Functional and structural investigation of spliceosomal snRNPs

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Georg-August-Universität zu Göttingen

vorgelegt von

Simon Trowitzsch

aus Northeim

Göttingen 2008

D7

Referent: Prof. Dr. R. Ficner

Korreferent: Prof. Dr. O. Einsle

Tag der mündlichen Prüfung: 2. Juli 2008

Table of contents

List of figures

List of tables

Abstract

The expression of most eukaryotic protein-encoding genes involves precursor messenger RNA (pre-mRNA) processing steps including pre-mRNA splicing. Pre-mRNA splicing is catalyzed by a multi-subunit RNA-protein enzyme, the spliceosome, which emerges from the stepwise recruitment of the U1, U2, U5 and U4/U6 small nuclear ribonucleoprotein particles (snRNPs) and numerous non-snRNP proteins to conserved sequences of the pre-mRNA substrate. The U2 snRNP is essential for splicing in yeast and humans, participating in the recognition/selection of the so-called branch site (BS) of the pre-mRNA during spliceosome assembly and also during the subsequent catalysis of splicing. 12S U2 snRNPs are composed of U2 snRNA, which is complexed by U2 snRNP specific proteins, U2-A' and U2-B", and seven Sm proteins. In catalytically active human 17S U2 snRNPs, two heteromeric splicing factors, SF3a and SF3b, are additionally found, which contain three (SF3a120, SF3a66 and SF3a60) and seven (SF3b155, SF3b145, SF3b130, SF3b49, SF3b14a/p14, SF3b14b and SF3b10) proteins, respectively. In the yeast *Sacharomyces cerevisiae*, SF3b complexes were shown to consist of yeast orthologs Hsh155p, Rse1p, Cus1p, Hsh49p, Rds3p, Rcp10p/Ysf3p, and at least two additional proteins Bud31p and Snu17p/Ist3p. Moreover, Snu17p is found in a complex with proteins Pml1p and Bud13p. Together, these three proteins form the pre-mRNA retention and splicing (RES) complex, which counteracts the escape of unspliced pre-mRNAs from the nucleus and activates splicing of a subset of Mer1p-dependent genes. A homologous complex is present in human pre-catalytic, activated and step 1 spliceosomes. Structural analyses by electron microscopy, nuclear magnetic resonance spectroscopy and X-ray crystallography gave important insight into structural arrangements among 17S U2 snRNP subcomplexes and proteins, but atomic models of 12S U2 snRNPs, SF3b particles and the RES complex remained elusive.

To derive atomic models of entire human 12S U2 snRNPs and SF3b particles, I set out to produce these particles for X-ray crystallographic analyses. I was able to optimize purification protocols for both natively isolated 12S U2 snRNPs and SF3b

Abstract

complexes. Since no crystals could be obtained from native 12S U2 snRNPs, presumed unstructured regions of U2 snRNA, Sm proteins Sm B/B' and U2-A' and U2-B" were successfully removed by a DNA-directed RNaseH-based cleavage and limited proteolysis, respectively. U2 snRNP particles with truncated snRNA or a truncated snRNA and truncated proteins Sm B/B' were subjected to crystallization at a concentration of at least 10 mg/ml. U2 snRNP particles with truncated snRNA, truncated proteins Sm B/B' and truncated U2-A' and U2-B" were obtained in analytical scale. The established purification protocols would allow preparative production in the future. SF3b particles were purified and screened for crystal growth at concentrations of at least 9 mg/ml. During the preparation of SF3b particles, a yet unknown complex could be isolated and characterized. It comprises at least 4 proteins, including the DEAD-box helicase DDX1 and proteins HSPC117, Family with sequence similarity 98/member B and CGI-99. Its association with SF3b particles is discussed.

To expand knowledge about Snu17p and its interaction partners, I set out to characterize yeast RES complexes by biophysical and biochemical methods, with the ultimate goal to elucidate the molecular architecture of the RES complex. GST pull-down experiments and size exclusion chromatography revealed that Snu17p constitutes the central platform of the complex, while Bud13p and Pml1p do not interact with each other. Fluorimetric structure probing showed the entire Bud13p and the N-terminal third of Pml1p to be natively disordered in isolation. Mutational analysis and tryptophan fluorescence confirmed that a conserved tryptophan-containing motif in the C-terminus of Bud13p binds to the core RRM of Snu17p, while a different interaction surface encompassing a C-terminal extension of the Snu17p RRM is required to bind an N-terminal peptide of Pml1p. Isothermal titration calorimetry revealed 1:1 interaction stoichiometries, large negative binding entropies and dissociation constants in the low nanomolar and micromolar ranges for the Snu17p-Bud13p and the Snu17p-Pml1p interactions, respectively. By performing peptide-scanning experiments, the Snu17p interacting regions of Bud13p and Pml1p could be further delineated with single amino acid resolution. Preliminary results from NMR spectroscopy and limited proteolysis indicate that Snu17p persists in a molten globule like structure and folds upon binding of at least one of the interating peptides. Thus, the non-canonical Snu17p RRM concomitantly binds multiple ligand

proteins *via* short, intrinsically unstructured peptide epitopes and thereby acts as a platform that displays functional modules of the ligands, such as a forkhead-associated domain of Pml1p and a conserved poly-lysine motif of Bud13p.

The crystal structures of full-length and N-terminally truncated Pml1p were determined and revealed the presence of a forkhead-associated (FHA) domain fold in the C-terminal region. FHA domains are small protein modules, which bind phosphorylated epitopes on proteins. The first 50 residues of Pml1p, encompassing the Snu17p-binding region, are disordered. A non-canonical N-terminal expansion runs across one β-sheet and, thereby, critically stabilizes the domain. Structure based alignments identified a similar expansion in the human protein NIPP1, which was previously shown to be associated with spliceosomes. A sulfate ion was found at the putative phosphopeptide-binding loops of full-length Pml1p while the truncated protein lacked a similar phosphopeptide mimic but retained an almost identical structure. A long loop neighboring the phosphopeptide-binding site was disordered in both structures and may confer ligand specificity. It is speculated that Pml1p most likely recognizes the phosphorylated amino acid of ligands by a lock-and-key mechanism, while specificity relies on induced-fit interactions. The results suggest that Snu17p recruits Pml1p as a phosphorylation sensor to the spliceosome.

Finally, a model of the molecular architecture of the RES complex is presented.

1.   Introduction

Throughout all cells, from prokaryotes to higher eukaryotes, the expression of information encoded in a genome follows the same fundamental principle. Genetic information is converted from linear desoxyribonucleic acid (DNA) sequences, via ribonucleic acids, to primary amino acid sequences. In contrast to prokaryotes, eukaryotic precursor-messenger RNAs (pre-mRNAs) undergo extensive co- and posttranscriptional processing prior to their translation into functional proteins by the ribosome. First, the pre-mRNA is transcribed in the nucleus and modified at the 5' end by attachment of a N7-methylguanosine cap (m^7G-cap) and by 2'-O-methylation of the first two nucleotides. Non-coding sequences are removed during splicing and the pre-mRNA is cleaved at the 3' end and polyadenylated. The mature mRNA is exported through the nuclear pore complex into the cytoplasm for translation. Anomalously processed or mutated pre- or mature mRNAs are eliminated via survey mechanisms, like non-sense-mediated decay (NMD), No-Go decay (NGD), nonstop decay (NSD), or ribosome extension-mediated decay (REMD) (Doma and Parker, 2006; Frischmeyer et al., 2002; Inada and Aiba, 2005; Isken and Maquat, 2007; Kong and Liebhaber, 2007; Maquat and Carmichael, 2001; van Hoof et al., 2002). Although each step of eukaryotic gene expression is driven by a different machinery, all processes seem to be physically and functionally coupled (Reed, 2003). The question remains, whether splicing takes place co- or posttranscriptionally (Gornemann et al., 2005; Nilsen, 2005; Tardiff et al., 2006).

During splicing in eukaryotes, non-coding intervening sequences (introns) are excised and the neighboring coding regions (exons) are ligated to produce mature mRNA that can serve as a template for protein biosynthesis. Each round of splicing entails two sequential transesterification reactions, which are carried out by a large RNA-protein enzyme, the spliceosome (Nilsen, 2003). The main task of the spliceosome lies in locating and bringing together the sites at which the cut-and-paste reactions have to proceed with single nucleotide precision. For each splicing event, a spliceosome is, therefore, assembled *de novo* in a stepwise fashion, undergoing a

multitude of compositional and conformational rearrangements, until a catalytically competent particle is correctly positioned on the pre-mRNA substrate (Brow, 2002; Will and Lührmann, 2006). Spliceosome assembly encompasses the ordered recruitment of complex RNA-protein subunits, small nuclear ribonucleoprotein particles (snRNPs), as well as many non-snRNP splicing factors (Jurica and Moore, 2003; Will and Lührmann, 2006). In total, five snRNAs and more than 150 proteins participate in this process (Jurica and Moore, 2003; Nilsen, 2003). More than 30-40 % of human genes are presumably alternatively spliced (Johnson et al., 2003; Modrek and Lee, 2002) and the evolution of a highly regulated splicing machinery may account for manifold *trans*-acting factors.

1.1   The mechanism of the pre-mRNA splicing reaction

Introns and intron/exon boundaries contain a set of specific sequences, which are required for splicing (Stephens and Schneider, 1992). In higher eukaryotes, the 5' splice site is characterized by the consensus sequence 5'-AG|GURAGU-3' (vertical tick - splice site; underlined - invariable ribonucleotides; R - purine base) (Will and Luhrmann, 1999). The 3' splice site in most of the introns of vertebrates follows the sequence 5'-YAG|-3' (Y - pyrimidine base) and a pyrimidine-rich, 10-12 nucleotide (nt) long region upstream of the AG dinucleotide. The AG dinucleotide thereby defines the end of an intron (Reed, 1989). The branch point sequence (BPS), containing a conserved adenosine, is normally located 18-40 nt upstream of the 3' splice site and comprises a degenerated sequence 5'-YNYUR**A**C-3' (**A** - branch adenosine; N - any nucleotide) (Reed and Maniatis, 1988). In the budding yeast *Saccharomyces cerevisiae* these elements are far more conserved and the branch site is invariant (5'-UACUA**A**C-3') (Lin et al., 1985; Rymond and Rosbash, 1992). A less abundant class of introns in metazoa, differing in their consensus sequences, was found (Jackson, 1991; Rymond and Rosbash, 1992). Among these introns, the sequences 5'-|AUAUCCUUU-3' and 5'-YAC|-3' represent highly conserved elements at 5' and 3' splice sites. These "ATAC"-introns account for less then 1 % of all introns and are removed by a distinct spliceosome, the so-called minor spliceosome (Hall and Padgett, 1994; Hall and Padgett, 1996; Patel and Steitz, 2003).

The introduction of cell-free systems allowed the unravelling of fundamental mechanisms of intron removal from precursor messenger RNAs catalyzed by the spliceosome (Beggs et al., 1995; Krainer et al., 1984; Lin et al., 1985; Padgett et al., 1983). All pre-mRNA introns are removed by an identical energy-independent mechanism, which involves two subsequent stereospecific transesterification reactions of the S_N2-type (Maschhoff and Padgett, 1993; Moore et al., 1993; Moore and Sharp, 1993; Ruskin et al., 1984). The 2' hydroxyl group of the branch point adenosine carries out a nucleophilic attack at the 5' splice site in step 1. The guanosine at the 5' end of the intron is thereby connected to the branch point adenosine via a 2'-5' phosphodiester. The free 5' exon and the intron-3'-exon lariat are formed as splicing intermediates. The second step results in the excision of the lariat intron and ligation of the 5' and 3' exons through the attack of the 5' exon's 3' OH group at the 3' splice site. The spliced out introns are usually debranched by nuclear 2'-5'-phosphodiesterase and subsequently degraded (Burge et al., 1999). The splicing reaction is proposed being catalyzed by RNA, since the consensus sequences at 5' and 3' splice sites of pre-mRNAs resemble those of self-splicing group II introns (Cech, 1986; Weiner, 1993). Furthermore, a general two-metal-ion mechanism for the phospho-transesterification reaction in splicing was described, in which the RNA positions the two catalytic metal ions and properly orient the substrate (Steitz and Steitz, 1993).

1.2   Components of the spliceosome

To date, two spliceosomes are known, the major or U2-dependent and the minor or U12-dependent spliceosomes (Dietrich et al., 1997; Sharp and Burge, 1997). Although major spliceosomes recognize canonical "GTAG"-introns and minor spliceosomes "ATAC"-introns, the selectivity is not only defined by the 3' and 5' splice site dinucleotides, but also through consensus sequences at the 5' splice sites and the branch points (Dietrich et al., 1997; Wu and Krainer, 1997). Since these *cis*-acting elements of the pre-mRNAs are rather short and weakly conserved (especially in higher eukaryotes), they are not sufficient to unambiguously define the exon-intron boundaries. As a consequence, spliceosome machineries evolved numerous *trans*-acting factors, which, in multiple steps, recognize the *cis*-acting elements of

pre-mRNA during spliceosome assembly, ensuring the fidelity of splicing (Will and Lührmann, 2006). Further *cis*-acting elements, like exonic and intronic splicing enhancers (ESEs and ISEs), or silencers (ESSs and ISSs), are thought to additionally modulate both constitutive and alternative splicing (Blencowe, 2000; Smith and Valcarcel, 2000). The *trans*-acting factors of the spliceosome are the U snRNPs and a multitude of non-snRNP proteins, which will be described in the following sections.

1.2.1   The spliceosomal U snRNPs

U2- and U12-dependent spliceosomes contain subunits, which encompass a specific uridin-rich small nuclear RNA (snRNA) and a set of common and specific proteins. The U2-dependent spliceosome consists of U1, U2 and U5 snRNPs and a U4/U6 di-snRNP. In U12-dependent spliceosomes U1, U2 and U4/U6 snRNPs are exchanged by U11, U12 and $U4_{ATAC}/U6_{ATAC}$ snRNPs and only U5 snRNPs are common to both particles. The U11 and U12 snRNPs are the functional analogs of the U1 and U2 snRNPs, respectively, whereas the $U4_{ATAC}/U6_{ATAC}$ snRNP is the functional analog of U4/U6 (Hall and Padgett, 1996; Kolossova and Padgett, 1997; Tarn and Steitz, 1996a; Tarn and Steitz, 1996b; Yu and Steitz, 1997). The U snRNAs are generally relatively short and their primary and secondary structures are phylogenetically conserved (Guthrie and Patterson, 1988)(Fig. 1.1). With the exception of U6 and $U6_{ATAC}$ snRNAs, all snRNAs are transcribed by RNA-polymerase II and have a co-transcriptionally attached N7-methylguanosine cap (m^7G cap) at their 5' termini, which is hypermethylated to a 2,2,7-trimethylguanosine cap ($m_3^{2,2,7}$G cap) during U snRNP maturation in the cytoplasm. This $m_3^{2,2,7}$G cap structure has been widely used to isolate U snRNPs from nuclear extracts via a specific antibody (Bach et al., 1990; Bochnig et al., 1987; Bringmann and Lührmann, 1986; Bringmann et al., 1983a; Bringmann et al., 1983b; Lührmann et al., 1982).
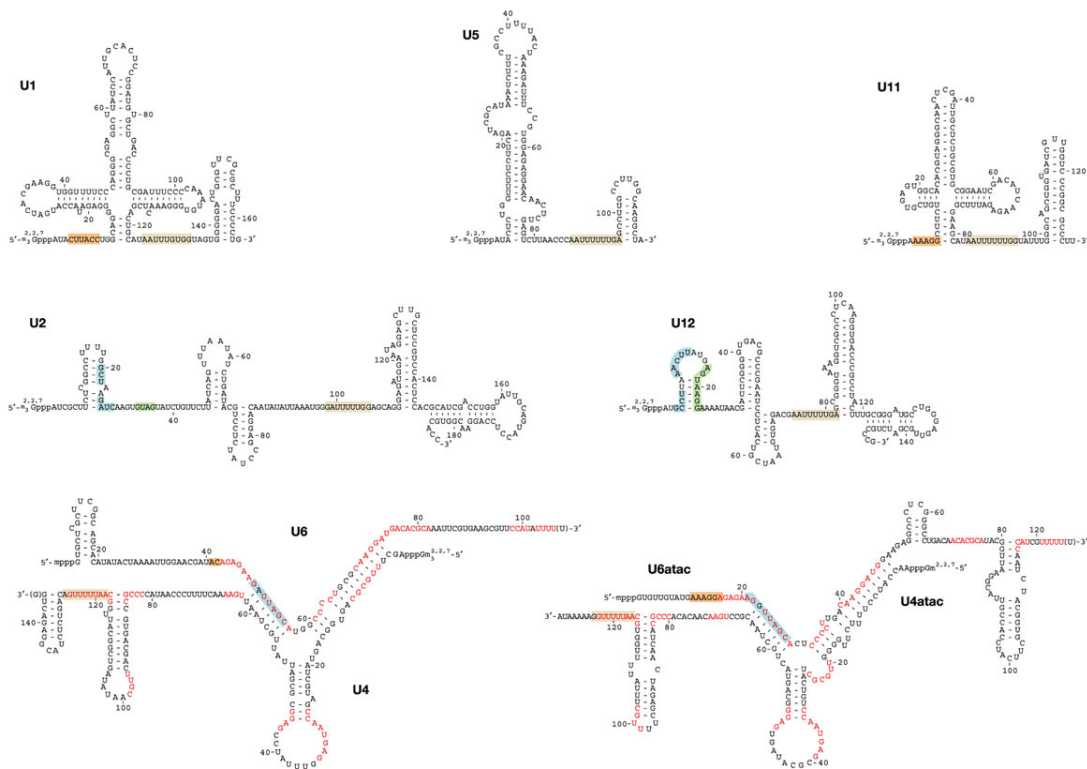
Fig. 1.1: Sequences and predicted secondary structures of human spliceosomal snRNAs. Sm sites are shaded in light yellow. Colored boxes indicate sequences that are predicted to be involved in intermolecular RNA-RNA base-pairing interactions: orange for interactions with the 5' splice site, green for interactions with the branch site, and blue for U2-U6 or U12-U6$_{ATAC}$ helix I interactions. Sequences in red represent stretches of four or more identical nucleotides between U4-U6 and U4$_{ATAC}$-U6$_{ATAC}$ (adopted from Patel and Steitz, 2003)

U6 and U6$_{ATAC}$ snRNAs are transcribed by RNA-polymerase III and carry a γ-monomethylphosphate at their 5' termini (Kunkel et al., 1986; Reddy et al., 1987; Singh and Reddy, 1989). Small nucleolar ribonuclear proteins (snoRNPs) guide 2'-O-methylation and pseudouridinylation of U snRNAs (Jady and Kiss, 2001; Kiss et al., 2004). Furthermore, some residues of U snRNAs show conversion to N[6]-methyl-adenosins. Also snoRNP-independent pseudouridinylation has been described (Ma et al., 2005).

With the exception of U6 and U6$_{ATAC}$ snRNAs, U snRNAs share the presence of a consensus sequence 5'-RAU$_{4-6}$GR-3', which is essential for the binding of a set of basic proteins, the Sm proteins (Branlant et al., 1982; Jarmolowski and Mattaj, 1993;

Jones and Guthrie, 1990; Raker et al., 1999; Raker et al., 1996). The Sm proteins derive their names from autoantibodies isolated from a patient called Smith and are core components of each snRNP (except U6 and U6$_{ATAC}$ snRNPs, which contain Sm-like - or LSm - proteins) and are loaded onto the highly conserved Sm-site with the help of the multiprotein-SMN complex in the cytoplasm (Fischer et al., 1997; Meister et al., 2001; Pellizzoni et al., 1998).

Seven U snRNP Sm-proteins, Sm B/B', D1, D2, D3, E, F and G, have been described. Sm B and B' proteins are isoforms created by alternative splicing of a single gene and differ only in their C-terminal 11 amino acids (van Dam et al., 1989). These proteins form subcomplexes *in vitro* (E-F-G, D1-D2 and B/B'-D3) and assemble *in vivo* at Sm-sites of snRNAs U1, U2, U4 and U5 to doughnut-shaped hetero-heptamers, the so-called Sm core RNPs (Hermann et al., 1995; Raker et al., 1999; Raker et al., 1996; Urlaub et al., 2001). All Sm proteins contain a conserved bipartite sequence motif, Sm 1 and Sm 2 (Hermann et al., 1995), forming their characteristic fold that comprises a bent β-sheet with 5 antiparallel β-strands as shown by X-ray crystallographic analyses (Kambach et al., 1999). Biochemical and structural data suggest a model of a seven-membered Sm-ring formed by Sm proteins in the order G-E-F-D2-D1-B-D3 with RNA-protein contacts in the inner surface of the ring (Heinrichs et al., 1992; Kambach et al., 1999; Nagai et al., 2001; Urlaub et al., 2001). The $m_3^{2,2,7}$G cap structures and a properly assembled core protein complex on the respective snRNA contribute to the bipartite nuclear localization signal, which drives reimport of the core U snRNPs into the nucleus. U6 and U6$_{ATAC}$ snRNAs seem to persist in the nucleus throughout the assembly of U6 and U6$_{ATAC}$ snRNPs (Fischer and Lührmann, 1990; Fischer et al., 1993).

Similarly to Sm-proteins, the LSm-proteins LSm2-8 associate with U6 and U6$_{ATAC}$ snRNAs in a highly ordered fashion as a heptameric ring on uridine-rich sequences at the 3' end of the snRNAs (Achsel et al., 1999; Ingelfinger et al., 2002; Séraphin, 1995; Vidal et al., 1999).

1.2.2   U1 snRNPs

Besides the Sm-proteins, the U1 snRNP contains particle-specific proteins U1-70k, U1-A and U1-C (k stands for kDa). Proteins U1-70k and U1-A interact via their canonical RNA binding motifs (RRMs) directly with loop-regions of hairpins I and II of U1 snRNA, respectively (Nelissen et al., 1994; Patton et al., 1989; Patton and Pederson, 1988; Query et al., 1989a; Query et al., 1989b; Urlaub et al., 2000). A high-resolution crystal structure of a complex between RRM 1 of U1-A and the snRNA stem loop served as a hallmark of RRM-RNA interactions (Oubridge et al., 1994). U1-C, which has a zinc finger-like motif, is associated with U1 snRNA in the presence of U1-70k and the Sm-proteins (Nelissen et al., 1991; Nelissen et al., 1994) and binds pre-mRNA directly (Du and Rosbash, 2002; Heinrichs et al., 1990). A high-resolution cryo-EM structure at 10 Å resolution (Stark et al., 2001) and localization of U1-70k, U1-A and the $m_3^{2,2,7}G$ cap by protein depletion and antibody binding experiments led to a rough model of the molecular architecture of U1 snRNP (Kastner et al., 1992; Kastner and Lührmann, 1989).

1.2.3   U2 snRNPs

The U2 snRNP plays an important role at the early stages of splicing. The particle contributes via its protein components to the formation of pre-spliceosomes and its RNA moiety probably contributes, together with U6 snRNA, to the active sites of the spliceosome (Burge et al., 1999; Will and Luhrmann, 2001b). U2 snRNPs were initially characterized as 12S particles consisting of the Sm-proteins and two particle specific proteins: U2-A' and U2-B" (Fig. 1.2).

Using mild purification conditions, a 17S particle was observed. Proteomic and structural studies identified twelve 17S U2 snRNP specific proteins, forming the main body of the particle, and an additional set of proteins, which are loosely associated with the complex (U2-related proteins) (Behrens et al., 1993; Kramer et al., 1999; Will et al., 2001; Will et al., 2002).
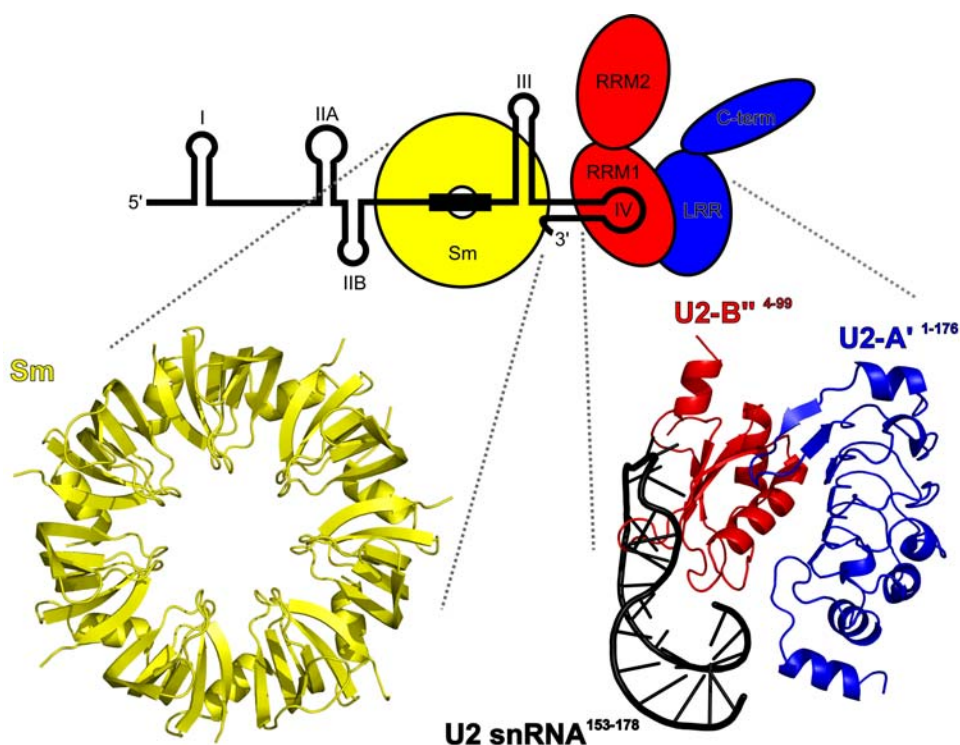
Fig. 1.2: 12S U2 snRNP. Schematic representation of 12S U2 snRNP with the U2 snRNA in black (Sm site is boxed), Sm proteins in yellow, U2-A' in blue and U2-B" in red. Predicted stem loops I, IIA, IIB, III and IV of the U2 snRNA are indicated. Molecular models in cartoon representation of an archeal Sm ring in yellow (left) (Toro et al., 2001) and of the U2-A'/U2-B"/U2 snRNA trimeric complex in blue, red and black, respectively (right) (Price et al., 1998). First and last amino acid residues of protein fragments of U2-A' and U2-B" and ribonucleotides of the U2 snRNA are indicated. RRM - RNA recognition motif; LRR - leucine-rich repeat; C-term - carboxy terminal region.

Under high salt conditions, 17S particles dissociate into 12S U2 snRNPs and two multi-protein splicing factors, the trimeric SF3a, consisting of proteins SF3a120, SF3a66 and SF3a60, and the heptameric SF3b, harboring proteins SF3b155, SF3b145, SF3b130, SF3b49, SF3b14a/p14, SF3b14b and SF3b10 (Will and Lührmann, 2006). 17S particles assemble stepwise from 12S U2 snRNPs by the interaction of SF3a with a 15S U2 snRNP intermediate, consisting of U2 snRNPs and SF3b (Brosi et al., 1993; Kramer et al., 1999). The SF3a and SF3b proteins are required for A complex formation and remain tightly associated during the first stages of the splicing cycle but seem to be destabilized at later stages (Bessonov et al., 2008; Brosi et al., 1993; Jurica and Moore, 2003). Cross-linking studies revealed that most of SF3a and SF3b proteins, except SF3b130, SF3b14b and SF3b10, are in close proximity to the branch site in pre-mRNAs and that SF3b14a/p14 could be

directly cross-linked to the branch point adenosine (Gozani et al., 1996; Gozani et al., 1998; Will et al., 2001). Besides U2-A' and U2-B'', proteins SF3a60, SF3b49, SF3b14a/p14 and Sm G and Sm B were shown to directly contact U2 snRNA (Dybkov et al., 2006). Subunits of SF3a and SF3b complexes are phylogenetically highly conserved and homologous proteins display very similar functions (Dziembowski et al., 2004; Pauling et al., 2000; Wang and Rymond, 2003).

EM studies of SF3b particles revealed a shell like structure with a central density (Golas et al., 2003). The shell is opened as two half shells upon integration of SF3b into U11/U12 di-snRNPs (Golas et al., 2005). A crystal structure of SF3b14a/p14 in complex with a fragment derived from SF3b155 could be modeled into the central density of the EM map (Schellenberg et al., 2006) and showed that a C-terminal expansion of the RRM of SF3b14a/p14 is necessary for the interaction with SF3b155 (Fig. 1.3).
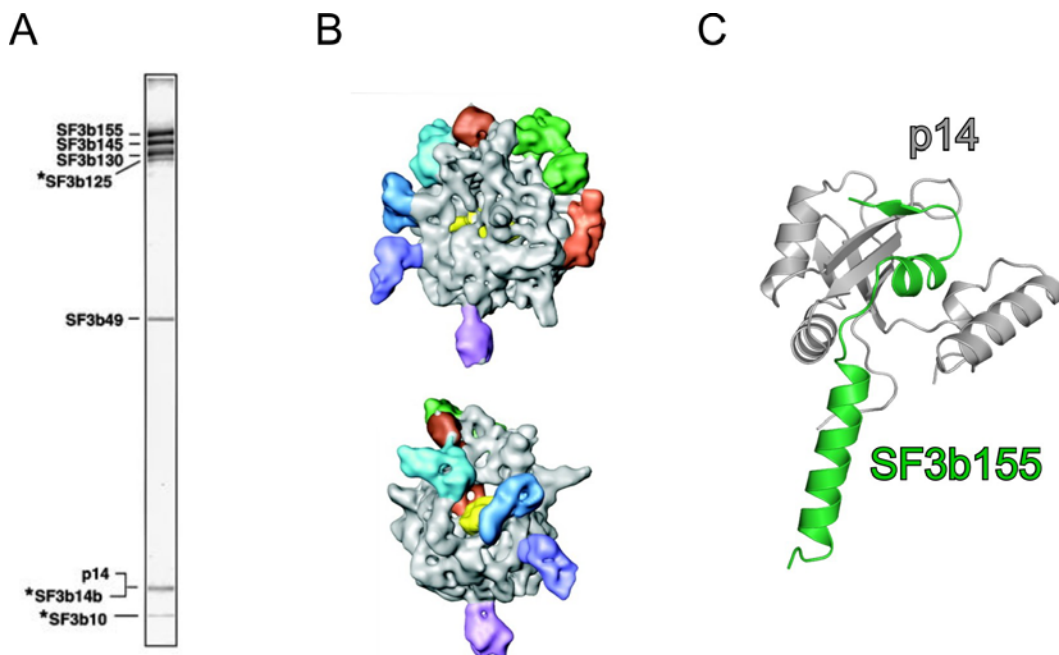
Fig. 1.3: Splicing factor 3b. **(A)** Protein composition of SF3b particles after affinity purification and glycerol gradient centrifugation. Protein identities are given (adopted from Will et al., 2002). **(B)** 3D surface representation of SF3b obtained by electron cryomicroscopy adopted from (Golas et al., 2003). Protuberances are colored; central density, corresponding to SF3b14a, is shown in yellow (adopted from Golas et al., 2003). **(C)** Crystal structure of a SF3b14a/SF3b155 dimer shown in cartoon representation. The C-terminal elongated RRM of SF3b14a(p14) is shown in grey and the fragment of SF3b155 in green (PDB code: 2F9D) (Schellenberg et al., 2006).

1.2.4   U4/U6.U5 tri-snRNPs

U4/U6.U5 tri-snRNPs are formed from 13S U4/U6 and 20S U5 snRNPs (Behrens and Lührmann, 1991; Black and Pinto, 1989). U5 snRNPs consist of particle-specific proteins with apparent molecular weights of 220k, 200k, 116k, 102k, 100k, 52k, 40k and 15k and 13S U4/U6 snRNPs comprise particle-specific proteins with apparent molecular weights of 90k, 61k, 60k, 20k (also known as CypH) and 15.5k (Bach et al., 1989; Behrens and Lührmann, 1991; Black and Pinto, 1989). Besides the Sm-proteins, which are associated with U4 and U5 snRNAs, U4/U6 snRNPs additionally contain the full set of LSm proteins (Vidal et al., 1999). Under physiological conditions (less than 250 mM salt), the 13S U4/U6 snRNP and 20S U5 snRNPs associate with three additional proteins (110k, 65k and 27k) to form the 25S U4/U6.U5 tri-snRNP. Upon tri-snRNP formation, the U5 snRNP specific protein 52k dissociates (Laggerbauer et al., 2005). The interactions among the proteins was studied by a yeast two-hybrid assay and revealed an extensive network of protein-protein interactions within tri-snRNPs (Liu et al., 2006). Crystal structures of a trimeric complex comprising 61k, 15.5k and U4 snRNA, as well as a fragment of 200k were recently solved (Liu et al., 2007; Pena et al., 2007).

1.3   Spliceosome assembly

The U2-type spliceosome is formed by the ordered interaction of U1, U2, and U4/U6.U5 snRNPs, as well as non-snRNP splicing factors, with the pre-mRNA (Reed and Palandjian, 1997). Both, across-intron and across-exon assemblies were described (Berget, 1995; Reed, 1996). Assembly intermediates of human spliceosomes that have been observed include the E, A, B, B* and C complexes. In the E complex, the U1 snRNP is recruited to the 5' splice site and in the subsequently observed A complex (or pre-spliceosome) the U2 snRNP stably associates with the BPS. A stable interaction of the U4/U6.U5 tri-snRNP is first observed at the time of B complex formation. Major rearrangements in RNA-RNA and RNA-protein interactions leading to the destabilization of the U1 and U4 snRNPs give rise to the catalytically activated spliceosome (B* complex). The latter is then converted into the C complex, in which the first of the two catalytic steps of splicing has occured.

After a second step, the spliceosome dissociates and the snRNPs are thought to take part in addtional rounds of splicing. Besides the model of stepwise snRNP recruitment, a model in which a pre-formed spliceosome engages the pre-mRNA has been described in yeast as the penta-snRNP model (Stevens et al., 2002). Since the formation of an active spliceosome is achieved by the ordered stabilization and destabilization of multiple RNA-RNA, RNA-protein and protein-protein interactions, the activation has been proposed to proceed by an allosteric cascade of interactions (Brow, 2002).

Assembly of the U12-dependent spliceosome is analogous to that of the U2-dependent spliceosome, with one major exception. In contrast to the U1 and U2 snRNPs, U11 and U12 bind as a stable, preformed U11/ U12 di-snRNP complex. An EM reconstruction of this di-snRNP was recently published (Golas et al., 2005). During the first step of U12-type spliceosome formation, the 5' splice site and branch site are recognized by the U11 and U12 snRNP, respectively, in a cooperative manner, forming the prespliceosome (A complex) (Frilander and Steitz, 1999). The minor $U4_{ATAC}/U6_{ATAC}.U5$ tri-snRNP subsequently binds, and after major conformational changes, a catalytically active U12-dependent spliceosome is formed (Frilander and Steitz, 2001; Tarn and Steitz, 1996b; Yu and Steitz, 1997).

Three dimensional EM reconstructions of spliceosomes trapped at various stages of the spliceosomal cycle, like complexes A, B, BΔU1 (B form after loss of U1 snRNP) or C, are available (Behzadnia et al., 2007; Boehringer et al., 2004; Jurica et al., 2004; Ohi et al., 2007; Sander et al., 2006).

1.4   Non-snRNP splicing factors

In contrast to other complex RNP enzymes, such as the ribosome, the spliceosome is primarily made up of proteins. More than 200 proteins were identified, which co-purified with human spliceosomal complexes (Rappsilber et al., 2002). Among the snRNP-associated complexes are the complex splicing factors SF3a and SF3b (Kramer et al., 1999), which harbor three and seven subunits in humans, respectively, and associate with 12S U2 snRNPs (Will et al., 2001; Will et al., 2002). Additionally, a stable tetrameric complex, comprising proteins 220k, 200k 116k and

40k, could be dissociated from U5 snRNPs (Achsel et al., 1998). The majority of splicing factors are not recruited as snRNPs to the spliceosome since only some 50 proteins were shown to be tightly associated with spliceosomal U snRNPs. The non-snRNP splicing factors are pre-organized as multimeric functional modules or are recruited as single proteins to the spliceosome.

The human non-snRNP CDC5/Prp19p complex comprises six core components that concomitantly join the spliceosome just prior to catalytic activation (Ajuh et al., 2000; Chen et al., 2002; Makarova et al., 2004; Tarn et al., 1994). A similar complex has been isolated from yeast (Tarn et al., 1994). Recently, a novel ternary non-snRNP protein complex that proved to be essential for the retention of unspliced pre-mRNAs in the nucleus has been identified in *Saccharomyces cerevisiae* (Dziembowski et al., 2004). Since the complex additionally enhanced splicing of a subset of pre-mRNAs, it was termed pre-mRNA retention and splicing complex, RES (Dziembowski et al., 2004). As work on the RES complex is part of this thesis, this complex will be introduced in some detail here.

The yeast RES complex consists of three proteins, Snu17p, Bud13p and Pml1p (Dziembowski et al., 2004). Component analyses of isolated human A and B spliceosomal complexes showed that putative human orthologs of Snu17p (CGI-79), Pml1p (SNIP1) and Bud13p (MGC13125) are present in purified B but not in A complexes, in line with the overall dramatic compositional differences of these two spliceosomal states (Behzadnia et al., 2007; Deckert et al., 2006). Consistent with its presence in B complex spliceosomes, it has been shown that yeast RES elicits effects on splicing prior to the first transesterification step (Dziembowski et al., 2004; Gottschalk et al., 2001). Knockout of Snu17p results in a dramatic reduction of the first step of splicing and leads to abnormally assembled spliceosomes, in which U1 and U4 snRNPs are stabilized. Thus, it is suggested, that loss of Snu17p inhibits the progression of spliceosome assembly prior to U1 snRNP release and after U4/U6.U5 tri-snRNP addition (Gottschalk et al., 2001). Additionally, the putative human orthologs of RES subunits are detected in C complex preparations (Bessonov et al., 2008) and MGC13125, the human Bud13p ortholog, is even detectable in core snRNPs, derived from high-salt treated C complexes (Bessonov et al., 2008).

A class of non-snRNP splicing factors, which are generally not assembled in higher order complexes, are DEXD/H-box ATPases/RNA helicases (with the exception of U5- or tri-snRNP-associated proteins). These helicases control the major compositional and structural transitions during spliceosome assembly and catalysis (Staley and Guthrie, 1998). Among these proteins are U5-220k, hPrp2, hPrp5, hPrp16, hPrp22, hPrp28, hPrp43 and UAP56. These RNA unwindases/RNPases function as motors to unwind short RNA duplexes or to displace protein factors (Bowers et al., 2006; Fuller-Pace, 1994; Staley and Guthrie, 1998; Will and Luhrmann, 2001a; Will et al., 2002). Some DEXD/H-box ATPases were shown to act as kinetic proofreading devices during spliceosomal quality control (Mayas et al., 2006; Valadkhan, 2007).

The superfamily of arginine/serine-rich (RS) domain-containing splicing factors is among the best-characterized non-snRNP factors (Fu, 1995). These proteins, the SR proteins, are required for constitutive and alternative splicing. Members of the SR family have a modular structure, usually consisting of an RNA recognition motif (RRM) and a C-terminal RS domain, rich in alternating serine and arginine residues. The RS domain functions in mediating protein-protein and protein-RNA interactions and as a nuclear localization signal, whereas the RRM determines RNA-binding specificity (Caceres et al., 1997; Lai et al., 2000). Members of the SR-like protein family, e.g. U1-70k, hPrp28, tri-snRNP proteins 110k, 65k and 27k, U2AF35 and U2AF65, have different domain structures as SR proteins and may or may not contain RRMs (Fetzer et al., 1997; Makarova et al., 2001; Teigelkamp et al., 1997; Woppmann et al., 1990; Zamore et al., 1992; Zhang et al., 1992).

1.5   Phosphorylation and splicing

Recently, several studies have underscored the crucial role of reversible phosphorylation for pre-mRNA splicing (Hastings and Krainer, 2001; Mathew et al., 2008; Misteli, 1999; Shi et al., 2006). Phosphorylation of splicing factor SF1 at Ser20 inhibits its interaction with U2AF65, blocking assembly of pre-spliceosomes (Wang et al., 1999). In contrast, phosphorylation of SF1 at Ser-Pro motifs C-terminal of the U2AF65-interacting region enhances its interaction with U2AF65 (Manceau et

al., 2006). Reversible phosphorylation is also known to control the activities of a family of SR proteins as splicing regulatory proteins (Manley and Tacke, 1996). Changes in physico-chemical properties upon introduction or removal of phosphate groups modulate the interactions of SR proteins with other spliceosomal constituents (Xiao and Manley, 1998). Apart from such regulation of spliceosome assembly, phosphorylation has a direct impact on splicing catalysis. For example, phosphorylation marks appear in the U2 snRNP-associated protein SF3b155 concomitantly with splicing catalysis and phosphorylated SF3b155 is detected only in functional spliceosomes (Wang et al., 1998).

1.5.1   The FHA domain as a modular phosphorylation sensor

Since the crystal structure of the Forkhead-associated (FHA) domain of Pmlp is presented in this study, FHA domains will be described in more detail here. These domains represent one type of phosphorylation sensors (Durocher et al., 2000; Yaffe and Cantley, 1999). FHA domains were described to comprise approximately 55-75 amino acids with three highly conserved blocks of residues, which are separated by more divergent spacer regions (Hofmann and Bucher, 1995). Several solution and crystal structures have shown, however, that the minimal autonomously folded unit is significantly larger and spans more than 90 residues (Durocher et al., 2000; Li et al., 2002; Liao et al., 1999; Westerholm-Parvinen et al., 2000). FHA domains are composed of an 11-stranded β-sandwich, which contains small helical insertions between the loops connecting the strands. Phospho-peptides are bound by loop regions on one side of the FHA-domain in an extended conformation, highly reminiscent of the binding of antigen epitopes on the complementarity-determining regions of antibodies (Durocher and Jackson, 2002; Durocher et al., 2000). As a prominent example, the FHA domain of mammalian polynucleotide kinase (PNK) targets the enzyme to sites of base excision repair and non-homologous end joining repair by interacting with phosphorylated XRCC1 and XRCC4, which are central components of the respective pathways (Caldecott, 2003; Koch et al., 2004; Loizou et al., 2004).

FHA domains are also found in proteins associated with pre-mRNA splicing. For example, NIPP1 harbors an FHA domain and interacts with CDC5L and SF3b155 in a phosphorylation-dependent manner (Boudrez et al., 2000; Boudrez et al., 2002). Mapping studies have revealed that the FHA domain of NIPP1 is required for the sub-nuclear localization of NIPP1 to spliceosomal compartments (Jagiello et al., 2000), suggesting that the NIPP1 FHA domain might have a targeting function. Dominant negative mutants of NIPP1 have been characterized that block spliceosome assembly at the step of catalytic activation (Beullens and Bollen, 2002).

1.6   The RNA recognition motif (RRM)

One of the most widespread interaction devices in spliceosomal proteins is the RNA recognition motif (RRM). Classical RRMs are 80-90 residue domains, comprising a four-stranded antiparallel β-sheet backed by two α-helices (helices A and B) on one side (Antson, 2000). They serve as RNA binding elements, often employing two conserved sequence motifs, RNP1 and RNP2, located on the central two β-strands to engage in RNA contacts (Maris et al., 2005).

The proteins U1-A and U2-B" comprise archaetypical RRMs, which were shown to bind to hairpins in their respective snRNAs (Oubridge et al., 1994; Price et al., 1998). Furthermore, RRMs were shown to interact with elongated RNA molecules as seen in the crystal structure of the RRMs of sex-lethal protein bound to the polypyrimidine tract of *tra* pre-mRNA (Handa et al., 1999). The crystal structure of U2-A', U2-B" and a U2 snRNA hairpin additionally revealed that a RRM can concomitantly act as a protein binding domain (Price et al., 1998).

RRMs carry out fundamental tasks in the splicing process. A pivotal role in the recognition of the 3' boundaries of introns is played by the splicing factors U2AF and SF1/BBP (Kramer, 1992; Ruskin et al., 1988; Zamore et al., 1992). The C-terminal RNA binding domains of the large subunit of U2AF (U2AF65) recognize the polypyrimidine tract and the small subunit U2AF35 concurrently specifies the conserved AG-dinucleotide at the 3' splice site (Merendino et al., 1999; Singh et al., 1995; Wu et al., 1999; Zorio and Blumenthal, 1999). U2AF thereby facilitates SF1/BBP binding to the adjacent branch point sequence (Berglund et al., 1998).

More recently, a number of proteins have been described, in which atypical RRMs mediate protein-protein interactions via a tryptophan-containing motif. Since the first example of such an interaction was obtained from the crystal structure of a U2AF65/U2AF35 complex, these atypical RRMs were named U2AF-homology motifs (UHMs) (Kielkopf et al., 2004). The linear peptide ligands were termed UHM-ligand motifs (ULMs) (Corsini et al., 2007). Besides the U2AF65/35 dimerization, UHM/ULM interactions were observed between the large subunit of U2AF and SF1, as well as SF3b155, and between SPF45 and SF3b155 (Corsini et al., 2007; Kielkopf et al., 2001; Selenko et al., 2003; Thickman et al., 2006). Further UHM-candidates were suggested from primary sequence analyses (Kielkopf et al., 2004).

Another unconventional RRM-containing protein is Y14, which is part of the exon-junction complex. In this complex, the Y14 RRM interacts via residues of its β-sheet with a helical part of the protein mago (Fribourg et al., 2003; Lau et al., 2003). A similar interaction was found in the crystal structure of the NMD-factors UPF2 and UPF3b (Kadlec et al., 2004).

Structural analyses of a SF3b14a-SF3b155 complex uncovered a RRM-protein binding mode, in which a C-terminally expanded RRM is necessary to facilitate the binding to a short, unstructured peptide, which adopts secondary structure upon binding (Kuwasako et al., 2007; Schellenberg et al., 2006; Spadaccini et al., 2006). A compilation of atomic models of selected canonical and atypical RRMs with their respective binding partners is shown in Fig. 1.4.
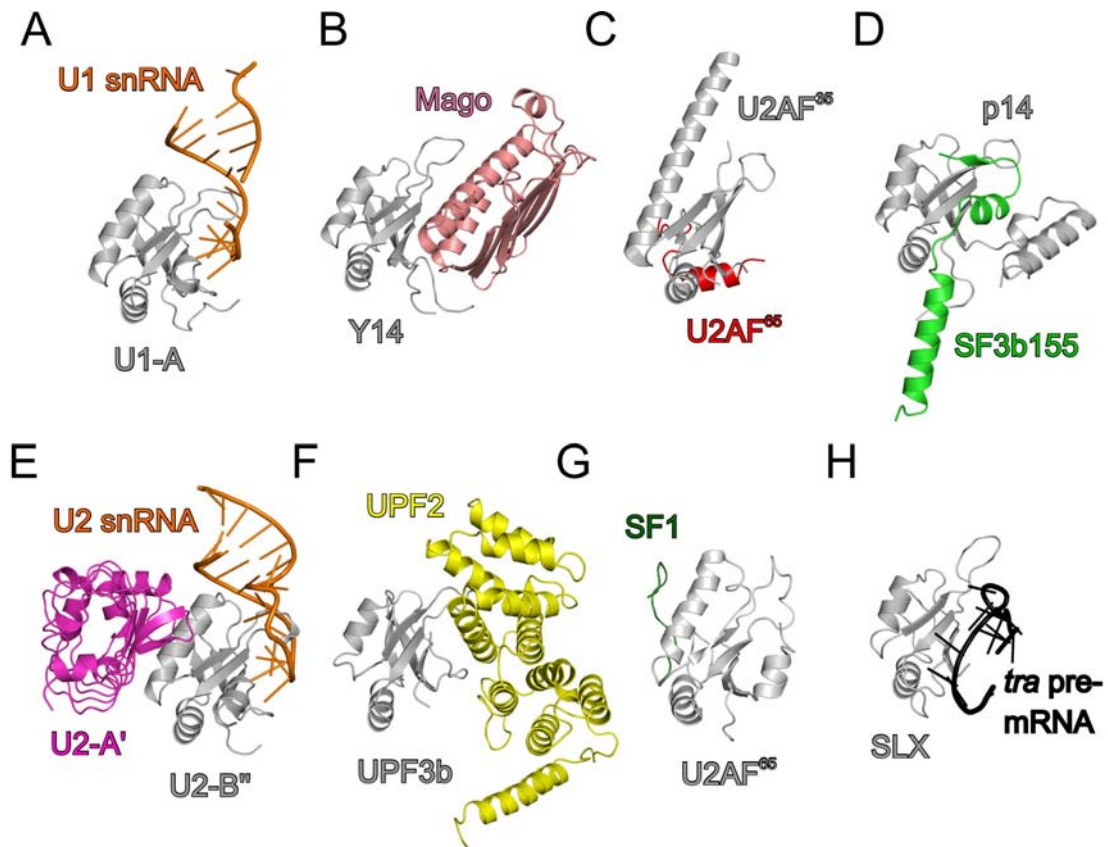
Fig. 1.4: Compilation of various RRMs and their interaction partners. Cartoon representations of **(A)** U1 snRNA (orange) bound by U1-A (grey), **(B)** Y14 (grey) in complex with mago (salmon), **(C)** U2AF35 (grey) in complex with U2AF65 (red), **(D)** SF3b14a/p14 (grey) in complex with SF3b155 (green), **(E)** U2 snRNA (orange) bound by a U2-A'/U2-B" dimer (grey and pink, respectively), **(F)** UPF3b (grey) in complex with UPF2 (yellow), **(G)** U2AF65 (grey) in complex with SF1 peptide (dark green) and **(H)** *tra* pre-mRNA (black) bound by sex-lethal protein (grey). The RRMs and UHMs are shown in grey and similar orientations (helices A and B on the left, b-sheet surface pointing to the right). Structures were adapted from Fribourg et al., 2003; Handa et al., 1999; Kadlec et al., 2004; Kielkopf et al., 2001; Oubridge et al., 1994; Price et al., 1998; Schellenberg et al., 2006; Selenko et al., 2003.

1.7   Pre-mRNA retention and splicing

Since introns often contain termination codons in frame with the upstream protein coding sequences, leakage of unspliced pre-mRNAs from the nucleus can result in the production of aberrant, truncated and potentially deleterious proteins. Several systems, which prevent pre-mRNA from being exported to the cytoplasm and being translated, have evolved.

In *Saccharomyces cerevisiae*, the nuclear envelope associated protein Mlp1 functions as a guard to retain pre-mRNAs in the nucleus (Galy et al., 2004). It has

been suggested, that Mlp1-dependent pre-mRNA retention is mediated via the 5' splice site of the pre-mRNA (Galy et al., 2004). Furthermore, SF1/BBP is known to bind the branch point sequence of intron-containing mRNAs, and defects in SF1 lead to an increase in pre-mRNA leakage (Rutz and Seraphin, 2000). Other splicing factors involved in early steps of spliceosome assembly were subsequently found to be involved in pre-mRNA retention, like *MUD*2, the homolog of U2AF65, and the cap binding complex CBC (Rain and Legrain, 1997). The RES complex, which presumably associates with the SF3b subcomplex of U2 snRNPs, has also been implicated in pre-mRNA retention in the nucleus (Dziembowski et al., 2004; Wang et al., 2005). Interestingly, a small molecule, spliceostatin A, was shown to inhibit pre-mRNA splicing and retention by interacting with subunits of SF3b, most likely through SF3b130 or SF3b155 (Kaida et al., 2007; Rymond, 2007). Another compound, pladienolide B, impaired *in vivo* splicing by binding to SF3b130 (Kotake et al., 2007). These results suggest that SF3b combines activities in pre-mRNA splicing and retention.

A role of purine-rich exonic splicing enhancers (ESEs) in nuclear retention of pre-mRNAs was recently uncovered (Taniguchi et al., 2007). ESEs usually stimulate splicing of the adjacent intron with suboptimal splice sites and are composed of purine-rich RNA sequences (Blencowe, 2000; Cartegni et al., 2002; Hastings and Krainer, 2001; Reed, 1996; Watakabe et al., 1993; Xu et al., 1993). Artificial purine-rich RNA sequences such as (GAA)*n* exert strong enhancer activity (Tacke et al., 1998; Tanaka et al., 1994). Factors like 17S U2 snRNPs and SR proteins were shown to associate with ESEs (Masuyama et al., 2007; Taniguchi et al., 2007), but do not seem to contribute to the nuclear retention activity. Thus, the identity of a nuclear retention factor associated with ESEs remains elusive (Taniguchi et al., 2007).

1.8   Objectives

The aims of this study were to structurally characterize human 12S U2 snRNPs, human splicing factor 3b (SF3b) and the yeast pre-mRNA retention and splicing (RES) complex by means of X-ray crystallography.

Existing purification protocols for nativlely isolated 12S U2 snRNPs and SF3b complexes from HeLa cells were to be improved or expanded. Special emphasis was to be put on preparating sufficient amounts of these low abundant complexes in pure form for crystallization trials. Flexible regions among 12S U2 snRNP components should be identified and removed. Different approches to process 12S U2 snRNP particles were to be established in analytical scale and to be adopted for preparative production. Nucleases and proteases should be used as biochemical "tools".

Heterologous production of RES complexes and components should be established in *E. coli* to achieve sufficient amounts for crystallization. Whenever possible, proteins or particles should be subjected to crystal screens. Further, the complexes should be characterized on a molecular scale by means of biochemical and biophysical methods. The results of this work should give insight into the architecture of the RES complex.

2.   Materials and Methods

2.1   Materials

2.1.1   Chemicals

Acetic acid	Merck, Darmstadt, D
Acrylamide solutions (ready to use):	Roth, Karlsruhe, D
Rotiphorese Gel 30 (30 % acrylamide, 0.8 % bisacrylamide)	
Rotiphorese Gel 40 (38 % acrylamide, 2 % bisacrylamide)	
Agarose (electrophoresis grade)	Invitrogen, Carlsbad, CA, USA
Ammonium peroxodisulfate (APS)	Merck, Darmstadt, D
Ammonium sulfate	Fluka, Buchs, CH
Ampicillin, sodium salt	Fluka, Buchs, CH
Borate	Merck, Darmstadt, D
Bradford reagent	Bio-Rad, München, D
Bromphenol blue, sodium salt	Merck, Darmstadt, D
Calcium chloride dihydrate	Merck, Darmstadt, D
Chloramphenicol	Roche, Mannheim, D
Complete-EDTA	Roche, Mannheim, D
Coomassie Brilliant Blue R250	Serva, Heidelberg, D
SYPRO Orange	Invitrogen, Carlsbad, CA, USA
Desoxynucleoside-5'-triphosphates (dNTPs)	New England Biolabs, Frankfurt, D
Dimethylsulfoxide (DMSO)	Roth, Karlsruhe, D
Dithiothreitol (DTT)	Roth, Karlsruhe, D
DNA molecular weight marker	GIBCO, New Zealand
Ethanol	Merck, Darmstadt, D
Ethidium bromide solution [10mg/ml]	Roth, Karlsruhe, D
Ethylenediamintetraacetate, disodium salt (EDTA)	Roth, Karlsruhe, D
Gel filtration standard	Bio-Rad, München, D
Glutathione, reduced	Sigma-Aldrich Chemie, Steinheim, D
Glycerol	Merck, Darmstadt, D
Glycogen	Roche, Mannheim, D
2-[4-(2-Hydroxyethyl)-1-piperazino]-ethansulfonic acid (HEPES)	Roth, Karlsruhe, D
Imidazole	Fluka, Buchs, CH
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Roth, Karlsruhe, D
Izit protein crystal stain	Hampton Research, Aliso Viejo, CA, USA
Kanamycin sulfate	Roth, Karlsruhe, D

LB-medium (capsules)	Q-Bio-Gene, Carlsbad, CA, USA
LB-agar-medium (capsules)	Q-Bio-Gene, Carlsbad, CA, USA
Lithium chloride	Fluka, Buchs, CH
$m_3^{2,2,7}$G cap	Amersham, Freiburg, D
Magnesium chloride hexahydrate	Fluka, Buchs, CH
2-Mercaptoethanol	Roth, Karlsruhe, D
Methanol	Merck, Darmstadt, D
2-Morpholinoethanesulfonic acid (MES)	Roth, Karlsruhe, D
NP-40 (Igepal CA 630)	Sigma-Aldrich Chemie, Steinheim, D
Orange-G	Fluka, Buchs, CH
Phenol / Chlorophorm / Isoamylalcohol (PCI); 50 % [v/v] Phenol, 48 % [v/v] Chloroform, 2 % [v/v] Isoamylalcohol, saturated in TE-Buffer	Roth, Karlsruhe, D
Phenylmethylsulfonyl fluoride (PMSF)	Roche, Mannheim, D
Polyethylene glycols 200-35000	Fluka, Buchs, CH
Polyethylene glycol 3350	Sigma-Aldrich Chemie, Steinheim, D
Potassium chloride	Merck, Darmstadt, D
Potassium hydroxide	J.T Baker, Deventer, NL
2-Propanol	Merck, Darmstadt, D
Protein molecular weight marker, unstained/prestained	Bio-Rad, München, D
Silver nitrate	Merck, Darmstadt, D
Sodium azide	Fluka, Buchs, CH
Sodium carbonate	Merck, Darmstadt, D
Sodium chloride	Merck, Darmstadt, D
Sodium dodecyl sulfate (SDS)	Serva, Heidelberg, D
Sodium hydroxide	Merck, Darmstadt, D
N,N,N',N'-Tetramethylethylenediamide (TEMED)	Sigma-Aldrich Chemie, Steinheim, D
Tris-(hydroxymethyl)aminomethane (Tris)	VWR international, Darmstadt, D
Triton-X-100	Merck, Darmstadt, D
Urea	Merck, Darmstadt, D
Xylene cyanol FF	Sigma-Aldrich Chemie, Steinheim, D

All other chemicals were purchased from Sigma, Fluka, or Merck.

2.1.2 Heavy atom derivatives

All heavy atom compounds used in this work were part of the "Heavy Atom Kit" from Hampton Research (Aliso Viejo, CA, USA) and were intended for derivatizing protein crystals in soaking experiments for structure solution by the SAD or MAD method. Derivatives were dissolved in reservoir solution right before use and stored no longer than 2-3 days at 4 °C. Heavy atom compounds used in this work are listed in Table 2.1.

Table 2.1: Heavy atom compounds used in this study.

Name	Formula
Potassium Tetrachloroplatinate (II)	K_2PtCl_4
Potassium Hexachloroplatinate (IV)	K_2PtCl_6
Potassium Tetracyanoplatinate (II)	$K_2Pt(CN)_4 \cdot xH_2O$
Gold (I) Potassium Cyanide	$KAu(CN)_2$
Potassium Tetrachloroaurate (III)	$KAuCl_4$
Mercury (II) Acetate	$Hg(OOCCH_3)_2$
Methylmercury (II) Chloride	CH_3HgCl
Mersalyl Acid	$HOHgCH_2CH(OCH_3)CH_2NHCOC_6H_4OCH_2CO_2H$
Baker's Dimercurial	$(CH_3CO)OHgCH_2CHOCH_3CHOCH_3CH_2HgO(COCH_3)$
Triethyl Lead Acetate	$(C_2H_5)_3Pb(CH_3COO)_2$
Potassium Hexachloroiridate (IV)	K_2IrCl_6
Potassium Hexachloroosmate (IV)	K_2OsCl_6
Sodium Tungstate dihydrate	Na_2WO_4
Samarium (III) Chloride hexahydrate	$SmCl_3$

2.1.3 Enzymes, proteins and peptides

Carboxypeptidase Y; dissolved in 100 mM citrate (pH 5.5)	Sigma-Aldrich Chemie, Steinheim, D
Chymotrypsin; dissolved in Milli-Q water	Roche, Grenzach-Whylen, D
cloned *Pfu* DNA polymerase (10 U/μl)	Stratagene, La Jolla, CA, USA
Elastase	Sigma-Aldrich Chemie, Steinheim, D
Endoproteinase Asp-N	Sigma-Aldrich Chemie, Steinheim, D
Endoproteinase Glu-C	Sigma-Aldrich Chemie, Steinheim, D
Endoproteinase Lys-C	Sigma-Aldrich Chemie, Steinheim, D
Goat α-rabbit antibody, horseradish peroxidase-conjugated	Jackson Immunoresearch, USA
Leucine-aminopeptidase, dissolved in 100 mM phosphate (pH 7.0)	Sigma-Aldrich Chemie, Steinheim, D
Monoclonal antibody 13E12 (α-SF3b155)	(Dybkov et al., 2006)
Monoclonal antibody H20 (α-$m_3^{2,2,7}$G cap)	(Bach et al., 1990)
Pefabloc SC	Roche, Mannheim, D
Peptide 155.1	(Will et al., 2001)
Peptide Bud13p[222-242]	Peptide Specialty Laboratories GmbH, Heidelberg, D
Peptide Bud13p[227-238]	Peptide Specialty Laboratories GmbH, Heidelberg, D
Peptide Pml1p[28-38]	Peptide Specialty Laboratories GmbH, Heidelberg, D
Precission protease1	selfmade, recombinant
Rabbit α-GST-antibody [GST (Z-5):sc-459]	Santa Cruz Biotechnology, INC. USA

Restriction enzymes   New England Biolabs, Frankfurt, D
Ribonuclease H (RNaseH), *E. coli*,   New England Biolabs, Frankfurt, D or
   recombinant   Applied Biosystems/Ambion, TX, USA
RNase A   New England Biolabs, Frankfurt, D
Subtilisin   Sigma-Aldrich Chemie, Steinheim, D
T4 DNA Ligase (400 U/µl)   New England Biolabs, Frankfurt, D
TEV-protease (1mg/ml)   selfmade, recombinant
Trypsin, dissolved in 1 M Tris-HCl   Roche, Grenzach-Whylen, D
   (pH 9.0), 100 mM $CaCl_2$

2.1.4   Oligonucleotides

Oligonucleotides used in this study were purchased from MWG (Ebersberg, Germany), Operon (Köln, Germany) or IBA (Göttingen, Germany) in lyophilized form and are listed in Table 2.2.

