

Absolute and relative quantification of proteins
in large protein-RNA assemblies
by mass spectrometry

Dissertation
for the award of the degree
“Doctor rerum naturalium” (Dr. rer. nat.)
Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

submitted by
Carla Schmidt
from Wolfenbüttel

Göttingen 2010

Thesis committee

Dr. Henning Urlaub (Reviewer)	Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Göttingen
Prof. Dr. Ralf Ficner (Reviewer)	Department of Molecular Structural Biology, Georg-August-Universität, Göttingen
Prof. Dr. Ivo Feußner	Department of Plant Biochemistry, Georg-August-Universität, Göttingen
Prof. Dr. Ulf Diederichsen	Department of Organic and Biomolecular Chemistry, Georg-August-Universität, Göttingen

Date of the oral examination: June 8, 2010

Affidavit

I hereby declare that my doctoral thesis entitled "Absolute and relative quantification of proteins in large protein-RNA assemblies by mass spectrometry" has been written independently and with no other sources and aids then quoted.

Carla Schmidt

May 14, 2010
Göttingen

Table of contents

1 Summary	1
2 Introduction	3
2.1 Mass spectrometry-based protein identification and quantification	3
2.1.1 MS-based protein identification (Proteomics)	3
2.1.2 Quantification by mass spectrometry	7
2.1.3 Absolute quantification by mass spectrometry	10
2.1.4 Relative quantification by mass spectrometry	13
2.1.5 Quantification by mass spectrometry to analyze dynamic protein transitions	19
2.1.6 Quantifying mass spectrometric measurements	20
2.2 The spliceosome	25
2.2.1 Eukaryotic pre-mRNA splicing	25
2.2.1.1 Structure of eukaryotic pre-mRNAs	25
2.2.1.2 The biochemical splice reaction	25
2.2.1.3 The catalytic machinery	27
2.2.2 Components of the spliceosome	27
2.2.2.1 The U snRNPs	27
2.2.2.2 non-snRNP components	29
2.2.3 The stepwise assembly of the spliceosome	31
2.3 Aim of this study	33
3 Materials and Methods	35
3.1 Materials	35
3.1.1 Chemicals	35
3.1.2 Enzymes and Enzyme inhibitors	36
3.1.3 Nucleotides	36
3.1.4 DNA oligonucleotides	36
3.1.5 Plasmids	37
3.1.6 Bacteria strains	37
3.1.7 Cell lines	37
3.1.8 Cell culture	37
3.1.9 Commercial reaction kits and buffers	38
3.1.10 Commonly used buffers	38
3.1.11 Standard peptides	39
3.1.12 Chromatography materials and consumables	39
3.1.13 Instruments, Equipment	39
3.2 Methods	40
3.2.1 Molecular biology methods	40
3.2.1.1 Concentration determination of nucleic acids	40
3.2.1.2 Phenol-Chloroform-Isoamylalcohol (PCI) extraction	40
3.2.1.3 Agarose gel electrophoresis of DNA	41
3.2.1.4 In vitro transcription	41
3.2.1.5 DNA amplification	42
3.2.1.6 Restriction digestion of DNA	43
3.2.1.7 Proteinase K digestion	43
3.2.1.8 Denaturing polyacrylamide gel electrophoresis	43
3.2.1.9 Silver staining of RNA	44

3.2.1.10 Native gel electrophoresis of RNA complexes	44
3.2.1.11 Generation of pre-mRNA mutants (deletion of 5'ss and BPS)	44
3.2.2 Protein biochemical methods	45
3.2.2.1 Protein concentration determination	45
3.2.2.2 Polyacrylamide gel electrophoresis (PAGE)	46
3.2.2.3 Colloidal Coomassie staining of proteins	46
3.2.3 Cell culture, nuclear extract, purification of spliceosomal complexes	46
3.2.3.1 Metabolic Labeling of HeLa S3 cells (SILAC labeling)	46
3.2.3.2 Preparation of splicing active HeLa cell nuclear extract	47
3.2.3.3 In vitro splicing	47
3.2.3.4 Purification of human Prp19/CDC5L complex	48
3.2.3.5 Purification of human tri-snRNP (U4/U6.U5)	48
3.2.3.6 Overexpression of MS2-MBP fusion protein	48
3.2.3.7 Purification of human B and C complexes	49
3.2.3.8 Investigating the protein composition during the spliceosomal assembly	49
3.2.4 Mass spectrometry methods	51
3.2.4.1 In-gel hydrolysis of proteins	51
3.2.4.2 Extraction of peptides	51
3.2.4.3 iTRAQ labeling of extracted peptides for relative quantification	52
3.2.4.4 In-solution hydrolysis of hPrp19/CDC5L complex	52
3.2.4.5 Comparison of hydrolysis protocols and selection of standard peptides (AQUA peptides) for absolute quantification	53
3.2.4.6 Internal standardization with the selected standard peptides	54
3.2.4.7 Absolute quantification by LC-offline MALDI-ToF/ToF-MS (Peak Area)	55
3.2.4.8 Absolute quantification by LC-online MS-MS/MS (Extracted Ion Chromatograms, XIC)	56
3.2.4.9 Absolute quantification by Multiple Reaction Monitoring	57
3.2.4.10 Relative quantification by LC-online MS-MS/MS (iTRAQ quantification)	58
3.2.4.11 Relative quantification by LC-online MS-MS/MS (SILAC quantification)	59
3.2.4.12 Statistical analysis	59
4 Results	60
4.1 Determination of the protein stoichiometry within the hPrp19/CDC5L complex by absolute quantification (AQUA)	60
4.1.1 Proteomic analysis of the hPrp19/CDC5L complex	60
4.1.2 Comparison of different protocols for in-solution hydrolysis	61
4.1.3 Selection of standard peptides for absolute quantification	66
4.1.4 Absolute quantification by LC-offline MALDI-ToF/ToF-MS (peak area)	69
4.1.5 Absolute quantification by LC-ESI-MS (extracted ion chromatograms)	71
4.1.6 Absolute quantification by Multiple Reaction Monitoring (MRM)	72
4.1.7 Comparison of results for absolute quantification in determining the Stoichiometry of the hPrp19/CDC5L complex	78
4.2 Relative quantification by iTRAQ-labeling of in-gel digested proteins	80
4.2.1 Optimization of iTRAQ-labeling of in-gel digested proteins	80
4.2.2 Relative quantification of different amounts of spliceosomal tri-snRNP proteins – a feasibility study	82
4.3 Relative quantification of spliceosomal B and C complexes – a comparative study	85
4.3.1 Purification of spliceosomal B and C complexes	85
4.3.2 Proteomic analysis of spliceosomal B and C complexes – spectral count	87
4.3.3 iTRAQ labeling of spliceosomal B and C complexes for relative quantification	93
4.3.4 SILAC quantification of spliceosomal B and C complexes	101
4.3.5 Relative quantification of spliceosomal B and C complexes	109
4.3.5.1 Proteins common to B and C complexes	109

4.3.5.2 Proteins predominantly associated with B complex	111
4.3.5.3 Proteins predominantly associated with C complex	113
4.3.6 Comparison of the three quantification methods – iTRAQ, SILAC, and spectral count	116
4.4 Protein assembly time line for spliceosomes by relative quantification	118
4.4.1 Generation of splicing inactive pre-mRNAs	118
4.4.2 Triple SILAC to monitor dynamic protein changes	123
4.4.3 Time-dependent protein assembly on splicing-active and splicing-inactive pre-mRNAs	126
4.4.3.1 Experimental setup	126
4.4.3.2 Normalization of the data	128
4.4.3.3 Protein assembly on splicing-active and splicing-inactive pre-mRNAs	129
4.4.4 Direct comparison of the protein assembly on splicing-active PM5 pre-mRNA and splicing-inactive 5'ss deleted PM5 pre-mRNA	134
4.4.4.1 Experimental setup	134
4.4.4.2 Normalization of the data	135
4.4.4.3 Protein assembly on PM5 pre-mRNA versus 5'ss-deleted PM5 pre-mRNA	136
5 Discussion	139
5.1 Determination of the protein stoichiometry within the hPrp19/CDC5L complex by absolute quantification (AQUA)	139
5.2 Relative quantification by iTRAQ-labeling of in-gel digested proteins	143
5.3 Relative quantification of spliceosomal B and C complexes – a comparative study	144
5.3.1 Methodical considerations	144
5.3.2 Functional considerations	148
5.4 Protein assembly time line for spliceosomes by relative quantification	150
6 References	156
7 Appendix	169
Additional information	169
List of abbreviations	200
Acknowledgements	203
Curriculum Vitae	205

List of figures

2.1	Protein identification by mass spectrometry	4
2.2	Principle of quantification by stable isotope labeling	8
2.3	Example MS spectrum of an endogenous and the corresponding standard peptide for quantification	11
2.4	Isotope labeled amino acids used for cell growth in SILAC labeling	15
2.5	Chemical labeling reagents	16
2.6	Structure of the 4-plex iTRAQ reagent	17
2.7	Example of Extracted ion chromatograms (XICs) to read out the signal intensities for quantification	23
2.8	MRM workflow	24
2.9	Schematic representation of the splicing process	26
2.10	The stepwise assembly of the spliceosome	32
3.1	Generation of 5'ss, BPS and BPS-ACTGA deleted PM5 pre-mRNA	45
3.2	In-house manufactured gel cutting device	51
4.1	1D-PAGE of the hPrp19/CDC5L complex	61
4.2	Miscleaved peptides obtained by hydrolysis in the presence of acetonitrile are significantly longer than miscleaved peptides obtained by hydrolysis in the presence of urea	65
4.3	Analysis of AD-002 and CTNNB1 sequences for presence of standard peptides	67
4.4	Solubility of standard peptides	68
4.5	Peak overlap is frequently observed in MALDI mass spectrometry	70
4.6	Specificity of the designed MRM transitions	73
4.7	Example of the MRM transitions for an endogenous and the corresponding standard peptide (LGLLGLPAPK derived from CDC5L protein)	74
4.8	Peptide ratios obtained from MRM analysis of the hPrp19/CDC5L complex	75
4.9	PRL 1 sequence	77
4.10	Protein stoichiometry within the hPrp19/CDC5L complex	79
4.11	Workflow for iTRAQ-labeling of in-gel digested proteins	81

4.12	Separation of different amounts of spliceosomal tri-snRNP proteins by gel electrophoresis	82
4.13	Example MS/MS spectrum of an iTRAQ labeled peptide	83
4.14	Example MS/MS spectrum of an iTRAQ labeled peptide	84
4.15	Purification of spliceosomal B and C complexes for iTRAQ and SILAC quantification	86
4.16	The proteomes of spliceosomal B and C complexes	87
4.17	Labeling efficiency achieved for iTRAQ labeling in two independent replicates	94
4.18	Example MS/MS spectrum for iTRAQ quantification of spliceosomal B and C complexes	94
4.19	Splicing kinetics, complex formation and RNA composition of B and C complexes as obtained using light and heavy SILAC nuclear extracts	102
4.20	Example MS spectrum for SILAC quantification of spliceosomal B and C complexes	103
4.21	Relative abundances of the Sm proteins in the B and C complexes determined by iTRAQ, SILAC, and spectral count	109
4.22	Relative protein abundances of the U2 and U5 snRNP specific proteins obtained by iTRAQ, SILAC, and spectral	110
4.23	Relative protein abundances of the U1, U4/U6, U4/U6.U5 snRNP specific proteins, and LSm proteins obtained by iTRAQ, SILAC, and spectral count	112
4.24	Protein ratios for hPrp19/CDC5L complex proteins obtained by iTRAQ, SILAC, and spectral count	114
4.25	Protein ratios of step 2 factors obtained by iTRAQ, SILAC, and spectral count	115
4.26	Scatter plot of SILAC and iTRAQ protein ratios	117
4.27	Sequence of the PM5 plasmid	119
4.28	The splicing kinetics and the spliceosomal complex formation using PM5 pre-mRNA and 5'ss- and BPS-deleted PM5 pre-mRNA	120
4.29	The splicing kinetics and the spliceosomal complex formation using PM5 pre-mRNA and BPS and BPS-ACTGA deleted PM5 pre-mRNAs	122
4.30	Splicing kinetics and spliceosomal complex formation using light, medium and heavy SILAC nuclear extracts	124
4.31	Comparison of the protein content within light, medium and heavy SILAC nuclear extracts	125

4.32	Experimental setup to monitor the time-dependent protein assembly during pre-mRNA splicing	127
4.33	Coomassie stained gels of combined samples from assembly studies on PM5 pre-mRNA, 5'ss-deleted PM5 pre-mRNA and BPS-ACTGA-deleted PM5 pre-mRNA	128
4.34	Normalization of the obtained protein ratios on the ribosomal proteins	129
4.35	Assembly of U1 snRNP specific proteins on PM5 pre-mRNA, 5'ss-deleted PM5 pre-mRNA, and BPS-ACTGA-deleted PM5 pre-mRNA over 30 minutes	130
4.36	Assembly of U2 snRNP specific proteins on PM5 pre-mRNA, 5'ss-deleted PM5 pre-mRNA and BPS-ACTGA-deleted PM5 pre-mRNA monitored over 30 minutes	131
4.37	Assembly of hPrp19/CDC5L complex proteins on PM5 pre-mRNA, 5'ss-deleted PM5 pre-mRNA and BPS-ACTGA-deleted PM5 pre-mRNA monitored over 30 minutes	132
4.38	Assembly of step 2 factors on PM5 pre-mRNA, 5'ss-deleted PM5 pre-mRNA, and BPS-ACTGA-deleted PM5 pre-mRNA monitored over 30 minutes	133
4.39	Experimental setup to compare the protein compositions of the sets of proteins assembled on PM5 and 5'ss-deleted PM5 pre-mRNA at different time points during pre-mRNA splicing	134
4.40	Coomassie stained gel of combined samples from comparison of the protein assembly on PM5 and 5'ss-deleted PM5 pre-mRNA at different time points during pre-mRNA splicing	135
4.41	Normalization of the protein ratios found for the cap binding proteins	136
4.42	Direct comparison of protein assembly on the PM5 pre-mRNA compared with corresponding assembly on the 5'ss-deleted PM5 pre-mRNA for U1 snRNP specific proteins and Sm proteins	137
4.43	Protein assembly of the hPrp19/CDC5L complex and the step 2 factors on the PM5 pre-mRNA compared with corresponding assembly on the 5'ss-deleted PM5 pre-mRNA	138
5.1	Endogenous counterparts of the selected standard peptides for CDC5L protein are located in highly structured regions of CDC5L	140
A.1	Miscleaved peptides generated in the presence of urea show higher peptide scores for the peptide sequences selected for quantification	172
A.2	Incompletely cleaved peptides obtained by hydrolysis in the presence of acetonitrile are significantly longer than incompletely cleaved peptides obtained by hydrolysis in the presence of urea	172

List of tables

2.1	Most important and popular MS-based quantification methods in proteomics	9
2.2	Composition of the human U snRNPs	30
3.1	Oligonucleotides used in this study	36
3.2	Isotope composition of amino acids used for different SILAC media	47
3.3	Used SILAC nuclear extracts for investigation of the protein assembly on pre-mRNAs during 30 min	50
3.4	Used SILAC nuclear extracts for direct comparison of the protein assembly on PM5 pre-mRNA and on a splicing inactive pre-mRNA during 30 min	50
4.1	Comparison of different hydrolysis protocols	62
4.2	Comparison of different hydrolysis protocols	63
4.3	Selected standard peptides	68
4.4	Relative protein stoichiometries within the hPrp19/CDC5L complex determined by LC-offline MALDI-ToF/ToF-MS	69
4.5	Relative protein stoichiometries within the hPrp19/CDC5L complex determined by LC-ESI-MS	71
4.6	Relative protein stoichiometries within the hPrp19/CDC5L complex as determined by MRM analyses	77
4.7	Summary of the protein stoichiometries obtained from different MS techniques	78
4.8	Protein ratios for the relative quantification of different amounts of spliceosomal tri-snRNP proteins	84
4.9	Spectral count for proteomic analysis of spliceosomal B and C complexes	88
4.10	Relative quantification of spliceosomal B and C complexes by iTRAQ	96
4.11	Relative quantification of spliceosomal B and C complexes by SILAC	104
A.1	Additional proteins identified in the hPrp19/CDC5L complex by LC-MS/MS	169
A.2	Spectral Count for peptides selected for quantification and their miscleaved versions	171
A.3	Three MRM transitions for each endogenous and standard peptide were designed	173
A.4	Absolute amounts of proteins of the hPrp19/CDC5L complex	174
A.5	Proteins identified in the first replicate of the relative quantification of spliceosomal B and C complexes by iTRAQ	175

A.6	Proteins identified in the second replicate of the relative quantification of spliceosomal B and C complexes by iTRAQ	186
A.7	Proteins that were found to be enriched in light (L), medium (M) or heavy (H) nuclear extracts, respectively	198
A.8	List of ribosomal proteins used for normalization	199
A.9	List of normalization factors calculated for the different assembly studies	199

1 Summary

The vast majority of cellular processes is driven by protein complexes. These emerge as multimeric protein assemblies or are complexed with protein ligands such as RNA or DNA molecules. To understand the details of the underlying cellular processes, an analysis of protein complexes is therefore a prerequisite. The analysis involves identification of the protein components, determination of protein stoichiometries, relative quantification of the components in different complex states, and the study of protein-protein and protein-ligand interactions. Several mass spectrometry-based techniques have been developed to tackle these problems and greatly facilitating the high throughput analysis of a variety of protein complexes. We have applied these methods to the analysis of the spliceosome. The spliceosome is a protein-RNA machinery that catalyzes the excision of introns and the ligation of exons during eukaryotic pre-mRNA splicing. It passes through sequential assembly states that dramatically differ in their RNA and protein composition. Mass spectrometry-based quantification methods are thus ideally suited to answer these questions. In this study, absolute and relative quantification by mass spectrometry was used to characterize different spliceosomal complexes or subcomplexes, in terms of their protein composition and protein stoichiometry. The aim of this study was to generate a complete as possible picture of the dynamic protein changes of the spliceosome during its assembly pathway.

One spliceosomal subcomplex, the hPrp19/CDC5L complex, consists of seven individual proteins (hPrp19, Hsp70, CTNNB1, PRL1, CDC5L, AD-002, and SPF27) and plays a crucial role in the assembly of a catalytically active spliceosome. The exact protein stoichiometries within this particular protein complex have not yet been investigated. We therefore set up a mass spectrometry-based quantification method and used the hPrp19/CDC5L complex to implement the methodology of absolute quantification (AQUA) with the help of synthetic standard peptides in combination with multiple reaction monitoring (MRM). Several conditions for complete hydrolysis of the protein complex were evaluated and found to be crucial for accurately determining the protein stoichiometries within this particular protein complex. In addition, the suitability of different standard peptides (AQUA peptides) was tested and different mass spectrometry techniques, to read out signal intensities for absolute quantification, were compared. The analyses revealed that the denaturing conditions used during protein hydrolysis were most crucial in obtaining a robust and reliable quantification of the different subcomponents of the hPrp19/CDC5L complex. Taking all independent experimental approaches into account, the hPrp19/CDC5L complex

consists of four copies of hPrp19, two copies of CDC5L, and one copy each of SPF27, PRL1, and CTNNBL1.

In a different series of investigation, we set out to make a high throughput relative quantification of two distinct spliceosomal complexes, namely the pre-catalytic and the catalytically active spliceosomes (spliceosomal B and C complexes). For this purpose, a relative quantification approach involving iTRAQ (isobaric tags for relative and absolute quantification) labeling of in-gel digested proteins was optimized and applied to the two complexes. The results were compared to relative quantification by metabolic labeling (SILAC; stable isotope labeling with amino acids in cell culture) and semi-quantitative spectral count from proteomic analysis of the B and C complexes. Thus, three independent relative quantification data sets were generated allowing for an in depth comparison of the methodologies involved. We found an overall good agreement for the used quantification methods. Although limits for relative quantification by spectral count were observed for some proteins. Several proteins were found to be pre-dominantly associated with spliceosomal B or C complexes, and only few proteins were found to be present in equal amounts within the two complexes. The high dynamics of the spliceosome during its assembly pathway was thus clearly demonstrated. Furthermore, the study revealed a novel C complex association of a DEAD box helicase, the DDX34 protein.

In a final scene of investigation, we analyzed the dynamic protein changes during the pre-mRNA splicing process. To this end, the protein compositions on splicing active and splicing inactive pre-mRNAs at different time points were compared by stable isotope labeling with amino acids in cell culture (SILAC). Splicing inactive pre-mRNAs were generated by deletion of the pre-mRNA's 5' splice site (5'ss) or the branchpoint site (BPS) to prevent the formation of a catalytically active spliceosome. Two approaches were then followed to monitor the spliceosomal assembly kinetics: (i) The time dependent protein assembly on the splicing active or splicing inactive pre-mRNAs, respectively, and (ii) the direct comparison of the assembled protein composition on a splicing active to a splicing inactive pre-mRNA at different time points during pre-mRNA splicing. The kinetics of the protein assembly on the different pre-mRNAs was analyzed for spliceosomal protein groups that are affected by deletion either of the 5'ss or the BPS. The results demonstrate that proteins specific for the pre-catalytic A complex could also be found on splicing inactive pre-mRNAs at a very early time point, whereas proteins, specific for the catalytically active spliceosome, show striking differences in their assembly characteristics on the different pre-mRNAs. The thus generated timelines for the assembly of whole groups of spliceosomal proteins during spliceosomal assembly and splicing catalysis are extremely helpful in understanding the dynamic process of pre-mRNA splicing.

2 Introduction

Proteins are involved in all cellular processes. They emerge as protein or protein-ligand complexes (e.g. complexed with RNA/DNA molecules, sugars, lipids, cofactors etc.) to form the catalytic core for different cellular actions. These cellular actions are completed either by the proteins alone or by the ligands that are arranged by the proteins within the catalytic core. The identification of active proteins within protein and protein-ligand complexes is therefore of great importance to understand different cellular processes. In the past, two fundamental breakthroughs facilitated protein identification within unknown protein mixtures enormously: (i) complete sequencing of the genomes, and (ii) the invention of two fundamental mass spectrometry approaches. As a result, mass spectrometry-based protein identification have become a popular and indispensable technique during the last years.

2.1 Mass spectrometry-based protein identification and quantification

2.1.1 MS-based protein identification (Proteomics)

By definition, the proteome is the “entire set of proteins expressed by a genome, cell, tissue or organism. More specifically, it is the set of expressed proteins in a given type of cells or an organism at a given time under defined conditions” (Marc Wilkins, 1994). Proteomics thus deals with the identification of proteins. An early breakthrough was the development of the Edman degradation enabling for N-terminal sequencing of unknown proteins by chemical cleavage in a step-wise manner (Edman, 1949). This discovery was rewarded in 1958 with the Nobel Prize in chemistry to Frederik Sanger for sequencing Insulin. Automation of Edman sequencing made it to a powerful technique for protein identification. However, this technique is very time-consuming, requires very homogenous samples, and a free amino terminus. Edman sequencing is limited to 20 - 30 amino acids and requires an intact N-terminus. To circumvent this problem, the proteins are hydrolyzed by endoproteinases and the generated peptides are then sequenced instead of whole proteins. Therefore, generated peptide mixtures need to be purified by reversed phase (RP) high pressure liquid chromatography (HPLC). This requires large amounts of samples, which poses a problem.

The invention of two soft ionization techniques, namely Matrix-assisted laser desorption/ionization (MALDI; Karas *et al.*, 1987; Karas and Hillenkamp, 1988; Tanaka *et al.*, 1988) and electrospray ionization (ESI; Fenn *et al.*, 1989), allowed the analysis of intact proteins and peptides by mass spectrometry (MS). This invention, together with the

knowledge of the genome sequence, allows protein identification by mass spectrometry using the following workflow: (i) hydrolysis of proteins by endoproteases into peptides, (ii) MS analysis of the peptides providing mass-to-charge ratios (m/z) of the intact peptides (MS experiment), (iii) selection of precursor ions (peptides) for tandem mass spectrometry (MS/MS), (iv) fragmentation of the selected precursor yielding the peptide sequence (MS/MS experiment), and (v) identification of the peptide and finally the protein by comparing the peptide mass and the obtained peptide sequence with *in silico* hydrolyzed proteins in a database (Figure 2.1).

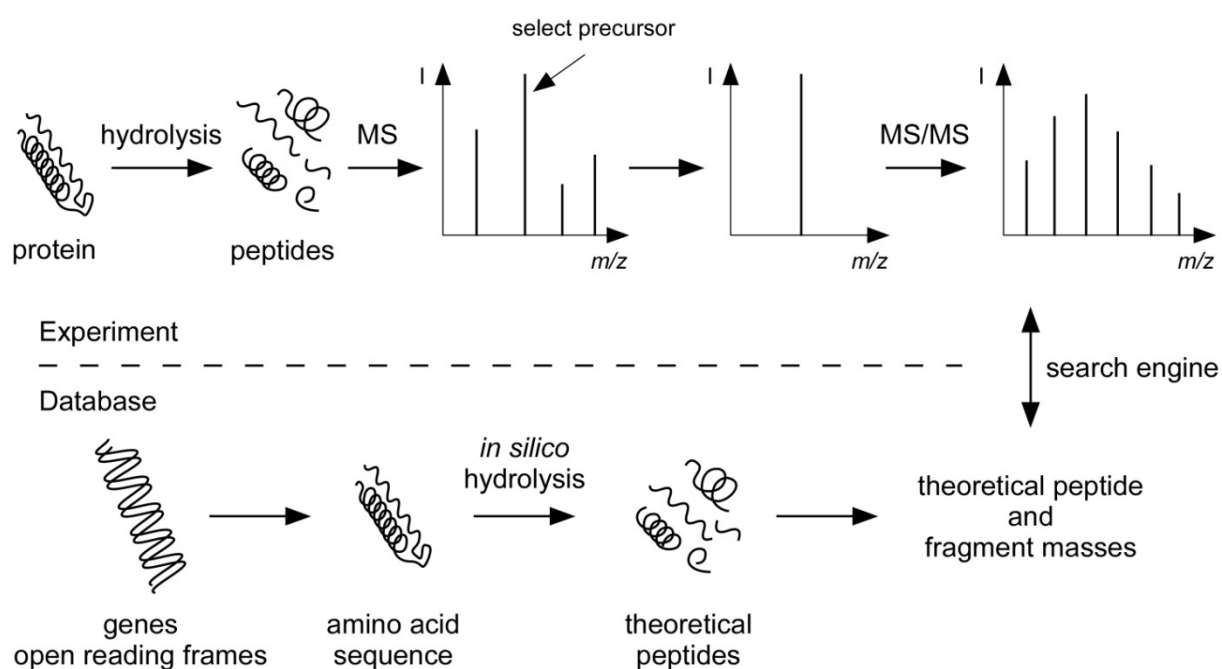


Figure 2.1: Protein identification by mass spectrometry. Proteins are hydrolyzed by specific endoproteases and the mass-to-charge ratios of the generated peptides are determined by mass spectrometry. Precursor ions are selected for fragmentation (MS/MS) and the protein is identified by comparing the experimentally determined peptide and fragment masses with theoretical peptide and fragment masses in a database.

The more proteins are present in a sample the more peptides are generated by hydrolysis. As even state-of-the-art mass spectrometers can only analyze a limited number of peptides at a given time, reduction of the sample complexity is one of the prerequisites for a successful proteome analysis. Coupling of reversed-phase liquid chromatography (RP-LC) and mass spectrometry is a powerful tool to reduce the sample complexity. Peptides are separated during chromatography and eluted peptides are subsequently analyzed in the mass spectrometer. However, sample reduction solely by RP-LC is still not sufficient for highly complex samples. There are two general strategies to further reduce sample complexity before the actual RP-LC: (i) additional separation of the generated peptides after hydrolysis of the proteins, and (ii) separation of the proteins before hydrolysis.

Separation of generated peptides

In situ hydrolysis of proteins has the advantage that the time-consuming protein separation step is bypassed, i.e. additional steps for sample preparation are omitted. Due to the high complexity of the generated peptide mixture, multidimensional peptide separation, which comprises the coupling of different separation techniques, is required. For this purpose, two dimensional liquid chromatography (2D-LC) using strong cation-exchange (SCX) chromatography in the first and RP-LC in the second dimension, has been shown to provide optimal results. Washburn *et al.*, 2001 successfully introduced the multidimensional protein identification technology (MuD-PIT), where generated peptides are separated on a biphasic microcapillary column packed with SCX and RP material and are subsequently eluted into the mass spectrometer (Washburn *et al.*, 2001). A different 2D-LC approach using RP in both dimensions but at different pH values has been described (Gilar *et al.*, 2005). The peptides are separated at high pH (10) in the first, followed by orthogonal separation at low pH (2) in the second dimension (Gilar *et al.*, 2005). The advantage of this set-up is to avoid SCX, whose resolution for peptides is relatively low and which involves the use of buffers with high salt concentrations that affect ionization of the sample. An alternative approach is the separation of generated peptides by isoelectric focusing (IEF) on immobilized pH gradient (IPG) strips. The IPG strips are cut into pieces, peptides are eluted and subsequently analyzed by reversed-phase liquid chromatography coupled mass spectrometry (Krijgsveld *et al.*, 2006).

Separation of proteins

Fractionation leads to a first reduction of sample complexity. The fractions are then applied to protein hydrolysis by specific endoproteinases. There are two different strategies for protein separation. Either, the proteins are separated by liquid chromatography and eluted fractions are collected for further sample preparation or, most popular, the proteins are separated by gel electrophoresis and excised proteins are hydrolyzed in-gel. Liquid chromatography-based protein separation is mostly performed by size exclusion chromatography (SEC) or ion exchange chromatography (IEC). The latter technique can be divided into cation and anion exchange chromatography depending on the chromatographic material used. Furthermore, the proteins can be separated according to their hydrophobicity using hydrophobic interaction chromatography (HIC) or hydrophilic interaction chromatography (HILIC). All these methods fractionate the sample in groups of proteins of similar size, charge or hydrophobicity / hydrophilicity, respectively. For protein separation by gel electrophoresis, several one and two dimensional methods are available (e.g. 1D (Laemmli, 1970), 2D (Klose, 1975; O'Farrell, 1975), Blue Native (Schagger *et al.*, 1988; Schagger and von Jagow, 1991), and 16BAC (Hartinger *et al.*, 1996; Macfarlane, 1989) gel electrophoresis). The combination of IEF and sodium dodecylsulfate (SDS) polyacrylamide gel electrophoresis (PAGE) is the most common approach, allowing for a resolution of up to 10,000 protein spots per gel (Klose, 1999). However, this approach is

limited to the very basic and acidic proteins and precipitation of the sample is sometimes observed. When working in experimental systems investigating samples with moderate complexity (e.g. purified cell compartments, purified protein complexes) 1D-SDS-PAGE (Laemmli, 1970) is the simplest and most adequate alternative.

Mass spectrometry As mentioned above, MALDI and ESI allow the determination of the molecular weight of intact proteins and peptides by mass spectrometry. NanoLC uses very low flow rates and is thus well suited for coupling to an ESI mass spectrometer. Consequently, the sample complexity is reduced and eluted peptides are immediately ionized and analyzed in the mass spectrometer. State-of-the-art mass spectrometers determine the mass-to-charge (m/z) ratio of peptides within a mixture (MS experiment) and isolate peptides (precursors) for subsequent fragmentation within the mass spectrometer (MS/MS experiment). Mass analysis of the fragment ions and the accurate monoisotopic peptide mass allows the identification of the corresponding protein by comparison to *in silico* hydrolyzed proteins in a database. For this purpose, search engines that compare experimentally determined peptide and fragment masses with *in silico* generated masses are employed (Figure 1).

Most of the mass spectrometers use data-dependent-acquisition (DDA). A DDA duty cycle consists of two steps: (i) a MS scan over a certain mass range (typically 350 - 1500 m/z) to detect masses and charge states of the peptides (MS experiment), and (ii) selection of peptide precursors and subsequent fragmentation (MS/MS experiment). While the mass spectrometer is occupied with MS/MS experiments, usually no MS scans can be performed (except for Orbitrap-FT mass analyzer). Consequently, instruments with a short duty cycle can sequence more peptides in a given time. In general, ion-trap instruments have shorter duty cycles than Qq-ToF instruments, whereas the quality of MS/MS spectra from Qq-ToF instruments is better, so that more MS/MS spectra can be correctly assigned to peptide sequences (Elias *et al.*, 2005). However, the sensitivity of a mass spectrometer is not only dependent on the acquisition speed, but also on the mass accuracy (i.e. the accuracy, with which the precursor and fragment masses are determined) and the peak resolution. The mass resolution is the ratio of the peak's mass and its width. The peak width is usually taken as the full width at half maximum intensity (FWHM). The more accurate the peptide and fragment masses are determined, the better can the peptide be assigned to the correct protein during database search. A high resolution of peaks allows to distinguish between co-eluting peaks of similar mass and leads to more data information. Currently, a mass accuracy of 0.2 ppm and a resolution of 100,000 is routinely achieved and is sufficient for most purposes. The recent development of a new, highly sensitive and fast mass spectrometer,

the LTQ-Orbitrap-FT (Hu *et al.*, 2005), greatly expedited the number of proteins identified by single proteome studies (Graumann *et al.*, 2008; Shi *et al.*, 2007).

2.1.2 Quantification by mass spectrometry

The development of highly sensitive and fast mass spectrometers leads to the analysis of post-translational modifications and protein quantification. During the last decade numerous MS-based quantification techniques have emerged and there is a clear trend in proteomic studies towards MS-based quantification (Bantscheff *et al.*, 2007; Ong *et al.*, 2003; Ong and Mann, 2005; Wilm, 2009).

Every peptide contains a certain amount of stable heavy isotopes (^{13}C , ^{15}N , ^{18}O , and ^2D) at their natural abundances. The isotope pattern of a peptide (also called the isotopic envelope) thus reflects abundance of stable isotopes within the peptide. Artificial incorporation of heavy stable isotopes (i.e. labeling with stable isotopes) induces a mass shift of the peptide's peak (including its isotopic envelope) in the mass spectrum. As mass spectrometry is not a quantitative method *per se*, incorporation of stable isotopes, such as ^{13}C , ^{15}N , ^{18}O or ^2D , can be used for a relative comparison of peptides or proteins from samples to be quantified. Importantly, differently labeled peptides show the same behavior in the mass spectrometer, but feature a mass shift in the MS or MS/MS spectrum according to the incorporated stable isotopes. The peak's signal intensity of the differently labeled peptides derived from various samples reflects the relative quantities of the particular peptide and therefore of the particular protein in the different samples (see Figure 2.2 for an overview).

Incorporation of stable isotopes is achieved by chemical, metabolic, and enzymatic labeling. Chemical labeling is performed on the protein or peptide level, whereas metabolic labeling takes place during cell growth and enzymatic labeling during protein hydrolysis. Performing metabolic labeling guarantees a nearly 100 % labeling efficiency in cell culture, whereas labeling efficiency using chemical labels is typically below 100 % and should be carefully checked during data analysis. Relative quantification is used to compare peptides and finally proteins from different cell states or different protein assemblies. Usually, proteins or peptides from different conditions are differently labeled, pooled in equal amounts and analyzed by LC-MS/MS. In the event of absolute quantification, the sample is spiked with synthetic standard peptides in known amounts during hydrolysis of the proteins or before LC-MS/MS analysis. Absolute quantification is thus used to determine the absolute amount (mass or mole) of proteins in a mixture. To date, there are various methods for relative and absolute quantification available (see Table 2.1 for a summary of the most important and popular methods).

Nonetheless, contrary to introduction of stable isotopes, several label-free quantification approaches, such as spectral count (Old *et al.*, 2005), emPAI (Ishihama *et al.*, 2005) and non-directed LC-MS/MS (Silva *et al.*, 2006), have been reported for relative and absolute MS-based quantification.

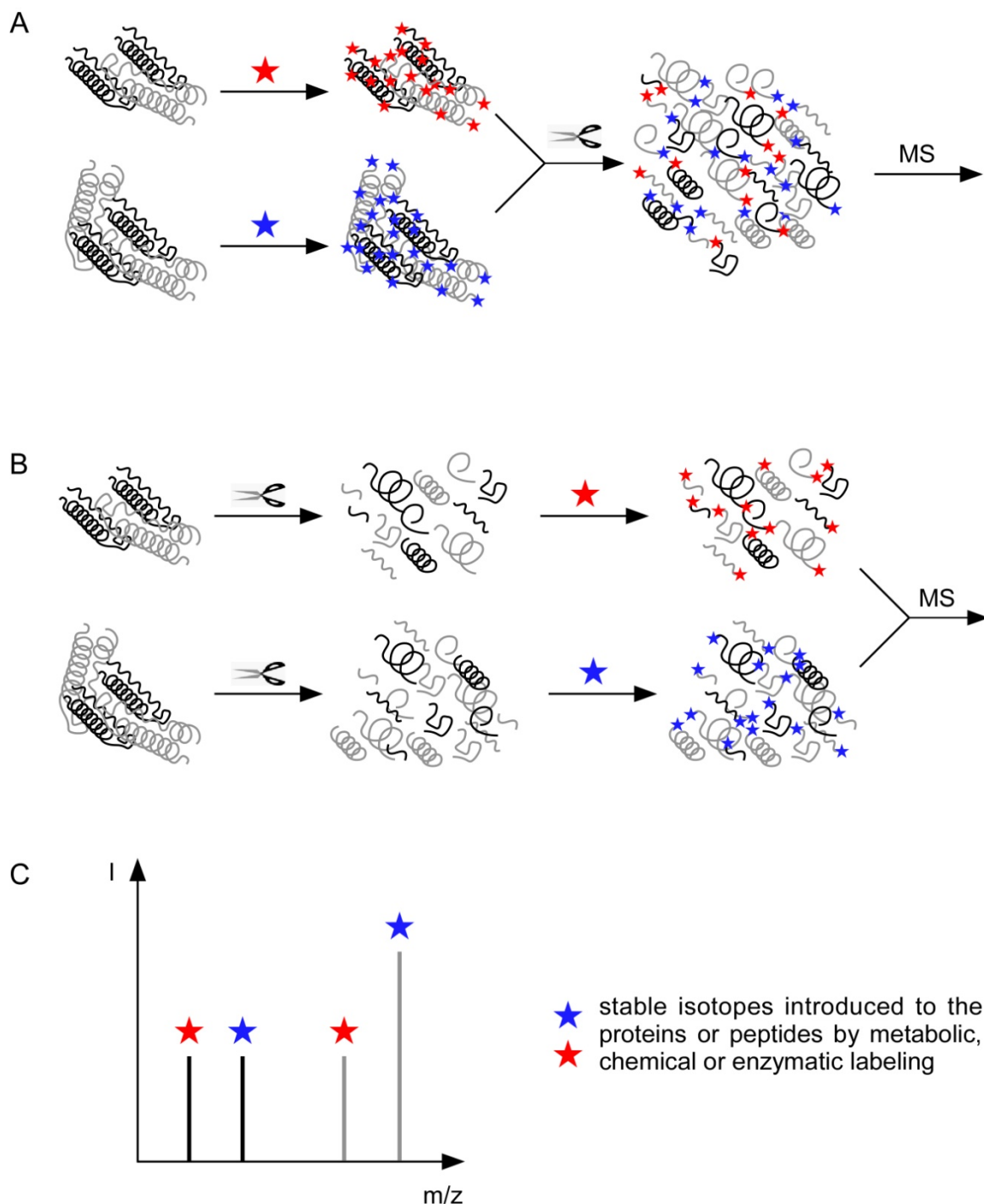


Figure 2.2: Principle of quantification by stable isotope labeling. Stable isotopes are introduced on the protein (A) or peptide (B) level. (A) Proteins from different samples are labeled with stable isotopes by metabolic or chemical labeling. After pooling, the proteins are hydrolyzed with a specific endoproteinase and generated peptides are subsequently analyzed by MS. (B) Proteins from different samples are hydrolyzed in parallel and stable isotopes are introduced to the generated peptides by chemical labeling. Labeled peptides are pooled and analyzed by MS. (C) Differently labeled peptides show same behavior in the mass spectrometer, but feature a distinct mass shift in the MS according to the incorporated stable isotopes.

Table 2.1: Most important and popular MS-based quantification methods in proteomics.

Method	Principle	Quantification	Advantages	Disadvantages	References
AQUA	addition of stable isotope labeled standard peptides	absolute	absolute quantification of proteins/peptides in complex mixtures	complete hydrolysis of the proteins is required, solubility of standard peptides	Desidero and Kai, 1983; Gerber <i>et al.</i> , 2003; Kirkpatrick <i>et al.</i> , 2005
¹⁵ N-labeling	metabolic labeling during cell growth	relative	complete introduction of stable isotopes (labeling efficiency nearly 100 %)	complex data analysis, extremely enriched nitrogen is needed	Oda <i>et al.</i> , 1999
SILAC	metabolic labeling during cell growth	relative	complete introduction of stable isotopes (labeling efficiency nearly 100 %)	no labeling in tissue	Ong <i>et al.</i> , 2002
ICAT	chemical labeling, cysteine specific	relative	enrichment of labeled peptides (reduced sample complexity)	no quantification of proteins containing no cysteine, partial separation of differently labeled peptides during LC	Gygi <i>et al.</i> , 1999
iTRAQ	chemical labeling with isobaric tags, amine specific, quantification in MS/MS	relative	multiplex (up to eight samples), quantification of proteins in tissue	labeling efficiency needs to be checked, specific software for data analysis is required	Ross <i>et al.</i> , 2004
TMT	chemical labeling with isobaric tags, amine specific, quantification in MS/MS	relative	multiplex (up to six samples), quantification of proteins in tissue	labeling efficiency needs to be checked, specific software for data analysis is required	Thompson <i>et al.</i> , 2003
dimethyl labeling	introduction of stable isotopes by dimethylation of N-termini and lysine side chains	relative	cheap reagents, nearly 100 % labeling efficiency, triplex	specific software for data analysis is required	Hsu <i>et al.</i> , 2003; Boersema <i>et al.</i> , 2008
¹⁸ O-labeling	introduction of ¹⁸ O during enzymatic hydrolysis	relative	cheap reagents, simple labeling protocol	incomplete labeling complicates data analysis enormously	Mirgorodskaya <i>et al.</i> , 2000; Reynolds <i>et al.</i> , 2002; Yao <i>et al.</i> , 2001
spectral count	quantification on the number of acquired MS/MS spectra	relative	no labeling required	semi-quantitative	Liu <i>et al.</i> , 2004

2.1.3 Absolute quantification by mass spectrometry

Determination of the absolute amount of proteins in proteomic studies is often derived from a measurement of the absolute peptide concentration. This is achieved by either providing standard peptides or by label-free approaches. In addition, a procedure for absolute quantification of intact proteins using labeled standard proteins has recently been described (Waanders *et al.*, 2007).

Absolute quantification using stable-isotope labeled standard peptides The use of synthetic standard peptides for absolute quantification was originally described in 1983, where enzymatically deuterium labeled standard peptides were used for absolute quantification of enkephalin in thalamus tissue (Desiderio and Kai, 1983). This approach has been refined by Gygi and co-workers, and is now commonly known as AQUA (absolute quantification; Gerber *et al.*, 2003; Kirkpatrick *et al.*, 2005). Heavy isotope labeled standard peptides (also termed AQUA peptides) are synthesized by incorporation of one ^{13}C - and/or ^{15}N -labeled amino acid. As a result, the standard and the endogenous peptide share the same physicochemical properties, including chromatographic co-elution, ionization efficiency, and fragmentation pattern during MS/MS experiments. Importantly, the endogenous and the standard peptide can be distinguished by a distinct mass shift in the MS spectrum caused by the incorporated heavy labeled amino acid. The endogenous and the corresponding standard peptide thus present a peak pair consisting of a light peak (the non-labeled endogenous peptide) and a heavy peak (belonging to the standard peptide harboring incorporated stable isotopes; see Figure 2.3 for an example). The signal intensities of the light (endogenous) and the heavy (standard) peptides reflect the relative amounts. Because the concentration of the standard peptide is known, the absolute amount of the endogenous peptide and finally the protein can thus be determined. Absolute quantification using standard peptides is often applied to measure the level of particular peptide modification (e.g. phosphorylation (Gerber *et al.*, 2003), ubiquitinylation (Kirkpatrick *et al.*, 2005)) or to analyze and validate biomarkers in clinical studies (Pan *et al.*, 2005). The selection of standard peptides is often empirically (Bantscheff *et al.*, 2007), i.e. the choice of synthetic standard peptides results from the analysis of endogenous peptides generated from the proteins under investigation. There are several aspects that have to be taken into account when selecting standard peptides, e.g. possible modification of amino acid residues (e.g. oxidation of methionine), chromatographic elution, ionization efficiency etc. A study about prediction of frequently detected tryptic peptides in a given proteomic platform (so-called proteotypic peptides) might help when selecting standard peptides for absolute quantification (Mallick *et al.*, 2007).

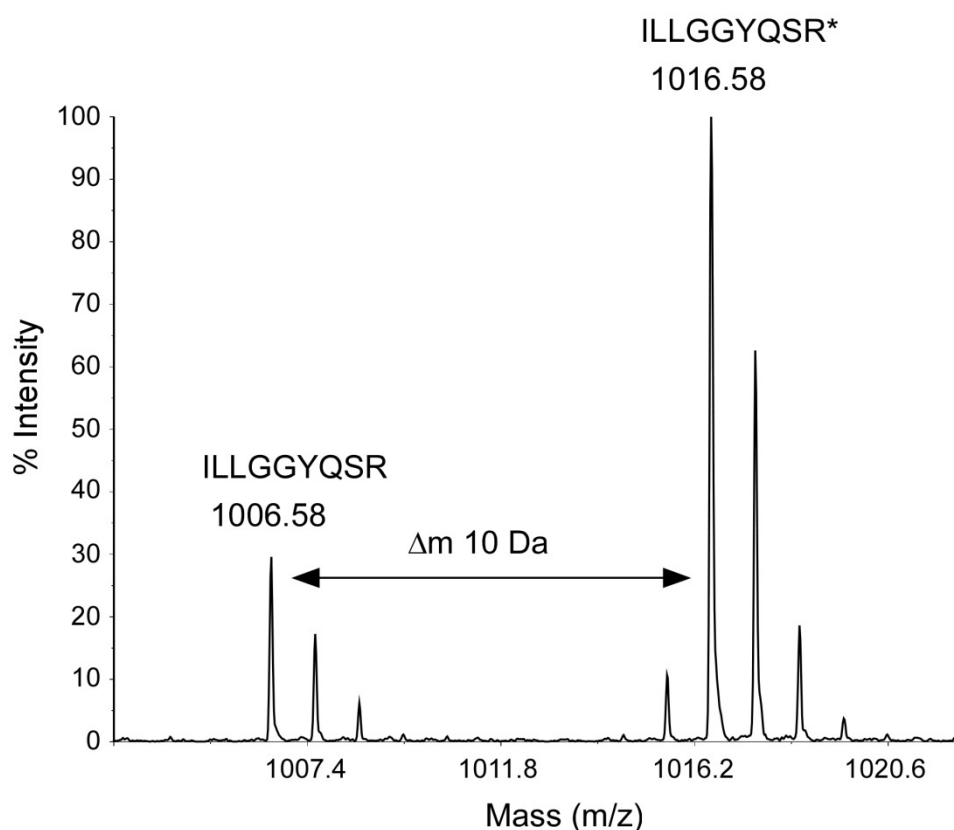


Figure 2.3: Example MS spectrum of an endogenous and the corresponding standard peptide for quantification. MS spectrum of ILLGGYQSR (CDC5L protein) and the corresponding synthetic standard peptide. The endogenous and the standard peptide show according to the incorporated stable isotope labeled amino acid (arginine) a distinct mass shift (10 Da). The spectrum was acquired by nanoLC-offline MALDI-ToF/ToF-MS.

There are however limitations for the absolute quantification using standard peptides. First, the addition of standard peptides to a proteome digest provides quantification of only single or few proteins of the sample. For reliable absolute quantification of a protein in a mixture, several standard peptides per protein have to be first selected and then provided during quantification to achieve more than one reference value per protein. Furthermore, incomplete protein hydrolysis is one of the most critical issue in the event of absolute quantification as it dramatically affects the final result.

To increase confidence in quantification experiments, addition of multiple peptides for each protein to be quantified is recommended. This can be simplified by using labeled standard proteins, which provide multiple peptides for absolute quantification after their hydrolysis. To this end, heavily labeled amino acids are incorporated into entire proteins resulting in heavily labeled standard proteins, which are then added to the sample under investigation. After protein hydrolysis of the endogenous and the standard proteins, standard peptides for all generated endogenous peptides are available. Several approaches using labeled standard peptides have recently been introduced (PSAQ, Protein Standard Absolute Quantification (Brun *et al.*, 2007); Absolute SILAC (Hanke *et al.*, 2008); FLEXIQuant, Full-length expressed

stable isotope-labelled proteins for quantification (Singh *et al.*, 2009)). Very similar is the use of artificial QconCAT proteins, which are assembled from different standard peptide sequences (concatenated signature peptides encoded by QconCAT genes; Pratt *et al.*, 2006). During hydrolysis of QconCAT proteins several standard peptides belonging to diverse proteins are generated allowing the quantification of different proteins in a sample.

Label-free approaches for absolute quantification

Label-free quantification is a method that determines relative or absolute protein amounts without using stable isotope containing compounds. The advantage of label-free quantification is that time-consuming steps of introducing labels to proteins or peptides are not required and that there are no costs for expensive labeling reagents. Furthermore, there is no limit as to the number of experiments to be compared and several peptides per protein are available for quantification. Mass spectral complexity is not increased as in the case of differently labeled samples what provides a higher analytical depth. Unfortunately, label-free approaches are least accurate among the mass spectrometric quantification techniques and they require a high reproducibility at each step, i.e. all experiments need to be accurately reproduced to achieve reliable quantification (Bantscheff *et al.*, 2007).

One possibility to determine absolute protein amounts without labeling of peptides or proteins is to make use of the number of observed and theoretically observable peptides. The protein abundance index (PAI) is then calculated by dividing the number of observed tryptic peptides by the number of theoretically observable peptides from a particular protein and gives an estimate for absolute protein amounts in a complex mixture (Rappsilber *et al.*, 2002). For absolute quantification, the PAI was later converted to an exponentially modified form (emPAI), which is proportional to the protein content in a protein mixture. Ishihama *et al.* have shown that the emPAI-abundances from the actual values are within 63 % on average. Nonetheless, emPAI values are easily calculated but provide only a rough estimate of the absolute protein amounts (Ishihama *et al.*, 2005).

Incomplete digestion is a critical issue for absolute quantification. One way to deal with this is to average the quantities of the three most abundant peptides of every protein. It is generally assumed that some parts of the protein are completely digested and thus the three most abundant peptides reflect the protein concentration. The protein mixture is spiked with an intact standard protein and after hydrolysis the average MS signal response of the standard protein is used to calculate an universal signal response factor (ion counts/mole of protein) for the particular experimental setup. This factor can then be used to determine the concentration of the analyzed proteins within the mixture (Silva *et al.*, 2006).

Top-down quantification of SILAC-labeled proteins Top-down is the analysis of intact proteins instead of generated peptides. To date, only one study about top-down absolute quantification has been released. Waanders *et al.*, 2007 introduced the quantification of intact SILAC-labeled proteins. During the SILAC (stable isotope labeling using amino acids in cell culture, see below) method, cells are grown in media containing different isotope labeled amino acids. When using heavy and light lysine and arginine for SILAC labeling, intact proteins do not interfere with peaks of different charge states between 10 and 200 kDa. The authors have shown that two SILAC proteins (light and heavily labeled) can be quantified with an average standard deviation of 6 % (Waanders *et al.*, 2007).

Absolute quantification to determine the protein stoichiometry within protein complexes By absolutely quantifying proteins in a purified protein complex, the protein stoichiometry of the quantified proteins can be established. In recent years, only few studies addressed the protein stoichiometries within protein complexes using absolute quantification. Two studies, combining chemical labeling of endogenous and standard peptides, were recently introduced. Hochleitner *et al.*, 2005 determined the protein stoichiometry of the spliceosomal U1 small nuclear ribonucleoprotein complex and Holzmann *et al.*, 2009 determined the stoichiometry of the MP1-p14 complex. A different approach has been designed for affinity-tag purified protein complexes (Wepf *et al.*, 2009). An amino acid sequence serving as standard peptide is embedded in the affinity tag and is released after tryptic digestion. The protein is then quantified by adding a stable isotope labeled reference peptide of the tag and the other proteins are quantified by correlational quantification to the tagged protein. This approach benefits from the fact that one stable isotope labeled standard peptide can be used for quantification of different proteins, but requires the protein of interest to be present as tagged protein in the sample.

2.1.4 Relative quantification by mass spectrometry

Several different approaches and techniques are available for relative MS-based quantification. Most of these approaches are based on the introduction of stable isotopes. This is performed by metabolic, chemical or enzymatic labeling. In addition, some label-free methods (e.g. spectral count) are available. As described for absolute quantification (see above), differently labeled peptides with the same physicochemical properties can be distinguished by a mass difference according to the introduced heavy stable isotopes. Relative quantification is then achieved by comparison of the peaks' signal intensities from differently labeled peptides (MS). In case of isobaric labeling reagents, i.e. labeling reagents that differ in their isotope composition but have the same mass, signal intensities from different reporter ions, which are released during fragmentation of the labeling reagent are

used for relative quantification (iTRAQ and TMT, see below). The most important and popular techniques for relative MS-based quantification will be discussed in detail in the following paragraphs.

Metabolic labeling The earliest time point to introduce stable isotopes into proteins is during cell growth (see also Table 2.1). This will reduce errors as samples to be quantified can be combined at a very early step during sample processing (Ong and Mann, 2005). Metabolic labeling followed by quantification was initially described for bacteria using ^{15}N -enriched cell culture medium (Oda *et al.*, 1999). Subsequently, mammalian cells (Conrads *et al.*, 2001) and even small organisms such as *C.elegans* and *D.melanogaster* (Krijgsveld *et al.*, 2003) have been fully labeled with ^{15}N . ^{15}N -labeling achieved complete incorporation of ^{15}N to all amino acids within the cells/organisms, thus providing a high number of peptides suitable for quantification. However, the mass difference between the labeled and unlabeled peptide depends on the number of nitrogen atoms within the amino acid sequence of the particular peptide. This complicates data analysis enormously, making ^{15}N -labeling not the method of choice. In addition, highly enriched ^{15}N labeled sources are required to avoid complicated isotopic distribution from partially labeled peptides (Ong and Mann, 2005).

As a simpler method, stable isotope labeling by amino acids in cell culture (SILAC), was introduced by Mann and co-workers (Ong *et al.*, 2002). In this approach, cells are grown in cell culture medium containing ^{13}C -, ^{15}N - and ^2D -labeled L-lysine and L-arginine. Incorporation of isotope-labeled lysine and arginine ensures a defined mass difference between the differently labeled peptides when using trypsin for proteolytic hydrolysis in proteome studies. As trypsin specifically cleaves proteins C-terminal of arginine and lysine (Olsen *et al.*, 2004) the mass difference of differentially labeled peptides is defined by the isotope incorporation into lysine and arginine provided during cell growth. Using different combinations of stable isotope labeled lysine and arginine allows quantification of up to three samples in one MS measurement (see Figure 2.4 for an overview of differently labeled lysines and arginines). The introduction of the SILAC method led to a high number of quantitative studies during the last years. It was also successfully combined with quantification of post-translational changes in different systems (e.g. Oellerich *et al.*, 2009; Olsen *et al.*, 2010; Pan *et al.*, 2009a). In addition, development of a new software (MaxQuant; Cox and Mann, 2008) facilitated data analysis and made SILAC to a very powerful technique for quantitative proteome studies. Since labeling occurs in cell culture, quantification of proteins in tissue is not possible. However, the production of a SILAC mouse using a “heavy diet” has recently been described (Kruger *et al.*, 2008). Nonetheless, labeling of animals is high priced, requires a big effort and is not in all cases achievable.

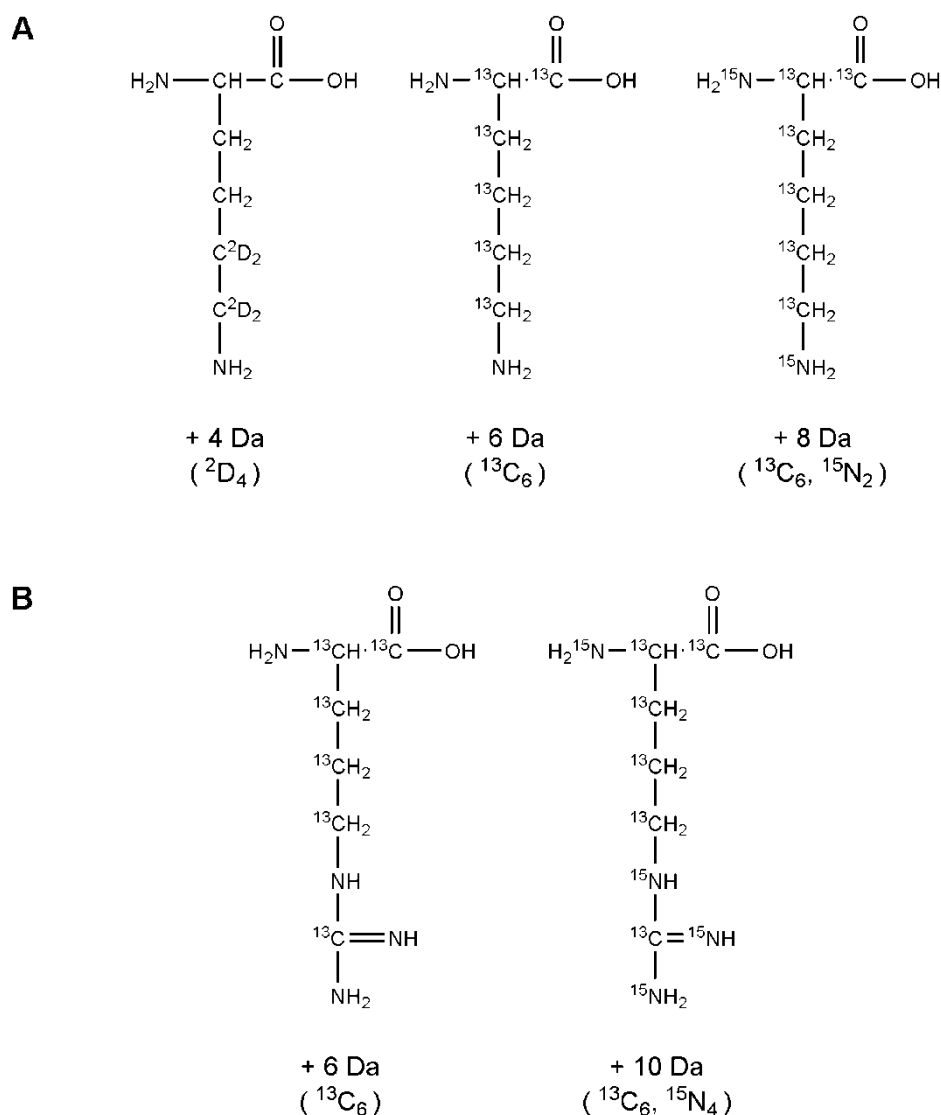


Figure 2.4: Isotope labeled amino acids used for cell growth in SILAC labeling. (A) Introduction of ^{13}C , ^{15}N and ^2D to lysine generates a mass difference of 4, 6, and 8 Da between the labeled and the non-labeled lysine. **(B)** Introduction of ^{13}C and ^{15}N generates a mass difference of 6 and 10 Da between labeled and non-labeled arginine.

Chemical labeling

During chemical labeling, the peptides or proteins are labeled with isotope containing reagents. At this, advantage is taken of the chemical reactivity of amino acid side chains or the peptide's N-terminus (for an overview see Table 2.1). The first reagent for chemical labeling was the isotope-coded affinity tag (ICAT; Gygi *et al.*, 1999). ICAT is a cysteine specific reagent consisting of a cysteine-reactive iodoacetyl group, a linker containing either zero (light ICAT) or eight (heavy ICAT) deuterium atoms and a biotin group for affinity purification of cysteine-derivatized peptides. Light and heavy ICAT-labeled protein samples are then combined and subsequently hydrolyzed. The advantage of ICAT is the reduced sample complexity after affinity purification, which allows protein quantification in complex samples. Complications arose with proteins containing none or only few cysteine

residues as these are excluded from quantification. Further, the presence of deuterium atoms led to partial separation by chromatography and the size of the label affected fragmentation in the mass spectrometer. Therefore, an improved reagent (cICAT) using ^{13}C labeling of the linker (resulting in a mass difference of 9 Da) and containing an acid-cleavable biotin group was developed (Figure 2.5 A).

Another approach is the isotope-coded protein label (ICPL; Schmidt *et al.*, 2005), which is based on the N-hydroxysuccinimide (NHS) chemistry targeting the epsilon-amino group of lysine residues in proteins (Figure 2.5 B). N-nicotinoyloxy-succinimide is used in a light (d0) and a heavy (d4) form allowing for relative quantification of two different samples. Further improve by incorporation of ^{13}C atoms to the nicotinoyl group resulted in a quadruplex reagent.

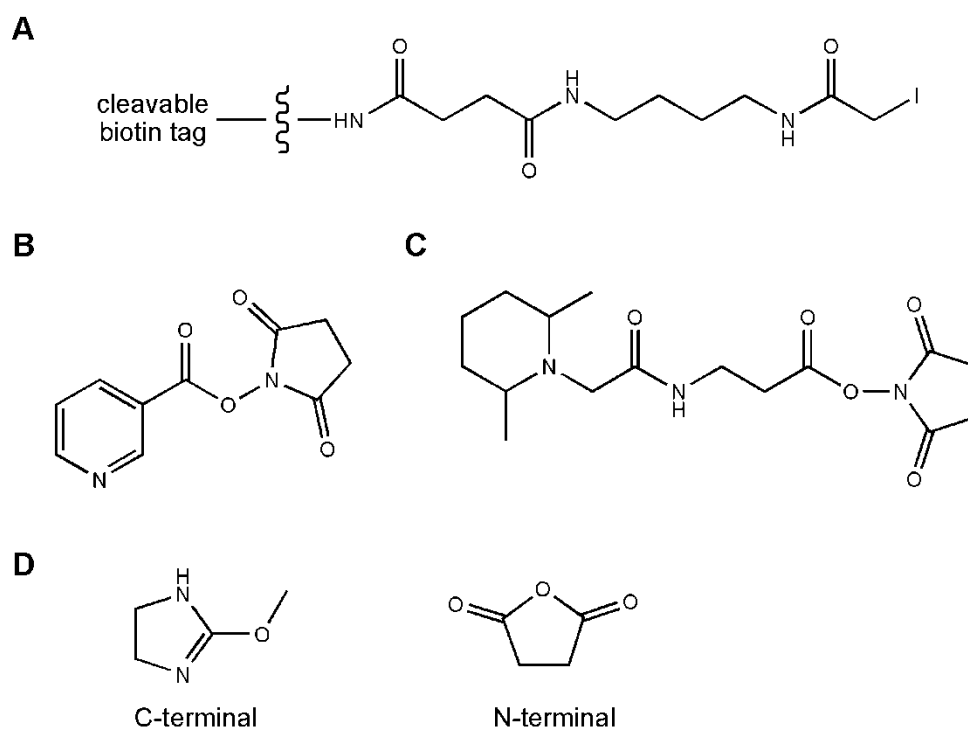


Figure 2.5: Chemical labeling reagents. (A) cICAT reagent. The exact structure of the cleavable biotin tag has not been publicly disclosed. (B) N-nicotinoyloxy-succinimide used for ICPL quantification. (C) Structure of the TMT reagent. (D) C-terminal and N-terminal reagents for IPTL. 2-Methoxy-4,5-dihydro-1H-imidazole (C-terminal) and succinic anhydride (N-terminal).

A different approach, based also on NHS chemistry, is iTRAQ (isobaric tags for relative and absolute quantification; Ross *et al.*, 2004). iTRAQ reagents are amine specific and label lysine side chains and amino termini of peptides. They are multiplexing reagents, i.e. due to the isotope composition of the reagents several samples can be compared relative to another in one experiment. iTRAQ 4-plex and 8-plex reagents are available allowing relative quantification of four and eight samples in one experiment. They consist of an amine reactive

group, a balance group and a reporter group (Figure 2.6). iTRAQ reagents are isobaric, i.e. all reagents have the same mass and all labeled peptides have a mass tag of 144.1 Da in case of the 4-plex iTRAQ reagents. Upon fragmentation in the MS/MS experiment, all isobaric tags release a marker ion (so-called reporter ions) of 114.1, 115.1, 116.1 and 117.1 Da, respectively, and a neutral fragment (28, 29, 30, 31 Da, respectively), which is not detected in the mass spectrometer (see Figure 2.6 for a detailed description of the 4-plex iTRAQ reagents). In contrast to the methods described above, iTRAQ-labeled peptides do not show a mass difference in the MS. Quantification of differently labeled peptides is therefore only achieved upon fragmentation. The described iTRAQ reagents were further modified so that labeling of eight samples can be performed simultaneously. iTRAQ has several advantages: (i) iTRAQ-labeling is multiplexing, i.e. up to four or eight samples can be compared in one quantification experiment. (ii) Since all labeled peptides have the same mass, the signal intensity in the MS is enhanced. (iii) The signal intensity of the fragment ions is enhanced as well because the mass tag is completely cleaved during fragmentation. (iv) iTRAQ labeling offers the opportunity to create an internal standard that contains a mixture of all samples, so that more than four or eight samples can be quantified in relation to the internal standard.

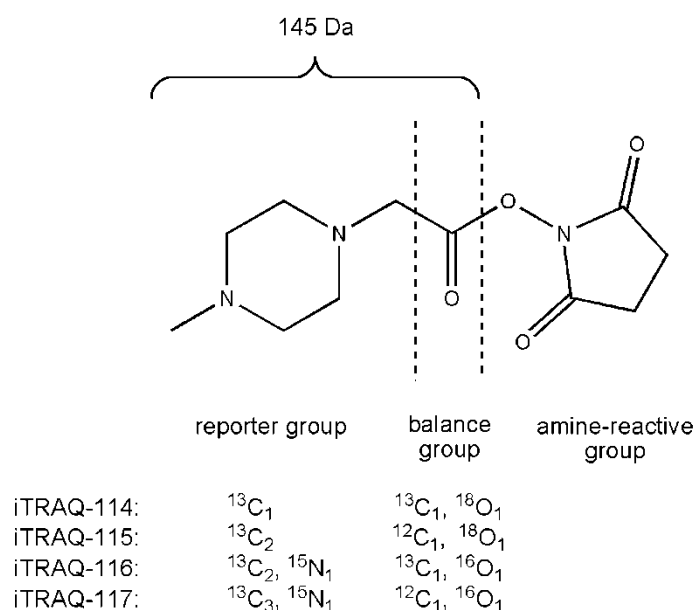


Figure 2.6: Structure of the 4-plex iTRAQ reagent. The reagent consists of an amine-reactive group, a balance group and a reporter group. The isobaric tags have a mass of 145 Da resulting in a mass difference of 144.1 Da for all differently labeled peptides. Upon fragmentation, the reporter group and a neutral fragment (balance group) are released. Due to the isotope composition of the reporter and the balance group, the generated reporter ions show different masses for the different iTRAQ reagents.

A very similar approach is the labeling with TMTs (tandem mass tags; Thompson *et al.*, 2003). The chemical structure resembles the one of iTRAQ reagents and the quantification

procedure is the same (Figure 2.5 C). Using TMTs, up to six samples can be compared in one quantification experiment.

A different MS/MS-based quantification approach is the isobaric peptide termini labeling (IPTL; Koehler *et al.*, 2009). Labeling of both peptide termini with tetradeuterated and non-deuterated reagents leads to isobaric peptides of the samples to be compared. After Lys-C hydrolysis of the proteins the peptides' C-termini are labeled with 2-methoxy-4,5-dihydro-1*H*-imidazole (MDHI) or tetradeuterated MDHI-d4 whereas the N-termini are labeled tetradeuterated succinic anhydride (SA-d4) or SA, respectively (Figure 2.5 D). Fragmentation during MS/MS causes the cleavage of the N-terminal succinic anhydride resulting in fragment ion pairs that contain either the C-terminal MDHI- or MDHI-d4-label. The advantage of this approach is the generation of several quantification data points for each peptide and the opportunity to use mass spectrometers with limited capabilities in the low molecular mass range.

Dimethyl labeling of peptides at the N-terminus and epsilon-amino groups of lysine residues has recently been introduced (Hsu *et al.*, 2003). The use of differentially ²D- and ¹³C-labeled formaldehyde and sodium cyanoborohydride allows triplex labeling (Boersema *et al.*, 2008). Dimethylation of amino groups causes a mass increase of 28, 32, and 34 Da, respectively, resulting in a mass difference of 4 and 6 Da between the differently dimethylated peptides. This labeling method has several advantages: (i) It uses inexpensive reagents and is thus a cost-effective labeling technique in comparison to other stable isotope-reagents. (ii) It is a reliable method providing a 100 % labeling efficiency in almost all cases. (iii) Different labeling protocols (in-solution, online and on-column) for different amounts of sample have been described allowing for automation and high-throughput proteomics (Boersema *et al.*, 2009). (iv) No reaction byproducts have been observed. (v) As dimethyl labeling is based on a simple chemical reaction, a large number of samples can be quantified by this method.

Enzymatic labeling Isotopic labels can also be incorporated during enzymatic proteolysis (Mirgorodskaya *et al.*, 2000; Reynolds *et al.*, 2002; Yao *et al.*, 2001). Proteolytic hydrolysis in "heavy" (H₂¹⁸O) and "normal" (H₂¹⁶O) water leads to introduction of two ¹⁸O atoms at the peptide's C-terminus resulting in a mass shift of 4 Da between the labeled (¹⁸O) and non-labeled (¹⁶O) peptide. Very common is also enzymatic labeling after proteolysis in a second incubation step with the protease. Suitable enzymes are Trypsin and Glu-C as they introduce two ¹⁸O atoms to the C-terminus resulting in a sufficient mass shift for differentiation of isotopomers (4 Da). Asp-N and other enzymes introduce only one ¹⁸O atom (Rao *et al.*, 2005) and should therefore be avoided. Advantageous over chemical labeling is that side reactions do not occur. A critical point is acid- and base-catalyzed back-exchange at extreme pH values (Schnolzer *et al.*, 1996), whereas mild acidic conditions during ESI-

and MALDI-MS analysis guarantee stability of the introduced label. However, incomplete labeling (i.e. incorporation of only one ^{18}O atom) complicates data analysis and requires correction of overlapping isotopic patterns (Johnson and Muddiman, 2004; Ramos-Fernandez *et al.*, 2007).

Label-free relative quantification

The observation that the more of a particular protein is present in a sample the more tandem MS spectra are collected during MS analysis led to the assumption that there is a correlation between number of spectra and the protein amount. Spectral count has therefore been applied for relative quantification in different studies (Gilchrist *et al.*, 2006; Washburn *et al.*, 2001). Liu *et al.*, 2004 have analyzed this correlation more in detail and found a linear correlation over 2 orders of magnitude between the number of spectra and the relative protein abundance whereas no correlation between relative protein amounts and peptide count or sequence coverage has been observed. Spectral count is less accurate for small changes between proteins, but was shown to be very accurate for large changes allowing a measure how protein changes overall (Liu *et al.*, 2004). However, dynamic exclusion of ions that already have been selected for fragmentation during MS analysis is disadvantageous for accurate quantification (Old *et al.*, 2005). Spectral counting is a very attractive approach but neglects that no physical property of a peptide is measured. Furthermore, the assumption that the response for every protein is the same is misleading (Bantscheff *et al.*, 2007). Due to different amino acid sequence and different properties of the generated peptides (e.g. chromatographic behavior) the number of spectra detectable varies for different proteins.

Certainly, all methods described for absolute quantification can also be applied for relative quantification by absolutely quantifying proteins in different samples and compare the absolute protein amounts in a relative manner.

2.1.5 Quantification by mass spectrometry to analyze dynamic protein transitions

Relative MS-based quantification can not only be used to compare two or multiple different cell stages, complexes etc., it can also be applied to monitor protein dynamics or protein assembly of multi-protein components. To date, only few studies about dynamic protein changes were published. All these studies use metabolic labeling and take advantage of relative peptide ratios at different time points to display protein changes over time.

Williamson and co-workers developed a method called pulse-chase monitored by quantitative mass spectrometry (PC/QMS; Talkington *et al.*, 2005; Williamson, 2005). They analyzed the assembly of the 30S ribosomal subunit using 16 rRNA and ^{15}N -labeled 30S proteins as starting products. At various time points, protein binding was chased with excess

of unlabeled proteins and completely formed 30S subunits were purified. The ratios of labeled to unlabeled peptides for each protein was then plotted as a function of time providing a progress curve for the binding of this specific protein during the assembly event. This method enabled a detailed and quantitative kinetic characterization of this specific assembly process.

Compared to ^{15}N -labeling, SILAC shows several advantages (see above). It is therefore not surprising that SILAC approaches were recently introduced to monitor protein changes. To this end, cells were grown in media supplemented with stable isotope labeled amino acids and differently labeled cells were treated with e.g. inhibitors or inducers. Triple SILAC allows the comparison of three different time points in one experiment. Performing multiple experiments using the zero time point as a reference allowed construction of dynamic profiles for single proteins. This method was successfully applied to analyze the nucleolar proteome dynamics (Andersen *et al.*, 2005; Lam *et al.*, 2007), endoplasmic reticulum stress response (Mintz *et al.*, 2008), and protein turnover rates in intact animals (Doherty *et al.*, 2005).

Pulsed SILAC (pSILAC) is a novel SILAC variant where cells cultivated in normal medium are pulse-labeled by transferring the cells in medium containing stable isotope labeled amino acids. Newly synthesized proteins are labeled with heavy stable isotopes whereas pre-existing proteins present before labeling remain in the light form. Protein changes upon treatment of the cells can thus be monitored. By this method, changes in protein synthesis induced by microRNAs (Selbach *et al.*, 2008) and translational regulation of cellular iron homeostasis (Schwanhausser *et al.*, 2009) have been successfully analyzed.

2.1.6 Quantifying mass spectrometric measurements

A mass spectrometer consists of an ion source, a mass analyzer that measures the mass-to-charge ratio (m/z) of the generated ions, and a detector that detects the number of ions at each m/z value. Two suitable techniques for the ionization of peptides and proteins are ESI and MALDI (see above). The analytes are ionized out of a solution (ESI) or out of a crystalline matrix via laser pulses (MALDI). Whereas these two techniques are set, a wider range of mass analyzers is available. Important parameters are sensitivity, resolution, mass accuracy, and the ability to generate information-rich mass spectra. The basic types of mass analyzers currently used in proteomic studies are time-of-flight (ToF), ion trap, quadrupole, Fourier transform ion cyclotron resonance (FT ICR), and the Orbitrap mass analyzers. MALDI is usually coupled to ToF analyzers whereas all other mass analyzers are commonly coupled to an ESI source. In addition, hybrid mass analyzers, such as hybrid quadrupole ToF

(Qq-ToF) mass analyzers, are available. Here, precursor ions for MS/MS experiments are selected in the first quadrupole, fragmented in a collision cell, and the fragment ion masses are analyzed in the ToF. These instruments are commonly coupled to an ESI source and have high sensitivity, resolution, mass accuracy, and - most important - generate information-rich fragment spectra.

Quantitative information can be obtained from MS or MS/MS signals (see above). The advantage of quantification from MS spectra is that several spectra are available and ion intensities are high. However, very low and very strong signals are problematic. Low signals are hard to distinguish from background noise and very strong signals can saturate the detector, what in turn limits precision of the measurement. The latter is more often observed in ToF and Qq-ToF instruments compared to ion traps as the latter can control the number of ions before detection. Using fragment ion intensities (MS/MS) for quantification, detector saturation and interference with background ions can be neglected. Low intensities are rather a problem as poor ion statistics may result in less robust quantification. However, limits to quantification of complex samples can often be attributed to interference of co-eluting components of similar masses.

Different specific mass spectrometric scanning modes are used to read out signal intensities during quantitative analysis. The commonly used and most powerful techniques will be described in the following paragraphs.

MALDI-ToF-MS Technical advances have enhanced the application of MALDI mass spectrometry for proteomics but also for quantitative studies. The investigation of tandem time-of-flight instruments (ToF-ToF) allows the fragmentation of precursors and thus unambiguously assigns the species to be quantified (Bienvenut *et al.*, 2002). Decoupled MS and MS/MS analysis allows for data-dependent MS/MS analysis, and the manner of sample preparation reserves most of the sample for repeated analysis. Furthermore, the generation of predominant singly charged ions during MALDI simplifies data analysis (Pan *et al.*, 2009b). During MALDI-ToF-MS quantitative information is often obtained from the area under the peaks to be quantified. However, LC-offline allows the generation of extracted ion chromatograms (XICs, see below) over the whole chromatographic timescale for the peptide of interest. However, in a MALDI spectrum there is often a large discrepancy between ion intensities and analyte concentration on the MALDI target. Ionization of the peptides occurs via proton transfer from the acidic matrix and the ionization efficiency is therefore dependent on the proton affinity of the different peptides. The presence of a peptide with a very high proton affinity can consequently influence the intensity of other ions through ion suppression effects (Knochenmuss *et al.*, 2000). For this reason, peptides below a mass of 3 kDa can only be quantified using standard peptides, which have the same chemical structure and thus

show the same behavior during the ionization process (see Figure 2.3 for an example of a spectrum of an endogenous and a standard peptide acquired by MALDI-ToF-MS). Peptides or proteins larger than 3 kDa have such a high proton affinity that ion suppression effects are very unlikely (Wilm, 2009).

