

Dreidimensionale elektronenmikroskopische Strukturanalyse des Hefe U1 snRNP

Von der Naturwissenschaftlichen Fakultät
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades
Doktorin der Naturwissenschaften
Dr.rer.nat.
genehmigte Dissertation
von

Dipl.-Biochem. Ilonka Bartoszek

geboren am 30.03.1978 in Lübeck

2010

Referent: Prof. Dr. Walter Müller
Korreferent: Prof. Dr. Bernd Otto
Tag der mündlichen Prüfung: 29. Januar 2010

Zusammenfassung

In dieser Arbeit wurde die dreidimensionale Struktur des U1 snRNP aus *S. cerevisiae* mittels der Einzelmolekül-Transmissionselektronenmikroskopie untersucht. Die gewählte Technik bietet dabei eine Möglichkeit die verschiedenen Strukturen eines bestimmten makromolekularen dynamischen Komplexes in seinen unterschiedlichen Konformationen zu ermitteln.

Das U1 snRNP ist der erste der fünf spleißosomalen Hauptkomponenten, die sich konsekutiv unter Beteiligung weiterer Faktoren auf der prä-mRNA zum Spleißosom formieren. Das katalytisch aktivierte Spleißosom, an dem das U1 snRNP nicht mehr beteiligt ist, schneidet ein Intron heraus und ligiert die Exone miteinander. Dieser Vorgang erfordert höchste Präzision und vorherige eindeutige Festlegung der Spleißstellen, um keine fehlerhaften mRNAs in den weiteren Prozess der eukaryotischen Genexpression zu entlassen. Das U1 snRNP übernimmt bei der Assemblierung des Spleißosoms die Erkennung der 5' Spleißstelle. Ein vollständiges Verständnis für die molekularen Vorgänge bei diesem Erkennungsprozess erfordert strukturelle Informationen über den U1 snRNP.

Das U1 snRNP, bestehend aus der U1 RNA, dem Sm-Ring und 10 weiteren spezifischen Proteinen wurde mit einem Affinitätstag am Snu71 Protein über die Tandem-Affinitäts-Aufreinigungsmethode aus totalem Hefeextrakt isoliert, wobei das Aufreinigungsprotokoll verschiedentlich modifiziert wurde, um die in den EM-Bildern des Partikels ersichtliche strukturelle Heterogenität des Komplexes einzudämmen. Aufgrund dieser anfänglich beobachteten Heterogenität wurden für die folgende 3D Strukturdetermination Algorithmen und Verfahrensweisen gewählt, mit denen es möglich ist, funktional unterschiedliche Konformationen und/oder Konfigurationen einer Struktur in einem Datensatz zu trennen. Als Ergebnis wurden tatsächlich mehrere 3D Rekonstruktionen aus den U1 snRNP Proben erhalten.

Die fünf kalkulierten 3D Strukturen des U1 snRNP wiesen große Unterschiede auf und spiegelten doch nur einen kleinen Teil der strukturellen Heterogenität wieder, wie anhand einiger einzelner Strukturmerkmale aufgezeigt werden konnte. Die geschätzten Massen der gezeigten Strukturen sind in Relation zueinander zu unterschiedlich, als dass sie Strukturen gleicher Konfiguration repräsentieren könnten. Daher wurde der Fehler, der durch eine nicht vollständig erfasste Heterogenität im Datensatz in den 3D Rekonstruktionen entsteht, mittels einer Modelltestreihe, die den Wegfall einer Domäne simulierte, abgeschätzt.

Die unterschiedlichen U1 snRNP Strukturen legen die Vermutung nahe, dass die Assemblierungsstadien des U1 snRNP biochemisch nicht ausreichend getrennt werden können, da das U1 snRNP in *S. cerevisiae* in mehreren fortgeschrittenen Assemblierungsstadien vorliegt, als beispielsweise im humanen Organismus. Die nur *in silico* trennbaren Assemblierungsstadien des U1 snRNP in *S. cerevisiae* spiegeln daher die Dynamik wieder, mit der das Partikel *in vivo* assembliert, um kotranskriptionell das Substrat, die prä-mRNA, dem Spleißprozess zu übergeben. Es kann nicht ausgeschlossen werden, dass mitunter weitere Proteine, die an der Assemblierung vom U1 snRNP zum Commitment Komplex oder an der Expression bestimmter Gene beteiligt sind, zur Heterogenität beitragen.

Schlagworte: U1 snRNP, *Saccharomyces cerevisiae*, Einzelmolekül-Elektronenmikroskopie

Abstract

Three-dimensional structure determination of the U1 snRNP in *S. cerevisiae* has been done using single-particle transmission electron microscopy. This technique allows the study of heterogeneous particle datasets and has therefore the potential to capture and analyse the structures of macromolecular complexes in their diverse conformational states.

The U1 snRNP is one of in total five main small nuclear ribonucleoprotein complexes which assemble on the eukaryotic pre-mRNA substrate consecutively in order to build the pre-catalytic spliceosome. In this process several other splicing factors are also involved. Before the spliceosome becomes catalytically active, U1 snRNP actually leaves the complex. The catalytically active spliceosome then cuts the intron and ligates the exons for correct gene expression later. As no erroneous mRNAs should be released for consequent gene expression this procedure needs to be highly accurate. The splice sites need to be defined precisely beforehand. The U1 snRNP is in charge of defining the 5' splice site cotranscriptionally. Furthermore it bears a crucial role in the early spliceosomal complexes, in which the intron is spanned to overcome the distance of 5' and 3' splice sites. A full mechanistic understanding of these processes requires structural information of the U1 snRNP complex.

The U1 snRNP consists of the U1 RNA, the 7 proteins of the Sm-core and 10 proteins specific to U1. It was purified via an established tandem-affinity-purification protocol from total yeast extract. This purification protocol was modified at different stages to reduce the observed heterogeneity of the U1 snRNP structure. The subsequent 3D structures were determined with computational algorithms with which it is, in general, possible to separate different configurations and/or conformations of a heterogeneous complex. As a result, indeed several 3D reconstructions of the U1 snRNP were obtained.

The five calculated 3D models of the U1 snRNP showed significant structural differences. Also, the estimated masses of these 3D structures when compared with each other were quite different to indicate a common structural configuration. Moreover, it could be shown that the five 3D models only reflected a small representation of the structural variety of the isolated sample. This representation could be shown on the basis of some structural details. Therefore, the inaccuracy of structure determination emerging from a high level of heterogeneity in the dataset was estimated. A 3D model set, which simulated loss of a structural domain, was used

for this purpose.

In summary these structures show that the states of assembly cannot be accurately separated, probably because the U1 snRNP in yeast exists in more various different states of assembly when compared to the human one. The computationally separable states of the yeast U1 snRNP may thus reflect the dynamic state in which the U1 snRNP assembles *in vivo* to commit the pre-mRNA to the splicing pathway cotranscriptionally. In addition or alternatively other proteins, which are involved in assembly of the U1 snRNP in a sub-stoichiometric manner, might also contribute to the particle's heterogeneity.

Keywords: U1 snRNP, *Saccharomyces cerevisiae*, Single-particle electron microscopy

Inhaltsverzeichnis

Zusammenfassung	iii
Abstract	v
Inhalt	vii
Abbildungsverzeichnis	xi
Tabellenverzeichnis	xiii
Abkürzungsverzeichnis	xv
Danksagung	xix
1 Einleitung	1
1.1 prä-mRNA Prozessierung	1
1.2 Der Spleißmechanismus	3
1.2.1 Die 2-stufige Spleißreaktion	3
1.2.2 Die Definition der Spleißstellen	3
1.3 Komponenten und Assemblierung des Spleißosoms	6
1.3.1 Der Spleißzyklus	6
1.3.2 U snRNP und weitere Spleißfaktoren	10
1.4 Das U1 snRNP in <i>S. cerevisiae</i>	16
1.4.1 Komposition des U1 snRNP in <i>S. cerevisiae</i>	16
1.4.2 Die Interaktionen mit der prä-mRNA im CC1 und CC2	25
1.4.3 Der E-Komplex	26
1.5 <i>S. cerevisiae</i> als Modellorganismus für den Spleißprozess	28
1.6 Strukturuntersuchungen an spleißosomalen Komplexen	28
1.7 Transmissionselektronenmikroskopie (TEM)	30
1.7.1 Bildentstehung im TEM	30
1.7.2 Das elektronenmikroskopische Bild	33
1.7.3 Phasenkontrastübertragungsfunktion	33
1.8 Ziel dieser Arbeit	34

2	Einleitung in die Methodik der Bildverarbeitung	35
2.1	Datensatz aus Einzelpartikelbildern	35
2.1.1	Aufnahmetechniken	35
2.1.2	Selektion der Partikel	36
2.1.3	Bandpassfilter	36
2.2	Datenprozessierung auf 2D Ebene	37
2.2.1	<i>Alignment</i>	38
2.2.2	Klassifizierung	40
2.2.3	Zyklische Datenprozessierung	44
2.3	Berechnungen auf 3D Struktur Ebene	44
2.3.1	<i>de novo</i> Rekonstruktion der 3D Struktur	47
2.3.2	Maximum Likelihood Methode (ML)	49
2.3.3	Verbesserung der 3D Struktur	50
2.3.4	Rekonstruktion der 3D Struktur via Rückprojektion	54
2.3.5	Visualisierung der 3D Struktur	56
2.3.6	Bestimmung der Auflösung einer 3D EM Struktur	56
3	Materialien und Methoden	59
3.1	Materialien	59
3.1.1	Computer Software und spezielle Hardware	59
3.1.2	Chemikalien und Pufferlösungen	60
3.1.3	Labormaterialien und Geräte	62
3.2	Biochemische Methoden	63
3.2.1	Aufreinigung des U1 snRNP aus <i>S. cerevisiae</i>	63
3.2.2	Markierung einzelner Proteine im <i>S. cerevisiae</i> U1 snRNP	67
3.3	Elektronenmikroskopie	68
3.3.1	Auftragen von Kohlelochfolien auf Kupfernetzträger	68
3.3.2	Herstellung kontinuierlicher Kohlefilme	69
3.3.3	Präparation negativ kontrastierter Proben	70
3.3.4	Aufnahmen am EM	72
3.3.5	Bildverarbeitung	72
4	Ergebnisse	75
4.1	Aufreinigung des U1 snRNP aus <i>S. cerevisiae</i>	75
4.1.1	Aufreinigung ohne Vorfraktionierung	75
4.1.2	Aufreinigung mit Vorfraktionierung	80
4.1.3	Markierung einzelner Proteine im U1 snRNP	85

4.2	3D Strukturen des U1 snRNP	90
4.2.1	<i>de novo</i> 3D Struktur-Rekonstruktion	90
4.3	Verbesserung und Validierungen der 3D Strukturen	94
4.4	Fehlerabschätzung der 3D Rekonstruktion bei Heterogenität der Probe .	103
4.4.1	Testdaten und Verfahren	103
4.4.2	Statistische Auswertung	106
5	Diskussion	117
5.1	Heterogenität des Hefe U1 snRNP Komplexes	117
5.1.1	Modifikationen der Aufreinigung	117
5.2	3D Strukturen des Hefe U1 snRNP	122
5.2.1	Auswirkung der Heterogenität auf die Lokalisation markierter Proteine im U1 snRNP	128
5.3	Schlussfolgerungen	131
5.3.1	U1 snRNP in <i>S. cerevisiae</i>	131
5.3.2	Ausblick	135
	Literaturverzeichnis	137
	I Appendix	159
	Massenspektroskopische Daten	161
	Erklärung/Affidavit	169
	Persönliche Daten	171

Abbildungsverzeichnis

1.1	Die Spleißreaktion	4
1.2	Vergleich der Konsensussequenzen der Introns in <i>S. cerevisiae</i> und Metazoa	5
1.3	Der Spleißzyklus	8
1.4	Die Proteinzusammensetzung der humanen und <i>S. cerevisiae</i> snRNP . .	11
1.5	Konservierte Regionen in den snRNA	12
1.6	Übersicht der Proteine des humanen und <i>S. cerevisiae</i> U1 snRNP . . .	17
1.7	Die Pimärsequenzen und vorgeschlagenen Sekundärstrukturen der humanen und <i>S. cerevisiae</i> U1 snRNA	18
1.8	Lebensfähige Deletionsmutanten der <i>S. cerevisiae</i> U1 snRNA	19
1.9	Entstehung von 2D Projektionsbildern im Transmissionselektronenmikroskop	31
2.1	Schematische Darstellung der Ausrichtung der EM-Einzelpartikelbilder	38
2.2	Referenz für das initiale translationale <i>Alignment</i>	40
2.3	Eigenbilder	42
2.4	Übersicht der zyklischen EM-Datenverarbeitung	45
2.5	Die Definition der Eulerwinkel	46
2.6	Die RCT-Methode	48
2.7	Äquivalenz von Bild- und Fourierraum	51
2.8	Ermittlung der gemeinsamen Linienprojektion	53
2.9	Illustration des Rückprojektionsverfahrens	55
3.1	Aufreinigung des U1 snRNP aus <i>S. cerevisiae</i>	64
3.2	Illustration der Präparation negativ kontrastierter Proben	71
4.1	Proteinanalyse des Gradienten aufgereinigten U1 snRNP	76
4.2	Charakteristische Klassensummen des U1 snRNP erhalten aus Glycerolgradientenfractionen ohne Fixativzugabe	78
4.3	Charakteristische Klassensummen des U1 snRNP erhalten aus Glycerolgradientenfractionen mit Fixativzugabe	79

4.4	Analyse der Gradientenfractionen der Vorfraktionierung	82
4.5	Proteinanalyse des Gradienten aufgereinigten U1 snRNP	83
4.6	Charakteristische Klassensummen des über Vorfraktionierung, TAP-Methode und fixierenden Glycerolgradienten aufgereinigten U1 snRNP .	84
4.7	Titration des anti-CBP Antikörpers gegen TAP-tag aufgereinigten U1 snRNP	87
4.8	Untersuchung der Expression des tDimer2 Fusionsproteins Snp1p im U1 snRNP	89
4.9	Elektronenmikroskopische Aufnahmen des U1 snRNP bei 0° und 45° geneigtem Probenhalter	91
4.10	<i>de novo</i> 3D Rekonstruktion des U1 snRNP mittels RCT-Methode . . .	92
4.11	3D Strukturen des U1 snRNP nach <i>ML-Alignment</i> und anschließender 3D Klassifizierung	93
4.12	<i>Refinement</i> -Zyklen zur Validierung einer U1 snRNP Substruktur	96
4.13	3D Strukturen des U1 snRNP aus <i>S. cerevisiae</i>	97
4.14	Auflösung der berechneten 3D Strukturen	99
4.15	Projektionswinkelverteilung in den Subdatensätzen I - III	100
4.16	Projektionswinkelverteilung in den Subdatensätzen IV und V	101
4.17	<i>Refinement</i> -Zyklen zur Validierung einer U1 snRNP Dömane	102
4.18	Visualisierung der 3D Modelle des Testdatensatzes	105
4.19	Trigonometrische Grundlagen der Fehlerberechnung	107
4.20	Mittlere Abweichung der Winkel Ψ in Grad in der Modellreihe	108
4.21	Häufigkeitsverteilung der Winkelabweichungen nach der <i>angular reconstruction</i> -Methodik	110
4.22	Häufigkeitsverteilung der Winkelabweichungen nach Normalisierung . .	111
4.23	Rekonstruierte Modelle	112
4.24	Mittlere Abweichungen des α -Winkels und des Ψ -Winkels nach <i>MR-Alignment</i>	114
4.25	FSC-Funktionen der berechneten 3D Strukturen der Testdatenreihe . .	115
5.1	Abhängigkeit der Salzkonzentrationen in der Aufreinigung des U1 snRNP	121
5.2	Schwierigkeiten bei der Auswertung von Positionen eines Labels	130
5.3	Modell der Biogenese des U1 snRNP	134
4	Verwendete Proteingelbanden für die massenspektroskopische Analyse der Proteine des U1 snRNP, aufgereinigt bei 200 <i>mM</i> KCl	162
5	Verwendete Proteingelbanden für die massenspektroskopische Analyse der Proteine des U1 snRNP, aufgereinigt bei 400 <i>mM</i> KCl	165

Tabellenverzeichnis

3.1	Verwendete Computersoftware und spezielle Hardware	59
3.2	Zusammensetzung der verwendeten Pufferlösungen	60
3.3	Verwendete Chemikalien, Enzyme und Antikörper	61
3.4	Verwendete Materialien	62
3.5	Genotypen der verwendeten Hefestämme	62
3.6	Verwendete Laborgeräte	63
3.7	Zur 3D Rekonstruktion von U1 snRNP verwendete Datensätze	72
4.1	3D Strukturen I - V, kalkuliert aus 107147 Einzelpartikelbildern einer kryo-negativ-kontrastierten U1 snRNP Probe.	98
4.2	3D Modelle des Testdatensatzes	104
4.3	Logarithmische Klasseneinteilung für die Häufigkeitsverteilung von Ψ .	109
1	Teil 1, Massenspektroskopische Analyse der Proteinbanden 1 – 14 aus Abb. 4. In die Liste aufgenommen wurden nur Proteine, die mit einem Score von ≥ 25 identifiziert wurden und aus <i>S. cerevisiae</i> stammen. . .	163
2	Teil 2, Fortsetzung der Tabelle 1: Massenspektroskopische Analyse der Proteinbanden 15 – 28 aus Abb. 4.	164
3	Teil 1, Massenspektroskopische Analyse der Proteinbanden 1 – 20 aus Abb. 5. In die Liste aufgenommen wurden nur Proteine, die mit einem Score von ≥ 25 identifiziert wurden und aus <i>S. cerevisiae</i> stammen. . .	166
4	Teil 2, Fortsetzung der Tabelle 3: Massenspektroskopische Analyse der Proteinbanden 21 – 41 aus Abb. 5.	167
5	Teil 3, Fortsetzung der Tabelle 3: Massenspektroskopische Analyse der Proteinbanden 42 – 54 aus Abb. 5.	168

Abkürzungsverzeichnis

Abkürzung	Begriff
1D	eindimensional
2D	zweidimensional
3D	dreidimensional
3' Ss	3' Spleißstelle
5' Ss	5' Spleißstelle
A	Adenosin
Abb.	Abbildung
Abschn.	Abschnitt
anti-CBP-AK	Calmodulinbindungsprotein-bindender Antikörper
approx.	approximativ
AS	Aminosäure
ATP	Adenosintriphosphat
BBP	<i>engl.</i> branchpoint binding protein, Verzweigungspunkt-Bindungsprotein
bp	<i>engl.</i> base pair, Basenpaare
C	Cytidin
CBP	Calmodulin-Bindungs-Peptid
CBC	<i>engl.</i> Cap-binding complex, (RNA) Kappen-bindender Komplex
CC	<i>engl.</i> Commitment complex, <i>Commitment</i> -Komplex, der die prä-mRNA in den Spleißprozess einweist
CCC	<i>engl.</i> cross correlation coefficient, Kreuzkorrelationskoeffizient
CCD	Charge Coupled Device
DNA	<i>engl.</i> deoxyribonucleic acid, Desoxyribonukleinsäure
DTT	1,4-Dithiothreitol
EDTA	Ethylendiaminotetraessigsäure
EGTA	Ethylenglykol-bis-(β -aminoethylether)-N,N,N',N'-tetraessigsäure
ELISA	<i>engl.</i> Enzyme-linked Immunosorbent Assay, Enzymimmunoassay
EM	Elektronenmikroskopie, elektronenmikroskopisch
<i>engl.</i>	englisch
FFT	Fourier Transformation (fast discrete)
FSC	<i>engl.</i> Fourier Shell Correlation, Fourier-Schalen-Korrelation

