







# Investigation of the structure of spliceosomal complexes from the yeast *S. cerevisiae*

Dissertation

For the award of the degree

**"Doctor of Philosophy (Ph.D.)"**

Division of Mathematics and Natural Sciences  
of the **Georg-August-Universität Göttingen**

Within the doctoral program Biology  
of the Georg-August University School of Science (GAUSS)

Submitted by

**Vinay Kumar**

from Delhi, INDIA

Göttingen, 2020

**Members of the Thesis Committee:**

**Prof. Dr. Reinhard Lührmann**

Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen

**Prof. Dr. Ralf Ficner**

Department for Molecular Structural Biology, Institute for Microbiology and Genetics, Georg-August-Universität, Göttingen

**Members of the Examination Board:**

**Prof. Dr. Reinhard Lührmann (Reviewer)**

Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen

**Prof. Dr. Ralf Ficner (Second Reviewer)**

Department for Molecular Structural Biology, Institute for Microbiology and Genetics, Georg-August-Universität, Göttingen

**Prof. Dr. Henning Urlaub**

Research group - Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen

**Prof. Dr. Holger Stark**

Department of Structural Dynamics, Max Planck Institute for Biophysical Chemistry, Göttingen

**Prof. Dr. Jörg Großhans**

Department of Developmental Biochemistry, Institute of Biochemistry and Molecular Cell Biology, University Medical Center, Göttingen

**Prof. Dr. Ralph Kehlenbach**

Department of Molecular Biology, Institute of Molecular Biology, University Medical Center, Göttingen

Date of submission: **February 7th, 2020**

Date of the oral examination: **March 20th, 2020**

*Harshey  
&  
My Parents*

## **Affidavit**

**I hereby declare that I prepared this doctoral thesis titled " Investigation of the structure of spliceosomal complexes from the yeast *S. cerevisiae*"  
*independently and with no other sources and aids than quoted***

**Göttingen, February- 2020**

---

**Vinay Kumar**

## Table of Contents

<b>Abstract</b> .....	<b>11</b>
<b>1 Introduction</b> .....	<b>12</b>
<b>1.1 Gene structure and pre-mRNA processing</b> .....	<b>12</b>
<b>1.2 Intron-exon defining conserved sequences in Pre-mRNA splicing</b> .....	<b>13</b>
<b>1.3 The basic chemistry behind the two-step mechanism of splicing</b> .....	<b>15</b>
<b>1.4 Step-by-step assembly of the spliceosome</b> .....	<b>17</b>
<b>1.5 The spliceosomal snRNPs</b> .....	<b>19</b>
1.5.1 The snRNA constituents of U snRNPs.....	20
1.5.2 The protein inventory of U snRNPs.....	24
1.5.3 The non-snRNP factors of the splicing machinery .....	25
<b>1.6 The U4/U6.U5 tri-snRNP</b> .....	<b>27</b>
1.6.1 A large building block of the spliceosome .....	27
1.6.2 The tri-snRNP undergoes large structural rearrangements upon spliceosome activation .....	30
1.6.3 The tri-snRNP proteins play a key role in the catalytic activation of the spliceosome and spliceosomal dynamics .....	35
<b>1.7 Structural analysis of spliceosomal complexes by cryo-electron microscopy</b> .....	<b>37</b>
<b>1.7.1 Structural differences in human and yeast tri-snRNP particles.</b> .....	<b>38</b>
1.7.2 The position of several proteins in human and yeast tri-snRNPs is dramatically different .....	39
1.7.3 Structural arrangement of Sad1 in human tri-snRNP and its potential role .....	41
1.7.4 Rationale for the proposed work .....	42
<b>2 Materials and methods</b> .....	<b>43</b>
<b>2.1 Materials</b> .....	<b>43</b>
2.1.1 Chemicals and media.....	43
2.1.2 Consumable laboratory material .....	45
2.1.3 Enzymes and enzyme inhibitors .....	46
2.1.4 Monoclonal and polyclonal antibodies .....	46
2.1.5 Nucleotides .....	46
2.1.6 Oligonucleotides.....	47

2.1.7 Yeast strains .....	47
2.1.8 Commercial reaction sets (kits) .....	47
2.1.9 Laboratory apparatuses.....	48
2.1.10 Cell culture.....	50
2.1.11 Common buffers .....	50
2.1.12 Commercial reaction sets (kits) .....	53
2.1.13 Software programs .....	53
<b>2.2 Methods .....</b>	<b>54</b>
2.2 Protein-biochemistry standard methods .....	54
2.2.1 Protein quantification .....	54
2.2.2 PCI (Phenol-Chloroform-Isoamylalcohol) extraction .....	54
2.2.3 Denaturing polyacrylamide gel electrophoresis (SDS-PAGE).....	55
2.2.4 Coomassie staining .....	55
2.2.5 Silver staining of protein gels.....	56
2.2.6 Western blot analysis.....	56
<b>2.3 Molecular biology .....</b>	<b>57</b>
2.3.1 Concentration determination of nucleic acids .....	57
2.3.2 Agarose gel electrophoresis of nucleic acids .....	57
2.3.3 Proteinase K digestion.....	58
2.3.4 Denaturing polyacrylamide gel electrophoresis of RNA .....	58
2.3.5 Silver staining of RNA gels .....	58
2.3.6 SYBR™ Gold nucleic acid staining.....	59
2.3.7 Radioactively labelled DNA-probes synthesis for Northern analysis .....	59
2.3.8 Northern blot analysis .....	60
<b>2.4 Cell culture methods .....</b>	<b>60</b>
2.4.1 Yeast cell culture .....	60
2.4.2 Extract preparation from yeast cells for snRNP purification using an electric grinder.....	61
<b>2.5 Special methods .....</b>	<b>62</b>
2.5.1 Tandem affinity purification of yeast snRNPs .....	62
2.5.2 Glycerol gradient sedimentation of purified snRNP particles .....	63
<b>2.6 Electron Microscopy .....</b>	<b>63</b>
2.6.1 Preparation of continuous carbon film and holey carbon grids.....	63
2.6.2 Sample preparation .....	64

2.6.3 Negative stain EM analysis .....	64
2.6.4 Cryo- EM image acquisition and analysis .....	65
2.6.5 Model fitting and building .....	65
2.6.6 Electron microscopy analysis of TAP purified snRNPs .....	65
<b>2.7 Mass Spectrometry .....</b>	<b>66</b>
2.7.1 Protein identification by mass spectrometry .....	66
2.7.2 Identification of Protein-protein crosslinks by mass spectrometry .....	67
<b>3 RESULTS .....</b>	<b>69</b>
<b>3.1 Tri-snRNP isolation with Tandem Affinity Purification Technique .....</b>	<b>69</b>
3.1.1 Purification of tri-snRNPs using TAP-tagged Brr2p .....	69
3.1.2 Brr2-mediated ATP sensitivity assay of purified Brr2p-TAP yeast tri- snRNPs under 150 mM KCl (high salt) conditions .....	72
3.1.3 Purification and ATP-sensitivity assay of Brr2-TAP tri-snRNPs under 75 mM KCl conditions .....	76
3.1.4.1 Purification of tri-snRNPs using a TAP-tagged Sad1p strain. ....	80
3.1.4.2 Purification of the Sad1-TAP tri-snRNP at 75 mM K <sup>+</sup> conditions .....	83
3.1.4.3 ATP-sensitivity assay for Sad1-TAP tri-snRNPs, purified under 75 mM KCl conditions .....	87
3.1.4.4 Sad1-TAP tri-snRNP purification at high salt conditions (150mM) .....	91
3.1.4.5 Resistance of purified Sad1-TAP tri-snRNPs to Brr2-mediated dissociation .....	95
3.1.4.6 Detailed analysis of protein distribution and tri-snRNP composition upon shift in salt concentration .....	96
<b>3.2 Electron microscopy of affinity-purified yeast tri-snRNPs .....</b>	<b>101</b>
3.2.1 Negative-stain 2D EM analysis of the yeast Sad1-TAP tri-snRNPs .....	102
3.2.2 3D Structure determination of the yeast Sad1-TAP tri-snRNP particles	107
3.2.2.1 Low-resolution negative-stain EM 3D model reconstruction .....	108
3.2.2.2 Cryo-EM 3D structure reconstruction of Sad1-TAP tri-snRNP .....	110
3.2.2.3 Placing available X-ray, NMR and EM-reconstructed structures of certain canonical factors in our tri-snRNP model .....	116
3.2.3 Novel features of the yeast tri-snRNP structure .....	118
<b>3.3 BS3 crosslinking of affinity-purified yeast Sad1-TAP tri-snRNPs .....</b>	<b>122</b>
3.3.1 Crosslinks supporting the Sad1 locations in the yeast tri-snRNPs .....	122

3.3.2 Building a human-like conformation model of the yeast tri-snRNP .....	129
<b>4 Discussion .....</b>	<b>133</b>
<b>4.1 Biochemical aspects .....</b>	<b>134</b>
4.1.1 An improved tri-snRNP is required to solve the question of significant differences in the published structures of tri-snRNPs.....	134
4.1.2 A new purification strategy for yeast tri-snRNPs. ....	135
4.1.3 The Prp38 – Snu23 – Spp381 group of proteins. ....	137
<b>4.2 Structural aspects .....</b>	<b>138</b>
4.2.1 Negative-stain EM of yeast tri-snRNPs. ....	138
4.2.2 Brr2 orientation in yeast Sad1-TAP tri-snRNP.....	141
4.2.3 The Sad1 organisation in yeast tri-snRNPs is similar to its human counterpart.....	141
4.2.4 The lack of Prp38 in yeast Sad1-tri-snRNPs may correlate with a missing region in the structure. ....	143
4.2.5 Prp28 in yeast tri-snRNPs shows similar interactions like its human counterpart.....	143
<b>5 REFERENCES.....</b>	<b>147</b>
<b>APPENDIX.....</b>	<b>157</b>
<b>SUPPLIMENTRY TABLE S1.....</b>	<b>161</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>183</b>



## Abstract

Nuclear pre-mRNA splicing is catalysed by the spliceosome, a multi-megadalton ribonucleoprotein (RNP) complex. It assembles de novo for each round of splicing on pre-mRNA intron with stepwise binding of five small nuclear ribonucleoprotein particles (snRNPs), and numerous proteins. The process of splicing initiates with the U1 and U2 snRNPs associating with the pre-mRNA's 5' splice site (SS) and branch site (BS), respectively, giving rise to the A complex. The newly formed A complex is further joined by the U4/U6.U5 tri-snRNP, which results in the formation of the precatalytic B complex. During the highly choreographed event of spliceosomal activation, Brr2 dissociates U4 RNA from U6 RNA, which are base-paired together in the tri-snRNP, and U6 RNA restructures and together with U2 RNA forms the active site. It has always been a major question, how Brr2 is prevented in isolated U4/U6.U5 tri-snRNPs from pre-maturely dissociating the U4/U6 RNAs. In the human tri-snRNP Brr2 is situated away (~10 nm) from its RNA substrate, the U4/U6 duplex. The human Sad1 protein has a key role in stabilizing the particle. Sad1 is present in stoichiometric amounts and it is located in a strategically important position at the interface between U4/U6 and U5 snRNP where it maintains numerous protein-protein contacts to act as a stabilizing clamp between the two snRNPs and at the same time inhibits premature access of Brr2 to the U4 RNA. In yeast instead, the reported tri-snRNP structures show that Brr2 is already loaded onto its U4 snRNA substrate, and ready to unwind the U4/U6 duplex. In fact, in the presence of ATP it does unwind and dissociate the tri-snRNP while the human tri-snRNP remains stable. The stability difference in the presence of ATP has been attributed to the presence of Sad1 in the human tri-snRNP, while Sad1 is usually not observed in isolated yeast tri-snRNPs. In order to prove the proposed role of Sad1 I set out to isolate and characterize yeast tri-snRNP particles which are similar to human tri-snRNPs in terms of protein composition and biochemical behaviour. The challenge rest in finding biochemical conditions and modified purification protocols, which would provide such yeast tri-snRNPs, which are then used for biochemical and structural comparison with the human tri-snRNP. Yeast tri-snRNPs were purified using a two-step procedure, using a Sad1-TAP tag and gradient centrifugation under suitable buffer conditions. I succeeded in purifying a tri-snRNP which is stable under ATP conditions. This tri-snRNP for the first time contained a stably associated Sad1 and Prp28. It appeared that both proteins may have an important role in stabilizing the yeast tri-snRNP. Additionally, I found that Sad1 and Prp38 are mutually exclusive proteins. Despite the drastically changed biochemical characteristics of the tri-snRNP purified under the novel conditions, Brr2 still appears to adopt the typical yeast-like conformation.

# 1 Introduction

## 1.1 Gene structure and pre-mRNA processing

Gene expression comprises the transfer of DNA encoded information either into the form of functionally important non-coding RNAs or into proteins. In order to make use of the DNA encoded blueprints for proteins during translation, the DNA is first transcribed into messenger RNA, which is extensively processed already in the nucleus and then in the cytoplasm after the passage through the nuclear membrane. The mature messenger RNA then serves as a template during ribosomal protein synthesis.

Primary RNA transcripts are the product of RNA polymerases. While the protein-coding DNA sequences in prokaryotic genes are continuous and the resulting messenger RNA (mRNA) transcripts are therefore readily available for protein translation, most genes of the higher eukaryotes show an alternating mosaic pattern of coding and non-coding DNA sequences. Their primary RNA transcripts (precursor-mRNAs) still contain the coding as well as the non-coding sequences and extensive processing of the precursor is therefore required before the RNA transcript can serve as a template during translation. In the case of eukaryotes, RNA polymerase II initially transcribes mRNA molecules as pre-mRNA.

The maturation process of a primary transcript begins with the co-transcriptional capping at the 5' end of the nuclear pre-mRNA. A 7-methylguanosine ( $m^7G$ ) is linked by an atypical 5'-5'-triphosphate bond to the ribose at the 5' end of the pre-mRNA (Shatkin, 1976). This cap-structure is required for mRNA export from the nucleus to the cytoplasm, protein synthesis initiation, and the stabilization of the mRNA (McCracken et al., 1997). A process known as polyadenylation occurs at the 3' end of the pre-mRNA during which poly(A)-polymerase adds a tail of 20 to 50 adenine residues .

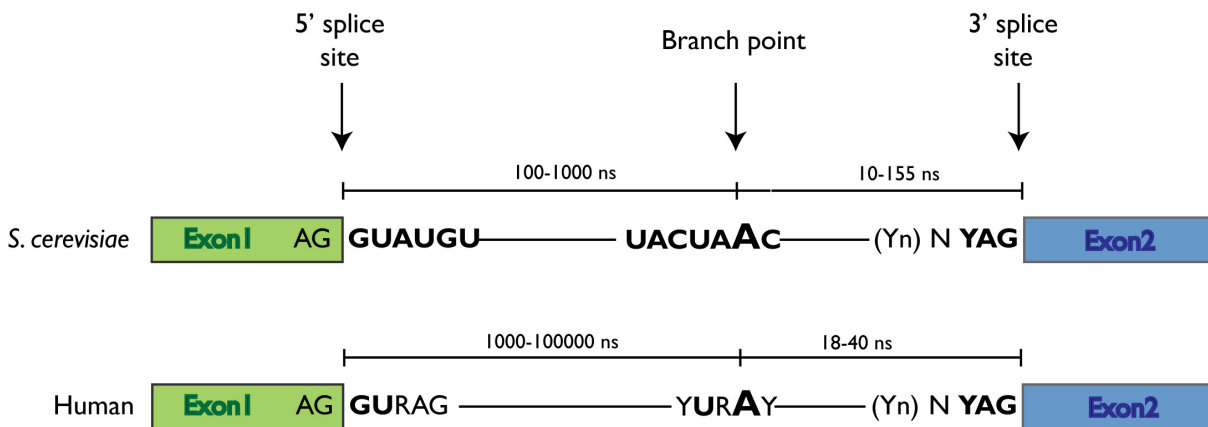
One of the most remarkable processing steps a primary transcript undergoes is the removal of non-coding sequences (introns). Since a eukaryotic gene in the vast majority of cases is a "gene in pieces", that is, coding sequences (exons) are interspersed by non-coding sequences (introns), only removal of introns from pre-mRNA and subsequent ligation of exons can create a continuous stretch of coding sequence, the mature mRNA (Berget et al., 1977; Chow et al., 1977; Gilbert, 1978; Tonegawa et al., 1978). This process is called splicing. This process occurs in the nucleus and is catalyzed by a highly complex multi-mega Dalton ribonucleoprotein complex known as the spliceosome (Will and Lührmann, 2011b)) Splicing not only results in a single defined continuous mature transcript, but since many eukaryotic pre-mRNAs harbor introns that

are sometimes included into, or excluded from the final mature mRNA, incorporation or exclusion of an intron in the final transcript during so-called alternative splicing events can create a variety of related, but distinct mature mRNAs originating from the same primary transcript of a gene. Alternative splicing therefore substantially increases the complexity of the eukaryotic proteome which comprises a much higher number of distinct protein species than the mere number of genes (Black, 2003).

## 1.2 Intron-exon defining conserved sequences in Pre-mRNA splicing

A precise recognition of a genuine splice site on the pre-mRNA is crucial because an error of only one single nucleotide would shift the reading frame forward or backward on the mature mRNA resulting in a non-functional protein product. Limited information content defining the area of exon-intron boundaries, vastly flexible intron lengths and sequences, make the task of splice site recognition very challenging. Each intron present on the pre-mRNA is recognized by three conserved sequences: the 5' splice site (5' SS), the branch-point (BP) sequence and the 3' splice site (3' SS) (fig: 1.1). In yeast (refers to *Saccharomyces cerevisiae*; *S. cerevisiae* in this thesis) these cis-acting recognition elements are highly conserved, however, less so in human, making the right selection of splice sites comparatively more complex. In yeast intron(s), the 5' splice site can be described by the consensus sequence **AAG|GUAUGU** (|indicating the exon-intron border, and the nucleotides shown in bold represent 90% or more conserved sequences in yeast introns (Ma and Xia, 2011);(Lopez and Séraphin, 1999). The branch-point (BP) sequence in yeast **UACUAAC** (Neuveglise et al., 2011), shown here with the branch point A, contains the BP adenosine residue crucial for step I catalysis of splicing (details in the following section). The BP adenosine in *S. cerevisiae* introns is located approximately 30 nucleotides upstream of the 3' SS (Neuveglise et al., 2011; Spingola et al., 1999). Finally, the consensus sequence found at the 3' splice site (3' SS) in yeast introns represented as **YAG|** (here Y stands for a pyrimidine) is highly conserved. Compared to yeast, in human the intron defining sequences are more degenerate. The 5' splice site sequence is **G|GURAG**, the branch site sequence is indicated by **YURAY**, and the consensus sequence at the 3' splice site (3' SS) in human introns can be described by **YAG|** (Zhang, 1998). Furthermore, in yeast as well as in human, upstream to the 3' SS a so-called polypyrimidine-tract (a pyrimidine-rich region) is present. The polypyrimidine tract in yeast can vary from 8-12 nucleotides and ends just one nucleotide upstream to the 3' SS. Because in human the branch-point sequences are less conserved, this region potentially plays an important role.

Also in the case of yeast, a preference for uridine-rich RNA sequences upstream to 3' SS has been reported (Patterson and Guthrie, 1991).



**Figure 1.1** Cartoon depiction of conserved intronic consensus sequences of *S. cerevisiae* and human.

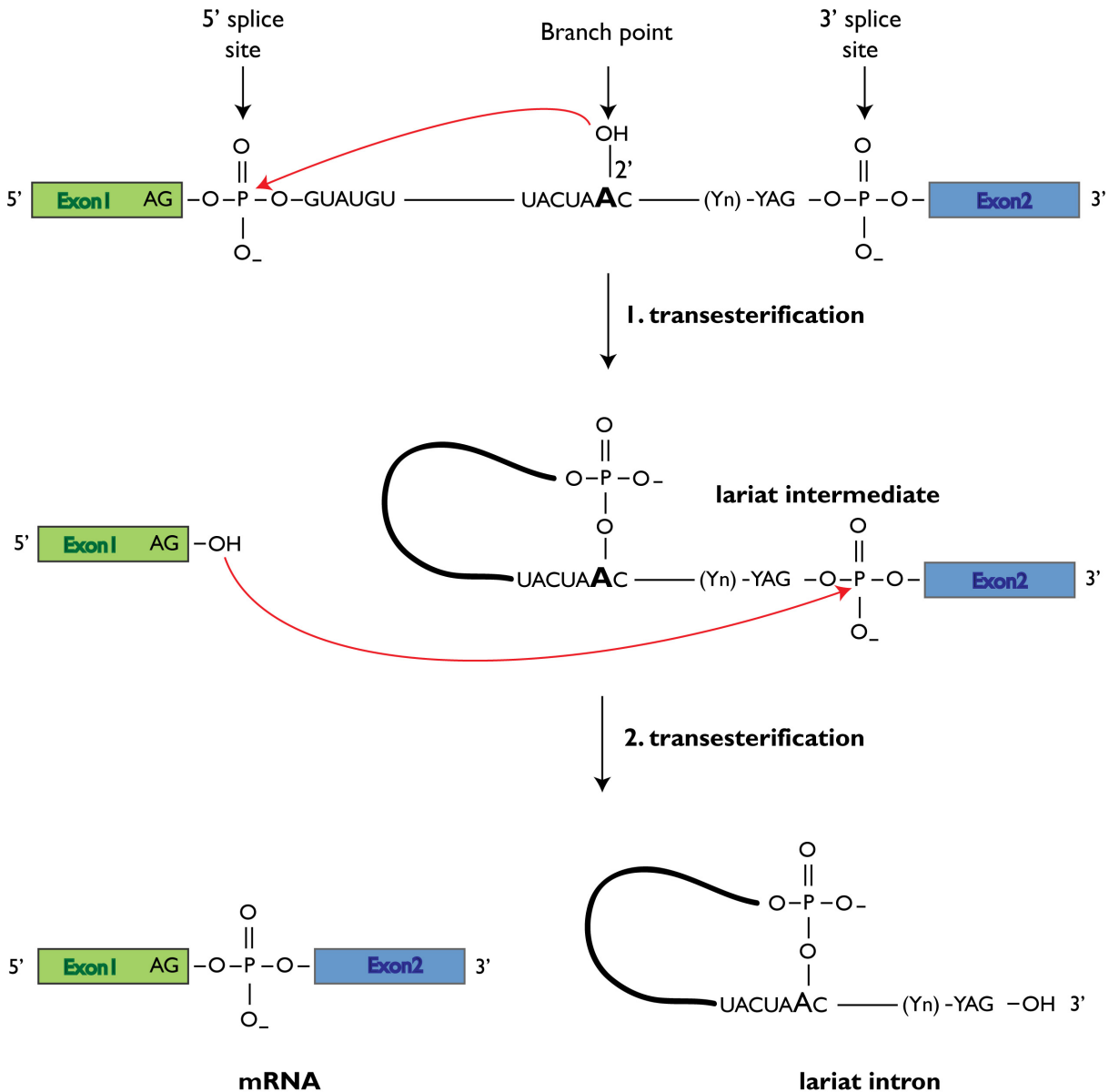
Rectangular boxes represent Exon1 (green) and Exon 2 (blue), consensus sequences with 90% or higher conservation are shown in bold. The intron between the two exons is shown as a continuous line. The BP adenosine is indicated here in larger font. The polypyrimidine tract is represented by (Yn); purines are shown with 'R'; ns = nucleotides. The scheme has not been drawn to scale. Distances between the 5' SS, BP and the 3' SS are based on the available bioinformatics data. Image was kindly provided by Dr. Patrizia Fabrizio.

Availability of the complete genomic sequence of the yeast *S. cerevisiae* (Goffeau et al., 1996) has provided us the substantial information about the chromosomal organization. Experimental and bioinformatics analysis has revealed interesting information about yeast introns. Studies revealed that in total the yeast genome contains 253 introns that are located on 248 genes out of 6000 genes. Although accounting for merely 3.8% of all the genes, 27.1% of all synthesized transcripts are spliced (Lopez and Séraphin, 1999; Spingola et al., 1999).

Yeast introns are usually short. In previous studies the length distribution has shown two peaks located mainly at 100 and 400 nucleotides. Very few introns have been reported to be 1000 nucleotides long (Spingola et al., 1999). On the other hand, in humans' genes contain multiple introns varying from 6-9 introns per gene with length varying from 20 – 100 000 nucleotides. In addition, < 0.01% of the introns are < 20 bp in length and < 10% of introns are more than 11,000 bp in length. (Sakharkar et al., 2004)

### 1.3 The basic chemistry behind the two-step mechanism of splicing

Excision of nuclear pre-mRNA introns during the splicing process entails two separate transesterification (phosphodiester transfer) reactions (Moore et al., 1993; Moore and Sharp, 1993). The first step begins with a nucleophilic attack by a 2' OH group, situated at the ribose moiety of the branch-point adenosine, on the phosphate bond, between the 5' splice site guanosine and the last nucleotide of the 5'-exon (fig: 1.2). The nucleophilic attack results in a free 5' exon with a 3' OH group, and a branched lariat of intron – 3' exon with a unique 2' – 5' phosphodiester bond. The subsequent second step involves a nucleophilic attack by the newly generated 3' OH group of the ribose moiety at the excised 5'-exon, on the phosphate bond between the 3' SS G and the first nucleotide of exon 2. With the ligation of 5' and 3' exons and the release of the intron-lariat, the second step of splicing results in the formation of mature mRNA. As a post splicing event, a liberated intron in the form of a lariat structure rapidly converts to a linear form (debranching) and is then degraded (Arenas and Hurwitz, 1987) (Moore et al., 1993). In yeast, as well as in other organisms, introns have been regarded as dispensable by-products of splicing as they get rapidly degraded after exon ligation; however, recent studies have suggested that in yeast, introns potentially play an important role in cell survival in starvation conditions. (Edwards and Johnson, 2019); (Parenteau et al., 2019); (Morgan et al., 2019)



**Figure 1.2 Basic chemistry behind the two-transesterification reactions in splicing.**

Splicing is carried out with two consecutive phosphoester exchange reactions resulting in the excision of the lariat-intron and the ligation of the two exons. Long curved arrows indicate the directions of nucleophilic attacks in each step by the respective nucleophilic agent (2'-OH group of the BP adenosine in step I reaction, and 3'-OH of the terminal guanosine nucleotide in the step II reaction). The formation of the intron lariat loop is shown here as a bold loop with the distinctive 2'—5' bond. The panel at the bottom shows two exons ligated after splicing and the free intron lariat. Details of the reaction are described in text section 1.3. The branchpoint adenosine is indicated in bold, the polypyrimidine tract is represented by (Y<sub>n</sub>).

## 1.4 Step-by-step assembly of the spliceosome

Splicing is catalysed by an intricate macromolecular machine, the spliceosome. The spliceosome not only has a complex parts list and an intricate assembly pathway but it also highly dynamic during its entire life cycle of catalysis. The spliceosome assembles from various small nuclear ribonucleoprotein particles (snRNPs); namely, U1, U2, U5, and U4/U6 snRNPs, and numerous non-snRNP splicing supporting factors (Will and Lührmann, 2011b)). The spliceosome assembles de-novo on each pre-mRNA intron in a step-wise manner for each splicing event. Fig: 1.3 depicts an overview of the splicing cycle.

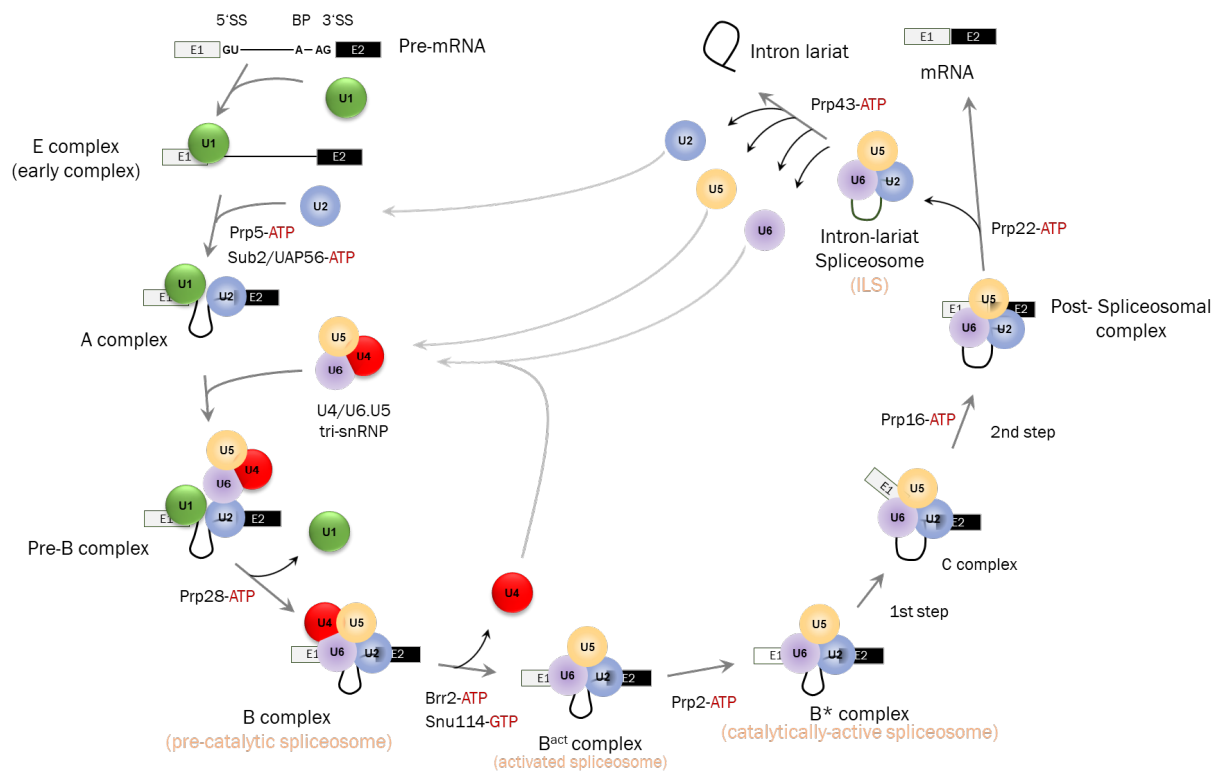
The spliceosome undergoes various highly dynamic structural changes during which many protein components are acquired or released from the complex. Many of these intermediate complexes are stable enough so that they can be isolated and their protein components be determined (Wahl et al., 2009; Will and Lührmann, 2011b). Spliceosomal E, A, pre-B, B B<sup>act</sup>, B\*, C, C\*, and post-splicing complexes have already been described using the various biochemical and genetics based experimental approach in human and yeast (Kastner et al., 2019)

The assembly starts with a stepwise association of various key components, and during the course of splicing structural changes occur which result in the formation of the spliceosomal active site. These remodelling events are primarily due to alterations in RNA-RNA and RNA-protein interactions, and are mainly driven by eight DEXD/H-box ATPases or RNA helicases. (Cordin and Beggs, 2013)

As a first step, the U1 snRNP binds to pre-mRNA via base-pairing interaction between the 5'SS and the 5' end of the U1 snRNA. This interaction is further enhanced by U1-associated proteins (Legrain et al., 1988)) (fig: 1.3). After, or concomitant with stable binding of the U1 snRNP to the 5' SS, the branch-point sequence and downstream sequences are recognized by Msl5/BBP and Mud2 (Berglund et al., 1997); (Abovich et al., 1994). Unstable binding of the U2 snRNP may occur at this point (Das et al., 2000). The complex E is now remodelled in an ATP-dependent manner by the helicases Sub2 and Prp5 resulting in a stable base pairing of U2 snRNA to the BPS (Fleckner et al., 1997); (Zhang and Green, 2001). The branch point adenosine is bulged out (Query et al., 1994). Now, the SF3a and SF3b sub-complexes of the U2 snRNP bind to a region upstream of the BPS, replacing the previous Msl5/BBP interaction and thus establishing the stable U2 snRNP anchoring they are yielding the so-called A complex (pre-spliceosome) (Gozani et al., 1996). Subsequently, the pre-formed U4/U6.U5 tri-snRNP complex (Stevens and Abelson, 1999) joins the A complex resulting in the formation of pre-B complex. Notably, in human the tri-snRNP is not yet stably bound (Boesler et al., 2016). This newly formed complex has all the snRNPs required for a complete spliceosome, however, it does still not have an active catalytic

center. An RNA helicase, Prp28, has to disrupt the U1/5'SS helix, which results in relocation of the 5'SS from U1 to U6 snRNA, which allows the base-pairing of the 5'SS with the conserved ACAGA box region of U6 snRNA (Staley and Guthrie, 1999). Consequently, U1 gets destabilized and the precatalytic B complex is formed. At this stage, the spliceosome is still catalytically inactive. Only after the action of Brr2, another RNA helicase, which unwinds the U4/U6 snRNA duplex (Cordin and Beggs, 2013), (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998) thus liberating U6 snRNA and releasing the U4 snRNP along with many associated proteins from the spliceosome, the liberated U6 snRNA can refold to form an intramolecular stem-loop and new base-pairing interactions with U2 snRNA. These new rearrangements finally result in the formation of the catalytic center in the heart of the spliceosome and promote the association of several factors such as nineteen complex (NTC), nineteen-related complex (NTR) (Chan et al., 2003), and numerous B<sup>act</sup> specific proteins. This newly generated complex is known as the activated complex or B<sup>act</sup> complex in the splicing cycle. Subsequently, another important helicase, Prp2 acts on the B<sup>act</sup> spliceosome and forms the catalytically activated B\* complex which catalyses step I of splicing (Kim and Lin, 1996; Warkocki et al., 2009). After step I catalysis the resulting C complex is characterized by the presence of a cleaved 5'-exon and the lariat-3'-exon. The Prp16 helicase now acts on the C complex and generates the C\* complex which accomplishes the second catalytic step (step II) of splicing (ligation of two exons) (Schwer and Guthrie, 1992; Tseng et al., 2011). Later remodelling generates the P (post-splicing) complex (Horowitz, 2012). The release of the ligated mature mRNA is carried out by the action of Prp22 helicase (Company et al., 1991; Schwer, 2008) which leads to the intron-lariat spliceosome (ILS), which is then disassembled by helicase Prp43 (Arenas and Abelson, 1997; Fourmann et al., 2013; Tsai et al., 2005). The U2, U5, and U6 snRNAs are recycled for the next round of splicing.





**Figure: 1.3 Schematic representation of the progressive spliceosome assembly and restructuring throughout a single round of splicing cycle.** Assembly starts with the joining of the U1 snRNP at the 5'SS of the pre-mRNA, which results in the E complex. Subsequent binding of the U2 snRNP at the BPS creates the A complex. The pre-assembled U4/U6.U5 snRNP joins the A complex resulting in pre-B complex formation. Upon stable integration of the tri-snRNP, U1snRNP leaves and the B complex is formed. The B complex is still pre-catalytic and only after the release of U4 becomes the activated complex (B<sup>act</sup>). Subsequent action of Prp2 (a DEAH box RNA helicase) converts the B<sup>act</sup> into a catalytically active B\* complex. The first transesterification reaction results in the formation of the C complex and, thereafter, the second splicing reaction leads to the formation of the mature mRNA and the intron lariat. During post-splicing, the spliceosomal components dissociate and are regenerated for the next round of splicing.

For the purpose of simplified depiction, a U snRNP is represented as a circle labelled with its respective U snRNA. The eight conserved DExD/H-box ATPases/helicases, which are essential for vital RNP remodelling events, are shown and their role is described in detail in section 1.4 and section 1.5.3. Exons and introns are indicated as coloured boxes (with E1 and E2) or bold lines, respectively. Recycling of snRNPs is indicated by light blue arrows. (Sandra Keiper kindly provided the image, which was further modified based on data from Will and Lührmann, 2011; Boesler *et al.*, 2015).

## 1.5 The spliceosomal snRNPs

The pre-formed major building blocks of the spliceosome are U snRNPs (ribonucleoprotein particles), namely, U1, U2, U5, and U4/U6. Each spliceosomal snRNP contains one uridine-rich small nuclear RNA (U snRNA) (with the only exception of the U4/U6 snRNP which contains two U snRNAs), an Sm protein ring (Sm-ring) with seven different proteins (except of U6 which

uses Lsm proteins instead), and various snRNP-specific proteins (fig: 1.6) (Will and Lührmann, 2011b).

### 1.5.1 The snRNA constituents of U snRNPs

U snRNAs are transcribed by RNA polymerase II from their respective genes. Only the U6 snRNA is transcribed by RNA polymerase III. The primary transcript of U5 snRNA gets cotranscriptionally cleaved at a 3' stem-loop by RNase III. This differential trimming in yeast gives rise to two different isoforms of U5 snRNA i.e., U5 long (U5L) and U5 short (U5S) snRNA (Chanfreau *et al.*, 1997). All U snRNAs (except U6, which carries a gamma-monomethyl cap [Didychuk et al 2019]) are modified in the nucleus with a 5' monomethylated m7G cap. With a total of 1175 nucleotides, U2 snRNA is the longest of all spliceosomal snRNAs in yeast (Table 1.1). Compared to yeast, human U2 snRNA is only 187 nucleotide long (however, it is the longest human snRNA). All U pre-snRNAs are transported into the cytoplasm for additional processing, during which their 5'-end regions now obtain a hypermethylated 2, 2, 7-trimethylguanosine (m3G) (Mattaj and De Robertis, 1985). This cap serves as an important signal for nuclear import.

U snRNAs (yeast)		U snRNAs (human)	
U snRNA	Length (ns)	U snRNA	Length (ns)
U1	568	U1	164
U2	1175	U2	187
U4	160	U4	145
U5L	214	U5	116
U5S	179	--	--
U6	112	U6	106

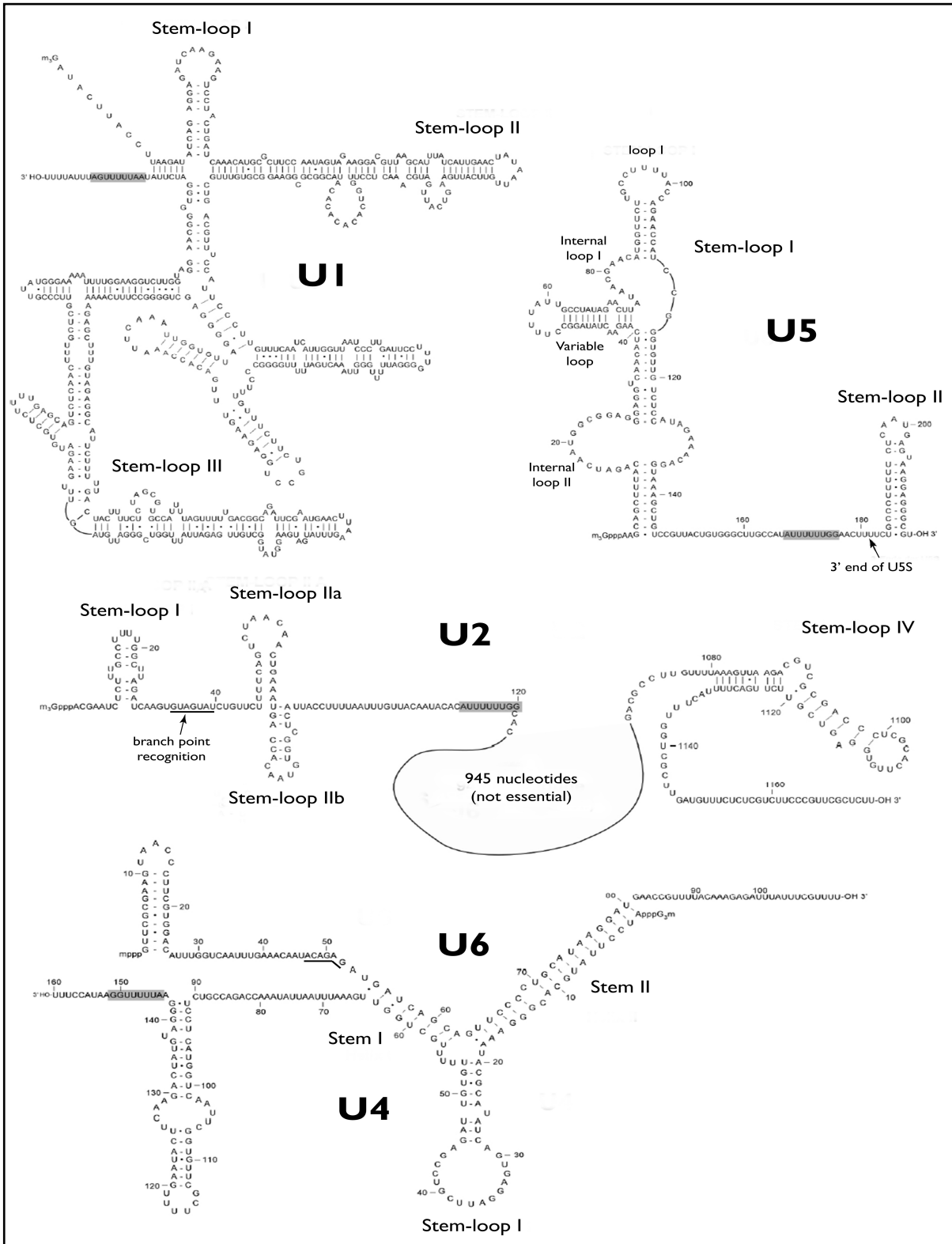
**Table: 1.1 Length comparison of different snRNAs of yeast and human**

The table shows the different types of snRNAs in yeast (in blue) and human (in green) with their respective length (ns). Two isoforms of U5 snRNA; U5L and U5S, are present in yeast due to 3' end trimming of U5 snRNA. ns - nucleotides

The snRNAs contain small sections of primary sequence, which are highly conserved (close to 100%) between different organisms (Frank et al., 1994). These small conserved sequences (fig: 1.4, and fig: 1.5) mostly reside on the single stranded regions that are in vicinity to the crucial

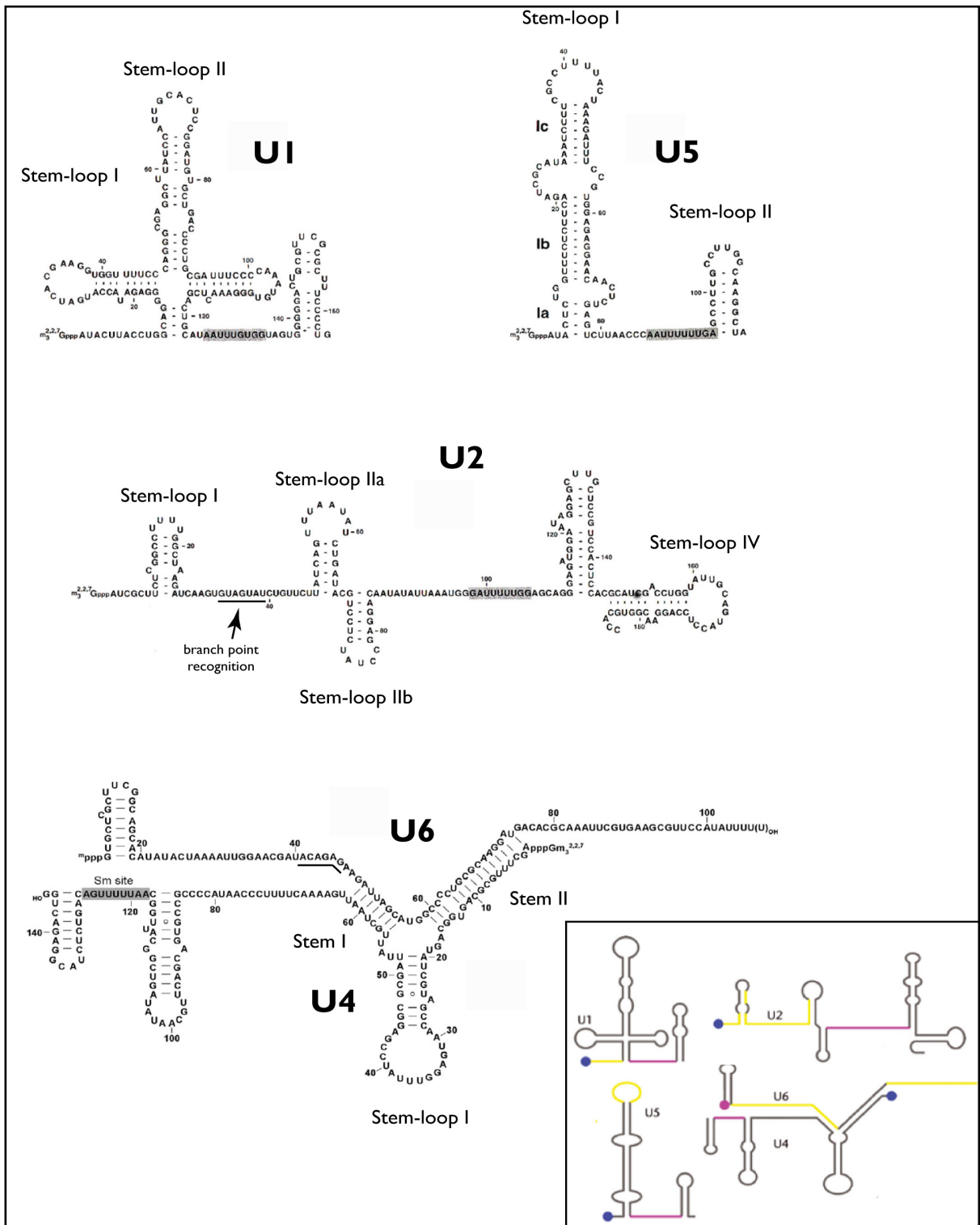
secondary structures. These little stretches play an important role in protein binding, snRNA and pre-mRNA base-pairing interactions, or in snRNA-snRNA interactions. Similarly, the secondary structures of the snRNAs, like loops and stems, are also highly conserved among different organisms (Patterson and Guthrie, 1991).

The U1, U2, U4, and U5 snRNAs are bound by the heptameric rings composed of seven Sm proteins (B, C, D1, D2, D3, E, and F) at the uridine-rich conserved sequence (Sm binding site) in the cytoplasm (Kambach et al., 1999);(Vidal et al., 1999). The U6 snRNA, which does not share the Sm binding consensus sequence, acquires a similar type of heptameric ring (seven protein ring) of Sm-like proteins ("Like-Sm" or, LSM 2-8) at its 3'-end (Achsel et al., 1999). The Sm ring at the Sm core RNP and the m<sub>3</sub>G cap of the snRNAs U1, U2, 4 and U5 serve as nuclear localisation signals for the nuclear import (Fischer et al., 1993). After entering the nucleus, the pre-snRNAs undergo extensive modifications via 2'-O-methylation and pseudouridylation in the Cajal bodies (Karijolich and Yu, 2010). The assembly of the U6 snRNP instead is thought to take place entirely within the nucleus.



**Figure 1.4 Yeast snRNA sequences with their secondary structures.**

The snRNAs harbour conserved secondary structures with single stranded regions and stem-loops. The Sm-sites for Sm protein binding are greyed out. The extensively base-paired U4/U6 snRNAs are shown at the bottom. Modification from (Kretzner et al., 1990) (U1); (Shuster and Guthrie, 1988) (U2); (Frank et al., 1994) (U5); (Brow and Guthrie, 1988) (U4/U6).



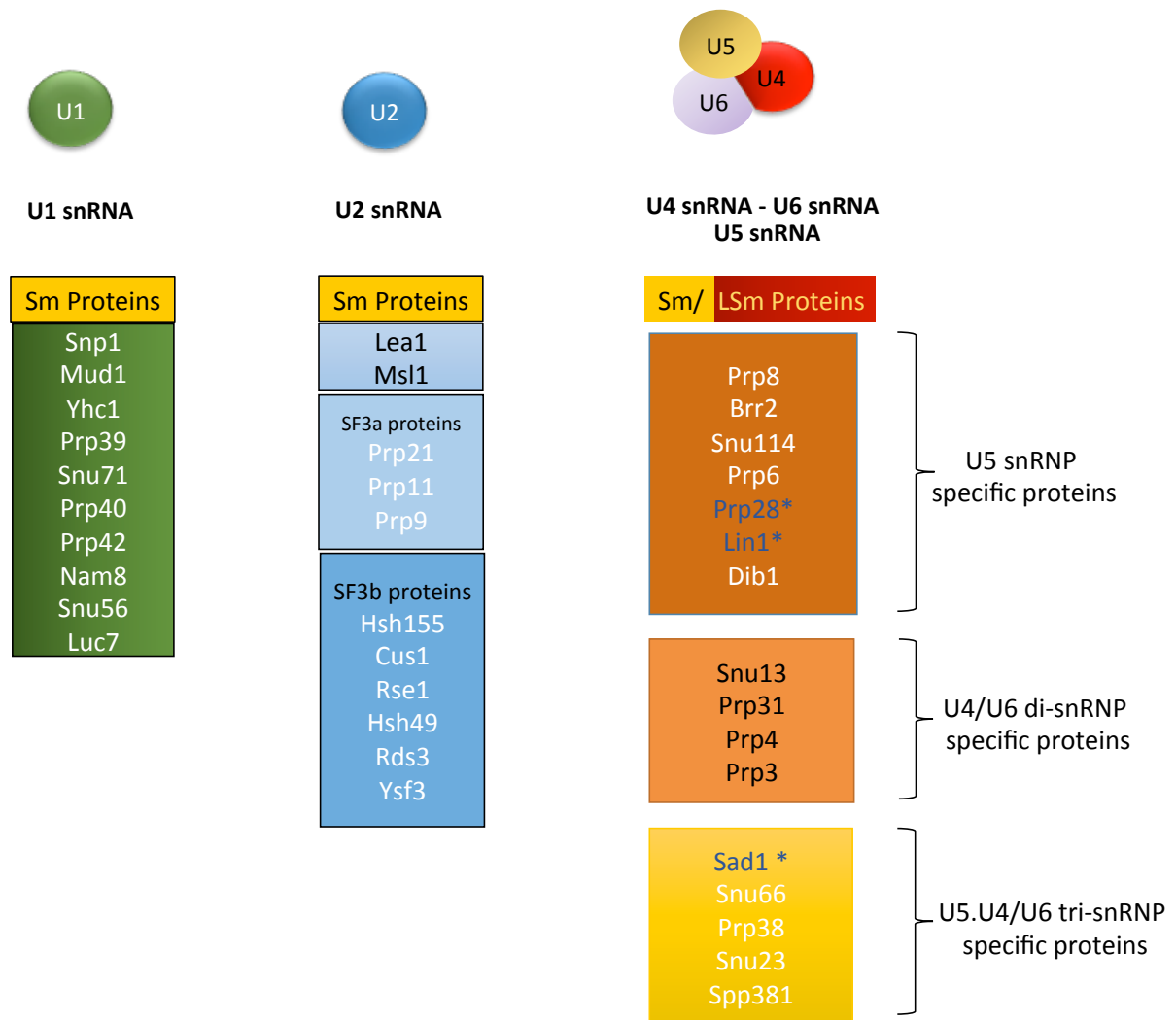
**Figure 1.5 HUMAN snRNA sequences with their secondary structures.**

The snRNAs harbour conserved secondary structures with single stranded regions and stem-loops. The Sm-sites for Sm protein binding are greyed out (U2); (Krol et al., 1981) (U5); (Brow and Guthrie, 1988) (U4/U6). **Inset:** conserved sequences participating in RNA-RNA interactions are depicted in yellow; the Sm-sites are shown in pink. The pink bold dots shows the  $\gamma$ -monome-thylguanosin cap on the U6 snRNA, the blue bold dots indicate the 2, 2, 7-trimethylguanosin cap. Modification from (Padgett, 2005).

### 1.5.2 The protein inventory of U snRNPs

As outlined in the section above, apart from a specific U snRNA, each snRNP comprises Sm/ or LSM proteins, forming the snRNP Sm core. Additionally, each individual snRNP also associates with several other proteins, which are specific for a particular snRNP (fig: 1.6). Together, all these proteins and snRNAs constitute a sophisticated structure, which provides a platform for RNA-RNA and RNA-protein interactions taking place within the snRNP or with other key players of splicing. The particle-specific and the common proteins, like their U snRNA compounds, are evolutionarily largely conserved between all eukaryotes.

In yeast, the U1 snRNP is comprised of 10 particle-specific and 7 Sm proteins (Gottschalk et al., 2001) (fig: 1.6). Likewise, the U2 snRNP contains of 18 distinct proteins, which include seven common proteins (Sm proteins) and 11 proteins that are specific to the U2 snRNP. These are Lea1, Msl1, the SF3a complex comprising Prp9, Prp11, and Prp21, and the SF3b complex comprising Rds3, Ysf3, Rse1, Cus1, Hsh49, and Hsh155 (Dziembowski et al., 2004); (Behrens et al., 1993). The U5 snRNP in yeast exists in two diverse forms: the 18S U5 snRNP and the 16S U5 snRNP. The 18S form, usually integrated into yeast tri-snRNP, contains several particle-specific proteins (Brr2, Prp8, Snu114, Prp28 (although the latter always in sub-stoichiometric amounts), Dib1, and Lin1) along with common Sm proteins. The 16S instead was co-purified along with U1 snRNPs (Gottschalk et al., 2001) consisting of only three particle-specific factors (Prp8, Snu114, and Aar2). A study in 2007 pointed out that the cytoplasmic Aar2-pre-U5 snRNP depends on a nuclear localization signal in Prp8 for nuclear import. In the nucleus it then completes its maturation process by acquiring Brr2 (Boon et al., 2007). In human as well as in yeast, the U4 and U6 snRNAs are firmly base-paired. The U4/U6 snRNAs, along with their respective common and particle-specific proteins, together form the U4/U6 di-snRNP, which then interacts with the U5 snRNP to give rise to U4/U6.U5 tri-snRNP. Further details on tri-snRNP particles are highlighted in section 1.6



**Fig: 1.6 protein composition of yeast spliceosomal snRNPs.**

The diagram shows the protein composition of yeast spliceosomal snRNPs. The list of snRNP specific proteins is shown. Common Sm proteins are indicated in yellow, while the Lsm are shown in maroon red. U1 snRNP proteins shown in green, U2 snRNP with shades of blue (SF3a and SF3b). The tri-snRNP proteins shown here: U5 snRNP proteins golden, U4/U6 di-snRNP proteins in gradient of purple and red, and U4/U6. U5 tri-snRNP specific proteins in triple gradient box. (Adaptation from Gottschalk *et al.*, 1999; Fabrizio *et al.*, 2009; Will and Lührmann, 2011) Asterisk- Peptides not detected in MS analyses in tri-snRNP samples in previous studies.

