.2.1.3 In vitro transcription

RNA *in vitro* transcription was carried out using DNA templates derived from linearized plasmids (3.2.1.2). Uniformly ^{32}P-labeled m^{7}G(5')ppp(5')G-capped pre-mRNA was synthesized *in vitro* by T7 (MINX) or SP6 (PM5-10) run-off transcription. The composition of a standard reaction is shown in Table 1:

Table 1: Standard in vitro transcription reaction

	Volume (µl)	Final concentration
5X T7 (SP6) transcription buffer	20	1X
0.1 M ATP	7.5	7.5 mM
0.1 M CTP	7.5	7.5 mM
0.1 M UTP	1.5	1.5 mM
0.1 M GTP	1.5	1.5 mM
0.15 M m7G(5')ppp(5')G-cap	3.3	5 mM
RNase inhibitor (40 U/µl)	3.7	1.5 U/ml
SP6 (T7) RNA polymerase (20 U/µl)	10	2.0 U/ml
DNA template (250 ng/µl)	10	25 ng/ml
α-[32P]-labelled UTP (10 mCi/ml, 3000 Ci/mmol)	5	
H2O	30	
Total	100	

The transcription was carried out at 37 °C for 3 h. The DNA template was digested by addition of 2 µl RQ1 DNase (1U/µl, NEB) and further incubation at 37 °C for 30 min. In order to purify the pre-mRNA from unincorporated NTPs, the synthesized RNA was passed through a ProbeQuant G-50 column (GE Healthcare) according to the manufacturer's protocol, followed by ethanol precipitation (3.2.1.5).

3.2.1.4 PCI extraction

Phenol-chloroform-isoamyl alcohol (PCI) extraction was used to separate nucleic acids from proteins. During extraction, proteins remain in the organic phase, while nucleic acids stay in the aqueous phase. Reactions were first adjusted to 200 µl final volume if not stated otherwise. Subsequently an equal volume of PCI was added and samples were thoroughly mixed on a shaker for 10 min at room temperature. The resulting suspension was centrifuged for 10 min at room temperature at 13000 rpm in a microfuge in order to separate the aqueous from the organic phase. Nucleic acids in the aqueous phase were precipitated in ethanol (3.2.1.5).

3.2.1.5 Ethanol Precipitation

Nucleic acids in aqueous solutions were precipitated by addition of 0.1 volume 3 M NaOAC, pH 4.7 and 3 volumes absolute ethanol. Samples were kept at -20 °C for at least 30 min and precipitated by centrifugation for 30 min at 4 °C and 13000 rpm in a microfuge. The pellet was washed once with 70% (v/v) ethanol, vacuum dried and finally resuspended in the desired solution.

3.2.1.6 Nucleic acid quantification

The concentration of nucleic acids was determined by measuring their absorption in an aqueous solution at a wavelength of 260 nm and 280 nm in comparison to the corresponding buffer without

nucleic acids. The purity of a nucleic acid solution was determined by the ratio of OD260/OD280: 2.0 corresponded to pure RNA or oligonucleotides and 1.8 to pure DNA. Lower ratios idnicate contamination with proteins or phenol, which requires an additional PCI extraction (3.2.1.4).

1 OD_{260} was considered to be equal to:

- 50 µg/ml of double-stranded DNA
- 33 µg/ml of single-stranded DNA
- 40 µg/ml of single-stranded RNA

3.2.1.7 Proteinase K digestion

To facilitate RNA recovery from protein rich samples, proteinase K treatment was performed prior to PCI extraction. Samples were incubated in 1x PK buffer in the presence of 0.2 µg/µl proteinase K for 15 min at 45 °C. The RNA was recovered by PCI extraction (3.2.1.4) followed by ethanol precipitation (3.2.1.5).

3.2.1.8 Denaturing polyacrylamide gel-electrophoresis

Denaturing polyacrylamide gel electrophoresis (PAGE) in the presence of 8 M urea was mainly used to separate snRNAs or intermediates and products of pre-mRNA *in vitro* splicing. Depending on the size of the RNAs, the gels contained 10-14% (v/v) polyacrylamide, 8 M urea and 1x TBE. Polymerization of the gels was initiated by addition of 300 µl 10% (w/v) APS and 30 µl TEMED to 50 ml of gel solution. The RNA samples were dissolved in RNA loading dye, denatured by incubating for 5 min at 96 °C and briefly chilled on ice before loading. The electrophoresis was performed at 15 W in 1x TBE buffer. The RNA was then visualized by SYBR Gold staining or autoradiography. Alternatively RNA was separated on a denaturing 4-12% NuPAGE gel (Invitrogen), run in 1x MOPS buffer (Invitrogen) according to the manufacturer's protocol, and then visualized by SYBR Gold staining or autoradiography.

3.2.1.9 SYBR Gold staining

SYBR Gold staining of RNA was performed after fixation of SDS denaturing gels in 40% (v/v) methanol / 10% (v/v) acetic acid. RNA was stained according to manufacturer's protocol and visualized with Fujifilm FLA-7000 laser scanner at a wavelength of 495 nm.

3.2.2 Protein-biochemistry standard methods

3.2.2.1 Protein quantification

To measure the concentration of proteins, a BCA™ Protein Assay Kit (Pierce) was used according to the manufacturer's protocol. This assay is a detergent-compatible method based on bicinchoninic acid (BCA) for the colorimetric detection. Quantification of total protein was analysed at an absorption maximum of 562 nm.

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

Proteins were resolved on denaturing 4-12% NuPAGE gels (Invitrogen), run in 1x MOPS buffer (Invitrogen) according to the manufacturer's protocol.

3.2.2.3 Coomassie staining

Coomassie staining of proteins was performed after fixation of SDS denaturing gels in 40% (v/v) methanol / 10 % (v/v) acetic acid. Proteins were stained with coomassie staining solution overnight at room temperature and were subsequently destained in water until protein bands were clearly visible.

3.2.2.4 Purification of MS2-MBP

The fusion protein, which was used for affinity-purification of spliceosomal complexes, was expressed in *E. coli* containing the plasmid encoding the MS2-MBP protein. Cells were grown to an OD_{600} of 0.3-0.6 in LB medium supplemented with ampicillin. The expression of the MS2-MBP protein was then induced by adding 0.1 M IPTG. The culture was further incubated for 2.5 h at 37 °C. The cells were pelleted by centrifugation in a Megafuge 1.0R (Heraeus) for 10 min at 2000 rpm. The supernatant was discarded and the cell pellet was washed with 40 ml of washing buffer (20 mM Tris-HCl, pH 7.6, 0.2 M NaCl) supplemented with one complete EDTA-free protease inhibitor cocktail tablet. Then, the cells were resuspended in 25 ml of washing buffer and subsequently lysed by sonification. The lysate was centrifuged in a FiberLite F14-14 × 50 rotor for 30 min at 16000 g. The supernatant was transferred into a MBPTrap HP column (GE Healthcare) and the amylose resin was washed with 10 column volumes of MBP150 buffer followed by 5 column volumes of 5 mM Na_2HPO_4, pH 7.0. The MS2-MBP protein was then eluted with 5 mM Na_2HPO_4, pH 7.0 containing 20 mM maltose and the eluate was subsequently loaded onto a pre-equilibrated HiTrap Heparin HP column (GE Healthcare). The column was washed with 10 volumes of 5 mM Na_2HPO_4, pH 7.0 and the protein was eluted fraction-wise with elution buffer (20 mM HEPES-KOH, pH 7.9, 100 mM KCl, 15 % (v/v) glycerol, 0.5 mM DTT). Peak fractions were pooled, shock frozen in liquid nitrogen and finally stored at -80 °C. The protein concentration was determined using the BCA™ Protein Assay Kit (3.2.2.1).

3.2.3 Special methods

3.2.3.1 Cell culture

HeLa S3 cells (Computer Cell Culture Center, Belgium) were grown in suspension in S-MEM media supplemented with 5% (v/v) newborn calf serum, 50 µg/ml penicillin and 100 µg/ml streptomycin to a density of 2.5-5x 10^5 cells/ml. Cultivation and harvesting of the cells was essentially performed as previously described in (Kastner 1998).

3.2.3.2 *Preparation of splicing active HeLa nuclear extract*

Splicing active nuclear extract was prepared from HeLa cells as described in (Dignam, Lebovitz et al. 1983). Six to eight liters of HeLa cells were grown to a density of 2.5-5 x 10^5 cells/ml (Kastner 1998). The cells were pelleted by centrifugation in a Megafuge 1.0R (Heraeus) for 10 min at 2000 rpm. The supernatant was discarded and the cells were washed three times with ice-cold 1x PBS, pH 7.4. Then, the cell pellet was resuspended in 1.25 volumes of 1x MC buffer supplemented with two complete EDTA-free protease inhibitor cocktail tablets per 50 ml of buffer. After an incubation of 5 min on ice, the cells were lysed with 18 strokes of a Dounce homogenizer at 4 °C. The mixture was transferred to 50 ml tubes and the nuclei were pelleted by centrifugation in a FiberLite F14-14 × 50 rotor for 5 min at 13000 g. Afterwards the supernatant was discarded and nuclei were treated in one of the following ways:

- According to the general procedure, the nuclei were resuspended in 1.3 volumes of 1x Roeder C buffer containing 0.5 mM DTE and 0.5 mM PMSF. For lysis of the nuclei the suspension was again subjected to 20 strokes of a Dounce homogenizer at 4 °C. The lysate was stirred for 40 min at 4 °C, followed by centrifugation in a FiberLite F14-14 × 50 rotor at 16000 rpm for 30 min to pellet particulate material. The supernatant was recovered and dialyzed against 50 volumes of 1x Roeder D buffer at 4 °C for 5 h with one change of buffer after 2.5 h. The dialyzed nuclear extract was recovered and centrifuged in a FiberLite F14-14 × 50 rotor at 10000 rpm for 10 min at 4 °C. The supernatant was aliquoted, frozen in liquid nitrogen and stored at -80 °C.
- According to the optimized for cryo-EM procedure, the nuclei were resuspended in 1.3 volumes of 1x Roeder D or 1x Roeder C buffer and homogenized with 20 strokes of a Dounce homogenizer at 4 °C. The lysate was stirred for 40 min at 4 °C, followed by centrifugation in a FiberLite F14-14 × 50 rotor at 16000 rpm for 30 min to pellet particulate material, if not stated otherwise. The supernatant was immediately used in an *in vitro* splicing reaction.

3.2.3.3 *In vitro splicing reactions*

A typical splicing reaction was carried out in the presence of 10 nM pre-mRNA and 40 % (v/v) HeLa nuclear extract in a buffer containing 20 mM HEPES-KOH, pH 7.9, 3 mM $MgCl_2$, 65 mM KCl, 2 mM ATP and 20 mM creatine phosphate. Alternatively, splicing was carried out in the presence of 5 nM pre-mRNA and 20% (v/v) HeLa nuclear extract (optimized procedure) in a splicing mixture containing 20 mM HEPES-KOH, pH 7.9, 3 mM $MgCl_2$, 65 mM KCl or 50 mM NaCl, 2 mM ATP and 20 mM creatine phosphate. Spliceosomal complexes were allowed to form by incubating at 30 °C. B^{act} complexes were assembled on PM5-10 pre-mRNA by incubating at 30 °C for 0-3 h. A 30-fold molar excess of DNA oligonucleotides complementray to nucleotides -6 to -18 and -12 to -24 relative to the 5'ss (M6 and M12 oligos, respectively) was added and the reaction was incubated at 30°C for an additional 20 min. After splicing, the reactions were chilled on ice.

3.2.3.4 Analysis of splicing complexes by native agarose gel-electrophoresis

Spliceosomal complexes A, B and C are formed in sequentially and show decreased mobility on a native agarose gel (Lamond, Konarska et al. 1987, Kent and MacMillan 2002). To resolve spliceosomal complexes, 20 µl splicing reactions were incubated at 30 °C for different time points. Heparin was added to 18 µl aliquots of the splicing reaction to a final concentration of 0.125 mg/ml and the mixtures were incubated for another 1 min at 30°C before addition of 2 µl of 4X agarose loading buffer. The samples were loaded onto a native agarose gel (26 cm x 15 cm) containing 1.5% (w/v) of low melting point agarose (Invitrogen) in 0.5 x TBE buffer. Electrophoresis was performed by applying 100 V for 5 h at room temperature. The gels were dried at 60 °C for 5 h and bands were visualized by exposing to a phosphoimager screen.

3.2.3.5 MS2 affinity-selection of splicing complexes

Spliceosomal complexes were isolated by MS2 affinity-selection as previously described (Bessonov, Anokhina et al. 2010). Pre-mRNA was incubated with a 20-fold molar excess of purified MS2-MBP fusion protein for 30 min at 4 °C prior to splicing. After performing in vitro splicing the reaction was equilibrated with HEPES-KOH, pH 7.9 and then loaded onto a MBPTrap HP column (GE Healthcare) equilibrated with 1x G-150 buffer (20 mM HEPES-KOH, pH 7.9, 1.5 mM $MgCl_2$, 150 mM NaCl). The column was washed with 40 volumes of 1x G-150 buffer and complexes were eluted with 1x G-150 buffer containing 1 mM maltose. Eluted complexes were loaded onto a 36 ml linear 10-30% (v/v) glycerol gradient containing G-150 buffer (20 mM HEPES- KOH, pH 7.9, 1.5 mM $MgCl_2$, 150 mM NaCl), centrifuged at 23000 rpm for 15 h at 4 °C in a Surespin 630 (Thermo Scientific) rotor and gradient fractions were harvested from the bottom. The distribution of ^{32}P-labeled pre-mRNA across the gradient was determined by Cherenkov counting. Fractions were analysed on denaturing 4-12% NuPAGE gels (Invitrogen), followed by autoradiography or fixation and staining (3.2.1.9 and 3.2.2.3). Peak fractions containing the activated spliceosomal complexes were pooled, concentrated by centrifugation with an Amicon Ultra concentration unit (Merck), diluted to decrease the glycerol concentration and reloaded on a second gradient with or without glutaraldehyde as fixative (for EM analysis and for biochemical sample validation, respectively). The RNA and protein compositions of purified complexes were analysed on denaturing 4-12% NuPAGE gels (Invitrogen).

3.2.3.6 ProteoPlex

A ProteoPlex screen, to identify the optimal buffer conditions for spliceosomal complex stability, was performed in 20 µl reactions in a 96-well plate (Bio-Rad) (Chari, Haselbach et al. 2015). Each individual 20 µl reaction was composed of 16 µl of the isolated complex at a concentration of 0.1 - 1 µM, 2 µl of SYPRO orange (Life Technologies) at a final concentration of 1x and 2 µL of the buffer/chemical under investigation. Buffer substance screens at various pH values were performed with a custom made buffer screen, where individual stock solutions were 1 M (Table 2). Each plate

contained control wells, which contained the purification buffer of the spliceosome complex and wells without complex. Unfolding transitions were recorded in a CFX connect real-time PCR machine (Bio-Rad). A reader program was utilized, where the entire plate was equilibrated to 30°C for 2 min followed by fluorescence measurement. Subsequently, the samples were incrementally heated in 1°C steps, equilibrated for 30 s and fluorescence was measured. Finally, the sample was held at 95°C for 2 min, followed by a final fluorescence measurement.

Table 2: Composition of custom made buffer screen

All individual stock solutions were 1 M. Column 12 is intentionally left blank for controls

	1	2	3	4	5	6	7	8	9	10	11	12
A	$SPG_{1)}$ pH 5.6	$MMT_{2)}$ pH 5.6	$PCB_{3)}$ pH 5.6	$MES_{4)}$ pH 5.5	Na Citrate pH 5.5	$BisTris_{5)}$ pH 5.8	Na/K Phosphate pH 5.8	$ADA_{6)}$ pH 6.0	$HEPES_{7)}$ pH 6.8	Imidazole pH 6.2	$Tris_{8)}$ pH 7.5	
B	$SPG_{1)}$ pH 6.0	$MMT_{2)}$ pH 6.0	$PCB_{3)}$ pH 5.0	$MES_{4)}$ pH 5.7	Na Citrate pH 5.75	$BisTris_{5)}$ pH 6.0	Na/K Phosphate pH 6.1	$ADA_{6)}$ pH 6.2	$HEPES_{7)}$ pH 7.0	Imidazole pH 6.45	$Tris_{8)}$ pH 7.7	
C	$SPG_{1)}$ pH 6.4	$MMT_{2)}$ pH 6.4	$PCB_{3)}$ pH 6.4	$MES_{4)}$ pH 5.9	Na Citrate pH 6.0	$BisTris_{5)}$ pH 6.2	Na/K Phosphate pH 6.4	$ADA_{6)}$ pH 6.4	$HEPES_{7)}$ pH 7.2	Imidazole pH 6.7	$Tris_{8)}$ pH 7.9	
D	$SPG_{1)}$ pH 6.8	$MMT_{2)}$ pH 6.8	$PCB_{3)}$ pH 6.8	$MES_{4)}$ pH 6.1	Na Citrate pH 6.25	$BisTris_{5)}$ pH 6.4	Na/K Phosphate pH 6.7	$ADA_{6)}$ pH 6.6	$HEPES_{7)}$ pH 7.4	Imidazole pH 6.95	$Tris_{8)}$ pH 8.1	
E	$SPG_{1)}$ pH 7.2	$MMT_{2)}$ pH 7.2	$PCB_{3)}$ pH 7.2	$MES_{4)}$ pH 6.3	Na Citrate pH 6.5	$BisTris_{5)}$ pH 6.6	Na/K Phosphate pH 7.0	$ADA_{6)}$ pH 6.8	$HEPES_{7)}$ pH 7.6	Imidazole pH 7.2	$Tris_{8)}$ pH 8.3	
F	$SPG_{1)}$ pH 7.6	MMT_{2} pH 7.6	$PCB_{3)}$ pH 7.6	$MES_{4)}$ pH 6.5	Na Citrate pH 6.75	$BisTris_{5)}$ pH 6.8	Na/K Phosphate pH 7.3	$ADA_{6)}$ pH 7.0	$HEPES_{7)}$ pH 7.8	Imidazole pH 7.45	$Tris_{8)}$ pH 8.5	
G	$SPG_{1)}$ pH 8.0	$MMT_{2)}$ pH 8.0	$PCB_{3)}$ pH 8.0	$MES_{4)}$ pH 6.7	Na Citrate pH 7.0	$BisTris_{5)}$ pH 7.0	Na/K Phosphate pH 7.6	$ADA_{6)}$ pH 7.2	$HEPES_{7)}$ pH 8.0	Imidazole pH 7.7	$Tris_{8)}$ pH 8.8	
H	$SPG_{1)}$ pH 8.4	$MMT_{2)}$ pH 8.4	$PCB_{3)}$ pH 8.4	$MES_{4)}$ pH 6.9	Na Citrate pH 7.25	$BisTris_{5)}$ pH 7.2	Na/K Phosphate pH 7.9	$ADA_{6)}$ pH 7.4	$HEPES_{7}$ pH 8.2	Imidazole pH 7.95	$Tris_{8)}$ pH 9.0	

1) 1M SPG Buffer contains: 0.125M Succinic Acid, 0.5M NaH_2PO_4, 0.375M Glycine adjusted to the appropriate pH with 10M NaOH.
2) 1M MMT Buffer contains: 0.2M DL-Malic Acid, 0.4M MES monohydrate, 0.4M Tris adjusted to the appropriate pH with either 10M HCl or 10M NaOH
3) 1M PCB Buffer contains: 0.4M Sodium Propionate, 0.2M Sodium Cacodylate trihydrate, 0.4M Bis-Tris Propane adjusted to the appropriate pH with 10M HCl
4) MES(2-(N-morpholino)ethanesulfonicacid)monohydrate
5) Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
6) N-(2-Acetamido)iminodiaceticacid
7) 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonicacid
8) 2-Amino-2-hydroxymethyl-propane-1,3-diol

3.2.4 Electron microscopy

The structure of isolated spliceosomal complexes was analyzed by electron microscopy (EM). Cryo EM experiments were performed in collaboration with Dr. David Haselbach (Department of Structural Dynamics, MPI-BPC). Carbon films and grids for cryo EM were kindly prepared by Frank Würriehausen (Department of Structural Dynamics, MPI-BPC).

3.2.4.1 Preparation of continuous carbon film and holey carbon grids

For continuous carbon films, mica (Plano G250- 1, 25x75 mm) was used as the coating carrier. To obtain a smooth surface that does not have beforehand any direct air contact, the mica was freshly split. The fresh side was indirectly exposed to carbon in an Edwards E12E vacuum coating unit (BOC Edwards, Kirchheim) by conducting electric current through two carbon rods (Ringsdorff Werke GmbH, Bonn).

Low-cost grids with irregular holes were made to evaluate sample quality. Copper grids were covered with self-made triafol film containing holes. Then carbon was directly sputtered onto the triafol film.

To dissolve and remove the triafol film between the carbon layer and copper grids, the grids were incubated in ethyl acetate overnight and air dried. In order to record images that were used for cryo EM analysis and 3D reconstruction, copper EM grids with a holey carbon film (R3.5/1) (Quantifoil Micro Tools GmbH, Jena) were used.

3.2.4.2 Sample preparation

For EM investigations, affinity-purified complexes were subjected to a second glycerol gradient centrifugation under mild-fixating GraFix conditions (Kastner, Fischer et al. 2008). Briefly, 6-30 pmol of spliceosomes were loaded onto a linear 5-20% sucrose and 0-0.1% (v/v) glutaraldehyde gradient containing G-150, if not stated otherwise. Gradients were centrifuged at 60000 rpm for 110 min in a TH660 rotor or at 23000 rpm for 15 h in TST 41.14 rotor at 4°C and harvested from the bottom in 200 or 500 µl fractions, respectively. The distribution of ^{32}P-labeled RNA across the gradient was determined by Cherenkov counting and peak fractions were used for EM studies. The purified spliceosomal complexes were allowed to adsorb on a thin carbon film prior to negative staining or rapid plunge freezing into liquid ethane at 100% humidity and 4°C.

3.2.4.3 Negative stain EM analysis

Negative staining provides high contrast with a relatively simple preparation procedure. Therefore, it was mostly used for a quick sample quality estimation. As a staining solution for the negative stain EM, we used saturated (~5% w/v) uranyl formate, dissolved in double distilled water. The solution was thoroughly mixed and then centrifuged at 13000 rpm for 20 min at 4°C in Biofuge pico (Heraeus) in order to pellet excess of uranyl formate. To adsorb isolated complexes, a piece of carbon coated mica was floated on the surface of the sample well in a custom made teflon block. Then, carbon film with absorbed sample was covered with a copper EM grid, transferred to the uranyl formate staining solution and incubated for 1 min.

For visualization of structural features, images were recorded at a magnification of 88000x, corresponding to a pixel size of 2.5 Å^2/pixel with a CM200 FEG electron microscope (Philips, Netherlands) equipped with a 4kx4k charge-coupled (CCD) camera (TVIPS) and a room temperature holder (Philips, Netherlands). For each data set 10000-15000 individual single-particle images were collected. Repeating rounds of image processing of these images were performed using the in-house software package (coweyes). After a reference-free alignment, images were subjected to multivariate statistical analysis and classification (van Heel and Frank 1981, van Heel and Harauz 1988, Dube, Tavares et al. 1993). The resulting class averages were used as reference images in subsequent rounds of alignment until the class averages were stable.

3.2.4.4 Cryo EM image acquisition and analysis

Cryo EM images were recorded at -193°C in a Titan Krios electron microscope (FEI Company, The Netherlands) on a Falcon 3EC direct electron detector at a nominal 59,000x magnification, resulting in a calibrated pixel size of 1.16 $Å^2$/pixel on the specimen level. 20 frames were recoded for each micrograph with an average dose of 2 e^- per frame per $Å^2$. Motion correction and spatial frequency weighed frame sum was achieved using the *unblur* software suite (Campbell, Cheng et al. 2012). Ctf parameters were determined with Gctf (Zhang 2016). Summed micrograph images were then evaluated based on CTF parameters. Gautomatch (http://www.mrc-lmb.cam.ac.uk/kzhang/) software was used to pick particles. As a reference I used 40° projections of the yeast B^{act} model (EMD-9524) filtered to 60 Å. In total ~0.52 million particles from the pre-sorted cryo-micrographs were collected and cleaned by 3D classification in RELION. The remaining ~0.48 million particles were subsequently applied to three separate rounds of 3D classification in RELION (Scheres 2012) resulting in two different forms of the human B^{act} complex. Best classes were further refined to achieve the final structures. Part of the lower resolution density of one of the forms was excluded with a soft mask with a cut-off of 7 voxel in the final rounds of the refinement. Resolution was determined by Fourier shell correlation that was calculated from two independent data sets with a threshold of 0.143. A local resolution estimation was made in ResMap (Kucukelbir, Sigworth et al. 2014).

3.2.4.5 Model fitting and building

Structures were visualized in UCSF Chimera (Pettersen, Goddard et al. 2004). Available X-ray or homology models of proteins were placed into the EM density mostly using Chimera. Individual models of substructures (e.g. domains or structural motifs) were further docked as rigid bodies by Coot (Emsley, Lohkamp et al. 2010). After visual inspection, the models were adjusted manually in the density. The U2/U6 helix II was refined by rigid-body fitting of an idealized double-stranded RNA helix. Final visualization was carried out with Chimera.

3.2.5 Mass spectrometry

3.2.5.1 Protein identification by mass spectrometry

For protein identification via mass spectrometry, proteins recovered from affinity-purified spliceosomal complexes were separated on 4-12% NuPAGE™ gradient gels (Invitrogen) and stained with Coomassie. Entire lanes were cut into 23 slices and proteins were digested in-gel with trypsin and extracted as described by Shevchenko *et al.* (Shevchenko, Wilm et al. 1996). The extracted peptides were analyzed on a LTQ-Orbitrap Velos or Q Exactive Mass Spectrometers (Thermo Scientific) under standard conditions. Proteins were identified by searching fragment spectra against the NCBI non-redundant database using Mascot as a search engine.

3.2.5.2 Crosslinking and crosslink identification by mass spectrometry

All crosslinking experiments were performed in collaboration with Dr. Olexandr Dybkov (Dep. of Cellular Biochemistry; MPI-BPC).

Purified B^{act} complexes were cross-linked with 150 µM BS3 for 40 min at 20°C and purified further by a second density gradient centrifugation step. Approximately 15 pmol of B^{act} complexes were pelleted by ultracentrifugation and analyzed essentially as described before (Leitner, Walzthoeni et al. 2014) with the following modifications: precipitated material was dissolved in a buffer containing 4 M urea and 50 mM ammonium bicarbonate, reduced with DTT, alkylated with iodoacetamide, diluted to 1 M urea and digested with trypsin (Promega) (1:20 w/w). Peptides were reverse-phase extracted and fractionated by gel filtration on a Superdex Peptide PC3.2/30 column (GE HealthCare). 50 µl fractions corresponding to an elution volume of 1.2–1.8 ml were analyzed on a Orbitrap Fusion Lumos Tribrid and Q Exactive HF Mass Spectrometer (Thermo Scientific). Protein-protein crosslinks were identified by pLink 1.23 search engine and filtered at FDR 1% (pfind.ict.ac.cn/software/pLink) according to the recommendations of the developer (Yang, Wu et al. 2012).

4 Results

4.1 Experimental strategy to isolate human activated spliceosomes

For the purification of activated (B^{act}) spliceosomes, I used a pre-mRNA substrate, termed PM5-10, that contains a single 5' exon, an intron with a branch site and an additional 10 nucleotides (nts) downstream of the BS, but lacking a 3' splice site (ss) and 3' exon. It has previously been shown that the PM5-10 construct allows the assembly of the B^{act} complex but that due to the short polypyrimidine tract (PPT) it cannot be catalytically activated (Bessonov, Anokhina et al. 2008). The PM5-10 construct contained at its 5' end three MS2 stem loops, which were co-transcriptionally added to allow affinity purification of the B^{act} complex on amylose agarose using a fusion protein of MS2 coat protein and maltose-binding protein (MBP).

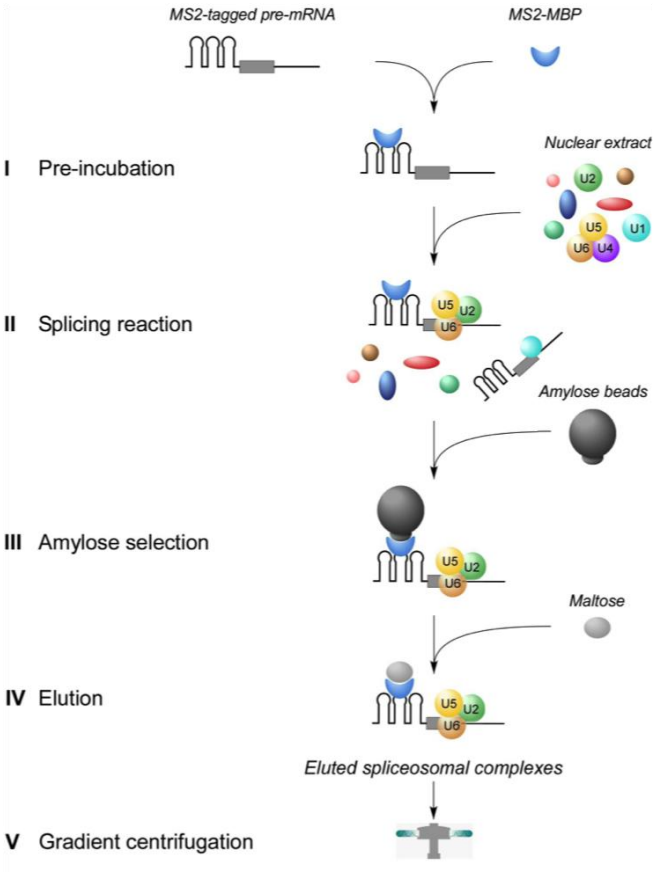


Figure 4.1: General strategy to isolate human spliceosomal complexes using an MS2-tagged pre-mRNA substrate

PM5-10 pre-mRNA tagged at its 5' end with three binding sites of the MS2 virus coat protein was preincubated with a MS2-MBP fusion protein (step I). Subsequently, nuclear extract was added and spliceosomes were allowed to form (step II). Complexes were then affinity selected using amylose beads (step III) and subsequently eluted with maltose (step IV). Eluted spliceosomal complexes were fractionated by size on a linear glycerol or sucrose gradient (step V).

The B^{act} purification scheme is illustrated in Figure 4.1. Briefly, MS2-PM5-10 pre-mRNA was incubated with a 20-fold molar excess of MS2-MBP fusion protein. Then, a standard splicing reaction was performed for 180 min. A 30-fold molar excess of DNA oligonucleotides, which are complementary to nts -6 to -18 and -12 to -24 (relative to the 5'ss) of the PM5-10 substrate was added and the reaction was incubated at 30°C for additional 20 min. During the latter incubation step, endogenous RNaseH

cleaves only pre-mRNA in assembly steps earlier than B^{act}, which allows the purification of B^{act} complexes largely uncontaminated by H, A or B complexes. B^{act} complexes are then purified by affinity chromatography using amylose beads and eluted from the amylose column under native conditions with an excess of maltose. Eluted complexes are subsequently fractionated by centrifugation on a linear glycerol or sucrose gradient, which allows separation of the 40-45S B^{act} spliceosomes from smaller MS2-MBP containing RNP particles. 40-45S peak fractions were either pooled or analysed individually.

4.2 Optimization of human B^{act} complex stability using a high-throughput buffer screen

Spliceosomes are known to be structurally dynamic, a property, which often hinders the reconstruction of 3D cryo-electron microscopy (cryo-EM) structures at high resolution (Stark and Lührmann 2006). Therefore I first tested numerous buffer conditions with the intention to biochemically stabilize the spliceosome in order to minimize its conformational heterogeneity and flexibility. A convenient screening method to assess the effect of numerous conditions (e.g. buffers and small molecule additives) on the stability of a protein or RNP complex of interest in a high-throughput format was recently established in the group of Prof. Stark (Dep. of Structural Dynamics, MPI-BPC) and was termed ProteoPlex (Chari, Haselbach et al. 2015). In the procedure, a hydrophobic dye (SYPRO Orange) interacts with exposed hydrophobic protein regions upon heat denaturation. Thus, the thermal stability of a protein complex is estimated at various temperatures by measuring the increase in fluorescent intensity as a result of binding of the fluorescent dye to unfolded proteins (Boivin, Kozak et al. 2013, Chari, Haselbach et al. 2015). For this purpose, B^{act} spliceosomes purified under standard splicing conditions were subjected to a Thermofluor screen under 88 different buffer conditions (Figure 4.2 A).

The screen revealed that the hB^{act} complex is stable under a variety of conditions from pH 6.4 to 8.2 and the particles were as stable under standard buffer conditions as in any other conditions tested. Therefore, I decided to validate the quality of the B^{act} complexes, purified under different buffer conditions, using transmission electron microscopy (TEM). A monodisperse distribution of spliceosomal particles was observed in a range of different pH values (Figure 4.2 B) and no or only little fragmentation of spliceosomes had occurred.

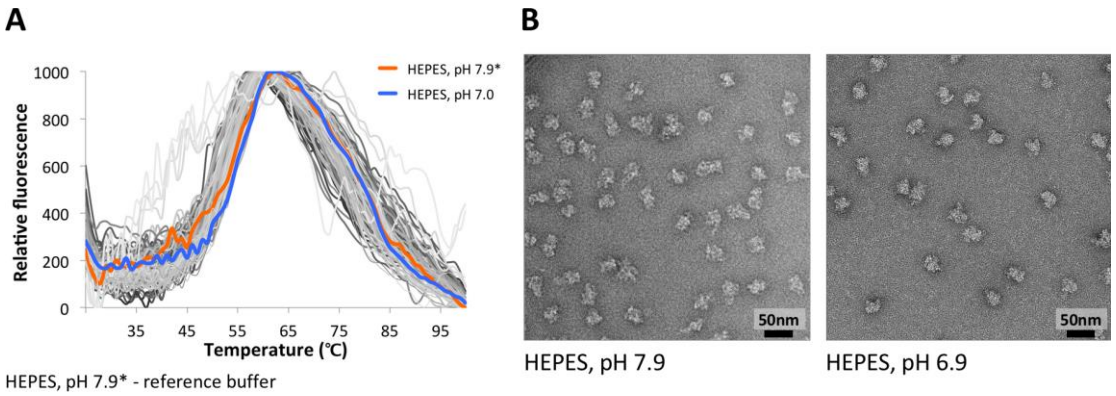


Figure 4.2: ProteoPlex approach for identifying stabilizing buffer conditions

(A) Normalized experimental melting curves for B^{act} complexes in all tested buffer conditions from a ProteoPlex pH screen. Standard buffer is shown in orange. One of the most optimal buffers as identified with ProteoPlex is indicated in blue. **(B)** Electron micrographs of the B^{act} complex under standard (left) and one of the most optimal conditions as determined by ProteoPlex (right).

Then I varied the buffer conditions during the splicing reaction / B^{act} complex formation, first by changing the pH from 6 to 9 (Figure 4.3 A). In brief, the *in vitro* splicing mixture was complemented with buffer at the pH of interest to a final concentration of 50 mM. After 3 hours of incubation the efficiency of B^{act} complex formation was initially analysed on a native agarose gel in the presence of heparin. Under pH conditions close to neutral, I observed that assembly of all spliceosomal intermediates including B^{act} was most efficient, suggesting that the complexes were more stable at neutral pH. At slightly acidic buffer conditions, however, I observed accumulation of earlier spliceosome forms, which could indicate that spliceosome assembly was slowed down. However, when the affinity purified B^{act} type complexes, which were assembled in a buffer at pH 6.9, were separated by gradient centrifugation, the spliceosomes exhibited a bimodal sedimentation behaviour, indicating some heterogeneity of the sample (Figure 4.3 B). For comparison, spliceosomes assembled and purified under standard splicing conditions (pH 7.9) were also subjected to gradient centrifugation, and fractionated as a largely symmetrical peak.

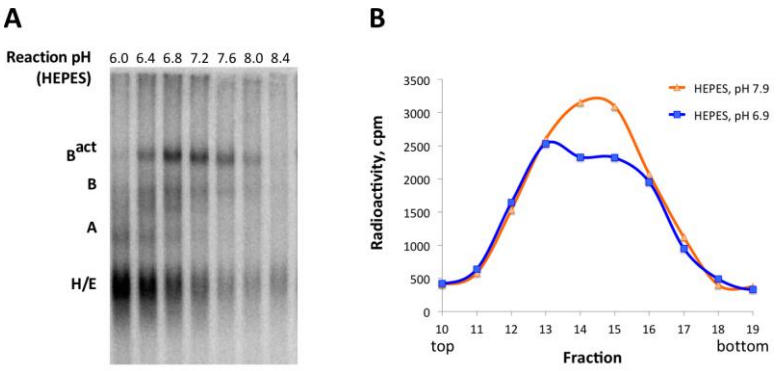


Figure 4.3: Optimization of reaction conditions

(A) In vitro splicing reaction of ^{32}P-labeled PM5-10 pre-mRNA performed in HeLa nuclear extract for 3 h. Spliceosomal complexes (positions indicated on the left) were analysed by agarose gel electrophoresis. Bands were visualized by autoradiography. **(B)** Sedimentation behaviour of spliceosomes formed under different buffer conditions on a 10-30% glycerol gradient.

RNA analysis of the gradient fractions showed that all peak fractions contain unspliced pre-mRNA and nearly equimolar amounts of U2, U5 and U6 snRNAs, but U1 and U4 snRNAs were largely absent, indicating that the vast majority of spliceosomes had undergone activation (Figure 4.4 A). Negative stain EM analysis of the spliceosomes from fractions 12/13 (low S value) and 16 (high S value), respectively, revealed that they displayed significantly different appearances. While faster sedimenting complexes contained largely round particles with an appearance similar to spliceosomes assembled and purified under standard conditions at pH 7.9, slower sedimenting particles displayed rather elongated shapes (Figure 4.4 B).

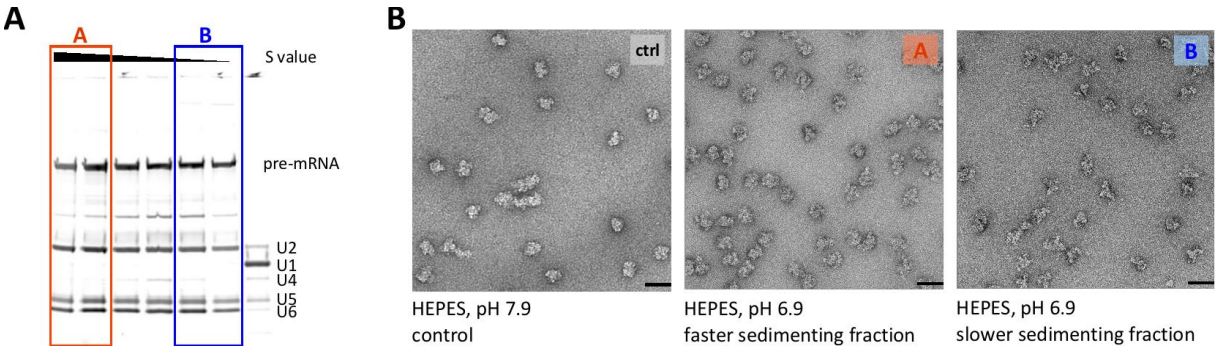


Figure 4.4: Change of the reaction conditions leads to an increase in conformational heterogeneity

(A) The peak gradient fractions of the complexes formed in an in vitro splicing reaction under neutral buffer conditions (HEPES pH 6.9) were analysed on a NuPAGE gel, followed by SYBR Gold staining. RNA identities are shown on the right. Fractions corresponding to faster and slower sedimenting particles are indicated in orange and blue boxes, respectively. **(B)** Negative stain electron micrographs of complexes from gradient fractions corresponding to lanes "A" and "B" of panel A compared to complexes assembled under standard conditions (ctrl, pH 7.9). Scale bar, 50 nm.

As the B^{act} complexes obtained in the standard splicing reaction appeared to be optimal spliceosomal particles, which did not aggregate and exhibited a homogeneous appearance under negative stain EM conditions, I decided to proceed with these B^{act} preparations for unstained cryo-EM analysis.

4.3 Structure determination of the activated human spliceosome

4.3.1 Purification and characterisation of B^{act} spliceosomes.

The final optimized purification strategy of B^{act} spliceosomes involved several steps. First, preparative amounts of B^{act} complex were assembled under standard splicing conditions in a 700 ml splicing reaction at pH 7.9 (Chapter 3.2.3.3). The affinity purified B^{act} complexes were fractionated on 38.6 ml 5-20% linear sucrose gradients in G-150 buffer (20 mM HEPES-KOH, pH 7.9, 150 mM NaCl, 1.5 mM $MgCl_2$). Peak fractions were split in two groups based on sedimentation behaviour (slower and faster sedimenting complexes). Particles in fractions 3-4 and 6-7, respectively, were pooled (Figure 4.5 A), concentrated by centrifugation with an Amicon 50kD cut-off unit and subjected separately to fractionation on a second sucrose gradient, which contained also 0 to 0.1% (v/v)

glutaraldehyde (Kastner, Fischer et al. 2008). Spliceomes from individual fractions from this gradient were concentrated by centrifugation with an Amicon 50kD cut-off unit and subjected to EM analysis.