Abkürzung	Begriff
G	Guanosin
GFP	Green Fluorescent Protein
HEPES	N-2-Hydroxyethylpiperazin-N-2-ethansulfonsäure
IgG	Immunglobulin G
Kap.	Kapitel
Lsm	<i>engl.</i> Like-Sm, Sm-Protein ähnlich
m_3G	2,2,7-Trimethylguanosin
ML- <i>Alignment</i>	Maximum-Likelihood- <i>Alignment</i>
MR- <i>Alignment</i>	Multi-Referenz- <i>Alignment</i>
mRNA	<i>engl.</i> messenger RNA, Boten-Ribonukleinsäure
PAGE	Polyacrylamidgelelektrophorese
PC	Personal Computer
PCA	<i>engl.</i> Principal Component Analysis, Hauptkomponenten-analyse
PCTF	Phasenkontrasttransferfunktion
PMSF	Phenylmethylsulfonylflorid
prä-mRNA	prä-Messenger-RNA
Prot A	<i>Staphylococcus aureus</i> Protein A
RCT	<i>engl.</i> Random Conical Tilt, Rekonstruktionsmethode
RNA	<i>engl.</i> ribonucleic acid, Ribonukleinsäure
RBD	<i>engl.</i> RNA binding domaine, RNA Bindungsdomäne
RRM	<i>engl.</i> RNA recognition motif, RNA-Erkennungsmotiv
rRNA	ribosomal RNA, ribosomale Ribonukleinsäure
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
RT	Raumtemperatur
SDS	<i>engl.</i> sodium dodecyl sulfate, Natriumdodecylsulfat
SNR	<i>engl.</i> signal-to-noise ratio, Signal-zu-Rausch Verhältnis
snRNA	small nuclear RNA, kleine nukleäre Ribonukleinsäure(n)
<i>S. pombe</i>	<i>Saccharomyces pombe</i>
sog.	sogenannt
Tab.	Tabelle
TAP	<i>engl.</i> tandem-affinity-purification, Tandem-Affinitäts-Aufreinigung
TEM	Transmissionselektronenmikroskopie und Transmissions-elektronenmikroskop
TEV Protease	Tabakmosaiketchvirus Protease
TPR	<i>engl.</i> tetratricopeptide repeat, Proteinstrukturmotiv
tRNA	Transfer-Ribonukleinsäure

Abkürzung	Begriff
U	Uridin
(U) snRNP	<i>engl.</i> (Uridin-rich) small nuclear ribonucleoparticle(s),(Uridinreiche) kleine(s) nucleäre(s) Ribonukleopartikel
UTR	<i>engl.</i> untranslated region, nicht-translatierter mRNA-Segmente
UV	Ultraviolett
vgl.	vergleiche
YFP	Yellow Fluorescent Protein
Zn	Zink

Abkürzung	Einheit
Å	Ångström
°C	Grad Celsius
<i>h</i>	Stunde
<i>kDa</i>	Kilo Dalton
<i>kV</i>	Kilo Volt
<i>l</i>	Liter
μ l	Mikroliter
<i>min</i>	Minute
<i>ml</i>	Milliliter
<i>mM</i>	millimolar, Millimol
<i>nm</i>	Nanometer
<i>rpm</i>	<i>rounds per minute</i> , Umdrehungen pro Minute
<i>s</i>	Sekunde
<i>S</i>	Svedberg

Danksagung

Ganz besonders danke ich Herrn Prof. Dr. Müller für die Bereitschaft, diese Doktorarbeit zu betreuen. So ist mir die Möglichkeit gegeben, an der Leibniz Universität Hannover zu promovieren, obwohl die experimentellen Arbeiten in Göttingen am Max-Planck-Institut für biophysikalische Chemie durchgeführt worden sind.

Vielen Dank den Professoren Dr. Holger Stark und Dr. Reinhard Lührmann, die mir das Thema dieser Dissertation vertrauensvoll zur Bearbeitung überließen. Bei Prof. Dr. Holger Stark bedanke ich mich darüber hinaus für die Einführung in die Arbeit mit dem Elektronenmikroskop und seine „offene Bürotür“.

Dr. Ramazan Karaduman danke ich für die gute Zusammenarbeit und die biochemische Vorarbeit, die für die Experimente nötig war.

Bei diversen Computer Hardware und/oder Softwareproblemen halfen mir Tobias Koske, Karl-Heinz Knauber und Dr. Mario Lüttich.

In verschiedenen Stadien dieser Arbeit gaben mir wertvolle Ratschläge und die Möglichkeit zur Diskussion von Ergebnissen die Doktoros Monika Golas, Björn Sander, Irina Häcker, Florian Hauer, Prakash Dube, Klaus Hartmut, Berthold Kastner und Tobias Schwarz.

Elmar Wolff, Michael Grote, Kohansal Hossein und Peter Kempkes überließen mir ihre Workbench zur Mitbenutzung bei meinen diversen Aufgaben im biochemischen Labor. Für eine entspannte Arbeitsatmosphäre sorgten auch Florian Platzmann, Dr. Burkhard Heisen, Dr. Niels Fischer, Dr. Martin Schmeißer, Boris Busche, Dr. Dietmar Riedel, Dr. Dirk Wenzel, Gudrun Heim und Andrius Krasauskas.

Ihnen allen danke ich herzlich.

Dankbar bin ich meinen Eltern für die großartige stetige Förderung aller meiner Vorhaben und Unterstützung in allen Lebensumständen.

Die vorliegende Arbeit entstand am

Max-Planck-Institut für Biophysikalische Chemie
Am Faßberg 11
37077 Göttingen

in der Arbeitsgruppe von

Prof. Dr. Holger Stark
Kryo-Elektronenmikroskopie

und in enger Kooperation mit der Abteilung von

Prof. Dr. Reinhard Lührmann
Zelluläre Biochemie

**An der Gottfried Wilhelm Leibniz Universität Hannover wurde die
Betreuung der Dissertation übernommen von**

Prof. Dr. Walter Müller
Institut für Physiologische Chemie
Medizinische Hochschule Hannover
Zentrum Biochemie - OE 4310
30625 Hannover

Kapitel 1

Einleitung

1.1 prä-mRNA Prozessierung

Genexpression umfasst alle Schritte, die der Konvertierung der in der DNA enthaltenen Informationen in primäre Aminosäuresequenzen dient, und zwar Transkription, RNA-Prozessierung und Translation. In eukaryontischen Zellen ist ein Teil der Gene mit nicht-kodierenden DNA-Abschnitten, den Introns, unterbrochen. Während nur die kodierenden Sequenzabschnitte, die Exons, die zu translatierende Information enthalten, werden die Introns dafür nicht benötigt. Der Anteil der Gene im Genom, die mit Introns unterbrochen sind, ist abhängig von der Spezies. In Säugetierzellen tragen die meisten Gene wenigstens ein Intron, das an einer beliebigen Stelle in die kodierende Sequenz integriert sein kann. In *Saccharomyces cerevisiae* (*S. cerevisiae*) gibt es nur wenige Gene, die von Introns unterbrochen sind, und diese sind zumeist kurz und liegen nahe dem 5' Ende des Gens [135]. In Zahlen konnten im Hefe Genom nur 353 Introns in 248 von insgesamt 6000 Hefegenen identifiziert werden [246]. Das heißt, dass nur 3,8% der Gene in Hefe ein Intron tragen. Da allerdings 27% der Transkripte in Hefe gespleißt werden, deutet dies auf eine hohe Expressionsrate der Gene mit Intron hin. Während die humanen Introns bis zu 100000 Nukleotide lang sein können, zeigten sich in einer Analyse der durchschnittlichen Länge der Introns in Hefe zwei Maxima bei 100 und 400 Nukleotiden. Im Nukleus wird ein Gen von der RNA Polymerase II komplett, bestehend aus Introns und Exons, in ein Primärtranskript, die so genannte prä-mRNA oder hnRNA (*engl.* heterogeneous nuclear RNA), transkribiert. Die prä-mRNA wird zum einen durch den Spleißprozess verändert, und zum anderen werden beide Enden, 3' und 5' Ende der prä-mRNA, chemisch modifiziert. Letzteres geschieht unter anderem mit der Funktion, das Transkript vor Degradation durch hydrolysierende Enzyme zu schützen. Am 5' Triphosphat-Ende vollzieht sich die Modifikation am entstehenden Transkript noch bevor es größer als 20 – 25 bp ist.

In einer dreistufigen Reaktion wird das 5' Triphosphat-Ende verändert, indem dort mittels einer spezifischen Guanosyl-Transferase ein Guanosintriphosphat in umgekehrter Orientierung unter Bildung einer 5'-5'-Triphosphat-Brücke angefügt wird [233]. In Säugerzellen weist die Guanosyl-Transferase Bifunktionalität auf, da sie mit ihrer Phosphataseaktivität zuvor das 5' Triphosphat-Ende um eine Phosphatgruppe dephosphoryliert. In *S. cerevisiae* wird für die Dephosphorylierung eine unabhängig kodierte RNA-5'-Triphosphatase benötigt [234]. Diese Kappen-Struktur (*engl.* cap-structure) wird durch Methylierung editiert. Die Methylierung an Position 7 des terminalen Guanins mit Hilfe der Guanin-7-Methyltransferase wird m^7G -Kappe (*engl.* Cap 0) genannt. Diese Methylierungsreaktion wird in allen Eukaryonten vorgenommen und ist in eukaryontischen Einzellern die einzige Veränderung. In den meisten anderen Eukaryonten wird allerdings zusätzlich an der vorherigen Base, der letzten Base des ursprünglichen Transkripts, an der 2'-O Position methyliert (*engl.* Cap 1).

Die Enzyme, die für die Bildung der 5' Kappenstruktur benötigt werden, binden an die phosphorylierte Form der Carboxyl-terminalen Domäne (CTD) der größten Untereinheit der RNA-Polymerase II, die für diese Polymerase einzigartig ist und die eine über Speziesgrenzen stark konservierte Hepta-Peptid-Konsensussequenz aufweist [161]. Diese Beobachtung erklärt sowohl die Spezifität der m^7G -Kappenstruktur für RNA-Polymerase II Transkripte, als auch ihre Synthese am naszierenden Transkript. Die 5' Kappenstruktur dient neben dem Schutz vor Degradation dem Spleißprozess (Abschn. 1.3.1), dem mRNA-Export [38] vom Nukleus ins Cytoplasma, und der Regulation und Initiation der Translation [243], [71]. Diese vielfältigen Funktionen werden vermittelt über nukleäre und cytosolische Proteine, die an die m^7G -Kappenstruktur binden. Im Nukleus bindet der sogenannte 'Kappen-bindende Komplex', CBC¹. Der CBC besteht aus einem Heterodimer, und zwar in humanen Zellen aus den Proteinen CBP20 und CBP80 und in *S. cerevisiae* aus den Proteinen Cbc2/Mud13 und Cbc1/Sto1 (Abschn. 1.4.2).

Am 3' Exon der prä-mRNA werden in einem zweistufigen Prozess bis zu 250 Adenosinphosphate addiert. Nachdem das wachsende Transkript an einer spezifischen Stelle geschnitten worden ist, wird die Polyadenylierung an der freien 3' OH Gruppe von der Poly(A)-Polymerase katalysiert [41].

Gleichfalls kotranskriptional werden die Introns mit Hilfe des Spleißosoms [26], [68], [86], einer komplexen dynamischen makromolekularen Maschine, die auf jeder prä-mRNA stets neu assembliert, herausgeschnitten und die Exons werden zusammengefügt. Dieser Vorgang wird Spleißen genannt. Durch unterschiedliche Kombinationen von Exons können, ausgehend vom selben Gen durch alternatives Spleißen [18], verschiedene Proteine gebildet werden.

¹*engl.* cap-binding complex, Kappen-bindender Komplex

Das Spleißprodukt, die mRNA, verlässt durch die Poren des Nukleus den Zellkern, wobei der Poly(A)Schwanz eine regulatorische Rolle spielt [284], und wird anschließend an den Ribosomen translatiert. Obwohl die Enden der reifen mRNA von Exons kodiert sind, tragen sie Informationen, die nicht in Aminosäuresequenz translatiert werden. Diese Segmente werden auch UTRs (*engl.* untranslated regions) genannt und beinhalten Sequenzen, die die Stabilität und die translationale Aktivität der mRNA regulieren.

1.2 Der Spleißmechanismus

1.2.1 Die 2-stufige Spleißreaktion

Die Spleißreaktion, um ein Intron herauszuschneiden und die Enden der Exons zu ligieren, umfasst zwei aufeinanderfolgende Umesterungen [164], (Abb 1.1). Im ersten Schritt wird eine 2',5'-Phosphodiesterbindung gebildet, und zwar zwischen der 2' Hydroxylgruppe eines internen Adenosins im Intron, am sogenannten Verzweigungspunkt (*engl.* branchpoint), und der 5' terminalen Phosphatgruppe desselben Introns. Dadurch wird das 5' Exon abgetrennt, und das Intron nimmt eine Lasso-Struktur ein. Das Adenosin des Verzweigungspunkts befindet sich innerhalb einer konservierten Sequenz aus 20 – 50 *bp* stromaufwärts der 3' Spleißstelle (3' Ss) im Intron. Am 5' Exon liegt nach der ersten Reaktion die 3' OH-Gruppe frei, die im zweiten Reaktionsschritt nukleophil die 3' Intron/Exon Verbindung angreift. Dabei wird eine Phosphodiesterbindung mit der 5' terminalen Phosphatgruppe des 3' Exons geknüpft, so dass die Exons ligiert werden und das Intron in Lasso-Form freigesetzt wird. In vivo wird das freie Intron schnell linearisiert und abgebaut [267]. Die zweistufige Spleißreaktion verläuft ohne Energiezufuhr. Da für jede gespaltene Phosphodiesterbindung eine neue gebildet wird, bleibt die Anzahl an Phosphodiesterbindungen gleich und die freie Energie wird auf diese Weise konserviert. Für die Spleißreaktionen scheint die Anwesenheit eines Metallions im katalytischen Zentrum notwendig [281], [251], was für die erste Spleißreaktion experimentell nachgewiesen werden konnte [244]. Zu beachten ist, dass für die Assemblierung des katalytisch aktiven Spleißosoms (Abschn. 1.3.1) sehr wohl Energiezufuhr durch die Hydrolyse von mehreren ATPs notwendig ist [28].

1.2.2 Die Definition der Spleißstellen

Um ein Intron korrekt und unter Beibehaltung des Leserasters aus der prä-mRNA zu spleißen, müssen die Spleißstellen zweifelsfrei definiert sein. Nur dann kann das Spleißosom an der richtigen Stelle auf der prä-mRNA assemblieren. Die Spleißstellen sind daher

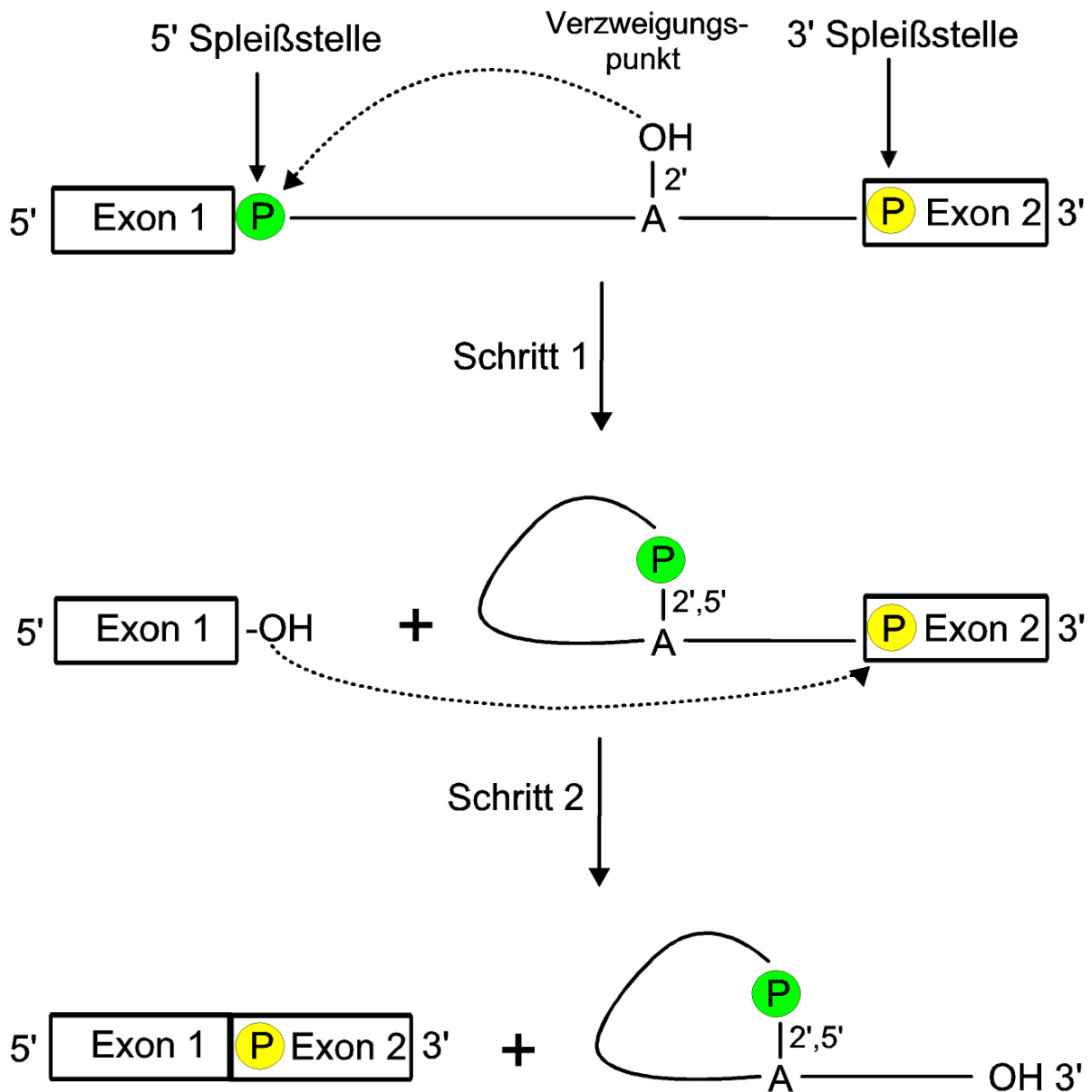


Abbildung 1.1: Die Spleißreaktion. In zwei aufeinander folgenden Reaktionsschritten wird das Intron entfernt und die Exone ligiert. **Schritt 1:** Die 2' OH-Gruppe des Adenosins (A) am Verzweigungspunkt des Introns knüpft eine Phosphodiesterbindung mit dem Phosphat der Phosphatgruppe an der 5' Spleißstelle (5' Ss). **Schritt 2:** Die im ersten Schritt erzeugte 3' OH-Gruppe am Exon 1 bildet eine neue Phosphodiesterbindung mit dem Phosphat der Phosphatgruppe des Exons 2 an der 3' Ss. (P) symbolisiert die beteiligten 5' Phosphatgruppen (PO_3)²⁻, wobei zur Unterscheidung die Phosphatgruppe des Introns grün und die des 3' Exons gelb unterlegt ist.

stark konservierte Basensequenzen an den 3' und 5' Enden des Introns, und auch am Verzweigungspunkt im Intron gibt es eine nötige Konsensussequenz [204].