### 1.5.3 The non-snRNP factors of the splicing machinery

Section 1.4 outlines that additional non-snRNP factors are important to restructure the spliceosome in a highly dynamic way. These additional factors either join the spliceosome in the form of pre-assembled sub-complexes or they are recruited individually to the splicing apparatus. The yeast nineteen complex (NTC) and the retention and splicing complex (RES) are two prominent examples of pre-assembled complexes of non-snRNP spliceosomal proteins in yeast. In yeast, the NTC exists as an independent sub-complex that can be purified from yeast whole

cell extract. It consists of eight different proteins, Prp19, Cef1, Syf1, Syf2, Syf3, Snt309, Isy1, and Ntc20. Its human homolog is the CDC5/Prp19 complex, which, however, shares only three out of eight proteins with the yeast NTC (Golas et al., 2010). The RES complex is a group of three proteins (Bud13, Pml1, and Ist3) (Dziembowski et al., 2004). (Fig: 1.9)

Other important spliceosomal factors that need special emphasis are DExD/H box proteins. These proteins belong to superfamily 2 of RNA/DNA helicases (Staley and Guthrie, 1998); (Bleichert and Baserga, 2007). Although named RNA helicases, several spliceosomal ATPases act preferentially as RNPsases, that is they displace proteins from their RNA binding sites. These factors require energy (from ATP hydrolysis) to perform their tasks and they apparently exert the crucial mechanical force on the spliceosome during the course of splicing. There are eight conserved DExD/H box proteins (Sub2, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22, and Prp43) involved in the splicing process. Only Brr2 and Prp28 join the spliceosome as snRNP-associated proteins.

The proteins of this family contain two RecA domains, and all members share many conserved sequences within the same conserved motif. Based on the conserved sequence motif in the RecA domains, these proteins can be further categorized into three groups (table: 1.2). The first group containing identical DEAD sequences in the RecA domain motif II consist of Sub2, Prp5, and Prp28. The second group contains Prp2, Prp16, Prp22, and Prp43 with similar DEAH sequence in RecA motif II. Lastly, Brr2 is the sole member of the third group with a DEIH sequence at the RecA motif II. (Pena et al., 2009) (Reviewed by Liu and Cheng, 2015)



<b>DExD/H-box protein</b>	<b>Motif II</b>	<b>Stage in splicing cycle</b>
Sub2	DEAD	Pre-spliceosome
Prp5		Pre-spliceosome
Prp28		Early step I activation
Brr2	DEIH	Early step I activation
Prp2	DEAH	catalytic activation for step I
Prp16		Step II activation
Prp22		mRNA release
Prp43		Disassembly of ILS

**Table 1.2 DExD/H box proteins of *S. cerevisiae***

Table showing the three different groups of DExD/H box proteins categorized based on the RecA domains Motif II conserved sequences. Each group contains proteins with similar motif II. Each protein's function in the spliceosomal cycle is indicated in column 3

## 1.6 The U4/U6.U5 tri-snRNP

### 1.6.1 A large building block of the spliceosome

The pre-assembled tri-snRNP is formed by the association of the U5 snRNP with a U4/U6 di-snRNP particle. It is a particularly protein-rich particle. Tri-snRNP contains two sets of Sm rings which are found associated at the Sm binding sequence at the 3' ends of U4 and U5 snRNAs. Additionally, the tri-snRNP contains a single set of seven Lsm proteins (Sm-like proteins 2-8) (Achsel et al., 1999; Vidal et al., 1999) (fig: 1.6). Each snRNP in the tri-snRNP particle contributes its RNP specific factors. These factors are highly conserved in human as well as in yeast. In yeast tri-snRNP, the U5 snRNA is associated with Prp8, Brr2, Snu114, Prp6, Prp28, Lin1, and Dib1. These proteins establish the structure of the particle and are involved in its structural rearrangements. (1.6.2; 1.6.3). Prp3, Prp4, Prp31, and Snu13 are associated with the U4/U6 di-snRNP. Yeast proteins Prp38, Snu66, and Snu23 join the tri-snRNP as so-called tri-snRNP specific proteins; they are not part of either the U5 or the U4/U6 snRNP. Published data has shown that in yeast Prp6 is a crucial factor for tri-snRNP particle integrity, as depletion of Prp6 from yeast whole cell extract results in a low yield of tri-snRNP, while the U4/U6 snRNP level remains the same (Galissou and Legrain, 1993). Likewise, in human Prp31 is important for tri-snRNP formation (Makarova et al., 2002). Later studies strongly indicated that Prp6p and

Prp31p function as connecting proteins in the tri-snRNP, joining the U5 snRNP with the U4/U6 snRNP (Liu et al., 2006; Makarova et al., 2002; Schaffert et al., 2004).

The human tri-snRNP contains some additional proteins, like the peptidyl-prolyl isomerase PPIH, aU4/U6 protein and the 40K U5 protein, as well as the tri-snRNP proteins RBM42 and 27K, which don't have a yeast ortholog. The human U5-100K, a DEAD-box RNA helicase, is the homolog of the yeast Prp28p. This factor has been shown to be involved in switching the U1 snRNA at the 5' SS against U6 snRNA in the course of spliceosome activation (Chen et al., 2001; Staley and Guthrie, 1999). Notably, Prp28 could not be identified by mass spectrometry in purified yeast tri-snRNPs so far (Gottschalk et al., 1999; Stevens and Abelson, 1999; Stevens et al., 2001), however, one study has shown that the yeast tri-snRNPs can be isolated via TAP-tagged Prp28p (Small et al., 2006), indicating that at least a minor sub-population of particles contains Prp28p.

Another important factor, which is notoriously elusive and so far, has never been shown to be purified along with yeast tri-snRNP particles, is the yeast Sad1 protein. In contrast, its human homolog, a 65 kDa protein, has been shown to be a stably integrated stoichiometric component. (Agafonov et al., 2016b); (Charenton et al., 2019). In a human splicing system Sad1 has been shown to be important for the recruitment of tri-snRNP particles into spliceosomal B complexes (Makarova *et al.*, 2001). On the other hand, in yeast splicing extract, Sad1 has been shown to be important in maintaining the tri-snRNP in an intact form (Huang *et al.*, 2014).

Before the first high-resolution structures of tri-snRNPs became available (Nguyen et al., 2015); (Yan et al., 2015); (Nguyen et al., 2016) speculations on the internal structure of the yeast tri-snRNP were based on experimentally observed protein-protein interactions (table: 1.3). Already this table, which summarizes several decades of research on protein-protein interactions in the yeast spliceosome, gives a hint how snRNP and non-snRNP proteins are organized as subgroups in the tri-snRNP and which proteins are possibly the connectors between the subdomains. If we take only proven direct contacts (for example in co-crystals) and yeast-2-hybrid data into consideration as evidence for one-to-one protein-protein interactions, the data reveals the high degree of these contacts between almost all U5 proteins (except Prp6 and Dib1 which interact with each other). In the U4/U6 snRNP Prp3 and Prp4 interact tightly, while the tri-snRNP specific proteins fall into two classes, the tightly interacting Snu66 and Sad1, and the trimeric complex of Spp31, Prp38 and Snu23. It is also remarkable that Snu66 and Sad1 proteins appear to be in close direct contacts with the U5 proteins. Now that 3D structures became available, many of these observations can be visualized in full detail. But also, the limitations for example of Y2H approaches became obvious. Many interactions we now have 3D-evidence for were never

visualized with Y2H experiments. This is for example due to the fact that several proteins, when used as isolated entities, never assume the proper conformation required for specific interactions with other partners. These conformations they would only assume once assembled into larger multi-protein complexes.

		U5							U4/U6				tri spec.				
		Prp8	Brr2	Snu114	Prp6	Prp28	Lin1	Dib1	Prp31	Prp3	Prp4	Snu13	Snu66	Sad1	Spp381	Prp38	Snu23
U5	Prp8		SL, SR(2), Y2H(2), dir.interact(5), SGD	SGD (4), Y2H		PS, dir.interaction, SR(3)	Y2H						SGD, SL			DL	
	Brr2			SL, dir.interact			SGD						SGD, SL, Y2H (2)				
	Snu114				NG (2)	SL, NG (2)	NG (3)	NG	NG (2)	NG (2)	NG (2)		NG (3), SL	Y2H, SL, NG (2)	NG	NG	NG (2)
	Prp6					NG (2)	NG (3)	Y2H (3)		NG (2)	NG (2)		NG (3), Y2H	NG (2)			NG
	Prp28						dir.interact	NG (3)		NG (2)		NG (2)	NG (3)	NG (2)	NG	NG (2)	
	Lin1								NG (3)	NG (3)	NG		NG (4)	NG (2)			NG (2)
	Dib1												NG	NG			
U4/U6	Prp31									NG (2)	NG (2)		NG (2)	NG (2)			
	Prp3										Y2H (2)	DR, Y2H (5), NG (2), SL	NG	NG (2)		NG	NG
	Prp4											NG	NG (4)	NG (2)	NG	NG	NG
	Snu13																
tri spec.	Snu66													Y2H, NG (2)	NG (3)	NG (2)	NG (2)
	Sad1															NG (2)	NG (2)
	Spp381															DR (2), NG (2), Y2H (2)	NG (3)
	Prp38																NG (2), Y2H
	Snu23																

**Table 1.3 Genetic interactions between pairs of yeast tri-snRNP proteins**

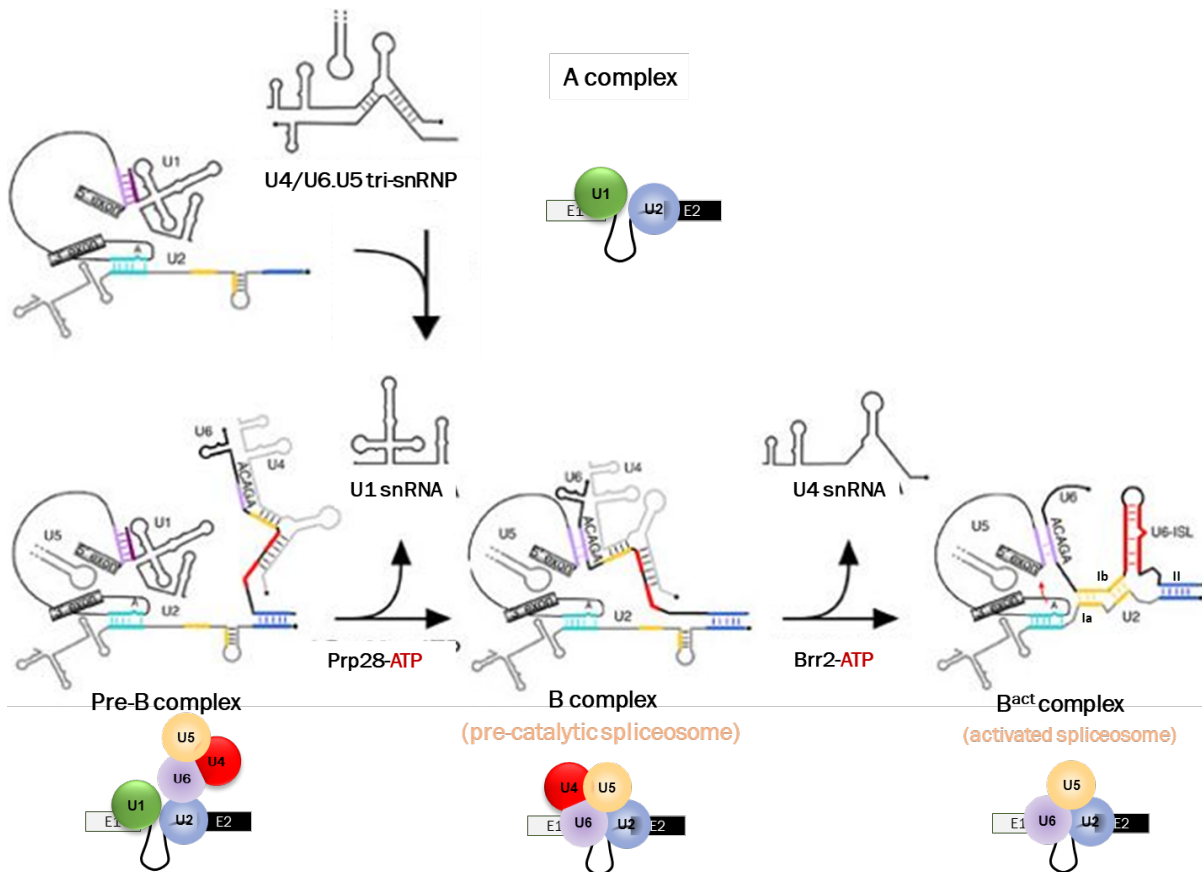
The canonical tri-snRNP proteins are listed on the margins of the grid. Yeast interactions denoted are SL synthetic lethal, SR synthetic rescue, Y2H yeast-2-hybrid, SGD synthetic growth defect, PS phenotypic suppression, direct interaction, DL dosage lethal, NG negative genetics, and DR dosage rescue. Numbers in brackets denote number of independent entries in the SGD database (as of February 2019). The internal U5 and U4/U6 snRNP interactions as well as the interactions between tri-snRNP-specific proteins are coloured.

### 1.6.2 The tri-snRNP undergoes large structural rearrangements upon spliceosome activation

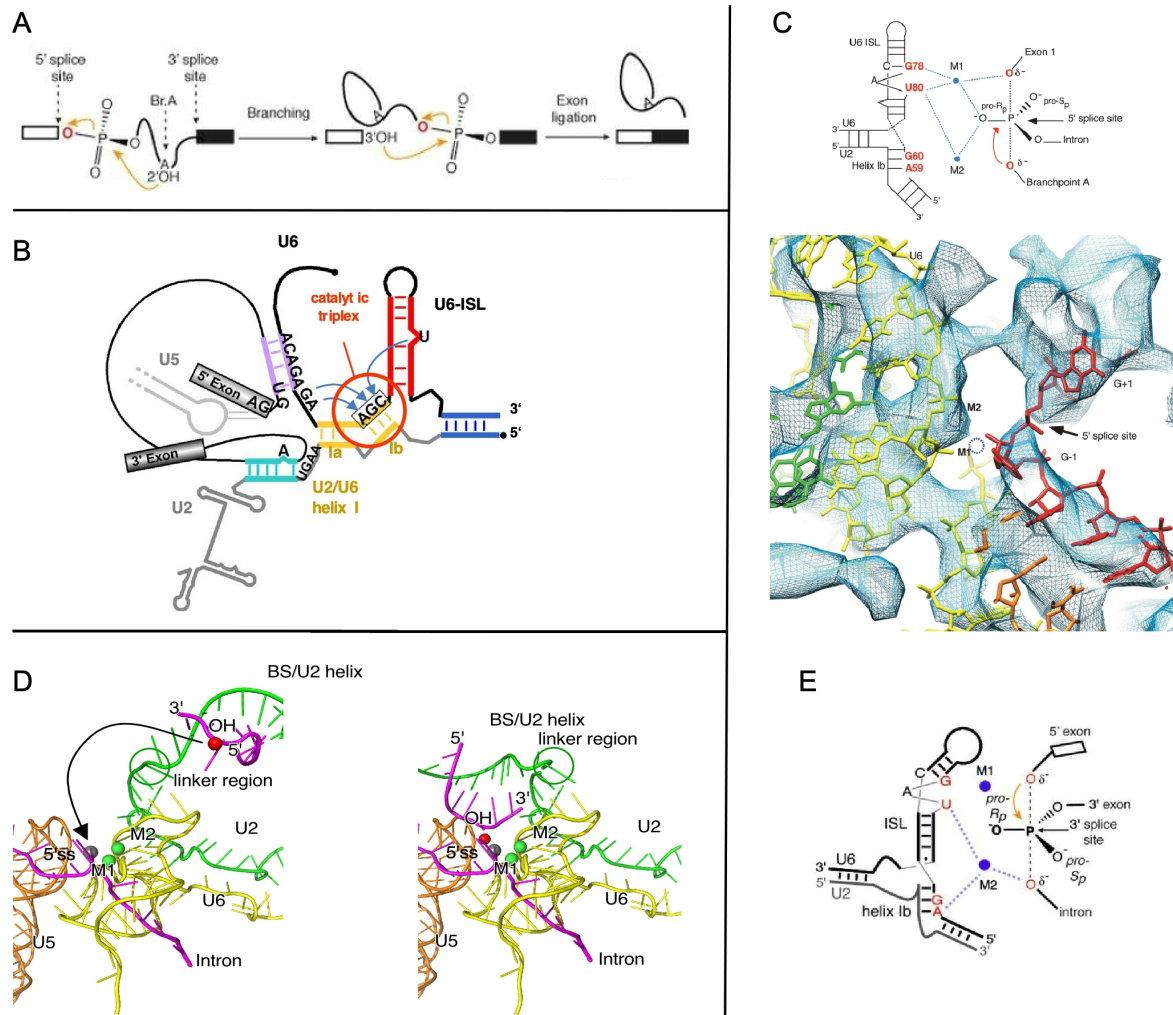
Significant changes in the network of RNA-RNA interactions during the process of spliceosome assembly contribute to the highly dynamic nature of the spliceosome. The activation of the spliceosome, that is, the transition of pre-B to B and then B<sup>act</sup> complex is a good example where many snRNA-snRNA and pre-mRNA-snRNA contacts are made and changed (Nilsen, 1998); (Will and Lührmann, 2011a). In the E complex the 5'SS, the BPS, and the 3'SS, are already arranged in close vicinity (Kent and MacMillan, 2002; Kent et al., 2005). During the catalytic activation the organized physical re-arrangements lead to the establishment of the U2/U6 and pre-mRNA network, in which, U2 and U6 snRNAs strongly contribute to the generation of the catalytic center of the spliceosome. As mentioned earlier in section 1.4, the tri-snRNP association with the A complex leads to the formation of the pre-B complex. The tri-snRNP is heavily involved in the catalytic activation of the spliceosome. The U4 and U6 snRNAs are extensively base-paired in the tri-snRNP which forms a 'Y' shaped interaction junction. This unique structural feature consists of U4/U6 stems I and II, separated by a U4 snRNA internal stem-loop (fig: 1.4, 1.5). After recruitment (in the pre-B complex), the 5' end of the U2 snRNA and the 3' end of the U6 snRNA of the U4/U6 di-snRNP form U2/U6 helix II. However, the stable integration of the tri-snRNP only occurs after action of Prp28 which leads to the displacement of the U1 snRNP from the 5'SS (Staley *et al.*, 1999) and thus permits the formation of new base pairing interactions between the highly conserved U6 ACAGAGA box sequences and the liberated 5'SS. The thus formed B complex contains the U6 snRNA base-paired with the U2 snRNA at U2 /U6 snRNA helix II, and with the 5'SS of pre-mRNA through the U6 ACAGA box. Extensive base-pairing also occurs with U4 (U4/U6 stem I and stem II) (fig: 1.4, 1.5) (Staley and Guthrie, 1999); (Boesler et al., 2015). Important constituents of the catalytic center are present in the U6 snRNA, which at this stage is still extensively base-paired with U4 snRNA. This inactive state is safeguarding against pre-mature catalysis. Only after this U4 – U6 interaction is dissolved by the action of Brr2 activation can occur (Brow and Guthrie, 1988); (Laggerbauer et al., 1998); (Raghunathan and Guthrie, 1998). Once the U4 snRNA is displaced, the U6 snRNA is now free to adopt novel conformations, whereby it forms an internal-stem-loop (U6-ISL) structure which is crucial for the catalytic step. Moreover, conserved sequences on U6 snRNA now become available to take part in new interactions with U2 snRNA (U2/U6 helix Ia, and Ib) (fig: 1.8). The newly formed U6 ISL structurally resembles domain V of self-splicing group II introns, and similarly coordinates catalytically important metal ions (Madhani and Guthrie, 1992); (Shukla and Padgett, 2002); (Anokhina et al., 2013). To perform step I catalysis,

the U2 and U6 snRNAs in the catalytically activated spliceosome (B\* complex) have assumed a complex 3D-structure (see Fig. 1.8) which is able to accommodate the splicing substrate and, in close proximity, the two metal ions which are required for catalysis. The most prominent RNA features are all located in U6 snRNA: the ACAGAGA box, the bulged-out uridine of the intramolecular stem-loop, ISL, of U6 snRNA and the so-called catalytic triad AGC of U6 in helix Ib. Together, these motifs form a catalytic triple helix (or catalytic-triplex), which coordinates the two magnesium ions at a specific distance, thereby, catalysing both steps of pre-mRNA splicing. The RNA structure of this catalytic center can only be formed and stabilized by a protein scaffold, including a central region of Prp8 complemented by proteins of the NTC. Spliceosomal proteins are instead not directly involved in catalysis.

Following further acquisition of new protein components and spliceosome remodelling after step I catalysis, the spliceosome uses the same catalytic center to catalyse step II of splicing. For more details see legend of Figs 1.7 and 1.8.



**Figure 1.7 RNA-RNA rearrangements during spliceosome activation.** Schematic diagram of the rearrangements of RNA-RNA exchanges during the course of spliceosome activation. Process starts with complex A in which the U1 snRNA is base-paired with the 5'-SS and the U2 snRNA with the BPS. Union of tri-snRNP with the A complex leads to the formation of the pre-B complex. During the now ensuing RNA restructuring the highly conserved ACAGAG sequence (highlighted in mauve) of the U6 snRNA substitutes U1 snRNA at the 5' SS. The U6 snRNA forms helices Ia/Ib (shown in yellow) and II with the U2 snRNA, and folds back on itself to form an intramolecular stem-loop (U6-ISL) (shown in red). Thus, the splice sites and the BP are positioned in close vicinity for the first and second transesterification reaction (complex C). U5 snRNA associates with nucleotides of the exon. Exons are represented as grey boxes, while introns are shown as a black line. (Adapted from Boesler et al., 2016)

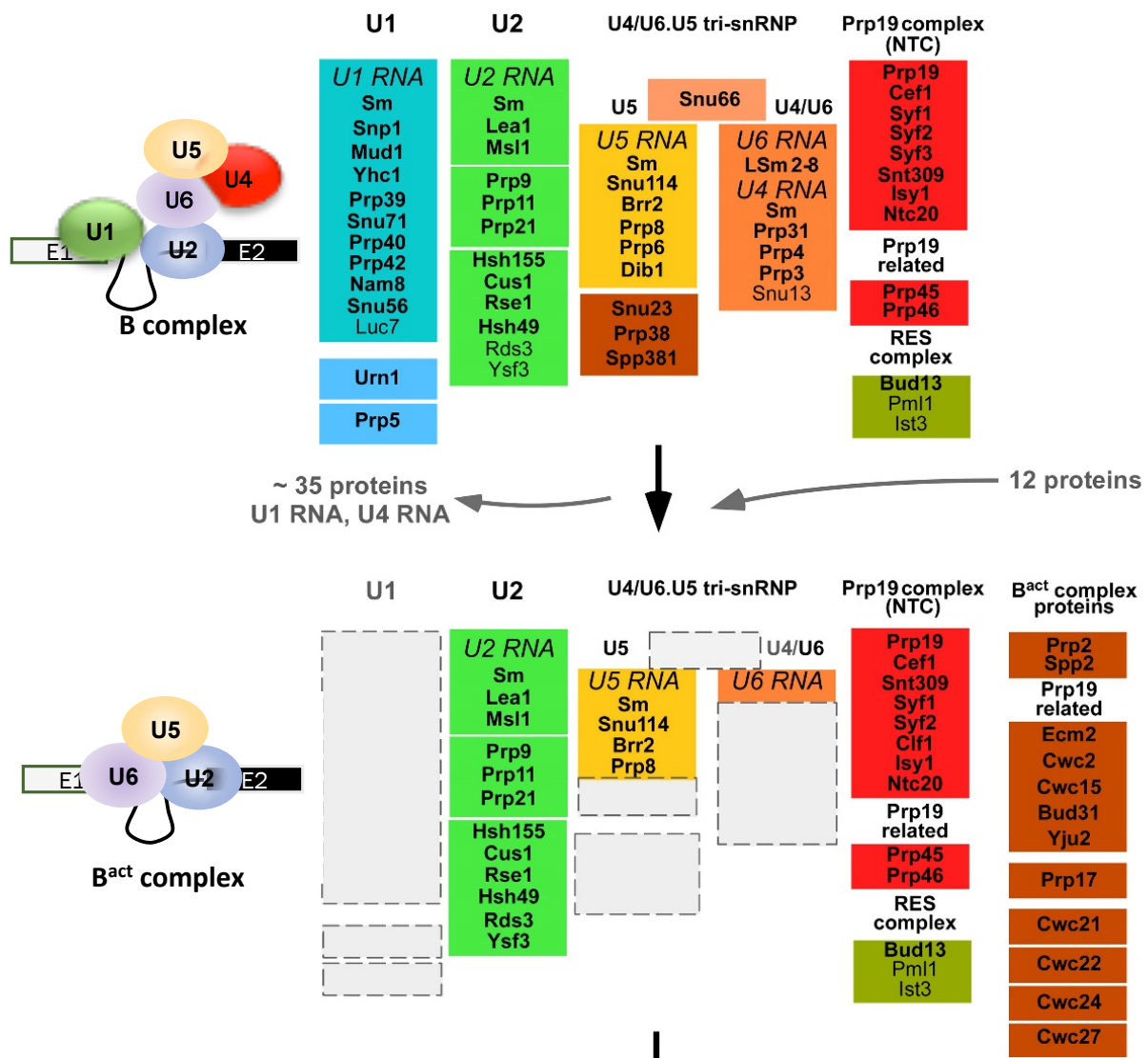


**Fig: 1.8** Panel **A** is a schematic drawing of the two-step reaction of splicing. Panel **B** presents an overview of the RNA network present in the  $B^{act}$  complex. Three conserved motifs to support the metal-mediated catalysis of splicing are indicated: the ACAGAGA box of U6 snRNA, the bulged-out uridine in the intramolecular stem-loop, ISL, of U6 snRNA and the so-called catalytic triad AGC of U6 in helix lb. Panel **C**, in the upper part shows the molecular geometry of the U2 and U6 snRNAs and the splice substrate in the active center during step I catalysis. The scissile phosphate of the 5'-splice site is shown in the transition state. ISL is the intramolecular stem-loop of U6 snRNA. Oxygens directly involved in the reaction are shown in red, their coordination in blue. The lower part of the panel shows the 3D-model of the 5'-splice site in the  $B^{act}$  complex just prior to the step I reaction (yellow U6, green U2, orange U5 and red pre-mRNA substrate). The putative position of the catalytic metal ions is indicated by circles. Panel **D** depicts a model how, during Prp2-mediated catalytic activation of  $B^{act}$ , the 5' splice site, which before activation is well separated from the catalytic center, moves into the catalytic center of the  $B^*$  complex. (The -OH group of the BS adenosine is shown in red, the 5'-splice site in grey, catalytic  $Mg^{2+}$  ions in green.) Panel **E** shows a model how the catalytic center accommodates the lariat-3'-exon pre-mRNA during step II catalysis. The scissile phosphate of the 5'-splice site is shown in the transition state. Oxygens directly involved in the reaction are shown in red, their coordination in blue. (Panel A and Dare from Fica et al. Nature 2013, 503:229-234; Panels B and C are from Rauhut et al. Science 2016, 353:1399-1405 and were kindly provided by Dr. B. Kastner)

The U5 snRNA, with the help of its loop I sequence, interacts with pre-mRNA exons. Together with the U5 Prp8 protein, it has been shown to be involved in proper positioning of the exons for the first and second step of splicing. (Grainger and Beggs, 2005). Interestingly, the U5 loop is not needed for the first step of splicing in yeast (O'Keefe and Newman, 1998), and in human for both steps of splicing *in vitro* (Segault et al., 1999), suggesting that loop I function may be compensated for by Prp8 in anchoring exons during the splicing reactions.

Physical rearrangements occurring at the time of spliceosome activation are not only limited to RNA-RNA interactions. Simultaneously, dramatic exchanges are happening also at the protein level (as described in section 1.6.3). During the B to B<sup>act</sup> transition (activation), the spliceosome loses the U1 proteins, the U4/U6 proteins, and the tri-snRNP-specific proteins (fig: 1.9). Of the original set of tri-snRNP proteins only the U5 specific proteins Prp8, Brr2, and Snu114 remain in the activated spliceosome (Fabrizio et al., 2009). Simultaneously NTC and NTC-related proteins become stably integrated (Chen et al., 2002; Makarova et al., 2004).



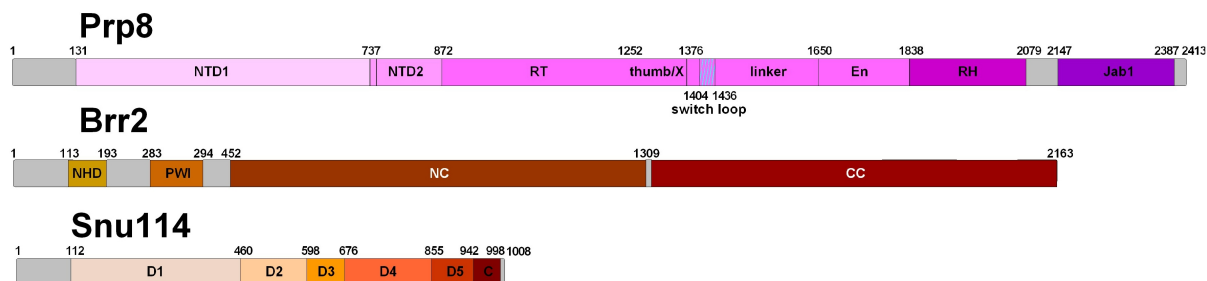


**Fig: 1.9 Exchange of key factors during spliceosomal activation**  
(Adapted from (Häcker et al., 2008); (Fabrizio et al., 2009)).

### 1.6.3 The tri-snRNP proteins play a key role in the catalytic activation of the spliceosome and spliceosomal dynamics

The U5-specific proteins Brr2, Snu114, and Prp8 remain bound to the spliceosome throughout the course of the splicing cycle (Ohi et al., 2002; Stevens et al., 2002). Fig. 1.10 shows the domain structures of the three largest tri-snRNP proteins. The huge Prp8 protein (279.5 KDa) has a complex domain structure. How Prp8 became the main organizer of the tri-snRNP is still not well understood. The U5 snRNP protein Prp8 plays a vital part in monitoring the activity of Brr2 in the spliceosome. Prp8 contains a Jab1/MPN domain in its C-terminal region that on the one hand activates the Brr2-dependent dissociation of U4/U6 snRNAs (Maeder et al., 2009). In addition, the C-terminal tail region of Prp8's Jab1/MPN domain has been shown to inhibit the

U4/U6 unwinding, ATPase, and RNA –binding activities of Brr2 by occupying the RNA-binding site of the active RNA helicase cassette of Brr2 (Mozaffari-Jovin et al., 2013).



**Fig. 1.10 Domain structure of tri-snRNP proteins Prp8, Brr2 and Snu114.** **Prp8:** NTD1 and NTD2, N-terminal domains 1 and 2; RT reverse-transcriptase-like; thumb/X linker region; En, endonuclease-like; RH, RNase H-like; Jab1, Jab1/MPN-like. **Brr2:** NHD, N-terminal helical domain; PWI, N-terminal, non-canonical PWI domain; NC/CC, N-terminal/C-terminal helicase cassette. **Snu114:** D1-D5 homologous to EF-G/EF-2. Figure from (Rauhut et al. Science 2016, 353:1399-1405) kindly provided by B. Kastner.

Although Brr2 (246.2 KDa) shows two helicase cassettes, only the N-terminal one is functionally active. Brr2 is a central component of the U5 snRNP and it is necessary for spliceosome activation. In in-vitro studies, Brr2 has been shown to be involved in the U4/U6 snRNA duplex unwinding (Laggerbauer et al., 1998) thus facilitating the release of U4 snRNA. It has been shown to translocate along U4 in the 3' to 5' direction after loading onto the single-stranded region of the U4 snRNA close to the U4/U6 duplex (Mozaffari-Jovin et al., 2013); (Cordin et al., 2012). As Brr2 is co-residing together with its U4/U6 snRNA substrate, it must be closely regulated during the splicing cycle (reviewed by Liu and Cheng, 2015).

Snu114 (114 KDa), despite of its similarity to the EF-G/EF-2 protein, is still not well understood concerning its function. Biochemical and genetic studies suggest a possible role of Snu114 in activation of the spliceosome, however, in the light of recent cryo-EM structures of the spliceosome, its exact role, if any, remain elusive.

As a general feature in all three proteins, considering their evolutionary origins, most domains are RNA-interacting domains, although many have lost their original functionality.

## 1.7 Structural analysis of spliceosomal complexes by cryo-electron microscopy

Structural studies of very large molecular machines in the megadalton range have been a challenge for structural biologists for many years. Due to the complex catalysis these machines have to perform they are not only large but they are also dynamic in their structural details. Both aspects, size and dynamic behaviour, are usually detrimental to classical approaches of structure analysis such as x-ray crystallography. This situation changed dramatically with the advent of single-particle electron microscopy (EM) a few years ago.

Crystallography requires highly homogeneous protein samples in sufficient amounts. The sample needs to crystallize providing diffracting crystals. Reflecting on the typical characteristics of a spliceosomal sample, which usually contains a distinct purified spliceosomal complex of the splice cycle with often significant contaminations of other spliceosomal complexes makes clear why early attempts on the X-ray crystallography of such huge dynamic RNA-protein complexes faced insurmountable difficulties. Dynamic and intricate assemblies such as spliceosomes, which are undergoing rapid conformational and compositional changes were ideally suited for this new technique of structure elucidation. In contrast to traditional techniques, EM poses far less restrictions on the size of the particle, its homogeneity, and the amounts required (usually picomolar quantities of complexes are sufficient).

In order to achieve a sufficient level of structure resolution for a biological sample, a step-by-step processing of image processing and classification is performed (Thuman-Commike, 2001; van Heel and Frank, 1981b). The process entails selection of tens of thousands of single particles from the electron micrographs. The process of particle selection is sometimes also referred to as single particle picking which is either done manually or by using automatic computerized programs. After image classification, based on the common features, particles are categorized and sorted into classes (class-average), which contain particles sharing similar features. Compiled data from representative class-averages produces refined images of the sample particles exposing detailed features with improved signal-to-noise ratio (SNR).

The purifications of tri-snRNPs from human and yeast were published 1991 and 1994, respectively (Behrens / Fabrizio). Several complexes were reconstructed at medium/low resolution, which provided initial insight into the overall organization of the spliceosomal snRNAs and B- and step I- spliceosomes (Boehringer et al., 2004; Deckert et al., 2006): the human U1 snRNP, human U2 snRNP SF3b sub-complex (Kastner *et al.*, 1992; Golas *et al.*, 2003), the human U11/U12 snRNP (Golas *et al.*, 2005), the yeast tri-snRNP (Sander *et al.*, 2006)

and yeast ILS (Ohi *et al.*, 2007). Methods being used for these investigations paved the road for the latest achievements in high-resolution cryo-EM. Until a few years ago, the dynamics and complexity of the spliceosome were still a major challenge for EM, however, after rapid progress and advancements in the field of EM detectors, the 3D cryo-EM technique in general (as reviewed by Kühlbrandt)(Kuhlbrandt, 2014), and in particular in the computational processing of the raw data, many spliceosome structures have been resolved at near atomic resolution. During the combined efforts of the Lührmann and Stark groups, and of the Nagai, Shi and Zhao/ Zhou groups now numerous structures of nearly all known spliceosomal intermediate complexes in yeast and in human have been resolved (Kastner *et al.*, 2019; Plaschka *et al.*, 2019; Shi *et al.*, 2019; Wilkinson *et al.*, 2018).

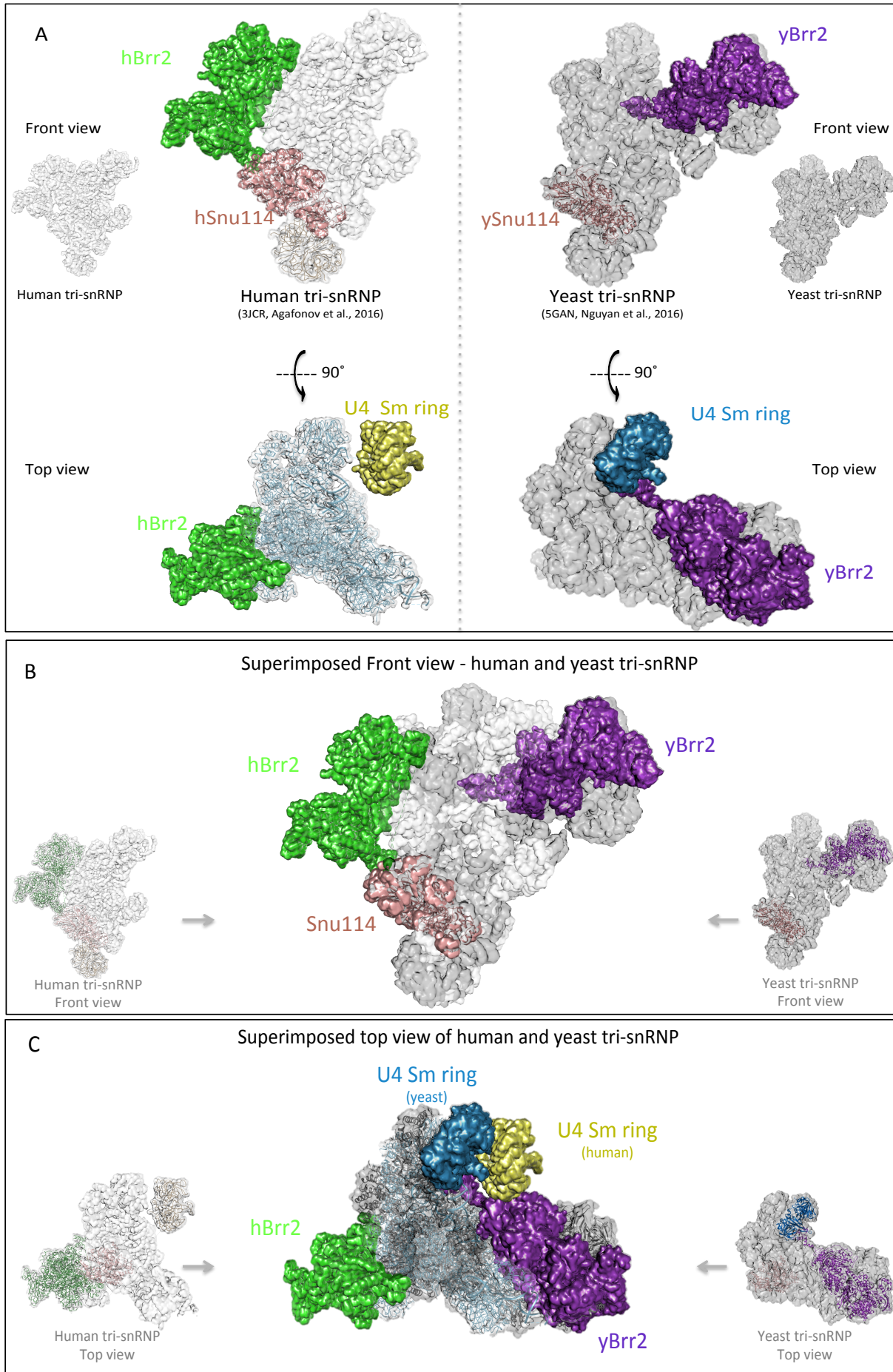
### 1.7.1 Structural differences in human and yeast tri-snRNP particles.

After decades of work on the biochemistry and genetics of tri-snRNP proteins and snRNAs a highly detailed picture of protein-protein, protein-RNA and RNA-RNA interactions in the tri-snRNP was achieved. This picture included also many functional aspects for some proteins. Even a rough picture of spatial distribution of proteins in the yeast tri-snRNP was elaborated (Häcker, Rigo, Chen-Fu). Only with the advent of high-resolution cryo-electronmicroscopy enabled us to finally see the 3D arrangement of all the main proteins. The recently reported structures of the yeast (Nguyen *et al.*, 2015 5.9Å and with better resolution in Nguyen *et al.*, 2016 3.7Å 3.8 Å; Wan *et al.* 2016) and human tri-snRNP (Agafonov *et al.*, 2016; and most recently in Charenton *et al.* 2019 2.9 Å) provided many important details about the spatial organization of these particles. In the yeast system, the 1.5 megadalton (MDa) tri-snRNP structure was resolved, using cryo-EM single particle reconstruction, at 5.9 Å resolution. The study used published X-ray or homology models of various known tri-snRNP constituent factors to fit into the cryo-EM density map to generate a final refined model. In case of the human tri-snRNP, the 1.8 MDa particle was resolved at an overall resolution of 7 Å using single particle cryo-EM. In addition to fitting available crystal structures and homology models into the generated cryo-EM density map, structural mass spectrometry was also used to validate the final refined model.

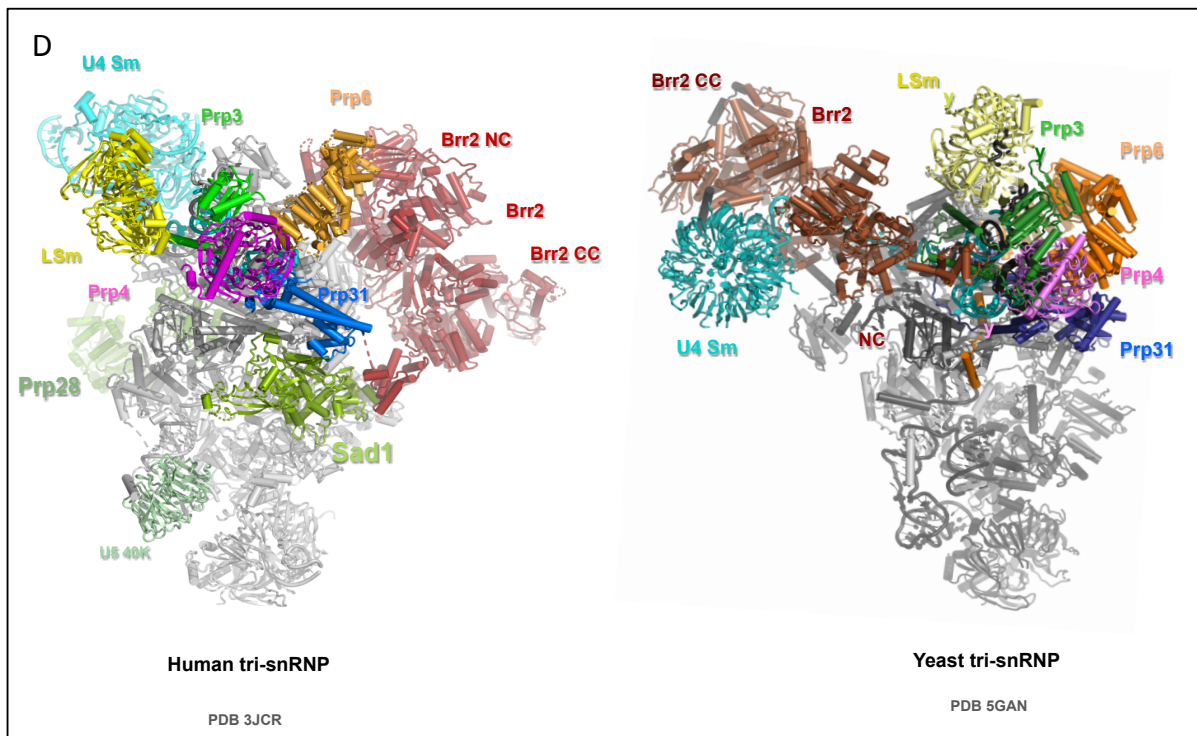
Comparison of the human and yeast tri-snRNP structures revealed several similarities but also substantial differences between the two particles (fig: 1.11).

### 1.7.2 The position of several proteins in human and yeast tri-snRNPs is dramatically different

In human tri-snRNPs, Brr2 is located in the particle at a very different position compared to the reported location in the yeast tri-snRNP structures (fig: 1.11) In the human tri-snRNP model the 245 kDa human Brr2 is positioned close to the tetratricopeptide-repeats (TPR) of Prp6 and Prp8's reverse transcriptase-like (RT) domain (Agafonov *et al.*, 2016; Charenton *et al.* 2019). In the yeast tri-snRNP model (Nguyen *et al.*, 2015 5.9Å and with better resolution in Nguyen *et al.*, 2016 3.7Å 3.8 Å; Wan *et al.* 2016) yBrr2 is located ~20 nm away from the location of the hBrr2 and about 180° rotated along the long axis of the human tri-snRNP model . Strikingly, in the yeast tri-snRNP model the U4-Sm core is found to be located at the space between the single-stranded region of U4 snRNA and the helicase cassette region of Brr2. This single-stranded region of U4 snRNA, which has been suggested to provide the docking site for Brr2, which then unwinds the U4/U6 duplex (Mozaffari-Jovin *et al.*, 2012), is located close to the Rec A domain of the active N-terminal cassette (NC) of yeast Brr2. In the human tri-snRNP structure instead, the active N-terminal cassette of Brr2 is situated about 8-10 nm away from the U4/U6 snRNA substrate (fig: 1.11 D) (Agafonov *et al.*, 2016).







**Fig: 1.11 Comparison of the overall 3D cryo-EM structures of yeast and human tri-snRNPs.** To highlight the differences in protein locations in the published tri-snRNP structures, the yeast and human tri-snRNP were first aligned side by side in corresponding views (Panel A) and then in superimposed views (Panel B and C). Snu114 served as a reference point for alignment. Superimposed front view as well as superimposed top view clearly show the dramatic difference of the Brr2 location in the two complexes. Panel D demonstrates that also proteins other than Brr2, like Prp31, Prp4, Prp6 and Prp3, take a different position in both complexes. See details in section 1.7.1 and 1.7.2

### 1.7.3 Structural arrangement of Sad1 in human tri-snRNP and its potential role

In purified human tri-snRNPs the evolutionarily conserved Sad1 protein is present in stoichiometric amounts (Behrens *et al.*, 1991). However in yeast, Sad1 is probably dissociated and lost during the purification of the complexes (Fabrizio *et al.*, 1994 and Stevens *et al.*, 2001).

In the human tri-snRNP structure, Sad1 is positioned at the interface between the U4/U6 and U5 snRNPs and potentially acts as a clamp which stabilizes the U4/U6 - U5 snRNP interaction by establishing contacts with the U4/U6 protein Prp31 and the U5 proteins Prp8, Snu114 and Brr2 (Agafonov *et al.*, 2016). The latter interaction with the Brr2 PWI domain might be crucial to keep Brr2 in a preactivation position, that is, in a safe distance to the U4/U6 duplex. Dissociation of Sad1 may in fact serve as a mechanism during human B complex activation, enabling the move of Brr2 into a conformation where it can interact with the U4/U6 duplex.

Also yeast Sad1 appears to be implicated in maintaining the tri-snRNP in an intact form (Huang *et al.*, 2014). The latter study showed that depletion of Sad1 protein from the yeast extract resulted in the complete dissociation of tri-snRNPs into U4/U6 di-snRNPs and U5 snRNPs upon incubation with 2 mM ATP. This further supports the potential role of Sad1 as a stabilizing factor for tri-snRNPs by tethering Brr2 in a pre-active conformation. Loss of Sad1 during purification of yeast tri-snRNPs may suffice to enable a move of Brr2 closer to the U4/U6 duplex, but without dissociating the tri-snRNP due to lack of ATP.

#### **1.7.4 Rational for the proposed work**

Although the tri-snRNP is an important, evolutionarily conserved component of the spliceosome and a key component in establishing the evolutionarily highly conserved network of RNA-RNA and RNA-protein interactions of the catalytic core of the spliceosome, the published yeast tri-snRNP structures revealed significant differences compared to the published structure of the human tri-snRNP. The most important major difference is the dramatically different position of Brr2. The reported structure of the yeast tri-snRNP showed a Brr2 RNA helicase already loaded onto its U4 snRNA substrate, and ready to unwind the U4/U6 duplex. In contrast, in the human tri-snRNP, Brr2 is positioned in a safe distance (~10 nm) to its U4 snRNA substrate. A major biochemical difference is the stability of human tri-snRNPs in the presence of ATP, while the yeast tri-snRNP undergoes Brr2-mediated dissociation under the same conditions. This is potentially due to differences in the protein composition of purified yeast and human tri-snRNPs. The yeast Sad1 protein has long been known to play an important role in tri-snRNP stability. While stoichiometrically present in purified human tri-snRNPs, it is usually absent in purified yeast tri-snRNPs. One may therefore speculate that it is the human ortholog of Sad1 which exerts a protective influence on human tri-snRNP stability by keeping Brr2 in a safe distance from the U4/U6 duplex.

I therefore set out to isolate and characterize yeast tri-snRNP particles which are similar to human tri-snRNPs in terms of protein composition and biochemical behaviour. Since a stable association of Sad1 and other evolutionarily conserved proteins may be the key to solving this problem, I will modify our established *in vitro* purification. Specifically, I modified existing tri-snRNPs purification protocols with the goal of obtaining particles that would contain Sad1 protein and would remain intact upon incubation with ATP. The yeast tri-snRNPs purified under the most gentle conditions were then subjected to protein cross-linking combined with mass spectrometry and to electron microscopic investigations to obtain insight into their 3D structure.