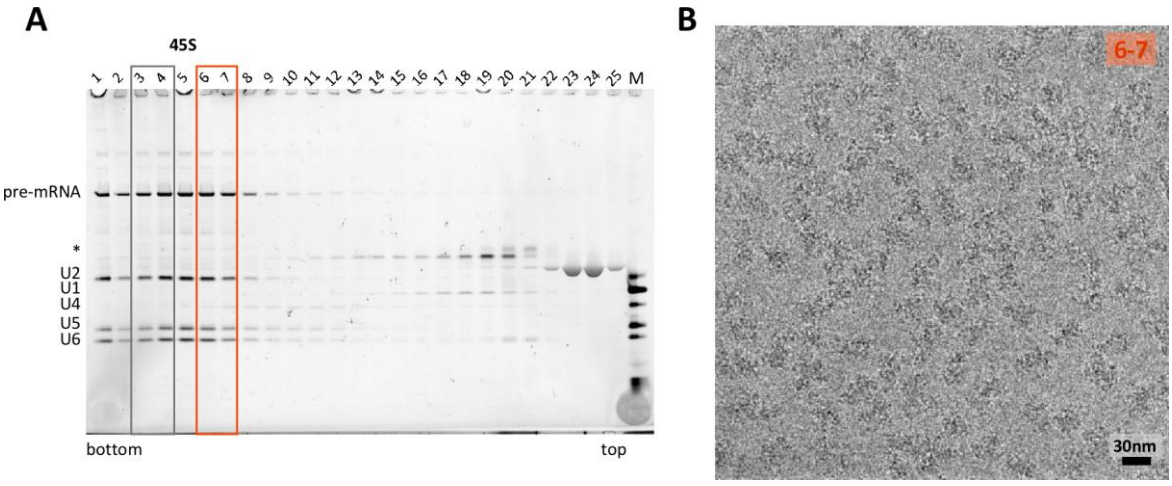


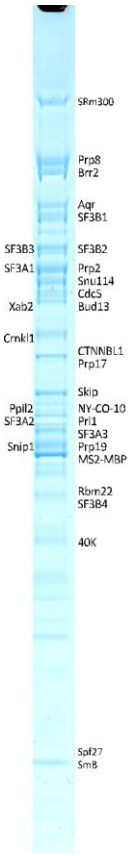
Figure 4.5: Purification and cryo-EM analysis of the human B^{act} complex

(A) The affinity-purified sample was further fractionated by sucrose gradient centrifugation. The total RNA from each fraction was analysed on a NuPAGE gel, followed by SYBR Gold staining. Fractions 3-4 and 6-7 were concentrated and analysed separately. The asterisk shows the position of RNaseH degraded pre-mRNA. **(B)** Typical cryo-EM raw image of unstained human B^{act} spliceosomes (fractions 6-7) recorded with a Titan Krios (FEI Company) electron microscope at a nominal magnification of 59000x with a Falcon 3EC direct electron detector resulting in a pixel size of 1.16 $Å^2$/pixel.

As shown in Figure 4.5 A, purified spliceosomes sedimenting with an S value of about 45S contain U2, U5 and U6 snRNAs in addition to pre-mRNA, consistent with the RNA composition of B^{act} complexes. I decided to split the peak fractions in two groups. Although no difference could be found between particles from separated pools, I processed the pools for cryo-EM analysis separately. Negatively stained and cryo-grids were prepared as described in Chapter 3.2.4. The B^{act} complexes from both pools were monodisperse in negative stain EM, as well as under cryogenic imaging conditions. An example of a cryo micrograph is shown in Figure 4.5 B.

Proteins present in isolated spliceosomes from corresponding fractions of non-fixed gradients were separated on a denaturing 4-12% NuPAGE gels (Invitrogen), followed by Coomassie staining. Proteins from a single lane were identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) in the Bioanalytical Mass Spectrometry Group as described in Chapter 3.2.5. The protein composition of purified spliceosomes is summarized in Figure 4.6. Mass-spectrometry identified more than 120 proteins, including snRNP and non-snRNP proteins (data not shown). These include proteins associated with U2 snRNP, some of the U5 proteins (Prp8, Brr2, Snu114, 40K and Sm core proteins), the Prp19/CDC5L protein complex and related factors, proteins of the intron-binding complex, components of the human RES complex, components of the exon junction complex (EJC) and proteins associated with the activated spliceosome through the first catalytic cycle, as previously described (Bessonov, Anokhina et al. 2010). Additional proteins found in isolated spliceosomes include MS2-MBP fusion protein, cap-binding proteins, a number of hnRNP and SR proteins (data not

shown). Consistent with previous studies, U1 and U4/U6 snRNP proteins, including LSm proteins, as well as B complex specific proteins were not identified by MS analysis or were underrepresented (Bessonov, Anokhina et al. 2010, Agafonov, Deckert et al. 2011).



Name used	Size kDa	*H. sapiens* name	Accession number	*S. cerevisiae* name	*S. pombe* name
U2 snRNP specific proteins					
U2-A'	38	SNRPA1	O73931	Lea1	lea1
U2-B''	25	SNRPB2	P08579	Msl1	msl1
SF3A1	89	SF3A1/SF3a120	Q15459	Prp21	sap114
SF3A2	49	SF3A2/SF3a66	Q15428	Prp11	sap62
SF3A3	59	SF3A3/SF3a60	Q12874	Prp9	sap61
SF3B1	146	SF3B1/SF3b155	O75533	Hsh155	prp10
SF3B2	100	SF3B2/SF3b145	Q13435	Cus1	sap145
SF3B3	135	SF3B3/SF3b130	Q15393	Rse1	prp12
SF3B4	44	SF3B4/SF3b49	Q15427	Hsh49	sap49
SF3B5	10	SF3B5/SF3b10	Q9BWJ5	Ysf3	UP C211.05
SF3B6	15	SF3B6/SF3b14a/p14	Q9Y3B4		p14-like
PHF5A	12	PHF5A/SF3b14b	Q7RTV0	Rds3	ini1
U5 snRNP specific proteins					
Prp8	274	PRPF8/220K	Q6P2Q9	Prp8	spp42
Brr2	244	SNRNP200/200K	O75643	Brr2	spp41
Snu114	109	EFTUD2/116K	Q15029	Snu114	cwf10
40K	39	SNRNP40/40K	Q96DI7		cwf17
Prp19/Cdc5L complex					
Prp19	55	PRPF19	Q9UMS4	Prp19	cwf8/prp19
Cdc5	92	CDC5L	Q99459	Cef1	cdc5
Spf27	26	BCAS2	O75934	Snt309	cwf7
Prl1	57	PLRG1	O43660	Prp46	cwf1/prp5
AD002	27	CWC15/AD-002	Q9P013	Cwc15	cwf15
Ctnnbl1	65	CTNNBL1	Q8WYA6		ctnnbl1
Hsp73	71	HSPA8/HSC70	P11142		hsp70
Prp19 related proteins					
Skip	61	SNW1	Q13573	Prp45	cwf13/prp45
Rbm22	47	RBM22	Q9NW64	Ecm2+Cwc2	cwf5+cwf2
Crnkl1	100	CRNKL1	Q9BZJ0	Clf1	cwf4
Ppil1	18	PPIL1	Q9Y3C6		ppi1
G10	17	BUD31	P41223	Bud31	cwf14

Name used	Size kDa	*H. sapiens* name	Accession number	*S. cerevisiae* name	*S. pombe* name
IBC complex					
Aquarius	171	Aquarius/KIAA0560	Q60306		cwf11
Xab2	100	XAB2	Q9HCS7	Syf1	cwf3
Isy1	33	ISY1	Q9ULR0	Isy1	cwf12/isy1
CypE	33	PPIE	Q9UNP9		UP C428.12c
Ccdc16	42	ZNF830	Q96NB3		
B^{act} proteins					
Cwc22	105	CWC22/KIAA1604	Q9HCG8	Cwc22	cwf22
Prp17	66	CDC40	O60508	Prp17	prp17
NY-CO-10	54	CWC27	Q6UX04	Cwc27	cwf27/cyp7
Rnf113A	39	RNF113A	O15541	Cwc24	cwf24
Ppil2	59	PPIL2	Q13356		cyp8
Ccdc12	19	CCDC12	Q8WUD4		
Prp2	119	DHX16	O60231	Prp2	cdc28
GPKOW	52	GPKOW	Q92917	Spp2	cwf28
SRm160	102	SRRM1/SRm160	Q8IYB3		
SRm300	300	SRRM2/SRm300	Q9UQ35	Cwc21	cwf21
PRCC	52	PRCC	Q92733		
RES complex					
Snip1	46	SNIP1	Q8TAD8	Pml1	
Bud13	70	BUD13	Q9BRD0	Bud13	cwf26
RbmX2	37	RBMX2/CGI-79	Q9Y388	Snu17	
EJC complex					
eIF4A3	47	eIF4A-III	P38919		
Y14	20	RBM8A	Q9Y5S9		
Magoh	17	MAGOH	P61326		
Sm proteins					
SmB	25	SNRPB	P14678	SmB	SmB
SmD1	13	SNRPD1	P62314	SmD1	SmD1
SmD2	14	SNRPD2	P62316	SmD2	SmD2
SmD3	14	SNRPD3	P62318	SmD3	SmD3
SmE	11	SNRPE	P62304	SmE	SmE
SmF	10	SNRPF	P62306	SmF	SmF
SmG	8	SNRPG	P62308	SmG	SmG

Figure 4.6: Characterization of the protein composition of B^{act} spliceosomes

Proteins isolated from purified B^{act} complexes were visualized on a NuPAGE gel by Coomassie blue staining **(left)**. Proteins present in human B^{act} complexes and their molecular masses are summarized in the table. Proteins were identified by LC-MS/MS. Standard names for human proteins, as well as names of their homologues in yeast *S. cerevisiae* and *S. pombe* are indicated. Accession numbers for human proteins are shown in blue.

4.3.2 Structure determination, model-building and overview of the B^{act} spliceosome structure

Images were recorded using a Falcon 3EC direct electron detector with a calibrated pixel size of 1.16 $Å^2$/pixel on the specimen level. The defocus range was varied from ~1.0 to ~3.0 µm. A total of 20 frames per stack were recorded with the total dose rate about 2 $e^-/Å^2$ for each frame. A total of 7 494 cryo-EM micrographs were collected. Frames in each image were aligned and summed using the *unblur* software suite (Campbell, Cheng et al. 2012). The defocus values of each image were determined by Gctf (Zhang 2016).

A total of ~520 000 particle images (henceforth termed particles) were autopicked using the reference-based particle picking routine in Gautomatch software (http://www.mrc-lmb.cam.ac.uk/kzhang/). As a template, I used 40° projections of the yeast

B^{act} complex map (EMD-9524) low-pass filtered to 60 Å. All subsequent 2D and 3D classifications as well as 3D auto-refinement runs were performed in RELION using particles binned to a pixel size of 2.32 $Å^2$/pixel.

An initial dataset of 50 000 particles was processed to inspect sample quality and composition and to define the image processing strategy that was applied to the entire data set (data not shown). In order to remove false positive particles from the initial data set, I performed reference-free 2D classification and subsequent 3D multi-reference refinement (unsupervised 3D classification) based on likelihood optimisation. The unsupervised 3D classification not only helps to remove junk-particles from a dataset, but was also shown to separate classes with large conformational differences (Scheres, Gao et al. 2007). Considering a high level of structural similarity between human and yeast spliceosomes, I started with the map of *S. cerevisiae* B^{act} complex filtered to 60 Å as an initial 3D reference (EMD-9524) (Yan, Wan et al. 2016). For the 3D classification I used 5 classes (K = 5) and performed 25 iterations of likelihood optimization using non-aligned particles.

3D classification yielded one poorly defined class comprising mostly junk particles that were discarded (class 3) (Figure 4.7). The other four classes displayed reasonable similarity to the yeast B^{act} structure and showed common structural features, such as an elongated central body connected to a foot and a top domain together with shallow and steep slopes. However, they also exhibited significant differences in peripheral regions, which could be interpreted in terms of two different conformations, henceforth termed as form A and form B, respectively (Figure 4.7). Classes 1 and 2, comprising ~122 000 and ~115 000 particles, respectively, represent form A, which has a globular density at the top right corner as a characteristic feature. Classes 4 and 5, comprising ~122 000 and ~123 000 particles, respectively, reflect the second form B, which has an elongated density element at the lower right side of the central body, as well as an additional globular density in the middle and at the top left corner, while density at the top right corner is, in contrast, much less visible (Figure 4.7).

Although independent processing of these four classes helped to improve the above mentioned forms, the final resolution was restricted due to insufficient number of particles. Therefore, the four selected classes were merged, yielding ~480 000 particles that were read back to the original images and re-extracted based on their refinement centre using RELION. Relative orientations of the selected particles were determined with improved accuracy in 3D auto-refinement runs with respective references 1 and 2, representing forms A and B, (Figure 4.7).

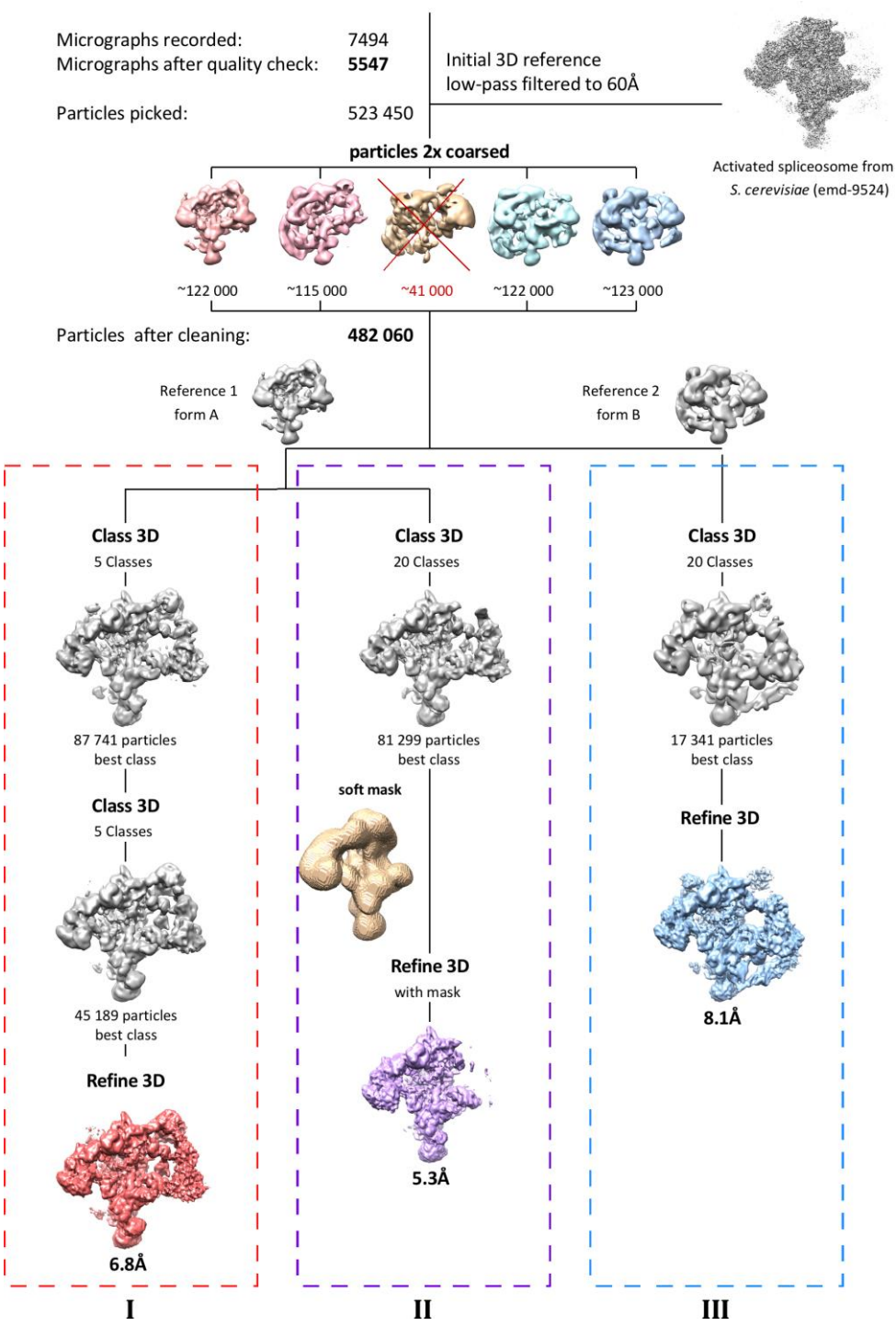


Figure 4.7: Computational sorting scheme of the human B^{act} complex

Using the particle picking software Gautomatch and 40° projections of the yeast B^{act} model (EMD-9524) filtered to 60 Å as a reference, I extracted ~520 000 particles from the pre-sorted cryo-micrographs and cleaned the dataset by 3D classification in RELION that resulted in separation of two conformationally different forms. Good classes comprising ~480 000 particles were merged and subsequently applied to three separate rounds of 3D classification in RELION featuring 5 or 20 classes each. **(I)** For the determination of the most complete form A, the 45 189 particles finally contributing to the best 3D class were used for the 3D auto-refinement revealing an 6.8 Å resolution structure (referred to as the unmasked EM density map of the form A). **(II)** For the highest resolution structure, I excluded part of lower resolved density representing the first conformational state in the final rounds of the refinement (form A with mask). A soft mask with a 7 voxel frame and an additional 7 voxel soft frame was used for the refinement and for the resolution determination. The final map was obtained with a resolution of 5.3 Å. **(III)** Exhaustive 3D classification targeting form B yielded a 3D class, consisting of 17 341 particles. Further refinement revealed an 8.1 Å resolution structure (form B).

First, I set out to resolve peripheral regions described above and to identify components of the corresponding densities. To obtain the most complete form A, a guided multi-reference classification procedure was applied again to all ~480 000 auto-refined particles. Two subsequent rounds of 3D classification with K = 5 followed by selection and 3D auto-refinement of the best class were performed yielding an average resolution of 6.8 Å (Figure 4.7, red, dashed box I).

For the highest resolution structure, ~480 000 particles, auto-refined with the reference 1 low-pass filtered to 40 Å were subjected to a 3D classification with K = 20 (Figure 4.7, violet, dashed box II). The best resulting class comprising ~81 000 particles yielded a ~10 Å resolution structure. While peripheral regions were somewhat lower resolved, I decided to mask them out and excluded part of the density referred to as the steep slope, applying a soft mask with a 7 voxel frame and an additional 7 voxel soft frame (Figure 4.7). Finally, I performed 3D auto-refinement resulting in a map with an average resolution of 5.3 Å. As can be seen in Figure 4.8 A, the angular distribution of particles used for the final reconstruction of the core B^{act} structure is reasonable. Resolution was determined by Fourier shell correlation calculated from two independent data sets with a threshold of 0.143 (Figure 4.8 B). Local resolution estimation using ResMap revealed that indeed peripheral regions are less well resolved but still at sub-nanometer resolution (Figure 4.9).

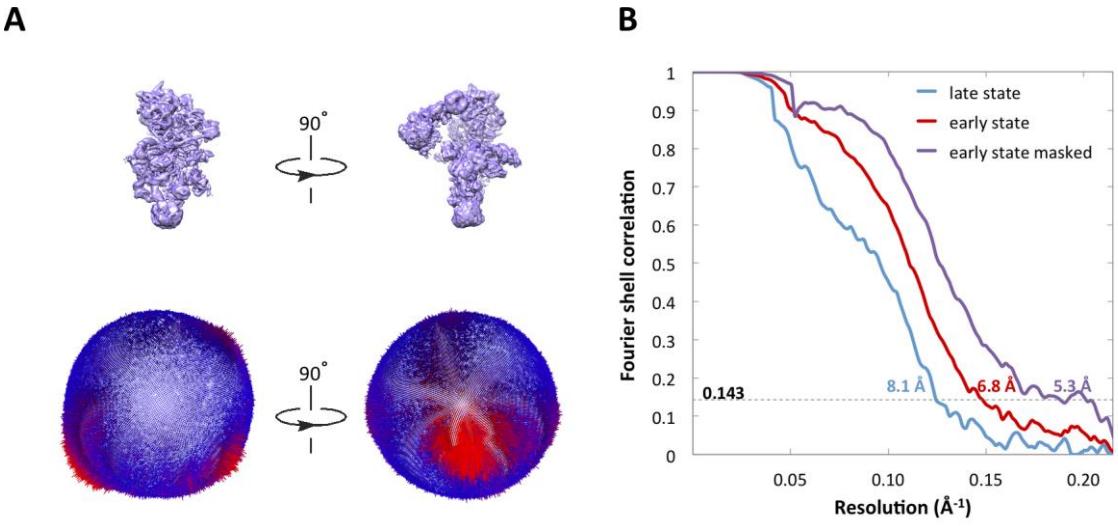


Figure 4.8: Cryo-EM analysis of the human B^{act} complex

(A) Two perpendicular orientations of the B^{act} complex are shown. Angular distribution of the particles used for the reconstruction of the spliceosomal B^{act} complex at 5.3 Å resolution. Each column represents one view and the height of the column is proportional to the number of particles for that view. Red columns indicate abundant views. **(B)** The resolution of the overall map of A and B forms and the core region of the A form is estimated to be 6.8 Å, 8.1 Å and 5.3 Å, respectively, on the basis of the FSC criterion of 0.143.

The next aim was to obtain the highest possible resolution for the peripheral regions of the form B. Due to considerable heterogeneity, the main challenge was to distinguish particles belonging to different conformational variants. Iterative multi-reference 3D classification targeting only those classes that exhibited form-B appearance lead to a significant decrease in the particle population,

indicating that either too many good particles were discarded during classification or too many variations of this conformational state can be distinguished. Therefore, I performed independent 3D auto-refinement for the entire dataset comprising ~480 000 particles with a low-pass filtered to 40 Å reference 2, representing form B. Subsequently, I performed exhaustive 3D classification with K = 20 resulting in few classes, characterizing the second conformational state. The best subset containing 17 341 particles was subjected to 3D refinement giving a structure with an average resolution of 8.1 Å, according to the FSC 0.143 criterion (Figure 4.8, blue, dashed box III). As can be seen in Figure 4.9 C, local resolution of the core part is better than 8 Å. The resulting map showed clear features of the elongated side density and the globular density in the middle part, while the globular density at the top peripheral region above the shallow slope could not be resolved at sub-nanometer resolution.

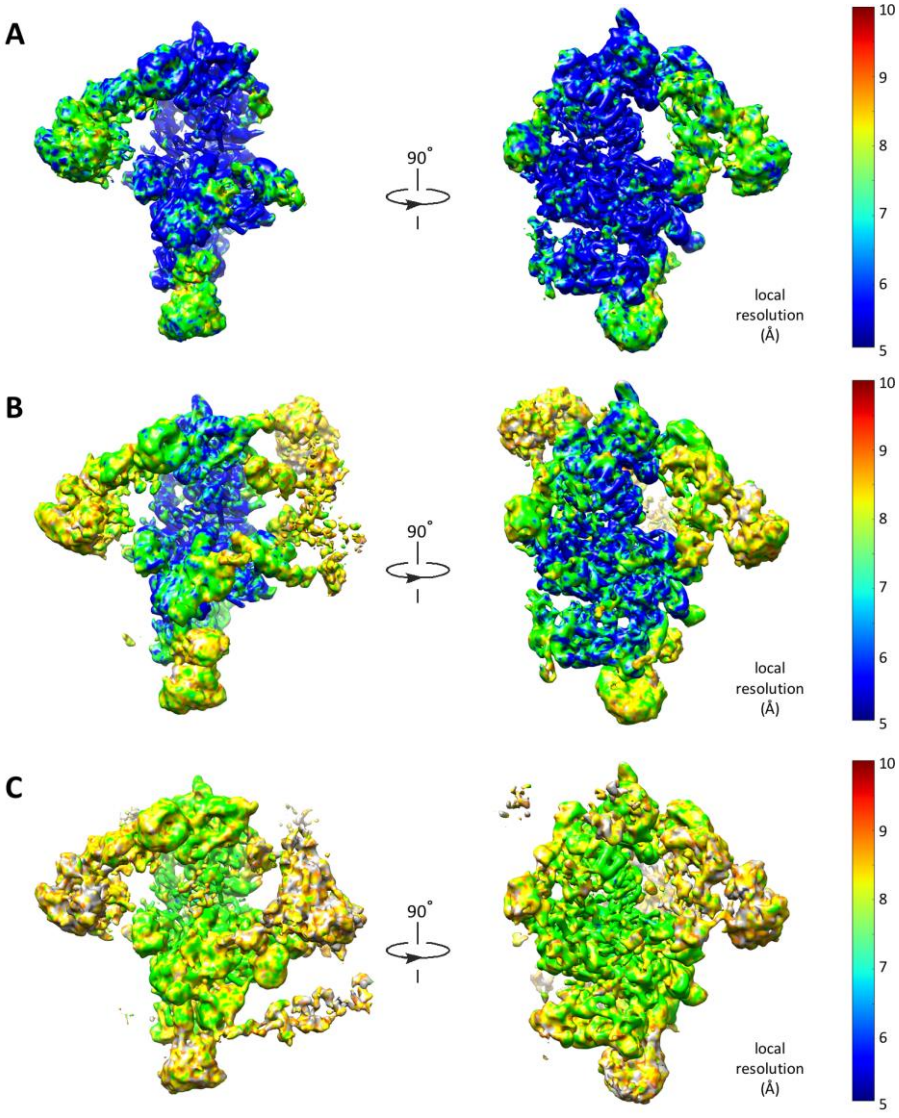


Figure 4.9: Local resolution plots of the B^{act} complexes

Two perpendicular views of the complexes are shown. The local resolutions are color-coded for different regions of the B^{act} complex and reveal a resolution distribution from ca. 5 to 10 Å with some less well-defined areas at the periphery of the complex. **(A)** Higher resolution regions (in blue, up to 5 Å) were obtained for the centrally-located core of the form A refined with the soft mask. **(B, C)** Peripheral regions that still have sub-nanometer resolution represent less stable areas in forms A (B) and B (C).

4.4 Structural overview of form A of the B^{act} complex

In the front view, the B^{act} spliceosome exhibits a mushroom-like shape with a central body connected to a foot and a top domain, as well as to a shallow and a steep slope (Figure 4.10). In the steep slope we distinguish a more elongated and a globular domains, termed UPD1 and 2 (unmasked peripheral domains 1 and 2). UPD1 is connected to the central domain by two massive density bridges (bridges 1 and 2), to the top domain by a thinner adjacent density bridge 3 and to a globular UPD2 by a bridge 4. Additionally, globular UPD2 is attached to the top domain by density bridges 5 and 6. Another unmasked domain, termed UPD3, is situated between bridges 2 and 3. Finally, a globular domain is attached to the lower part of the top domain (UPD4) and another one to the lower part of the central domain (UPD5).

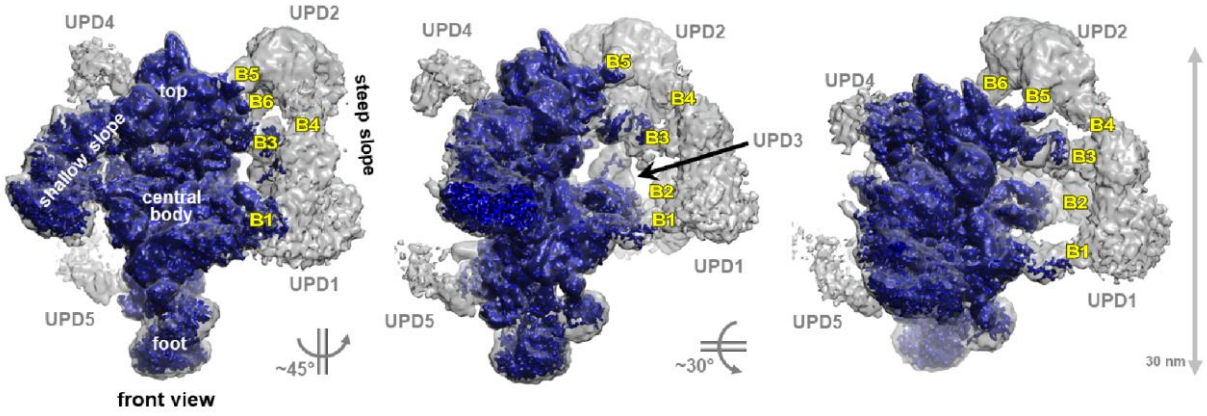


Figure 4.10: cryo-EM structure of the human B^{act} complex

Different views of the B^{act} complex EM density map. Better resolved regions are shown in blue and five additional domains, termed unmasked peripheral domain UPD 1-5 are shown in grey. Density bridges B1-B6 connect UPD 1-3 to the central body and a top domain. Scale bar is located at the right side.

4.4.1 Structural organisation of U5 snRNP proteins

U5 snRNP proteins that remain associated with the human B^{act} complex include four large proteins, Prp8, Brr2, Snu114 and U5 40K, in addition to the U5 Sm core and the U5 snRNA. The domain organization of Prp8, Brr2 and Snu114 is shown in Figure 4.11. The U5 Sm core, including the 3'-terminal U5 snRNA Sm site, is located at the foot of the 5.3 Å EM density map of B^{act}. The GTP-binding protein Snu114 shows a high degree of homology with the ribosomal elongation factor EF-2/EF-G (Fabrizio, Laggerbauer et al. 1997); it is located at the lower end of the main body (Figure 4.11, E). Domains D1 to D5 of Snu114 adopt the same compact structure as in the human (h) tri-snRNP and C* complex (Agafonov, Kastner et al. 2016, Bertram, Agafonov et al. 2017, Zhang, Yan et al. 2017).

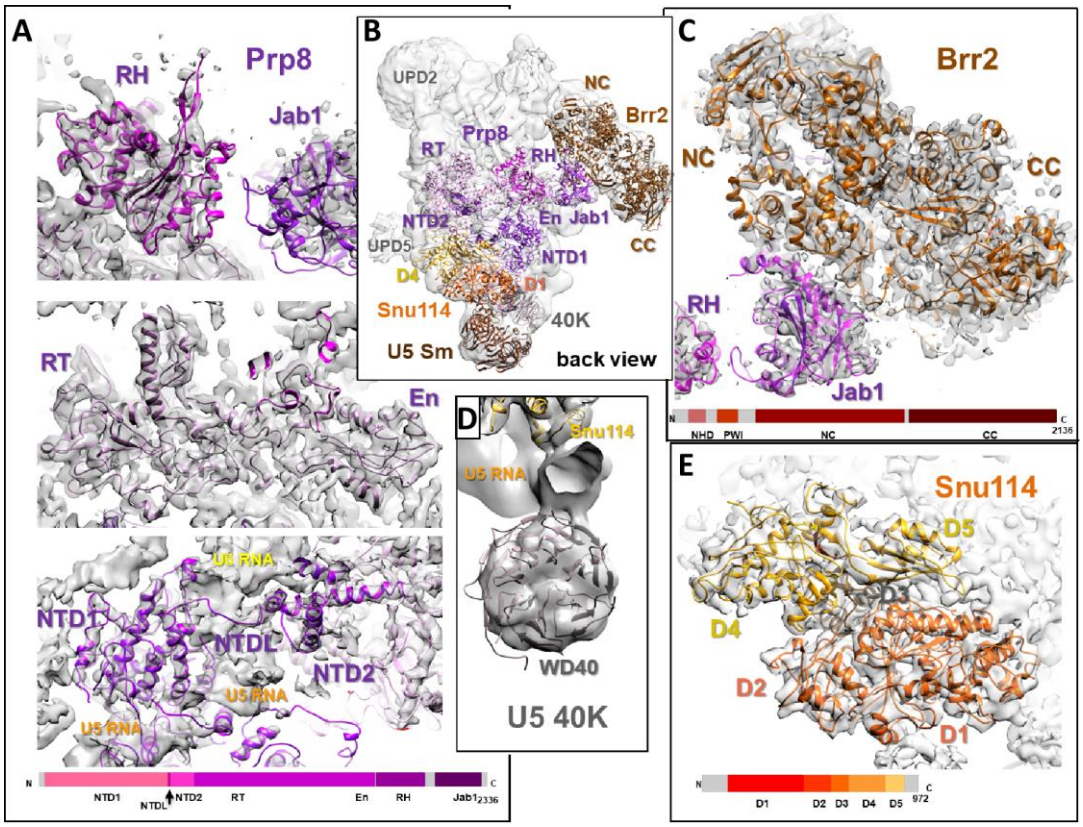


Figure 4.11: Structures and positions of the U5 snRNP proteins

(A) Fit and structural organisation of Prp8. Prp8 is comprised of the following domains: NTD1, NTD2: N-terminal domains 1 and 2; NTDL, linker between NTD1 and 2; RT, reverse-transcriptase-like; En, endonuclease-like; RH, RNase H-like; Jab1, Jab1/MPN-like. **(B)** Location of Prp8, Brr2, Snu114, U5 40K and U5 Sm proteins. **(C)** Expanded view showing fit and structural organisation of Brr2 (NC/CC, N-terminal/C-terminal helicase cassettes) with the bound Prp8 Jab1 domain. **(D)** Fit of the U5 40K WD40 domain. **(E)** Fit and structural organisation of Snu114 (domains D1 to D5 homologous to EF-G/EF-2).

The 274 kDa protein Prp8 is a major scaffolding protein situated at the heart of the spliceosome's catalytic core RNP. The large central RT/En domain of Prp8 lies with its long axis almost perpendicular to the axis of the main structural domain of B^{act}, such that the RT domain lies above Snu114 (Figure 4.11, B). The RH domain is closely associated with the "back side" of the En domain, whereby the RH RNA-binding β-sheet faces the En domain and the RH β-hairpin-loop points upwards. The Prp8 NTD1 has essentially the same structure as in the *S. cerevisiae* cryo-EM models; it is in contact with the U5 snRNA and Snu114 in the lower part of the main domain of B^{act} (Figure 4.11, lower A and B). The Prp8 RT/En domain is found at different positions relative to the NTD1 in the B^{act} complex compared with the human tri-snRNP (Figure 4.12). While in the tri-snRNP Prp8 has an open conformation, in which the En domain is located ca. 5 nm away from the NTD1, in the B^{act} complex it adopts a closed conformation. That is, the En domain is very close to NTD1, leaving a narrow channel between the two domains that appear to accommodate the 5' exon (see below). Thus, already at the B^{act} stage, Prp8 has undergone the major rearrangements that are required to create the pocket between the thumb/X, linker, NTD1 and NTD2, into which the catalytic RNA network docks (Figure 4.12, B).

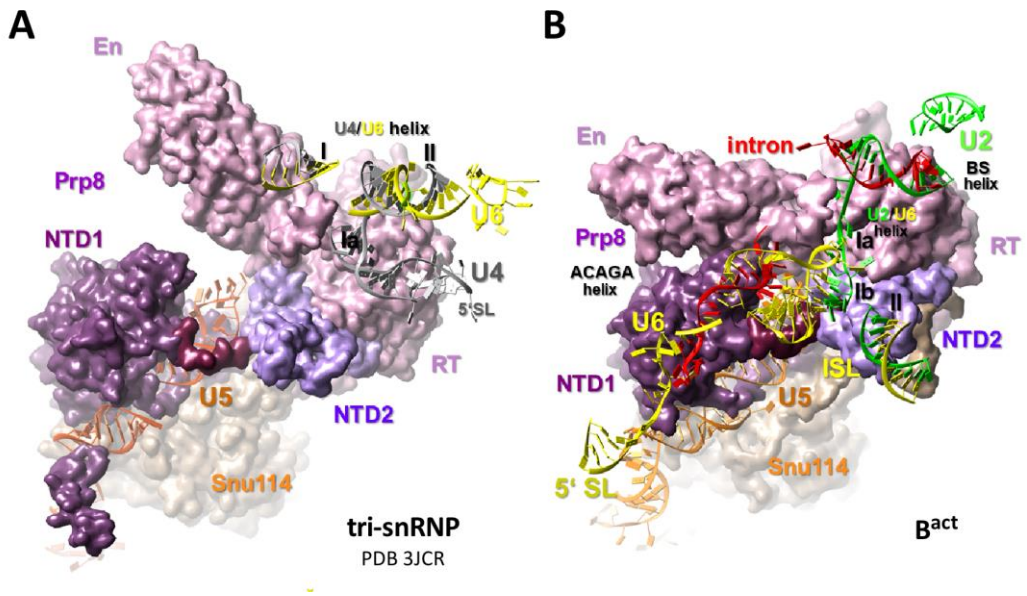


Figure 4.12: Conformational differences in Prp8's RT/En domains between the human tri-snRNP and B^{act} complex

Protein surfaces of Prp8 domains (NTD1, NTD2: N-terminal domains 1 and 2; RT, reverse-transcriptase-like; En, endonuclease-like) and Snu114 together with snRNA models are shown. **(A)** Open conformation of Prp8 in the human tri-snRNP (PDB 3JCR) (Agafonov, Kastner et al. 2016), where the En domain is spatially separated from NTD1. SL, stem loop **(B)** In the B^{act} structure Prp8 adopts a closed conformation. The catalytic U2/U6 RNA network and U2/BS helix are docked into the active site pocket of the Prp8 protein. ISL, U6 internal stem loop; BS, branch site of the pre-mRNA.

I was able to locate the complete 200 kDa helicase unit of Brr2 in complex with the Prp8's Jab1 domain in the shallow slope of the B^{act} structure, close to the tip of the Prp8's En domain (Figure 4.13). Although Brr2 contains two tandemly organised helicase cassettes, only the N-terminal cassette (NC) actively unwinds the U4/U6 duplex during spliceosome activation (Santos, Jovin et al. 2012). Interestingly, the C-terminal tail of Prp8's Jab1 domain, which regulates Brr2 helicase activity (Maeder, Kutach et al. 2009, Mozaffari-Jovin, Wandersleben et al. 2013), appears to be docked in the RNA-binding channel of the NC, as previously observed in the Brr2-Jab1 co-crystal (Mozaffari-Jovin, Wandersleben et al. 2013); thus, after U4/U6 unwinding, Brr2 appears to be once again repressed (Nguyen, Galej et al. 2015, Nguyen, Galej et al. 2016, Wan, Yan et al. 2016).

Brr2 shares an interface with Prp8's En domain *via* the helix-loop-helix (HLH) domain of its N-terminal helicase cassette (NC); this interaction appears to be stabilized by the PPIase domain of the B^{act} protein NY-CO-10 (hCwc27), which connects Brr2's NC (*via* its HLH domain) and Prp8's NTD1 and En domain (Figure 4.13, B and C), consistent with our crosslinking data (Supplementary Table S1). As in the human tri-snRNP Brr2 is located close to Prp8's RT domain (Agafonov, Kastner et al. 2016), hBrr2 must undergo a dramatic repositioning after integration of the human tri-snRNP into the human spliceosome (see also below).

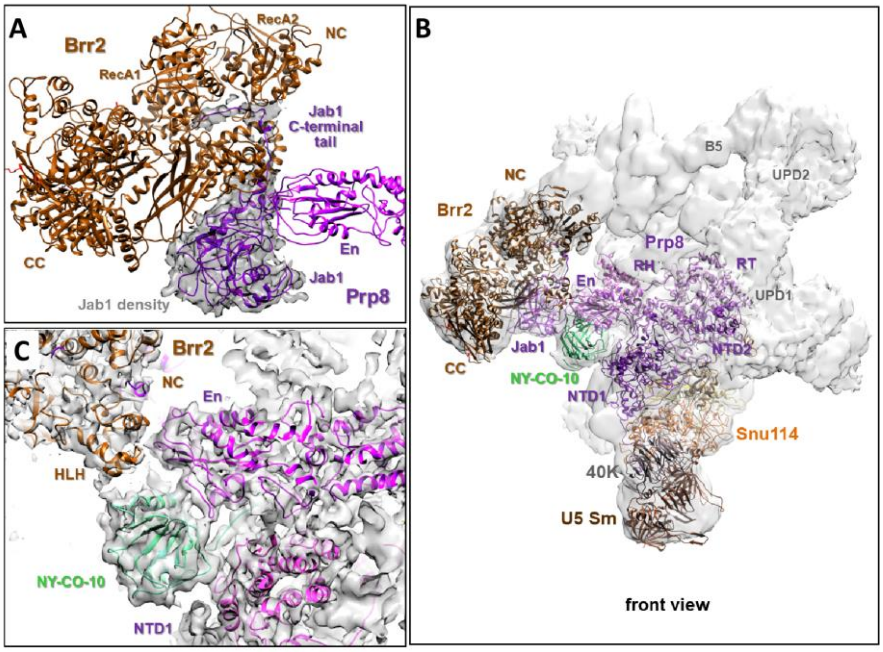


Figure 4.13: Location of the RNA helicase Brr2 in complex with Prp8's Jab1 domain and its interaction with the Prp8 and NY-CO-10 proteins

(A) Overall structure of hBrr2 in complex with Prp8's Jab1 domain. EM density map for the Jab1 domain is shown. **(B)** Brr2 is placed in the shallow slope of the B^{act} density, while Prp8, Snu114, U5 40K and NY-CO-10 are positioned in the central body, and U5 Sm proteins are located at the foot of the B^{act} density map. **(C)** Helix-loop-helix (HLH) domain of hBrr2 interacts with both Prp8's En domain and the PPIase domain of the NY-CO-10 protein.

4.4.2 Organization of U5 snRNA and the catalytic U2-U6 RNA network in the B^{act} spliceosome

Figure 4.14 shows the location of high density RNA helical elements within the B^{act} structure. The major stem-loop (SL) of U5 snRNA fits into the long RNA helical density element such that the 3'-terminal U5 snRNA Sm site is located at the foot of the 5.3 Å EM density map of B^{act}, whereas loop 1 of the U5 snRNA is found in the upper region of the elongated main body. U5 snRNA has essentially the same structure as in the human C* complex and makes multiple contacts with Prp8's NTD1 domain (Figure 4.14, C).

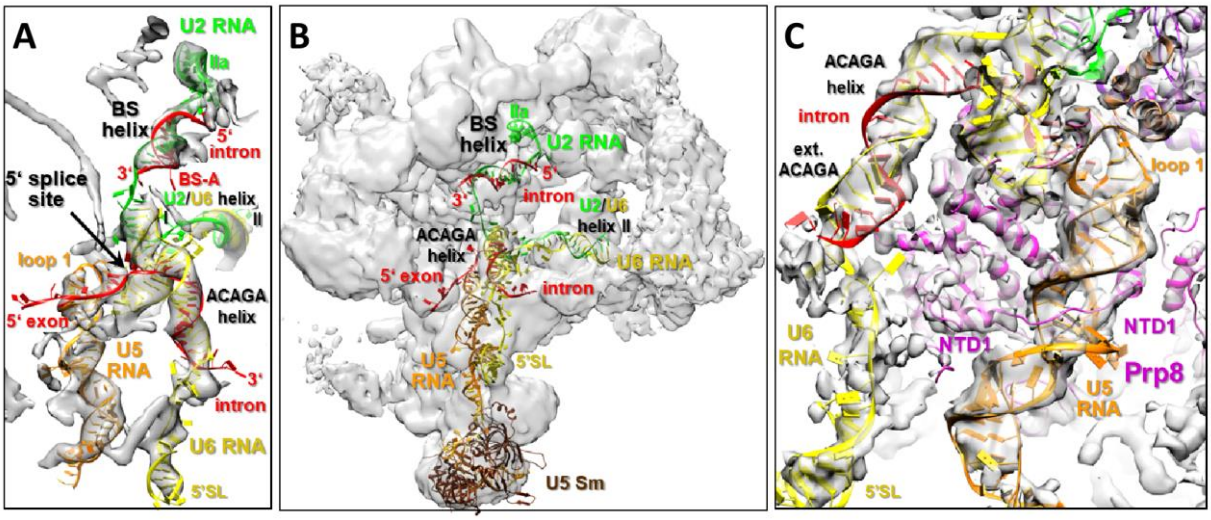


Figure 4.14: Structure of catalytic core elements in the human B^{act} complex

(A) Regions of the highest density within the central domain of the hB^{act} 5.3 Å EM density map. BS-A, branch site adenosine. **(B)** An overview of the RNA architecture in the hB^{act}. **(C)** A close-up view of the U5 snRNA surrounded by Prp8's NTD1 and U6 snRNA elements including the extended U6 ACAGA/5'ss helix and U6 5'SL.

