Purification and crystallization of spliceosomal snRNPs

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August-Universität Göttingen

vorgelegt von

Gert Weber

aus Homberg/Efze

Göttingen, den 27.5.2008

Contents

List of Figures

List of Tables

VIII

Abstract

In eukaryotes, the non-coding sequences (introns) of precursor messenger RNA are excised by a large ribonucleoprotein complex, the spliceosome. The core components of the spliceosome are small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4, U5 and U6. Besides the U1 snRNA, U1 snRNP comprises three particle-specific proteins, U1-70k, U1-A and U1-C, and a set of seven proteins, also present in other snRNPs, termed Sm proteins. The Sm proteins form a ring-like structure, the Sm-core. U2 snRNP is associated with two multi-protein factors, SF3a and SF3b, forming a 17S particle. SF3a comprises three protein factors, SF3a 60, 66 and 120. U4, U5 and U6 snRNPs constitute the 25S U4/U6·U5 tri-snRNP, with 36 proteins and three RNAs. Electron microscopic reconstructions of HeLa U1 snRNP, SF3a and tri-snRNP are available and several substructures solved by NMR and X-ray crystallography are known. Atomic structures of SF3a, tri-snRNP and U1 snRNP would yield fundamental insight into splicing. The objective of this work was to obtain U1 snRNP, tri-snRNP and SF3a preparations suitable for X-ray crystallography.

In the work presented here, SF3a was purified from HeLa cells, concentrated and subjected to crystallization. HeLa tri-snRNP was sensitive to ultrafiltration or pelleting. Concentration was achieved by ammonium sulfate precipitation followed by dialysis. Highly concentrated and stable tri-snRNP fractions were obtained and subjected to crystallization. Previously, HeLa U1 snRNP crystals were obtained, but not reproducible, and diffracted X-rays to 20 Å. The deliberate inclusion of proteases to the crystallization setup (*in situ* proteolysis) dramatically increased size and reproducibility of U1 snRNP crystals. An analysis of the contents of the crystals demonstrated that a truncation of several U1 snRNP proteins was a prerequisite for crystal formation. The purification protocol of U1 snRNP was improved by entirely separating U1 and U2 snRNPs. This allowed a combination of a high throughput screening approach with *in situ* proteolysis in the presence of various U1 snRNP ligands. The combined strategies led to the reproducible production of U1 snRNP crystals, diffracting to 4 Å resolution. Entire datasets to a resolution of 4.5 Å were collected of U1 snRNP crystals, complexed with a short DNA-oligonucleotide, mimicking a 5' splice site. Initial phase information for

these crystals was obtained. Three heavy atom derivatives were characterized at low resolution and a potential molecular replacement solution, encompassing U1-A and a model of the Sm-core, was found. The different phasing strategies were confirmed by cross-validation. These results represent a fundamental step towards the structure at atomic resolution of a major part of U1 snRNP.

1 Introduction

1.1 The molecular basis of pre-mRNA splicing

1.1.1 Pre-mRNA splicing in the context of gene expression

According to the central dogma of molecular biology, which was elaborated in the 1960s and 1970s, genetic information is stored in DNA (desoxyribonucleic acid) and copied to messenger RNA (mRNA) in a process termed transcription. Finally, the mRNA is translated to proteins with the help of adapter molecules, transfer RNAs (tRNAs). Translation is carried out by the ribosome, a ribonucleoprotein complex, composed of mainly ribosomal RNA (rRNA) and a minor part of proteins. Since the prokaryotic cell lacks a nucleus, translation occurs simultaneously with transcription and in a collinear fashion, which means that the informational content of a transcriptional unit matches that of the corresponding proteins.

In eukaryotes, transcription occurs in the nucleus and is separated from the translational machinery, therefore mRNA has to be exported through the nuclear membrane to the cytosol. In 1977, two different approaches revealed the existence of 'split genes' in eukaryotic genomes. It means that coding DNA is interspersed with sequences that are not translated to proteins but are excised at the level of pre-mRNA (Berget *et al*., 1977; Chow *et al*., 1977). These non-coding sequences were termed introns (intervening sequences) whereas the coding parts of the DNA and RNA were named exons (expressed sequences). This finding, termed splicing, fundamentally changed the understanding of gene organization in eukaryotes and was awarded with the Nobel prize in 1993 (Berget *et al*., 1977; Chow *et al*., 1977). Later, it was found that exons can be skipped or combined in a non-linear fashion in a process termed alternative splicing (Cáceres and Kornblihtt, 2002). So the principle of collinear gene expression holds not true for eukaryotes. Pre-mRNA splicing allows for a multitude of combinations for joining different exons and increases the number of possible proteins in a cell. This mechanism might have evolved since higher organisms have to administrate a huge genome size

($> 2 \times 10^7$ nucleotides) in various cell types or tissues and, therefore, require a complex proteome.

The spatial and temporal separation of transcription and translation coincides with a far more complex protein biosynthesis in eukaryotes. In contrast to prokaryotes there are three RNA polymerases that fulfill defined tasks in the nucleus. RNA polymerase I exerts the transcription of rRNAs and RNA polymerase III is transcribing tRNA, 5S rRNA and some snRNA (small nuclear RNA) genes. The products of the latter polymerases are not translated whereas RNA polymerase II is synthesizing mRNA destined for translation. In addition, RNA polymerase II transcribes untranslated snoRNA (small nucleolar RNA) and snRNA genes. RNA polymerase II has a multitude of auxiliary factors that interact with the core enzyme to regulate gene expression.
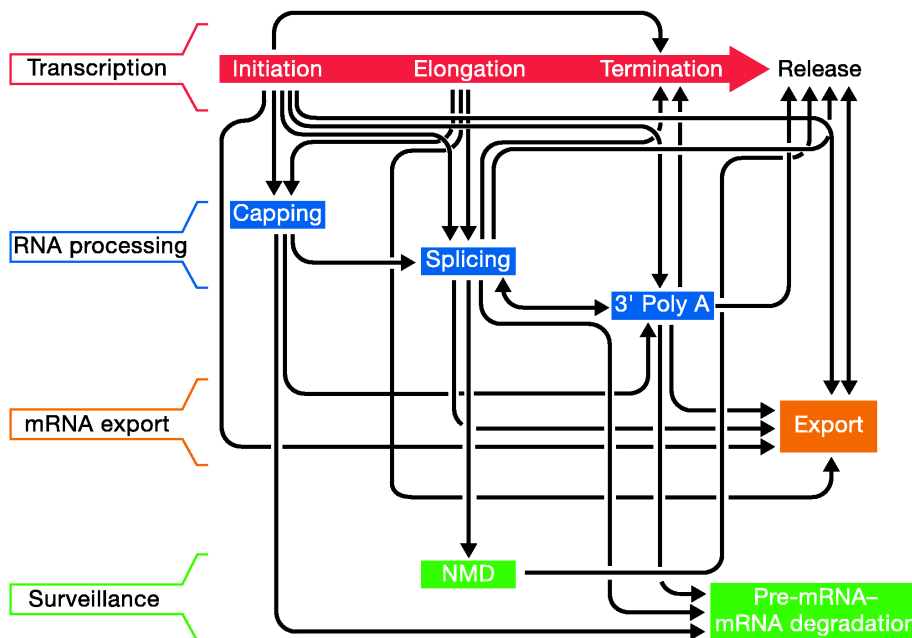
Figure 1.1 A complex network of directly and indirectly coupled interactions in gene expression. Transcription, RNA processing, mRNA export and surveillance are interlinked and regulated processes that direct the maturation of pre-mRNA and its nuclear export. (Maniatis and Reed, 2002)

As shown in figure 1.1, apart from pre-mRNA splicing, further modifications are introduced into the nascent mRNA transcript. Concomitantly with the pre-mRNA exiting RNA polymerase II, the 5' end of the RNA is modified with a m_7G cap (m_7GpppN). The latter process, in turn, promotes the assembly of the splicing machinery on the substrate to excise the intronic sequences. Splicing has multiple variations like exon skipping

or trans-splicing, an exon ligation between different mRNA molecules. Like regular splicing, these mechanisms are modulated by the pre-mRNA sequence and additional factors. In parallel to multiple splicing events a poly(A) tail is sequentially added to the 3' end of the transcript. The enzymes responsible for pre-mRNA maturation are mainly composed of multi-protein factors that may contain small RNAs like the RNA editing machinery or the spliceosome.

As depicted in figure 1.1 each of these cotranscriptional mRNA modifications and their combination increases the stability of the mRNA and facilitates its export to the cytosol. The absence of modifications can trigger degradation of the mRNA. The presence of a protein complex, termed exon-exon-junction complex, on yet unspliced stretches of pre-mRNA can trigger nonsense mediated decay (NMD) (Le Hir *et al.*, 2001; Chang *et al.*, 2007). A fully processed and modified mRNA is therefore termed mature mRNA and exported to the cytosol for translation.

The processes that occur in the nucleus during pre-mRNA maturation are either directly coupled to transcriptional events like capping, splicing and polyadenylation or communicate via RNAs and RNA-associated protein factors that may in addition carry post-translational modifications for signaling.

1.1.2 Pre-mRNA splicing is carried out by the spliceosome in a two-step reaction

After the discovery of nuclear pre-mRNA splicing by Sharp, Roberts and coworkers, other splicing mechanisms have been identified. In lower eukaryotes, so called group I introns exist, where a single guanosine nucleotide is involved in catalysis (Zaug and Cech, 1980). Furthermore group II and III introns were identified which convert the excised intron to a lariat structure (van der Veen *et al.*, 1986; Woolford and Peebles, 1992). Both, group I and II introns do not require any protein factors for catalysis and therefore belong to the class of ribozymes. These are enzymes composed of nucleic acid and they presumably represent relics of an ancient 'RNA world', with proteins acting as cofactors at most (Sharp, 1985; Szathmáry, 1999).

Nuclear introns in metazoa are removed by a complex ribonucleoprotein machinery, termed the spliceosome. In the context of an RNA world it has been discussed whether the spliceosome-dependent intron removal is carried out by RNA or protein. RNA-catalysis may imply an evolutionary relation to the autocatalytic excision of group I introns in *Tetrahymena thermophilia* (Zaug and Cech, 1980). Furthermore group I introns employ two magnesium ions for catalysis, which is hypothesized to be true for

the spliceosome-dependent reaction and group II introns as well (Stahley and Strobel, 2006; Valadkhan, 2007).

With increasing genome size and complexity of cellular tasks the frequency of nuclear intronic sequences increases during evolution. While baker's yeast only has a small percentage of split genes, mostly ribosomal proteins, in *Homo sapiens* nearly all genes are interspersed with one or several intronic sequences. The combinatorial possibilities of interpreting genes with multiple introns as well as different modes of splicing give rise to a great diversity of the resulting proteins. On a given pre-mRNA one or several exons can be left out (alternative splicing) or even exons from different pre-mRNAs can be ligated in a process termed trans-splicing (Cáceres and Kornblihtt, 2002; Garcia Blanco, 2003). The research on alternative splicing opens up the way to potential gene therapies by silencing specific splice sites on the pre-mRNA by antisense nucleotides (Garcia Blanco, 2006).

As shown in figure 1.2, the nuclear splicing reaction is dependent on specific recognition sequences on the pre mRNA substrate, such as the 5' splice site, the branch point adenosine with a proximate polypyrimidine tract and the 3' splice site. The distance of 5' splice site and branch point can be up to 10^5 nucleotides (nt), whereas the branchpoint and polypyrimidine tract are located close to the 3' splice site.

In the first step of the splicing reaction, the OH-group of the branch point adenosine ribose attacks the phosphodiester bond of the 5' splice site. Concomitantly, the guanosine at the 5' splice site is linked to the hydroxyl group of the branch point adenosine as a 2',5' phosphodiester bond. The first reaction yields a 5' exon as well as the 3' exon with the intron. In the second step of the reaction the 3' hydroxyl group at the 5' exon end attacks the phosphodiester bond at the 3' splice site. Finally the spliced pre-mRNA and the intron, whose internal cyclic structure resembles a lariat, are released from the reaction (Will and Lührmann, 1999).

Both reactions follow a S_N2 type reaction mechanism and are dependent on ATP (Ruskin *et al.*, 1984; Moore and Sharp, 1993). The consumption of ATP is presumably related to RNA helicase and protein kinase activities that are required for the reaction (Staley and Guthrie, 1998). As mentioned previously, there is evidence that magnesium ions are involved in the reaction (Valadkhan, 2007).
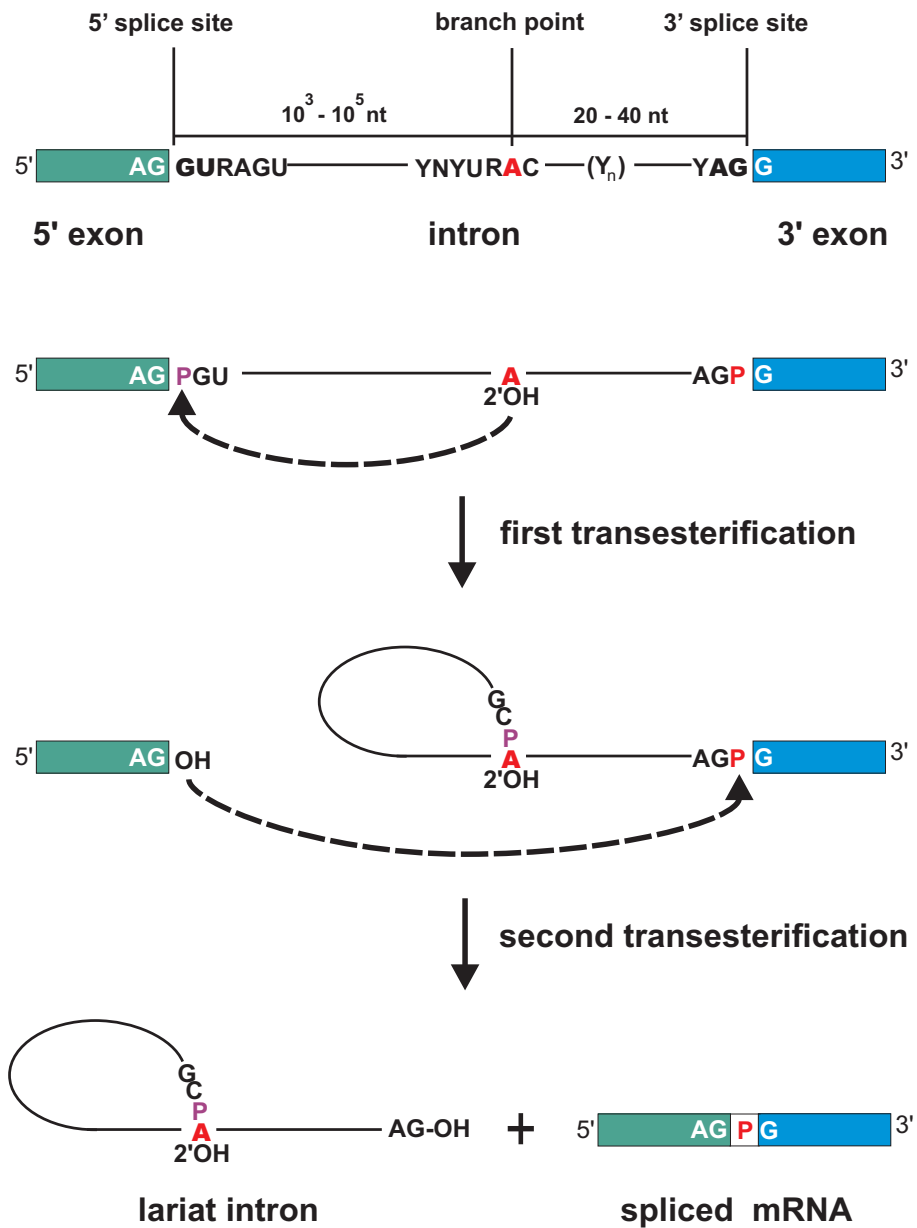
Figure 1.2 The spliceosome ligates exon ends by two consecutive transesterification steps. Dashed arrows indicate a nucleophilic attack. The branch point adenosine (A) and the 3' splice site phosphate (P) are indicated by red letters. The phosphate of the 5' splice site is shown in purple letters. Modified after Will and Lührmann (1999)

1.1.3 The spliceosome is composed of RNAs and proteins, which form ribonucleoprotein particles

The spliceosome contains five uridine-rich snRNAs (U1, U2, U5 and U4/U6) and more than 100 protein factors. Several of the protein factors and the snRNAs form complexes, termed small nuclear ribonucleoprotein complexes (snRNPs) (Lerner and Steitz, 1979). Figure 1.3 shows the secondary structures of the spliceosomal snRNAs and their length in nucleotides. All snRNAs comprise several modifications in their nucleotides, like pseudouridinylation, adenosine methylation or 2'-O-ribose methylation. U6 snRNA carries a γ-monomethylphosphate on its 5' end, since it is a transcript of RNA polymerase III (Krol *et al.*, 1987; Singh and Reddy, 1989). The other spliceosomal snRNAs derive from RNA polymerase II transcription and have a trimethylated guanosine ($m_3^{2,2,7}$GpppN) cap at their 5' end. This hypermethylated cap-structure is originating from a m_7G cap, to which additional methylations are introduced by a methyltransferase (Girard *et al.*, 2008). Constituting a unique feature of snRNAs, the $m_3^{2,2,7}$G cap has been successfully employed for an immunoaffinity-based purification of snRNPs (Bringmann *et al.*, 1983). U4 and U6 snRNAs form a single ribonucleoprotein particle due to a large region of complementary nucleotides (Hashimoto and Steitz, 1984). In eukaryotic nuclei, U4/U6 snRNP is associated with the U5 snRNP, forming the 25S U4/U6·U5 tri-snRNP.

Nuclear pre-mRNA splicing requires multiple protein factors that either maintain the structure or facilitate rearrangements of the spliceosome (Staley and Guthrie, 1998). A major fraction of spliceosomal proteins contains an RNA recognition motif (RRM) which is primarily implicated in single strand RNA binding, but can also contact proteins. An RRM combined with other domains may direct a specific factor or an activity to a specific snRNA or the pre-mRNA sequence. Another well known motif employed by proteins in splicing is the RS domain. It is hallmarked by consecutively alternating arginines and serines, the latter can be reversibly phosphorylated. This domain is often combined with an RRM.

The U4/U6·U5 tri-snRNP comprises a set of ATP-dependent RNA helicases that are responsible for the rearrangement of snRNAs. The Brr2 DEIH box helicase, for example, is unwinding U4 and U6 snRNAs during the first step of splicing. Other helicases like Prp2 or Prp22 are not associated with the snRNPs but also act on the snRNAs or pre-mRNA (Laggerbauer *et al.*, 1998; Schwer and Meszaros, 2000).

Apart from the canonical snRNPs, a subset of spliceosomal particles was named by its recognition sequence on the pre-mRNA, the 'ATAC' spliceosome. This so called minor spliceosome excises 0.2 % of all vertebrate introns (Tarn and Steitz, 1996). In
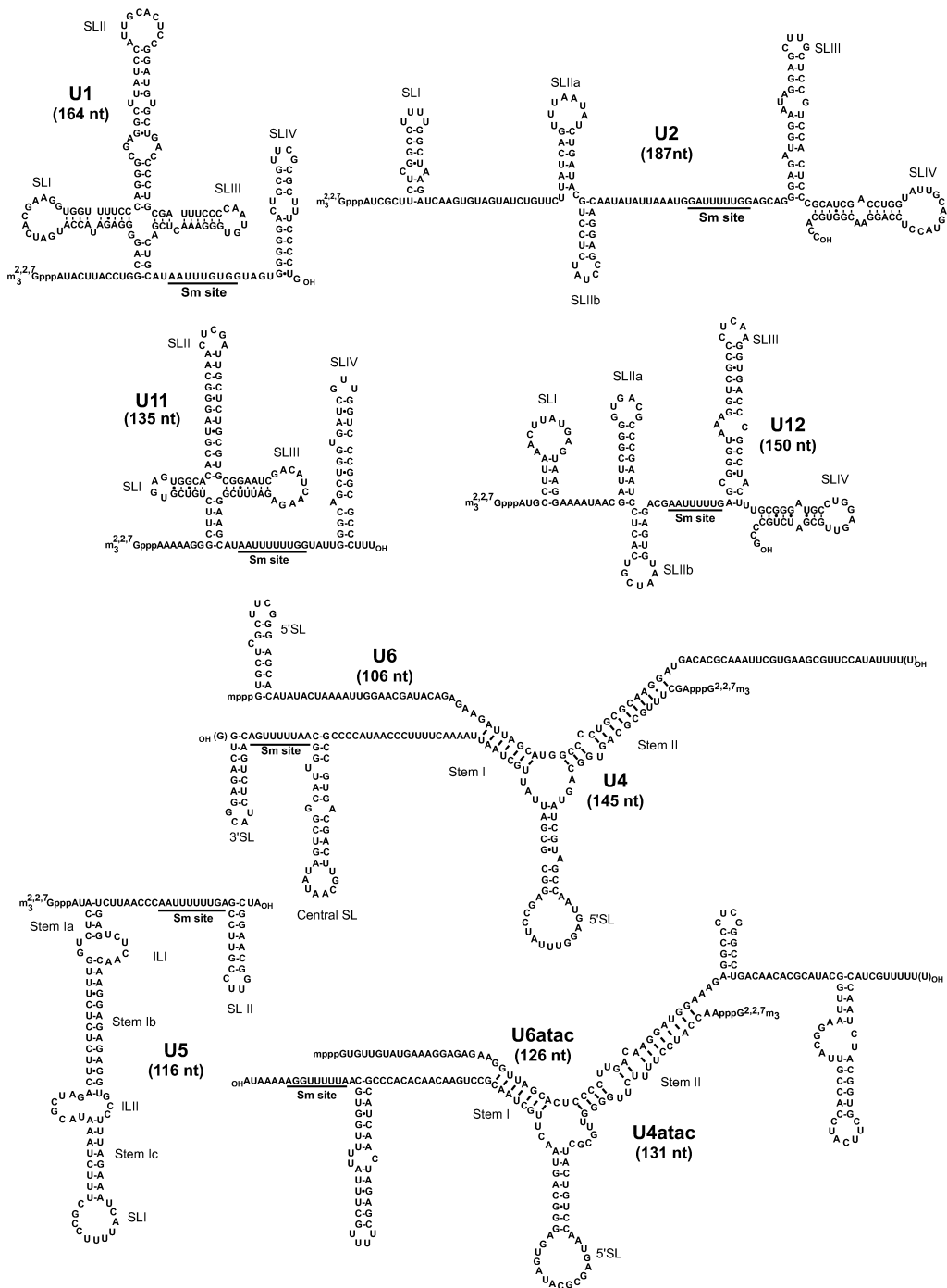
Figure 1.3 Sequence and secondary structure models of the human snRNAs of the major and minor spliceosome. The conserved Sm sites are underlined, stem loop structures (SL) are indicated (Will and Lührmann, 2005; Guthrie and Patterson, 1988; Wassarman and Steitz, 1992; Padgett and Shukla, 2002)

the minor spliceosome, U1, U2 and U4/U6 snRNPs are replaced by the U11, U12 and U4/U6 ATAC particles, respectively. In HeLa cells, U1 and U2 snRNP represent two distinct moieties, whereas U11 and U12 snRNPs are bound to each other by protein factors (Benecke *et al*., 2005).

1.2 SnRNPs as core components of the spliceosome

1.2.1 The Sm core proteins

U1, U2, U5 and U4 snRNPs share a common set of factors which are the Sm proteins B/B', D1, D2, D3, E, F and G. The term Sm proteins refers to the proteins reactivity with serum of patients afflicted with systemic *lupus erythematosus* (SLE), an autoimmune disease (Lerner and Steitz, 1979). These proteins are important for snRNP biogenesis and bind to an uridine-rich region (Sm site) on the snRNAs. Most likely, the seven Sm proteins form a heptameric ring-like structure on the snRNAs (Achsel *et al*., 1998, 1999; Kambach *et al*., 1999b). The snRNAs are presumably threaded through the pore, in the center of the ring. In contrast to other snRNAs, U6 binds an alternative set of Sm proteins, termed Sm-like (LSm). The RNA-binding mode of LSm proteins is presumably similar to Sm proteins. SmB and B' are coded by one transcript, the difference is a result of alternative splicing where the C-terminus of the B' protein comprises an additional repeat of a proline-rich motif (van Dam *et al*., 1989). It has been shown that the arginine glycine (RG) motifs in D1, D3 and B/B' contain symmetrical dimethylarginines. These methylations are required for snRNP biogenesis (Brahms *et al*., 2001). The binding of Sm proteins to the RNA is conferred by a bipartite Sm motif with a linker of variable size (Achsel *et al*., 1999). Accounting for the presence of a distinct motif, Sm proteins constitute a protein family of their own. In the absence of snRNA, complexes of D3 and B/B', D1 and D2 and furthermore E, F and G have been described (Hermann *et al*., 1995; Raker *et al*., 1996; Lehmeier *et al*., 1994).

Figure 1.4 represents a hypothetical model of the Sm core doughnut-shaped heptamer. B/B' and D3 as well as D1 and D2 have been shown to interact with each other by co-expression studies and crystal structures (Kambach *et al*., 1999a; Raker *et al*., 1999). The relative positions of E, F and G can be concluded from immunoprecipitation assays and yeast two-hybrid studies (Camasses *et al*., 1998; Fury and Andersen, 1997; Fury *et al*., 1999). The fold of the Sm motifs is an anti-parallel five stranded β-sheet, furthermore
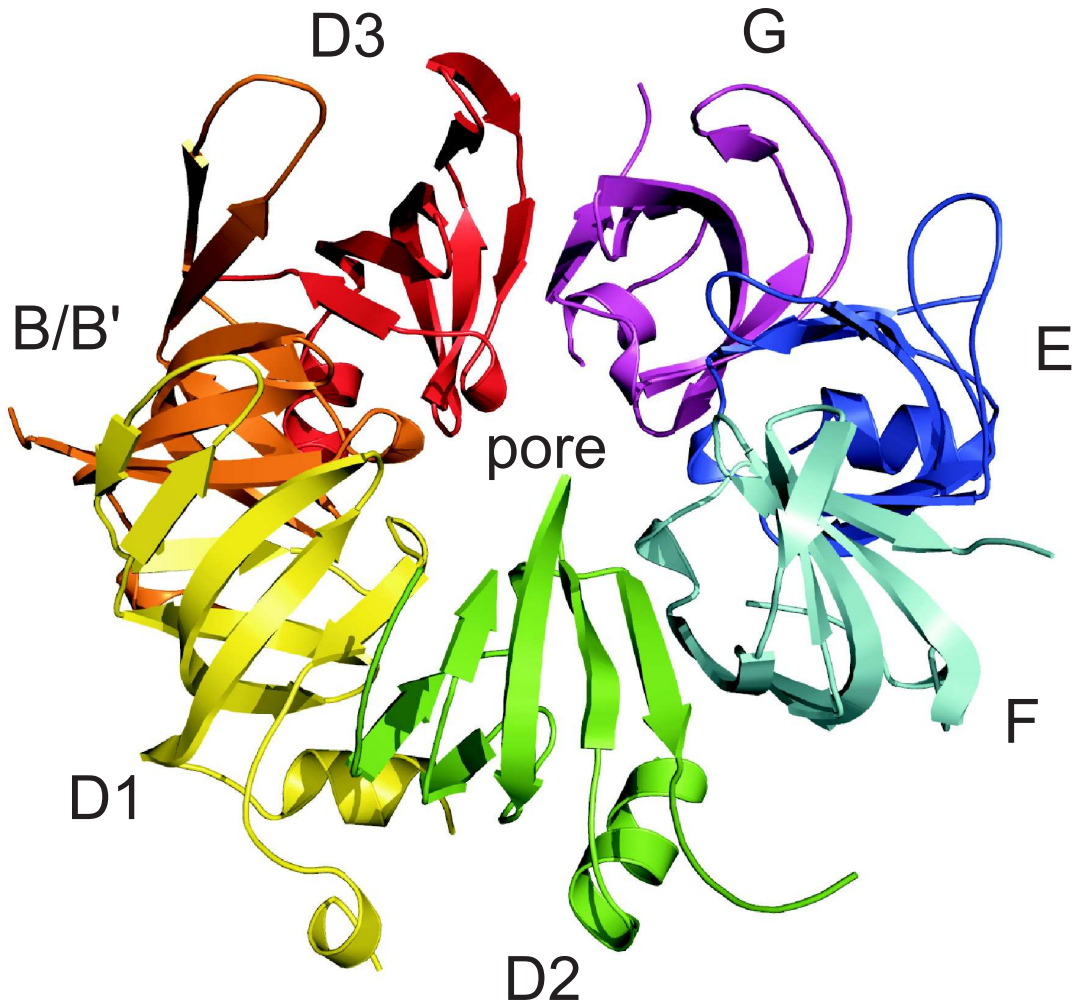
Figure 1.4 Hypothetical model of the Sm core heptamer. SmB/B' is shown in orange, D1 in yellow, D2 in green, D3 in red, E in blue, F in cyan and G in purple. The model is based on the X-ray structures of D3 complexed with B/B' and D1 with D2 (PDB accession codes 1D3B and 1B34 respectively). Proteins E, F and G were modeled according to the structure of SmF (PDB accession code 1N9S) from yeast by the online-tool HHpred (http://toolkit.tuebingen.mpg.de/hhpred). The single proteins were aligned to the structure of an homoheptameric Sm ring of *archaeoglobus fulgidus* (PDB accession code 1I5L) in an order derived from crystal structures, biochemical experiments and yeast two-hybrid studies (see text for details). The original structures contain the Sm motif only, loop L4 of D2 has been truncated. Clashes, which occurred mainly on the termini were removed manually by deleting the respective residues (Kambach *et al.*, 1999b; Törö *et al.*, 2001; Collins and Guthrie, 2000)

.

the protein-protein interfaces continue the β-sheet structure. The C-termini of some Sm proteins fold to α-helices. The pore of the Sm core as well as the extended β-sheets of D2 and B/B' are basic and presumably interact with the acidic backbone of the snRNA.

1.2.2 U1 snRNP

In HeLa cells, U1 snRNP is the most abundant spliceosomal particle, encompassing a 164 nt snRNA and 10 proteins. A complex of U1 snRNA and seven Sm core proteins is termed U1 core snRNP. This particle provides a platform for the association of the U1 snRNP specific proteins U1-70k, U1-A and U1-C. U1 snRNP has been identified to bind the 5' splice site by base complementarity to its 5' terminus, although this interaction is presumably dispensable for the excision of certain introns (Lazar *et al.*, 1982; Tarn and Steitz, 1994). Concerning the constitution of U1 snRNP a one to one stoichiometry of single protein components and RNA has been experimentally determined. (Hochleitner *et al.*, 2005). The components of U1snRNP and their corresponding molecular masses are listed in table 1.1. U1 snRNA forms four stem loop structures as depicted in figure 1.3 (see page 7). Stem loop I has been shown to interact with U1-70k and stem loop II is tightly associated with the first RRM of U1-A (Patton and Pederson, 1988; Scherly *et al.*, 1989). Several Sm proteins have been shown to interact with the Sm site on the U1 snRNA (Heinrichs *et al.*, 1990; Raker *et al.*, 1999; Urlaub *et al.*, 2001). The Sm core proteins are bound to the Sm motif, spanning from nucleotides 125 to 133. This interaction is likely to be stabilized by stem loop IV. Several distinct species of U1 snRNA with different nucleotide exchanges have been recently identified in HeLa cells (Kyriakopoulou *et al.*, 2006).

The U1snRNP-specific 70k protein contains an N-terminal RRM and two C-terminal SR-like domains that are rich in alternating serine and arginine, glutamate and arginine as well as aspartate and arginine residues. It has to be noted that the arginine and serine contents of SR like domains are lower than in typical RS domains. Apart from making RNA contacts, U1-70k binds to U1-C, SmD2 and SmB/B' as shown by interaction studies as well as by crosslinks. The Sm core proteins are contacted by a sequence N-terminal of the U1-70k RRM allowing an RNA-independent association to the core particle. Surprisingly, this interaction was not found at U2 or U5 snRNPs, suggesting fundamental differences among the Sm core particles (Nelissen *et al.*, 1994).

The SR-like domain of U1-70k is believed to function in the recruitment of other proteins to promote the 5' splice site binding of U1 snRNP and to mediate protein contacts to the 3' splice site via interactions with other SR-proteins. It has been shown that

Table 1.1 Protein composition of U1 snRNP from *Homo sapiens*, the molecular weight (MW) of the corresponding proteins and the RNA is given in kDa. Alternatively spliced proteins and different snRNA species are not accounted for.

Protein	MW	motifs	yeast ortholog
U1 snRNA	54.2	stem loop I, II, III, IV, $m_3^{2,2,7}$G cap	U1 snRNA
SmB/B'	24.6	Sm motif, proline-rich	Smb1
SmD1	13.3	Sm motif, RG repeats	Smd1
SmD2	13.5	Sm motif	Smd2
SmD3	13.9	Sm motif, RG repeats	Smd3
SmE	10.8	Sm motif	Sme1
SmF	9.7	Sm motif	Smf1
SmG	8.4	Sm motif	Smg1
U1-70k	51.5	RRM, RS-like domain	Snp1
U1-A	31.1	2 RRMs	Mud1
U1-C	17.3	C_2H_2-type zinc finger motif	Yhc1

several factors directly interact with the SR-like domain of U1-70k, like the SR proteins ASF/SF2 or Sc35. These contacts are modulated by kinases, e.g. SFRS protein kinase I (SRPK1) that reversibly phosphorylates serines in RS domains. The binding site for ASF/SF2 on the 70k protein was narrowed down to 20 consecutive amino acids and the interface locates to the SR domains of both proteins (Cao *et al.*, 1997; Wu and Maniatis, 1993). U1-70k localizes to the nucleus independently of U1 snRNA (Romac *et al.*, 1994).

The N-terminal RRM of U1-A contacts the second stem loop of U1 snRNA in a sequence-specific fashion. This has been studied in detail by binding assays and X-ray crystallography (Jessen *et al.*, 1991; Oubridge *et al.*, 1994a). A schematic view of the structure is shown in figure 1.5. A hexanucleotide sequence of U1 snRNA interacts extensively with the conserved RRM and its C-terminal extension. The ten-nucleotide RNA loop binds to the surface of the four-stranded beta-sheet of the RRM. The binding is mediated by RNA bases, aromatic protein side chains and an extensive hydrogen bonding network.

The structure of the second RRM of U1-A has been solved by nuclear resonance spectroscopy (NMR). Although typical RNA binding features are present, an affinity for snRNAs or random sequences was not detectable (Lu and Hall, 1995, 1997).

The two RRMs of U1-A are interspersed with a proline-rich region of unknown function. This region is absent from the U2-B" protein, which is related to U1-A in architecture, sequence and presumably function (Nelissen *et al.*, 1991b). Two molecules of U1-A
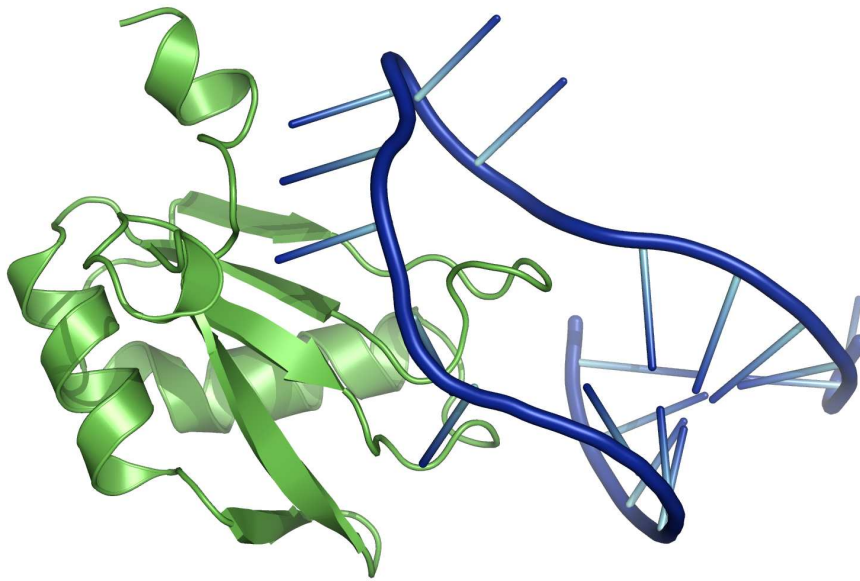
Figure 1.5 Crystal structure of U1-A (green), complexed with a 21 nucleotide hairpin of U1 snRNA (blue) (Oubridge *et al.*, 1994a)

.

can dimerize in the presence of U1-A mRNA. This results in an interface, that has an autoregulatory function by inhibiting poly(A) polymerase (Guan *et al.*, 2003). It has been also shown that in the absence of RNA binding, residues 103-115 of U1-A mediate dimerization (Klein Gunnewiek *et al.*, 2000).

The U1-C protein contains a C_2H_2-type zinc finger motif and does not bind to the snRNA but contacts U1-70k and B/B'. The interaction to U1-70k is mediated by the N-terminal zinc finger, which is not accessible to antibodies when U1-C is bound to U1 snRNP (Dumortier *et al.*, 1998). Furthermore, U1-C has been implicated in 5' splice site binding by UV crosslinking assays (Nelissen *et al.*, 1994, 1991a; Rossi *et al.*, 1996; Du and Rosbash, 2002). Like U1-A the C protein is dispensable for exon bridging and has been shown to dimerize in the absence of U1 snRNP (Gunnewiek *et al.*, 1995). The protein TIA-1 interacts with U1-C to enhance U1 snRNP binding to weak 5' splice sites (Förch *et al.*, 2002). It was recently shown that U1-C is a target for methylation by the methyltransferase CARM1 (Cheng *et al.*, 2007). The structure of the U1-C zinc finger was solved by NMR (Muto *et al.*, 2004). The stoichiometric assembly of the entire U1 snRNP from subunits expressed in *E. coli* and *in vitro* transcribed RNA has been reported (Muto *et al.*, 2001). The isolation of U snRNPs by immunoaffinity chromatography permitted structural studies of natively isolated U1 snRNP by electron
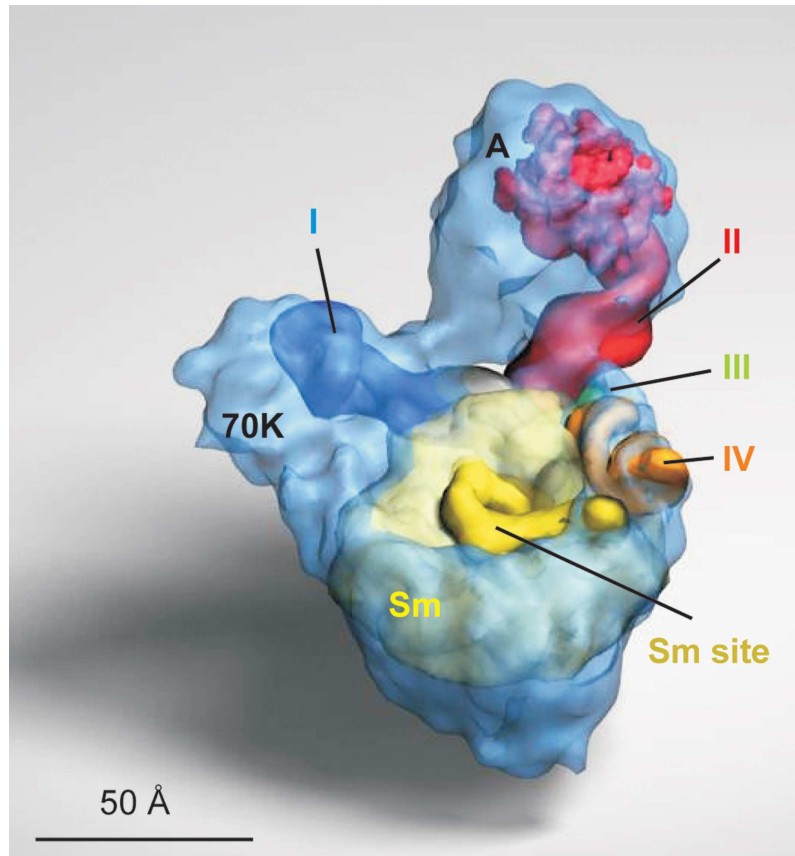
Figure 1.6 3D cryo-EM structure of HeLa U1 snRNP at 10 Å resolution. The numbers I, II, III and IV indicate the presumed location of the respective stem loops, 70k and A the U1 snRNP specific proteins (Stark and Lührmann, 2006)

.

microscopy (Bringmann *et al.*, 1983). It was furthermore possible to localize U1-A, U1-70k and the $m_3^{2,2,7}G$ cap by protein depletion and antibody binding experiments (Kastner and Lührmann, 1989; Kastner *et al.*, 1992). The structure of U1 snRNP from HeLa cells has been determined at 10-14 Å resolution by cryo-electron microscopy (cryo-EM) (Stark *et al.*, 2001). Two prominent protuberances could be attributed to the presence of U1-A and U1-70k including the respective stem loops as shown in figure 1.6. The hypothesized heptameric Sm core ring can be entirely placed in the EM-envelope.

The biogenesis of U1 snRNP (U1) involves an export of the nascent snRNA to the cytoplasm which is dependent on the m_7G cap. In the cytoplasm, Sm proteins are loaded on the U1 snRNA, which is followed by a hypermethylation of the m_7G cap. The latter depends on the presence of the Sm proteins. The recognition of the $m_3^{2,2,7}G$ cap structure and the nuclear import of the particle are carried out by snurportin1 (Spn1) and importin beta, respectively (Strasser *et al.*, 2005).

1.2.3 U2 snRNP

12S U2 snRNP consists of a 187 nt U2 snRNA (see figure 1.3 on page 7), the Sm core proteins, and two U2-specific proteins, U2-A' and U2-B". The X-ray crystal structure of U2-A' and U2-B" in complex with a U2 snRNA hairpin revealed that the binding mode of U2-B" to an U2 snRNA hairpin resembles the interaction of U1-A with U1 snRNA (Price *et al.*, 1998). The U2 specific, as well as the Sm core proteins, are located in the 3' region of the U2 snRNA whereas the 5' region of U2 snRNA is heavily modified and these modifications are a prerequisite for splicing (Dönmez *et al.*, 2007).

In addition to 12S U2, a 17S U2 snRNP was isolated containing the core U2 snRNP, a set of additional proteins as well as two multi-protein complexes essential for splicing, termed SF3a and SF3b. (Krämer, 1988; Krämer and Utans, 1991). The 17S U2 snRNP

Table 1.2 Protein composition of U2 snRNP, SF3a and SF3b from *Homo sapiens*. The molecular weight (MW) of the corresponding proteins and the RNA is given in kDa. Alternatively spliced proteins and different snRNA species are not accounted for.

