

Aus dem Max-Planck-Institut für Biophysikalische Chemie in Göttingen

Abteilung Zelluläre Biochemie

Direktor: Prof. Dr. Reinhard Lührmann

**The Molecular Architecture and Structure of the Human
Prp19/CDC5L Complex and 35S U5 snRNP**

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August-Universität zu Göttingen

vorgelegt von

Michael Grote

aus Bottrop

Göttingen 2010

D7

Referent: Prof. Dr. Ralf Ficner

Korreferent: Prof. Dr. Ernst A. Wimmer

Tag der mündlichen Prüfung: 16. Februar 2011

Meiner Familie

TABLE OF CONTENTS

TABLE OF CONTENTS	I
LIST OF FIGURES	VI
LIST OF TABLES	VIII
ABSTRACT	IX
1 INTRODUCTION	1
1.1 The protein coding regions of eukaryotic genes are split	1
1.2 Pre-mRNA splicing is catalyzed by the spliceosome	2
1.3 The spliceosome assembles stepwise in a highly dynamic manner	4
1.4 The spliceosome is a particularly protein-rich machinery	8
1.4.1 Roles of proteins in pre-mRNA splicing	10
1.5 Prp19 is part of a heteromeric complex in yeast and human	11
1.6 The Prp19 complex plays a key role in the catalytic activation of human and yeast spliceosomes	14
1.7 The 35S U5 snRNP appears to be a product of spliceosome disassembly	16
1.8 Towards the structure of the spliceosome	18
1.9 Aim of this study	21
2 MATERIALS AND METHODS	23
2.1 Materials	23
2.1.1 Chemicals	23
2.1.2 Oligonucleotides	26

2.1.3	Plasmids	27
2.1.4	Enzymes and enzyme buffers	28
2.1.5	Peptides	29
2.1.6	Antibodies	29
2.1.7	Bacterial strains	30
2.1.8	Cell lines	30
2.1.9	Common buffers	31
2.1.10	Commercial reagent sets	34
2.1.11	Working equipment	35
2.1.12	Instruments	36
2.1.13	Software	38
2.2	Methods	39
2.2.1	Molecular biological standard methods	39
2.2.1.1	Concentration determination of nucleic acids	39
2.2.1.2	Agarose gel electrophoresis of DNA	39
2.2.1.3	Phenol/Chloroform/Isoamyl alcohol extraction	39
2.2.1.4	Precipitation of nucleic acids	40
2.2.1.5	Purification of DNA from agarose gels	40
2.2.2	Molecular cloning in plasmid vectors	40
2.2.2.1	Polymerase chain reaction (PCR)	40
2.2.2.2	Endonucleolytic restriction of DNA	41
2.2.2.3	DNA ligation	41
2.2.2.4	Preparation of electrocompetent <i>E. coli</i> cells	42
2.2.2.5	Electrotransformation	42
2.2.2.6	Plasmid purification	43
2.2.2.7	Sequencing	43
2.2.3	RNA biochemical standard methods	43
2.2.3.1	Denaturing polyacrylamide gel electrophoresis of RNA	43
2.2.3.2	Silver staining of RNA in polyacrylamide gels	44
2.2.3.3	Autoradiography of radioactively labeled RNA in polyacrylamide gels	44
2.2.3.4	Northern blot	44
2.2.3.5	Purification of RNA from polyacrylamide gels	45
2.2.3.6	Purification of RNA by gel filtration	45
2.2.3.7	Proteinase K digestion	46
2.2.4	<i>In vitro</i> transcription of radiolabeled pre-mRNA	46

2.2.4.1	DNA template production by PCR	46
2.2.4.2	<i>In vitro</i> transcription reaction	47
2.2.5	Protein biochemical standard methods	47
2.2.5.1	Concentration determination of proteins in solution	47
2.2.5.2	Precipitation of proteins	47
2.2.5.3	SDS polyacrylamide gel electrophoresis of proteins	48
2.2.5.4	Coomassie Brilliant Blue staining of proteins in polyacrylamide gels	48
2.2.5.5	Silver staining of proteins in polyacrylamide gels	48
2.2.5.6	Sypro Ruby staining of proteins in polyacrylamide gels	49
2.2.5.7	Semi-dry blotting of proteins from polyacrylamide gels	49
2.2.5.8	Wet blotting of proteins from polyacrylamide gels	49
2.2.6	<i>In vitro</i> translation	50
2.2.6.1	Preparation of a pEU3-based vector for protein expression and purification	50
2.2.6.2	Preparation of template mRNA	50
2.2.6.3	<i>In vitro</i> translation of proteins in a wheat germ-based system	51
2.2.6.4	Liquid scintillation counting	51
2.2.7	Immunological methods	52
2.2.7.1	Affinity purification of anti-peptide antibodies	52
2.2.7.2	Western blotting	52
2.2.7.3	Far western analysis	53
2.2.8	Cell culture	54
2.2.8.1	Cultivation of HeLa cells	54
2.2.8.2	Freezing of HeLa cells	54
2.2.8.3	Establishing HeLa stable cell lines	55
2.2.8.4	Fluorescence microscopy	56
2.2.8.5	Preparation of HeLa nuclear extract	56
2.2.9	<i>In vitro</i> splicing reaction	57
2.2.10	Affinity-based purification methods	57
2.2.10.1	Purification of the MS2-MBP fusion protein	57
2.2.10.2	Pulldown analysis of protein-protein interactions	58
2.2.10.3	Purification of the hPrp19/CDC5L complex from HeLa nuclear extract	59
2.2.10.4	Purification of the 35S U5 snRNP from HeLa nuclear extract	60
2.2.10.5	MS2 affinity selection of spliceosomal C complexes	61
2.2.11	Special Methods	61
2.2.11.1	[¹⁴ C]Iodoacetamide labeling of hPrp19/CDC5L complex proteins	61

2.2.11.2	Dissociation of native hPrp19/CDC5L complexes	62
2.2.11.3	Sulfo-MBS crosslinking in native hPrp19/CDC5L complexes	62
2.2.11.4	Limited proteolysis of native hPrp19/CDC5L complexes	62
2.2.11.5	Mass spectrometry	63
2.2.11.6	Electron microscopy	65
2.2.11.7	X-ray crystallography	68
2.2.11.8	Analytical ultracentrifugation	70
3	RESULTS	72
3.1	The molecular architecture of the human Prp19/CDC5L complex	72
3.1.1	Establishing HeLa cell lines stably expressing FLAG/HA-AD002 or FLAG/HA-SPF27	72
3.1.2	Isolation of the human Prp19/CDC5L complex	75
3.1.3	Analysis of the stoichiometry of the native hPrp19/CDC5L complex reveals hPrp19 tetramerization	77
3.1.4	hPrp19, CDC5L, PRL1 and SPF27 form a salt-stable core of the hPrp19/CDC5L complex	80
3.1.5	Crosslinking delineates protein-protein interactions within the native hPrp19/CDC5L complex	84
3.1.6	Analysis of protein-protein interactions between hPrp19/CDC5L components with <i>in vitro</i> translated proteins	87
3.1.7	Limited proteolysis of the native hPrp19/CDC5L complex delineates protein interaction domains	91
3.1.8	Electron microscopic analysis of the human Prp19/CDC5L complex	98
3.1.9	Establishing stable cell lines expressing an eGFP-AD002 fusion protein	99
3.2	The crystal structure of the WD40 domain of human Prp19	102
3.2.1	Structure determination	103
3.2.2	Overall structure description	105
3.2.3	Comparison with other WD repeat proteins	108
3.2.4	Identification of a peptide bound to the β propeller	110
3.2.5	Surface analysis of the β propeller and location of functional sites	112
3.3	Purification and EM structure the human 35S U5 snRNP	114
3.3.1	Purification and MS analysis of the human 35S U5 snRNP	114
3.3.2	EM analysis of the human 35S U5 snRNP	119
4	DISCUSSION	121

4.1	Establishing HeLa stable cell lines	121
4.2	Purification of the hPrp19/CDC5L complex	122
4.3	The stoichiometry of the hPrp19/CDC5L complex	124
4.4	Identification of a highly stable core of the hPrp19/CDC5L complex	125
4.5	Elucidation of domains of the hPrp19/CDC5L proteins involved in protein-protein interactions	126
4.6	The human Prp19/CDC5L complex has an asymmetric and elongated morphology	127
4.7	Insight into a major building block of the RNP core of the step I spliceosome	128
4.8	The crystal structure of the WD40 domain of hPrp19	129
4.9	Purification of the 35S U5 snRNP	131
4.10	The proteome of the 35S U5 snRNP resembles the composition of the salt-stable RNP core of the spliceosome	132
4.11	Comparison of the EM structure of the 35S U5 snRNP with that of the spliceosomal C complex	134
5	REFERENCES	136
6	APPENDIX	148
6.1	Abbreviations	148
6.2	Acknowledgements	150
6.3	Erklärung/Affidavit	152

LIST OF FIGURES

Fig. 1.1: Consensus sequences of mammalian pre-mRNAs.	2
Fig. 1.2: Schematic illustration of the two-step mechanism of pre-mRNA splicing.	3
Fig. 1.3: Schematic representation of the composition of the human spliceosomal U snRNPs.	5
Fig. 1.4: Assembly/disassembly pathway of the spliceosome.	6
Fig. 1.5: Dynamic network of RNA interactions in the spliceosome.	7
Fig. 1.6: Compositional dynamics of human spliceosomal A, B, and C complexes.	9
Fig. 1.7: Comparison of the human Prp19/CDC5L complex with the yeast Nineteen complex (NTC).	12
Fig. 1.8: Stable interaction of the hPrp19/CDC5L complex with the U5 snRNP during the splicing cycle.	15
Fig. 1.9: The U4/U6.U5 tri-snRNP is extensively remodeled during splicing.	16
Fig. 1.10: Comparison of the protein composition of the human 35S U5 snRNP with the 1 M salt treated RNP core of the C complex.	17
Fig. 1.11: 3D reconstitutions of spliceosomal complexes.	19
Fig. 1.12: Functional interpretation of spliceosomal complexes by EM.	20
Fig. 2.1: Illustration of the GraFix procedure.	66
Fig. 3.1: Characterization of HeLa cell lines stably expressing FLAG/HA-tagged AD002 or SPF27.	73
Fig. 3.2: The presence of a FLAG/HA-tag does not affect pre-mRNA splicing.	74
Fig. 3.3: Purification of the hPrp19/CDC5L complex from HeLa cell lines stably expressing FLAG/HA-tagged AD002 or SPF27.	76
Fig. 3.4: Size exclusion chromatography of affinity purified hPrp19/CDC5L complexes.	77
Fig. 3.5: Determination of the protein stoichiometry of the native human Prp19/CDC5L complex reveals four copies of hPrp19.	78
Fig. 3.6: Analytical ultracentrifugation of the hPrp19/CDC5L complex reveals its molecular mass.	80
Fig. 3.7: CDC5L, PRL1, hPrp19 and SPF27 form a salt-stable subcomplex.	82
Fig. 3.8: Analysis of the salt stability of the human Prp19/CDC5L complex.	83
Fig. 3.9: Protein crosslinking in purified hPrp19/CDC5L complex with different crosslinking reagents.	84
Fig. 3.10: Schematic representation of protein crosslinking with the heterobifunctional crosslinking agent Sulfo-MBS.	85
Fig. 3.11: Identification of protein-protein interactions in the hPrp19/CDC5L complex via crosslinking.	86
Fig. 3.12: Plasmids used for the cell-free translation of proteins in wheat germ lysate.	88
Fig. 3.13: Far western analysis demonstrates interactions among the hPrp19/CDC5L complex core components and between AD002 and CTNNBL1, and CTNNBL1 and CDC5L.	89

Fig. 3.14: Co-immunoprecipitation confirms interactions between SPF27 and both CDC5L and hPrp19, between CTNNB1 and AD002, and between CDC5L and hPrp19.	90
Fig. 3.15: Schematic overview of the protein-protein interactions within the hPrp19/CDC5L complex.	90
Fig. 3.16: Limited proteolysis of the hPrp19/CDC5L complex.	91
Fig. 3.17: Limited protease digestion delineates those domains of hPrp19/CDC5L complex proteins involved in protein-protein interactions.	93
Fig. 3.18: Exhaustive MS analyses of hPrp19/CDC5L complex proteins after limited proteolysis and subsequent trypsin digestion.	94
Fig. 3.19: Limited proteolysis of the hPrp19/CDC5L complex with increasing amounts of Subtilisin.	96
Fig. 3.20: Schematic representation of the topology of the hPrp19/CDC5L complex.	97
Fig. 3.21: Electron microscopy of native human Prp19/CDC5L complexes.	98
Fig. 3.22: Characterization of HeLa cell lines stably expressing FLAG/StrepII-tagged eGFP-AD002.	100
Fig. 3.23: FLAG/StrepII-tagged eGFP-AD002 colocalizes with SC35 in nuclear speckles.	101
Fig. 3.24: Purification of the hPrp19/CDC5L complex from HeLa cell lines stably expressing FLAG/StrepII-tagged eGFP-AD002.	102
Fig. 3.25: Distribution of the seven WD repeats in the sequence of hPrp19.	103
Fig. 3.26: Purification and crystallization of N-terminally truncated hPrp19.	104
Fig. 3.27: Ramachandran map of the hPrp19 WD40 domain.	105
Fig. 3.28: Overall structure of the hPrp19 WD40 domain.	106
Fig. 3.29: Sequence alignment of Prp19 proteins from different species.	107
Fig. 3.30: Sequence alignment of different WD40 repeat domains.	108
Fig. 3.31: Superposition of the hPrp19 WD40 domain with other WD40 repeat-containing proteins reveals overall similarity.	109
Fig. 3.32: The two “structural tetrads” of the β propeller of hPrp19.	109
Fig. 3.33: X-ray structure comparison of the WD40 domains from human and yeast Prp19.	110
Fig. 3.34: Electron density map for the hPrp19 WD40 domain.	111
Fig. 3.35: The electron density of the additional peptide and the very N-terminus of the hPrp19 WD40 domain is not well defined.	112
Fig. 3.36: The surface of the hPrp19 WD40 domain.	113
Fig. 3.37: Purification of the human 35S U5 snRNP from HeLa cell lines stably expressing FLAG/HA-tagged AD002.	116
Fig. 3.38: The 35S U5 snRNP also contains low amounts of U6 snRNA.	117
Fig. 3.39: Proteins present in the 35S U5 snRNP in significant amounts.	119
Fig. 3.40: Electron microscopy of native human 35S U5 snRNPs.	120
Fig. 3.41: 3D EM structure of the human 35S U5 snRNP.	120
Fig. 4.1: Comparison of the protein and RNA composition of the catalytically active step I spliceosome, the salt-stable RNP core of the C complex and the 35S U5 snRNP.	133
Fig. 4.2: Comparison of the 3D EM structures of the 35S U5 snRNP with the C complex.	134

LIST OF TABLES

Table 2.1: Molecular weight (MW) and V_{bar} values of the hPrp19/CDC5L complex proteins.	71
Table 3.1: Summary of the crystallographic analysis.	104
Table 3.2: Protein composition of the human 35S U5 snRNP.	118

ABSTRACT

Pre-mRNA splicing, the two consecutive transesterification reactions leading to intron removal and exon ligation, is catalyzed by the spliceosome, a highly dynamic, multi-megadalton molecular machinery. The spliceosome is comprised of the U1, U2, U4, U5, and U6 small nuclear ribonucleoprotein particles (snRNPs), plus many non-snRNP proteins. Spliceosome assembly occurs in a stepwise and highly dynamic manner. Initially, the U1 snRNP binds the 5' splice site, followed by the ATP-dependent recognition of the pre-mRNA's branch point sequence by the U2 snRNP, forming the prespliceosome or A complex. The assembly of snRNPs on the pre-mRNA is completed by the addition of the U4/U6.U5 tri-snRNP, generating the pre-catalytic B complex, which is still catalytically inactive. In order to catalyze the first step of splicing, the spliceosome must undergo dramatic compositional and structural remodeling events, culminating in the destabilization of the U1 and U4 snRNPs and the formation of the catalytically activated spliceosome (B* complex). The first transesterification reaction then occurs, generating the C complex, which in turn catalyzes the second step of splicing. After catalysis, the spliceosome dissociates, releasing the mRNA, U2 and U6 snRNPs and the U5 snRNP in the form of a post-spliceosomal 35S U5 snRNP.

The spliceosome contains numerous non-snRNP proteins, many of which play essential roles during splicing and are recruited to the spliceosome in the form of a heteromeric protein complex. One such protein is Prp19, an evolutionary highly conserved splicing factor required for the activation of the spliceosome. Prp19 is present in cells as part of a stable heteromeric complex both in yeast (i.e., the nineteen complex, NTC) and in humans (i.e., the hPrp19/CDC5L complex) and Prp19 and its related proteins are major components of the spliceosome's catalytic core RNP. To learn more about the spatial organization of the human Prp19/CDC5L complex, we have affinity-purified preparative amounts of native hPrp19/CDC5L complexes from HeLa cell lines stably expressing FLAG-tagged AD002 or SPF27, both stable components of this complex. Consistent with previous results from our laboratory, mass spectrometric (MS) analysis of the affinity-purified hPrp19/CDC5L complexes revealed they contain seven proteins, namely hPrp19, CDC5L, PRL1, AD002, SPF27, CTNNBL1, and HSP73. Subsequent analysis of the stoichiometry of the hPrp19/CDC5L complex by analytical ultracentrifugation, fluorescent staining of hPrp19/CDC5L complex proteins after SDS-PAGE and [¹⁴C]iodoacetamide labeling indicated that it contains four

copies of the hPrp19 protein and likely single copies of all other components. Furthermore, analytical ultracentrifugation revealed that the hPrp19/CDC5L complex has a frictional ratio value of 2.1, indicating that the particle's shape is elongated.

The molecular organization of the hPrp19/CDC5L complex was analyzed in more detail by treatment of the complex with high concentrations of NaCl, NaSCN, or heparin. These studies revealed a highly salt resistant core composed of hPrp19, CDC5L, PRL1 and SPF27. Furthermore, fractions of AD002 and CTNNBL1, as well as CTNNBL1 and HSP73 appeared to co-migrate after salt treatment, suggesting that these proteins interact with each other. A complex protein-protein interaction network involving the hPrp19/CDC5L complex core proteins, as well as the less stably associated AD002 and CTNNBL1 proteins, was also elucidated by crosslinking studies with purified hPrp19/CDC5L complexes or by far western blotting and pulldown experiments with *in vitro* translated hPrp19/CDC5L complex proteins. Further probing of the structural organization of the native hPrp19/CDC5L complex by limited proteolysis confirmed our salt dissociation data and showed that full-length SPF27, the C-terminus of CDC5L, and the N-termini of PRL1 and hPrp19 also form a protease-resistant core complex. CTNNBL1, HSP73 and the N-terminus of CDC5L co-fractionated after protease treatment, and an interaction between AD002 and CTNNBL1 was identified by co-purification and crosslinking. Thus, association of these less stably bound hPrp19/CDC5L subunits appears to be mediated by contacts between CTNNBL1 and CDC5L.

The human Prp19 protein contains at least two recognized domains: an N-terminal U-box domain and a C-terminal WD40 repeat domain. Limited proteolysis of the native hPrp19/CDC5L complex allowed the isolation of the WD40 domain of hPrp19 and we subsequently determined its crystal structure at 1.8 Å resolution. This WD40 domain folds into a canonical seven-bladed β-propeller structure with each blade composed of four antiparallel β sheets.

Finally, by performing negative stain electron microscopy (EM), we analyzed the overall 2D structure of the hPrp19/CDC5L complex. Consistent with the data obtained from analytical ultracentrifugation, purified hPrp19/CDC5L complexes exhibit an elongated, asymmetric shape with a maximum dimension of ~20 nm. Considering that the hPrp19/CDC5L complex contains an hPrp19 tetramer but apparently single copies of the other components of the complex, the asymmetric structure of the complex suggests an asymmetric assembly pathway. Together, our findings on the hPrp19/CDC5L complex not only elucidate the molecular organization of the hPrp19/CDC5L complex but also provide insights into potential protein-protein interactions at the core of the catalytically active spliceosome and additionally pave the way for future functional and high-resolution structural analyses of this essential complex that is part of the spliceosome's catalytically active RNP core.

After completion of the splicing reaction, the spliceosome disassembles and the hPrp19/CDC5L complex dissociates from the spliceosome as part of the post-spliceosomal 35S U5 snRNP. To learn more about the structure of the human 35S U5 snRNP, we isolated this snRNP from HeLa cells stably expressing FLAG-tagged AD002. MS analyses revealed a protein composition similar to that previously described for this complex, which includes most U5 snRNP proteins, proteins of the hPrp19/CDC5L complex plus its related proteins and several additional proteins that are known components of the spliceosomal C complex. Thus, the 35S U5 snRNP contains a set of proteins highly similar to that of the salt-stable core of the spliceosomal C complex. 2D EM of negatively stained 35S U5 snRNP particles revealed that they have a maximum dimension of about 27 nm and appear typically trapezoidal or triangular. A 3D reconstruction of the 35S U5 snRNP at a resolution of 2.4-2.8 nm was also generated by performing unstained cryo-EM. Subsequent comparison of the EM structure of the 35S U5 snRNP, with that of the native, spliceosomal C complex revealed striking similarities between both complexes and allowed the localization of functionally important domains of the step I spliceosome. Thus, the structural information obtained for the human 35S U5 snRNP, as well as the hPrp19/CDC5L complex, not only have contributed to the localization of functionally important components of the step I spliceosome, but also provide first insights into the molecular architecture of part of the spliceosome's core domain.

1 INTRODUCTION

1.1 The protein coding regions of eukaryotic genes are split

The most fundamental process in molecular biology is gene expression, the transfer of information between the three major biopolymers DNA, RNA, and protein (Crick, 1970). With the exception of RNA viruses, the genetic information in all organisms is encoded within a DNA genome in distinct units, called genes. In the most general case, information from a DNA gene is transcribed into mRNA, which is further translated into the amino acid sequence of a protein.

In contrast to prokaryotes, a gene in higher organisms can be split (Berget et al., 1977; Chow et al., 1977). In this case, the protein coding sequences of DNA, the exons, are interrupted by sequences that do not contain protein information, the introns. Such a gene resembles a mosaic (Roberts and Sharp, 1997). In order to enable a split gene to direct the synthesis of a protein, a special genetic mechanism is required to physically connect the exon encoded information. This is achieved by the process of pre-mRNA splicing that takes place in the nucleus. The pre-mRNA is a conform copy of a gene and contains both exons and introns. During nuclear pre-mRNA splicing the introns are removed from the pre-mRNA sequence and the flanking exons are joined to form the spliced mRNA. Lower eukaryotes, such as yeast, have only few and usually small introns. In contrast, in higher metazoans introns can be as long as hundreds of thousands of nucleotides and the coding sequence of a gene is often distributed over several exons. The multiple intron-exon structure of a gene renders the intriguing possibility to alternatively select different combinations of exons at the stage of pre-mRNA splicing to generate multiple protein isoforms (reviewed in Blencowe, 2006). Thus, alternative splicing greatly increases the proteomic and functional diversity of metazoan organisms. Recent studies with high-throughput sequencing indicated that the extent of alternative splicing in humans is much greater than previously expected. These studies estimate that over 90% of all multi-exon genes are alternatively spliced (Pan et al., 2008; Wang et al., 2008). The choice between two different splicing patterns is highly regulated and depends on cell type, developmental state or in response to external stimuli. Indeed, alternative splicing is an essential component of gene regulation. Furthermore, defects in splicing are the cause or a severity modifier of a growing number of diseases (reviewed in Cooper et al., 2009).

Pre-mRNA splicing is performed in two consecutive, energy independent S_N2 -type transesterification reactions (Moore and Sharp, 1993). As illustrated in Fig. 1.2, in the first step of splicing, the phosphodiester bond at the 5' ss is nucleophilically attacked by the 2'-hydroxyl group of an adenosine of the BPS in the intron, generating a free 5' exon and an intron lariat-3' exon, the so-called splicing intermediates. In the second step, the 3'-hydroxyl group of the free 5' exon nucleophilically attacks the phosphodiester bond at the 3' ss, which leads to exon ligation and excision of the intron in a lariat conformation.

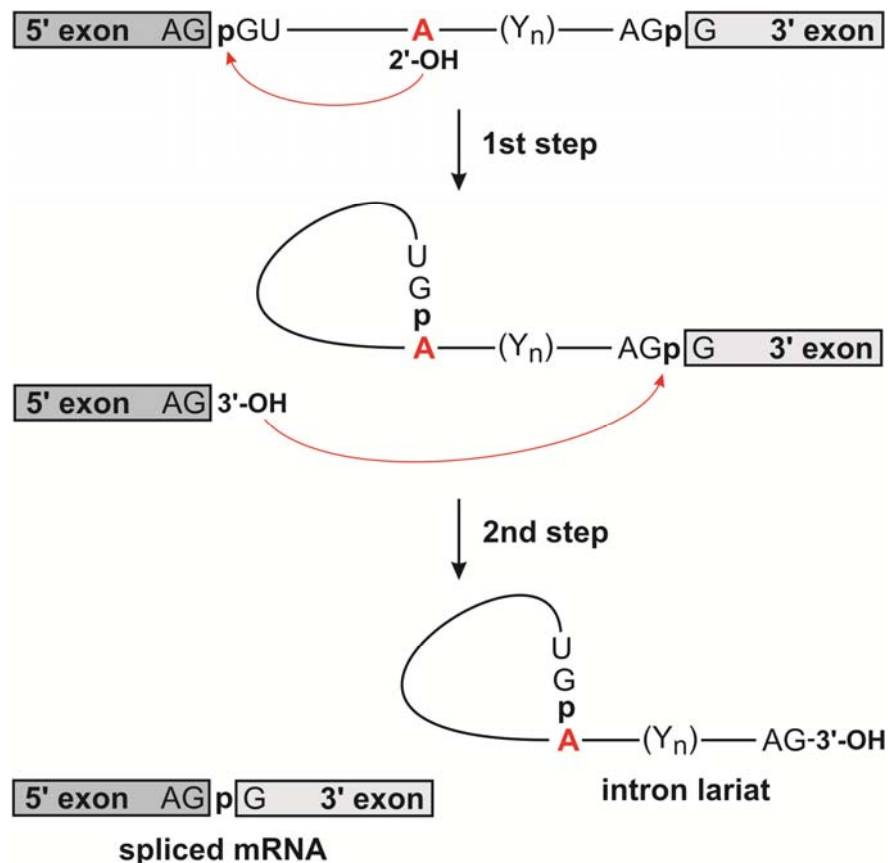


Fig. 1.2: Schematic illustration of the two-step mechanism of pre-mRNA splicing. In the first step, the 2' hydroxyl group of the conserved branchpoint adenosine attacks the phosphate at the 5' ss resulting in a free 5' exon containing a 3' terminal hydroxyl group and a branched lariat intermediate which contains the intron and 3' exon. In the second step, the 3' hydroxyl group of the free 5' exon attacks the phosphate at the 3' ss, thereby ligating the 5' and 3' exons and removing the intron lariat. The nucleophilic attack in both steps is indicated by red arrows. The consensus sequences are only partially shown (see Fig. 1.1), with the branch point adenosine highlighted in red. The reactive groups of the pre-mRNA are indicated with the bold letter "p" for the phosphate groups, and "OH" for the hydroxyl group, together with their position on the ribose.

The splicing reaction is catalyzed by the spliceosome, a highly dynamic, multimegadalton molecular machinery (reviewed in Wahl et al., 2009). The main subunits of the spliceosome are the uridine-rich small nuclear ribonucleoprotein particles (U snRNPs). Spliceosomes additionally contain numerous non-snRNP proteins (see section 1.4). The major spliceosome is composed of the U1, U2,

U4, U5, and U6 snRNP. The latter two are usually found together as the U4/U6 di-snRNP, which stably associates with the U5 snRNP to form the U4/U6.U5 tri-snRNP prior to spliceosome assembly. Each snRNP consists of a unique RNA moiety, the snRNA, and a variable set of particle-specific proteins (reviewed in Will and Lührmann, 2006). In addition, the U1, U2, U4, and U5 snRNPs all contain seven Sm-proteins (B/B', D1, D2, D3, E, F, and G) that form a heptameric ring complex around the Sm binding site of each snRNA (reviewed in Will and Lührmann, 2001). The U6 snRNA does not contain a conventional Sm binding site but instead associates with a set of seven Sm-like proteins (LSm2-LSm8) (Achsel et al., 1999). The composition of the snRNPs is depicted in Fig. 1.3. The U1 snRNP contains three particle-specific proteins, U1-70K, U1-A and U1-C. Glycerol gradient centrifugation revealed a sedimentation coefficient for the U1 snRNP of 12 S. The 17S U2 snRNP is comprised of the particle-specific proteins U2-A' and U2-B'', and the heteromeric protein complexes SF3a and SF3b, which contain subunits with molecular weights of 120, 66 and 60 kDa, and 155, 145, 130, 49, 14 and 10 kDa, respectively. In the 13S U4/U6 di-snRNP, particle specific proteins with apparent molecular weights of 15.5, 61, 20, 60, and 90 kDa are found. The list of 20S U5-specific proteins includes proteins of 220, 200, 116, 102, 100, 52, 40, and 15 kDa. With exception of the U5-52 kDa protein, all proteins of the 20S U5 snRNP and 13S U4/U6 di-snRNP are also part of the U4/U6.U5 tri-snRNP. Furthermore, upon association of the U4/U6 di-snRNP and the U5 snRNP, three additional proteins with molecular weights of 27, 65, and 110 kDa bind to complete the formation of the 25S U4/U6.U5 tri-snRNP.

1.3 The spliceosome assembles stepwise in a highly dynamic manner

Whereas ribosomal subunits possess a preformed active center, none of the snRNPs contains the active site responsible for the catalysis of pre-mRNA splicing by the spliceosome (reviewed in Wahl et al., 2009). In contrast, during the stepwise process of spliceosome assembly, the active site evolves *de novo* on each pre-mRNA. Due to the ordered interaction of the snRNPs and other factors with the pre-mRNA several assembly intermediates of the human spliceosome can be distinguished, including the E, A, B, B*, and C complexes (Fig. 1.4).

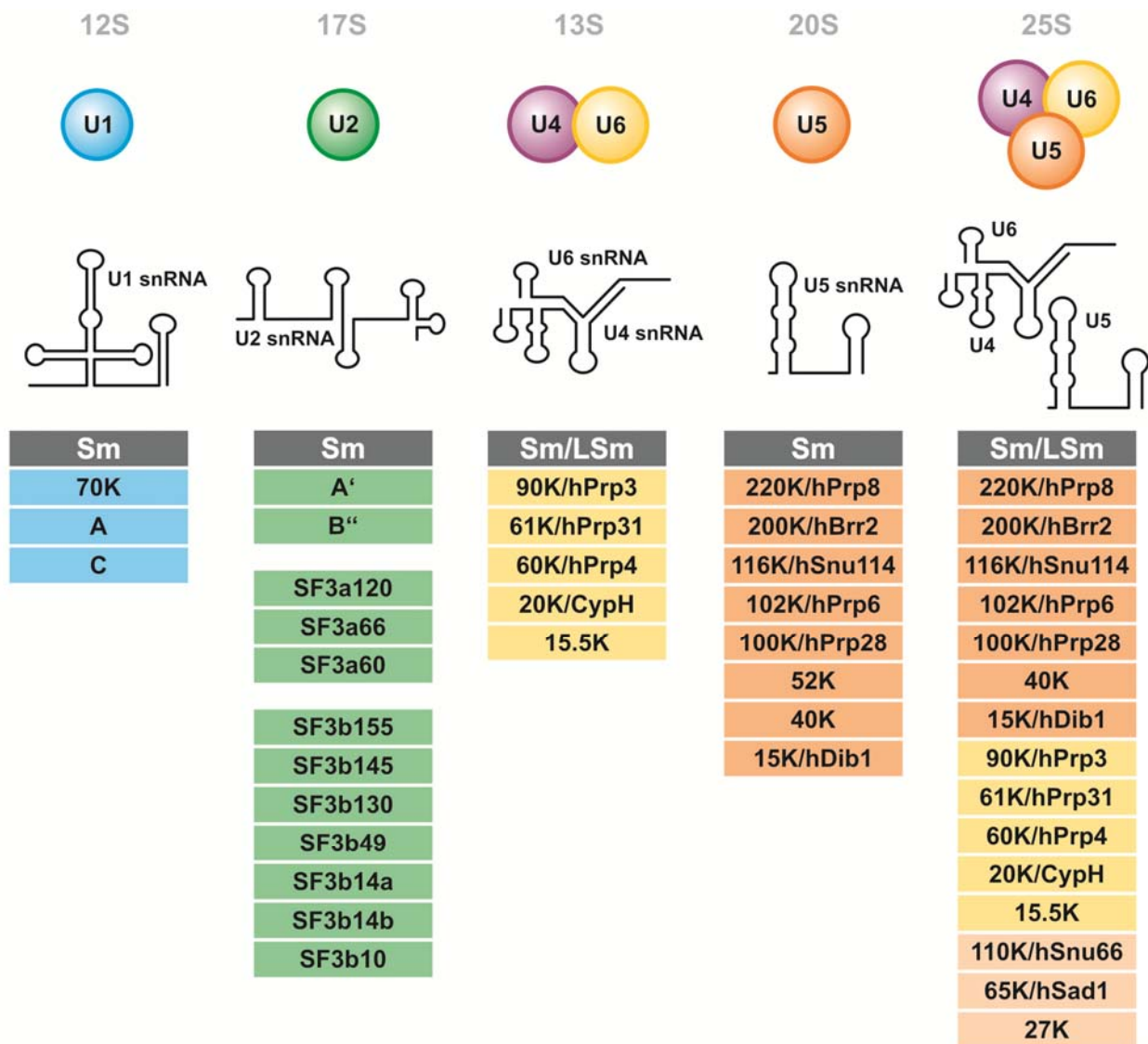


Fig. 1.3: Schematic representation of the composition of the human spliceosomal U snRNPs. Each snRNP is composed of a unique U snRNA (illustrated by a black line in its representative secondary structure) together with a particle-specific set of proteins (listed below each snRNA in a particle specific color-code). The sedimentation coefficient in Svedberg units (S) of each snRNP is shown on top. The U4 and U6 snRNPs are illustrated in the form of the 13S U4/U6 di-snRNP and together with the U5 snRNP in the 25S U4/U6.U5 tri-snRNP. Sm and LSm indicate the whole set of Sm- or LSm-proteins, respectively. The letter K indicates the apparent molecular weight of the certain protein in kilodalton (K and kDa are interchangeably used throughout the text).

Whereas the splicing reaction as such does not require an external source of energy, many steps during the incremental association and rearrangement of spliceosomal factors on the pre-mRNA are driven by the hydrolysis of ATP. At the beginning of spliceosome assembly, the U1 snRNP binds the pre-mRNA in an ATP-independent manner through base pairing interactions of the 5' end of the U1 snRNA to the 5' ss of the pre-mRNA, generating the early (E) complex, which also contains loosely-attached U2 snRNP. In a subsequent step, ATP-dependent base pairing of U2 snRNA with the pre-mRNA's BPS takes place and the pre-spliceosome or A complex is formed, in which the U2

snRNP is now stably associated. The assembly of snRNPs on the pre-mRNA is completed by the recruitment of the preformed U4/U6.U5 tri-snRNP, which generates the pre-catalytic B complex that is still catalytically inactive. In order to facilitate the first step of splicing, the spliceosome must undergo dramatic compositional and structural remodeling events, culminating in the destabilization and dissociation of the U1 and U4 snRNPs and the formation of the activated spliceosome (B* complex). The first transesterification reaction then occurs generating the spliceosomal C complex, which in turn catalyses the second step of splicing. After catalysis, the post-spliceosomal complex dissociates, which releases the mRNA. The snRNPs are then recycled for subsequent rounds of splicing.

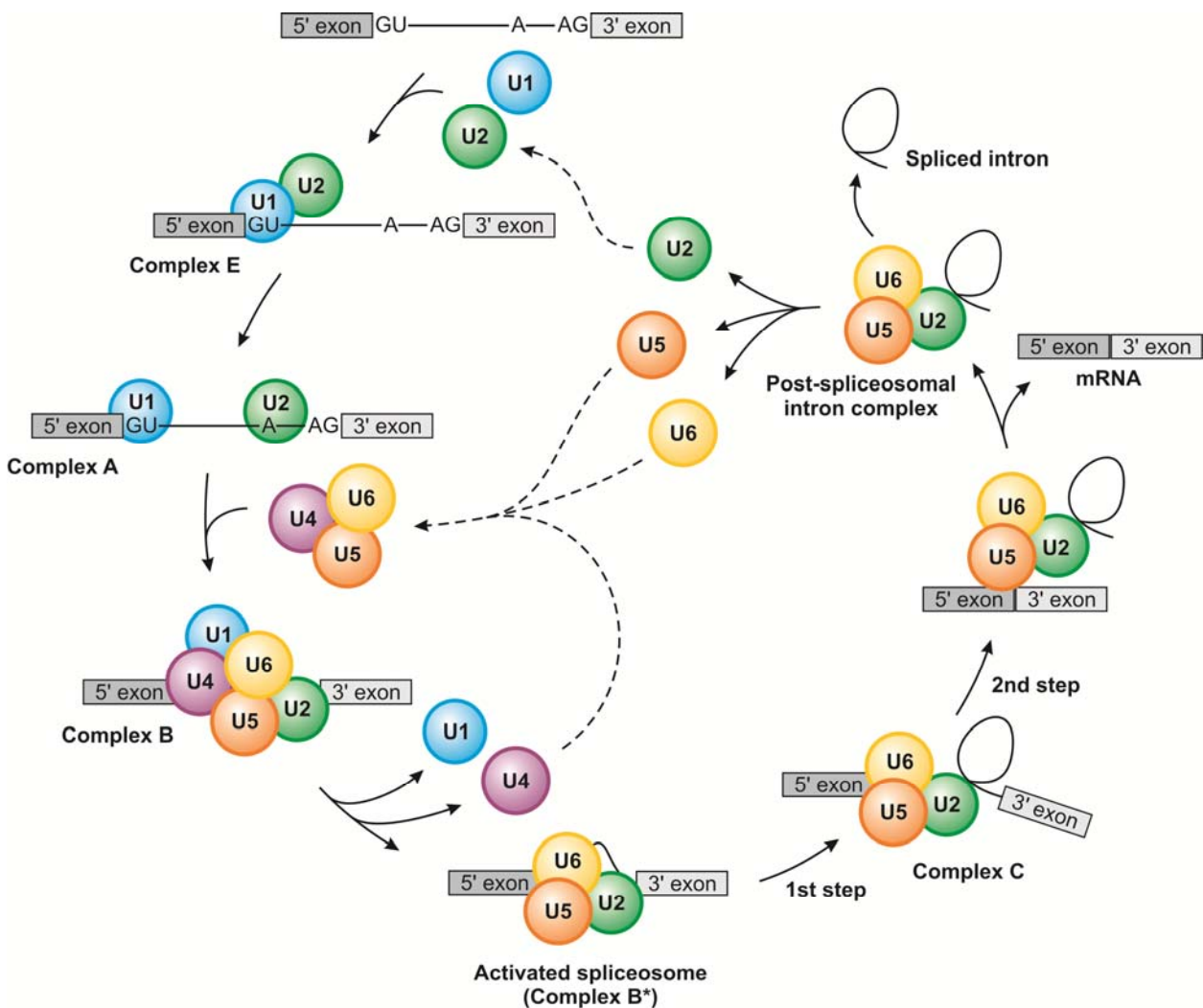


Fig. 1.4: Assembly/disassembly pathway of the spliceosome. The spliceosome is formed by the stepwise assembly of the spliceosomal U snRNPs (indicated by colored circles) on the pre-mRNA. Spliceosome assembly is initiated by the binding of U1 and U2 snRNPs to the 5' ss, followed by the stable association of U2 snRNP with the branch site. After binding of the U4/U6.U5 tri-snRNP the spliceosome is catalytically activated, leading to the dissociation of U1 and U4 snRNPs. Next, the splicing reactions occur. After splicing, the spliceosome disassembles, and the snRNPs are recycled for subsequent rounds of splicing. The distinct spliceosomal complexes are indicated.

An extensive RNA-RNA network is formed during spliceosome assembly and maturation (Fig. 1.5; reviewed in Nilsen, 1998). During the early stages of spliceosome assembly, U1 snRNA engages in base pairing interactions with the 5' ss and U2 snRNA with the BPS of the pre-mRNA. In the latter case, the branch point adenosine is bulged out in order to be sterically configured for the first step of splicing. After recruitment of the U4/U6.U5 tri-snRNP to the spliceosome, the U5 snRNA contacts nucleotides of the 5' and 3' exon. Catalytic activation of the spliceosome entails disruption of the U4/U6 base pairing interaction and binding of the conserved ACAGAG box of U6 snRNA to intron nucleotides at the 5' ss, thereby displacing the U1 snRNA. Furthermore, the U6 and U2 snRNAs form several short snRNA-snRNA duplexes. Additionally, U6 is refolded, forming an intramolecular stem loop (U6-ISL) (Fortner et al., 1994) that plays a crucial role in splicing catalysis (Yean et al., 2000). The establishment of an elaborate RNA-RNA network is a major design principle that ensures a spatial arrangement of the pre-mRNA's reactive groups compatible for both splicing reactions. Significantly, conformational rearrangements in the snRNA/pre-mRNA network are accompanied by intricate changes in the spliceosome's protein composition and in its RNP network, as will be discussed in the following chapters.

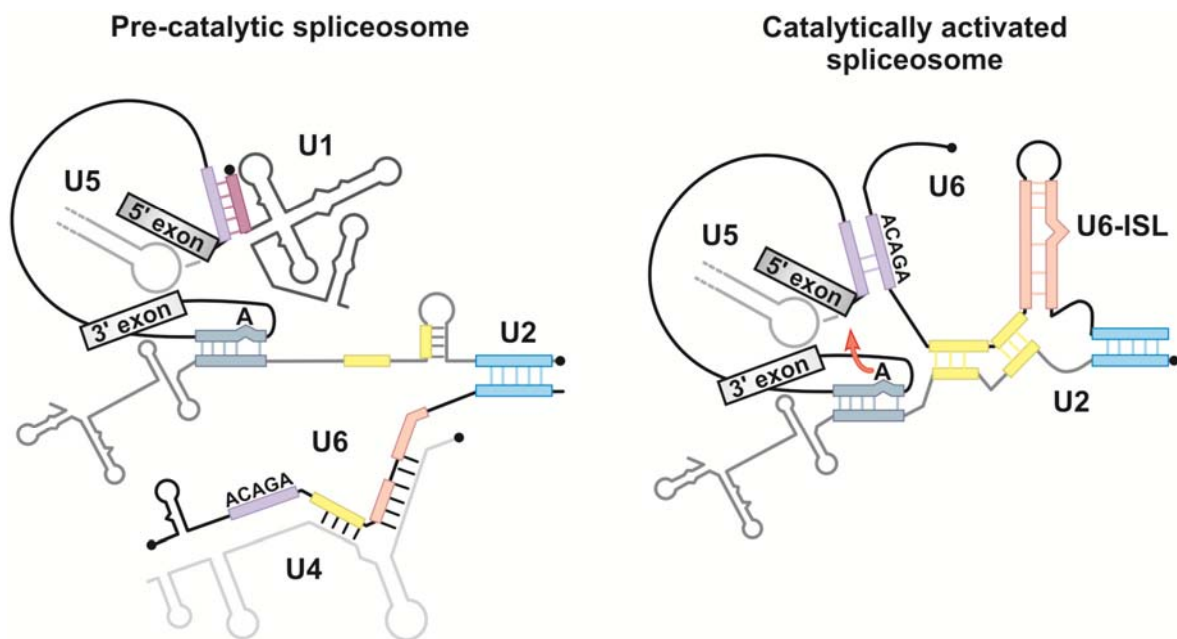


Fig. 1.5: Dynamic network of RNA interactions in the spliceosome. Comparison of RNA-RNA interactions in the pre-catalytic (left) and activated spliceosome (right). During catalytic activation, the interactions of U2 and U6 snRNAs and the pre-mRNA are dramatically remodeled. The conserved ACAGAG box at the 5' end of U6 displaces U1 at the 5' ss (purple boxes). Subsequently, U1 and U4 dissociate from the spliceosome. U6 refolds and forms the U6 intramolecular stem loop (U6-ISL, light red boxes) and engages in new base pairing interactions with U2 (yellow boxes). The branch point adenosine is shown and the red arrow indicates it is poised for the first step of splicing. Each snRNA is shown in its representative secondary structure and the 5' end of each snRNA is indicated by a black ball. Important regions of snRNAs involved in base pairing interactions are highlighted with colored boxes. Only stem loop I of U5 is shown. Adapted from Wahl et al., 2009.

1.4 The spliceosome is a particularly protein-rich machinery

Intermediates and products of pre-mRNA splicing shown in Fig. 1.2 are highly similar to those generated during the removal of group II self-splicing introns (Cech, 1986). Furthermore, several of the described interactions formed between the pre-mRNA and the U2, U5, and U6 snRNAs are related to those structures formed by group II self-splicing introns (Madhani and Guthrie, 1992; Weiner, 1993). Recently, the crystal structure of a hydrolytic group IIC intron revealed that its U6-like ISL coordinates two metal ions (Toor et al., 2008). By analogy to the spliceosomal U6-ISL, this U6-like stem loop might mediate the transesterification reactions by positioning the metal ions for catalysis (Sontheimer et al., 1997; Yean et al., 2000). Furthermore, base paired fragments of U6 and U2 snRNA perform a two-step reaction with two short oligonucleotides as substrates that is chemically identical to the splicing reaction and leads to the formation of one linear RNA product (Valadkhan et al., 2009). Although it is still unclear whether group II self-splicing introns and the spliceosome share identical catalytic strategies (Weiner, 1993), it is clear that a vast interplay of pre-mRNA and snRNAs is essential for forming the catalytically active structures of the spliceosome. Nevertheless, splicing of pre-mRNA in the nucleus of a eukaryotic cell requires a sophisticated interplay of both RNA and protein. Several spliceosomal proteins play key roles in splice site recognition and pairing (reviewed in Will and Lührmann, 2006), the formation of the RNA/RNP structures required for splicing catalysis (Kramer, 1996), and might even be directly involved in splicing catalysis (Pena et al., 2008). Thus, the spliceosome may be regarded as a protein-assisted ribozyme.

Approximately 45 unique proteins are recruited to the spliceosome as stable components of the snRNPs. In addition to these proteins, the spliceosome also contains numerous proteins that are not associated with snRNA, many of which play essential roles during splicing. Indeed, in the case of short introns, protein constitutes more than two-thirds of the mass of a spliceosome. Mass spectrometric (MS) analyses of human spliceosomes indicate that over 170 proteins associate with the spliceosome throughout its assembly and catalytic activity (Wahl et al., 2009). Many proteins associate transiently and thus individual complexes (B or C) contain only ~110 proteins. Yeast spliceosomes are less complex (e.g. ~80-50 proteins in the yeast C complex compared to ~110 in its human counterpart) (Fabrizio et al., 2009). Nonetheless, those proteins thought to comprise the core of the spliceosome are conserved between yeast and human.

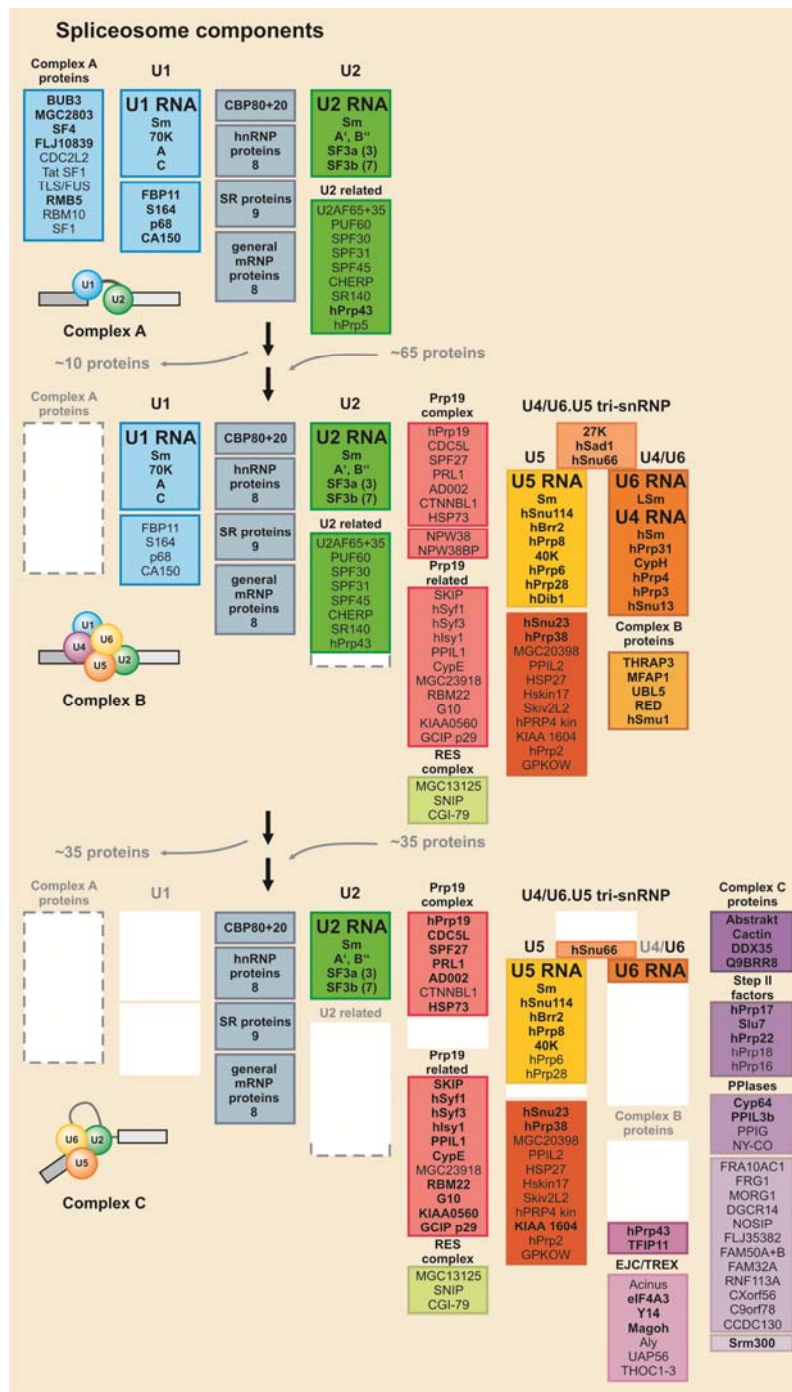


Fig. 1.6: Compositional dynamics of human spliceosomal A, B and C complexes. The protein composition of the human A complex (Behzadnia et al., 2007), the human B complex (Bessonov et al., 2008; Deckert et al., 2006) and the human C complex (Bessonov et al., 2008), as determined by MS. Proteins are grouped according to snRNP association, function, presence in a stable heteromeric complex or association with a particular spliceosomal complex, as indicated. The relative abundance of proteins is indicated by light (substoichiometric amounts) or dark (stoichiometric amounts) lettering and is based on the relative amounts of peptides sequenced or, in some cases, also by immunoblotting experiments. The numbers indicate the total number of individual proteins in a particular group (where individual proteins are not listed). Adapted from Wahl et al., 2009.

Comparative MS analyses of spliceosomal A, B and C complexes purified under native conditions have revealed a dynamic exchange of many proteins during spliceosome maturation (Fig. 1.6). Human A complexes contain nearly all U1 and U2 proteins plus ~50 non-snRNP proteins (Behzadnia et al., 2007). During the A complex to B complex transition, ~25 proteins enter the spliceosome as part of the U4/U6.U5 tri-snRNP, plus additional ~40 non-snRNP proteins (Deckert et al., 2006). At the same time ~10 proteins that are unique for the A complex leave the spliceosome. During conversion of the B complex into a step I spliceosome (C complex) ~35 proteins are lost, including all U4/U6-associated proteins, whereas approximately the same

number of proteins enter the spliceosome (Bessonov et al., 2008). These include, among others, so-called step 2 factors, that function just prior to or during the second transesterification reaction, a number of proteins designated C complex specific, together with several DExD/H-box helicases and peptidyl-prolyl cis/trans-isomerases. Recent studies indicate that the dynamic exchange of spliceosomal proteins is also evolutionary conserved from yeast to human (Fabrizio et al., 2009).

1.4.1 Roles of proteins in pre-mRNA splicing

Proteins play critical roles in the early recognition of the 5' and 3' ss of an intron. For example, proteins containing serine-arginine-rich domains (SR proteins) and proteins of the U1 snRNP stabilize initial interactions of the U1 snRNA with the 5' ss (reviewed in Will and Lührmann, 1997). Indeed, most RNA-RNA interactions in the spliceosome are very short and thus require spliceosomal proteins for their stability. Furthermore, the proteins SF1/BBP and U2 auxiliary factor (U2AF) cooperatively bind the BPS and polypyrimidine tract, respectively (Berglund et al., 1998). Subsequent binding of the U2 snRNP to the BPS is facilitated by the U2-associated SF3a and SF3b complexes, as well as the proteins Sub2/UAP56 and Prp5, both members of the DExD/H-type RNA-dependent ATPases/helicases (Fleckner et al., 1997; O'Day et al., 1996). Members of this family are found associated with the spliceosome during all steps of spliceosome assembly and maturation. These additionally include, among others, U5-100K/Prp28, U5-200K/Brr2, Prp2, Prp16, Prp22, and Prp43. They catalyze the dynamic rearrangements in the spliceosomal RNA-RNA and RNA-protein network (reviewed in Schwer, 2001; Staley and Guthrie, 1998). For example, U5-100K/Prp28 and U5-200K/Brr2 are key players in rearrangements leading to the catalytic activation of the spliceosome (Maeder et al., 2009; Staley and Guthrie, 1999), whereas rearrangements required for step I and II are catalyzed by Prp2 and Prp16, respectively (Roy et al., 1995; Schwer and Guthrie, 1992; Teigelkamp et al., 1994). Protein kinases, such as SRPK2 and Prp4 kinase (Mathew et al., 2008; Schneider et al., 2010; Schwelnus et al., 2001), and phosphatases (Shi and Manley, 2007) also play crucial roles during splicing.

A number of proteins are located at or near the reactive groups of the pre-mRNA and are therefore candidates for participating in catalysis of the splicing reaction. The U2-specific SF3b14a/p14 protein contacts the branch point adenosine during the early stages of spliceosome formation and also until the first catalytic step (Will et al., 2001). The U5-specific protein 220K/Prp8 occupies a central position in the spliceosome's catalytic core. After integration of the U4/U6.U5 tri-snRNP during B complex formation, the U5-220K/Prp8 protein can be crosslinked to all reactive groups of the pre-mRNA (5' ss, 3' ss and BPS) (reviewed in Boon et al., 2006; Grainger and Beggs, 2005). Recent X-ray crystallography structures of the 5' ss interacting region in the human U5-

220K/hPrp8 protein and its yeast homolog revealed an RNase H-like domain (Pena et al., 2008; Ritchie et al., 2008; Yang et al., 2008), suggesting that it both orients the substrate and may participate in the chemical catalysis of splicing (reviewed in Abelson, 2008).

1.5 Prp19 is part of a heteromeric complex in yeast and human

Genetic and biochemical analysis in the yeast *Saccharomyces cerevisiae* identified the Prp19 protein as an essential splicing factor that is not associated with snRNA (Cheng et al., 1993; Tarn et al., 1993). Prp19 is evolutionary highly conserved, and it is found as part of a stable heteromeric complex in yeast (the Nineteen complex, NTC; Tarn et al., 1994) and human (the hPrp19/CDC5L complex; Makarova et al., 2004). MS analysis of immunoaffinity purified human Prp19/CDC5L complexes showed that they consist of seven proteins: CDC5L, HSP73, CTNBL1, PRL1, hPrp19, AD002 and SPF27 (Makarova et al., 2004). Five of these proteins, that is, Prp19, CDC5L, HSP73, PRL1 and SPF27, are also core components of several CDC5L-containing complexes that were isolated from HeLa nuclear extract either by immunoaffinity chromatography or by size exclusion (Ajuh et al., 2000). The organization of the human Prp19/CDC5L complex differs considerably from the yeast NTC complex (Fig. 1.7A). The latter complex contains at least eight proteins: Cef1 (NTC85), Prp19, Snt309 (NTC25), Syf1 (NTC90), Syf2 (NTC31), Syf3 (NTC77), Isy1 (NTC30) and NTC20 (Chan et al., 2003), but several additional proteins are also thought to be present, including Prp46 (Ntc50) and Cwc2 (NTC40) (Tarn et al., 1994). Of the described NTC components, only three have clear homologs in the human Prp19/CDC5L complex, namely Prp19, Cef1 (CDC5L in human) and Prp46 (PRL1 in human). Whereas Cef1 and Prp19 are stably bound in the yeast NTC, Prp46 dissociates from the complex upon gradient centrifugation, suggesting that it is only loosely associated (Tarn et al., 1994). Furthermore, although of low sequence homology, the NTC component Snt309 is thought to be a functional homolog of the *Schizosaccharomyces pombe* Cwf7p, which in turn is homologous to the human Prp19/CDC5L-complex protein SPF27 (Ohi and Gould, 2002).

Whereas little is known about the molecular organization of the human Prp19/CDC5L complex, some data has accumulated about interactions among NTC components. Prp19 interacts directly with a number of NTC components, including Cef1, Snt309 and Cwc2, as identified by several methods, including yeast two hybrid, far western, or immunoprecipitation assays (Chen et al., 2002; Chen et al., 1998; Ohi and Gould, 2002; Tsai et al., 1999). Growing evidence suggests that Prp19 is required for the integrity of the NTC and that Snt309 is a modulator of the required interactions (Chen et al., 1999).

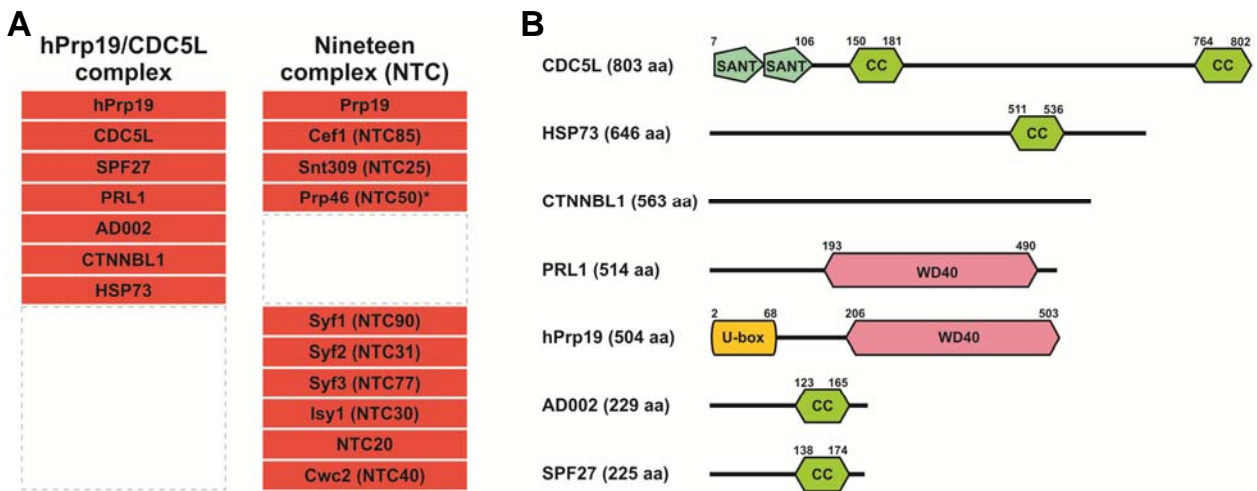


Fig. 1.7: Comparison of the human Prp19/CDC5L complex with the yeast Nineteen complex (NTC). (A) Both complexes share the Prp19 and CDC5L (Cef1) proteins. Yeast Snt309 is a functional counterpart of *S. pombe* Cwf7 that is homologous to human SPF27 (Ohi and Gould, 2002). Yeast Prp46 is homologous to human PRL1, but it is only loosely associated with the yeast NTC and therefore marked with an asterisk. The dotted grey line indicates those proteins that are not shared between both complexes. (B) Predicted domain structure of the human Prp19/CDC5L complex proteins. The total number of amino acids is shown in parentheses. Domains were predicted with the SMART 6 program (Letunic et al., 2009) and are indicated with their amino acid position. CC – coiled-coil domain; U-box – U-box domain; WD40 – WD40 domain; SANT – SANT domain. For further explanation, see text below.

Yeast Prp19 contains three known protein motifs that contribute to the structure and function of the NTC. These are an N-terminal U-box domain, followed by a predicted coiled-coil domain and a C-terminal WD40 domain (Ohi et al., 2005). EM analysis together with analytical ultracentrifugation revealed that recombinant Prp19 from yeast forms a tetramer *in vivo* and *in vitro* and the domain required for the tetramerization was mapped to the central coiled-coil domain (Ohi et al., 2005). Furthermore, the crystal structure of the N-terminal U-box domain of yeast Prp19 revealed a dimerization of the U-box in the context of the Prp19 tetramer (Vander Kooi et al., 2006). Although an oligomerization is likely for the human Prp19 protein as well (Grillari et al., 2005), it is presently unclear whether it also forms a tetramer in the native human Prp19/CDC5L complex.

The U-box domain is a structural feature of ubiquitin E3 ligases and is stabilized by a system of salt-bridges and hydrogen bonds (Aravind and Koonin, 2000; Ohi et al., 2003). Post-translational ubiquitin-modification of a protein targets it for degradation by the proteasome (Ciechanover, 1998), but was also shown to be a mediator of protein activity, protein-protein interactions, and subcellular localization (Aguilar and Wendland, 2003). Ubiquitin ligation is a multi-step process that involves three classes of enzymes. The E1 enzymes first charge ubiquitin in an ATP-dependent manner to form an activated E1-ubiquitin thioester intermediate. Ubiquitin is then transferred to a cysteine residue of an E2 enzyme and finally to an E3 ligase. The E3 ligase interacts directly with the substrate and transfers ubiquitin to a lysine residue of the substrate.