Close to U5 RNA loop I a tripartite RNA density element is located, into which the rearranged U2/U6 catalytic RNA network can be placed. The ca. 5.3 Å resolution in this part of the B^{act} complex allows the unambiguous placement of the loop and kinked stem of the U6 snRNA internal stem loop (ISL), the two U2/U6 helices Ia and Ib (which are stacked on each other), and the sharp turn separating the ACAGA box helix of U6 snRNA and helix Ia of U2/U6 (Figure 4.15).

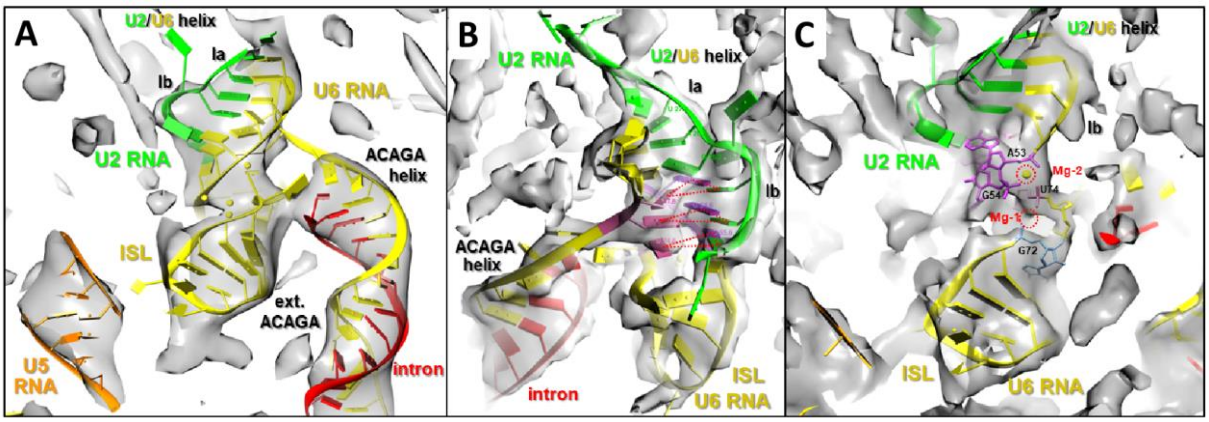


Figure 4.15: RNA network of the human B^{act} complex

(A) EM density map and fit of the U6 ISL, extended U6/5'ss helix, U2/U6 helix I and U5 stem I. **(B)** The helical stack of U2/U6 helix I. U6 snRNA nucleotides of the catalytic triplex are shown in purple. **(C)** Position of metal-coordinating U6 snRNA nucleotides A53, G54, G72 and U74. Putative catalytic Mg^{2+} ions (M1 and M2) positions are represented by red, dashed circles.

The topography of the RNA density is consistent with the existence of the catalytic triplex, as found in the hC* complex. The topography of U6 nts U74 and G72, as well as of A53 and G54, is such that they could potentially coordinate Mg^{2+} ions 1 and 2, respectively (which I cannot discern at 5.3 Å resolution). This would be consistent with biochemical studies showing metal ion coordination of the corresponding nts in yeast U6 (Fica, Tuttle et al. 2013). It is interesting to note, however, that there is density for a possible location of only the second catalytic Mg^{2+} ion (M2) but not for M1. This is similar to the situation in the yeast B^{act} complex, where M2 is already in place while density for M1 is missing (Yan, Wan et al. 2016). Thus, while the catalytic RNA triplex appears to be established in the hB^{act} complex, the catalytic centre is not yet in an active state. However, only minor rearrangements of the M1-coordinating nucleotides would be required to allow binding of M1 and catalysis of step 1 of splicing.

4.4.3 The 5'ss is located close to the catalytic centre, but is occluded by proteins Rnf113A and U2 SF3A2

In the B^{act} model, the RNA helix formed by the interaction of the U6 ACAGA box and the 5' end of the intron is located in the vicinity of the U2/U6 snRNA network (Figure 4.16); thus the 5'ss is already positioned close to the catalytic centre in the B^{act} complex. A continuation of the intron sequence of

the U6/intron helix (towards the 5'ss) leads to a density element that makes a U-turn close to the U2/U6 catalytic triplex and along U5 snRNA loop 1 (Figure 4.14, A). This density accommodates well the kinked 5'ss RNA element from the yeast B^{act} cryo-EM structure (Rauhut, Fabrizio et al. 2016, Yan, Wan et al. 2016) with the scissile bond of the pre-mRNA 5'ss oriented towards the catalytic centre (Figure 4.16). Furthermore, the GU dinucleotide at the intron's 5' end is in close contact with a short protein density element as shown in Figure 4.16. In the yeast B^{act} spliceosome, the 5'-terminal GU nts of the intron are in close contact with aromatic amino acids of the C-terminal zinc finger (ZnF) of Cwc24 (Rnf113A) and with N-terminal residues of the U2 SF3A protein SF3A2 (Prp11). Additionally, these protein-RNA interactions appear to be stabilized by Prp8's α-finger (Yan, Wan et al. 2016).

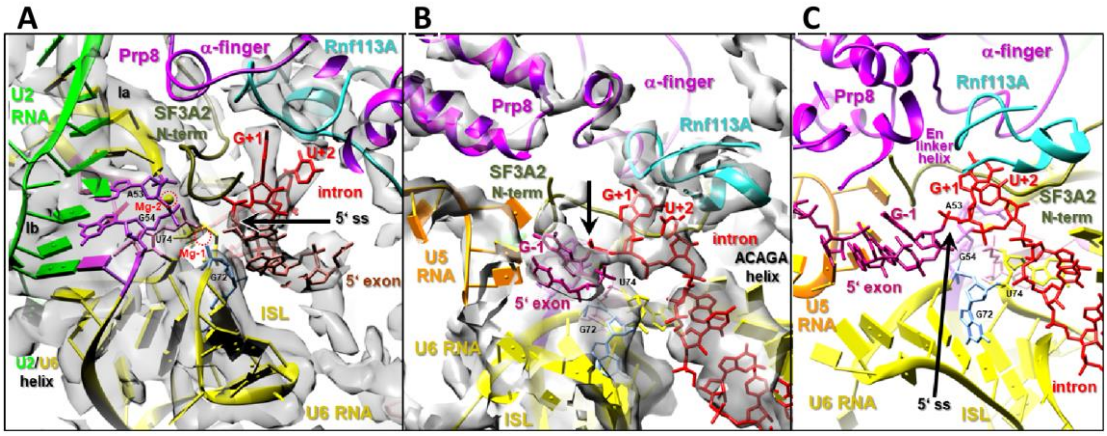


Figure 4.16: Organisation of the 5'ss in the human B^{act} complex

(A) EM density map of the RNA elements at the catalytic centre. The scissile phosphodiester bond is located close to the catalytic triplex of U6 snRNA, which is shown in purple. Putative location of the catalytic Mg^{2+} ions is indicated by red, dashed circles. **(B)** Fit of RNA and protein elements into EM density map. The 5' exon of the pre-mRNA is anchored to U5 snRNA, whereas the 5'ss is recognized by the ACAGA box of U6 snRNA. The GU dinucleotide of the 5'ss interacts with the Rnf113A and SF3A2 proteins. Prp8's α-finger is located close to the 5'ss **(C)** A close-up view of the 5'ss at the catalytic centre.

In the human B^{act} structure a homology model of the C-terminal ZnF of the evolutionarily conserved human Cwc24 protein (Rnf113A) can be fit into the density associated with the GU dinucleotide of the 5'ss. Major parts of SF3A2 are similarly organized as in the yeast B^{act} structure, and it is also possible to trace its N-terminal part in a corresponding density close to the 5'ss GU nts. Moreover, the α-finger of the human Prp8 protein is also located close to the ZnF of Rnf113A (Figure 4.16, B). Therefore, it is likely that also SF3A2 may contact the 5'ss in the human B^{act} structure. Taken together, the 5'ss GU nts of the pre-mRNA are located close to the catalytic centre of the human B^{act} structure but are occluded by interaction with the ZnF of Rnf113A (yCwc24) and probably also by residues of SF3A2, a situation which closely resembles that in the yeast B^{act} spliceosome. As these proteins not only spatially separate the 5'ss from the catalytic centre, but at the same time hinder access of the BS adenosine (the step 1 nucleophile) to the 5'ss, they must be rearranged to liberate the 5'ss for its final docking into the catalytic centre. Consistent with this idea, Cwc24 is dissociated

from the spliceosome during the Prp2-mediated catalytic activation, and SF3A proteins including SF3A2, are destabilized, i.e. structurally rearranged (Warkocki, Odenwälder et al. 2009, Lardelli, Thompson et al. 2010, Ohrt, Prior et al. 2012).

While it is possible to trace an additional ca. 10 more N-terminally located amino acids of Rnf113A in a corresponding density extending from the ZnF region towards the top domain of the B^{act} structure, there is no density for the RING finger domain of Rnf113A, close to the U2 SF3B3 protein (where the homologous domain of yCwc24 is positioned in the yB^{act} structure (Yan, Wan et al. 2016), presumably due to flexibility of this domain in the hB^{act} structure. It is interesting to note, however, that this region of the protein is close to the top domain and to the Brr2 helicase domain, as indicated by multiple crosslinks of Rnf113A to the U2 SF3B and Brr2 proteins (Supplementary Table S1).

It is also interesting to note that the N-terminal elongated part of NY-CO-10 is located adjacent to the part of Rnf113A (just N-terminal of its ZnF domain) and directly interacts with it, consistent with crosslinks between these protein regions, providing a possible molecular basis for the biochemical observation that NY-CO-10 dissociates from the spliceosome at the same time as Rnf113A (Bessonov, Anokhina et al. 2008).

4.4.4 The U2/U6 catalytic RNA network interacts closely with Prp8 and numerous B^{act} proteins

While the 3D structure of the rearranged U2/U6 RNA network in the spliceosome is very similar to the structure of the catalytic centre of the group II self-splicing intron, it can only form, and is only stable, in the presence of the closed form of Prp8 and numerous additional proteins of the B^{act} complex including the Prp19/CDC5 complex. Figure 4.17 shows how the U2/U6 RNA network fits tightly with one face into the active site of PRP8. For example, the NTD2 of PRP8 packs against the U2 snRNA part of U2/U6 helix Ib and is also in contact with the U6 ISL, and the linker between PRP8 NTD1 and 2 is located in the major groove of U6 ISL.

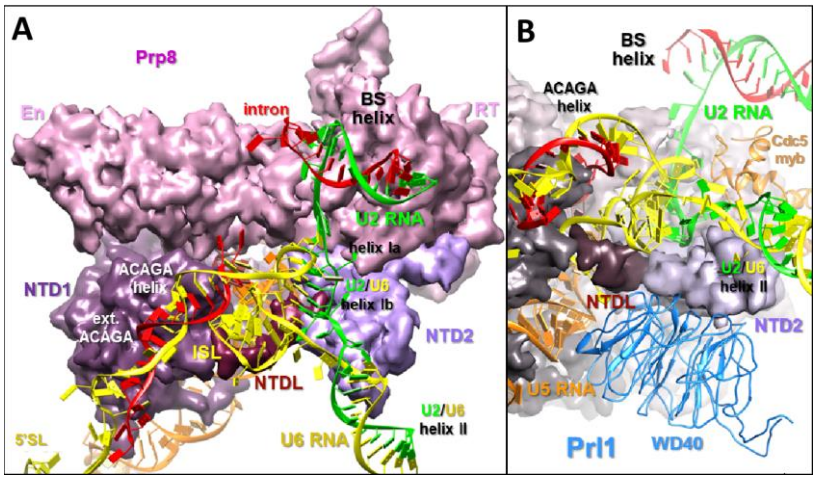


Figure 4.17: Proteins that stabilize the U2/U6 RNA network

(A) The catalytic U2/U6 RNA network docked in the active site pocket of Prp8 (shown as a space filling model) in the hB^{act} structure. The conserved interaction of the linker between NTD 1 and 2 (NTDL) with the U6 ISL is shown. **(B)** Myb domain of Cdc5 interacts with U2/U6 helices Ia and II. WD40 domain of Prl1 contacts the U5 stem Ic and U6 ISL.

In addition to Prp8, the following B^{act} proteins also interact with the U2/U6 RNA network. The WD40 domain of Prl1 interacts with U5 stem Ic and the U6 ISL loop (Figure 4.17, B; Figure 4.18, A). The N-terminal Myb domains of Cdc5 are attached to the Prp8 RT domain and interact with U2/U6 helices Ia and II (Figure 4.17, B; Figure 4.18, B). Interestingly, an α-helix of Cdc5 located closer to the C terminus, which runs along the U6 ACAGA/5'ss helix in the human C* complex, is not visible and, therefore, is probably very flexible in the B^{act} complex. N-terminal parts of Skip (hPrp45) are also in contact with the U6 ISL, while other parts of this protein meander around the Prl1 WD40 domain (Figure 4.18, B and C). While in the human C* complex, the N-terminal amino acids ~100-130 of Skip could be found, in the B^{act} they appear to be flexible and thus not visible (Figure 4.18, D). More C-terminal regions of Ad002 are organized in a similar manner as in C*, indicating that this protein is correctly integrated into the hB^{act} complex (Figure 4.18, A). Finally, the N-terminal HAT (Half a TPR) repeats of the HAT protein hSyf3 (Crnkl1) bind to the lower stem of the U6 ISL, while its HAT repeats 1-2 are bound to Prp8's NTD2 and HAT repeats 3-5 interact with U2/U6 helix II (Figure 4.18, C).

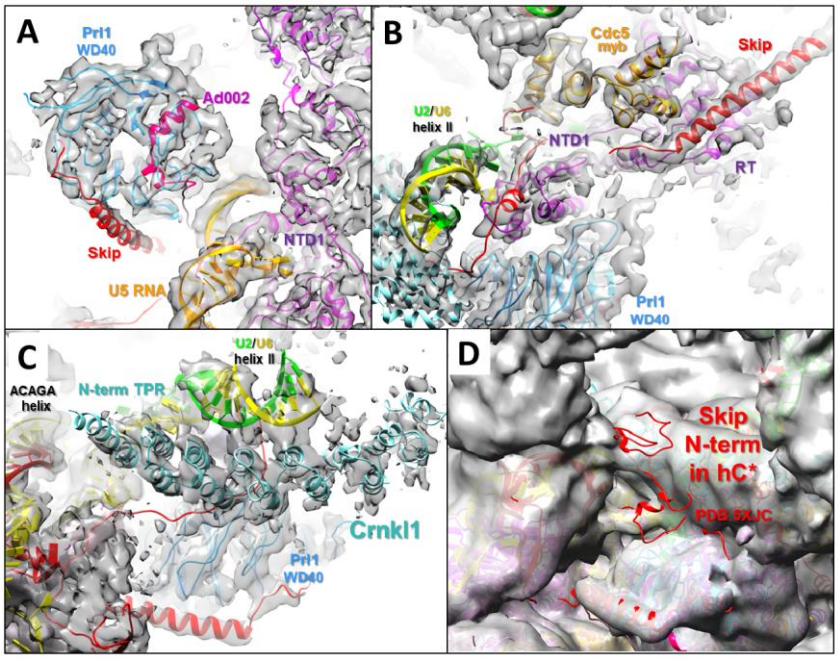


Figure 4.18: EM density map of proteins interacting with the U2/U6 catalytic RNA network

(A) A close-up view of Prl1's WD40 domain. The interactions of the Prl1's WD40 domain with Skip and Ad002 and U5 stem Ic are shown. **(B)** Fit of proteins surrounding U2/U6 helix II. The meandering path of Skip is shown. **(C)** A close-up view of the region between the U6 ACAGA/5'ss helix and U2/U6 helix II, bound to HAT repeats of Crnkl1. Skip bridges Crnkl1 and the WD40 domain of Prl1. **(D)** Skip protein from the hC* complex, placed into EM density map of hB^{act} (PDB 5XJC; Zhang, Yan et al. 2017). Missing density for the N-terminal region of Skip in the hB^{act} map suggests that this part of the protein is highly flexible.

4.4.5 An extended U6 ACAGA/5'ss helix and a rearranged 5'-terminal region of U6 snRNA in the human B^{act} complex

Interestingly, U6 nucleotides 30-40, which are immediately upstream of the ACAGA box, also form base pairs (including non-canonical ones) with additional intron nucleotides downstream of the 5'ss (Figure 4.19). This resulting helix is hereinafter termed the extended U6 ACAGA/5'ss helix. As an extended U6 ACAGA/5'ss helix is also present in the human C* complex (Bertram, Agafonov et al. 2017, Zhang, Yan et al. 2017), but was not observed in yeast spliceosomes (Rauhut, Fabrizio et al.

2016, Yan, Wan et al. 2016, Fica, Oubridge et al. 2017, Plaschka, Lin et al. 2017, Yan, Wan et al. 2017), it may be a structural feature of the spliceosome in higher eukaryotes only. Such extended helical elements may help to stabilize short RNA helices, such as the human U6 ACAGA box/5'ss helix, which consists of only a few canonical base pairs.

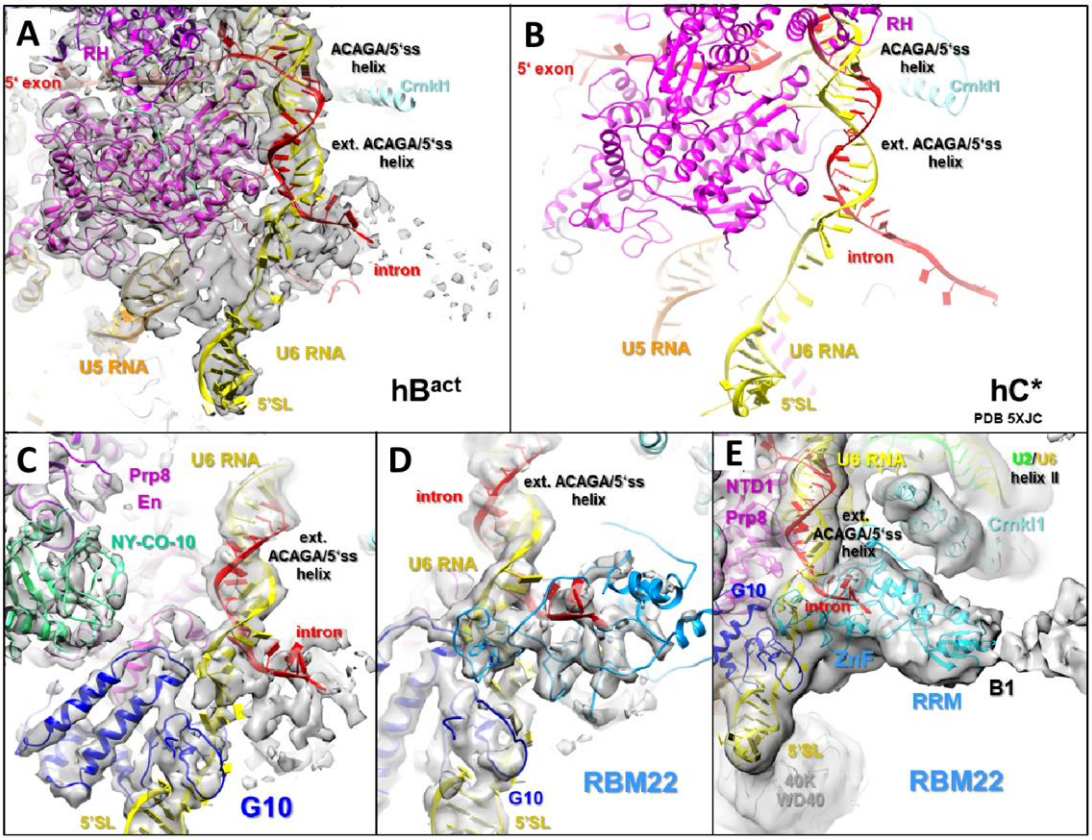


Figure 4.19: Structure and interaction partners of the U6 snRNA 5'-terminal region in hB^{act} complex

(A-B) Position and fit of the extended U6 ACAGA/5'ss helix and U6 5'SL in hB^{act} complex compared to hC* (PDB 5XJC) (Zhang, Yan et al. 2017). **(C)** Close-up view of the G10 protein, which tightly interacts with the 5'SL of U6. **(D)** N-terminal ZnF domains of Rbm22 directly bind to the extended U6 ACAGA/5'ss helix and lock the intron in a channel. **(E)** C-terminally located RRM domain of Rbm22 is part of the bridge 1 (B1).

In human B^{act}, the 5'-terminal SL of U6 is located in the lower part of the central domain in essentially the same position as in the human C* complex (Figure 4.19, A and B). Consistent with the similar position of the U6 snRNA 5' end in the human B^{act} and C* complexes, the human G10 (Bud31) protein also interacts with U6 snRNA close to the U6 5'SL and anchors it to the Prp8 NTD1 domain (Figure 4.19, C). The ZnF domains of Rbm22 also bind tightly to the single-stranded region of U6 snRNA on the opposite side of G10, close to the U6 5'SL (Figure 4.19, D). The C-terminally located RRM domain of Rbm22 reaches out to the UPD1 domain and is thus part of bridge 1 (B1) (Figure 4.19, E; and also below).

4.4.6 The 5' exon-binding channel and the location of the EJC-binding protein Cwc22

While the first three 3'-terminal nts of the 5' exon interact with nts of the U5 snRNA loop 1, I can trace an additional ca. 10 nts of the exon RNA that thread through a narrow channel; this channel connects the Prp8 En domain and the NTD1 (Figure 4.20, A). Consistent with it accommodating RNA, the channel contains many positively charged amino acids. The positioning of the Cwc22 protein at the end of the proposed exon-binding channel supports indirectly the postulated path of the 5' exon in the B^{act} complex. Cwc22 contains an N-terminal MIF4G-like domain, which in turn contains five HEAT repeats and a centrally positioned α-helical MA3 domain. Guided by crosslinks, the two domains of Cwc22 can be located on both sides of the exon channel (Figure 4.20 B and D). In the human spliceosome hCwc22's MIF4G domain binds the eIF4A3 helicase of the exon junction complex (EJC), and it therefore aids in the deposition of the EJC upstream of the spliced exon junction (Le Hir, Izaurralde et al. 2000). The position of the MIF4G domain at the end of the exon-binding channel is thus consistent with the function of this domain in depositing the EJC complex. Indeed, there is additional less well resolved density associated with the MIF4G domain, and this density may represent part of eIF4A3 of the EJC complex. At least it is possible to dock the RecA domains of eIF4A3 into this density (Figure 4.20, E). In the B^{act} model, the α-helical MA3 domain of Cwc22 is located next to the RT domain of Prp8 and thus lies on the opposite side of the exon-binding channel as the MIF4G domain.

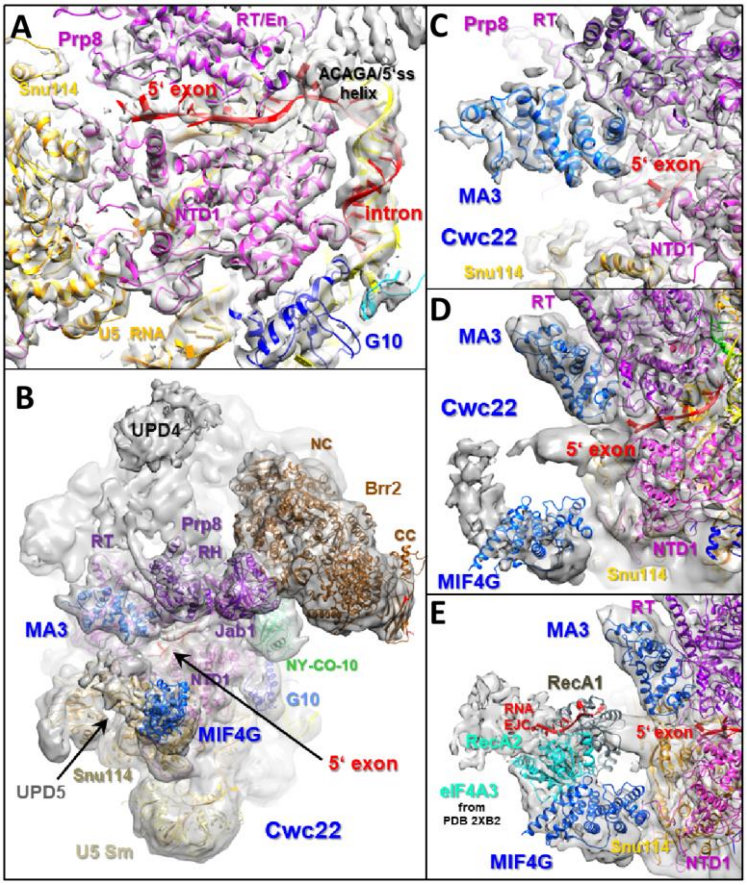


Figure 4.20: Organisation of proteins comprising the exon-binding channel

(A) A close-up view of the 5'ss and surrounding regions of the pre-mRNA. Fit of the Prp8 domains forming the exon-binding channel. **(B)** An overview of proteins forming the exon-binding channel in the hB^{act} structure. **(C)** A close-up view of the MA3 domain of Cwc22 and its fit in the B^{act} EM density. **(D)** Position of Cwc22 domains around the 5' exon. Cwc22's MA3 domain is bound to the Prp8 RT/En domain and its C-terminal MIF4G domain is attached to Snu114's domain 1 **(E)** Location of the Cwc22 MA3 and MIF4G domains and the RecA domains of the EJC helicase eIF4A3 (PDB 2XB2) (Buchwald, Ebert et al. 2010), placed in the UPD5 density element of the unmasked B^{act} model. Model of an extended 5' exon RNA (red) positioned relative to the RecA domains of eIF4A3 in the B^{act} complex.

4.4.7 The U2/U6 helix II adopts a unique conformation in the human B^{act} complex and is sandwiched between proteins Crnkl1 and Ppil2

Comparison of the conformation of the rearranged U2/U6 RNA network, the U2/U6 helix II and the 5'SL of U6 snRNA in the yeast (B^{act}, C, C*) and human (C*) cryo-EM models reveals that they are largely similar, except that the human C* complex has an extended U6 ACAGA helix (Zhang, Yan et al. 2017). In stark contrast, the U2/U6 RNA helix II adopts a unique conformation in the human B^{act} complex. While in the human C* and all other yeast complexes mentioned above, the end of U2/U6 helix II points upwards, in the hB^{act} complex it is not only situated somewhat lower in the spliceosome, but in addition it is rotated by about 90˚, with the end of the helix lying in a plane roughly perpendicular to the long axis of the central body. Furthermore, the protein interactions of U2/U6 helix II also differ significantly in the hB^{act} complex (Figure 4.21).

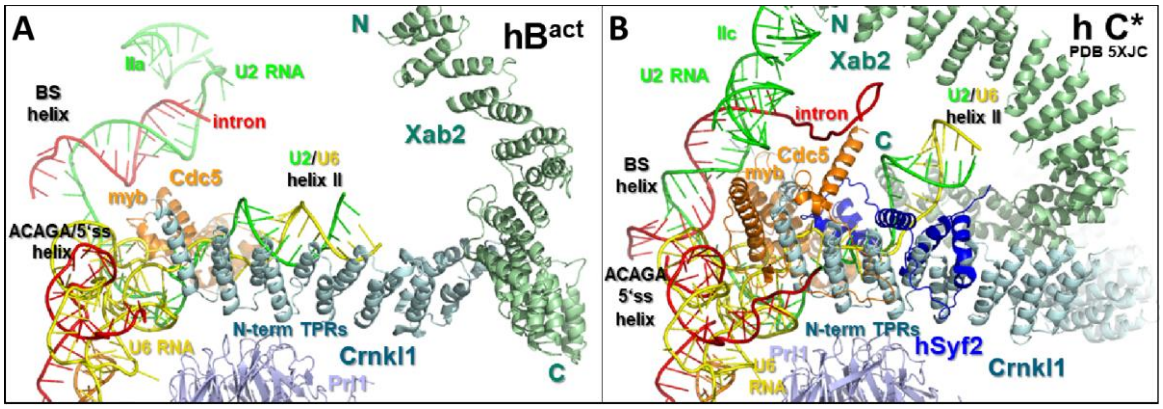


Figure 4.21: Comparison of the position/orientation of the U2/U6 helix II in the human B^{act} and C* complexes

Structural comparison of the U2/U6 helix II between the hB^{act} **(A)** and hC* **(B)** complexes. Compared to C* (PDB 5XJC) (Zhang, Yan et al. 2017), the U2/U6 helix II in the B^{act} complex is rotated and situated at a lower position, interacting directly with Crnkl1 and, thus, leaving no space for integration of the hSyf2 protein.

In the yeast B^{act} and all other later spliceosomal complexes, U2/U6 helix II is located close to the crossing point of the HAT proteins Clf1 (Syf3) and Syf1, but does not directly interact with these proteins. Instead, several α-helices of the Syf2 protein embrace U2/U6 helix II and are in direct contact with it (Figure 4.21, B). In contrast, in hB^{act} the HAT repeats 3 to 5 of Crnkl1 (Syf3) preceding the crossing point of Crnkl1 with Xab2 (Syf1), interact directly with one side of U2/U6 helix II, forming part of the B2 bridge. The other side of U2/U6 helix II is in contact with a globular density, which also has a thin connection to bridge 3. The X-ray structure of the PPIase domain of Ppil2 (Davis, Walker et al. 2010) can be placed into this density element, consistent with protein-protein crosslinks (Supplementary Table S1) (Figure 4.22, A).

In addition to the PPIase domain, Ppil2 contains two C-terminally located U box domains. Surprisingly, these are spatially separated from the PPIase domain by about 10 nm and are associated with domain IV of hSnu114 (U box 1) and an adjacent region of hPrp8 comprising α-helix

amino acids 1100 to 1120 (U box 2) (Figure 4.22); this is consistent with results from crosslinking (Supplementary Table S1). Unfortunately, the linker region of Ppil2 connecting its U boxes and PPIase domain cannot be traced in the hB^{act} cryo-EM density.

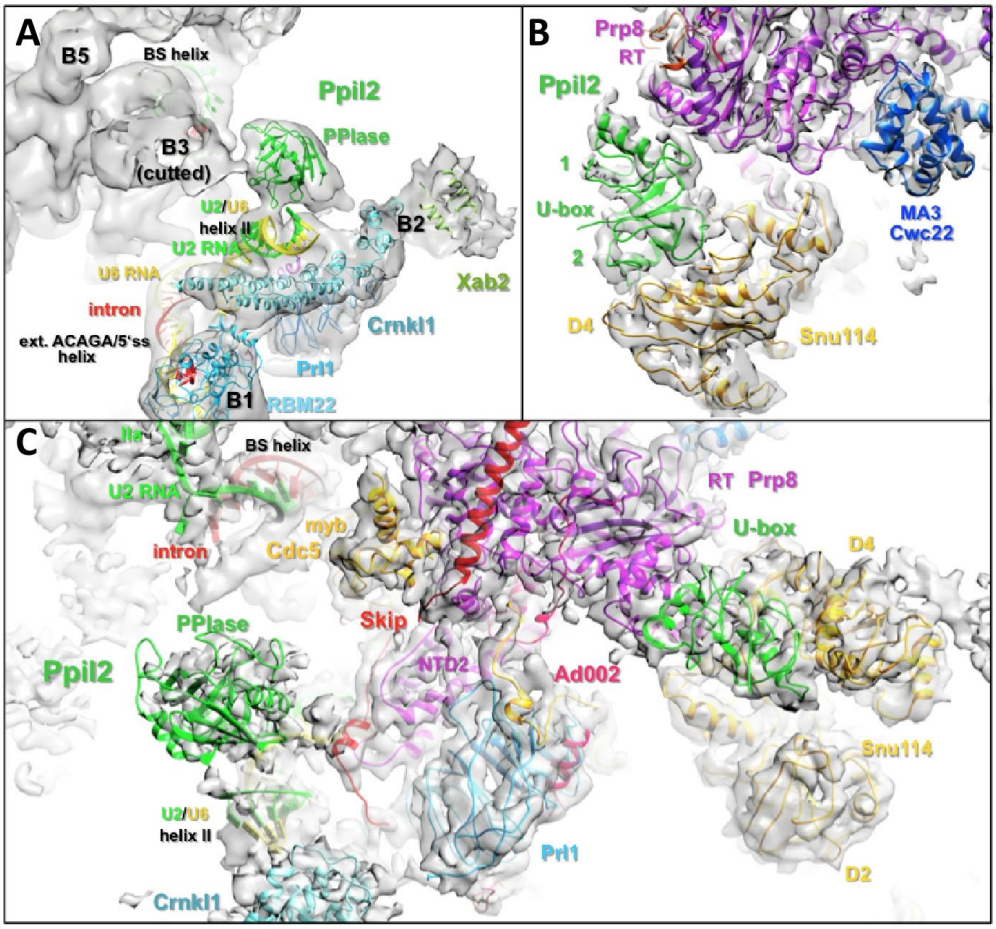


Figure 4.22: Location and interactions of Ppil2 in the hB^{act} complex

(A) The UPD3 PPIase domain of Ppil2 interacts with the upper side of the U2/U6 helix II. **(B)** Close-up view of the fit of the Ppil2 U box domains and their neighbouring proteins into the B^{act} EM density. **(C)** Overview of the location and interactions of spatially separated domains of Ppil2 in the B^{act} structure.

Owing to the unique conformation of the U2/U6 helix II and its sandwich between Crnkl1 and the PPIase domain of Ppil2, there is no space left for the integration of hSyf2 (GCIP interacting protein p29, GCIPp29) into the hB^{act} spliceosome. This explains, at least in part, why hSyf2 integrates into the human spliceosome not together with the other NTC proteins (as in yeast), but only concomitantly with or after catalytic activation of the hB^{act} complex (Agafonov, Deckert et al. 2011). As in the human C* complex, hSyf2 and the U2/U6 helix II have adopted essentially the same structure as in the corresponding yeast B^{act} or C complexes, it is likely that the recruitment of hSyf2 may be coordinated with the structural rearrangement of U2/U6 helix II during the transformation of the hB^{act} complex to the hC or hC* complex.

4.4.8 Organisation of the HAT proteins Crnkl1 and Xab2 and the RNA helicase Aquarius

The Crnkl1 (hSyf3) protein has altogether 17 HAT repeats, of which the N-terminal 5 repeats are attached to the central domain of hB^{act} and to U2/U6 helix II. HAT repeats 6 to 7 bind strongly to the HAT protein Xab2 (hSyf1). The C-terminal HAT repeats of Crnkl1 are not visible in the unmasked density of hB^{act} (Figure 4.22, A). The Xab2 protein forms on both sides of the crossing with Crnkl1 α-helical solenoids. While density for most of the N-terminal HAT repeats is present, it is visible only for a few of the HATs of the C-terminal wing of the curved solenoid (Figure 4.23).

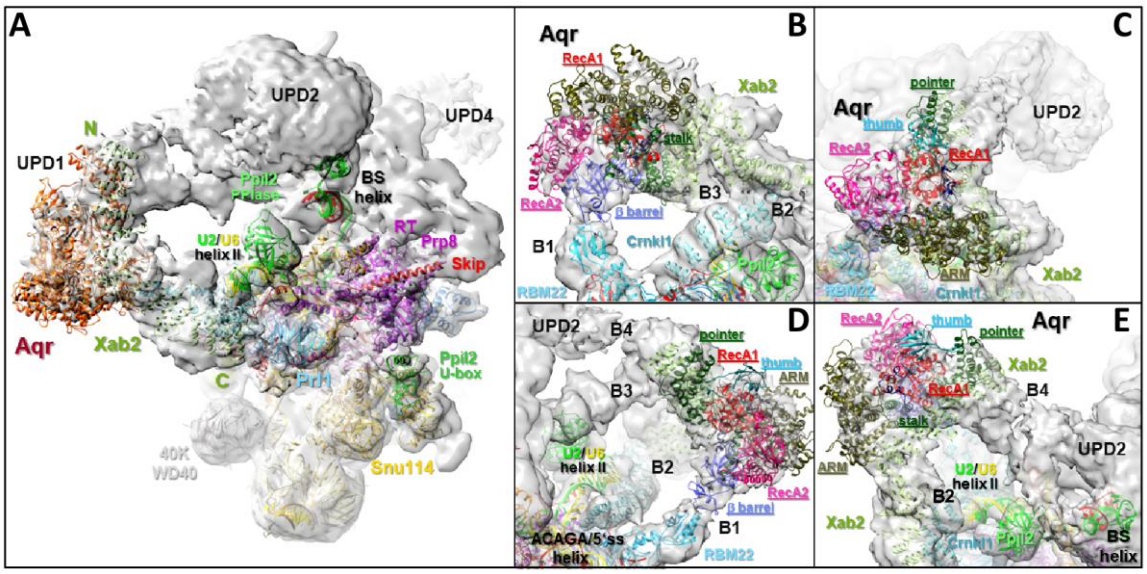


Figure 4.23: Organisation of the UPD1

(A) Overview of the location of UPD1, made up by Aquarius (Aqr) and the N-terminal region of its interaction partner Xab2. **(B-E)** EM density fit of the RNA helicase Aqr in the UPD1 density in the unmasked B^{act} model. UPD1 is connected to the central domain by two bridges (B1 and B2), to the top domain by B3 and to a globular UPD2 by B4.

The large elongated density element UPD1 is made up by the RNA helicase Aquarius (Aqr). The Xab2 N-terminal HAT repeats 1 to 8 share a large interface with the Aquarius protein (Figure 4.23 A). Compared with the crystal structure of the isolated Aqr protein (De, Sessonov et al. 2015), several domains of Aqr are rearranged in the B^{act} structure. Most prominent is the repositioning of the β-barrel of Aqr by about 2 nm from its position in the crystal structure, in the direction of the pointer domain in the hB^{act} structure (Figure 4.24). In the hB^{act} complex, the β-barrel is part of the bridge 1 and is probably in direct contact with the RRM of RBM22, consistent with protein crosslinks (Supplementary Table S1) (Figure 4.23, B and D). The massive ARM domain of Aqr is located at the opposite side of B4, which connects UPD1 and 2 and is partially comprised of the N-terminal region of Xab2 and the pointer that protrudes from the RecA1 domain of Aqr (Figure 4.23, C and E).

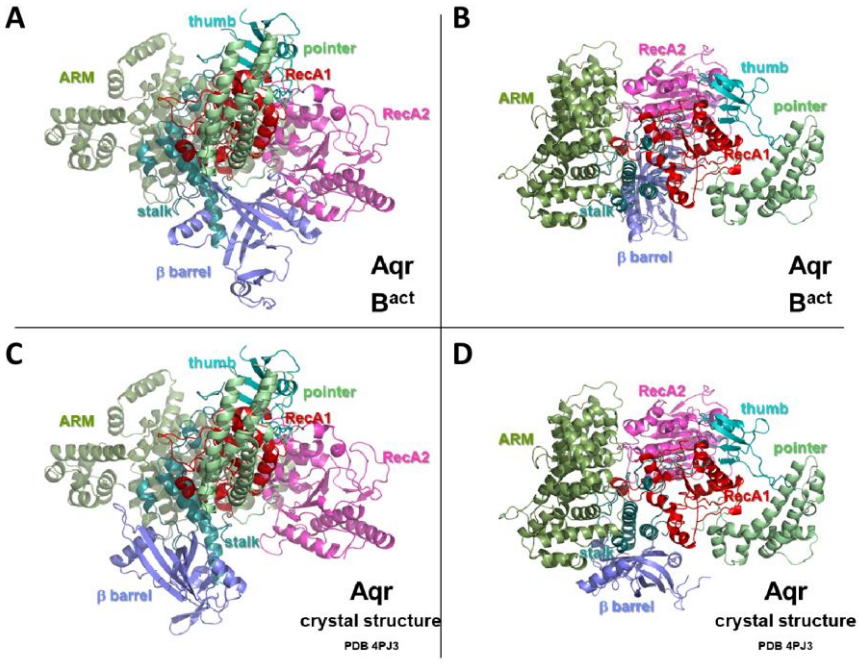


Figure 4.24: Comparison of the overall architecture of Aquarius in the human B^{act} complex and crystal structure

(A-B) Ribbon representation of Aquarius, viewed from two different orientations (rotated perpendicularly) in the human B^{act} complex and **(C-D)** crystal structure (PDB 4PJ3) (De, Bessonov et al. 2015).

4.4.9 The U2 SF3B protein complex is located at the top of hB^{act} and bridges Prp8 and Brr2

Major building blocks of the B^{act} spliceosome include the U2 snRNP SF3B and SF3A protein complexes, which are destabilized or entirely absent in later complexes of the catalytic cycle of the spliceosome. In human B and B^{act} spliceosomes, SF3A/B proteins are in contact with the pre-mRNA intron at or near the BS, and they stabilise the U2/BS helix. Of particular interest is the SF3B1 protein of the SF3B complex, which contains a C-terminal HEAT repeat domain and which can be crosslinked within spliceosomes to the BS, as well as to nucleotides upstream of the BS and along the entire 3' end of the intron (Gozani, Potashkin et al. 1998, Will, Schneider et al. 2001). Thus, SF3B1 should be close to the U2/BS helix and its location in the B^{act} complex structure should also help to define the path of the end of the intron (see also BS-3'ss, below). As described below, a major portion of SF3B is located in the top domain of the 5.3 Å cryo-EM structure of B^{act} complex. In the recently solved crystal structure of a protease-resistant human SF3B core complex (Cretu, Schmitzová et al. 2016) (Figure 4.25 B), the C-terminal domain of human SF3B1 is composed of 20 HEAT repeats and is organised as a relatively relaxed, right-handed superhelix with a pitch of about 65 Å and a mean diameter of about 100 Å. In the B^{act} structure, there are 20 well-defined density elements arranged in a ring-like manner, located above the Prp8 RT/En domain, into which all 20 HEAT repeats of the human SF3B1 crystal structure fit.