Protein	MW	motifs	yeast ortholog
U2 snRNA	60.7	stem loop I, IIa, IIb III, IV, $m_3^{2,2,7}$G cap	U2 snRNA
SmB/B'	24.6	Sm motif, proline-rich	Smb1
SmD1	13.3	Sm motif, RG repeats	Smd1
SmD2	13.5	Sm motif	Smd2
SmD3	13.9	Sm motif, RG repeats	Smd3
SmE	10.8	Sm motif	Sme1
SmF	9.7	Sm motif	Smf1
SmG	8.4	Sm motif	Smg1
U2-A'	31	Leucine rich repeat	Lea1
U2-B"	28.5	2 RRMs	Msl1
SF3a-120	88.9	2 SURP domains, proline-rich domain, ubiquitin domain	Prp21
SF3a-66	49.2	Zinc finger, proline-rich	Prp11
SF3a-60	58.9	Zinc finger, glutamate-rich	Prp9
SF3b-155	145.8	HEAT repeats	Hsh155
SF3b-145	97.7	Proline rich, SAP domain	Cus1
SF3b-130	135.6		Rse1
SF3b-125	31.1	DEAD box helicase, Q motif	-
SF3b-49	44.4	2RRMs	Hsh49
SF3b-14a	15.5	RRM	Snu17
SF3b-14b	12.4	PHD finger-like	Rds3
SF3a-10	10.1	-	Rcp10

has been studied by electron microscopy, which allowed a global assignment of its additional large substructures in relation to 12S U2 snRNP (Krämer *et al.*, 1999). Table 1.2 compiles the proteins and the motifs present in 17S U2 snRNP.

SF3a and SF3b are both stable at high salt but dissociate from U2 snRNP at moderate salt concentrations. SF3a as well as SF3b are known to contact U2 snRNP with multiple proteins like SF3b-49, SF3b-14 or SF3a-60 (Dybkov *et al.*, 2006). The entire SF3b particle from HeLa cells has been solved by 3D cryo-EM at 7 Å resolution and as part of the U11/U12 di-snRNP. (Golas *et al.*, 2005, 2003). These structures allow conclusions about global architecture of the particles, though it is difficult to localize single protein components or a conformational change. So far two protein-protein-interfaces of SF3a have been mapped in *Homo sapiens* (Krämer *et al.*, 1995; Nesic and Krämer, 2001). In SF3a a relatively small fraction of the 120k protein constitutes a binding platform for 60k and 66k. An N-terminal peptide of 60k interacts with the second SURP domain and 66k binds to a region C-terminal of the second SURP domain of 120k (Krämer *et al.*, 1995). Only little is known about the three-dimensional architecture of SF3a. Recently, structures of the first SURP domain and of an interface of the second SURP domain, complexed with a peptide of SF3a-60, have been determined by NMR (Kuwasako *et al.*, 2006).

1.2.4 U4/U6·U5 tri-snRNP

The U4/U6·U5 tri-snRNP consists of 16 particle-specific proteins, a double set of Sm core proteins as well as a set of LSm proteins and three RNAs (see figure 1.3 on page 7). At higher salt concentrations, the tri-snRNP dissociates into two particles, U4/U6 and U5 snRNP. 13S U4/U6 di-snRNP consists of the particle-specific proteins 90k, 60k, 20k, 61k, 15.5k as well as Sm and LSm proteins. U4 and U6 snRNAs are interacting due to a large stretch of base complementarity, whereas their association with U5 snRNP is mediated through protein-protein interactions (Black and Pinto, 1989). All components of U4/U6·U5 tri-snRNP and known motifs are listed in table 1.3.

Several interactions in tri-snRNP have been mapped by yeast-two hybrid assays, pull-down experiments or salt treatment of the particles (Liu *et al.*, 2006; Achsel *et al.*, 1998). A specific function in splicing could be attributed to various tri-snRNP proteins, several have been shown to be essential for splicing *in vitro* and *in vivo*, for example 110k, 90k or 60k (Hu *et al.*, 1994; Makarova *et al.*, 2001; Makarov *et al.*, 2000). The 15.5k protein is a nucleation factor for U4/U6 di-snRNP biogenesis and contacts U4 snRNA as well as 61k. These interactions have been characterized in detail by X-ray

Table 1.3 Protein composition of U4/U6·U5 tri-snRNP from *Homo sapiens*, the molecular weight (MW) of the corresponding proteins and the RNAs is given in kDa. Alternatively spliced proteins and different snRNA species are not accounted for.

Protein	MW	motifs	yeast ortholog
U4 snRNA	47.8	3', 5' central stem loop, $m_3^{2,2,7}$G cap	U4 snRNA
U5 snRNA	38.2	stem loop I and II, $m_3^{2,2,7}$G cap	U5 snRNA
U6 snRNA	34.9	5' stem loop, γ-monomethylphosphate cap	U6 snRNA
SmB/B'	24.6	Sm motif, prolin-rich	Smb1
SmD1	13.3	Sm motif, RG repeats	Smd1
SmD2	13.5	Sm motif	Smd2
SmD3	13.9	Sm motif, RG repeats	Smd3
SmE	10.8	Sm motif	Sme1
SmF	9.7	Sm motif	Smf1
SmG	8.4	Sm motif	Smg1
LSm2	10.8	Sm motif	Lsm2
LSm3	11.8	Sm motif	Lsm3
LSm4	15.3	Sm motif	Lsm4
LSm5	9.9	Sm motif	Lsm5
LSm6	9.1	Sm motif	Lsm6
LSm7	11.6	Sm motif	Lsm7
LSm8	10.4	Sm motif	Lsm8
220k	273.6	JAB1/MPN, RRM	Prp8
200k	244.5	2 DEAD box helicase domains	Brr2
116k	109.4	GTPase domain (EF-2-like)	Snu114
110k	90.3	SR-like, coiled coil	Snu66
102k	106.9	TPR elements	Prp6
100k	95.6	SR domain, glutamate-rich, DEAD box helicase	Prp28
90k	77.5	PWI domain	Prp3
65k	65.3	SR-like, ubiquitin hydrolase fold	Sad1
61k	55.5	NOP domain	Prp31
60k	58.4	WD repeats	Prp4
40k	39.9	WD repeats	-
27k	18.9	SR domain	-
20k	19.2	PPIase domain	-
15.5k	14.2	-	Snu13
15k	16.8	Thioredoxin-like	Dib1

crystallography (Nottrott *et al*., 1999; Liu *et al*., 2007). The 20k, 60k and 90k proteins form a salt-stable complex and are components of U4/U6 di-snRNP (Wang and Manley, 1997). 20k comprises a cyclophilin fold, but the function of cyclophilins in splicing is yet unclear. The 102k protein contacts 61k and seems to constitute a bridge between 20S U5 and 13S U4/U6 snRNPs (Makarov *et al*., 2000). Apart from U4/U6·U5 tri-snRNP a 20S U5 snRNP can be isolated from nuclear extract by immunoaffinity chromatography. It is still unclear if this particle is a breakdown product of the splicing reaction or a precursor for tri-snRNP formation. Apart from the tri-snRNP specific proteins, that are bound to U5 snRNP, an additional factor, the 52k protein, was found in 20S U5 snRNP. The 52k protein contacts 15k with its GYF domain and is essential for splicing (Laggerbauer *et al*., 2005; Liu *et al*., 2006). The 200k protein is a DEAD box helicase and was shown to be essential for the second step of splicing. Its presumable function is the ATP-dependent dissociation of U4 snRNA at the transition of B and B* complexes (Lauber *et al*., 1996; Laggerbauer *et al*., 1998). Furthermore the protein is part of an RNA-free protein complex that has been isolated by sodium thiocyanate treatment of U5 snRNP. This tetrameric complex consists of 220k, 200k, 116k and 40k (Achsel *et al*., 1998). The 220k protein has been shown to contact several regions of the pre-mRNA during splicing as well as the U5 snRNA (Reyes *et al*., 1999; Chiara *et al*., 1996; Urlaub *et al*., 2000). Furthermore the U5 specific 220k protein contains a JAB/MPN domain and a conserved RRM. Several mutations in 220k that occur in *Homo sapiens* are linked to *renititis pigmentosa*, a disease of the eyes leading to impaired vision or blindness (Pena *et al*., 2007). The U5-116k protein resembles the fold of a ribosomal translocase, EF-2. 116k presumably consumes GTP and is likely to be involved in snRNP biogenesis (Brenner and Guthrie, 2006). Three proteins containing an SR-like domain are specific for U4/U6·U5 tri-snRNP, these are 110k, 65k and 27k. 65k and 110k have been shown to be essential for splicing, their function could be the tethering of tri-snRNP to the A complex (Fetzer *et al*., 1997; Makarova *et al*., 2001).

The presumed function of U4/U6·U5 tri-snRNP in pre-mRNA splicing could be a remodeling of the catalytic core by RNA helicase activity, since the involvement of these helicases in several steps of the reaction has been documented (Staley and Guthrie, 1998). An EM structure of tri-snRNP was recently published that revealed the global architecture of the particle also allowing its placement in higher order structures like the A complex (Sander *et al*., 2006). A minor portion of the entire complex has been characterized in atomic detail, but the topological arrangement of most U4/U6·U5 tri-snRNP proteins is yet unknown (Liu *et al*., 2007; Reuter *et al*., 1999; Pena *et al*., 2007).

1.3 Nuclear pre-mRNA splicing is a highly dynamic and complex process

During nuclear pre-mRNA splicing several structural and compositional rearrangements of the spliceosomal components take place (Makarov *et al*., 2002). It was observed early that the first 10-12 nucleotides of U1 snRNA are complementary to the 5' splice site of the pre-mRNA substrate (Avvedimento *et al*., 1980). Later it was verified that U1 snRNP binds to the 5' splice site by base pairing and thus contacts the nascent RNA polymerase II transcript (Lazar *et al*., 1982). This interaction is accompanied by a multitude of hnRNPs (heterogeneous nuclear ribonucleoproteins) binding to the pre-mRNA substrate. This early spliceosomal complex (E complex) recruits further protein factors like U2AF or SF2/ASF to the branch point, the polypyrimidine tract and the 3' splice site (Zamore *et al*., 1992; Berglund *et al*., 1998). An auxiliary class of factors, termed SR proteins are believed to bring the 5' and 3' splice sites in proximity to each other for the reaction to follow (Wu and Maniatis, 1993).

When U2 snRNP contacts the 3'splice site, it is complexed with two multi-protein subunits, termed SF3a and SF3b, forming a 17S particle. Whether SF3a and SF3b are permanently associated with U2 snRNP is yet unclear. The protein factors U2AF and presumably SF1 promote the recruitment of U2 snRNP to the 3' splice site (Krämer, 1992; Krämer *et al*., 1990; Brosi *et al*., 1993). Upon U2 snRNP binding, U2AF and SF1 dissociate. It has been hypothesized that a catalytically important residue, the branch point adenosine, is bulged out of the RNA by U2 snRNP to promote catalysis (Berglund *et al*., 2001; Pascolo and Séraphin, 1997; Query *et al*., 1994). As shown in figure 1.7, the stage of assembly, where U1 and U2 snRNPs are bound to the pre-mRNA, is termed A complex and its formation is ATP dependent. Further the recruitment of U4/U6·U5 tri-snRNP to Complex A follows and is likely to be mediated by SR-Proteins. (Roscigno and Garcia Blanco, 1995; Bindereif and Green, 1987). The assembly state, where all snRNAs are present, is termed B complex and constitutes an entry point to first major structural and compositional rearrangements. U1 and U4 snRNPs dissociate from the spliceosome which is promoted by ATP-dependent RNA helicases and concomitantly a protein complex, the Prp19-associated splicing complex, is recruited. The function of the Prp19 complex during splicing has not been elucidated in detail yet (Chan *et al*., 2003; Tarn *et al*., 1993). During the rearrangements of the spliceosome, B complex is converted to B* complex. Base pairing contacts between U2 snRNA and U6 snRNA apparently facilitate the formation of a catalytic center that promotes the first transes-
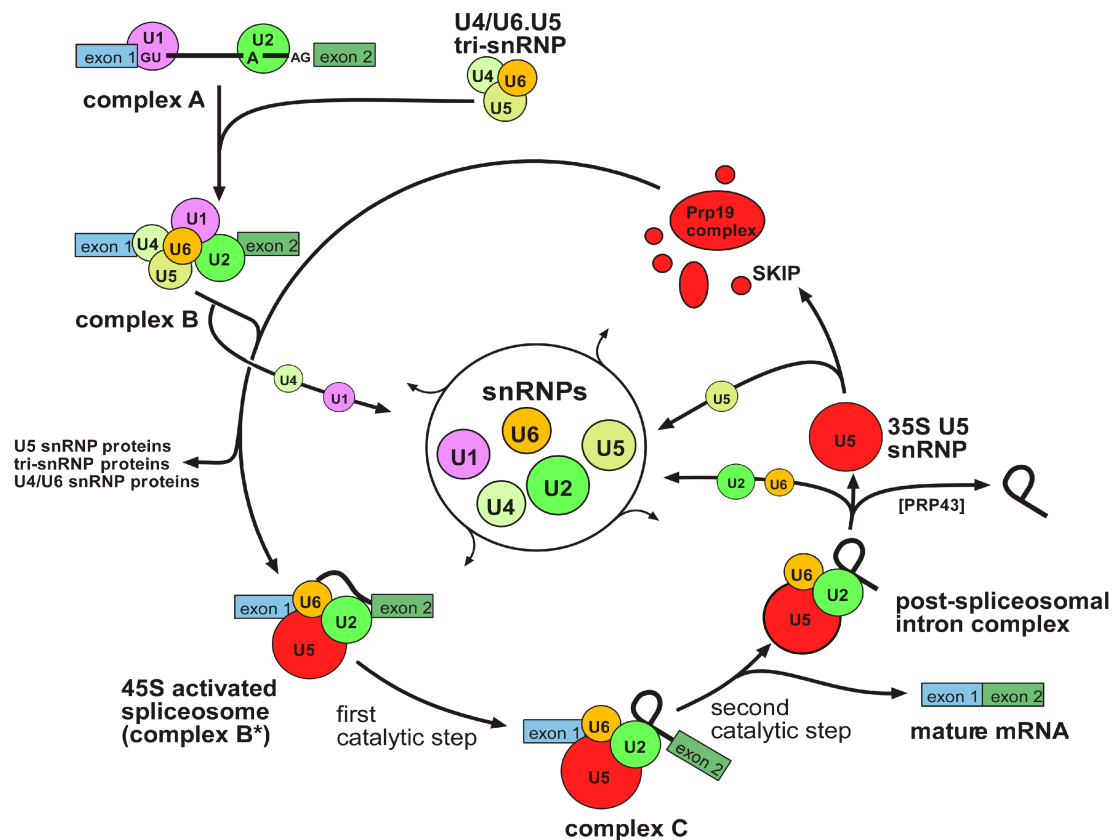
Figure 1.7 Schematic model of the spliceosomal cycle. Ribonucleocomplexes are color coded as follows. U1 snRNP is shown in purple, U2 snRNP in dark green, U4 snRNP in medium green, U5 snRNP in light green, U6 snRNP in orange. U5snRNP, with Prp19 complex bound, as well as Prp19 complex and its components are shown in red (Makarov *et al*., 2002)

.

terification reaction. Thereby U6 snRNA and U2 snRNA are in the vicinity of the 5' splice site (Rhode *et al*., 2006). It is hypothesized that the protein factor Prp2 arranges the branch point adenosine in the vicinity of the 5' splice site (Roy *et al*., 1995).

The transition from complex B* to complex C requires ATP. At this stage the U4/U6·U5 tri-snRNP proteins 116k, 220k, 110k as well as Slu7 contact the 3' splice site (Chiara *et al*., 1997; Chua and Reed, 1999). After the second transesterification is carried out, the mature pre-mRNA and a post-spliceosomal intron complex comprising the lariat, U6, U5 and U2 snRNP exit the reaction. The lariat as well as U6 and U2 snRNPs are presumably released from this complex resulting in a 35S particle, containing U5 snRNP and the Prp19 complex (Makarov *et al*., 2002).

1.4 Basic principles of X-ray crystallography

1.4.1 Obtaining structures from crystals by X-ray diffraction

X-rays are electromagnetic radiation and were discovered by Wilhelm Conrad Röntgen in 1895. It was thereafter realized that they can be used for the structure determination of salt crystals, small molecules and later protein crystals. Protein crystals can be grown from supersaturated solutions of the respective molecule in the presence of organic and inorganic precipitants. Since a large chemical space is explored by screening these precipitant conditions, protein-crystallography requires huge amounts of material. The major bottleneck of this procedure is still to obtain a crystal, that has sufficient diffraction capabilities. The following section is compiled from regular textbooks about solid state physics and X-ray crystallography (Rhodes, 2006; Ashcroft and Mermin, 1976).

A crystal can be regarded as a regular periodic array of atoms or molecules. The fundamental principle underlying crystallography is the scattering of X-rays by the electrons in molecules that form a crystal lattice. The emitted or scattered radiation is a spherical wave that is enhanced by the periodicity of the crystal lattice and results in a defined periodical distribution of electromagnetic waves surrounding the crystal. Therefore a crystal exposed to X-rays produces a diffraction pattern of regularly spaced spots on a detector or X-ray film. With current X-ray sources and detectors, diffraction resulting from a single molecule would hardly be detectable. Taken together the main goal of X-ray crystallography is to determine the density of electrons throughout the crystal. According to Bragg's Law a set of parallel planes with index hkl and interplanar spacing of d_{hkl} produces a diffracted beam if X-rays of wavelength λ impinge upon the planes at an angle θ and are reflected on the same angle, if θ meets the condition

$$2d_{hkl} \sin \theta = \lambda n$$

in which n is an integer. In a diffraction experiment the intensity (I_{hkl}) and the position (hkl) of each reflection, which depends on the angle of the X-rays relative to the source beam, is measured. Each X-ray that is recorded as a reflection represents the sum of the contribution to the scattering of all scatterers in the unit cell. The sum that describes a diffracted X-ray at position hkl is called structure factor F and can be expressed as

$$F_{hkl} = fA + fB + fC \ldots$$

in which A, B and C represent different atoms in the unit cell.

Electromagnetic radiation used in X-ray crystallography is a complex wave that can be approximated by a *Fourier sum*, which is composed of single wave equations or *Fourier terms*. According to Fourier, for any function $f(x)$ there exists another function $F(x)$, which is also used in X-ray crystallography to convert real space into reciprocal space. All variables of $F(x)$ have reciprocal values. This explains the reciprocal relation between atomic coordinates in the crystal lattice and the diffraction pattern. The contribution of a single atom j to the structure factor is called atomic structure factor and can be expressed as

$$f_{hkl} = f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

where f_j is the scattering factor of the respective atom, x, y and z are the fractional atomic coordinates in the unit cell and h, k and l are the indices or frequencies in the x, y or z direction. Fourier sums in crystallography represent the distribution of scatterers in a given unit cell by a complex three-dimensional wave function like

$$f(x, y, z) = \sum_{h,k,l} F_{hkl} e^{2\pi i(h_x + k_y + l_z)}.$$

Each term in the sum is a three-dimensional wave whose frequency is h in the x-direction, k in the y-direction and l in the z-direction. For each possible set of values, h, k, and l the associated wave has amplitude F_{hkl} and phase α_{hkl}. To calculate the electron density σ at a given position of the crystal lattice x, y and z, the volume of the unit cell is included in the Fourier sum.

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l} F_{hkl} e^{-2\pi i(h_x + k_y + l_z)}.$$

Two of the three parameters constituting this complex wave can be recorded in a diffraction experiment, that is amplitude and frequency. The amplitude is proportional to the square root of the intensity and can thus be calculated. The frequency of a structure factor is equal to d_{hkl} and can be deduced from the diffraction pattern. The phase of the reflections is unknown and must be determined experimentally to solve the structure. This can be achieved in several ways. According to Friedel's law the intensity of I_{hkl} equals the intensity of $I_{\overline{hkl}}$ but F_{hkl} does not equal $F_{\overline{hkl}}$, since there is a difference in phase angles (Rhodes, 2006).

This can be expressed as a complex term with an imaginary number i and a phase angle α

$$F = |F|e^{i\alpha}$$

This term can be explicitly included in the equation describing the distribution of electron density ρ for a specific reflection α_{hkl} as

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l} |F_{hkl}|e^{-2\pi i(h_x+k_y+l_z)-\alpha'}$$

where the phase angle α is expressed as $2\pi\alpha'$.

1.4.2 Solutions to the phase problem

An approach to solve the phase problem is the derivatization of the crystal with a heavy atom like mercury or lead that scatters more powerfully than the regular atoms in the unit cell which are mainly carbon, nitrogen, oxygen, sulphur, phosphorus and hydrogen. In order to locate the heavy atoms that give rise to the difference in the diffraction intensities of a native and derivatized crystal, Patterson maps are calculated from structure factors analogous to the electron density map, but without phases. The Patterson map displays the length of all interatomic vectors of the unit cell with the vector tail at the origin and the head resulting in a peak at the position u, v and w. By subtracting the structure factors obtained from the derivatized crystal from that of the native crystal, a difference Patterson function, or a Fourier transform of the structure factor differences, is calculated. The interatomic vectors of the protein (P) are removed, which reveals peaks of the heavy atom (H) that are emphasized in the map as H-H vectors. H-P vectors and P-P vectors are down-scaled in the map. This way the location of the heavy atom in the unit cell and the calculation of the heavy atom structure factors including the phase angles is possible. With this information the phases of the structure factor contribution of the protein can be obtained.

Anomalous scattering occurs when the wavelength of the incident beam lies near the absorption peak of a heavy atom, which is positioned in the unit cell. A fraction of the radiation is emitted with an altered phase. This breaks Friedel's Law or the symmetrical relation between F_{hkl} and $F_{\overline{hkl}}$ in matters of intensity and phase. From this disparity in phase information of single heavy atoms, the remaining unit cell content can be determined by Patterson methods to locate heavy atom substructures. From that

and the knowledge of protein and protein-heavy atom diffraction intensities, the protein structure factors can be calculated.

In molecular replacement, known structures can be used as phasing models if they are similar to the target structure. This is achieved by calculating structure factors F_{calc} from the model and comparing those with the target structure factors F_{obs}. If a correlation is found, phases from the model α_{calc} are extracted and used to initially phase the target structure and to obtain an interpretable map (Rhodes, 2006).

1.5 Structural studies of macromolecular spliceosomal assemblies

1.5.1 Structural studies of spliceosomal particles

X-ray crystallography is a powerful tool to determine three-dimensional structures of biological macromolecules at atomic resolution. This strongly depends on the crystal's diffraction capabilities and the accessibility of phase information. Hitherto several macromolecular assemblies have been studied by X-ray crystallography like the ribosome or RNA polymerase (Ban *et al.*, 1998, 2000; Cramer *et al.*, 2000). These structures gave detailed insight into the architecture and mechanists of cellular multi-megadalton machineries that exert metabolic key processes. For protein crystallography comparably huge amounts of the target molecule are required. Some macromolecular assemblies like the ribosome or RNA pol II cannot be assembled from recombinant subunits and, therefore, had to be natively purified from thermophilic bacteria or yeast. Using this approach, a heterogeneity of the particles in terms of posttranslational modifications or additional/lacking subunits cannot be avoided, rendering the growth of suitable crystals difficult.

Likewise it is technically demanding to assemble spliceosomal snRNPs from recombinantly produced subunits and *in vitro* transcripts of the respective RNA. Nevertheless it has been reported that for example U4 core snRNP or U1 snRNPs can be stoichiometrically assembled obtaining yields suitable for X-ray crystallography (Muto *et al.*, 2004). Apart from a recombinant approach, spliceosomal particles can be isolated from a native source like HeLa cells by immunoaffinity chromatography (Bringmann *et al.*, 1983). Up to now, purification protocols for these particles have been constantly improved (Kastner and Lührmann, 1999).

1.5.2 Aims of this work

The goals of this work were the isolation, concentration, crystallization and structure solution of U1 snRNP, tri-snRNP and SF3a. All particles were immunoaffinity-purified in the Lührmann department (Kastner and Lührmann, 1999; Dybkov *et al.*, 2006) from HeLa cells. The isolation of native snRNPs was a prerequisite to set up first crystallization-tion trials of U1 snRNP 10 years ago. Small U1 snRNP crystals were obtained soon, but in an irreproducible manner and with a low diffraction quality of 20 Å (personal communication B. Kastner, U. Reidt). To screen a large number of crystallization conditions for U1 snRNP with a selection of different ligands, larger amounts of the particles are required. The current purification protocol of U1 snRNP should be optimized, to obtain higher yields of particles for high throughput crystallization. Especially the separation of U1 snRNPs and U2 snRNPs has to be improved. To obtain highly diffracting crystals of the particles, potential ligands of U1 snRNP like RNA and DNA oligos, mimicking the 5' splice site or recombinant snurportin1, (see page 13) should be assessed as cofactors for co-crystallization trials.

Some parts of U1 snRNP are assumed to be flexible, which might interfere with crystallization or reduce the diffraction quality of the crystals. U1-70k has a C-terminal SR-like domain that is presumably unstructured and B/B' has a proline-rich extension. These domains are unfavorable for crystallization and should be removed by proteolytic treatment of U1 snRNP. It has been reported that Caspase 3, for example, has a specific cleavage site in U1-70k, removing a large portion of the SR-like domain from the particles (Degen *et al.*, 2000). Also other proteases could be tested. The 5' splice site can be truncated by DNA-guided RNAse H cleavage. This protocol has been utilized previously to inactivate U1 snRNP in nuclear extract and is also applicable for the production of truncated U1 snRNP particles (Behzadnia *et al.*, 2006).

To solve the atomic structure of U1 snRNP, phases have to be obtained in addition to highly diffracting crystals. Heavy atom clusters should be used to obtain initial phase information of U1 snRNP crystals, also at low resolution. U1 snRNP contains several GU-wobble basepairs that represent a binding site for iridium or osmium hexammine compounds (Cate and Doudna, 1996). Several components of U1 snRNP have been solved at atomic resolution (Muto *et al.*, 2004; Kambach *et al.*, 1999a; Oubridge *et al.*, 1994a). If a resolution high enough is achieved for U1 snRNP crystals, molecular replacement can be employed. So far there is no example of a high resolution structure comprising RNA and several proteins, natively isolated from HeLa cells. Therefore the atomic structure of U1 snRNP states the main goal of this work.

A second target for crystallization is U4/U6·U5 tri-snRNP, that can be, like U1 snRNP, isolated by immunoaffinity chromatography. This particle is more complex than U1 snRNP and has a molecular weight of around 1.5 million kDa (see table 1.3 on page 16). So far it has not been tried to purify tri-snRNP for crystallization, since it is sensitive to high salt and aggregates in solution under certain conditions. Therefore, a further goal of this work is to concentrate tri-snRNP and make it accessible for crystallization trials. In line with this, possible reasons for aggregation or dissociation have to be dealt with. A complex biogenesis is required to yield a functional tri-snRNP particle. Therefore the native purification of the particles is necessary to subject U4/U6·U5 tri-snRNP to crystallization trials. Spliceosomal protein particles, like SF3a, can be isolated by immunoaffinity chromatography in amounts suitable for structure determination. This particle, which associates with U2 snRNP under low salt conditions is another target to be purified for X-ray crystallography. For SF3a, a suitable purification protocol has to be devised.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

For stock solutions, chemicals were dissolved in Milli-Q water and sterile-filtered if not stated otherwise. The pH was titrated with either 37 % HCl or 10 M NaOH at room temperature. The chemicals used in this work are listed in table 2.1.

Table 2.1 Chemicals

Compound	Supplier
Acetic acid	Merck, Darmstadt, DE
Acrylamide solutions:	Roth, Karlsruhe, DE
Rotiphorese Gel 30 (30 % acryl-amide, 0.8 % bis-acrylamide)	
Rotiphorese Gel 40 (38 % acryl-amide, 2 % bis-acrylamide)	
Agarose, electrophoresis grade	Invitrogen, Carlsbad, CA ,USA
Ammonium peroxodisulfate (APS)	Merck, Darmstadt, DE
Ammonium sulfate	Fluka, Buchs, CH
Ampicillin, sodium salt	Fluka, Buchs, CH
Adenosine triphosphate (ATP)	Sigma-Aldrich, Steinheim, DE
Borate	Merck, Darmstadt, DE
Bradford reagent	Bio-Rad, Munich, DE
Bromphenol blue, sodium salt	Merck, Darmstadt, DE
Calcium chloride dihydrate	Merck, Darmstadt, DE
Chloramphenicol	Roche, Mannheim, DE
Coomassie Brilliant Blue	Serva, Heidelberg, DE
Complete-EDTA	Roche, Mannheim, DE
Dithiothreitol	Roth, Karlsruhe, DE
Didesoxynucleotides	New England Biolabs, Frankfurt, DE
Ethanol	Merck, Darmstadt, DE

Table 2.1 Chemicals continued ...

Compound	Supplier
Ethylenediamminetetraacetate di-sodium salt (EDTA)	Roth, Karlsruhe, DE
Ethidium bromide solution (10 mg/ml)	Roth, Karlsruhe, DE
Glycerol	Merck, Darmstadt, DE
Glycylglycylglycine	Sigma-Aldrich, Steinheim, DE
Glutathione, reduced	Sigma-Aldrich, Steinheim, DE
2-[4-(2-Hydroxyethyl)-1-piperazino]-ethansulfonic acid (HEPES)	Roth, Karlsruhe, DE
Imidazole	Fluka, Buchs, CH
Isopropyl-β-thiogalactopyranoside (IPTG)	Roth, Karlsruhe, DE
Kanamycin sulfate	Roche, Mannheim, DE
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, DE
Magnesium chloride hexahydrate	Fluka, Buchs, CH
Maltose	Fluka, Buchs, CH
2-Mercaptoethanol	Roth, Karlsruhe, DE
Methanol	Merck, Darmstadt, DE
2-Morpholinoethanesulfonic acid (MES)	Roth, Karlsruhe, DE
NP-40 (Igepal CA 630)	Sigma-Aldrich, Steinheim, DE
Orange-G	Fluka, Buchs, CH
Phenol / Chloroform / Isoamylalcohol (PCI); 50 % [v/v] Phenol, 48 % [v/v] Chloroform, 2 % [v/v] Isoamylalcohol in TE buffer	Roth, Karlsruhe, DE
Phenylmethylsulfonyl fluoride (PMSF)	Roche, Mannheim, DE
Polyethylene glycol 200-35000	Fluka, Buchs, CH
Potassium chloride	Merck, Darmstadt, DE
Potassium hydroxide	J.T Baker, Deventer, NL
2-Propanol	Merck, Darmstadt
Sodium azide	Fluka, Buchs, CH
Sodium carbonate	Merck, Darmstadt, DE
Sodium chloride	Merck, Darmstadt, DE
Sodium hydroxide	Merck, Darmstadt, DE
Sodiumdodecyl sulfate	Serva, Heidelberg, DE
Sodium thiocyanate (Rhodanite)	Fluka, Buchs, CH

Table 2.1 Chemicals continued . . .

Compound	Supplier
N,N,N',N'-Tetraethylenediamide (TEMED)	Sigma-Aldrich, Steinheim, DE
Tris-(hydroxymethyl)aminomethane (Tris)	VWR international, Darmstadt, DE
Urea	Merck, Darmstadt, DE
Xylenecyanol FF	Sigma-Aldrich, Steinheim, DE

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

2.1.2 Heavy atom derivatives

All heavy atom derivatives used in this work were dissolved in either Milli-Q water or in a buffer used for crystallization. Heavy atom solutions were freshly prepared before use and stored at maximum for two weeks at 4 °C. Heavy atom clusters and hexammines are not commercially available and were generously provided by Prof. Robert Huber, MPI of biochemistry, Munich and Gleb Bourenkov, EMBL, DESY, Hamburg (see table 2.2). All other heavy atom compounds used in this work were obtained from Hampton Research in heavy atom kits.

Table 2.2 Heavy atom derivatives

Derivates	Source
$Na_{27}(NaAs_4W_{40}O_{140})$ or W40	Prof. R. Huber, Munich, DE
$(NH_4)_{14}(NH_5P_5W_{30})$ or W30	Prof. R. Huber, Munich, DE
$Cs_6(W_5P_2O_{20}) \times 2H_2O$ or W5	Prof. R. Huber, Munich, DE
$K_6(HSiW_9V_3O_{40}) \times 3H_2O$ or W9	Prof. R. Huber, Munich, DE
$W_6Br_{14} \times 2H_2O$ or W6	Prof. R. Huber, Munich, DE
$[W_3O_2(O_2CCH_3)_6][3CF_3(SO)_3]_2 \times 2H_2O$ or W-trimer	Prof. R. Huber, Munich, DE
$K_6Mo_7O_{22}(O_2)_2 \times 8H_2O$ or Mo7	Prof. R. Huber, Munich, DE
$(NH_4)_4[Ir_3N(SO_4)_6H_2O_3] \times 2H_2O$ or Ir3	Prof. R. Huber, Munich, DE
$[(C_6H_5)_3PAu]_3OBF_4$ or Au3	Prof. R. Huber, Munich, DE
Tetrakis(acetoxymercury)methane or TAMM	Prof. R. Huber, Munich, DE
Ta_6Br_{14} or TaBr	Prof. R. Huber, Munich, DE

Table 2.2 Heavy atom derivatives continued ...

Derivates	Source
$K_{14}(NaP_5W_{30}O_{110}) \times 31H_2O$, or W30A	Gleb Bourenkov, Hamburg, DE
$Na_{16}[(O_3PCH_2PO_3)_4W_{12}O_{36}] \times 40H_2O$ or W12	Gleb Bourenkov, Hamburg, DE
$Li_7[P_2W_{17}O_{61}Co(NC_5H_5)] \times 14H_2O$ or W17	Gleb Bourenkov, Hamburg, DE
CoW_12O_40, or CoW12	Gleb Bourenkov, Hamburg, DE
$[Os(NH_3)_6]Cl_3$ or OsHex	Gleb Bourenkov, Hamburg, DE
$[Ir(NH_3)_6]Cl_3$ or IrHex	Gleb Bourenkov, Hamburg, DE
$[Os(N_2)(NH_3)_5]Cl_2$ or OsPen	Gleb Bourenkov, Hamburg, DE

2.1.3 Buffers and solutions

All buffers and solutions were prepared with Milli-Q water and were sterile-filtered. The pH of buffers was adjusted by the addition of buffered stock solutions or by titration with 37 % HCl or 10 M NaOH if not stated otherwise. All solution components of the buffers are given as [v/v], if not stated otherwise. The buffers and solutions used in this work are listed in table 2.3.

Table 2.3 Buffers and solutions

Buffer	Composition
Coomassie staining solution	42.5 % ethanol, 10 % acetic acid, 0.6 % Coomassie Brilliant Blue R250
Crystallization buffer	10 mM Tris-HCl, pH 7.9, 100 mM NaCl, 1 mM DTT
Crystallization buffer 120	20 mM Tris-HCl, pH 7.9, 120 mM NaCl, 1 mM DTT
Crystallization buffer ATP	20 mM Tris-HCl, pH 7.9, 120 mM NaCl, 1 mM DTT, 1.5 mM $MgCl_2$, 1.5 mM ATP
Crystallization buffer AS	20 mM Tris-HCl, pH 7.9, 100 mM NaCl, 100 mM $Na_2(SO_4)$ 1 mM DTT
Destaining solution I	45 % ethanol, 10 % acetic acid
Destaining solution II	20 % acetic acid

Table 2.3 Buffers and solutions continued . . .

Buffer	Composition
Ethylenediamminetetraacetate Na / K, pH 8.0	0.5 M EDTA, titrated to pH 8.0 with NaOH or KOH
Formamide RNA loading-dye	90 % formamide, 0.025 % [w/v] bromphenol blue, 0.025 % [w/v] xylenecyanol, 50 mM EDTA
G_{600}	20 mM HEPES-NaOH-KOH, pH 7.9, 1.5 mM $MgCl_2$, 600 mM NaCl, 0.5 mM DTT, 0.25 mM PMSF
Glycerol gradient buffer	20 mM HEPES-NaOH-KOH, pH 7.9, 1.5 mM $MgCl_2$, 150 mM KCl, 0.5 mM DTT, 0.25 mM PMSF
H_0 buffer	20 mM Tris-HCl, pH 7.9, 1.5 mM $MgCl_2$, 1 mM DTT
H_{50} buffer	20 mM Tris-HCl, pH 7.9, 50 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT
H_{1000} buffer	20 mM Tris-HCl, pH 7.9, 1000 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT
Imidazole	Imidazole, titrated with NaOH to pH 8.0
10 × PBS	1.5 M NaCl, 25 mM KCl, 80 mM Na_2HPO_4, 15 mM KH_2PO_4, pH 7-8
6 × protein loading dye	60 mM Tris-HCl, pH 6.8, 50 mM DTT, 1 mM EDTA, 15 % glycerol, 2 % [w/v] SDS, 0.1 % [w/v] bromphenol blue
4 × resolving gel buffer	1.5 M Tris-HCl, pH 8.8, 0.4 % SDS
Protein purification buffer	20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 2 mM DTT
Roeder A buffer	20 mM Hepes-NaOH-KOH, 3 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTE, the pH was adjusted to 7.9 with KOH
Roeder C250 buffer	20 mM Hepes-NaOH-KOH, 250 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTE, the pH was adjusted to 7.9 with KOH
Roeder C420 buffer	20 mM Hepes-NaOH-KOH, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTE, the pH was adjusted to 7.9 with KOH

Table 2.3 Buffers and solutions continued ...

Buffer	Composition
U1 derivatization buffer	50 mM MES-NaOH, pH 6.5, 100 mM NaCl, 10 % Polyethylene glycol (PEG) 4000
10 × SDS running buffer	250 mM Tris, 2 M Glycine, 1 % SDS
4 × stacking-gel buffer	500 mM Tris-HCl, pH 6.8, 0.4 % [w/v] SDS
10 × TBE, pH 8.3	1 M Tris, 0.82 M boric acid, 25 mM EDTA
10 × trypsin buffer	1 M Tris-HCl, pH 9.0, 100 mM $CaCl_2$

2.1.4 Oligonucleotides

The DNA- and RNA oligonucleotides listed in table 2.4 were used in this work. DNA oligonucleotides for crystallization were ordered at IBA Göttingen and dissolved in Milli-Q water at a concentration of $1 - 2$ nmol/µl . RNA oligonucleotides were ordered at Dharmacon, Lafayette, CO, USA and dissolved in Milli-Q water at a concentration of $10 - 50$ µg/µl. Some of the sequencing primers were directly available at Seqlab.

Table 2.4 Oligonucleotides

Name	Type	Sequence	Supplier
DNA1	DNA	CAGGTAAGTAT	MWG
DNA2	DNA	AGGTAAGTATCCCCCGGGG	IBA
DNA3	DNA	AGGTAAGT	IBA
DNA4	DNA	AGGTAAGTA	IBA
RNA1	RNA	AGGUAUGU	Dharmacon
RNA2	RNA	CAGGUAAGU	Dharmacon
RNA3	RNA	AGGUAAGUAU	Dharmacon
RNA4	RNA	AGGUAAGUAUG	Dharmacon
RNA5	RNA	AGGUAAGU	Dharmacon
RNA6	RNA	AGGUAGG	Dharmacon
RNA7	RNA	GGUAAGU	Dharmacon
RNA8	RNA	AGGUAAGUA	Dharmacon
RNA9	RNA	CUGCCAGGUAAGU	Dharmacon
RNA10	RNA	GUAAGU	Dharmacon

2.1.5 Kits

2.1.5.1 Crystallization Kits

All kits for crystallization screening used in this work are listed in table 2.5. The solutions were transferred to 96-well crystallization plates for automated pipetting.

Table 2.5 Crystallization kits

Kit	Supplier
Nextal tubes (Classics Suite, Classics Lite Suite, Cryos Suite, PEGs Suite, AmSO4 Suite, MPD Suite, Anions Suite, Cations Suite, pHClear Suite, pHClear II Suite, MBClass Suite, MBClass II Suite, Protein Complex Suite, PEGs II Suite, ComPAS Suite, PACT Suite, Nucleix Suite, JCSG+ Suite, Classics II Suite, SM II suite, SM III suite)	Qiagen, Hilden, DE
Hampton I, Hampton II, Index, Additives I-III, Detergents I-III, Additive 96-well	Hampton research, Aliso Viejo, CA, US
Heavy atom screens Au, Pt, Hg, M1, M2	Hampton research, Aliso Viejo, CA, US
Wizard I, Wizard II	Molecular dimensions, Suffolk, UK
Izit protein crystal stain	Hampton research, Aliso Viejo, CA, US
Optimize - Low meltingpoint agarose	Hampton research, Aliso Viejo, CA, US

2.1.6 Enzymes and proteins

All enzymes used in this work are listed in table 2.6. Proteases were dissolved in Milli-Q water if not stated otherwise.

Table 2.6 Enzymes and proteins

Protein	Supplier
BSA (10 mg/ml)	New England Biolabs, Frankfurt/Main, DE
Calpain, dissolved in Milli-Q water	Sigma-Aldrich Chemie, Steinheim, DE
carboxypeptidase Y dissolved in 100 mM citrate, pH 5.5	Sigma-Aldrich Chemie, Steinheim, DE
Chymotrypsin (1 mg/ml) dissolved in Milli-Q water	Roche, Grenzach-Wyhlen, DE
cloned *Pfu* DNA Polymerase (10 U/µl)	Stratagene, La Jolla, CA, USA
Elastase	Sigma-Aldrich Chemie, Steinheim, DE
Endoproteinase Arg-C dissolved in 1 mM HCl	Sigma-Aldrich Chemie, Steinheim, DE
Endoproteinase Asp-N	Sigma-Aldrich Chemie, Steinheim, DE
Endoproteinase Glu-C	Sigma-Aldrich Chemie, Steinheim, DE
Endoproteinase Lys-C	Sigma-Aldrich Chemie, Steinheim, DE
Endoproteinase Pro-C, recombinant	Sigma-Aldrich Chemie, Steinheim, DE
Leucine-aminopeptidase dissolved in 100 mM phosphate, pH 7.0	Sigma-Aldrich Chemie, Steinheim, DE
Papain	Sigma-Aldrich Chemie, Steinheim, DE
Prolidase dissolved in 50 mM Tris-HCl, pH 8.0	Sigma-Aldrich Chemie, Steinheim, DE
Proteinase K (10 mg/ml)	Sigma-Aldrich Chemie, Steinheim, DE
Ribonuclease H (RNase H), *E. coli*, recombinant	Ambion, Austin, TX, US
Subtilisin dissolved in 50 mM Tris, pH 8.0	Sigma-Aldrich Chemie, Steinheim, DE
T4 DNA ligase (400 U/µl)	New England Biolabs, Frankfurt/Main, DE
Thermolysin	Sigma-Aldrich Chemie, Steinheim, DE
TEV-protease (1 mg/ml)	selfmade, recombinant
Trypsin (1 mg/ml) dissolved in trypsin buffer	Roche, Grenzach-Wyhlen, DE

2.1.7 Antibodies

The monoclonal antibody H20, binding to the $m_3^{2,2,7}$G cap has been described in (Bach *et al.*, 1990) and was coupled to NHS-activated sepharose for immunoaffinity chromatography. The monoclonal antibody directed against SF3a 120k is described in (Dybkov *et al.*, 2006). It was crosslinked to a Protein A sepharose column for immunoaffinity chromatography.

2.1.8 Chromatography media and columns

All regular chromatography media and columns are listed in table 2.7.