Extracted ion chromatograms (XICs) Electrospray ionization is well suited for quantitative measurements if the flow rates are 100 nL/min or lower. In this case, spectral intensities correspond to the analyte concentrations very well. For higher flow rates the electrospray is unsteady and ion intensities become irregular and do not reflect molecular concentrations (Wilm, 2009). In addition, ion suppression has been observed for higher flow rates (Schmidt *et al.*, 2003). The peptide's signal in the MS analysis can be plotted over time while the peptide is eluting from the chromatography column, i.e. so-called extracted ion chromatograms (XICs) for defined peptides can be generated. The XIC signal is related to the relative amount for the same peptide at the same experimental conditions and can therefore be used for comparison of the same peptide in different samples. XICs are usually generated from samples analyzed by LC-ESI-MS whereas generation of XICs from LC-offline MALDI-MS analyses is also possible (see above). Using high mass accuracy mass spectrometers, greater than two-fold changes of a peptide can be measured. As one peptide in a complex mixture is not always selected for fragmentation in different MS runs (Kuster *et al.*, 2005) it is critical to find and quantify the correct peptide in the different analyses to be compared. Development of required software and normalization of the runs to be compared by spiked-in calibrants can overcome this problem. The great advantage of XIC-based quantification is that no labeling strategy is used and almost every different MS analyses can be compared as long as they were performed under the same conditions. However, this requires a very high reproducibility during sample preparation, chromatography, and MS analysis. One alternative to circumvent this problem is the use of stable isotope labeled standard peptides in known amounts. As the labeled and the non-labeled peptide show the same behavior during chromatography and MS generated XICs from both peptides show the same chromatographic retention time. The area under the XICs can then be used for absolute or relative quantification (see Figure 2.7 for an example of XICs of a labeled and a non-labeled peptide).

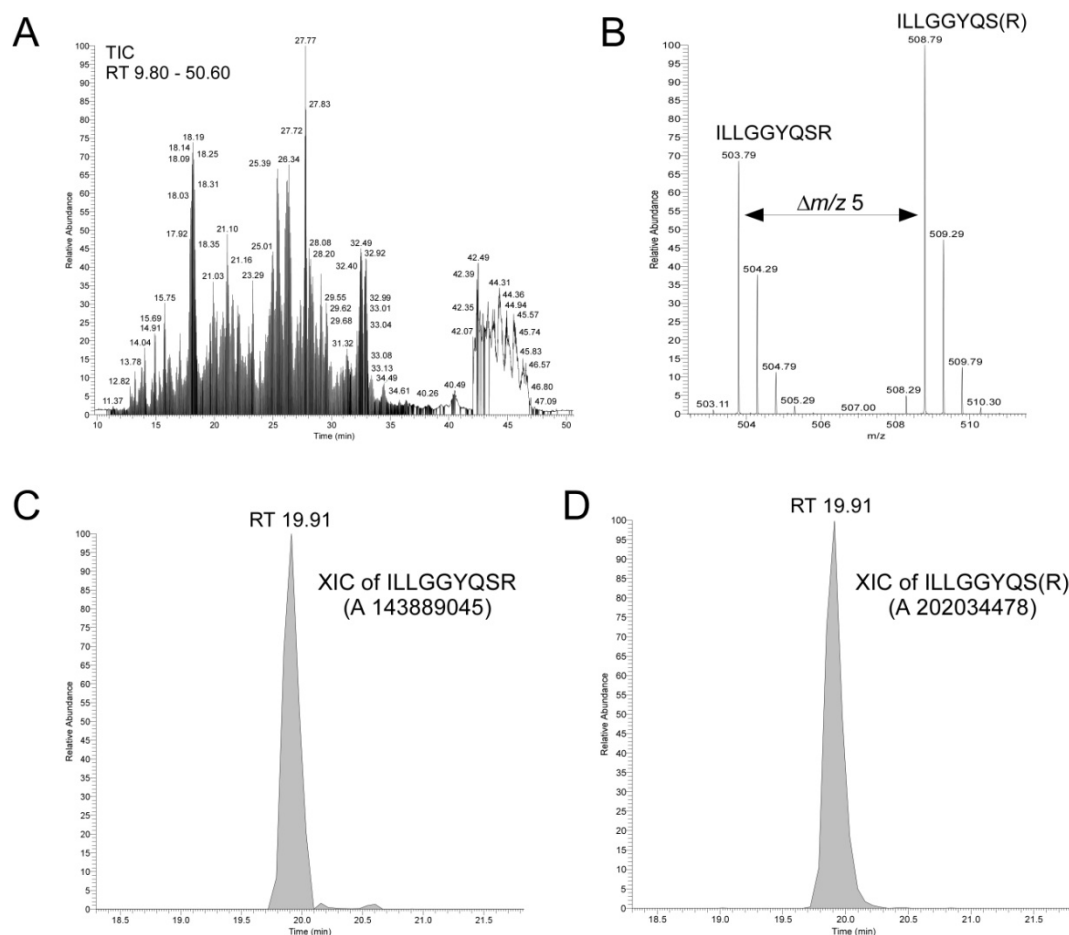


Figure 2.7: Example of Extracted ion chromatograms (XICs) to read out the signal intensities for quantification. (A) Total ion count of the hydrolyzed hPrp19/CDC5L complex. **(B)** MS spectrum of ILLGGYQSR (CDC5L protein) and the stable isotope labeled standard peptide. According to the incorporated stable isotope labeled amino acid (arginine) a mass difference of 5 m/z between the doubly charged peptides is observed. **(C)** Extracted ion chromatogram of ILLGGYQSR. **(D)** Extracted ion chromatogram of the standard peptide ILLGGYQS(R). The endogenous and the standard peptide show the same retention time. The peak area of the signals can be used for quantification.

Multiple Reaction Monitoring (MRM)

A quadrupole mass analyzer can be operated as a mass filter allowing only one specific m/z to pass the quadrupole. This feature is utilized to detect the specific transition from a given precursor to a user-defined fragment ion (selected reaction monitoring, SRM) in a triple quadrupole mass spectrometer. To this end, the precursor mass is selected in quadrupole Q1, whereupon (after fragmentation in q2) the user-defined fragment ion is detected in Q3 (SRM transition). This technique has been extended to the detection of multiple fragment ions per precursor and is then called multiple reaction monitoring (MRM; Anderson and Hunter, 2006; Kuhn *et al.*, 2004; Stahl-Zeng *et al.*, 2007, Figure 2.8). The MRM signal is quantitative over 4-5 orders of magnitude (Wolf-Yadlin *et al.*, 2007) and can be used for relative and for absolute quantification as outlined in section 2.1.3. One run, in which all MRM transitions are monitored once, is called a duty cycle and its length is dependent on the dwell time (i.e. the time to accumulate ions in the quadrupole), the

number of precursors, and the number of MRM transitions per precursor. The duty cycle is repeated consistently during the MRM analysis. It should be repeated several times while the peptide is eluting from the chromatography column to achieve a certain number of data points, which are required to record a sufficient MRM signal. The length of the duty cycle is therefore not unlimited and needs to be adjusted for every analysis. MRM is a very sensitive and, as two mass filters are connected in series, a very specific method. For this reason, it is not only a quantification technique but also very well suited for targeted proteomics, such as biomarker verification and validation.

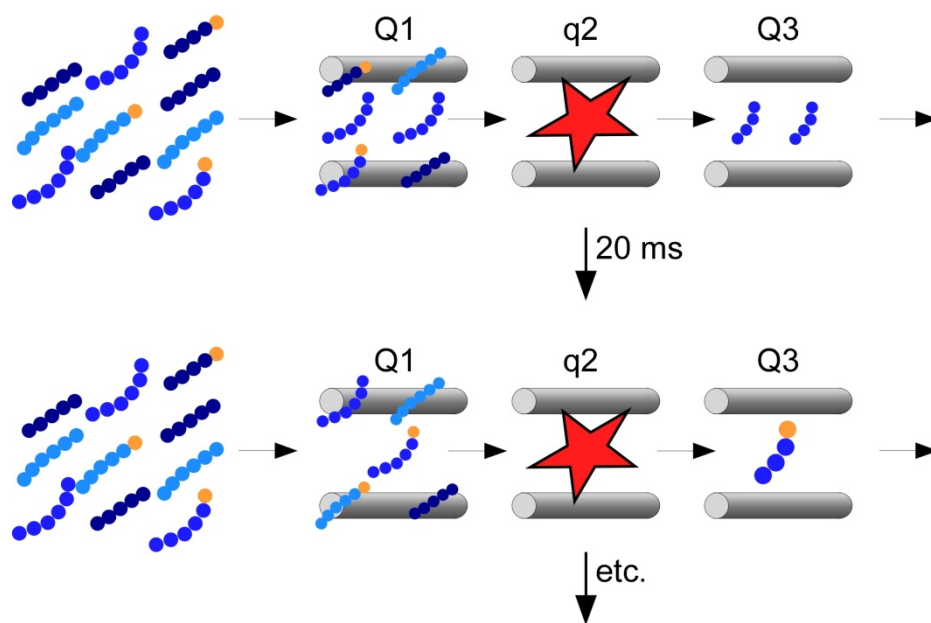


Figure 2.8: MRM workflow. A peptide mixture containing endogenous and stable isotope labeled standard peptides is analyzed by MRM in a triple quadrupole mass analyzer. In quadrupole 1 (Q1) a defined precursor ion is selected and, upon fragmentation in q2, a specific fragment ion of the selected precursor is detected in Q3. After a defined time (e.g. 20 ms) the next precursor is selected for MRM analysis. One duty cycle comprises the analysis of all MRM transitions scheduled for the peptide sample to be analyzed. The obtained signal for the MRM transitions is quantitative and can be used for relative and absolute quantification of the peptides.

Parallel fragmentation (MS^E)

A very different approach to analyze and quantify peptides is parallel fragmentation (also known as MS^E ; Silva *et al.*, 2006). Here, all precursors, entering the mass spectrometer at a particular time, are fragmented without selecting an individual precursor ion. For this purpose, the mass spectrometer switches continuously between MS and MS/MS mode, thus delivering an almost complete data set of the sample. Fragment ions are assigned retrospectively to their precursors by their identical time profiles. However, this requires a well-resolved chromatographic system and the assignment of fragment ions to their precursor might fail due to fragment ions that are generated by several precursors simultaneously. Therefore, the MS^E scanning method is performed on Qq-ToF mass spectrometers that might compensate probable mis-

assignments by their high mass resolution in MS/MS. Once the experimental data are obtained, the data set can be interrogated for a specific ion and a set of specific fragment ions (pseudo-MRM; Niggeweg *et al.*, 2006). This interrogation is not as specific as a real MRM experiment but has been proven for direct, label-free quantification (Wilm, 2009).

2.2 The spliceosome

2.2.1 Eukaryotic pre-mRNA splicing

Gene expression is the transcription of DNA in messenger RNA (mRNA) and the translation of mRNA into proteins. Most eukaryotic genes are expressed as precursor mRNAs (pre-mRNAs) consisting of coding sequences (exons) and non-coding sequences (introns). One of the major steps during gene expression is therefore the excision of introns and the ligation of adjacent exons to yield mature mRNA (pre-mRNA splicing).

2.2.1.1 Structure of eukaryotic pre-mRNAs

Eukaryotic pre-mRNAs consist of exons and introns. The introns are defined by very short, conserved sequences at the 5'exon/intron and 3'intron/exon junctions (5' and 3' splice sites, respectively) as well as the branch point site (BPS), which is located 18-40 nucleotides upstream of the 3' splice site (3'ss) and involves a conserved adenosine (branch point adenosine). In addition, pre-mRNAs of higher eukaryotes usually contain a polypyrimidine tract consisting of 10-15 pyrimidines located between the BPS and the 3'ss (Figure 2.9 A). These *cis*-acting elements are conserved in yeast but are degenerated in higher eukaryotes with only GU (5'ss) and AG (3'ss) at the intron ends being invariable. Two types of spliceosomes that catalyze pre-mRNA splicing co-exist in most eukaryotes. There is a U2-dependent spliceosome, which catalyzes splicing of U2-type introns and which is found in all eukaryotes (major spliceosome). In addition, there is a U12-dependent spliceosome (the minor spliceosome), which is only present in a subset of eukaryotes.

2.2.1.2 The biochemical splice reaction

The splicing process follows two subsequent transesterification reactions (Figure 2.9 B): First, the 2' hydroxyl group of the branch point adenosine attacks the phosphodiester bond at the 5' splice site (5'ss). This results in a 2'-5' phosphodiester bond formed between the branch point adenosine and the first nucleotide of the intron, resulting in an intron-exon 2 lariat and a free exon 1. In the second step of splicing, the 3' hydroxyl group of the free exon 1 attacks the phosphodiester bond of the 3' splice site (3'ss). Exon 1 and exon 2 are thus

ligated and the intron lariat is released (Burge *et al.*, 1999; Green, 1991; Moore *et al.*, 1993; Moore and Sharp, 1993; Nilsen, 1998).

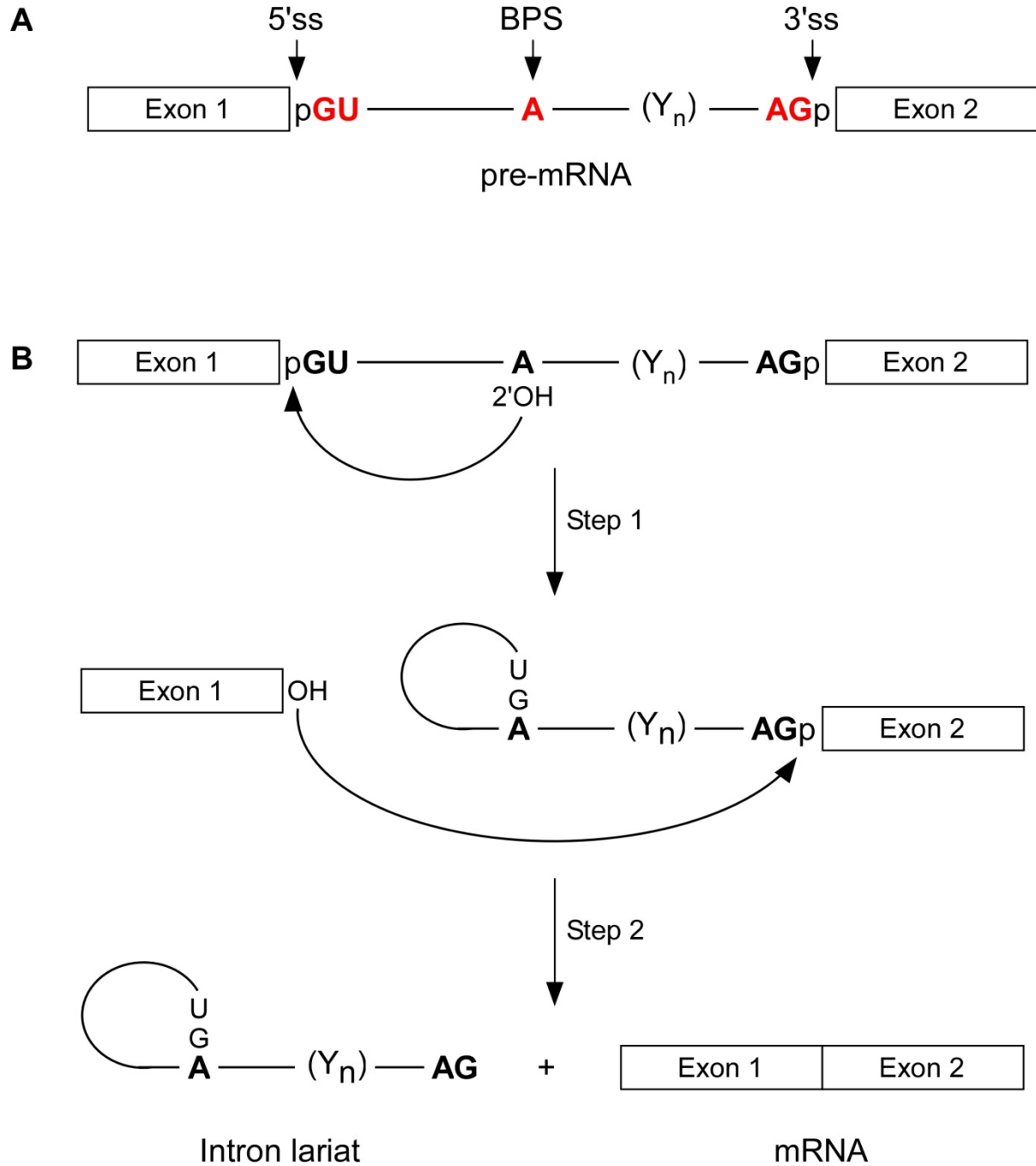


Figure 2.9: Schematic representation of the splicing process. (A) pre-mRNAs in higher eukaryotes consist of exons and interjacent introns. The introns are defined by a 5' and a 3' splice site, a branch point site containing a conserved adenosine and in most cases a polypyrimidine tract (Y_n). Conserved nucleotides of U2-dependent introns in higher eukaryotes are highlighted in bold red. Note that conserved nucleotides of U12-dependent introns differ from U2-dependent introns. (B) The two-step mechanism of splicing. See text for details.

2.2.1.3 The catalytic machinery

The spliceosome is responsible for the recognition and base pairing of the 5' and 3' splice site and for catalysis of the excision of the defined introns and the ligation of adjacent exons. It assembled in a stepwise fashion on each intron that is processed. The catalytic site of the spliceosome is formed by an extensive RNA-RNA, RNA-protein and protein-protein network. The RNA network is designed by several inter- and intra-snRNA interactions, as well as interactions between snRNAs and the pre-mRNA (reviewed by Nilsen, 1998). The spliceosome is highly dynamic and undergoes several structural rearrangements during formation of an active spliceosome. These structural changes are facilitated by proteins of the DExD/H-box protein family, which are able to rearrange RNP and RNA-RNA interactions while hydrolyzing ATP (reviewed by Schwer, 2001; Staley and Guthrie, 1998). Due to the extensive RNA network, which is designed during spliceosomal activation, the spliceosome resembles an RNA enzyme with active sites comprised of RNA. However, the catalytically active structure requires the assembly of numerous spliceosomal proteins.

2.2.2 Components of the spliceosome

2.2.2.1 The U snRNPs

A large number of *trans*-acting factors interact with the pre-mRNA during pre-mRNA splicing. The major spliceosome assembles from U1, U2, U4/U6 and U5 snRNPs and several non-snRNP proteins. All U snRNPs consist of a specific uridine-rich RNA (U snRNA) and a set of particle-specific proteins (Table 2.2). The sequence of the U snRNAs is highly conserved evolutionary (Guthrie and Patterson, 1988) and all U snRNAs (except U6 snRNA) possess a conserved Sm binding site (Liautard *et al.*, 1982). U1, U2, U4 and U5 snRNAs are transcribed by RNA polymerase II and transported to the cytoplasm, where seven Sm proteins are loaded onto the Sm site of the RNAs. They are further modified by addition of a m₃G (2,2,7-trimethylguanosine) cap. U6 snRNA is transcribed by RNA polymerase III and carries a gamma-monomethyl phosphate cap (Kunkel *et al.*, 1986; Singh and Reddy, 1989). In addition, all snRNAs are post-transcriptionally modified by pseudouridines, 6-methyladenosines and 2'-O-methylation at the ribose residues (Massenet *et al.*, 1998). They are involved in RNA-RNA and RNA-protein interactions and thus play an important role in the formation of a catalytically active spliceosome (see above).

Common to all U snRNPs (except U6) are the seven Sm proteins (E, F, G, D1, D2, D3, and B/B'). They form a ring-shaped heptamer and bind the U snRNA at the Sm site via a Sm motif to form the Sm core RNPs (Raker *et al.*, 1999; Raker *et al.*, 1996; Urlaub *et al.*, 2001).

U6 snRNA does not contain a Sm binding site but associates with a group of related proteins, called Sm-like proteins (LSm2-LSm8). They also form a heptameric ring and bind the U6 snRNA at the 3' end (Achsel *et al.*, 1999; Vidal *et al.*, 1999). In addition to Sm and LSm proteins, all U snRNPs contain a specific set of proteins, which contribute to the functionality of the U snRNPs during pre-mRNA splicing. Biochemical and immunological procedures allowed for purification of the snRNPs (Kastner and Luhrmann, 1999).

The 12S U1 snRNP contains beside the Sm proteins three specific proteins: The U1-70K, U1-A and U1-C proteins. From these, the U1-70K and the U1-A bind directly to the snRNA (Patton *et al.*, 1989; Patton and Pederson, 1988; Query *et al.*, 1989; Scherly *et al.*, 1989; Urlaub *et al.*, 2000), whereas the U1-C protein only binds to the U1 snRNP in the presence of the Sm core and U1-70K (Nelissen *et al.*, 1994). The U1-C protein is important for splicing activity, as it directly contacts the pre-mRNA near the 5'ss stabilizing snRNA-pre-mRNA interactions (Heinrichs *et al.*, 1990; Pomeranz Krummel *et al.*, 2009).

The U2 snRNP includes several proteins. It was first described as 12S U2 snRNP consisting of the U2 snRNA, the Sm proteins and two other proteins, namely A' and B'', and was later found to be present as splicing active 17S U2 snRNP including two further heteromeric splicing factors, called SF3a and SF3b (Behrens *et al.*, 1993; Brosi *et al.*, 1993; Will *et al.*, 2002). SF3a is composed of three proteins with apparent molecular weight of 120, 66 and 60 kDa, whereas SF3b consist of seven proteins with molecular weight of 10, 15 (SF3b14a and b), 49, 130, 145, 155 kDa. Almost all SF3a and SF3b proteins contact the pre-mRNA near the branch point site and are thus essential for the spliceosomal assembly (Gozani *et al.*, 1996; Kramer *et al.*, 1999). In addition to SF3a and SF3b, several other proteins (e.g. hPrp5, SPF45 and CHERP) were identified in immunoaffinity purified U2 snRNP (Will *et al.*, 2002).

The 20S U5 snRNP contains eight U5 specific proteins with an apparent molecular weight of 15, 40, 52, 100, 102, 116, 200, and 220 kDa (Bach *et al.*, 1989). A couple of these proteins are involved in structural rearrangements, mostly in the first step of splicing (Staley and Guthrie, 1998). A stable RNA-free sub-complex of U5-220K, -200K, -116K, and -40K has been isolated (Achsel *et al.*, 1998) and protein interactions with the U5 snRNA could only be detected for the 220K protein (Urlaub *et al.*, 2000). This specific protein plays an important role during the splicing process, as it contacts the 5'ss of the pre-mRNA and is thought to align the 5' and 3' splice sites after the first step of splicing before the second step (Collins and Guthrie, 1999). The U5 snRNP is recruited to the spliceosome after tri-snRNP formation with U4/U6 and is remodeled during its activation, at which hPrp19/CDC5L and related proteins associate with the U5 snRNP to form the 35S U5 snRNP (Makarov *et al.*, 2002).

The 13S U4/U6 di-snRNP consists of two snRNAs (U4 and U6), Sm proteins and LSM proteins. It forms a stable heterodimer as a result of base pairing between U4 and U6 snRNAs. There are five specific proteins within the U4/U6 di-snRNP: the 15.5K, 20K (CypH), 60K, 61K, and 90K proteins.

Under physiological conditions the 13S U4/U6 di-snRNP and the 20S U5 snRNP form the 25S tri-snRNP (U4/U6.U5; Behrens and Luhrmann, 1991; Black and Pinto, 1989), which is integrated into the spliceosome during B complex formation (see below). The U4/U6.U5 involves all U5 and U4/U6 specific proteins except U5-52K, which dissociates during its formation (Laggerbauer *et al.*, 2005). It further contains three tri-snRNP specific proteins (27K, 65K, and 110K) which are required for integration into the spliceosome (Makarova *et al.*, 2001).

2.2.2.2 non-snRNP components

There are several non-snRNP protein components that bind to the spliceosome during its assembly pathway and play an important role during pre-mRNA splicing. One group of essential splicing factors are the SR proteins (reviewed by Manley and Tacke, 1996; Sanford *et al.*, 2003). SR proteins involve a various number of C-terminal SR dipeptides (serine/arginine-rich (SR) domain) and one or two RNA recognition motifs (RRMs) at the N-terminus (Birney *et al.*, 1993; Graveley, 2000). The RRM allows interaction with the pre-mRNA whereas the SR domain is responsible for protein-protein interactions. The SR proteins are thus able to function as bridge between the pre-mRNA and protein splicing factors.

Another group of proteins binding to the pre-mRNA are the heterogeneous ribonucleoproteins (hnRNPs; reviewed by Dreyfuss *et al.*, 1993). They are among the most abundant proteins in the nucleus and more than 20 major hnRNP proteins and several isoforms have been identified to date. The hnRNP proteins are designated alphabetically, starting from low molecular weight (A1 (34 kDa) to U (120 kDa)). All hnRNP proteins contain one or more RNA recognition motifs (RRM) and auxiliary domains that mediate protein-protein interactions. Although the precise function of hnRNP proteins is not clear, several hnRNP proteins have been shown to regulate splicing. The hnRNP C proteins have been shown to inhibit cleavage at the 5'ss (Choi *et al.*, 1986), hnRNP I/PTB has been shown to be essential for splicing (Patton *et al.*, 1991) and hnRNP A1 effects a switch from a 5'ss to an upstream splice site in pre-mRNAs that contain multiple splice sites (Mayeda and Krainer, 1992).

Table 2.2: Composition of the human U snRNPs. U1, U2, U4 and U5 snRNPs contain Sm proteins, whereas the U6 snRNP contains LSm proteins. All U snRNPs consist of additional snRNP specific proteins.

Proteins	Name	apparent MW [kDa]	12S U1	12S U2	17S U2	20S U5	13S U4/U6	25S U4/U6.U5
Sm proteins	SmB/B'	28/29	◆	◆	◆	◆	◆	◆◆
	SmD1	16	◆	◆	◆	◆	◆	◆◆
	SmD2	16.5	◆	◆	◆	◆	◆	◆◆
	SmD3	18	◆	◆	◆	◆	◆	◆◆
	SmE	12	◆	◆	◆	◆	◆	◆◆
	SmF	11	◆	◆	◆	◆	◆	◆◆
	SmG	9	◆	◆	◆	◆	◆	◆◆
LSm proteins	LSm2	10					◆	◆
	LSm3	15					◆	◆
	LSm4	15					◆	◆
	LSm5	10					◆	◆
	LSm6	8					◆	◆
	LSm7	13					◆	◆
	LSm8	13					◆	◆
U1 snRNP	70K	70	◆					
	A	34	◆					
	C	22	◆					
U2 snRNP	A'	31		◆	◆			
	B''	28.5		◆	◆			
	SF3a120	110			◆			
	SF3a66	66			◆			
	SF3a60	60			◆			
	SF3b155	160			◆			
	SF3b145	150			◆			
	SF3b130	120			◆			
	SF3b49	53			◆			
	SF3b14a/p14	15			◆			
	SF3b14b	15			◆			
SF3b10	9			◆				
U2-related	hPrp5	140			◆			
	SR140	140			◆			
	CHERP	130			◆			
	hPrp43	90			◆			
	SPF45	50			◆			
	SPF31	33			◆			
	SPF30	31			◆			
U5 snRNP	220K	220				◆		◆
	200K	200				◆		◆
	116K	116				◆		◆
	102K	120				◆		◆
	100K	100				◆		◆
	52K	52				◆		
	40K	40				◆		◆
15K	15				◆		◆	
U4/U6 snRNP	90K	90					◆	◆
	61K	61					◆	◆
	60K	60					◆	◆
	20K/CypH	20					◆	◆
U4/U6.U5 snRNP	15.5K	15.5					◆	◆
	110K	110						◆
	65K	65						◆
	27K	27						◆

The hPrp19/CDC5L complex is a non-snRNP protein complex, which binds to the spliceosome during its activation. It consists of seven proteins (CDC5L, Hsp70, CTNNBL1, PRL1, hPrp19, AD-002, and SPF27; Ajuh *et al.*, 2000; Makarova *et al.*, 2004) and associates with additional related proteins with the U5 snRNP to form the remodeled 35S U5 (Makarova *et al.*, 2004). It thus plays a crucial role in the assembly of a catalytically active spliceosome, presumably by stabilizing the RNA interaction network in the catalytic core (Ajuh *et al.*, 2000). The architecture and the protein stoichiometry of this particular protein complex in yeast and human has been recently analyzed (Grote *et al.*, 2010; Tarn *et al.*, 1994). All studies revealed that the hPrp19 protein is present as a tetramer within this particular protein complex.

Several splicing factors belong to the DExD/H-box protein family. These proteins are able to rearrange RNP and RNA-RNA interactions and are therefore required for structural rearrangements during the spliceosomal assembly. Some of them are U snRNP specific proteins (see Table 2.2), whereas others are non-snRNP specific. There are eight DExD/H box proteins that are important for pre-mRNA splicing: hPrp5, U5-200K, U5-100K, UAP56, hPrp2, hPrp16, hPrp22, and hPrp43 (Schwer, 2001). However, the precise function has only been established for few of these proteins.

Numerous additional proteins have been detected, predominantly by mass spectrometry, within the different spliceosomal transition states (Behzadnia *et al.*, 2007; Bessonov *et al.*, 2008; Deckert *et al.*, 2006; Hartmuth *et al.*, 2002). From these, Prp16, Prp17, Prp18, Prp22, and Slu7 (reviewed by Umen and Guthrie, 1995) have been reported to bind the spliceosome after the first step of splicing to function at the second step of splicing (Figure 2.9). Additionally, several factors specific for the catalytically active spliceosome (see below) have been identified (Bessonov *et al.*, 2008).

2.2.3 The stepwise assembly of the spliceosome

Pre-mRNA splicing is a highly dynamic process, which involves the assembly of the U snRNPs and additional non-snRNP splicing factors, the dissociation of these factors upon structural changes, the completion of the two transesterification steps, and finally the release of the generated mRNA. The spliceosome assembles on the pre-mRNA in a stepwise manner passing through different functional intermediate states (Figure 2.10).

In the first assembly step, the U1 snRNP binds to the 5'ss of the pre-mRNA by base pairing of U1 snRNA and the pre-mRNA. Formation of the thus formed E complex (the early complex) is mediated by interactions of U1-C and the 5'ss (Heinrichs *et al.*, 1990; Will *et al.*, 1996). Recruitment of U2 snRNP leads then to formation of the A complex (the pre-

spliceosome). At this stage, the U2 snRNA base pairs with the pre-mRNA in the branch point site region bulging out the branch point adenosine, which is thus enabled to carry out the nucleophilic attack of the first step of splicing (Query *et al.*, 1994). Binding of U2 is promoted by splicing factors U2AF and SF3a and SF3b (Gozani *et al.*, 1996; Valcarcel *et al.*, 1996). Upon integration of the U4/U6.U5 tri-snRNP, which is mediated by tri-snRNP specific proteins 110K and 65K (Makarova *et al.*, 2001), the pre-catalytic spliceosome (B complex) is developed. Structural RNA and protein rearrangements within the B complex induce the dissociation of U1 and U4 snRNPs. Dissociation of U1 and U4, as well as remodeling of U5 by the hPrp19/CDC5L complex (see above), generate the activated spliceosome (B*), in which the first catalytic step of splicing occurs. The complex that forms during this process is the catalytically active C complex, which performs the second step of splicing. The final step is the release of the generated mRNA, dissociation of the post-spliceosomal intron complex, and the reconstitution of the splicing factors.

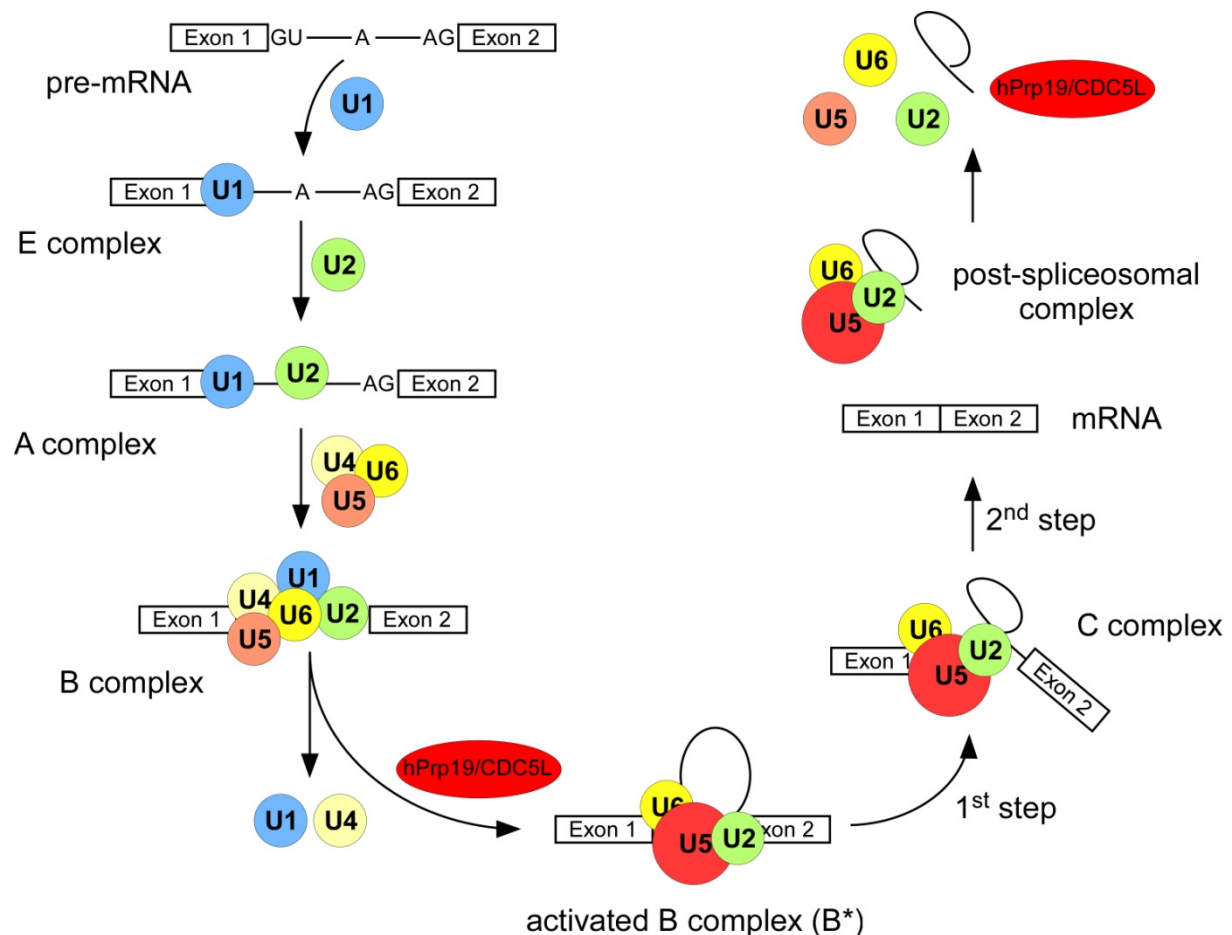


Figure 2.10: The stepwise assembly of the spliceosome. In the first step, U1 snRNP binds to the 5'ss forming the E complex followed by binding of U2 snRNP (A complex formation). Recruitment of the preassembled tri-snRNP (U4/U6.U5) leads to formation of the pre-catalytic B complex. Upon structural rearrangements U1 and U4 snRNPs dissociate and incorporation of the hPrp19/CDC5L complex leads to remodeling of U5 generating the activated B complex. The first step of splicing occurs in this intermediate assembly yielding the C complex, in which the second step of splicing is carried out. The generated mRNA and the post-spliceosomal complex are released and the splicing factors are reconstituted.

Different strategies for affinity purification of the functional intermediates during the spliceosomal assembly have been developed: (i) affinity purification of spliceosomal intermediates using peptide antibodies specific for particular proteins, (ii) isolation of spliceosomal complexes using tobramycin-tagged pre-mRNA, and (iii) affinity selection of spliceosomal complexes using MS2-tagged pre-mRNA and a MS2-MBP fusion protein. From these, the latter has proven to be most suitable. The method has been used to purify A complexes (Behzadnia *et al.*, 2007), B complexes (Deckert *et al.*, 2006), C complexes (Bessonov *et al.*, 2008) and activated B complexes (B^{*}) from yeast (Fabrizio *et al.*, 2009). Mass spectrometric analyses of the purified complexes revealed a complex and highly dynamic proteome of the spliceosome in its different functional states.

In addition to the stepwise assembly model of the spliceosome, a second model comprising the recruitment of a pre-assembled spliceosome followed by structural rearrangements on the pre-mRNA, exists. Strongest evidence of a pre-assembled spliceosome gave the isolation of a 45S RNP complex in yeast consisting of all spliceosomal snRNPs, which is functional in extracts (Stevens *et al.*, 2002). Furthermore, a 200S RNP consisting of all spliceosomal snRNPs was isolated from HeLa nuclear extracts and was shown to assemble on RNA containing a 5'ss (Malca *et al.*, 2003). The two models are not necessarily contradictory. As the spliceosomal assembly intermediates (A, B, B^{*} and C complexes) are purified under relatively stringent conditions, they reflect distinct stabilization states of the spliceosomal snRNPs (Will and Luhrmann, 2005). The stepwise assembly model therefore represents a proper model showing at which time during splicing the different U snRNPs are stably associated with the pre-mRNA.

2.3 Aim of this study

The spliceosome assembles from different uridine-rich small nuclear ribonucleoprotein particles (U snRNPs) and additional non-snRNP proteins. The RNA and protein composition of the involved U snRNPs has been well characterized in the recent past. Therefore, the identification of the non-snRNP splicing factors in the splicing intermediates is of great interest. In recent studies, mass spectrometry allowed identification of several proteins within different spliceosomal complexes. Changes in the protein abundance between these assembly states will provide an insight into the functionality of the different protein components. As protein abundances cannot be determined by simple mass spectrometric analyses, quantification of the complexes' proteins is required to detect differences within the different spliceosomal intermediates.

In the course of investigating different intermediates of *in vitro* spliceosomal assembly, problems have arisen that require quantification of proteins within different spliceosomal complexes. Therefore, the aim of this work was the establishment of robust mass spectrometry-based quantification methods for the quantitative analysis of proteins in the different spliceosomal complexes to tackle the following problems: (i) Determination of the protein stoichiometry within the spliceosomal hPrp19/CDC5L complex, which is recruited to the spliceosome during its activation, by absolute quantification of the proteins within this particular complex. (ii) Relative comparison of the pre-catalytic and the catalytically active spliceosomal complexes (i.e. B and C complexes) to show differences in their protein composition and in this manner to determine proteins, which are necessary for splicing activity. (iii) Monitoring the dynamic protein changes of spliceosomal proteins during the assembly on the pre-mRNA by relative quantification.

To achieve the aim, different mass spectrometry-based quantification strategies, namely AQUA, iTRAQ and SILAC, were to be set up. At the same time, sample preparation and data analysis had to be implemented. For absolute quantification, the complete hydrolysis of proteins needed to be achieved and the suitability of synthetic standard peptides and different methods to read out the quantitative signals were to be tested to achieve most accurate results for reliable quantification. For relative comparison of different protein complexes, a workflow comprising iTRAQ labeling of in-gel digested proteins had to be established. This method was then set out to be compared to other quantification techniques, such as SILAC or spectral count, which is a simple and often applied but only semi-quantitative quantification method. In view of protein changes during the spliceosomal assembly pathway, metabolic labeling (SILAC) was to be applied to display the protein dynamics of the spliceosome during pre-mRNA splicing. An experimental setup allowing the construction of assembly time lines needed to be implemented. To this end, different pre-mRNAs lacking the 5'ss and the BPS, respectively, were to be generated and tested for their splicing behavior. These were then used to compare the protein assembly on different pre-mRNAs to understand the dynamic process of pre-mRNA splicing.

3 Materials and methods

3.1 Materials

3.1.1 Chemicals

Acetic acid	Merck, Darmstadt
Acetonitrile, LiChrosolv	Merck, Darmstadt
Agarose (low melting point)	Invitrogen, USA
Agarose, Ultra Pure	Invitrogen, USA
Ammonium bicarbonate	Fluka, Switzerland
Ammonium peroxodisulfate (APS)	Merck, Darmstadt
Ammonium sulfate	Merck, Darmstadt
Boric acid	Merck, Darmstadt
Bovine Serum Albumin (BSA)	New England Biolabs, USA
Bromophenol blue (sodium salt)	Merck, Darmstadt
Calcium chloride dihydrate	Merck, Darmstadt
Chloroform	Merck, Darmstadt
Cleland's Reagent (DTT, for MS analyses)	Calbiochem, USA
Coomassie Brilliant Blue G-250	Fluka, Switzerland
Creatine phosphate	Sigma-Aldrich, Steinheim
α -Cyano-4-hydroxycinnamic acid	Sigma-Aldrich, Steinheim
Dimethylsulfoxide (DMSO)	Roth, Karlsruhe
di-Potassium hydrogen phosphate trihydrate	Merck, Darmstadt
Dithioerythrol (DTE)	Roth, Karlsruhe
Dithiothreitol (DTT)	Roth, Karlsruhe
Dodecyl sulfate sodium salt (SDS)	Merck, Darmstadt
Ethanol	Merck, Darmstadt
Ethidium bromide solution 10 mg/ml	Roth, Karlsruhe
Ethylenediaminetetraacetic acid (EDTA, disodium salt)	Roth, Karlsruhe
Formaldehyde	Merck, Darmstadt
Formamide	Merck, Darmstadt
Formic acid	Fluka, Switzerland
Glycerol	Merck, Darmstadt
Glycine	Merck, Darmstadt
Glycogen	Merck, Darmstadt
Heparin	Sigma-Aldrich, Steinheim
4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid (HEPES)	Merck, Darmstadt
Iodoacetamide	Sigma-Aldrich, Steinheim
Magnesium chloride	Sigma-Aldrich, Steinheim
Maltose	Merck, Darmstadt
Methanol	Merck, Darmstadt
Methanol, LiChrosolv	Merck, Darmstadt
N,N,N',N'-Tetramethylethylenediamid (TEMED)	Sigma-Aldrich, Steinheim
N,N-Dimethylformamide	Merck, Darmstadt
ortho-Phosphoric acid	Merck, Darmstadt
Phenol/Chloroform/Isoamylalcohol (PCI)	Roth, Karlsruhe
Phenylmethylsulfonyl fluoride (PMSF)	Roche, Mannheim
Potassium chloride	Merck, Darmstadt
Potassium dihydrogen phosphate	Merck, Darmstadt
RapiGest	Waters, UK
Rotiphorese Gel 40 (38% Acrylamide, 2% Bis-Acrylamide)	Roth, Karlsruhe

Silver nitrate	Merck, Darmstadt
Sodium acetate	Merck, Darmstadt
Sodium carbonate	Merck, Darmstadt
Sodium chloride	Merck, Darmstadt
Sodium dihydrogen phosphate monohydrate	Merck, Darmstadt
Trifluoric acid	Fluka, Switzerland
Tris-(hydroxymethyl) aminoethane (Tris)	Roth, Karlsruhe
Urea	Merck, Darmstadt
Water, LiChrosolv	Merck, Darmstadt
Xylene Cyanol FF	Fluka, Switzerland

3.1.2 Enzymes and enzyme inhibitors

BamH I	New England Biolabs, USA
LysC	Roche, Mannheim
Proteinase Inhibitor Cocktail Complete, EDTA free	Roche, Mannheim
Proteinase K	Sigma-Aldrich, Steinheim
RNasin (40U/μl)	Promega, Mannheim
RQ DNase I	Promega, Mannheim
SP6 polymerase	Promega, Mannheim
T4 polynucleotide kinase	New England Biolabs, USA
T4 DNA ligase	New England Biolabs, USA
Trypsin	Roche, Mannheim
Trypsin	Promega, Mannheim
Turbo Pfu DNA polymerase	Stratagene, Heidelberg

3.1.3 Nucleotides

adenosine 5'-triphosphate (rATP, 100 mM)	Promega, Mannheim
cytidine 5'-triphosphate (rCTP, 100 mM)	Promega, Mannheim
guanosine 5'-triphosphate (rGTP, 100 mM)	Promega, Mannheim
uridine 5'-triphosphate (rUTP, 100 mM)	Promega, Mannheim
Easy Tides Uridine 5'-triphosphate, [α - ³² P]	Perkin Elmer, USA
2'-deoxyadenosine-5'-triphosphate (dATP, 100 mM)	New England Biolabs, USA
2'-deoxycytidine-5'-triphosphate (dCTP, 100 mM)	New England Biolabs, USA
2'-deoxyguanosine-5'-triphosphate (dGTP, 100 mM)	New England Biolabs, USA
2'-deoxythymidine-5'-triphosphate (dTTP, 100 mM)	New England Biolabs, USA
m7G(5')ppp(5')G cap	Kedar, Poland

3.1.4 DNA oligonucleotides

Table 3.1: Oligonucleotides used in this study

Name	Sequence (5' → 3')	Description
M6	GGCGGTCTCGTC	RNAseH digestion
M12	CTCGTCGGCAGC	RNAseH digestion
PM5-5'ss-del_for	ATCAAGCTTACAAGACAGCTTT	deletion of the 5'ss within the PM5 plasmid
PM5-5'ss-del_rev	TGGCGGCGGTCTCGTCG	deletion of the 5'ss within the PM5 plasmid
PM5-BPS-del_for	AAATACATAAGAATCAGGTAGTG	deletion of the BPS within the PM5 plasmid

PM5-BPS-del_rev	CTGTGTTTTTTTTGCTACTTTTTTTT	deletion of the BPS within the PM5 plasmid
PM5-BPS_ACTGA-del_for	TAGAACACTACCTGATTCTTATG	deletion of duplicate BPS sequence within the PM5 plasmid
PM5-BPS_ACTGA-del_for	CACACTCCACACACATTCCA	deletion of duplicate BPS sequence within the PM5 plasmid

M6 and M12 oligos were purchased from IBA (Göttingen) and all other oligonucleotides were purchased from Eurofins MWG Operon (Ebersberg).

3.1.5 Plasmids

PM5
pMS2-MBP

Anderson and Moore, 1997
Das *et al.*, 2000

3.1.6 Bacteria strains

The following *Escherichia coli* strains were used in this study:

BL21-Rosetta
DH5 α

3.1.7 Cell lines

HeLa S3 cells

3.1.8 Cell culture

Ampicillin
Chloramphenicol
DMEM High Glucose w/o Arginine, w/o Lysine
Foetal Bovine Serum, dialyzed
IPTG (Isopropyl-b-D-1-thiogalactopyranoside)
L-Arginine
L-Arginine, $^{13}\text{C}_6$
L-Arginine, $^{13}\text{C}_6$, $^{15}\text{N}_4$
LB Medium
LB-Agar
L-Lysine
L-Lysine, $^{13}\text{C}_6$
L-Lysine, $^{13}\text{C}_6$, $^{15}\text{N}_2$
L-Lysine, $^2\text{D}_4$
100 \times Penicillin/Streptomycin Concentrate

Sigma-Aldrich, Steinheim
Roth, Karlsruhe
PAA Laboratories, Cölbe
PAA Laboratories, Cölbe
Roth, Karlsruhe
Sigma-Aldrich, Steinheim
Eurisotop, Saarbrücken
Eurisotop, Saarbrücken
MP Biomedicals, USA
MP Biomedicals, USA
Sigma-Aldrich, Steinheim
Eurisotop, Saarbrücken
Eurisotop, Saarbrücken
Eurisotop, Saarbrücken
PAA Laboratories, Cölbe

3.1.9 Commercial reaction kits and buffers

Bradford solution, Bio-Rad Protein Assay Buffer 4	Bio-Rad, München
10 × cloned Pfu DNA Polymerase Buffer	New England Biolabs, USA
Invitrogen PureLink HiPure Plasmid Filter Maxiprep kit	Stratagene, Heidelberg
iTRAQ reagent Multi-Plex Kit	Invitrogen, USA
NucleoSpin Extract II kit	ABSciex, Darmstadt
NuPAGE Antioxidant	Machery-Nagel
NuPAGE LDS Sample Buffer (4 ×)	Invitrogen, USA
NuPAGE MOPS SDS Running Buffer (10 ×)	Invitrogen, USA
NuPAGE Sample Reducing Agent (10 ×)	Invitrogen, USA
10 PNK Buffer	New England Biolabs, USA
QIAprep Spin Miniprep kit	Qiagen, Hilden
T4 DNA Ligase Buffer	New England Biolabs, USA
Transcription Optimized 5 × Buffer	Promega, Mannheim
Triethylammonium bicarbonate buffer, 1M, pH 8.5	Sigma-Aldrich, Steinheim

3.1.10 Commonly used buffers10 × D⁻ Buffer

20 mM HEPES pH 7.9
1.5 mM MgCl₂
0.2 mM EDTA pH 8.0

Na-P buffer

5 mM Na₂HPO₄

1 × PBS

130 mM NaCl
20 mM K-PO₄

2 × PK-Buffer

200 mM Tris pH 7.5
25 mM EDTA pH 8.0
2% (m/v) SDS

Roeder C buffer

25% (v/v) Glycerol
20 mM HEPES
420 mM NaCl
1.5 mM MgCl₂
0.2 mM EDTA

Roeder D buffer

10% (v/v) Glycerol
20 mM HEPES
100 mM KCl
1.5 mM MgCl₂
0.2 mM EDTA
0.5 mM DTT
0.5 mM PMSF

1 × TBE, pH 8.3

0.1 M Boric acid
0.1 M Tris
0.1 M EDTA

3.1.11 Standard peptides

[Glu ¹]-Fibrinopeptide B	Sigma-Aldrich, Steinheim
EAAAA-LVEEET(¹³ C ₆ ¹⁵ N ₄ -R)	Sigma-Genosys, USA
FVDILG(¹³ C ₆ ¹⁵ N-L)R	Thermo Fisher Scientific, Ulm
HYTFASGSPDN(¹³ C ₆ ¹⁵ N-I)K	Thermo Fisher Scientific, Ulm
ILLGGYQS(¹³ C ₆ ¹⁵ N ₄ -R)	Sigma-Genosys, USA
LGLLGLPAP(¹³ C ₆ ¹⁵ N ₂ -K)	Sigma-Genosys, USA
NVVV(¹³ C ₉ ¹⁵ N-F)DK	Thermo Fisher Scientific, Ulm
TGYN(¹³ C ₉ ¹⁵ N-F)QR	Thermo Fisher Scientific, Ulm
TIVQLENEIYQ-(¹³ C ₆ ¹⁵ N-I)K	Thermo Fisher Scientific, Ulm
TLQLDNNFEV(¹³ C ₆ ¹⁵ N ₂ -K)	Sigma-Genosys, USA
TVPEELVKPEELS(¹³ C ₆ ¹⁵ N ₂ -K)	Sigma-Genosys, USA
YADLL(¹³ C ₆ ¹⁵ N-L)EK	Thermo Fisher Scientific, Ulm

3.1.12 Chromatography materials and consumables

Amylose Resin	New England Biolabs, USA
anti-FLAG-M2-Agarose beads	Sigma-Aldrich, Steinheim
Bio-Spin Disposable Chromatography Columns	Bio-Rad, München
FLAG peptide	Sigma-Aldrich, Steinheim
Heparin Sepharose 6Fast Flow	GE Healthcare, UK
m-Precolumn™ Cartridge, Acclaim PepMap100 C18, 300µm i.d. × 5 mm	LC Packings, Netherlands
Nucleosil 100-5 C18, 5 µm	Machery-Nagel, Düren
NuPAGE Novex Bis-Tris Mini Gels	Invitrogen, USA
PepMap C18, 300 µm, 5 µm	Dionex, Idstein
PepMap C18, 75 µm, 15 cm	Dionex, Idstein
Reprosil-Pur Basic C18-HD, 5 µm	Maisch, Ammerbuch

3.1.13 Instruments, Equipment

Blank Opti-TOF, 123 mm × 81 mm, MALDI target	ABSciex, Darmstadt
CAP-LC system	Waters, UK
HP1100 series chromatography system	Agilent, USA
LTQ-Orbitrap XL	Thermo Scientific, Dreieich
4800 MALDI TOF/TOF Analyzer	ABSciex, Darmstadt
Phosphorimager Typhoon 8600	Molecular Dynamics, Switzerland
Probot, robotic spotting device	Dionex, Idstein
Q-ToF Ultima mass spectrometer	Waters, UK
4000 QTRAP mass spectrometer	ABSciex, Darmstadt
Scintillation counter LS1701/TRI-CARB 2100TR	Beckmann/Packard, USA
Sorvall SA-600 rotor	Kendro, USA
Sorvall SS-34 rotor	Kendro, USA
Tempo 1D chromatography system	ABSciex, Darmstadt
Ultimate chromatography system	Dionex, Idstein
Ultracentrifuges	Sorvall/Beckmann
XCell Sure Lock Mini Cell	Invitrogen, USA

3.2 Methods

3.2.1 Molecular biology methods

3.2.1.1 Concentration determination of nucleic acids

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

$$1 \text{ OD}_{260} = 50 \text{ } \mu\text{g/ml double stranded DNA} = 0.15 \text{ mM (in nucleotides)}$$

$$1 \text{ OD}_{260} = 33 \text{ } \mu\text{g/ml single stranded DNA} = 0.10 \text{ mM (in nucleotides)}$$

$$1 \text{ OD}_{260} = 40 \text{ } \mu\text{g/ml single stranded RNA} = 0.11 \text{ mM (in nucleotides)}$$

The concentration of α - ^{32}P UTP-labeled RNA was determined by the ratio of α - ^{32}P UTP and ^{31}P -UTP, the isotope concentration of α - ^{32}P UTP, the absolute number of uridines in the transcript and an instrument constant using the following equations:

$$\text{mixing ratio} = \frac{[^{31}\text{P} - \text{UTP}]}{[\alpha - ^{32}\text{P}]\text{UTP}}$$

$$\frac{\text{isotope concentration} \left(\frac{\text{Ci}}{\text{mmol}} \right)}{\text{mixing ratio}} \times \# \text{ Uridines} = \text{radioactivity} \left(\frac{\text{Ci}}{\text{mmol}} \right)$$

$$\text{specific activity} \left(\frac{\text{dpm}}{\text{mmol}} \right) = \text{radioactivity} \left(\frac{\text{Ci}}{\text{mmol}} \right) \times 2.2 \cdot 10^6 \left(\frac{\text{dpm}}{\text{Ci}} \right)$$

$$\text{specific activity} \left(\frac{\text{cpm}}{\text{mmol}} \right) = \text{specific activity} \left(\frac{\text{dpm}}{\text{mmol}} \right) \div 2$$

3.2.1.2 Phenol-Chloroform-Isoamylalcohol (PCI) extraction

Phenol-chloroform-isoamylalcohol (PCI) extraction was used to purify and separate proteins and nucleic acids. The sample was mixed with 1 vol. of PCI and 1 μl of 10 $\mu\text{g}/\mu\text{l}$ glycogen and vigorously agitated on a vortex. Aqueous and organic phases were separated by centrifugation for 5 min at room temperature (13,000 rpm). The aqueous phase, containing nucleic acids, was transferred to a new tube. Proteins (organic phase) were precipitated with 5 vol. 100 % (v/v) acetone at $-20 \text{ }^\circ\text{C}$ for at least 2 h. Nucleic acids (aqueous phase) were further purified by addition of 1 vol. of chloroform. After mixing and centrifugation as above,

the nucleic acids were precipitated from the aqueous phase in 3 vol. of 100 % (v/v) ethanol and 1/10 vol. of 0.3 M sodium acetate pH 5.3 for at least 2 h at -20 °C. Precipitated proteins or nucleic acids were collected by centrifugation for 30 min at 4 °C (13,000 rpm). Proteins or nucleic acids were washed with 80 % (v/v) ethanol, collected by centrifugation (see above) and dried in a vacuum centrifuge.

Phenol-Chloroform-Isoamylalcohol (PCI):

50 % (v/v) Phenol
48 % (v/v) Chloroform
2 % (v/v) Isoamylalcohol

3.2.1.3 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was used to analyze and purify DNA. Gels were prepared with 1 % (w/v) agarose and 0.5 µg/ml ethidium bromide to visualize DNA by UV light. DNA samples were diluted with 5 × DNA loading dye and allowed to migrate horizontally at 120 V using 1 × TBE as running buffer until the dye migrated through the gel. DNA was visualized under a UV light at a wavelength of 365 nm.

Gel solution:

1 % (w/v) Agarose
1 × TBE
0.5 µg/ml Ethidium bromide

DNA loading dye:

30 % (w/v) Glycerol
0.25 % (w/v) Bromophenol blue

3.2.1.4 In vitro transcription

Pre-mRNAs were synthesized by *in vitro* transcription using SP6 polymerase and linearized plasmid DNA as template. The reaction was incubated at 40 °C for 3-4 h. Template DNA was subsequently digested with 1U RQ1 DNase/µg template DNA at 37 °C for 20 min. RNA transcripts were purified by gel purification using 5 % polyacrylamide gels containing 8 M urea. Unlabeled RNA was visualized by UV-shadowing (254 nm) and α-[³²P]UTP-labeled RNA was visualized by exposure to an X-ray film (1 min). Bands were excised from the gel and extracted by incubation with RNA extraction buffer overnight. Extracted RNA was further purified by PCI extraction and ethanol precipitation. The purified RNA was resuspended with RNase-free water.

Preparation of capped PM5-pre-mRNA:

	α -[³² P]UTP -labeled	non-labeled
5 × transcription buffer	10.00 μ l	30.00 μ l
0.1 M ATP	3.75 μ l	11.25 μ l
0.1 M UTP	3.75 μ l	11.25 μ l
0.1 M CTP	3.75 μ l	11.25 μ l
0.01 M GTP	6.50 μ l	19.50 μ l
152 mM m ⁷ GpppG cap	1.64 μ l	4.92 μ l
³² P- α UTP (3000 Ci/mmol)	5.00 μ l	-
1M MgCl ₂	1.05 μ l	3.15 μ l
1M DTT	0.50 μ l	1.50 μ l
10 mg/ml BSA	0.50 μ l	1.50 μ l
40 U/ μ l RNAsin	1.25 μ l	3.75 μ l
SP6 polymerase	5.33 μ l	16.00 μ l
DNA template	5.00 μ g	15.00 μ g
The volume was adjusted to (RNase-free water)	50 μl	150 μl

8 M urea-5 % Polyacrylamide Gel Solution (100 ml):

12.5 ml Rotiphorese Gel 40
 42 g Urea
 10 ml 10 × TBE
 Adjust to 100 ml (RNase free water)

Per 100 ml gel solution:
 10 μ l TEMED
 100 μ l 10 % (w/v) APS

3.2.1.5 DNA amplification

100 μ l chemical competent dH5 α cells were transfected with 100 ng of plasmid DNA by the heat shock method according to standard protocols (Sambrook *et al.*, 1989). To this end, competent bacteria were thawed on ice and mixed with plasmid DNA. After incubation on ice for 30 min, cells were heat shocked at 42°C for 90 s and subsequently cooled on ice for 2 min. 1 ml of LB medium was added followed by incubation at 37 °C for 1 h. To select transformed bacteria, cells were plated on agar plates containing the appropriate antibiotic and incubated at 37 °C overnight. Single clones were selected and grown at 37 °C in LB medium. DNA was recovered using QIAprep Spin Miniprep kit or Invitrogen PureLink HiPure Plasmid Filter Maxiprep kit.

3.2.1.6 Restriction digestion of DNA

Restriction digestion was used to generate desired ends of template DNA for *in vitro* transcription from plasmid DNA. 200 µg of plasmid DNA, 100 U of *Bam*HI, 140 µl of 10 × Buffer 4 were adjusted to a final volume of 1400 µl with RNase-free water. The mixture was incubated at 37 °C overnight and linearized DNA was recovered by PCI extraction. Linearization was assayed by agarose gel electrophoresis.

3.2.1.7 Proteinase K digestion

Proteinase K digestion was performed to improve RNA recovery from splicing reactions. 125 µl of proteinase K mix were added to one 20 µl aliquot of splice reaction followed by incubation for 1h at 37 °C. The volume was adjusted to a final volume of 200 µl with 35 µl 1 × D⁻ buffer and 20 µl of 10 % (v/v) SDS. The RNA was recovered by PCI extraction and ethanol precipitation.

Proteinase K mix:

62.5 µl 2 × PK buffer
1 µl 10 mg/ml Glycogen
57.5 µl RNase-free water
4 µl 10 mg/ml Proteinase K

3.2.1.8 Denaturing polyacrylamide gel electrophoresis

Denaturing gel electrophoresis was used to analyze snRNA and pre-mRNA splicing products. RNA samples were dissolved in RNA sample buffer, heated for 3-5 min at 96 °C and loaded onto 0.5 mm thick 10 % polyacrylamide gels (37.5:1 acrylamide to bis-acrylamide ratio) containing 8 M urea. Electrophoresis was performed vertically at 20 W using 1 × TBE as running buffer.

8 M urea-10 % Polyacrylamide Gel Solution (100 ml):

25.0 ml Rotiphorese Gel 40
42 g Urea
10 ml 10 × TBE
Adjust to 100 ml (RNase free water)
Per 100 ml gel solution:
10 µl TEMED
100 µl 10 % (w/v) APS

RNA Sample Buffer:

80 % (w/v) Formamide
1 mM EDTA
0.05 % (w/v) Bromophenol blue
0.05 % (w/v) Xylene cyanol

3.2.1.9 Silver staining of RNA

RNA was visualized by silver staining according to Schoenle *et al.*, 1984. Briefly, RNA was fixed in 40 % (v/v) methanol/10 % (v/v) acetic acid for at least 30 min and washed twice with 10 % (v/v) ethanol/5 % (v/v) acetic acid for 15 min. The gel was briefly rinsed with water and incubated with 12 mM silver nitrate for 30 min. After 3 short rinsings with water, the gel was incubated with 280 mM sodium carbonate/0.0185 % (v/v) formaldehyde until the desired staining intensity was reached. Staining was then stopped immediately with 5 % (v/v) acetic acid.

3.2.1.10 Native gel electrophoresis of RNA complexes

Splicing complex formation was analyzed by native gel electrophoresis. 0.7 µl Heparin (5 mg/ml, final concentration 170 mg/ml) was added to 20 µl aliquots of splicing complexes and heated for 1 min at 30 °C. Samples were diluted with 5 × native loading dye and loaded onto 1.5 % agarose gels. Electrophoresis was performed horizontally at 40 V for 15 h using 0.5 % TBE as running buffer. The gels were fixed with 10 % (v/v) acetic acid/10 % (v/v) methanol for 30 min and dried under vacuum at 60 °C for 4.5 h. Radioactively labeled RNA was visualized by exposing the dried gels to phosphorimager screens for approx. 1 hr and scanning using Typhoon PhosphorImager.

Agarose Gel Solution:

1.5 % (v/v) Low Melting Point Agarose
0.5 × TBE

Native loading dye:

1 × TBE
30 % Glycerol
0.02 % Bromophenol blue

3.2.1.11 Generation of pre-mRNA mutants (deletion of 5'ss and BPS)

5'splice site (5'ss)-, branch point site (BPS)- and BPS-ACTGA-deleted PM5 pre-mRNA was generated as described before (Dönmez, 2006; see also Figure 3.1 for an overview). PCR primers were used to exclude the 5'SS or BPS region within the PM5 plasmid. The PCR product was purified by agarose gel electrophoresis and recovered using the NucleoSpin Extract II kit (Machery-Nagel). The linear plasmid was kinased with T4 polynucleotide kinase (T4 PNK) in the presence of 1mM ATP. The reaction was incubated for 1 h at 37 °C and the kinased plasmid was purified using the NucleoSpin Extract II kit. The plasmid was then ligated with T4 DNA ligase for 30 min at 37 °C followed by incubation at 16 °C overnight. The ligase was deactivated by incubation for 10 min at 65 °C. The ligated plasmid was purified by ethanol precipitation, dissolved in water and used for transformation of dH5α cells. Amplified

DNA was isolated using the QIAprep Spin Miniprep kit and cloned plasmids were sequenced (Seqlab Sequence Laboratories Göttingen GmbH).

PCR reaction: (100 μ l)

0.5 μ l PM5 plasmid (20 μ g/ml)
 1 μ l Turbo Pfu
 1.6 μ l dNTPs (25 mM each)
 1 μ l forward primer (100 μ M)
 1 μ l reverse primer (100 μ M)
 10 μ l cloned Pfu buffer
 7.5 μ l DMSO
 77.4 μ l H₂O

PCR program:

		5'ss deletion	BPS / BPS-ACTGA deletion
step 1	90 s	95 °C	95 °C
step 2 (32	30 s	95 °C	95 °C
	45 s	53 °C	50 °C
	7 min	68 °C	68 °C
step 3	hold	4 °C	4 °C

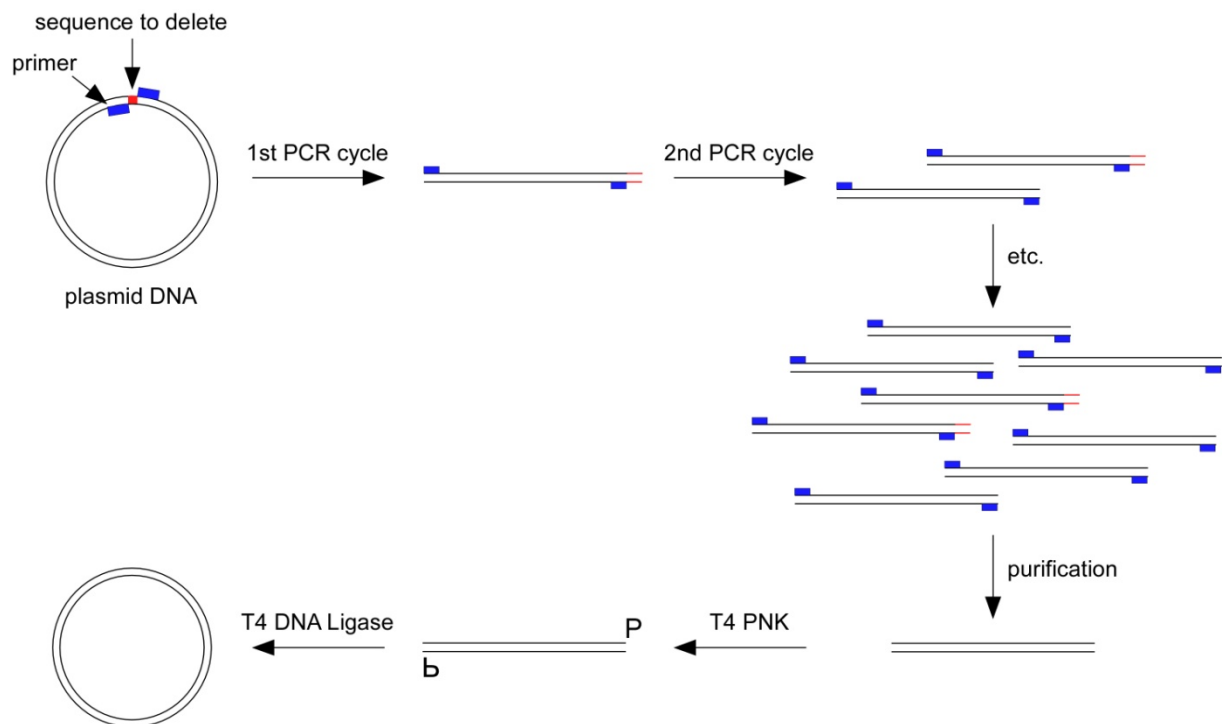


Figure 3.1: Generation of 5'ss, BPS and BPS-ACTGA deleted PM5 pre-mRNA. PCR primers were used to exclude the sequence to be deleted. After purification of the linear PCR product, the DNA was kinased using T4 PNK and subsequently ligated with T4 DNA Ligase. The obtained plasmid DNA does not include the deleted nucleotide sequence.

3.2.2 Protein biochemical methods

3.2.2.1 Protein concentration determination

The concentration of proteins in solution was determined by using the Bradford Protein Assay (Bearden, 1978). The protein sample was diluted with water to a final volume of 800 μ l

followed by addition of 200 μ l of Bradford solution (Bio-Rad Protein Assay) and incubation at room temperature for 10 min. The absorbance at 595 nm was then measured and the protein concentration was determined by comparison to a standard dilution series of BSA (0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0 μ g).

3.2.2.2 Polyacrylamide gel electrophoresis (PAGE)

Protein samples were separated using the NuPAGE[®] system (Invitrogen) according to manufacturer's protocols. Briefly, precipitated and dried protein samples were dissolved in 4 \times sample buffer, 10 \times reducing agent and water. Dissolved samples were loaded onto 4-12 % Bis-Tris pre-cast gels (1.0 mm \times 10 well) and gel electrophoresis was performed for approx. 50 min at 200 V using MOPS running buffer and antioxidant.

Buffers and reagents:

- NuPAGE LDS Sample Buffer (4 \times)
- NuPAGE Sample Reducing Agent (10 \times)
- NuPAGE MOPS SDS Running Buffer (20 \times)
- NuPAGE Antioxidant

3.2.2.3 Colloidal Coomassie staining of proteins

Separated proteins were stained using colloidal coomassie solution (Neuhoff *et al.*, 1988). To this end, gels were incubated with staining solution overnight and background staining was removed by rinsing the gels with water.

Colloidal Coomassie staining solution:

0.08 % (m/v) Coomassie Brilliant Blue G 250
20 % (v/v) Methanol
1.6 % (v/v) ortho-Phosphoric acid
8 % (m/v) Ammonium sulfate

3.2.3 Cell culture, nuclear extract, purification of spliceosomal complexes

3.2.3.1 Metabolic Labeling of HeLa S3 cells (SILAC labeling)

HeLa S3 cells were grown in custom made DMEM medium containing normal (light) or stable isotope labeled (medium and heavy) L-arginine and L-lysine (see table 3.2). Cells were grown for at least six passages at smaller volumes and then expanded to 2.0 l in spinner

flasks ($0.5\text{-}1.0 \times 10^6$ cells/ml). The cells were transferred to a 2.5 l fermenter and grown under standard conditions ($2.5\text{-}5.0 \times 10^6$ cells/ml).

SILAC medium:

500 ml DMEM w/o Arginine, w/o Lysine
 50 ml dialyzed FBS
 5ml $100 \times$ Penicillin/Streptomycin
 50 mg/l L-Arginine
 50 mg/l L-Lysine

Table 3.2: Isotope composition of amino acids used for different SILAC media. Due to the number of incorporated ^{13}C , ^{15}N or ^2D atoms a mass difference between the light amino acid and the stable isotope labeled amino acid is obtained.

		L-Arginine	Δm	L-Lysine	Δm
duplex SILAC	light	-	0	-	0
	heavy	$^{13}\text{C}_6, ^{15}\text{N}_4$	+10 Da	$^{13}\text{C}_6$	+6 Da
triple SILAC	light	-	0	-	0
	medium	$^{13}\text{C}_6$	+6 Da	$^2\text{D}_4$	+4 Da
	heavy	$^{13}\text{C}_6, ^{15}\text{N}_4$	+10 Da	$^{13}\text{C}_6, ^{15}\text{N}_2$	+8 Da

3.2.3.2 Preparation of splicing-active HeLa cell nuclear extract

Cells from a fermenter were used to prepare nuclear extract according to Dignam *et al.*, 1983. Briefly, cells were harvested by centrifugation for 5 min at 2000 rpm (Cryofuge 6000i, Heraeus) and washed with ice-cold PBS. The cells were resuspended in 1.25 vol. MC buffer supplemented with 1/500 vol. 0.25 M DTE and 1/100 vol. EDTA free protease inhibitor cocktail, incubated on ice for 5 min and lysed in a Dounce homogenizer (18 strokes) at 4 °C. Nuclei were pelleted by centrifugation for 5 min at $18,000 \times g$ (Sorvall SS34 rotor) and dounced (20 strokes) at 4 °C in 1.3 vol. of Roeder C buffer supplemented with 1/500 vol. 0.25 M DTE and 1/200 vol. 0.1 M PMSF. The mixture was stirred for 40 min at 4 °C followed by centrifugation for 30 min at 16,000 rpm (Sorvall SS34 rotor). The supernatant was dialyzed three times for 2 h against 50 vol. of Roeder D buffer. The dialysate was centrifuged for 2 min at $9000 \times g$ (Sorvall SA600 rotor) and aliquots of the supernatant were frozen in liquid nitrogen, stored at -80 °C and tested for splicing activity.

3.2.3.3 In vitro splicing

In vitro splicing was performed using m7G(5')ppp(5')G-capped, ^{32}P -labeled and MS2-tagged pre-mRNA. Splicing reactions contained 50 % (v/v) HeLa nuclear extracts, 65 mM KCl, 3 mM

MgCl₂, 2 mM ATP, 20 mM creatine phosphate and 30 nM pre-mRNA and were incubated for different time intervals at 30 °C. Assembled spliceosomal complexes were analyzed by native gel electrophoresis. RNA was recovered by PCI extraction and analyzed by denaturing gel electrophoresis.

3.2.3.4 Purification of human Prp19/CDC5L complex

hPrp19/CDC5L complex was isolated from AD-002-FLAG/HA-tagged HeLa nuclear extract. Briefly, HeLa S3 cell lines stably expressing FLAG/HA-tagged human AD-002 were generated according to general protocols. Nuclear extract from these cells was prepared according to Dignam *et al.*, 1983. The hPrp19/CDC5L complex was affinity-purified from nuclear extract by using anti-FLAG-M2-Agarose beads and FLAG peptide. Isolated complexes were purified on a 5–20 % glycerol gradient. Protein tagging, stable cell line construction and hPrp19/CDC5L complex purification were performed by M. Grote (Dept. for Cellular Biochemistry; Grote *et al.*, 2010).

3.2.3.5 Purification of human tri-snRNP (U4/U6.U5)

Total snRNPs were isolated from HeLa cell nuclear extract by immunoaffinity chromatography using monoclonal anti-m₃G-specific H20-antibody (Bach *et al.*, 1990; Bringmann *et al.*, 1983). Total snRNPs were separated on 10-30 % glycerol gradients. 1.5 ml fractions were collected yielding the single snRNP components (12S U1 snRNP, 17S U2 snRNP, 20S U5 snRNP and 25S U4/U6.U5 snRNP) in different fractions. This was always verified by analyzing RNA and protein composition after PCI extraction by gel electrophoresis. All fractions were routinely frozen in liquid nitrogen and peak fractions containing only the U4/U6.U5 snRNP were used for the experiments described herein. Isolation of total snRNPs was performed on a routine basis by P. Kempkes and H. Kohansal (Dept. for Cellular Biochemistry) and purification of U4/U6.U5 snRNP was performed by M. Raabe (Bioanalytical Mass Spectrometry Group).

3.2.3.6 Overexpression of MS2-MBP fusion protein

The MS2-MBP fusion protein was used to affinity purify spliceosomal complexes assembled on MS2-tagged pre-mRNA. To overexpress the MS2-MBP protein, 50 µl chemical competent BL21 cells were transfected with 50 ng MS2-MBP fusion protein plasmid DNA by

electroporation and subsequently transferred to LB medium containing ampicillin and chloramphenicol. After incubation at 37 °C overnight, cells were diluted with LB medium, grown at 37 °C until they reached an OD of 0.7 and induced with IPTG. When the cells reached an OD of 1.6 they were harvested by centrifugation for 30 min at 5000 rpm (SL6000). Cells were washed with PBS and the cell pellet was resuspended in Tris-HCl pH 7.5 supplemented with 200 mM NaCl and protease inhibitor cocktail. The cells were solubilized by sonication and centrifuged for 15 min at 4000 rpm (Megafuge) at 4 °C. The supernatant was incubated overnight with amylose beads and the MS2-MBP fusion protein was eluted from the beads with 15 mM maltose in Na-P buffer. The protein was further purified on Heparin beads using 20 mM Hepes pH 7.9, 100 mM KCl, 15 % (v/v) glycerol, 0.5 M DTT and 0.2 M PMSF as elution buffer. The protein concentration of the collected fractions was determined by the Bradford assay (section 3.2.2.1) and the fractions containing MS2-MBP fusion protein were identified by PAGE.

3.2.3.7 Purification of human B and C complexes

Spliceosomal B and C complexes were purified as described before (Bessonov *et al.*, 2008; Deckert *et al.*, 2006). Briefly, ³²P-labeled pre-mRNA was incubated with a 20-fold molar excess of MS2-MBP fusion protein and used without further purification in a standard splicing reaction. This contained 10 nM (B complex) or 30 nM (C complex) pre-mRNA and was incubated at 30 °C for 6 min (B) or 180 min (C). A 30-fold molar excess of M6 and M12 DNA oligonucleotides was added and the reaction was incubated for further 2 min (B) or 20 min (C). Assembled complexes were separated on 10-30 % glycerol gradients and 40-45 S gradient fractions were affinity purified on amylose beads. Spliceosomal B and C complexes were purified in collaboration with S. Bessonov and J. Deckert (Dept. for Cellular Biochemistry).

3.2.3.8 Investigating the protein composition during the spliceosomal assembly

To investigate the time dependent protein assembly of spliceosomal proteins, two different approaches were followed: (i) The time dependent assembly of proteins on different pre-mRNAs, and (ii) the direct comparison of the time dependent protein assembly on PM5 pre-mRNA and on a splicing-inactive pre-mRNA.