Table 2.2: Oligonucleotides used in this study.

Name	Sequence [5' - 3']	Description
pACYCDuetUP1	GGATCTCGACGCTCTCCCT	Sequencig
DuetUP2	TTGTACACGGCCGCATAATC	Sequencig
DuetDOWN1	GATTATGCGGCCGTGTACAA	Sequencig
DuetT7term	GCTAGTTATTGCTCAGCGG	Sequencig
pETupstream	ATGCGTCCGGCGTAGA	Sequencig
T7 promotor	TAATACGACTCACTATAGGG	Sequencig
pGEX 5'	GGGCTGGCAAGCCACGTTTGGTG	Sequencig
BUD_1_for	ATAAGACCATGGCATTGCATCAGTATTTATCAG	Cloning (*Nco*I)
BUD_34_for	ATAAGACCATGGACAAAACTTCTTTGATAGTAAAAGAACG	Cloning (*Nco*I)
BUD_65_for	ATAAGACCATGGGACAAAAAAGCAAGAATATATGGAAGAACC	Cloning (*Nco*I)
BUD_97_for	ATAAGACCATGGGAAAGAACGATCTAAAGGAAATCAGGG	Cloning (*Nco*I)
BUD_128_for	ATAAGACCATGGGTCACAAGATTCAGGAAGATTCC	Cloning (*Nco*I)
BUD_163_for	ATAAGACCATGGGAAATATGGGAGACGTGCAAAAGC	Cloning (*Nco*I)
BUD_202_for	ATAAGACCATGGAAAGAACTGTAAAAACATCTTTACTGGG	Cloning (*Nco*I)
BUD_33_rev	TATTCTTAGCGGCCGCTTATGAGTTAGCGTCTGATTTTGACTCTTTC	Cloning (*Not*I)
BUD_64_rev	TATTCTTAGCGGCCGCTTATTTGTCAAACTTGCTGAATGAAGC	Cloning (*Not*I)
BUD_96_rev	TATTCTTAGCGGCCGCTTAGCTTTCGTTGCCAGTAATTGACG	Cloning (*Not*I)
BUD_127_rev	TATTCTTAGCGGCCGCTTATTGAGCGTCTCTGTATATGGTTTTTC	Cloning (*Not*I)
BUD_162_rev	TATTCTTAGCGGCCGCTTACAAATTTTTCAGGTATTGCTCTCTTTC	Cloning (*Not*I)
BUD_201_rev	TATTCTTAGCGGCCGCTTATTTGTCATGTGTAAATGTTATTGCAGG	Cloning (*Not*I)
BUD_266_rev	TATTCTTAGCGGCCGCTCAATAATCCTCCTGTAGGGTGTAGC	Cloning (*Not*I)
PML_1_for	CATGCCGTCTCCCATGTTTCACAGACGCAAAAGACC	Cloning (*Bsm*BI[*Nco*I])
PML_34_for	ATAAGACCATGGGACCTTCCGGTCGTTAGAGCTGG	Cloning (*Nco*I)
PML_62_for	ATAAGACCATG GATAACTACATGGATATGCTAGGACTGG	Cloning (*Nco*I)
PML_90_for	ATAAGACCATGGGAAAAGGACCGTGGAAAAGATACG	Cloning (*Nco*I)
PML_113_for	ATAAGACCATGGGAAGTCTGGATACAGATCTCGACGATAG	Cloning (*Nco*I)
PML_137_for	ATAAGACCATGGGATCAAAGCAGCACTGCGTCATC	Cloning (*Nco*I)
PML_163_for	ATAAGACCATGGGATCCAATGGGACATGTTTGAACAAC	Cloning (*Nco*I)
PML_33_rev	TATTCTTAGCGGCCGCTTAGCTGAAGTCTGGCATAATATCGATG	Cloning (*Not*I)
PML_61_rev	TATTCTTAGCGGCCGCTTATGGAGAAATAGCGTCCTGCG	Cloning (*Not*I)
PML_89_rev	TATTCTTAGCGGCCGCTTAATCCTTGTCATTCTTTCTGTATATCACTAG	Cloning (*Not*I)
PML_112_rev	TATTCTTAGCGGCCGCTTAGTGGCCCAGTTCACGACCG	Cloning (*Not*I)
PML_136_rev	TATTCTTAGCGGCCGCTTAGGATGTTTCTTCGGGTATACCAATG	Cloning (*Not*I)
PML_161_rev	TATTCTTAGCGGCCGCTTACGAATCCAAATCCATTACATAACATTTC	Cloning (*Not*I)
PML_204_rev	TATTCTTAGCGGCCGCTTATACATTCATGAAGATGAGTTCGTAATC	Cloning (*Not*I)
SNU_1_for	CATGCCGTCTCCCATGAACAAAATTCAGCAAATCAACG	Cloning (*Bsm*BI[*Nco*I])
SNU_9_for	ATAAGACCATGGATAAAGAACTACAAAGTGGAATTCTTAGC	Cloning (*Nco*I)
SNU_25_for	ATAAGACCATGGGAAATGAATATAAGGATAATGCTTATATATATCGG	Cloning (*Nco*I)
SNU_106_rev	TATTCTTAGCGGCCGCTTAATGATCAATTTTGAGGGCTCTACC	Cloning (*Not*I)

Table 2.2 continued ...

SNU_125_rev	TATTCTTAGCGGCCGCTTATTCTTCCTTTACAGCTTCATAGTATTTTTG	Cloning (*Not*I)
SNU_138_rev	TATTCTTAGCGGCCGCTTATTTCTCAGCATTGTTCTTCGAAAC	Cloning (*Not*I)
SNU_148_rev	TATTCTTAGCGGCCGCTTAATTTGGTTGGTCCTTTTTAGCC	Cloning (*Not*I)
BUD_GST_for	ATAGGATCCATGGCATTGCATCAGTATTTATCAG	Cloning (*Bam*HI)
PML_GST_for	ATAGGATCCATGTTTCACAGACGCAAAAGACC	Cloning (*Bam*HI)
SNU_GST_for	ATAGGATCCATGAACAAAATTCAGCAAATCAACG	Cloning (*Bam*HI)
SNU_A_for	CATGCCGTCTCCCATGAACAAAATTCAGCAAATCAACGATAAAGAAC	Cloning (*Bsm*BI[*Nco*I])
SNU_C_for	CGCGGATCCGATGAACAAAATTCAGCAAATC	Cloning (*Bam*HI)
SNU_B_for	CGCGGATCCGAGCGAGAATCTTTATTTTCAGGGCATGAACAAAATTCA-GCAAATCAACGATAAAGAAC	Cloning (*Bam*HI) + TEV
SNU_A_rev	TATTCTTAGCGGCCGCTTAATTTGGTTGGTCCTTTTTAGCC	Cloning (*Not*I)
PML_A_for	CGCGGATCCGAGCGAGAATCTTTATTTTCAGGGCATGTTTCACAGACG-CAAAAGACC	Cloning (*Bam*HI) + TEV
PML_B_for	CATGCCGTCTCCCATGTTTCACAGACGCAAAAGAC	Cloning (*Bsm*BI[*Nco*I])
PML_A_rev	TATTCTTAGCGGCCGCTTATACATTCATGAAGATGAGTTC	Cloning (*Not*I)
Pml1p$^{51-204}$for	ATAAGACCATGGGAAAACATGTGGAGCCGCAG	Cloning (*Nco*I)
Pml1p$^{51-204}$rev	TATTCTTAGCGGCCGCTTATACATTCATGAAGATGAGTTCGTAATC	Cloning (*Not*I)
Bud13p$^{125-266}$for	GGAATTCCATATGGACGCTCAAGGTCACAAGATTCAGG	Cloning (*Nde*I)
Bud13p$^{125-266}$rev	GCCGGTACCTCAATAATCCTCCTGTAGGGTGTAGCTTTG	Cloning (*Kpn*I)
BUD_A_for	CATGCCATGGCATTGCATCAGTATTTATC	Cloning (*Nco*I)
BUD_A_rev	ATAAGAATGCGGCCGCTCAATAATCCTCCTGTAGGG	Cloning (*Not*I)
Bud_mut_for	CCATTATGCCTGGGTCAAGACGGACGGTGTCCACAGATCAAATG	Mutagenesis (W232A)
Bud_mut_rev	CCATTTGATCTGTGGACACCGTCCGCTCTTGACCCAGGCATAATG	Mutagenesis (W232A)
Oligo1	CCTATTCCATCTCCCTGCTCC	RNaseH reaction
Oligo2	CATTTAATATATTGTCCTCGG	RNaseH reaction
Oligo3	TGGAGGTACTGCAATACCA	RNaseH reaction
OligoA	CATTTAAT	RNaseH reaction
OligoB	CATTTAATATATT	RNaseH reaction
OligoC	ATTTAATATATTGTCCTCGG	RNaseH reaction
DNAzyme	ATTTAATATAGGCTAGCTACAACGATGTCCTCGGA	DNAzyme

2.1.5 Plasmids

Plasmids, which were used in this study, are listed in Table 2.3. If available in the vector sequence, the identity of the multiple cloning site (MCS) is indicated. If not otherwise stated, full-length genes were cloned. Subscripts indicate the appropriate numbers of amino acids, encoded by the cloned gene fragments.

Table 2.3: Plasmids used in this study.

Name	Description	Reference
pGEX-6P-1	T7-based *E. coli* vector for overexpression of chimeric genes coding for an N-terminal GST-tag, *Precission* protease cleavage site and the protein of interest; Amp^R	Amerscham Pharmacia
pETM-11	T7-based *E. coli* vector for overexpression of chimeric genes coding for an N-terminal hexahistidine-tag, a TEV protease cleavage site and the protein of interest; Kan^R	Gunther Stier, EMBL, Heidelberg, D
pETM-10	T7-based *E. coli* vector for overexpression of chimeric genes coding for an N-terminal hexahistidine-tag and the protein of interest; Kan^R	Gunther Stier, EMBL, Heidelberg, D
pCDFDuet-1	T7-based *E. coli* vector for co-expressions of two genes; MCS1 and MCS2; Strp^R	Novagen, Darmstadt, D
pRSFDuet-1	T7-based *E. coli* vector for co-expressions of two genes; MCS1 and MCS2; Kan^R	Novagen, Darmstadt, D

Table 2.3 continued ...

pETDuet-1	T7-based *E. coli* vector for co-expressions of two genes; MCS1 and MCS2; Amp^R	Novagen, Darmstadt, D
pETM-11_Snu17p$^{25-138}$	His$_6$-TEV-site; cloned via *Nco*I/*Not*I	This study
pETM-11_Snu17p$^{25-106}$	His$_6$-TEV-site; cloned via *Nco*I/*Not*I	This study
pETM-11_Pml1p$^{51-204}$	His$_6$-TEV-site; cloned via *Nco*I/*Not*I	This study
pETM-10_Snu17p$^{1-148}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Snu17p$^{9-148}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Snu17p$^{25-148}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Snu17p$^{1-138}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Snu17p$^{9-138}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Snu17p$^{25-138}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Snu17p$^{1-125}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Snu17p$^{9-125}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Snu17p$^{25-125}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Snu17p$^{1-106}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Snu17p$^{9-106}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Snu17p$^{25-106}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{1-64}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{1-96}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{1-127}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{1-162}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{1-266}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{34-266}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{65-266}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{97-266}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{128-266}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{163-266}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Bud13p$^{202-266}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Pml1p$^{1-61}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Pml1p$^{1-89}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Pml1p$^{1-112}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Pml1p$^{1-136}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Pml1p$^{1-161}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Pml1p$^{1-204}$	His$_6$; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pETM-10_Pml1p$^{34-204}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Pml1p$^{62-204}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Pml1p$^{90-204}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pETM-10_Pml1p$^{113-204}$	His$_6$; cloned via *Nco*I/*Not*I	This study
pCDFDuet-1_Snu17p_TEV	His$_6$-TEV-site; MCS1; cloned via *Bam*HI/*Not*I	This study
pRSFDuet-1_Snu17p	His$_6$; MCS1; cloned via *Bam*HI/*Not*I	This study
pCDFDuet-1_Snu17p	- ; MCS1; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pCDFDuet-1_Pml1p_TEV	His$_6$-TEV-site; MCS1; cloned via *Bam*HI/*Not*I	This study
pETDuet-1_Pml1p	- ; MCS1; cloned via *Bsm*BI(*Nco*I)/*Not*I	This study
pCDFDuet-1_Bud13p	- ; MCS1; cloned via *Nco*I/*Not*I	This study
pCDFDuet-1_Bud13p$^{125-266}$	- ; MCS2; cloned via *Nde*I/*Kpn*I	This study
pGEX-6P-1_Snu17p	GST-*Precission*-site; cloned via *Bam*HI/*Not*I	This study
pGEX-6P-1_Pml1p	GST-*Precission*-site; cloned via *Bam*HI/*Not*I	This study
pGEX-6P-1_Bud13p	GST-*Precission*-site; cloned via *Bam*HI/*Not*I	This study

2.1.6   Cell lines

HeLa S3 cells   (Computer cell culture centre, Belgium)

(human cervical cancer cells)

2.1.7   Bacterial strains

Escherichia coli strains DH5α, HB101 and XL-1 blue were used for propagation of plasmid DNA and BL21 (DE3) Rosetta2 cells were used for the heterologous production of proteins.

DH5α	[F⁻ *endA1 hsdR17* (r_k^- m_k^-) *supE44 thi-1 recA1 gyrA* (Nal^R) *relA* Δ(lacZYA-argF)U169 φ80lacZΔM15]	(Hanahan, 1983)
HB101	[*SupE44* Δ(*mcr*C-*mrr*) *recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl5 mtl-1 leu*E6 *thi*-1]	Promega, Mannheim, D
XL-1 blue	[*recA1 endA1 gyrA96 thi-1 hsdR17 SupE44 relA1 lac* [F′ *proAB lacI^{q}ZΔM15* Tn*10* (*Tet^R*)]]	Stratagene, La Jolla, USA
BL21 (DE3) Rosetta2	[F⁻, *ompT, hsdSB* ($r_B^-m_B^-$), *gal, dcm,* (DE3), pRARE2 (Cam^R)]	Novagen,Wisconsin, USA

2.1.8   Yeast strains

Genomic DNA from yeast strain YPH499α was used as template DNA for the amplification of the genes YIR005W (*SNU*17), YLR016C (*PML*1) and YGL174W (*BUD*13).

2.1.9   Commercial kits and crystallization screens

Crystallization kits: Qiagen, Hilden, D
    The Classics Suite
    The Classics II Suite
    The Classics Lite Suite
    The Cryos Suite
    The AmSO4 Suite
    The MPD Suite
    The Anions Suite
    The Cations Suite
    The pHClear Suite
    The pHClear II Suite
    The MbClass Suite
    The MbClass II Suite
    The Protein Complex Suite
    The PEGs Suite
    The PEGs II Suite
    The ComPAS Suite
    The PACT Suite
    The Nucleix Suite

The JCSG+ Suite
The SM I Suite
The SM II Suite
The SM III Suite
Crystallization kits: Hampton Research, Aliso Viejo, CA, USA
   Hampton I
   Hampton II
   Index
   Salt RX
   Additives Screen (I-III)
   Detergents Screen (I-III)
   Heavy Atom Kit
     (Au, Pt, Hg, M1, M2)
Crystallization kits: Molecular Dimensions, Suffolk, UK
   Wizard I
   Wizard II
   MD Structure Screen I&II
ECL Western lightning kit PLUS   Perkin Elmar, MA, USA
GFX PCR DNA and Gel Band   Amersham Biosciences, Freiburg, D
   Purification Kit
QIAprep spin miniprep kit   Qiagen, Hilden, D
QuickChange site-directed   Stratagene, Heidelberg, D
   mutagenesis kit

2.1.10 Chromatographic materials and consumables

Concentrators Vivaspin 0.5, 4, 15   Sartorius, Göttingen, D
Crystallization plates, 96 well, sitting   Greiner-Bio-One, Frickehausen, D
   drop, round bottom
Cryschem plates, 24 well, sitting drop   Hampton Research, Aliso Viejo, CA,
                                                 USA
Cryschem plates, 24 well, hanging drop   Hampton Research, Aliso Viejo, CA,
                                                 USA
Econo-Pac columns   Bio-Rad, München, D
Gloves, Dermatril   KCL, Fulda, D
Gloves, safeskin, PFE   Kimberly-Clark, Mainz, D
Glutathione Sepharose 4FF   Amersham Biosciences, Freiburg, D
HiTrap columns (1ml):   Amersham Biosciences, Freiburg, D
   DEAE, CM, ANX, Heparin, Q
HiTrap heparin Sepharose (5ml)   Amersham Biosciences, Freiburg, D
HiPrep desalting 16/26   Amersham Biosciences, Freiburg, D
NAP columns (5, 10, 25)   Amersham Biosciences, Freiburg, D
Needles   Henke Sass Wolf, Tuttlingen, D
Ni^{2+}-NTA agarose   Qiagen, Hilden, D
NHS-activated Sepharose FF   Amersham Biosciences, Freiburg, D
PC tube ASSY   Hitachi, Osaka, JP

PD-10 columns   Amersham Biosciences, Freiburg, D
Pipet tips   Sarstedt, Nürnbrecht, D
Polyprep columns   Bio-Rad, München, D
Polycarbonate flanged tube, 50 ml   Thermo Fisher Scientific, Waltham, USA

Protein A Sepharose CL 4B   Amersham Biosciences, Freiburg, D
Reaction tubes (0.5 ml, 1.5 ml, 2 ml) Sarstedt, Nürnbrecht, D
Reaction tubes (15 ml, 50 ml)   Greiner-Bio-One, Frickehausen, D
Sequencing cups   Perkin-Emer, Wellesley, MA, USA
Slide-A-Lyzer dialysis units (MWCO   Pierce, USA
    1000 Da; volume 0.5-3 ml)
Sterile filters (0.22 µm, 0.45 µm)   Sarstedt, Nürnbrecht, D
Superdex200 26/60   Amersham Biosciences, Freiburg, D
Superdex200 10/300   Amersham Biosciences, Freiburg, D
Superdex200 PC 3.2   Amersham Biosciences, Freiburg, D
Superdex75 26/60   Amersham Biosciences, Freiburg, D
Superdex75 10/300   Amersham Biosciences, Freiburg, D
Superdex75 PC 3.2   Amersham Biosciences, Freiburg, D
Superose 6 10/300   Amersham Biosciences, Freiburg, D
Superose 6 PC 3.2   Amersham Biosciences, Freiburg, D
Syringes   Braun, Melsungen, D
Tube, PA, 36 ml   Beckmann, Krefeld, D
Tube, PA, 11x66 ml   Thermo Fisher Scientific, Waltham, USA

Whatman 3MM   Whatman plc, Maidstone, UK
Zeba-spin columns   Pierce, USA

2.1.11 Commonly used buffers and solutions

Media, buffers, and solutions were prepared with deionized water (Millipore) and autoclaved if necessary (121 °C, 20 min, 1 bar). Solutions with heat-labile components were filter-sterilized (0.22 µm). The pH was adjusted by the addition of NaOH or HCl if not otherwise stated.

2.1.12 Equipment

ÄKTA Explorer   Amersham Biosciences, Freiburg, D
ÄKTA Purifier   Amersham Biosciences, Freiburg, D
ÄKTA Prime   Amersham Biosciences, Freiburg, D
Beamline PX2   Swiss Light Source, SLS, Villigen, CH
Beamline BW6   DESY, Hamburg, D
Beamline 14-2   BESSY, Berlin, B

Biobench bioreactor	Applikon, CA, USA
Biofuge fresco / pico	Heraeus, Hanau, D
Cartesian crystallization robot, 4 channels	Zinsser analytic, Frankfurt, D
Chirascan Circular Dichroism Spectrometer	Applied Photophysics, UK
Cryofuge 6000i	Heraeus, Hanau, D
DNA Engine Opticon^{TM} System	MJ Research Inc, MA, USA
FR591 rotating anode generator	Bruker ASX B.V., Delft, NL
Gel documentation system	Bio-Rad, München, D
Geldryer model 583	Bio-Rad, München, D
Gel electrophoresis equipment	Bio-Rad, München, D
Glass ware	Schott, Mainz, D, Merck, Darmstadt, D
Gradient master	Biocomp Fredericton, CA
Heating blocks	Eppendorf, Hamburg, D
FluoroLog$^®$-3 spectrofluorometer	HORIBA Jobin Yvon, NJ, USA
HT multitron culture shaker	Infors, Bottmingen, CH
Ice machine	Ziegra, Stockport, UK
Milli-Q synthesis A10	Millipore, Bedford, MA, USA
Magnetic stirrer, heatable	IKA, Staufen, D
Mar345 image plate detector	Marresearch GmbH, Norderstedt, D
Megafuge 1.0R, swing out rotor type 2704	Heraeus, Hanau, D
Microliter syringes	Hamilton, Bonaduz, CH
12-way multichannel pipette	Eppendorf, Hamburg, D
PCR cycler T-Professional	Biometra, Göttingen, D
Pharmacia LKB Frac 200	Amerscham Biosciences, Freibug, D
pH-meter seven easy	Mettler Toledo, Giessen, D
Photometer DU 530	Beckmann, Krefeld, D
Pipettes Pipetman P1000, P200, P20, P10, P2	Gilson, Middleton, WI, USA
Power supplies	Bio-Rad, München, D
Reflex, mass spectrometer	Bruker ASX B.V., Delft, NL
Rocking platform	Biometra, Göttingen, D
Scales BP4100, BP410S	Sartorius, Göttingen, D
SMART, HPLC	Amersham Biosciences, Freiburg, D
Sonifier II 250 Digi	Heinemann, Schwäbisch-Gmünd, D
Sorvall Discovery 90	Thermo Fisher Scientific, Waltham, USA
Sorvall Discovery M150	Thermo Fisher Scientific, Waltham, USA
Sorvall Evolution RC	Thermo Fisher Scientific, Waltham, USA
Sorvall SA 600 rotor	Thermo Fisher Scientific, Waltham, USA
Sorvall SLC6000 rotor	Thermo Fisher Scientific, Waltham, USA
Sorvall T-865 rotor	Thermo Fisher Scientific, Waltham, USA
Spectrophotometer Ultrospec 3000 pro	Amersham Biosciences, Freiburg, D
Speed vac concentrator 5301	Eppendorf, Hamburg, D
Surespin 630 rotor	Thermo Fisher Scientific, Waltham, USA

Spinocan needle	Braun, Melsungen, D
S100AT rotor	Thermo Fisher Scientific, Waltham, USA
Thermomixer comfort	Eppendorf, Hamburg, D
Vortex Genie 2	Scientific Industries, Bohemia, USA
Varioklav 300 EPZ	H+P Labortechnik, Oberschleißheim, D

2.1.13 Software and Programs

Advanced Image Data Analyzer software	http://www.raytest.de/
CNS	(Brunger et al., 1998)
Collaborative Computational Project Number4 (CCP4i) program suite	(Potterton et al., 2003)
Coot	(Emsley and Cowtan, 2004)
Corel Draw	http://www.corel.com/
GraphPad Prism	http://www.graphpad.com/prism/
HKL2000	(Otwinowski and Minor, 1997)
HKL2MAP	(Pape and Schneider, 2004)
Kaleidagraph	http://www.synergy.com/
Mapman	(Kleywegt and Jones, 1996b)
Moleman (2)	(Kleywegt and Jones, 1996a)
Phenix suite	(Adams et al., 2004)
Pymol	http://pymol.sourceforge.net/
Shelx suite	(Sheldrick, 2008)
XDS	(Kabsch, 1988)

2.2   Methods

2.3   Nucleic acid biochemistry

2.3.1   Concentration determination of nucleic acids

To determine the concentration of nucleic acids, the extinction in an aqueous solution was measured at a wavelength of 260 nm in comparison to a reference. The following equations were used to determine concentrations (Sambrook et al., 1989):

1 OD_{260} = 50µg/ml doubled stranded DNA = 0.15 mM (in nucleotides)

1 OD_{260} = 33µg/ml single stranded DNA = 0.10 mM (in nucleotides)

1 OD_{260} = 40µg/ml single stranded RNA = 0.11 mM (in nucleotides)

2.3.2   Phenol-Chloroform-Isoamylalcohol (PCI) extraction

To purify and separate proteins and nucleic acids, the phenol-chloroform-isoamylalcolol (PCI) extraction was used. Equal volumes of sample were mixed with PCI and were vortexed for 2-5 min. Sample volumes smaller than 100 µl were supplemented with ddH_2O to at least 100 µl before addition of PCI. Centrifugation in a tabletop microfuge at 13000 rpm (RT) separated aqueous and organic phases. The upper phase was transferred into a new reaction tube and nucleic acids were precipitated by the addition of 2.5 volumes of 100 % ethanol and 0.1 volume of 3 M sodium acetate (pH 5.2). At low sample concentrations, glycogen (15 µg) was added as a carrier. The organic phase was mixed with 5 volumes of 100 % acteone and both samples were either stored at -20°C over night, for 5 h at -80°C or for 10 min in liquid nitrogen. The precipitate was subsequently collected by centrifugation (13000 rpm, 4°C, 20 min). If necessary, pellets were further washed with 80 % [v/v] ethanol after removal of the supernatant. After centrifugation (13000 rpm, 4°C, 10 min) samples were dried in a speed vac.

2.3.3 Denaturing polyacrylamide gel electrophoresis (Urea-PAGE)

U2 snRNAs and RNA fragments were usually resolved by denaturing Urea-PAGE. For this purpose 10 or 14 % polyacrylamide gels (37.5:1 acrylamide to bis-acrylamide ratio) containing 7-8 M urea were used. RNA samples were denatured in RNA sample buffer at 95 °C for 3 min prior to electrophoresis. Samples were separated by vertical electrophoresis in 0.5 x TBE running buffer. A 50 ml 10 % polyacrylamid gel solution was generated by mixing 16.7 ml Rotiphorese Gel 30 % solution, 21 g urea, 5 ml 10 x TBE (add ddH2O to 50 ml) and the addition of 300 µl 10 % APS and 30 µl TEMED. RNA sample buffer contained 0.5 x TBE, 90 % formamide, 0.05 % [w/v] bromphenol blue, 0.05 % [w/v] xylene cyanol FF and 1 mM EDTA (pH 8.0). RNAs were visualized by silver stain (Merril and Goldman, 1982) or ethidium bromide.

2.3.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used for the analysis and purification of DNA molecules e.g. after restriction digestions and PCR reactions. Depending on the size of the molecules to be separated, gels containing 0.8 - 2 % agarose and 1 x TBE buffer were poured. 0.4 µg/ml ethidium bromide were added prior to solidification of the gel for visualization of the bands under UV light. Samples were supplemented with DNA loading dye (5x DNA loading dye: 30 % [v/v] glycerol, 0.25 % [w/v] bromophenol blue, 0.25 % [w/v] xylene cyanol FF or 50 % [w/v] saccharose, 1 mg/ml bromophenol blue, 2 mg/ml xylene cyanol FF, 2 mg/ml Orange G) and were separated using a horizontal electrophoresis.

2.3.5 DNA extraction from agarose gels

DNA bands of interest were excised with a sterile razor blade from ethidium bromide-stained gels illuminated with UV light at 365 nm. DNA exraction from gel pieces was performed using the GFX PCR DNA and Gel Band Purification Kit according to the manufacturers' instructions.

2.3.6   Polymerase chain reaction (PCR)

PCR was used for the amplification of genes or gene fragments as well as for site directed mutagenesis. To circumvent the appearance of mutations, *Pfu* polymerase was used for all reactions. A standard PCR approach contained 5 µl 10 x *Pfu* buffer, 2 µl dNTP mix (10 mM of each nucleotide-5'-triphosphate), 5-50 ng template DNA, 100 pmol sense and antisense primer, and 1-2 µl cloned *Pfu* polymerase (2.5 U/µl). A temperature protocol included an initial denaturation step (step 1; 95 °C, 2 min), another denaturation step (step 2; 95 °C, 30 sec), a primer annealing step (step 3; 2-5 °C below the melting temperature of the complementary primer sequences, 30 sec), an elongation step (step 4; 72 °C, 1-2 min per kilobase product), and a final step (step 5; 72 °C, 5 min). Steps 2-4 were repeated 25 times.

2.3.7   Site directed mutagenesis

To substitute Trp232 by an alanine in Bud13p$^{202-266}$ fragments, a mutation in the respective codon was introduced using the QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturers' instructions. The resulting clones were verified by DNA sequencing.

2.3.8   Restriction digestion of DNA

Restriction digestion was used to generate desired ends in vectors and PCR products for subsequent ligation reactions and for analytical test cleavages of plasmids. All cleavage reactions were performed according to the enzyme suppliers' suggestions. Generally, 1 U of restriction enzyme was used to cleave 1 µg of DNA in 1 h at the appropriate temperature.

2.3.9   DNA ligation

DNA fragments containing 5'-phosphates and 3'-hydroxyl groups were ligated using T4 DNA ligase. In a cloning procedure, vector and PCR product (insert) were preferentially cut with the same pair of enzymes and 100 ng of cleaved vector were mixed with a 3-fold molar excess of insert in a 20 µl reaction volume, containing 2

μl of 10x T4 DNA ligase buffer and 1 μl T4 DNA ligase. The reaction was incubated for at least 4 h at 16 °C prior to transformation of chemically competent cells.

2.3.10 Isolation of chromosomal DNA from *Saccharomyces cerevisiae*

DNA was extracted from yeast cells, which were grown in 5 ml culture volume with YPD medium (1% yeast extract, 2 % peptone, 2 % glycose). The method was generally adopted from (Sherman et al., 1986). Briefly, cells were grown at 30°C for 24 h, harvested by centrifugation and the supernatant was discarded. Cells were resuspended in 0.5 ml 1 M sorbitol solution and 25 μl lyticase solution were added prior to incubation for 30 min at 37°C. After pelleting, spheroplasts were resuspended in 0.5 ml buffer (50 mM Tris-HCl (pH 7.5), 50 mM EDTA), 25 μl 20 % [w/v] SDS-solution were added, mixed by inverting the tube, and incubated at 65°C for 20 min. 400 μl 5 M poassium acetate solution were added, the tube was inverted for mixing and set on ice for 30 min. After centrifugation in a microfuge, the supernatant was transferred to a fresh tube and isopropanol was added in a 1:1 volume ratio. The sample was briefly centrifuged, the supernatant discarded and the pellet was dried. 300 μl TE buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA) plus 50 μl RNAse A [1 mg/ml] solution were added, incubated at 37°C for 45 min and the genomic DNA was again precipitated by isopropanol and finally resuspended in 125 μl TE buffer.

2.3.11 Preparation of plasmid DNA

Plasmid DNA was extracted from bacteria cells using the QIAprep spin miniprep kit (Qiagen), according to the manufacturers' instructions. Plasmid DNA was eluted in ddH$_2$O and stored at -20°C.

2.3.12 General cloning strategies

For cloning of the genes YIR005w, YGL174w and YLR016c, coding for Snu17p, Bud13p, and Pml1p, respectively, chromosomal DNA was isolated from the yeast *Saccharomyces cerevisiae* by standard techniques. Coding sequences were amplified by PCR and cloned into pGEX-6P-1 (GE Healthcare) via *Bam*HI/*Not*I restriction

sites. Other constructs encoding portions of Snu17p, Bud13p and Pml1p were amplified from the respective full-length genes, cloned by similar strategies and also verified by DNA sequencing.

For GST pull-down and ITC experiments, gene fragments coding for Snu17p$^{1-148}$, Snu17p$^{9-148}$, Snu17p$^{25-148}$, Snu17p$^{1-138}$, Snu17p$^{9-138}$, Snu17p$^{25-138}$, Snu17p$^{1-125}$, Snu17p$^{9-125}$, Snu17p$^{25-125}$, Snu17p$^{1-106}$, Snu17p$^{9-106}$, Snu17p$^{25-106}$, Bud13p$^{1-64}$, Bud13p$^{1-96}$, Bud13p$^{1-127}$, Bud13p$^{1-162}$, Bud13p$^{1-266}$, Bud13p$^{34-266}$, Bud13p$^{65-266}$, Bud13p$^{97-266}$, Bud13p$^{128-266}$, Bud13p$^{163-266}$, Bud13p$^{202-266}$, Pml1p$^{1-61}$, Pml1p$^{1-89}$, Pml1p$^{1-112}$, Pml1p$^{1-136}$, Pml1p$^{1-161}$, Pml1p$^{1-204}$, Pml1p$^{34-204}$, Pml1p$^{62-204}$, Pml1p$^{90-204}$, and Pml1p$^{113-204}$ were cloned into the pETM-10 vector (EMBL Heidelberg, Germany) via *Nco*I/*Not*I restriction sites, allowing the production of His$_6$-tagged fusion proteins.

For purification of proteins suitable for crystallization, the gene coding for full-length Pml1p was cloned into the pCDFDuet-1 vector (Novagen) via *Bam*HI/*Not*I restriction sites. A TEV-site was engineered between the His$_6$-tag and Pml1p via the primers. The gene-fragment coding for the Pml1p$^{51-204}$ was cloned into the pETM-11 vector (EMBL Heidelberg, Germany) via *Nco*I/*Not*I restriction sites.

For co-expression experiments, full-length genes coding for Snu17p, Bud13p and Pml1p were cloned into co-expression vectors (Novagen) pRSFDuet-1, pCDFDuet-1 and pETDuet-1 via *Bam*HI/*Not*I, *Nco*I/*Not*I and *Nco*I/*Not*I restriction sites, respectively. Snu17p will have a N-terminal His$_6$-tag. The *SNU17* and *PML1* genes were additionally clone into pCDFDuet-1 vectors via restriction sites *Nco*I/*Not*I and *Bam*HI/*Not*I, respectively, allowing the expression of untagged Snu17p and Pml1p with cleavable His$_6$-tag (engineered via the primers).

The Bud13p$^{125-266}$ coding sequence was cloned via *Nde*I/*Kpn*I restriction sites into pCDFDuet-1 vector.

2.3.13 DNA sequencing

All plasmid constructs generated during this work were analyzed for the presence of the correct insert by analytical restriction cleavage and checked for the correct

sequence of the respective inserts by DNA sequencing. For a sequencing reaction, 700 ng template DNA were butanol-precipitated and sequenced at SeqLab (Göttingen, Germany) with the respective sequencing primer.

2.4   Protein biochemistry

2.4.1   Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)

To analyze proteins, high-TEMED SDS-polyacrylamide gels were used. According to the protein sizes to be separated, 10-17 % (final acrylamide concentration) gels were poured and run vertically in SDS-PAGE running buffer (25 mM Tris-HCl (pH 8.8), 192 mM glycine, 0.1 % SDS). SDS-PAGE was basically performed as described elsewhere (Laemmli, 1970; Weber et al., 1972). In brief, a typical gel contained a stacking part, consisting of 1x stacking gel buffer (125 mM Tris-HCl (pH 6.8), 0.1 % [w/v] SDS) and 16 % [v/v] Rotiphorese Gel 30 solution, polymerized with 0.3 % [v/v] APS and 0.03 % [v/v] TEMED, and a separating part, consisting of 1x separating gel buffer (375 mM Tris-HCl (pH 8.8), 0.1 % [w/v] SDS) and 30-70 % [v/v] Rotiphorese Gel 30 solution, polymerized with 0.3 % [v/v] APS and 0.3 % [v/v] TEMED. Protein samples were denatured in protein loading buffer (60 mM Tris-HCl (pH 6.8), 50 mM DTT, 1 mM EDTA, 15 % glycerol, 2 % [w/v] SDS, 0.1 % [w/v] bromophenol blue) and heated to 95°C for 5 min. Electrophoresis was typically stopped when the bromophenol blue border reached the bottom of the gel. Subsequently, the gel was boiled in Coomassie blue solution (42.5 % [v/v] ethanol, 10 % [v/v] acetic acid, 0.6 % [w/v] Coomassie brilliant blue R250), incubated for 5-15 min and destained by the sequential addition of 1st destaining solution (40 % [v/v] ethanol; 10 % [v/v] acetic acid) until bands became visible. The gel was destained further with 2nd destaining solution (5 % [v/v] acetic acid). To locate RNAs in Coomassie blue stained gels, a subsequent silver stain was performed (Merril and Goldman, 1982).

2.4.2   Concentration determination of proteins

Protein concentrations in solutions were either determined by the Bradford method (Bradford, 1976) or by measuring the absorption at 280 nm. For Bradford measurements, protein samples were diluted in up to 800 µl ddH$_2$O and 200 µl of Bradford solution was added. The absorption of the sample was measured at 595 nm, corrected for a reference and determined by the use of values from protein standards. For exact protein quantifications, 10 µl of protein sample were mixed with 140 µl of denaturing buffer (50 mM Na/K-PO$_4$ (pH 7.4), 6 M guanidin hydrochloride) and measured in quartz cuvettes (1 cm path length). The absorption was corrected by subtracting of a reference and protein concentrations were calculated using the appropriate extinction coefficents.

2.4.3   Heterologous protein production in *E. coli*

For protein production, *E. coli* BL21 (DE3) Rosetta2 cells were transformed with the appropriate plasmid DNA, carrying the protein coding open reading frame (ORF), and selected for positive clones on LB-agar plates supplemented with suitable antibiotics. Antibiotics were added in all subsequent cultivation steps. Pre-cultures of 50 ml LB medium were inocculated with a single colony and incubated over night at 37 °C on a shaker. Main cultures (1-6 L of LB-medium) were inocculated with 10-20 ml pre-culture per 1 L main culture. Gene expression was induced by the addition of 1 mM IPTG. Depending on the needs of the particular protein, induction was done at OD$_{600}$ of 0.6-1.2 and protein production was carried out for 4-24 h at 30 °C, 25 °C or 16 °C. Next, cells were harvested by cenrifugation in a SLC6000 rotor for 20 min at 4000 rpm and 4 °C. Cell pellets were resuspended in 1/100 of the culture volume with lysis buffer (50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 1mM DTT) and flash-frozen in liquid nitrogen. Frozen cell suspensions were stored at -80 °C. Lysis buffer was supplemented with 10 mM imidazole (pH 7.5) when His$_6$-tagged proteins were produced. For SDS-PAGE analysis, cell samples corresponding to 50 µl of a culture with OD$_{600}$ of 0.8 were taken prior to and after induction. Cell pellets of the samples were dissolved in 100 µl protein sample buffer and 10 µl were used for gel electrophoresis. Soluble extracts from overproductions

were obtained by thawing cell suspension, addition of a dash of lysozyme, incubation on ice for 20 min and subsequent sonication (Branson sonifier, 4°C) with short pulses until a homogeneous suspension was obtained. The cell lysate was cleared by centrifugation using a SA-600 rotor, spun for 40 min at 12000 rpm and 4 °C.

Production of selenomethionine-substituted proteins was done according to a protocol of Fusinita van den Ent & Jan Loewe (MRC Laboratory of Molecular Biology, Cambridge), which was modified from (Van Duyne et al., 1993). The protocol allows expression in any *E. coli* strain and makes use of a feedback inhibition of methionine biosynthesis by the addition of amino acids to minimal medium prior to induction. This protocol was used to produce selenomethionine-substituted Pml1p from pCDFDuet-1_Pml1p_TEV vector (coding for a N-terminal His$_6$-tag followed by an engineered TEV-site).

Isotopic   labeling   of   Snu17p   was   done   as   described   at http://www.embl.de/nmr/sattler/lab/protocols/13c_15n_print.html,   with modifications. In brief, one liter medium contained 100 ml 10 x M9 medium (per liter: 60 g Na$_2$HPO$_4$, 30 g KH$_2$PO$_4$, 5g NaCl), 10 ml 100 x trace elements (per liter: 5g EDTA, 0.8 g FeCl$_3$, 0.05 g ZnCl$_2$, 0.01 g CuCl$_2$, 0.01 g CoCl$_2$, 0.01 g H$_3$BO$_3$, 1.6 g MnCl$_2$, some NiSO$_4$, some molybdic acid, brought to pH 7.0), 1 ml 1000 x vitamins (per 500 ml: 0.5 g riboflavin, 0.5 g niacinamide, 0.5 g pyridoxine monohydrate, 0.5 g thiamine), 1 mM MgSO$_4$, 0.3 mM CaCl$_2$, 1 g ^{15}NH$_4$Cl, 4 g ^{13}C-glucose and appropriate antibiotics. Pre-cultures were grown from the same minimal medium with unlabeled NH$_4$Cl and glucose. The *SNU17* gene was expressed from a pCDFDuet-1_Snu17p_TEV vector (coding for a N-terminal His$_6$-tag followed by an engineered TEV-site).

2.4.4   Chromatographic separation of proteins on HPLC systems

Preparative ion exchange and size exclusion chromatography steps were typically performed on ÄKTA prime, purifier or explorer systems according to the manufacturers' instructions. Sample volumes for size exclusion chromatography were between 1 % and 4 % of the total column volume. If not stated otherwise, all protein samples used in this work eluted within the separation range of the respective

gel filtration column and appeared to be monodispers without unspecific aggregation. All chromatographic steps were performed at 4 °C.

For analytical purposes, size exclusion chromatography was performed on a SMART HPLC/FPLC system utilizing appropriate PC 3.2 gel filtration columns. For a typical run, 50 µl of sample were loaded on the column at a flow rate of 40 µl/min. 40 µl fractions were collected during the run and protein samples were analyzed by SDS-PAGE with subsequent Coomassie blue staining of the gel.

2.4.5 Immunoaffinity purification of spliceosomal snRNPs from nuclear extract

Nuclear extracts were prepared from HeLa cells essentially as described elsewhere (Dignam et al., 1983b). All steps were carried out at 4 °C. Six to eight liters of Hela cells (2.5-5×10^5 cells/ml) were harvested by centrifugation for 10 min at 2000 rpm in a Cryofuge 6000i and washed 3 times with ice-cold 1x PBS (130 mM NaCl, 20 mM K-PO$_4$ (pH 8.0)) for 10 min each. Cells were resuspended in 2 cell pellet volumes of Roeder A buffer (20 mM Hepes-KOH (pH 7.9), 3 mM MgCl$_2$, 10 mM KCl, 0.5 mM DTE) and allowed to swell for 5 min on ice. Cells were pelleted in a Megafuge 1.0R for 10 min at 2000 rpm and resuspended in Roeder A buffer supplemented with 0.5 mM phenylmethylsulfonylfluoride (PMSF). The suspension was dounced with 15 strokes to disrupt the cells and nuclei were sedimented in a SA-600 rotor (10 min at 2700 rpm). 1.5 pellet volumes of Roeder C250 buffer (20 mM Hepes-KOH (pH 7.9), 250 mM NaCl, 1.5 mM MgCl$_2$, 0.2 mM EDTA, 0.5 mM DTE) supplemented with 0.5 mM PMSF were added to the nuclear pellet and nuclei were disrupted by douncing with 15 strokes. The solution was slowly stirred on a magnetic stirrer for 30 min at 4 °C. The homogenate was transferred to 70Ti tubes, overlaid with N-heptane and cleared by centrifugation using a Beckman 45Ti rotor (40000 rpm, 30 min, 4 °C). The supernatant was recovered with a syringe. If not used immediately, the extract was flash-frozen in liquid nitrogen and stored at -80 °C. To achieve a final glycerol concentration of 5 %, the nuclear extract was diluted with glycerol-free Roeder C250 buffer (supplemented with 0.5 mM PMSF) and cleared by centrifugation (45Ti rotor, 40000 rpm, 30 min, 4 °C) and filtration. A 50

ml Sepharose-column carrying covalently coupled H20 antibodies was pre-equilibrated in Roeder C250 buffer and the cleared extract was loaded. After washing the column with 20-30 column volumes of Roeder C250 buffer (with 0.5 mM PMSF), bound snRNPs were eluted with 2 column volumes of Roeder C250 buffer, supplemented with 0.5 mM PMSF and 5 mM of $m_3^{2,2,7}$G cap in fractions of 2 ml. Samples were analyzed via SDS-PAGE, stored at -80 °C and the column was regenerated with Roeder C250 buffer supplemented with 6M urea.

2.4.6 Glycerol gradient centrifugation of spliceosomal snRNPs

Glycerol gradient centrifugation was essentially performed as described elsewhere (Kastner and Lührmann, 1999). All steps were done at 4 °C. Six preparative glycerol gradients were prepared by layering 15 ml of 30 % glycerol buffer (20 mM Hepes-KOH (pH 7.9), 1.5 mM MgCl$_2$, 150 mM KCl, 0.5 mM DTT, 0.25 mM PMSF, 30 % [v/v] glycerol) under 15 ml of 10 % glycerol buffer (20 mM Hepes-KOH (pH 7.9), 1.5 mM MgCl$_2$, 150 mM KCl, 0.5 mM DTT, 0.25 mM PMSF, 10 % [v/v] glycerol) using a syringe and a spinocam needle. The gradient was subsequently formed by a gradient mixer. After 30 min of incubation, 6 ml buffer were carefully removed from the top of each gradient and replaced with the sampe volume of a 1 mg/ml snRNP solution derived from H20 fractions. Gradients were spun in a Surespin630 rotor for 17 h at 27000 rpm and harvested in 1.5 ml fractions from the bottom of each tube. To ensure reproducibility, a peristaltic pump and a fraction collector were used for harvesting. An input sample and samples from the 24 gradient fractions were subjected to gel-electrophoretic analysis and protein concentrations of each fraction were determined by the Bradford method. Fractions were flash-frozen in liquid nitrogen and stored at -80 °C.

2.4.7 Isolation of 12S U2 snRNPs from glycerol gradient fractions

All steps described in the following were performed at 4 °C utilizing pre-chilled buffers. Glycerol gradient fractions containing 12S particles of U1 and U2 snRNPs were pooled and diluted to a final salt concentration of 50 mM with buffer 0 (20 mM Tris-HCl (pH 7.9), 1.5 mM MgCl$_2$, 1 mM DTT). The mixture was loaded on a 5 ml

HiTrap heparin Sepharose column, equilibrated in buffer 50 (20 mM Tris-HCl (pH 7.9), 50 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT) using an ÄKTA prime system. The best separation of U1 and U2 snRNPs was achieved by applying a linear salt gradient from 50-800 mM NaCl over a range of 60-70 ml. 20 µl samples were taken from 1.5 ml peak fractions and analyzed together with an input sample by SDS-PAGE. After the run, the column was washed with buffer 1000 (20 mM Tris-HCl (pH 7.9), 1000 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT). 12S U2 snRNPs usually eluted between 200 and 250 mM NaCl, whereas U1 snRNPs eluted around 450 mM NaCl. 12S U2 snRNPs were concentrated to 12 mg/ml in buffer 10 mM Tris-HCl (pH 7.9), 250 mM NaCl, 1 mM DTT.

2.4.8 Purification of Δ5' U2 snRNPs

5 mg 12S U2 snRNPs from heparin-Sepharose column runs were pooled and OligoC was added in 10 fold molar excess. The mixture was incubated on ice for 10 min prior to 1:1 dilution with buffer comprising 20 mM Tris-HCl (pH 7.9), 5 mM DTT, 8 mM $MgCl_2$. 550 U of RNAseH [10U/µl] (Ambion) were added and the reaction was incubated for at least 3 h at 30°C until a complete cleavage has occurred. NaCl was added from a 5 M stock up to a final salt concentration of 250 mM. Cleaved U2 snRNPs were passed twice over a 5 ml H20 antibody-coupled Sepharose column, pre-equilibrated in buffer comprising 20 mM Tris-HCl (pH 7.9), 250 mM NaCl, 2 mM DTT. Flow-through and two wash steps (5 ml each) with the equilibration buffer were collected. Flow-through and wash fractions were concentrated to 600 µl using a VIVAscience ultrafiltration concentrator 15 (MWCO 30000). Two gel filtration runs were performed on an ÄKTA purifier system using a Superdex 200 10/300 column with a buffer comprising 10 mM Tris-HCl (pH 7.9), 250 mM NaCl and 1 mM DTT. Peak fractions from gel filtration runs containing cleaved U2 snRNPs were collected and concentrated to 11.1 mg/ml in VIVAscience ultrafiltration concentrators 0.5 (MWCO 30000). Finally particles were subjected to crystallization screens at 4°C (672 conditions) or flash frozen in liquid nitrogen and stored at -80°C. Samples of each purification step were collected, proteins and RNAs separated by PCI and analyzed by SDS-PAGE or Urea-PAGE.

To verify the presence of the 3'-RNA part in Δ5'-U2 snRNPs, snRNAs from 12S U2 snRNPs and Δ5' U2 snRNPs were isolated by PCI extraction. 10 pmol snRNAs, 300 pmol oligo (oligo1, oligo2 or oligo3) and 1U RNAseH (New England Biolabs) were used per reaction. Each reaction was incubated for 1 h at 37°C in a total volume of 20 µl. RNAs were again extracted by PCI, separated by Urea-PAGE and visualized with ethidium bromide.

2.4.9   Purification of Δ5' Sm B$^{\Delta C\text{-term}}$ U2 snRNPs

Δ5' U2 snRNPs were purified as described in section 2.4.8 but were mixed with 40 µg Carboxypeptidase Y prior to the final gel filtration step. Proteolysis was performed at 30 °C for at least 3 h and the products were analyzed by SDS-PAGE. In contrast to the protocol described in section 2.4.8, concentration of the particles was achieved by pelleting essentially as described for SF3b particles (SAT100 rotor, 4 h, 60000 rpm). Finally, particles with a concentration of 9 mg/ml were subjected to nano-scale crystallization screens at 4°C (672 conditions) or flash frozen in liquid nitrogen and stored at -80°C.

2.4.10   Cleavage of snRNAs in 12S U2 snRNPs by a DNAzyme

To remove the presumably flexible 5'-part of 12S U2 snRNAs under native conditions, a DNAzyme was engineered carrying the central catalytic motif of 10-23 DNAzymes (Santoro and Joyce, 1998) and flanking regions complementary to U2 snRNA region 76-96. 50 µl 12S U2 snRNPs (180 pmol) were incubated with 100x molar excess of DNAzyme molecules in 30 µl reaction volume for 1 h at various temperatures (10, 20 and 30°C) in a buffer comprising 50 mM Tris-HCl (pH 7.9), 250 mM NaCl, 1 mM DTT and 2 mM $MnCl_2$ or 10 mM $MgCl_2$. Samples were PCI extracted and RNAs were ethanol precipitated prior to Urea-PAGE. RNAs were stained with ethidium bromide.

2.4.11   Purification of SF3b particles

To obtain SF3b particles for crystallization trials, the procedure described in section 2.4.5 was followed with the exception of the final immunoaffinity step, in

which the H20 antibody-coupled column was run with Roeder C420 buffer (20 mM Hepes-KOH (pH 7.9), 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTE). About 200 ml flow-through of the H20 antibody-coupled column were collected, adjusted to a final NaCl concentration of 600 mM and passed over a 2 ml α-SF3b155 antibody-coupled column, equilibrated in G_{600} buffer (20 mM HEPES (pH 7.9), 600 mM NaCl, 1.5 mM $MgCl_2$, 5 % glycerol), with a peristaltic pump. After washing the column with 6x 15 ml G_{600}^+ buffer (20 mM HEPES (pH 7.9), 600 mM NaCl, 1.5 mM $MgCl_2$, 5 % glycerol, 0.5 mM DTE, 0.5 mM PMSF), SF3b particles were eluted by the sequential addition of 0.1 M 155.1-peptide in 18 ml G_{600}^+ buffer. 25 µl of the input and 50 µl of eluate fractions were analyzed by SDS-PAGE and Coomassie blue stain.

Peak fractions of four H20 antibody column runs were pooled and concentrated by pelleting in a mini ultracentrifuge (Sorvell Discovery M150) using thick-walled polycarbonate tubes (PC tube ASSY). Typically 2.5 ml SF3b were sedimented for 6 h at 60000 rpm. Deceleration was set to minimum to avoid disturbance of the pellet. Supernatant was removed carefully and SF3b particles were recovered from the remaining ~50-100 µl per tube. 250 µl of concentrated SF3b were passed over a Superose 6 10/300 column on an Äkta Explorer system. The running buffer comprised 10 mM Tris-HCl (pH 7.9), 150 mM NaCl and 1 mM DTT. 10 µl of peak fractions were analyzed on a 10-13 % SDS-polyacrylamide (PAA) step gel, which was subsequently stained with Coomassie blue. Fractions containing SF3b particles were pooled and again sedimented in thick-walled polycarbonate tubes until a final concentration of at least 9 mg/ml was obtained. Particles were screened in a nano-scale crystallization approach with 576 conditions (4 °C). A sample from the final pool was additionally analyzed on a 10-13 % SDS-PAA step gel. SF3b samples not subjected to crystallization experiments were flash frozen in liquid nitrogen and stored at -80°C.

2.4.12 Purification of Pml1p

Full-length Pml1p ($Pml1p^{FL}$) was expressed in *Escherichia coli* Rosetta2 (DE3) cells from the pCDFDuet-1_Pml1p_TEV expression vector. Transformed cells were

grown in LB medium supplemented with chloramphenicol/streptomycin to an OD_{600} of 0.8-1.2. After induction with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG), cells were incubated for 3-4 h at 25 °C before harvesting. The target protein was captured from soluble extract *via* Ni^{2+}-NTA resin and eluted by addition of imidazole. The tag was removed by addition of Tobacco Etch Virus (TEV) protease (mass ratio protease/protein 1:40). After buffer exchange to 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT *via* a 16/10 HighPrep column (GE Healthcare), the protein solution was loaded onto CM-Sepharose (GE Healthcare), washed and eluted in a linear gradient to the same buffer plus 600 mM NaCl. Pooled fractions were further purified by size exclusion chromatography on a Superdex 75 26/60 column (GE Healthcare) in 10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT. The purified protein was concentrated to 16 mg/ml with spin concentrators (VIVAscience). The final sample was > 98 % pure as judged by SDS-PAGE analysis. Selenomethionine-derivatized $Pml1p^{FL}$ was obtained using the same purification protocol. The purified protein was concentrated to 14 mg/ml with spin concentrators (VIVAscience). The final sample was > 98 % pure as judged by SDS-PAGE analysis. Unlike the native protocol, 2 mM DTT were used in all purification steps.

An N-terminally truncated version of Pml1p encompassing residues 51-204 ($Pml1p^{51-204}$), also bearing an N-terminal, cleavable His_6-tag, was expressed in *E. coli* Rosetta2 (DE3) cells from a pETM-11 vector. The protein was purified as described for full-length Pml1p but loaded directly on a Superdex 75 26/60 column (GE Healthcare) after TEV-cleavage and was subsequently cycled over pre-equilibrated Ni^{2+-}NTA agarose to remove residual uncleaved fusion protein. The final $Pml1p^{51-204}$ pool was concentrated to 34 mg/ml.

2.4.13 Production of RES complexes, subcomplexes and single components

All proteins and protein complexes were heterologously produced in *E. coli* as described in section 2.4.3. All chromatographic steps were done as described in section 2.4.4. Purifications were followed by SDS-PAGE analyses essentially as

described under 2.4.1. Rough estimates and precise measurements of protein concentrations were done as depicted in section 2.4.2.

Genes, coding for RES complex components Snu17p, Pml1p and Bud13p, were co-expressed, or expressed individually, from vectors pRSFDuet-1_Snu17p (His$_6$-tagged), pETDuet-1_Pml1p and pCDFDuet-1_Bud13p. Assembled RES complexes were captured on Ni^{2+}-NTA agarose, beads were washed extensively with buffer (50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 1 mM DTT, 10 mM Imidazole) and particles were eluted by the addition of 300 mM Imidazole to the same buffer. The eluat was buffer-exchanged using a HiPrep Desalting 26/10 column into 50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 1 mM DTT. The protein solution was mixed with pre-equilibrated CM-Sepharose slurry (50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 1 mM DTT) and the mixture was diluted 1:1 with buffer (50 mM Tris-HCl (pH 7.5), 1 mM DTT). CM-Sepharose beads were washed with buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT) and particles were eluted by the addition of the same buffer with 800 mM NaCl. Alternatively, particles were buffer-exchanged into 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM DTT, 1 mM MgCl$_2$, 10 μM ZnCl$_2$, applied on a CM-Separose column and eluted by a salt gradient. Particles were further purified via a Superdex 200 26/60 gel filtration column, which was run in 10 mM Tris-HCl (pH 7.5), 800 mM NaCl, 1 mM DTT. RES complexes were concentrated to ~60 mg/ml by ultrafiltration using VIVAscience concentrators (10000 MWCO).

A truncated RES complex, comprising proteins His$_6$-Snu17p, Pml1p and Bud13p$^{125-266}$, was produced from vectors pRSFDuet-1_Snu17p, pETDuet-1_Pml1p and pCDFDuet-1_Bud13p$^{125-266}$ essential as described for the full-length complex. Particles were buffer-exchanged into 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM DTT prior to loading on a CM-Sepharose column (25 ml bead volume). Particles were eluted by a linear salt gradient (50 mM Tris-HCl (pH 7.5), 50-600 mM NaCl, 1 mM DTT) and truncated RES complex-containing fractions were further chromatographed on a Superdex 75 26/60 gelfiltartion column (10 mM Tris-HCl (pH 7.5), 400 mM NaCl, 1 mM DTT). Particles were concentrated to ~18 mg/ml by ultrafiltration using VIVAscience concentrators (10000 MWCO).

Dimeric complexes, comprising His$_6$-Snu17p and Pml1p or Snu17p and Pml1p with a cleavable His$_6$-tag, were produced from vectors pRSFDuet-1_Snu17p and pETDuet-1_Pml1p or pCDFDuet-1_Snu17p and pCDFDuet-1_Pml1p_TEV, respectively. Genes were expressed individually. Buffers were supplemented with 1 mM MgCl$_2$ and 10 µM ZnCl$_2$. Particles were captured on Ni^{2+}-NTA agarose as described above and eluates were desalted via a HiPrep Desalting 26/10 column into buffer (50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2 mM DTT, 1 mM MgCl$_2$, 10 µM ZnCl$_2$). The His$_6$-tag of Pml1p was removed by overnight incubation of the Ni^{2+}NTA-eluate with TEV protease in a particle/protease mass ratio of 40:1 prior to bufferexchange. Particles were further purified on a Superdex 75 26/60 gelfiltartion column (10 mM Tris-HCl (pH 7.5), 300 mM NaCl, 1 mM DTT). Fractions of a dimer, comprising His$_6$-Snu17p and Pml1p, were concentrated to 13 mg/ml by ultrafiltration using VIVAscience concentrators (10000 MWCO). Pml1p protein, eluting as a second single peak from the gel filtration column, was concentrated to 20 mg/ml by ultrafiltration. The Snu17p-Pml1p dimer was concentrated to 24 mg/ml by ultrafiltration using VIVAscience concentrators (10000 MWCO) in buffer 10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT.

A dimeric complex, comprising His$_6$-Snu17p and Bud13p$^{125-266}$ was produced from vectors pRSFDuet-1_Snu17p and pCDFDuet-1_Bud13p$^{125-266}$. Particles were captured via Ni^{2+}-NTA agarose, bufferexchanged and chromatographed on a DEAE-Sepharose column (20 ml bead volume; (50 mM Tris-HCl (pH 7.5), 50-600 mM NaCl, 2 mM DTT, 1 mM MgCl$_2$, 10 µM ZnCl$_2$). Complex containing fractions were further purified via gel filtration and concentrated to 25 mg/ml.

Gene fragments, coding for Snu17p$^{25-106}$ and Bud13p$^{202-266}$, were expressed individually from pETM-10_Snu17p$^{25-106}$ and pETM-10_Bud13p$^{202-266}$ vectors and the His$_6$-tagged proteins were purified individually via Ni^{2+}-NTA agarose. Eluates were combined and chromatographed on a Superdex 75 26/60 gel filtration column in a buffer comprising 10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT. Particles were concentrated to 12 mg/ml by ultrafiltration (VIVAscience concentrator 5000 MWCO).

Proteins Snu17p$^{25-138}$ and Snu17p$^{25-106}$ were produced from pETM-11_ Snu17p$^{25-138}$ and pETM-11_Snu17p$^{25-106}$ vectors, respectively. The proteins were captured on Ni^{2+}-NTA agarose and His$_6$-tags of proteins in eluate fractions were removed by overnight incubation with TEV protease in a particle/protease mass ratio of 40:1. Proteins were further purified on a Superdex 75 26/60 gelfiltartion column (10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT). Mini-RES complexes were reconstituted from Snu17p$^{25-138}$ and peptides Bud13p$^{228-237}$ and Pml1p$^{28-38}$ in a 1:2 molar ratio of protein/peptides and concentrated to 34 or 64 mg/ml using ultrafiltration (VIVAscience concentrator 5000 MWCO). Masses of 13477.0 (Snu17p$^{25-138}$), 1166.6 (Bud13p$^{228-237}$) and 1176.6 (Pml1p$^{28-38}$) from the final concentrated samples were identified by mass spectrometry, indicating the presence of the polypeptides. A Snu17p$^{25-106}$/Bud13p$^{222-242}$ dimer was produced in the same way as mini-RES complexes (20 mg/ml).