A mutation in the hydrogen-bonding network of the Prp19 U-box domain destabilizes Prp19's protein structure, and disrupts interactions with its binding partners, causing NTC instability (Chen et al., 1998; Ohi et al., 2003; Vijayraghavan et al., 1989). However, mutations that are predicted to interrupt the interactions with an E2 enzyme, but do not alter the U-box structure are unable to rescue Prp19 function *in vivo*, suggesting that the enzymatic activity of the Prp19 U-box domain is essential for the cell viability (Vander Kooi et al., 2006). Indeed, human Prp19 was shown to possess ubiquitin ligase activity *in vitro* (Hatakeyama et al., 2001). Very recently it was shown that the human Prp19 complex ubiquitinates the U4 snRNP component Prp3 (Song et al., 2010) and that ubiquitination of Prp3 increases its affinity for the U5 component 220K/Prp8 thereby stabilizing the U4/U6.U5 tri-snRNP. Furthermore, yeast Prp8 is reversibly ubiquitinated and its ubiquitinated form represses Brr2 (Bellare et al., 2008). Thus, it was suggested that Prp19 may play a role also in the ubiquitination of Prp8, but formal proof of this is currently lacking.

In addition to the U-box, Prp19 contains seven WD40 repeats. WD40-repeat containing proteins were shown to have several critical roles in many biological functions (reviewed in Smith et al., 1999). As demonstrated first by the crystal structure of the G protein β subunit, all WD40-repeat proteins are thought to form a circularized β propeller (Sondek et al., 1996; Wall et al., 1995). However, the crystal structure of recombinant yeast Prp19 (Fig. 3.33B) revealed a geometry that is atypical from known β propeller structures (Vander Kooi et al., 2010), as discussed below. To date, the only protein known to interact with Prp19 via the WD40-domain is Cwc2 (Ohi and Gould, 2002). Interestingly, Cwc2 can be crosslinked directly to the U6 snRNA, before and after the first step of splicing (McGrail et al., 2009). Therefore, it was proposed that Cwc2 may link the NTC to the spliceosome during pre-mRNA splicing through the U6 snRNA (McGrail et al., 2009).

As shown in Fig. 1.7B, some structural features of the hPrp19 protein are repeatedly present also in the other subunits of the hPrp19/CDC5L complex. For example, the PRL1 protein also contains a WD40 domain that is build up by seven WD40 repeats. This domain was shown to be involved in an interaction with the C-terminus of CDC5L (Ajuh et al., 2001). CDC5L is the largest protein that is present in the hPrp19/CDC5L complex. In addition to its interaction with PRL1, a self-interaction was proposed for CDC5L (Graub et al., 2008), as well as for its yeast counterpart Cef1 (Tsai et al., 1999). At its very N-terminus, CDC5L contains two SANT domains (also cMyb domain; Switching-defective protein 3 – Swi3, Adaptor 2 – Ada2, Nuclear receptor corepressor – N-CoR, Transcription factor IIIb) that in other proteins were shown to be engaged in chromatin remodeling by functioning as a unique histone-interaction module that couples histone binding to enzyme catalysis (reviewed in Boyer et al., 2002). SANT domains possess strong similarity to the DNA-binding domain of Myb-related proteins and contain tandem repeats of three α -helices that are arranged in a helix-turn-helix motif (Grune et al., 2003). Furthermore, CDC5L also contains two

coiled-coil domains and coiled-coil domains were also predicted for the AD002, SPF27, and HSP73 proteins (Fig. 1.7B). HSP73 is a molecular chaperone and a constitutively expressed member of the 70 kDa heat shock protein family. HSP73 binds to unfolded polypeptides (Beckmann et al., 1990) to facilitate membrane translocation of the polypeptide or proper folding, perhaps in the presence of additional subunits (reviewed in Rothman, 1989). Furthermore, HSP73 hydrolyses ATP and the structure of the ATPase domain was solved (O'Brien and McKay, 1993).

1.6 The Prp19 complex plays a key role in the catalytic activation of human and yeast spliceosomes

Comparative proteomic analyses of the B Δ U1 and B* spliceosomal complexes that were purified from HeLa nuclear extract under identical stringent conditions indicated that the human Prp19/CDC5L complex enters the spliceosome during or immediately prior to catalytic activation of the spliceosome (Deckert et al., 2006; Makarov et al., 2002; Makarova et al., 2004). Isolation of native B complexes revealed the presence of substoichiometric amounts of the hPrp19/CDC5L complex already in the B complex, as compared to a step I spliceosome that was purified under the same native conditions (Bessonov et al., 2008; Deckert et al., 2006). Together with the hPrp19/CDC5L complex, a set of proteins designated Prp19-related, that either interact with human or yeast Prp19 complex subunits or are present in the human post-spliceosomal 35S U5 snRNP, associate with the spliceosome at this stage (Ajuh et al., 2000; Chen et al., 2002; Makarov et al., 2002; Ohi et al., 2002), but are more stably integrated during the B to C complex transition (Bessonov et al., 2008). These proteins include SKIP, hSyf1, hSyf3, hIsy1, PPIL1, CypE, MGC23918, RBM22, G10, KIAA0560 and GCIP/p29. Immunodepletion/complementation studies with HeLa nuclear extract confirmed that the hPrp19/CDC5L complex is required for pre-mRNA splicing prior to the first step of splicing (Makarova et al., 2004). This is consistent with data obtained from yeast, where a temperature sensitive *prp19* mutant accumulated unspliced pre-mRNA at the nonpermissive temperature (Tarn et al., 1993).

As schematically indicated in Fig. 1.8, the hPrp19/CDC5L complex subunits and Prp19-related proteins stably interact with the U5 snRNP during catalytic activation of the spliceosome and remain associated during both catalytic steps (Makarov et al., 2002). Significantly, the hPrp19/CDC5L complex and related proteins, together with U5 proteins, are still part of a 1 M NaCl resistant RNP core of the C complex (Bessonov et al., 2008). This suggests that they are tightly incorporated components of the C complex and may play an important role in maintaining its catalytically active RNA network. Accordingly, the hPrp19/CDC5L complex leaves the spliceosome together with U5 as part of the post-spliceosomal 35S U5 snRNP (Makarov et al., 2002).

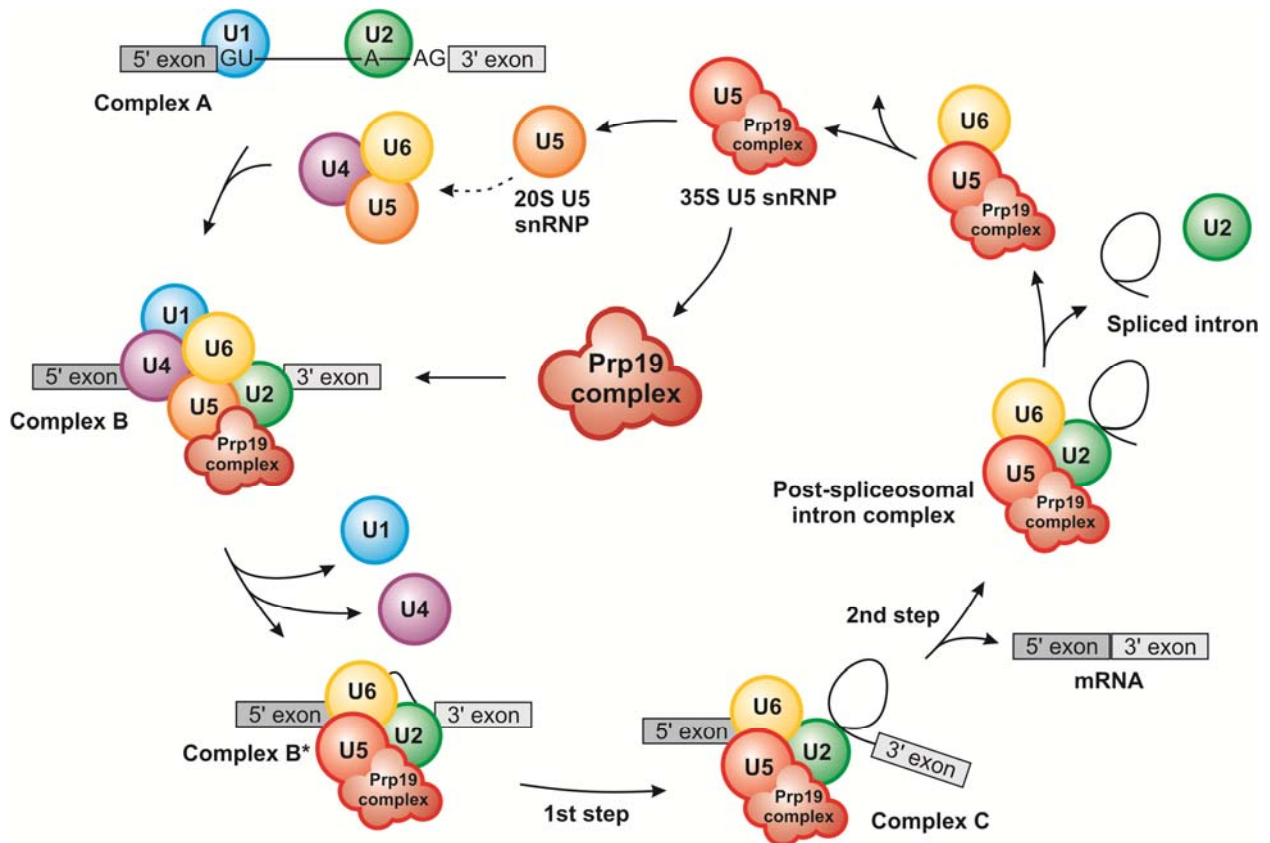


Fig. 1.8: Stable interaction of the hPrp19/CDC5L complex with the U5 snRNP during the splicing cycle. The hPrp19/CDC5L complex and Prp19-related proteins enter the spliceosome during B complex formation. During catalytic activation, they stably interact with the U5 snRNP, leading to a reconstruction of the U5 particle. The interaction is thought to persist during both catalytic steps. U5 is thought to dissociate after splicing in the form of a 35S particle and must then be converted to a 20S form before it can reassociate with the U4/U6 snRNP. Simultaneously, the hPrp19/CDC5L complex is released from the 35S U5 snRNP to partake in another round of splicing. Whether the 35S U5 snRNP is initially released from the post-spliceosomal complex together with U6 is presently not clear. Adapted from Makarov et al., 2002 with modifications.

Stable association of the hPrp19/CDC5L-complex and Prp19-related proteins with the U5 snRNP contributes to the massive remodeling of the U4/U6.U5 tri-snRNP during activation. The U4/U6.U5 tri-snRNP enters the spliceosome as a preformed unit. Whereas U4 and U6 snRNAs are extensively base paired with each other, no evidence exists for base pairing interactions between U4 or U6 snRNAs and the U5 snRNA. Therefore the connection between U4/U6 and U5 is likely to be mediated by protein-protein interactions (Liu et al., 2006; Makarova et al., 2002). Upon catalytic activation, the U4/U6 base pairing is disrupted by the DExD/H-box protein U5-200K/Brr2, thereby allowing the annealing of U2 with U6 snRNA. Furthermore, the U4 snRNA and all pre-catalytic U4/U6-proteins dissociate from the spliceosome, along with proteins important for tethering the U5 snRNP to the U4/U6 snRNP. Subsequently, the U5 and U6 snRNAs engage in new base pairing interactions with the pre-mRNA. As U6 has lost its protein binding partners at this stage, it likely engages in new interactions, which most likely involve the hPrp19/CDC5L complex. Remodeling of

35S U5 snRNP, but not 20S U5 snRNP, and are released from the spliceosome as a part of it (Fig. 1.8). Significantly, the 35S U5 snRNP contains most proteins found in the salt-stable RNP core of the C complex (Fig. 1.10) (Bessonov et al., 2008; Makarov et al., 2002). Since these proteins are already present in the activated spliceosome, the remodeled 35S U5 snRNP appears to persist throughout both catalytic steps of splicing and is released upon dissociation of the spliceosome (Fig. 1.8). Thus, the 35S form of the U5 snRNP probably represents a disassembly intermediate of the spliceosome and is likely released together with the U2 and U6 snRNP as part of the post-spliceosomal complex. In support of this model, a recently characterized 40S post-splicing intron-containing complex isolated from HeLa cells was shown to contain U2, U5 and U6 snRNAs and also several proteins of the 35S U5 snRNP (Yoshimoto et al., 2009). Furthermore, endogenous yeast spliceosomes that likely represent a post-catalytic form of the splicing machinery and contain solely U2, U5 and U6 snRNPs, exhibit a protein composition strikingly similar to that of the human 35S U5 snRNP (Ohi et al., 2002).

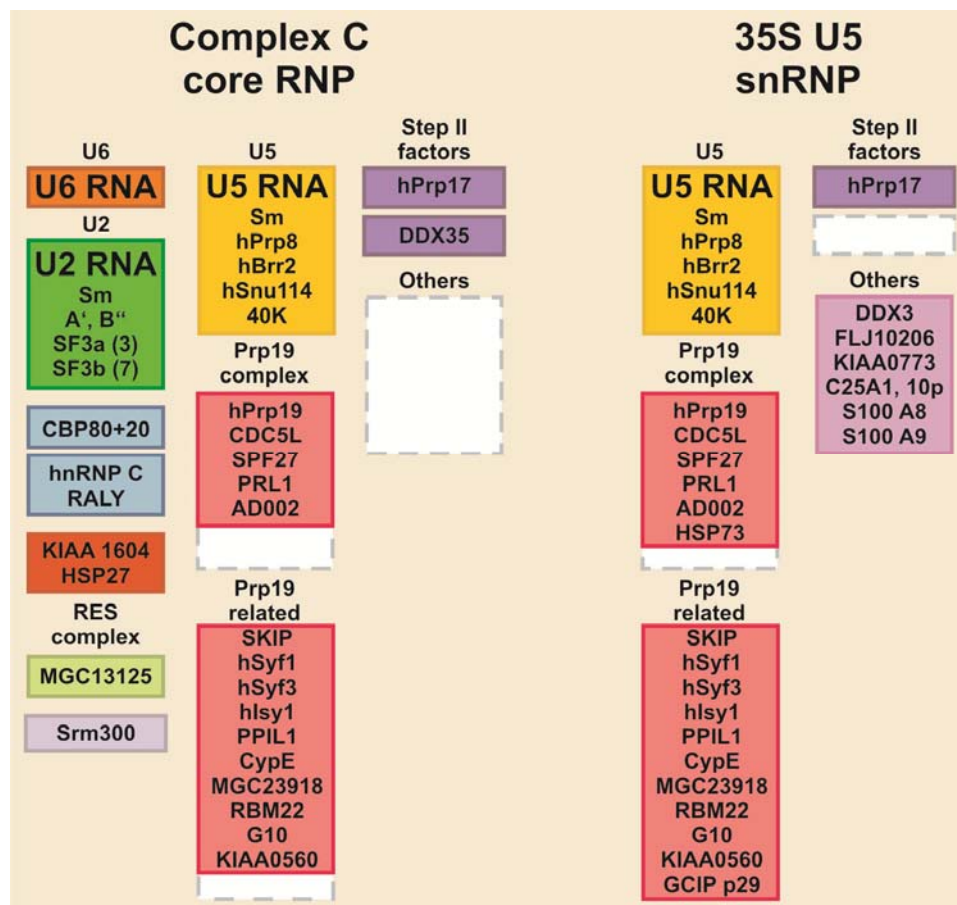


Fig. 1.10: Comparison of the protein composition of the human 35S U5 snRNP with the 1 M salt treated RNP core of the C complex. The protein composition of the 1 M salt treated human C complex (Bessonov et al., 2008), and the human 35S U5 snRNP (Makarov et al., 2002), as determined by MS. Proteins are grouped according to snRNP association, function, presence in a stable heteromeric complex or association with a particular spliceosomal complex, as indicated.

In a subsequent step, the 35S U5 snRNP is released from the post-spliceosomal intron-containing complex. Whether U5 is released as part of a larger complex containing U6, as shown in Fig. 1.8, is presently unclear. Finally, the 35S U5 snRNP must be converted back into a 20S form. This requires dissociation of the hPrp19/CDC5L complex and Prp19-related proteins, and binding of the U5-specific 102K, 100K and 15K proteins, before it can reassociate with the U4/U6 snRNP to build a new U4/U6.U5 tri-snRNP and partake in subsequent rounds of splicing.

1.8 Towards the structure of the spliceosome

Most cellular processes rely on the concerted action of macromolecular assemblies (Gavin et al., 2006) and knowledge of the three-dimensional (3D) architecture of these machines is crucial for understanding their function. Much of what we know about the spliceosome's structural organization has come from the characterization of its main subunits, namely the snRNPs. For example, the recently determined crystal structure of a reconstituted human U1 snRNP at 5.5 Å resolution gives a first close-up view of the hierarchical architecture of spliceosomal snRNPs (Pomeranz Krummel et al., 2009). The solved structure contains not less than nine proteins together with the U1 snRNA, and provides insights into the role of both the U1C protein and U1 snRNA in 5' ss recognition.

Considering the large size and highly dynamic nature of the spliceosome, analyzing a eukaryotic spliceosome by means of X-ray crystallography is an enormous challenge. With respect to the different assembly intermediates of a spliceosome, it is not possible to purify the spliceosome from cellular extract. In contrast, spliceosomes must be assembled *in vitro* on an exogenously added pre-mRNA and stalled at a particular assembly stage in order to obtain a homogenous population of complexes. Furthermore, reconstitution of a whole spliceosome with recombinant proteins and RNAs, as in the case of the human U1 snRNP (Pomeranz Krummel et al., 2009), does not seem feasible considering the high number of proteins constituting a spliceosome. Therefore, it is challenging to prepare larger spliceosomal complexes in amounts and concentrations sufficient for crystallization trials. On the other hand, electron microscopy (EM) can be used to obtain a more global, although less detailed, view of spliceosome assemblies and a number of splicing-related complexes have been studied by EM in the last years (Fig. 1.11) (reviewed in Jurica, 2008; Stark and Lührmann, 2006). Low resolution 3D structures (~14-40 Å) of human snRNPs and their subcomplexes including the U1 snRNP (Stark et al., 2001), a protein subcomplex of the U2 snRNP (Golas et al., 2003, 2005), as well as the U4/U6.U5 tri-snRNP and its U5 and U4/U6 snRNP subunits (Sander et al., 2006) were obtained. Also, structures of endogenous yeast and human splicing complexes have been reported (Azubel et al., 2004; Ohi et al., 2007). Moreover, recent progress has

been made in elucidating the 3D architecture of the human A and B Δ U1 complexes (Behzadnia et al., 2007; Boehringer et al., 2004). Additionally, the 3D structure of a human C complex isolated at stringent conditions has been presented (Jurica et al., 2004). However, the resolution of the structural data presently available for these spliceosomal complexes is too low to allow a functional interpretation of the observed structure. Thus, subunit labels can be extremely helpful to obtain a better understanding of the overall arrangement and subunit organization.

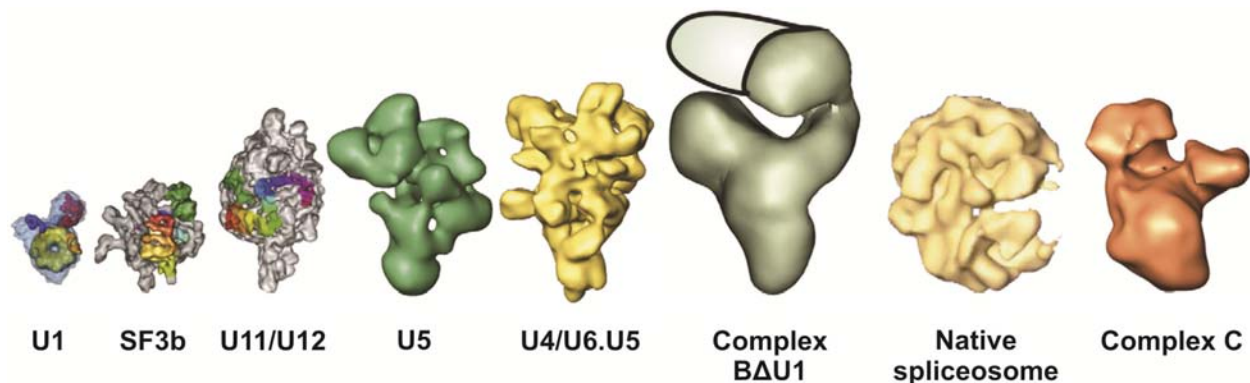


Fig. 1.11: 3D reconstitutions of spliceosomal complexes. Several macromolecular assemblies involved in the removal of an intron in pre-mRNA to generate mRNA have been visualized by EM and single particle reconstruction including snRNP complexes and subcomplexes (SF3b, U1, U5, U4/U6.U5 tri-snRNP, U11/U12 di-snRNP), splicing-related complexes (B Δ U1) and structural intermediates of the spliceosome (complex C, “native” spliceosome). “Native” spliceosomes are endogenously purified complexes and the identity of these particles is not entirely defined. Adapted from Lührmann and Stark, 2009.

As shown in Fig. 1.12, different labeling strategies have been applied to identify functional sites in several spliceosomal complexes. A gold-labeling method was adopted to locate the 5' and 3' exons and the intron of the pre-mRNA, as well as the U2-associated protein SF3b155, at the surface of affinity-purified, human B complexes (Fig. 1.12A-C) (Wolf et al., 2009). Therefore, either the spliceosomal protein SF3b155 or the MBP protein that was bound to pre-mRNA indirectly by insertion of an MS2 stem-loop and subsequent binding of an MS2-MBP fusion protein were labeled specifically with antibodies. To facilitate visualization of the bound antibody by EM, the latter was additionally labeled with colloidal gold coated with protein A. A similar approach was also used for the localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast U4/U6.U5 tri-snRNP (Hacker et al., 2008).

A different strategy was utilized for structural mapping of the human U4/U6.U5 tri-snRNP. In this case, the structures of the U4/U6.U5 snRNP, the U5 snRNP and the U4/U6 di-snRNP were determined by EM at resolutions of 20 Å, 25 Å and 35 Å, respectively (Sander et al., 2006). The U5 snRNP and the U4/U6 di-snRNP were then fitted into the U4/U6.U5 tri-snRNP structure (Fig. 1.12D). In such a fit, the U4/U6 snRNP could be localized toward the broader top region of the U5 snRNP and only one structural domain of U5 did not fit completely into the U4/U6.U5 density.

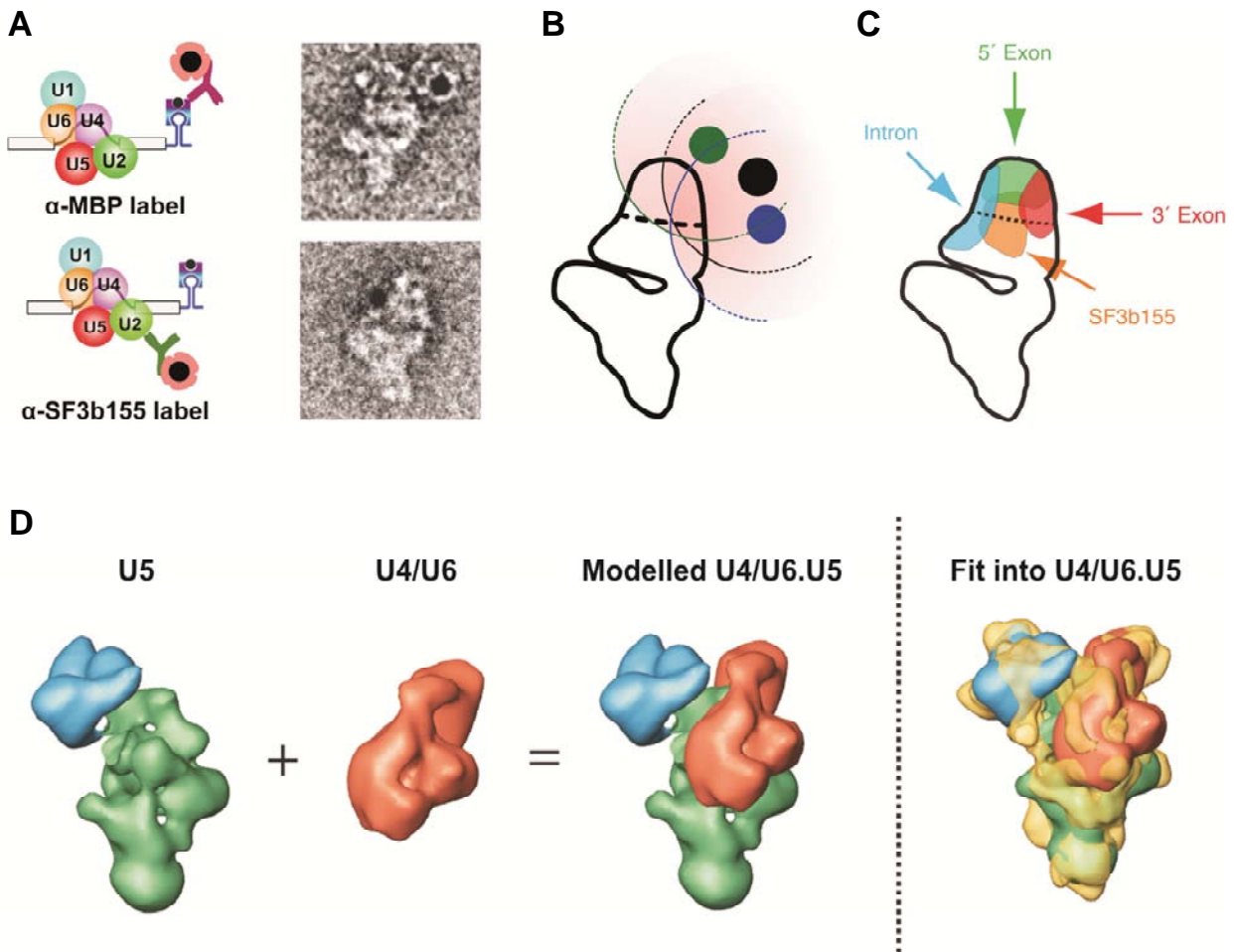


Fig. 1.12: Functional interpretation of spliceosomal complexes by EM. (A-C) EM localization of the pre-mRNA and SF3b155 in human B complexes. Adapted from Wolf et al., 2009. (A) Purified B complexes assembled on MS2 hairpin-tagged pre-mRNA were incubated with α -MBP (upper left panel) or α -SF3b155 (lower left panel) antibodies, labeled by incubating with colloidal gold coated with protein A, and visualized by EM. EM images of the pre-mRNA's 3' end labeling (upper right panel) and SF3b155 labeling (lower right panel) are shown. (B) Mapping of the 3' end of the 3' exon in the head domain of the B complex. The diagram illustrates the maximum distance (~17 nm; shaded in pink) from the centre of the gold label (green, black and blue) to the antibody-binding site. The overlapping area of the 17-nm circles determined by analysis of several EM images indicates the region wherein the 3' end of the 3' exon is located. (C) Comparison of the mapped position of the three pre-mRNA positions and SF3b155, as indicated. (D) 3D reconstruction of the U4/U6.U5 tri-snRNP. The orientation of the head domain (blue) of low-salt purified U5 snRNP (green) was slightly modified and combined with the reconstruction of the U4/U6 di-snRNP (red). The modeled U4/U6.U5 tri-snRNP was then fitted into the 3D map of the U4/U6.U5 tri-snRNP. Adapted from Stark and Lührmann, 2006.

Combining several structural approaches can be beneficial to tackle a system with the complexity of a spliceosome. Thus, using the 3D structures of the U1 snRNP (Stark et al., 2001) and the U2-specific complex SF3b (Golas et al., 2003) obtained by single-particle cryo-EM, detailed 3D models of the spatial organization of individual proteins could be built for each complex, utilizing all available structural data, including X-ray structures of individual proteins (reviewed in Stark and Lührmann, 2006).

1.9 Aim of this study

A unique feature of the spliceosome is that it forms an active site anew during each round of splicing. Formation of the catalytically active site of the spliceosome requires extensive remodeling events leading to major structural and compositional rearrangements. Protein complexes containing Prp19 play a central role in these processes in both human and yeast spliceosomes, and Prp19 and its related proteins are major components of the spliceosome's catalytic core RNP. Consequently, it was shown recently in our laboratory that the human Prp19/CDC5L complex is still associated with the 1 M NaCl treated core of a catalytically active spliceosomal C complex (Bessonov et al., 2008). In addition, the protein composition of the salt treated core of the spliceosome shows remarkable similarities to the post-spliceosomal 35S U5 snRNP. Thus, investigating the composition, organization and structure of the hPrp19/CDC5L complex and 35S U5 snRNP will provide first insights into the possible architecture of proteins at the catalytic core of the spliceosome and lay a foundation for a better understanding of the RNP core structure that also provides a framework for future high resolution structural studies.

Although some data is available about the organization and interactions of proteins in the yeast NTC, there is relatively little information available about the structure and molecular organization of the human Prp19/CDC5L complex and the 35S U5 snRNP. In previous studies, both of these complexes were isolated from HeLa nuclear extracts by immunoaffinity chromatography (Ajuh et al., 2000; Makarov et al., 2002; Makarova et al., 2004). However, the use of polyclonal antibodies for the purification of these complexes limited the amounts that could be purified, thereby preventing an in depth analysis of the structure and molecular organization of the human Prp19/CDC5L complex and 35S U5 snRNP.

Therefore, the first aim of this study was to purify preparative amounts of the human Prp19/CDC5L complex and 35S U5 snRNP utilizing HeLa cell lines that stably express the hPrp19/CDC5L complex proteins AD002 and SPF27 that contain an N-terminal FLAG affinity purification tag. As AD002 and SPF27 are both bona fide components of the hPrp19/CDC5L complex and the post-spliceosomal 35S U5 snRNP, a similar purification strategy was used for both complexes. The use of FLAG-tagged proteins allowed the use of monoclonal anti-FLAG antibodies for the affinity purification of both human complexes. During these studies purification protocols were established for large-scale purification of the human Prp19/CDC5L complex from nuclear extract with high reproducibility.

After purification of sufficient amounts of the hPrp19/CDC5L complex, the complex was analyzed in terms of its composition, stoichiometry, molecular organization and structure. Protein-protein interactions were analyzed using the endogenous hPrp19/CDC5L complex (e.g. via salt dissociation, glycerol gradient centrifugation and chemical crosslinking) and also via classical co-precipitation

assays utilizing *in vitro* translated proteins. Furthermore, limited proteolysis of the hPrp19/CDC5L complex was used to analyze the structural organization of the hPrp19/CDC5L complex in more detail. These studies provided information about functional domains involved in protein-protein interactions. A further aim of this study was to initialize high resolution structure determination studies of functionally important domains of the hPrp19/CDC5L complex that can be purified in high amounts and with reasonable structural and compositional homogeneity.

Due to the complexity of the 35S U5 snRNP, initial studies dealt mainly with the purification of the complex and its characterization on a proteomic and structural level. Structure analyses were first performed by negative stain two-dimensional (2D) EM, which provided a first view of the higher order structure and homogeneity of the complex. This information is a prerequisite for a subsequent 3D image reconstitution of the 35S U5 snRNP or even high resolution structural studies of the complex.

Taken together, this work increases our understanding of the structure of the hPrp19/CDC5L complex and 35S U5 snRNP, which in turn provide insights into the organization and structure of the catalytic core of the spliceosome.

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals

Chemical	Manufacturer
Acetic acid (glacial), 100%, anhydrous	Merck, Darmstadt
Acetone, for analysis	Merck, Darmstadt
Adenosine-5'-triphosphate (ATP)	Jena Biosciences, Jena
Agarose, UltraPure	Invitrogen, Carlsbad, USA
Ammonium bicarbonate (NH ₄ HCO ₃)	Fluka, Buchs, CH
Amonium persulfate (APS)	Merck, Darmstadt
Ampicillin	Sigma, Taufkirchen
Amylose resin	NEB, Ipswich, USA
ANTI-FLAG M2 Affinity Gel	Sigma, Taufkirchen
[γ - ³² P]-ATP	PerkinElmer, Waltham, USA
Boric acid (H ₃ BO ₃)	Merck, Darmstadt
Bovine serum albumin (BSA), acetylated	Sigma, Taufkirchen
Bradford reagent	Bio-Rad, Munich
Brij35	Serva, Heidelberg
Bromophenol blue	Merck, Darmstadt
Cacodylic acid	Fluka, Buchs, CH
Casytone	Schärfe System, Reutlingen
Chloramphenicol	Roth, Karlsruhe
Coomassie Brilliant Blue (CBB) G-250	Sigma, Taufkirchen
Complete protease-inhibitor cocktail, EDTA-free	Roche, Basel, Switzerland
Creatine phosphate	Sigma, Taufkirchen
Cysteine	Sigma, Taufkirchen
Cytidine-5'-triphosphate (CTP)	Jena Biosciences, Jena
2'-Deoxynucleoside-5'-triphosphate mix (dNTPs)	Invitrogen, Carlsbad, USA
Dimethyl sulfoxide (DMSO)	Roth, Karlsruhe
Dipotassium phosphate (K ₂ HPO ₄)	Merck, Darmstadt

Disodium carbonate (Na ₂ CO ₃)	Merck, Darmstadt
Disodium phosphate (Na ₂ HPO ₄)	Merck, Darmstadt
Dithioerythritol (DTE)	Roth, Karlsruhe
Dithiothreitol (DTT)	Roth, Karlsruhe
Dulbecco's Modified Eagle Medium (DMEM)	Biochrom, Berlin
Ethanol	Merck, Darmstadt
Ethidium bromide, 1 %	Serva, Heidelberg
Ethylenediaminetetraacetic acid (EDTA)	Merck, Darmstadt
Ficoll 400	Sigma, Taufkirchen
Formaldehyde solution, min. 37%	Merck, Darmstadt
Formamide	Merck, Darmstadt
G418-sulphate (Neomycine)	PAA, Pasching, Austria
GeneRuler™ 100 bp DNA Ladder	Fermentas, Burlington, CA
GeneRuler™ 1 kb DNA Ladder	Fermentas, Burlington, CA
Glutaraldehyde, 25% solution, EM grade	EMS, Hatfield, USA
Glycerol, for analyses	Merck, Darmstadt
Glycerol, about 87%, for analyses	Merck, Darmstadt
Glycinamide hydrochloride	Sigma, Taufkirchen
Glycine	Merck, Darmstadt
GlycoBlue	Applied Biosystems, Foster City, USA
Glycogen for mol. biol.	Roche, Basel, Switzerland
Guanidinium hydrochloride	Fluka, Buchs, CH
Guanosine-5'-triphosphate (GTP)	Jena Biosciences, Jena
Heparin, sodium salt	Roth, Karlsruhe
Heparin-Sepharose 6 Fast Flow	GE Healthcare, Chalfont St Giles, UK
HEPES, free acid	Sigma, Taufkirchen
Herring sperm DNA	Sigma, Taufkirchen
Hydrochloric acid, fuming, 37% (HCl)	Merck, Darmstadt
Iodoacetamide	Sigma, Taufkirchen
[¹⁴ C]Iodoacetamide (56 Ci/mmol)	GE Healthcare, Chalfont St Giles, UK
Isopropanol	Merck, Darmstadt
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Roth, Karlsruhe
L-Amino acids	Sigma, Taufkirchen
LB-Agar medium	MP Biomedicals, Solon, USA
L-[U- ¹⁴ C]Leucine (318 mCi/mmol)	GE Healthcare, Chalfont St Giles, UK
L-[³⁵ S]Methionine, Redivue (>1000 Ci/mmol)	GE Healthcare, Chalfont St Giles, UK
Magnesium acetate (Mg(Ac) ₂)	Merck, Darmstadt
Magnesium chloride (MgCl ₂)	Sigma, Taufkirchen
m-Maleimidobenzoyl sulfo succinimide-ester	Thermo Scientific, Waltham, USA

Maltose-monohydrate	Merck, Darmstadt
Methanol	Merck, Darmstadt
m ₇ G(5')ppp(5')G (m ₇ G-cap)	Kedar, Poland
Milk powder, dried, fat-free	Heirler, Radolfzell
Mowiol 4-88	Roth, Karlsruhe
New born calf serum (NCS)	PAA, Linz, Austria
n-Butanol	Merck, Darmstadt
n-Heptane	Merck, Darmstadt
Nonidet P-40, IGEPa CA-630 (NP-40)	Sigma, Taufkirchen
OPTI-MEM, Gibco	Invitrogen, Carlsbad, USA
Paraformaldehyde (PFA)	Sigma, Taufkirchen
pBR322 DNA-MspI Digest (BIM Marker)	NEB, Ipswich, USA
Penicillin/Streptomycin	Biochrom, Berlin
Perfluoropolyether	Hampton Research, Aliso Viejo, USA
Phenol/Chloroform/Isoamyl alcohol, Roti-Phenol/C/I (PCI)	Roth, Karlsruhe
Phenylmethylsulphonyl fluoride (PMSF)	Roche, Basel, Switzerland
Polyvinylpyrrolidone	Sigma, Taufkirchen
Ponceau S solution, 0.2% in 3% TCA	Serva, Heidelberg
Potassium acetate (KAc)	Merck, Darmstadt
Potassium chloride (KCl)	Merck, Darmstadt
Potassium dihydrogen phosphate (KH ₂ PO ₄)	Merck, Darmstadt
Potassium hydroxide (KOH)	Merck, Darmstadt
Precision Plus Protein Standard	Bio-Rad, Munich
Precision Plus Protein Standard, All Blue	Bio-Rad, Munich
Rotiphorese Gel 30 for protein gels (30% Acrylamide, 0.8% Bisacrylamide)	Roth, Karlsruhe
Rotiphorese Gel 40 for RNA gels (38% Acrylamide, 2% Bisacrylamide)	Roth, Karlsruhe
Rotiszont EcoPlus	Roth, Karlsruhe
Sephadex G-25 Superfine DNA Grade	GE Healthcare, Chalfont St Giles, UK
Silver nitrate (AgNO ₃)	Merck, Darmstadt
Sodium acetate (NaAc)	Merck, Darmstadt
Sodium azide (NaN ₃)	Roth, Karlsruhe
Sodium chloride (NaCl)	Merck, Darmstadt
Sodium citrate (NaCitrate)	Merck, Darmstadt
Sodium dihydrogen phosphate (NaH ₂ PO ₄)	Merck, Darmstadt
Sodium dodecyl sulfate (SDS)	Serva, Heidelberg
Sodium hydroxide (NaOH)	Merck, Darmstadt
Sodium thiocyanate (NaSCN)	Merck, Darmstadt

Sodium thiosulfate (Na ₂ S ₂ O ₃)	Merck, Darmstadt
Spermidine	Sigma, Taufkirchen
StrepTactin-Sepharose	IBA, Göttingen
Sypro Ruby staining solution	Invitrogen, Carlsbad, USA
Template Suppression Reagent (TSR)	Applied Biosystems, Foster City, USA
Transfection reagent, FuGENE	Roche, Basel, CH
Trichloroacetic acid (TCA)	Merck, Darmstadt
N, N, N', N'-Tetramethylethylenediamine (TEMED)	Sigma, Taufkirchen
Tris(hydroxymethyl)aminomethane (Tris)	BDH Chemicals, Poole, UK
Triton X-100	Sigma, Taufkirchen
Tween20	Roth, Karlsruhe
Uranyl formate	Polysciences, Warrington, USA
Urea	Merck, Darmstadt
Uridine-5'-triphosphate (UTP)	Jena Biosciences, Jena
[α - ³² P]-UTP	PerkinElmer, Waltham, USA
Xylene cyanol	Sigma, Taufkirchen

2.1.2 Oligonucleotides

All primers used in this study are listed below. Primers MG1-MG24 were used for routine cloning procedures as indicated in 2.1.3. Primers M7-F and M7-R were used for the amplification of MINX pre-mRNA template DNA (2.2.4.1) and oligonucleotides m6 and m12 for the purification of spliceosomal C complexes (2.2.10.5). All oligonucleotides were purchased from MWG Biotech, Ebersberg.

Oligo	Sequence (5' to 3')	Restriction Site
MG1	GGC CGC GGG CGG CAT CGA GGG CAG GGG CGG CGG CTG GTC CAA CCC ACA ATT CGA GAA GCC CGG GTA ATA AC	
MG2	CAT GGT TAT TAC CCG GGC TTC TCG AAT TGT GGG TGG GAC CAG CCG CCG CCG CCG CCC CTG CCC TCG ATG CCG CCC GC	
MG3	CCG GAT ATC ATG ACA ACA GCA GCC AGG CC	EcoRV
MG4	AAG GAA AAA AGC GGC CGC TTT AAT ATA TTT CTC CAT GAA CTT TTT GTG	NotI
MG5	CCG GAT ATC ATG CCT CGA ATT ATG ATC AAG	EcoRV
MG6	AAG GAA AAA AGC GGC CGC GAA TTT TGA CTT TAA AGT CTC TTT CTC	NotI
MG7	CCG GAT ATC ATG GAC GTG GGC GAA CTT CTG	EcoRV
MG8	CTT CCC GAT AGC GTC CGC GTT CTC GAG	
MG9	CTC GAG AAC GCG GAC GCT ATC GGG AAG	
MG10	ATA AGA ATG CGG CCG CGA AGT TCT CCA GCA AGC CCA G	NotI
MG11	CCG GAT ATC ATG TCC CTA ATC TGC TCC ATC TCT AAC	EcoRV
MG12	AAG GAA AAA AGC GGC CGC CAG GCT GTA GAA CTT GAG GCT TCT GTC	NotI

MG13	CCGGATATCATGGTCGAGGAGGTACAGAAAC	EcoRV
MG14	AAG GAA AAA AGC GGC CGC AAA TCT CTT TCT CTT GAT AAT TTC TG	NotI
MG15	CCG GAT ATC TCC AAG GGA CCT GCA GTT G	EcoRV
MG16	AAG GAA AAA AGC GGC CGC ATC AAC CTC TTC AAT GGT GGG C	NotI
MG17	CCG GAT ATC ATG GCG GGC ACA GGT TTG GTG	EcoRV
MG18	AAG GAA AAA AGC GGC CGC GAA GTC TTG CCG GAT GTT TTC TTT G	NotI
MG19	CGC GGT ACC ATG GGA ACA ACA GCA GCC AGG CCA AC	KpnI
MG20	AAG GAA AAA AGC GGC CGC TTT AAT ATA TTT CTC CAT GAA C	NotI
MG21	CGC GGT ACC ATG GTG AGC AAG GGC GAG GAG C	KpnI
MG22	CTA GCT AGC ACA ACA GCA GCC AGG CCA AC	NheI
MG23	CCG CTC GAG TTA TTT AAT ATA TTT CTC CAT GAA C	XhoI
MG24	CTA GCT AGC GTG AGC AAG GGC GAG GAG C	NheI
IL1	TAC GCT AGC ATG ACA ACA GCA GCC AGG CC	NheI
IL2	CTA GTG GAT CCC TAT TTA ATA TAT TTC TCC ATG	BamHI
IL3	TAC GCT AGC ATG GCG GGC ACA GGT TTG G	NheI
IL4	CTA GTG GAT CCC TAG AAG TCT TGC CGG ATG TTT TC	BamHI
M7-F	GGG GTA CCT AAT ACG ACT CAC TAT AGG GAG ACG GAA TTC GAG CTC GCC C	
M7-R	CGC GGA TCC CCA CTG GAA AGA CC	
m6	GGC GGT CTC GTC	
m12	CTC GTC GGC AGC	

2.1.3 Plasmids

The plasmids that were used in this study are listed below, together with restriction sites and primers used for the cloning procedure. The pEU3b vector was obtained from Invitrotech Inc. (Kyoto, Japan). The pEU3b derived plasmids were used for cell-free translation with wheat germ lysate (2.2.6). pcDNA3-N/C-SF-TAP and pIRESneo vectors were kindly provided by Dr. Marius Ueffing (Helmholtz Zentrum, München) and Dr. Gunter Meister (MPI for Biochemistry, Martinsried), respectively, and were used for the heterologous expression of proteins in HeLa cells (2.2.8.3). pMS2-MBP was a kind gift of Prof. Robin Reed (Harvard Medical School, Boston, USA) and was used for the expression of the MS2-MBP protein in *E. coli* (2.2.10.1). The PM5 plasmid was prepared by Dr. Sergey Bessonov, using the PYP pre-mRNA as a template, and served as a template for pre-mRNA production for the purification of spliceosomal C complexes (2.2.10.5). The PYP pre-mRNA was a kind gift of Prof. Chris Smith, University of Cambridge, Cambridge, UK (Wollerton et al., 2004). The pMINX vector was also utilized as a template to produce pre-mRNA substrate for *in vitro* splicing analysis (2.2.9). The plasmids pIRESneo-FLAG/HA-AD002 and pIRESneo-FLAG/HA-SPF27 were cloned by Dr. Ira Lemm. All other plasmids are based on the above mentioned vectors and were generated during this study. All plasmids contain an Ampicillin selection marker.

Plasmid	Primers	Restriction Sites	Reference
pEU3b			(Sawasaki et al., 2002)
pEU3b-SII	MG1, MG2	NotI/NcoI	
pEU3b-SII-AD002	MG3, MG4	EcoRV/NotI	
pEU3b-SII-CDC5L	MG5, MG6	EcoRV/NotI	
pEU3b-SII-CTNNBL1	MG7, MG8, MG9, MG10	EcoRV/NotI	
pEU3b-SII-hPrp19	MG11, MG12	EcoRV/NotI	
pEU3b-SII-PRL1	MG13, MG14	EcoRV/NotI	
pEU3b-SII-HSP73	MG15, MG16	EcoRV/NotI	
pEU3b-SII-SPF27	MG17, MG18	EcoRV/NotI	
pcDNA3-C-SF-TAP			(Gloeckner et al., 2007)
pcDNA3-C-SF-TAP-AD002	MG19, MG20	NotI/KpnI	
pcDNA3-C-SF-TAP-eGFP-AD002	MG20, MG21	NotI/KpnI	
pcDNA3-N-SF-TAP			(Gloeckner et al., 2007)
pcDNA3-N-SF-TAP-AD002	MG22, MG23	NheI/XhoI	
pcDNA3-N-SF-TAP-eGFP-AD002	MG23, MG24	NheI/XhoI	
pIRESneo-FLAG/HA			(Malik and Roeder, 2003)
pIRESneo-FLAG/HA-AD002	IL1, IL2	NheI/BamHI	
pIRESneo-FLAG/HA-SPF27	IL3, IL4	NheI/BamHI	
pMINX			(Zillmann et al., 1988)
pMS2-MBP			(Das et al., 2000)
pm5			(Anderson and Moore, 1997)

2.1.4 Enzymes and enzyme buffers

Enzymes & Buffers	Manufacturer
Alkaline Phosphatase (CIP), 10 U/ μ l	NEB, Ipswich, USA
Chymotrypsin	Sigma, Taufkirchen
Creatine Kinase	Roche, Basel, CH
DNA Ligase buffer, 10 \times	NEB, Ipswich, USA
Lys-C, endoproteinase	Roche, Basel, CH
PfuCloned DNA Polymerase, 2.5 U/ μ l	Stratagene, La Jolla, USA
PfuTurbo DNA Polymerase, 2.5 U/ μ l	Stratagene, La Jolla, USA
Pfu reaction buffer, 10 \times	Stratagene, La Jolla, USA
Proteinase K (PK), 20 U/ μ l	Sigma, Taufkirchen
Restriction enzymes, 20 U/ μ l	NEB, Ipswich, USA
Restriction enzyme buffer, 10 \times	NEB, Ipswich, USA

RNasin Ribonuclease Inhibitor, 40 U/ μ l	Promega, Madison, USA
T4 DNA Ligase, 400 U/ μ l	NEB, Ipswich, USA
T4 DNA Ligase buffer, 10 \times	NEB, Ipswich, USA
T7 RNA Polymerase	Marjan Askarian, MPI for BPC
Trypsin	Promega, Madison, USA
RQ1 RNase free DNase, 1 U/ μ l	Promega, Madison, USA
SP6 RNA Polymerase, 20 U/ μ l	NEB, Ipswich, USA
Subtilisin	Sigma, Taufkirchen
Taq DNA Polymerase, 50 U/ μ l	NEB, Ipswich, USA
Taq DNA Polymerase buffer, 10 \times	Jena Bioscience, Jena

2.1.5 Peptides

Peptide	Sequence (N to C)	Source/Reference
AD002 ₇₁₋₈₆	CAREKNRDRPTREHTTS	Dept. of Cellular Biochemistry (Makarova et al., 2004)
CDC5L ₁₀₆₋₁₂₄	KAAQRDNEEETDDPRKLKC	Dept. of Cellular Biochemistry (Makarova et al., 2004)
FLAG	DYKDDDDK	Sigma, Taufkirchen
hPrp19 ₁₇₆₋₁₉₁	LQDKATVLTTERKKRGC	Dept. of Cellular Biochemistry (Makarova et al., 2004)
PRL1 ₁₄₆₋₁₆₀	C+SEYRHPGASDRPQPT	Dept. of Cellular Biochemistry
SPF27 ₃₇₋₅₀	EEETRRYRPTKNYL+C	Dept. of Cellular Biochemistry

2.1.6 Antibodies

Antibody	Species	Purification	Source/Reference
anti-AD002 ₇₁₋₈₆	Rabbit, polyclonal	Affinity purified via peptide	Dept. of Cellular Biochemistry (Makarova et al., 2004)
anti-CDC5L ₁₀₆₋₁₂₄	Rabbit, polyclonal	Affinity purified via peptide	Dept. of Cellular Biochemistry (Makarova et al., 2004)
anti-CTNBL1	Rabbit, polyclonal	Antiserum	Dr. Dierk Ingelfinger, DKFZ, Heidelberg
anti-eGFP	Mouse, monoclonal	Affinity purified	Santa Cruz Biotech., Santa Cruz, USA
anti-FLAG	Rabbit, polyclonal	Affinity purified via peptide	Sigma, Taufkirchen
anti-hPrp19 ₁₇₆₋₁₉₁	Rabbit, polyclonal	Affinity purified via peptide	Dept. of Cellular Biochemistry
anti-HSC70 (K-19)	Goat, polyclonal	Affinity purified via peptide (C-terminus)	Santa Cruz Biotech., Santa Cruz, USA
anti-Mouse-HRP	Goat, polyclonal	Affinity purified	Jackson Inc., West Grove, USA
anti-PRL1 ₁₄₆₋₁₆₀	Rabbit, polyclonal	Affinity purified via peptide	Dep. Cellular Biochemistry

anti-Rabbit-HRP	Goat, polyclonal	Affinity purified	Jackson Inc., West Grove, USA
anti-Mouse-Cy5	Donkey, polyclonal	Affinity purified	Biomeda, Foster City, USA
anti-SC35	Mouse, monoclonal	Affinity purified	Santa Cruz Biotech., Santa Cruz, USA
anti-SPF27 ₃₇₋₅₀	Rabbit, polyclonal	Affinity purified via peptide	Dept. of Cellular Biochemistry
anti-StrepII	Mouse, monoclonal	Affinity purified	IBA, Göttingen

2.1.7 Bacterial strains

Escherichia coli strains HB101 and XL-1 blue were used for propagation of plasmid DNA (2.2.2) and BL21 (DE3) Rosetta2 cells were used for the expression of the MS2-MBP protein (2.2.10.1).

Bacterial strain	Genotype	Source
<i>E. coli</i> HB101	F ⁻ <i>proA2 hsdS20</i> (r _B ⁻ m _B ⁻) <i>recA13 ara-14 lacY1 galK2 rpsL20 supE44 rpsL20 supE44 proA2 xyl-5 mtl-1</i>	Promega, Mannheim
<i>E. coli</i> XL-1 blue	<i>recA1 endA1 gyrA96 thi-1 hsdR-17 supE44 relA1 lac[F' proAB lacI^q ΔM15 Tn10(Tet^R)]</i>	Stratagene, La Jolla, USA
<i>E. coli</i> BL21 (DE3) Rosetta2	[F ⁻ <i>ompT hsdS_B</i> (r _B ⁻ m _B ⁻), <i>gal [dcm] [lon]</i> (DE3), pRARE2 (Cm ^R)]	Merck, Darmstadt

2.1.8 Cell lines

Cell line	Description	Source/Reference
HeLa S3	HeLa S3 cell line is a subclone of the HeLa cell line. This line can be adapted to grow in suspension.	(Fisher and Puck, 1956) Computer Cell Culture Centre, Seneffe, Belgium
FLAG-AD002	HeLa S3 derived cell line stably expressing N-terminally FLAG/HA-tagged AD002	This study, generated by Dr. Ira Lemm
FLAG-SPF27	HeLa S3 derived cell line stably expressing N-terminally FLAG/HA-tagged SPF27	This study, generated by Dr. Ira Lemm
AD002-C-SF-TAP	HeLa S3 derived cell line stably expressing C-terminally FLAG/StrepII-tagged AD002	This study
AD002-N-SF-TAP	HeLa S3 derived cell line stably expressing N-terminally FLAG/StrepII-tagged AD002	This study
eGFP-AD002-C-SF-TAP	HeLa S3 derived cell line stably expressing C-terminally FLAG/HA-tagged eGFP-AD002	This study
eGFP-AD002-N-SF-TAP	HeLa S3 derived cell line stably expressing N-terminally FLAG/HA-tagged eGFP-AD002	This study

2.1.9 Common buffers

Commonly used media, buffers and solutions were prepared according to Sambrook and Russell, 2001. Deionized water was from a Millipore apparatus. Solutions were autoclaved (121°C, 20 min, 1 bar) or filter-sterilized (0.22 µm), if necessary. The pH was adjusted using HCl or NaOH if not stated otherwise.