Die Anforderungen an die Konsensusabschnitte sind in Hefe und Mensch allerdings unterschiedlich und spiegeln die verschiedene Art und Weise der Organisation der Gene und deren Expression wieder (Abb. 1.2). Im humanen System tragen viele Gene meist mehrere Introns und durch alternatives Spleißen wird die Vielfalt zusätzlich erhöht. Dies erfordert eine komplexe Regulation durch verschiedene Faktoren. Eine geringere Sequenzkonservierung an den Spleißstellen [171] könnte dieser Vielfalt Rechnung tragen. Mehr als 98% der Introns im humanen Genom tragen am 5' Exon das Dinukleotid

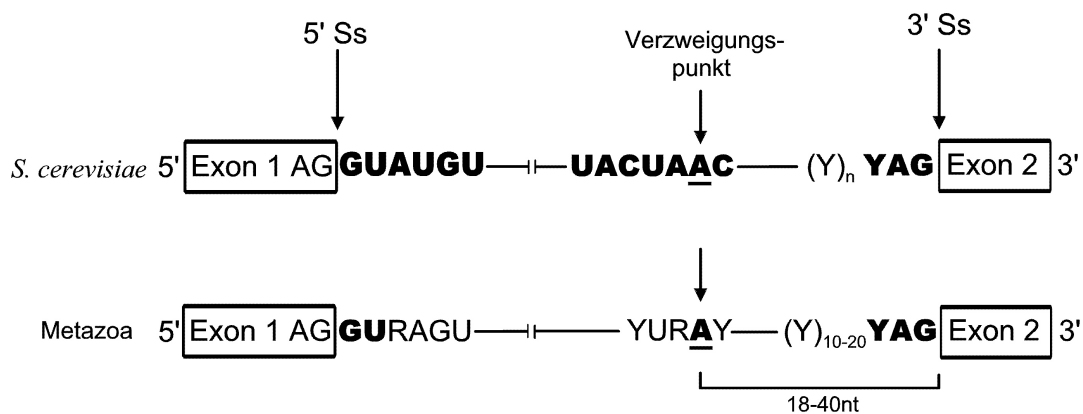


Abbildung 1.2: Vergleich der Konsensussequenzen der Introns in *S. cerevisiae* und Metazoa. Nukleotide, die zu mehr als 90% konserviert sind, sind fett gedruckt. Das Adenosin am Verzweigungspunkt ist unterstrichen; Purine sind mit 'R' und Pyrimidine mit 'Y' bezeichnet.

GU und am 3' Exon AG [180]. Eine geringe Zahl Introns ($< 1\%$) trägt am 5' Exon als Variante das Dinukleotid GC. Damit qualifizieren sie sich für die Erkennung vom majoren Spleißosom (Abschn. 1.3). Eine sehr geringere Anzahl Introns (0,1%), an den Enden mit den Dinukleotiden AU (5' Exon) und AC (3' Exon) markiert, werden vom minoren Spleißosom erkannt, das dem majoren Spleißosom zwar funktional und im prinzipiellen Aufbau analog ist, jedoch bei den einzelnen snRNP-Komponenten (*engl.* small nuclear ribonucleoprotein particles, Abschn. 1.3) eine geringe Sequenzhomologie aufweist [259], [182]. Die evolutionsgeschichtliche Beziehung der beiden Spleißsysteme ist noch unklar. Die Introns werden zur Unterscheidung U2-Typ- und U12-Typ-Introns genannt, wobei erstere vom majoren und letztere vom minoren Spleißosom prozessiert werden. In *S. cerevisiae* ist die 5' Ss durch den längeren Sequenzabschnitt R/**GUAUGU** (/markiert die Exon/Intron Grenze, R ist ein Purin) definiert, wobei die Nukleotide zu 90% und mehr konserviert sind [151]. Der Verzweigungspunkt im Intron mit dem konservierten Adenosin (hervorgehoben) liegt in Hefe in der Sequenzbox UACUAAC

[151]. In höheren Eukaryonten ist dieser Bereich nicht so stark konserviert, liegt aber in einem pyrimidinreichen Sequenzabschnitt $Y_{80}NY_{80}Y_{87}R_{75}AY_{95}$ (Y = Pyrimidinbase, tiefgestellte Zahlen geben die statistische Häufigkeit dieser Base an der Stelle wieder). Da der Verzweigungspunkt stets 18 – 40 Nukleotide stromaufwärts von der 3' Ss liegt, ist er auch wichtig, um eben genau diese, dem Verzweigungspunkt nächst gelegene 3' Ss, als Substrat für die Ligation mit dem 5' Exon zu identifizieren. Zusätzlich ist der Abschnitt zwischen Verzweigungspunkt und 3' Ss häufig reich an Pyrimidin Basen (U und C; *engl.* poly-pyrimidine tract). Dieser pyrimidinreiche Abschnitt ist 8 – 12 Nukleotide lang und endet ein Nukleotid stromaufwärts vor der 3' Ss. Die Sequenz am 3' Ende des Introns lautet YAG. Der Polypyrimidinabschnitt scheint im humanen Genom aufgrund des weniger stark konservierten Verzweigungspunkts eine wichtigere Rolle zu spielen als in Hefe.

1.3 Komponenten und Assemblierung des Spleißosoms

Das gesamte Spleißosom wird gebildet aus fünf Subkomplexen, den sogenannten 'Uridine-reich small nuclear ribonucleoparticles', kurz U snRNP oder nur snRNP, und verschiedenen weiteren Spleiß-Faktoren (*engl.* non-small ribonucleoproteins). Den U1-, U2-, U4-, U5- und U6- snRNP Komplexen ist ein prinzipieller Aufbau aus RNA und Proteinen (Abschn. 1.3.2) gemein. Sie sind die spleißosomalen Hauptkomponenten, die in allen eukaryontischen Spezies universal auftreten, im prinzipiellen Aufbau stark über die Speziesgrenzen konserviert sind, und die die große Mehrheit aller prä-mRNA Introns in der Zelle spleißen [28], [288]. Da dieses Spleißsystem in *S. cerevisiae* als einziges vorhanden ist, beziehen sich im Folgenden alle Beschreibungen auf das Spleißosom, das Introns des U2-Typs spleißt.

1.3.1 Der Spleißzyklus

Das Modell der sequenziellen Assemblierung des Spleißosoms beschreibt den Spleißzyklus über die sukzessive Zusammenlagerung der spleißosomalen snRNP Komplexe auf der prä-mRNA zu Intermediaten und schließlich zum aktiven Spleißosom und den Zerfall in die Einzelkomponenten nach der Katalyse (Abb. 1.3). Die snRNP Subkomplexe, die während der Assemblierung entstehen, und die Reihenfolge ihres Auftretens im Spleißzyklus wurden aus humanen Zellextrakten mit nativer Gelelektrophorese [123], [124] und Affinitätschromatographie [85] bestimmt und vergleichbare Ergebnisse wurden auch in *S. cerevisiae* mit Hilfe der Gelelektrophorese gefunden [190], [39]. Im Folgenden wird die mehr gebräuchliche Nomenklatur der humanen Subkomplexe angegeben und die Be-

zeichnungen der Komplexe in *S. cerevisiae* hinzugefügt.

Die Assemblierung des Spleißosoms beginnt mit der ATP-unabhängigen Bindung des U1 snRNP an die 5' Ss auf der prä-mRNA [206], [209], wobei die U1 snRNA in *S. cerevisiae* typischerweise 5 – 7 komplementäre Basenpaare mit der prä-mRNA an der 5' Ss ausbildet [206]. Diese U1 snRNA:5' Ss Basenpaarung ist allerdings nicht ausreichend, um die 5' Ss zu definieren [231]. Vielmehr konnte gezeigt werden, dass das U1 snRNP auch ohne komplementäre Basenpaarung an die 5' Ss der prä-mRNA bindet [53]. Offenbar assistiert das U1 snRNP spezifische Protein Yhc1 (U1C) bei der Selektion der 5' Ss, indem es Sequenz-spezifisch die 5' Ss noch vor der snRNA-mRNA Basenpaarung erkennt [54]. Da die nun vorliegende prä-mRNA, verglichen mit freier RNA, bevorzugt den Spleißprozess unterläuft, wird dieser Zustand als CC für *engl.* Commitment Complex² (human: E-Komplex) [28] bezeichnet, wobei in *S. cerevisiae* zwei CC unterschieden werden, die sich in ihrer Komposition aus weiteren non-snRNP Spleißfaktoren unterscheiden (Abschn. 1.4.2). In diesem Stadium ist das humane U2 snRNP wohl schon schwach, aber noch unabhängig von der Sequenz am Verzweigungspunkt mit der prä-mRNA assoziiert [46]. Die stabile Bindung von U2 snRNP führt zur Bildung des A Komplexes (*engl.* Pre-Spliceosome; *S. cerevisiae*: B Komplex), in dem Watson-Crick Basenpaare zwischen der U2 snRNA mit einer Sequenz am Verzweigungspunkt bestehen. Das für den nukleophilen Angriff der ersten Spleißreaktion erforderliche Adenosin bleibt aber ungepaart [191]. Der Prozess dorthin ist ATP-abhängig [135] und erfordert mehrere non-snRNP Spleißfaktoren, u.a. das heterodimere SR Protein U2AF (U2AF35/U2AF65) und den Proteinfaktor SF1, sowie den U2 RNA assoziierten Komplex SF3 [127]. Anschließend bindet der zuvor assemblierte Tri-snRNP-Komplex U4/U6.U5 ebenfalls an und es entsteht der B-Komplex (*S. cerevisiae*: A2-1-Komplex). Im Tri-snRNP-Komplex sind die RNAs von U4 und U6 über eine Basenpaarung miteinander verbunden, während U5 über Protein-Protein-Interaktion bindet.

In Two-Hybrid-Assays wurden Interaktionen des Tri-snRNP mit U1 snRNP Proteinen gefunden, die möglicherweise die Bindung des Tri-snRNP an die 5' Ss stabilisieren [27]. Snp1p interagiert mit den U5 Proteinen Prp44 und Prp8 [9], [69], [277] und auch mit einem Spleißfaktor, Exo84. Letzterer wiederum wechselwirkt mit Prp8 und den U1 snRNP Proteinen Prp39p und Prp40. Tatsächlich ist in Hefezell-extrakten ein - allerdings um einige U5 snRNP Proteine reduzierter - U1/U5 bi-snRNP gefunden worden [208], [82]. Das mit dem Zugang des Tri-snRNP entstandene komplette Spleißosom ist katalytisch noch inaktiv. Zunächst destabilisiert das U5 snRNP Protein Prp28, eine DExD/H-box Helikase, die U1 RNA:5' Ss Bindung, so dass das U1 snRNP schließlich dissoziiert. Bei einer veränderten U1 RNA oder eines mutierten YHC1 Gens [36], ist Prp28 nicht mehr

²CC: *engl.* Commitment Complex, *Commitment*-Komplex

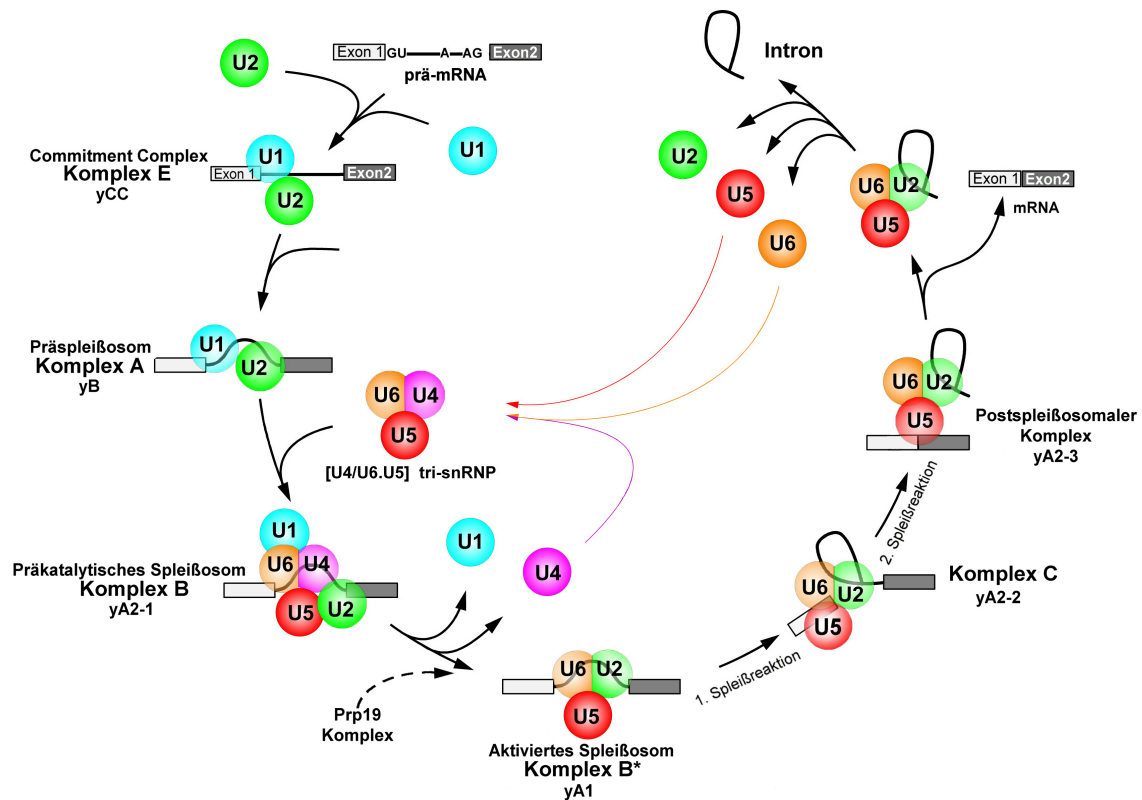


Abbildung 1.3: Der Spleißzyklus. Der generelle Zyklus der Assemblierung des Spleißosoms ist in niedrigen und höheren Eukaryonten hoch konserviert. Die aufgeführten Funktionszustände, die vom Spleißosom durchlaufen werden, sind dementsprechend über Speziesgrenzen allgemein gültig. Die Bezeichnungen der Komplexe unterscheiden sich jedoch in *S. cerevisiae* von denen im humanen System. Die mehr gebräuchlichen Namen der humanen Komplexe sind fett gedruckt hervorgehoben, darunter folgt die Bezeichnung für den Komplex in *S. cerevisiae* nach [39], [210], [60].

essentiell, eine Bestätigung, dass Yhc1p der Stabilisierung der U1 RNA:5' Ss - Basenpaarung dient [99].

Der verbleibende Komplex (B Δ U1-Komplex) unterliegt weiteren strukturellen Änderungen, die in die Dissoziation der U4-U6 RNA-Helices, der Freisetzung von U4 und schließlich der Basenpaarung der U6 RNA mit der 5' Ss münden. An der Trennung der U4/U6 RNAs ist das DExD/H-box Protein Prp44/Brr2) beteiligt [120], [134], [199]. Das U6 snRNP assoziierte Protein Prp24 wirkt möglicherweise bei diesem Prozess unterstützend, wie Studien mit Mutationen in U4 und U6 RNA vermuten lassen [232], [279]. Ebenso übernehmen die Tri-snRNP assoziierten Proteine Prp38p [293]) und Spp381 [157] wahrscheinlich eine Funktion bei der U4/U6 Trennung. Die U6 RNA bindet dann das Intron zwei Basen neben dem 5' GU Dipeptid. Die Region der U6 RNA, die mit der U4 RNA gepaart war, bildet mehrere kurze Helices, die mit U2 wechselwirken, um so den Verzweigungspunkt mit der 5' Ss einander näher zu bringen. Die zentrale Region der U6 RNA faltet sich in eine intramolekulare Stamm-und-loop Struktur (U6-ILS), die eine Bindungstasche für ein divalentes Metallion bildet, das für den ersten Katalyseschritt benötigt wird [219], [294].

Während die U1 und U4 snRNP den Komplex verlassen, bindet eine Gruppe Proteine, die in Hefe überwiegend im Prp19-Komplex organisiert sind. In humanen Zellen ist Prp19 mit wenigstens sechs anderen Proteinen in einem Komplex assembliert [3], ohne den der Spleißprozess in nuklearen Hela-Zellextrakten nach der Tri-snRNP Assoziation unterbrochen wird [159]. Der Prp19p Komplex übernimmt bei der Aktivierung des Spleißosoms offensichtlich eine wichtige Funktion [32]. Im katalytisch aktivierten Spleißosom B* (*S. cerevisiae*: A1), in dem schließlich die U6 RNA mit der U2 RNA gepaart vorliegt, erfolgt die erste Spleißreaktion. Dies mündet in den C-Komplex (*S. cerevisiae*: A2-2-Komplex), der den zweiten Schritt der Spleißreaktion katalysiert. Ist die Reaktion abgeschlossen, wird das mRNA-Produkt freigesetzt. Es verbleibt ein post-spleißosomaler Komplex, der das Intron in Lasso-Struktur und die snRNA U2,U5 und U6 beinhaltet und der alsbald aufgelöst wird.

Das Holospleißosom-Modell beschreibt die weitgehende Assemblierung der U snRNP zu einem Holoenzym vor Bindung an die prä-mRNA möglicherweise als penta-snRNP [253] oder als tetra-snRNP, nachdem das nur schwach assoziierte U1 snRNP zuvor gebunden hat. Die Interaktionen der verschiedenen U snRNP untereinander und mit der prä-mRNA sind jedoch in beiden Modellen hoch dynamisch und das Spleißosom unterliegt während des gesamten Spleißzyklus sowohl Änderungen in seiner Proteinkomposition, als auch in seiner Konformation. Da die spleißosomalen Subkomplexe im Gegensatz zum penta-snRNP bei relativ stringenten Bedingungen erhalten worden sind, stellen sie bestimmte stabilisierte Kernzustände des Spleißosoms dar. Die Vorstellung der schritt-

weisen Assemblierung des Spleißosoms bleibt ein adequates Modell, um die Vorgänge beim Spleißen auf molekularer Ebene zu erforschen [289].

1.3.2 U snRNP und weitere Spleißfaktoren

Die Subkomplexe des Spleißosoms, die U snRNP, bestehen aus einer an Uridinnukleotiden reichen snRNA (*engl.* small nuclear RNA) und einem gemeinsamen Set aus sieben Sm Proteinen [254]. Nur U6 RNA tritt assoziiert mit sieben Sm-ähnlichen, sog. Lsm Proteinen (*engl.* like-Sm) auf [160].

Neben der snRNA und dem Sm Komplex enthält jeder snRNP weitere spezifische Proteine, die in *S. cerevisiae* häufig mit 'Prp' (*engl.* precursor mRNA processing) bezeichnet werden, um auf ihre Funktion in der prä-mRNA Prozessierung hinzuweisen [210] (Abb. 1.4).

Neben den snRNP gibt es weitere Proteine, die als Spleißfaktoren zum Aufbau des Spleißosoms beitragen, aber nicht Bestandteil eines der snRNP sind. Sie werden zusammenfassend *engl.* non-snRNP proteins genannt. Viele dieser Spleißfaktoren sind an den strukturellen Umbildungen des Spleißosoms beteiligt, sowohl vor, als auch nach den zwei katalytischen Reaktionen, oder tragen zur Regulation der Assemblierung bei. Auch in dieser Gruppe treten Proteine mit der Bezeichnung 'Prp' auf.