Methylation of Snu17p$^{25-138}$ was done as described in Walter et al. (2006).

2.4.14 Mass spectrometry

Mass spectrometric analyses were carried out by U. Pleßmann and M. Raabe in the group of Dr. H. Urlaub at the Max-Planck-Institute for Biophysical Chemistry (Göttingen, Germany). Measurements of tryptic protein fragments from excised gel bands were done according to Shevchenko et al. (1996). The measurements allowed the identification of components of particle X (co-purified with SF3b) and of fragments of U2 snRNP proteins after limited proteolysis by their peptide mass fingerprint. MALDI MS analyses were carried out on a Bruker Reflex IV mass spectrometer (Bruker Daltonics, Bremen, D) and data were evaluated with the program Mascot.

For the identification of Snu17p-interacting peptides of Bud13p, 40 µg of purified His$_6$-Bud13p$^{125-266}$ were fragmented by the addition of Glu-C (mass protease/protein ratio of 1:20) and incubation at 37 °C in a buffer comprising 20 mM Tris-HCl (pH 7.5), 250 mM NaCl, 1 mM DTT. The reaction was followed by SDS-PAGE until a complete cleavage has occured. Glu-C activity was inhibited by the addition of Pefabloc SC (1 mM final concentration). Half of the Bud13p sample was mixed with

20 µg His$_6$-Snu17p and incubated for 2 h on ice. The protein/peptide mixture was passed over a Superdex 75 PC3.2 column, equilibrated in buffer (20 mM Tris-HCl (pH 7.5), 250 mM NaCl, 1 mM DTT), using a SMART system (see section 2.4.4). Samples of Glu-C treated Bud13p and of peak fractions derived from analytical gel filtration were acetone-precipitated prior to MS measurements. Thin-layer preparations using α-cyano-4-hydroxycinnamonic acid (CHCA) and nitrocellulose as matrix were done for MALDI measurements. For ESI experiments on a Q-ToF Ultima mass spectrometer (Micromass, GB), dried samples were dissolved in 10 % [v/v] acetonitrile and 0.15 % [w/v] formic acid.

2.4.15 Electron microscopy

Negative stain electron micrographs of RNaseH and Carboxypeptidase Y treated 12S U2 snRNPs and of particle X samples were imaged by Dr. P. Dube and F. Hauer, respectively, who are both members of the group of Dr. Holger Stark at the Max-Planck-Institute for Biophysical Chemistry (Göttingen, Germany).

Negative staining of the particles was carried out by standard double carbon film method with uranyl formate (Kastner et al., 1990). Highly concentrated U2 snRNP samples were diluted 50-fold in SEC buffer prior to adsorption to EM-grids. Particle X samples were applied on a 10-30 % glycerol GraFix gradient before imaging (Kastner et al., 2008). GraFix gradients and subsequent sample preparations for EM were performed by F. Hauer. Low dose electron micrographs were recorded on a 4k x 4k CCD camera in a Phillips CM200 FEG using an acceleration voltage of 160 kV at 88000-fold magnification.

2.4.16 Limited proteolysis

Limited proteolysis experiments were usually performed with 5-10 µg target protein or complex in 10-20 µl reaction volumes. Experiments were carried out with different protease concentrations (typically a series of 1:10 dilutions per protease) over a given time window or were carried out at a fixed protease concentration and samples were taken at different time points. The reactions were stopped by the addition of 10 µl SDS loading dye per sample and heating to 95 °C for 3 min.

2.4.17 Peptide array binding experiments

The peptide array analysis was used for the precise identification of Snu17p-interacting peptides derived from proteins Bud13p and Pml1p. Overlapping 15-, 20-, and 25-mers (1 residue offset) covering amino acids 1-60 of Pml1p and 200-266 of Bud13p were synthesized and assembled in duplicate as an array of spots on cellulose membranes (kind gift of Dr. C. Freund, FMP, Berlin, Germany). Membranes were washed for 10 min in 100 % ethanol and, subsequently, three times for 10 min with 1x SPOT-TBS to remove residual contaminants. Membranes were blocked for 3 h in blocking buffer and again washed for 10 min in 1x SPOT-TBS. GST-fused Snu17p protein, or GST alone as a control, were diluted in blocking buffer to a final concentration of 40 µg/ml. Membranes were incubated over night in protein solution at 4 °C on a shaker. Unbound proteins were removed by washing membranes three times with 1x SPOT-TBS. Membranes were incubated for 3 h at RT with primary antibody (anti-GST-antibodies, diluted 1:1000 in SPOT-blocking buffer), washed three times with 1x SPOT-TBS and incubated 1.5 h with secondary antibody (horseradish peroxidase-conjugated anti-rabbit IgG; diluted 1:25000 in SPOT-blocking buffer). The membranes were washed 5 times with 1x SPOT-TBS and subsequently treated with the Western Lightning^{TM} Chemiluminescence Reagent Plus. Chemiluminescence was detected with a CCD camera and analyzed by the AIDA (Advanced Image Data Analyzer) software.

10x SPOT-TBS:	2 g KCl, 80 g NaCl, 61 g Tris, ad 1L H_2O, adjust to pH 8.0
SPOT-blocking buffer:	2% [w/v] milk powder, 5 ml 10x SPOT-TBS, 2.5 g saccharose, filtered

2.4.18 GST pull-down assays

GST-fusions of Snu17p, Bud13p, Pml1p and GST alone were expressed in *E. coli* Rosetta2 (DE3) cells (Novagen) from the respective pGEX-6P-1 plasmids and His$_6$-tagged polypeptides of Snu17p, Bud13p and Pml1p were expressed from pETM-10 vectors. GST and the GST-fusion proteins were captured on glutathione-coated Sepharose beads (GE Healthcare) and washed extensively with buffer A (50 mM

Tris-HCl (pH 7.5), 300 mM NaCl and 1 mM DTT). His$_6$-tagged fragments were pre-purified via Ni^{2+}-NTA agarose (Qiagen) and eluted in buffer A supplemented with 300 mM imidazole. 20 μl protein-coated glutathione-beads were used per binding reaction and incubated with an excess of the respective His$_6$-tagged protein for 2 h on ice. After extensive washing with buffer (50 mM Tris-HCl (pH 7.5), 250 mM NaCl, 1mM DTT) proteins were eluted by addition of SDS sample buffer and half of the fractions were analyzed by SDS-PAGE.

2.4.19 Mutational analysis of the Snu17p-Bud13p complex

A Trp232Ala point mutation was introduced into His$_6$-tagged Bud13p$^{202\text{-}266}$ by using the QuikChange site-directed mutagenesis kit (Stratagene) and confirmed by sequencing. Wild type and mutant fragments were produced in *E. coli* and purified via Ni^{2+}-NTA resin. Three-fold molar excess of either His$_6$-tagged Bud13p$^{202\text{-}266}$ or Bud13p$^{202\text{-}266}$[Trp232Ala] were incubated with His$_6$-tagged Snu17p for 3 h at 4 °C. 50 ml of the protein mix were applied on a Superdex75 10/300 PC column and chromatographed using the SMART protein purification system (GE Healthcare) at a flow rate of 40 ml/min. 40 ml fractions were collected and analyzed by SDS-PAGE.

2.5   Cells and cell culture

2.5.1   Cultivation of *Escherichia coli*

E. coil cells were cultivated in LB medium, M9 minimal medium or on LB agar plates. To maintain adequate aeration in preparative protein production setups, cells were grown in chicane flasks. Optical densities (ODs) of cultures were measured in plastic cuvettes with 1 cm path-length in an Ultrospec 3000 pro spectral photometer at 600 nm wavelength with respective plain media as a reference.

2.5.2   Transformation of chemically competent bacteria

Chemically competent *E. coli* cells were prepared according to the CaCl$_2$-method (Chung and Miller, 1993) and stored in 50 μl aliquots at -80°C. In a standard approach, 5 μl of a ligation reaction were mixed with 50 μl of chemically competent *E. coli* cells and the cells were incubated for 30 min on ice. Bacteria were heat-

shocked for 90 sec at 42°C and subsequently cooled on ice for 2 min. 800 µl of LB medium were added and the cells were further incubated at 37°C for 30 min on a shaker. Successfully transformed cells were selected on LB-agar plates supplemented with appropriate antibiotics. Final antibiotic concentrations in media were 50 µg/ml (streptomycin), 30 µg/ml (kanamycin), 34 µg/ml (chloramphenicol) and 100 µg/ml (ampicillin).

2.5.3   Cultivation of HeLa cells

For the cultivation of HeLa S3 cells (Computer cell culture centre, Belgium), S-MEM medium (Gibco BRL), supplemented with 5 % [v/v] newborn calf serum (Gibco BRL), 50 µg/ml penicillin and 100 µg/ml streptomycin were used. Cells were grown in suspension (2.5-5 x 10^6 cells/ml medium) at logarithmic growth rate in a bioreactor. HeLa cell cultivation and harvesting was done essentially as described elsewhere (Dignam et al., 1983a; Dignam et al., 1983b; Kastner, 1998).

2.6   Crystallographic methods

2.6.1   General crystallography setup

In order to identify initial crystallisation conditions, diverse commercially available crystallization screens were performed using a Cartesian liquid dispensing robot for nanoliter-setups. This procedure allowed minimization of sample volumes and testing of a broad range of different chemical compounds. All screens used are listed under "Commercial kits and crystallization screens" (section 2.1.9). Pre-screens were performed by the sitting-drop vapor diffusion technique in a 96-well plate format with reservoir volumes of 100 µl and a drop volume of 200 nl. Conditions of initial hits were typically refined in 24-well linbro plates by small variations of pH values and precipitant or salt concentrations. Conditions were further improved by the performance of commercial additive screens (Hampton) in 96-well sitting drop plates using the Cartesian liquid dispensing robot or in 24-well linbro plates. Refinements were also done using the hanging drop vapor-diffusion technique.

Diffraction capabilities of refined crystals were tested at 100 K on a Bruker Nonius FR591 home source and datasets of native and derivatized crystals were typically collected on synchrotron beamlines. If reservoir solutions showed ice formation upon freezing, various cryo-protectants like PEG400, glycerol, polypropylene glycol, etc. were tested or crystals were transferred into oils to prevent ice formation.

2.6.2   Crystallization of Pml1p

Full-length Pml1p was crystallized by the sitting drop vapor diffusion method using a reservoir comprising 100 mM Tris-HCl (pH 7.5), 200 mM LiSO$_4$, 21-24 % PEG3350 and IPTG as an additive. For experimental phasing, Pml1p crystals were incubated in reservoir solution containing 1 mM of Baker's dimercurial (Hampton Research Inc.). Soaked crystals were transferred into reservoir solution containing 1 mM Baker's dimercurial and 6 % polypropylene glycol (PG) and were flash frozen in liquid nitrogen for data collection. Native crystals were transferred into perfluoropolyether (Hampton Research Inc.) and trace amounts of reservoir were removed before flash freezing in liquid nitrogen.

Pml1p$^{51-204}$ was crystallized by the hanging drop vapor diffusion technique using a reservoir comprising 75 mM HEPES (pH 7.1), 17 % PEG 10000, 15 % glycerol and 6.5 % ethylene glycol. Crystals appeared after five days at 20 °C. For diffraction experiments, crystals were directly frozen in a liquid nitrogen stream.

2.6.3   Diffraction data collection and structure determination

A native dataset of a full-length Pml1p (Pml1p^{FL}) crystal was collected at 100 K on beamline BW6 (DESY, Hamburg, Germany). Crystal symmetry and unit cell dimensions (Table 3.3) suggested two Pml1p molecules per asymmetric unit. Crystals soaked with Baker's dimercurial were measured at the high-energy side of the Hg K-edge on beamlines PXII (SLS, Villigen, Switzerland) and 14-2 (BESSY, Berlin, Germany) and were non-isomorphous with the native crystals. Diffraction data were processed using DENZO/Scalepack (Otwinowski and Minor, 1997). The structure was solved by the single anomalous dispersion (SAD) technique using the

SHELX program collection (Sheldrick, 2008). Scaled diffraction data were analyzed using SHELXC. The heavy atom substructure was solved using SHELXD. Initial phases were calculated using SHELXE and improved by solvent flattening. The SAD electron density map allowed building of β-sheets that could be used as a search model for phasing of the native dataset. The model was completed by iterative cycles of manual model building in Coot (Emsley and Cowtan, 2004) and automated refinement with the program Phenix (Adams et al., 2004). During initial refinement cycles, the model was restrained by a twofold non-crystallographic symmetry.

Diffraction data of a $Pml1p^{51-204}$ crystal were measured on beamline PXII of SLS and processed as above. Crystal symmetry and unit cell dimensions (Table 3.3) suggested one $Pml1p^{51-204}$ molecule per asymmetric unit. The structure was solved by molecular replacement with the program Molrep (Vagin and Teplyakov, 2000) using the structure coordinates of the full-length protein. Automated refinement and building of the water structure were done with Phenix and the model was manually adjusted by using Coot.

The final structural models were validated with CNS (Brunger et al., 1998) by calculating simulated annealing composite omit maps covering the entire asymmetric units of the two crystal forms. Secondary structure elements were assigned with the program DSSP (Kabsch and Sander, 1983). Coordinates and structure factors of $Pml1p^{FL}$ and $Pml1p^{51-204}$ will be deposited in the Protein Data Bank (http://www.pdb.org) upon publication of the manuscript.

2.7   Biophysical methods

2.7.1   Isothermal titration calorimetry (ITC)

His_6-tagged proteins were expressed, captured on Ni^{2+}-NTA resin, washed and eluted as described. All proteins were further purified via gel filtration on a Superdex 75 26/60 column (GE Healthcare) in buffer (10 mM Na/K PO_4 (pH 7.5), 150 mM NaCl, 1 mM β-mercaptoethanol). An additional ion exchange chromatographic step *via* CM-Sepharose was required to remove residual impurities from $Pml1p^{1-204}$ and $Bud13p^{1-266}$. All purified samples were dialyzed twice against the same batch of

buffer (10 mM Na/K PO$_4$ (pH 7.5), 150 mM NaCl, 1 mM β-mercaptoethanol) overnight. SDS-PAGE analysis suggested purities above 95 % for each polypeptide. All protein concentrations were determined based on absorbance at 280 nm using calculated extinction coefficients.

The thermodynamics of binding of full-length proteins and minimal fragments of Bud13p and Pml1p (Bud13p$^{1-266}$, Bud13p$^{202-266}$, Bud13p$^{202-266}$[Trp232Ala], Pml1p$^{1-204}$ and Pml1p$^{1-61}$; employed as titrants) to full-length or minimal fragments of Snu17p (Snu17p$^{1-148}$, Snu17p$^{25-138}$ and Snu17p$^{25-106}$; employed as analytes) were determined using a VP-ITC microcalorimeter (MicroCal Inc.). A typical experiment involved 60 µM titrant in a 300ml injection syringe and 1300 µl of analyte at 6 µM in the sample cell. ITC analyses were carried out at 20°C with 20 injections of 15 ml at 5 min intervals. Plain buffer was injected into protein solutions as a control. The heat releases from the control experiments were subtracted from the experimental data before evaluation. Data were analyzed with Microcal Origin 7.0 to extract the enthalpies and entropies of binding (ΔH_a and ΔS_a, respectively), the equilibrium dissociation constants (K_d) and the interaction stoichiometries (n; Table 3.2).

2.7.2   Circular dichroism (CD) spectroscopy

Fragments His$_6$-Pml1p$^{1-61}$ and His$_6$-Bud13p$^{202-266}$ were obtained as described above but were further purified *via* a Superdex-75 gel filtration (GE Healthcare). Measurements were performed on a Chirascan Circular Dichroism Spectrometer (Applied Photophysics). Peptides were dissolved in buffer (10 mM Na/K PO$_4$ (pH 7.5), 150 mM NaF). Measurements were performed at 20°C, data were corrected for the buffer signal and analyzed using the manufacturer's software.

2.7.3   Intrinsic tryptophan fluorescence measurements

Binding studies showed that a Snu17p fragment (residues 25-106) can interact with a Bud13p peptide (residues 202-266). Within the Bud13p sequence a tryptophan residue (Trp^{232}) seems to play an important role, since mutation of this tryptophan to an alanine abolishes the interaction with Snu17p. To elucidate a direct involvement of Trp^{232} in the interaction with Snu17p, its intrinsic fluorescence was measured in

the bound an unbound state. Purified His$_6$-Snu17p$^{25-106}$ fragments, comprising the RRM, have been dialyzed against binding buffer (20 mM Na/K PO$_4$ (pH 7.5), 150 mM NaCl) and concentrated to 39.2 µM using a VIVAScience Ultrafiltration concentrator (5000 MWCO). Bud13p peptides (residues 222-242; purchased from Peptide Specialty Laboratories GmbH, Heidelberg, Germany) were diluted to a final concentration of 4 mM in double distilled water. As a control, 1.2 ml of binding buffer were pipetted into a quarz cuvette and a magnetic stirrer was added. The cuvette was placed in a HORIBA Jobin Yvon Fluorolog3 system and a buffer spectrum was measured after equilibration of the solution for 2 min at 20°C. All spectra were collected from 304 to 450 nm in increments of 2 nm at an excitation wavelength of 295 nm. Bud13p peptides were added to a final concentration of 0.5 µM and a tryptophan fluorescence spectrum of the peptide in its free form was measured. His$_6$-Snu17p$^{25-106}$ protein fragments were added to a final concentration of 0.5 µM and a tryptophan fluorescence spectrum of the peptide in its bound state was measured. A fluorescence spectrum of His$_6$-Snu17p$^{25-106}$ protein fragments alone was measured in a new cuvette in a volume of 1.2 ml with a protein concentration of 0.5 µM. After subtracting the buffer spectrum, the emission spectra were plotted as fluorescence intensity vs. wavelength.

For a titration experiment, the change in tryptophan fluorescence was measured at a fixed wavelength of 345 nm. Bud13p peptide was dissolved in 1.2 ml of buffer (20 mM Na/K PO$_4$ (pH 7.5), 150 mM NaCl) to a final concentration of 400 nM and tryptophan fluorescence of the unbound state was measured. 22 titration steps were performed and 40 pmol (4 µl) of His$_6$-Snu17p were added per step. An equilibration phase of 1 min was set prior to each tryptophan fluorescence measurement. Data were corrected for dilution effects and plotted as fluorescence intensity against titrant concentration. The dissociation constant for the His$_6$-Snu17p$^{25-106}$/Bud13p$^{222-242}$ interaction was determined essentially as described in (Fan et al., 2007).

All data were analyzed using GraphPad Prism$^®$ non-linear regression software. In brief, the concentration of Snu17p at each titration point was entered as the *X*-value and the fluorescence absorbance values (at λ = 345 nm) were entered as *Y*-values. The first titration point in the fluorescence absorbance represents the signal from

Bud13p$^{222-242}$ alone in the absence of His$_6$-Snu17p$^{25-106}$ (F$_0$, see below). All calculations were based on the observed 1:1 binding stoichiometry for full-length proteins in ITC measurements (section 3.3.11). Equation (1) relates fluorescence of Bud13p$^{222-242}$ to the fractional saturation of Bud13p$^{222-242}$ (A) with His$_6$-Snu17p$^{25-106}$ (T). Equation (2) relates the dissociation constant K$_d$ to A and T. Equation (3) relates T to fractions of bound and unbound T.

$$[AT]/A_{tot} = (F_0 \text{-} F)/(F_0 \text{-} F_\infty) \qquad\qquad (1)$$

$$K_d = [A][T]/[AT] \qquad\qquad (2)$$

$$T_{tot} = [AT] + [T] \qquad\qquad (3)$$

T (titrant) is the concentration of His$_6$-Snu17p$^{25-106}$ and A (analyte) is the concentration of Bud13p$^{222-242}$. T$_{tot}$ is the total concentration of His$_6$-Snu17p$^{25-106}$ at each titration point (X). F$_0$-F is the change in fluorescence following each addition of His$_6$-Snu17p$^{25-106}$ (Y). F$_\infty$ is the fluorescence intensity at saturating His$_6$-Snu17p$^{25-106}$ concentration and F$_0$ is the fluorescence intensity of Bud13p$^{222-242}$ alone.

Combining equations (1)-(3) produces equation (4), from which the equilibrium binding constant was calculated using GraphPad Prism software:

$$F_0 \text{-} F = [(F_0 \text{-} F_\infty)/2A_{tot}]\{K_d + A_{tot} + T_{tot} - [(K_d + A_{tot} + T_{tot})^2 - 4A_{tot}T_{tot}]^{1/2}\} \qquad (4)$$

2.7.4 Fluorimetric structure probing

In order to probe the fold of a protein, a fluorimetric approach was used (modified from (Vedadi et al., 2006)). His$_6$-tagged purified proteins Snu17p$^{1-148}$, Pml1p$^{1-204}$, and Bud13p$^{1-266}$ (see above) were diluted in buffer (20 mM Na/K-PO$_4$ (pH 7.5), 200 mM NaCl, 0.5 mM DTT) to a final concentration of 30 μM in a 40 μl reaction volume and pipetted into a 96-well plate. After applying 4 μl of the fluorophor SYPRO Orange to each reaction, the change in fluorescence emission upon increasing the temperature was monitored on a DNA Engine Opticon^{TM} system in steps of 1°C. Starting temperature was 15°C and final temperature 95°C with hold steps of 30 sec between reads. Fluorescence intensities were plotted as a function of temperature and corrected for the signal of the dye in buffer.

2.7.5   Nuclear Magnetic Resonance (NMR)

All NMR experiments were performed by Shengqi Xiang in the group of Dr. Markus Zweckstetter at the Max-Planck-Institute for Biophysical Chemistry (Göttingen, Germany). Hetreonuclear single quantum coherence (HSQC) measurements were performed on a 600 MHz Bruker Avance[TM] III NMR spectrometer equipped with a triple-resonance probehead at 15 °C. Data were processed by NMRPipe software (Delaglio et al., 1995) and displayed with Sparky - NMR assignment and integration software (T.D. Goddard and D.G. Kneller, University of California, San Francisco, USA).

2.8   Bioinformatic and -computational methods

2.8.1   Multiple sequence alignments

Multiple sequence alignments were generated with ClustalX software (http://bips.u-strasbg.fr/fr/Documentation/ClustalX/) or by the VAST homology server (Gibrat et al., 1996).

2.8.2   Structural modeling

A global architectural model of the RES complex was devised by combining the present crystal structure of Pml1p and the model of the Snu17p RRM in complex with binding peptides of Pml1p and Bud13p. The peptide representing Pml1p was connected to the crystal structure of the folded portion of Pml1p by manually inserting the corresponding number of alanine residues. The Pml1p FHA domain and the Snu17p RRM were oriented arbitrarily relative to each other. The peptide representing bound Bud13p was manually extended to the length of Bud13p by addition of alanine residues in a random coil conformation. All manual modeling was done with Coot.

3.   Results

3.1   Native 12S U2 snRNPs

Previous studies had shown that 17S U2 snRNP particles could be isolated from HeLa cell nuclei for biochemical purposes (Behrens et al., 1993; Kramer et al., 1999; Will et al., 2001; Will et al., 2002). The protein compositions of these particles were highly dependent on the ionic strength of the buffers used. At high salt concentrations particles dissociate into three independent complexes: SF3a, SF3b and 12S U2 snRNPs (Behrens et al., 1993; Brosi et al., 1993).

In the following sections, attempts to establish purification protocols for native 12S U2 snRNPs, which are feasible for crystallization experiments, are described. Besides the purification of untreated 12S U2 snRNPs, further strategies to remove presumably unstructured regions, that may be unfavorable for the crystallization process, were worked out. As targets for this controlled trimming procedure, the U2 snRNA as well as unstructured parts of the proteins, e.g. the C-terminal regions of the Sm proteins B/B', were chosen. Of particular interest was (i) to maintain the integrity of the particles and (ii) to cleave to completion. RNases, such as RNaseA or RNaseH, a DNAzyme, or Exo- and Endoproteases, like Carboxypeptidase-Y and Asp-N, were used for trimming.

A considerable challenge was the low yield of 12S U2 snRNP particles per preparation. The amount of 12S U2 snRNP was only about 1/5 to 1/10 of the amount of U1 snRNP. Additionally, 12S U2 snRNP peak fractions were overlapping with U1 snRNPs during ion exchange chromatography on a DEAE-Sepharose column. Therefore, experiments to optimize the yield of 12S U2 snRNPs per preparation were performed first.

3.1.1 Purification and crystallization trials of native 12S U2 snRNPs

To obtain 12S U2 snRNPs for crystallization, an established protocol with minor modifications was used (Bach et al., 1990; Behzadnia et al., 2006). Briefly, the protocol includes cultivation of HeLa cells in a fermenter with subsequent cell harvesting, preparation of nuclear extract and an affinity chromatography step utilizing a non-commercial antibody-coupled Sepharose column. The coupled antibody, termed H20, recognizes the $m_3^{2,2,7}$G-cap present at the 5' ends of the U1, U2, U4 and U5 snRNAs. Peak fractions of eluate of the antibody-column were pooled and applied on a 10-30% glycerol gradient. In this fashion, populations of tri-snRNPs, U5 snRNPs and U1/U2 snRNPs are separated (Fig. 3.1).
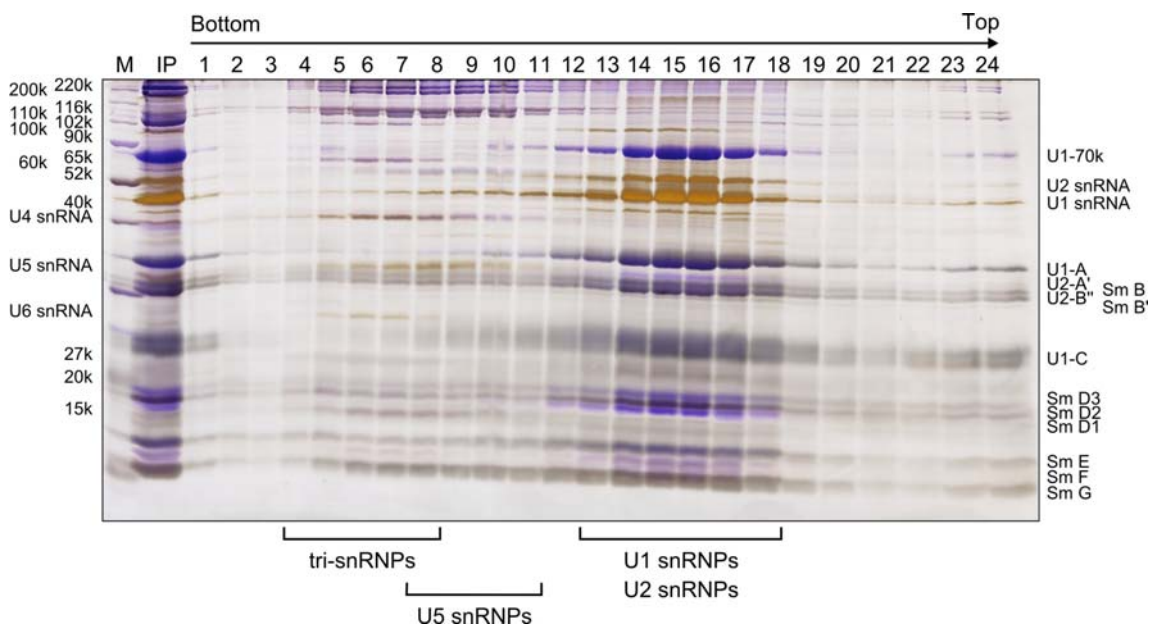
Fig. 3.1: Glycerol gradient centrifugation of spliceosomal complexes. 50 µl samples of preparative 10-30 % glycerol gradient fractions were analyzed on a 13.5 % SDS polyacrylamide (PAA) gel. Proteins were visualized by Coomassie blue staining and after destaining, RNAs were visualized by silver stain. Proteins and RNAs are indicated on the left and right of the gel. LSm proteins are not shown for clarity. Fraction numbers are shown above the gel and particle distributions along the gradient are indicated below. M - protein size marker; IP - input.

To separate 12S U2 snRNPs from U1 snRNPs, a DEAE-Sepharose column was used but peak fractions usually contained a mixture of both particles (Fig. 3.2 A). Therefore different anion exchange chromatography media were tested for their

ability to separate U1 from U2 snRNPs. As illustrated in figure 3.2 (B), the use of an ANX-Sepharose column improved the separation but was still insufficient. By using a Heparin-Sepharose column on an ÄKTA prime system, we were able to completely separate U1 from U2 snRNPs, which originated from pooled and diluted gradient fractions (Fig. 3.2 C). 12S U2 snRNPs usually eluted at an ionic strength of approximately 220-250 mM salt under the conditions used. The 12S U2 snRNPs could subsequently be buffer-exchanged and concentrated by ultrafiltration to 12 mg/ml for crystallization trials.

Although 1152 different conditions were screened for crystal growth at 4 °C, no crystals or crystalline precipitates were observed. Therefore a new strategy of pre-processing, described in section 3.1.2, was developed.
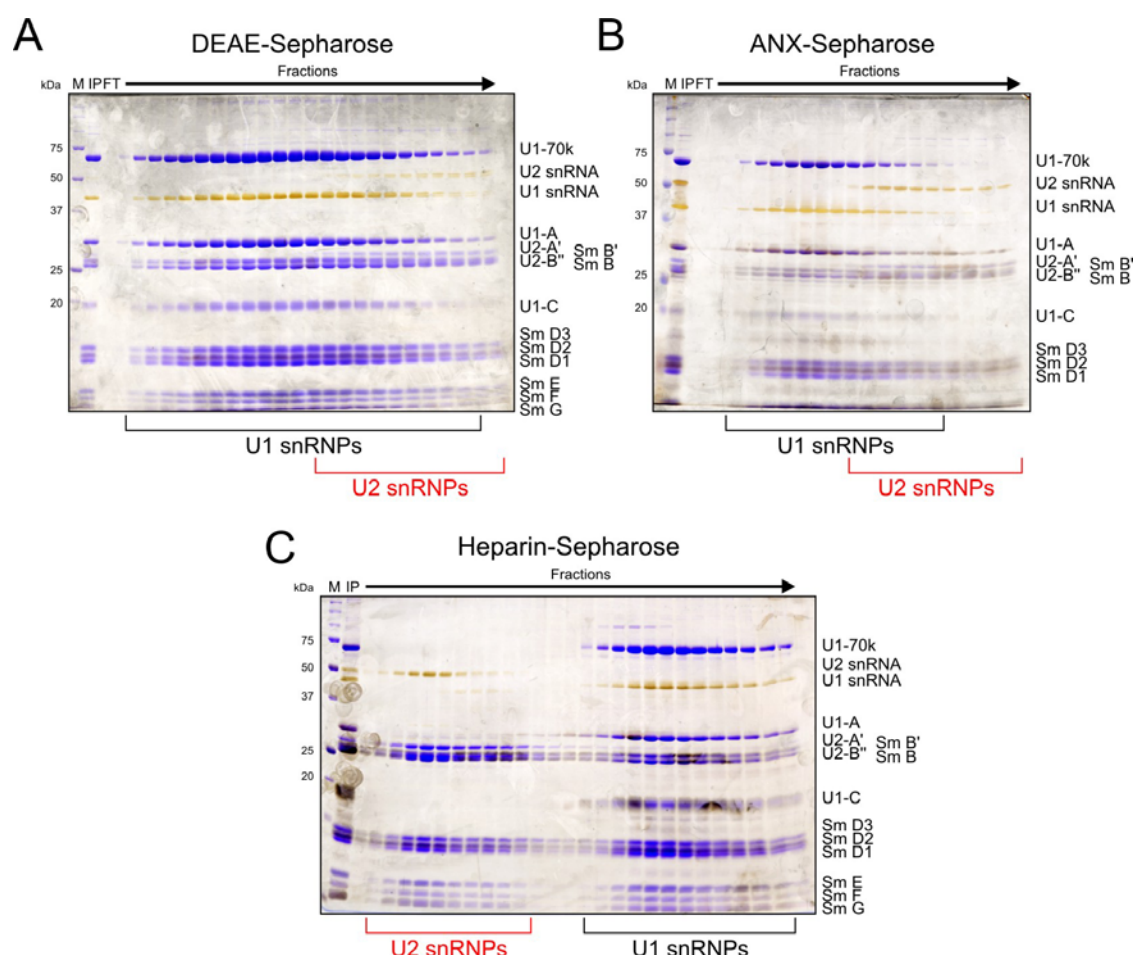
For figure description see next page...

Fig. 3.2: Separation of U1 and U2 snRNPs via various anion exchange chromatography media. 50 µl samples of peak fractions from salt gradient eluates (50-1000 mM NaCl) were analyzed on 13.5% SDS-PAA gels. Proteins were visualized by coomassie brilliant blue staining and after removal of unbound coomassie, RNAs were colored by silver staining. Proteins and RNAs are indicated on the right of the gels. **(A)** A U1/U2 snRNP mixture is separated on a DEAE-Sepharose column. Peaks of both particles are overlapping to a large extent. The separation is slightly improved by the use of an ANX-Sepharose column **(B)** and particles can be completely separated by the use of a heparin-Sepharose column **(C)**. Protein size markers are indicated on the left. Particle distributions of U1 snRNPs (black) and U2 snRNPs (red) are shown below each gel. M - marker, IP - input, FT - flow-through.

3.1.2   Purification and crystallization trials of Δ5' U2 snRNPs

Crystallization trials with full-length 12S U2 snRNPs remained unsuccessful raising the question, whether flexible regions are present in 12S U2 snRNP particles, which prevent crystallization. Among the most obvious unstructured or flexible regions is the first half (5' part) of the U2 snRNA. This region serves as a binding platform for SF3a and SF3b particles, which are absent in 12S U2 snRNPs (Dybkov et al., 2006; Kramer et al., 1999). Therefore, we focused on removing this region of the U2 snRNA. Trimming of the U2 snRNA by endonucleases, such as RNase A or RNase T1, was attempted under numerous conditions but did not yield processed U2 snRNP particles in a reproducible manner. Also an *in situ* crystallization approach (modified from Dong et al., 2007), in which RNase trimming was attempted during the crystallization process in drops, failed to produce crystalline material. Therefore, we focused on an RNaseH-based approach to remove 5' parts of U2 snRNAs.

DNA oligonucleotide-directed RNaseH cleavages were successfully used for the depletion of snRNA components in nuclear extracts and to site-specifically cleave RNAs *in vitro* (Gunzl et al., 2002; Inoue et al., 1987; Lapham and Crothers, 1996). At first, I designed three oligonucleotides (oligo A, B and C) and tested them for their effectiveness in RNaseH cleavage reactions. All three oligos were complementary to a region 5' of the Sm-binding site with the longest extending into loop IIb of U2 snRNA (Fig. 3.3 A). No cleavage occurred when the shortest oligonucleotide (oligo A) was used (Fig. 3.3 B, lane 3), whereas cleavage could be observed with both longer oligos B and C (Fig. 3.3 B, lane 4, 5). Since the cleavage reaction with the longest oligonuleotide (oligo C) was the most efficient, oligo C was

chosen for further experiments. In order to optimize the cleavage reaction, several titration experiments were performed. First, the impact of different amounts of RNaseH on the cleavage efficiency was investigated. The efficiency was significantly improved by addition of increased amounts of RNaseH (Fig. 3.3 C, lanes 2-5). In contrast, titration experiments with increasing amounts of DNA oligonucleotides showed that higher concentrations of oligonucleotides had only minor effects on the cleavage efficiency (Fig. 3.3 D, lanes 3-5). A longer incubation time also improved the cleavage efficiency (Fig. 3.3 E, lanes 4-6), whereas refreshing of the reducing agent (DTT) during the cleavage reaction was not of importance (Fig. 3.3 E, lane 3).
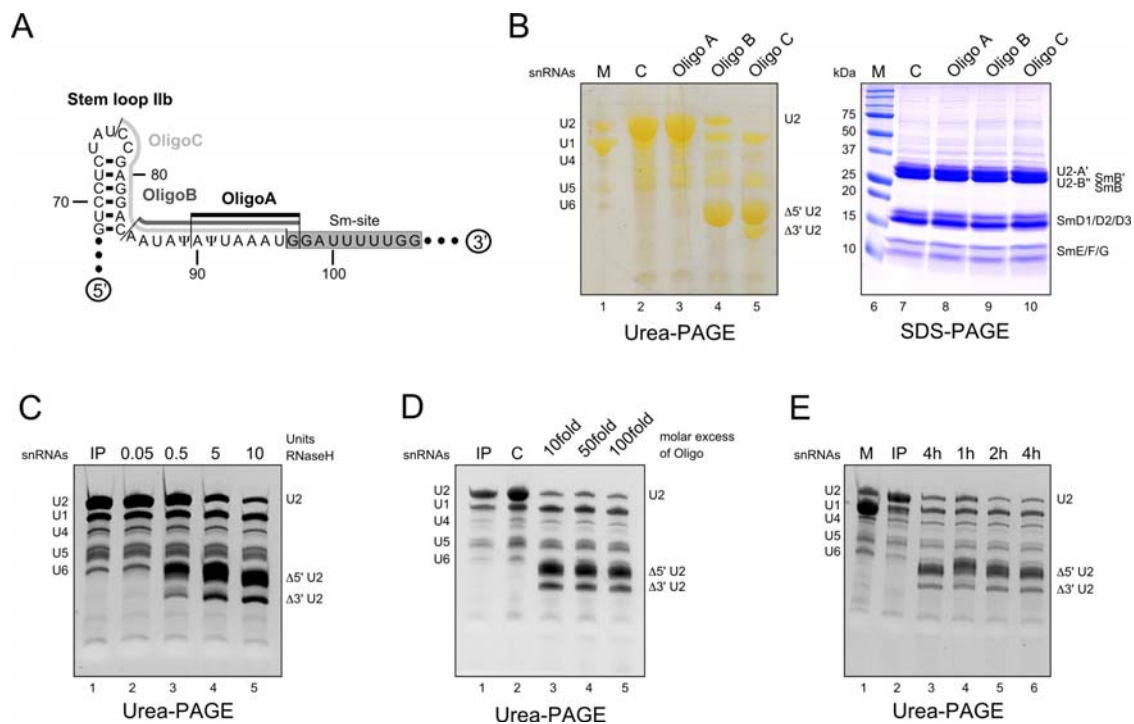
Fig. 3.3: Titration experiments to optimize the RNaseH cleavage of U2 snRNA. (A) Schematic representation of a region of U2 snRNA (nucleotides 68-106). Binding regions for oligo A (black), oligo B (anthracide) and oligo C (grey) as well as the Sm site (boxed) are indicated. **(B)** DNA-directed RNaseH cleavage with oligos described in (A) show that no cleavage of U2 snRNA occurs with oligo A (lane 3), a cleavage with moderate efficency is observed with oligo B (lane 4) and a complete cleavage occurs with oligo C (lane 5). Proteins are not affected (lanes 8-10). **(C-E)** Titration experiments with increasing amounts of RNaseH **(C)**, oligo C **(D)** and with different incubation times **(E)** show that different RNAseH concentrations have the strongest effect on the cleavage reaction (see text). All reactions were done with purifed, native 12S U2 snRNPs. RNAs and proteins were separated by PCI extraction and analyzed by Urea-PAGE or SDS-PAGE, respectively. RNAs were visualized by silver and ethidium bromide stain (B and C-E, respectively) and proteins were stained by Coomassie blue. Identities of snRNAs and proteins are indicated. M - marker; C - control; IP - input.

Since isolated 12S U2 snRNP particles are kept at 30 °C for several hours to achieve an efficient RNaseH cleavage reaction, we asked whether the particles remain intact during the treatment. To address this question, particles were incubated in the presence of RNaseH and oligo C under optimized conditions until a complete cleavage was achieved and were subsequently passed over an analytical size exclusion chromatography (SEC) column. Δ5' U2 snRNPs migrated as globular particles with an estimated molecular weight of ~170 kDa, in agreement with the nominal molecular weight of ~210 kDa (Fig. 3.4 A).
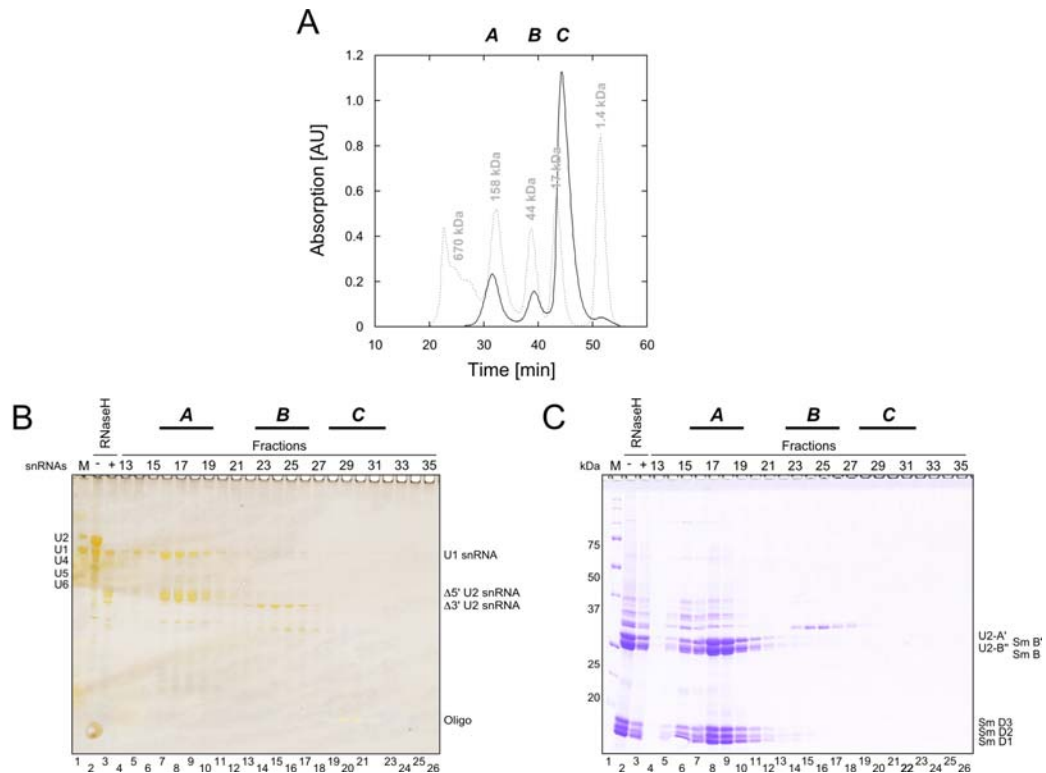
Fig. 3.4: Δ5' U2 snRNP particles are intact and monodispers. (A) 12S U2 snRNPs were cleaved by RNAseH using oligo C and components were separated on an analytical Superdex 200 PC3.2 column in a buffer comprising 20 mM Tris-HCl (pH 7.9), 250 mM NaCl and 1 mM DTT. The absorption profile at 280 nm is shown in (**A**, black line). The absorption profile is characterized by the appearance of three major peaks *A*, *B* and *C*. For comparison, the absorption profile of a standard is shown as dashed grey line and molecular weights of size markers are indicated. Samples from the SEC run were subjected to PCI extraction and RNAs were analyzed by Urea-PAGE with subsequent silver staining (**B**), whereas proteins were separated by SDS-PAGE and visualized by Coomassie blue stain (**C**). Peak *A* corresponds to Δ5' U2 snRNPs (panel **B/C**, lanes 6-10), peak *B* to Δ3' U2 snRNA and a protein with similar size as U1-A and peak *C* to oligo C. Protein and RNA identities are depicted on the right of each gel. AU - absorption units; M - marker; minus - before RNAseH cleavage; plus - after RNAseH cleavage.

Gel electrophoretic analysis of peak fractions showed the presence of intact U2 snRNP core particles co-migrating with the longer RNA species derived from the RNaseH cleavage reaction (Fig. 3.4 B/C, lanes 6-11). A second, shorter RNA species eluted later, indicating that this RNA is the cleaved 5' part of U2 snRNA (Fig. 3.4 B/C, lanes 14-17).

Besides the integrity of Δ5' U2 snRNP particles, the chromatographic analysis showed that contaminating U1 snRNA co-migrates with processed U2 snRNPs, most probably as Sm core-containing RNPs. The presence of the latter RNPs differs from preparation to preparation. Since these RNPs might obstruct the crystallization process, a further purification step was introduced (Fig. 3.5 A). I reasoned that an H20 antibody-coupled Sepharose column should capture U1 core snRNPs and the cleaved 5' part of U2 snRNAs, whereas Δ5' U2 snRNP particles should remain in the flow-through. To test this idea, 12S U2 snRNPs were cleaved by RNaseH as described and loaded on the H20 antibody-coupled Sepharose column. Contaminating U1 snRNAs and the shorter U2 snRNA species were retained on the column and could be specifically eluted by the addition of an $m_3^{2,2,7}$G-cap analog (Fig. 3.5 B/C, lanes 7, 8). Cleaved Δ5' U2 snRNP particles were found exclusively in the flow-through and wash fractions (Fig. 3.5 B/C, lanes 4-6).
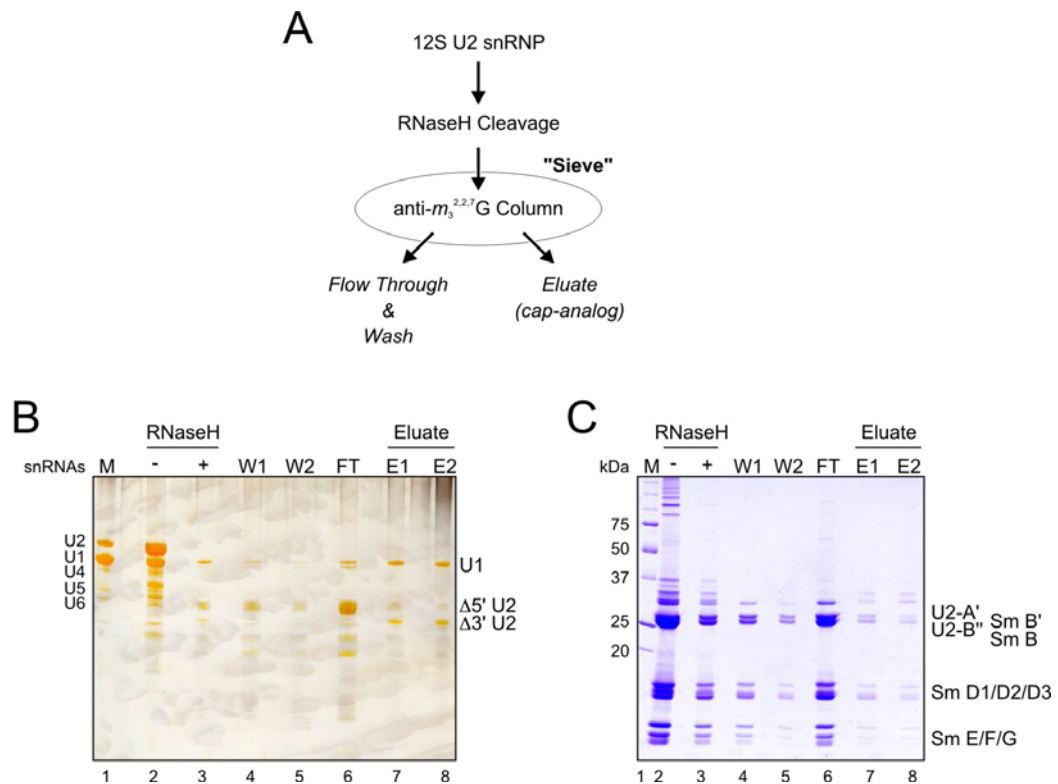
Fig. 3.5: Δ5' U2 snRNP particles can be separated from snRNP core particles. (A) Schematic representation of the purification step. See text for further details. Proteins and RNAs from all samples were separated by PCI extraction prior to electrophoresis. **(B)** Analysis of RNA fragments by Urea-PAGE. RNA fragments before (lane 2) and after (lane 3) DNA-directed RNAseH cleavage reaction, as well as wash (lane 4, 5) and flow-through fractions (lane6) of the H20 antibody-coupled column are shown. Samples of the $m_3^{2,2,7}$G cap analog eluates are given in lanes 7 and 8. Note that putative Δ3' U2 snRNA fragments can be specifically eluted by the addition of the cap analog. **(C)** SDS-PAGE analysis of the proteins samples. Arrangement of the lanes is as in **(B)**. Proteins and RNAs are indicated on the right of the gel. Coomassie staining of the gel reveals the presence of all 12S U2 snRNP proteins. M -marker, W - wash fractions, FT - flow-through fractions, E - eluate fractions.

To purify Δ5' U2 snRNP particles in a preparative way for crystallization trials, preparation of 12S U2 snRNPs, RNaseH cleavage, affinity chromatography and size exclusion chromatography were combined. Analyses of RNA purities after each step (starting from pre-purified 12S U2 snRNPs) and the final protein composition of purified particles are compiled in figure 3.6. In brief, 12S U2 snRNPs were isolated from total snRNP mixtures as described in section 3.1.1, cleaved by RNase H (Fig. 3.6 A), passed over a H20 antibody-coupled Sepharose column (Fig. 3.6 B) and, subsequently, over a SEC column (Fig. 3.6 C). Protein compounds of each SEC fraction were also analyzed via gel electrophoresis and it turned out that all 12S U2

snRNP proteins were present and co-eluted with the RNA (Fig. 3.6 C/D, lanes 5-9). Ultracentrifugation, a method used for concentrating ribosomes and U1 snRNPs (Spiess, 1978) (B. Kastner, personal communication), was used before and after the final purification step.
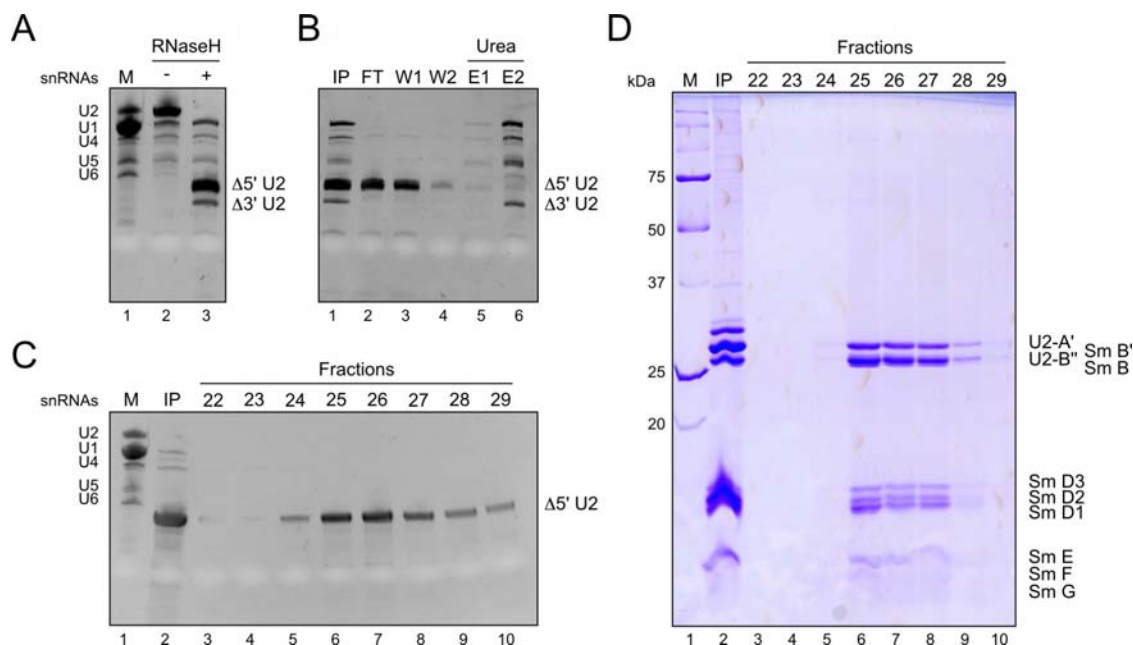
Fig. 3.6: Δ5' U2 snRNP particle preparation for crystallization attempts. (A) PCI extracted snRNAs prior to (lane 2) and after DNA-directed RNAseH cleavage (lane 3) of 12S U2 snRNPs were separated by Urea-PAGE and stained with ethidium bromide. Presumed 5' and 3' parts of the RNA are indicated on the right. **(B)** Urea-PAGE with subsequent ethidium bromide staining reveals the improved purity of Δ5' U2 snRNAs in H20 antibody-column flow-through (lane 2) and wash fractions (lane 3, 4). Contaminants remain bound to the antibody-column during the purification procedure as show by the analysis of urea wash fractions (lane 5, 6). SEC further improves the purity of the protein/RNA complex and confirms its monodispersity. Proteins and RNAs from SEC peak fractions were PCI extracted and analyzed by Urea-PAGE **(C)** (lanes 3-10) and SDS-PAGE **(D)** (lanes 3-10), respectively. Proteins and RNAs are depicted on the right of each gel. M - marker, IP - input, FT - flow-through, W1/W2 - wash fractions, E1/E2 - fractions after elution with urea.

The putative 3' part of U2 snRNA co-migrated with U2 snRNP specific proteins and had lost the ability to interact with H20 antibodies (Fig. 3.6). These observations are consistent with the longer RNA species to comprise the 3' part of U2 snRNA. In order to further verify the nature of this RNA, an RNaseH-based approach was used. DNA oligos complementary to certain regions of the U2 snRNA were used to site-specifically cleave isolated full-length U2 snRNA and the presumed Δ5' U2 snRNA (Fig. 3.7 A). Sizes of full-length U2 snRNA or Δ5' U2 snRNA fragments to be

expected after RNaseH treatment are listed in figure 3.7. Oligo1 generated 100 and 60 residue fragments from full-length U2 snRNAs (Fig. 3.7 B, lane 6), oligo2 generated 75 and 90 residue fragments (Fig. 3.7 B, lane 9) and oligo3 generated 155 and 10 residue fragments (Fig. 3.7 B, lane 12). Fragments of ~62 and ~9 residues generated by oligo1 from Δ5' U2 snRNA were hardly detectable on the gel (Fig. 3.7 B, lane 7). Δ5' U2 snRNA was not affected by oligo2 (Fig. 3.7 B, lane 10), whereas oligo3 generated a fragment of ~60 residues (Fig. 3.7 B, lane 13). These results may indicate that the RNA species of the Δ5' U2 snRNP is the 3' part of the U2 snRNA.
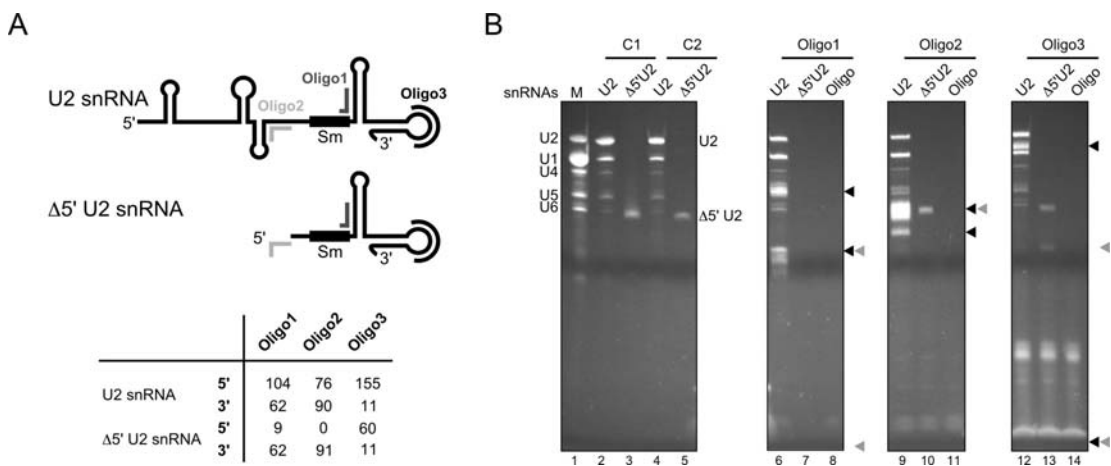
Fig. 3.7: The 3' part of U2 snRNA is present in isolated, truncated particles. (A) Schematic representation of the U2 snRNA of full-length (U2 snRNA) and truncated paticles (Δ5' U2 snRNA). Binding regions for oligo 1 (anthracide), oligo 2 (grey) and oligo 3 (black) as well as the Sm-site (boxed) are indicated. Expected fragment sizes in residues generated by each oligo are listed (lower panel). **(B)** RNAs of U2 snRNPs (U2) and Δ5' U2 snRNPs (Δ5'U2) were extracted by PCI and were treated by RNAseH in the absence of oligo (lanes 4 and 5, respectively) and in the presence of oligo 1 (lanes 6 and 7, respectively), oligo 2 (lanes 9 and 10, respectively) and oligo 3 (lanes 12 and 13, respectively) and analyzed by Urea-PAGE. Untreated RNAs (lanes 2 and 3) and oligos alone (lanes 8, 11 and 14) were also analyzed. Arrowheads indicate positions of expected bands of full-length U2 snRNAs (black) and Δ5' U2 snRNAs (grey) after RNaseH treatment. Sm - Sm site; M - marker; C1 - control 1 (w/o RNaseH and oligo); C2 - control 2 (with RNaseH, w/o oligo).

672 crystallization trials with Δ5' U2 snRNPs at a final concentration of ~11 mg/ml were performed at 4° C, but the assay did not lead to the formation of crystals or crystalline material.

3.1.3   Purification of 12S Δ5' U2 snRNPs with truncated Sm B/B' proteins

Since no crystals could be obtained with purified Δ5' U2 snRNPs, 12S U2 snRNP particles were inspected for further regions, which might inhibit crystal growth. Additional problematic regions are the long and heterogeneous C-terminal tails of the Sm proteins, Sm B and Sm B'. Both proteins are products of an alternatively spliced single gene (Chu and Elkon, 1991; van Dam et al., 1989). These C-terminal tails of Sm B and Sm B' consist of PPP-PGMR repeats, which are thought to serve as unstructured protein-protein interaction regions (Kofler et al., 2004). Furthermore, C-terminal tails of Sm B/B' were shown to be highly methylated (Brahms et al., 2001). Biochemical data (Hermann et al., 1995) and crystal structure analysis (Kambach et al., 1999) showed that a truncated version of Sm B (residues 1-91, corresponding to the Sm motif) is sufficient to interact with its binding partner Sm D3. Consequently, I tried to specifically remove these potential unstructured regions by a limited proteolysis approach.

Among the commercially available proteases, Carboxypeptidase Y acts specifically as an exopeptidase on C-terminal residues of proteins (Hayashi et al., 1973). To test whether C-terminal regions of Sm B/B' can be removed in the context of Δ5' U2 snRNPs, these particles were treated with Carboxypeptidase Y and remaining proteins were analyzed via SDS-PAGE. Interestingly, a band of an apparent molecular weight of ~12 kDa appeared after over night digestion (Fig. 3.8 A, lane 7). Mass spectrometric analysis of the emerging band showed that it matches the region corresponding to the Sm-motif (residues 2-94) of Sm B (Fig. 3.8 C). We reproducibly observed peptides corresponding to a short C-terminally elongated region beyond the Sm motif of Sm B. Thus, we speculate that the very C-terminal regions of proteins Sm B/B' are not involved in protein/protein or protein/RNA interactions in the context of Δ5' U2 snRNPs. C-terminal residues close to the Sm-motif may be stabilized through protein-protein or protein-RNA interactions and therefore may not be accessible for Carboxypeptidase Y activity.
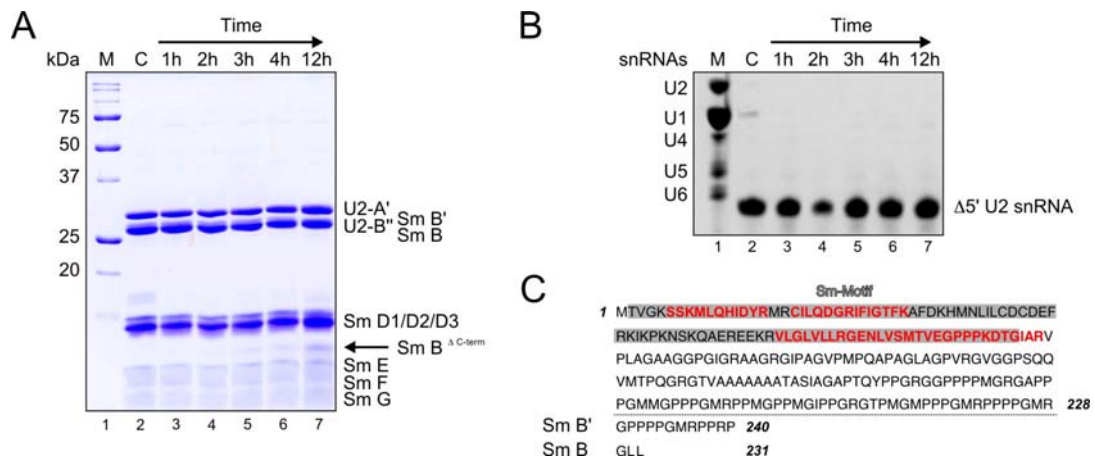
Fig. 3.8: Sm B/B' proteins are processed by Carboxypeptidase Y in Δ5' U2 snRNPs. 10 µg samples of purified Δ5' U2 snRNPs with a concentration of 0.6 mg/ml were mixed with Carboxypeptidase Y in a protease/particle mass ratio of 1:100 and incubated at 25 °C for 1, 2, 3, 4 and 12h. Proteins and RNAs from each time point and from a control reaction without protease were separated by PCI extraction and analyzed by SDS-PAGE **(A)** or Urea-PAGE **(B)**, respectively. **(A)** Proteins, visualized by coomassie blue stain, are indicated on the right of the gel. Note that a band of the size of ~12 kDa (arrow) appears after 3 h of the digestion. **(B)** Urea-PAGE and subsequent ethidium bromide stain showed that Δ5' U2 snRNAs are not affected by the protease treatment (lanes 3-7). **(C)** Peptide mass fingerprint of the appearing band in **(A)** (black arrow) identified peptides (red letters) of the N-terminal part of Sm B/B', covering the Sm motif. The Sm motif (grey boxes) and the C-terminal variation among the two proteins Sm B/B' are shown. M - marker, C - control.