## Materials and methods

### 2.1 Materials

#### 2.1.1 Chemicals and media

Agarose	Invitrogen, Netherlands
Ammonium peroxodisulphate (APS)	Merck, Germany
Ampicillin	Sigma-Aldrich, Germany
Bacto agar	BD, USA
Bacto peptone	BD, USA
Bacto yeast extract	BD, USA
BisTris (Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane)	Sigma-Aldrich, Germany
Boric acid	Merck, Germany
Bovine Serum Albumin (BSA)	Sigma-Aldrich, Germany
Brilliant Blue G-Colloidal concentrate	Sigma-Aldrich, Germany
Bromophenol blue	Merck, Germany
BS3 (Bis[sulfosuccinimidyl]suberate)	Thermo Fisher Scientific,
Calmodulin affinity resin	Stratagene, USA
Complete supplement mixture (CSM)	Bio 101 Inc., USA
Coomassie brilliant blue G-250	Serva, Germany
D(+)-glucose monohydrate	Merck, 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
DOBA-powder (dextrose/galactose)	Bio 101 Inc., USA
EDTA (Disodium salt dihydrate)	Roth, Germany
Ethanol	Merck, Germany
Ethidium bromide	Roche, Germany
Ficoll <sup>®</sup> PM 400	Sigma-Aldrich, Germany
Fish sperm DNA (10mg/ml)	Roche, Germany
Formamide	Merck, Germany
Glutaraldehyde	Electron Microscopy

Glycerol	Merck, Germany
Glycine	Merck, Germany
Glycoblue	Ambion, USA
Glycogen	Roche, Germany
HEPES (N-2-Hydroxyethylpiperazin-N-2-ethanesulfonic acid)	Calbiochem, USA
Hydrochloric acid (HCl)	Merck, Germany
IgG-sepharose	GE Healthcare, UK
Imidazole	Merck, Germany
IPTG	Merck, Germany
Isopropanol	Merck, Germany
LB-Agar	Q-Bio-gene, USA
LB-liquid media	Q-Bio-gene, USA
Magnesium acetate	Merck, Germany
Magnesium chloride	Merck, Germany
Maltose	Merck, Germany
Methanol	Merck, Germany
Milk powder, dry, instant	Heirler, Germany
MOPS	Invitrogen, Netherlands
Paraformaldehyde	Merck, Germany
Phenol-chloroform-isoamyl (25:24:1) alcohol (PCI)	Roth, Germany
PMSF (Phenylmethylsulfonyl fluoride)	Roche, Germany
Ponceau S	Serva, 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
Protein A-sepharose CL-4B	GE Healthcare, UK
PVP (Polyvinylpyrrolidone)	Merck, Germany
Roti-Phenol-Chloroform-Isoamyl alcohol (PCI)	Roth, Germany
Rotiphorese Gel 30 solution	Roth, Germany
Rotiphorese Gel 40 solution	Roth, Germany
snRNA marker	In-house
Sodium acetate	Merck, Germany

Sodium chloride	Merck, Germany
Sodium dodecyl sulfate (SDS)	Serva, Germany
Spermidine	Sigma-Aldrich, Germany
SYBR Gold Nucleic Acid Gel Stain	Thermo Fisher Scientific,
TEMED (N, N, N', N'-Tetramethylethylenedia mine)	Sigma-Aldrich, Germany
Tris [Tris-(hydroxymethyl)aminomethane]	VWR International,
Triton X-100	Sigma-Aldrich, Germany
tRNA <i>E. coli</i>	Boehringer, Germany
Tween 20	Sigma-Aldrich, Germany
Uranyl formate	In-house
Urea	Merck, Germany
Xylene cyanol	Fluka, Switzerland
$\beta$ -Mercaptoethanol	Roth, Germany

### 2.1.2 Consumable laboratory material

96 well Plates	Bio-Rad, Germany
Amicon Ultra 0.5 ml, 15 ml (MWCO 50 and 100 kDa)	Merck, Germany
Amylose resin	New England Biolabs, Germany
Carbon rods, highest grade	Ringsdorff Werke GmbH, Bonn
Centrifuge tubes	Beraneck, Germany
Copper EM grids Quantifoil R2/2 and R3.5/1, Cu, 200 mesh	Quantifoil Micro Tools GmbH, Jena
Copper EM grids, 200 mesh square fine bar	Science Services, Munich, Germany
Ethane (liquid)	Messer, Sulzbach, Germany
Filter paper, 90 mm	Whatman (GE Healthcare)
Megafuge 1.0R	Kendro, USA
Mica, 75x25 mm	Plano, Wetzlar, Germany
Microfuge tubes (0.5 ml, 1.5 ml and 2.0 ml)	Eppendorf, Germany
Microfuge tubes (15 ml and 50 ml)	Sarstedt, Germany
Milli-Q-water supply apparatus	Millipore, USA
Needles BD Spinal 18GA 3.50 IN 1.2 x 90 mm	Becton Dickinson, Spain
Nitrogen (liquid)	Air Liquide, Paris, France
NuPAGE™ precast gels (1.5 mm, 4-12%)	Invitrogen, Netherlands

Nylon membrane Hybond XL	Amersham Pharmacia, Germany
Parafilm	Roth, Germany
pH-Meter	Mettler Toledo, Switzerland
Pipette tips (1.0, 0.2, 0.01 ml)	Roth, Germany
Pipettes	Eppendorf, Germany
Poly-Prep columns 0.8 x 4-cm	Bio-Rad, USA
ProbeQuant™ G-25 micro columns	GE Healthcare, UK
ProbeQuant™ G-50 micro columns	GE Healthcare, UK
Protran Nitrocellulose membrane 0.2 µm	Schleicher & Schüll, Germany
SnakeSkin™ Dialysis Tubing, 10K MWCO, 22 mm	Thermo Fisher Scientific Inc., USA
Sterile filters (0.2 µm or 0.45 µm)	Sarstedt, Germany
Surgical blades	Martin, Germany
X-ray films BioMax MR	Kodak, USA

### 2.1.3 Enzymes and enzyme inhibitors

AcTEV Protease, recombinant (10 U/µl)	Invitrogen, Netherlands
Complete™ protease inhibitor tablets, EDTA-free	Roche, Germany
Phenylmethylsulfonylfluoride (PMSF)	Merck, Germany
Proteinase K	Sigma-Aldrich, Germany
Sequencing grade trypsin, porcine	Promega, Germany
T4 Polynucleotide Kinase	New England Biolabs, Germany

### 2.1.4 Monoclonal and polyclonal antibodies

Peroxidase-antiperoxidase complex (PAP)	Sigma, Germany
Anti-TAP antibody ('anti-CBP')	Biocat, Germany
Goat anti-rabbit/anti-mouse antibodies,	Jackson ImmunoResearch, USA

### 2.1.5 Nucleotides

Nucleoside-5'-triphosphate (100 mM): (ATP, CTP, GTP, UTP)	Pharmacia, Germany
--	--------------------

**Radiolabelled nucleotides**

Radionucleotides: [*alpha*-<sup>32</sup>P] dATP (3000Ci/mmol) PerkinElmer, Germany

Radionucleotides: [*gamma*-<sup>32</sup>P] ATP (6000Ci/mmol) PerkinElmer, Germany

**2.1.6 Oligonucleotides****DNA oligonucleotides for end labelling**

Sequences are displayed in 5' → 3' orientation.

snRNA	Complementary Sequence
U1	CACGCCTTCCGCGCCGT
U2	CTACACTTGATCTAAGCCAAAAG
U4	AGGTATTCCAAAAATTCCC
U5	AAGTTCCAAAAAATATGGCAAGC
U6	ATCTCTGTATTGTTTCAAATGACCAA

**2.1.7 Yeast strains**

Name	Genotype	Description
YEK2	<i>MATa; trp1-Δ1; his3-Δ; ura3-52; lys2-801; ade2-101; BRR2::TAP Tag-TRP1 C-terminus</i>	haploid yeast strain like TR2a, in which TAP tag was inserted at the C-terminus of <i>BRR2</i> using plasmid pBS1479 (Karagöz, 2006)
YFR005C	<i>SC0000; MATa; ura3-52; leu2-3,112; YFR005c::TAP-K1URA3</i>	Yeast strain with TAP tag at the C-terminus (Euroscarf)

**2.1.8 Commercial reaction sets (kits)**

BCA protein assay kit	Pierce., USA
Brilliant Blue G-Colloidal Concentrate	Sigma, Germany
ECL Western Blot Detection Kit	GE Healthcare, UK
Bradford Assay	Bio-Rad, Germany

### 2.1.9 Laboratory apparatuses

Autoclaves	H+P Labortechnik, Germany
Balances	Sartorius, Germany
Biofuge fresco centrifuge	Kendro, USA
Biofuge pico centrifuge	Kendro, USA
Centrifuge Megafuge 1.0R	Heraeus, Germany
Cryo-electron microscope CM200 FEG	Philips, Netherlands
Cryo-electron microscope Titan Krios	FEI, Eindhoven, Netherlands
DNA Thermal Cycler	Hybaid Omni Gene, UK
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, USA
Gel electrophoresis apparatus	In-house
Geldryer Model 583	Bio-Rad, USA
Gel electrophoresis apparatus	In-house
Glassware	VWR International, Germany
Gradient master model 106	BioComp Instruments, Canada
Gradient Master	BioComp Instruments, Canada
Head-over-tail Rotor	Cole-Parmer, USA
Heating blocks	Eppendorf, Germany
Hybridization oven	Hybaid Biometra, UK
Ice machine	Ziegra, UK
Incubators	Heraeus, Germany
Liquid Scintillation Analyzer Tri-Carb 2100 T	Packard, USA
Liquid scintillation analyzer	Packard, USA
LTQ-Orbitrap Velos	Thermo Fisher Scientific, Germany
Magnetic Stirrer	IKA Staufen, Germany
Megafuge 1.0R	Kendro, USA
Microwave oven	Bosch, USA

Milli-Q-water supply unit	Millipore, USA
Mortar Grinder RM100	Retsch, Germany
Orbitrap Fusion Lumos Tribrid Mass Spectrometer	Thermo Fisher Scientific, Germany
Peristaltic Pump P-1	Pharmacia Biotech, Sweden
pH-Meter	Thermo Fisher Scientific, Germany
Phosphorimager screens	GE Healthcare, Germany
Phosphorimager Typhoon Trio+	GE Healthcare, Germany
Pipettes	Eppendorf, Germany
Pipetting Robot Microlab Star LET	Hamilton, USA
Power supply EPS 2A 2000	Hofer Pharmacia Biotech, USA
Power supply EPS 2A200	Hofer Pharmacia Biotech, USA
Power supply EPS 3501/XL	Amersham Pharmacia, Germany
ProbeQuant™ G-25/ G-50 micro columns	GE Healthcare, UK
Q Exactive HF Hybrid Quadrupol-Orbitrap Mass-Spectrometer	Thermo Fisher Scientific, Germany
Rocking platform MAX Q3000 and Q2000	Thermo Fisher Scientific, Germany
Room temperature EM specimen holder	Philips, Eindhoven, Netherlands
SnakeSkin™ Dialysis Tubing, 10K MWCO, 22 mm	Thermo Fisher Scientific Inc., USA
Sorvall HB-6 rotor	Kendro, USA
Sorvall SA800 AT4 rotor	Kendro, USA
Sorvall SLC-6000 rotor	Kendro, USA
Sorvall SS-34 rotor	Kendro, USA
Sorvall T647.5 Rotor	Kendro, USA
Sorvall T865 Rotor	Kendro, USA
Sorvall TH660 Rotor	Kendro, USA
Sorvall TST41.14 rotor	Kendro, USA
Spectrophotometer Nanodrop ND-1000	Thermo Fisher Scientific, Germany
Spectrophotometer Ultrospec 3000 pro	Amersham Pharmacia, Germany
Speed Vac Concentrator 5301	Eppendorf, Germany
SureSpin630 Rotor	Thermo Scientific, Germany
Tabletop centrifuge 5415D	Eppendorf, Germany
Tabletop centrifuges	Heraeus, Germany
Teflon blocks	In-house
Thermomixer	Eppendorf, Germany

Trans-Blot Cell	Bio-Rad, USA
Ultracentrifuge Discovery 90 SE	Sorvall/Kendro, USA
Ultracentrifuge Discovery M150	Sorvall/Kendro, USA
Ultracentrifuge Evolution	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
X-ray film developer X-Omat 2000	Kodak, USA
Film exposure cassettes	Kodak, USA

### 2.1.10 Cell culture

Media for yeast culture was prepared as described by Sambrook et al., (Sambrook et al., 1989) using deionized water from a Millipore apparatus and were autoclaved at 121°C and 1 bar for 15 min. For the preparation of culture plates, the sugars contained in the yeast media were added before autoclaving and then poured into Petri plates under a laminar flow hood. For the preparation of the liquid medium, the sterile filtrated sugars solutions were added to the medium after autoclaving.

### 2.1.11 Common buffers

Commonly used media, buffers, and solutions were primarily prepared, as described (Sambrook, 1989; Sambrook et al., 1989). Deionized water was from a Millipore apparatus. Solutions were autoclaved if necessary (121°C, 20 min, and 1 bar). Thermolabile components were filter-sterilized (0.22 µm). The pH was adjusted using HCl or NaOH.

6x DNA loading dye	60% glycerol (v/v) 10 mM EDTA, pH 7.5 0.25% (w/v) bromophenol blue 0.25% (w/v) xylene cyanol
50x Denhardt's solution	1% (w/v) Ficoll 400



	1% (w/v) Polyvinylpyrrolidone (PVP)
	1% (w/v) Bovine serum albumin
10x G-150 buffer:	200 mM HEPES 1.5 M KCl 15 mM MgCl <sub>2</sub> Adjust to pH 7.5 at 4°C
1x GK-150 buffer:	10% (v/v) 10 x G-150 buffer 5% (v/v) 100% glycerol
10x G-75 buffer:	200 mM HEPES 0.75 M KCl 15 mM MgCl <sub>2</sub> Adjust to pH 7.5 at 4°C
1x GK-75 buffer:	10% (v/v) 10 x G-75 buffer 5% (v/v) 100% glycerol
1 M K-PO <sub>4</sub>	1 M K <sub>2</sub> HPO <sub>4</sub> 1 M KH <sub>2</sub> PO <sub>4</sub> pH 7.5 at 4°C
20x MOPS buffer	1 M MOPS 1 M Tris-Base 2% SDS 20mM EDTA Adjust to pH 7.7
10x PBS buffer	1.3 M NaCl 188 mM K <sub>2</sub> HPO <sub>4</sub> 12 mM KH <sub>2</sub> PO <sub>4</sub> pH 8.0
1x Protein loading dye	75 mM Tris-HCl, pH 6.8

	1.25 mM EDTA, pH 8.0
	20% (v/v) glycerol
	2.5% (w/v) SDS
	0.125% (w/v) bromophenol blue
	50 mM DTT
RNA loading dye (FA dye)	80% formamide
	1 mM EDTA pH 8.0
	0.05% (w/v) bromophenol blue
	0.05% (w/v) xylene cyanol
1x Running buffer SDS-PAGE	25 mM Tris-HCl, pH 6.8
	192 mM glycine
	1% (w/v) SDS
4x Separating gel buffer	1.5 M Tris
	0.4% (w/v) SDS
	adjust to pH 8.8
4x Stacking gel buffer	0.5 M Tris
	0.4% (w/v) SDS
	adjust pH to 6.8
Slab 4 Buffer	50 mM Tris
	105 mM glycine
	0.1% (w/v) SDS
20x SSC	3M NaCl
	0.3 M NaCitrate
10x TBE	0.89 M Tris
	0.89 M boric acid
	25 mM EDTA pH 8.0

10x TBS	200 mM Tris 1.37 M NaCl adjust to pH 7.6
10x TE	100 mM Tris-HCl, pH 8.0 10 mM EDTA, pH 8.0
1x Western blotting buffer (Transfer buffer)	1.5 L Slab4 Buffer 0.6 L methanol 0.9 L ddH <sub>2</sub> O

### 2.1.12 Commercial reaction sets (kits)

BCA protein assay kit	Thermo Fisher Scientific Inc., USA
Bradford Assay	Bio-Rad, Germany
Brilliant Blue G-Colloidal Concentrate	Sigma, Germany
ECL Western Blot Detection Kit	GE Healthcare, UK
Expand Long Template PCR Mix	Roche, Germany
Prime It II Random Primer Labeling Kit	Stratagene, USA

### 2.1.13 Software programs

Program	Source
Coweyes	Custom made (Department of Structural Dynamics, MPI-BPC)
Eman 2	<a href="http://blake.bcm.edu/emanwiki/EMAN2">http://blake.bcm.edu/emanwiki/EMAN2</a> (Tang et al., 2007)
Gautomatch	<a href="http://www.mrc-lmb.cam.ac.uk/kzhang/">http://www.mrc-lmb.cam.ac.uk/kzhang/</a>
Gctf	<a href="http://www.mrc-lmb.cam.ac.uk/kzhang/">http://www.mrc-lmb.cam.ac.uk/kzhang/</a> (Zhang, 2016)
ImageJ	<a href="http://imagej.nih.gov/ij/">http://imagej.nih.gov/ij/</a> (Schneider et al., 2012)
JohnHenry	Custom made (Department of Structural Dynamics, MPI-BPC)
Relion 1.4	<a href="http://www2.mrc-lmb.cam.ac.uk/relion">http://www2.mrc-lmb.cam.ac.uk/relion</a> (Scheres, 2012)
UCSF Chimera	<a href="http://www.cgl.ucsf.edu/chimera/">http://www.cgl.ucsf.edu/chimera/</a> (Pettersen et al., 2004)
PyMOL by Schrödinger	The PyMOL Molecular Graphics System, Version 2.3 Schrödinger, LLC. <a href="https://pymol.org/2/">https://pymol.org/2/</a>

## Methods

### 2.2 Protein-biochemistry standard methods

#### 2.2.1 Protein quantification

To quantitatively determine the concentration of proteins, the BCA<sup>TM</sup> Protein Assay Kit (Pierce) was used according to the manufacturer's instructions. It is a detergent-compatible formulation based on bicinchoninic acid (BCA) for the colorimetric detection and quantification of total protein.  $\text{Cu}^{2+}$  is reduced to  $\text{Cu}^{1+}$  by protein in an alkaline medium (Biuret Reaction). The BCA chelates  $\text{Cu}^{1+}$  ions forming purple-coloured complexes with an absorption maximum at 562 nm (Smith et al., 1985).

Another method that was sometimes employed to determine the protein contents was Bradford protein assay. It involves the binding of Coomassie Brilliant Blue G-250 dye to proteins. The dye exists in three forms: cationic (red), neutral (green), and anionic (blue). In acidic conditions, the dye is primarily in the cationic form with absorption at 470 nm. When the dye binds to a protein, it is changed to the blue colour, and the absorption changes to 595 nm. Using a protein standard (such as BSA in known concentration), it is possible to estimate the protein concentration of the sample solution.

#### 2.2.2 PCI (Phenol-Chloroform-Isoamylalcohol) extraction

The Phenol-Chloroform-Isoamylalcohol extraction (PCI-extraction) is a method that is used for separating proteins from nucleic acids. The sample solution was supplied with the same volume of PCI-solution and thoroughly mixed. The mixed suspension was centrifuged for 5 min at room temperature at 13000 rpm to separate the aqueous and organic phases. The two phases were separated as the upper aqueous phase and the lower organic phase. Carefully remove the upper phase in a separate microfuge without mixing the two phases. The nucleic acids in the upper aqueous phase were precipitated by the addition of 1/10 volume of a 3 M NaOAc pH 5.2 and 2.5 volume of 100% ethanol. To increase precipitation of RNAs, 15-30  $\mu\text{g}$  of glycoblue was added as a carrier. The proteins in the lower organic phase were precipitated with 2.5 to 3 vol. of cold 100% ethanol. Nucleic acids or proteins were precipitated overnight at  $-20^{\circ}\text{C}$  or one hour at  $-80^{\circ}\text{C}$ . Subsequently, they were pelleted for 20-30 min at  $4^{\circ}\text{C}$ , 13000 rpm. The pellet was washed

with 70% (v/v) cold ethanol and pelleted again. The recovered pellet was dried for 2-3 min under vacuum and resuspended in an appropriate buffer or RNAase free Milli-Q H<sub>2</sub>O.

### 2.2.3 Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated by denaturing polyacrylamide gel electrophoresis (SDS-PAGE) adapted from (Laemmli, 1970) or by commercially available pre-cast NuPAGE™ Bis-Tris gels (4-12%), for MS analyses. Protein samples were dissolved in 1x SDS loading dye and proteins were denatured for 5 min at 95°C and loaded onto a high TEMED (0.33% (v/v)) polyacrylamide gel (PAA) (acrylamide (AA):bisacrylamide (BA) = 37.5:1). 0.033% (w/v) APS was used for polymerization. The separating gel used varied from 12% to 8%. The stacking gel was 5% PAA. Gels were run in a vertical gel running unit in a 1x SDS-PAGE running buffer until the bromophenol blue dye reached the bottom. Electrophoresis was done at 200V for 45-50 min using Bio-Rad mini gel systems. The protein bands were visualized either by Coomassie or by silver staining.

5% Stacking gel		8% separating gel	
4x stacking gel buffer	375.0 µl	4x separating gel buffer	1250.0 µl
Rotiphorese Gel 30	250.5 µl	Rotiphorese Gel 30	1332.5 µl
H <sub>2</sub> O	874.5 µl	H <sub>2</sub> O	2417.5 µl
APS	10.0 µl	APS	16.75 µl
TEMED	1.0 µl	TEMED	16.75µl

For commercial gels, the manufacturer's instructions were obeyed.

### 2.2.4 Coomassie staining

This protein staining technique exploits the Coomassie Brilliant Blue G-250, which is a triphenylmethane dye forming stable noncovalent complexes with proteins. Coomassie staining of protein gels was basically performed as described by Sambrook, Russell (Sambrook et al., 1989). Coomassie staining of proteins was performed after fixation of SDS denaturing gels in 40% (v/v)

Methanol / 10 % (v/v) acetic acid for 30 min at room temperature on a shaker. After rehydration of the gel in Milli-Q H<sub>2</sub>O, the pre-heated staining solution was added on top of the gel. Incubation was continued for at least 1 h at room temperature with gentle shaking. After that, the gel was destained with water with mild shaking until protein bands were clearly visible. Finally,

the gel was scanned, and individual lanes were analysed by mass spectrometry (Research group Mass Spectrometry -MPI-BPC, Göttingen)

### 2.2.5 Silver staining of protein gels

Silver staining of protein gel was mainly performed as described by Blum et al. (Blum et al., 1987). Before staining, all gels were fixed using the fixing solution of 40% (v/v) methanol and 10% (v/v) acetic acid for at least 30 min. During the entire process, all the solutions were kept at least 10 times the gel volume, and all steps were performed on a shaker. After fixation, the gel was washed twice with 50% ethanol and once with 30% ethanol for 20 min each. The gel was then submerged into a 0.8 mM Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> solution for precisely 60 seconds. Immediately afterward, the gel was washed three times shortly with ddH<sub>2</sub>O to remove the thiosulfate from the surface of the gels. After that, the gel was incubated with 0.012 M AgNO<sub>3</sub>, 0.026% formaldehyde solution for 30 min. After carefully removing the staining solution, the gel was washed at least three times with ddH<sub>2</sub>O before applying the developing solution containing 0.56 M Na<sub>2</sub>CO<sub>3</sub>, 0.0185% formaldehyde, 16 μM Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>. Developing was stopped with a solution of 40% methanol, 10% acetic acid. The gel was then transferred to a Whatman paper and dried under vacuum at 80°C for 45 min.

### 2.2.6 Western blot analysis

After separating proteins by SDS-PAGE, proteins were transferred to a nitrocellulose membrane by the wet-blotting procedure, according to Burnette et al. (Burnette, 1981). The transfer was carried out for 2 h at 70 V and 4°C in 1x western blotting buffer (1x transfer buffer). Membranes were washed with ddH<sub>2</sub>O, (stained with 'Ponceau S' if necessary) and blocked for 2 hr at room temperature (RT) or overnight at 4°C in 1x TBS, 0.1% (v/v) Tween-20, 5% (w/v) milk powder. Immunoblotting was carried out for 1 hr at RT or overnight at 4°C with primary antibodies specified according to standard protocols (dilutions as per supplier's protocol). Horseradish peroxidase (HRP)-conjugated anti-rabbit antibodies were used as the secondary antibody (1 hr at RT, dilution 1:50000), and proteins were detected by enhanced chemiluminescence using an ECL kit according to the manufacturer's instructions. Where TAP tag detection was required, the Peroxidase-antiperoxidase complex (PAP) was used to detect Protein A of the TAP tag, and no secondary antibody was required.

## 2.3 Molecular biology

### 2.3.1 Concentration determination of nucleic acids

The concentration determination of nucleic acids is based on the absorption maximum of the aromatic rings present in the nucleic acids at a wavelength of 260 nm. To determine the concentration of nucleic acids, its absorption extinction in an aqueous solution was measured at wavelengths 260 nm and 280 nm in comparison to the corresponding buffer without nucleic acids using a NanoDrop spectrophotometer according to the manufacturer's instructions. The concentration was then calculated using pre-determined absorption values at 260 nm, as described (Sambrook et al., 1989). The ratio of  $OD_{260}/OD_{280}$  determined the purity of the nucleic acid solution, which is usually 2.0 for pure RNA or oligonucleotides, and 1.8 for pure DNA. Lower ratios indicate the presence of protein, phenol, or other impurities, which must be removed by PCI-extraction (2.2.2).

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

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

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

### 2.3.2 Agarose gel electrophoresis of nucleic acids

To analyse DNA probes against different snRNAs, agarose gel electrophoresis of nucleic acids was performed as described (Sambrook et al., 1989). Depending on the length of the oligos, our gels contained 1 % (w/v) agarose and 0.4  $\mu\text{g}/\text{ml}$  ethidiumbromide in 1x TBE buffer. Samples were supplemented with a 6x DNA loading dye. The nucleic acids were separated in a horizontal gel running assembly at 100V in 1x TBE buffer. The bands were visualized with UV-light at 254 nm.

Migration rates of the marker dyes through agarose gels		
% agarose (w/v)	xylene cyanol	bromophenol blue
0.7-1-7	4000 bp	300 bp
2.5-3.0	800 bp	100 bp

### 2.3.3 Proteinase K digestion

To obtain a protein-free RNA sample, proteinase K digestion was performed. To each sample reaction, 2  $\mu$ l of the stock solution (proteinase K mix) was added. The proteinase K digestion contained final concentration as - 0.2% (w/v) SDS, 10 mM EDTA, and 0.3 mg/ml proteinase K. All sample reactions were incubated for 30 min at 37°C. RNA was extracted by PCI-extraction and precipitated.

### 2.3.4 Denaturing polyacrylamide gel electrophoresis of RNA

To separate different snRNAs present in the tri-snRNP fractions, denaturing polyacrylamide gel-electrophoresis was used. The gels contained 8 M urea, 0.5x TBE, and according to the size of the snRNA, the concentration of polyacrylamide was kept 8%. The gels were polymerized with 50  $\mu$ l 10% (w/v) APS and 5  $\mu$ l TEMED per 10 ml gel solution. RNA samples were resuspended in 10  $\mu$ l of RNA loading buffer and denatured for 5 min at 95°C and immediately transferred on ice. The electrophoresis was carried out in 1x TBE buffer at 500V until the xylene cyanol reached the bottom of the gel. The RNA fragments on the gel were visualized by silver staining or by SYBR<sup>TM</sup> Gold nucleic acid staining method.

<b>Migration rates of marker dyes in denaturing polyacrylamide gels</b>		
<b>polyacrylamide (%)</b>	<b>xylene cyanol</b>	<b>bromophenol blue</b>
5.0	130 bases	35 bases
6.0	106 bases	29 bases
8.0	76 bases	26 bases
10.0	55 bases	12 bases
20.0	28 bases	8 bases

### 2.3.5 Silver staining of RNA gels

Silver staining of RNA gels was performed according to Merrill et al. (Merrill et al., 1981). During the procedure, at least a 10-times the gel volume was used for all solutions. Sample containing gel was fixed in a solution of 40% methanol and 10% acetic acid for at least 30 min. After fixating,



the gel was washed twice with a solution containing 10% ethanol, 5% acetic acid for 15 min to eliminate interfering substances. Then, the gel was washed with ddH<sub>2</sub>O once and subsequently stained in 12 mM AgNO<sub>3</sub> solution for 30 min. After incubation with the staining solution, the gel was washed twice with ddH<sub>2</sub>O twice. Then, the gel was quickly treated with the developing solution (0.28 M Na<sub>2</sub>CO<sub>3</sub>, 0.0185% formaldehyde) to remove excess silver from the surface. Carefully discard the solution and again added a fresh developing solution. Developing was stopped by adding a 5% acetic acid solution. The gel was transferred to a Whatman paper and dried under vacuum at 80°C for 1 hr.

### **2.3.6 SYBR™ Gold nucleic acid staining**

SYBR™ Gold staining of RNA gels was performed according to manufacturer's protocol after fixation of in 40% (v/v) Methanol / 10% (v/v) acetic acid. After staining the gel, RNA was visualized with Fujifilm FLA-7000 laser scanner at a wavelength of 495 nm.

### **2.3.7 Radioactively labelled DNA-probes synthesis for Northern analysis**

To precisely detect tri-snRNP U snRNAs by the Northern blot analysis, radioactively labelled DNA probes were produced from their corresponding DNA templates with the help of the 'Prime It II Random Primer Labeling Kit' (Stratagene). DNA templates were (kindly provided by Dr. Patrizia Fabrizio from our department at MPI-BPC) analysed for degradation by an agarose gel. For DNA probes synthesis, 25 - 50 ng of DNA template was mixed with 23 µl ddH<sub>2</sub>O and 10 µl of random ninemer primers and were denatured at 95°C for 5 min. Subsequently, the mix was cooled down at room temperature, followed by short centrifugation in order to allow hybridization. The reaction mixture was then supplied with 10 µl of dATP primer buffer, 5 µl of [ $\alpha$ -<sup>32</sup>P] dATP (3000 Ci/mmol), and five units of Klenow enzyme. Then, the reaction mixture was incubated for 10 min at 40°C. After completion of the required time period, 2 µl of 0.5 mM EDTA solution was added to stop the reaction. The DNA-probes were then isolated from free radioactive dATP by using a 'ProbeQuant™ G-50' columns as per the manufacturers' directions. For Northern blot analysis, typically 10-20x 10<sup>6</sup> cpm of the labelled DNA-probes were used.

Alternatively, oligonucleotides were 5' end-labelled with [*gamma*-<sup>32</sup>P] ATP (6000Ci/mmol) using the T4 PNK (Polynucleotide Kinase) from NEB, as per the manufacturer's instructions.

### 2.3.8 Northern blot analysis

Northern blot analysis was carried out to detect a specific snRNA in the tri-snRNP sample. For this, samples were first separated by denaturing polyacrylamide gel electrophoresis (as described in section 2.2.2.4) and transferred onto a Biodyne nylon membrane (Pall Corp.) using a semi-dry transferring method (PeqLab Biotechnologie GmbH). For semi-dry transfer, 2 Whatman paper sheets soaked in 0.5 x TBE buffer were placed on the cathode plate, followed by the pre-wet nylon membrane, gel, and again 2 more pre-soaked Whatman filter paper sheets. The sequence of forming the transfer sandwich was carefully followed precisely in the described order. After gently removing the bubbles from the transfer setup, finally, the anode plate was softly placed onto the transfer sandwich. Gradually screws were tightened, and the transfer unit was connected with the power supply. The transfer was carried out for 2 hr at room temperature at a constant current supply of 2 mA/cm<sup>2</sup>. After completion of the RNA transfer, the unit was carefully opened to retrieve the nylon membrane. With the RNA side facing up, the membrane was exposed to UV irradiation using the Stratlinker 2400 (from Stratagene, USA) to crosslink the RNA with the membrane.

Radiolabelled DNA probes were generated using the protocol described in section 2.2.2.7. The RNA containing membrane was pre-hybridized for at least 2hr at 37°C in the hybridization buffer containing denatured fish sperm DNA. After that, denatured radiolabelled probes were added, and the incubation was continued overnight at 42°C. The next morning, the membrane was washed two times with 25 ml of 2xSSC, and 0.5% SDS solution for 1 min. And twice again, with 25 ml 2xSSC, and 0.1% SDS solution. All washings were done at room temperature. In the final step, the washed membrane was dried and wrapped in the transparent plastic foil. The membrane was then visualized by exposing onto the phosphorimaging screen or the X-ray film for sufficient duration.

## 2.4 Cell culture methods

### 2.4.1 Yeast cell culture

*S. cerevisiae* cells were cultivated according to standard methods (Sambrook et al., 1989).

**Cell culture on plates:** the medium (YPD) was autoclaved for 15 min at 121°C and cooled down to 50°C To prepare plates. Subsequently, sterile 50% D (+)-glucose was added to the autoclaved medium to a final concentration of 2%. The medium was poured into Petri dishes and cooled down at RT until the agar in the medium was solidified. The plates were stored at room temperature. Yeast cells scraped from a glycerol stock were streaked on the plates and incubated 2-4 days at 25°C or 30°C, depending on the yeast strain. For more extended storage, the plates were sealed adequately with a parafilm and kept at 4°C for up to 3 months.

**Liquid cell culture in the flask:** The liquid cultures of yeast cells, usually 5-20 ml of pre-culture in YPD medium were prepared and kept overnight at the suitable temperature in the shaking incubator. The OD<sub>600</sub> of the cells in pre-culture was measured, and usually, 15 OD of cells were inoculated into 2 litres of YPD culture in a 5 L Erlenmeyer flask. After 16-17 hr, the cells reached OD<sub>600</sub> of 3-4.5/ml. Alternatively, 100L fermenter was often used to grow yeast cells in preparative quantities. Fermentation units were operated by Mr. Thomas Schulz (Facility for crystallography- MPI-BPC, Göttingen)

**Glycerol stocks for yeast:** For long time storage, 800 µl of a fresh yeast pre-culture in YPD medium were mixed with 400 µl of sterile 50% glycerol and kept at -80°C.

#### **2.4.2 Extract preparation from yeast cells for snRNP purification using an electric grinder**

Initially, whole-cell extract from yeast cells for tri-snRNP purification was prepared as described by my previous lab members (Gottschalk, 1999); Häcker et al., 2008), however later certain modifications were introduced which are mentioned in the following section. Yeast cells were grown overnight at 30°C up to OD<sub>600</sub> of 3.5-4/ml. Cells were pelleted by centrifugation for 10 min at 5000 rpm using an SLC-6000 rotor. The cell pellet from 2 L yeast culture was washed with 200 ml of cold ddH<sub>2</sub>O and with 50 ml AGK buffer (20 mM HEPES/KOH pH 7.5, 200 mM KCl, 1.5 mM MgCl<sub>2</sub>, 10% (v/v) glycerol). The cells were resuspended in equal volume (w/v) of the AGK buffer (additionally containing 0.5 mM DTT, 0.5 mM PMSF, and protease inhibitor mixture) per 2 L culture and transferred into a 50 ml Falcon tube. KCl was adjusted to 200 mM, and the cell suspension was dropped into liquid nitrogen, using either a glass pipette or a dropping funnel. The cell beads were stored at -80°C or broken using an electric grinder (Retsch). In one round of grinding, beads from 8 L culture were ground for 15 min at the highest pressure. The resulting cell powder was either stored at 4°C to liquefy slowly. The cell lysate was centrifuged for 30 min at 17000 rpm (25000 g) using an SS-34 rotor to remove cell debris. The

supernatant was then transferred into a new tube and further centrifuged for 1 h at 37000 rpm (100000 g) using a T-865 rotor. Alternatively, for large amounts of cells, a T647.5 rotor was used for 50 min at 42000 rpm (100000 g). After the high-speed centrifugation step, the cell lysate separated into three different phases: the uppermost white layer included lipids and liposomes; the middle yellow phase contained total cell extract, and the pellet comprises ribosomes, fine cell debris, and genomic DNA. The middle phase was recovered with a long spinal needle and dialyzed three times for 1 h each time against Buffer D (20 mM HEPES/KOH pH 7.9, 50 mM KCl, 0.2 mM EDTA, 0.4 mM MgCl<sub>2</sub>, 20% (v/v) glycerol, 0.5 mM DTT, 0.5 mM PMSF). The dialyzed extract was shock-frozen in liquid nitrogen and stored at -80°C.

## 2.5 Special methods

### 2.5.1 Tandem affinity purification of yeast snRNPs

The tandem affinity purification (TAP) is a two-step affinity purification employing first an IgG matrix and then a density gradient separation. Yeast strains expressing TAP-tagged snRNP proteins were cultivated, and frozen yeast whole-cell extract from these strains was prepared as described in section 2.4.2 above. The frozen extract was thawed in a water bath at 4°C. Initially, the buffer composition of the extract was adjusted to 10 mM Tris/HCl pH 8.0, 100 mM NaCl (in addition to the 50 mM KCl resulting from dialysis), and 0.1% NP-40. However, in later experiments which were performed under low salt condition (75mM K<sup>+</sup> condition), the step of buffer composition adjustment was abolished entirely from the protocol and only was followed where KCl conditions were required to be adjusted. IgG-sepharose beads (300 µl beads for 30ml extract of yeast culture) were washed with 10 ml of 1x GK150 or 1x GK 75 buffer and incubated with yeast extract in a 50 ml Falcon tube for 2 h at 4°C by using a head-over-tail rotor. After binding to the IgG-sepharose beads, the extract was transferred into 10 ml (0.8 x 4 cm) Bio-Rad Poly-Prep columns such that each column contained 300 µl of IgG beads. The columns were drained by gravity flow. The beads were washed three times with 10 ml of 1x GK 150 or 1x GK 75 buffer. Then the TAP tag was cleaved with 300 units TEV protease per 30ml culture in 0.6 ml of TEV cleavage buffer/ 1xGK 150 / 1x GK 75 buffer for 2 h at 16°C, rotating head-over-tail. Finally, the eluate was collected by gravity flow into a new microfuge and further subjected to a glycerol gradient centrifugation after determining the protein and RNA contents.

## 2.5.2 Glycerol gradient sedimentation of purified snRNP particles

After TAP-purification, yeast snRNP particles were further purified for electron microscopy analysis and mass spectrometry by glycerol gradient centrifugation. The particles were layered on a linear 10-30% glycerol gradient. 100 ml of the gradient solutions contained 10 or 30% (v/v) glycerol, 20 mM HEPES/KOH pH 7.5, 150 mM KCl / 75 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.5 mM PMSF, and 0.5 mM DTT. The gradient solutions were filtered before use. When particles were fixed during the gradient centrifugation for EM, glutaraldehyde was added to a final concentration of 0.04 or 0.1% (w/v) to the 30% glycerol gradient solution. Gradients (TH660) were prepared using the manufacturer's instructions by a pre-installed program on the gradient master machine. The centrifugation was carried out in an ultracentrifuge for 15 hr at 4°C and 29000 rpm for tri-snRNP particles using a Sorvall TH660 rotor. After centrifugation, 24 fractions of ~185 µl were collected. Fraction collection was done from the bottom of the gradient tube by using a fraction-collector.

## 2.6 Electron Microscopy

### 2.6.1 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. The mica was freshly split to obtain a smooth surface that does not have beforehand any direct air contact. 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 a self-made triafol film containing holes. Then carbon was directly sputtered onto the triafol film. The grids were incubated in ethyl acetate overnight and air-dried to dissolve and remove the triafol film between the carbon layer and copper grids. 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.

### 2.6.2 Sample preparation

For EM investigations, affinity-purified complexes were subjected to a glycerol gradient centrifugation under mild-fixating GraFix conditions (Kastner, Fischer et al. 2008). The affinity-purified tri-snRNP particles were loaded onto a linear 10-30% glycerol, and 0-0.1% (v/v) glutaraldehyde gradient prepared in GK-150 / GK 75, specified as per the experiments. Gradients were centrifuged at 29000 rpm for 15 hr in a TH660 rotor and harvested from the bottom in 175 to 185  $\mu$ l fractions ( $\sim$ 23-24 fractions in total). A denaturing RNA-PAGE analysis determined the distribution of UsnRNP across the gradient and based on the RNA gel, and the peak fractions were used for EM studies. The purified spliceosomal complexes were allowed to adsorb on a thin carbon film before negative staining or rapid plunge freezing into liquid ethane at 100% humidity and 4°C

### 2.6.3 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 ( $\sim$ 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. A piece of carbon-coated mica was floated on the surface of the sample well in a custom-made Teflon block to adsorb isolated complexes. Then, carbon film with the 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<sup>2</sup>/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 dataset 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, 1981a); (van Heel et al., 1996);(Dube et al., 1993). The resulting class averages were used as reference images in subsequent rounds of alignment until the class averages were stable.

#### 2.6.4 Cryo- EM image acquisition and analysis

Cryo- EM images were recorded at  $-193^{\circ}\text{C}$  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  $1.16 \text{ \AA}^2/\text{pixel}$  on the specimen level. 20 frames were recorded for each micrograph with an average dose of  $2 \text{ e}^-$  per frame per  $\text{\AA}^2$ . Motion correction and spatial frequency weighed frameshift were achieved using the unblur software suite (Campbell, Cheng et al. 2012). Ctf parameters were determined with Gctf (Zhang 2016). Sum of micrograph images was 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  $40^{\circ}$  projections of the yeast Bact model (EMD-9524) filtered to  $60 \text{ \AA}$ . In total  $\sim 0.52$  million particles from the pre-sorted cryo-micrographs were collected and cleaned by 3D classification in RELION. The remaining  $\sim 0.48$  million particles were subsequently applied to three separate rounds of 3D classification in RELION (Scheres 2012) resulting in two different forms of the human Bact 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 voxels in the final rounds of the refinement. The 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).

#### 2.6.5 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.

#### 2.6.6 Electron microscopy analysis of TAP purified snRNPs

Image acquisition was performed in a transmission electron microscope operated at 160 kV and equipped with a field emission gun (CM200 FEG, FEI, Eindhoven, The Netherlands). A 4kx4k CCD camera (TemCam F415, TVIPS, Gauting, Germany) was used with 2-fold binning of the pixels and a magnification of 122000. The specimens were prepared as described (Golas et al., 2003) and imaged at room temperature. Briefly, a carbon film was floated on the surface of the

particle solution allowing adsorption over 15 min. The carbon film was then transferred to a second well filled with a 2% uranyl formate solution and incubated for 2 min. Subsequently, the carbon film with adsorbed particles was attached to a 200-mesh copper grid on which a carbon film containing holes of  $\sim 1\text{--}2\ \mu\text{m}$  diameter had previously been mounted. Finally, another carbon film floated onto a second solution of uranyl formate was used to form a sandwich. The grids were stored under dry conditions until image acquisition. Of each sample, a few hundred up to a few thousand particle images were manually selected. Class averages were computed by three to five rounds of reference-free alignment (Dube et al., 1993) using exhaustive alignment via polar coordinates (Sander et al., 2003) and multivariate statistical analysis followed by hierarchical ascendant classification with moving elements refinement as implemented in the IMAGIC5 software (van Heel et al., 1996).

## 2.7 Mass Spectrometry

### 2.7.1 Protein identification by mass spectrometry

For MS analyses of proteins associated with tri-snRNP complexes, the peak fractions of the glycerol gradient were combined (depending on the concentration) in a low-protein binding 1.5 ml microfuge (Eppendorf). Approximately 185  $\mu\text{l}$  of a fraction (in case of an individual single fraction) was supplemented with 20  $\mu\text{g}$  glycoblue, 18.5  $\mu\text{l}$  of 3 M NaOAc (pH 5.2) and 600  $\mu\text{l}$  of absolute ethanol. After proper mixing, the solution was kept at  $-20^\circ\text{C}$  for overnight and then the pellet was recovered by centrifugation for 20 min at 13,000 rpm and  $4^\circ\text{C}$ . Afterward, the recovered pellet was washed with 70 % ethanol. Later, the washed pellet was dried in a vacuum dryer and further resuspended in 1X loading buffer from NuPAGE<sup>TM</sup> Invitrogen. The proteins were then directly separated with the Unpaged<sup>®</sup> pre-cast gel system from Invitrogen (1 mm, 4-12% Bis-Tris gel) as per the manufacturer's instructions. After running the gel, proteins were stained using Coomassie staining, which was done overnight. The gel was scanned and analysed by MS in collaboration with Dr. H. Urlaub's research group (Research group Mass Spectrometry -MPI-BPC, Göttingen). The whole lane of the Coomassie -stained gel was cut into many slices, and then, in gel digestion of proteins was carried out with trypsin as described by Shevchenko et al. (Shevchenko et al., 1996). After the extraction of peptides, they were analysed in a liquid chromatography-coupled electrospray ionization quadrupole time of flight (Q-ToF Ultima;



Waters) mass spectrometer under standard conditions. Different proteins were recognized by searching the spectra of sequenced peptides against the NCBI non-redundant database using Mascot as search engine.

### **2.7.2 Identification of Protein-protein crosslinks by mass spectrometry**

All crosslinking experiments were performed in collaboration with Dr. Olexandr Dybkov (Dep. of Cellular Biochemistry; MPI-BPC). Affinity column purified tri-snRNP complexes were cross-linked with 150  $\mu\text{M}$  BS3 for 45 min at 4°C and purified further by a density gradient centrifugation step. Approximately 10 pmol of tri-snRNP complexes were pelleted by 17 hr long ultracentrifugation and then further analysed, basically as described earlier (Leitner, Walzthoeni et al. 2014) with the mentioned modifications: Sample pellet 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 extracted by the reverse-phase method and fractionated by gel filtration on a Superdex Peptide column (PC-3.2/30 column from GE Health Care). Then, 50  $\mu\text{l}$  fractions, which corresponded to an elution volume of 1.2–1.5 ml, were analysed on an Orbitrap Fusion Lumos Tribrid and Q Exactive HF Mass Spectrometer (Thermo Scientific). Identification of protein-protein crosslinks was done by pLink 1.23 search engine and then was filtered at FDR 1% ([pfind.ict.ac.cn/software/pLink](http://pfind.ict.ac.cn/software/pLink)) according to the recommendations of the developer (Yang, Wu et al. 2012).



## 3 RESULTS

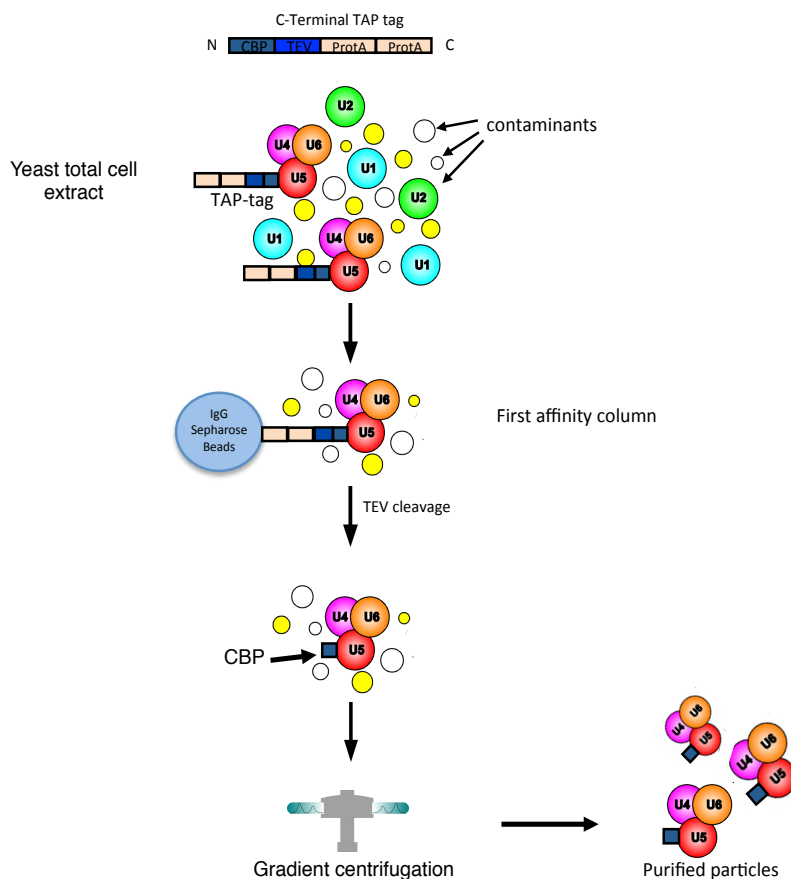
### 3.1 Tri-snRNP isolation with Tandem Affinity Purification Technique

To isolate tri-snRNPs from *S.cerevisiae* whole cell extract, we have used the well-established tandem affinity purification (TAP) technique, which utilizes genetically tagged fusion proteins and their incorporation into native snRNP complexes (Puig et al., 2001). The TAP tag comprises IgG binding domains from *Staphylococcus aureus* protein A (ProtA) and a calmodulin-binding peptide (CBP) separated by a Tobacco Etch Virus (TEV) protease cleavage site. The modified strain of *S. cerevisiae* used for the isolation of tri-snRNPs contains the TAP tag, which is genetically fused to the C-terminus of a tri-snRNP protein, enabling the purification of particles at natural expression levels and without need for overexpression of the tagged protein which might lead to unwanted protein complexes. In the following text "yeast" will be used to denote *S. cerevisiae*.

#### 3.1.1 Purification of tri-snRNPs using TAP-tagged Brr2p

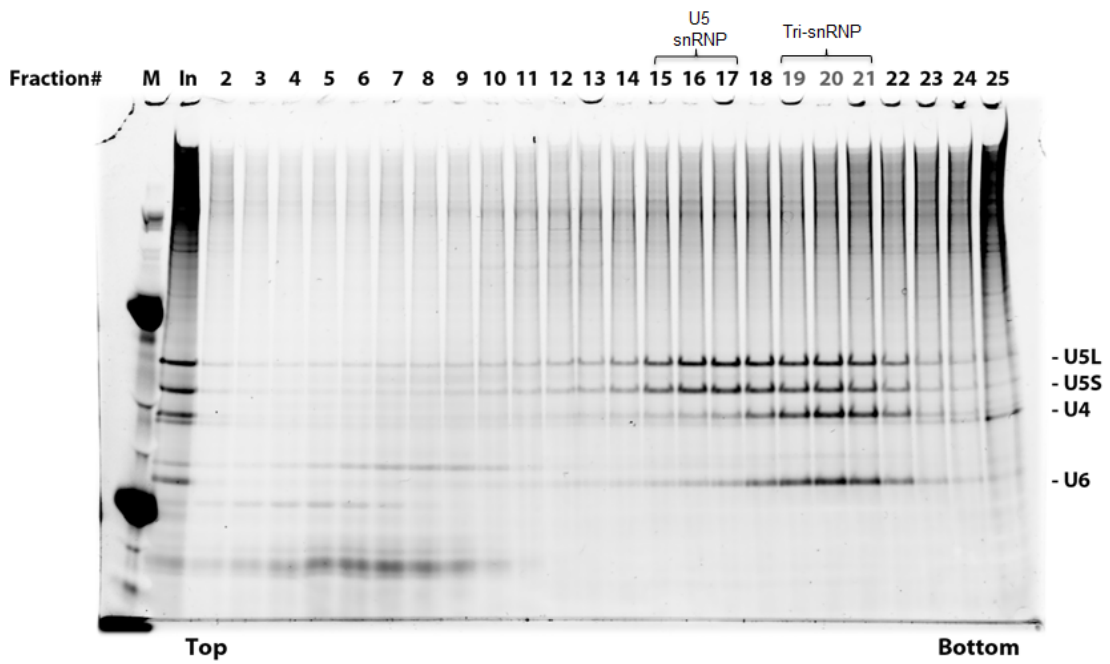
For purifications of yeast tri-snRNPs, TAP purifications have been successfully used in our lab before (Häcker et al., 2008). I initially used a TAP-tagged Brr2p strain (kindly provided by P. Fabrizio). The initial purification scheme involved two steps, a first, binding of snRNPs to an IgG-column under gravity flow and release of particles from the column using TEV protease, followed by a second, namely binding to calmodulin-Sepharose and elution with EGTA/CaCl<sub>2</sub> (Häcker et al., 2008). Initial experiments revealed a very low yield of purified tri-snRNPs; therefore, to minimize the loss of the particles during the second affinity purification step, the calmodulin-Sepharose affinity-based purification step was omitted from the final scheme of purification experiments. That is, tri-snRNPs were only affinity purified via the Brr2-TAP tag bound to the IgG Sepharose column and eluted using acTEV protease cleavage. Figure 3.1 depicts the final layout of the purification strategy being used for the tri-snRNPs purification. Particles from this one-step purification were then subjected to glycerol gradient centrifugation to obtain highly purified tri-snRNP fractions for mass-spectrometry (MS) and electron microscopy studies. In order to avoid contamination with low molecular weight compounds from the top of the gradient, fractions were collected from the bottom of the gradient. The initial experiments were carried out with buffers containing 150 mM KCl, for affinity purification and gradient centrifugations.

Figure 3.1

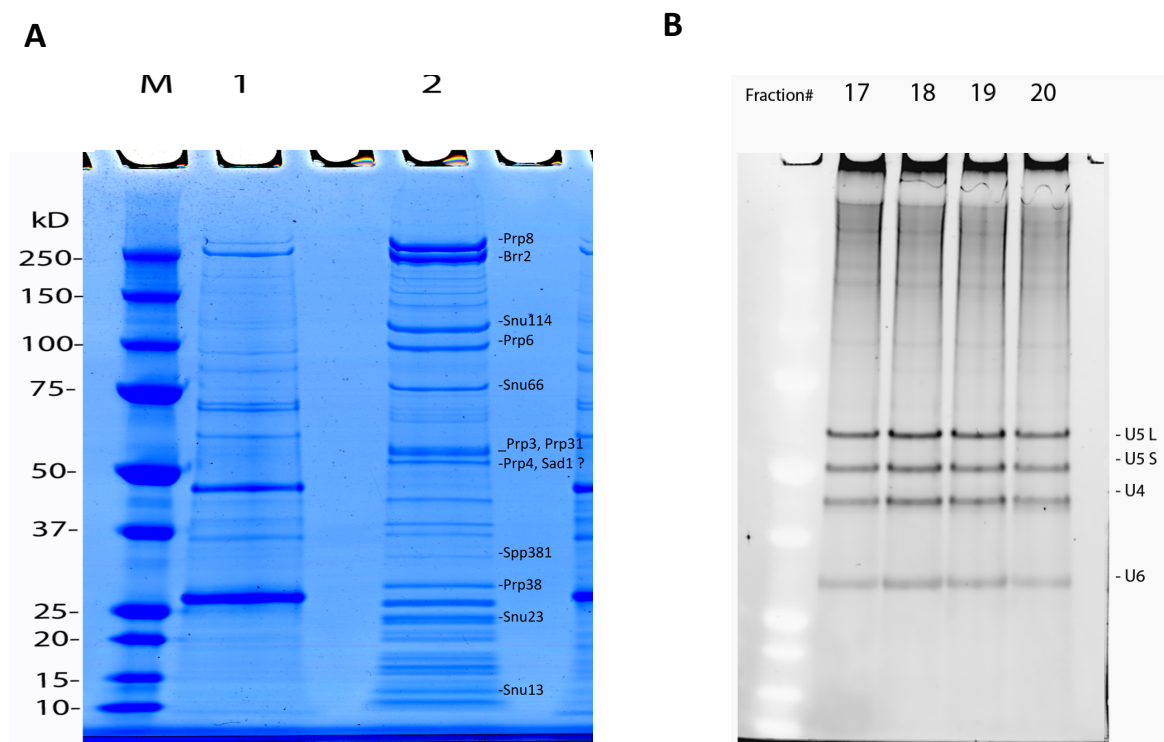


**Figure 3.1 Overview of TAP purification methods.** In the first step, complexes containing TAP tagged proteins are bound to IgG-Sepharose beads. Contaminating non-complex proteins are removed during washes, TEV protease treatment (which cleaves at the TEV protease cleavage site sequence situated between the ProtA and CBP) releases the complexes from the column. In the final step, eluted complexes are subjected to gradient centrifugation and fractions of the purified tri-snRNPs are collected.

The RNA content of collected fractions was analysed by SYBR Gold staining, while their protein components were separated by SDS gel electrophoresis and then subjected to mass-spectrometry (MS) analysis. The snRNA analysis revealed similar amounts of U5L, U5S, U4 and U6 snRNAs in the tri-snRNP fractions 18-21 (fig. 3.2). The analysis of the corresponding lanes by SDS-PAGE and MS (fig. 3.3 A, lane 2) revealed that the tri-snRNPs contained the typical tri-snRNP-associated proteins (Gottschalk et al., 1999; Stevens and Abelson, 1999; Stevens et al., 2001), however the relatively low peptide numbers of Snu23p, Spp381p and Sad1p found during MS analysis indicated that these proteins were underrepresented in the tri-snRNP preparation (Table: 3.1).



**Figure 3.2 Denaturing gel showing RNA analyses of all gradient fractions of the purified Brr2-TAP tri-snRNP.** Migration pattern of the purified tri-snRNP on the glycerol gradient shows the presence of U5L, U5S, U4, and U6 snRNA in fractions # 18, 19, 20, and 21. M, Molecular weight markers. In, sample input before gradient.