To analyze the time dependent assembly of proteins on the tagged pre-mRNA during 30 min, ³²P-labeled pre-mRNA was first incubated with a 20-fold molar excess of MS2-MBP fusion protein. Several standard splicing reactions, each containing 20 pmol pre-mRNA, were then

assembled and incubated at 30 °C for different time intervals using triple (i.e. light, medium and heavy) SILAC extracts (as outlined in detail in Table 3.3). Assembled complexes were affinity purified on amylose beads and the concentration of samples from different time points to be compared (Table 3.3) was determined according to the radioactivity of the used pre-mRNA. Samples from different time points to be compared were pooled in equal molar amounts and proteins were precipitated with ethanol.

To directly compare the protein assembly on PM5 pre-mRNA and on a splicing-inactive pre-mRNA, ³²P-labeled pre-mRNA was incubated with a 20-fold molar excess of MS2-MBP fusion protein. Standard splicing reactions were incubated at 30 °C for different time intervals during 30 min using duplex SILAC extracts (Table 3.4). Assembled complexes from different time points were affinity purified on amylose beads and the concentration was determined according to the radioactivity of the pre-mRNAs. Samples from the same time point but assembled on different pre-mRNAs were pooled in equal molar amounts and proteins were subsequently precipitated with ethanol.

Precipitated proteins were separated by PAGE, hydrolyzed with trypsin as described in section 3.2.4.1 except that 100 mM ammonium bicarbonate buffer was used instead of 50 mM TEAB and analyzed by LC-MS/MS.

Table 3.3: Used SILAC nuclear extracts for investigation of the protein assembly on pre-mRNAs during 30 min. The concentration of samples from different time points to be compared was determined according to the radioactivity contained within the used pre-mRNAs and samples were then pooled in equal molar amounts (Pool 1, Pool 2, Pool 3).

time	0'	2'	5'	0'	10'	15'	0'	20'	30'
NE	light	medium	heavy	light	medium	heavy	light	medium	heavy
	Pool 1			Pool 2			Pool 2		

Table 3.4: Used SILAC nuclear extracts for direct comparison of the protein assembly on PM5 pre-mRNA and on a splicing-inactive pre-mRNA during 30 min. The concentration of samples from different time points to be compared was determined according to the radioactivity contained within the used pre-mRNAs and samples were then pooled in equal molar amounts (Pools 1 - 8).

time	PM5 pre-mRNA	Splicing-inactive pre-mRNA	
0'	heavy NE	light NE	→ Pool 1
2'	heavy NE	light NE	→ Pool 2
5'	heavy NE	light NE	→ Pool 3
10'	heavy NE	light NE	→ Pool 4
15'	heavy NE	light NE	→ Pool 5
20'	heavy NE	light NE	→ Pool 6
25'	heavy NE	light NE	→ Pool 7
30'	heavy NE	light NE	→ Pool 8

3.2.4 Mass spectrometry methods

3.2.4.1 In-gel hydrolysis of proteins

In-gel hydrolysis was performed as previously described with few modifications (Shevchenko *et al.*, 1996). All incubation steps were carried out at 26 °C in a thermomixer at 1050 rpm for 15 min unless otherwise stated. The solutions were removed after incubation steps. Gel slices were cut from entire gel lanes using an in-house manufactured gel cutting device (see Figure 3.2). Gel slices were cut into small pieces and washed with 150 µl water, followed by dehydration with 150 µl acetonitrile. Gel pieces were dried in a vacuum centrifuge and proteins were reduced by addition of 100 µl 100 mM DTT (in 50 mM triethylammonium bicarbonate buffer (TEAB)) and incubation at 56 °C for 50 min. After dehydration with 150 µl acetonitrile, reduced cysteine residues were alkylated by addition of 100 µl 60 mM iodoacetamide (in 50 mM TEAB) and incubation at 26 °C for 20 min. The gel pieces were incubated with 150 µl of 50 mM TEAB for 15 min, followed by addition of 150 µl acetonitrile. After dehydration with 150 µl acetonitrile, the gel pieces were dried in a vacuum centrifuge and rehydrated on ice with buffer 1. The gel pieces were covered with buffer 2 and tryptic digestion was carried out overnight at 37 °C.

Buffer 1:

50 µl 50 mM TEAB
50 µl water
15 µl 0.1 µg/µl Trypsin (Roche)

Buffer 2:

50 µl 50 mM TEAB
50 µl water
-

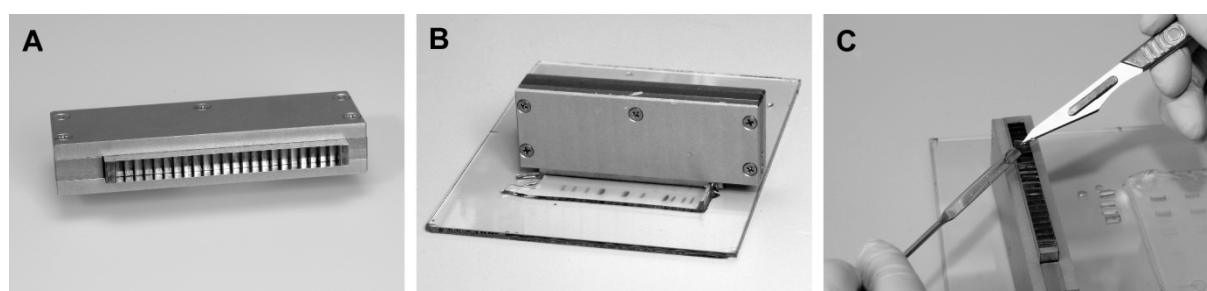


Figure 3.2: In-house manufactured gel cutting device. (A) Side view of the gel cutting device. Entire lanes of NuPAGE pre-cast gels are cut into 23 slices of equal size. **(B)** Gel cutting device attached to a NuPAGE pre-cast gel. **(C)** Cutting gel slices into small pieces.

3.2.4.2 Extraction of peptides

Extraction of peptides from in-gel hydrolysis was performed as described before (Shevchenko *et al.*, 1996). All incubation steps were carried out at 37 °C in a thermomixer at

1050 rpm for 15 min. Briefly, gel pieces were incubated with 50 µl water, followed by addition of 50 µl acetonitrile. The supernatant containing tryptic peptides was removed and collected in a new eppendorf tube. In a second extraction step, 50 µl of 5 % (v/v) formic acid (FA) were added, followed by addition of 50 µl acetonitrile. The supernatant was removed and pooled with the first supernatant. In a third extraction step, 50 µl of acetonitrile were added to the gel pieces to ensure complete extraction of peptides. The supernatant was collected and pooled with the other supernatants. Supernatants were evaporated to dryness in a vacuum centrifuge and the pellet could be stored for prolonged periods at -20 °C.

3.2.4.3 iTRAQ labeling of extracted peptides for relative quantification

For iTRAQ labeling the extracted peptides were dissolved in 20 µl 100 mM TEAB buffer. Internal standards were prepared by mixing 5 µl aliquots of samples to be compared and obtained from different gel lanes, resulting in a final volume of 15 µl (if 2 samples were compared 5 µl TEAB buffer was added). iTRAQ reagents were reconstituted at room temperature in 70 µl ethanol per vial. 5 µl of iTRAQ reagents were added to each sample and incubated at room temperature for 1 h by gentle mixing. Internal standards were labeled with iTRAQ reagent 114 and corresponding samples to be compared were labeled with iTRAQ reagents 115, 116 and 117, respectively. After the reaction, the remaining reagent was quenched by addition of 5 µl of 50 mM glycine and incubation at room temperature for 30 min by gentle mixing. Samples to be compared, i.e. peptides labeled with iTRAQ reagents 115, 116, and 117, and belonging internal standards (labeled with iTRAQ reagent 114) were pooled and dried in a SpeedVacc.

3.2.4.4 In-solution hydrolysis of hPrp19/CDC5L complex

5 µg of purified hPrp19/CDC5L complex was diluted with 25 mM Tris-HCl pH 7.9 to a final volume of 200 µl. 20 µl of 0.5 M sodium acetate buffer pH 5.3 and 600 µl of ice-cold 100 % (v/v) ethanol were added and the sample was incubated for at least 2 h at -20 °C. After centrifugation (30 min, 4 °C, 13,300 rpm) the pellet was washed with 80 % (v/v) ice-cold ethanol followed again by centrifugation (see above). The pellet was dried in a vacuum centrifuge for approximately 5 min. The different in-solution protocols were performed at room temperature with gentle mixing.

(i) In-solution hydrolysis in the presence of urea Hydrolysis of the protein complex in the presence of urea was performed as described before with a few modifications (Gruhler *et al.*, 2005). The protein pellet from ethanol precipitation was dissolved in 20 µl 8 M urea (in 25

mM Tris-HCl pH 7.9) and incubated at room temperature for 15 min. Proteins were reduced by addition of 20 μ l 10 mM DTT (in 8 M urea/25 mM Tris-HCl pH 7.9) and incubation for 30 min. Alkylation of cysteine residues was subsequently performed by addition of 20 μ l 60 mM iodoacetamide (in 8 M urea/25 mM Tris-HCl pH 7.9) and incubation for 30 min. Lys-C (0.6 μ g) was added followed by incubation for 3 h at room temperature. Before tryptic digestion, the sample was diluted to 2 M urea with 100 mM ammonium bicarbonate pH 7.9. Trypsin (1 μ g) was added and proteolysis was continued overnight at room temperature. Aliquots of the hydrolyzed protein complex were stored at -80 °C.

(ii) In-solution digestion in the presence of acetonitrile The protein pellet from ethanol precipitation was dissolved in 50 μ l 80 % (v/v) acetonitrile/20 mM ammonium bicarbonate pH 7.9 and incubated at room temperature for 15 min. Proteins were reduced by addition of 50 μ l 10 mM DTT (in 80 % (v/v) acetonitrile/20 mM ammonium bicarbonate pH 7.9) and incubation for 30 min. Alkylation of cysteine residues was subsequently performed by addition of 50 μ l 60 mM iodoacetamide (in 80 % (v/v) acetonitrile/20 mM ammonium bicarbonate pH 7.9) and incubation for 30 min. Lys-C and Trypsin digestion was then performed as described above except dilution of the used buffer. Aliquots of the hydrolyzed protein complex were stored at -80 °C.

(iii) In-solution digestion in the presence of RapiGest The protein pellet from ethanol precipitation was dissolved in 10 μ l 1 % RapiGest/25 mM ammonium bicarbonate pH 8.5. Proteins were reduced by addition of 10 μ l 50 mM DTT (in 25 mM ammonium bicarbonate) and incubation for 1 h at 37 °C. Alkylation was subsequently performed by addition of 10 μ l 100 mM iodoacetamide (in 25 mM ammonium bicarbonate) and incubation for 1 hr at 37 °C. 0.25 μ g Trypsin in 70 μ l ammonium bicarbonate were added for proteolysis followed by incubation overnight at 37 °C. To decompose RapiGest 20 μ l 5 % (v/v) trifluoroacetic acid (TFA) were added followed by incubation for 2 h at 37 °C. The sample was centrifuged for 30 min at 13,000 rpm. The supernatant was transferred to another tube and dried in a SpeedVacc.

3.2.4.5 Comparison of hydrolysis protocols and selection of standard peptides (AQUA peptides) for absolute quantification

Prior to performing AQUA experiments, the hydrolyzed protein complex was analyzed qualitatively by high-resolution mass spectrometry to determine completeness of digestion and to identify peptide sequences suitable for absolute quantification. Aliquots were analyzed both by nanoLC-ESI-MS/MS on a hybrid Linear Ion Trap-Orbitrap mass spectrometer (LTQ-Orbitrap XL, Thermo Scientific) and by offline nanoLC-MALDI-MS/MS on a Tandem-ToF

mass spectrometer (4800 MALDI TOF/TOF Analyzer, Applied Biosystems/MDS Sciex). Peak lists were generated from raw data using Mascot Daemon (Matrix Science) or the 4000 Series Explorer Remote Client Software (Applied Biosystems/MDS Sciex). The lists were searched against an NCBI non-redundant database (2007, Oct 8th; 5539442 sequences) by using Mascot v2.2.04. For offline nanoLC-MALDI-MS/MS the mass accuracy filter used was 150 ppm for precursor ions and 0.2 Da for product ions. For the Orbitrap mass spectrometer this was 5 ppm for precursor and 0.5 Da for product ions. Peptides with none or maximally two missed cleavage sites were defined as tryptic peptides. Carbamidomethylation of cysteines, oxidation of methionine residues and carbamylation of lysine residues and N-terminal carbamylation for hydrolysis in the presence of urea were allowed as variable modifications.

Up to three peptide sequences per protein were selected from qualitative analysis of the hPrp19/CDC5L complex according to their intensities in MS and MS/MS spectra and the detailed requirements for AQUA peptides as suggested by the manufacturer (Sigma-Genosys and Thermo Fisher Scientific; see below). Where possible, proteotypic peptides (Mallick *et al.*, 2007) were selected. The following is a list of important criteria for standard peptides:

(i) The peptides must resolve well by HPLC; (ii) The peptides must not be too hydrophobic. (A peptide is classified as hydrophobic if more than 50 % of its amino acids were either Ile, Leu, Val, Phe, Trp or Met.); (iii) The peptides must not be too hydrophilic; (iv) The peptides must ionize well to ensure detection in the mass spectrometer; (v) The peptides must not contain chemically reactive amino acids (Cys, Met, Trp); (vi) The peptides must not contain chemically unstable sequences (N-terminal Asn, N-terminal Gln, Asp-Gly); (vii) The peptides' length must be limited to 15 amino acids; (viii) The peptides must contain amino acids that are well suited for labeling with stable isotopes (e.g. Arg, Lys, Phe, Ile).

The different hydrolysis protocols were compared for sequence coverage of the proteins observed, for the number of missed cleavages and the peptide score observed for the endogenous counterparts of the peptides selected for AQUA quantification.

3.2.4.6 Internal standardization with the selected standard peptides

The following peptides were obtained from Sigma-Genosys: TVPEELVKPEELS(¹³C₆¹⁵N₂-K), TLQLDNNFEV(¹³C₆¹⁵N₂-K), ILLGGYQS(¹³C₆¹⁵N₄-R), LGLLGLPAP(¹³C₆¹⁵N₂-K) and EAAAA-LVEEET(¹³C₆¹⁵N₄-R). They were delivered as 1 nmol lyophilized peptide and were dissolved in 20 µl dimethyl formamide by extensive vortexing and sonication in a sonication bath.

Dissolved peptides were diluted with 180 μl of 20 % (v/v) acetonitrile/0.1 % (v/v) FA, resulting in a concentration of 5 pmol/ μl . Small aliquots of the dissolved peptides were stored at $-20\text{ }^{\circ}\text{C}$.

The following peptides were obtained from Thermo Fisher Scientific: FVDILG($^{13}\text{C}_6$ ^{15}N -L)R, HYTFASGSPDN($^{13}\text{C}_6$ ^{15}N -I)K, NVVV($^{13}\text{C}_9$ ^{15}N -F)DK, YADLL($^{13}\text{C}_6$ ^{15}N -L)EK, TIVQLENIYQ-($^{13}\text{C}_6$ ^{15}N -I)K and TGYN($^{13}\text{C}_9$ ^{15}N -F)QR. They were delivered in 5 % (v/v) acetonitrile at a concentration of 5 pmol/ μl . After extensive vortexing and sonication, aliquots were stored at $-20\text{ }^{\circ}\text{C}$. Before use, peptides were dried in a vacuum centrifuge and redissolved in 100 % acetonitrile followed by extensive vortexing and sonication as above to ensure complete solubilization.

Dissolved (Sigma Genosys) and redissolved (Thermo Fisher Scientific) peptides were diluted 2-fold with the appropriate loading buffer (see sections that deal with MS analyses) and again vortexed and sonicated. Peptide mixtures with concentrations of 100, 50, and 25 fmol/ μl of each peptide were prepared. Aliquots from in-solution hydrolysis of the hPrp19/CDC5L complex containing ~ 250 ng were diluted with loading buffer and peptide mixture to give final amounts of ~ 70 ng, 35 ng and 17.5 ng of protein complex and 100 fmol, 50 fmol and 25 fmol of standard peptides per injection (injection volume 5 μl).

3.2.4.7 Absolute quantification by LC-offline MALDI-ToF/ToF-MS (Peak Area)

The sample containing endogenous peptides of the hPrp19/CDC5L complex and standard peptides was separated offline by reversed-phase nanoflow chromatography (Ultimate, Dionex) using 3.5 % (v/v) acetonitrile/0.1 % (v/v) TFA as loading buffer, 0.1 % (v/v) TFA as mobile phase A and 60 % (v/v) acetonitrile/0.1 % (v/v) TFA as mobile phase B. The peptides were loaded on a trap column (μ -PrecolumnTM Cartridge, Acclaim PepMap100 C18, 300 μm i.d. \times 5 mm, LC Packings) at a flow rate of 5 $\mu\text{l}/\text{min}$ and separated at a flow rate of 300 nl/min on an analytical capillary C18 column packed in-house (15 cm, 360 μm o.d., 75 μm i.d., Nucleosil 100-5 C18), with a gradient of 10–60 % buffer B over 60 min. Separated peptides were mixed with α -cyano-4-hydroxycinnamic acid (10 mg/ml in 70 % (v/v) acetonitrile/0.1 % (v/v) TFA) containing 10 fmol/ μl Glu-fibrinogen peptide as internal standard and delivered at a flow rate of 0.9 $\mu\text{l}/\text{min}$. Fractions were spotted every 15 s onto blank stainless steel MALDI targets (Applied Biosystems/MDS Sciex) using a robotic spotting device (Probot, Dionex). The spotted peptide fractions were analyzed by MALDI-Tandem-ToF mass spectrometry (4800 MALDI TOF/TOF Analyser, Applied Biosystems/MDS Sciex) in positive-ion mode. MS spectra were generated with a total of 1000 shots. Up to 20 of the most intense peptide precursors per spot were selected for subsequent MS/MS analysis. For MS/MS a maximum

of 5000 shots were accumulated per precursor using dynamic stop criteria depending on spectral quality. MS/MS spectra were acquired using a collision energy (potential difference between source acceleration voltage and collision cell) of 1 kV, with air at a pressure of 1×10^{-6} torr as collision gas.

The peptides were quantified by their peak areas obtained from single mass spectra using Data Explorer software (Applied Biosystems/Sciex MDS, Foster City). The ratios of the peak areas for the endogenous and the corresponding standard peptides were calculated manually. Protein stoichiometries were determined by comparing the peptide ratios obtained for different proteins.

3.2.4.8 Absolute quantification by LC-online MS-MS/MS (Extracted Ion Chromatograms, XIC)

The sample containing endogenous peptides of the hPrp19/CDC5L complex and standard peptides was independently analyzed by online reversed-phase nanoflow chromatography (HP 1100 series, Agilent; mobile phase A, 0.1 % (v/v) FA; mobile phase B, 95 % (v/v) acetonitrile/0.1% (v/v) FA coupled to a hybrid Linear Ion Trap-Orbitrap mass spectrometer (Thermo Scientific). To this end, the peptides were loaded on a manually packed trap column (1.5 cm, 360 μ m o.d., 75 μ m i.d., Nucleosil 100-5 C18, Macherey-Nagel) and separated with a flow rate of 300 nl/min on an analytical C18 capillary column (30 cm, 360 μ m o.d., 75 μ m i.d., Nucleosil 100-5 C18) with a gradient of 0–38 % mobile phase B over 30 min. Eluted peptides were analyzed directly in the mass spectrometer (LTQ-Orbitrap XL, Thermo Scientific). The LTQ-Orbitrap was operated in data-dependent mode. Survey full scan MS spectra were acquired in the orbitrap (m/z 350–1600) with a resolution of 30,000 at m/z 400 and an automatic gain control target at 10^6 . The five most intense ions were selected for CID (collision induced dissociation) MS/MS fragmentation and detection in the linear ion trap with previously selected ions dynamically excluded for 90 s. Singly charged ions as well as ions with unrecognized charge state were also excluded. Internal calibration of the orbitrap was performed by using the lock mass option (lock mass: m/z 445.120025; Olsen *et al.*, 2005).

The peptides were quantified by generating extracted ion chromatograms (XICs) of the endogenous and the corresponding standard peptide. XICs were generated from the single MS spectra collected over time using the Qual Browser feature of Xcalibur software suite (Thermo Scientific) with a mass tolerance of 5 ppm and a mass precision of 0.001 amu. Signals in the XICs were inspected by eye to make sure that the endogenous and the standard peptides showed the same retention time. The peptide ratios of the endogenous and the corresponding standard peptides were further calculated in a spread sheet (Excel)

from the peak area of the extracted ions. The ratios calculated were used to determine the stoichiometry of the proteins within the complex.

3.2.4.9 Absolute quantification by Multiple Reaction Monitoring

MRM mass spectrometry was carried out on a hybrid Triple Quadrupole/Linear Ion Trap mass spectrometer (4000 QTRAP LC-MS/MS System, Applied Biosystems/MDS Sciex). MRM transitions for each peptide were first designed by nanoSpray direct-infusion mass spectrometry. To this end, each standard peptide was diluted with 40 % (v/v) acetonitrile/0.1 % (v/v) FA to a final concentration of 200 fmol/ μ l and analyzed by Enhanced Resolution (ER) single MS and Enhanced Product Ion (EPI) MS/MS scans to first establish the exact molecular weight and the fragmentation pattern. For each peptide the monoisotopic m/z of the doubly charged precursor was then chosen as Q1 mass, while the three most abundant fragments with an m/z above that of the doubly charged precursor were chosen as Q3 masses. Q1 and Q3 were both set to unit resolution (0.7 FWHM). For each MRM transition the declustering potential (DP), entrance potential (EP), collision energy (CE), and collision cell exit potential (CXP) were finally optimized by ramping the parameters and choosing the values displaying highest signal intensity. MRM transitions for the corresponding endogenous peptides were obtained by Q1/Q3 mass transition using the expected mass differential from the standard peptides.

Samples containing endogenous peptides together with the standard peptides were separated by reversed-phase nanoflow chromatography (Tempo 1D, Applied Biosystems/MDS Sciex) using 2 % (v/v) acetonitrile/0.1 % (v/v) FA as mobile phase A and 98 % (v/v) acetonitrile/0.1 % (v/v) FA as mobile phase B. The peptides were loaded on a trap column (Dionex PepMap C18, 300 μ m, 5 mm) at a flow rate of 20 μ l/min and washed with loading buffer (2 % (v/v) acetonitrile/0.5 % (v/v) FA) for 5 minutes. The peptides were separated on a capillary column (Dionex PepMap C18, 75 μ m, 15 cm) at a flow rate of 300 nl/min with a gradient of 5–40 % mobile phase B over 30 min. Eluted peptides were analyzed directly in the 4000 QTRAP mass spectrometer. For each standard peptide and each endogenous peptide three MRM transitions were monitored with a dwell time of 20 ms per transition. The peptide ratios were obtained by automatic integration of peak areas for each endogenous peptide MRM transition and its AQUA counterpart using MultiQuant 1.0 Software (Applied Biosystems/MDS Sciex). The peptide ratios thus obtained were then used to determine the protein stoichiometry within the complex.

3.2.4.10 Relative quantification by LC-online MS-MS/MS (iTRAQ quantification)

Samples were analyzed on a CAP-LC system coupled to Q-ToF Ultima mass spectrometer (Waters, Manchester, UK). To this end, peptides were dissolved in 10% acetonitrile/0.15 % FA and separated online by reversed phase chromatography using 0.1 % (v/v) FA as mobile phase A and 80 % (v/v) acetonitrile/0.15 % (v/v) FA as mobile phase B. The peptides were loaded on a trap column (μ -PrecolumnTM Cartridge, Acclaim PepMap100 C18, 300 μ m i.d. \times 5 mm, LC Packings) and separated with a flow rate of 200 nl/min on an analytical column (C₁₈, Reprosil, Maisch, Germany; packed in-house) with a gradient of 7-40 % mobile phase B over 50 min. Eluted peptides were directly analyzed in the Q-ToF mass spectrometer in a data dependent manner. MS scans were acquired for 1 s followed by three MS/MS spectra for 3 s each with an ion mass window set to 2.5 Da. MS to MS/MS switch was set to 15 counts/s, and MS/MS to MS was set to an intensity below a threshold of 2 counts/s. Charge state recognition was used to estimate the collision energy for the selected precursors. Scan time and interscan time were set to 0.9 s and 0.1 s, respectively. Data analysis was performed using MassLynx v4.0 software. Peak lists were generated from raw-data using the following settings: smooth window 4.00, number of smooth 2, smooth mode Savitzky-Golay, percentage of peak height to calculate centroid spectra: 80 % with no baseline subtraction. Generated peak lists were searched against NCBI non-redundant database (2007, Oct 8th; 5539442 sequences) using Mascot v.2.2.04 as search engine. Mass accuracy was set to 0.2 Da for the parent and fragment ions. Peptides with no or maximal two missed cleavage sites were defined as tryptic peptides. Carbamidomethylation of cysteines and oxidation of methionine residues were allowed as variable modifications.

Non-normalized peptide ratios for iTRAQ quantification were obtained from Mascot v2.2.04 for unique peptides with a minimum peptide score of 20. Proteins were quantified from the main bands by calculating the mean ratio after manual removal of outliers. Data normalization was performed on proteins known to be present in a 1:1 ratio (5' pre-mRNA cap binding proteins CBP20 and CBP80 for comparison of B and C complexes). Obtained protein ratios were further validated by three independent procedures: (i) calculation of the labeling efficiency for each protein in each band, (ii) using same amounts of non-modified trypsin (Roche) resulting in a 1:1 ratio for autoproteolytic Trypsin peptides, and (iii) analyzing peak intensities of the reporter ions for the internal standards (iTRAQ-114) of low scoring peptides ($\text{Intensity}(\text{iTRAQ-114}) = \frac{1}{3} \text{Intensity}(\text{iTRAQ-115}) + \frac{1}{3} \text{Intensity}(\text{iTRAQ-116}) + \frac{1}{3} \text{Intensity}(\text{iTRAQ-117})$).

3.2.4.11 Relative quantification by LC-online MS-MS/MS (SILAC quantification)

The samples from in-gel digestion were dissolved in 18 μ l 5 % (v/v) acetonitrile/1 % (v/v) FA. Aliquots of 5 μ l were separated by reversed-phase nanoflow chromatography (HP 1100 series, Agilent) and eluted peptides were analyzed directly on a LTQ-Orbitrap XL (Thermo Scientific). For detailed information on experimental conditions see the sections that deal with absolute quantification (3.2.4.6).

Raw data were analyzed using MaxQuant software. Generated peak lists were searched against IPI human decoy database. Peptides with no or maximal two missed cleavage sites were defined as tryptic peptides. Carbamidomethylation of cysteines and oxidation of methionine residues were allowed as variable modifications. Depending on the nuclear extract used (see section 3.2.3.8, Table A and B), stable isotope lysine and arginine residues were set as fixed modifications.

3.2.4.12 Statistical analysis

Statistical analyses were performed using software R version 2.8.0.

To analyze differences in the peptides length and peptide scores, boxplots were generated and the variances were tested for heterogeneity. Differences of mean values were tested for statistical significance by applying Welch's two sample t-test.

To analyze the correlation between protein ratios obtained by iTRAQ and SILAC, the protein ratios were \log_2 transformed and plotted in a scatter plot. The correlation was tested according to Pearson.

4. Results

4.1 Determination of the protein stoichiometry within the hPrp19/CDC5L complex by absolute quantification (AQUA)

The hPrp19/CDC5L complex consists of CDC5L, HSP70, CTNNBL1, PRL1, hPrp19, AD-002, and SPF27 proteins. Previous studies of the yeast homologue of the hPrp19/CDC5L complex revealed that Prp19 forms a tetramer within this complex. For this reason, the stoichiometry within the human complex was analyzed. The protein stoichiometry within protein complexes can be determined by absolute quantification of the complex's proteins. For this purpose absolute protein amounts have to be determined and compared in a relative manner. In proteomic studies absolute protein amounts are often deduced by the peptide concentrations. Here, stable isotope labeled standard peptides were used to determine the concentration of endogenous peptides generated from the hPrp19/CDC5L complex (AQUA). Different mass spectrometric techniques have been applied to obtain signal intensities for endogenous and standard peptides which reflect the relative abundance between these two peptide species. By addition of known amounts of standard peptides absolute amounts of the endogenous peptides and finally the proteins can be determined. The workflow for absolute quantification of the hPrp19/CDC5L complex thus comprised (i) proteomic analysis of the hPrp19/CDC5L complex, (ii) optimization of the hydrolysis conditions of the proteins, (iii) selection of suitable standard peptides for absolute quantification, (iv) synthesis of selected standard peptides by incorporation of stable isotopes, (v) addition of defined amounts of the standard peptides to the hydrolyzed protein complex, (vi) determination of the abundance ratios between the labeled standard peptides and the corresponding endogenous peptides by mass spectrometry, and (vii) comparison of the determined ratios to obtain the protein stoichiometry within the hPrp19/CDC5L complex.

4.1.1 Proteomic analysis of the hPrp19/CDC5L complex

The hPrp19/CDC5L complex was affinity purified from nuclear extract of HeLa cells expressing FLAG/HA-tagged AD-002. A homogeneous preparation of the complex was obtained after glycerol gradient centrifugation. The protein composition of the purified complex in the corresponding fraction was visualized by 1D-PAGE and Coomassie staining of the proteins (Figure 4.1). The proteins were identified by LC-MS/MS of the generated peptides after in-gel digestion of the visible protein bands and, in parallel, by LC-MS/MS of

generated peptides after in-solution hydrolysis (see below) of the gradient fraction containing the purified complex. Consistent with previous studies (Grote *et al.*, 2010; Makarova *et al.*, 2004), mass spectrometric analyses revealed that the hPrp19/CDC5L complex consists of seven proteins, namely CDC5L, Hsp70, CTNNBL1, PRL1, hPrp19, AD-002, and SPF27. Additional proteins could be detected but were represented with only few peptides and not in all biological replicates (for further information on additionally identified proteins see Table A1 in the Appendix). For Hsp70, different isoforms have been identified by MS analysis.

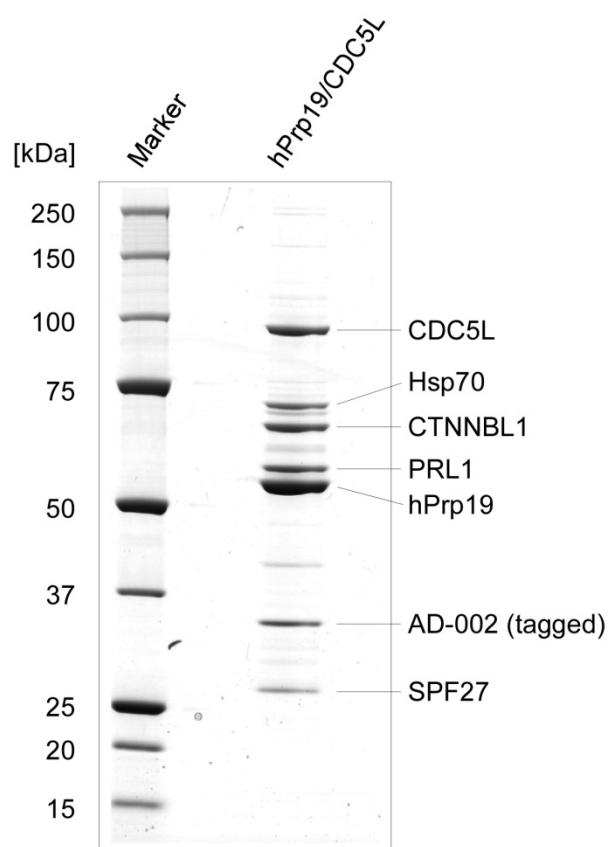


Figure 4.1: 1D-PAGE of the hPrp19/CDC5L complex. The protein composition of the affinity purified complex was visualized by 1D-PAGE and Coomassie staining. Proteomic analysis, i.e. qualitative analysis, revealed that the protein complex consists of seven proteins. In-gel digestion and MS analysis allowed assignment of the proteins to different protein bands.

4.1.2 Comparison of different protocols for in-solution hydrolysis

There are two major prerequisites for absolute quantification using isotope labeled standard peptides: (i) the complete hydrolysis of the proteins to be investigated, and (ii) complete dissolving of the synthetic standard peptides with which the sample is spiked. In this study, three different sets of conditions for enzymatic hydrolysis of the proteins of the hPrp19/CDC5L complex were compared and evaluated: (i) in the presence of urea, (ii) in the presence of acetonitrile, and (iii) in the presence of RapiGest (Waters, Manchester). The different hydrolysis conditions were compared by LC-ESI-MS/MS and LC-offline MALDI-ToF/ToF analysis. To this end, a total amount of 100 ng of hydrolyzed hPrp19/CDC5L complex was analyzed in two to three replicates. As criteria for complete hydrolysis we used

the sequence coverage of the proteins and the presence of missed cleavage site containing peptides within the endogenous quantification proxy (see below).

After hydrolysis in the presence of urea and by the use of LC-ESI-MS/MS we identified all seven proteins of the hPrp19/CDC5L complex with a sequence coverage of 40-70 % (Tables 4.1 and 4.2 A). After hydrolysis in the presence of acetonitrile the proteins were identified with a sequence coverage of 42-85 % (Tables 4.1 and 4.2 A). However, after hydrolysis in the presence of RapiGest the proteins were only identified with a sequence coverage of 24-53 % (Table 4.1). Since the sequence coverage in the presence of RapiGest was significantly lower compared to urea and acetonitrile, hydrolysis in the presence of RapiGest was not used in further experiments.

Table 4.1: Comparison of different hydrolysis protocols. The sequence coverage (SC), the number of unique peptide sequences (UPS), and the number of peptides containing missed cleavage sites (MC) is specified for the hydrolysis in the presence of RapiGest, urea and acetonitrile analyzed by LC-ESI-MS/MS. The obtained sequence coverage is significantly lower for hydrolysis in RapiGest compared to urea and acetonitrile. The parameters listed take all detected protein isoforms into account.

Protein	RapiGest			8M/2M Urea			80 % (v/v) Acetonitrile		
	SC [%]	# UPS	# MC	SC [%]	# UPS	# MC	SC [%]	# UPS	# MC
CDC5L	45	59	34	70	78	45	79	74	37
Hsp70	36/19/24	27/14/8	9/0/5	62/30/39	44/25/16	17/9/6	54/22/41	39/24/11	14/5/1
CTNNBL1	32	37	20	45	53	35	49	50	29
PRL1	49	27	6	69	42	16	72	36	5
hPrp19	24	17	7	50	27	7	60	32	8
AD-002	47	18	11	55	20	15	60	22	15
SPF27	53	17	7	60	25	10	70	22	8

Under all the hydrolysis conditions tested, the overall sequence coverage of the single proteins was higher when LC-ESI-MS/MS was applied as compared with LC-offline MALDI-ToF/ToF-MS (Table 4.2 A and B), presumably due to the higher sensitivity of the used ESI mass spectrometer in MS/MS (LTQ-Orbitrap XL). Comparing the hydrolysis in the presence of urea and acetonitrile, hydrolysis in the presence of acetonitrile resulted in an approximately 10% higher sequence coverage for all proteins (Table 4.2 A and B). An exception is AD-002 which revealed a higher sequence coverage in only one replicate. In contrast, the sequence coverage of two proteins, namely CTNNBL1 and AD-002, was significantly lower (only ~50 % sequence coverage) as compared to other proteins of the

hPrp19/CDC5L complex. This is due to the low number of tryptic cleavage sites within both protein sequences (see below).

Table 4.2: Comparison of different hydrolysis protocols. (A) The sequence coverage (SC), the number of unique peptide sequences (UPS), and the number of peptides containing missed cleavage sites (MC) is specified for three independent replicates after hydrolysis in 8M/2M urea or 80 % (v/v) acetonitrile analyzed by LC-ESI-MS/MS. **(B)** The sequence coverage, the number of unique peptide sequences, and the number of peptides containing missed cleavage sites is specified for two independent replicates after hydrolysis in 8M/2M urea or 80 % (v/v) acetonitrile analyzed by LC-offline MALDI-MS/MS. The parameters listed take all detected protein isoforms into account.

A	LC-ESI-MS/MS (LTQ Orbitrap XL)								
	1			2			3		
	SC [%]	# UPS	# MC	SC [%]	# UPS	# MC	SC [%]	# UPS	# MC
8M/2M Urea									
CDC5L	65	71	42	75	86	46	77	80	43
Hsp70	58/32/27	45/24/14	19/12/4	54/32	38/22/9	13/7/3	50/37	40/27	15/9
CTNBL1	41	57	33	47	48	30	41	47	32
PRL1	48	37	17	69	41	15	68	38	12
hPrp19	40	28	10	59	35	10	60	29	8
AD-002	54	24	19	63	24	18	59	24	18
SPF27	56	23	10	60	28	13	71	26	10
80 % (v/v) Acetonitrile									
CDC5L	76	78	31	85	98	56	82	79	40
Hsp70	58/49/35	40/27/14	12/7/2	52/34	38/20/7	14/5/0	53/38	41/25	17/9
CTNBL1	51	52	28	50	57	34	42	51	31
PRL1	69	33	7	70	45	18	70	43	16
hPrp19	66	36	10	71	44	20	63	42	18
AD-002	55	22	15	55	22	15	49	20	13
SPF27	78	22	7	83	29	15	81	28	12
B	LC-offline MALDI-MS/MS (4800 MALDI ToF/ToF Analyzer)								
	1			2					
	SC [%]	# UPS	# MC	SC [%]	# UPS	# MC			
8M/2M Urea									
CDC5L	33	29	14	43	32	12			
Hsp70	21	12	4	28	15	5			
CTNBL1	23	21	9	33	27	17			
PRL1	21	11	4	33	15	5			
hPrp19	23	12	1	28	12	1			
AD-002	18	5	3	27	7	5			
SPF27	40	8	4	48	15	6			
80 % (v/v) Acetonitrile									
CDC5L	43	31	7	52	36	12			
Hsp70	33/14/27	18/6/7	5/1/1	37	17	3			
CTNBL1	33	21	9	35	25	13			
PRL1	37	19	3	48	17	4			
hPrp19	33	16	3	47	18	4			
AD-002	29	8	5	29	5	3			
SPF27	43	13	7	50	13	6			

Hydrolysis in the presence of urea always risks artifactual carbamylation of the generated peptides. This is because isocyanic acid is generated upon decomposition of urea which in turn can carbamylate the peptide's N-terminus. Carbamylated peptides show different properties in liquid chromatography and the mass spectrometer when compared to the non-modified peptides. Consequently, the amount of the endogenous peptide available for quantification is reduced and does not reflect the actual concentration. Quantification of this specific peptide is then not possible. Performing the hydrolysis at room temperature reduced the amount of carbamylated peptides to 1-4 % and possible carbamylation of peptides can then be neglected.

The two hydrolysis conditions, i.e. hydrolysis in the presence of urea and in the presence of acetonitrile, were also compared regarding peptides containing missed cleavage sites for the particular endoproteinase used in the hydrolysis. As we used a highly sensitive mass spectrometer (LTQ-Orbitrap XL) more peptides containing missed cleavage sites (~10 %) were detected by LC-ESI-MS/MS as compared to LC-offline MALDI-ToF/ToF-MS (compare Tables 4.2 A and B). Peptides containing missed cleavage sites were more frequently detected when they were generated in the presence of urea as compared to acetonitrile (Table 4.2 A and B). They also showed higher peptide scores for the peptide sequences selected for quantification (see below) as compared to hydrolysis in the presence of acetonitrile, i.e. they were sequenced with a higher probability when generated from hydrolysis in the presence of urea (Figure A.1 in the Appendix). Importantly, some peptides were very frequently detected with missed cleavage site (e.g. EAAAALVEEETR/R and MK/ILLGGYQSR derived from proteins SPF27 and CDC5L, respectively). Close inspection of the missed cleavage site-containing peptides showed that in the presence of acetonitrile incompletely cleaved peptides are significantly longer than incompletely cleaved peptides generated in the presence of urea. To further investigate this observation, the peptide length (number of amino acids) of missed cleavage site-containing peptides obtained by the two different hydrolysis protocols was statistically analyzed using the software R (version 2.8.0). Differences of mean values for peptide length were tested for statistical significance by applying Welch's two sample t-test (Figure 4.2). A p-value of 0.035 confirms the observation that incompletely cleaved peptides in the presence of acetonitrile are longer when compared to incompletely cleaved peptides generated in urea. On average, missed cleavage site-containing peptides generated in the presence of acetonitrile are 14.1 amino acids long whereas missed cleavage site-containing peptides generated in the presence of urea are 12.8 amino acids long (the statistical output is shown in Figure A.2 in the Appendix). However, these peptides were only detected by LC-ESI-MS/MS and were much less frequently detected than their correctly cleaved counterparts (as determined by spectral count; for further information see Table A2 in the Appendix).

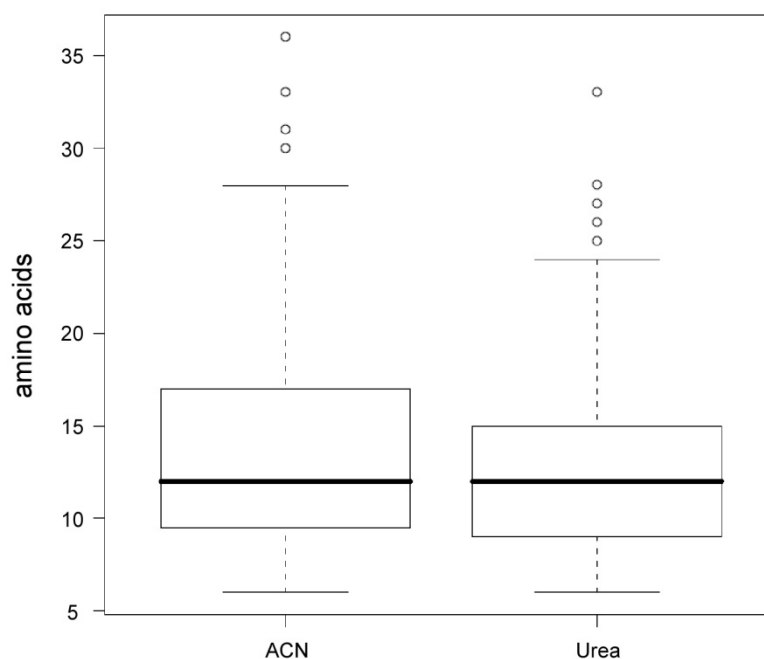


Figure 4.2: Miscleaved peptides obtained by hydrolysis in the presence of acetonitrile are significantly longer than miscleaved peptides obtained by hydrolysis in the presence of urea. Boxplot of peptide length (number of amino acids) for miscleaved peptides obtained by hydrolysis in the presence of acetonitrile and urea. On average, incompletely cleaved peptides generated in the presence of acetonitrile are 14.1 amino acids long, whereas incompletely cleaved peptides generated in urea are 12.8 amino acids long. The output of the statistical analysis is shown in Figure A.2 in the Appendix.

In summary, both protocols for hydrolysis of the hPrp19/CDC5L complex produced similar results in the database search. As described above, the effects of carbamylation and the generation of missed cleavage site-containing peptides can be neglected. Consequently, the hPrp19/CDC5L complex was analyzed after hydrolysis according to both protocols.

4.1.3 Selection of standard peptides for absolute quantification

After optimization of hydrolysis conditions of the proteins to be investigated, the next step for absolute quantification involved the selection of suitable standard peptides. Sequenced endogenous peptides were evaluated whether they meet the selection criteria for standard peptides (see section 3.2.4.3). To this end, endogenous peptides were analyzed for their amino acid composition, length and hydrophobicity. The MS and MS/MS spectra of suitable peptides were inspected for sufficient intensities and a proper retention time in chromatography, i.e. elution of the different peptides in a well-separated range and no peak overlap with other peptides. Finally, the chosen peptides were checked to see whether they meet the criteria for proteotypic peptides (Mallick *et al.*, 2007). The latter are peptides that are repeatedly and consistently identified for any given protein in a mixture and are thus suitable to function as standard peptides during absolute quantification. The selection of proteotypic peptides increases the possibility that the peptide of interest (here, the peptide selected for quantification) is identified (sequenced) during MS analysis and thus allows the unambiguous assignment of the identified (sequenced) peptide to the protein of interest.

Even though a high number of peptides derived from the hPrp19/CDC5 complex proteins could be identified by MS and MS/MS analysis, only a very few peptide sequences come into consideration as standard peptides for absolute quantification. For reliable absolute quantification it is recommended to select up to three standard peptides per protein. For most of the proteins, size and amino acid sequence limit the number of peptides that are suitable to serve as standard peptides. Small proteins do not generate a high number of peptides during hydrolysis, thus reducing the selection. The amino acid sequence further limits the selection of standard peptides as certain amino acids are unsuitable (e.g. chemically reactive amino acids). Figure 4.3 shows the protein sequences of AD-002 and CTNNBL1 and theoretically expected tryptic peptides that can be analyzed by mass spectrometry (> 700 Da). According to the above mentioned criteria, no peptide for AD-002 and only one peptide of CTNNBL1 could be selected for quantitative analysis.

The standard peptides were synthesized by incorporation of stable isotope labeled amino acids and synthetic peptides obtained were checked mass and sequence by MALDI-ToF/ToF-MS. Synthesized standard peptides were obtained by Sigma-Genosys or Thermo Fisher Scientific (for further information see section 3.2.4.4). Table 4.3 shows selected standard peptides for absolute quantification of the hPrp19/CDC5L complex.

A AD-002 sequence

1
MTTAARPTFE PARGGRGKGE GDLSQLSKQY SSRDLPSHTK IKYRQTTQDA PEEVRNRDFR RELEERERAA 51
 AREKNRDRPT REHTTSSSVS KKPRLDQIPA ANLDADDPLT DEEDEDFFEE SDDDDTAALL AELEKIKKER
 101
 151 201
AEEQARKEQE QKAEERIRM ENILSGNPLL NLTGPSQPQA NFKVKRRWDD DVVFKNCAKG VDDQKKDKRF
 VNDTLRSEFH KKFMEKYIK

B

[M+H] ⁺	Position	Peptide sequence	Remarks
1448.73	1-13	<u>MTTAARPTFEPAR</u>	proteotypic, contains methionine
1033.52	19-28	<u>GEGDLSQLSK</u>	not proteotypic
797.42	34-40	<u>DLPSHTK</u>	not proteotypic
1273.60	45-55	<u>QTTQDAPEEVR</u>	N-terminal glutamine, not proteotypic
1062.51	82-91	<u>EHTTSSSVSK</u>	not proteotypic
4535.00	95-135	<u>LDQIPAAANLDADDPLTDEEDE-</u> <u>DFEEESDDDDTAALLAELEK</u>	too long
703.34	141-146	<u>AEEQAR</u>	too short, not proteotypic
2583.32	160-183	<u>MENILSGNPLLNLGPSQPQANFK</u>	too long
1023.48	188-195	<u>WDDDVVFK</u>	proteotypic, contains tryptophan

C CTNNBL1 sequence

1
 MVDGQVVALL VQNLERLDES VKEEADGVHN TLAIVENMAE FRPEMCTEGA QQGLLQWLLK RLKAKMPFDA 51
 NKLYCSEVLA ILLQDNDENR ELLGELDGID VLLQQLSVFK RHNPSTAAEQ EMMENLFDLSL CSCMLSSNR
 101
 151 201
ERFLKGEGLQ LMNLMLEK ISRSSALKVL DHAMIGPEGT DNCHK FVDIL GLR TIFPLFM KSPRKIKKVG
 251
 TTEKEHEEHV CSILASLLRN LRGQQRTRLL NKFTENDSEK VDRLMELHFK YLGAMQVADK KIEGEKHDVM
 301
 RRGEIIDNDT EEEFYLRRLD AGLFLVQHC YIMAEICNAN VPQIRQRVHQ ILNMRGSSIK IVRHIIKEYA
 351
 ENIGDGRSPE FRENEQKRIL GLENF

D

[M+H] ⁺	Position	Peptide sequence	Remarks
1783.97	1-16	<u>MVDGQVVALLVQNLER</u>	contains methionine, not proteotypic
4271.06	23-60	<u>EEADGVHNTLAIVENMAEFRPE-</u> <u>MCTEGAQQGLLQWLLK</u>	too long
822.38	66-72	<u>MPFDANK</u>	contains methionine, not proteotypic
2108.03	73-90	<u>LYCSEVLAILLQDNDENR</u>	too long, not proteotypic
2229.24	91-110	<u>ELLGELDGIDVLLQQLSVFK</u>	too long, not proteotypic
3316.40	112-140	<u>HNPSTAAEQEMMENLFDLSLCS-</u> <u>CLMLSSNR</u>	too long
1374.72	146-157	<u>GGLQLMNLMLR</u>	proteotypic, contains methionine
1836.84	169-185	<u>VLDHAMIGPEGTDNCHK</u>	proteotypic, contains methionine
932.56	186-193	<u>FVDILGLR</u>	proteotypic
996.56	194-201	<u>TIFPLFMK</u>	proteotypic, contains methionine
1735.88	215-229	<u>EHEEHVCSILASLLR</u>	proteotypic, contains cysteine
969.42	243-250	<u>FTENDSEK</u>	not proteotypic
917.49	254-260	<u>LMELHFK</u>	proteotypic, contains methionine
1095.55	261-270	<u>YLGAMQVADK</u>	contains methionine, not proteotypic
1842.84	283-297	<u>GEIIDNDTEEEFYLR</u>	not proteotypic
3044.55	299-325	<u>LDAGLFLVQHICYIMAEICNANV-</u> <u>PQIR</u>	too long
1010.56	328-335	<u>VHQILNMR</u>	contains methionine, not proteotypic
1123.50	348-357	<u>EYAENIGDGR</u>	not proteotypic

Figure 4.3: Analysis of AD-002 and CTNNBL1 sequences for presence of standard peptides. (A) AD-002 sequence. Proteotypic peptides are underlined. **(B)** Theoretically expected tryptic peptides > 700 kDa of AD-002 that could be detected by MS. Peptides were inspected according to the selection criteria of standard peptides. No peptide of AD-002 could be selected for quantitative analysis. **(C)** CTNNBL1 sequence. Proteotypic peptides are underlined. **(D)** Theoretically expected tryptic peptides > 700 Da of CTNNBL1 that could be detected by MS. Only one peptide could be selected for quantitative analysis (FVDILGLR, highlighted in bold).

Table 4.3: Selected standard peptides. Up to three peptide sequences were selected from proteomic analysis of the hPrp19/CDC5L complex. Owing to protein size and sequence the number of peptides that fulfilled the requirements to serve as standard peptides was limited. No peptide for AD-002 was available.

protein	peptide sequence	mass [Da]		Δm [Da]	proteotypic
		endogenous	standard		
CDC5L	ILLGGYQS(R)	1005	1015	10	◆
CDC5L	LGLLGLPAP(K)	977	985	8	◆
CDC5L	YADLL(L)EK	963	970	7	◆
CTNBL1	FVDILG(L)R	931	938	7	◆
PRL1	HYTFASGSPDN(I)K	1435	1442	7	◆
PRL1	TGYN(F)QR	884	894	10	◆
hPrp19	TLQLDNNFEV(K)	1319	1327	8	
hPrp19	NVVV(F)DK	819	829	10	◆
SPF27	EAAAALVEEET(R)	1287	1297	10	
SPF27	TIVQLENEIQ(I)K	1589	1596	7	◆

The chromatographic behavior of the synthetic standard peptides was investigated whether they reveal the same behavior as their endogenous counterparts. Surprisingly, standard peptides that were delivered in 5 % (v/v) acetonitrile and subsequently stored in smaller aliquots at -20 °C did not reveal sufficient signal intensity in nanoLC. These peptides were only detectable in LC after drying the aliquots in a vacuum centrifuge and re-dissolving them in 100 % (v/v) acetonitrile before use. Complete solubility of the standard peptides is a major prerequisite for reliable and accurate absolute quantification. During this study, we found that an initial LC analysis of peptides dissolved under different conditions gave an easy estimate on the solubility and chromatography behavior of the synthetic standard peptides.

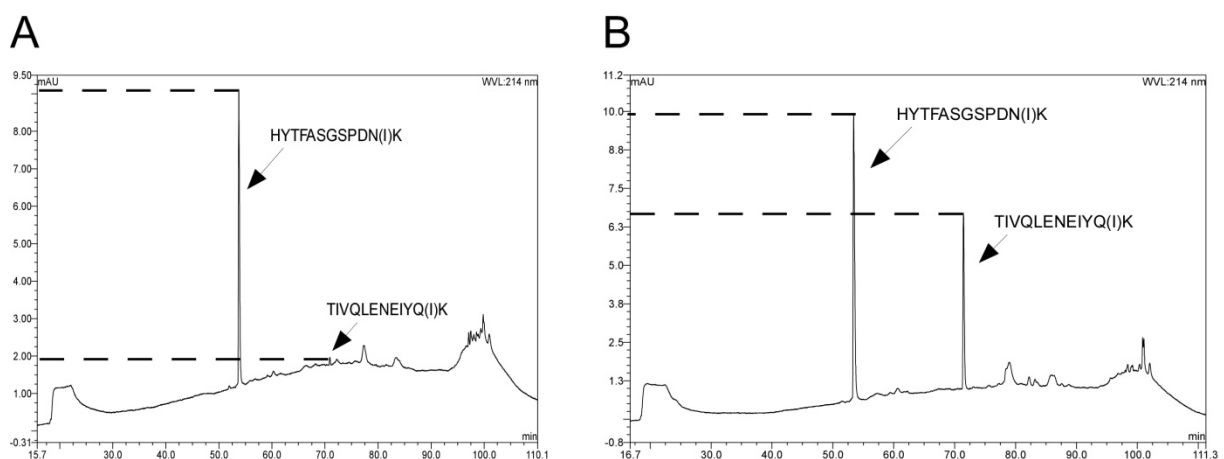


Figure 4.4: Solubility of standard peptides. After storage at -20°C, peptides delivered in 5 % (v/v) acetonitrile did not reveal sufficient signal intensity in LC. **(A)** The peptides HYTFASGSPDN(I)K and TIVQLENEIQ(I)K were diluted with 3.5 % (v/v) acetonitrile/0.1 % (v/v) TFA, 1.25 pmol of each peptide were subjected to nanoLC and detected at 214 nm. **(B)** The peptides were dried in a vacuum centrifuge and re-dissolved in 100 % (v/v) acetonitrile. They were diluted with 3.5 % (v/v) acetonitrile/0.1 % (v/v) TFA and 1.25 pmol of each peptide were subjected to nanoLC.

4.1.4 Absolute quantification by LC-offline MALDI-ToF/ToF-MS (peak area)

In this part, different mass spectrometry techniques have been applied to obtain signal intensities for endogenous and standard peptides of the hPrp19/CDC5L complex proteins. The use of MALDI mass spectrometry features several advantages. First, the generation of singly charged ions facilitates data analysis and the peak area of the peptides can be easily obtained. Second, the coupling of nanoLC and MALDI-ToF/ToF-MS allows reinvestigation of the samples, i.e. low abundance precursors that are used for quantification can be detected. This means that the sample separated by nanoLC and spotted on a MALDI target can be manually reinvestigated if precursors have been missed during the analysis.

For absolute quantification, various amounts of hydrolyzed hPrp19/CDC5L complex were supplemented with equal amounts of all standard peptides (Table 4.3) and *vice versa*. Peak area ratios of endogenous and standard peptides were obtained from individual MS spectra (for an example spectrum see Figure 2.3 in the Introduction). Protein ratios were calculated from average peptide ratios of three technical replicates. Protein stoichiometries within the complex were determined by comparing protein ratios. They are displayed by the ratio of hPrp19 to each of the single proteins of the complex (Table 4.4).

For both hydrolysis conditions (i.e. urea and acetonitrile) the ratios indicate that hPrp19 is present in a higher stoichiometry than the other proteins. However, the values obtained for the two hydrolysis protocols are not consistent. Hydrolysis in the presence of urea yielded a stoichiometry of 2.5 for hPrp19 relative to CDC5L, PRL1 and CTNNBL1, and of 5.5 relative to SPF27. Hydrolysis in acetonitrile yielded a stoichiometry of 2:1 for hPrp19 relative to CDC5L and CTNNBL1, and of approximately 4:1 relative to PRL1 and SPF27 (Table 4.4).

Table 4.4: Relative protein stoichiometries within the hPrp19/CDC5L complex determined by LC-offline MALDI-ToF/ToF-MS. Average peptide ratios of three replicates were used to calculate protein ratios. Protein stoichiometries are displayed by the ratio of hPrp19 to the relevant protein showing the stoichiometry of hPrp19 within the hPrp19/CDC5L complex. Values in parentheses are not included in the statistics (apparent outliers).

Protein complex [ng] standard peptides [fmol]	70	35	17.5	70	70	Average
8M/2M urea						
Protein stoichiometry						
hPrp19/CDC5	2.48	2.24	2.38	2.86	2.74	2.54 ± 0.256
hPrp19/SPF27	2.44	2.54	3.26	1.87	(0.67)	2.53 ± 0.571
hPrp19/PRL1	4.35	4.76	6.00	6.75	5.91	5.55 ± 0.979
hPrp19/CTNNBL1	2.30	1.76	2.01	2.80	2.58	2.29 ± 0.419
80% (v/v) acetonitrile						
Protein stoichiometry						
hPrp19/CDC5	1.76	1.64	1.86	2.01	1.85	1.82 ± 0.136
hPrp19/SPF27	5.36	3.08	4.79	(1.96)	(2.19)	4.41 ± 1.187
hPrp19/PRL1	3.79	4.28	4.79	5.03	6.15	4.81 ± 0.890
hPrp19/CTNNBL1	1.92	1.19	2.01	2.69	2.34	2.17 ± 0.559

One reason for this clear discrepancy might have been the frequently observed peak overlap in LC-offline MALDI-MS. In case of the selected peptide sequence derived from the protein CTNNBL1, peak overlap of the corresponding standard peptide was observed in all spectra. Since only one peptide for this protein is available for quantification (see Table 4.3), no further reference value was present and the protein could thus not be quantified reliably by this method. Figure 4.3 shows MALDI MS spectra of peptide FVDILGLR (generated from CTNNBL1) in three consecutive spots during LC-offline MALDI-ToF/ToF-MS (Figure 4.5 A-C). Peak overlap was observed in the spot with highest intensity of the peptides. The calculated peptide ratio within this particular spot differs from the calculated peptide ratio (Figure 4.5 D) within the adjacent spots presumably due to peak overlap of the standard peptide.

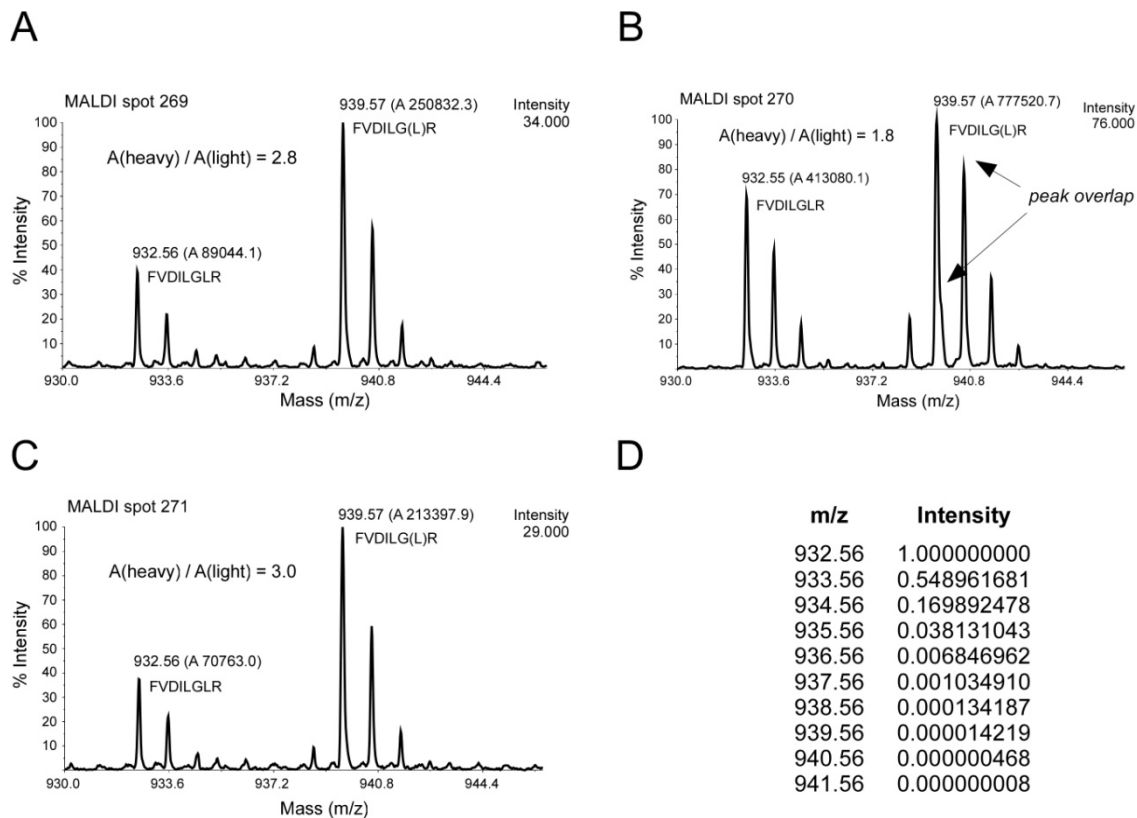


Figure 4.5: Peak overlap is frequently observed in MALDI mass spectrometry. Upon closer inspection of the MALDI MS spectra for FVDILGLR and the corresponding standard peptide, peak overlap was observed in almost all analyses. (A-C) show MS spectra of FVDILGLR in three consecutive spots (269-271). (A) FVDILGLR and the corresponding standard peptide first appeared in spot 269. The isotopic patterns of both peptides are consistent with the calculated isotopic pattern on FVDILGLR (D). The calculated ratio of the endogenous and the standard peptide is 2.8. (B) The peptides show highest intensity in spot 270. The isotopic pattern of the standard peptide is different from the endogenous and the calculated isotopic patterns. In addition, the monoisotopic peak shows a shoulder. These findings lead to the conclusion that two peaks overlap. The calculated peptide ratio is 1.8. (C) In the subsequent spot (271) the peptides show lower intensity but again show the theoretical isotopic pattern and a ratio of 3.0 comparable with spot 269. (D) Calculated intensities of the isotopic pattern of FVDILGLR.

In summary, quantification by LC-offline MALDI-TOF/TOF-MS does not show unambiguously interpretable protein stoichiometries within the hPrp19/CDC5L complex although the data do

suggest that hPrp19 is present in more than one copy. It is surprising that although similar values for the sequence coverage have been observed for both hydrolysis protocols (i.e., urea and acetonitrile) the determined protein stoichiometries differ significantly. These data therefore indicate that the hydrolysis conditions are a critical parameter in the absolute quantification using standard peptides.

4.1.5 Absolute quantification by LC-ESI-MS (extracted ion chromatograms)

Since MALDI mass spectrometry did not reveal clear protein stoichiometries within the hPrp19/CDC5L complex, LC-ESI-MS was used as a complementary method. In this study, extracted ion chromatograms (XICs) were used to read out signal intensities for quantification. Various amounts of hydrolyzed hPrp19/CDC5L complex were supplemented with equal amounts of all standard peptides (Table 4.3) and *vice versa*. XICs were generated for endogenous and corresponding standard peptides for the single proteins and resulting peak areas were used to calculate peptide ratios (for an example see Figure 2.7 in the Introduction). Protein ratios were calculated from average peptide ratios from three replicates.

Protein ratios are displayed as the ratio of hPrp19 to other proteins of the complex for both hydrolysis conditions (i.e. hydrolysis in the presence of urea and acetonitrile, respectively). Hydrolysis in the presence of urea yielded a stoichiometry of approximately 3:1 for hPrp19 relative to CDC5L, SPF27, and CTNNBL1 and of 9:1 relative to PRL1. Hydrolysis in the presence of acetonitrile yielded a stoichiometry of 2:1 for hPrp19 relative to CDC5L, 4:1 relative to CTNNBL1, and approximately 5:1 relative to PRL1 and SPF27 (Table 4.5).

Table 4.5: Relative protein stoichiometries within the hPrp19/CDC5L complex determined by LC-ESI-MS. Average peptide ratios of three replicates were used to calculate protein ratios. Protein stoichiometries are displayed by the ratio of hPrp19 to the relevant protein showing the stoichiometry of hPrp19 within the hPrp19/CDC5L complex. Values in parentheses are not included in the statistics (apparent outliers).

Protein complex [ng] standard peptides [fmol]	70 100	35 100	17.5 100	70 50	70 25	Average
8M/2M urea						
Protein stoichiometry						
hPrp19/CDC5	3.87	3.61	3.62	3.39	3.42	3.58 ± 0.193
hPrp19/SPF27	3.66	2.66	2.80	(1.70)	(1.04)	3.04 ± 0.541
hPrp19/PRL1	7.59	8.11	10.03	9.33	9.79	8.97 ± 1.069
hPrp19/CTNNBL1	(6.52)	3.8	3.57	2.99	3.06	3.36 ± 0.393
80 % (v/v) acetonitrile						
Protein stoichiometry						
hPrp19/CDC5	1.87	2.37	2.38	2.31	2.28	2.24 ± 0.212
hPrp19/SPF27	4.77	4.76	4.91	(8.67)	(7.48)	4.81 ± 0.084
hPrp19/PRL1	3.42	6.56	7.53	4.51	5.77	5.56 ± 1.628
hPrp19/CTNNBL1	3.05	4.44	4.71	4.12	3.51	3.97 ± 0.680

The values yielded by the two different hydrolysis protocols are again inconsistent. In both hydrolysis procedures, the values for PRL1 protein differ significantly among the different analyses. PRL1 shows higher values for both hydrolysis protocols compared with the other proteins; these might be due to incomplete digestion or to interference with co-eluting peptides. Indeed, selected peptides for PRL1 co-eluted with other peptides in the analysis and thus affected the quantification.

As for the MALDI analysis, no clear protein stoichiometry could be determined by LC-ESI-MS using XIC signal intensities for quantification. A method which specifically identifies and quantifies endogenous and standard peptides is therefore required. One such method is multiple reaction monitoring (MRM) in which two quadrupoles of a triple quadrupole mass analyzer are operated as mass filters to detect the specific transition of a precursor ion to a specific fragment ion.

4.1.6 Absolute quantification by Multiple Reaction Monitoring (MRM)

The use of a triple quadrupole mass analyzer allows the detection of the specific transition from a given precursor to a user-defined fragment ion (single reaction monitoring). The precursor mass is selected in quadrupole Q1, fragmentation takes place in q2, and the fragment ion is detected in Q3. Multiple reaction monitoring (MRM) allows the detection of multiple fragment ions specific for one precursor. Signals from MRM transitions of endogenous and standard peptides are well suited for absolute quantification of the peptide/protein under investigation. In previous studies, MRM with standard peptides has proven to be a suitable method for absolute quantification of proteins in a mixture (Abbatiello *et al.*, 2008; Langenfeld *et al.*, 2009). In LC-coupled MRM experiments, peak overlaps caused by co-eluting peptides can be neglected under defined conditions, namely by choosing several MRM transitions specific for a certain precursor. Operating the first and the third quadrupole as mass filters guarantees that only the specific MRM transitions are monitored and co-eluting peptides do not influence the quantitative signal.

For absolute quantification of the hPrp19/CDC5L complex, three MRM transitions for each selected peptide sequence were designed. In all cases, the doubly charged precursor was chosen as Q1 mass, and the three most intense y-type fragment ions with an m/z above that of the precursor were chosen as Q3 masses. For the selected MRM transitions, the declustering potential (DP), entrance potential (EP), collision energy (CE), and collision cell exit potential (CXP) of the instrument were first optimized (for information about MRM transitions and optimized instrument parameters see Table A.3 in the Appendix) and then tested by analyzing the standard peptides and the endogenous peptides separately. For this

purpose, 70 ng of hydrolyzed hPrp19/CDC5L complex or 100 fmol of standard peptides were separated by LC, and the MRM transitions were monitored using the optimized parameters. All MRM transitions were well separated and showed sufficient intensity for quantitative analysis (Figure 4.6). The analysis of the hydrolyzed hPrp19/ CDC5L complex showed no MRM transition for the heavy counterparts (Figure 4.6 A) and *vice versa* (Figure 4.6 B). We infer that the chosen MRM transitions are highly peptide-specific and can therefore be used for their investigation.

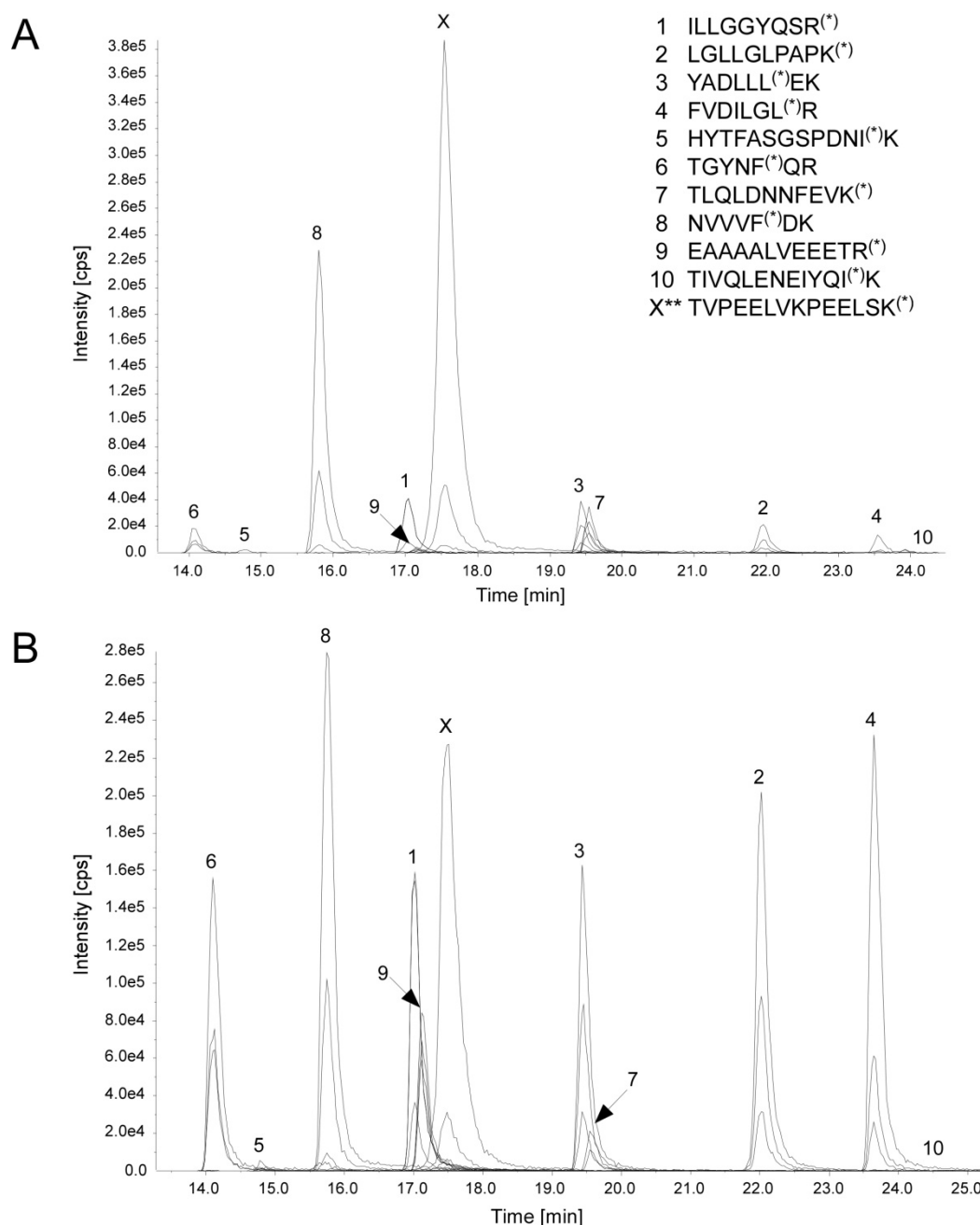


Figure 4.6: Specificity of the designed MRM transitions. Labeled standard and endogenous peptides were analyzed separately. The detected MRM transitions show sufficient intensity for quantitative analysis. **(A)** 70 ng of hydrolyzed hPrp19/CDC5L complex were separated by LC and analyzed by MRM. Traces for MRM transitions of standard peptides are empty. **(B)** 100 fmol of standard peptides were separated by LC and analyzed by MRM. Traces for MRM transitions of endogenous peptides are empty. ** TVPEELVKPEELSK was not used for absolute quantification. (*) labeled amino acid (standard peptides).

As described for MALDI and ESI analyses, various amounts of hydrolyzed hPrp19/CDC5L complex were supplemented with equal amounts of all standard peptides and *vice versa*. In the different experiments, three MRM transitions were monitored for each standard and endogenous peptide. Figure 4.7 shows an example of the total of six MRM transitions for the coeluting standard and endogenous peptides (LGLLGLPAPK derived from CDC5L). Ratios between standard and endogenous peptides were obtained by integration of the peak areas of the corresponding transitions and the values were then used to calculate protein ratios from three technical replicates.

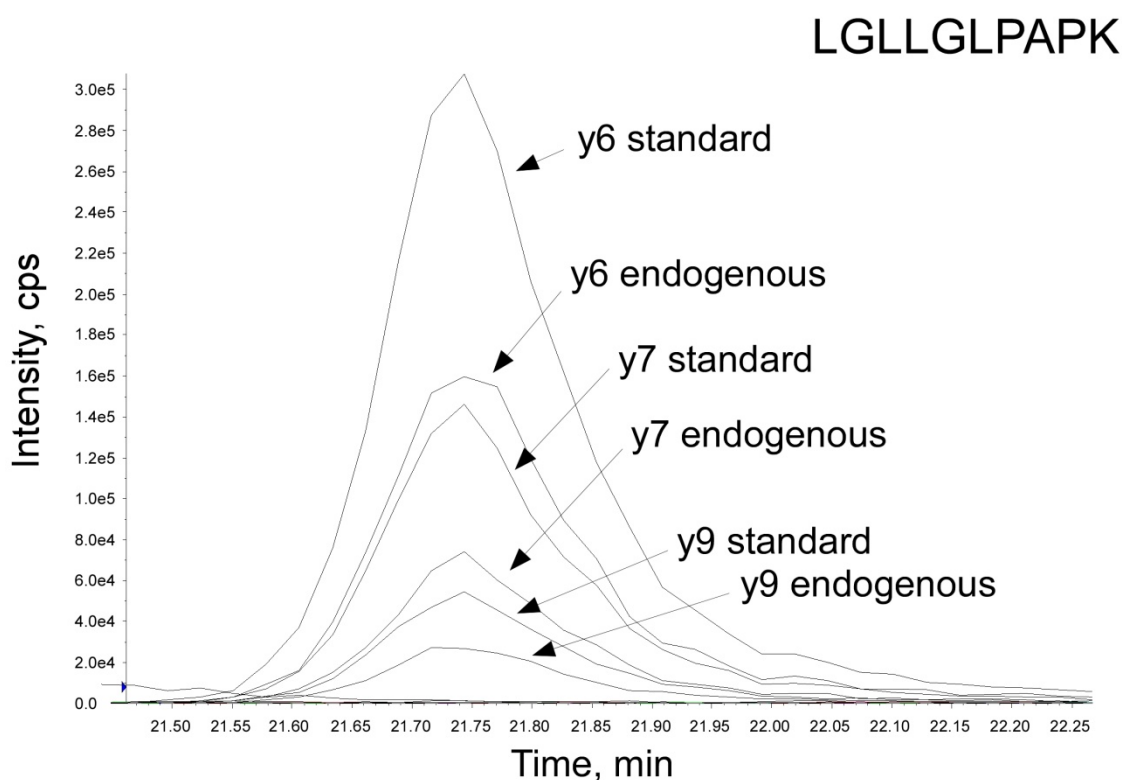


Figure 4.7: Example of the MRM transitions for an endogenous and the corresponding standard peptide (LGLLGLPAPK derived from CDC5L protein). The doubly charged precursor mass was selected as Q1 mass and the y6, y7, and y9 fragment ions were chosen as Q3 masses. Peptide ratios were calculated from integrated peak areas of the corresponding transitions.

The peptide ratios based on the three MRM transitions for the individual peptides are compared in bar Figure 4.8. Clearly, peptide ratios obtained for a given protein differed significantly when hydrolysis was performed in the presence of urea as compared with the corresponding ratios as obtained from hydrolysis in the presence of acetonitrile (hPrp19, SPF27, and PRL1, Figure 4.8 A). Peptide ratios obtained from hydrolysis in acetonitrile were more consistent for the same protein, i.e. peptide ratios obtained for peptides generated from the same protein show comparable values (Figure 4.8 B). The standard deviations of the protein ratios derived after hydrolysis in the presence of acetonitrile were lower than the corresponding values after digestion in the presence of urea (Table 4.6).