U2 snRNPs with a truncated RNA and truncated Sm B/B' proteins (termed Δ5' Sm B$^{ΔC-term}$ U2 snRNPs) could be prepared in a preparative way. Gel electrophoretic analyses of Δ5' Sm B$^{ΔC-term}$ U2 snRNPs revealed the presence of the proteins U2-A', U2-B'', Sm D1, D2, D3, E, F, G and the N-terminal fragment of Sm B/B', as well as the 3' part of the U2 snRNA (Fig. 3.9 A/B, lanes 3). To test whether the processed particles were intact, they were analyzed by electron microscopy. On negatively stained electron micrographs most of the Δ5' Sm-B$^{ΔC-term}$ U2 snRNPs appeared as particles with a doughnut shaped main body and small protuberances (Fig. 3.9 C). Similar images were obtained by electron microscopy (EM) with untreated 12S U2 snRNP (Kramer et al., 1999). Thus, the Δ5' Sm B$^{ΔC-term}$ U2 snRNPs seem to be intact. Unfortunately, crystallization trials under 768 conditions with Δ5' Sm B$^{ΔC-term}$ U2 snRNPs remained unsuccessful.
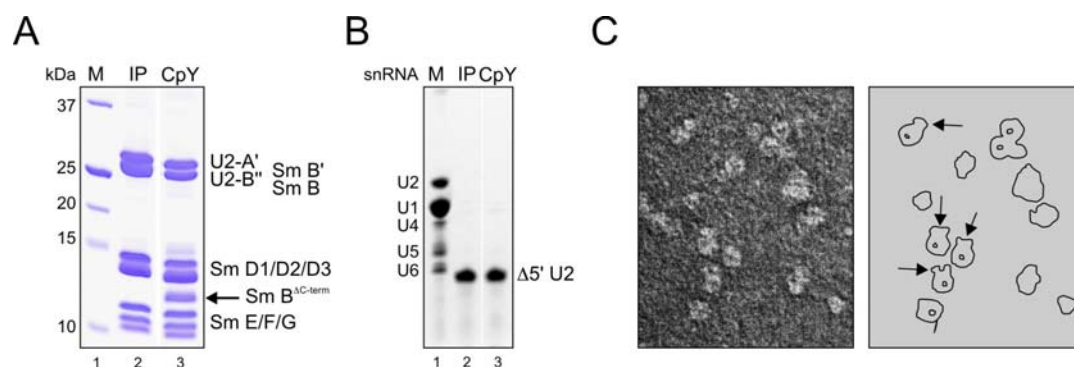
Fig. 3.9: Δ5' Sm B$^{ΔC-term}$ U2 snRNP particle preparation for crystallization attempts and particle analysis by negative stain EM. Δ5' U2 snRNPs were treated with Carboxypeptidase Y and PCI extracted proteins and RNAs were analyzed by SDS-PAGE and Coomassie stain **(A)** or Urea-PAGE and ethidium bromide stain **(B)**, respectively. After proteolysis, the Sm B$^{ΔC-term}$ band is clearly visible (**A**, lane 3, black arrow) and the Δ5' U2 snRNA remains unaffected (**B**, lane 3). **(C)** A section of the electron micrograph (left) of purified Δ5' Sm B$^{ΔC-term}$ U2 snRNPs shows doughnut-shaped particles with small protuberances. A schematic representation of the electron micrograph is given on the right. Protuberances are indicated by arrows. Proteins are indicated. M - marker, IP - input, CpY - Carboxypeptidase Y. EM imaging was done by Dr. P. Dube.

3.1.4 U2-A' and U2-B" proteins as further targets for a preparative, limited proteolysis

Further targets for a controlled trimming via limited proteolyses were searched. Two good candidates are the 12S U2 snRNP specific proteins U2-A' and U2-B" (see Fig. 1.2, introduction). U2-A' belongs to a family of leucine-rich repeat (LRR) proteins which generally fold into a horseshoe-like shape (Enkhbayar et al., 2004; Kobe and Deisenhofer, 1995). In U2 snRNPs, the LRR region (residues 1-176) of U2-A' folds on α-helix A and strand β2 of the first RRM of U2-B" (Price et al., 1998). The crystal structure showed that the first RRM of U2-B" spans residues 4-99 (Price et al., 1998). Thus, U2-A' and U2-B" might encompass flexible regions, e.g. the C-terminal half of U2-A', or flexibly linked domains, e.g. the second RRM of U2-B". In agreement with this notion, the linker region connecting RRM1 and RRM2 of U2-B" was suggested to be flexible and cleavable by proteases (Sillekens et al., 1987). Residues 161-180 of U2-A' were shown to be susceptible to trypsin and chymotrypsin in isolation (Price et al., 1998). Hence, suitable proteases for the cleavage of flexible regions within U2-A' and U2-B" proteins in the context of 12S U2 snRNPs were systematically explored.

Below, two examples will be described. The endoprotease Lys-C hydrolyses peptide bonds C-terminally at accessible lysine residues. Lys-C would most probably cleave U2-B" between residues 101 and 111 due to high incidence of lysine residues. Potential cleavage sites can be assumed for U2-A' C-terminal of Lys^{165}, Lys^{172}, or Lys^{179}. The endoprotease Asp-N hydrolyses peptide bonds N-terminally at accessible aspartic acid or cystein residues. Suitable Asp-N sites locate N-terminal of Asp^{100} in U2-B" and N-terminal of Asp^{173} in U2-A'. Proteins Sm B and Sm B' do not provide appropriate cleavage sites outside the structured Sm motif, neither for Lys-C, nor for Asp-N, and should therefore stay unchanged during proteolysis. Preliminary titration experiments of Lys-C with 12S U2 snRNPs revealed that U2-A' and U2-B" were cleaved as expected, but Sm proteins D1-G were cleaved in a disorderly fashion in addition (Fig. 3.10, lanes 28-31). Sm B/B' proteins were still visible even at high protease concentrations (Fig. 3.10, lane 31, grey dots). Asp-N treatment, however, caused emerging of three bands with apparent molecular weights of ~12, ~16 and ~17.5 kDa (Fig. 3.10, lanes 12-15, black dots). Simultaneously, two bands corresponding to U2-A' and U2-B" disappeared. Thus, the three new bands are most likely stable fragments of proteins U2-A' and U2-B". Proteins Sm B and Sm B' were left unaffected (Fig. 3.10, lane 15, grey dots).
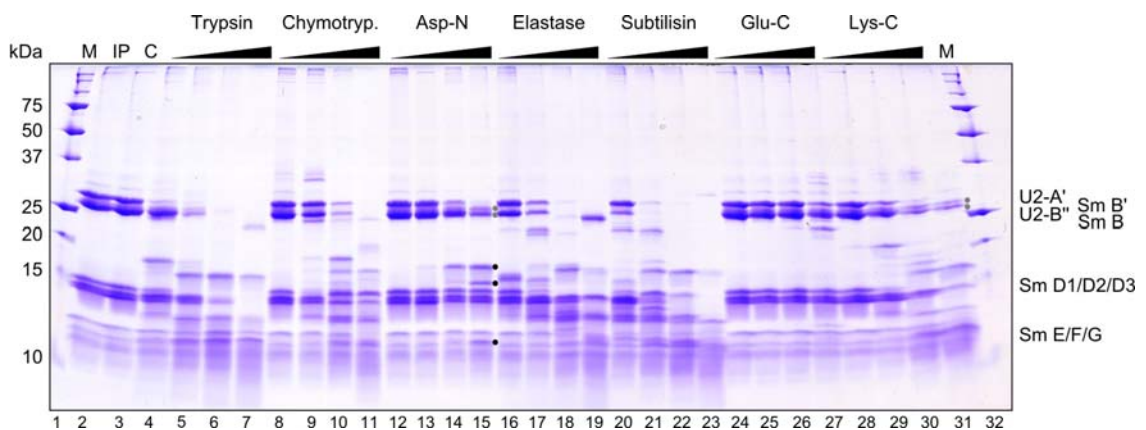
Fig. 3.10: Limited proteolysis experiments with 12S U2 snRNPs. 12S U2 snRNPs were treated with increasing amounts of trypsin (lanes 4-7), chymotrypsin (lanes 8-11), Asp-N (lanes 12-15), elastase (lanes 16-19), subtilisin (lanes 20-23), Glu-C (lanes 24-27) and Lys-C (lanes 28-31). 10 μg of 12S U2 snRNPs were used per reaction and the amount of protease was increased by a factor of ten at each step, achieving a protease/protein mass ratio of 1:10 at the highest protease concentration. Mixtures were incubated at 20 °C for 1 h. Proteins were subsequently separated by SDS-PAGE and visualized by Coomassie blue stain. Protein identities and sizes of a protein marker are given on the right and left of the gel, respectively. M - marker; IP - input; C - control

On account of the above findings, Δ5' U2 snRNPs were processed with Carboxypeptidase Y and Asp-N, and remaining proteins were separated on denaturing polyacrylamide gels (Fig. 3.11 B). Mass spectrometric analysis of a semi-processed and gel-filtrated sample (Fig. 3.11 C) revealed that peptides of U2-A' were present in trimmed particles, which covered the region between residues 2-165 of the protein, corresponding to the LRR region (Price et al., 1998). Furthermore, peptides matching to a C-terminally truncated Sm B protein (covering residues 2-94) and to the first RRM of U2-B" (covering residues 2-95) were found. To summarize, these results are in agreement with solved structures of subcomplexes of U2 snRNPs and provide a starting point for further preparations of minimal U2 snRNPs suitable for crystallization trials.
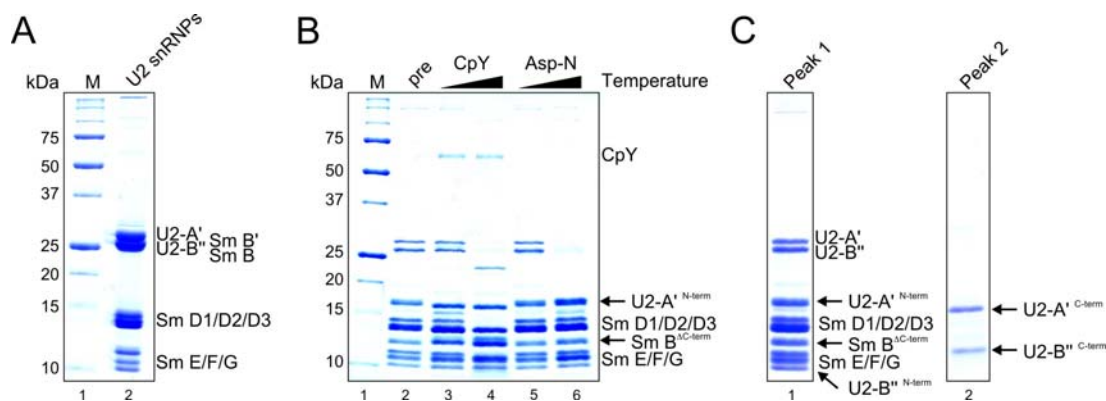
Fig. 3.11: Carboxypeptidase Y and Asp-N treatment of Δ5' U2 snRNPs. (A) SDS-PAGE and subsequent Coomassie blue stain shows the protein composition of an untreated 12S U2 snRNP sample. **(B)** 12S U2 snRNPs were pre-treated with Carboxypeptidase Y and Asp-N (lane 2) and further incubated at 4 (lanes 3, 5) or 30 °C (lanes 4, 6) after supplementing Carboxypeptidase Y (lanes 3,4) or Asp-N (lanes 5,6). Proteins were visualized by SDS-PAGE and Coomassie blue stain. **(C)** A sample of pre-processed 12S U2 snRNPs (**(B),** lane 2) was run on an analytical Superdex 200 SEC column and proteins from peak fractions were separated on a SDS-PAA gel, which was subsequently stained with Coomassie blue. Proteins of peak 1 (lane 1) and peak 2 (lane 2) were analyzed by MS (see text). Protein identities, derived from MS analyses are indicated on the right of each gel (**(B)** an **(C)**). M - marker; CpY - Carboxypeptidase.

3.1.5   Suitability of DNAzymes for trimming of 12S U2 snRNPs

Although U2 snRNAs were efficiently cleaved by DNA oligonucleotide directed RNaseH treatment as shown in section 3.1.2, the 5' ends of the 3' parts of the RNA may be heterogeneous. To safeguard against this problem, chimeric oligonucleotides consisting of a 4 nt DNA and a long 2'-O-methyl-RNA stretch (5'-3' direction),

complementary to the target sequence, could have been used. Such oligonucleotides in combination with RNaseH have been shown to produce a precise, single cut in target RNAs (Lapham and Crothers, 1996). The RNA molecule is cleaved between ribonucleotides, which correspond to nucleotides 1 and 2 of the complementary DNA stretch. Since these chimeras are very expensive and huge amounts would be necessary to cleave U2 snRNAs in a preparative way, we focused on the use of DNAzymes first. Two types of DNAzymes have been engineered, the so-called 8-17 and 10-23 DNAzymes, which only differ in their catalytic motifs (Santoro and Joyce, 1997). Both DNAzymes consist of two flanking binding arm regions, which base pair to a target RNA, and a central looped-out motif forming the catalytically active site (Santoro and Joyce, 1997). The catalytically active motif of 10-23 DNAzymes bound to its target is schematically depicted in figure 3.12 (A). Besides flanking arms with lengths of 7-13 residues, a further necessity for 10-23 DNAzymes to cut is the presence of a purine followed by a pyrimidine (in 5'-3' direction) at cleavage sites in target RNA sequences (Santoro and Joyce, 1997; Santoro and Joyce, 1998). Sequence inspection of the U2 snRNA 5' of the Sm-binding site revealed the presence of one A-U dinucleotide (arrow) at position 86/87 appropriate for the cleavage reaction (Fig. 3.12 B). In contrast to 8-17 DNAzymes, which would only allow an A-G dinucleotide at the cleavage site (Santoro and Joyce, 1997), 10-23 DNAzymes are thus applicable for the sequence specific cleavage of U2 snRNAs 5' of the Sm site. Consequently, a DNAzyme targeting A^{86}-U^{87} with binding arm lengths of 10 and 9 nucleotides was designed (Fig. 3.12 B).
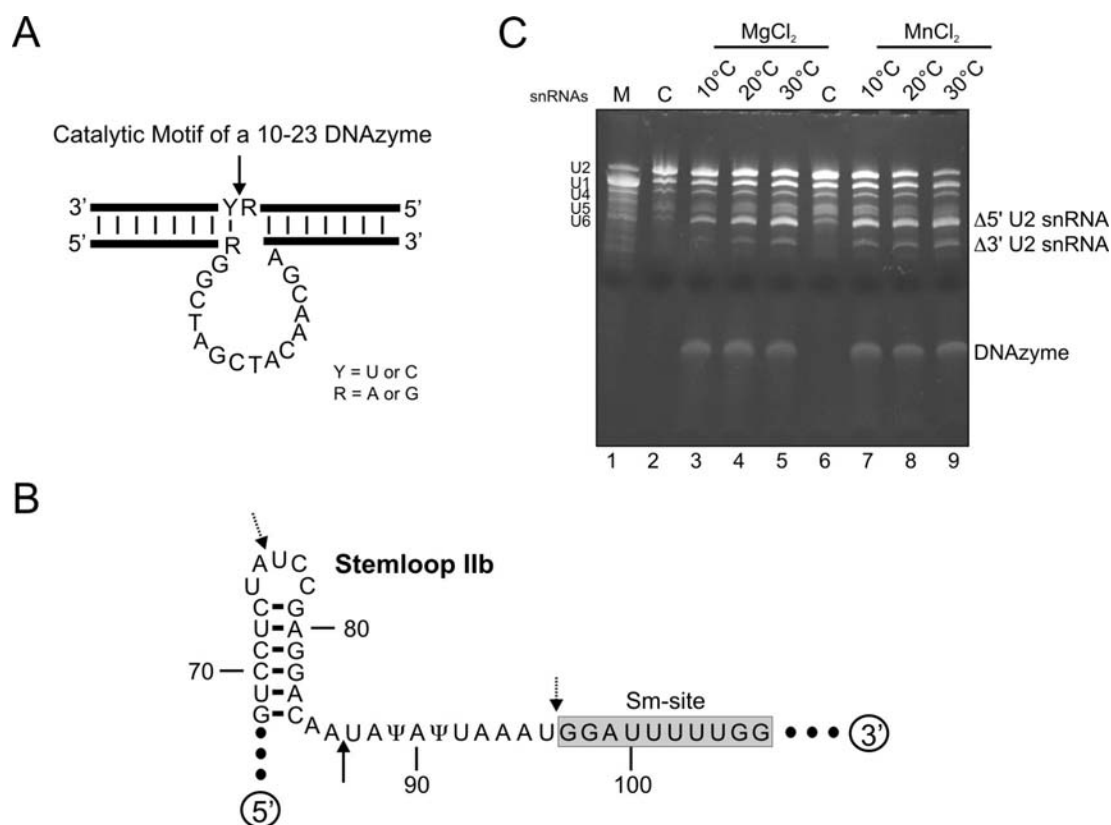
Fig. 3.12: Cleavage of U2 snRNA by a DNAzyme. (A) Schematic representation of a 10-23 DNAzyme (lower part) with indicated cleavage site in the target molecule (upper part) (adapted from (Santoro and Joyce, 1997)). Complementary regions are illustrated as lines. **(B)** Region around stem loop IIb of the U2 snRNA is shown with start and end points of the complementary region of the DNAzyme (dashed arrows). The cleavage site between A^{86}-U^{87} is indicated by an arrow. The Sm site is highlighted by a grey box. **(C)** 12S U2 snRNPs were incubated in the presence of DNAzyme at indicated temperatures, RNAs were extracted by PCI and separated by Urea-PAGE. RNAs were visualized by ethidium bromide stain (see text). Marker snRNAs are indicated on the left of the gel, RNAs on the right. M - marker; C - control (without DNAzyme).

Initial titration experiments showed that the optimal reaction temperature is 30°C and that Mn^{2+}-ions are slightly preferred over Mg^{2+}-ions (Fig. 3.12 C). As reaction products, two RNA species are to be expected, a longer one with a length of 101 nucleotides and a shorter one with a length of 86 nucleotides. Interestingly, both RNA species derived from DNAzyme cleavage reactions show a similar migration behavior in gel electrophoresis as the ones observed after DNA-directed RNaseH cleavage with oligo C (compare Fig. 3.12 C, lane 5, with Fig. 3.3 E, lane 6), providing further evidence for the shorter RNA fragment of the RNaseH reaction to represent the 5' part of U2 snRNA.

In summary, these experiments clearly show that sequence specific cleavage of U2 snRNA by a DNAzyme under native conditions is possible but needs further optimization. Improvement of the cleavage reaction has been shown to be achievable by the concomitant use of manganese ions and EPPS buffer (Santoro and Joyce, 1998). The results presented here could be used as a starting point for future studies to optimize the cleavage reaction, making it suitable for preparative production of Δ5' U2 snRNPs with a defined 5' end in the snRNA.

3.2   Splicing Factor 3b (SF3b)

Human splicing factors 3b (SF3b) and 3a (SF3a) associate with the 12S U2 snRNP to form a 17S particle. 17S U2 snRNP represents the active form of U2 snRNPs that binds to the pre-mRNA during spliceosome assembly. SF3b particles have been isolated biochemically and their protein content has been characterized (Will et al., 2002 and references therin). They consist of 7 core proteins, SF3b155, SF3b145, SF3b130, SF3b49, SF3b14a (previously known as p14), SF3b14b and SF3b10 and a loosely associated factor, a DExH-box helicase domain containing protein, SF3b125. SF3b14a is important for the U2 snRNP to recognize the branch site adenosine and can be crosslinked to this adenosine during the entire splicing process (Query et al., 1996; Will et al., 2001). A cryo-electron microscopy structure showed that the SF3b particle consists of a globular shell, with several protuberances pointing outwards, and a core structure attached to the inner wall of the shell (Golas et al., 2003). A model of SF3b14a bound to a short fragment of SF3b155, derived from X-ray crystal structure analysis, could be placed into this core density of the EM model (Golas et al., 2003; Schellenberg et al., 2006). Furthermore, HEAT-repeats present in the C-terminal portion of SF3b155 were assigned to ladder-like structures around the hollow body of the EM model (Golas et al., 2003). Finally, the RRM domains of SF3b49 were located to protuberances on the surface of the shell (Golas et al., 2003). To gain more insights into the structural details of SF3b particles, we set out to establish a purification protocol to obtain samples for crystallization trials. The first section deals with the purification and crystallization tests of natively isolated, human SF3b particles, whereas the second section copes with the characterization of a yet unknown protein complex, which co-purifies with human SF3b particles.

3.2.1 Purification and crystallization trials of native SF3b particles

A protocol for the purification of SF3b particles via affinity chromatography using an α-SF3b155 antibody-coupled Sepharose column has already been established in our lab (Will et al., 2001). This method includes a final glycerol gradient run, which separates contaminating proteins from SF3b particles. Since high concentrations of glycerol should be avoided in samples for crystallization, a mild alternative may be to use size exclusion chromatography (SEC) instead. SEC has several advantages; (i) it is a powerful method to separate particles with different sizes/shapes; (ii) the samples will elute in a defined buffer, thereby increasing reproducibility; and (iii) it will give an idea about the monodispersity of the sample.

At high salt concentration (600 mM), SF3b particles dissociate from U2 snRNPs, therefore this ionic strength has been used during all previous purification steps (Will and Lührmann, 2006). To check whether SF3b particles survive a size exclusion chromatographic step, they were passed over an analytical SEC column equilibrated in high salt buffer and peak fractions were analyzed for their protein content. As shown in figure 3.13 (A), two peaks were observed during a high salt run. SDS-PAGE analysis revealed that the first peak corresponds to SF3b and the second to a particle of unknown identity (referred to as particle X in the following). To test if SF3b particles would survive low salt conditions (150 mM), the SEC experiment was repeated with a low salt buffer. Again two peaks appeared during gel filtration of which the first corresponded to SF3b particles and the second to particle X (Fig. 3.13 B). Hence, SF3b particles are stable during gel filtration and survive a low salt treatment.
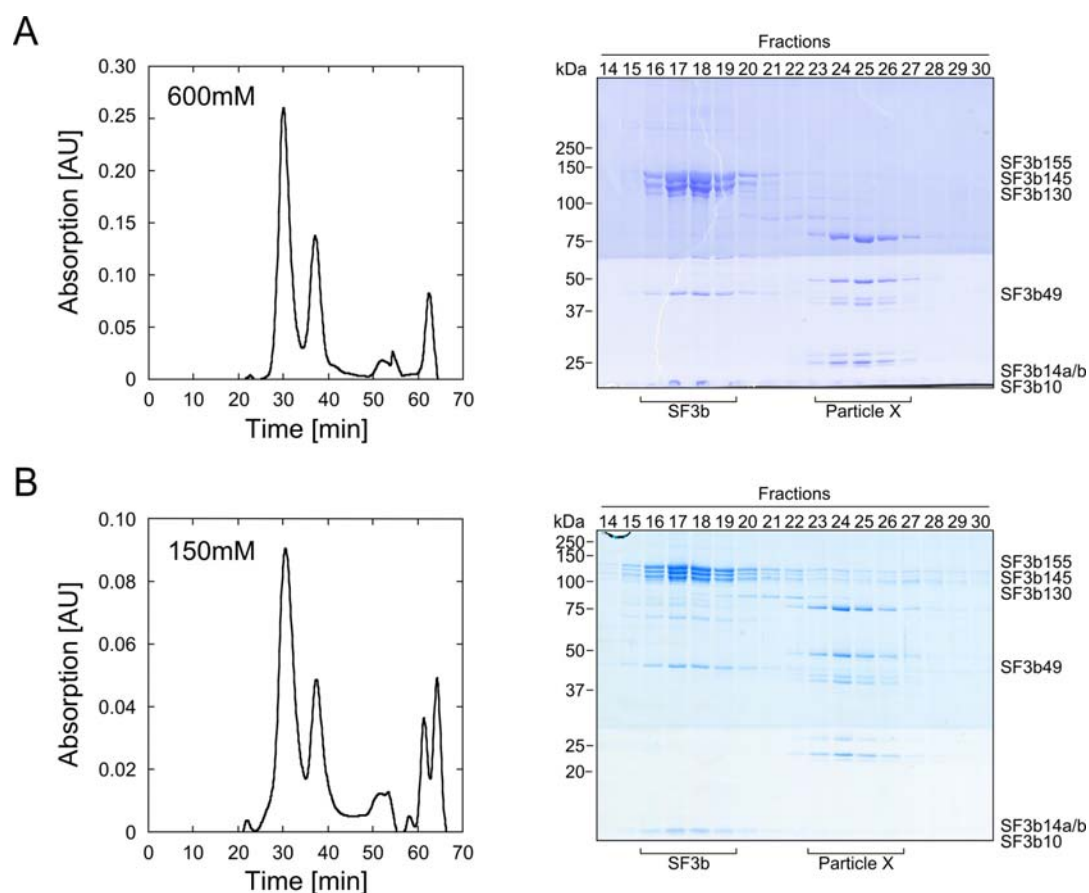
Fig. 3.13: Analytical SEC of SF3b particles at high and low salt conditions. Affinity-purified SF3b particles were applied on an analytical Superose 6 PC3.2 column and run in either high-salt buffer (20 mM Tris-HCl (pH 7.9), 600 mM NaCl, 1 mM DTT) **(A)** or low-salt buffer (20 mM Tris-HCl (pH 7.9), 150 mM NaCl, 1 mM DTT) **(B)**. SEC elution profiles, monitored by absorption at 280 nm, are shown on the left. Analyses of peak fractions by SDS-PAGE with subsequent Coomassie stain are shown on the right. Protein identities are indicated on the right and size markers on the left of each gel. The distributions of SF3b and particle X are denoted below each gel. Note that 1/3 of the input for the high-salt run was used for the low-salt run.

As already described in section 3.1.2, ultracentrifugation is a powerful tool to increase the concentration of biological samples. Using ultracentrifugation, SF3b could be concentrated to more than 12 mg/ml, suitable for crystallization trials. The final purification protocol of SF3b particles included preparation of nuclear extract, two affinity chromatographic steps using antibody-coupled matrices, size exclusion chromatography and two ultracentrifugation steps (one prior and one subsequent to SEC). The SDS-PAGE analyses of the whole purification procedure are illustrated in figure 3.14.
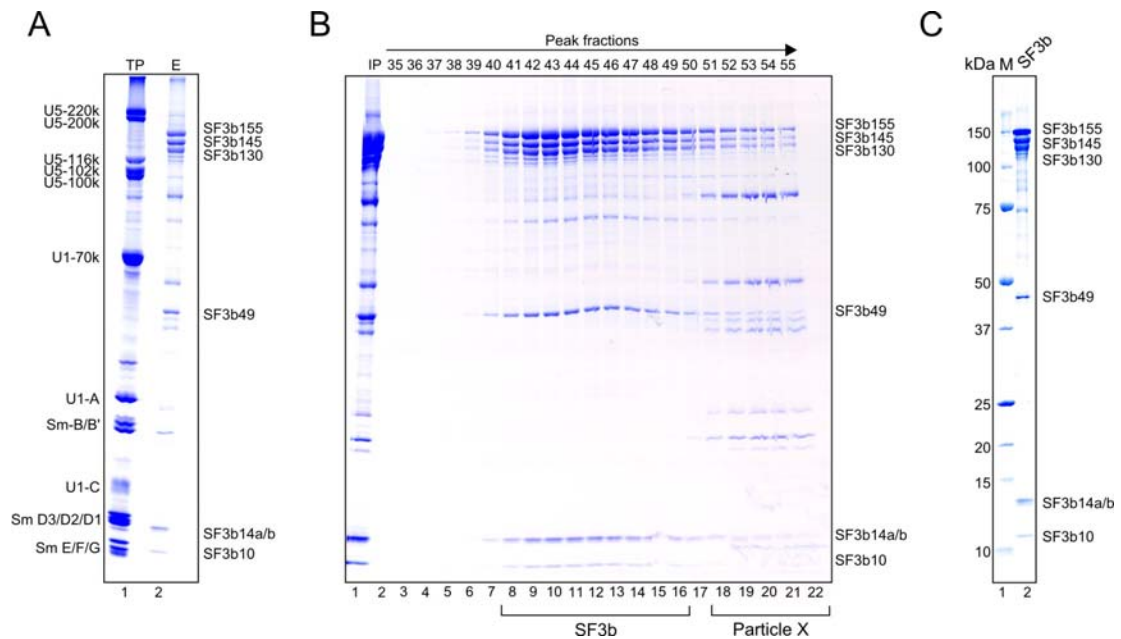
Fig. 3.14: SF3b particle preparation for crystallization. The purity of SF3b particles during the purification procedure was followed by SDS-PAGE. **(A)** Purity of affinity-purified SF3b particles (lane 2). **(B)** A preparative SEC run using a Superose 6 column separates SF3b (lanes 8-16) from particle X (lanes 18-22). **(C)** Purity of the final SF3b sample after pelleting (lane 2), which was used for crystallization. Identities of SF3b components are shown on the right of each gel; size markers on the left. Proteins were stained by Coomassie blue. TP - total proteins (spliceosomal protein mix); IP - input; M - marker.

Samples of SF3b particles eluted from the α-SF3b155 antibody-coupled Sepharose column (Fig. 3.14 A), samples of SEC peak fractions (Fig. 3.14 B) and a sample of purified and concentrated SF3b particles (Fig. 3.14 C) revealed the presence of all 7 SF3b components. A minor amount of contaminating protein with an apparent molecular weight of ~74 kDa could not be removed during all purification steps. 576 crystallization trials were set up with a final SF3b particle concentration of ~9 mg/ml at 4 °C, but neither crystals nor crystalline material were found during inspection of the crystallization plates.

3.2.2 A yet unknown particle co-elutes with human SF3b complexes from α-SF3b155 antibody columns

During the purification of SF3b particles, a yet unknown complex has been isolated by gel filtration (Fig. 3.14 B). Six prominent non-SF3b proteins are visualized in Coomassie stained SDS-polyacrylamide gels (Fig. 3.15 A). To identify

proteins forming this complex, the bands with apparent molecular weights of ~85 (band 1), ~54 (band 2), ~45 (band 3), ~44 (band 4), ~29 (band 5) and ~26 kDa (band 6) were analyzed by mass spectrometry. A compilation of all peptides found in prominent bands is shown in Appendix 6.2.
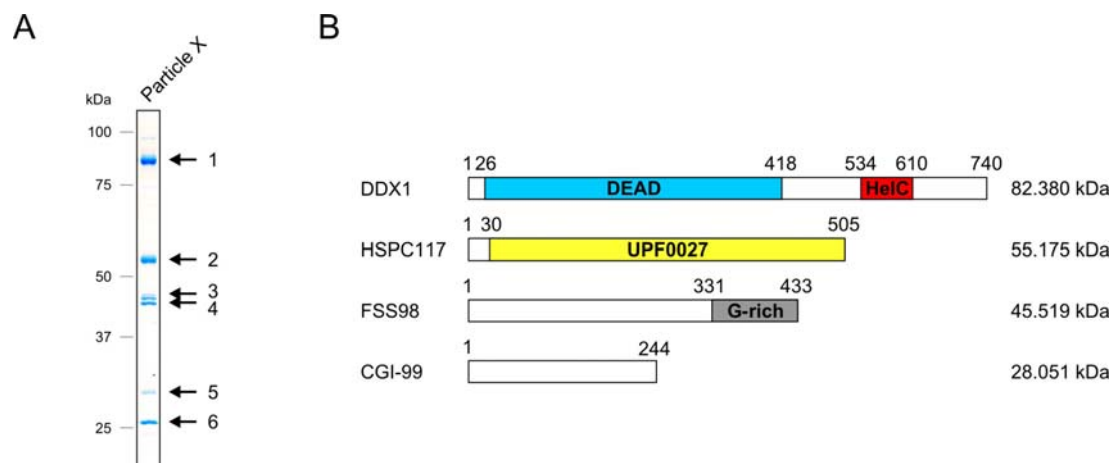
Fig. 3.15: Identification of protein components of particle X by mass spectrometry. (A) Particle X components were separated on a SDS-PAA gel. Bands 1-6 were excised from the gel and analyzed by mass spectrometry via their peptide mass fingerprint (see text). **(B)** Domain structures of particle X proteins. Calculated molecular weigths for each polypeptide are given on the right. Domains, identified by primary sequence inspection (SMART server (http://smart.embl-heidelberg.de/) (Letunic et al., 2006; Schultz et al., 1998) and Pfam (http://pfam.sanger.ac.uk/) (Finn et al., 2006)), are indicated. Protein names are shown on the left. Numbers above each box indicate amino acids. DEAD - (Asp-Glu-Ala-Asp)-box motif; HelC - Helicase superfamily C-terminal domain; UPF - uncharacterized protein family ;G-rich - glycine rich.

Interestingly, band 1 corresponds to DDX1, a protein of the DEAD-box family of RNA helicases with a calculated molecular mass of 82.380 kDa. Peptides derived from band 2 identified the protein as HSPC117, with a predicted UPF0027 domain and a calculated molecular mass of 55.175 kDa. Bands 3 and 4 both correspond to one protein, which is known as Family with Sequence Specificity 98/member B (FSS98B; calculated molecular mass of 45.519 kDa). This protein is only characterized by a glycine-rich region found in the very C-terminal part of the protein. Band 6 is the protein CGI-99, which has a calculated mass of 28.051 kDa. The protein belonging to band 5 seems to be substoichiometrically associated with the complex and is a hypothetical protein (LOC79074).

To get an idea about the shape of the isolated complexes, a sample was run on a GraFix gradient (Kastner et al., 2008) and fixed complexes were visualized by

negative-stain electron microscopy. Although the sample density was very low, some particle-like structures could be observed (Fig. 3.16). The findings suggest that particle X is a tetrameric complex, probably suitable for structural studies by electron microscopy or X-ray crystallography.
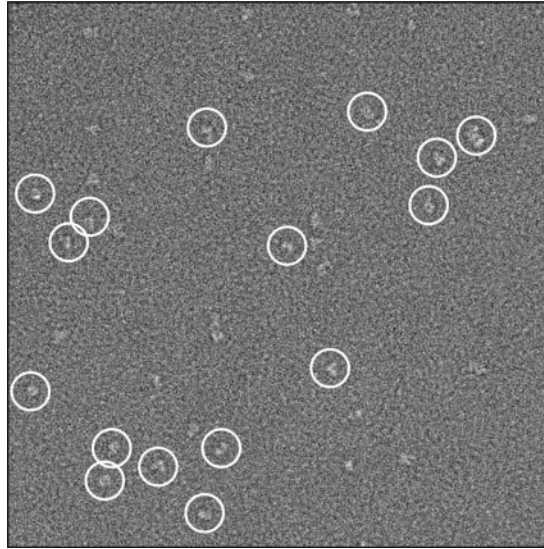
Fig. 3.16: Electron micrograph of particle X. A particle X sample obtained from gel filtration was pelleted by ultracentrifugation to increase the concentration of the particles and was run on a standard GraFix gradient (Kastner et al., 2008). Particles of the peak fraction were adsorbed onto a single-layered carbon film, sandwiched beneath another carbon film and stained by 2 % [w/v] uranyl-formate. White circles mark presumed particle X complexes. Image was taken by F. Hauer.

3.3   Pre-mRNA retention and splicing complex (RES)

The protein composition of SF3b complexes from the budding yeast *Saccaromyces cerevisiae* was recently described (Wang et al., 2005). Besides known yeast orthologs like Hsh155p (SF3b155), Cus1p (SF3b145), Rse1p (SF3b130) and Hsh49p (SF3b49), the proteins Snu17p (Ist3p), Rcp10 (Ysf3p), and Bud31p have been identified in tandem affinity purified particles (Wang et al., 2005; Wang and Rymond, 2003). Snu17p was also shown to coexist in a trimeric complex with Pml1p and Bud13p (Dziembowski et al., 2004), interacting with Bud13p (Uetz et al., 2000). Since deletion experiments showed that removal of proteins Snu17p and Bud13p had an impact on splicing efficiency and that removal of Pml1p caused pre-mRNA leakage from the nucleus, this complex was termed pre-mRNA retention and splicing complex (RES) (Dziembowski et al., 2004). Although nonessential, Snu17p was thought to be the yeast ortholog of the branch point interacting protein SF3b14a/p14. Consistent with this idea, deletion of the *SNU17* gene resulted in the formation of stalled spliceosomes, harboring all five snRNAs (Gottschalk et al., 2001). To gain more information about Snu17p and its interaction partners, I set out to characterize RES complexes by biophysical and biochemical methods, with the ultimate goal to elucidate the molecular architecture of the RES complex. The following sections deal with the heterologous production of recombinant RES complex components, the characterization of interacting regions among the three proteins, and their structural classification.

3.3.1   *In vitro* reconstitution of RES complexes

Several strategies to produce protein complexes heterologously in *E. coli* are illustrated in figure 3.17. Co-expression has been shown to be a powerful tool to produce complexes in *E. coli* when individual gene expression fails or leads to the production of insoluble protein (Stebbins et al., 1999)(Fig. 3.17 A). If possible, genes can be expressed individually and crude extracts containing protein products can be mixed afterwards to assemble and isolate complexes (Fig. 3.17 B). Last, individually

produced and purified proteins can be used to investigate specific associations among complex components in binding assays (Fig. 3.17 C).
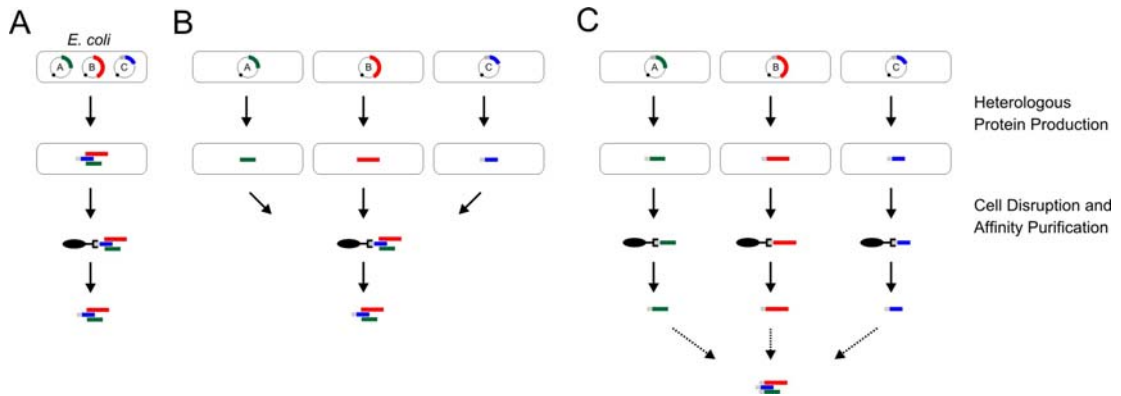
Fig. 3.17: Heterologous protein production strategies of RES complexes and subcomplexes. **(A)** *E. coli* cells are transformed with plasmids coding for complex components A (green), B (red) and a tagged version of C (grey and blue). Cells harboring all three plasmids are isolated on selective media. Co-expression leads to protein production and proteins assemble to complexes within a cell. After cell disruption protein complexes can be isolated by an affinity step via the tag. **(B)** Genes are expressed individually. Before disruption, cells are combined and protein complexes can be isolated by an affinity step (as in (A)). **(C)** *E. coli* cells are transformed with plasmids coding for tagged versions of complex components A (green), B (red) and C (blue) (tag in grey). Genes are expressed individually and proteins are purified individually. Complex assembly can be tested by sequential addition of each component (dashed arrows).

To test whether recombinant RES components, Snu17p, Pml1p and Bud13p, assemble *in vitro*, corresponding genes were cloned from yeast chromosomal DNA into a vector series utilizable for co-expression. A N-terminal His$_6$-tag was engineered for Snu17p. Using affinity chromatography via Ni^{2+}-NTA agarose, ion exchange chromatography and gel filtration, a stable trimeric complex consisting of His$_6$-tagged Snu17p, Pml1p and Bud13p can be isolated (Fig. 3.18).

The complexe is stable under intermediate (250 mM NaCl) and high salt (800 mM NaCl) conditions (Fig. 3.19, B and D, respectively) and seems to be generally unaffected by low (6.0) and high (9.0) pH (Fig. 3.19, A and C, respectively). Gel electrophoretic analysis suggests that this heterologously produced complex is the same as the TAP-tag isolated complex, which was described previously (Dziembowski et al., 2004).
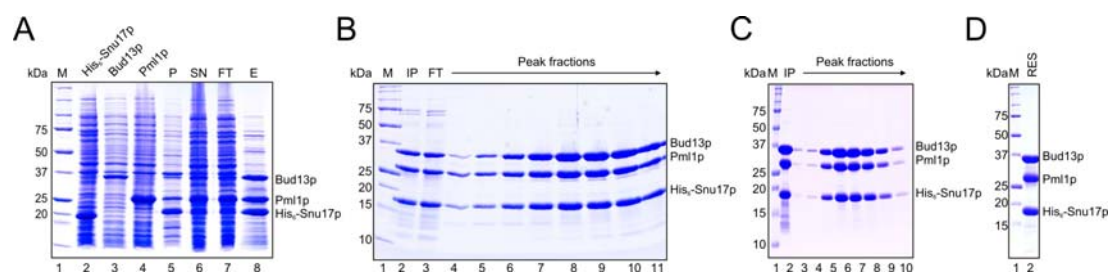
Fig. 3.18: Purification of full-length RES complexes. (A) Indicated proteins were produced individually (lanes 2-4) in *E. coli* and cells were combined after harvesting. Unsoluble and soluble fractions (lanes 5 and 6, respectively) as well as samples from the flow-through (lane 7) and eluate (lane 8) fraction of the first purification step using Ni^{2+}-NTA agarose are shown. Proteins were analyzed via SDS-PAGE with subsequent Coomassie stain. **(B)** The Ni^{2+}-NTA agarose eluate was buffer-exchanged and loaded on CM-Sepharose. RES complexes were eluted by a salt gradient and peak fractions were analyzed on a SDS-PAA gel. **(C)** RES particles from CM-eluate were further purified via SEC and peak fractions were analyzed by SDS-PAGE. **(D)** Protein products from a co-expression approach were similarly purified and show the same protein composition after the final gel filtration step (lane 2). M - marker; P - pellet; SN - supernatant; FT - flow-through; E - eluate.

In a second strategy, proteins His_6-Snu17p, Pml1p and Bud13p were expressed individually. These individually produced proteins assembled into similar complexes with alike biochemical properties (Fig. 3.18). Thus, co-expression and expression of single genes are suitable for the production of RES complexes or proteins that assemble into RES complexes, respectively.
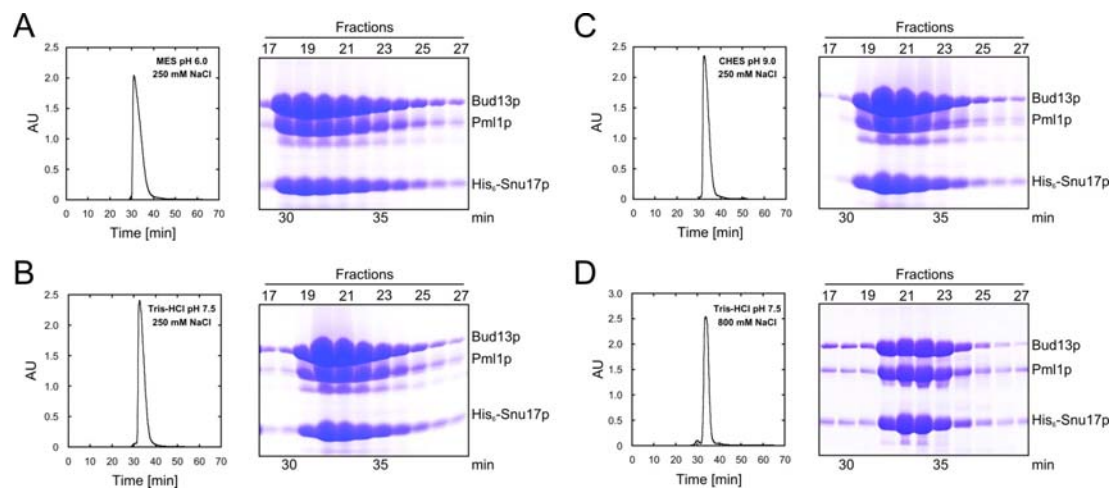
Fig. 3.19: Salt and pH stability of RES complexes. Elution profiles of analytical SEC runs on a Superdex 200 PC3.2 column were monitored by the absorption at 280 nm and are shown on the left of each panel. Peak fractions of each run were analyzed by SDS-PAGE. Gels were stained by Coomassie blue and are shown on the right of each panel. Buffer conditions were as follows: 20 mM MES (pH 6.0), 250 mM NaCl, 1 mM DTT **(A)**; 20 mM Tris-HCl (pH 7.5), 250 mM NaCl, 1 mM DTT **(B)**; 20 mM CHES (pH 9.0), 250 mM NaCl, 1 mM DTT **(C)**; 20 mM Tris-HCl (pH 7.5), 800 mM NaCl, 1 mM DTT **(D)**. AU - absorption units.

3.3.2   Global arrangement of RES proteins within the complex

To derive a global interaction map of the RES complex, binding of each subunit to the respective other two components was tested by GST pull-down experiments. All three proteins were produced either as GST-tagged or His$_6$-tagged fusion proteins. Immobilized GST-Snu17p brought down both Bud13p and Pml1p (Fig. 3.20, lanes 17 and 12), whereas GST-Bud13p and GST-Pml1p only interacted with Snu17p (Fig. 3.20, lanes 8 and 9). No interaction could be observed between Bud13p and Pml1p (Fig. 3.20, lanes 14 and 18).
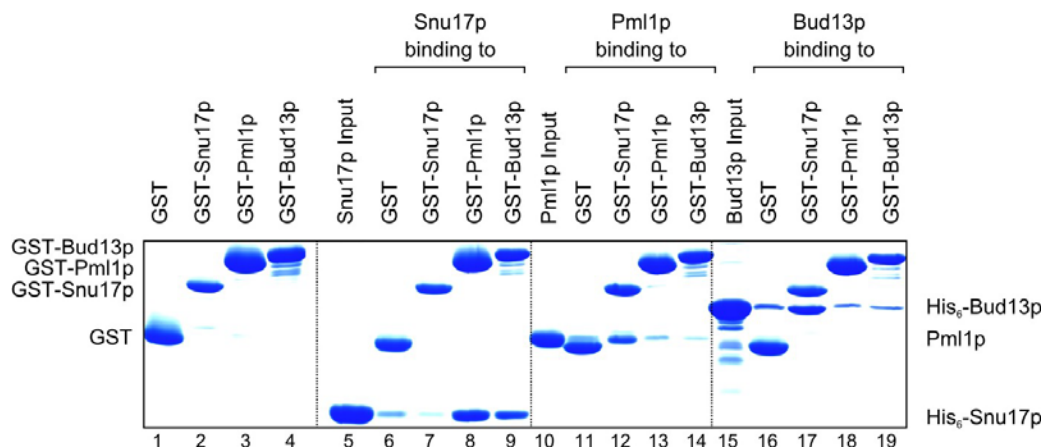
Fig. 3.20: *In vitro* **reconstitution of the RES complex.** For GST pull-down assays with full-length proteins, GST-fusion proteins of Snu17p, Pml1p and Bud13p were produced in *E. coli* and bound to glutathione-Sepharose beads. Pre-purified His$_6$-tagged Snu17p and Bud13p and untagged Pml1p were mixed with GST-fusion protein-coated glutathione-Sepharose and protein fractions bound to the beads were analyzed by SDS-PAGE. Lanes 1-4 represent the bait input controls. Lanes 5, 10, 15 are input controls for the His$_6$-tagged proteins Snu17p, Pml1p and Bud13p, respectively, used as preys. Lanes 6, 11, 16 show controls with GST alone, lanes 7, 12, 17, lanes 8, 13, 18 and lanes 9, 14, 19 show pull-downs with GST-Snu17p, GST-Pml1p, and GST-Bud13p, respectively, used as baits. (taken from Trowitzsch *et al.* 2008).

To confirm this molecular organization, pairwise combinations of the proteins were analyzed by analytical SEC (Fig. 3.21). Snu17p formed hetero-dimers with either Bud13p or Pml1p, which were stable during gel filtration even in the absence of the respective third component (Fig. 3.21 E, D). In line with the GST pull-down experiments, no stable Bud13p-Pml1p hetero-dimer could be isolated (Fig. 3.21 F).

Results
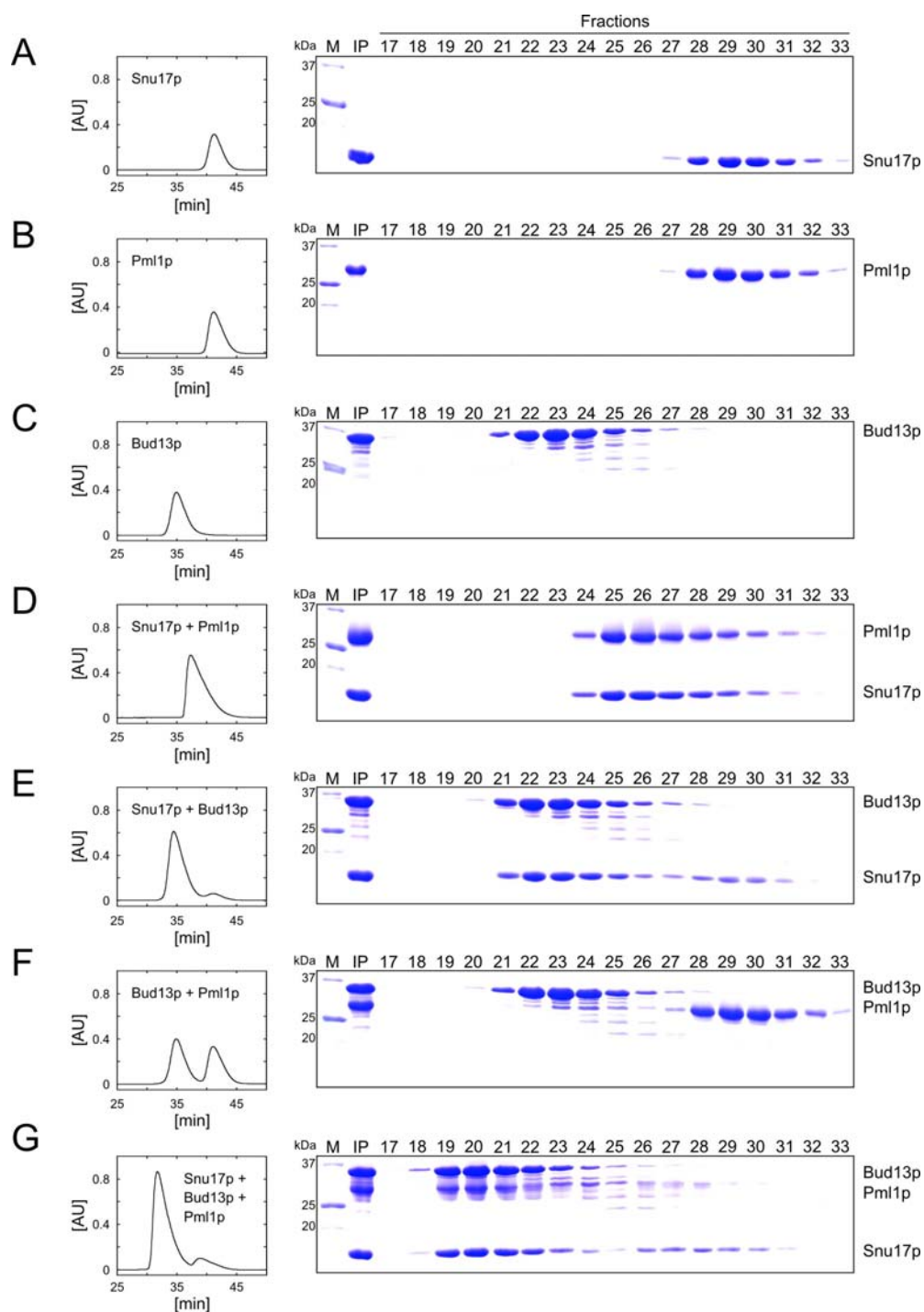
Fig. 3.21: Interaction mapping by gel filtration. Analysis of RES complex assembly by analytical SEC reveals that Snu17p provides a binding platform for Pml1p and Bud13p. Left panels: Elution profiles. Right panels: SDS gel analysis of the elutions. **(A-C)** Gel filtration analyses of isolated His$_6$-tagged Snu17p **(A)**, Pml1p **(B)** and Bud13p **(C)**. **(D-F)** Gel filtration analyses of two-component mixtures Snu17p + Pml1p **(D)**, Snu17p + Bud13p **(E)** and Bud13p + Pml1p **(F)**. **(G)** Gel filtration analysis of the fully assembled RES complex. All components were present at a concentration of 40 µM and were run in a buffer comprising 20 mM Tris-HCl (pH 7.5), 250 mM NaCl, 1 mM DTT. AU - absorption units. (taken from Trowitzsch *et al.* 2008).

3.3.3 Domain organizations of the three RES proteins analyzed by a fluorescence based approach

Examination of the intrinsic structural organization of the three RES complex proteins by primary sequence analysis indicated that Snu17p harbors a central RNA recognition motif (RRM), which is flanked by presumably unstructured regions (Fig. 3.22). Pml1p most probably consists of a C-terminal forkhead-associated domain and an unstructured N-terminal region (Fig. 3.22). In Bud13p, no folded domain is detectable, although highly conserved portions of 10-20 residues are found close to the N- and C-termini of the protein (Fig. 3.22). Consistent with the primary sequence analysis, Snu17p, with a molecular weight of 17.1 kDa, and Pml1p, with a molecular weight of 23.7 kDa, eluted at the expected retention volumes for approximately globular proteins in analytical gel filtration runs (Fig. 3.21 A,B). In contrast, Bud13p (30.5 kDa) showed an abnormal migration behavior and eluted earlier than expected (Fig. 3.21 C).
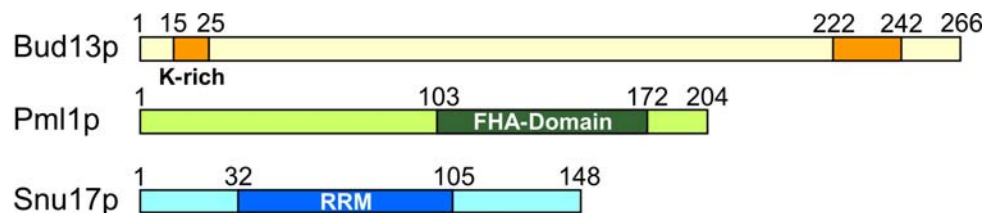
Fig. 3.22: Intrinsic structural organization of RES complex components. The conserved lysine-rich and C-terminal regions (residues 15-25 and 222-242, respectively) of Bud13p are shown in orange. Predicted structured parts of Pml1p (FHA-domain; residues 103-172) and of Snu17p (RRM; residues 32-105) are colored green and blue, respectively, and presumed unstructured regions are indicated in light green (Pml1p) and light blue (Snu17p). Bud13p is predicted to be unstructured throughout. Unstructured regions were predicted by the FoldIndex server (http://bioportal.weizmann.ac.il/fldbin/findex) and structured domains by the SMART server (http://smart.embl-heidelberg.de/).

To experimentally test the presence of folded domains in the RES proteins, a modified fluorescence-based approach was used (Vedadi et al., 2006). Upon increase in temperature, stable three-dimensional folds in proteins exhibit cooperative unfolding. Binding of a fluorescent dye to exposed hydrophobic surfaces of unfolded proteins will change the fluorescent properties of the dye, a process, which can be

monitored. Figure 3.23 shows a plot of fluorescence intensity against temperature for the proteins Snu17p, Pml1p and Bud13p. Snu17p (Fig. 3.23, solid black line) and Pml1p (Fig. 3.23, dashed black line) showed cooperative unfolding upon increase in temperature, whereas Bud13p did not (Fig. 3.23, grey line). These results indicate that Snu17p and Pml1p comprised folded domains and Bud13p lacked regular tertiary structure in solution. The steeper slope of the curve of Pml1p in comparison to Snu17p curve suggests that the protein may be more stably folded than Snu17p.
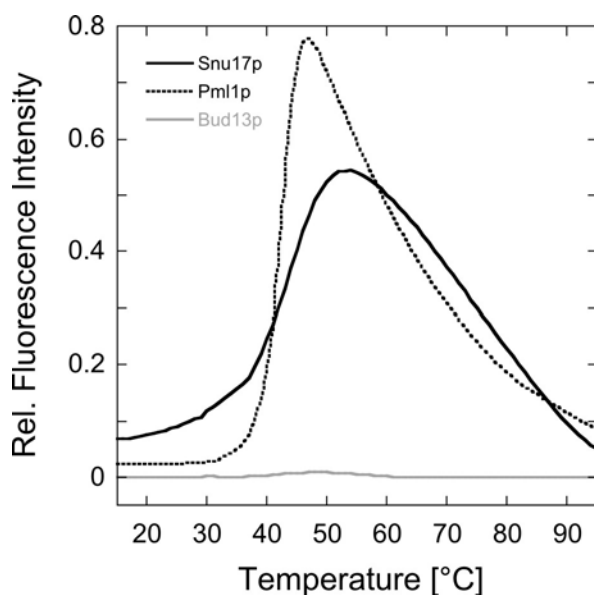
Fig. 3.23: Global fold analysis. Thermal unfolding of full-length Snu17p (solid black line), Pml1p (dashed line) and Bud13p (gray line) monitored by a fluorescence-based approach. Snu17p and Pml1p exhibit cooperative transitions, indicative of folded domains. Bud13p seems to be completey unfolded, since no transition is recognizable (taken from Trowitzsch *et al*. 2008).

3.3.4   Binding regions between RES components

Snu17p represents the central binding platform for Pml1p and Bud13p (section 3.3.2). To delineate the minimal regions of Snu17p sufficient for binding of Bud13p and Pml1p, several His_6-tagged N- and C-terminal truncation mutants were generated (Fig. 3.24 A). Importantly, none of the deletions corrupted the predicted core RRM domain, which spans residues 25-106. Binding of the Snu17p deletion mutants to GST-Pml1p and GST-Bud13p in pull-down experiments showed that N-terminal truncations of Snu17p (Snu17p[9-148] and Snu17p[25-148]) neither affected binding to GST-Bud13p (Fig. 3.24 B, lanes 8, 11), nor to GST-Pml1p (Fig. 3.24 C,

lanes 8, 11). Short truncations from the C-terminus of Snu17p (Snu17p$^{1-138}$, Snu17p$^{1-125}$ and Snu17p$^{1-106}$) did not abolish binding to Bud13p (Fig. 3.24 C, lanes 14, 23, 32). However, binding to GST-Pml1p was completely abolished when more than ten residues were deleted from the C-terminus of Snu17p (Snu17p$^{1-125}$ and Snu17p$^{1-106}$)(Fig. 3.24 C, lanes 23, 32). These data demonstrate that the core RRM of Snu17p (Snu17p$^{25-106}$) is sufficient to form a complex with Bud13p but that a C-terminal expansion beyond the core RRM (Snu17p$^{25-138}$) is required for Pml1p binding.
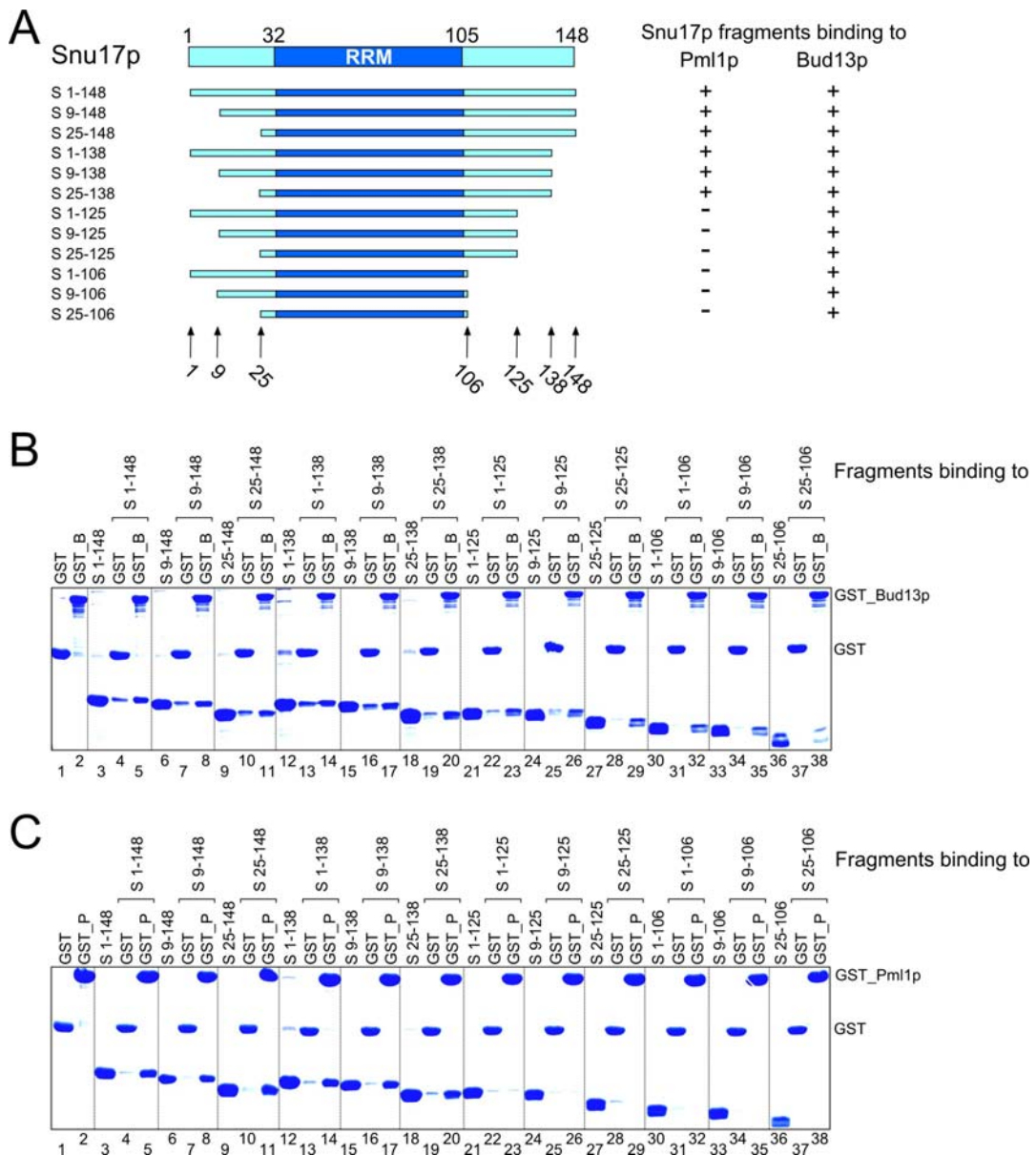
See following page for figure description ...