Table 2.7 Chromatography media and columns

Matrix	Supplier
Glutathione sepharose FF	Amersham Biosciences, Freiburg, DE
HiTrap columns 1 ml DEAE, SP, Q, CM, ANX, heparin, phenyl sepharose, butyl sepharose, octyl sepharose	Amersham Biosciences, Freiburg, DE
HiTrap heparin sepharose 5 ml	Amersham Biosciences, Freiburg, DE
HiPrep desalting 16/26	Amersham Biosciences, Freiburg, DE
NAP 5, NAP 10, NAP 25 columns	Amersham Biosciences, Freiburg, DE
NHS-activated sepharose FF	Amersham Biosciences, Freiburg, DE
PD10 columns	Amersham Biosciences, Freiburg, DE
Protein A sepharose CL 4B	Amersham Biosciences, Freiburg, DE
Superdex200 26/60	Amersham Biosciences, Freiburg, DE
Superdex200 10/300	Amersham Biosciences, Freiburg, DE
Superdex200 PC 3.2	Amersham Biosciences, Freiburg, DE
Superdex75 26/60	Amersham Biosciences, Freiburg, DE
Superdex75 10/300	Amersham Biosciences, Freiburg, DE
Superdex75 PC 3.2	Amersham Biosciences, Freiburg, DE
Superose 6 10/300	Amersham Biosciences, Freiburg, DE
Superose 6 XK 16/70	Amersham Biosciences, Freiburg, DE
Superose 6 PC 3.2	Amersham Biosciences, Freiburg, DE

2.1.9 Devices

All devices used in this work are listed in table 2.8.

Table 2.8 Devices

Device	Manufacturer
Äkta Explorer	Amersham Biosciences, Freiburg, DE
Äkta Purifier	Amersham Biosciences, Freiburg, DE
Äkta Prime	Amersham Biosciences, Freiburg, DE
Beamline ID23-1	ESRF, Grenoble, CH
Beamlines PX1 and PX2	Swiss Light Source, SLS, Villigen, CH
Beamline BW6	DESY, Hamburg, DE
Biofuge pico and Biofuge fresco	Heraeus, Hanau
Cartesian crystallization robot, 4 channels	Zinsser analytic, Frankfurt, DE
Cryofuge 6000i	Heraeus, Hanau, DE
FR591 rotating anode generator	Bruker ASX B.V., Delft, NL
Free mounting system	Rigaku Europe, Kent, UK
Gel documentation system	Bio-Rad, Munich, DE
Geldryer model 583	Bio-Rad, Munich, DE
Gelelectrophoresis equipment	Bio-Rad, Munich, DE
Glass ware	Schott, Mainz, DE, Merck, Darmstadt, DE
Gradient master	Biocomp Fredericton, CA
HT multitron culture shaker	Infors, Bottmingen, CH
Ice machine	Ziegra, Stockport, UK
Milli Q synthesis A10	Millipore, Bedford, MA, US
Magnetic stirrer, heatable	IKA, Staufen, DE
Mar345 image plate detector	Marresearch GmbH, Norderstedt, DE
Megafuge 1.0R, swing out rotor type 2704	Heraeus, Hanau, DE
Microliter syringes	Hamilton, Bonaduz, CH
12-way multichannel pipette	Eppendorf, Hamburg, DE
PCR cycler T-Professional	Biometra, Göttingen, DE
Pharmacia LKB Frac 200	Amersham Biosciences, Freiburg, DE
pH-meter seven easy	Mettler Toledo, Giessen, DE
Photometer DU 530	Beckmann, Krefeld, DE
Pipettes Pipetman P1000 ,P200, P20, P10, P2	Gilson, Middleton, WI, US
Power supplies	Bio-Rad, Munich, DE
Reflex, mass spectrometer	Bruker ASX B.V., Delft, NL
Rocking platform	Biometra, Göttingen, DE
Scales BP4100, BP410S	Sartorius, Göttingen, DE

Table 2.8 Devices continued . . .

Device	Manufacturer
SMART, HPLC	Amersham Biosciences, Freiburg, DE
Sonifier II 250 Digi	Heinemann, Schwäbisch-Gmünd, DE
Sorvall Discovery 90	Thermo fisher Scientific, Waltham, US
Sorvall Discovery M150	Thermo Fisher Scientific, Waltham, US
Sorvall Evolution RC	Thermo Fisher Scientific, Waltham, US
Sorvall SA 600 rotor	Thermo Fisher Scientific, Waltham, US
Sorvall SLC6000 rotor	Thermo Fisher Scientific, Waltham, US
Sorvall T-865 rotor	Thermo Fisher Scientific, Waltham, US
Speed vac concentrator 5301	Eppendorf, Hamburg, DE
Surespin 630 rotor	Thermo Fisher Scientific, Waltham, US
Spinocan needle	Braun, Melsungen, DE
S100AT rotor	Thermo Fisher Scientific, Waltham, US
Thermomixer comfort	Eppendorf, Hamburg, Germany
Ultrospec 3000 pro spectral photometer	Amersham Biosciences, Freiburg, DE
Vortex Genie 2	Scientific Industries, Bohemia, US
Varioklav 300 EPZ	H+P Labortechnik, Oberschleissheim, DE

2.1.10 Consumption items

The consumption items are listed in table 2.9.

Table 2.9 Consumption Items

Item	Supplier
Concentrators Vivaspin 1.5, 4, 15	Sartorius, Göttingen, DE
Crystallization plates, 96 well, sitting drop, round bottom	Greiner-Bio-One, Frickenhausen, DE
Cryschem plates, 24 well, sitting drop	Hampton research, Aliso Viejo, CA, US
Econo-Pac columns	Bio-Rad, Munich, DE
Eppendorf-cups, 1.5 ml, 2.5 ml	Sarstedt, Nürnbrecht, DE
Gloves, Dermatril	KCL, Fulda, DE
Gloves, Safeskin, PFE	Kimberly-Clark, Mainz, DE
Needles	Henke Sass Wolf, Tuttlingen, DE
PC tube ASSY	Hitachi, Osaka, JP
Polyprep columns	Bio-RAD, Munich, DE

Table 2.9 Consumption Items continued . . .

Item	Supplier
Polycarbonate flanged tube, 50 ml	Hitachi, Osaka, JP
Pipet tips	Sarstedt, Nürnbrecht, DE
Sequencing cups	Perkin-Elmer, Wellesley, MA, US
Sterile filters 0.22 μm, 0.45 μm	Sarstedt, Nürnbrecht, DE
Syringes	Braun, Melsungen, DE
Tubes, 50 ml, 15ml	Greiner-Bio-One, Frickenhausen, DE
Tube, PA, 36 ml	Beckmann, Krefeld, DE
Tube, PA, 11×66 ml	Thermo Fisher Scientific, Waltham, US
Whatman 3MM	Whatman plc, Maidstone, UK

2.1.11 Software

Table 2.10 lists the software used in this work.

Table 2.10 Software

Software	Reference
Collaborative Computational Project Number 4 (CCP4i) pro-gramme suite	(Potterton *et al.*, 2003)
Coot	(Emsley and Cowtan, 2004)
Gromacs	(Lindahl *et al.*, 2006)
HKL2000	(Minor *et al.*, 2006)
Mapman	(Kleywegt and Jones, 1996b)
Moleman(2)	(Kleywegt and Jones, 1996a)
Phenix suite	(Adams *et al.*, 2004)
Pymol	http://pymol.sourceforge.net/
Sharp/Autosharp	(Bricogne *et al.*, 2003)
Shelx suite	(Sheldrick, 2008)
Solve/Resolve	(Terwilliger, 2004)
XDS	(Kabsch, 1988)

2.2 Methods

2.2.1 Cells and cell cultures

2.2.1.1 *Escherichia coli*

The cultivation of *E. coli* cells was carried out in LB medium or on LB agar in the presence of antibiotics to maintain selective markers. Chemically competent cells were prepared for plasmid transformation according to (Chung and Miller, 1993). The cell strains HB 101 (Promega, Mannheim, Germany) and DH5α (Hanahan, 1983) were used for plasmid propagation. The strains BL21 (DE3) Rosetta and Rosetta 2 (Novagen, Madison, WI, USA) for protein production by T7 RNA polymerase induced expression. The culture density of *E. coli* cells was measured with an Ultrospec 3000 pro spectral photometer in a plastic cuvette at OD_{600} with the respective medium as a reference. Antibiotics were used in the following concentrations, ampicillin 100 μg/ml, chloramphenicol 20 μg/ml in 100 % ethanol, streptomycin 30 μg/ml and kanamycin 20 μg/ml.

2.2.1.2 HeLa cell culture

HeLa S3 cells (Computer cell culture centre, Belgium) were cultivated in S-MEM medium (Gibco) supplemented with 5 % (v/v) newborn calf serum, 50 μg/ml penicillin and 100 μg/ml streptomycin. For large scale cultivation a biobench (Applikon) was used for a continuous culture. The density of the culture was kept at $2.5 - 5 \times 10^6$ cells per ml. The HeLa culture was harvested according to Dignam *et al.* (1983).

2.2.2 Nucleic acid biochemistry

2.2.2.1 Urea polyacrylamide gel electrophoresis (Urea-PAGE)

RNAs were extracted from snRNPs by phenol-chloroform extraction according to (Sambrook *et al.*, 1989). The RNA pellets were dissolved in 20 μl RNA loading buffer and heated to 95 °C for 3 min. A typical gel was composed of $12 - 14$ % polyacrylamide and 8 M urea in 1 × TBE. The gel was polymerized by the addition of 1 % [v/v] APS and 1 % [v/v] TEMED. After loading the gel was run in 0.5 × TBE with a current of 1 mA/cm^2 for a gel of 1 mm thickness for $2 - 3$ h. The RNA was visualized by silver staining according to Merril and Goldman (1982) or by ethidium bromide staining.

2.2.2.2 DNA-guided digestion of RNA by RNase H

According to Günzl *et al.* (2002), RNase H can be used to specifically cleave RNA directed by DNA oligonucleotides complementary to the RNA sequence. This technique is not restricted to analytical probing of RNA but can also be used for the removal of the 5' splice site binding region of U1 snRNP (Seiwert and Steitz, 1993). For a quantitative reaction optimal cleavage conditions were determined to set up a large scale reaction of 10 ml volume. This reaction comprised 22 nmol U1 snRNP, 110 nmol DNA1-oligonucleotide and 500 U RNase H in the presence of 20 mM Tris-HCl, pH 7.9, 120 mM NaCl, 2 mM DTT and 3 mM $MgCl_2$. The reaction was incubated for 3 h at 30 °C.

2.2.3 Protein biochemistry

All protein solutions were kept at 4 °C and were frozen in liquid nitrogen for long term storage. Protein concentrations were determined according to Bradford (1976) with the respective buffer as a reference.

2.2.3.1 Preparation of HeLa nuclear extract

HeLa nuclear extract was prepared according to Dignam *et al.* (1983). Hela cells were cultivated in a 20 liter scale as described in 2.2.1.2. The following steps were performed at 4 °C. The cells were harvested with a Cryofuge 6000i for 10 min at 2000 rpm and washed three times with PBS in a Megafuge 1.0R spinning 10 min at 2000 rpm each wash. After washing, the cell pellet was resuspended in 2 cell pellet volumes Roeder A buffer and incubated for 5 min. The 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 PMSF. To disrupt the cells, the suspension was dounced with 15 strokes. The nuclei were pelleted by centrifugation in a SA-600 rotor for 10 min at 2700 rpm and were resuspended in 1.5 pellet volumes of Roeder C buffer, supplemented with 0.5 mM PMSF. The nuclei were opened by douncing with 15 strokes and slowly stirred with a magnetic stirrer for $20 - 30$ min. In the following step, the homogenate was overlaid with n-heptane to separate the aqueous phase from lipids. The nuclear extract was cleared by centrifugation with a T-865 rotor at 40 000 rpm for 30 min. Finally, the extract was harvested with a syringe and stored at −80 °C.

2.2.3.2 Protein expression and preparation of bacterial cell extract

E. coli cells were cultivated as described in 2.2.1.1 on page 39. From a single colony of an expression strain carrying an expression plasmid, a pre-culture of $50 - 100$ ml was inoculated. After overnight growth, the cells were spun down in a Megafuge 1.0R for 10 min at 4000 rpm and resuspended in $1 - 2$ ml of fresh medium. This suspension was used for the inoculation of the main culture. Protein expression was carried out either in autoinducing medium according to Studier (2005) or with IPTG induction and LB medium. In case of LB medium, expression of the target gene was induced by the addition of 1 mM IPTG at an OD_{600} of 0.4 - 0.6. Typically protein production was induced over night at 20 ℃. For autoinducing cultures, cells were grown until they reached the stationary phase. Cells were harvested in a SLC6000 rotor for 20 min at 4000 rpm or in a Megafuge 1.0R for 10 min at 4000 rpm depending on the culture size. The cell pellet was then resuspended in about 2 ml protein purification buffer per liter culture at OD_{600} of 1. For bacterial cell suspensions the buffer was supplemented with 1 tablet of Complete without EDTA per 50 ml of buffer. The cell suspension was frozen in liquid nitrogen for long term storage and to disrupt the cells. Bacterial cells were lysed by sonication with a Branson sonifier. Depending on the volume of the suspension an appropriate sonication program was used for one pass or several passes, in case the viscosity of the solution did not decrease. The suspension was cleared either for 1 h in a SA600 rotor at 12 000 rpm or in a tabletop centrifuge at 13 000 rpm for 30 min, depending on the total volume. 100 μl of cells at an OD_{600} 0.6 before and after induction 10 μl of supernatant and the corresponding amount of pellet were analyzed by SDS-PAGE (see 2.2.3.5, page 42).

2.2.3.3 Immunoaffinity purification of snRNPs

The H20 antibody, which is covalently crosslinked to NHS-activated sepharose, specifically binds the $m_3^{2,2,7}G$ cap of snRNPs. The bound material can be eluted with an excess of synthesized m_7G cap. HeLa nuclear extract (see 2.2.3.1) was diluted with glycerol-free Roeder C buffer supplemented with 0.5 mM PMSF to achieve a final glycerol concentration of 5 % for loading the column. Before loading, the extract was cleared with a T-865 rotor at 40 000 rpm for 30 min and passed through a filter. After loading, the column was washed with 20-30 column volumes of Roeder C buffer with 0.5 mM PMSF. The bound snRNPs were then eluted with two column volumes of Roeder C buffer supplemented with 0.5 mM of PMSF and 5 mM of m_7G cap in 2 ml fractions. The fractions were analyzed by SDS-PAGE (2.2.3.5) and stored at -80 ℃. Typically

one column run yielded $30-40$ mg of total snRNPs, where the peak fractions had a concentration of $1.5-2$ mg/ml. The H20 column was regenerated with 6 M urea.

2.2.3.4 Immunoaffinity purification of SF3a

An immunoaffinity column coupled to anti-SF3a-120 antibodies was equilibrated with 20 ml of buffer G_{600} supplemented with 5 % glycerol. About 200 ml flow through of an H20 column, which was run at 420 mM NaCl, were brought to 600 mM NaCl and passed over the anti-SF3a-120 column over night with a peristaltic pump. The column was washed with 15 ml of G_{600} buffer with 5 % glycerol. The bound material was eluted with 18 ml G_{600} buffer with 5 % glycerol, supplemented with the peptide for elution (Dybkov *et al.*, 2006). The column was regenerated with 20 ml 10 mM $NaPO_4$, pH 7.2, 10 ml 3.5 M $MgCl_2$ in 10 mM $NaPO_4$, pH 7.2 and 30 ml PBS, pH 8, supplemented with 30 µl 3 % sodium azide. In addition to $MgCl_2$ the column was regenerated with 4 M urea. The input, flowthrough (25 µl) and fractions (50 µl) were analyzed by SDS-PAGE.

2.2.3.5 SDS-PAGE

SDS-PAGE was essentially performed according to Laemmli (1970) and (Weber *et al.*, 1972). The acrylamide content of the gels was adjusted to the size of the proteins to be resolved. A typical gel contained a stacking part, cast of stacking-gel buffer and 16 % of Rotiphorese Gel 30 solution, polymerized with 0.3 % w/v APS and 0.03 % TEMED. Proteins were resolved on the lower gel, containing resolving-gel buffer and $25-50$ % of Rotiphorese Gel 30 solution, polymerized with 0.3 % APS and 0.3 % TEMED. Samples were heated to 95 °C for 2 min with SDS loading buffer and loaded on the gel, sided to a molecular weight marker (Biorad, unstained). Polyacrylamide gels were 1 mm thick and were run at 1 mA/cm^2 in SDS running buffer until the front of the loading dye reached the bottom of the gel. Gels were stained with Coomassie staining solution for 30 min and destained with destaining solutions I & II.

2.2.3.6 Glycerol gradient centrifugation

Glycerol gradient centrifugation is a gentle method to separate snRNPs (Kastner and Lührmann, 1999). For snRNP preparation, six preparative glycerol gradients of 36 ml each were run. The top phase comprised 15 ml glycerol gradient buffer with 10 % glycerol. The bottom phase contained 15 ml glycerol gradient buffer in 30 % glycerol

and was layered under the top phase with a syringe and a spinocam needle. All steps were carried out at 4 °C. The gradient was formed with a gradient master and then incubated for 1 h. A volume of 6 ml was removed from the top of each gradient and 6 ml replaced by snRNPs from H20 fractions (see 2.2.3.3) with a concentration of 1 mg/ml. The gradients were typically spun for 17 h at 27 000 rpm in a Surespin 630 rotor and harvested with a peristaltic pump and a fraction collector from the gradient bottom. During fractionation, UV absorbance at 280 nm was recorded. All fractions including the input were analyzed by SDS-PAGE (see 2.2.3.5).

Analytical gradients of 4.4 ml were run in a TH660 rotor with different parameters. These gradients were harvested manually from the top with a pipette and analyzed by Bradford assay and SDS-PAGE (see 2.2.3.5).

2.2.3.7 Ion exchange chromatography

Ion exchange chromatography was performed with HiTrap columns at analytical and preparative scale. For a standard run, the column was equilibrated with 10 column volumes of low salt buffer, typically containing 50 mM NaCl. For elution, a linear gradient of 10-15 column volumes with up to 1 M salt was applied to the column. Fractions of half a column volume were collected and analyzed by SDS-PAGE (see 2.2.3.5).

U1 snRNP and U2 snRNPs were entirely separated on a 5 ml HiTrap heparin sepharose column with an Äkta prime system. The best separation performance was achieved using an Äkta prime and applying a linear gradient of $60 - 70$ ml up to 800 mM NaCl. Fractions of 1.5 ml were taken and 40 µl of each fraction and the input were analyzed by SDS-PAGE (see 2.2.3.5).

2.2.3.8 Size exclusion chromatography and desalting

Size exclusion chromatography and desalting was performed with SMART, Äkta purifier and Äkta explorer or disposable desalting columns according to the manufacturer's instructions. Columns were pre-equilibrated with 1.5 column volumes of running buffer in case of gel filtration. Fractions typically had $1 - 2$ % of the total column volume. At maximum $1 - 3$ % of sample in relation to the total column volume were loaded on gel filtration columns.

In case of analytical SMART runs with PC 3.2 gel filtration columns, 50 µl of sample were injected and run at a flow rate of 40 µl/min. Fractions of 40 µl were taken for gel analysis by SDS-PAGE (see 2.2.3.5).

For buffer exchange of U1 snRNP, 15 ml of protein sample were applied on a HiPrep 26/16 desalting column, equilibrated with crystallization buffer. The flow rate was set to 3 ml/min. The protein peak was fractionated in one fraction according to the UV absorbance at 280 nm or until the conductivity started to increase due to the eluting salt.

For native SF3a preparation a 120 ml Superose 6 XK 16/70 column was used. For the run, a flow rate of 1 ml/min was set. 4 ml of protein solution were loaded on the column and fractions of 2 ml were taken. 20 µl of each fraction as well as of the input were analyzed by SDS-PAGE (see 2.2.3.5).

2.2.3.9 Concentration of protein solutions

The concentration of protein solutions is a critical prerequisite for the setup of crystallization screens. As a standard procedure, concentrators from Sartorius and Millipore were used, which contain an ultrafiltration membrane with a specific molecular cutoff. The sample was passed through the membrane by centrifugal force, retaining larger components like proteins or particles. Typically, a cutoff lower than half of the protein size was chosen.

12 S U1 snRNP, which was sensitive to all membrane types tested, was concentrated by pelleting in a mini ultracentrifuge (Sorvall Discovery M150) using thick-walled polycarbonate tubes (PC tube ASSY). For one pelleting run, 1.6 ml of U1 snRNP were sedimented for 3 h at 55 000 rpm with maximum acceleration and deceleration set to 1. The supernatant was taken off and roughly 90 % of the sedimented U1 snRNPs were recovered carefully as a viscous pellet. This procedure was repeated until a protein concentration of 20 mg was reached. The pellet was carefully resuspended, transferred to a 1.5 ml tubes and spun down for one minute in a tabletop centrifuge at 4 ℃. The supernatant was carefully transferred to a fresh tube and directly used for crystallization or frozen in liquid nitrogen and stored at −80 ℃.

U4/U6·U5 tri-snRNP was concentrated by ammonium sulfate precipitation followed by microdialysis. Ammonium sulfate powder was added to 25 S fractions of glycerol gradients which mainly contained U4/U6·U5 tri-snRNP to a concentration of 2 M. The increase of volume due to the addition of ammonium sulfate was neglected. The solution was then incubated on a roller at 4 ℃ for 30 min and spun down in a Megafuge 1.0R at 4000 rpm for 1 h. After the entire removal of the supernatant, the highly viscous whitish ammonium sulfate protein pellet was recovered with a pipette. The pellet was carefully transferred on a millipore dialysis membrane, which was pre-equilibrated with

crystallization buffer AS. After the first step of dialysis, the solution was transferred to a fresh dialysis membrane, equilibrated with crystallization buffer AS. During the dialysis, the protein solution cleared up. Both dialysis steps were carried out for 1 h with 0.5 l of buffer. As controls to monitor the buffer exchange, a smaller amount of tri-snRNP was dialyzed against either crystallization buffer AS, lacking ammonium sulfate, or crystallization buffer AS with 50 mM sodium chloride only. In both cases the solution precipitated after initial clearing up. On a preparative scale a final protein concentration of $7 - 10$ mg/ml was achieved.

2.2.3.10 Limited proteolysis

A prerequisite for proteins and protein-RNA complexes to crystallize is the presence of a folded structure. Flexible domains, linkers and loops as well as the entire absence of a fold hinder crystallization. Limited proteolysis is a powerful method to characterize protein fragments that have the potential to crystallize, since proteolytic activity primarily targets accessible protein regions, unfavorable for crystallization. Limited proteolysis was employed in early stages of protein crystallography when techniques like molecular cloning were not available (Gast *et al.*, 1977; Kabsch *et al.*, 1977).

For limited proteolysis experiments, several proteases were employed (see table 2.6 on page 34). The amount of protease used depended on its activity and purity. Typically an experiment was either carried out with different protease concentrations and one fixed time point or a series of time points with a fixed protease concentration at 4 °C. The reactions were stopped by the addition of SDS loading dye. The samples were analyzed by SDS-PAGE (see 2.2.3.5).

2.2.3.11 Purification of a chymotrypsin treated U1 snRNP for crystallization

5 mg of U1 snRNP at a concentration of 0.79 mg/ml were mixed with a 1:2 molar ratio with DNA4 oligonucleotide and 50 µg of chymotrypsin. The reaction was incubated for 14 h at 4 °C and directly loaded on a 26/60 Superdex 200 column in crystallization buffer, containing 500 mM NaCl. The column was run at 2 ml/min and 5 ml fractions were taken. Gel filtration resulted in a single peak, comprising three fractions. The fractions were concentrated to a volume of 2 ml with a vivaspin concentrator (MWCO 30000) and diluted to 10 ml with crystallization buffer, containing no salt. The truncated U1 snRNP did not precipitate during concentration unlike the full length particle. After

ultrafiltration, around 120 μl of truncated U1 snRNP with a concentration of 16 mg/ml were obtained.

2.2.3.12 Mass spectrometry

Mass spectrometry of protein bands from Coomassie-stained gels was carried out by U. Plessmann and M. Raabe in the group of Dr. H. Urlaub according to **Shevchenko** *et al.* (1996). Matrix assisted laser desorption/ionisation (MALDI) time of flight (TOF) analyses were carried out with a Bruker Reflex mass spectrometer, data were evaluated with the program Mascot. Proteins were identified by the size of their tryptic fragments. An alignment of the identified fragments with the full length protein sequence allowed conclusions about the localization of protein fragments with respect to the full length protein.

2.2.4 Crystallography and computational methods

2.2.4.1 *In situ* proteolysis

Apart from proteolytic analysis of proteins in solution, it is possible to include proteases in crystallization droplets. This approach, termed *in situ* proteolysis, has greatly enhanced the reproducibility of U1 snRNP crystals. Several proteases, mainly chymotrypsin, were included in the crystallization setup by pre-mixing the protease with protein solution in mass ratios ranging from 1:10000 to 1:100. Furthermore the dilution of the protease had a direct impact on the time that was required for the crystals to appear. The resulting crystals were extensively washed with reservoir solution and analyzed by SDS-PAGE (see 2.2.3.5). An analogous approach for *in situ* proteolysis has been recently described by Dong *et al.* (2007).

2.2.4.2 Crystallography setup

Protein and snRNP samples with a concentration of $5 - 20$ mg/ml were subjected to an initial round of crystallization trials. For screening, a Cartesian crystallization robot was employed, which reliably sets 100 nl plus 100 nl sitting drops in a 96-well format. Crystal growth is then eventually promoted upon equilibration of the crystallization drop with 100 μl of reservoir solution. The nanoliter-setup greatly reduces the amount of protein spent, but still allows crystal harvesting in many cases. The crystallization screens used in the first place are listed in table 2.5 on page 33.

First hits were refined in a 24-well linbro plate with a drop size of 1 µl plus 1 µl and a reservoir of 400 µl. Typically the pH and precipitant concentration were varied around the original condition. If not stated otherwise U1 snRNP was used for crystallization in crystallization buffer at 20 mg/ml.

In case of no growth of crystals or no improvement of the initial hit, an additive screen was performed with the Cartesian crystallization robot. For this purpose, 100 nl of protein solution and 100 nl of the original screening solution were set up in the presence of 20 nl of one of 96 different additives and a fixed reservoir of $50 - 100$ µl. For the screening of additives, a customized screen as well as a commercial hampton additive screen in a 96-well format were used. The crystallization screens that were carried out during this work are listed in table 2.11.

Table 2.11 Crystallization screens in 96-well format. All proteases were used in an 1:2000 dilution by mass, relative to the crystallization target, if not stated otherwise under 'remarks'. All oligonucleotides were used in an 1:2 molar excess. All screens were performed at 4 ℃, unless stated otherwise under 'remarks'

Screen No.	Name/Oligo	Protease	Screens	remarks
1	SF3a	none	Index, MBClass I, Classics Lite, Nucleix, Pro-Complex, JCSG	
2	tri-snRNP	none	Index, Classics, Classics Lite, MBClass I + II, Nucleix, JCSG, SM II, PEG/ION, AmSO4	
3	tri-snRNP	Asp-N	Index, Nucleix,	
4	tri-snRNP	Glu-C	Index, Nucleix,	
5	tri-snRNP	Chymotrypsin	Index, Nucleix,	
6	RNA1	Chymotrypsin	Index, Classics Lite, MBClass I + II, Salt RX, Nucleix, SM I, SM III	
7	RNA4 -C	none	Classics, Classics Lite, MBClass I + II, Nucleix, SM I, AmSO4, Anions, Cations, pH-Clear I + II	
8	RNA4 +C	Chymotrypsin	Classics, Classics Lite, MBClass I + II, Nucleix, SM I, AmSO4, Anions, Cations, pH-Clear I + II	1:1000 Chymotrypsin
9	RNA2	Chymotrypsin	Classics, Classics Lite, MBClass I + II, SM I	
10	RNA3	Chymotrypsin	Index, Classics, MBClass I + II, Nucleix, SM I, Anions, Cations	
11	U1-RNase H	Chymotrypsin	Index, MBClass I + II, Classics lite, Nucleix, SM I, SM III, Salt RX	
12	RNA5 -C	none	Classics, MBClass I + II, Nucleix, SM I, AmSO4, Anions, Cations, pHClear I + II, PEGs	

Table 2.11 Crystallization screens continued. . .

Screen No.	Name/Oligo	Protease	Screens	remarks
13	RNA5 +C	Chymotrypsin	Classics, Classics Lite, MBClass I + II, Nucleix, SM I, AmSO4, Anions, Cations, pH-Clear I + II, PEGs	1:1000 Chymotrypsin
14	RNA5 +C RT	Chymotrypsin	Index, Classics, Classics Lite, MBClass I, MbClass II, Nucleix, SM I, Anions, Cations	1:25000 Chymotrypsin at room temperature
15	RNA6	Chymotrypsin	Salt RX, Classics Lite, MBClass I, SM I, SM III, Nucleix	
16	RNA7	Chymotrypsin	Index, Salt RX, Classics Lite, MBClass I, MbClass II, SM I, SM III, Nucleix	
17	RNA8 +C RT	Chymotrypsin	MBClass I, MBClass II, Classics Lite, Nucleix, Anions	1:25000 Chymotrypsin at room temperature
18	RNA8	Chymotrypsin	Index, Classics, MBClass I, MBClass II, Classics Lite, SM I, Nucleix, Anions, Cations	
19	DNA4	Chymotrypsin	Index, Classics, MBClass I, MBClass II, Classics Lite, Salt RX, SM I	
20	DNA2	Chymotrypsin	Index, Classics, MBClass I, MBClass II, Classics Lite, Salt RX, SM I	
21	DNA3	Chymotrypsin	Index, Classics, MBClass I, MBClass II, Classics Lite, Salt RX, SM I	
22	DNA4	Asp-N	Index, MBClass I, MBClass II, Classics Lite, ProComplex, JCSG, Nucleix	1:200 Asp-N

Table 2.11 Crystallization screens continued...

Screen No.	Name/Oligo	Protease	Screens	remarks
23	DNA4	Thermolysin	Index, MBClass I, MBClass II, Classics Lite, ProComplex, JCSG, Nucleix	1:1000 Thermolysin
24	DNA4	Trypsin	Index, MBClass I, MBClass II, Classics Lite, ProComplex, JCSG, Nucleix	1:10000 Trypsin
25	DNA4	Glu-C	Index, MBClass I, MBClass II, Classics Lite, ProComplex, JCSG, Nucleix	1:200 Glu-C
26	DNA4	Lys-C	Index, MBClass I, MBClass II, Classics Lite, ProComplex, JCSG, Nucleix	1:1000 Lys-C
27	DNA4	Elastase	Index, MBClass I, MBClass II, Classics Lite, ProComplex, JCSG, Nucleix	1:1000 Elastase
28	RNA9	Chymotrypsin	MBClass I, MBClass II, Salt RX, Classics Lite, SM I, SM III, Nucleix	
29	U1 SPN1	none	Index, Wizard, Salt RX	
30	U1 SPN1	Chymotrypsin	Index, Classics, Classics Lite, Wizard, Salt RX, SM I	
31	RNA10	Chymotrypsin	Index, Classics, MBClass I, MBClass II, Classics Lite, SM I, Cations, Anions	

2.2.4.3 Assessing crystal quality by X-ray diffraction

The quality of protein crystals is typically assessed by exposing them to X-ray radiation. Radiation damage can be minimized by cooling the crystals to 100 K during the measurement (Owen *et al.*, 2006). Owing to the high solvent content of most protein crystals, the formation of ice during freezing has to be avoided by the addition of cryoprotectants like glycerol, sugars, propylene glycol, 2,3-butanediol or oils. Only few crystallization conditions provide cryoprotection. In case of U1 snRNP crystals, the addition of $15 - 20$ % propylene glycol has proven to preserve diffraction and guarantee cryoprotection. Only low resolution diffraction was observed when U1 snRNP crystals were mounted in a capillary or via the free mounting system (FMS) at room temperature. U1 snRNP crystals were tested at 100 K on a Bruker Nonius FR591 home source as well as on several synchrotron beamlines, listed in table 2.8 on page 36. Suitable crystals were used for the collection of native, SAD and MAD datasets at synchrotron beamlines.

2.2.4.4 Heavy atom derivatization

Heavy atom derivatives, listed in tables 2.5 and 2.2, were dissolved in U1 derivatization buffer. The concentrations of the heavy atom compounds ranged from $2 - 10$ mM. $U1_{DNA4}$ co-crystals in U1 crystallization buffer were grown in a solution of 100 mM tri-sodium citrate, pH 5.6, 100 mM NaCl, $8 - 10$ % PEG 4000, 2 mM EDTA, pH 8.0 and 1 % Anapoe-X-305. The plate-like crystals were transferred to a sitting drop of 10 µl U1 derivatization buffer, which was pre-equilibrated with 400 µl of the respective reservoir solution. Crystals were then incubated for 24 h. Afterwards the crystals were transferred to a new pre-equilibrated drop of U1 derivatization buffer, containing $0.05 - 2$ mM of the heavy atom compound. The incubation time of derivatization varied between 10 min and seven days. It has to be noted that NaCl was replaced by $((NH_4)_2)(SO_4)$ of equivalent concentration, when mercury compounds were used. For iridium and osmium hexammines, the original solution from the crystallization condition containing EDTA and citrate, was used. Finally, the crystals were cryoprotected in a new drop with the same solution used for derivatization, but supplemented with $15 - 17$ % propylene glycol. Derivatized crystals were then flash-frozen in liquid nitrogen. Before datasets of these crystals were taken, the presence of the heavy atom compound as well as the peak, inflection point and remote wavelengths were determined by a fluorescence scan. Furthermore, the anomalous scattering factors f' and f'' were obtained.

2.2.4.5 Data processing and phasing

Datasets were processed with different software packages, depending on the beamline and detector used (see table 2.10 on page 38). In case of the Mar225 and Mar165 CCD detectors, data were processed and scaled with the HKL2000 suite and XDS. Friedel's law was considered false in the case of anomalous data. For the Pilatus pixel detector, data integration was only possible with XDS and specific correction files (Broennimann *et al.*, 2006). The Pilatus detector and accompanying processing software are still in its beta-phase, thus initial XDS versions produced artifacts which dramatically reduced data quality. Since XDS version 12/2007, reliability in data integration seems greatly enhanced.

Various applications like data reduction, molecular replacement, refinement, experimental phasing, map and mask manipulations, reflection data and coordinate manipulations were performed with the CCP4i software suite. Furthermore, pdb-files and electron density maps were manipulated with the programs Moleman and Mapman, respectively. Refmac5 and the Phenix software suite were used for refinement of molecular replacement solutions. For heavy atom searches, substructure solution and experimental phasing approaches, the Shelx suite, hkl2map, Sharp, Phenix and Solve/Resolve were employed. For the visual inspection of electron density maps and pdb coordinate files, the programs Coot and Pymol were used. All figures displaying coordinate files and electron density maps were generated with Pymol.

3 Results

3.1 U1 snRNP

3.1.1 Purification of U1 snRNP

The protocol for purification and concentration of U1 snRNP was constantly improved during this work. The yield of purified U1 snRNP was thereby increased by a factor of 3-4. Hela cell cultivation and the purification of total snRNPs by immunoaffinity chromatography has been described before (Kastner and Lührmann, 1999) and is also found in the methods section (2.2.3.1 and 2.2.3.3 respectively).

3.1.1.1 Glycerol gradient centrifugation of total snRNPs

Total snRNPs derived from Hela cells (2.2.3.1, 2.2.3.3) were separated on a $10 - 30$ % glycerol gradient in a preparative scale (2.2.3.6). Typically, six gradients of 36 ml each, were run in parallel for the separation of about 36 mg of total snRNPs. In figure 3.1, the RNA content of each gradient fraction, extracted by phenol chloroform extraction (PCI), was visualized by silver staining on a denaturing polyacrylamide gel (2.2.2.1). At 12 S, U1 snRNP and U2 snRNP co-migrated corresponding to the fractions number 13 to 18, clearly separated from the residual snRNPs. In fractions 5, 6 and 7, 25 S tri-snRNP was present, whereas fractions number 8, 9 and 10 contained exclusively U5 snRNP.

The proteins of the gradient fractions were isolated by PCI and applied on a 13.5 % SDS gel, shown in figure 3.2. In the gradient, proteins co-migrated with the U snRNAs to which they are binding to. This way all particles were clearly identified by their protein composition. Free U5 snRNP was distinguished from tri-snRNP by the 52k protein which is specific for 20 S U5. The 52k protein was most abundant in fractions 7 to 10, indicating an overlap of tri-snRNP and U5 snRNP peaks in fraction 7, in some gradients extending to fractions 6 and 8. The U1-70k protein seemed slightly degraded in this run. Taken together, the gradient fractions 4-6 could be regarded as tri-snRNP,
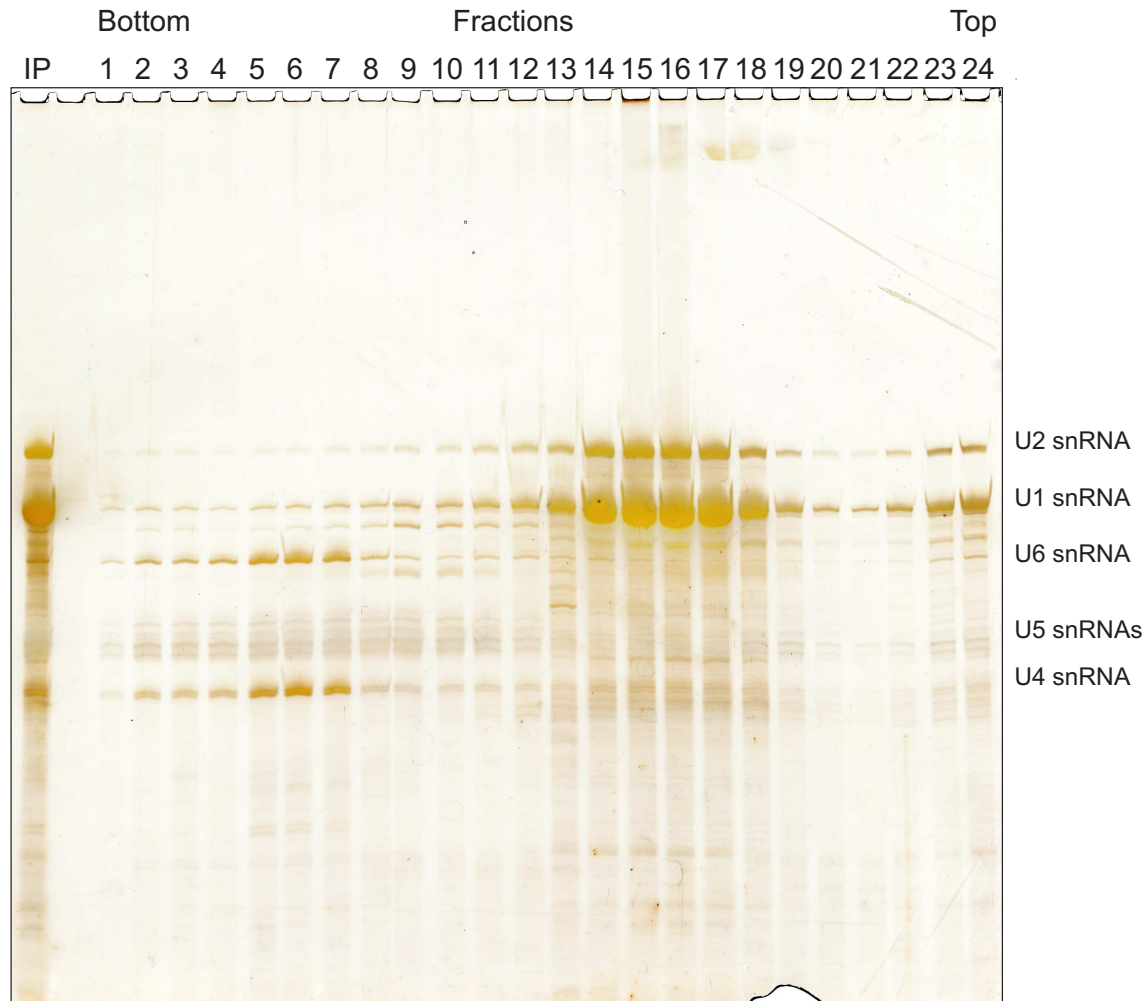
Figure 3.1 Separation of U snRNPs and their corresponding RNAs by glycerol gradient centrifugation. A $10 - 30$ % preparative glycerol gradient of 36 ml, containing 6 mg total snRNPs, was run for 17 h at $27\,000$ rpm in glycerol gradient buffer. The RNA of 50 µl of each 2 ml fraction was extracted by PCI and resolved on a 12 % polyacrylamide gel containing 8 M urea. The RNA species were visualized by silver staining. IP is 50 µl of the input material, fractions 1 to 23 are numbered from gradient bottom to top. The identities of the RNA species are denoted on the right side. The image was generously provided by Olexandr Dybkov.

fractions 8 to 10 contained U5 snRNP and fractions 13 to 18 a mixture of U1 and U2 snRNPs. Typical yields were $4 - 5$ mg for tri-snRNP, $5 - 6$ mg for U5 snRNP and $15 - 20$ mg of U1/U2 snRNPs from 36 mg input material.

3.1.1.2 Separation of U1 and U2 snRNPs by ion exchange chromatography

Initially U1 and U2 snRNPs were separated by a DEAE or MonoQ column. These protocols only allowed for a limited separation of the particles, yielding a minor fraction
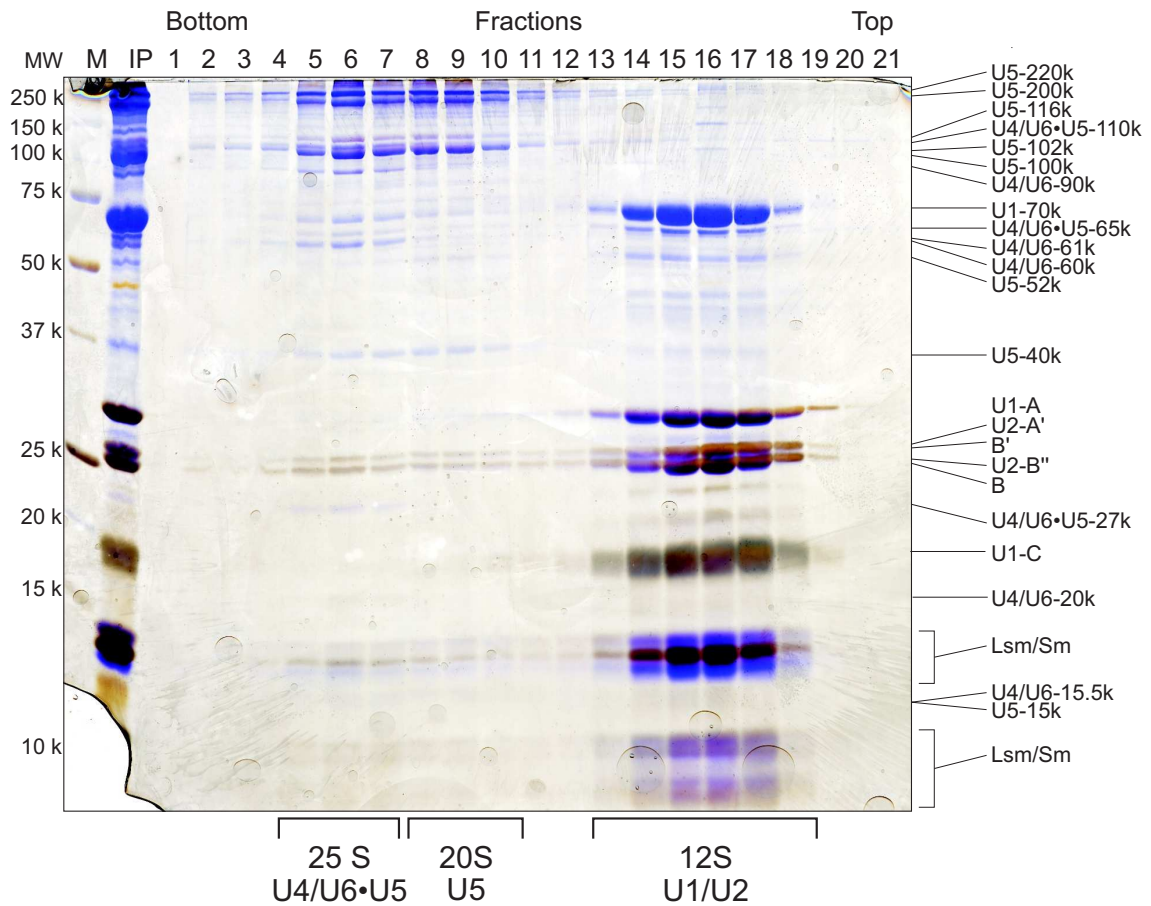
Figure 3.2 Separation of U snRNPs and their corresponding proteins by glycerol gradient centrifugation. A $10 - 30\,\%$ preparative glycerol gradient of 36 ml, containing 6 mg total snRNPs, was run for 17 h at $27\,000$ rpm in glycerol gradient buffer. The proteins of 50 μl of each 2 ml fraction were extracted by PCI and resolved on a 13.5 % polyacrylamide gel. The proteins were visualized by Coomassie staining. The molecular weight (MW) of the marker (M) proteins is given on the left side of the gel. IP is 50 μl of the input material, fractions 1 to 21 are numbered from gradient bottom to top. The identities of the proteins are denoted on the right side of the gel. The identities of the U snRNP complexes and their sedimentation coefficients in S are marked below the gel.

of U1 snRNPs and a second mixed fraction of U1 and U2 snRNPs. To improve this limiting step of the protocol, the separation capacities of other ion exchange resins were explored by individual FPLC runs (not shown). The most suitable ion exchange resin for the separation of U1 and U2 snRNPs was heparin sepharose, which entirely separated the particles.