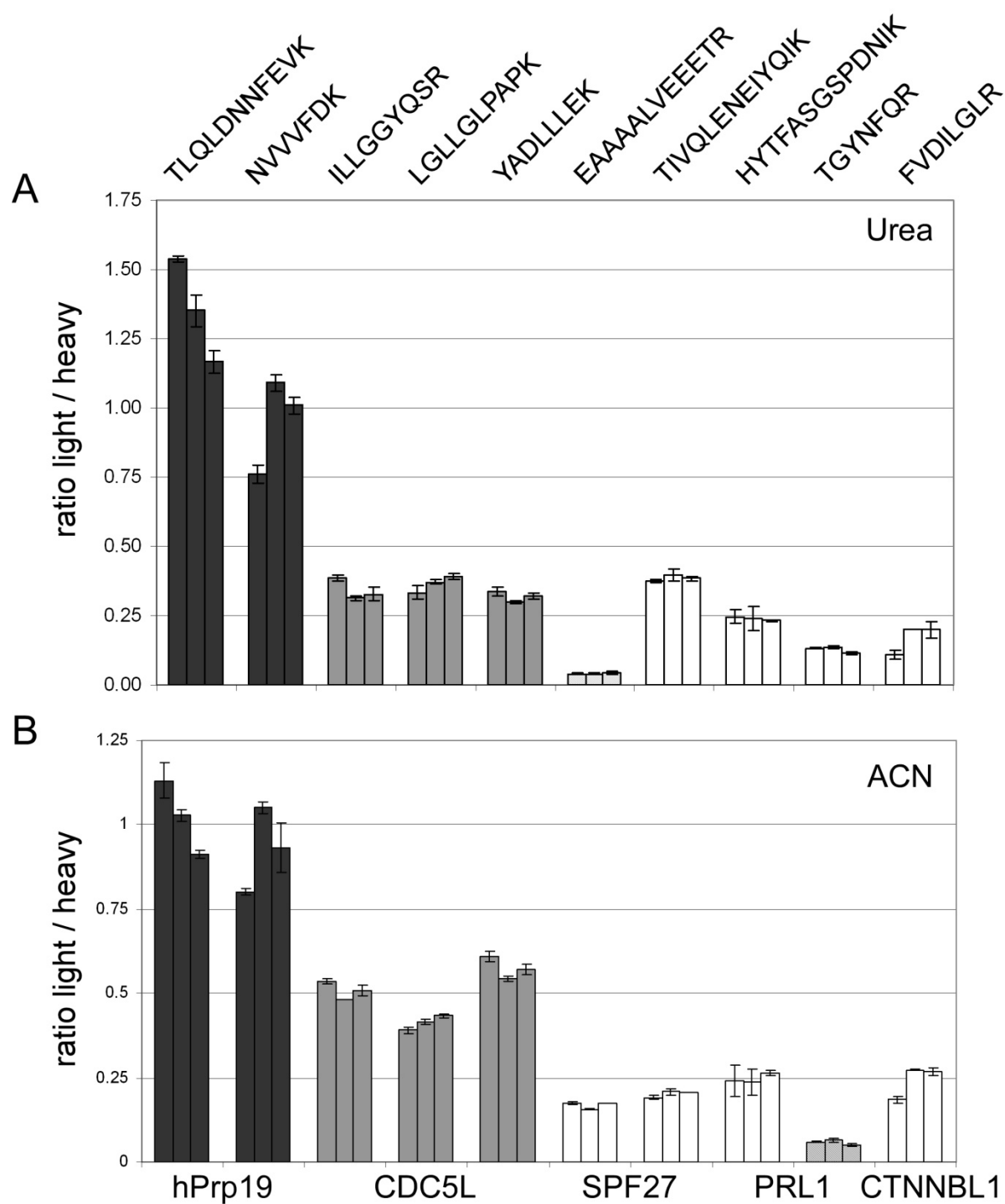


Figure 4.8: Peptide ratios obtained from MRM analysis of the hPrp19/CDC5L complex. Peptide ratios from three replicates of three different MRM transitions for each peptide are plotted in bar diagrams for hydrolysis in the presence of urea (A) or acetonitrile (B). **(A)** Peptide ratios obtained from hydrolysis in urea show no clear protein stoichiometry. **(B)** Peptide ratios obtained from hydrolysis in acetonitrile are more consistent than those obtained from hydrolysis in urea.

Two peptides (EAAAALVEEETR derived from SPF27 after hydrolysis in the presence of urea, Figure 4.8 A; and TGYNFQR derived from PRL1 after hydrolysis in the presence of acetonitrile, Figure 4.8 B) consistently revealed a very low peptide ratio. The low value for this peptide is probably caused by low abundance of the endogenous peptide when compared with the isotopically labeled standard peptide suggesting that the endogenous peptide might be underrepresented. Indeed, both of these peptides were also hardly detectable in MALDI-TOF/TOF and ESI-LC-MS/MS experiments and could thus not be used for calculation of the protein ratio in the previous experiments. The low values are indeed consistent with the presence of a slightly longer form of the peptide EAAAALVEEETR (SPF27) containing a missed cleavage site after hydrolysis in the presence of urea as monitored in the initial analysis (see above). The adjacent protein sequence of this particular SPF27 peptide shows several tryptic C-terminal cleavage sites (EAAAALVEEETR/R/YR/) that might increase the possibility of missed cleavages in urea, but not in acetonitrile (compare peptide ratios for SPF27 in Figure 4.8 A and B). Conversely, the peptide derived from PRL1 (TGYNFQR), which shows a low value after hydrolysis using acetonitrile, does not have any additional adjacent tryptic sites. The nearest tryptic sites are 43 positions in the N-terminal and 27 positions in the C-terminal direction. Thus, the selected peptide sequence is located in a protein region that contains only very few tryptic cleavage sites (Figure 4.9). A very long tryptic peptide containing a missed cleavage might have been generated in the presence of acetonitrile (consistently with the observation in early experiments that longer peptides with missed cleavages are generated preferentially when acetonitrile is used for digestion; see above, Figure 4.2), and its detection might be hampered in ESI and MALDI analyses. As the presence of missed cleavage site-containing peptides causes underrepresentation of the endogenous peptide to be quantified, both these peptides were excluded from the calculation of the protein stoichiometry of the hPrp19/CDC5L complex after hydrolysis in urea and acetonitrile, respectively. Since one of the peptides (TGYNFQR derived from PRL1) represents a proteotypic peptide (Table 4.3), the results for the missed cleavage sites of tryptic peptides in the presence of different denaturing agents highlight the need for thorough evaluation of experimental data before standard peptides for absolute quantification are selected.

PRL1 sequence

```

1
MVEEVQKHSV HTLVFRSLKR THDMFVADNG KPVLDEESH KRKMAIKLRN EYGPVLHMPT SKENLKEKGP
101
QNATDSYVHK QYPANQGQEV EYFVAGTHPY PPGPGVALTA DTKIQRMPS SAAQSLAVAL PLQTKADANR
151
TAPSGSEYRH PGASDRPQPT AMNSIVMETG NTKNSALMAK KAPTMPKPQW HPPWKLYRVI SGHLGWVRCI
201
AVEPGNQWFV TGSADRTIKI WDLASGKCLK SLTGHISTVR GVIIVSTRSPY LFSCGEDKQV KCWDLEYNKV
251
IRHYHGHLA VYGLDLHPTI DVLVTCRDS TARIWDVRTK ASVHTLSGHT NAVATVRCQA AEPQIITGSH
301
351
DTTIRLWDLV AGKTRVTLTN HKKSVRAVVL HPRHYTFASG SPDNIKQWKF PDGSFIQNLS GHNAINTLT
401
VNSDGVLVSG ADNGTMHLWD WRTGYNFQR HAAVQPGSLD SESGIFACAF DQSESRLLLTA EADKTIKVYR
451
501
EDDTATEETH PVSWKPEIIK RKR

```

Figure 4.9: PRL 1 sequence. The peptide TGYNFQR (underlined) is located in a protein region that contains no further tryptic cleavage sites. The nearest tryptic sites are 43 positions in N-terminal and 27 positions in C-terminal direction.

As described above, protein ratios were calculated from average peptide ratios of three technical replicates and three different MRM transitions for each peptide. The MRM experiments after hydrolysis of the hPrp19/CDC5L complex in the presence of urea revealed a stoichiometry of approximately 4:1 for hPrp19 relative to CDC5L and SPF27, 5:1 relative to PRL1, and 9:1 relative to CTNNB1. Hydrolysis in the presence of acetonitrile and subsequent analysis by MRM resulted in a stoichiometry of 2:1 for hPrp19 relative to CDC5L, 5:1 relative to SPF27, and 4:1 relative to PRL1 and CTNNB1 (Table 4.6).

Table 4.6: Relative protein stoichiometries within the hPrp19/CDC5L complex as determined by MRM analyses. Peptide ratios were calculated from MRM signals of endogenous and standard peptides. Average peptide ratios from replicates from three different MRM transitions for each peptide were used to calculate protein ratios. Protein stoichiometries are displayed by the ratio of hPrp19 to the relevant protein showing the stoichiometry of hPrp19 within the hPrp19/CDC5L complex. Values in parentheses were omitted for calculation of the average values. For SPF27 no ratio could be determined for dilution of the standard peptides.

Protein complex [ng] standard peptides [fmol]	70	70	35	17.5	70	70	Average
8M/2M urea							
Protein stoichiometry							
hPrp19/CDC5	3.54	3.37	3.46	3.73	3.09	3.70	3.48 ± 0.236
hPrp19/SPF27	3.32	2.99	5.44	(11.92)	/	/	3.92 ± 1.330
hPrp19/PRL1	5.35	4.80	5.21	6.62	4.37	4.96	5.22 ± 0.767
hPrp19/CTNNB1	10.00	6.78	13.03	9.03	6.39	(3.94)	9.05 ± 2.691
80 % (v/v) acetonitrile							
Protein stoichiometry							
hPrp19/CDC5	1.97	1.96	2.01	1.95	1.87	1.82	1.93 ± 0.071
hPrp19/SPF27	4.69	4.83	5.58	6.11	/	/	5.30 ± 0.665
hPrp19/PRL1	4.31	3.94	3.77	3.37	3.82	3.72	3.82 ± 0.306
hPrp19/CTNNB1	4.24	4.03	3.74	3.59	(1.97)	(2.33)	3.90 ± 0.106

4.1.7 Comparison of results for absolute quantification in determining the Stoichiometry of the hPrp19/CDC5L complex

A comparison of the different MS analysis approaches MALDI-ToF/ToF, LC-ESI-MS, and MRM (Table 4.7 and Figure 4.10) shows that the MRM method yielded the most consistent and reproducible results when judging the experimentally determined peptide and resulting protein ratios from the standard deviation. The large variability in MALDI-ToF/ToF-MS (peak area) or ESI-MS (XIC) experiments could be explained by co-eluting and thus contaminating peptides that cause peak overlap in the chromatograms and spectra as described above.

Although MRM experiments produce consistent values for the single proteins (as a result of the highly specific monitoring of the investigated peptides), the protein ratios obtained for hPrp19 vs. CDC5L still vary with the hydrolysis conditions used (Figure 4.10 A and B). A stoichiometry of 1:1:1:2:4 (CTNNB1/PRL1/SPF27/CDC5L/hPrp19) is calculated for these proteins within the hPrp19/CDC5L complex after hydrolysis in the presence of acetonitrile (Figure 4.10 B and Figure 4.8 B) whereas the values obtained for the complex after hydrolysis in the presence of urea do not allow a clear determination of the stoichiometry (Figure 4.10 A and Figure 4.8 A). As no peptide of AD-002 could be selected for quantitative analysis (see above) AD-002 could not be quantified in this study. Furthermore, Hsp70, which was detected during proteomic analysis of the hPrp19/CDC5L complex (Figure 4.1), could not be quantified because the analysis revealed that Hsp70 is underrepresented in the protein complex analyzed under optimized conditions using standard peptides and LC-ESI-MS and XICs (Table A.4 in the Appendix).

Table 4.7: Summary of the protein stoichiometries obtained from different MS techniques. Peptide ratios were obtained from peak areas, XICs and MRM transitions, respectively (Tables 4.3, 4.4, and 4.5). Protein ratios from different experiments were averaged and listed as the ratio of hPrp19 to the relevant protein showing the stoichiometry of hPrp19 within the hPrp19/CDC5L complex.

	LC-offline MALDI-ToF/ToF-MS (peak area)		LC-online ESI-MS/MS (XIC)		LC-online ESI-MS/MS (MRM)	
	urea	ACN	urea	ACN	urea	ACN
Hydrolysis						
Protein stoichiometry						
hPrp19/CDC5	2.54 ± 0.26	1.82 ± 0.14	3.58 ± 0.19	2.24 ± 0.21	3.48 ± 0.24	1.93 ± 0.07
hPrp19/SPF27	2.53 ± 0.57	4.41 ± 1.19	3.04 ± 0.54	4.81 ± 0.08	3.92 ± 1.33	5.30 ± 0.67
hPrp19/PRL1	5.55 ± 0.98	4.81 ± 0.89	8.97 ± 1.07	5.56 ± 1.63	5.22 ± 0.77	3.82 ± 0.31
hPrp19/CTNNB1	2.29 ± 0.42	2.17 ± 0.56	3.36 ± 0.39	3.97 ± 0.68	9.05 ± 2.69	3.90 ± 0.11

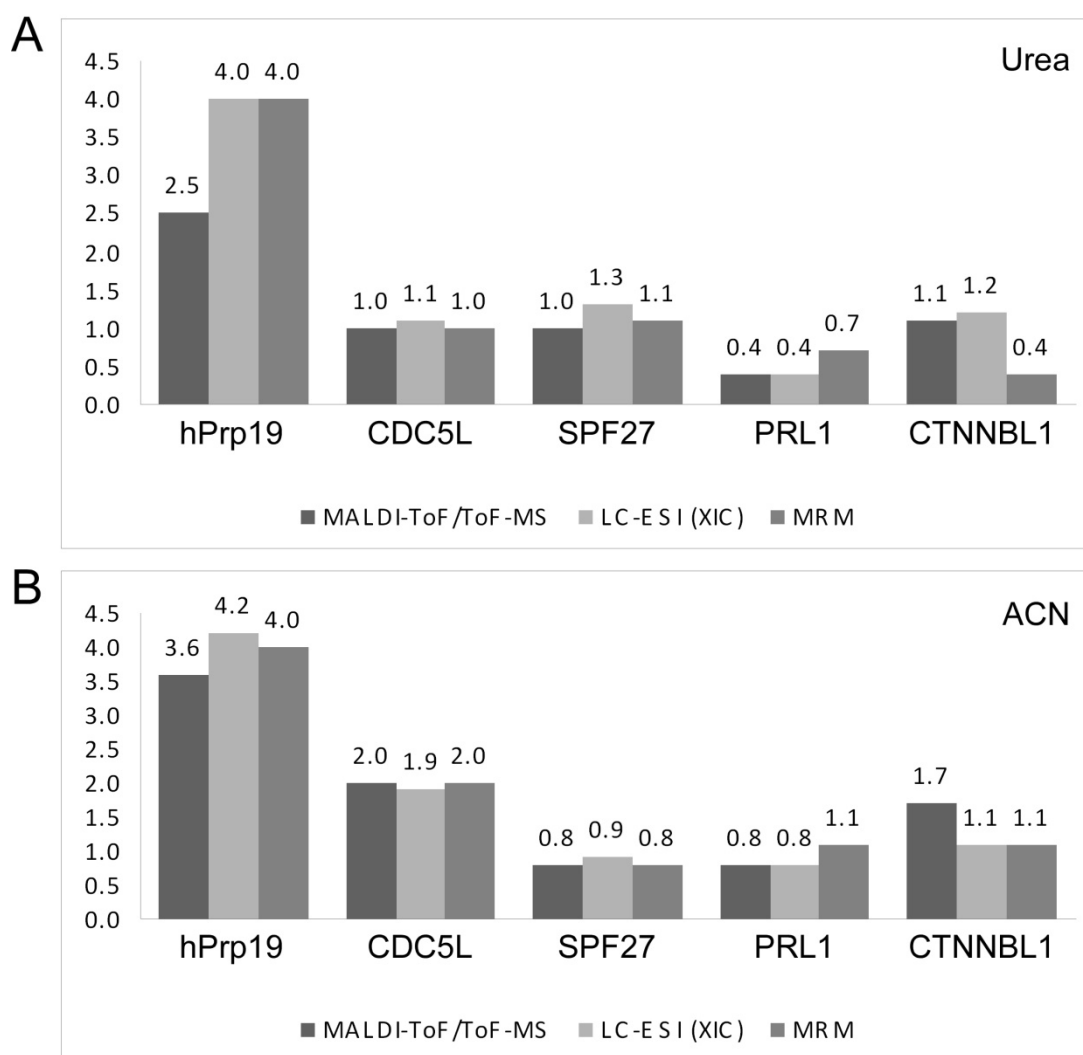


Figure 4.10: Protein stoichiometry within the hPrp19/CDC5L complex. The values are normalized to the number of copies of hPrp19 obtained by MRM analysis. For each protein, the stoichiometry is given for all three MS techniques tested. (A) Protein stoichiometry within the hPrp19/CDC5L complex obtained by hydrolysis in urea. The values obtained do not allow unambiguous determination of the stoichiometries. (B) Protein stoichiometry within the hPrp19/CDC5L complex obtained by hydrolysis in acetonitrile. A stoichiometry of 4:2:1:1:1 (hPrp19/CDC5L/SPF27/PRL1/CTNNBL1) is calculated for these proteins in the hPrp19/CDC5L complex.

In previous studies, the stoichiometry of Prp19p in the yeast nineteen complex (NTC, the yeast homologue of the hPrp19/CDC5L complex) was analyzed by electron microscopy and density gradient centrifugation (Ohi *et al.*, 2005). Ohi *et al.*, 2005 found that Prp19p forms a tetramer and that oligomerization is essential for its function in vivo. In addition, the authors also analyzed the protein-protein interactions of Prp19p and Cef1p (the yeast homologue of CDC5L) by the yeast two-hybrid assay. They found interactions between Prp19p and Cef1p and self-association of Prp19 but no self-interaction of CDC5L. Consistent with a similar study by Grillari *et al.*, 2005, the possibility of dimerization of CDC5L was recently discussed by (Graub *et al.*, 2008). Using cross-linking with glutaraldehyde and subsequent immunoblotting these authors demonstrated the presence of CDC5L homodimers. By

mutation of CDC5L phosphorylation sites, they found that homodimerization is independent of phosphorylation, suggesting that CDC5L might be present as a homodimer. Our results are consistent with extensive biochemical studies characterizing the molecular architecture of the Prp19/CDC5L (Grote *et al.*, 2010). In this study, four copies of Prp19 and one copy of the proteins PRL1, AD-002, and SPF27 were found by [¹⁴C]-iodoacetamide labeling and Sypro-Ruby staining of the denatured proteins of the complex and subsequent quantification of radioactivity and fluorescence, respectively. However, as observed here in the presence of urea, CDC5L was only present in one copy.

4.2 Relative quantification by iTRAQ-labeling of in-gel digested proteins

A robust mass spectrometry-based quantification of spliceosomal proteins for characterization of the spliceosome's protein composition at different intermediate states is required. Various methods for MS-based relative quantification exist (for an overview see Table 2.1 in the Introduction). iTRAQ reagents are chemical labels that specifically label amino termini and lysine side chains of peptides. iTRAQ quantification has some clear advantages: (i) iTRAQ reagents are multiplexing (i.e. up to four or eight samples can be compared in one experiment). (ii) As the iTRAQ reagents are amine specific, the peptides' N-termini and lysine side chains are labeled thus providing many quantification data points per protein. (iii) Differentially labeled peptides are isobaric and the intensity in MS is thus enhanced. (iv) The different mass tags are completely cleaved during fragmentation leading to an enhanced fragment ion intensity. (v) iTRAQ-labeling offers the opportunity to create an internal standard by mixing aliquots of all samples to be compared, allowing comparison of even more samples than four (four-plex iTRAQ) or eight (eight-plex iTRAQ) samples.

4.2.1 Optimization of iTRAQ-labeling of in-gel digested proteins

The ABSciex iTRAQ labeling protocol comprises in-solution hydrolysis of the proteins, subsequent labeling of the peptides with iTRAQ reagents, and final removal of excess reagent by strong cation exchange chromatography (SCX). Here, we optimized iTRAQ labeling for in-gel digested proteins. This alteration has the major advantage that sample complexity is reduced before hydrolysis and even small quantities can be quantified as sample loss during SCX is avoided. For this purpose, the buffers used during in-gel hydrolysis of the proteins and the amount of iTRAQ reagents have been adjusted to the in-gel hydrolysis protocol. An additional reaction step to quench the excess of iTRAQ reagents by addition of glycine was included. Furthermore, an internal standard was prepared by

mixing equal amounts of all samples to be compared and labeling with one of the iTRAQ reagents. This allows controlling of the labeling reaction because the reporter ion intensity of the internal standard is related to the intensities of the reporter ions of the differently labeled samples. The optimized workflow thus comprises (i) the separation of the samples to be compared by gel electrophoresis, (ii) cutting entire gel lanes into gel slices of equal size, (iii) in-gel digestion of the proteins and extraction of the peptides, (iv) preparation of an internal standard from samples to be compared (i.e. from peptides extracted from gel slices of the same molecular weight region within the different samples), (v) iTRAQ-labeling of the internal standard and the samples with the different iTRAQ reagents and subsequent quenching of iTRAQ reagent excess, (vi) pooling of the samples and their respective internal standard, (vii) LC-MS/MS analysis, and (viii) identification (database search) and quantification of the peptides and finally the proteins (Figure 4.11).

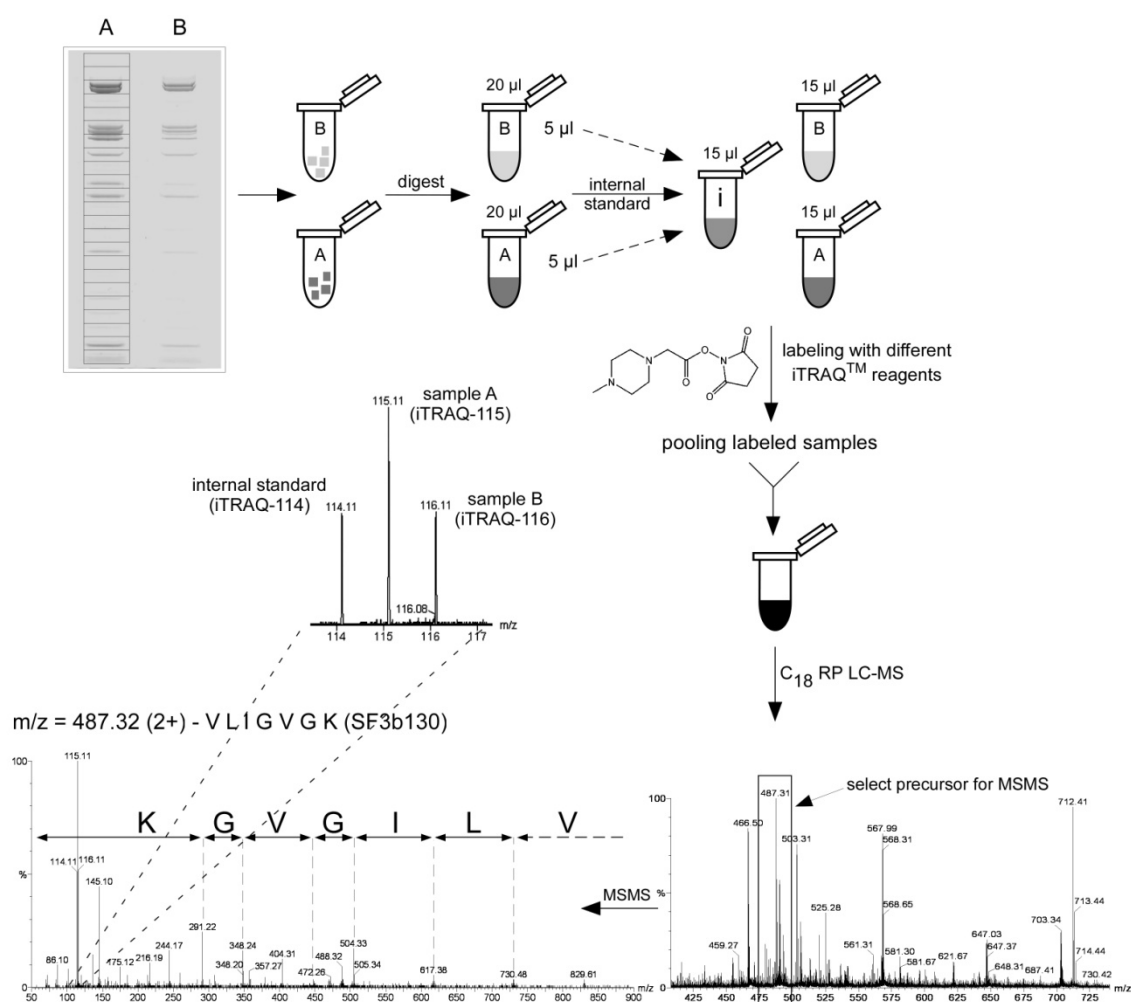


Figure 4.11: Workflow for iTRAQ-labeling of in-gel digested proteins. Entire gel lanes of samples to be analyzed and quantified are cut into gel slices of equal size and gel slices are manually cut into smaller pieces. Proteins are digested with trypsin within the gel and generated peptides are extracted. Extracted peptides are re-dissolved in 20 μl TEAB and an internal standard is prepared by pooling 5 μl of each sample. The internal standard and the samples to be compared are labeled with different iTRAQ reagents. After pooling, the samples are analyzed by LC-MS/MS and quantification is done by comparing the peak areas of individual reporter ions.

4.2.2 Relative quantification of different amounts of spliceosomal tri-snRNP proteins – a feasibility study

To validate the established iTRAQ workflow, different amounts of spliceosomal tri-snRNP proteins were quantified using iTRAQ labeling. To this end, 5.0 and 2.5 μg of purified human tri-snRNP were separated by gel electrophoresis (Figure 4.12) and entire gel lanes were cut into gel slices of equal size. In addition, an empty gel lane (blank) was cut and processed together with the other samples to show accuracy of the established iTRAQ protocol.

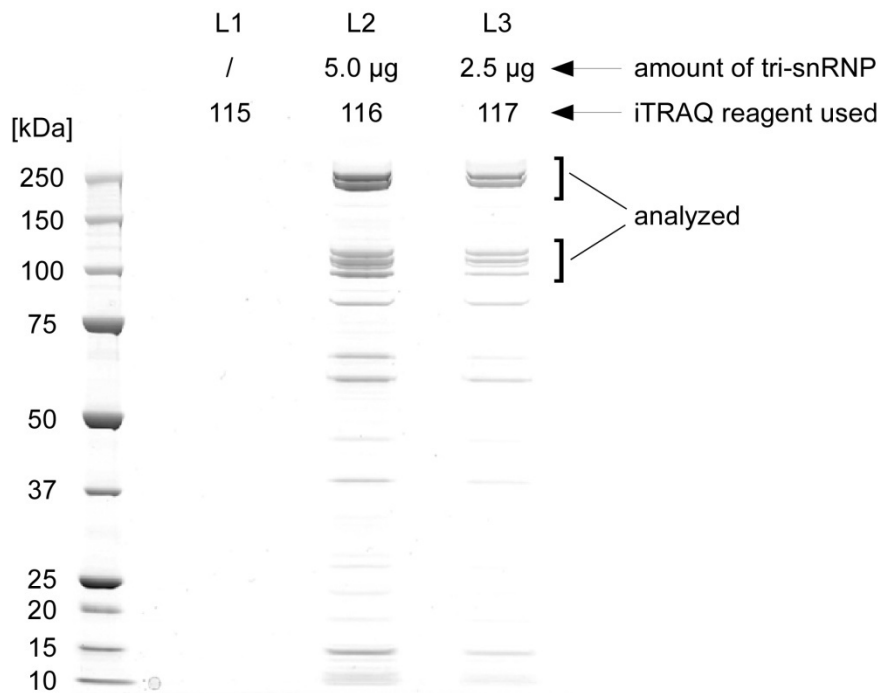


Figure 4.12: Separation of different amounts of spliceosomal tri-snRNP proteins by gel electrophoresis. 5.0 μg and 2.5 μg tri-snRNP were separated by gel electrophoresis and entire gel lanes were cut into slices of equal size. An empty gel lane (L1) was cut and processed with the other samples.

The proteins from two gel slices at different molecular weight (approximately 250 kDa and 100 kDa), showing proper stained protein bands at both concentrations, were hydrolyzed with trypsin in-gel. An internal standard was prepared from the samples to be compared (i.e. L1, L2 (5.0 μg tri-snRNP) and L3 (2.5 μg tri-snRNP)) and the peptides of the different samples and the internal standard were labeled separately with iTRAQ reagents. The internal standard was labeled with iTRAQ reagent 114 and the different samples were labeled with iTRAQ reagents 115-117 (see also Figure 4.12). Excess iTRAQ reagents were quenched by adding glycine. Samples to be compared (i.e. L1-L3) and their corresponding internal standard were then pooled and subsequently analyzed by LC-MS/MS. Quantification was performed by comparing the peak areas of generated reporter ions after fragmentation. Figure 4.13 shows a typical MS/MS spectrum.

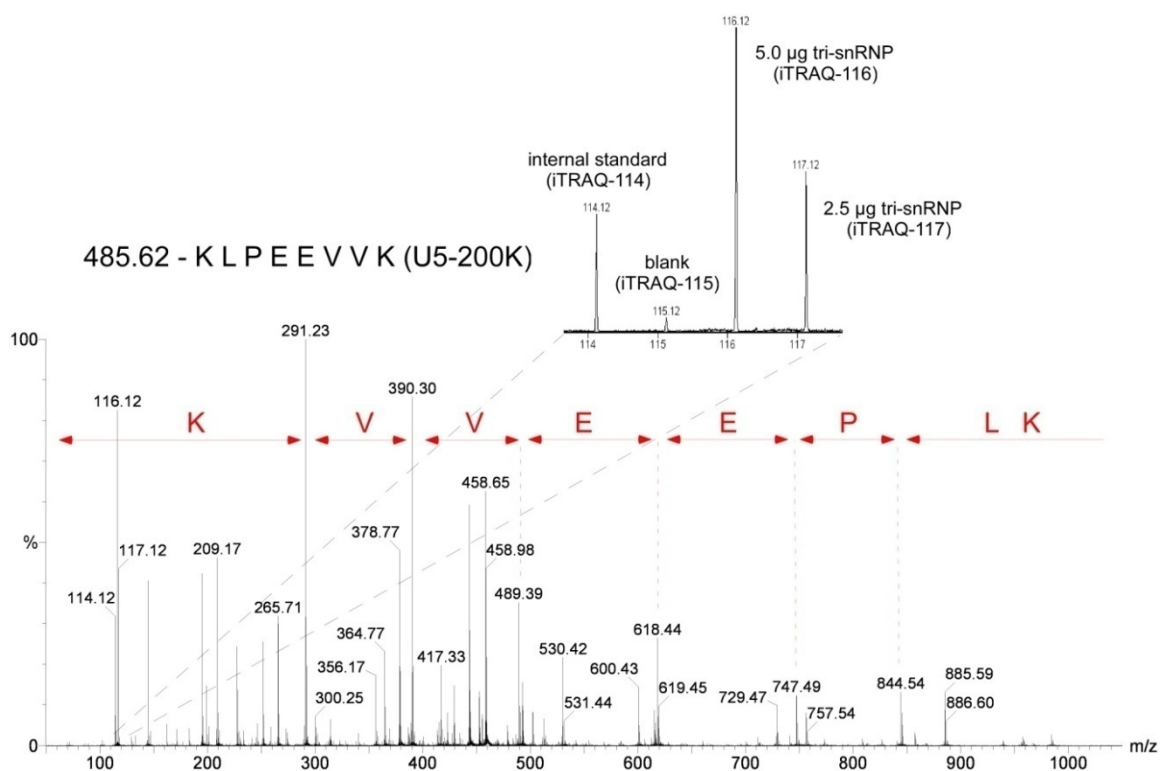


Figure 4.13: Example MS/MS spectrum of an iTRAQ labeled peptide. During MS/MS KLPEEVVK (U5-200K) was sequenced and the reporter ions were released. The reporter region (m/z 114-117) is magnified and shows the different reporter ions that could be used for quantification.

The different tri-snRNP specific proteins within the two molecular weight regions were identified. For quantification, the protein ratios are calculated relatively to the internal standard, i.e. by dividing the peak area of reporter ions of the blank sample (iTRAQ-115), 5.0 µg (iTRAQ-116), and 2.5 µg (iTRAQ-117) tri-snRNP, respectively, by the peak area of the internal standard (iTRAQ-114). In addition, the protein ratio comparing the different amounts of tri-snRNP was also calculated (iTRAQ-116/iTRAQ-117).

A protein ratio of approximately 2.5 for the different amounts of tri-snRNP (iTRAQ-116/iTRAQ-117) was obtained for almost all proteins. The protein ratio of the blank sample (iTRAQ-115/iTRAQ-114) was in all cases approximately 0.1 or lower (Table 4.8). The obtained protein ratios are plotted in bar diagrams to visualize the relative protein amounts within the different gel lanes (Figure 4.14). As chemical labeling always risks being incomplete, the labeling efficiency (i.e. the percentage of labeled peptides) was calculated for all identified proteins. For this purpose, the obtained precursor masses (MS) and fragment ion masses (MS/MS) were searched against the database allowing for iTRAQ-labels as variable modifications. This yields labeled and non-labeled peptides and the labeling efficiency could be calculated from the number of identified iTRAQ-labeled peptides and the total number of identified peptides. In pilot experiments, the labeling efficiency was found to

be one of the major issues for reliable quantification, which was only achieved for labeling efficiencies > 90 %. For this proof of principle, a labeling efficiency of 92 % or higher was obtained (Table 4.8).

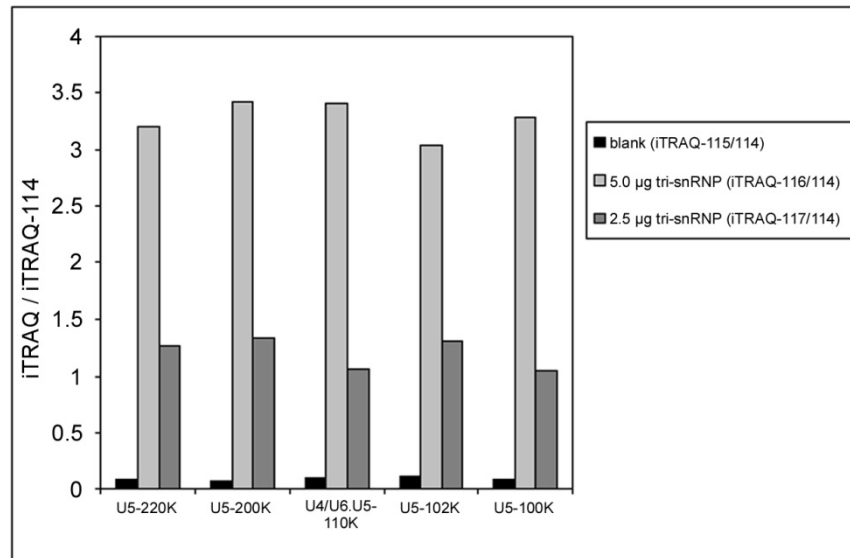


Figure 4.14: iTRAQ protein ratios for various identified tri-snRNP proteins. Ratios were calculated by dividing the peak area of reporter ions of the blank sample (iTRAQ-115), 5.0 µg (iTRAQ-116), and 2.5 µg (iTRAQ-117) tri-snRNP, respectively, by the peak area of the internal standard (iTRAQ-114).

Table 4.8: Protein ratios for the relative quantification of different amounts of spliceosomal tri-snRNP proteins. iTRAQ ratios relative to the internal standard and one iTRAQ ratio comparing the different tri-snRNP amounts were calculated. The labeling efficiency is indicated in percent.

Protein	iTRAQ ratios				labeling efficiency [%]
	115/114	116/114	117/114	116/117	
U5-220K	0.085	3.192	1.269	2.515	95.27
U5-200K	0.073	3.413	1.341	2.545	94.15
U4/U6.U5-110K	0.104	3.407	1.057	3.223	100.00
U5-102K	0.106	3.036	1.300	2.335	97.27
U5-100K	0.092	3.283	1.048	3.133	92.86

4.3 Relative quantification of spliceosomal B and C complexes – a comparative study

The spliceosome is a highly dynamic protein-RNA machinery. During pre-mRNA splicing, it passes through different functional states that differ in their protein and RNA composition. In this study, the proteomes of the pre-catalytic and the catalytically active spliceosome (i.e. the spliceosomal B and C complexes) were analyzed by relative quantification to show differences in their protein abundances. To this end, affinity purified spliceosomal B and C complexes were labeled with iTRAQ reagents and analyzed by liquid chromatography-coupled mass spectrometry. In addition, B and C complexes prepared from metabolically labeled nuclear extracts were affinity purified and quantified by mass spectrometry (SILAC quantification). Protein ratios showing protein abundances within B and C complexes obtained by iTRAQ and SILAC quantification were then evaluated and compared. In addition, these results based on stable isotope labeling made an evaluation of semi-quantitative spectral count from proteomic analysis of the two spliceosomal complexes possible.

4.3.1 Purification of spliceosomal B and C complexes

Spliceosomal B and C complexes were purified from HeLa nuclear extract using MS2-tagged radioactively labeled PM5 pre-mRNA as described in previous studies (Bessonov *et al.*, 2008; Deckert *et al.*, 2006). Briefly, radioactively labeled PM5 pre-mRNA was pre-incubated with MS2-MBP fusion protein and spliceosomal complexes were allowed to assemble in nuclear extract *in vitro*. B complexes were assembled by kinetic control for 6 minutes, whereas C complexes were allowed to accumulate for 180 min. Assembled complexes were separated by glycerol gradient centrifugation and B or C complexes were affinity purified from fractions that contained the corresponding complex. For proteomic analysis (spectral count) and iTRAQ quantification, equal amounts of affinity purified B and C complexes were separated by gel electrophoresis (Figure 4.15 A). For SILAC quantification, B and C complexes were assembled in metabolically labeled light (C complex) and heavy (B complex) nuclear extracts, respectively. Equal amounts of B and C complexes were then pooled and the proteins were separated by gel electrophoresis (Figure 4.15 B). B and C complexes for iTRAQ analysis were purified and provided by S. Bessonov. Purification of B and C complexes for SILAC quantification was performed by J. Deckert.

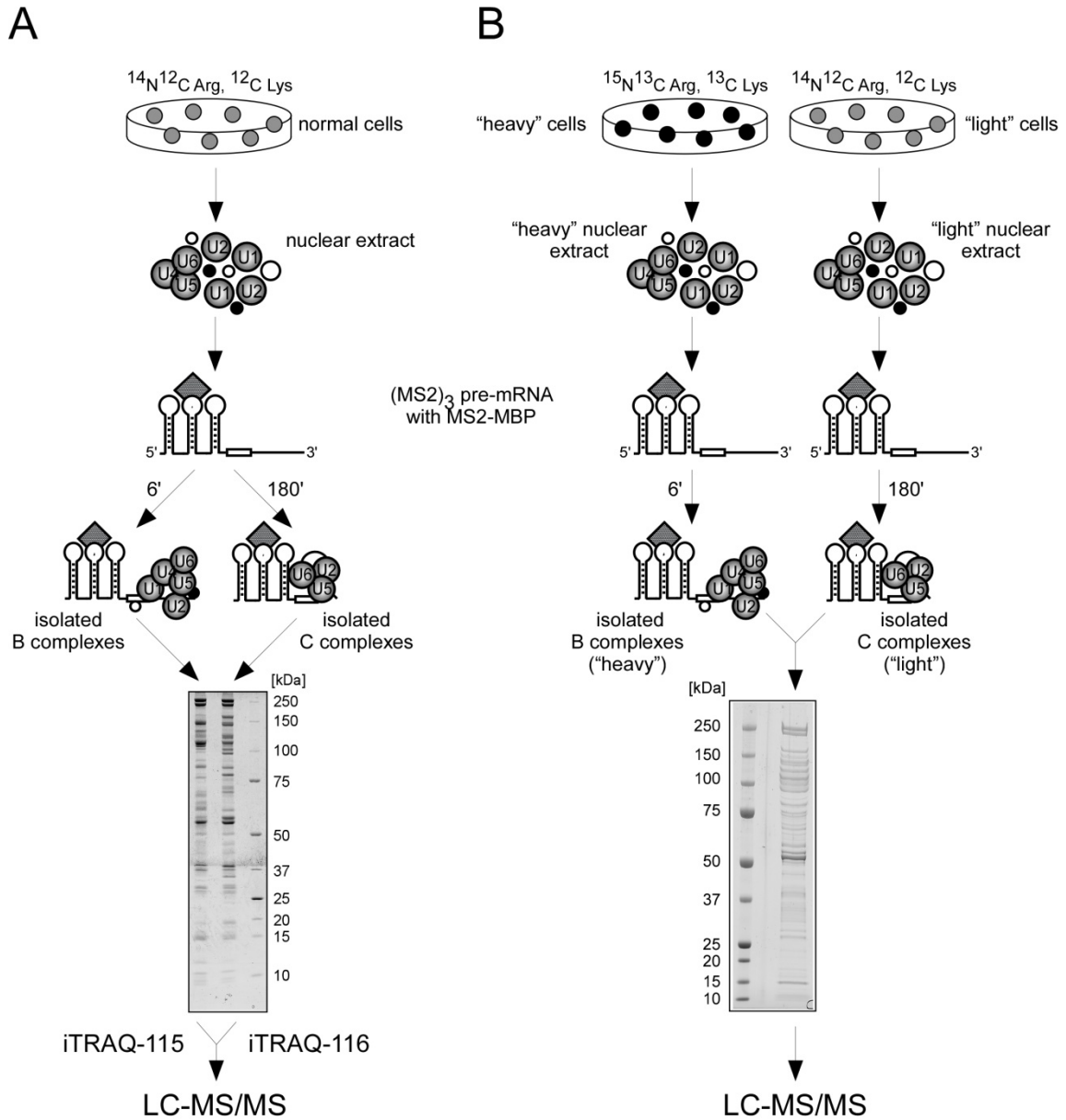


Figure 4.15: Purification of spliceosomal B and C complexes for iTRAQ and SILAC quantification. (A) B and C complexes were allowed to assemble on MS2-tagged pre-mRNA for 6 and 180 min, respectively. The complexes were isolated by gradient centrifugation and affinity purification and proteins were separated by gel electrophoresis. After in-gel digestion of the proteins, peptides generated from B complex were labeled with iTRAQ reagent 115 and peptides generated from C complex proteins were labeled with iTRAQ reagent 116. After pooling the samples were analyzed by LC-MS/MS. **(B)** B and C complexes were purified from light (C complex) and heavy (B complex) nuclear extract, respectively. Isolated complexes were pooled in equal amounts, the proteins were separated by gel electrophoresis and generated peptides were analyzed by LC-MS/MS.

4.3.2 Proteomic analysis of spliceosomal B and C complexes – spectral count

The proteomes of spliceosomal B and C complexes have been analyzed in previous studies (Bessonov *et al.*, 2008; Deckert *et al.*, 2006). Approximately 150 proteins were identified within the two complexes. From these, several are snRNP specific whereas others are non-snRNP specific proteins. Figure 4.16 shows the proteomes of B and C complexes separated by gel electrophoresis.

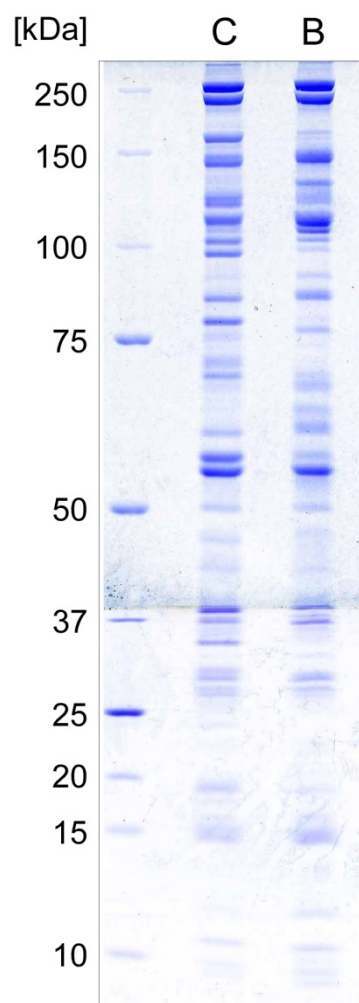


Figure 4.16: The proteomes of spliceosomal B and C complexes. Equal amounts of B and C complexes were loaded onto the gel and the complexes' proteins were separated by gel electrophoresis.

A comparison of the protein composition of the precatalytic (B complex) versus the catalytic (C complex) spliceosome has been performed by Bessonov *et al.*, 2008. For this purpose, the absolute number of peptides was used to determine protein abundances within B and C complexes. According to their abundance within the complexes, several snRNP specific and non-snRNP specific proteins have been assigned to be specific for B or C complexes.

The correlation between the relative protein abundance and the number of tandem MS spectra (spectral count), the number of identified peptides (peptide count), and the obtained sequence coverage has recently been analyzed (Liu *et al.*, 2004). A linear correlation was found between the relative protein abundance and the number of acquired MS/MS spectra, but not between the protein amount and the number of identified peptides or the sequence coverage. The data set analyzed in Bessonov *et al.*, 2008 was therefore reanalyzed using the software Scaffold 2 to achieve unweighted spectral count for relative comparison of spliceosomal B and C complexes from the proteomic analysis of the two complexes (Table 4.9).

Table 4.9: Spectral count for proteomic analysis of spliceosomal B and C complexes. The number of acquired tandem MS spectra for two independent purifications of B and C complexes (B1, B2, C1, and C2, respectively) is given for all spliceosomal proteins identified in Bessonov *et al.*, 2008. The average number of spectra from the two replicates and the resulting protein ratio showing the protein abundances within B and C complexes (B/C) were calculated. The proteins are grouped according to Bessonov *et al.*, 2008. For proteins identified solely in B complexes no protein ratio could be calculated ("/"). For proteins that were only or predominantly identified in C complexes the calculated protein ratio is 0 or close to 0.

Protein	MW [kDa]	accession no.	B1	B2	B	C1	C2	C	B / C
U1 snRNP									
U1-A	31.3	gi 4759156	6	4	5	0	0	0	/
U1-C	17.4	gi 4507127	1	0	0.5	0	0	0	/
U1-70K	51.6	gi 29568103	4	3	3.5	0	0	0	/
17S U2 snRNP									
U2A'	28.4	gi 50593002	12	30	21	17	17	17	1.24
U2B''	25.4	gi 4507123	8	12	10	8	7	7.5	1.33
SF3a120	88.9	gi 5032087	33	31	32	5	13	9	3.56
SF3a66	49.3	gi 21361376	17	2	9.5	1	2	1.5	6.33
SF3a60	58.5	gi 5803167	30	18	24	7	2	4.5	5.33
SF3b155	145.8	gi 54112117	89	77	83	18	15	16.5	5.03
SF3b145	100.2	gi 55749531	41	25	33	7	4	5.5	6.00
SF3b130	135.5	gi 54112121	182	93	137.5	42	19	30.5	4.51
SF3b49	44.4	gi 5032069	1	3	2	1	1	1	2.00
SF3b14a (p14)	14.6	gi 7706326	9	12	10.5	6	7	6.5	1.62
SF3b14b	12.4	gi 14249398	13	7	10	7	1	4	2.50
SF3b10	10.1	gi 13775200	3	1	2	0	0	0	/
17S U2 related									
U2AF65	53.5	gi 6005926	1	2	1.5	0	0	0	/
U2AF35	27.9	gi 5803207	1	1	1	0	0	0	/
hPRP43	90.9	gi 68509926	36	26	31	19	19	19	1.63
SPF45	45.0	gi 14249678	6	3	4.5	0	0	0	/
SR140	118.2	gi 122937227	8	2	5	0	0	0	/
CHERP	100.0	gi 119226260	2	6	4	0	0	0	/
SF3b125	103.0	gi 45446747	10	8	9	0	0	0	/
U5 snRNP									
220K	273.3	gi 3661610	227	184	205.5	158	80	119	1.73
200K	244.5	gi 45861372	282	144	213	220	93	156.5	1.36
116K	109.4	gi 41152056	109	94	101.5	100	79	89.5	1.13
40K	39.3	gi 4758560	26	12	19	19	5	12	1.58
102K	106.9	gi 40807485	138	39	88.5	36	8	22	4.02
15K	16.8	gi 5729802	8	4	6	0	0	0	/
100K	95.6	gi 41327771	59	17	38	29	8	18.5	2.05
52K	37.6	gi 5174409	4	7	5.5	3	1	2	2.75
U4/U6 snRNP									
90K	77.6	gi 4758556	54	33	43.5	7	3	5	8.70
60K	58.4	gi 45861374	45	25	35	6	1	3.5	10.00
20K	20.0	gi 5454154	10	17	13.5	3	2	2.5	5.40
61K	55.4	gi 40254869	47	15	31	8	0	4	7.75
15.5K	14.2	gi 4826860	7	10	8.5	2	0	1	8.50
U4/U6.U5 snRNP									
110K	90.2	gi 13926068	44	32	38	10	4	7	5.43
65K	65.4	gi 56550051	30	10	20	18	5	11.5	1.74
hPRP38	37.5	gi 24762236	19	8	13.5	4	0	2	6.75
TFIP11	96.8	gi 8393259	6	4	5	6	5	5.5	0.91

Protein	MW [kDa]	accession no.	B1	B2	B	C1	C2	C	B / C
L_{Sm} Proteins									
LSM 2	10.8	gi 10863977	8	8	8	1	0	0.5	16.00
LSM 3	11.8	gi 7657315	2	1	1.5	1	0	0.5	3.00
LSM 4	15.4	gi 6912486	9	4	6.5	0	0	0	/
LSM 6	9.1	gi 5919153	9	14	11.5	1	0	0.5	23.00
LSM 7	11.6	gi 7706423	2	4	3	0	0	0	/
LSM 8	10.4	gi 7706425	6	7	6.5	1	0	0.5	13.00
S_m Proteins									
B	24.6	gi 4507125	15	20	17.5	15	14	14.5	1.21
D1	13.3	gi 5902102	16	9	12.5	17	9	13	0.96
D2	13.5	gi 29294624	48	21	34.5	42	17	29.5	1.17
D3	13.9	gi 4759160	30	11	20.5	38	12	25	0.82
E	10.8	gi 4507129	9	8	8.5	9	9	9	0.94
F	9.7	gi 4507131	16	2	9	8	2	5	1.80
G	8.5	gi 4507133	5	6	5.5	2	5	3.5	1.57
hPRP19/CDC5L complex									
hPrp19	55.2	gi 7657381	56	14	35	85	46	65.5	0.53
CDC5L	92.2	gi 11067747	36	32	34	83	86	84.5	0.40
SPF27	21.5	gi 5031653	8	9	8.5	9	25	17	0.50
PRL1	57.2	gi 4505895	22	9	15.5	36	26	31	0.50
Hsp70	70.4	gi 5729877	3	3	3	14	14	14	0.21
AD-002	26.6	gi 7705475	1	4	2.5	4	9	6.5	0.38
CTNBL1	65.1	gi 18644734	7	4	5.5	5	2	3.5	1.57
Npw38BP	70.0	gi 7706501	8	7	7.5	0	0	0	/
Npw38	30.5	gi 74735456	0	2	1	0	0	0	/
hPRP19/CDC5L related									
hSYF1	100.0	gi 55770906	22	31	26.5	65	66	65.5	0.40
CRNKL1	100.6	gi 30795220	52	24	38	128	75	101.5	0.37
hlsy1	33.0	gi 20149304	4	2	3	19	16	17.5	0.17
SKIP	51.1	gi 6912676	43	12	27.5	65	46	55.5	0.50
RBM22	46.9	gi 8922328	7	12	9.5	31	28	29.5	0.32
Cyp-E	33.4	gi 5174637	4	6	5	8	17	12.5	0.40
PPIL1	18.2	gi 7706339	0	0	0	33	15	24	0.00
KIAA0560	171.3	gi 38788372	30	17	23.5	107	72	89.5	0.26
G10	17.0	gi 32171175	6	7	6.5	15	18	16.5	0.39
hRES complex proteins									
SNIP1	45.8	gi 21314720	10	5	7.5	9	5	7	1.07
MGC12135	70.5	gi 14249338	13	13	13	10	10	10	1.30
CGI-79	39.7	gi 4929627	4	0	2	4	1	2.5	0.80
Proteins recruited to A complex									
RBM39	59.4	gi 4757926	27	9	18	6	1	3.5	5.14
p68 (DDX5)	69.2	gi 4758138	13	4	8.5	6	2	4	2.13
ELAV-like 1 (HuR)	36.1	gi 38201714	60	24	42	8	9	8.5	4.94
p72/DDX17	80.5	gi 3122595	13	3	8	4	0	2	4.00
Proteins recruited to B complex									
MFAP1	51.9	gi 50726968	38	15	26.5	6	2	4	6.63
RED	65.6	gi 10835234	41	16	28.5	2	0	1	28.50
hSmu-1	57.5	gi 8922679	61	31	46	12	0	6	7.67
PPIL2	59.5	gi 7657473	15	6	10.5	9	10	9.5	1.11
hPRP2 (DDX16)	119.2	gi 4503293	11	11	11	17	11	14	0.79
hPRP4-Kinase	117.1	gi 89276756	24	1	12.5	33	1	17	0.74

Protein	MW [kDa]	accession no.	B1	B2	B	C1	C2	C	B / C
THRAP3	108.6	gi 4827040	26	12	19	9	1	5	3.80
PABP1	70.5	gi 46367787	11	5	8	7	4	5.5	1.45
SKIV2L2	117.8	gi 39930353	6	3	4.5	6	8	7	0.64
PABPN1	32.6	gi 4758876	3	1	2	2	1	1.5	1.33
RNF113A	38.8	gi 5902158	1	5	3	3	1	2	1.50
NY-CO-10	53.8	gi 64276486	5	3	4	2	7	4.5	0.89
KIAA1604	105.5	gi 55749769	4	12	8	34	29	31.5	0.25
hsp27	22.8	gi 4504517	2	1	1.5	1	4	2.5	0.60
GCFC	104.7	gi 22035565	0	0	0	2	2	2	0.00
UBL5	8.5	gi 13236510	4	3	3.5	0	0	0	/
CCDC16	42.0	gi 49472814	2	3	2.5	4	1	2.5	1.00
CCDC12	19.2	gi 21389497	1	6	3.5	3	8	5.5	0.64
HsKin17	45.2	gi 13124883	7	7	7	1	0	0.5	14.00
Step 2 factors									
hPRP22	139.3	gi 4826690	4	2	3	107	68	87.5	0.03
hPRP18	39.9	gi 4506123	0	0	0	2	2	2	0.00
hPRP17	65.5	gi 7706657	8	6	7	30	25	27.5	0.25
hPRP16	140.5	gi 17999539	0	0	0	3	1	2	0.00
hSLU7	68.4	gi 27477111	0	0	0	44	17	30.5	0.00
Proteins recruited to C complex									
Abstrakt	69.8	gi 21071032	1	0	0.5	64	54	59	0.01
GCIP p29	28.7	gi 46371998	0	0	0	16	18	17	0.00
DDX35	78.9	gi 20544129	3	0	1.5	22	20	21	0.07
Q9BRR8	103.3	gi 74732921	0	0	0	6	10	8	0.00
c19orf29 (NY-REN-24)	88.6	gi 126723149	1	0	0.5	52	16	34	0.01
PPIase-like 3b	18.6	gi 19557636	0	0	0	10	6	8	0.00
PPWD1	73.6	gi 24308049	0	0	0	33	15	24	0.00
MORG1	34.3	gi 153791298	1	0	0.5	2	9	5.5	0.09
FRG1	29.2	gi 4758404	0	0	0	2	8	5	0.00
NOSIP	33.2	gi 7705716	0	0	0	2	7	4.5	0.00
GPKOW	52.1	gi 15811782	0	1	0.5	4	1	2.5	0.20
C1orf55	39.3	gi 148664216	0	0	0	24	8	16	0.00
FAM32A	13.1	gi 7661696	0	0	0	4	3	3.5	0.00
RACK1 (GNB2L1)	35.1	gi 5174447	0	0	0	0	1	0.5	0.00
Tip-49	50.2	gi 4506753	0	0	0	3	1	2	0.00
Potential C complex specific proteins									
PPIG	88.5	gi 42560244	1	0	0.5	8	3	5.5	0.09
FAM50A	40.1	gi 4758220	0	0	0	17	6	11.5	0.00
FAM50B	38.6	gi 6912326	0	0	0	12	0	6	0.00
C9orf78	33.7	gi 7706557	0	0	0	6	4	5	0.00
C10orf4	37.5	gi 24432067	0	0	0	3	5	4	0.00
CXorf56	25.6	gi 11545813	0	0	0	17	16	16.5	0.00
DGCR14	52.4	gi 13027630	0	0	0	5	11	8	0.00
CCDC130	44.7	gi 13540614	0	0	0	4	1	2.5	0.00
TOE1	56.4	gi 156564398	0	0	0	1	0	0.5	0.00
NKAP	47.0	gi 13375676	0	0	0	7	2	4.5	0.00
ZCCHC10	18.4	gi 8923106	0	0	0	10	3	6.5	0.00
CDK10	35.4	gi 16950647	0	0	0	7	0	3.5	0.00
TTC14	88.2	gi 33457330	0	0	0	9	4	6.5	0.00
WDR70	73.2	gi 8922301	0	0	0	1	2	1.5	0.00
NFKBIL1	43.1	gi 26787991	0	0	0	0	3	1.5	0.00

Protein	MW [kDa]	accession no.	B1	B2	B	C1	C2	C	B / C
EJC/mRNP									
eIF4A3	46.9	gi 7661920	11	12	11.5	51	24	37.5	0.31
Magoh	17.2	gi 4505087	1	2	1.5	11	4	7.5	0.20
Y14	19.9	gi 4826972	2	2	2	8	5	6.5	0.31
Pinin	81.6	gi 33356174	5	1	3	2	2	2	1.50
RNPS1	34.2	gi 6857826	4	0	2	4	0	2	1.00
Acinus	151.8	gi 7662238	23	14	18.5	15	9	12	1.54
SAP18	17.4	gi 5032067	4	3	3.5	1	0	0.5	7.00
Aly/REF (THOC4)	26.9	gi 55770864	5	0	2.5	4	2	3	0.83
UAP56	49.1	gi 18375623	9	1	5	7	4	5.5	0.91
TREX									
THOC1	75.6	gi 154448890	3	5	4	2	5	3.5	1.14
THOC2	169.6	gi 125656165	1	1	1	3	8	5.5	0.18
THOC3	38.8	gi 14150171	0	2	1	0	0	0	/
KIAA0983 (THOC5)	78.4	gi 50959110	7	3	5	5	4	4.5	1.11
WDR58 (THOC6)	37.4	gi 31543164	0	0	0	2	4	3	0.00
pre-mRNA / mRNA binding proteins									
CBP20	18.0	gi 110349727	15	7	11	23	10	16.5	0.67
CBP80	91.8	gi 4505343	69	26	47.5	67	36	51.5	0.92
NF45	43.0	gi 24234747	6	8	7	2	4	3	2.33
ZC3H18	104.0	gi 31377595	2	4	3	2	3	2.5	1.20
YB-1	35.9	gi 34098946	12	1	6.5	7	5	6	1.08
ELG	38.9	gi 8923771	6	5	5.5	7	6	6.5	0.85
DDX3	73.3	gi 87196351	8	3	5.5	2	0	1	5.50
ASR2B	100.0	gi 33383233	14	11	12.5	14	5	9.5	1.32
BCLAF1	107.2	gi 7661958	20	3	11.5	3	0	1.5	7.67
RBM7	30.5	gi 4503293	1	0	0.5	2	4	3	0.17
HSP70	70.0	gi 5123454	3	3	3	14	14	14	0.21
Miscellaneous proteins									
BAG2	23.4	gi 4757834	0	1	0.5	1	2	1.5	0.33
RBBP6	197.2	gi 33620716	1	0	0.5	1	0	0.5	1.00
RBM42	50.3	gi 21359951	1	0	0.5	1	0	0.5	1.00
SR related proteins									
SRm160	102.5	gi 42542379	1	0	0.5	1	0	0.5	1.00
SRm300	300.0	gi 4759098	0	7	3.5	6	29	17.5	0.20
SR proteins									
SF2/ASF	27.8	gi 5902076	11	21	16	21	15	18	0.89
9G8	27.4	gi 72534660	44	25	34.5	36	11	23.5	1.47
SRp20	19.4	gi 4506901	10	2	6	8	4	6	1.00
SRp30c	25.5	gi 4506903	10	11	10.5	10	15	12.5	0.84
SRp38	31.3	gi 5730079	22	9	15.5	34	15	24.5	0.63
SRp40	31.3	gi 3929378	22	5	13.5	15	0	7.5	1.80
SRp46	31.2	gi 15055543	0	1	0.5	4	0	2	0.25
SRp55	39.6	gi 20127499	30	6	18	24	7	15.5	1.16
SC35 (SFRS2)	25.5	gi 47271443	0	1	0.5	4	0	2	0.25
hTra-2 alpha	32.7	gi 9558733	13	6	9.5	6	7	6.5	1.46
hTra-2 beta (SFRS10)	33.7	gi 4759098	21	9	15	18	5	11.5	1.30
hnRNP									
hnRNP A1	38.7	gi 4504445	12	2	7	1	2	1.5	4.67
hnRNP A3	39.6	gi 34740329	10	0	5	1	0	0.5	10.00
hnRNP AB	36.0	gi 12803583	4	1	2.5	0	0	0	/
hnRNP A2/B1	37.4	gi 14043072	11	1	6	1	0	0.5	12.00

Protein	MW [kDa]	accession no.	B1	B2	B	C1	C2	C	B / C
hnRNP C	33.3	gi 4758544	36	18	27	21	22	21.5	1.26
hnRNP D	38.4	gi 14110420	3	0	1.5	1	0	0.5	3.00
hnRNP F	45.7	gi 148470406	4	0	2	0	0	0	/
hnRNP G	47.4	gi 56699409	22	12	17	10	17	13.5	1.26
hnRNP H1	49.1	gi 5031753	5	0	2.5	3	1	2	1.25
hnRNP H3	36.9	gi 14141157	4	0	2	0	0	0	/
hnRNP K	51.0	gi 14165435	6	7	6.5	0	0	0	/
hnRNP M	77.5	gi 14141152	17	3	10	8	1	4.5	2.22
hnRNP Q	69.6	gi 15809590	7	2	4.5	6	0	3	1.50
hnRNP R	70.9	gi 5031755	15	0	7.5	11	0	5.5	1.36
hnRNP U	90.6	gi 14141161	2	2	2	0	1	0.5	4.00
PCBP1	37.5	gi 5453854	24	7	15.5	4	2	3	5.17
PCBP2	38.1	gi 14141166	24	14	19	4	8	6	3.17
RALY	32.5	gi 8051631	4	3	3.5	5	10	7.5	0.47

The number of acquired tandem MS spectra (MS/MS spectra) for two independent purifications of B and C complexes (B1, B2, C1 and C2, respectively) are given for the spliceosomal proteins identified in Bessonov *et al.*, 2008. For both complexes, the average number of spectra for the two purifications and the protein ratio (B/C) was calculated. The protein ratio displays the protein abundance within the two complexes. A protein ratio of 1 indicates that the distinct protein is present in equal amounts within both complexes. Protein ratios > 1 indicate a higher abundance for these proteins in spliceosomal B complexes, whereas protein ratios < 1 indicate a higher abundance in C complexes. For proteins identified solely in B complexes no protein ratio could be calculated (indicated by "/"). For proteins that were solely or predominantly identified in C complexes the calculated protein ratio is 0 or close to 0 (Table 4.9).

This reanalysis shows that the overall protein assignment was basically correct. The calculated B/C ratio allows for clear assignment of the proteins to B or C specific proteins. Several proteins show a high abundance in B or C complexes. All proteins grouped as potential C specific proteins and proteins recruited to C complexes show a protein ratio of 0 or close to 0 and almost all proteins grouped as B specific proteins show protein ratios higher than one. However, some proteins assigned to be recruited to B complexes show low protein ratios indicating their enrichment in C complexes (e.g. GCFC, hsp27 and KIAA1604, Table 4.9). Surprisingly, only few proteins (e.g. CBP80, U5-116K, U2-A') appear to be present in a 1:1 ratio within B and C complexes.

On the basis of semi-quantitative analysis, we then compared two quantification techniques that are based on stable isotope labeling: (i) Quantification by chemical labeling with iTRAQ reagents, and (ii) quantification of B and C complexes assembled in metabolically labeled nuclear extracts (SILAC quantification).

4.3.3 iTRAQ labeling of spliceosomal B and C complexes for relative quantification

Spliceosomal B and C complexes were purified as described above (Figure 4.15 A). Equal amounts of both complexes were separated by gel electrophoresis (Figure 4.16) and entire lanes were cut into gel slices of equal size. iTRAQ labeling was then performed according to the optimized workflow described in section 4.2.1 (see also Figure 4.11). An internal standard was prepared for all samples to be compared (i.e. from peptides generated from gel slices of B and C complexes cut at the same molecular weight region). Peptides generated from B complex proteins were labeled with iTRAQ reagent 115 and peptides generated from C complexes were labeled with iTRAQ reagent 116. The internal standard was labeled with iTRAQ reagent 114. The samples to be compared and the corresponding internal standard were pooled and the samples were subsequently analyzed by LC-MS/MS. For relative quantification of B and C complexes, two independent replicates were performed. Purified complexes were kindly provided by S. Bessonov.

Pilot experiments showed that the labeling efficiency is crucial for reliable quantification by iTRAQ labeling. To estimate the degree of iTRAQ labeling, first the labeling efficiency was calculated for all protein hits identified in the two replicates. To this end, the peptide and fragment masses were searched against the NCBI nr database allowing iTRAQ labels as variable modification. This yielded the total number of peptides and the number of iTRAQ-labeled peptides. The labeling efficiency was then calculated from the number of labeled peptides and the total number of peptides identified. Figure 4.17 shows the labeling efficiency achieved for the two independent replicates. A labeling efficiency of 95-100 % was achieved for approximately 55 and 75 % of the protein hits detected in the two replicates. Approximately 40 and 20 % of the protein hits show a labeling efficiency between 80 and 95 %. A labeling efficiency of 80 % or lower was calculated for less than 10 % of the protein hits. A high labeling efficiency for iTRAQ labeling was thus achieved for both replicates and reliable quantification can therefore be expected.

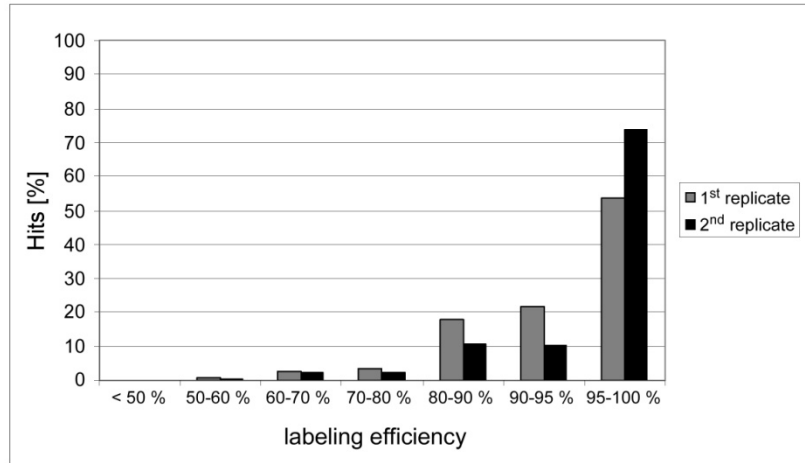


Figure 4.17: Labeling efficiency achieved for iTRAQ labeling in two independent replicates. The labeling efficiency was calculated from the number of peptides that were labeled with iTRAQ reagents and the total number of peptides that were identified for every protein. The number of protein hits is given in percent.

The proteins within B and C complexes were identified by searching peptide and fragment masses against the NCBI nr database using Mascot as search engine. The samples were quantified by comparing the peak areas of iTRAQ reporter ions after fragmentation of the differently labeled peptides (i.e. peptides generated from B and C complexes). A typical spectrum is shown in Figure 4.18. Peptide ratios for all identified proteins were obtained from the Mascot search. Protein ratios were calculated from these after manual removal of outliers.

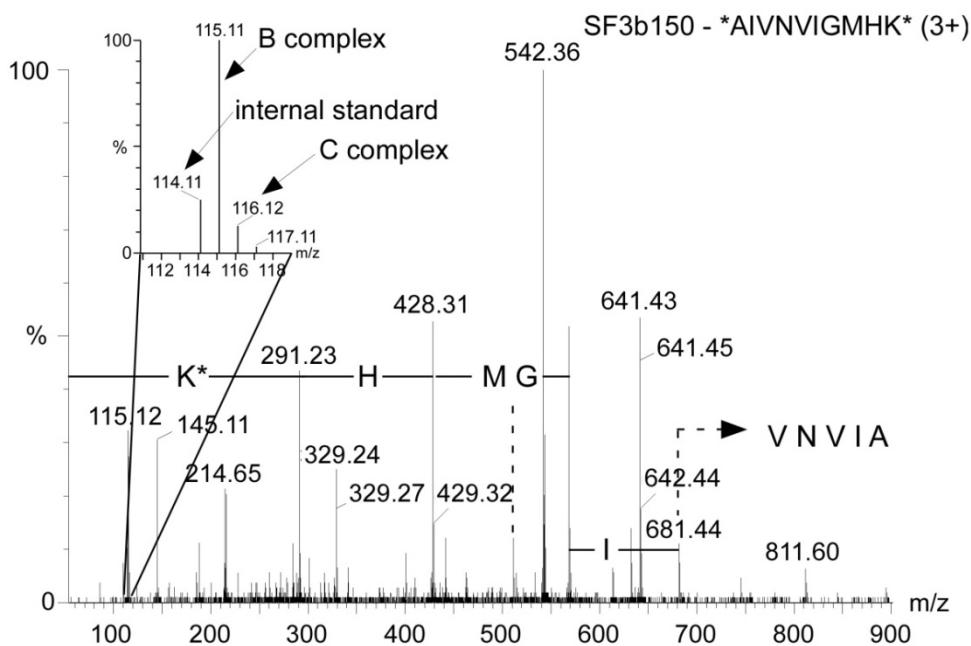


Figure 4.18: Example MS/MS spectrum for iTRAQ quantification of spliceosomal B and C complexes. The peptide AIVNVIGMHK (SF3b130) was labeled (*) with iTRAQ reagents at the N-terminus and at the C-terminal lysine. The iTRAQ reporter region (m/z 114-117) is magnified showing the reporter ions. The peptide was quantified by comparing the area of iTRAQ reporter ions generated from B and C complexes after fragmentation.

In total, 265 proteins were identified and quantified by iTRAQ from two independent replicates. From these, 186 proteins were detected in both replicates. (For the complete list of identified and quantified proteins in the two replicates see Tables A.5 and A.6 in the Appendix.) Spliceosomal proteins identified and quantified within the two replicates were combined in groups of proteins according to their particle or functional association (Bessonov *et al.*, 2008; Deckert *et al.*, 2006).

Calculated protein ratios were normalized on four proteins known to be present with the same abundance in B and C complexes. These are the two cap binding proteins CBP20 and CBP80 and the U5 snRNP specific proteins U5-220K and U5-200K. A normalization factor was calculated from protein ratios of these proteins and applied to all protein ratios obtained by iTRAQ. Table 4.10 shows the list of spliceosomal proteins quantified in B and C complexes. The protein ratio (B/C), the standard deviation of the calculated protein ratio, and the number of peptides used for quantification for the two replicates as well as the average values are given.

Table 4.10: Relative quantification of spliceosomal B and C complexes by iTRAQ. The protein ratio (B/C), the standard deviation (StDev), and the number of peptides used to calculate the protein ratio (#) are given for two replicates (iTRAQ 1 and 2). The mean protein ratio and the corresponding standard deviation were calculated from the two replicates (iTRAQ mean). Proteins without assigned value could not be quantified.