Zusammenfassend konnten in massenspektroskopischen Analysen von spleißosomalen Subkomplexen (Review [109]), die mit affinitätschromatographischen Methoden aufgereinigt worden waren, insgesamt ungefähr 200 Proteine identifiziert werden, die, abhängig vom Funktionszustand, mit dem Spleißosom assoziierten [48], [96], [110], [158], [203], [299]. Neben den bereits bekannten snRNP und Spleißfaktoren konnten auch Proteine identifiziert werden, die an anderen Vorgängen der (i) Genexpression, vor allem der Transkription, beteiligt sind und (ii) Proteine mit noch unbekannter Funktion. Inwieweit die Mitglieder der zuletzt genannten beiden Gruppen spezifisch zum Spleißprozess beitragen, muss im einzelnen gezeigt werden.

snRNA

Die RNA Komponente ist je nach snRNP verschieden, in der Regel jedoch 60 bis 300 Nukleotide lang. Die Nummerierung der RNAs erfolgte in der Reihenfolge ihrer Entdeckung, und sie gaben dem entsprechenden U snRNP auch ihren Namen: U1, U2, U4, U5 und U6 [179] (U3 snRNA ist an der ribosomalen RNA Prozessierung beteiligt). Die U1, U2, U4 und U5 snRNA werden von der RNA Polymerase II transkribiert und erhalten an ihrem 5' Exon eine ungewöhnliche sogenannte m^3G -Kappe, ein 2,2,7-hypermethyliertes Trimethylguanosin [29]. U6 snRNA hingegen wird von der RNA Po-

<i>S.cerevisiae</i> snRNPs								humane snRNPs					
Proteine	U1	U2	U4	U5	U6	U4/U6	U4/U6.U5	Proteine	U1	U2	U5	U4/U6	U4/U6.U5
SmB	●	●	●	●		●	● ●	SmB/B'	●	●	●	●	● ●
SmD1	●	●	●	●		●	● ●	SmD1	●	●	●	●	● ●
SmD2	●	●	●	●		●	● ●	SmD2	●	●	●	●	● ●
SmD3	●	●	●	●		●	● ●	SmD3	●	●	●	●	● ●
SmE	●	●	●	●		●	● ●	SmE	●	●	●	●	● ●
SmF	●	●	●	●		●	● ●	SmF	●	●	●	●	● ●
SmG	●	●	●	●		●	● ●	SmG	●	●	●	●	● ●
Lsm2					●	●	●	Lsm2				●	●
Lsm3					●	●	●	Lsm3				●	●
Lsm4					●	●	●	Lsm4				●	●
Lsm5					●	●	●	Lsm5				●	●
Lsm6					●	●	●	Lsm6				●	●
Lsm7					●	●	●	Lsm7				●	●
Lsm8					●	●	●	Lsm8				●	●
Snp1	●							U1-70k	●				
Mud1	●							U1-A	●				
Yhc1	●							U1-C	●				
Luc7	●							LUC7-B1					
Nam8	●							TIA1/TIAR					
Prp39	●							FBP11					
Prp40	●												
Prp42	●												
Snu56	●												
Snu71	●							S164					
Lea1		●						U2A'		●			
Msl1		●						U2B''		●			
Prp21		●						SF3a120		●			
Prp11		●						SF3a66		●			
Prp9		●						SF3a60		●			
Hsh155		●						SF3b155		●			
Cus1		●						SF3b145		●			
Rse1		●						SF3b130		●			
Hsh49		●						SF3b49		●			
Snu17		●						SF3b14a		●			
Rds3		●						SF3b14b		●			
Ysf3		●						SF3b10		●			
Prp3			●			●	●	90k				●	●
Prp4			●			●	●	60k				●	●
Snu13			●			●	●	15.5k				●	●
Prp8				●			●	220k			●		●
Brr2				●			●	200k			●		●
Snu114				●			●	116k			●		●
Prp6				●			●	102k			●		●
Prp28				●			●	100k			●		●
Lin1				●			●	52k			●		●
Dib1				●			●	40k			●		●
								15k			●		●
Prp24					●								
Prp31						●	●	61k				●	●
Prp38						●	●	hPrp38				●	●
Snu23						●	●	hSnu23					
Snu66						●	●	110k					●
Spp381						●	●	TFIP11					
								20k				●	●
Sad1							●	65k					●
								27k					●

Abbildung 1.4: Die Proteinzusammensetzung der humanen und *S. cerevisiae* snRNP. In einer Reihe sind die jeweils homologen Proteine aus humanen und *S. cerevisiae* Zellen gegenübergestellt und es ist jeweils angegeben, von welchem snRNP diese Bestandteil sind. Der Farbcode dient der Unterscheidung der Proteine im U4/U6 und U4/U6.U5 Komplex; die zusätzlichen Proteine in diesen Komplexen sind grau dargestellt.

lymerase III synthetisiert und das 5' Exon ist mit einem γ -Monomethylguanodin, das am letzten (γ) Phosphat eine Methylgruppe tragt, modifiziert [241]. Ein monoklonaler Antikorper, H-20, erkennt die m^3G -Kappe als Epitop [20], so dass mit diesem Antikorper eine Methode etabliert werden konnte, uber Immunoaffinituschromatographie snRNP aufzureinigen [117] und zu charakterisieren. Die snRNA U4 und U6 zeigen hohe Sequenz-Komplementaritat und bilden leicht Basenpaare aus, so dass U4 und U6 in der Zelle hauptsachlich als di-snRNP Komplex assembliert auftreten [25], [97]. In der Hefe werden die snRNA snR19 (U1 snRNP), snR20 (U2 snRNP), snR14 (U4 snRNP), snR7 (U5 snRNP) und snR6 (U6 snRNP) genannt [39]. Die Strukturen der snRNA in Hefe unterscheiden sich zwar von den humanen, jedoch in bioinformatischen Sequenzvergleichsstudien der snRNA konnte gezeigt werden, dass es in phylogenetisch weit auseinanderliegenden Spezies kurze primare Sequenzabschnitte gibt, die stark uber Speziesgrenzen konserviert sind [91], (Abb. 1.5). Diese Sequenzabschnitte sind haufig

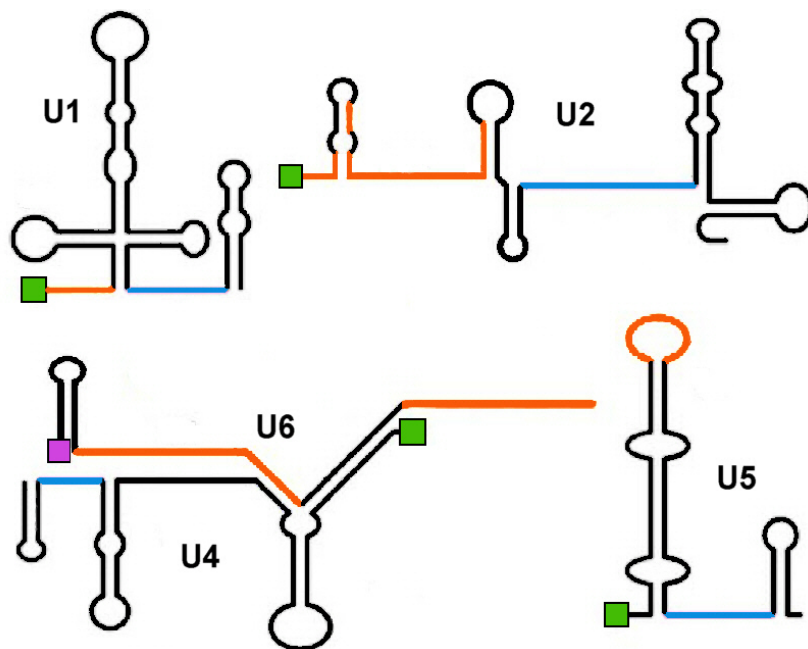


Abbildung 1.5: Konservierte Regionen in den snRNA. Schematisierte Sekundarstrukturen der humanen snRNA U1, U2, U5 und U4/U6 mit farblich hervorgehobenen Regionen, die uber Speziesgrenzen konserviert sind. Orange: Konservierte Sequenzabschnitte, die in RNA-RNA Interaktionen involviert sind. Blau: Bindungsstelle fur den Sm Ring. Rechteck grun: 5' Kappenstruktur 2,2,7-Trimethylguanodin. Rechteck pink: 5' Kappenstruktur γ -Monomethylguanodin.

bei Protein-RNA oder RNA-RNA Interaktionen beteiligt. Auch bestimmte Sekundarstrukturelemente (loops und stems) sind homolog, so dass die Analogien der snRNA als zweifelsfrei gelten konnen [91], und im Folgenden werden daher die humanen Bezeich-

nungen auch als Synonyme für die snRNA in *S. cerevisiae* verwendet.

Sm Proteine

Auf den snRNA, mit Ausnahme von U6, binden an die konservierte Sequenz RAU₃₋₆GR (R: beliebiges Nukleotid) sieben homologe Sm-Proteine, SmG, SmF, SmE, SmD1, SmD2, SmD3 und SmB/B' [91], wobei SmB und SmB' zwei alternativ gespleißte Produkte desselben Gens sind [201]. Humanes SmB und SmB' unterscheiden sich in 11 Aminosäureresten am Carboxyterminus [269]. Die Sm Proteine sind weit über Speziesgrenzen konserviert. Anti-Sm Antikörper, die in Sera von Patienten mit der Autoimmunerkrankung SLE (systemischer Lupus erythematoses) diagnostiziert wurden, wurden benutzt, um spleißosomale snRNP von verschiedenen Spezies, darunter Invertebraten (*Drosophila*), Hefe und Pflanzen zu präzipitieren, wenn auch in manchen Fällen ineffizient [228]. Gemeinsame Eigenschaften auf Sekundärstrukturebene in einer als Sm Domäne bezeichneten Region könnte die Kreuzreaktivität der Sm-Antikörper begründen [228]. Auf Primärstrukturebene teilen die Sm Proteine zwei evolutiv hoch konservierte Sequenzabschnitte, 32 und 14 Aminosäuren lang, die Sm-Motiv 1 und 2 genannt werden [101]. Die Sm-Motive scheinen bei Sm Protein-Protein Interaktionen untereinander wichtig zu sein [101]. Drei der Sm Proteine, SmD1, B und D3, haben lange positiv geladene C-Termini, die vermutlich bei der Bindung von U1 snRNP an das prä-mRNA Substrat im CC (Abschn. 1.3.1 & 1.4.2) mit der prä-mRNA interagieren und so zur Stabilität des Komplexes beitragen [297]. Die Sm Proteine bilden gemeinsam eine heptamere ringförmige Struktur, wobei die RNA am oder durch das zentrale Loch geführt wird und die Sm Proteine in der Reihenfolge G-E-F-D2-D1-B-D3 angeordnet sind [112], [288].

Auf der U6 snRNA hingegen binden sieben Sm-ähnliche (*engl.* Like-Sm) Lsm Proteine (Lsm 2-8) an eine uridinreiche Sequenz am 3' Exon der RNA [278] und bilden analog zu den Sm-Proteinen heteromere Komplexe [160], wahrscheinlich gleichfalls in Form einer Ringstruktur, wie EM und kristallographische Daten zeigten [2], [166]. Homologe der Sm und Lsm Proteine treten in allen Spezies auf, die Spleißosom-vermittelt prä-mRNA prozessieren. Die Sm/Lsm Proteine, die auch bei höherer Ionenstärke mit der snRNA assoziiert bleiben [201], [105], [107], bilden mit der snRNA den Kern der U snRNP (*engl.* core structure).

Prp19 Komplex

Eine Gruppe von Spleißfaktoren bilden die Proteine, die in *S. cerevisiae* mit dem Spleißfaktor Prp19p assoziieren. Dieser Komplex wurde erstmalig in *S. cerevisiae* beschrieben [262] und wird in diesem eukaryontischen Einzeller auch als NTC³ bezeichnet. Der

³*engl.* Nineteen Complex

Komplex besteht in *S. cerevisiae* neben Prp19p aus zumindest 10 weiteren Proteinen [175], von denen 9 bereits identifiziert worden sind: Ntc90 (Syf1p), Ntc85 (Cef1p), Ntc77 (Clf1p), Ntc50 (Prp46p), Ntc40 (Cwc2p), Ntc31 (Syf2p), Ntc30 (Isy1p), Ntc25 (Snt309p) und Ntc20p [35], [33], [266], [51], [34]. In *S. pombe* konnte ebenfalls ein Multiprotein Komplex isoliert werden, der in der Proteinkomposition mit dem NTC aus *S. cerevisiae* vergleichbar ist [175]. Für Prp19p, das *in vitro* und *in vivo* zu einem Tetramer oligomerisiert [176], konnte gezeigt werden, dass es kein Bestandteil eines der snRNP ist [40]. Es assoziiert mit dem Spleißosom, nachdem U2 snRNP gebunden hat, wahrscheinlich gleichzeitig oder kurz nach der Dissoziation von U4 snRNP und bevor das Spleißosom die erste Katalysereaktion vollzieht [260], [261]. Aus humanen Zellextrakten konnte auch ein Proteinkomplex isoliert werden, der ein dem Prp19p aus *S. cerevisiae* homologes Protein enthält [159], [3]. Unter den ungefähr 30 kopräzipitierten Proteinen bleiben jedoch die meisten vorerst uncharakterisiert. Bisher konnten neben CDC5 (Cef1p) nur vier humane Proteine identifiziert werden, die im *S. pombe* NTC Komplex Homologe haben [175].

DExH/D-box Proteine

Eine weitere wichtige Gruppe von non-snRNP bilden die DExH/D-box RNA Helikasen, eine vom Bakterium zum Menschen hoch konservierte Proteinsuperfamilie. Sie teilen sieben bis acht charakteristische Sequenzmotive [257], nach denen sie in Untergruppen klassifiziert werden. Das charakteristische Motiv II nimmt meistens in Unilettercode die Aminosäureprimärsquenz DEAD, DEAH oder DExH an, wonach sie zunächst in drei Subgruppen untergliedert werden. Die DExH/D-box Proteine sind an allen möglichen Prozessen, in denen RNA Moleküle benötigt werden, beteiligt, unter anderem auch Transkription, RNA Editierung und Translation. Im Spleißzyklus (Abschn. 1.3.1) werden sie für die zahlreichen Restrukturierungen der snRNA vor und nach den zwei katalytischen Reaktionen gebraucht [247]. Wenige dieser DExH/D-box Proteinfaktoren sind Bestandteile eines snRNP. Für diesen Fall sind Brr2p und Prp28p Beispiele. Einige der DExH/D-box Proteine zeigen RNA-abhängige NTPase Aktivität und/oder entwinden doppelsträngige RNA *in vitro* [147], [227], [282], [286], wie beispielsweise das zu Brr2p homologe humane Protein U5-200kD [134]. Da jedoch die zelluläre RNA selten als nackte doppelsträngige Helix vorliegt, sondern meistens im Komplex mit Proteinen, stützen andere Ergebnisse die Vorstellung, dass DExH/D-box Proteine Bindungen lösen, indem sie tertiäre Strukturelemente, die für RNA-Protein Bindungen nötig sind, destabilisieren. Ein möglicher Mechanismus für die Zerschlagung der strukturellen Umgebung könnte die ATP-abhängige RNA induzierte Konformationsänderung des DExH/D-box Proteins sein, so wie es für das DEAD-box Protein und Translations-Initiationsfaktor, eIF4A, beschrieben wurde [152]. Allgemeiner formuliert könnte ein Protein, das kritisch

für die Stabilität eines größeren Teils eines Komplexes ist, Substrat für ein DExH/D-box Protein sein, und dann durch DExH/D-box-Protein-vermittelte Konformationsänderung die Integrität dieses gesamten Teils zerfallen lassen [104]. So löst wahrscheinlich Prp28p das U1 snRNP von der prä-mRNA über die Destabilisierung der Interaktion von U1-C mit der 5' Ss [36].

SR Proteine

Zu einer Klasse regulatorischer Spleißfaktoren gehören die SR Proteine, die in Metazoa in neun Familien unterteilt werden, von denen sechs Familien zwei oder mehr Mitglieder haben [202]. In *S. cerevisiae* allerdings gibt es keine kanonischen SR Proteine, in *S. pombe* nur zwei, und in anderen einzelligen Organismen treten ebenfalls nur ein oder zwei SR Gene auf [11]. Die Diversität der SR Proteine scheint sich mit multizellulären Organismen entwickelt zu haben. In *S. cerevisiae* enthält jedes Intron jegliche Informationen, die für korrektes Spleißen notwendig sind (*engl.* intron definition) [90], [75]. In Metazoa hingegen sind die Introns nicht mehr ausreichend gut definiert (Abschn. 1.2.2), und es werden auch Exon Sequenzen für die Definition der korrekten Spleißstellen benötigt (*engl.* exon definition). Dabei erleichtern SR Proteine die Erkennung dieser weniger gut definierten Introns, typischerweise durch Bindung an die der Spleißstelle benachbarten Exon Sequenzen. Vermutlich übernehmen sie auch eine wichtige Funktion in der Regulation des alternativen Spleißprozesses [202], [7]. Die SR Proteine tragen am N-Terminus ein oder zwei RNA-bindende RRM⁴ Domänen und am C-Terminus eine RS-Domäne, die reich an Arginine und Serine Aminosäuren ist und die Protein-Protein Interaktionen vermittelt [89]. Die RS-Domänen sind an den Serine-Aminosäureresten stark phosphoryliert und die Veränderung des Phosphorylierungsgrades kann die Protein-Protein Interaktion, und so die Funktion des SR Proteins, beeinflussen [292], [285].

In *S. cerevisiae* gibt es drei Proteine, Npl3, Hrb1 und Gbp2, die die grundlegende Struktur von SR Proteinen aufweisen [73], [94] und als *engl.* SR-like bezeichnet werden. Npl3 und Gbp2 können von einer Serine-Kinase, Sky1, phosphoryliert werden, die hohe strukturelle Gemeinsamkeiten mit der SR Protein Kinase 1 (SRPK1) in Säugetieren aufweist [73]. Npl3, Hrb1 und Gbp2 wurden ursprünglich in ihrer Funktion als RNA-Export Faktoren identifiziert. Für Npl3 konnte nun auch gezeigt werden, dass es für effizienteres Spleißen von prä-mRNAs in *S. cerevisiae* benötigt wird [128]. Interessanterweise konnte Npl3p bereits früher mit dem U1 snRNP in *S. cerevisiae* ko-isoliert werden, wies allerdings nur schwache, Salz-sensitive Assoziation auf [79]. Auch wenn es in *S. cerevisiae* keine kanonischen SR Proteine gibt, so könnten jedoch einige der zusätzlichen U1 snRNP

⁴RRM: *engl.* RNA recognition motif, RNA Erkennungsmotiv

spezifischen Proteine SR-Protein-ähnliche Funktionen tragen (Abschn. 1.4).

1.4 Das U1 snRNP in *S. cerevisiae*

1.4.1 Komposition des U1 snRNP in *S. cerevisiae*

Das U1 snRNP aus *S. cerevisiae* besteht neben der snR19-RNA [129], [130], [239] aus insgesamt siebzehn Proteinen (Abb. 1.6), und zwar aus den sieben konservierten Sm Proteinen (Abschn. 1.3.2) sowie zehn U1 snRNP spezifischen Proteinen [169], [79], [64], [205]. Drei der spezifischen Proteine, Snp1, Mud1 und Yhc1, sind den humanen U1 snRNP spezifischen Proteinen U1-70K [242], U1A [144] und U1-C homolog [256]. Desweiteren sind bisher zu vier der weiteren Proteine im U1 snRNP aus *S. cerevisiae* Homologe im humanen System gefunden worden, und zwar für Nam8p (TIA1/TIAR), Snu71p (S164), Prp40p (FBP11) und Luc7p (LUC7B1). Insgesamt hat die Proteinkomponente des U1 snRNP in *S. cerevisiae* ein Molekulargewicht von 610,5 *kDa* im Gegensatz zum humanen U1 snRNP mit 194,3 *kDa* (Abb. 1.6).