Fig. 3.24: Binding regions of Snu17p for Bud13p and Pml1p. **(A)** Schematic representation of Snu17p and of Snu17p fragments engineered for GST pull-down assays. Start and end amino acid residues of the Snu17p fragments are designated on the left. Predicted unstructured regions and the central RRM are colored light blue and blue, respectively. Interactions of the Snu17p fragments with either Pml1p (left) or Bud13p (right) are indicated by plus (binding) and minus (no binding) on the right. **(B)** Pull-down of His_6-tagged Snu17p fragments (A) with glutathione-Sepharose beads pre-coated with recombinant GST or GST-Pml1p fusion protein shows that Snu17p binding is dependent on the presence of a C-terminal expansion of the core RRM. Input and bound fractions were analyzed by SDS-PAGE and stained with Coomassie. Lanes 1 and 2 - GST input controls; lanes 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36; lanes 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37; and lanes 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38 - His_6-tagged inputs used as preys, the GST control pull-downs and pull-downs with GST-Bud13p, respectively. **(C)** Pull-down of His_6-tagged Snu17p fragments as in (B) but using glutathione-Sepharose beads pre-coated with recombinant GST or GST-Pml1p fusion protein. (taken from Trowitzsch *et al.* 2008).

3.3.5   A conserved C-terminal motif of Bud13p interacts with Snu17p

Multiple sequence alignments of putative Bud13p orthologs revealed two highly conserved regions (Fig. 3.22 and Fig. 3.25 A). The first, at the very N-terminus of the protein, contains a high fraction of lysine residues. The second region comprises about 20 residues close to the C-terminus (Fig. 3.25 A). To test whether either of the conserved regions in Bud13p constitute an interaction element for Snu17p, various N- and C-terminally truncated fragments of Bud13p were generated and inspected for their binding to immobilized GST-Snu17p fusion protein by GST pull-down experiments (Fig. 3.25, B, C). N-terminal truncations of Bud13p, which removed the lysine-rich conserved element, had no influence on binding to GST-Snu17p (Fig. 3.25 C, lanes 20, 23, 26, 29, 32, 35). In contrast, none of the Bud13p fragments that lacked the C-terminal 65 residues could be precipitated by immobilized GST-Snu17p fusion protein (Fig. 3.25 C, lanes 5, 8, 11, 14). The very C-terminal fragment of Bud13p alone ($Bud13p^{202-266}$) was sufficient to sustain binding to GST-Snu17p (Fig. 3.25 C, lane 35). These results demonstrate that the Bud13p C-terminus encompassing the second highly conserved region is involved in the interaction with Snu17p.
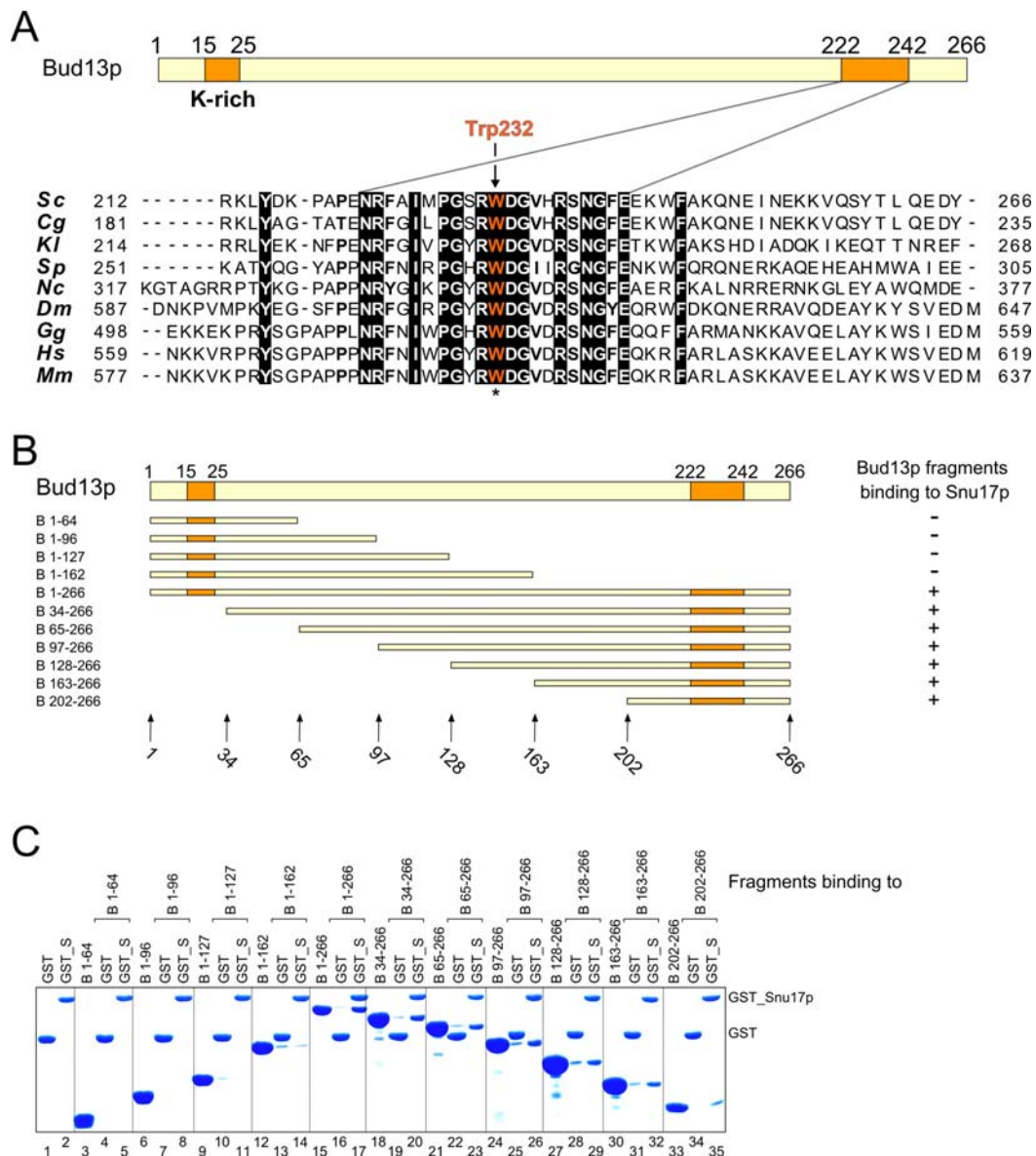
Fig. 3.25: The Snu17p-binding region of Bud13p. (A) Upper panel - schematic representation of Bud13p with conserved regions depicted as range boxes. Lower panel - multiple sequence alignment of the C-terminal region of *Saccharomyces cerevisiae* (*Sc*) Bud13p with orthologs from *Candida glabrata* (*Cg*), *Klyveris lactis* (*Kl*), *Schizosaccharomyces pombe* (*Sp*), *Neurospora crassa* (*Nc*), *Drosophila melanogaster* (*Dm*), *Gallus gallus* (*Gg*), *Homo sapiens* (*Hs*) and *Mus musculus* (*Mm*) shows an Arg-Trp-Asp-Gly motif with a central Trp present in all sequences analyzed. Identical residues are boxed. Numbers on the left and right of the alignment indicate the start and end residues of the sequences. The alignment was generated using the ClustalW algorithm (http://www.ebi.ac.uk/Tools/clustaw/) (Thompson et al., 1994). **(B)** Schematic representation of Bud13p fragments engineered for GST pull-down assays with start and end amino acids designated on the left. Interactions of the Bud13p fragments with GST-Snu17p are indicated by plus (binding) and minus (no binding) on the right. B - Bud13p. **(C)** Pull-down of His$_6$-tagged Bud13p fragments (A) with glutathione-Sepharose beads pre-coated with recombinant GST or GST-Snu17p fusion protein shows that Bud13p binding is dependent on the presence of the C-terminal portion (residues 202-266) harboring the highly conserved C-terminal region. Experimental conditions were as in figure 3.24. B - Bud13p; S - Snu17p. (taken from Trowitzsch *et al.* 2008).

3.3.6   Identification of a Snu17p binding epitope in Bud13p via a mass spectrometry based approach

Mass spectrometry is a powerful tool to identify peptides in mixtures or in isolation (Fenn, 1993; Karas and Hillenkamp, 1988; Tanaka et al., 1988). GST pull-down experiments suggested that Snu17p binds the C-terminal portion of Bud13p (residues 202-266). To verify this result, we combined proteolysis, size-exclusion chromatography and mass spectrometry as illustrated in figure 3.26.
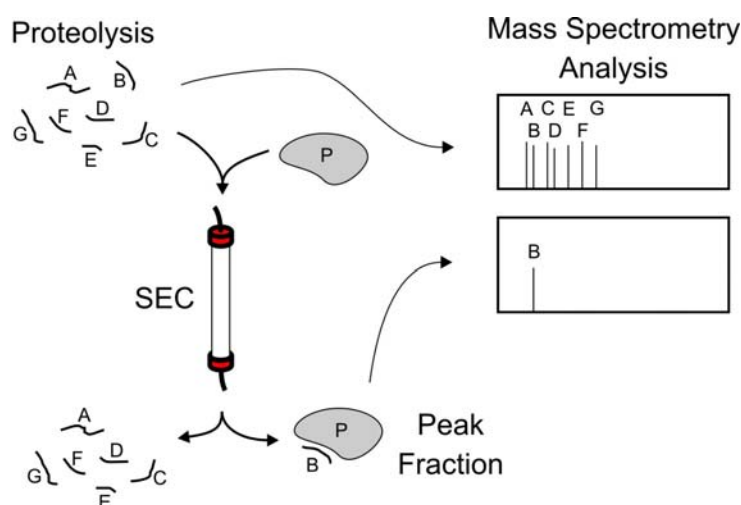
Fig. 3.26: Schematic illustration of a mass spectrometric approach to identify Snu17p peptide ligands of Bud13p. In brief, both interaction partners of interest are available in purified form. One component is entirely proteolyzed (upper left) and fragments are identified by mass spectrometry in a control experiment. After incubation of peptides with the target protein P, the protein-peptide mixture is separated by size on an analytical gel filtration column. Mass spectrometric analyses of peak fractions, which contain the target molecules P, identify bound peptides.

A fragment of Bud13p (residues 125-266) was entirely cleaved by the addition of Glu-C endoprotease and resulting peptides were analyzed by mass spectrometry. *m/z* values could be assigned to Bud13p peptides [192]DPAITFTHDKE[202], [203]RTVKTSLLGRKLYDKPAPE[221], [222]NRFAIMPGSRWDGVHRSNGFE[242], [222]NRFAIMPGSRWDGVHRSNGFEE[243], [244]KWFAKQNE[251], [255]KKVQSYTLQE[264] and [255]KKVQSYTLQEDY[266] (Table 3.1). Peptides covering the N-terminal region (residues 125-191) could not be identified. The pool of Bud13p peptides was incubated in the presence of Snu17p and protease inhibitors and proteins/peptides were subsequently separated on an analytical SEC column. Mass spectrometric

analyses of Snu17p-containing peak fractions revealed the presence of peptide [222]NRFAIMPGSRWDGVHRSNGFEE[243] of Bud13p, whereas no *m/z* values fitting to other Bud13p peptides were found. Since peptide [222]NRFAIMPGSRWDGV-HRSNGFEE[243] of Bud13p can be isolated from the peptide pool together with Snu17p via gel filtration, it seems to interact specifically with Snu17p. Significantly, isolation of a specific tryptic Bud13p fragment failed. Trypsin cleavage would occur after Arg223 and Arg231 of Bud13p. Thus, further fragmentation of the Bud13p[222-243] peptide may have destructed a potential Snu17p-binding motif of Bud13p. This technique may be developed into a general method with use of a set of proteases. Especially interactions with unfolded proteins, likely to present short linear epitopes, may be probed by this method.

Table 3.1: Compilation of Bud13p peptides. Monoisotopic peptide masses (Mass), found in a control experiment of Glu-C-treated Bud13p[125-266] (peptide pool), are compared with masses of co-isolated peptides, observed in peak fractions containing Snu17p after gel filtration (co-isolated).

Peptide	Mass [M+H]$^+$	peptide pool	co-isolated
[126]AQGHKIQE[133]	910.4741	-	-
[134]DSKIDDSSFSRSKYE[148]	1763.8082	-	-
[149]DE[150]	263.0874	-	-
[151]KAAE[154]	418.2296	-	-
[155]RE[156]	304.1615	-	-
[157]QYLKNLNMGDVQKLGINVDAHD-KKKNQTASSLTIE[191]	3914.0440	-	-
[192]DPAITFTHDKE[202]	1273.6059	yes	-
[203]RTVKTSLLGRKLYDKPAPE[221]	2172.2499	yes	-
[222]NRFAIMPGSRWD[233]	1449.7055	yes	-
[222]NRFAIMPGSRWDGVHRSNGFE[242]	2433.1629	yes	-
[222]NRFAIMPGSRWDGVHRSNGFEE[243]	2562.2054	yes	yes
[244]KWFAKQNE[251]	1050.5367	yes	-
[252]INE[254]	375.1874	-	-
[255]KKVQSYTLQE[264]	1223.6630	yes	-
[255]KKVQSYTLQEDY[266]	1501.7533	yes	-

3.3.7 Mutational analysis of the Snu17p/Bud13p interaction

Among the highly conserved residues at the C-terminus of Bud13p, a central Trp (Trp232) stands out (Fig. 3.27 A). Similar conserved tryptophans are found in UHM ligand motifs (ULMs; Fig. 3.27 A), where they are crucial for the interaction with U2AF homology motifs (UHMs) (Kielkopf et al., 2004). To test whether the Bud13p[202-266] fragment might act as an ULM, using its Trp-containing motif to associate with Snu17p, we mutated Trp232 of Bud13p to alanine and inspected

hetero-dimer formation in mixtures of Snu17p with either excess of the wild type (wt) Bud13p C-terminal fragment (Bud13p[202-266]) or the mutated fragment (Bud13p[202-266][Trp232Ala]) by analytical gel filtration. Whereas a stable interaction was observed between wt Bud13p[202-266] and Snu17p, the Trp232Ala substitution completely abrogated binding of the fragment to Snu17p (Fig. 3.27 B). Thus, Trp232 of Bud13p represents a critical latching point for the interaction with Snu17p, reminiscent of UHM-ULM interactions.
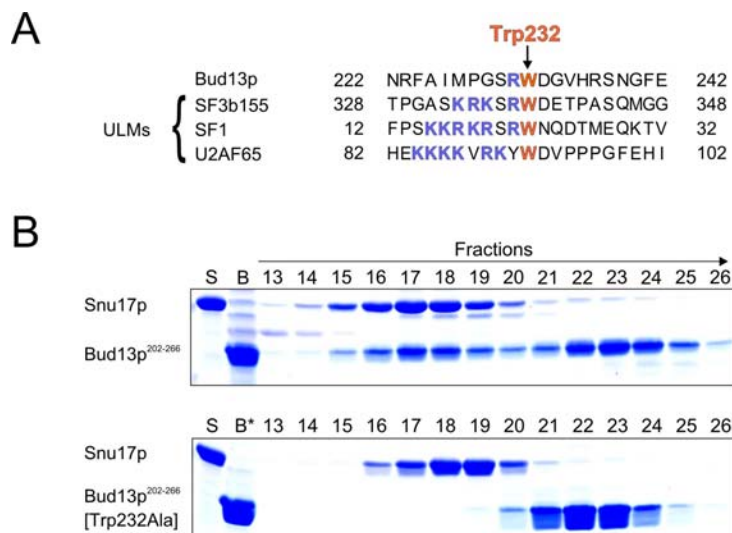
Fig. 3.27: Role of Trp232 in Bud13p for binding to Snu17p. (A) Multiple sequence alignment of the conserved C-terminal portion of Bud13p with ULMs from SF3b155 (ULM5), SF1 and U2AF[65]. Trp232 of Bud13p and the ULM tryptophans are in orange, preceding basic residues in blue. **(B)** Gel filtration analysis shows that a dimeric complex forms upon mixing Snu17p and Bud13p[202-266] fragments (upper panel; fractions 15-20). In contrast, Bud13p[202-266][Trp232Ala] does not co-migrate with Snu17p during gel filtration (lower panel), indicating a crucial role of Trp232 in the Snu17p-Bud13p interaction. Protein-containing fractions were analyzed by SDS-PAGE and proteins were visualized by Coomassie staining. Input controls before mixing are on the left. S - Snu17p; B - Bud13p[202-266]; B* - Bud13p[202-266][Trp232Ala]. (taken from Trowitzsch *et al.* 2008).

3.3.8 Intrinsic tryptophan fluorescence shows that Trp232 of Bud13p binds to a hydrophobic pocket of Snu17p

Known UHMs sequester the Trp residue of the bound ULMs in a hydrophobic pocket between their two α-helices (Corsini et al., 2007; Kielkopf et al., 2004). In order to test whether Snu17p positions Trp232 of Bud13p in a similar environment, we monitored the intrinsic Trp fluorescence of chemically synthesized Bud13p peptide encompassing residues 222-242 upon interaction with the core RRM of

Snu17p (Snu17p$^{25-106}$). Snu17p$^{25-106}$ does not contain a Trp residue, so Trp232 of Bud13p can be selectively monitored. Unbound Bud13p$^{222-242}$ showed an emission maximum at 355 nm, which shifted to 345 nm upon addition of equimolar amounts of Snu17p$^{25-106}$ (Fig. 3.28 A). In addition to the blue-shift, the fluorescence quantum yield increased upon complex formation (Fig. 3.28 A). The spectral changes are consistent with the transfer of Bud13p Trp232 from a polar (aqueous) to a non-polar (proteinaceous) environment, again supporting an UHM-ULM-type binding mode between Snu17p and Bud13p.

To derive the dissociation constant for Snu17p$^{25-106}$/Bud13p$^{222-242}$ complexes, Snu17p$^{25-106}$ was titrated into a Bud13p$^{222-242}$ solution and the change in intensity of fluorescence emission was monitored at a constant wavelength of 345 nm. The fluorescence intensity change was plotted against the concentration of Snu17p$^{25-106}$ (Fig. 3.28 B). From the titration experiment, a dissociation constant (K_d) of 137 nM could be deduced (Fig. 3.28 B), which is in the range of known dissociation constants described for UHM-ULM interactions (Corsini et al., 2007; Kielkopf et al., 2004; Kielkopf et al., 2001; Selenko et al., 2003).
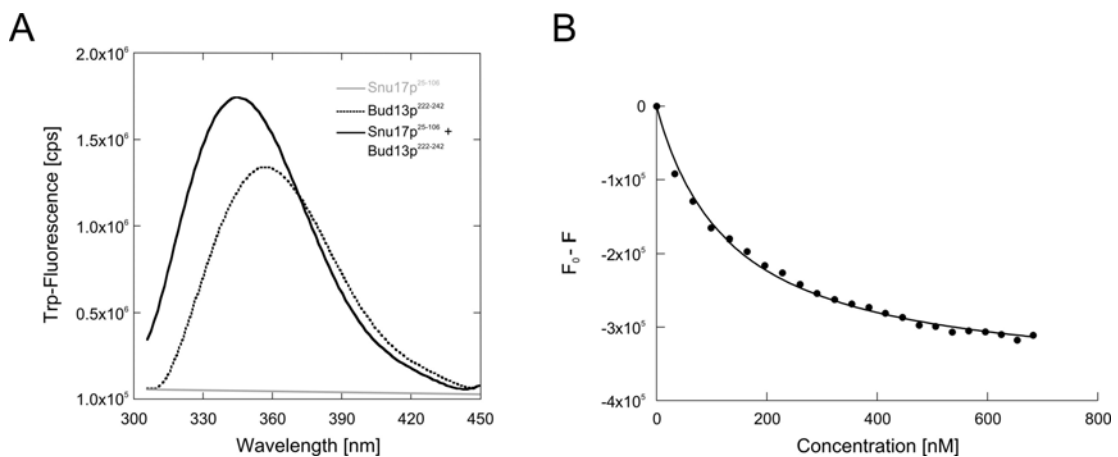
Fig. 3.28: Binding of Snu17p and Bud13p analyzed by intrinsic tryptophan fluorescence (A) Intrinsic Trp fluorescence emission spectra of unbound Bud13p$^{222-242}$ (dashed curve) or bound to Snu17p$^{25-106}$ (black curve) suggest burial of Bud13p Trp232 in a hydrophobic pocket upon complex formation. Snu17p$^{25-106}$ does not contain Trp (gray curve). All spectra were corrected against a buffer spectrum. (Panel A taken from Trowitzsch *et al.* 2008). **(B)** Change of tryptophan fluorescence of Bud13p$^{222-242}$ peptide upon titration of Snu17p$^{25-106}$ was monitored at constant wavelength (345 nm) and plotted versus titrant concentration. Analysis of the data using GraphPad Prism$^{®}$ software (http://www.graphpad.com/prism/) allowed calculation of the dissociation constant of the Snu17p$^{25-106}$/Bud13p$^{222-242}$ dimer (see "Materials and Methods").

3.3.9   The N-terminus of Pml1p is required for Snu17p binding

To address the question, which part of Pml1p mediates interaction with Snu17p, various N- and C-terminal truncation mutants of Pml1p were engineered (Fig. 3.29 A) and tested for their binding to recombinant GST-Snu17p in pull-down assays. Some truncations that cut into the predicted C-terminal FHA domain of Pml1p were partially insoluble (Fig. 3.29 B), suggesting that in these fragments the proper fold of the FHA domain was corrupted. Irrespectively, all polypeptide chains harboring the first 61 amino acid residues were brought down by GST-Snu17p in pull-down assays (Fig. 3.29 C, lanes 5, 8, 11, 14, 17, 20). These data suggest that the Snu17p-binding capacity of Pml1p does not rely on an intact FHA domain. Since FHA domains often function as phosphopeptide-binding elements (Durocher et al., 1999; Durocher and Jackson, 2002), the Snu17p-Pml1p interaction is apparently independent of Snu17p phosphorylation. Contrary to the dispensable FHA domain, Pml1p fragments lacking the N-terminal 61 residues failed to interact with Snu17p (Fig. 3.29 C, lanes 26, 29, 30) and even a deletion of the first 33 residues of Pml1p abrogated the binding (Fig. 3.29 C, lane 23).
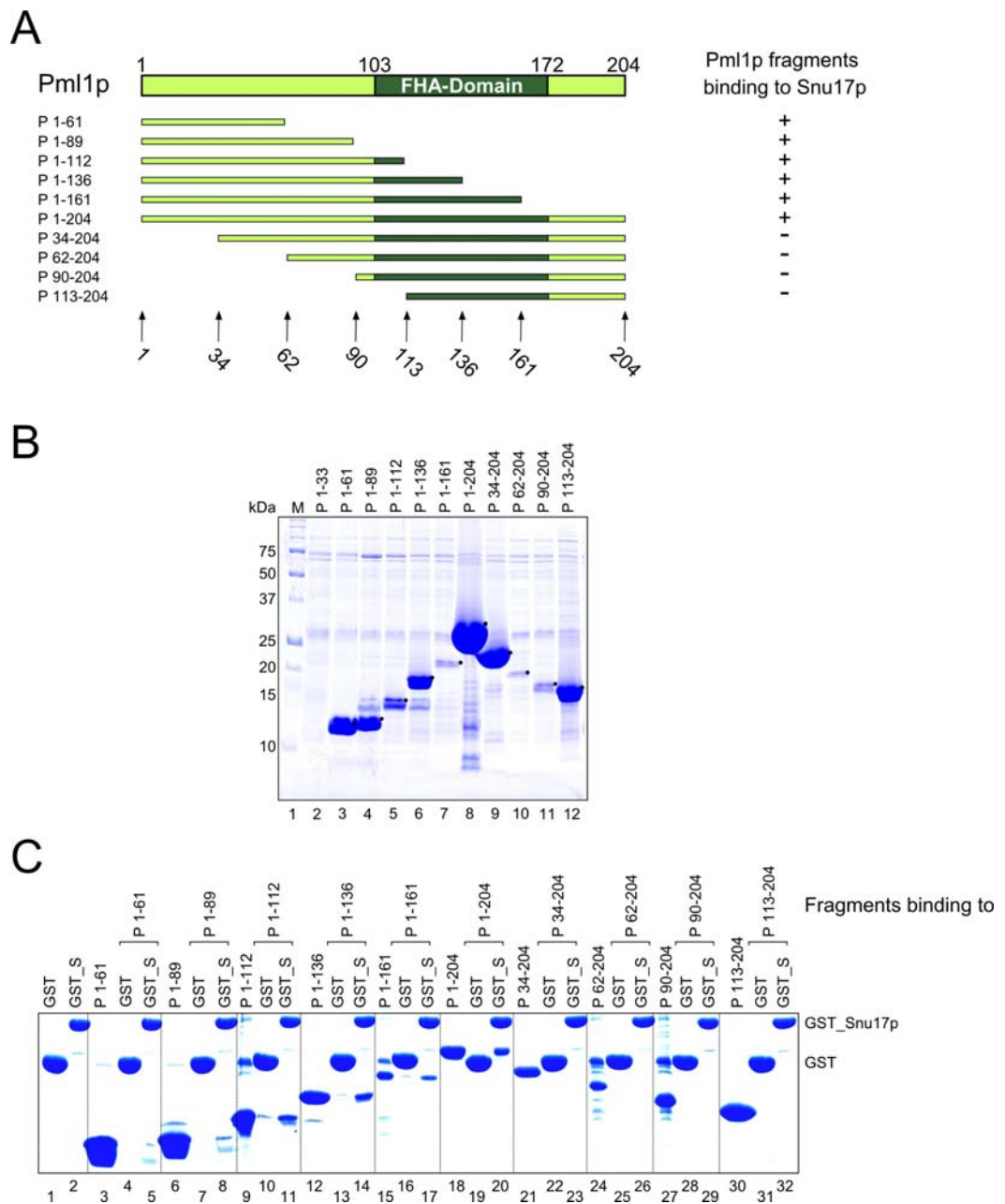
Fig. 3.29: The Snu17p-binding region of Pml1p. (A) Schematic representation of Pml1p and Pml1p fragments engineered for GST pull-down assays with start and end amino acid residues designated on the left. The predicted unstructured N-terminal part of Pml1p is shown in light green. The predicted C-terminal FHA-domain is in dark green. Interactions of the Pml1p fragments with GST-Snu17p fusion protein are indicated by plus (binding) and minus (no binding) on the right. P – Pml1p. **(B)** Solubility of Pml1p constructs analyzed by SDS-PAGE. All fragments were expressed under identical conditions and fragments were treated equally during purification. Fragments P1-112, P1-161, P62-204, and P90-204 were weakly soluble. Bands of correct constructs are indicated by black dots. **(C)** Pull-down of His$_6$-tagged Pml1p fragments **(A)** by glutathione-Sepharose beads pre-coated with recombinant GST or GST-Snu17p fusion protein shows that Pml1p binding is dependent on the presence of the N-terminal portion (residues 1-33), that residues 1-61 of Pml1p are sufficient for binding to Snu17p and that the Pml1p FHA-domain is dispensable. Bands in lanes 9, 15, 24 and 27, which are migrating at the size of GST, are contaminations. Experimental conditions were as in figure 3.24. P – Pml1p; S - Snu17p. (taken from Trowitzsch *et al*. 2008).

3.3.10 Structure probing of Snu17p-interacting peptides

Disorder predictions suggested that Snu17p interacting regions of Bud13p and Pml1p lack secondary structure (Prilusky et al., 2005). Equilibrium circular dicroism (CD) spectra of His_6-Bud13p$^{202-266}$ and His_6-Pml1p$^{1-61}$ peptides indicated random coil structures for these parts of the proteins (Fig. 3.30, grey line and black dotted line). In contrast, the equilibrium CD spectrum of a RRM containing fragment of Snu17p (Snu17p$^{25-138}$) was characterized by two minima, one at 208 nm and another at 222 nm, both indicative for the presence of structured parts in the protein (Fig. 3.30, black line). Thus, intrinsically unstructured portions in isolation at the C-terminus of Bud13p (residues 202-266) and the N-terminus of Pml1p (residues 1-61) are sufficient for interacting with Snu17p.
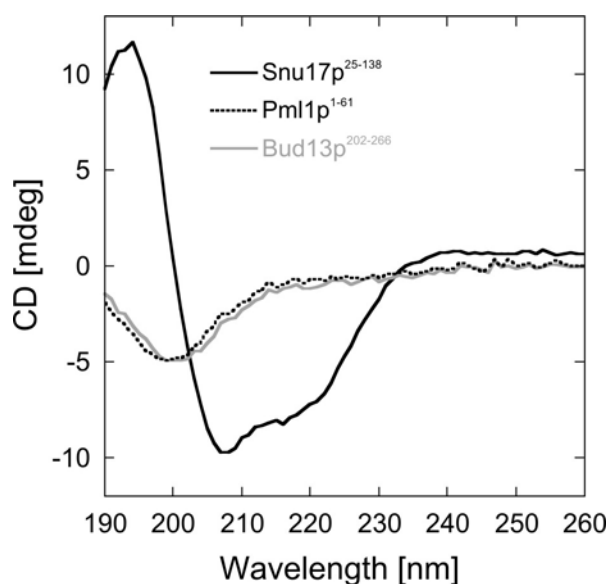
Fig. 3.30: Structure probing of Snu17p-interacting peptides by CD spectroscopy. Far-UV CD spectra of Snu17p$^{25-138}$ (black line), His_6-Pml1p$^{1-61}$ (dashed line) and His_6-Bud13p$^{202-266}$ (gray line) suggest that both the Bud13p and Pml1p portions are disordered in solution. Snu17p$^{25-138}$, which binds both Bud13p and Pml1p, shows a mixed α/β-structure indicated by minima at 208 and 222 nm. (taken from Trowitzsch *et al.* 2008).

3.3.11 Thermodynamic characterization of minimal hetero-dimers

To test whether the outlined interacting regions of Bud13p and Pml1p contain all Snu17p-contacting elements of the proteins, we compared thermodynamic binding parameters of the full-length and the minimal complexes using isothermal titration calorimetry (ITC). All protein samples (His$_6$-Snu17p$^{1-148}$, His$_6$-Snu17p$^{25-138}$, His$_6$-Snu17p$^{25-106}$, His$_6$-Pml1p$^{1-204}$, His$_6$-Pml1p$^{1-61}$, His$_6$-Bud13p$^{1-266}$, His$_6$-Bud13p$^{202-266}$ and His$_6$-Bud13p$^{202-266}$[Trp232Ala]) were checked for their purities via SDS-PAGE (Fig. 3.31).
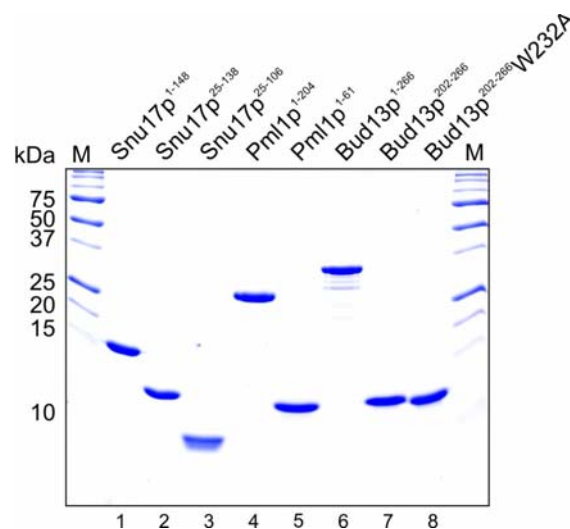
Fig. 3.31: Purities of ITC samples. 4 μg of each His$_6$-tagged protein, indicated on top of the gel, were separated on a 17 % SDS-PAA gel. Proteins were visualized by Coomassie blue staining. The analysis revealed purities of more than 95 % for each sample. M - marker.

ITC measurements yielded an estimated dissociation constant in the low nanomolar range for the interaction of Snu17p with full-length Bud13p and a value in the micromolar range for the binding of Snu17p to full-length Pml1p (Table 3.2; Appendix Fig. 6.1-3). Comparable dissociation constants were observed when the shortest interacting fragments of Bud13p (His$_6$-Bud13p$^{202-266}$) and Pml1p (His$_6$-Pml1p$^{1-61}$) were employed (Table 3.2). All proteins tested interacted in 1:1 stoichiometries (Table 3.2). Furthermore, the large negative interaction entropies for all protein pairs support the view that the interacting regions of Bud13p and Pml1p are intrinsically unstructured and become at least partially immobilized or fold upon interaction with Snu17p. Taken together, these data show or corroborate (i) that all

relevant Snu17p-binding elements are maintained within the first 61 residues of Pml1p and within the last 65 residues of Bud13p; (ii) that both Bud13p and Pml1p employ intrinsically unstructured regions to bind to Snu17p; and (iii) that RES components interact stoichiometrically. Furthermore, the particularly strong interaction between Snu17p and Bud13p suggest a permanent association between the two proteins *in vivo*. This notion is supported by the observation that deletions of the two proteins show similar phenotypes in yeast (Ni and Snyder, 2001; Scherrer and Spingola, 2006; Spingola et al., 2004). In contrast, the intermediate strength Snu17p/Pml1p complex may be indicative of a transient or facultative, and thus possibly regulated, association.

Table 3.2: Thermodynamic characterization of hetero-dimers. Raw data are shown in the appendix (Fig. 6.1-3). n.d. – not detected (taken from Trowitzsch *et al*. 2008).

Analyte	**Titrant**	K_d	ΔH [kcal mol^{-1}]	ΔS [cal K^{-1} mol^{-1}]	**Stoichiometry (n)**
Snu17p$^{1-148}$	**Bud13p$^{1-266}$**	2.5 nM	-76	-222	1.1
	Bud13p$^{202-266}$	2.6 nM	-61	-170	1.0
	Pml1p$^{1-204}$	1.1 μM	-15	- 24	0.9
	Pml1p$^{1-61}$	0.9 μM	-17	- 29	0.9
Snu17p$^{25-138}$	**Pml1p$^{1-204}$**	1.0 μM	-16	- 28	0.9
	Pml1p$^{1-61}$	0.6 μM	-19	- 37	0.9
Snu17p$^{25-106}$	**Bud13p$^{1-266}$**	6.8 nM	-53	-143	1.0
	Bud13p$^{202-266}$	5.3 nM	-58	-159	0.9
	Bud13p$^{202-266}$ [Trp232Ala]	n. d.	-	-	-

3.3.12 Identification of Snu17p-interacting epitopes by a peptide array system

To further delineate minimal binding epitopes among the RES components, peptide array experiments were performed. Arrays with 15-, 20- and 25-mer peptides derived from residues 202-266 of Bud13p and 1-60 of Pml1p were chemically synthesized with an offset of one residue and immobilized as spots on membranes. Recombinant GST-Snu17p bound strongly to consecutive peptides in all arrays, albeit spot intensities in Bud13p arrays appeared to be more unevenly distributed (Fig. 3.32).
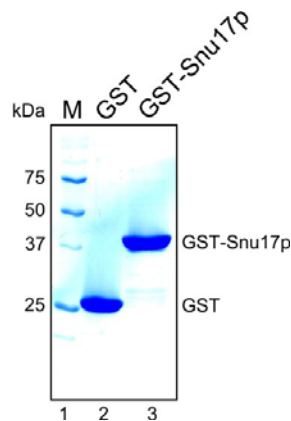
Fig. 3.32: Purity of samples for peptide scanning experiments. Samples of gel filtrated GST (lane 2) and GST-Snu17p (lane 3) were analyzed on a 15 % SDS-PAA gel and proteins were stained by Coomassie blue. The analysis reveals purities of more than 95 % for both protein preparations. M - marker. Protein size markers are indicated on the left of the gel; protein identities on the right.

Control experiments with GST alone did not show any specific binding (Fig. 3.33). Minimal binding regions representing amino acids ^{229}GSRWDGVHR^{237}, ^{224}FAIMPGSRWDGVHR^{237} and ^{227}MPGSRWDGVHR^{237} could be deduced from 15-, 20- and 25-mer peptide arrays of Bud13p, respectively (Fig. 3.33 A, B, C), when all interacting peptides of the array were taken into account.

Minimal binding regions in Pml1p, embodying amino acids ^{28}IMPDFSPSGLL^{38}, could be worked out from 15- and 25-mer peptide arrays, whereas the 20-mer array indicated residues ^{29}MPDFSPSGLL^{38} to be the minimal binding region (Fig. 3.33 D, E, F).

Despite a minor discrepancy of five residues in the definition of N-termini between the three Bud13p-experiments, the C-terminus seems to be fixed by Arg239. All sequences of interacting peptides cover Trp232, which was shown to be indispensable for the interaction (sections 3.3.7 and 3.3.8). The Snu17p-binding epitope of Pml1p is more unambiguously defined with just a difference of one residue at the N-terminus among all three peptide-scanning experiments. Thus, Snu17p seems to bind a very short epitope of Pml1p with approximately 10-11 residues in length and to a region of at least 9 residues in Bud13p.
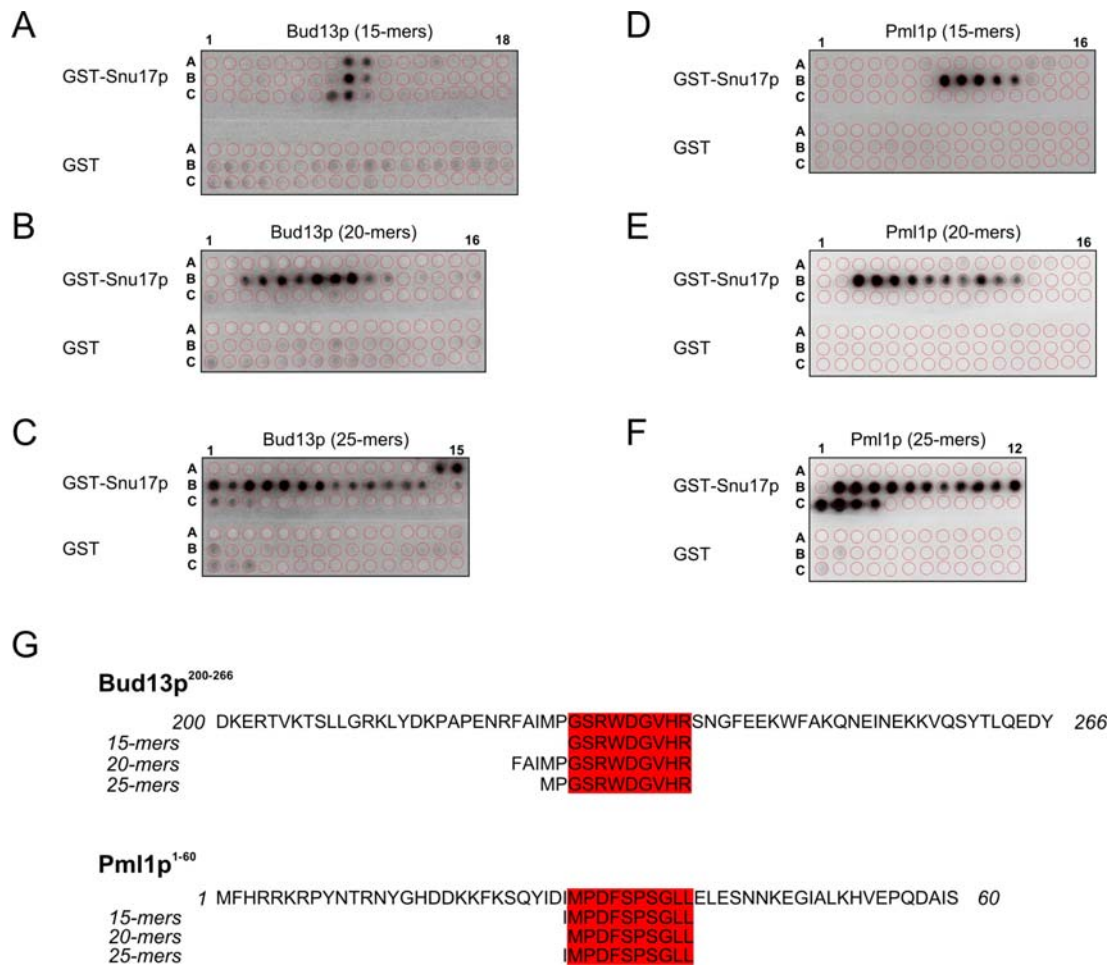
Fig. 3.33: Snu17p binds short epitopes of Bud13p and Pml1p in peptide scanning experiments. (A-C) Peptide scan analysis of the Bud13p-Snu17p interaction. Cellulose-bound peptide arrays were successively incubated with GST-Snu17p fusion protein, rabbit anti-GST antibodies and horseradish peroxidase-coupled goat anti-rabbit antibodies. CCD images were obtained after luminol reaction. Arrays of 15- (A), 20- (B) and 25-mers (C) of Bud13p[202-266] were used with an offset of one resiude. **(D-F)** Experimental conditions were as in (A) but arrays of 15- (D), 20- (E) and 25-mers (F) of Pml1p[1-60] were used instead. Spot identities are given for each array. **(G)** Minimal binding epitopes, deduced from each peptide array, are shown. Peptide regions common to all sequences within one array system are highlighted by a red box. Start and end amino acids are indicated by italic numbers.

3.3.13 The susceptibility of Snu17p to chymotrypsin is altered by the presence of interacting peptides

Since Snu17p binds short epitopes of Bud13p and Pml1p (section 3.3.12), their impact on the stability of Snu17p was tested by a limited proteolysis approach (Cleveland et al., 1977). The susceptibility of Snu17p to chymotrypsin in the presence or absence of peptides [228]PGSRWDGVHR[237] (Bud13p) and

^{28}IMPDFSPSGLL^{38} (Pml1p) was analyzed by varying protease concentration and incubation time. Protein fragments remaining from protease treatment were visualized by SDS-PAGE and coomassie staining (Fig. 3.34 A, B) and were further analyzed by mass spectrometry (Fig. 3.34 C). The presence of equimolar amounts of peptides stabilized a larger protein fragment of Snu17p as compared to the control experiment without peptides (Fig. 3.34 A, lanes 4, 9). Mass spectrometry analysis of the remaining bands after limited proteolysis showed, that the longer Snu17p fragment consists at least of residues 29-138 (Fig. 3.34 C, middel panel), whereas the shorter Snu17p fragment lacked C-terminal residues and only peptides covering the region including residues 29-113 could be identified (Fig. 3.34 C, bottom panel). Although a control experiment with unrelated peptides was not performed, these data suggest that the presence of Snu17p-interacting peptides presumably stabilizes the structure of Snu17p and thereby renders the protein less susceptible to proteases. Interestingly, the longer Snu17p fragment harboring residues 29-138 fits well to the minimal Snu17p fragment (Snu17p$^{25-138}$), which was shown to be capable of binding to Bud13p and Pml1p in GST pull-down experiments (Fig. 3.24 B, C, lanes 20).
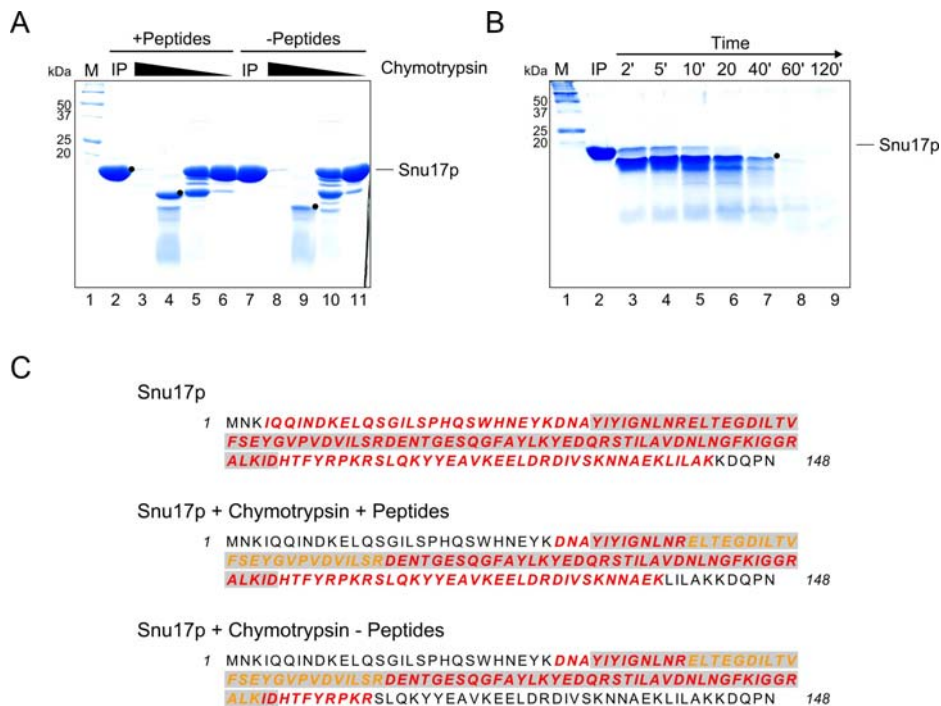
Fig. 3.34: Limited proteolysis of Snu17p in the presence and absence of interacting peptides.
(**A**) Free recombinant Snu17p protein (lanes 7-11) and Snu17p protein in the presence of equimolar
Bud13p [228]PGSRWDGVHR[237] and Pml1p [28]IMPDFSPSGLL[38] peptides (lanes 3-6) were incubated 45
min at 20°C with endoprotease chymotrypsin at protease/protein mass ratios of 1:20 (lanes 3, 8),
1:200 (lanes 4, 9), 1:2000 (lanes 5, 10) and 1:20000 (lanes 6, 11), subsequently separated by SDS-
PAGE and visualized by Coomassie blue staining. Lanes 1 and 6 show input controls. Full-length
Snu17p is indicated on the right, protein size markers on the left. (**B**) Time course of the proteolysis.
Snu17p protein in the presence of peptides (as in **A**) and chymotrypsin (protease/protein mass ration
of 1:100) were incubated for 2, 5, 10, 20, 40, 60 and 120 min and subsequently analyzed as in **A**. Lane
1 shows the input control. (**C**) Peptide mass fingerprint analyses of Snu17p fragments. Bands were
excised from gels shown in A and B and tryptic fragments were identified by MS (red, italic letters).
Sequences of Snu17p shown in italic, yellow letters could not be assigned. Note that free Snu17p is
further cleaved from the C-terminus as compared to the peptide bound state. Similar peptide
coverages were found in fragments derived from the peptide bound state of Snu17p in both
experiments (A, lane 3 and B, lane 7). M - marker, IP - input, black dots - fragments analyzed by MS,
grey boxes highlight the predicted RRM.

3.3.14 Nuclear magnetic resonance spectroscopy of Snu17p alone and in complex with interacting peptides

In order to investigate the nature of the structural changes in Snu17p induced by
short peptides derived from Pml1p ([28]IMPDFSPSGLL[38]) or Bud13p
([228]PGSRWDGVHR[237]), a collaborative work with the nuclear magnetic resonance
(NMR) group of Markus Zweckstetter at the MPI Göttingen was initiated and is still
ongoing. Heteronuclear single quantum coherence (HSQC) experiments were

performed with isotopically-labeled ^{15}N-Snu17p in the presence or absence of interacting peptides. An HSQC experiment with ^{15}N-Snu17p alone shows, that peaks are crowded for H-dimension (Fig 3.35 A). Most peaks locate within 1 ppm and just a few are to be found out of this range (Fig 3.35 A). This experiment indicated, that Snu17p on its own adopts a molten globule-like structure with some features of secondary structure elements. Upon addition of a short Bud13p peptide (Bud13p$^{228-237}$) the spectrum is basically unchanged, indicating that Bud13p$^{228-237}$ does not bind or that binding occurs in a part that is broadened, beyond detection due to chemical exchange (Fig. 3.35 B). This finding indicates, that presumably a longer fragment of Bud13p seems to be necessary to efficiently bind to Snu17p, which is consistent with data derived from intrinsic Trp fluorescence measurements (section 3.3.8). Upon binding of peptide Pml1p$^{28-38}$ to Snu17p a strong change in the spectrum is observed (Fig. 3.35 C). It appears, that the RRM motif of Snu17p is fully formed, since peaks locate in a range of at least 3 ppm in H-dimension. This stabilizing effect of Pml1p peptides was also reflected in longer shelf times of the samples at 293 K, which is advantagous for NMR data aquisition. Upon addition of peptide Bud13p$^{228-237}$ to the sample containing both Snu17p and Pml1p$^{28-38}$ further chemical shift changes are observed (Fig. 3.35 D) and suggest, that the short Bud13p peptide may bind to the pre-formed RRM of Snu17p.

To further elucidate the above findings, isotopically double-labeled ^{15}N/^{13}C-Snu17p was produced and purified (Fig. 3.36). At the moment, backbone assignments are on their way to illuminate, which residues of Snu17p are important for the interaction with Bud13p and Pml1p.
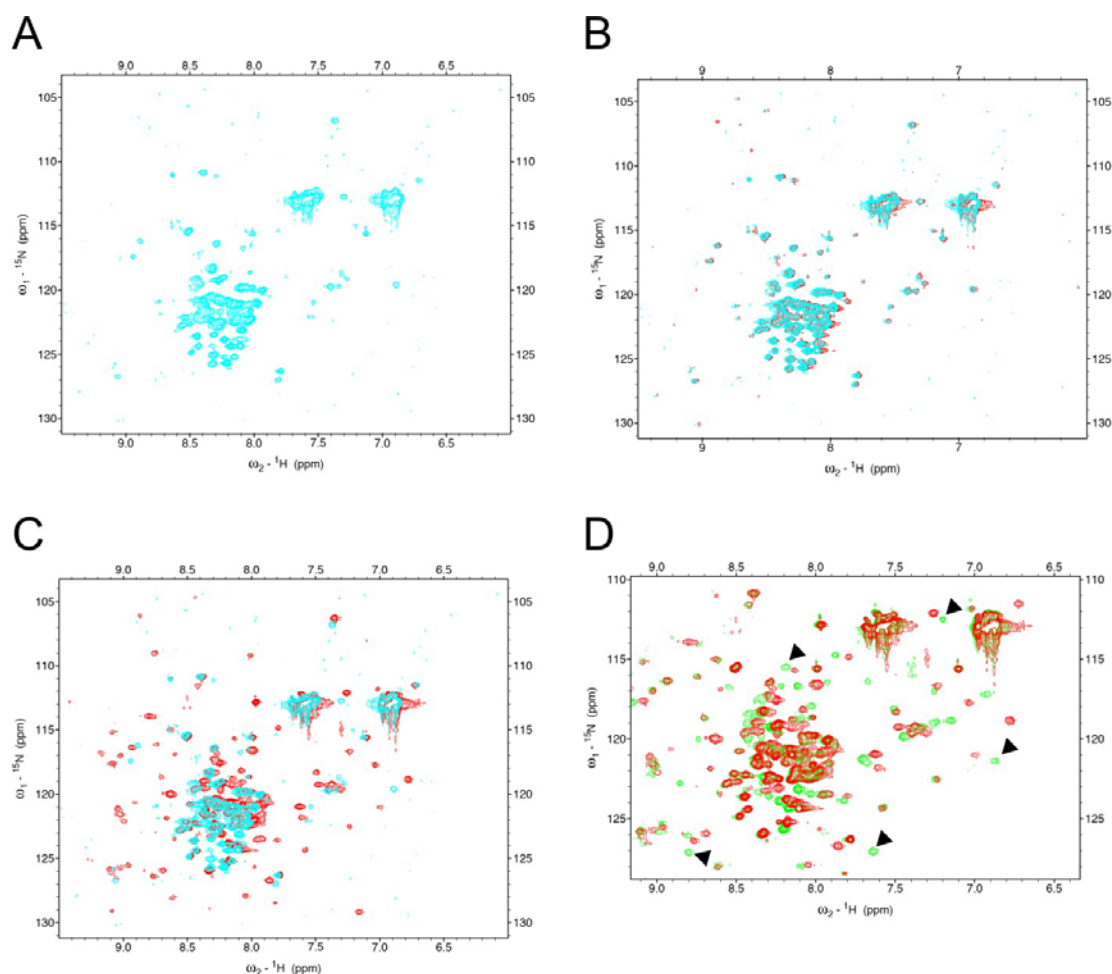
Fig. 3.35: NMR titration experiments with Snu17p and minimal Bud13p- and Pml1p-peptides. (A) ^{1}H, ^{15}N correlation NMR spectrum of ^{15}N-labeled Snu17p alone (cyan). The chemical shifts indicate a molton globule state of Snu17p in isolation. The Snu17p concentration was 117 μM in a buffer comprising 20 mM Na/K PO$_4$ (pH 7.5), 250 mM NaCl and 1 mM DTT at 15 °C. **(B)** ^{1}H, ^{15}N correlation NMR spectrum of ^{15}N-labeled Snu17p in the presence of ~3-fold molar excess of Bud13p$^{228-237}$ (red) as compared to the spectrum of Snu17p alone (cyan). The presence of Bud13p$^{228-237}$ hardly effects the structure of Snu17p, since no chemical shift changes are to be observed. **(C)** ^{1}H, ^{15}N correlation NMR spectrum of ^{15}N-labeled Snu17p in the presence of ~10-fold molar excess of Pml1p$^{28-38}$ (red) in comparison to the spectrum of Snu17p alone (cyan). The chemical shift changes indicate that Snu17p folds upon binding to Pml1p$^{28-38}$. **(D)** ^{1}H, ^{15}N correlation NMR spectrum of ^{15}N-labeled Snu17p in the presence of ~10-fold molar excess of Pml1p$^{28-38}$ and ~3-fold molar excess of Bud13p$^{228-237}$ (green) as compared to the spectrum of Snu17p/Pml1p$^{28-38}$ complexes (red). Further chemical shift changes upon titration of Bud13p$^{228-237}$-peptides to Snu17p/Pml1p$^{28-38}$ complexes (green) are indicated by arrows. All NMR experiments were performed by Shengqi Xiang.

To summarize, NMR titration experiments suggest that (i) Snu17p in isolation exists in a molten globule-like state; (ii) that the protein adopts a rigid structure upon binding of a short peptide from Pml1p and that (iii) several residues of Snu17p experience further chemical-shift changes upon addition of Bud13p peptide.
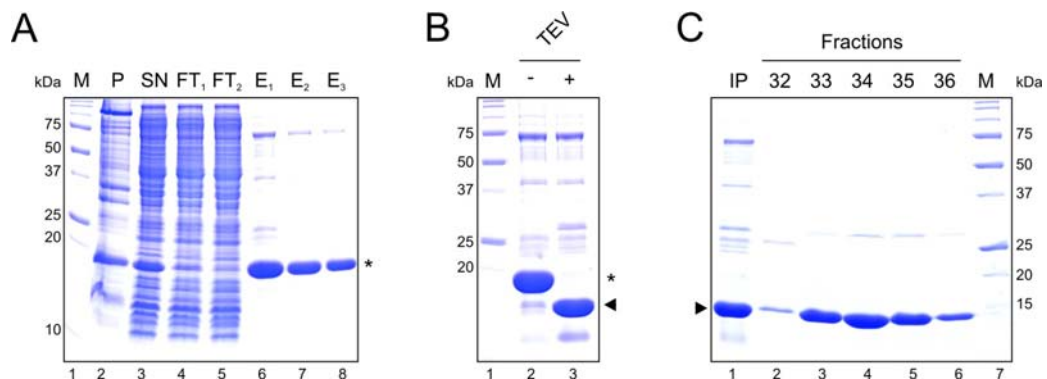
Fig. 3.36: Purification of isotopically-labeled ^{15}N/^{13}C-Snu17p. Panels **(A)**-**(C)** show SDS-PAGE analyses of the purification procedure. Proteins were stained with Coomassie blue. **(A)** Lanes 2 and 3 show samples from unsoluble and soluble fractions from cells after sonication, respectively. Lanes 4/5 and 6-8 show flowthrough (FT$_{1/2}$) and elution (E$_{1-3}$) fractions from the first purification step via Ni^{2+}-NTA agarose, respectively. The asterisc indicates His$_6$-tagged Snu17p protein. **(B)** Analysis of His$_6$-tag removal by TEV protease treatment. Lane 2 shows a sample of Snu17p before, lane 3 after TEV cleavage. The arrowhead indicates the cleaved Snu17p product. **(C)** Analysis of peak fractions from the final gel filtration step. The protein is more than 95 % pure. Lane 1 shows an input sample, lanes 2-6 samples of peak fractions. M - marker; P - pellet; SN - supernatant; FT - flowthroug; E - eluate; IP - input. Protein size markers are indicated.

3.3.15 Purification and crystallization of hetero-di- and trimeric RES complexes

The results from biochemical and biophysical mapping suggested various options for crystallization trials of RES particles, sub-complexes and single proteins. In this section, attempts to crystallize components of the RES complex are described. In order to keep this section short, only gel-electrophoretic analyses of final polishing steps will be illustrated. Descriptions of purification procedures can be found under "Materials and Methods".

At first, full-length RES complexes consisting of Bud13p, Pml1p and His$_6$-tagged Snu17p (Fig. 3.37 A) were produced and used for crystallization trials at concentrations up to 50 mg/ml. 1536 conditions were screened but no crystals were obtained. A truncated complex (Fig. 3.37 B), comprising His$_6$-tagged Snu17p, Pml1p and a C-terminal fragment of Bud13p (Bud13p$^{125-266}$) did not form any crystals either.
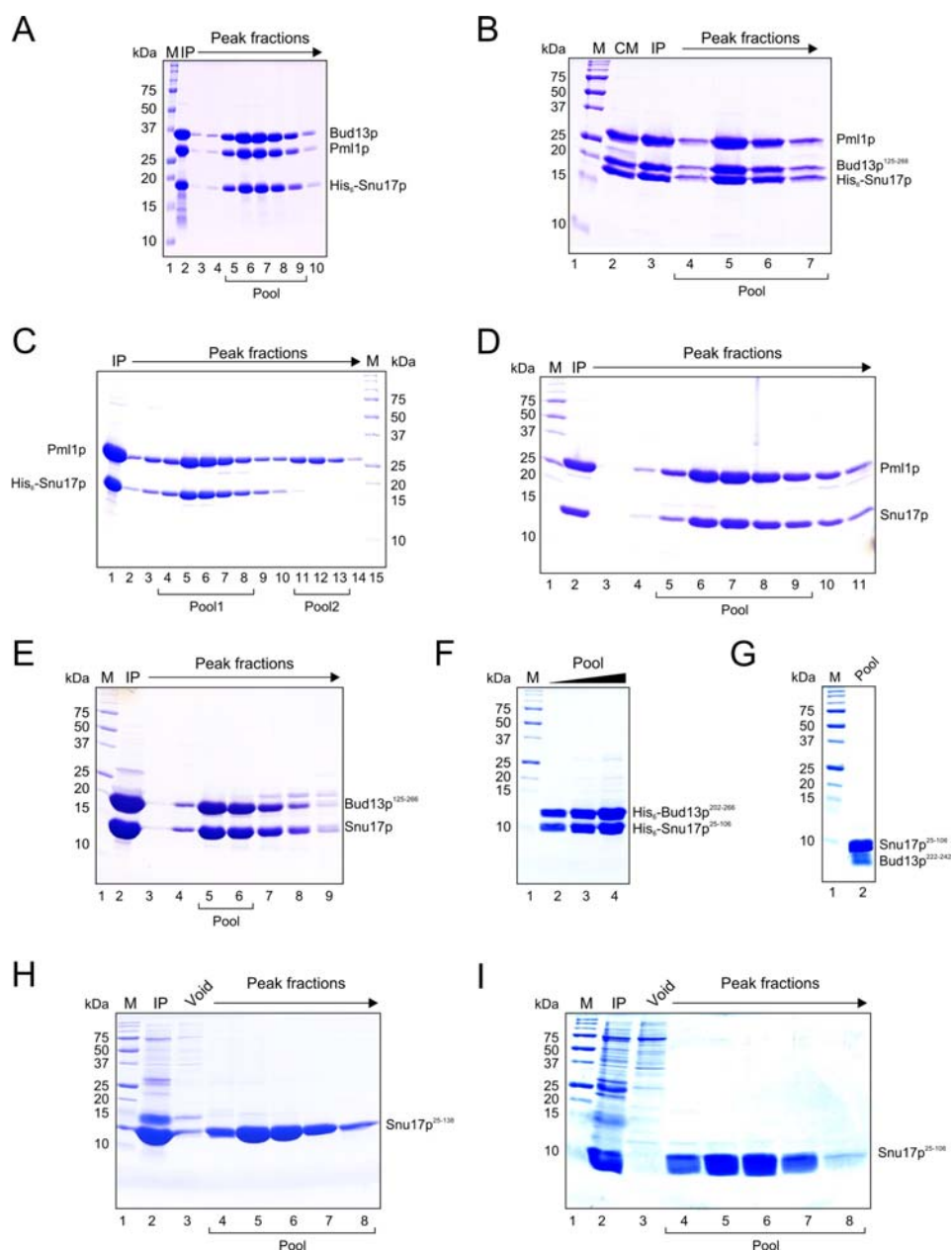
Fig. 3.37: Manifold RES particles and components subjected to crystallization attempts.
SDS-PAGE analyses of peak fractions of final preparative gel filtrations. Proteins were stained by
Coomassie blue. **(A)** Preparation of full-length RES complexes as in figure 3.18. **(B)** Preparation of a
truncated RES complex comprising full-length proteins Snu17p and Pml1p and a Bud13p$^{125-266}$
fragment. Preparation of a Snu17p-Pml1p dimer with **(C)** and without His$_6$-tag **(D)**. **(E)** Preparation of
a Snu17p-Bud13p dimer comprising a Bud13p$^{125-266}$ fragment. **(F)** Preparation of a Snu17p-Bud13p
dimer comprising His$_6$-tagged Bud13p$^{202-266}$ and Snu17p$^{25-106}$ fragments. Different amounts of the
final concentrated sample were loaded on the gel. **(G)** Preparation of a Snu17p-Bud13p dimer
comprising a Snu17p$^{25-106}$ fragment and a Bud13p$^{222-242}$ peptide. Preparation of a Snu17p$^{25-138}$ **(H)** and
a Snu17p$^{25-106}$ fragment **(I)**. Fractions, which were pooled prior to ultrafiltration, are indicated. Protein
size markers are indicated. M- marker; IP - input; CM - sample after ion exchange chromatography
using a CM-Sepharose column (see "Materials and Methods" for details).