Figure 3.3 shows the separation of U1 and U2 snRNPs by ion exchange chromatography with heparin sepharose (2.2.3.7). For one FPLC run, around $10 - 15$ mg U1 and U2 snRNPs from six glycerol gradients were diluted to a salt concentration of 100 mM
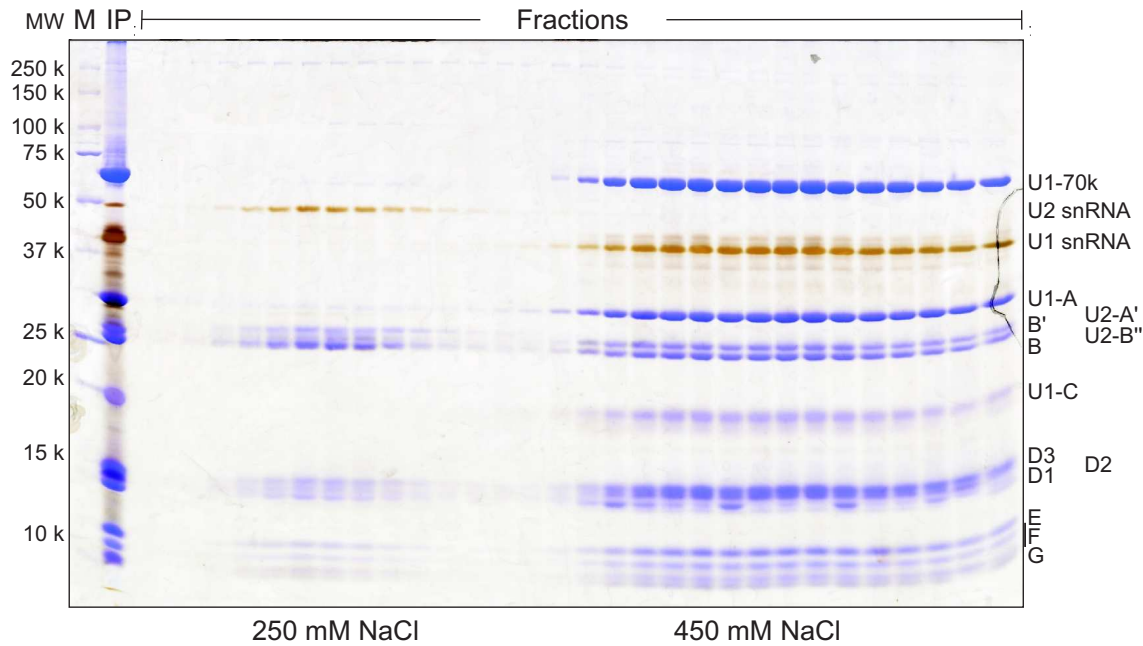
Figure 3.3 Separation of 12 S U1 and U2 snRNPs by ion exchange chromatography on heparin sepharose. 12 S fractions derived from glycerol gradients were loaded on a 5 ml heparin sepharose column at 100 mM NaCl. A linear gradient of 12 column volumes up to 800 mM NaCl was applied to the column and fractions of 1.5 ml were collected. 40 µl of the undiluted input (IP) and fractions containing protein were resolved on a 13.5 % denaturing polyacrylamide gel and visualized by Coomassie stain for proteins and silver stain for RNAs. The identities of proteins and RNA are given on the right side of the gel. The molecular weight (MW) of the marker (M) proteins is given on the left side of the gel. The salt concentrations 250 mM and 450 mM, where U2 and U1 snRNPs elute, respectively, are denoted below the gel.

NaCl and loaded on a 5 ml HiTrap heparin column. 40 µl of each fraction and the undiluted input were applied on a 13.5 % denaturing polyacrylamide gel. Proteins were visualized by Coomassie, RNA was stained with silver nitrate.

U1 and U2 snRNPs were nicely separated and eluted at different salt concentrations. All core snRNP and particle-specific proteins as well as the RNAs were present. The introduction of a heparin column, replacing DEAE, increased the yield of U1 snRNP from about $2-3$ mg to $8-10$ mg per FPLC run. Furthermore, a limited amount of pure U2 snRNPs became available for crystallization trials.

3.1.1.3 Buffer exchange and concentration of U1 snRNP

Due to the increased amounts of U1 snRNP, which were available, a 16/60 desalting column was used for buffer exchange instead of disposable PD10 desaltig columns (2.2.3.8). The U1 snRNP peak eluting from the heparin column was buffer exchanged

to crystallization buffer in one or two runs, if necessary. The desalted protein was manually fractionated according to the UV absorbance readout of the FPLC system. The buffer-exchanged particles were then pelleted in a mini-ultracentrifuge and fractionated from top to bottom with decreasing fraction size (see 2.2.3.9). Usually, two rounds of pelleting were carried out, where the lower fractions of the first run were subjected to a second pelleting step. Typically around 300 μl of U1 snRNP with a concentration of 20 mg/ml - in total 6 mg - were obtained after the second pelleting step.

3.1.2 Limited proteolysis of U1 snRNP in the absence and presence of oligonucleotides

3.1.2.1 Limited proteolysis of U1 snRNP

Several U1 snNRP proteins are known to contain domains of low complexity, that might be unstructured and unfavorable for crystallization. U1-70k comprises an RS-domain that is assumed to be flexible according to its amino acid composition. The first RRM of U1-A tightly binds to U1 snRNA, the second RRM of the protein is connected to the first with a flexible linker. The second RRM of U1-A does not contact U1 snRNA (Lu and Hall, 1995, 1997). In addition, the N or C-termini of some of the Sm proteins, like for example B/B', comprise low complexity regions that are presumably unstructured. Initial U1 snRNP crystals appeared irreproducibly after 4 weeks. When these crystals were dissolved and analyzed by SDS-PAGE, several proteins were truncated. This indicated that a removal of the flexible domains of the particles was favorable for crystallization of U1 snRNP.

Proteases were used to treat U1 snRNP and prepare truncated particles for crystallization. Judging from the pattern of the in-solution proteolysis, particles were either digested in a preparative scale in solution or suitable proteases were selected for *in situ* proteolysis experiments (see 2.2.4.1). In an *in situ* proteolysis experiment proteases were added to the crystallization setup.

U1 snRNP was digested for 72 h at 4 °C with ten-fold dilution series of trypsin, chymotrypsin, Arg-C, Pro-C and Asp-N. The reaction was stopped with SDS-loading dye and was visualized on a SDS-gel as shown in figure 3.4. Some of the bands resulting from partial proteolysis were submitted to mass spectrometry. The protein bands could not be identified due to technical problems.

Digestion with trypsin produced a homogeneous pattern with the largest fragment at around 25 kDa, (figure 3.4, lanes 7,8). The proteins U1-70k, U1-A, U1-C, D1, D2, D3
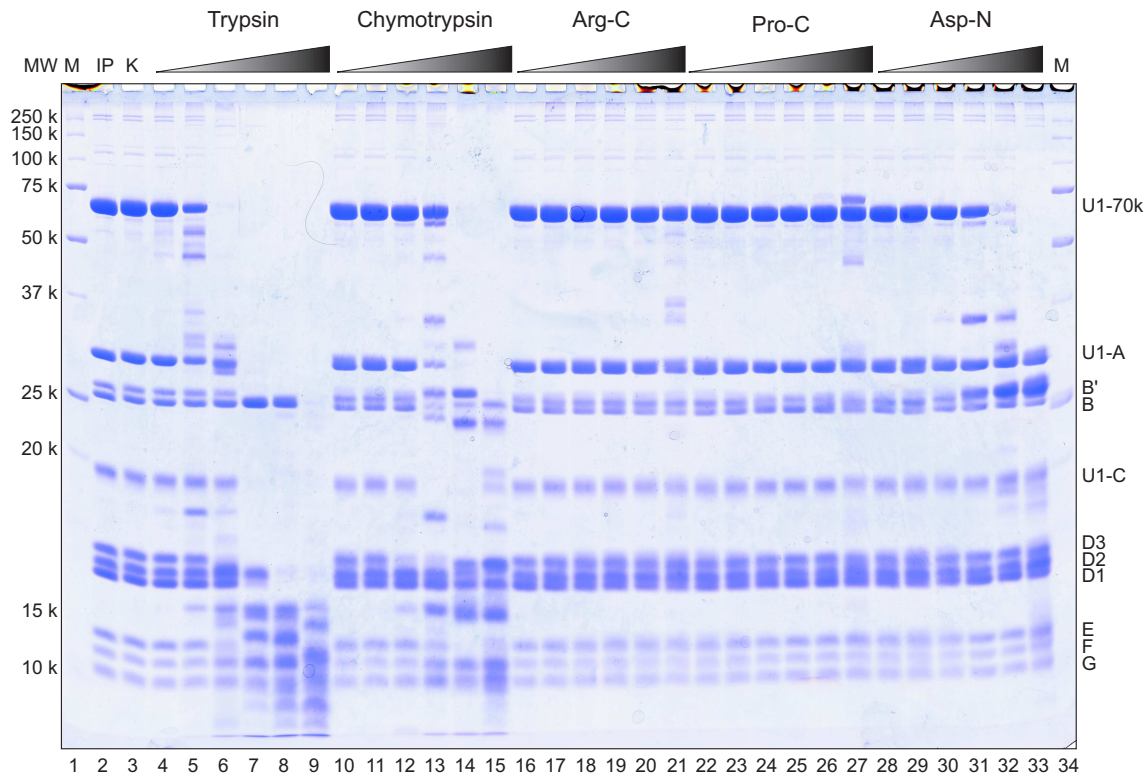
Figure 3.4 Limited proteolysis of U1 snRNP with trypsin, chymotrypsin, Arg-C, Pro-C and Asp-N. 20 µg of U1 snRNP were treated with 1:10 dilutions of proteases for 72 h at 4 ℃. The reactions were stopped by adding SDS-loading dye and applied on a 13.5 % polyacrylamide gel. The proteins were visualized by Coomassie staining. Lane 1 is the molecular weight marker, lane 2 is the U1 snRNP input, lane 3 is the control which was incubated in the absence of proteases. Lanes 4-9 are a series of ten-fold dilutions of digestions from 5 pg to 0.5 µg of trypsin, lanes 10-15 are a series of ten-fold dilutions of digestions from 5 pg to 0.5 µg of chymotrypsin, lanes 16-21 are a series of ten-fold dilutions of digestions from 2 pg to 0.2 µg of Arg-C, lanes 22-27 are a series of ten-fold dilutions of digestions from 0.000 002 U to 0.2 U of Pro-C and lanes 28-33 are a series of ten-fold dilutions of digestions from 1 pg to 0.1 µg of Asp-N. The molecular weight (MW) of the marker proteins is given on the left side of the gel. The identities of the full-length proteins are denoted on the right side of the gel.

and B/B' were affected by proteolysis, since the bands corresponding to the respective full-length proteins disappeared. The total mass of the remaining fragments in lane 8, estimated by their migration behavior, comprised around 100 kDa, which was around half of the total protein mass of U1 snRNP. The largest fragment was entirely digested by 0.5 µg of trypsin (figure 3.4, lane 9), reducing the particle's total protein mass to around 90 kDa. Chymotrypsin digestion produced a similar pattern like trypsin, but with two larger stable fragments at a molecular weight of 26 kDa and 24 kDa. Comparing lanes 12 and 13, B/B' can be excluded from being one of the larger fragments, since it is smaller than the 26 kDa band and degraded before the appearance of the 24k fragment
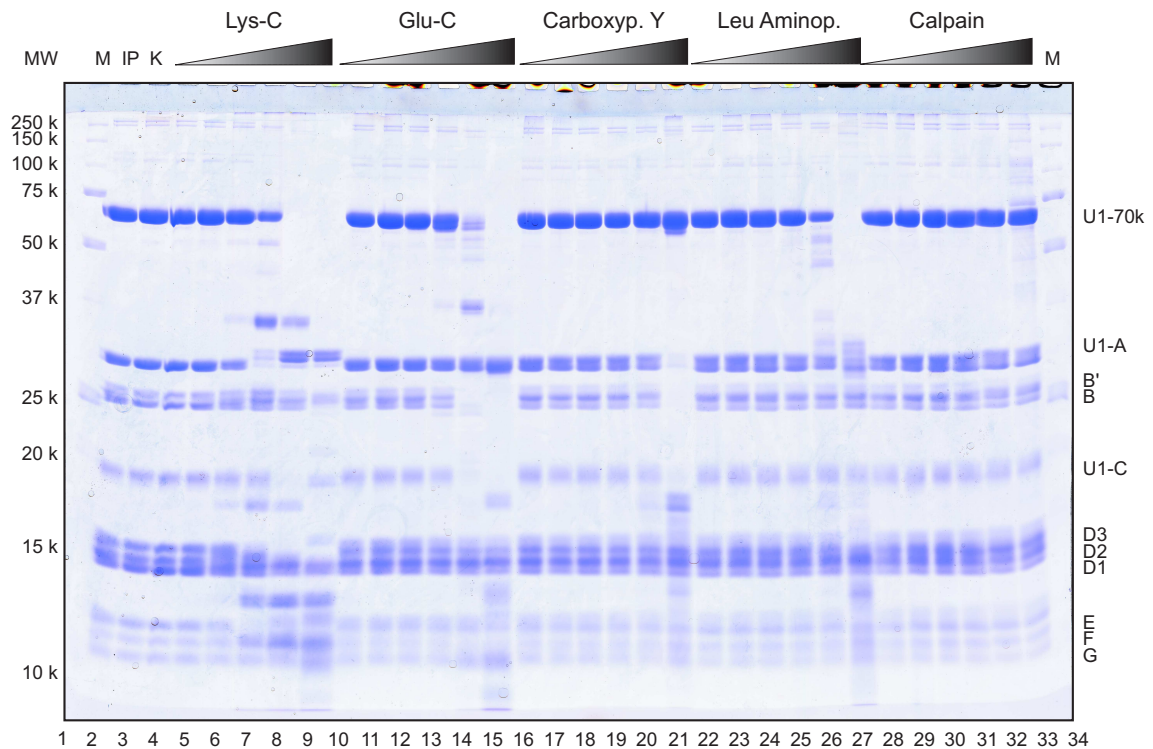
Figure 3.5 Limited proteolysis of U1 snRNP with Lys-C, Glu-C, carboxypeptidase Y, leucine-aminopeptidase and Calpain. 20 µg of U1 snRNP were treated with 1:10 dilutions of proteases for 72 h at 4 °C. The reactions were stopped with SDS loading dye and applied on a 13.5 % polyacrylamide gel. The proteins were visualized by Coomassie staining. Lane 1 is the molecular weight marker, lane 2 is the U1 snRNP input, lane 3 is the control which was incubated in the absence of proteases. Lanes 4-9 are a series of ten-fold dilutions of digestions from 2 pg to 0.2 µg of Lys-C, lanes 10-15 are a series of ten-fold dilutions of digestions from 5 pg to 0.5 µg of Glu-C, lanes 16-21 are a series of ten-fold dilutions of digestions from 5 pg to 0.5 µg of carboxypeptidase Y, lanes 22-27 are a series of ten-fold dilutions of digestions from 10 pg to 1 µg of leucine-aminopeptidase and lanes 28-33 are a series of ten-fold dilutions of digestions from 5 pg to 0.5 µg of Calpain. The molecular weight (MW) of the marker proteins is given on the left side of the gel. The identities of the full-length proteins are denoted on the right side of the gel.

(lane 14). Mass spectrometry analysis identified the band appearing at 15 kDa in lane 13 to be either B/B' or U1-A. Therefore the 24 kDa, and 26 kDa, fragments in lanes 13-15 can be attributed to U1-A or U1-70k.

The proteases Arg-C and Pro-C did not affect the integrity of any of the U1 snRNP proteins, only Pro-C slightly degraded U1-70k (lane 27). Proteolysis with Asp-N almost exclusively degraded U1-70k (lane 33), resulting in a fragment of about 26 kDa, which resembles the largest band from proteolysis with chymotrypsin (lane 14).

20 µg of U1 snRNP were digested for 72 h at 4 °C with a series of ten-fold dilutions of

Lys-C, Glu-C, carboxypeptidase Y, leucine-aminopeptidase and Calpain. The proteolysis experiment was visualized on a SDS-gel as shown in figure 3.5.

Lys-C affected U1-70k, U1-A, U1-C, B/B', D2, D3 and E. Concluding from lanes 7-9, the presence of $2 - 200$ ng Lys-C truncated U1-70k to a size of about 32 kDa, apparently forming a stable fragment. U1-A was truncated to 18 kDa, since the full-length protein disappeared at a Lys-C concentration of 5 ng. The size of two fragments at 13 kDa, and 18 kDa, which appeared concomitantly, together added up to the size of full-length U1-A.

According to lanes 14 and 15 in figure 3.5, Glu-C digestion affected U1-70k, B/B' and U1-C. The pattern in lane 14 indicated that U1-C and B/B' were truncated to a size below 15 kDa. U1-70k presumably shifted to the size of 32 kDa, since the intensity of the U1-A band increased slightly. U1-A itself could be also truncated to fragments corresponding to those of the digestion by Lys-C.

Carboxypeptidase Y treatment of U1 snRNP resulted in a truncation of U1-A, U1-C and B/B' (lane 21, figure 3.5). U1-70k was not affected by the protease. It is not clear whether the other Sm proteins were digested by carboxypeptidase Y. Interestingly, previously observed fragments of 13 kDa, and 18 kDa, appeared, which could be attributed to U1-A or B/B'. Leucine-aminopeptidase treatment resulted in a truncation of U1-70k (lane 26) and U1-A (lane 27). A band of 13 kDa, appeared upon proteolysis of U1-A, indicating a stable N-terminal fragment of the protein. The latter fragment could also originate from U1-70k.

Several U1 snRNP proteins were truncated by proteolysis. In particular U1-70k, B/B' and U1-A were digested by a broad range of proteases. These results were used to determine suitable proteases for *in situ* and in-solution proteolysis. Chymotrypsin was selected for initial *in situ* proteolysis experiments. Chymotrypsin or Asp-N were selected to prepare truncated U1 snRNPs in a large scale for crystallization.

3.1.2.2 Limited proteolysis of U1 snRNP in the presence of 5' splice site mimic oligonucleotides

Crystallization of U1 snRNP was carried out in the absence and presence of DNA and RNA oligonucleotides which bind to the particle as a 5' splice site mimic. It was tested if the presence of oligonucleotides leads to a different chymotrypsin proteolysis pattern of U1 snRNP in solution. A change in the proteolysis pattern could indicate an interaction of components of the particle with the oligonucleotide. As shown in figure 3.6 limited proteolysis experiments with 1:6 dilutions of chymotrypsin in the absence
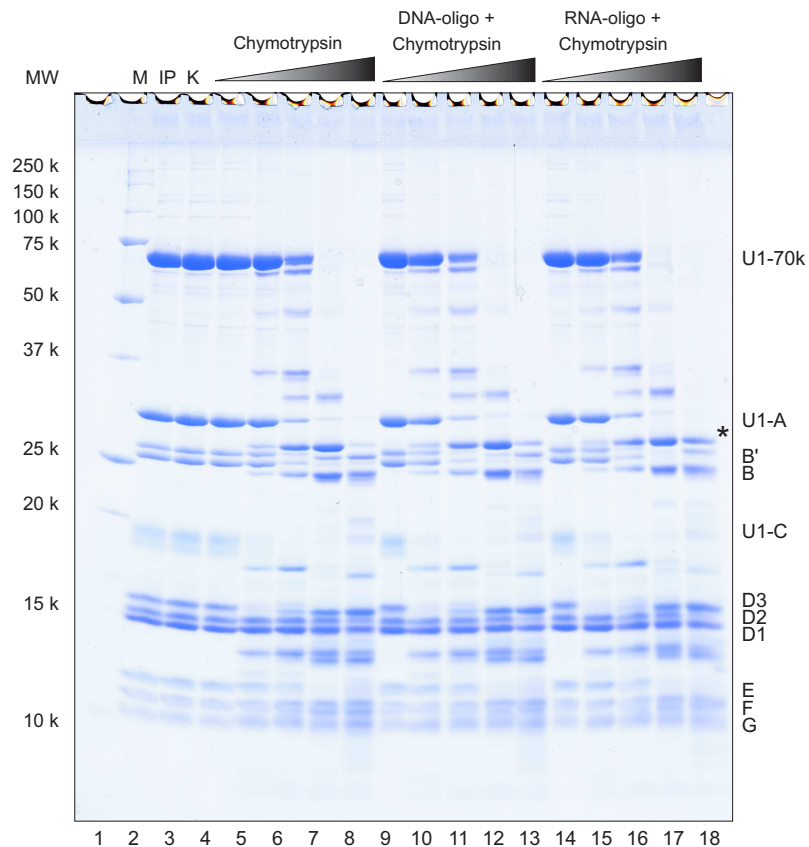
Figure 3.6 Limited proteolysis of U1 snRNP with chymotrypsin in the presence of a DNA or RNA-oligonucleotide mimicking a 5' splice site. 20 µg of U1 snRNP were treated with 1:6 dilutions of proteases for 22 h at 4 ℃. The reactions were stopped by adding SDS loading dye and applied on a 13.5 % polyacrylamide gel. The proteins were visualized by Coomassie staining. Lane 1 is the molecular weight marker, lane 2 is the U1 snRNP input, lane 3 is the control which was incubated in the absence of proteases. Lanes 4-8 are a series of six-fold dilutions of U1 snRNP digestions from 380 pg to 0.5 µg of chymotrypsin, lanes 9-13 are a series of six-fold dilutions of U1 snRNP digestions from 380 pg to 0.5 µg of chymotrypsin in the presence of the oligonucleotide DNA4, lanes 14-18 are a series of six-fold dilutions of U1 snRNP digestions from 380 pg to 0.5 µg of chymotrypsin in the presence of the oligonucleotide RNA8. The molecular weight (MW) of the marker proteins is given on the left side of the gel. The identities of the full-length proteins are denoted on the right side of the gel.

(lanes 5-9) and presence of DNA (lanes 10-14) and RNA (lanes 15-19) oligonucleotides were conducted.

The presence of DNA (figure 3.6, lane 14) or RNA (lane 19) does not alter the general pattern which was generated by chymotrypsin in the absence of oligonucleotides (lane 9). In the case of the RNA oligonucleotide present (lane 19), a fragment of about 34 kDa, (lane 19, asterisk) was stabilized, which could originate from U1-A or U1-70k. In the presence of both oligonucleotides, the D2 protein was destabilized and affected by

proteolysis (compare figure 3.6, lanes 9, 14 and 19). Taken together the presence of DNA and RNA oligonucleotides affects the pattern of chymotrypsin proteolysis of U1 snRNP. The proteins U1-70k or U1-A and D2 are differently truncated by chymotrypsin.

3.1.3 Crystallization of U1 snRNP by *in situ* proteolysis

Initial trials to produce crystals of full-length U1 snRNP in a reproducible fashion were not successful. In some cases, tiny crystals grew after 4 - 8 weeks. These crystals spanned around 20 μm at the longest dimension and diffraction was hardly detectable. Most of the attempts to crystallize U1 snRNP failed.

When dissolved crystals were analyzed by SDS-PAGE, it was noticed that several proteins of the particle were truncated. This truncation could be attributed to a removal of flexible domains of the proteins and regions of low complexity. Concluding from these observations, it was hypothesized that U1 snRNP crystals were only obtained, when a contaminating protease was present in the crystallization droplets, that truncated the particles. Therefore the deliberate inclusion of proteases in the crystallization setup was
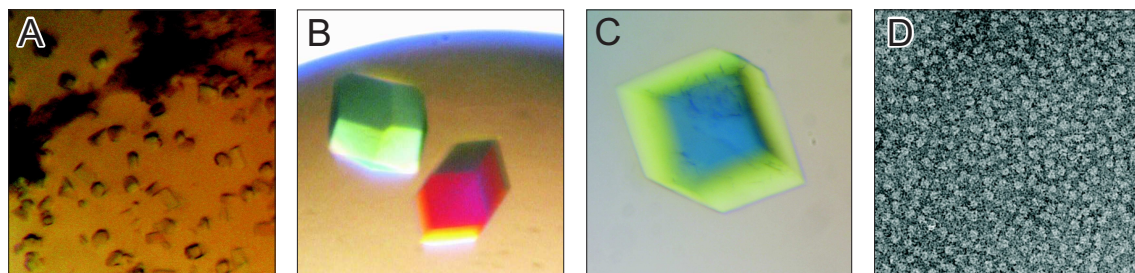
Figure 3.7 U1 snRNP crystals grown in the presence of chymotrypsin. **A** Initial U1 snRNP crystals were grown in a solution of 13.6 mg/ml U1 snRNP in crystallization buffer 120 ATP and reservoir solution containing 100 mM MES, pH 6.5, 10 % PEG 6000 and 10 mM glycylglycyl glycine. Crystals of 15 μm × 15 μm × 15 μm size appeared after 3 days at 4 °C in the presence of 13 ng chymotrypsin. **B** U1 snRNP crystals were grown in a solution of 20 mg/ml U1 snRNP in crystallization buffer 120 ATP and reservoir solution containing 100 mM MES, pH 6.5, 8 % PEG 6000, 0.8 M lithium sulfate, 40 mM glycylglycyl glycine and 20 mM DTT. Three-dimensional crystals of about 150 μm × 100 μm × 100 μm appeared after 30 days at 4 °C in the presence of 20 ng chymotrypsin. **C** U1 snRNP crystals were grown in a solution of 15 mg/ml U1 snRNP in crystallization buffer 120 ATP and reservoir solution containing 100 mM cacodylate-HCl, pH 6.5, 8 % PEG 6000, 0.6 M ammonium sulfate, 40 mM glycylglycyl glycine and 20 mM DTT. Crystals with dimensions of about 200 μm × 200 μm × 150 μm appeared after 30 days at 4 °C in the presence of 15 ng chymotrypsin. **D** Crystals shown in panel B were washed twice directly in the well, transferred to a new drop and dissolved in U1 crystallization buffer. The dissolved crystals were analyzed by electron microscopy according to (Kastner and Lührmann, 1989).

tested to obtain the crystals in a reproducible fashion. This approach has been also recently described in Dong *et al.* (2007) for single proteins.

Three proteases were included in the crystallization setup in different concentrations (see 2.2.4.1). A crystal screen was set up with a 13.6 mg/ml solution of U1 snRNP in crystallization buffer ATP, using linbro sitting drop plates. Initially, a crystallization condition was used, where U1 snRNP crystal growth was previously observed after 4-6 weeks (personal communication U. Reidt). After three days at 4 ℃ small crystals grew in the presence of chymotrypsin, shown in figure 3.7 A. The crystallization setup at 4 ℃ was absolutely required for crystal growth. Since all further attempts to obtain U1 snRNP crystals at room temperature failed all following screens were carried out at 4 ℃.

To visualize the crystal's contents and verify the integrity of U1 snRNP in the crystal lattice, electron microscopic studies were carried out. Crystals as shown in figure 3.7, panel B, were washed with 20 μl of reservoir solution, dissolved in crystallization buffer ATP and analyzed by electron microscopy according to (Kastner and Lührmann, 1989). U1 snRNP-like particles were visible, presumably missing parts of one or both protuberances, that were attributed to U1-70k and U1-A. However the typical doughnut

Space Group	P222$_1$
Unit Cell	156 Å, 178 Å, 220 Å; 90 , 90 , 90
Complexes per Asym. Unit	2
Resolution (Å)	50.0 – 9.0
Completeness (%)	99.6 (100)
R$_{sym}$ (%)	7.0 (38.3)
I/σ (I)	13.9 (3.4)
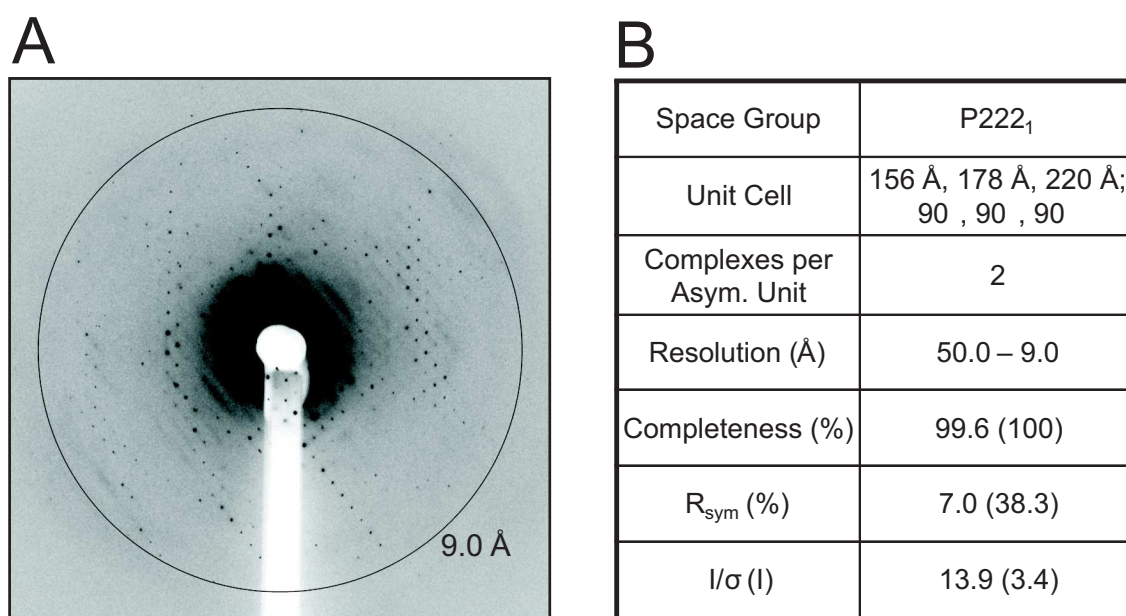
Figure 3.8 Diffraction image and statistics of a dataset collected from of U1 snRNP crystals. **A** U1 snRNP crystals grown as described in figure 3.7 C, were cryoprotected in a solution containing reservoir and 15 % propylene glycol. The crystals were flash-frozen in liquid nitrogen and diffracted to 9 Å at PX1 at SLS, Villigen, CH. **B** Data collection statistics of the U1 snRNP crystals.

shaped Sm core was present in the particles, that were obtained from dissolved crystals. Initial small U1 snRNP crystals were improved by the addition of 20 mM DTT, 40 mM glycylglycyl glycine and 0.8 M lithium sulfate or 0.6 M ammonium sulfate. Crystals of around 200 μm by 200 μm were grown in a reproducible fashion (see figure 3.7 B and C respectively). Crystals were flash frozen in reservoir solution with $15 - 18$ % propylene glycol for cryoprotection. The diffraction capability of the crystals was tested at the beamlines BW6 Hamburg and PX1/PX2 at SLS Villigen, CH. The crystals shown in figure 3.7 produced a diffraction pattern to around 9 Å resolution at PX1, SLS, CH (see figure 3.8 A). The space group of these crystals was determined as $P222_1$ with unit cell constants of a = 156, b = 178, c = 220 Å and $\alpha = \beta = \gamma = 90°$ (see figure 3.8 B). According to the cell constants, two molecules of the truncated particle were assumed in the asymmetric unit, assuming that proteolysis reduced the mass of the particles by $30 - 40$ %. All attempts to enhance the diffraction power of these crystals including controlled dehydration and further additives, failed.

3.1.4 Crystallization of U1 snRNP with a truncated 5' splice site by *in situ* proteolysis

The diffraction of *in situ* proteolyzed U1 snRNP crystals (described in 3.1.3) could not be improved further. Apart from the proteins, U1 snRNA can be modulated to obtain crystals with better diffraction power. The 5' region of U1 snRNA is an important feature of U1 snRNP and is accessible for interaction of the particle with the 5'splice site via base-pairing Lazar *et al*. (1982). Since this region of U1 snRNA is presumably located at the surface of the particle, it can be modulated. Particles with an altered 5' splice site binding region might lead to changes in crystal packing and result in different crystal forms. The 5' splice site binding region can be base-paired to short oligonucleotides. An alternative option was to entirely remove the The 5' splice site binding region. U1 snRNA can be truncated by a DNA-oligo directed RNase H treatment as described in Seiwert and Steitz (1993); Rossi *et al*. (1996). The reaction had to be scaled up to meet the demands of crystallographic studies.

Different amounts of RNase H concentration, $MgCl_2$ concentration and temperature were tested in a small scale and transferred to a reaction with 5.5 mg of U1 snRNP according to 2.2.2.2. After incubation the sample was buffer-exchanged with a 16/60 desalting column to crystallization buffer according to 2.2.3.8.

The sample was then passed over a 2 ml H20 immunoaffinity column, which was pre-equilibrated with 30 ml of crystallization buffer. This step was included to remove
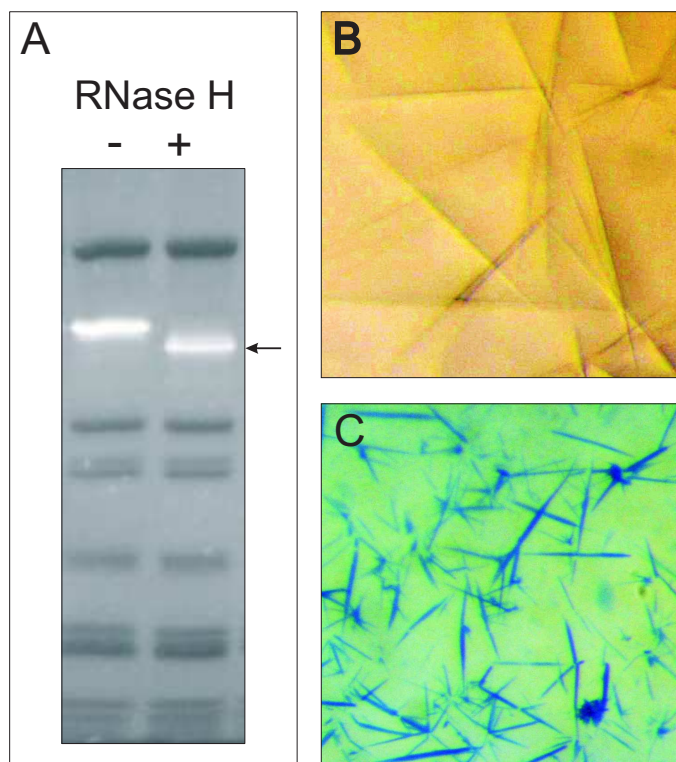
Figure 3.9 Crystallization of U1 snRNP with a truncated 5' splice site by *in situ* proteolysis. **A** U1 snRNA was truncated (indicated by an arrow) at the 5' splice site binding region by DNA-oligo directed RNase H treatment, visualized by Coomassie staining of a denaturing 13.5 % SDS-polyacrylamide gel. The lane '-' represents untreated particles, '+' is the RNase H treated U1 snRNP. **B** U1 snRNP crystals were grown in a solution of 15 mg/ml U1 snRNP with a truncated 5'splice site binding region and a reservoir solution containing 100 mM HEPES, pH 7.0, 10 % PEG 5500 monomethyl ether, 5 % taccsimate and 7.5 μg of chymotrypsin. Needle-like crystals appeared after 4 days. **C** Crystals described under 'B' were stained with IZIT crystal stain.

particles, in which the 5' splice site binding region was not truncated. Truncation of the RNA was verified on a denaturing 13.5 % SDS-polyacrylamide gel, as shown in figure 3.9 A. UV shadows of the protein bands demonstrate, that the protein components of U1 snRNP were not affected by the RNase H treatment.

Finally the particles were concentrated by pelleting according to 2.2.3.9 and a final concentration of 15 mg/ml was achieved. The material was supplemented with a 1:2000 mass ratio of chymotrypsin and used for a first round of screening in a 96-well format. The screens INDEX, MbClass I, MbClass II, Salt RX, Classics lite, Nucleix, SM I and SM III were used. Only one condition, INDEX F3, yielded needle-shaped crystals, which were reproduced in a refinement screen, shown in figure 3.9 B, and stained with IZIT crystal stain (see figure 3.9 C). The stain demonstrated that the crystals contained

protein. The crystals shown in figure 3.9 B were cryoprotected by adding 16 % propylene glycol and flash frozen in liquid nitrogen. At the beamlines PX2/SLS and ID23-1/ESRF no diffraction was detectable.

3.1.5 Co-crystallization of U1 snRNP with snurportin 1 employing *in situ* proteolysis

Snurportin 1 (SPN1) interacts with the m_3G cap of U1 snRNP and is therefore suitable for co-crystallization with U1 snRNP. The U1 snRNP-SPN1-complex might produce different crystal forms than U1 snRNP alone, that might diffract to a better degree. Full-length SPN1 was expressed with the plasmid pGEX-SPN1 (Strasser *et al.*, 2004) and purified as described in 2.2.3.2, 2.2.3.7 and 2.2.3.8 according to Strasser *et al.* (2004).

Figure 3.10 Purification of snurportin 1 (SPN1). Snurportin 1 from *Homo sapiens* was expressed and purified according to Strasser *et al.* (2004). About 10 μg of purified protein were applied on a 13.5 % SDS-polyacrylamide gel. The Coomassie stained gel shows the marker (M) and snurportin 1 (SPN1). The molecular weight (MW) of the relevant marker proteins is given on the left side of the gel.
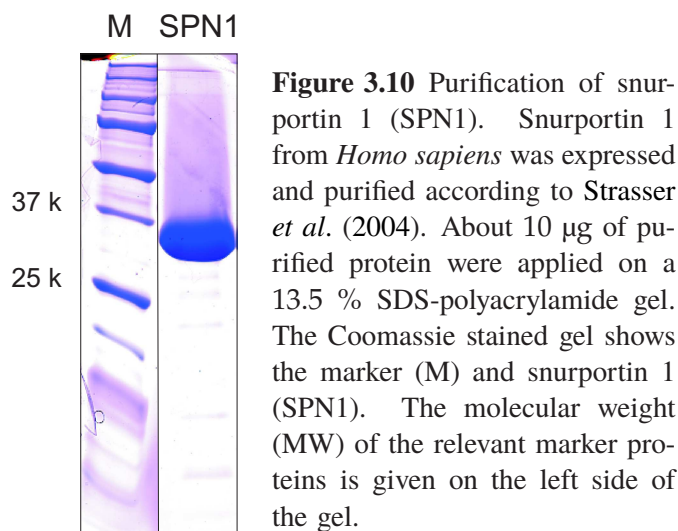
Figure 3.10 shows the purified snurportin 1 from *Homo sapiens* visualized on a Coomassie-stained 13.5 % SDS-polyacrylamide gel. The protein migrates at the expected size of around 37k.

To assess binding of SPN1 to purified U1 snRNP, a gel filtration of U1 snRNP, mixed with SPN1 in a 1:2 molar ratio, was carried out. Crystallization buffer was used as running buffer. The input and the fractions of the run were applied on a 13.5 % SDS-polyacrylamide gel and visualized by Coomassie stain (see figure 3.11). During purification SPN1 eluted as a single peak from a gel filtration column at the expected molecular weight (not shown). When SPN1 was premixed with U1 snRNP and applied on a gel filtration column, SPN1 co-eluted with U1 snRNP, indicating an association
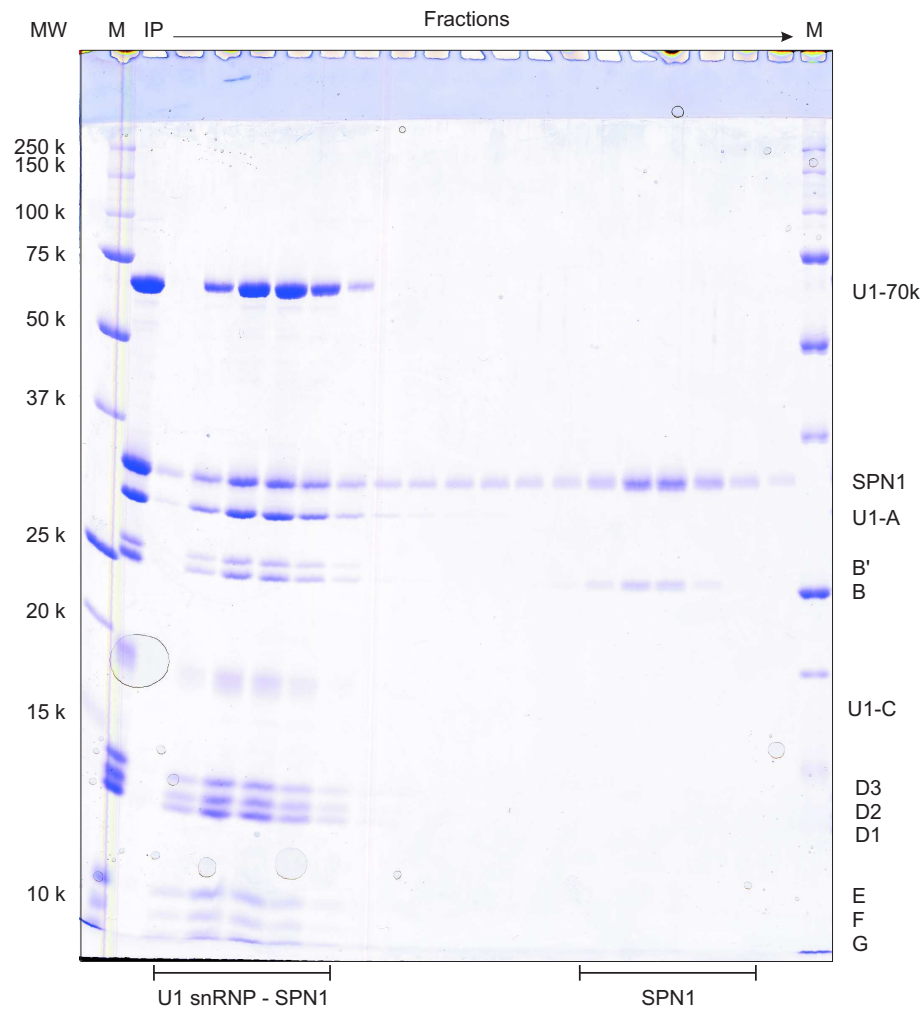
Figure 3.11 Gel filtration of recombinantly expressed SPN1 and U1 snRNP. Full-length snurportin 1 from *homo sapiens* was expressed and purified according to Strasser *et al*. (2004). The protein was mixed in a 2:1 molar ratio with U1 snRNP. The Coomassie stained gel shows the marker (M), the input (IP) consisting of 9 μg U1 snRNP and a two-fold molar excess of SPN1. In total about 50 μg of protein were applied on the Superdex 200 PC 3.2 column at a flow rate of 40 μl/min. The fraction size was 40 μl. The molecular weight (MW) of the relevant marker proteins is given on the left side, the identities of the U1 snRNP proteins and SPN1 on th right side of the gel.

of both SPN1 and U1 snRNP (see figure 3.11). Surplus SPN1 eluted in the lower molecular weight fractions.