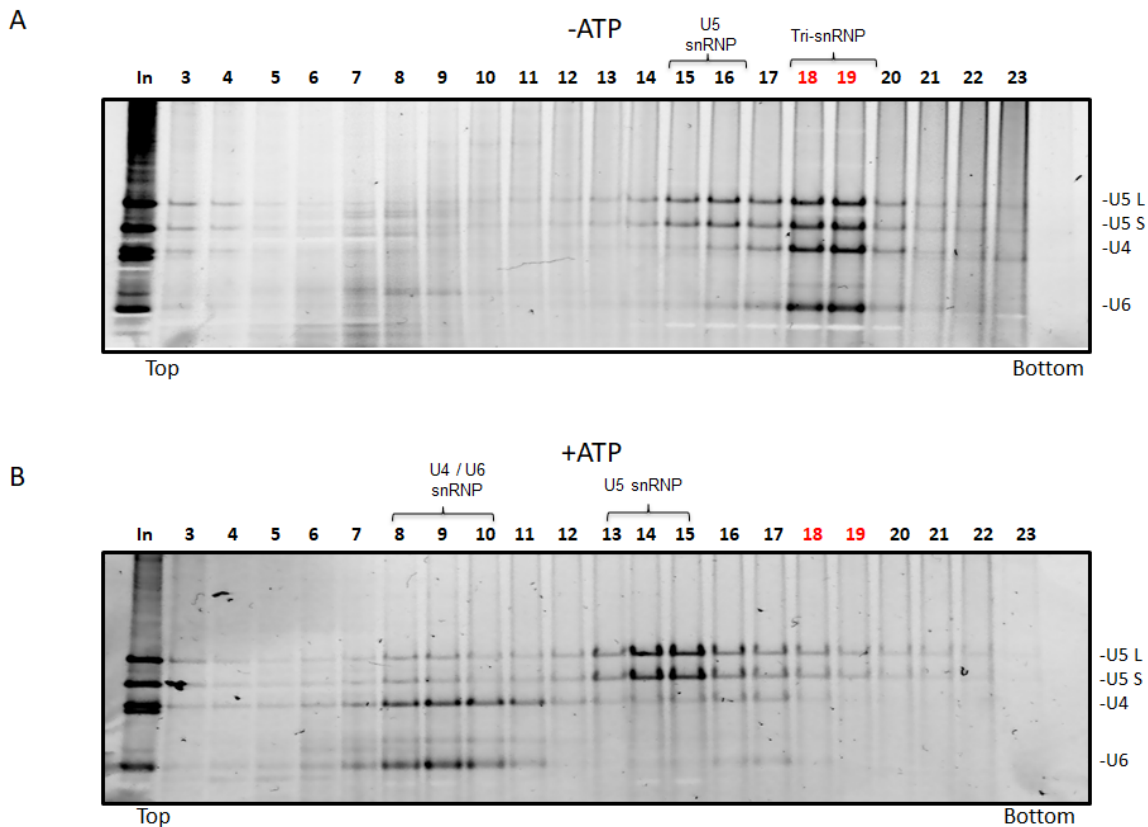
**Figure 3.3 Protein and RNA gels of purified Brr2-TAP tri-snRNP.** Affinity purified Brr2-TAP tri-snRNPs at 150mM  $K^+$  salt conditions were further separated on a 10-30% glycerol gradient keeping the salt concentration at 150mM. Equal fractions were collected from bottom to top and were analysed for protein and RNA by separation on a 4-15% MOPS-SDS precast gel, and stained using sensitive fluorescent nucleic acid stain (SYBR® Gold stain). **A:** Protein profile of the pooled fraction from the collected gradient fractions showing in lane 1 combined fractions # 2 and 3, in lane 2 pooled fractions from the tri-snRNPs peak (fraction# 17, 18 and 19). M- Molecular weight marker. **B:** RNA analysis of fractions 17 - 20 representing the tri-snRNP peak showing U5L, U5S, U4 and U6 snRNA on the gel.

### 3.1.2 Brr2-mediated ATP sensitivity assay of purified Brr2p-TAP yeast tri-snRNPs under 150 mM KCl (high salt) conditions

Experiments were performed in order to test whether the purified tri-snRNP is resistant to ATP-dependent dissociation. There are reports, which shows that the yeast tri-snRNP, in the Sad1 depleted whole cell extract, when incubated with ATP concentrations above 1 mM is rapidly dissociated (Huang et al., 2014). However, these experiments were performed using the whole cell yeast extracts and not the purified tri-snRNPs. In the published study, Huang et al. concluded that the presence of the yeast Sad1p protein improves the formation of a stable tri-snRNP, probably by counteracting the ATP-dependent Brr2p-mediated dissociation of the tri-snRNP. As Sad1 was clearly underrepresented in my tri-snRNP preparations, shown in fig. 3.3, I investigated

whether the tri-snRNPs remain intact in the presence of ATP or dissociate. Thus, a 350  $\mu$ l sample of this purified tri-snRNP was incubated for 30 min at 25°C in the absence or presence of 2 mM ATP and 2 mM MgCl<sub>2</sub>. The reaction mixture was then loaded on a 10 – 30 % glycerol gradient and fractionated (2 x 4.2 ml gradient, 29 000 rpm, type TH660 rotor, 15 h). The gradient buffers also contained 150 mM KCl. All fractions were subjected to RNA and protein analysis. Fig. 3.4 shows RNA analysis, MS analysis of Brr2-TAP tri-snRNPs incubated with/without 2mM ATP is shown in Table 3.1.

RNA analyses for the tri-snRNP incubated without ATP indicate that the tri-snRNP remained intact (Fig. 3.4 A). The tri-snRNP proteins also co-migrated along the respective snRNAs in gradient fractions 17/18 (column B in Table 3.1). Thus, in the absence of ATP tri-snRNPs are stable. The low amounts of Sad1p in the purified tri-snRNPs are displaced from the snRNP during the gradient centrifugation step (see column A in Table 3.1) and migrate in the top fractions of the gradient.



**Figure 3.4 ATP-dependent, Brr2p-mediated dissociation assay of yeast tri-snRNP.**

Approx. 102  $\mu\text{g}$  purified Brr2-TAP tri-snRNP in a total volume of 450  $\mu\text{l}$  were incubated with 2 mM  $\text{Mg}^{2+}$  in the absence (gel A), and presence of 2mM ATP (gel B). Gel **A**: tri-snRNPs incubated without ATP show intact tri-snRNP in fractions 18 and 19 (equimolar amounts of U5L, U5S, U4, and U6 snRNA). Gel **B**: tri-snRNPs incubated in the presence of 2 mM ATP showing dissociated U4 and U6 snRNA in fractions 8, 9, and 10 and the U5L and U5S snRNA in fractions 13, 14, and 15. In- sample input before gradient.

In contrast, when the isolated tri-snRNP is incubated in the presence of 2 mM ATP the RNA analysis after gradient centrifugation (Fig. 3.4 B) clearly shows pronounced dissociation of tri-snRNPs. That is, now the majority of U4 and U6 snRNAs now co-migrate in fractions 8, 9, and 10 (indicating that both snRNAs interact) and free 18S U5 snRNPs migrate in fraction 13 to 15. (Fig. 3.4, B)

The MS analysis (Table 3.1) across the gradient indicate that the majority of U5 proteins co-migrate with U5 snRNA at about 18S (see fraction 15) while the majority of typical U4/U6 proteins co-migrate with U4/U6 RNAs (see fraction 9), suggesting that the tri-snRNP dissociates in an ATP-dependent and very likely Brr2 helicase-dependent manner into U4/U6 di-snRNPs



and U5 snRNPs. A similar behaviour was observed in a previous study (Huang et al., 2014), which was performed using the whole cell yeast extract.

It is not clear, whether Brr2 dissociates intermittently U4 and U6 followed by fast re-association, or whether the tri-snRNP dissociates in an aberrant manner. In view of the latter possibility, it is interesting to note that Brr2 appears to dissociate in significant amounts from U5 snRNPs and relocates together with U4/U6 di-snRNPs. Tri-snRNP proteins such as Sad1 and Prp38 are also found in the U4/U6 di-snRNP and top fractions.

Table 3.1 No ATP 2 mM ATP

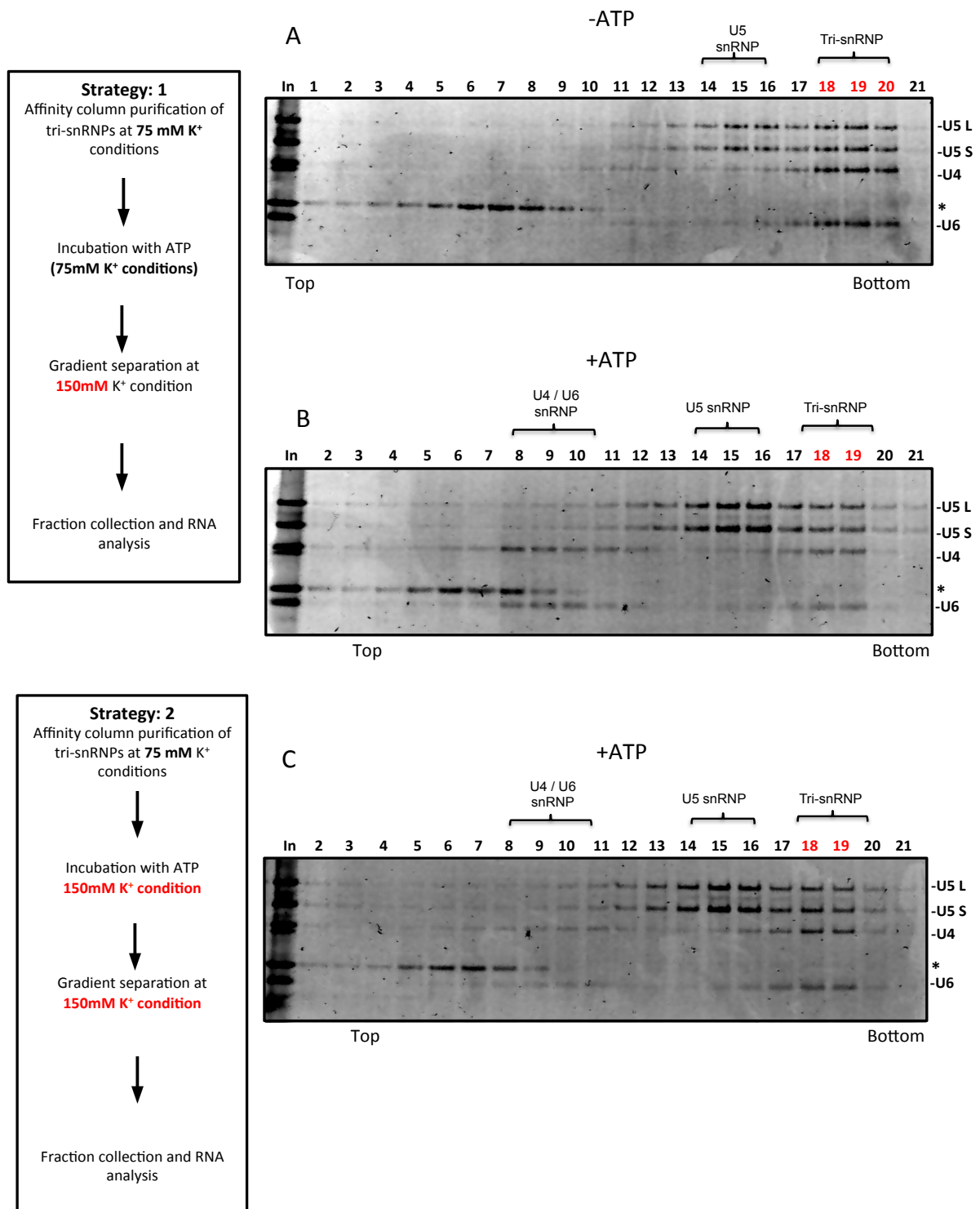
Protein	MW	A	B	C	D	E
Gradient Fractions		#6	#17	#6	#9	#15
<b>Sm proteins</b>						
B	22.4	4	37	9	30	28
D1	16.3	3	10	2	6	5
D2	12.8	4	24	9	16	14
D3	11.2	3	29	5	13	19
E	10.4	-	-	-	-	-
F	9.6	-	-	-	-	-
G	8.5	1	7	2	2	3
<b>U5 snRNP proteins</b>						
Prp8	279.5	60	497	34	66	410
Brr2	246.2	196	488	127	461	375
Snu114	114	28	202	12	21	148
Prp6	104.2	14	138	40	32	81
Prp28	66.6	39	4	17	17	67
Lin1	40.4	7	8	2	2	36
Dib1	16.7	4	22	13	8	16
<b>U4/U6 snRNP proteins</b>						
Prp31	56.3	1	55	19	34	11
Prp3	56	3	63	16	62	11
Prp4	52.4	5	47	20	52	11
Snu13	13.6	1	4	4	5	3
<b>U4/U6.U5 snRNP proteins</b>						
Snu66	66.4	10	75	10	27	30
<b>Sad1</b>	<b>52.2</b>	<b>18</b>	<b>0</b>	<b>12</b>	<b>17</b>	<b>3</b>
Spp381	34	0	4	6	0	1
<b>Prp38</b>	<b>28</b>		<b>10</b>	<b>11</b>	<b>1</b>	<b>5</b>
Snu23	23	0	8	21	2	7
<b>LSm proteins</b>						
LSm2	11.2	1	23	7	24	5
LSm3	10	-	-	-	-	-
LSm4	21.3	2	8	6	9	1
LSm5	10.4	0	1	2	4	3
LSm6	9.4	0	6	3	6	2
LSm7	13	3	7	4	12	2
LSm8	12.4	0	11	5	9	1

**Table 3.1 MS analysis of yeast Brr2-TAP tri-snRNPs isolated in the presence of 150 mM K<sup>+</sup> incubated in the absence or presence of 2 mM ATP.** Columns A and B show the peptide counts in fractions 6 and 17 of the gradient in Fig. 3.4 A (no ATP), columns C to E the peptide counts in fractions 6, 9 and 15 of the gradient in Fig. 3.4 B (2 mM ATP added).

### **3.1.3 Purification and ATP-sensitivity assay of Brr2-TAP tri-snRNPs under 75 mM KCl conditions**

As outlined in the introduction it was our goal to screen for conditions where Sad1p remains stably bound to the isolated tri-snRNP to gain insight into a possible role of Sad1p in modulating tri-snRNP stability in the presence of ATP. Thus, in a modified protocol, the  $K^+$  concentration of buffers used during affinity purification was lowered to 75 mM KCl. Keeping the KCl concentration at 75 mM, the purified Brr2-TAP tri-snRNPs were then incubated with or without ATP and the sample was run on a 10 – 30% glycerol gradient (at 150 mM  $K^+$ ) with subsequent RNA and protein (MS) analysis of the fractions (Fig. 3.5 A, B).

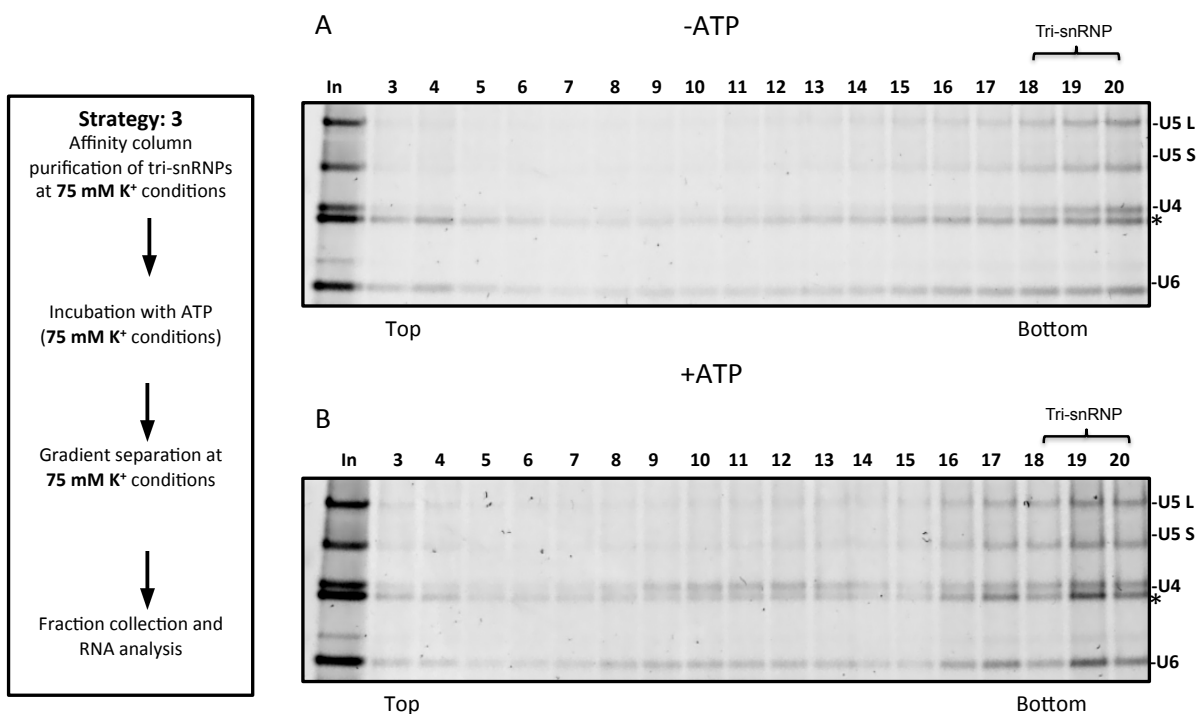
In a separate (strategy 2) approach, the tri-snRNP was first isolated at 75 mM KCl conditions, readjusted to 150 mM  $K^+$  prior to 2 mM ATP incubation and the gradient run at 150 mM  $K^+$  (Fig. 3.5 C).



**Fig. 3.5 RNA analysis for Brr2-TAP tri-snRNP, strategies 1 and 2.** Panels **A**, **B**: Brr2-TAP tri-snRNP, affinity-purified at 75 mM K<sup>+</sup> and at 75 mM K<sup>+</sup> incubated with 2 mM Mg<sup>2+</sup> in the absence (panel A) or presence (panel B) of 2 mM ATP. Gradient at 150 mM K<sup>+</sup>. Panel **C**: tri-snRNP affinity-purified at 75 mM K<sup>+</sup>, adjusted to 150 mM K<sup>+</sup> and incubated with 2 mM Mg<sup>2+</sup> and 2 mM ATP. Gradient at 150 mM K<sup>+</sup>. Asterisk \* indicates a degradation product of U4 snRNA lacking the 3' end (Anthony et al., 1997).

Panels B and C of Figure 3.5 clearly show that with low-salt purified tri-snRNPs (75 mM K<sup>+</sup>) a substantially higher amount of apparently intact tri-snRNP is observed in the tri-snRNP peak fractions 18, 19 upon incubation with 2 mM ATP when compared to tri-snRNP purified at 150 mM K<sup>+</sup> (Fig. 3.4 B). This was also true when the low-salt purified tri-snRNPs were incubated with ATP at 150mM KCl conditions (fig: 3.5, panel C)

Finally, in a third variant, the tri-snRNP was isolated, incubated with 2 mM ATP and gradient fractionated always at 75 mM K<sup>+</sup>. RNA and protein of collected fractions were analysed with RNA gels and MS. RNA analysis is depicted in Fig. 3.6.



**Fig. 3.6 RNA analysis of ATP-sensitivity assay for Brr2-TAP tri-snRNP under 75 mM K<sup>+</sup> conditions.** Purification, ATP incubation, and gradient separation were performed at 75 mM K<sup>+</sup> conditions. Panel A shows the RNA profile under -ATP conditions, panel B the profile when incubating with 2 mM. Tri-snRNP is observed in fractions 18 – 20. In- sample input before gradient. Asterisk \* indicates a degradation product of U4 snRNA lacking the 3' end (Anthony et al., 1997).

Comparing panels A and B in Fig. 3.6 indicates that the Brr2-TAP tri-snRNPs which were purified and incubated with ATP under low salt conditions (75 mM KCl) are more resistant to ATP-dependent dissociation. Furthermore, the MS analysis shows that the tri-snRNPs, which

were incubated with or without ATP, revealed a similar protein composition with comparable peptide numbers for all tri-snRNP proteins (Table 3.2).

extract/sample		no ATP	2mM ATP
Protein	MW	Brr2-TAP Tri-snRNPs Fraction 19 (Peak)	Brr2-TAP Tri-snRNPs + ATP Fraction 19 (Peak)
<b>Sm proteins</b>			
B	22.4	21	26
D1	16.3	7	10
D2	12.8	45	51
D3	11.2	48	38
E	10.4	3	3
F	9.6	4	3
G	8.5	2	3
<b>U5 snRNP proteins</b>			
Prp8	279.5	571	487
Brr2	246.2	798	692
Snu114	114	330	218
Prp6	104.2	241	203
Prp28	66.6	114	93
Lin1	40.4	10	23
Dib1	16.7	28	30
<b>U4/U6 snRNP proteins</b>			
Prp31	56.3	67	57
Prp3	56	38	36
Prp4	52.4	102	67
Snu13	13.6	9	5
<b>U4/U6.U5 snRNP proteins</b>			
Snu66	66.4	158	124
<b>Sad1</b>	52.2	<b>31</b>	<b>80</b>
Spp381	34	12	10
<b>Prp38</b>	28	18	12
Snu23	23	22	16
<b>LSm proteins</b>			
LSm2	11.2	23	15
LSm3	10		
LSm4	21.3	14	14
LSm5	10.4	6	3
LSm6	9.4	8	5
LSm7	13	5	4
LSm8	12.4	12	7

**Table 3.2 Mass-spectrometry analysis for Brr2-TAP tri-snRNPs purified, ATP treated and glycerol-gradient fractionated under 75 mM K<sup>+</sup> conditions.** Tri-snRNP was purified at 75 mM K<sup>+</sup>, incubated (30 min at 25°C) with or without 2 mM ATP, and subjected to glycerol gradient (10 – 30%, GK75mM buffer) centrifugation at the same salt concentration. The left column (in grey) shows peptide counts of all tri-snRNP proteins when incubated without ATP, the right (column shown in orange) when incubated with 2 mM ATP (in both cases peak fraction 19 was used for MS analysis).

It is remarkable that peptide counts for all proteins of the so-called yeast tri-snRNP-specific proteins are considerably higher in both samples (with and without ATP treatment) when compared to purifications of the tri-snRNP at 150 mM KCl conditions (compare Table 3.1

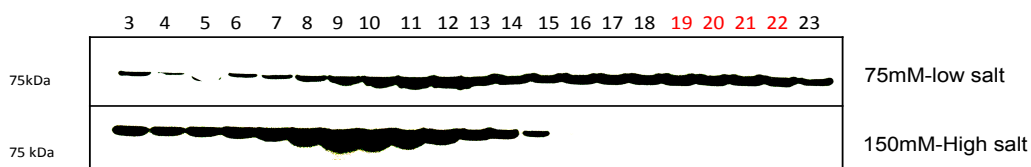
columns B and E with Table 3.2). This is particularly remarkable for Sad1p, which is significantly more abundant compared to the purifications at 150 mM K<sup>+</sup> conditions. Tri-snRNPs isolated at low salt conditions are clearly more resistant to ATP-dependent dissociation (Fig. 3.6), and this appears to correlate with the abundance of Sad1 in the purified particles.

### 3.1.4.1 Purification of tri-snRNPs using a TAP-tagged Sad1p strain.

As Sad1 is not very stably bound to the yeast tri-snRNP, the purification strategy used so far, i.e., via a TAP tagged Brr2 protein, appears to be sub-optimal, as these tri-snRNP preparations may always contain a considerable fraction of particles which lack Sad1.

Therefore, I considered using a yeast strain, which expressed a TAP-tagged Sad1 protein instead of TAP-tagged Brr2. The Sad1-TAP strain was obtained from Euroscarf. Initially, I investigated whether the presence of a TAP tag at the C-terminus of SAD1 would prohibit the association of the protein with tri-snRNPs. Therefore, I investigated the migration behaviour of Sad1 during gradient centrifugation of yeast whole cell splicing extract at 75 or 150 mM KCl concentrations.

After gradient centrifugation and fraction collection, Western Blotting with an anti-TAP antibody was performed to visualize the Sad1 distribution in the gradients (Fig. 3.7.1). The upper graph in Fig. 3.7.1 shows that at 75mM salt, a major fraction of Sad1p migrates in the 25S region of the gradient (fractions 16 to 21), suggesting an association with tri-snRNP particles. To account for the amount of Sad1p in fractions at the top of the gradient one needs to keep in mind that whole cell extract was fractionated, not purified tri-snRNP. Under 150 mM K<sup>+</sup> conditions (lower graph of Fig. 3.7.1), however, most of Sad1p migrates in the top of the gradient (fractions 8 to 11), indicating loss of Sad1p from intact tri-snRNPs at 150 mM, a salt concentration which does not allow formation of a stable tri-snRNP.

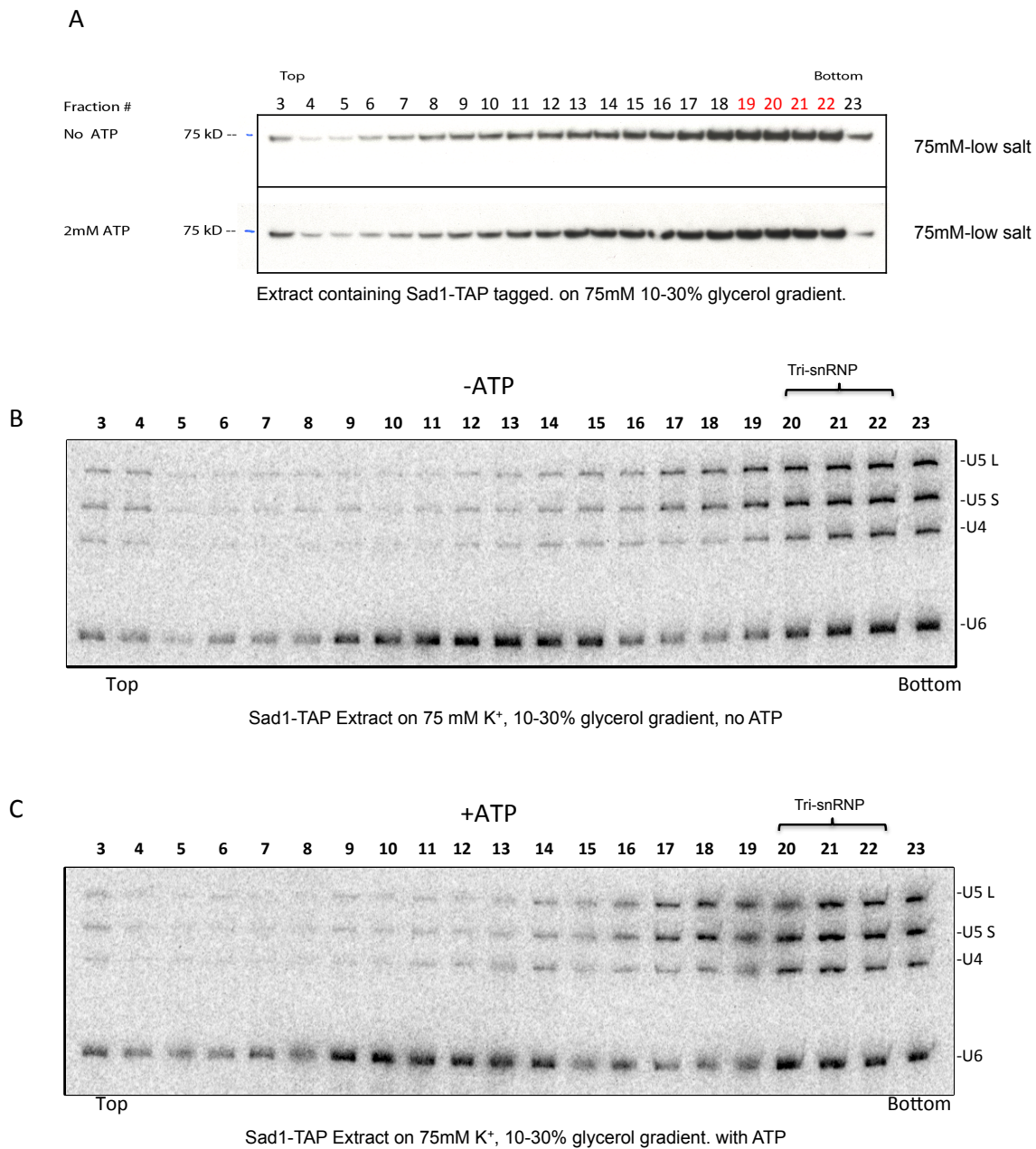


Extract containing Sad1-TAP tagged. 10-30% glycerol gradient.

**Figure 3.7.1 Distribution of Sad1-TAP protein from yeast whole cell extract in glycerol gradients**  
Western blot of Sad1-TAP protein under 75 mM (upper graph) and 150 mM K<sup>+</sup> (lower graph).

Next, I asked whether at 75 mM salt conditions, the tri-snRNPs in the splicing extract derived from the Sad1-TAP strain would remain intact upon incubation with 2mM ATP. Hence, I performed a quick experiment using glycerol gradient centrifugation for the fractionation of the extract and tri-snRNP components. (Fig: 3.7.2). Panel A of fig: 3.7.2 shows that the majority of TAP-Sad1 migrates in fractions 17 to 22 (which contains the tri-snRNPs, see panel below) with or without prior ATP incubation.

Importantly, northern blot analysis showed that the tri-snRNPs remained largely intact upon incubation of the splicing extract with ATP (compare panel B and C in fig: 3.7.2). This indicates that the TAP-Sad1 remains associated with tri-snRNPs in the splicing extract at 75 mM KCl, likely inhibiting dissociation of the tri-snRNPs in the presence of ATP.



### Figure 3.7.2 Distribution of snRNAs from yeast whole cell extract in glycerol gradients

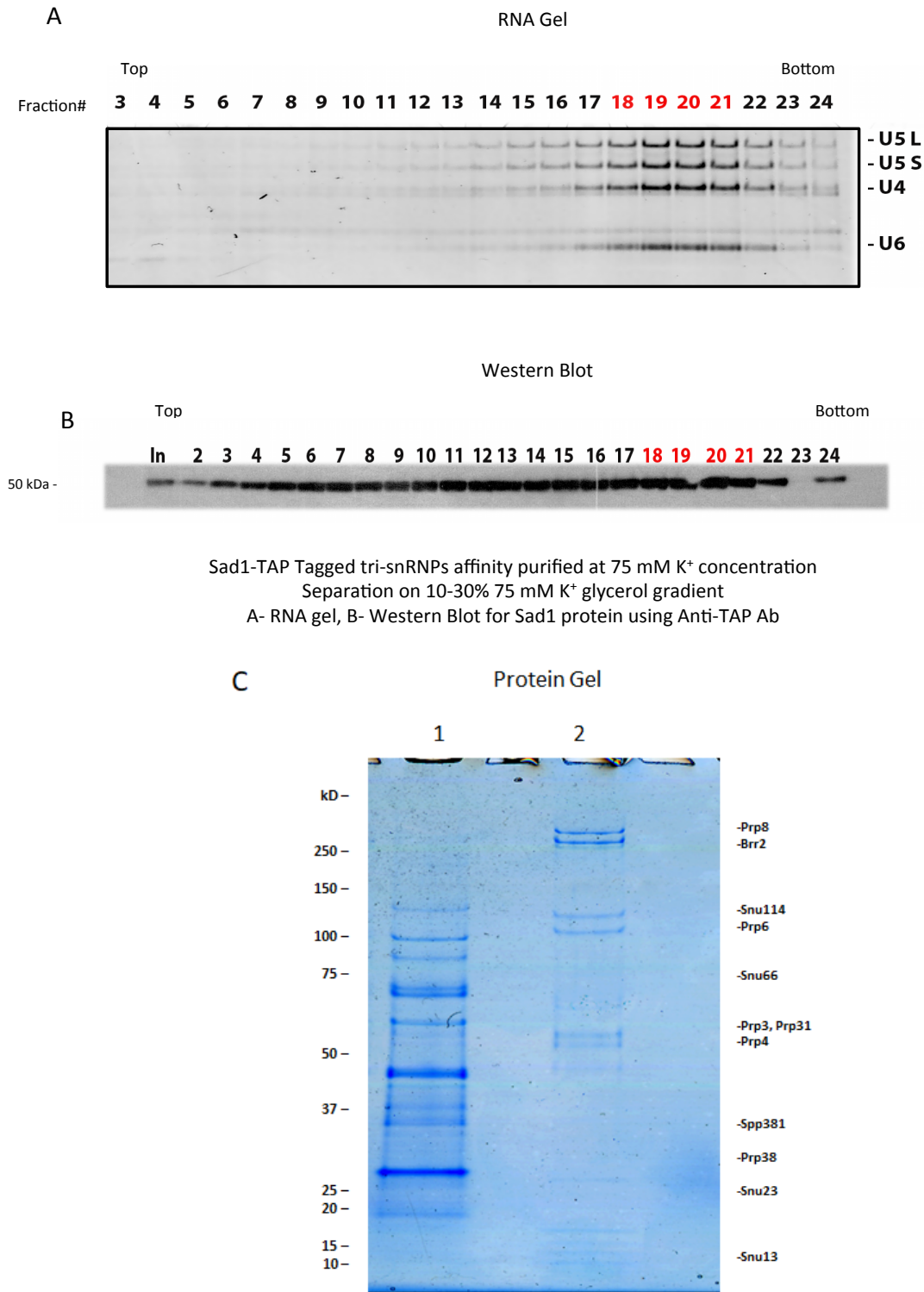
**Panel A:** Western blot of Sad1-TAP protein after 2 mM ATP treatment (upper panel without, lower panel with ATP). Both gradients with 75 mM K<sup>+</sup>. **Panel B and C:** Northern blot visualization of snRNA distribution in a (10 – 30%) glycerol gradient at 75 mM K<sup>+</sup>. Sad1-TAP yeast whole cell extract was loaded on a (10 – 30%) glycerol gradient containing 75 mM K<sup>+</sup> without (Panel B) or with (Panel C) prior incubation with 2 mM ATP (25°C, 30 min). The prominent band in the panel B and C in fractions 9 to 15 is likely free U6 snRNAs, which is commonly in excess in yeast.



#### **3.1.4.2 Purification of the Sad1-TAP tri-snRNP at 75 mM K<sup>+</sup> conditions**

Experiments described in sections above indicate that the TAP-Sad1 yeast strain should be promising for the purification of tri-snRNPs that may contain stoichiometric amounts of Sad1. First, I affinity purified the tri-snRNPs at 75 mM K<sup>+</sup> concentration and fractionated them by glycerol gradient centrifugation.

Yeast whole cell extract preparation, affinity chromatography on IgG-Sepharose, and glycerol-gradient separation were performed following the standard protocol at 75 mM KCl conditions. After high-speed centrifugation, gradient fractions were collected and subjected to RNA and protein analysis (Fig. 3.8 and Table 3.3). The quality of the observed snRNA and protein patterns was excellent and without any degradation products (Fig 3.8 A, C).



**Figure 3.8 Protein and RNA analyses of purified Sad1-TAP tri-snRNPs purified at 75 mM buffer conditions.** Sad1-TAP tri-snRNPs, affinity-purified at 75 mM K<sup>+</sup> conditions, separated on a 10-30% glycerol gradient with K<sup>+</sup> concentration at 75 mM. Fractions were collected from bottom to top and protein and RNA were analysed by separating on a 4-12% MOPS-SDS precast denaturing gel. **A:** RNA profile of the complete gradient. Fractions 18-21 represent the tri-snRNP peak and display the characteristic snRNAs. **B:** Western blot of gradient fractions with anti-TAP antibodies visualizing the

distribution of Sad1-TAP protein in the gradient. **C:** Coomassie stained protein gel showing, in lane 1 fraction 4 from top/lighter part of the gradient, and in lane 2 fraction 19 from the tri-snRNP peak.

RNA gel of the gradient fractions (Fig. 3.8 A) shows the snRNAs of the intact, clean and degradation-free tri-snRNPs in fractions 18 to 21.

After checking the snRNAs of the tri-snRNP peak fractions, I decided to examine the distribution of the Sad1p in the gradient fractions; hence, I further performed a western blot (fig: 3.8 B) using anti-TAP antibodies – with affinity to the TAP tag fused to Sad1p.

Western blot analysis shows that a substantial amount of Sad1p is present in the fractions associated with the tri-snRNP particles (i.e., fraction 18 to 21). Considerable amounts of Sad1 smearing to the top gradient fractions are due to the fact that also free Sad1 is purified from the extract.

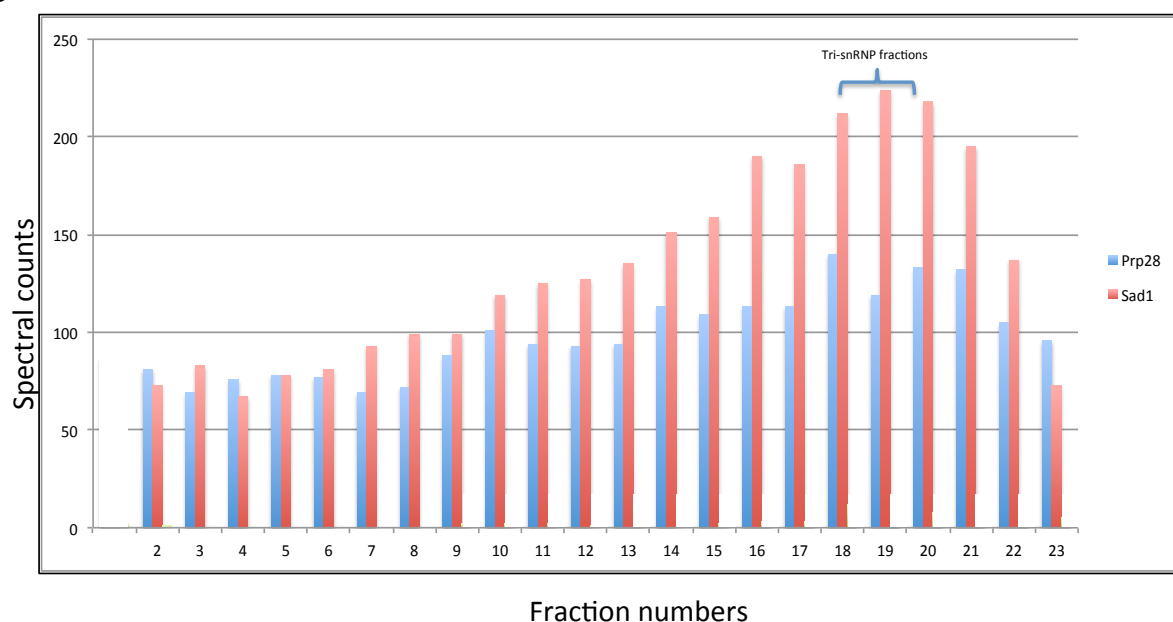
For a more detailed protein analysis we used MS of the gradient fractions. A total amount of 400 µg of affinity purified tri-snRNPs were loaded on a 10-30% GK 75 mM buffer, glycerol gradient. After density gradient separation, fractions were collected and all the fractions were subjected to MS analyses. (See table 3.3).

A

Fraction #	M.W	10-30% glycerol gradient (75mM K <sup>+</sup> )																		Tri-snRNPs		Bottom	
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21		22
<b>US snRNP proteins</b>																							
Prp8	279.5	210	142	126	124	168	185	163	156	192	215	239	253	326	384	451	477	530	738	760	843	543	400
Brr2	246.2	197	169	147	144	146	143	137	162	197	206	220	269	304	385	453	483	548	680	722	776	586	379
Snu114	114	119	95	94	86	91	83	88	97	114	120	115	153	193	208	293	302	371	411	538	532	439	237
Prp6	104.2	97	72	56	66	77	81	80	102	113	100	130	137	175	203	229	285	324	398	447	444	329	190
<b>Prp28</b>	<b>66.6</b>	<b>81</b>	<b>69</b>	<b>76</b>	<b>78</b>	<b>77</b>	<b>69</b>	<b>72</b>	<b>88</b>	<b>101</b>	<b>94</b>	<b>93</b>	<b>94</b>	<b>113</b>	<b>109</b>	<b>113</b>	<b>113</b>	<b>140</b>	<b>119</b>	<b>133</b>	<b>132</b>	<b>105</b>	<b>96</b>
Lin1	40.4	23	22	33	44	26	21	18	13	15	11	9	18	25	27	31	27	21	11	9	9	16	12
Dib1	16.7	0	0	0	0	0	1	4	2	3	0	3	1	5	5	5	6	10	9	9	11	9	5
<b>U4/U6 snRNP proteins</b>																							
Prp31	56.3	40	34	16	22	21	21	31	36	33	40	48	47	55	57	67	91	108	155	163	153	140	79
Prp3	56	43	51	34	38	29	29	39	34	44	44	52	66	69	69	80	112	140	182	198	193	146	79
Prp4	52.4	58	48	41	37	36	32	29	37	43	51	58	62	64	69	77	105	130	191	175	192	162	87
Snu13	13.6	5	4	3	4	4	4	4	5	6	6	6	7	7	7	6	7	11	9	13	11	9	4
<b>U4/U6.U5 snRNP proteins</b>																							
Snu66	66.4	52	53	38	39	34	30	37	42	40	37	43	43	49	50	55	64	92	84	91	101	88	67
<b>Sad1</b>	<b>52.2</b>	<b>73</b>	<b>83</b>	<b>67</b>	<b>78</b>	<b>81</b>	<b>93</b>	<b>99</b>	<b>99</b>	<b>119</b>	<b>125</b>	<b>127</b>	<b>135</b>	<b>151</b>	<b>159</b>	<b>190</b>	<b>186</b>	<b>212</b>	<b>224</b>	<b>218</b>	<b>195</b>	<b>137</b>	<b>73</b>
Spp381	34												0				2	3	2	3	1	1	0
Prp38	28	1	0	0	0	0			2	2	2	1	3	1	2	3	3	12	11	8	5	4	5
Snu23	23													1	0	2	4	5	4	5	4	2	1

Sad1-TAP Tagged tri-snRNPs affinity purified at 75mM K<sup>+</sup> concentration  
Separation on 10-30% 75mM K<sup>+</sup> glycerol gradient

B



**Table 3.3 MS analysis of glycerol gradient fractions from the Sad1-TAP tri-snRNP purification under 75 mM K<sup>+</sup> conditions. Panel A:** MS data of the protein content of fractions collected from the 10-30% glycerol gradient. **Panel B:** Graphical representation of MS peptide counts for Sad1p (red) and Prp28p (blue). About 400  $\mu$ g of tri-snRNP (as determined in the affinity column eluate) were loaded on the gradient.

Table: 3.3-A shows the protein composition of the collected gradient fractions. MS data shows that all U5 snRNP proteins, under the 75 mM buffer conditions, are present in the peak fractions with a higher percentage of the total peptide count than in the rest of the gradient. Notably, Prp28p is somewhat trailing throughout the gradient but shows a clear peak in the tri-snRNP fraction (table: 3.3, panel B). U4/U6 snRNP proteins can be seen confined to the tri-snRNP peak under 75 mM K<sup>+</sup> conditions. Within the group of tri-snRNP specific proteins, the Sad1p is primarily associated with tri-snRNP. Thus, MS analysis further proves that using the TAP tagged Sad1p is the method of choice for purifying tri-snRNPs, yielding reasonable quantities of tri-snRNPs with a stably attached set of all canonical tri-snRNP proteins.

### **3.1.4.3 ATP-sensitivity assay for Sad1-TAP tri-snRNPs, purified under 75 mM KCl conditions**

Similar experiment, as described in sections 3.1.3, which show the results for the ATP-dependent dissociation assays of purified Brr2-TAP tri-snRNPs, were performed using the Sad1-TAP tri-snRNPs purified under 75 mM K<sup>+</sup> concentration. Experiments mentioned in section 3.1.4.1 above have already indicated that the yeast Sad1-TAP tri-snRNPs remained largely intact upon incubation of the splicing extract with ATP, however, these experiments were carried out using the whole cell yeast extract, and not the purified tri-snRNP particles. Therefore, I decided to perform the ATP-sensitivity assay on the Sad1-TAP tri-snRNPs, purified at 75 mM KCl conditions.

In a first experiment, affinity purified Sad1-TAP tri-snRNPs under 75 mM K<sup>+</sup>, were incubated with 2 mM ATP at 25°C (30 min) and then separated on a 10-30% glycerol gradient containing 75 mM K<sup>+</sup> (Fig. 3.9). RNA and protein composition of the collected fractions was analysed using denaturing RNA gel, and MS, respectively.

RNA analysis of the gradient fractions shows (Fig. 3.9 A, B) that the Sad1-TAP tri-snRNPs did not dissociate during ATP incubation. A tri-snRNP peak fraction of the glycerol gradient was then subjected to MS analysis (Table 3.4).



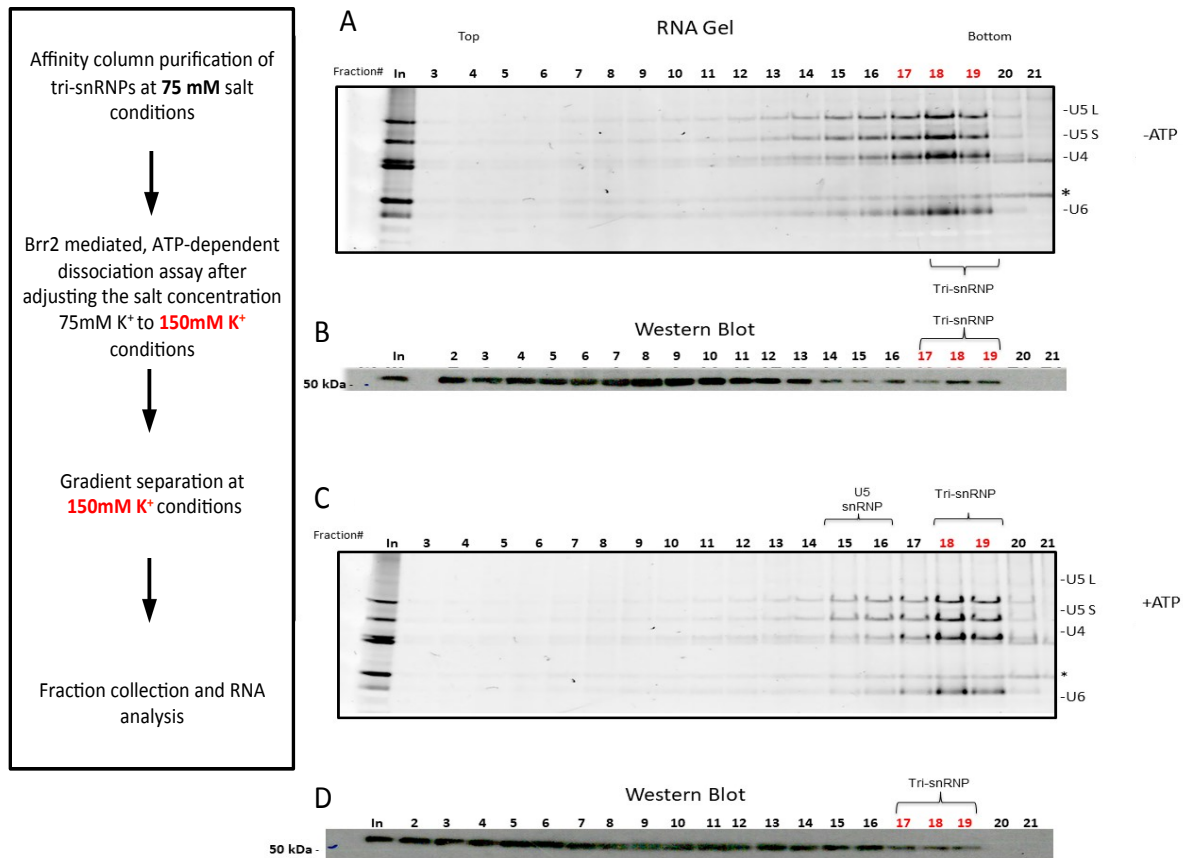
Extract used		-ATP	+ATP
extract/sample		-ATP	+ATP
Gradient fraction		Peak	Peak
Protein	MW	Fraction #18	Fraction #18
<b>Sm proteins</b>			
B	22.4	15	18
D1	16.3	4	6
D2	12.8	22	31
D3	11.2	21	22
E	10.4	3	3
F	9.6	2	3
G	8.5	4	4
<b>U5 snRNP proteins</b>			
Prp8	279.5	268	340
Brr2	246.2	339	400
Snu114	114	151	193
Prp6	104.2	99	115
Prp28	66.6	46	57
Lin1	40.4	7	6
Dib1	16.7	12	19
<b>U4/U6 snRNP proteins</b>			
Prp31	56.3	39	53
Prp3	56	39	46
Prp4	52.4	34	49
Snu13	13.6	6	8
<b>U4/U6.U5 snRNP proteins</b>			
Snu66	66.4	61	73
Sad1	52.2	35	39
Spp381	34	3	3
Prp38	28	6	3
Snu23	23	9	7
<b>LSm proteins</b>			
LSm2	11.2	12	13
LSm3	10		
LSm4	21.3	10	13
LSm5	10.4	5	5
LSm6	9.4	7	7
LSm7	13	4	7
LSm8	12.4	5	5

**Table: 3.4** Mass-spectrometry data for peak gradient fractions of Sad1-TAP tri-snRNPs incubated with or without 2 mM ATP.

Based on these results it was evident that Sad1-TAP tri- snRNPs purified under low salt conditions, not only stably retained TAP-tagged Sad1p, but also do not dissociate when incubated in the presence of ATP.

Next, we asked, whether ATP incubation under comparatively high salt conditions might have an influence on the ATP sensitivity of the Sad1-TAP tri-snRNPs. Therefore, in a second variant of the ATP-sensitivity assay, the Sad1-TAP tri-snRNPs purified at 75 mM K<sup>+</sup> were incubated with 2

mM ATP (25° C 30 min) at 150 mM K<sup>+</sup> and subsequently separated on a glycerol gradient containing 150 mM K<sup>+</sup>. The results are shown in Fig. 3.10.



**Figure 3.10 ATP-sensitivity assay of purified Sad1-TAP tri-snRNPs in the absence or presence of ATP.** (Side panel) Sad1-TAP tri-snRNPs purified at 75mM KCl conditions, after ATP dissociation at 150 mM K<sup>+</sup>, were separated on a 10-30% glycerol gradient at 150 mM K<sup>+</sup>. Fractions were collected from bottom to top, and the RNA was analysed by separating on a 4-12% MOPS-SDS precast denaturing gel and staining with fluorescent nucleic acid stain (SYBR<sup>®</sup> Gold stain). **A:** RNA gel showing the snRNA profile of gradient fractions after mock incubation without 2 mM ATP. **B:** Western blot analysis of a gradient shown in B with antibody against the TAP-tag. **C:** snRNA distribution of gradient fractions after incubation with 2 mM ATP (25°C 30 min). **D:** Western blot analysis of a gradient shown in C with antibody against the TAP-tag. Asterisk \* indicates a degradation product of U4 snRNA lacking the 3' end (Anthony et al., 1997). 'In'- Input before gradient.

While the snRNA profiles (Fig. 3.10, Panel A and C) do not show dissociation of the tri-snRNP upon incubation with or without 2 mM ATP, the Western blot analyses of the respective gradients with an antibody against the TAP-tag show that Sad1 is partially released from the tri-



snRNP under the comparatively harsh 150 mM K<sup>+</sup> conditions of the glycerol gradient (fig: 3.10, Panel B and D).

In summary, it can be stated that the Sad1-TAP tri-snRNPs purified under 75 mM K<sup>+</sup> conditions are highly stable when incubated under dissociative conditions with 2 mM ATP.

#### **3.1.4.4 Sad1-TAP tri-snRNP purification at high salt conditions (150mM)**

After examining the purified tri-snRNP from a Sad1-TAP strain at 75 mM K<sup>+</sup> conditions we performed a purification using the 150 mM K<sup>+</sup> conditions in order to obtain reliable information how the Sad1-TAP strain by itself might influence the characteristics of the tri-snRNP upon altering the salt conditions. Hence, tri-snRNPs were affinity purified from the whole cell extract of the Sad1-TAP yeast strain under 150 mM K<sup>+</sup> conditions. Later steps and glycerol gradients also contained GK150mM buffer (150 mM K<sup>+</sup>). RNA and proteins of each fraction were analysed as shown previously (Fig. 3.8 and Table 3.3).

A remarkably high quality of snRNAs and proteins was observed, without any degradation products (Fig 3.11 A, B). Based on the RNA gel, an intact tri-snRNP peak was observed in fraction 18 – 20. While typical tri-snRNP purifications using the Brr2-TAP strain at 75 or 150 mM K<sup>+</sup> conditions showed a yield of about 180 µg tri-snRNP (measured in the eluate), the Sad1-TAP strain results in a yield of 300 µg (at 150 mM K<sup>+</sup>) to 400 µg (at 75 mM K<sup>+</sup>).



Analogous to section 3.1.4.2 above, for a detailed analysis of protein distribution throughout the collected fractions, all fractions were subjected to MS analysis. A total amount of 200  $\mu\text{g}$  of affinity purified tri-snRNP were loaded on a 10-30% GK 150mM glycerol gradient. After glycerol gradient centrifugation, all collected fractions were processed to isolate proteins, and then examined by MS analyses. (See table 3.5)

Notably, since this is a Sad1-TAP strain, while doing affinity purification a certain presence of Sad1p is to be expected even when Sad1p falls off the tri-snRNP during glycerol gradient centrifugation. The amount of Sad1 not associated with tri-snRNPs is substantial, which is not surprising given the 150 mM  $\text{K}^+$  conditions. Prp28p, in yeast a notoriously loosely associated tri-snRNP protein, is clearly present in the tri-snRNP peak. It is also noteworthy that the protein group formed by Spp381/Prp38/Snu23 is absent.