As shown in section 3.3.12, minimal Snu17p-binding regions in the N-terminus of Pml1p and in the C-terminus of Bud13p were identified by peptide arrays. Consequently, a particle (termed mini-RES), comprising Snu17p$^{25-138}$, Bud13p$^{228-237}$ and Pml1p$^{28-38}$, was generated. Since short peptides are hardly visible in denaturing polyacrylamide gels, mass spectrometric analysis was used to reveal the presence of all components in the final concentrated sample. Expected masses of the complex components Snu17p$^{25-138}$, Bud13p$^{228-237}$ and Pml1p$^{28-38}$ were detected (not shown).

In total 1152 conditions were screened for crystal growth at 20 °C with a concentration of mini-RES particles between 34 and 64 mg/ml. Three conditions were found, in which thin plates or a needle cluster were grown (Fig. 3.38 A). Unfortunately, both crystal forms could not be reproduced. Therefore, the needle cluster as such was tested for its ability to diffract X-rays. The diffraction pattern of the cluster indicated that it most likely consisted of protein compounds (not shown). X-ray diffraction capability of plate-like crystals could not be analyzed, since they were tightly attached to the walls of the wells. Further preparative fine-tuning will be necessary to obtain bigger crystals in a reproducible manner, which may then be used for structure determination.
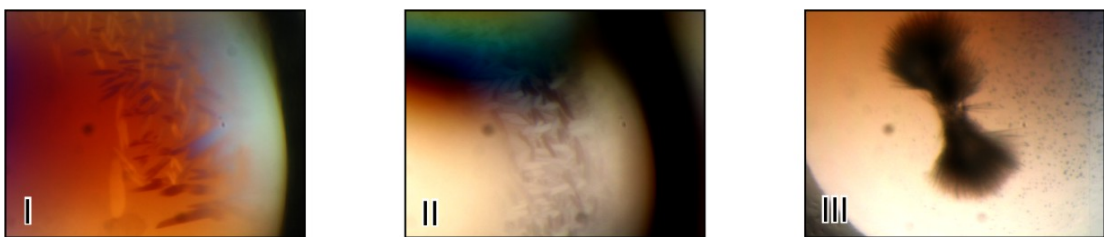
Fig. 3.38: Initial crystals of mini-RES complexes. Images of crystals derived from a nano-scale screening approach with mini-RES omplexes at a concentration of 34 mg/ml under condition I (0.085 M Tris-HCl (pH 8.5), 0.17 M LiSO$_4$, 25.5 % PEG 4000, 15 % glycerol), condition II (0.085 M Tris-HCl (pH 8.5), 0.17 M Na acetate, 25.5 % PEG 4000, 15 % glycerol) and condition III (0.05 M Tris-HCl (pH 7.5), 0.01 M MgCl$_2$, 1.6 M AmSO$_4$). Crystals were visualized by polarizing light.

As demonstrated in section 3.3.3 hetero-dimers and single components of RES complexes can be isolated. Therefore, dimeric complexes consisting of Snu17p and Pml1p (Fig. 3.37 C, D) or Snu17p and Bud13p$^{125-266}$ (Fig. 3.37 E) were assayed for crystal growth. Furthermore, complexes consisting of His$_6$-tagged Snu17p$^{25-106}$ and

Bud13p[202-266] (Fig. 3.37 F), or untagged Snu17p[25-106] and a Bud13p[222-242] peptide (Fig. 3.37 G) were tested. Additionally, truncated versions of Snu17p, e.g. Snu17p[25-138] (Fig. 3.37 H) and Snu17p[25-106] (Fig. 3.37 I), were subjected to crystallization trials. Since attempts to crystallize these fragments failed, chemical modifications of surface residues were tried. Reductive methylation of surface lysine residues was reported to be a potent rescue strategy, facilitating crystallization of otherwise intractable proteins (Walter et al., 2006). Although no crystals could be observed with methylated Snu17p[25-138], the chemical modification changed the properties of the protein drastically. Methylated Snu17p[25-138] could be brought to much higher concentrations compared to unmethylated samples and migrated faster in analytical gel filtration (not shown). At higher concentrations (above 0.8 µM), Snu17p[25-138] tends to form gels, which are only soluble in guanidin hydrochloride solution. Methylated Snu17p[25-138], however, did not form such gels. Finally, full-length Pml1p was subjected to crystallization trials and crystals could be obtained within one week. The crystallization and structure determination of Pm1p will be described in detail in section 3.4.

In summary, first steps towards an atomic structure of core RES complexes were made. Crystals of Pml1p could be obtained. Crystallization of a minimal RES complex, capable of diffracting X-rays, was achieved but need further improvement. Assignment of critical residues of Snu17p and of ligand-peptides derived from Bud13p and Pml1p by NMR spectroscopy is in progress. These definitions may open up new possibilities for the production of core RES complexes, which crystallize in a reproducible manner.

3.4   Crystal structure of Pml1p

In the preceding sections, a detailed interaction map among the RES components was presented. Snu17p harbors an unusual RNA recognition motif (RRM), which presents two separate binding sites for short peptides in the C-terminus of Bud13p and in the N-terminus of Pml1p (section 3.3.4 ff). Most likely, Snu17p is also the main anchor of the complex to the spliceosome. This organization leaves the bulk of Bud13p and Pml1p available to engage in other activities.

Sequence analysis suggested that the C-terminal region of Pml1p, which is not involved in Snu17p binding, fold into a forkhead-associated (FHA) domain (Fig. 3.22). In the following, the elucidation of crystal structures of full-length Pml1p and of a truncated form lacking the first 50 residues is presented.

3.4.1   Structure determination and quality of Pml1p models

Full-length Pml1p (Pml1p^{FL}) was produced heterologously in *E. coli* and could be purified to homogeneity using affinity and size exclusion chromatography (Fig. 3.39 A). Pml1p^{FL} crystallized in primitive hexagonal space group P3$_2$21 with two monomers per asymmetric unit. Since crystals grown from initial screening conditions diffracted x-rays relatively poorly, an additive screen was performed and large, single crystals were obtained by the addition of isopropyl-β-D-galactopyranoside (IPTG) (Fig. 3.39 B). The improved crystals allowed data collection to 2.4 Å resolution at beamline PXII at the Swiss Light Source (Fig. 3.39 C).
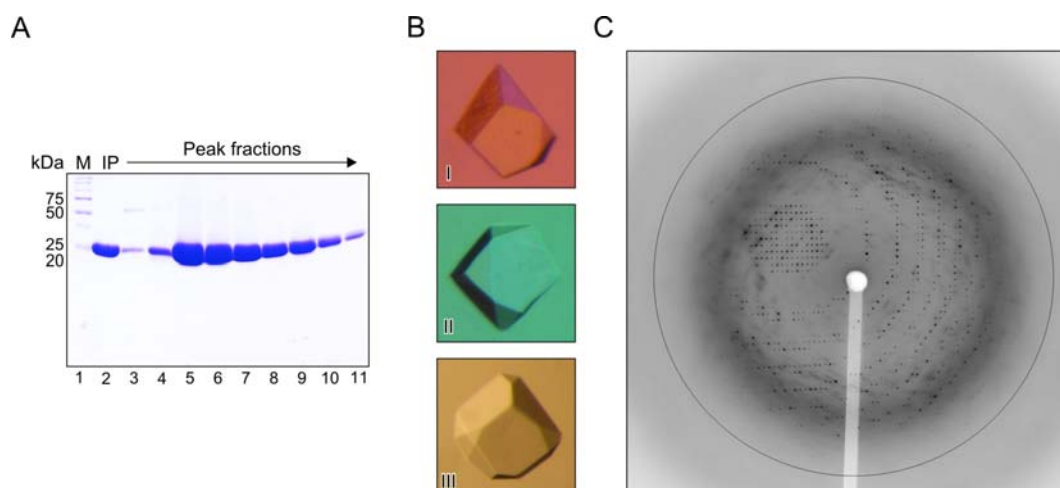
Fig. 3.39: Purification and crystallization of Pml1p^{FL} and diffraction capability of Pml1p^{FL} crystals. (A) SDS-PAGE analysis of the final gel filtration step shows purity of more than 95 % of Pml1p after Coomassie staining. Lane 2 shows the input sample, lanes 3-11 samples of peak fractions. M - marker; IP - input. Protein size markers are indicated. **(B)** Images of crystals grown from 0.1 M Tris-HCl (pH 7.5), 200 mM LiSO$_4$, 23 % PEG 3350 and IPTG as an additive (I, II, III), visualized by polarizing light. **(C)** Diffraction image of a single crystal from (B). The circle indicates a resolution of 2.4 Å. Image was taken over a Δφ range of 0.5° at beamline PXII, SLS.

Attempts to solve the structure by molecular replacement using FHA domain models from various proteins failed. Therefore, crystals were grown from selenomethionine-substituted Pml1p proteins (Fig. 3.40 A). Although fluorescence scans revealed the presence of selenium (Fig. 3.40 B), the poor diffraction capability of the crystals did not allow collection of useful data sets.
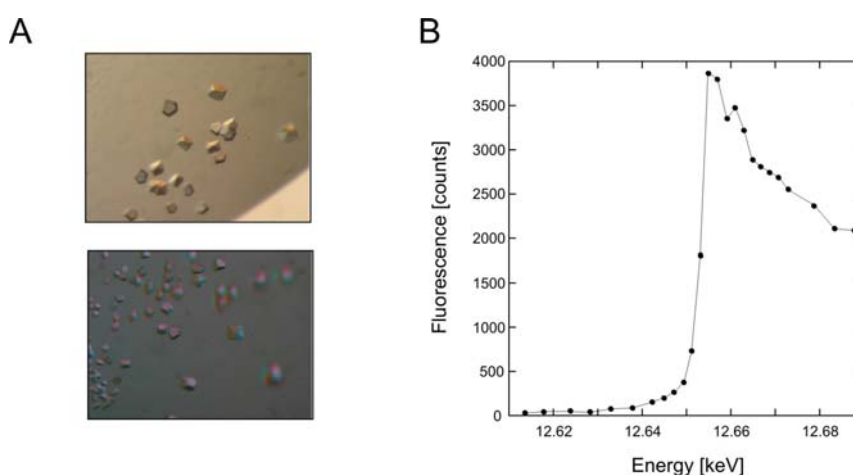
Fig. 3.40: Characterization of crystals grown from seleno-methionine derivatized Pml1p molecules. (A) Images of crystals grown from 0.1 M Tris-HCl (pH 7.5), 200 mM LiSO$_4$, 23 % PEG 3350 and IPTG as an additive, visualized by polarizing light. **(B)** A fluorescence scan around the K edge of selenium shows a fluorescence peak at the theoretical excitation energy (12.6578 keV), indicating the presence of the selenium compound in the crystals.

Screening of various heavy atom compounds identified Bakers' dimercurial as a promising candidate. Figure 3.41 shows a diffraction image (A) and the plot of a fluorescence scan (B) with decomposed f' and f'' values (C) of a Pml1p crystal, derivatized with Bakers' dimercurial.
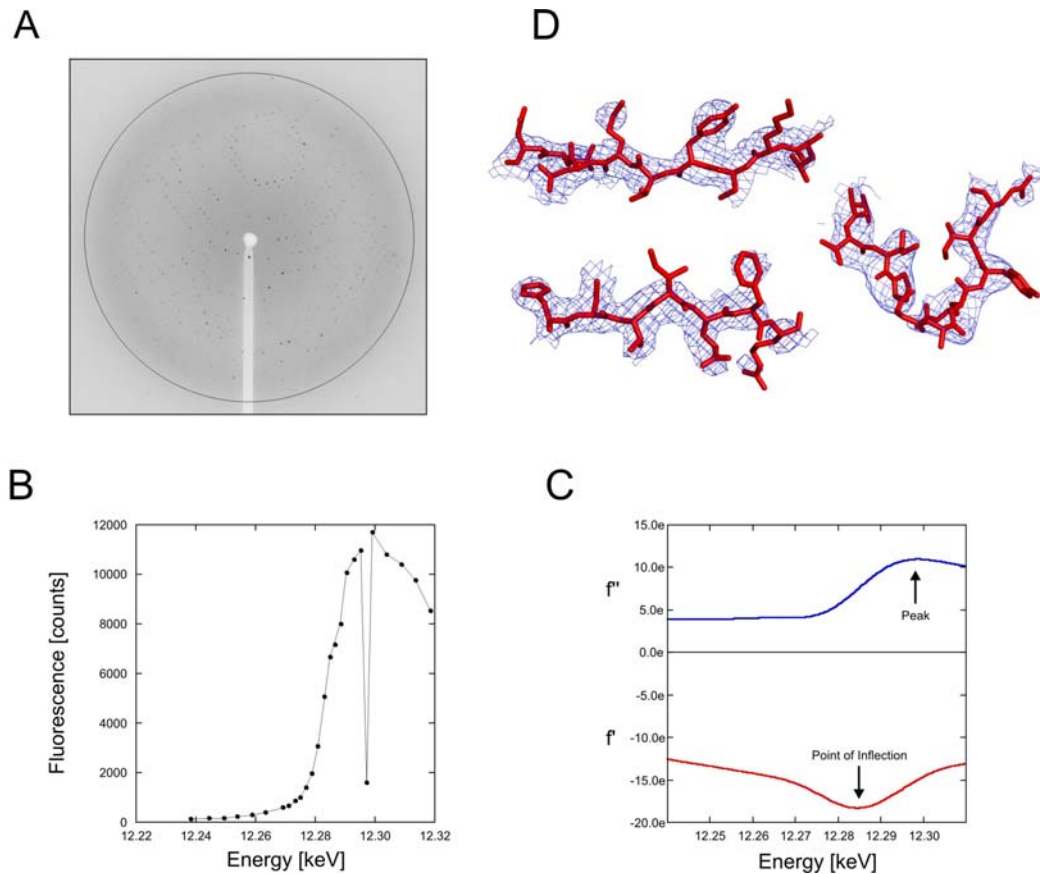
Fig. 3.41: Characterization of Pml1p^{FL} crystals soaked with Bakers' dimercurial. (A) Diffraction image of a single crystal from Pml1p^{FL} after soaking in with Bakers' dimercurial. The circle indicates a resolution of ~3.0 Å. Image was taken over a $\Delta\varphi$ range of 0.5° at beamline PXII, SLS. **(B)** A fluorescence scan of Pml1p^{FL} crystals, soaked with Bakers' dimercurial. The scan was taken around the L-III edge of mercury. **(C)** Decomposition of the fluorescence into imaginary compound f'' (experimental) and real compound f' indicates a peak at an energy of 12.2986 keV (f' = -15.080; f'' = 10.074) and an inflection point at an energy of 12.2845 keV (f' = -18.035; f'' = 7.416). **(D)** Quality of the experimentally determined electron density map after single-wavelength anomalous dispersion (SAD) phasing and solvent flattening. Regions of the map covering three representative portions of the protein are displayed at a contour level of 1 σ. Parts of the final refined model of Pml1p (red) are displayed as stick model.

A dataset collected at the high-energy side of the Hg L-III-edge allowed solving the Pml1p structure at 3.3 Å resolution *via* the single-wavelength anomalous

dispersion (SAD) method, taking advantage of the anomalous scattering of the mercury derivative (Table 3.3). Six heavy atom sites with significant occupancies were found and allowed calculation of initial phases (Fig. 3.42). β-sheets could be built in the solvent flattened SAD electron density map and were used as a search model to phase the native dataset, which proved non-isomorphous to the Hg-set. The final model was refined to 2.4 Å resolution and exhibited R_{work} and R_{free} values of 21.5 % and 27.2 %, respectively. Most residues were in the favored or additionally allowed regions of the Ramachandran plot (Table 3.3 and Appendix Fig. 6.5 / 6.6).
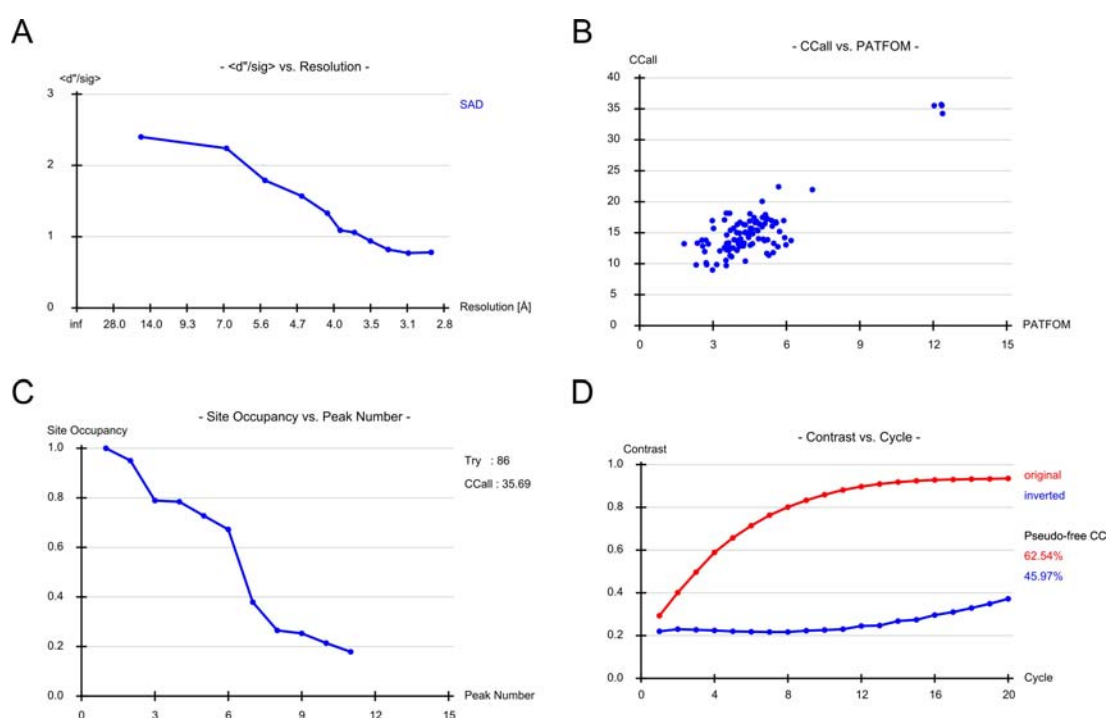
Fig. 3.42: ShelX statistics of the structure solution process using a mercury SAD dataset. (A) Plot of the anomalous differences versus resolution shows significant anomalous signal up to a resolution of 3.3 Å. **(B)** Plot of overall correlation coefficients versus Patterson figure of merit indicates a distinct solution. **(C)** Plot of heavy atom site occupancies (peak-hight) versus peak number suggests 6 heavy atom sites with high occupancies (0.7-1.0) and 2 with lower occupancies (0.25-0.38). **(D)** Plot of contrast versus density modification cycle for the two enantiomorphs. The contrast values for the original and inverted substructure are shown in red and blue, respectively. Here, the original substructure is the correct one. All plots were generated by the program HKL2MAP (Pape and Schneider, 2004).

In parallel with the full-length protein, the crystal structure of a truncated version, Pml1p[51-204], lacking the N-terminal 50 residues, was determined. Crystals were obtained from a highly pure protein sample under different conditions (Fig. 3.43 A,

B). The best diffracting crystals were found in a condition containing polyethylene glycol (PEG) 10000, glycerol and ethylene glycol (Fig. 3.43 B), that allowed collection of a data set to 1.8 Å resolution (Fig. 3.43 C). Pml1p$^{51-204}$ crystallized in primitive rhombohedral space group R32 with one molecule per asymmetric unit. The structure was solved by molecular replacement using the full-length structure as a search model and was refined at 1.8 Å resolution with R$_{work}$ and R$_{free}$ values of 18.7 % and 23.5 %, respectively, retaining good stereochemistry (Table 3.3).
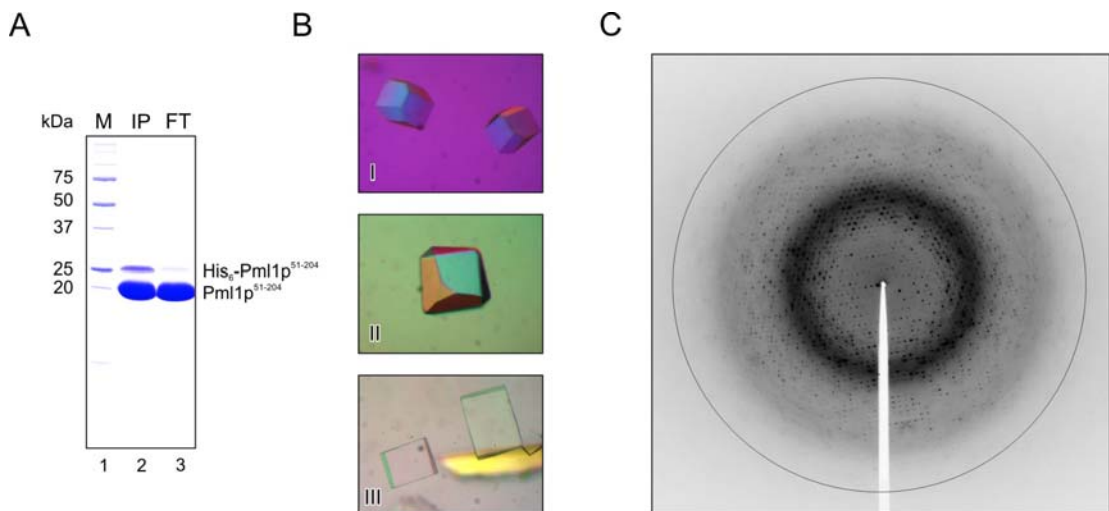
Fig. 3.43: Purification and crystallization of a truncated Pml1p$^{51-204}$ and diffraction capability of Pml1p$^{51-204}$ crystals. (A) SDS-PAGE analysis of the final purification step ("Ni^{2+}-cycling") reveal a purity of Pml1p$^{51-204}$ of more than 95 % after Coomassie staining. Lane 2 shows a sample of Pml1p$^{51-204}$ before and lane 3 after Ni^{2+}-cycling. M - marker; IP - input; FT - flow-through. Protein size markers are indicated. **(B)** Images of crystals obtained from condition I [0.085 M HEPES (pH 7.1), 17 % PEG 10000, 6.8 % ethylene glycole and 15 % glycerol], condition II [0.075 M Tris-HCl (pH 7.4), 1.4 M AmSO$_4$, 25 % glycerol] and condition III [0.16 M MgSO4, 16 % PEG 3350] visualized in polarizing light. **(C)** Diffraction image of a single crystal from condition I (B) used for data collection. The black circle depicts a resolution of ~1.7 Å. The image was taken over a Δφ range of 0.5° at beamline PXII, SLS.

Results

Table 3.3: Crystallographic Data

Data Collection	Pml1p[FL]	Pml1p[FL]-Hg	Pml1p[51-204]
Wavelength (Å)	0.9802	1.0095	0.9800
Temperature (K)	100	100	100
Space Group	$P3_221$	$P3_221$	H32
Unit Cell Parameters (Å)	a=b=86.89	a=b=86.20	a=b=100.61
	c=96.70	c=99.11	c=102.20
Resolution (Å)	40.7-2.4 (2.53-2.40)[a]	30.0-2.98 (3.08-2.98)[a]	29.0-1.80 (2.04-1.80)
Reflections			
Unique	16896 (2473)	16703 (1667)	18621 (5539)
Completeness (%)	99.9 (100)	100 (100)	99.7 (100)
Redundancy	6.3 (6.2)	8.9 (8.7)	10.6 (10.3)
I/σ (I)	19.7 (1.6)	15.4 (3.7)	33.7 (7.7)
Rsym (*I*)[b]	0.083 (0.81)	0.235 (0.590)	0.048 (0.283)
Phasing			
Resolution (Å)		3.3	
Heavy Atom Sites		8	
Correlation Coefficients[c]			
SHELXD CC_{all}/CC_{weak}		34.93/16.22	
CC_{free} Left/Right Hand		62.5 % / 46.0 %	
PATFOM[d]		13.92	
Refinement			
Resolution (Å)	40.7-2.4 (2.55-2.4)		29.0-1.8 (1.9-1.8)
Reflections			
Number	16670 (2601)		18614 (2617)
Completeness (%)	99.6 (98.0)		99.6 (99.0)
Test Set (% / reflections)	5.06/853		5.1/949
R_{work}[e]	21.5 (30.2)		18.7 (23.9)
R_{free}[e]	27.2 (32.6)		23.5 (26.9)
ESU (Å)[f]	0.43		0.27
Contents of A.U.[g]			
Protein Molecules/Residues/Atoms	2/280/2327		1/145/1200
Water Oxygens	30		177
Ligand (SO_4^{2-})	2		-
Mean B-Factors (Å^2)			
Wilson	71.71		30.36
Protein	76.95		35.87
Water	70.68		48.52
Ramachandran Plot[h]			
favored (%, residues/residues$_{toltal}$)	93.2 (261/280)		98.6 (143/145)
allowed (%, residues/residues$_{toltal}$)	97.0 (274/280)		100.0 (145/145)
Rmsd[i] from Target Geometry			
Bond Length (Å)	0.007		0.008
Bond Angles (°)	1.147		1.179
Chirality	0.090		0.089
Planarity	0.004		0.005
Dihedral	19.483		16.169

[a] Data for the highest resolution shell in parentheses

[b] $R_{sym}(I) = \Sigma_{hkl}\Sigma_i |I_i(hkl) - <I(hkl)>| / \Sigma_{hkl}\Sigma_i |I_i(hkl)|$; for *n* independent reflections and *i* observations of a given reflection; $<I(hkl)>$ – average intensity of the *i* observations

[c] $CC = [\Sigma wE_oE_c\Sigma w - \Sigma wE_o\Sigma wE_c]/\{[\Sigma wE_o^2\Sigma_w - (\Sigma wE_o)^2] [\Sigma wE_c^2\Sigma w - (\Sigma wE_c)^2]\}^{1/2}$; *w* – weight (see http://shelx.uni-ac.gwdg.de/SHELX/shelx_de.pdf for full definitions).

[d] FOM – figure of merit = $|F(hkl)_{best}| / |F(hkl)|$; $\mathbf{F}(hkl)_{best} = \Sigma_\alpha P(\alpha)\mathbf{F}_{hkl}(\alpha) / \Sigma_\alpha P(\alpha)$

[e] $R = \Sigma_{hkl} ||F_{obs}| - |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$; R_{work} – hkl $\notin$ T; R_{free} – hkl $\in$ T; T – test set

[f] ESU – estimated overall coordinate error based on maximum likelihood

[g] A.U. – asymmetric unit

[h] Calculated with MolProbity (http://molprobity.biochem.duke.edu/)

[i] Rmsd – root-mean-square deviation

3.4.2   Overall structure

Apart from a long, flexible loop (residues 112-121; denoted Δ10 loop hereafter), amino acids 52-204 could be modeled unequivocally into the electron density map of Pml1p^{FL} (Fig. 3.44 A). The remaining N-terminal part (residues 1-51) could not be traced, although the crystals contained the intact polypeptide chain (Fig. 3.45 A) This finding confirmed the biochemical analyses described in section 3.3, suggesting that the first 51 residues of Pml1p are unstructured in the isolated protein.

A VAST homology search (Gibrat et al., 1996) showed that the ordered portion of Pml1p resembles an FHA domain fold (Fig. 3.44 A). Thus, it comprises a short N-terminal α-helix and a sandwich of two twisted β-sheets (strands β1/β2/β7/β8/β10/β11 and strands β3/β4/β5/β6/β9; Fig. 3.44 A). Strands β3 and β4 run parallel to each other, all remaining β-strands are anti-parallel. Two short 3_{10}-helices are inserted between the N-terminal α-helix and strand β1 and between strands β10 and β11.

The structures of the full-length and truncated Pml1p were very similar (Fig. 3.44), exhibiting a root-mean-square deviation (RMSD) of about 0.7 Å for 144 Cα atoms. The flexible loop (Δ10 loop) of full-length Pml1p suspended between strands β3 and β4 was likewise disordered in the structure of the truncated protein.
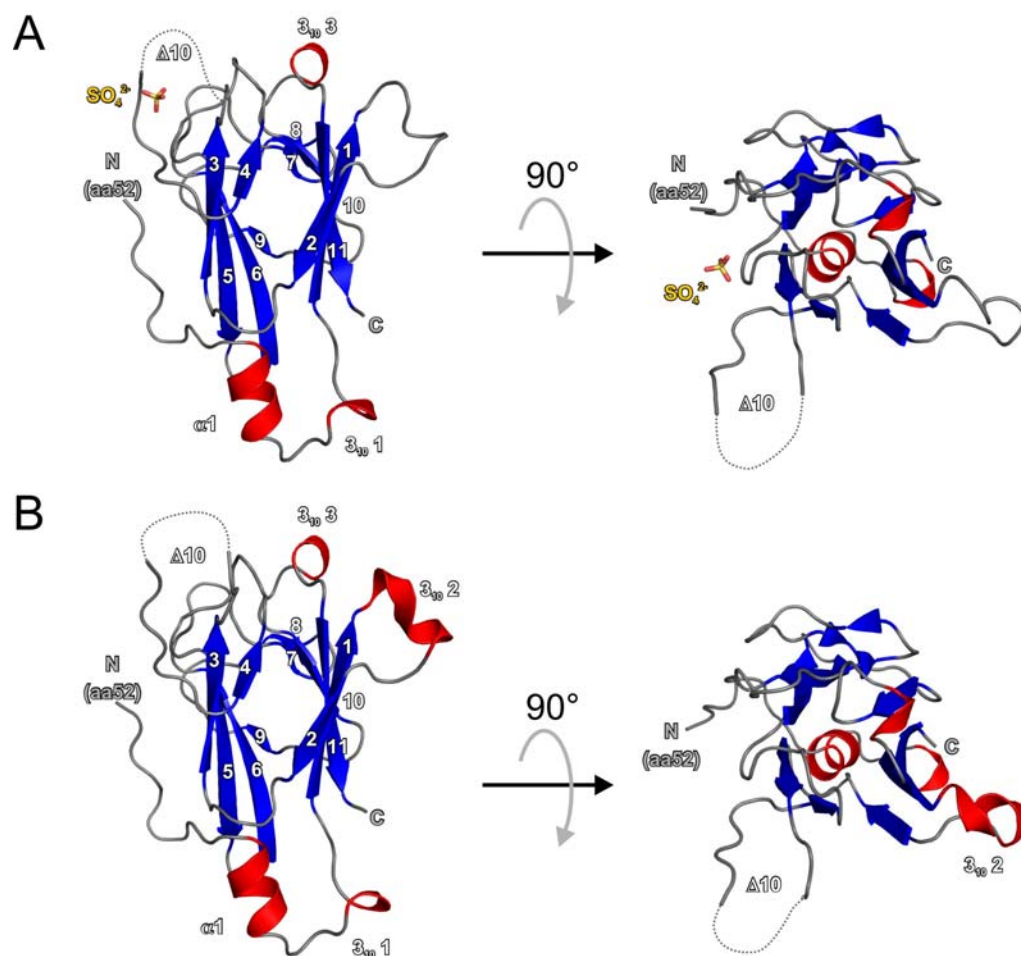
Fig. 3.44: Overview of crystal structures of full-length (Pml1p^{FL}) and truncated Pml1p (Pml1p$^{51-204}$). (A) Cartoon representation of the Pml1p^{FL} structure. β-strands - blue; helices - red; loops - gray. Secondary structure elements are numbered. The coordinated sulfate ion is shown as sticks (sulfur – yellow; oxygen – red). Unstructured regions in the protein showing no electron density are drawn as dashed lines. C - C-terminus; N - N-terminus. **(B)** Structure of truncated Pml1p$^{51-204}$ (residues 51-204) as cartoon representation. Color-code and inscription as in (A). The view on the right is rotated by 90° as indicated. All structure figures were prepared using the program PyMOL (http://www.pymol.org). (taken from Trowitzsch *et al*. 2009).

Crystals of full-length Pml1p contained two molecules in an asymmetric unit that were related by a local twofold axis, forming a V-shaped pseudo-dimer (Fig. 3.45 B). Crystals of the truncated protein showed a similar mode of dimerization via a crystallographic 2-fold. Nevertheless, the apparent higher oligomeric states of Pml1p are most likely a consequence of crystal packing, since gel filtration indicated that the protein is monomeric in solution (Fig. 3.45 C).
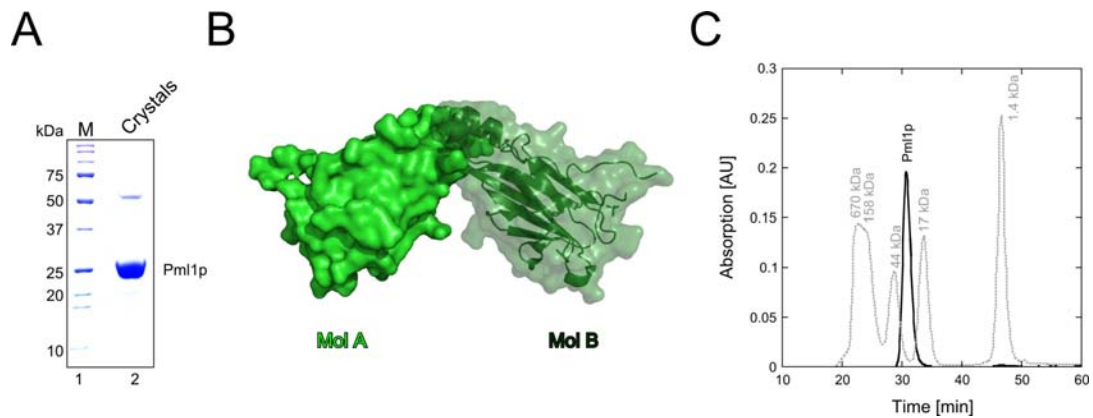
Fig. 3.45: Pml1p crystals spring from monomeric, full-length protein. (A) Crystals of Pml1p molecules were collected, washed, dissolved and proteins were separated on a 15% SDS polyacrylaminde gel. The coomassie stained gel reveals that crystals contained full-length Pml1p molecules (lane 2). Protein size markers are indicated. **(B)** Surface representation of an asymmetric unit of Pml1p crystals. The two molecules (MolA and MolB) of a V-shaped dimer in the crystals are colored green and light green, respectively. The right molecule is shown as a ribbon diagram wrapped by a semi-transparent surface. **(C)** Size exclusion chromatographic analysis of purified full-length Pml1p reveals a monomeric state of the protein in solution. Pml1p - black trace; molecular mass standard - dashed gray trace. The molecular weights of the standard components are indicated. (taken from Trowitzsch *et al.* 2009).

3.4.3 The Pml1p FHA domain is expanded by non-canonical elements

The Pml1p FHA domain (152 amino acids) is considerably larger than most other FHA domain structures determined so far. For example, FHA domains from the yeast checkpoint protein Rad53p (FHA1; (Durocher et al., 2000)), its human homolog Chk2 (Li et al., 2002), the mitotic checkpoint protein Chfr (Stavridi et al., 2002), mammalian PNK (Bernstein et al., 2005), the antigen identified by monoclonal antibody Ki-67 (Ki67; (Byeon et al., 2005)), *Arabidopsis thaliana* kinase-associated protein phosphatase (KAPP) (Lee et al., 2003a) and the putative transcriptional regulator EmbR (Alderwick et al., 2006) encompass between 95 and 123 residues. The increased length of the Pml1p FHA domain is accounted for by several structural insertions and appendices (Fig. 3.46).
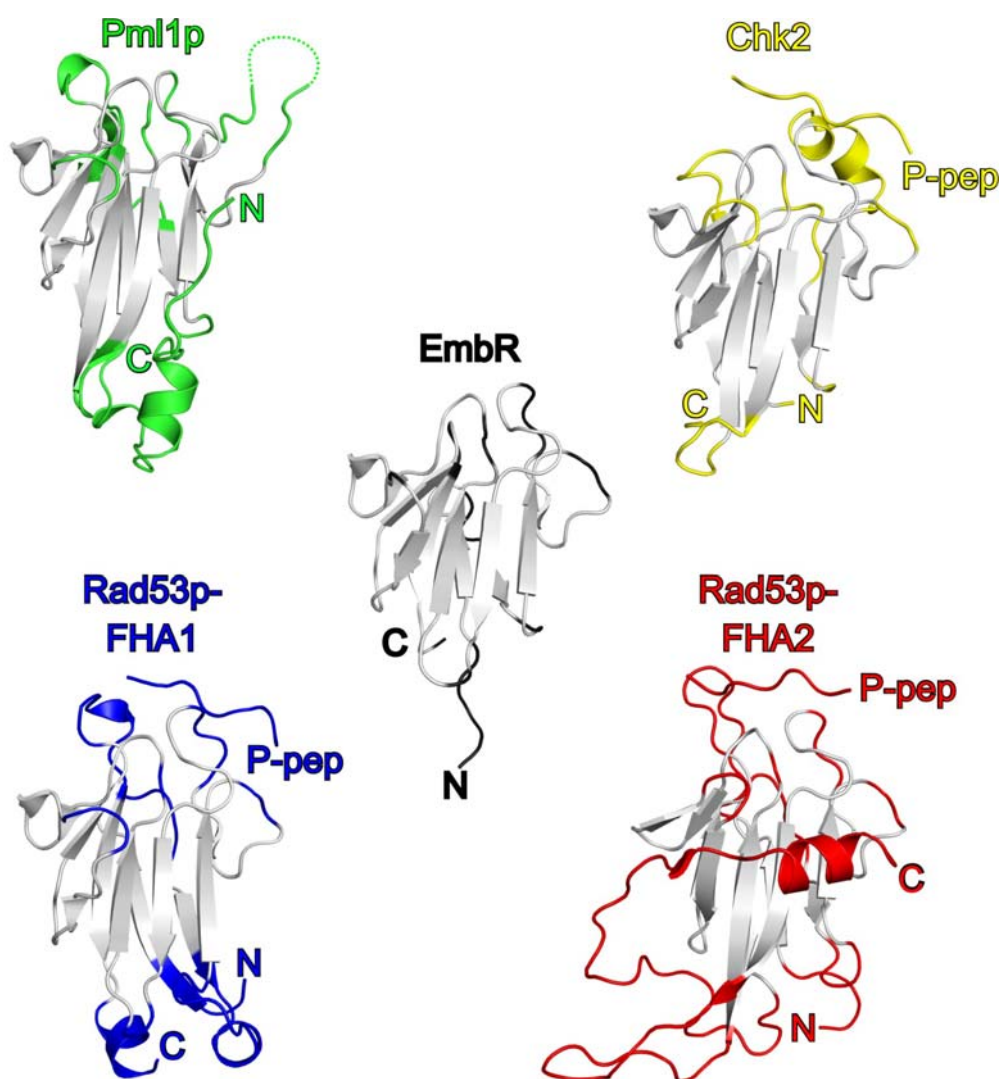
Fig. 3.46: Global comparison of FHA domains from various proteins. FHA domains from Pml1p (green), Chk2 (yellow), EmbR (black) and FHA domains 1 (blue) and 2 (red) from Rad53p are displayed in cartoon representation. Aligning regions in FHA domains after a VAST structure based analysis (Gibrat et al., 1996) are shown in gray. Non-overlapping residues are color-coded. Secondary structure elements were assigned using the DSSP algorithm and displayed in PyMOL. Note the N- and C-terminal extensions in Pml1p, Rad53p FHA1 and Rad53p FHA2 and the helical insertion in Chk2. The small EmbR FHA domain has neither extensions nor insertions. C - C-terminus; N - N-terminus; P-pep - phospho-peptide. Models are as in figure 3.44 A (left), but rotated 70° around x- and 10° around y-axis. (modified from Trowitzsch *et al.* 2009).

Compared to one of the shortest known FHA domain structure, the EmbR FHA domain, Pml1p features a loop insertion between β1 and β2, another long unstructured loop insertion between strands β3 and β4 (Δ10 loop) and a helical insertion between β10 and β11 (Fig. 3.44, Fig. 3.46). In addition, residues 52-61 form an ordered N-terminal appendix that lacks regular secondary structure and lies

like a clamp on top of strands β3 and β5 (Fig. 3.46, Fig. 3.47 A). This N-terminal peptide thereby positions helix α1, which in turn stabilizes the loop between β2 and β3 by hydrogen bonds. Interestingly, a similar N-terminal expansion is seen in the NIPP1 FHA domain but not in other FHA domain proteins analyzed to date (Fig. 3.47 A). Crucial residues, by which the peptide is fastened to the β-sheet, are conserved in NIPP1 and Pml1p (Fig. 3.47 A, Fig. 3.48). I tested the importance of the N-terminal peptide for the integrity of the FHA domain, by generating a truncated Pml1p variant lacking the first 61 residues. $Pml1p^{62-204}$ exhibited dramatically reduced solubility compared to $Pml1p^{FL}$ or $Pml1p^{51-204}$ (Fig. 3.47 B). Therefore, the N-terminal appendix confers critical stability on the Pml1p FHA domain. Similar stabilizing appendices that wrap around a core domain have recently been seen in the Prp8p Jab1/MPN domain (Pena et al., 2007).
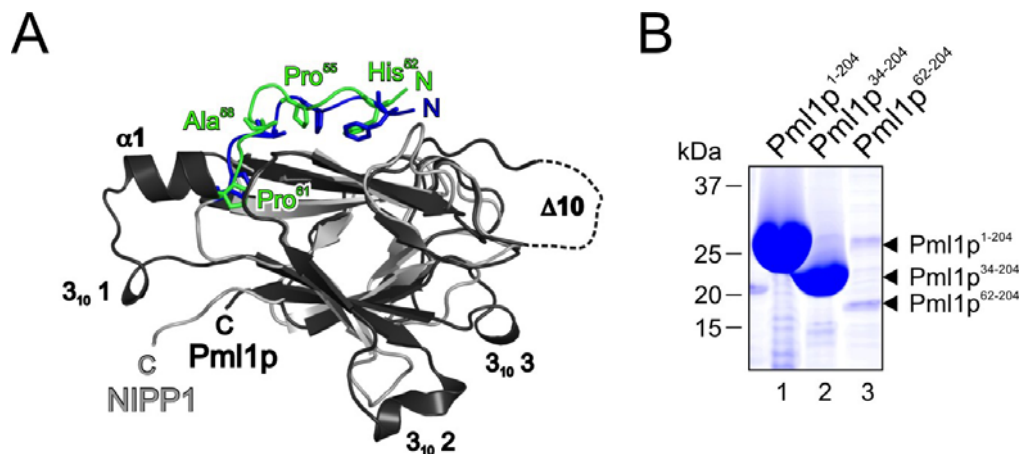
Fig. 3.47: A N-terminal extension relates Pml1p to NIPP1 and is important for Pml1p's integrity. (A) Superposition of Pml1p (anthracite) and NIPP1 (gray) reveals a structurally conserved N-terminal extension present in both structures. Highly conserved residues are shown as sticks (green - Pml1p; blue - NIPP1). Secondary structure elements of Pml1p are indicated for clarity. C = C-terminus; N = N-terminus. Models are as in figure 3.44 A (left), but rotated 100° around z-axis. **(B)** N-terminally truncated, His_6-tagged fragments of Pml1p ($Pml1p^{1-204}$, lane 1; $Pml1p^{34-204}$, lane 2 and $Pml1p^{62-204}$, lane 3; detailed view from figure 3.29 B) were produced under identical conditions in *E. coli*. Proteins were captured on Ni^{2+}-NTA agarose from soluble extract and eluted proteins were analyzed via SDS-PAGE and Coomassie staining. Note that removal of the N-terminal extension of Pml1p (residues 51-61) renders the protein unsoluble. Protein size markers are indicated. (taken from Trowitzsch *et al.* 2009).

Other FHA domains can be expanded by different structural elements (Fig. 3.48). For example, Rad53p FHA2 shows an extended α-helical C-terminal region, which

runs across one of the β-sheets (Wang et al., 2000), and a long insertion connecting strands β4 and β5.
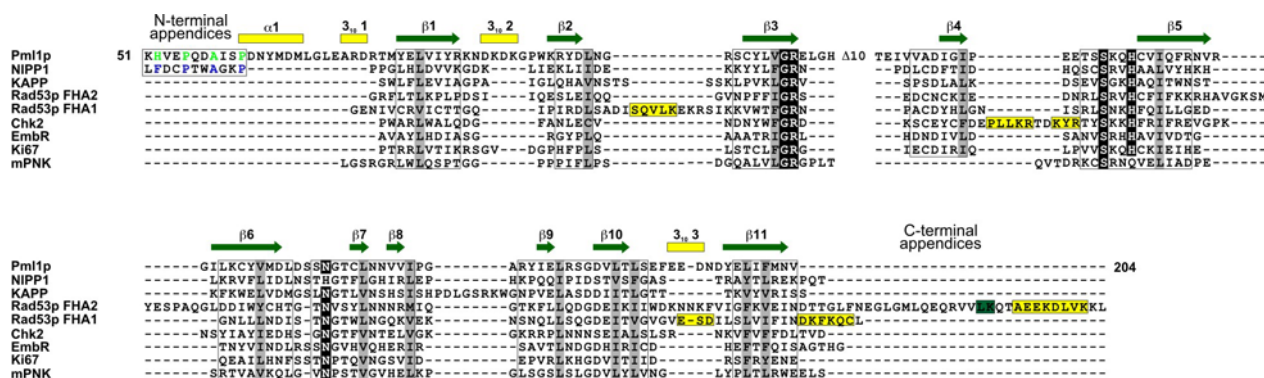
Fig. 3.48: Structure based alignment. Sequences of FHA domain structures from Pml1p (this study), NIPP1 (pdb code 2JPE), KAPP (pdb code 1MZK), Rad53p FHA1 (pdb code 1G6G), Rad53p FHA2 (pdb code 1FHR), Chk2 (pdb code 1GXC), EmbR (pdb code 2FF4), Ki67 (pdb code 2AFF) and mPNK (pdb code 1YJM) were aligned using the VAST server and adjusted manually. N- and C-terminal appendicies are indicated. Secondary structure elements of Pml1p$^{51-204}$ are shown as green arrows (b-sheets) and yellow boxes (a-helices) above the alignment. Secondary structure elements among the other proteins are shown as green boxes (b-sheets) and yellow boxes (a-helices). Structurally aligned residues are boxed (compare also with figure 3.46, gray regions). Gray background corresponds to highly conserved, hydrophobic residues and black background to conserved motifs important for phospho-peptide interaction. (modified from Trowitzsch *et al.* 2009).

3.4.4 The Pml1p FHA domain exhibits hallmarks of a phosphopeptide-binding element

Several FHA domains were shown to function as phosphopeptide-binding modules (Durocher and Jackson, 2002). Six residues that are highly conserved among phosphopeptide-binding FHA proteins are also found in Pml1p (Gly107, Arg108, Ser137, His140, Asn163, and Asn168). With the exception of Asn168, which is thought to stabilize the structure, they lie in surface loops that come together at one edge of the FHA domain (Fig. 3.49 A). Significantly, a sulfate ion is bound to this region in the full-length Pml1p structure, mimicking a bound phospho-amino acid group (Fig. 3.49 A). Other FHA domains have also been observed to coordinate phosphate mimics in the absence of phosphopeptide ligands, such as a tungstate ion in the case of Chfr (Stavridi et al., 2002) (Fig. 3.49 A).

Gly107, Ser137, His140 and Asn163 do not directly interact with the sulfate ion but reinforce the productive conformation of the sulfate-binding region. The highly

conserved Arg108 in loop β3-β4 engages in two salt bridges to sulfate oxygens. The orientation of Arg108 with respect to the sulfate ion is very similar to the disposition of analogous arginines in phosphopeptide-binding FHA domains (Durocher et al., 2000; Stavridi et al., 2002). The conformation of the putative phosphopeptide-binding loops of Pml1p is almost identical in full-length and truncated Pml1p structures, although the latter lacks a sulfate ion (Fig. 3.49 B). Independent of the sulfate, Arg108 seems to be held in position by Ser136.
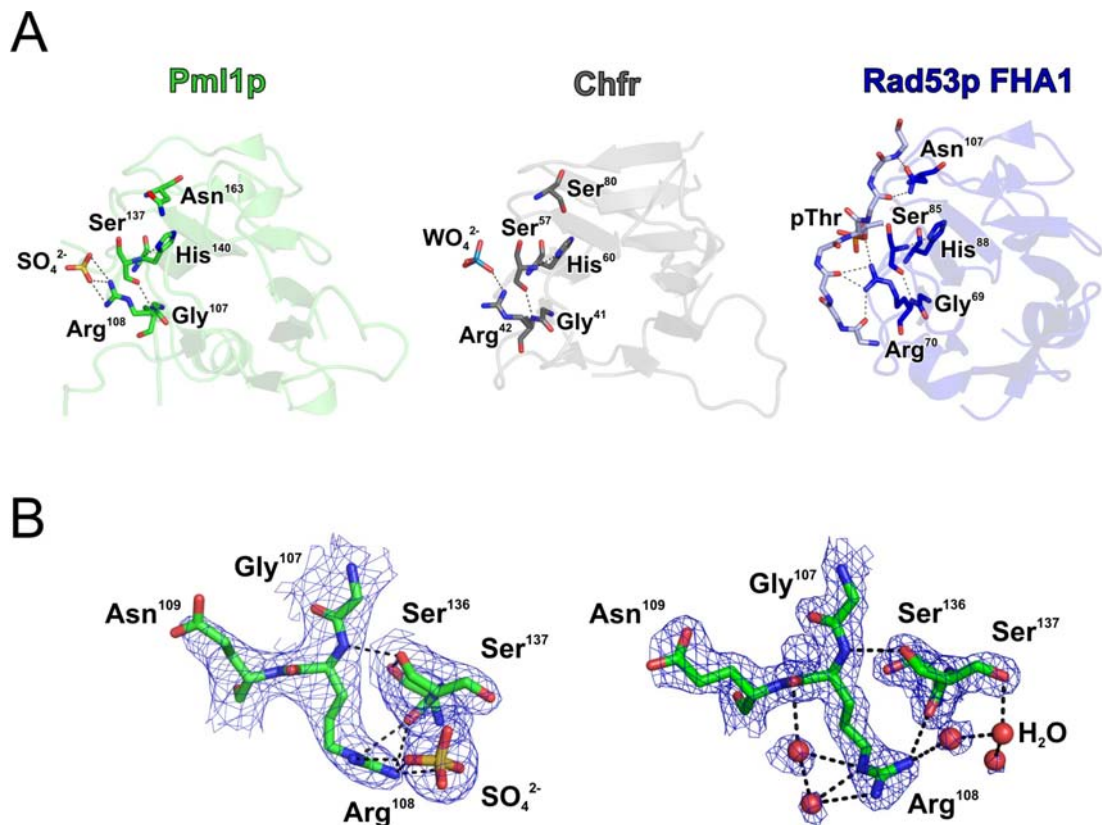
Fig. 3.49: Comparison of the presumed phospho-amino acid binding region of Pml1p FHA. **(A)** Residues of Pml1p (left) and Chk2 (middle) presumably involved in phospho-peptide binding are shown as sticks. Residues of Rad53p FHA1 involved in the interaction with a phospho-peptide (shown as backbone stick model with a phospho-threonin in lightblue) are displayed. Hydrogen bonds are indicated by dashed, black lines. Carbons are colored green (Pml1p), gray (Chfr) and blue/lightblue (Rad53p FHA1), Nitrogens are in blue, oxygens in red, sulfur in yellow and tungstate in cyan. Models are as in figure 3.44 A (left), but rotated 130° around y-axis. **(B)** Electron density of the sulfate-bound (left) and unbound (right) structures displayed 5 Å around Arg108. Dashed black lines represent hydrogen bonds and salt-bridges. Amino acids are shown as sticks. Electron density maps were generated with Coot and displayed in PyMOL at 1 sigma. Carbon – colored green; nitrogen – blue; oxygen – red; sulfur – yellow. In both structures the backbone oxygen of Ser136 is engaged in positioning Arg108. (taken from Trowitzsch *et al.* 2009).

Work on the FHA domain of the receptor kinase-associated protein phosphatase from *A. thaliana* suggested that loops forming the phosphopeptide-binding region rigidify upon ligand binding (Ding et al., 2005). Similarly, loop regions in the FHA domain of Ki67 adopt a fixed structure in the presence of a binding partner (Byeon et al., 2005). NMR chemical shift changes have indicated a similar situation in NIPP1, where several loops are restructured upon phosphopeptide binding (Kumeta et al., 2008). One of these loops corresponds to the flexible Δ10 loop of Pml1p, which neighbors the sulfate-binding site (Fig. 3.50). We therefore suggest that the Δ10 loop helps to determine ligand specificity by engaging in direct contacts with the peptide.
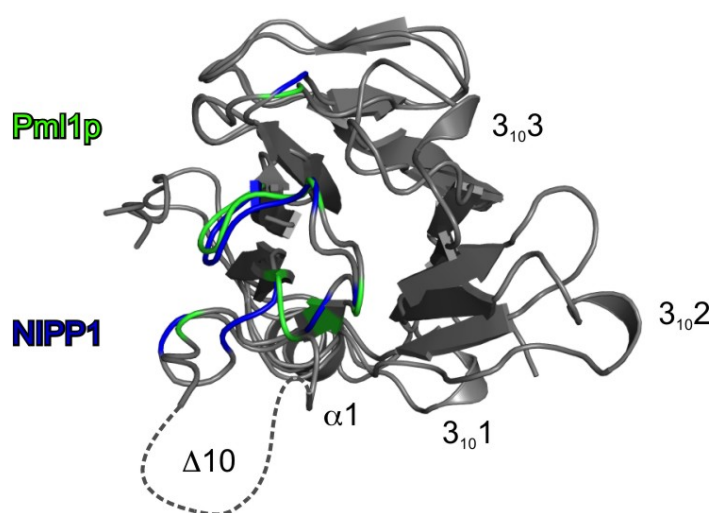
Fig. 3.50: Hotspots of phospho-peptide interaction. Models are as in figure 3.44 A (left), but rotated 70° around x-axis. Residues in NIPP1, which are affected by binding of a phospho-peptide in NMR experiments (Kumeta et al., 2008), are highlighted in blue. Corresponding residues in Pml1p are colored green. Note that an elongated phospho-peptide bound by Pml1p in a similar fashion as to NIPP1 would point towards the unstructured Δ10-loop region. (taken from Trowitzsch *et al.* 2009).

4.   Discussion

4.1   Isolation of native 12S U2 snRNPs and SF3b complexes for crystallographic purposes

12S U2 snRNPs in complex with SF3b and SF3a were isolated biochemically as 17S particles (Kramer et al., 1999; Will et al., 2002). The 17S U2 snRNP binds pre-mRNA, specifically interacting with the conserved branch point sequence (BPS). The branch point adenosine, which is the nucleophile for the first catalytic step of splicing, is thereby specified (Staley, 2002). Multiple cooperative interactions stabilize and position 17S U2 snRNPs on the pre-mRNA, which include protein-protein interactions e.g. binding of SF3b155 to polypyrimidine tract bound U2AF65 (Gozani et al., 1998) and RNA-RNA interactions e.g. duplex formation between the BPS and the U2 snRNA ((Madhani and Guthrie, 1994). Further, protein RNA interactions between several of the 17S U2 snRNP specific proteins and the pre-mRNA are thought to anchor U2 at the BPS (Champion-Arnaud and Reed, 1994; Dybkov et al., 2006; Gozani et al., 1996; Staknis and Reed, 1994).

Biochemical and electron microscopic studies on 17S, 15S and 12S U2 snRNPs have revealed the global architecture of these particles (Kramer et al., 1999). More recently, a high resolution 3D cryo-EM reconstruction of SF3b particles allowed the localization of the components SF3b14a/p14 and SF3b155 (Golas et al., 2003), which were shown to directly contact the branch site and flanking nucleotides, respectively (Gozani et al., 1998).

X-ray crystallographic analyses of components of the 17S U2 snRNP provide excellent structural insights into the interactions between U2-A', U2-B'' and stem loop IV of the U2 snRNA (Price et al., 1998), the arrangement of Sm proteins (Kambach et al., 1999; Toro et al., 2001) and the interaction between SF3b14a/p14 and a N-terminal fragment of SF3b155 (Schellenberg et al., 2006). High resolution X-ray chrystallographic structures of larger particles like SF3b or 12S U2 snRNPs, which would give valuable knowledge about how intra-particle interactions are

established, are not available at present. Thus, one major aim of this study was to establish purification strategies for 12S U2 snRNPs and SF3b particles that are suitable for crystallographic studies.

4.1.1 Purification of 12S U2 snRNP

Useful amounts of 12S U2 snRNPs could be obtained from HeLa nuclear extracts by affinity purification, glycerol gradient centrifugation and ion exchange chromatography using a heparin-Sepharose column (Fig. 3.2). Via this procedure, U1 and U2 snRNP particles could be completely separated, while previous procedures including a DEAE-Sepharose column were inefficient. The heparin-Sepharose column was chosen, since both particles comprise proteins with potential RNA-binding affinity, e.g U2-B" and U1-A. The large, naked portion of U2 snRNA (5' half), which presumably does not contact proteins, most probably contributed to the earlier elution of U2 snRNPs (~250 mM) from heparin-Sepharose columns as compared to U1 snRNPs (~450 mM). Except the 5' splice site-interacting region, no long stretch of naked RNA is present in U1 snRNP. This new purification procedure did not alter the protein composition of 12S U2 snRNPs, since all proteins and the RNA were found in samples of purified particles as judged from SDS-PAGE.

Since samples need to have a significant concentration for crystallization trials (typically 5-20 mg/ml), ultrafiltration of 12S U2 snRNP solutions was attempted as a concentration step. This method turned out to be suitable for increasing the concentration of U2 snRNPs to at least 12 mg/ml. Alternatively, a pelleting approach proved to be as efficient and was also utilized. Interestingly, the behavior of 12S U2 snRNPs in these procedures clearly differs from that of U1 snRNP, which can only be concentrated by pelleting (G. Weber, personal communication). Final purification steps were performed at 250 mM salt, since concentrated 12S U2 snRNP particles tend to precipitate at salt concentrations lower than 120 mM. The aggregation process was reversible, since turbid 12S U2 snRNP solutions could be cleared by the addition of salt from a 5 M stock solution. The tendency of 12S U2 snRNP to aggregate might result from accessible interaction interfaces on U2 snRNP, which are usually occupied by binding partners SF3a and SF3b.