U1 snRNA

Die U1 snRNA in *S. cerevisiae* (snR19) ist mit 567 Nukleotiden 3,5 mal länger als die humane U1 snRNA mit 164 Nukleotide (Abb. 1.7). Die Sequenzhomologien zur humanen U1 RNA beschränken sich auf funktional wichtige Sequenzabschnitte wie der 5' Ss Erkennungssequenz, der Bindungssequenz für den Sm Ring und der Loopregion des RNA Stamms II (loop A) [91], [240], der für die Bindung der humanen U1 snRNP Proteine U1-A und U1-70k nötig ist [95], [183], [197]. Trotz der geringen Primärsequenzhomologien weisen die ersten 50 Nukleotide der beiden snRNA Sekundärstrukturhomologien auf [129]. Die snRNA-Stämme V, VII und VIII, sowie ein kurzer Abschnitt von Stamm VI der U1-RNA aus *S. cerevisiae* scheinen konserviert in Hefe Spezies mit vergleichbar langen U1 snRNA zu sein, da dieser Bereich Homologien zu den U1 Molekülen in *S. uvarum* und *Klyveromyces lactis* aufweist [130], [145]. Die Stämme IX, X und die Enden der Stämme III und VI hingegen sind auch unter diesen drei Hefe U1 RNAs nicht konserviert und die Strukturvorhersage ist spekulativ aufgrund fehlender Daten. Weite Bereiche der U1 snRNA in *S. cerevisiae* scheinen funktional nicht essentiell zu sein. Eine Deletion der Nukleotide 192 – 507 jedenfalls hatte kaum negativen Einfluss auf die Viabilität der Hefezellen [240]. Erst eine Deletion von 60 % der U1 Nukleotide zwischen den Positionen 184 und 540 führte zu einer fünffach verlangsamten Verdopplungsrate verglichen mit dem Wildtyp (Abb. 1.8). Deletionen, die am 5' Ende unterhalb

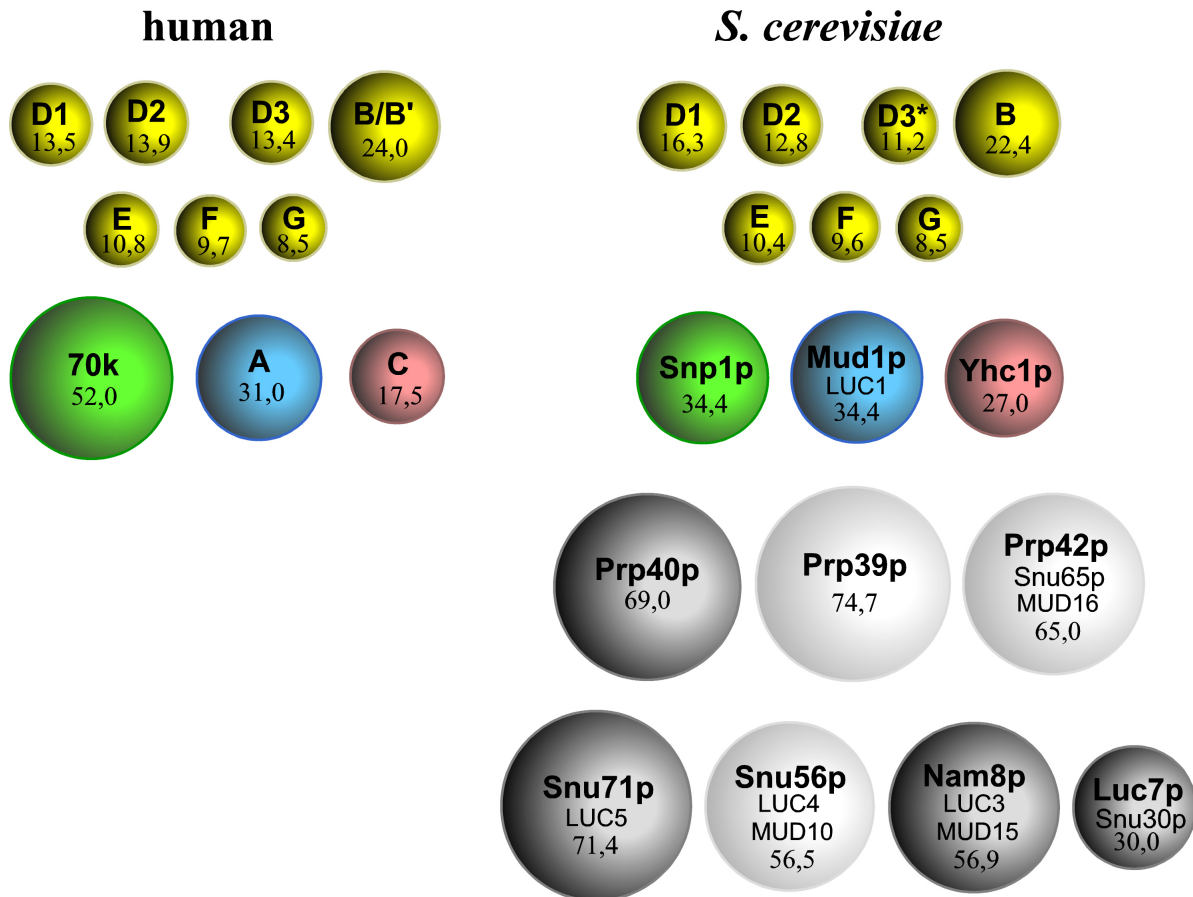


Abbildung 1.6: Übersicht der Proteine des humanen und *S. cerevisiae* U1 snRNP. Die homologen Proteine sind jeweils einander gegenübergestellt und farbig entsprechend gekennzeichnet. Die Größe der Kreise ist ins Verhältnis zur Molekularmasse des jeweils dargestellten Proteins gesetzt. **Gelb:** Sm Proteine; B' bezieht sich auf die alternativ gespleißte Variante des SmB im humanen System. '*' am D3: LUC6. **Grün, blau und rot:** Die drei homologen U1 snRNP spezifischen Proteine. **Grau:** U1 snRNP Proteine in *S. cerevisiae*, für die im humanen System homologe Proteine unter den Spleißfaktoren gefunden worden sind (vgl. Abb. 1.4). **Weiss:** U1 snRNP Proteine in *S. cerevisiae*, für die keine homologen Proteine im humanen System beschrieben worden sind.

Für die Proteine in *S. cerevisiae* sind weitere Proteinnamen angegeben. Die Bezeichnungen LUC ('lethal unless CBC is produced') und MUD ('mutant U die') weisen auf genetische Screens hin. In diesen wirkte die Deletion des Genlocus des entsprechenden Proteins in einem Hefestamm ohne CBC Komplex (LUC-Screening) oder mit mutierter U1 RNA (MUD-Screening) letal.

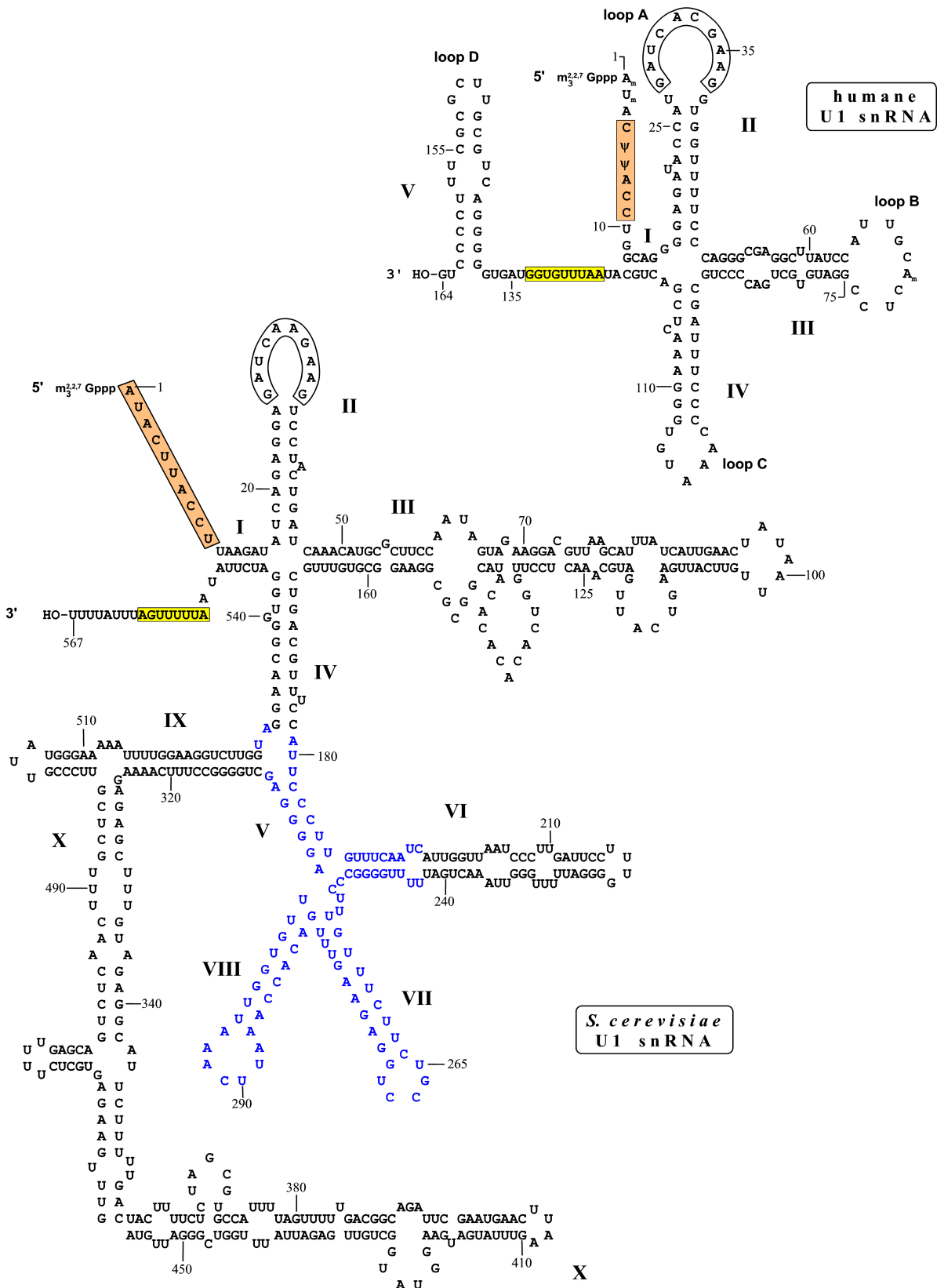


Abbildung 1.7: Die Primärsequenzen und vorgeschlagenen Sekundärstrukturen der humanen und *S. cerevisiae* U1 snRNA. Die Sekundärstrukturvorschläge sind [130], [240] (*S. cerevisiae*) und [131] (human) entnommen. Box gelb: Bindungsstelle für den Sm Ring. Box orange: An der prä-mRNA Basenpaarung beteiligte Nukleotide. Box farblos: hohe Sequenzkonservierung. blau: In Hefespezies konservierte Region. Römische Zahlen: Nummerierung der Sekundärstrukturstämme. $m_3^{2,2,7}G$: 2,2,7-Trimethylguanodin. Ψ : Pseudouridin. Index m: methylierte Form.

von Position 170 begannen, waren letal oder zeigten einen erheblich Wachstumsdefekt, was mit der Zerstörung von Bindungsdomänen für die Proteine U1-70K und U1-A erklärt werden kann. Die minimale essentielle U1 RNA Sequenz mit der Deletion 185 – 539 kann eine Sekundärstruktur annehmen, die der U1 RNA in höheren Eukaryonten ähnlich ist [240], [130].

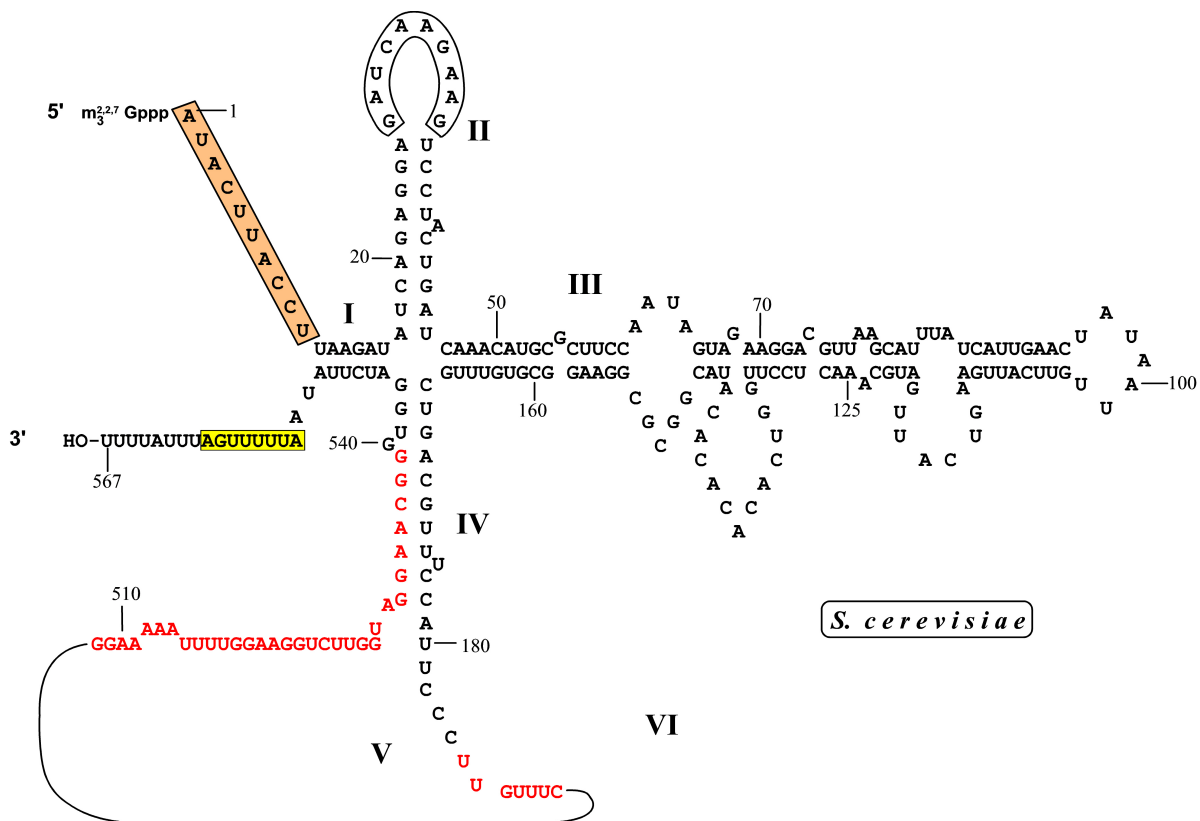


Abbildung 1.8: Lebensfähige Deletionsmutanten der *S. cerevisiae* U1 snRNA. Bei einer Deletion der Nukleotide $\Delta 192-507$ der U1 snRNA ist die Verdopplungsrate der Hefezellen bei 30 °C kaum verringert (1,1-fach), verglichen mit dem Wildtyp. Wird die U1 snRNA um weitere Nukleotide (rot) $\Delta 185-539$ verkürzt, leben die Zellen zwar noch, zeigen aber einen erheblichen Wachstumsdefekt gegenüber dem Wildtyp (Verdopplungsrate 5-fach reduziert bei 30 °C) [240]. Die deletierte U1 snRNA zeigt Ähnlichkeiten mit der Sekundärstruktur der humanen U1 snRNA.

Snp1p, Mud1p und Yhc1

Die drei Proteine, Snp1p, Mud1p und Yhc1p sind zwar mit den humanen U1 snRNP Proteinen U1-70K [242], U1A [144] und U1-C homolog [256], doch gibt es aufgrund der evolutiv weit auseinander liegenden Spezies auch hier beachtenswerte Unterschiede.

Snp1p, dessen Primärsequenz mit 300 Aminosäuren (AS) deutlich kürzer ist als die seines humanen Homologs U1-70k mit 437 AS, ist mit letzterem zu 30% identisch. Be-

reiche, die sich besonders ähneln, umfassen ein RNA-Erkennungsmotiv, RRM, und eine anschließende glycinreiche Sequenz. Hingegen sind die C-Termini beider Proteine sehr unterschiedlich in Länge und Ladung. Während der C-Terminus von Snp1p 95 AS lang und nur moderat geladen ist, ist jener von U1-70k 238 Aminosäuren lang und trägt zwei argininreiche Sequenzabschnitte [242]. Die Deletion des SNP1 Gens wirkt zwar nicht letal [102], allerdings zeigen die Δ SNU1 Hefestämme stark verlangsamtes Wachstum und sind temperatursensitiv. Einige prä-mRNA Transkripte akkumulieren in diesen Zellen, oder aber die Menge an reifer mRNA ist von bestimmten Transkripten reduziert [102]. Dieser Phänotyp der Deletionsmutanten konnte nur mit der Expression von künstlichen Snp1p Varianten ausgeglichen werden, die auch die N-terminale Domäne enthielten. Immunpräzipitationsexperimente zeigten, dass die N-terminale Domäne mit der U1 RNA assoziiert [102]. Für die direkte und spezifische Bindung von Snp1p an den Stamm II der U1 RNA von *S. cerevisiae* [113], scheinen allerdings die RRM Domäne und die glycinreichen Sequenzen nötig zu sein [102]. Da für das U1-70k Protein gezeigt werden konnte, dass es an den Loop A (Stamm II in Abb. 1.7) der humanen U1 RNA bindet, und dass die N-terminalen AS 1 – 97 mit dem U1 snRNP Kern Komplex (U1 snRNP ohne die 3 spezifischen Proteine) assoziieren [168], ist das Bindungsverhalten von Snp1p mit dem von U1-70k vergleichbar.

Mud1p wurde in einem genetischen Screening entdeckt, bei dem die Einführung von Mutationen in jeweils einem bestimmten Gen auf die Letalität eines ansonsten lebensfähigen Hefestamms, der allerdings ausschließlich eine veränderte U1 snRNA exprimiert, getestet wurde. Die veränderte U1 RNA des verwendeten Hefestamms trug eine Deletion in den Stämmen V bis X und eine Punktmutation im konservierten Loop A (mud: *engl. mutant-U-die*, vgl. Abb. 1.6, [144]). Mit Computer-basierten Methoden konnte die Homologie von Mud1p zu U1-A geklärt werden [144]. Beide Proteine enthalten zwei RNA-Bindungsdomänen (RBD), getrennt durch eine vergleichbar lange Aminosäuresequenz. Die C-terminale RBD weist insgesamt höhere Sequenzidentität auf als die N-terminale. Allerdings sind in der N-terminalen RBD interessanterweise gerade jene Aminosäuren über beide Organismen stark konserviert, für die im Falle des humanen U1-A gezeigt werden konnte, dass sie den Kontakt zur snRNA herstellen [223], [224], [106], [178]. Die N-terminale RBD von U1-A bindet an den Loop B der humanen U1 snRNA [222], und auch Mud1p bindet an eine organisatorisch vergleichbare Stelle, und zwar an den proximalen Teil des Stamms III der U1 snRNA in *S. cerevisiae* [255]. Zwar ist Mud1p in *S. cerevisiae* überraschenderweise im Wildtyp nicht essentiell, es konnte aber in Reporter-genessays gezeigt werden, dass die Spleißeffizienz von artifiziellen Intronsequenzen mit Mud1p erheblich erhöht ist im Vergleich zum Experiment vor einem Mud1p deletierten genetischen Hintergrund. Dabei übernimmt die C-terminalen RBD von Mud1p

wohl eine zentrale Funktion [255].

Das *S. cerevisiae* Protein Yhc1p ist mit dem humanen U1-C in der N-terminalen Region zu ungefähr 50 % identisch, wobei die hoch konservierte Sequenz auf die ersten 38 Aminosäuren beschränkt ist [256]. In diesem Teil liegt ein Zn-Finger-ähnliches Motiv mit den für dieses Motiv kritischen AS Cystein und Histidin. Der C-terminale Teil des Proteins divergiert stark zwischen den Spezies und ist im humanen U1-C reich an Prolin und in *S. cerevisiae* reich an Lysin, Serin und Arginin. Eine vollständige Deletion von Yhc1p wirkt letal, so dass Yhc1p essentiell ist. Eine konditionelle Deletion von Yhc1p in Abhängigkeit der Galaktosekonzentration im Medium zeigte erhebliche Beeinträchtigung im Spleißen *in vivo* und *in vitro*. In Extrakten dieses Hefestamms wurden nur stark reduzierte Mengen an CC und Spleißosomen detektiert [256]. Da die Assoziation von Mud1p und Snp1p mit der U1 snRNA durch die konditionelle Deletion von Yhc1p in diesem Hefestamm nicht beeinträchtigt wurde, scheint die weitgehende Assemblierung und Stabilität des U1 snRNP unabhängig von Yhc1p gewährleistet zu sein [256]. In Abwesenheit von Yhc1p ist das 5' Ende der U1 snRNA mit der 5' Ss Erkennungssequenz hypersensitiv für partiellen RNase I Verdau [256]. Mittlerweile konnte auch die Funktion von Yhc1p bei der Erkennung der 5' Ss auf der prä-mRNA beschrieben werden [53], [54].