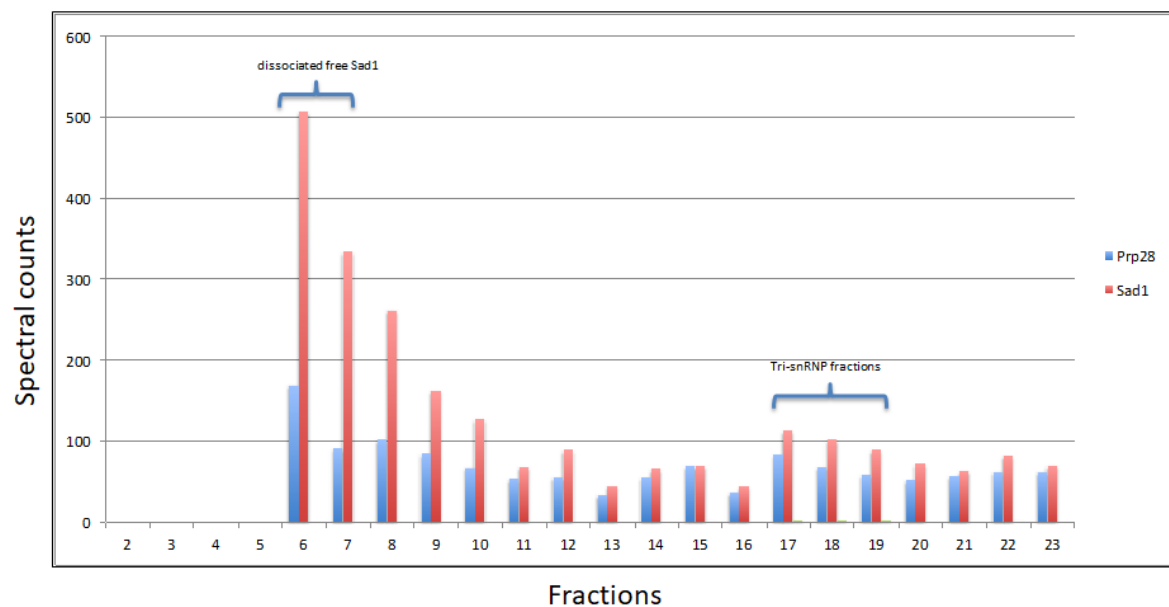
Taken together these experiments clearly show that the MS data from table: 3.5 have to be compared directly with those from table: 3.3 (purification at 75 mM  $\text{K}^+$ ). A considerable improvement can be noted for all U5 snRNP proteins under the 75 mM conditions, with a higher percentage of the total peptide count located in the peak fractions. This effect is particularly strong for Snu114p, Lin1p and Dib1p. Another factor, Prp28p, is somewhat trailing throughout the gradient but at 75mM  $\text{K}^+$  shows a clear peak in the tri-snRNP fraction. U4/U6 snRNP proteins are more confined to the tri-snRNP peak under 75 mM  $\text{K}^+$  as well. Within the group of tri-snRNP specific proteins the improvement for the Sad1p is extraordinary. The Sad1-TAP tri-snRNP under 75 mM  $\text{K}^+$  conditions shows a much more pronounced confinement of Sad1 to the tri-snRNP when compared with the rest of the gradient. As mentioned above, interactions for Spp381p, Prp38p and Snu23p significantly differ in case of tri-snRNP purification at 75 mM  $\text{K}^+$  (table: 3.3), with substantial peptide counts becoming visible in the tri-snRNP only under these low salt conditions.

A

Protein	M.W	10-30% glycerol gradient (150mM)																Tri-snRNPs					Bottom	
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
<b>U5 snRNP proteins</b>																								
Prp8	279.5	225	115	113	119	111	143	193	117	148	129	186	129	442	442	533	825	745	586	192	81	73	70	
Brr2	246.2					117	123	190	181	225	187	301	331	561	478	598	919	768	609	239	94	89	77	
Snu114	114	157	96	89	90	134	81	98	107	100	73	128	167	165	198	268	496	483	339	140	72	84	72	
Prp6	104.2					80	75	93	78	57	48	79	130	159	188	189	462	452	315	98	39	41	36	
<b>Prp28</b>	<b>66.6</b>					<b>169</b>	<b>92</b>	<b>102</b>	<b>85</b>	<b>67</b>	<b>54</b>	<b>55</b>	<b>34</b>	<b>55</b>	<b>69</b>	<b>36</b>	<b>84</b>	<b>68</b>	<b>59</b>	<b>52</b>	<b>57</b>	<b>62</b>	<b>62</b>	
Lin1	40.4									2	1	5	5	11	9	5	5							
Dib1	16.7					0	0	2	2	0	3	2	0	3	3	3	7	10	5	2	0	0	0	
<b>U4/U6 snRNP proteins</b>																								
Prp31	56.3					32	14	23	28	24	21	27	21	36	53	39	102	105	86	36	18	29	28	
Prp3	56					169	92	102	85	36	30	33	25	43	58	51	144	139	110	40	28	31	22	
Prp4	52.4					50	37	54	72	57	57	57	43	58	78	73	149	155	123	60	55	55	41	
Snu13	13.6					5	3	5	6					4	5	1	6	6	6	4	3	3	2	
<b>U4/U6.U5 snRNP proteins</b>																								
Snu66	66.4					32	21	31	40	24	16	29	20	40	42	27	96	82	66	37	39	35	37	
<b>Sad1</b>	<b>52.2</b>					<b>507</b>	<b>334</b>	<b>261</b>	<b>162</b>	<b>127</b>	<b>68</b>	<b>89</b>	<b>45</b>	<b>66</b>	<b>69</b>	<b>44</b>	<b>113</b>	<b>102</b>	<b>89</b>	<b>72</b>	<b>63</b>	<b>82</b>	<b>70</b>	
Spp381	34																							
Prp38	28													0			2	2	2				0	
Snu23	23																							

Sad1-TAP Tagged tri-snRNPs affinity purified at 150mM concentration  
Separation on 10-30% 150mM glycerol gradient

B



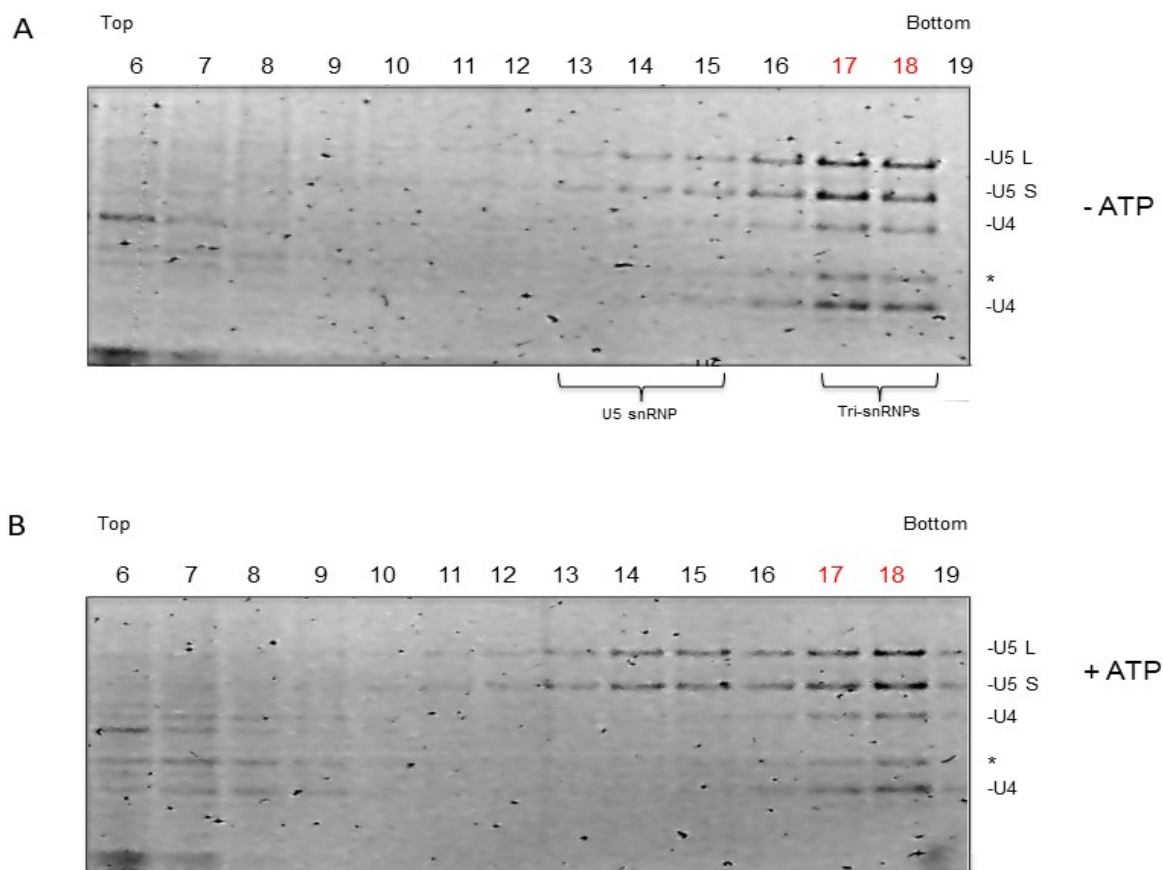
**Table 3.5 MS data of all gradient fractions of Sad1-TAP tri-snRNPs purified at 150 mM K<sup>+</sup>**

**Panel A:** Peptide counts for tri-snRNP proteins in glycerol gradient fractions. **Panel B:** Graphical representation of MS peptide counts for Sad1p (red) and Prp28p (blue). About 200 µg of tri-snRNP (as determined in the affinity column eluate) were loaded on the gradient.

It is evident from these comparative MS analyses that the Sad1-TAP tri-snRNPs, purified under 75 mM K<sup>+</sup> conditions, not only shows a complete and strong set of all the constitutive parts of this snRNP, the particle also remains highly stable when incubated under dissociative conditions with 2 mM ATP (as shown in section: 3.1.4.3, and fig: 3.9)

#### **3.1.4.5 Resistance of purified Sad1-TAP tri-snRNPs to Brr2-mediated dissociation**

After evaluating the protein composition of the Sad1-TAP tri-snRNPs, purified at 150 mM K<sup>+</sup> conditions, I decided to examine their ATP sensitivity as well. In order to examine the effect of ATP incubation, tri-snRNPs were affinity purified from the whole cell extract of the Sad1-TAP yeast strain under 150 mM K<sup>+</sup> conditions (see section: 3.1.4.4). Thus, isolated tri-snRNPs were then subjected to Brr2-mediated ATP-sensitivity assay that involved incubating the affinity-purified particle with or without 2 mM ATP, 2 mM Mg<sup>++</sup> and 150 mM K<sup>+</sup> at 25°C for 30 min. The samples were then separated by glycerol gradient centrifugation (10 – 30 %, 150 mM K<sup>+</sup>) and fractions were subjected to protein and snRNA analysis (Fig. 3.12, Panel A and Panel B). The comparison of both gels shows that upon ATP incubation apparently no substantial increase of a U4/U6 type degradation product occurs.



**Figure 3.12 ATP-dependent dissociation of Sad1-TAP tri-snRNPs purified at 150 mM K<sup>+</sup>.**

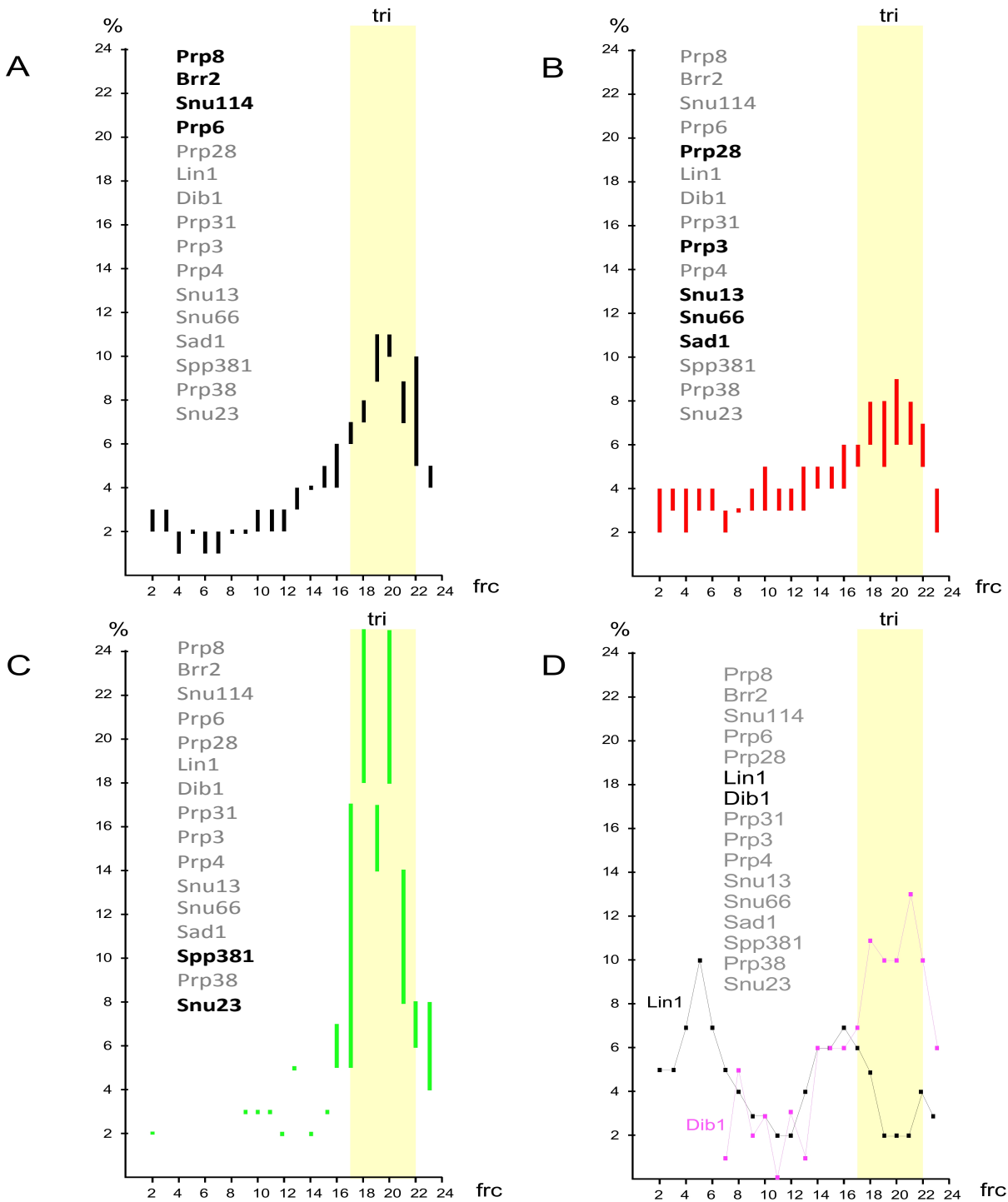
Sad1-TAP tri-snRNP was purified at 150 mM K<sup>+</sup>, incubated with (**Panel A**) or without (**Panel B**) 2 mM ATP and the sample was separated on a 10 – 30 % glycerol gradient under 150 mM K<sup>+</sup>. Gradient fractions were collected from bottom to top. SnRNAs were analysed by separation on 4-12% MOPS-SDS pre-cast denaturing gels and stained with fluorescent nucleic acid stain (SYBR<sup>®</sup> Gold stain). Asterisk \* indicates a degradation product of U4 snRNA lacking the 3' end (Anthony et al., 1997).

### 3.1.4.6 Detailed analysis of protein distribution and tri-snRNP composition upon shift in salt concentration.

To simplify the comparison of the shifts in protein distribution upon changing of salt conditions we converted the numbers from table: 3.3 and 3.4 into a percentage distribution throughout the gradient for each protein, the peptide count over all fractions set as 100 %.

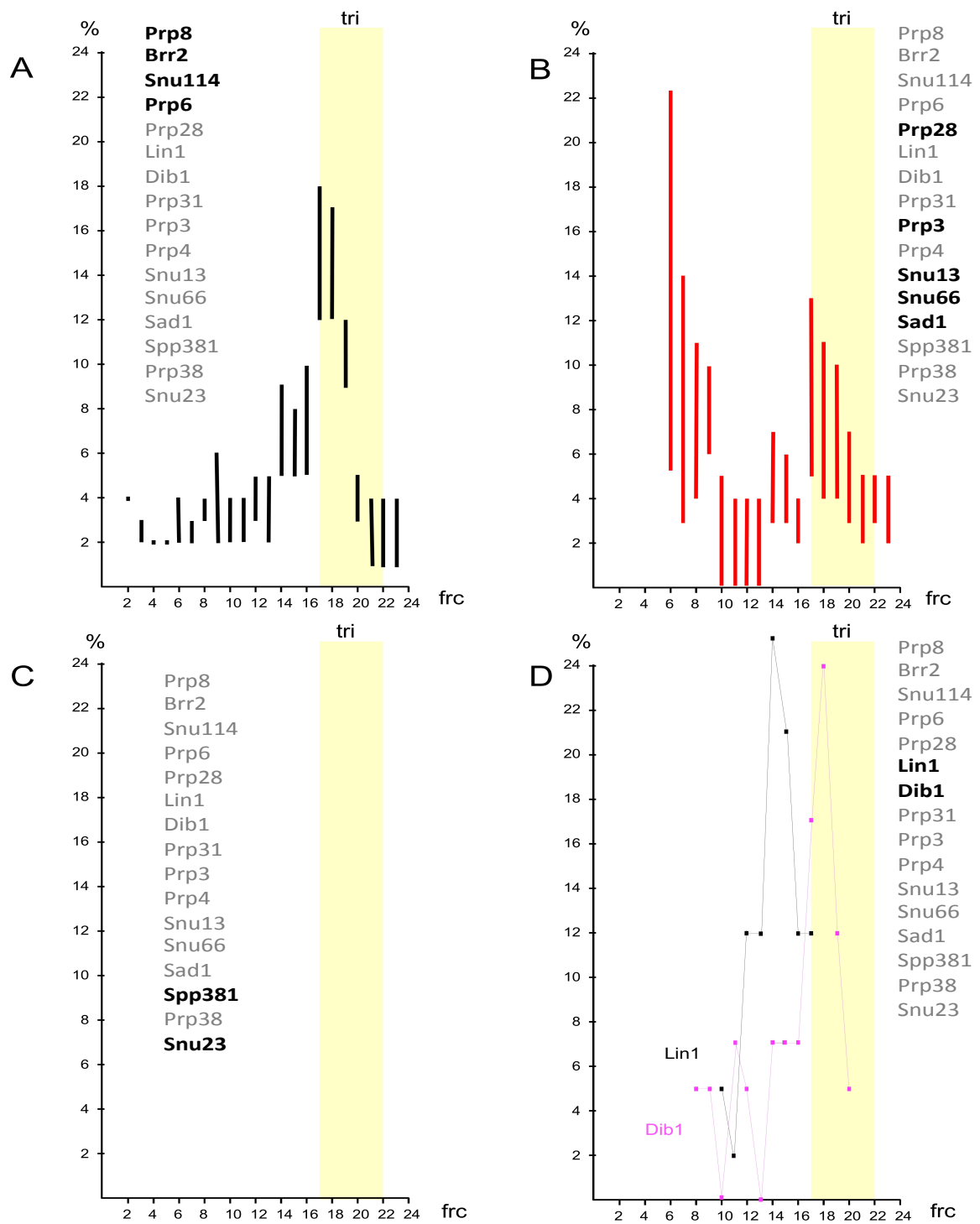
First, we look at the situation for the Sad1-TAP tri-snRNP purified and gradient-fractionated under 75 mM conditions (Fig: 3.13). When analysed with a percentage profile, three protein groups become immediately discernible, each with a distinct distribution profile. First, there is the

group comprising the U5 proteins (except Prp28, Lin1 and Dib1) and U4/U6 proteins Prp31, Prp3 and Prp4, with a strong presence in the tri-snRNP position and a rather low baseline in the top of the gradient (Fig. 3.13). A second group with a common profile comprises the U5 protein Prp28p, the U4/U6 protein Snu13p and tri-snRNP-specific proteins Snu66p and Sad1p (Fig. 3.13, B). Compared with the previous group, the profile is similar but the baseline is somewhat higher and the protein presence in the tri-snRNP position slightly reduced. Most dramatic is the situation in the third group; here we see a strict confinement of three tri-snRNP-specific proteins, Spp381p, Prp38p and Snu23p, to the tri-snRNP position with only minute amounts in the gradient. The two U5 proteins Lin1p and Dib1p defy easy assignment to any of the three groups: the non-essential U5 component Lin1p shows two peaks outside of the tri-snRNP position, while the essential U5 component Dib1p peaks well in the tri-snRNP peak with a relatively strong presence in the middle of the gradient (Fig.3.13, D). Taken together, again the picture of a compact, highly integer tri-snRNP emerges.



**Figure 3.13 Percentage distribution of protein and tri-snRNP composition under 75 mM K<sup>+</sup> conditions.** For each gel fraction peptide counts from Tab. 3.3 were calculated as percentage of total counts for a particular protein. Vertical axis shows percent value, horizontal axis the fraction number. **Panel A** Distribution of U5 and U4/U6 proteins (see insert and text). **Panel B** Prp28p/Snu13p/Snu66p/Sad1p group (see text). **Panel C** Spp381/ Snu23 group (see text). **Panel D** Distribution of Lin1p and Dib1p. The bars in A – C indicate the min/max percent value for the protein group in a particular fraction of the gradient. The tri-snRNP position is indicated in yellow.





**Figure 3.14 Percentage distribution of protein and tri-snRNP composition under 150 mM K+ conditions.** For each gel fraction peptide counts from Tab. 3.5 were calculated as percentage of total counts for a particular protein. Vertical axis shows percent value, horizontal axis the fraction number. **Panel A** Distribution of U5 and U4/U6 proteins (see insert and text). **Panel B** Prp28p/Snu13p/Snu66p/Sad1p group (see text). **Panel C** Spp381/Snu23 group (see text). **Panel D** Distribution of Lin1p and Dib1p. The bars in A – C indicate the min/max percent value for the protein group in a particular fraction of the gradient. The tri-snRNP position is indicated in yellow.

If we now subject the Sad1-TAP tri-snRNP to 150 mM  $K^+$  during purification and gradient, remarkable changes occur with severe consequences for the tri-snRNP (Fig. 3.14). The group of U5 proteins (except Prp28p, Lin1p and Dib1p) and U4/U6 proteins Prp31p and Prp4 is still strongly present in the tri-snRNP (Fig. 3.12, A). Prp3p instead starts bleeding from the tri-snRNP and appears in top of the gradient, grouping now rather with the second set of proteins (U5 protein Prp28p, the U4/U6 protein Snu13p and tri-snRNP-specific proteins Snu66p and Sad1p). Compared to 75 mM  $K^+$  this second group now shows a much stronger presence outside of the tri-snRNP position in the gradient (Fig. 3.14, B). Under 150 mM  $K^+$ , proteins become destabilized from the tri-snRNP (Fig. 3.14). Sad1p and Snu66p are still present in the Sad1-TAP tri-snRNP, but start to dissociate. The Spp381p, Prp38p and Snu23p group of tri-snRNP-specific proteins is virtually absent from the tri-snRNP and even from the top fractions of the gradient (Fig. 3.14, C).

In any case, it is evident that the integrity of a yeast tri-snRNP purified under 150 mM  $K^+$  is severely compromised due to bleeding or complete loss of proteins. It was recently shown with reconstitution assays that the human orthologs of yeast Spp381p, Prp38p, and Snu23p, namely MFAP1, hPRP38 and hSNU23 could form a stable subcomplex (Ulrich and Wahl, 2017). This human subcomplex depends on the same, evolutionarily highly conserved, interactions as observed between Spp381p, Prp38p, and Snu23p (Ulrich and Wahl, 2017). Most remarkably, the human MFAP1 protein is a B complex-specific protein and the hSNU23 – hPRP38 – MFAP1 complex indeed enters the human spliceosome at the level of the B-complex and independent of the human tri-snRNP. It has been suggested that in yeast the Snu23p – Prp38p – Spp381p subcomplex is part of the tri-snRNP and enters the spliceosome with the tri-snRNP (Jia and Sun, 2018), however, it is also important to note that so far, purified yeast tri-snRNPs are shown to be devoid of loosely associated factors such as, Sad1 (Häcker et al., 2008), and Prp28 (Nguyen et al., 2015; Wan et al., 2016). This may hint towards the mutually exclusive presence of Sad1/Prp28 and Prp38 – Snu23 – Spp381 in the yeast tri-snRNP (as observed in the case of Brr2-TAP tri-snRNPs as well, table 3.1- column: B). Interaction behaviour of Dib1p and Lin1p with the tri-snRNP particles at 150 mM  $K^+$  appears similar to as under 75 mM  $K^+$ , with Lin1p outside of the tri-snRNP peak and Dib1p inside and outside (Fig. 3.14, D).

Based on the conclusions of the biochemical experiments and comprehensive MS analyses, it is apparent that the Sad1-TAP tri-snRNP, purified at 75 mM K<sup>+</sup> concentration, not only is remarkably resistant to dissociative conditions in the presence of ATP but also shows a complete and strong set of all the constitutive parts of this snRNP. These results adequately support the interesting possibility that such a Sad1-TAP tri-snRNP may have a different conformational state when compared to tri-snRNPs obtained from previous purifications used for structure determinations. Therefore, I set out to investigate the structural features of the Sad1-TAP tri-snRNPs using cryo-electron microscopy.

### **3.2 Electron microscopy of affinity-purified yeast tri-snRNPs**

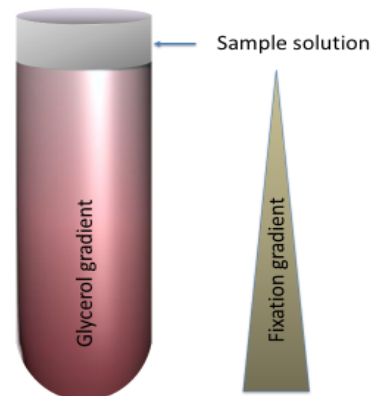
Conclusions of the ATP sensitivity assays and comprehensive MS analyses clearly proved that the yeast Sad1-TAP tri-snRNPs, which were purified using my optimised purification method, significantly differed from the published yeast tri-snRNP in terms of their biochemical behaviour as well as the protein composition. In fact, these isolated Sad1-TAP tri-snRNP particles are more closely comparable to their human counterparts. We observed that similar to human tri-snRNPs, yeast Sad1-TAP tri-snRNPs when purified under optimal conditions, were not only resistant to ATP dependent dissociation but also showed stable integration of crucial factors like Sad1p.

Understandably, the biochemical evidence presented in sections above raises the possibility that this yeast tri-snRNP may be structurally different from the yeast tri-snRNP structures published until now (Nguyen et al., 2015- PDB 5gan; Wan et al., 2016 – PDB 3jcm) and may in fact resemble the human tri-snRNP structure. This different conformational state possibly resembles the reported human tri-snRNP structures, including features like the different arrangement of Brr2p. Hence, I purified Sad1-TAP tri-snRNP particles, and in collaboration with my colleagues, employed single-particle electron microscopy with the aim to inspect the structural organization of Sad1-TAP tri-snRNP particles closely.

The objective of the structural investigation is to resolve the 3D structure of the intact tri-snRNP particle to a sufficiently high resolution that will enable the fitting of the various tri-snRNP protein models, which have already been resolved by x-ray crystallography or NMR.

Thus, purified particles (method section 2) were analysed by single-particle EM in collaboration with Norbert Rigo and Ilya Komarov from our group (Department of cellular biochemistry, MPI-BPC). For negative stain EM, particles were fixed during gradient centrifugation using the ‘GraFix’ method (Kastner et al., 2008). The method entails sedimentation of particles in a density gradient with a mild chemical fixation agent. The fixation procedure protects the particles and

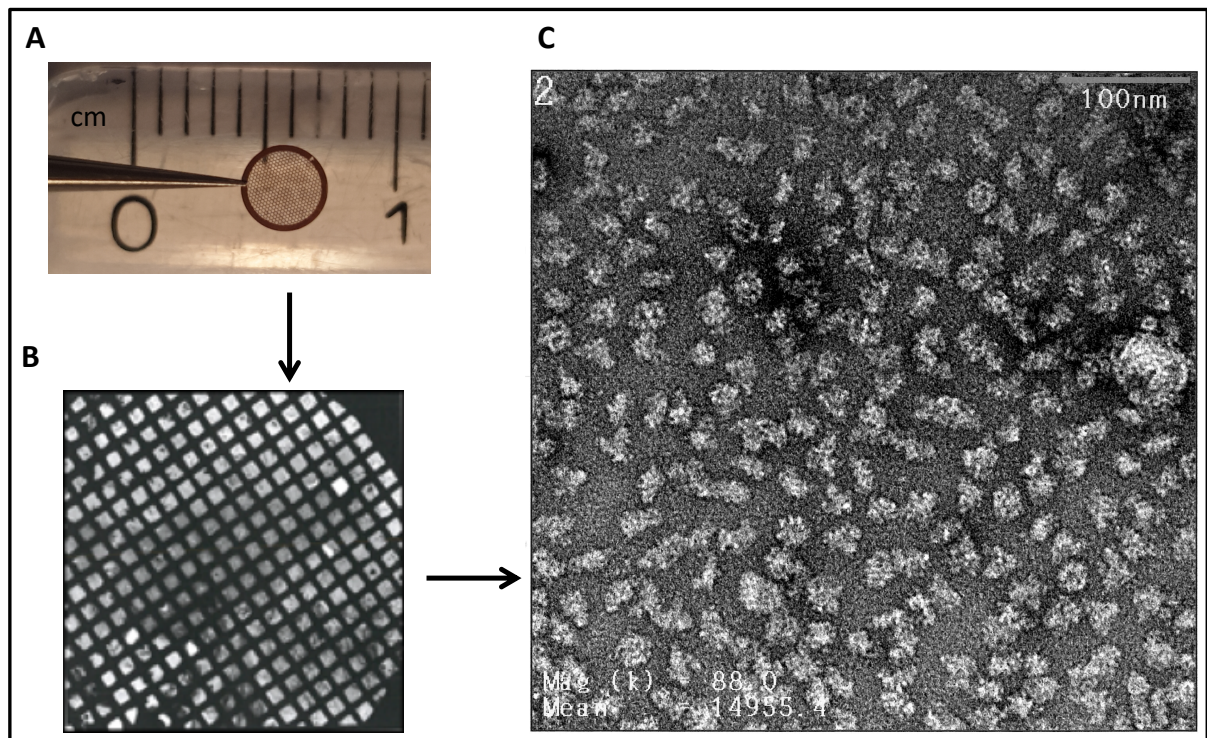
stabilises them against unspecific aggregation or fragmentation caused by the various forces acting on the particles during sample preparation for EM (Kastner et al., 2008). Most widely used fixative agents for EM are aldehydes. The aldehyde group, upon reacting with an amino group of protein results in the formation of a Schiff base, and thereby crosslinks the proteins. Here we have used glutaraldehyde (a di-aldehyde) to perform GraFix. After crosslinking, complexes can be directly used for negative-stain EM, cryo negative-stain EM or for cryo-EM without staining after doing buffer-exchange to remove glycerol.



**Figure: 3.15 Cartoon depiction of 'GraFix'.** The particles present in the sample migrate into the increasing glutaraldehyde concentration as they pass through the denser region of the glycerol gradient. This gradient of glutaraldehyde is prepared by adding it to the denser gradient solution. Optionally, a cushion of the lighter gradient solution is sometimes used for buffer exchange during the initial period of gradient centrifugation. This practice helps to remove primary amines that are often present in the sample buffer, and are incompatible with aldehyde crosslinking (Kastner et al., 2008).

### 3.2.1 Negative-stain 2D EM analysis of the yeast Sad1-TAP tri-snRNPs

The negative-stain EM analysis procedure started with isolation of the tri-snRNP particles. Isolated tri-snRNPs were fixed using a maximum concentration of 0.1% of glutaraldehyde at the bottom of the gradient. High-speed ultracentrifugation conditions, as already mentioned in the Methods section (see section 2), were selected in such a way that the tri-snRNP particles sediment in the lower third of the glycerol-gradient, a region of the gradient, which corresponds to approximately 0.075% concentration of glutaraldehyde. The tri-snRNP particles were visualised by negative-stain EM using 3.5% uranyl formate and the single carbon film method (Kastner et al., 1990). Negative-stain EM grids were prepared from the tri-snRNP peak fractions of the gradient according to our optimised protocols described in the methods section.



**Figure: 3.16 Image depicting the relative size perspective of the grids and adsorbed sample**

**A:** A copper grid used for sample preparation, shown here with the scale in the background (in centimetres). **B:** Magnified image of the copper grid with adsorbed sample on the carbon foil. **C:** A micrograph of negatively-stained particles recorded on a CM200 FEG electron microscope (magnification- 78 000 – 80 000 folds)

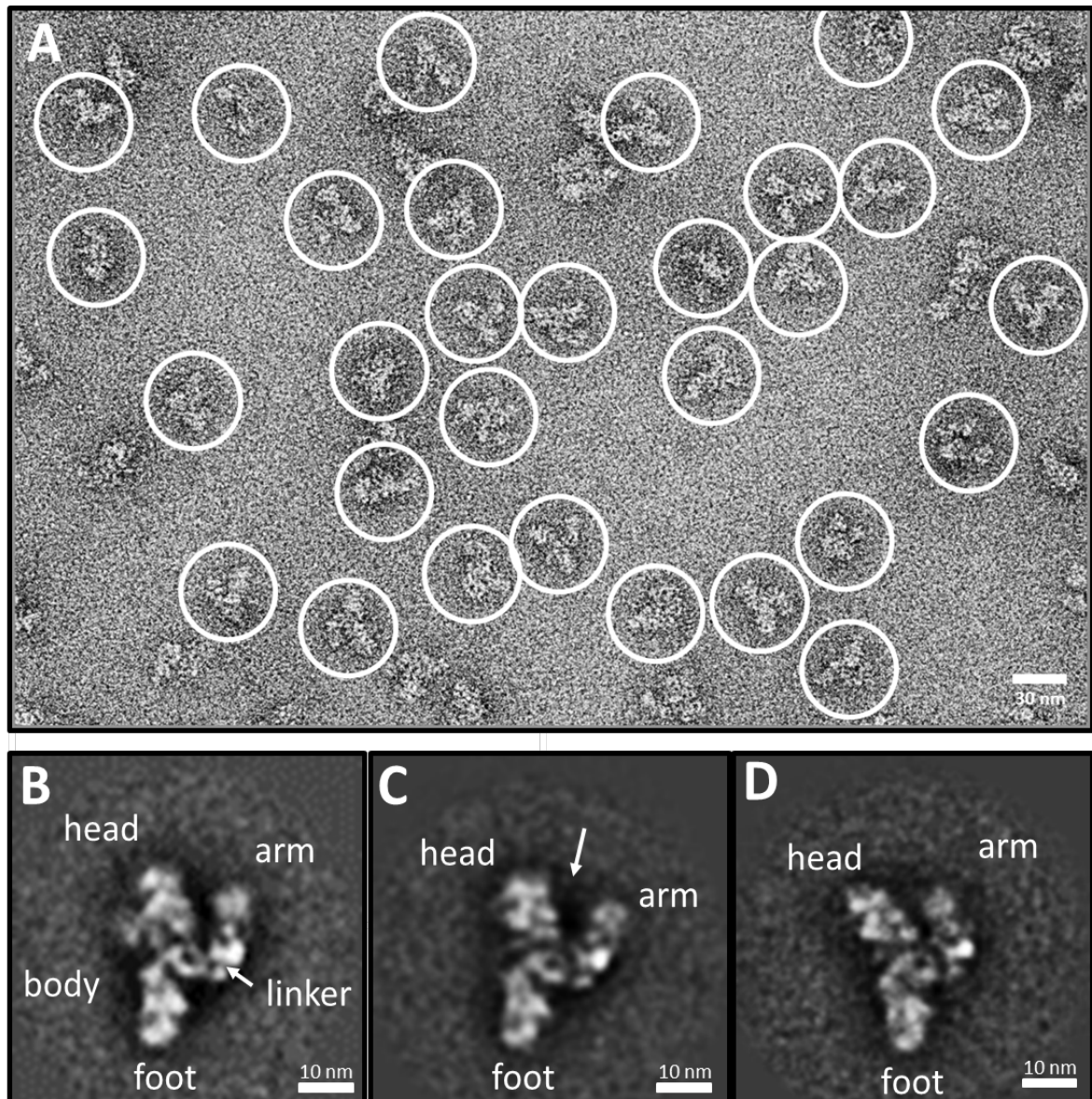
Negative-stain EM micrographs showed a well-dispersed population of triangular-shaped, elongated particles with an approximate dimension of 30-35 nm (fig: 3.17, A)

Roughly, 15000 individual single-particle images were collected. Using our in-house software package (Coweyes), a repeated round of image processing was carried out. After obtaining a reference-free alignment (Dube et al., 1993), images were further subjected to multivariate statistical analysis and classification (van Heel and Frank, 1981a); (van Heel et al., 1996);(Dube et al., 1993). Resulting class averages were used as the reference images in subsequent rounds of alignment until the class averages were stable. In total, four rounds of alignment were accomplished to get the final representative class averages (fig: 3.18).

After a final round of alignment, resultant predominant views (class-averages) of the particle (fig: 3.17, B), representative class averages, revealed the tri-snRNP's domain architecture at a high signal to noise ratio. The most prominent view shows an elongated main body with an upper head and a lower foot domain, and a small arm connected to the central portion of the main

body by a linker domain. The domains of the prominent view mentioned here are comparable to those from a previous study from our group, which focused on the mapping of tri-snRNP proteins (Häcker *et al.*, 2008).

Analysis of the class averages revealed primarily two types of predominant positions of the head and arm region relative to the main body (fig: 3.17, B), the 'open conformation' state, and the 'closed conformation' state. In the 'open conformation' state, the head and arm regions are well separated from each other (fig: 3.17, C). In the 'closed conformation' instead, the two regions, head and arm, are in close proximity to one another with no visible space in between, resulting, thus in a more compact conformation (fig: 3.17, D).



**Figure: 3.17 EM of Sad1-TAP purified yeast tri-snRNPs.** Sad1-TAP tagged purified tri-snRNPs were separated by density gradient centrifugation. Gradient fractions were collected from bottom to top; the tri-snRNP particles in the selected peak fractions of the gradient were adsorbed to a carbon film, negatively stained, and the carbon film was attached to a copper grid for EM analysis. **A:** EM overview image of negatively stained Sad1-TAP tri-snRNP showing a well-distributed population of triangular complexes. **B:** A representative class average reveals the tri-snRNP's domain architecture at a high signal to noise ratio comprising an elongated main body with an upper head and a lower foot domain, and a small arm, which can be seen connected to the central portion of the main body by a linker domain. **C:** A class-average showing the open conformation. Here the head domain is well separated from the arm domain. **D:** A representative class average is showing the closed conformation state. Here the arm domain is in close proximity to the head domain.

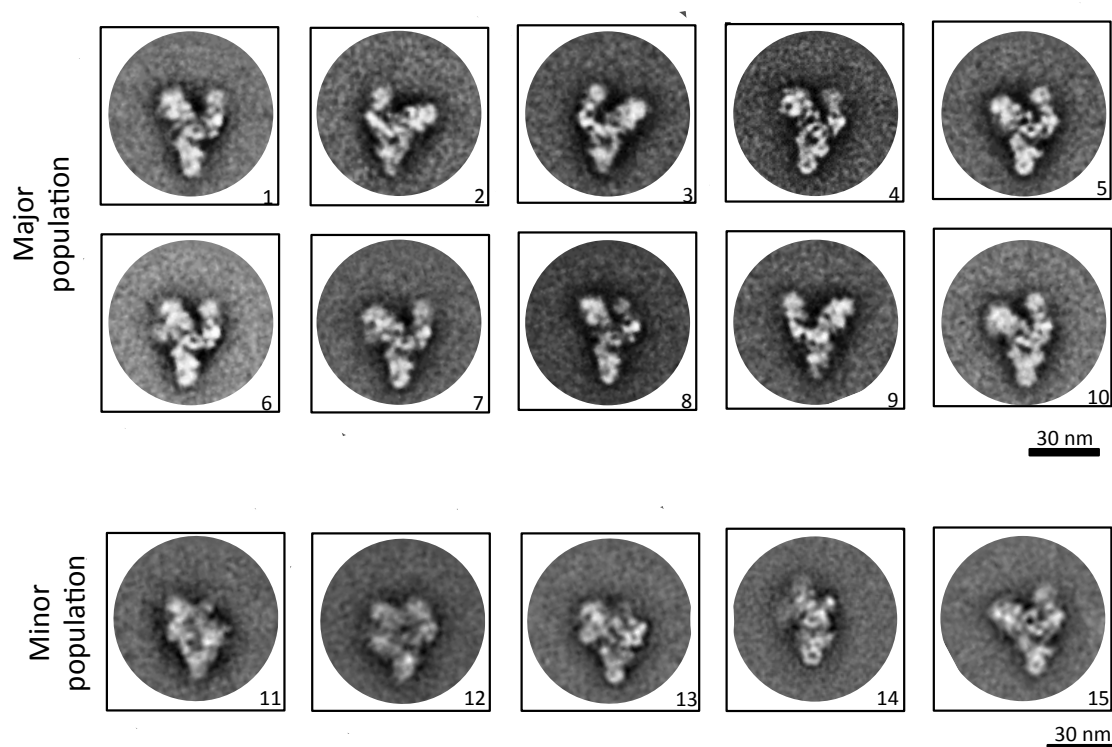


During previous investigations (Häcker *et al.*, 2008) of yeast tri-snRNPs, the EM specimens were tilted, and it was shown that the overall appearance of individual open or closed state particles does not change upon tilting. Closed and open state confirmations are therefore not just artefacts caused by different orientations of particles on the grid.

Besides these two predominant views (fig: 3.18, upper and middle panel box # 1 to 10), other subclasses were also detected to a minor extent (fig: 3.18 lower panel box # 11 to 15). Nearly in all attempts of sample preparations, the closed and open conformations were very consistent. Importantly, the addition of the TAP tag to the Sad1 did not alter the appearance of the particles in the electron microscope, suggesting that the TAP tag had no influence on the overall structure of the tri-snRNPs.

The overall view of these class averages suggested that the Sad1-TAP tri-snRNP, regardless of showing different biochemical behaviour and stably associated Sad1, still preserved a similar arrangement of the head domain, which corresponds to Brr2 (Häcker *et al.*, 2008), as in the published yeast tri-snRNP structure (Nguyen *et al.*, 2015- PDB 5GAN; Wan *et al.*, 2016 – PDB 3JCM). Nevertheless, since the MS data had already shown that the Sad1-TAP tri-snRNP stably contained Sad1p, hence, it was intriguing to find out the location and organisation of Sad1 in the tri-snRNP structure.





**Figure: 3.18 Representative set of 2D class-averages of Sad1-TAP purified yeast tri-snRNPs.**

Sad1-TAP-purified tri-snRNPs were separated on a glycerol gradient. Gradient fractions collected from bottom to top; the tri-snRNP particles in the selected peak fractions of the gradient were adsorbed to a carbon film, negatively stained, and the carbon film was attached to a copper grid for EM analysis. Approx. 15 000 individual single-particle images were collected. Using our in-house software package (Coweyes), a repeated round of image processing was carried out. After obtaining a reference-free alignment, images were further subjected to multivariate statistical analysis and classification (van Heel and Frank 1981, van Heel and Harauz 1988, Dube, Tavares et al. 1993). Resulting class averages were used as reference images in subsequent rounds of alignment until the class averages were stable. Four rounds of alignments were done to get these final representative class averages. Classes shown here in upper panels (box # 1 to 10) represent the prominent view containing the majority (~95%) of the classified particles. Lowermost panel (11 to 15) shows some of the few classes with particles compact in overall appearance, which represents the minor population of the particles (~5%).

### 3.2.2 3D Structure determination of the yeast Sad1-TAP tri-snRNP particles

In section 3.2.1 we observed that the negative-stain 2D EM class averages of the Sad1-TAP tri-snRNP revealed an overall view, which closely resembled the already known yeast-like conformation of the tri-snRNP. However, class-averages only provide the overall view of the particles, and simply on the basis of the class-averages, it is extremely difficult to distinguish

conformational changes. Even minor conformational differences can be of great significance and can only be visualized by a high-resolution 3D model. Additionally, the determination of a 3D structure was essential to localise and understand the interactions of factors, such as Sad1p, which have not been annotated before in any of the previously published yeast tri-snRNP structures. We, therefore, decided to determine the 3D structure of the Sad1-TAP tri-snRNP particles by cryo-EM. To achieve this objective, we first isolated Sad1-TAP tri-snRNPs and performed negative-stain EM 3D reconstruction to ensure their quality and to get a low-resolution negative-stain 3D model. Afterwards, particles were analysed by cryo-EM in order to obtain a high-resolution 3D structure.

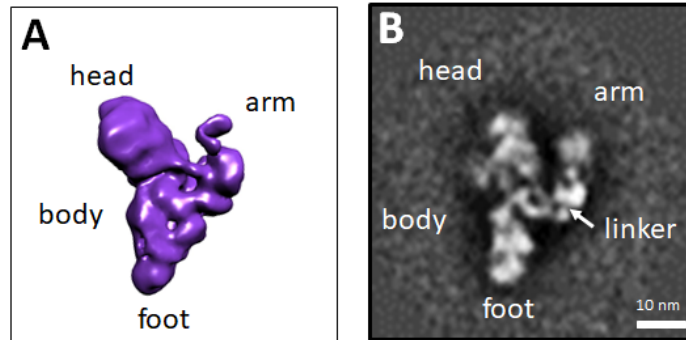
### 3.2.2.1 Low-resolution negative-stain EM 3D model reconstruction

Procedures for affinity purification, fixation, and negative-stain grid preparation were similar as described in section 3.2.1, except this time, data was further classified and processed using in addition RELION software, which enabled the *ab initio* 3D model reconstruction.

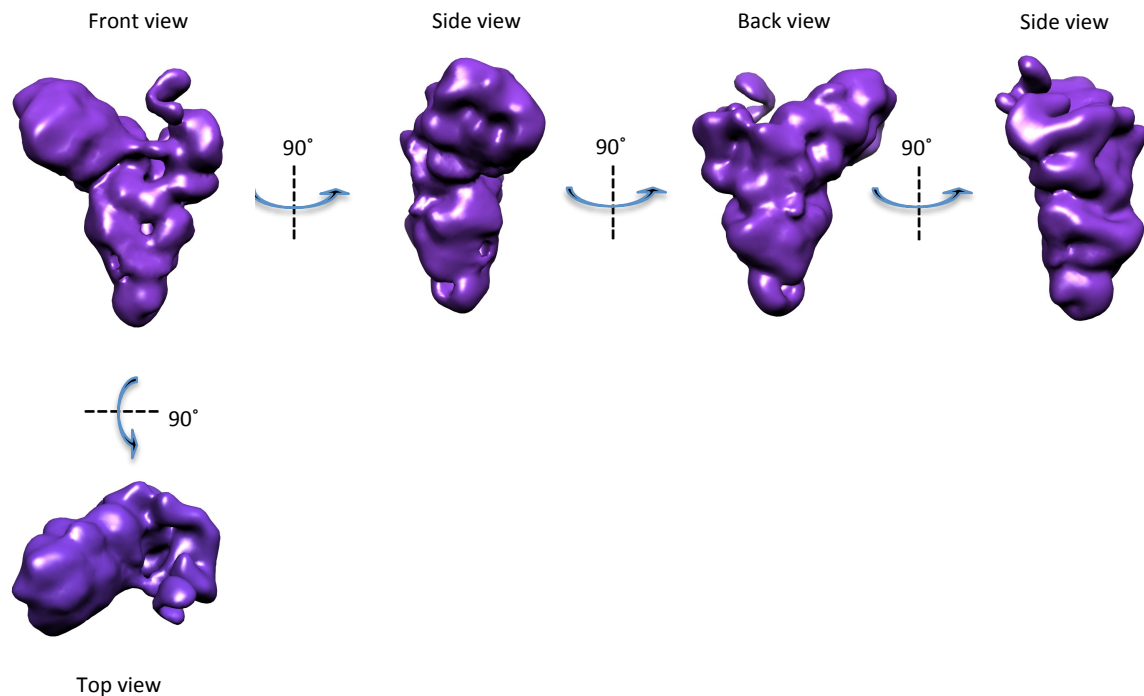
The negative-stain EM grids were screened using a CM200 FEG electron microscope, and micrographs were recorded to gather a sufficient number of particles (~15 000 particles) for generating a low-resolution negative stain-EM 3D model. The classification of particles, including 2D and 3D classifications, as well as 3D auto-refinement runs, were carried out in RELION using particles binned pixel size of ~5.02 Å/pixel. The low-resolution negative-stain EM model was processed with similar approaches (reference-free particle picking) mentioned earlier in section 3.2.1.

After the final refinement, the low-resolution 3D model showed the standard view of the published yeast tri-snRNP particles with similar features. These features were comparable to prominent domains depicted in the representative 2D class-averages (Fig: 3.19), and this applies to the yeast-like Brr2 orientation as well.

The limited resolution of the reconstructed negative-stain 3D model did not allow the placement of the individual constituent factors; hence, further analysis was carried out using the cryo-EM technique with the goal to get a 3D model with sufficiently high-resolution. The cryo-EM 3D model would potentially enable the fitting of the known crystal structures of individual constituent factors that have been found in the MS analysis of the Sad1-TAP tri-snRNP particles.



**Figure: 3.19 Low-resolution negative-stain EM structure of Sad1-TAP tri-snRNP shows similar architecture domains as seen in prominent class-averages.** Image shows similar domain architecture (head, body, foot and arm domain) in the negative-stain EM model of Sad1-TAP tri-snRNP (A), as were observed previously in 2D classes (section 3.18 and Häcker et al., 2008) (B).

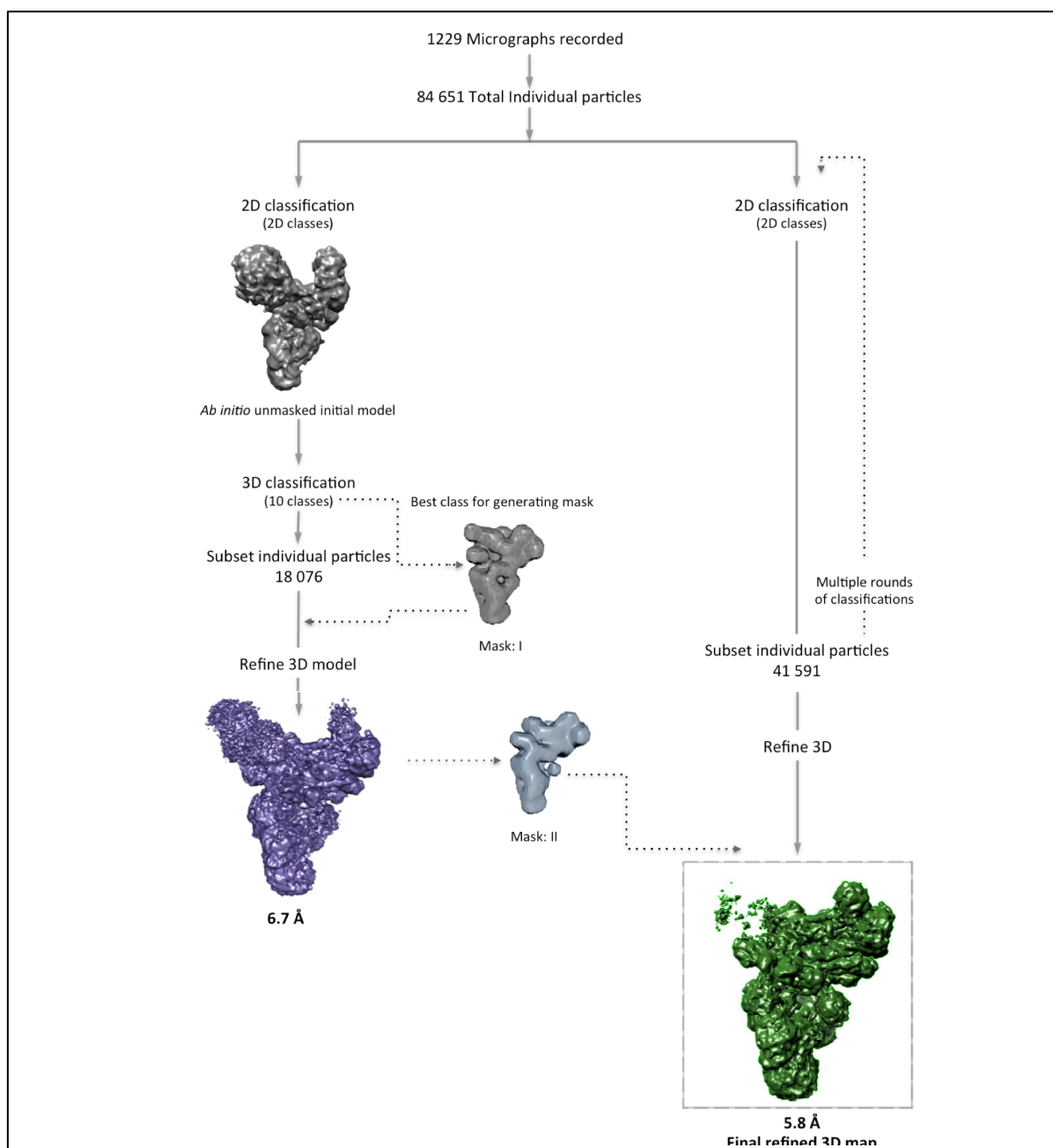


**Figure: 3.20 Different orientations of the low-resolution 3D structure of the Sad1-TAP tri-snRNP reconstructed using negative-stained particles.** Image showing front, side, and top views of the 3D structure of Sad1-TAP tagged tri-snRNP particles. Curved arrows denote the direction and the degree of rotation ( $90^\circ$ ) in each turn. The vertical dotted line represents the plane of rotation (along the Y-axis in the upper four images, and along X-axis in the case of the bottom image). The final low-resolution 3D model was reconstructed using negative-stain particles. The micrographs, obtained from CM200, were processed by the selection of particles and then generating the class-averages in order to reconstruct the final low-resolution 3D model (for details, please see methods section: 2).