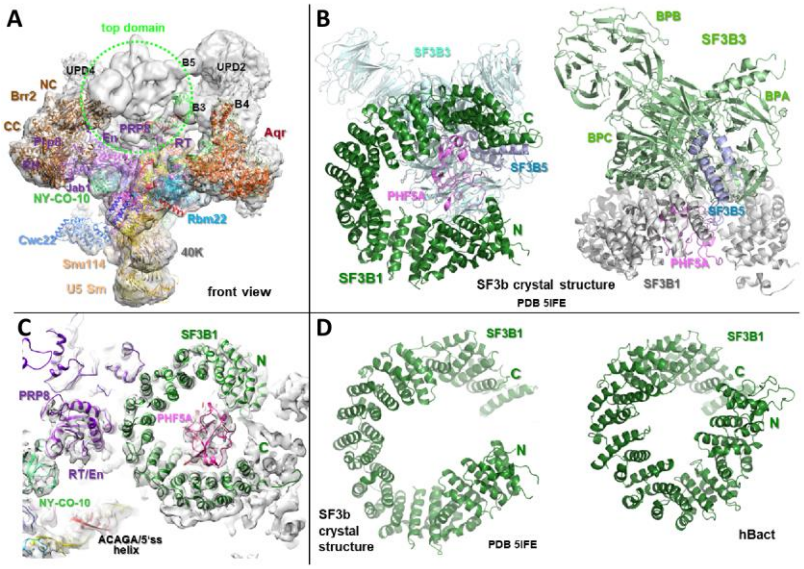


Figure 4.25: Comparison of the overall architecture of the SF3B protein complex in human B^{act} and the crystal structure of a protease-resistant human SF3B core complex

A) Overview of the location of SF3B proteins in the human B^{act} structure. **(B)** Ribbon representation of the SF3B core complex, viewed from two perpendicular orientations in the crystal structure of the isolated SF3B protein complex (PDB 5IFE) (Cretu, Schmitzová et al. 2016). **(C)** A close-up of the fit of the SF3B1 HEAT domain and PHF5A. **(D)** A comparison of the overall arrangement of the SF3B1 HEAT domain between the crystal structure of the human SF3B core complex and the hB^{act} complex (right).

However, the overall arrangement of the HEAT domain in the B^{act} complex differs significantly from the crystal structure; in B^{act} it appears much more condensed, with a lower pitch, such that the HEAT repeats form a ring-like structure with a diameter of ~60 Å. As the N-terminal HEAT repeats H1–H7 are rotated downwards, the first and last HEAT repeats lie nearly on top of each other, and are separated by only 18 Å. The SF3B1 HEAT domain is in contact with the top of the Prp8 RT/En domain through HEAT repeats 13 and 14, and the top of the RT domain, through its C-terminal α-helices (Figure 4.25).

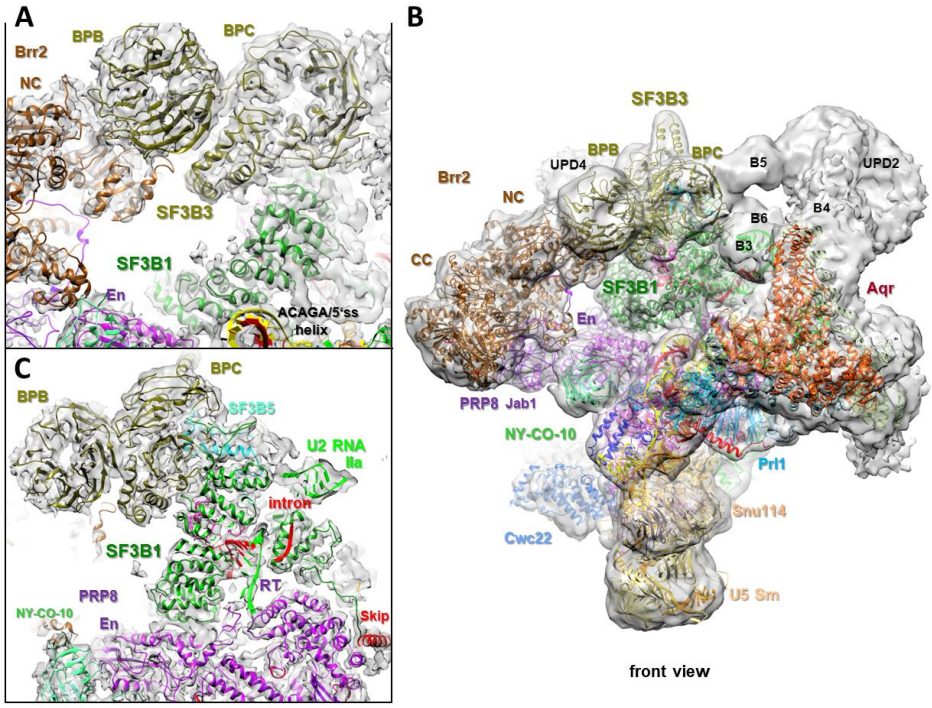


Figure 4.26: Localization of the SF3B complex proteins in the hB^{act} structure

(A) A close-up view of the fit of SF3B3 in the hB^{act} EM density map. Interface of SF3B3's β-propeller BPB with RecA domains of Brr2's NC cassette. **(B)** Front view of the hB^{act} complex and location of SF3B proteins. **(C)** The SF3B1 HEAT domain is located on top of the Prp8 RT/En domain. SF3B5 tightly interacts with SF3B3.

In the human SF3B core complex crystal structure, SF3B3 adopts an intertwined three-propeller (WD40) cluster form, which contains two tightly coupled β propellers (BPA+BPC) with a large clam-shaped pocket between them, into which the three-helix structure of SF3B5 (SF3B10) is tightly bound (Figure 4.26). Both SF3B3 and PHF5A have a main interface with SF3B1's C-terminal HEAT repeats, but they also contact N-terminal HEAT repeats (Figure 4.25; Figure 4.26). The intertwined β-propellers BPA und BPC of SF3B3 are located at the top of the B^{act} model and are aligned with the long axis of the main body. The bottom part of BPB faces Brr2's NC cassette and has a major interface with the RecA domain of the latter (Figure 4.26). In addition, BPA und BPC also contact Brr2's helicase domain. Thus, SF3B1's's HEAT and SF3B3's WD40 β-propeller domains bridge Prp8 and Brr2 in the B^{act} complex.

4.4.10 Path of the intron's 3' region across the HEAT domain

The PM5-10 construct, which was used for the assembly and purification of the human B^{act} complex, contains 10 nts downstream of the BS. The cryo-EM map of the B^{act} structure reveals the presence of a density element that could accommodate a single-stranded RNA, starting at the bottom part of PHF5A, where the 3'-terminal nucleotide of the U2/BS helix is positioned, along the concave side of HEAT repeats 3–5 and ending just short of the bottom of HEAT repeats 6 and 7 (Figure 4.27). The length of this density element is sufficient to accommodate 10 nts in a stretched conformation. If the intron RNA were continued it would pass along the RES protein CGI-79 (Snu17) (see below), very similar to the situation in the yeast B^{act} complex (Rauhut, Fabrizio et al. 2016, Yan, Wan et al. 2016)

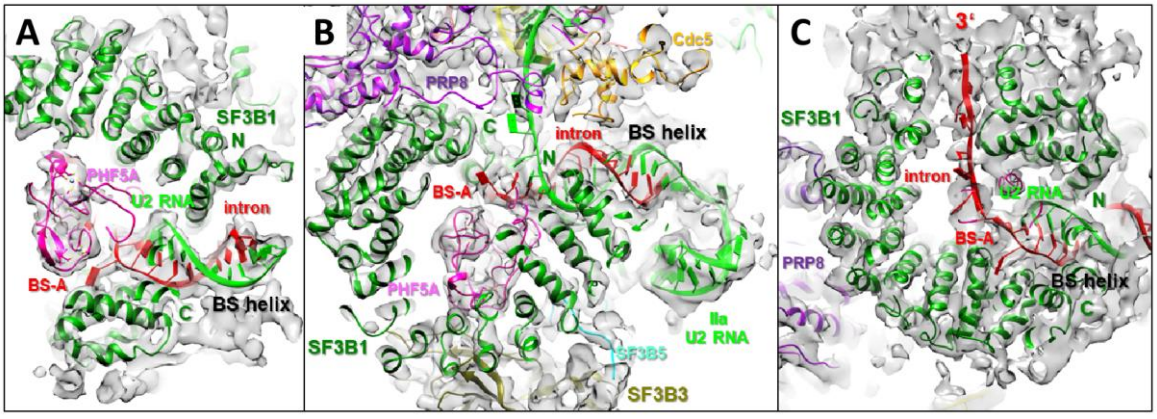


Figure 4.27: Location of the BS/U2 helix and path of the intron's 3' region

(A) The BS/U2 helix could be placed into the density element located at the opening between the terminal HEAT repeats of SF3B1. **(B)** Upstream of the BS, the intron is base-paired to U2 snRNA, forming an extended BS/U2 RNA helix. **(C)** Close-up view of the intron path downstream of the BS.

4.4.11 The BS/U2 RNA helix is sequestered between the SF3B1 N- and C-terminal HEAT repeats, ca. 20 Å away from the RNA catalytic centre

In the hB^{act} model, the pre-mRNA BS/U2 helix (from human C*) could be placed into the density element located at the opening between the terminal HEAT repeats of SF3B1 (Figure 4.27). The BS RNA faces the C-terminal HEAT repeats and the density of the BS adenosine (BS-A) is clearly visible. The latter is flipped out of the branch helix and is located in a protein pocket built by the β-helices of HEAT repeats H15–H16 and capped by PHF5A (SF3B14b). The close contact of HEAT repeats H15–H16 and PHF5A with the BS RNA is consistent with the inaccessibility of these nucleotides in the B^{act} complex (including the bulged-out adenosine residue) to chemical modification (P. Bao and K. Hartmuth, personal communication). On the other side of the BS/U2 helix, the HEAT repeats 1 and 2 pack against the U2 snRNA. Upstream of the BS, ca. 10 nts of the intron are base-paired to U2 snRNA, including non-canonical base pairs, forming an extended BS/U2 RNA helix (Figure 4.27).

The orientation of the BS/U2 helix is such that the 5'-terminal nt of U2 is ca. 20 Å above the 3'-terminal U2 nt of U2/U6 helix Ia (Figure 4.28). This is the greatest distance that can be bridged by the 4-nt-long U2 linker that connects these two helices. The 2'-hydroxyl of the bulged BS-A is spatially separated from the scissile bond of the 5'ss by about 5 nm (Figure 4.28, B). The 3'-terminal U2 nt of the BS/U2 helix lies at the upper end of the HEAT repeat H1 and is connected through a 9-nts-long linker to the U2 stem-loop IIa (Figure 4.28, C). In summary, these results show that the BS/U2 RNA helix is held at the greatest possible distance from the catalytic RNA centre by the terminal repeats of SF3B1's HEAT domain. At the same time, the BS-A is occluded in a SF3B protein pocket and thus cannot attack the 5'ss at this stage.

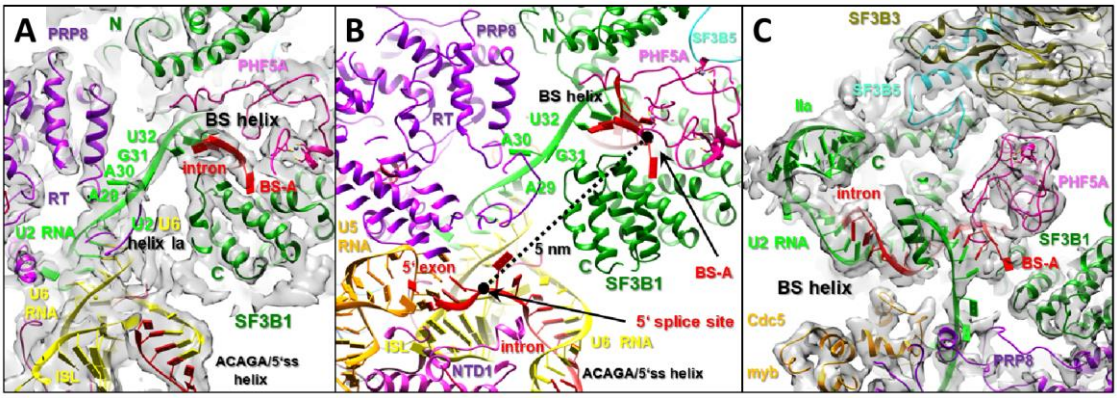


Figure 4.28: The BS/U2 helix is spatially separated from the active site

(A) 5'-terminal nt of U2 is ca. 20 Å above the 3'-terminal U2 nt of U2/U6 helix Ia. (B) The BS-A is spatially separated from the scissile bond of the 5'ss by about 5 nm. (C) The 3'-terminal U2 nt of the BS/U2 helix is located at the upper end of the HEAT repeat H1 and is connected via a 9-nts-long linker to U2 SL IIa.

4.4.12 The U2 SF3B6 (p14) protein is spatially separated from the BS adenosine

In spliceosomal A and B complexes the SF3B6 RRM domain protein has been shown to crosslink to the BS-A, suggesting that it is involved in the BS recognition in the early spliceosome (Will, Schneider et al. 2001). The SF3B6 was shown to bind to the N-terminal region of SF3B1, comprising amino acids (aa) 373-415; moreover, SF3B6 in complex with a corresponding SF3B1 peptide could be crystallized (Schellenberg, Edwards et al. 2006, Spadaccini, Reidt et al. 2006). In the B^{act} cryo-EM structure no density close to the U2/BS helix was found, which could accommodate the SF3B6 protein. Instead it is located at the convex side of SF3B1's HEAT domain, close to the N-terminal HEAT repeats, where an α-helical region of SF3B1 bridges SF3B6 to the HEAT repeats 1 and 2, consistent with protein crosslinks (Supplementary Table S1). A short α-helix spanning aa Pro398 to Val414 of SF3B1 is firmly attached to the central part of the SF3B6 domain (Figure 4.29, B). The putative adenosine-binding region of SF3B6 (Schellenberg, Edwards et al. 2006) is spatially separated from the flipped-out BS-A by about 5 nm (Figure 4.29, D). Thus, if SF3B6 is indeed close to the BS in early spliceosomes, a major rearrangement of SF3B6 together with its SF3B1 binding region has to occur during the transition of the early spliceosome to the B^{act} complex.

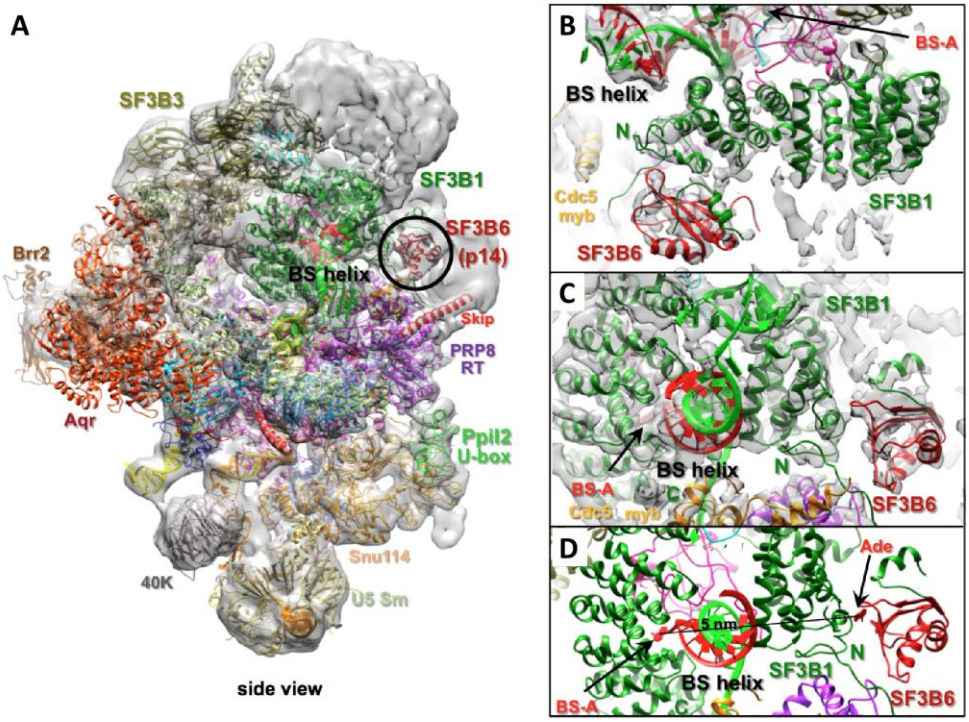


Figure 4.29: Localization of the SF3B6 protein

(A) Overview of the SF3B6 position in the hB^{act} complex. **(B)** A close-up of the fit of SF3B6 into the hB^{act} EM density map. SF3B6 is attached to SF3B1 at the convex side of SF3B1's HEAT domain. **(C)** EM density map and **(D)** ribbon representation of the SF3B protein complex, showing the adenosine-binding region (Ade) of SF3B6, spatially separated from the flipped-out BS-A by about 5 nm.

4.4.13 RES proteins and RNA helicase Prp2 are in contact with distinct regions of SF3B1's HEAT repeats on the side opposite of the BS/U2 RNA helix

The *S. cerevisiae* proteins Snu17, Bud13 and Pml1 form a heteromeric complex (RES) that is important both for splicing and for the retention of unspliced pre-mRNA in the nucleus (Gottschalk, Bartels et al. 2001, Dziembowski, Ventura et al. 2004). Recently, a solution structure of the yeast RES core complex consisting of the RRM protein Snu17, the C-terminal region of Bud13 and the N-terminal region Pml1, was determined (Wysoczanski, Schneider et al. 2014). In view of the high evolutionary conservation of human RES proteins, it is likely that they are also similarly organized as a hetero-trimeric complex as the yeast counterpart. Indeed, guided by crosslinks, it was possible to fit the RES core structure below the SF3B1 HEAT ring, close to HEAT repeats H7 to H9, where RbmX2 (hSnu17) interacts with the bottom edge of H6–H8 (Figure 4.30) (Supplementary Table S1). The RES core structure contains a short α-helix of hBud13, which in the B^{act} complex points downwards from RbmX2 toward the RT domain, and an additional region of hBud13 (aa 592-618), which was modelled as an α-helix, extends to the lower part of Prp8's RT domain (Figure 4.30). The C-terminal FHA domain of Snip1 (hPml1) also lies next to the RT domain (Figure 4.30). A long alpha-helical region of Skip is associated with Snip1. These observations are consistent with the idea that the RES complex plays a part in bridging the SF3B1 HEAT and Prp8 RT domains.

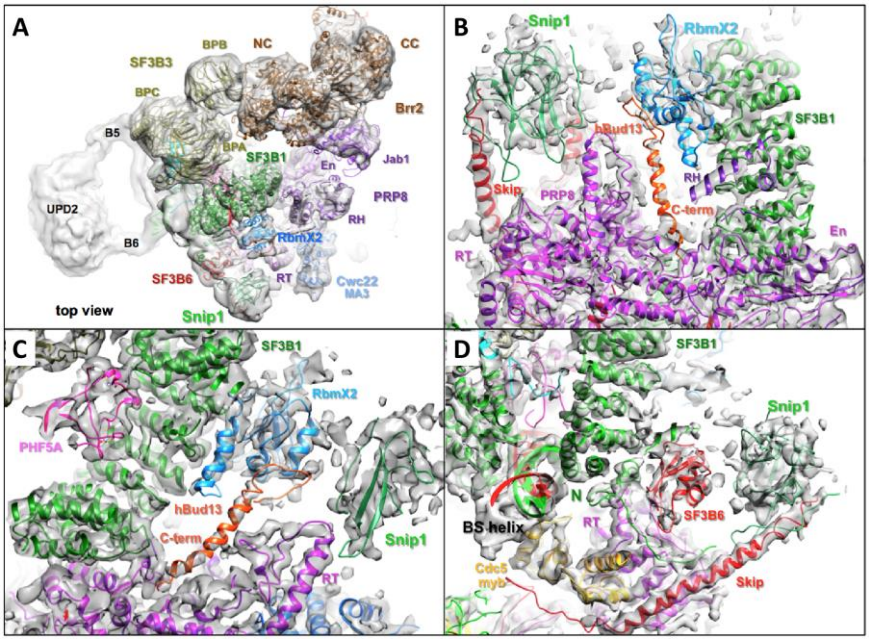
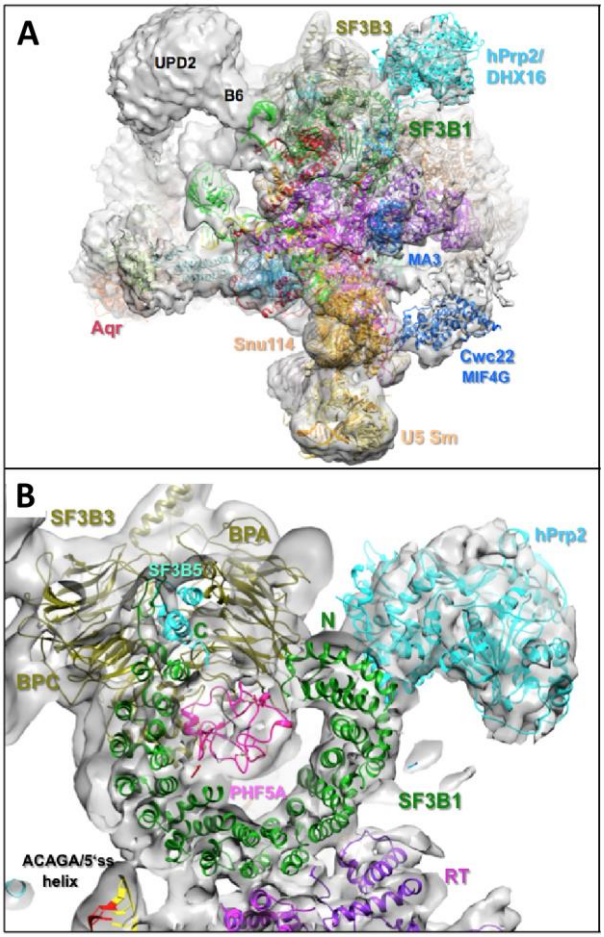


Figure 4.30: The location of RES proteins

(A) Overview of the position of the RES complex in the human B^{act} structure. **(B)** Density fit of the RES core complex on the convex side of SF3B1's HEAT domain. **(C)** The hBud13 C-terminal helix occupies a density element that continues down to Prp8's RT/En domain. (**D**) A long alpha-helical region of Skip is associated with Snip1.

Interestingly, in the hB^{act} 3D model, the RNA helicase hPrp2 interacts closely (through its C-terminal OB fold domain) with the upper part of HEAT repeats H7–H9 of SF3B1, and its RecA domains point away from the HEAT domain, facing the RES complex (Figure 4.31). This is consistent with the C-terminal domain playing an essential role in mediating Prp2's interaction with the spliceosome (Edwalds-Gilbert, Kim et al. 2004). It is interesting to note that hPrp2 is stably bound to the HEAT



domain even in the absence of a long 3'-end of the intron, demonstrating that RNA-protein interactions are not essential for complex formation of hPrp2 with the human spliceosome. Similar findings have been made for the yeast spliceosome (Warkocki, Odenwälder et al. 2009). On the basis of protein-protein crosslinks, Prp2's cofactor GPKOW, appears to be located between Prp2 and the RES complex (Supplementary Table S1).

Figure 4.31: Location of the hPrp2 RNA helicase

(A) Back view of the hB^{act} complex and fit of hPrp2 in the UPD4 of the B^{act} model. **(B)** hPrp2 interacts closely with the upper part of HEAT repeats H7–H9 of SF3B1. Organisation of the 3'-domain of U2 snRNP and the intricate network of SF3A complex proteins with other parts of the spliceosome

4.4.14 Organization and interactions of the U2 SF3A protein complex

The 3'-terminal Sm core structure of U2 snRNP, with its associated A' and B'' heterodimer proteins, is spatially separated from the central SF3B protein complex and is located together with a large part of the SF3A protein complex in the UPD2 domain of the unmasked hB^{act} cryo-EM model. It is connected by several bridges to the SF3B-containing top domain and to UPD1. Starting from the 3'-terminal U2 nt of the extended U2/BS helix, the U2 snRNA nts can be traced in a density close to the convex side of the HEAT domain that runs up to a density, which accommodates the U2 stem-loop IIa. U2 stem-loop IIb then forms the bridge B6, connecting the U2 Sm site. Proteins of the U2 SF3A complex, and the SF3B2 and SF3B4 proteins (hCus1 and hHsh49), form an intricate protein network, connecting numerous parts of the spliceosome including the U2 snRNP 3' domain.

Bridge B5 contains two RRM domains of the U2 B" protein, with RRM2 being close to the WD40-C domain of SF3B3. Thus, consistent with the finding of crosslinks between RRM2 of U2 B" and the C-terminal domain of SF3A3, which in turn crosslinks to SF3B3's WD40-C domain, SF3A3's C-terminal domain can be placed into the density between RRM2 of U2B" and the WD40-C domain of SF3B3 (Figure 4.32, B; Figure 4.33, B).

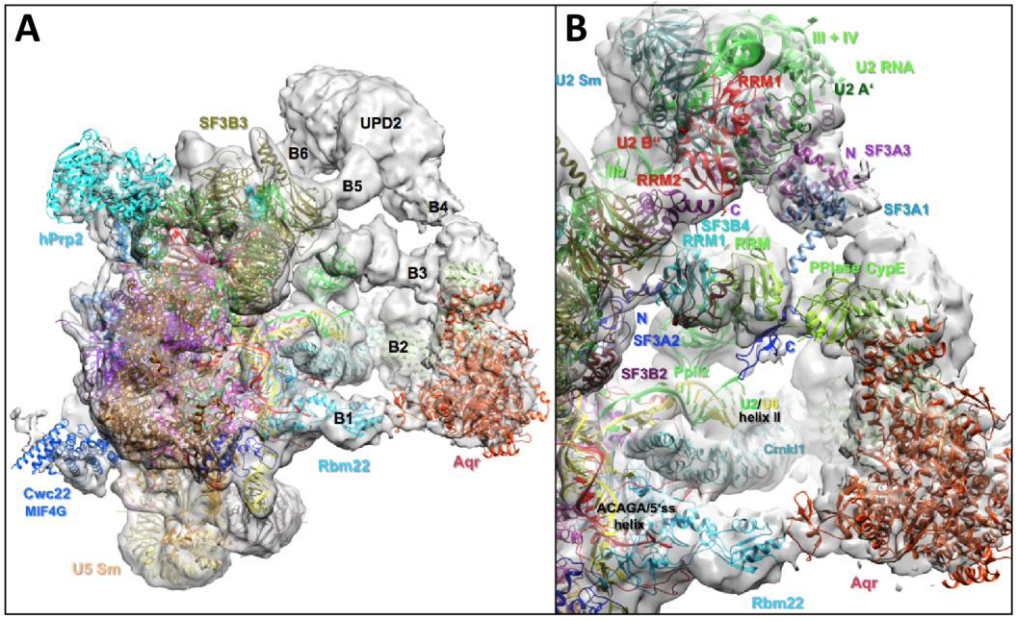


Figure 4.32: Organization of U2 SF3A protein complex

(A) Globular UPD2 is connected to the U2 SF3B complex located in the top domain *via* thin bridges B5 and B6 and to the UPD1, containing Aqr and Xab2, *via* B4. **(B)** Top view of the unmasked B^{act} model and density elements in UPD2 attributed to the U2 Sm core RNP and part of the SF3A protein complex. Main connections are provided between the 3'-terminal U2 snRNP domain and the top domain of B^{act} through the SF3A proteins, and SF3B2 and SF3B4.

Guided by crosslinks, a large part of the crystal structure of the SF3A core comprising the N-terminal region of SF3A3 and an α-helical domain of SF3A1 can be placed into a density element at the top of UPD2 (Figure 4.33). Here, one part of SF3A3 (aa 124 to 353) interacts with Sm proteins and stem-loop III of U2 snRNA, while the N-terminal part (aa 1 to 123) binds to the U2A' protein and together with the helical region of SF3A1, is connected to the N-terminal HAT repeats of Xab2 in bridge B4 (Figure 4.33). Evidence from crosslinks indicates that hIsy1 may possibly also be part of this bridge.

Beneath bridge B4 runs an additional bridge B3, which connects the N-terminal HAT repeats 2-4 of Xab2, through protein CypE, to the RRM1 of SF3B4, whereby CypE's PPIase domain is attached to Xab2 and its RRM domain interacts with RRM1 of SF3B4, which in turn also interacts with a C-terminally located domain of SF3B4 (Figure 4.33, A). As mentioned above, the PPIase domain of Ppil2 appears to interconnect bridge B3 *via* contacts with SF3B4's RRM1 domain with U2/U6 helix II (Figure 4.22). A SF3A2 domain is also part of bridge B3 and is located close to the linker region between CypE's PPIase and RRM domain (Figure 4.33). Finally, the central ZnF domain of SF3A2 is associated with the C-terminal HEAT repeats of SF3B1 and contacts closely the backbone of the extended U2/BS helix (Figure 4.33), while its N-terminal part reaches out to the 5'ss GU nts (see above), a setting which closely mirrors the organization and RNA contacts of its counterpart Prp11 in the yeast B^{act} complex (Yan, Wan et al. 2016).

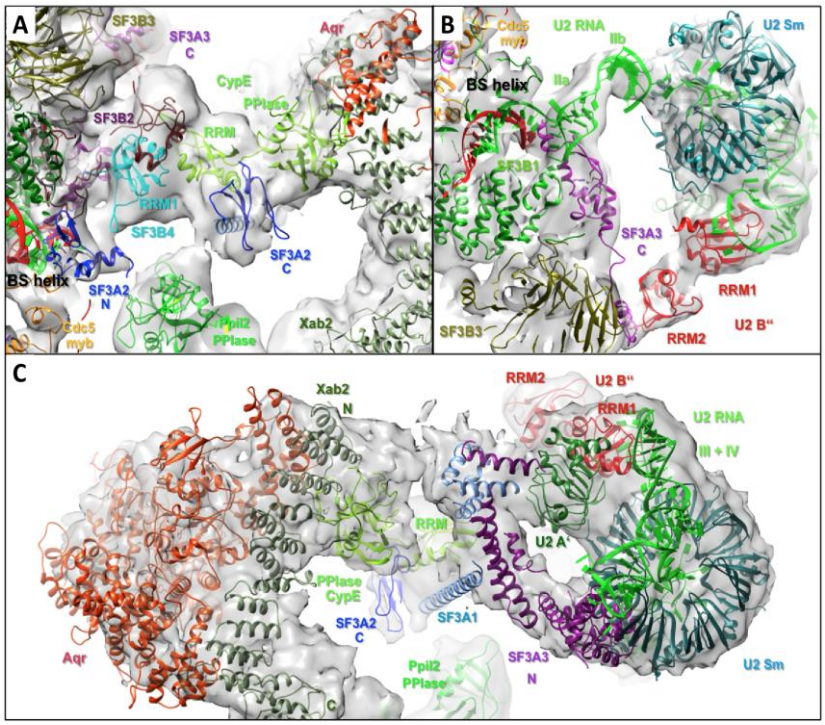


Figure 4.33: SF3A/B proteins bridge UPD1 and 2 to the top domain and also to each other

(A) A close-up view of the fit of proteins in bridge B3, connecting UPD1 to the top domain. B3 contains the CypE protein, whose PPIase domain is attached to Xab2, whereby its RRM domain interacts with the RRM domain of SF3B4. The C-terminally located region of SF3A2, which is also part of B3, is placed between domains of CypE, while its N-terminal domain interacts with SF3B1 and the extended U2/BS helix. **(B)** UPD2 is attached to the top domain *via* the 3' end of U2 snRNA (B6) and the two RRM domains of U2 B". **(C)** Organization of UPD2, comprised of the 3' end of U2 snRNA, the U2 Sm ring, U2 proteins A' and B", as well as parts of the SF3A complex proteins SF3A1 and SF3A3.

Similar to the yeast B complex structure, a large part of the SF3B2 protein forms an extended polypeptide domain that is associated with the C-terminal HEAT repeats of SF3B1 (Figure 4.34). A globular region of SF3B2 forms an intricate complex with the RRM1 domain of SF3B4 and an extended C-terminally located α-helical part of SF3A3 (Figure 4.34).

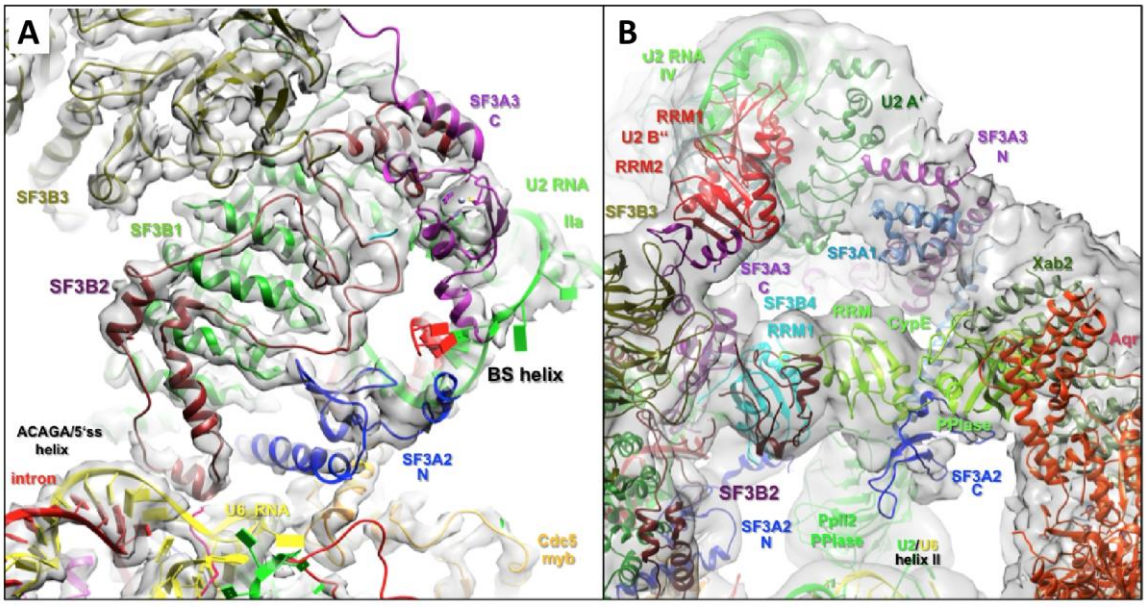


Figure 4.34: A close-up view on the SF3B2 structure in the human B^{act} complex

(A) SF3B2 forms an extended polypeptide region that folds onto SF3B1 HEAT repeats H16–H20 and the SF3B3 BPC domain. **(B)** A globular region of SF3B2 interacts with SF3B4's RRM1 and nearby ZnF domains of SF3A2 and SF3A3.

4.5 Cryo-EM structure of form B of the B^{act} complex

The final 3D classification of the hB^{act} particles yielded a second major form of the hB^{act} complex at an average resolution of 8.1 Å. Close inspection of this B^{act} form reveals that the structure of most of the central domains, the U2 SF3B core structure, hBrr2 and most of the 3' domain of U2 snRNP is very similar to that in form A of hB^{act}. This also applies to the organisation of the U2/U6 catalytic RNP core, including the missing density for the catalytic metal M1.

However, forms A and B also show significant differences in other regions of the structure. The most striking difference is the presence of an elongated density element at the side of the hB^{act} main body in form B, which is not visible in form A. This elongated density comprises a helical bundle, which is formed by the coiled-coil regions of the four copies of hPrp19, Spf27 and the C-terminal helical part of Cdc5. The four Prp19 WD40 domains are not visible, presumably because they are highly flexible (Figure 4.35).

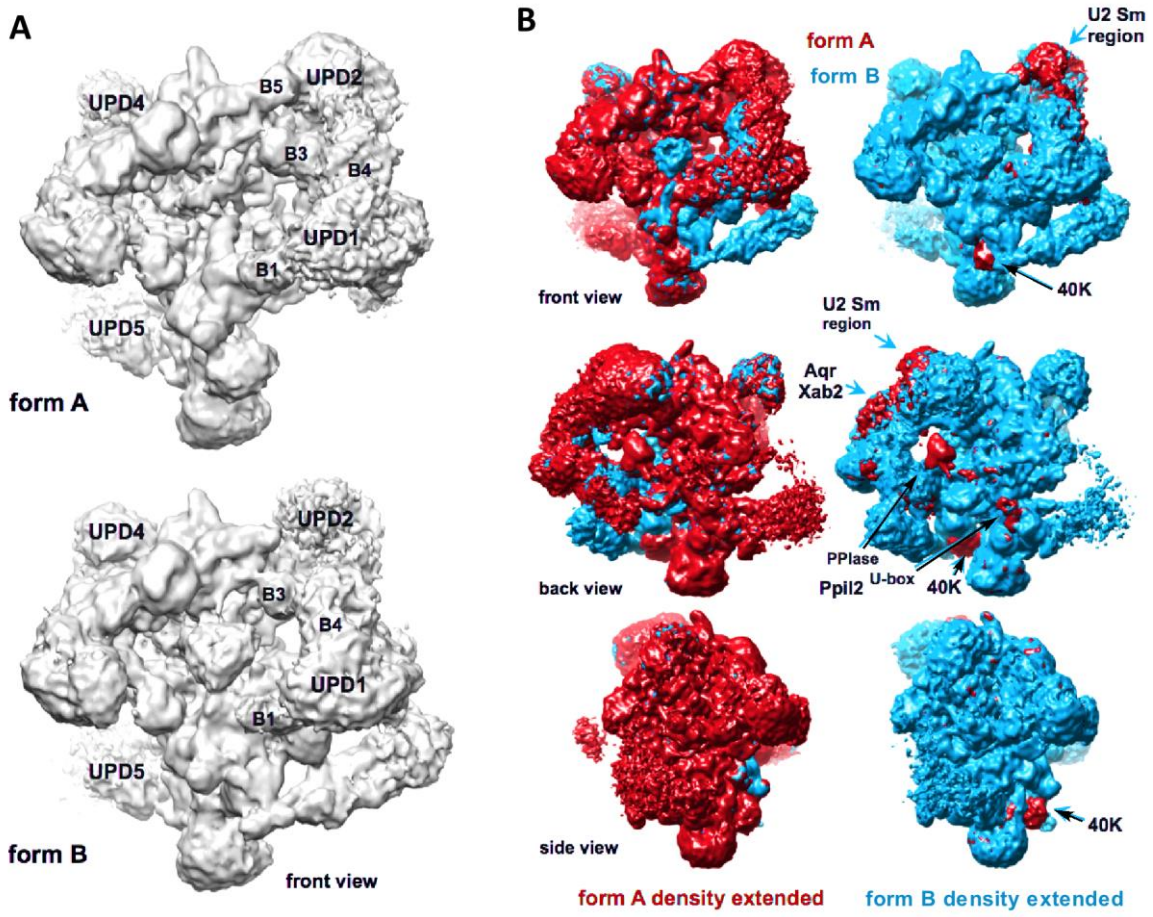


Figure 4.35: Two major conformationally different forms of the human B^{act} complex

(A) Front view of the two major forms of hB^{act}. **(B)** Comparison of the overall EM density of the two forms of B^{act}. The U2 Sm region (UPD2) is repositioned together with the Aqr-Xab2 protein complex (UPD1). Simultaneously, rearrangement of the U5 40K WD40 domain and Ppil2 domains occurs. Form B, which has an elongated density element at the lower right side of the central body, as well as an additional globular density in the middle, likely represents a later transition state of the spliceosome towards its activation.

In form B the Prp19 helical bundle is connected to the main body of the hB^{act} complex by three major bridges. One attachment point is the WD40 domain of the U5 40K protein, which is associated with the lower end of the helical bundle, most probably through the Spf27 helix. Interestingly, the position of the WD40 domain of U5 40K has rearranged between the two forms of the hB^{act} structure, such that in form B it has moved upwards, away from the U5 snRNA close to the foot of the main body (Figure 4.36). The second anchor point of the Prp19 helical bundle at the main body is a density bridge between a roughly spherical density element at the main body and the centre of the helical bundle (Figure 4.36). This spherical density element comprises the PPIase domain of Ppil1, the bridge that was also observed in the cryo-EM structure of the hC* complex (Bertram, Agafonov et al. 2017, Zhang, Yan et al. 2017).