4.1.2   Purification of Δ5' U2 snRNP

Although 12S U2 snRNPs were well behaved during the optimized purification procedure, crystal growth was not observed. The lack of crystals can be accounted for by the presence of unstructured or flexible regions within the 5' part of U2 snRNA. As a consequence, we set to develop strategies to remove the 5' part of the snRNA by various approaches. I could show that a DNA-directed RNaseH cleavage is a powerful method to cut the U2 snRNA 5' of the Sm binding site (section 3.1.2, Fig. 3.3-3.7). Remaining particles were intact, comprising the 3' part of the snRNA and all 12S U2 snRNP proteins, and the digested particle migrated in gel filtrations with an expected size of ~170-220 kDa. A U2 snRNA cleavage approach utilizing an engineered DNAzyme showed a similar pattern of product snRNA fragments as compared to the RNaseH approach. In both cases, the cleaved 5' part of the snRNA was less intensively stained by ethidium bromide after Urea-PAGE, thus less abundant, which was probably a consequence of trace amounts of RNases in the mixtures. Evidence for the identities of the two RNA species was manifold. First, the two RNA species differed in migration behavior during gel electrophoresis (5' RNA part is expected to be shorter). Second, the longer RNA species did not interact with a $m_3^{2,2,7}G$ cap-binding H20 antibody-coupled column, whereas the shorter variant remained bound to the column. Third, the longer RNA species co-migrated with U2 snRNP specific proteins during gel filtration. Last, the longer RNA variant was susceptible to DNA directed RNaseH cleavage with specific oligonucleotides. Thus, truncated 12S U2 snRNPs can be generated comprising half of the U2 snRNA and all U2 snRNP-specific proteins. These processed particles could be further concentrated by both approches, using ultrafiltration devices and ultracentrifugation, and were subjected to crystallization at a concentration of ~12 mg/ml. Crystal growth, however, was not observed so far.

4.1.3   Purification of Δ5' U2 snRNPs with truncated Sm B/B' proteins

Limited proteolysis experiments conducted during this work indicated, that the Sm proteins Sm B/B' are sensitive to an exopeptidase, Carboxypeptidase Y, acting

on the C-termini of polypeptide chains, whereas all other 12S U2 snRNP proteins remained unaffected (section 3.1.3). Thus, C-terminal regions of proteins Sm B/B' most probably lack secondary structure in isolated 12S U2 snRNPs. With the exception of a proline-rich motif in the very C-terminal region of Sm B' (van Dam et al., 1989), proteins Sm B/B' share arginine-glycine (RG) motifs, which contain symmetrical dimethylarginines. The dimethylarginines were shown to be required for binding of the SMN tudor domain (Brahms et al., 2001). Additionally, it was shown, that the U5-52k protein interacts via its GYF-domain with the poly-proline region of Sm B/B' in co-precipitation assays (Kofler et al., 2004). For both modes of binding, interacting peptides of Sm B/B' proteins do not require secondary or tertiary structure, consistent with their high flexibility in isolated U2 snRNPs. It is noteworthy, that a slight difference in processivity of Carboxypeptidase Y could be observed with 12S U2 snRNPs and Δ5' U2 snRNPs allowing the speculation, that C-terminal regions of Sm B/B' proteins might engage in intra-particle interactions with U2 snRNA regions 5' of the Sm binding site.

Although a complete cleavage of C-terminal regions in Sm B/B' by Carboxypeptidase Y can be assumed from gel electrophoretic analysis of processed Δ5' U2 snRNPs (Fig. 3.9), the treatment did not help to obtain crystallizeable complexes. Western blot analysis of processed and unprocessed particles might be necessary to unequivocally validate the absence of C-terminal regions of Sm B/B' in purified Δ5' Sm B$^{\Delta C\text{-term}}$ U2 snRNPs.

4.1.4 U2-A' and U2-B'' as further candidates for controlled proteolysis

Since none of the attempts to crystallize 12S U2 snRNPs so far led to crystal growth, I tested the susceptibility of the U2 snRNP-specific proteins U2-A' and U2-B'' to a broad range of proteases. In isolation, U2-A' was described to harbor a flexible region, which is hydrolyzable by endoproteases (Price et al., 1998). In U1-A, the two RRMs are connected via a long, protease-sensitive linker region (Oubridge et al., 1994). The strong sequence homology of U2-B'' and U1-A suggests a protease-sensitive region between the two RRMs of U2-B'' (Burd and Dreyfuss, 1994).

Protease Asp-N turned out to be perfectly suited for selective hydrolysis of flexible linker sequences of U2-A' and U2-B" in the context of 12S U2 snRNPs (section 3.1.4). By a combination of Carboxypeptidase Y and Asp-N, particles with processed Sm B/B', U2-A' and U2-B" could be obtained. Mass spectrometric analyses of the remaining bands indicated that Sm B/B' were trimmed to the N-terminal Sm motif, and U2-A' and U2-B" to fragments matching the fragments modeled in a crystal structure of the latter two proteins in complex with a short RNA stem loop derived from U2 snRNA (Price et al., 1998). SEC of partially proteolyzed 12S U2 snRNPs further indicated that the C-terminal domains of U2-A' and U2-B" are not involved in protein-protein or protein-RNA contacts and are probably dispensable for the integrity of 12S U2 snRNPs. Interestingly, the yeast homolog of U2-B", YIB9p, lacks the C-terminal RRM (Tang et al., 1996).

4.1.5   Prospectives of 12S U2 snRNP crystallization

Besides flexible regions of the 12S U2 snRNP-specific proteins, the relatively high salt concentration of 250 mM NaCl used during sample preparations might prohibit crystal growth. In order to stabilize highly concentrated U2 snRNPs at low salt concentration, small molecules might be screened in a fluorescence-based approach (Vedadi et al., 2006). Solely by the addition of glycerol, higher particle concentrations were achieved in another case, in which protein concentration was critical for crystal growth (V. Pena, personal communication). Therefore, it is worth trying to find conditions that stabilize U2 snRNPs to achieve higher particle concentrations. As mentioned in the previous chapter, a preparative approach with combined DNA-directed RNaseH or DNAzyme-, Carboxypeptidase Y- and Asp-N-treatments might generate particles, which would lead to crystals suitable for structure solution. Interestingly, sequence inspection of U2 snRNA allows speculation about a presumable intra-U2 snRNA basepairing interaction between loop IIa (ribonucleotides ^{53}UUUAAUAU^{60}) and a region 5' of the Sm motif (ribonuleotides ^{88}AUAUUAAA^{95}). Cross-linking experiments with deproteinized snRNAs isolated from HeLa cells are in agreement with this notion, since crosslinks were found between U^{41}-U^{46} and G^{105}/G^{106} and U^{53} and G^{97}/G^{98} (Datta and Weiner, 1992). Therefore, generating particles with intact stemloops IIa and IIb (cleaving U2

snRNAs 3' of the branch point interacting region [BPiR]) might be a further option to expand the repertoire of potentially crystallizeable material. Figure 4.1 illustrates a summary of processing strategies employed during this work to achieve 12S U2 snRNP particles for crystallization trials.

As a final note, the main bottleneck in the production line of native 12S U2 snRNPs is the relative low yield (1.5 mg from 10^{11} cells) as compared to U1 snRNPs (12 mg from 10^{11} cells). Therefore, an *in vitro* reconstitution approach with recombinant proteins and *in vitro* transcribed RNA, which would allow rational manipulation of each component individually, might be envisioned. A reconstitution approach has already been successfully established for U1 snRNPs (Muto et al., 2001).
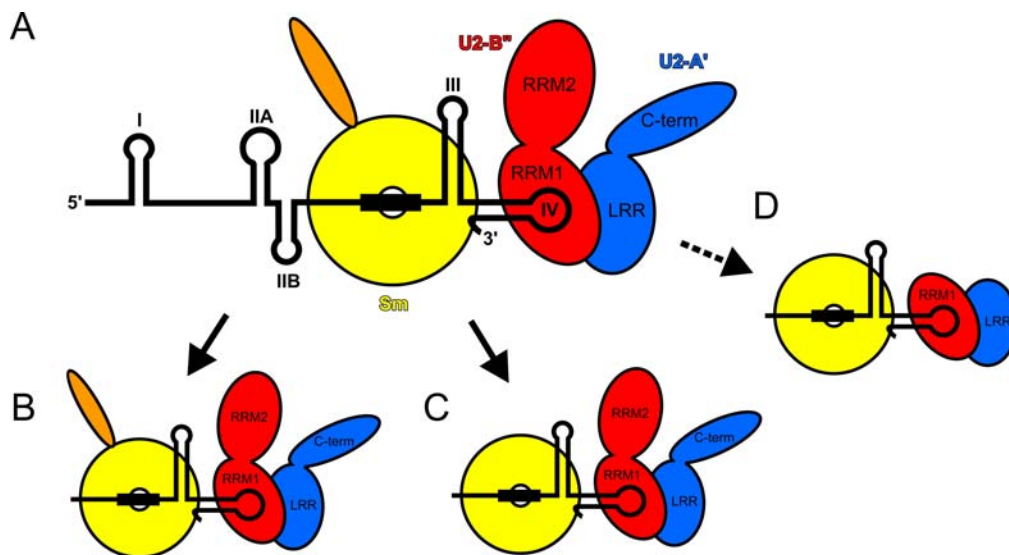
Fig. 4.1: Particle preparation strategies for 12S U2 snRNPs. (A) Schematic representation of 12S U2 snRNPs. U2 snRNA is depicted as black line with the Sm site as a black box. The Sm core is shown in yellow and the C-terminal tail of Sm B/B' is indicated as an orange ellipse. U2-B" and U2-A' are colored red and blue, respectively, and protease -sensitive regions are laced in. Stem loop structures are numbered. **(B)** Schematic representation of Δ5' U2 snRNPs. The snRNA is cleaved 5' of the Sm site. **(C)** Schematic representation of Δ5' Sm B$^{\Delta C\text{-term}}$ U2 snRNPs. **(D)** Schematic representation of Δ5' U2 snRNPs with processed Sm B/B', U2-A' and U2-B" proteins. These particles have not been generated in a preparative manner so far. All color-codes are as in (A).

4.1.6   Splicing factor 3b (SF3b)

SF3b particles could be obtained from HeLa nuclear extracts by a three-step purification protocol including two affinity purification steps and size exclusion chromatography. Protocols for the first as well as the second affinity purification step via an anti-SF3b155 antibody have been established previously in our lab (Will et al., 2001; Will et al., 2002). Details of the purification were kindly provided by Cindy Will. The optimized particle preparation allowed screening of 576 crystallization conditions in nano-drop scale (section 3.2). As for the 12S U2 snRNP, a pelleting protocol could be used for increasing the concentration of SF3b particles. Additionally, the final purification step via glycerol gradient centrifugation as described in (Will et al., 2001) was successfully replaced by a SEC step. Gel filtration combined a purification step by size and a buffer exchange step, which would have been necessary to remove the high glycerol concentration from SF3b-containing gradient fractions. Previous purifications were done in the presence of 600 mM salt (Will et al., 2001), whereas here it was possible to reduce the salt concentration to 150 mM, which is more suitable for crystallization purposes.

Unfortunately, up to now no crystals of SF3b particles have appeared. The lack of crystal growth can be accounted for by the presence of numerous potentially unstructured and low complexity regions among SF3b proteins and heterogeneity of the particle preparation. The N-terminus of SF3b155 was shown to be highly flexible in isolation (Thickman et al., 2006) acting to anchor 17S U2 snRNPs to the pre-mRNA via the splicing factor U2AF (Cass and Berglund, 2006; Gozani et al., 1998). Preliminary limited proteolysis experiments indicated that the N-terminus of SF3b155 is susceptible to the exopeptidase leucin-aminopeptidase in isolated SF3b particles (not shown). Furthermore, SF3b155 is known to be differentially phosphorylated (de Graaf et al., 2006; Wang et al., 1998) and would thereby introduce heterogeneity into the particle preparations. SF3b49 contains a proline-rich region in the C-terminal part of the protein presumably lacking secondary structure.

Evidence has been presented, that SF3b adopts different conformations in isolation and when incorporated into U11/U12 di-snRNPs (Golas et al., 2005). Similar rearrangements within SF3b particles might occur during splicing catalyzed

by the major spliceosome. Two small molecules, Spliceostatin A and pladienolide, were shown to bind to SF3b components and thereby impair splicing (Kaida et al., 2007; Kotake et al., 2007). It is possible that certain conformational states of SF3b components are stabilized upon binding of these small molecules rendering SF3b particles more amenable to crystallization. Detailed studies by limited proteolysis could help to differentiate between Spliceostatin A- or pladienolide-bound and unbound states of SF3b particles and additionally might help in identifying flexible regions of SF3b proteins.

4.1.7 Particle X

Besides SF3b, another particle, denoted particle X, was purified by gel filtration (section 3.2.2). Particle X comprises at least 4 proteins. Peptide mass fingerprints of proteins derived from particle X revealed the presence of the DEAD-box helicase DDX1 and three other stoichiometric proteins: HSPC117, Family with Sequence Specificity 98/member B (FSS98B) and CGI-99.

It has been shown that DDX1 binds the transcriptional control-region of JC viruses and regulates the proliferation of JC viruses *in vitro* through transcriptional activation (Sunden et al., 2007a; Sunden et al., 2007b). CGI-99 belongs to the group of ninein-interacting proteins and blocks phosphorylation of a centrosomal-associated protein, hNinein (Howng et al., 2004). It acts further as an interaction partner of a subunit of the Influenza virus polymerase complex, potentially as a transcriptional activator (Huarte et al., 2001).

Interestingly, the three proteins, DDX1, HSPC117, and CGI-99, were described as part of RNA-transporting, kinesin-associated granules and were predominantly found in the cytoplasm (Kanai et al., 2004). The interaction of DDX1 and CGI-99 was additionally shown to be RNase-resistant (Kanai et al., 2004).

The question of why these proteins co-purifiy with SF3b particles from nuclear extracts, presently cannot be explained yet, because (i) sequence inspection of particle X proteins did not indicate the presence of the α-SF3b155 antibody epitope, (ii) α-SF3b155 antibody recognizes only SF3b155 in nuclear extracts and (iii) none of the proteins belonging to particle X have been identified in A, B or C complex

preparations (Behzadnia et al., 2007; Bessonov et al., 2008; Deckert et al., 2006). Thus, it is still questionable if particle X associates with SF3b particles *in vivo*. Nonetheless, future biochemical and structural work might disclose interesting features of the particle's architecture.

4.2   Pre-mRNA retention and splicing complex (RES)

4.2.1   The RES complex is organized around an unconventional RRM that binds multiple proteins

Work presented here showed that Snu17p is the central subunit of the RES complex, which binds both Bud13p and Pml1p, while the latter two proteins do not interact with each other (section 3.3.1-3.3.2). Bud13p and Pml1p interact with Snu17p exclusively via their C-terminal 65 residues and their N-terminal 61 residues, respectively. Mutational analysis demonstrated that a conserved Trp in Bud13p is essential for complex formation. Structure probing by CD spectroscopy revealed that the interacting portions of both ligand proteins are intrinsically unstructured. The large negative binding entropies detected by ITC suggest that upon complex formation some parts of Bud13p and Pml1p become immobilized and perhaps locally fold into regular secondary structures. Whether larger portions of the otherwise completely unstructured Bud13p fold upon interaction with Snu17p is presently unclear. While Bud13p binds to the core RRM of Snu17p (residues 25-106), Pml1p requires a C-terminal extension in Snu17p for a stable interaction (residues 25-138).

As further detailed below, these findings show that the Snu17p RRM is unconventional in that it supports two different modes of protein-protein interactions at the same time. On the one hand, Snu17p resembles a UHM domain that interacts with an ULM-like motif in Bud13p (Kielkopf et al., 2004). On the other hand, the binding mode between Snu17p and Pml1p recapitulates features of the complex between the human SF3b14a/p14 protein and a peptide of SF3b155 (Schellenberg et al., 2006; Spadaccini et al., 2006).

Preliminary data obtained from NMR analysis and limited proteolysis experiments suggest that Snu17p seems to persist *in vitro* in a molten globule state and completely folds upon binding of a short Pml1p peptide ([28]IMPDFSPSGLL[38]). It is noteworthy that refolded Snu17p molecules failed to interact with Pml1p and Bud13p in GST pull-down experiments and that natively isolated Snu17p molecules did not bind to an N-terminal fragment of Hsh155p[1-117] (not shown). Further experiments are therefore necessary to test whether natively isolated and refolded Snu17p have different binding preferences. These observations would shed new light on how a single protein might act within different environments.

4.2.2   The complex of Snu17p and Bud13p exhibits hallmarks of a UHM-ULM interaction

Compared to canonical RRMs, UHMs exhibit poor conservation of the RNP1 and RNP2 motifs, an Arg-X-Phe motif in the loop preceding the core RRM fold, conserved acidic residues in helix A and a low isoelectric point (the average pI of known UHMs is ~ 4.5) (Kielkopf et al., 2004). All UHM-ULM interactions studied in molecular detail exhibited an invariant Trp of the ULM that was inserted into a hydrophobic pocket between helices A and B of the UHM domain. Structure-based mutational analyses demonstrated that this Trp is essential for the interaction in all examples studied (Selenko et al., 2003; Thickman et al., 2006). In many but not all ULMs, an Arg-Ser dipeptide preceding the invariant Trp is additionally important for the binding to UHMs (Corsini et al., 2007; Kielkopf et al., 2001; Selenko et al., 2003; Thickman et al., 2006). Phosphorylation of the serine residue can modulate the UHM-ULM interaction in some cases. For instance, phosphorylation of Ser20 of splicing factor 1 (SF1) inhibits interaction with U2AF65 and impedes spliceosome assembly (Wang et al., 1999).

The interaction of Snu17p with Bud13p clearly resembles an UHM-ULM interaction. A Trp232Ala mutation in Bud13p completely abolished the interaction with Snu17p. Spectral changes of Trp fluorescence upon complex formation and the high salt stability of the complex suggest that Trp232 comes to lie in a hydrophobic pocket upon binding to Snu17p. Furthermore, as observed for UHMs, the core RRM

of Snu17p (residues 25-106) is sufficient for binding of Bud13p. In addition, the Snu17p-Bud13p complex exhibits a dissociation constant in the nanomolar range (Table 3.2), similar to the K_d determined for the interaction between some other UHM-ULM pairs; e.g., the K_d of the U2AF65-U2AF35 interaction is 1.7 nM (Kielkopf et al., 2001). Finally, Snu17p exhibits a low pI (5.92).

On the other hand, some other typical UHM sequence characteristics are absent from Snu17p. E.g., aromatic amino acid residues of the RNP1 and RNP2 motifs, known to mediate RNA binding in canonical RRMs, are conserved, while the UHM-characteristic Arg-X-Phe motif is lacking. Therefore, we assume that Snu17p belongs to a new subfamily of RRM-containing proteins, which combine features of RNA-binding RRMs and protein-binding UHMs. An important question for the future is whether Snu17p is also able to interact with RNA.

4.2.3 Ablation of the Snu17p-Bud13p interaction has severe cellular consequences

The C-terminal Snu17p-binding portion of Bud13p is the most highly conserved region in the protein, indicating that the interaction between Snu17p and Bud13p is functionally important. In yeast, Bud13p is encoded by the *slc7* gene that has been identified in a screen for mutants defective in bud site selection (Ni and Snyder, 2001). Interestingly, deletion of the gene encoding Snu17p was also shown to result in a similar budding defect (Ni and Snyder, 2001). Furthermore, a Bud13p mutant, encoded by the *slc7-1* allele, showed a synthetic growth defect when combined with a mutant of an essential splicing factor, Clf1 (Vincent et al., 2003). The *slc7-1* allele introduces a nonsense codon into the conserved C-terminal segment of Bud13p at the position of Trp232. Based on our analysis, a mutant Bud13p lacking Trp232 will not be able to interact stably with Snu17p. The splicing defect of the *slc7-1* allele and the budding defect of a Bud13p deletion mutant are, therefore, possibly a consequence of impaired pre-mRNA retention or splicing, brought about by the disruption of the Snu17p-Bud13p interaction. Although little is known about the function of Bud13p in other organisms, RNAi experiments have shown that it is required for *Caenorhabditis elegans* embryogenesis (Jiang et al., 2001). Whether the destruction of the Bud13p-Snu17p complex and a consequent

defect in pre-mRNA retention or splicing are the main cause for these effects remains to be tested.

4.2.4 The interaction of Snu17p and Pml1p resembles the complex between SF3b14a and SF3b155

A short region beyond the core RRM of Snu17p up to residue 138 is required for stable binding of Pml1p. Previously, a very similar portion of SF3b14a, a suggested Snu17p ortholog in the human SF3b complex, was shown to be required for binding of a peptide of SF3b155 (Schellenberg et al., 2006; Spadaccini et al., 2006). Upon interaction with SF3b14a, the otherwise unstructured SF3b155 peptide adopts a mixed α/β structure (Schellenberg et al., 2006), reminiscent of the behavior of the N-terminal Snu17p-binding region of Pml1p. SF3b14a and the SF3b155 peptide interact *via* a stable hydrophobic interface (Schellenberg et al., 2006). The high salt-resistance of the Snu17p-Pml1p complex is also consistent with a strongly hydrophobic interaction mode. Therefore, we suggest that in analogy to the SF3b14a-SF3b155 complex, Snu17p interacts with Pml1p *via* the surface opposite of helices A and B.

I devised a molecular model in order to visualize how Snu17p could concomitantly bind to the two peptide ligands. The expanded RRM motif of Snu17p was homology modeled utilizing the Bioinformatics Toolkit server (http://toolkit.tuebingen.mpg.de/hhpred; Soding et al., 2005) based on the atomic coordinates of the SF3b14a subunit of the SF3b14a-SF3b155 complex crystal structure (PDB ID 2F9D; Schellenberg et al., 2006). The SF3b155 fragment and an SF1 ULM peptide from a complex with the C-terminal UHM domain of U2AF65 (PDB ID 1OPI; Selenko et al., 2003) were then aligned with the model of the Snu17p RRM (Fig. 4.2). In the model, the SF3b155 peptide mimicks the N-terminal Snu17p-binding portion of Pml1p, while the SF1 peptide mimicks the C-terminal Snu17p-binding region of Bud13p. The two ligands come to lie on two neighboring and non-overlapping interaction surfaces (Fig. 4.2).

4.2.5 Snu17p displays functional elements of Bud13p and Pml1p

Notably, the Snu17p-Bud13p interaction makes use of only a short, conserved region in the C-terminus of Bud13p. We suggest that this interaction, therefore, serves to tether other functionally important regions of Bud13p to complexes involved in pre-mRNA retention or splicing. E.g., although Bud13p orthologs show considerable length variation, all exhibit a stretch of consecutive Lys or Arg residues in the N-termini. While the role of this region is presently unclear, its conservation clearly points to an important function.

A large C-terminal portion of Pml1p is dispensable for the interaction with Snu17p. As described in chapter 3.4, crystal structure analysis shows that this C-terminal portion of Pml1p folds into an authentic FHA domain, which exhibits hallmarks of a phosphopeptide binding module. Recapitulating its mode of interaction with Bud13p, Snu17p binds a short peptide of Pml1p and thereby holds on to a functional motif, the FHA domain, in the ligand protein. It is noteworthy that Snu17p and Pml1p interact with a K_d in the micromolar range (Table 3.2), significantly higher than the nanomolar K_d of the Snu17p-Bud13p interaction. Therefore, it is possible that Snu17p recruits Pml1p intermittently as a phosphorylation sensor during pre-mRNA retention or splicing.

4.2.6 Linking pre-mRNA retention and splicing via the RES complex

Presently, the molecular mechanisms, by which unspliced pre-mRNAs are retained in the nucleus are not well understood (Saguez et al., 2005; Sommer and Nehrbass, 2005). Conceivably, a retention factor recognizes portions of the RNAs that mark them as premature transcripts. Snu17p of the RES complex is a candidate for this function. Although we characterized the Snu17p RRM as a protein-binding module, previous examples show that protein interaction does not *a priori* exclude the possibility for concomitant RNA binding by an RRM. E.g., although SPF45 constitutes a UHM that forms a complex with a ULM in SF3b155 (Corsini et al., 2007), it additionally interacts with an AG dinucleotide at the 3'-splice site and promotes its utilization for the second catalytic step (Lallena et al., 2002). Similarly,

the crystal structure of human SF3b14a with a SF3b155 peptide revealed that although the RRM β-sheet of SF3b14a is largely occluded upon complex formation, the branch site adenosine of pre-mRNA could still be cross-linked to Tyr22 of the protein (Schellenberg et al., 2006). In analogy to these examples, Snu17p could have RNA binding activity and might be specifically targeted to unspliced pre-mRNA molecules in the context of RES.

Interestingly, purine-rich exonic splicing enhancers (ESEs), i.e. sequences in an exon that facilitate regulated splicing in higher eukaryotes, were recently found to support retention of RNAs in the nucleus through a saturable nuclear retention factor (Taniguchi et al., 2007). Intron-containing pre-mRNAs use the same retention factor as ESEs (Taniguchi et al., 2007). Perhaps significantly, yeast Snu17p also exhibits splicing enhancer activity (Spingola et al., 2004). While it remains to be seen, whether this function is paralleled by Snu17p orthologs in higher eukaryotes, where regulated splicing takes place, it is conceivable that nuclear retention and splicing enhancement by Snu17p is achieved by binding to the same RNA elements.

Pre-mRNA retention is thought to be mediated by a spliceosome-associated factor (Saguez et al., 2005; Sommer and Nehrbass, 2005). Consistently, the RES complex activates splicing (Dziembowski et al., 2004; Scherrer and Spingola, 2006; Spingola et al., 2004) and was found in precatalytic human spliceosomes (Deckert et al., 2006) and C-complexes (Bessonov et al., 2008). Significantly, Snu17p and Bud13p were found associated with the SF3b complex of yeast U2 snRNP (Wang et al., 2005) and one human ortholog of Snu17p appears to be the SF3b14a/p14 subunit of SF3b, affording a direct link to the spliceosome. Finally, Snu17p knockout leads to a pre-mRNA splicing defect and to abnormally assembled spliceosomes (Gottschalk et al., 2001). Therefore, one important task for the future will be to delineate the surfaces on the RES complex, through which it is tethered to the spliceosome.

4.2.7   Crystal structures of Pml1p

Crystal structure analysis showed, that Pml1p exhibits a FHA domain with non-canonical expansions (Fig. 3.44-46), and, that the ordered portion of the protein is preceded by an intrinsically unstructured Snu17p-interacting region. A short peptide

N-terminal of the canonical FHA domain lies like a clamp across one of the β-sheets and protects the domain against aggregation. A similarly positioned and highly homologous peptide is only found in NIPP1 (Kumeta et al., 2008), another FHA domain protein with a function in pre-mRNA splicing.

FHA domain proteins often contain additional functional elements. For instance, in the *A. thaliana* KAPP, a central FHA domain is flanked by an N-terminal type I membrane anchor and a C-terminal protein phosphatase type 2C (PP2C) catalytic domain (Stone et al., 1994). A flexible linker probably allows free rotational movement between the FHA and PP2C domains (Lee et al., 2003b). Similarly, the PNK FHA domain is thought to be flexibly linked to the kinase/phosphatase catalytic domain (Bernstein et al., 2005). As in KAPP and PNK, tethering of Pml1p to Snu17p *via* a flexible hinge would allow largely unrestricted movement of the FHA domain with respect to the scaffold, even if Snu17p or the Snu17p-Bud13p heterodimer were docked to a massive spliceosome. The FHA domain could thus reorient in order to interact with binding partners. A similar situation was observed in the lateral L7/12 stalk of the 50S ribosomal subunit, in which L7/12 C-terminal domains are flexibly linked to the stalk base allowing them to 'search' neighboring areas for elongation factors (Diaconu et al., 2005).

FHA domains mediate a variety of signal transduction processes by binding phosphoproteins (Durocher and Jackson, 2002). Several structures of FHA domains alone and in complex with phosphopeptide ligands are available (Bernstein et al., 2005; Byeon et al., 2005; Byeon et al., 2001; Durocher et al., 2000; Lee et al., 2003a; Li et al., 2002; Liao et al., 1999; Stavridi et al., 2002; Wang et al., 2000; Yuan et al., 2001). In these structures, phosphopeptides bind in an extended conformation at one side of the domain and the phospho-amino acid-binding site is formed by a set of highly conserved residues. Data presented here suggest that the Pml1p FHA domain can also act as a phosphopeptide-binding domain. First, interaction of Pml1p with Snu17p leaves the canonical phosphopeptide-binding site on the FHA domain available for further interaction with a phosphoprotein (Fig. 4.2). Second, the Pml1p FHA domain exhibits all residues critical for phosphopeptide binding (Fig. 3.49). Third, Pml1p can bind a phosphate-mimicking sulfate ion at the putative phospho-

amino acid binding site (Fig. 3.49). Finally, as shown by yeast two-hybrid analysis, a portion of Pml1p encompassing most of the FHA domain interacts with Spo12, a putative phosphoprotein that regulates meiotic nuclear divisions in budding yeast (Grether and Herskowitz, 1999; Shah et al., 2001).

Conservation of the structure between full-length (sulfate-bound) and truncated (lacking sulfate) Pml1p in the immediate vicinity of the sulfate-binding site strongly argues for a lock-and-key recognition mode for the phospho-amino acid moiety of a putative phophopeptide ligand. On the other hand, the unstructured Δ10 loop is ideally positioned to bind neighboring regions of the phosphopeptide ligand by an induced fit mechanism, thereby increasing specificity. A similar mixed binding strategy has recently been revealed in complexes of the spliceosomal Prp31 protein (Liu et al., 2007).

On the basis of the biochemical and structural analyses presented herein, a structural model of the RES complex, centered on Snu17p, is devised. The FHA domain of Pml1p is anchored with the N-terminal peptide *via* a short flexible linker to the C-terminal expanded RRM of Snu17p (Fig. 4.2). Additionally a model of the intrinsically unstructured Bud13p illustrates the molecular dimensions of the RES complex (Fig. 4.2). This model exemplifys how Snu17p serves as a holding device for functional elements in Pml1p (FHA domain) and possibly in Bud13p (highly conserved N-terminal lysine-rich motif).
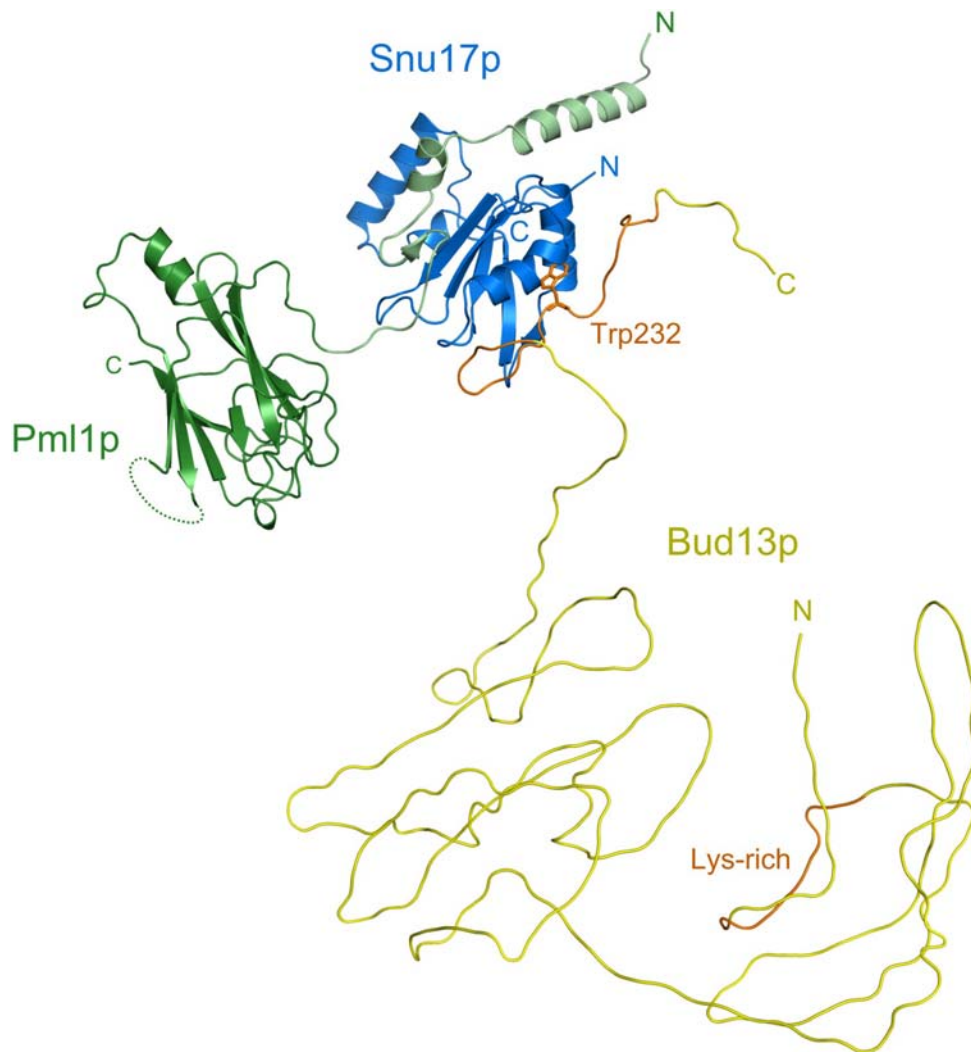
Fig. 4.2: Global model of the RES complex (see Methods). Snu17p, Pml1p and Bud13p are shown as cartoons in blue, green and yellow colors, respectively. The dark green portion of the Pml1p structure has been determined experimentally (residues 51-204). The light green N-terminal part was modeled using the X-ray crystal structure of human SF3b14a-SF3b155 complex as a template (Schellenberg et al., 2006). The Snu17p-Bud13p interaction was modeled according to the solution structure of the U2AF65(RRM3)-SF1 complex (Selenko et al., 2003). Highly conserved regions of Bud13p are colored orange. N - N-termini; C - C-termini.

Work presented herein is a first step towards the elucidation of signaling networks, which may ultimately lead to an understanding of how the RES complex is involved in the regulation of pre-mRNA retention and splicing. Further structural and functional studies will be required to characterize the targets recognized by Pml1p and possible binding partners of the long and unstuctured part of Bud13p.

5.   References

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

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

Adams, P.D., Gopal, K., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Pai, R.K., Read, R.J., Romo, T.D., Sacchettini, J.C., Sauter, N.K., Storoni, L.C. and Terwilliger, T.C. (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. *J Synchrotron Radiat*, **11**, 53-55.

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J.C., Mann, M. and Lamond, A.I. (2000) Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *Embo J*, **19**, 6569-6581.

Alderwick, L.J., Molle, V., Kremer, L., Cozzone, A.J., Dafforn, T.R., Besra, G.S. and Futterer, K. (2006) Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*, **103**, 2558-2563.

Antson, A.A. (2000) Single-stranded-RNA binding proteins. *Curr Opin Struct Biol*, **10**, 87-94.

Bach, M., Bringmann, P. and Lührmann, R. (1990) Purification of small nuclear ribonucleoprotein particles with antibodies against modified nucleosides of small nuclear RNAs. *Methods Enzymol*, **181**, 232-257.

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

Beggs, J.D., Teigelkamp, S. and Newman, A.J. (1995) The role of PRP8 protein in nuclear pre-mRNA splicing in yeast. *J Cell Sci Suppl*, **19**, 101-105.

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

Behrens, S.E., Tyc, K., Kastner, B., Reichelt, J. and Lührmann, R. (1993) Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. *Mol Cell Biol*, **13**, 307-319.

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

Behzadnia, N., Hartmuth, K., Will, C.L. and Luhrmann, R. (2006) Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP. *Rna*, **12**, 1738-1746.

Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J Biol Chem*, **270**, 2411-2414.

Berglund, J.A., Abovich, N. and Rosbash, M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev*, **12**, 858-867.

Bernstein, N.K., Williams, R.S., Rakovszky, M.L., Cui, D., Green, R., Karimi-Busheri, F., Mani, R.S., Galicia, S., Koch, C.A., Cass, C.E., Durocher, D., Weinfeld, M. and Glover, J.N. (2005) The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Mol Cell*, **17**, 657-670.

Bertani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J Bacteriol*, **62**, 293-300.

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

Beullens, M. and Bollen, M. (2002) The protein phosphatase-1 regulator NIPP1 is also a splicing factor involved in a late step of spliceosome assembly. *J Biol Chem*, **277**, 19855-19860.

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

Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci*, **25**, 106-110.

Bochnig, P., Reuter, R., Bringmann, P. and Lührmann, R. (1987) A monoclonal antibody against 2,2,7-trimethylguanosine that reacts with intact, class U, small nuclear ribonucleoproteins as well as with 7- methylguanosine-capped RNAs. *Eur J Biochem*, **168**, 461-467.

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

Boudrez, A., Beullens, M., Groenen, P., Van Eynde, A., Vulsteke, V., Jagiello, I., Murray, M., Krainer, A.R., Stalmans, W. and Bollen, M. (2000) NIPP1-mediated interaction of protein phosphatase-1 with CDC5L, a regulator of pre-mRNA splicing and mitotic entry. *J Biol Chem*, **275**, 25411-25417.

Boudrez, A., Beullens, M., Waelkens, E., Stalmans, W. and Bollen, M. (2002) Phosphorylation-dependent interaction between the splicing factors SAP155 and NIPP1. *J Biol Chem*, **277**, 31834-31841.

Bowers, H.A., Maroney, P.A., Fairman, M.E., Kastner, B., Luhrmann, R., Nilsen, T.W. and Jankowsky, E. (2006) Discriminatory RNP remodeling by the DEAD-box protein DED1. *Rna*, **12**, 903-912.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, **72**, 248-254.

Brahms, H., Meheus, L., de Brabandere, V., Fischer, U. and Luhrmann, R. (2001) Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. *Rna*, **7**, 1531-1542.

Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Haendler, B. and Jacob, M. (1982) U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *Embo J*, **1**, 1259-1265.

Bringmann, P. and Lührmann, R. (1986) Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.*, **5**, 3509-3516.

Bringmann, P., Reuter, R., Rinke, J., Appel, B., Bald, R. and Lührmann, R. (1983a) 5'-terminal caps of snRNAs are accessible for reaction with 2,2,7-trimethylguanosine-specific antibody in intact snRNPs. *J. Biol. Chem.*, **258**, 2745-2747.

Bringmann, P., Rinke, J., Appel, B., Reuter, R. and Lührmann, R. (1983b) Purification of snRNPs U1, U2, U4, U5 and U6 with 2,2,7-trimethylguanosine-specific antibody and definition of their constituent proteins reacting with anti-Sm and anti-(U1)RNP antisera. *EMBO J.*, **2**, 1129-1135.

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

Brow, D.A. (2002) Allosteric cascade of spliceosome activation. *Annu Rev Genet*, **36**, 333-360.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**, 905-921.

Burd, C.G. and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615-621.

Burge, C.B., Tuschl, T. and Sharp, P.A. (1999) Splicing of Precursors to mRNAs by the Spliceosomes. In Gesteland, R.F., Cech, T.R., & Atkins, J.F. (ed.), *The*

RNA world II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY., pp. 525-560.

Byeon, I.J., Li, H., Song, H., Gronenborn, A.M. and Tsai, M.D. (2005) Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67. *Nat Struct Mol Biol*, **12**, 987-993.

Byeon, I.J., Yongkiettrakul, S. and Tsai, M.D. (2001) Solution structure of the yeast Rad53 FHA2 complexed with a phosphothreonine peptide pTXXL: comparison with the structures of FHA2-pYXL and FHA1-pTXXD complexes. *J Mol Biol*, **314**, 577-588.

Caceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L. and Krainer, A.R. (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol*, **138**, 225-238.

Caldecott, K.W. (2003) DNA single-strand break repair and spinocerebellar ataxia. *Cell*, **112**, 7-10.

Cartegni, L., Chew, S.L. and Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet*, **3**, 285-298.

Cass, D.M. and Berglund, J.A. (2006) The SF3b155 N-terminal domain is a scaffold important for splicing. *Biochemistry*, **45**, 10092-10101.

Cech, T.R. (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell*, **44**, 207-210.

Champion-Arnaud, P. and Reed, R. (1994) The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev*, **8**, 1974-1983.

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. *Nucleic Acids Res*, **30**, 1029-1037.

Chu, J.L. and Elkon, K.B. (1991) The small nuclear ribonucleoproteins, SmB and B', are products of a single gene. *Gene*, **97**, 311-312.

Chung, C.T. and Miller, R.H. (1993) Preparation and storage of competent Escherichia coli cells. *Methods Enzymol*, **218**, 621-627.

Cleveland, D.W., Fischer, S.G., Kirschner, M.W. and Laemmli, U.K. (1977) Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J Biol Chem*, **252**, 1102-1106.

Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcarcel, J. and Sattler, M. (2007) U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat Struct Mol Biol*, **14**, 620-629.

Datta, B. and Weiner, A.M. (1992) Cross-linking of U2 snRNA using nitrogen mustard. Evidence for higher order structure. *J. Biol. Chem.*, **267**, 4497-4502.

de Graaf, K., Czajkowska, H., Rottmann, S., Packman, L.C., Lilischkis, R., Luscher, B. and Becker, W. (2006) The protein kinase DYRK1A phosphorylates the splicing factor SF3b1/SAP155 at Thr434, a novel in vivo phosphorylation site. *BMC Biochem*, **7**, 7.

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.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, **6**, 277-293.

Diaconu, M., Kothe, U., Schlunzen, F., Fischer, N., Harms, J.M., Tonevitsky, A.G., Stark, H., Rodnina, M.V. and Wahl, M.C. (2005) Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell*, **121**, 991-1004.

Dietrich, R.C., Incorvaia, R. and Padgett, R.A. (1997) Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol Cell*, **1**, 151-160.

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

Dignam, J.D., Martin, P.L., Shastry, B.S. and Roeder, R.G. (1983b) Eukaryotic gene transcription with purified components. *Methods Enzymol*, **101**, 582-598.

Ding, Z., Lee, G.I., Liang, X., Gallazzi, F., Arunima, A. and Van Doren, S.R. (2005) PhosphoThr peptide binding globally rigidifies much of the FHA domain from Arabidopsis receptor kinase-associated protein phosphatase. *Biochemistry*, **44**, 10119-10134.

Doma, M.K. and Parker, R. (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, **440**, 561-564.

Dong, A., Xu, X., Edwards, A.M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., Filippova, E., Gu, J., Guthrie, J., Ignatchenko, A., Joachimiak, A., Klostermann, N., Kim, Y., Korniyenko, Y., Minor, W., Que, Q., Savchenko, A., Skarina, T., Tan, K., Yakunin, A., Yee, A., Yim, V., Zhang, R., Zheng, H., Akutsu, M., Arrowsmith, C., Avvakumov, G.V., Bochkarev, A., Dahlgren, L.G., Dhe-Paganon, S., Dimov, S., Dombrovski, L., Finerty, P., Jr., Flodin, S., Flores, A., Graslund, S., Hammerstrom, M., Herman, M.D., Hong, B.S., Hui, R., Johansson, I., Liu, Y., Nilsson, M., Nedyalkova, L., Nordlund, P., Nyman, T., Min, J., Ouyang, H., Park, H.W., Qi, C., Rabeh, W., Shen, L., Shen, Y., Sukumard, D., Tempel, W., Tong, Y., Tresagues, L., Vedadi, M., Walker, J.R., Weigelt, J., Welin, M., Wu, H., Xiao, T., Zeng, H. and Zhu, H. (2007) In

situ proteolysis for protein crystallization and structure determination. *Nat Methods*, **4**, 1019-1021.

Du, H. and Rosbash, M. (2002) The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature*, **419**, 86-90.

Durocher, D., Henckel, J., Fersht, A.R. and Jackson, S.P. (1999) The FHA domain is a modular phosphopeptide recognition motif. *Mol Cell*, **4**, 387-394.

Durocher, D. and Jackson, S.P. (2002) The FHA domain. *FEBS Lett*, **513**, 58-66.

Durocher, D., Taylor, I.A., Sarbassova, D., Haire, L.F., Westcott, S.L., Jackson, S.P., Smerdon, S.J. and Yaffe, M.B. (2000) The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol Cell*, **6**, 1169-1182.

Dybkov, O., Will, C.L., Deckert, J., Behzadnia, N., Hartmuth, K. and Luhrmann, R. (2006) U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. *Mol Cell Biol*, **26**, 2803-2816.

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.

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.

Enkhbayar, P., Kamiya, M., Osaki, M., Matsumoto, T. and Matsushima, N. (2004) Structural principles of leucine-rich repeat (LRR) proteins. *Proteins*, **54**, 394-403.

Fan, J.Y., Zhou, J. and Tremethick, D.J. (2007) Quantitative analysis of HP1alpha binding to nucleosomal arrays. *Methods*, **41**, 286-290.

Fenn, J.B. (1993) Ion formation from charged droplets: roles of geometry, energy and time. *J. Am. Soc. Mass Spectrom.*, 524–535.

Fetzer, S., Lauber, J., Will, C.L. and Luhrmann, R. (1997) The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated protein kinase. *Rna*, **3**, 344-355.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer, E.L. and Bateman, A. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res*, **34**, D247-251.

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

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

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.

Fribourg, S., Gatfield, D., Izaurralde, E. and Conti, E. (2003) A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat Struct Biol*, **10**, 433-439.

Frilander, M.J. and Steitz, J.A. (1999) Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. *Genes Dev*, **13**, 851-863.

Frilander, M.J. and Steitz, J.A. (2001) Dynamic Exchanges of RNA Interactions Leading to Catalytic Core Formation in the U12-Dependent Spliceosome. *Mol. Cell*, **7**, 217-226.

Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R. and Dietz, H.C. (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science*, **295**, 2258-2261.

Fu, X.D. (1995) The superfamily of arginine/serine-rich splicing factors. *Rna*, **1**, 663-680.

Fuller-Pace, F.V. (1994) RNA helicases: modulators of RNA structure. *Trends Cell Biol*, **4**, 271-274.

Galy, V., Gadal, O., Fromont-Racine, M., Romano, A., Jacquier, A. and Nehrbass, U. (2004) Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell*, **116**, 63-73.

Gibrat, J.F., Madej, T. and Bryant, S.H. (1996) Surprising similarities in structure comparison. *Curr Opin Struct Biol*, **6**, 377-385.

Golas, M.M., Sander, B., Will, C.L., Lührmann, R. and Stark, H. (2003) Molecular Architecture of the Multiprotein Splicing Factor SF3b. *Science*, **300**, 980-984.

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

Gornemann, J., Kotovic, K.M., Hujer, K. and Neugebauer, K.M. (2005) Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell*, **19**, 53-63.

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

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.

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

Grether, M.E. and Herskowitz, I. (1999) Genetic and biochemical characterization of the yeast spo12 protein. *Mol Biol Cell*, **10**, 3689-3703.

Gunzl, A., Palfi, Z. and Bindereif, A. (2002) Analysis of RNA-protein complexes by oligonucleotide-targeted RNase H digestion. *Methods*, **26**, 162-169.

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

Hall, S.L. and Padgett, R.A. (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J Mol Biol*, **239**, 357-365.

Hall, S.L. and Padgett, R.A. (1996) Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science*, **271**, 1716-1718.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*, **166**, 557-580.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature*, **398**, 579-585.

Hastings, M.L. and Krainer, A.R. (2001) Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol*, **13**, 302-309.

Hayashi, R., Moore, S. and Stein, W.H. (1973) Carboxypeptidase from yeast. Large scale preparation and the application to COOH-terminal analysis of peptides and proteins. *J Biol Chem*, **248**, 2296-2302.

Heinrichs, V., Bach, M. and Lührmann, R. (1990) U1-specific protein C is required for efficient complex formation of U1 snRNP with a 5' splice site. *Mol Biol Rep*, **14**, 165.

Heinrichs, V., Hackl, W. and Lührmann, R. (1992) Direct binding of small nuclear ribonucleoprotein G to the Sm site of small nuclear RNA. Ultraviolet light cross-linking of protein G to the AAU stretch within the Sm site (AAUUUGUGG) of U1 small nuclear ribonucleoprotein reconstituted in vitro. *J Mol Biol*, **227**, 15-28.

Hermann, H., Fabrizio, P., Raker, V.A., Foulaki, K., Hornig, H., Brahms, H. and Lührmann, R. (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J.*, **14**, 2076-2088.

Hofmann, K. and Bucher, P. (1995) The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem Sci*, **20**, 347-349.

Howng, S.L., Hsu, H.C., Cheng, T.S., Lee, Y.L., Chang, L.K., Lu, P.J. and Hong, Y.R. (2004) A novel ninein-interaction protein, CGI-99, blocks ninein phosphorylation by GSK3beta and is highly expressed in brain tumors. *FEBS Lett*, **566**, 162-168.

Huarte, M., Sanz-Ezquerro, J.J., Roncal, F., Ortin, J. and Nieto, A. (2001) PA subunit from influenza virus polymerase complex interacts with a cellular protein with homology to a family of transcriptional activators. *J Virol*, **75**, 8597-8604.

Inada, T. and Aiba, H. (2005) Translation of aberrant mRNAs lacking a termination codon or with a shortened 3'-UTR is repressed after initiation in yeast. *Embo J*, **24**, 1584-1595.

Ingelfinger, D., Arndt-Jovin, D.J., Lührmann, R. and Achsel, T. (2002) The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. *RNA*, **8**, 1489-1501.

Inoue, H., Hayase, Y., Iwai, S. and Ohtsuka, E. (1987) Sequence-dependent hydrolysis of RNA using modified oligonucleotide splints and RNase H. *FEBS Lett*, **215**, 327-330.

Isken, O. and Maquat, L.E. (2007) Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev*, **21**, 1833-1856.

Jackson, I.J. (1991) A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res*, **19**, 3795-3798.

Jady, B.E. and Kiss, T. (2001) A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *Embo J*, **20**, 541-551.

Jagiello, I., Van Eynde, A., Vulsteke, V., Beullens, M., Boudrez, A., Keppens, S., Stalmans, W. and Bollen, M. (2000) Nuclear and subnuclear targeting sequences of the protein phosphatase-1 regulator NIPP1. *J Cell Sci*, **113 Pt 21**, 3761-3768.

Jarmolowski, A. and Mattaj, I.W. (1993) The determinants for Sm protein binding to Xenopus U1 and U5 snRNAs are complex and non-identical. *EMBO J*, **12**, 223-232.

Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V. and Kim, S.K. (2001) Genome-wide analysis of developmental and sex-regulated gene expression profiles in Caenorhabditis elegans. *Proc Natl Acad Sci U S A*, **98**, 218-223.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R. and Shoemaker, D.D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141-2144.

Jones, M.H. and Guthrie, C. (1990) Unexpected flexibility in an evolutionarily conserved protein-RNA interaction: genetic analysis of the Sm binding site. *EMBO J*, **9**, 2555-2561.

Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*, **12**, 5-14.

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

Kabsch, W. (1988) Evaluation of single-crystal X-ray di raction data from a position-sensitive detector. **21**, 916–924.

Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Kadlec, J., Izaurralde, E. and Cusack, S. (2004) The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat Struct Mol Biol*, **11**, 330-337.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., Tani, T., Horinouchi, S. and Yoshida, M. (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol*, **3**, 576-583.

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.

Kanai, Y., Dohmae, N. and Hirokawa, N. (2004) Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron*, **43**, 513-525.

Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*, **60**, 2299-2301.

Kastner, B. (1998) Purification and electron microscopy of spliceosomal snRNPs. *J. Schenkel (ed.), RNP Particles Splicing and Autoimmune Diseases. Springer Lab Manual, Heidelberg*, 95-104.

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., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J.M., Poerschke, D., Luhrmann, R. and Stark, H. (2008) GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods*, **5**, 53-55.

Kastner, B., Kornstadt, U., Bach, M. and Lührmann, R. (1992) Structure of the small nuclear RNP particle U1: identification of the two structural protuberances with RNP-antigens A and 70K. *J. Cell. Biol.*, **116**, 839-849.

References

Kastner, B. and Lührmann, R. (1989) Electron microscopy of U1 small nuclear ribonucleoprotein particles: shape of the particle and position of the 5' RNA terminus. *EMBO J.*, **8**, 277-286.

Kastner, B. and Lührmann, R. (1999) Purification of U small nuclear ribonucleoprotein particles. *Methods Mol Biol*, **118**, 289-298.

Kielkopf, C.L., Lucke, S. and Green, M.R. (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes Dev*, **18**, 1513-1526.

Kielkopf, C.L., Rodionova, N.A., Green, M.R. and Burley, S.K. (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell*, **106**, 595-605.

Kiss, A.M., Jady, B.E., Bertrand, E. and Kiss, T. (2004) Human box H/ACA pseudouridylation guide RNA machinery. *Mol Cell Biol*, **24**, 5797-5807.

Kleywegt, G.J. and Jones, T.A. (1996a) Phi/psi-chology: Ramachandran revisited. *Structure*, **4**, 1395-1400.

Kleywegt, G.J. and Jones, T.A. (1996b) xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr D Biol Crystallogr*, **52**, 826-828.

Kobe, B. and Deisenhofer, J. (1995) Proteins with leucine-rich repeats. *Curr Opin Struct Biol*, **5**, 409-416.

Koch, C.A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A., Weinfeld, M. and Durocher, D. (2004) Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *Embo J*, **23**, 3874-3885.

Kofler, M., Heuer, K., Zech, T. and Freund, C. (2004) Recognition sequences for the GYF domain reveal a possible spliceosomal function of CD2BP2. *J Biol Chem*, **279**, 28292-28297.

Kolossova, I. and Padgett, R.A. (1997) U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA*, **3**, 227-233.

Kong, J. and Liebhaber, S.A. (2007) A cell type-restricted mRNA surveillance pathway triggered by ribosome extension into the 3' untranslated region. *Nat Struct Mol Biol*, **14**, 670-676.

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

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

Kramer, A. (1992) Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. *Mol Cell Biol*, **12**, 4545-4552.

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

Kumeta, H., Ogura, K., Adachi, S., Fujioka, Y., Tanuma, K., Kikuchi, K. and Inagaki, F. (2008) The NMR structure of the NIPP1 FHA domain. *J Biomol NMR*, **40**, 219-224.

Kunkel, G.R., Maser, R.L., Calvet, J.P. and Pederson, T. (1986) U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc Natl Acad Sci U S A*, **83**, 8575-8579.

Kuwasako, K., Dohmae, N., Inoue, M., Shirouzu, M., Taguchi, S., Guntert, P., Seraphin, B., Muto, Y. and Yokoyama, S. (2007) Complex assembly mechanism and an RNA-binding mode of the human p14-SF3b155 spliceosomal protein complex identified by NMR solution structure and functional analyses. *Proteins*, **71**, 1617-1636.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

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

Lai, M.C., Lin, R.I., Huang, S.Y., Tsai, C.W. and Tarn, W.Y. (2000) A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J Biol Chem*, **275**, 7950-7957.

Lallena, M.J., Chalmers, K.J., Llamazares, S., Lamond, A.I. and Valcarcel, J. (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell*, **109**, 285-296.

Lapham, J. and Crothers, D.M. (1996) RNase H cleavage for processing of in vitro transcribed RNA for NMR studies and RNA ligation. *Rna*, **2**, 289-296.

Lau, C.K., Diem, M.D., Dreyfuss, G. and Van Duyne, G.D. (2003) Structure of the Y14-Magoh core of the exon junction complex. *Curr Biol*, **13**, 933-941.

Lee, G.I., Ding, Z., Walker, J.C. and Van Doren, S.R. (2003a) NMR structure of the forkhead-associated domain from the Arabidopsis receptor kinase-associated protein phosphatase. *Proc Natl Acad Sci U S A*, **100**, 11261-11266.

Lee, G.I., Li, J., Walker, J.C. and Van Doren, S.R. (2003b) 1H, 13C and 15N resonance assignments of the kinase-interacting FHA domain of Arabidopsis thaliana kinase-associated protein phosphatase. *J Biomol NMR*, **25**, 253-254.

Letunic, I., Copley, R.R., Pils, B., Pinkert, S., Schultz, J. and Bork, P. (2006) SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res*, **34**, D257-260.

Li, J., Williams, B.L., Haire, L.F., Goldberg, M., Wilker, E., Durocher, D., Yaffe, M.B., Jackson, S.P. and Smerdon, S.J. (2002) Structural and functional

versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. *Mol Cell*, **9**, 1045-1054.

Liao, H., Byeon, I.J. and Tsai, M.D. (1999) Structure and function of a new phosphopeptide-binding domain containing the FHA2 of Rad53. *J Mol Biol*, **294**, 1041-1049.

Lin, R.J., Newman, A.J., Cheng, S.C. and Abelson, J. (1985) Yeast mRNA splicing in vitro. *J. Biol. Chem.*, **260**, 14780-14792.

Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Luhrmann, R., Carlomagno, T. and Wahl, M.C. (2007) Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science*, **316**, 115-120.

Liu, S., Rauhut, R., Vornlocher, H.P. and Luhrmann, R. (2006) The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *Rna*, **12**, 1418-1430.

Loizou, J.I., El-Khamisy, S.F., Zlatanou, A., Moore, D.J., Chan, D.W., Qin, J., Sarno, S., Meggio, F., Pinna, L.A. and Caldecott, K.W. (2004) The protein kinase CK2 facilitates repair of chromosomal DNA single-strand breaks. *Cell*, **117**, 17-28.

Lührmann, R., Appel, B., Bringmann, P., Rinke, J., Reuter, R., Rothe, S. and Bald, R. (1982) Isolation and characterization of rabbit anti-m3 2,2,7G antibodies. *Nucleic Acids Res.*, **10**, 7103-7113.

Ma, X., Yang, C., Alexandrov, A., Grayhack, E.J., Behm-Ansmant, I. and Yu, Y.T. (2005) Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *Embo J*, **24**, 2403-2413.

Madhani, H.D. and Guthrie, C. (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet*, **28**, 1-26.

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

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

Manceau, V., Swenson, M., Le Caer, J.P., Sobel, A., Kielkopf, C.L. and Maucuer, A. (2006) Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65. *Febs J*, **273**, 577-587.

Manley, J.L. and Tacke, R. (1996) SR proteins and splicing control. *Genes Dev*, **10**, 1569-1579.

Maquat, L.E. and Carmichael, G.G. (2001) Quality control of mRNA function. *Cell*, **104**, 173-176.

Maris, C., Dominguez, C. and Allain, F.H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *Febs J*, **272**, 2118-2131.

Maschhoff, K.L. and Padgett, R.A. (1993) The stereochemical course of the first step of pre-mRNA splicing. *Nucleic Acids Res*, **21**, 5456-5462.

Masuyama, K., Taniguchi, I., Okawa, K. and Ohno, M. (2007) Factors associated with a purine-rich exonic splicing enhancer sequence in Xenopus oocyte nucleus. *Biochem Biophys Res Commun*, **359**, 580-585.

Mathew, R., Hartmuth, K., Mohlmann, S., Urlaub, H., Ficner, R. and Luhrmann, R. (2008) Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nat Struct Mol Biol*.

Mayas, R.M., Maita, H. and Staley, J.P. (2006) Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol*, **13**, 482-490.

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

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

Merril, C.R. and Goldman, D. (1982) Quantitative two-dimensional protein electrophoresis for studies of inborn errors of metabolism. *Clin Chem*, **28**, 1015-1020.

Misteli, T. (1999) RNA splicing: What has phosphorylation got to do with it? *Curr Biol*, **9**, R198-200.

Modrek, B. and Lee, C. (2002) A genomic view of alternative splicing. *Nat Genet*, **30**, 13-19.

Moore, M.J., Query, C.C. and Sharp, P.A. (1993) Splicing of precursors to mRNA by the spliceosome. In Gesteland, A. (ed.), *RNA World*. Cold Spring Harbor Labratory Press, Cold Spring Harbor, New York, 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.

Muto, Y., Krummel, D.A., Kambach, C., Hernandez, H., Robinson, C.V. and Nagai, K. (2001) Reconstitution of the spliceosomal U1 snRNP from all recombinant subunits and its characterisation by ionspray Q-tof mass-spectrometry. *Nucleic Acids Res Suppl*, 275-276.

Nagai, K., Muto, Y., Pomeranz Krummel, D.A., Kambach, C., Ignjatovic, T., Walke, S. and Kuglstatter, A. (2001) Structure and Assembly of the Spliceosomal snRNPs. *Biochem Soc Trans*, **29**, 15-26.

Nelissen, R.L., Heinrichs, V., Habets, W.J., Simons, F., Lührmann, R. and van Venrooij, W.J. (1991) Zinc finger-like structure in U1-specific protein C is essential for specific binding to U1 snRNP. *Nucleic Acids Res.*, **19**, 449-454.

References

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

Ni, L. and Snyder, M. (2001) A genomic study of the bipolar bud site selection pattern in Saccharomyces cerevisiae. *Mol Biol Cell*, **12**, 2147-2170.

Nilsen, T.W. (2003) The spliceosome: the most complex macromolecular machine in the cell? *Bioessays*, **25**, 1147-1149.

Nilsen, T.W. (2005) Spliceosome assembly in yeast: one ChIP at a time? *Nat Struct Mol Biol*, **12**, 571-573.

Ohi, M.D., Ren, L., Wall, J.S., Gould, K.L. and Walz, T. (2007) Structural characterization of the fission yeast U5.U2/U6 spliceosome complex. *Proc Natl Acad Sci U S A*, **104**, 3195-3200.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology*, **276**, 307-326.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. and Nagai, K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432-438.

Padgett, R.A., Hardy, S.F. and Sharp, P.A. (1983) Splicing of adenovirus RNA in a cell-free transcription system. *Proc Natl Acad Sci U S A*, **80**, 5230-5234.

Pape, T. and Schneider, T.R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.*, 843-844.

Patel, A.A. and Steitz, J.A. (2003) Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*, **4**, 960-970.

Patton, J.R., Habets, W., van Venrooij, W.J. and Pederson, T. (1989) U1 small nuclear ribonucleoprotein particle-specific proteins interact with the first and second stem-loops of U1 RNA, with the A protein binding directly to the RNA independently of the 70K and Sm proteins. *Mol Cell Biol*, **9**, 3360-3368.

Patton, J.R. and Pederson, T. (1988) The Mr 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 5' stem-loop of U1 RNA and interacts with Sm domain proteins. *Proc Natl Acad Sci U S A*, **85**, 747-751.