Prp39p und Prp42p

Prp39p und Prp42p gehören zu jenen spezifischen Proteinen im *S. cerevisiae* U1 snRNP, für die bisher kein homologes im humanen System gefunden worden ist, wodurch gewisse Hinweise auf die Funktion dieser beiden Proteine fehlen. Beide Proteine enthalten mehrere Kopien eines Sequenzmotivs, das *engl.* tetratricopeptide repeat motif (TPR-Motiv) genannt wird [162]. Dieses Motiv besteht aus zwei aufeinanderfolgenden 34 AS langen Sequenzen, die ein Paar von antiparallelen α -Helices bilden in der Anordnung Helix-Turn-Helix, wie mit der Kristallstruktur der TPR-Motiv Domäne der Phosphatase PP5 gezeigt werden konnte [45]. Zuerst war das TPR-Motiv in den Genen *cdc16*, *cdc23* und *cdc27* gefunden worden. Diese kodieren Zellzyklus-Proteine, die Bestandteil des APC⁵ sind (Abb. 2.6), [103], [238], [121]. Es zeigte sich, dass das Strukturmotiv in fast allen Organismen, auch Prokaryonten, auftritt und in Proteinen, die an den verschiedensten Zellprozessen beteiligt sind [44]. Die TPR-Motive sind in Proteinen meist in Folgen von zumindest 3 bis 16 α -Helixpaaren angeordnet, können aber auch über die gesamte Proteinsequenz verstreut liegen. Die TPR Motive vermitteln Protein-Protein Interaktionen und Proteine, die mehrere Kopien dieses Motivs tragen, könnten als Gerüst Proteine (*engl.* scaffolding proteins) der Assemblierung von Komplexen mit mehreren Untereinheiten dienen [193]. In *S. cerevisiae* sind beide Proteine essentiell für die Funk-

⁵APC: *engl.* Anaphase Promoting Complex

tion des U1 snRNP und den Spleißprozess, da die jeweilige Deletion eines der Proteine die Bildung der CC1/CC2 Komplexe unterband [162], [149]. Bei der Untersuchung von Prp39p wurde allerdings eine Subpopulation des U1 snRNP gefunden, der kein Prp39p enthielt [149]. Möglicherweise stellte diese U1 snRNP Subpopulation ein Intermediat bei der Biogenese des U1 snRNP dar, da keine anderen Indizien darauf hinwiesen, dass es funktionell unterschiedliche U1 snRNP gibt, von denen in der einen Form auf Prp39p verzichtet werden könnte. Letzteres kann dennoch nicht vollends ausgeschlossen werden.

Prp40p

Prp40p enthält zwei N-terminale WW Domänen gefolgt von vier FF Domänen und gehört damit zu einer Unterklasse von FF Domänen tragenden Proteinen. Eine FF Domäne umfasst ungefähr 50 AS und enthält zwei konservierte namensgebende Phenylalanine. Hingegen ist die circa 40 AS lange Sequenz der WW Domäne, die als Sekundärstruktur ein 3-strängiges (*engl.* triple-stranded) β -Faltblatt bildet, durch zwei Tryptophane in definiertem Abstand charakterisiert. In der WW/FF- Proteinfamilie sind drei Mitglieder bekannt, die die CDT Domäne der RNA-Polymerase II erkennen (vgl. Abschn. 1.1), und zwar der humane Transkriptionselongationsfaktor CA150 [31], der humane Spleißfaktor FBP11 [4] und sein homologes Protein in *S. cerevisiae* Prp40p [165], wobei für CA150 in [31] gezeigt werden konnte, dass die FF Domänen den Kontakt zur phosphorylierten CDT Domäne vermitteln. Prp40p scheint in *S. cerevisiae* als U1 snRNP Komponente [114] die Funktion eines Verbindungsglieds zwischen Transkription und RNA Spleißen zu haben und liefert eine Erklärung für den kotranskriptionellen Erkennungsprozess der 5' Ss durch das U1 snRNP (Abschn. 1.4.2). In einer Studie [58] konnte gezeigt werden, dass Prp40p mit Hilfe seiner FF Domänen Protein-Protein Wechselwirkung mit Snu71p und Luc7p vermittelt, und es wird vermutet, dass Prp40p mit Snu71p und Luc7p innerhalb des U1 snRNP einen Subkomplex bildet. Folgende weitere Beobachtung stützt diese Vorstellung: In Zellextrakten aus einem temperatursensitiven *luc7* Hefestamm (*luc7-1*) assoziierte das mutierte Luc7p nicht mehr mit dem U1 snRNP oder mit U1 snRNP Proteinen unabhängig davon, ob die Zellen zuvor bei tolerierter oder erhöhter Temperatur gewachsen waren. Gleichzeitig waren auch Snu71p, Prp40p und Prp42p nicht mehr stabil an das U1 snRNP gebunden [64].

Nam8p

Das Nam8 Gen, das ein Protein von 523 Aminosäuren Länge kodiert, war ursprünglich entdeckt worden, weil es im Falle der Überexpression fähig ist, die mitochondrialen Spleißdefizite in einem Hefestamm, der Mutationen im zweiten Intron des Cytochrom b Gens und im ersten Intron der ersten Untereinheit der Cytochrom Oxidase trägt, zu

supprimieren [13]. Eine Nam8p Deletion in *S. cerevisiae* ist nicht letal und Nam8p ist nicht essentiell für das vegetative Wachstum der Zelle [57]. Es zeigte sich allerdings, dass Nam8p essentiell während der Meiose ist. Es wird dort zum effizienten Spleißen des Meiose-spezifischen prä-mRNA-Transkripts MER2, das keine kanonische 5' Ss enthält, benötigt [174], [167]. Möglicherweise wird Nam8p auch noch für weitere Meiose-spezifische Transkripte benötigt [167].

Sequenzanalyse zeigte, dass Nam8p reich an den Aminosäuren Serin und Threonin ist und drei RNA Bindungsdomänen (RBD⁶, RRM) trägt [79], [57]. Letzteres legte die Vermutung nahe, dass Nam8p an die prä-mRNA bindet. Es konnte gezeigt werden, dass Nam8p eine Komponente sowohl des CC1, also auch des CC2 ist [194]. In der gleichen Studie wurde in Crosslink-Versuchen festgestellt, dass Nam8p mit der prä-mRNA nahe der 5' Ss an nicht-konservierten Intronsequenzen bindet, was in [296] bestätigt wurde. In anderen Experimenten führten uridinreiche Intronsequenzen zu einer erhöhten Spleißaktivität im Wildtyp verglichen mit einem Δ Nam8 Stamm [194]. Diese Ergebnisse liefern eine Unterstützung für die These, dass in Hefe nicht-konservierte Sequenzen zur Spleiß-Effizienz beitragen können [189].

Außerdem sorgt Nam8p *in vivo* wahrscheinlich für das effiziente Spleißen von CBC freien Transkripten, wie in einem Versuch mit dem Reportergen prä-tRNA-Tyr, das von der RNA Polymerase III transkribiert wird und somit *per se* CBC frei ist, gezeigt werden konnte [194].

Das Protein Nam8p ist mit dem humanen Spleißfaktor TIA-1, der ebenfalls drei RNA Bindungsdomänen und eine C-terminale Q-reiche Domäne trägt, homolog. Auch ist von TIA-1 gleichfalls bekannt, dass es an uridinreiche Sequenzen stromabwärts der 5' Ss bindet [49], [62], [136]. Es konnte in affinitätschromatographischen Experimenten gezeigt werden, dass TIA-1 mit dem U1 snRNP assoziiert und im Komplex scheinbar direkt mit dem U1-C Protein interagiert [63]. Während TIA-1 mit den RRM 2 und 3 die prä-mRNA zu erkennen vermag [63], könnten die Q-reiche Domäne am C-Terminus und die RRM 1 Domäne, die stärker negativ geladenen Aminosäurereste trägt, das U1 snRNP via U1-C rekrutieren. In diesem Szenario wirkt TIA-1 regulierend auf den Spleißprozess.

Luc7p

Luc7p wurde als letztes spezifisches Protein dem U1 snRNP zugeordnet, nachdem das U1 snRNP mit der TAP-tag Proteinpurifikationsmethode [205] aufgereinigt worden war. Das 261 Aminosäuren lange Polypeptid enthält zwei Zn-Finger Domänen, eine vom Typ C3H am N-Terminus und die andere vom Typ C2H2 am C-Terminus. Datenbankanalysen von Proteinsequenzen ergab, dass Luc7p homologe Proteine in höheren Eukaryonten

⁶RBD: *engl.* RNA binding domain, RNA Bindungsdomäne oder RRM: *engl.* RNA recognition motif

hat. Scheinbar wurde LUC7 zu einem frühen Zeitpunkt während der Evolution dupliziert, was in höheren Eukaryonten zur Bildung zweier Subfamilien Luc7A und Luc7B führte. Die Aminosäuresequenzen der Zn-Finger Domänen sind in der gesamten Proteinfamilie besonders stark konserviert. In Metazoa tragen interessanterweise alle Mitglieder der Luc7 Familie Extensionen am C-Terminus mit mehreren Arginin-Serin (RS) oder Arginin-Glutamat (RE) Wiederholungen, charakteristisch für eine große Anzahl von Spleißfaktoren in Metazoa [170]. Im humanen System gibt es drei homologe Proteine, hLuc7A, hLuc7B1 und hLuc7B2. Antikörper gegen hLuc7A vermögen aus humanen HeLa Zellextrakten auch das U1 snRNP zu präzipitieren [196], was ein Indiz für die ebenfalls auftretende stabile Assoziation dieses Spleißfaktors mit dem U1 snRNP in höheren Eukaryonten ist.

Das Protein Luc7p ist nicht nur wichtig für die Integrität des U1 snRNP, wie das Assoziationsverhalten einer mutierten Luc7p Variante mit dem U1 snRNP in einem temperatursensitiven Hefestamm (*luc7p-1*) zeigte (siehe unter Prp40p), sondern auch für die Bildung der Komplexe CC1 und CC2 [64]. Ursprünglich wurde LUC7 nämlich in einem genetischen Screen entdeckt, der nach Genen suchte, deren Mutation in Abwesenheit des CBC zur Letalität führte (LUC- lethal unless CBC is produced) [65]. Diese genetische Interaktion konnte auf Proteinebene mit einem Reportergenessay verifiziert werden. Dieser Essay zeigte nämlich, dass in Anwesenheit von Wildtyp Luc7p die CBC proximale 5' Ss bevorzugt gespleißt wurde [64]. In anderen Versuchen, Crosslink-Studien, konnte gezeigt werden, dass Luc7p die prä-mRNA mit dem N-terminalen Zn-finger Motiv bindet, wobei die C-terminale Zn-Finger Domäne zur Stabilität beitragen könnte [196]. In der Hefe ist das Gen LUC7 essentiell für das Wachstum [64], [150].

Snu71p und Snu56p

Snu71 ist ein Protein von 620 AS Länge. Die Sequenz enthält 19 Arginin-Serin, Arginin-Aspartat oder Arginin-Glutamat Dipeptide, die allerdings verstreut über die gesamte Sequenz auftreten [79]. Es fehlt somit das Charakteristikum einer Arginin-Serin reichen Domäne (RS-Domäne), wie sie für eine Superfamilie von Spleißfaktoren in Metazoa beschrieben wurde [70].

Snu56p ist ein Protein von 492 AS und trägt ebenfalls keine besonderen Sequenzmotive. Zwischen den AS 310 und 321 treten acht Serinreste gehäuft auf [79].

Sowohl Snu71p, für das es ein humanes homologes Protein, S164, gibt, als auch Snu56p sind keine Proteine, deren Sequenzmotive Schlüsse auf ihre Funktion zuließen. Sie sind aber beide unerlässlich für die Zellvitalität, da die jeweilige Deletion der Gene SNU71 oder SNU56 letal wirken [79]. Snu56p konnte in einer Protein-prä-mRNA Crosslink-Studie (vgl. Abschn. 1.4.2) als eines der Proteine, die die prä-mRNA im CC1 Komplex

banden, identifiziert werden [296]. Dabei schien Snu56p wie Nam8p Intronsequenzen nahe der 5' Ss zu binden. In [54] konnte ebenfalls gezeigt werden, dass Snu56p die prä-mRNA bindet, wobei die Interaktion von Snu56p mit der prä-mRNA anscheinend abhängig von der Interaktion mit U1-C war.

1.4.2 Die Interaktionen mit der prä-mRNA im CC1 und CC2

Das U1 snRNP ist der erste der fünf spleißosomalen snRNP Subkomplexe, der an die prä-mRNA bindet, und zwar an die 5' Ss in einem ATP-unabhängigen Prozess, wobei das Protein Yhc1 assistiert (Abschn. 1.3.1). Dieser Erkennungsprozess geschieht noch während der Transkription wie in *S. cerevisiae* gezeigt werden konnte [126], [132]. Es konnten in *S. cerevisiae* zwei frühe Komplexe, CC1 und CC2⁷, detektiert werden, die das U1 snRNP beinhalten [229] und die für ihre Bildung den CBC (Abschn. 1.1) auf der prä-mRNA benötigen [42], [142]. Beide Komplexe können sich bei Mutationen an der 5' Ss nicht bilden, doch nur die Formation des CC2 wird bei Mutationen des Verzweigungspunkts zerstört [230]. Dieser Befund und die Tatsache, dass der CC2 auf nativen Gelen langsamer läuft, deutete auf zusätzliche Komponenten im CC2 hin. Neben den siebzehn U1 snRNP Proteinen (Abb. 1.6) und den zwei Proteinen des CBC, namentlich Sto1 (Cbc1, human CBP80) und Cbc2 (human CBP20), im CC1 sind die Proteine BBP⁸ und Mud2p zusätzlich Komponenten im CC2 Komplex [211], [181]. BBP und Mud2p sind mit den humanen Proteinen SF1 und U2AF65 homolog [1], [211]. BBP (human SF1) ist das erste Protein, dass die Sequenz am Verzweigungspunkt erkennt [6], [15], wobei dessen KH Domäne den Verzweigungspunkt und die flankierenden Nukleotide spezifisch bindet [17], [148], [184], [200]. Die Bindung von Mud2p ist anscheinend hoch kooperativ mit BBP, wie im Falle von U2AF65 und SF1 gezeigt werden konnte [16]. Während das humane U2AF65 die Polypyrimidinsequenz bindet und ein Heterodimer mit U2AF35, [295], das das AG Dinukleotid an der 3' Ss erkennt [163], [291], bildet, gibt es in *S. cerevisiae* kein offensichtliches U2AF35 Homolog. Um den ersten Katalyseschritt auszuführen, bedarf es allerdings in *S. cerevisiae* zumindest *in vitro* auch nicht der Erkennung der 3' Ss [212].

Für die Formation und Stabilität in den CC sorgen mehrere Protein-RNA und RNA-RNA Wechselwirkungen. Der CBC scheint dabei eine wichtige Funktion zu übernehmen. In *S. cerevisiae* war die zu vermindertem Wachstum führende Deletion der CBC1 und CBC2 Gene in Kombination mit Mutationen in bestimmten U1 snRNP Proteinen letal

⁷CC1 & CC2: *Commitment*-Komplexe 1 & 2

⁸BBP *engl.* branchpoint binding protein, Verzweigungspunkt-Bindungsprotein

(Abb. 1.6) [65]. Ferner konnte in einer UV-crosslink Studie gezeigt werden, dass sich Sto1 (Cbc1) mit der prä-mRNA verbinden lässt [296]. In der gleichen Studie konnten auch mehrere U1 snRNP Proteine identifiziert werden, die an die prä-mRNA binden und die daher *in vivo* wohl zur Stabilität der CC beitragen: Nam8p, Snu56p, Snp1p, Yhc1, SmB, Smd1p und Smd2p. Außer diesen vermag auch Luc7p die prä-mRNA zu binden [196].

Die volle Stabilität im CC erfordert die konservierten Sequenzen sowohl an der 5' Ss, als auch am Verzweigungspunkt [139], [209], so dass möglicherweise das U1 snRNP kooperativ mit BBP und Mud2p an die prä-mRNA bindet [27]. Protein-Protein Wechselwirkungen erleichtern wahrscheinlich die Überwindung der Distanz zwischen den Proteinen, die an der 5' Ss und denen, die am Verzweigungspunkt gebunden haben. So interagiert BBP mit dem U1 snRNP Protein Prp40 [1], [114], während Mud2p möglicherweise den CBC kontaktiert [65]. In humanen Zellen sind in diesen E Komplexen außerdem verschiedene SR Proteine beteiligt [88]. So konnte für SC35 gezeigt werden, dass dieses Protein zwischen U1-70k und U2AF35 den Kontakt vermittelt. SF2/ASF hingegen fördert bereits die Formation von Prä-Spleißosomen, indem es mit seiner SR Domäne U1-70k bindet.

1.4.3 Der E-Komplex

Für die stabile Assoziation von U2 snRNP, dessen snRNA mit der prä-mRNA am Verzweigungspunkt Basenpaarung eingeht, sind zwei DExD/H Box Proteine, Sub2 (human UAP56; [298]) und Prp5, nötig. Sub2 scheint bei diesem Vorgang die Funktion zu haben, ATP-abhängig Mud2p und/oder BBP vom Verzweigungspunkt zu entfernen [27], um dem U2 snRNP die Bindung zu ermöglichen. Das humane Homolog UAP56 wurde durch seine Interaktionen mit U2AF65 (Mud2) gefunden und wird für *in vitro* Spleißen und die Assoziation von U2 snRNP mit der prä-mRNA benötigt [61]. Ähnlich wie die genetische Interaktion von Prp28 mit Yhc1 (Abschn. 1.3.1), hebt die Deletion von MUD2 in *S. cerevisiae* die Letalität einer Deletion von SUB2 auf [122], was ebenfalls die These einer funktionellen Interaktion unterstützt. Da die Inaktivierung von Sub2 in Hefezellextrakten bereits zu geringerer Formation von CC2 führte [298], hat Sub2 möglicherweise auch andere Funktionen in den frühen Schritten der Assemblierung des Spleißosoms inne [146].

Damit die U2 RNA die Region um den Verzweigungspunkt zu binden vermag, muss sie in der geeigneten Konformation vorliegen und es bedarf der U2 snRNP assoziierten Spleißfaktoren, die unerlässlich für die Formation des E-Komplexes sind und die in zwei Subkomplexen organisiert sind, SF3a und SF3b. In *S. cerevisiae* besteht SF3a aus

den Proteinen Prp9, Prp11 und Prp21 und SF3b aus Cus1, Hsh49, Hsh155, Rse1 und Snu17 (Abb. 1.4). Diese beiden Komplexe stabilisieren die Bindung von U2 snRNP an die prä-mRNA durch die Bindung flankierender prä-mRNA Abschnitte [83], [84]. Im humanen System können alle SF3 Proteine mit Ausnahme von SF3b130 im E-Komplex in Crosslink-Experimenten mit einer 20 Nukleotide umfassenden Region gerade 5' vom Verzweigungspunkt verbunden werden [88]. Das DExD/H Box Protein Prp5 scheint dann in Verbindung mit SF3a und SF3b für die Konformationsänderungen und die Zugänglichkeit der nötigen Sequenzabschnitte auf der U2 RNA zu sorgen [173], [287], entweder indem es das U2 snRNP Protein Cus2 vollständig destabilisiert oder indem es über Cus2 in Cus2-stabilisierten U2-RNA Abschnitten Änderungen induziert [188]. Während der Assoziation des U2 snRNP wird im humanen System SF1 durch das SF3b Protein SF3b14a/p14 am Verzweigungspunkt ersetzt [290]. Snu17p ist wahrscheinlich das Homolog in *S. cerevisiae* [81]. Es konnte jedoch noch nicht gezeigt werden, dass es den Verzweigungspunkt bindet.