Purified U1 snRNP and snurpotin 1 were mixed in a 1:1.5 ratio and were subjected to crystallization screens in the absence or the presence of 1:2000 chymotrypsin. The screens used are listed in the methods section (see table 2.11, screens 29 and 30). Crystals were found in crystallization conditions that were related to those of U1 snRNP alone and contained chymotrypsin. Also the morphology of these crystals was nearly

identical to that of U1 snRNP crystals grown in the absence of SPN1. Some of these crystals were frozen in liquid nitrogen in the presence of various cryoprotectants and propylene glycol and tested at synchrotron beamlines, but diffraction was restricted to $25 - 30$ Å (not shown). Due to the similarity of the SPN1-U1 snRNP co-crystals and original crystals and the failure to improve diffraction power, this approach was not followed further.

3.1.6 Co-crystallization of U1 snRNP with RNA-oligonucleotides employing *in situ* proteolysis

The 5' splice site constitutes a substrate for U1 snRNP binding, since it is complementary to the 5' end of the U1 snRNA (Lazar *et al*., 1982; Black *et al*., 1985). Truncation of the 5' splice site binding region had an effect on the crystallization of U1 snRNP. It dramatically reduced the number of conditions that produced crystals (see 3.1.4). The crystals that were obtained, did not diffract. An alternative option to modulate the 5' splice site binding region of U1 snRNP was to employ oligonucleotides for co-crystallization, that mimic a 5' splice site. The conformational change of the U1 snRNA 5' region upon oligonucleotide binding might lead to novel crystal contacts, novel crystal forms and crystals with higher diffraction power.

RNA-oligonucleotides of varying length, that base pair to the 5'splice site of U1 snRNP were employed for co-crystallization in the absence and presence of chymotrypsin. Table 2.11 lists all crystallization screens which were carried out with U1 snRNP and RNA oligonucleotides. Without chymotrypsin and in the presence of the RNA oligonucleotide RNA5, crystal formation was observed in 1 out of 1050 conditions tested (see table 3.1, condition 1). Crystals were small, plate-like, fragile and irreproducible. Therefore they could not be tested for diffraction, and no further screens in the absence of proteases were carried out. Likewise crystal growth was entirely abolished at room temperature also in the presence of chymotrypsin (see table 3.1). At 4 °C in the presence of chymotrypsin, crystals grew with all RNA-oligonucleotides which were tested. In the following only crystals that had a distinct morphology are mentioned.

Crystals that were obtained in 96-well format, were refined for size and morphology by variation in precipitants and pH or supplementation of additives. Figure 3.12 depicts the co-crystals obtained of U1 snRNP and different RNA-oligonucleotides. One typical class of U1 snRNP crystals which appeared frequently had a hexagonal rod-like shape. The latter crystals were improved in size to dimensions up to 500 μm length 100 μm diameter (see figure 3.12 C, D). The growth and appearance of these

crystals depended on the oligonucleotide, which was used for co-crystallization and was furthermore restricted to reservoir solutions containing ammonium phosphate, glycerol or ethylene glycol. Furthermore a buffer or salt was often absent in these crystallization conditions, indicating that the reduction of salt concentration might have contributed to crystal growth. Although the morphological features of these crystals were excellent, they did not have sufficient diffraction power for clear indexing. Hexagonal crystals which grew in ethylene glycol were cryoprotected by adding ethylene glycol up to a final concentration of 35 % and tested for diffraction at PX2/SLS. The space group of these crystals, which diffracted to around 20 Å, was determined to be C222 with cell constants of a = 481, b = 583, c = 335 and $\alpha = \beta = \gamma = 90°$. The presence of a screw axis remained uncertain (C222$_1$). A higher symmetry than orthorhombic could not be excluded. Other cryoprotectants than ethylene glycol did not improve diffraction.

Table 3.1 Co-crystallization of U1 snRNP with RNA-oligos. The list contains only crystallization conditions, which led to crystals, which were unique in morphology, chemical composition of the crystallization buffer or the oligonucleotide used for co-crystallization. 'No.' refers to the number of the crystallization condition, 'screen' refers to table 2.11 on page 48 .

No.	Screen	Condition	Buffer	Morphology
1	12	Cations C4	0.1 M sodium acetate anhydrous pH 4.6, 1.25 M Lithium acetate dihydrate	multiple, plates
2	13	MbClass II F10	0.1 M tri-sodium citrate pH 5.6, 12 % PEG 4000, 0.1 M sodium chloride	thin plates
3	13	Classics lite A11	0.1 M Tris-HCl pH 8.5, 15 % isopropanol, 0.2 M ammonium acetate	spherulites
4	13	pHClear I G11	0.1 M Tris-HCl pH 8.0, 20 % 2,4,MPD,	spherulites
5	8	Nucleix C2	0.05 M cacodylate-NaOH pH 5.0, 10 % PEG 8000, 0.2 M KCl 0.1 M magnesium acetate	plate-like, multiple
6	8	pHClear II H4	0.1 M Bicine pH 9.0, 10 % PEG 6000, 0.8 M LiCl	plate-like, multiple
7	8	MbClass II A12	1 M di-ammonium hydrogen phosphate,	rod-like, small
8	8	Classics lite D9	0.8 M sodium citrate tribasic dihydrate pH 6.5	plate-like
9	9	Index B6	0.49 M sodium di-hydrogen phosphate, 0.91 M di-potassium hydrogen phosphate	hexagonal rods, small
10	9	Classics lite D9	0.8 M sodium citrate tribasic	plate-like
11	9	Mb II G3	0.1 M Tris-HCl pH 8.5, 12 % PEG 4000, 0.2 M lithium sulfate	plate-like
12	10	Mb I F2	1 M KCl, 5 % PEG 4000	hexagonal rods
13	10	Classics lite B2	12.5 % ethylene glycol	hexagonal rods
14	18	Classics lite C1	0.2 M ammonium phosphate	hexagonal rods
15	31	MbClass I F2	1 M potassium chloride, 5 % PEG 4000	hexagonal rods
16	31	MbClass I G12	0.05 M potassium phosphate, 12 % PEG 8000, 10 % 2,4-MPD	hexagonal rods

Table 3.1 Co-crystallization of U1 snRNP with RNA-oligos ...

No.	Screen	Condition	Buffer	Morphology
17	31	MbClass I G12	12.5 % ethylene glycol	hexagonal rods
18	6	Classics lite C1	0.2 M ammonium phosphate	hexagonal rods
19	6	Classics lite G8	0.05 M potassium phosphate, 10 % PEG 8000	hexagonal rods
20	6	MbClass II F10	0.1 M tri-sodium citrate pH 5.6, 12 % PEG 4000, 0.1 M sodium chloride	plates
21	28	Index C1	3.5 M sodium formate pH 7.0,	pyramids
22	15	Classics lite A7	0.1 M tri-sodium citrate pH 5.6, 10 % PEG 4000, 10 % isopropanol	flat needles
23	16	Classics lite A7	0.1 M tri-sodium citrate pH 5.6, 10 % PEG 4000, 10 % isopropanol	flat needles
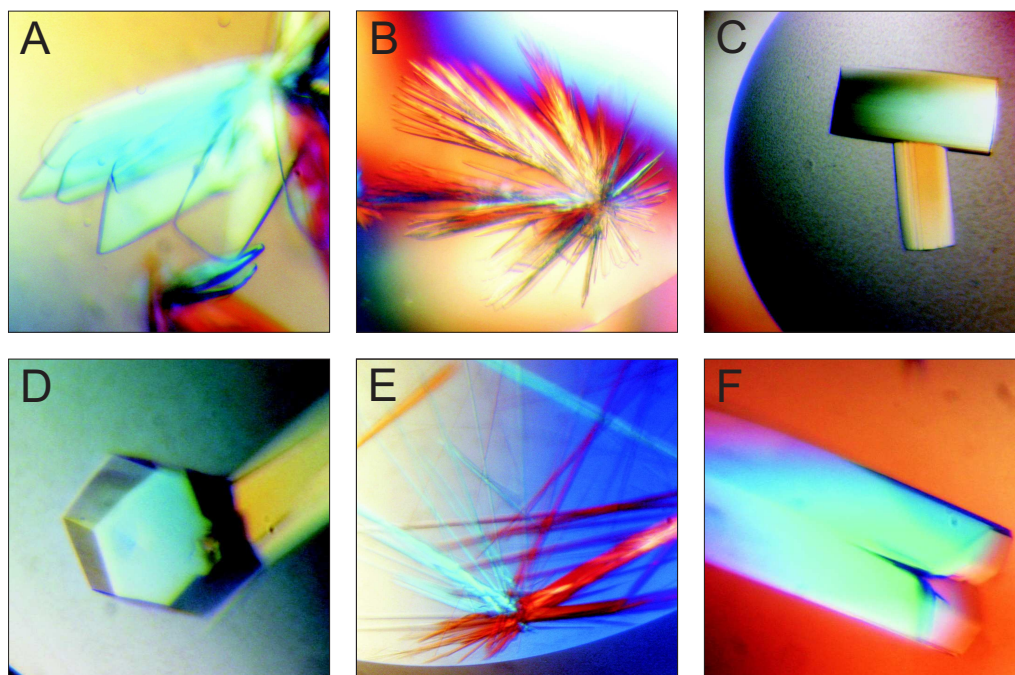
Figure 3.12 Co-crystals of U1 snRNP and RNA-oligonucleotides. U1 snRNP was co-crystallized with a 1:2 molar ratio of different RNA-oligonucleotides and a 1:2000 molecular weight ratio of chymotrypsin. **A** Crystals obtained from a refinement screen of crystallization condition No. 2 with RNA-oligonucleotide RNA5. Crystal growth was observed after 4 weeks in 0.1 M tri-sodium citrate pH 5.6, 8 − 10 % PEG 4000 and 0.1 M sodium chloride. **B** Crystals obtained from an additive refinement screen of crystallization condition No. 2 with RNA-oligonucleotide RNA5. Crystal growth was observed after 4 weeks in 0.1 M tri-sodium citrate pH 5.6, 8 − 10 % PEG 4000, 0.1 M sodium chloride and 10 mM EDTA. **C** Crystals obtained from a refinement screen of crystallization condition No. 13 with RNA-oligonucleotide RNA3. Crystal growth was observed after 8 weeks in 21 % ethylene glycol. **D** Crystals obtained from a refinement screen of crystallization condition No. 14 with RNA-oligonucleotide RNA8. Crystal growth was observed after 1 day in 0.2 M ammonium phosphate. **E** Crystals obtained from a refinement screen of crystallization condition No. 14 with RNA-oligonucleotide RNA6. Crystal growth was observed after 1 week in 0.1 M tri-sodium citrate pH 5.6, 10 % PEG 4000, 10 % isopropanol. **F** Crystals obtained from a refinement screen of crystallization condition No. 5 with RNA-oligonucleotide RNA4. Crystal growth was observed after 4 weeks in 0.05 M cacodylate-NaOH pH 5.0, 10 % PEG 8000, 0.2 M KCl, 0.1 M magnesium acetate.

A second major class of U1 snRNP RNA-oligonucleotide co-crystals is shown in figure 3.12 A, B, E and F. These crystals depended on the presence of carboxylic acids like citrate, acetate or formate (see table 3.1, e.g. conditions No. 2, 5, 8, 10, 20, 21, 22 and 23). The best diffracting crystals were obtained from crystallization condition 2 and refinement screens thereof. These crystals were typically cryoprotected with the addition of 15 − 17 % propylene glycol and diffracted to a resolution of 8 Å at PX1 and PX2 at SLS. The space group was determined as C2 with cell constants of a = 358,

b = 88, c = 152, $\alpha = \gamma = 90°$ and $\beta = 112°$. The diffraction was further improved to $6 - 7$ Å with the addition of 2 mM EDTA. The addition of EDTA coincided with an increased frequency of multiple crystals (compare figure 3.1 A and B).

3.1.7 Analysis of U1 snRNP-RNA co-crystals by SDS-PAGE and mass spectrometry

U1 snRNP crystallized on its own and in the presence of RNA-oligonucleotides. Crystallization depended on the addition of chymotrypsin, and the truncation of U1 snRNP proteins. Several RNA-oligonucleotides were employed as U1 snRNP ligands and morphologically different U1 snRNP co-crystals were obtained, depending on the oligonucleotides, that were used. In the following it was tested, whether the morphology of the crystals corresponded to different degrees of truncation of U1 snRNP.

To analyze the extent of proteolysis, crystals were washed in reservoir solution and applied on a 13.5 % and 15 % composite SDS-gel to monitor the extent of proteolysis. Figure 3.13 shows the protein and RNA-content of U1 snRNP crystals visualized on a denaturing gel. Bands of interest, which are marked by a letter and a pointer, were excised from the gel and analyzed by mass spectrometry.

Figure 3.13 demonstrates that full-length RNA was present in all crystals. The band above the full-length RNA was presumably a gel artifact. U1 snRNP crystals that were grown in the absence of RNA-oligonucleotides (lanes 2 and 3) contained truncated versions of all U1-snRNP-specific proteins. Sm proteins like B/B' and D3 were also affected by the protease. The proteolysis patterns of the crystals grown with ammonium sulfate (see figure3.7 C) or lithium chloride (see 3.7 B) were indifferent. In lanes 4 and 5, co-crystals of U1 snRNP and RNA5 were applied. Like for regular U1 snRNP crystals, proteolysis affected all U1 snRNP-specific proteins as well as B/B' and D3. In the presence of the oligonucleotide a different pattern was generated by chymotrypsin, where a prominent band at 26 kDa (a) was absent and new bands at 25 kDa (j) and 24 kDa (k) appeared. The material in lane 6 represents dissolved crystals of U1 snRNP co-crystallized with RNA4 and does essentially not differ from pure U1 snRNP crystals (lanes 1 and 2). The proteolysis patterns of U1 snRNP co-crystals with RNA8 or RNA3, only slightly differed from each other, but contained other fragments than the residual crystals described, especially in the molecular weight range around 30 kDa.

According to mass spectrometry analyses, bands 'a' and 'w' clearly comprised U1-70k. All peptides found, localized to the N-terminal half of the protein, spanning from residue 16 to 180. Bands 'b' and 'm' were identified as U1-A, represented by peptides
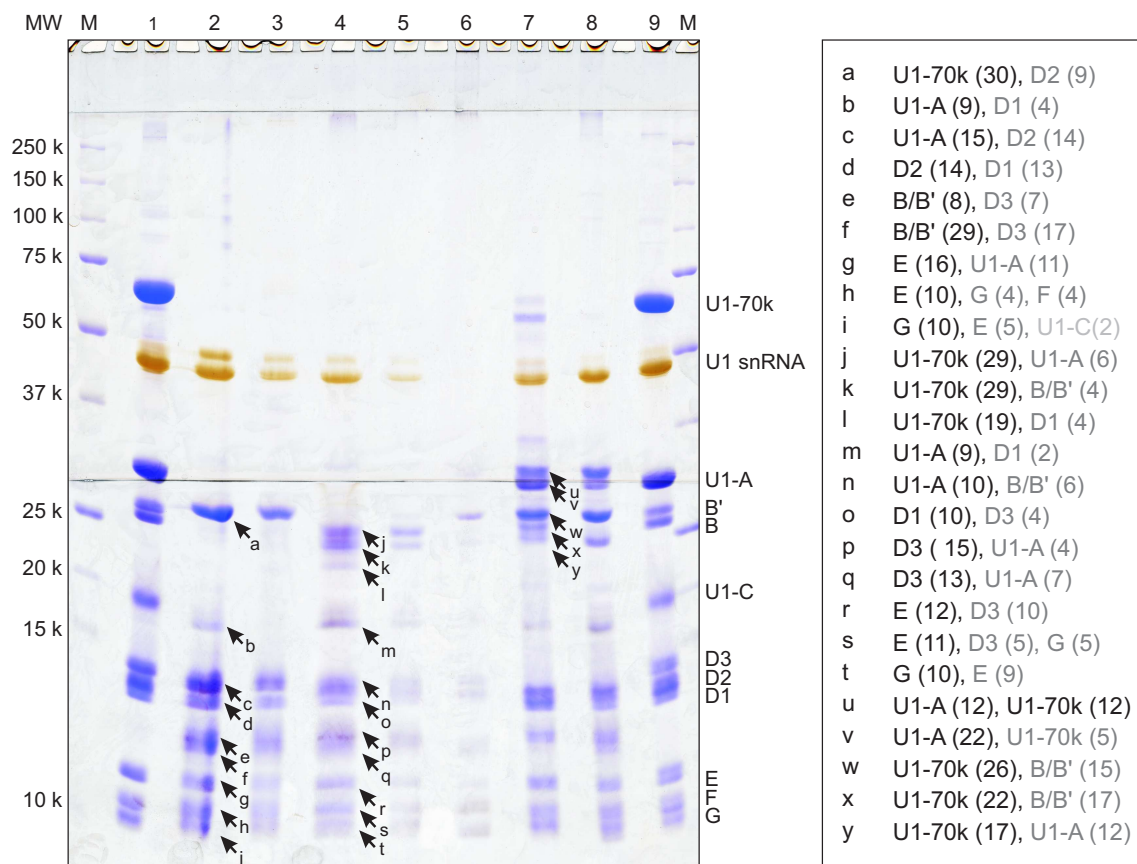
Figure 3.13 Gelelectrophoretic analysis of dissolved U1 snRNP crystals. 20-30 U1 snRNP crystals were transferred to 20 µl reservoir solution, washed with 20 µl of reservoir solution and resuspended in 10 µl SDS-loading dye. The proteins were resolved on an SDS-gel with an upper phase containing 13.5 % and a lower phase with 15 % polyacrylamide. The gel was stained with Coomassie and silver to visualize proteins and RNA respectively. Several bands were excised and analyzed by mass spectrometry, denoted by letters and arrows in the gel. The result of the mass spectrometry analysis is shown in a box right of the gel. For each band analyzed, the most prominent hit (black) as well as secondary hits (gray, light gray) are listed. The number of peptides found is shown in brackets. Identities of U1 snRNP proteins are shown on the right, size of the marker proteins (M) on the left side of the gel. Lanes 1 and 9 are 40 µg of U1 snRNP as input, lanes 2 and 3 represent U1 snRNP crystals without oligonucleotide, shown in figure 3.7 C and B, respectively. Lanes 4 and 5 represent co-crystals of U1 snRNP and oligonucleotide RNA5, also shown in figure 3.12 A. Lane 6 represents co-crystals of U1 snRNP and oligonucleotide RNA4, also shown in figure 3.12 F. Lane 7 represents co-crystals of U1 snRNP and oligonucleotide RNA8, also shown in figure 3.12 D. Lane 8 represents co-crystals of U1 snRNP and oligonucleotide RNA3, also shown in figure 3.12 C.

in the region from amino acid 50 to 152. Bands 'c' and 'n' could be attributed to the N-terminus of U1-A, which binds to the U1 snRNA. Respective peptides in the region of amino acids 28 - 96 were identified. Bands 'd' and 'o' mainly contained peptides of D1 and D2. The size of the bands indicated, that these proteins were not truncated. A double band (e, f, p, q) at around 12k contained fragments of B/B' and D3, presumably representing the Sm core domain of these proteins. All smaller fragments which were analyzed (g, h, i, r, s, t) comprised the Sm proteins E, F and G. Two peptides of U1-C were found in band i. They were assigned to the N-terminus of U1-C, representing the Zn-finger of the protein.

Upon the addition of oligonucleotide RNA5 (lanes 4 and 5) the U1-70k fragment (band a) was further truncated to a size of 24 kDa and 23 kDa (bands j, k). Co-crystals of U1 snRNP and oligonucleotides RNA8 and RNA3 (lanes 7 and 8, respectively) contained larger fragments at around 25 kDa to 30 kDa. Band 'v' was clearly identified to be U1-A, which seemed present as a full-length protein. The residual bands of these crystals (u, w, x, y), which were analyzed, could be attributed to U1-70k.

Taken together, the morphology of th U1 snRNP crystals depended on the oligonucleotide, that was used for crystallization as well as the crystallization buffer composition. The protein and RNA contents of U1 snRNP crystals and RNA-co-crystals were analyzed by SDS-PAGE. Each morphologically distinct crystal, that was analyzed, produced a specific proteolysis pattern, regarding U1 snRNP proteins. The different oligonucleotides changed crystallization behavior and modulated the *in situ proteolysis* pattern of U1 snRNP. Thus, the 5' splice site region was identified as being sensitive to modifications, concerning crystallization behavior of the particles. Therefore this region is likely to be exposed and can be extensively utilized to obtain better diffracting crystals of U1 snRNP.

3.1.8 Co-crystallization of U1 snRNP with DNA-oligonucleotides by *in situ* proteolysis

The 5' splice site binding region was identified as a sensitive region for the crystal formation of truncated U1 snRNP. Its removal nearly abolished U1 snRNP crystallization. Co-crystallization with oligonucleotides, that mimic a 5' splice site, and bind to the 5' end of U1 snRNA, produced morphologically different U1 snRNP crystals.. As shown in 3.1.6, U1 snRNP-RNA co-crystals, diffracting to around 6 Å resolution, were obtained in a reproducible fashion. The diffraction power of these crystals could not be improved by other cryoprotectants, by capillary mounting or the Rigaku free mounting system

(FMS). Therefore DNA oligonucleotides as well as other proteases than chymotrypsin were tested for co-crystallization and *in situ* proteolysis experiments with U1 snRNP, respectively.

U1 snRNP was mixed in a 1:2 molar ratio with the DNA-oligonucleotides DNA2, DNA3 or DNA4. The latter is an oligonucleotide which contains a self-complementary stretch. This oligonucleotide base pairs to the 5' splice site and can dimerize via base complementarity and recruit a second oligonucleotide to the 5' splice site binding region. This second oligonucleotide may be also base paired to U1 snRNP. Moreover the effect of the DNA2 oligonucleotide is further influenced by the crystallization condition and the stoichiometry to U1 snRNP.

Table 3.2 Co-crystallization of U1 snRNP with DNA-oligos and different proteases. The list contains only crystallization conditions, which led to crystals, which were unique in morphology, chemical composition of the crystallization buffer or the oligonucleotide used for co-crystallization. 'No.' refers to the number of the crystallization condition, 'screen' refers to table 2.11 on page 48 .

No.	Screen	Condition	Buffer	Morphology
24	21	Classics lite, A5	0.1 M HEPES pH 7.0, 10 % PEG 4000, 5 % isopropanol	plates
25	19	MbClassI, F2	0.1 M potassium chloride, 5 % PEG 4000	rods
26	19	MbClassII, D10	0.1 M tri-sodium citrate pH 5.6 0.1 M sodium chloride, 12 % PEG 4000	plates
27	19	MbClassII, F11	0.1 M ADA, pH 6.5 0.1 M lithium sulfate, 12 % PEG 4000	plates
28	19	Classics lite A7	0.1 M tri-sodium citrate pH 5.6, 10 % PEG 4000, 10 % isopropanol	round plates
29	19	Index, H10	0.2 M sodium citrate, 20 % PEG 3350	plates
30	20	SM III, H10	0.1 M HEPES-NaOH, pH 7.0, 1 M Na-K-tartrate	cubes
31	20	Index, C10	1 M succinic acid, pH 7.0, 0.1 M HEPES-NaOH, pH 7.0, 1 % PEG 2000 MME	cubes
32	20	SaltRX, C9	1.2 M D-L-malic acid, pH 7.0, 0.1 M bis-tris propane, pH 7.0	cubes
33	22	MbClass I, E8	0.05 M Tris-HCl, pH 7.5, 12 % PEG 2000 MME, 0.5 M sodium chloride	small rods
34	22	Classics lite, B10	0.1 M HEPES-NaOH, pH 7.5, 35 % 2,4-MPD	prolate spheroids
35	22	MbClass II, B7	0.1 M tri-sodium citrate, pH 5.6, 30 % PEG 400, 0.1 M sodium chloride	prolate spheroids
36	23	MbClass I, E9	0.05 M tri-sodium citrate, pH 5.6, 10 % PEG 3350, 0.15 M sodium chloride	plates
37	23	Pro Complex, D2	0.1 M MES, pH 6.5, 10 % PEG 5000 MME, 12 % 1-propanol	thin rods
38	23	MbClassII, D10	0.1 M tri-sodium citrate pH 5.6 0.1 M sodium chloride, 12 % PEG 4000	plates

Table 3.2 Co-crystallization of U1 snRNP with DNA-oligos …

No.	Screen	Condition	Buffer	Morphology
39	24	MbClass I, E8	0.05 M Tris-HCl, pH 7.5, 12 % PEG 2000 MME, 0.5 M sodium chloride	small rods
40	24	Index, E5	0.1 M Tris-HCl, pH 8.5, 45 % 2,4-MPD, 0.2 M ammonium acetate	prolate spheroids
41	24	JCSG, D5	0.1 M HEPES-NaOH, pH 7.5, 70 % 2,4-MPD,	small rods
42	25	Classics lite, B9	0.1 M Tris-HCl, pH 8.5, 25 % 2,4-MPD, 0.2 M ammonium phosphate	small spheres
43	25,26	Pro Complex, H9	0.1 M imidazole, pH 7.0, 50 % 2,4-MPD	undefined
44	27	MbClass II, A12	1 M di-ammonium phosphate	spherical
45	27	Nucleix, A1	0.1 M MES pH 6.5 2 M lithium sulfate, 0.01 M magnesium chloride	needle clusters
46	27	MbClass I, E7	0.1 M tri-sodium citrate pH 5.6, 10 % PEG 2000 MME, 3 % PEG 200, 20 % glycerol	needle clusters
47	27	MbClass I, F5	0.1 M HEPES-NaOH pH 7.5, 10 % PEG 4000, 0.1 M ammonium sulfate	rhombic rods
48	27	MbClass I, F6	0.1 M Tris-HCl pH 7.5, 10 % PEG 4000, 0.5 M sodium chloride	plates
49	27	MbClass I, F10	0.05 M MOPS pH 7.0, 12 % PEG 4000, 0.5 M potassium chloride 20 % glycerol	rhombic rods
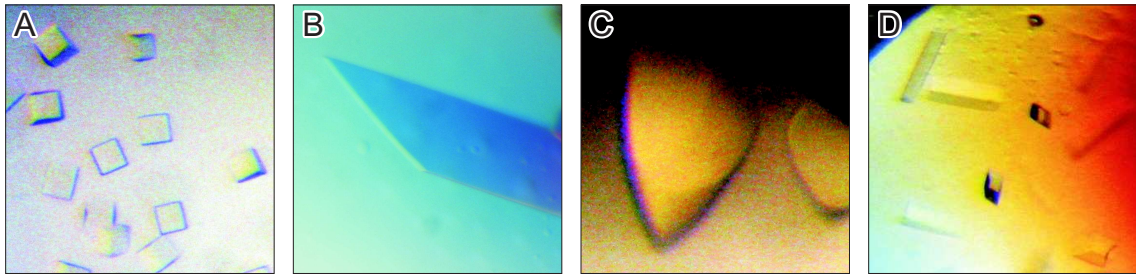
Figure 3.14 Co-crystals of U1 snRNP and DNA-oligonucleotides. U1 snRNP was co-crystallized with a two-fold molar excess of different DNA-oligonucleotides and several proteases. **A** Crystals obtained from a refinement screen of crystallization condition No. 30 with DNA-oligonucleotide DNA2. Crystal growth was observed after 6 weeks in 0.1 M HEPES-NaOH, pH 7.0, 1 M Na-K-tartrate in the presence of 20 ng chymotrypsin **B** Crystals obtained from an additive refinement screen of crystallization condition No. 38 with DNA-oligonucleotide DNA4. Crystal growth was observed after 4 weeks in 0.1 M tri-sodium citrate pH 5.6, $8 - 10$ % PEG 4000, 0.1 M sodium chloride, 2 mM EDTA and $1 - 2$ % Anapoe-X-305 in the presence of 20 ng chymotrypsin. **C** Crystals obtained from a refinement screen of crystallization condition No. 35 with DNA-oligonucleotide DNA4 in the presence of 2 ng Asp-N. Crystal growth was observed after 8 weeks in 0.1 M tri-sodium citrate, pH 5.6, 30 % PEG 400 and 0.1 M sodium chloride. **D** Crystals obtained from crystallization condition No. 47 with DNA-oligonucleotide DNA4. Crystal growth was observed after 4 weeks in 0.1 M HEPES-NaOH pH 7.5, 10 % PEG 4000 and 0.1 M ammonium sulfate in the presence of 20 ng elastase.

Figure 3.14 illustrates crystals obtained from U1 snRNP with different DNA-oligonucleotides. The cube-like U1 snRNP co-crystals were obtained with the oligonucleotide DNA2, which was used in a 2-fold molar excess, and chymotrypsin as a protease (see table 3.2, conditions 30-32). These crystals were dependent on the presence of carboxylic acids like tartrate, malate or succinate (see figure 3.14 A). This resembled the buffer requirements of plate-like co-crystals of the RNA-oligonucleotide RNA5, as previously described. The best diffraction was observed at ESRF/ID23-1 when the cube-like crystals were cryoprotected with 20 % propylene glycol. The diffraction pattern extended to a maximum resolution of 20 Å but was not sufficient for indexing.

The so far most promising RNA co-crystals were plates, grown in the presence of chymotrypsin and the oligonucleotide RNA5 (table 3.1, condition 2). The same crystals with identical cell constants could be reproduced with the DNA-oligonucleotide DNA4. In an analogous fashion, EDTA enhanced diffraction but led to the growth of multiple crystals. The addition of $1 - 2$ % Anapoe-X-305 or glycylglycyl glycine suppressed this negative effect of EDTA in most cases. The exchange of DNA for RNA increased resolution to $5 - 4.5$ Å. At SLS/PX2 one out of 5 - 10 crystals of this kind diffracted to 5 Å resolution. Accordingly a dataset of 180° with $\Delta\phi = 1°$ and 1 s unattenuated

exposure was obtained at SLS/PX2. Data collection statistics were significant to around 5 Å. An over-exposure of 10 s of the crystals increased the maximum resolution, diffraction spots were visible to around $3.8 - 4$ Å (see figure 3.15 A). After at maximum 5 images, or 50 s unattenuated total exposure, these highly resolved reflections were not visible anymore.

A

3.9 Å

B

Space Group	C2
Unit Cell	359 Å, 87 Å, 152 Å 90°, 112°, 90°
Complexes per Asym. Unit	2
Resolution (Å)	50.0 - 4.5 (4.6 - 4.5)
Completeness (%)	98.5 (98.1)
R-factor (%)	4.8 (81.2)
I/σ	11.8 (1.74)
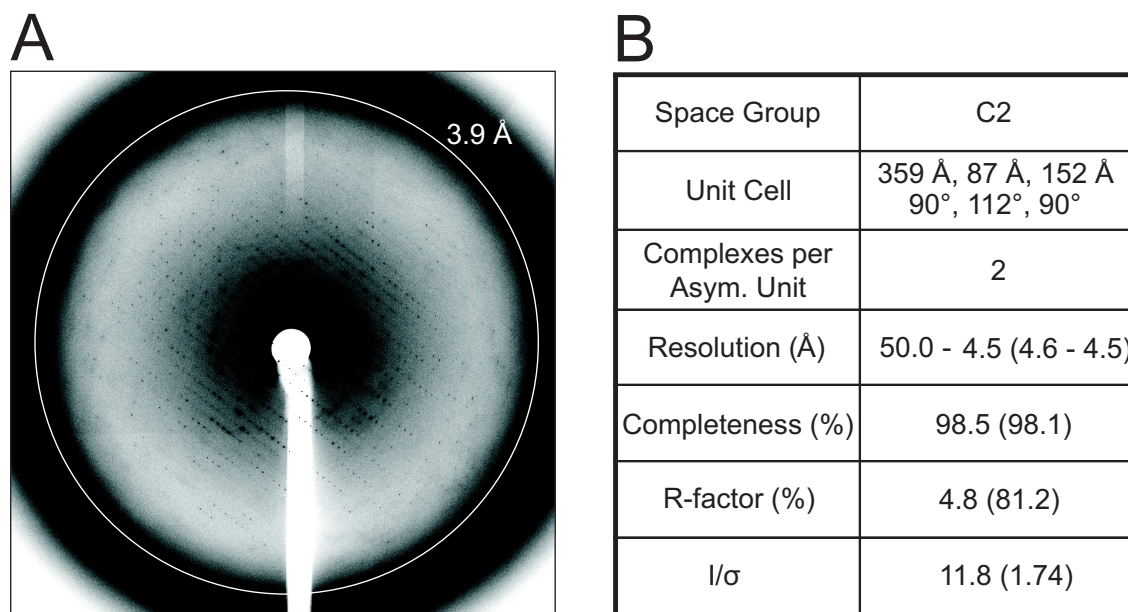
Figure 3.15 Diffraction of co-crystals obtained from the DNA-oligonucleotide DNA4 and U1 snRNP, grown in the presence of chymotrypsin. **A** A co-crystal of U1 snRNP and DNA4 (see 3.14 B) was cryoprotected with 15 % propylene glycol and exposed for 10 s at SLS/PX2 unattenuated with $\Delta\phi = 0.5°$ Å resolution of 3.9 Å is indicated by a circle. **B** Data collection statistics of a co-crystal obtained from the DNA-oligonucleotide DNA4 and U1 snRNP, grown in the presence of chymotrypsin (see 3.14 B). The crystal was cryoprotected with 15 % propylene glycol. The parameters for data collection over a rotation range of 180° were $\Delta\phi = 0.1°$ and 0.1 s exposure per image, attenuated by 50 %. Data were collected at SLS/PX1, employing the Pilatus 6M pixel detector. Data were integrated and scaled with XDS. The values in brackets represent the highest resolution shell between 4.5 Å and 4.6 Å.

The Pilatus detector has a lower background than regular CCD detectors and allowed for the collection of a complete dataset of the respective U1 snRNP crystals to 4.5 Å resolution (see figure 3.15 B) (Broennimann *et al.*, 2006). The parameters for data collection over a rotation range of 180° were $\Delta\phi = 0.1°$ and 0.1 s exposure per image, with the beam attenuated by 50 %. Taken together the refined U1$_{DNA4}$ co-crystals, originating from crystallization condition 2 (see table 3.1), produced plate-like crystals that diffracted to around 4.5 Å. The addition of EDTA and Anapoe-X-305 or glycylglycyl glycine enhanced morphology and diffraction of the crystals. Since these crystals were

the main thread for U1 crystallization, they are referred to as $U1_{DNA4}$ in the following. Other DNA-oligonucleotides of different length as well as the inclusion of other additives to the screens did not improve the diffraction. Growing crystals in 0.1 % to 0.3 % agarose or controlled dehydration with the FMS did not have a positive effect on diffraction either.

3.1.9 Co-crystallization of U1 snRNP with DNA4 oligonucleotide by *in situ* proteolysis with different proteases

Several proteases were assessed for in solution proteolysis of U1 snRNP (see 3.1.2.1). Each protease generated a different proteolysis pattern. A different truncation of the U1 snRNP proteins by *in situ* proteolysis may result in the growth of different U1 snRNP crystals. Six different proteases were used for *in situ* proteolysis of U1 snRNP, co-crystallized with the DNA oligonucleotide DNA4 in a two-fold molar excess. Table 3.2 summarizes a selection of the crystals obtained in these screens (see table 2.11, on page 48).

The crystals shown in figure 3.14 C were obtained by including the endopeptidase Asp-N in a screen with the oligonucleotide DNA4 (see table 3.2, condition 35) . The crystals required two months to grow and were already cryoprotected by 30 % PEG 400. The crystals were tested at SLS/PX1 (Pilatus detector) and diffracted to around 8 Å resolution. The space group was determined to be C222 with a = 138, b = 238, c = 228, $\alpha = \gamma = \beta$ 90°The presence of a screw axis ($C222_1$) could not be excluded. Cryoprotected crystals of the same input material and with the same morphology appeared with 2,4-MPD as a precipitant, but with lower diffraction power (not shown).

Figure 3.14 D illustrates U1 snRNP DNA4 crystals obtained in the presence of elastase. These crystals had a distinct morphology, rhombic rods, which had not been observed before. Two crystallization conditions (see table 3.2, conditions 47, 49) produced these crystals. Those of condition 47 were cryoprotected with 17 % propylene glycol and tested at SLS/PX2, but no diffraction was observed. The nearly cryoprotected crystals resulting of condition 49 were not tested for diffraction yet.

Only few morphologically undefined crystals of U1 snRNP were observed in the presence of the proteases Glu-C and Lys-C. The remaining crystals, which appeared in the DNA-oligonucleotide as well as the protease screens, were similar in morphology or the buffer

condition to the ones observed in previous co-crystallization trials with RNA. Most of these crystals have not been tested extensively yet.

3.1.10 Purification of a chymotrypsin treated U1 snRNP for crystallization

According to 3.1.8 crystals of U1 snRNP were obtained which diffract X-rays to a resolution of 4 Å. *In situ* proteolysis generates protein fragments that might disturb the crystal lattice and decrease the quality of the U1 snRNP crystals. Since *in situ* proteolysis was a prerequisite for crystal formation it was tested if U1 snRNP could be purified after proteolysis in solution. A gel filtration at 500 mM NaCl was included to remove fragments from proteolysis that were still bound to the particle.

The proteolysis experiment was adjusted to the condition that produced the best U1 snRNP crystals so far, including the oligonucleotide and protease. In the following a preparative proteolysis combined with purification and concentration is described. The
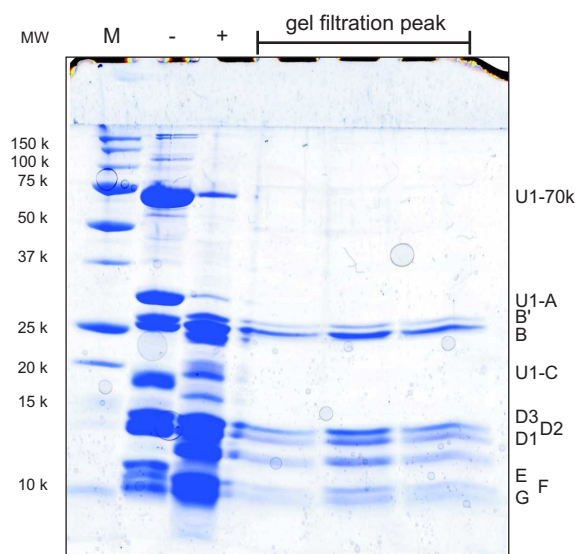
Figure 3.16 Purification of a chymotrypsin-treated U1 snRNP for crystallization. 5 mg of U1 snRNP were mixed with the oligonucleotide DNA4 in a 1:2 molar ratio and treated with 50 µg of chymotrypsin for 14 h at 4 °C. The resulting particles were immediately loaded on a Superdex 200 gel filtration column. The column was run in crystallization buffer with 500 mM NaCl at a flowrate of 2 ml/min, 5 ml fractions were taken. 20 µl of the U1 snRNP input (-) and the digested particles (+) next to the three peak fractions from gel filtration (gel filtration peak) were applied on a 13.5 % SDS-gel. The proteins were visualized by Coomassie stain. The molecular weight (MW) of the marker (M) proteins is given on the left side of the gel. Identities of the U1snRNP proteins are given on the right side of the gel.

appropriate experimental conditions for optimal proteolysis were first tested in a small scale (not shown). The large scale experiment was carried out as described in 2.2.3.11. The result of proteolysis and gel filtration is shown in figure 3.16. U1 snRNP specific proteins are truncated by chymotrypsin as observed in previous experiments (compare lane + and - with figure 3.6, lanes 3 and 13). Gel filtration resulted in a single peak, eluted over three fractions (gel filtration peak in figure 3.16). No absorption was detected in the excluded volume or in other fractions than the peak (not shown). After concentration with a Vivaspin concentrator, around 120 µl of truncated U1 snRNP with a concentration of 16 mg/ml were obtained. In contrast to the full length U1 snRNP, the truncated particle was not sensitive to membranes. The gel filtration at 500 mM NaCl separated several contaminating bands, especially between 15 kDa and 25 kDa from the particle. The bands have not been analyzed by mass spectrometry yet.

The concentrated and truncated U1 snRNP was used for a refinement screen and an additive screen of the DNA4-condition which produced the best crystals so far (table 3.2, condition 38). Roundish plate-like crystals appeared after 2 days in several drops of the refinement screen which contained Anapoe-X-305 or glycylglycyl glycine as an additive. Supplementation with oligonucleotide or chymotrypsin was not required for crystal growth. Diffraction of these crystals, cryoprotected in $15 - 17$ % propylene glycol and 25 % glycerol, was tested at ESRF and was limited to $15 - 20$ Å.

3.1.11 Calculation of a Self-rotation function of U1$_{DNA4}$ co-crystals

When no phase information is available, a dataset can still be used to calculate a Patterson map which is a Fourier transform, but without phases. The Patterson map can be rotated and superimposed on itself, which is called a self-rotation function. When the Patterson map of the unit cell is superimposable on itself in a different orientation, symmetry axes of crystallographic and non-crystallographic origin can be detected.