Protein	MW [kDa]	accession no.	iTRAQ 1			iTRAQ 2			iTRAQ mean	
			B/C	StDev	#	B/C	StDev	#	B/C	StDev
U1 snRNP										
U1-A	31.3	gi 4759156				22.35	7.32	5	22.35	
U1-C	17.4	gi 4507127								
U1-70K	51.6	gi 29568103	3.24	1.42	12	7.08	1.10	7	5.16	2.71
17S U2 snRNP										
U2A'	28.4	gi 50593002	0.48	0.18	23	1.39	0.20	24	0.93	0.65
U2B''	25.4	gi 4507123	0.53	0.25	14	1.74	0.42	12	1.14	0.86
SF3a120	88.9	gi 5032087	2.66	0.58	26	2.64	0.44	29	2.65	0.01
SF3a66	49.3	gi 21361376	3.94	0.71	7	8.09	0.43	6	6.02	2.93
SF3a60	58.5	gi 5803167	4.91	1.28	24	6.68	0.65	14	5.80	1.25
SF3b155	145.8	gi 54112117	3.98	2.36	68	3.66	2.15	118	3.82	0.22
SF3b145	100.2	gi 55749531	4.63	1.53	21	4.13	0.93	34	4.38	0.35
SF3b130	135.5	gi 54112121	5.01	1.85	98	3.52	0.72	49	4.27	1.05
SF3b49	44.4	gi 5032069	5.67	1.26	2	5.33	2.56	4	5.50	0.24
SF3b14a (p14)	14.6	gi 7706326	4.25	1.31	11	4.37	0.60	14	4.31	0.08
SF3b14b	12.4	gi 14249398	3.75	1.04	10	2.00	0.25	2	2.87	1.23
SF3b10	10.1	gi 13775200								
17S U2 related										
U2AF65	53.5	gi 6005926	3.63	0.80	2				3.63	
U2AF35	27.9	gi 5803207	5.59	2.69	8				5.59	
hPRP43	90.9	gi 68509926	5.98	2.21	28	2.59	0.35	37	4.28	2.39
SPF45	45.0	gi 14249678	5.38	1.25	5	6.22	0.84	6	5.80	0.60
SR140	118.2	gi 122937227	3.04	0.17	3	4.82	0.80	5	3.93	1.26
CHERP	100.0	gi 119226260	5.36	0.05	2	4.27	1.06	5	4.82	0.77
SF3b125	103.0	gi 45446747	10.06	4.64	6	6.19	2.07	6	8.12	2.73
U5 snRNP										
220K	273.3	gi 3661610	1.18	0.20	154	1.00	0.15	158	1.09	0.13
200K	244.5	gi 45861372	1.05	0.24	211	1.06	0.20	168	1.05	0.01
116K	109.4	gi 41152056	0.96	0.28	52	1.09	0.27	73	1.03	0.10
40K	39.3	gi 4758560	1.03	0.18	20	1.33	0.18	16	1.18	0.22
102K	106.9	gi 40807485	2.77	0.78	55	1.81	0.54	64	2.29	0.68
15K	16.8	gi 5729802	7.03	1.46	3	4.73	0.45	3	5.88	1.63
100K	95.6	gi 41327771	1.57	0.38	48	1.09	0.16	33	1.33	0.34
52K	37.6	gi 5174409	1.06	0.57	3	3.50	1.17	5	2.28	1.72
U4/U6 snRNP										
90K	77.6	gi 4758556	8.27	3.88	37	4.37	1.59	39	6.32	2.75
60K	58.4	gi 45861374	8.18	2.98	26	5.83	0.85	20	7.00	1.66
20K	20.0	gi 5454154	2.50	0.33	3	2.55	0.40	4	2.53	0.04
61K	55.4	gi 40254869	7.43	3.30	26	5.50	1.46	29	6.46	1.37
15.5K	14.2	gi 4826860	9.12	2.96	2	16.96	0.28	2	13.04	5.54
U4/U6.U5 snRNP										
110K	90.2	gi 13926068	4.50	1.96	23	3.60	1.09	19	4.05	0.63
65K	65.4	gi 56550051	1.78	0.36	21	1.63	0.21	19	1.70	0.11
27K (RY1)	18.9	gi 24307919								
hPRP38	37.5	gi 24762236	4.95	1.88	11	4.04	0.85	7	4.50	0.64
TFIP11	96.8	gi 8393259	6.50	1.91	12	1.01	0.08	22	3.75	3.88

Protein	MW [kDa]	accession no.	iTRAQ 1			iTRAQ 2			iTRAQ mean	
			B/C	StDev	#	B/C	StDev	#	B/C	StDev
LSm Proteins										
LSM 2	10.8	gi 10863977	5.24	1.28	4	4.47	0.39	8	4.86	0.54
LSM 3	11.8	gi 7657315	3.45		1	3.21		1	3.33	0.17
LSM 4	15.4	gi 6912486	4.78	1.60	8	5.52	2.31	2	5.15	0.53
LSM 6	9.1	gi 5919153	5.64	2.12	4	6.16	1.02	3	5.90	0.36
LSM 7	11.6	gi 7706423	6.61	3.62	8	3.70	0.08	2	5.15	2.06
LSM 8	10.4	gi 7706425	3.55	1.53	3	4.14	0.70	2	3.85	0.41
Sm Proteins										
B	24.6	gi 4507125	1.69	0.28	13	1.70	0.29	15	1.70	0.01
D1	13.3	gi 5902102	1.39	0.26	6	1.86	0.51	7	1.62	0.34
D2	13.5	gi 29294624	1.62	0.66	35	1.88	0.30	22	1.75	0.18
D3	13.9	gi 4759160	1.40	0.32	32	2.08	0.44	19	1.74	0.48
E	10.8	gi 4507129	1.20	0.17	11	2.03	0.47	7	1.61	0.59
F	9.7	gi 4507131	2.91	0.93	2	1.40	0.37	6	2.15	1.07
G	8.5	gi 4507133	1.52	0.44	4	1.19	0.09	5	1.36	0.23
hPRP19/CDC5L complex										
hPrp19	55.2	gi 7657381	0.62	0.36	61	0.57	0.12	62	0.60	0.04
CDC5L	92.2	gi 11067747	0.23	0.14	45	0.19	0.12	30	0.21	0.03
SPF27	21.5	gi 5031653	0.53	0.13	18	0.55	0.12	15	0.54	0.01
PRL1	57.2	gi 4505895	0.69	0.22	33	1.06	0.22	6	0.87	0.26
Hsp70	70.4	gi 5729877	0.25	0.14	15	1.19	0.22	10	0.72	0.67
AD-002	26.6	gi 7705475	0.22	0.08	6	0.15	0.08	3	0.18	0.05
CTNBL1	65.1	gi 18644734	1.44	0.46	4	2.29	0.34	11	1.86	0.60
Npw38BP	70.0	gi 7706501	7.31	0.72	6	8.77	2.32	17	8.04	1.04
Npw38	30.5	gi 74735456				7.68	1.60	3	7.68	
hPRP19/CDC5L related										
hSYF1	100.0	gi 55770906	0.27	0.09	62	0.38	0.08	65	0.33	0.08
CRNKL1	100.6	gi 30795220	0.47	0.52	88	0.32	0.07	82	0.39	0.11
hlsy1	33.0	gi 20149304	0.20	0.07	16	0.61	0.10	17	0.40	0.28
SKIP	51.1	gi 6912676	0.58	0.17	58	0.63	0.10	42	0.60	0.03
RBM22	46.9	gi 8922328	0.45	0.13	35	0.53	0.11	39	0.49	0.06
Cyp-E	33.4	gi 5174637	0.13	0.05	8	0.58	0.14	5	0.36	0.32
PPIL1	18.2	gi 7706339	0.42	0.18	10	0.45	0.08	8	0.43	0.02
KIAA0560	171.3	gi 38788372	0.09	0.05	96	0.39	0.13	186	0.24	0.21
G10	17.0	gi 32171175	0.39	0.19	14	0.88	0.02	4	0.64	0.35
hRES complex proteins										
SNIP1	45.8	gi 21314720	0.80	0.11	10	1.66	0.15	2	1.23	0.61
MGC12135	70.5	gi 14249338	1.97	0.71	10	0.94	0.38	10	1.46	0.73
CGI-79	39.7	gi 4929627	0.76	0.11	3	1.36	0.10	2	1.06	0.42
Proteins recruited to A complex										
RBM39	59.4	gi 4757926	4.08	0.89	17	4.07	0.73	13	4.07	0.01
p68 (DDX5)	69.2	gi 4758138	1.52	0.55	8	2.30	0.45	15	1.91	0.55
ELAV-like 1 (HuR)	36.1	gi 38201714	4.76	2.29	27	5.65	0.82	18	5.20	0.63
p72/DDX17	80.5	gi 3122595	1.07		1	2.30	0.73	2	1.69	0.87

Protein	MW [kDa]	accession no.	iTRAQ 1			iTRAQ 2			iTRAQ mean	
			B/C	StDev	#	B/C	StDev	#	B/C	StDev
Proteins recruited to B complex										
MFAP1	51.9	gi 50726968	3.54	1.00	38	4.77	0.56	11	4.15	0.87
RED	65.6	gi 10835234	4.64	2.17	23	7.21	2.03	20	5.92	1.82
hSmu-1	57.5	gi 8922679	5.67	2.36	44	4.00	0.65	4	4.83	1.18
PPIL2	59.5	gi 7657473	1.33	0.88	19	2.96	0.74	32	2.15	1.15
hPRP2 (DDX16)	119.2	gi 4503293	0.54	0.16	19	2.10	0.29	20	1.32	1.10
hPRP4-Kinase	117.1	gi 89276756	0.59	0.19	25	0.37	0.07	18	0.48	0.15
THRAP3	108.6	gi 4827040	2.04	0.67	30	1.19	0.08	4	1.62	0.60
PABP1	70.5	gi 46367787	1.70	0.48	8	4.72	2.46	4	3.21	2.13
SKIV2L2	117.8	gi 39930353	1.13	0.20	12	1.15	0.19	8	1.14	0.01
PABPN1	32.6	gi 4758876	1.21	0.09	3	0.79	0.00	2	1.00	0.29
RNF113A	38.8	gi 5902158	0.43	0.15	2	1.20	0.13	3	0.81	0.55
NY-CO-10	53.8	gi 64276486	0.63	0.22	12	0.68	0.05	2	0.66	0.03
KIAA1604	105.5	gi 55749769	0.35	0.15	25	0.46	0.23	27	0.40	0.08
hsp27	22.8	gi 4504517	1.02		1	0.47	0.02	2	0.74	0.39
GCFC	104.7	gi 22035565	2.12	0.42	16	1.36	0.26	24	1.74	0.54
UBL5	8.5	gi 13236510	17.04		1	10.08	6.92	2	13.56	4.92
CCDC16	42.0	gi 49472814	0.93	0.37	11	1.38	0.25	22	1.16	0.31
CCDC12	19.2	gi 21389497	0.53	0.17	6	0.51	0.05	4	0.52	0.02
HsKin17	45.2	gi 13124883	3.10	1.36	6	2.99	0.57	8	3.04	0.08
Step 2 factors										
hPRP22	139.3	gi 4826690	0.13	0.12	55	0.29	0.15	89	0.21	0.11
hPRP18	39.9	gi 4506123	0.38		1	0.39	0.18	4	0.39	0.01
hPRP17	65.5	gi 7706657	0.19	0.08	32	0.45	0.15	57	0.32	0.19
hPRP16	140.5	gi 17999539				0.43	0.11	12	0.43	
hSLU7	68.4	gi 27477111	0.10	0.10	19	0.20	0.09	38	0.15	0.07
Proteins recruited to C complex										
Abstrakt	69.8	gi 21071032	0.08	0.06	48	0.17	0.14	90	0.12	0.06
GCIP p29	28.7	gi 46371998	0.18	0.14	11	0.18	0.07	7	0.18	0.00
DDX35	78.9	gi 20544129	0.32	0.17	28	0.17	0.07	22	0.25	0.11
Q9BRR8	103.3	gi 74732921	0.40	0.09	3	0.19	0.07	9	0.29	0.15
c19orf29 (NY-REN-24)	88.6	gi 126723149	0.17	0.13	21	0.17	0.08	35	0.17	0.00
PPIase-like 3b	18.6	gi 19557636	0.17	0.14	8	0.31	0.19	3	0.24	0.10
PPWD1	73.6	gi 24308049	0.10	0.12	39	0.20	0.09	53	0.15	0.07
MORG1	34.3	gi 153791298	0.10		1	0.69	0.09	3	0.39	0.42
FRG1	29.2	gi 4758404				0.58	0.17	6	0.58	
NOSIP	33.2	gi 7705716	0.11	0.08	9	0.19	0.09	3	0.15	0.05
GPKOW	52.1	gi 15811782	0.66	0.21	6				0.66	
C1orf55	39.3	gi 148664216	0.08	0.04	19	0.13	0.06	14	0.10	0.03
FAM32A	13.1	gi 7661696	0.11	0.06	2				0.11	
RACK1 (GNB2L1)	35.1	gi 5174447								
Tip-49	50.2	gi 4506753								
Potential C complex specific proteins										
PPIG	88.5	gi 42560244	0.20	0.09	6				0.20	
FAM50A	40.1	gi 4758220	0.15	0.08	15	0.29	0.12	5	0.22	0.10
FAM50B	38.6	gi 6912326	0.13	0.03	2	0.36	0.05	3	0.25	0.16

Protein	MW [kDa]	accession no.	iTRAQ 1			iTRAQ 2			iTRAQ mean	
			B/C	StDev	#	B/C	StDev	#	B/C	StDev
SR related proteins										
SRm160	102.5	gi 42542379	1.12	0.16	5				1.12	
SRm300	300.0	gi 4759098	0.31	0.13	7	1.63	0.85	4	0.97	0.93
SR proteins										
SF2/ASF	27.8	gi 5902076	0.80	0.19	14	2.46	0.37	8	1.63	1.17
9G8	27.4	gi 72534660	1.16	0.32	24	2.58	0.38	9	1.87	1.00
SRp20	19.4	gi 4506901	1.65	0.44	8	2.53		1	2.09	0.62
SRp30c	25.5	gi 4506903	0.99	0.20	16	0.53		1	0.76	0.33
SRp38	31.3	gi 5730079	0.91	0.52	15	1.44	0.32	7	1.18	0.38
SRp40	31.3	gi 3929378	0.85	0.07	4	1.83	0.26	5	1.34	0.70
SRp46	31.2	gi 15055543	1.85	1.08	2				1.85	
SRp55	39.6	gi 20127499	1.35	0.23	17	0.88	0.18	7	1.11	0.33
SRp75	56.8	gi 21361282				1.16	0.17	2	1.16	
SC35 (SFRS2)	25.5	gi 47271443								
hTra-2 alpha	32.7	gi 9558733	1.44	0.70	9	2.80	0.47	2	2.12	0.97
hTra-2 beta	33.7	gi 4759098	1.37	1.30	22	1.22	0.37	9	1.30	0.10
hnRNP										
hnRNP A1	38.7	gi 4504445	3.44	1.81	6	5.21	0.13	2	4.32	1.25
hnRNP A3	39.6	gi 34740329	4.62	1.12	6	5.14		1	4.88	0.37
hnRNP AB	36.0	gi 12803583								
hnRNP A2/B1	37.4	gi 14043072	4.63	1.96	6	3.60	0.18	4	4.11	0.73
hnRNP C	33.3	gi 4758544	1.72	0.47	23	1.97	0.30	16	1.84	0.17
hnRNP D	38.4	gi 14110420								
hnRNP F	45.7	gi 148470406				2.94		1	2.94	
hnRNP G	47.4	gi 56699409	1.52	0.59	32	2.21	0.35	6	1.87	0.49
hnRNP G-T	42.7	gi 153252068								
hnRNP H1	49.1	gi 5031753	1.27	0.30	4	1.84	0.27	13	1.56	0.40
hnRNP H3	36.9	gi 14141157								
hnRNP K	51.0	gi 14165435	2.97		1	4.01	0.73	2	3.49	0.74
hnRNP M	77.5	gi 14141152	1.32	0.31	8	1.37	0.38	4	1.34	0.04
hnRNP Q	69.6	gi 15809590								
hnRNP R	70.9	gi 5031755	0.88	0.32	8	1.09		1	0.98	0.15
hnRNP U	90.6	gi 14141161	3.86	2.17	10				3.86	
PCBP1	37.5	gi 5453854	5.02	1.56	5	4.66	1.45	10	4.84	0.26
PCBP2	38.1	gi 14141166	2.58	0.29	4	3.71	1.04	16	3.15	0.80
RALY	32.5	gi 8051631	0.80	0.20	4	2.48	0.68	3	1.64	1.19

Several proteins that are more abundant in B or C complexes, i.e. they show a protein ratio > 1 (high abundance in B complexes) or < 1 (high abundance in C complexes), were identified and quantified by iTRAQ. In addition, some proteins show a protein ratio of approximately 1 meaning that they are present in equal amounts within the two complexes compared. The proteins that were used for normalization of this data set show protein ratios of 1.09, 1.05, 0.82, and 1.21 (U5-220K, U5-200K, CBP20, and CBP80, respectively) confirming that these proteins are indeed present in a 1:1 ratio within B and C complexes and can be used to normalize the protein ratios. Nonetheless, some proteins that have been

identified during the proteomic analysis of the B and C complexes were only quantified in one of the two replicates and few proteins could not be quantified by iTRAQ.

4.3.4 SILAC quantification of spliceosomal B and C complexes

Spliceosomal B and C complexes for SILAC quantification were purified as described above (Figure 4.15 B). To this end, heavy and light nuclear extracts (SILAC nuclear extracts) were prepared from metabolically labeled HeLa cells. B complexes were then purified from heavy nuclear extract and C complexes were purified from light nuclear extract. The purified complexes were pooled in equal amounts and the proteins were separated by gel electrophoresis. After hydrolysis of the proteins, generated peptides were analyzed by LC-MS/MS. B and C complexes assembled in metabolically nuclear extracts were purified by J. Deckert and SILAC quantification was performed by M. Grønborg.

To prove that the different metabolically labeled nuclear extracts (i.e. light and heavy SILAC nuclear extracts) show same behavior during pre-mRNA splicing, splicing kinetics, spliceosomal complex formation, and the RNA composition of the purified complexes were monitored for the differently labeled nuclear extracts (Figure 4.19). Analysis from splicing revealed no differences between the two SILAC nuclear extracts (Figure 4.19 A). In both cases, i.e. using light or heavy nuclear extract, splicing products first appeared after 10 minutes incubation time and the amount of pre-mRNA was reduced during incubation over 180 minutes. Spliceosomal complex formation was identical for light and heavy nuclear extracts (Figure 4.19 B). H/E complexes were rapidly formed, whereas A and B complexes appeared after 2 to 4 minutes. C complexes were first observed after 10 to 15 minutes of incubation. RNase H digestion using DNA oligonucleotides led to degradation of early spliceosomes. The RNA composition of the purified B and C complexes is shown in Figure 14.9 C. Purified B complexes contained equimolar amounts of U1, U2, U4, U5, and U6 snRNA (Figure 14.9 C, lane 1), while C complexes contained only U2, U5, and U6 snRNA (Figure 14.9 C, lane 3). Splicing products and reduced amounts of pre-mRNA were detected in C complexes (Figure 14.9 C, lane 4), whereas B complexes contained high amounts of pre-mRNA and no splicing products (Figure 14.9 C, lane 2). As the light and heavy nuclear extracts show same behavior during pre-mRNA splicing, the proteomes of purified B and C complexes can be relatively compared by SILAC using the tested nuclear extracts. Analysis of the splicing kinetics and the complex formation was performed by J. Deckert.

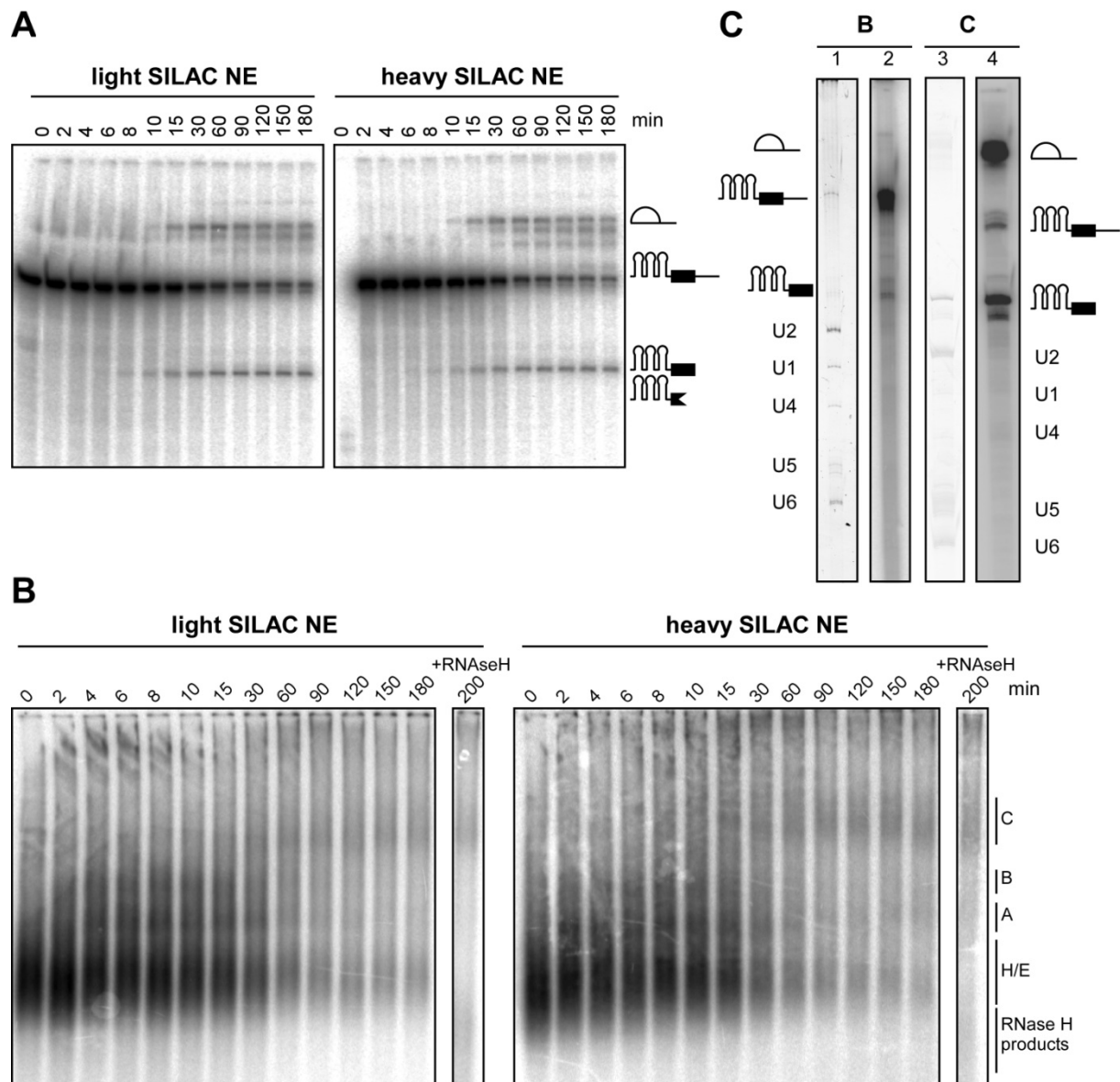


Figure 4.19: Splicing kinetics, complex formation and RNA composition of B and C complexes as obtained using light and heavy SILAC nuclear extracts. The splicing kinetics and the complex formation are identical for light and heavy SILAC nuclear extracts. This figure was kindly provided by J. Deckert. **(A)** The splicing kinetics were analyzed by denaturing gel electrophoresis of aliquots from splicing reactions over 180 min. Pre-mRNA and splicing products were visualized by autoradiography. Splicing products first appeared after 10 min of incubation. **(B)** The spliceosomal complex formation was assayed by native agarose gel electrophoresis and visualized by autoradiography. A and B complex formation was observed after 2 to 4 min whereas C complexes first appeared after 10 to 15 min. **(C)** The RNA composition of purified B (heavy nuclear extract) and C (light nuclear extracts) complexes was analyzed by denaturing gel electrophoresis and visualized by silver staining (lanes 1 and 3) or autoradiography (lanes 2 and 4). B complexes contained U1, U2, U4, U5, and U6 snRNA (lane 1) and a high amount of pre-mRNA (lane 2). C complexes contained U2, U5, and U6 snRNA (lane 3), splicing products and reduced amounts of pre-mRNA (lane 4).

Proteins from SILAC labeled B and C complexes were identified by searching peptide and fragment masses against NCBI nr database using Mascot as search engine. Peptides generated from the two complexes were quantified by comparing the peak area of the differently labeled peptides in MS spectra. A typical MS spectrum is shown in Figure 4.20. Peptide and resulting protein ratios were obtained using the MS Quant software. In total, 266 proteins could be identified and quantified in the two independent replicates. Of these, 160 proteins were common to both replicates. Obtained protein ratios were normalized as described for iTRAQ (see above) using protein ratios of CBP20, CBP80, U5-220K, and U5-200K proteins, which are expected to be present in equal amounts in both complexes. The data are shown in Table 4.11. The protein ratio (B/C), the standard deviation, and the number of peptides used for quantification for the two replicates as well as the average values are given.

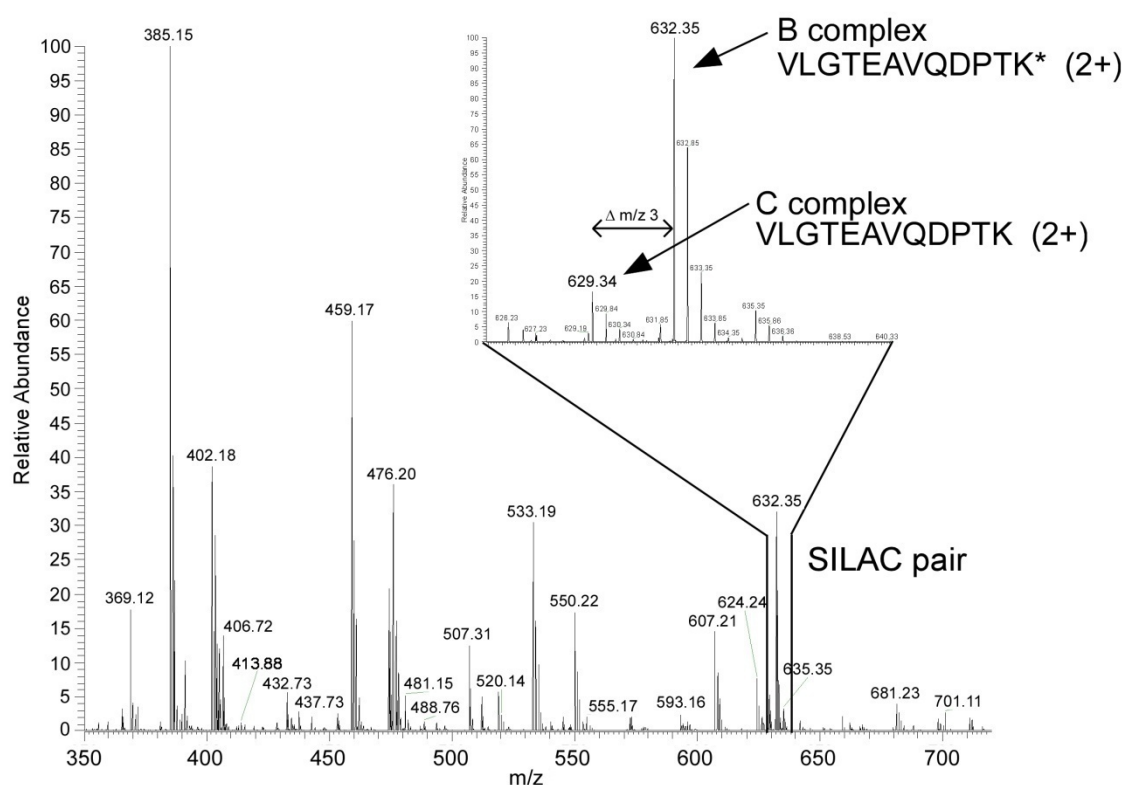


Figure 4.20: Example MS spectrum for SILAC quantification of spliceosomal B and C complexes. The peptide VLGTEAVQDPTK (U4/U6-90K) was metabolically labeled by incorporation of stable isotope labeled lysine (*). The SILAC pair (i.e. the peptide derived from B and C complexes purified from differently labeled nuclear extracts) is magnified showing different intensities for this particular peptide within the two complexes. The peptide generated from B and C complexes was quantified by comparing the peak area of the two differently labeled peptides.

Table 4.11: Relative quantification of spliceosomal B and C complexes by SILAC. The protein ratio (B/C), the standard deviation (StDev), and the number of peptides used to calculate the protein ratio (#) are given for two replicates (SILAC 1 and 2). The mean protein ratio and the corresponding standard deviation (SILAC mean) were calculated from the two replicates. Proteins without assigned value could not be quantified.

Protein	MW [kDa]	accession no.	SILAC #1			SILAC #2			SILAC mean	
			B/C	StDev	#	B/C	StDev	#	B/C	StDev
U1 snRNP										
U1-A	31.3	gi 4759156	22.18	3.04	3	29.73		1	25.95	5.34
U1-C	17.4	gi 4507127				27.15		1	27.15	
U1-70K	51.6	gi 29568103	3.19	1.40	5	11.83	0.41	6	7.51	6.11
17S U2 snRNP										
U2A'	28.4	gi 50593002	1.07	0.12	24	0.78	0.03	25	0.92	0.21
U2B''	25.4	gi 4507123	0.96	0.45	14	0.90	0.03	7	0.93	0.04
SF3a120	88.9	gi 5032087	3.14	0.04	11	3.15	0.12	32	3.15	0.00
SF3a66	49.3	gi 21361376	3.04	0.59	13	3.30	0.01	2	3.17	0.19
SF3a60	58.5	gi 5803167	3.16	0.77	37	3.35	0.18	9	3.26	0.13
SF3b155	145.8	gi 54112117	3.24	0.42	32	4.60	1.41	71	3.92	0.96
SF3b145	100.2	gi 55749531	3.29	0.25	25	2.95	0.20	42	3.12	0.24
SF3b130	135.5	gi 54112121	3.40	0.35	31	3.40	0.14	94	3.40	0.00
SF3b49	44.4	gi 5032069	3.58	0.27	6	2.83	0.03	3	3.20	0.54
SF3b14a (p14)	14.6	gi 7706326	5.01	0.65	6	2.80	0.09	10	3.90	1.56
SF3b14b	12.4	gi 14249398				3.88	0.11	4	3.88	
SF3b10	10.1	gi 13775200				3.43	0.11	4	3.43	
17S U2 related										
U2AF65	53.5	gi 6005926	1.51		1				1.51	
U2AF35	27.9	gi 5803207	2.86		1				2.86	
hPRP43	90.9	gi 68509926	4.69	0.65	51	4.15	0.20	25	4.42	0.38
SPF45	45.0	gi 14249678	10.09	0.00	7	10.78	0.41	5	10.43	0.49
SR140	118.2	gi 122937227	12.48	3.02	52	12.23	3.32	13	12.35	0.18
CHERP	100.0	gi 119226260	4.12	0.00	10	10.60	0.54	5	7.36	4.58
SF3b125	103.0	gi 45446747	5.15	0.00	1	17.58	1.46	3	11.36	8.79
U5 snRNP										
220K	273.3	gi 3661610	0.98	0.02	107	1.13	0.07	179	1.05	0.10
200K	244.5	gi 45861372	1.00	0.14	128	1.10	0.06	229	1.05	0.07
116K	109.4	gi 41152056	0.93	0.12	92	1.08	0.05	65	1.00	0.11
40K	39.3	gi 4758560	0.94	0.13	6	1.05	0.07	23	0.99	0.08
102K	106.9	gi 40807485	3.32	0.24	18	3.48	0.40	53	3.40	0.11
15K	16.8	gi 5729802				13.85	1.99	2	13.85	
100K	95.6	gi 41327771	2.78	0.68	28	2.38	0.18	23	2.58	0.29
52K	37.6	gi 5174409	2.02	0.54	10	2.33	0.08	8	2.17	0.21
U4/U6 snRNP										
90K	77.6	gi 4758556	16.28	7.39	42	19.38	1.84	24	17.83	2.19
60K	58.4	gi 45861374	14.21	5.44	16	17.08	0.93	4	15.64	2.02
20K	20.0	gi 5454154	7.17	1.90	5	10.18	0.36	8	8.67	2.12
61K	55.4	gi 40254869	17.30	7.67	13	21.88	1.49	3	19.59	3.23
15.5K	14.2	gi 4826860				21.50	1.02	2	21.50	
U4/U6.U5 snRNP										
110K	90.2	gi 13926068	7.25	1.72	51	10.78	0.67	29	9.01	2.50
65K	65.4	gi 56550051	3.22	0.33	35	3.28	0.12	11	3.25	0.04
27K (RY1)	18.9	gi 24307919				6.25	0.22	3	6.25	
hPRP38	37.5	gi 24762236				2.70	0.67	11	2.70	
TFIP11	96.8	gi 8393259	1.71	0.18	11	1.45	0.04	8	1.58	0.18

Protein	MW [kDa]	accession no.	SILAC #1			SILAC #2			SILAC mean	
			B/C	StDev	#	B/C	StDev	#	B/C	StDev
L_{Sm} Proteins										
LSM 2	10.8	gi 10863977	23.42		1	18.35	0.77	7	20.88	3.58
LSM 3	11.8	gi 7657315	4.04	0.00	2	14.13	0.29	3	9.08	7.13
LSM 4	15.4	gi 6912486	9.60	4.06	3	16.35	0.91	3	12.98	4.77
LSM 5	9.9									
LSM 6	9.1	gi 5919153	15.80	0.00	2	19.85	1.58	6	17.82	2.87
LSM 7	11.6	gi 7706423	19.62	8.34	2				19.62	
LSM 8	10.4	gi 7706425	10.70	5.53	5				10.70	
S_m Proteins										
B	24.6	gi 4507125	1.70	0.09	5	1.50	0.15	19	1.60	0.14
D1	13.3	gi 5902102	1.96	0.28	11	1.70	0.07	8	1.83	0.18
D2	13.5	gi 29294624	1.98	0.25	21	1.50	0.06	21	1.74	0.34
D3	13.9	gi 4759160	1.89	0.24	17	1.73	0.06	9	1.81	0.12
E	10.8	gi 4507129	2.03	0.23	8	1.73	0.04	7	1.88	0.21
F	9.7	gi 4507131	1.82	0.00	2	1.73	0.04	7	1.77	0.07
G	8.5	gi 4507133	1.94	0.00	2	1.78	0.03	8	1.86	0.12
hPRP19/CDC5L complex										
hPrp19	55.2	gi 7657381	0.35	0.07	59	0.38	0.06	29	0.36	0.02
CDC5L	92.2	gi 11067747	0.41	0.13	38	0.43	0.16	66	0.42	0.01
SPF27	21.5	gi 5031653	0.33	0.07	28	0.35	0.01	17	0.34	0.01
PRL1	57.2	gi 4505895	0.33	0.10	26	0.33	0.01	11	0.33	0.01
Hsp70	70.4	gi 5729877	0.19	0.12	21	0.15	0.01	4	0.17	0.03
AD-002	26.6	gi 7705475	0.21	0.02	8	0.33	0.01	9	0.27	0.08
CTNBL1	65.1	gi 18644734	0.59	0.05	4	0.75	0.02	10	0.67	0.11
Npw38BP	70.0	gi 7706501	3.10	0.49	2	3.95	0.15	2	3.53	0.60
Npw38	30.5	gi 74735456				3.55		1	3.55	
hPRP19/CDC5L related										
hSYF1	100.0	gi 55770906	0.22	0.04	86	0.28	0.05	42	0.25	0.04
CRNKL1	100.6	gi 30795220	0.27	0.10	78	0.25	0.01	44	0.26	0.02
hlsy1	33.0	gi 20149304	0.17	0.17		0.23	0.02	15	0.20	0.04
SKIP	51.1	gi 6912676	0.29	0.05	85	0.33	0.03	34	0.31	0.03
RBM22	46.9	gi 8922328	0.28	0.10	20	0.30	0.02	14	0.29	0.02
Cyp-E	33.4	gi 5174637	0.18	0.01	16	0.23	0.01	10	0.20	0.03
PPIL1	18.2	gi 7706339	0.25	0.03	8	0.38	0.02	21	0.31	0.09
KIAA0560	171.3	gi 38788372	0.22	0.07	122	0.28	0.06	86	0.25	0.04
G10	17.0	gi 32171175								
hRES complex proteins										
SNIP1	45.8	gi 21314720				0.73	0.03	7	0.73	
MGC12135	70.5	gi 14249338	0.74	0.07	10	0.75	0.03	14	0.75	0.01
CGI-79	39.7	gi 4929627	0.52	0.06	8	0.73	0.02	6	0.62	0.15
Proteins recruited to A complex										
RBM39	59.4	gi 4757926	3.32	0.75	17	3.65	0.11	5	3.48	0.24
p68 (DDX5)	69.2	gi 4758138	3.38	0.23	5	3.05	0.39	4	3.21	0.23
ELAV-like 1 (HuR)	36.1	gi 38201714	9.86	4.68	28	10.50	0.76	16	10.18	0.45
p72/DDX17	80.5	gi 3122595	2.90		1				2.90	

Protein	MW [kDa]	accession no.	SILAC #1			SILAC #2			SILAC mean	
			B/C	StDev	#	B/C	StDev	#	B/C	StDev
Proteins recruited to B complex										
MFAP1	51.9	gi 50726968	2.13	0.24	10	3.18	0.16	11	2.65	0.74
RED	65.6	gi 10835234	5.62	0.88	11	5.40	0.16	11	5.51	0.15
hSmu-1	57.5	gi 8922679	2.74	0.32	9	9.50	0.27	8	6.12	4.78
PPIL2	59.5	gi 7657473	0.48	0.14	12	0.80	0.03	17	0.64	0.23
hPRP2 (DDX16)	119.2	gi 4503293	0.04		1	0.58	0.02	23	0.31	0.38
hPRP4-Kinase	117.1	gi 89276756	1.41	0.24	15				1.41	
THRAP3	108.6	gi 4827040	3.56	0.57	5	16.85	1.17	11	10.21	9.39
PABP1	70.5	gi 46367787	0.30	0.04	7				0.30	
SKIV2L2	117.8	gi 39930353	3.98	0.72	13	2.70	0.06	7	3.34	0.91
PABPN1	32.6	gi 4758876	0.36	0.01	4	1.08	0.03	6	0.72	0.51
RNF113A	38.8	gi 5902158				0.43	0.02	8	0.43	
NY-CO-10	53.8	gi 64276486	0.33	0.10	8	0.70	0.04	8	0.51	0.26
KIAA1604	105.5	gi 55749769	0.16	0.05	17	0.23	0.02	26	0.19	0.04
hsp27	22.8	gi 4504517				0.25	0.00	2	0.25	
GCFC	104.7	gi 22035565	1.60	0.26	10	1.50	0.09	11	1.55	0.07
UBL5	8.5	gi 13236510								
CCDC16	42.0	gi 49472814	0.41	0.06	5	0.73	0.04	5	0.57	0.22
CCDC12	19.2	gi 21389497				0.28	0.07	13	0.28	
HsKin17	45.2	gi 13124883	1.25	0.14	7	1.40	0.06	5	1.33	0.10
Step 2 factors										
hPRP22	139.3	gi 4826690	0.14	0.10	49	0.10	0.07	48	0.12	0.03
hPRP18	39.9	gi 4506123				0.10		1	0.10	
hPRP17	65.5	gi 7706657	0.29	0.10	28	0.23	0.02	29	0.26	0.04
hPRP16	140.5	gi 17999539				0.48	0.01	2	0.48	
hSLU7	68.4	gi 27477111	0.24	0.08	6	0.40	0.53	14	0.32	0.11
Proteins recruited to C complex										
Abstrakt	69.8	gi 21071032	0.19	0.06	25	0.13	0.15	24	0.16	0.05
GCIP p29	28.7	gi 46371998	0.02	0.20	11	0.15	0.15	18	0.09	0.09
DDX35	78.9	gi 20544129	0.07	0.08	13	0.05	0.01	15	0.06	0.01
Q9BRR8	103.3	gi 74732921				0.05	0.01	7	0.05	
c19orf29 (NY-REN-24)	88.6	gi 126723149				0.05	0.02	21	0.05	
PPIase-like 3b	18.6	gi 19557636	0.01	0.00	2	0.03	0.01	7	0.02	0.01
PPWD1	73.6	gi 24308049	0.08	0.03	18	0.08	0.05	18	0.08	0.00
MORG1	34.3	gi 153791298	0.07	0.01	2	0.05	0.00	5	0.06	0.02
FRG1	29.2	gi 4758404	0.22	0.02	3	0.23	0.01	2	0.22	0.00
NOSIP	33.2	gi 7705716				0.10	0.05	2	0.10	
GPKOW	52.1	gi 15811782	0.09	0.04	5	0.33	0.04	12	0.21	0.17
C1orf55	39.3	gi 148664216	0.19	0.00	2	0.05	0.03	25	0.12	0.10
FAM32A	13.1	gi 7661696	0.12		1	0.03	0.02	4	0.07	0.07
RACK1 (GNB2L1)	35.1	gi 5174447								
Tip-49	50.2	gi 4506753				0.13	0.01	3	0.13	
Potential C complex specific proteins										
PPIG	88.5	gi 42560244	0.54	0.15	2	0.20	0.06	5	0.37	0.24
FAM50A	40.1	gi 4758220	0.15		1	0.03	0.00	7	0.09	0.09
FAM50B	38.6	gi 6912326				0.05	0.03	2	0.05	

Protein	MW [kDa]	accession no.	SILAC #1			SILAC #2			SILAC mean	
			B/C	StDev	#	B/C	StDev	#	B/C	StDev
C9orf78	33.7	gi 7706557	0.04	0.01	2	0.03	0.01	2	0.03	0.01
C10orf4	37.5	gi 24432067				0.10	0.05	5	0.10	
CXorf56	25.6	gi 11545813	0.18	0.07	3	0.08	0.06	8	0.13	0.08
DGCR14	52.4	gi 13027630	0.25	0.07	8	0.03	0.01	8	0.14	0.16
CCDC130	44.7	gi 13540614				0.10	0.02	2	0.10	
TOE1	56.4	gi 156564398								
NKAP	47.0	gi 13375676	0.10		1	0.03	0.00	5	0.06	0.05
ZCCHC10	18.4	gi 8923106	0.16	0.00	2	0.08	0.03	2	0.12	0.06
CDK10	35.4	gi 16950647				0.10	0.04	6	0.10	
TTC14	88.2	gi 33457330	0.34		1	0.08	0.02	8	0.21	0.19
WDR70	73.2	gi 8922301								
NFKBIL1	43.1	gi 26787991				0.08	0.01	3	0.08	
JUP	81.6	gi 12056468	0.08	0.08	4				0.08	
EJC/mRNP										
eIF4A3	46.9	gi 7661920	0.26	0.08	18	0.23	0.02	21	0.24	0.03
Magoh	17.2	gi 4505087	0.21	0.01	6	0.28	0.00	3	0.24	0.04
Y14	19.9	gi 4826972	0.13	0.13	2	0.18	0.04	2	0.15	0.03
Pinin	81.6	gi 33356174	0.31		1	1.23	0.02	4	0.77	0.65
RNPS1	34.2	gi 6857826	0.63	0.02	3	2.28	0.08	5	1.45	1.17
Acinus	151.8	gi 7662238	0.50	0.11	11	2.75	0.33	13	1.63	1.59
SAP18	17.4	gi 5032067				2.25	0.03	2	2.25	
Aly/REF (THOC4)	26.9	gi 55770864	0.98	0.26	2	3.20	0.07	5	2.09	1.57
UAP56	49.1	gi 18375623	0.28	0.03	5	6.08	0.07	4	3.18	4.10
TREX										
THOC1	75.6	gi 154448890	0.07	0.02	2				0.07	
THOC2	169.6	gi 125656165	0.15	0.02	5	3.43	0.09	3	1.79	2.31
THOC3	38.8	gi 14150171								
KIAA0983 (THOC5)	78.4	gi 50959110				3.15		1	3.15	
WDR58 (THOC6)	37.4	gi 31543164				1.25	0.11	3	1.25	
pre-mRNA/mRNA binding proteins										
CBP20	18.0	gi 110349727	0.88	0.06	10	1.20	0.02	7	1.04	0.23
CBP80	91.8	gi 4505343	0.89	0.11	64	1.20	0.07	53	1.04	0.22
NF45	43.0	gi 24234747	2.86	0.44	12	4.90	0.22	3	3.88	1.44
ZC3H18	104.0	gi 31377595	3.74	0.54	13	4.15	0.17	7	3.95	0.29
YB-1	35.9	gi 34098946	0.56	0.12	14	0.35	0.02	11	0.45	0.15
ELG	38.9	gi 8923771	1.37	0.07	5	10.50	0.06	3	5.93	6.46
DDX3	73.3	gi 87196351	4.71	1.13	4				4.71	
ASR2B	100.0	gi 33383233	3.29	0.55	51	4.18	0.32	22	3.73	0.63
BCLAF1	107.2	gi 7661958	0.29		1	12.80	0.33	2	6.54	8.85
DBPA	40.1	gi 20070160				0.33		1	0.33	
RBM7	30.5	gi 4503293								
HSP70	70.0	gi 5123454	0.19	0.12	21	0.15	0.01	4	0.17	0.03
Miscellaneous proteins										
BAG2	23.4	gi 4757834				0.50	0.01	5	0.50	
RBBP6	197.2	gi 33620716				1.93	0.10	2	1.93	
RBM42	50.3	gi 21359951	12.35	12.21	2				12.35	

Protein	MW [kDa]	accession no.	SILAC #1			SILAC #2			SILAC mean	
			B/C	StDev	#	B/C	StDev	#	B/C	StDev
SR related proteins										
SRm160	102.5	gi 42542379	0.14	0.05	3	0.50	0.06	3	0.32	0.25
SRm300	300.0	gi 4759098	0.26	0.12	13	0.40	0.09	26	0.33	0.10
SR proteins										
SF2/ASF	27.8	gi 5902076	1.62	0.23	44	5.43	0.73	23	3.52	2.69
9G8	27.4	gi 72534660	1.91	0.32	14	3.98	0.98	19	2.94	1.46
SRp20	19.4	gi 4506901	2.89	0.00	2	9.15	0.00	2	6.02	4.42
SRp30c	25.5	gi 4506903	1.33	0.12	16	1.30	0.25	18	1.32	0.02
SRp38	31.3	gi 5730079	0.95	0.18	6				0.95	
SRp40	31.3	gi 3929378	0.70	0.19	10	1.90	0.58	10	1.30	0.85
SRp46	31.2	gi 15055543				4.15	0.15	2	4.15	
SRp55	39.6	gi 20127499	1.41	0.83	4	1.58	0.44	15	1.49	0.12
SRp75	56.8	gi 21361282				5.93	0.17	2	5.93	
SC35 (SFRS2)	25.5	gi 47271443								
hTra-2 alpha	32.7	gi 9558733	2.90		15	4.33	0.15	6	3.61	1.01
hTra-2 beta	33.7	gi 4759098	3.97	0.94	2	5.53	0.97	16	4.75	1.10
hnRNP										
hnRNP A1	38.7	gi 4504445	12.92	4.92		12.13	1.44	12	12.52	0.56
hnRNP A3	39.6	gi 34740329	5.83	2.37	4	5.03	0.11	3	5.43	0.57
hnRNP AB	36.0	gi 12803583	7.38	0.05	3	7.05	0.68	3	7.21	0.23
hnRNP A2/B1	37.4	gi 14043072	2.93		1	9.28	1.75	8	6.10	4.49
hnRNP C	33.3	gi 4758544	1.34	0.14	41	1.68	0.11	28	1.51	0.24
hnRNP D	38.4	gi 14110420	9.75	0.00	2	10.40	2.42	2	10.08	0.46
hnRNP F	45.7	gi 148470406				3.85	0.07	2	3.85	
hnRNP G	47.4	gi 56699409	3.30	0.66	19	3.75	0.12	11	3.52	0.32
hnRNP G-T	42.7	gi 153252068				0.08		1	0.08	
hnRNP H1	49.1	gi 5031753	2.63	0.89	3	2.65	0.16	6	2.64	0.01
hnRNP H3	36.9	gi 14141157								
hnRNP K	51.0	gi 14165435	12.05	4.94	13				12.05	
hnRNP M	77.5	gi 14141152	4.30	1.59	8	2.83		1	3.56	1.04
hnRNP Q	69.6	gi 15809590	2.49	0.07	3				2.49	
hnRNP R	70.9	gi 5031755	0.87	0.15	12	1.23	0.07	4	1.05	0.25
hnRNP U	90.6	gi 14141161	12.82	0.00	2	9.65	1.39	2	11.24	2.24
PCBP1	37.5	gi 5453854	5.32	1.27	11	5.53	0.08	2	5.42	0.14
PCBP2	38.1	gi 14141166	5.44	0.81	12	3.50	0.01	2	4.47	1.37
RALY	32.5	gi 8051631				1.25	0.11	8	1.25	

The proteins used for normalization show a protein ratio of 1, confirming the suitability of these proteins for normalization. As described for the iTRAQ analysis, several proteins showing a high abundance in B and C complexes and only few proteins present in equal amounts within the two complexes were identified.

4.3.5 Relative quantification of spliceosomal B and C complexes

Several snRNP specific proteins and numerous non-snRNP specific proteins identified in spliceosomal B and C complexes have been quantified by iTRAQ, SILAC, and spectral count and all three approaches give information about the protein abundance within the pre-catalytic and the catalytic spliceosome (i.e. B and C complexes). In the following paragraphs the results will be compared and discussed more in detail.

4.3.5.1 Proteins common to B and C complexes

Confidence for the quantification approaches was achieved by the observed 1:1 ratio of the two 5' pre-mRNA cap binding proteins CBP20 and CBP80 (see Tables 4.9, 4.10, and 4.11). The cap binding proteins interact with the 5' cap structure of the pre-mRNA and should be present in equal amounts in B and C complexes. For all approaches (i.e. iTRAQ, SILAC, and spectral count) a ratio of approximately 1 within B and C complexes was observed for these proteins (except CBP20, which shows a protein ratio of 0.67 for spectral count; see Table 4.9).

The Sm proteins are common to all U snRNP except for U6 snRNP. Therefore, four copies of Sm proteins are expected in B complex. Upon C complex formation, U1 and U4 are dissociated from the spliceosome and only two copies are left. A comparison of the protein ratios for the Sm proteins obtained by the different methods is shown in Figure 4.21. For all seven Sm proteins the iTRAQ and SILAC procedures yield protein ratios close to the expected value of 2. In contrast, spectral count yielded the correct value only for SmF and SmG and clearly gave a wrong result for the other Sm proteins.

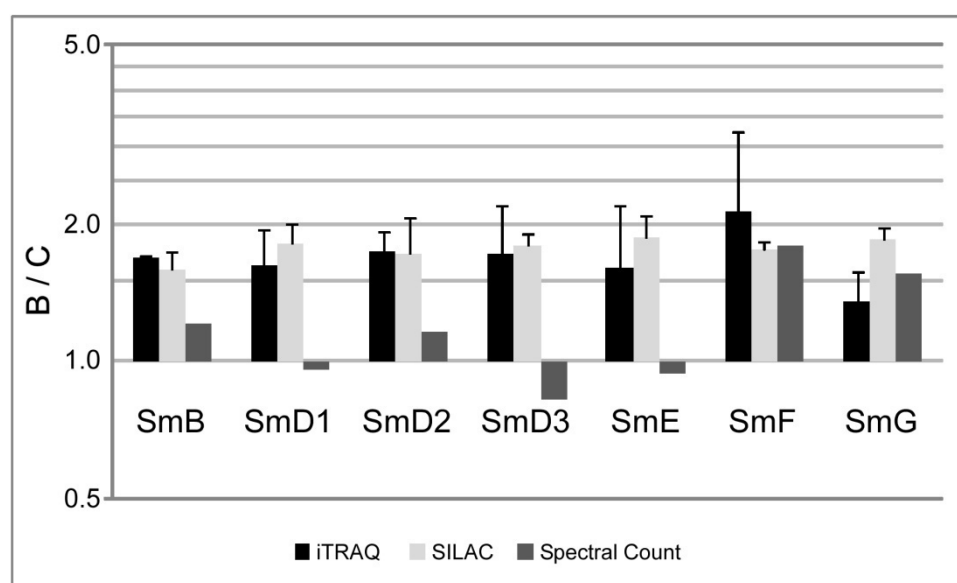


Figure 4.21: Relative abundances of the Sm proteins in the B and C complexes determined by iTRAQ, SILAC, and spectral count. iTRAQ and SILAC quantification reveal an average protein ratio of 1.75 for the Sm proteins, which is very close to the expected protein ratio of 2.

Although the U2 and U5 snRNP are stable associated to B and C complexes, the obtained protein ratios show that only selected members of U2 and U5 are present in equal amounts in B and C complexes (Figure 4.22). U2-A' and U2-B'' show a protein ratio of 1 for iTRAQ and SILAC analysis and a slightly higher value (approximately 1.3) for spectral count (Figure 4.22 A). In contrast, the two U2 snRNP associated splicing factors SF3a and SF3b show a much higher abundance in B complex. SF3a and SF3b likely dissociate from the spliceosome during transition from B to C complex. As they were found to be present in the activated spliceosome (Makarov *et al.*, 2002), our data thus give compelling evidence for their dissociation upon activation of the spliceosome. Comparing iTRAQ, SILAC, and spectral count, the obtained protein ratios are consistent for all SF3a and SF3b proteins (except SF3b49 and SF3b14a).

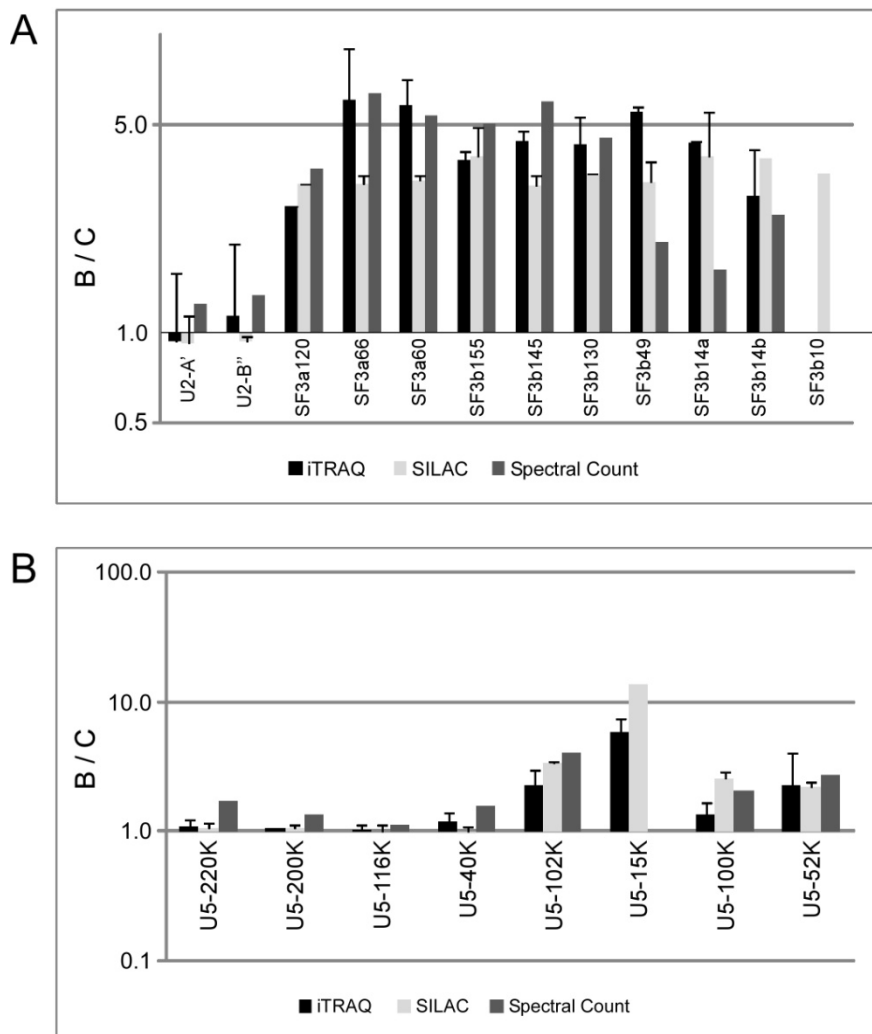


Figure 4.22: Relative protein abundances of the U2 and U5 snRNP specific proteins obtained by iTRAQ, SILAC, and spectral. (A) Protein ratios for U2 snRNP specific proteins. (B) Protein ratios for U5 snRNP proteins. A protein ratio of U5-15K could not be determined from the spectral count.

Although U5 snRNA is stably associated with both the B and C complexes, only four of the eight U5 snRNP proteins were found to be present in a 1:1 ratio (U5-220K, U5-200K, U5-116K, and U5-40K; Figure 4.22 B). The U5 snRNP proteins 102K, 100K, 52K, and 15K show higher ratios (from 2 to 8.5) indicating their dissociation from the spliceosome during transition from B to C complex. These findings are in agreement with previous studies (Makarov *et al.*, 2002), where the absence of U5 snRNP proteins 100K, 52K, and 15K in the activated spliceosome is already discussed. Interestingly, the 15K protein shows a very high enrichment in B complexes as compared to C complexes. A protein ratio for this particular protein could not be determined from the spectral count as it was totally absent from the C complex (Table 4.9).

The RES complex, which binds to the spliceosome before the first step of splicing, was found to be necessary for efficient intron removal and nuclear pre-mRNA retention (Dziembowski *et al.*, 2004). Our data are consistent with these findings and association with B and C complexes without much change in relative quantification was obtained by all three quantification techniques (see Tables 4.9, 4.10, and 4.11).

Most of the proteins were either specific for B or C complex, underlying the highly dynamic nature of the spliceosome during its assembly pathway.

4.3.5.2 Proteins predominantly associated with B complex

U1 and U4 snRNP dissociate from the spliceosome during transition from B to C complex. This is clearly observed in our analysis, where all the U1 and U4 proteins show high protein ratios indicating their specificity to B complex (see Figure 4.23). For U1-A, U1-C, and U1-70K proteins high protein ratios were obtained by iTRAQ and SILAC indicating their high abundance in B complex (Figure 4.23 A). Furthermore, these proteins were totally absent in C complex and no protein ratios could be determined from spectral count (see Table 4.9).

Also U4/U6 snRNP specific proteins show high protein ratios for all three quantification approaches (Figure 4.23 B) showing that these proteins together with the U4 snRNA dissociate from the spliceosome. In addition, tri-snRNP (U4/U6.U5) specific proteins show (with few exceptions) high protein ratios (Figure 4.23 B). This observation reveals that, consistent with previous studies (Makarov *et al.*, 2002), some tri-snRNP specific proteins dissociate from the spliceosome during transition from B to C complex.

The LSm proteins are associated with U6 snRNP and the protein ratios for these proteins corroborate their dissociation from the spliceosome during transition from B to C complex, although the U6 snRNA remains associated (Figure 4.23 C). These protein ratios were high,

irrespective of the method used. Consistent with Chan *et al.*, 2003 association of the NTC (nineteen complex, the yeast homologue of the human hPrp19/CDC5L complex) during activation of the spliceosome leads to destabilization of LSM proteins and U6 snRNA, what further supports this observation.

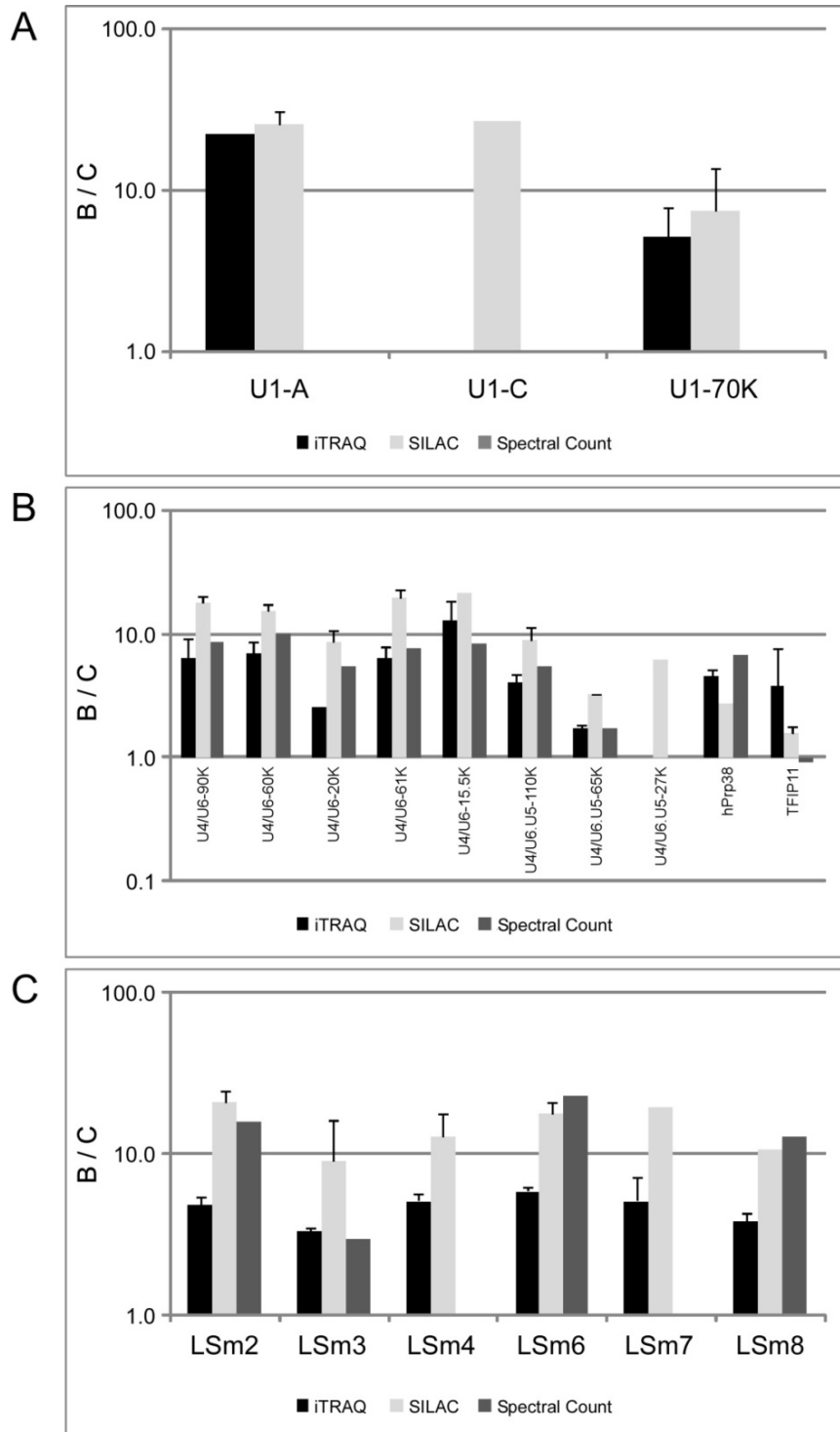


Figure 4.23: Relative protein abundances of the U1, U4/U6, U4/U6.U5 snRNP specific proteins, and LSM proteins obtained by iTRAQ, SILAC, and spectral count. (A) Protein ratios for U1 snRNP specific proteins. No protein ratios were obtained from spectral count as these proteins were totally absent in C complex. **(B)** Protein ratios for U4/U6 and U4/U6.U5 snRNP specific proteins. **(C)** Protein ratios for LSM proteins. For LSM7, no protein ratio was obtained from spectral count as this particular protein was totally absent in C complex.

The U2 snRNP related proteins are without exceptions highly enriched in B complex. Protein ratios obtained by iTRAQ and SILAC are between 1.5 and 12 indicating their high abundance in B complexes (Tables 4.10 and 4.11). During proteomic analysis, U2 snRNP related proteins were only identified in B complex (with the exception of hPrp43, Table 4.9).

In addition to the snRNP specific proteins, several other proteins are known to be highly enriched in B complex. One group of proteins combined as “proteins recruited to B complexes” were identified in both, B and C complexes (Table 4.9). Of these, only few were highly abundant in B complexes. For MFAP1, RED, hSmu-1, THRAP3 and UBL5 all three methods yielded similarly abundance for the B complex. However, all other proteins show no preferential association to either B or C complexes (see Tables 4.9, 4.10, and 4.11). Furthermore, some other proteins, which were previously detected in A complex (Behzadnia *et al.*, 2007) are found to be highly abundant in B complex. These proteins were never found in C complex.

The SR and hnRNP proteins are found to be more abundant in B complex. However, according to iTRAQ and spectral count, the SR proteins do not show high protein ratios (Table 4.9 and 4.10) whereas SILAC analysis yielded relatively high protein ratios (between 3 and 6) for some of the SR proteins (see Table 4.11). Most of the hnRNP proteins are according to all three quantification approaches highly abundant in B complexes. Exceptions are hnRNP R and RALY, which show a 1:1 ratio for B and C complexes obtained by iTRAQ, SILAC, and spectral count.

4.3.5.3 Proteins predominantly associated with C complex

The hPrp19/CDC5L protein complex is essential for pre-mRNA splicing (Ajuh *et al.*, 2000). It associates with the spliceosome during its activation prior to the first catalytic step of splicing. Together with U5 snRNP it forms a remodeled 35S U5 complex (Makarov *et al.*, 2002). In previous studies characterizing the precatalytic A and B complexes (Behzadnia *et al.*, 2007; Deckert *et al.*, 2006), the hPrp19/CDC5L complex proteins have been already detected in A and B complexes. In this study, most of the hPrp19/CDC5L proteins show a higher abundance in C complex (Figure 4.24). An exception is CTNNBL1, which does not show a clear protein ratio. For this particular protein, an enrichment in the C complex was found by SILAC whereas iTRAQ and spectral count revealed a slightly higher abundance in B complex. The two proteins Npw38 and Npw38BP appear to be present predominantly in B complex. This is evidenced by the high protein ratios obtained by iTRAQ and SILAC (Figure 4.24). During proteomic analysis, these two proteins were only detected in B complex, making it impossible to determine the protein ratios for spectral count. Both these proteins

were co-isolated with the hPrp19/CDC5L complex but were found on top when applied to a glycerol gradient (Makarova *et al.*, 2004). Their enrichment in B complex thus corroborates their dissociation from the hPrp19/CDC5L complex. The proteins related to the hPrp19/CDC5L complex are no exception and show a high abundance to C complex (Tables 4.9, 4.10, and 4.11).

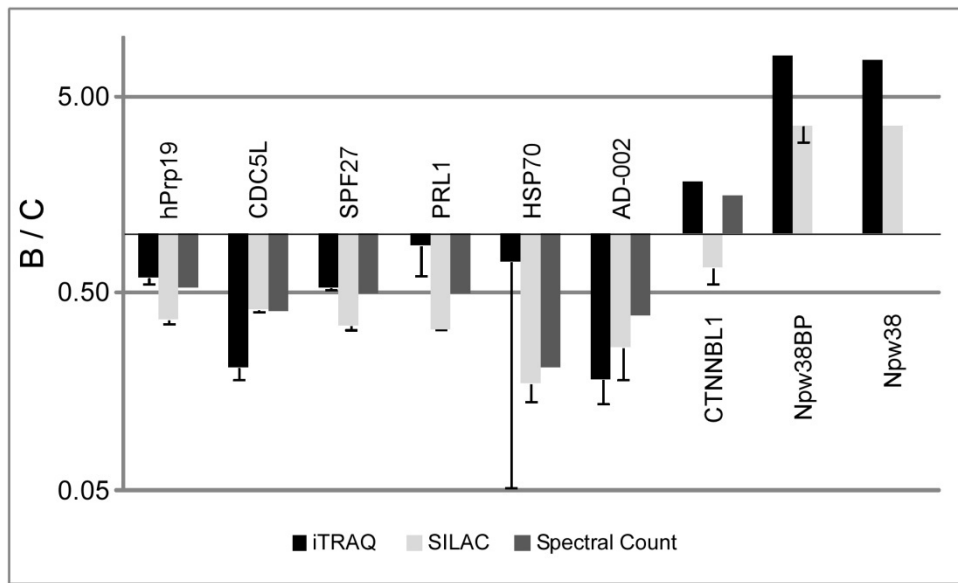


Figure 4.24: Protein ratios for hPrp19/CDC5L complex proteins obtained by iTRAQ, SILAC, and spectral count. Almost all proteins show a clear enrichment in C complex (protein < 1). For CTNNB1 no clear association to B or C complexes could be determined. Npw38BP and Npw38 show a high abundance in the B complex. For these proteins no protein ratio for spectral count could be calculated as both proteins were solely detected in B complex during proteomic analysis.

All proteins required for the second step of splicing (so-called Step 2 factors, namely hPrp22, hPrp18, hPrp17, hPrp16 and hSLU7) show a B/C ratio of less than 0.5 suggesting their association to the catalytically active spliceosome (C complex). This is in agreement with their function in the second catalytic step during pre-mRNA splicing, which occurs in the C complex. Comparing protein ratios obtained by iTRAQ, SILAC, and spectral count, iTRAQ and SILAC revealed protein ratios between 0.5 and 0.1 (Figure 4.25). The spectral count yielded very low B/C ratios (close to 0) for the step 2 factors, consistent with their function in C complex (see also Table 4.9).

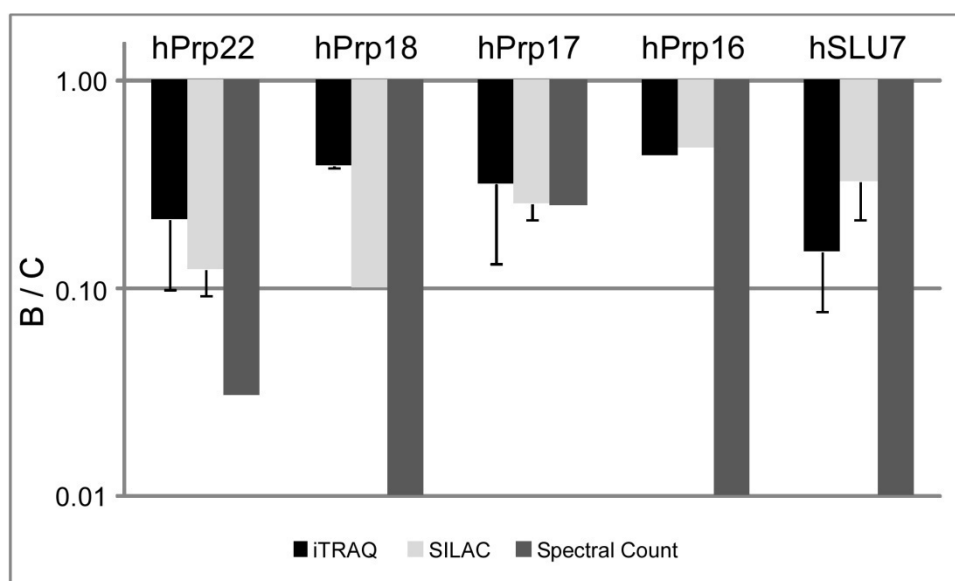


Figure 4.25: Protein ratios of step 2 factors obtained by iTRAQ, SILAC, and spectral count. iTRAQ and SILAC protein ratios are between 0.5 and 0.1, whereas spectral count yielded very low protein ratios close to 0.

The EJC proteins eIF4A3, Magoh, and Y14 show B/C ratios of less than 0.4 with all three quantification approaches (iTRAQ, SILAC, and spectral count) indicating that they are recruited to the spliceosome during C complex formation. However, the other EJC/mRNP proteins (Pinin, RNPS1, Acinus, SAP18, and Aly) show a B/C ratio of approximately 1 or slightly higher, implying that they are present in both B and C complexes. UAP56 yielded a protein ratio of approximately 3 by iTRAQ and SILAC, indicating that it is specific for B complex (Tables 4.9, 4.10, and 4.11).

In addition to the above discussed protein groups previously classified as “recruited to C complex” and “potential C complex specific proteins” (Bessonov *et al.*, 2008) were quantified in this study. These proteins display B/C ratios below 0.7 in all three quantification methods (Tables 4.9, 4.10, and 4.11). Comparing the two stable isotope labeling quantification methods (iTRAQ and SILAC), iTRAQ protein ratios are slightly higher (compare Tables 4.10 and 4.11). Spectral count from proteomic analysis clearly shows, that “potential C specific proteins” are only detected in C complex (with the exception of one peptide generated of PPIG identified in one replicate of the B complex analyses). Some representatives of the “proteins recruited to C complex” are represented with few peptides also in B complex (Table 4.9), but still show a clear association with C complex.

For all other groups of proteins (e.g. TREX, miscellaneous proteins etc.) no clear association with either the B or the C complex could be found. Proteins within these groups differ in their abundances in B or C complexes. Interestingly, the DDX34 protein has not been found in any spliceosomal complex before, but was in this study clearly identified to be highly enriched in

C complex by iTRAQ (see Tables A.5 and A.6) and SILAC. During proteomic analysis of the C complex, DDX34 was identified with 14 and 3 tandem MS spectra within the two replicates, respectively, whereas no DDX34 peptides were found in the B complex.

4.3.6 Comparison of the three quantification methods – iTRAQ, SILAC, and spectral count

In general, quantitative information about protein abundances within spliceosomal B and C complexes could be obtained from all three quantification approaches. The protein ratios achieved by iTRAQ and SILAC yielded for all proteins comparable B/C ratios with only few exceptions (e.g. CTNNBL1, Figure 4.24). However, the B/C ratios obtained by SILAC are in most cases slightly higher or slightly lower for proteins enriched in B or C complexes, respectively, when compared to iTRAQ. This phenomenon can be explained by background signals of iTRAQ reporter ions due to the width of the precursor selection window for MS/MS fragmentation. This is not 100 % selective for one precursor and co-eluting peptides might contribute to the iTRAQ reporter signals (Bantscheff *et al.*, 2007). Background signals for iTRAQ reporter ions thus cause slightly lower protein ratios for proteins that are enriched in B complexes and slightly higher values for proteins that are enriched in C complexes.

Comparing iTRAQ and SILAC to semi-quantitative spectral count, in almost all cases the correct abundance trend was already observed by the spectral count procedure. Most of the B/C ratios fit reasonably to the values obtained by the stable isotope labeling procedures (i.e. iTRAQ and SILAC). However, spectral count appears to be limited to proteins showing major changes within the two states that were compared. For proteins that are present in equal amounts in the two complexes, SILAC and iTRAQ yielded accurate protein ratios of 1 in most cases, whereas protein ratios obtained by spectral count deviated within a narrow range. In addition, accurate quantification of the Sm proteins could not be achieved by spectral count. The correct protein ratios were, however, readily obtained by iTRAQ and SILAC. For proteins that were only detected in either the B or the C complexes, spectral count cannot yield the B/C ratios. This suggests total absence of a particular protein in one of the compared complexes but might not reflect the actual association of the protein with the analyzed complex. Thus a peptide is not selected for sequencing during MS/MS, it is not necessarily absent in the sample to be analyzed. Therefore some of the extreme values obtained by spectral count might be misleading in terms of relative protein abundances in the different samples.

Finally, a correlation analysis of iTRAQ and SILAC was performed to see, if the obtained protein ratios are comparable. To this end, protein ratios achieved by iTRAQ and SILAC were \log_2 transformed and plotted against one another. The scatter plot confirms the observation that iTRAQ shows lower or higher protein ratios for proteins enriched in B or C complexes, respectively. The correlation of SILAC and iTRAQ protein ratios was tested by calculating the Pearson correlation coefficient using software R. A correlation factor of 0.815 and a p-value of $2.2 \cdot 10^{-16}$ were determined showing that a correlation between these two data sets exists.

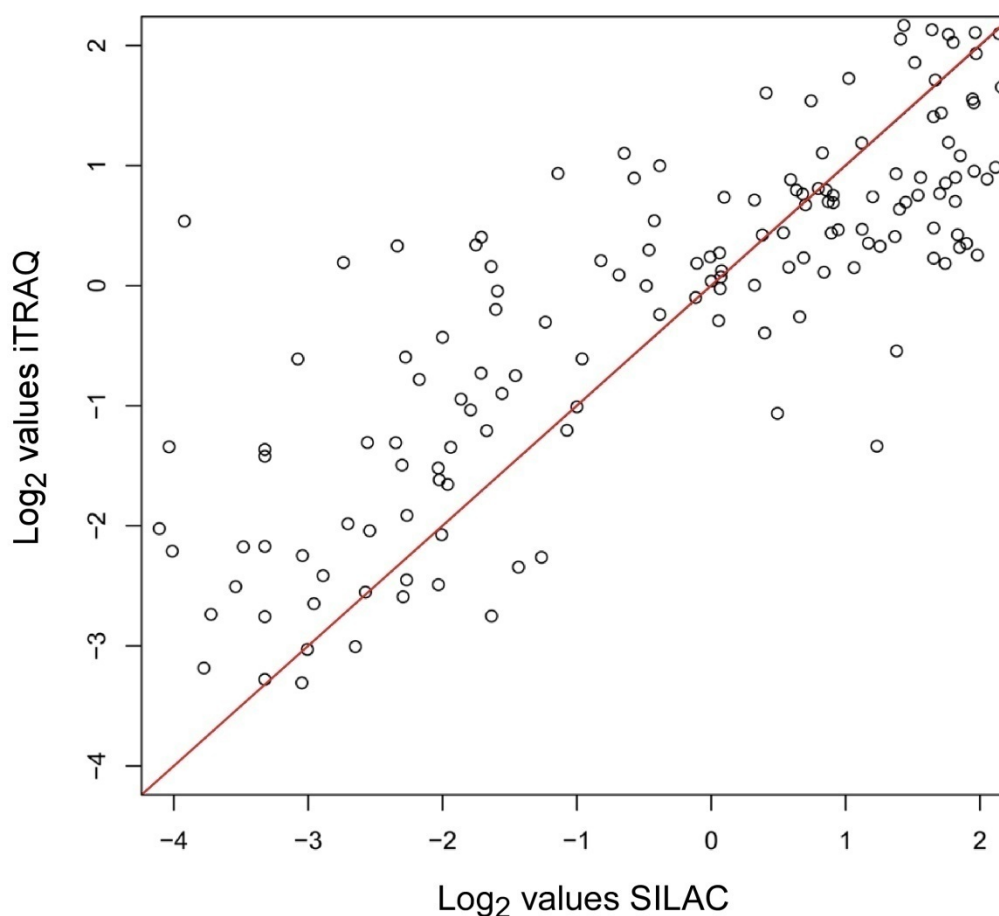


Figure 4.26: Scatter plot of SILAC and iTRAQ protein ratios. The \log_2 values of iTRAQ and SILAC protein ratios, respectively, were plotted in a scatter plot. The red line shows optimal correlation between the two quantification methods. As described above, lower or higher iTRAQ ratios for proteins enriched in B or C complexes, respectively, were obtained when compared to SILAC values.

For the lack of meaningful B/C ratios in cases where the protein was missing in either the B or the C complex, the spectral count data could not be included in the statistical analysis.

4.4 Protein assembly time line for spliceosomes by relative quantification

During pre-mRNA splicing the spliceosome assembles on the pre-mRNA and passes sequentially through functional states that differ in their protein and RNA compositions (E, A, B, and C complexes). The relative quantification of the isolated spliceosomal complexes shows differences in the protein abundances between the different assembly states, but does not reveal how the transition from one complex to the other takes place. Mass spectrometry-based relative quantification can also be applied to monitor dynamic protein changes. Triple SILAC allows comparison of three different samples in one MS experiment and thus provides the opportunity to compare three different time points. Performing multiple experiments using the zero time point as a reference allows the construction of dynamic profiles for single proteins. In this study, the protein assembly on splicing-active and splicing-inactive pre-mRNAs was analyzed and the protein compositions at different time points were compared. Two approaches were followed: (i) Comparison of the time-dependence of protein assembly on splicing-active or splicing-inactive pre-mRNAs, and (ii) the direct comparison of the protein composition on a splicing-active to a splicing-inactive pre-mRNA at different time points during pre-mRNA splicing.

4.4.1 Generation of splicing inactive pre-mRNAs

The protein assembly was investigated on splicing-active and splicing-inactive pre-mRNAs. The PM5 pre-mRNA was used as splicing-active pre-mRNA. It lacks the 3' splice site and the 3' exon and has been shown to be well-suited to isolate catalytically active step 1 C complexes (Bessonov *et al.*, 2008). When incubated under splicing conditions with HeLa nuclear extract, PM5 undergoes 5' splice site cleavage and formation of the intron lariat but no exon ligation takes place.

Two splicing-inactive pre-mRNAs were considered for comparison of the protein assembly on splicing-active (PM5) and splicing-inactive pre-mRNAs: (i) PM5 pre-mRNA lacking the 5' splice site (5'ss) and, (ii) PM5 pre-mRNA lacking the branch point site (BPS). As the U1 and U2 snRNP bind to the pre-mRNA's 5'ss and BPS, respectively, during the formation of A complex, the use of pre-mRNAs from which the 5'ss and BPS have been deleted will hamper the spliceosomal assembly. These pre-mRNAs are therefore splicing-inactive and may be expected to prevent spliceosomal complex formation.

The sequence of the PM5 plasmid is shown in Figure 4.27. The nucleotide sequence AGGTATGT or ACTGA (highlighted in red) was deleted to generate 5'ss- or BPS-deleted PM5 pre-mRNAs, respectively. Generation of the pre-mRNAs was performed as outlined in

section 3.2.1.11 using the PM5 plasmid as template. Two PM5 pre-mRNAs, one lacking the 5'ss and one lacking the BPS, respectively, were obtained.

MS2

GAATACAAGCTCATCCGATATCC**GTACACCATCAGGGTACGA**GCTAGCCCATGG

MS2 **MS2** EcoR1

CGTACACCATCAGGGTACGACTAGTAGATCT**CGTACACCATCAGGGTAC**GGAAT

Exon1

TCTCTAGAG**TCGAGGAGGACATCTCAGCAAAGAGAAGCTGCTGCGGGCGTCGG**

AGGACGAGCGGGACCGGGTGCTGGAGGAGCTGCACAAGGCAGAGGACAGCCTGC

↓ **5' ss** Hind3

TGGCTGCCGACGAGACCGCCGCCAAGGTATGTATCAAGCTTACAAGACAGCTTT

Xho1

AAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGCC**TCGA**

GAAACCTGTA**ACTGGAATGTGTGTGGAGTGTGACTGAT**AGAACACTACCTGATT

BPS (Y_n)

CTTATGTATTT**ACTGA**CCTGTG**TTTTTTTGCTACTTTTTTTCTTTTCTCCCTT**

BamH1

CCCCTTCCCTATTTTTTTTCTTGCCCTGATCCGGAATTTGGATCC

Figure 4.27: Sequence of the PM5 plasmid. MS2 binding sites (MS2), exon 1, branch point site (BPS) and polypyrimidine tract (Y_n) are shown in bold. Restriction sites of DNA restriction enzymes are shown in blue. Deleted nucleotide sequences are shown in red.