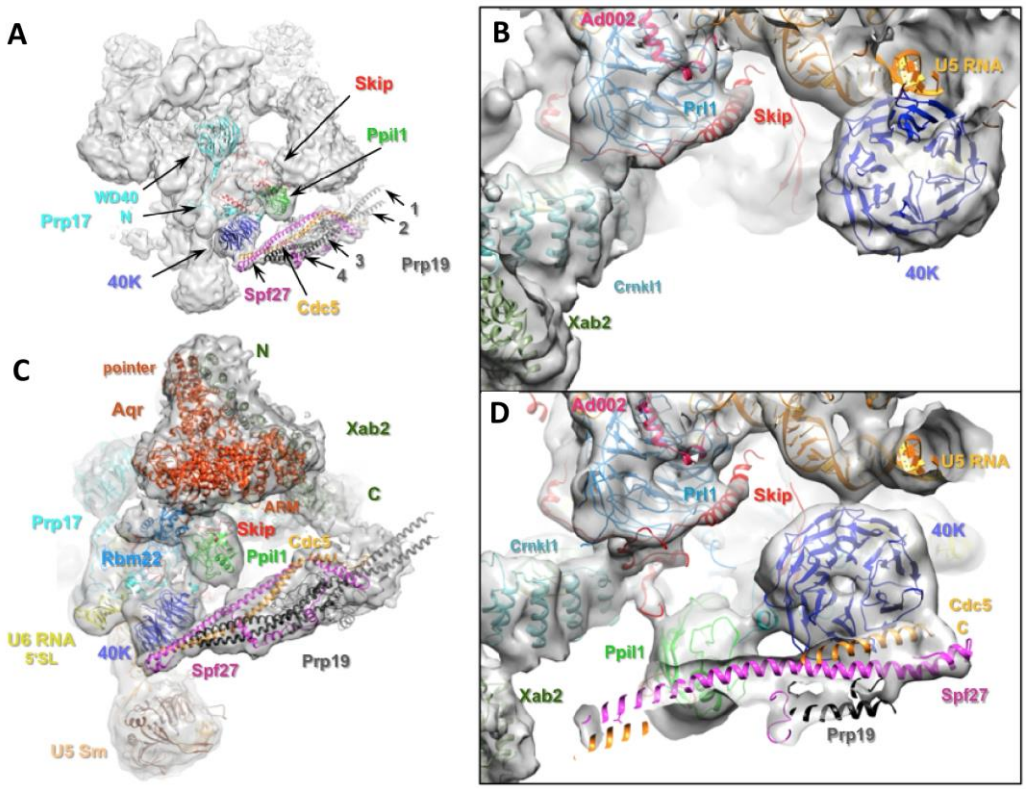


Figure 4.36: Overview of the organisation of additional density elements in form B of the human B^{act} complex.

(A) Front view of the form B unmasked EM density map. Arrows indicate the location of protein domains that are either not visible in the form A (Prp17, Prp19, Spf27, Cdc5, Ppil1, Skip) or reorganised (U5 40K). **(B, D)** A close-up of the fit of the U5 40K WD40 domain in forms A **(B)** and B **(D)** of hB^{act}. **(C)** Overview of the structural architecture and interactions of the Prp19 helical bundle with the central body and Aqr-Xab2 protein complex.

Strikingly, density for the PPIase domain of Ppil1 is not present in form A of the hB^{act} complex (Figure 4.35). The absence of density for Ppil1 in form A raises the question of whether the protein is physically absent from the B^{act} particles representing form A, or whether it is already associated with the B^{act} particles but simply not stably bound and, therefore, not visible in the cryo-EM reconstruction. For the following reasons it is likely that Ppil1 is already bound to the form A B^{act} particles but is still in a flexible state. Ppil1 has previously been shown to bind stably to a region

of the Skip protein encompassing aa 59-79 (Wang, Zhang et al. 2010). Thus it is highly likely that in nuclear extracts Skip and Ppil1 form a complex, which is integrated into spliceosome as a heterodimer. Evidence for this idea was indeed provided recently by the protein composition of a purified human B complex that, in addition to the typical B complex proteins, also contained Skip and Ppil1 (Bertram, Agafonov et al. 2017, in press). A close inspection of the form A B^{act} structure revealed that the region of the Skip protein, which is presumably attached to Ppil1, is indeed also not visible, indicating that this part of Skip together with Ppil1 is flexible in form A. In contrast, in form B, not only is the density for Ppil1 well defined, but the density of the part of Skip that is attached to Ppil1 is also visible, at least to some extent (Figure 4.36).

A third bridge between the helical bundle and the main body of B^{act} is located at the top of the helical bundle (Figure 4.36) and probably comprises part of Cdc5. This bridge connects Cdc5's C-terminal helical region with its N-terminal Myb domains, which are associated with the Prp8 RT domain in form B as in form A of the B^{act} structure.

The third significant difference between forms A and B is the absence of density for the PPIase domain of Ppil2 in form B (Figure 4.35). At the same time the N-terminal U box domains of Ppil2 are still well defined in form B and have not rearranged in comparison with form A. This indicates that Ppil2 is still present in the B^{act} particles representing form B of the B^{act} structure, but that Ppil2's PPIase domain has become destabilized and flexible. It is interesting to note that in form B the conformation of U2/U6 helix II has changed slightly, with the end of the helix turned slightly upwards; such a rearrangement is probably only possible as a result of destabilization of Ppil2's PPIase domain in form B. Moreover, bridge B3 is slightly rearranged in form B as compared with form A (Figure 4.37).

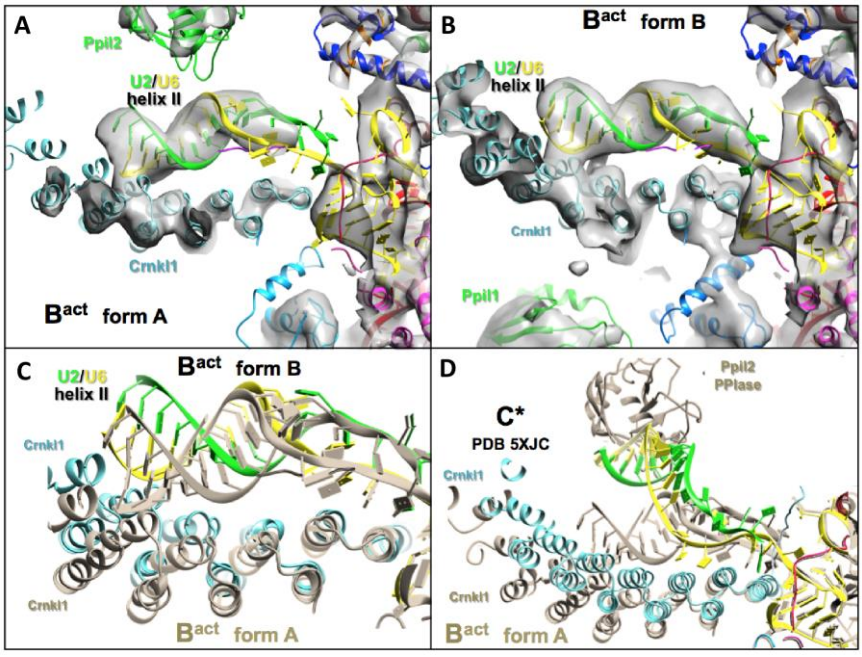


Figure 4.37: Comparison of the structural organization of U2/U6 helix II in the two major forms of B^{act} and the human C* complex

(A-B) Close-up views of the fit of U2/U6 helix II and its interaction partner Crnkl1 in forms A **(A)** and B **(B)** of the hB^{act} complex. **(C)** Ribbon representation of the superimposed structures of U2/U6 helix II in the two major forms of hB^{act}. Components of the form A are shown in light brown. **(D)** Comparison of the U2/U6 helix II structure between form A of B^{act} and the C* complex.

As in form A the Ppil2 PPIase domain is also connected to bridge B3, the rearrangement of bridge B3 may be caused by the destabilization of Ppil2's PPIase domain in form B. The Aqr/Xab2 protein complex has also slightly rearranged in form B, while all other bridges and the 3' domain of U2_appear to be organised in a very similar way in the two forms of the B^{act} structure. Finally, these two forms of the B^{act} structure differ significantly with respect to the visibility of the WD40 domain of hPrp17; in form A density for this domain is missing, while in form B it is well defined (Figure 4.38).

In the human C* complex, the C-terminal part of hPrp17 forms several short α-helical elements and meanders through the peripheral part of the main domain, connecting several proteins, including Skip and CypE (Bertram, Agafonov et al. 2017, Zhang, Yan et al. 2017). In form B of the hB^{act} structure, it is possible to trace the path of hPrp17 and to place several α-helices by rigid-body docking into corresponding densities of the hB^{act} structure.

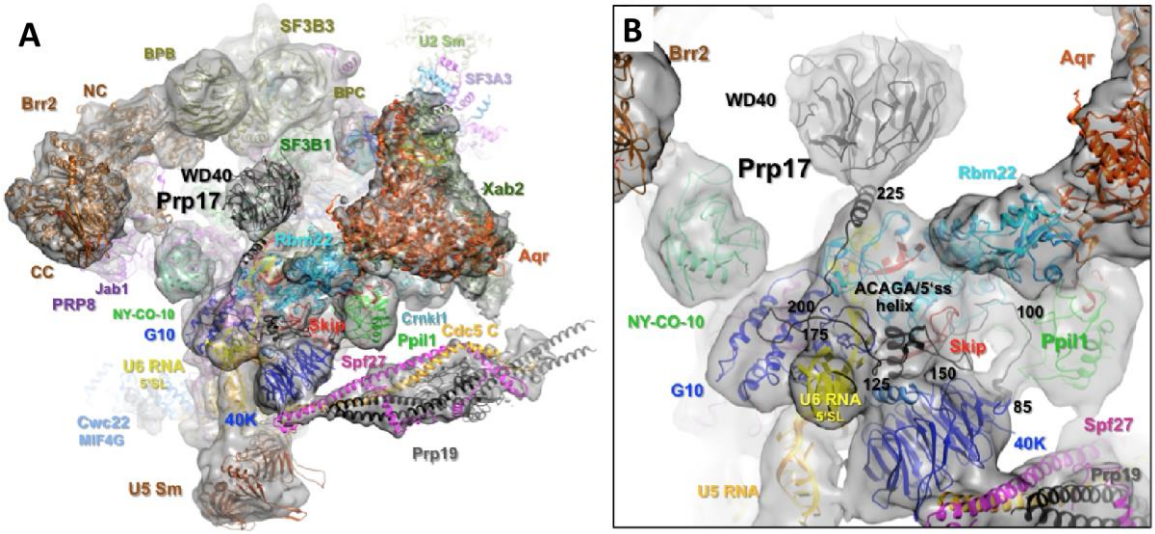


Figure 4.38: The location of the step II splicing factor Prp17 in form B of the human B^{act} complex

(A) Overview of the structure of the form B. **(B)** A close-up view of the location of Prp17 in the EM density map of form B. The C-terminal WD40 doman of Prp17 is placed in the globular density between the central and top domains. The N-terminal portion of Prp17 adopts an extended conformation.

In summary, the presence and absence of certain densities in the two forms of the cryo-EM B^{act} structure, is in all probability not due to the physical presence and absence of the respective proteins or protein domains. Instead, the two forms very probably represent two forms of the B^{act} complex, which differ in the conformational state of several proteins or domains thereof. Interestingly, the rearrangement of some of the protein domains appears to be coordinated. This is most likely to be the case for the movement of the U5 40K WD40 domain, and for the stable docking of Ppil1 and the Prp19 helical bundle, both to each other and to the central domain of the B^{act} structure.

5 Discussion

In this work the cryo-EM structure of the human activated spliceosome was determined for the first time. Following exhaustive 3D classifications two major populations of images were finally obtained, forms A and B, at resolutions of 5.3 Å and 8.1 Å, respectively. While the overall structures of the two major populations were largely very similar, several proteins or protein domains exhibited distinct conformational states in these two populations; this is probably due to the differential conformational flexibility of these proteins. In the following discussion I will first discuss the main features of the hB^{act} structure, which are shared by both populations. In the last part I will discuss the conformational dynamics of the human B^{act} complex.

5.1 The catalytic U2/U6 RNP core structure of the hB^{act} complex is evolutionarily highly conserved

For many years it was debated whether, during activation of the human spliceosome, initially a 4-way U2-U6 RNA junction forms at the catalytic centre, rather than a 3-way junction as observed in the yeast spliceosome (Madhani and Guthrie 1992, Sun and Manley 1995, Sashital, Cornilescu et al. 2004). The cryo-EM structure of the human B^{act} complex shows clearly that the catalytic U2/U6 RNA network forms a U2/U6 3-way junction and adopts a 3-dimensional structure that not only closely resembles the 3D structure of group II self-splicing introns, but also the 3D structure of the catalytic RNA network found in the *S. cerevisiae* B^{act} spliceosome. For example, the topology of the EM density map of the catalytic core RNA elements of the hB^{act} structure is entirely consistent with the formation of the catalytic triplex in the human activated spliceosome (Figure 4.15).

For catalysis of the first (and second) steps of splicing, two Mg^{2+} ions have to be positioned ca. 4 Å apart within the catalytic RNA network (Steitz and Steitz 1993), a process that is accomplished by their coordination through nucleotides of the U6 ISL bulged loop and the catalytic triad of U6 snRNA in U2/U6 helix 1b (Fica, Tuttle et al. 2013). This arrangement of the two catalytic metal ions was indeed observed in the 3D cryo-EM structure of the yeast C complex and the human and yeast C* complexes (Fica, Oubridge et al. 2017, Yan, Wan et al. 2017, Zhang, Yan et al. 2017). In the human B^{act} structure, however, density was only observed for the catalytic metal ion M2 at the catalytic centre, but not for M1 (Figure 4.15 C; Figure 4.16 A) – even though the nucleotides that would coordinate M1 appeared to be largely positioned in the correct conformation. Interestingly, the same situation was also observed in the cryo-EM structure of the yeast B^{act} complex, which likewise showed density for the catalytic Mg^{2+} M2, but not for M1 (Yan, Wan et al. 2016). As in both

B^{act} complexes an N-terminal region of the SF3A2/Prp11 protein is located close to the nucleotides that would coordinate M1, this protein may possibly hinder the correct positioning of M1. Consistent with this idea, SF3A2/Prp11 is rearranged during the catalytic activation of the yeast and human B^{act} spliceosomes (Bessonov, Anokhina et al. 2008, Warkocki, Odenwälder et al. 2009, Lardelli, Thompson et al. 2010).

The catalytic U2/U6 RNA network of the human B^{act} spliceosome is embedded in the same protein environment as it is in the yeast B^{act} spliceosome. The proteins that surround it include Prp8, Cdc5, Prl1, Crnkl1 and Skip. Indeed, the same conserved protein regions are in contact with equivalent RNA sites/regions of the U2/U6 RNA network in these two spliceosomes (Figure 5.1) (Rauhut, Fabrizio et al. 2016, Yan, Wan et al. 2016). The only difference between them concerns the N-terminal region of Ad002/Cwc15, which in the yeast B^{act} complex is in contact with the U6 ISL (Yan, Wan et al. 2016), but in the human B^{act} complex appears to be flexible. However, as it is in contact with U6 ISL in the human C* complex (Bertram, Agafonov et al. 2017), it seems very likely that the Ad002 N-terminal region will be stably positioned at the U6 ISL during catalytic activation of the B^{act} complex for step 1 catalysis. In summary, the structure of the catalytic U2/U6 RNP core is strikingly conserved between yeast and human. Moreover, the intricate protein-RNA network at the catalytic RNP core also clearly indicates that catalytic U2/U6 RNA network will only form in the presence of the spliceosomal proteins, which in turn explains why previous attempts to reconstitute an RNA-only catalytic pre-mRNA network largely failed (Valadkhan and Manley 2001, Valadkhan and Manley 2003).

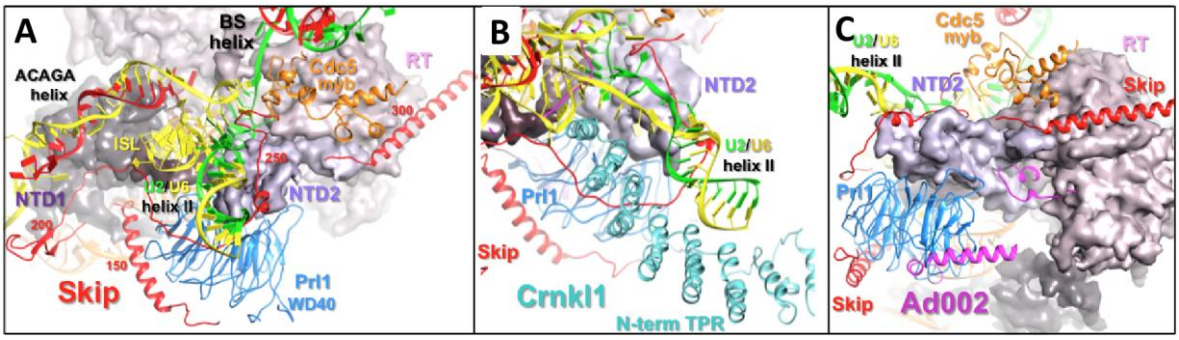


Figure 5.1: The U2/U6 catalytic RNA network closely interacts with proteins in the hB^{act} structure

The catalytic U2/U6 RNA network docked in the active site pocket of Prp8 (shown as a space filling model) interacts with Cdc5, Prl1, Skip **(A)** and Crnkl1 **(B)** proteins. **(C)** Ad002 does not interact with the U6 ISL

5.2 The first-step reactants are spatially separated in the human B^{act} complex and are occluded by proteins in an evolutionarily conserved manner

Consistent with earlier biochemical crosslinking studies (Sontheimer and Steitz 1993), the 3' end of the 5' exon and the ACA nts of the U6 ACAGA box are base-paired to U5 snRNA loop 1 and the 5' end of the intron, respectively (Figure 5.2). This induces a U-turn conformation at the 5'ss GU di-nucleotide and positions these two nucleotides close to the catalytic centre. The 5'ss, however, is occluded by interactions of the GU nucleotides with the N-terminal ZnF region of Rnf113A/hCwc24 (Figure 4.16 B and C). In the yeast B^{act} complex the 5' terminal GU nucleotides are in contact with the side chains of protein Cwc24 and the N-terminus of protein Prp11 (Yan, Wan et al. 2016). As the N-terminus of SF3A2 (the human homologue of Prp11) is also visible, at least in part, close to the 5'ss in the human B^{act} structure, it is likely that it interacts with the 5'ss nucleotides in a similar manner as Prp11 in the yeast B^{act} structure (Yan, Wan et al. 2016).

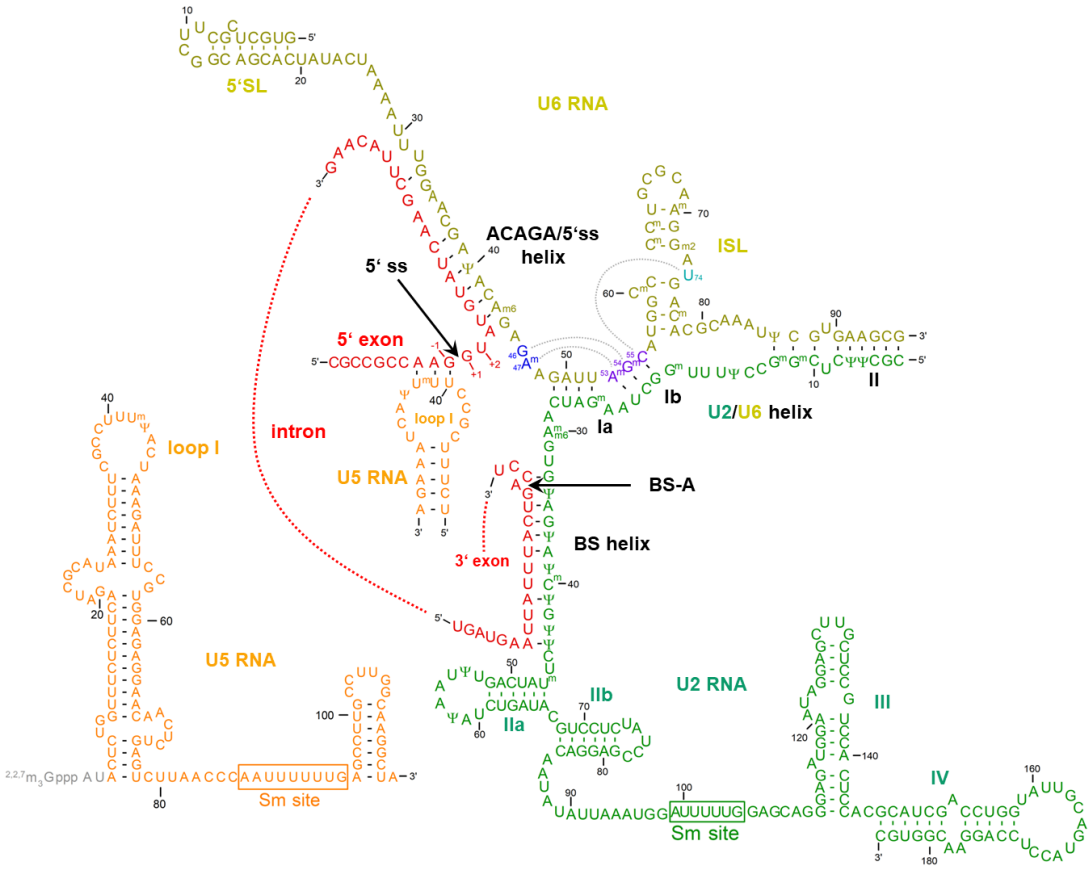


Figure 5.2: RNA secondary structure in the B^{act} spliceosome

Schematic representation of the RNA network. The complete secondary structure of the U5 snRNA is shown on the left. Tertiary interactions are indicated by grey, dotted lines.

While the conformation of the 5'ss and its base-pairing interactions, as well as interactions with proteins are highly conserved between yeast and human B^{act} complexes, the first ca. 10 nts of the intron downstream of the U6 ACAGA/5'ss helix are involved in the formation of an extended helix in the human B^{act}, but not in the yeast spliceosome. As such, an extended ACAGA/5'ss helix was not observed in the cryo-EM structures of other yeast assembly stages, but was present in both human B (Bertram et al. 2017, in the press) and C* complexes (Bertram, Agafonov et al. 2017), despite the fact that they were assembled on distinct pre-mRNAs. This suggests that an extended U6 ACAGA/5'ss helix may be specific for spliceosomes of higher eukaryotes.

The BS region of the pre-mRNA is organised in the human B^{act} structure in a manner very similar to its organisation in the yeast B^{act} complex. Thus, in both complexes the BS is base-paired to U2 snRNA in an extended BS/U2 RNA helix and the BS/U2 helix itself is clamped between the terminal HEAT repeats of the superhelical HEAT domain of SF3B1 (Figure 5.2; Figure 4.27), with the BS-A occluded in a protein pocket comprised of the C-terminal HEAT repeats and the PFH5 protein. Moreover, the distance between the BS-A and the catalytic centre is ca. 5 nm in both spliceosomes. Finally, the C-terminal ZnF domains of SF3A2/Prp11 make contact with the extended part of the U2/BS helix in a structurally homologous manner.

In summary, the human and yeast spliceosomes follow structurally the same strategy in generating a pre-catalytic activated spliceosome. In both spliceosomes (i) a catalytic U2/U6 RNP core is assembled whose 3D structure resembles closely the RNP core of a catalytically active spliceosome, except that the catalytic M1 metal is not yet stably coordinated at the catalytic centre, (ii) the 5'ss is already positioned close to the catalytic centre but is occluded by homologous proteins (Rnf113A/Cwc24 and SF3A2/Prp11) and (iii) the BS-A is bulged out from the U2/BS helix, but is occluded by U2 SF3B proteins and is spatially separated by ca. 5 nm from the catalytic centre. Together, these structural features contribute synergistically to a "safeguard" system, which prevents premature nucleophilic attack on the scissile bond of the 5'ss at this stage of the spliceosome's assembly.

5.3 Similarities and possible differences in the catalytic activation pathways between human and yeast spliceosomes

For the catalytic activation of the B^{act} complex, several major structural rearrangements have to occur in the spliceosome. First, the catalytic metal ion M1 has to be stably bound at the catalytic U2/U6 RNA centre at a distance of about 4 Å from metal ion M2. Second, Rnf113A/Cwc24 and SF3A2/Prp11 have to be rearranged to make the 5'ss available for nucleophilic attack by the 2'-hydroxyl group of the BS-A. Finally, for the latter attack to happen, SF3B1/Hsh155 has to be rearranged such as to liberate the U2/BS helix; at the same time SFA2/Prp11 has to be displaced

from the extended U2/BS helix to allow the latter to swing into the U2/U6 catalytic RNA centre for step I catalysis. In the yeast system, it is clear that the only driving force for these events is the ATP-dependent action of the RNA helicase Prp2 in co-operation with its co-factor Spp2 (Warkocki, Odenwälder et al. 2009). Also in human, hPrp2 together with its co-factor GPKOW is essential for the catalytic activation of the spliceosome (Silverman, Maeda et al. 2004). However, there is evidence that Prp2 action may not be sufficient. For example, it was recently shown that the ATPase activity of the RNA helicase Aqr, which is absent in yeast, is also required to activate catalytically the human spliceosome (De, Bessonov et al. 2015). Whether Prp2 and Aqr act co-ordinately or sequentially, is currently not clear. Moreover, at least two additional RNA helicases, DHX35 and Abstrakt, which are also not present in yeast, have been identified in human C complexes, and these could also play a part in the catalytic activation process.

With the caveat in mind that Aqr and, perhaps, additional RNA helicases contribute to the final catalytic activation of the human spliceosome, it is nevertheless reasonable to assume that the human Prp2 may trigger the same or similar structural rearrangements of the hB^{act} spliceosome as yeast Prp2. In yeast, Prp2 not only displaces Cwc24 and Cwc27 and a fraction of the RES proteins from the spliceosome, but also rearranges the U2 SF3B and SF3A proteins, such that their binding to the spliceosome becomes salt-sensitive (Warkocki, Odenwälder et al. 2009, Lardelli, Thompson et al. 2010). How can these dramatic events be triggered by a single RNA helicase? An unexpected scenario opened up when the cryo-EM structure of the yeast B^{act} complex revealed that Prp2 was bound to the convex side of the Hsh155 HEAT domain, close to the exit site of the 3'-end of the intron from the HEAT domain, but opposite the point where the BS/U2 helix is located within the HEAT domain (Rauhut, Fabrizio et al. 2016, Yan, Wan et al. 2016). As this architectural design precluded the possibility that Prp2 could access the BS/U2 helix directly, the more likely scenario is that Prp2 binds to the intron's 3'-end and induces, in an ATP-dependent manner, a structural change in the HEAT domain, such that the terminal HEAT repeats liberate the BS/U2 RNA helix. Moreover, as Cwc24, Prp11 and RES proteins – as well as other U2 SF3A and SF3B proteins, including Cus1 and Prp9 – contact the HEAT domain at distinct HEAT repeats, the Prp2-mediated conformational change of the HEAT domain may also destabilize the binding of the abovementioned proteins to Hsh155 and thus lead to their dissociation from the spliceosome.

Thus, in summary, the Hsh155 HEAT domain is a major scaffolding domain, which not only occludes the U2/BS helix, but is also indirectly connected to the 5'ss by an intricate protein-protein network, such that a Prp2-mediated structural change within the HEAT domain would liberate the first-step reactants in a highly coordinated manner and pave the way for the docking of the BS-A into the catalytic centre. For the following reasons it is reasonable to assume that a similar scenario as

depicted for the yeast spliceosome may also apply to the Prp2-mediated part of the catalytic activation pathway in the human spliceosome. First of all, all the protein-protein and protein-RNA interactions associated with the 5'ss and the U2/BS helix and their interconnectivity with the HEAT domain is highly conserved between the yeast and human B^{act} spliceosomes. Secondly, the human and yeast Prp2 RNA helicases are both bound to the homologous HEAT repeats 7 and 8 on the convex side of the HEAT domain of the SF3B1/Hsh155 protein in the human and yeast B^{act} complexes. In the yeast spliceosome the intron end leaves the HEAT domain and passes along the RbmX2 (hSnu17) RRM domain, and is finally bound by the helicase domain of Prp2 (Schneider, Agafonov et al. 2015, Rauhut, Fabrizio et al. 2016, Yan, Wan et al. 2016). The human B^{act} complex was assembled onto a pre-mRNA that contained only 10 nucleotides downstream of the BS – an RNA stretch just long enough to transverse the inner part of the HEAT domain close to the HEAT repeats 5 to 7. In the human B^{act} structure, RbmX2 is bound at a site equivalent to that in the yeast complex, and the distance and orientation of RbmX2 and the HEAT domain-bound hPrp2 are highly similar to those in yeast B^{act}. Therefore, we may reasonably infer that in human B^{act} a wild-type intron 3' end would be accommodated by RbmX2 and hPrp2 in the same way. Thus hPrp2 may also rearrange the SF3B and SF3A proteins, as well as Rnf113A and NY-CO-10 (hCwc27), as it does in the yeast spliceosome.

5.4 Rearrangement of the U2/U6 helix II and recruitment of hSyf2 as a prerequisite for catalytic activation of the human spliceosome?

The question remains why, in the human system, the RNA helicase Prp2 is not sufficient to activate the hB^{act} complex catalytically, instead requiring at least also the ATPase activity of Aqr. In this context it is important to note that the yeast and human B^{act} structures differ significantly in respect of the conformation of the U2/U6 helix II. Moreover, the human homologue of the yeast NTC protein Syf2 is not integrated into the human B^{act} complex; this is probably due to the peculiar conformation of U2/U6 helix II in the human B^{act} complex, which simply leaves no space to bind hSyf2 stably (Figure 4.21) (Rauhut, Fabrizio et al. 2016, Yan, Wan et al. 2016). The observations (i) that in the yeast B^{act} complex the U2/U6 helix II has adopted the canonical upright conformation and (ii) that Syf2 is present in the B^{act} spliceosome and binds close to the bottom of the U2/U6 helix together suggest that in the human spliceosome the U2/U6 helix II has to be rearranged in order to allow the integration of the human Syf2. Indeed, in the human C* complex U2/U6 helix II has adopted the conventional conformation and hSyf2 is bound to the bottom of the helix in a manner very similar to that in the yeast B^{act} complex (Bertram, Agafonov et al. 2017, Zhang, Yan et al. 2017). At present it is not known exactly when hSyf2 is recruited to the spliceosome. Therefore, it is possible that the rearrangement of the U2/U6 helix II and the incorporation of hSyf2 may occur before – or

concomitantly with – catalytic activation and may require the coordinated action of the Aqr and hPrp2 RNA helicases (and perhaps even an additional RNA helicase such as DHX35). Consistent with this scenario is the observation that in form B of the B^{act} complex the PPIase domain of Ppil2 is already dissociated from its binding site at the U2/U6 helix II. A rearranged U2/U6 helix II and a stably bound hSyf2 at the base of the helix could in turn also cause a rearrangement of other proteins of the central domain of the B^{act} complex.

In summary, the structure of the human B^{act} complex indicates that the catalytic activation pathway of the human spliceosome is more sophisticated and more highly regulated than in yeast. The absence of hSyf2 from the hB^{act} complex and the different conformations of the HAT proteins Xab2 (hSyf1) and Crnkl1 (hClf1) in the human and yeast B^{act} structures explain, at least in part, why in human cells these proteins are organised differently as compared with yeast. While in yeast they are all recruited to the spliceosome as part of the large NTC complex, in human Syf2 and Crnkl1 exist as single proteins, while Xab2 is part of the IBC. Therefore, in human spliceosomes these proteins can be sequentially recruited, allowing staggered assembly and thus the possibility of additional steps of regulation, as discussed for the integration of hSyf2.

5.5 Residues in human SF3b155's HEAT repeats whose mutation is linked to cancer are close to RES, Prp2 and the 3' end of the intron

Mutations in certain amino acids of the HEAT repeats of human SF3B1 are associated with various cancers (Bonnal, Vigevani et al. 2012, Darman, Seiler et al. 2015, Alsafadi, Houy et al. 2016). Most of those mutations that occur very frequently in patients (i.e. "hot spot" mutations) are located close to or within the intra-repeat loops of HEAT repeats H4–H7. Thus, cancer-related hot-spot mutations are located close to the likely binding region of the intron's 3' end and the binding sites for the RES core complex and Prp2 in the human B^{act} complex (Figure 5.3).

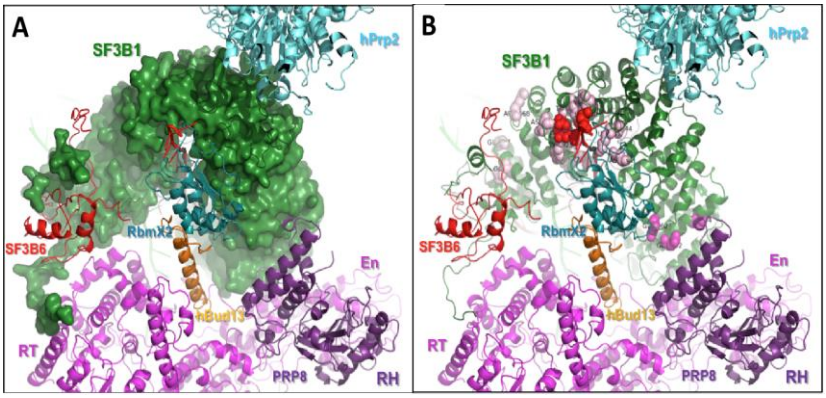


Figure 5.3: Cancer-related mutations in the human SF3b155

(A) hPrp2 interacts closely with the upper part of HEAT repeats H7–H9 of SF3B1. **(B)** Cancer-related mutations in the human SF3B1. The mutated amino acids are shown as pink space-filling spheres. The amino acids shown in red correspond to hot spot mutations that map close to the exit site of the BS-3'ss RNA from the HEAT domain and to the binding sites of RES and Prp2.

Mutations in SF3B1 HEAT repeats lead, by an unknown mechanism, to the selection of alternative branch sites and consequently to the use of cryptic pre-mRNA 3' splice sites and thus aberrantly spliced mRNAs (Alsafadi, Houy et al. 2016). Hot-spot mutations have been proposed to lead to a change in the curvature of the HEAT solenoid (Darman, Seiler et al. 2015). Although we do not know whether the spatial organization is different at earlier stages of spliceosome assembly (i.e. during A complex formation, when BS selection is thought to take place), based on the yeast B^{act} EM model, a structural change in the HEAT domain could indirectly destabilize the binding of the BS/U2 helix through the terminal HEAT repeats. Hot-spot mutations could potentially also alter binding of the 3' end of the intron to the HEAT domain, and thereby directly influence the choice of the BS. Alternatively, or in addition, these mutations could affect the interaction of RES proteins, and alter splicing efficiency – and thus the levels of certain mRNAs – in cancer cells.

5.6 The role of the U2 SF3B complex protein p14/SF3B6 remains enigmatic

While six of the seven subunits of the U2 SF3B protein complex are conserved between human and yeast, the RRM domain protein p14/SF3B6 is absent in yeast, while being conserved among higher eukaryotes (Will, Schneider et al. 2001). The SF3B6 protein has been crosslinked to the BS-A in early human spliceosomal complexes (Query, Moore et al. 1994, Will and Luhrmann 2001). Therefore, it was surprising that in the human B^{act} structure SF3B6 is located on the convex side of the SF3B1 HEAT domain, close to the N-terminal HEAT repeats, but separated from the BS-A by ca. 5 nm (Figure 4.29). Indeed, the BS-A is nested into a protein pocket composed of the C-terminal HEAT repeats and PHF5A/hRds3 in such a way that it is impossible for the BS-A to crosslink to SF3B6, even if it was flexible in solution. The U2/BS helix is occluded by the hHsh155 HEAT domain and Rds3, in a very similar manner, in the yeast B complex (Plaschka, Lin et al. 2017), and it is likely that it is also similarly organised already in the A complex. However, the crystal structure of the isolated human SF3B core protein complex (lacking SF3B6) revealed an open, more relaxed left-handed superhelical HEAT domain (Cretu, Schmitzová et al. 2016). It is likely that in the 17S U2 snRNP the HEAT domain may also adopt the open conformation, as the U2 snRNA will have to engage first with the BS of the pre-mRNA intron during the E complex state of the spliceosome, while the BS-A is still recognised by splicing factor SF1 (Wahl, Will et al. 2009). After formation of the U2/BS helix and concomitant displacement of SF1, probably by the action of UAP56 (Staley and Guthrie 1998), the SF3B1 HEAT domain will have to adopt the closed conformation so as to clamp the U2/BS helix tightly. This may possibly occur during the Prp5 RNA-helicase-mediated structural change that is required to form a stable A complex. It is possible that in higher eukaryotes, where the selection of the correct BS by the spliceosome is regulated, the SF3B6 subunit may play a part in this process and contact the BS-A during the transition from the E to the A complex. Concomitantly with formation of a stable

A complex, SF3B6 may rearrange and adopt its position on the convex side of the HEAT domain. Cryo-EM structures of the A complex should soon reveal whether this model is correct.

5.7 An intricate protein-protein interaction network comprising U2 SF3A proteins and peptidyl-prolyl isomerases

The 3' domain of the U2 snRNP, comprising the Sm core structure and the U2 A' and B'' proteins, is spatially separated from the core of the SF3B complex by an intricate network of protein-protein interactions in which the SF3A proteins play a major part (Figure 4.33). This is reminiscent of the organisation of the 17S U2 snRNP, where the SF3A protein complex also bridges the 3'-domain of U2 snRNP and the SF3B protein complex; the latter is bound to the 5'-terminal part of U2 snRNA (Kramer, Gruter et al. 1999). In addition, several SF3A and SF3B proteins are involved in forming bridges between the 3'-domain of the U2 snRNA, the core domain of the SF3B protein complex and the Aqr/Xab2 protein complex (Chapter 4.4.14).

The published cryo-EM structures of the yeast B^{act} complex are not well resolved in the peripheral domains containing the 3' domain of U2 and the Xab2/Crnkl1 protein basket. Therefore, it is difficult to compare these parts between the yeast and human B^{act} structures. In yeast, Xab2 appears to be primarily connected to the main body of B^{act} by its binding to the proteins Crnkl1 and hSyf2. In contrast, in the human B^{act} complex, a more sophisticated protein interaction network exists, which appears to stabilize the peripheral Aqr/Xab2 protein complex through several protein bridges to the main body of the B^{act} complex. These bridges contain, among other components, a class of proteins that is largely missing in the yeast spliceosome, i.e. PPIases. Altogether, the proteome of the human splicing machinery comprises at least eight PPIases, only one of which, NY-CO-10, has a counterpart in *S. cerevisiae* spliceosome (Cwc27) (Fabrizio, Dannenberg et al. 2009). The human PPIases are conserved in higher eukaryotes; nevertheless, their roles in pre-mRNA splicing have remained largely enigmatic. Four out of the eight PPIases (CypE, Ppil1, Ppil2 and NY-CO-10) are present in the human B^{act} complex. Strikingly, all of these are engaged in numerous protein-protein interactions. For example, CypE, which contains an RRM in addition to the PPIase domain, is part of the bridge B3, whereby its PPIase domain is bound to the N-terminal HATs of Xab2 and its RRM domain interacts with the RRM1 of SF3B4 and possibly also with the C-terminal domain of SF3A2. The PPIase domain of Ppil2 connects bridge B3 with the U2/U6 helix II, a case where a PPIase domain is in direct contact with RNA. The N-terminal U box domains of Ppil2 are spatially separated from its PPIase domain by about 10 nm and bind to hSnu114 and a nearby region of the hPrp8 RT domain.

Ppil1 anchors the Prp19 helical bundle to the main body of the B^{act} complex in form B of B^{act} and the PPIase domain of NY-CO-10 connects the NTD1 domain to the Prp8 RT domain in a manner similar to the way in which Cwc27 does this in the yeast B^{act} complex. Thus, one obvious function of the numerous PPIases in the human B^{act} complex appears to be of serving as bridges between various protein modules within the spliceosome. Whether the PPIase domains also fulfil enzymatic functions by isomerizing proline-peptide bonds is not clear at all. Inspection of the catalytic centres of the various PPIases suggests that at least Ppil1 and CypE may be active enzymes and could act as molecular switches; however, this remains to be investigated. It is also unclear whether the U boxes of Ppil2 play a part in ubiquitination reactions, in addition to their function as protein interaction modules.

5.8 Protein dynamics in the human B^{act} structure

An interesting structural feature of the second major form B of the B^{act} complex is that it differs from form A in respect of the presence and absence of several protein domains, mostly in the peripheral regions of the B^{act} structure. The structure of the central domain, including the catalytic U2/U6 RNP core, is largely similar in both forms. The most dramatic difference is the absence of density for the large elongated Prp19 helical bundle in form A and the presence of its well defined density in form B of the B^{act} structure (Figure 4.35). At the same time, density for Ppil1 appears in form B, forming a bridge between the centre of the helical bundle and the central body of B^{act}. Moreover, the U5 40K WD40 domain has moved away from its site close to the lower stem of U5 snRNA in form A and engages in interactions with the lower end of the helical bundle. It is highly likely that the rearrangements of U5 40K and Ppil1 are coordinated with the stable docking of the helical bundle on the lower side of the central domain of B^{act}. It is not clear whether the rearrangements of the PPIase domain of Ppil2 (which is absent in form B) and the WD40 domain of the hPrp17 (which is present in form B but absent in form A of B^{act}) are also somehow coordinated with each other and/or with the other rearrangements discussed above, or whether they occur sequentially.

In Chapter 4.5, I provided several pieces of evidence to indicate that the absence or presence of the above-mentioned protein domains in the two forms of the cryo-EM B^{act} structure is not due to the physical presence/absence of the respective proteins. Instead, in all likelihood, the differences indicate conformational flexibility of the protein domains. If we consider the conformational states in the cryo-EM structure of the human C* complex of those proteins which differ between forms A and B of hB^{act}, it is striking that there is a significant similarity in these conformations between form B of B^{act} and that of the hC* structure. This is best illustrated for the stable docking of the Prp19 helical bundle to the main body of hB^{act} and C* through interactions with U5 40K and Ppil1 proteins at almost equivalent positions (Figure 4.36). Therefore, it is reasonable to assume that form A is a

precursor to form B of the B^{act} structure and not just a dead-end complex. Thus, excitingly, it was possible to capture by cryo-EM analysis two conformational states of the human B^{act} complex that differ in their degree of conformational maturation towards the catalytic activation step. Consistent with the idea that form B is closer to the mature state of the B^{act} spliceosome than form A, is the finding that Ppil2 is dissociated from the spliceosome during the catalytic activation or step 1 formation, as it is absent in the purified C complex (Bessonov, Anokhina et al. 2008). Thus we may speculate that Ppil2 becomes destabilized in the spliceosome in at least two steps. First the PPIase domain in the last stage of B^{act} complex maturation (as observed in form B), followed by the dissociation of the interaction of the Ppil2 U boxes with the central B^{act} domain during catalytic activation (Figure 4.35).