The CCP4i program molrep (Vagin and Teplyakov, 2000) was used to calculate a self-rotation function of a dataset, collected from U1$_{DNA4}$ crystals. Figure 3.17 shows the plot of a self-rotation function calculated at $\chi = 90°$ and $\chi = 180°$ using data between 48.6 and 4.4 Å resolution. Peaks were only observed at $\chi = 180°$. The peak at zero rotation presumably represents the crystallographic 2-fold axis along the b-axis as a consequence of the symmetry operators in the space group C2. A second peak was observed at $\theta = 100.58°$ $\phi = -180°$ and $\chi = 180°$. This second peak was observed at all native datasets and indicated the presence of a non-crystallographic symmetry.
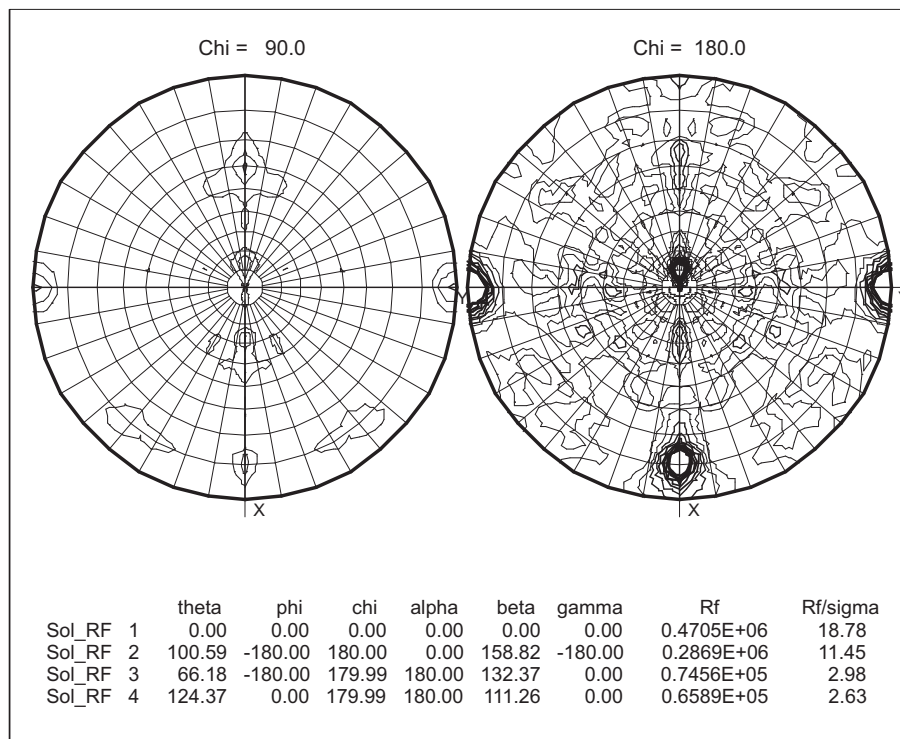
Figure 3.17 Self-rotation function of a dataset, collected from U1$_{DNA4}$ co-crystals. The dataset was collected at SLS/PX1, using the Pilatus pixel detector. Data were integrated and scaled with XDS. The self-rotation function with $\chi = 90°$ and $\chi = 180°$ was calculated with the program Molrep, a part of the CCP4i suite. Data between 48.6 and 4.4 Å were included. Below the plot of the rotation function, the first 4 peaks with $\chi = 0°$ (1) and $\chi = 180°$ (2,3,4) are shown as a list.

If the unit cell constants of the crystal and the size of the crystallized particle or protein are known, the number of particles or proteins per asymmetric unit can be calculated (Kantardjieff and Rupp, 2003). The crystal volume per unit of protein molecular weight, known as Matthews coefficient, is determined and compared to an average value. Unit cell content analyses were carried out by the program Phaser. According to the results of the limited proteolysis and mass spectrometric analyses of dissolved crystals, the molecular weight of truncated U1 snRNP was estimated to be about 180 kDa. Since the RNA was present as full-length in the crystals, the molecular weight contribution of RNA was estimated to be 60 kDa. Assuming a typical solvent content of protein

crystals, two molecules per asymmetric unit were predicted with a solvent content of 64 %. This result was corroborated by the self-rotation function that suggests the presence of a non-crystallographic symmetry in the asymmetric unit.

3.1.12 Approaches to the structure solution of U1 snRNP

3.1.12.1 Molecular replacement (MR)

Several components of U1 snRNP have been solved at atomic resolution by X-ray crystallography and NMR (Oubridge *et al.*, 1994b; Kambach *et al.*, 1999a; Muto *et al.*, 2004). These models could be employed to attempt structure solution by molecular replacement (MR). The crystal structure of U1-A in complex with an oligonucleotide, that represents stem loop II of U1 snRNA, and the solution structure of U1-C were directly used for molecular replacement. An entire structure of the human heptameric Sm core is not available. Only two hetero-dimers of the Sm core, D1 complexed with D2 and D3 complexed with B/B', have been solved at atomic resolution by X-ray crystallography. SmB/B' complexed with D3 and D1 complexed with D2 contain Sm-Sm interfaces, which could be used to form a heptameric ring by stepwise superposition. Neither B/B'-D3 nor D1-D2 were suitable models to generate a ring-like pseudo-symmetrical structure. Repetitive superposition of the molecules led to an open ring-like structure that lost seven-fold pseudo-symmetry.

Several structures of archeal homo-heptameric Sm cores are available at atomic resolution. These structures could be employed as a scaffold for the superimposition of known hetero-heptameric subunits from *Homo sapiens*.

A homo-heptameric Sm ring of *Archaeoglobus fulgidus* (PDB accession code 1I5L) was used as a scaffold to build a hetero-heptameric Sm ring, as described in figure 1.4 in chapter 1. This model is named Sm core model 1 in the following. Since it was artificially assembled, this model still contained amino acid clashes at its subunit-interfaces. These clashing amino acids are energetically unfavorable. Several tools for energy minimization of protein-protein interfaces are available. Sm core model 1 was energy-minimized with the Gromacs online-tool, the energy-minimized model is termed Sm core model 2 in the following.

In a homology-based search model, like employed here, the conformation of amino acid side chains is not known and might introduce errors. In a third search model these side chains were truncated to serines. The heptameric Sm ring from *archaeoglobus fulgidus* was converted to a poly-serine model with the program Moleman. All amino

acid residues were converted to serines, except for alanine. In the following this model is termed Sm core model 3.

If the unit cell constants of the crystal and the size of the crystallized particle or protein are known, the number of particles or proteins per asymmetric unit can be calculated (Kantardjieff and Rupp, 2003). As the unit cell contents, a molecular weight contribution of 120 kDa from protein and 60 kDa from RNA/DNA was assumed. Molecular replacements with Phaser were carried out for the models Sm core 1 - 3. Two heptameric Sm cores and in addition two molecules of U1-A complexed with a 21-mer of RNA (PDB accession code 1URN, chains B and Q) were used as a search model.

The replacement was carried out in a sequential fashion, starting with a consecutive search of two Sm heptamers, followed by U1-A complexed with RNA. An estimate of the correct placement of a component in the unit cell is the Z-score. Solutions with Z-scores higher than 8 can be considered as correct. For each run (Sm core 1-3) all components, except the first Sm core model that was placed, were found with relatively high Z-scores of 10-20, indicating correct positioning in the unit cell. Best scores were obtained with Sm core model 3, followed by Sm core model 1 and Sm core model 2. In each case the components were placed into the unit cell in an identical fashion. For models 1 and 2 one unique solution was found, the poly-serine model yielded several solutions. These solutions comprised rotations around the seven-fold axis of the homo-heptameric Sm core. The energy minimized model of the Sm core was placed with a lower confidence than the residual models. In the following the energy-minimized Sm hetero-heptamer (model 2) was not used further.

Figure 3.18 depicts Sm core model 1 and U1-A complexed with RNA positioned in the unit cell by Phaser. In line with the self-rotation function and the estimation of the unit cell content (see 3.1.11), the molecular replacement solution confirmed the assumption of two molecules in the asymmetric unit. The molecular weight of the truncated particles was estimated to be 360 kDa per asymmetric unit. The search models that were placed by phaser comprised around about 50 % of the estimated protein and nucleic acid content of the unit cell. The crystal packing of the molecules appeared reasonable but comprised large solvent channels. These solvent channels might be bridged by U1 snRNP components that were not found by MR.

To validate the molecular replacement solution, the rotational operators of the NCS were deduced from the self rotation function and compared to those of the MR-model. The rotational and translational NCS-operators of the molecular replacement solution were calculated with the program TOP of the CCP4i suite. The rotation angles derived of
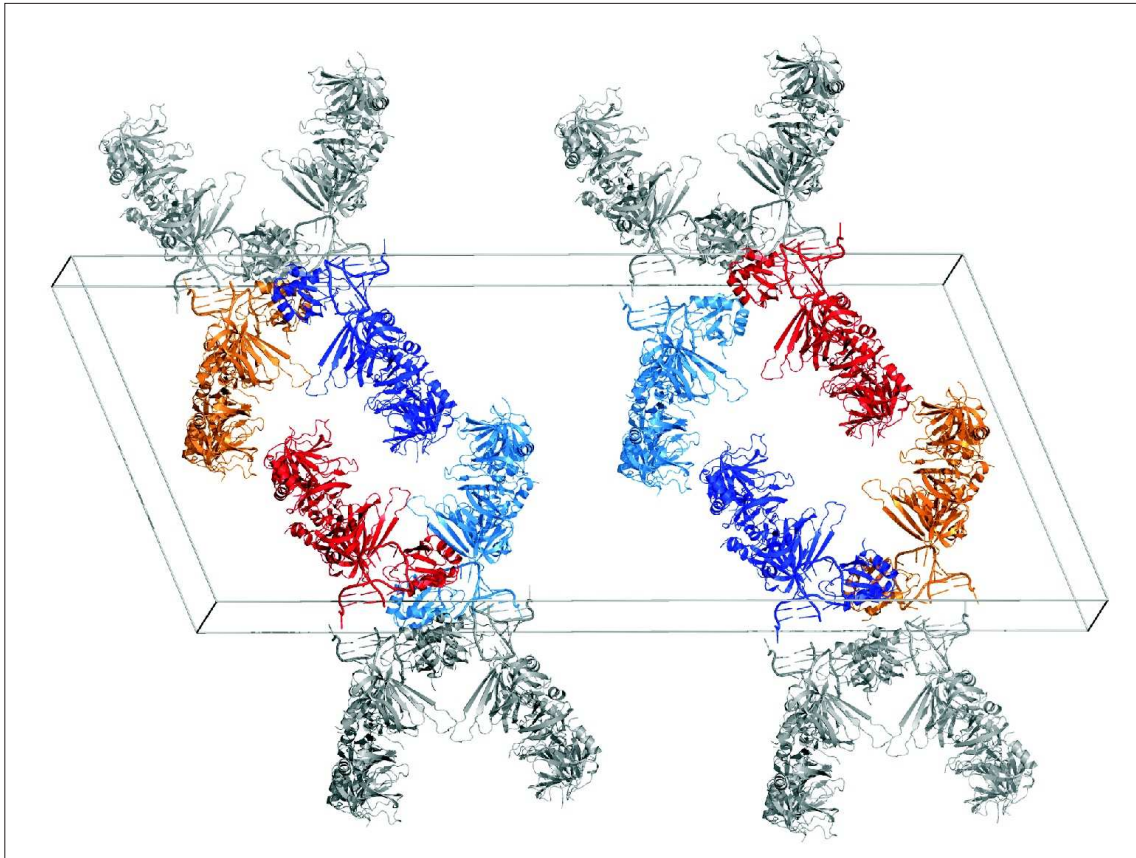
Figure 3.18 Crystal packing and unit cell content of a potential molecular replacement solution. A molecular replacement was carried out with Phaser, starting with a search of two Sm hetero-heptamers, followed by U1-A complexed with RNA. All components were placed in the unit cell with a relatively high Z-score. NCS-related particles are colored red and orange, the crystallographically related ones blue and cyan. The entire C2 unit cell is visualized as a gray box.

the self rotation function and the MR-solution were identical, which strongly supported the molecular replacement model. Furthermore the placement of U1-A and the RNA relative to the Sm core proteins fulfilled the expected NCS.

As shown in figure 3.19 phasing by molecular replacement with models generated maps that were restricted to the model that was used for replacement. No additional electron density for either RNA or protein components was recognizable.

To improve the map that was generated by molecular replacement with the Sm core and U1-A/RNA, other structurally known components were searched on top of the actual solution with phaser. A homology model of U1-70k RRM was generated with the online tool HHpred (http://toolkit.tuebingen.mpg.de/hhpred/) with U1-A (PDB accession code 1URN) as a template. The homology model of the U1-70k RRM could not be placed by
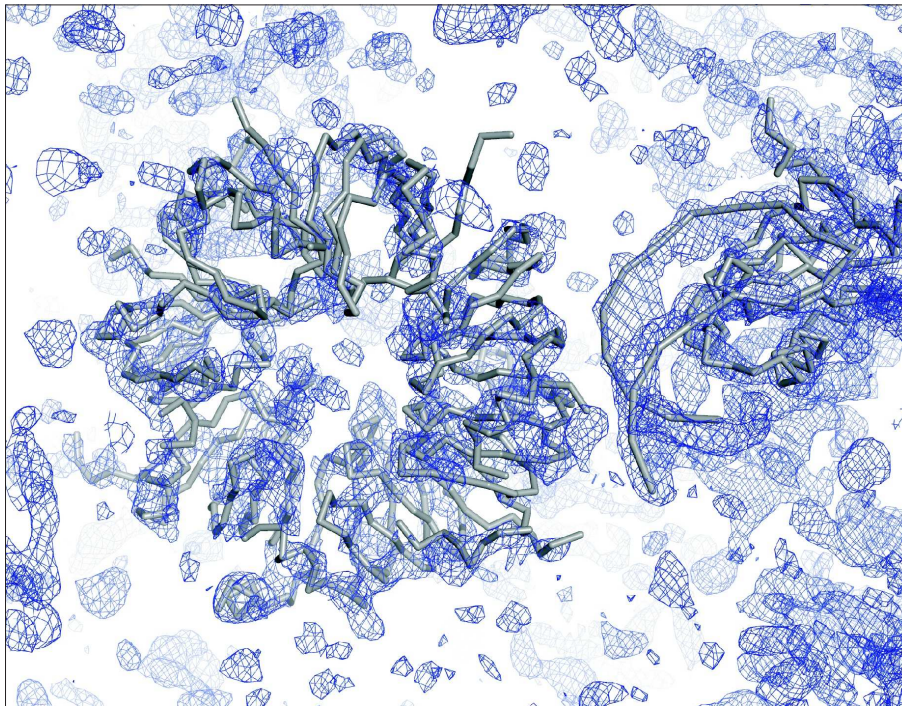
Figure 3.19 Electron density of the potential molecular replacement solution. One Sm heptamer (gray ribbons) and U1-A with U1 snRNA stem loop II (gray ribbons) is shown with the corresponding map, calculated from weighted structure factors and the molecular replacement phases. The map (blue) was displayed by Coot and was contoured at 2 σ.

MR. Also the first ensemble of the NMR-structure of U1-C was employed as a search model (Muto *et al.*, 2004) but was not placed by a MR-search on top of the partial solution.

U1 snRNA was shown to be a component of the truncated U1 snRNP, that crystallized, and it is available as a three-dimensional *in silico* model (Krol *et al.*, 1990). Stem loops and the four-way junction were extracted from the U1 snRNA model and used for a molecular replacement search. These molecules yielded no solutions (not shown).

Other programs than Phaser are available for molecular replacement. The CCP4-related programs Molrep and Amore did not find a solution with any of the search models employed.

The density that was generated by initial molecular replacement searches was restricted to the input models, placed by Phaser. The residual components of U1 snRNP were not identified in the map. Different refinement strategies were applied to the molecular replacement solution. In all refinement protocols the NCS was used. First, a standard restrained refinement with phenix.refine was carried out. Additional density did not appear. Likewise a rigid body refinement, defining each chain as a NCS-restrained rigid

body or defining larger portions like the Sm core as NCS-restrained rigid bodies, did not yield any additional density for the lacking components of the particle. To validate a model, an omit map can be calculated. When a single chain, like SmF, was omitted from refinement, also the corresponding density entirely disappeared.

3.1.12.2 Derivatization of U1$_{DNA4}$ co-crystals with heavy atom compounds

Although a potential molecular replacement solution for parts of U1 snRNP was obtained, phase information was not sufficient for the interpretation of the lacking components. Therefore experimental phasing with heavy atom clusters as well as with osmium and iridium hexammines was tried.

Heavy atom clusters were shown to provide sufficient phasing power also at lower resolution, which has been demonstrated for the ribosome (Weinstein and Steitz, 1999). Since U1 snRNP contains a significant portion of RNA osmium and iridium hexammine were used for derivatization. These compounds bind to GU-wobble sites (Cate and Doudna, 1996), of which four are present in U1 snRNA.

The initial crystallization condition for U1$_{DNA4}$ crystals contained the chelators EDTA and citrate, which might interfere with heavy atom cluster derivatization. When the presence of EDTA and citrate during derivatization was avoided, the longest axis of the derivatized crystals shrunk by around $2-3$ % and the β-angle was expanded by 3 - 4 degree. The influence of buffer devoid of EDTA and citrate on the crystals has not been tested yet and could not be excluded as a reason for a change of cell constants. Even brief soaking with derivatization buffer led to an immediate change of the unit cell parameters, as described above.

Table 3.3 summarizes a selection of datasets collected from derivatized U1$_{DNA4}$ co-crystals. Crystals were derivatized according to 2.2.4.4.

Table 3.3 Derivatization of U1 snRNP crystals with heavy atom compounds. The list comprises a selection of datasets collected from derivatized U1$_{DNA4}$ crystals. All datasets listed, except the native, were collected at SLS/PX2 covering a total $\Delta\phi$ range of 360°. The beam was attenuated to $5-50$ % of the total intensity available. The degree of attenuation was set to a value which allowed indexing and processing of the reflections. All heavy atom compounds used are listed in 2.2 and 2.5. St. refers to the soaking time, mM indicates the concentration of the heavy atom compound, $\Delta\phi$ is the rotation angle per frame and λ is the wavelength. Deriv. stands for the derivative name and d for the resolution, where I/σ dropped below 2. All derivative datasets were processed and scaled as described in 2.2.4.5.

No.	Deriv.	c [mM]	St.	$\Delta\phi$ [°]	λ [Å]	a [Å]	b [Å]	c [Å]	β [°]	d [Å]	Remarks
1	native	-	-	0.1	0.9999	359.09	88.38	151.19	111.87	4.5	
2	W40	0.05	2 h	2	1.2145	344.20	88.65	152.08	115.75	8.3	inflection
3	W40	0.05	2 h	2	1.2141	346.52	88.65	152.75	115.75	8.3	
4	W40	0.05	2 h	2	1.2023	346.67	88.75	152.56	115.80	8.3	high remote
5	W30	0.2	5 d	2	1.2146	346.67	89.22	153.48	115.80	8.0	inflection
6	W30	0.2	5 d	2	1.2140	346.61	89.15	153.46	115.83	8.5	
7	W30	0.2	5 d	2	1.2022	346.73	89.38	153.68	115.79	8.8	high remote
8	W5	0.05	5 d	2	1.2140	350.35	88.14	153.87	115.98	8.0	
9	W6	0.05	5 d	2	1.2140	344.19	88.36	150.81	116.20	8.5	
10	W3	0.4	5 d	2	1.2140	345.79	89.00	153.96	115.81	7.8	
11	W30	0.4	2 h	2	1.2140	345.25	88.80	151.71	115.79	10.0	
12	W5	0.05	2 h	2	1.2140	346.54	88.42	154.30	115.80	10.5	
13	W6	0.05	2 h	2	1.2146	343.96	88.56	152.19	115.58	8.5	inflection
14	W6	0.05	2 h	2	1.2140	344.06	88.54	151.97	115.6	8.5	
15	W6	0.05	2 h	2	1.2023	344.08	88.57	152.36	115.58	8.5	high remote
16	W3	0.4	2 h	2	1.2140	342.90	89.36	150.21	115.79	8.5	
17	W18	0.4	2 h	3	1.2146	345.46	88.72	152.67	115.76	8.6	inflection
18	W18	0.4	2 h	3	1.2140	347.53	89.08	153.96	115.52	8.6	
19	W18	0.4	2 h	3	1.2023	345.45	88.76	152.79	115.78	8.6	high remote
20	W17	0.4	2 h	3	1.2146	347.04	88.54	152.34	115.93	9.0	inflection

Table 3.3 Derivatization of U1 snRNP crystals with heavy atom compounds continued ...

No.	Deriv.	c [mM]	St.	$\Delta\phi$ [°]	λ [Å]	a [Å]	b [Å]	c [Å]	β [°]	d [Å]	Remarks
21	W17	0.4	2 h	3	1.2140	347.05	88.79	152.32	115.94	8.6	
22	W17	0.4	2 h	3	1.2023	345.45	88.76	152.79	115.78	8.6	high remote
23	TAMM	0.05	0.5 h	3	1.0106	342.17	88.24	152.05	115.05	9.0	
24	Ir3	0.4	5 d	3	1.1033	347.46	89.69	152.59	116.53	9.3	
25	Ir3	0.4	5 d	3	1.1033	346.34	88.95	152.09	115.99	8.0	
26	Mo7	0.4	5 d	3	1.1033	343.12	89.63	150.74	115.97	8.0	high remote
27	Mo7	0.4	5 d	3	1.1033	343.91	89.74	151.28	116.09	7.6	high remote
28	Oshex	0.5	5 d	2	1.1384	361.02	86.65	150.74	113.07	7.3	
29	Oshex	0.5	5 d	2	1.1364	360.98	86.65	151.03	113.13	7.0	high remote
30	Irhex	0.5	5 d	2	1.1056	360.09	86.91	150.45	112.7	6.5	

Derivatization with heavy atom clusters (see table 3.3) clearly shrank the a-axis of the unit cell and expanded the β-angle from 111 to 115 degrees. The resolution dropped from $4.5 - 5$ Å to around $8 - 9$ Å. Still datasets at low resolution could be collected and used for cross-validation with other datasets and heavy atom searches. In the case of non-isomorphism, single isomorphous or multiple isomorphous replacement phasing (SIR and MIR, respectively) cannot be employed.

The presence of iridium and osmium hexammines changed the cell constants of $U1_{DNA4}$ co-crystals. While the length of the b-axis slightly shrank, the β-angle expanded from 111 to 113 degrees. Resolution also dropped to $6 - 7$ Å. Datasets were collected at $6 - 7$ Å, interacted and cross-validated with other datasets.

A cross-comparison of the structure factors of all datasets collected was carried out with the CCP4i program FHSCAL, as shown in table 3.4. The program scales the datasets and cross-compares the difference of the structure factors, expressed as an R-factor. When comparing datasets, a slightly increased R-factor might indicate heavy atom binding. Larger R-factor differences can be either attributed to non-isomorphism or may also indicate binding of a heavy atom compound.

Initially several datasets like W40 or W5 did not scale to the other derivative data, although they had matching cell constants (see dataset W5* in table 3.4). The self rotation functions calculated of these datasets also deviated from the one shown in figure 3.17 on page 84 and were indicative for these datasets. In space group C2 different choices for the unit cell are possible, therefore tungsten derivative datasets that did not match to the others like W40, W5, Ir3 and Mo7 were reindexed with

$$h_{(new)} = h_{(old)} + 2l_{(old)}$$

$$k_{(new)} = -k_{(old)}$$

$$l_{(new)} = -l_{(old)}.$$

After reindexing, all space group settings matched to the native dataset. The structure factor differences, expressed by the R-factor dropped to a degree, where all tungsten derivative crystals could be considered isomorphous. Only the W5 cluster scaled to the other tungsten derivatives with a higher R-factor. This might indicate the presence of a bound heavy atom cluster.

Table 3.4 Isomorphous differences of U1$_{DNA4}$ native and derivative datasets. Statistics of selected derivative datasets are listed in table 3.3, datasets 3, 6, 8, 9, 10, 18, 21, 23, 25, 27, 28 and 30. The datasets listed in both tables were scaled to a native U1$_{DNA4}$ dataset with FHSCAL using standard parameters, a standard setting of the unit cell and a resolution range from $4 - 164$ Å. The table lists the isomorphous difference of the datasets, expressed by the R-factor of the corresponding structure factors. The last line (W5*) represents a W5 dataset that was not reindexed.

	native	Irhex	Oshex	Ir3	Mo7	W5	W3	W30	TAMM
native		0.227	0.266	0.509	0.563	0.587	0.553	0.526	0.499
Irhex	0.227		0.151	0.467	0.518	0.544	0.501	0.476	0.452
Oshex	0.266	0.151		0.466	0.513	0.535	0.490	0.471	0.437
Ir3	0.509	0.467	0.466		0.221	0.409	0.194	0.186	0.247
Mo7	0.563	0.518	0.513	0.221		0.415	0.212	0.210	0.299
W5	0.587	0.544	0.535	0.409	0.415		0.340	0.366	0.413
W3	0.553	0.501	0.490	0.194	0.212	0.340		0.148	0.258
W30	0.526	0.476	0.471	0.186	0.210	0.366	0.148		0.220
TAMM	0.499	0.452	0.437	0.247	0.299	0.413	0.258	0.220	

	native	W3	W5	W6	W17	W18	W30	W40	Irhex
native		0.553	0.587	0.518	0.514	0.507	0.526	0.509	0.227
W3	0.553		0.345	0.181	0.173	0.159	0.148	0.146	0.465
W5	0.587	0.345		0.375	0.345	0.360	0.366	0.368	0.493
W6	0.518	0.181	0.375		0.201	0.151	0.168	0.141	0.448
W17	0.514	0.173	0.345	0.201		0.154	0.158	0.172	0.437
W18	0.507	0.159	0.360	0.151	0.154		0.145	0.110	0.443
W30	0.526	0.148	0.366	0.168	0.158	0.145		0.126	0.446
W40	0.509	0.146	0.368	0.141	0.172	0.110	0.126		0.453
Irhex	0.227	0.465	0.493	0.448	0.437	0.443	0.446	0.453	
W5*	0.559	0.568	-	0.395	0.581	0.570	0.582	0.385	0.661

In contrast to the heavy atom cluster compounds, the addition of hexammines to the crystals did not significantly alter cell constants which was also mirrored by the low R-factors. This implied a high degree of isomorphism, rendering osmium and iridium hexammines suitable for isomorphous replacement experiments, also using anomalous signals.

Several programs for heavy atom searches are available. All datasets were searched for heavy atoms with hkl2map (ShelX E) or Solve/Autosolve. Of all datasets anomalous difference Patterson maps were calculated. In case of isomorphism, difference Patterson maps of pairs of derivative and native datasets were calculated and evaluated. Fur-

thermore automated structure solution with Autosharp was tried, supplying unscaled datasets. A W5 heavy atom substructure with four sites was found with hkl2map. The search of the residual derivatized datasets did not result in the localization heavy atom sites.

3.1.12.3 Localization of W5-sites in U1$_{DNA4}$ derivative co-crystals

Of all heavy atom clusters tested, only W5 derivative U1$_{DNA4}$ co-crystals looked promising since a solution with hkl2map was found, using the SAD method.
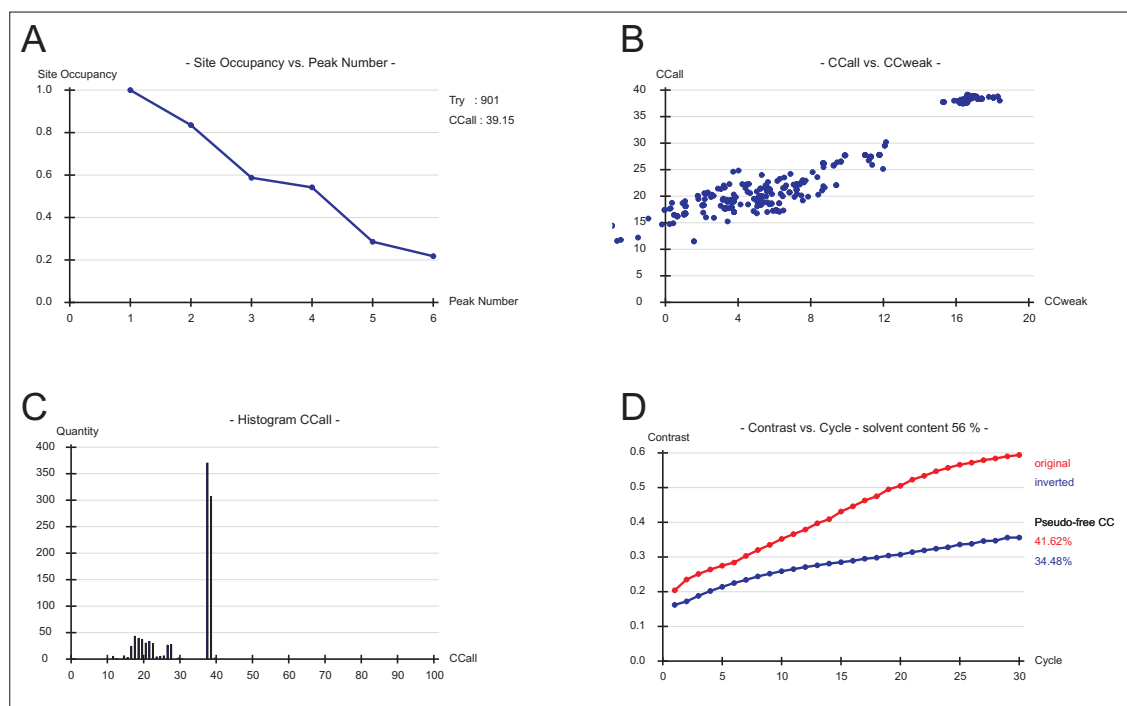
Figure 3.20 ShelX statistics obtained from a W5-derivative U1$_{DNA4}$ crystal, using SAD. Data to 8.5 Å resolution were included. **A** The plot of heavy atom site occupancies (peak-hight) versus peak number suggests 4 heavy atom sites with high occupancies (0.5-1.0). The plot of CCall versus CCweak ((B)) and a histogram of the quantity of CCall (**C**) suggest a clear solution. **D** Plot of contrast versus density modification cycle for the two enantiomorphs. Density modification with a solvent content of 58 % identified the original hand (blue) to have a higher contrast than the inverted (red).

Figure 3.20 shows the statistics of heavy atom search, phasing and density modification, indicating a solution for four heavy atom sites. As shown in table 3.4 isomorphism of W5 to other datasets was low, and likely therefore, methods like MIR or SIR with a native dataset as a reference were not successful.

The W5-sites that were found were cross-validated with the molecular replacement
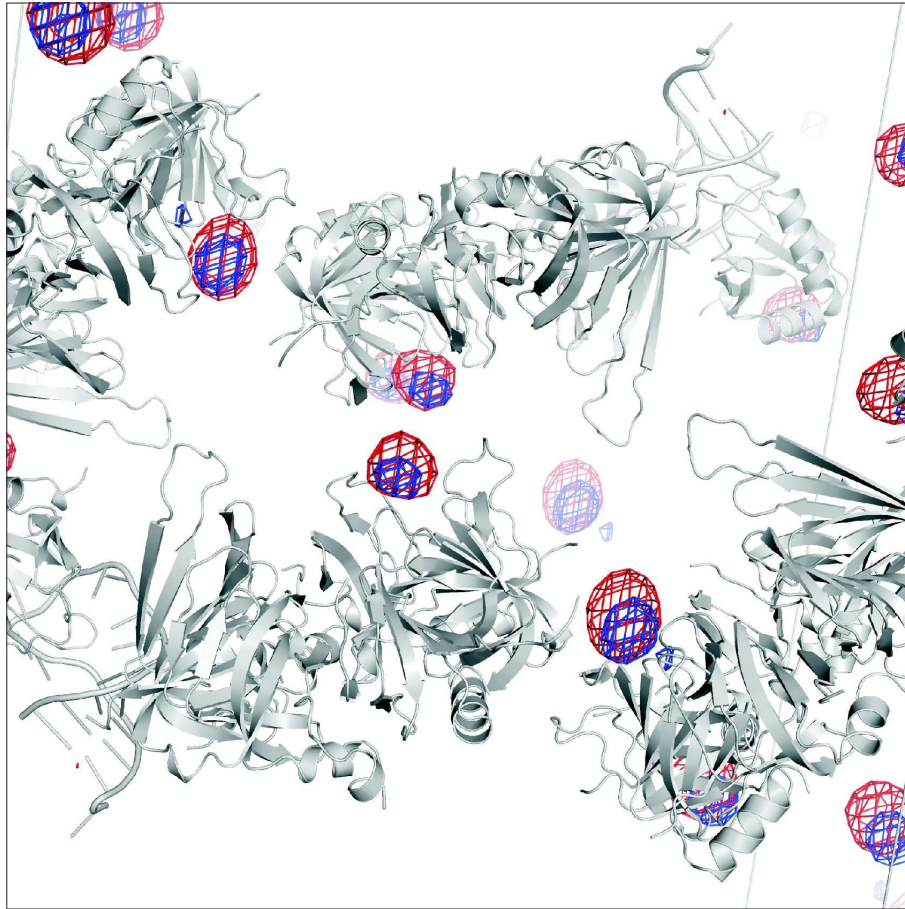
Figure 3.21 Comparison of W5 derivative sites obtained by molecular replacement and SAD phasing. Phases (PHIC) from a molecular replacement solution and the anomalous differences (DANO) from the W5 derivative dataset were used to calculate a phased anomalous difference map with FFT (shown in blue, contoured at 4 σ). A Patterson-based heavy atom search and phasing of the W5 derivative data with hkl2map yielded four W5 sites per asymmetric unit. A map was calculated from these experimentally determined phases and the anomalous differences of the W5 derivative data (shown in red, contoured at 5 σ). All peaks that were found in both maps overlapped with a slight offset along the b axis. Two peaks were related by the two-fold NCS that was found by molecular replacement. The molecular replacement solution is shown in gray.

solution. First, the W5 derivative dataset was used for molecular replacement. Therefore a model, derived from the initial molecular replacement solution (see 3.1.12.1) was employed. A model of one molecule of U1-A complexed to the RNA, and the proximate Sm core was sufficient for molecular replacement (see figure 3.19). The components were placed in the unit cell in an identical fashion like for the native data. The phases (PHIC) from the molecular replacement solution of the W5 dataset and the anomalous difference structure factors (DANO) from the W5 derivative dataset were

used to calculate a phased anomalous difference map with FFT (CCP4i suite). This MR-phased map showed 4 peaks, or potential W5 sites, per asymmetric unit. Next, the experimental phases, obtained from shelxE density modification with the original hand (see figure 3.20), were used to calculate a map with DANO of the W5 derivative data with FFT. Both maps, calculated with phases either originating from molecular replacement or from an experimental phasing approach and DANO of the W5 data were compared, as shown in figure 3.21.

A total overlap of the major peaks cross-validated the experimentally phased substructure as well as the molecular replacement solution. The molecular replacement solution was further corroborated by the fact that at least two of the sites found in the asymmetric unit were related by NCS. Apart from hkl2map, other programs, like Sharp/Autosharp were employed for phasing. SAD of the W5-derivative or SIR(AS) with other tungsten derivatives did not produce interpretable electron density maps. Density modification with the programs Solomon or DM was not successful.

3.1.12.4 Localization of iridium and osmium hexammine-sites in U1$_{DNA4}$ derivative crystals

U1 snRNP has potential binding sites for osmium or iridium hexammine. A derivatization of U1$_{DNA}$ crystals with these compounds might provide sufficient phase information to solve the structure of truncated U1 snRNP. As shown in table 3.3, datasets of U1$_{DNA4}$ co-crystals, derivatized with iridium and osmium hexammines, were taken at SLS/PX2 at around 7 Å resolution. These crystals were considered isomorphous to the native ones, according to cell constants and R-factors (see table 3.4). Heavy atom searches by difference Patterson methods or automated phasing programs like Sharp/Autosharp did not result in the localization of heavy atom sites. Likewise it was not successful to employ one of the derivative datasets as a 'pseudo native' for SIR or SIRAS in automated phasing programs. Like previously described for the W5 derivative, the phases derived from the molecular replacement solution of the native data were used to search for potential heavy atom sites.

Both derivative datasets were scaled to a native dataset (see table 3.4), which contained the phases of the molecular replacement solution. In the next step a MR-phased isomorphous difference map (MR phased SIR) was calculated of the native data and the W5 derivative data with FFT. This map contained five peaks contoured higher than 4.5 σ in the case of osmium hexammine and nine peaks higher than 4.5 σ in the case of iridium hexammine. Nearly all of these potential sites of the iridium compound matched to

those of the osmium compound and vice versa. Though it has to be kept in mind that these sites could be an artifact introduced by the molecular replacement phases.

The model derived from the molecular replacement solution was superimposed with the MR-phased SIR map. The most prominent Ir/Os-sites were located to a potential Os/Ir hexammine binding site in stem loop II of U1 snRNA. This supported a correct positioning of the molecule by molecular replacement. Figure 3.22 shows the RNA hairpin that is complexed with U1-A as well as the two prominent potential osmium hexammine and iridium hexammine sites. The osmium and iridium hexammine sites were related by the same NCS, like the W5 derivative sites and the molecular replacement model.

All sites that were found in both hexammine derivative datasets were used for phasing with the program Mlphare. Isomorphous replacement combined with anomalous scattering (MIRAS) yielded the best phasing statistics. The phasing protocol was followed
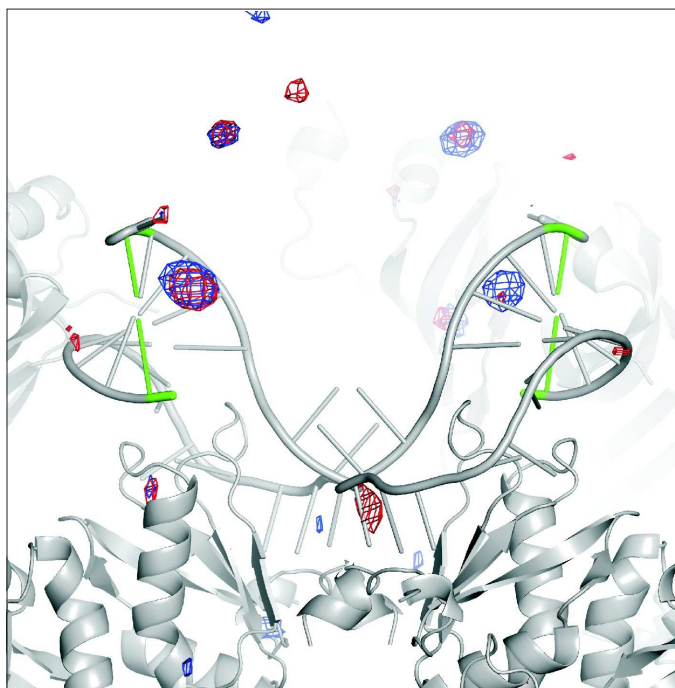
Figure 3.22 Osmium and iridium hexammine sites in the potential vicinity of a GU wobble basepair, found in MR-phased difference Fourier maps. Osmium and iridium hexammine derivative datasets were scaled to a native dataset (see table 3.4) which contained the phases of the molecular replacement solution. In the next step, a phased difference map was calculated of the native data and the hexammine derivative data with FFT. The difference density of the osmium hexammine derivative is shown in red, the difference density of iridium hexammine is shown in blue. Both maps are contoured at 4 σ. The structure depicts two NCS-related molecules of the U1-A protein, complexed to stem loop II of U1 snRNA as placed by molecular replacement with Phaser (shown in gray). The GU-wobble basepair in U1 snRNA (here a GA basepair derived from PDB file 1URN) is colored green.

by density modification and NCS-averaging by DM. The matrix for NCS-averaging was obtained from the molecular replacement solution (see also 3.1.11). However, an interpretable map could not be obtained.

3.2 U4/U6·U5 tri-snRNP

3.2.1 Purification of *Homo sapiens* U4/U6·U5 tri-snRNP for crystallization

U4/U6·U5 tri-snRNP is a challenging crystallization target among the snRNPs. A crystal structure of tri-snRNP would yield great insight into spliceosome architecture and could be combined with electron microscopic data (Sander *et al*., 2006). The particle has a size of roughly 1.5 MDa and is not stable in glycerol-free solution. Tri-snRNP particles are sensitive to membranes and disintegrate at salt concentrations higher than 450 mM NaCl. The glycerol gradient fractions that contain tri-snRNP have a protein concentration of less than 0.3 mg/ml. The particle had to be stabilized and concentrated to setup initial crystallization trials.

The U4/U6·U5 tri-snRNP was purified and separated from other snRNPs by glycerol gradient centrifugation (see 3.2, 3.1). As mentioned in 3.1.1.1, tri-snRNP was not entirely separated from U5 snRNP. Various concentration protocols, like ultrafiltration or pelleting, were tried to obtain particles suitable for crystallization. These strategies resulted in a loss of nearly all input material. As an alternative, a protocol based on ammonium sulphate precipitation was developed. Buffer conditions and ammonium sulphate requirement were tested in a smaller scale beforehand (not shown). In the following a preparative protocol is described.

For tri-snRNP preparation, 7.2 mg of tri-snRNP with an average concentration of 0.09 mg/ml from two glycerol gradients were pooled. The pooled fractions were concentrated by ammonium sulfate precipitation according to 2.2.3.9. It was avoided to include fractions containing the 52k protein, which is a marker for U5 snRNP.

After ammonium sulfate precipitation, it was necessary to include 100 mM NaCl and 50 mM ammonium sulfate to the dialysis buffer. If ammonium sulfate was omitted, the particles precipitated. 450 µl of tri-snRNP with a concentration of 9 mg/ml were obtained from dialysis. This represented 56 % of the input material. A major fraction of this preparation was used for crystallization trials at 4 °C. In total 864 conditions in
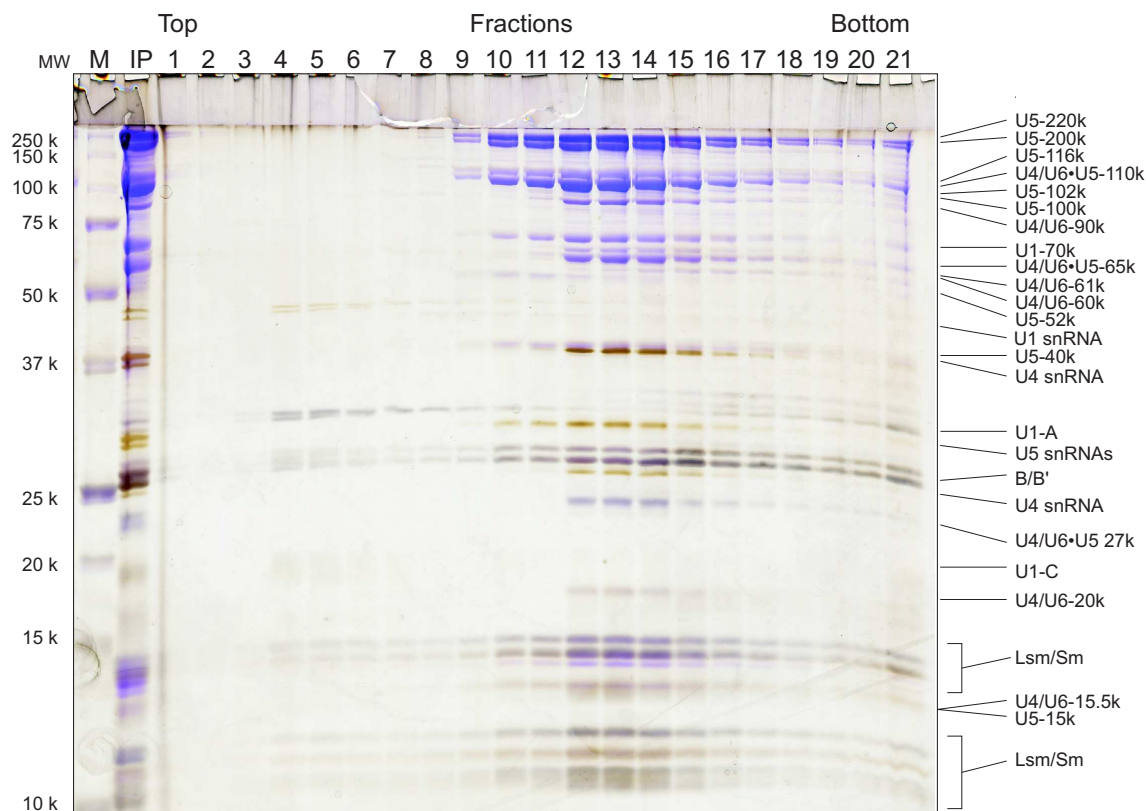
Figure 3.23 Glycerol gradient analysis of ammonium sulfate-precipitated U4/U6·U5 tri-snRNP. 28 µl of U4/U6·U5 tri-snRNP with a concentration of 4 mg/ml were applied on a 4.4 ml $10 - 30$ % glycerol gradient in crystallization buffer AS. The gradient was run at 25 000 rpm for 17 h. The gradient was harvested manually in 225 µl fractions (1-21), which were acetone precipitated and applied on an 13.5 % SDS-gel, next to 10 µl of the input (IP) material. Proteins were visualized by Coomassie stain, RNA was stained by silver stain. The molecular weight (MW) of the marker (M) proteins is given on the left side of the gel. Identities of the tri-snRNP proteins are given on the right side of the gel.

the absence and 576 conditions in the presence of proteases were screened (see 2.11 on page 48, screen 2-5). Crystal growth was not observed so far.