Buffer	Ingredients
CE buffer	10 mM Cacodylic acid (pH 7.0) 0.2 mM EDTA (pH 8.0)
CIP buffer, 2×	50 mM Tris-HCl (pH 8.0)
Coomassie Brilliant Blue (CBB) G-250 staining solution	0.1 mM CBB G-250 35 mM HCl
Denhardt's reagent, 100×	2% (w/v) Polyvinylpyrrolidone 2% (w/v) Ficoll 400 2% (w/v) BSA
DNA loading buffer, 5×	1× TBE 30% (v/v) Glycerol 0.05 % (w/v) Xylene cyanol 0.05 % (w/v) Bromophenol blue
Gel fixation solution	40% (v/v) Methanol 10% (v/v) Acetic Acid
G150 buffer, 10×	200 mM HEPES-KOH, pH 7.9 15 mM MgCl ₂ 1.5 M NaCl
G250 buffer, 10×	200 mM HEPES-KOH 15 mM MgCl ₂ 2.5 M NaCl
30% (v/v) Glycerol in G150 buffer	10% (v/v) 10× G150 buffer 30% (v/v) Glycerol
10% (v/v) Glycerol in G150 buffer	10% (v/v) 10× G150 buffer 10% (v/v) Glycerol
5% (v/v) Glycerol in G150 buffer	10% (v/v) 10× G150 buffer 5% (v/v) Glycerol
Heparin elution buffer	20 mM HEPES-KOH (pH 7.9) 100 mM KCl 15% (v/v) Glycerol 0.5 mM DTT 0.2 mM PMSF
IPP150 buffer	20 mM HEPES-KOH (pH 7.9) 150 mM NaCl 1.5 mM MgCl ₂ 0.5 mM DTT 0.5% (w/v) NP-40

LB-Medium	4 capsules dissolved in 1000 ml ddH ₂ O (10 g Tryptone, 5 g Yeast Extract, 10 g NaCl)
LB-Agar medium	40 g dissolved in 1000 ml ddH ₂ O (10 g Tryptone, 5 g Yeast Extract, 10 g NaCl, 15 g Agar)
MC buffer, 10×	100 mM HEPES-KOH (pH 7.6) 100 mM KAc 5 mM Mg(Ac) ₂
MC buffer, 1×	10% (v/v) 10× MC buffer 5 mM DTE 1 Complete protease inhibitor tab / 50 ml
Na-PO ₄ , 25 mM (pH 6.5)	7.5 mM Na ₂ HPO ₄ 17.5 mM NaH ₂ PO ₄
NET ₁₅₀ buffer	50 mM Tris-HCl (pH 7.9) 150 mM NaCl 0.1% Triton X-100
NET ₃₀₀ buffer	50 mM Tris-HCl (pH 7.9) 300 mM NaCl 0.1% Triton X-100
PBS (pH 8.0), 10×	188 mM K ₂ HPO ₄ 12 mM KH ₂ PO ₄ 1.3 M NaCl
PBS (pH 7.4), 10×	160 mM K ₂ HPO ₄ 40 mM KH ₂ PO ₄ 1.3 M NaCl
PBS (pH 7.2), 10×	144 mM K ₂ HPO ₄ 56 mM KH ₂ PO ₄ 1.3 M NaCl
PK buffer, 2×	200 mM Tris-HCl (pH 7.5) 25 mM EDTA (pH 8.0) 2% (w/v) SDS
Pre-Hybridization Buffer	25 mM Na-PO ₄ (pH 6.5) 6× (v/v) SSC 5× (v/v) Denhardt's reagent 0.5% (w/v) SDS 50% (v/v) Formamide, deionised 0.1 mg/ml denatured herring sperm DNA
Protein electrophoresis buffer, 10×	250 mM Tris 1.92 M Glycine 1% (w/v) SDS
RNA gel extraction buffer	20 mM Tris-HCl (pH 7.5) 0.2 mM EDTA (pH 8.0) 150 mM NaCl 0.5% (w/v) SDS

RNA sample buffer (FA buffer)	80% (v/v) Formamide 1 mM EDTA (pH 8.0) 0.05% (w/v) Bromophenol blue 0.05% (w/v) Xylene cyanol
Roeder A buffer, 10×	100 mM HEPES-KOH (pH 7.9) 100 mM KCl 15 mM MgCl ₂
Roeder C buffer, 1×	20 mM HEPES-KOH (pH 7.9) 420 mM NaCl 1.5 mM MgCl ₂ 0.2 mM EDTA (pH 8.0) 25 % (v/v) Glycerin
Roeder D buffer, 10×	200 mM HEPES-KOH (pH 7.9) 1 M KCl 2 mM EDTA (pH 8.0) 15 mM MgCl ₂
Roeder D buffer, 1×	10% (v/v) 10× Roeder D buffer 10% (v/v) Glycerol 0.5 mM DTT 0.5 mM PMSF
SEC100 buffer	10 mM Tris-HCl (pH 8.0) 100 mM NaCl 1 mM DTT
SDS gel loading buffer, 6×	244 mM Tris-HCl (pH 6.8) 2.4% (w/v) SDS 244 mM DTT 21.2% (v/v) Glycerol 0.025 % (w/v) Bromophenol blue
Semi-dry blotting buffer	48 mM Tris 39 mM Glycine 0.1% (w/v) SDS 20% (v/v) Methanol
SLAB4	50 mM Tris-HCl (pH 8.5) 380 mM Glycine 0.1% (w/v) SDS
Sonication buffer	20 mM Tris-HCl (pH 7.6) 200 mM NaCl 1 Complete protease inhibitor tab / 50 ml
SSC buffer, 20×	3 M NaCl 0.3 M Na Citrate
T7 RNA polymerase buffer, 5×	1 M Hepes-KOH (pH 7.5) 160 mM MgCl ₂ 200 mM DTT 10 mM Spermidine

TBE, 10×	900 mM Tris 890 mM Boric acid 20 mM EDTA (pH 8.0)
TBS, 10×	200 mM Tris-HCl (pH 7.5) 1.5 M NaCl
Trypsin/EDTA	0.05% Trypsin 0.02% EDTA (pH 8.0) in 1 × PBS (pH 7.4)

2.1.10 Commercial reagent sets

Reagent Set (Kit)	Manufacturer
AmpliTaq FS Big Dye Terminator Kit	Applied Biosystems, Foster City, USA
BCA Protein Assay Kit	Thermo Scientific, Waltham, USA
Crystallization Kits	Qiagen, Hilden
The AmSO ₄ Suite	
The Classics Suite	
The JCSG+ Suite	
The PACT Suite	
The PEGs Suite	
The pH Clear Suite	
The Protein Complex Suite	
The SmI Suite	
Crystallization Kits	Hampton Research, Aliso Viejo, USA
Index	
Additive Screen I	
DyeEx Kit	Qiagen, Hilden
ECL Advance Western Blotting Detection Kit	GE Healthcare, Chalfont St Giles, UK
JET star 2.0, Plasmid Purification MAXI Kit	Genomed, Löhne
Prime It II Random Primer Labeling Kit	Stratagene, La Jolla, USA
QIAprep Spin Miniprep Kit	Qiagen, Hilden
QIAprep Maxiprep Kit	Qiagen, Hilden
QIAquick Gel Extraction Kit	Qiagen, Hilden
QIAquick PCR Purification Kit	Qiagen, Hilden
SulfoLink Immobilization Kit for Peptides	Thermo Scientific, Waltham, USA
Zoengene Wheat Germ Cell-Free Protein Synthesis System	Zoengene, Japan

2.1.11 Working equipment

Working Material	Manufacturer
Bio-Spin chromatography columns (1 ml, 2 ml)	Bio-Rad, Munich
Centrifuge tubes (TH660, TST41.14, 45 Ti, S100AT)	Beranek, Weinheim
Concentrator Ultracel – 100 kDa (15 ml)	Millipore, Billerica, USA
Concentrator Ultracel – 100 kDa (4 ml)	Millipore, Billerica, USA
Copper grids	Agar Scientific, Stansted, UK
Corex glass tube	Thermo Scientific, Waltham, USA
Cover slip	Menzel-Gläser, Braunschweig
Cryo freezing container “Mr. Frosty”	Thermo Scientific, Waltham, USA
Cryovials (1 ml, 5 ml)	Greiner bio-one, Kremsmünster, Austria
Dialyses membrane (MWCO 6-8 kD)	SpectrumLabs, Rancho Dominguez, USA
Dounce Homogenizer	Kimble/Kontes, Vineland, USA
Dewar flasks	KGW-Isotherm, Karlsruhe
Electroporation cuvettes, 2 mm	Bio-Rad, Munich
Falcon Tubes (15 ml, 50 ml)	BD Biosciences, San Jose, USA
Fluor-coated TLC Plate	Applied Biosystems, Foster City, USA
Glass ware	Schott, Mainz
LINBRO 24 well plate, hanging drop	Hampton Research, Aliso Viejo, USA
Microfuge tubes	Eppendorf, Hamburg
Micro-Bio Spin 6 gel filtration columns	Bio-Rad, Munich
Microscope slide	ICN Biomedicals, Aurora, USA
MRC 96 well plate, sitting drop	Hampton Research, Aliso Viejo, USA
Nesco-/Parafilm	Roth, Karlsruhe
NuPAGE gel (1 mm, 4-12%)	Invitrogen, Carlsbad, USA
Nylonmembrane Hybond XL	GE Healthcare, Chalfont St Giles, UK
Petri dish	Sarstedt, Nümbrecht
Pipette tips (10 µl, 200 µl, 1000 µl)	Starlab, Milton Keynes, UK
Protran nitrocellulose membrane	Schleicher & Schüll, Dassel
Scintillation hinge cap vial, 8 ml	Perkin Elmer, Woodbridge, CA
Slide-A-Lyzer, dialysis cassette	Thermo Scientific, Waltham, USA
Spinner flasks (25 ml-1000 ml)	BellCo, Vineland, USA
Sterile filter (0.2 µm, 0.45 µm)	Sarstedt, Nümbrecht
Superose 6 PC 3.2/30	GE Healthcare, Chalfont St Giles, UK
Superose 6 GL 10/300	GE Healthcare, Chalfont St Giles, UK
Tissue culture plastic	Greiner bio-one, Kremsmünster, A
Whatman 3MM paper	GE Healthcare, Chalfont St Giles, UK
X-ray film, Amersham Hyperfilm ECL	GE Healthcare, Chalfont St Giles, UK
X-ray film, BioMax MR	Kodak, Rochester, USA

2.1.12 Instruments

Instrument	Manufacturer
4800 MALDI ToF/ToF Analyzer	Applied Biosystems, Foster City, USA
ABI Prism 310 Genetic Analyzer	Applied Biosystems, Foster City, USA
Äkta Explorer	GE Healthcare, Chalfont St Giles, UK
Äkta Purifier	GE Healthcare, Chalfont St Giles, UK
Autoclaves	Thermo Scientific, Waltham, USA
Balance, special accuracy	Sartorius, Göttingen
Beamline PX2	Swiss Light Source, SLS, Villigen, CH
Beckman 45Ti rotor	Beckman Coulter, Fullerton, USA
Beckman XL-A Analytical Ultracentrifuge	Beckman Coulter, Fullerton, USA
Biofuge fresco	Thermo Scientific, Waltham, USA
Biofuge pico	Thermo Scientific, Waltham, USA
Bioreactor	Applikon Biotek, Dresden
Cartesian crystallization robot, 4 channels	Zinsser analytic, Frankfurt
Cell incubator, BBD 6220	Thermo Scientific, Waltham, USA
CM 200 FEG electron microscope	Philips, Eindhoven, NL
Confocal fluorescence microscope, LSM 510 Meta	Carl Zeiss, Jena
Cryofuge 6000i	Thermo Scientific, Waltham, USA
DNA Thermal Cycler, Hybaid OmniGene	Thermo Scientific, Waltham, USA
DNA Thermal Cycler, T3 Thermocycler	Biometra, Göttingen
Electrophoresis power supply, 2A 2000	Hoefer Inc., Holliston, USA
Electrophoresis power supply, 3501/XL	GE Healthcare, Chalfont St Giles, UK
Electrophoresis power supply, 200/2.0	Bio-Rad, Munich
Fluorescence imager, FLA-7000	Fujifilm Corp., Japan
Flow cytometer, Casy	Schärfe System, Reutlingen
Fraction collector Frac-100	GE Healthcare, Chalfont St Giles, UK
Gel electrophoresis equipment (horizontal)	Bio-Rad, Munich
Gel electrophoresis equipment (vertical)	In-house, MPI for BPC
Gel documentation unit	Bio-Rad, Munich
Geldryer model 583	Bio-Rad, Munich
Gradient master model 106	BioComp Instr., Fredericton ,Canada
“head over tail” rotor	Cole-Parmer, Vernon-Hills, USA
Hybridization oven, Hybaid Biometra	Thermo Scientific, Waltham, USA
Ice machine	Ziegra, Stockport, UK
Incubator, BK-600	Thermo Scientific, Waltham, USA
Incubator, HT Multitron culture shaker	Infors, Bottmingen, CH
Incubator, Kelvitron kp	Thermo Scientific, Waltham, USA
Inverted light microscope	Carl Zeiss, Jena

Laboratory freezer, -20°C	Liebherr, Bulle, CH
Laboratory freezer, -80°C	Thermo Scientific, Waltham, USA
Laboratory refridgerator, 4°C	Liebherr, Bulle, CH
Laminar flow cabinet, HERAsafe	GE Healthcare, Chalfont St Giles, UK
Liquid nitrogen freezer, K-Series	Taylor-Wharton, Theodore, USA
LTQ-Orbitrap	Thermo Scientific, Waltham, USA
Magnetic stirrer/heater	IKA, Staufen
Megafuge 1.0R	Thermo Scientific, Waltham, USA
Micromass qToF II Ultima	Waters, Milford, USA
MicroPulser	Bio-Rad, Munich
Microwave	Sharp Electronics, Uxbridge, UK
Milli-Q-water supply apparatus	Millipore, Billerica, USA
Nanodrop ND1000	Kisker Biotech, Steinfurt
Peristaltic pump P-1	GE Healthcare, Chalfont St Giles, UK
pH Meter	Mettler-Toledo, Giessen
Phosphorimager Typhoon 8600	GE Healthcare, Chalfont St Giles, UK
Phosphor screen	GE Healthcare, Chalfont St Giles, UK
Pipette	Eppendorf, Hamburg
Pipetus	Hirschmann Laborgeräte, Eberstadt
Rocking platform, 3019	GFL, Großburgwedel
Rocking platform, Unimax 1010	Heidolph, Kelheim
Rocking platform, WT 16	Biometra, Göttingen
Scintillation counter	Beckman Coulter, Fullerton, USA
Semi-dry blot unit	Peqlab, Erlangen
Single path monitor UV-1	GE Healthcare, Chalfont St Giles, UK
SMART system HPLC	GE Healthcare, Chalfont St Giles, UK
Sonifier II 250 Digi	G.Heinemann, Schwäbisch Gmünd
Sorvall HB-6 rotor	Thermo Scientific, Waltham, USA
Sorvall S100AT rotor	Thermo Scientific, Waltham, USA
Sorvall SLA-1500 rotor	Thermo Scientific, Waltham, USA
Sorvall SS-34 rotor	Thermo Scientific, Waltham, USA
Sorvall TH660 rotor	Thermo Scientific, Waltham, USA
Sorvall TST 41.14 rotor	Thermo Scientific, Waltham, USA
Spectrophotometer Ultrospec 3000 pro	GE Healthcare, Chalfont St Giles, UK
Speed Vac Concentrator 5301	Eppendorf, Hamburg
Thermomixer comfort	Eppendorf, Hamburg
TemCam-F415, 4kx4k CCD camera	TVIPS, Gauting
Triple Quadrupole/Linear Ion Trap	Applied Biosystems, Foster City, USA
Ultracentrifuge, Sorvall Discovery 90	Thermo Scientific, Waltham, USA

Ultracentrifuge, Sorvall Discovery M150	Thermo Scientific, Waltham, USA
Ultracentrifuge, Sorvall Evolution RC	Thermo Scientific, Waltham, USA
UV lamps, 254 nm/365 nm	Bachofer, Reutlingen
UV-Stratalinker 2400	Stratagene, La Jolla, USA
Vortex Genie 2	IKA, Staufen
Wet-Blot unit	Bio-Rad, Munich
X-ray film developer X-Omat 2000	Kodak, Rochester, USA

2.1.13 Software

Software	Source/Reference
Collaborative Computational Project 4 (CCP4i) Suite	(Potterton et al., 2003)
Coot	(Emsley and Cowtan, 2004)
HHpred	(Soding et al., 2005)
HKL2000	(Otwinowski and Minor, 1997)
Image Gauge v4.22	Fujifilm Corp., Japan
IMAGIC	(van Heel et al., 1996)
MASCOT	(Perkins et al., 1999)
MultiQuant 1.0	Applied Biosystems/MDS Sciex, Foster City, USA
Pymol	http://pymol.sourceforge.net/
SEDFIT	(Schuck, 2000)
SEDNTERP	(Schuck, 2003)
SEDPHAT	(Schuck, 2003)
STAMP	(Russell and Barton, 1992)
XDS	(Kabsch, 1988)

2.2 Methods

2.2.1 Molecular biological standard methods

2.2.1.1 Concentration determination of nucleic acids

To determine the concentration of nucleic acids, the absorption in an aqueous solution was measured at wavelengths of 260 nm and 280 nm in units of optical density (OD). The concentration was then calculated using pre-determined absorption values at 260 nm as described (Sambrook and Russell, 2001). The ratio of OD_{260}/OD_{280} determined the purity of a nucleic acid solution, which is 2.0 for pure RNA or oligonucleotides, and 1.8 for pure DNA. Lower ratios show protein contamination, which had to be removed by phenol/chloroform/isoamyl alcohol extraction (2.2.1.3).

1 OD_{260} :	50 $\mu\text{g/ml}$ double-stranded DNA
1 OD_{260} :	33 $\mu\text{g/ml}$ single-stranded DNA
1 OD_{260} :	40 $\mu\text{g/ml}$ single-stranded RNA

2.2.1.2 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was the method of choice for the size separation of DNA and was used both for analytical purposes and to purify preparative amounts of DNA. Depending on the length of PCR products or plasmids, gels contained 1-2% agarose in $1\times$ TBE buffer containing 0.1 $\mu\text{g/ml}$ ethidium bromide. Samples were supplemented with $5\times$ DNA loading dye prior to loading. Gels were run in self-made 7 cm \times 10 cm gel chambers at 130 Volts for approximately 1 hr in $1\times$ TBE buffer. 100 or 1,000 bp DNA ladders (GeneRuler) were used as a size standard. DNA bands were visualized by illumination with UV-light (254 nm, analytical scale; or 365 nm, preparative scale).

2.2.1.3 Phenol/Chloroform/Isoamyl alcohol extraction

The phenol/chloroform/isoamyl alcohol extraction (PCI extraction) was used for separating proteins from nucleic acids in an aqueous solution. The solution was extracted with the same volume of PCI solution and thoroughly mixed. To separate the aqueous and organic phases, the suspension was centrifuged for 5 min at room temperature (RT) and 13,000 rpm in a microfuge. The nucleic acids in the upper aqueous phase were recovered.

2.2.1.4 Precipitation of nucleic acids

Precipitation was performed by addition of 1/10 volume 3 M NaAc (pH 5.2) and 2.5 volumes 100% ice-cold ethanol (p.a.). To improve precipitation of small amounts of nucleic acids, 10 µg of glycogen were added. Nucleic acids were precipitated either for at least 1 hr at -80°C or for at least 2 hrs at -20°C. Subsequently, they were pelleted by centrifugation for 20 min at 4°C and 13,000 rpm in a microfuge. The pellet was washed twice with 70% (v/v) ice-cold ethanol and pelleted again for 10 min at 4°C and 13,000 rpm. The RNA or DNA pellet was finally air-dried and resuspended in an appropriate buffer or ddH₂O.

2.2.1.5 Purification of DNA from agarose gels

In order to isolate DNA from agarose gels (2.2.1.2), DNA bands illuminated with UV-light at a wavelength of 365 nm were cut out of the gel using a sterile razor blade. DNA was extracted from the gel using the QIAquick Gel Extraction Kit according to the manufacturer's manual.

2.2.2 Molecular cloning in plasmid vectors

Cloning a DNA fragment (insert) into a plasmid vector involved several steps. First, the insert was generated by the polymerase chain reaction (PCR) (2.2.2.1). The primers for PCR were designed in a way that produces compatible termini with the plasmid vector. The insert and also the circular plasmid DNA were cleaved with one or more restriction enzymes (2.2.2.2), purified via agarose gel electrophoresis (2.2.1.2, 2.2.1.5) and ligated *in vitro* (2.2.2.3). The products of the ligation reaction were then used to transform (2.2.2.5) a competent strain of *E. coli* (2.2.2.4). The resulting transformed colonies, grown under selective conditions, were screened by digestion with restriction enzymes (2.2.2.2) to identify those that carry the desired DNA sequence. Finally, the plasmid was purified in an appropriate scale (2.2.2.6) and sequenced (2.2.2.7).

2.2.2.1 Polymerase chain reaction (PCR)

PCR was used to amplify a particular region of DNA across several orders of magnitude, generating multiple copies of the specific DNA sequence (Saiki et al., 1988). For the amplification of inserts for molecular cloning, the Pfu DNA polymerase system was used according to the manufacturer's instructions. In brief, a typical 100 µl reaction contained 30 ng DNA template, 25 pmol of both primers (forward/reverse), 0.6 mM dNTPs, 10 µl of 10× Pfu reaction buffer and 2 µl of 2.5 U/µl PfuCloned (or PfuTurbo) DNA polymerase. The PCR was performed in a PCR cycler using the following program:

1×	95°C	2 min	initial denaturation
35×	95°C	30 sec	denaturation
	50-70°C	30 sec	annealing
	72°C	1 min / 1 kb	elongation
1×	72°C	10 min	final elongation

The annealing temperature was usually chosen to be 5°C below the melting temperature of the used primers. The elongation time was always 1 min per 1 kb length of the amplification sequence with the minimum of 1 min incubation (for sequences shorter than 1 kb). After PCR, the product was purified from unincorporated nucleotides and DNA polymerase with the QIAquick PCR Purification Kit, as described in the manufacturer's manual and finally analyzed by agarose gel electrophoresis (2.2.1.2).

2.2.2.2 Endonucleolytic restriction of DNA

Restriction endonucleases cleave double stranded DNAs at a specific recognition sequence (restriction site). Restriction enzymes can cut both DNA strands at different positions, producing overhangs ("sticky ends") or at the same position generating DNA fragments without overhangs ("blunt ends"). Endonucleolytic restriction of plasmid vectors or inserts generated by PCR was performed with commercially available buffer solutions and restriction enzymes (2.1.4). A typical 50 µl reaction contained up to 10 µg of DNA, 5 µl 10× restriction enzyme buffer, 5 µl 10× BSA solution (optionally) and the restriction enzyme in an appropriate amount (e.g., 1 unit of the enzyme per 1 µg of DNA). The solution was incubated for 12 hrs at a temperature recommended for each enzyme by the manufacturer. In the case of analytical digestion screens (<500 ng DNA), the reaction volume was lowered to 15 µl and the incubation time was 2 hrs.

2.2.2.3 DNA ligation

Ligation of an insert, generated by PCR, and a linearized plasmid vector was usually performed at a molar ratio of 1:1 or 5:1 (insert relative to vector). A typical ligation reaction was performed in a volume of 20 µl and contained 100 ng of plasmid DNA and a respective amount of insert, together with 2 µl 10× T4 DNA ligase buffer and 1 µl T4 DNA ligase. The ligation reaction was incubated at 16°C for at least 12 hrs. The ligated plasmid was always purified by butanol precipitation. Therefore, 1 µl 10 mg/ml glycogen was added to the reaction, together with 10 volumes of n-butanol. After thorough mixing, the solution was kept for at least 2 hrs at -20°C and then pelleted by centrifugation for 20 min at 4°C and 13,000 rpm in a microfuge. The DNA pellet was washed twice with 70% (v/v) ice-cold ethanol and pelleted again for 10 min at 4°C and 13,000 rpm in a microfuge. The DNA was finally air-dried and resuspended in 8 µl ddH₂O.

2.2.2.4 Preparation of electrocompetent *E. coli* cells

For preparation of electrocompetent cells, *E. coli* HB101 or XL1-blue strains were used. First, 5 ml of LB medium were inoculated with a single colony of the respective *E. coli* strain and incubated overnight at 37°C, while rapidly shaking at 250 rpm. Subsequently, the overnight culture was used to inoculate 500 ml of LB medium and incubated at 37°C with 250 rpm agitation. When the OD₆₀₀ of the culture reached 0.4-0.7, the flasks were transferred to ice and cooled down for 15 min. For maximum efficiency of transformation all further steps were performed on ice or at 4°C. Cells were harvested by centrifugation for 10 min at 5,000 rpm and 4°C in a Sorvall SLA-1500 rotor, using a Sorvall evolution RC centrifuge. The cell pellet was carefully washed with 50 ml of ice-cold water. To remove salt, the pelleted cells were resuspended in 500 ml of ice-cold water and centrifuged again under the described conditions. The pelleted cells were resuspended in 50 ml of ice-cold 15% (w/v) glycerol and transferred into 50 ml plastic tubes. The cells were harvested by centrifugation for 15 min at 4,000 rpm and 4°C in a Megafuge 1.0R centrifuge. Finally, cells were resuspended in 1.5-fold pellet volumes of ice-cold 15% (w/v) glycerol solution, aliquoted into 60 µl aliquots, frozen in liquid nitrogen and stored at -80°C.

2.2.2.5 Electrotransformation

During transformation, the plasmid vector is taken up by the bacterial cell. Exposure of the *E. coli* cell to an electrical charge destabilizes the membranes and induces the formation of transient membrane pores, through which the DNA molecule can pass (Sambrook and Russell, 2001). For electroporation, 4 µl of the ligated plasmid DNA (2.2.2.3) were added to 60 µl electrocompetent *E. coli* cells on ice. The mixture was transferred to a pre-cooled electroporation cuvette and electroporated in a MicroPulser (2.5 kV, 5 ms). Subsequently, cells were transferred into 1 ml pre-warmed LB medium at 37°C and incubated for 1 hr with shaking. Afterwards, up to 200 µl of the suspension were plated onto LB agar containing the respective antibiotics for selection (e.g. 75 µg/ml ampicillin, 25 µg/ml chloramphenicol) and incubated overnight at 37°C. Single colonies were used to inoculate 5 ml LB medium containing the corresponding antibiotic, and the culture was incubated at 37°C overnight while shaking at 250 rpm. Plasmid DNA was recovered with the QIAprep Spin Miniprep Kit according to the manufacturer's protocol and subsequently screened by digestion with appropriate restriction enzymes (2.2.2.2), to identify those plasmid vectors that carry the desired DNA sequence.

2.2.2.6 Plasmid purification

For large scale plasmid purification, 5 ml of LB medium were inoculated with a single colony of the respective *E. coli* strain possessing the desired plasmid and incubated overnight at 37°C with 250 rpm agitation. The overnight culture was used to inoculate 250 ml of LB medium, which were incubated overnight at 37°C with 250 rpm agitation. Plasmid DNA was recovered from the cells with the QIAprep Maxiprep Kit or the JET Star 2.0 MAXI Kit according to the manufacturer's protocol.

2.2.2.7 Sequencing

For sequencing of new constructs, 300 ng of the plasmid DNA, 0.5 µl DMSO, 3.2 pmol sequencing primer, 2 µl of 5× Big Dye buffer and 4 µl Big Dye mix (AmpliTaq FS Big Dye Terminator Kit) were used in a 20 µl reaction and subjected to the following sequencing program:

1×	95°C	2 min	initial denaturation
25×	95°C	30 sec	denaturation
	55°C	20 sec	annealing
	60°C	4 min	elongation

Purification was performed with the DyeEx Kit according to the manufacturer's protocol. The eluate was redissolved in 25 µl of TSR-buffer (Template suppression reagent) and sequenced on an ABI PRISM 310 Genetic Analyzer. Alternatively, sequencing was performed at Seqlab, Sequencing Laboratories, Göttingen.

2.2.3 RNA biochemical standard methods

2.2.3.1 Denaturing polyacrylamide gel electrophoresis of RNA

Denaturing polyacrylamide gel electrophoresis in the presence of urea was used to size separate RNA for analytical purposes and also to preparatively purify *in vitro* transcribed, radiolabeled pre-mRNA. According to the length of nucleic acids, the polyacrylamide content varied between 5 and 12%. Denaturing gels typically contained 8 M urea, 1× TBE buffer and an appropriate concentration of Rotiphorese Gel 40. Polymerization was accomplished by adding 10 µl 10% (w/v) APS solution and 1 µl TEMED per 1 ml of the gel solution. Gel electrophoresis was performed in 1× TBE buffer at 20 W for up to 2 hrs. Usually, the RNA samples were treated with Proteinase K (2.2.3.7) prior to polyacrylamide gel analysis. The RNA pellet was then dissolved in RNA loading buffer and denatured for 3 min at 96°C before loading. RNA was visualized by silver staining (2.2.3.2), autoradiography (2.2.3.3) or northern blotting (2.2.3.4).

2.2.3.2 Silver staining of RNA in polyacrylamide gels

Silver staining of denaturing gels was adapted from Blum et al., 1987. During the procedure, minimally a 10-fold volume of each solution (relative to the gel volume) was used and all steps were performed on a shaker. First, the gel was fixed in a solution of 40% (v/v) methanol, 10% (v/v) acetic acid for at least 30 min or overnight. After fixation, the gel was washed twice with 500 ml of 10% (v/v) ethanol, 5% (v/v) acetic acid for 15 min in order to remove interfering substances. After that, the gel was washed briefly with water and subsequently stained in 250 ml of 12 mM AgNO₃ solution for 30 min. The stained gel was then washed twice with water and once with 500 ml of the developing solution [0.28 M Na₂CO₃, 0.0185% (v/v) formaldehyde] to remove excess silver from the surface of the gel. The brownish solution was immediately discarded and the gel was developed in 500 ml fresh developing solution. After appearance of the bands, the developing solution was discarded and the reaction was stopped with 5% (v/v) acetic acid.

2.2.3.3 Autoradiography of radioactively labeled RNA in polyacrylamide gels

For the detection of radiolabeled RNA in polyacrylamide gels, the gels were fixed in 40% (v/v) methanol/10% (v/v) acetic acid for at least 30 min and then either stained with silver (2.2.3.2) and subsequently washed with water for 10 min or directly washed with water without staining. After washing, the gel was transferred to a Whatman paper and dried under vacuum at 80°C for 1 hr. The dried gel was exposed for an appropriate time interval either to an X-ray film (e.g. at -80°C for 24 hrs), or to a phosphor screen (e.g. at room temperature for 12 hrs). Afterwards, the autoradiography film was developed in a Kodak X-Omat developer or the phosphor screen was scanned at 633 nm in a Typhoon variable mode imager.

2.2.3.4 Northern blot

For detection of RNAs by northern blotting, a radioactive probe was prepared using the Prime It II Random Primer Labeling Kit. In this procedure, DNA is radioactively labeled with [α^{32} P]-dATP using randomly hybridizing 9-mer oligonucleotides as primer. As a template, U5 and U6 snRNA PCR products synthesized from a plasmid containing the gene of these snRNAs were used. Initially, a mixture of 25 ng of template DNA, 8 μ l of a 9-mer oligonucleotide mixture (27 OD units/ml) and 17 μ l ddH₂O were mixed and denatured at 96°C for 5 min. Then 8 μ l 5 \times dATP buffer, 5 μ l [α^{32} P]-dATP (6,000 Ci/mmol), and 1 μ l Klenow polymerase were added, mixed, and incubated at 40°C for 10 min. Next, RNA was separated by denaturing PAGE and transferred to a Hybond XL membrane using the semi-dry blotting method. The apparatus for semi-dry blotting consists of two electrode plates (anode on bottom, cathode on top), separated by a horizontal sandwich of 5 layers of

Whatman paper, the membrane, gel and again 5 layers of Whatman paper (direction from bottom to top). Whatman papers and membrane were pre-soaked in 1× TBE. After assembling the sandwich, air bubbles trapped between the layers were removed by rolling a glass pipette over the sandwich. Blotting was performed at 2 mA/cm² of the membrane at a maximum of 25 V for 2 hrs at room temperature. After the transfer, the top layers of Whatman paper were removed and the membrane marked on the side touching the gel. The membrane was irradiated with 1,200 mJ UV in a Stratalinker apparatus, subsequently washed once with 25 mM sodium phosphate buffer (pH 6.5) and pre-hybridised in 50 ml pre-hybridization buffer for at least 1 hr at 42°C. Subsequently, the radiolabeled probe was added to the sample and incubated overnight at 42°C. The membrane was washed twice for 10 min with 100 ml 2× SSC/0.5% (w/v) SDS and then twice for 10 min with 100 ml of 2× SSC/0.1% (w/v) SDS and twice for 10 min with 100 ml of pre-warmed 2× SSC/0.1% (w/v) SDS at 50°C. Signals were visualized by autoradiography (2.2.3.3). To reuse the blot with another probe, the hybridized probe was removed by incubating the blot for 30 min with 100 ml 0.1× SSC/0.1% (w/v) SDS at 100°C.

2.2.3.5 Purification of RNA from polyacrylamide gels

To purify RNA from contaminating nucleic acids, nucleotides and proteins, RNA was separated on a 5% denaturing polyacrylamide gel (2.2.3.1). To recover the RNA from the gel, it must be localized in the gel. In the case of radiolabeled RNA, this was done by exposing the gel to an X-ray film. Unlabeled RNA was localized by UV shadowing, using shortwave UV light (254 nm) and a fluor-coated TLC plate to visualize the RNA in the gel. The gel piece containing the RNA was excised with a sterile razor blade, cut into small pieces and incubated with 500 µl RNA gel extraction buffer overnight at room temperature with agitation. The RNA was further purified by PCI extraction (2.2.1.3) and recovered from the aqueous phase by ethanol precipitation (2.2.1.4). After precipitation the RNA pellet was dried and dissolved in CE buffer.

2.2.3.6 Purification of RNA by gel filtration

Alternatively, RNA was purified via gel filtration. Therefore, Sephadex G-25 Superfine DNA Grade was pre-soaked for 3 hrs in an excess of ddH₂O at room temperature. An empty 2 ml Bio-Spin chromatography column was placed into a 2 ml microfuge tube and 1 ml bed volume of the soaked material was filled into the column. The assembly was placed into a 15 ml plastic tube and centrifuged for 1 min at 1,900 rpm and 4°C in a Megafuge 1.0R centrifuge to remove residual water. After centrifugation, the column was transferred to a new 2 ml microfuge tube and 100 µl of the transcription reaction were pipetted on top of the gel bed and again centrifuged for 2 min at 1,900

rpm and 4°C. The RNA was recovered from the eluate by ethanol precipitation (2.2.1.4). After precipitation, the RNA pellet was dried and dissolved in CE buffer or water.

2.2.3.7 Proteinase K digestion

To purify RNA from contaminating proteins, samples were digested with Proteinase K (PK). In a standard PK digestion of a splicing reaction, 15 µl of the reaction were supplemented with 11 µl ddH₂O, 30 µl 2× PK buffer, 2 µl 10 mg/ml glycogen, and 2 µl 10 mg/ml PK, and incubated for 30 min at 45°C. After addition of 130 µl ddH₂O, the RNA was subjected to PCI extraction (2.2.1.3), followed by ethanol precipitation to recover the RNA (2.2.1.4).

2.2.4 *In vitro* transcription of radiolabeled pre-mRNA

Radiolabeled MINX pre-mRNA was used for *in vitro* splicing reactions. The MINX RNA contains a duplicate of the leader 2 sequence of the major late transcription unit of adenovirus type 2, separated by one intron (Zillmann et al., 1988). For the assembly and purification of spliceosomal C complexes, the PM5 pre-mRNA was used (Bessonov et al., 2008). The PM5 pre-mRNA does not contain a 3' exon, but a prolonged polypyrimidine tract, and is preceded by three replications of the MS2 coat protein binding sequence (MS2 RNA aptamer) at its 5' end.

2.2.4.1 DNA template production by PCR

MINX pre-mRNA transcription template was generated by PCR from pMINX plasmid. Reaction mixtures contained 1× Taq DNA Polymerase buffer, 0.4 mM dNTPs, 4 mM MgCl₂, 1 µM M7-F, 1 µM M7-R, 100 ng pMINX and 200 U/ml Taq DNA Polymerase. The following PCR cyclers conditions were used for amplification:

1×	95°C	2 min	initial denaturation
35×	95°C	30 sec	denaturation
	55°C	1 min	annealing
	72°C	1 min	elongation
1×	72°C	5 min	final elongation

The final template was obtained by restriction digest of the PCR template with the *Xba* I restriction enzyme as described (2.2.2.2) and purified by agarose gel electrophoresis (2.2.1.5). The PM5 pre-mRNA transcription template was generated by restriction digest of the PM5 plasmid with the *Bam*H I restriction enzyme as described (2.2.2.2) and purified by agarose gel electrophoresis (2.2.1.5).

2.2.4.2 *In vitro* transcription reaction

Uniformly [³²P]-labeled, m₇G(5')ppp(5')G-capped pre-mRNA was synthesized *in vitro* by runoff transcription using bacteriophage T7 RNA polymerase. The reaction conditions were adapted from Pokrovskaya and Gurevich, 1994. A typical 100 µl transcription reaction contained 20 µl 5× T7 RNA polymerase buffer, ATP and CTP (7.5 mM each), UTP and GTP (1.5 mM each), 0.42 µM [α-³²P]-UTP (3,000 Ci/mmol), 0.5-2 µg of DNA template, 5 mM m₇G(5')ppp(5')G cap, 2.5 µl RNasin (40 U/µl) and 2.5 µl T7 RNA Polymerase. The reaction was incubated at 37°C for at least 3 hrs. The DNA template was removed by addition of 5.6 µl of RQ1 DNase (1 U/µl) and further incubation at 37°C for 20 min. Finally, the transcribed RNA was purified via preparative polyacrylamide gel electrophoresis (2.2.3.5).

2.2.5 Protein biochemical standard methods

2.2.5.1 Concentration determination of proteins in solution

To determine the concentration of proteins, the BCA Protein Assay Kit was used according to the manufacturer's instructions. This assay is a detergent-compatible formulation based on bicinchoninic acid (BCA) for the colorimetric detection and quantification of total protein (Smith et al., 1985). First, Cu²⁺ from cupric sulfate is reduced in a temperature dependent reaction to Cu⁺ by the peptide bonds of a protein in an alkaline medium (Biuret Reaction). Thus, the amount of Cu²⁺ reduced is proportional to the amount of protein in the solution. The BCA chelates only the Cu⁺ ions, forming purple-colored complexes with an absorbance maximum at 562 nm. The amount of protein present in the solution can be quantified by comparing the absorption of the solution with dilutions of known concentrations of bovine serum albumine (BSA) treated under exactly the same conditions.

2.2.5.2 Precipitation of proteins

For precipitation of proteins from a buffered solution, 1/10 volume 3 M NaOAc (pH 5.2) and either at least 5 volumes of 100% ice-cold acetone (p.a.) or at least 3 volumes 100% ice-cold ethanol (p.a.) were added. To improve precipitation of small amounts of protein, up to 40 µg glycogen were added. Proteins were precipitated overnight at -20°C. Subsequently, the proteins were pelleted by centrifugation for 20 min at 13,000 rpm at 4°C in a microfuge. The pellet was washed twice with 70% (v/v) ice-cold ethanol and pelleted again for 10 min at 13,000 rpm and 4°C. The pellet was finally air-dried and resuspended in 2× SDS gel-loading buffer.

2.2.5.3 SDS polyacrylamide gel electrophoresis of proteins

Analytical electrophoresis of proteins was carried out in polyacrylamide gels under denaturing conditions (Laemmli, 1970) in a discontinuous buffer system as described in (Sambrook and Russell, 2001). The polyacrylamide gel is composed of a resolving gel of high pH and high acrylamide concentration at the bottom of the gel that is poured first. After polymerization of the resolving gel, a stacking gel of low pH and low acrylamide concentration is layered on top of the resolving gel. The stacking gel contained 5% acrylamide (Rotiphorese Gel 30), 125 mM Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 0.1% APS and 1 μ l/ml TEMED. The resolving gel contained 8-15% acrylamide (Rotiphorese Gel 30), 375 mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.1% APS and 0.4 μ l/ml TEMED. 6 \times SDS gel-loading buffer was added to the protein samples and the samples were denatured for 5 min at 96°C, before loading. The gel was run at 25 mA in protein electrophoresis buffer until the samples migrated into the resolving gel and was further run at 50-70 mA until the bromophenol blue of the gel-loading buffer reached the bottom of the gel. Proteins were visualized by Coomassie staining (2.2.5.4), silver staining (2.2.5.5), Sypro Ruby staining (2.2.5.6) or western blotting (2.2.7.2).

2.2.5.4 Coomassie Brilliant Blue staining of proteins in polyacrylamide gels

Coomassie Brilliant Blue (CBB) is an aminotriarylmethane dye that forms strong but not covalent complexes with proteins. The CBB G-250 staining solution was prepared by dissolving 80 mg CBB G-250 in 1000 ml ddH₂O and stirring for at least 2 hrs. Afterwards 3.5 ml 37% HCl was added and the solution was filled and stored in a UV resistant amber glass bottle. Polypeptides separated on a SDS polyacrylamide gel were first fixed with 40% (v/v) methanol/10% (v/v) acetic acid for at least 30 min while gently shaking. After rehydration of the gel in water, the gel was immersed for at least 30 min in at least 5 gel volumes CBB G-250 staining solution on a slowly rocking platform. For an optimal result, the CBB G-250 staining solution was pre-heated and then added to the gel. Destaining was performed by soaking the gel in water for at least 1 hr or until the blue background completely disappeared. To make a permanent record, the gels were scanned and then dried under vacuum for 1 hr at 80°C.

2.2.5.5 Silver staining of proteins in polyacrylamide gels

Silver staining of proteins was performed essentially as described (Blum et al., 1987). Polypeptides separated by SDS polyacrylamide gels were first fixed with 40% (v/v) methanol/10% (v/v) acetic acid for at least 30 min on a slowly rocking platform. The gels were incubated twice in 500 ml of 50% (v/v) ethanol and once in 30% (v/v) ethanol for 15 min each, followed by a brief wash in water. Gels were treated exactly 60 sec with 0.8 mM Na₂S₂O₃ and washed briefly with water.

Subsequently, gels were incubated in 250 ml of 12 mM AgNO₃ and 0.026% (v/v) formaldehyde for 20 min. Again, the gel was briefly washed with water and then incubated in 0.56 M Na₂CO₃, 0.0185% (v/v) formaldehyde and 16 μM Na₂S₂O₃ until the desired visualization of protein bands was achieved. The developing solution was removed and the reaction was finally stopped by addition of 250 ml of 5% (v/v) acetic acid. For a permanent record, the gels were scanned and then dried under vacuum for 1 hr at 80°C.

2.2.5.6 Sypro Ruby staining of proteins in polyacrylamide gels

Sypro Ruby fluorescence staining of proteins was performed according to the manufacturer's instructions (Invitrogen) by Dr. Dmitry E. Agafonov at the group of Cellular Biochemistry at the Max Planck Institute (MPI) for Biophysical Chemistry. Polypeptides separated on SDS polyacrylamide gels were first fixed with 50% (v/v) methanol/7% (v/v) acetic acid for at least 30 min on a slowly rocking platform. After rehydration of the gel in water, the gel was immersed in Sypro Ruby staining solution overnight and then washed with 10% (v/v) methanol/7% (v/v) acetic acid. Fluorescence was imaged with the Fujifilm FLA-7000 imager and quantified using the Fujifilm Image Gauge v4.22 program.

2.2.5.7 Semi-dry blotting of proteins from polyacrylamide gels

Proteins with a molecular weight below 100 kDa that were separated by polyacrylamide gel electrophoresis were usually transferred to a nitrocellulose membrane using the semi-dry blotting method. The apparatus for semi-dry blotting consists of two electrode plates (Anode on bottom, Cathode on top) separated by a horizontal sandwich of five layers of Whatman paper, nitrocellulose membrane, polyacrylamide gel and again five layers of Whatman paper. Whatman paper and nitrocellulose membrane are pre-soaked in 1× semi-dry blotting buffer. After assembling the sandwich, air bubbles trapped between the layers were removed by rolling a glass pipette over the sandwich. Blotting was performed by applying a constant current (1 mA per cm² gel size) for 60-120 min at room temperature. Afterwards, the transferred proteins were usually stained by soaking the blot for 5 min in Ponceau S solution, followed by destaining with water.

2.2.5.8 Wet blotting of proteins from polyacrylamide gels

Proteins with a molecular weight above 100 kDa that were separated by polyacrylamide gel electrophoresis were usually transferred to a nitrocellulose membrane using the wet blotting method. Wet blotting electrophoretic transfer devices use large vertical electrodes submerged in a tank of transfer buffer (1× SLAB4, 20% (v/v) methanol) in a plastic cradle surrounding the gel and

the membrane. As described for the semi-dry blot, gel and membrane are surrounded by 3-5 layers of Whatman paper saturated with transfer buffer in a sandwich and air bubbles trapped between the layers were removed by rolling a glass pipette over the sandwich. The sandwich was placed into the transfer device in such a way, that the membrane faces the Anode, whereas the gel faces the Cathode. The transfer was performed by applying a constant voltage of 70 V for 2 hrs at 4°C. Afterwards, the transferred proteins were usually stained by soaking the blot for 5 min in Ponceau S solution followed by a short destaining with water.

2.2.6 *In vitro* translation

Cell-free translation was performed with a translation system based on wheat germ lysate (2.2.6.3). Wheat germ lysate efficiently translates exogenously added mRNA that was produced by *in vitro* transcription as described in 2.2.6.2. The RNA template was generated from a pEU3b cell-free expression vector (Sawasaki et al., 2002) containing the desired cDNA. The vector contains the 5' leader (Omega) and 3' untranslated region of the tobacco mosaic virus, which were shown to result in highly efficient translation in the wheat germ translation system (Sawasaki et al., 2002).

2.2.6.1 Preparation of a pEU3-based vector for protein expression and purification

A pEU3b-based vector containing a C-terminal StrepII-purification tag (Voss and Skerra, 1997), preceded by a Factor Xa protease cleavage site (Nagai et al., 1985), was generated by conventional cloning techniques (2.2.2) and named pEU3b-StrepII. For this purpose, complementary oligonucleotides (MG1 and MG2) that contain both sequence elements were mixed (0.05 μ M each) and allowed to hybridize by heating for 1 min at 91°C and slowly cooling down to room temperature. The hybridized oligonucleotides were directly used for *Not* I/*Nco* I (both NEB) dependent ligation into the pEU3b vector. Subsequently, the cDNA of interest was inserted.

2.2.6.2 Preparation of template mRNA

The mRNAs for protein synthesis by the wheat germ cell-free translation system were synthesized by *in vitro* transcription of linearized pEU3b-derived vector carrying the cDNA of interest downstream of a bacteriophage T7 RNA polymerase promoter. Linearization of pEU3b-derived vectors was performed by restriction enzyme digestion as described in 2.2.2.2, usually with the enzymes *Apa*L I and *Pci* I (both NEB). When untagged protein was produced from a pEU3b-StrepII-derived vector, the plasmid was linearized with *Not* I, generating a template that ends directly after the coding sequence of the protein of interest. For increased protein production, a cap structure was introduced by incorporating the m⁷G-cap dinucleotide at the 5' end of the mRNA. The reaction

conditions were adapted from Pokrovskaya and Gurevich, 1994. A typical 100 μ l transcription reaction contained 20 μ l 5 \times T7 transcription buffer, ATP, CTP, and UTP (7.5 mM each), 0.6 mM GTP, 0.5 – 2 μ g of DNA template, 3 mM m⁷G(5')ppp(5')G cap, 2.5 μ l RNasin (40 U/ μ l) and 1.25 μ l T7 RNA polymerase. The reaction was carried out at 37°C for 4 hrs. The RNA was purified by gel filtration (2.2.3.6) and used for cell free translation.

2.2.6.3 *In vitro* translation of proteins in a wheat germ-based system

Cell-free protein synthesis using wheat germ extract was performed in a batch reaction for 4 hrs at 26°C. Buffers and wheat germ lysate were obtained from Zoegene (Japan) and the reaction conditions were adopted as recommended by the manufacturer, with minor modifications. For each mRNA template the optimum concentration was initially determined by titrating increasing concentrations of mRNA (0.01-0.8 mg/ml) into a small-scale translation reaction (e.g. 25 μ l). A typical 100 μ l reaction contained 20 μ l wheat germ lysate, 10.4 μ l buffer 3 (Zoegene, Japan), 9.6 μ l amino acid mix (Zoegene, Japan), 0.1% (w/v) Brij35, 0.4 μ g/ μ l creatine kinase, 0.64 U/ μ l RNasin, and 0.01-0.8 mg/ml mRNA template. For the synthesis of [¹⁴C]-radiolabeled proteins, an amino acid mix lacking L-leucine was used and the reaction was supplemented with 0.3 mM L-leucine and 6.5-13 μ M L-[U-¹⁴C]leucine (318 mCi/mmol). When proteins were labeled with L-[³⁵S]methionine, a 100 μ l translation reaction was carried out in the absence of unlabeled L-methionine (0.3 mM of the remaining L-amino acids) and in the presence of 3 μ l redivue L-[³⁵S]methionine (>1,000 Ci/mmol). Translation was monitored by liquid scintillation counting (2.2.6.4).

2.2.6.4 Liquid scintillation counting

Whatman 3MM paper was pre-soaked in 10% (v/v) TCA at room temperature and dried. For liquid scintillation counting, samples were applied onto the pre-soaked Whatman paper and then incubated for 10 min in 100 ml of boiling 10% (v/v) TCA, 10 min in 100 ml of 10% (v/v) TCA at room temperature and 10 min in 100 ml ice-cold acetone. After drying, the Whatman paper was transferred to plastic tubes containing 2.5 ml of liquid scintillant (Rotiszint EcoPlus) and radioactivity was monitored using the Liquid Scintillation Analyzer.

2.2.7 Immunological methods

2.2.7.1 Affinity purification of anti-peptide antibodies

Anti-peptide antibodies were affinity purified using a SulfoLink column containing immobilized peptide that was generated with a SulfoLink Kit. For this purpose, the peptide was first coupled to the column resin, followed by affinity purification of the antibody.

First, 2 ml of SulfoLink coupling gel were equilibrated 2× with 12 ml of SulfoLink coupling buffer [50 mM Tris, 5 mM EDTA (pH 8.0)] at room temperature. After dissolving 2 mg of sulfhydryl containing peptide in 2 ml of SulfoLink coupling buffer, the solution was filtered (0.2 μm) and added to the column. After incubating at room temperature for 15 min while rotating head over tail, the column was incubated for another 30 min without mixing. Subsequently, the buffer was drained off the column, and the column was washed with 6 ml of SulfoLink coupling buffer. In order to block non-specific binding sites on the SulfoLink coupling gel, 2 ml of 0.05 M L-cysteine/HCl in SulfoLink coupling buffer was added, followed by an incubation step for 15 min at room temperature while rotating head-over-tail and another 30 min incubation without mixing. After draining off the buffer, the column was washed 4× with 4 ml of SulfoLink wash buffer (1 M NaCl) and 3× with 4 ml of 0.05% (v/v) sodium azide (NaN₃).

Second, affinity purification of the antibody was performed as follows. The column containing the coupled peptide was equilibrated by washing with at least 6 ml of 1× PBS (pH 8.0). Washing steps were monitored with Bradford assay reagent according to the manufacturer's instructions. 2-8 ml of filtered (0.45 μm) serum [containing 1× PBS (pH 8.0)] were added per 1 ml of SulfoLink coupling gel resin with peptide. After incubating for 60 min at room temperature while rotating head-over-tail, the flowthrough was collected and tested for complete antibody removal by western blotting (2.2.7.2). Subsequently, the column was washed with 16 ml of 1× PBS (pH 8.0) (washing steps were again monitored by Bradford assay). Finally, antibodies were eluted by applying 8× stepwise 1 ml of 100 mM glycine buffer (pH 2.75). Collected fractions were immediately neutralized with 50 μl of 1 M Tris-HCl (pH 9.5). Antibody distribution in the fractions was monitored by Bradford assay. Peak fractions containing the affinity purified antibody were pooled and dialyzed overnight against 1× PBS (pH 8.0) at 4°C.

2.2.7.2 Western blotting

Immunoblotting (western blotting) was used to identify an antigen (protein) that reacts with a specific antibody. Proteins were first separated by electrophoresis through SDS polyacrylamide gels (2.2.5.3) and then transferred electrophoretically from the gel to a nitrocellulose membrane

(2.2.5.7, 2.2.5.8). The unreacted binding sites of the membrane were blocked by immersing the membrane for 1 hr in 1× TBS, 0.1% (w/v) Tween20, 5% (w/v) fat-free dried milk at room temperature to suppress nonspecific adsorption of antibodies. Next, the immobilized proteins were incubated with a specific antibody at an appropriate dilution (1:5,000-1:100) in 1× TBS, 0.1% (w/v) Tween20, 1% (w/v) fat-free dried milk for 1 hr at room temperature. After washing the membrane 3× for 10 min with 1× TBS, 0.1% (w/v) Tween20, 1% (w/v) fat-free dried milk at room temperature, the antigen-antibody complexes were incubated with horseradish peroxidase-conjugated secondary antibodies (dilution 1:50,000) in 1× TBS, 0.1% (w/v) Tween20, 1% (w/v) fat-free dried milk for 1hr at room temperature. The signal was detected by enhanced chemiluminescence using an ECL Advance Western Blotting Detection Kit according to the manufacturer's instructions.

2.2.7.3 Far western analysis

Far western analysis is a method very similar to western blotting that is used for the screening of protein-protein interactions between a radiolabeled protein and proteins present on nitrocellulose membrane. The radiolabeled protein was first synthesized by *in vitro* translation in the presence of L-[³⁵S]methionine (2.2.6.3). The protein was purified from unincorporated L-[³⁵S]methionine by gel filtration using Micro Bio-Spin 6 columns according to the manufacturer's instructions and the protein concentration was determined by scintillation counting. The proteins to be probed were isolated by affinity selection (2.2.10.3), separated by SDS-PAGE (2.2.5.3) and transferred to nitrocellulose membrane by wet blotting (2.2.5.8). Before probing, the proteins on the nitrocellulose membrane (which are denatured due to methanol in the blotting buffer) were renatured by washing (for 15 min at 4°C) the blot in NET₁₅₀ (or NET₃₀₀) containing 6 M guanidine-HCl, then in NET₁₅₀ (or NET₃₀₀) containing 2 M guanidine-HCl, and finally in NET₁₅₀ (or NET₃₀₀) containing 0.66 M guanidine-HCl. The nitrocellulose membrane was then blocked for 3 hrs with NET₁₅₀ (or NET₃₀₀) containing 5% fat-free dried milk, washed with NET₁₅₀ (or NET₃₀₀), and incubated overnight at 4°C with the probe protein in NET₁₅₀ (or NET₃₀₀). The blot was washed three times for 15 min with NET₁₅₀ (or NET₃₀₀) and bands were visualized by autoradiography with X-ray film. The same concentration of each protein was employed, if several proteins were used as probes in one experiment.

2.2.8 Cell culture

In this work, cell culture refers to cultivation of HeLa S3 cells under controlled conditions. To prevent bacterial contamination, all steps were carried out under a laminar flow cabinet and all materials used to manipulate cells were sterilized.

2.2.8.1 Cultivation of HeLa cells

HeLa S3 cells (Computer Cell Culture Centre, Belgium) were grown and maintained in DMEM supplemented with 10% (v/v) new born calf serum (NCS) at 37°C, 5% CO₂, and 95% relative humidity in an incubator. Cells were grown in suspension or adherent cultures. Adherent cells were cultivated in plastic tissue culture dishes and passaged when they reached semi-confluency. To adapt adherent cells for suspension growth, cells grown to semi-confluency in a 10 cm dish were used to inoculate 25 ml DMEM/10% (v/v) NCS. In order to transfer the cells to a bioreactor, the culture was passaged stepwise to a 1000 ml suspension with a cell density of 10⁶ cells/ml. The cell density was measured by flow cytometry using a Casy flow cytometer according to the manufacturer's instructions.

HeLa S3 cells were cultivated in suspension (2.5-5 × 10⁶ cells/ml medium) at a logarithmic growth rate using a bioreactor, essentially as described by Kastner, 1998, with the help of Thomas Conrad. For cell cultivation, DMEM medium supplemented with 5% (v/v) NCS, 50 µg/ml penicillin and 100 µg/ml streptomycin was used. For the cultivation of HeLa stable cell lines, the media was additionally supplemented with 1 mg/ml G418. Harvesting of HeLa cells was performed according to Dignam et al., 1983.

2.2.8.2 Freezing of HeLa cells

HeLa S3 cells were frozen in cryo-media containing DMEM supplemented with 10% DMSO and 20% NCS. Semi-confluent cells from a 10 cm dish were washed with 1 ml of 1 × PBS (pH 7.4) and trypsinized with 1 ml Trypsin-EDTA. The reaction was stopped by the addition of 3 ml DMEM/10% (v/v) NCS. Alternatively, 100 ml of cells were directly taken from a suspension culture (~5 × 10⁵ cells/ml). The cells were carefully pelleted by centrifugation for 5 min at 1200 rpm and 4°C in a Megafuge 1.0R and washed once with 10 ml 1 × PBS (pH 7.4). The pellet was dissolved in 1 ml (10 cm dish) or 5 ml (100 ml suspension culture) cryo-media and filled into 1 ml or 5 ml cryo tubes, respectively. The cells were slowly frozen to -80°C using a cryo freezing container filled with isopropanol that provides the critical -1°C/minute cooling rate required for successful cell cryopreservation and recovery. For long term storage, cells were transferred to a liquid nitrogen freezer. For thawing, cells were heated to 37°C in a water bath until they were no longer frozen.

Subsequently, the cells were transferred to a 15 ml plastic tube and the volume was adjusted to 10 ml with DMEM/10% (v/v) NCS that had been heated to 37°C. The cells were spun down for 5 min at 1,200 rpm and 4°C and the pellet was dissolved in 8 ml DMEM/10% (v/v) NCS and either transferred to a 10 cm dish (1 ml aliquot) or directly used to inoculate a 100 ml suspension culture (5 ml aliquot).

2.2.8.3 Establishing HeLa stable cell lines

HeLa S3 stable cell lines were established that constitutively express recombinant AD002 or SPF27 proteins. The genes encoding AD002 and SPF27 were cloned into the pIRESneo-FLAG/HA vector that introduces an N-terminal FLAG/HA double affinity tag to the protein of interest. Furthermore, the genes encoding AD002 and an eGFP-AD002 fusion protein were cloned into the pcDNA3.0-N/C-SF-TAP vectors that introduce either an N- or C-terminal FLAG/StrepII-tag to the protein of interest. These vectors encode ampicillin resistance for selection in bacteria and Neomycin resistance (G418) for selection in mammalian cells. HeLa S3 cells were pipetted 24 hrs before transfection into 6 well plates to a density of 5×10^5 cells per well in 2.5 ml DMEM/10% (v/v) NCS and incubated overnight at 37°C (5% CO₂, 95% humidity). The following day, a transfection pre-mix was prepared containing 50 µl Optimem and 0.5 µg, 1 µg, or 2 µg of the expression vector. The pre-mix was thoroughly mixed with a second pre-mix comprised of 700 µl Optimem and 14 µl of the transfection reagent, and incubated for 15 min at room temperature. Next, the culture media was replaced with 2.5 ml Optimem and the transfection mix was added and mixed well. Cells were then incubated for 2 hrs at 37°C. After incubation, the transfection solution was replaced with 2.5 ml DMEM supplemented with 10% NCS and incubated overnight at 37°C. The next day, cells were washed with 1 ml of 1× PBS (pH 7.4) and trypsinized by incubating 5 min with 1 ml of Trypsin-EDTA. The reaction was stopped by adding 3 ml of DMEM supplemented with 10% NCS, and different amounts (0.5 ml, 1 ml, 2 ml) of the cells were pipetted into three 10 cm dishes. The volume was adjusted to 8 ml with DMEM/10% NCS and the cells were incubated for 6 hrs at 37°C. To select for stable cell lines, the standard media was replaced by 8 ml DMEM/10% NCS, containing 100 U/ml Pencillin/Streptomycin and 1.5 mg/ml G418 and cells were incubated at 37°C. Within 1 to 2 weeks, those cells that stably incorporated the plasmid enabling Neomycin resistance were selected, generating colonies of stable cells. During this time, the selection medium was exchanged every day. Individual colonies were then pipetted into a sterile 200 µl pipette tip under an inverted light microscope, and transferred to 24-well dishes containing 1 ml selection media. The cells were grown until semi-confluency and then split into 6 well plates. At this time, 10^5 cells were taken from each well and boiled in 3× SDS loading buffer for 5 min at 96°C. The whole cell extract was separated on a SDS polyacrylamide gel (2.2.5.3) and analyzed by western blotting for the presence

of the recombinant protein (2.2.7.2). The clones that stably expressed the protein of interest were passaged from the 6-well plate into at least three 10 cm dishes, grown to confluency and frozen as described (2.2.8.2).

2.2.8.4 Fluorescence microscopy

Fluorescence microscopy was performed to visualize the distribution of the eGFP-tagged AD002 and the SC35 protein (via immunofluorescence) in HeLa cells. First, cover slips were placed in the wells of a 24 well plate and sterilized under UV light. Approximately 10^5 HeLa S3 cells (expressing eGFP-AD002-C-SF-TAP) were added to each sterilized well in 0.5 ml of DMEM, containing 10% (v/v) NCS, 100 U/ml Pencillin/Streptomycin and 1.5 mg/ml G418. Cells were grown overnight in an incubator at 37°C to allow the cells to settle on the cover slip. All further steps were performed at room temperature. The following day, medium was removed and each well was washed twice with 0.5 ml of 1× PBS (pH 7.4). Next, 300 µl of 3% PFA in PBS was added and incubated for 10 min. Cells were washed twice with PBS and subsequently permeabilized by incubating for 10 min with 0.5 ml of 0.2% Triton-X-100 in PBS. After permeabilization, cells were washed once with PBS and then blocked for 1 hr with 0.5 ml of 2% BSA in PBS. The blocking solution was removed and the cells were incubated for 1 hr with anti-SC35 antibody (dilution 1:1,000) in 0.5 ml of 2% BSA in PBS. Next, cells were washed 3× with PBS and then incubated for 1 hr with anti-mouse-Cy5 conjugated antibody (1:15,000) in 0.5 ml of 2% BSA in PBS. Finally, cells were washed 3× with PBS and fixed with the coverslip on a microscope slide by adding one drop of Mowiol and drying for 5-6 hrs. The endogenous eGFP-fluorescence and the immunofluorescence of the Cy5-conjugated antibody were analyzed by confocal fluorescence microscopy.

2.2.8.5 Preparation of HeLa nuclear extract

Nuclear extracts were prepared from HeLa cells essentially as described by Dignam et al., 1983. In particular, six to eight liters of HeLa cells were grown in suspension culture in DMEM media supplemented with 5% (v/v) NCS, 50 µg/ml penicillin, and 100 µg/ml streptomycin to a density of $2.5-5 \times 10^5$ cells/ml (Kastner, 1998). Cells were harvested by centrifugation for 10 min at 2,000 rpm in a Cryofuge 6000i centrifuge and washed 3 times by resuspending in ice-cold 1× PBS (pH 7.4) and centrifuging for 10 min at 4°C and 2,500 rpm in a Megafuge 1.0R centrifuge. The washed cells were resuspended in 1.25 volumes of 1× MC buffer and allowed to swell for 5 minutes on ice. Cells were lysed with 18 strokes of a Dounce homogenizer at 4°C. Nuclei were pelleted in Corex tubes at $13,000 \times g$ in a Sorvall SS34 rotor for 5 min at 4°C (for C complex preparation, alternatively at $18,000 \times g$). After removing the supernatant, the pellet was resuspended in 1.3 volumes of 1×

Roeder C buffer and the nuclei were disrupted with 20 strokes in a Dounce homogenizer. The lysed nuclei were stirred for 40 min at 4°C and then transferred to ultracentrifuge tubes. N-heptane was added on top, followed by centrifugation for 30 min at 40,000 rpm and 4°C using a Beckman 45Ti rotor (for C complex preparation, alternatively for 30 min at 16,000 rpm and 4°C in a Sorvall SS34 rotor). The supernatant was recovered and dialyzed twice for 2.5 hrs against 50–100 volumes of 1× Roeder D buffer. The dialyzed nuclear extract was cleared by centrifugation at 10,000 rpm at 4°C in an HB-6 rotor. The precipitate was discarded and the HeLa nuclear extract was frozen in aliquots in liquid nitrogen, stored at -80°C and tested for splicing activity (2.2.9).

2.2.9 *In vitro* splicing reaction

A typical splicing reaction was performed in solution containing 5-20 nM pre-mRNA substrate and 40% (v/v) HeLa nuclear extract in 1× Roeder D buffer (2.2.8.5), and was supplemented with 25 mM KCl, 3 mM MgCl₂, 2 mM ATP and 20 mM creatine phosphate. The reaction was incubated at 30°C for various times. After incubation, samples were cooled down on ice and the RNA was recovered by proteinase K digestion (2.2.3.7) and analyzed by denaturing polyacrylamide gel electrophoresis (2.2.3.1).

2.2.10 Affinity-based purification methods

2.2.10.1 Purification of the MS2-MBP fusion protein

MS2-MBP is a bifunctional fusion protein of the bacteriophage MS2 coat protein that binds MS2 RNA aptamers (2.2.4), and the maltose binding protein (MBP) that binds maltose and amylose with high efficiency. MS2-MBP was used for the purification of spliceosomal C complexes (2.2.10.5). The expression plasmid pMS2-MBP was kindly provided by Prof. Robin Reed (Das et al., 2000). The plasmid was transformed into *E. coli* BL21-rosetta cells by electroporation (2.2.2.5) and plated onto LB-Agar under selective conditions (75 µg/ml Ampicillin, 25 µg/ml chloramphenicol). A single colony was used to inoculate 5 ml of LB medium (containing Ampicillin and chloramphenicol), which were incubated overnight at 37°C while rapidly shaking at 250 rpm. On the next day, the overnight culture was used to inoculate 1 liter of LB medium, and this culture was incubated at 37°C with 250 rpm agitation. When the culture reached an OD₆₀₀ of 0.5, the MS2-MBP expression was induced by addition of 400 µl 250 mM IPTG solution and the cells were allowed to grow for a further 3 hrs under the same conditions. Then the cells were pelleted by centrifugation (10 min at 5,000 rpm and 4°C in a Sorvall SLA-1500 rotor) and resuspended in 40 ml lysis buffer [20 mM Tris/HCl (pH 7.6), 200 mM NaCl] and 1 tablet of complete protease inhibitor cocktail. Subsequently,

cells were lysed by sonication with 8 pulses of 10 s duration (output: 5, duty: 30%). The cell lysate was cleared by centrifugation (10 min at 4,000 rpm and 4°C in a Megafuge 1.0R). The cleared supernatant was taken and incubated with 1.5 ml amylose resin (bed volume) for 2 hrs at 4°C while rotating the tube. After incubation, the amylose resin was transferred into a chromatography column and drained by gravity flow. Subsequently, the matrix was washed with 10 ml MBP-150 buffer [20 mM Hepes-KOH (pH 7.9), 150 mM NaCl, 0.05 % (v/v) NP-40] and 1 ml 5 mM phosphate buffer (pH 7.9). The bound MS2-MBP was eluted with 1.5 ml of 5 mM phosphate buffer (pH 7.9) containing 15 mM maltose. To remove nucleic acids from the eluate, it was applied to 700 µl of Heparin-Sepharose in a chromatography column. The flow-through was collected and applied again to the same column to increase the recovery. The Heparin-Sepharose was washed with 5 ml of 5 mM phosphate buffer (pH 7.9) and finally eluted stepwise with 0.3 ml and 1.7 ml elution buffer [20 mM Hepes-KOH (pH 7.9), 100 mM KCl, 15 % (v/v) glycerol, 0.5 mM DTT, 0.2 mM PMSF]. Purified MS2-MBP was analyzed by SDS polyacrylamide gel electrophoresis (2.2.5.3) and the protein concentration determined as described (2.2.5.1).

2.2.10.2 Pulldown analysis of protein-protein interactions

Protein-protein interactions were analyzed by performing StrepII fusion protein pulldowns. By analogy to the well-described GST fusion protein pulldown technique (Kaelin et al., 1991), this technique uses the affinity of the StrepII-tag for StrepTactin-coupled Sepharose to screen for interacting proteins in solution. First, the StrepII-tagged protein was expressed in the wheat germ cell-free translation system (2.2.6) and mixed with the same amount (e.g. 40 µl) of a [¹⁴C]leucine-labeled target protein that was also produced by *in vitro* translation. The mixture was incubated for at least 30 min on ice. Subsequently, 200 µl of IPP150 buffer and 15 µl (bed volume) StrepTactin-Sepharose were added and the mixture was incubated for 90 min at 4°C with head-over-tail rotation. In a negative control, the StrepII-tagged protein was omitted from the incubation, but otherwise identical conditions were employed. Afterwards, the beads were pelleted by centrifugation for 1 min at 1,000 rpm and 4°C in a microcentrifuge. The supernatant was withdrawn and the beads were washed 5 times with 500 µl of IPP150 buffer. Finally, the beads were dried under vacuum and dissolved in 40 µl of 3× SDS gel-loading buffer. After incubation for 5 min at 96°C, the supernatant was analyzed by SDS polyacrylamide gel electrophoresis (2.2.5.3) and the gels were stained with Coomassie (2.2.5.4) and dried under vacuum for 1 hr at 80°C. The co-precipitated, radiolabeled target proteins were visualized by autoradiography.