The kinetics of splicing and spliceosomal complex formation were investigated for PM5 pre-mRNA and 5'ss- and BPS-deleted PM5 pre-mRNAs (Figure 4.28). When PM5 pre-mRNA was used, splicing products (intron lariat and MS2-exon) first appeared after 10 minutes. The amount of pre-mRNA was reduced during incubation. As RNaseH digestion leads to degradation of early spliceosomes, the pre-mRNA disappeared and only the splicing products and an RNaseH digestion product were visible after RNaseH digestion. For PM5 5'ss-deleted pre-mRNA, no splicing products were formed, and the amount of pre-mRNA remained constant. After 180 minutes incubation followed by RNaseH digestion a digestion product was observed. However, using BPS-deleted PM5 pre-mRNA the appearance of splicing products after 30 minutes and 180 minutes followed by RNaseH digestion was indicated (Figure 4.28 A). Spliceosomal complex formation was assayed by native agarose gel electrophoresis, and the complexes formed were visualized by autoradiography. On PM5

pre-mRNA, H/E complexes were rapidly formed, whereas A and B complexes first appeared after 2-4 minutes. C complex formation was first observed after 10 minutes. Surprisingly, A and B complex formation was also observed on PM5 5'ss deleted pre-mRNA, although B complex formation seemed to be delayed about 2 minutes. However, C complex formation was not observed on the 5'ss-deleted PM5 pre-mRNA. Complex formation on BPS-deleted PM5 pre-mRNA suggests also formation of A and B complexes. Strikingly, after 180 minutes incubation followed by RNaseH digestion the presence of C complexes was indicated, whereas this is not clearly seen (Figure 4.28 B).

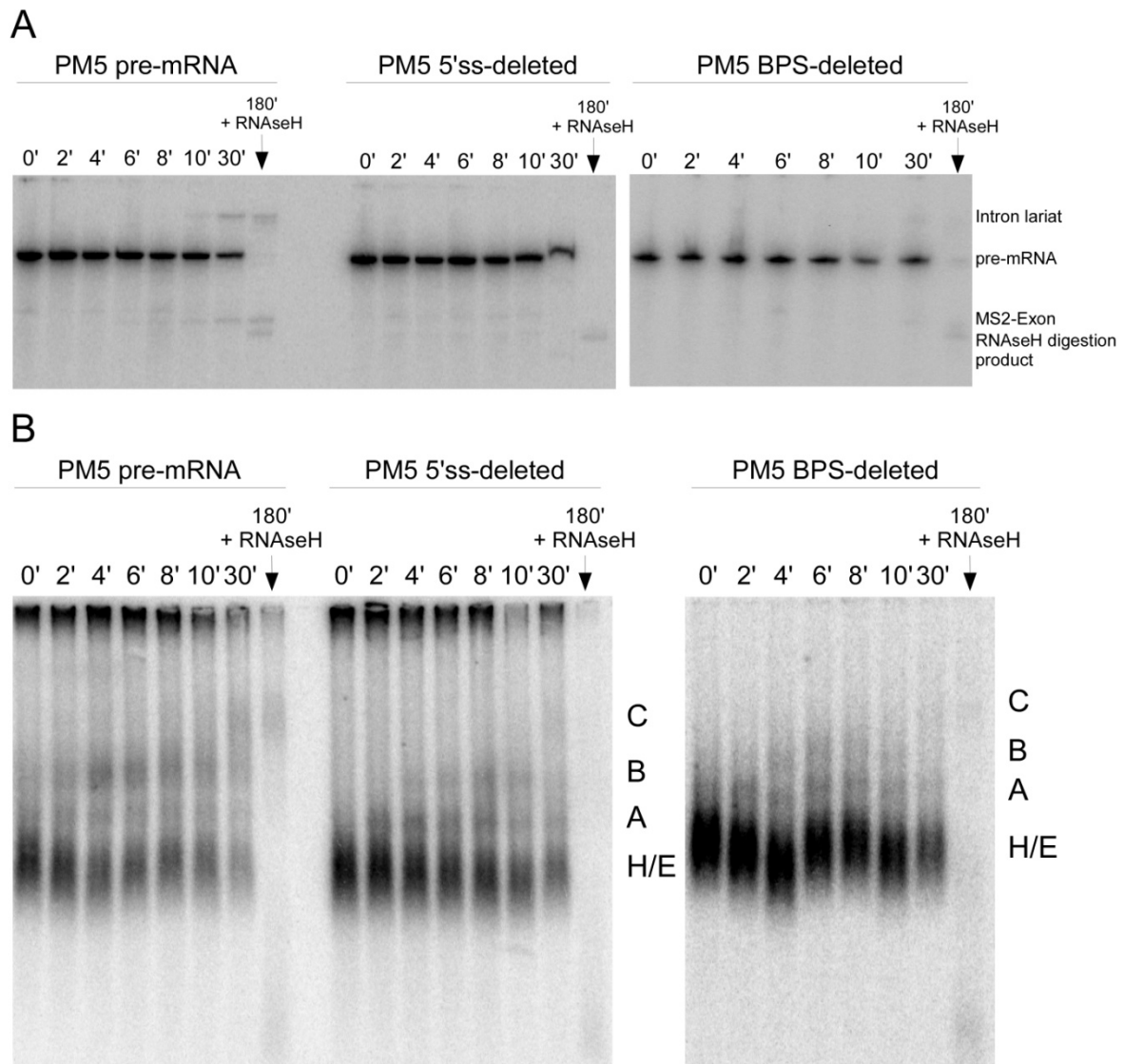


Figure 4.28: The splicing kinetics and the spliceosomal complex formation using PM5 pre-mRNA and 5'ss- and BPS-deleted PM5 pre-mRNA. (A) Splicing kinetics were followed by denaturing gel electrophoresis. Radioactively labeled pre-mRNA and splicing products were visualized by autoradiography. With PM5 pre-mRNA, splicing products first appeared after 10 minutes and the amount of pre-mRNA was reduced during incubation. With 5'ss-deleted PM5 pre-mRNA, no splicing products were observed, whereas the appearance of splicing products after 30 minutes was indicated for BPS-deleted PM5 pre-mRNA. **(B)** Spliceosomal complex formation was assayed by native agarose gel electrophoresis, and the complexes formed were visualized by autoradiography. On PM5 pre-mRNA, the formation of H/E, A, B and C complexes was observed. Surprisingly, A and B complex formation was also observed on 5'ss-deleted PM5 pre-mRNA. On BPS-deleted PM5 pre-mRNA, A and B complex formation was indicated and after 180 minutes incubation followed by RNaseH digestion C complex formation was suggested.

The splicing kinetics revealed that no pre-mRNA splicing occurred when 5'ss-deleted PM5 pre-mRNA was used. Surprisingly, A and B complex formation was observed, even though the 5'ss, with which the U1 snRNP is known to make contact during A complex formation, was deleted. Nonetheless, C complex formation was not detected, confirming that this pre-mRNA is indeed splicing-inactive. Strikingly, splicing products appeared after 30 and 180 minutes when using BPS-deleted PM5 pre-mRNA. During complex formation the presence of C complex after RNaseH digestion is indicated, although this is not clearly visible. Inspection of the PM5 plasmid sequence revealed that the deleted nucleotide sequence (ACTGA) exists once more in the PM5 plasmid. This duplicate is located 28 nucleotides downstream of the branch point site within the intron of the PM5 pre-mRNA (see Figure 4.27) and might serve as an alternative branch point. This would explain the presence of splicing products that were observed when the splicing kinetics and complex formation were analyzed.

For this reason, an additional pre-mRNA was produced in which the duplicate BPS sequence was also deleted (in the following referred to as "BPS-ACTGA-deleted PM5 pre-mRNA"). The splicing kinetics and the spliceosomal complex formation were then investigated for the BPS- and the BPS-ACTGA-deleted pre-mRNA (Figure 4.29). As described above, when PM5 pre-mRNA was used, splicing products were observed after 10 minutes and the amount of pre-mRNA present decreased during the incubation. The splice assay for BPS-deleted PM5 pre-mRNA again revealed the presence of splicing products after 30 and 180 minutes. The "double-deleted" pre-mRNA (the BPS-ACTGA-deleted PM5 pre-mRNA) did not show splicing products, confirming that this pre-mRNA mutant is splicing-inactive (Figure 4.29 A). Spliceosomal complex formation for PM5 pre-mRNA again showed H/E, A, B and C complex formation. As described above, A and B complex formation was also observed for the BPS-deleted PM5 pre-mRNA; moreover C complex formation after 30 and 180 minutes was suggested, although it was not clearly visible. For the BPS-ACTGA-deleted PM5 pre-mRNA no comparable complex formation was achieved. However, native agarose gel electrophoresis indicated the formation of some complexes on this particular pre-mRNA (Figure 4.29 B). The complexes formed differed in their migration behavior in native agarose gel electrophoresis and thus appear to have a different RNA and/or protein composition as compared with the spliceosomal A, B and C complexes. Nonetheless, as the "double-deleted" pre-mRNA (BPS-ACTGA-deleted PM5 pre-mRNA) did not lead to splicing products and did not show C complex formation, it is splicing-inactive. Therefore, in this study the PM5 pre-mRNA and the 5'ss- and the BPS-ACTGA-deleted PM5 pre-mRNAs were used for comparison of the protein assembly on splicing-active and splicing-inactive pre-mRNAs.

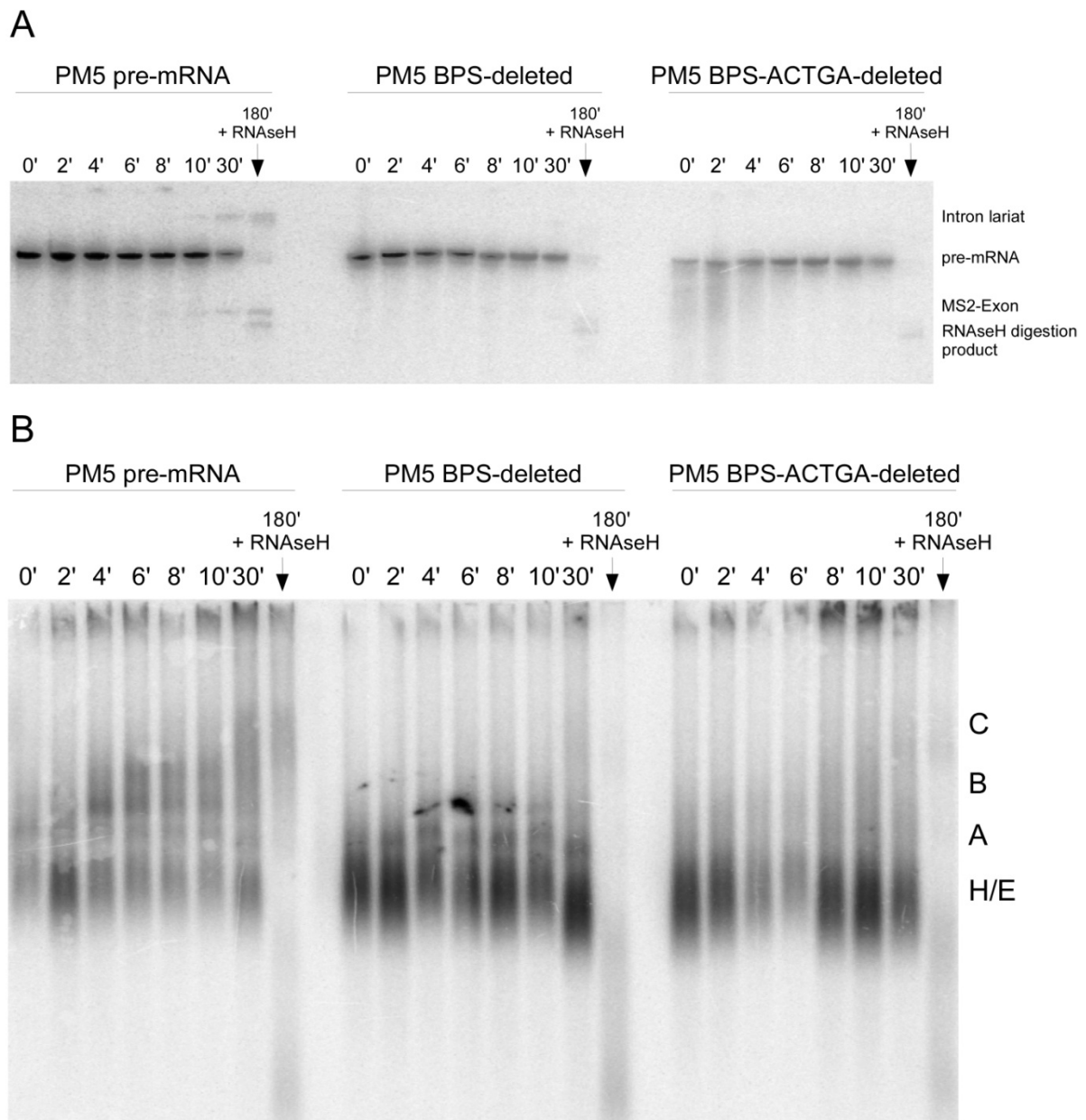


Figure 4.29: The splicing kinetics and the spliceosomal complex formation using PM5 pre-mRNA and BPS and BPS-ACTGA deleted PM5 pre-mRNAs. (A) Splicing kinetics were followed by denaturing gel electrophoresis. Pre-mRNA and splicing products were visualized by autoradiography. With PM5 pre-mRNA, splicing products first appeared after 10 minutes and the amount of pre-mRNA was reduced during incubation. With the BPS-deleted pre-mRNA splicing products had appeared after 30 minutes. For the “double-deleted” BPS-ACTGA-deleted PM5 pre-mRNA no splicing products were observed. **(B)** Spliceosomal complex formation was assayed by native agarose gel electrophoresis and the complexes formed were visualized by autoradiography. On PM5 pre-mRNA, the formation of H/E, A, B and C complexes was observed. On BPS-deleted PM5 pre-mRNA, A and B complex formation was detected, and after 180 minutes incubation followed by RNaseH digestion some C complex formation was detected. For BPS-ACTGA-deleted PM5 pre-mRNA no comparable complex formation was observed, whereas formation of some complexes is indicated.

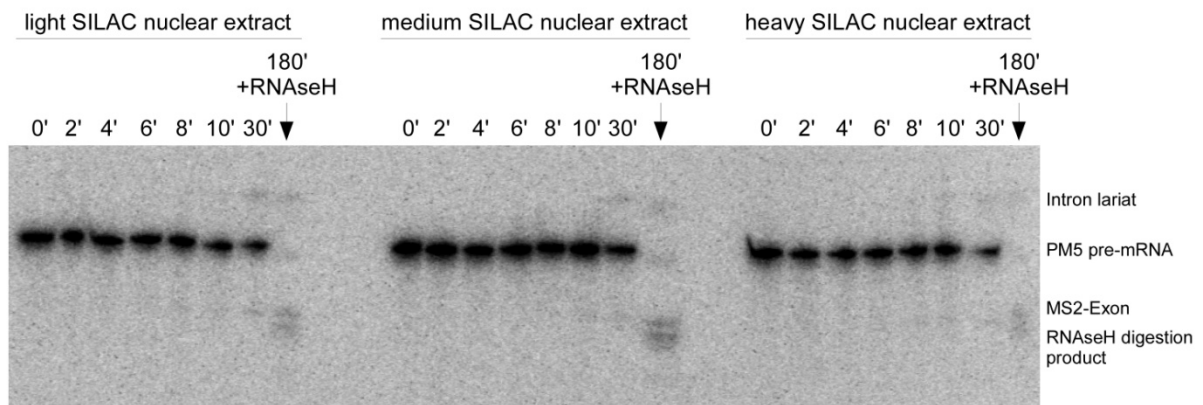
4.4.2 Triple SILAC to monitor dynamic protein changes

Triple SILAC allows quantitative comparison of three samples in one MS experiment and thus comparison of three time points during dynamic changes of protein composition. Using the zero time point as a reference point allows the construction of assembly timelines when performing multiple experiments. To make possible triple SILAC experiments for the investigation of the spliceosomal dynamic protein changes, triple SILAC nuclear extracts were prepared. To this end, HeLa cells were grown in the presence of isotope-labeled lysine and arginine. The combination of different isotope-labeled lysine and arginine allows generation of three differentially labeled cells (light, medium and heavily labeled cells; for experimental details see section 3.2.3.1). Nuclear extracts were then prepared and tested for their splicing activity. Figure 4.30 shows the splicing kinetics and the spliceosomal complex formation for PM5 pre-mRNA using light, medium and heavy SILAC nuclear extracts.

For all nuclear extracts (i.e. light, medium and heavy SILAC nuclear extracts) splicing products were first observed after 10 minutes. The intensity of the splicing products in heavy SILAC nuclear extracts is very low compared to that of the splicing products generated in light and medium extracts. After RNaseH digestion, pre-mRNA had been degraded and the splicing products - as well as an RNaseH digestion product - were visible (Figure 4.30 A). The formation of spliceosomal complexes was assayed by native agarose gel electrophoresis. The yields of complex formation for the three SILAC nuclear extracts were comparable. H/E complexes were formed rapidly, and A complexes first appeared after approximately 2 minutes. B complexes were formed after 2-4 minutes, and C complex formation was first observed after 10 minutes. After RNaseH digestion, early spliceosomes were degraded and only C complexes remained. However, B complex formation seemed to be delayed by about 2 minutes when medium and heavy nuclear extracts were compared to light nuclear extract (Figure 4.30 B).

Although the intensity of the splicing products was very low when heavy SILAC nuclear extract was used, the three nuclear extracts were considered to be identical in splicing activity, as C complex formation was clearly observed. The low intensity of the splicing products might have been caused by sample loss during sample preparation, or by sample loading for gel electrophoresis. Nonetheless, it was inferred the delayed B complex formation should be kept in mind when performing time-dependent experiments.

A



B

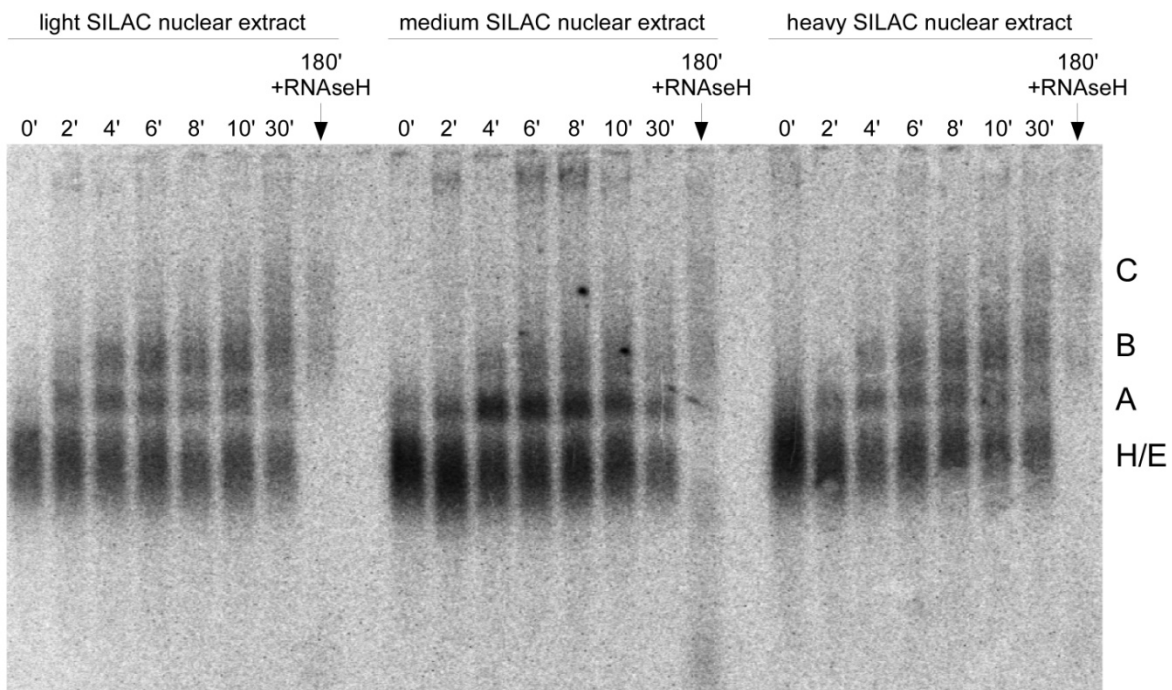


Figure 4.30: Splicing kinetics and spliceosomal complex formation using light, medium and heavy SILAC nuclear extracts. The splicing kinetics and the complex formation are comparable for light, medium and heavy SILAC nuclear extracts. **(A)** The splicing kinetics were followed by denaturing gel electrophoresis. Pre-mRNA and splicing products were visualized by autoradiography. Splicing products first appeared after 10 minutes incubation for all SILAC nuclear extracts. However, the intensity of the splicing products obtained from heavy SILAC nuclear extracts was very low compared with those from light and medium nuclear extracts. **(B)** Spliceosomal complex formation was assayed by native agarose gel electrophoresis and visualized by autoradiography. A and B complex formation was observed after 2-4 minutes, whereas C complexes were first observed after 10 minutes. B complex formation seems to be delayed about 2 minutes in medium and heavy nuclear extracts compared with light nuclear extract.

As the protein dynamics during pre-mRNA splicing were investigated in this study, the nuclear extracts also had to be compared regarding their protein content. The protein content within the nuclear extracts should be comparable, in order to exclude influences on the spliceosomal protein assembly during pre-mRNA splicing.

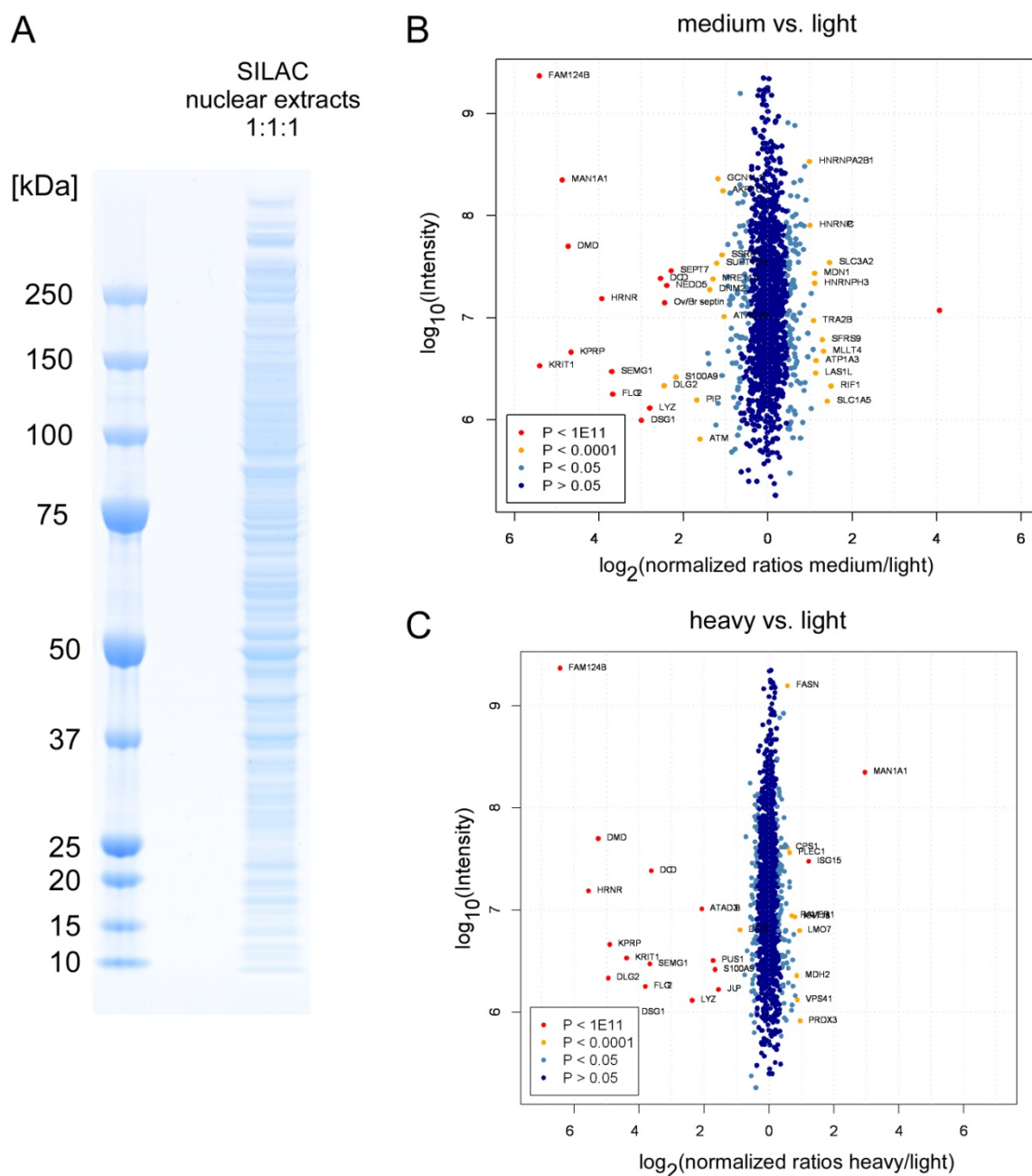


Figure 4.31: Comparison of the protein content within light, medium and heavy SILAC nuclear extracts. (A) Equal amounts of SILAC nuclear extracts (light, medium and heavy) were pooled and proteins were separated by PAGE. (B) Comparison of medium and light nuclear extract. Intensities were plotted against the \log_2 protein ratios showing that most of the proteins are of similar abundance within the two nuclear extracts and only few proteins were enriched in light nuclear extract. (C) Comparison of heavy and light nuclear extract. As described for comparison of medium and light nuclear extracts, only few proteins were enriched in light nuclear extract.

To compare the protein content of the different SILAC nuclear extracts, the total protein concentrations were obtained by using the Bradford protein assay. Protein concentrations of 16.14, 15.08 and 14.05 $\mu\text{g}/\mu\text{l}$ were found for the light, medium and heavy nuclear extract, respectively. Equal amounts of proteins from the light, medium and heavy nuclear extracts were pooled, and the proteins were separated by PAGE (Figure 4.31 A). The entire gel lane

was excised and the proteins were hydrolyzed with trypsin. The peptides thus generated were analyzed by LC-MS/MS and peptides and proteins were subjected to relative quantification quantified using MaxQuant software (Cox and Mann, 2008). The protein ratios obtained (medium/light and heavy/light) were \log_2 transformed and the intensities (\log_{10}) were plotted against the \log_2 protein ratios (Figure 4.31 B and C). A \log_2 protein ratio of 0 represents no difference in the protein abundance between the samples compared. A \log_2 ratio of 1 shows a twofold enrichment, whereas a \log_2 ratio of -1 shows that the protein amount is halved compared with the other sample. Figure 4.31 B shows the comparison of the medium with the light nuclear extract. Most of the proteins show a \log_2 protein ratio of 0, i.e. most of the proteins are present in equal amounts in the two nuclear extracts. Only few proteins were found to be enriched in the light nuclear extract. However, these proteins are in most cases contaminating proteins (e.g. Hornerin, Filaggrin) and they are not involved in pre-mRNA splicing. Figure 4.31 C shows the comparison of \log_2 protein ratios between heavy and light nuclear extracts. As described for comparison of medium and light nuclear extracts, most of the proteins were of similar abundance in the heavy and the light nuclear extract. Again, proteins that were enriched in the light nuclear extract are not involved in pre-mRNA splicing and are therefore not expected to affect the incorporation of proteins into the spliceosome. An overview of the proteins that were enriched in one of the SILAC nuclear extracts is given in Table A7 in the Appendix.

In summary, the SILAC nuclear extracts show no differences in their protein content and show the same splicing activity when incubated with pre-mRNA under splicing conditions. They are therefore expected to show the same response when analyzing the spliceosomal protein assembly at different time points. Changes between the different time points to be compared can thus be concluded to reflect differences in the spliceosomal assembly and are not due to differences between the different SILAC nuclear extracts.

4.4.3 Time-dependent protein assembly on splicing-active and splicing-inactive pre-mRNAs

4.4.3.1 Experimental setup

To investigate the time-dependent protein assembly on different splicing-active and splicing-inactive pre-mRNAs, triple SILAC was used according to the experimental setup described below (Figure 4.32). SILAC nuclear extracts were prepared from differentially labeled SILAC HeLa cells (light, medium and heavy) and spliceosomal assembly was performed by incubation of a MS2-tagged and radioactively labeled pre-mRNA (PM5 pre-mRNA and 5'ss- and BPS-ACTGA-deleted PM5 pre-mRNA) with different SILAC nuclear extracts under

splicing conditions. The splicing reactions were assembled for different time intervals and the assembled complexes were affinity purified (see section 3.2.3.8 for details). According to the radioactivity of the pre-mRNA, the assembled complexes from two different time points were mixed in equal molar amounts with complexes assembled at time point zero (Figure 4.32). To allow monitoring of the time-dependent protein assembly on the different pre-mRNAs, assembled complexes were not further purified by glycerol gradient centrifugation as described in previous studies.

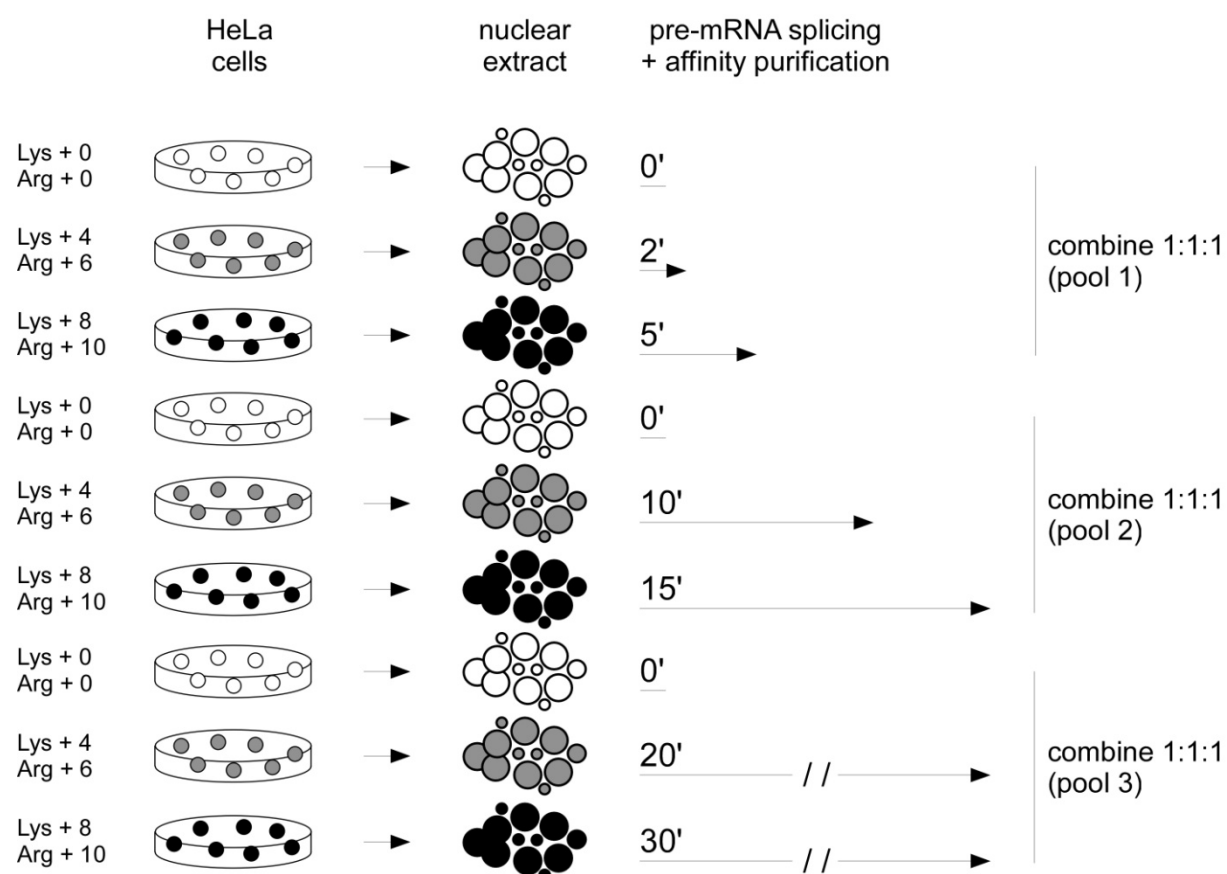


Figure 4.32: Experimental setup to monitor the time-dependent protein assembly during pre-mRNA splicing. SILAC nuclear extracts were prepared from differentially labeled HeLa cells (light, medium, heavy). Spliceosomal protein assembly was performed by incubation of a pre-mRNA (splicing-active or splicing-inactive) with the different nuclear extracts under splicing conditions. Several splicing reactions were assembled for different time intervals. Assembled complexes from two time points were pooled in equal amounts with complexes assembled at time point zero.

Assembled proteins within the combined samples (pools 1–3) were separated by PAGE (Figure 4.33). The MS2-MBP protein which was used for affinity purification shows equal intensity within the combined samples (pools 1–3) for assembly on the different pre-mRNAs confirming the mixing procedure based on the radioactively labeled pre-mRNA. However, significant differences in the protein assembly on the splicing-active and splicing-inactive pre-mRNAs were not observed in the Coomassie stained gels (Figure 4.33).

After separation by PAGE the proteins were hydrolyzed with trypsin, and the peptides generated were subsequently analyzed by LC-MS/MS (see section 3.2.4.12). The peptides and finally the proteins were quantified by using MaxQuant software (Cox and Mann, 2008). Protein ratios for the time point of analysis versus the reference time point were calculated and used for the construction of assembly time lines for individual proteins.

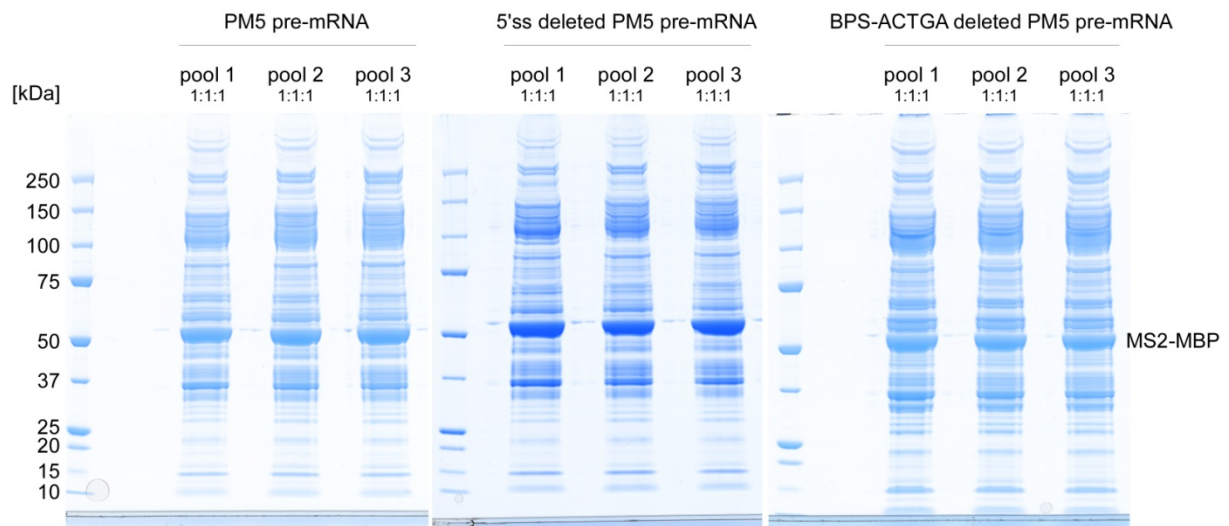


Figure 4.33: Coomassie stained gels of combined samples (pools 1–3) from assembly studies on PM5 pre-mRNA, 5'ss-deleted PM5 pre-mRNA and BPS-ACTGA-deleted PM5 pre-mRNA. Assembled complexes taken from the reaction at different time points were pooled in equal molar amounts with complexes assembled at time point zero and proteins were separated by PAGE.

4.4.3.2 Normalization of the data

Before constructing time-dependent assembly timelines for proteins that assemble on the analyzed pre-mRNAs, the data was normalized to compensate for errors that occurred owing to mixing of the samples or at any other step during sample-handling. The protein ratios obtained were normalized to those found for ribosomal proteins, which are contaminating proteins co-purified during the affinity purification of assembled complexes, and which are expected to be present in equal amounts within all samples. All ribosomal proteins were checked for enrichment in one of the extracts when the different SILAC nuclear extracts were mixed in equal amounts (see above) and only ribosomal proteins that were present in equal amounts within the light, medium and heavy nuclear extract were chosen for normalization (see Table A.8 for a complete list of the ribosomal proteins used for data normalization in this work). The normalization is exemplified for the assembly of proteins on PM5 pre-mRNA (Figure 4.34).

The protein ratios of 19 ribosomal proteins were plotted for every time point (using zero minutes as reference time point) showing the abundance of the ribosomal proteins in protein complexes assembled on PM5 pre-mRNA as a function of time over the interval investigated (Figure 4.34 A). The abundance of the ribosomal proteins is expected to be the same at every time point. Thus, a normalization factor for every time point could be calculated from the average value of the ribosomal protein ratios (Figure 4.34 C). The calculated normalization factors for the different time points were then applied to the protein ratios. Figure 4.34 B shows the assembly of the normalized ribosomal protein ratios, which are the same at every time point. The calculated normalization factors were used to normalize the protein ratios of every protein quantified in the data set. The normalization factors determined for the assembly studies on different pre-mRNAs are listed in Table A9.

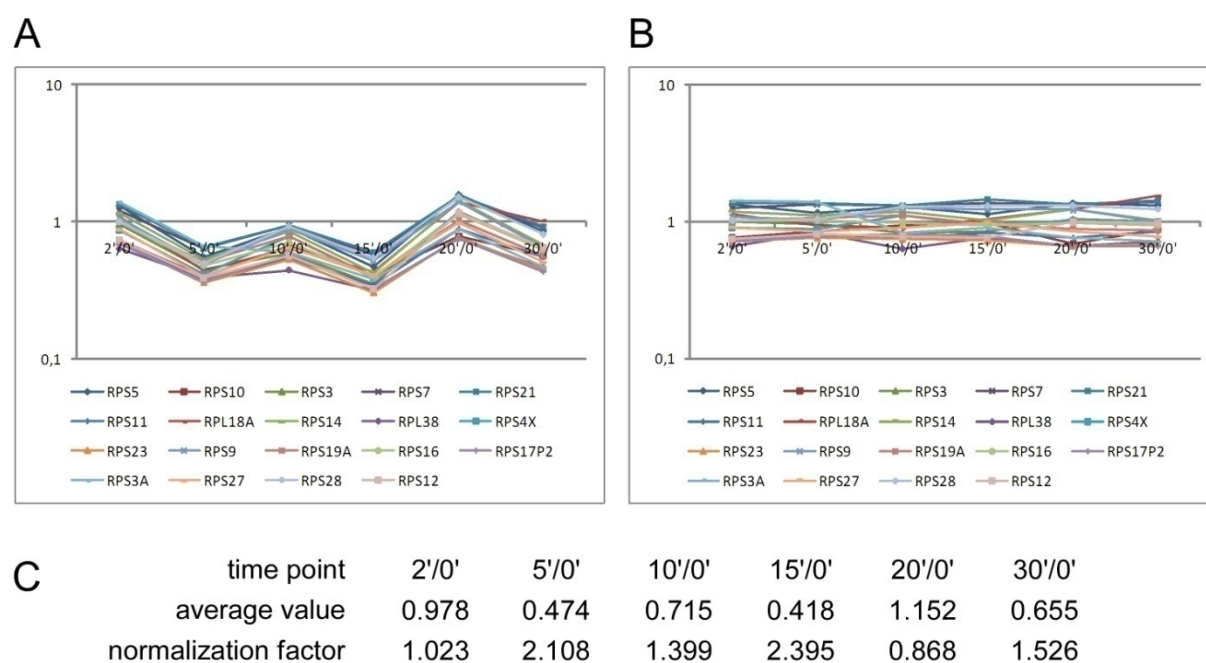


Figure 4.34: Normalization of the obtained protein ratios on the ribosomal proteins. (A) Obtained protein ratios for 19 ribosomal proteins were plotted for every time point using the zero time point as a reference. (B) The normalized protein ratios for the 19 ribosomal proteins were plotted for every time point. (C) For every time point the average protein ratios for the 19 ribosomal proteins and the calculated normalization factors are given.

4.4.3.3 Protein assembly on splicing-active and splicing-inactive pre-mRNAs

The protein assembly during pre-mRNA splicing was investigated on various splicing-active and splicing-inactive pre-mRNAs: (i) PM5 pre-mRNA, (ii) 5'ss-deleted PM5 pre-mRNA, and (iii) BPS-ACTGA-deleted PM5 pre-mRNA (see above). Several splicing reactions were started on the different pre-mRNAs for different time intervals, and the assembled complexes were mixed in equal molar amounts using the time point zero as a reference value (see Figure 4.32 for an overview). After normalization of the protein ratios (see above), time lines

for the different proteins were constructed. In the following section the protein assembly of the U1 and U2 snRNP specific and some C specific proteins, for which an effect during the assembly on the splicing-inactive pre-mRNAs is expected, will be described.

Figure 4.35 shows the protein assembly of the U1 snRNP specific proteins on PM5, 5'ss-deleted PM5 and BPS-ACTGA-deleted PM5 pre-mRNAs over a time of 30 minutes. Inspection of the protein assembly on PM5 pre-mRNA reveals that the U1-A, U1-C and U1-70K protein ratios showed a maximum at 5 minutes (Figure 4.35 A). After 5 minutes the U1-A and the U1-70K proteins showed nearly constant protein ratios, whereas the U1-C assembly time line decreased. On the 5'ss-deleted PM5 pre-mRNA (Figure 4.35 B), the U1-A and U1-70K protein ratios again showed a maximum at 5 minutes and are constant thereafter. However, the U1-C protein, which is known to bind the 5'ss of the pre-mRNA (Heinrichs *et al.*, 1990; Pomeranz Krummel *et al.*, 2009), was only detected at few time points and did not show a time-dependent assembly on this pre-mRNA. The protein ratios obtained for the three U1 snRNP specific proteins during assembly on the BPS-ACTGA-deleted pre-mRNA again showed a maximum at 5 minutes, but thereafter varied substantially with time.

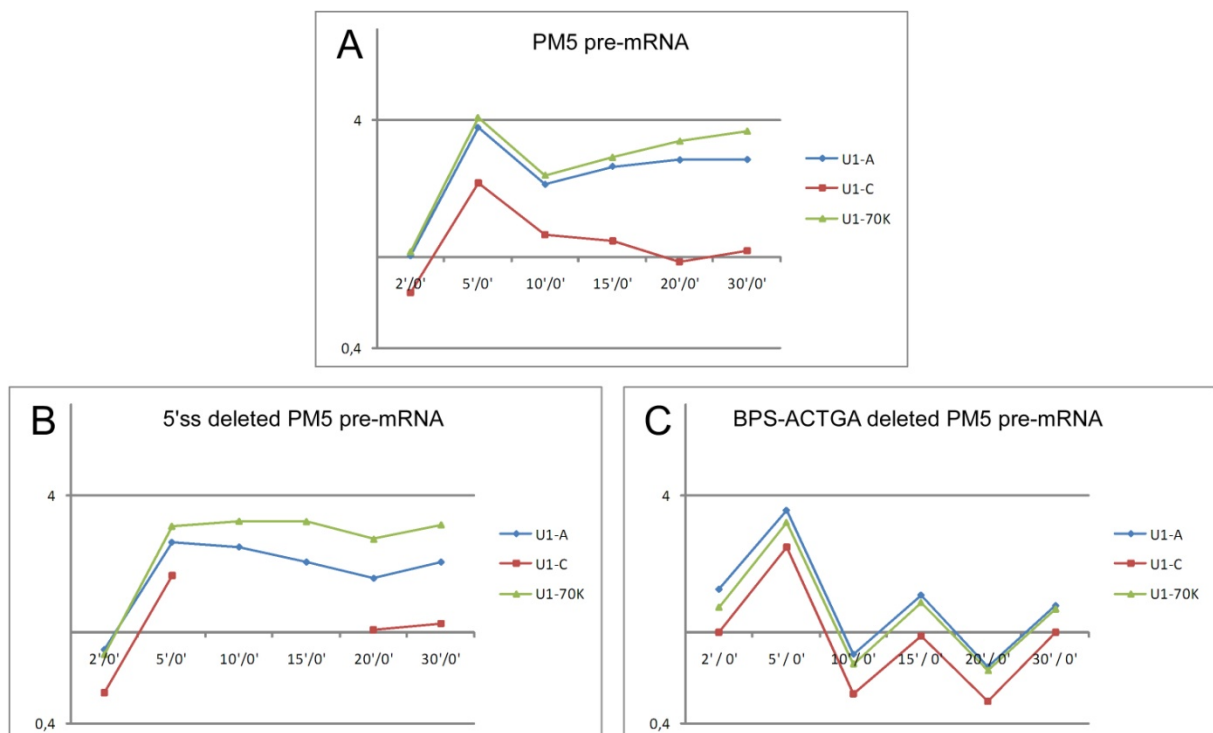


Figure 4.35: Assembly of U1 snRNP specific proteins on PM5 pre-mRNA, 5'ss-deleted PM5 pre-mRNA, and BPS-ACTGA-deleted PM5 pre-mRNA over 30 minutes. (A) Assembly of U1-A, U1-C and U1-70K proteins on PM5 pre-mRNA. **(B)** Assembly of U1-A, U1-C and U1-70K proteins on 5'ss deleted PM5 pre-mRNA. **(C)** Assembly of U1-A, U1-C and U1-70K proteins on BPS-ACTGA deleted PM5 pre-mRNA.

The U2 snRNP is known to bind the BPS of the pre-mRNA during pre-mRNA splicing. As the BPS was deleted in one of the splicing-inactive pre-mRNAs, the assembly of the U2 snRNP specific proteins on the different pre-mRNAs was investigated. Figure 4.36 A shows the protein assembly of U2 snRNP specific proteins on the PM5 pre-mRNA. Constructed time lines for all proteins show the same time course. All proteins were enriched on the PM5 pre-mRNA at the different time points compared with at zero minutes. The protein ratios reached a maximum at 5 minutes and were nearly constant afterwards (Figure 4.36 A). The protein assembly on the 5'ss-deleted PM5 pre-mRNA showed a time course comparable to that of the splicing-active PM5 pre-mRNA. The protein ratios showed a maximum at 5 minutes and were constant thereafter (Figure 4.36 B). Surprisingly, the assembly time course of the U2 snRNP specific proteins on the BPS-ACTGA-deleted PM5 pre-mRNA resembled that of the U1 snRNP specific proteins on this particular pre-mRNA (see above). The protein ratios varied substantially with time and no effect on the binding of the U2 snRNP proteins was observed (Figure 4.36 C). However, the time lines constructed for protein assembly on this pre-mRNA differed from that of the PM5 and 5'ss-deleted PM5 pre-mRNA.

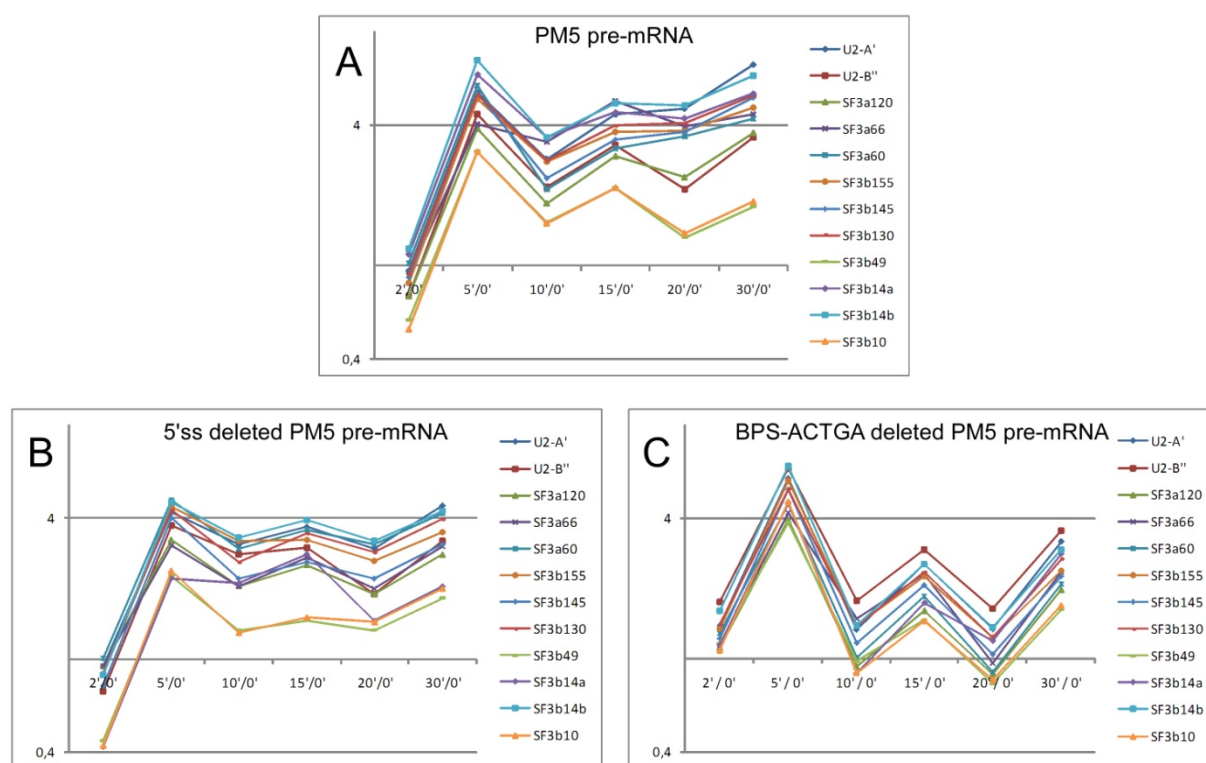


Figure 4.36: Assembly of U2 snRNP specific proteins on PM5 pre-mRNA (A), 5'ss-deleted PM5 pre-mRNA (B) and BPS-ACTGA-deleted PM5 pre-mRNA (C) monitored over 30 minutes.

The hPrp19/CDC5L complex is recruited during spliceosomal activation and its proteins have been found to be enriched in spliceosomal C complex (see section 4.3). However, in earlier studies hPrp19/CDC5L complex proteins were already detected in spliceosomal A and B

complexes (Behzadnia *et al.*, 2007; Deckert *et al.*, 2006). The protein assembly of the hPrp19/CDC5L complex on the splicing-active pre-mRNA is shown in Figure 4.37 A. Indeed, the time courses of all hPrp19/CDC5L proteins showed a maximum at 15 minutes when the spliceosomal C complex was assembled (see above). However, consistently with previous studies, the proteins were found to be enriched at an earlier time point, i.e. the protein ratios showed an initial increase after 5 minutes. Surprisingly, after 15 minutes the protein ratios decreased followed by an increase at 30 minutes. The protein assembly on 5'ss-deleted PM5 pre-mRNA was found to differ for these proteins (Figure 4.37 B). The proteins showed no temporary maximum at 5 minutes but increased continuously over 30 minutes. However, a maximum at 15 minutes was also indicated. In contrast to the protein assembly on PM5 pre-mRNA two proteins, namely AD-002 and CTNNBL1, were first observed on the 5'ss-deleted pre-mRNA after 10 minutes. The protein assembly on the BPS-ACTGA-deleted PM5 pre-mRNA resembled the assembly of the U1 and U2 snRNP specific proteins on this particular pre-mRNA variant (see above, Figure 4.35 C and 4.36 C). It again showed a decreasing and increasing time course with two maxima at 5 and 15 minutes. However, in contrast to the assembly of the U2 snRNP proteins, a maximum was observed at 15 minutes instead of 5 minutes, suggesting that this specific protein group is more highly enriched on the pre-mRNA after 15 minutes (Figure 4.37 C). Proteins ratios for Hsp70, which is an additional component of the hPrp19/CDC5L complex (see above), were not monitored, as for this particular protein several isoforms have been detected and quantified.

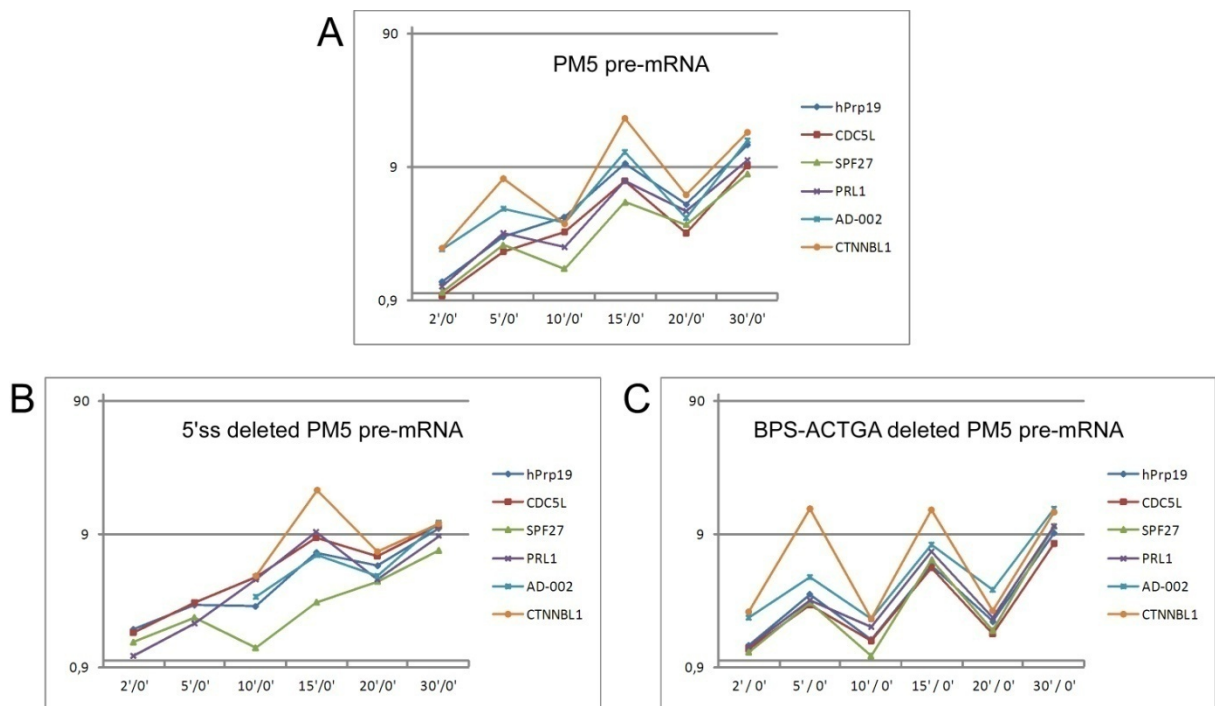


Figure 4.37: Assembly of hPrp19/CDC5L complex proteins on PM5 pre-mRNA (A), 5'ss-deleted PM5 pre-mRNA (B) and BPS-ACTGA-deleted PM5 pre-mRNA (C) monitored over 30 minutes.

The step 2 factors are essential for the second catalytic step of pre-mRNA splicing. By relative quantification of B versus C complexes (see section 4.3), they were found to be highly enriched in C complex. The time course for the assembly of these proteins onto PM5 pre-mRNA indeed shows a maximum at 15 minutes. Before 10 minutes the protein ratios showed almost no increase. As for the hPrp19/CDC5L complex proteins, the protein ratios showed a decrease at 20 minutes that was followed by an increase at 30 minutes. The hPrp17 protein showed a slightly different time course, with protein ratios continuously increasing over 30 minutes (Figure 4.38 A). The protein assembly of the step 2 factors on the 5'ss-deleted pre-mRNA differs from that of the splicing-active PM5 pre-mRNA. The protein ratios (except that for hPrp17) showed almost no increase (at most twofold) before 30 minutes. hPrp17 again showed a continuous increase over 30 minutes. The hPrp16 protein was only detected at two time points (Figure 4.38 B). The assembly of the step 2 factors on the BPS-ACTGA-deleted pre-mRNA was found to be similar to that of the splicing-active PM5 pre-mRNA. A maximum was observed at 15 minutes followed by a minimum at 20 minutes and a second maximum at 30 minutes (Figure 4.38 C). The hPrp18 protein, which also belongs to the step 2 factors, was not detected and is therefore not included in these analyses.

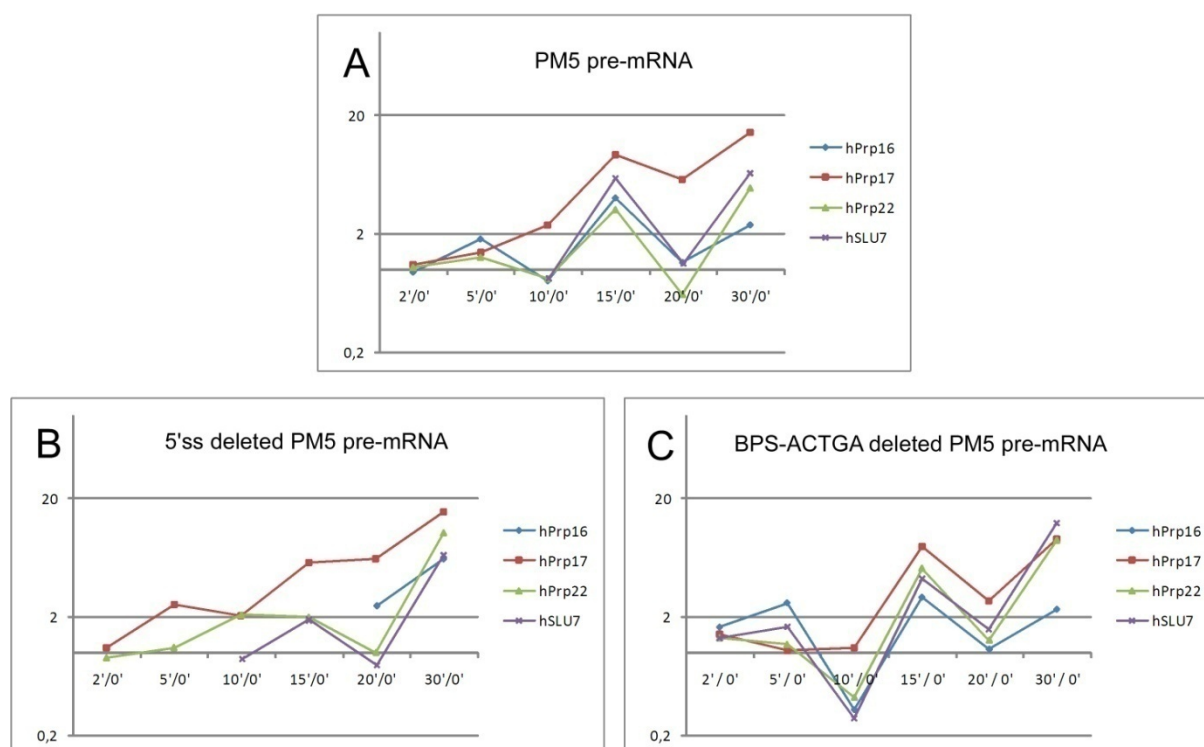


Figure 4.38: Assembly of step 2 factors on PM5 pre-mRNA (A), 5'ss-deleted PM5 pre-mRNA (B), and BPS-ACTGA-deleted PM5 pre-mRNA (C) monitored over 30 minutes.

4.4.4 Direct comparison of the protein assembly on splicing-active PM5 pre-mRNA and splicing-inactive 5'ss deleted PM5 pre-mRNA

4.4.4.1 Experimental setup

In addition to the time-dependent protein assembly on different splicing-active and splicing-inactive pre-mRNAs, the protein compositions at different time points on the splicing-active (PM5 pre-mRNA) and on a splicing-inactive pre-mRNA were compared directly. As the 5'ss-deleted PM5 pre-mRNA showed differences in the assembly of the U1 snRNP specific proteins, hPrp19/CDC5L complex proteins and step 2 factors, this pre-mRNA was chosen for direct comparison of the protein assembly. SILAC nuclear extracts (light and heavy) were prepared from differentially labeled HeLa cells (light and heavy), and spliceosomal assembly was performed in these two extracts in parallel by incubation of an MS2-tagged and radioactively labeled PM5 pre-mRNA and 5'ss-deleted PM5 pre-mRNA with light and heavy SILAC nuclear extract under splicing conditions. Splicing reactions were assembled for different time intervals and affinity purified complexes from the same time points assembled on PM5 or 5'ss-deleted PM5 pre-mRNA, respectively, were mixed in equal molar amounts (Figure 4.39).

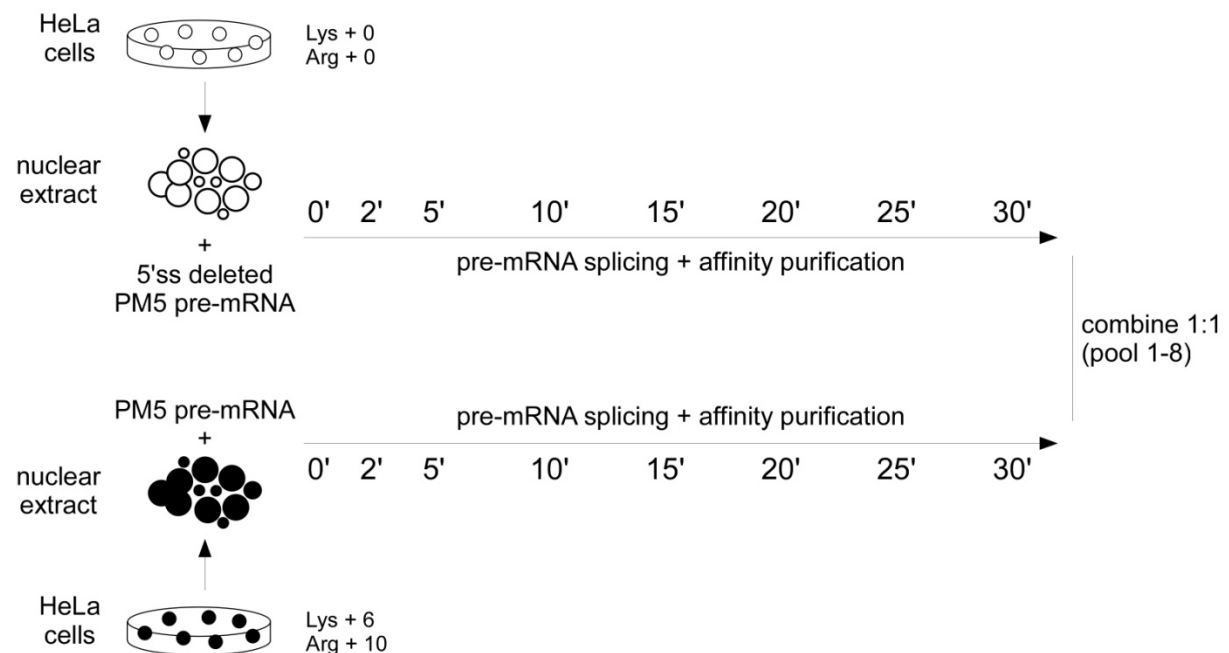


Figure 4.39: Experimental setup to compare the protein compositions of the sets of proteins assembled on PM5 and 5'ss-deleted PM5 pre-mRNA at different time points during pre-mRNA splicing. Splicing reactions were assembled on PM5 and 5'ss deleted PM5 pre-mRNA using light and heavy nuclear extracts, respectively. Assembled complexes from the same time point but assembled on the different pre-mRNAs were pooled in equal amounts.

The proteins within the assembled complexes purified at the same time point but assembled on different pre-mRNAs were separated by PAGE (Figure 4.40). The MS2-MBP protein, which was used for affinity purification showed nearly equal intensity in the combined samples (pools 1–8) for the protein assembly on the different pre-mRNAs confirming the mixing procedure based on the radioactively labeled pre-mRNA. Minor differences in the protein amounts and in the abundances of individual protein bands were observed in the various samples, i.e. at different time points during spliceosomal assembly (compare pools 1–8, Figure 4.40).

Entire gel lanes were excised and the proteins were hydrolyzed with trypsin. The peptides generated were subsequently analyzed by LC-MS/MS. The peptides and finally the proteins were quantified by using MaxQuant software (Cox and Mann, 2008). Protein ratios for the assembly on the splicing-active versus the splicing-inactive pre-mRNA were calculated for every time point and assembly time lines for individual proteins were plotted.

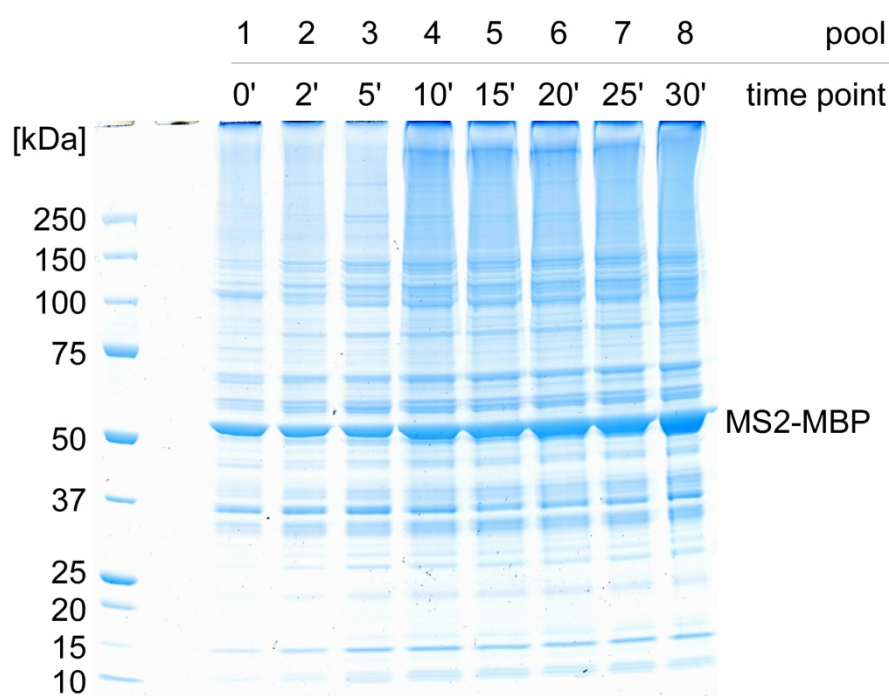


Figure 4.40: Coomassie stained gel of combined samples (pools 1–8) from comparison of the protein assembly on PM5 and 5'ss-deleted PM5 pre-mRNA at different time points during pre-mRNA splicing. Affinity purified complexes assembled for the same time intervals on PM5 and 5'ss-deleted PM5 pre-mRNA were pooled in equal amounts and the proteins were separated by PAGE.

4.4.4.2 Normalization of the data

To compensate for errors that occurred in the mixing of the samples or in any other step during sample handling the protein ratios obtained were normalized. For normalization of

these ratios, two proteins were chosen that are expected to bind to both pre-mRNAs in equal amounts. These are the cap binding proteins CBP20 and CBP80, which bind to the pre-mRNA cap and are not involved in pre-mRNA splicing.

The protein ratios of CBP20 and CBP80 obtained for every time point were plotted to show their assembly on the PM5 pre-mRNA in comparison with 5'ss-deleted PM5 pre-mRNA over the time investigated (Figure 4.41 A). As expected, the two proteins showed the same abundance on both pre-mRNAs, as verified by protein ratios close to 1. However, some fluctuations were observed and normalization of the data is necessary to compensate for these fluctuations. For this purpose, the average protein ratio of the two proteins was calculated for every time point. These values were then used to calculate the different normalization factors (Figure 4.41 C). Figure 4.41 B shows the normalized protein ratios for CBP20 and CBP80. The normalization factors obtained for the different time points were applied to the protein ratios of individual proteins quantified in the data set.

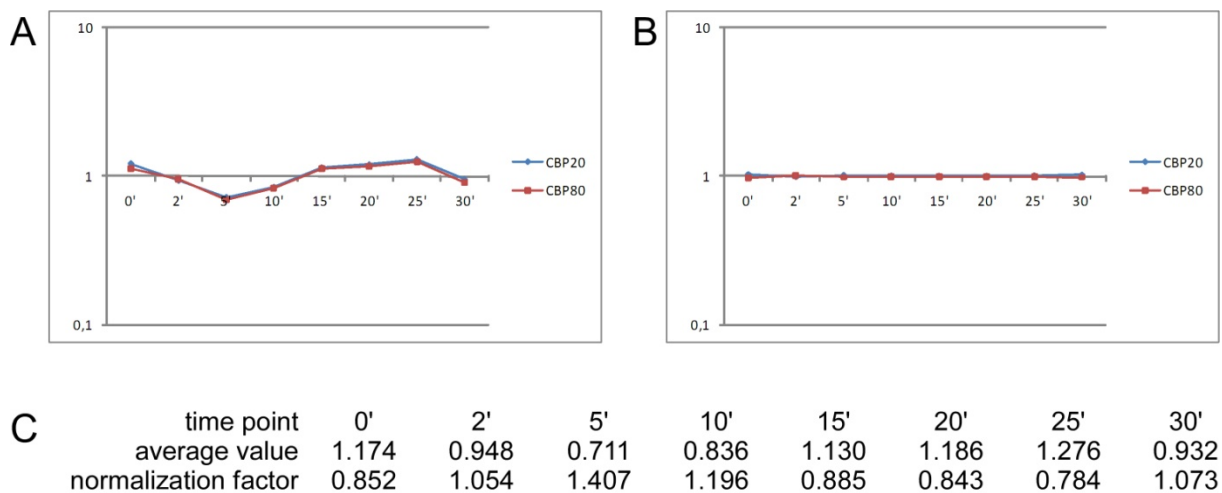


Figure 4.41: Normalization of the protein ratios found for the cap binding proteins. (A) Protein ratios obtained for CBP20 and CBP80 were plotted for every time point. (B) The normalized protein ratios for CBP20 and CBP80 were plotted for every time point. (C) For every time point the average protein ratio of the two cap binding proteins and the calculated normalization factor is given.

4.4.4.3 Protein assembly on PM5 pre-mRNA versus 5'ss-deleted PM5 pre-mRNA

Protein assembly on PM5 and 5'ss-deleted PM5 pre-mRNAs was compared. Several splicing reactions were assembled for different time intervals in light and heavy SILAC nuclear extracts (Figure 4.39). Assembled complexes from the same time points were pooled and analyzed as described above. Protein ratios obtained for the different time intervals were plotted to show the protein assembly on PM5 pre-mRNA in comparison with the 5'ss-deleted PM5 pre-mRNA.

The U1 snRNP is known to bind to the 5' splice site during spliceosomal assembly (Heinrichs *et al.*, 1990; Pomeranz Krummel *et al.*, 2009). Differences in protein abundances on the two pre-mRNAs investigated were therefore expected for U1 snRNP specific proteins. The protein ratios obtained showed clearly that between zero and five minutes the U1-A, U1-C and U1-70K protein were highly enriched on the PM5 pre-mRNA as compared with the 5'ss-deleted PM5 pre-mRNA (Figure 4.42 A). At zero minutes, the proteins are 5-6 times enriched on the PM5 pre-mRNA, whereas protein ratios decreased during incubation. After 10 minutes the protein ratios of the U1 snRNP specific proteins were constant but showed a continuing enrichment on the PM5 pre-mRNA as compared with the 5'ss-deleted PM5 pre-mRNA (Figure 4.42 A). Interestingly, the assembly time line found for the Sm proteins resembles the one found for the U1 snRNP proteins (Figure 4.42 B). At zero minutes the Sm proteins are enriched threefold on the PM5 pre-mRNA compared with on the 5'ss-deleted PM5 pre-mRNA. This can be explained by the binding of U2 snRNP on the 5'ss-deleted PM5 pre-mRNA. Two sets of Sm proteins (belonging to U1 and U2 snRNPs) bind to the PM5 pre-mRNA, whereas only one set of Sm proteins (belonging to U2 snRNP) assembles on the 5'ss-deleted PM5 pre-mRNA. The protein ratio of the Sm proteins at zero minutes is thus halved as compared with the U1 snRNP specific proteins. Nonetheless, the protein ratios decreased between zero and 10 minutes and were constant after 10 minutes (Figure 4.42 B). The effect of the 5' splice site deletion upon the binding of U1 snRNP is thus clearly demonstrated, although binding of U1 proteins on the 5'ss-deleted pre-mRNA was also observed.

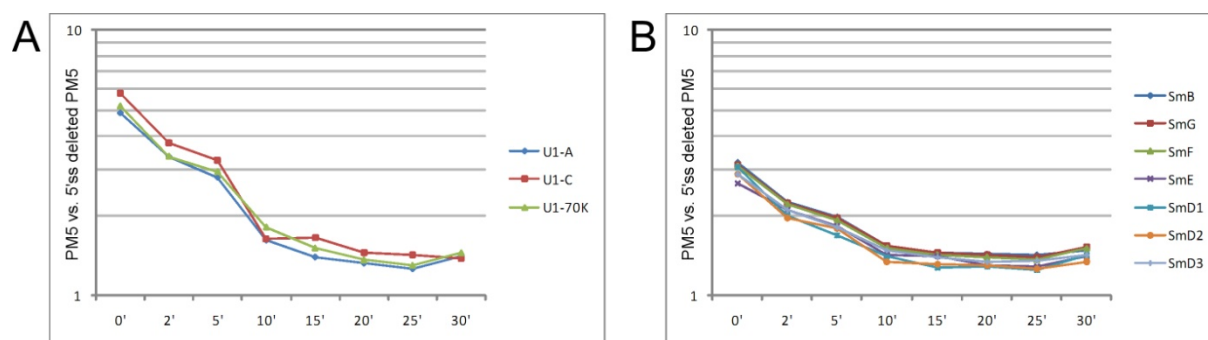


Figure 4.42: Direct comparison of protein assembly on the PM5 pre-mRNA compared with corresponding assembly on the 5'ss-deleted PM5 pre-mRNA for U1 snRNP specific proteins (A) and Sm proteins (B).

As a next step, the effect of the 5'ss deletion on other proteins essential for pre-mRNA splicing was studied. The hPrp19/CDC5L complex and the step 2 factors are enriched in the C complex compared with the B complex (see section 4.3). These proteins are essential for pre-mRNA splicing and are therefore of great interest. Constructed time lines for the direct

comparison of the protein assembly showed that these proteins were highly enriched on the splicing-active PM5 pre-mRNA as compared with the splicing-inactive variant of the pre-mRNA (Figure 4.43). The hPrp19/CDC5L complex protein ratios showed a substantial increase at 5 minutes and were then nearly constant over the whole period (Figure 4.43 A). They are thus highly enriched on the PM5 pre-mRNA from the time point when they predominantly associate with the pre-mRNA. In contrast, assembly time lines of the step 2 factors reach a maximum at 15 minutes but decrease again after 15 minutes. These proteins thus show a time-dependent assembly on the splicing-active compared with the splicing-inactive pre-mRNA. It is noteworthy that, hSLU7 and hPrp22 are first observed after 10 and 15 minutes, respectively. hPrp18 was only detected at two time points, but it too shows high enrichment on the PM5 pre-mRNA compared with the 5'ss-deleted pre-mRNA. hPrp16, which also belongs to the step 2 factors, was not detected in these analyses.

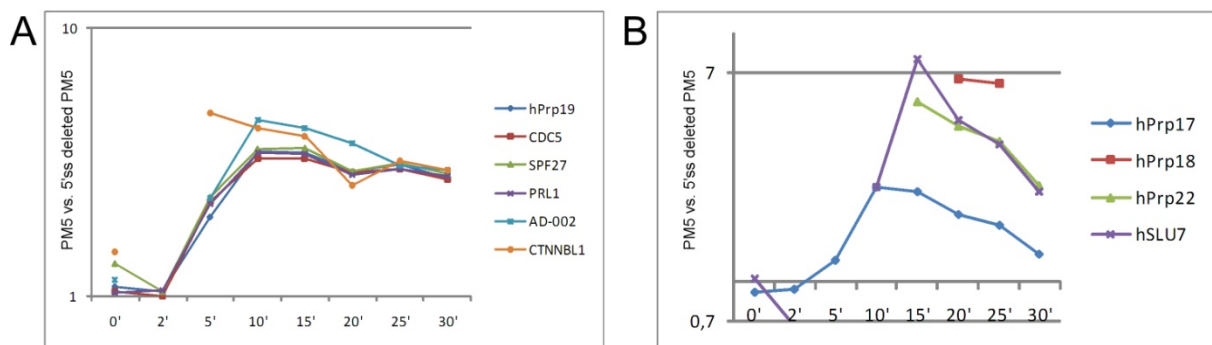


Figure 4.43: Protein assembly of the hPrp19/CDC5L complex (A) and the step 2 factors (B) on the PM5 pre-mRNA compared with corresponding assembly on the 5'ss-deleted PM5 pre-mRNA.

The direct comparison of the protein assembly on the splicing-active (PM5) and the splicing-inactive (5'ss deleted PM5) pre-mRNA yields information about the abundance of proteins on these two pre-mRNAs. Although binding of the groups of proteins investigated on the splicing-inactive pre-mRNA was observed (see section 4.4.3), a clear difference between this and the splicing-active pre-mRNA was found. Proteins that had been expected to be affected by deletion of the 5'ss are highly enriched on the splicing-active pre-mRNA, and an effect of the 5'ss deletion is thus clearly observed.