To the best of our present-day knowledge, no ATP-requiring steps are involved in the transition of the B^{act} complex from form A to form B. This would mean that the observed conformational changes of the various protein domains are facilitated ("catalysed") simply by the thermal energy of the system. This, in turn, would suggest further that the dissociation of certain protein-protein interfaces would require the formation of new protein-protein interactions, which is exactly what one observes when comparing the structures of the two forms of the B^{act} spliceosome.
For all these reasons, it is reasonable to assume that by capturing the two states of the B^{act} structure by 3D classification, I have visualized conformational dynamics of some protein domains of the B^{act} complex. It is likely that these represent only a small window at the end of a long trajectory on which the B complex is transformed into a B^{act} complex. Following the dissociation of the U4/U6 duplex of the B complex by Brr2 RNA helicase, an exchange of more than 40 proteins occurs to form the mature B^{act} structure and it is likely that this exchange is also facilitated largely by thermal energy. By exhaustive 3D classification of intermediate assembly steps (between the B to B^{act} transition), it may thus be possible in the near future to obtain insight into the myriad of conformational changes that the numerous B^{act} proteins have to undergo to finally produce a mature B^{act} complex structure, as reconstructed in the work described here.

6 Bibliography

Achsel, T., H. Brahms, B. Kastner, A. Bachi, M. Wilm and R. Lührmann (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 J **18**(20): 5789-5802.

Aebi, M., H. Hornig, R. A. Padgett, J. Reiser and C. Weissmann (1986). "Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA." Cell **47**(4): 555-565.

Agafonov, D. E., J. Deckert, E. Wolf, P. Odenwälder, S. Bessonov, C. L. Will, H. Urlaub and R. Lührmann (2011). "Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method." Mol Cell Biol **31**(13): 2667-2682.

Agafonov, D. E., B. Kastner, O. Dybkov, R. V. Hofele, W. T. Liu, H. Urlaub, R. Luhrmann and H. Stark (2016). "Molecular architecture of the human U4/U6.U5 tri-snRNP." Science **351**(6280): 1416-1420.

Agafonov, D. E., M. van Santen, B. Kastner, P. Dube, C. L. Will, H. Urlaub and R. Lührmann (2016). "ATPγS stalls splicing after B complex formation but prior to spliceosome activation." RNA **22**(9): 1329-1337.

Alsafadi, S., A. Houy, A. Battistella, T. Popova, M. Wassef, E. Henry, F. Tirode, A. Constantinou, S. Piperno-Neumann, S. Roman-Roman, M. Dutertre and M. H. Stern (2016). "Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage." Nat Commun **7**: 10615.

Bach, M., G. Winkelmann and R. Lührmann (1989). "20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition." Proc Natl Acad Sci U S A **86**(16): 6038-6042.

Behrens, S. E. and R. Luhrmann (1991). "Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells." Genes Dev **5**(8): 1439-1452.

Behzadnia, N., M. M. Golas, K. Hartmuth, B. Sander, B. Kastner, J. Deckert, P. Dube, C. L. Will, H. Urlaub, H. Stark and R. Lührmann (2007). "Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes." EMBO J **26**(6): 1737-1748.

Berget, S. M., C. Moore and P. A. Sharp (1977). "Spliced segments at the 5' terminus of adenovirus 2 late mRNA." Proc Natl Acad Sci U S A **74**(8): 3171-3175.

Berglund, J. A., K. Chua, N. Abovich, R. Reed and M. Rosbash (1997). "The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC." Cell **89**(5): 781-787.

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

Bertram, K., D. E. Agafonov, W. T. Liu, O. Dybkov, C. L. Will, K. Hartmuth, H. Urlaub, B. Kastner, H. Stark and R. Luhrmann (2017). "Cryo-EM structure of a human spliceosome activated for step 2 of splicing." Nature **542**(7641): 318-323.

Bessonov, S., M. Anokhina, A. Krasauskas, M. M. Golas, B. Sander, C. L. Will, H. Urlaub, H. Stark and R. Luhrmann (2010). "Characterization of purified human B^{act} spliceosomal complexes reveals

compositional and morphological changes during spliceosome activation and first step catalysis." RNA **16**(12): 2384-2403.

Bessonov, S., M. Anokhina, C. L. Will, H. Urlaub and R. Luhrmann (2008). "Isolation of an active step I spliceosome and composition of its RNP core." Nature **452**(7189): 846-850.

Bessonov, S., M. Anokhina, C. L. Will, H. Urlaub and R. Lührmann (2008). "Isolation of an active step I spliceosome and composition of its RNP core." Nature **452**(7189): 846-850.

Birney, E., S. Kumar and A. R. Krainer (1993). "Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors." Nucleic Acids Res **21**(25): 5803-5816.

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

Black, D. L. and A. L. Pinto (1989). "U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6." Mol Cell Biol **9**(8): 3350-3359.

Boehringer, D., E. M. Makarov, B. Sander, O. V. Makarova, B. Kastner, R. Lührmann and H. Stark (2004). "Three-dimensional structure of a pre-catalytic human spliceosomal complex B." Nat Struct Mol Biol **11**(5): 463-468.

Boesler, C., N. Rigo, M. M. Anokhina, M. J. Tauchert, D. E. Agafonov, B. Kastner, H. Urlaub, R. Ficner, C. L. Will and R. Luhrmann (2016). "A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity." Nat Commun **7**: 11997.

Boesler, C., N. Rigo, M. M. Anokhina, M. J. Tauchert, D. E. Agafonov, B. Kastner, H. Urlaub, R. Ficner, C. L. Will and R. Lührmann (2016). "A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity." Nat Commun **7**: 11997.

Boivin, S., S. Kozak and R. Meijers (2013). "Optimization of protein purification and characterization using Thermofluor screens." Protein Expr Purif **91**(2): 192-206.

Bonnal, S., L. Vigevani and J. Valcárcel (2012). "The spliceosome as a target of novel antitumour drugs." Nat Rev Drug Discov **11**(11): 847-859.

Bringmann, P. and R. Luhrmann (1986). "Purification of the Individual Snrnps U1, U2, U5 and U4/U6 from Hela-Cells and Characterization of Their Protein Constituents." Embo Journal **5**(13): 3509-3516.

Brosi, R., H. P. Hauri and A. Kramer (1993). "Separation of splicing factor SF3 into two components and purification of SF3a activity." J Biol Chem **268**(23): 17640-17646.

Brow, D. A. and C. Guthrie (1988). "Spliceosomal RNA U6 is remarkably conserved from yeast to mammals." Nature **334**(6179): 213-218.

Buchwald, G., J. Ebert, C. Basquin, J. Sauliere, U. Jayachandran, F. Bono, H. Le Hir and E. Conti (2010). "Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex." Proc Natl Acad Sci U S A **107**(22): 10050-10055.

Burge, C. B., T. Tuschl and P. A. Sharp (1999). 20 Splicing of Precursors to mRNAs by the Spliceosomes.

Campbell, M. G., A. Cheng, A. F. Brilot, A. Moeller, D. Lyumkis, D. Veesler, J. Pan, S. C. Harrison, C. S. Potter, B. Carragher and N. Grigorieff (2012). "Movies of ice-embedded particles enhance resolution in electron cryo-microscopy." Structure **20**(11): 1823-1828.

Chanfreau, G., S. A. Elela, M. Ares, Jr. and C. Guthrie (1997). "Alternative 3'-end processing of U5 snRNA by RNase III." Genes Dev **11**(20): 2741-2751.

Chari, A., D. Haselbach, J. M. Kirves, J. Ohmer, E. Paknia, N. Fischer, O. Ganichkin, V. Möller, J. J. Frye, G. Petzold, M. Jarvis, M. Tietzel, C. Grimm, J. M. Peters, B. A. Schulman, K. Tittmann, J. Markl, U. Fischer and H. Stark (2015). "ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space." Nat Methods **12**(9): 859-865.

Chow, L. T., R. E. Gelinas, T. R. Broker and R. J. Roberts (1977). "An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA." Cell **12**(1): 1-8.

Colgan, D. F. and J. L. Manley (1997). "Mechanism and regulation of mRNA polyadenylation." Genes Dev **11**(21): 2755-2766.

Company, M., J. Arenas and J. Abelson (1991). "Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes." Nature **349**(6309): 487-493.

Coolidge, C. J., R. J. Seely and J. G. Patton (1997). "Functional analysis of the polypyrimidine tract in pre-mRNA splicing." Nucleic Acids Res **25**(4): 888-896.

Cordin, O., D. Hahn and J. D. Beggs (2012). "Structure, function and regulation of spliceosomal RNA helicases." Current Opinion in Cell Biology **24**(3): 431-438.

Cretu, C., J. Schmitzová, A. Ponce-Salvatierra, O. Dybkov, E. I. De Laurentiis, K. Sharma, C. L. Will, H. Urlaub, R. Lührmann and V. Pena (2016). "Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations." Mol Cell **64**(2): 307-319.

Dalbadie-McFarland, G. and J. Abelson (1990). "PRP5: a helicase-like protein required for mRNA splicing in yeast." Proc Natl Acad Sci U S A **87**(11): 4236-4240.

Darman, R. B., M. Seiler, A. A. Agrawal, K. H. Lim, S. Peng, D. Aird, S. L. Bailey, E. B. Bhavsar, B. Chan, S. Colla, L. Corson, J. Feala, P. Fekkes, K. Ichikawa, G. F. Keaney, L. Lee, P. Kumar, K. Kunii, C. MacKenzie, M. Matijevic, Y. Mizui, K. Myint, E. S. Park, X. Puyang, A. Selvaraj, M. P. Thomas, J. Tsai, J. Y. Wang, M. Warmuth, H. Yang, P. Zhu, G. Garcia-Manero, R. R. Furman, L. Yu, P. G. Smith and S. Buonamici (2015). "Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point." Cell Rep **13**(5): 1033-1045.

Davis, T. L., J. R. Walker, V. Campagna-Slater, P. J. Finerty, R. Paramanathan, G. Bernstein, F. MacKenzie, W. Tempel, H. Ouyang, W. H. Lee, E. Z. Eisenmesser and S. Dhe-Paganon (2010). "Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases." PLoS Biol **8**(7): e1000439.

De, I., S. Bessonov, R. Hofele, K. dos Santos, C. L. Will, H. Urlaub, R. Luhrmann and V. Pena (2015). "The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes." Nature Structural & Molecular Biology **22**(2): 138-144.

De, I., J. Schmitzova and V. Pena (2016). "The organization and contribution of helicases to RNA splicing." Wiley Interdisciplinary Reviews-RNA **7**(2): 259-274.

De, I., S. Sessonov, R. Hofele, K. dos Santos, C. L. Will, H. Urlaub, R. Luhrmann and V. Pena (2015). "The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes." Nature Structural & Molecular Biology **22**(2): 138-144.

Deckert, J., K. Hartmuth, D. Boehringer, N. Behzadnia, C. L. Will, B. Kastner, H. Stark, H. Urlaub and R. Luhrmann (2006). "Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions." Mol Cell Biol **26**(14): 5528-5543.

Deckert, J., K. Hartmuth, D. Boehringer, N. Behzadnia, C. L. Will, B. Kastner, H. Stark, H. Urlaub and R. Lührmann (2006). "Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions." Mol Cell Biol **26**(14): 5528-5543.

Dignam, J. D., R. M. Lebovitz and R. G. Roeder (1983). "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei." Nucleic Acids Res **11**(5): 1475-1489.

Dube, P., P. Tavares, R. Lurz and M. van Heel (1993). "The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry." EMBO J **12**(4): 1303-1309.

Dziembowski, A., A. P. Ventura, B. Rutz, F. Caspary, C. Faux, F. Halgand, O. Laprevote and B. Seraphin (2004). "Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing." EMBO J **23**(24): 4847-4856.

Edwalds-Gilbert, G., D. H. Kim, E. Silverman and R. J. Lin (2004). "Definition of a spliceosome interaction domain in yeast Prp2 ATPase." RNA **10**(2): 210-220.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Fabrizio, P., J. Dannenberg, P. Dube, B. Kastner, H. Stark, H. Urlaub and R. Luhrmann (2009). "The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome." Mol Cell **36**(4): 593-608.

Fabrizio, P., J. Dannenberg, P. Dube, B. Kastner, H. Stark, H. Urlaub and R. Lührmann (2009). "The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome." Mol Cell **36**(4): 593-608.

Fabrizio, P., B. Laggerbauer, J. Lauber, W. S. Lane and R. Lührmann (1997). "An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2." EMBO J **16**(13): 4092-4106.

Fica, S. M., M. A. Mefford, J. A. Piccirilli and J. P. Staley (2014). "Evidence for a group II intron-like catalytic triplex in the spliceosome." Nature Structural & Molecular Biology **21**(5): 464-471.

Fica, S. M., C. Oubridge, W. P. Galej, M. E. Wilkinson, X. C. Bai, A. J. Newman and K. Nagai (2017). "Structure of a spliceosome remodelled for exon ligation." Nature.

Fica, S. M., C. Oubridge, W. P. Galej, M. E. Wilkinson, X. C. Bai, A. J. Newman and K. Nagai (2017). "Structure of a spliceosome remodelled for exon ligation." Nature **542**(7641): 377-380.

Fica, S. M., N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley and J. A. Piccirilli (2013). "RNA catalyses nuclear pre-mRNA splicing." Nature **503**(7475): 229-234.

Fica, S. M., N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley and J. A. Piccirilli (2013). "RNA catalyses nuclear pre-mRNA splicing." Nature **503**(7475): 229-+.

Fischer, U., V. Sumpter, M. Sekine, T. Satoh and R. Luhrmann (1993). "Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap." EMBO J **12**(2): 573-583.

Fleckner, J., M. Zhang, J. Valcárcel and M. R. Green (1997). "U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction." Genes Dev **11**(14): 1864-1872.

Fourmann, J. B., J. Schmitzova, H. Christian, H. Urlaub, R. Ficner, K. L. Boon, P. Fabrizio and R. Luhrmann (2013). "Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system." Genes Dev **27**(4): 413-428.

Galej, W. P., C. Oubridge, A. J. Newman and K. Nagai (2013). "Crystal structure of Prp8 reveals active site cavity of the spliceosome." Nature **493**(7434): 638-643.

Galej, W. P., M. E. Wilkinson, S. M. Fica, C. Oubridge, A. J. Newman and K. Nagai (2016). "Cryo-EM structure of the spliceosome immediately after branching." Nature **537**(7619): 197-201.

Galej, W. P., M. E. Wilkinson, S. M. Fica, C. Oubridge, A. J. Newman and K. Nagai (2016). "Cryo-EM structure of the spliceosome immediately after branching." Nature **537**(7619): 197-+.

Gilbert, W. (1978). "Why genes in pieces?" Nature **271**(5645): 501.

Golas, M. M., B. Sander, C. L. Will, R. Lührmann and H. Stark (2003). "Molecular architecture of the multiprotein splicing factor SF3b." Science **300**(5621): 980-984.

Golas, M. M., B. Sander, C. L. Will, R. Lührmann and H. Stark (2005). "Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy." Mol Cell **17**(6): 869-883.

Gottschalk, A., C. Bartels, G. Neubauer, R. Luhrmann and P. Fabrizio (2001). "A novel yeast U2 snRNP protein, Snu17p, is required for the first catalytic step of splicing and for progression of spliceosome assembly." Mol Cell Biol **21**(9): 3037-3046.

Gozani, O., R. Feld and R. Reed (1996). "Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A." Genes Dev **10**(2): 233-243.

Gozani, O., J. Potashkin and R. Reed (1998). "A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site." Mol Cell Biol **18**(8): 4752-4760.

Grainger, R. J. and J. D. Beggs (2005). "Prp8 protein: at the heart of the spliceosome." RNA **11**(5): 533-557.

Graveley, B. R. (2000). "Sorting out the complexity of SR protein functions." RNA **6**(9): 1197-1211.

Graveley, B. R. and T. Maniatis (1998). "Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing." Mol Cell **1**(5): 765-771.

Guthrie, C. and B. Patterson (1988). "Spliceosomal snRNAs." Annu Rev Genet **22**: 387-419.

Hang, J., R. Wan, C. Yan and Y. Shi (2015). "Structural basis of pre-mRNA splicing." Science.

Hang, J., R. X. Wan, C. Y. Yan and Y. G. Shi (2015). "Structural basis of pre-mRNA splicing." Science **349**(6253): 1191-1198.

Heinrichs, V., M. Bach, G. Winkelmann and R. Luhrmann (1990). "U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site." Science **247**(4938): 69-72.

Hinterberger, M., I. Pettersson and J. A. Steitz (1983). "Isolation of Small Nuclear Ribonucleoproteins Containing U1, U2, U4, U5, and U6 Rnas." Journal of Biological Chemistry **258**(4): 2604-2613.

Jurica, M. S., D. Sousa, M. J. Moore and N. Grigorieff (2004). "Three-dimensional structure of C complex spliceosomes by electron microscopy." Nat Struct Mol Biol **11**(3): 265-269.

Kastner, B. (1998). Purification and Electron Microscopy of Spliceosomal snRNPs. J. Schenkel (ed.), RNP Particles, Splicing and

Autoimmune Diseases., Springer, Berlin, Heidelberg**:** 95 - 140

Kastner, B., N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F. Herzog, J. M. Peters, D. Poerschke, R. Lührmann and H. Stark (2008). "GraFix: sample preparation for single-particle electron cryomicroscopy." Nat Methods **5**(1): 53-55.

Kent, O. A. and A. M. MacMillan (2002). "Early organization of pre-mRNA during spliceosome assembly." Nat Struct Biol **9**(8): 576-581.

Kiss, T. (2001). "Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs." EMBO J **20**(14): 3617-3622.

Kiss, T. (2004). "Biogenesis of small nuclear RNPs." Journal of Cell Science **117**(25): 5949-+.

Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco and J. L. Manley (1994). "Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors." Nature **368**(6467): 119-124.

Kondo, Y., C. Oubridge, A. M. van Roon and K. Nagai (2015). "Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition." Elife **4**.

Kramer, A., P. Gruter, K. Groning and B. Kastner (1999). "Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP." J Cell Biol **145**(7): 1355-1368.

Kretzner, L., A. Krol and M. Rosbash (1990). "Saccharomyces cerevisiae U1 small nuclear RNA secondary structure contains both universal and yeast-specific domains." Proc Natl Acad Sci U S A **87**(2): 851-855.

Krämer, A. and U. Utans (1991). "Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs." EMBO J **10**(6): 1503-1509.

Kucukelbir, A., F. J. Sigworth and H. D. Tagare (2014). "Quantifying the local resolution of cryo-EM density maps." Nat Methods **11**(1): 63-65.

Kühlbrandt, W. (2014). "Biochemistry. The resolution revolution." Science **343**(6178): 1443-1444.

Laggerbauer, B., T. Achsel and R. Lührmann (1998). "The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro." Proc Natl Acad Sci U S A **95**(8): 4188-4192.

Laggerbauer, B., S. Liu, E. Makarov, H. P. Vornlocher, O. Makarova, D. Ingelfinger, T. Achsel and R. Luhrmann (2005). "The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation." RNA **11**(5): 598-608.

Lamond, A. I., M. M. Konarska and P. A. Sharp (1987). "A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation." Genes Dev **1**(6): 532-543.

Lardelli, R. M., J. X. Thompson, J. R. Yates and S. W. Stevens (2010). "Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing." RNA **16**(3): 516-528.

Lauber, J., G. Plessel, S. Prehn, C. L. Will, P. Fabrizio, K. Gröning, W. S. Lane and R. Lührmann (1997). "The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p." RNA **3**(8): 926-941.

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

Leitner, A., T. Walzthoeni and R. Aebersold (2014). "Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline." Nat Protoc **9**(1): 120-137.

Li, J., A. K. Leung, Y. Kondo, C. Oubridge and K. Nagai (2016). "Re-refinement of the spliceosomal U4 snRNP core-domain structure." Acta Crystallogr D Struct Biol **72**(Pt 1): 131-146.

Long, J. C. and J. F. Caceres (2009). "The SR protein family of splicing factors: master regulators of gene expression." Biochem J **417**(1): 15-27.

Madhani, H. D. and C. Guthrie (1992). "A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome." Cell **71**(5): 803-817.

Maeder, C., A. K. Kutach and C. Guthrie (2009). "ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8." Nat Struct Mol Biol **16**(1): 42-48.

Makarova, O. V., E. M. Makarov, S. Liu, H. P. Vornlocher and R. Luhrmann (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 J **21**(5): 1148-1157.

Martin, A., S. Schneider and B. Schwer (2002). "Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome." J Biol Chem **277**(20): 17743-17750.

Matera, A. G., R. M. Terns and M. P. Terns (2007). "Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs." Nat Rev Mol Cell Biol **8**(3): 209-220.

Matera, A. G. and Z. Wang (2014). "A day in the life of the spliceosome." Nat Rev Mol Cell Biol **15**(2): 108-121.

McPheeters, D. S. and P. Muhlenkamp (2003). "Spatial organization of protein-RNA interactions in the branch site-3' splice site region during pre-mRNA splicing in yeast." Mol Cell Biol **23**(12): 4174-4186.

Montemayor, E. J., E. C. Curran, H. H. Liao, K. L. Andrews, C. N. Treba, S. E. Butcher and D. A. Brow (2014). "Core structure of the U6 small nuclear ribonucleoprotein at 1.7-angstrom resolution." Nature Structural & Molecular Biology **21**(6): 544-551.

Moore, M. J. and P. A. Sharp (1993). "Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing." Nature **365**(6444): 364-368.

Mozaffari-Jovin, S., T. Wandersleben, K. F. Santos, C. L. Will, R. Luhrmann and M. C. Wahl (2013). "Inhibition of RNA Helicase Brr2 by the C-Terminal Tail of the Spliceosomal Protein Prp8." Science **341**(6141): 80-84.

Mozaffari-Jovin, S., T. Wandersleben, K. F. Santos, C. L. Will, R. Luhrmann and M. C. Wahl (2014). "Novel regulatory principles of the spliceosomal Brr2 RNA helicase and links to retinal disease in humans." RNA Biology **11**(4): 298-312.

Möhlmann, S., R. Mathew, P. Neumann, A. Schmitt, R. Lührmann and R. Ficner (2014). "Structural and functional analysis of the human spliceosomal DEAD-box helicase Prp28." Acta Crystallogr D Biol Crystallogr **70**(Pt 6): 1622-1630.

Nguyen, T. H., W. P. Galej, X. C. Bai, C. Oubridge, A. J. Newman, S. H. Scheres and K. Nagai (2016). "Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 A resolution." Nature.

Nguyen, T. H., W. P. Galej, X. C. Bai, C. G. Savva, A. J. Newman, S. H. Scheres and K. Nagai (2015). "The architecture of the spliceosomal U4/U6.U5 tri-snRNP." Nature **523**(7558): 47-52.

Nguyen, T. H., J. Li, W. P. Galej, H. Oshikane, A. J. Newman and K. Nagai (2013). "Structural basis of Brr2-Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site." Structure **21**(6): 910-919.

Ohrt, T., M. Prior, J. Dannenberg, P. Odenwälder, O. Dybkov, N. Rasche, J. Schmitzová, I. Gregor, P. Fabrizio, J. Enderlein and R. Lührmann (2012). "Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS." RNA **18**(6): 1244-1256.

Parker, R., P. G. Siliciano and C. Guthrie (1987). "Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA." Cell **49**(2): 229-239.

Peebles, C. L., P. S. Perlman, K. L. Mecklenburg, M. L. Petrillo, J. H. Tabor, K. A. Jarrell and H. L. Cheng (1986). "A self-splicing RNA excises an intron lariat." Cell **44**(2): 213-223.

Pena, V., S. Liu, J. M. Bujnicki, R. Lührmann and M. C. Wahl (2007). "Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa." Mol Cell **25**(4): 615-624.

Pena, V., A. Rozov, P. Fabrizio, R. Lührmann and M. C. Wahl (2008). "Structure and function of an RNase H domain at the heart of the spliceosome." EMBO J **27**(21): 2929-2940.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.

Plaschka, C., P. C. Lin and K. Nagai (2017). "Structure of a pre-catalytic spliceosome." Nature **546**(7660): 617-621.

Pomeranz Krummel, D. A., C. Oubridge, A. K. Leung, J. Li and K. Nagai (2009). "Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution." Nature **458**(7237): 475-480.

Pyle, A. M. (2008). "Translocation and unwinding mechanisms of RNA and DNA helicases." Annu Rev Biophys **37**: 317-336.

Query, C. C., M. J. Moore and P. A. Sharp (1994). "Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model." Genes Dev **8**(5): 587-597.

Raghunathan, P. L. and C. Guthrie (1998). "RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2." Curr Biol **8**(15): 847-855.

Rauhut, R., P. Fabrizio, O. Dybkov, K. Hartmuth, V. Pena, A. Chari, V. Kumar, C. T. Lee, H. Urlaub, B. Kastner, H. Stark and R. Luhrmann (2016). "Molecular architecture of the Saccharomyces cerevisiae activated spliceosome." Science **353**(6306): 1399-1405.

Reed, R. (1989). "The organization of 3' splice-site sequences in mammalian introns." Genes Dev **3**(12B): 2113-2123.

Reed, R. (1990). "Protein composition of mammalian spliceosomes assembled in vitro." Proc Natl Acad Sci U S A **87**(20): 8031-8035.

Reyes, J. L., E. H. Gustafson, H. R. Luo, M. J. Moore and M. M. Konarska (1999). "The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site." RNA **5**(2): 167-179.

Reyes, J. L., P. Kois, B. B. Konforti and M. M. Konarska (1996). "The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome." RNA **2**(3): 213-225.

Ritchie, D. B., M. J. Schellenberg, E. M. Gesner, S. A. Raithatha, D. T. Stuart and A. M. Macmillan (2008). "Structural elucidation of a PRP8 core domain from the heart of the spliceosome." Nat Struct Mol Biol **15**(11): 1199-1205.

Ruskin, B., P. D. Zamore and M. R. Green (1988). "A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly." Cell **52**(2): 207-219.

Sakharkar, M. K., V. T. Chow and P. Kangueane (2004). "Distributions of exons and introns in the human genome." In Silico Biol **4**(4): 387-393.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. **2nd ed.**

Santos, K. F., S. M. Jovin, G. Weber, V. Pena, R. Lührmann and M. C. Wahl (2012). "Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome." Proc Natl Acad Sci U S A **109**(43): 17418-17423.

Sashital, D. G., G. Cornilescu, C. J. McManus, D. A. Brow and S. E. Butcher (2004). "U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction." Nat Struct Mol Biol **11**(12): 1237-1242.

Schaffert, N., M. Hossbach, R. Heintzmann, T. Achsel and R. Luhrmann (2004). "RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies." EMBO J **23**(15): 3000-3009.

Schellenberg, M. J., R. A. Edwards, D. B. Ritchie, O. A. Kent, M. M. Golas, H. Stark, R. Lührmann, J. N. Glover and A. M. MacMillan (2006). "Crystal structure of a core spliceosomal protein interface." Proc Natl Acad Sci U S A **103**(5): 1266-1271.

Scheres, S. H. (2012). "RELION: implementation of a Bayesian approach to cryo-EM structure determination." J Struct Biol **180**(3): 519-530.

Scheres, S. H., H. Gao, M. Valle, G. T. Herman, P. P. Eggermont, J. Frank and J. M. Carazo (2007). "Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization." Nat Methods **4**(1): 27-29.

Schneider, C., D. E. Agafonov, J. Schmitzová, K. Hartmuth, P. Fabrizio and R. Lührmann (2015). "Dynamic Contacts of U2, RES, Cwc25, Prp8 and Prp45 Proteins with the Pre-mRNA Branch-Site and 3' Splice Site during Catalytic Activation and Step 1 Catalysis in Yeast Spliceosomes." PLoS Genet **11**(9): e1005539.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Methods **9**(7): 671-675.

Schneider, M., H. H. Hsiao, C. L. Will, R. Giet, H. Urlaub and R. Lührmann (2010). "Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation." Nat Struct Mol Biol **17**(2): 216-221.

Shatkin, A. J. and J. L. Manley (2000). "The ends of the affair: capping and polyadenylation." Nat Struct Biol **7**(10): 838-842.

Shevchenko, A., M. Wilm, O. Vorm, O. N. Jensen, A. V. Podtelejnikov, G. Neubauer, P. Mortensen and M. Mann (1996). "A strategy for identifying gel-separated proteins in sequence databases by MS alone." Biochem Soc Trans **24**(3): 893-896.

Silverman, E. J., A. Maeda, J. Wei, P. Smith, J. D. Beggs and R. J. Lin (2004). "Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing." Mol Cell Biol **24**(23): 10101-10110.

Small, E. C., S. R. Leggett, A. A. Winans and J. P. Staley (2006). "The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase." Mol Cell **23**(3): 389-399.

Smith, C. W. and J. Valcarcel (2000). "Alternative pre-mRNA splicing: the logic of combinatorial control." Trends Biochem Sci **25**(8): 381-388.

Sontheimer, E. J. and J. A. Steitz (1993). "The U5 and U6 small nuclear RNAs as active site components of the spliceosome." Science **262**(5142): 1989-1996.

Spadaccini, R., U. Reidt, O. Dybkov, C. Will, R. Frank, G. Stier, L. Corsini, M. C. Wahl, R. Lührmann and M. Sattler (2006). "Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing." RNA **12**(3): 410-425.

Spingola, M., L. Grate, D. Haussler and M. Ares, Jr. (1999). "Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae." RNA **5**(2): 221-234.

Staley, J. P. and C. Guthrie (1998). "Mechanical devices of the spliceosome: motors, clocks, springs, and things." Cell **92**(3): 315-326.

Stark, H., P. Dube, R. Lührmann and B. Kastner (2001). "Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle." Nature **409**(6819): 539-542.

Stark, H. and R. Lührmann (2006). "Cryo-electron microscopy of spliceosomal components." Annu Rev Biophys Biomol Struct **35**: 435-457.

Steitz, T. A. and J. A. Steitz (1993). "A general two-metal-ion mechanism for catalytic RNA." Proc Natl Acad Sci U S A **90**(14): 6498-6502.

Stephens, R. M. and T. D. Schneider (1992). "Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites." J Mol Biol **228**(4): 1124-1136.

Strauss, E. J. and C. Guthrie (1994). "PRP28, a 'DEAD-box' protein, is required for the first step of mRNA splicing in vitro." Nucleic Acids Res **22**(15): 3187-3193.

Sun, J. S. and J. L. Manley (1995). "A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing." Genes Dev **9**(7): 843-854.

Tang, G., L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees and S. J. Ludtke (2007). "EMAN2: an extensible image processing suite for electron microscopy." J Struct Biol **157**(1): 38-46.

Tarn, W. Y. and J. A. Steitz (1996). "A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro." Cell **84**(5): 801-811.

Teigelkamp, S., A. J. Newman and J. D. Beggs (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 J **14**(11): 2602-2612.

Teigelkamp, S., E. Whittaker and J. D. Beggs (1995). "Interaction of the yeast splicing factor PRP8 with substrate RNA during both steps of splicing." Nucleic Acids Res **23**(3): 320-326.

Tonegawa, S., A. M. Maxam, R. Tizard, O. Bernard and W. Gilbert (1978). "Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain." Proc Natl Acad Sci U S A **75**(3): 1485-1489.

Umen, J. G. and C. Guthrie (1995). "Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing." RNA **1**(6): 584-597.

Valadkhan, S. and J. L. Manley (2001). "Splicing-related catalysis by protein-free snRNAs." Nature **413**(6857): 701-707.

Valadkhan, S. and J. L. Manley (2003). "Characterization of the catalytic activity of U2 and U6 snRNAs." RNA **9**(7): 892-904.

Valcárcel, J., R. K. Gaur, R. Singh and M. R. Green (1996). "Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]." Science **273**(5282): 1706-1709.

van Heel, M. and J. Frank (1981). "Use of multivariate statistics in analysing the images of biological macromolecules." Ultramicroscopy **6**(2): 187-194.

van Heel, M. and G. Harauz (1988). "Biological macromolecules explored by pattern recognition." Scanning Microsc Suppl **2**: 295-301.

Wahl, M. C., C. L. Will and R. Luhrmann (2009). "The spliceosome: design principles of a dynamic RNP machine." Cell **136**(4): 701-718.

Wan, R., C. Yan, R. Bai, G. Huang and Y. Shi (2016). "Structure of a yeast catalytic step I spliceosome at 3.4 A resolution." Science **353**(6302): 895-904.

Wan, R., C. Yan, R. Bai, L. Wang, M. Huang, C. C. Wong and Y. Shi (2016). "The 3.8 A structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis." Science **351**(6272): 466-475.

Wang, X., S. Zhang, J. Zhang, X. Huang, C. Xu, W. Wang, Z. Liu, J. Wu and Y. Shi (2010). "A large intrinsically disordered region in SKIP and its disorder-order transition induced by PPIL1 binding revealed by NMR." J Biol Chem **285**(7): 4951-4963.

Warkocki, Z., P. Odenwälder, J. Schmitzová, F. Platzmann, H. Stark, H. Urlaub, R. Ficner, P. Fabrizio and R. Lührmann (2009). "Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components." Nat Struct Mol Biol **16**(12): 1237-1243.

Will, C. L. and R. Luhrmann (2001). "Spliceosomal UsnRNP biogenesis, structure and function." Curr Opin Cell Biol **13**(3): 290-301.

Will, C. L. and R. Luhrmann (2006). "Spliceosome Structure and Function." RNA World, Third Edition **43**: 369-400.

Will, C. L. and R. Luhrmann (2011). "Spliceosome structure and function." Cold Spring Harb Perspect Biol **3**(7).

Will, C. L., C. Schneider, A. M. MacMillan, N. F. Katopodis, G. Neubauer, M. Wilm, R. Lührmann and C. C. Query (2001). "A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site." EMBO J **20**(16): 4536-4546.

Will, C. L., H. Urlaub, T. Achsel, M. Gentzel, M. Wilm and R. Lührmann (2002). "Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein." EMBO J **21**(18): 4978-4988.

Wu, J. and J. L. Manley (1989). "Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing." Genes Dev **3**(10): 1553-1561.

Wysoczanski, P., C. Schneider, S. Xiang, F. Munari, S. Trowitzsch, M. C. Wahl, R. Luhrmann, S. Becker and M. Zweckstetter (2014). "Cooperative structure of the heterotrimeric pre-mRNA retention and splicing complex." Nat Struct Mol Biol **21**(10): 911-918.

Yan, C., J. Hang, R. Wan, M. Huang, C. C. Wong and Y. Shi (2015). "Structure of a yeast spliceosome at 3.6-angstrom resolution." Science.

Yan, C., R. Wan, R. Bai, G. Huang and Y. Shi (2016). "Structure of a yeast activated spliceosome at 3.5 A resolution." Science **353**(6302): 904-911.

Yan, C., R. Wan, R. Bai, G. Huang and Y. Shi (2017). "Structure of a yeast step II catalytically activated spliceosome." Science **355**(6321): 149-155.

Yan, C. Y., J. Hang, R. X. Wan, M. Huang, C. C. L. Wong and Y. G. Shi (2015). "Structure of a yeast spliceosome at 3.6-angstrom resolution." Science **349**(6253): 1182-1191.

Yan, C. Y., R. X. Wan, R. Bai, G. X. Y. Huang and Y. G. Shi (2016). "Structure of a yeast activated spliceosome at 3.5 angstrom resolution." Science **353**(6302): 904-911.

Yang, B., Y. J. Wu, M. Zhu, S. B. Fan, J. Lin, K. Zhang, S. Li, H. Chi, Y. X. Li, H. F. Chen, S. K. Luo, Y. H. Ding, L. H. Wang, Z. Hao, L. Y. Xiu, S. Chen, K. Ye, S. M. He and M. Q. Dong (2012). "Identification of cross-linked peptides from complex samples." Nat Methods **9**(9): 904-906.

Yang, K., L. Zhang, T. Xu, A. Heroux and R. Zhao (2008). "Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins." Proc Natl Acad Sci U S A **105**(37): 13817-13822.

Zamore, P. D., J. G. Patton and M. R. Green (1992). "Cloning and domain structure of the mammalian splicing factor U2AF." Nature **355**(6361): 609-614.

Zhang, K. (2016). "Gctf: Real-time CTF determination and correction." J Struct Biol **193**(1): 1-12.

Zhang, M. (1998). "Statistical features of human exons and their flanking regions." Human Molecular Genetics **7**(5): 919-932.

Zhang, X., C. Yan, J. Hang, L. I. Finci, J. Lei and Y. Shi (2017). "An Atomic Structure of the Human Spliceosome." Cell **169**(5): 918-929.e914.

Zhuang, Y. and A. M. Weiner (1986). "A compensatory base change in U1 snRNA suppresses a 5' splice site mutation." Cell **46**(6): 827-835.

Zillmann, M., M. L. Zapp and S. M. Berget (1988). "Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles." Mol Cell Biol **8**(2): 814-821.

7 Appendix

7.1 Abbreviations

3'ss	3'splice site
5'ss	5'splice site
A	Adenosine
aa	Amino acid
Amp	Ampicillin
APS	Ammonium peroxodisulfate
ATP	Adenosine triphosphate
bp	Base pair
BS	Branch site
BSA	Bovine serum albumin
BS-A	Branch site adenosine
°C	Degree celsius
C	Cytosine
CTP	Cytosine triphosphate
Ci	Curie
cpm	Counts per minute
Da	Dalton
dd	Double distilled
DExD/H	Consensus sequence of helicases
DMSO	Dimethyl sulfoxide
DNA	Desoxyribonucleic acid
DTT	Dithiothreitol
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*
G	Guanosine
g	Gram / centrifugal force
GTP	Guanosine triphosphate
H. sapiens	*Homo sapiens*
h	Hour / human
HCl	Hydrochloric acid
HEPES	N-2-Hydroxyethylpiperazin-N-2-ethansulfonic acid

hn	Heterogenous nuclear
IBC	Intron binding complex
ILS	Internal-stem loop
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
k	Kilo
kDa	Kilodalton
l	Liter
LB	Luria Bertani
LC	Liquid chromatography
Lsm	Like-Sm
M	Molar
m	Milli
µ	Micro
min	Minutes
mM	Millimolar
mRNA	Messenger RNA
MS	Mass spectrometry
MW	Molecular weight
nm	nanometer
NTC	NineTeen complex
NTP	Nucleoside triphosphate
nts	Nucleotides
OD	Optical density
P	Phosphate
PAGE	Polyacrylamide gel-electrophoresis
PCI	Phenol-chloroform-isoamyl alcohol
pH	Preponderance of hydrogen ions
PMSF	Phenylmethylsulfonylfluoride
PPT	Polypyrimidine tract
Pre-mRNA	Precursor-mRNA
%	Percent
Py	Pyrimidine base
R	Purine base
RES	Retention and splicing
RNA	Ribonucleic acid
RNase	Ribonuclease
RNP	Ribonucleoprotein
rpm	Revolutions per minute
RRM	RNA recognition motif

RT	Room temperature
S	Svedberg unit
s	Second
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	Sodium dodecylsulfate
Sm	‘Smith’, patient in which Sm proteins were first discovered
snRNA	Small nuclear ribonucleic acid
snRNP	Small nuclear ribonucleoparticle
SR	Serine arginine rich
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	Pyrimidine base

7.2 Acknowledgements

I would like to express my deepest gratitude to Prof. Dr. Reinhard Lührmann for providing an opportunity to work on this project, for his advice, fruitful discussions, and constant support. Special thanks for the energy that he invested to motivate me and for his help to keep me focused.

I am grateful to Prof. Dr. Ralf Ficner and Prof. Dr. Claudia Höbartner for helpful comments on my work during the progress reports. Many thanks to Prof. Dr. Holger Stark, Prof. Dr. Henning Urlaub and Prof. Dr. Wolfgang Wintermeyer for being members of my examination committee.

My special thanks to Dr. Berthold Kastner for discussing my results and suggestions in preparing the figures. I appreciate the help of Dr. Reinhard Rauhut and Dr. Cindy L. Will with reviewing the thesis.

I also would like to thank those people whose contribution made this PhD thesis possible:

Dr. Vlad Pena for introducing me to the world of splicing and Dr. Sergey Bessonov for his supervision during my first months in the department. I am also thankful to Dr. Dmitry Agafonov for his supervision and memorable lessons. My work and I profited from discussions with them.

For my introduction to the world of Electron Microscopy, I would like to thank Prof. Dr. Holger Stark and Dr. David Haselbach. Many thanks go to Dr. Prakash Dube, Dr. Berthold Kastner and Dr. Norbert Rigo for hands-on help with the electron microscopy in the initial phase.

Thomas Conrad & Thomas Schulz for cultivation of HeLa cells and Hossein Kohansal for preparing HeLa nuclear extract.

Annika Kühn and Monika Raabe for mass spectrometry analyses.

Dr. Olex Dybkov for his invaluable suggestions and the excellent collaboration in performing crosslinking experiments and crosslink identification by mass spectrometry.

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

I thank Claudia Fahlbusch, Gabi Heyne, Marion Killian, Gertrud Nowak, Irene Öchsner, Ulrich Steuerwald, and Winfried Lendeckel for the excellent technical support. I appreciate Juliane Moses help in solving a multitude of everyday problems, as well as all members of the Department for their everyday help in all aspects of scientific life and providing a friendly atmosphere.

My very warm appreciation goes also to my friends: Almu, Cristoph, Csaba and Judit, Greg, Inessa

and Tales, Katya and Timur, Leyla, Masha K., Myro and Vinay.

Finally, I would never be able to complete my studies without the support and commitment of my dear family.

7.3 Affidavit

I declare that my Doctoral thesis entitled “Investigation of the 3D structure of the human activated spliceosome by cryo-electron microscopy” has been written independently and with no other sources and aids than quoted. This thesis (wholly or in part) has not been submitted elsewhere for any academic award or qualification.

Ilya Komarov Göttingen, July 2017

7.4 Curriculum Vitae

Personal data

Name	Ilya Komarov
Date of birth	August 2^{d} 1988
Place of birth	Volgograd Oblast, USSR

Education

09.1994 – 06.2005	General Education School N^{o} 9, Volgograd Oblast, Russia
09.2005 – 06.2011	Volgograd State Medical University, Volgograd, Russia
	Faculty of Medical Biochemistry
09.2009 – 05.2011	Graduate school in Institute of Protein Research of Russian Academy of Sciences, Laboratory of Viral RNA Biochemistry, Pushchino, Russia
	Diploma Title: Mechanism of RNA Recombination carried out by RNA-Dependent RNA Polymerase Qbeta Bacteriophage
09.2011 – 06.2012	Internship in Volgograd State Medical University, Volgograd, Russia
	Primary specialization in clinical laboratory diagnostics
02.2012 – 01.2013	Macromolecular Crystallography Group, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany
02.2013 – present	PhD project, Department of Cellular Biochemistry, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany

7.5 Supplementary materials

Supplementary Table S1: BS3-crosslinks between known structured regions of human B^{act}

Statistics (Spectral Counts and $Score_{max}$) of the CX-MS data for the proteins of the purified human B^{act}. Numbers in the Residue 1 and 2 columns indicate the position of the crosslinked lysine or N-terminal methionine residue. Samples were analysed on Orbitrap Fusion Lumos Tribrid (Lumos) and Thermo Scientific Q Exactive HF (QExHF) mass spectrometers. Only intermolecular (Inter) protein-protein crosslinks are shown. Proteins are named according to UniProt Entry names.