Pauling, M.H., McPheeters, D.S. and Ares, M., Jr. (2000) Functional Cus1p is found with Hsh155p in a multiprotein splicing factor associated with U2 snRNA. *Mol. Cell Biol.*, **20**, 2176-2185.

Pellizzoni, L., Kataoka, N., Charroux, B. and Dreyfuss, G. (1998) A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell*, **95**, 615-624.

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

Potterton, E., Briggs, P., Turkenburg, M. and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr*, **59**, 1131-1137.

Price, S.R., Evans, P.R. and Nagai, K. (1998) Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, **394**, 645-650.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I. and Sussman, J.L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, **21**, 3435-3438.

Query, C.C., Bentley, R.C. and Keene, J.D. (1989a) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell*, **57**, 89-101.

Query, C.C., Bentley, R.C. and Keene, J.D. (1989b) A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. *Mol Cell Biol*, **9**, 4872-4881.

Query, C.C., Strobel, S.A. and Sharp, P.A. (1996) Three recognition events at the branch-site adenine. *Embo J*, **15**, 1392-1402.

Rain, J.C. and Legrain, P. (1997) In vivo commitment to splicing in yeast involves the nucleotide upstream from the branch site conserved sequence and the Mud2 protein. *Embo J*, **16**, 1759-1771.

Raker, V.A., Hartmuth, K., Kastner, B. and Lührmann, R. (1999) Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol. Cell Biol.*, **19**, 6554-6565.

Raker, V.A., Plessel, G. and Lührmann, R. (1996) The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *EMBO J.*, **15**, 2256-2269.

Rappsilber, J., Ryder, U., Lamond, A.I. and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res*, **12**, 1231-1245.

Reddy, R., Henning, D., Das, G., Harless, M. and Wright, D. (1987) The capped U6 small nuclear RNA is transcribed by RNA polymerase III. *J Biol Chem*, **262**, 75-81.

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

Reed, R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev*, **6**, 215-220.

Reed, R. (2003) Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol*, **15**, 326-331.

Reed, R. and Maniatis, T. (1988) The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev*, **2**, 1268-1276.

References

Reed, R. and Palandjian, L. (1997) Spliceosome assembly. *Eukaryotic mRNA processing*, 103-129.

Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) Excision of an intact intron as a novel lariat structure during pre- mRNA splicing in vitro. *Cell*, **38**, 317-331.

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

Rutz, B. and Seraphin, B. (2000) A dual role for BBP/ScSF1 in nuclear pre-mRNA retention and splicing. *Embo J*, **19**, 1873-1886.

Rymond, B. (2007) Targeting the spliceosome. *Nat Chem Biol*, **3**, 533-535.

Rymond, B.C. and Rosbash, M. (1992) The molecular and cellular biology of the yeast Saccaromyces. *Broach, J.R. Cold Spring Harbor, New York, Cold Spring Harbor Lab. ,* **2**, 143-192.

Saguez, C., Olesen, J.R. and Jensen, T.H. (2005) Formation of export-competent mRNP: escaping nuclear destruction. *Curr Opin Cell Biol*, **17**, 287-293.

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

Sander, B., Golas, M.M., Makarov, E.M., Brahms, H., Kastner, B., Luhrmann, R. and Stark, H. (2006) Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol Cell*, **24**, 267-278.

Santoro, S.W. and Joyce, G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci U S A*, **94**, 4262-4266.

Santoro, S.W. and Joyce, G.F. (1998) Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry*, **37**, 13330-13342.

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

Scherrer, F.W., Jr. and Spingola, M. (2006) A subset of Mer1p-dependent introns requires Bud13p for splicing activation and nuclear retention. *Rna*, **12**, 1361-1372.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A*, **95**, 5857-5864.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A. and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell*, **11**, 965-976.

Séraphin, B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.*, **14**, 2089-2098.

Shah, R., Jensen, S., Frenz, L.M., Johnson, A.L. and Johnston, L.H. (2001) The Spo12 protein of Saccharomyces cerevisiae: a regulator of mitotic exit whose cell cycle-dependent degradation is mediated by the anaphase-promoting complex. *Genetics*, **159**, 965-980.

Sharp, P.A. and Burge, C.B. (1997) Classification of introns: U2-type or U12-type. *Cell*, **91**, 875-879.

Sheldrick, G.M. (2008) A short history of SHELX. *Acta Crystallogr A*, **64**, 112-122.

Sherman, F., Fink, G.R. and Hicks, J. (1986) Methods in Yeast Genetics. *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.*, 128.

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., Reddy, B. and Manley, J.L. (2006) PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. *Mol Cell*, **23**, 819-829.

Sillekens, P.T., Habets, W.J., Beijer, R.P. and van Venrooij, W.J. (1987) cDNA cloning of the human U1 snRNA-associated A protein: extensive homology between U1 and U2 snRNP-specific proteins. *Embo J*, **6**, 3841-3848.

Singh, R. and Reddy, R. (1989) Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc Natl Acad Sci U S A*, **86**, 8280-8283.

Singh, R., Valcarcel, J. and Green, M.R. (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science*, **268**, 1173-1176.

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

Soding, J., Biegert, A. and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, **33**, W244-248.

Sommer, P. and Nehrbass, U. (2005) Quality control of messenger ribonucleoprotein particles in the nucleus and at the pore. *Curr Opin Cell Biol*, **17**, 294-301.

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

Spiess, E. (1978) The influence of concentrating methods on electron microscopical imaging of negatively-stained 50 S ribosomal subunits of Escherichia coli. *FEBS Lett*, **91**, 289-292.

References

Spingola, M., Armisen, J. and Ares, M., Jr. (2004) Mer1p is a modular splicing factor whose function depends on the conserved U2 snRNP protein Snu17p. *Nucleic Acids Res*, **32**, 1242-1250.

Staknis, D. and Reed, R. (1994) Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol Cell Biol*, **14**, 2994-3005.

Staley, J.P. (2002) Hanging on to the branch. *Nat Struct Biol*, **9**, 5-7.

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

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

Stavridi, E.S., Huyen, Y., Loreto, I.R., Scolnick, D.M., Halazonetis, T.D., Pavletich, N.P. and Jeffrey, P.D. (2002) Crystal structure of the FHA domain of the Chfr mitotic checkpoint protein and its complex with tungstate. *Structure*, **10**, 891-899.

Stebbins, C.E., Kaelin, W.G., Jr. and Pavletich, N.P. (1999) Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science*, **284**, 455-461.

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

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

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.

Stone, J.M., Collinge, M.A., Smith, R.D., Horn, M.A. and Walker, J.C. (1994) Interaction of a protein phosphatase with an Arabidopsis serine-threonine receptor kinase. *Science*, **266**, 793-795.

Sunden, Y., Semba, S., Suzuki, T., Okada, Y., Orba, Y., Nagashima, K., Umemura, T. and Sawa, H. (2007a) DDX1 promotes proliferation of the JC virus through transactivation of its promoter. *Microbiol Immunol*, **51**, 339-347.

Sunden, Y., Semba, S., Suzuki, T., Okada, Y., Orba, Y., Nagashima, K., Umemura, T. and Sawa, H. (2007b) Identification of DDX1 as a JC virus transcriptional control region-binding protein. *Microbiol Immunol*, **51**, 327-337.

Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, **93**, 139-148.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y. and Yoshida, T. (1988) Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-flight Mass Spectrometry. *Rapid Comm. Mass Spectrom.*, **2**, 151–153.

Tanaka, K., Watakabe, A. and Shimura, Y. (1994) Polypurine sequences within a downstream exon function as a splicing enhancer. *Mol Cell Biol*, **14**, 1347-1354.

Tang, J., Abovich, N. and Rosbash, M. (1996) Identification and characterization of a yeast gene encoding the U2 small nuclear ribonucleoprotein particle B" protein. *Mol Cell Biol*, **16**, 2787-2795.

Taniguchi, I., Masuyama, K. and Ohno, M. (2007) Role of purine-rich exonic splicing enhancers in nuclear retention of pre-mRNAs. *Proc Natl Acad Sci U S A*, **104**, 13684-13689.

Tardiff, D.F., Lacadie, S.A. and Rosbash, M. (2006) A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol Cell*, **24**, 917-929.

Tarn, W.Y., Hsu, C.H., Huang, K.T., Chen, H.R., Kao, H.Y., Lee, K.R. and Cheng, S.C. (1994) Functional association of essential splicing factor(s) with PRP19 in a protein complex. *Embo J*, **13**, 2421-2431.

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

Tarn, W.Y. and Steitz, J.A. (1996b) Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science*, **273**, 1824-1832.

Teigelkamp, S., Mundt, C., Achsel, T., Will, C.L. and Luhrmann, R. (1997) The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *Rna*, **3**, 1313-1326.

Thickman, K.R., Swenson, M.C., Kabogo, J.M., Gryczynski, Z. and Kielkopf, C.L. (2006) Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. *J Mol Biol*, **356**, 664-683.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.

Toro, I., Thore, S., Mayer, C., Basquin, J., Séraphin, B. and Suck, D. (2001) RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J*, **20**, 2293-2303.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. and Rothberg, J.M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, **403**, 623-627.

Urlaub, H., Hartmuth, K., Kostka, S., Grelle, G. and Lührmann, R. (2000) A general approach for identification of RNA-protein crosslinking sites within native

human spliceosomal snRNPs: Analysis of RNA-protein contacts in native U1 and [U4/U6.U5] snRNPs. *J. Biol. Chem.*

Urlaub, H., Raker, V.A., Kostka, S. and Lührmann, R. (2001) Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J*, **20**, 187-196.

Vagin, A. and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr*, **56**, 1622-1624.

Valadkhan, S. (2007) The spliceosome: caught in a web of shifting interactions. *Curr Opin Struct Biol*, **17**, 310-315.

van Dam, A., Winkel, I., Zijlstra-Baalbergen, J., Smeenk, R. and Cuypers, H.T. (1989) Cloned human snRNP proteins B and B' differ only in their carboxy-terminal part. *Embo J*, **8**, 3853-3860.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol*, **229**, 105-124.

van Hoof, A., Frischmeyer, P.A., Dietz, H.C. and Parker, R. (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science*, **295**, 2262-2264.

Vedadi, M., Niesen, F.H., Allali-Hassani, A., Fedorov, O.Y., Finerty, P.J., Jr., Wasney, G.A., Yeung, R., Arrowsmith, C., Ball, L.J., Berglund, H., Hui, R., Marsden, B.D., Nordlund, P., Sundstrom, M., Weigelt, J. and Edwards, A.M. (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A*, **103**, 15835-15840.

Vidal, V.P., Verdone, L., Mayes, A.E. and Beggs, J.D. (1999) Characterization of U6 snRNA-protein interactions. *Rna*, **5**, 1470-1481.

Vincent, K., Wang, Q., Jay, S., Hobbs, K. and Rymond, B.C. (2003) Genetic interactions with CLF1 identify additional pre-mRNA splicing factors and a link between activators of yeast vesicular transport and splicing. *Genetics*, **164**, 895-907.

Walter, T.S., Meier, C., Assenberg, R., Au, K.F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I. and Grimes, J.M. (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure*, **14**, 1617-1622.

Wang, C., Chua, K., Seghezzi, W., Lees, E., Gozani, O. and Reed, R. (1998) Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes Dev*, **12**, 1409-1414.

Wang, P., Byeon, I.J., Liao, H., Beebe, K.D., Yongkiettrakul, S., Pei, D. and Tsai, M.D. (2000) II. Structure and specificity of the interaction between the FHA2 domain of Rad53 and phosphotyrosyl peptides. *J Mol Biol*, **302**, 927-940.

Wang, Q., He, J., Lynn, B. and Rymond, B.C. (2005) Interactions of the yeast SF3b splicing factor. *Mol Cell Biol*, **25**, 10745-10754.

Wang, Q. and Rymond, B.C. (2003) Rds3p is required for stable U2 snRNP recruitment to the splicing apparatus. *Mol Cell Biol*, **23**, 7339-7349.

Wang, X., Bruderer, S., Rafi, Z., Xue, J., Milburn, P.J., Kramer, A. and Robinson, P.J. (1999) Phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly. *Embo J*, **18**, 4549-4559.

Watakabe, A., Tanaka, K. and Shimura, Y. (1993) The role of exon sequences in splice site selection. *Genes Dev*, **7**, 407-418.

Weber, K., Pringle, J.R. and Osborn, M. (1972) Measurement of molecular weights by electrophoresis on SDS-acrylamide gel. *Methods Enzymol*, **26 PtC**, 3-27.

Weiner, A.M. (1993) mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell*, **72**, 161-164.

Westerholm-Parvinen, A., Vernos, I. and Serrano, L. (2000) Kinesin subfamily UNC104 contains a FHA domain: boundaries and physicochemical characterization. *FEBS Lett*, **486**, 285-290.

Will, C.L. and Luhrmann, R. (1999) Splice Sites. *Encyclopedia of Molecular Biology. Creighton, T.E., New York, John Wiley & Sons, Inc.*, 2411-2413.

Will, C.L. and Luhrmann, R. (2001a) Molecular biology. RNP remodeling with DExH/D boxes. *Science*, **291**, 1916-1917.

Will, C.L. and Luhrmann, R. (2001b) Spliceosomal UsnRNP biogenesis, structure and function. *Current Opinion in Cell Biology.*, **13**, 290-301.

Will, C.L. and Lührmann, R. (2006) Spliceosome structure and function. In Gesteland, R.F., Cech, T.R. and Atkins, J.F. (eds.), *The RNA world*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 369–400.

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

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

Woppmann, A., Patschinsky, T., Bringmann, P., Godt, F. and Luhrmann, R. (1990) Characterisation of human and murine snRNP proteins by two-dimensional gel electrophoresis and phosphopeptide analysis of U1-specific 70K protein variants. *Nucleic Acids Res*, **18**, 4427-4438.

Wu, Q. and Krainer, A.R. (1997) Splicing of a divergent subclass of AT-AC introns requires the major spliceosomal snRNAs. *RNA*, **3**, 586-601.

Wu, S., Romfo, C.M., Nilsen, T.W. and Green, M.R. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, **402**, 832-835.

References

Xiao, S.H. and Manley, J.L. (1998) Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *Embo J*, **17**, 6359-6367.

Xu, R., Teng, J. and Cooper, T.A. (1993) The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol Cell Biol*, **13**, 3660-3674.

Yaffe, M.B. and Cantley, L.C. (1999) Signal transduction. Grabbing phosphoproteins. *Nature*, **402**, 30-31.

Yu, Y.T. and Steitz, J.A. (1997) Site-specific crosslinking of mammalian U11 and u6atac to the 5' splice site of an AT-AC intron. *Proc. Natl. Acad. Sci. USA*, **94**, 6030-6035.

Yuan, C., Yongkiettrakul, S., Byeon, I.J., Zhou, S. and Tsai, M.D. (2001) Solution structures of two FHA1-phosphothreonine peptide complexes provide insight into the structural basis of the ligand specificity of FHA1 from yeast Rad53. *J Mol Biol*, **314**, 563-575.

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

Zhang, M., Zamore, P.D., Carmo-Fonseca, M., Lamond, A.I. and Green, M.R. (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc Natl Acad Sci U S A*, **89**, 8769-8773.

Zorio, D.A. and Blumenthal, T. (1999) Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature*, **402**, 835-838.

6. Appendix

6.1 List of abbreviations

°C	degree centigrade
α	alpha or anti
AIDA	Advanced Image Data Analyzer
Amp^R	ampicillin resistance
APS	ammonium peroxodisulfate
ATAC introns	introns, recognized by minor spliceosomes
ATP	adenosintriphosphate
BBP	branch point binding protein
BER	base excision repair
bp	base pair
BP	branch point
BPS	branch point sequence
BS	branch site
BSiR	Branch site interacting region
Bud13p	bud site selection 13
C	cytosine, carbon, carboxy
C.e.	*Caenorhabditis elegans*
Cg	*Candida glabra*
ca.	circa
Cam^R	chloramphenicol resistance
CCD	charge-coupled device
CD	circular dichroism
CDC5	cell division control protein 5
cDNA	chromosomal DNA
Chfr	checkpoint with FHA and RING finger
Chk2	checkpoint protein 2
CNS	Crystallography & NMR System
Dm	*Drosophila melanogaster*
Da	dalton
dd	double distilled
ddH$_2$O	double distilled water
DExD/H, DEAD	consensus sequences of helicases
DMSO	dimethylsulfoxide
DNA	desoxyribonucleic acid
dNTP	desoxynucleotide-5'-triphosphate
DTE	dithioerythrol
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediamintetraacetate
EM	electron microscopy

EmbR	ethambultol resistance
EPPS buffer	4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid
ESE	exonic splicing enhancer
ESI	electrospray ionisation
ff	following pages
FHA	forkhead-associated
Gg	*Gallus gallus*
GST	Glutahion-S-transferase
HEPES	2-[4-(2-Hydroxyethyl)-1-piperazino]-ethansulfonic acid
Hs	*Homo sapiens*
HPLC	high pressure liquid chromatography
hPrp31	pre-mRNA processing factor 31
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl-β-D-thiogalactopyranoside
ITC	isothermal titration calorimetry
Kan^R	kanamycin resistance
KAPP	kinase-associated protein phosphatase
Ki67	antigen identified by monoclonal antibody Ki-67
Kl	*Klyveris lactis*
LB	lysogeny broth (Bertani, 1951)
M	molar
M. m.	*Mus musculus*
MALDI	matrix-assisted laser desorption/ionisation
MES	2-morpholinoethanesulfonic acid
min	minutes
mM	milli molar
MS	mass spectrometry
Mm	*Mus musculus*
MWCO	molecular weight cut off
N	amino
NEB	New England Biolabs
Nc	*Neurospora crassa*
ng	nano gramms
NGD	No-Go decay
NHEJ	non-homologous end joining
NIPP1	nuclear inhibitor of protein phosphatase 1
nm	nano meter
NMD	non-sense-mediated decay
NMR	nuclear magnetic resonance
NSD	nonstop decay
NTA	Nitrilo-triacetic acid
OD	optic density
OH	hydroxyl
ORF	open reading frame
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis
PCI	phenol-chloroform-isoamylalcolol

PCR	polymerase chain raection
PEG	polyethylene glycole
PG	propylene glycol
Pml1p	pre-mRNA leakage 1
PMSF	phenylmethylsulfonylfluoride
PNK	polynucleotide kinase
PP2C	protein phosphatase type 2C
pre-mRNA	precursor messenger RNA
Prp19p	pre-mRNA processing factor 19
REMD	ribosome extension-mediated decay
RES	pre-mRNA retention and splicing
RMSD	root-mean-square deviation
RNA	Ribonucleic acid
RNaseH	Ribonuclease H
rpm	revlutions per minute
RRM	RNA recognition motif
RT	room temperature
S. c. / *Sc*	*Saccharomyces cherevisiae*
SAD	single-wavelength anomalous dispersion
SDS	sodium dodecyl sulfate
sec	seconds
SF1	splicing factor 1
SF3a	splicing factor 3a
SF3b	splicing factor 3b
SF3b155	splicing factor 3b subunit of 155 kDa
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle
Snu17p	small nuclear ribonucleoprotein associated 17
Sp	*Schizosaccharomyces pombe*
SR-protein	arginine-serine dipeptide rich splicing factor
Strep^R	streptomycin resistance
TAP	tandem affinity purification
TBS	Tris-buffered saline
TEMED	N,N,N',N'-Tetramethylethylenediamide
TEV	Tobacco Etch Virus
Tris	Tris-(hydroxymethyl)aminomethane
U2AF	U2 auxilliary factor
U2AF35	U2 auxiliary factor 35 kDa
U2AF65	U2 auxiliary factor 65 kDa
UHM	U2AF homology motif
ULM	UHM ligand motif
USA	United States of America
v/v	volume per volume
w/v	weight per volume
μg	micro grams

6.2 Compilation of peptides identified in components of particle X by mass spectrometry

The following tables (6.1-6.4) list all peptides, which were found with a peptide score higher then 20. Peptides found in band 1 matched to DDX1 (Table 6.1), peptides found in band 2 matched to HSPC117 (Table 6.2), peptides found in bands 3 and 4 matched to Family with sequence similarity 98 / member B (Table 6.3) and peptides found in band 6 matched to CGI-99 (Table 6.4). Peptides are numbered (#) and the observed as well as calculated masses (Mr calc) are indicated. Further the mass difference (Delta) and the peptide score (Score) is given. (M) = Oxidation; (C) = Carbamidomethyl.

Table 6.1: Peptides matching to DDX1.

#	observed	Mr (expt)	Mr (calc)	Delta	Score	Expect	Peptide
1	415.2258	828.4370	828.4817	-0.0447	42	0.029	K.LNLSQVR.F
2	418.7145	835.4144	835.4439	-0.0295	39	0.11	K.DGFVALSK.A
3	442.7447	883.4748	883.5127	-0.0378	47	0.0062	K.ALIVEPSR.E
4	474.2398	946.4650	946.4971	-0.0321	64	0.00019	R.LDDLVSTGK.L
5	487.7261	973.4376	973.4657	-0.0280	69	5.8e-005	K.FGFGFGGTGK.K
6	503.2166	1004.4186	1004.4814	-0.0628	56	0.001	K.VPVDEFDGK.V
7	509.7734	1017.5322	1017.5607	-0.0285	45	0.015	K.YIDNPKLR.E
8	524.7964	1047.5782	1047.6076	-0.0294	45	0.012	K.ILKGEYAVR.A
9	556.3319	1110.6492	1110.6760	-0.0268	55	0.0012	R.ELLIIGGVAAR.D
10	573.7464	1145.4782	1145.5029	-0.0246	46	0.01	K.FNFGEEEFK.F
11	616.8351	1231.6556	1231.6846	-0.0289	67	7.1e-005	R.MGLAISLVATEK.E
12	411.5652	1231.6738	1231.7136	-0.0398	54	0.0016	K.TTIKTGASVLNK.W
13	636.8094	1271.6042	1271.6721	-0.0678	69	4.2e-005	R.ELAEQTLNNIK.Q
14	727.8456	1453.6766	1453.6983	-0.0217	59	0.00041	R.MHNQIPQVTSDGK.R
15	734.3736	1466.7326	1466.7591	-0.0265	63	0.00019	K.DLGLAFEIPPHMK.N
16	497.2654	1488.7744	1488.8221	-0.0477	59	0.0004	R.MGLAISLVATEKEK.V
17	537.5892	1609.7458	1609.7994	-0.0537	69	4e-005	R.MHNQIPQVTSDGKR.L
18	539.2488	1614.7246	1614.7718	-0.0472	36	0.084	K.FNFGEEEFKFPPK.D
19	838.4392	1674.8638	1674.8940	-0.0302	61	0.00024	R.ELAEQTLNNIKQFK.K
20	560.6224	1678.8454	1678.8929	-0.0476	28	0.5	K.DGFVALSKAPDGYIVK.S
21	561.2712	1680.7918	1680.8358	-0.0441	26	0.75	K.VPVDEFDGKVTYGQK.R
22	854.8986	1707.7826	1707.8138	-0.0311	86	9.3e-007	R.VGWSTMQASLDLGTDK.F
23	879.4763	1756.9380	1756.9683	-0.0302	65	9.6e-005	R.LDDLVSTGKLNLSQVR.F
24	627.2884	1878.8434	1878.9046	-0.0612	86	8e-007	K.TGASVLNKWQMNPYDR.G
25	635.3360	1902.9862	1903.0414	-0.0552	53	0.0013	K.GHVDILAPTVQELAALEK.E
26	984.5074	1967.0002	1967.0323	-0.0320	56	0.00067	R.DQLSVLENGVDIVVGTPGR.L
27	668.2885	2001.8437	2001.9214	-0.0777	44	0.0099	K.DNTRPGANSPEMWSEAIK.I
28	506.9974	2023.9605	2024.0326	-0.0721	33	0.15	K.GEDSVPDTVHHVVVPVNPK.T
29	693.6896	2078.0470	2078.1047	-0.0578	49	0.0032	R.GIDIHGVPYVINVTLPDEK.Q
30	743.6999	2228.0779	2228.1185	-0.0406	39	0.029	K.APDGYIVKSQHSGNAQVTQTK.F

Table 6.2: Peptides matching to HSPC117.

#	observed	Mr (expt)	Mr (calc)	Delta	Score	Expect	Peptide
1	415.2221	828.4296	828.4527	-0.0231	38	0.07	K.GVIPMNAK.D
2	469.2304	936.4462	936.4738	-0.0276	37	0.074	K.LMFEELR.N
3	471.7251	941.4356	941.4640	-0.0284	72	2.4e-005	R.SSMTFLTR.Q
4	480.2556	958.4966	958.5269	-0.0303	28	0.72	K.LADMGIAIR.V
5	486.7648	971.5150	971.5400	-0.0249	42	0.03	R.DKIIVNDR.Q
6	545.7743	1089.5340	1089.5640	-0.0300	44	0.018	R.GGGVGGFLPAMK.Q
7	603.7830	1205.5514	1205.5928	-0.0413	33	0.21	R.NLDFQDVLDK.L
8	610.8001	1219.5856	1219.6084	-0.0228	52	0.0027	R.IASPEGQDYLK.G
9	648.2999	1294.5852	1294.6115	-0.0262	58	0.0006	K.LVMEEAPESYK.N
10	432.5498	1294.6276	1294.6629	-0.0353	49	0.0048	K.VEQHVVDGKER.T
11	656.2997	1310.5848	1310.6064	-0.0215	56	0.0009	K.LVMEEAPESYK.N + (M)
12	443.8882	1328.6428	1328.6870	-0.0443	45	0.012	R.MLQADPNKVSAR.A
13	672.8241	1343.6336	1343.6336	-0.0232	58	0.00055	R.TNLDESDVQPVK.E
14	449.2220	1344.6442	1344.6820	-0.0378	58	0.00061	R.MLQADPNKVSAR.A + (M)
15	690.8160	1379.6174	1379.6404	-0.0229	80	3.9e-006	K.GMAAAGNYAWVNR.S
16	693.3157	1384.6168	1384.6510	-0.0342	44	0.013	R.SYNDELQFLEK.I
17	698.8016	1395.5886	1395.6353	-0.0467	59	0.00044	K.GMAAAGNYAWVNR.S + (M)
18	482.2691	1443.7855	1443.8310	-0.0455	34	0.14	K.QIGNVAALPGIVHR.S
19	547.2735	1638.7987	1638.7987	-0.0412	41	0.023	R.GLGHQVATDALVAMEK.A

Table 6.3: Peptides matching to FSS98 / Member B.

#	observed	Mr (expt)	Mr (calc)	Delta	Score	Expect	Peptide
1	487.2731	972.5316	972.5604	-0.0287	51	0.0034	R.EDLSKIIR.T
2	599.8192	1197.6238	1197.6605	-0.0366	63	0.00018	K.GPLLEEQALTK.A
3	649.8485	1297.6824	1297.7176	-0.0352	81	2.9e-006	K.TTITMAHLLAAR.E
4	755.3538	1508.6930	1508.7259	-0.0329	68	5.4e-005	R.LDVTVQSFGWSDR.A
5	765.3043	1528.5940	1528.6286	-0.0346	77	6.1e-006	R.INDALSCEYECR.R + 2 (C)
6	768.8518	1535.6890	1535.7249	-0.0359	78	5.8e-006	K.SLCNLEESITSAGR.D + (C)
7	789.8394	1577.6642	1577.6991	-0.0349	84	1.3e-006	K.MDLNSEQAEQLER.I + (M)
8	555.9337	1664.7793	1664.8270	-0.0478	44	0.013	K.RLDVTVQSFGWSDR.A
9	841.8949	1681.7752	1681.8055	-0.0302	37	0.053	K.EMACPYSVLISGDIK.D + (C)
10	570.2728	1707.7966	1707.8580	-0.0614	50	0.0035	R.LDVTVQSFGWSDRAK.V
11	893.9177	1785.8208	1785.8567	-0.0358	42	0.018	K.STTSDIPHMLNQVESK.V
12	901.9142	1801.8138	1801.8516	-0.0378	25	0.83	K.STTSDIPHMLNQVESK.V + (M)
13	624.3147	1869.9223	1869.9982	-0.0759	29	0.38	K.TTITMAHLLAAREDLSK.I

Table 6.4: Peptides matching to CGI-99.

#	observed	Mr (expt)	Mr (calc)	Delta	Score	Expect	Peptide
1	440.1976	878.3806	878.3956	-0.0150	29	0.57	R.DVNCPFK.I + (C)
2	444.2095	886.4044	886.4396	-0.0351	37	0.096	K.DLVPDNSK.T
3	471.2619	940.5092	940.5229	-0.0137	38	0.057	K.EGLPVALDK.H
4	511.7866	1021.5586	1021.5920	-0.0333	38	0.058	R.LLHIEELR.E
5	519.7388	1037.4630	1037.4665	-0.0035	50	0.0039	R.LEYGDNAEK.Y
6	542.2534	1082.4922	1082.5145	-0.0222	34	0.15	R.NIHSSDWPK.F
7	567.2746	1132.5346	1132.5587	-0.0240	42	0.028	R.HDDYLVMLK.A
8	589.7932	1177.5718	1177.5979	-0.0260	40	0.036	K.YKDLVPDNSK.T
9	646.3252	1290.6358	1290.6608	-0.0249	39	0.046	R.NFIVWLEDQK.I
10	665.3061	1328.5976	1328.6248	-0.0271	32	0.22	R.LEYGDNAEKYK.D
11	463.2402	1386.6988	1386.7467	-0.0479	48	0.0056	R.LTQDAVAKANQTK.E
12	495.2606	1482.7600	1482.8042	-0.0442	27	0.65	K.ANQTKEGLPVALDK.H
13	749.4080	1496.8014	1496.8497	-0.0482	90	3.3e-007	K.AGVMALANLLQIQR.H
14	530.2497	1587.7273	1587.7740	-0.0467	48	0.0054	K.DLVPDNSKTADNATK.N
15	832.9724	1663.9302	1663.9508	-0.0206	63	0.00018	K.INEAIVAVQAIIADPK.T
16	906.9427	1811.8708	1811.9053	-0.0345	52	0.003	K.NAEPLINLDVNNPDFK.A

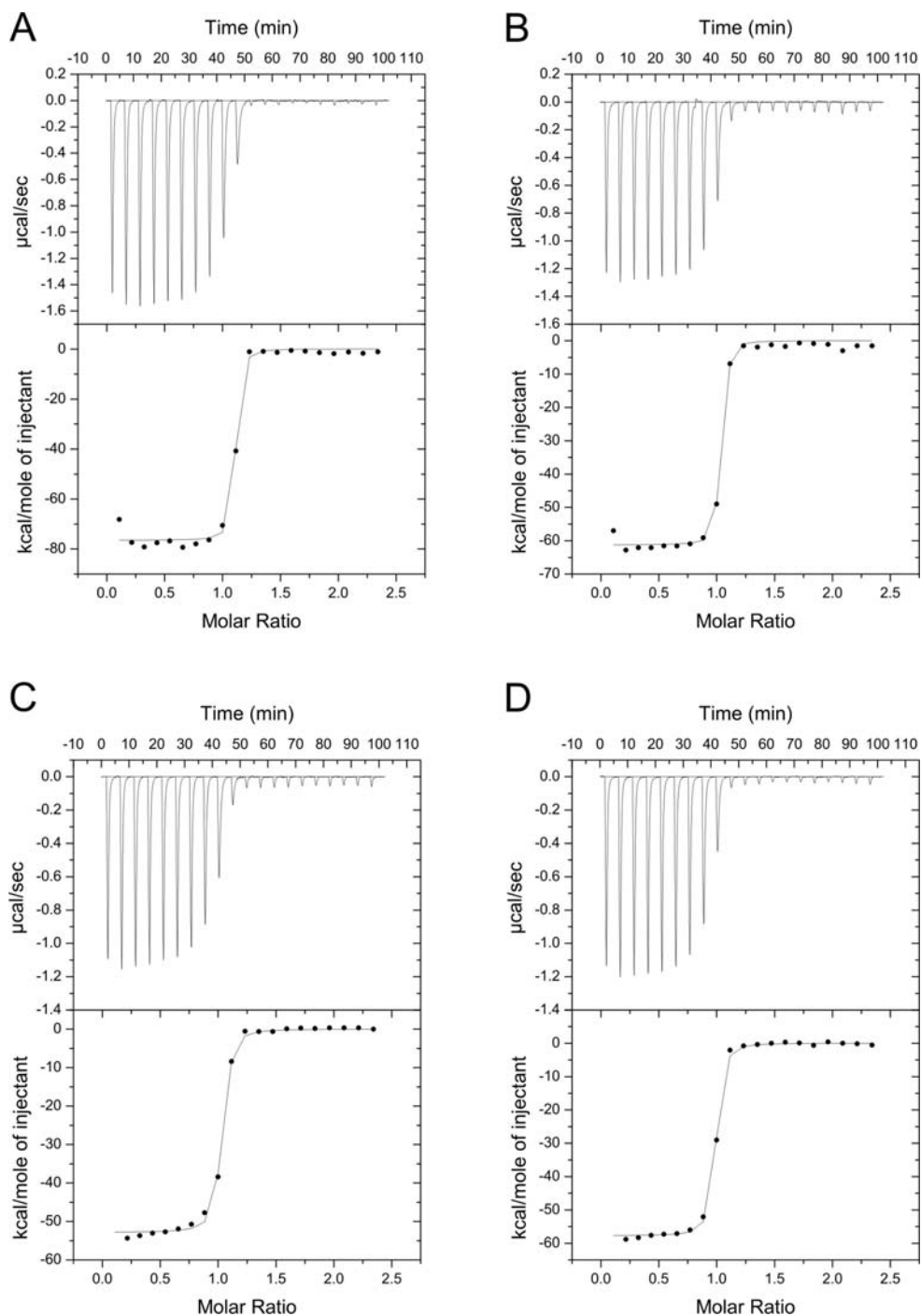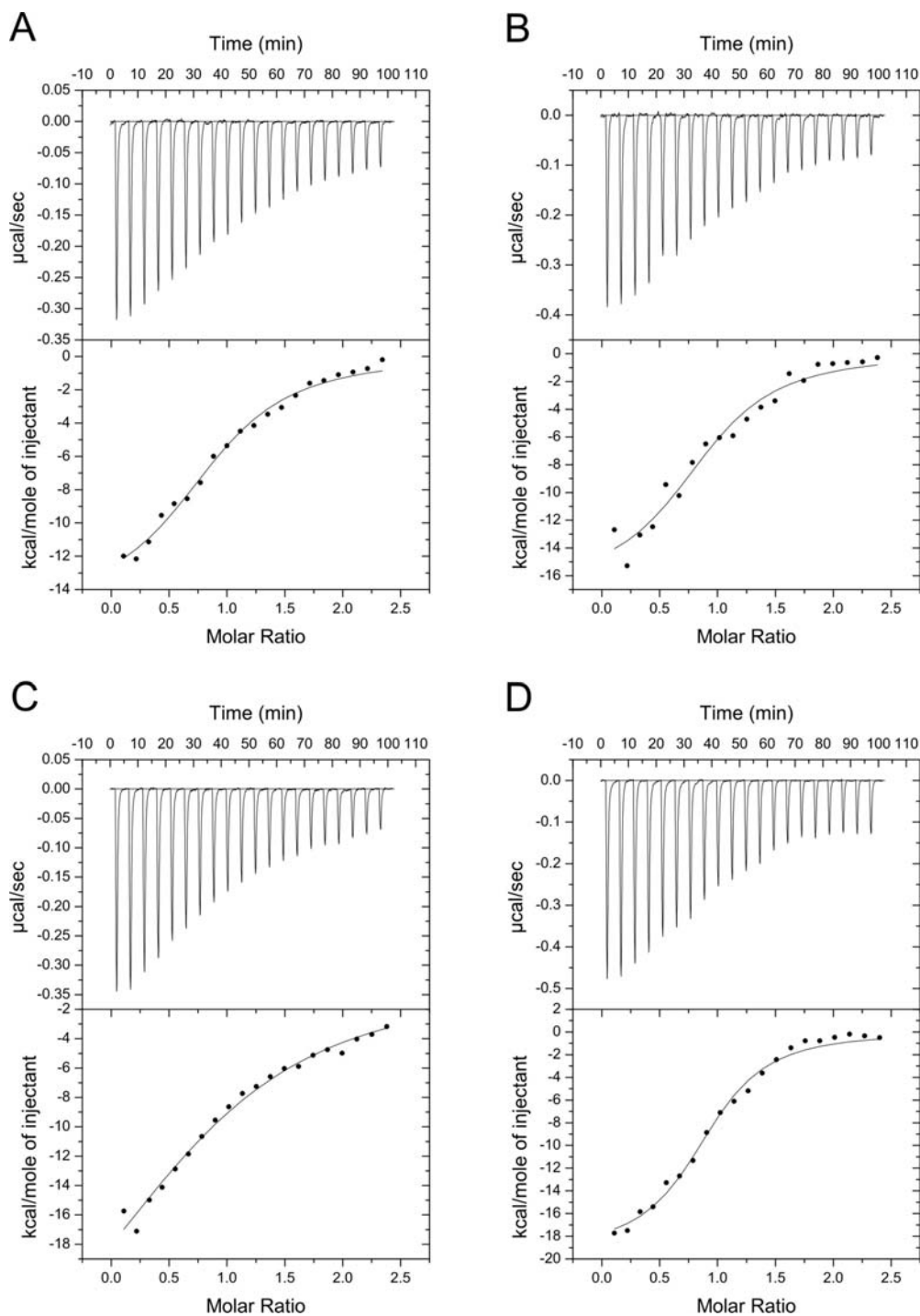## 6.3 Isothermal titration calorimetry data

Fig. 6.1: Isothermal titration calorimetric data of the Bud13p/Snu17p interaction. All Snu17p constructs were used as analytes at 6 µM and all Bud13p constructs were used as titrants at 60 µM concentration. **(A)** Snu17p$^{1-148}$-Bud13p$^{1-266}$ interaction; **(B)** Snu17p$^{1-148}$-Bud13p$^{202-266}$ interaction; **(C)** Snu17p$^{25-106}$-Bud13p$^{1-266}$ interaction; **(D)** Snu17p$^{25-106}$-Bud13p$^{202-266}$ interaction.

Fig. 6.2: Isothermal titration calorimetric data of the Pml1p/Snu17p interaction. All Snu17p constructs were used as analytes at 6 µM and all Pml1p constructs were used as titrants at 60 µM concentration. **(A)** Snu17p$^{1-148}$-Pml1p$^{1-204}$ interaction; **(B)** Snu17p$^{1-148}$-Pml1p$^{1-61}$ interaction; **(C)** Snu17p$^{25-138}$-Pml1p$^{1-204}$ interaction; **(D)** Snu17p$^{25-138}$-Pml1p$^{1-61}$ interaction.
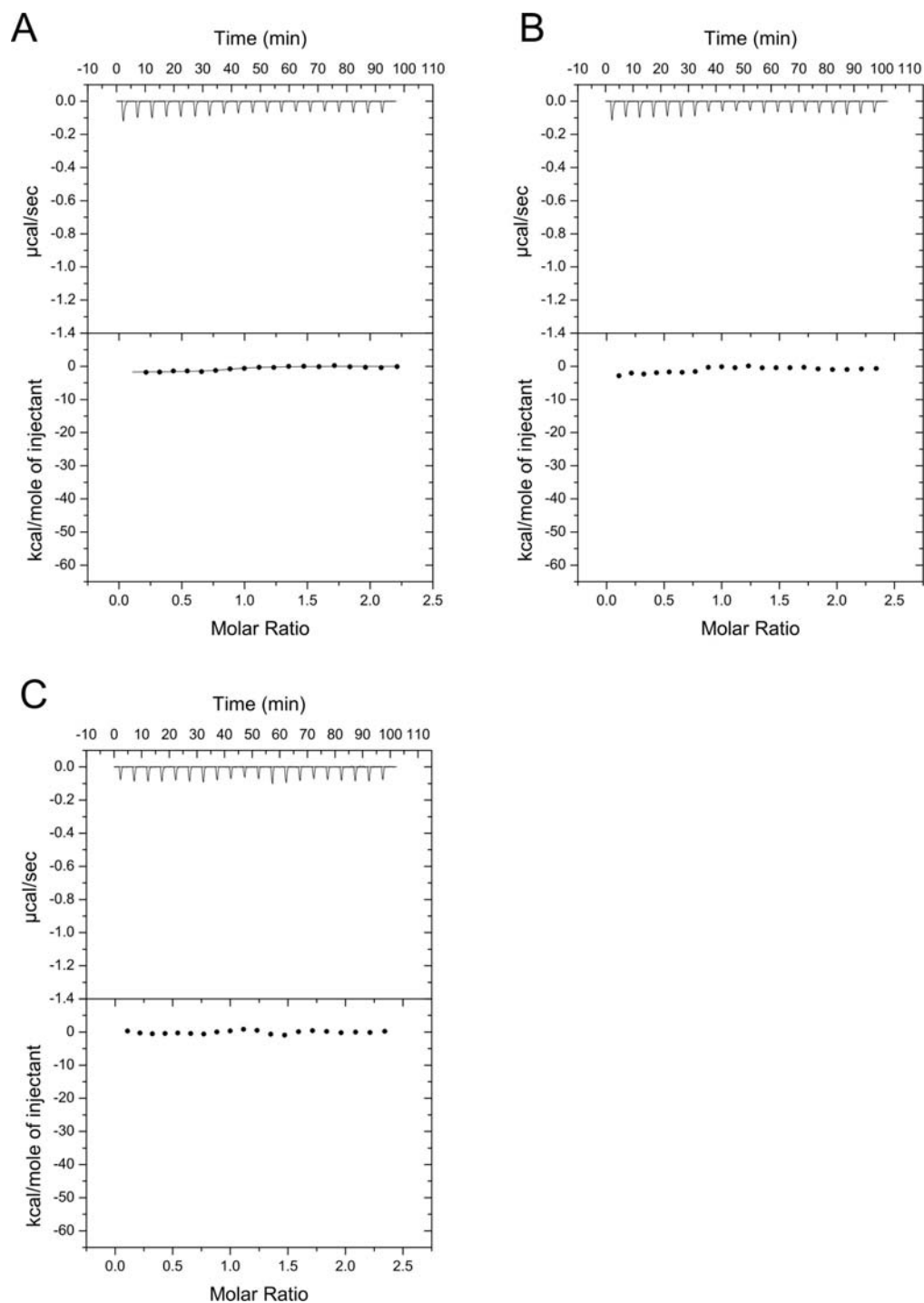
Fig. 6.3: Isothermal titration calorimetric data of control experiments. (A) Snu17p$^{25-106}$-Bud3p$^{202-266}$[Trp232Ala] interaction (6 and 60 μM, respectively); **(B)** Pml1p$^{1-204}$-Bud3p$^{1-266}$ interaction (6 and 60 μM, respectively); **(C)** Bud3p$^{1-266}$ into buffer (30μM).

6.4   B-factor distribution in Pml1p models
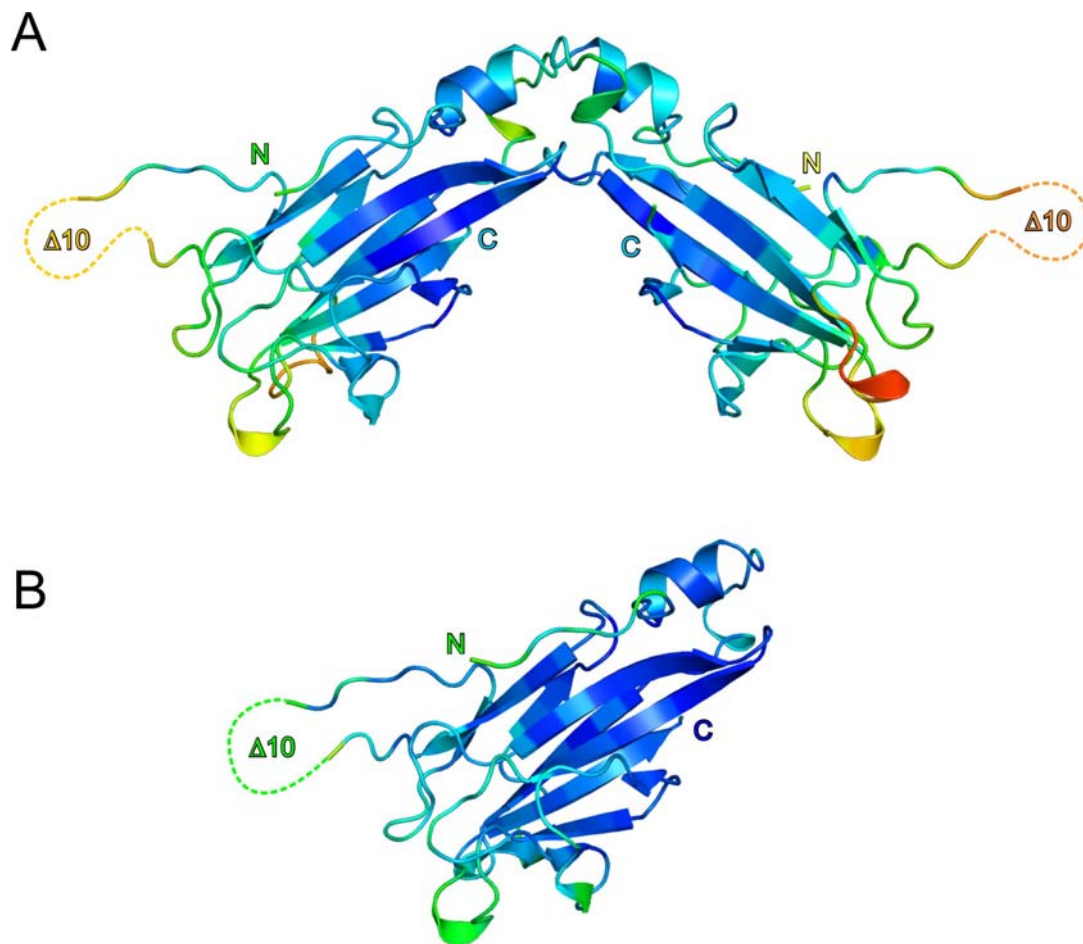
Fig. 6.4: Crystallographic *B*-factors. Models covering the asymmetric units of Pml1p[FL] (**A**) and Pml1p[51-204] crystals (**B**) are shown in cartoon representation and colored according to the C_α atom *B* factor value. In the color gradient, lowest *B* factors are blue, highest red. Solvent exposed loops display higher *B* factors than the protein core. Packing interactions in Pml1p[51-204] crystals seem to rigidify the molecules.

6.5   Ramachandran plots of Pml1p models
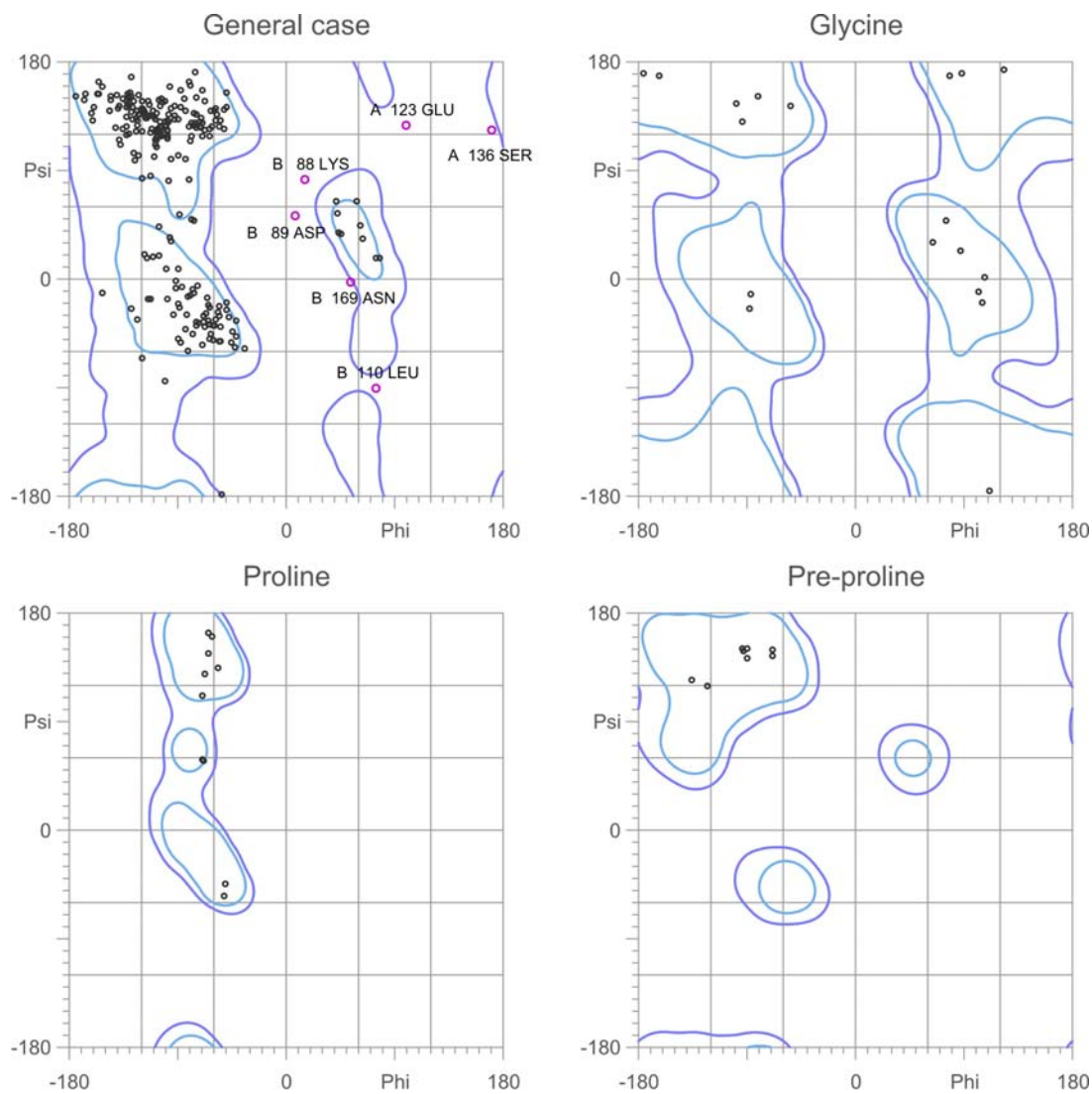
Fig. 6.5: Ramachandran plots of the full-length Pml1p models calculated by the MolProbity server. 93.2% (261/280) of all residues were in favored (98%) regions; 97% (274/280) of all residues were in allowed (>99.8) regions. There were six outliers: A123 Glu, A136 Ser, B88 Lys, B89 Asp, B110 Leu and B169 Asn. (http://kinemage.biochem.duke.edu).
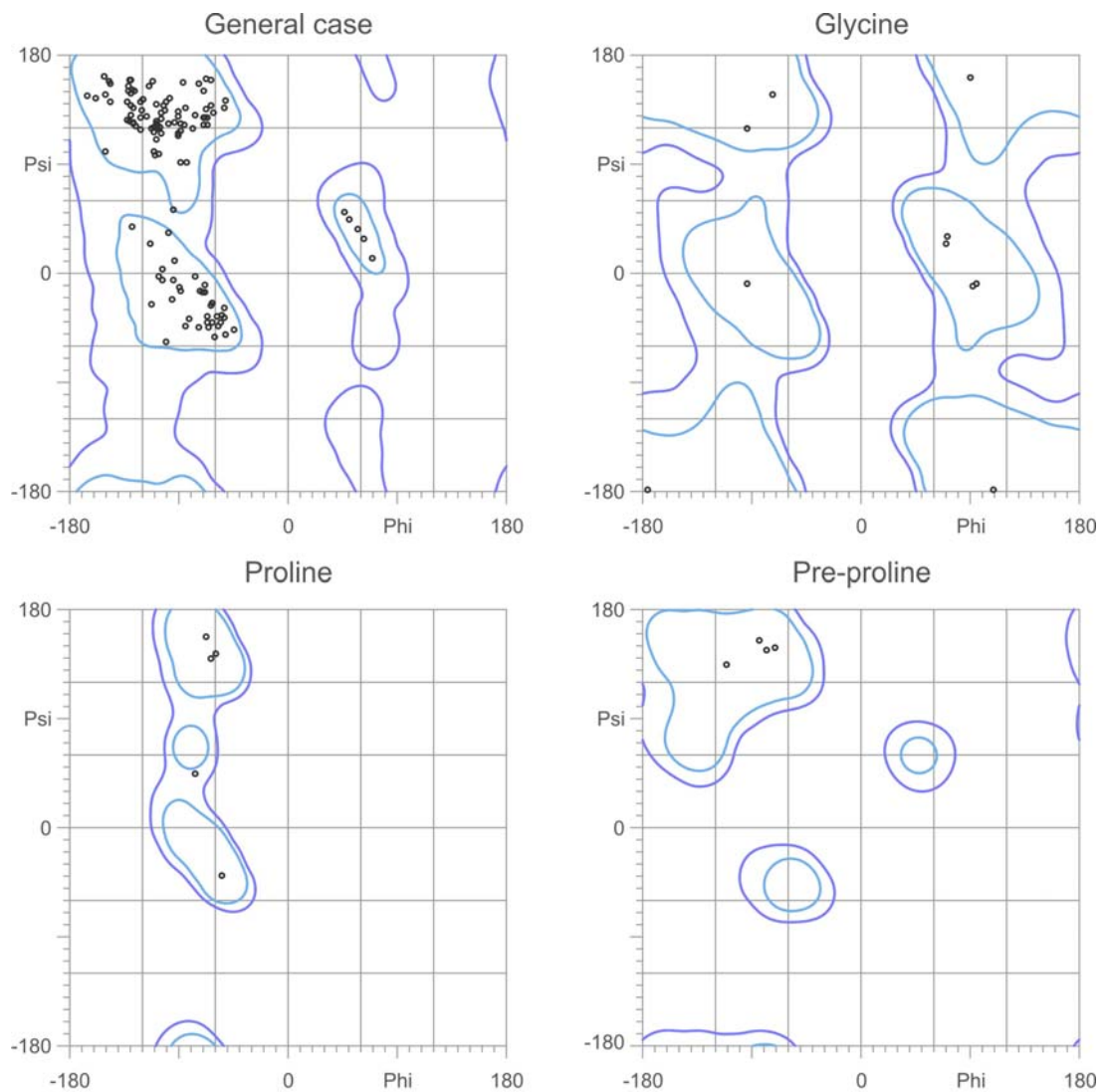
Fig. 6.6: Ramachandran plots calculated by the MolProbity server. 98.6% (143/145) of all residues were in favored (98%) regions; 100% (145/145) of all residues were in allowed (>99.8) regions (http://kinemage.biochem.duke.edu).

6.6 Topology plots of Pml1p models
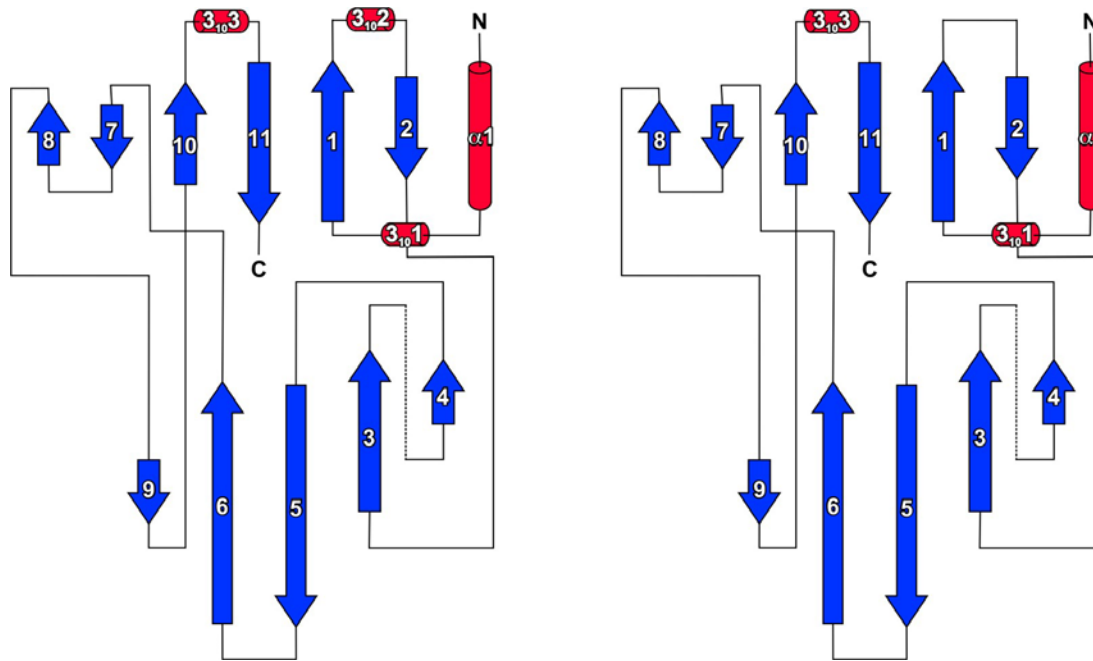
Fig. 6.7: Topology plots of Pml1p$^{51-204}$ and Pml1p^{FL}. Topology plots of Pml1p$^{51-204}$ (left) and Pml1p^{FL} (right) are shown with β-strands as blue arrows and helices as red columns. 3$_{10}$-helices are indicated. Consecutive numbering of secondary structure elements is from the N-terminus (N) to the C-terminus (C). Dashed lines indicate unstructured regions not present in the models.

6.7 Acknowledgements

Parts of this work were published in the following papers:

2008   Trowitzsch, S., Weber, G., Lührmann, R., and Wahl, M.C. (2009) Crystal
            structure of the Pml1p subunit of the yeast pre-mRNA retention and splicing
            complex, *JMB*, **385**, 531-541

2008   Trowitzsch, S., Weber, G., Lührmann, R., and Wahl, M.C. (2008) An
            unusual RNA recognition motif acts as a scaffold for multiple proteins in
            the pre-mRNA retention and splicing complex, *JBC*, **283**, 32317-32327

6.8   Curriculum vitae

Personal data

Name	Simon Trowitzsch
Date of birth	August 16th 1977
Place of birth	Northeim, Germany

Education

08/1990 - 06/1997	Gymnasium Corvinianum, Northeim
06/1997	Abitur
09/1997 - 09/1998	Social service (German Red Cross)
10/1998 - 03/2004	Diploma Study: Biology at the Georg-August-University of Göttingen, Germany
10/2000	Degree: Pre-Diploma (sehr gut)
03/2004	Degree: Diploma (sehr gut)
since 04/2004	Doctoral Studies at the Max-Planck-Institute for Biophysical Chemistry, Department of Cellular Biochemistry / X-ray Crystallography, Göttingen, Germany

Practical Education

03/2003 - 02/2004	Diploma work at the Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany
since 04/2004	Doctoral Studies at the Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany
01/2005 - 03/2005	Trainee-Supervision within the International Master/PhD program "Molecular Biology", Georg-August-University of Göttingen, Germany
01/2006 - 03/2006	Trainee-Supervision within the International Master/PhD program "Molecular Biology", Georg-August-University of Göttingen, Germany
01/2006 - 01/2006	Tutor of the International Master/PhD program "Molecular Biology", Georg-August-University of Göttingen, Germany
01/2007 - 01/2007	Tutor of the International Master/PhD program "Molecular Biology", Georg-August-University of Göttingen, Germany

Publications

2005   Andresen, M., Wahl, M.C., Stiel, A.C., Gräter, F., Schäfer, L.V., Trowitzsch, S., Weber, G., Eggeling, C., Grubmüller, H., Hell, S.W., Jakobs, S. (2005) Structure and mechanism of the reversible photoswitch of a fluorescent protein, *Proc. Natl. Acad. Sci. U S A*, **13**, 102

2006   Savarese, E., Chae, O.W., Trowitzsch, S., Weber, G., Kastner, B., Akira, S., Wagner, H., Schmid, R.M., Bauer, S., Krug, A. (2006) U1 small nuclear ribonucleoprotein immune complexes induce type I interferon in plasmacytoid dentritic cells through TLR7, *Blood*, **15**, 107

2007   Stiel, A.C.[*], Trowitzsch, S.[*], Weber, G.[*], Andresen, M., Eggeling, C., Hell, S.W., Jakobs, S., Wahl, M.C. (2007) 1.8 A bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants, *Biochem. J.*, **15**, 402
          [*] These authors contribute equally to this work.

2007   Andresen, M., Stiel, A.C., Trowitzsch, S., Weber, G., Eggeling, C., Wahl, M.C., Hell, S.W., Jakobs, S. (2007) Structural basis for reversible photoswitching in Dronpa, *Proc. Natl. Acad. Sci. U S A*, **7**, 104

2008   Trowitzsch, S., Weber, G., Lührmann, R., and Wahl, M.C. (2009) Crystal structure of the Pml1p subunit of the yeast pre-mRNA retention and splicing complex, *JMB*, **385**, 531-541

2008   Trowitzsch, S., Weber, G., Lührmann, R., and Wahl, M.C. (2008) An unusual RNA recognition motif acts as a scaffold for multiple proteins in the pre-mRNA retention and splicing complex, *JBC*, **283**, 32317-32327

Contributions to conferences

05/2007   "12th Annual Meeting of the RNA society", Madison, WI, USA

                 Oral presentation: "Architecture of the pre-mRNA Retention and Splicing Complex from Yeast"

09/2007   "Horizons in Molecular Biology", Göttingen, Germany

                 Poster presentation: "Architecture of the Yeast pre-mRNA Retention and Splicing Complex"