Sind U1 und U2 snRNP gebunden, sind sowohl 5' Ss, als auch 3' Ss des jeweiligen Introns definiert. Für die relativ kurzen Introns in einfachen Eukaryonten, die stets in derselben Weise gespleißt werden, ist dies ein adäquater Mechanismus. In höheren Eukaryonten dagegen sind Introns meistens sehr viel länger als die Exons, und sie treten in höherer Anzahl in einer prä-mRNA auf (Abschn. 1.1). Dort wird die Erkennung der 3' Ss verbunden mit der Bindung des U1 snRNP an die 5' Ss des folgenden Introns [14], [204]. Für den weiteren Spleißprozess müssen sich diese Interaktionen über ein Exon wieder zu Gunsten von Interaktion über ein Intron verschieben [283]. Die Definition der Exons wird bei diesem Vorgang vor allem von SR Proteinen und hnRNP Proteinen (*engl.* heterogeneous nuclear Ribonucleoproteins) reguliert. Sie binden bestimmte Sequenzen auf der prä-mRNA, die abgekürzt ISE und ESE (*engl.* Intronic or Exonic Splicing Enhancers) oder ISS und ESS (*engl.* Intronic or Exonic Splicing Silencers) genannt werden. Über die Bindung der regulatorischen Proteine fördern ('Splicing Enhancer') oder unterbinden ('Splicing Silencer') diese Sequenzen den Spleißprozess. Während ESEs häufig von SR Proteine gebunden werden, werden ESSs typischerweise von hnRNPs gebunden. Dieses Konzept der Definition der Exons ist somit sehr wichtig für den alternativen Spleißprozess.

1.5 *S. cerevisiae* als Modellorganismus für den Spleißprozess

In der modernen Biologie ist Hefe aus mehreren Gründen ein wichtiger Modellorganismus geworden. Zuerst ist die einfache Handhabung der Zellkulturen im Labor zu nennen. Das Vorhandensein von haploiden und diploiden Zuständen, die relativ gute genetische Stabilität, die Verfügbarkeit von einer großen Anzahl von Selektionsmarkern und die Effizienz sowie Präzision von homologer Rekombination, die es ermöglicht, modifizierte Gene direkt in die chromosomale DNA zu integrieren, machen *S. cerevisiae* zu einem beliebten Organismus für genetische und proteinbiochemische Studien. Es entfällt das Nebeneinander von artifiziellem und endogenem Protein und die Problematik der Überexpression von Proteinen, die auch zu unerwünschten Proteinaggregaten führen kann. *S. cerevisiae* verfügt als eukaryontischer Einzeller über die Mechanismen, die notwendig sind, um prä-mRNA zu spleißen. Wie festgestellt werden konnte, sind die grundlegenden Bestandteile und die Funktionsweise des Spleißosoms von der Hefe bis zum Mensch konserviert [28]. Mitte der 90er Jahre lag das Genom von *S. cerevisiae* als erstes eines eukaryontischen Organismus sequenziert vor [74]. Dies ermöglichte die Suche nach homologen Proteinen mit bioinformatischen Methoden, und Erkenntnisse, die unter Anwendung der eleganten genetischen Techniken im Hefeorganismus gewonnen werden, können zum Verständnis des Spleißprozesses im Menschen beitragen. Da in *S. cerevisiae* der Mechanismus des alternativen Spleißens und die damit verbundene Vielzahl an zusätzlichen regulatorischen Proteinen fehlt, ist die Komplexität der spleißosomalen Assemblierungskaskade auf ein Mindestmaß reduziert, was die Erkenntnisfindung prinzipiell erleichtern sollte. Unterschiede, die in den Systemen gefunden werden, tragen desweiteren zum Verständnis der evolutiven Entwicklung auf molekularer Ebene bei.

1.6 Strukturuntersuchungen an spleißosomalen Komplexen

Um die mechanistischen Vorgänge während der Assemblierung des Spleißosoms und des Spleißprozesses umfassend zu verstehen, sind Informationen über die räumliche Anordnung der Komponenten im Ganzen und zueinander in den verschiedenen Funktionszuständen des Spleißosoms unerlässlich [283]. Allerdings sind die Strukturuntersuchungen gerade dadurch, dass es so viele unterschiedliche Konfigurationen und Konformationen gibt, erschwert. Die Einzelmolekül-Transmissionselektronenmikroskopie (*engl.* Single-Particle-Electronmicroscopy, im folgenden gekürzt mit EM bezeichnet) ist eine geeignete

Technik, um von den im Spleißzyklus auftretenden Komplexen in der Größenordnung von 240 *kDa* bis ≈ 5 *MDa*, Strukturen zu erhalten [248]. Diese Technik hat, verglichen mit der Röntgenkristallographie, den Vorteil, dass die Notwendigkeit entfällt, von den Komplexen gut streuende Kristalle zu erhalten. Vor allem dadurch bedingt sind die benötigten Proteinmengen um einen Faktor 100 – 1000 geringer und der Probenpräparationsmaßstab muss in der Regel nicht für die Strukturbestimmung angepasst werden [248], was ein weiterer Vorteil bei den nur in geringen Mengen in der Zelle auftretenden spleißosomalen Komplexen ist.

Wie die zahlreichen Strukturuntersuchungen am Ribosom zeigen, können grundsätzlich auch mittels der Kryo-Elektronenmikroskopie Strukturauflösungen von 10 – 20 Å von asymmetrischen biologischen Komplexen erreicht werden. Kürzlich erschienen vom Ribosom EM-Strukturen bei 7 Å Auflösung [226], [280]. Die dynamische Natur und Funktionsweise des Spleißosoms führt jedoch zu starker Heterogenität der Probe, wodurch die erreichbare Auflösung der EM-Strukturen limitiert wird. Für die spleißosomalen Komplexe B Δ U und C wurden Auflösungen um die 40 Å erreicht [22], [111]. In diesen EM-Strukturen war es nicht möglich, die relative Position der einzelnen snRNP anzugeben.

An den spleißosomalen Komplexen U11/U12 di-snRNP, dem Subkomplex SF3b und dem humanen U1 snRNP konnten Auflösungen um die 10 Å erreicht werden [76], [77], [249]. Im Falle von SF3b konnten einige charakteristische Sekundärstrukturelemente erkannt werden, die es ermöglichten, 3 Proteine aufgrund ihrer strukturellen Domänen in der EM-Struktur zu positionieren. Über das humane U1 snRNP, der zudem neben dem Sm-Ring eine geringe Anzahl weiterer Proteine trägt, lagen reichlich biochemische und strukturelle Daten vor, sowie elektronenmikroskopische 2D Aufnahmen mit Antikörper-Bindung [116], [118], so dass die Proteine U1-70k und U1-A in der EM-Struktur lokalisiert werden konnten. Von diesem snRNP ist mittlerweile sogar eine Röntgenkristallstruktur veröffentlicht worden [192].

Von dem humanen tri-snRNP konnte erst nach Etablierung des GraFix Protokolls [119], [248] eine EM-Struktur von 20-30 Å erhalten werden [218], wobei mit EM-Strukturen des U5 snRNP und des U4/U6 di-snRNP der grobe Aufbau des tri-snRNP beschrieben werden konnte.

Um die Positionen einzelner Proteine in der EM-Struktur zu lokalisieren, wurden im Tri-snRNP und im U6 snRNP jeweils im Zellorganismus *S. cerevisiae* einige Proteine genetisch mit einem 50 *kDa* und/oder 30 *kDa* Proteinlabel versehen [93], [115]. Elektronenmikroskopische Aufnahmen zeigten zusätzliche Dichte im Bereich des genetischen Labels, womit die Proteine im Komplex nachgewiesen werden konnten.

1.7 Transmissionselektronenmikroskopie (TEM)

Die Verwendung eines Elektronenstrahls für die Mikroskopie basiert auf den Erkenntnissen de Broglies⁹, der den Welle-Teilchen-Dualismus eines Elektrons postulierte:

$$\lambda = \frac{h}{p} \quad (1.1)$$

wobei λ die Wellenlänge, h das Plancksche Wirkungsquantum und p der Impuls aus Masse m und Frequenz ν ($p = m \times \nu$) ist. Demnach können Elektronen als Wellen mit Wellenlängen unter 1 nm beschrieben werden. Als generelles Prinzip in der Mikroskopie ist nach dem Gesetz von Abbe¹⁰ die Wellenlänge λ ein limitierender Faktor in der Punktauflösung d eines Lichtmikroskops. Gleichfalls abhängig ist die Auflösung vom Brechungsindex n des Mediums und dem halben Öffnungswinkel der Objektivlinse α (Aperturwinkel):

$$d = \frac{0.61 * \lambda}{n * \sin \alpha} \quad (1.2)$$

Diese Gleichung gilt prinzipiell auch für das Transmissionselektronenmikroskop¹¹, wobei die Wellenlänge λ der beschleunigten Elektronen mit der Beschleunigungsspannung des TEM U_b angenähert werden kann: $\lambda \approx 3,7 * U_b^{-0,6}$.

Mit modernen TEM, wie sie im Bereich der Materialforschung eingesetzt werden, werden maximale Auflösungen um $0,1\text{ nm}$ erreicht, so dass einzelne Atome abgebildet werden können. Die ersten Elektronenmikroskope wurden von E. Ruska¹², M. Knoll¹³ und B. von Borries¹⁴ Anfang der 30er Jahre des letzten Jahrhunderts konstruiert.

1.7.1 Bildentstehung im TEM

Bei der Bildentstehung im TEM kommt es wie bei der Lichtmikroskopie auch zu grundsätzlich zwei unterschiedlichen Interaktionen des einfallenden Strahls mit der Probe, der elastischen und der inelastischen Streuung.

Bei der elastischen Streuung wird ein eintreffendes Elektron durch elektrostatische Wechselwirkung mit einem Atomkern der Probe von seiner Bahn abgelenkt. Diese Ablenkung ist desto größer, je stärker die Coulomb-Kraft wirkt, das heißt je näher das Elektron dem positiv geladenen Atomkern kommt, je stärker positiv geladen der Atomkern ist (Kernladungszahl Z ; Kontrastierung mit Schwermetallen) und je langsamer die Geschwindigkeit

⁹Louis de Broglies, 1892 - 1987

¹⁰Ernst Abbe, 1840 - 1905

¹¹Transmissionselektronenmikroskop & -mikroskopie: beides abgekürzt mit TEM

¹²Ernst Ruska, 1906 - 1988

¹³Max Knoll, 1897 - 1969

¹⁴Bodo von Borries, 1905 - 1956

des Strahlelektrons ist, was wiederum abhängig von der angelegten Beschleunigungsspannung ist. Das Elektron erfährt bei dieser Wechselwirkung keinen Energieverlust (Bild rechts in Abb. 1.9).

Die biologischen Präparate, die hauptsächlich aus Atomen niedriger Ordnungszahl bestehen, verhalten sich wie schwache Phasenobjekte. Das gilt auch in Näherung noch für die Proteinkomplexe, die mit Schwermetallsalzen negativ kontrastiert wurden. Das bedeutet, dass der Elektronenstrahl durch das Objektpotential (entspricht dem Brechungsvermögen eines Präparats in der Lichtmikroskopie) eines biologischen Präparats durch elastische Streuung eine Phasenverschiebung unter Beibehaltung der Wellenlän-

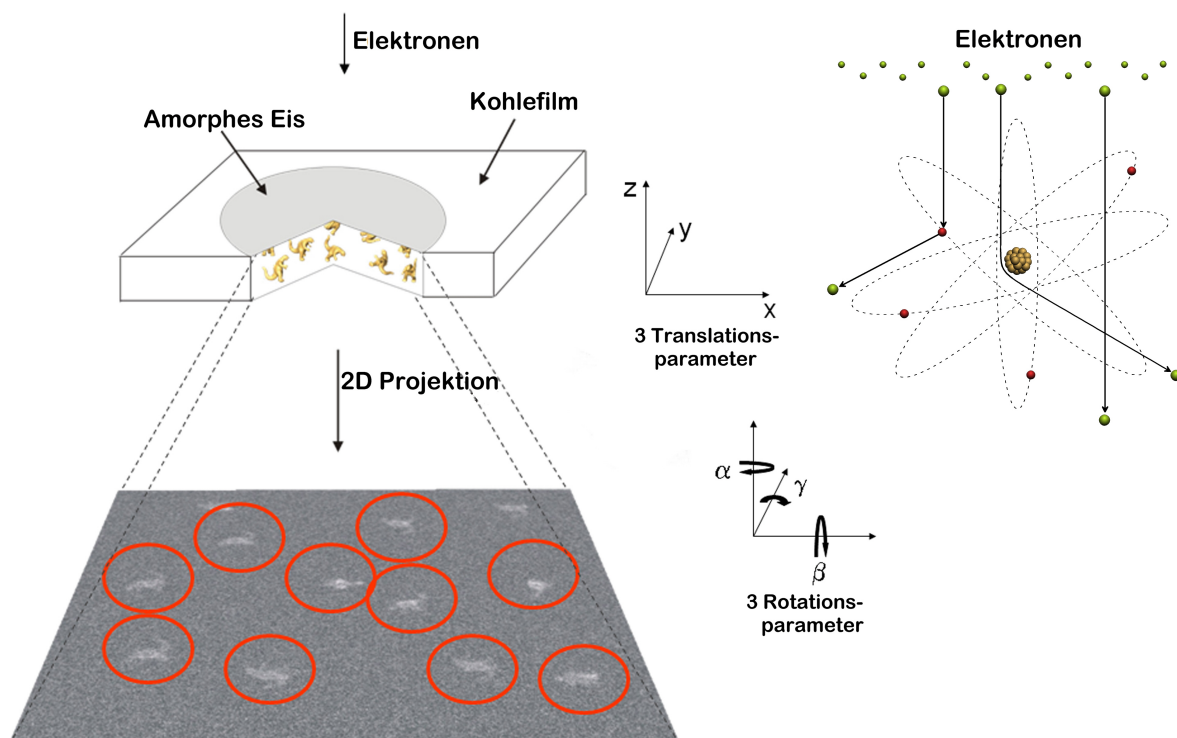


Abbildung 1.9: Entstehung von 2D Projektionsbildern im Transmissionselektronenmikroskop. Ein EM Projektionsbild entsteht aufgrund der elastischen Streuung von Elektronen an den Atomen des Präparates, wodurch sie eine Phasenverschiebung erfahren (Phasenkontrast). Sehr stark elastisch gestreute Elektronen treffen auf die Kontrastblende und tragen zum Amplitudenkontrast bei. Die inelastische Streuung bei Auftreffen eines Strahlelektrons auf ein Hüllelektron der Probe führt letztlich zu einer Verringerung des Kontrasts, da sich die Wellenlänge des Elektrons ändert, die Kohärenz des Elektronenstrahls insgesamt abnimmt, aber die elastische Streuung abhängig von der Wellenlänge ist. Betroffen von letzterem Effekt sind insbesondere in amorphem Eis eingebettete Proteinkomplexe.

Die Proteinkomplexe der Probe haben 3 translationale Freiheitsgrade und auch 3 Freiheitsgrade der Rotation. Bildteile freundlich zur Verfügung gestellt von Prof. H. Stark.

ge und einer kaum merklich reduzierten Amplitude erfährt. Der Phasenunterschied der

gebeugten gegenüber der einfallenden Welle ist sehr viel geringer als die Wellenlänge der Elektronen ($\delta \ll \lambda$) und beträgt unter diesen Annahmen $\approx \frac{\pi}{2} = \frac{\lambda}{4}$. Da Phasenunterschiede für das Auge und das Photomaterial nicht sichtbar sind, ist das Ziel, die beiden Elektronenstrahlen, den Nullstrahl und den gebeugten Strahl, so um $\pi/2$ zu verschieben, dass die Amplituden beider Wellen phasengleich aufeinander fallen. Diese konstruktive Interferenz, bei der die resultierende Amplitude gleich der Summe beider Einzelamplituden ist, wird wahrgenommen und als Phasenkontrast bezeichnet. Während diese Transformation im Phasenkontrastmikroskop durch eine Zernicke-Phasenplatte¹⁵ ($\frac{\lambda}{4}$ -Plättchen) im Strahlengang des Nullstrahls leicht erreicht werden kann, wird dies im TEM mit der Verschiebung des gebeugten Strahls durch Defokussierung der Objektivlinse gelöst (die Probe wird aus der Fokusebene der Objektivlinse gefahren), wobei eine Verschiebung um $n\frac{\pi}{2}$ einen positiven und um $-n\frac{\pi}{2}$ einen negativen Phasenkontrast ergibt. Abhängig ist diese Phasenverschiebung im TEM auch noch von der nicht korrigierbaren sphärischen Abberation der Objektivlinse, die mit der Konstanten C_s beschrieben wird, und außerdem von der Elektronenwellenlänge, bestimmt durch die Beschleunigungsspannung des Mikroskops. Die optimale (opt) Einstellung, bei der die Phasenverschiebung der gebeugten Welle über einen weiten Beugungswinkelbereich $\frac{\pi}{2}$ beträgt und der Phasenkontrast somit maximal ist, liegt im schwachen Unterfokusbereich bei $\Delta z_{opt} = (\sqrt{C_s \lambda})$ (Scherzer, 1949) und wird nach dem Elektronenmikroskopiker O. Scherzer¹⁶ Scherzer-Fokus genannt. Da die optimale Phasenverschiebung nicht für alle Beugungswinkel erreicht werden kann, ist die Folge eine Phasenkontrastübertragungsfunktion (Abschn. 1.7.3).

Der Amplitudenkontrast entsteht durch elastisch stark gestreute Strahlelektronen, die auf die Objektivblende (Kontrastblende) treffen, da so die Intensität I des Elektronenstrahls gemindert wird. Die Absorption der Elektronen durch das Präparat selbst ist bei den dünnen Objekten in der Einzelmolekül-EM sehr gering, so dass der Amplitudenkontrast, der auf diese Weise entsteht, unerheblich ist.

Inelastische Streuung entsteht durch das Auftreffen von Strahlelektronen auf die Elektronenhülle eines Atoms. Dabei findet gleichzeitig ein Energietransfer statt, der die kinetische Energie der Strahlelektronen verlangsamt. Die weitgehende Kohärenz des Elektronenstrahls weist nach dem Objektdurchgang ein Wellenlängenspektrum auf. Da die elastische Streuung abhängig von der Wellenlänge ist, führt dies letztlich zu einer Verringerung des Kontrasts. Besonders betroffen davon sind massereiche Präparate wie in amorphes Eis eingebettete Proteinkomplexe.