5 Discussion

5.1 Determination of the protein stoichiometry within the hPrp19/CDC5L complex by absolute quantification (AQUA)

This study has shown that for determination of the protein stoichiometry within a protein complex using synthetic standard peptides and mass spectrometry it is crucial to achieve complete enzymatic hydrolysis of all the proteins within the complex. The choice of conditions for complete hydrolysis strongly affected the results for the stoichiometry (Figure 4.10). The consistency of the results obtained in all the analyses after the hydrolysis of the complex in the presence of acetonitrile led to the conclusion that the Prp19/CDC5L protein complex has a stoichiometry of 4 × Prp19, 2 × CDC5L, 1 × SPF27, 1 × PRL1, 1 × CTNBL1 (Figure 4.10 B). Two other proteins found in the same fraction of the complex were not quantified: AD-002, because of the lack of suitable standard peptides for this protein, and Hsp70, because of initial results obtained by MS with suitable standard peptides suggesting that this protein is not represented in most copies of the complex (Table A.4 in the Appendix).

For similar studies with other protein complexes, there is thus an absolute requirement for initial experiments to determine (i) the solubility and elution profile of the standard peptides and (ii) the optimum hydrolysis conditions of the complex. Recent studies in absolute quantification using standard peptides address the question of the determination of the absolute amount of the synthesized standard peptides before analysis. Standard peptides, whose total amounts have not been correctly determined lead to false results. Amino acid analysis (AAA) is therefore widely recommended for the determination of the total amount of standard peptides in such experiments. The experiments in this study have shown that the solubility of the peptides is the critical issue, rather than putative false amounts of synthesized peptides. The results revealed that once the various synthesized standard peptides are brought into solution the results are highly consistent for the different proteins. Therefore, the quantity itself does not represent the bottleneck of the analysis.

The analyses demonstrate clearly that the digestion conditions are indeed the most critical issue. By a change in the denaturing conditions, the apparent number of copies of the CDC5L protein in the hPrp19/CDC5L complex could be made to vary from one to two. This result was unexpected and cannot easily be explained. As less endogenous peptides of CDC5L were identified after hydrolysis in the presence of urea than after hydrolysis in the

presence of acetonitrile, incomplete digestion of CDC5L in the presence of urea is presumed. Indeed, protection of the protein against proteolysis was observed in particular within those regions that harbor the endogenous counterparts of the selected standard peptides (Grote *et al.*, 2010; Figure 5.1), suggesting that these regions are highly structured and therefore cannot be completely denatured by urea.

```

1                                     51
MPRIMIKGGV WRNTEDEILK AAVMKYGKNQ WSRIASLLHR KSAKQCKARW YEWLDPSIKK
                                     101
TEWSREEEEEK LLHLAKLMPT QWRTIAPIIG RTAAQCLEHY EFLLDKAAQR DNEEETDDDP
                                     151
RKLKPGEIDP NPETKPARPD PIDMDEDELE MLSEARARLA NTQGKKAKRK AREKQLEEAR
                                     201
RLAALQKRRE LRAAGIEIQK KRKRKRGVDY NAEIPFEKKP ALGFYDTSEE NYQALDADFR
                                     251
KLRQQDL DGE LRSEKEGRDR KKDKQHLKRK KESDLPSAIL QTSGVSEFTK KRSKLVLPAP
301                                     351
QISDAELQEV VKVGQASEIA RQTAEESGIT NSASSTLLSE YNVTNNSVAL RTPRTPASQD
                                     401
RILQEAQNL M ALTNVDTPLK GGLNTP LHES DFGSVTPQRQ VVQTPNTVLS TPFRTPSNGA
                                     451
EGLTPRS GTT PKPVINSTPG RTPLRDK LNI NPEDGMADYS DPSYVKQMER ESREHLRLGL
                                     501
LGLPAPKNDF EIVLPENAEK ELEEREIDDT YIEDAADVDA RKQAIRDAER VKEMKRMHKA
                                     551
VQKDLPRPSE VNETILRPLN VEPPLTDLQK SEELIKKEMI TMLHYDLLHH PYEPSGNKKG
601                                     651
KTVGFGTNNS EHITYLEHNP YEKFSKEELK KAQDVLVQEM EVVKQGM SHG ELSSEAYNQV
                                     701
WEECYSQVLY LPGQSR YTRA NLASKKDRIE SLEKRLEINR GHMTTEAKRA AKMEKKMKIL
                                     751
LGGYQSRAMG LMKQLNDLWD QIEQAHLELR TFEELKKHED SAIPRRLECL KEDVQRQQR
                                     801
EKELQHRYAD LLEKETLKS KF

```

Figure 5.1: Endogenous counterparts of the selected standard peptides for CDC5L protein are located in highly structured regions of CDC5L. The amino acid sequence of CDC5L is shown. Regions that are protected against proteolysis (Grote *et al.*, 2010) are labeled in light red. Endogenous counterparts of selected standard peptides are highlighted in bold red. All peptide sequences selected as standard peptides are located in a highly structured region of CDC5L.

In fact, only a few limited methods are available to determine the complete hydrolysis of a complex. Denaturing PAGE, even at the highest possible resolution, as achieved by Schagger gels (Schagger and von Jagow, 1987), reveals incompletely digested fragments only up to a size of 1 kDa. Moreover, PAGE conditions are not necessarily compatible with the digestion of proteins in the presence of urea, so that the risk of sample loss during desalting cannot be excluded. Accordingly, larger residual protein fragments after hydrolysis of the Prp19/CDC5L complex, either in the presence of acetonitrile or in the presence of urea, cannot be monitored reliably. MS *per se* can monitor larger fragments, but the sensitivity is dependent on the fragments' size, and multiply charged ions in ESI can be

suppressed by double and triply charged ones. An alternative might be the use of monolithic columns that allow one to separate and elute intact proteins. However, not all proteins elute even under organic and denaturing conditions from these columns.

The amount of peptides containing missed cleavage sites can influence the absolute quantification. For this reason, the use of more than one standard peptide for each protein is highly recommended to provide a confirmatory control and to detect any deviations that might arise. On the basis of miscleavage, two peptides were excluded from the quantification (TGYNFQR derived from PRL1 for hydrolysis in acetonitrile and EAAAALVEEETR derived from SPF27 for hydrolysis in urea; Figure 4.8). The amount of a miscleaved peptide can ultimately only be determined if one also has standard peptides for this peptide and adds these to the sample. However, the generation of isotope labeled standards is expensive and the generation of standard peptides and additional standard peptides containing possible missed cleavage sites might not be justified.

Importantly, the different accessibilities of certain protein regions within proteins are a further critical issue in the determination of protein stoichiometry when the investigated protein complex is, under different cellular conditions, an integral part of another complex comprising even more proteins. Integration into another complex might lead to structural rearrangement of the proteins and therefore to changes in the accessibility of the proteins' regions against which standard peptides have been generated for absolute quantification. Previous proteome analyses have demonstrated that the human Prp19/CDC5L complex associates with the pre-mRNA splicing machinery throughout the various steps of splicing. This complex is present within the so-called pre-catalytic spliceosomal B complex, the catalytically activated B complex (B*), and the step 1 C complex. Furthermore, it was shown that it is part of a post-spliceosomal 35S U5 complex. Although the hydrolysis conditions for the isolated Prp19/CDC5L complex have been established and, on this basis, proteotypic standard peptides have been selected, the possibility that the major changes in protein-protein interactions that take place during the transition of the spliceosomal B complex to the C complex also affect the proteins of the Prp19/CDC5L complex, cannot be ruled out. For instance, the proteotypic peptides selected for the protein PRL1 are located in a region that is readily accessible toward proteases and has therefore been suggested to be highly flexible and unstructured in the Prp19/CDC5L complex (Grote *et al.*, 2010). These regions might become structured upon additional protein-protein interactions in larger spliceosomal complexes. Thus, when extending the studies on the above mentioned spliceosomal complexes, complete hydrolysis conditions must be established for all of the complexes (B, B*, and C complex), which consist no longer of seven proteins but generally of about 125 proteins (see Wahl *et al.*, 2009 for review). Therefore, the key question to be asked is

whether, although the strategy applied here produces highly reproducible results under the improved experimental conditions (hydrolysis in acetonitrile and analysis by MRM) for a complex of moderate complexity, the cost in time and financial resources is justified when such studies are extended to complexes of much higher complexity.

Another critical issue for the determination of the protein stoichiometry within a complex is its homogeneity. Here, the affinity-purified hPrp19/CDC5L complex was subjected to glycerol-density centrifugation in order to obtain the highest possible purity and homogeneity. However, Hsp70 has been identified in the gradient fraction of the hPrp19/CDC5L complex, but was found to be of very low abundance. This raises the question whether gradient centrifugation under these conditions is not sufficient to shift a minor portion of the hPrp19/CDC5L complex that contains Hsp70 toward a different sedimentation coefficient or, alternatively, whether Hsp70 forms a complex by itself that was co-purified and co-migrates with Prp19/CDC5L complex. Alternatively, on the assumption that the very low amount of Hsp70 represents the lowest possible numbers of copies of a single protein with the hPrp19/CDC5L complex, then the copy numbers of all other proteins must be adapted accordingly, thus suggesting that several copies of the hPrp19/CDC5L complex with the stoichiometry reported here are assembled on a single copy of Hsp70. This question cannot be answered unequivocally by the applied method. Rather, the determination of the entire mass of the complex would be required. In addition, different Hsp70 isoforms were detected during proteomic analysis of the hPrp19/CDC5L complex. The composition of these isoforms in the complex has not been investigated so far and might also influence the quantification of this particular protein.

Very recently, the use of standard peptides for absolute quantification of affinity-tagged proteins within protein interaction networks was reported (Wepf *et al.*, 2009). Isotopically labeled standard peptides encompassing a peptide derived from the introduced tag were designed and used for the quantification of several proteins in various affinity-purified protein complexes. In this manner, labeled standard peptides were used to calibrate for any other protein within the complexes. However, since the authors compared the stoichiometry of different affinity-purified complexes without any further purification, no conclusion was drawn about the protein stoichiometry within a single protein complex.

In a similar way, differing only in technical details, absolute quantification is frequently used in clinical proteomics. In such studies, standard peptides have been used to determine by MS the absolute amount of defined proteins (mass or mole number) in samples (Abbatiello *et al.*, 2008; Langenfeld *et al.*, 2009). However, even though within these studies samples from different sources were compared in terms of their absolute protein amount, the quantification is actually a relative one. Thus, when comparing the absolute amount of protein and/or

protein complexes in a sample, hydrolysis conditions are not crucial as long as they are consistent among the different samples and as long as the proteins in question are not present in different complexes in the samples of different origin.

In summary, the complete hydrolysis of the proteins under investigation and the complete solubility of the standard peptides have been proven to be the major prerequisites for successful absolute quantification. The protein stoichiometry within the hPrp19/CDC5L complex could be determined although discrepancies between different hydrolysis conditions arised. Hydrolysis in the presence of acetonitrile in combination with MRM analysis revealed the most consistent results, thus representing optimal experimental conditions. However, when transferring this method to other (possibly larger) protein complexes, testing of different hydrolysis conditions is mandatory. In addition, selection of suitable standard peptides emerged as not trivial. The protein sequences as well as structural information about the proteins under investigation (if available) should be taken into account to successfully select suitable standard peptides for absolute quantification.

5.2 Relative quantification by iTRAQ-labeling of in-gel digested proteins

The traditional iTRAQ workflow comprises in-solution digestion of proteins, labeling of the generated peptides, removal of excess reagent by strong cation exchange chromatography (SCX) and finally the analysis by LC-MS/MS (Ross *et al.*, 2004). To reduce the sample complexity before protein hydrolysis and the possibility of sample loss during SCX, a modified protocol for iTRAQ labeling of in-gel digested proteins was established. This has the advantage that even of small protein amounts can be quantified. The modified workflow was tested by relative quantification of selected proteins of spliceosomal tri-snRNP particles.

5.0 µg and 2.5 µg of tri-snRNP was separated by PAGE and quantified using the modified iTRAQ-labeling workflow. The established method showed consistent results for the selected tri-snRNP specific proteins. However, the expected protein ratio of 2.0 was not achieved. The obtained ratio for different tri-snRNP amounts is 2.5 instead. As all quantified proteins show the same protein ratio of 2.5, this deviation might be due to sample loading or sample preparation for gel electrophoresis and is a matter of normalization of the sample. The blank sample that was processed along with the other samples showed a ratio of 0.1. The observation that background noise is often obtained during chemical labeling was recently discussed (Bantscheff *et al.*, 2007). This background noise does not depend on the mass resolution of the mass spectrometer but rather on the size of the m/z window chosen for isolation of peptides for fragmentation (typically between 2-6 m/z). All peptides present in this

window will contribute to the obtained signal. In addition, the stable isotopes incorporated to the used iTRAQ reagents do not show 100 % purity. This is typically 98-99 % and contributes to the obtained signal intensities of the lower reporter ions.

During pilot experiments the labeling efficiency was found to be one of the major issues for reliable relative quantification. It can be calculated from the total number of peptides and the number of labeled peptides identified in the same analysis. Using chemical labeling approaches complete labeling of the peptides or proteins is rarely observed (Bantscheff *et al.*, 2007). However, for successful relative quantification, the labeling efficiency should be as high as possible (i.e. nearly 100 %). In this study, only for experiments with a labeling efficiency higher than 90 % accurate quantification was achieved (i.e. expected protein ratios were obtained). For the proteins identified and quantified within this proof of principle experiment the labeling efficiency was 92 % and the resulting protein ratios were very consistent among the different proteins.

5.3 Relative quantification of spliceosomal B and C complexes – a comparative study

5.3.1. Methodical considerations

This work is the first proteome study that compares directly iTRAQ and SILAC for relative quantification of proteins. We performed relative quantification of proteins derived from spliceosomal B and C complexes and found that iTRAQ and SILAC yielded similar result. However, when comparing the obtained protein ratios, iTRAQ showed in general slightly lower values for proteins that are enriched in the B complex and slightly higher values for proteins that are enriched in the C complex. This might be due to the fact that the precursor selection for MS/MS is not 100 % selective and co-eluting peptides thus might contribute to the iTRAQ reporter ion intensity as it has been recently discussed (Bantscheff *et al.*, 2007). Therefore, SILAC protein ratios show extremely high values or low values for proteins that are enriched or underrepresented, respectively, in one of the samples when compared to iTRAQ.

Protein ratios obtained from the SILAC experiments showed a lower standard deviation *per se* as compared to iTRAQ. Indeed, metabolic labeling should be a more reliable method for labeling proteins as compared to chemical labeling, as it should guarantee a 100 % labeling efficiency of proteins and thus of the peptides. Moreover, samples can be pooled at an earlier stage and therefore variation in the quantification ratios by e.g. sample losses, can be

neglected. Nonetheless, depending on the cellular system, sometimes complete labeling cannot or can only hardly be achieved by SILAC. In particular for those cells grown in culture that do not rapidly divide (e.g. embryonic stem cells; Graumann *et al.*, 2008) or only proliferate in cell culture without dividing (e.g. primary neurons; Liao *et al.*, 2008). Normally in such cells, a labeling efficiency of maximal 80 % is achieved. However, such incomplete labeling can be handled by novel computational methods (Liao *et al.*, 2008).

Of course, chemical labeling always involves the risk of incomplete labeling. In our initial experiments a low labeling efficiency (< 80 %) was found to drastically influence the quantification results. Therefore, the labeling efficiency is one of the major issues for obtaining accurate quantification when using chemical labeling approaches and the initial experiments were performed to achieve the highest possible labeling efficiency. Nonetheless, the great advantage of all chemical labeling approaches is that proteins from almost every source (cells, tissue, body fluids etc.) can be quantified.

An alternative chemical labeling approach to iTRAQ labeling has recently been introduced by Boersema *et al.*, 2008. Dimethylation of the peptides' N-termini and lysine side chains using different stable isotope labeled reagents allows relative quantification of three samples in one MS analysis. Dimethyl labeling is based on a simple chemical reaction without any observed byproducts and provides a 100 % labeling efficiency in almost all cases. In addition, it uses inexpensive reagents and is thus a cost-effective labeling technique in comparison to other stable isotope reagents. However, we have not performed this particular labeling strategy.

Among the different chemical labeling strategies, isobaric reagents, such as iTRAQ reagents or TMTs (Thompson *et al.*, 2003), have the advantage that quantification is performed during MS/MS analysis so that sample complexity is not enhanced, as the differently labeled peptides show the same mass in the MS. As the reporter ions, which are used for quantification of the different samples, are released during MS/MS, the analytical depth of the analysis is higher as compared to analysis of differently labeled samples that show peak pairs in the MS (e.g. SILAC or dimethyl labeling).

Until now, isolated spliceosomal complexes have been compared by peptide count after LC-MS/MS analysis (Behzadnia *et al.*, 2007; Bessonov *et al.*, 2008; Deckert *et al.*, 2006; Kuhn *et al.*, 2009). The correlation between the relative protein abundances and the number of acquired tandem MS spectra (spectral count), the number of identified peptides (peptide count) and the obtained sequence coverage has recently been compared by Liu *et al.*, 2004 in a study with defined standard proteins. A linear correlation was found between the relative protein amount and the number of acquired MS/MS spectra (spectral count), but not between the relative protein amount and the number of identified peptides (peptide count) or the

sequence coverage. In addition to iTRAQ and SILAC, we therefore evaluated spectral count for relative quantification of spliceosomal B and C complexes. Spectral count is a simple quantification technique that enables the comparison of almost every sample from a proteome study without additional sample preparation. As it does not require labeling of the peptides or proteins, no expensive labeling reagents are needed and as many samples as desired can be quantified relative to each other. However, it requires highly reproducible LC-MSMS analyses (see also below).

In our study, we found a good overall agreement between spectral count and quantification with iTRAQ and SILAC. In all three quantitative analyses, several proteins were clearly identified to be more abundant in spliceosomal B or C complexes and some proteins were found to be present in equal amounts within both complexes (see Tables 4.9 – 4.11). As discussed above, iTRAQ and SILAC yielded consistent results for relative quantification of B and C complexes. Strikingly, for most of the quantified proteins, the same quantitative trend, i.e. enrichment in one of the complexes or same abundance in both complexes, was also obtained by spectral count. Importantly, spectral count exhibits some discrepancies compared to iTRAQ or SILAC. In particular for small proteins (< 20 kDa), accurate quantification could not at all or only roughly be achieved by spectral count, as only a limited number of peptides were generated. For example, the relative quantification of the Sm proteins within both, the spliceosomal B and C complexes, using spectral count did not unambiguously reveal the expected ratios, i.e. a two-fold difference of Sm proteins in the B vs. C complex.

As the Sm proteins are common to all U snRNPs except for U6 snRNP, four copies of Sm proteins are expected in the B complex. Upon transition from the B to the C complex, only two copies of Sm proteins are left, because U1 and U4 snRNPs are destabilized/dissociated. For all seven Sm proteins, protein ratios close to 2 are obtained by iTRAQ and SILAC, whereas spectral count yielded the correct value only for two of the Sm proteins (SmF and SmG). The other Sm proteins show protein ratios of approximately 1 by spectral count.

Another example for the lower accuracy of spectral count is the quantification of proteins that should be present in a 1:1 ratio or where iTRAQ and SILAC clearly showed such a ratio. These are, for example, U5-220K, U5-40K, and CBP20 (for a complete list, see Tables 4.9 – 4.11). With spectral count, higher or lower protein ratios (approximately between 0.6 and 1.7) were obtained. These results are consistent with the observations of Liu *et al.*, 2004 that small quantitative changes among proteins in different samples cannot be accurately monitored by spectral counting. The quantification of large proteins within the B and C complex by spectral count yielded results very similar to iTRAQ and SILAC (e.g. with the

SF3a and SF3b proteins), except when proteins are present in nearly equimolar amounts in both samples (e.g. U5-220K and U5-200K).

Another discrepancy between spectral count and labeling approaches is observed with the protein ratios obtained for proteins that are pre-dominantly associated with one of the two complexes. These, like the hPrp18, hPrp22, and DDX35 proteins, show extreme values after spectral count, thus suggesting their complete absence in one of the complexes. Such extreme protein ratios might be misleading regarding the presence or absence of proteins within the different samples. If a peptide is not selected for sequencing it does not necessarily mean that the peptide is not present in the sample. Such low abundance peptides, which might escape detection by spectral count, are still detectable by iTRAQ and SILAC quantification. During iTRAQ analysis, samples are pooled and the differently labeled peptides are isobaric and are thus selected for sequence analysis irrespective of whether the actual non-labeled peptides are of low or high abundance. Quantification is then performed on the MS/MS level, where even low abundance peptides produce the corresponding reporter ions. SILAC quantification is based on the correct assignment of the mass pairs that were generated upon the incorporation of stable isotopes. In this manner, the low abundance peptides – if present – will be recognized by the software (e.g. MSQuant, Schulze and Mann, 2004; or MaxQuant, Cox and Mann, 2008), through the assignment of the corresponding highly abundance peptide.

Spectral count has further limitations that are due to the following technical requirements: (i) A high reproducibility of the chromatography system is required to obtain comparable elution profiles of the peptides during sample separation. (ii) The spectral count response for every protein is not the same, i.e. due to the protein's amino acid sequence and the different properties of the generated peptides (e.g. chromatographic behavior) the number of detectable spectra varies for every protein. As discussed above, smaller proteins generate only few peptides and relative quantification by spectral count is limited for these proteins. (iii) Different co-eluting peptides in the respective samples can affect the acquisition of distinct MS/MS spectra and thus influence the quantification process. (iv) Dynamic exclusion of precursor masses - that is selection of a precursor that already has been selected for fragmentation before and is thus subsequently not selected again for fragmentation (within a certain time window) - is usually used during LC-MS/MS analyses. Although the analytical depth is enhanced by using dynamic exclusion, dynamic exclusion negatively influences accurate quantification by spectral count, because different peptides are always selected for MS/MS fragmentation. It is important to note that spectral count not only takes the number of unique peptides into account, but also the overall number of spectra, i.e. the same peptide is selected several times for fragmentation. Consequently, when working with dynamic

exclusion, the quantification of highly abundant peptides that show a longer retention time during their elution from the LC is not considered adequately. When working without dynamic exclusion, the spectral count is more reliable but, on the other hand, the analytic depth is drastically reduced and therefore only a limited number of proteins can be quantified.

The same holds true for SILAC and other labeling procedures that are based on peptide intensities on the MS level. Incorporation of stable isotopes (by metabolic or chemical labeling) generates peptides of different masses and the sample complexity is consequently increased. The analytical depth for the analysis of complex samples is thus reduced and only a limited number of proteins can be quantified.

Importantly, heavily modified proteins, protein isoforms and truncated proteins escape quantification. As the modified and the unmodified peptides show different masses, these proteins might yield false quantification values, in particular, when a protein becomes significantly modified during transition from the B to the C complex. Protein ratios of different protein isoforms within the quantified complexes might also not be correctly assigned and might thus affect the quantification for these proteins.

There are several reasons why (semi-quantitative) spectral count in our study, i.e. the comparison of highly purified spliceosomal complexes, is still applicable for relative comparison: (i) The analyzed spliceosomal complexes were highly purified under stringent conditions, thus minimizing the number of contaminating proteins during the analyses; (ii) the complexes are of moderate complexity and consist of only a limited number of proteins; (iii) The use of always the same LC system coupled front-end to the ESI Q-ToF mass spectrometer (Q-ToF Ultima, Waters) in all of the performed previous studies (Behzadnia *et al.*, 2007; Bessonov *et al.*, 2008; Deckert *et al.*, 2006; Kuhn *et al.*, 2009) fulfilled the prerequisite to generate comparable results and ensures high reproducibility such that analyses can be compared. However, when adapting the above described relative quantification approaches to biological systems other than spliceosomal complexes, the critical aspects discussed above have to be taken into account.

5.3.2. Functional considerations

Our results are in general very consistent with the findings from Bessonov *et al.*, 2008. In this study, they compared the proteomes of B and C complexes using the number of peptides identified during LC-MS/MS analysis. In addition, the abundances of individual proteins in the respective complexes were monitored by immunoblotting. Peptide count and immunoblotting revealed that both complexes (i.e. B and C complexes) contained U2 and U5 snRNP proteins, the hPrp19/CDC5L complex and related proteins and the RES complex. U1 and U4

snRNP proteins, U6 LSM proteins and several non-snRNP proteins were found to be underrepresented in the C complexes. Other proteins like the second-step factors, DEAD-box helicases Abstrakt and DDX35 and several peptidyl-prolyl isomerases (e.g. PPWD1, PPIG, and PPIL3b) were found solely or predominantly in the C complex. Strikingly, immunoblotting showed that various SF3a/b proteins were less abundant in C complex.

Our relative quantification approaches applied to isolated B and C complexes clearly show that U1 and U4/U6 snRNP specific proteins, as well as LSM proteins are underrepresented in the C complex. However, quantification by stable isotope-labeling turned out to be more reliable as exemplified by the stoichiometry of the Sm proteins, the cap binding proteins and, as previously suggested, a stable complex of U5 snRNP specific proteins (Achsel *et al.*, 1998). The Sm proteins show a two-fold enrichment in B vs. C complex (see above), and the cap binding proteins CBP20 and CBP80 and various U5 snRNP proteins (namely U5-220K, U5-200K, U5-116K, and U5-40K) were unambiguously found in equimolar amounts within both complexes. This was also shown to be the case for U5-116K by immunoblotting (Bessonov *et al.*, 2008).

Comparison of the canonical U-snRNP specific proteins in the analysis of Bessonov *et al.*, 2008 revealed that U2 snRNP specific SF3a and SF3b proteins are less abundant in the C complex as compared to the B complex. Our data are clearly consistent with their previous analysis and thus support strongly the hypothesis that the SF3b and SF3a proteins either partially dissociate from the U2 snRNA during the transition of B to C complex, or that structural rearrangements cause a destabilization of these proteins so that they, upon the purification of the complexes, partially dissociate. Importantly, the U2-snRNP specific proteins U2-A' and U2-B'' show in our analysis a clear equimolar ratio between both complexes. This proves that both of these U2 snRNP specific proteins, together with the Sm proteins, remain stably bound to the U2 snRNA. It remains to be elucidated whether U2 snRNP within the C complex indeed consists only of the so-called 12S U2-snRNP (U2 snRNA, Sm proteins and U2-A' and U2-B''; Behrens *et al.*, 1993) or not.

The hPrp19/CDC5L complex (Ajuh *et al.*, 2000; Makarova *et al.*, 2004) was found to be associated already with the B complex, but was shown to be more abundant in the C complex. Our results are consistent with this observation and protein ratios (B vs. C) of approximately 0.5 show a higher abundance in the C complex. A loose association of these proteins with the B complex and a more stable association with the C complex is therefore suggested.

Several other proteins like hSLU7, Abstrakt, and DDX35 show significantly low B/C protein ratios suggesting that they are only present in the C complex. Among these C complex

specific proteins, we identified by our methods a hitherto unknown protein associated with the C complex, namely DDX34. This protein is a probable ATP-dependent RNA helicase (gi|38158022), but has so far not been functionally described, neither in yeast nor in human. In yeast, several proteins (PRP43, PRP22, PRP16, PRP2, DHR2, YLR419W, and ECM16) with sequences homologous to DDX34 have been identified; however, the sequence homology is limited to the DEAD-box motifs. In addition, core components of the exon junction complex, namely eIF4A3, Magoh, and Y14, were also found to be clearly associated with the C complex.

The RES (retention and splicing) complex was found to be present in equimolar amounts within B and C complexes. In a study from yeast, this complex was shown to be essential for pre-mRNA splicing (Dziembowski *et al.*, 2004). Its inactivation caused pre-mRNA leakage from the nucleus, confirming our results that RES complex proteins are present in equal amounts in both, B and C complexes.

5.4 Protein assembly time line for spliceosomes by relative quantification

Metabolic labeling using SILAC is widely accepted as a technique for monitoring quantitative changes between different cellular states (for review see Ong and Mann, 2005). Here, we have used SILAC to monitor quantitatively the assembly over time of spliceosomal proteins on distinct pre-mRNAs.

To date, only a very few studies have investigated dynamic changes in protein composition by taking quantitative mass spectrometric approaches. All these also used metabolic labeling to follow such protein changes over time. The first study of this kind was the analysis of proteome dynamics in the nucleoli of eukaryotic cells (Andersen *et al.*, 2005), where the response of the nucleolar proteome to transcription inhibition was observed by the SILAC approach. Making use of three differentially labeled HeLa cell lines, Andersen *et al.*, 2005 were able to describe the turnover kinetics of several nucleolar factors influenced by inhibitors of transcription.

Later on, the same group modified the quantitative mass spectrometric approach to protein turnover within the nucleolus by introducing a “pulsed” incorporation of stable isotopes (Lam *et al.*, 2007). For this purpose, unlabeled cells were transferred to SILAC media and cells were labeled for different time intervals. Nucleoli were isolated and the protein changes within the nucleoli were compared with those in the unlabeled cells (i.e., cells taken from the same preparation at the moment the reaction was started). By this approach, ribosomal

proteins were found to be synthesized and to accumulate in nucleoli more rapidly than any other nucleolar factors.

This approach has been further pursued by Selbach *et al.*, 2008 to study changes in protein translation under the influence of certain microRNAs. It is termed pulsed SILAC (pSILAC). In their study, non-labeled cells were transfected with miRNAs and transferred to cell-culture media containing amino acids labeled with stable isotopes. Subsequently, newly synthesized proteins were labeled with heavy stable isotopes, and thus changes in protein composition could be monitored by LC-MS/MS. A similar approach with pSILAC, by the same group, was applied to investigate the translational response upon cell stimulation (Schwanhausser *et al.*, 2009). However, all these approaches addressed solely the question of which proteins are newly synthesized and which are degraded over time.

The only study so far to address the question of protein assembly by quantitative mass spectrometry was performed by Williamson and co-workers, who analyzed the assembly of the 30S ribosomal subunit *in vitro* by pulse-chase mass spectrometry (Talkington *et al.*, 2005; Williamson, 2005). The assembly of the proteins on 16S ribosomal RNA to give a functionally active 30S ribosomal subunit was investigated by using ¹⁵N- and ¹⁴N-containing ribosomal proteins. In pulse-chase experiments, the 16S ribosomal RNA was pulsed with ¹⁵N-labeled ribosomal proteins which were then chased with an excess of ¹⁴N-labeled proteins. After a certain period of time, the fully assembled 30S ribosomal subunits were isolated by density-gradient centrifugation and the degrees of incorporation of ¹⁴N- and ¹⁵N-labeled proteins were measured by mass spectrometry. By varying the pulse over time and keeping the assembling time constant they were able to observe the kinetics of assembly of the various ribosomal proteins into 30S ribosomal subunit.

A similar approach, but with nuclear extract containing differently labeled amino acids lysine and arginine (see section 4.4.2) would have been of benefit, to monitor the kinetics of assembly of the spliceosomal proteins into complex A, B and/or C.

However, there are some restrictions that might have complicated the adoption of the above-mentioned strategy for observing the assembly of spliceosomal complexes and thus such (more complex) experiments were omitted during the period of this work. (i) Assembly of the spliceosomal proteins and a pre-mRNA to yield spliceosomes is a process in which the spliceosome passes through different functional states to generate finally a catalytically active spliceosome (see Introduction). Once the pre-mRNA is spliced and the mature mRNA is formed, the intron lariat is released and the spliceosome dissociates; a new spliceosome is formed on each pre-mRNA to start a new round of splicing (for review, see Wahl *et al.*, 2009). In this manner, complexes are continuously assembled on a new pre-mRNA in

contrast to the studies of the 30S ribosomal subunit where the assembly on the 16S rRNA is complete once the subunit has formed. Re-assembly of the spliceosome can be avoided when a pre-mRNA is used that lacks the 3' splice site, like the PM5 pre-mRNA (Anderson and Moore, 1997; Bessonov *et al.*, 2008) and the assembly of the spliceosome halts at the state of the C complex. (ii) Pulse-chase experiments require the isolation of formed complexes, e.g. by density-gradient centrifugation, to allow elucidation of the ratios of differently labeled proteins. If the PM5 pre-mRNA is used, assembly of the C complex requires 180 min plus time for the additional RNaseH cleavage step. The C complexes generated are then subjected to density-gradient centrifugation for preparative purification. For such experiments and the subsequent mass spectrometry-based protein analysis, a relatively large amount of nuclear extract is required, and in the case of pulse-chase experiments the amount would be multiplied by the number of time points chosen for pulsing. The same is true for monitoring the formation of A and/or B complexes.

For these reasons – i.e. the dynamic nature of the assembly pathway of spliceosomes and the fact that for preparative purification of assembled spliceosomal complexes a large quantity of nuclear-extracted (labeled and non-labeled) material is needed – together with constraints of time, this strategy was not followed.

Instead, it was decided to compare quantitatively the assembly of spliceosomal proteins on different pre-mRNAs. In the first round of experiments we compared quantitatively the assembly of the proteins on the different pre-mRNAs at distinct time points in a triple-label SILAC experiment (see Results, section 4.4.3). By using a double-labeling SILAC approach, in the second round of experiments we directly compared quantitatively the assembly of the proteins over time on two pre-mRNAs (30 min; see Results, section 4.4.4). Moreover, experiments of this type did not require the purification of assembled spliceosomal complexes by density-gradient centrifugation, but simply an affinity purification through the MS2-tag on the pre-mRNA. In this manner we obtained first quantitative insights into the assembly of various proteins on pre-mRNA.

Our initial results on this system will be discussed briefly in the following paragraphs:

In general these results show that relative mass spectrometry-based quantification is indeed suitable for displaying protein changes during the spliceosomal assembly pathway over time. Although we observed assembly of spliceosomal proteins on splicing-active (PM5) as well as on splicing-inactive pre-mRNAs (5' ss- and BPS-deleted PM5 pre-mRNAs), our studies clearly reveal differences between the splicing-active and -inactive pre-mRNAs. The observed quantitative differences in the assembly of the proteins on different pre-mRNAs will help in subsequent studies aimed at a complete understanding of protein assembly during

the process of pre-mRNA splicing *in vitro*. We further believe that this approach is more straightforward than fluorescence techniques such as FRET (Ohrt *et al.*, 2006) and – once all kinetic studies including the above-mentioned pulse-chase experiments have been performed – will yield dissociation rate constants for a multitude of spliceosomal proteins.

We observed quantitative differences in the association of U1 snRNP specific proteins with the three different pre-mRNAs (see Results, section 4.4.3). While in all three experiments a significant amount of U1 snRNP specific proteins was found to be associated with the pre-mRNA after 5 min under splicing conditions, the amount of U1 snRNP specific proteins only drastically decreases over time on the BPS-deleted PM5 pre-mRNA (Results 4.4.3), whereas on the PM5 and on the 5'ss-deleted PM5 pre-mRNA the amount of proteins seems to remain constant over time. Strikingly, the deletion of the 5'ss within the pre-mRNA shows the most significant effect on the protein assembly. The U1 snRNP specific protein C was absent after 2 min on the 5'ss-deleted PM5 pre-mRNA; however, even on the PM5 pre-mRNA its amount was clearly lowered during assembly time. Protein U1-C is crucial for E complex formation (Heinrichs *et al.*, 1990) and it has been suggested that this protein stabilizes base-pairing between U1 snRNA and the 5' splice site of the pre-mRNA (Pomeranz Krummel *et al.*, 2009; and references therein). Its absence on the 5'ss-deleted PM5 pre-mRNA – despite the fact that the other U1 snRNP proteins are present on the 5'ss-deleted pre-mRNA – sheds light on its important function in correct 5'ss recognition and perhaps in the subsequent formation of “correct” E and A complexes. However, the observation that the amount of associated U1-C protein also decreases on PM5 pre-mRNA is not easily explained. We suggest that a part of U1 snRNP also associates unspecifically or loosely with pre-mRNA, irrespective of whether or not a canonical 5'ss is present. This would explain the similar assembly pattern over time of the U1 snRNP-specific proteins (except for U1-C) between PM5 and 5'ss-deleted PM5 pre-mRNA. Importantly, the overall assembly kinetics of the BPS-deleted PM5 pre-mRNA clearly reveals less (or no) association of the U1 snRNP specific proteins. This is consistent with the previous observation that interaction of U2 snRNP with the branch point sequence also stabilizes significantly the binding of U1 snRNP to the pre-mRNA.

We observed very similar association kinetics of U2 snRNP specific proteins on the three different pre-mRNAs as compared with U1 snRNP specific proteins. The association of the U2 snRNP specific proteins does not seem to be altered by the 5'ss-deletion, but it is affected significantly – as expected – by the BPS-deletion (see above).

At the current state of the analysis, the results for the assembly of the proteins of hPrp19/CDC5L complex and second-step splicing factors are difficult to interpret in a comparison of the three different pre-mRNA constructs. On all three pre-mRNAs we

observed a significant amount of Prp19/CDC5L complex proteins and step 2 factors associated with the RNA after 30 min of incubation under splicing conditions. Overall, the association kinetics of all these proteins show a saw-tooth-like pattern with a maximum at 5 min, a minimum at 10 min, a maximum at 15 and so forth (see Results). An exception is clearly the step 2 factor hPrp17 (Neer *et al.*, 1994), the amount of which associated with PM5 and 5'ss-deleted PM5 steadily increases over time. However, on the basis of the current literature surprisingly little is known about the biological function of Prp17, except for a role in the late splicing process and more recently also in the cell cycle (Ben Yehuda *et al.*, 1998). Mutational analysis of the 3'ss of pre-mRNA in yeast demonstrated that the association of Prp17 and Prp16 with the spliceosome is not affected and suggested that both these factors function at the second catalytic step of splicing but prior to 3'ss recognition (Zhou and Reed, 1998). Proteins Prp16 and Prp17 associate with the spliceosome independently of one another, and cross-linking experiments have suggested that Prp17 acts subsequently to the association of Prp16 with the spliceosome. Indeed, our data suggest that the behavior of Prp17 differs from that of the other step II factors.

In our second quantitative analysis we have compared the amounts of proteins associated on the different pre-mRNAs over time (see Results, section 4.4.4). This comparison reflects more drastically the differences in the association of spliceosomal proteins with the different pre-mRNAs. Strikingly, U1 snRNP-specific proteins readily associate with PM5 pre-mRNA even upon very short incubation under splicing conditions (0 min) but much less with the 5'ss-deleted PM5 pre-mRNA. As expected, the amount of U1 snRNP on the PM5 pre-mRNA decreases over time, owing the fact that over time the first step of splicing takes place and, concomitantly with it, there occur structural changes that release U1 snRNPs from the spliceosome (for review see Wahl *et al.*, 2009). Importantly, the same pattern is observed with the Sm proteins; this is consistent with the dissociation of U1 (see above) and U4 snRNP from the spliceosome upon rearrangement of the U4/U6 di-snRNP to form new base pairs between U6 and U2 snRNP (Hausner *et al.*, 1990; Sashital *et al.*, 2004; Sun and Manley, 1995).

Even more impressive is the association of the Prp19/CDC5L proteins and the step 2 factors on the PM5 pre-mRNA as compared with the 5'ss-deleted PM5 pre-mRNA. After 5 min under splicing condition we observe a dramatic increase in the amounts of those proteins associated with the PM5 but not with 5'ss-deleted PM5 pre-mRNA.

When comparing these quantitative results with those obtained for single pre-mRNA, where we observed some minor differences in the assembly of the proteins on the various pre-mRNAs, we have to conclude that – at least under the conditions that we describe here (i.e. incubation of nuclear extract with pre-mRNAs under splicing conditions and affinity

purification without further density-gradient purification of the spliceosomal complexes formed) – the kinetics of assembly of the proteins on the different pre-mRNAs are surprisingly similar but that proteins assemble in much higher yield on splicing-active pre-mRNA than on the splicing-inactive mutants.

This can be explained by the observation that both the mutants (5'ss and BPS-deleted PM5 pre-mRNA) show a very low remaining activity in splicing (see Results, section 4.4.1). Indeed, mass spectrometry is such a sensitive technique that it is able to monitor those proteins on the pre-mRNA mutants that reflect the residual splicing activity.

However, final conclusions can only be drawn after the evaluation of the remaining other spliceosomal proteins, e.g. the U5 snRNP-specific proteins and the A and B complex specific proteins and – more importantly – after having performed the above-mentioned pulse-chase experiment with the final isolation and mass spectrometric analysis of the assembled A, B, and C complexes.

6 References

Abbatiello, S.E., Pan, Y.X., Zhou, M., Wayne, A.S., Veenstra, T.D., Hunger, S.P., Kilberg, M.S., Eyler, J.R., Richards, N.G., and Conrads, T.P. (2008). Mass spectrometric quantification of asparagine synthetase in circulating leukemia cells from acute lymphoblastic leukemia patients. *J Proteomics* 71, 61-70.

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

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Luhrmann, 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.

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.

Andersen, J.S., Lam, Y.W., Leung, A.K., Ong, S.E., Lyon, C.E., Lamond, A.I., and Mann, M. (2005). Nucleolar proteome dynamics. *Nature* 433, 77-83.

Anderson, K., and Moore, M.J. (1997). Bimolecular exon ligation by the human spliceosome. *Science* 276, 1712-1716.

Anderson, L., and Hunter, C.L. (2006). Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* 5, 573-588.

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

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

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 389, 1017-1031.

Bearden, J.C., Jr. (1978). Quantitation of submicrogram quantities of protein by an improved protein-dye binding assay. *Biochim Biophys Acta* 533, 525-529.

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

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

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2007). Composition and three-

dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J* 26, 1737-1748.

Ben Yehuda, S., Dix, I., Russell, C.S., Levy, S., Beggs, J.D., and Kupiec, M. (1998). Identification and functional analysis of hPRP17, the human homologue of the PRP17/CDC40 yeast gene involved in splicing and cell cycle control. *RNA* 4, 1304-1312.

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

Bienvenut, W.V., Deon, C., Pasquarello, C., Campbell, J.M., Sanchez, J.C., Vestal, M.L., and Hochstrasser, D.F. (2002). Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics* 2, 868-876.

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

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

Boersema, P.J., Aye, T.T., van Veen, T.A., Heck, A.J., and Mohammed, S. (2008). Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* 8, 4624-4632.

Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* 4, 484-494.

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

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

Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. (2007). Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol Cell Proteomics* 6, 2139-2149.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of precursors to mRNA by the Spliceosome. In *The RNA world*, R.F. Gesteland, T.R. Cech, and T.R. Atkins, eds. (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press), pp. 525-560.

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.

Choi, Y.D., Grabowski, P.J., Sharp, P.A., and Dreyfuss, G. (1986). Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science* 231, 1534-1539.

Collins, C.A., and Guthrie, C. (1999). Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev* 13, 1970-1982.

Conrads, T.P., Alving, K., Veenstra, T.D., Belov, M.E., Anderson, G.A., Anderson, D.J., Lipton, M.S., Pasa-Tolic, L., Udseth, H.R., Chrisler, W.B., Thrall, B.D., and Smith, R.D. (2001). Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and ¹⁵N-metabolic labeling. *Anal Chem* 73, 2132-2139.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26, 1367-1372.

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

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

Desiderio, D.M., and Kai, M. (1983). Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed Mass Spectrom* 10, 471-479.

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.

Dönmez, G. (2006). Investigation of the higher order structure of the spliceosomal RNA network (Göttingen, Georg-August-Universität Göttingen).

Doherty, M.K., Whitehead, C., McCormack, H., Gaskell, S.J., and Beynon, R.J. (2005). Proteome dynamics in complex organisms: using stable isotopes to monitor individual protein turnover rates. *Proteomics* 5, 522-533.

Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., and Burd, C.G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* 62, 289-321.

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

Edman, P. (1949). A method for the determination of amino acid sequence in peptides. *Arch Biochem* 22, 475.

Elias, J.E., Haas, W., Faherty, B.K., and Gygi, S.P. (2005). Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* 2, 667-675.

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

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64-71.

Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* 100, 6940-6945.

Gilar, M., Olivova, P., Daly, A.E., and Gebler, J.C. (2005). Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci* 28, 1694-1703.

Gilchrist, A., Au, C.E., Hiding, J., Bell, A.W., Fernandez-Rodriguez, J., Lesimple, S., Nagaya, H., Roy, L., Gosline, S.J., Hallett, M., Paiement, J., Kearney, R.E., Nilsson, T., and Bergeron, J.J. (2006). Quantitative proteomics analysis of the secretory pathway. *Cell* 127, 1265-1281.

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

Graub, R., Lancero, H., Pedersen, A., Chu, M., Padmanabhan, K., Xu, X.Q., Spitz, P., Chalkley, R., Burlingame, A.L., Stokoe, D., Bernstein, H.S. (2008). Cell cycle-dependent phosphorylation of human CDC5 regulates RNA processing. *Cell Cycle* 7, 1795-1803.

Graumann, J., Hubner, N.C., Kim, J.B., Ko, K., Moser, M., Kumar, C., Cox, J., Scholer, H., and Mann, M. (2008). Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol Cell Proteomics* 7, 672-683.

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

Green, M.R. (1991). Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu Rev Cell Biol* 7, 559-599.

Grillari, J., Ajuh, P., Stadler, G., Loscher, M., Voglauer, R., Ernst, W., Chusainow, J., Eisenhaber, F., Pokar, M., Fortschegger, K., Grey, M., Lamond, A.I., and Katinger, H. (2005). SNEV is an evolutionarily conserved splicing factor whose oligomerization is necessary for spliceosome assembly. *Nucleic Acids Res* 33, 6868-6883.

Grote, M., Wolf, E., Will, C.L., Lemm, I., Agafonov, D.E., Schomburg, A., Fischle, W., Urlaub, H., and Luhrmann, R. (2010). The Molecular Architecture of the Human Prp19/CDC5L Complex. *Mol Cell Biol* 30, 2105-2119

Gruhler, A., Olsen, J.V., Mohammed, S., Mortensen, P., Faergeman, N.J., Mann, M., and Jensen, O.N. (2005). Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 4, 310-327.

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

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17, 994-999.

Hanke, S., Besir, H., Oesterhelt, D., and Mann, M. (2008). Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *J Proteome Res* 7, 1118-1130.

Hartinger, J., Stenius, K., Hogemann, D., and Jahn, R. (1996). 16-BAC/SDS-PAGE: a two-dimensional gel electrophoresis system suitable for the separation of integral membrane proteins. *Anal Biochem* 240, 126-133.

Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Luhrmann, R. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A* *99*, 16719-16724.

Hausner, T.P., Giglio, L.M., and Weiner, A.M. (1990). Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev* *4*, 2146-2156.

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

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

Holzmann, J., Pichler, P., Madalinski, M., Kurzbauer, R., and Mechtler, K. (2009). Stoichiometry determination of the MP1-p14 complex using a novel and cost-efficient method to produce an equimolar mixture of standard peptides. *Anal Chem* *81*, 10254-10261.

Hsu, J.L., Huang, S.Y., Chow, N.H., and Chen, S.H. (2003). Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* *75*, 6843-6852.

Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. *J Mass Spectrom* *40*, 430-443.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* *4*, 1265-1272.

Johnson, K.L., and Muddiman, D.C. (2004). A method for calculating ¹⁶O/¹⁸O peptide ion ratios for the relative quantification of proteomes. *J Am Soc Mass Spectrom* *15*, 437-445.

Karas, M., Bachmann, D., Bahr, U., and Hillenkamp, F. (1987). Matrix-assisted ultra-violet laser desorption of non-volatile compounds. *Int J Mass Spectrom Ion Proc* *78*, 53-58.

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

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

Kirkpatrick, D.S., Gerber, S.A., and Gygi, S.P. (2005). The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* *35*, 265-273.

Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* *26*, 231-243.

Klose, J. (1999). Large-gel 2-D electrophoresis. *Methods Mol Biol* *112*, 147-172.

Knochenmuss, R., Stortelder, A., Breuker, K., and Zenobi, R. (2000). Secondary ion-molecule reactions in matrix-assisted laser desorption/ionization. *J Mass Spectrom* **35**, 1237-1245.

Koehler, C.J., Strozynski, M., Kozielski, F., Treumann, A., and Thiede, B. (2009). Isobaric peptide termini labeling for MS/MS-based quantitative proteomics. *J Proteome Res* **8**, 4333-4341.

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

Krijgsveld, J., Gauci, S., Dormeyer, W., and Heck, A.J. (2006). In-gel isoelectric focusing of peptides as a tool for improved protein identification. *J Proteome Res* **5**, 1721-1730.

Krijgsveld, J., Ketting, R.F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C.P., Plasterk, R.H., and Heck, A.J. (2003). Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat Biotechnol* **21**, 927-931.

Kruger, M., Moser, M., Ussar, S., Thievensen, I., Lubber, C.A., Forner, F., Schmidt, S., Zanivan, S., Fassler, R., and Mann, M. (2008). SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **134**, 353-364.

Kuhn, A.N., van Santen, M.A., Schwienhorst, A., Urlaub, H., and Luhrmann, R. (2009). Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. *RNA* **15**, 153-175.

Kuhn, E., Wu, J., Karl, J., Liao, H., Zolg, W., and Guild, B. (2004). Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and ¹³C-labeled peptide standards. *Proteomics* **4**, 1175-1186.

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

Kuster, B., Schirle, M., Mallick, P., and Aebersold, R. (2005). Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol* **6**, 577-583.

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

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

Lam, Y.W., Lamond, A.I., Mann, M., and Andersen, J.S. (2007). Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Curr Biol* **17**, 749-760.

Langenfeld, E., Zanger, U.M., Jung, K., Meyer, H.E., and Marcus, K. (2009). Mass spectrometry-based absolute quantification of microsomal cytochrome P450 2D6 in human liver. *Proteomics* **9**, 2313-2323.

Liao, L., Park, S.K., Xu, T., Vanderklish, P., and Yates, J.R., 3rd (2008). Quantitative proteomic analysis of primary neurons reveals diverse changes in synaptic protein content in *fmr1* knockout mice. *Proc Natl Acad Sci U S A* 105, 15281-15286.

Liautard, J.P., Sri-Widada, J., Brunel, C., and Jeanteur, P. (1982). Structural organization of ribonucleoproteins containing small nuclear RNAs from HeLa cells. Proteins interact closely with a similar structural domain of U1, U2, U4 and U5 small nuclear RNAs. *J Mol Biol* 162, 623-643.

Liu, H., Sadygov, R.G., and Yates, J.R., 3rd (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76, 4193-4201.

Macfarlane, D.E. (1989). Two dimensional benzyldimethyl-n-hexadecylammonium chloride---sodium dodecyl sulfate preparative polyacrylamide gel electrophoresis: a high capacity high resolution technique for the purification of proteins from complex mixtures. *Anal Biochem* 176, 457-463.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298, 2205-2208.

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

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

Malca, H., Shomron, N., and Ast, G. (2003). The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. *Mol Cell Biol* 23, 3442-3455.

Mallick, P., Schirle, M., Chen, S.S., Flory, M.R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B., and Aebersold, R. (2007). Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* 25, 125-131.

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

Massenet, S., Mougin, A., and Branlant, C. (1998). Posttranscriptional modifications in the U small nuclear RNAs. In *The Modification and Editing of RNA*, H. Grosjean, and R. Benne, eds. (Washington D.C., ASM Press), pp. 201-227.

Mayeda, A., and Krainer, A.R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* 68, 365-375.

Mintz, M., Vanderver, A., Brown, K.J., Lin, J., Wang, Z., Kaneski, C., Schiffmann, R., Nagaraju, K., Hoffman, E.P., and Hathout, Y. (2008). Time series proteome profiling to study endoplasmic reticulum stress response. *J Proteome Res* 7, 2435-2444.

Mirgorodskaya, O.A., Kozmin, Y.P., Titov, M.I., Korner, R., Sonksen, C.P., and Roepstorff, P. (2000). Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards. *Rapid Commun Mass Spectrom* 14, 1226-1232.

Moore, M.J., Query, C.C., and Sharp, P.A. (1993). Splicing of precursors to mRNA by the spliceosome. In *RNA world*, R.F. Gesteland, and T.R. Atkins, eds. (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press), pp. 303-357.

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

Neer, E.J., Schmidt, C.J., Nambudripad, R., and Smith, T.F. (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature* 371, 297-300.

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

Neuhoff, V., Arold, N., Taube, D., and Ehrhardt, W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 9, 255-262.

Niggeweg, R., Kocher, T., Gentzel, M., Buscaino, A., Taipale, M., Akhtar, A., and Wilm, M. (2006). A general precursor ion-like scanning mode on quadrupole-TOF instruments compatible with chromatographic separation. *Proteomics* 6, 41-53.

Nilsen, T.W. (1998). RNA-RNA interactions in nuclear pre-mRNA splicing. In *RNA structure and function*, R.W. Simmons, and M. Grunerg-Manaro, eds. (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press), pp. 1793-1309.

Oda, Y., Huang, K., Cross, F.R., Cowburn, D., and Chait, B.T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* 96, 6591-6596.

Oellerich, T., Gronborg, M., Neumann, K., Hsiao, H.H., Urlaub, H., and Wienands, J. (2009). SLP-65 phosphorylation dynamics reveals a functional basis for signal integration by receptor-proximal adaptor proteins. *Mol Cell Proteomics* 8, 1738-1750.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007-4021.

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.

Ohrt, T., Merkle, D., Birkenfeld, K., Echeverri, C.J., and Schwill, P. (2006). In situ fluorescence analysis demonstrates active siRNA exclusion from the nucleus by Exportin 5. *Nucleic Acids Res* 34, 1369-1380.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., and Ahn, N.G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 4, 1487-1502.

Olsen, J.V., Ong, S.E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* 3, 608-614.

Olsen, J.V., de Godoy, L.M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 4, 2010-2021.

- Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., Brunak, S., and Mann, M. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* 3, ra3.
- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1, 376-386.
- Ong, S.E., Foster, L.J., and Mann, M. (2003). Mass spectrometric-based approaches in quantitative proteomics. *Methods* 29, 124-130.
- Ong, S.E., and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 1, 252-262.
- Pan, C., Olsen, J.V., Daub, H., and Mann, M. (2009a). Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics. *Mol Cell Proteomics* 8, 2796-2808.
- Pan, S., Aebersold, R., Chen, R., Rush, J., Goodlett, D.R., McIntosh, M.W., Zhang, J., and Brentnall, T.A. (2009b). Mass spectrometry based targeted protein quantification: methods and applications. *J Proteome Res* 8, 787-797.
- Pan, S., Zhang, H., Rush, J., Eng, J., Zhang, N., Patterson, D., Comb, M.J., and Aebersold, R. (2005). High throughput proteome screening for biomarker detection. *Mol Cell Proteomics* 4, 182-190.
- Patton, J.G., Mayer, S.A., Tempst, P., and Nadal-Ginard, B. (1991). Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. *Genes Dev* 5, 1237-1251.
- Patton, J.R., Habets, W., van Venrooij, W.J., and Pederson, T. (1989). U1 small nuclear ribonucleoprotein particle-specific proteins interact with the first and second stem-loops of U1 RNA, with the A protein binding directly to the RNA independently of the 70K and Sm proteins. *Mol Cell Biol* 9, 3360-3368.
- Patton, J.R., and Pederson, T. (1988). The Mr 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 5' stem-loop of U1 RNA and interacts with Sm domain proteins. *Proc Natl Acad Sci U S A* 85, 747-751.
- 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.
- Pratt, J.M., Simpson, D.M., Doherty, M.K., Rivers, J., Gaskell, S.J., and Beynon, R.J. (2006). Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat Protoc* 1, 1029-1043.
- Query, C.C., Bentley, R.C., and Keene, J.D. (1989). A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. *Mol Cell Biol* 9, 4872-4881.
- Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev* 8, 587-597.
- Raker, V.A., Hartmuth, K., Kastner, B., and Luhrmann, R. (1999). Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol Cell Biol* 19, 6554-6565.

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

Ramos-Fernandez, A., Lopez-Ferrer, D., and Vazquez, J. (2007). Improved method for differential expression proteomics using trypsin-catalyzed 18O labeling with a correction for labeling efficiency. *Mol Cell Proteomics* 6, 1274-1286.

Rao, K.C., Carruth, R.T., and Miyagi, M. (2005). Proteolytic 18O labeling by peptidyl-Lys metalloendopeptidase for comparative proteomics. *J Proteome Res* 4, 507-514.

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

Reynolds, K.J., Yao, X., and Fenselau, C. (2002). Proteolytic 18O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *J Proteome Res* 1, 27-33.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A., and Pappin, D.J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3, 1154-1169.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular cloning - a Laboratory Manual* (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Sanford, J.R., Longman, D., and Caceres, J.F. (2003). Multiple roles of the SR protein family in splicing regulation. *Prog Mol Subcell Biol* 31, 33-58.

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

Schagger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* 166, 368-379.

Schagger, H., Aquila, H., and Von Jagow, G. (1988). Coomassie blue-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for direct visualization of polypeptides during electrophoresis. *Anal Biochem* 173, 201-205.

Schagger, H., and von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem* 199, 223-231.

Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J., and Mattaj, I.W. (1989). Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J* 8, 4163-4170.

Schmidt, A., Karas, M., and Dulcks, T. (2003). Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? *J Am Soc Mass Spectrom* 14, 492-500.

Schmidt, A., Kellermann, J., and Lottspeich, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 5, 4-15.

- Schnolzer, M., Jedrzejewski, P., and Lehmann, W.D. (1996). Protease-catalyzed incorporation of ^{18}O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* 17, 945-953.
- Schoenle, E.J., Adams, L.D., and Sammons, D.W. (1984). Insulin-induced rapid decrease of a major protein in fat cell plasma membranes. *J Biol Chem* 259, 12112-12116.
- Schulze, W.X., and Mann, M. (2004). A novel proteomic screen for peptide-protein interactions. *J Biol Chem* 279, 10756-10764.
- Schwanhausser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009). Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* 9, 205-209.
- Schwer, B. (2001). A new twist on RNA helicases: DExH/D box proteins as RNAPases. *Nat Struct Biol* 8, 113-116.
- Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58-63.
- 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, R., Kumar, C., Zougman, A., Zhang, Y., Podtelejnikov, A., Cox, J., Wisniewski, J.R., and Mann, M. (2007). Analysis of the mouse liver proteome using advanced mass spectrometry. *J Proteome Res* 6, 2963-2972.
- Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P., and Geromanos, S.J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* 5, 144-156.
- Singh, R., and Reddy, R. (1989). Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc Natl Acad Sci U S A* 86, 8280-8283.
- Singh, S., Springer, M., Steen, J., Kirschner, M.W., and Steen, H. (2009). FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *J Proteome Res* 8, 2201-2210.
- Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R., and Domon, B. (2007). High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* 6, 1809-1817.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315-326.
- Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell* 9, 31-44.
- Sun, J.S., and Manley, J.L. (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev* 9, 843-854.
- Talkington, M.W., Siuzdak, G., and Williamson, J.R. (2005). An assembly landscape for the 30S ribosomal subunit. *Nature* 438, 628-632.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1988). Laser ionization Time-of-flight mass spectrometry. *Rapid Comm Mass Spectrom* 2, 151-153.

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.

Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75, 1895-1904.

Umen, J.G., and Guthrie, C. (1995). The second catalytic step of pre-mRNA splicing. *RNA* 1, 869-885.

Urlaub, H., Hartmuth, K., Kostka, S., Grelle, G., and Luhrmann, R. (2000). A general approach for identification of RNA-protein cross-linking sites within native human spliceosomal small nuclear ribonucleoproteins (snRNPs). Analysis of RNA-protein contacts in native U1 and U4/U6.U5 snRNPs. *J Biol Chem* 275, 41458-41468.

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

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

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

Waanders, L.F., Hanke, S., and Mann, M. (2007). Top-down quantitation and characterization of SILAC-labeled proteins. *J Am Soc Mass Spectrom* 18, 2058-2064.

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

Washburn, M.P., Wolters, D., and Yates, J.R., 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19, 242-247.

Wepf, A., Glatter, T., Schmidt, A., Aebersold, R., and Gstaiger, M. (2009). Quantitative interaction proteomics using mass spectrometry. *Nat Methods* 6, 203-205.

Will, C.L., and Luhrmann, R. (2005). Spliceosome structure and function. In *RNA world III*, R.F. Gesteland, and T.R. Cech, eds. (CSH laboratory press), pp. 369-400.

Will, C.L., Rumpler, S., Klein Gunnewiek, J., van Venrooij, W.J., and Luhrmann, R. (1996). In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res* 24, 4614-4623.

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

Williamson, J.R. (2005). Assembly of the 30S ribosomal subunit. *Q Rev Biophys* 38, 397-403.

Wilm, M. (2009). Quantitative proteomics in biological research. *Proteomics* 9, 4590-4605.

Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D.A., and White, F.M. (2007). Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci U S A* 104, 5860-5865.

Yao, X., Freas, A., Ramirez, J., Demirev, P.A., and Fenselau, C. (2001). Proteolytic ¹⁸O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* 73, 2836-2842.

Zhou, Z., and Reed, R. (1998). Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *EMBO J* 17, 2095-2106.

7 Appendix

Additional information

Table A.1: Additional proteins identified in the hPrp19/CDC5L complex by LC-MS/MS. The number of unique peptides for every protein identified after hydrolysis in 80 % (v/v) acetonitrile and 8M/2M urea, respectively, is given for three replicates.

		80 % (v/v) Acetonitrile			8M / 2M Urea		
		# Unique Peptides			# Unique Peptides		
Protein	Accession no.	# 1	# 2	# 3	# 1	# 2	# 3
ABP130	gi 6009492			1			1
alpha-tubulin	gi 37492	5			5	2	
alpha-tubulin	gi 32015		1				
aquarius	gi 38788372	3		1	2		
BC273239_1	gi 4559318						1
BCL2-associated athanogene 2	gi 4757834		3	1	1		
beta actin variant	gi 62897625					5	
coiled-coil domain containing 12	gi 21389497	3			3		1
crn	gi 27372168	4		9	7	4	10
cyclophilin-33A	gi 2828149	1	2		1		
DEAH (Asp-Glu-Ala-His) box polypeptide 8	gi 4826690				1		
dermcidin preproprotein	gi 16751921	2	2	1		2	
Dsc1a precursor	gi 457464	1					
E2F-associated phosphoprotein	gi 7020780	1					
eIF-4B	gi 124219	1					
filaggrin family member 2	gi 62122917						2
FLJ00137 protein	gi 18676480			1			
GTPase activating Rap/RanGAP domain-like 3	gi 119608065		1				
hCG1984029	gi 119572483	1					
HEF like Protein	gi 9650711						1
hornerin	gi 28557150	2			2		
hU1-70K-like protein (216 AA)	gi 36100			3	2		3
hypothetical protein LOC445577	gi 149363702	1					
ISY1 protein	gi 13938521	3	1	2	1	1	
KIAA0788 protein	gi 20521660			2			
KIAA1177	gi 6330235	5			3		
KIAA1620 protein	gi 10047317				1		
microtubule-associated protein tau isoform 2	gi 1790878	1					
MMP27	gi 37182623	1					

mutant beta-actin	gi 28336	5	4		5		
Myomesin-2	gi 1709093	1					
Nogo-A protein	gi 9408096					1	
prostate differentiation factor	gi 2290972			1			
PRP8 protein	gi 91208426	7	5	3	3	5	5
putative ORF	gi 763429	1					
sirtuin 1	gi 7657575	3		1			1
Skb 1 Hs	gi 2323410	5	6		5		
SKI-interacting protein	gi 6912676	1	2		4		
small nuclear ribonucleoprotein polypeptide A	gi 4759156				1		
SmD2	gi 4759158			2			1
SmF	gi 4507131	1					1
SmG	gi 4507133	1			2		
SYF2 homolog, RNA splicing factor isoform 1	gi 7661636				1	2	
tubulin 5-beta	gi 35959		1				
tubulin beta	gi 223429				3		
U5-116K	gi 24474791	3	3	3	8	4	3
U5-200K	gi 40217847	3	3		4	1	2
WD repeat domain 77	gi 13129110				4		
Williams-Beuren syndrome chromosome region 27	gi 30795190			1			
ZNF461 protein	gi 20306351			1			

Table A.2: Spectral Count for peptides selected for quantification and their miscleaved versions. The number of spectra is given for different replicates for hydrolysis in 8M/2M urea and 80 % (v/v) acetonitrile analyzed by LC-ESI-MS/MS and LC-offline MALDI-ToF/ToF. The number in parentheses represents the mascot score obtained from the different spectra.

Protein	Peptide sequence	hydrolysis in the presence of 8M/2M urea						hydrolysis in the presence of 80 % (v/v) acetonitrile					
		# spectra (Mascot scores)						# spectra (Mascot scores)					
		LC-ESI-MS/MS			LC-offline MALDI-ToF/ToF			LC-ESI-MS/MS			LC-offline MALDI-ToF/ToF		
	YADLILLEK	1 (54)	1 (54)	2 (34; 47)	1 (51)	1 (51)	1 (54)	1 (51)	1 (51)	2 (42; 47)	1 (62)	1 (51)	
CDC5L	YADLILLEKETLK	2 (40; 68)	2 (34; 64)	4 (17; 36; 44; 61)	1 (80)	1 (80)	2 (15; 72)	3 (23; 29; 68)	2 (42; 60)	2 (7; 48)	1 (32)	1 (32)	
	ELOHRYADLLEK												
	LGLLGLPAPK	2 (8; 51)	1 (41)	3 (17; 62; 69)	1 (53)	1 (58)	1 (40)	1 (56)	2 (32; 66)	1 (62)	1 (58)	1 (58)	
	ESREHLR.LGLLGLPAPK			1 (18)				1 (28)	1 (53)				
CDC5L	EHLR.LGLLGLPAPK							1 (5)	3 (12; 15; 21)				
	LGLLGLPAPK.NDFEIVLPENAEK						1 (27)	1 (31)	1 (29)				
	ILGGYQSR	1 (52)	1 (49)	3 (16; 22; 48)	1 (43)	1 (51)	1 (57)	1 (44)	2 (35; 51)	1 (62)	1 (41)	1 (41)	
CDC5L	MK.IILGGYQSR	1 (34)					1 (14)	1 (64)	2 (50; 50)				
	NWVFDK	1 (15)	1 (25)	4 (7; 12; 14; 26)	1 (43)	1 (30)	1 (13)	1 (22)	1 (17)	1 (38)	1 (31)	1 (31)	
hPp19	ILTGADK.NWVFDK	2 (10; 84)											
	TLQLDNNVEFK	2 (19; 54)	1 (43)	12 (6; 10; 19; 20; 20; 25; 41; 42; 46; 48; 49)	1 (71)	1 (93)	2 (62; 32)	1 (63)	4 (15; 23; 44; 46)	1 (75)	1 (69)	1 (69)	
hPp19	NFK.TLQLDNNFEVK								1 (41)				
PRL1	TGYNFQR	1 (45)	1 (50)	1 (44)		1 (10)	1 (47)		1 (43)				
	HYTFASGSPDNIK	2 (13; 64)	2 (26; 64)	2 (38; 72)	1 (85)	1 (99)	2 (31; 64)	2 (23; 66)	2 (45; 72)	1 (96)	1 (88)	1 (88)	
PRL1	AVLLHPR.HYTFASGSPDNIK							1 (3)	1 (15)				
	EAAAALVEETR	1 (82)	1 (103)	1 (64)			1 (97)	1 (97)	1 (94)	1 (59)	1 (22)	1 (22)	
SPF27	EAAAALVEEETR.R	2 (18; 30)	2 (31; 38)	4 (22; 29; 34; 44)	1 (81)	1 (43)	2 (18; 37)	2 (19; 34)	2 (14; 34)	1 (79)	1 (41)	1 (41)	
	TIVQLENEYQIK	1 (53)	2 (50; 50)	6 (7; 23; 35; 44; 53; 66)			2 (57; 60)	2 (47; 60)	3 (33; 60; 64)	1 (71)	1 (65)	1 (65)	
SPF27	TIVQLENEYQIK.QQHGEANK							1 (30)	1 (29)				
	FVDILGLR	1 (47)	1 (59)	1 (54)	1 (17)	1 (24)	1 (52)	1 (67)	1 (49)	1 (32)	1 (26)	1 (26)	
CTNBL1	VLDHAMIGPE3TDNCHK.FVDILGLR								2 (4; 36)				
	FVDILGLR.TIFPLFMK							2 (15; 35)					

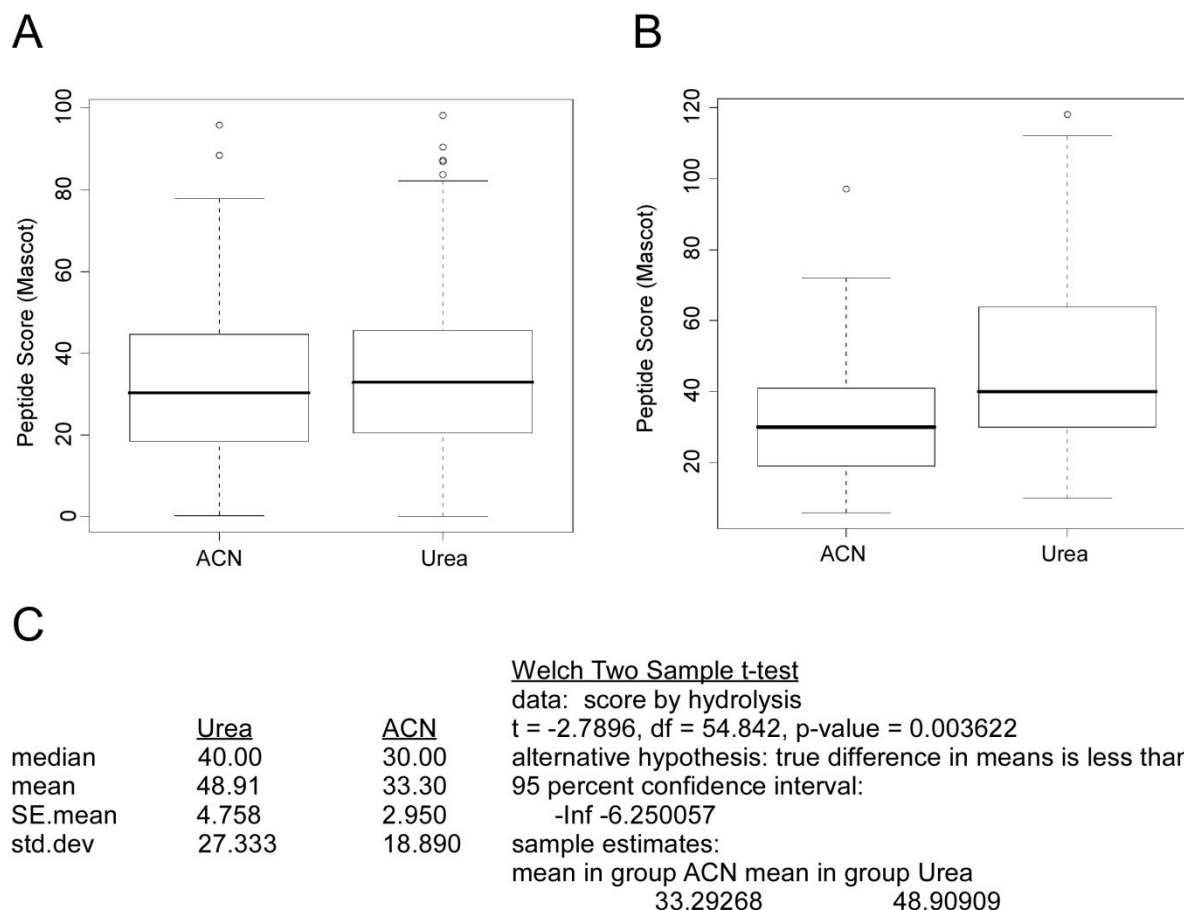


Figure A.1: Miscleaved peptides generated in the presence of urea show higher peptide scores for the peptide sequences selected for quantification. (A) The boxplot does not show any significant difference in the Mascot peptide scores for miss cleavage site-containing peptides generated in the presence of urea or acetonitrile. (B) Strikingly, the boxplot for Mascot peptide scores for miscleaved peptides selected as standard peptides for quantification generated in the presence of urea and acetonitrile shows that Mascot peptide scores for miscleaved peptides obtained by hydrolysis in urea are higher. (C) Output of the statistical analysis. Differences of mean peptide scores for miscleaved peptides selected for quantification were tested for statistical significance by applying Welch's two sample t-test (R 2.8.0). The observed difference is statistically significant (p-value < 0.5).

	<u>ACN</u>	<u>Urea</u>	<u>Welch Two Sample t-test</u>
nbr.val	123	167	data: length by Hydrolysis
min	6	6	t = 1.8197, df = 225.351, p-value = 0.03506
max	36	33	alternative hypothesis: true difference in means is greater than 0
median	12.000	12.000	95 percent confidence interval:
mean	14.073	12.760	0.1212455 Inf
SE.mean	0.597	0.405	sample estimates:
var	43.790	27.448	mean in group ACN mean in group Urea
std.dev	6.617	5.239	14.07317 12.76048

Figure A.2: Incompletely cleaved peptides obtained by hydrolysis in the presence of acetonitrile are significantly longer than incompletely cleaved peptides obtained by hydrolysis in the presence of urea. Output of the statistical analysis.

Table A.3: Three MRM transitions for each endogenous and standard peptide were designed. In all cases, the doubly charged precursor was chosen as Q1 mass, and the three most intense fragment ions with an m/z above that of the precursor were chosen as Q3 masses. For the selected MRM transitions declustering potential (DP), entrance potential (EP), collision energy (CE), and collision cell exit potential (CXP) were optimized.

Protein	Peptide sequence	Q1	Q3	Fragment	DP [V]	EP [V]	CE [V]	CXP [V]
CDC5L	ILLGGYQSR light	503.79	610.29	y5	95	14	29	18
	ILLGGYQSR light	503.79	667.32	y6	95	14	30	19
	ILLGGYQSR light	503.79	780.4	y7	95	14	28	7
	ILLGGYQSR heavy	508.79	620.3	y5	95	14	29	18
	ILLGGYQSR heavy	508.79	677.32	y6	95	14	30	19
	ILLGGYQSR heavy	508.79	790.41	y7	95	14	28	7
CDC5L	LGLLGLPAPK light	489.82	582.36	y6	105	13	25	17
	LGLLGLPAPK light	489.82	695.45	y7	105	13	26	21
	LGLLGLPAPK light	489.82	865.55	y9	105	13	28	23
	LGLLGLPAPK heavy	493.82	590.38	y6	105	13	25	17
	LGLLGLPAPK heavy	493.82	703.46	y7	105	13	26	21
	LGLLGLPAPK heavy	493.82	873.56	y9	105	13	28	23
CDC5L	YADLLEK light	482.77	615.41	y5	97	13	28	17
	YADLLEK light	482.77	730.43	y6	97	13	24	21
	YADLLEK light	482.77	801.47	y7	97	13	26	23
	YADLLEK heavy	486.28	622.42	y5	97	13	28	17
	YADLLEK heavy	486.28	737.45	y6	97	13	24	21
	YADLLEK heavy	486.28	808.49	y7	97	13	26	23
CTNBL1	FVDILGLR light	466.78	571.39	y5	73	11	32	15
	FVDILGLR light	466.78	686.42	y6	73	11	25	15
	FVDILGLR light	466.78	785.49	y7	73	11	27	15
	FVDILGLR heavy	470.21	578.41	y5	73	11	32	15
	FVDILGLR heavy	470.21	693.44	y6	73	11	25	15
	FVDILGLR heavy	470.21	792.51	y7	73	11	27	15
PRL1	HYTFASGSPDNIK light	718.84	817.41	y8	135	10	41	30
	HYTFASGSPDNIK light	718.84	888.44	y9	135	10	41	30
	HYTFASGSPDNIK light	718.84	1136.56	y11	135	10	39	30
	HYTFASGSPDNIK heavy	722.35	824.42	y8	135	10	41	30
	HYTFASGSPDNIK heavy	722.35	895.46	y9	135	10	41	30
	HYTFASGSPDNIK heavy	722.35	1143.58	y11	135	10	39	30
PRL1	TGYNFQR light	443.21	564.29	y4	80	11	27	16
	TGYNFQR light	443.21	727.35	y5	80	11	25	16
	TGYNFQR light	443.21	784.37	y6	80	11	28	16
	TGYNFQR heavy	448.23	574.32	y4	80	11	27	16
	TGYNFQR heavy	448.23	737.38	y5	80	11	25	16
	TGYNFQR heavy	448.23	794.4	y6	80	11	28	16
hPrp19	TLQLDNNFEVK light	660.84	865.41	y7	123	8	34	8
	TLQLDNNFEVK light	660.84	978.49	y8	123	8	33	9.4
	TLQLDNNFEVK light	660.84	1106.55	y9	123	8	32	11
	TLQLDNNFEVK heavy	664.85	873.42	y7	123	8	34	8
	TLQLDNNFEVK heavy	664.85	986.5	y8	123	8	33	9.4
	TLQLDNNFEVK heavy	664.85	1114.56	y9	123	8	32	11
hPrp19	NVVVFDK light	410.73	508.28	y4	85	10	23	19
	NVVVFDK light	410.73	607.35	y5	85	10	20	18
	NVVVFDK light	410.73	706.41	y6	85	10	23	20
	NVVVFDK heavy	415.75	518.3	y4	85	10	23	19
	NVVVFDK heavy	415.75	617.37	y5	85	10	20	18
	NVVVFDK heavy	415.75	716.44	y6	85	10	23	20

SPF27	EAAAALVEEETR light	644.82	663.29	y5	110	10	31	20
	EAAAALVEEETR light	644.82	762.36	y6	110	10	34	23
	EAAAALVEEETR light	644.82	875.45	y7	110	10	34	25
	EAAAALVEEETR heavy	649.83	673.3	y5	110	10	31	20
	EAAAALVEEETR heavy	649.83	772.37	y6	110	10	34	23
	EAAAALVEEETR heavy	649.83	885.46	y7	110	10	34	25
SPF27	TIVQLENEYQIK light	795.94	907.49	y7	145	10	40	8.2
	TIVQLENEYQIK light	795.94	1036.53	y8	145	10	40	9.5
	TIVQLENEYQIK light	795.94	1149.62	y9	145	10	40	11
	TIVQLENEYQIK heavy	799.45	914.51	y7	145	10	40	8.2
	TIVQLENEYQIK heavy	799.45	1043.55	y8	145	10	40	9.5
	TIVQLENEYQIK heavy	799.45	1156.63	y9	145	10	40	11

Table A.4: Absolute amounts of proteins of the hPrp19/CDC5L complex. The protein complex was hydrolyzed in the presence of acetonitrile and a mix of standard peptides (100 fmol each) was added to the sample. Absolute amounts were then obtained by generating XICs of the endogenous and standard peptides. Hsp70 was found in reduced amounts when compared with other proteins of the hPrp19/CDC5L complex.

hPrp19/CDC5L complex [ng]	70	70	35
	[fmol]	[fmol]	[fmol]
hPrp19	252.0	255.0	128.9
CDC5	85.8	84.2	42.7
SPF27	31.9	38.2	16.5
PRL1	38.6	33.1	19.6
CTNNBL1	34.2	28.7	13.9
Hsp70	13.9	7.65	3.7

Table A.5: Proteins identified in the first replicate of the relative quantification of spliceosomal B and C complexes by iTRAQ. The accession number, the protein database description, the obtained protein score (Mascot), the protein mass, the determined protein ratio (B/C), the standard deviation (StDev), and the number of peptides used for quantification (#) is given for all proteins found. Proteins used for normalization are highlighted in red.