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
Inter	AQR	CRNL1	524	175	1	1	4,32	3,50	2	4,32
			604	175	1		5,93		1	5,93
				182		3		5,66	3	5,66
		CWC22	550	530	1		0,36		1	0,36
		ISY1	604	101		1		3,08	1	3,08
			762	179	2	4	14,97	18,53	6	18,53
		PPIL1	558	80	1		5,94		1	5,94
		PRP19	667	192	2		5,42		2	5,42
			759	192		1		0,95	1	0,95
				200		4		2,05	4	2,05
		RBM22	604	229	1	1	4,79	4,10	2	4,79
		SF3A2	662	42		2		2,80	2	2,80
		SYF1	950	2	3	3	11,71	11,71	6	11,71
			954	2	12	16	6,32	8,25	28	8,25
			1002	2	5	1	6,34	5,28	6	6,34
		ZN830	201	339	2		5,02		2	5,02
				343	5	1	9,34	6,06	6	9,34
Inter	BUD13	CRNL1	558	460	1	1	3,71	2,03	2	3,71
		DHX16	417	389	2	2	3,70	3,97	4	3,97
			418	733	2		5,60		2	5,60
		GPKOW	14	334		1		4,01	1	4,01
			104	334	1		3,16		1	3,16
			466	201	2	3	3,46	3,21	5	3,46
			474	299	2		3,11		2	3,11
		PHF5A	596	108		1		5,72	1	5,72
		PLRG1	558	41	1		2,88		1	2,88
		PRP8	560	1514		1		3,73	1	3,73
		RBMX2	483	119	11	9	16,25	18,46	20	18,46
			486	119		1		2,82	1	2,82
		SF3B1	499	816	3		6,12		3	6,12
		SF3B2	339	326		1		1,38	1	1,38
		SNW1	474	503	1	1	2,71	3,47	2	3,47
			604	441	2	2	5,54	5,35	4	5,54
			605	441	7	5	14,66	17,06	12	17,06
		U520	68	975	1		8,76		1	8,76
Inter	BUD31	CWC27	28	35	25	18	6,78	6,48	43	6,78
		PRP8	9	218		2		10,33	2	10,33
			125	36	1	2	7,01	6,39	3	7,01
		SPF27	42	177	1		0,43		1	0,43
Inter	CDC5L	CRNL1	60	229	1		0,47		1	0,47
			174	229		4		8,41	4	8,41
			200	229	7	12	8,91	12,49	19	12,49
			255	388	1	1	4,96	2,20	2	4,96
			466	568	3	5	6,38	5,72	8	6,38
				569		1		3,47	1	3,47
			522	602		1		8,57	1	8,57
		CTBL1	174	458		1		5,33	1	5,33
			219	31		1		7,53	1	7,53
			294	534		2		6,34	2	6,34
			685	56	2	1	6,87	5,91	3	6,87
		CWC27	165	340		1		3,61	1	3,61
		ISY1	187	55	1	2	6,75	8,61	3	8,61
		PLRG1	576	19	2	1	5,38	5,95	3	5,95
		PPIL2	70	226	2		3,17		2	3,17
			312	507	2	2	5,01	4,32	4	5,01
			466	507	7	8	11,57	8,58	15	11,57
				508	5	2	8,80	7,99	7	8,80

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
				509	3	3	8,66	6,16	6	8,66
		PRCC	487	139	1	2	4,89	4,02	3	4,89
			522	216		2		10,98	2	10,98
		PRP17	187	333	9	8	9,06	9,56	17	9,56
			200	333	13	13	13,19	8,36	26	13,19
				551		1		5,67	1	5,67
		PRP19	522	179	2	3	6,70	6,95	5	6,95
				192	1	3	3,60	5,50	4	5,50
				244	2	1	7,59	4,25	3	7,59
			570	122	26	21	11,62	10,10	47	11,62
			631	122		4		8,06	4	8,06
				244	2	1	8,69	5,97	3	8,69
			685	179	4	6	7,23	7,71	10	7,71
				266	2	2	3,30	6,55	4	6,55
			686	179		1		6,48	1	6,48
			771	244		2		6,66	2	6,66
			782	179	1	1	4,39	4,09	2	4,39
				192		1		2,39	1	2,39
				244	11	9	10,80	11,86	20	11,86
				261		4		9,75	4	9,75
				266	4	1	9,38	5,73	5	9,38
			795	244	1	4	5,11	6,77	5	6,77
		PRP8	7	892	29	19	10,16	8,77	48	10,16
			28	727	1		5,79		1	5,79
			291	533		1		4,41	1	4,41
			539	85	1	2	5,32	5,37	3	5,37
		RBMX2	187	243		1		2,12	1	2,12
		SF3B2	174	577		1		2,47	1	2,47
			187	543	1	2	4,15	7,04	3	7,04
				570		1		1,76	1	1,76
		SF3B6	106	29	1	2	3,04	5,82	3	5,82
		SMD1	291	86	2		8,38		2	8,38
		SMD2	264	98	1		1,02		1	1,02
			782	118		1		9,25	1	9,25
		SMD3	201	104		1		2,93	1	2,93
		SNR40	771	18	7	6	6,41	7,68	13	7,68
				275		2		6,92	2	6,92
			782	270	2	2	2,95	6,68	4	6,68
				275		8		10,60	8	10,60
			795	270	17	11	9,39	5,90	28	9,39
		SNW1	28	258	1		5,25		1	5,25
				266	3	9	8,05	9,99	12	9,99
			124	311	1		5,37		1	5,37
		SPF27	312	168	3		5,87		3	5,87
			685	97	11	6	5,63	4,70	17	5,63
			771	177	69	69	15,74	16,94	138	16,94
			782	191	12	10	10,53	10,17	22	10,53
			795	218	2	2	4,79	4,37	4	4,79
			799	218		2		3,49	2	3,49
		SYF1	290	654	7	9	5,96	11,85	16	11,85
			294	654	31	23	22,28	18,10	54	22,28
			312	677	1	1	8,19	11,57	2	11,57
			522	532	12	11	6,94	7,18	23	7,18
		U5S1	466	446	4	6	4,12	4,94	10	4,94
		ZN830	255	137	1		4,97		1	4,97
				149	1	2	3,25	3,74	3	3,74
Inter	CRNL1	AQR	175	524	1	1	4,32	3,50	2	4,32
				604	1		5,93		1	5,93
			182	604		3		5,66	3	5,66
		BUD13	460	558	1	1	3,71	2,03	2	3,71
		CDC5L	229	60	1		0,47		1	0,47
				174		4		8,41	4	8,41
				200	7	12	8,91	12,49	19	12,49
			388	255	1	1	4,96	2,20	2	4,96
			568	466	3	5	6,38	5,72	8	6,38
			569	466		1		3,47	1	3,47
			602	522		1		8,57	1	8,57
		CWC27	730	462	1		4,13		1	4,13

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
		ISY1	213	127	2	2	7,46	8,08	4	8,08
		KIN17	229	33		1		5,10	1	5,10
		PLRG1	397	173		2		7,99	2	7,99
				180	3	6	9,36	8,12	9	9,36
			602	62		2		4,25	2	4,25
				68		1		2,60	1	2,60
		PPIL2	445	454		3		7,57	3	7,57
				460	1	2	4,62	5,74	3	5,74
				462	10	11	7,95	11,37	21	11,37
			460	462		3		5,38	3	5,38
			475	462		2		3,97	2	3,97
			485	462	1	3	3,88	14,03	4	14,03
			531	507		2		4,22	2	4,22
			568	490	1		11,43		1	11,43
			569	490	7	10	11,20	8,61	17	11,20
			574	490		1		1,68	1	1,68
			602	490	20	13	17,22	14,25	33	17,22
		PRCC	229	434		1		4,15	1	4,15
				437		1		2,42	1	2,42
				442	1		8,40		1	8,40
				443	2	2	8,26	10,53	4	10,53
			357	193	1	1	7,80	6,74	2	7,80
			388	193		2		4,62	2	4,62
				207		9		12,28	9	12,28
				216	1		7,19		1	7,19
			397	193	8	10	9,89	7,82	18	9,89
			427	193	3	7	12,94	13,17	10	13,17
			724	137		4		4,98	4	4,98
		PRP17	182	541		1		4,90	1	4,90
			213	547		3		4,95	3	4,95
				551	4	7	13,42	8,92	11	13,42
		PRP19	832	179		1		3,59	1	3,59
				192		1		3,46	1	3,46
		RBM22	224	114		1		8,39	1	8,39
		RSMB	607	57		1		2,67	1	2,67
		SF3A2	175	118	1	1	4,00	1,27	2	4,00
			680	194	1		0,89		1	0,89
		SF3B2	175	570	2	2	3,94	3,93	4	3,94
			178	570		1		1,57	1	1,57
			182	570	5	4	8,03	11,74	9	11,74
		SNW1	229	217		2		3,02	2	3,02
			357	48		1		5,04	1	5,04
			397	108	5	5	9,26	8,29	10	9,26
				110	8	5	8,54	9,53	13	9,53
				115	3	3	10,11	5,18	6	10,11
		SYF1	427	482	15	9	9,90	6,95	24	9,90
			460	482		1		7,93	1	7,93
			461	482	4	9	5,37	7,04	13	7,04
			787	794	4	2	6,49	7,70	6	7,70
		ZN830	445	216	2	2	5,16	5,44	4	5,44
				228	5	8	9,15	7,06	13	9,15
				234	18	20	10,46	13,76	38	13,76
			460	234	11	16	12,03	10,60	27	12,03
			475	228	2	1	3,24	3,04	3	3,24
				234	8	6	11,26	11,79	14	11,79
			485	234	1		2,84		1	2,84
			771	83		2		6,42	2	6,42
Inter	CTBL1	CDC5L	31	219		1		7,53	1	7,53
			56	685	2	1	6,87	5,91	3	6,87
			458	174		1		5,33	1	5,33
			534	294		2		6,34	2	6,34
		CWC15	95	183	3	1	5,31	5,10	4	5,31
			102	183		4		5,70	4	5,70
			252	18		2		7,24	2	7,24
			347	91	1	2	4,93	4,75	3	4,93
			372	91		1		3,16	1	3,16
			527	28		1		3,28	1	3,28
			534	18		3		6,77	3	6,77

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		Score$_{max}$		Total spec. count	Best Score$_{max}$
					Lumos	QExHF	Lumos	QExHF		
		DHX16	16	69		3		5,39	3	5,39
			27	69		2		9,40	2	9,40
			31	69	8	8	7,46	6,27	16	7,46
			56	69	1		2,95		1	2,95
			57	69		1		6,35	1	6,35
			91	69	2		7,24		2	7,24
			102	69	4		6,46		4	6,46
			297	1028	1		4,56		1	4,56
		GPKOW	447	299		1		1,29	1	1,29
		KIN17	56	297		1		3,79	1	3,79
		PPIL2	56	82		1		4,82	1	4,82
		PRP19	534	192	2		5,45		2	5,45
		PRP8	91	2070	1		2,53		1	2,53
		RBM22	554	170		1		0,44	1	0,44
		RU2B	83	1		1		2,36	1	2,36
			84	1	1		3,22		1	3,22
		SF3B1	31	6	2	2	6,24	5,44	4	6,24
			56	6	2	2	10,60	11,18	4	11,18
			57	6		1		14,05	1	14,05
			83	6	16	21	19,29	17,06	37	19,29
			84	6	10	7	10,38	11,45	17	11,45
			91	6	9	6	9,13	6,39	15	9,13
				21	5	5	8,40	5,87	10	8,40
			95	6	10	8	8,97	6,16	18	8,97
				21	2		6,69		2	6,69
		ZN830	31	117		1		4,87	1	4,87
				149		1		4,97	1	4,97
			56	78	5	7	8,63	8,82	12	8,82
				104		7		18,19	7	18,19
Inter	CWC15	CTBL1	18	252		2		7,24	2	7,24
				534		3		6,77	3	6,77
			28	527		1		3,28	1	3,28
			91	347	1	2	4,93	4,75	3	4,93
				372		1		3,16	1	3,16
			183	95	3	1	5,31	5,10	4	5,31
				102		4		5,70	4	5,70
		ISY1	183	260	1		2,08		1	2,08
		PLRG1	18	363	3	4	12,71	9,86	7	12,71
		PPIL2	92	82	1		4,29		1	4,29
			183	72	3	3	12,12	9,24	6	12,12
			195	188		1		6,52	1	6,52
			199	72	1		11,72		1	11,72
				180		3		10,65	3	10,65
				188	1	2	3,60	4,65	3	4,65
			221	418		1		2,04	1	2,04
		PRP8	91	1144	5	8	7,47	11,71	13	11,71
			92	1144		1		4,10	1	4,10
			205	1158	1		4,23		1	4,23
			226	1132		3		7,91	3	7,91
		RBM22	18	76	11	8	19,88	18,51	19	19,88
			28	40		2		4,60	2	4,60
		SF3B2	91	556	2		7,23		2	7,23
		SNW1	18	153	21	24	11,11	11,74	45	11,74
			152	311		3		11,67	3	11,67
		ZN830	40	272	1		0,60		1	0,60
			147	26	3	3	5,54	5,95	6	5,95
			183	28	1		2,89		1	2,89
Inter	CWC22	AQR	530	550	1		0,36		1	0,36
		CWC27	235	366	13	1	10,34	5,05	14	10,34
			243	429	1		4,36		1	4,36
				455		2		5,77	2	5,77
			250	434		2		10,55	2	10,55
			330	432	1	1	3,92	5,30	2	5,30
				434	11	13	10,02	12,87	24	12,87
				455	1		5,12		1	5,12
			397	364		1		6,00	1	6,00
		DHX16	604	93		2		3,88	2	3,88
			648	93	16	19	13,10	13,52	35	13,52

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		Score$_{max}$		Total spec. count	Best Score$_{max}$
					Lumos	QExHF	Lumos	QExHF		
		IF4A3	116	374	1	3	6,09	10,71	4	10,71
			118	252		2		7,15	2	7,15
			119	252	2	1	9,22	7,95	3	9,22
			243	374		1		3,10	1	3,10
			250	60	10	4	9,65	7,62	14	9,65
				374	3	2	9,48	10,91	5	10,91
			330	60	2	3	11,12	7,87	5	11,12
				152		1		5,98	1	5,98
		PRP8	330	366	19	20	11,47	13,57	39	13,57
			584	1392		1		1,86	1	1,86
		RUXE	745	80		1		2,02	1	2,02
		SMD2	149	8	1		0,50		1	0,50
			250	50		1		0,20	1	0,20
		SNW1	648	376	1		9,05		1	9,05
		SPF27	598	155		2		3,80	2	3,80
		SYF1	330	654	1		5,83		1	5,83
		U520	584	1146		1		1,41	1	1,41
		U5S1	330	359	1		6,22		1	6,22
Inter	CWC27	BUD31	35	28	25	18	6,78	6,48	43	6,78
		CDC5L	340	165		1		3,61	1	3,61
		CRNL1	462	730	1		4,13		1	4,13
		CWC22	364	397		1		6,00	1	6,00
			366	235	13	1	10,34	5,05	14	10,34
			429	243	1		4,36		1	4,36
			432	330	1	1	3,92	5,30	2	5,30
			434	250		2		10,55	2	10,55
				330	11	13	10,02	12,87	24	12,87
			455	243		2		5,77	2	5,77
				330	1		5,12		1	5,12
		IF4A3	432	374	2	1	8,45	10,24	3	10,24
			434	60	3	8	9,18	8,02	11	9,18
				70		3		3,84	3	3,84
				374	13	13	14,06	10,87	26	14,06
			455	19		1		6,90	1	6,90
				70	2	9	7,59	8,08	11	8,08
				74	2	4	6,07	8,28	6	8,28
				242		3		6,75	3	6,75
		PPIE	372	83	1		4,24		1	4,24
		R113A	192	107	1		3,03		1	3,03
		RSMB	183	50	4	2	6,76	6,56	6	6,76
			184	50	17	16	11,34	8,20	33	11,34
			186	50	5	2	8,01	8,26	7	8,26
		SF3A1	192	115		2		4,67	2	4,67
		SF3A2	184	103	1	1	4,88	3,44	2	4,88
		SYF1	180	539	1	1	4,39	3,50	2	4,39
		U5S1	393	352	6	2	14,28	5,73	8	14,28
				359		1		5,13	1	5,13
Inter	DHX16	BUD13	389	417	2	2	3,70	3,97	4	3,97
			733	418	2		5,60		2	5,60
		CTBL1	69	16		3		5,39	3	5,39
				27		2		9,40	2	9,40
				31	8	8	7,46	6,27	16	7,46
				56	1		2,95		1	2,95
				57		1		6,35	1	6,35
				91	2		7,24		2	7,24
				102	4		6,46		4	6,46
			1028	297	1		4,56		1	4,56
		CWC22	93	604		2		3,88	2	3,88
				648	16	19	13,10	13,52	35	13,52
		GPKOW	860	201		3		7,39	3	7,39
			1019	50	1		2,70		1	2,70
				61	4	6	7,74	9,03	10	9,03
		PPIL2	73	90	16	9	10,56	9,37	25	10,56
				95	21	23	14,23	13,94	44	14,23
		PRP19	1032	192		1		0,77	1	0,77
		R113A	1027	21		2		4,28	2	4,28
		RBMX2	93	246		1		4,56	1	4,56
			860	72	1		4,73		1	4,73

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
		SF3B1	69	6	21	14	10,16	6,71	35	10,16
				21	5	2	8,24	6,29	7	8,24
				22	2	7	4,67	6,28	9	6,28
			73	22		1		8,12	1	8,12
			860	733	2	3	11,31	11,23	5	11,31
		SF3B3	1019	126	12	8	12,64	11,21	20	12,64
				296	3	1	5,58	9,21	4	9,21
			1027	126	34	48	11,86	11,97	82	11,97
		SMD2	228	98	1		1,71		1	1,71
		SNIP1	860	216	3	2	9,62	9,69	5	9,69
		SRRM2	69	130		3		5,97	3	5,97
				131		1		3,77	1	3,77
Inter	GPKOW	BUD13	201	466	2	3	3,46	3,21	5	3,46
			299	474	2		3,11		2	3,11
			334	14		1		4,01	1	4,01
				104	1		3,16		1	3,16
		CTBL1	299	447		1		1,29	1	1,29
		DHX16	50	1019	1		2,70		1	2,70
			61	1019	4	6	7,74	9,03	10	9,03
			201	860		3		7,39	3	7,39
		PRP8	192	2244		2		2,22	2	2,22
			334	1831		1		4,97	1	4,97
			349	2108		1		4,16	1	4,16
		SF3A1	349	210		1		2,09	1	2,09
		U520	50	14		1		3,05	1	3,05
				60	1	3	4,91	6,76	4	6,76
			61	695		2		6,13	2	6,13
			192	14		1		3,71	1	3,71
				46	3	1	6,91	4,54	4	6,91
				73	4	5	5,89	5,80	9	5,89
				83	3	6	4,73	6,77	9	6,77
			346	155	1	1	3,67	3,87	2	3,87
			372	1049		1		9,98	1	9,98
Inter	IF4A3	CWC22	60	250	10	4	9,65	7,62	14	9,65
				330	2	3	11,12	7,87	5	11,12
			152	330		1		5,98	1	5,98
			252	118		2		7,15	2	7,15
				119	2	1	9,22	7,95	3	9,22
			374	116	1	3	6,09	10,71	4	10,71
				243		1		3,10	1	3,10
				250	3	2	9,48	10,91	5	10,91
		CWC27	19	455		1		6,90	1	6,90
			60	434	3	8	9,18	8,02	11	9,18
			70	434		3		3,84	3	3,84
				455	2	9	7,59	8,08	11	8,08
			74	455	2	4	6,07	8,28	6	8,28
			242	455		3		6,75	3	6,75
			374	432	2	1	8,45	10,24	3	10,24
				434	13	13	14,06	10,87	26	14,06
		SMD3	15	99	2	1	5,81	3,26	3	5,81
		U5S1	152	244	4	7	11,23	12,38	11	12,38
				694	1	1	4,58	3,56	2	4,58
Inter	ISY1	AQR	101	604		1		3,08	1	3,08
			179	762	2	4	14,97	18,53	6	18,53
		CDC5L	55	187	1	2	6,75	8,61	3	8,61
		CRNL1	127	213	2	2	7,46	8,08	4	8,08
		CWC15	260	183	1		2,08		1	2,08
		PPIE	92	218		3		11,00	3	11,00
			101	218		1		9,28	1	9,28
		PRP19	7	192	1		3,85		1	3,85
				244		4		6,46	4	6,46
		RBM22	92	199	1		1,09		1	1,09
		RU2A	259	205		2		6,82	2	6,82
			260	192	1		5,06		1	5,06
				193	1		16,77		1	16,77
				205	20	16	8,98	8,82	36	8,98
				221	2		4,90		2	4,90
			266	56	10	10	17,92	11,80	20	17,92

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
				179	1		3,95		1	3,95
				191	12	6	10,33	8,40	18	10,33
				192		1		7,19	1	7,19
				193	46	48	16,71	22,31	94	22,31
				221	29	28	19,59	20,12	57	20,12
		SF3A1	7	2	1	1	6,62	5,50	2	6,62
			55	115	4	4	8,45	6,34	8	8,45
			84	2		1		4,80	1	4,80
				30	1		3,78		1	3,78
				37	1	1	3,99	4,08	2	4,08
				115	5	5	7,39	7,41	10	7,41
			101	37	1		4,45		1	4,45
			105	37	6	3	7,82	5,61	9	7,82
				115	2	2	5,94	4,29	4	5,94
			121	2		1		5,27	1	5,27
				115		1		7,04	1	7,04
				131	1		3,26		1	3,26
			127	30	2		6,59		2	6,59
				37	1	1	5,55	5,21	2	5,55
			260	30	1	1	2,57	1,77	2	2,57
		SF3A3	266	69	2	1	23,89	5,77	3	23,89
		SF3B2	41	556		1		1,94	1	1,94
		SMD2	259	98	2		3,99		2	3,99
			260	98	3	3	5,24	4,86	6	5,24
		SNW1	121	317		1		1,68	1	1,68
		SRRM2	259	169		1		1,87	1	1,87
		SYF1	7	794	1		4,32		1	4,32
			41	794		4		8,69	4	8,69
			44	794		2		3,17	2	3,17
			84	257		1		3,12	1	3,12
			138	257	11	8	11,39	12,61	19	12,61
			188	83	5	3	5,37	4,77	8	5,37
			190	2	5	3	5,76	6,21	8	6,21
			260	2	10	11	10,22	12,04	21	12,04
			266	2	2	1	4,43	5,57	3	5,57
		U520	260	2091		1		3,02	1	3,02
Inter	KIN17	CRNL1	33	229		1		5,10	1	5,10
		CTBL1	297	56		1		3,79	1	3,79
		PRP8	129	2		1		1,80	1	1,80
			171	609		1		0,53	1	0,53
		SF3A2	157	48		1		2,23	1	2,23
		SF3B3	231	26	11	5	12,74	9,88	16	12,74
				296		3		9,47	3	9,47
		SMD1	33	9	1		0,01		1	0,01
		U520	272	1169		1		3,73	1	3,73
			293	1146		2		3,19	2	3,19
		U5S1	10	341	1	3	3,79	3,53	4	3,79
Inter	PHF5A	BUD13	108	596		1		5,72	1	5,72
		PRP8	108	853	1	1	4,16	7,98	2	7,98
		SF3B3	13	137		1		4,82	1	4,82
			95	137		1		4,86	1	4,86
Inter	PLRG1	BUD13	41	558	1		2,88		1	2,88
		CDC5L	19	576	2	1	5,38	5,95	3	5,95
		CRNL1	62	602		2		4,25	2	4,25
			68	602		1		2,60	1	2,60
			173	397		2		7,99	2	7,99
			180	397	3	6	9,36	8,12	9	9,36
		CWC15	363	18	3	4	12,71	9,86	7	12,71
		PPIL1	66	80		2		6,55	2	6,55
			68	80	1	1	4,82	4,25	2	4,82
			135	80		3		6,41	3	6,41
		PPIL2	505	462	2	5	4,74	7,58	7	7,58
		PRCC	173	193		1		4,76	1	4,76
		PRP19	1	122	4	9	9,20	9,50	13	9,50
			180	32		2		4,66	2	4,66
		PRP8	271	1132		1		0,66	1	0,66
		RUXE	47	80		1		2,44	1	2,44
		SF3A2	320	118		1		2,42	1	2,42

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		Score$_{max}$		Total spec. count	Best Score$_{max}$
					Lumos	QExHF	Lumos	QExHF		
		SNW1	173	110		1		4,48	1	4,48
			180	110	3	5	8,32	9,25	8	9,25
			181	110	1		3,51		1	3,51
			320	158	18	15	14,27	12,23	33	14,27
			363	153		1		6,44	1	6,44
				158	4	4	8,96	4,65	8	8,96
				217		2		3,31	2	3,31
		SYF1	135	420	1	1	5,77	5,48	2	5,77
			237	2		1		1,77	1	1,77
		U520	487	453		1		2,88	1	2,88
Inter	PPIE	CWC27	83	372	1		4,24		1	4,24
		ISY1	218	92		3		11,00	3	11,00
				101		1		9,28	1	9,28
		PPIL1	185	158	2	1	3,97	2,97	3	3,97
		RU2A	114	193	1		4,11		1	4,11
			123	172	2	4	6,97	6,56	6	6,97
			134	172	4	2	11,41	5,01	6	11,41
				179	1		4,45		1	4,45
				193	1	1	4,64	3,79	2	4,64
		SF3A1	88	210	7	2	9,14	12,51	9	12,51
			104	131		3		4,62	3	4,62
				210	9	5	8,65	8,89	14	8,89
				223	1	2	3,84	7,60	3	7,60
			108	223	2	3	5,30	3,69	5	5,30
				230	2		7,92		2	7,92
				231	4	3	10,01	7,41	7	10,01
				251	1		5,99		1	5,99
			114	223	1	4	2,88	4,01	5	4,01
				230	1		6,53		1	6,53
				231		2		5,50	2	5,50
			123	230	4	1	10,80	4,93	5	10,80
				231	3	3	11,17	8,80	6	11,17
			134	231	1	4	11,97	6,50	5	11,97
			218	2	8	5	10,00	9,59	13	10,00
		SF3A2	284	42		1		3,99	1	3,99
		SF3A3	134	219	5	1	6,68	2,68	6	6,68
		SF3B2	88	275		2		6,70	2	6,70
			104	275	1		7,39		1	7,39
		SYF1	104	76	1	1	3,93	5,43	2	5,43
			108	36	3	4	5,59	7,83	7	7,83
				76	5	9	4,14	5,19	14	5,19
				212	1		3,66		1	3,66
			114	36	1	1	5,93	3,81	2	5,93
				212		1		4,71	1	4,71
			123	76	1	1	3,11	2,68	2	3,11
			134	76	3		6,00		3	6,00
			218	2	3	11	6,32	10,13	14	10,13
Inter	PPIL1	AQR	80	558	1		5,94		1	5,94
		PLRG1	80	66		2		6,55	2	6,55
				68	1	1	4,82	4,25	2	4,82
				135		3		6,41	3	6,41
		PPIE	158	185	2	1	3,97	2,97	3	3,97
		PRP17	37	63	1		6,01		1	6,01
		RBM22	80	286		3		6,71	3	6,71
		SF3A3	52	97		1		2,39	1	2,39
		SNR40	80	1	5	6	5,97	7,97	11	7,97
				8		1		6,01	1	6,01
				18	3	1	4,39	3,94	4	4,39
			161	18	1		3,56		1	3,56
		SNW1	52	193	1		2,66		1	2,66
			58	81	1		12,15		1	12,15
			80	193	3	11	6,74	7,82	14	7,82
			158	97	28	20	14,71	12,35	48	14,71
		SPF27	80	168	10	12	9,74	8,90	22	9,74
Inter	PPIL2	CDC5L	226	70	2		3,17		2	3,17
			507	312	2	2	5,01	4,32	4	5,01
				466	7	8	11,57	8,58	15	11,57
			508	466	5	2	8,80	7,99	7	8,80

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
			509	466	3	3	8,66	6,16	6	8,66
		CRNL1	454	445		3		7,57	3	7,57
			460	445	1	2	4,62	5,74	3	5,74
			462	445	10	11	7,95	11,37	21	11,37
				460		3		5,38	3	5,38
				475		2		3,97	2	3,97
				485	1	3	3,88	14,03	4	14,03
			490	568	1		11,43		1	11,43
				569	7	10	11,20	8,61	17	11,20
				574		1		1,68	1	1,68
				602	20	13	17,22	14,25	33	17,22
			507	531		2		4,22	2	4,22
		CTBL1	82	56		1		4,82	1	4,82
		CWC15	72	183	3	3	12,12	9,24	6	12,12
				199	1		11,72		1	11,72
			82	92	1		4,29		1	4,29
			180	199		3		10,65	3	10,65
			188	195		1		6,52	1	6,52
				199	1	2	3,60	4,65	3	4,65
			418	221		1		2,04	1	2,04
		DHX16	90	73	16	9	10,56	9,37	25	10,56
			95	73	21	23	14,23	13,94	44	14,23
		PLRG1	462	505	2	5	4,74	7,58	7	7,58
		PRP8	26	1132	23	42	10,77	12,63	65	12,63
			166	1144		1		4,98	1	4,98
			180	1144		1		4,70	1	4,70
			226	727	31	32	16,31	9,03	63	16,31
			231	727	5	4	5,91	4,06	9	5,91
			232	727	2		3,60		2	3,60
		SF3A2	392	101		1		4,79	1	4,79
			393	101		3		9,76	3	9,76
			418	91	10	20	11,08	9,21	30	11,08
			423	107	7	16	7,66	8,46	23	8,46
		SF3B1	82	6		1		3,27	1	3,27
			90	6	1		4,32		1	4,32
			198	496		1		2,70	1	2,70
		SMD2	230	98	1		3,66		1	3,66
		SNW1	277	240	7	3	8,35	11,94	10	11,94
		SYF1	418	50	1		3,79		1	3,79
		ZN830	277	262		1		7,82	1	7,82
			312	234	1		4,53		1	4,53
			313	182	4		7,49		4	7,49
				190	5	14	11,12	11,80	19	11,80
				216		2		7,50	2	7,50
				228	12	14	10,15	8,92	26	10,15
				234	29	33	13,86	14,99	62	14,99
			450	190		2		4,28	2	4,28
				228	8	15	7,97	7,69	23	7,97
				234	4	7	5,51	7,02	11	7,02
			454	228	2	4	4,88	9,32	6	9,32
				234	10	8	8,24	8,84	18	8,84
			460	234	4		7,98		4	7,98
			462	182	1		8,86		1	8,86
				190	2	1	5,23	3,06	3	5,23
				228	7	7	10,95	10,98	14	10,98
				234	13	9	12,81	12,31	22	12,81
Inter	PRCC	CDC5L	139	487	1	2	4,89	4,02	3	4,89
			216	522		2		10,98	2	10,98
		CRNL1	137	724		4		4,98	4	4,98
			193	357	1	1	7,80	6,74	2	7,80
				388		2		4,62	2	4,62
				397	8	10	9,89	7,82	18	9,89
				427	3	7	12,94	13,17	10	13,17
			207	388		9		12,28	9	12,28
			216	388	1		7,19		1	7,19
			434	229		1		4,15	1	4,15
			437	229		1		2,42	1	2,42
			442	229	1		8,40		1	8,40

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
			443	229	2	2	8,26	10,53	4	10,53
		PLRG1	193	173		1		4,76	1	4,76
		RBM22	428	78	2		8,00		2	8,00
				104		2		4,89	2	4,89
			437	149		1		5,77	1	5,77
			442	286		1		3,25	1	3,25
			479	149	1	1	5,49	4,18	2	5,49
		SF3A2	414	121		1		1,89	1	1,89
		SNW1	137	48	4		3,86		4	3,86
			139	48	2		3,32		2	3,32
			193	97		2		9,21	2	9,21
				110		2		7,98	2	7,98
			207	110		1		5,50	1	5,50
			414	193	3	10	9,29	8,51	13	9,29
		SYF1	193	423		2		7,77	2	7,77
				458	13	10	10,82	11,29	23	11,29
			207	420	3	5	7,47	7,31	8	7,47
				458	5	2	8,55	7,76	7	8,55
			216	420	2	2	9,41	7,17	4	9,41
				458	1	1	4,41	3,41	2	4,41
			223	420	1	3	5,40	8,12	4	8,12
Inter	PRP17	CDC5L	333	187	9	8	9,06	9,56	17	9,56
				200	13	13	13,19	8,36	26	13,19
			551	200		1		5,67	1	5,67
		CRNL1	541	182		1		4,90	1	4,90
			547	213		3		4,95	3	4,95
			551	213	4	7	13,42	8,92	11	13,42
		PPIL1	63	37	1		6,01		1	6,01
		PRP8	161	36		1		5,69	1	5,69
			399	50		1		2,62	1	2,62
			551	162	1		2,71		1	2,71
		SF3B2	440	570	4	4	9,59	5,60	8	9,59
Inter	PRP19	AQR	192	667	2		5,42		2	5,42
				759		1		0,95	1	0,95
			200	759		4		2,05	4	2,05
		CDC5L	122	570	26	21	11,62	10,10	47	11,62
				631		4		8,06	4	8,06
			179	522	2	3	6,70	6,95	5	6,95
				685	4	6	7,23	7,71	10	7,71
				686		1		6,48	1	6,48
				782	1	1	4,39	4,09	2	4,39
			192	522	1	3	3,60	5,50	4	5,50
				782		1		2,39	1	2,39
			244	522	2	1	7,59	4,25	3	7,59
				631	2	1	8,69	5,97	3	8,69
				771		2		6,66	2	6,66
				782	11	9	10,80	11,86	20	11,86
				795	1	4	5,11	6,77	5	6,77
			261	782		4		9,75	4	9,75
			266	685	2	2	3,30	6,55	4	6,55
				782	4	1	9,38	5,73	5	9,38
		CRNL1	179	832		1		3,59	1	3,59
			192	832		1		3,46	1	3,46
		CTBL1	192	534	2		5,45		2	5,45
		DHX16	192	1032		1		0,77	1	0,77
		ISY1	192	7	1		3,85		1	3,85
			244	7		4		6,46	4	6,46
		PLRG1	32	180		2		4,66	2	4,66
			122	1	4	9	9,20	9,50	13	9,50
		RSMB	32	57		1		2,84	1	2,84
		SF3A1	179	2	2	1	8,21	9,01	3	9,01
		SF3A2	425	103	1		3,41		1	3,41
		SF3B1	266	816	1		3,87		1	3,87
		SMD2	192	51	3	1	3,33	3,10	4	3,33
		SNR40	32	18	1	1	3,35	3,58	2	3,58
			122	1	2	8	8,37	12,84	10	12,84
				8		2		5,93	2	5,93
				18	2	1	8,17	6,13	3	8,17

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
				270	1	7	2,13	7,02	8	7,02
				275	32	30	13,41	10,74	62	13,41
				322	1		3,68		1	3,68
			179	1		2		5,34	2	5,34
				270		2		2,35	2	2,35
			192	270	1		3,37		1	3,37
				275	8	5	12,96	12,89	13	12,96
			244	275	1	6	3,79	8,32	7	8,32
		SPF27	122	168		2		9,89	2	9,89
			179	47	1	1	3,43	5,71	2	5,71
				168	3	5	11,24	11,35	8	11,35
			244	47	2	3	4,13	6,97	5	6,97
				168	5	6	9,50	9,97	11	9,97
				177		1		3,56	1	3,56
				218		4		5,22	4	5,22
			266	168		2		5,50	2	5,50
		SYF1	192	708	2	2	5,70	8,37	4	8,37
			244	708	1	1	3,48	4,26	2	4,26
			266	708		6		12,79	6	12,79
		ZN830	122	104	1		3,81		1	3,81
				149	2	2	8,36	7,80	4	8,36
			425	149	1	2	3,63	3,56	3	3,63
Inter	PRP8	BUD13	1514	560		1		3,73	1	3,73
		BUD31	36	125	1	2	7,01	6,39	3	7,01
			218	9		2		10,33	2	10,33
		CDC5L	85	539	1	2	5,32	5,37	3	5,37
			533	291		1		4,41	1	4,41
			727	28	1		5,79		1	5,79
			892	7	29	19	10,16	8,77	48	10,16
		CTBL1	2070	91	1		2,53		1	2,53
		CWC15	1132	226		3		7,91	3	7,91
			1144	91	5	8	7,47	11,71	13	11,71
				92		1		4,10	1	4,10
			1158	205	1		4,23		1	4,23
		CWC22	366	330	19	20	11,47	13,57	39	13,57
			1392	584		1		1,86	1	1,86
		GPKOW	1831	334		1		4,97	1	4,97
			2108	349		1		4,16	1	4,16
			2244	192		2		2,22	2	2,22
		KIN17	2	129		1		1,80	1	1,80
			609	171		1		0,53	1	0,53
		PHF5A	853	108	1	1	4,16	7,98	2	7,98
		PLRG1	1132	271		1		0,66	1	0,66
		PPIL2	727	226	31	32	16,31	9,03	63	16,31
				231	5	4	5,91	4,06	9	5,91
				232	2		3,60		2	3,60
			1132	26	23	42	10,77	12,63	65	12,63
			1144	166		1		4,98	1	4,98
				180		1		4,70	1	4,70
		PRP17	36	161		1		5,69	1	5,69
			50	399		1		2,62	1	2,62
			162	551	1		2,71		1	2,71
		R113A	442	180	2		9,88		2	9,88
			609	126	6	2	12,06	8,48	8	12,06
			623	139	2	3	9,02	5,64	5	9,02
		RBM22	68	301	1		3,65		1	3,65
		RUXE	267	72	2		0,28		2	0,28
		SF3B1	1505	943	14	7	5,51	3,95	21	5,51
		SF3B2	609	770		1		1,97	1	1,97
			1636	323	1		2,74		1	2,74
		SF3B3	68	942	3	1	4,89	3,32	4	4,89
			1732	2	1		1,43		1	1,43
		SNR40	29	286	3	14	6,07	8,40	17	8,40
			36	286	1	2	8,50	6,65	3	8,50
			43	286	3	2	6,84	4,94	5	6,84
			48	349		2		3,99	2	3,99
			50	349	3	2	10,85	6,83	5	10,85
		SNW1	666	158	5	7	7,07	6,94	12	7,07

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
			702	236	3	2	6,90	7,33	5	7,33
			727	266	1	1	2,73	4,97	2	4,97
			796	266	1	3	3,95	3,63	4	3,95
			847	416	4	6	7,92	6,30	10	7,92
			1505	441	7	2	9,55	11,45	9	11,45
			1801	476	3	3	6,63	6,48	6	6,63
				503	2	2	6,07	6,16	4	6,16
			1838	503	6	5	6,75	5,33	11	6,75
			1984	441		1		3,00	1	3,00
			1993	452		2		5,75	2	5,75
				468	8	10	6,00	6,42	18	6,42
			1994	468		2		3,96	2	3,96
			2031	468		2		5,19	2	5,19
			2049	468	1		3,33		1	3,33
				476	1		4,12		1	4,12
			2113	158		1		2,30	1	2,30
		SPF27	1516	155		1		2,03	1	2,03
		SRRM2	1831	1467		1		2,96	1	2,96
		U520	1210	14	2		7,73		2	7,73
			1300	1557		1		3,59	1	3,59
			1306	14	2	3	6,62	6,62	5	6,62
			1344	14	15	14	14,63	16,45	29	16,45
			1636	1146	4	2	6,86	10,16	6	10,16
			1649	14	4		17,94		4	17,94
			1993	105	3	2	7,46	3,35	5	7,46
				144	5		7,11		5	7,11
				147		1		2,33	1	2,33
			1994	73	29	24	19,19	11,51	53	19,19
			2031	60	1		3,77		1	3,77
				107	1	1	6,25	7,19	2	7,19
				177	3	1	5,59	8,71	4	8,71
				178	1	1	4,54	7,63	2	7,63
				733	6	6	8,07	8,06	12	8,07
				1039	3	1	9,00	4,87	4	9,00
				1049		2		3,88	2	3,88
			2034	60	6	7	14,84	13,43	13	14,84
				177	8	6	11,94	8,73	14	11,94
				178	2		10,24		2	10,24
				733	9	6	7,20	6,80	15	7,20
				770	2	3	6,56	8,85	5	8,85
				1039	15	14	8,48	8,07	29	8,48
				1049	4	6	5,89	10,30	10	10,30
			2049	60	11	8	17,24	17,13	19	17,24
				107	1	1	5,90	4,04	2	5,90
				177	2	2	7,92	5,21	4	7,92
				178	1	1	2,78	1,98	2	2,78
				770	10	13	13,07	11,70	23	13,07
				1034		1		8,06	1	8,06
				1039	17	19	14,60	12,83	36	14,60
				1049	9	7	18,15	20,74	16	20,74
				1134	2	2	5,37	7,43	4	7,43
				1142	4	9	5,86	6,97	13	6,97
				1146	6	7	9,53	10,28	13	10,28
			2070	1034	1		4,40		1	4,40
				1039	12	14	13,61	14,76	26	14,76
				1049	3	5	9,54	13,06	8	13,06
				1145		1		4,63	1	4,63
			2244	256	1		5,36		1	5,36
			2249	14	1	3	5,80	7,57	4	7,57
		U5S1	218	405	50	45	18,31	15,17	95	18,31
				409	33	18	13,05	10,81	51	13,05
			366	359	37	33	14,67	13,05	70	14,67
			1210	712		1		1,29	1	1,29
		ZN830	477	102	1		3,62		1	3,62
Inter	R113A	CWC27	107	192	1		3,03		1	3,03
		DHX16	21	1027		2		4,28	2	4,28
		PRP8	126	609	6	2	12,06	8,48	8	12,06
			139	623	2	3	9,02	5,64	5	9,02