2.2.10.3 Purification of the hPrp19/CDC5L complex from HeLa nuclear extract

Human Prp19/CDC5L complexes were affinity-purified from nuclear extract prepared from HeLa cells stably expressing FLAG/HA-AD002 and FLAG/HA-SPF27. All steps were carried out at 4°C. For the preparative isolation of the hPrp19/CDC5L complex, 250 ml nuclear extract were mixed with an equal volume of G250 buffer containing 0.5 mM DTE and 0.5 mM PMSF, and incubated overnight. Next, the extract was cleared by centrifugation for 30 min at 40,000 rpm in a 45Ti rotor. The supernatant was loaded onto 5 ml of anti-FLAG-M2-Agarose beads using a peristaltic pump with a flow rate of 25 ml/hr. Subsequently, the beads were washed with 100 ml G250 buffer containing 2.5% (v/v) glycerol, 0.5 mM DTE, and 0.5 mM PMSF, at a flow rate of 25 ml/hr. The bound material was eluted with 20 ml of G250 buffer containing 2.5% (v/v) glycerol, 1 mM DTT, and 0.2 mg/ml FLAG peptide at a flow rate of 3 ml/hr and 1 ml fractions were collected. The absorption of each elution fraction was measured at 280 nm and further analyzed by SDS polyacrylamide gel electrophoresis (2.2.5.3) and Coomassie staining (2.2.5.4). The peak fractions containing the affinity-purified hPrp19/CDC5L complexes were pooled and concentrated to approximately 500 µl volume with an Amicon Ultra-15 centrifugal filter device (MWCO 100 kDa), by centrifuging at 4,000 rpm in a Megafuge 1.0R. The concentrated material was further purified by size exclusion chromatography on a Superose 6 GL 10/300 column using the Äkta purifier or explorer system according to the manufacturer's instructions. Gel filtration was performed with SEC100 buffer at a flow rate of 0.3 ml/min and 0.5 ml fractions were collected. The absorption of each fraction was measured at 254 nm and 280 nm with a photometer and recorded. The peak fractions were pooled and either directly used or concentrated by ultracentrifugation in a S100AT rotor for 13 hrs at 60,000 rpm and 4°C to a final concentration of up to 10 mg/ml. If not used directly, particles were flash frozen in liquid nitrogen and stored at -80°C. The FLAG-M2-Agarose column was regenerated by washing with 50 ml of G250 buffer, followed by 15 ml of 100 mM L-glycine (pH 3.5) and again 50 ml of G250 buffer containing 2.5% (v/v) glycerol and 0.02% (v/v) NaN₃.

For analytical purposes, size exclusion chromatography was performed on a SMART HPLC/FPLC system using a Superose 6 PC 3.2/30 gel filtration column. For a typical run, 50 µl of the sample were loaded on the column at a flow rate of 30 µl/min. 40 µl fractions were collected during the run and protein samples were analyzed by SDS polyacrylamide gel electrophoresis (2.2.5.3), with subsequent Coomassie staining (2.2.5.4).

Alternatively, affinity-purified hPrp19/CDC5L complexes were separated on a 4 ml linear 5-20% (v/v) glycerol gradient containing G150 buffer by centrifugation in a TH660 rotor (Sorvall) for 13-16 hrs at 40,000 rpm and 4°C. Gradients were fractionated by hand from top to bottom into 175 µl aliquots and the proteins from each fraction were isolated and analyzed by SDS polyacrylamide gel

electrophoresis (2.2.5.3) and Coomassie staining (2.2.5.4). For EM analysis of the hPrp19/CDC5L complex, the particles were centrifuged under GraFix conditions (Kastner et al., 2008). Therefore, the gradients contained 0-0.1% (v/v) glutaraldehyde, and a 200 μ l cushion containing 2.5% (v/v) glycerol in G150 buffer (but no glutaraldehyde) was loaded onto the gradient prior to loading the sample. After centrifugation, the gradients were fractionated from bottom to top using a peristaltic pump. Fractions were collected and the absorption profile of the gradient was measured at 254 nm and 280 nm with a photometer. The peak fractions containing the purified, fixed complexes were used for negative stain EM.

2.2.10.4 Purification of the 35S U5 snRNP from HeLa nuclear extract

For the preparative isolation of the 35S U5 snRNP, a 12 ml splicing reaction containing 50% nuclear extract prepared from the HeLa FLAG/HA-AD002 stable cell line was prepared as described in 2.2.9, but without pre-mRNA, and incubated at 30°C for 10 min. Heparin was then added to a final concentration of 0.5 mg/ml and the incubation at 30°C was continued for an additional 5 min. All further procedures were carried out at 4°C. The splicing reaction was incubated for at least 2 hrs with head-over-tail rotation with 150 μ l of anti-FLAG-M2-Agarose. Subsequently, the beads were washed 3 times with 10 ml of G150 buffer and transferred to a 1.5 ml microfuge tube and washed an additional 5 times with 1 ml G150 buffer. The bound material was eluted in two consecutive elution steps by incubating each time for at least 1 hr with 200 μ l of G150 buffer containing 0.2 mg/ml of FLAG peptide, while rotating head-over-tail. The eluate was loaded onto a 4 ml linear 10-30% glycerol gradient containing G150 buffer and centrifuged in a TH660 rotor for 14 hrs at 25,000 rpm. Gradients were fractionated by hand from top to bottom into 175 μ l aliquots. RNA and protein were isolated from each fraction by PCI extraction (2.2.1.3). The upper aqueous phase containing the RNA was precipitated with ethanol (2.2.1.4) and analyzed by denaturing polyacrylamide gel electrophoresis (2.2.3.1) and stained with silver (2.2.3.2) or analyzed by northern blotting (2.2.3.4). Proteins in the phenol phase plus interphase were precipitated with acetone and separated on a SDS polyacrylamide gel (2.2.5.3) and stained with silver (2.2.5.5) or Coomassie (2.2.5.4). For EM analysis of the 35S U5 snRNP, the particles were centrifuged under GraFix conditions (Kastner et al., 2008). Therefore, the gradients contained 0-0.1% (v/v) glutaraldehyde, and a 200 μ l cushion containing 5% (v/v) glycerol in G150 buffer (but no glutaraldehyde) was loaded onto the gradient before addition of the sample. After centrifugation, the gradients were fractionated from bottom to top using a peristaltic pump. Fractions were collected and the absorption profile of the gradient was measured at 254 nm and 280 nm with a photometer. The peak fractions containing the purified, fixed complexes were analyzed by negative stain EM.

2.2.10.5 MS2 affinity selection of spliceosomal C complexes

MS2 affinity selection of spliceosomal C complexes was carried out as described by Bessonov et al., 2008. In particular, PM5 pre-mRNA (2.2.4) was incubated with a 20-fold molar excess of purified MS2-MBP fusion protein (2.2.10.1) at 4°C for 30 min in 20 mM HEPES-KOH (pH 7.9), 60 mM KCl. To affinity purify preparative amounts of the C complex, a 12 ml standard splicing mixture (2.2.9) containing 15 nM of [³²P]-labeled PM5 pre-mRNA (specific activity 8,000 cpm/pmol) was carried out. C complexes were assembled by incubating at 30°C for 2.5 hrs. The reaction was incubated for an additional 30 min at 30°C in the presence of the m6 and m12 DNA oligonucleotides (final concentration 1 μM). After incubation, the splicing reaction was cooled down on ice and then 2 ml aliquots of it were loaded onto six 14 ml, linear 10-30% (v/v) glycerol gradients in 1× G150 buffer. Gradients were centrifuged for 16 hrs at 25000 rpm in a Sorvall TST 41.14 rotor and harvested manually in 500 μl fractions from the top. Radioactivity in each fraction was measured in a scintillation counter and the peak fractions containing the C complexes were pooled, divided into two fractions and loaded (by gravity flow) onto two separate columns containing 250 μl of amylose resin pre-equilibrated with 1× G150 buffer. The matrix was washed 6× with 1 ml of 1× G150 buffer each. Spliceosomal complexes were eluted dropwise with 900 μl of elution buffer (1× G150 buffer containing 12 mM maltose). Proteins and RNAs were recovered and analyzed by polyacrylamide gel electrophoresis (2.2.5.3, 2.2.3.1).

2.2.11 Special Methods

2.2.11.1 [¹⁴C]Iodoacetamide labeling of hPrp19/CDC5L complex proteins

Iodoacetamide modification of human Prp19/CDC5L complexes was performed in collaboration with Dr. Dmitry E. Agafonov at the group of Cellular Biochemistry at the MPI for Biophysical Chemistry. For iodoacetamide modification, 50 μg of the affinity-purified, gel-filtrated hPrp19/CDC5L complex were initially treated with 8M Urea and 10 mM β-mercaptoethanol for 90 min at 37°C. Subsequently, the pH was increased to 9.0 with 100 mM Tris-HCl buffer and [¹⁴C]iodoacetamide (56 mCi/mmol, GE Healthcare) was added to a concentration of 50 mM. The reaction was incubated for 6 min at 37°C and subsequently stopped by adding 0.5 M β-mercaptoethanol. Subsequently, 6× SDS gel loading buffer was added and the polypeptides were separated by SDS polyacrylamide gel electrophoresis (2.2.5.3). The radioactivity of the proteins in the SDS gel was monitored by autoradiography and quantified using the Fuji Image Gauge v4.22 program.

2.2.11.2 Dissociation of native hPrp19/CDC5L complexes

Dissociation of native hPrp19/CDC5L complexes was performed to identify stable subcomplexes. Therefore, 25 μg of the purified, gel filtrated hPrp19/CDC5L complexes (2.2.10.3) were incubated for 30 min on ice in the presence of 150 mM-1200 mM NaCl, or 150 mM NaCl and 0.1 mg/ml heparin, or 150 mM NaCl and 0.4M NaSCN. Subsequently, the treated complexes were loaded onto of a 4 ml linear 5-20% (v/v) glycerol gradient containing 20 mM HEPES-KOH (pH 7.9), 1.5 mM MgCl_2 and a corresponding concentration of NaCl (150 mM-1,200 mM) or 150 mM NaCl and 0.1 mg/ml heparin, or 150 mM NaCl and 0.4 M NaSCN, and centrifuged in a Sorvall TH660 rotor for 16 hrs at 40,000 rpm and 4°C. The gradients were harvested from the top by hand in 175 μl fractions and proteins present in each gradient fraction were precipitated with ethanol (2.2.5.2), analyzed by SDS gel electrophoresis (2.2.5.3), and stained with Coomassie (2.2.5.4).

2.2.11.3 Sulfo-MBS crosslinking in native hPrp19/CDC5L complexes

Sulfo-MBS crosslinking was used to identify protein-protein interaction partners in the native hPrp19/CDC5L complex. The complexes were purified as described in 2.2.10.3. Initially, the buffer of 50 μg of the purified hPrp19/CDC5L complex was exchanged to 1 \times PBS (pH 7.2) using a pre-equilibrated Bio-Gel P6 Column. After buffer exchange, the complex was divided into three equal aliquots and incubated for 30 min at 20°C with 0 μM , 1 μM , or 10 μM of the heterobifunctional crosslinking agent Sulfo-MBS in a final volume of 100 μl of 1 \times PBS (pH 7.2). The reaction was stopped by the addition of 12 μl of 0.5 M glycylamide-hydrochloride and 8 μl of 150 mM DTT, and then incubated for 30 min at 37°C. Next, proteins were precipitated with acetone (2.2.5.2), separated by SDS polyacrylamide gel electrophoresis (2.2.5.3), and transferred to a nitrocellulose membrane (2.2.7.2). The membrane for each condition (0 μM , 1 μM , and 10 μM Sulfo-MBS) was cut into stripes of identical size and probed separately with antibodies against each hPrp19/CDC5L complex protein (2.2.7.2).

2.2.11.4 Limited proteolysis of native hPrp19/CDC5L complexes

The structure of the hPrp19/CDC5L complex was probed by limited proteolysis. Furthermore, limited proteolysis was performed to prepare stable hPrp19/CDC5L subcomplexes for X-ray crystallography. In an initial screen, 15 μg of affinity-purified, gel filtrated complexes (2.2.10.3) were treated with 4 μl of three dilutions (1/10, 1/100, 1/1000) of the protease Subtilisin (0.5 mg/ml) or Chymotrypsin (0.5 mg/ml) in SEC100 buffer. The reactions were incubated for 30 min on ice and stopped by the addition of 1 μl of 100 mM PMSF (in isopropanol) and 5 μl of 6 \times SDS

loading buffer, and then incubated for 5 min at 96°C. The samples were separated on a pre-casted 4-12% Novex Bis-Tris polyacrylamide gel and stained with Coomassie (2.2.5.4).

Semi-preparative, limited proteolysis was performed with Subtilisin by incubating 150 µg of the hPrp19/CDC5L complex separately with 0.02 µg, 0.05 µg, 0.1 µg, 0.2 µg, or 0.4 µg Subtilisin in a reaction volume of 60 µl on ice. The reaction was stopped by the addition of 1 µl of 100 mM PMSF (in isopropanol) and further analyzed by size exclusion chromatography on a SMART HPLC/FPLC system using a Superose 6 PC 3.2/30 gel filtration column (GE Healthcare) as described (2.2.10.3). Gel filtration fractions were pooled pairwise and the proteins were precipitated with acetone (2.2.5.2). The pellet was dissolved in 30 µl of 3× SDS gel loading buffer and the proteolyzed components were separated by SDS polyacrylamide gel electrophoresis (2.2.5.3), stained with Coomassie (2.2.5.4) and analyzed by MS (2.2.11.5).

For preparative isolation of the protease resistant subcomplexes, 2.5 mg of affinity-purified, gel filtrated hPrp19/CDC5L complex were treated with 1.67 µg Subtilisin in a final volume of 167 µl for 30 min on ice. The reaction was stopped by the addition of 6 µl of 0.1 M PMSF (in isopropanol). The proteolyzed components were separated by size exclusion chromatography on a Superose 6 GL 10/300 GL column as described (2.2.10.3). The peak fractions were pooled and concentrated by ultracentrifugation in a S100AT rotor for 13 hrs at 60,000 rpm and 4°C. A final concentration of up to 10 mg/ml was obtained and concentrated samples were used for X-ray crystallography.

2.2.11.5 Mass spectrometry

The technological basis for most proteomic studies is mass spectrometry (MS) and the development of electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) techniques has been key instrumental for the direct analysis of complex protein mixtures (reviewed in Aebersold and Mann, 2003). MS analysis of a large biological sample usually starts with a pre-separation of the protein components via SDS polyacrylamide gel electrophoresis (2.2.5.3). In a consecutive step either Coomassie or silver stained protein bands are excised from the gel or the whole gel lane is cut into pieces. MS for protein identification relies on the digestion of the protein samples into peptides by a sequence-specific protease, such as trypsin. Peptides are much more accessible to MS analysis and can be easily eluted from the gel. For analysis, the extracted peptides are often separated directly out of the solution on a C18 reversed phase chromatography column and subsequently delivered to the mass spectrometer via integrated liquid-chromatography coupled to ESI (LC online ESI-MS). Alternatively, MALDI sublimates and ionizes the samples out of a dry, crystalline matrix via laser pulses after liquid-chromatographic separation (LC offline MALDI-MS). ESI generates multiply charged molecules and ESI-MS systems

are therefore preferred for the analysis of more complex samples, whereas MALDI-MS is normally suited to analyze relatively simple peptide mixtures. After ionization, the charged peptides enter the mass analyzer, where they are separated according to their mass-to-charge (m/z) ratio. There exist many types of mass analyzers, using either static or dynamic fields, or magnetic or electric fields. For example, a time-of-flight (TOF) analyzer accelerates the ionized peptides in an electric field and measures the time they take to reach the detector. This time is directly proportional to the molecular weight of the peptides and lighter peptides will reach the detector first. Quadrupole mass analyzers select ions by oscillating electric fields, which permit a stable trajectory only for ions of a particular m/z at a time (ion trap). A mass spectrometer can be instructed to first mass measure all the peptides and then select a number of peptide ions in turn for fragmentation in a collision cell. In a collision cell the peptide ions are accelerated by electrical potential to high kinetic energy and allowed to collide with inert gas molecules in order to break peptide bonds. The mass of the resulting fragments is again measured in a mass analyzer and the resulting spectrum is called a tandem or MS/MS spectrum and contains the sequence information of the fragmented peptide that can be used to identify the peptide in a protein database.

MALDI is usually coupled to TOF analyzers that measure the mass of intact peptides. However, the combination of two TOF analyzers, interconnected by a collision cell also permits peptide sequencing. The resulting fragment ion spectra are often more extensive and informative than those generated in a trapping instrument. Furthermore, ionization from a crystalline matrix allows a sample to be measured again at a later time, thereby providing very high peptide coverage throughout the whole protein sequence. Therefore, this technique is particularly suitable for sequencing a protein sample, e.g. after protease digestion.

MS was used to identify the protein composition of the purified human Prp19/CDC5L complex and 35S U5 snRNP. Furthermore, LC offline MALDI tandem MS was performed to identify those regions of the proteins of the Prp19/CDC5L complex present after protease digestion (2.2.11.4). All MS analyses described in this work were performed in collaboration with the Bioanalytical Mass Spectrometry Group at the MPI for Biophysical Chemistry headed by Dr. Henning Urlaub. In particular, quantitative analysis of the hPrp19/CDC5L complex was performed by Carla Schmidt. For MS analysis, proteins were separated by SDS gel electrophoresis (2.2.5.3) and either Coomassie or silver stained protein bands were excised from the gel, or alternatively the whole gel lane was cut into equal pieces. Subsequently, proteins were digested into peptides by the sequence-specific protease trypsin, as described by Shevchenko et al., 1996. For analysis, the extracted peptides (in solution) were separated on a C18 reversed phase chromatography column and subsequently delivered to the mass spectrometer via LC online ESI-MS/MS, using either a hybrid Linear Ion Trap-Orbitrap mass spectrometer (LTQ-Orbitrap XL) or a high resolution Quadrupole-ToF analyzer

(Micromass qToF II Ultima). Alternatively, dried droplets were prepared from LC peptide fractions as described by Ons, Proteomics, 2009 and analyzed by LC offline MALDI-MS/MS with a tandem time of flight mass analyzer (4800 MALDI ToF/ToF Analyzer, Applied Biosystems/MDS Sciex). Proteins were identified by searching fragment spectra of sequenced peptides against the NCBI database restricted to human or a human RefSeq database using a MASCOT database search (Perkins et al., 1999).

2.2.11.6 Electron microscopy

Electron microscopy (EM) and single particle reconstruction are the methods of choice for visualizing the architecture of entire splicing complexes and their snRNP components (reviewed in Stark and Lührmann, 2006). In recent years, much technical progress has been made in EM instrumentation, computer hardware and new image analysis software. However, sample preparation is still a critical step in the structural analysis of macromolecular complexes. Reliable and high-resolution structure determination of a macromolecular complex by EM requires a structurally and compositionally homogeneous sample. Notably, recent development of the GraFix procedure has considerably improved sample quality for negative stain, cryo-negative stain and, particularly, unstained cryo-EM (Kastner et al., 2008). As schematically shown in Fig. 2.1, during GraFix a macromolecular complex is centrifuged into an increasing concentration of a chemical fixation reagent (e.g. glutaraldehyde) in a glycerol gradient to prevent aggregation and to stabilize individual macromolecules.

As most biological material shows little contrast relative to their buffered solution, a first insight into the higher-order structure of a macromolecular complex is often achieved by negatively staining the specimen. Therefore, the complexes are initially allowed to adsorb from a liquid solution to a thin carbon film. The immobilized complexes are transferred to a copper grid covered with a perforated carbon support film. Subsequently, the grid is transferred to the staining solution, usually a heavy metal salt of high atomic number (e.g. uranyl formate) that serves to scatter the electrons from those regions covered with the stain. In the negative staining procedure, the stain surrounds the complex, but is excluded from the volume occupied by the sample. Thus, during transmission EM (TEM), electron beams that are passing through the complex are less scattered than those passing through stain-rich regions (Reimer, 1997). Electrons that are highly scattered by the stain are then filtered out by the objective aperture located below the sample and only electrons that pass through the complex are involved in the formation of the final image.

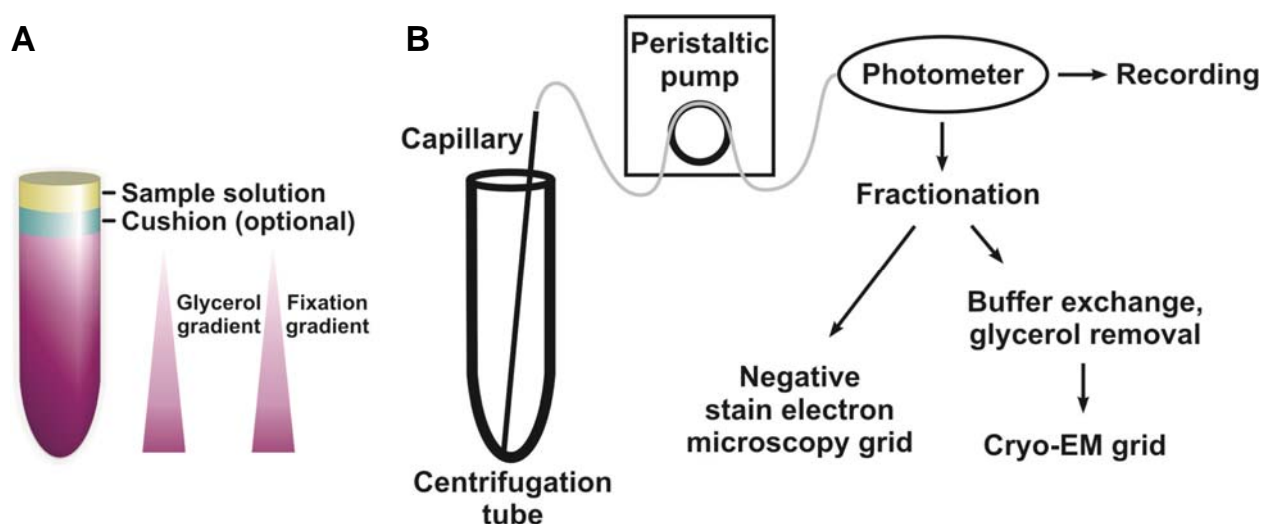


Fig. 2.1: Illustration of the GraFix procedure. (A) Preparation of the GraFix gradient. The fixation reagent is added only to the denser glycerol solution. Thus, during gradient formation, a fixation gradient is formed along with the density gradient. The sample containing the purified macromolecular complex is added on top of the gradient. To prevent premature fixation of the complexes, a cushion that does not contain the fixation reagent may be optionally added between gradient and sample. (B) GraFix sample preparation. Gradients are continuously fractionated from bottom to top using a peristaltic pump. Fractions are collected and the absorption profile of the gradient is measured at 254 nm and 280 nm with a photometer and recorded. The peak fractions containing the purified, fixed complexes are used either directly for negative stain EM or are purified away from glycerol for unstained cryo-EM grid preparation. Adapted from Kastner et al., 2008.

In this way, every single complex that initially adsorbes to the carbon film can be visualized and, dependent on its orientation on the film, different views of the complex may be observed. For different reasons, the signal to noise ratio of a single particle image is rather low. To increase the signal, a variety of complex images are selected (e.g. $\geq 5,000$ “raw images”) and subjected to image processing (Frank, 2006). Therefore, the individual images need to be aligned and classified. Alignment involves placing the images of particles into a similar orientation by rotating and shifting each image on the screen (Dube et al., 1993). During particle classification, similar images are grouped and separated from images that are distinct (van Heel, 1989). The images of each group are now “added up” to create an averaged 2D image (“class average”, or “class sum”) with an increased signal to noise ratio, in which the structural details are highly accentuated compared to the raw image. These days, aligning, classification and averaging is performed bias-free and automated using computer-assisted algorithms (van Heel et al., 1996). Based on high-quality 2D class averages it is possible under certain conditions to reconstruct also the 3D structure of a macromolecular complex.

The higher order 2D structure of human Prp19/CDC5L complexes and 35S U5 snRNPs was analyzed by EM. EM of the hPrp19/CDC5L complex was performed by Dr. Elmar Wolf at the group of Cellular Biochemistry at the MPI for Biophysical Chemistry. EM analysis of the 35S U5 snRNP was carried out by Dr. Monika Mariola Golas and Dr. Björn Sander at the 3D Electron Cryomicroscopy Group at the MPI for Biophysical Chemistry, headed by Prof. Holger Stark. For EM analysis, the particles were fixed by the GraFix procedure (Kastner et al., 2008) on a linear glycerol gradient containing 0-0.1% glutaraldehyde, as described in 2.2.10.3 for the hPrp19/CDC5L complex, and 2.2.10.4 for the 35S U5 snRNP. Negative staining was carried out by the perforated double carbon film method as described in Kastner, 1998. Briefly, complexes in solution were allowed to adsorb to a thin carbon film. The immobilized complexes were then transferred to a copper grid covered with a perforated carbon support film and stained for 2 min in a 2.5% (w/v) uranyl formate solution. Subsequently, they were overlaid with a second perforated carbon film. Images were recorded with a Philips CM 200 FEG electron microscope (Philips, Eindhoven, Netherlands) at a primary magnification of 105,000 \times with a TVIPS 4k \times 4k CCD camera (TemCam-F415; Gauting, Germany) using a Philips holder. Images were taken under low dose conditions at \sim 1.5 μ m defocus. Imagic-5 was used for image processing of \sim 5000 individual molecular images (van Heel et al., 1996). Briefly, after a “reference-free” alignment procedure (alignment by classification) (Dube et al., 1993), images were subjected to multivariate statistical analysis (MSA) (van Heel and Frank, 1981) and classification (van Heel, 1989). Based on the classification, \sim 30 individual molecular images were averaged. The resulting class averages were used as reference images in subsequent rounds of alignment, until the computed class averages were stable.

3D reconstruction of the 35S U5 snRNP was performed by Dr. Monika Mariola Golas and Dr. Björn Sander at the 3D Electron Cryomicroscopy Group at the MPI for Biophysical Chemistry headed by Prof. Holger Stark. In brief, 7890 unstained tilt pairs were selected manually, and iterative rounds of multi-reference alignment using resampling to polar coordinates and multivariate statistical analysis were performed in the context of IMAGIC-5. 3D maps of each class average were calculated according to the RCT method and a weighted average of these was determined. For refinement using unstained cryo-EM, 21476 images were selected manually, defocus-corrected and subjected to projection matching. Following the corrim-based alignment, the iteratively refined reference projections had an angular separation of 0.5 $^\circ$ in the final alignment round. 3D maps are presented as isosurface views (AmiraDev 2.3, TGS Europe, Merignac Cedex, France).

2.2.11.7 X-ray crystallography

Determination of the complex arrangement of atoms within a molecule, such as a protein or nucleic acid, is the main goal of structural biology and provides the basis to explain or predict their function and interaction with other molecules. X-ray crystallography is a powerful tool to obtain high-resolution structures of biological macromolecules and the growing number of crystal structures has deepened our understanding of various cellular processes. X-ray crystallographic structure determination of a biological macromolecule involves three basic steps (Drenth, 2007). The first and also less understood step is the crystallization of the macromolecule. Crystals are 3D ordered structures that can be described as a repetition of identical unit cells. The unit cell is made up of the smallest possible volume that represents the full symmetry of the entire crystal. The quality of the crystal is of vital importance for the success of the experiment. Crystals should have a sufficient size, a pure composition and a regular structure, and should not contain significant internal imperfections. Therefore, crystal growth usually requires a sample of highest possible purity and concentration.

The most frequently used crystallization method is the vapor diffusion technique. In this method, a small volume of the concentrated sample is initially mixed with the same volume of a crystallizing solution (e.g. buffer, salt and precipitant) and placed in a sealed reservoir containing the crystallization solution, either on a cover slip placed above the reservoir (hanging drop) or on a vessel in the reservoir (sitting drop). The difference in concentration between the sample solution and the reservoir causes vapor diffusion until equilibrium is reached. During equilibration, the sample is brought to supersaturation, which may lead to sample crystallization. A standard crystallization setup contains several hundred different conditions that influence crystal growth, in order to screen for the ideal crystallization conditions (e.g. buffer, salt, precipitant, temperature, drop size and shape). In the second step, a collimated beam of monochromatic X-rays is directed through the crystal. Traditionally, X-rays are produced by the impact of high-velocity electrons on a target of pure metal (e.g. copper) emitting "bremsstrahlung". Today, the brightest and most useful X-rays are generated in a synchrotron light source, allowing much faster data collection and often a higher resolution. On their way through the crystal, the X-rays are scattered in all directions by the electrons surrounding the nuclei of the atoms in the molecule, with a magnitude proportional to the size of its electron complement. The regular image of reflections is recorded on a detector plate, thereby generating the diffraction pattern. The magnitudes and phases of the scattered waves that contribute to the diffraction pattern are strictly dependent on the atomic numbers of the atoms and their relative position in the crystal. During the measurement, the crystal is gradually rotated in the X-ray beam and the diffraction pattern is recorded at various orientations of the crystal. In the third

step, the collected diffraction data is combined computationally to produce and refine a model of the arrangement of the atoms within the crystal. However, the diffraction pattern does not contain any information about the phases of the scattered waves, which is strictly required to solve the crystal structure. Several methods have been developed to solve the phase problem, either experimentally (e.g. Multiple Isomorphous Replacement, Single Wavelength Anomalous Dispersion), or by using the phases of similar and already solved structures (Molecular Replacement). Knowing the phases allows an initial model to be built that when refined leads to the crystal structure of the molecule.

X-ray crystallography and structure determination was performed in collaboration with the X-ray crystallography group at the MPI for Biophysical Chemistry headed by Dr. Vladimir Pena. In order to identify successful crystallization conditions, diverse commercially available crystallization screens were performed using a Cartesian liquid dispensing robot for nanoliter-setups. This procedure allowed the minimization of sample volumes and testing of a broad range of different chemical compounds. All screens used are listed under "Commercial kits and crystallization screens" (2.1.10). Pre-screens were performed by the sitting-drop vapor diffusion technique in a 96-well plate format with reservoir volumes of 100 μ l and a drop volume of 100 nl. Initial crystallization conditions were typically refined in 24-well linbro plates by small variations in the pH values and in precipitant or salt concentrations. Conditions were further improved by performing commercial additive screens in 96-well sitting drop plates using the Cartesian liquid dispensing robot or in 24-well linbro plates. Refinements were also carried out using the hanging drop vapor-diffusion technique.

The WD40 domain of Prp19 was crystallized by the sitting drop vapor diffusion method using a reservoir containing 25% (w/v) PEG 2000 MME in 0.1 M Tris-HCl (pH 8.5). The crystals were cryoprotected with perfluoropolyether and subsequently flash frozen in liquid nitrogen for data collection. Diffraction data were measured on beamline PXII of the Synchrotron Light Source at the Paul Scherrer Institute (Villingen, Switzerland). Diffraction data were processed using XDS (Kabsch, 1993). The structure was solved by molecular replacement with the program Molrep (Vagin and Teplyakov, 2000) using the structure coordinates from a WD40 domain structural model generated with the HHpred interactive server for protein homology detection and structure prediction (Soding et al., 2005). Automated refinement and building of the water structure were done with Refmac (CCP4i suite, (Potterton et al., 2003)) and the model was manually adjusted by using Coot (Emsley and Cowtan, 2004). The final structural models were validated with PROCHECK (Laskowski et al., 1993). Secondary structure elements were assigned using the program PROMOTIF (Hutchinson and Thornton, 1996). Coordinates and structure factors of the Prp19 WD40 domain will be deposited in the Protein Data Bank (<http://www.pdb.org>) upon publication.

2.2.11.8 Analytical ultracentrifugation

Analytical ultracentrifugation is a paramount technique to characterize the dynamic interactions of macromolecules in solution (reviewed in Howlett et al., 2006). During analytical ultracentrifugation, the macromolecular complex is subjected to adjustable radial forces in an ultracentrifuge. Data acquisition in analytical ultracentrifuges is usually obtained by real-time recording of an optical signal of the complex (e.g. UV-VIS) as a function of radial distance. In general, two basic principles exist in analytical ultracentrifugation. Static associations are most commonly studied by the hydrodynamic technique of sedimentation velocity (Schuck, 2003). In a sedimentation velocity experiment, macromolecular complexes are fractionated at high centrifugal force. The time-dependent sedimentation rate of the complex gives rise to the sedimentation coefficient and characterizes the sedimentation behaviour of the complex during centrifugation. The sedimentation coefficient that is measured in Svedberg units (S) is defined as the ratio of the maximal sedimentation velocity of the particle to the acceleration that is applied to it. The sedimentation rate provides information about the homogeneity of a particle's mass and conformation, its stoichiometry and overall shape (e.g. spherical, elongated or rod-like). It also gives an estimation of the particle's molecular mass. Furthermore, it helps to detect aggregates in the sample and is able to quantify the rate of aggregation. A thermodynamic approach is applied in a sedimentation equilibrium experiment (Vistica et al., 2004). This method is typically used to determine dynamic associations of molecules that are rapidly reversible. Sedimentation equilibrium is reached, if the rate of sedimentation of the particle due to centrifugation equals the rate of diffusion in the opposite direction. Thus, a concentration gradient is distributed across the centrifugation cell. At sedimentation equilibrium, the concentration distribution depends only on the molecular mass of the particle and is entirely independent of its shape. Therefore, it is possible to determine the molecular weight of a macromolecular complex with high accuracy by performing analytical ultracentrifugation.

Analytical ultracentrifugation of the hPrp19/CDC5L complex was performed in collaboration with Adrian Schomburg from the Chromatin Biochemistry Group at the MPI for Biophysical Chemistry headed by Dr. Wolfgang Fischle. For this purpose, affinity-purified hPrp19/CDC5L complexes (2.2.10.3) were further purified by size exclusion chromatography using a Superose 6 10/300 GL column in 1× PBS (pH 8.0). The peak fractions containing highly purified hPrp19/CDC5L complexes were pooled and the OD₂₈₀ was adjusted to the required concentration with 1× PBS (pH 8.0). Sedimentation was monitored on a Beckmann XL-A analytical ultracentrifuge, making use of the protein intrinsic absorbance at 280 nm. Two-sector 12 mm cells were loaded with 400 µl of the hPrp19/CDC5L complex solution for sedimentation velocity or 180 µl for sedimentation

equilibrium analysis. Sedimentation velocity runs were performed at four different concentrations (0.15, 0.3, 0.6, or 1.2 OD₂₈₀) by centrifugation at 40,000 rpm and 20°C until all material was pelleted. The raw data was analyzed using the SEDFIT software (Schuck, 2000) with a continuous $c(s)$ distribution model (Schuck et al., 2002). Parameters were set to a range of 2–40 S with a maximal resolution of 150 S values. Curves for $c(s)$ were plotted using maximum entropy regularization at 1 σ (SIGMA). The density of the buffer (1.00652 g/ml) and viscosity (1.021 centipoises), as well as protein partial specific volumes (V_{bar} , see Table 2.1) were calculated from primary sequences using the SEDNTERP software (Schuck, 2003). Fits were calculated while floating the meniscus, frictional coefficient and baseline, but keeping the buffer parameters constant, until the deviations of fits to raw data were adequate. Sedimentation equilibrium data was acquired at 3500 and 5500 rpm with a complex concentration of 1.2 OD₂₈₀.

Table 2.1: Molecular weight (MW) and V_{bar} values of the hPrp19/CDC5L complex proteins. Sequence and MW information was acquired from the NCBI database. V_{bar} was calculated using the SEDNTERP program (Schluck, Anal Biochem, 2003).

Protein	MW / [Da]	V_{bar} / [g/ml]
CDC5L	92250	0.7340
HSP73	70898	0.7362
CTNBL1	65173	0.7372
PRL1	57193	0.7332
hPrp19	55180	0.7402
AD002	26624	0.7168
SPF27	26131	0.7272

The sedimentation equilibrium approach was monitored using the SEDFIT software at 6 hrs intervals, and final scans were recorded when no systematic deviations between subsequent scans could be observed. Species analysis was performed using the SEDPHAT software (Schuck, 2003). For initial analysis, a single species was initialized at 11.8 S and 600 kDa, which did not result in a meaningful description of the raw data. Iterations between the Simplex and Marquard Levenberg routines, as well as the inclusion of two additional species that became apparent from the $c(s)$ analysis resulted in significantly better fits when floating the molecular weight of the 11.8 S complex, the cell bottom, as well as the baseline. Monte Carlo analysis using 1,000 iterations was employed to estimate the error of analysis.

3 RESULTS

The spliceosome is a highly dynamic molecular machine that assembles anew onto each intron. After the stable integration of all U snRNPs and the formation of the pre-catalytic B complex, major conformational rearrangements occur, leading to the formation of the active center of the spliceosome. In addition to U snRNPs, the spliceosome also contains several non-snRNP proteins and many of these proteins play important roles during splice site recognition, spliceosome assembly, and formation of the active sites. Of particular importance are those proteins that are recruited immediately prior to or during catalytic activation of the spliceosome, suggesting that they are key players in remodeling the spliceosome's active sites for splicing catalysis. Recent studies have shown that the human Prp19/CDC5L complex is first recruited during B complex formation and that its interaction with the spliceosome is stabilized during catalytic activation of the spliceosome (Bessonov et al., 2008; Deckert et al., 2006; Makarov et al., 2002). These studies also demonstrated that the hPrp19/CDC5L core components remain stably associated in a 1 M salt treated RNP core of the spliceosome (Bessonov et al., 2008), as well as in the post-spliceosomal 35S U5 snRNP (Makarov et al., 2002), indicating that there is no significant remodeling of the spliceosome's core after the first step of splicing has occurred. Thus, the elucidation of the molecular organization and structure of the hPrp19/CDC5L complex and 35S U5 snRNP can provide important information about the possible architecture of the RNP core of a catalytically active spliceosome.

3.1 The molecular architecture of the human Prp19/CDC5L complex

3.1.1 Establishing HeLa cell lines stably expressing FLAG/HA-AD002 or FLAG/HA-SPF27

In order to purify preparative amounts of the human Prp19/CDC5L complex and the 35S U5 snRNP, we established HeLa S3 cell lines stably expressing N-terminally FLAG/HA-tagged AD002 and SPF27, both bona fide integral components of the hPrp19/CDC5L complex, as well as the 35S U5 snRNP. For this purpose, the protein coding regions of the AD002 and SPF27 genes were cloned into the mammalian expression vector pIRESneo-FLAG/HA (Malik and Roeder, 2003) (Fig. 3.1A).

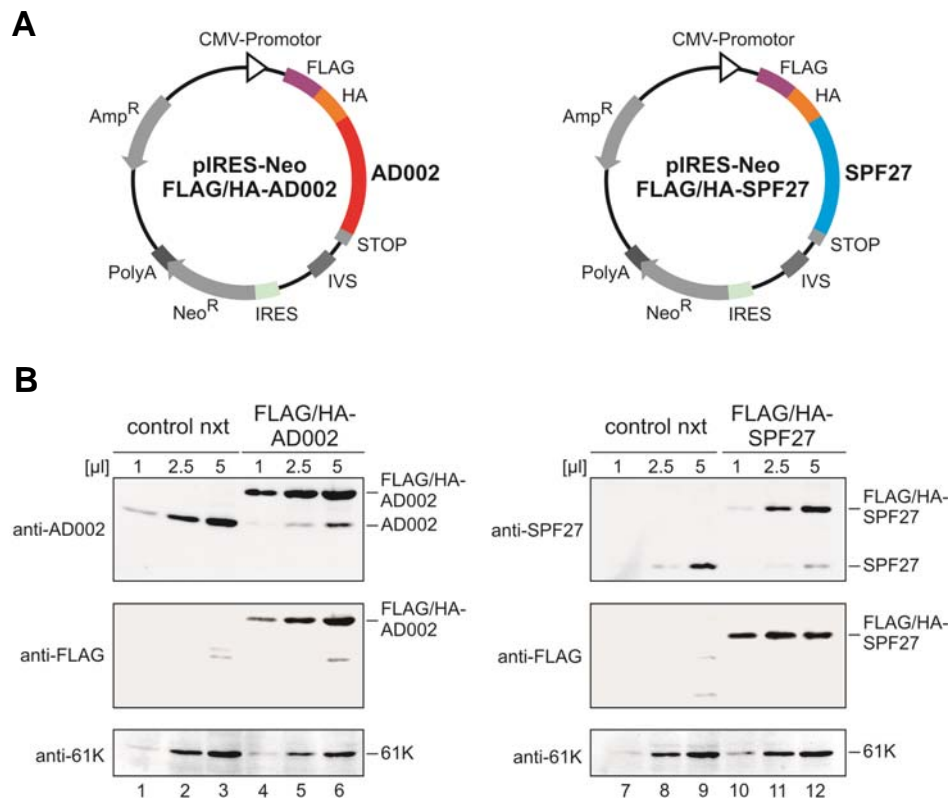


Fig. 3.1: Characterization of HeLa cell lines stably expressing FLAG/HA-tagged AD002 or SPF27. (A) Plasmids that were used for the stable transfection of HeLa cells contain the FLAG/HA-tagged AD002 or SPF27 open reading frames under the control of a Cytomegalovirus (CMV) promoter. Alternatively, HeLa cells were transfected with a plasmid carrying only the antibiotic resistance but no protein-coding gene (control, not shown). Amp^R – Ampicillin resistance gene; IRES – Internal ribosomal entry site; IVS – Intervening sequence; Neo^R – Neomycin resistance gene; PolyA – Poly(A) tail; STOP – stop codon. (B) Nuclear extract (nxt) was prepared from FLAG/HA-AD002, FLAG/HA-SPF27, or control cell lines as indicated above. Proteins were recovered from increasing amounts of nuclear extract (as indicated above each lane) and separated by SDS-PAGE. Proteins were transferred to a nitrocellulose membrane and further analyzed by western blotting with anti-FLAG (lanes 1-12, middle panels), anti-AD002 (lanes 1-6, upper panel) and anti-SPF27 antibodies (lanes 7-12, upper panel). As a loading control, all membranes were also probed with anti-61K antibodies (lanes 1-12, bottom panel). The position of the endogenous and the FLAG/HA-tagged protein is indicated on the right of each panel. HeLa cell lines stably expressing FLAG/HA-tagged AD002 or SPF27 were initially established by Dr. Ira Lemm (Cellular Biochemistry group, MPI for Biophysical Chemistry).

The vector introduces both FLAG (Chiang and Roeder, 1993) and HA (Pati, 1992) epitope tags to the N-terminus of the expressed proteins, under the control of the constitutive CMV promoter that allows for high expression rates in mammalian cells. These plasmids were used for the stable transfection of HeLa cells. During stable transfection, the plasmid enters the HeLa cell and integrates into the nuclear genome of the cell. As only a few cells integrate the foreign genetic material into their genome, the plasmids also contain a Neomycin resistance gene as a selectable marker that is co-transfected together with the gene of interest. By addition of the antibiotic G418 to cell culture, only those cells with the marker gene integrated into their genomes will be able to proliferate. Once a positive clone was identified, it was expanded, and adapted to suspension

culture. Dignam-type nuclear extracts (Dignam et al., 1983) were prepared and used as a source for further analysis. HeLa cell lines stably expressing FLAG/HA-tagged AD002 or SPF27 were initially established by Dr. Ira Lemm (Cellular Biochemistry group, MPI for Biophysical Chemistry).

For analysis of the overexpression of FLAG/HA-tagged AD002 and SPF27, proteins from nuclear extracts were first separated by SDS-PAGE and transferred to nitrocellulose membrane. Subsequent western blotting with antibodies against either protein (AD002 or SPF27) or the FLAG epitope confirmed the presence and overexpression of the tagged proteins in nuclear extract prepared from these cell lines (Fig. 3.1B). Interestingly, the concentration of endogenous AD002 or SPF27 is reduced in cells overexpressing FLAG/HA-AD002 or FLAG/HA-SPF27, respectively, compared to wildtype cells (Fig. 3.1B, upper panels). Furthermore, the overall protein concentration in nuclear extracts prepared from the control or stable cell lines is similar, as was shown by immunoblotting with anti-61K antibodies (Fig. 3.1B, lower panel). This suggests there may be a feedback mechanism that controls the level of expression of both proteins in the cell.

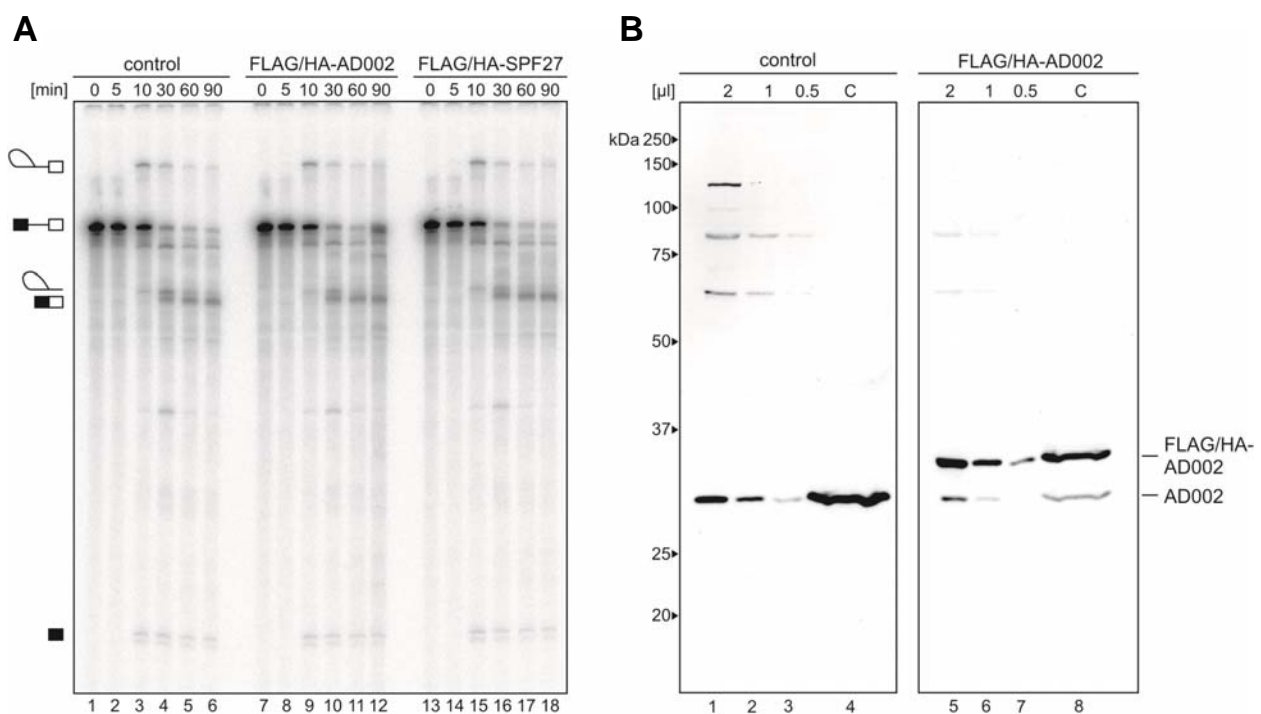


Fig. 3.2: The presence of a FLAG/HA-tag does not affect pre-mRNA splicing. (A) Splicing was performed with [32 P]-labeled MINX pre-mRNA in HeLa nuclear extract prepared from the control cell line (lanes 1-6), FLAG/HA-AD002 cell-line (lanes 7-12), or FLAG/HA-SPF27 cell-line (lanes 13-18) for 0 to 90 min as indicated. RNA was recovered and separated on a 8.3M urea-10% polyacrylamide gel. The [32 P]-labeled pre-mRNA and splicing intermediates or products were detected by autoradiography, and their positions are indicated on the left. (B) FLAG/HA-tagged AD002 incorporates into spliceosomes to the same extent as the wildtype protein. Spliceosomal C complexes were affinity-purified from nuclear extracts produced from FLAG/HA-AD002 producing HeLa cells, as well as wildtype cells. Western blotting was performed with anti-FLAG antibodies and nitrocellulose stripes containing either different amounts of nuclear extract (lanes 1-3 and lanes 5-7) or purified C complexes (lanes 4 and 8). The molecular mass (kDa) of marker proteins is indicated on the left.

Next, we determined whether overexpression of FLAG/HA-AD002 or FLAG/HA-SPF27 influences the splicing efficiency of MINX pre-mRNA *in vitro*. As shown in Fig. 3.2A, nuclear extract from either stable cell line exhibited the same splicing activity as the wildtype extract. To further analyze the functionality of the tagged proteins, we purified spliceosomal C complexes via MS2 selection (Bessonov et al., 2008) from FLAG/HA-AD002 stable cell lines or wildtype cells and analyzed the proteins by western blotting, using antibodies against AD002 (Fig. 3.2B). Significantly, the tagged protein was incorporated into affinity-purified C complexes to the same extent as the wildtype protein (Fig. 3.2B, lane 8). Thus, assuming that all AD002 enters the spliceosome as part of the hPrp19/CDC5L complex, the presence of a FLAG/HA-tag does not alter the function of the AD002 protein and appears to have little or no effect on the structure of the hPrp19/CDC5L complex.

3.1.2 Isolation of the human Prp19/CDC5L complex

Previously, hPrp19-containing complexes were isolated from HeLa nuclear extracts by immunoaffinity chromatography using polyclonal antibodies against AD002 or CDC5L (Ajuh et al., 2000; Makarova et al., 2004). In the latter study, affinity purification with anti-CDC5L antibodies led to the isolation of a well-defined hPrp19/CDC5L heteromeric complex, consisting of seven proteins: hPrp19, CDC5L, PRL1, HSP73, β -catenin-like 1 (CTNNB1), AD002 and SPF27 (Makarova et al., 2004). However, only limited amounts could be obtained using this method mainly because of the lack of antibodies. To overcome these problems, we used monoclonal anti-FLAG antibodies coupled to agarose-beads for the immunoaffinity selection of human Prp19/CDC5L complexes from nuclear extract prepared from cell lines stably expressing FLAG/HA-AD002 or FLAG/HA-SPF27 (Fig. 3.3A). Complexes were eluted under native conditions with an excess of FLAG peptide, and further purified on a 5-20% glycerol gradient. Proteins were isolated from fractions across the gradient and analyzed by SDS-PAGE (Fig. 3.3B and C). In both cases, a multi-component complex peaking in fractions 15-18 was observed. This affinity-purified complex was comprised of hPrp19, CDC5L, HSP73, CTNNB1, PRL1, AD002 and SPF27, as determined by MS (Fig. 3.3A and B, lanes 15-18), confirming it is the hPrp19/CDC5L complex. MS analysis was performed in collaboration with the Bioanalytical Mass Spectrometry group of Dr. Henning Urlaub (MPI for Biophysical Chemistry). Interestingly, a fraction of FLAG/HA-AD002 and CTNNB1 (Fig. 3.3B, lane 7), as well as of FLAG/HA-SPF27 and hPrp19 (Fig. 3.3C, lane 12), co-sedimented on the gradient, indicating that these proteins can interact with one another. Furthermore, a large amount of free FLAG/HA-AD002 co-purified with the hPrp19/CDC5L complexes and was found in the top fractions of the gradient (Fig. 3.3B, lanes 2-6), showing that this protein was indeed overexpressed. However, no free FLAG/HA-SPF27 protein was isolated from the FLAG/HA-SPF27 HeLa stable cell line (Fig. 3.3C).

The entire FLAG/HA-SPF27 protein was either found in hPrp19/CDC5L complexes (Fig. 3.3C, lane 15-18) or associated together with hPrp19 (Fig. 3.3C, lanes 11-13).

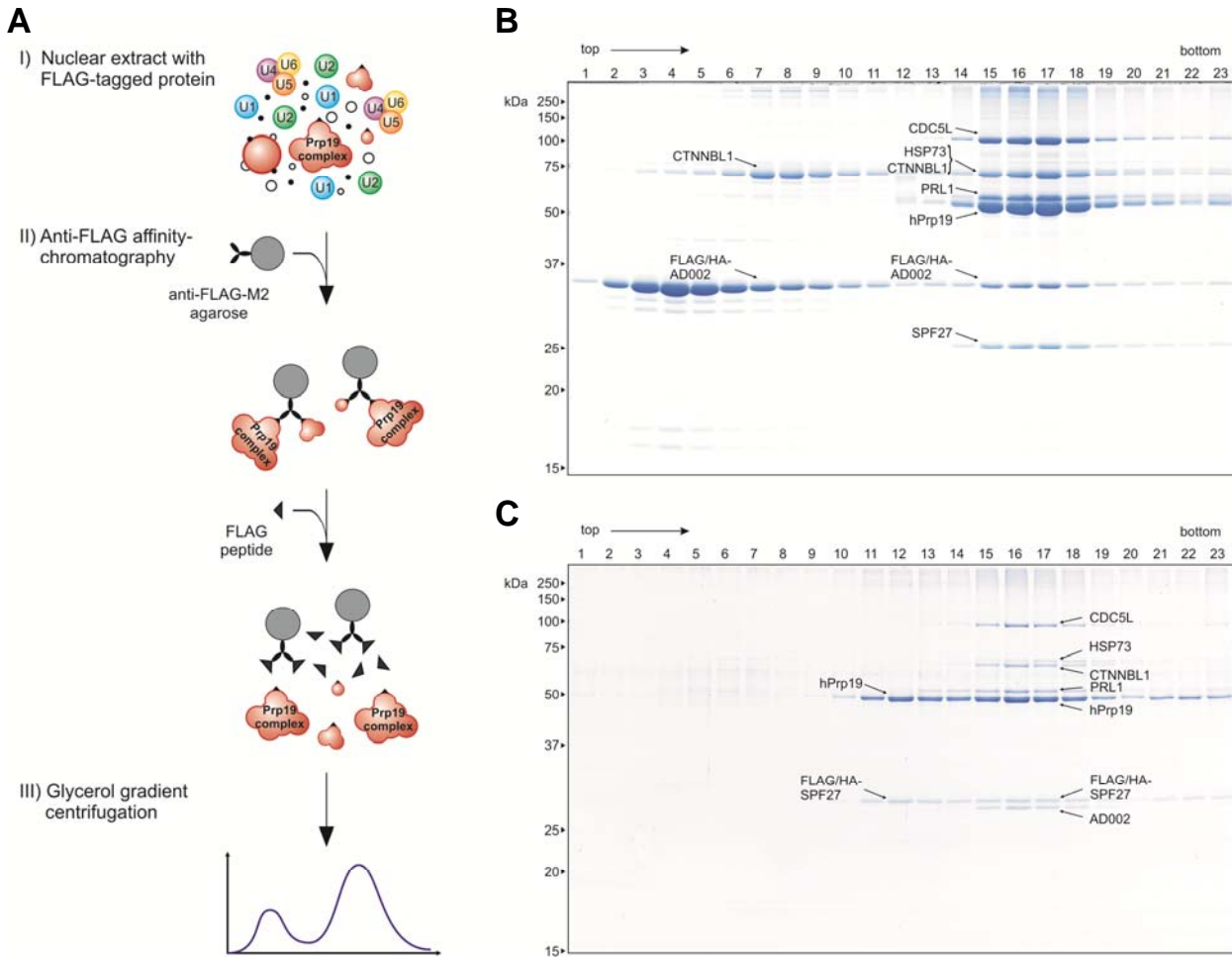


Fig. 3.3: Purification of the hPrp19/CDC5L complex from HeLa cell lines stably expressing FLAG/HA-tagged AD002 or SPF27. (A) Schematic diagram of the anti-FLAG immunoaffinity purification strategy. Complexes present in nuclear extract prepared from either cell line were immunoaffinity selected using anti-FLAG-M2-Agarose and subsequently eluted under native conditions with FLAG peptide. Affinity purified complexes were then fractionated on a linear 5 to 20% glycerol gradient. Proteins present in each gradient fraction after affinity-purification from (B) the FLAG/HA-AD002 cell line and (C) the FLAG/HA-SPF27 cell line were analyzed by SDS-PAGE and stained with Coomassie. The molecular mass (kDa) of marker proteins is indicated on the left. Proteins were identified by MS and are indicated by arrows: CDC5L (gi|11067747), HSP73 (gi|5729877), CTNNBL1 (gi|18644734), PRL1 (gi|4505895), hPrp19 (gi|7657381), AD002 (gi|6523797), and SPF27 (gi|5031653). MS analysis was performed in collaboration with the Bioanalytical Mass Spectrometry group of Dr. Henning Urlaub.

The presence of CTNNBL1 in the top fractions after gradient centrifugation raised the question, whether it is indeed part of a binary complex together with AD002 or instead dissociated alone from the hPrp19/CDC5L complex, due to the high centrifugal forces of glycerol gradient centrifugation. To analyze this, we also separated affinity-purified hPrp19/CDC5L complexes via gel filtration (Fig. 3.4). Using buffer with a low salt concentration (e.g. 100 mM NaCl) we could identify CTNNBL1 co-migrating together with AD002 after Superose 6 (Fig. 3.4A, lane 13) as well as

Sephadex 200 chromatography (Fig. 3.4B, lane 10-11). Thus, it is likely that CTNNB1 is indeed part of a binary complex together with AD002. On the other hand, we still cannot exclude that each of these proteins is present as a monomer.

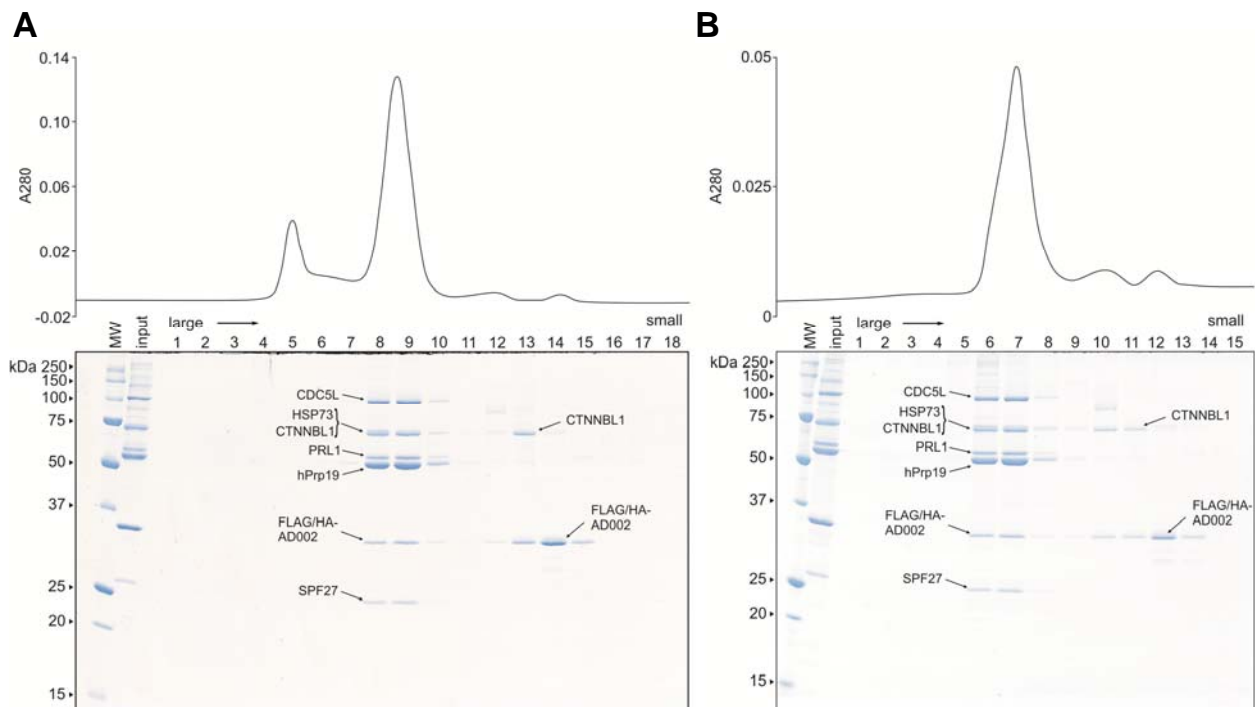


Fig. 3.4: Size exclusion chromatography of affinity purified hPrp19/CDC5L complexes. Complexes present in nuclear extract prepared from the FLAG/HA-AD002 stable cell line were immunoaffinity selected using anti-FLAG-M2-Agarose and subsequently eluted under native conditions with FLAG peptide. Affinity purified complexes were then subjected to (A) Superose 6 or (B) Sephadex 200 gel filtration. The absorption at 280 nm of proteins eluting from the gel filtration column was plotted against time (upper panel). Fractions after gel filtration were collected, analyzed by SDS-PAGE and stained with Coomassie (lower panel). The molecular mass (kDa) of marker proteins is indicated on the left. MW – molecular weight marker proteins.

3.1.3 Analysis of the stoichiometry of the native hPrp19/CDC5L complex reveals hPrp19 tetramerization

In the yeast *S. cerevisiae* it was shown that Prp19 forms a tetramer *in vivo* and *in vitro* (Ohi et al., 2005). An oligomerization was reported for the human Prp19 protein as well (Grillari et al., 2005), but it is presently unclear whether it also forms a tetramer in the native hPrp19/CDC5L complex. Furthermore, nothing is currently known about the stoichiometry of the other components of the hPrp19/CDC5L complex, as well as the yeast NTC. Therefore, we next determined the stoichiometry of the human Prp19/CDC5L complex. Analysis of purified hPrp19/CDC5L complexes by SDS-PAGE and subsequent staining with Coomassie showed a broad and intensively stained band for the hPrp19 protein, indicating that it may indeed be present in a higher copy number compared to the remainder of the hPrp19/CDC5L complex proteins (Fig. 3.3B and C). To analyze the stoichiometry of the hPrp19/CDC5L complex in more detail, we stained SDS-PAGE separated hPrp19/CDC5L

complex proteins with Sypro Ruby (Fig. 3.5A), or labeled denatured proteins with [^{14}C]iodoacetamide prior to SDS-PAGE (Fig. 3.5B). Fluorescent staining as well as iodoacetamide labeling was performed in collaboration with Dr. Dmitry E. Agafonov (Cellular Biochemistry group, MPI for Biophysical Chemistry). The fluorescent dye Sypro Ruby is a highly sensitive ruthenium complex-based stain that persistently interacts with proteins in a non-covalent fashion. Sypro Ruby offers several advantages over commonly used silver staining of proteins in SDS-gels, including linearity and stoichiometric staining of proteins, compatibility with the microchemical preparation of proteins for identification by MS techniques, and assessment of the staining endpoint (Berggren et al., 2000). Sypro Ruby has a bimodal excitation spectrum, with a UV peak of 280 nm and a visible peak of 450 nm. The emission maximum is relatively broad and centered at 610 nm. Thus, the stain can be visualized using commonly used excitation sources and due to its high linear dynamic range, which extends over three orders of magnitude, the stain signal is directly proportional to the protein concentration, relative to its size. Iodoacetamide is an alkylating sulfhydryl reagent.