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
9G8	gi 72534660	splicing factor, arginine/serine-rich 7 [Homo sapiens]	171	28687	1.16	0.32	24
abstrakt	gi 21071032	DEAD-box protein abstrakt [Homo sapiens]	604	77694	0.08	0.06	48
Acinus	gi 109082928	PREDICTED: apoptotic chromatin condensation inducer 1 isoform 2 [Macaca mulatta]	148	77764	1.35	0.42	13
Acinus	gi 7662238	apoptotic chromatin condensation inducer 1 [Homo sapiens]	218	166097	1.03	0.23	9
Aquarius	gi 58257729	KIAA0560 protein [Homo sapiens]	380	187968	2.24	0.64	23
Aquarius	gi 38788372	aquarius [Homo sapiens]	2120	183588	0.09	0.05	96
ASF/SF2	gi 179074	alternative	209	33984	0.80	0.19	14
ASR2B	gi 13383501	ASR2B [Homo sapiens]	182	110669	1.00	0.27	15
UAP56	gi 4758112	HLA-B associated transcript 1 [Homo sapiens]	182	53651	2.01	0.44	11
BCLAF1	gi 7661958	BCL2-associated transcription factor 1 isoform 1 [Homo sapiens]	84	119552	1.61	0.24	23
C10orf4	gi 24432067	FRA10AC1 protein [Homo sapiens]	61	44908	0.10	0.06	2
C1orf55	gi 40255125	hypothetical protein LOC163859 [Homo sapiens]	73	54043	0.08	0.04	19
C9orf78	gi 7706557	chromosome 9 open reading frame 78 [Homo sapiens]	66	38613	0.14	0.05	6
Cactin (C19orf29)	gi 3253120	R31449_3 [Homo sapiens]	465	101398	0.17	0.13	21
CBP20	gi 19923387	nuclear cap binding protein subunit 2, 20kDa isoform 1 [Homo sapiens]	110	20001	0.71	0.16	8
CBP80	gi 24987336	Chain C, Structure Of The Human Nuclear Cap-Binding-Complex (Cbc)	805	97287	1.41	0.41	33
CCAP2	gi 6841518	HSPC148 [Homo sapiens]	75	30099	0.22	0.08	6
CCNK	gi 8980825	cyclin K [Homo sapiens]	29	44722	2.56		1
CDC2L5	gi 5870326	similar to KIAA0904; similar to AAA58424 (PID:g180492) [Homo sapiens]	23	86103	1.23	0.29	2
CDC5L	gi 11067747	CDC5-like [Homo sapiens]	703	103618	2.26	0.75	30
CDC5L	gi 11067747	CDC5-like [Homo sapiens]	361	103618	0.23	0.14	45
CLK1	gi 67551261	CDC-like kinase 1 [Homo sapiens]	117	63472	1.08	0.45	4
CLK4	gi 10190706	CDC-like kinase 4 [Homo sapiens]	42	62901	0.85		1

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
CDK10(PISSLRE)	gi 6226784	Cell division protein kinase 10 (Serine/threonine-protein kinase PISSLRE)	66	45362	0.20	0.09	3
CHERP	gi 1770394	DAN26 [Homo sapiens]	51	55656	5.36	0.05	2
CIP29	gi 119617243	hCG2016179, isoform CRA_d [Homo sapiens]	28	25459	3.57		1
CRNKL1/hSYF3	gi 50949465	hypothetical protein [Homo sapiens]	1070	96523	0.47	0.52	88
CTNBL1	gi 18644734	beta catenin-like 1 [Homo sapiens]	83	70832	1.44	0.46	4
CUG	gi 5729794	CUG triplet repeat, RNA-binding protein 1 isoform 1 [Homo sapiens]	31	55437	1.97	0.02	2
CXorf56	gi 11545813	hypothetical protein LOC63932 [Homo sapiens]	416	29821	0.08	0.08	12
CyclinM	gi 24308243	hypothetical protein LOC92002 [Homo sapiens]	65	26090	0.09		1
CypE	gi 45439318	peptidylprolyl isomerase E isoform 3 [Homo sapiens]	113	29821	0.13	0.05	8
DDX21	gi 2135315	RNA helicase Gu - human (fragment)	144	103155	1.67	0.58	7
DDX3	gi 2580550	dead box, X isoform [Homo sapiens]	51	78159	1.05	0.16	2
DDX35	gi 10439270	unnamed protein product [Homo sapiens]	380	84743	0.32	0.17	28
DDX9	gi 33878473	DHX9 protein [Homo sapiens]	54	73628	1.34		1
DGCR14	gi 12804313	Similar to expressed sequence 2 embryonic lethal [Homo sapiens]	147	56774	0.09	0.03	3
DHX34	gi 38158022	DEAH (Asp-Glu-Ala-His) box polypeptide 34 [Homo sapiens]	90	135588	0.49	0.33	7
EEF1A1	gi 31092	unnamed protein product [Homo sapiens]	80	57288	0.32	0.15	4
eIF4A3	gi 496902	translation initiation factor [Homo sapiens]	495	50492	0.15	0.07	31
ELAV	gi 1022961	HuR RNA binding protein	312	38915	4.76	2.29	27
ELG	gi 8923771	ELG protein [Homo sapiens]	92	43508	3.11	0.54	5
NOSIP	gi 7705716	eNOS interacting protein [Homo sapiens]	217	37167	0.11	0.08	9
ERH	gi 4758302	enhancer of rudimentary homolog [Homo sapiens]	154	13542	2.04	0.53	5
EXOSC2	gi 19923403	exosome component 2 [Homo sapiens]	41	35402	0.40	0.17	3
EXOSC5	gi 14043511	Exosome component 5 [Homo sapiens]	30	27406	0.52		1
FAM32A	gi 7661696	hypothetical protein LOC26017 [Homo sapiens]	57	16773	0.11	0.06	2
FAM50A	gi 4758220	XAP-5 protein [Homo sapiens]	158	46701	0.15	0.08	15

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
FAM50B	gi 6912326	family with sequence similarity 50, member B [Homo sapiens]	55	42766	0.13	0.03	2
FBP11	gi 34222504	Pre-mRNA-processing factor 40 homolog A (Formin-binding protein 3) (Formin-binding protein 11) (Huntingtin-interacting protein HYP A) (Huntingtin yeast partner A) (Fas ligand-associated factor 1) (NY-REN-6 antigen)	121	123665	1.26	0.21	4
FUSE3	gi 1575609	FUSE binding protein 3 [Homo sapiens]	72	67621	3.18	1.54	4
FUSIP1/SRp38	gi 5730079	FUS interacting protein (serine-arginine rich) 1 isoform 1 [Homo sapiens]	114	23166	0.91	0.52	15
Q9BRR8 (GPATC1)	gi 21361684	G patch domain containing 1 [Homo sapiens]	119	116769	0.40	0.09	3
G10	gi 32171175	G10 protein [Homo sapiens]	166	19898	0.39	0.19	14
GCIP p29	gi 7661636	SYF2 homolog, RNA splicing factor isoform 1 [Homo sapiens]	219	32930	0.18	0.14	11
GPKOW	gi 15811782	G patch domain and KOW motifs [Homo sapiens]	107	55983	0.66	0.21	6
GU2	gi 5174447	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1 [Homo sapiens]	50	38016	0.85		1
H2A	gi 32111	unnamed protein product [Homo sapiens]	0	16181	1.23		1
hECM2 (RBM22)	gi 8922328	RNA binding motif protein 22 [Homo sapiens]	419	51786	0.45	0.13	35
hISY1	gi 20149304	ISY1 splicing factor homolog [Homo sapiens]	247	37197	0.20	0.07	16
hnRNP A0	gi 1911429	A0=heterogeneous nuclear ribonucleoprotein [human, placenta, Peptide, 305 aa]	33	33758	1.25	0.67	2
hnRNP A1	gi 36102	unnamed protein product [Homo sapiens]	127	37105	3.44	1.81	6
hnRNP A2/B1	gi 119614244	heterogeneous nuclear ribonucleoprotein A2/B1, isoform CRA_d [Homo sapiens]	56	33161	4.63	1.96	6
hnRNP A3	gi 34740329	heterogeneous nuclear ribonucleoprotein A3 [Homo sapiens]	104	42925	4.62	1.12	6
hnRNP C1/C2	gi 306875	C protein	90	36316	1.72	0.47	23
hnRNP G	gi 542850	heterogeneous nuclear ribonucleoprotein G - human	322	50301	1.52	0.59	32
hnRNP H1	gi 5031753	heterogeneous nuclear ribonucleoprotein H1 [Homo sapiens]	160	51878	1.27	0.30	4

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
hnRNP K	gi 460789	transformation upregulated nuclear protein [Homo sapiens]	30	54441	2.97		1
hnRNP L	gi 133274	Heterogeneous nuclear ribonucleoprotein L (hnRNP L)	25	64788	5.84		1
hnRNP M	gi 187281	M4 protein	72	83739	1.32	0.31	8
hnRNP Q2	gi 15809588	hnRNP Q2 [Homo sapiens]	68	72196	0.84	0.12	3
hnRNP R	gi 73950226	PREDICTED: similar to heterogeneous nuclear ribonucleoprotein R isoform 1 [Canis familiaris]	85	65593	0.88	0.32	8
hnRNP U	gi 39645240	HNRPU protein [Homo sapiens]	44	87577	3.86	2.17	10
hPRP17	gi 7706657	cell division cycle 40 homolog [Homo sapiens]	654	73381	0.19	0.08	32
hPRP18	gi 4506123	PRP18 pre-mRNA processing factor 18 homolog [Homo sapiens]	32	45161	0.38		1
hPRP19	gi 7657381	PRP19/PSO4 pre-mRNA processing factor 19 homolog [Homo sapiens]	674	59981	0.62	0.36	61
hPRP2	gi 4503293	DEAH (Asp-Glu-Ala-His) box polypeptide 16 [Homo sapiens]	283	128919	0.54	0.16	19
hPRP22	gi 4826690	DEAH (Asp-Glu-Ala-His) box polypeptide 8 [Homo sapiens]	924	153030	0.13	0.12	55
hPRP38	gi 24762236	PRP38 pre-mRNA processing factor 38 (yeast) domain containing A isoform 2 [Homo sapiens]	137	41240	4.95	1.88	11
hPRP43	gi 68509926	DEAH (Asp-Glu-Ala-His) box polypeptide 15 [Homo sapiens]	554	98436	5.98	2.21	28
hPRP4-Kinase	gi 89276756	serine/threonine-protein kinase PRP4K [Homo sapiens]	390	135781	0.59	0.19	25
HsKin17	gi 13124883	HsKin17 protein [Homo sapiens]	78	52728	3.10	1.36	6
hSLU7	gi 4249705	step II splicing factor SLU7 [Homo sapiens]	339	79345	0.10	0.10	19
hSmu-1	gi 7023065	unnamed protein product [Homo sapiens]	815	62752	5.67	2.36	44
hSnu23	gi 13385046	zinc finger, matrin type 2 [Mus musculus]	34	28778	5.48	1.95	2
Hsp27	gi 662841	heat shock protein 27 [Homo sapiens]	0	23414	1.02		1
Hsp70	gi 5729877	heat shock 70kDa protein 8 isoform 1 [Homo sapiens]	314	78964	0.25	0.14	15
hSYF1	gi 55770906	XPA binding protein 2 [Homo sapiens]	1201	106931	0.27	0.09	62
hTra-2-alpha	gi 9558733	transformer-2 alpha [Homo sapiens]	201	33724	1.44	0.70	9

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
hTra-2-beta	gi 4377849	transformer-2-beta isoform 3 [Homo sapiens]	245	22734	1.37	1.30	22
KIAA0073	gi 559713	KIAA0073 [Homo sapiens]	630	80918	0.10	0.12	39
KIAA1604	gi 10047283	KIAA1604 protein [Homo sapiens]	232	120523	0.35	0.15	25
LOC124245	gi 31377595	conserved nuclear protein NHN1 [Homo sapiens]	58	117864	0.71	0.42	11
LOC51325	gi 22035565	GC-rich sequence DNA-binding factor candidate isoform 1 [Homo sapiens]	198	115914	2.12	0.42	16
Lsm2	gi 10863977	LSM2 homolog, U6 small nuclear RNA associated [Homo sapiens]	184	12217	5.24	1.28	4
Lsm3	gi 7657315	Lsm3 protein [Homo sapiens]	46	12414	3.45		1
Lsm4	gi 6912486	U6 snRNA-associated Sm-like protein 4 [Homo sapiens]	54	17207	4.78	1.60	8
Lsm6	gi 5901998	Sm protein F [Homo sapiens]	47	10176	5.64	2.12	4
Lsm7	gi 7706423	U6 snRNA-associated Sm-like protein LSm7 [Homo sapiens]	55	13272	6.61	3.62	8
Lsm8	gi 7706425	U6 snRNA-associated Sm-like protein LSm8 [Homo sapiens]	113	10684	3.55	1.53	3
Magoh	gi 4505087	mago-nashi homolog [Homo sapiens]	135	19216	0.15	0.06	10
MFAP1	gi 50726968	microfibrillar-associated protein 1 [Homo sapiens]	302	58411	3.54	1.00	38
MGC13125	gi 14249338	BUD13 homolog [Homo sapiens]	132	78018	1.97	0.71	10
MGC20398 (CCDC16)	gi 74732532	Coiled-coil domain-containing protein 16	160	47290	0.93	0.37	11
MGC23918 (CCDC12)	gi 21389497	coiled-coil domain containing 12 [Homo sapiens]	149	22674	0.53	0.17	6
MORG1	gi 14150114	mitogen-activated protein kinase organizer 1 [Homo sapiens]	72	37126	0.10		1
NKAP	gi 13375676	NF-kappaB activating protein [Homo sapiens]	85	55386	0.22	0.11	6
N-myc and STAT interactor	gi 4758814	N-myc and STAT interactor [Homo sapiens]	43	39988	4.96		1
Npw38BP	gi 7706501	WW domain binding protein 11 [Homo sapiens]	131	77303	7.31	0.72	6
NUFIP1	gi 6912542	nuclear fragile X mental retardation protein interacting protein 1 [Homo sapiens]	42	64004	0.22		1
NY-CO-10	gi 64276486	serologically defined colon cancer antigen 10 [Homo sapiens]	122	61734	0.63	0.22	12
NY-REN-37	gi 10433149	unnamed protein product [Homo sapiens]	56	51438	1.16		1
p68(DDX5)	gi 226021	growth regulated nuclear 68 protein	114	72338	1.52	0.55	8

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
p72/DDX17	gi 119580652	DEAD (Asp-Glu-Ala-Asp) box polypeptide 17, isoform CRA_h [Homo sapiens]	74	64210	1.07		1
PABP	gi 119612222	poly(A) binding protein, cytoplasmic 1, isoform CRA_c [Homo sapiens]	192	52768	1.70	0.48	8
PABPC4	gi 4504715	poly A binding protein, cytoplasmic 4 [Homo sapiens]	46	77931	1.52	0.35	4
PABPN1	gi 4758876	poly(A) binding protein, nuclear 1 [Homo sapiens]	77	34406	1.21	0.09	3
PCBP1	gi 5453854	poly(rC) binding protein 1 [Homo sapiens]	56	40077	5.02	1.56	5
PCBP2	gi 14141166	poly(rC)-binding protein 2 isoform b [Homo sapiens]	352	41113	2.58	0.29	4
Pinin	gi 3021392	nuclear protein SDK3 [Homo sapiens]	35	89790	2.11	0.58	5
PPIG	gi 42560244	peptidyl-prolyl isomerase G (cyclophilin G) [Homo sapiens]	109	104685	0.20	0.09	6
PPIL1	gi 7706339	peptidylprolyl isomerase-like 1 [Homo sapiens]	116	19902	0.42	0.18	10
PPIL2	gi 7657473	peptidylprolyl isomerase-like 2 isoform a [Homo sapiens]	331	66360	1.33	0.88	19
PPIL3	gi 14043400	Peptidylprolyl isomerase (cyclophilin)-like 3 [Homo sapiens]	134	20214	0.17	0.14	8
PRCC	gi 14714625	Papillary renal cell carcinoma (translocation-associated) [Homo sapiens]	54	57721	5.20	0.41	2
PRKRIP1	gi 13375901	PRKR interacting protein 1 (IL11 inducible) [Homo sapiens]	201	24875	0.13	0.09	7
PRL1	gi 4505895	pleiotropic regulator 1 (PRL1 homolog, Arabidopsis) [Homo sapiens]	603	62478	0.69	0.22	33
PUF60	gi 1809248	siah binding protein 1 [Homo sapiens]	48	62853	2.34		1
RALY	gi 8051631	RNA binding protein (autoantigenic, hnRNP-associated with lethal yellow) long isoform [Homo sapiens]	90	35804	0.80	0.20	4
RBBP6	gi 33620716	retinoblastoma-binding protein 6 isoform 2 [Homo sapiens]	33	229462	1.38		1
RBM10	gi 34785044	RBM10 protein [Homo sapiens]	48	64274	6.49		1
RBM14	gi 5454064	RNA binding motif protein 14 [Homo sapiens]	23	72469	1.26		1
RBM15(OTT)	gi 10433990	unnamed protein product [Homo sapiens]	27	108615	1.41	0.19	2
RBM5/LUCA15	gi 1244404	putative tumor suppressor	56	100212	6.73	0.86	2

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
RBM7	gi 7023641	unnamed protein product [Homo sapiens]	0	32759	1.14		1
RED	gi 119582432	IK cytokine, down-regulator of HLA II, isoform CRA_a [Homo sapiens]	438	75361	4.64	2.17	23
RNF113A	gi 5902158	ring finger protein 113A [Homo sapiens]	33	43684	0.43	0.15	2
RNPC2(CAPER)	gi 4757926	RNA binding motif protein 39 isoform b [Homo sapiens]	350	64182	4.08	0.89	17
RNPS1	gi 3253165	SR protein [Homo sapiens]	49	38412	0.63	0.17	2
S164	gi 55741709	RNA binding motif protein 25 [Homo sapiens]	50	113081	0.73	0.35	3
SAFB-like	gi 62825862	SLTM protein [Homo sapiens]	55	86766	4.10	0.84	5
SAP18	gi 5032067	Sin3A-associated protein, 18kDa [Homo sapiens]	99	19469	1.37	0.21	6
SF3a120	gi 5032087	splicing factor 3a, subunit 1, 120kDa isoform 1 [Homo sapiens]	444	97523	2.66	0.58	26
SF3a60	gi 551450	splicing factor SF3a60 [Homo sapiens]	602	65068	4.91	1.28	24
SF3a66	gi 409219	spliceosomal protein	121	52572	3.94	0.71	7
SF3b125	gi 3435312	RNA helicase-related protein [Homo sapiens]	126	84712	10.06	4.64	6
SF3b130	gi 6006515	spliceosomal protein SAP 130 [Homo sapiens]	1144	144018	5.01	1.85	98
SF3b145	gi 14043240	SF3B2 protein [Homo sapiens]	403	83704	4.63	1.53	21
SF3b14a/p14	gi 7706326	splicing factor 3B, 14 kDa subunit [Homo sapiens]	41	16541	4.25	1.31	11
SF3b14b	gi 14249398	PHD-finger 5A [Homo sapiens]	84	14868	3.75	1.04	10
SF3b155	gi 54112117	splicing factor 3b, subunit 1 isoform 1 [Homo sapiens]	474	158585	3.98	2.36	68
SF3b49	gi 5032069	splicing factor 3b, subunit 4 [Homo sapiens]	105	46421	5.67	1.26	2
SFRS12	gi 28703790	Similar to expressed sequence AI450757 [Homo sapiens]	35	46956	6.41		1
SFRS17A	gi 187242	550 amino acids MW=61kDa, glycosylated=75 kDa; expressed on endothelium, activated lymphocytes and syncytiotrophoblast, contains leucine zipper and basic region	58	71752	1.36	0.20	4
SKIP	gi 6912676	homologous to myc; 721P SKI-interacting protein [Homo sapiens]	651	69284	0.58	0.17	58
SKIV2L2	gi 6633995	KIAA0052 protein [Homo sapiens]	206	131978	1.13	0.20	12
SmB	gi 190247	snRNP polypeptide B	192	31604	1.69	0.28	13

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
SmD1	gi 5902102	small nuclear ribonucleoprotein D1 polypeptide 16kDa [Homo sapiens]	205	14858	1.39	0.26	6
SmD2	gi 4759158	small nuclear ribonucleoprotein polypeptide D2 [Homo sapiens]	361	16204	1.62	0.66	35
SmD3	gi 4759160	small nuclear ribonucleoprotein polypeptide D3 [Homo sapiens]	249	15584	1.40	0.32	32
SmE	gi 4507129	small nuclear ribonucleoprotein polypeptide E [Homo sapiens]	292	11851	1.20	0.17	11
SmF	gi 4507131	small nuclear ribonucleoprotein polypeptide F [Homo sapiens]	114	10629	2.91	0.93	2
SmG	gi 4507133	small nuclear ribonucleoprotein polypeptide G [Homo sapiens]	122	9545	1.52	0.44	4
SNIP	gi 21314720	Smad nuclear interacting protein [Homo sapiens]	58	49687	0.80	0.11	10
SON3	gi 17046381	SON DNA binding protein isoform E [Homo sapiens]	144	240773	1.13	0.22	5
Spen	gi 14790190	spen homolog, transcriptional regulator [Homo sapiens]	251	440442	4.77	1.53	16
SPF27	gi 5031653	breast carcinoma amplified sequence 2 [Homo sapiens]	305	28080	0.53	0.13	18
SPF30	gi 5032113	survival motor neuron domain containing 1 [Homo sapiens]	37	31300	30.16		1
SPF45	gi 14249678	RNA binding motif protein 17 [Homo sapiens]	64	50450	5.38	1.25	5
SR140	gi 2224605	KIAA0332 [Homo sapiens]	73	134163	3.04	0.17	3
SRm160	gi 23274133	Serine/arginine repetitive matrix 1 [Homo sapiens]	89	117116	1.12	0.16	5
SRm300	gi 5821153	RNA binding protein [Homo sapiens]	133	318973	0.31	0.13	7
SRp20	gi 2125864	Srp20 [Mus musculus]	101	14955	1.65	0.44	8
SRp30c	gi 4506903	splicing factor, arginine/serine-rich 9 [Homo sapiens]	186	26915	0.99	0.20	16
SRp40	gi 55640963	PREDICTED: splicing factor, arginine/serine-rich 5 isoform 6 [Pan troglodytes]	70	33844	0.85	0.07	4
SRp46	gi 14141216	SRp46 splicing factor [Homo sapiens]	65	33004	1.85	1.08	2
SRp55	gi 62087532	arginine/serine-rich splicing factor 6 variant [Homo sapiens]	87	33759	1.35	0.23	17

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
TARDBP	gi 6678271	TAR DNA binding protein [Homo sapiens]	0	48013	0.99	0.33	3
TCERG1	gi 21327715	transcription elongation regulator 1 isoform 1 [Homo sapiens]	81	138273	2.10	0.26	3
TCP1	gi 1800303	HIV-1 Nef interacting protein [Homo sapiens]	0	50858	0.36	0.13	3
TFIP11	gi 8393259	tuftelin interacting protein 11 [Homo sapiens]	150	106591	6.50	1.91	12
TFIP11	gi 8393259	tuftelin interacting protein 11 [Homo sapiens]	257	106591	0.66	0.15	16
THOC1	gi 37999906	THO complex subunit 1 (Tho1) (Nuclear matrix protein p84)	136	83277	1.13	0.42	4
THOC2	gi 125656165	THO complex 2 isoform 1 [Homo sapiens]	148	211268	0.62	0.17	13
Aly/REF (THOC4)	gi 55770864	THO complex 4 [Homo sapiens]	183	28601	1.11	0.27	4
KIAA0983 (THOC5)	gi 40789009	KIAA0983 protein [Homo sapiens]	59	88032	1.44	0.46	5
WDR58 (THOC6)	gi 22761350	unnamed protein product [Homo sapiens]	55	35141	0.51	0.25	2
THRAP3	gi 114555524	PREDICTED: thyroid hormone receptor associated protein 3 isoform 1 [Pan troglodytes]	207	123201	2.04	0.67	30
TOE1	gi 10436256	unnamed protein product [Homo sapiens]	35	61544	0.84	0.37	3
TPR14 (TTC14)	gi 33457330	tetratricopeptide repeat domain 14 isoform a [Homo sapiens]	99	99198	0.22	0.14	8
tubulin alpha-2	gi 34740335	tubulin, alpha 2 [Mus musculus]	614	53554	0.73	0.11	2
tubulin beta	gi 18088719	Tubulin, beta [Homo sapiens]	381	52313	0.93	0.23	2
tubulin beta-2	gi 5174735	tubulin, beta, 2 [Homo sapiens]	255	52473	1.15		1
U1-70K	gi 36100	unnamed protein product [Homo sapiens]	83	73878	3.24	1.42	12
U2-A'	gi 50593002	small nuclear ribonucleoprotein polypeptide A~ [Homo sapiens]	227	31660	0.48	0.18	23
U2AF35	gi 5803207	U2 small nuclear RNA auxillary factor 1 isoform a [Homo sapiens]	49	29853	3.63	0.80	2
U2AF65	gi 228543	splicing factor U2AF:SUBUNIT=large	107	57879	5.59	2.69	8
U2-B''	gi 4507123	small nuclear ribonucleoprotein polypeptide B~ [Homo sapiens]	194	29217	0.53	0.25	14
U4/U6.U5-110K	gi 10863889	squamous cell carcinoma antigen recognized by T cells 1 [Homo sapiens]	629	101434	4.50	1.96	23

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
U4/U6.U5-65K	gi 13926071	U4/U6.U5 tri-snRNP-associated 65 kDa protein [Homo sapiens]	411	72324	1.78	0.36	21
U4/U6-15.5K	gi 4826860	NHP2 non-histone chromosome protein 2-like 1 [Homo sapiens]	62	15934	9.12	2.96	2
U4/U6-20K	gi 5454154	peptidylprolyl isomerase H [Homo sapiens]	110	21299	2.50	0.33	3
U4/U6-60K	gi 2708305	U4/U6 small nuclear ribonucleoprotein hPrp4 [Homo sapiens]	623	64027	8.18	2.98	26
U4/U6-61K	gi 114678987	PREDICTED: pre-mRNA processing factor 31 homolog isoform 2 [Pan troglodytes]	419	61514	7.43	3.30	26
U4/U6-90K	gi 4758556	PRP3 pre-mRNA processing factor 3 homolog [Homo sapiens]	280	89147	8.27	3.88	37
U5-100K	gi 41327771	DEAD (Asp-Glu-Ala-Asp) box polypeptide 23 [Homo sapiens]	720	108337	1.57	0.38	48
U5-102K	gi 4103604	putative mitochondrial outer membrane protein import receptor [Homo sapiens]	666	118065	2.77	0.78	55
U5-116K	gi 40788951	KIAA0031 [Homo sapiens]	511	119251	8.38	3.17	22
U5-116K	gi 40788951	KIAA0031 [Homo sapiens]	1294	119251	0.96	0.28	52
U5-15K	gi 5729802	thioredoxin-like 4A [Homo sapiens]	50	18452	7.03	1.46	3
U5-200K	gi 45861372	200 kDa U5 snRNP-specific spliceosomal protein [Homo sapiens]	3198	265598	1.05	0.24	211
U5-220K	gi 73967172	PREDICTED: similar to Pre-mRNA processing splicing factor 8 (Splicing factor Prp8) (PRP8 homolog) (220 kDa U5 snRNP-specific protein) (p220) isoform 2 [Canis familiaris]	2542	296532	1.18	0.20	154
U5-40K	gi 109000921	PREDICTED: WD repeat domain 57 (U5 snRNP specific) [Macaca mulatta]	527	42536	1.03	0.18	20
U5-52K	gi 5174409	CD2 antigen (cytoplasmic tail) binding protein 2 [Homo sapiens]	52	40021	1.06	0.57	3
UBL5	gi 13236510	ubiquitin-like 5 [Homo sapiens]	0	10074	17.04		1
USP42	gi 79750944	ubiquitin specific protease 42 [Homo sapiens]	86	158266	0.59	0.35	7
Y14	gi 4826972	RNA binding motif protein 8A [Homo sapiens]	163	21796	0.40	0.14	7
YB-1	gi 181486	DNA-binding protein B	147	42404	1.55	0.69	9
YT521	gi 16551831	unnamed protein product [Homo sapiens]	26	82609	0.98	0.34	2
ZC3H11A	gi 29387160	ZC3H11A protein [Homo sapiens]	32	83552	2.64	1.36	2

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
ZCCHC10	gi 8923106	zinc finger, CCHC domain containing 10 [Homo sapiens]	60	21419	0.66	0.71	2
ZCCHC19	gi 51243065	zinc finger CCHC-type and RNA binding motif 1 [Homo sapiens]	90	29083	0.48	0.21	4
ZCCHC8	gi 7018505	hypothetical protein [Homo sapiens]	49	55697	1.15	0.15	2

Table A.6: Proteins identified in the second replicate of the relative quantification of spliceosomal B and C complexes by iTRAQ. The accession number, the protein database description, the obtained protein score (Mascot), the protein mass, the determined protein ratio (B/C), the standard deviation (StDev), and the number of peptides used for quantification (#) is given for all proteins found. Proteins used for normalization are highlighted in red.

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
9G8	gi 72534660	splicing factor, arginine/serine-rich 7 [Homo sapiens]	146	28687	2.58	0.38	9
Abstrakt	gi 21071032	DEAD-box protein abstrakt [Homo sapiens]	382	77694	0.17	0.14	90
Acinus	gi 109082928	PREDICTED: apoptotic chromatin condensation inducer 1 isoform 2 [Macaca mulatta]	164	77764	1.84	0.38	17
Acinus	gi 7513059	hypothetical protein KIAA0670 - human (fragment)	99	158783	1.76	0.37	8
ALB	gi 28592	serum albumin [Homo sapiens]	79	79864	7.91	0.97	3
alpha tubulin	gi 340021	alpha-tubulin	363	53554	1.14	0.20	26
Aquarius	gi 38788372	aquarius [Homo sapiens]	1599	183588	0.39	0.13	186
ASR2B	gi 19879862	arsenite-resistant protein ASR2 [Homo sapiens]	357	99344	1.55	0.20	32
BAG2	gi 4757834	BCL2-associated athanogene 2 [Homo sapiens]	0	26057	0.50		1
BCLAF1	gi 7582386	Bcl-2-associated transcription factor short form [Homo sapiens]	47	112929	2.23	0.31	6
BCR	gi 116666695	Chain A, Crystal Structure Of Full-Length 3~-Exonuclease	0	45012	1.38		1
beta-tubulin	gi 1297274	beta-tubulin [Homo sapiens]	90	53303	0.67	0.10	5
BRIX	gi 55770900	BRIX [Homo sapiens]	0	47945	1.39	0.31	2
c10orf4/FRA10AC1	gi 24432067	FRA10AC1 protein [Homo sapiens]	49	44908	0.11	0.05	6
C1orf55	gi 40255125	hypothetical protein LOC163859 [Homo sapiens]	178	54043	0.13	0.06	14
c9orf78	gi 6808233	hypothetical protein [Homo sapiens]	40	32417	0.26	0.01	2
Cactin	gi 91208260	Uncharacterized protein C19orf29 (Renal carcinoma antigen NY-REN-24)	571	87199	0.17	0.08	35
Cas-Br-M	gi 13376204	Cas-Br-M (murine) ecotropic retroviral transforming sequence-like 1 [Homo sapiens]	66	57712	1.94	0.08	2
CBP20	gi 19923387	nuclear cap binding protein subunit 2, 20kDa isoform 1 [Homo sapiens]	123	20001	0.93	0.16	10
CBP80	gi 4505343	nuclear cap binding protein subunit 1, 80kDa [Homo sapiens]	580	99715	1.01	0.12	32

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
CCAP2	gi 6841518	HSPC148 [Homo sapiens]	78	30099	0.15	0.08	3
CCAR1	gi 27497118	death inducer with SAP domain DIS [Homo sapiens]	38	147586	5.02	1.24	2
CCDC55	gi 14149807	coiled-coil domain containing 55 isoform 1 [Homo sapiens]	0	76765	1.17		1
CCNK	gi 8980825	cyclin K [Homo sapiens]	46	44722	2.33	1.10	2
CDC5L	gi 11067747	CDC5-like [Homo sapiens]	716	103618	1.42	0.39	53
CDC5L	gi 11067747	CDC5-like [Homo sapiens]	360	103618	0.19	0.12	30
CDK10(PISSLRE)	gi 6226784	Cell division protein kinase 10 (Serine/threonine-protein kinase PISSLRE)	69	45362	0.24	0.10	4
CGI-79	gi 4929627	CGI-79 protein [Homo sapiens]	44	46601	1.36	0.10	2
CHERP	gi 2058691	ERPROT 213-21 [Homo sapiens]	73	105842	4.27	1.06	5
CRNKL1	gi 50949465	hypothetical protein [Homo sapiens]	757	96523	0.32	0.07	82
CTNNBL1	gi 18644734	beta catenin-like 1 [Homo sapiens]	120	70832	2.29	0.34	11
CUG	gi 5729794	CUG triplet repeat, RNA-binding protein 1 isoform 1 [Homo sapiens]	76	55437	3.42	0.53	7
CXorf56	gi 11545813	hypothetical protein LOC63932 [Homo sapiens]	167	29821	0.24	0.13	11
Cyclin M	gi 88989669	PREDICTED: similar to CG31232-PA, isoform A [Homo sapiens]	39	27359	0.18	0.03	2
CypE	gi 45439318	peptidylprolyl isomerase E isoform 3 [Homo sapiens]	71	29821	0.58	0.14	5
DDX34	gi 38158022	DEAH (Asp-Glu-Ala-His) box polypeptide 34 [Homo sapiens]	460	135588	0.26	0.17	20
DDX35	gi 20544129	DEAH (Asp-Glu-Ala-His) box polypeptide 35 [Homo sapiens]	505	84842	0.17	0.07	22
DDX9	gi 307383	RNA helicase A	69	152784	1.43	0.14	3
DGCR14	gi 12804313	Similar to expressed sequence 2 embryonic lethal [Homo sapiens]	105	56774	0.29	0.16	8
E1B-AP5	gi 3319956	E1B-55kDa-associated protein [Homo sapiens]	55	101542	1.62	0.10	4
Ecm29	gi 2224677	KIAA0368 [Homo sapiens]	22	175362	1.77	0.17	2
eEF1A1	gi 31092	unnamed protein product [Homo sapiens]	35	57288	1.07	0.51	2
ELAV	gi 119589356	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R), isoform CRA_b [Homo sapiens]	305	53703	5.65	0.82	18

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
eIF4A3	gi 496902	translation initiation factor [Homo sapiens]	166	50492	0.55	0.06	9
ELG	gi 8923771	ELG protein [Homo sapiens]	0	43508	1.08		1
EXOSC4	gi 9506689	exosome component 4 [Homo sapiens]	63	27317	1.67		1
EXOSC5	gi 14043511	Exosome component 5 [Homo sapiens]	30	27406	0.86		1
EXOSC7	gi 473949	KIAA0116 [Homo sapiens]	0	34928	2.51		1
Fam50A(XAP-5)	gi 4758220	XAP-5 protein [Homo sapiens]	44	46701	0.29	0.12	5
FAM50B	gi 6912326	family with sequence similarity 50, member B [Homo sapiens]	35	42766	0.36	0.05	3
FBP11	gi 5360087	NY-REN-6 antigen [Homo sapiens]	88	59273	2.70	0.50	4
FNBP4	gi 6808095	hypothetical protein [Homo sapiens]	86	119110	2.71	0.29	4
FRG1	gi 4758404	FSHD region gene 1 [Homo sapiens]	100	35148	0.58	0.17	6
FUSE3	gi 100816392	far upstream element (FUSE) binding protein 3 [Homo sapiens]	214	65481	3.78	0.65	17
FUSIP1/SRp38	gi 5730079	FUS interacting protein (serine-arginine rich) 1 isoform 1 [Homo sapiens]	54	23166	1.44	0.32	7
G10	gi 32171175	G10 protein [Homo sapiens]	85	19898	0.88	0.02	4
GCIP p29	gi 7661636	SYF2 homolog, RNA splicing factor isoform 1 [Homo sapiens]	104	32930	0.18	0.07	7
hECM2 (RBM22)	gi 8922328	RNA binding motif protein 22 [Homo sapiens]	380	51786	0.53	0.11	39
hISY1	gi 6330157	KIAA1160 protein [Homo sapiens]	261	43813	0.61	0.10	17
hnRNP A1	gi 36102	unnamed protein product [Homo sapiens]	56	37105	5.21	0.13	2
hnRNP A2/B1	gi 4504447	heterogeneous nuclear ribonucleoprotein A2/B1 isoform A2 [Homo sapiens]	77	38480	3.60	0.18	4
hnRNP A3	gi 34740329	heterogeneous nuclear ribonucleoprotein A3 [Homo sapiens]	0	42925	5.14		1
hnRNP C1/C2	gi 306875	C protein Chain A, Nmr Structure Of The Third Qrrm Domain Of Human	193	36316	1.97	0.30	16
hnRNP F	gi 112491343	Hnrnp F	47	15509	2.94		1
hnRNP G	gi 3256007	hnRNP G protein [Homo sapiens]	80	44396	2.21	0.35	6
hnRNP H1	gi 5031753	heterogeneous nuclear ribonucleoprotein H1 [Homo sapiens]	123	51878	1.84	0.27	13

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
hnRNP K	gi 55958547	heterogeneous nuclear ribonucleoprotein K [Homo sapiens]	31	44558	4.01	0.73	2
hnRNP M	gi 187281	M4 protein	33	83739	1.37	0.38	4
hnRNP Q2	gi 15809588	hnRNP Q2 [Homo sapiens]	41	72196	0.96	0.44	2
hnRNP R	gi 5031755	heterogeneous nuclear ribonucleoprotein R [Homo sapiens]	71	77614	1.09		1
hPRP16	gi 17999539	DEAH (Asp-Glu-Ala-His) box polypeptide 38 [Homo sapiens]	84	152201	0.43	0.11	12
hPRP17	gi 7706657	cell division cycle 40 homolog [Homo sapiens]	535	73381	0.45	0.15	57
hPRP18	gi 4506123	PRP18 pre-mRNA processing factor 18 homolog [Homo sapiens]	55	45161	0.39	0.18	4
hPRP19	gi 7657381	PRP19/PSO4 pre-mRNA processing factor 19 homolog [Homo sapiens]	226	59981	0.57	0.12	62
hPRP2	gi 14250712	DEAH (Asp-Glu-Ala-His) box polypeptide 16 [Homo sapiens]	266	129236	2.10	0.29	20
hPRP22	gi 4826690	DEAH (Asp-Glu-Ala-His) box polypeptide 8 [Homo sapiens]	863	153030	0.29	0.15	89
hPRP38	gi 24762236	PRP38 pre-mRNA processing factor 38 (yeast) domain containing A isoform 2 [Homo sapiens]	107	41240	4.04	0.85	7
hPRP43	gi 68509926	DEAH (Asp-Glu-Ala-His) box polypeptide 15 [Homo sapiens]	566	98436	2.59	0.35	37
HsKin17	gi 13124883	HsKin17 protein [Homo sapiens]	76	52728	2.99	0.57	8
hSLU7	gi 4249705	step II splicing factor SLU7 [Homo sapiens]	286	79345	0.20	0.09	38
hSnu23	gi 13385046	zinc finger, matrin type 2 [Mus musculus]	0	28778	12.22	3.17	4
Hsp27	gi 4504517	heat shock 27kDa protein 1 [Homo sapiens]	54	23967	0.47	0.02	2
Hsp70	gi 5729877	heat shock 70kDa protein 8 isoform 1 [Homo sapiens]	156	78964	1.19	0.22	10
Hsp70	gi 5729877	heat shock 70kDa protein 8 isoform 1 [Homo sapiens]	314	78964	0.24	0.11	18
hSYF1	gi 55770906	XPA binding protein 2 [Homo sapiens]	1131	106931	0.38	0.08	65
hTra-2-alpha	gi 9558733	transformer-2 alpha [Homo sapiens]	59	33724	2.80	0.47	2
hTra-2-beta	gi 4377849	transformer-2-beta isoform 3 [Homo sapiens]	179	22734	1.22	0.37	9
KIAA0073	gi 559713	KIAA0073 [Homo sapiens]	588	80918	0.20	0.09	53

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
KIAA0983 (THOC5)	gi 40789009	KIAA0983 protein [Homo sapiens]	49	88032	1.35	0.26	5
KIAA1604	gi 10047283	KIAA1604 protein [Homo sapiens]	310	120523	0.46	0.23	27
LOC124245	gi 31377595	conserved nuclear protein NHN1 [Homo sapiens]	134	117864	1.68	0.36	27
LOC150383	gi 49533621	hypothetical protein LOC150383 [Homo sapiens]	34	15480	0.07		1
LOC51325	gi 22035565	GC-rich sequence DNA-binding factor candidate isoform 1 [Homo sapiens]	284	115914	1.36	0.26	24
LOC540543	gi 84000355	hypothetical protein LOC540543 [Bos taurus]	43	38096	2.45	1.21	2
LRC	gi 24308289	leukocyte receptor cluster (LRC) member 1 [Homo sapiens]	0	33537	0.37		1
LSm2	gi 10863977	LSM2 homolog, U6 small nuclear RNA associated [Homo sapiens]	129	12217	4.47	0.39	8
LSm3	gi 7657315	Lsm3 protein [Homo sapiens]	38	12414	3.21		1
Lsm4	gi 6912486	U6 snRNA-associated Sm-like protein 4 [Homo sapiens]	43	17207	5.52	2.31	2
LSm6	gi 5901998	Sm protein F [Homo sapiens]	56	10176	6.16	1.02	3
LSm7	gi 7706423	U6 snRNA-associated Sm-like protein LSm7 [Homo sapiens]	0	13272	3.70	0.08	2
LSm8	gi 7706425	U6 snRNA-associated Sm-like protein LSm8 [Homo sapiens]	57	10684	4.14	0.70	2
LUC7A	gi 8922297	LUC7-like isoform a [Homo sapiens]	34	42306	4.20	0.88	2
Matrin3	gi 57162250	novel protein [Homo sapiens]	0	16315	1.09		1
Magoh	gi 4505087	mago-nashi homolog [Homo sapiens]	65	19216	0.20	0.05	5
MFAP1	gi 50726968	microfibrillar-associated protein 1 [Homo sapiens]	216	58411	4.77	0.56	11
MGC13125	gi 14249338	BUD13 homolog [Homo sapiens]	121	78018	0.94	0.38	10
MGC20398 (CCDC16)	gi 74732532	Coiled-coil domain-containing protein 16	333	47290	1.38	0.25	22
MGC23918 (CCDC12)	gi 21389497	coiled-coil domain containing 12 [Homo sapiens]	92	22674	0.51	0.05	4
MORG1	gi 14150114	mitogen-activated protein kinase organizer 1 [Homo sapiens]	120	37126	0.69	0.09	3
NCL	gi 189306	nucleolin	0	89313	1.30		1
NF45	gi 532313	NF45 protein	63	47880	1.94	0.32	4

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
N-myc and STAT inetractor	gi 4758814	N-myc and STAT interactor [Homo sapiens]	43	39988	4.03	1.19	4
NOSIP	gi 7705716	eNOS interacting protein [Homo sapiens]	79	37167	0.19	0.09	3
Npw38	gi 5031957	polyglutamine binding protein 1 [Homo sapiens]	69	33814	7.68	1.60	3
Npw38BP	gi 7706501	WW domain binding protein 11 [Homo sapiens]	236	77303	8.77	2.32	17
NUFIP1	gi 6912542	nuclear fragile X mental retardation protein interacting protein 1 [Homo sapiens]	48	64004	0.33	0.02	2
Numa 1	gi 35119	NuMA protein [Homo sapiens]	30	260907	3.11		1
NY-CO-10	gi 3170184	antigen NY-CO-10 [Homo sapiens]	27	49863	0.68	0.05	2
p68(DDX5)	gi 226021	growth regulated nuclear 68 protein	151	72338	2.30	0.45	15
p72/DDX17	gi 119580652	DEAD (Asp-Glu-Ala-Asp) box polypeptide 17, isoform CRA_h [Homo sapiens]	49	64210	2.30	0.73	2
PABP	gi 35570	unnamed protein product [Homo sapiens]	55	76372	4.72	2.46	4
PABP	gi 35570	unnamed protein product [Homo sapiens]	42	76372	0.83	0.25	2
PABPC4	gi 4504715	poly A binding protein, cytoplasmic 4 [Homo sapiens]	0	77931	0.99		1
PABPN1	gi 4758876	poly(A) binding protein, nuclear 1 [Homo sapiens]	33	34406	0.79	0.00	2
PCBP1	gi 5453854	poly(rC) binding protein 1 [Homo sapiens]	133	40077	4.66	1.45	10
PCBP2	gi 14141166	poly(rC)-binding protein 2 isoform b [Homo sapiens]	121	41113	3.71	1.04	16
Pinin	gi 3021392	nuclear protein SDK3 [Homo sapiens]	23	89790	1.89	0.06	3
plakophilin	gi 6005830	plakophilin 3 [Homo sapiens]	0	91432	2.49		1
PPIL1	gi 7706339	peptidylprolyl isomerase-like 1 [Homo sapiens]	68	19902	0.45	0.08	8
PPIL2	gi 7657473	peptidylprolyl isomerase-like 2 isoform a [Homo sapiens]	477	66360	2.96	0.74	32
PPIL3b	gi 14043400	Peptidylprolyl isomerase (cyclophilin)-like 3 [Homo sapiens]	58	20214	0.31	0.19	3
PPIL4	gi 20911035	peptidylprolyl isomerase-like 4 [Homo sapiens]	42	65817	2.91		1
PRCC	gi 14714625	Papillary renal cell carcinoma (translocation-associated) [Homo sapiens]	98	57721	2.94	0.18	4

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
PRKRIP1	gi 13375901	PRKR interacting protein 1 (IL11 inducible) [Homo sapiens]	36	24875	0.11	0.06	4
PRL1	gi 4505895	pleiotropic regulator 1 (PRL1 homolog, Arabidopsis) [Homo sapiens]	26	62478	1.06	0.22	6
hPRP4-Kinase	gi 23271009	PRP4 pre-mRNA processing factor 4 homolog B (yeast) [Homo sapiens]	268	135785	0.37	0.07	18
PUF60	gi 6176532	poly-U binding splicing factor PUF60 [Homo sapiens]	331	64428	4.26	1.35	19
Q9BRR8	gi 21361684	G patch domain containing 1 [Homo sapiens]	141	116769	0.19	0.07	9
RALY	gi 3334899	e1b-55kDa-associated protein [Homo sapiens]	45	33780	2.48	0.68	3
RBM10	gi 1469167	KIAA0122 [Homo sapiens]	98	120110	5.09	1.05	4
RBM15(OTT)	gi 10433990	unnamed protein product [Homo sapiens]	108	108615	1.91	0.26	9
RBM18	gi 14916461	RNA binding motif protein 18 [Homo sapiens]	62	25042	0.42		1
RBM5/LUCA	gi 1244404	putative tumor suppressor	28	100212	4.99	0.36	2
RED	gi 119582433	IK cytokine, down-regulator of HLA II, isoform CRA_b [Homo sapiens]	280	75577	7.21	2.03	20
RNF 113A	gi 5902158	ring finger protein 113A [Homo sapiens]	50	43684	1.20	0.13	3
RNPC2(CAPER)	gi 4757926	RNA binding motif protein 39 isoform b [Homo sapiens]	295	64182	4.07	0.73	13
RNPS1	gi 3253165	SR protein [Homo sapiens]	41	38412	2.09		1
RSRC1	gi 14714462	RSRC1 protein [Homo sapiens]	75	35833	1.69	0.54	4
S100 A9	gi 4506773	S100 calcium-binding protein A9 [Homo sapiens]	0	15009	89.37		1
S164	gi 55741709	RNA binding motif protein 25 [Homo sapiens]	47	113081	3.22	0.48	4
SAP18	gi 5032067	Sin3A-associated protein, 18kDa [Homo sapiens]	33	19469	1.19	0.17	6
SEMG I	gi 487420	SEMG I	133	58184	14.26	5.79	4
SEMG II	gi 4506885	semenogelin II precursor [Homo sapiens]	48	73999	28.49		1
SF2/ASF	gi 179074	alternative splicing factor 3a,	138	33984	2.46	0.37	8
SF3a120	gi 5032087	subunit 1, 120kDa isoform 1 [Homo sapiens]	465	97523	2.64	0.44	29

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
SF3a60	gi 551450	splicing factor SF3a60 [Homo sapiens]	277	65068	6.68	0.65	14
SF3a66	gi 409219	spliceosomal protein	162	52572	8.09	0.43	6
SF3b125	gi 3435312	RNA helicase-related protein [Homo sapiens]	83	84712	6.19	2.07	6
SF3b130	gi 6006515	spliceosomal protein SAP 130 [Homo sapiens]	294	144018	7.46	1.45	20
SF3b130	gi 54112121	splicing factor 3b, subunit 3 [Homo sapiens]	498	144003	3.52	0.72	49
SF3b145	gi 2498883	Splicing factor 3B subunit 2 (Spliceosome-associated protein 145) (SAP 145) (SF3b150) (Pre-mRNA-splicing factor SF3b 145 kDa subunit)	441	109792	4.13	0.93	34
SF3b14a/p14	gi 7706326	splicing factor 3B, 14 kDa subunit [Homo sapiens]	179	16541	4.37	0.60	14
SF3b14b	gi 14249398	PHD-finger 5A [Homo sapiens]	44	14868	2.00	0.25	2
SF3b155	gi 54112117	splicing factor 3b, subunit 1 isoform 1 [Homo sapiens]	765	158585	3.66	2.15	118
SF3b49	gi 5032069	splicing factor 3b, subunit 4 [Homo sapiens]	63	46421	5.33	2.56	4
SFRS11	gi 4759100	splicing factor, arginine/serine-rich 11 [Homo sapiens]	50	61240	3.21		1
SFRS12	gi 28703790	Similar to expressed sequence AI450757 [Homo sapiens] 550 amino acids MW=61kDa, glycosylated=75 kDa; expressed on endothelium, activated lymphocytes and syncytiotrophoblast, contains leucine zipper and basic region homologous to myc; 721P	95	46956	2.80	0.59	6
SFRS17A	gi 187242	SKI-interacting protein [Homo sapiens]	33	71752	0.41	0.01	2
SKIP	gi 6912676	KIAA0052 protein [Homo sapiens]	268	69284	0.63	0.10	42
SKIV2L2	gi 6633995	snRNP polypeptide B small nuclear ribonucleoprotein D1 polypeptide 16kDa [Homo sapiens]	78	131978	1.15	0.19	8
SmB	gi 190247		84	31604	1.70	0.29	15
SmD1	gi 5902102		111	14858	1.86	0.51	7

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
SmD2	gi 4759158	small nuclear ribonucleoprotein polypeptide D2 [Homo sapiens]	94	16204	1.88	0.30	22
SmD3	gi 4759160	small nuclear ribonucleoprotein polypeptide D3 [Homo sapiens]	118	15584	2.08	0.44	19
SmE	gi 4507129	small nuclear ribonucleoprotein polypeptide E [Homo sapiens]	179	11851	2.03	0.47	7
SmF	gi 4507131	small nuclear ribonucleoprotein polypeptide F [Homo sapiens]	126	10629	1.40	0.37	6
SmG	gi 4507133	small nuclear ribonucleoprotein polypeptide G [Homo sapiens]	145	9545	1.19	0.09	5
hSmu-1	gi 7023065	unnamed protein product [Homo sapiens]	93	62752	4.00	0.65	4
SNIP1	gi 21314720	Smad nuclear interacting protein [Homo sapiens]	26	49687	1.66	0.15	2
SON3 (DBP-5)	gi 17046381	SON DNA binding protein isoform E [Homo sapiens]	44	240773	1.38	0.06	2
Spen	gi 14790190	spen homolog, transcriptional regulator [Homo sapiens]	55	440442	7.31	2.57	4
SPF27	gi 5031653	breast carcinoma amplified sequence 2 [Homo sapiens]	202	28080	0.55	0.12	15
SPF45	gi 14249678	RNA binding motif protein 17 [Homo sapiens]	70	50450	6.22	0.84	6
SR140	gi 2224605	KIAA0332 [Homo sapiens]	55	134163	4.82	0.80	5
SRm300	gi 71891780	KIAA0324 protein [Homo sapiens]	150	323720	1.63	0.85	4
SRp20	gi 2125864	Srp20 [Mus musculus]	40	14955	2.53		1
SRp30c	gi 4506903	splicing factor, arginine/serine-rich 9 [Homo sapiens]	21	26915	0.53		1
SRp40	gi 55640963	PREDICTED: splicing factor, arginine/serine-rich 5 isoform 6 [Pan troglodytes]	40	33844	1.83	0.26	5
SRp55	gi 20127499	arginine/serine-rich splicing factor 6 [Homo sapiens]	87	43546	0.88	0.18	7
SRp75	gi 2914669	SRP0001LB [Homo sapiens]	25	24207	1.16	0.17	2
TARDBP	gi 6678271	TAR DNA binding protein [Homo sapiens]	33	48013	5.26	1.78	2
tat SF1	gi 1667611	Tat-SF1 [Homo sapiens]	30	96255	1.13		1

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
TCERG1	gi 21327715	transcription elongation regulator 1 isoform 1 [Homo sapiens]	27	138273	4.52	0.71	6
TCP1	gi 1800303	HIV-1 Nef interacting protein [Homo sapiens]	31	50858	0.64	0.09	2
TF IIB	gi 8392875	transcription factor IIB [Homo sapiens]	25	25348	3.61		1
TFIP11	gi 8393259	tuftelin interacting protein 11 [Homo sapiens]	341	106591	1.01	0.08	22
THOC2	gi 52486999	THO complex 2 isoform 2 [Homo sapiens]	57	196891	1.54	0.26	7
THOC1	gi 37999906	THO complex subunit 1 (Tho1) (Nuclear matrix protein p84)	74	83277	1.77	0.17	3
THOC7	gi 34783006	THOC7 protein [Homo sapiens]	31	26105	1.81	0.47	2
THRAP3	gi 4827040	thyroid hormone receptor associated protein 3 [Homo sapiens]	86	123229	1.19	0.08	4
TPR14 (TTC14)	gi 37182643	DRDL5813 [Homo sapiens]	130	82362	0.32	0.10	11
tubulin alpha-2	gi 34740335	tubulin, alpha 2 [Mus musculus]	341	53554	1.09	0.22	33
tubulin beta-5	gi 7106439	tubulin, beta 5 [Mus musculus]	372	52313	0.94	0.25	29
U1-70K	gi 36100	unnamed protein product [Homo sapiens]	61	73878	7.08	1.10	7
U1-A	gi 4759156	small nuclear ribonucleoprotein polypeptide A [Homo sapiens]	117	34574	22.35	7.32	5
U2-A'	gi 50593002	small nuclear ribonucleoprotein polypeptide A~ [Homo sapiens]	259	31660	1.39	0.20	24
U2-B''	gi 4507123	small nuclear ribonucleoprotein polypeptide B~~ [Homo sapiens]	122	29217	1.74	0.42	12
U4/U6.U5-110K	gi 10863889	squamous cell carcinoma antigen recognized by T cells 1 [Homo sapiens]	338	101434	3.60	1.09	19
U4/U6.U5-65K	gi 13926071	U4/U6.U5 tri-snRNP-associated 65 kDa protein [Homo sapiens]	353	72324	1.63	0.21	19
U4/U6-15.5K	gi 4826860	NHP2 non-histone chromosome protein 2-like 1 [Homo sapiens]	107	15934	16.96	0.28	2
U4/U6-20K	gi 5454154	peptidylprolyl isomerase H [Homo sapiens]	83	21299	2.55	0.40	4
U4/U6-60K	gi 2708305	U4/U6 small nuclear ribonucleoprotein hPrp4 [Homo sapiens]	364	64027	5.83	0.85	20

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
U4/U6-61K	gi 114678987	PREDICTED: pre-mRNA processing factor 31 homolog isoform 2 [Pan troglodytes]	341	61514	5.50	1.46	29
U4/U6-90K	gi 4758556	PRP3 pre-mRNA processing factor 3 homolog [Homo sapiens]	372	89147	4.37	1.59	39
U5-100K	gi 41327771	DEAD (Asp-Glu-Ala-Asp) box polypeptide 23 [Homo sapiens]	526	108337	1.09	0.16	33
U5-102K	gi 40807485	PRP6 pre-mRNA processing factor 6 homolog [Homo sapiens]	443	117877	1.81	0.54	64
U5-116K	gi 40788951	KIAA0031 [Homo sapiens]	1226	119251	1.09	0.27	73
U5-15K	gi 5729802	thioredoxin-like 4A [Homo sapiens]	51	18452	4.73	0.45	3
U5-200K	gi 45861372	200 kDa U5 snRNP-specific spliceosomal protein [Homo sapiens]	1609	265598	1.06	0.20	168
U5-220K	gi 73967172	PREDICTED: similar to Pre-mRNA processing splicing factor 8 (Splicing factor Prp8) (PRP8 homolog) (220 kDa U5 snRNP-specific protein) (p220) isoform 2 [Canis familiaris]	1581	296532	1.00	0.15	158
U5-40K	gi 109000921	PREDICTED: WD repeat domain 57 (U5 snRNP specific) [Macaca mulatta]	592	42536	1.33	0.18	16
U5-52K	gi 5174409	CD2 antigen (cytoplasmic tail) binding protein 2 [Homo sapiens]	52	40021	3.50	1.17	5
UAP56	gi 2739119	BAT1 [Homo sapiens]	31	36331	4.54	1.00	3
UBL5	gi 13236510	ubiquitin-like 5 [Homo sapiens]	29	10074	10.08	6.92	2
UKp68	gi 40804742	nuclear protein UKp68 isoform 1 [Homo sapiens]	45	91387	1.05	0.23	6
USP36	gi 7023072	unnamed protein product [Homo sapiens]	31	114660	0.30		1
USP42	gi 49616863	ubiquitin specific protease 42 [Homo sapiens]	54	159352	0.57	0.17	5
WDR58 (THOC6)	gi 13111899	THOC6 protein [Homo sapiens]	23	37448	1.50		1
WDR70	gi 8922301	WD repeat domain 70 [Homo sapiens]	82	81246	1.19	0.18	5
Y14	gi 4826972	RNA binding motif protein 8A [Homo sapiens]	55	21796	0.11	0.03	2
YB-1	gi 181486	DNA-binding protein B	61	42404	2.27	0.02	2
ZBT46	gi 30354537	ZBTB46 protein [Homo sapiens]	46	69412	1.14		1

Protein	accession number	Protein database description	Protein score	Protein mass [Da]	B/C	StDev	#
ZCCHC19	gi 51243065	zinc finger CCHC-type and RNA binding motif 1 [Homo sapiens]	32	29083	1.38	0.01	2
ZCCHC8	gi 14042579	unnamed protein product [Homo sapiens]	79	85705	0.94	0.09	6
ZNF	gi 14330434	putative zinc finger protein [Homo sapiens]	0	127231	0.57		1
ZNF432	gi 40788368	KIAA0798 protein [Homo sapiens]	0	89866	1.19		1

Table A.7: Proteins that were found to be enriched in light (L), medium (M) or heavy (H) nuclear extracts, respectively. The protein ratios obtained for comparison of medium to light nuclear extract (M/L) and heavy to light nuclear extract (H/L) are given.

Protein name	Gene name	Protein ratio		enriched in		
		M/L	H/L	L	M	H
ATPase family AAA domain-containing protein 3B	ATAD3B	0.490	0.239	•		
Dermcidin	DCD	0.173	0.081	•		
Disks large homolog 2	DLG2	0.183	0.032	•		
Dystrophin	DMD	0.038	0.026	•		
Desmoglein-1	DSG1	0.126	0.055	•		
Protein FAM124B	FAM124B	0.023	0.012	•		
Filaggrin-2	FLG2	0.079	0.071	•		
Hornerin	HRNR	0.066	0.021	•		
Interferon-induced 17 kDa protein	ISG15	1.424	2.343			•
Junction plakoglobin	JUP	0.574	0.342	•		
Keratinocyte proline-rich protein	KPRP	0.039	0.033	•		
Krev interaction trapped protein 1	KRIT1	0.024	0.048	•		
Lysozyme C	LYZ	0.145	0.194	•		
Mannosidase, alpha, class 1A, member 1, isoform CRA_a	MAN1A1	0.034	7.796	•		•
highly similar to Septin-2	NEDD5	0.190	0.691	•		
highly similar to Septin-9	Ov/Br septin	0.184	0.695	•		
tRNA pseudouridine synthase A	PUS1	1.341	0.304	•		
Protein S100-A9	S100A9	0.222	0.317	•		
Semenogelin-1	SEMG1	0.078	0.079	•		
Septin-7	SEPT7	0.205	0.701	•		

Table A.8: List of ribosomal proteins used for normalization. The protein ratios obtained for comparison of medium to light nuclear extract (M/L) and heavy to light nuclear extract (H/L) are given. The significance for the proteins to be up- or down regulated within the different nuclear extracts is listed.

Ribosomal Protein	Protein Ratio M/L	Significance M/L	Protein Ratio H/L	Significance H/L
RPS7	1.076	0.3603	0.881	0.2092
RPS27	1.076	0.3588	0.845	0.1410
RPS20	1.088	0.3389	0.781	0.0569
RPS14	1.090	0.3352	0.948	0.3668
RPS19	1.102	0.3134	0.815	0.0952
RPS11	1.109	0.3027	0.859	0.1658
RPS3	1.115	0.2933	0.811	0.0898
RPS21	1.130	0.2684	0.825	0.1097
RPS16	1.137	0.2587	0.834	0.1231
RPS4X	1.141	0.2526	0.828	0.1138
RPS17	1.141	0.2517	0.870	0.1862
RPS25	1.142	0.2512	0.849	0.1475
RPS10	1.148	0.2420	0.832	0.1194
RPS23	1.169	0.2133	0.884	0.2161
RPS15A	1.177	0.2028	0.843	0.1376
RPS28	1.183	0.1955	0.926	0.3111
RPS5	1.190	0.1874	0.830	0.1173
RPS3A	1.207	0.1680	0.915	0.2844
RPS12	1.211	0.1635	0.831	0.1176
RPS9	1.264	0.1144	0.898	0.2466
RPL8	1.274	0.1066	0.828	0.1144
RPL18A	1.299	0.0890	0.864	0.1745
RPL38	1.328	0.0717	0.863	0.1732

Table A.9: List of normalization factors calculated for the different assembly studies. For every time point of the protein assembly on different pre-mRNAs a normalization factor was calculated. The normalization factors were applied to the protein ratios obtained.

protein assembly on:	2'0'	5'0'	10'0'	15'0'	20'0'	30'0'
PM5 pre-mRNA	1.023	2.108	1.399	2.395	0.868	1.526
5'ss-deleted PM5 pre-mRNA	1.139	1.484	1.171	1.249	1.114	2.031
BPS-deleted PM5 pre-mRNA	2.244	3.448	1.116	1.753	0.986	2.171

List of abbreviations

%	percent
1D	one dimensional
2D	two dimensional
3'ss	3' splice site
5'ss	5' splice site
AAA	amino acid analysis
ACN	acetonitrile
approx.	Approximately
APS	Ammonium peroxodisulfate
AQUA	absolute quantification
ATP	adenosine 5'-triphosphate
16BAC	benzyltrimethyl-n-hexadecylammonium chloride
BPS	branch point site
BSA	Bovine serum albumin
°C	degree Celsius
<i>C.elegans</i>	<i>Caenorhabditis elegans</i>
CE	collision energy
Ci	Curie
cICAT	Cleavable ICAT
cm	centimeter
cpm	counts per minute
CTP	cytidine 5'-triphosphate
CXP	collision cell exit potential
<i>D.melanogaster</i>	<i>Drosophila melanogaster</i>
Da	Dalton
DDA	data-dependent acquisition
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethylsulfoxide
DNA	deoxyribonucleic acid
DP	declustering potential
dpm	disintegrations per minute
DTE	Dithioerythrol
DTT	Dithiothreitol
e.g.	for example, <i>exempli gratia</i>
EDTA	Ethylenediaminetetraacetic acid
emPAI	Exponentially modified PAI
EP	entrance potential
EPI	Enhanced product ion
ER	enhanced resolution
ESI	electrospray ionization
<i>et al.</i>	and others, <i>et alii</i>
FA	formic acid
FBS	foetal bovine serum
FLEXIQuant	Full-length expressed stable isotope-labeled proteins for quantification
fmol	femtomole
FT	fourier transformation
FT ICR	Fourier transform ion cyclotron resonance
FWHM	full width at half maximum intensity
g	gram / gravitational acceleration
GTP	guanosine 5'-triphosphate
h	Hours
H	heavy SILAC nuclear extract
HEPES	4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid
HIC	hydrophobic interaction chromatography

HILIC	hydrophilic interaction chromatography
HPLC	high pressure liquid chromatography
i.d.	inner diameter
i.e.	that is, <i>id est</i>
ICAT	isotope-coded affinity tag
ICPL	isotope-coded protein label
IEC	ion exchange chromatography
IEF	isoelectric focussing
IPG	immobilized pH gradient
IPI	International Protein Index
IPTG	Isopropyl-b-D-1-thiogalactopyranoside
iTRAQ	isobaric tags for relative and absolute quantification
kDa	Kilodalton
kV	kilovolts
l	liter
L	light SILAC nuclear extract
LB	Luria-Bertani
LC	liquid chromatography
µg	microgram
µM	micromole
µm	Micrometer
M	molar / medium SILAC nuclear extract
<i>m/z</i>	mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionization
MC	missed cleavages
MDHI	2-methoxy-4,5-dihydro-1 <i>H</i> -imidazole
mg	milligram
min	minutes
ml	milliliter
mM	millimolar
mm	Millimeter
mmol	millimole
MOPS	3-(<i>N</i> -morpholino)propanesulfonic acid
MRM	multiple reaction monitoring
MS	mass spectrometry
ms	milliseconds
MS/MS	tandem mass spectrometry
MuD-PIT	multidimensional protein identification technology
MW	molecular weight
NCBI	National Center for Biotechnology Information
NE	nuclear extract
ng	nanogram
NHS	<i>N</i> -hydroxysuccinimide
nl	nanoliter
nm	nanometer
nM	nanomole
no.	number
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PAI	protein abundance index
PBS	Phosphate buffered saline
PC/QMS	pulse-chase/quantitative mass spectrometry
PCI	Phenol/Chloroform/Isoamylalcohol
pmol	picomole
PMSF	Phenylmethylsulfonyl fluoride
PNK	polynucleotide kinase

ppm	parts per million
pre-mRNA	pre-messenger RNA
PSAQ	Protein Standard Absolute Quantification
Q/q	quadrupole
QconCAT	concatenated signature peptides encoded by QconCAT genes
Q-ToF	Quadrupole-time-of-flight
RNA	ribonucleic acid
RNP	ribonucleoprotein
RP	reversed phase
rpm	revolutions per minute
RRM	RNA recognition motif
rRNA	Ribosomal RNA
RT	retention time
S	Svedberg
s	seconds
SA	succinic anhydride
SC	Sequence coverage
SCX	strong cation-exchange
SDS	sodium dodecylsulfate
SDS	Dodecyl sulfate sodium salt
SEC	size exclusion chromatography
SILAC	stable isotope labeling with amino acids in cell culture
snRNA	small-nuclear RNA
SRM	selected reaction monitoring
StDev	standard deviation
TBE	Tris-Borate-EDTA buffer
TEAB	triethylammonium bicarbonate
TEMED	N,N,N',N'-Tetramethylethylenediamid
TFA	trifluoroacetic acid
TIC	total ion current/count
TMTs	tandem mass tags
ToF	time-of-flight
Tris	Tris-(hydroxymethyl) aminoethane
TTP	thymidine-5'-triphosphate
U	units
U snRNA	uridine-rich small-nuclear RNA
UPS	unique peptide sequences
UTP	uridine 5'-triphosphate
UV	ultraviolet
V	volts
vol.	volumes
vs.	versus
w/o	without
XIC	extracted ion chromatogram
Y _n	Polypyrimidine tract

Acknowledgements

I would like to thank many people, who have helped and supported me during the last years.

First of all, I would like to thank Dr. Henning Urlaub for supervision of this work, and my thesis and examination committee - Prof. Dr. Ralf Ficner, Prof. Dr. Ivo Feußner, Prof. Dr. Ulf Diederichsen, Prof. Dr. Kai Tittmann, and Prof. Dr. Dirk Görlich - for support, thesis committee meetings, and review of this work.

I thank Prof. Dr. Reinhard Lührmann for the opportunity to work on these interesting projects, and Sergey Bessonov, Jochen Deckert, Mads Grønborg and Michael Grote for collaboration on the absolute and relative quantification projects – in addition, Sergey Bessonov and Jochen Deckert for introducing me to the field of complex purification.

I am thankful to Dr. Christof Lenz for help, support, and introducing me to the world of MRM!

My very special thanks go to the Bioanalytical Mass Spectrometry Group. Thanks for the great atmosphere, the nice time in and outside the lab, help, support, ... ! Monika, Johanna and Uwe - thanks for excellent (!!) technical assistance and help with diverse problems! I would like to thank He-Hsuan for help in solving different problems and Miro for discussions about data analysis and statistics. I would also like to thank the former members of our group; in particular, Mads for the nice time, fun and support. Last but not least, Henning – for continuous support, never ending enthusiasm, encouragement, discussions, ... , biking, first aid, Summer Schools, conferences!

I am more than grateful to Klaus Hartmuth for reading (almost) every word I wrote during the last year, for discussions, advice and small talk.

Many thanks are addressed to...

...Thomas Conrad for fermentation of SILAC cells and answering all my questions.

...Peter Odenwälder, Norbert Rigo and Thomas Oellerich for help with HeLa cells.

...Paul Woolley and Cindy Will for reading parts of this thesis.

...Juliane Moses for solving all kinds of problems.

...Hossein Kohansal for help in preparing nuclear extracts.

...Gabi Heyne, Peter Kempkes and Irene Öchsner for help and technical assistance.

...Ursula Drössler and Gertrud Nowak for support with daily needs.

...Julia and Ramazan for answering diverse questions, help, support.

...the whole Lührmann department for help with various problems and the nice atmosphere.

...the GGNB office for help in solving all problems.

...the GGNB for a travel grant and a bridging fund.

At last, I would like to thank my family and friends - my parents, who always supported me during the last years, my brother Carsten and my sister Cornelia for the nice meetings, fun, support, encouragement and help; all my friends for continuous help and support, especially Simone and Jennifer for reading this thesis, interest, sympathy, endless discussions – and Simone for the nice time in Göttingen (and Leipzig)!

Curriculum Vitae

Personal data

Name	Carla Schmidt
Date of birth	August 15, 1981
Place of birth	Wolfenbüttel

Education

11/2006 – present	PhD project, Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Göttingen;
11/2006	Diploma <i>“Charakterisierung hirnspezifischer Proteine im Zusammenhang mit der Alzheimer-Krankheit”</i>
01/2006 – 10/2006	Diploma project; Group of Bioanalytics, Center for Biotechnology and Biomedicine, University Leipzig;
03/2005 – 06/2005	Exchange student (Erasmus) at the University of Uppsala, Sweden; <i>“Protein structure and function”</i>
09/2004	Examination of the basic study period
09/2003	Intermediate diploma
10/2001 – 10/2006	Studies of chemistry, University Leipzig
06/2001	Abitur
08/1994 – 06/2001	Gymnasium im Schloß, Wolfenbüttel
08/1992 – 07/1994	Werla-Schule Schladen
08/1988 – 07/1992	Grundschule Börßum

Publications

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, 2784-96.

Schmidt, C.*, Gronbørg, M.*, Deckert, J., Bessonov, S., Lemm, I., Conrad, T., Lührmann, R., and Urlaub, H. Quantification of spliceosomal B and C complexes by mass spectrometry, *in preparation*

Invited publications

Marcus, K., Schmidt, C., and Urlaub, H. (2009). Proteomic basics - Quantification and post-translational modifications of proteins: The 3rd European summer school in Kloster Neustift. *J Proteomics [Epub ahead of print]*

Schmidt, C. and Urlaub, H. (2009). iTRAQ-labeling of in-gel digested proteins for relative quantification. *Methods Mol Biol* 564, 207-26.

Marcus, K., Kühn-Hölsken, E., Schmidt, C., Schulenburg, T., and Urlaub, H. (2008). 'Proteomic basics--sample preparation and separation': the 1st European Summer School in Kloster Neustift, 12-18 August, 2007 Brixen/Bressanone, South Tyrol, Italy. *Proteomics* 8, 230-3.

Schmidt, C., Urlaub, H., and Lenz, C. (2008). Protein-Quantifizierung nach In-Gel Verdau und iTRAQ-Markierung, *Laborwelt*, 9.Jahrgang, Nr4/2008.

Schmidt, C. and Urlaub, H. Analysis of protein complexes by (quantitative) mass spec analysis. *Methods Mol Biol*, *submitted*.

Conference contributions

Schmidt, C., Grote, M., Lenz, C., Lührmann, R., Urlaub, H. (2009) Evaluation of Protein Stoichiometry within Protein-RNA Complexes by Multiple Reaction Monitoring in Comparison to MS Standard Methods; 57th ASMS Conference on Mass Spectrometry and Allied Topics, Philadelphia, Pennsylvania

Schmidt, C., Grote, M., Lenz, C., Lührmann, R., Urlaub, H. (2009) Establishing the protein stoichiometry in Protein Complexes by Multiple Reaction Monitoring, *Proteomic Forum 2009*, Berlin

Schmidt, C., Grote, M., Lenz, C., Lührmann, R., Urlaub, H. (2009) Bestimmung der Stöchiometrie von Proteinen in Protein-Komplexen mittels Multiple Reaction Monitoring; 42. Jahrestagung der Deutschen Gesellschaft für Massenspektrometrie, Konstanz

Schmidt, C., Grønborg, M., Deckert, J., Bessonov, S., Lemm, I., Lührmann, R., Urlaub, H. (2008) Quantitation of native catalytically active spliceosomal B and C complexes by chemical (iTRAQ) and metabolic (SILAC) labeling; 2nd European Summer School "Proteomic Basics", Brixen/Bressanone, South Tyrol, Italy

Schmidt, C., Grønborg, M., Deckert, J., Bessonov, S., Lemm, I., Lührmann, R., Urlaub, H. (2008) Quantitation of native catalytically active spliceosomal B and C complexes by chemical (iTRAQ) and metabolic (SILAC) labeling; 56th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, Colorado

Schmidt, C., Grønborg, M., Karaduman, R., Deckert, J., Bessonov, S., Kastner, B., Stark, H., Lührmann, R., Urlaub, H. (2007) Relative and Absolute Quantitation of Spliceosomal Proteins by Mass Spectrometry; 1st European Summer School "Proteomic Basics", Brixen/Bressanone, South Tyrol, Italy

Invited oral presentations

Absolute Quantitation of Proteins in Protein Complexes by nano LC-MRM, QTRAP Pro Workshop 2009, Applied Biosystems, Darmstadt

Relative und absolute Quantifizierung spleißosomaler Komplexe mittels Massenspektrometrie, Seminar am Medizinischen Proteom Center, Bochum

Stipends

Proteomic Forum 2009, Berlin; Travel Grant, Deutsche Gesellschaft für Proteom Forschung
57th ASMS Conference on Mass Spectrometry and Allied Topics, Philadelphia, Pennsylvania, USA; Travel Grant, GGNB
Bridging Fund, 03/2010 – 05/2010; GGNB

Memberships

American Society for Mass Spectrometry (ASMS)
Deutsche Gesellschaft für Massenspektrometrie (DGMS)
Deutsche Gesellschaft für Proteomforschung (DGPF)
Gesellschaft für Biochemie und Molekularbiologie (GBM)

Conference organization

Co-organization of the

1st European Summer School “Proteomic Basics”, 2007

2nd European Summer School “Proteomic Basics”, 2008

3rd European Summer School “Proteomic Basics”, 2009

in Brixen, Bressanone, South Tyrol, Italy.