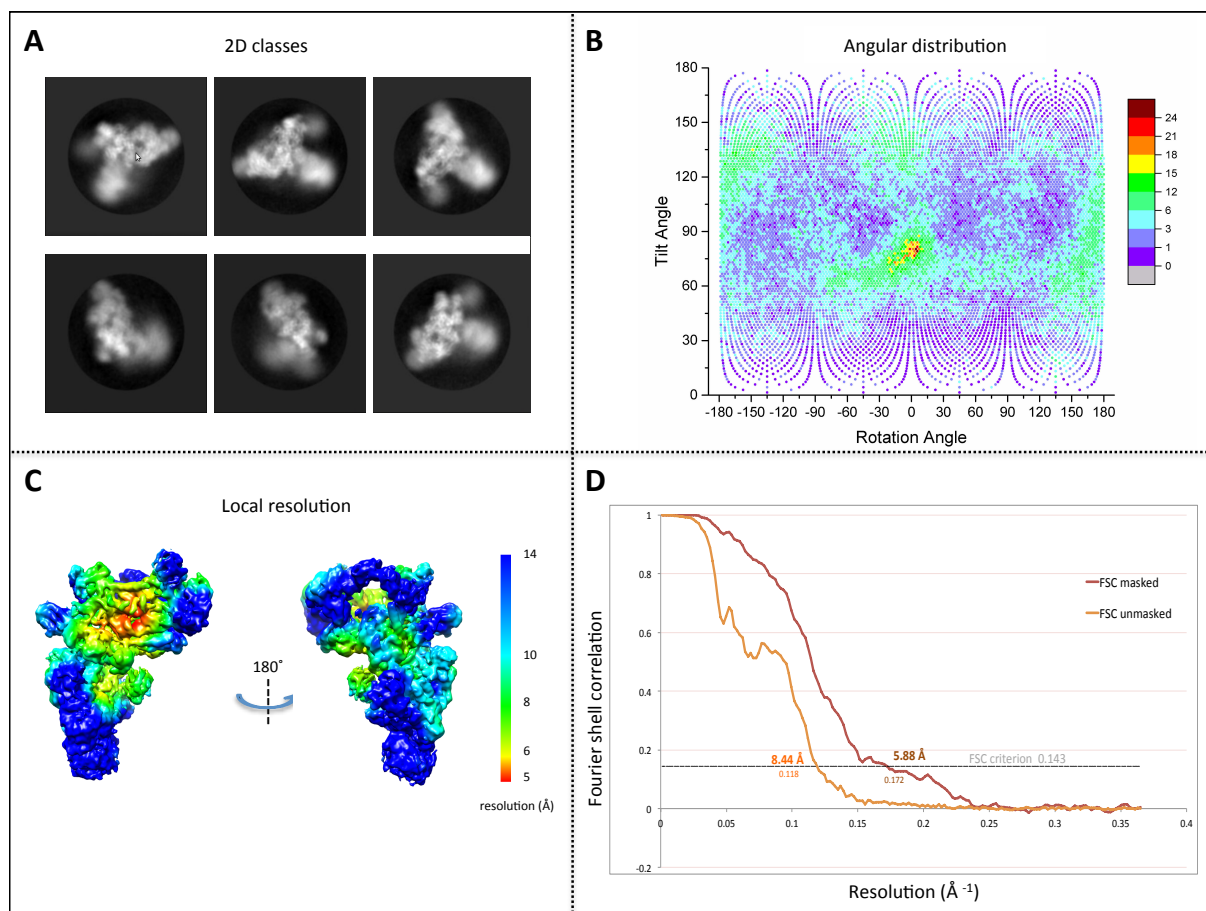
### 3.2.2.2 Cryo-EM 3D structure reconstruction of Sad1-TAP tri-snRNP

Next, we affinity purified tri-snRNPs, fixed by GraFix method, and adsorbed them on an EM grid with a thin carbon film at 100% humidity and 4°C, and immediately plunged into liquid ethane. In order to reconstruct a 3D cryo-EM model, 18 cryo-grids were prepared under varied conditions of freezing and absorption. Detailed protocols for purification, fixation of particles and data processing including the type of detector, calibrated pixel size, defocus range, frames per stack, and dose rate for data collection have all been described in the Methods section (see section 2).

Cryo-images were recorded by using a K2 direct electron detector mounted on a Titan Krios electron microscope (FEI, Netherlands). Nominal magnification resulted in the calibrated pixel size of 1.37 Å/ pixel at the specimen level. A total sum of ~85 000 particle images (from here onwards addressed as a particle in this section) was auto-picked from the recorded micrographs (~ 1 200 micrographs) by using the reference-free particle picking routine in Warp software. The latter was also used for estimation of contrast transfer function (CTF) parameters as well as for global and local movement correction during alignment of the dose-fractionated image frames. We applied multiple sorting steps at the 2D and 3D level. Multivariate statistics and classification helped us in the identification and elimination of bad classes from the final round of processing. After processing and several rounds of refinements in RELION, the final structure was generated using ~41 000 particles (see particle sorting scheme in fig: 3.21) after applying a soft mask. Overall global resolution of 5.8 Å was calculated by Fourier-shell correlation function using the standard 0.143 cut-off FSC criterion (fig: 3.22, D). Consistent with the low-resolution negative-stain EM model (see Section 3.2.2.1), the cryo-EM model depicts the same domain structure, but with a higher resolution (fig: 3.23). The tri-snRNP shows the well-known triangular shape, composed of a body domain, a bulky head domain and an arm domain.



**Figure: 3.21 Flow diagram for computational sorting of yeast tri-snRNP particles.** Starting with 1229 micrographs, 84 651 individual particles were picked with Warp software. Initially, in two separate processing attempts, data was processed for generating 2D classes. In a first attempt, the entire dataset was used, and the best class was selected to generate an *ab initio* 3D map without applying any soft mask or solvent mask. Subsequently, using this primary model, we again processed the data to obtain 3D classes. From the resulting 10 classes, a suitable 3D class was selected to generate a mask (Mask: I) that was further used to refine the model. The processed data resulted in a 6.7 Å 3D model which was constructed using ~21 % of the entire dataset. To further refine our cryo-EM model, a separate attempt was made, and the complete dataset of 84 651 particles was again processed to generate 2D classes. 2D multivariate statistics and classification was applied to the non-aligned images of the particles and subsequently to the aligned particle images. In each round of classification, corrupt classes were eliminated. This sorting procedure resulted in a database of 41 591 particles, which were subsequently subjected to refine 3D. Final processing using this dataset resulted in a 3D map with the maximum resolution of 5.8 Å, which was created using a soft mask (mask II) that was generated from the previously obtained 6.7 Å model.

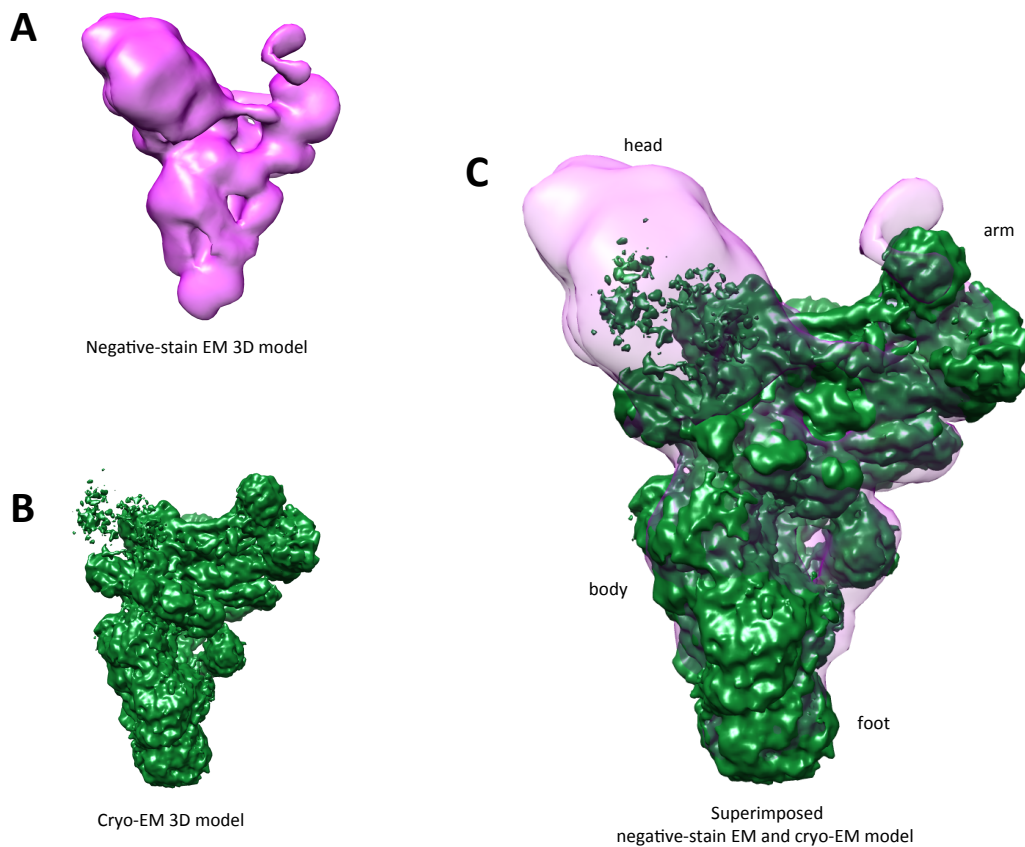


**Figure: 3.22 Cryo-EM analysis of yeast tri-snRNP particles.** Image showing Cryo-EM data processing for the Sad1-TAP tri-snRNP particles. **A** Several representative 2D class averages showing different views of Sad1-TAP tri-snRNPs. **B** Euler angle distribution of all the particle images that were being used in the final 3D map reconstruction. Coordinates on the X-axis and Y-axis denote the Tilt angle and the angle of rotation, respectively. The coloured stripe on the extreme right specifies the colour of the dot that corresponds to the number of particles at a given Euler angle/coordinates. **C** Local resolution distribution plot shows that the range of local resolution varied from 5 to above 14 Å in different regions of the map denoted by different colours. **D** Fourier-shell correlation function was calculated after the final round of particle sorting. The refined selected dataset was split into two independent groups, and the global resolution of 5.8 Å was calculated using the FSC cut-off criterion of 0.143.

Electron density maps, obtained from the two independent data sets (negative-stain EM electron density map and 3D cryo-EM electron density map), were imported in UCSF Chimera software (Yang *et al.*, 2012) and aligned to one another, relative to the fixed reference regions (relative to lower and central body region i.e., region corresponding to Snu114 and the Prp8 NTD1 domain)(see fig: 3.23). The Superimposed view of the EM density maps revealed the missing

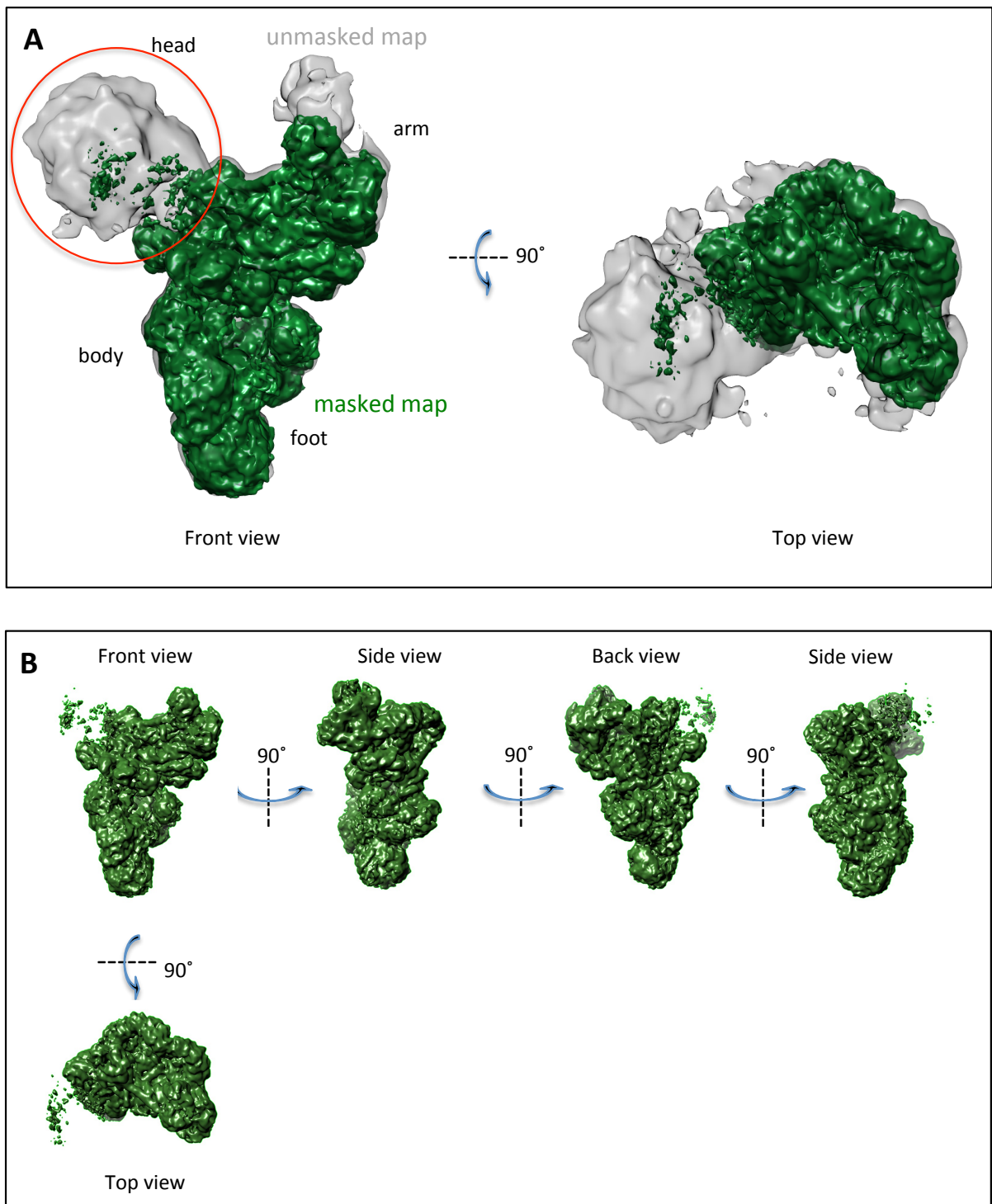
head region of the Sad1-TAP tri-snRNP particles in the 5.8 Å cryo-EM structure model (see fig: 3.23, C). The head domain, due to its high mobility, needed to be masked and is therefore not included in the 5.8 Å structure. It is instead present in the 6.7 Å structure and can thus be fitted back into the 5.8 Å structure to complete the picture of the tri-snRNP. Notably, this highly flexible region/ head domain has been already shown (Häcker *et al.*, 2008) to correspond to the region of yeast Brr2. The absence of the head domain is likely due to factors such as a prominently variable degree of movement of the head domain, and probably a preferred specific adsorption orientation of the particles on the grid relative to the rest of the body. It was difficult to resolve this region at a considerably higher resolution. The poor density for the Brr2 region in our 3D model (fig: 3.23, B and fig: 3.24) is also observed in the published yeast model (PDB 5gan), also this model lacks the proper density in the head domain.

Furthermore, this observation remained unchanged even by overlapping the electron density map of the initial unmasked low-resolution cryo-EM model with the electron density map of the masked final refined 3D cryo-EM. Our cryo-EM model shows that the head region (corresponding to Brr2) appears to be in a similar position as shown in the published yeast tri-snRNP structure (fig: 3.24, A fig: 3.27).



**Figure: 3.23 Comparing the negative-stain EM structure with the 3D cryo-EM structure of the yeast Sad1-TAP tri-snRNPs.** Comparison of EM density models, obtained from negative-stain EM analysis (**A**) and the cryo-EM analysis (**B**). (**C**) Superimposed view of two structure models (A and B) aligned in UCSF Chimera.





**Figure: 3.24 Cryo-EM 3D structure of Sad1-TAP tri-snRNP particles.** Image showing the reconstructed 3D structure of Sad1-TAP tagged tri-snRNP particles. Final 5.8 Å resolution 3D model was reconstructed using unstained particle images obtained from micrographs recorded on the cryo-electron microscope. **A:** Image depicts two superimposed cryo-EM models, with the unmasked cryo-EM 3D model shown in grey colour, and final refined masked cryo-EM model shown in dark green colour. Image shows the front view of the tri-snRNP particle on the left-side and the top view on the right side of the image. The red circle in the upper image shows the unresolved region in the masked structure (corresponding to head domain/ Brr2 in the unmasked 3D model. For further comparison refer to negative stain EM model shown in fig: 3.19, A and B). **B:** Image showing the front, side, and top view of masked refined final 3D reconstructed structure of Sad1-TAP tagged tri-snRNP particles. Curved arrows denote the direction, and the degree of rotation (90°) for each turn and the vertical

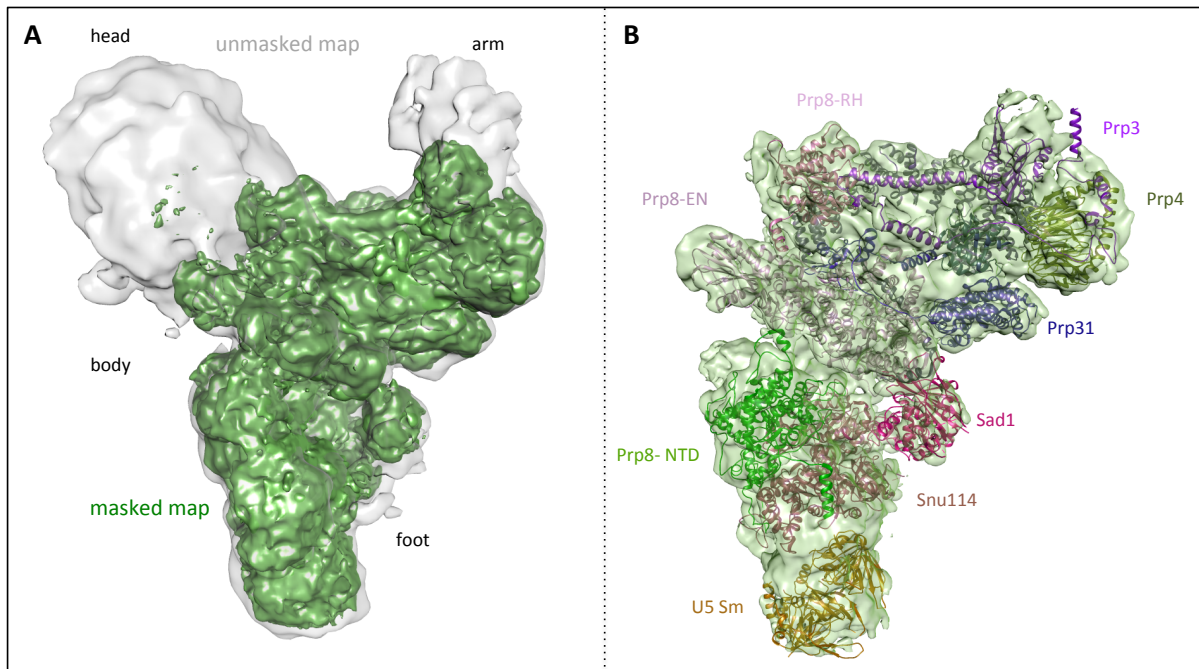
dotted line represents the plane of rotation (along the Y-axis in upper four images, and along X-axis in the case of the bottom image).

We can conclude that our cryo-EM model resembles the published yeast structures (Nguyen et al., 2015- PDB 5gan; Wan et al., 2016 – PDB 3jcm) in terms of Brr2 location and overall shape. Our 6.7 Å cryo-EM model of the yeast tri-snRNP helps us to confirm the position of Brr2 in the most prominent population of the tri-snRNP and also enables us to localise the position for the Sad1 and other proteins (see Section 3.2.3). In this way, we can analyse possible interactions of factors uniquely present in our particle, which have not been annotated in previously published tri-snRNP models before.

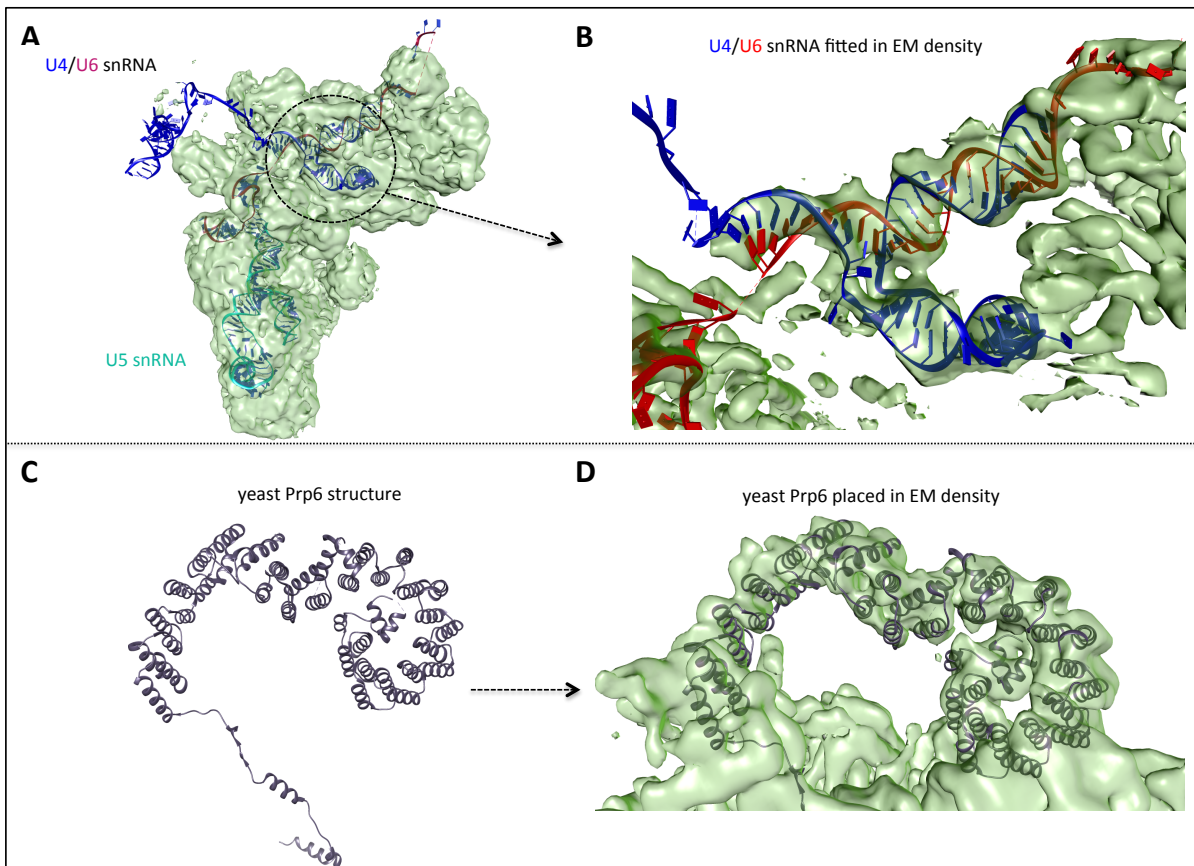
Prior to the arrival of high-resolution cryo structures we used immuno-localization of tagged spliceosomal proteins during negative-stain electron microscopy to determine protein positions in the yeast tri-snRNP and yeast B and B<sup>act</sup> complexes (Häcker et al., 2008); (Sun et al., 2016); (Rigo et al., 2015). It is remarkable how well the positions elucidated with this method match the protein positions revealed after determining the 3D cryo structures of the same complexes. The position of Prp8 and Snu114 in the central body of the tri-snRNP, the location of Brr2 in the head domain, as well as the location of U4/U6 snRNP proteins Prp3 and Prp4 in the arm domain was well confirmed in the cryo structure.

### **3.2.2.3 Placing available X-ray, NMR and EM-reconstructed structures of certain canonical factors in our tri-snRNP model**

In order to understand the arrangement of different proteins in our 3D reconstructed cryo-EM model of the Sad1-TAP tri snRNP, we utilised the auto-aligning tool in the UCSF chimera software. We employed rigid-body fitting of available X-ray, NMR, and EM density-based structures of certain canonical proteins in the electron density map of our 3D cryo-EM model. The software automatically aligned the majority of the protein structures in the electron density map with minor fine adjustments (fig: 3.25, B). In certain regions, where the resolution was sufficiently high (in the core region), proteins and RNA structures were aligned based on their secondary structural elements that were available in the electron density map (fig: 3.26).



**Figure: 3.25 Sad1-TAP tri-snRNP's 3D cryo-EM structure and overview of the molecular arrangement of its proteins.** **A:** Image depicts the two superimposed cryo-EM models: the 6.7 Å resolution unmasked cryo-EM 3D model shown in grey colour, and the final refined 5.8 Å 3D model (obtained after masking the head domain) shown in green colour. Image shows the front view of the tri-snRNP model with head, arm, central body, and foot region. **B:** Image shows the molecular organisation of some of the known tri-snRNP proteins in the 3D cryo-EM map of Sad1-tri-snRNP.



**Figure: 3.26 Sad1-TAP tri-snRNP's 3D cryo-EM structure and overview of the molecular arrangement of its RNA (panel A and B) and Proteins (panel C and D) based on secondary structural features** **A:** Image depicts snRNAs organisation in our cryo-EM tri-snRNP model, with U4 snRNA shown in blue, U6 snRNA in red, and the U5 snRNA is shown in turquoise. **B:** An enlarged view of the U4/U6 snRNA placed in the clearly visible secondary structure elements. **C:** Yeast Prp6 PDB model (Nguyen et al., 2015- PDB 5gan; Wan et al., 2016 – PDB 3jcm; UniProt P19735). **D:** The enlarged view of the yeast Prp6 placed into a clearly visible secondary structure element in the Sad1-TAP tri-snRNP 3D cryo-EM map.

These images convincingly show the proper fitting of different proteins like Prp3, Prp4, Prp6, Prp8, Snu114 and the Sm-ring of U5 snRNA into the electron density and support the reliability of our reconstructed model, as secondary structure elements and electron density match nicely in highly resolved regions. We can now analyse our EM map and compare it with the published tri-snRNP model to identify the unique structural features of the Sad1-TAP tri-snRNP.

### 3.2.3 Novel features of the yeast tri-snRNP structure

As mentioned in the previous section, highly flexible and vigorously moving structural elements are difficult to resolve by EM analysis. In contrast, stable and fixed regions of the structure are relatively easy to resolve. While we were able to reconstruct the highly flexible head domain only

in our 6.7 Å Cryo-EM density map, comparatively stable and fixed structural features (body, foot, and linker domain) were relatively quickly resolved in our 5.8 Å 3D cryo-EM density map. Nevertheless, our structure revealed some novel features of the Sad1-TAP tri-snRNP structure.

In order to compare our 3D cryo-EM model of Sad1-TAP tri-snRNP with the published yeast tri-snRNP model (Nguyen et al., 2015, Wan et al., 2016; Nguyen et al., 2016), we superimposed our cryo-EM electron density map (low-resolution unmasked model in this case) on the published yeast tri-snRNP model (Nguyen et al., 2016; PDB - 5gan) taking the well-resolved lower region as fixed reference region (fig: 3.27) (i.e., relative to Snu114 and the Prp8 NTD1 domain).

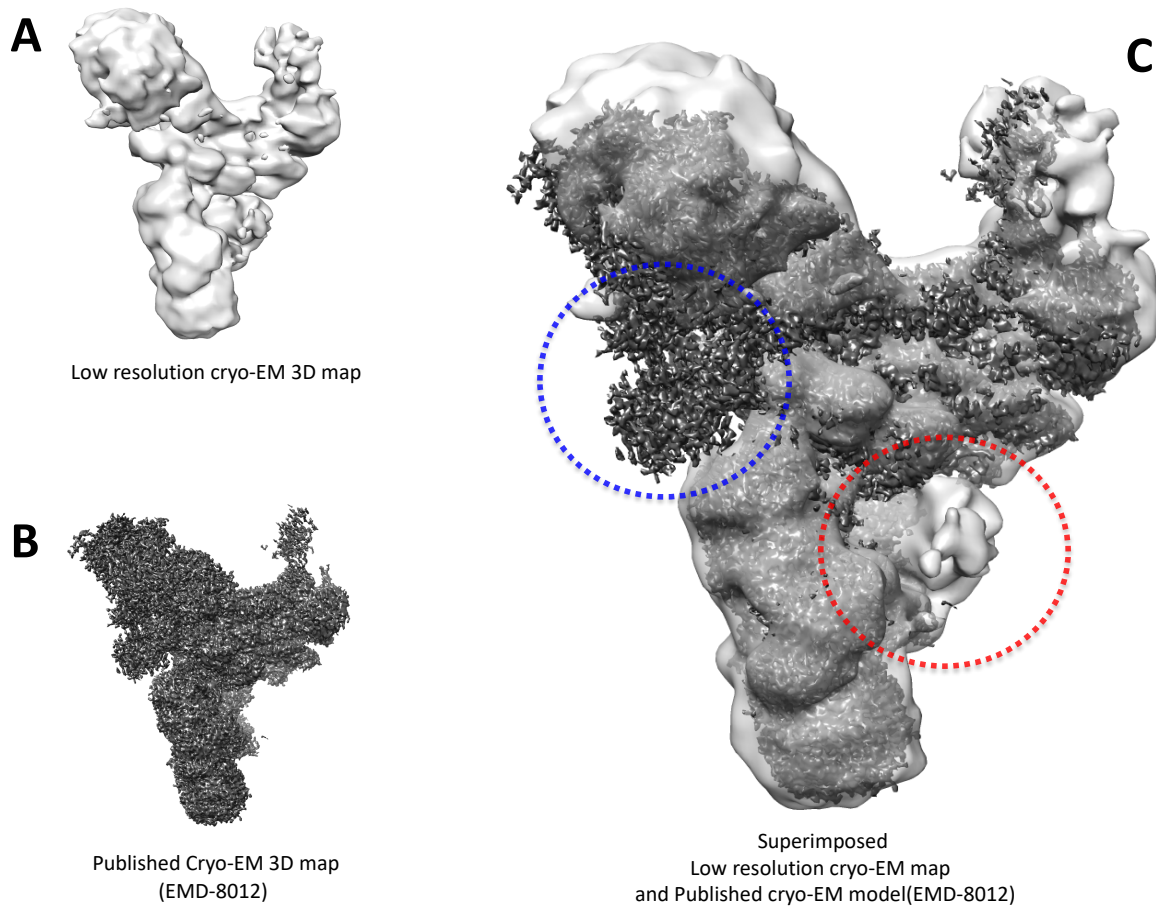
A comparison of the two models revealed some structural features that, for the first time, were exclusively observed in the Sad1-TAP tri-snRNP structure. While comparing with the published structure of the yeast tri-snRNP (Nguyen et al., 2016), we noted that the Sad1-TAP tri-snRNP has a missing density in one of its regions. The missing density region of Sad1-TAP tri-snRNP is located at the base of the head domain – a density region that has been annotated in the published structure (PDB-5gan) as that of the yeast Snu66 protein (fig: 3.27, C; represented by a blue dotted circle). Notably, in a recent study, my colleagues (Bertram et al., 2017) compared the molecular architecture of the yeast tri-snRNP reported by Wan et al. (2016) and Nguyen et al. (2016) respectively, with the human spliceosomal B complex as determined by cryo-EM (Bertram et al., 2017; PDB-5O9Z). The aforementioned study not only showed that the structural organisation of Prp8, Brr2, Prp6 and the U4/U6 proteins in the reported yeast tri-snRNP is analogous to the organisation in the human B complex, but also, that the density which was tentatively assigned to helical regions of yeast Snu66 by Nguyen et al. (2016) more likely comprises the yeast Prp38 complex (Prp38/ Snu23/ Spp381 protein complex), as this position is equivalent to the hPrp38 complex in the human B complex and can even be superimposed (Bertram et al., 2017; PDB-5o9z). Later that year another study which was related to the yeast B complex structure (Plaschka et al., 2017), showed that, indeed, this density was comprised of the yeast Prp38/ Snu23/ Spp381 protein complex.

Secondly, we observed an extra density in the Sad1-TAP tri-snRNP when compared with the same published tri-snRNP structure (fig: 3.27, C; represented by a red dotted circle).

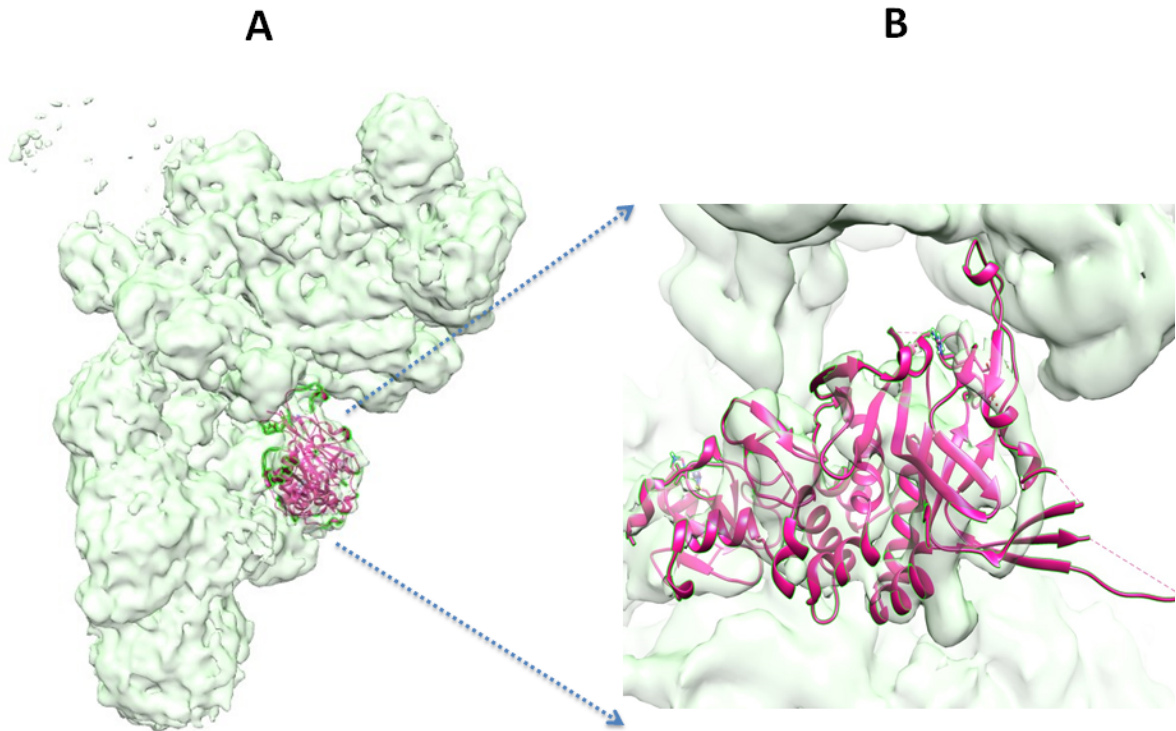
To identify the likely occupant of this extra density in the Sad1-TAP tri-snRNP, we compared the yeast Sad1-TAP tri-snRNP model with the available human tri-snRNP model (Agafonov et al., 2016a). In the case of the human tri-snRNP, the equivalent region was predicted to harbour the human Sad1 protein (PDB-3jcr). To test the possibility of having a similar situation in the yeast tri-snRNP, we used the X-ray crystal structure of the yeast Sad1 protein (PDB-4msx;



(Hadjivassiliou et al., 2014) and performed a rigid body placement in our Sad1-TAP tri-snRNP cryo-EM map. We found that the crystal structure of the Sad1 protein fits very nicely into the unique extra density of the Sad1-TAP tri-snRNP model (fig. 3.28).



**Figure: 3.27 comparing the Sad1-TAP tri-snRNP's 3D cryo-EM structure with the published cryo-EM structure of the yeast tri-snRNPs.** The images show the structure comparison using superimposed features. **A** The low-resolution (8.8 Å, unmasked cryo-EM model of the Sad1-TAP tri-snRNP particle. **B** Published 3D cryo-EM model of the yeast tri-snRNP (EMD-8012). **C** Superimposed view of the two structures (i.e., our low-resolution cryo-EM map with published tri-snRNP map) revealing the missing density (shown here in a blue dotted circle) and the electron density (shown by a red dotted circle) exclusively present in our cryo-EM model of the Sad1-TAP tri-snRNP.



**Figure: 3.28 Sad1-TAP tri-snRNP's 3D cryo-EM structure with ySad1 location. A:** Image showing our cryo-EM model of Sad1-TAP tri-snRNP with the X-ray crystal structure of Sad1 protein placed in the “extra density”. **B:** Enlarged view of the crystal structure of published yeast Sad1 protein (PDB-4msx; Hadjivassiliou *et al.*, 2014) fitted in the extra density.

Though an extra density appeared in our 3D cryo-EM model, the current level of resolution did not allow us to place the Sad1p crystal structure on the basis of the secondary structure elements. Therefore, we have performed protein-protein crosslinking experiments (see section: 3.3) to independently verify our tentative positioning of the Sad1p in this extra density. Crosslinking data clearly supported the stable presence of Sad1 protein in the Sad1-TAP tri-snRNP and its placement in our cryo-EM map. These findings were also in-line with the previously presented biochemical experiments (section: 3.1.4.1) and MS data (table: 3.4 and table: 3.5) in earlier chapters, which showed that the Sad1-TAP tri-snRNP indeed retained Sad1p.

### 3.3 BS3 crosslinking of affinity-purified yeast Sad1-TAP tri-snRNPs

Chemical crosslinking of proteins together with mass spectrometry analysis (CXMS) can provide valuable information about the location of proteins and the identity of their next neighbours. In theory, at least, CXMS can help define the general architecture of a large protein complex by identifying direct binding partners within the complex and localising the binding interface. CXMS also provides unique information about the flexible regions, and therefore it is often combined with other structural techniques (e.g. cryo-EM, X-ray crystallography, NMR) to resolve the structure of large macromolecular assemblies. CXMS can be performed for assemblies of any complexity and often plays an essential role in resolving their compositional and conformational heterogeneity. The numerous strategies of the general CXMS strategy primarily diverge in the structure and the functional groups of the crosslinking substances, as well as in the strategy for data analysis. In this study, two-step purified Sad1-TAP tri-snRNPs were cross-linked with a homobifunctional, lysine-reactive compound BS3 (Leitner et al., 2014) and the protein-protein crosslinks were identified by pLink software (Fan et al., 2015). The CXMS results were filtered to false discovery rate of 1% (FDR 1%) and shown in table S1. High confidence crosslinks are supported by multiple CSMs (Cross-linked peptide - Spectrum Matches, that is, computational matches between measured and *in silico* generated mass spectra returned by specialized CXMS software) and usually display higher score. Thus, obtained crosslinks represents a set of low-resolution structural (distance) restraints for proteins in Sad1-TAP tri-snRNPs that we used to characterise these particles by integrative modelling. In this way, supplementary evidence from a second, independent experimental approach was used to verify structural assumptions being made based on our cryo-EM analysis of the yeast Sad1-TAP tri-snRNP.

#### 3.3.1 Crosslinks supporting the Sad1 locations in the yeast tri-snRNPs

This section presents the details of BS3 crosslinks that were found for Sad1 residues. In human tri-snRNP, Sad1 is stably associated, and from the published human tri-snRNP structure (Agafonov et al., 2016; PDB-3jcr; Charenton et al., 2019; PDB- 6qw6) we know its precise location. On the contrary, in yeast Sad1 was consistently missing from the available published yeast tri-snRNP structures (5gan; 5zwo). Therefore, a direct structural comparison with a published yeast structure is not possible. Our dataset though is compatible with a human-like position of Sad1 in the yeast tri-snRNP.



As an alternative approach, I went back to our cryo-EM model reconstruction of the Sad1-TAP tri-snRNP described in section 3.2.3, fig: 3.28. There we had placed Sad1 in the extra density that was exclusively observed in this cryo-EM density map (fig: 3.27 C, 3.28). The crosslinking experiments were used as an independent means to verify the correct placement of Sad1 in our tri-snRNP model. After measuring the distances between the residues that were cross-linked, we analysed the location of Sad1 relative to its neighbouring proteins. As stated above, when analysing the crosslinking data, we assume a Sad1 location in the yeast tri-snRNP as observed in the human tri-snRNP. For this purpose, the crystal structure of yeast Sad1 is modelled into the published yeast tri-snRNP structure (Nguyen et al., 2016; EMD: 8012) according to its relative position in the human tri-snRNP. We then determined the relative distances of Sad1 to its neighbouring proteins. Crosslinks obtained for Sad1 residues and neighbouring proteins are shown in the table: 3.6

Protein 1 (A)	Residue 1 (B)	Protein 2 (C)	Residue 2 (D)	CSMs (E)	Calculated Distance between residue 1 and residue 2 (in Å) (F)
Sad1	207	Brr2	1529	1	45.5
Sad1	266	Prp6	89	3	< 20
Sad1	270	Prp6	89	10	< 15
Sad1	315	Prp8	810	6	< 20.4
Sad1	319	Prp8	810	7	< 20.9
Sad1	322	Prp8	810	1	< 21.0
Sad1	403	Prp8	1242	12	13
Sad1	73	Snu114	520	3	< 25
Sad1	220	Snu114	617	30	8.1
Sad1	263	Snu114	60	11	< 10

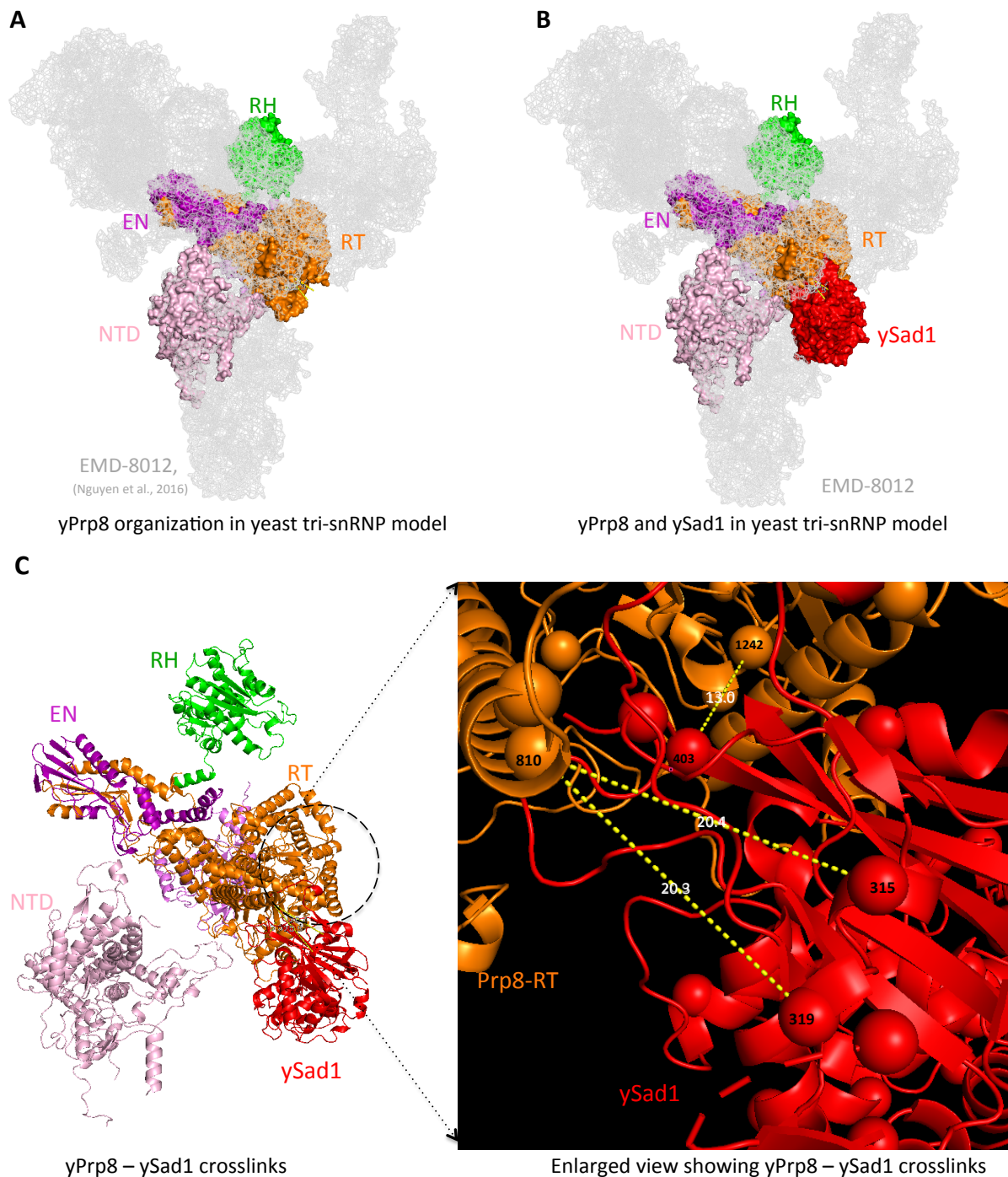
**Table: 3.6 BS3-crosslinks of ySad1 in the yeast tri-snRNP.** CSMs (Cross-linked peptide - Spectrum Matches) of the CXMS data for proteins of the purified yeast Sad1-TAP tri-snRNP. The table only shows some selected protein-protein crosslinks of Sad1 with other proteins. Numbers in the columns B and D indicate the position of the cross-linked lysine. Euclidian C $\alpha$ -C $\alpha$  distances between cross-linked residues are given in Angstrom ("Å") in the column F (calculated distance between the residue-1 of Protein-1 given in column B and residue-2 of protein-2 in the column D). Approximation (</>) of distance was used for regions of the protein that are not in the crystal structure or the EM structure. Column E shows the CSMs, which reflect the number of times a single cross-linked peptide pair was identified in the search query.

The table includes crosslinks of only selected proteins of the tri-snRNP complex, for complete table refer to Supplementary data Table: S1, which includes all protein crosslinks even if they were not observed/modelled into the EM density.

The crosslinking reagent BS3 can crosslink primary amino groups present in lysines and protein N-terminal residues, when their  $\alpha$ -carbon atoms are separated by about 3 to 30 Å. Thus, by visualizing the experimentally observed crosslinks in PyMOL or UCSF Chimera and analysing the distances between cross-linked residues of Sad1 and other tri-snRNP proteins, we obtained constraints, which allowed us to conclude whether the postulated location of Sad1 in our yeast tri-snRNP model is feasible.

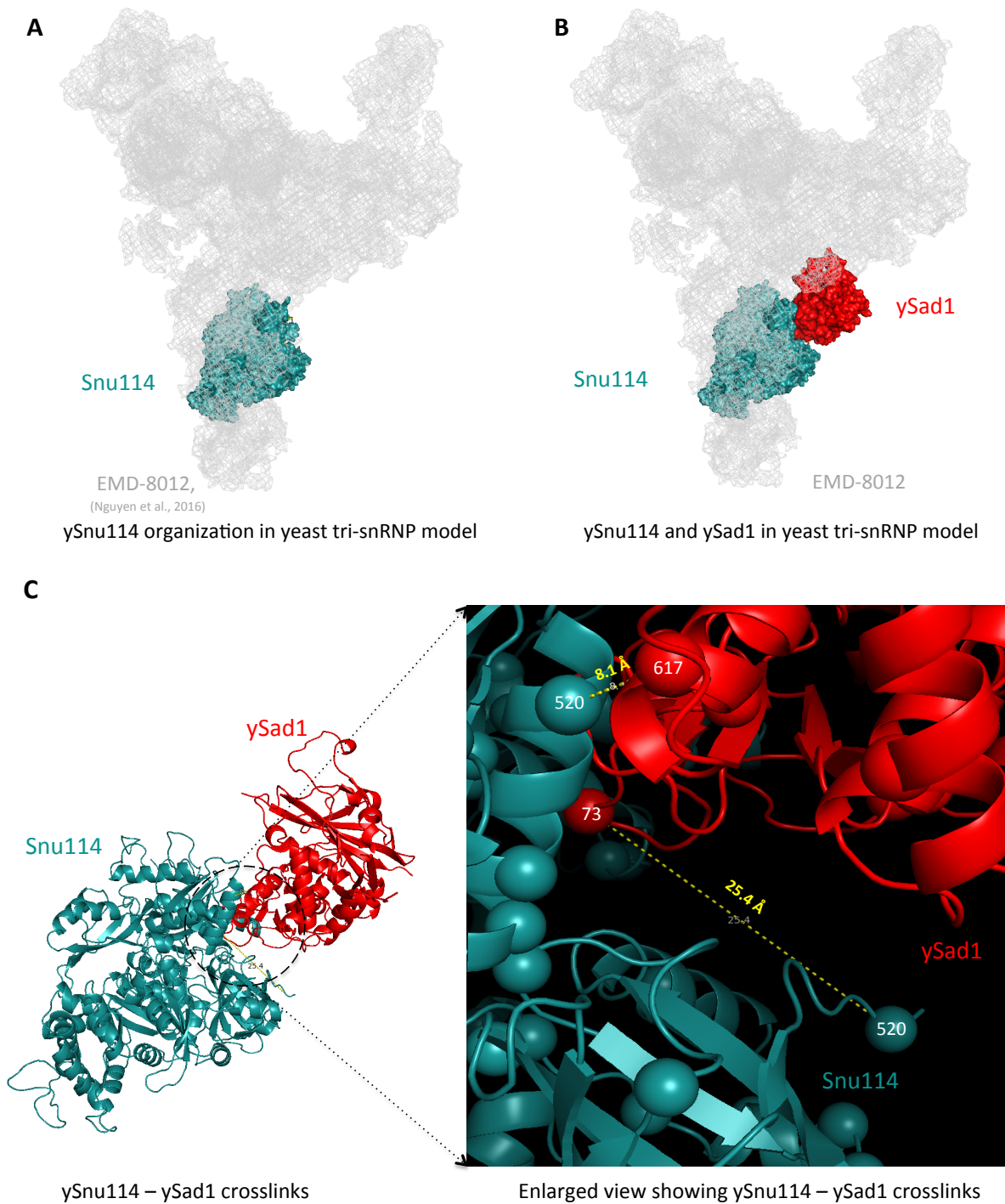
We analysed first the crosslinks between Sad1 and Prp8: all crosslinks between the modelled residues of these proteins (fig: 3.29, C) were in a distance range expected for BS3 crosslinking (see table 3.6), and therefore supported the correct placement of Sad1 in our tri-snRNP model. As can be seen from Table 3.6, all Sad1--Prp8 crosslinks correspond to feasible distances. For example, Sad1 lysine residue 315 cross-linked to Prp8 lysine residue 810, which are separated in our model by approximately 20.4Å, while residue 403 of Sad1 was 13 Å away from the lysine 1242 of Prp8 (table 3.6, and fig: 3.29, C).

Similarly, we analysed crosslinks between Sad1 and Snu114. Fig: 3.30 shows the Sad1—Snu114 crosslinks correspond to feasible distances. For example, Sad1 lysine residue 220 cross-linked to Snu114 lysine residue 617, shown at approximately 8.1 Å, while Sad1 residue 73 was 25.1 Å away from the lysine 520 of Snu114 (table 3.6, and fig: 3.30, C).



**Figure: 3.29 Yeast Sad1 residues cross-linked with Prp8 residues.**

**A:** Image shows the arrangement of different domains of yPrp8 protein (Prp8 domains EN/ RH/ RT/ NTD, in different colours) in the published yeast tri-snRNP EM density map (Nguyen et al., 2016; EMD- 8012), shown in grey. **B:** Description as for image A, except the probable position of the ySad1 is shown in the published yeast tri-snRNP model (EMD-8012, Nguyen et al., 2016). For placement of Sad1, see section 3.3.1. **C:** PDB model of yPrp8 and ySad1 with designated crosslinks between the RT domain and ySad1. The enlarged view of the region of interaction shows the respective lysine residues of ySad1 and yPrp8 with distances between them (shown with yellow dotted lines). For example, the Sad1 lysine residue 315 is cross-linked with yPrp8 lysine residue 810 (the model predicts a distance of 20.4 Å).