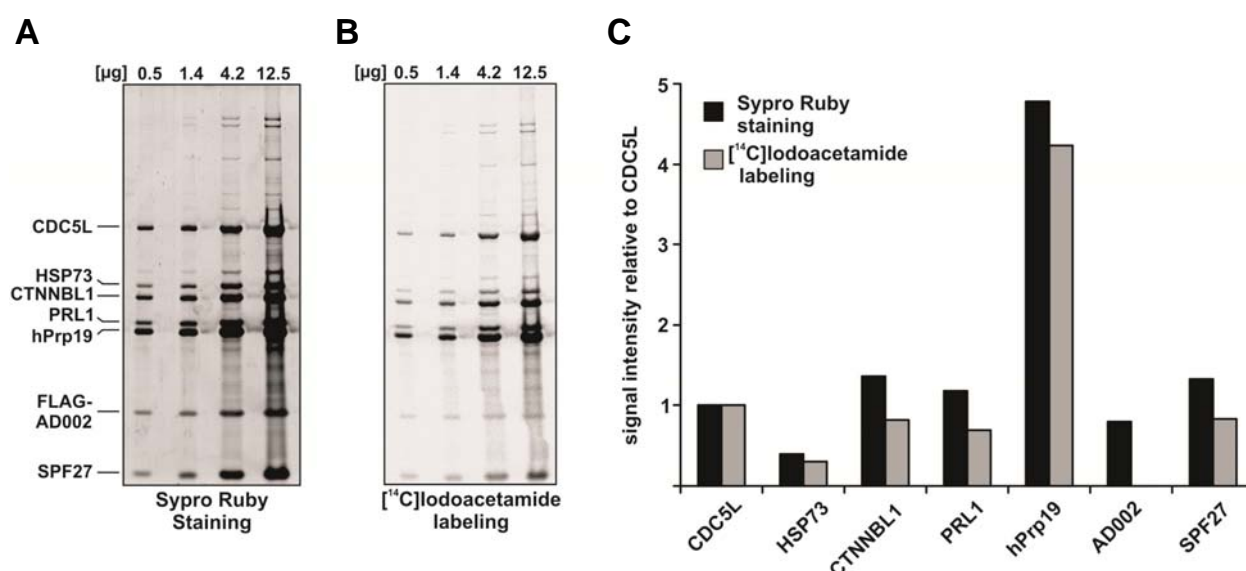


Fig. 3.5: Determination of the protein stoichiometry of the native human Prp19/CDC5L complex reveals four copies of hPrp19. (A) Determination of the protein stoichiometry by fluorescence staining. Four different amounts of hPrp19/CDC5L complex proteins (as indicated above each lane) were separated by SDS-gel electrophoresis and stained with Sypro Ruby. The fluorescence was detected by fluorescence imaging (FLA-7000, Fujifilm). (B) Determination of the protein stoichiometry by radioactive labeling. Denatured proteins (0.5-12.5 μg , as indicated) were labeled with [^{14}C]iodoacetamide, separated by SDS-PAGE and detected by autoradiography. Sypro Ruby staining, as well as iodoacetamide labeling was performed in collaboration with Dr. Dmitry E. Agafonov (Cellular Biochemistry group, MPI for Biophysical Chemistry). (C) Band intensities obtained after Sypro Ruby staining (black bars) and [^{14}C]iodoacetamide labeling (grey bars) were quantified using Fuji Image Gauge v4.22 and normalized against the CDC5L signal.

It is commonly used to acetylate cysteine residues of a protein, thereby preventing the re-formation of disulfide bonds after denaturation with SDS. Therefore, the use of [^{14}C]iodoacetamide allows radioactive labeling of proteins at cysteine residues. Exhaustive labeling of cysteine residues is

achieved by using a high excess of the label. Thus, the radioactive signal is proportional to the number of cysteine residues in the protein. As the number of cysteine residues for each protein in the hPrp19/CDC5L complex is known, it is possible to determine the relative stoichiometry of the proteins in the complex.

Band intensities of Sypro Ruby stained gels and the radioactivity signal caused by [¹⁴C]iodoacetamide acetylation of cysteine residues were quantified by fluorescent imaging and autoradiography, respectively. Subsequently, both signals were normalized against the CDC5L signal (Fig. 3.5C). To avoid effects due to non-linearity of the signal at a certain concentration of a protein that may lead to misinterpretation of the data, increasing amounts of the hPrp19/CDC5L complex were used for quantification (Fig. 3.5A and B). Only those signals were used for stoichiometry determination of the hPrp19/CDC5L complex that were linear over the whole concentration range. The linearity of the signals was further confirmed by analyzing increasing amounts of the protein BSA in the same way as described for the hPrp19/CDC5L complex (data not shown). As the signal obtained via iodoacetamide labeling of AD002 was not linear over the concentration range analyzed (Fig. 3.5B), its relative amount could not be quantified using this method. As shown in Fig. 3.5C, both methods revealed that approximately four times more hPrp19 is present in the hPrp19/CDC5L complex relative to the CDC5L protein. Except for HSP73, which seems to be underrepresented, all other components appear to be present in equimolar amounts.

To confirm these results, we also performed sedimentation velocity and equilibrium analytical ultracentrifugation (Fig. 3.6A and B). Analytical Ultracentrifugation of hPrp19/CDC5L complexes was performed by Adrian Schomburg (Chromatin Biochemistry group, MPI for Biophysical Chemistry). Analysis of the velocity data at 4 different concentrations showed that 84% of the hPrp19/CDC5L complex exhibited a weight-averaged sedimentation coefficient $S_{(20, \text{water})}$ of 11.8 (Fig. 3.6A). Furthermore, two additional species were identified, that sediment at 3.9 S (2.1%) and 17.6 S (5%). However, over the range of analyzed concentrations, no concentration dependent associations were observable. For these data a single species model leads to an apparent molecular mass estimate of the 11.8 S complex of approximately 600 kDa. Interestingly, the hPrp19/CDC5L complex has a frictional ratio value of 2.1, indicating that the particle's shape is elongated. To analyze the molecular weight of the hPrp19/CDC5L complex in more detail, we performed multi-speed sedimentation equilibrium analysis at 2 different rotor speeds (Fig. 3.6B). Using the information of the sedimentation velocity analysis at four different concentrations, a molecular weight of 608 kDa \pm 16.4 kDa for the 11.8 S species was obtained using the Sedphat program (Schuck, 2003).

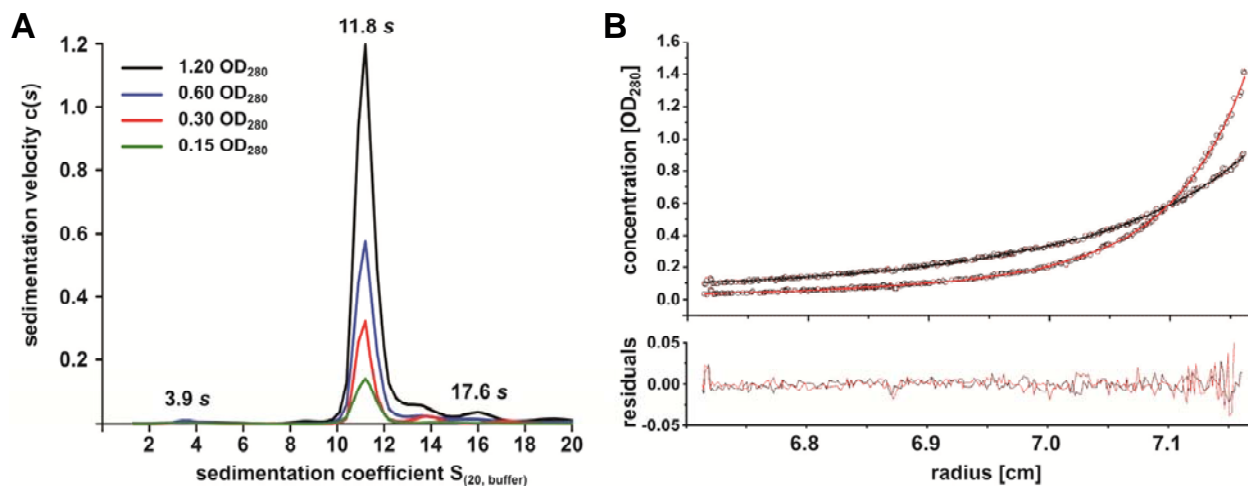


Fig. 3.6: Analytical ultracentrifugation of the hPrp19/CDC5L complex reveals its molecular mass. (A) Continuous size distribution analysis of the sedimentation velocity $c(S)$ of the hPrp19/CDC5L complex. Four different concentrations of the complex were analyzed (OD_{280} values of 1.2, 0.6, 0.3, and 0.15) as indicated in the figure. The calculated $c(S)$ is plotted versus the sedimentation coefficient (S). The S values in the panel are corrected according to standard conditions ($S_{(20,w)}$). (B) Multi speed sedimentation equilibrium analysis of the hPrp19/CDC5L complex at two different rotor speeds (red: 5500 rpm; black: 3500 rpm). Using the information of the $c(S)$ analysis, all 3 species (3.9 S, 11.8 S, and 17.6 S) are included in the species analysis performed with the program Sedphat (Schuck, 2003). The OD_{280} along the cell was recorded and is shown in the upper panel. The lower panel shows the residuals between the model (straight line in the upper panel) and the raw data (dots in the upper panel). Analytical ultracentrifugation was performed by Adrian Schomburg (Chromatin Biochemistry group, MPI for Biophysical Chemistry).

Taken together, the data obtained from analytical ultracentrifugation indicates that the human Prp19/CDC5L complex has a molecular mass of approximately 600 kDa. This is consistent with the presence of four copies of hPrp19 and single copies of the other proteins that sums up to approximately 560 kDa. Thus, within the hPrp19/CDC5L complex, hPrp19 appears to be present as a tetramer.

3.1.4 hPrp19, CDC5L, PRL1 and SPF27 form a salt-stable core of the hPrp19/CDC5L complex

Knowing the stoichiometry of the proteins in the hPrp19/CDC5L complex, we next investigated how these proteins are organized and which protein-protein interactions contribute to the structure of the complex. Information about the structural organization of proteins within the hPrp19/CDC5L complex may also provide important insights into the biogenesis of the complex. As shown already for the ribosome, the order of protein dissociation may reflect the order of their association, thus directing the process of subunit self-assembly (El-Baradi et al., 1984). Thus, we treated affinity-purified, gel filtrated hPrp19/CDC5L complexes with buffer containing increasing concentrations of NaCl, 0.1 mg/ml heparin or with 0.4 M of the chaotropic salt NaSCN, respectively (Fig. 3.7 and Fig. 3.8). Subsequently, we performed glycerol gradient centrifugation under the same

conditions. Dissociation of protein-protein interactions with kosmotropic ions like NaCl depends on the ionic character of the interaction. Those interactions with the weakest ionic strength are disrupted already at low salt concentration, whereas those molecules that have a very strong ionic interaction require a higher salt concentration for dissociation. Heparin is a linear, unbranched, highly sulfated polysaccharide that also interferes with the electrostatic association of proteins. In contrast, chaotropic agents interfere with stabilizing intermolecular interactions mediated by non-covalent forces such as hydrogen bonds, van der Waals forces and hydrophobic effects.

As shown in Fig. 3.7, hPrp19, CDC5L, PRL1 and SPF27 co-migrate after treatment with 1.2 M NaCl (Fig. 3.7A, lane 10-13), 0.1 mg/ml heparin (Fig. 3.7B, lane 17-19) or 0.4 M NaSCN (Fig. 3.7C, lane 12-13), indicating that they form a stable subcomplex. The intense staining of the hPrp19 band, together with the fact that no free hPrp19 is observed at the top of the gradient, indicates that the hPrp19 tetramer is present in this stable subcomplex. In contrast, AD002, CTNNBL1 and HSP73 are found in the top fractions of the gradients, demonstrating that they have dissociated and are thus more loosely-associated components. The diverse migration behaviour of the hPrp19/CDC5L sub-complexes between these three conditions is caused by the use of different salts and salt concentrations that were added to the gradient buffer. Together with the glycerol, salt also contributes to increased density of the gradient solution. Thus, the same complex migrates faster in a gradient with the same glycerol concentration, but with a lower salt concentration. Interestingly, a portion of AD002 and CTNNBL1 still co-migrated after treatment with 0.1 mg/ml heparin (Fig. 3.7B, lane 8) or 300 mM NaCl (Fig. 3.8A, lane 6-8), indicating that they interact. However, at higher NaCl concentrations (i.e. already at 400 mM and 600 mM NaCl, see Fig. 3.8B and C) and at 0.4 M NaSCN (Fig. 3.7C) the interaction between both proteins was completely disrupted (i.e., they no longer co-migrated during gradient centrifugation). We also note that CTNNBL1 and HSP73 appear to co-migrate after dissociation from the hPrp19/CDC5L complex, suggesting that they interact with one another. However, as CTNNBL1 and HSP73 have nearly identical molecular masses and we cannot distinguish between ~75 kDa (single proteins) and ~150 kDa (dimers) in this region of the gradient, and as they also do not appear to be present in equimolar amounts, it is also possible that CTNNBL1 and HSP73 dissociate as individual proteins. Taken together, salt dissociation of the human Prp19/CDC5L complex demonstrates that the salt-stable core of the complex is comprised of hPrp19, CDC5L, PRL1 and SPF27. Stability of this complex up to a salt concentration of at least 1.2 M NaCl, as well as 0.1 mg/ml heparin and 0.4 M NaSCN suggests several electrostatic and hydrophobic protein-protein interactions among these proteins contributing to its salt stability. Furthermore, these data indicate that a less salt-stable complex consisting of AD002, CTNNBL1 and HSP73 is associated with the salt-stable core via a salt sensitive interaction that likely involves only a few protein-protein interactions.

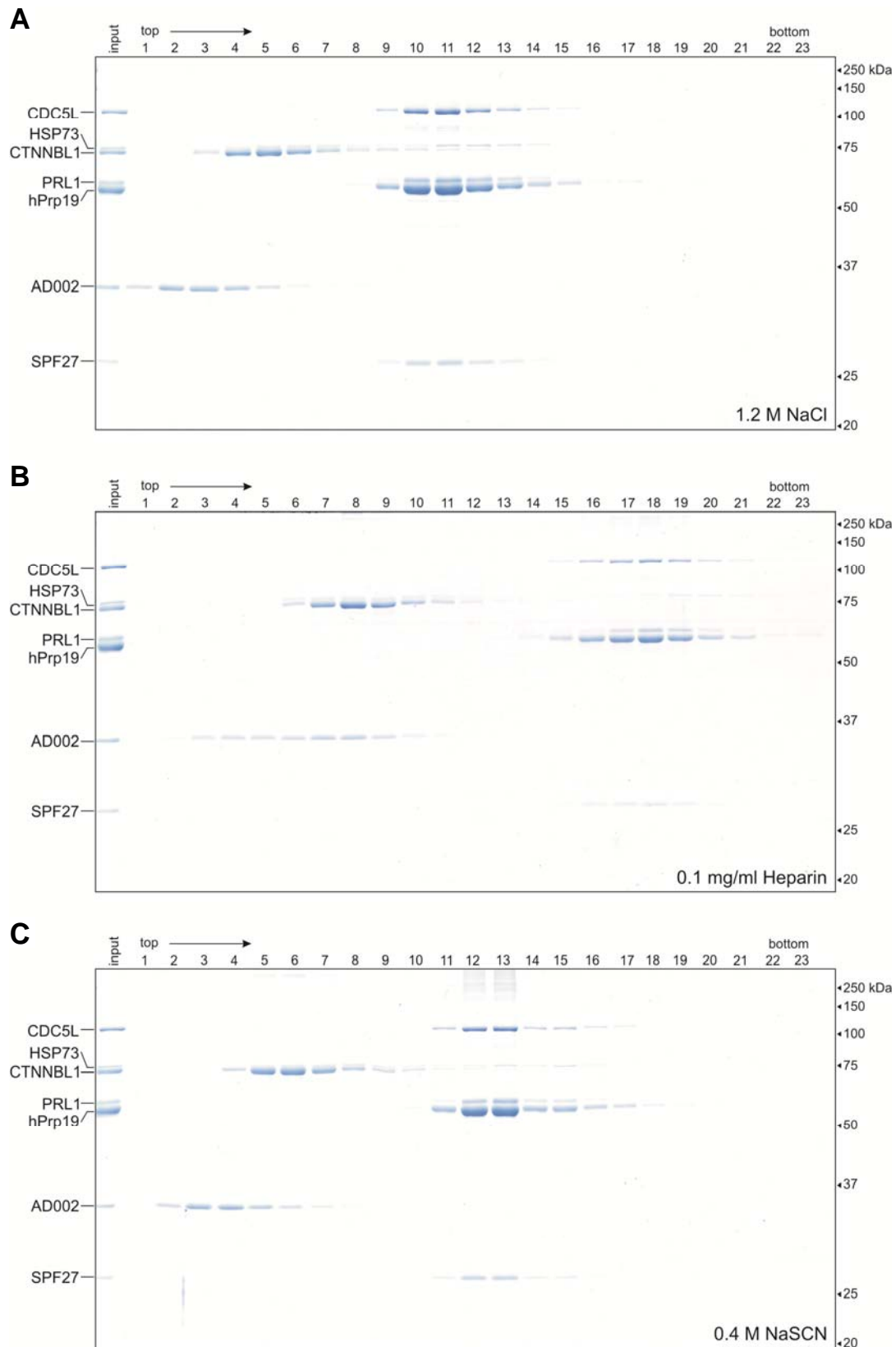


Fig. 3.7: CDC5L, PRL1, hPrp19 and SPF27 form a salt-stable subcomplex. Purified hPrp19/CDC5L complexes were incubated with (A) 1.2 M NaCl, (B) 0.1 mg/ml heparin, or (C) 0.4 M NaSCN for 30 min on ice. Subsequently, the complexes were fractionated on a linear 5 to 20% glycerol gradient containing the same salt or heparin concentration. Proteins across the gradient were analyzed by SDS-PAGE and stained with Coomassie. The position of the proteins is indicated on the left and the molecular mass (kDa) of marker proteins is indicated on the right.

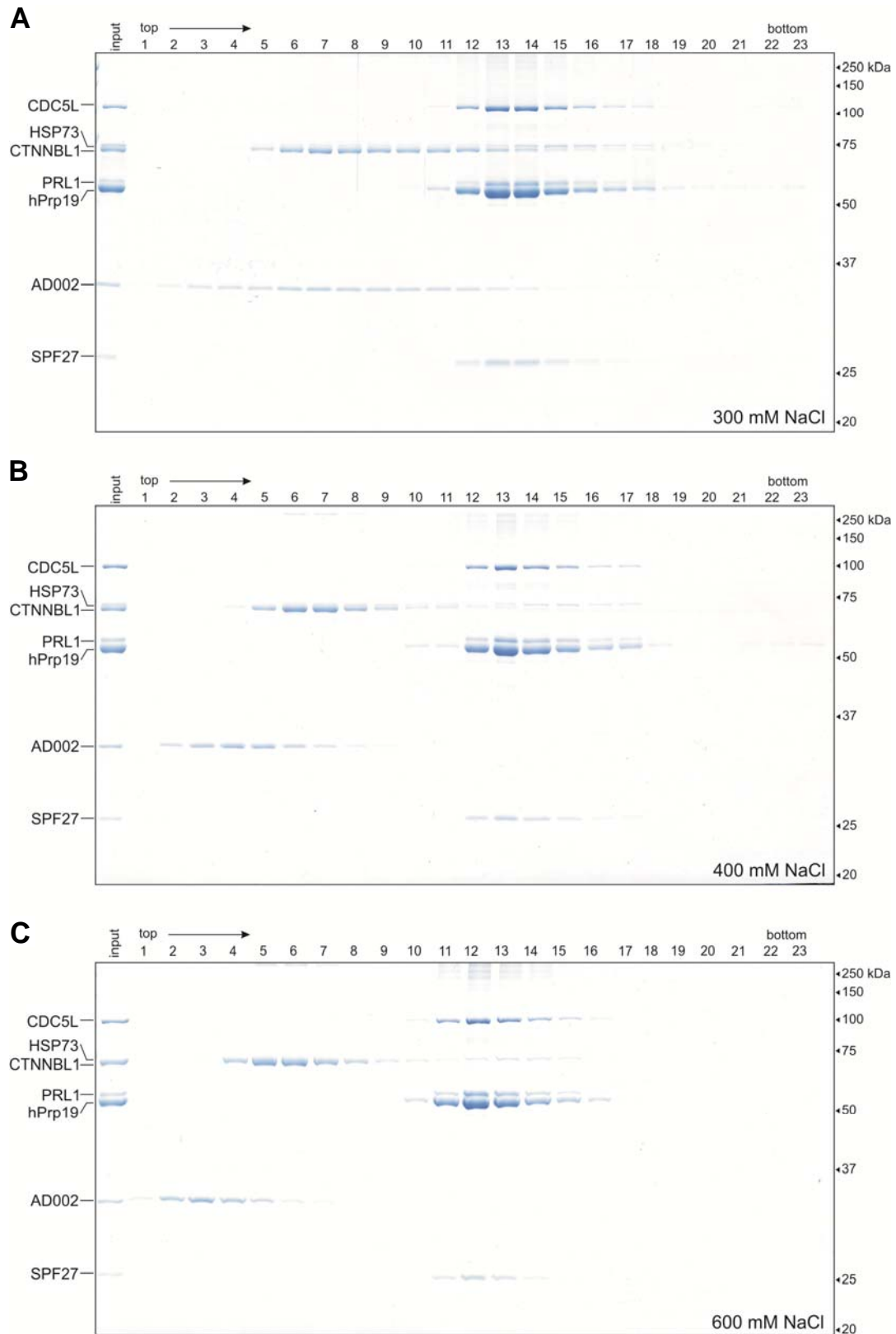


Fig. 3.8: Analysis of the salt stability of the human Prp19/CDC5L complex. Purified hPrp19/CDC5L complexes were incubated with (A) 300 mM NaCl, (B) 400 mM NaCl, or (C) 600 mM NaCl for 30 min on ice. Subsequently, the complexes were fractionated on a linear 5 to 20% glycerol gradient containing the same salt concentration. Proteins across the gradient were analyzed by SDS-PAGE and stained with Coomassie. The position of the proteins is indicated on the left and the molecular mass (kDa) of marker proteins is indicated on the right.

3.1.5 Crosslinking delineates protein-protein interactions within the native hPrp19/CDC5L complex

To further analyze interactions among the hPrp19/CDC5L complex proteins we performed protein-protein crosslinking within native complexes. Crosslinking can provide important information about the near-neighbor organization of the proteins within a native protein complex.

Initially, a variety of crosslinking reagents, including glutaraldehyde (GA), Butadien diepoxide (DEB), Dimethyl pimelimidate (DMP) and N-Maleimidobenzoyl-N-hydroxysulfosuccinimide ester (Sulfo-MBS) were tested in a small scale experiment for their ability to crosslink proteins within the hPrp19/CDC5L complex (Fig. 3.9). Crosslinked proteins were separated by SDS-PAGE and analyzed by staining with Coomassie. Whereas no effect could be observed even at high concentrations of DEB (Fig. 3.9, lane 8-10) and DMP (Fig. 3.9, lane 11-13), crosslinking with glutaraldehyde (Fig. 3.9, lane 5-7) was exhaustive already at low concentration, showing only one band at the very top of the gel and indicating that also intermolecular crosslinking took place. In the case of Sulfo-MBS several distinct bands could be distinguished from those of the hPrp19/CDC5L complex (Fig. 3.9, lane 2-4). Thus, Sulfo-MBS was used for an in depth analysis of protein-protein interactions in the hPrp19/CDC5L complex.

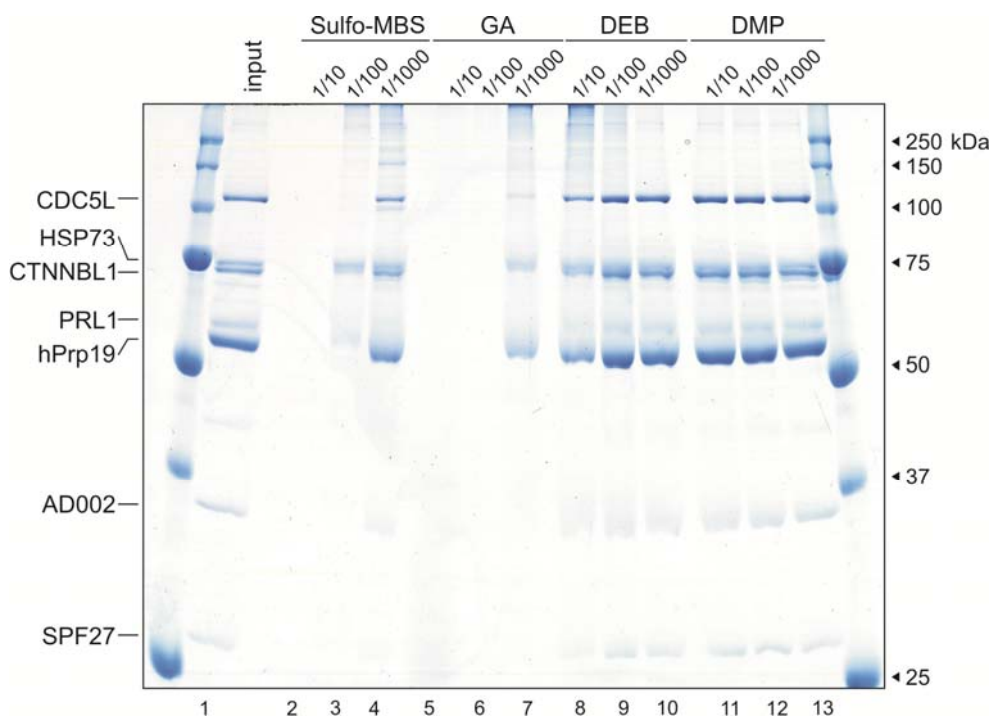


Fig. 3.9: Protein crosslinking in purified hPrp19/CDC5L complex with different crosslinking reagents. For each crosslinking reagent, three different dilutions were used. The starting concentrations were as follows: 1 mM Sulfo-MBS; 10% (w/v) glutaraldehyde (GA); 10% (w/v) Butadien diepoxide (DEB); 1 mM Dimethyl pimelimidate (DMP). The reaction was performed for 30 min at room temperature and stopped by adding 50 mM glycine-HCl (pH 7.5). Samples were analyzed by SDS-PAGE and stained with Coomassie. The position of the proteins is indicated on the left and the molecular mass (kDa) of marker proteins is indicated on the right. MW – molecular weight marker proteins.

Sulfo-MBS is a water soluble, non-cleavable, heterobifunctional crosslinking reagent that links a thiol group to an amino group at neutral pH (Myers et al., 1989). In Sulfo-MBS conjugation, side reactions are rare and usually not observed. The chemistry for this reagent is well characterized, providing a stable intermolecular chemical bond. Heterobifunctional crosslinkers are probably the most notable and useful protein crosslinking reagents because they have two distinct reactive groups. This allows the protein conjugation reaction to progress in a controlled, two-step reaction, thereby minimizing the formation of intramolecular dimers and polymers. As shown in Fig. 3.10, the protein provides the thiol group in the form of a cysteine residue and the amino group in the form of a lysine. First, Sulfo-MBS is activated by reaction of its succinimide moiety with the amino group. Then the thiol group of a cysteine residue reacts with the maleimide moiety of the activated Sulfo-MBS. Although there exist a number of possible side reactions involving other amino acids than lysine, such as histidine, the desired reaction is quite fast and problems owing to side reactivity are rare. The linker length of Sulfo-MBS is 7.3 Å, ensuring that only proteins are crosslinked that are in close proximity to each other.

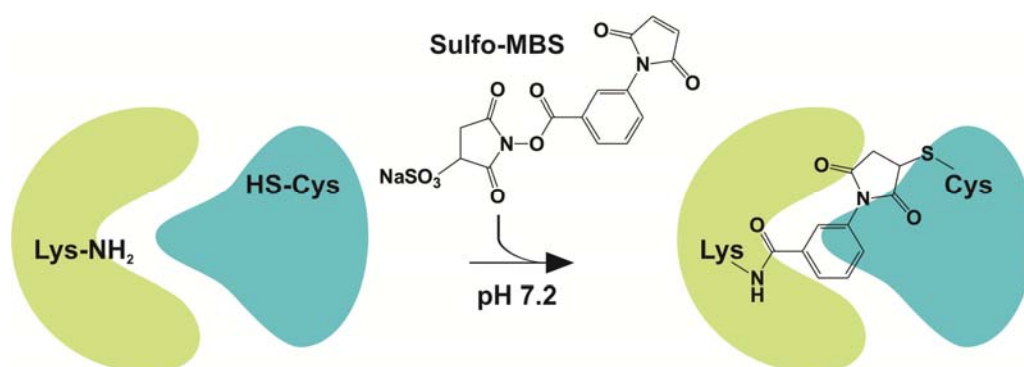


Fig. 3.10: Schematic representation of protein crosslinking with the heterobifunctional crosslinking agent Sulfo-MBS. Two interacting proteins are represented in green and blue, respectively. Sulfo-MBS preferentially reacts with the amino group of Lys residues and the thio-group of Cys residues in a two-step mechanism at neutral pH.

After treating hPrp19/CDC5L complexes with Sulfo-MBS and performing SDS-PAGE, crosslinks were identified by western blotting with antibodies specific for each protein of the complex (Fig. 3.11, lanes 1-7). At 1 μ M Sulfo-MBS a strong crosslink between the PRL1 and SPF27 protein was detected (Fig. 3.11, lanes 11 and 14). Furthermore, crosslinks were observed between CDC5L and SPF27 (Fig. 3.11, lanes 8 and 14) and between CTNNBL1 and AD002 (Fig. 3.11, lanes 10 and 13). These data demonstrate physical interactions between each of the crosslinked protein pairs within the hPrp19/CDC5L complex. At 10 μ M Sulfo-MBS, several high molecular weight crosslinks were observed. One crosslink appeared to contain solely hPrp19 (Fig. 3.11, lane 19), although a weak signal at this position was also observed with anti-PRL1 antibodies (Fig. 3.11, lane 18). Based on its size (\sim 200 kDa) it could represent the hPrp19 tetramer or alternatively three copies of hPrp19 plus

PRL1, as both proteins exhibit molecular masses of ~ 50 kDa. An additional band at 250 kDa was recognized by antibodies against hPrp19 and PRL1, suggesting an interaction of the hPrp19 tetramer with the PRL1 protein (Fig. 3.11, lanes 18 and 19). Finally, a very high molecular weight crosslinked complex was detected with antibodies against CDC5L, PRL1 and hPrp19 (Fig. 3.11, lanes 15, 18 and 19), and to a lesser extent by antibodies against CTNNBL1 and SPF27 (Fig. 3.11, lanes 18 and 21). Although only a faint signal can be seen with the anti-SPF27 antibodies, the overall signal for SPF27 is generally lower (Fig. 3.11, lane 21). Therefore, it is likely that this crosslinked complex corresponds to the hPrp19/CDC5L core complex alone or together with CTNNBL1. Taken together, these data confirm physical links between PRL1-SPF27, CDC5L-SPF27 and hPrp19-PRL1. In addition, the interaction between AD002 and CTNNBL1 is consistent with the presence of an AD002-CTNNBL1 binary complex in nuclear extract from the FLAG/HA-AD002 stable cell line. We were not able to identify the exact position of the observed crosslinks by MS. That is, even after protease digestion the crosslinked fragments were too large to be sequenced with a mass spectrometer.

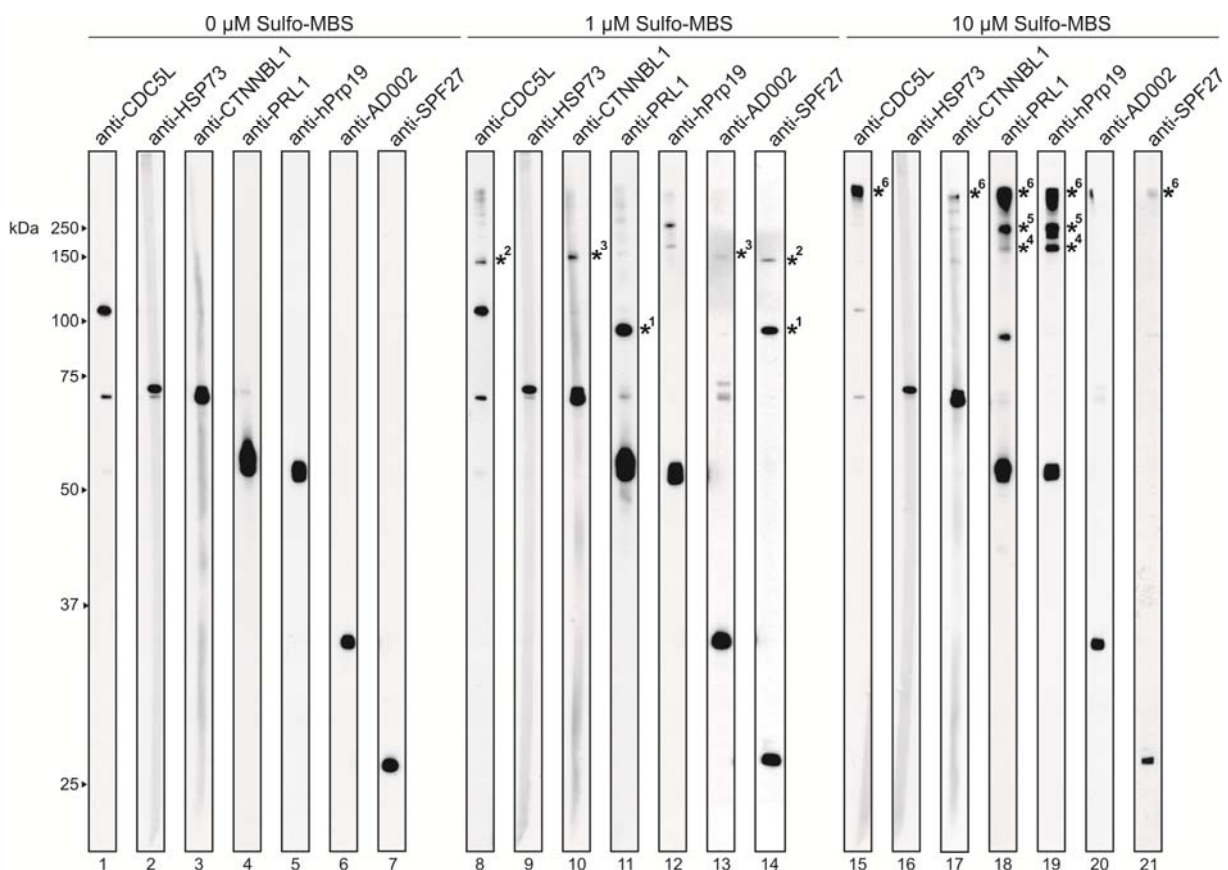


Fig. 3.11: Identification of protein-protein interactions in the hPrp19/CDC5L complex via crosslinking. Purified hPrp19/CDC5L complexes were incubated for 30 min at 25°C with coupling buffer (lanes 1-7), or 1 μ M Sulfo-MBS (lanes 8-14), or 10 μ M Sulfo-MBS (lanes 15-21). Crosslinked complexes were analyzed by western blotting with the antibodies indicated above each lane. Crosslinked proteins are indicated by an asterisk with a number, where proteins in the same crosslinked complex share the same number.

3.1.6 Analysis of protein-protein interactions between hPrp19/CDC5L components with *in vitro* translated proteins

Sulfo-MBS crosslinking in the native hPrp19/CDC5L complex could mainly identify protein-protein interactions within the salt-stable core of the complex, plus an AD002-CTNNB1 interaction. However, further interactions linking the HSP73 protein to the AD002-CTNNB1 dimer or providing information about the association of this potential trimeric complex with the stable core of the hPrp19/CDC5L complex could not be observed.

To elucidate additional protein-protein interactions in the hPrp19/CDC5L complex, we performed protein-protein interaction studies with *in vitro* translated proteins. It was not possible to translate hPrp19/CDC5L complex components using conventional *E. coli* based translation extracts (data not shown). Thus, we used a translation system derived from wheat germ that has been used widely for the efficient translation of viral and eukaryotic messenger RNAs in a heterologous cell-free system (reviewed in Endo and Sawasaki, 2003). For *in vitro* translation of a protein, the corresponding gene was cloned into the specific cell-free expression vector pEU3b (Fig. 3.12A). The vector contains a viral 5' untranslated region (UTR), the so-called omega sequence from the tobacco mosaic virus, and a 3' UTR, that were shown to result in highly efficient translation in the wheat germ translation system (Sawasaki et al., 2002). The message is supplied to the wheat germ system as *in vitro* transcribed mRNA. Therefore, the vector contains a T7-promotor. However, plasmids providing purification tags to the protein of interest were not developed so far. Therefore, we introduced a StrepII-tag (Voss and Skerra, 1997), preceded by a Factor Xa protease cleavage site (Fig. 3.12B and C) into the pEU3b vector, generating pEU3b-StrepII. Using conventional cloning procedures it was now possible to generate mRNAs coding for each hPrp19/CDC5L complex protein containing a C-terminal StrepII-tag. Using the *Not I* restriction enzyme and subsequent run-off transcription, it was also possible to generate mRNAs without the StrepII-tag, thereby generating StrepII-tagged and untagged proteins from the same plasmid (Fig. 3.12B). Initially, the translation system was optimized in terms of incubation time and mRNA concentration, indicating optimal translation efficiency at mRNA concentrations of 0.1-0.2 $\mu\text{g}/\mu\text{l}$ after 4 hrs incubation (data not shown). As shown in Fig. 3.13A, all hPrp19/CDC5L complex proteins could be produced with the wheat germ cell-free translation system.

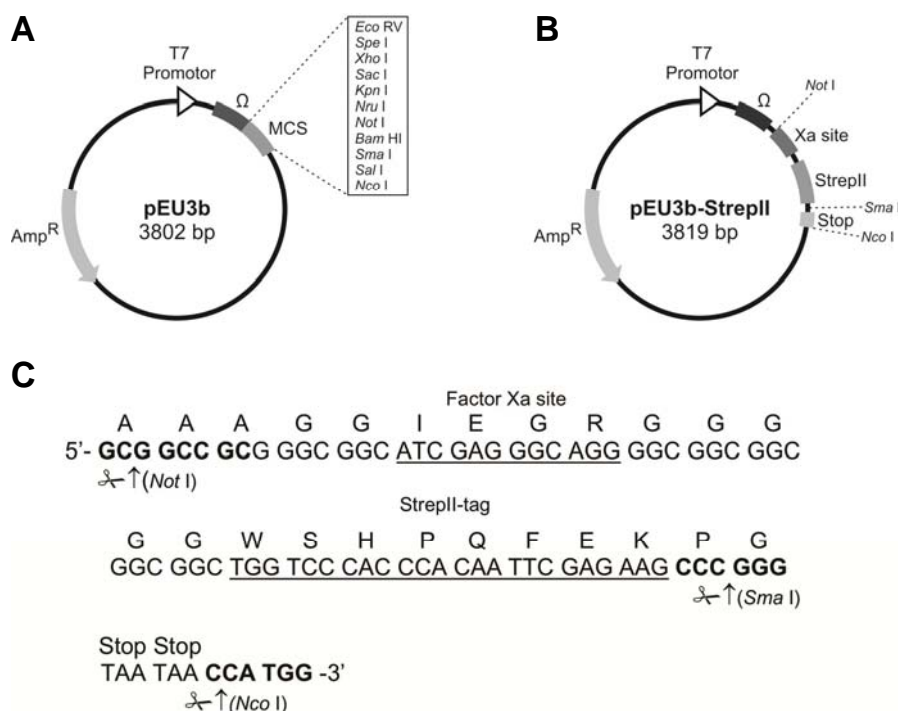


Fig. 3.12: Plasmids used for the cell-free translation of proteins in wheat germ lysate. Schematic illustration of (A) the pEU3b vector (Sawasaki et al., 2002) and (B) the pEU3b-StrepII vector. The main features are shown in grey boxes and the restriction sites are indicated with dotted lines. Amp^R – Ampicillin resistance gene; MCS – Multiple cloning site; Stop – Termination codon; Ω - Omega sequence. (C) Nucleic acid (bottom) and the resulting protein sequence (single letter code, top) that was introduced into pEU3b to generate pEU3b-StrepII. The sequence contains a StrepII affinity purification tag and a Factor Xa protease cleavage site to potentially remove the tag from the expressed protein. The main restriction sites are shown in bold and indicated with arrows.

To detect additional protein-protein interactions, we first performed far western analyses. Thus, native hPrp19/CDC5L complex proteins were separated by SDS-PAGE, transferred to nitrocellulose (Fig. 3.13B, lane 1) and then incubated with *in vitro* translated, [³⁵S]-labeled proteins (Fig. 3.13A). Consistent with the aforementioned results, at 300 mM salt [³⁵S]-labeled SPF27 interacted with CDC5L and hPrp19 (Fig. 3.13B, lane 8). Reciprocally, [³⁵S]-labeled CDC5L and hPrp19 interact with SPF27 (Fig. 3.13B, lanes 1 and 6). Furthermore, [³⁵S]-labeled CDC5L also interacted with hPrp19 and vice versa (Fig. 3.13B, lanes 1 and 6). In addition to these strong interactions, we also detected interactions between CTNNBL1 and AD002, and also between CTNNBL1 and CDC5L, which were only observed at 150 mM salt (Fig. 3.13B, lane 11). This provides a physical link between AD002/CTNNBL1 and a protein of the hPrp19/CDC5L complex core. Similar to previously published results (Ajuh et al., 2001), we also detected an interaction between PRL1 and CDC5L at 150 mM, but not at 300 mM NaCl (compare Fig. 3.13B, lanes 5 and 12). Finally, CDC5L interacted with itself, but only at 150 mM salt (Fig. 3.13B, lane 9). As only one copy of CDC5L is present in our hPrp19/CDC5L complexes, CDC5L dimerization and CDC5L association with the hPrp19 complex appear to be mutually exclusive.

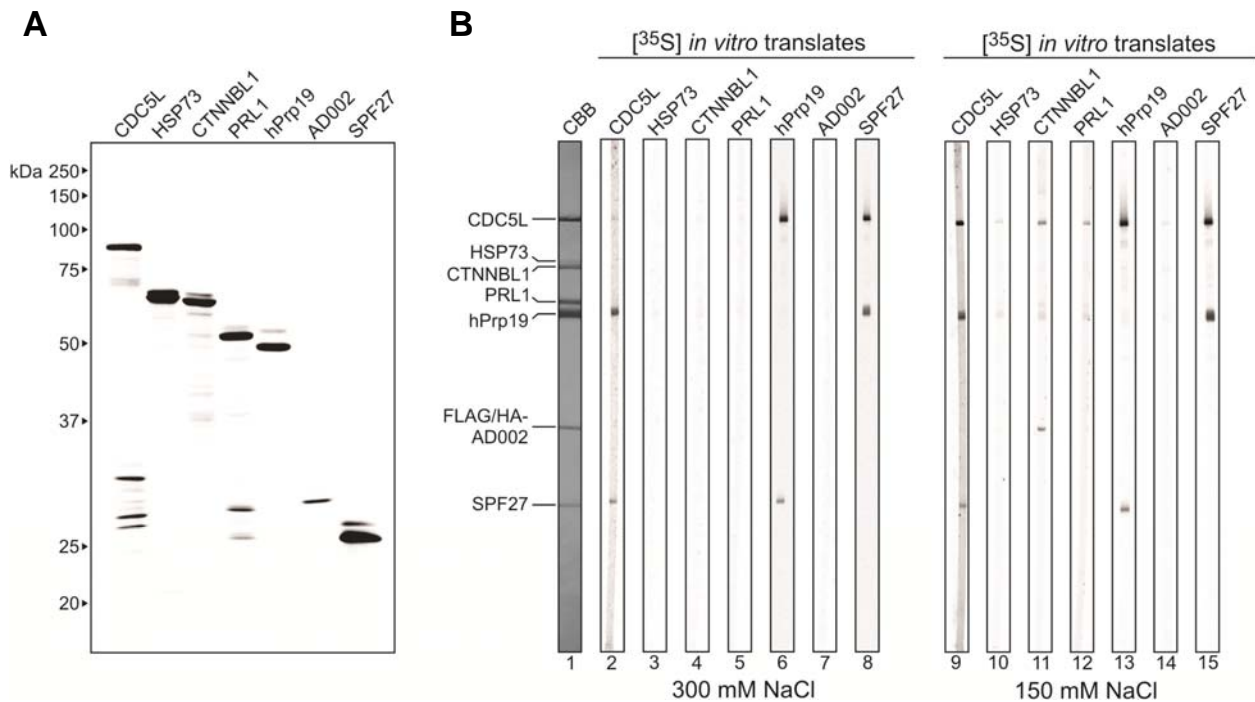


Fig. 3.13: Far western analysis demonstrates interactions among the hPrp19/CDC5L complex core components and between AD002 and CTNNBL1, and CTNNBL1 and CDC5L. (A) SDS-PAGE analysis of *in vitro* translated [^{35}S]-labeled proteins (as indicated above each lane). Proteins were visualized by autoradiography. The molecular mass (kDa) of marker proteins is indicated on the left. (B) Far western overlay analysis of protein-protein interactions between individual hPrp19/CDC5L complex proteins. SDS-PAGE separated hPrp19/CDC5L complex proteins were blotted to nitrocellulose stripes and visualized by staining with Coomassie (lane 1), or were incubated with [^{35}S]-labeled, *in vitro* translated hPrp19/CDC5L complex proteins in buffer containing 300 mM (lanes 2-8) or 150 mM NaCl (lanes 9-15) and visualized by autoradiography (lanes 2-15). hPrp19/CDC5L complex proteins are indicated on the left.

Interaction partners in the hPrp19/CDC5L complex were also analyzed by *in vitro* pulldown assays. For this purpose, AD002, SPF27 and hPrp19 containing a StrepII-tag were translated *in vitro* and incubated with *in vitro* translated [^{14}C]-labeled CDC5L, hPrp19, or CTNNBL1. Proteins containing the StrepII-tag were then precipitated with StrepTactin-Sepharose and coprecipitating proteins were analyzed by SDS-PAGE. As shown in Fig. 3.14, we could confirm interactions between AD002 and CTNNBL1 (lane 3), SPF27 and CDC5L (lane 7), SPF27 and hPrp19 (lane 9), as well as CDC5L and hPrp19 (lane 12). No interaction with the StrepTactin-Sepharose alone was observed with radiolabeled CDC5L, hPrp19 or CTNNBL1 (lanes 1, 4, 5 and 10).

A summary of interactions within the hPrp19/CDC5L complex is shown in Fig. 3.15. Taken together, these data indicate a strong interaction network among proteins of the hPRP19/CDC5L complex, in particular among proteins comprising its highly stable core. However, based on our current data, it is still unclear how HSP73 associates with the hPrp19/CDC5L complex.

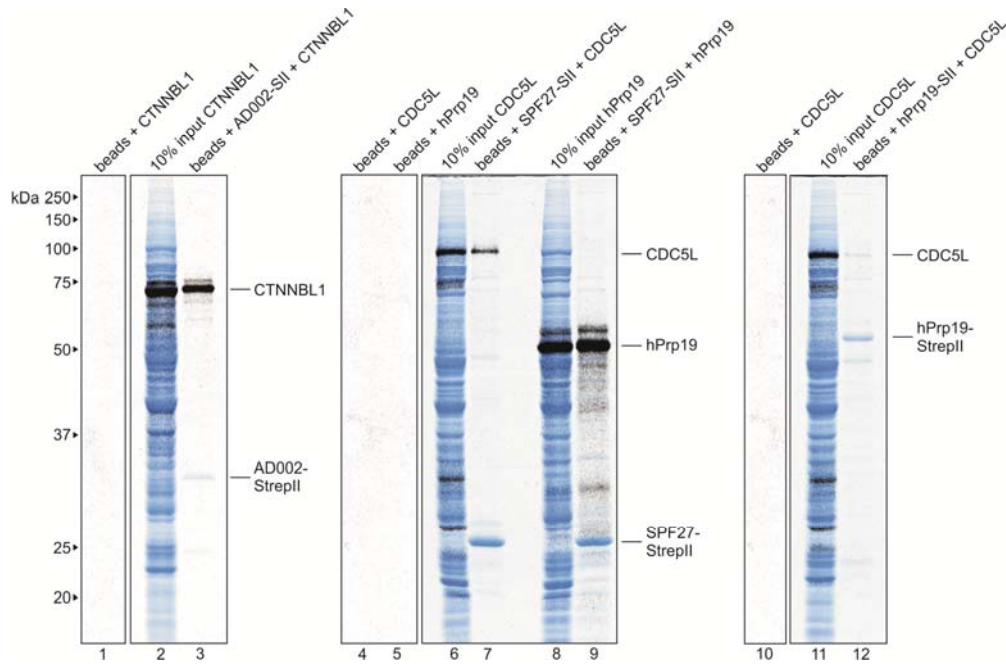


Fig. 3.14: Co-immunoprecipitation confirms interactions between SPF27 and both CDC5L and hPrp19, between CTNNBL1 and AD002, and between CDC5L and hPrp19. [¹⁴C]-labeled, *in vitro* translated CTNNBL1, CDC5L and hPrp19, as indicated above each lane, were incubated alone (lanes 1, 4, 5, 10) or with StrepII-tagged AD002 (lane 3), SPF27 (lanes 7, 9) or hPrp19 (lane 12) and pulldowns were subsequently performed with Strep-Tactin-Sepharose beads. Coprecipitated proteins or 10% of the input (lanes 2, 6, 8, 11) were analyzed by SDS-PAGE and visualized by autoradiography (black). The presence of the StrepII-tagged protein was confirmed by Coomassie staining (blue). The identity of the co-precipitated proteins is indicated on the right. The molecular mass (kDa) of marker proteins is indicated on the left.

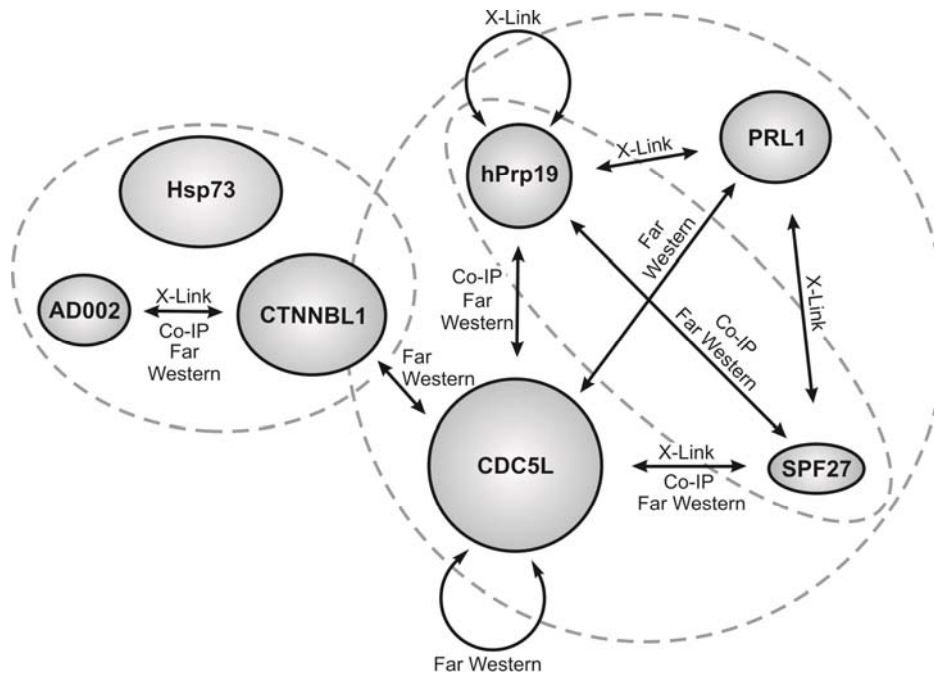


Fig. 3.15: Schematic overview of the protein-protein interactions within the hPrp19/CDC5L complex. Interactions identified in this study are indicated by an arrow together with the methods demonstrating a given interaction. The dashed circles enclose those proteins that were found in stable subcomplexes. Co-IP – Co-immunoprecipitation; X-Link – Crosslinking.

3.1.7 Limited proteolysis of the native hPrp19/CDC5L complex delineates protein interaction domains

As summarized in Fig. 3.15, we were able to identify various protein-protein interactions within the hPrp19/CDC5L complex. Furthermore, dissociation of the hPrp19/CDC5L complex with salt and heparin identified a stable core of the complex. To further probe the structural organization in the native hPrp19/CDC5L complex and identify contact regions between the various proteins, we next performed limited proteolysis.

During limited proteolysis the digestion reaction is restricted and protease activity is preferentially restricted to a few, specific peptide bonds. Conformational parameters, such as accessibility and solvent exposition correlate quite well with limited proteolytic sites (Fontana et al., 1997; Price and Johnson, 1989). As the main structural feature of interacting proteins in a complex is that the interface regions are inaccessible for the solvent and therefore also for the protease, limited proteolysis and the subsequent identification of protease-resistant fragments by MS offer an experimental route for the determination of exact boundaries of the interacting domains.

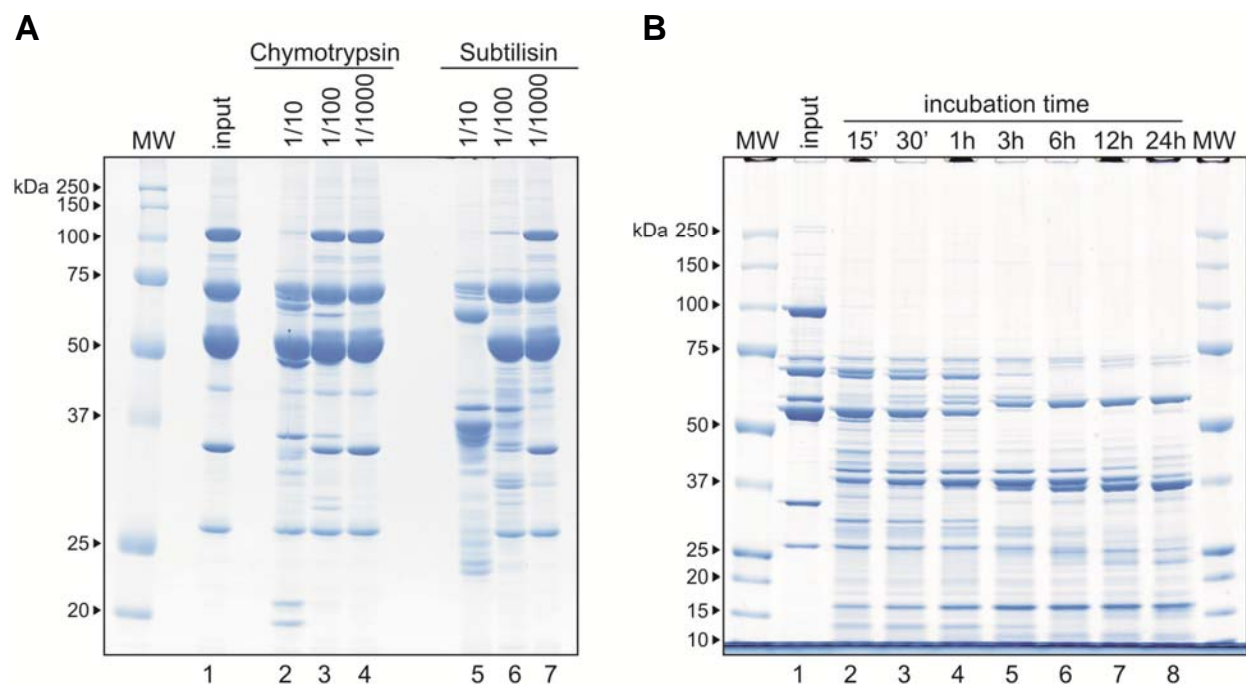


Fig. 3.16: Limited proteolysis of the hPrp19/CDC5L complex. (A) Purified complexes were treated with different concentrations of the proteases Subtilisin (stock concentration 0.5 mg/ml) and Chymotrypsin (stock concentration 0.5 mg/ml) for 30 min on ice. The digested proteins were separated by SDS-PAGE and stained with Coomassie. The position of the proteins is indicated on the left and the molecular mass (kDa) of marker proteins is indicated on the left. (B) Kinetic of Subtilisin digestion of purified hPrp19/CDC5L complexes. Aliquots were taken at different incubation times, separated by SDS-PAGE and stained with Coomassie. The position of the proteins is indicated on the left and the molecular mass (kDa) of marker proteins is indicated on the left. MW - molecular weight marker proteins.

In an initial experiment, the reaction conditions were carefully selected in order to ensure stability of the hPrp19/CDC5L complex and activity of the protease used. In addition to the choice of the protease itself, the pH value, digestion time and incubation temperature, an important parameter in a limited proteolysis experiment is the enzyme to substrate ratio. Therefore, we initially tested two different proteases, Subtilisin and Chymotrypsin, over a broad concentration range (Fig. 3.16A). Both proteases were chosen, because of their broad substrate specificity and undemanding buffer requirements for enzymatic activity. Chymotrypsin selectively catalyzes the hydrolysis of peptide bonds on the C-terminal side of Tyr, Phe, Trp and Leu. However, some significant cleavage will also occur on the C-terminal side of Met, Ile, Ser, Thr, Val, His, Gly and Ala (Sweeney and Walker, 1993). Subtilisin is a non-glycosylated single protease without disulfide bonds and of a molecular weight of 27 kDa. It is a member of the Serine S8 Endoproteinase family and has broad specificity similar to Chymotrypsin with a preference for aromatic side chains in the P1 position, the position preceding the cleavage site (Schechter and Berger, 1967, 1968), but will also accept non-aromatic apolar side chains and even charged side chains at P1, when the substrate residue at P3 or P4 is hydrophobic (Donlon, 2007; Kraut, 1977). As shown in Fig. 3.16A, the digestion pattern of Subtilisin and Chymotrypsin was very similar (compare lanes 2 and 6), indicating that independent from the protease specificity the accessible regions of the polypeptide chains are cleaved. Because Subtilisin showed a higher activity compared to Chymotrypsin, it was used for the further experiments. After further optimization of the incubation time (Fig. 3.16B), we incubated affinity-purified, gel filtered hPrp19/CDC5L complexes for 30 min with 30 nM Subtilisin and fractionated the proteolysed components via gel filtration (Fig. 3.17A). SDS-PAGE analysis of the gradient fractions identified a large protease resistant complex (Fig. 3.17B, peak 1). Exhaustive LC-offline MALDI tandem MS was performed to delineate regions of each protein present in this complex (Fig. 3.18) (performed in collaboration with the Bioanalytical Mass Spectrometry group of Dr. Henning Urlaub, MPI for Biophysical Chemistry). MALDI is usually coupled to TOF analyzers that measure the mass of intact peptides. However, the combination of two TOF analyzers interconnected by a collision cell also permits peptide sequencing. The resulting fragment ion spectra are often more extensive and informative than those generated in a trapping instrument. Furthermore, ionization from a crystalline matrix allows a sample to be measured again at a later time, thereby providing very high peptide coverage throughout the whole protein sequence. Therefore, this technique is particularly suitable for sequencing a protein sample, e.g. after protease digestion.