Next it was assessed, if U4/U6·U5 tri-snRNP disintegrated upon the ammonium sulfate treatment and dialysis. Therefore an analytical glycerol gradient was run in crystallization buffer AS at 25 000 rpm for 17 h, using material from an analogous small-scale concentration. In an ammonium sulfate-free gradient that was run as a control U4/U6·U5 tri-snRNP was entirely pelleted, most likely due to aggregation or precipitation (not shown). The presence of 50 mM ammonium sulfate was required for tri-snRNP stability after ammonium sulfate precipitation. In Figure 3.23 the SDS-gel of the glycerol gradient fractions is shown. The tri-snRNP clearly migrated as one particle in the gradient in fractions 12 to 16, a minor fraction was found in the pellet. All particle-

specific proteins as well as the three RNAs were present in these fractions. Apart from 25S tri-snRNP a contamination of U5 snRNP was present in fractions 9 - 12. This contamination was not present in larger preparations that were used for crystallization.

3.2.2 Limited proteolysis of U4/U6·U5 tri-snRNP

It is yet unclear whether U4/U6·U5 tri-snRNP comprises flexible domains and extensions which might prevent crystallization. A global flexibility of protein extensions loops or entire domains cannot be excluded (Sander *et al.*, 2006).

For individual proteins it can only be assumed that for example RS-domains like in 100k or low complexity regions are unstructured. To find suitable proteases for pre-crystallization treatment or *in situ* proteolysis, U4/U6·U5 tri-snRNP, purified and con-
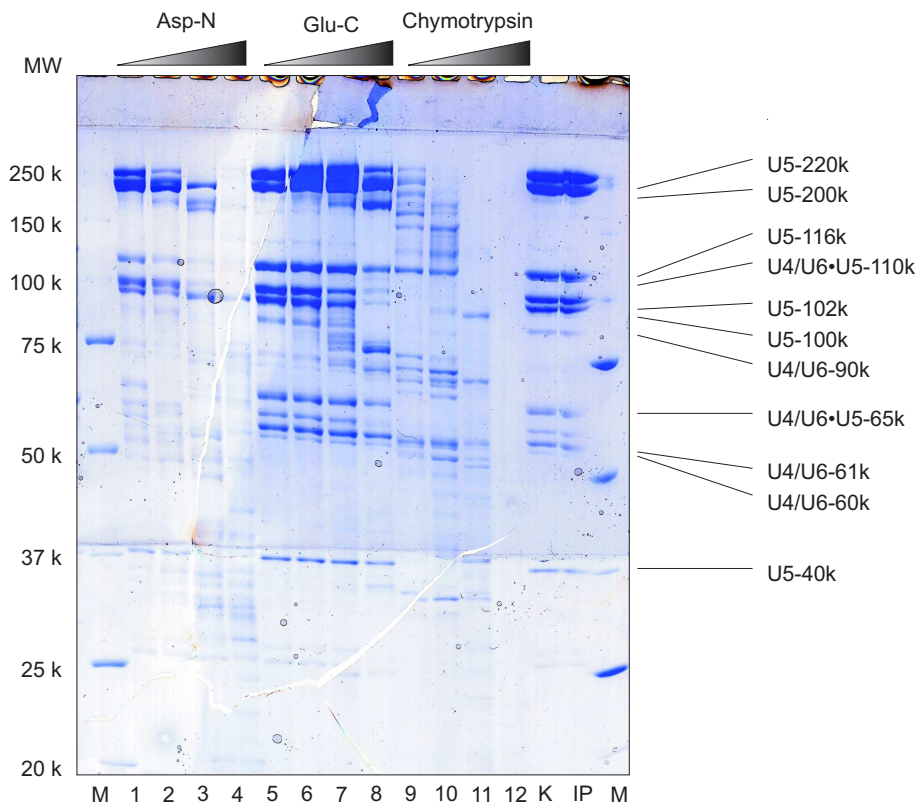
Figure 3.24 Limited proteolysis of U4/U6·U5 tri-snRNP. 23 μg of tri-snRNP (IP) were subjected to limited proteolysis. Lane K denotes the control, which was incubated in the absence of proteases. 23 μg of tri-snRNP per reaction were incubated in a series of ten-fold dilutions of Asp-N from 0.0001 μg (lane 1) to 0.1 μg (lane 4), Glu-C from 0.0005 μg (lane 5) to 0.5 μg (lane 8) and chymotrypsin from 0.001 μg (lane 9) to 1 μg (lane 12). The proteins were visualized by Coomassie stain. The molecular weight (MW) of the marker (M) proteins is given on the left side of the gel. Identities of the tri-snRNP proteins are given on the right side of the gel.

centrated (see 3.23), was treated with proteases. Samples were subjected to a limited proteolysis experiment with chymotrypsin, Asp-N and Glu-C in a series of ten-fold dilutions with 23 µg of tri-snRNP per reaction. The reaction was incubated at 4 °C for 12 h and stopped by the addition of SDS-loading dye. The samples were visualized on a 10 % SDS-gel (see figure 3.24).

In this experiment the applicability of several proteases for the treatment of tri-snRNP was judged by the pattern of proteolysis. Asp-N degraded all larger proteins of the particle, only fragments corresponding to 100k and 200k seemed to be stable (lanes 1-3) but disappeared in lane 4.

A proteolysis by Glu-C resulted in a truncation of 220k, whereas other proteins like 200k and 116k or 110k seemed to be stable. In the range of 60k, the 60k and 65k proteins were not affected by the protease. On the other hand two major fragments appeared. The band below 200k (lane 9) could originate from 220k, since the largest protein was degraded to a major degree, whereas 200k seems unaffected by the protease. A second stable fragment appeared at 75k with unknown origin. The pattern which resulted from chymotrypsin proteolysis, did not contain any stable fragments and suggests a high sensitivity of all tri-snRNP proteins to chymotrypsin - in contrast to U1 snRNP.

3.3 SF3a

3.3.1 Purification of *Homo sapiens* SF3a for crystallization

SF3a is associated with U2 snRNP in the cell. The particle comprises three proteins and can be isolated from nuclear extract by immunoaffinity chromatography in a one-step procedure (Dybkov *et al*., 2006). The 17S U2 snRNP has been studied by electron microscopy, which allowed a global assignment of its additional large substructures SF3a and SF3b in relation to 12S U2 snRNP (Krämer *et al*., 1999). Substructures of SF3a have been solved by NMR, which comprise a relatively small fraction of SF3a in total (Kuwasako *et al*., 2006). A structure of the entire SF3a states an objective of this work. The U2-associated particle SF3a was natively purified from HeLa cells by immunoaffinity chromatography and gel filtration as described (2.2.3.4, 2.2.3.8). Figure 3.25 shows an immunoaffinity column run, visualized by SDS-PAGE, which has been performed several times to obtain enough material for further purification and crystallization trials. The input (IP) material was applied next to the flowthrough (FT) of the SF3a-120 immunoaffinity column . All fractions in the elution contain SF3a, composed of SF3a-60, SF3a-66 and SF3a-120. Due to the presence of the peptide it was not possible to
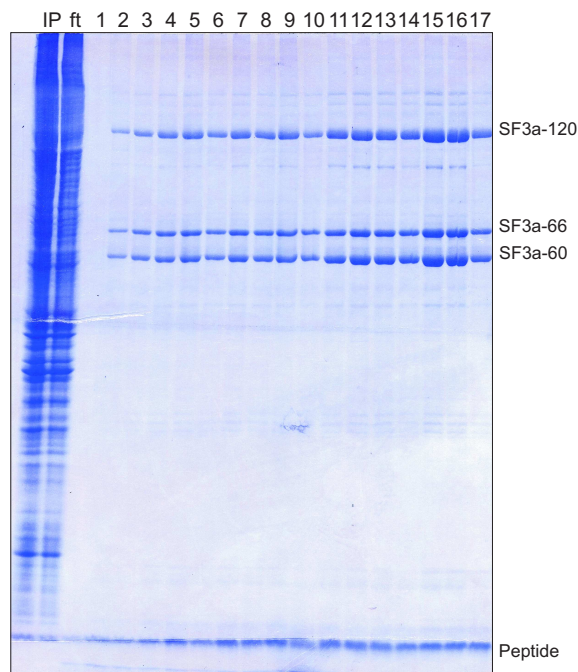
Figure 3.25 Immunoaffinity purification of SF3a from HeLa cells. SF3a was purified by immunoaffinity chromatography as described in 2.2.3.4 from HeLa nuclear extract which was passed over a H20 immunoaffinity column. 25 µl of the input (IP) and the flowthrough (ft) were applied on a 13.5 % SDS gel, next to 50 µl of each 1 ml fraction which was collected during elution (lanes 1-17). The proteins were visualized by Coomassie stain. The molecular weight (MW) of the marker (M) proteins is given on the left side of the gel. Identities of the SF3a proteins are given on the right side of the gel.

determine the yield of one run. For preparative scale, fractions of four immunoaffinity column runs were pooled and concentrated to 2.9 mg/ml. In total 1.4 mg of SF3a were subjected to gel filtration with a Superose 6 column as described in 2.2.3.8. The input and fractions of the gel filtration run are shown in figure 3.26.

The peptide used for the elution of the particle in the immunoaffinity step was the only major contaminant and migrated in fractions 31 and 32, clearly separated from the particle. Fractions 16 - 21 contained SF3a and were pooled. SF3a was finally concentrated to around 12 mg/ml for crystallization screening in a Vivaspin concentrator (MWCO 30000). The volume of concentrated material obtained was 150 µl used for screening of 576 conditions (see table 2.11 on page 48). Crystal growth has been not observed so far.
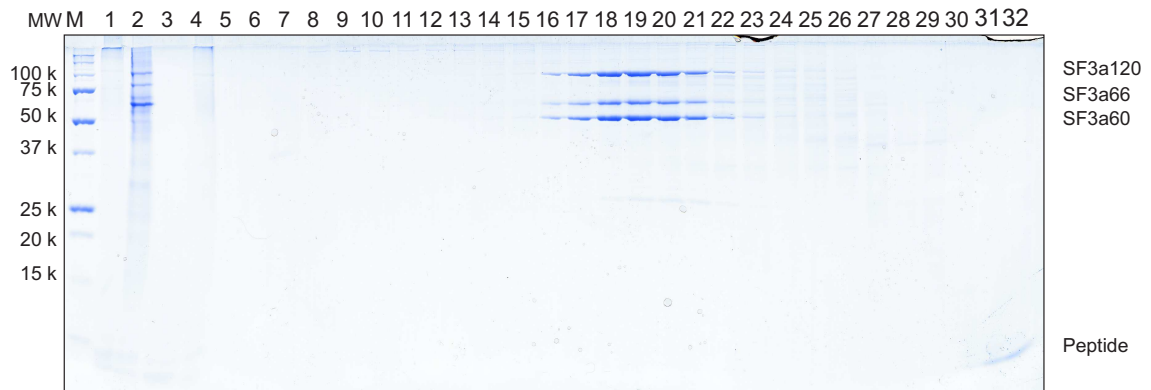
Figure 3.26 Gel filtration of purified Sf3a. 4 mg of SF3a isolated from HeLa cells were applied on a Superose 6 XK 16/70 gel filtration column and run with 0.2 ml/min of a buffer containing 20 mM Tris-HCl, pH 7.9, 150 mM NaCl, 1 mM DTT. 10 µl of the input material (lane 1) is applied on a 13.5 % SDS-gel next to 5.8 µg of the concentrated gel filtration input (lane 2). In lane 3-32, 20 µl of the 1 ml fractions are applied. Proteins were stained with Coomassie brilliant blue. The molecular weight (MW) of the marker (M) proteins is given on the left side of the gel. Identities of the SF3a proteins are given on the right side of the gel.

4 Discussion

SnRNPs function as key players to form a complex and dynamic machinery, termed the spliceosome. This multi-megadalton assembly, and even its main subunits, the snRNPs, are of comparable size to the bacterial ribosome subunits or RNA polymerase 2. The structures of the bacterial ribosome as well as of RNA polymerase 2 have been solved and yielded detailed insight in key cellular processes (Ban *et al*., 1998, 2000; Cramer *et al*., 2000). The generation of these structures required a tremendous effort in improving purification as well as crystallization protocols and crystallographic data collection techniques over decades (Yonath *et al*., 1998). During this work all aspects of purification, crystallization and phasing of the macromolecular assemblies were addressed.

Crystals of native HeLa U1 snRNPs were available at the beginning of this work. These crystals diffracted X-rays to 25 Å and grew in an irreproducible fashion (personal communication B. Kastner, M. Wahl). The amounts of U1 snRNP particles were not sufficient for a high throughput crystal screening approach. To address this fundamental prerequisite, purification strategies for U1 snRNP were improved to a degree that allowed to exhaustively screen crystallization conditions, in the presence and absence of ligands. Initial U1 snRNP crystallization was regarded as a random event, depending on the quality of the input material (personal communication, U. Reidt). In this work, it was shown that protein truncation was required for U1 snRNP crystal formation. In order to truncate U1 snRNP, proteases were deliberately added to the crystallization setup, which resulted in the growth of U1 snRNP crystals in a reproducible fashion. Now, a major objective of this work could be addressed, which was to improve crystal diffraction to a resolution, rendering structure solution possible. This was achieved by extensive crystal screening and the addition of ligands that were known to bind to the particles. With a combination of high throughput crystal screening and the latest synchrotron technologies, the diffraction power of the crystals was extended from 25 Å to 4.5 A. Although atomic resolution was not obtained so far, strategies to obtain suitable phases for the present crystals were developed. The identification of heavy atom sites of several compounds as well as a molecular replacement solution now bring the solution of native

U1 snRNP in reach at medium resolution. A further improvement of the crystals would be required to obtain a crystallographic map at atomic resolution.

The advances that were made required a combinatorial effort to address multi-threaded problems and also to evaluate strategies. The variation of one parameter or the introduction of a new method like *in situ* proteolysis for U1 snRNP crystallization often resulted in a multitude of possible threads to follow. However, the most promising threads were followed, though it is reasonable that not all could be explored in an exhaustive fashion. Strategies of U1 purification and crystallization can now be transferred to the other spliceosomal particles that have been isolated in this work. Splicing factor 3a and U4/U6·U5 tri-snRNP were purified from HeLa cells and concentrated to a degree, that allowed initial crystallization screening. So far no crystal growth of these larger particles was observed. U1 snRNP only formed crystals upon deliberate protease treatment, a strategy that can now be transferred to tri-snRNP and SF3a. Initial proteolysis experiments for U4/U6·U5 tri-snRNP have been carried out to identify proteases suitable for *in situ* treatment.

4.1 Purification of snRNPs

4.1.1 Improvement of the purification protocol for snRNPs

Total snRNPs were purified from Hela nuclear extract by immunoaffinity chromatography followed by glycerol gradient centrifugation (Bringmann *et al*., 1983; Kastner and Lührmann, 1999). These purification protocols were established in the Lührmann department and yielded amounts that are suitable for X-ray crystallography studies. In this work, pre-purified particles from glycerol gradient fractions were further purified, concentrated and subjected to crystallization trials. The protocols for further purification and concentration were improved during this work.

Immunoaffinity purification and glycerol gradient centrifugation were already optimized and were carried out as described in the literature (3.1.1.1). The yield of snRNPs in glycerol gradient centrifugation was at least 70-80 % of the input and could be only roughly estimated due to errors of the Bradford assay. These errors were introduced by different glycerol concentrations and the presence of the m_3G cap in the buffer systems. Furthermore the contribution of the RNA to the particle's masses was not covered by the Bradford assay.

According to previous protocols, U1 and U2 snRNPs were separated by DEAE or MonoQ resins, obtaining only a minor fraction of pure U1 snRNP. Most of the material

was obtained as U1/U2 snRNP mixture after FPLC runs. Testing of several column materials revealed heparin sepharose to be suitable for full separation of the particles (3.1.1.2). This on one hand increased the yield of U1 snRNP from around 2 up to 10 mg per preparation and on the other hand yielded up to 2 mg U2 snRNP per FPLC run. U1 snRNP was now available in larger amounts for crystallography. Side fractions could be used for biochemical and immunological experiments.

U1 snRNPs were previously buffer exchanged by disposable PD10 desalting columns. Due to the increased amounts of U1 snRNP per preparation these columns were replaced by a 16/10 desalting column (3.1.1.3). Nearly all the input material was recovered after buffer exchange. The amounts of U1 snRNP that were available necessitated several rounds of pelleting to concentrate the material. During the pelleting steps nearly all input material could be recovered. Concentrations of up to 40 mg/ml were obtained, but 20 mg/ml were sufficient for crystallization. Crystal quality did not change upon freezing and thawing the concentrated material.

4.1.2 Purification of snRNPs for biochemical and immunological studies

In this work, U1 and U2 snRNPs were sucessfully separated by anion exchange chromatography and made available in large amounts. This material was used for biochemical and immunological studies. In a collaboration with Dr. Anne Krug, LMU Munich, it was shown that murine plasmacytoid dendritic cells (PDCs) were activated by purified U1snRNP/anti-Sm immuno complexes (ICs) to produce interferon-α and proinflammatory cytokines and to up-regulate costimulatory molecules. The U1snRNA component of U1snRNP immune complexes, found in patients with systemic *lupus erythematosus* (SLE), acted as an endogenous "self" ligand for toll-like receptor 7 and triggered interferon-α and interleukin-6 production in PDCs. PDCs produce interferon-α in response to autoimmune complexes containing nuclear antigens and are thought to be critically involved in the pathogenesis of systemic *lupus erythematosus* (SLE) (Savarese *et al.*, 2006).

Apart from U1 snRNP, 12 S U2 snRNP was also available for biochemical experiments and crystallization (Behzadnia *et al.*, 2006). Usually U1 and U2 snRNPs are equally abundant in Hela cells (Venrooij, 1987; Montzka and Steitz, 1988). The relatively low yield of U2 snRNP in contrast to U1 snRNP could be explained due to the fact that the m_3G cap of 17S U2 snRNP is not accessible by the H20 antibody in contrast to the

12 S particles. This way a major fraction of U2 snRNPs was in the flowthrough of the H20 immunoaffinity column and thus not part of the snRNP isolate.

4.2 Limited proteolysis of U1 snRNP

Limited proteolysis is a powerful tool to determine borders between structured and unstructured regions in proteins. The analysis of the proteolysis pattern by SDS-PAGE and mass spectrometry allows conclusions about the proteins that are truncated. Furthermore protease-resistant protein-fragments can be identified. The results of limited proteolysis experiments have to be treated with caution. A protein that is affected by proteolysis can be truncated in a stepwise fashion or nicked in the backbone. In the case of U1 snRNP, even small protein fragments that are not resolved by SDS-PAGE, can still be bound to the particle. A band resulting from a truncation can either overlap with other bands or be not stainable by the Coomassie dye.

A comparison of the results of the limited proteolyses (3.1.2.1) and the mass spectrometry data with a schematic representation of the U1 snRNP associated proteins (see 4.1) allows several conclusions about unstructured regions in U1 snRNP proteins.

The U1-70k protein contains a large C-terminal region, rich in arginines, glutamates and serines that can be reversibly phosphorylated (Xiao and Manley, 1997). The U1-70k protein was truncated by almost every protease used except carboxypeptidase Y (see figure 3.4 on page 58). This result suggested either an ordered region at the very C-terminus of the U1-70k protein or a lacking capability of carboxypeptidase Y to process phosphorylated serines. On the other hand leucine-aminopeptidase digested exclusively U1-70k. This may be the result of contamination by other proteases, since proteins with presumably unstructured N-terminal extensions like SmD2 or SmE remained intact. The entire leucine-aminopeptidase experiment unfortunately allowed no conclusions that matched to the presumed architecture of U1-70k. Asp-N digestion truncated only U1-70k to a size of 26 kDa (see figure 3.4 on page 58), resembling fragments of similar size obtained by other proteases. These fragments were identified by mass spectrometry in dissolved crystals and comprised the proteins N-terminus which was presumably protected by interactions with the RNA and protein. In the case of Asp-N cleavage the residual C-terminal fragments of U1-70k might be rapidly degraded, as the preferred cleavage site of Asp-N is highly abundant in U1-70k's C-terminus. The comparison of chymotrypsin and trypsin digestion of U1-70k allowed conclusions about cleavage at hydrophobic and basic charged residues respectively. One of the two large fragments
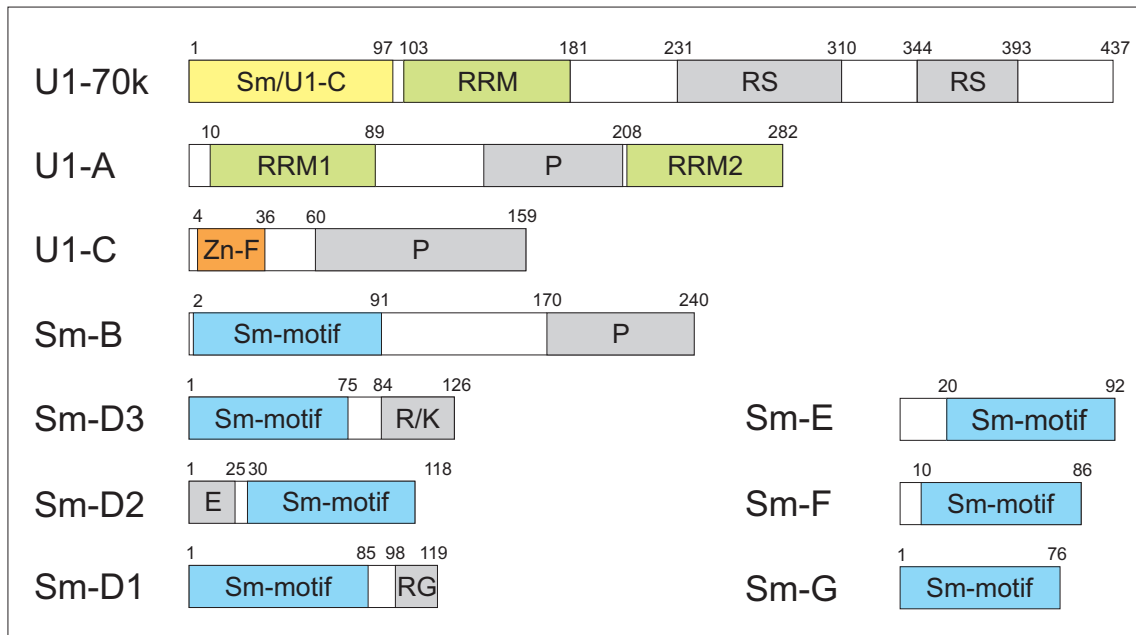
Figure 4.1 Schematic representation of the protein components of U1 snRNP. Proteins are represented by bars with a length relative to the number of the amino acids of the proteins. In addition amino acid positions at domain borders are denoted above the bars. Yellow color indicates the region, N-terminal of the U1-70k RRM, that is presumably contacting the U1-C and the Sm proteins B/B' and D2 (Nelissen *et al.*, 1994). Green color represents RNA-recognition motifs (RRMs) that contact the U1 snRNA, red color shows the second RRM of U1-A which makes no RNA-contacts, the zinc-finger of U1-C is colored orange. The Sm motif of the Sm core proteins is represented by blue color. All domains of low complexity or putatively unstructured domains are colored gray, letters represent the predominant amino acids within these regions in single letter amino acid code.

generated by chymotrypsin likely represented U1-70k. The other fragment that was absent in trypsin treated sample could be either U1 A, a differently truncated N-terminus or the entire C-terminus of U1-70k. Since the latter region comprises rather basic charged than hydrophobic residues it would be rapidly degraded by trypsin, Asp-N or Lys-C, but not by chymotrypsin. Taken together U1-70k was truncated by several proteases to a fragment of around $25 - 26$ kDa that contained the N-terminus of the protein (see figure 3.4 on page 58 and figure 3.5 on page 59). This fragment presumably comprised the interaction sites to U1-C, SmB/B' and SmD2 as well as the RRM.

For U1-A a protease sensitivity of the linker between the two RRMs has been noted (Nagai *et al.*, 2001; Oubridge *et al.*, 1994a). U1-A was truncated by all proteases except Asp-N Glu-C and leucine-aminopeptidase (see figure 3.4 on page 58 and figure 3.5 on page 59). Interestingly the preferred cleavage sites for Asp-N and Glu-C, asparagin and glutamate respectively are only found within 10 residues proximate to the RRMs

and therein, but not in the linker region. The notion that these residues are protected by the RRMs is in agreement with lacking cleavage of the protein by Asp-N and Glu-C. Irrespective of the truncation of U1-70k by leucine-aminopeptidase, U1-A is not expected to be cleaved by the protease, since its N-terminal RRM is involved in U1 snRNA binding. Carboxypeptidase on the other hand degrades the protein to a size of 18 kDa. The C-terminal RRM of U1-A has been shown not to interact with U1 snRNA and should be more sensitive to carboxypeptidase Y, when not bound to RNA (Lu and Hall, 1995, 1997).

The U1-C protein is affected by every protease, except leucine-aminopeptidase and Asp-N. Interestingly the entire proline-rich region of U1-C spanning from residues 51 to 158 does not contain aspartate residues which constitute the cleavage site of Asp-N. Aspartates in the vicinity of the N-terminal zinc-finger might be protected from cleavage. The degradation of the C-terminus of U1-C is further supported by the fact that the protein is cleaved by carboxypeptidase Y, but not by leucine-aminopeptidase. The U1-C zinc finger comprises around 30-40 residues and is too small to be visualized on the SDS gel. However it has been shown that the N-terminus of U1-C, comprising the Zn-finger, is buried in the U1 snRNP particle (Dumortier *et al.*, 1998). Depending on whether the U1-70k binding site was truncated or not, an N-terminal fragment of U1-C was presumably part of a degraded U1 snRNP. It has to be noted that U1-C is hardly detected by mass spectrometry in general (personal communication E. Kühn).

For the Sm proteins it is difficult to hypothesize about truncated fragments, since a disappearing full-length protein band could coincide with a emerging truncation product of a larger protein. Several results, obtained from mass spectrometric analyses of truncated Sm proteins were too ambiguous to draw conclusions about stable fragments. SmB/B' was truncated by all proteases except leucine-aminopeptidase and Asp-N (see figure 3.4 on page 58 and figure 3.5 on page 59). Like for U1-A and U1-C no aspartate is present in the presumably unstructured region at the C-terminus of B/B'. One truncation product of B/B' could be the fragment, appearing at around 15 kDa in several proteolysis experiments, since it was identified to contain B/B' fragments by mass spectrometry. It was expected that a truncation product of B/B' still comprised the Sm domain, which is complexed to the Sm domain of SmD3 (Kambach *et al.*, 1999a). Carboxypeptidase Y degraded B/B' to a size of at least 15 k, but the protein was not truncated in the presence of leucine-aminopeptidase. The Sm motifs in general form stable interfaces. The residues that comprise these interfaces that are likely to be protected from proteolysis.

The presence of DNA or RNA oligonucleotides changed the proteolysis pattern generated

by chymotrypsin (see figure 3.6 on page 61). Candidates for an interaction with a base-paired 5'splice site mimic would be the Sm proteins and primarily U1-C (Du and Rosbash, 2002). As mentioned previously, a stable fragment of the U1-C zinc-finger could not be identified on the gels used for limited proteolysis, due to its small size. A difference in the patterns of any of the Sm proteins was not observed. SmD3 was truncated by Lys-C, trypsin and chymotrypsin presumably at the C-terminus which is of low-complexity. D2 was only truncated by trypsin and Lys-C, presumably leaving a stable fragment of the Sm domain. However, for the smaller Sm proteins it was hard to judge which sites were accessible to the proteases and which fragments resulted from the proteolyis.

After comparing the different proteolysis results with the mass spectrometry analyses of dissolved U1 snRNP crystals a rough model of a truncated U1 snRNP could be deduced that might be favorable for crystallization in contrast to the full-length particle. As an alternative to *in situ* proteolysis, particles could be treated with proteases and re-purified before crystallization (pre-crystallization treatment). The proteolysis experiments could be used to select and combine proteases for pre-crystallization treatment. Possible candidates were for example Asp-N, chymotrypsin or carboxypeptidase Y.

4.3 Crystallization of U1 snRNP

4.3.1 Crystallization of U1 snRNP by *in situ* proteolysis

Initial U1 snRNP crystals had low reproducibility, and diffraction was hardly detectable. When these crystals were analyzed by SDS-PAGE, a truncation of several proteins was noted. Also in solution, proteases truncated U1 snRNP (see 4.2). By deliberately including proteases to the crystallization droplets, reproducibility was increased and crystal size was dramatically improved. This approach was also recently described by Dong *et al*. (2007), but for single proteins. This study demonstrated that the addition of proteases to the crystallization setup can be extended to multi-component assemblies, as postulated by Dong *et al*. (2007). Especially in the case of U1snRNP, which was natively isolated, proteolysis is a way to influence the shape of the particles.

A proteolysis experiment of U1 snRNP in solution showed, that the particles are truncated by a broad range of proteases. For this initial proteolysis experiment, mass spectrometric data were not available. A detailed analysis of dissolved U1 snRNP crystals by SDS-PAGE and mass spectrometry showed that presumed flexible regions of the proteins were removed by chymotrypsin. The degree of truncation depended on the

oligonucleotide used for co-crystallization. The U1 snRNP 70k protein comprises an alternatively spliced mini exon that results in a heterogeneity of the particles (Hornig *et al.*, 1989). This region was most likely removed by proteolysis of the particles; parts of it that are still attached to U1 snRNP by protein-protein interaction cannot be excluded. Mass spectrometry results confirmed the presence of the U1-70k N-terminus including the RRM in the U1 snRNP crystals. The N-terminus of U1-70k is involved in protein-protein contacts with U1-C, SmB/B' and D2 (Nelissen *et al.*, 1994). Taken together a truncation of the SR-domain of U1-70k during *in situ* proteolysis seemed to be a prerequisite for crystal growth

A second alternatively spliced region in U1 snRNP is the C-terminus of SmB/B' (van Dam *et al.*, 1989). SmB/B' was truncated to a size matching to the Sm core, respective peptides were found by mass spectrometry. Thus, a second heterogenous region of U1 snRNP was removed by proteolysis. Of the Sm proteins in general, mainly D3 was clearly affected by proteolysis, shifting to a size of the Sm core.

For U1-A, a removal of the second RRM of the material that crystallized, was assumed. In mass spectrometric analyses no peptides of the corresponding C-terminal region were found. In contrast, several peptides found contained the N-terminus of U1-A that is bound to the U1 snRNA (Scherly *et al.*, 1989). For the growth of the hexagonally shaped crystals, a truncation of U1-A seemed not necessary, since the presence of full-length protein was identified in the SDS-PAGE and mass spectrometry analyses of the respective crystals (see 3.1.7). Likewise for the hexagonal crystals, all Sm proteins, except D3 and maybe B/B' were not affected by the proteolysis. A strategy to improve these crystals could be the selective truncation of U1-70k, B/B' and D3 before crystallization.

U1 snRNP crystals were dissolved and analyzed by electron microscopy. The most prominent feature of U1 snRNP, the Sm core, was clearly identified in the electron microscopic images. Apart from the Sm core U1 snRNP has two characteristic protuberances, representing U1-A and U1-70k. These protuberances could not be identified on the crystallized particles (Kastner and Lührmann, 1989). The reduction of the protuberances could be a result of U1-A and U1-70k truncation, introduced by proteolysis.

4.3.2 Crystallization of U1 snRNP as a truncated particle and in the presence of ligands

Several attempts were made to improve the diffraction of the $P222_1$ crystals of U1 snRNP, obtained in the absence of ligands. Neither additive screens nor the mounting of the crystals on a FMS for controlled dehydration (not shown) led to better diffraction. Likewise other dehydration protocols, different cryoprotectants or the inclusion of heavy atom clusters did not improve diffraction. Propylene glycol as a cryoprotectant can be excluded from being the limiting factor for resolution, since measurements in a capillary at 4 °C did not show an improvement. When an oligonucleotide, complementary to the 5' splice site, was added to the crystals in a 1:2 molar ratio, they immediately dissolved, indicating the 5' terminus to be crucial for either global U1 snRNP conformation or crystal contacts. The $P222_1$ crystals were abandoned at an early phase of this work, since diffraction could not be improved. To improve the diffraction power of U1 snRNP crystals, the particle was truncated at the 5' splice site binding region. As an alternative strategy, binding partners like RNA and DNA oligonucleotides or snurportin 1 were added for co-crystallization. The removal of the 5' splice site led to reproducible U1 snRNP crystals of distinct morphology. This corroborated the importance of the 5' splice site binding region for U1 snRNP crystallization. The observation that only one condition out of 776 screened produced crystals, indicated that the 5'-splice site might also be important for the formation of crystal contacts.

In a different approach snurportin 1 (SPN1) was bound to the particles (Strasser *et al.*, 2005). Although the protein binds to core U snRNPs during nuclear transport, the presence of the U1 specific proteins did not abolish binding of snurpotin 1 to the particles (see figure 3.11). The addition of SPN1 to U1 snRNP crystallization led to crystals, presumably identical to initial $P222_1$ crystals, that only diffracted to 25 Å. Since crystal growth was only observed in the presence of chymotrypsin it cannot be excluded that SPN1 was entirely degraded by the protease, or dissociated from the particle, allowing for growth of the pyramidal crystals, previously observed. To further assess the effect of SPN1 in U1 snRNP crystallization, other proteases than chymotrypsin should be tested. In the case of SPN1 crystals, described in Strasser *et al.* (2004), Glu-C did not entirely degrade the protein. In addition there is a shortened version of the protein available, which lacks the N-terminal 66 amino acids, that are presumably flexible as a part of the Importin β binding domain. A combination of a N-terminally shortened SPN1 with other proteases than chymotrypsin could be a future strategy for U1 snRNP SPN1 co-crystals.

The addition of oligonucleotides to the crystallization setup had a fundamental impact on the morphology of the crystals. Table 4.1 lists all U1 snRNP crystals that were characterized in this work.

Table 4.1 Characteristics of U1 snRNP crystals. The respective maximum resolution (Resol.) is that of refined and additive-supplemented crystals. The space group (SG) is given, but it has to be noted that space group determination of low resolution datasets is ambiguous.

Crystal	SG	a, b, c (Å)	α, β, γ (°)	Resol. (Å)
U1 snRNP, pyramidal	P222$_1$	156, 178, 220	90, 90, 90	8-9
U1 snRNP, hexagonal	C222	481, 583, 335	90, 90, 90	20
U1 snRNP, plate, RNA	C2	358, 88, 152	90, 112, 90	6
U1 snRNP, prolate spheroid	C222	138, 238, 228	90, 90, 90	8
U1 snRNP, plate, DNA	C2	358, 88, 152	90, 112, 90	4.5

None of the U1 snRNP-oligonucleotide co-crystals shared morphological features with crystals that were obtained in the absence of oligonucleotides. Most RNA or DNA-co-crystals could be assigned to two or three classes according to their morphology and the reservoir composition. Previous observations implicated the 5' splice site in U1 snRNP crystal formation. The novel crystal morphologies that occurred upon nucleotide addition are likely a result of the modulation of an exposed region, the 5' end of U1 snRNA, that might mediate crystal contacts. Bound oligonucleotides presumably induce the formation of a helical structure. In addition a conformational change of the entire U1 snRNP was hypothesized.

The first class comprised hexagonally-shaped crystals that mostly contained neither salt nor crystallization buffer in the reservoir solution. It was hypothesized that lacking salt led to precipitation and crystal formation. The occurrence of these crystals was restricted to specific oligonucleotides, for example there were no hexagonal crystals observed with the oligonucleotide RNA5, but with RNA1, RNA2, RNA3, RNA8 or RNA9. Interestingly some of these crystals, like the ones shown in figure 3.12 D appeared after one day, indicating that a lesser extent of proteolysis had occurred. This was corroborated by SDS-PAGE analysis (see 3.13) where a lesser extent of proteolysis was observed in contrast to the other crystals analyzed. These crystals were not refined further, since the initial diffraction was about 20 Å.

A second class of U1 snRNP crystals was observed with the oligonucleotide DNA4, *in situ* digested with Asp-N instead of chymotrypsin. Although these crystals did not have a clear morphology, they produced a homogeneous diffraction pattern to around

8 Å. These crystals were barely reproducible, probably due to variations in the quality of Asp-N endoproteinase.

A third major class of U1 snRNP crystals exhibited a plate-like shape and depended on the presence of carbonic acids like citrate or acetate in the crystallization buffer. These crystals became the main focus of this work, since they had the best diffraction capabilities observed so far. Several rounds of refinement and an exchange of the oligonucleotide from RNA to DNA led to crystals that produced a dataset to 4.5 Å resolution. Overexposure of these crystals produced reflections up to a resolution of 3.8 Å. Thus, a merged dataset of around 4 Å resolution could be obtained from a multitude of crystals. Although a total ϕ-range of 180° would have to be collected, the crystals have a comparably low mosaicity of 0.4, and every second reflection of the longest axis is absent, due to the space group C2. Both facts limit the number of images that would have to be collected. The residual crystals described in this work were of low diffraction quality or have not been analyzed to an extent that allowed conclusions or relation to the crystals already characterized.

Taken together extensive screening of U1 snRNP oligonucleotide ligands increased the diffraction power of the crystals from 9 Å to 4.5 Å which potentially allowed initial molecular replacement trials and phasing approaches.

Some of the cell axes of the unit cells of the U1 snRNP crystals were related to each other by a factor of two or three. For example the a-axis of the $P222_1$ crystals was double length of the b-axis of the C2 crystals and vice versa. The c-axis of both crystal types seemed not related, but the c-axis of the C2 crystals was similar to the length of the a- axis of the $P222_1$ crystals. The hexagonal and prolate spheroid crystals were at least not obviously related. Similar cell-axes might indicate similar packing of the particles.

The best diffracting U1 snRNP crystals so far were obtained in the presence of the oligonucleotide DNA4 and chymotrypsin. *In situ* proteolysis generates protein fragments that are present during the crystal formation. Peptides or protein fragments of U1 snRNP, resulting from *in situ* proteolysis, might bind to the particles and disturb the crystal lattice. As a result the diffraction power could be severely impaired. It has been demonstrated that it is possible to obtain truncated U1 snRNPs that nearly match the crystals with respect to the pattern of proteolysis (see 3.11). A gel filtration at 500 mM sodium chloride resulted in the removal of several fragments that were present in the crystals, but apparently loosely bound. Due to time limitations the identities of the bands that were removed from the particles have not been analyzed by mass

spectrometry yet. The purified and truncated U1 snRNP particles crystallized. The crystals resulting from protease pre-treatment (see 2.2.3.11), only diffracted to 15 Å. In this case an improvement was not achieved. However, this approach only addressed the most successful crystallization condition so far and could be extended to other crystal screens or carried out in the absence of an oligonucleotide. Furthermore the salt concentration of the gel filtration run could be lowered to prevent dissociation of fragments that support the crystal packing.

4.4 Approaches to phase the U1 snRNP structure

Several strategies can be employed to assess crystallographic datasets in the first place to obtain information of unit cell contents. The self rotation function calculated of $U1_{DNA4}$ datasets showed a second peak next to the crystallographic axes, which pointed at the presence of two or an integer multiple of two complexes of U1 snRNP per asymmetric unit (see 3.1.11 on page 83). The assumption of two complexes per asymmetric unit was corroborated by the calculation of the packing density, assuming also two molecules per asymmetric unit (Kantardjieff and Rupp, 2003). The truncation of U1 snRNP by the protease was taken into account when the molecular weight of the particles was estimated.

4.4.1 A potential and partial solution by molecular replacement

Obtaining data to a resolution of 4.5 Å opened up the possibility of molecular replacement with structurally known subunits of U1 snRNP. Successful cases for molecular replacement at an equivalent resolution have been reported, but final structure solution at 4.5 Å depended on the presence of experimental phases (DeLaBarre and Brunger, 2003, 2006).

As a prominent feature of the U1 snRNP particle, two copies of the Sm core heptamer were placed in the unit cell using the program Phaser (see 3.1.12.1 on page 85). One heptamer was placed with low reliability, the second produced significant Z-scores and LL-gain values. Also a polyserine model of an *Archaeoglobus fulgidus* Sm heptamer was placed with a higher reliability in the unit cell. Apparently the highly conserved fold of the Sm core was sufficient for finding a solution at this intermediate resolution. Apart from the polyserine model, a model was generated, based on the structure from

Archaeoglobus fuligidus, but with the archaeal proteins replaced by known structures from *Homo sapiens* or respective homology models (see figure 1.4 on page 9). Although the polyserine model produced better Z-scores than the synthetic hetero-heptamer, the latter was placed in the unit cell according to the expected NCS-rotation derived from the self-rotation function. Therefore the polyserine model was not tested exhaustively for further replacement runs. In a second search U1-A complexed with the stem loop II of U1 snRNA was placed in the unit cell with Z-scores up to 19 and LL-gain values of 500 - 600. Like for the Sm heptamer the two positioned U1-A-RNA complexes were related by the experimentally observed NCS, derived from the self-rotation function.

When the most prominent feature of the Sm core, the extended β sheet of B/B' was removed in a MR-search, the rotational positioning of the hetero-heptamer did not change. Also a search, lacking the RNA of the U1-A model resulted in the same positioning of the molecule. Although further substructures of U1 snRNP were available, neither the zinc-finger of U1-C nor stretches of the modeled RNA were found (Muto *et al*., 2004; Krol *et al*., 1990). It has been observed previously that NMR models like U1-C, have low efficiency as a search model for molecular replacement (Chen *et al*., 2000). The stretches of the U1 snRNA, a component definitely present in the crystals, that were searched were probably too small in size or the model did not have sufficient quality. In addition, the RNA closely contacts the proteins, so a slight misplacement of the protein components due to the moderate resolution, might prevent a correct placement of the RNA. Interestingly a solution was only found with Phaser, but not with Molrep or Amore. This could be attributed to the Log-likelihood approach that is employed by Phaser and which has been positively reviewed in the crystallographic community (CCP4i bulletin board).

As already mentioned in the results part, different refinement protocols with phenix.refine or Refmac were not successful. The missing components of the particle were not interpretable in the electron density. This was most likely due to the relatively low resolution for a molecular replacement search. Several strategies for the refinement of low resolution data have been proposed by DeLaBarre and Brunger (2006) and were applied to the U1$_{DNA4}$ data, but no additional electron density was detectable. Omit maps of the MR-solution were calculated. The electron density of subunits of the model that were removed during refinement, disappeared. Thus, the density of lacking subunits of U1 snRNP could not be expected to appear after refinement.