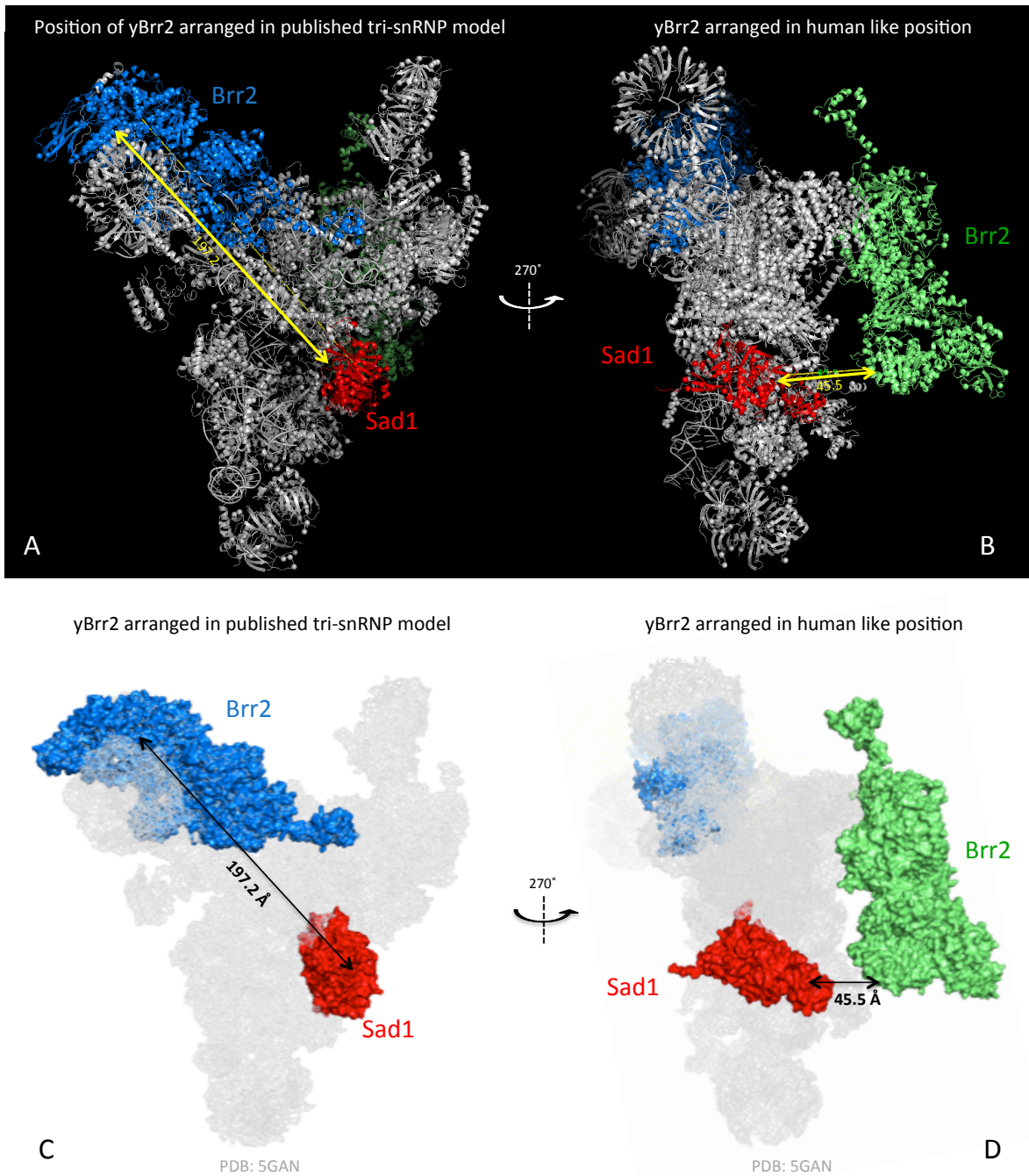
### Figure: 3.30 Yeast Sad1 residues cross-linked with Snu114 residues

**A:** Image shows the arrangement of ySnu114 protein in the published yeast tri-snRNP EM density map (Nguyen et al., 2016; EMD- 8012), shown in grey. **B:** Description as for image A, except the probable position of the ySad1 is shown in the published yeast tri-snRNP model (EMD-8012, Nguyen et al., 2016). For placement of Sad1, see section 3.3.1. **C:** PDB model of ySnu114 and ySad1 with designated crosslinks. The enlarged view of the region of interaction shows the respective lysine residues of ySad1 and Snu114 with distances between them (shown with yellow dotted lines). For example, the Sad1 lysine residue 617 is cross-linked with Snu114 lysine residue 520 (the model predicts a distance of 8.1 Å).

Analysis of crosslinks of Sad1-Brr2 was particularly interesting, as can be seen in fig: 3.31, which shows an example of a crosslink analysis between Sad1 – Brr2 residues. Placing Brr2 in a yeast-like and in a human-like orientation respectively, gives two rather different distances for an amino acid pair. The actual observed crosslinks between Sad1 and Brr2, in this case, can only occur when Brr2 is placed in a human-like position (see fig: 3.31, B and D). In other words, when we placed Brr2 in the human tri-snRNP like position, the distance between the residues of the amino acid pair is much shorter (45.5 Å, and closer to a feasible range of crosslinking distance) than the distance between the same amino acids when placing Brr2 in the yeast-like conformation (the distance is then too large for crosslinking to occur, i.e., ~197.2 Å, see fig: 3.31, A, and C).

Unfortunately, the protein regions of Sad1, which contain the cross-linked residues, are not available in the PDB file, so we cannot measure the distance with certainty. Nevertheless, these observations are compatible with a model, in which at least a sub-population of particles contains a Brr2, which is arranged in the human like conformation.





**Figure: 3.31 Sad1 residue crosslinks with Brr2 residues**

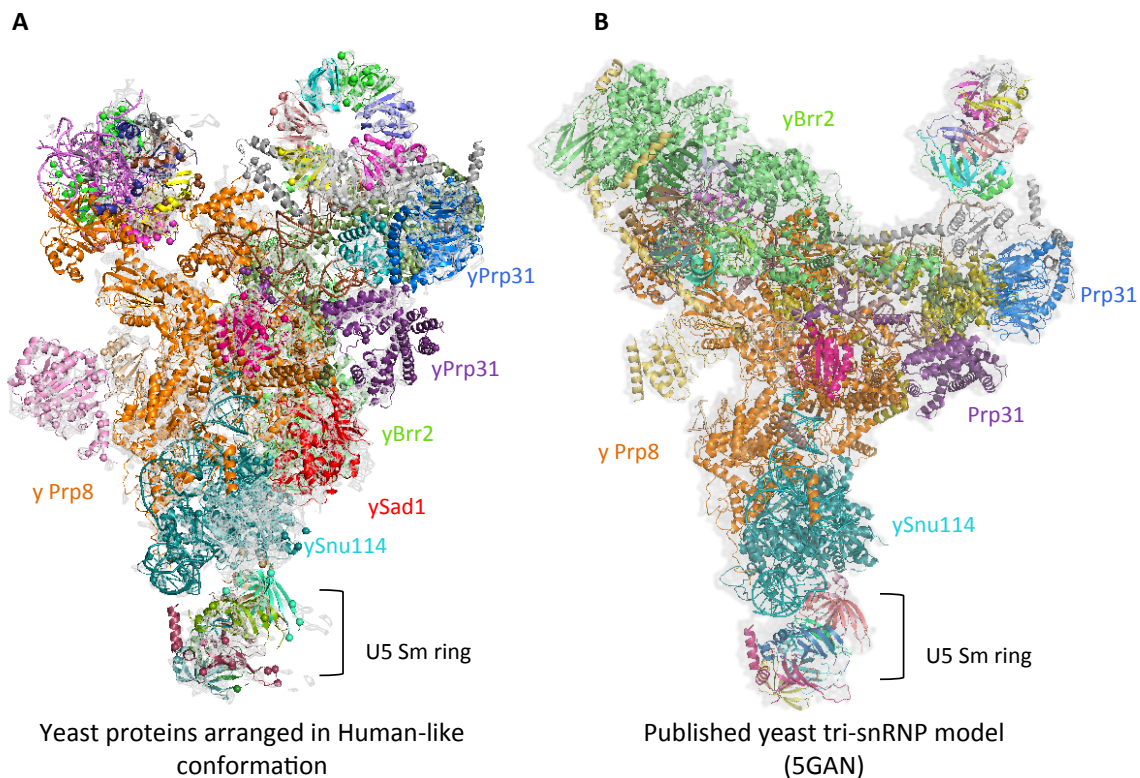
**A:** Image shows the arrangement of yBrr2 in the published yeast tri-snRNP (PDB-5gan) structure (highlighted in blue), the Sad1 crystal structure (PDB: 4msx- shown in red), and yBrr2 placed in the position of its human ortholog (shown in green). Distances between the Sad1 residue and the yBrr2 residue are indicated by a yellow arrow ( $\sim 197.2 \text{ \AA}$ ). **B:** The image shows the Sad1 residue cross-linked to yBrr2 residues. The distance between residues is highlighted by a yellow arrow ( $45.5 \text{ \AA}$ ). The curved arrows denote the direction and the degree of rotation ( $270^\circ$ ) when turning the model, and the vertical dotted line represents the plane of rotation (along the Y-axis). Image C and D are simplified views of image A, and image B, respectively.

Based on these results, we can place Sad1 with some confidence in a region of the yeast tri-snRNP which corresponds to the position described for the human tri-snRNP structure (Agafonov et al., 2016; PDB-3jcr; Charenton et al., 2019; PDB- 6qw6). In section 3.2.3 we have shown that our 3D cryo-EM structure (fig: 3.27, C) of the purified Sad1-TAP tri-snRNP contains an extra density in this location, analogous to the human Sad1 location. The crosslinking results strongly support this placement of Sad1 in our 3D cryo-EM structure of the Sad1-TAP tri-snRNP.

### 3.3.2 Building a human-like conformation model of the yeast tri-snRNP

As mentioned in the previous section, the observed Sad1-Brr2 crosslink was not consistent with our tri-snRNP model. It was also noted in this and previously published cryo-EM studies (Nagai, Shi ref) that Brr2 is not stably fixed and prone to movement in tri-snRNPs. Therefore, we asked whether our purified tri-snRNP particles represented a mixed population with differently positioned Brr2. Here we applied an in-silico approach involving model reconstruction in the UCSF Chimera software to generate an alternative, human-like model of the yeast Sad1-TAP tri-snRNP and analysed it against the our BS3 crosslinking data (see methods section: 2).

The comparison process entailed the following steps: First, all available known yeast tri-snRNP protein structures (X-ray crystal structures or EM structures) were imported as PDB files in PyMol (<http://www.pymol.org>) or the UCSF Chimera software. Afterwards, the PDB structure of each single yeast tri-snRNP protein was structurally aligned to its human ortholog (the human tri-snRNP structure PDB-3jcr was used as a reference). In this way, we rearranged the yeast tri-snRNP proteins in a human-like form. We used this in-silico **“human-like yeast tri-snRNP model”** (fig: 3.32, A) for comparisons with the our (Section 3.2.2.2) and published yeast tri-snRNP structures (PDB-5gan) (fig: 3.29, B).



**Figure: 3.32 Rearranging the yeast proteins in the “human-like conformation” for comparing crosslinks.**

First, the published PDB file of the human tri-snRNP model (Agafonov et al., 2016) was taken as the starting point. Then various known yeast tri-snRNP protein structures were arranged in this PDB according to the position of their human counterparts. Thus, the available protein structures were aligned in a map that resulted in a final model, which had human-like overall conformation, however, contained yeast tri-snRNP proteins. We named this model "Human-like yeast tri-snRNP model". **B**: The published yeast tri-snRNP model (Nguyen et al., 2016, PDB: 5gan).

Using PyMOL I mapped crosslinks from our CXMS data set, determined and compared the distances between residue pairs (protein A residue 1-protein B residue 2) as found in the “human-like yeast tri-snRNP model” and the published model (Nguyen et al., 2016, PDB: 5gan). Hence, by determining actual distances between two cross-linked residues in two distinct proteins for alternative models, we could argue in favour or against one of them. (See colour-coded columns F and G of the table: 3.7). As expected, the majority of the crosslinks was found to be compatible with both models, i.e., the published yeast-like tri-snRNP conformation as well as the “human-like tri-snRNP” conformation.

Intriguingly, certain crosslinks, such as the Brr2 residue 445 cross-linked with residue 307 of Prp3, are only possible if Brr2 and Prp3 are arranged as in the published yeast tri-snRNP



structure and not as in the human-like conformation. On the other hand, Brr2 crosslinks with Prp6 and Prp8 can only be explained by assuming a human-like arrangement of these two proteins. The fact that we have found crosslinks supporting both conformations are indicative that yeast SAD1-TAP tri-snRNP preparations may contain a subpopulation of tri-snRNPs with human-like conformations.

Protein 1 (A)	Residue 1 (B)	Protein 2 (C)	Residue 2 (D)	CSMs (E)	Distance between residue 1 and residue 2 (in Å)	
					Yeast model (5gan) (F)	Human-like yeast tri-snRNP model (G)
Brr2	445	Prp3	307	2	16.3	100.2
Brr2	457	Prp3	261	6	22.6	114.4
Brr2	717	Prp3	261	8	28.8	107.3
Brr2	414	Prp6	183	4	45.5	122.6
Brr2	445	Prp6	351	2	102.3	9.7
Brr2	445	Prp6	356	8	98.8	7
Brr2	445	Prp6	360	11	93.9	11.3
Brr2	454	Prp6	356	2	105.4	15.2
Brr2	454	Prp6	360	20	100.6	16.6
Brr2	1676	Prp6	570	2	162.1	104.6
Brr2	414	Prp8	1864	9	70.3	54
Brr2	1055	Prp8	2154	18	92.6	22.9
Brr2	1060	Prp8	2149	2	98	16.6
Brr2	1529	Sad1-TAP	207	1	197.2	45.5
Brr2	956	Snu114	520	1	> 200	< 100
Brr2	1623	Snu114	947	1	> 200	< 90
Brr2	1634	Snu66	486	131	> 200	NA

**Table: 3.7 BS3-crosslinks of yBrr2 in the tri-snRNP.** CSMs (Crosslink-Spectrum Match) of the CXMS data for proteins of the purified yeast Sad1-TAP tri-snRNPs. Table only shows some selected intermolecular crosslinks of yBrr2 proteins with other factors. Numbers in Residue 1 (column B) and Residue 2 (column E) indicate the position of the cross-linked lysine residues. Euclidian Ca-Ca distances between cross-linked residues are given in Angstroms ("Å") in column (F) (as measured in the published yeast model), and in column G (in the human-like yeast tri-snRNP model). The table includes only crosslinks of Brr2 with selected proteins of the tri-snRNP complex, for a complete table see Supplementary data Table: S1 which includes all the protein crosslinks, also those not observed/mapped into the EM density. NA- Not Annotated (Protein was not modelled). Green boxes- crosslinks supportive of the human-like conformation of the yeast tri-snRNP. Yellow boxes – crosslinks only feasible in the yeast-like conformation. Red boxes – crosslinks support other arrangements not supported by either model. Approximations (</>) were used for regions of the protein that are not in the crystal structure or the EM structure.



## 4 Discussion

I would first like to briefly recapitulate and review the current thinking about tri-snRNP-related topics before discussing the work presented here. When the first cryo-EM structures of yeast tri-snRNPs were published in 2015/2016 by Nagai and colleagues (2x Nguyen et al.), a purification protocol using 150 mM K<sup>+</sup> was adopted. All tri-snRNP proteins were at least partially modelled, except for Lin1, Prp38, Snu23, Spp381, Sad1, Prp28 and apparently Prp38. The latter protein was correctly annotated only later, because the EM density was first erroneously assigned to Snu66. The authors of this structure were well aware of the surprising fact that Brr2 is already preloaded on U4 snRNA and in close contact with the U4 snRNA region between the 3' stem loop and stem1. The purified tri-snRNP in fact disintegrates upon contact with ATP, indicating that Brr2 is already in an active state.

Shi and colleagues published a structure of the yeast tri-snRNP (Wan et al. 2016) with an overall structure as reported by the Nagai lab. They also used 150 mM K<sup>+</sup>, the tri-snRNP disintegrates in the presence of ATP and for Prp38, Snu23, Spp381 and Sad1 no peptides are detected during MS analysis.

2016, Lührmann and colleagues published the tri-snRNP structure of the human tri-snRNP (Agafanov et al., 2016). In this tri-snRNP, Brr2 is in a dramatically different position compared to the yeast tri-snRNP. The active N-terminal cassette of the Brr2 helicase is kept at a safe distance of about 10 nm to the U4/U6 duplex. Given the usually high degree of evolutionary conservation observed in the splicing machinery, this is a surprising difference to yeast. Most remarkably, the human tri-snRNP is resistant during ATP incubation. The authors speculate that this may be due to the stoichiometric presence of the Sad1 protein in the purified complex, a model which is supported by the strategically position Sad1 occupies in the tri-snRNP structure, a position which clearly stabilizes the U4/U6 – U5 interface and tethers Brr2 in its inactive state. The authors further suggest, that during activation of the human B complex, Sad1 release allows Brr2 to assume a new conformation and position close to the U4/U6 duplex substrate so that U4 release and Prp28 mediated transfer of the 5' SS from U1 to the U6 (ACAGAGA) coincide. The authors suggest that the structural investigation of a Sad1 containing yeast tri-snRNP may provide a solution to this problem.

In 2017/2018 several publications discussed this obvious difference between the yeast and human tri-snRNP in the larger context of tri-snRNP integration during the formation and

activation of pre-B and B complexes. A scenario emerged in which the yeast tri-snRNP may even first assume a human-like structure (i.e. with a 'securely locked' Brr2 protein), containing Prp28 and Sad1. Upon contact of the U1 snRNP with the tri-snRNP, the RecA domains of Prp28 clamp the U1 snRNA thus disrupting the U1 interaction with the 5' SS. The latter contacts the U6 ACAGAGA motif, U1 leaves, Prp28 dissociates and its position is taken by Prp38 and Snu23. The latter contact the ACAGAGA/5' SS duplex, Brr2 assumes its active conformation and loading its active helicase cassette onto the U4 RNA it melts the destabilized U4/U6 duplex (reviewed in Kastner et al. 2019 CSH).

In the yeast tri-snRNP, as well as in the yeast and human B complexes, Brr2 is already loaded on its U4 snRNA substrate (2x Plaschka et al. / Bertram et al., 2017). It has therefore been suggested that in yeast, Brr2 may already assume this conformation at the level of the tri-snRNP, with or without a brief period of a conformation as observed in the isolated human tri-snRNP.

## 4.1 Biochemical aspects

Spliceosomal complexes in the past were investigated primarily during biochemical and functional studies. However, following the resolution revolution in electron microscopy, numerous spliceosomal complexes were investigated using EM. Structural biologists, working in the field of pre-mRNA splicing have seen a sharp rise in EM based structural studies of spliceosomal complexes, providing insights far beyond the level of single protein structures.

The human tri-snRNP first is only loosely attached to the 37S pre-B complex. The stable integration of the tri-snRNP following Prp28 action is accompanied by substantial structural rearrangements of most proteins turning the pre-B complex into the 45 S B complex. The fact that Brr2 in yeast is already loaded on the U4 snRNA in the tri-snRNP, not only gave rise to speculations about possible differences in mechanism, but also raised doubts about the quality of tri-snRNP preparations used.

### 4.1.1 An improved tri-snRNP is required to solve the question of significant differences in the published structures of tri-snRNPs.

One of the most significant biochemical differences in the biochemical behaviour of yeast and human tri-snRNPs concerns their stability upon incubation with ATP. While the purified human

tri-snRNP is stable (Behrens et al., 1991; Stevens et al., 1999; Absmeier et al., 2015), the yeast tri-snRNP is completely dissociated due to ATP-dependent Brr2-mediated dissociation. The latter may be interpreted simply as a functional difference between yeast and human tri-snRNPs. But it may as well be due to the obvious differences in the biochemical composition of the human and yeast tri-snRNPs as purified using the currently available protocols. This in turn may have a severe effect on the structure of tri-snRNPs. Interest in the structure and composition of the tri-snRNPs stems also from the fact that the answer to these questions holds the key for understanding the detailed mechanisms by which the tri-snRNP joins the A complex and thus forms and integrates into the nascent B complex. I will return to this point later in this discussion.

It was long known that the tri-snRNP associated yeast protein Sad1 is critical for splicing of nearly all intron-containing genes and required for the assembly of U4 snRNA into the U4/U6 particle (Lygerou et al., 1999; Hadjivassiliou et al., 2014). That Sad1 is important for maintaining the integrity of the yeast tri-snRNP was first observed when, in the presence of ATP, yeast tri-snRNPs completely dissociated into U4/U6 di-snRNPs and U5 snRNP upon depletion of Sad1 from yeast whole cell extracts (Huang et al., 2014).

Human Sad1, which is stoichiometrically present in the purified tri-snRNP, locks Brr2 in a pre-activated position away from U4 snRNA (Agafonov et al., 2016). The yeast Sad1 is mostly lost during purification of yeast tri-snRNPs (Fabrizio et al., 1994). Its presence indeed was not reported for both published yeast tri-snRNP structures, since the traditional protocol was used. It therefore became obvious that an improved purification of yeast tri-snRNPs, possibly with a stable presence of Sad1, needed to be a primary goal. Such a tri-snRNP has the potential for rather different structural features compared to those tri-snRNPs originating from a conventional purification procedure.

#### **4.1.2 A new purification strategy for yeast tri-snRNPs.**

The original purification protocol for yeast tri-snRNPs used a TAP-tagged Brr2 strain for extract preparation and a three-step purification, which comprises IgG- and calmodulin-affinity chromatography and a final glycerol gradient fractionation (Häcker et al. 2008), purification is performed under 150 mM [K<sup>+</sup>] conditions. The published reports of yeast tri-snRNP cryo-EM structures used essentially the same buffer and salt conditions:

Lab. group (organism)	<b>Nagai's Group</b> <i>S. cerevisiae</i> (Nguyen et al., 2015, 2016)	<b>Shi's Group</b> <i>S. cerevisiae</i> (Wan et al., 2016)	<b>Lührmann Group</b> Human tri-snRNP (Agafonov et al., 2016)
Purification strategy	Brr2 TAP tag 2 affinity steps + glycerol gradient	Prp6 TAP tag gel filtration + 2 affinity steps	Zonal centrifugation (sucrose gradient) + m3G-cap Ab affinity step + sucrose gradient.
Salt conditions	150 mM [Na <sup>+</sup> ], Hepes pH 7.9	150 mM [Na <sup>+</sup> ], Hepes pH 7.9	150 mM [K <sup>+</sup> ], Hepes pH 7.9
ATP sensitivity	dissociates with ATP	dissociates with ATP	stable with ATP
Features	no Prp28, no Sad1, no Prp38* [* was identified later]	no Prp28, no Sad1, no Prp38	hPrp28 and hSad1, no Prp38

My revised protocol now uses a Sad1-TAP-tagged strain for extract preparation and only a two-step purification, comprising an IgG-affinity step and the final glycerol gradient fractionation. Moreover, the protocol uses 75 mM [K<sup>+</sup>] in a Hepes pH 7.3 (4°C) buffer. Quality criteria applied were the appearance of the particle during EM visualization, mass spectrometric analysis of the protein composition, visualization of snRNAs and the stability of the particle in the presence of ATP.

Lab. group (organism)	<b>This work</b> <i>S. cerevisiae</i>
Purification strategy	Sad1 TAP tag IgG affinity step + glycerol gradient
Salt conditions	75 mM [K <sup>+</sup> ], Hepes pH 7.3
ATP sensitivity	stable in the presence of ATP
Features	all canonical proteins including Sad1 and Prp28 (but very little Prp38)

The mass-spec analysis of particles purified in this way shows a surprising picture when compared to all previously reported protein inventories of the yeast tri-snRNP (including those for the available cryo-EM structures). The list of canonical tri-snRNP proteins was first firmly established in 1999 (Gottschalk et al. EMBO J. 1999; Stevens and Abelson, PNAS 1999). Only two other proteins were later added to this list, the notoriously elusive Prp28 and Sad1 proteins

(Staley and Guthrie 1999; Lygerou et al., 1999). Our purified yeast tri-snRNP indeed contains all of the known canonical yeast tri-snRNP proteins in clearly detectable quantities, even Sad1 and Prp28 ! Fig. 3.13 clearly shows that nearly all tri-snRNP proteins are clearly present in the tri-snRNP peak (except perhaps for Lin1) with a very low background in the rest of the gradient.

#### 4.1.3 The Prp38 – Snu23 – Spp381 group of proteins.

In yeast, the tri-snRNP specific proteins Prp38, Spp381 and Snu23 form a group of tightly interacting proteins (see Table 1.3). Until now it was always assumed that the three proteins enter the yeast tri-snRNP and stay associated with the tri-snRNP during tri-snRNP integration and B complex formation and leave during the transition to B<sup>act</sup> (Fabrizio et al. 2009). Details about the exact sequence of entering the tri-snRNP are not known. In human instead, the corresponding orthologs hPrp38, hSpp381 and MFAP (Ulrich and Wahl 2017) are present exclusively at the B complex stage. They enter the spliceosome independent of the human tri-snRNP and are therefore classified as B-specific proteins (Agafonov et al 2011; Boesler et al. 2016). All three yeast proteins are present in our purified tri-snRNP, even though with lower peptide counts when compared to Prp28 and Sad1. The mode of recruiting these three proteins to the spliceosome appears therefore to be intrinsically different between yeast and human. Prp38 is a protein required for the release of U4 snRNA (Lybarger 1999). It interacts with Spp381 (which in turn contacts Brr2), Snu23 and the NTD of Prp8.

The integrity of this particle is furthermore confirmed by quantitative and qualitative analysis of snRNAs U4, U5 and U6. The high quality of the purified tri-snRNP is due to the fact that Sad1 is stably integrated and that in turn the Sad1-tag can be used for purification. Stable Sad1 integration is due to the overall design of the purification in particular to the use of only 75 mM [K<sup>+</sup>]. The detailed analyses shown in section 3.1.4.6 clearly show the detrimental effects of elevated salt concentrations for the integrity of the yeast tri-snRNP:

Under 75 mM [K<sup>+</sup>] conditions all U4/U6 and U5 proteins as well as the tri-snRNP specific proteins are firmly centered in the tri-snRNP peak with little material in the rest of the gradient. The purification of course is based on Sad1 protein affinity, but it is nevertheless remarkable that the presence of debris and free Sad1 is minimal. The fact that also the yield is significantly improved indicates a real net increase in Sad1 containing tri-snRNP.

With increased salt concentrations Sad1 and other tri-snRNP-specific proteins start to migrate into the higher fractions of the gradient indicated the gradual 'meltdown' of the tri-snRNP

(Fig.3.14). In the human tri-snRNP structure the hSad1 occupies a position where it can act as a bridging protein between the U5 and U4/U6 proteins by interacting with Snu114 - Brr2, and Prp31 – Prp8, respectively (Agafonov et al. 2016). Such a behaviour is consistent with data shown in Fig 3.14. The evaluation of tri-snRNP quality is therefore not simply a question of presence/absence of Sad1, the absence of Sad1 has also far reaching consequences for the stable attachment of other proteins and the orientation they assume in the complex. The stabilizing role of Sad1, proposed already years ago, is therefore fully confirmed (Huang et al., 2014).

Yeast tri-snRNPs purified according to the new procedure provide an ideal starting material for attempts to solve the structure of the yeast tri-snRNP using cryo-EM techniques.

## 4.2 Structural aspects

### 4.2.1 Negative-stain EM of yeast tri-snRNPs.

We gathered sufficient biochemical evidence, which compelled us to modify the typical standard yeast tri-snRNP protein profile. The tri-snRNP now contains stably associated Sad1 and Prp28, while the amounts of Prp38 remain comparatively low. Also the fact that the yeast tri-snRNP is now stable under ATP conditions makes the yeast particle more similar to its human counterpart. Therefore it was interesting to see whether the Brr2 would assume a conformation in our Sad1-TAP tri-snRNP like observed for Brr2 in the human tri-snRNP.

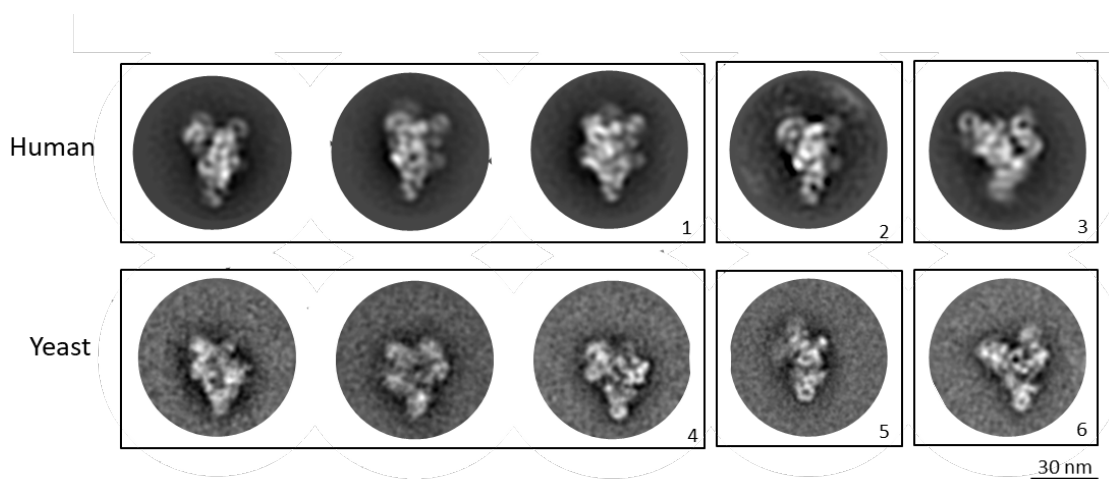
We therefore set out to investigate the structure of the yeast tri-snRNP purified via Sad1-Tag. We started preliminary structure investigation with negative-stained EM analysis. Tri-snRNP particles were accordingly purified, fixed, and negatively stained using a well-established protocol for negative EM analysis. Fixed purified particles were adsorbed on EM grids, stained for suitable duration, and later inspected under a CM 200 electron microscope. In collaboration with my colleagues, we performed EM analysis, and collected negative stain-EM micrographs of Sad1-TAP tri-snRNP particles. Micrographs collected from the negative stain-EM were subjected to software-based as well as manual particle selection. The refined data set enabled us to obtain a number of 2D class-averages of the Sad1-Tap tri-snRNP. The entire dataset generated close to 250 different 2D class-averages with each class consisting of particles with a similar view.

The 2D classes from negative-stained micrographs displayed structural details of the triangular particles with high contrast. Visible is a main 'body' with pointed lower end and a broad 'head'



structure along with a small 'arm' like structure connected to the main central body by a tiny linker region (fig: 3.17, B). The most prominent classes showed similar structural features that resemble an overall view of the low resolution tri-snRNP model, which had already been reported by our laboratory (Häcker et al., 2008). The two major class-averages showed two different prominent views: one with the 'closed' conformation, and the other with a more 'open' conformation. 'Open' and 'closed' conformation relates to a different positioning of the head and the arm domains relative to each other on a fixed central main body (fig: 3.17, C and D). These results are consistent with the previous findings supporting the dynamics of tri-snRNP particles with respect to the highly flexible 'head' domain.

Later we compared the 2D class-averages of the yeast SAD1-TAP tri-snRNP with the 2D class-averages obtained from negative EM studies of the human tri-snRNP (human negative stained EM 2D classes were kindly provided by Cole Townsend, Department for Structural Dynamics, MPI-BPC).



**Figure 4.1 Comparison of 2D class-averages of Sad1-TAP purified yeast tri-snRNPs with human tri-snRNP 2D class-averages.** Upper panel; shown here are various predominant views (2D classes) obtained from the negative-stain EM 2D data analysis of human tri-snRNP particles. With three different views for somewhat similar and slightly rotated views in upper panel box-1, and in box 2 and 3, other prominent views in the negative-stain EM human tri-snRNP data set. Lower panel: shown here are the selected final representative 2D class averages of yeast Sad1-TAP tri-snRNP particles that may bear similar prominent features as observed in the case of human tri-snRNPs. Lower panel box 4 contains three different views that are potentially in a similar orientation as shown for the human tri-snRNP in the upper panel box 1. Lower panel box 5 and box 6 also show striking resemblance to representative views of the human tri-snRNPs in box 2 and box 6 respectively (human tri-snRNP 2D class average pics were kindly provided by Cole Townsend - Department of Structural Dynamics, MPI-BPC)

Interestingly, 2D comparison revealed the existence of particles that populate some subclasses with few members only. These subclasses harbour particles with compact topographies, which resemble the typical features of the most prominent class-averages of the negatively-stained human tri-snRNPs particles (see fig: 4.1.). One has to keep in mind though, that 2D similarities do not permit any conclusions related to possible 3D similarities. The predominant views in the class-averages of Sad1-TAP tri-snRNP particles show an overall view similar to the reported yeast tri-snRNP particles. These observations, with all due caution, could possibly be interpreted as an equilibrium between a dominant yeast-like conformation of the tri-snRNP and a minor population of human-like tri-snRNPs. Support for such a scenario stems from crosslinking data collected for the Sad1-TAP tri-snRNP and described in sections 3.3.1. and 3.3.2. Not only were these crosslinks helpful in determining the location of Sad1, but certain crosslinks and crosslinking distances were clearly only compatible with a sub-population of yeast tri-snRNPs in an assumed human-like conformation. These crosslinks made it even possible to design *in silico* a "human-like" yeast tri-snRNP.

Using the same negative-stain EM dataset, we used the 3D electron density map of the Sad1-TAP tri-snRNP to interpret structural features with more confidence. The dataset resulted in the reconstruction of a 3D model very similar to the published 3D structures of the yeast tri-snRNP. It was therefore even more interesting to pin-point the location of factors (such as Sad1 and Prp28) which are exclusively associated with our Sad1-TAP tri-snRNP purified according to the new protocol. The 3D model reconstruction with negatively-stained EM particles had the disadvantage that we could only observe the overall features. In order to understand the location and orientation of certain protein factors which had never been annotated in published yeast tri-snRNP models before, we needed to further analyse the Sad1-TAP tri-snRNPs by cryo-EM with hoping to get a higher resolution 3D model.

The 3D model derived from negative-stained EM of Sad1-TAP tri-snRNPs, again confirmed that Sad1-TAP tri-snRNP particles have a Brr2 in a similar orientation as already shown in the published yeast tri-snRNP model. However, we could not reconstruct a 3D model using the limited number of particles in minor sub-classes. (particles with unique human-like orientation).

#### **4.2.2 Brr2 orientation in yeast Sad1-TAP tri-snRNP**

In order to obtain a higher resolution 3D model of our Sad1-TAP tri-snRNP, a cryo-EM based approach with protocols that have been successfully employed in our lab, was used. During the initial stage of data refining, we obtained a structure with a resolution of 6.7 Å. Comparing it with the known tri-snRNP models (Agafonov et al., 2016a; Nguyen et al., 2015), even at this relatively low resolution we could clearly see, that despite of having an integral yeast tri-snRNP resistant to ATP-dependent degradation, the determined structure still is governed by a vast majority of particles containing Brr2 already loaded on its U4/U6 substrate. This might be due to a number of reasons: first, the surrounding ion concentration in the buffer environment still favours such a flip of Brr2 relative to the position in human tri-snRNPs. Secondly, certain tethering protein factors may still be unable to fully exert their stabilising effect on Brr2 to prevent it from flipping, although the tri-snRNP is already sufficiently stabilised to prevent ATP-dependent Brr2-mediated dissociation. The resistance of the tri-snRNP to ATP-dependent degradation alone does not give cause to a stabilisation of Brr2 in a position far away from the U4/U6 duplex.

To exclude some more trivial factors like temperature and other in vitro conditions as reasons for the preferred Brr2 orientation in yeast, would require some additional efforts to optimize purification conditions.

The yeast tri-snRNP particle may simply exist in at least two conformational states with respect to Brr2 conformation; an early one with a human-like Brr2 conformation, and another, dominant state as reported in published structures of the yeast tri-snRNP. Brr2, as observed in the yeast tri-snRNP, already assumes a conformation which the human Brr2 only assumes after incorporation into the pre-B spliceosome (Agafonov, Kastner et al. 2016 ; Charenton, Wilkinson et al. 2019). Our 3D reconstruction of the Sad1-TAP tri-snRNP model offers an opportunity to analyse the location of uniquely associated factors of the Sad1-TAP tri-snRNP.

#### **4.2.3 The Sad1 organisation in yeast tri-snRNPs is similar to its human counterpart**

Whereas the position and orientation of Brr2 was fairly clear, other vital regions of the low-resolution 3D cryo-EM model required additional efforts to achieve a better resolution in order to localize certain factors that have been found to be uniquely associated with the Sad1-TAP tri-

snRNP (section: 3.1.4.6). Further comprehensive refining of the cryo-EM data, including the application of stable regional masks resulted in a 3D electron density map with an overall resolution of 5.8 Å. This level of resolution allowed us to fit in the structures (PDB models) of some of the known yeast tri-snRNP proteins with a certain degree of confidence (section: 3.2.2.3). While placing known protein PDB model in our electron density map we took into consideration the known positions of analogous proteins or their orthologs in the human tri-snRNP. The completed construction of our 3D cryo-EM model of the yeast Sad1-TAP tri-snRNP, was then superimposed with the published yeast tri-snRNP model (Nguyen et al., 2015, Nguyen et al., 2016; PDB - 5gan), taking the well-resolved lower region of the particle as fixed reference point (fig: 3.27) (i.e., relative to Snu114 and the Prp8 NTD1 domain).

Interestingly, this comparison of the two models not only confirmed the position of Brr2 as described in the preceding section, but also revealed some differences observed in the cryo-EM map of the Sad1-TAP tri-snRNP (section: 3.2.3). One of the differences was the presence of an extra density in the Sad1-TAP tri-snRNP (fig: 3.27, C). The identification of a likely resident of this extra density was assisted by the available human tri-snRNP model (Agafonov et al., 2016). The additional density region uniquely present in our Sad1-TAP tri-snRNP likely corresponds to the equivalent region in the human tri-snRNP that harbours the human Sad1 protein (Agafonov et al., 2016; PDB-3jcr; Charenton et al., 2019; PDB- 6qw6). Equipped with this information, a rigid body placement of the available X-ray crystal structure of the yeast Sad1 protein (PDB-4msx; Hadjivassiliou et al., 2014) was carried out, revealing a reasonable fit into the extra density region of our 3D cryo-EM map (fig: 3.28). Afterwards, we independently confirmed its correct positioning in our EM map with the help of BS3 cross-linking experiments that have been described in the results section 3.3. Data analyses of the BS3 crosslinking experiments further confirmed that the relative positioning of the Sad1 in our EM density model with reference to several adjacent neighbouring proteins was indeed correct. Distinct crosslinks between lysine residues of Sad1 and Prp8 (fig: 3.29), as well as between Sad1 and Snu114 (fig: 3.30) lysine residues (fig: 3.29, C) were detected. These were in a distance range expected for BS3 crosslinking (see table 3.6), and were only plausible when Sad1 was placed in the defined extra density region. Selected crosslink examples that were vital for the validation of our Sad1 positioning have already been discussed in section: 3.3.1.

#### **4.2.4 The lack of Prp38 in yeast Sad1-tri-snRNPs may correlate with a missing region in the structure.**

Beside the extra density, the comparison with the published yeast tri-snRNP also revealed that our cryo-EM model was lacking a density in the region located at the base of the head domain – a region that had been annotated in the published structure (PDB-5gan) as that of the yeast Snu66 protein (fig: 3.27, C; shown by a blue dotted circle). In a recent study Bertram et al. compared the molecular architecture of the yeast tri-snRNP reported by Wan et al. (2016) and Nguyen et al. (2016) respectively, with the human spliceosomal B complex as determined by cryo-EM (Bertram et al., 2017; PDB-5o9z). The aforementioned study not only showed that the structural organisation of Prp8, Brr2, Prp6 and the U4/U6 proteins in the reported yeast tri-snRNP is analogous to the organisation in the human B complex, but also, that the density, which was tentatively assigned to helical regions of yeast Snu66 by Nguyen et al. (2016), instead most likely comprises the yeast Prp38 complex (Prp38/ Snu23/ Spp381 protein complex), as this position is equivalent to the hPrp38 complex in the human B complex and can even be superimposed (Bertram et al., 2017; PDB-5o9z). Another study (Plaschka et al., 2017) indeed showed that this density comprises the yeast Prp38/ Snu23/ Spp381 protein complex.

MS analyses of our Sad1-TAP tri-snRNP have shown that Prp38, Snu23 and Spp381 are all drastically reduced, while the incorporation of Sad1 and Prp28 was increased. This may be interpreted as a mutually exclusive presence of Prp38 and Sad1 and appears to be reflected in our cryo-EM model as well. We have clearly shown above that the presence of an extra density in the Sad1-TAP tri-snRNP is consistent with the presence of Sad1. It is feasible that the absence of Prp38 and other proteins causes a missing density in our cryo-EM model of the Sad1-TAP tri-snRNP. Such a mutually exclusive presence of Sad1 and Prp38 would clearly bear a resemblance to the human tri-snRNP situation. Even with Sad1 and Prp28 in place, concomitant with the absence of Prp38 (or the Prp38 complex as a whole), the majority of particles harbours a Brr2 already loaded onto the U4 snRNA. The shift of the Brr2 position may therefore not be regulated by only one factor, but rather by a set of factors.

#### **4.2.5 Prp28 in yeast tri-snRNPs shows similar interactions like its human counterpart**

The human U5-100K protein, the ortholog of the yeast DEAD-box protein Prp28, is an integral component of the human U5 snRNP and the U4/U6.U5 tri-snRNP (Teigelkamp et al., 1997).

Prp28 has been shown to be involved in switching the U1 snRNA at the 5' SS against U6 snRNA in the course of spliceosome activation (Chen et al., 2001; Staley and Guthrie, 1999). While the hPrp28 protein is stably associated with the human tri-snRNP (Behrens et al., 1991; Teigelkamp et al., 1997) and its location in the complex is known (Agafonov et al., 2016a), its yeast ortholog could neither be identified by mass spectrometry in previously purified yeast tri-snRNPs (Gottschalk et al., 1999; Stevens and Abelson, 1999; Stevens et al., 2001) nor has it been annotated in any of the published yeast tri-snRNP structures (Nguyen et al., 2015, 2016; Wan et al., 2016). One study, however, has shown that the yeast tri-snRNP can be isolated via TAP-tagged Prp28p (Small et al., 2006) suggesting that at least a tiny sub-population of particles contains Prp28p. In contrast, Prp28, was routinely detected together with Sad1 in our MS investigations of Sad1-TAP tri-snRNPs (table: 3.2, 3.3 and 3.4: Section: 3.1.4.6).

Regardless of detecting numerous peptides of Prp28 in our MS data, our medium-resolution cryo-EM 3D model, regrettably, could not address the structural arrangement of Prp28 in Sad1-TAP tri-snRNP particles. However, back in 2016, my colleagues (Agafonov et al., 2016a) tackled the same problem by utilizing BS3 cross-linking data in order to describe the probable arrangement and location of hPrp28 in the human tri-snRNP. The placement of the Prp28 structure in our 3D cryo-EM map, unlike the Sad1 placement, could not use an extra EM density in our model, and therefore, crosslinking data was instrumental for a correct placement of Prp28 in our cryo-EM map. Cross-linking data for yeast Prp28 (in conjunction with the previous data for hPrp28) allowed us to discuss the most probable arrangement of the protein in the Sad1-TAP tri-snRNP. Our cross-linking data shows that lysine residue 27 in the N-terminal region of Prp28 strongly crosslinks to lysine 1529 located in the CC cassette of Brr2. Similar to the RecA2 domain crosslinks of hPrp28 with the NTD1 and RT/EN domain of hPrp8 (Agafonov et al. 2016), we observed crosslinks (table: S1) in the yPrp28 region aa424 - 580 with the yPrp8-NTD (e.g., lysine559 of Prp28 with aa235 of Prp8). Likewise, crosslinks with the corresponding region of yPrp8 were observed (Lys 1713 yPrp8 with aa429 of Prp28). Furthermore, the N-terminal region of yPrp28 (aa30-80) appears to contact a region of Prp8 (aa1150 to 1299) which corresponds to the thumb/X linker region of hPrp8) and apparently communicates with region aa718-939 of Snu114.

Using a modified purification strategy, we were able to purify a yeast tri-snRNP characterized by a stably associated Sad1 protein, which can be localized in the 3D structure. Sad1 renders the yeast tri-snRNP stable during incubation with ATP, thus resembling the human tri-snRNP. Surprisingly, the 3D structure of this particle revealed that also in this stable yeast tri-snRNP the

Brr2 has already predominantly flipped into the vicinity of U4 in the U4/U6 substrate, a conformation observed in the human system only when forming the B complex (Bertram et al.). But there is also evidence for a subpopulation in which Brr2 appears to be in the orientation observed in the human tri-snRNP.

Given the substantial differences between the human and yeast tri-snRNPs, their composition and mode of integration into the B complex, and taking into consideration that the human system uses a large set of B-specific proteins without orthologs in yeast, it may well be possible that the yeast tri-snRNP enters a pre-B complex indeed with an already flipped Brr2. The true role of Sad1 remains somewhat enigmatic. It is not present in human as well as in yeast B complexes. Sad1 may have a more general stabilizing role during the early assembly steps of the yeast tri-snRNP. A human-like tri-snRNP form with a non-flipped Brr2 may exist just at this early stage. It may be worth to remember the history of yeast Sad1 discovery: it was first described by Lygerou et al. (MCB 1999) as *snRNP assembly defective 1*, an essential, non-snRNP protein, affecting the *assembly of newly synthesized U4/U6 particles rather than the stability/reassembly of the existing U4/U6*. Furthermore, an involvement in the first step of splicing was suspected. When in 2014 Hadjivassilion et al. published the crystal structure of yeast Sad1, the authors speculated about a possible interaction of the protein with pre-spliceosome and tri-snRNP or mere tri-snRNP stability. In any case the presence of Sad1 was very transient. We now know that conditions exist where Sad1 remains firmly attached to the yeast tri-snRNP.

Sad1 and Prp28 are carried into the human pre-B complex as cargo of the human tri-snRNP, where the tri-snRNP and Sad1 are only loosely attached (Boesler et al.). hPrp38 instead enters as a B complex specific protein and as part of the trimeric complex MFAP1(=ySpp381), hPRP38 and hSNU23. Evidence for Sad1 in the yeast pre-B complex remains inconclusive, because the authors placed it in their structure only in analogy to human (Plaschka et al. 2018). In any case, current knowledge would have it, that Prp28, Prp38 and Sad1 being delivered to a pre-B complex as cargo of the yeast tri-snRNP. In human and in yeast (Bertram et al., Plaschka et al. 2017) Prp28 and Prp38 are mutually exclusive proteins using the same binding site. It would therefore create a particularly difficult situation if both proteins are cargo of the same yeast tri-snRNP. In the original papers describing the yeast tri-snRNP for the first time, Prp28 and Sad1 were not described and not even MS peptides were detected (Gottschalk et al. 1999, Fabrizio et al. 1994). Prp28 in fact became a tri-snRNP protein only because it was detected in the U5 snRNP (Stevens et al. 2001). Prp38 instead was always readily detectable. The mutual relations of Sad1, Prp28 and Prp38 are rather different in the human and yeast system. A significant observation in our preparations was the simultaneous strong presence of Sad1 and Prp28 accompanied by a very

low Prp38 count in the isolated tri-snRNP. This was particularly strong under 75 mM K<sup>+</sup> conditions, but also clearly visible under 150 mM K<sup>+</sup>. The sub-complex of Spp381p, Prp38p and Snu23p showed only very low peptide counts under both conditions.

It is during the B to B<sup>act</sup> transition that Prp8 assumes the closed conformation of its cavity, snugly accommodating the U2/U6 RNA network, the U6 ILS, U2/U6 helix Ib and the U2/BS helix (Kastner et al. CSH). In the human tri-snRNP, Dim1 and the RecA2 domain of Prp28 are located between the Prp8's RT/En and NTP domains, stabilizing the open conformation of Prp8 in the tri-snRNP (Agafonov et al.). In yeast, Prp28 binds only with low affinity to the tri-snRNP and is therefore easily dissociated and only a weak competitor of Prp38 when competing for the overlapping binding site in the Prp8 NTD. This may in fact be the main reason that Prp8 in the yeast tri-snRNP assumes the partially closed conformation. This in turn may lead to a rearrangement of the rest of the tri-snRNP structure. While originally assuming that Sad1 may be the only stabilizing factor in the yeast tri-snRNP required to keep the yeast Brr2 in the human tri-snRNP-like Brr2 conformation, it now appears as if Prp28 may be as important, if not more. Further progress will depend on a high resolution cryo-structure of the Sad1-TAP-tri-snRNP and possibly other tags in order to gain an unbiased picture.



## 5 REFERENCES

- Abovich, N., Liao, X.C., and Rosbash, M. (1994). The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes Dev* 8, 843-854.
- Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'- end of U6 snRNA, thereby facilitating U4/U6 duplex formation *in vitro*. *EMBO J* 18, 5789-5802.
- Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.T., Urlaub, H., Lührmann, R., and Stark, H. (2016a). Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* 351, 1416-1420.
- Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.T., Urlaub, H., Lührmann, R., and Stark, H. (2016b). Molecular Architecture of the Human U4/U6.U5 Tri-snRNP. *Science* 351, 1416.
- Anokhina, M., Bessonov, S., Miao, Z., Westhof, E., Hartmuth, K., and Lührmann, R. (2013). RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J*.
- Anthony, J.G., Weidenhammer, E.M., and Woolford, J.L., Jr. (1997). The yeast Prp3 protein is a U4/U6 snRNP protein necessary for integrity of the U4/U6 snRNP and the U4/U6.U5 tri-snRNP. *RNA* 3, 1143-1152.
- Arenas, J., and Hurwitz, J. (1987). Purification of a RNA debranching activity from HeLa cells. *J Biol Chem* 262, 4274-4279.
- Arenas, J.E., and Abelson, J.N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci USA* 94, 11798-11802.
- Behrens, S.E., Galisson, F., Legrain, P., and Lührmann, R. (1993). Evidence that the 60-kDa protein of 17S U2 small nuclear ribonucleoprotein is immunologically and functionally related to the yeast PRP9 splicing factor and is required for the efficient formation of prespliceosomes. *PNAS* 90, 8229-8233.
- Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* 74, 3171-3175.
- Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89, 781-787.
- Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* 170, 701-713 e711.
- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72, 291-336.
- Bleichert, F., and Baserga, S.J. (2007). The long unwinding road of RNA helicases. *Mol Cell* 27, 339-352.
- Blum, H., Beier, H., and Gross, H.J. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8, 93-99.

- Boehringer, D., Makarov, E.M., Sander, B., Makarova, O.V., Kastner, B., Lührmann, R., and Stark, H. (2004). Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat Struct Mol Biol* *11*, 463-468.
- Boesler, C., Rigo, N., Agafonov, D.E., Kastner, B., Urlaub, H., Will, C.L., and Luhrmann, R. (2015). Stable tri-snRNP integration is accompanied by a major structural rearrangement of the spliceosome that is dependent on Prp8 interaction with the 5' splice site. *RNA* *21*, 1993-2005.
- Boesler, C., Rigo, N., Anokhina, M.M., Tauchert, M.J., Agafonov, D.E., Kastner, B., Urlaub, H., Ficner, R., Will, C.L., and Luhrmann, R. (2016). A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nat Commun* *7*, 11997.
- Boon, K.L., Grainger, R.J., Ehsani, P., Barrass, J.D., Auchynnikava, T., Inglehearn, C.F., and Beggs, J.D. (2007). Prp8 Mutations That Cause Human Retinitis Pigmentosa Lead to a U5 snRNP Maturation Defect in Yeast. *Nature structural & molecular biology* *14*, 1077.
- Brow, D.A., and Guthrie, C. (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* *334*, 213-218.
- Burnette, W.N. (1981). 'Western Blotting': Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal Biochem* *112*, 195-203.
- Chan, S.P., Kao, D.I., Tsai, W.Y., and Cheng, S.C. (2003). The Prp19p-associated complex in spliceosome activation. *Science* *302*, 279-282.
- Charenton, C., Wilkinson, M.E., and Nagai, K. (2019). Mechanism of 5' splice site transfer for human spliceosome activation. *Science* *364*, 362-367.
- Chen, C.-H., Yu, W.-C., Tsao, T.Y., Wang, L.-Y., Chen, H.-R., Lin, J.-Y., Tsai, W.-Y., and Cheng, S.-C. (2002). Functional and physical interactions between components of the Prp19p-associated complex. *Nucl Acids Res* *30*, 1029-1037.
- Chen, J.Y., Stands, L., Staley, J.P., Jackups, R.R., Jr., Latus, L.J., and Chang, T.H. (2001). Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol Cell* *7*, 227-232.
- Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An Amazing Sequence Arrangement at the 5' Ends of Adenovirus 2 Messenger RNA. *Cell* *12*, 1.
- Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev* *11*, 2755-2766.
- Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* *349*, 487-493.
- Cordin, O., and Beggs, J.D. (2013). RNA helicases in splicing. *RNA Biol* *10*, 83-95.
- Cordin, O., Hahn, D., and Beggs, J.D. (2012). Structure, function and regulation of spliceosomal RNA helicases. *Curr Opin Cell Biol* *24*, 431-438.
- Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell* *5*, 779-787.
- Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* *26*, 5528-5543.

- Dube, P., Tavares, P., Lurz, R., and van Heel, M. (1993). The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J* *12*, 1303-1309.
- Dziembowski, A., Ventura, A.P., Rutz, B., Caspary, F., Faux, C., Halgand, F., Laprevote, O., and Seraphin, B. (2004). Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J* *23*, 4847-4856.
- Edwards, S.R., and Johnson, T.L. (2019). Intron RNA sequences help yeast cells to survive starvation. *Nature* *565*, 578-579.
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* *36*, 593-608.
- Fan, S.B., Meng, J.M., Lu, S., Zhang, K., Yang, H., Chi, H., Sun, R.X., Dong, M.Q., and He, S.M. (2015). Using pLink to Analyze Cross-Linked Peptides. *Curr Protoc Bioinformatics* *49*, 8 21 21-19.
- Fischer, U., Sumpter, V., Sekine, M., Satoh, T., and Lührmann, R. (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*, 573-583.
- Fleckner, J., Zhang, M., Valcarcel, J., and Green, M.R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev* *11*, 1864-1872.
- Fourmann, J.B., Schmitzova, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.L., Fabrizio, P., and Lührmann, R. (2013). Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes Dev* *27*, 413-428.
- Frank, D.N., Roiha, H., and Guthrie, C. (1994). Architecture of the U5 small nuclear RNA. *Mol Cell Biol* *14*, 2180-2190.
- Galisson, F., and Legrain, P. (1993). The biochemical defects of prp4-1 and prp6-1 yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. *Nucleic Acids Res* *21*, 1555-1562.
- Gilbert, W. (1978). Why genes in pieces? *Nature* *271*, 501.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., *et al.* (1996). Life with 6000 genes. *Science* *274*, 546, 563-547.
- Golas, M.M., Sander, B., Bessonov, S., Grote, M., Wolf, E., Kastner, B., Stark, H., and Lührmann, R. (2010). 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. *Mol Cell* *40*, 927-938.
- Gottschalk, A. (1999). Umfassende Analyse der Protein-Komponenten spleißosomaler Ribonukleoprotein-Komplexe aus *Saccharomyces cerevisiae*. In Institut für Molekularbiologie und Tumorforschung (Marburg: Philipps-Universität).
- Gottschalk, A., Kastner, B., Lührmann, R., and Fabrizio, P. (2001). The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p. *RNA* *7*, 1554-1565.
- Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Lührmann, R., and Fabrizio, P. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J* *18*, 4535-4548.

- Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* *10*, 233-243.
- Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. *RNA* *11*, 533-557.
- Häcker, I., Sander, B., Golas, M.M., Wolf, E., Karagoz, E., Kastner, B., Stark, H., Fabrizio, P., and Lührmann, R. (2008). Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nat Struct Mol Biol* *15*, 1206-1212.
- Hadjivassiliou, H., Rosenberg, O.S., and Guthrie, C. (2014). The crystal structure of *S. cerevisiae* Sad1, a catalytically inactive deubiquitinase that is broadly required for pre-mRNA splicing. *RNA* *20*, 656-669.
- Horowitz, D.S. (2012). The mechanism of the second step of pre-mRNA splicing. *Wiley Interdiscip Rev RNA* *3*, 331-350.
- Huang, Y.H., Chung, C.S., Kao, D.I., Kao, T.C., and Cheng, S.C. (2014). Sad1 counteracts Brr2-mediated dissociation of U4/U6.U5 in tri-snRNP homeostasis. *Mol Cell Biol* *34*, 210-220.
- Jia, X., and Sun, C. (2018). Structural dynamics of the N-terminal domain and the Switch loop of Prp8 during spliceosome assembly and activation. *Nucleic Acids Res* *46*, 3833-3840.
- Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Lührmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* *96*, 375-387.
- Karagöz, E.G. (2006). Localization of functionally important proteins of the yeast *S. cerevisiae* U4/U6.U5 tri-snRNP spliceosomal particle using specific labeling tags and Electron Microscopy. In *Biologische Fakultät (Göttingen: Georg August Universität)*.
- Karijolich, J., and Yu, Y.T. (2010). Spliceosomal snRNA modifications and their function. *RNA Biol* *7*, 192-204.
- Kastner, B., Bach, M., and Lührmann, R. (1990). Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family. *Proc Natl Acad Sci USA* *87*, 1710-1714.
- Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., *et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods* *5*, 53-55.
- Kastner, B., Will, C.L., Stark, H., and Lührmann, R. (2019). Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb Perspect Biol*.
- Kent, O.A., and MacMillan, A.M. (2002). Early organization of pre-mRNA during spliceosome assembly. *Nat Struct Biol* *9*, 576-581.
- Kent, O.A., Ritchie, D.B., and Macmillan, A.M. (2005). Characterization of a U2AF-independent commitment complex (E') in the mammalian spliceosome assembly pathway. *Mol Cell Biol* *25*, 233-240.
- Kim, S.H., and Lin, R.J. (1996). Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol Cell Biol* *16*, 6810-6819.
- Kretzner, L., Krol, A., and Rosbash, M. (1990). *Saccharomyces cerevisiae* U1 small nuclear RNA secondary structure contains both universal and yeast-specific domains. *Proc Natl Acad Sci USA* *87*, 851-855.

- Krol, A., Branlant, C., Lazar, E., Gallinaro, H., and Jacob, M. (1981). Primary and secondary structures of chicken, rat and man nuclear U4 RNAs. Homologies with U1 and U5 RNAs. *Nucleic Acids Res* *9*, 2699-2716.
- Kuhlbrandt, W. (2014). Cryo-EM enters a new era. *eLife* *3*, e03678.
- Laemmli, U.K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* *227*, 680-685.
- Laggerbauer, B., Achsel, T., and Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proc Natl Acad Sci USA* *95*, 4188-4192.
- Legrain, P., Séraphin, B., and Rosbash, M. (1988). Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol Cell Biol* *8*, 3755-3760.
- Liu, S., Rauhut, R., Vornlocher, H.P., and Lührmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA* *12*, 1418-1430.
- Lopez, P.J., and Séraphin, B. (1999). Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition [letter]. *RNA* *5*, 1135-1137.
- Ma, P., and Xia, X. (2011). Factors affecting splicing strength of yeast genes. *Comp Funct Genomics* *2011*, 212146.
- Madhani, H.D., and Guthrie, C. (1992). A Novel Base-Pairing Interaction Between U2 and U6 snRNAs Suggests a Mechanism for the Catalytic Activation of the Spliceosome. *Cell* *71*, 803.
- Maeder, C., Kutach, A.K., and Guthrie, C. (2009). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat Struct Mol Biol* *16*, 42-48.
- Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P., and Lührmann, R. (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6\*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J* *21*, 1148-1157.
- Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J* *23*, 2381-2391.
- Mattaj, I.W., and De Robertis, E.M. (1985). Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell* *40*, 111-118.
- McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Program, A.E.S.T., Shuman, S., *et al.* (1997). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* *11*, 3306-3318.
- Merril, C.R., Goldman, D., Sedman, S.A., and Ebert, M.H. (1981). Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* *211*, 1437-1438.
- Moore, M.J., Query, C.C., and Sharp, P.A. (1993). Splicing of precursors to mRNA by the spliceosome. In *RNA World*, A. Gesteland, ed. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 303-357.
- Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* *365*, 364-368.
- Morgan, J.T., Fink, G.R., and Bartel, D.P. (2019). Excised linear introns regulate growth in yeast. *Nature* *565*, 606-611.

- Mozaffari-Jovin, S., Wandersleben, T., Santos, K.F., Will, C.L., Lührmann, R., and Wahl, M.C. (2013). Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* *341*, 80-84.
- Neueglise, C., Marck, C., and Gaillardin, C. (2011). The intronome of budding yeasts. *C R Biol* *334*, 662-670.
- Nguyen, T.H., Galej, W.P., Bai, X.C., Oubridge, C., Newman, A.J., Scheres, S.H., and Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature*.
- Nguyen, T.H., Galej, W.P., Bai, X.C., Savva, C.G., Newman, A.J., Scheres, S.H., and Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* *523*, 47-52.
- Nilsen, T.W. (1998). RNA-RNA interactions in nuclear pre-mRNA splicing. In *RNA Structure and Function*, M. Grundber-Manago, and R.W. Simons, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 279-307.
- O'Keefe, R.T., and Newman, A.J. (1998). Functional Analysis of the U5 snRNA Loop 1 in the Second Catalytic Step of Yeast Pre-mRNA Splicing. *EMBO J* *17*, 565.
- Ohi, M.D., Link, A.J., Ren, L., Jennings, J.L., McDonald, W.H., and Gould, K.L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol Cell Biol* *22*, 2011-2024.
- Padgett, R.A. (2005). mRNA Splicing: Role of snRNAs. In *Encyclopedia of Life Sciences* (Chichester: John Wiley & Sons, Ltd).
- Parenteau, J., Maignon, L., Berthoumieux, M., Catala, M., Gagnon, V., and Abou Elela, S. (2019). Introns are mediators of cell response to starvation. *Nature* *565*, 612-617.
- Patterson, B., and Guthrie, C. (1991). A U-rich tract enhances usage of an alternative 3' splice site in yeast. *Cell* *64*, 181-187.
- Pena, V., Jovin, S.M., Fabrizio, P., Orłowski, J., Bujnicki, J.M., Lührmann, R., and Wahl, M.C. (2009). Common design principles in the spliceosomal RNA helicase Brr2 and in the Hel308 DNA helicase. *Mol Cell* *35*, 454-466.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* *25*, 1605-1612.
- Plaschka, C., Lin, P.C., and Nagai, K. (2017). Structure of a pre-catalytic spliceosome. *Nature* *546*, 617-621.
- Plaschka, C., Newman, A.J., and Nagai, K. (2019). Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. *Cold Spring Harb Perspect Biol* *11*.
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* *24*, 218-229.
- Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev* *8*, 587-597.
- Ragunathan, P.L., and Guthrie, C. (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol* *8*, 847-855.

- Rigo, N., Sun, C., Fabrizio, P., Kastner, B., and Lührmann, R. (2015). Protein localisation by electron microscopy reveals the architecture of the yeast spliceosomal B complex. *EMBO J* *34*, 3059-3073.
- Sakharkar, M.K., Chow, V.T., and Kanguane, P. (2004). Distributions of exons and introns in the human genome. *In silico biology* *4*, 387-393.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*, 2nd edn (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
- Sambrook, J., Russell, D.W., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, Second edn (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
- Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T., and Lührmann, R. (2004). RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO J* *23*, 3000-3009.
- Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* *180*, 519-530.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods* *9*, 671-675.
- Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell* *30*, 743-754.
- Schwer, B., and Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J* *11*, 5033-5039.
- Segault, V., Will, C.L., Polycarpou-Schwarz, M., Mattaj, I.W., Branlant, C., and Lührmann, R. (1999). Conserved loop I of U5 small nuclear RNA is dispensable for both catalytic steps of pre-mRNA splicing in HeLa nuclear extracts. *Mol Cell Biol* *19*, 2782-2790.
- Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. *Cell* *9*, 645-653.
- Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* *68*, 850-858.
- Shi, Y., Su, Z., Yang, H., Wang, W., Jin, G., He, G., Siddique, A.N., Zhang, L., Zhu, A., Xue, R., *et al.* (2019). Alternative splicing coupled to nonsense-mediated mRNA decay contributes to the high-altitude adaptation of maca (*Lepidium meyenii*). *Gene* *694*, 7-18.
- Shukla, G.C., and Padgett, R.A. (2002). A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol Cell* *9*, 1145-1150.
- Shuster, E.O., and Guthrie, C. (1988). Two conserved domains of yeast U2 snRNA are separated by 945 nonessential nucleotides. *Cell* *55*, 41-48.
- Small, E.C., Leggett, S.R., Winans, A.A., and Staley, J.P. (2006). The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol Cell* *23*, 389-399.
- Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985). Measurement of protein using bicinchoninic acid. *Anal Biochem* *150*, 76-85.
- Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* *5*, 221-234.

- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* *92*, 315-326.
- Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* *3*, 55-64.
- Stevens, S.W., and Abelson, J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc Natl Acad Sci USA* *96*, 7226-7231.
- Stevens, S.W., Barta, I., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2001). Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA* *7*, 1543-1553.
- Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and Functional Characterization of the Yeast Spliceosomal Penta-snRNP. *Mol Cell* *9*, 31-44.
- Sun, C., Rigo, N., Fabrizio, P., Kastner, B., and Luhrmann, R. (2016). A protein map of the yeast activated spliceosome as obtained by electron microscopy. *RNA* *22*, 1427-1440.
- Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* *157*, 38-46.
- Thuman-Commike, P.A. (2001). Single particle macromolecular structure determination via electron microscopy. *FEBS Lett* *505*, 199-205.
- Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O., and Gilbert, W. (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*, 1485-1489.
- Tsai, R.T., Fu, R.H., Yeh, F.L., Tseng, C.K., Lin, Y.C., Huang, Y.H., and Cheng, S.C. (2005). Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev* *19*, 2991-3003.
- Tseng, C.K., Liu, H.L., and Cheng, S.C. (2011). DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* *17*, 145-154.
- Ulrich, A.K., and Wahl, M.C. (2017). Human MFAP1 is a cryptic ortholog of the *Saccharomyces cerevisiae* Spp381 splicing factor. *BMC Evol Biol* *17*, 91.
- van Heel, M., and Frank, J. (1981a). Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy* *6*, 187-194.
- van Heel, M., and Frank, J. (1981b). Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy* *6*, 187-194.
- van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J Struct Biol* *116*, 17-24.
- Vidal, V.P., Verdone, L., Mayes, A.E., and Beggs, J.D. (1999). Characterization of U6 snRNA-protein interactions. *RNA* *5*, 1470-1481.
- Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* *136*, 701-718.
- Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C.C., and Shi, Y. (2016). The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* *351*, 466-475.



- Warkocki, Z., Odenwalder, P., Schmitzova, J., Platzmann, F., Stark, H., Urlaub, H., Ficner, R., Fabrizio, P., and Luhrmann, R. (2009). Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat Struct Mol Biol* *16*, 1237-1243.
- Wilkinson, M.E., Lin, P.C., Plaschka, C., and Nagai, K. (2018). Cryo-EM Studies of Pre-mRNA Splicing: From Sample Preparation to Model Visualization. *Annu Rev Biophys* *47*, 175-199.
- Will, C.L., and Luhrmann, R. (2011a). Spliceosome structure and function. In *RNA Worlds*, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 181-203.
- Will, C.L., and Luhrmann, R. (2011b). Spliceosome structure and function. *Cold Spring Harb Perspect Biol* *3*, 1-23.
- Yan, C., Hang, J., Wan, R., Huang, M., Wong, C.C., and Shi, Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* *349*, 1182-1191.
- Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C.C., Pettersen, E.F., Goddard, T.D., Meng, E.C., Sali, A., and Ferrin, T.E. (2012). UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *J Struct Biol* *179*, 269-278.
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J Struct Biol* *193*, 1-12.
- Zhang, M., and Green, M.R. (2001). Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes Dev* *15*, 30-35.
- Zhang, M.Q. (1998). Statistical features of human exons and their flanking regions. *Hum Mol Genet* *7*, 919-932.



# APPENDIX

## Abbreviations

2D	two-dimensional
3D	three-dimensional
3' SS	3' splice site
5' SS	5' splice site
°C	degree Celsius
A	adenosine
Å	Ångström
AA	acrylamide
ADP	adenosine-5'-diphosphate
APS	ammonium peroxy sulfate
ATG	Start codon
ATP	adenosine-5'-triphosphate
bp	base pair
<i>C. albicans</i>	<i>Candida albicans</i>
CBP	calmodulin binding peptide
Ci	Curie
Da	Dalton
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide-5'-triphosphates
DTT	1,4-dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediamine-N, N, N', N'-tetraacetic acid
EM	electron microscopy
fmole	femtomole
G	guanosine
GDP	guanosine-5'-diphosphate
GFP	green fluorescent protein
GTP	guanosine-5'-triphosphate

hr.	hour
HIS	Histidine
IgG	immunoglobulin G
kb	kilo base
kDa	kilo dalton
kV	kilo volt
L	liter
LB	Luria Bertani
Lsm	Like-Sm
$\mu$	micro
$\mu$ l	microliter
$\mu$ m	micrometer
$\mu$ M	micromolar
$\mu$ mol	micromol
M	molar
m <sub>3</sub> G	2,2,7-trimethylguanosine
min	minute
ml	milliliter
mM	millimolar
mRNA	messenger RNA
MS	mass spectrometry
MSA	multivariate statistical analysis
n	nano
nm	nanometer
nM	nanomolar
NMR	nuclear magnetic resonance spectroscopy
Nmol	nanomol
NP-40	Nonidet P-40
ns	nucleotides
NTPs	nucleotide-5'-triphosphates
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCI	phenol-chloroform-isoamyl alcohol

PCR	polymerase chain reaction
PCV	packed cell volume
PEG	polyethylene glycol
pmole	picomole
poly (A)	poly-adenine
poly (U)	poly-uridine
pre-mRNA	pre-messenger RNA
ProtA	<i>Staphylococcus aureus</i> protein A
RNA	ribonucleic acid
RNP	ribonucleoprotein
RRM	RNA recognition motif
RT	room temperature
RT-mix	reverse transcriptase mix
S	Svedberg
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	sodium dodecyl sulfate
Sm	Smith (Stefanie); systemic lupus erythematosus patient in which Sm proteins were first discovered as antigens
snRNA	small nuclear ribonucleic acid
snoRNP	small nucleolar ribonucleoprotein
SNR	signal-to-noise ratio
<i>S. pombe</i>	<i>Schizosaccharomyces pombe</i>
T	thymidine
TAP	tandem affinity purification
TBE	Tris-borate-EDTA buffer
TBS	Tris-buffered saline
TE	Tris-EDTA buffer
TEMED	N, N, N', N'-tetramethylethylenediamine
TEV protease	<i>tobacco etch virus</i> protease
TPR	tetratricopeptide repeat
TRP	tryptophane
U	uridine
URA	uracil
UsnRNA	uridine-rich small nuclear ribonucleic acid

UsnRNP	uridine-rich small nuclear ribonucleoprotein
UV	ultra violet
YPD	yeast extract, peptone, dextrose

Supplimentary data to section 3 Cross-linking data for Sad1-TAP trisnRNP  
For codes please refer to section 3

Supplimentary Data Table:S1

Protein 1	XL_type	Protein 2	Residue1	Residue2	CSMs	Values Count of Scan
<b>BRR2</b>	<b>inter- molecular</b>	DIB1	<b>597</b>	34	<b>3</b>	3
		LIN1	<b>1589</b>	219	<b>4</b>	4
		LSM7	<b>276</b>	32	<b>1</b>	1
		LSM8	<b>1404</b>	82	<b>2</b>	2
			<b>1634</b>	52	<b>2</b>	2
		PRP28	<b>2</b>	61	<b>2</b>	2
				163	<b>3</b>	3
				429	<b>1</b>	1
				458	<b>3</b>	3
				471	<b>12</b>	12
				555	<b>1</b>	1
				559	<b>4</b>	4
				573	<b>2</b>	2
			<b>7</b>	559	<b>1</b>	1
			<b>28</b>	471	<b>5</b>	5
				555	<b>2</b>	2
				573	<b>1</b>	1
				575	<b>1</b>	1
			<b>688</b>	279	<b>1</b>	1
			<b>1315</b>	27	<b>2</b>	2
			<b>1529</b>	27	<b>16</b>	16
		PRP3	<b>445</b>	307	<b>2</b>	2
			<b>457</b>	261	<b>6</b>	6
			<b>717</b>	261	<b>8</b>	8
			<b>720</b>	261	<b>1</b>	1
			<b>1529</b>	355	<b>1</b>	1
			<b>1676</b>	242	<b>3</b>	3
		PRP4	<b>439</b>	3	<b>4</b>	4
			<b>445</b>	3	<b>69</b>	69
			<b>454</b>	3	<b>6</b>	6
			<b>457</b>	17	<b>64</b>	64
			<b>470</b>	3	<b>20</b>	20
			<b>769</b>	17	<b>32</b>	32
		PRP6	<b>2</b>	39	<b>1</b>	1
				45	<b>1</b>	1
				231	<b>5</b>	10

				262	6	6
			59	261	4	4
			71	231	18	18
				256	26	26
						Values
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
				261	1	1
				262	34	34
			74	231	55	55
				235	4	4
				256	18	18
				261	1	1
				262	44	44
			82	231	41	41
				235	61	61
				256	12	12
				262	58	58
			85	231	26	26
					30	30
				235	28	28
				256	12	12
				262	30	30
			90	231	5	5
				235	11	11
				256	5	5
				261	2	2
				262	1	1
					16	16
			91	231	23	23
				235	45	45
				256	24	24
				261	2	2
				262	76	76
				293	3	3
				329	2	2
			166	231	1	1
			168	231	1	1
				262	2	2
			414	183	4	4
			417	183	1	1
				193	1	1
			445	351	1	1
					2	2



				356	<b>3</b>	3
					<b>8</b>	8
				360	<b>11</b>	11
			<b>454</b>	356	<b>2</b>	2
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
					<b>15</b>	15
				360	<b>20</b>	20
			<b>778</b>	31	<b>2</b>	2
				39	<b>14</b>	14
			<b>795</b>	39	<b>14</b>	14
			<b>1653</b>	183	<b>1</b>	1
			<b>1676</b>	570	<b>2</b>	2
			<b>2095</b>	39	<b>6</b>	6
		PRP8	<b>2</b>	152	<b>3</b>	3
				235	<b>5</b>	5
					<b>20</b>	20
				351	<b>3</b>	3
				517	<b>1</b>	1
				586	<b>1</b>	1
				681	<b>1</b>	1
				1416	<b>8</b>	8
				1693	<b>1</b>	1
				1713	<b>18</b>	18
				1725	<b>1</b>	1
				2122	<b>16</b>	16
				2124	<b>1</b>	1
			<b>7</b>	235	<b>6</b>	6
				1416	<b>1</b>	1
				1713	<b>6</b>	6
				2122	<b>10</b>	10
				2124	<b>1</b>	1
			<b>11</b>	2167	<b>5</b>	5
			<b>25</b>	1416	<b>1</b>	1
				1693	<b>1</b>	1
				1713	<b>3</b>	3
			<b>28</b>	235	<b>4</b>	4
				600	<b>21</b>	21
				681	<b>3</b>	3
				1416	<b>14</b>	14
				1435	<b>1</b>	1
				1713	<b>13</b>	13
			<b>59</b>	1299	<b>23</b>	23
			<b>71</b>	1435	<b>6</b>	6

					<b>8</b>	8
			<b>74</b>	1299	<b>3</b>	3
				1416	<b>17</b>	17
				1435	<b>8</b>	8
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
			<b>82</b>	131	<b>1</b>	1
				1416	<b>19</b>	19
				1435	<b>1</b>	1
				1903	<b>2</b>	2
			<b>85</b>	1416	<b>29</b>	29
				1435	<b>1</b>	1
				1903	<b>5</b>	5
			<b>90</b>	1416	<b>11</b>	11
			<b>91</b>	1416	<b>19</b>	19
				1435	<b>1</b>	1
				1903	<b>12</b>	12
			<b>187</b>	2097	<b>1</b>	1
			<b>190</b>	2094	<b>1</b>	1
			<b>257</b>	1148	<b>2</b>	2
				1713	<b>1</b>	1
				2167	<b>10</b>	10
				2187	<b>2</b>	2
				2191	<b>1</b>	1
					<b>2</b>	2
			<b>259</b>	1148	<b>1</b>	1
			<b>304</b>	1864	<b>2</b>	2
				2167	<b>3</b>	3
				2187	<b>5</b>	5
			<b>398</b>	1864	<b>1</b>	1
			<b>414</b>	1864	<b>9</b>	9
			<b>748</b>	2108	<b>6</b>	6
				2122	<b>1</b>	1
			<b>795</b>	2108	<b>50</b>	50
				2122	<b>14</b>	14
					<b>19</b>	19
				2124	<b>5</b>	5
			<b>1051</b>	2108	<b>1</b>	1
			<b>1055</b>	2108	<b>19</b>	19
				2149	<b>1</b>	1
				2154	<b>18</b>	18
			<b>1060</b>	2122	<b>4</b>	4
				2124	<b>3</b>	3
				2149	<b>2</b>	2
			<b>1158</b>	2124	<b>8</b>	8
			<b>1441</b>	98	<b>1</b>	1

		RSMB	<b>2</b>	138	<b>1</b>	1
			<b>28</b>	145	<b>3</b>	3
			<b>257</b>	135	<b>1</b>	1
			<b>470</b>	186	<b>1</b>	1
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
			<b>720</b>	145	<b>7</b>	7
			<b>758</b>	132	<b>2</b>	2
				138	<b>2</b>	2
				145	<b>5</b>	5
				186	<b>1</b>	1
			<b>769</b>	132	<b>1</b>	1
				135	<b>3</b>	3
				138	<b>3</b>	3
					<b>6</b>	6
				145	<b>23</b>	23
			<b>778</b>	117	<b>1</b>	1
				131	<b>1</b>	1
				132	<b>3</b>	3
				135	<b>3</b>	3
				138	<b>1</b>	1
					<b>7</b>	7
				145	<b>13</b>	13
				186	<b>5</b>	5
			<b>782</b>	100	<b>3</b>	3
				105	<b>43</b>	43
				117	<b>8</b>	8
				121	<b>1</b>	1
				138	<b>1</b>	1
				145	<b>2</b>	2
					<b>3</b>	3
			<b>795</b>	100	<b>1</b>	1
				105	<b>6</b>	6
				113	<b>3</b>	3
				114	<b>27</b>	27
				116	<b>4</b>	4
				117	<b>31</b>	31
				121	<b>50</b>	50
				124	<b>9</b>	9
				127	<b>3</b>	3
				129	<b>6</b>	6
				131	<b>35</b>	35
				132	<b>17</b>	17
				135	<b>1</b>	1
					<b>12</b>	12
				138	<b>3</b>	3

					<b>19</b>	19
				145	<b>44</b>	44
				186	<b>5</b>	5
			<b>1138</b>	3	<b>23</b>	23
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
				105	<b>2</b>	2
				117	<b>5</b>	5
			<b>1150</b>	105	<b>5</b>	5
			<b>1158</b>	117	<b>4</b>	4
				121	<b>1</b>	1
			<b>1187</b>	117	<b>10</b>	10
				121	<b>1</b>	1
				124	<b>1</b>	1
			<b>1634</b>	132	<b>1</b>	1
		RUXE	<b>351</b>	1	<b>7</b>	7
			<b>1529</b>	40	<b>2</b>	2
		SAD1TAP	<b>376</b>	6	<b>1</b>	1
			<b>1529</b>	207	<b>1</b>	1
			<b>1589</b>	564	<b>1</b>	1
				568	<b>1</b>	1
				614	<b>1</b>	1
			<b>1623</b>	564	<b>2</b>	2
				568	<b>2</b>	2
		SMD1	<b>1896</b>	111	<b>16</b>	16
			<b>1904</b>	111	<b>7</b>	7
		SMD2	<b>546</b>	73	<b>75</b>	75
			<b>549</b>	73	<b>42</b>	42
				93	<b>6</b>	6
			<b>1060</b>	27	<b>1</b>	1
			<b>1904</b>	27	<b>9</b>	9
				59	<b>36</b>	36
			<b>2109</b>	10	<b>6</b>	6
		SMD3	<b>1150</b>	99	<b>1</b>	1
			<b>1187</b>	85	<b>8</b>	8
		SN114	<b>2</b>	955	<b>62</b>	62
			<b>7</b>	955	<b>28</b>	28
			<b>9</b>	955	<b>4</b>	4
					<b>8</b>	8
			<b>25</b>	955	<b>28</b>	28
			<b>28</b>	955	<b>23</b>	23
			<b>82</b>	955	<b>4</b>	4
			<b>259</b>	99	<b>1</b>	1
			<b>956</b>	520	<b>1</b>	1
			<b>1623</b>	947	<b>1</b>	1
		SNU66	<b>74</b>	87	<b>4</b>	4

			<b>82</b>	35	<b>1</b>	1
				87	<b>13</b>	13
			<b>85</b>	87	<b>7</b>	7
			<b>90</b>	87	<b>1</b>	1
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
			<b>91</b>	3	<b>1</b>	1
				18	<b>2</b>	2
				22	<b>1</b>	1
				30	<b>3</b>	3
				35	<b>14</b>	14
				53	<b>17</b>	17
				73	<b>4</b>	4
				83	<b>1</b>	1
				87	<b>18</b>	18
				131	<b>4</b>	4
					<b>5</b>	5
				135	<b>3</b>	3
			<b>166</b>	73	<b>10</b>	10
				83	<b>1</b>	1
				87	<b>21</b>	21
			<b>168</b>	73	<b>13</b>	13
				80	<b>6</b>	6
				83	<b>22</b>	22
				87	<b>93</b>	93
			<b>181</b>	73	<b>2</b>	2
			<b>204</b>	87	<b>12</b>	12
			<b>304</b>	123	<b>5</b>	5
				135	<b>2</b>	2
				157	<b>12</b>	12
				158	<b>3</b>	3
				276	<b>1</b>	1
				279	<b>4</b>	4
				284	<b>3</b>	3
				286	<b>1</b>	1
				292	<b>1</b>	1
			<b>380</b>	157	<b>4</b>	4
			<b>398</b>	157	<b>1</b>	1
				158	<b>2</b>	2
			<b>414</b>	157	<b>20</b>	20
				158	<b>6</b>	6
				306	<b>2</b>	2
			<b>417</b>	157	<b>48</b>	48
				158	<b>1</b>	1
			<b>457</b>	397	<b>1</b>	1
			<b>470</b>	397	<b>7</b>	7

				486	<b>2</b>	2
			<b>478</b>	337	<b>2</b>	2
				395	<b>1</b>	1
				397	<b>23</b>	23
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
				404	<b>8</b>	8
				486	<b>5</b>	5
				498	<b>4</b>	4
			<b>537</b>	395	<b>2</b>	2
				397	<b>9</b>	9
				468	<b>1</b>	1
				471	<b>5</b>	5
				486	<b>2</b>	2
			<b>546</b>	375	<b>45</b>	45
				395	<b>19</b>	19
				397	<b>2</b>	2
					<b>50</b>	50
				404	<b>51</b>	51
				415	<b>13</b>	13
				456	<b>15</b>	15
				457	<b>227</b>	227
				468	<b>79</b>	79
				471	<b>38</b>	38
				486	<b>26</b>	26
				496	<b>6</b>	6
				498	<b>22</b>	22
				518	<b>25</b>	25
				528	<b>7</b>	7
				535	<b>5</b>	5
				542	<b>6</b>	6
				547	<b>2</b>	2
				551	<b>10</b>	10
			<b>548</b>	457	<b>4</b>	4
			<b>549</b>	375	<b>18</b>	18
				397	<b>6</b>	6
				456	<b>7</b>	7
				457	<b>42</b>	42
				486	<b>1</b>	1
			<b>717</b>	397	<b>2</b>	2
			<b>795</b>	286	<b>5</b>	5
			<b>967</b>	486	<b>3</b>	3
				518	<b>7</b>	7
				535	<b>4</b>	4
				542	<b>8</b>	8
				547	<b>8</b>	8

				551	<b>5</b>	5
				556	<b>7</b>	7
				564	<b>2</b>	2
				573	<b>5</b>	5
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
			<b>1055</b>	208	<b>6</b>	6
			<b>1372</b>	535	<b>32</b>	32
			<b>1392</b>	297	<b>1</b>	1
				301	<b>1</b>	1
			<b>1414</b>	551	<b>3</b>	3
			<b>1437</b>	518	<b>2</b>	2
			<b>1441</b>	251	<b>1</b>	1
				256	<b>2</b>	2
				306	<b>12</b>	12
				486	<b>4</b>	4
				498	<b>4</b>	4
				518	<b>20</b>	20
				528	<b>1</b>	1
				535	<b>1</b>	1
				542	<b>24</b>	24
				547	<b>7</b>	7
				551	<b>17</b>	17
				556	<b>7</b>	7
				564	<b>6</b>	6
				573	<b>8</b>	8
			<b>1529</b>	3	<b>1</b>	1
				276	<b>14</b>	14
					<b>75</b>	75
				279	<b>30</b>	30
				281	<b>2</b>	2
				282	<b>2</b>	2
				284	<b>53</b>	53
				286	<b>17</b>	17
				292	<b>12</b>	12
				457	<b>1</b>	1
			<b>1589</b>	375	<b>9</b>	9
				395	<b>1</b>	1
			<b>1623</b>	251	<b>3</b>	3
				256	<b>9</b>	9
				258	<b>1</b>	1
				395	<b>1</b>	1
				397	<b>6</b>	6
				486	<b>29</b>	29
				496	<b>2</b>	2
				498	<b>19</b>	19

				518	<b>8</b>	8
				535	<b>3</b>	3
				542	<b>13</b>	13
				546	<b>1</b>	1
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
				547	<b>6</b>	6
				551	<b>62</b>	62
				556	<b>2</b>	2
				564	<b>3</b>	3
				573	<b>1</b>	1
			<b>1634</b>	375	<b>34</b>	34
				395	<b>35</b>	35
				397	<b>13</b>	13
					<b>94</b>	94
				404	<b>38</b>	38
				468	<b>77</b>	77
				471	<b>80</b>	80
				486	<b>131</b>	131
				496	<b>48</b>	48
				498	<b>56</b>	56
				518	<b>52</b>	52
				528	<b>18</b>	18
				535	<b>11</b>	11
				542	<b>33</b>	33
				547	<b>7</b>	7
				551	<b>20</b>	20
				556	<b>11</b>	11
				559	<b>1</b>	1
				564	<b>4</b>	4
				573	<b>2</b>	2
			<b>1676</b>	535	<b>5</b>	5
				542	<b>2</b>	2
				564	<b>1</b>	1
			<b>1904</b>	375	<b>37</b>	37
				395	<b>5</b>	5
				397	<b>10</b>	10
				471	<b>1</b>	1
				486	<b>18</b>	18
				496	<b>3</b>	3
				498	<b>10</b>	10
				518	<b>17</b>	17
			<b>2003</b>	375	<b>12</b>	12
				397	<b>12</b>	12
				468	<b>6</b>	6
				471	<b>6</b>	6



Protein 1	XL_type	Protein 2	Residue1	Residue2	CSMs	Count of Scan
				486	<b>29</b>	29
				498	<b>4</b>	4
				518	<b>9</b>	9
				542	<b>2</b>	2
						<b>Values</b>
			<b>2007</b>	486	<b>4</b>	4
				518	<b>5</b>	5
				542	<b>2</b>	2
				547	<b>7</b>	7
				551	<b>4</b>	4
			<b>2109</b>	297	<b>1</b>	1
				337	<b>30</b>	30
				348	<b>1</b>	1
				354	<b>4</b>	4
					<b>11</b>	11
			<b>2115</b>	297	<b>4</b>	4
		SP381	<b>2108</b>	264	<b>1</b>	1
	<b>intra-molecular</b>	BRR2	<b>2</b>	25	<b>13</b>	13
				28	<b>19</b>	19
					<b>20</b>	20
				31	<b>16</b>	16
				50	<b>6</b>	6
				59	<b>3</b>	3
				74	<b>6</b>	6
				82	<b>2</b>	2
				91	<b>2</b>	2
				795	<b>15</b>	15
			<b>7</b>	11	<b>8</b>	8
				12	<b>1</b>	1
				25	<b>2</b>	2
				28	<b>13</b>	13
				31	<b>4</b>	4
				74	<b>1</b>	1
				417	<b>1</b>	1
				795	<b>4</b>	4
			<b>9</b>	28	<b>2</b>	2
					<b>11</b>	11
				417	<b>1</b>	1
			<b>11</b>	7	<b>8</b>	8
				772	<b>1</b>	1
			<b>12</b>	7	<b>1</b>	1
			<b>25</b>	2	<b>13</b>	13
				7	<b>2</b>	2
				31	<b>10</b>	10

			<b>28</b>	2	<b>19</b>	19
					<b>20</b>	20
				7	<b>13</b>	13
				9	<b>2</b>	2
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
					<b>11</b>	11
				50	<b>2</b>	2
				74	<b>4</b>	4
				82	<b>2</b>	2
				91	<b>6</b>	6
			<b>31</b>	2	<b>16</b>	16
				7	<b>4</b>	4
				25	<b>10</b>	10
				50	<b>1</b>	1
				74	<b>1</b>	1
				90	<b>1</b>	1
				1414	<b>1</b>	1
			<b>50</b>	2	<b>6</b>	6
				28	<b>2</b>	2
				31	<b>1</b>	1
				59	<b>2</b>	2
				71	<b>2</b>	2
				74	<b>4</b>	4
				91	<b>7</b>	7
			<b>59</b>	2	<b>3</b>	3
				50	<b>2</b>	2
				74	<b>13</b>	13
				82	<b>6</b>	6
				85	<b>1</b>	1
				91	<b>12</b>	12
				982	<b>2</b>	2
			<b>71</b>	50	<b>2</b>	2
				82	<b>5</b>	5
				85	<b>14</b>	14
				90	<b>7</b>	7
				91	<b>19</b>	19
					<b>20</b>	20
			<b>74</b>	2	<b>6</b>	6
				7	<b>1</b>	1
				28	<b>4</b>	4
				31	<b>1</b>	1
				50	<b>4</b>	4
				59	<b>13</b>	13
				85	<b>36</b>	36
				90	<b>11</b>	11

				91	<b>1</b>	1
					<b>35</b>	35
				1529	<b>5</b>	5
			<b>82</b>	2	<b>2</b>	2
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
				28	<b>2</b>	2
				59	<b>6</b>	6
				71	<b>5</b>	5
				90	<b>32</b>	32
				91	<b>89</b>	89
				1529	<b>1</b>	1
			<b>85</b>	59	<b>1</b>	1
				71	<b>14</b>	14
				74	<b>36</b>	36
				91	<b>73</b>	73
			<b>90</b>	31	<b>1</b>	1
				71	<b>7</b>	7
				74	<b>11</b>	11
				82	<b>32</b>	32
				1589	<b>1</b>	1
			<b>91</b>	2	<b>2</b>	2
				28	<b>6</b>	6
				50	<b>7</b>	7
				59	<b>12</b>	12
				71	<b>19</b>	19
					<b>20</b>	20
				74	<b>1</b>	1
					<b>35</b>	35
				82	<b>89</b>	89
				85	<b>73</b>	73
				166	<b>2</b>	2
				168	<b>6</b>	6
				204	<b>2</b>	2
				1051	<b>2</b>	2
			<b>152</b>	190	<b>85</b>	85
			<b>166</b>	91	<b>2</b>	2
				717	<b>1</b>	1
				720	<b>1</b>	1
			<b>168</b>	91	<b>6</b>	6
				190	<b>1</b>	1
				717	<b>12</b>	12
				720	<b>1</b>	1
			<b>187</b>	1676	<b>1</b>	1
			<b>190</b>	152	<b>85</b>	85
				168	<b>1</b>	1

				304	1	1
				1676	1	1
			204	91	2	2
			215	659	1	1
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
			255	257	11	11
				259	6	6
				398	1	1
			256	304	1	1
			257	255	11	11
				304	11	11
				1187	1	1
			259	255	6	6
			271	304	3	3
				351	12	12
				398	3	3
				414	1	1
			276	304	6	6
				351	15	15
				380	11	11
				398	5	5
				414	1	1
				417	8	8
				1733	3	3
			304	190	1	1
				256	1	1
				257	11	11
				271	3	3
				276	6	6
				339	4	4
					54	54
				414	7	7
				417	7	7
				967	3	3
				1315	1	1
				1316	3	3
				1529	11	11
			339	304	4	4
					54	54
			351	271	12	12
				276	15	15
				398	10	10
			364	417	1	1
			380	276	11	11
			390	398	4	4

				414	<b>3</b>	3
				417	<b>14</b>	14
				1529	<b>1</b>	1
			<b>398</b>	255	<b>1</b>	1
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
				271	<b>3</b>	3
				276	<b>5</b>	5
				351	<b>10</b>	10
				390	<b>4</b>	4
				414	<b>7</b>	7
				417	<b>12</b>	12
				457	<b>1</b>	1
				967	<b>6</b>	6
				1138	<b>1</b>	1
				1733	<b>2</b>	2
			<b>414</b>	271	<b>1</b>	1
				276	<b>1</b>	1
				304	<b>7</b>	7
				390	<b>3</b>	3
				398	<b>7</b>	7
				417	<b>19</b>	19
					<b>46</b>	46
				967	<b>33</b>	33
			<b>417</b>	7	<b>1</b>	1
				9	<b>1</b>	1
				276	<b>8</b>	8
				304	<b>7</b>	7
				364	<b>1</b>	1
				390	<b>14</b>	14
				398	<b>12</b>	12
				414	<b>19</b>	19
					<b>46</b>	46
				967	<b>68</b>	68
			<b>431</b>	956	<b>6</b>	6
					<b>13</b>	13
			<b>439</b>	956	<b>4</b>	4
			<b>457</b>	398	<b>1</b>	1
				760	<b>4</b>	4
			<b>470</b>	505	<b>42</b>	42
				527	<b>3</b>	3
				772	<b>3</b>	3
			<b>505</b>	470	<b>42</b>	42
				688	<b>1</b>	1
			<b>527</b>	470	<b>3</b>	3
			<b>546</b>	549	<b>58</b>	58

				1634	<b>50</b>	50
			<b>548</b>	1529	<b>2</b>	2
			<b>549</b>	546	<b>58</b>	58
				1904	<b>28</b>	28
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
				1907	<b>3</b>	3
			<b>564</b>	584	<b>3</b>	3
				1138	<b>15</b>	15
			<b>584</b>	564	<b>3</b>	3
				597	<b>1</b>	1
					<b>11</b>	11
			<b>597</b>	584	<b>1</b>	1
					<b>11</b>	11
				1589	<b>4</b>	4
				1896	<b>86</b>	86
				1904	<b>34</b>	34
			<b>611</b>	1138	<b>1</b>	1
			<b>659</b>	215	<b>1</b>	1
				967	<b>22</b>	22
			<b>688</b>	505	<b>1</b>	1
			<b>717</b>	166	<b>1</b>	1
				168	<b>12</b>	12
			<b>720</b>	166	<b>1</b>	1
				168	<b>1</b>	1
				760	<b>2</b>	2
			<b>748</b>	758	<b>1</b>	1
				795	<b>157</b>	157
				1088	<b>6</b>	6
			<b>758</b>	748	<b>1</b>	1
				772	<b>4</b>	4
				778	<b>62</b>	62
				782	<b>91</b>	91
				795	<b>142</b>	142
			<b>760</b>	457	<b>4</b>	4
				720	<b>2</b>	2
			<b>769</b>	778	<b>7</b>	7
			<b>772</b>	11	<b>1</b>	1
				470	<b>3</b>	3
				758	<b>4</b>	4
				782	<b>2</b>	2
				795	<b>6</b>	6
			<b>778</b>	758	<b>62</b>	62
				769	<b>7</b>	7
				795	<b>149</b>	149
				2007	<b>1</b>	1

			<b>782</b>	758	<b>91</b>	91
				772	<b>2</b>	2
				795	<b>87</b>	87
			<b>795</b>	2	<b>15</b>	15
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
				7	<b>4</b>	4
				748	<b>157</b>	157
				758	<b>142</b>	142
				772	<b>6</b>	6
				778	<b>149</b>	149
				782	<b>87</b>	87
				795	<b>2</b>	4
			<b>956</b>	431	<b>6</b>	6
					<b>13</b>	13
				439	<b>4</b>	4
			<b>967</b>	304	<b>3</b>	3
				398	<b>6</b>	6
				414	<b>33</b>	33
				417	<b>68</b>	68
				659	<b>22</b>	22
				1529	<b>1</b>	1
			<b>982</b>	59	<b>2</b>	2
			<b>1051</b>	91	<b>2</b>	2
				1055	<b>3</b>	3
				1423	<b>1</b>	1
			<b>1055</b>	1051	<b>3</b>	3
			<b>1060</b>	2121	<b>1</b>	1
			<b>1088</b>	748	<b>6</b>	6
				1138	<b>11</b>	11
				1150	<b>2</b>	2
				1158	<b>5</b>	5
			<b>1118</b>	1729	<b>5</b>	5
				1733	<b>3</b>	3
			<b>1133</b>	2090	<b>2</b>	2
			<b>1138</b>	398	<b>1</b>	1
				564	<b>15</b>	15
				611	<b>1</b>	1
				1088	<b>11</b>	11
				1150	<b>79</b>	79
				1423	<b>1</b>	1
			<b>1150</b>	1088	<b>2</b>	2
				1138	<b>79</b>	79
			<b>1158</b>	1088	<b>5</b>	5
				1187	<b>13</b>	13
			<b>1187</b>	257	<b>1</b>	1

				1158	<b>13</b>	13
				1589	<b>1</b>	1
			<b>1201</b>	2154	<b>1</b>	1
			<b>1315</b>	304	<b>1</b>	1
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
				1529	<b>2</b>	2
			<b>1316</b>	304	<b>3</b>	3
			<b>1372</b>	1404	<b>18</b>	18
				1676	<b>56</b>	56
			<b>1404</b>	1372	<b>18</b>	18
				1653	<b>4</b>	4
				1676	<b>18</b>	18
			<b>1414</b>	31	<b>1</b>	1
				1423	<b>6</b>	6
				1427	<b>1</b>	1
				1653	<b>2</b>	2
			<b>1423</b>	1051	<b>1</b>	1
				1138	<b>1</b>	1
				1414	<b>6</b>	6
			<b>1427</b>	1414	<b>1</b>	1
			<b>1437</b>	2116	<b>5</b>	5
				2121	<b>3</b>	3
					<b>102</b>	102
			<b>1441</b>	2121	<b>11</b>	11
			<b>1529</b>	74	<b>5</b>	5
				82	<b>1</b>	1
				304	<b>11</b>	11
				390	<b>1</b>	1
				548	<b>2</b>	2
				967	<b>1</b>	1
				1315	<b>2</b>	2
			<b>1589</b>	90	<b>1</b>	1
				597	<b>4</b>	4
				1187	<b>1</b>	1
				1896	<b>24</b>	24
				1904	<b>122</b>	122
			<b>1600</b>	1634	<b>40</b>	40
			<b>1603</b>	1634	<b>189</b>	189
			<b>1634</b>	546	<b>50</b>	50
				1600	<b>40</b>	40
				1603	<b>189</b>	189
			<b>1653</b>	1404	<b>4</b>	4
				1414	<b>2</b>	2
				1676	<b>5</b>	5
			<b>1676</b>	187	<b>1</b>	1



				190	<b>1</b>	1
				1372	<b>56</b>	56
				1404	<b>18</b>	18
				1653	<b>5</b>	5
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
			<b>1729</b>	1118	<b>5</b>	5
			<b>1733</b>	276	<b>3</b>	3
				398	<b>2</b>	2
				1118	<b>3</b>	3
			<b>1896</b>	597	<b>86</b>	86
				1589	<b>24</b>	24
				1904	<b>10</b>	10
			<b>1904</b>	549	<b>28</b>	28
				597	<b>34</b>	34
				1589	<b>122</b>	122
				1896	<b>10</b>	10
			<b>1907</b>	549	<b>3</b>	3
			<b>2007</b>	778	<b>1</b>	1
			<b>2070</b>	2115	<b>9</b>	9
			<b>2090</b>	1133	<b>2</b>	2
			<b>2115</b>	2070	<b>9</b>	9
			<b>2116</b>	1437	<b>5</b>	5
			<b>2121</b>	1060	<b>1</b>	1
				1437	<b>3</b>	3
					<b>102</b>	102
				1441	<b>11</b>	11
			<b>2137</b>	2163	<b>1</b>	1
			<b>2154</b>	1201	<b>1</b>	1
			<b>2163</b>	2137	<b>1</b>	1
<b>SAD1TAP</b>	<b>inter-molecular</b>	BRR2	<b>6</b>	376	<b>1</b>	1
			<b>207</b>	1529	<b>1</b>	1
			<b>564</b>	1589	<b>1</b>	1
				1623	<b>2</b>	2
			<b>568</b>	1589	<b>1</b>	1
				1623	<b>2</b>	2
			<b>614</b>	1589	<b>1</b>	1
		PRP6	<b>266</b>	89	<b>3</b>	3
					<b>6</b>	6
				101	<b>5</b>	5
			<b>270</b>	89	<b>10</b>	10
				183	<b>1</b>	1
		PRP8	<b>315</b>	810	<b>6</b>	6
			<b>319</b>	810	<b>7</b>	7
			<b>322</b>	810	<b>1</b>	1

			<b>389</b>	817	<b>2</b>	2
				819	<b>2</b>	2
			<b>403</b>	1242	<b>12</b>	12
		RSMB	<b>251</b>	145	<b>3</b>	3
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
			<b>263</b>	138	<b>2</b>	2
				145	<b>5</b>	5
			<b>266</b>	131	<b>1</b>	1
				132	<b>2</b>	2
				135	<b>16</b>	16
				138	<b>7</b>	7
					<b>20</b>	20
				145	<b>28</b>	28
				186	<b>15</b>	15
				194	<b>1</b>	1
			<b>270</b>	131	<b>2</b>	2
				132	<b>1</b>	1
				135	<b>4</b>	4
					<b>6</b>	6
				138	<b>10</b>	10
					<b>11</b>	11
				145	<b>20</b>	20
				186	<b>6</b>	6
			<b>284</b>	138	<b>3</b>	3
				145	<b>3</b>	3
			<b>322</b>	186	<b>1</b>	1
			<b>457</b>	145	<b>1</b>	1
		SN114	<b>6</b>	461	<b>9</b>	9
				520	<b>24</b>	24
				587	<b>2</b>	2
				749	<b>19</b>	19
				790	<b>5</b>	5
				792	<b>3</b>	3
			<b>15</b>	461	<b>1</b>	1
				520	<b>66</b>	66
				587	<b>4</b>	4
				749	<b>1</b>	1
				790	<b>1</b>	1
			<b>20</b>	520	<b>90</b>	90
				587	<b>1</b>	1
				790	<b>6</b>	6
			<b>21</b>	520	<b>7</b>	7
			<b>23</b>	520	<b>120</b>	120
				790	<b>8</b>	8
				835	<b>4</b>	4

			<b>73</b>	520	<b>3</b>	3
			<b>220</b>	617	<b>30</b>	30
			<b>261</b>	60	<b>1</b>	1
			<b>263</b>	60	<b>11</b>	11
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
			<b>322</b>	59	<b>2</b>	2
				60	<b>15</b>	15
				72	<b>1</b>	1
			<b>446</b>	60	<b>1</b>	1
			<b>452</b>	60	<b>3</b>	3
				81	<b>2</b>	2
			<b>457</b>	59	<b>11</b>	11
				60	<b>9</b>	9
				72	<b>3</b>	3
				81	<b>7</b>	7
		SNU66	<b>564</b>	375	<b>3</b>	3
			<b>568</b>	375	<b>7</b>	7
		SP381	<b>2</b>	264	<b>1</b>	1
			<b>284</b>	142	<b>1</b>	1
	<b>intra-molecular</b>	SAD1TAP	<b>2</b>	6	<b>1</b>	1
				20	<b>1</b>	1
				322	<b>1</b>	1
			<b>6</b>	2	<b>1</b>	1
				15	<b>57</b>	57
				20	<b>1</b>	1
					<b>136</b>	136
				23	<b>20</b>	20
				204	<b>3</b>	3
			<b>15</b>	6	<b>57</b>	57
				21	<b>19</b>	19
				23	<b>64</b>	64
			<b>20</b>	2	<b>1</b>	1
				6	<b>1</b>	1
					<b>136</b>	136
				23	<b>120</b>	120
			<b>21</b>	15	<b>19</b>	19
			<b>23</b>	6	<b>20</b>	20
				15	<b>64</b>	64
				20	<b>120</b>	120
			<b>118</b>	204	<b>4</b>	4
			<b>128</b>	204	<b>1</b>	1
				457	<b>5</b>	5
			<b>204</b>	6	<b>3</b>	3
				118	<b>4</b>	4

				128	<b>1</b>	1
			<b>220</b>	403	<b>6</b>	6
			<b>259</b>	457	<b>1</b>	1
			<b>261</b>	457	<b>1</b>	1
						<b>Values</b>
<b>Protein 1</b>	<b>XL_type</b>	<b>Protein 2</b>	<b>Residue1</b>	<b>Residue2</b>	<b>CSMs</b>	<b>Count of Scan</b>
			<b>263</b>	270	<b>2</b>	2
				322	<b>18</b>	18
				457	<b>8</b>	8
			<b>266</b>	284	<b>46</b>	46
				319	<b>10</b>	10
				322	<b>13</b>	13
			<b>270</b>	263	<b>2</b>	2
				284	<b>2</b>	2
				322	<b>13</b>	13
			<b>284</b>	266	<b>46</b>	46
				270	<b>2</b>	2
				319	<b>1</b>	1
			<b>319</b>	266	<b>10</b>	10
				284	<b>1</b>	1
			<b>322</b>	2	<b>1</b>	1
				263	<b>18</b>	18
				266	<b>13</b>	13
				270	<b>13</b>	13
				452	<b>7</b>	7
				457	<b>13</b>	13
			<b>357</b>	389	<b>59</b>	59
			<b>376</b>	457	<b>4</b>	4
			<b>380</b>	457	<b>3</b>	3
			<b>389</b>	357	<b>59</b>	59
			<b>403</b>	220	<b>6</b>	6
			<b>414</b>	457	<b>4</b>	4
			<b>446</b>	457	<b>16</b>	16
			<b>452</b>	322	<b>7</b>	7
				456	<b>7</b>	7
				457	<b>10</b>	10
			<b>456</b>	452	<b>7</b>	7
			<b>457</b>	128	<b>5</b>	5
				259	<b>1</b>	1
				261	<b>1</b>	1
				263	<b>8</b>	8
				322	<b>13</b>	13
				376	<b>4</b>	4
				380	<b>3</b>	3
				414	<b>4</b>	4

## Acknowledgements

I would like to express my warmest gratitude to Prof. Dr. Reinhard Lührmann for providing me an opportunity to work on this project. I shall always remain thankful for his unconditional support, guidance, and illuminating discussions, and above all his down to earth approachable persona, which I enjoy during the completion of this work. I wholeheartedly thank Prof. Dr. Ralf Ficner for his wonderful suggestions and lively discussions during the course of our meeting. His kind and approachable nature always encouraged my scientific thinking and immensely boosted the quality of my work.

Most humble thanks go to Prof. Dr. Henning Urlaub, Prof. Dr. Hoger Stark, Prof. Dr. Jörg Großhans, Prof. Dr. Ralph Kehlenbach for accepting my invitation to participate in my examination committee.

Here I am trying my best to express in words, however, I know for certain that it is impossible for me to express in words, my deepest thanks to Dr. Patrizia Fabrizo and Dr. Reinhard Rauhut for teaching me valuable tools and techniques, and for guiding me throughout this project. I will always remember them for shaping my scientific aptitude and enriching my work thesis discussions. Their critical evaluations, endless support, ingenious inputs to tackle challenges helped me in achieving the aims of this project.

I appreciate the help of Dr. Norbert Rigo, Dr. Ilya Komarov, and Dr. Olex Dybkov, Dr. Leelaram Majety for their wonderful collaboration and support as without their expertise this work could not have been finished on time.

Further thanks go to very special people whose motivation and support helped in the completion of this Ph.D. thesis: Especially to Dr. Berthold Kastner and Dr. Prakash Dube, and to Dr: Klaus Hartmuth, Dr. Leyla El Ayoubi, Dr. Dimitry Agafonov, and Dr. Cindy Will

I would like to thank Thomas Conrad & Thomas Schulz for the cultivation of yeast cells. Gertrud Nowak and Claudia Fahlbusch, Winfried Lendeckel, Gabi Heyne, Marion Killian, Irene Öchsner, Ulrich Steuerwald, for their excellent technical support.

My very special and heartfelt thanks go to my dear Juliane Moses for being so much supportive, fun, and doing multitasking for managing all formal and administrative work involved.

To my friends, I offer my love for making my Ph.D. journey a little more fun, enjoyable and to make me feel like home from home...Dear Almu, Csaba and Judit, Norbert, Sebastian, Ilya, Sandra, Evelina, Carsten, Sergej, Masha, and Michael, I thank you all!!

I could not have completed this work without the support of my wife. Please allow me to share credits for this work with my loving wife Harsh, and my parents for their constant motivation, support, and unconditional affection.