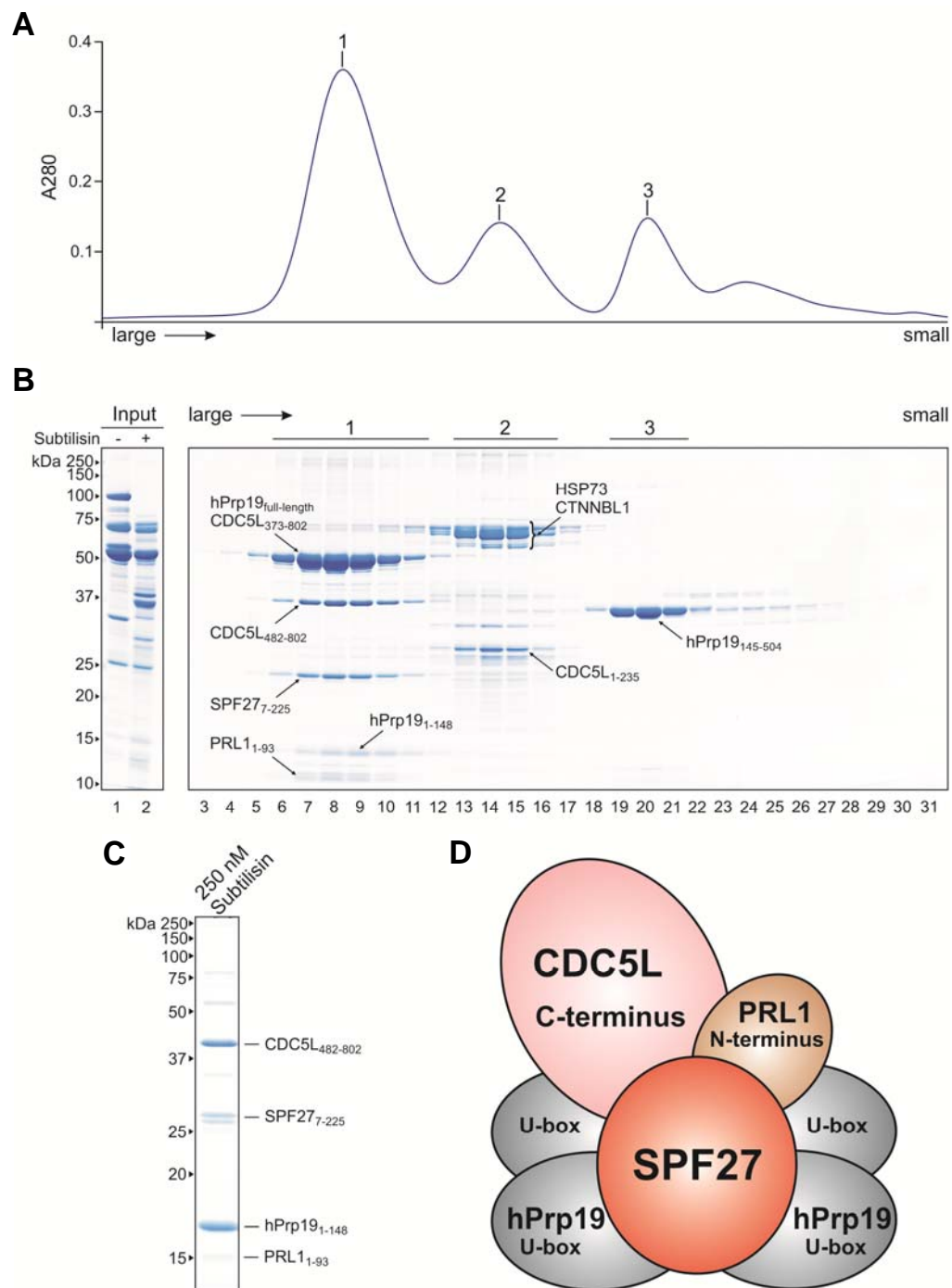


Fig. 3.17: Limited protease digestion delineates those domains of hPrp19/CDC5L complex proteins involved in protein-protein interactions. Purified hPrp19/CDC5L complexes were incubated for 30 min at 0°C with 30 nM Subtilisin and the proteolysed fragments were fractionated by size via gel filtration. **(A)** Three protease resistant complexes can be separated by gel filtration. The absorption at 280 nm of proteins eluting from the gel filtration column was plotted against time. **(B)** SDS-PAGE analysis of the undigested hPrp19/CDC5L complex (lane 1) or the protease-treated complex prior to gel filtration (lane 2) or after gel filtration fractionation (lanes 3-31). Exhaustive LC offline tandem MS was performed to identify protein fragments in the protease resistant complexes. Proteins and their protease resistant amino acids delineated by MS are indicated. The molecular mass (kDa) of marker proteins is indicated on the left. MS analysis was performed in collaboration with the Bioanalytical Mass Spectrometry group of Dr. Henning Urlaub. **(C)** Composition of the largest protease resistant complex observed after incubation with 250 nM Subtilisin. **(D)** Schematic model of interaction domains in the protease resistant core of the hPrp19/CDC5L complex.

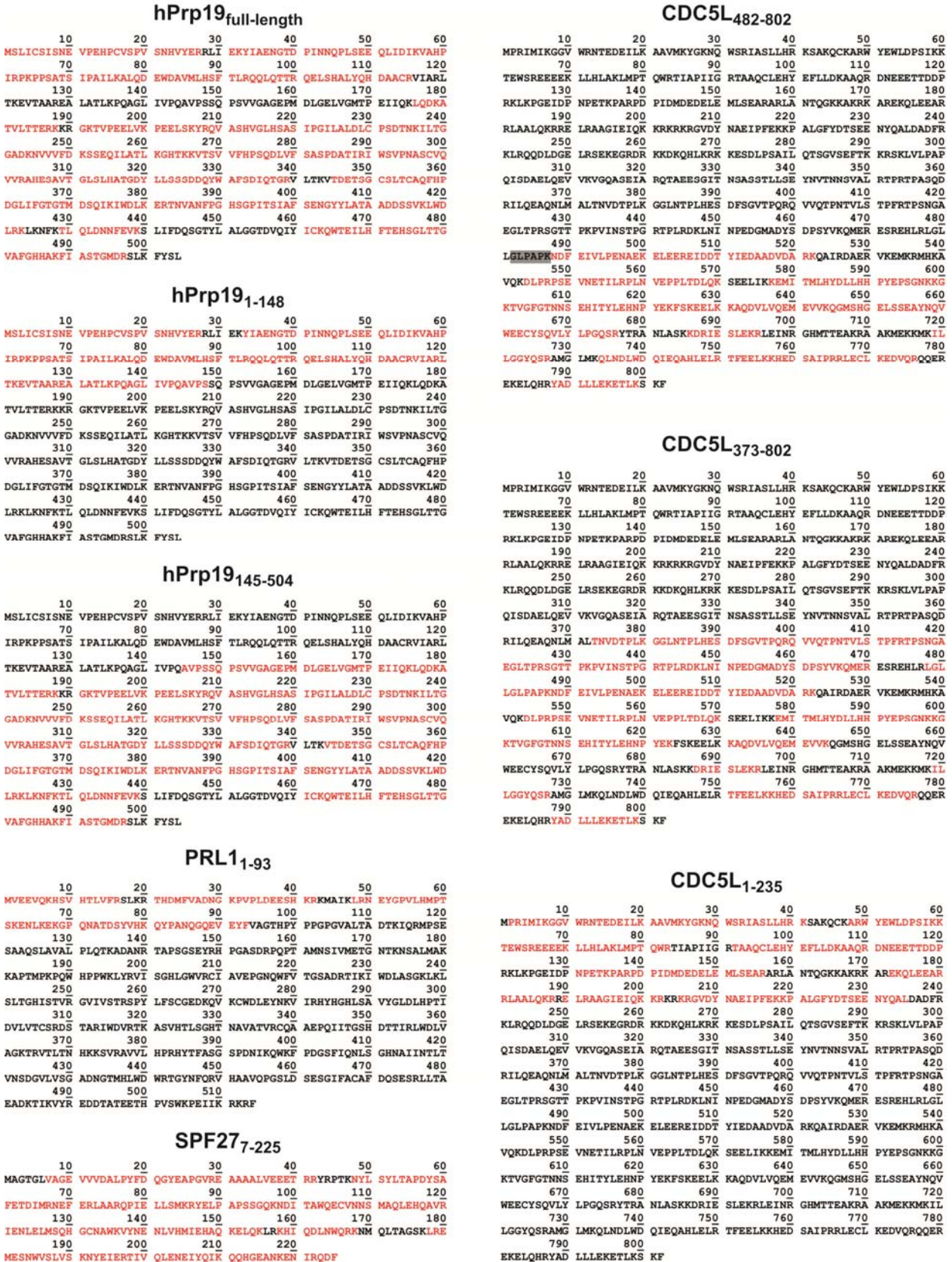


Fig. 3.18: Exhaustive MS analyses of hPrp19/CDC5L complex proteins after limited proteolysis and subsequent trypsin digestion. The peptides identified via MALDI-MS/MS are highlighted in red. The presence of the N-terminal peptide of CDC5L (482-802), indicated with a grey box, was verified by N-terminal Edman sequencing. MS analysis was performed in collaboration with the Bioanalytical Mass Spectrometry group of Dr. Henning Urlaub (MPI for Biophysical Chemistry).

Nearly full-length SPF27 (i.e., lacking only the N-terminal most six amino acids) and full-length hPrp19, together with an N-terminal domain of hPrp19 that contains a U-box domain (aa 1-148) were identified in the largest complex. An overview of predicted domains in the hPrp19/CDC5L complex proteins was already shown in Fig. 1.7B. Two fragments of the CDC5L protein, both consisting of the C-terminal half of the protein (aa 373-802 and 482-802, respectively), were also detected; the precise N-terminus of the latter could be confirmed by Edman sequencing. Furthermore, the very N-terminus (aa 1-93) of the PRL1 protein could be identified by MS. Thus, consistent with our salt dissociation data, fragments of these proteins form a protease resistant core complex.

In addition, a second smaller complex comprised of full-length HSP73 and CTNNBL1 (and slightly truncated versions thereof) was identified after Subtilisin treatment of the hPrp19/CDC5L complex (Fig. 3.17A and B, peak 2). A 150 kDa marker protein also peaked at this position in a separate gradient (data not shown), indicating that HSP73 and CTNNBL1 are most likely present in this peak in the form of a complex. Interestingly, we could also identify an N-terminal fragment of CDC5L (aa 1-235) co-migrating with these proteins in the same peak. This is consistent with our far western blot analysis, where CDC5L interacts with CTNNBL1, and indicates that the AD002-CTNNBL1 dimer binds the hPrp19/CDC5L complex core via the N-terminus of CDC5L. The third peak solely contains a 35 kDa fragment of the hPrp19 protein (Fig. 3.17A and B, peak 3). MS analysis demonstrated that it is the C-terminal part of hPrp19 (aa 145-504), containing the WD40 domain of the protein (compare with Fig. 1.7B).

To analyze the protease stability of the hPrp19/CDC5L complex in more detail, we treated the complex with increasing amounts of Subtilisin (Fig. 3.19) and the proteolysed components were then fractionated via gel filtration and analyzed by SDS-PAGE (Fig. 3.19A and B). MS analysis showed that the largest protein resistant complex, after treatment with 125 nM Subtilisin, contains a similar set of protein fragments as the 30 nM Subtilisin treated complex. However, the complex elutes from the gel filtration column at a later time point (Fig. 3.19A), indicating that it is smaller. Indeed, Coomassie staining intensities (Fig. 3.19B) suggest that the amounts of full-length hPrp19 and CDC5L (aa 373-802) are largely reduced. Simultaneously, the protein bands of the N-terminal region of hPrp19 (aa 1-148) and the C-terminal half of CDC5L (aa 482-802) appear more intensively stained. Thus, these fragments of hPrp19 and CDC5L appear to be highly resistant against protease cleavage. Interestingly, also the C-terminal part of hPrp19 (aa 145-504) accumulates in the third peak (Fig. 3.19B) and is still stable after treatment with 250 nM Subtilisin, as indicated by an increased absorption intensity (Fig. 3.19A), as well as an increased Coomassie staining intensity of the corresponding protein band in the SDS gel (Fig. 3.19B).

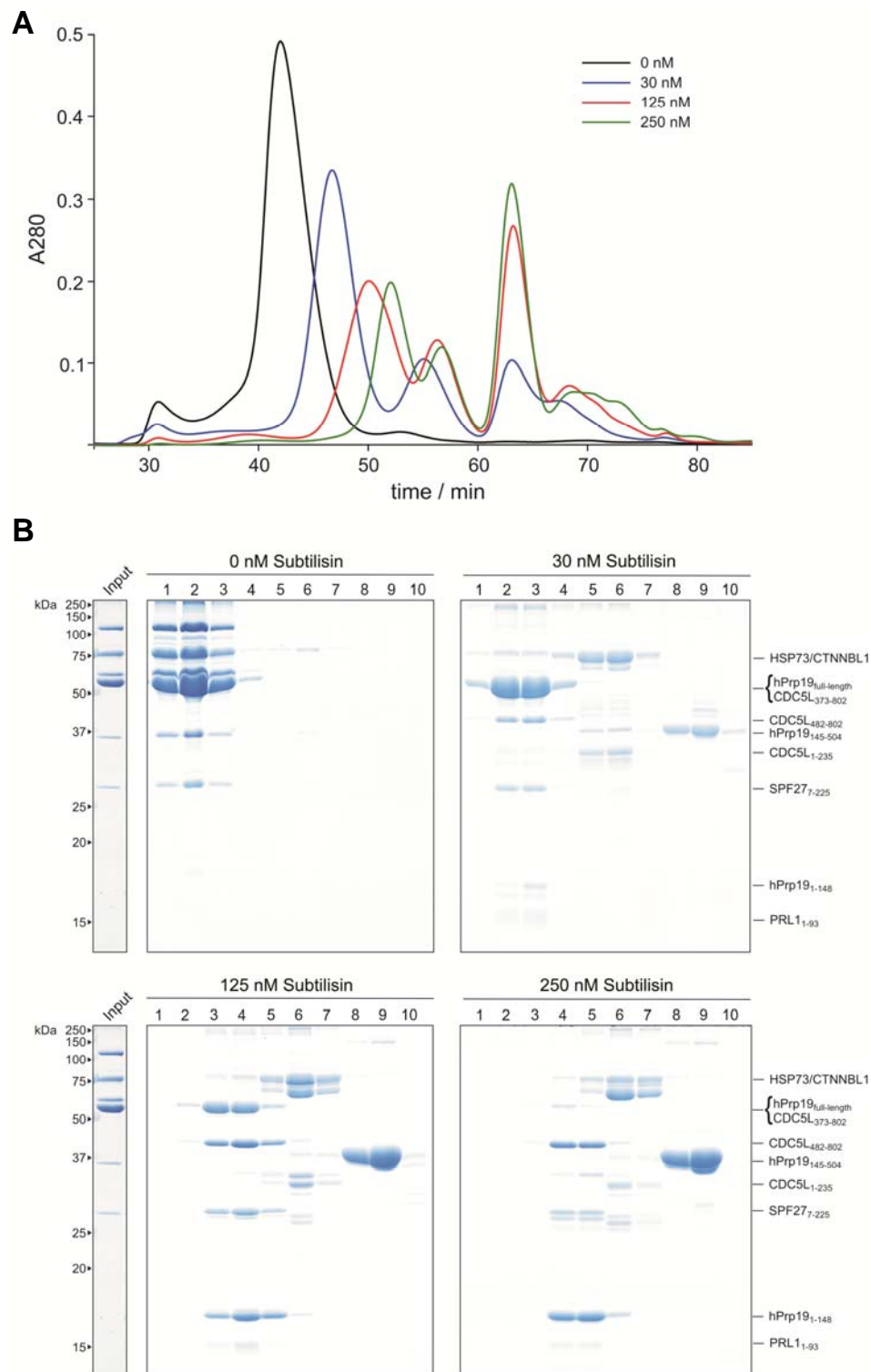


Fig. 3.19: Limited proteolysis of the hPrp19/CDC5L complex with increasing amounts of Subtilisin. Affinity-selected, gel-filtrated complexes were incubated for 30 min at 0 °C with 0, 30 nM, 125 nM and 250 nM of the protease Subtilisin. The proteolysed fragments were fractionated by size via gel filtration. **(A)** Gel filtration profile after proteolysis with three different concentrations of Subtilisin, as compared to the untreated hPrp19/CDC5L complex. The Subtilisin concentrations are indicated. **(B)** SDS-PAGE analysis of Subtilisin treated hPrp19/CDC5L complex at three different concentrations of the protease compared to the untreated complex after gel filtration. The Subtilisin concentration is indicated above each gel. Protein bands were analyzed by MS and the identity of the protein bands is indicated on the right. MS analysis was performed in collaboration with the Bioanalytical Mass Spectrometry group of Dr. Henning Urlaub.

When Subtilisin was increased to 250 nM, only the N-terminal region of hPrp19 (aa 1-148), containing the U-box domain, was present in the largest protease resistant complex, together with an N-terminal fragment of PRL1 (aa 1-93) and a C-terminal fragment of CDC5L (aa 482-802) (Fig. 3.17C and Fig. 3.19B). This complex also contained nearly full-length SPF27 lacking only the N-terminal most six amino acids. Notably, *in vitro* translated SPF27 on its own is completely digested by 250 nM Subtilisin (data not shown). Thus, SPF27 is highly resistant to protease digestion within the hPrp19/CDC5L complex, apparently due to its interaction with all other core components. The staining intensity of the hPrp19 N-terminal fragment further indicates that multiple copies of it are present in the protease-resistant core of the hPrp19/CDC5L complex. Indeed, those amino acids known to mediate hPrp19 self-interaction (i.e., aa 56-74) (Grillari et al., 2005) are still present.

Taken together, these data suggest that the N-terminal fragment of the hPrp19 protein is present as a tetramer in the protease resistant core of the hPrp19/CDC5L complex together with SPF27, the N-terminus of PRL1 and the C-terminus of CDC5L (Fig. 3.17D). The entire topology of the hPrp19/CDC5L complex is summarized in Fig. 3.20. In this model, the U-box containing domain of hPrp19, which forms the hPrp19 tetramer (Grillari et al., 2005; Vander Kooi et al., 2006), interacts together with SPF27, the N-terminus of PRL1 and the C-terminus of CDC5L, forming the stable core of the hPrp19/CDC5L complex. In addition, the potential trimeric subcomplex composed of AD002, CTNNBL1 and HSP73 is associated with the core complex via the interaction of CTNNBL1 with the N-terminus of CDC5L.

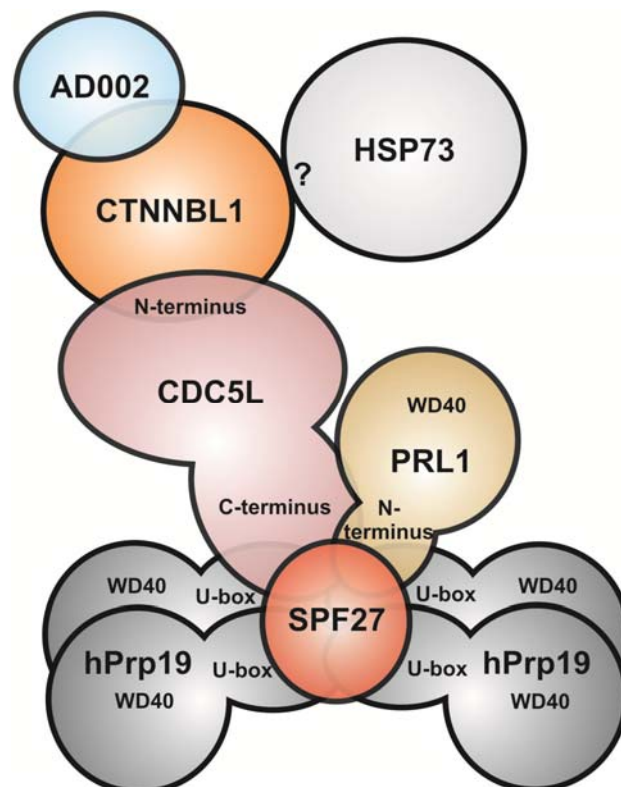


Fig. 3.20: Schematic representation of the topology of the hPrp19/CDC5L complex.

3.1.8 Electron microscopic analysis of the human Prp19/CDC5L complex

The number of proteins of the splicing machinery that are required for maintaining the catalytically active RNP network of the human C complex consists of about 35-40 proteins (Bessonov et al., 2008), including nearly all subunits of the hPrp19/CDC5L complex. Thus, the hPrp19/CDC5L complex is a major component of the catalytically active center of the step I spliceosome. Structure determination of this complex by EM may in part provide a first low resolution view into the possible organization of part of the spliceosome's catalytic core.

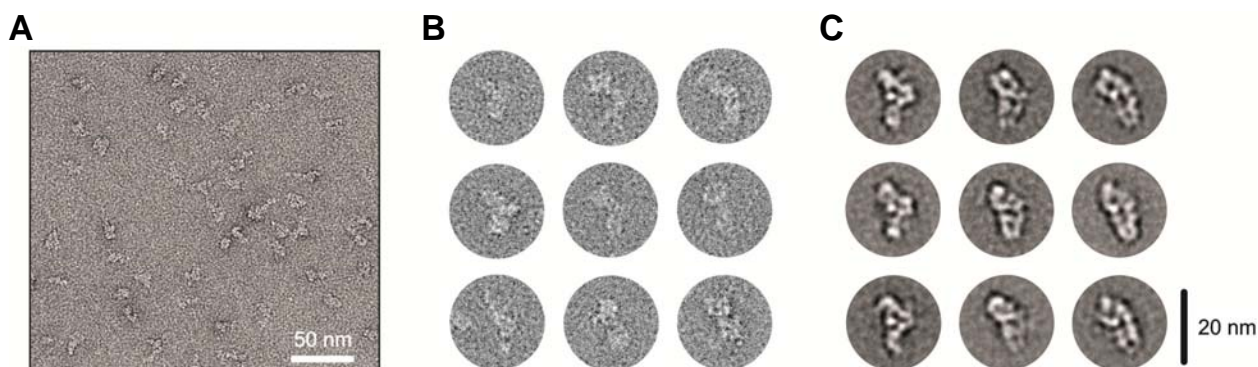


Fig. 3.21: Electron microscopy of native human Prp19/CDC5L complexes. (A) Electron micrograph of negatively-stained hPrp19/CDC5L complexes. The images were taken at a primary magnification of 105,000 under low-dose conditions. Bar, 50 nm. (B) A representative selection of individual raw images. (C) A gallery of individual class averages of native hPrp19/CDC5L complexes obtained by averaging individual raw images. Bar, 20 nm. EM analysis of hPrp19/CDC5L complexes was performed by Elmar Wolf (Cellular Biochemistry group, MPI for Biophysical Chemistry).

Given its size of approximately 600 kDa and its homogeneous composition, the hPrp19/CDC5L complex is well suited for structural analysis by EM. To determine the structure of native hPrp19/CDC5L complexes, FLAG affinity-purified particles were subjected to GraFix (Kastner et al., 2008), negatively stained with uranyl formate and analyzed by EM. EM analysis of hPrp19/CDC5L complexes was performed by Dr. Elmar Wolf (Cellular Biochemistry group, MPI for Biophysical Chemistry) in collaboration with the 3D Electron Cryo-Microscopy group of Prof. Dr. Holger Stark at the MPI. The raw images (Fig. 3.21A and B) reveal a homogeneous, well-defined population of complexes similar in size and shape. 2D class averages of the hPrp19/CDC5L complex were obtained by averaging individual images after “reference-free” image alignment, MSA, and classification (Fig. 3.21C). The most frequent class averages reveal an elongated, asymmetric shape with a maximum dimension of ~ 20 nm and a width of ~ 9 nm. This is consistent with our sedimentation velocity data that showed that the hPrp19/CDC5L complex has a frictional ratio value of 2.1, indicating an elongated shape for the complex. Additionally, some class averages appear triangular and branched with a slightly reduced maximal length (Fig. 3.21B and C, first column); these likely represent different views of the hPrp19/CDC5L complex, as opposed to

structural heterogeneity in the complex. Stoichiometric analyses of the hPrp19/CDC5L complex showed that the hPrp19/CDC5L complex contains four copies of the hPrp19 protein (Fig. 3.6). However, the presence of single copies of all other hPrp19/CDC5L complex subunits indicates an asymmetric organization of the fully assembled complex, despite the presence of a tetrameric hPrp19 protein that is suggestive of a high symmetry. The asymmetric shape of the hPrp19/CDC5L complex is consistent with an asymmetric assembly pathway, as discussed below.

3.1.9 Establishing stable cell lines expressing an eGFP-AD002 fusion protein

It was shown previously that the hPrp19/CDC5L complex binds the spliceosome during B complex formation and it stably interacts first during catalytic activation of the spliceosome (Bessonov et al., 2008). However, little is known about the remodeling events taking place during catalytic activation of the spliceosome. To determine the structural organization of the hPrp19/CDC5L complex within the spliceosome, fluorescently labeled proteins can be used to directly measure their interaction with other proteins via fluorescence resonance energy transfer (FRET). This should lead to information about the molecular neighbor of a given protein in the nm range and therefore specify its relative position within the spliceosome.

As the hPrp19/CDC5L complex is part of the catalytic core of the spliceosome, mapping the position of a component of the hPrp19/CDC5L complex within the EM structure of the spliceosome would provide important information about the localization of its catalytic core. In yeast, LSm proteins were topographically tagged with the 30 kDa protein yECitrine and used for localizing the LSm protein positions in the LSm ring of the U6 snRNP (Karaduman et al., 2008). eGFP is of similar size as yECitrine and thus could potentially be used for both localization of the hPrp19/CDC5L complex in the EM structure of a human spliceosome, as well as fluorescence studies.

To be able to perform such experiments in the future, we established a HeLa cell line that stably expresses an eGFP-AD002 fusion protein that additionally carries a tandem affinity purification tag consisting of a FLAG and two sequential copies of the StrepII-tag (Voss and Skerra, 1997) (Fig. 3.22A). For characterization of the stable cell line, proteins were first separated by SDS-PAGE and transferred to a nitrocellulose membrane. Subsequent western blotting with antibodies against the AD002 protein or the FLAG epitope confirmed the presence and overexpression of the tagged protein in HeLa nuclear extract prepared from these cell lines (Fig. 3.22B). Western blotting was also performed with antibodies against eGFP and the StrepII-tag, which confirmed the aforementioned results (data not shown). Interestingly, as was already shown for the FLAG/HA-AD002 and FLAG/HA-SPF27 cell lines (Fig. 3.1B), the concentration of endogenous AD002 is reduced also in cells overexpressing eGFP-AD002, compared to wildtype cells (Fig. 3.22B, upper

panel). Furthermore, the overall protein concentration in nuclear extracts prepared from the control or stable cell lines is similar, as was shown by immunoblotting with anti-PRL1 antibodies (Fig. 3.22B, lower panel). Additionally, western blotting with anti-PRL1 antibodies suggests that the overall concentration of hPrp19/CDC5L complexes is the same in the control or the stable cell line.

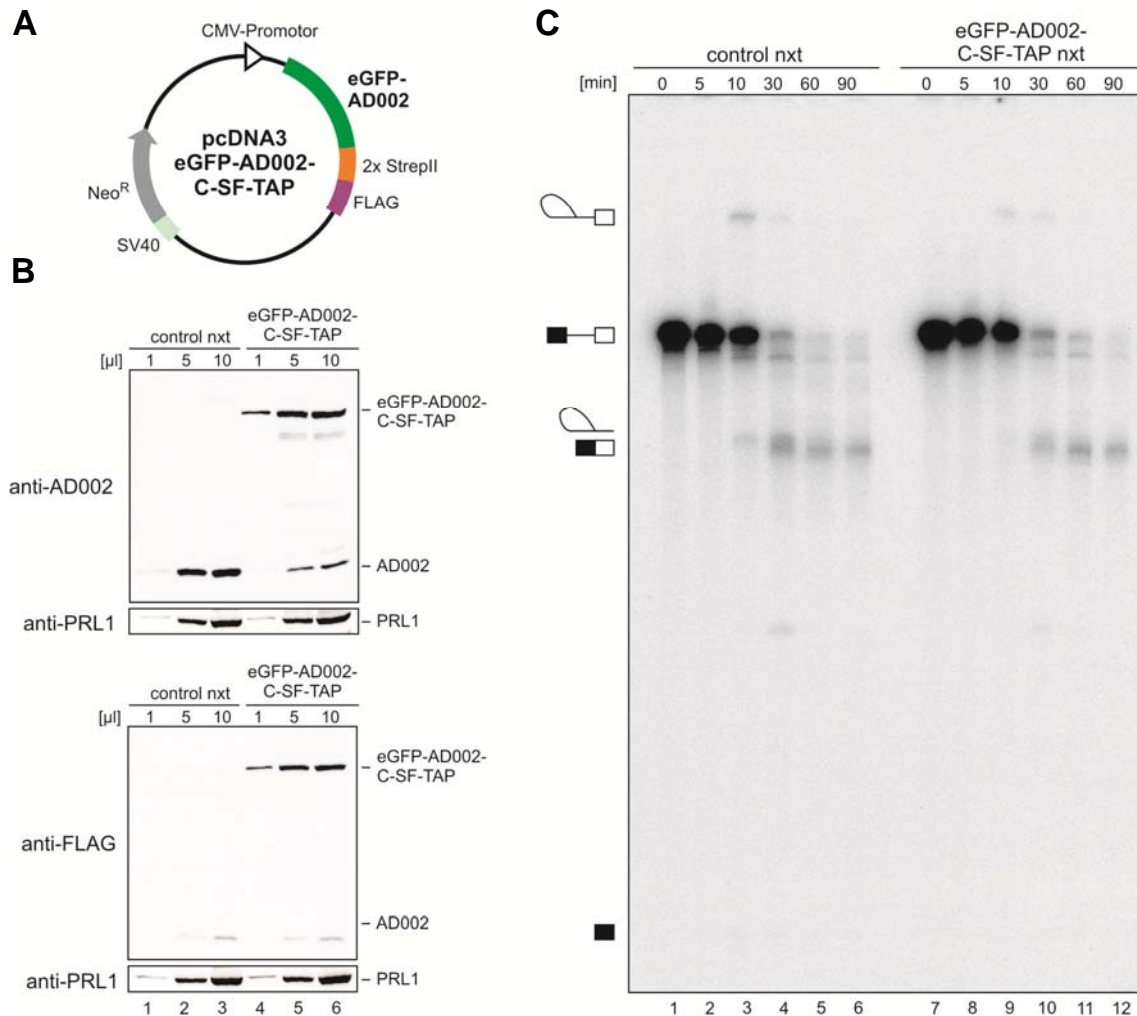


Fig. 3.22: Characterization of HeLa cell lines stably expressing FLAG/StrepII-tagged eGFP-AD002. (A) Plasmid that was used for the stable transfection of HeLa cells. The plasmid contains the FLAG/StrepII-tagged eGFP-AD002 open reading frame under the control of a cytomegalovirus (CMV) promoter. Alternatively, HeLa cells were transfected with a plasmid carrying only the antibiotic resistance but no protein-coding gene (control, not shown). SV40 – SV40 promoter; Neo^R – Neomycin resistance gene; SF – StrepII/FLAG; TAP – Tandem affinity purification. (B) Nuclear extract (nxt) was prepared from eGFP-AD002 or control cell lines, as indicated above. Proteins were recovered from increasing amounts of nuclear extract (as indicated above each lane) and separated by SDS-PAGE. Proteins were transferred to nitrocellulose membrane and further analyzed by western blotting with anti-AD002 (lanes 1-6, upper panel) or anti-FLAG antibodies (lanes 1-6, lower panel). As a loading control, all membranes were also probed with anti-PRL1 antibodies (lanes 1-6, below upper/lower panel). The position of the endogenous and the FLAG/StrepII-tagged protein is indicated on the right of each panel. (C) The presence of a FLAG/StrepII- and eGFP-tag does not affect pre-mRNA splicing. Splicing was performed with [³²P]-labeled MINX pre-mRNA and HeLa nuclear extract prepared from the control (lanes 1-6) or eGFP-AD002 cell line (lanes 7-12) for 0 to 90 min as indicated. RNA was recovered and separated on a 8.3M urea-10% polyacrylamide gel. The [³²P]-labeled pre-mRNA and splicing intermediates or products were detected by autoradiography, and their positions are indicated on the left.

Next, we determined whether overexpression of eGFP-AD002 influences the efficiency of pre-mRNA splicing *in vitro*. As shown in (Fig. 3.22C), nuclear extract from the cell line expressing eGFP-AD002 exhibited the same splicing activity as the wildtype extract. To further analyze the functionality of the tagged proteins, we determined the localization of eGFP-AD002 in HeLa cells by direct immunofluorescence. As shown in Fig. 3.23 (left panel), staining was restricted to the nucleus. Further confirmation of the nuclear localization was obtained by localizing the SC35 protein by indirect immunofluorescence using anti-SC35 antibodies (Fig. 3.23, middle panel). The SC35 protein is commonly used as a marker protein for nuclear speckles, which are subnuclear storage places for snRNPs and additional splicing factors (Lamond and Earnshaw, 1998). Significantly, eGFP-AD002 colocalized with SC35 in nuclear speckles (Fig. 3.23, right panel). Thus, these data indicate that eGFP-AD002 specifically localizes in the nuclear speckles, consistent with the idea that it potentially functions in pre-mRNA splicing process *in vivo*.

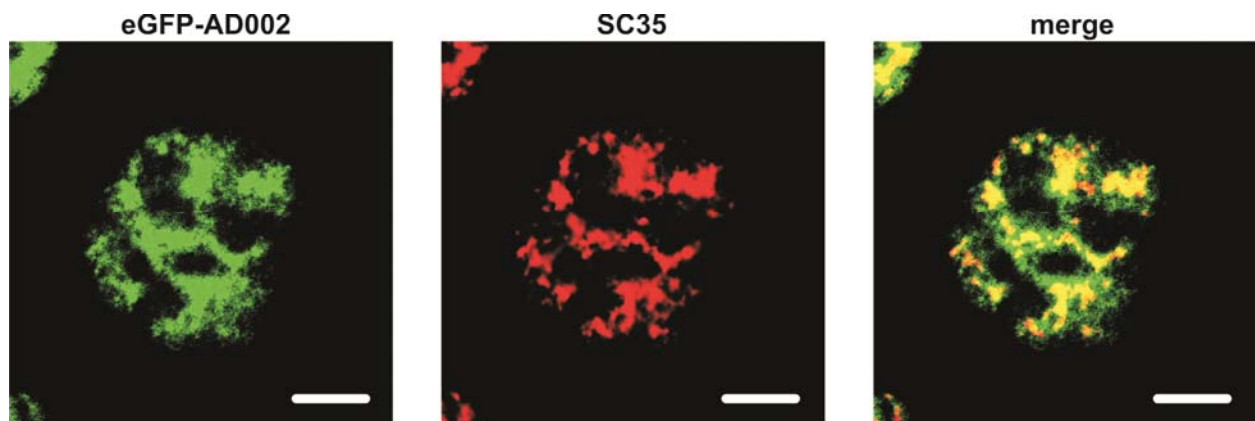


Fig. 3.23: FLAG/StrepII-tagged eGFP-AD002 colocalizes with SC35 in nuclear speckles. eGFP-AD002 expressing HeLa S3 cells were stained with antibodies against SC35 and examined by indirect immunofluorescence (red, middle panel). The eGFP fluorescence is shown in green (left panel). The right panel shows the merge picture of the two panels to its left, where yellow indicates overlying signals. Bar, 5 μm .

Finally, we analyzed whether it is possible to purify hPrp19/CDC5L complexes from the eGFP-AD002 stable cell line. Therefore, we used the same strategy as previously described for the FLAG/HA-AD002 and FLAG/HA-SPF27 cell lines (3.1.2). In brief, after anti-FLAG affinity selection of complexes containing FLAG/StrepII-tagged eGFP-AD002, the eluted complexes were subjected to 5-20% glycerol gradient centrifugation. Proteins were isolated from fractions across the gradient and analyzed by SDS-PAGE (Fig. 3.24). Indeed, the hPrp19/CDC5L complex peaked in fraction 15-17, and contained all of the proteins previously shown to be present in the complex (Fig. 3.3). The purified complex contains eGFP-AD002 in stoichiometric amounts, as estimated by Coomassie staining.

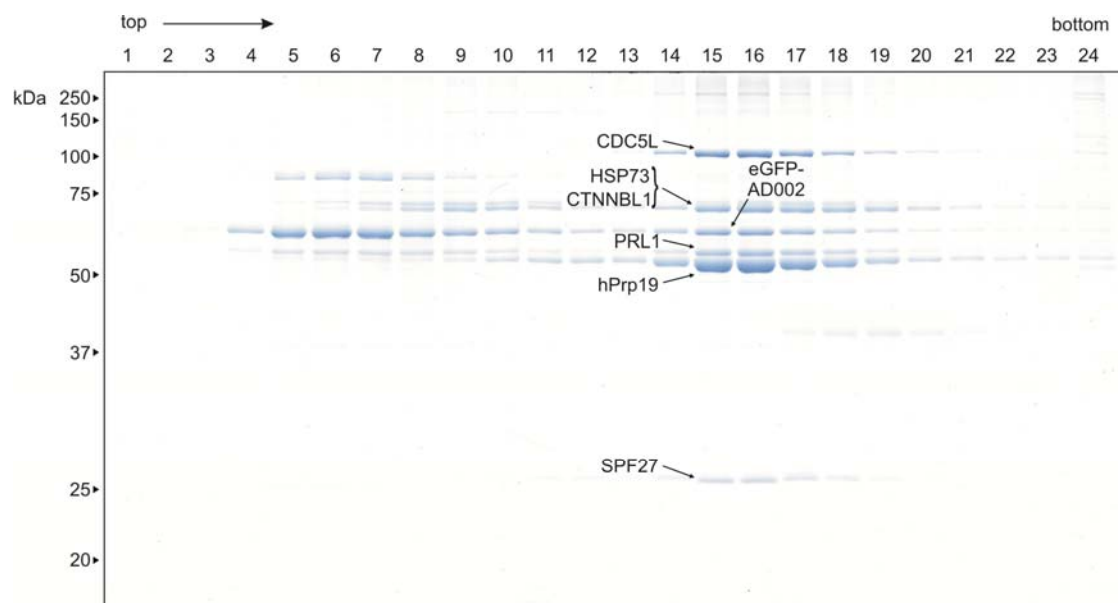


Fig. 3.24: Purification of the hPrp19/CDC5L complex from HeLa cell lines stably expressing FLAG/StrepII-tagged eGFP-AD002. The purification procedure was as described in Fig. 3.3A. Proteins present in each gradient fraction (after glycerol gradient centrifugation) were analyzed by SDS-PAGE and stained with Coomassie. The molecular mass (kDa) of marker proteins is indicated on the left.

Taken together, we were able to establish a HeLa cell line that stably expresses the eGFP-AD002 fusion protein containing a C-terminal FLAG/StrepII-tag and nuclear extract prepared from this cell line exhibited the same splicing activity as control extract. Furthermore, the distribution of eGFP-AD002 in HeLa cells indicates a subcellular localization common for splicing factors. Most importantly, eGFP-AD002 is incorporated into hPrp19/CDC5L complexes and could be used to isolate the hPrp19/CDC5L complex from nuclear extracts. Thus, we suggest that eGFP-AD002 exhibits a similar activity as the endogenous protein and may thus be used for further studies.

3.2 The crystal structure of the WD40 domain of human Prp19

Limited proteolysis identified a 35 kDa protease resistant fragment of human Prp19 (Fig. 3.17, peak 3) and MS analysis revealed that it is the C-terminus of the protein, which contains a WD40 domain (Fig. 3.18). As predicted with the SMART program (Letunic et al., 2009), the WD40 domain of hPrp19 contains seven WD repeats (Fig. 3.25). The WD repeat is a motif containing about 40-60 amino acids usually ending in a tryptophan-aspartic acid (WD) dipeptide at the C-terminus. The repeating unit, first recognized in the β subunit of the GTP-binding protein transducin (Fong et al., 1986), has been identified in many proteins of diverse functions, and is thought to mediate protein-protein interactions (Smith et al., 1999). As shown in Fig. 3.25, considerable variations from the WD consensus sequence can be observed in some repeats of the WD40 motif of hPrp19. Only repeats 2 and 5 contain the canonical WD dipeptide at its C-terminal end. However, in the remaining repeats,

tryptophane is replaced by Phe, Ser, Cys or Tyr, whereas Asp is less redundant and is substituted only by Ser or Lys. In addition, the LxGH motif, often observed at the beginning of a WD repeat (Fong et al., 1986) was found only in repeat 2 of hPrp19 (Fig. 3.25). Thus, considering its variation from the WD40 consensus sequence, we determined whether the WD40 domain of hPrp19 also folds into the well characterized seven-bladed β propeller fold.

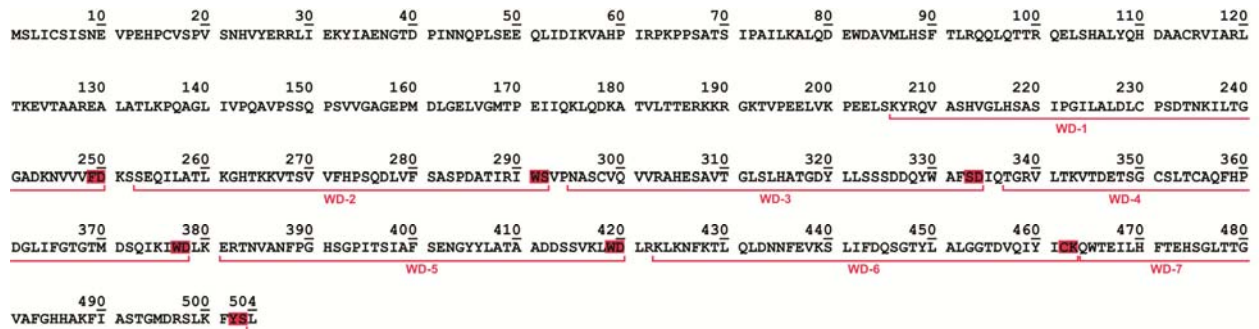


Fig. 3.25: Distribution of the seven WD repeats in the sequence of hPrp19. The WD repeats were predicted with the SMART program (Letunic et al., 2009) and are indicated below the sequence. The WD dipeptide motifs are indicated by red boxes. The amino acid position is shown above the sequence with every tenth residue marked.

3.2.1 Structure determination

The C-terminal fragment of hPrp19 (aa 145-504; compare Fig. 3.26B) was highly resistant to protease cleavage; even after treatment with 250 nM Subtilisin, the fragment remained uncleaved (Fig. 3.17D), indicating that it is a stable structural domain. Thus, after gel filtration (Fig. 3.26A) and concentrating up to 10 mg/ml, this fragment was subjected to crystallization trials. X-ray crystallography and structure determination was performed in collaboration with Dr. Vladimir Pena (X-ray crystallography group, MPI for Biophysical Chemistry). Crystals with a size of $\sim 60 \mu\text{m}$ were observed in the initial screening (Fig. 3.26C). The best diffracting crystals were obtained with a solution containing polyethylene glycol 2000 monomethyl ether and Tris-buffer (pH 8) and allowed collection of a data set to 1.8 Å resolution (Fig. 3.26D). The C-terminal fragment of hPrp19 crystallized in a primitive hexagonal space group ($P3_121$) with one molecule per asymmetric unit (Table 3.1). The structure was solved by molecular replacement using a structural model of the hPrp19 WD40 domain, generated with the HHpred program (Soding et al., 2005). The final model was refined to a resolution of 1.8 Å, with working and free R factors of 18.2% and 23.6%, respectively, retaining good stereochemistry (Table 3.1). Most residues were in the favored or additionally allowed regions of the Ramachandran plot (Fig. 3.27). The statistics of structure determination and refinement are summarized in Table 3.1.

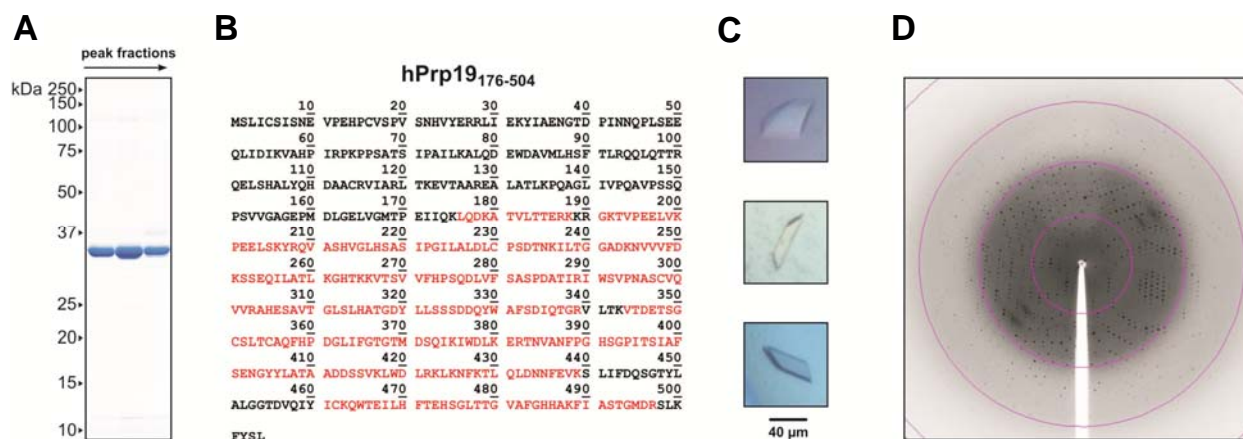


Fig. 3.26: Purification and crystallization of N-terminally truncated hPrp19. The hPrp19/CDC5L complex was subjected to limited proteolysis and the proteolyzed fragments were separated by gel filtration (**A**) SDS-PAGE analysis of the peak gel filtration fractions containing the C-terminal fragment of hPrp19. (**B**) MS analysis of truncated hPrp19 from **A**. Peptides identified by MALDI tandem MS are shown in red. (**C**) Images of crystals grown in 0.1 M Tris-HCl (pH 8.0), 20% (w/v) polyethylene glycol 2000 monomethyl ether. (**D**) Diffraction image of a single crystal from **C**. The outer red circle indicates a resolution of 1.8 Å. Images were taken over a $\Delta\phi$ range of 0.5° at the beamline PXII, SLS.

Table 3.1: Summary of the crystallographic analysis.

<i>Data collection</i>	
<i>Space group</i>	$P3_121$
<i>Cell dimensions</i>	
<i>a, b, c</i> (Å)	82.97, 82.97, 75.27
α, β, γ (°)	90, 90, 120
Wavelength (Å)	0.999
Resolution	20-1.80 (1.847-1.800)
R_{meas}^*	(32.7)
$I/\sigma(I)^*$	26.1 (7.04)
Completeness (%)	97.0 (100)
Number of observed reflections	213,453 (57,892)
Redundancy	7.59 (7.69)
<i>Refinement</i>	
Resolution (Å)	20-1.80
Number of reflections	26,723 (1,946)
$R_{\text{work}}/R_{\text{free}}^\ddagger$	0.182/0.236
<i>Number of atoms</i>	
Protein	2606
<i>B-factors</i>	
Protein	17.407
<i>R.m.s. deviations</i>	
Bond length (Å)	0.013
Bond angles (°)	1.411

*as specified in XDS (Kabsch, 1993)

‡as specified in REFMAC (Murshudov et al., 1997)

Values for the highest resolution shell are given in parentheses

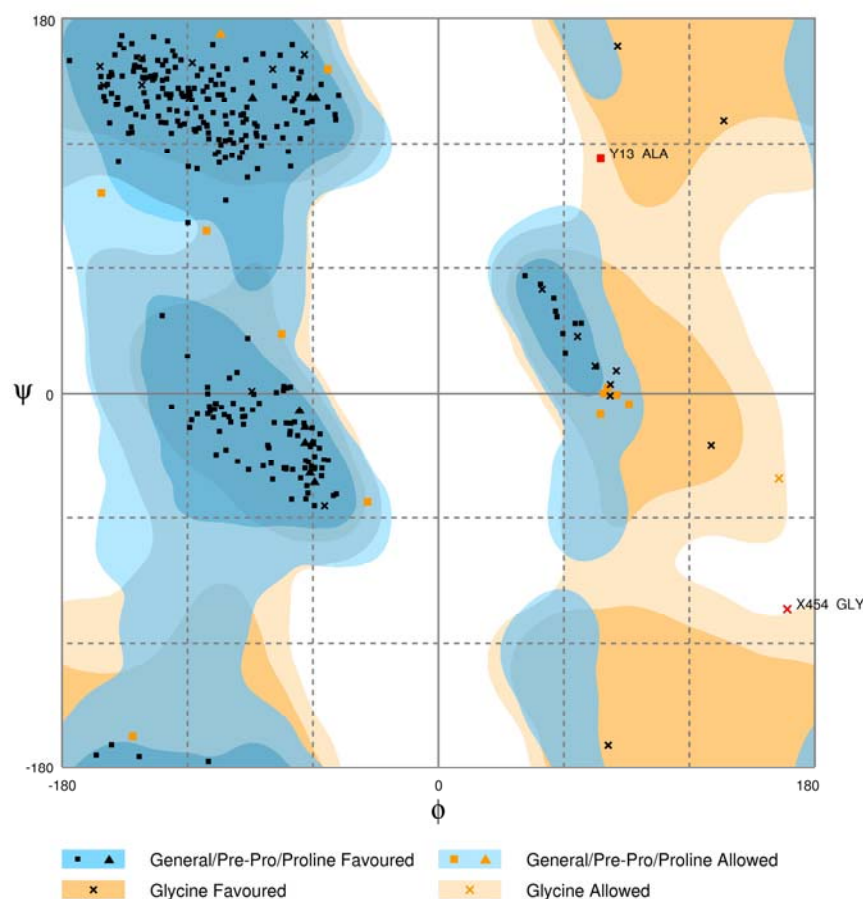


Fig. 3.27: Ramachandran map of the hPrp19 WD40 domain.

The plot calculation was performed with the RAMPAGE program (Lovell et al., 2003). Favoured and allowed regions are colored in blue and orange, as indicated below the plot. 303 residues are located in the favoured region (95.3%), 13 residues are in allowed regions (4.1%) and 2 residues are in outlier regions (0.6%).

3.2.2 Overall structure description

As shown in Fig. 3.28, the polypeptide chain of the truncated hPrp19 folds into a seven-bladed β propeller similar to that observed in the β subunits of heterotrimeric G proteins. The structure is characterized by seven blades that are symmetrically arranged around a central axis. Each blade consists of a four-stranded antiparallel β sheet (labeled A-D), with the strands in each blade ranging from the inside to the outside of the propeller. The center of the propeller is formed by the edges of the seven "A" strands, forming a narrow axial channel that runs through the center of the propeller. The channel is filled with a large number of water molecules (see Fig. 3.34A). Thus, each repeat sequence corresponds to a structural repeat of four β strands. However, the repeat structure is not equivalent to each propeller blade; rather, each propeller blade contains the last three strands of the repeat unit and the first strand of the consecutive repeat unit (Fig. 3.29). Therefore, only blades 1-6 are formed by contiguous segments of the polypeptide chain, whereas in blade 7, the innermost three strands come from the very C-terminus while the outermost strand is provided by the first β strand from the N-terminus. This clamping of the first β strand in the last blade closes the β propeller and contributes to its structural stability. Thus, although of low sequence conservation, the predicted WD40 domain of hPrp19 indeed forms a β propeller structure.

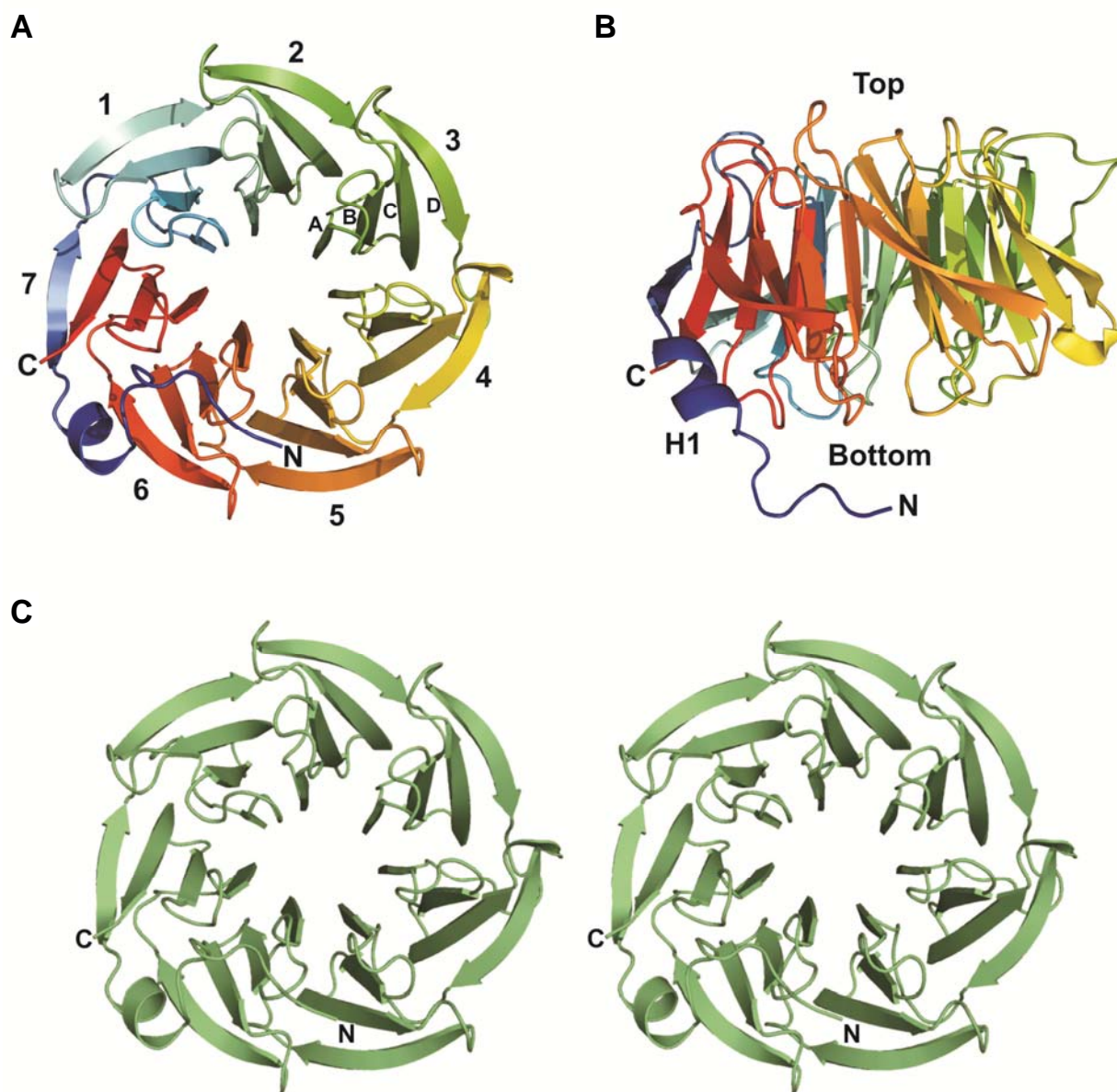


Fig. 3.28: Overall structure of the hPrp19 WD40 domain. (A) Ribbon diagram showing the seven-bladed β propeller structure. Each blade consists of a four-stranded β -sheet (labeled A-D). The view is looking up the central axis of the propeller onto the “bottom” surface. The structure is rainbow colored going from the N-terminus (N, blue) to the C-terminus (C, red). (B) View rotated 90° about the horizontal and vertical axis in A, looking at the propeller from the side. The top face and bottom face are marked according to the convention for WD40-repeat protein structure. The Helix H1 at the N-terminus is indicated. (C) Stereo view of a ribbon diagram of the hPrp19 WD40 domain. The orientation is identical to that in A.

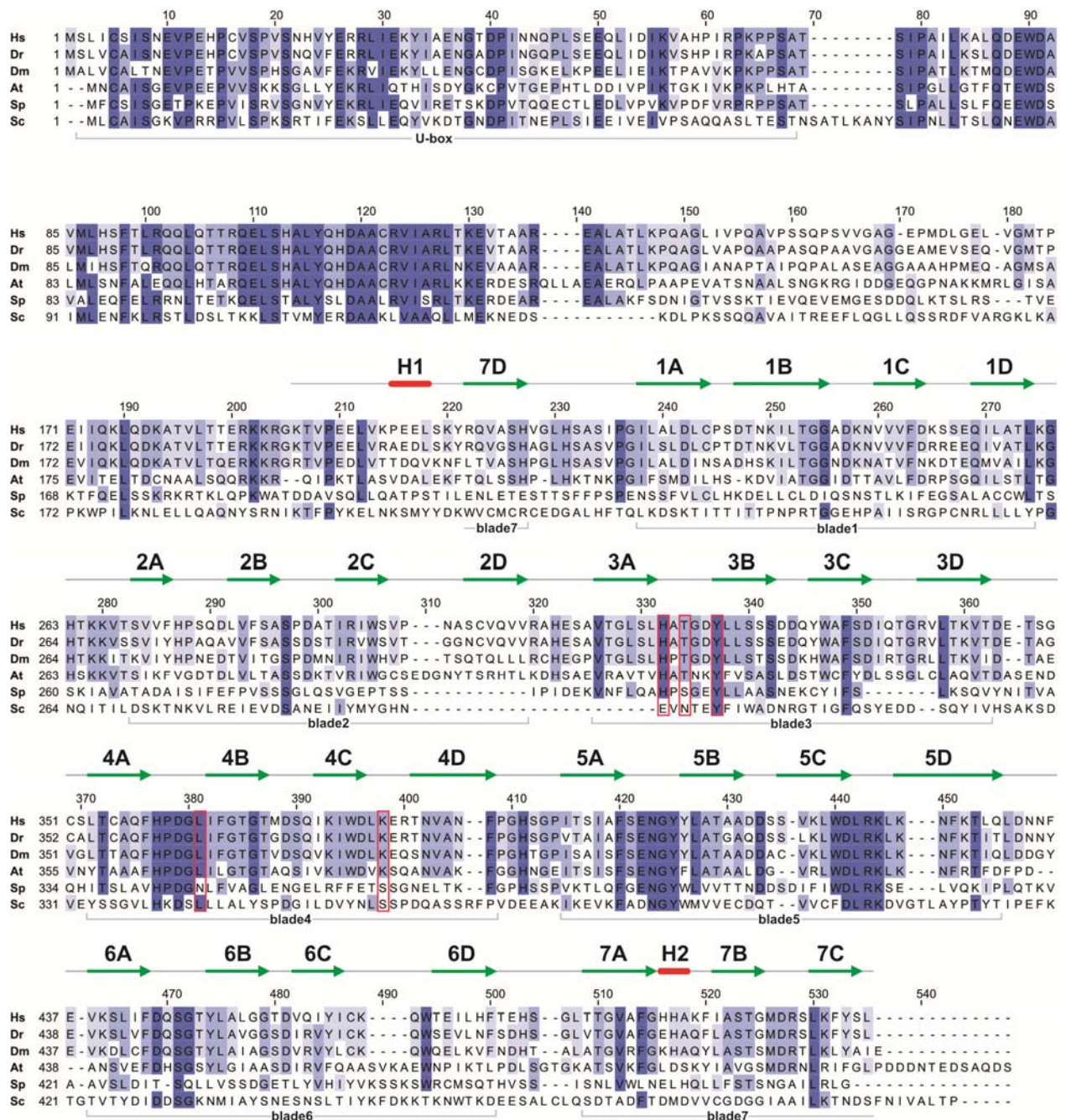


Fig. 3.29: Sequence alignment of Prp19 proteins from different species. The ClustalW 2.0.12 program (Larkin et al., 2007) was used to align hPrp-related proteins from human (Hs), *Danio rerio* (Dr), *Drosophila melanogaster* (Dm), *Arabidopsis thaliana* (At), *Schizosaccharomyces pombe* (Sp) and *Saccharomyces cerevisiae* (Sc). Residues are highlighted in different shades of blue according to conservation, where dark blue indicates a high conservation and light blue a low conservation across these species. The secondary structures of human Prp19 are shown above the sequence and numbered as indicated in Fig. 3.28A. The β propeller blades are indicated below the sequence and numbered from 1-7 as described. Residues that are thought to play a role in binding the unidentified peptide are indicated by red squares. The conserved U-box domain is labeled.

3.2.3 Comparison with other WD repeat proteins

Since its discovery, the WD repeat motif has been identified in many proteins. To date, there are at least 123 proteins known with at least four WD repeats (Smith et al., 1999). The structure of the hPrp19 WD40 domain closely resembles those of other representative members of WD repeat proteins. Although the sequence of the hPrp19 WD40 domain shows only 23% identity with the β subunit of heteromeric G proteins (G_β) and 24% identity with the histone binding protein WDR5 (Fig. 3.30), the β propeller of hPrp19 superimposes strikingly well with G_β , with an RMSD of 1.7 Å over 335 C α atoms, and with WDR5, with an RMSD of 1.4 Å over 335 C α atoms (Fig. 3.31). Besides the structure of the single sheets, which are highly conserved between the three proteins, also the loop regions superimpose well.

Despite these similarities, some structural differences exist between the β propeller of hPrp19 and those of G_β and WDR5. The most notable differences occur in the N-terminus, where the propeller of G_β was superimposed with a shift of one blade with the propellers of hPrp19 and WDR5 (Fig. 3.31). Furthermore, the N-terminal extensions of G_β and WDR5 form longer loop regions, whereas the N-terminus of hPrp19 forms a short α -helix. Taken together, the overall structure of the WD40 domain of hPrp19 is structurally highly conserved and can be grouped into the WD40-repeat like superfamily of seven-bladed β propellers.

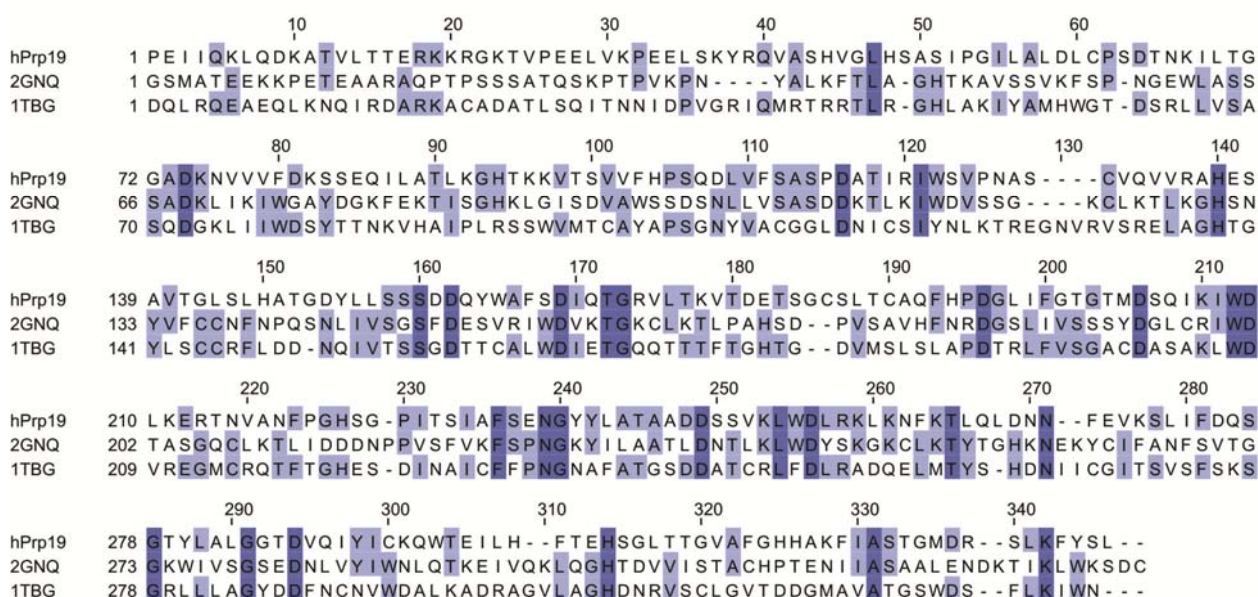


Fig. 3.30: Sequence alignment of different WD40 repeat domains. The ClustalW 2.0.12 program (Larkin et al., 2007) was used to align WD40 domains from hPrp19, WDR5 (2gnq) and the β subunit of heterotrimeric G proteins (1tbg). Residues are highlighted in different shades of blue according to conservation, where dark blue indicates a high conservation and light blue a low conservation across the species.