The molecular replacement model was sufficient to localize initial heavy atom sites of iridium and osmium hexammines as well as W5. The model was further supported by

experimental SAD phasing, which yielded a W5 substructure that was identical to the substructure obtained with the MR-phases. In some aspects the MR-model can be put to question; however it seemed to provide sufficient information for initial phasing. In addition it can be transferred to non-isomorphous derivative datasets like W5 and used to combine this phase information with the native or other derivative datasets with DM Multi (CCP4i suite). This has not been exhaustively tested yet.

4.4.2 Solution of three heavy atom substructures of U1$_{DNA4}$ co-crystals

The phase information that was obtained from the molecular replacement solution, was apparently not sufficient to interpret the residual U1 snRNP structure. U1-70k as well as about 140 nucleotides were not found in the refined MR-electron density maps but clearly present in the crystals according to SDS-PAGE analysis. For other structures at equivalent resolution, experimental phase information was required for final structure solution although a large part of the structure could be interpreted by molecular replacement and NCS was present(DeLaBarre and Brunger, 2006). Initial trials to derivatize U1$_{DNA4}$ crystals with heavy atoms failed, presumably because the crystallization condition contained EDTA and citrate, two well known chelators. When both substances, EDTA and citrate, were avoided by using a synthetic buffer, resolution dropped dramatically and the longest cell axis was reduced by around $1 - 3$ %. The change of cell axes could not only be attributed to the heavy atom presence but could also be a side effect of buffer change. This has not been clarified yet.

In contrast to the other datasets, derivatization with W5 led to a slightly increased length of the a-axis (see table 3.3). This could be an indicator for binding of the compound to U1 snRNP. By heavy atom searches using the anomalous data of the W5 cluster, four W5 sites were found by the program hkl2map (see 3.20 on page 94).

The W5 sites that were found are listed in table 4.2. In addition to the experimental approach the W5-substructure was confirmed by a Fourier map, calculated of the anomalous difference of the W5 structure factors and phases that were obtained from a molecular replacement solution of the W5 data. The minimal search model for molecular replacement encompassed one Sm heptamer and U1-A complexed with RNA.

Two of the W5-sites found, correlated with the NCS that was found by molecular replacement. One cluster was located at a special position in the vicinity of U1-A. The fourth site was found in the vicinity of the origin, not related by a two-fold NCS. SAD phasing and density modification including NCS averaging were tried but were not

successful. The resulting density was not interpretable. Also including further heavy atom cluster derivative datasets as 'pseudo native' did not yield a solution. A reason for this was presumably the low resolution in general and that no true native dataset for the derivatized crystals was available. It has been shown that cross-linking of protein crystals with glutaraldehyde prior to derivatization might prevent the change of cell constants. This would retain isomorphism when the crystals are derivatized (Lusty, 1999). This technique could also be applied to U1 snRNP$_{DNA4}$ crystals.

Iridium and osmium hexammines were successfully used for phasing of the large ribosomal subunit, where they bound to GU wobble basepairs mainly (Ban *et al*., 1998, 2000). U1 snRNP contains 4 predicted GU wobble basepairs, according to the secondary structure of its RNA (Guthrie and Patterson, 1988).

Table 4.2 Heavy atom sites of U1$_{DNA4}$ crystals. The sites of W5, Oshex and Irhex derivatives are given in fractional coordinates.

Crystal	SG	x (fractional)	y (fractional)	z (fractional)	atom
U1$_{DNA4}$	C2	0.0214	0.2709	0.4372	W5 1
U1$_{DNA4}$	C2	0.3470	0.1895	0.4605	W5 2
U1$_{DNA4}$	C2	0.2930	0.1415	0.0998	W5 3
U1$_{DNA4}$	C2	0.2572	0.1056	0.5263	W5 4
U1$_{DNA4}$	C2	0.3524	0.0119	0.0781	Irhex 1
U1$_{DNA4}$	C2	0.3448	0.3672	0.0096	Irhex 2
U1$_{DNA4}$	C2	0.3506	0.0136	0.0778	Oshex 1
U1$_{DNA4}$	C2	0.3422	0.3657	0.0121	Oshex 2

Neither osmium nor iridium hexammine changed cell constants dramatically, though the β angle slightly increased. Both datasets scaled to the native U1$_{DNA4}$ crystals with comparably low R-factors. Heavy atom sites were only found in a difference Fourier map of the derivative and the native datasets. For calculation of the Fourier map, phases of the molecular replacement solution of the native dataset were used. Two identical osmium and iridium hexammine sites were found in all maps calculated. The map was compared to the molecular replacement solution, obtained by Phaser. The sites localized to a GU wobble basepair in stem loop II of U1 snRNA. The potential osmium and iridium hexammine sites are listed in table 4.2. Since the hexammine sites localized to both NCS-related GU wobble base pairs, the molecular replacement was further corroborated.

Since osmium and iridium hexammine derivative datasets were isomorphous to the native

data, phasing was tried with both derivatives and in combination with the native data. MIR and MIRAS phasing protocols were not successful. Also a phase combination of experimental phases with the MR-phases of the native data with the program sigma A was not sufficient to see additional density for the lacking parts of U1 snRNP. The programs Mlphare and Sharp/Autosharp have an option to include molecular replacement phases. Both programs did not produce an interpretable electron density map. The resolution of the osmium and iridium hexammine derivative datasets was only $6 - 7$ Å. For the osmium and iridium hexammine derivative crystals, an improvement of the resolution could be sufficient to localize the sites by SAD or SIR. In a next step the derivatization protocol for these crystals has to be improved to obtain a better resolution for these crystals.

4.5 Purification and concentration of U4/U6·U5 tri-snRNP for crystallization

U4/U6·U5 tri-snRNP consists of 16 particle-specific proteins, a double set of Sm core proteins as well as a set of LSm proteins and three RNAs (see table 1.3 on page 7). During the spliceosomal cycle, its components are involved in both catalytic steps of the splicing reaction. Especially the 220k protein has been shown to contact several regions of the pre-mRNA during splicing as well as the U5 snRNA (Reyes *et al.*, 1999; Chiara *et al.*, 1996; Urlaub *et al.*, 2000). A crystal structure of tri-snRNP would yield deep insight into spliceosome architecture and function and could be combined with electron microscopic data (Sander *et al.*, 2006).

The particles were pre-purified as shown in figure 3.2 according to a protocol that was established in the Lührmann department. The tri-snRNP concentration of the peak fractions of the glycerol gradient was 0.3 mg/ml. To obtain tri-snRNP for crystallization trials, two questions had to be addressed. For crystallographic studies, proteins are recommended to have a concentration of at least $2 - 5$ mg/ml (Rhodes, 2006). In addition, tri-snRNP was isolated in the presence of relatively high glycerol concentrations. Glycerol was unfavorable as a component of a general crystallization buffer. Glycerol was shown to act as a crystallization inhibitor in some cases (Sedgwick *et al.*, 2007). Therefore it was an intermediate goal of tri-snRNP crystallization to obtain highly concentrated particles, preferably in a glycerol-free buffer. This material would be suitable to set up initial crystallization trials.

Initial approaches to obtain a concentrated fraction of tri-snRNP failed. The material

precipitated almost entirely when it was pelleted in a mini-ultracentrifuge or when standard protein concentrators were used. Likewise, buffer exchange resulted in loss of nearly all material. Notably, tri-snRNP was stable in the presence of the other snRNPs, but not in isolation. Even in the glycerol gradient fractions (see figure 3.2) the particles tended to precipitate over time. Therefore, a novel protocol was developed that relied on ammonium sulfate precipitation, followed by a buffer exchange to a glycerol-free crystallization buffer (see 3.2.1). The ammonium sulfate precipitation was carried out with solid ammonium sulfate powder and led to the formation of a white viscous suspension. The precipitate was sedimented by centrifugation. The viscous pellet had a protein concentration of about $18 - 20$ mg/ml and was carefully resuspended and transferred to a dialysis membrane. For buffer exchange, it was crucial to use floating membranes that were routinely used for the desalting of DNA. Several other membranes and desalting columns were tested, but precipitated most of the material irreversibly.

A comparatively low amount of ammonium sulfate had to be preserved in the dialysis buffer that was also used for crystallization screening. When the dialysis buffer was lacking ammonium sulfate, U4/U6·U5 tri-snRNP entirely precipitated during dialysis. During dialysis the volume of the drop increased by about $50 - 60$ %, and a final protein concentration of $7 - 10$ mg/ml was obtained. With this protocol the particles were brought to a reasonable concentration, suitable for crystallization screening. Precipitation of the particles was not observed when the material was stored at 4 ℃.

As a quality control, concentrated and buffer exchanged tri-snRNP was loaded on an analytical glycerol gradient. U4/U6·U5 tri-snRNP migrated as one particle, encompassing all proteins and the RNA. The crystal screening did not produce any crystals. Like U1 snRNP, tri-snRNP comprises several proteins that encompass flexible domains. For crystallization it is favorable to remove these domains, as they might interfere with crystal packing. At this stage the strategies, that were applied to obtain U1 snRNP crystals, were transferred to U4/U6·U5 tri-snRNP.

A limited proteolysis was set up in solution with three proteases to identify proteases that were suitable for *in situ* or in solution treatment. In this experiment, Glu-C produced the most promising proteolysis pattern with several fragments that appeared stable (see figure 3.24). Taken together, suitable input material of U4/U6·U5 tri-snRNP for crystallization was obtained and limited proteolysis experiments identified a suitable protease for digestion of the particles *in situ* or prior to crystallization. In a next step, these results can be applied for further crystallization screening.

4.6 Purification of SF3a for crystallization

SF3a is associated with U2 snRNP at salt concentrations lower than 175 mM. The particle comprises three proteins and can be isolated from nuclear extract by immunoaffinity chromatography in a one-step procedure. The 17S U2 snRNP has been studied by electron microscopy, which allowed a global assignment of its additional large substructures SF3a and SF3b in relation to 12S U2 snRNP (Krämer *et al.*, 1999). Recently, structures of the first SF3a-120k SURP domain and of an interface of the second SURP domain, complexed with a peptide of SF3a60, have been determined by NMR. This comprises a relatively small fraction of SF3a in total (Kuwasako *et al.*, 2006). A structure of the entire SF3a was one objective of this work.

After purification and concentration of the particle (see 2.2.3.4) only 50 % of the input material could be recovered, still representing a reasonable yield for one purification step and concentration. So far no crystals of SF3a were obtained. At this stage the limited proteolysis protocols that were devised for U1 snRNP could be employed. If stable fragments could be identified by mass spectrometry, it would be possible to produce a truncated version of SF3a in *E. coli*, as described in Kuwasako *et al.* (2006). Furthermore *in situ* proteolysis would be an option to increase the chance of crystallization of SF3a. The purified SF3a could further be used for reconstitution assays with 12S U2 snRNP and purified SF3b to form a particle that resembles 17S U2 snRNP. (Dybkov *et al.*, 2006). In 17S U2 snRNP, flexible domains of SF3a might get structured upon binding which is favorable for crystallization.

5 Bibliography

Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Lührmann, R. 1998. The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol Cell Biol*, 18(11):6756–66.

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(20):5789–802.

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(Pt 1):53–55.

Ashcroft, N. and Mermin, N. 1976. *Solid State Physics*. Saunders College Publishing.

Avvedimento, V. E., Vogeli, G., Yamada, Y., Maizel, J. V., Pastan, I., and de Crombrugghe, B. 1980. Correlation between splicing sites within an intron and their sequence complementarity with U1 RNA. *Cell*, 21(3):689–696.

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*, 181:232–57.

Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R. A., Sweet, R., Frank, J., Moore, P. B., and Steitz, T. A. 1998. A 9 A resolution X-ray crystallographic map of the large ribosomal subunit. *Cell*, 93(7):1105–1115.

Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, 289(5481):905–920.

Behzadnia, N., Hartmuth, K., Will, C. L., and Lührmann, R. 2006. Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP. *Rna*, 12(9):1738–46.

Benecke, H., Lührmann, R., and Will, C. L. 2005. The U11/U12 snRNP 65K protein acts as a molecular bridge, binding the U12 snRNA and U11-59K protein. *EMBO J*, 24(17):3057–3069.

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

Berglund, J. A., Abovich, N., and Rosbash, M. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev*, 12(6):858–67.

Berglund, J. A., Rosbash, M., and Schultz, S. C. 2001. Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA*, 7(5):682–691.

Bindereif, A. and Green, M. R. 1987. An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *Embo J*, 6(8):2415–24.

Black, D. L., Chabot, B., and Steitz, J. A. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell*, 42(3):737–50.

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(8):3350–9.

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*, 72:248–54.

Brahms, H., Meheus, L., de Brabandere, V., Fischer, U., and Lührmann, 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(11):1531–42.

Brenner, T. J. and Guthrie, C. 2006. Assembly of Snu114 into U5 snRNP requires Prp8 and a functional GTPase domain. *RNA*, 12(5):862–871.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr*, 59(Pt 11):2023–2030.

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

Broennimann, C., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H., and Wagner, A. 2006. The PILATUS 1M detector. *J Synchrotron Radiat*, 13(Pt 2):120–130.

Brosi, R., Groning, K., Behrens, S. E., Lührmann, R., and Kramer, A. 1993. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science*, 262(5130):102–5.

Camasses, A., Bragado Nilsson, E., Martin, R., Séraphin, B., and Bordonné, R. 1998. Interactions within the yeast Sm core complex: from proteins to amino acids. *Mol Cell Biol*, 18(4):1956–1966.

Cao, W., Jamison, S. F., and Garcia Blanco, M. A. 1997. Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. *RNA*, 3(12):1456–1467.

Cate, J. H. and Doudna, J. A. 1996. Metal-binding sites in the major groove of a large ribozyme domain. *Structure*, 4(10):1221–1229.

Cáceres, J. F. and Kornblihtt, A. R. 2002. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet*, 18(4):186–193.

Chan, S.-P., Kao, D.-I., Tsai, W.-Y., and Cheng, S.-C. 2003. The Prp19p-associated complex in spliceosome activation. *Science*, 302(5643):279–282.

Chang, Y.-F., Imam, J. S., and Wilkinson, M. F. 2007. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem*, 76:51–74.

Chen, Y. W., Dodson, E. J., and Kleywegt, G. J. 2000. Does NMR mean "not for molecular replacement"? Using NMR-based search models to solve protein crystal structures. *Structure*, 8(11):R213–R220.

Cheng, D., Côté, J., Shaaban, S., and Bedford, M. T. 2007. The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol Cell*, 25(1):71–83.

Chiara, M. D., Gozani, O., Bennett, M., Champion Arnaud, P., Palandjian, L., and Reed, R. 1996. Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. *Mol Cell Biol*, 16(7):3317–26.

Chiara, M. D., Palandjian, L., Feld Kramer, R., and Reed, R. 1997. Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. *Embo J*, 16(15):4746–59.

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

Chua, K. and Reed, R. 1999. Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. *Genes Dev*, 13(7):841–50.

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

Collins, C. A. and Guthrie, C. 2000. The question remains: is the spliceosome a ribozyme? *Nat Struct Biol*, 7(10):850–4.

Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*, 288(5466):640–649.

Degen, W. G., Aarssen, Y., Pruijn, G. J., Utz, P. J., and van Venrooij, W. J. 2000. The fate of U1 snRNP during anti-Fas induced apoptosis: specific cleavage of the U1 snRNA molecule. *Cell Death Differ*, 7(1):70–79.

DeLaBarre, B. and Brunger, A. T. 2003. Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. *Nat Struct Biol*, 10(10):856–863.

DeLaBarre, B. and Brunger, A. T. 2006. Considerations for the refinement of low-resolution crystal structures. *Acta Crystallogr D Biol Crystallogr*, 62(Pt 8):923–932.

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

Dönmez, G., Hartmuth, K., Kastner, B., Will, C. L., and Lührmann, R. 2007. The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. *Mol Cell*, 25(3):399–411.

Dong, A., Xu, X., Edwards, A. M., , M. C. f. S. G., , S. G. C., 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., Flodin, S., Flores, A., Gräslund, S., Hammerström, 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(12):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(6902):86–90.

Dumortier, H., Klein Gunnewiek, J., Roussel, J. P., van Aarssen, Y., Briand, J. P., van Venrooij, W. J., and Muller, S. 1998. At least three linear regions but not the zinc-finger domain of U1C protein are exposed at the surface of the protein in solution and on the human spliceosomal U1 snRNP particle. *Nucleic Acids Res*, 26(23):5486–5491.

Dybkov, O., Will, C. L., Deckert, J., Behzadnia, N., Hartmuth, K., and Lührmann, R. 2006. U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. *Mol Cell Biol*, 26(7):2803–16.

Emsley, P. and Cowtan, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 12 Pt 1):2126–2132.

Fetzer, S., Lauber, J., Will, C. L., and Lührmann, 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(4):344–55.

Förch, P., Puig, O., Martínez, C., Séraphin, B., and Valcárcel, J. 2002. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *EMBO J*, 21(24):6882–6892.

Fury, M., Andersen, J., Ponda, P., Aimes, R., and Zieve, G. W. 1999. Thirteen anti-Sm monoclonal antibodies immunoprecipitate the three cytoplasmic snRNP core protein precursors in six distinct subsets. *J Autoimmun*, 12(2):91–100.

Fury, M. G. and Andersen, J. 1997. In vitro interaction of U2 snRNA with cytoplasmic 6S protein complexes. *FEBS Lett*, 404(1):70–74.

Garcia Blanco, M. A. 2003. Messenger RNA reprogramming by spliceosome-mediated RNA trans-splicing. *J Clin Invest*, 112(4):474–480.

Garcia Blanco, M. A. 2006. Alternative splicing: therapeutic target and tool. *Prog Mol Subcell Biol*, 44:47–64.

Gast, W. H., Kabsch, W., Wittinghofer, A., and Leberman, R. 1977. Crystals of a large tryptic peptide (fragment A) of elongation factor EF-Tu from Escherichia coli. *FEBS Lett*, 74(1):88–90.

Girard, C., Verheggen, C., Neel, H., Cammas, A., Vagner, S., Soret, J., Bertrand, E., and Bordonné, R. 2008. Characterization of a short isoform of human tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *J Biol Chem*, 283(4):2060–2069.

Günzl, A., Palfi, Z., and Bindereif, A. 2002. Analysis of RNA-protein complexes by oligonucleotide-targeted RNase H digestion. *Methods*, 26(2):162–169.

Golas, M. M., Sander, B., Will, C. L., Lührmann, R., and Stark, H. 2003. Molecular architecture of the multiprotein splicing factor SF3b. *Science*, 300(5621):980–4.

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(6):869–83.

Guan, F., Palacios, D., Hussein, R. I., and Gunderson, S. I. 2003. Determinants within an 18-amino-acid U1A autoregulatory domain that uncouple cooperative RNA binding, inhibition of polyadenylation, and homodimerization. *Mol Cell Biol*, 23(9):3163–3172.

Gunnewiek, J. M., van Aarssen, Y., Wassenaar, R., Legrain, P., van Venrooij, W. J., and Nelissen, R. L. 1995. Homodimerization of the human U1 snRNP-specific protein C. *Nucleic Acids Res*, 23(23):4864–4871.

Guthrie, C. and Patterson, B. 1988. Spliceosomal snRNAs. *Annu Rev Genet*, 22:387–419.

Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*, 166(4):557–580.

Hashimoto, C. and Steitz, J. A. 1984. U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res*, 12(7):3283–93.

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(2-3):165.

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(9):2076–88.

Hochleitner, E. O., Kastner, B., Fröhlich, T., Schmidt, A., Lührmann, R., Arnold, G., and Lottspeich, F. 2005. Protein stoichiometry of a multiprotein complex, the human spliceosomal U1 small nuclear ribonucleoprotein: absolute quantification using isotope-coded tags and mass spectrometry. *J Biol Chem*, 280(4):2536–2542.

Hornig, H., Fischer, U., Costas, M., Rauh, A., and Lührmann, R. 1989. Analysis of genomic clones of the murine U1RNA-associated 70-kDa protein reveals a high evolutionary conservation of the protein between human and mouse. *Eur J Biochem*, 182(1):45–50.

Hu, J., Xu, Y., Schappert, K., Harrington, T., Wang, A., Braga, R., Mogridge, J., and Friesen, J. D. 1994. Mutational analysis of the PRP4 protein of Saccharomyces cerevisiae suggests domain structure and snRNP interactions. *Nucleic Acids Res*, 22(9):1724–1734.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., and Nagai, K. 1991. Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J*, 10(11):3447–3456.

Kabsch, W. 1988. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. 21(6):916–924.

Kabsch, W., Gast, W. H., Schulz, G. E., and Leberman, R. 1977. Low resolution structure of partially trypsin-degraded polypeptide elongation factor, EF-TU, from Escherichia coli. *J Mol Biol*, 117(4):999–1012.

Kambach, C., Walke, S., and Nagai, K. 1999a. Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr Opin Struct Biol*, 9(2):222–30.

Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Lührmann, R., Li, J., and Nagai, K. 1999b. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, 96(3):375–87.

Kantardjieff, K. A. and Rupp, B. 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci*, 12(9):1865–1871.

Kastner, B., Kornstädt, 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(4):839–849.

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(1):277–286.

Kastner, B. and Lührmann, R. 1999. Purification of U small nuclear ribonucleoprotein particles. *Methods Mol Biol*, 118:289–98.

Klein Gunnewiek, J. M., Hussein, R. I., van Aarssen, Y., Palacios, D., de Jong, R., van Venrooij, W. J., and Gunderson, S. I. 2000. Fourteen residues of the U1 snRNP-specific U1A protein are required for homodimerization, cooperative RNA binding, and inhibition of polyadenylation. *Mol Cell Biol*, 20(6):2209–2217.

Kleywegt, G. J. and Jones, T. A. 1996a. Phi/psi-chology: Ramachandran revisited. *Structure*, 4(12):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(Pt 4):826–828.

Krämer, A. 1988. Presplicing complex formation requires two proteins and U2 snRNP. *Genes Dev*, 2(9):1155–67.

Krämer, A. 1992. Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. *Mol Cell Biol*, 12(10):4545–52.

Krämer, 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(7):1355–68.

Krämer, A., Mulhauser, F., Wersig, C., Gröning, K., and Bilbe, G. 1995. Mammalian splicing factor SF3a120 represents a new member of the SURP family of proteins and is homologous to the essential splicing factor PRP21p of Saccharomyces cerevisiae. *RNA*, 1(3):260–272.

Krämer, A. and Utans, U. 1991. Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. *Embo J*, 10(6):1503–9.

Krämer, A., Utans, U., Keller, W., and Lamond, A. 1990. Components involved in nuclear pre-mRNA splicing. *Mol Biol Rep*, 14(2-3):199–200.

Krol, A., Carbon, P., Ebel, J. P., and Appel, B. 1987. Xenopus tropicalis U6 snRNA genes transcribed by Pol III contain the upstream promoter elements used by Pol II dependent U snRNA genes. *Nucleic Acids Res*, 15(6):2463–2478.

Krol, A., Westhof, E., Bach, M., Lührmann, R., Ebel, J. P., and Carbon, P. 1990. Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res*, 18(13):3803–3811.

Kuwasako, K., He, F., Inoue, M., Tanaka, A., Sugano, S., Güntert, P., Muto, Y., and Yokoyama, S. 2006. Solution structures of the SURP domains and the subunit-assembly mechanism within the splicing factor SF3a complex in 17S U2 snRNP. *Structure*, 14(11):1677–1689.

Kyriakopoulou, C., Larsson, P., Liu, L., Schuster, J., Söderbom, F., Kirsebom, L. A., and Virtanen, A. 2006. U1-like snRNAs lacking complementarity to canonical 5' splice sites. *RNA*, 12(9):1603–1611.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(259):680–5.

Laggerbauer, B., Achsel, T., and Lührmann, R. 1998. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. *Proc Natl Acad Sci U S A*, 95(8):4188–92.

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(5):598–608.

Lauber, J., Fabrizio, P., Teigelkamp, S., Lane, W. S., Hartmann, E., and Lührmann, R. 1996. The HeLa 200 kDa U5 snRNP-specific protein and its homologue in Saccharomyces cerevisiae are members of the DEXH-box protein family of putative RNA helicases. *Embo J*, 15(15):4001–15.

Lazar, E., Jacob, M., Krol, A., and Branlant, C. 1982. Accessibility of U1 RNA to base pairing with a single-stranded DNA fragment mimicking the intron extremities at the splice junction. *Nucleic Acids Res*, 10(4):1193–201.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M. J. 2001. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *Embo J*, 20(17):4987–97.

Lehmeier, T., Raker, V., Hermann, H., and Lührmann, R. 1994. cDNA cloning of the Sm proteins D2 and D3 from human small nuclear ribonucleoproteins: evidence for a direct D1-D2 interaction. *Proc Natl Acad Sci U S A*, 91(25):12317–12321.

Lerner, M. R. and Steitz, J. A. 1979. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci U S A*, 76(11):5495–9.

Lindahl, E., Azuara, C., Koehl, P., and Delarue, M. 2006. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res*, 34(Web Server issue):W52–W56.

Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Lührmann, 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(5821):115–20.

Liu, S., Rauhut, R., Vornlocher, H. P., and Lührmann, R. 2006. The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *Rna*, 12(7):1418–30.

Lu, J. and Hall, K. B. 1995. An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J Mol Biol*, 247(4):739–752.

Lu, J. and Hall, K. B. 1997. Tertiary structure of RBD2 and backbone dynamics of RBD1 and RBD2 of the human U1A protein determined by NMR spectroscopy. *Biochemistry*, 36(34):10393–10405.

Lusty, C. J. 1999. A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. *Journal of Applied Crystallography*, 32(1):106–112.

Makarov, E. M., Makarova, O. V., Achsel, T., and Lührmann, R. 2000. The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions. *J Mol Biol*, 298(4):567–75.

Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., and Lührmann, R. 2002. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, 298(5601):2205–8.

Makarova, O. V., Makarov, E. M., and Lührmann, R. 2001. The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *Embo J*, 20(10):2553–63.

Maniatis, T. and Reed, R. 2002. An extensive network of coupling among gene expression machines. *Nature*, 416(6880):499–506.

Merril, C. R. and Goldman, D. 1982. Quantitative two-dimensional protein electrophoresis for studies of inborn errors of metabolism. *Clin Chem*, 28(4 Pt 2):1015–1020.

Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. 2006. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr*, 62(Pt 8):859–866.

Montzka, K. A. and Steitz, J. A. 1988. Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc Natl Acad Sci U S A*, 85(23):8885–9.

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(6444):364–8.

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*, (1):275–6.

Muto, Y., Pomeranz Krummel, D., Oubridge, C., Hernandez, H., Robinson, C. V., Neuhaus, D., and Nagai, K. 2004. The structure and biochemical properties of the human spliceosomal protein U1C. *J Mol Biol*, 341(1):185–198.

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. Novartis Medal Lecture. *Biochem Soc Trans*, 29(Pt 2):15–26.

Nelissen, R. L., Heinrichs, V., Habets, W. J., Simons, F., Lührmann, R., and van Venrooij, W. J. 1991a. Zinc finger-like structure in U1-specific protein C is essential for specific binding to U1 snRNP. *Nucleic Acids Res*, 19(3):449–454.

Nelissen, R. L., Sillekens, P. T., Beijer, R. P., Geurts van Kessel, A. H., and van Venrooij, W. J. 1991b. Structure, chromosomal localization and evolutionary conservation of the gene encoding human U1 snRNP-specific A protein. *Gene*, 102(2):189–196.

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(17):4113–25.

Nesic, D. and Krämer, A. 2001. Domains in human splicing factors SF3a60 and SF3a66 required for binding to SF3a120, assembly of the 17S U2 snRNP, and prespliceosome formation. *Mol Cell Biol*, 21(19):6406–6417.

Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., and Lührmann, R. 1999. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *Embo J*, 18(21):6119–33.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. 1994a. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, 372(6505):432–8.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. 1994b. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, 372(6505):432–438.

Owen, R. L., Rudiño Piñera, E., and Garman, E. F. 2006. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc Natl Acad Sci U S A*, 103(13):4912–4917.

Padgett, R. A. and Shukla, G. C. 2002. A revised model for U4atac/U6atac snRNA base pairing. *Rna*, 8(2):125–8.

Pascolo, E. and Séraphin, B. 1997. The branchpoint residue is recognized during commitment complex formation before being bulged out of the U2 snRNA-pre-mRNA duplex. *Mol Cell Biol*, 17(7):3469–3476.

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(3):747–51.

Pena, V., Liu, S., Bujnicki, J. M., Lührmann, 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(4):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(Pt 7):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(6694):645–50.

Query, C. C., Moore, M. J., and Sharp, P. A. 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev*, 8(5):587–97.

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(10):6554–65.

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(9):2256–69.

Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R., and Ficner, R. 1999. Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *J Mol Biol*, 294(2):515–25.

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

Rhode, B. M., Hartmuth, K., Westhof, E., and Lührmann, R. 2006. Proximity of conserved U6 and U2 snRNA elements to the 5' splice site region in activated spliceosomes. *EMBO J*, 25(11):2475–2486.

Rhodes, G. 2006. *Crystallography Made Crystal Clear*. Elsevier Academic Press, third edition edition.

Romac, J. M., Graff, D. H., and Keene, J. D. 1994. The U1 small nuclear ribonucleoprotein (snRNP) 70K protein is transported independently of U1 snRNP particles via a nuclear localization signal in the RNA-binding domain. *Mol Cell Biol*, 14(7):4662–4670.

Roscigno, R. F. and Garcia Blanco, M. A. 1995. SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA*, 1(7):692–706.

Rossi, F., Forné, T., Antoine, E., Tazi, J., Brunel, C., and Cathala, G. 1996. Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5' splice site-U1 snRNP interaction. *J Biol Chem*, 271(39):23985–23991.

Roy, J., Zheng, B., Rymond, B. C., and Woolford, J. L. 1995. Structurally related but functionally distinct yeast Sm D core small nuclear ribonucleoprotein particle proteins. *Mol Cell Biol*, 15(1):445–455.

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(1):317–31.

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., Lührmann, R., and Stark, H. 2006. Organization of core spliceosomal components U5 snRNA loop

I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol Cell*, 24(2):267–78.

Savarese, E., Chae, O.-w., Trowitzsch, S., Weber, G., Kastner, B., Akira, S., Wagner, H., Schmid, R. M., Bauer, S., and Krug, A. 2006. U1 small nuclear ribonucleoprotein immune complexes induce type I interferon in plasmacytoid dendritic cells through TLR7. *Blood*, 107(8):3229–3234.

Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J., and Mattaj, I. W. 1989. Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *Embo J*, 8(13):4163–70.

Schwer, B. and Meszaros, T. 2000. RNA helicase dynamics in pre-mRNA splicing. *Embo J*, 19(23):6582–91.

Sedgwick, H., Cameron, J. E., Poon, W. C. K., and Egelhaaf, S. U. 2007. Protein phase behavior and crystallization: effect of glycerol. *J Chem Phys*, 127(12):125102.

Seiwert, S. D. and Steitz, J. A. 1993. Uncoupling two functions of the U1 small nuclear ribonucleoprotein particle during in vitro splicing. *Mol Cell Biol*, 13(6):3135–3145.

Sharp, P. A. 1985. On the origin of RNA splicing and introns. *Cell*, 42(2):397–400.

Sheldrick, G. M. 2008. A short history of SHELX. *Acta Crystallogr A*, 64(Pt 1):112–122.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem*, 68(5):850–8.

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(21):8280–3.

Stahley, M. R. and Strobel, S. A. 2006. RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr Opin Struct Biol*, 16(3):319–326.

Staley, J. P. and Guthrie, C. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, 92(3):315–26.

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(6819):539–42.

Stark, H. and Lührmann, R. 2006. Cryo-electron microscopy of spliceosomal compo-
nents. *Annu Rev Biophys Biomol Struct*, 35:435–457.

Strasser, A., Dickmanns, A., Lührmann, R., and Ficner, R. 2005. Structural basis for
m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *EMBO
J*, 24(13):2235–2243.

Strasser, A., Dickmanns, A., Schmidt, U., Penka, E., Urlaub, H., Sekine, M., Lührmann,
R., and Ficner, R. 2004. Purification, crystallization and preliminary crystallographic
data of the m3G cap-binding domain of human snRNP import factor snurportin 1.
Acta Crystallogr D Biol Crystallogr, 60(Pt 9):1628–31.

Studier, F. W. 2005. Protein production by auto-induction in high density shaking
cultures. *Protein Expr Purif*, 41(1):207–234.

Szathmáry, E. 1999. The origin of the genetic code: amino acids as cofactors in an
RNA world. *Trends Genet*, 15(6):223–229.

Tarn, W. Y., Lee, K. R., and Cheng, S. C. 1993. Yeast precursor mRNA processing pro-
tein PRP19 associates with the spliceosome concomitant with or just after dissociation
of U4 small nuclear RNA. *Proc Natl Acad Sci U S A*, 90(22):10821–10825.

Tarn, W. Y. and Steitz, J. A. 1994. SR proteins can compensate for the loss of U1
snRNP functions in vitro. *Genes Dev*, 8(22):2704–17.

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

Terwilliger, T. 2004. SOLVE and RESOLVE: automated structure solution, density
modification and model building. *J Synchrotron Radiat*, 11(Pt 1):49–52.

Törö, 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(9):2293–2303.

Urlaub, H., Hartmuth, K., Kostka, S., Grelle, G., and Lührmann, R. 2000. A general
approach for identification of RNA-protein cross-linking sites within native human
spliceosomal small nuclear ribonucleoproteins (snRNPs). Analysis of RNA-protein
contacts in native U1 and U4/U6.U5 snRNPs. *J Biol Chem*, 275(52):41458–68.

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(1-2):187–96.

Vagin, A. and Teplyakov, A. 2000. An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr*, 56(Pt 12):1622–1624.

Valadkhan, S. 2007. The spliceosome: a ribozyme at heart? *Biol Chem*, 388(7):693–7.

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(12):3853–60.

van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F., and Grivell, L. A. 1986. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell*, 44(2):225–234.

Venrooij, W. J. V. 1987. The protein components of snRNPs in mammalian cells. *Molecular Biology Reports*, Volume 12, Number 3:151–153.

Wang, J. and Manley, J. L. 1997. Regulation of pre-mRNA splicing in metazoa. *Curr Opin Genet Dev*, 7(2):205–11.

Wassarman, D. A. and Steitz, J. A. 1992. Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. *Science*, 257(5078):1918–25.

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.

Weinstein, L. B. and Steitz, J. A. 1999. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol*, 11(3):378–84.

Will, C. and Lührmann, R. 1999. *RNA splicing. Encyclopedia of molecular biology*, volume 1-4, pages 2230–2417. John Wiley & Sons Inc.

Will, C. L. and Lührmann, R. 2005. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem*, 386(8):713–724.

Woolford, J. L. and Peebles, C. L. 1992. RNA splicing in lower eukaryotes. *Curr Opin Genet Dev*, 2(5):712–719.

Wu, J. Y. and Maniatis, T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, 75(6):1061–1070.

Xiao, S. H. and Manley, J. L. 1997. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev*, 11(3):334–344.

Yonath, A., Harms, J., Hansen, H. A., Bashan, A., Schlünzen, F., Levin, I., Koelln, I., Tocilj, A., Agmon, I., Peretz, M., Bartels, H., Bennett, W. S., Krumbholz, S., Janell, D., Weinstein, S., Auerbach, T., Avila, H., Piolleti, M., Morlang, S., and Franceschi, F. 1998. Crystallographic studies on the ribosome, a large macromolecular assembly exhibiting severe nonisomorphism, extreme beam sensitivity and no internal symmetry. *Acta Crystallogr A*, 54(Pt 6 Pt 1):945–955.

Zamore, P. D., Patton, J. G., and Green, M. R. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, 355(6361):609–14.

Zaug, A. J. and Cech, T. R. 1980. In vitro splicing of the ribosomal RNA precursor in nuclei of Tetrahymena. *Cell*, 19(2):331–338.

6 Appendix

6.1 Abbreviations

°C	degree Celsius
A	adenosine
Å	Ångstrom
ATP	adenosine triphosphate
Arg	arginine
Asp	aspartate
cm	centimeter
dATP	2'-deoxyadenosine 5'-triphosphate
DANO	anomalous difference structure factors
dCTP	2'-deoxycytosine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
DM	density modification
DNA	desoxyribonucleic acid
dTTP	2'-deoxythymidine 5'-triphosphate
e.g.	*exempli gratia*
EM	electron microscopy
et al.	*et altera*
FFT	fast fourier transform
FMS	free mounting system
FPLC	fast performance liquid chromatography
FT	flowthrough
G	guanosine
Glu	glutamate
GTP	guanosine triphosphate
GST	glutathione-S-transferase
h	hour

hnRNP	hetero nucleolar ribonucleo particle
IP	input
IPTG	isopropyl-β-D-thiogalactopyranosid
k	kilo
K	Kelvin
kDa	kilodalton
l	liter
Lys	lysine
M	marker
MALDI	Matrix Assisted Laser Desorption/Ionisation
mA	milliampere
mg	milligram
min	minute(s)
MIR	multiple isomorphous replacement
MIRAS	multiple isomorphous replacement using anomalous data
ml	milliliter
mm	millimeter
mM	millimolar
mRNA	messenger RNA
MW	molecular weight
NCS	non crystallographic symmetry
ng	nanogram
nl	nanoliter
nm	nanomolar
NMD	nonsense mediated decay
nmol	nanomol
NMR	Nucleic magnet resonance
nt	nucleotide
OD	optical density
PCI	phenol-chloroform-isopropanol
PCR	polymerase chain reaction
PHIC	calculated phases
Pro	proline
tRNA	transfer ribonucleic acid
rRNA	ribosomal ribonucleic acid

rpm	rounds per minute
RRM	RNA recognition motif
s	seconds
S	Svedberg
SAD	single-wavelength anomalous dispersion
SDS	sodium dodecyl sulphate
SF	splicing factor
SIR	single isomorphous replacement
SIRAS	single isomorphous replacement using anomalous data
SLE	systemic *lupus erythematosus*
snRNP	small nuclear ribonucleoprotein complex
snoRNA	small nucleolar ribonucleic acid
SPN1	snurportin 1
U	Unit
v	volume
V	volt(s)
µg	micro gram
µl	micro liter

6.2 Acknowledgements

I would like to thank Prof. Markus Wahl and Prof. Reinhard Lührmann for the constant support, encouragement, advice and a great atmosphere in the department. Especially, I would like to thank Markus for the supervision of this external PhD thesis, constant support and many fruitful discussions. I also would like to thank Prof. R. Ficner and Prof. O. Einsle for the supervision of this external Ph.D. thesis and taking part in the thesis committee. Furthermore I would like to thank Prof. S. Pöggeler, Prof. M. Oppermann, Dr. W. Kramer and Dr. R. Daniel for being members of my extended thesis committee.

Especially, I would like to thank Thomas Conrad, Peter Kempkes and Hossein Kohansal for HeLa cell fermentation and the preparation of HeLa nuclear extract. I am grateful to Elke Penka for constant support with cloning. I would like to thank Prof. Holger Stark and his group for sharing their expertise in electron microscopy. I would like to thank Dr. Henning Urlaub, Uwe Plessmann and Monika Raabe for mass spectrometry support.

I would like to thank all former and present members of the department of cellular biochemistry and especially the X-ray crystallography group for a great atmosphere. I would like to thank Christian Stegmann for maintaining all crystallographic software. I am grateful to Prof. S. Urshumtsev and Dr. Gleb Bourenkov for their help with low resolution and medium resolution phasing approaches. I would like to thank the beamline staff at SLS/PX1 and PX2 for their constant support. Furthermore I would like to say thank you to Hanne and Ron Smith, Gudrun Weber, Kristin Sauer and Karine Fernanda dos Santos for corrections.

I would like to thank Simon Trowitzsch, for several years of collaboration, friendship and fruitful discussions during my entire studies. I am especially grateful to my parents for their unconditional support during my education. I thank my brothers, Gunter and especially Georg, who encouraged me to write the thesis with LaTeX and for constant help with the program.

I would finally like to thank Imke for her love and support.

6.3 Curriculum Vitae

Gert Weber

Dipl. Biol.

Contact address

Am Steinsgraben 13

37085 Göttingen

Germany

Date of birth: 21.12.1977

Place of birth: Homberg/Efze

Tel.: 0551/5314233

Email: gweber@gwdg.de

Education

since 2004	Ph.D. thesis
	„Purification and crystallization of spliceosomal snRNPs"
	Max-Planck-Institute of Biophysical Chemistry, Göttingen
	X-ray crystallography group
2004	Diploma (sehr gut)
	Max-Planck-Institute of Biophysical Chemistry, Göttingen
	X-ray crystallography group
1998 – 2004	Diploma studies in Biology
	Georg-August-University Göttingen
1997 – 1998	Military service
1988 – 1997	Abitur (Grade: 1.7)
	Bundespräsident-Theodor-Heuss-Schule Homberg/Efze

Practical education

2006	Tutor of the international Master/PhD program
	molecular biology", Georg-August-University, Göttingen
2006	Hercules specialized course (HSC2) 9.-13.10.2006, "structural
	genomics", ESRF/EMBL Grenoble, France
2007	Tutor of the international Master/PhD program
	"molecular biology", Georg-August-University, Göttingen
2008	Tutor of the international Master/PhD program of
	"molecular biology", Georg-August-University, Göttingen

Publications Gert Weber

Trowitzsch, S., Weber ,G., Lührmann, R., and Wahl, M. 2008a. An unusual RNA recognition motif acts as a scaffold for multiple proteins in the pre-mRNA retention and splicing complex. *manuscript in preparation*.

Trowitzsch, S., Weber, G., Lührmann, R., and Wahl, M. 2008b. Crystal structure of the Pml1p subunit of the yeast pre-mRNA retention and splicing complex. *manuscript in preparation*.

Luo, X., Hsiao, H.-H., Bubunenko, M., Weber, G., Court, D., Gottesman, M., Urlaub, H., and Wahl, M. 2008. Structural and Functional Analysis of E. coli NusB-S10 Transcription Antitermination Complex. *submitted*.

Andresen, M., Stiel, A. C., Trowitzsch, S., Weber, G., Eggeling, C., Wahl, M. C., Hell, S. W., and Jakobs, S. 2007. Structural basis for reversible photoswitching in Dronpa. *Proc Natl Acad Sci U S A*, 104(32):13005-13009.

Stiel, A. C., Trowitzsch, S., Weber, G., Andresen, M., Eggeling, C., Hell, S. W., Jakobs, S., and 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*, 402(1):35-42.

Savarese, E., Chae, O., Trowitzsch, S., Weber, G., Kastner, B., Akira, S., Wagner, H., Schmid, R. M., Bauer, S., and Krug, A. 2006. U1 small nuclear ribonucleoprotein immune complexes induce type I interferon in plasmacytoid dendritic cells through TLR7. *Blood*, 107(8):3229–3234.

Andresen, M., Wahl, M. C., Stiel, A. C., Gräter, F., Schäfer, L. V., Trowitzsch, S., Weber, G., Eggeling, C., Grubmüller, H., Hell, S. W., and Jakobs, S. 2005. Structure and mechanism of the reversible photoswitch of a fluorescent protein. *Proc Natl Acad Sci U S A*, 102(37):13070-13074.