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		Score$_{max}$		Total spec. count	Best Score$_{max}$
					Lumos	QExHF	Lumos	QExHF		
			180	442	2		9,88		2	9,88
		SF3B3	276	1206	30	15	15,80	9,30	45	15,80
				1207		3		9,26	3	9,26
		SF3B4	311	2		1		1,74	1	1,74
		SNR40	164	349	1		2,60		1	2,60
		U520	25	2026	6	5	7,69	6,44	11	7,69
				2087	4	4	4,70	4,95	8	4,95
			319	557	3	5	5,79	4,99	8	5,79
				770	12	6	6,29	3,97	18	6,29
Inter	RBM22	AQR	229	604	1	1	4,79	4,10	2	4,79
		CRNL1	114	224		1		8,39	1	8,39
		CTBL1	170	554		1		0,44	1	0,44
		CWC15	40	28		2		4,60	2	4,60
			76	18	11	8	19,88	18,51	19	19,88
		ISY1	199	92	1		1,09		1	1,09
		PPIL1	286	80		3		6,71	3	6,71
		PRCC	78	428	2		8,00		2	8,00
			104	428		2		4,89	2	4,89
			149	437		1		5,77	1	5,77
				479	1	1	5,49	4,18	2	5,49
			286	442		1		3,25	1	3,25
		PRP8	301	68	1		3,65		1	3,65
		SMD2	185	88		2		2,41	2	2,41
		SNW1	40	153	2	1	4,22	3,40	3	4,22
			76	193	2	1	4,41	4,37	3	4,41
			114	193	1	1	5,63	7,06	2	7,06
			139	193		4		5,63	4	5,63
			286	81	8	8	12,64	21,62	16	21,62
Inter	RBMX2	BUD13	119	483	11	9	16,25	18,46	20	18,46
				486		1		2,82	1	2,82
		CDC5L	243	187		1		2,12	1	2,12
		DHX16	72	860	1		4,73		1	4,73
			246	93		1		4,56	1	4,56
		SF3B1	8	120		1		4,28	1	4,28
				786	1	3	5,29	7,54	4	7,54
			77	700		1		3,48	1	3,48
				741	19	12	12,12	9,04	31	12,12
			101	120	1	5	4,07	8,28	6	8,28
				121		1		6,48	1	6,48
			119	741	1		8,55		1	8,55
			204	741	1		2,06		1	2,06
		SNIP1	71	216	2		5,03		2	5,03
			72	216	3	1	6,18	4,84	4	6,18
		SRRM2	101	228		1		3,23	1	3,23
		U5S1	197	609		1		2,67	1	2,67
			263	352	1		4,35		1	4,35
Inter	RSMB	CRNL1	57	607		1		2,67	1	2,67
		CWC27	50	183	4	2	6,76	6,56	6	6,76
				184	17	16	11,34	8,20	33	11,34
				186	5	2	8,01	8,26	7	8,26
		PRP19	57	32		1		2,84	1	2,84
		SF3A3	54	175	3		4,98		3	4,98
		SF3B2	57	320		1		7,58	1	7,58
		SMD1	5	41	6	9	8,10	7,27	15	8,10
				86		2		5,65	2	5,65
			8	41	3	8	12,27	11,47	11	12,27
		SMD3	5	87	3		4,40		3	4,40
			88	78	1	2	3,23	5,23	3	5,23
				84		3		5,63	3	5,63
				87		1		3,28	1	3,28
		SNR40	57	1	1	1	4,10	4,65	2	4,65
Inter	RU2A	ISY1	56	266	10	10	17,92	11,80	20	17,92
			179	266	1		3,95		1	3,95
			191	266	12	6	10,33	8,40	18	10,33
			192	260	1		5,06		1	5,06
				266		1		7,19	1	7,19
			193	260	1		16,77		1	16,77
				266	46	48	16,71	22,31	94	22,31

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		Score$_{max}$		Total spec. count	Best Score$_{max}$
					Lumos	QExHF	Lumos	QExHF		
			205	259		2		6,82	2	6,82
				260	20	16	8,98	8,82	36	8,98
			221	260	2		4,90		2	4,90
				266	29	28	19,59	20,12	57	20,12
		PPIE	172	123	2	4	6,97	6,56	6	6,97
				134	4	2	11,41	5,01	6	11,41
			179	134	1		4,45		1	4,45
			193	114	1		4,11		1	4,11
				134	1	1	4,64	3,79	2	4,64
		RU2B	30	101		1		3,90	1	3,90
				111	2	4	9,04	15,53	6	15,53
		SF3A1	67	264		1		1,49	1	1,49
			193	316		2		8,13	2	8,13
			205	223	3	3	5,52	4,84	6	5,52
			221	188	3		7,24		3	7,24
		SF3A3	56	69		1		5,33	1	5,33
			172	64	1	1	5,84	3,43	2	5,84
				69	4	4	5,87	4,42	8	5,87
			179	64		1		9,27	1	9,27
				69	11	21	15,81	11,99	32	15,81
			221	64		3		10,20	3	10,20
				69		5		11,00	5	11,00
		SF3B1	191	175	4	4	4,88	5,22	8	5,22
				177	2		4,32		2	4,32
			192	175		1		4,31	1	4,31
				177	1		3,34		1	3,34
		SF3B2	179	280	1	4	3,71	4,03	5	4,03
		SYF1	205	2	39	20	11,40	8,14	59	11,40
Inter	RU2B	CTBL1	1	83		1		2,36	1	2,36
				84	1		3,22		1	3,22
		RU2A	101	30		1		3,90	1	3,90
			111	30	2	4	9,04	15,53	6	15,53
		SF3A3	103	463	3	3	11,65	9,14	6	11,65
			111	463	2	1	4,12	4,46	3	4,46
			224	496	8	5	10,73	11,11	13	11,11
			225	496	1	1	11,23	10,13	2	11,23
		SF3B2	85	352		1		2,95	1	2,95
			93	352		1		3,52	1	3,52
			101	352		1		6,45	1	6,45
			103	352		1		4,32	1	4,32
			111	352	8	3	9,25	6,22	11	9,25
		SF3B3	111	1074	7	7	13,76	13,15	14	13,76
		SMD3	101	31		1		4,43	1	4,43
		SNIP1	111	79		1		3,11	1	3,11
Inter	RUXE	CWC22	80	745		1		2,02	1	2,02
		PLRG1	80	47		1		2,44	1	2,44
		PRP8	72	267	2		0,28		2	0,28
		SF3B2	80	323		1		3,86	1	3,86
Inter	SF3A1	CWC27	115	192		2		4,67	2	4,67
		GPKOW	210	349		1		2,09	1	2,09
		ISY1	2	7	1	1	6,62	5,50	2	6,62
				84		1		4,80	1	4,80
				121		1		5,27	1	5,27
			30	84	1		3,78		1	3,78
				127	2		6,59		2	6,59
				260	1	1	2,57	1,77	2	2,57
			37	84	1	1	3,99	4,08	2	4,08
				101	1		4,45		1	4,45
				105	6	3	7,82	5,61	9	7,82
				127	1	1	5,55	5,21	2	5,55
			115	55	4	4	8,45	6,34	8	8,45
				84	5	5	7,39	7,41	10	7,41
				105	2	2	5,94	4,29	4	5,94
				121		1		7,04	1	7,04
			131	121	1		3,26		1	3,26
		PPIE	2	218	8	5	10,00	9,59	13	10,00
			131	104		3		4,62	3	4,62
			210	88	7	2	9,14	12,51	9	12,51

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		Score$_{max}$		Total spec. count	Best Score$_{max}$
					Lumos	QExHF	Lumos	QExHF		
				104	9	5	8,65	8,89	14	8,89
			223	104	1	2	3,84	7,60	3	7,60
				108	2	3	5,30	3,69	5	5,30
				114	1	4	2,88	4,01	5	4,01
			230	108	2		7,92		2	7,92
				114	1		6,53		1	6,53
				123	4	1	10,80	4,93	5	10,80
			231	108	4	3	10,01	7,41	7	10,01
				114		2		5,50	2	5,50
				123	3	3	11,17	8,80	6	11,17
				134	1	4	11,97	6,50	5	11,97
			251	108	1		5,99		1	5,99
		PRP19	2	179	2	1	8,21	9,01	3	9,01
		RU2A	188	221	3		7,24		3	7,24
			223	205	3	3	5,52	4,84	6	5,52
			264	67		1		1,49	1	1,49
			316	193		2		8,13	2	8,13
		SF3A2	102	118		2		4,23	2	4,23
		SF3A3	131	219	5	3	8,40	6,04	8	8,40
				220	8		23,37		8	23,37
			188	89	15	9	7,82	8,08	24	8,08
				92	26	19	11,18	8,92	45	11,18
			231	212	1	2	2,73	7,00	3	7,00
			251	29	13	6	10,18	8,09	19	10,18
			258	29	19	26	14,88	12,53	45	14,88
			259	29	3		7,84		3	7,84
		SF3B1	467	522	1		1,52		1	1,52
		SF3B2	115	466	1		3,09		1	3,09
			210	275	5	2	10,50	8,25	7	10,50
			217	275	2		5,37		2	5,37
		SYF1	2	257		1		7,77	1	7,77
				708	13	14	13,35	10,58	27	13,35
				794	2	1	6,04	5,25	3	6,04
			55	549	12	6	12,34	15,68	18	15,68
				590		1		8,43	1	8,43
				794	4		6,87		4	6,87
			115	590	4		10,10		4	10,10
				794		2		4,31	2	4,31
			131	794	1		4,59		1	4,59
			223	76	2		4,04		2	4,04
			399	708		2		9,76	2	9,76
			533	708		1		5,14	1	5,14
Inter	SF3A2	AQR	42	662		2		2,80	2	2,80
		CRNL1	118	175	1	1	4,00	1,27	2	4,00
			194	680	1		0,89		1	0,89
		CWC27	103	184	1	1	4,88	3,44	2	4,88
		KIN17	48	157		1		2,23	1	2,23
		PLRG1	118	320		1		2,42	1	2,42
		PPIE	42	284		1		3,99	1	3,99
		PPIL2	91	418	10	20	11,08	9,21	30	11,08
			101	392		1		4,79	1	4,79
				393		3		9,76	3	9,76
			107	423	7	16	7,66	8,46	23	8,46
		PRCC	121	414		1		1,89	1	1,89
		PRP19	103	425	1		3,41		1	3,41
		SF3A1	118	102		2		4,23	2	4,23
		SF3A3	77	390	1		3,78		1	3,78
			91	390	28	21	14,70	13,47	49	14,70
		SF3B4	88	46	1	2	2,63	2,80	3	2,80
			91	2		1		5,41	1	5,41
		SNIP1	121	342	1		4,92		1	4,92
Inter	SF3A3	ISY1	69	266	2	1	23,89	5,77	3	23,89
		PPIE	219	134	5	1	6,68	2,68	6	6,68
		PPIL1	97	52		1		2,39	1	2,39
		RSMB	175	54	3		4,98		3	4,98
		RU2A	64	172	1	1	5,84	3,43	2	5,84
				179		1		9,27	1	9,27
				221		3		10,20	3	10,20

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
			69	56		1		5,33	1	5,33
				172	4	4	5,87	4,42	8	5,87
				179	11	21	15,81	11,99	32	15,81
				221		5		11,00	5	11,00
		RU2B	463	103	3	3	11,65	9,14	6	11,65
				111	2	1	4,12	4,46	3	4,46
			496	224	8	5	10,73	11,11	13	11,11
				225	1	1	11,23	10,13	2	11,23
		SF3A1	29	251	13	6	10,18	8,09	19	10,18
				258	19	26	14,88	12,53	45	14,88
				259	3		7,84		3	7,84
			89	188	15	9	7,82	8,08	24	8,08
			92	188	26	19	11,18	8,92	45	11,18
			212	231	1	2	2,73	7,00	3	7,00
			219	131	5	3	8,40	6,04	8	8,40
			220	131	8		23,37		8	23,37
		SF3A2	390	77	1		3,78		1	3,78
				91	28	21	14,70	13,47	49	14,70
		SF3B2	69	280	19	17	7,06	5,96	36	7,06
			420	514	1	3	6,50	5,19	4	6,50
			466	604		1		4,30	1	4,30
		SF3B3	89	1074	1		3,50		1	3,50
			463	1074	27	36	9,17	11,96	63	11,96
			466	980		1		6,24	1	6,24
				984		4		5,27	4	5,27
				1074	17	16	10,78	8,64	33	10,78
			496	974	6	8	11,29	12,14	14	12,14
		SF3B4	29	183	3	1	7,38	5,48	4	7,38
		SMD2	157	118	1		8,17		1	8,17
		SRRM2	308	1428	1		0,88		1	0,88
Inter	SF3B1	BUD13	816	499	3		6,12		3	6,12
		CTBL1	6	31	2	2	6,24	5,44	4	6,24
				56	2	2	10,60	11,18	4	11,18
				57		1		14,05	1	14,05
				83	16	21	19,29	17,06	37	19,29
				84	10	7	10,38	11,45	17	11,45
				91	9	6	9,13	6,39	15	9,13
				95	10	8	8,97	6,16	18	8,97
			21	91	5	5	8,40	5,87	10	8,40
				95	2		6,69		2	6,69
		DHX16	6	69	21	14	10,16	6,71	35	10,16
			21	69	5	2	8,24	6,29	7	8,24
			22	69	2	7	4,67	6,28	9	6,28
				73		1		8,12	1	8,12
			733	860	2	3	11,31	11,23	5	11,31
		PPIL2	6	82		1		3,27	1	3,27
				90	1		4,32		1	4,32
			496	198		1		2,70	1	2,70
		PRP19	816	266	1		3,87		1	3,87
		PRP8	943	1505	14	7	5,51	3,95	21	5,51
		RBMX2	120	8		1		4,28	1	4,28
				101	1	5	4,07	8,28	6	8,28
			121	101		1		6,48	1	6,48
			700	77		1		3,48	1	3,48
			741	77	19	12	12,12	9,04	31	12,12
				119	1		8,55		1	8,55
				204	1		2,06		1	2,06
			786	8	1	3	5,29	7,54	4	7,54
		RU2A	175	191	4	4	4,88	5,22	8	5,22
				192		1		4,31	1	4,31
			177	191	2		4,32		2	4,32
				192	1		3,34		1	3,34
		SF3A1	522	467	1		1,52		1	1,52
		SF3B2	1292	486	16	14	9,55	9,77	30	9,77
		SF3B3	1292	965	11	7	7,24	8,07	18	8,07
		SF3B6	163	125	3	9	9,96	10,75	12	10,75
			175	116	1	1	3,56	4,55	2	4,55
			182	116	1		5,64		1	5,64

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
			195	116	9	6	12,64	9,99	15	12,64
			656	125		3		9,77	3	9,77
			700	125	4	4	11,70	8,15	8	11,70
		SNIP1	182	301		3		4,75	3	4,75
			700	223	3		15,42		3	15,42
			816	216	4		6,48		4	6,48
		SNW1	6	311	1	1	4,52	4,61	2	4,61
			22	379		1		3,30	1	3,30
			816	456	17	12	9,20	7,41	29	9,20
				468		3		3,28	3	3,28
			943	441	15	7	7,71	7,07	22	7,71
		U520	21	343		1		5,52	1	5,52
Inter	SF3B2	BUD13	326	339		1		1,38	1	1,38
		CDC5L	543	187	1	2	4,15	7,04	3	7,04
			570	187		1		1,76	1	1,76
			577	174		1		2,47	1	2,47
		CRNL1	570	175	2	2	3,94	3,93	4	3,94
				178		1		1,57	1	1,57
				182	5	4	8,03	11,74	9	11,74
		CWC15	556	91	2		7,23		2	7,23
		ISY1	556	41		1		1,94	1	1,94
		PPIE	275	88		2		6,70	2	6,70
				104	1		7,39		1	7,39
		PRP17	570	440	4	4	9,59	5,60	8	9,59
		PRP8	323	1636	1		2,74		1	2,74
			770	609		1		1,97	1	1,97
		RSMB	320	57		1		7,58	1	7,58
		RU2A	280	179	1	4	3,71	4,03	5	4,03
		RU2B	352	85		1		2,95	1	2,95
				93		1		3,52	1	3,52
				101		1		6,45	1	6,45
				103		1		4,32	1	4,32
				111	8	3	9,25	6,22	11	9,25
		RUXE	323	80		1		3,86	1	3,86
		SF3A1	275	210	5	2	10,50	8,25	7	10,50
				217	2		5,37		2	5,37
			466	115	1		3,09		1	3,09
		SF3A3	280	69	19	17	7,06	5,96	36	7,06
			514	420	1	3	6,50	5,19	4	6,50
			604	466		1		4,30	1	4,30
		SF3B1	486	1292	16	14	9,55	9,77	30	9,77
		SF3B3	602	1074		1		4,37	1	4,37
			604	1074	14	14	13,82	9,05	28	13,82
			605	1074	22	20	20,59	19,63	42	20,59
			877	1206	1		4,22		1	4,22
		SMD3	352	31	2	1	7,40	6,95	3	7,40
		SNR40	891	270	1		3,30		1	3,30
		SNW1	877	193		1		11,22	1	11,22
		SYF1	275	36	2	1	4,38	4,67	3	4,67
		U520	877	469		1		5,67	1	5,67
				1874	4	2	12,50	10,24	6	12,50
Inter	SF3B3	DHX16	126	1019	12	8	12,64	11,21	20	12,64
				1027	34	48	11,86	11,97	82	11,97
			296	1019	3	1	5,58	9,21	4	9,21
		KIN17	26	231	11	5	12,74	9,88	16	12,74
			296	231		3		9,47	3	9,47
		PHF5A	137	13		1		4,82	1	4,82
				95		1		4,86	1	4,86
		PRP8	2	1732	1		1,43		1	1,43
			942	68	3	1	4,89	3,32	4	4,89
		R113A	1206	276	30	15	15,80	9,30	45	15,80
			1207	276		3		9,26	3	9,26
		RU2B	1074	111	7	7	13,76	13,15	14	13,76
		SF3A3	974	496	6	8	11,29	12,14	14	12,14
			980	466		1		6,24	1	6,24
			984	466		4		5,27	4	5,27
			1074	89	1		3,50		1	3,50
				463	27	36	9,17	11,96	63	11,96

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
				466	17	16	10,78	8,64	33	10,78
		SF3B1	965	1292	11	7	7,24	8,07	18	8,07
		SF3B2	1074	602		1		4,37	1	4,37
				604	14	14	13,82	9,05	28	13,82
				605	22	20	20,59	19,63	42	20,59
			1206	877	1		4,22		1	4,22
Inter	SF3B4	R113A	2	311		1		1,74	1	1,74
		SF3A2	2	91		1		5,41	1	5,41
			46	88	1	2	2,63	2,80	3	2,80
		SF3A3	183	29	3	1	7,38	5,48	4	7,38
Inter	SF3B6	CDC5L	29	106	1	2	3,04	5,82	3	5,82
		SF3B1	116	175	1	1	3,56	4,55	2	4,55
				182	1		5,64		1	5,64
				195	9	6	12,64	9,99	15	12,64
			125	163	3	9	9,96	10,75	12	10,75
				656		3		9,77	3	9,77
				700	4	4	11,70	8,15	8	11,70
		SMD2	106	98	1		0,89		1	0,89
		SNIP1	125	216	1		6,66		1	6,66
Inter	SMD1	CDC5L	86	291	2		8,38		2	8,38
		KIN17	9	33	1		0,01		1	0,01
		RSMB	41	5	6	9	8,10	7,27	15	8,10
				8	3	8	12,27	11,47	11	12,27
			86	5		2		5,65	2	5,65
		SMD2	1	37		1		2,15	1	2,15
			86	18	2	2	3,91	6,35	4	6,35
		SNR40	44	8		1		2,89	1	2,89
		ZN830	44	102	1		4,67		1	4,67
Inter	SMD2	CDC5L	98	264	1		1,02		1	1,02
			118	782		1		9,25	1	9,25
		CWC22	8	149	1		0,50		1	0,50
			50	250		1		0,20	1	0,20
		DHX16	98	228	1		1,71		1	1,71
		ISY1	98	259	2		3,99		2	3,99
				260	3	3	5,24	4,86	6	5,24
		PPIL2	98	230	1		3,66		1	3,66
		PRP19	51	192	3	1	3,33	3,10	4	3,33
		RBM22	88	185		2		2,41	2	2,41
		SF3A3	118	157	1		8,17		1	8,17
		SF3B6	98	106	1		0,89		1	0,89
		SMD1	18	86	2	2	3,91	6,35	4	6,35
			37	1		1		2,15	1	2,15
		SRRM2	98	2667	1		4,89		1	4,89
Inter	SMD3	CDC5L	104	201		1		2,93	1	2,93
		IF4A3	99	15	2	1	5,81	3,26	3	5,81
		RSMB	78	88	1	2	3,23	5,23	3	5,23
			84	88		3		5,63	3	5,63
			87	5	3		4,40		3	4,40
				88		1		3,28	1	3,28
		RU2B	31	101		1		4,43	1	4,43
		SF3B2	31	352	2	1	7,40	6,95	3	7,40
		U5S1	87	421	1	4	4,15	6,06	5	6,06
Inter	SNIP1	DHX16	216	860	3	2	9,62	9,69	5	9,69
		RBMX2	216	71	2		5,03		2	5,03
				72	3	1	6,18	4,84	4	6,18
		RU2B	79	111		1		3,11	1	3,11
		SF3A2	342	121	1		4,92		1	4,92
		SF3B1	216	816	4		6,48		4	6,48
			223	700	3		15,42		3	15,42
			301	182		3		4,75	3	4,75
		SF3B6	216	125	1		6,66		1	6,66
		SNW1	342	339	49	45	15,91	15,15	94	15,91
				344	28	25	14,37	16,79	53	16,79
				376		1		3,63	1	3,63
				379	13	18	4,50	5,08	31	5,08
			353	339	12	15	9,17	10,08	27	10,08
			355	319	1		3,29		1	3,29
				323	14	12	12,54	10,15	26	12,54

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
				330	1		5,76		1	5,76
				339	7	15	7,39	9,38	22	9,38
				344		2		5,68	2	5,68
				379	2	1	6,72	3,36	3	6,72
			359	339	4		9,68		4	9,68
		U520	2	2080	1		3,60		1	3,60
		U5S1	257	788	1		2,73		1	2,73
Inter	SNR40	CDC5L	18	771	7	6	6,41	7,68	13	7,68
			270	782	2	2	2,95	6,68	4	6,68
				795	17	11	9,39	5,90	28	9,39
			275	771		2		6,92	2	6,92
				782		8		10,60	8	10,60
		PPIL1	1	80	5	6	5,97	7,97	11	7,97
			8	80		1		6,01	1	6,01
			18	80	3	1	4,39	3,94	4	4,39
				161	1		3,56		1	3,56
		PRP19	1	122	2	8	8,37	12,84	10	12,84
				179		2		5,34	2	5,34
			8	122		2		5,93	2	5,93
			18	32	1	1	3,35	3,58	2	3,58
				122	2	1	8,17	6,13	3	8,17
			270	122	1	7	2,13	7,02	8	7,02
				179		2		2,35	2	2,35
				192	1		3,37		1	3,37
			275	122	32	30	13,41	10,74	62	13,41
				192	8	5	12,96	12,89	13	12,96
				244	1	6	3,79	8,32	7	8,32
			322	122	1		3,68		1	3,68
		PRP8	286	29	3	14	6,07	8,40	17	8,40
				36	1	2	8,50	6,65	3	8,50
				43	3	2	6,84	4,94	5	6,84
			349	48		2		3,99	2	3,99
				50	3	2	10,85	6,83	5	10,85
		R113A	349	164	1		2,60		1	2,60
		RSMB	1	57	1	1	4,10	4,65	2	4,65
		SF3B2	270	891	1		3,30		1	3,30
		SMD1	8	44		1		2,89	1	2,89
		SNW1	1	153		1		5,48	1	5,48
				193	2	2	10,60	9,60	4	10,60
			18	48	1		4,59		1	4,59
				97	2	2	5,04	5,42	4	5,42
				193	2	1	8,73	4,68	3	8,73
			349	170	2	1	5,73	4,68	3	5,73
		SPF27	18	158	1	4	6,09	11,91	5	11,91
				168	2	6	7,38	8,82	8	8,82
				177		2		4,80	2	4,80
			270	168		1		1,69	1	1,69
			275	168	27	28	16,45	17,59	55	17,59
				177	1	3	7,65	4,58	4	7,65
				191	4	4	8,97	8,43	8	8,97
			322	168	1		4,43		1	4,43
		SYF1	8	62		1		1,44	1	1,44
Inter	SNW1	BUD13	441	604	2	2	5,54	5,35	4	5,54
				605	7	5	14,66	17,06	12	17,06
			503	474	1	1	2,71	3,47	2	3,47
		CDC5L	258	28	1		5,25		1	5,25
			266	28	3	9	8,05	9,99	12	9,99
			311	124	1		5,37		1	5,37
		CRNL1	48	357		1		5,04	1	5,04
			108	397	5	5	9,26	8,29	10	9,26
			110	397	8	5	8,54	9,53	13	9,53
			115	397	3	3	10,11	5,18	6	10,11
			217	229		2		3,02	2	3,02
		CWC15	153	18	21	24	11,11	11,74	45	11,74
			311	152		3		11,67	3	11,67
		CWC22	376	648	1		9,05		1	9,05
		ISY1	317	121		1		1,68	1	1,68
		PLRG1	110	173		1		4,48	1	4,48

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
				180	3	5	8,32	9,25	8	9,25
				181	1		3,51		1	3,51
			153	363		1		6,44	1	6,44
			158	320	18	15	14,27	12,23	33	14,27
				363	4	4	8,96	4,65	8	8,96
			217	363		2		3,31	2	3,31
		PPIL1	81	58	1		12,15		1	12,15
			97	158	28	20	14,71	12,35	48	14,71
			193	52	1		2,66		1	2,66
				80	3	11	6,74	7,82	14	7,82
		PPIL2	240	277	7	3	8,35	11,94	10	11,94
		PRCC	48	137	4		3,86		4	3,86
				139	2		3,32		2	3,32
			97	193		2		9,21	2	9,21
			110	193		2		7,98	2	7,98
				207		1		5,50	1	5,50
			193	414	3	10	9,29	8,51	13	9,29
		PRP8	158	666	5	7	7,07	6,94	12	7,07
				2113		1		2,30	1	2,30
			236	702	3	2	6,90	7,33	5	7,33
			266	727	1	1	2,73	4,97	2	4,97
				796	1	3	3,95	3,63	4	3,95
			416	847	4	6	7,92	6,30	10	7,92
			441	1505	7	2	9,55	11,45	9	11,45
				1984		1		3,00	1	3,00
			452	1993		2		5,75	2	5,75
			468	1993	8	10	6,00	6,42	18	6,42
				1994		2		3,96	2	3,96
				2031		2		5,19	2	5,19
				2049	1		3,33		1	3,33
			476	1801	3	3	6,63	6,48	6	6,63
				2049	1		4,12		1	4,12
			503	1801	2	2	6,07	6,16	4	6,16
				1838	6	5	6,75	5,33	11	6,75
		RBM22	81	286	8	8	12,64	21,62	16	21,62
			153	40	2	1	4,22	3,40	3	4,22
			193	76	2	1	4,41	4,37	3	4,41
				114	1	1	5,63	7,06	2	7,06
				139		4		5,63	4	5,63
		SF3B1	311	6	1	1	4,52	4,61	2	4,61
			379	22		1		3,30	1	3,30
			441	943	15	7	7,71	7,07	22	7,71
			456	816	17	12	9,20	7,41	29	9,20
			468	816		3		3,28	3	3,28
		SF3B2	193	877		1		11,22	1	11,22
		SNIP1	319	355	1		3,29		1	3,29
			323	355	14	12	12,54	10,15	26	12,54
			330	355	1		5,76		1	5,76
			339	342	49	45	15,91	15,15	94	15,91
				353	12	15	9,17	10,08	27	10,08
				355	7	15	7,39	9,38	22	9,38
				359	4		9,68		4	9,68
			344	342	28	25	14,37	16,79	53	16,79
				355		2		5,68	2	5,68
			376	342		1		3,63	1	3,63
			379	342	13	18	4,50	5,08	31	5,08
				355	2	1	6,72	3,36	3	6,72
		SNR40	48	18	1		4,59		1	4,59
			97	18	2	2	5,04	5,42	4	5,42
			153	1		1		5,48	1	5,48
			170	349	2	1	5,73	4,68	3	5,73
			193	1	2	2	10,60	9,60	4	10,60
				18	2	1	8,73	4,68	3	8,73
		SYF1	48	420	2		7,80		2	7,80
		U520	236	151	1		5,86		1	5,86
			452	1176	1		5,23		1	5,23
		ZN830	311	13	3	1	6,34	5,53	4	6,34
			315	13	4	3	8,84	8,25	7	8,84

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		Score$_{max}$		Total spec. count	Best Score$_{max}$
					Lumos	QExHF	Lumos	QExHF		
Inter	SPF27	BUD31	177	42	1		0,43		1	0,43
		CDC5L	97	685	11	6	5,63	4,70	17	5,63
			168	312	3		5,87		3	5,87
			177	771	69	69	15,74	16,94	138	16,94
			191	782	12	10	10,53	10,17	22	10,53
			218	795	2	2	4,79	4,37	4	4,79
				799		2		3,49	2	3,49
		CWC22	155	598		2		3,80	2	3,80
		PPIL1	168	80	10	12	9,74	8,90	22	9,74
		PRP19	47	179	1	1	3,43	5,71	2	5,71
				244	2	3	4,13	6,97	5	6,97
			168	122		2		9,89	2	9,89
				179	3	5	11,24	11,35	8	11,35
				244	5	6	9,50	9,97	11	9,97
				266		2		5,50	2	5,50
			177	244		1		3,56	1	3,56
			218	244		4		5,22	4	5,22
		PRP8	155	1516		1		2,03	1	2,03
		SNR40	158	18	1	4	6,09	11,91	5	11,91
			168	18	2	6	7,38	8,82	8	8,82
				270		1		1,69	1	1,69
				275	27	28	16,45	17,59	55	17,59
				322	1		4,43		1	4,43
			177	18		2		4,80	2	4,80
				275	1	3	7,65	4,58	4	7,65
			191	275	4	4	8,97	8,43	8	8,97
		ZN830	168	149	4		8,13		4	8,13
Inter	SRRM2	DHX16	130	69		3		5,97	3	5,97
			131	69		1		3,77	1	3,77
		ISY1	169	259		1		1,87	1	1,87
		PRP8	1467	1831		1		2,96	1	2,96
		RBMX2	228	101		1		3,23	1	3,23
		SF3A3	1428	308	1		0,88		1	0,88
		SMD2	2667	98	1		4,89		1	4,89
		U520	101	599	1		4,79		1	4,79
			169	1134		1		3,61	1	3,61
			2680	1176	2	1	5,43	4,57	3	5,43
		ZN830	161	216		1		2,96	1	2,96
Inter	SYF1	AQR	2	950	3	3	11,71	11,71	6	11,71
				954	12	16	6,32	8,25	28	8,25
				1002	5	1	6,34	5,28	6	6,34
		CDC5L	532	522	12	11	6,94	7,18	23	7,18
			654	290	7	9	5,96	11,85	16	11,85
				294	31	23	22,28	18,10	54	22,28
			677	312	1	1	8,19	11,57	2	11,57
		CRNL1	482	427	15	9	9,90	6,95	24	9,90
				460		1		7,93	1	7,93
				461	4	9	5,37	7,04	13	7,04
			794	787	4	2	6,49	7,70	6	7,70
		CWC22	654	330	1		5,83		1	5,83
		CWC27	539	180	1	1	4,39	3,50	2	4,39
		ISY1	2	190	5	3	5,76	6,21	8	6,21
				260	10	11	10,22	12,04	21	12,04
				266	2	1	4,43	5,57	3	5,57
			83	188	5	3	5,37	4,77	8	5,37
			257	84		1		3,12	1	3,12
				138	11	8	11,39	12,61	19	12,61
			794	7	1		4,32		1	4,32
				41		4		8,69	4	8,69
				44		2		3,17	2	3,17
		PLRG1	2	237		1		1,77	1	1,77
			420	135	1	1	5,77	5,48	2	5,77
		PPIE	2	218	3	11	6,32	10,13	14	10,13
			36	108	3	4	5,59	7,83	7	7,83
				114	1	1	5,93	3,81	2	5,93
			76	104	1	1	3,93	5,43	2	5,43
				108	5	9	4,14	5,19	14	5,19
				123	1	1	3,11	2,68	2	3,11

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
				134	3		6,00		3	6,00
			212	108	1		3,66		1	3,66
				114		1		4,71	1	4,71
		PPIL2	50	418	1		3,79		1	3,79
		PRCC	420	207	3	5	7,47	7,31	8	7,47
				216	2	2	9,41	7,17	4	9,41
				223	1	3	5,40	8,12	4	8,12
			423	193		2		7,77	2	7,77
			458	193	13	10	10,82	11,29	23	11,29
				207	5	2	8,55	7,76	7	8,55
				216	1	1	4,41	3,41	2	4,41
		PRP19	708	192	2	2	5,70	8,37	4	8,37
				244	1	1	3,48	4,26	2	4,26
				266		6		12,79	6	12,79
		RU2A	2	205	39	20	11,40	8,14	59	11,40
		SF3A1	76	223	2		4,04		2	4,04
			257	2		1		7,77	1	7,77
			549	55	12	6	12,34	15,68	18	15,68
			590	55		1		8,43	1	8,43
				115	4		10,10		4	10,10
			708	2	13	14	13,35	10,58	27	13,35
				399		2		9,76	2	9,76
				533		1		5,14	1	5,14
			794	2	2	1	6,04	5,25	3	6,04
				55	4		6,87		4	6,87
				115		2		4,31	2	4,31
				131	1		4,59		1	4,59
		SF3B2	36	275	2	1	4,38	4,67	3	4,67
		SNR40	62	8		1		1,44	1	1,44
		SNW1	420	48	2		7,80		2	7,80
		U520	677	699		1		4,92	1	4,92
		ZN830	482	279	2	5	5,46	8,41	7	8,41
			708	117	1		8,06		1	8,06
				137	5	4	7,48	9,08	9	9,08
				140	2	1	7,00	5,29	3	7,00
			794	140	1		4,10		1	4,10
				149		1		3,84	1	3,84
Inter	U520	BUD13	975	68	1		8,76		1	8,76
		CWC22	1146	584		1		1,41	1	1,41
		GPKOW	14	50		1		3,05	1	3,05
				192		1		3,71	1	3,71
			46	192	3	1	6,91	4,54	4	6,91
			60	50	1	3	4,91	6,76	4	6,76
			73	192	4	5	5,89	5,80	9	5,89
			83	192	3	6	4,73	6,77	9	6,77
			155	346	1	1	3,67	3,87	2	3,87
			695	61		2		6,13	2	6,13
			1049	372		1		9,98	1	9,98
		ISY1	2091	260		1		3,02	1	3,02
		KIN17	1146	293		2		3,19	2	3,19
			1169	272		1		3,73	1	3,73
		PLRG1	453	487		1		2,88	1	2,88
		PRP8	14	1210	2		7,73		2	7,73
				1306	2	3	6,62	6,62	5	6,62
				1344	15	14	14,63	16,45	29	16,45
				1649	4		17,94		4	17,94
				2249	1	3	5,80	7,57	4	7,57
			60	2031	1		3,77		1	3,77
				2034	6	7	14,84	13,43	13	14,84
				2049	11	8	17,24	17,13	19	17,24
			73	1994	29	24	19,19	11,51	53	19,19
			105	1993	3	2	7,46	3,35	5	7,46
			107	2031	1	1	6,25	7,19	2	7,19
				2049	1	1	5,90	4,04	2	5,90
			144	1993	5		7,11		5	7,11
			147	1993		1		2,33	1	2,33
			177	2031	3	1	5,59	8,71	4	8,71
				2034	8	6	11,94	8,73	14	11,94

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		Score$_{max}$		Total spec. count	Best Score$_{max}$
					Lumos	QExHF	Lumos	QExHF		
				2049	2	2	7,92	5,21	4	7,92
			178	2031	1	1	4,54	7,63	2	7,63
				2034	2		10,24		2	10,24
				2049	1	1	2,78	1,98	2	2,78
			256	2244	1		5,36		1	5,36
			733	2031	6	6	8,07	8,06	12	8,07
				2034	9	6	7,20	6,80	15	7,20
			770	2034	2	3	6,56	8,85	5	8,85
				2049	10	13	13,07	11,70	23	13,07
			1034	2049		1		8,06	1	8,06
				2070	1		4,40		1	4,40
			1039	2031	3	1	9,00	4,87	4	9,00
				2034	15	14	8,48	8,07	29	8,48
				2049	17	19	14,60	12,83	36	14,60
				2070	12	14	13,61	14,76	26	14,76
			1049	2031		2		3,88	2	3,88
				2034	4	6	5,89	10,30	10	10,30
				2049	9	7	18,15	20,74	16	20,74
				2070	3	5	9,54	13,06	8	13,06
			1134	2049	2	2	5,37	7,43	4	7,43
			1142	2049	4	9	5,86	6,97	13	6,97
			1145	2070		1		4,63	1	4,63
			1146	1636	4	2	6,86	10,16	6	10,16
				2049	6	7	9,53	10,28	13	10,28
			1557	1300		1		3,59	1	3,59
		R113A	557	319	3	5	5,79	4,99	8	5,79
			770	319	12	6	6,29	3,97	18	6,29
			2026	25	6	5	7,69	6,44	11	7,69
			2087	25	4	4	4,70	4,95	8	4,95
		SF3B1	343	21		1		5,52	1	5,52
		SF3B2	469	877		1		5,67	1	5,67
			1874	877	4	2	12,50	10,24	6	12,50
		SNIP1	2080	2	1		3,60		1	3,60
		SNW1	151	236	1		5,86		1	5,86
			1176	452	1		5,23		1	5,23
		SRRM2	599	101	1		4,79		1	4,79
			1134	169		1		3,61	1	3,61
			1176	2680	2	1	5,43	4,57	3	5,43
		SYF1	699	677		1		4,92	1	4,92
		U5S1	284	682		1		8,05	1	8,05
		ZN830	1134	44		1		2,63	1	2,63
Inter	U5S1	CDC5L	446	466	4	6	4,12	4,94	10	4,94
		CWC22	359	330	1		6,22		1	6,22
		CWC27	352	393	6	2	14,28	5,73	8	14,28
			359	393		1		5,13	1	5,13
		IF4A3	244	152	4	7	11,23	12,38	11	12,38
			694	152	1	1	4,58	3,56	2	4,58
		KIN17	341	10	1	3	3,79	3,53	4	3,79
		PRP8	359	366	37	33	14,67	13,05	70	14,67
			405	218	50	45	18,31	15,17	95	18,31
			409	218	33	18	13,05	10,81	51	13,05
			712	1210		1		1,29	1	1,29
		RBMX2	352	263	1		4,35		1	4,35
			609	197		1		2,67	1	2,67
		SMD3	421	87	1	4	4,15	6,06	5	6,06
		SNIP1	788	257	1		2,73		1	2,73
		U520	682	284		1		8,05	1	8,05
Inter	ZN830	AQR	339	201	2		5,02		2	5,02
			343	201	5	1	9,34	6,06	6	9,34
		CDC5L	137	255	1		4,97		1	4,97
			149	255	1	2	3,25	3,74	3	3,74
		CRNL1	83	771		2		6,42	2	6,42
			216	445	2	2	5,16	5,44	4	5,44
			228	445	5	8	9,15	7,06	13	9,15
				475	2	1	3,24	3,04	3	3,24
			234	445	18	20	10,46	13,76	38	13,76
				460	11	16	12,03	10,60	27	12,03
				475	8	6	11,26	11,79	14	11,79

Type	Protein 1	Protein 2	Residue 1	Residue 2	Spectral count		$Score_{max}$		Total spec. count	Best $Score_{max}$
					Lumos	QExHF	Lumos	QExHF		
				485	1		2,84		1	2,84
		CTBL1	78	56	5	7	8,63	8,82	12	8,82
			104	56		7		18,19	7	18,19
			117	31		1		4,87	1	4,87
			149	31		1		4,97	1	4,97
		CWC15	26	147	3	3	5,54	5,95	6	5,95
			28	183	1		2,89		1	2,89
			272	40	1		0,60		1	0,60
		PPIL2	182	313	4		7,49		4	7,49
				462	1		8,86		1	8,86
			190	313	5	14	11,12	11,80	19	11,80
				450		2		4,28	2	4,28
				462	2	1	5,23	3,06	3	5,23
			216	313		2		7,50	2	7,50
			228	313	12	14	10,15	8,92	26	10,15
				450	8	15	7,97	7,69	23	7,97
				454	2	4	4,88	9,32	6	9,32
				462	7	7	10,95	10,98	14	10,98
			234	312	1		4,53		1	4,53
				313	29	33	13,86	14,99	62	14,99
				450	4	7	5,51	7,02	11	7,02
				454	10	8	8,24	8,84	18	8,84
				460	4		7,98		4	7,98
				462	13	9	12,81	12,31	22	12,81
			262	277		1		7,82	1	7,82
		PRP19	104	122	1		3,81		1	3,81
			149	122	2	2	8,36	7,80	4	8,36
				425	1	2	3,63	3,56	3	3,63
		PRP8	102	477	1		3,62		1	3,62
		SMD1	102	44	1		4,67		1	4,67
		SNW1	13	311	3	1	6,34	5,53	4	6,34
				315	4	3	8,84	8,25	7	8,84
		SPF27	149	168	4		8,13		4	8,13
		SRRM2	216	161		1		2,96	1	2,96
		SYF1	117	708	1		8,06		1	8,06
			137	708	5	4	7,48	9,08	9	9,08
			140	708	2	1	7,00	5,29	3	7,00
				794	1		4,10		1	4,10
			149	794		1		3,84	1	3,84
			279	482	2	5	5,46	8,41	7	8,41
		U520	44	1134		1		2,63	1	2,63