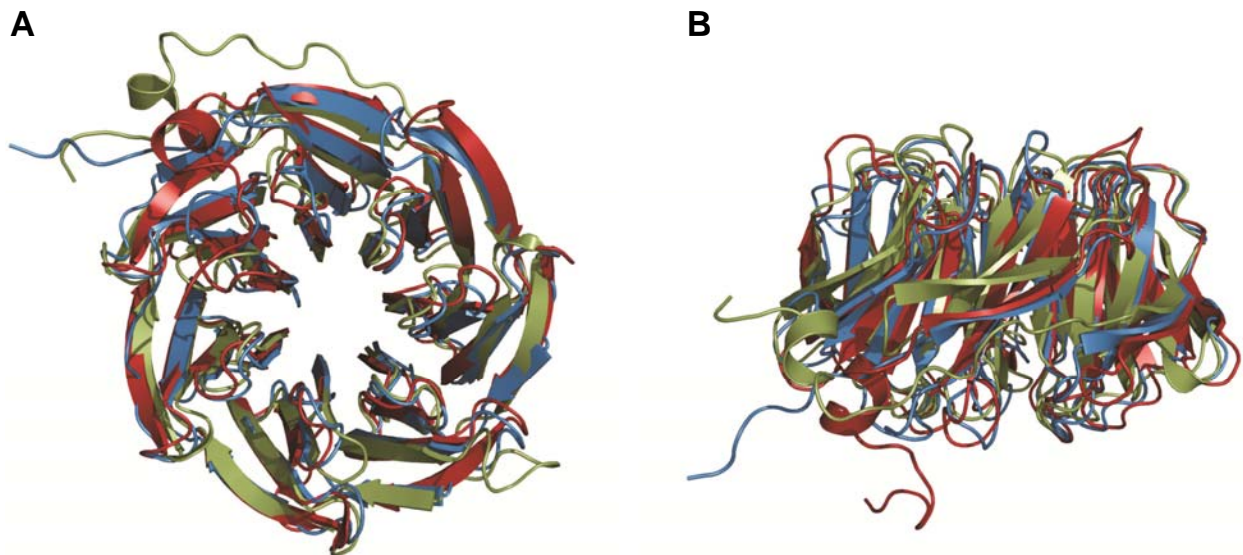


Fig. 3.31: Superposition of the hPrp19 WD40 domain with other WD40 repeat-containing proteins reveals overall similarity. (A) View from the bottom of the WD40 domain structures. The superimposed WD40 domain structures are shown in red (hPrp19), blue (WDR5; 2GNQ) and green (G β subunit; 1TBG). (B) Side view of the same molecules. The superposition was generated with the STAMP program (Russell and Barton, 1992).

Strikingly, a hydrogen bonding network, also known as a “structural tetrad”, which superficially resembles the catalytic triad found in serine proteases was also observed, but only in blades 2 and 5 of hPrp19 (Fig. 3.32). This tetrad was first observed in the G β subunit and is formed between Trp in strand C, Ser/Thr in strand B, His in the DA loop and the nearly invariant Asp in the tight turn between strands B and C (Wall et al., 1995).

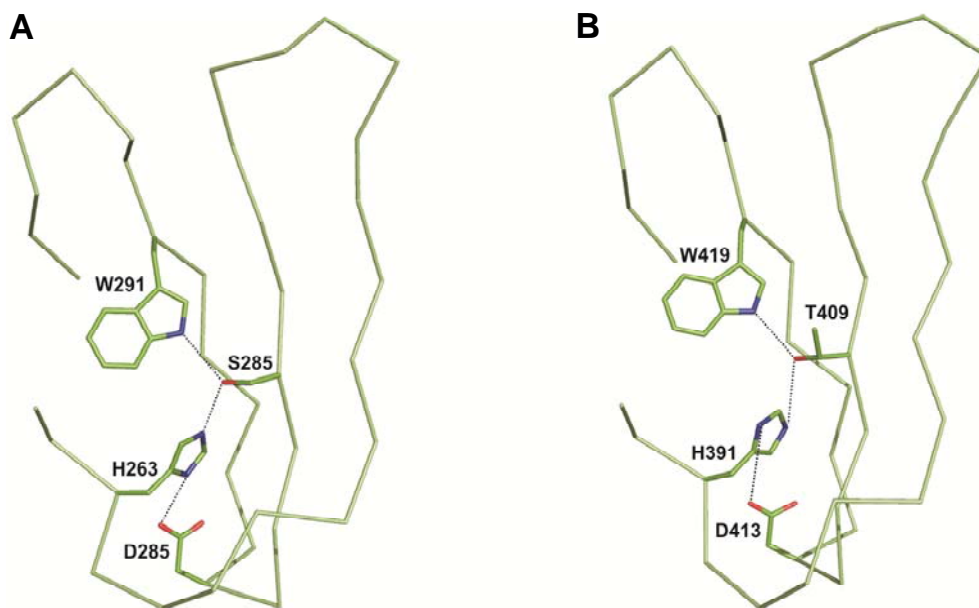


Fig. 3.32: The two “structural tetrads” of the β propeller of hPrp19. Backbone representation of (A) blade 2 and (B) blade 5, illustrating the side chains involved in the formation of the “structural tetrad”. Hydrogen bonds are indicated with dashed lines. The position of the amino acids refers to the full-length hPrp19 protein.

Interestingly, the crystal structure of the yeast Prp19 WD40 domain was solved very recently at 2.6 Å resolution (Vander Kooi et al., 2010). However, this WD40 domain shows an atypical folding (Fig. 3.33B). Instead of a nearly circular geometry, as in the human WD40 domain, the yeast counterpart has a distorted seven bladed WD40 architecture with significant asymmetry due to irregular packing of blades one and seven into the core of the WD40 domain (Fig. 3.33), as discussed below.

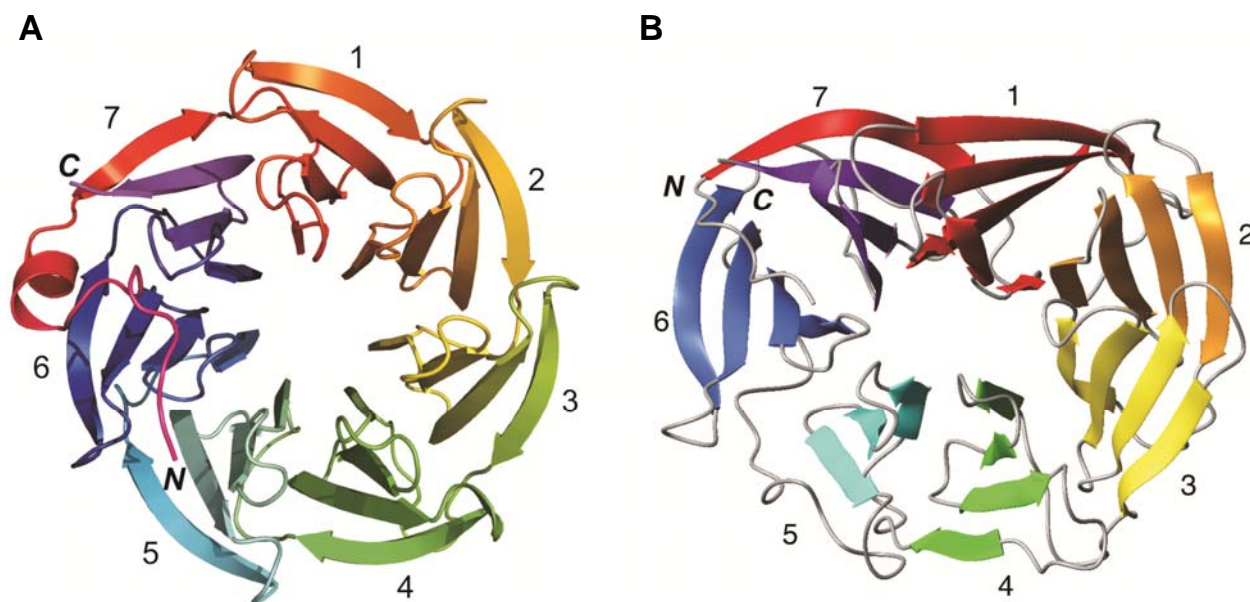


Fig. 3.33: X-ray structure comparison of the WD40 domains from human and yeast Prp19. (A) Human structure as described in this work. (B) Yeast structure (3LRV), adapted from Vander Kooi et al., 2010.

3.2.4 Identification of a peptide bound to the β propeller

A closer look at the electron density map of the crystal structure of hPrp19's WD40 domain revealed the existence of a nine residue peptide that is bound to the β propeller surface (Fig. 3.34). This peptide adopts a turned conformation with a part of it folded as a α -helix that binds to the bottom of the β propeller. Of the nine residues within the peptide, only the side chain of an Asp (or Asn) residue at position 5 assumes a well defined conformation in the electron density map, whereas the remaining side chains do not appear (Fig. 3.35A). Nonetheless, the directionality of the peptide could be unambiguously determined from the electron density in the 1.8 Å structure. After refinement, the carbonyl oxygen bumps became apparent in the electron density map, as well as the side chain branches from the $C\alpha$ atoms. However, the electron density map was too weak in this region to determine the peptide sequence and consequently, except for the Asp/Asn residue, the peptide model was built as a poly-Ala chain. Thus, we currently can only speculate about the origin of this peptide. One possibility is that it is a part of the N-terminus of the protease resistant fragment of the hPrp19 protein. MS identified 360 residues in the protease resistant fragment of hPrp19 (Fig. 3.26B). However, in the 1.8 Å structure, the N-terminal 47 amino acids are missing.

Furthermore, since the crystalized fragment is a part of the entire hPrp19/CDC5L complex that was digested with protease, another possibility is that the peptide belongs to a second hPrp19/CDC5L complex protein that interacts with the hPrp19 WD40 domain in this region.

Specific recognition of the peptide might in part be mediated by contacts with the peptide's Asp/Asn residue. The carboxylate side chain fits well into an amphipathic pocket build up by His315, Thr317, Tyr320, Leu363 and Lys380 of the β propeller (Fig. 3.35A). These residues are highlighted in Fig. 3.29 and they are part of the turn regions 3A3B, 4A4B and 4C4D. Furthermore, the highly conserved Lys380 is the first residue of the 3B β sheet. However, due to the weak electron density we were not able to specify interactions between the peptide and the β propeller of hPrp19 in more detail, but most certainly it involves a number of hydrophilic and hydrophobic contacts.

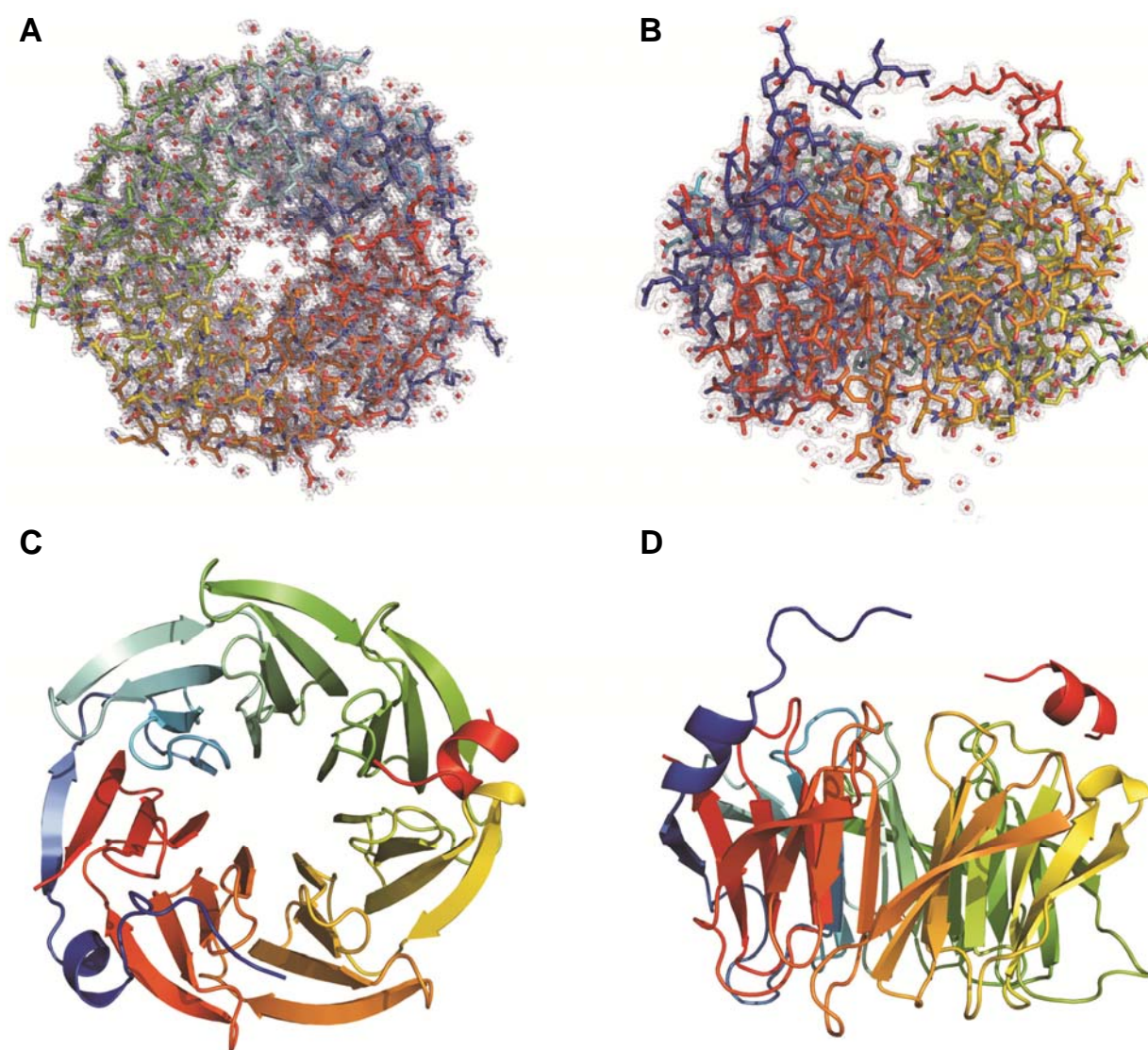


Fig. 3.34: Electron density map for the hPrp19 WD40 domain. The electron density mesh is shown in light blue (1.1 sigma). The polypeptide chain is shown in rainbow colors, starting with blue (N-terminus) and ending with red (C-terminus). The undefined peptide is shown in red. (A and C) View from the bottom of the β propeller. (B and D) View rotated 90° about the horizontal axis showing the β propeller from the side. Also notice loop 5D6A (orange), which clearly bulges out of the top surface of the β propeller.

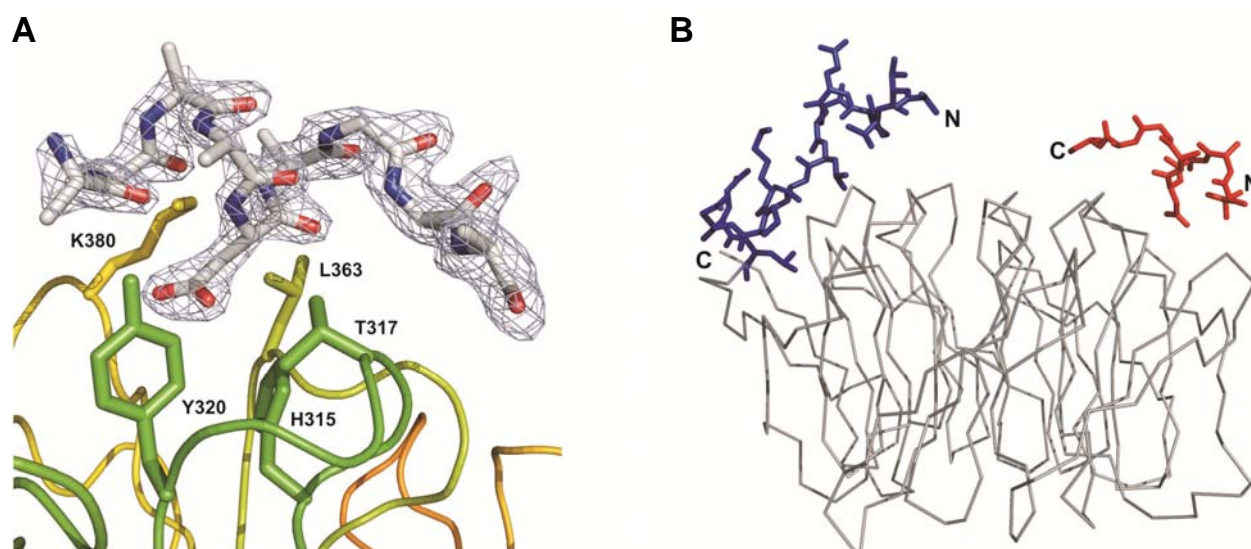


Fig. 3.35: The electron density of the additional peptide and the very N-terminus of the hPrp19 WD40 domain is not well defined. (A) Close-up view of the interface between the undefined peptide (shown together with the electron density in 1.1 sigma) and the WD40 domain of hPrp19. The side chains of those residues from the WD40 domain involved in interactions with the peptide are shown and labeled. The position of the amino acids refers to the full-length hPrp19 protein. **(B)** Position of the peptide (red) relative to the N-terminus of the hPrp19 WD40 domain (blue).

3.2.5 Surface analysis of the β propeller and location of functional sites

Despite its low sequence identity the WD40 β propeller shows a high structural conservation. This conservation can be expanded also to the surface electrostatics, which were used previously to cluster WD repeat β propeller proteins into different subfamilies (Valeyev et al., 2008). It is known that electrostatics play an important role in molecular interactions and surface electrostatic patches on proteins affect the specificity of protein-protein interactions (Honig and Nicholls, 1995). Electrostatic potential is a protein property that can only be examined with structural knowledge (Valeyev et al., 2008). As shown in Fig. 3.36, the electrostatic potential of hPrp19's β propeller is mainly negative. However, some residues protruding out of the surface are also positively charged. The central channel is almost completely negatively charged, with only a small region at the top surface being positively charged, which extends over the side of the β propeller surface to its bottom (Fig. 3.36B). Between the charged areas, there are also numerous grooves that are generally uncharged, suggesting potential binding pockets. Interestingly, the unassigned peptide binds into such a hydrophobic cavity (Fig. 3.36C). This suggests that hydrophobic interactions contribute to its interaction with the hPrp19 β propeller.

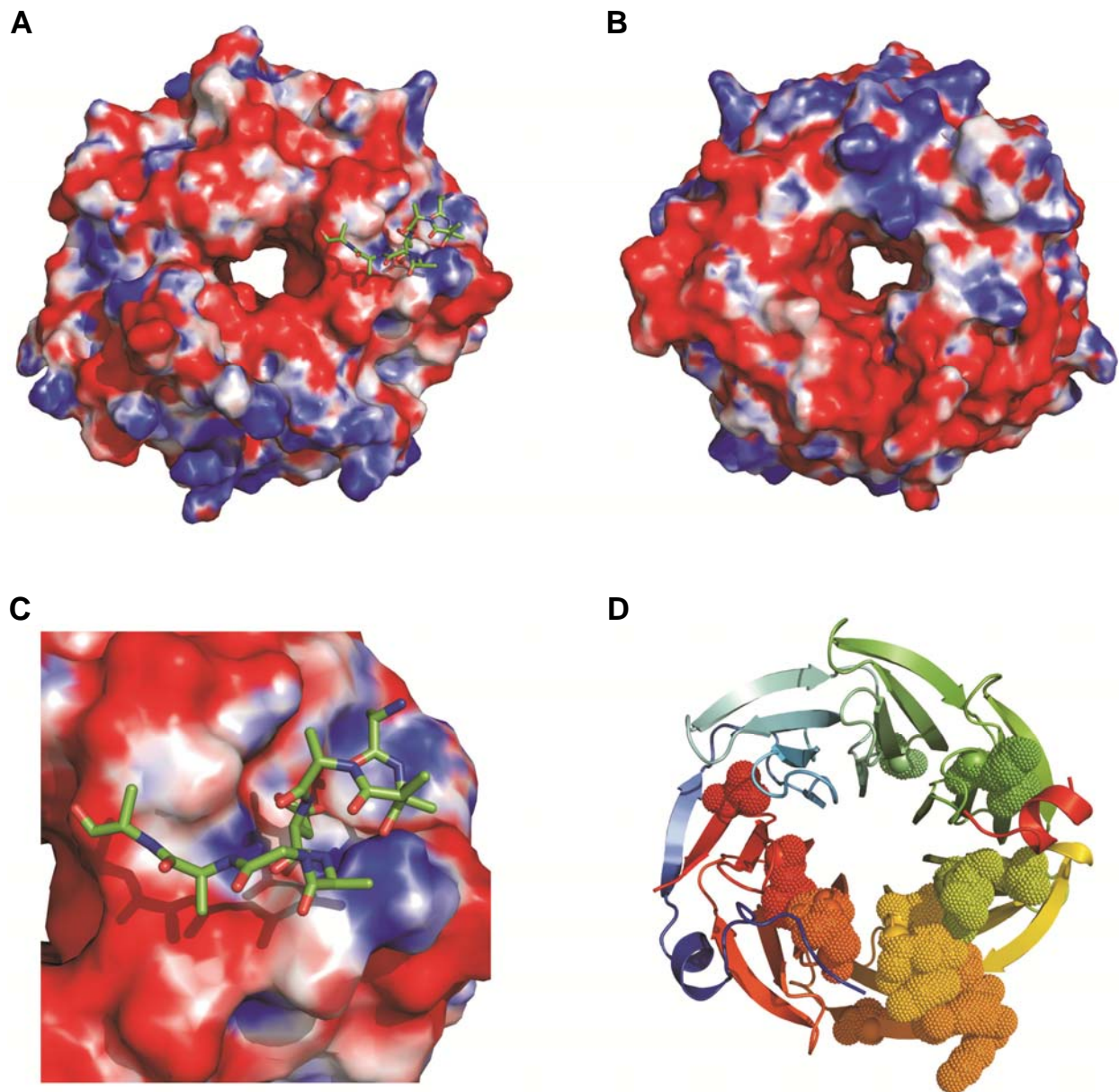


Fig. 3.36: The surface of the hPrp19 WD40 domain. (A-C) Surface colored according to its electrostatic potential. Negative charge is shown in red, positive charge is colored blue and white indicates uncharged regions. The map was calculated with the program APBS (Baker et al., 2001) and displayed with PyMOL (www.pymol.org). (A) View from the bottom of the β propeller's surface as shown also in Fig. 3.28A, showing the position of the unassigned peptide on the surface of the hPrp19 WD40 domain. (B) View rotated 180° about the vertical axis in A, showing the top surface of the β propeller. (C) Close-up view on the unassigned peptide in front of the surface of the β propeller. (D) Location of the highly conserved residues on the surface of the β propeller. View from the bottom of the β propeller's surface. Colors are in accordance with Fig. 3.28.

A sequence alignment with Prp19 proteins from different species showed that several residues of the Prp19 WD40 domain are conserved between *S. cerevisiae* and higher eukaryotic organisms (Fig. 3.29). These include Ser283, Tyr320, Phe332, His359, Asp361, Leu363, Phe400, Asn403, Tyr405, Asp420, Leu421, Arg422, Leu423, Thr429, Asp444, Ser446, Phe483 and Leu499. Intriguingly, these residues show a higher sequence conservation than the WD-motif itself. As indicated in Fig. 3.36D,

nearly all of these residues are located at the bottom surface of the β propeller. Interestingly, Tyr320 and Leu363, which might be involved in binding of the unassigned peptide, are also highly conserved. Additionally, Fig. 3.36D reveals high sequence conservation in blades 4 and 5, especially in the connecting area. This suggests that this region might be involved in additional protein-protein interactions and is consistent with recent data from the yeast Prp19 WD40 domain (Vander Kooi et al., 2010). For the latter domain, the authors could identify a highly conserved surface centered around blade five in a structure-based mutagenesis approach. Residues Arg403 and Lys404 (representing Arg422 and Lys423 in the human protein) were shown to be required for the physical interaction between Prp19 and Cwc2, an essential splicing factor that is part of the yeast NTC. However, previous studies reported that functional residues are located mainly on the top surface of the β propellers (Valeyev et al., 2008). In contrast, our data and data from the yeast Prp19 WD40 domain (Vander Kooi et al., 2010) suggest a possible involvement of the bottom surface of the β propeller surface in protein-protein interactions.

Residues that are clearly exposed on the surface may also be functionally important. Here, the 5D6A loop should be emphasized, as it clearly bulges out of the top surface of the β propeller (Fig. 3.34B and D), presenting Asp433, Asn434 and Asn435 to the solvent. Taken together, the surface of the hPrp19 β propeller provides a continuum of potential protein-protein interactions.

3.3 Purification and EM structure the human 35S U5 snRNP

After completion of the first step of splicing, the spliceosomal C complex possesses a relatively small subset of proteins (~35 of its full complement of ~110) that remain stably bound together with all of its RNAs – the U2, U5 and U6 snRNAs, as well as both pre-mRNA splicing intermediates (excised 5' exon and intron lariat), even at high salt concentrations, at which many RNPs are unstable (Bessonov et al., 2008). This stable RNP core of the C complex contains, among others, most of the constituents of a U snRNP particle that is thought to be formed within the spliceosome during catalytic activation and then released after completion of splicing, namely the U5 snRNA containing 35S U5 snRNP (Makarov et al., 2002).

3.3.1 Purification and MS analysis of the human 35S U5 snRNP

To learn more about the structure of the human 35S U5 snRNP, we set out to isolate this snRNP from HeLa nuclear extract. Similar to the hPrp19/CDC5L complex, 35S U5 snRNPs were purified from a HeLa cell line that stably expresses AD002, one of the protein components of the 35S U5 snRNP, with an N-terminal FLAG-tag. 35S U5 snRNP particles were anti-FLAG affinity-selected from

heparin-treated nuclear extract (incubated under splicing conditions) and further purified by glycerol gradient centrifugation (Fig. 3.37A). Purified 35S U5 snRNPs peaked in fractions 15-22 (Fig. 3.37B and C), as evidenced by the distribution of the U5 snRNA and known protein components of the 35S U5 snRNP (e.g. U5-220K/hPrp8 and U5-200K/hBrr2). Purified 35S U5 snRNPs contained not only the U5 snRNA (present in several isoforms), but also very low amounts of the U6 snRNA (Fig. 3.37C, lane 15-20), as verified by northern blotting (Fig. 3.38).

To determine the polypeptide composition of the 35S U5 snRNP, proteins from affinity-selected, gel filtrated 35S particles were separated by SDS-PAGE and subsequently characterized by MS. For this purpose, the polyacrylamide gel lane containing 35S U5 snRNP proteins (e.g. Fig. 3.37, lane 17+18) was cut into 24 pieces of equal size, and proteins were recovered from the gel, digested with a protease and subjected to liquid chromatography-tandem MS (LC-MS/MS). The proteome of the human 35S U5 snRNP was analyzed from two independent preparations and is summarized in Table 3.2. For MS analysis of the two independent preparations, two different mass analyzers were used (Q-ToF in preparation #1, and OrbiTrap in preparation #2). As the latter is more sensitive, the number of peptides sequenced for a given protein was higher. MS analysis of the 35S U5 snRNP revealed that the 35S U5 snRNP has a protein complement fundamentally different from that of the smaller 20S U5 particle, known to enter the splicing cycle as part of the 25S U4/U6.U5 tri-snRNP (Makarov et al., 2002). That is, unlike the 20S U5 snRNP, the 35S U5 snRNP contains proteins of the hPrp19/CDC5L complex and hPrp19/CDC5L complex-related proteins and several proteins found also in the spliceosomal B/B* and/or C complex (e.g. hPrp22, hPrp16, hSlu7 and PPIase-like 3b). All proteins identified were consistently found in both 35S U5 snRNP preparations and most proteins were also previously found in an earlier 35S U5 snRNP preparation from our laboratory (Makarov et al., 2002). Thus, they clearly represent 35S U5 snRNP components. However, several proteins that were not present in earlier preparations of the 35S U5 snRNP were detected in our anti-FLAG affinity purified 35S U5 snRNPs. These proteins include, among others, the human Prp19/CDC5L complex component CTNNBL1, the U5 snRNP proteins 52K, 100K/hPrp28 and 102K/hPrp6, and several proteins known to be stably associated with the human C complex (hSlu7, hPrp16, hPrp22, DDX35, Q9BRR8, Morg1, PPIase-like 3b and GCIPp29). However, for some of these proteins, only a small number of peptides were sequenced by MS, a first indication that they may be present in substoichiometric amounts.

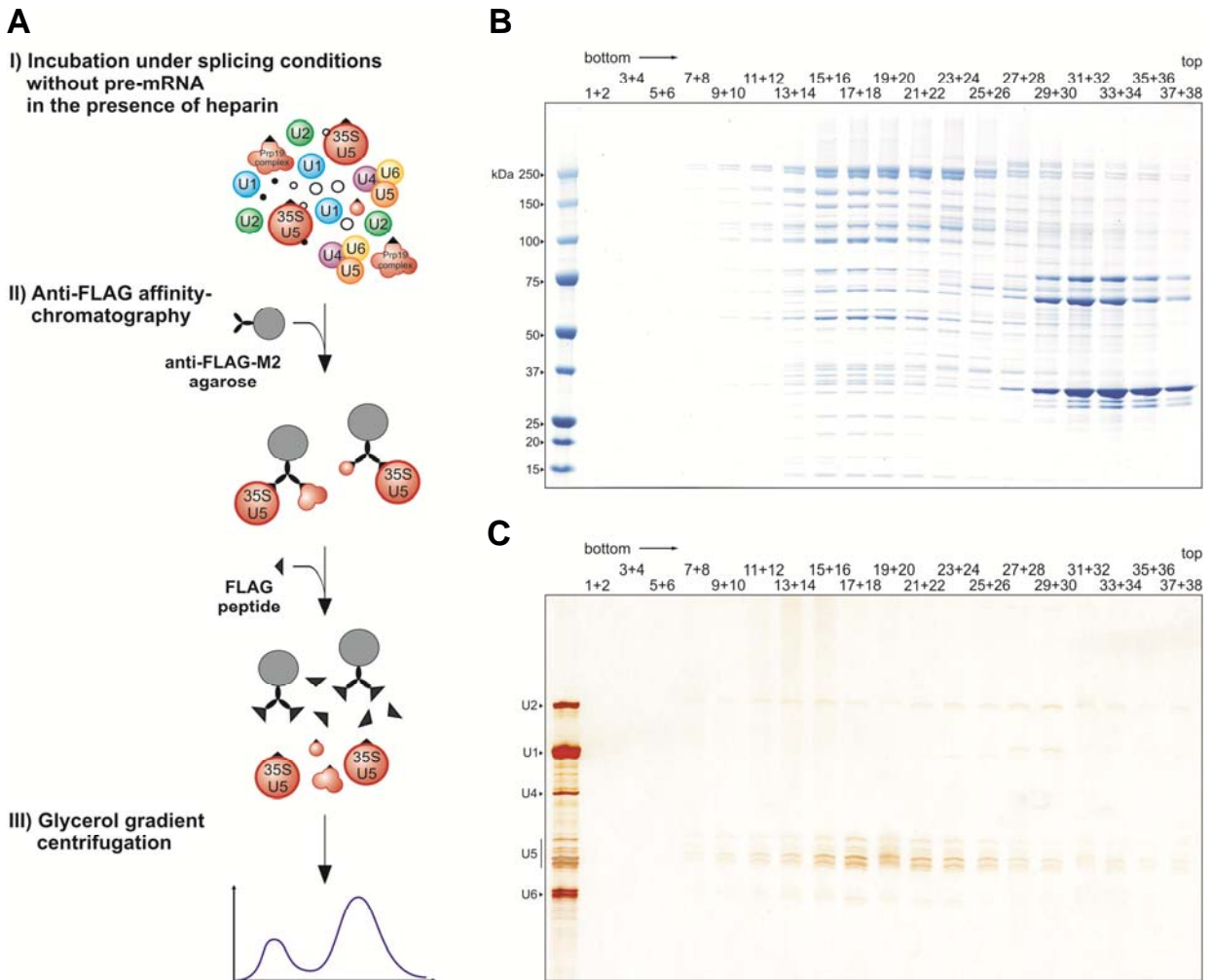


Fig. 3.37: Purification of the human 35S U5 snRNP from HeLa cell lines stably expressing FLAG/HA-tagged AD002. (A) Schematic diagram of the anti-FLAG immunoaffinity purification strategy. Complexes were isolated from a splicing reaction prepared with nuclear extract from the FLAG/HA-AD002 stable cell line without pre-mRNA. After treatment with heparin, complexes were immunoaffinity selected using anti-FLAG-M2-Agarose and complexes were eluted under native conditions with FLAG peptide. Affinity purified complexes were then fractionated on a linear 10 to 30% glycerol gradient. (B) Proteins present in two subsequent gradient fractions were pooled, analyzed by SDS-PAGE and stained with Coomassie. The molecular mass (kDa) of marker proteins is indicated on the left. (C) Analysis of the snRNAs present in each gradient fraction by denaturing PAGE and subsequent silver staining. The migration position of marker snRNAs is indicated on the left.

To analyze whether these additional proteins are likely stoichiometric (or near stoichiometric) components of the human 35S U5 snRNP, only those proteins present in visible bands after SDS-PAGE were analyzed by MS. Thus, only Coomassie or silver stained bands were cut from the polyacrylamide gel, thereby avoiding proteins that are present in substoichiometric amounts (fig. 3.39). Silver staining was performed to identify low molecular weight proteins that were poorly visualized by Coomassie staining. Thus, only 35S U5 snRNP proteins with a molecular weight lower than 30 kDa were also analyzed by MS from the silver stained polyacrylamide gel.

Most proteins identified by MS, when the entire lane of the polyacrylamide gel was analyzed (Table 3.2), were also found when only visible bands were analyzed (ig. 3.39). However, some proteins for which only a few sequenced peptides were identified in the whole lane approach were not detected by analyzing single bands (e.g. U5-100K/hPrp28 or U5-52K), suggesting that they are either underrepresented in the 35S U5 snRNP preparation or are contaminants. Similar to previous results from our laboratory (Makarov et al., 2002), a total of ~37 proteins were found to be present in significant amounts in the 35S U5 snRNP. With the exception of FLJ10206, KIAA0773, C25A1.10.p and S100A9 all previously identified 35S U5 snRNP proteins were also present in the 35S U5 snRNP purified from the FLAG/HA-AD002 stable cell line, including mainly components of the U5 snRNP, the hPrp19/CDC5L complex and hPrp19/CDC5L complex-related proteins. However, some proteins that were not described in this previous study appear to be major components in our 35S U5 snRNP preparation. These include mainly those proteins that were also found in the catalytically active human C complex, namely hPrp22, hPrp16, hSlu7 and PPIase-like 3b. Furthermore, one protein, DNAJC17, that was not identified in preparations of spliceosomal complexes so far, was also present in the 35S U5 snRNP and might represent a new splicing factor. However, its function is not known and requires further analysis.

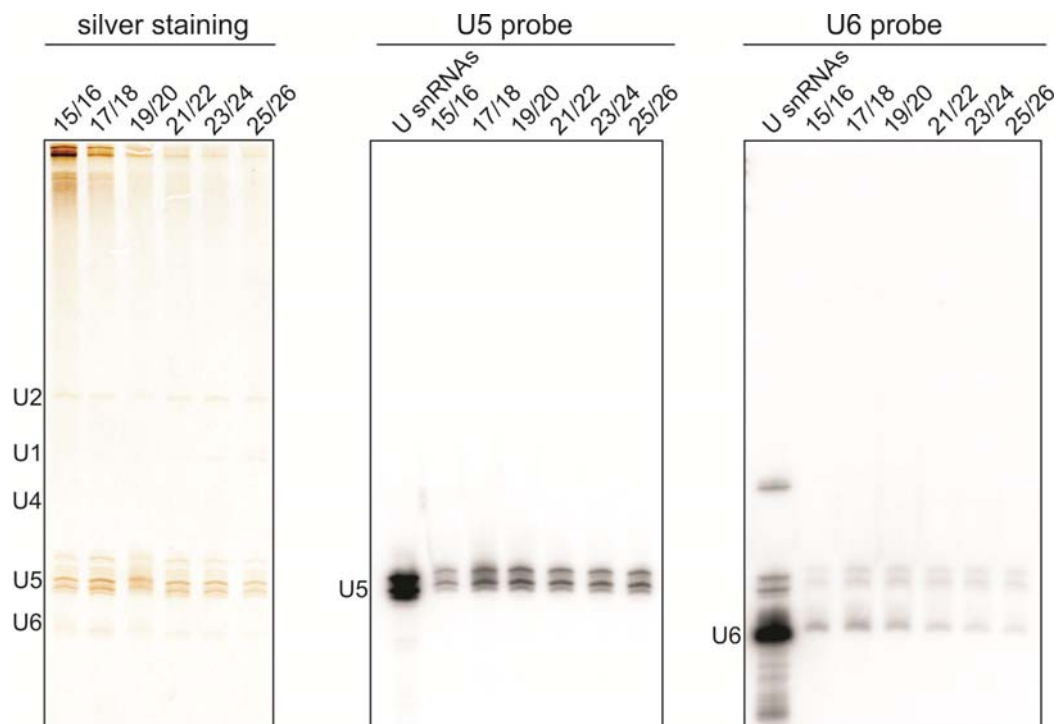


Fig. 3.38: The 35S U5 snRNP also contains low amounts of U6 snRNA. Northern blot analysis of the snRNA composition of purified 35S U5 snRNPs. Affinity-purified complexes were further purified by glycerol gradient centrifugation, and RNA was recovered from those fractions containing the 35S U5 snRNPs and separated by denaturing PAGE. Fractions are numbered as shown in Fig. 3.37 and are indicated at the top of each gel. The silver stained gel is shown on the left. Northern blotting was performed with U5 (middle) and U6 (right) probes, and bands were visualized by autoradiography. The position of marker snRNAs is indicated on the left.

Table 3.2: Protein composition of the human 35S U5 snRNP. Two different preparations of the 35S U5 snRNP were analyzed (#1 and #2). For each protein, the total and unique number of peptides identified by MS are indicated. Preparation #2 was analyzed by an OrbiTrap mass analyzer (as opposed to Q-TOF) resulting in a higher number of peptides sequenced. The presence of each protein in the preparation from Makarov et al., 2002 is indicated with a black dot in the right column.

	MW [kDa]	GenBank accession no.	#1 (Q-TOF)		#2 (OrbiTrap)		Present in Makarov et al., 2002
			total peptides	unique peptides	total peptides	unique peptides	
SM proteins							
Sm B	24.6	gi 4507125	6	5	26	8	•
Sm D1	13.3	gi 5902102	4	2	7	4	•
Sm D2	13.5	gi 29294624	6	4	23	9	•
Sm D3	13.9	gi 4759160	10	3	13	7	•
Sm E	10.8	gi 4507129	7	3	16	4	•
Sm F	9.7	gi 4507131	5	3	11	3	•
Sm G	8.5	gi 4507133	3	2	4	4	•
U5 snRNP							
200K/hBrr2	273.7	gi 45861372	135	63	433	104	•
220K/hPrp8	244.5	gi 3661610	95	58	594	141	•
116K/hSnu114	109.4	gi 41152056	68	30	172	48	•
102K/hPrp6	106.9	gi 40807485	17	15	35	27	
40K	39.3	gi 4758560	11	6	26	15	•
100K/hPrp28	95.6	gi 41327771	3	1	35	25	
52K	37.6	gi 5174409	5	5	11	6	
hPrp19/CDC5L complex							
CDC5L	92.2	gi 11067747	49	25	187	53	•
PRL1	57.2	gi 4505895	40	26	78	29	•
hPrp19	55.2	gi 7657381	36	17	116	24	•
SPF27	21.5	gi 5031653	15	9	42	17	•
HSP73	70.4	gi 5729877	10	10	35	23	•
AD002	26.6	gi 7705475	10	6	37	12	•
CTNBNB1	65.1	gi 18644734	10	7	4	4	
hPrp19/CDC5L complex-related							
CRNKL1/hSYF3	100.6	gi 30795220	78	32	156	54	•
hSYF1 (XAB2)	100	gi 55770906	71	35	262	54	•
KIAA0560 (AQR)	171.3	gi 38788372	70	35	293	71	•
SKIP	51.1	gi 6912676	44	19	106	36	•
hlsy1	33	gi 20149304	24	15	36	25	•
PPIase-like 1 (PPIL1)	18.2	gi 7706339	18	6	26	8	•
RBM22	46.9	gi 8922328	12	8	23	15	•
Cyp-E	33.4	gi 5174637	9	6	36	1	•
G10	17	gi 32171175	4	2	11	6	•
Npw38	30.5	gi 5031957	1	1	5	4	
Complex C							
hPRP22	139.3	gi 4826690	51	30	195	64	
hPRP17	65.5	gi 7706657	21	18	57	40	•
DDX35	78.9	gi 20544129	17	16	35	23	•
GCIP p29	28.7	gi 7661636	14	11	29	18	•
hPRP16	140.5	gi 17999539	11	9	38	28	
Q9BRR8	103.3	gi 21361684	4	4	24	21	•
hSlu7	68.4	gi 27477111	4	4	18	14	
PPIase-like 3b	18.6	gi 19557636	2	1	5	3	
Complex B/B*							
MGC23918	19.2	gi 21389497	15	8	30	14	•
hSmu-1	57.5	gi 8922679	12	10	18	15	
RED	65.6	gi 10835234	5	5	11	10	
Hsp27	22.8	gi 4504517	2	2	4	2	
MGC20398	42	gi 49472814	2	2	19	15	
MFAP1	51.9	gi 50726968	1	1	3	3	
35S U5 snRNP specific							
DNAJC17	34.7	gi 296434239	18	13	58	23	
S100A8	10.8	gi 21614543	2	1	7	2	•
Other spliceosomal proteins							
MORG1	34.3	gi 153791298	16	10	16	13	
TFIP11	96.8	gi 8393259	14	13	65	29	
RACK1	35.1	gi 5174447	9	8	11	9	
GCFC	104.8	gi 22035565	8	5	73	38	
p68 (DDX5)	69.2	gi 4758138	3	3	17	15	
RBM10	103.5	gi 12644371	1	1	13	12	
110K/hSnu66	90.2	gi 10863889	1	1	10	10	

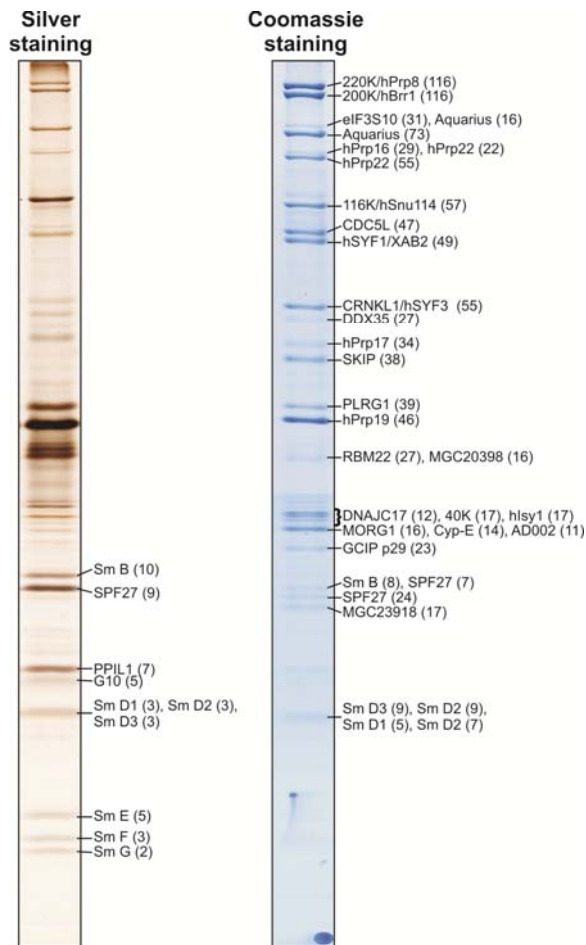


Fig. 3.39: Proteins present in the 35S U5 snRNP in significant amounts. 35S U5 snRNP proteins were separated by SDS-PAGE and stained with silver or Coomassie. MS analysis was performed with all visible bands in the Coomassie stained gel and with bands in the low molecular weight region (<30 kDa) in the silver stained gel. The identified proteins are indicated together with the number of peptides sequenced by MS in parentheses.

3.3.2 EM analysis of the human 35S U5 snRNP

To determine the structure of the 35S U5 snRNP, affinity-purified particles were subjected to negative staining with uranyl formate using the GraFix method (Fig. 2.1) and subsequently analyzed by EM. EM analysis of the 35S U5 snRNP was performed in the 3D Electron Cryomicroscopy group of Prof. Dr. Holger Stark (MPI for Biophysical Chemistry) by Drs. Monika M. Golas and Björn Sander. Typical EM raw images of the 35S U5 snRNP show a homogenous, monodisperse population of particles with a maximum dimension of about 24–27 nm (Fig. 3.40A). A total of 59,380 molecular images were subjected to image analysis. Unbiased single-particle image clustering indicated a preferential orientation of the particles on the carbon film. The most frequent class averages in the main view, presented in Fig. 3.40B, appear typically trapezoidal or triangular. As suggested by the class averages, the 35S U5 snRNP appears to be built up around an elongated central domain with surrounding structural elements, including a hook-like extrusion (Fig. 3.40B, indicated by an arrow).

As might be expected from their different protein compositions, the 35S U5 snRNP differs also structurally from the 20S U5 snRNP. As shown previously, EM images of U5 snRNPs purified by H2O immunoaffinity chromatography from HeLa nuclear extract, followed by glycerol-gradient

centrifugation, showed an elongated shape with a maximum diameter of 26.5 nm (Sander et al., 2006). Although of similar length, the 20S U5 snRNP appears very narrow, whereas the 35S U5 particle is more compact. Thus, EM analysis of the more complex 35S U5 snRNP reveals additional structural elements not present in the 20S U5 snRNP.

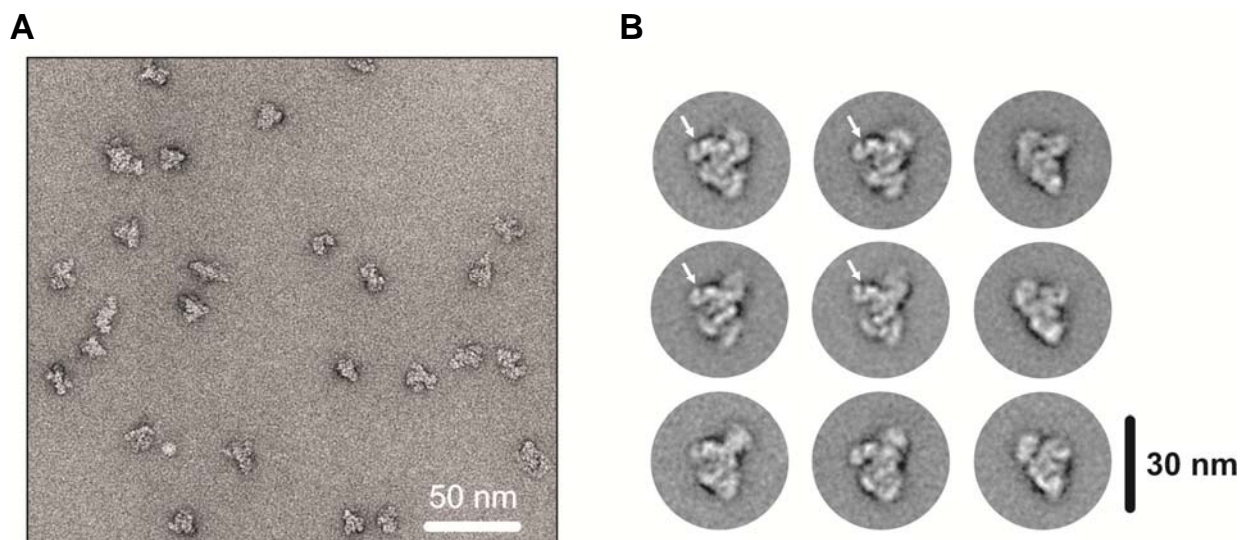


Fig. 3.40: Electron microscopy of native human 35S U5 snRNPs. (A) Electron micrograph of negatively-stained 35S U5 snRNPs. The images were taken at a primary magnification of 105,000 under low-dose conditions. Bar, 50 nm. (B) A gallery of individual class averages of native 35S U5 snRNPs obtained by averaging individual raw images. Bar, 30 nm. A hook-like extrusion from the central domain is indicated by an arrow. EM analysis of the 35S U5 snRNP was performed in the 3D Electron Cryomicroscopy group of Prof. Dr. Holger Stark (MPI for Biophysical Chemistry) by Drs. Monika M. Golas and Björn Sander.

The 3D map of the 35S U5 snRNP was reconstructed to a resolution of 2.4-2.8 nm by Drs. Monika M. Golas and Björn Sander (3D Electron Cryomicroscopy group of Prof. Dr. Holger Stark, MPI for Biophysical Chemistry), by performing unstained cryo-EM (Fig. 3.41). Overall, the 35S U5 snRNP has dimensions of about 27 nm × 26 nm × 23 nm and appears to be built up around an elongated central domain with surrounding structural elements. The 3D structure corresponds to a particle with an estimated enclosed molecular mass of ~2.4 MDa, in agreement with the composition of the particle.

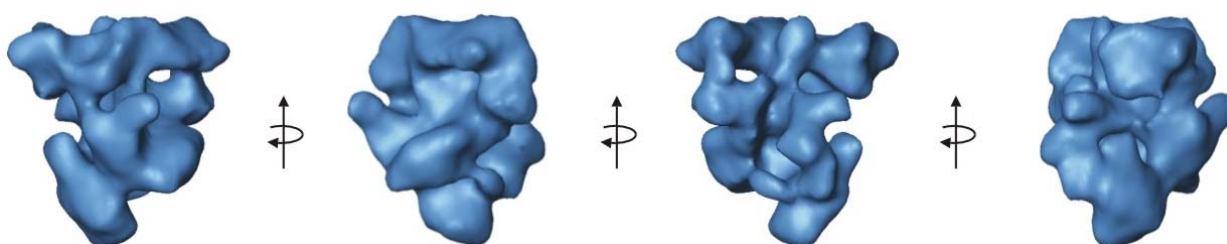


Fig. 3.41: 3D EM structure of the human 35S U5 snRNP. The structure was determined from unstained, vitrified cryo images. The figure shows four images separated by 90° rotations as indicated. 3D reconstruction of the 35S U5 snRNP was performed in the 3D Electron Cryomicroscopy group of Prof. Dr. Holger Stark (MPI for Biophysical Chemistry) by Drs. Monika M. Golas and Björn Sander.

4 DISCUSSION

In this study, we affinity purified the human Prp19/CDC5L complex from HeLa cell lines stably expressing FLAG-AD002 or FLAG-SPF27 and determined its molecular architecture and EM structure. By analyzing the stoichiometry of the hPrp19/CDC5L complex we could show that, like in the yeast NTC, the human Prp19/CDC5L complex contains four copies of hPrp19. All other components are present in single copies. Salt treatment revealed a stable core complex composed of CDC5L, PRL1, hPrp19 and SPF27. Indeed, we identified an extensive protein-protein interaction network among the core components of the hPrp19/CDC5L complex by a variety of methods, including protein-protein crosslinking and co-precipitation studies. Furthermore, interactions of the more loosely associated components AD002, CTNNB1 and HSP73 with the hPrp19/CDC5L complex core were also elucidated. In addition, protease treatment coupled with MS allowed us to delineate those domains of the core proteins involved in protein-protein interactions, as well as to identify structurally stable domains that are good candidates for high resolution structural investigation. Consequently, we determined the crystal structure of the WD40 domain of hPrp19 at a resolution of 1.8 Å. Finally, we affinity-purified the post-spliceosomal 35S U5 snRNP and determined its structure by EM.

Taken together, our data provide first insights into the molecular organization and structure of the human Prp19/CDC5L complex with interesting implications for the assembly of this heteromeric complex. As the core components of the hPrp19/CDC5L complex are also part of the stable RNP core of the spliceosomal C complex, our data also potentially provide information about the organization of the spliceosome's catalytic RNP core. Furthermore, a comparison of the 35S U5 snRNP's EM structure with that of the C complex allowed the localization of functionally important components of the step I spliceosome, as described below.

4.1 Establishing HeLa stable cell lines

Recent advances in cloning and cell line technologies have allowed the application of affinity purification tags to metazoan proteins. This provides a powerful tool for the purification and characterization of mammalian multi-protein complexes that are not amenable for overexpression in orthologous systems. Historically, epitope tags have been added to the termini of proteins to

allow their purification with anti-epitope antibodies or affinity resins. Several objectives have to be considered before generating a stable cell line for the purification of protein complexes like the hPrp19/CDC5L complex. In general, the choice of a subunit for epitope-tagging and subsequent affinity purification is largely an empirical matter. Nonetheless, a prerequisite for the heterologous expression of a protein in HeLa cells is that overexpression of the protein is not toxic for the cell. Furthermore, the protein (and in particular the epitope tag) must be accessible to the antibody or affinity resin (i.e., the protein must not be buried deep within the multiprotein complex). In addition, the selected protein must be a bona fide, integral subunit such that sufficient amounts of relatively homogenous complexes can potentially be purified. The AD002 and SPF27 proteins fulfill all of these requirements and were successfully used for the establishment and propagation of HeLa cell lines stably expressing FLAG/HA-tagged AD002 or SPF27. However, it was not possible to generate HeLa cell lines stably expressing other components of the hPrp19/CDC5L complex (e.g. hPrp19, CDC5L or PRL1; Dr. Ira Lemm, personal communication). As hPrp19 and CDC5L also play important roles in several other cellular processes (Bernstein and Coughlin, 1997; Lei et al., 2000), the overexpression of these proteins might impair cell growth and therefore be toxic for the cell.

4.2 Purification of the hPrp19/CDC5L complex

In earlier studies, hPrp19 and CDC5L containing complexes were isolated via immunoaffinity chromatography using polyclonal antibodies raised against AD002 or CDC5L (Ajuh et al., 2000; Makarova et al., 2004). However, the availability and quality of these antibodies varied between different antibody preparations and only small amounts of these complexes could be obtained by these methods. This limited structural investigations of the human Prp19/CDC5L complex. As an alternative strategy we isolated the hPrp19/CDC5L complex from HeLa cell lines stably expressing FLAG/HA-AD002 or FLAG/HA-SPF27 using anti-FLAG immunoaffinity purification. For unknown reasons, it was not possible to purify the complex utilizing the HA-tag (data not shown). Since the HA-tag immediately proceeds the N-termini of both proteins (see Fig. 3.1), we suggest that the tag is sterically hidden within the folded protein and therefore not available for purification. However, the FLAG-tag is located at the very N-terminus of the protein sequence preceding the HA-tag. In this way, the HA-tag likely functions as a linker that provides enough space for the FLAG-tag to be accessible to the anti-FLAG antibodies. Importantly, extracts prepared from both stable cell lines exhibited the same splicing activity compared to wild-type extract (Fig. 3.2A). Furthermore, both FLAG-tagged proteins are incorporated into spliceosomes to the same extent as their non-tagged endogenous counterparts (Fig. 3.2B). Thus, insertion of a tag at the N-terminus of AD002 or SPF27 does not appear to alter their activity, nor that of the hPrp19/CDC5L complex, indicating it is also

structurally intact. Indeed, MS analysis of anti-FLAG affinity-purified hPrp19/CDC5L complexes after glycerol gradient centrifugation revealed a protein composition identical to that previously described for immunoaffinity purified complexes (Makarova et al., 2004) (Fig. 3.3B and C). However, the proteins NPW38 and NPW38BP that previously appeared to co-purify with the human Prp19/CDC5L complex (but dissociate upon gradient centrifugation) were not found.

Interestingly, dimers consisting of AD002-CTNNB1 or SPF27-hPrp19 were co-isolated together with hPrp19/CDC5L complexes from nuclear extracts of the FLAG/HA-AD002 or FLAG/HA-SPF27 stable cell line, respectively. This was a first indication that these proteins interact with one another; indeed AD002-CTNNB1 and SPF27-hPrp19 interactions within the hPrp19/CDC5L complex were confirmed via crosslinking, far western and/or co-immunoprecipitation assays. A stable interaction between Prp19 and Snt309 (the yeast homolog of SPF27) was also found in the yeast *S. cerevisiae* (Chen et al., 1999). So far, we cannot rule out that both binary complexes dissociated from the hPrp19/CDC5L complex during gradient centrifugation. However, an AD002/CTNNB1 complex was not observed after glycerol gradient centrifugation of complexes purified from the FLAG/HA-SPF27 containing nuclear extract (Fig. 3.3C). Conversely, a binary complex comprised of SPF27/hPrp19 was not detected after centrifugation of complexes purified from FLAG/HA-AD002 containing nuclear extract (Fig. 3.3B). Furthermore, a AD002/CTNNB1 binary complex was identified also after gel filtration (as opposed to gradient centrifugation) of affinity-purified hPrp19/CDC5L complexes, suggesting that this binary complex is indeed present in nuclear extract from the FLAG/HA-AD002 stable cell line (Fig. 3.4) and did not dissociate due to the high centrifugal forces used for glycerol gradient centrifugation. Thus, the presence of both binary complexes in nuclear extracts raises the interesting possibility that they represent pre-formed building blocks used during the biosynthesis of the hPrp19/CDC5L complex.

When FLAG/HA-AD002 was overproduced in HeLa cells, free FLAG/HA-AD002 was found in large amounts at the top of the gradient (Fig. 3.3B). In contrast, free FLAG/HA-SPF27 was not observed (Fig. 3.3C). Rather, all SPF27 was found either in the hPrp19/CDC5L complex or in association with the hPrp19 protein. This suggests there may be a feedback mechanism to prevent free SPF27 from accumulating in cells or alternatively that free SPF27 forms large aggregates and is lost during nuclear extract preparation. A similar phenomenon was reported for the yeast Snt309 protein (the yeast homolog of SPF27) (Chen et al., 1999). In this case, the majority of the protein was found in the cell debris pellet during preparation of splicing extracts. When Snt309 was overproduced together with Prp19, the protein became soluble in the extract, presumably in the form of a binary Snt309/Prp19 complex. The reason why free Snt309 is insoluble in the yeast total cell extracts is not clear. One possibility is that Snt309 is prone to self-interaction and thereby forms large

aggregates when not in complex with Prp19. However, no self-interaction could be observed for Snt309 nor for its human counterpart SPF27.

4.3 The stoichiometry of the hPrp19/CDC5L complex

Quantification of the hPrp19/CDC5L complex stoichiometry by radioactive iodoacetamide labeling and Sypro Ruby staining, along with analytical ultracentrifugation and crosslinking consistently indicated that, like in the yeast NTC (Ohi et al., 2005), there are also four copies of hPrp19 in the human Prp19/CDC5L complex (Fig. 3.5). HSP73 appears to be underrepresented; thus, it may be very loosely associated and partially lost during purification. All other proteins appear to be present in single copies. These results are generally consistent with quantitative MS analyses of the hPrp19/CDC5L complex (employing labeled internal standard peptides for absolute quantification), except that CDC5L was found in 1 or 2 copies depending on the conditions used for protein denaturation prior to protease digestion and subsequent absolute quantification by MS (Schmidt et al., 2010). Interestingly, CDC5L can interact with itself, at least when it is present as an individual protein (Fig. 3.11; Graub et al., 2008). In the yeast NTC, Cef1 (the yeast homolog of human CDC5L), appears to be present as a single copy (Ohi et al., 2005). Given that only one copy of CDC5L is present in our hPrp19/CDC5L complexes, CDC5L dimerization and CDC5L association with the hPrp19/CDC5L complex would appear to be mutually exclusive. As CDC5L appears to have multiple roles in the cell (Bernstein and Coughlin, 1997; Lei et al., 2000), dimerization may be important for one or more of CDC5L's additional functions.

Iodoacetamide modifies cysteine residues of a protein. AD002 contains only one cysteine, and due to the low signal intensity, the signal obtained via iodoacetamide labeling of AD002 was not linear over the concentration range analyzed (Fig. 3.5B). Thus, a reliable quantification based on iodoacetamide labeling was not possible for AD002. Nevertheless, Sypro Ruby staining revealed the presence of a single copy of AD002 in the native hPrp19/CDC5L complex. The Sypro Ruby dye is a permanent fluorescent stain comprised of ruthenium as part of an organic complex that interacts noncovalently with proteins. This protein stain is particularly suitable for quantitation of protein amount in polyacrylamide gels because of its high sensitivity and broad linear dynamic range (Berggren et al., 2000). Furthermore, the presence of a single copy of AD002 and SPF27 in the native hPrp19/CDC5L complex is also supported by our affinity-purification results (Fig. 3.3). That is, only tagged AD002 or tagged SPF27, but not their endogenous, untagged counterparts, were present in purified complexes, indicating that the hPrp19/CDC5L complex contains only one copy of each of these proteins.

As already described earlier, we also detected a binary complex consisting of hPrp19 and SPF27. The migration behaviour of the hPrp19-SPF27 binary complex during glycerol gradient centrifugation, as well as gel filtration, indicates that it is larger than the AD002-CTNBL1 complex (compare Fig. 3.3B and C). Thus, it is feasible that hPrp19 is present as a tetramer already in this binary complex. In support of this finding it was shown for human, as well as for yeast, Prp19 that it does not require any other protein factor for oligomerization *in vitro* (Grillari et al., 2005; Ohi et al., 2005).

4.4 Identification of a highly stable core of the hPrp19/CDC5L complex

A subcomplex of the hPrp19/CDC5L complex, consisting of CDC5L, PRL1, hPrp19 and SPF27, was stable in the presence of 1.2 M NaCl, 0.4 M NaSCN and 0.1 mg/ml heparin (Fig. 3.7 and Fig. 3.8). In addition, fragments of these proteins were found together in a protease resistant complex. The existence of an extensive interaction network between the stable core components of the hPrp19/CDC5L complex (CDC5L, PRL1, hPrp19 and SPF27) was confirmed by several independent approaches, including crosslinking, far western blots and *in vitro* pulldown analysis. Prior to our analyses, the only known molecular interaction involving proteins of the human Prp19/CDC5L complex core was that between the C-terminus of CDC5L and the WD40 domain of PRL1 (Ajuh et al., 2001). In addition to this known interaction, we observed several novel interactions (i.e., hPrp19-CDC5L, hPrp19-PRL1, SPF27-hPrp19, SPF27-PRL1 and SPF27-CDC5L) by a variety of methods. A model of the human Prp19/CDC5L complex, which incorporates all of the information presented here, is shown in Fig. 3.15. Interestingly, some, but not all, of the aforementioned intermolecular interactions are also observed between the yeast homologs of CDC5L, hPrp19, PRL1 and SPF27, which are all components of the yeast NTC (Chen et al., 2002; Tarn et al., 1994). For example, Cef1 (the yeast homolog of human CDC5L) interacts with both Prp19 and Prp46 (the yeast homolog of human PRL1), but it does not interact with Snt309 (the yeast homologue of SPF27). Snt309 interacts with the Prp19 tetramer and Prp46 interacts with Cef1 (Chen et al., 2002; Ohi and Gould, 2002), but neither of these proteins interacts with the yeast homologues of any of the other core components of the human Prp19/CDC5L complex. Thus, protein-protein interactions involving these common proteins appear to differ somewhat between yeast and humans, and the overall architecture of the NTC core, at least in part, may also differ. Indeed, the yeast NTC contains seven additional proteins not found in the human complex (Syf1, Syf2, Syf3, Isy1, NTC20 and Cwc2), some of which also contact Prp19, Cef1, or Prp46 (Ohi and Gould, 2002) and at present it is not clear which of the yeast NTC proteins comprise its salt-stable core. However, Prp46 is readily lost from the NTC upon gradient centrifugation (Ohi and Gould, 2002; Tarn et al., 1994) and thus it is not a highly stable NTC component (in contrast to its human homologue PRL1). Taken together, these

results point to clear differences in the organization of the stable core of human versus yeast Prp19 complexes. Although four proteins are shared by the NTC and hPrp19/CDC5L complex, a completely different set of additional proteins is recruited to each complex. This likely is a consequence of the presence of additional or different domains in the human proteins. Whereas the length and domain structure of Prp19 is highly conserved between yeast and man, this is not the case for CDC5L and SPF27. For example, human CDC5L is 803 amino acids long, compared to yeast Cef1 which is only 590 amino acids in length. While the N-terminal half of these proteins shares over 40% homology and contains two SANT (c-Myb) domains, the C-terminal half is much less conserved, with CDC5L containing a second coiled-coil domain at its very C-terminus that is lacking in Cef1 (Fig. 1.7B; data not shown). Likewise, SPF27 and Snt309 share very limited homology (24%) and SPF27 contains a coiled-coil domain that is not found in Snt309 (Fig. 1.7B; data not shown). These differences also likely contribute to the dissimilarities in interactions between SPF27 versus Snt309 (or CDC5L vs Cef1) and the other shared proteins of both complexes.

AD002, CTNNB1 and HSP73 dissociate from the core complex already at moderate salt concentrations (e.g. 300 mM NaCl). In contrast to the components of the stable hPrp19/CDC5L complex core, homologs of human CTNNB1, AD002 and HSP73 either do not exist in yeast (CTNNB1) or are not found in the yeast NTC (AD002 and HSP73), but are nonetheless present in the yeast spliceosome (Chen et al., 2002; Fabrizio et al., 2009; Tarn et al., 1994). AD002 and CTNNB1 interact with one another (Fig. 3.3, Fig. 3.7, Fig. 3.11, Fig. 3.13, Fig. 3.14), and CTNNB1 also interacts with CDC5L (Fig. 3.11, Fig. 3.13). Furthermore, CTNNB1 is found together with the N-terminus of CDC5L in a protease resistant complex (Fig. 3.17). Thus, CTNNB1 (and its interaction partner AD002) are likely linked to the hPrp19/CDC5L complex core, at least in part, via the N-terminus of CDC5L (Fig. 3.20). No direct interaction involving HSP73 could be observed via crosslinking, far western, or co-immunoprecipitation assays. Nonetheless, HSP73 appears to migrate in a complex together with CTNNB1 and the N-terminus of CDC5L after protease digestion, suggesting it associates with the hPrp19/CDC5L complex via interacting with one or more of these proteins (Fig. 3.20).

4.5 Elucidation of domains of the hPrp19/CDC5L proteins involved in protein-protein interactions

Limited protease digestion, followed by gel filtration and MS analyses revealed a protease resistant complex of the human Prp19/CDC5L complex that is composed of nearly full-length SPF27, the N-terminal U-box containing domain of hPrp19, the N-terminus of PRL1 and a C-terminal fragment of CDC5L. Interestingly, the U-box domain and ~80 amino acids downstream of it are highly

homologous in both humans and yeast Prp19 (Fig. 3.29), and these regions were shown to be sufficient for tetramerization of yeast Prp19 (Ohi et al., 2005), with the U-box domain alone capable of forming a dimer *in vitro* (Ohi et al., 2003). This supports the conclusion that the N-terminal fragment of hPrp19 is indeed present in the form of a tetramer in this protease resistant complex. The C-terminus of CDC5L could contact both hPrp19 and SPF27 in this complex. In contrast, as the C-terminus of CDC5L was previously shown to bind the C-terminal WD40 domain of PRL1 (and this region is lacking in the protease resistant complex), it likely does not contact PRL1. Indeed, the C-terminal WD40 domain of PRL1 is completely digested after treatment with low concentrations (30 nM) of Subtilisin (Fig. 3.16, Fig. 3.17). The N-terminus of PRL1, which contains no known motifs (Fig. 1.7B), potentially contacts the N-terminus of hPrp19 and/or SPF27. Finally, SPF27 potentially contacts all other components of the protease resistant core. Indeed, in contrast to free SPF27, it remains nearly intact after protease treatment, suggesting that it is either well-protected against protease digestion (and thus potentially buried within the protease-resistant complex) or that it becomes highly-structured (and thus more resistant to protease digestion) upon integration into the hPRP19/CDC5L complex. This is further consistent with the idea that it potentially plays an important role as a scaffold protein (together with hPrp19) and/or in maintaining the integrity of the hPrp19/CDC5L complex. That is, in contrast to yeast, where Snt309 seems to interact solely with the Prp19 tetramer (Chen et al., 1999), and is only required for yeast cell growth at elevated temperatures (Chen et al., 1998), the human SPF27 protein might play a more central role, interacting with all other core components, apparently via those regions identified by MS in the protease resistant complex.

4.6 The human Prp19/CDC5L complex has an asymmetric and elongated morphology

EM of the human Prp19/CDC5L complex revealed an elongated, asymmetric shape with a maximum dimension of ~20 nm and a width of ~9 nm. This agrees well with our analytical ultracentrifugation data, which revealed a frictional ratio value of 2.1 for the complex, indicating that it is indeed elongated. Previous studies showed that recombinant Prp19 from the yeast *S. cerevisiae* forms a tetramer and EM analyses indicated that this tetrameric complex is an elongated particle with a side length of ~40 nm (Ohi et al., 2005). The reason for this apparent difference in size between the yeast Prp19 tetramer and the entire human Prp19/CDC5L complex is not clear. Prp19 is also present as a tetramer in the human Prp19/CDC5L complex and both proteins have a similar molecular weight and overall domain structure. Thus, the presence of much smaller and/or fewer copies of Prp19 in the human complex cannot explain this difference. It is possible, that the human

complex is very compact and that the hPrp19 tetramer is highly condensed due to its interaction with other components. Furthermore, in contrast to the yeast Prp19 tetramer, the human Prp19/CDC5L complex was prepared under GraFix conditions, which was shown to stabilize individual macromolecules, and thereby provide reliable and high-resolution structures of macromolecular complexes by EM (Kastner et al., 2008). The presence of the crosslinking reagent thus might fix the human complex in its native, but more compact conformation.

The asymmetric shape of the hPrp19/CDC5L complex, together with the presence of an hPrp19 tetramer but apparently single copies of the other components of the complex, suggests an asymmetric assembly pathway. That is, although four copies of hPrp19 are present, they are not recognised by the same interaction partner multiple times. This suggests that hPrp19 tetramerization leads to the formation of a unique interaction interface which recruits one or more of the other core components of the hPrp19/CDC5L complex. As a dimer consisting of SPF27 and what appears to be tetrameric hPrp19 (based on its migration behaviour and the staining intensity of the hPrp19 band) was present in nuclear extract prepared from cells expressing FLAG-tagged SPF27 (Fig. 3.3C), it is tempting to speculate that SPF27 initially binds the hPrp19 tetramer and that this interaction in turn would promote the interaction of CDC5L and PRL1. Indeed, the association of Cef1 (the yeast homolog of human CDC5L) with the yeast NTC requires Prp19 oligomerization (Tarn et al., 1994) and Snt309 (the yeast homolog of human SPF27) is required to maintain the Prp19-Cef1 interaction, at least at elevated temperatures (Chen et al., 1999). However, as we showed that CDC5L and PRL1 interact with hPrp19 on their own *in vitro*, it is also conceivable that they bind independently of SPF27 and even potentially prior to hPrp19 tetramerization. Finally, the AD002-CTNNB1 dimer alone or possibly together with HSP73, likely binds after formation of the stable hPrp19/CDC5L core. However, future *in vitro* reconstitution studies with purified recombinant proteins are needed to shed more light on the assembly pathway of the hPrp19/CDC5L complex.

4.7 Insight into a major building block of the RNP core of the step I spliceosome

Taken together, this work provides a detailed view of the architecture of the human Prp19/CDC5L complex and provides a framework for future high resolution structural analyses. Furthermore, the elucidation of the molecular organization of the hPrp19/CDC5L complex provides a first insight into the possible architecture of a major building block of the RNP core of the catalytically active spliceosome. That is, all proteins of the highly stable core of the hPrp19/CDC5L complex, in addition to AD002, plus ~30 other proteins, are present in the 1M salt-stable RNP core of the human

spliceosomal C complex (Bessonov et al., 2008), and they are thus thought to help maintain its catalytically active RNA network. In addition, with the exception of CTNNB1, all proteins of the hPrp19/CDC5L complex are also present in the 35S U5 snRNP, a presumed dissociation product of the post-spliceosomal intron-containing complex (Makarov et al., 2002). Based on its exceptional stability, it is highly likely that the complex protein interaction network found in hPrp19/CDC5L complex core is also conserved in the core of the step I spliceosome.

4.8 The crystal structure of the WD40 domain of hPrp19

WD40-repeat containing proteins are abundant and play important roles in many biological functions, including cell cycle regulation, transcription, epigenetic regulation of histones, DNA damage repair, protein ubiquitination and apoptosis (reviewed in Smith et al., 1999). It is well known that WD40-repeat proteins mainly function by interacting with other proteins. At present, only a small number of crystal structures of WD40 repeat proteins in complex with their binding protein are available. Thus, it is poorly understood how WD40 domain containing proteins recognize their binding partners. The propeller structure provides extensive surface exposure of three surfaces: upper, lower and the circumference. The surfaces of this structure appear to provide a stable platform for several simultaneous protein-protein interactions (reviewed in Li and Roberts, 2001).

Limited proteolysis of the human Prp19/CDC5L complex identified a ~37 kDa stable domain of the complex that accumulated with increased protease concentration and MS analysis revealed that it is the WD40 domain of the hPrp19 protein. Solving the 1.8 Å crystal structure of this hPrp19 domain revealed that the WD40 domain of hPrp19 folds into a classical seven-bladed β -propeller structure, as observed for other WD40-repeat proteins. Recent structure determination of the yeast Prp19 WD40 domain (Vander Kooi et al., 2010), on the other hand, revealed some structural variability. As visualized in Fig. 3.33, the human WD40 domain is nearly symmetrically folded, whereas its yeast counterpart shows some deviations from this ideal geometry. Although the yeast WD40 domain also folds into a seven-bladed β propeller, the domain has a distorted architecture. This is mainly caused by an irregular packing of blades one and seven that leads to a significant asymmetry in its architecture. The human Prp19 WD40 domain was purified from native hPrp19/CDC5L complexes whereas the WD40 repeat domain of yeast Prp19 was subcloned, expressed and purified from *E. coli* cells. It cannot be excluded that expression in an orthologous system in the case of the yeast protein leads to the structural differences between both domains.

The presence of a peptide of unknown identity that binds to the bottom surface of the β -propeller of human Prp19 indicates that the bottom surface functions as a protein-protein interaction site

(Fig. 3.35). This contrasts with several other WD-repeat proteins that use their top faces for interacting with their protein binding partners (Valeyev et al., 2008). However, we cannot exclude that the top surface of the hPrp19 β -propeller is also engaged in interactions with other proteins. So far, we can only speculate about the origin of the peptide. Since the WD40 domain is a cleavage product of the entire hPrp19/CDC5L complex, it is possible that the peptide interacting with it is derived from one of the other components of the complex. However, MS analysis revealed the presence of an additional 47 amino acids at the N-terminus of the WD40 domain that are not represented in the electron density map. Thus, this peptide could also be part of this region of the hPrp19 protein that folds back onto the bottom surface of the β -propeller.

An intriguing structural feature of WD40 domains is the presence of an Asp-His-Ser/Thr-Trp (D-H-S/T-W) hydrogen-bonded structural tetrad. An analysis of 20 available WD40-repeat proteins revealed that each protein has at least one D-H-S/T-W hydrogen-bonded tetrad and some proteins have up to seven such tetrads, one for each blade of the β -propeller (Wu et al., 2010). In addition to the formation of strong hydrogen bonds, the relative positions of the four residues in such tetrads are also highly conserved. The WD40 domain of human Prp19 contains two structural tetrads (Fig. 3.32). A comparison of the two structural tetrads that are located in blade 2 and blade 5 of the WD40 domain clearly supports the high conservation of this motif. Interestingly, the hydrogen bonding network of the D-H-S/T-W tetrad closely resembles the hydrogen bonding of the catalytic triads found in serine/threonine proteases that are essential for peptide hydrolysis by deprotonating Ser/Thr to make it a stronger nucleophile through the hydrogen-bond network. Furthermore, noncatalytic triads were found to function as protein conformational determinants in proteins (Marx et al., 2008). Thus, the hydrogen-bond network formed by the tetrad may play an important role in maintaining the structural integrity of the WD40 domain.

Determination of the stoichiometry of the human Prp19/CDC5L complex showed that four copies of hPrp19 are present in the complex. We could further delineate several interactions of hPrp19 with a number of hPrp19/CDC5L complex components and also with itself. The ability of hPrp19 to interact with numerous hPrp19/CDC5L components suggests that the hPrp19 tetramer acts as an essential scaffold for the organization of the entire complex and perhaps other associated factors. At present, the only protein known to interact with yeast Prp19 via the WD40-domain is Cwc2 (Ohi and Gould, 2002; Vander Kooi et al., 2010) and this interaction is mediated by blade five (Vander Kooi et al., 2010). Interestingly, this blade shows the highest sequence conservation among Prp19 WD40 domains from different species (Fig. 3.29). In the hPrp19 WD40 domain several amino acid residues are also highly conserved. Consistent with the yeast structure, these residues are exposed on the bottom surface of the propeller. This suggests a function for these residues in mediating protein-protein interactions also in the human spliceosome. Yeast Cwc2 shows significant sequence

conservation with the human spliceosomal protein RBM22. Thus, analyzing whether RBM22 binds the hPrp19 WD40 domain represents an interesting future direction that may provide insight into the association of the hPrp19/CDC5L complex with other spliceosome components. Indeed, recent data showed that yeast Cwc2 can be directly crosslinked to the U6 snRNA, before and after the first step of splicing (McGrail et al., 2009). Therefore, it was proposed that Cwc2 may link the yeast NTC to the spliceosome during pre-mRNA splicing via the U6 snRNA (McGrail et al., 2009). Although we cannot exclude that the WD40 domain of human Prp19 interacts with a component of the hPrp19/CDC5L complex, it is likely that it provides a binding platform for interaction with other spliceosomal proteins. The ability to cleave the WD40 domain already at low protease concentration suggests that the globular WD40 domain is flexibly tethered to the central domain of hPrp19. This flexibility may be important for both binding and integrating the hPrp19 complex into the spliceosome or to facilitate the numerous transitions that occur during formation of the catalytically active spliceosome after binding of the hPrp19/CDC5L complex. In the latter case, this flexibility could allow hPrp19 to accommodate the structural rearrangements occurring during the dynamic reorganization of spliceosome components without influencing binding interactions with other components within the hPrp19/CDC5L complex.

In addition to its WD40 domain, human Prp19 also contains a U-box domain at its N-terminus and hPrp19 was shown to possess ubiquitin ligase activity *in vitro* (Hatakeyama et al., 2001). Interestingly, WD40-repeat containing proteins were shown to be able to recruit substrates for ubiquitination (Angers et al., 2006). Thus, it is possible that hPrp19 combines both functions in one protein; on the one hand substrate recognition via the WD40 domain and on the other hand ubiquitination via its U-box domain. Tetramerization of hPrp19 would further increase its ability to recognize substrates for ubiquitination. In support of this idea, it was shown recently that the human Prp19 complex is responsible for Prp3 ubiquitination (Song et al., 2010). Furthermore, yeast Prp8 is reversibly ubiquitylated and it was suggested that Prp19 may also play a role in the ubiquitination of yeast Prp8 (Bellare et al., 2008), but formal proof of this is currently lacking.

4.9 Purification of the 35S U5 snRNP

As described in detail for the hPrp19/CDC5L complex, the human 35S U5 snRNP was purified from HeLa nuclear extract containing FLAG/HA-tagged AD002. Consistent with previous studies from our lab, MS analysis of the 35S U5 snRNP showed that AD002 and also SPF27 are both bona fide components of the 35S U5 snRNP. However, only small amounts of 35S U5 snRNPs could be obtained from the FLAG/HA-SPF27 containing extract (data not shown). This observation suggests

that SPF27 is deeply embedded in the 35S U5 snRNP, thereby masking the tag for immunoaffinity purification.

35S U5 snRNPs purified by anti-FLAG immunochromatography mainly contained U5 snRNA and exhibited essentially the same protein composition as found in our earlier study (Makarov et al., 2002). Only a few proteins, such as S100A8/A9, KIAA0773 and FLJ10206, which were present in substoichiometric amounts in our earlier preparations (Makarov et al., 2002), were absent from the preparations described here. In fact, these proteins have never been observed in native spliceosomal complexes, indicating that they were contaminants. On the other hand, the proteins Morg1, Q9BRR8, hPrp22 and Slu7, which were identified in the 35S U5 snRNP described here, were absent from the previous preparations. These minor differences are likely due to the use of different antibodies during immuno-affinity purification of the 35S U5 snRNP.

4.10 The proteome of the 35S U5 snRNP resembles the composition of the salt-stable RNP core of the spliceosome

MS analysis of the 35S U5 snRNP revealed striking similarities between its protein composition and that of the salt-stable RNP core of a catalytically active step I spliceosome (Bessonov et al., 2008). For example, both complexes contain the four U5 proteins 220K/hPrp8, 200K/hBrr2, 116K/hSnu114 and 40K, as well as the hPrp19/CDC5L complex and its related factors (Fig. 4.1). Furthermore, the 35S U5 snRNP contains only a handful of proteins – such as Slu7, hPrp22, DDX35, Q9BRR8, Morg1 and GCIPp29 – that are not present in the RNP core of the C complex. A major compositional difference between the 35S U5 and the C complex's salt-stable RNP core particle is the absence of U2 snRNP and its associated proteins, most of the U6 snRNA and the pre-mRNA splicing intermediates; notably, regions of these RNAs comprise the catalytically active RNA network and thus the active site of the C complex (Nilsen, 1998). However, the catalytic core also comprises the U5 snRNA and the protein U5-220K/hPrp8 present in both complexes. The high similarity between the protein composition of the 35S U5 snRNP and the salt-stable C complex core supports the idea that the 35S U5 snRNP is a remodeled form of the U5 snRNP that is generated during spliceosome activation and is released after exon ligation as a post-spliceosomal complex. Consistent with this idea a 40S post-splicing intron-containing complex isolated from HeLa cells containing U2, U5 and U6 snRNAs, also contained most proteins of the 35S U5 snRNP (Yoshimoto et al., 2009). This hypothesis is further backed by the presence of small amounts of U6 snRNA in our anti-FLAG affinity-purified 35S U5 snRNPs (Fig. 3.38).

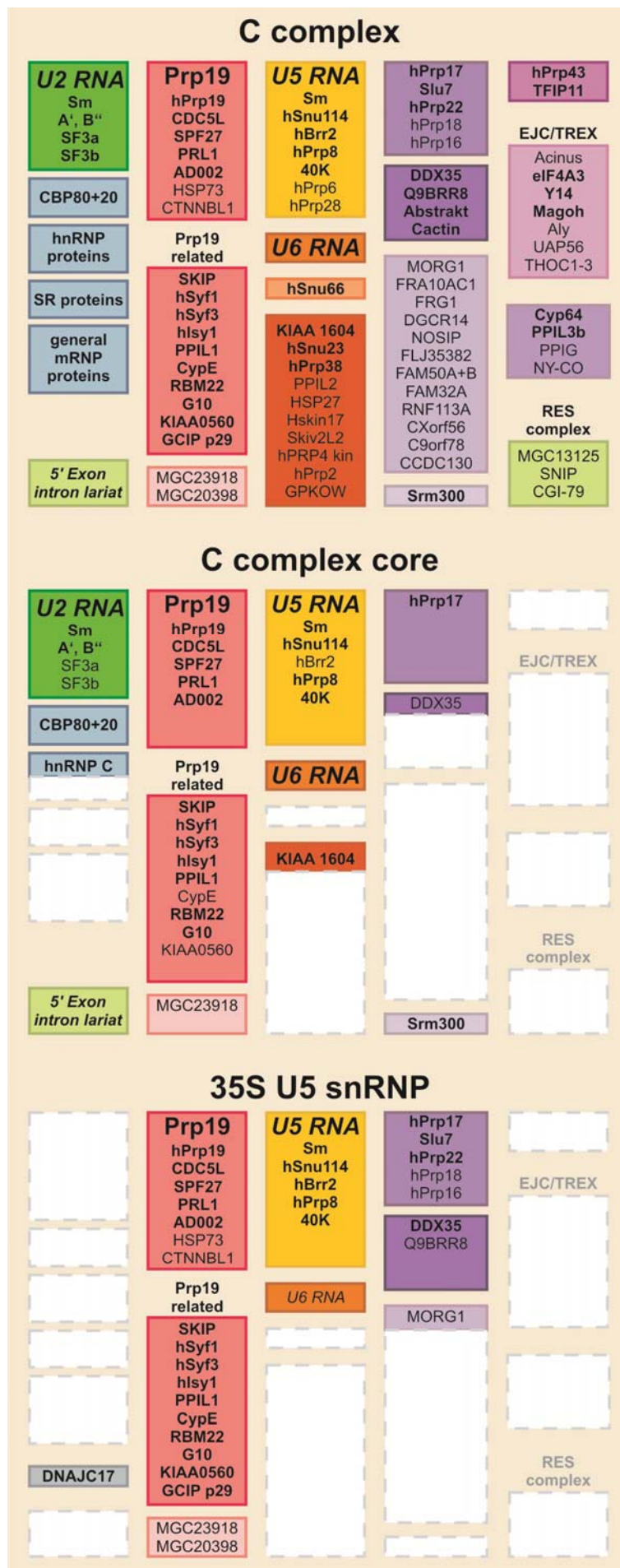


Fig. 4.1: Comparison of the protein and RNA composition of the catalytically active step I spliceosome, the salt-stable RNP core of the C complex and the 35S U5 snRNP. The figure shows proteins grouped according to function or complex association, along with the RNA content of the respective complex. The relative abundance of a protein is indicated by light (substoichiometric amounts) or dark (abundant) lettering and is based on the relative amounts of peptides sequenced by MS. SR proteins identified in the step I spliceosome include ASF/SF2, 9G8, hTra-2, SRp30c, SRp38, SRp55 and to a minor extent SRp20, SRp40 and SRp75. hnRNP proteins include predominantly hnRNP C, G, M and PCBP1. mRNP proteins include predominantly ASR2B, NFAR, BCLAF1, YB-1, HuR and LOC124245.

4.11 Comparison of the EM structure of the 35S U5 snRNP with that of the spliceosomal C complex

Recently, the cryo-EM 3D structure of the catalytically active human C complex was determined by our group (Golas et al., 2010). To locate functionally important components of the step I spliceosome the authors performed cryo-EM studies also on the 35S U5 snRNP that was purified during this study (Fig. 3.41). The 3D EM structures of the C complex and the 35S U5 snRNP particles are represented in Fig. 4.2A. In line with its reduced proteome, the 35S U5 snRNP complex is smaller than the C complex and shows fewer density elements.

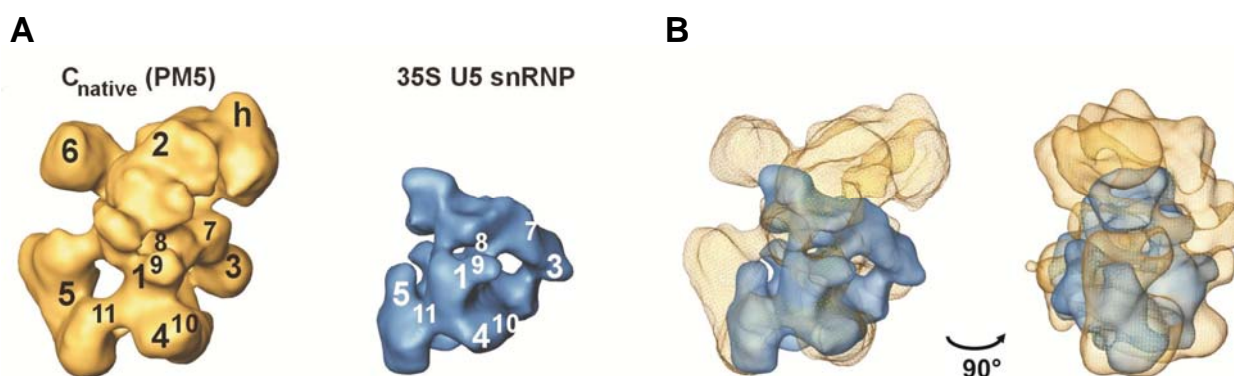


Fig. 4.2: Comparison of the 3D EM structures of the 35S U5 snRNP with the C complex. (A) 3D EM structures of the C complex (golden) and the 35S U5 snRNP (blue). The numbers indicate structural domains found in both complexes. (B) The comparison of the C complex (golden) and the 35S U5 snRNP (blue) on the 3D level reveals that in addition to domains 1, 3, 4 and the lower part of 5, additional fine structural details are consistently found in both complexes (labeled 7–11). The fit is shown from two different orientations separated by 90° as indicated. Adapted from Golas M.M., Sander B., Bessonov S., Grote M., Wolf E., Kastner B., Stark H., Lührmann R., submitted manuscript.

A fit of the 35S U5 snRNP into the C complex structure revealed that the 35S U5 snRNP resides in the lower 2/3 of the latter comprising domains 1 and 3-5 (Fig. 4.2B, blue regions in the two panels to the right). Thus, this region of the C complex contains the U5 proteins 220K/hPrp8, 200K/hBrr2 and 116K/hSnu114, which are known to function in unwinding U4/U6 during catalytic activation (Grainger and Beggs, 2005; Staley and Guthrie, 1998; Wahl et al., 2009), as well as the Prp19/CDC5L complex and its related factors, plus the U5 snRNA. These and other 35S U5 proteins are needed to maintain the catalytically active RNP core of the C complex. The 35S U5 snRNP also contains components that are thought to comprise part of the C complex's active site, namely loop 1 of the U5 snRNA, which orients the exons for their subsequent ligation (McConnell and Steitz, 2001; Newman et al., 1995) and very probably also the RNase H-like domain of Prp8 (Pena et al., 2008; Ritchie et al., 2008; Yang et al., 2008). Thus, a portion of the active site of the step I spliceosome should be located in the lower two-thirds of the 3D structure, encompassed by the 35S U5 snRNP. The clear and unambiguous fit of the 35S U5 snRNP into the lower part of the C complex shows also

that this part of the C complex must consist more or less only of components of the 35S U5 snRNP. The most obvious difference between the 35S U5 snRNP and the C complex is present in the upper part of the C complex. Domains h and 6, and a large portion of domain 2 are missing in the 35S U5 snRNP. In the lower region only the upper part of domain 5 is solely present in the C complex. Taken together, the isolation of the 35S U5 snRNP and its subsequent structure determination by EM have allowed initial insights into the potential location of functionally important components of the step I spliceosome.

5 REFERENCES

- Abelson, J. (2008). Is the spliceosome a ribonucleoprotein enzyme? *Nat Struct Mol Biol* 15, 1235-1237.
- Achsel, T., Brahm, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J* 18, 5789-5802.
- Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198-207.
- Aguilar, R.C., and Wendland, B. (2003). Ubiquitin: not just for proteasomes anymore. *Curr Opin Cell Biol* 15, 184-190.
- Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J.C., Mann, M., and Lamond, A.I. (2000). Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO J* 19, 6569-6581.
- Ajuh, P., Sleeman, J., Chusainow, J., and Lamond, A.I. (2001). A direct interaction between the carboxyl-terminal region of CDC5L and the WD40 domain of PLRG1 is essential for pre-mRNA splicing. *J Biol Chem* 276, 42370-42381.
- Anderson, K., and Moore, M.J. (1997). Bimolecular exon ligation by the human spliceosome. *Science* 276, 1712-1716.
- Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T., and Zheng, N. (2006). Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* 443, 590-593.
- Aravind, L., and Koonin, E.V. (2000). The U box is a modified RING finger - a common domain in ubiquitination. *Curr Biol* 10, R132-134.
- Azubel, M., Wolf, S.G., Sperling, J., and Sperling, R. (2004). Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol Cell* 15, 833-839.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98, 10037-10041.
- Beckmann, R.P., Mizzen, L.E., and Welch, W.J. (1990). Interaction of Hsp 70 with newly synthesized proteins: implications for protein folding and assembly. *Science* 248, 850-854.
- Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., *et al.* (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J* 26, 1737-1748.
- Bellare, P., Small, E.C., Huang, X., Wohlschlegel, J.A., Staley, J.P., and Sontheimer, E.J. (2008). A role for ubiquitin in the spliceosome assembly pathway. *Nat Struct Mol Biol* 15, 444-451.

- 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*, 3171-3175.
- Berggren, K., Chernokalskaya, E., Steinberg, T.H., Kemper, C., Lopez, M.F., Diwu, Z., Haugland, R.P., and Patton, W.F. (2000). Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. *Electrophoresis* *21*, 2509-2521.
- Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev* *12*, 858-867.
- Bernstein, H.S., and Coughlin, S.R. (1997). Pombe Cdc5-related protein. A putative human transcription factor implicated in mitogen-activated signaling. *J Biol Chem* *272*, 5833-5837.
- Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature* *452*, 846-850.
- Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. *Cell* *126*, 37-47.
- Blum, H., Beier, H., and Gross, H.J. (1987). Improved silver staining of plant proteins, RNA and DNA polyacrylamide gels. *Electrophoresis* *8*, 93-99.
- Boehringer, D., Makarov, E.M., Sander, B., Makarova, O.V., Kastner, B., Lührmann, R., and Stark, H. (2004). Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat Struct Mol Biol* *11*, 463-468.
- Boon, K.L., Norman, C.M., Grainger, R.J., Newman, A.J., and Beggs, J.D. (2006). Prp8p dissection reveals domain structure and protein interaction sites. *RNA* *12*, 198-205.
- Boyer, L.A., Langer, M.R., Crowley, K.A., Tan, S., Denu, J.M., and Peterson, C.L. (2002). Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. *Mol Cell* *10*, 935-942.
- Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of Precursors to mRNAs by the Spliceosomes. In *The RNA World*, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press), pp. 525-560.
- Cech, T.R. (1986). The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* *44*, 207-210.
- Chan, S.P., and Cheng, S.C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *J Biol Chem* *280*, 31190-31199.
- Chan, S.P., Kao, D.I., Tsai, W.Y., and Cheng, S.C. (2003). The Prp19p-associated complex in spliceosome activation. *Science* *302*, 279-282.
- Chen, C.H., Yu, W.C., Tsao, T.Y., Wang, L.Y., Chen, H.R., Lin, J.Y., Tsai, W.Y., and Cheng, S.C. (2002). Functional and physical interactions between components of the Prp19p-associated complex. *Nucleic Acids Res* *30*, 1029-1037.
- Chen, H.R., Jan, S.P., Tsao, T.Y., Sheu, Y.J., Banroques, J., and Cheng, S.C. (1998). Snt309p, a component of the Prp19p-associated complex that interacts with Prp19p and associates with the

spliceosome simultaneously with or immediately after dissociation of U4 in the same manner as Prp19p. *Mol Cell Biol* 18, 2196-2204.

Chen, H.R., Tsao, T.Y., Chen, C.H., Tsai, W.Y., Her, L.S., Hsu, M.M., and Cheng, S.C. (1999). Snt309p modulates interactions of Prp19p with its associated components to stabilize the Prp19p-associated complex essential for pre-mRNA splicing. *Proc Natl Acad Sci U S A* 96, 5406-5411.

Cheng, S.C., Tarn, W.Y., Tsao, T.Y., and Abelson, J. (1993). PRP19: a novel spliceosomal component. *Mol Cell Biol* 13, 1876-1882.

Chiang, C.M., and Roeder, R.G. (1993). Expression and purification of general transcription factors by FLAG epitope-tagging and peptide elution. *Pept Res* 6, 62-64.

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, 1-8.

Ciechanover, A. (1998). The ubiquitin-proteasome pathway: on protein death and cell life. *EMBO J* 17, 7151-7160.

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

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561-563.

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

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

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

Donlon, J. (2007). Subtilisin. In *Industrial Enzymes Structure, Function and Applications*, J. Polaina, and A.P. MacCabe, eds. (Dordrecht, Springer), pp. 197-206.

Drenth, J. (2007). *Principles of Protein X-Ray Crystallography* (New York, Springer Science+Business Media, LLC).

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

El-Baradi, T.T., Raue, H.A., De Regt, C.H., and Planta, R.J. (1984). Stepwise dissociation of yeast 60S ribosomal subunits by LiCl and identification of L25 as a primary 26S rRNA binding protein. *Eur J Biochem* 144, 393-400.

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

Endo, Y., and Sawasaki, T. (2003). High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. *Biotechnol Adv* 21, 695-713.

- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* *36*, 593-608.
- Fisher, H.W., and Puck, T.T. (1956). On the Functions of X-Irradiated "Feeder" Cells in Supporting Growth of Single Mammalian Cells. *Proc Natl Acad Sci U S A* *42*, 900-906.
- Fleckner, J., Zhang, M., Valcarcel, J., and Green, M.R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev* *11*, 1864-1872.
- Fong, H.K., Hurley, J.B., Hopkins, R.S., Miake-Lye, R., Johnson, M.S., Doolittle, R.F., and Simon, M.I. (1986). Repetitive segmental structure of the transducin beta subunit: homology with the CDC4 gene and identification of related mRNAs. *Proc Natl Acad Sci U S A* *83*, 2162-2166.
- Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., and Scaramella, E. (1997). Probing the conformational state of apomyoglobin by limited proteolysis. *J Mol Biol* *266*, 223-230.
- Fortner, D.M., Troy, R.G., and Brow, D.A. (1994). A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev* *8*, 221-233.
- Frank, J. (2006). *Three-dimensional electron microscopy of macromolecular assemblies*, 2nd edn (Oxford, Oxford University Press).
- Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., *et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* *440*, 631-636.
- Gloeckner, C.J., Boldt, K., Schumacher, A., Roepman, R., and Ueffing, M. (2007). A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. *Proteomics* *7*, 4228-4234.
- Golas, M.M., Sander, B., Bessonov, S., Grote, M., Wolf, E., Kastner, B., Stark, H., Lührmann, R. (2010). 3D Cryo-EM Structure of an Active Step I Spliceosome and Localization of its Catalytic Core. *Mol Cell* *40*, 927-938.
- Golas, M.M., Sander, B., Will, C.L., Lührmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. *Science* *300*, 980-984.
- Golas, M.M., Sander, B., Will, C.L., Lührmann, R., and Stark, H. (2005). Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol Cell* *17*, 869-883.
- Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. *RNA* *11*, 533-557.
- Graub, R., Lancero, H., Pedersen, A., Chu, M., Padmanabhan, K., Xu, X.Q., Spitz, P., Chalkley, R., Burlingame, A.L., Stokoe, D., *et al.* (2008). Cell cycle-dependent phosphorylation of human CDC5 regulates RNA processing. *Cell Cycle* *7*, 1795-1803.
- Grillari, J., Ajuh, P., Stadler, G., Loscher, M., Voglauer, R., Ernst, W., Chusainow, J., Eisenhaber, F., Pokar, M., Fortschegger, K., *et al.* (2005). SNEV is an evolutionarily conserved splicing factor whose oligomerization is necessary for spliceosome assembly. *Nucleic Acids Res* *33*, 6868-6883.

- Grune, T., Brzeski, J., Eberharter, A., Clapier, C.R., Corona, D.F., Becker, P.B., and Muller, C.W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol Cell* *12*, 449-460.
- Häcker, I., Sander, B., Golas, M.M., Wolf, E., Karagoz, E., Kastner, B., Stark, H., Fabrizio, P., and Lührmann, R. (2008). Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nat Struct Mol Biol* *15*, 1206-1212.
- Hatakeyama, S., Yada, M., Matsumoto, M., Ishida, N., and Nakayama, K.I. (2001). U box proteins as a new family of ubiquitin-protein ligases. *J Biol Chem* *276*, 33111-33120.
- Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* *268*, 1144-1149.
- Howlett, G.J., Minton, A.P., and Rivas, G. (2006). Analytical ultracentrifugation for the study of protein association and assembly. *Curr Opin Chem Biol* *10*, 430-436.
- Hutchinson, E.G., and Thornton, J.M. (1996). PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Sci* *5*, 212-220.
- Jurica, M.S. (2008). Detailed close-ups and the big picture of spliceosomes. *Curr Opin Struct Biol* *18*, 315-320.
- Jurica, M.S., Sousa, D., Moore, M.J., and Grigorieff, N. (2004). Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat Struct Mol Biol* *11*, 265-269.
- Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J Appl Cryst* *21*, 916-924.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* *26*, 795-800.
- Kaelin, W.G., Jr., Pallas, D.C., DeCaprio, J.A., Kaye, F.J., and Livingston, D.M. (1991). Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product. *Cell* *64*, 521-532.
- Karaduman, R., Dube, P., Stark, H., Fabrizio, P., Kastner, B., and Lührmann, R. (2008). Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSM complex as revealed by electron microscopy. *RNA* *14*, 2528-2537.
- Kastner, B. (1998). Purification and Electron Microscopy of Spliceosomal snRNPs. In *RNP Particles, Splicing and Autoimmune Diseases*, J. Schenkel, ed. (Heidelberg, Springer Lab Manual), pp. 95-140.
- Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., *et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods* *5*, 53-55.
- Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* *65*, 367-409.
- Kraut, J. (1977). Serine proteases: structure and mechanism of catalysis. *Annu Rev Biochem* *46*, 331-358.

- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Lamond, A.I., and Earnshaw, W.C. (1998). Structure and function in the nucleus. *Science* 280, 547-553.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J App Cryst* 26, 283-291.
- Lei, X.H., Shen, X., Xu, X.Q., and Bernstein, H.S. (2000). Human Cdc5, a regulator of mitotic entry, can act as a site-specific DNA binding protein. *J Cell Sci* 113 Pt 24, 4523-4531.
- Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Res* 37, D229-232.
- Li, D., and Roberts, R. (2001). WD-repeat proteins: structure characteristics, biological function, and their involvement in human disease. *Cell Mol Life Sci* 58, 2085-2097.
- Liu, S., Rauhut, R., Vornlocher, H.P., and Lührmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA* 12, 1418-1430.
- Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* 50, 437-450.
- Lührmann, R., and Stark, H. (2009). Structural mapping of spliceosomes by electron microscopy. *Curr Opin Struct Biol* 19, 96-102.
- Madhani, H.D., and Guthrie, C. (1992). A novel base pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71, 803-817.
- Maeder, C., Kutach, A.K., and Guthrie, C. (2009). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C-terminus of Prp8. *Nat Struct Mol Biol* 16, 42-48.
- Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Lührmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298, 2205-2208.
- Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H.P., and Lührmann, R. (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J* 21, 1148-1157.
- Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J* 23, 2381-2391.
- Malik, S., and Roeder, R.G. (2003). Isolation and functional characterization of the TRAP/mediator complex. *Methods Enzymol* 364, 257-284.

- Marx, J.C., Poncin, J., Simorre, J.P., Ramteke, P.W., and Feller, G. (2008). The noncatalytic triad of alpha-amylases: a novel structural motif involved in conformational stability. *Proteins* 70, 320-328.
- Mathew, R., Hartmuth, K., Mohlmann, S., Urlaub, H., Ficner, R., and Lührmann, R. (2008). Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nat Struct Mol Biol* 15, 435-443.
- McConnell, T.S., and Steitz, J.A. (2001). Proximity of the invariant loop of U5 snRNA to the second intron residue during pre-mRNA splicing. *EMBO J* 20, 3577-3586.
- McGrail, J.C., Krause, A., and O'Keefe, R.T. (2009). The RNA binding protein Cwc2 interacts directly with the U6 snRNA to link the nineteen complex to the spliceosome during pre-mRNA splicing. *Nucleic Acids Res* 37, 4205-4217.
- Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* 365, 364-368.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240-255.
- Myers, D.E., Uckun, F.M., Swaim, S.E., and Vallera, D.A. (1989). The effects of aromatic and aliphatic maleimide crosslinkers on anti-CD5 ricin immunotoxins. *J Immunol Methods* 121, 129-142.
- Nagai, K., Perutz, M.F., and Poyart, C. (1985). Oxygen binding properties of human mutant hemoglobins synthesized in *Escherichia coli*. *Proc Natl Acad Sci U S A* 82, 7252-7255.
- Newman, A.J., Teigelkamp, S., and Beggs, J.D. (1995). snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *RNA* 1, 968-980.
- Nilsen, T.W. (1998). RNA-RNA interactions in nuclear pre-mRNA splicing. In *RNA Structure and Function*, R.W. Simons, ed. (Cold Spring Harbor, New York, Cold Spring Harbor Press), pp. 279-308.
- O'Brien, M.C., and McKay, D.B. (1993). Threonine 204 of the chaperone protein Hsc70 influences the structure of the active site, but is not essential for ATP hydrolysis. *J Biol Chem* 268, 24323-24329.
- O'Day, C.L., Dalbadie-McFarland, G., and Abelson, J. (1996). The *Saccharomyces cerevisiae* Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J Biol Chem* 271, 33261-33267.
- Oh, M.D., and Gould, K.L. (2002). Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. *RNA* 8, 798-815.
- Oh, M.D., Link, A.J., Ren, L., Jennings, J.L., McDonald, W.H., and Gould, K.L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol Cell Biol* 22, 2011-2024.
- Oh, M.D., Ren, L., Wall, J.S., Gould, K.L., and Walz, T. (2007). Structural characterization of the fission yeast U5.U2/U6 spliceosome complex. *Proc Natl Acad Sci U S A* 104, 3195-3200.
- Oh, M.D., Vander Kooi, C.W., Rosenberg, J.A., Chazin, W.J., and Gould, K.L. (2003). Structural insights into the U-box, a domain associated with multi-ubiquitination. *Nat Struct Biol* 10, 250-255.

- Ohi, M.D., Vander Kooi, C.W., Rosenberg, J.A., Ren, L., Hirsch, J.P., Chazin, W.J., Walz, T., and Gould, K.L. (2005). Structural and functional analysis of essential pre-mRNA splicing factor Prp19p. *Mol Cell Biol* 25, 451-460.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* 276, 307-326.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40, 1413-1415.
- Patel, A.A., and Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* 4, 960-970.
- Pati, U.K. (1992). Novel vectors for expression of cDNA encoding epitope-tagged proteins in mammalian cells. *Gene* 114, 285-288.
- Pena, V., Rozov, A., Fabrizio, P., Lührmann, R., and Wahl, M.C. (2008). Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO J* 27, 2929-2940.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.
- Pokrovskaya, I.D., and Gurevich, V.V. (1994). In vitro transcription: preparative RNA yields in analytical scale reactions. *Anal Biochem* 220, 420-423.
- Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* 458, 475-480.
- Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* 59, 1131-1137.
- Price, N.C., and Johnson, C.M. (1989). Proteinases as probes of conformation of soluble proteins. In *Proteolytic Enzymes A Practical Approach*, R.J. Beynon, and J.S. Bond, eds. (Oxford, IRL Press), pp. 163-180.
- Reimer, L. (1997). *Transmission Electron Microscopy* (Berlin, Springer Verlag).
- Ritchie, D.B., Schellenberg, M.J., Gesner, E.M., Raithatha, S.A., Stuart, D.T., and Macmillan, A.M. (2008). Structural elucidation of a PRP8 core domain from the heart of the spliceosome. *Nat Struct Mol Biol* 15, 1199-1205.
- Roberts, R.J., Sharp, P.A. (1997). *The Nobel Prize in Physiology or Medicine 1993*. Published in *Nobel Lectures, Physiology or Medicine 1991-1995*. (Editor Nils Ringertz, World Scientific Publishing Co., Singapore).
- Rothman, J.E. (1989). Polypeptide chain binding proteins: catalysts of protein folding and related processes in cells. *Cell* 59, 591-601.
- Roy, J., Kim, K., Maddock, J.R., Anthony, J.G., and Woolford, J.L., Jr. (1995). The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. *RNA* 1, 375-390.

- Russell, R.B., and Barton, G.J. (1992). Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* 14, 309-323.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487-491.
- Sambrook, J., and Russell, D.W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd edition edn (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).
- Sander, B., Golas, M.M., Makarov, E.M., Brahm, 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, 267-278.
- Sawasaki, T., Ogasawara, T., Morishita, R., and Endo, Y. (2002). A cell-free protein synthesis system for high-throughput proteomics. *Proc Natl Acad Sci U S A* 99, 14652-14657.
- Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27, 157-162.
- Schechter, I., and Berger, A. (1968). On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem Biophys Res Commun* 32, 898-902.
- Schmidt, C., Lenz, C., Grote, M., Lührmann, R., and Urlaub, H. (2010). Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring. *Anal Chem* 82, 2784-2796.
- Schneider, M., Hsiao, H.H., Will, C.L., Giet, R., Urlaub, H., and Lührmann, R. (2010). Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. *Nat Struct Mol Biol* 17, 216-221.
- Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78, 1606-1619.
- Schuck, P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal Biochem* 320, 104-124.
- Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J., and Schubert, D. (2002). Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys J* 82, 1096-1111.
- Schwelnus, W., Richert, K., Opitz, F., Gross, T., Habara, Y., Tani, T., and Kaufer, N.F. (2001). Fission yeast Prp4p kinase regulates pre-mRNA splicing by phosphorylating a non-SR-splicing factor. *EMBO Rep* 2, 35-41.
- Schwer, B. (2001). A new twist on RNA helicases: DExH/D box proteins as RNPs. *Nat Struct Biol* 8, 113-116.
- Schwer, B., and Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J* 11, 5033-5039.
- Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68, 850-858.

- Shi, Y., and Manley, J.L. (2007). A complex signaling pathway regulates SRp38 phosphorylation and pre-mRNA splicing in response to heat shock. *Mol Cell* 28, 79-90.
- Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985). Measurement of protein using bicinchoninic acid. *Anal Biochem* 150, 76-85.
- Smith, T.F., Gaitatzes, C., Saxena, K., and Neer, E.J. (1999). The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* 24, 181-185.
- Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33, W244-248.
- Sondek, J., Bohm, A., Lambright, D.G., Hamm, H.E., and Sigler, P.B. (1996). Crystal structure of a G-protein beta gamma dimer at 2.1A resolution. *Nature* 379, 369-374.
- Song, E.J., Werner, S.L., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., Harper, J.W., Elledge, S.J., Kirschner, M.W., and Rape, M. (2010). The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes Dev* 24, 1434-1447.
- Sontheimer, E.J., Sun, S., and Piccirilli, J.A. (1997). Metal ion catalysis during splicing of pre-messenger RNA. *Nature* 388, 801-805.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315-326.
- Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* 3, 55-64.
- Stark, H., Dube, P., Lührmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* 409, 539-542.
- Stark, H., and Lührmann, R. (2006). Cryo-electron microscopy of spliceosomal components. *Annu Rev Biophys Biomol Struct* 35, 435-457.
- Sweeney, P.J., and Walker, J.M. (1993). Proteolytic enzymes for peptide production. *Methods Mol Biol* 16, 277-303.
- Tarn, W.Y., Hsu, C.H., Huang, K.T., Chen, H.R., Kao, H.Y., Lee, K.R., and Cheng, S.C. (1994). Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO J* 13, 2421-2431.
- Tarn, W.Y., Lee, K.R., and Cheng, S.C. (1993). The yeast PRP19 protein is not tightly associated with small nuclear RNAs, but appears to associate with the spliceosome after binding of U2 to the pre-mRNA and prior to formation of the functional spliceosome. *Mol Cell Biol* 13, 1883-1891.
- Teigelkamp, S., McGarvey, M., Plumpton, M., and Beggs, J.D. (1994). The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. *EMBO J* 13, 888-897.
- Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008). Crystal structure of a self-spliced group II intron. *Science* 320, 77-82.
- Tsai, W.Y., Chow, Y.T., Chen, H.R., Huang, K.T., Hong, R.I., Jan, S.P., Kuo, N.Y., Tsao, T.Y., Chen, C.H., and Cheng, S.C. (1999). Cef1p is a component of the Prp19p-associated complex and essential for pre-mRNA splicing. *J Biol Chem* 274, 9455-9462.

- Vagin, A., and Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* 56, 1622-1624.
- Valadkhan, S., Mohammadi, A., Jaladat, Y., and Geisler, S. (2009). Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc Natl Acad Sci U S A* 106, 11901-11906.
- Valeyev, N.V., Downing, A.K., Sondek, J., and Deane, C. (2008). Electrostatic and Functional Analysis of the Seven-Bladed WD beta-Propellers. *Evol Bioinform Online* 4, 203-216.
- van Heel, M. (1989). Classification of very large electron microscopical image data sets. *Optik* 82, 114-126.
- van Heel, M., and Frank, J. (1981). Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy* 6, 187-194.
- van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J Struct Biol* 116, 17-24.
- Vander Kooi, C.W., Ohi, M.D., Rosenberg, J.A., Oldham, M.L., Newcomer, M.E., Gould, K.L., and Chazin, W.J. (2006). The Prp19 U-box crystal structure suggests a common dimeric architecture for a class of oligomeric E3 ubiquitin ligases. *Biochemistry* 45, 121-130.
- Vander Kooi, C.W., Ren, L., Xu, P., Ohi, M.D., Gould, K.L., and Chazin, W.J. (2010). The Prp19 WD40 domain contains a conserved protein interaction region essential for its function. *Structure* 18, 584-593.
- Vijayraghavan, U., Company, M., and Abelson, J. (1989). Isolation and characterization of pre-mRNA splicing mutants of *Saccharomyces cerevisiae*. *Genes Dev* 3, 1206-1216.
- Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R.A., Rouault, T.A., and Schuck, P. (2004). Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal Biochem* 326, 234-256.
- Voss, S., and Skerra, A. (1997). Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng* 10, 975-982.
- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701-718.
- Wall, M.A., Coleman, D.E., Lee, E., Iniguez-Lluhi, J.A., Posner, B.A., Gilman, A.G., and Sprang, S.R. (1995). The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell* 83, 1047-1058.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470-476.
- Weiner, A.M. (1993). mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell* 72, 161-164.
- Will, C.L., and Lührmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol* 9, 320-328.

- Will, C.L., and Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol* *13*, 290-301.
- Will, C.L., and Lührmann, R. (2006). Spliceosome Structure and Function. In *The RNA World*, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press), pp. 369-400.
- Will, C.L., Schneider, C., MacMillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Lührmann, R., and Query, C.C. (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J* *20*, 4536-4546.
- Wolf, E., Kastner, B., Deckert, J., Merz, C., Stark, H., and Lührmann, R. (2009). Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J* *28*, 2283-2292.
- Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A., and Smith, C.W. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol Cell* *13*, 91-100.
- Wu, X.H., Zhang, H., and Wu, Y.D. (2010). Is Asp-His-Ser/Thr-Trp tetrad hydrogen-bond network important to WD40-repeat proteins: a statistical and theoretical study. *Proteins* *78*, 1186-1194.
- Yang, K., Zhang, L., Xu, T., Heroux, A., and Zhao, R. (2008). Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc Natl Acad Sci U S A* *105*, 13817-13822.
- Yean, S.L., Wuenschell, G., Termini, J., and Lin, R.J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* *408*, 881-884.
- Yoshimoto, R., Kataoka, N., Okawa, K., and Ohno, M. (2009). Isolation and characterization of post-splicing lariat-intron complexes. *Nucleic Acids Res* *37*, 891-902.
- Zillmann, M., Zapp, M.L., and Berget, S.M. (1988). Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol Cell Biol* *8*, 814-821.

6 APPENDIX

6.1 Abbreviations

2D	two-dimensional	EDTA	Ethylendiamintetraacetate
3D	three-dimensional	EM	Electron Microscopy
Å	Angstrom	engl.	english
A	Adenosine	ESI	electrospray ionization
Ac	Acetate	EtOH	Ethanol
Amp	Ampicillin	et al.	et alii
APS	Ammoniumperoxodisulfate	f	femto
ATP	Adenosintriphosphate	Fig.	figure
BCA	bicinchoninic acid	FRET	fluorescence resonance energy transfer
bp	base pair	G	Guanosine
BPS	branchpoint sequence	g	Gram/centrifugal force
BSA	Bovine serum albumin	GA	Glutaraldehyde
C	Cytosine	GraFix	gradient fixation
°C	degree Centigrade	h	hour/human
ca.	circa	HCl	Hydrochlorid
CCD	charged-coupled device	HEPES	N-2-Hydroxyethyl-piperazine-N-2-ethane-sulfonic acid
cDNA	complementary DNA		
Ci	Curie	hn	heterogenous nuclear
cm	Centimeter	I	Intensity
CMV	Cytomegalovirus	IRES	internal ribosomal entry site
cpm	counts per minute	ISL	intermolecular stem loop
d	desoxy	IVS	intervening sequence
Da	Dalton	k	kilo
dd	double distilled	K	Kilodalton
DEB	Butadien diepoxide	kb	kilo bases
DExD/H	consensus sequence of helicases	kDa	Kilodalton
DMP	Dimethyl pimelimidate	l	Liter
DMSO	Dimethylsulfoxide	LC	liquid chromatography
DNA	Desoxyribonucleic acid	M	molar
ds	double stranded	m	milli
DTE	1,4-Dithioerythrol	μ	micro
DTT	Dithiothreitol	MALDI	matrix-assisted laser desorption ionization
ECL	enhanced chemiluminiscence		

m3G	N ₂ ,N ₂ ,N ₇ - Trimethylguanosine	rpm	revolutions per minute
m7G	N ₇ -Monomethylguanosine	RRM	RNA recognition motif
min	Minutes	rRNA	ribosomal RNA
mM	millimolar	RT	Room temperature
mRNA	messenger RNA	S	Svedbert-unit
MS	mass spectrometry	s	Second
MW	molecular weight	SDS	Sodium-Dodecyl-Sulfate
m/z	mass over charge	SF	StrepII/FLAG-tag
n	nano	siRNA	small interfering RNA
NCBI	National Center of Biotechnology Information	sn	small nuclear
Neo ^R	Neomycin resistance	S _N 2	bimolecular nucleophilic substitution
NTA	Nitrilo-triacetic acid	sno	small nucleolar
NTC	Nineteen complex	SR	Serine Arginine rich
nt	Nucleotide	ss	Splice site
NTPs	Nucleosid-5'-Triphosphate	Sulfo-MBS	N-Maleimidobenzoyl-N- hydroxysulfosuccinimide ester
OAc	Acetate	SV40	Simian virus 40
OD	Optical density	T	Thymidine
OH	hydroxyl	Tab.	Table
O/N	overnight	TAP	Tandem affinity purification
ORF	open reading frame	TBE	Tris-Borat-EDTA-solution
P	Phosphate	TE	Tris-EDTA
p	pico	TEM	transmission electron microscopy
PAGE	polyacrylamide gel- electrophoresis	TEMED	N, N, N', N'-Tetramethyl- ethylenediamine
PBS	Phosphate Buffered Saline	T _m	melting temperature
PCI	Phenol, Chloroform, Isoamylalkohol	TOF	Time of flight
PCR	polymerase chain reaction	Tris	Tris-(hydroxymethyl)- aminomethane
PFA	Paraformaldehyde	tRNA	transfer RNA
pH	preponderance of hydrogen ions	U	Uracil/Unit
PMSF	Phenylmethylsulfonyl- fluoride	U snRNA	uridine-rich small nuclear RNA
Pre-mRNA	precursor-mRNA	U snRNP	uridine-rich small nuclear ribonucleoprotein
Prp	pre-mRNA processing	UTP	Uridintriphosphate
Py	Pyrimidine base	UTR	untranslated region
R	Purine base	UV	Ultraviolet
R _{free}	free R-factor	V	Volt
R _{work}	working R-factor	v/v	volume per volume
R.m.s.	Root mean square	W	Watts
RMSD	Root mean square deviation	WB	Western Blot
RNA	Ribonucleic acid	w/v	weight per volume
RNAi	RNA-Interference	Y	Pyrimidine base
RNase	Ribonuclease		
RNP	Ribonucleoprotein		

6.2 Acknowledgements

At the end of this work I would like to take the opportunity to express my heartfelt gratitude to all the people to continuously support me along my way to this PhD thesis. It is a pleasure to thank those who made this thesis possible.

I am grateful to Prof. Dr. Reinhard Lührmann for his guidance and support throughout my time in his laboratory. I highly appreciate that I could present my studies with a poster at the annual meeting of the RNA society in Berlin (Germany) and with a talk at the annual meeting of the RNA society in Madison (USA).

I also would like to thank Prof. Dr. Ralf Ficner and Prof. Dr. Ernst A. Wimmer for their supervision of this external Ph.D. thesis. In addition, many thanks to Prof. Dr. Stefanie Pöggeler, Prof. Dr. Gerhard Braus, Prof. Dr. Kai Tittmann and Prof. Dr. Tomas Pieler for being members of my extended thesis committee.

My special thanks go to Cindy. This work would not have been possible without your encouragement and brilliant advice. I would like to thank you for the many fruitful discussions and your help with the paperwork on the different manuscripts and submission of our paper, and especially for spending a holiday together with me in the lab to prepare our talk for the RNA meeting.

I would like to extend my special thanks to Ira. Establishing HeLa stable cell lines was a turning point during my PhD thesis. I am also grateful for the many cheerful discussions and your always optimistic and positive attitude.

I am thankful to Henning and his Mass Spectrometry group, especially to Monika who helped me so much with the difficult interpretation of the MS data that greatly contributed to the success of this work.

I am also thankful to Vlad who gave me the fantastic opportunity to deepen my knowledge in X-ray crystallography. It was a great experience to go through all the steps starting with crystallization, collecting the data set at the synchrotron and finally solving the structure.

I am grateful to Holger and his 3D Electron Cryomicroscopy group, especially Björn and Monika, for all their technical help, scientific discussions and profound advices.

I would like to thank Dmitry who introduced me into the laboratory and together with Adrian and Dr. Wolfgang Fischle also contributed to this work.

I am indebted to Peter, Kami, Gaby, Thomas, Juliane, Irene, Marion, Gertrud and Uschi. With your hard work you keep the place running and make life so much easier for us.

I would like to extend my thanks to my lab colleagues Elmar and Genia, but also to Jochen, Berthold and Norbert for the friendly working atmosphere and all the good discussions we have had.

I also would like to thank the "Fonds der Chemischen Industrie" for financial support by a "Kekulé Stipendium".

Last, but not least I would like to express my deepest gratitude to my family. Without their continuous support and love during the numerous ups and downs in the past, it would not be possible to write down these lines now.

6.3 Erklärung/Affidavit

Hiermit erkläre ich an Eides Statt, dass die Dissertation mit dem Titel „The Molecular Architecture and Structure of the Human Prp19/CDC5L Complex and 35S U5 snRNP“ selbständig und ohne unerlaubte Hilfe angefertigt wurde.

I declare that my Doctoral Thesis “The Molecular Architecture and Structure of the Human Prp19/CDC5L Complex and 35S U5 snRNP” has been written independently and with no other sources and aids then quoted.

Michael Grote, Göttingen, September 2010

Parts of this thesis were published in:

Grote, M., Wolf, E., Will, C.L., Lemm, I., Agafonov, D.E., Schomburg, A., Fischle, W., Urlaub, H., and Lührmann, R. (2010) *Molecular architecture of the human Prp19/CDC5L complex*. MCB, Vol. 30, No. 9, 2105-2119.

Schmidt, C., Lenz, C., **Grote, M.**, Lührmann, R., and Urlaub, H. (2010) *Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring*. Anal Chem, Vol. 82, No. 7, 2784-2796.

Golas, M.M., Sander, B., Bessonov, S., **Grote, M.**, Wolf, E., Kastner, B., Stark, H., Lührmann, R. (2010) *3D Cryo-EM Structure of an Active Step I Spliceosome and Localization of its Catalytic Core*. Mol Cell, Vol. 40, No. 6, 927-938.