¹⁵Frederik Zernicke, 1888 - 1966

¹⁶Otto Scherzer, 1909 - 1982

1.7.2 Das elektronenmikroskopische Bild

Das elektronenmikroskopische Bild entsteht aus einer Vielzahl von Wellen, einem sogenannten Wellenpaket $\Psi(x, t)$. Im EM Bild bleiben die Strukturinformationen des Objekts in den Raumfrequenzen (Schwingungen pro Längeneinheit) erhalten, wobei die Kontrastübertragung des Mikroskops die abgebildete Funktion verändert (Abschn. 1.7.3). Das Reziproke der Raumfrequenz, die Wellenlänge λ , entspricht einer bestimmten Strecke im realen Bild und die Amplitude der maximalen Dichteschwankung über diese betrachtete Strecke λ . Die Amplitude wird auch Strukturfaktor genannt. Dementsprechend und den Erkenntnissen des Mathematikers Fourier folgend wird die Struktur des abgebildeten Objekts im Bild aus Einzelfunktionen von periodisch harmonischen Schwingungen, also Sinus- und Kosinusfunktionen, mit ihrem gesamten kontinuierlichem Spektrum an Frequenzen ν_i und dazugehörigen Amplituden k_i und Phasen (Phasenverschiebungen) $\Delta\varphi_i$ präsentiert. Eine Darstellung unendlich vieler Funktionen ist nicht praktikabel. Da das Bild digitalisiert aus einer endlichen Anzahl von Pixeln n besteht, wird das Bild mit $\frac{n}{2}$ verschiedenen Funktionen ($\frac{n}{2}$ verschiedenen Raumfrequenzen) beschrieben, und zwar hat die kleinste Frequenz $\frac{1}{\lambda}$ die Wellenlänge $\lambda = n$ und die kleinstmögliche Wellenlänge beträgt 2 Pixel, was der größtmöglichen Frequenz entspricht, der sog. Nyquist-Frequenz. Mit der Bedeutung von x als Strecke im Bild lautet die Funktion $F(x)$ des abgebildeten Objekts:

$$F(x) = \sum [a_i \cos(2\pi\nu_i x) + b_i \sin(2\pi\nu_i x)] \quad (1.3)$$

mit den Gewichtungsfaktoren $a_i = k_i \sin(\Delta\varphi)$ und $b_i = k_i \cos(\Delta\varphi)$. sowie x als Strecke im Bild. Mit den Methoden der Fourier-Analyse wird die Funktion $F(x)$ in ihre Einzelkomponenten zerlegt, und es wird eine Frequenzverteilung erhalten. Die unterschiedlichen Frequenzen sind durch die Berechnung der Strukturfaktoren und Phasenverschiebungen so zu wichten, dass die Funktionen in ihrer Summe die Dichteverteilung im Bild wiedergeben. Während niedrige Ortsfrequenzen die grobe Objektstruktur wiedergeben, repräsentieren hohe Ortsfrequenzen feine Strukturdetails wie Sekundärstrukturelemente. Für die Überprüfung der Qualität des EM-Bildes wird die Verteilung und Größe der Strukturfaktoren (Amplituden) als Funktion der Raumfrequenz genutzt, dargestellt im sogenannten Powerspektrum eines Bildes, das die Quadrate der Amplituden enthält.

1.7.3 Phasenkontrastübertragungsfunktion

In der Realität sind magnetische ebenso wie lichteoptische Linsen nicht ideal gebaut. Ein Fehler der auftritt, die sphärische Abberation C_s der Objektivlinse, führt zu unter-

schiedlich starker Beugung des Elektronenstrahls beim Passieren der Linse außerhalb des Zentrums der optischen Achse. Die Brennweite ist kürzer für Strahlenverläufe durch die Peripherie der Linse. Dies hat unterschiedlich lange Strahlenwege und somit Phasenverschiebungen zur Folge, die nicht durch Einstellung am Mikroskop korrigierbar sind. Da die Phasenverschiebungen zu einer Änderung der Amplituden führen, ändert sich der Kontrast. Die Folge ist eine Phasenkontrastübertragungsfunktion, mit der die Funktion der Objektstruktur gefaltet wird. Die Phasenkontrastübertragungsfunktion ergibt sich aus der Phasenverschiebung in Abhängigkeit von Defokus $\Delta\hat{z}$ und Ortsfrequenzen \hat{k} zu:

$$PCTF = \sin\pi[-\Delta\hat{z}\hat{k}^2 + \frac{1}{2}\hat{k}^4] \quad (1.4)$$

und gilt mit den Konventionen normalisierter Größen für Defokus $\Delta\hat{z}$ (1.5) und Raumfrequenz \hat{k} (1.6) für alle Transmissionselektronenmikroskope unabhängig von Hochspannung und sphärischer Aberration.

$$\Delta\hat{z} = \frac{\Delta z}{\sqrt{C_s\lambda}} \quad (1.5)$$

$$\hat{k} = k\sqrt[4]{C_s\lambda^3} \quad (1.6)$$

1.8 Ziel dieser Arbeit

Die Experimente und Datenverarbeitung im Rahmen dieser Arbeit hatten zum Ziel, die 3D Struktur des U1 snRNP aus *S. cerevisiae* zu berechnen.

Um einen möglichst homogenen U1 Komplex als Probe für die elektronenmikroskopische 3D Struktur Bestimmung zur Verfügung zu haben, sollte zunächst die Aufreinigungsmethode des U1 snRNP in Zusammenarbeit mit der Abteilung Zelluläre Biochemie von Prof. R. Lührmann (Max-Planck-Institut für Biophysikalische Chemie, Göttingen) optimiert werden. Anschließend sollte unter Verwendung elektronenmikroskopischer Datenverarbeitungsverfahren, die die Rekonstruktion mehrerer unterschiedlicher 3D Strukturen aus einem Datensatz ermöglichen, die strukturelle Heterogenität des Komplexes berücksichtigt werden, um so Einblicke in die Dynamik der strukturellen Organisation des U1 snRNP zu gewinnen. Um ein besseres Verständnis für das Gefüge der Proteine innerhalb des Komplexes zu erhalten, sollte in weiteren Experimenten mittels Antikörperbindung und genetisch fusionierten Proteintags einzelne Proteine im U1 snRNP Komplex in der EM-Struktur lokalisiert werden. Darüber hinaus ließen sich möglicherweise über die erhaltenen Strukturdaten Erkenntnisse über den Initiationsprozess des Spleißzyklus in *S. cerevisiae* erlangen.

Kapitel 2

Einleitung in die Methodik der Bildverarbeitung

Die Bildverarbeitung in der Einzelpartikel-TEM dient der Rekonstruktion der 3D Struktur des ausgewählten Proteinkomplexes aus seinen im EM-Bild abgebildeten 2D Projektionen [263], [276]. Zunächst ist es notwendig, einen Datensatz zu generieren, der aus Einzelbildern des Proteinkomplexes besteht. Es folgen Datenprozessierungszyklen auf 2D Ebene vor der eigentlichen Rekonstruktion der 3D Struktur. Weitere iterativ oder zyklisch angelegte Datenprozessierungsmethoden dienen sowohl der Validierung, als auch der Verbesserung der 3D Struktur.

Für die im folgenden beschriebene Datenverarbeitung standen im wesentlichen 2 große Programmpakete zur Verfügung, und zwar Imagic-5 [275] und CowHide, das im Hause entwickelt wird.

2.1 Datensatz aus Einzelpartikelbildern

2.1.1 Aufnahmetechniken

Für die Aufnahme der Bilder stehen grundsätzlich zwei Verfahren zur Verfügung, und zwar die konventionelle Aufnahmetechnik auf Film und die heutzutage vielgenutzte digitale Aufnahmemöglichkeit mit einer CCD-Kamera unter Anwendung der Spot Scan Technik [52]. Die direkte Aufnahme mit einem CCD-Detektor birgt den Vorteil, dass die nachträgliche für die weitere Datenprozessierung notwendige Digitalisierung der analogen Filmaufnahmen entfällt. Desweiteren konnte gezeigt werden, dass CCD-Detektoren im mittleren Auflösungsbereich (10 - 25 Å) eine bessere Übertragung der Bildinformationen liefern [217], was einen Beitrag zur Qualität und Richtigkeit von *de novo* 3D Strukturen leistet. Da die untersuchten biologischen Proteinkomplexe sensitiv auf die Elektronenstrahlung reagieren und bei Strahlexposition Schaden nehmen [100], werden die

Aufnahmen mit einer geringen Elektronendosis $< 20 e^-/\text{\AA}^2$ gemacht [8]. Dies führt zu Bildern mit einem geringen Signal zu Rausch Verhältnis (SNR^1 , Signal-zu-Rausch Verhältnis).

2.1.2 Selektion der Partikel

Die unprozessierten EM-Bilder enthalten viele 2D Projektionsbilder der einzelnen Proteinkomplexe, die in mehr oder weniger zufälliger Orientierung auf dem Kohlefilm liegen. Diese Partikelansichten müssen zunächst selektiert werden, um letztlich einen Datensatz aus Einzelpartikelbildern zu erhalten. Eine visuell-manuelle Selektion am Computerbildschirm mit einem dafür geeigneten Computerprogramm gewährleistet, dass die Partikelansichten zentriert ausgewählt und demzufolge zentriert extrahiert werden können. Insbesondere für nicht kugelförmige Proteinkomplexe und solche, die möglicherweise strukturelle Heterogenitäten aufweisen, ist diese zeitintensive Methode vorzuziehen. Es stehen allerdings auch Computerprogramme zur Verfügung, die eine (semi-)automatische Auswahl der Partikel ermöglichen (z.B. Signature [37], Boxer aus dem EMAN Programmpaket [154]). In diesen Fällen gibt der Anwender dem Programm mit einer Vorauswahl von meistens 20 – 100 Partikeln im Bild Referenzen vor. Existiert bereits eine 3D Struktur des Proteinkomplexes, ist es auch möglich, Referenzen, abgeleitet aus dieser Struktur, zu verwenden. Der Anwender überprüft bei jeder der automatischen Methoden, dass keine Artefakte ausgewählt wurden. Artefakte können beispielsweise Kontaminationen von gefrorenem Eis oder kleine Schmutzpartikel auf dem Kohlefilm sein. Auch muss überprüft werden, dass der verwendete Algorithmus sensitiv genug ist und alle Subpopulationen des Partikels berücksichtigt werden. Nachdem die Koordinaten bestimmt wurden, werden diese benutzt, um den Partikel in einem kleinen Pixelrahmen zu extrahieren. Für das 2D *Alignment* sollte die Größe des Pixelrahmens so gewählt werden, dass der Partikel idealerweise $\frac{2}{3}$ der Kantenlänge einnimmt.

2.1.3 Bandpassfilter

Hohe Ortsfrequenzen repräsentieren zwar feine Strukturdetails (Abschn. 1.7.2), müssen aber vom Rauschen in den Bildern, das keine Feinstruktur des Objekts beschreibt, getrennt werden. Andernfalls könnten sie das nachfolgende 2D *Alignment* fehlerleiten. Sehr niedrige Ortsfrequenzen hingegen resultieren nicht aus den Objekteigenschaften, sondern beispielsweise aus Variationen des Kohlefilms bedingt durch Eis- und/oder Kontrastmitteldicke. Aus diesem Grund werden die Einzelpartikelbilder mit einem Bandpassfilter

¹SNR: *engl.* Signal-to-Noise Ratio

versehen, der sowohl die hohen, als auch die niedrigen Ortsfrequenzen entfernt. Im zyklischen Verlauf der Datenprozessierung können verstärkt wieder höhere Ortsfrequenzen zugelassen werden, um feinere Objektstrukturdetails zurück zu gewinnen, es bleibt aber bei der Filterung von niedrigen und hohen Ortsfrequenzen.

Um für das nachfolgende 2D *Alignment* und der damit verbundenen Suche nach den besten Kreuzkorrelationskoeffizienten (Abschn. 2.2.1) das umliegende Rauschen des Partikels so gering wie möglich zu halten, wird außer der Anwendung des Bandpassfilters eine Kreismaske um das Partikel gelegt. Die Benutzung des Bandpassfilters im Imagic-5 Softwarepaket nimmt außerdem eine Normalisierung der Grauwerte vor. Können mit dem Datensatz höhere 3D Strukturauflösungen erreicht werden (Verbesserung einer bereits vorhandenen 3D Struktur), ist es notwendig, vor Anwendung des Bandpassfilters die Bildinformationen mit der Phasenkontrasttransferfunktion (PCTF) zu korrigieren [216].

2.2 Datenprozessierung auf 2D Ebene

Da im EM die Proteinkomplexe in zufälliger Orientierung aufgenommen werden, liegen die 2D Projektionsansichten der Proteinkomplexe in den Einzelbildern ebenfalls in zufälliger Orientierung vor. Das geringe SNR in den Bildern führt zu einem hohen Rauschsignal, so dass Details in den Partikelansichten schwer erkennbar sind. Die für die 3D Rekonstruktion notwendige Zuweisung der Projektionswinkel erfolgt sinnvoller Weise daher nicht für jedes einzelne Bild, sondern für eine Gruppe (Klasse) von Bildern bestehend aus in der Regel 10 – 40 ähnlichen Ansichten. Summiert ergeben diese eine Klassensumme, die ein erhöhtes SNR aufweist und eine Subpopulation von Partikelansichten im Datensatz repräsentiert. In eine Klasse fallen nämlich alle Partikel, die ungefähr mit derselben Seite an den Kohlefilm gebunden haben oder in derselben Lage im Eis liegen, aber mit verschiedenen Winkeln α um eine Achse senkrecht zur Trägerebene gedreht sind (Abb. 1.9). Für die Berechnung der Klassen/Klassensummen ist daher eine vorherige Rotationsausrichtung der Projektionsansichten unerlässlich. Ebenfalls notwendig ist eine gleichzeitige translationale Zentrierung der Einzelpartikelansichten, da die Selektion der Koordinaten auf der EM-Aufnahme für die Extraktion der Einzelpartikelbilder nicht genau genug im Zentrum des jeweiligen Partikels erfolgen konnte. Beide Arten der Ausrichtung der Einzelpartikelbilder werden im sog. *Alignment* (Abb. 2.1 & Abschn. 2.2.1) vorgenommen.

2.2.1 *Alignment*

Für das *Alignment* stehen zwei verschiedene Methoden zur Verfügung, das direkte *Alignment* und das *Alignment* über Polarkoordinaten, die nachfolgend kurz beschrieben

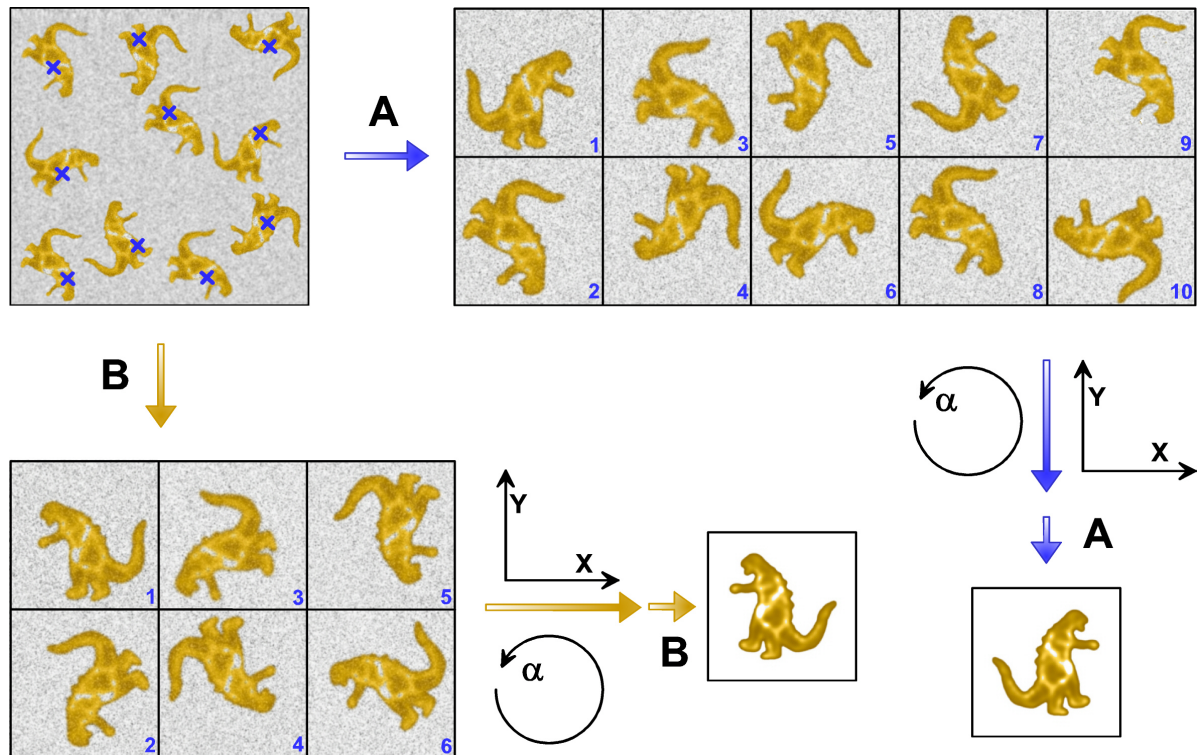


Abbildung 2.1: Schematische Darstellung der Ausrichtung der EM-Einzelpartikelbilder. Um nach Selektion und Extraktion der Einzelpartikelbilder (**A**, blaue Pfeile und **B**, gelbe Pfeile) gleiche EM-Projektionsansichten (1 – 10 und 1 – 6, in diesem Beispiel Bild und Spiegelbild) nachfolgend in eine Klasse **A** und **B** sortieren zu können, müssen sie in einem *Alignment*-Prozedere durch translationale Ausrichtung zentriert werden und durch Rotation in gleicher Weise im Pixelrahmen ausgerichtet werden. Die Klassensummen weisen ein verbessertes SNR auf.

werden. Sie basieren beide darauf, durch Verschiebung und Rotation des Bildes, die Ähnlichkeit des 2D Projektionsbildes zu einem Referenzbild zu maximieren. Dabei sind sich zwei Bilder dann ähnlich, wenn sich die Grauwerte der Pixel des einen Bildes ähnlich verändern wie die des anderen. Um die höchste Korrelation von Bild zum Referenzbild zu erhalten, wird die nötige Operation der Kreuzkorrelation bei beiden Methoden rechenstechnisch im Fourierraum umgesetzt, da sie sich dort einfacher berechnen lässt [265].

Direktes *Alignment*

Die ursprüngliche Methode ist ein sogenanntes direktes *Alignment* [250], das separat nach dem jeweils höchsten Wert für den Translations- und den Rotationskorrelationsko-

effizienten von Bild und Referenz sucht. Die Ermittlung dieser beiden Korrelationspeaks erfolgt abwechselnd und iterativ. Ein Algorithmus dieser Art, implementiert in das Imagic-5 Programmpaket, wurde für den initialen translationalen *Alignment*-Schritt verwendet.

***Alignment* über Polarkoordinaten**

Im Gegensatz zum direkten *Alignment* wurde für das Multi-Referenz-Alignment (s. u.) in dieser Arbeit ein Algorithmus des sogenannten *engl. exhaustive alignments* verwendet [108], [215]. Insbesondere für Bilder mit geringem SNR liefert dieser Algorithmus bessere Ergebnisse, ist allerdings auch erheblich rechenintensiver. Zunächst wird das Bild in Polarkoordinaten transformiert, und zwar für jedes mögliche Zentrum des Partikels im Pixelrahmen. Dieser Bereich der möglichen Zentren wird benutzerdefiniert sinnvoll eingeschränkt, um die nötige Rechenleistung so gering wie möglich zu halten. Anschließend wird für jede dieser translationalen Positionen der jeweils höchste Rotationskorrelationskoeffizient zu einer Referenz gesucht. Der von allen ermittelte höchste Rotationskorrelationskoeffizient liefert so neben dem Rotationswinkel als Ergebnis auch die translationalen Parameter. Eine getrennte Berechnung der Translationsparameter wie bei der direkten Methode ist in diesem Fall nicht mehr erforderlich.

Initiales Referenzalignment

Zu Beginn ist bei der *de novo* Rekonstruktion eines Proteinkomplexes weder eine 2D Referenz noch eine 3D Struktur, aus der Referenzen für ein 2D *Alignment* generiert werden könnten, verfügbar. Der erste *Alignment*-Schritt erfolgt daher gegen eine kreisförmige Referenz, gebildet aus einer rotationsgemittelten totalen Summe der Einzelpartikelbilder des Datensatzes und wird auf die Ausrichtung der translationalen Parameter beschränkt. Diese initiale Massenzentrierung wird auch als referenzfrei (*engl. reference-free Alignment*) bezeichnet [55]. Da die Einzelpartikelbilder mit einer gewissen Zentrierung im Pixelrahmen vorliegen, befinden sich in der (Rotations-gemittelten) totalen Summe aller Bilder die hellen Pixel in der Mitte und die dunklen am Rand (Abb. 2.2 A & B). Diese Art der initialen Referenz birgt den Vorteil, dass sie faktisch frei von Detailinformationen ist und somit keinen Bias im ersten *Alignment*-Schritt verursacht. Nachfolgend können Klassensummen gebildet werden (Abschn. 2.2.2), aus denen neue Referenzen für das nächste *Alignment* gewonnen werden.

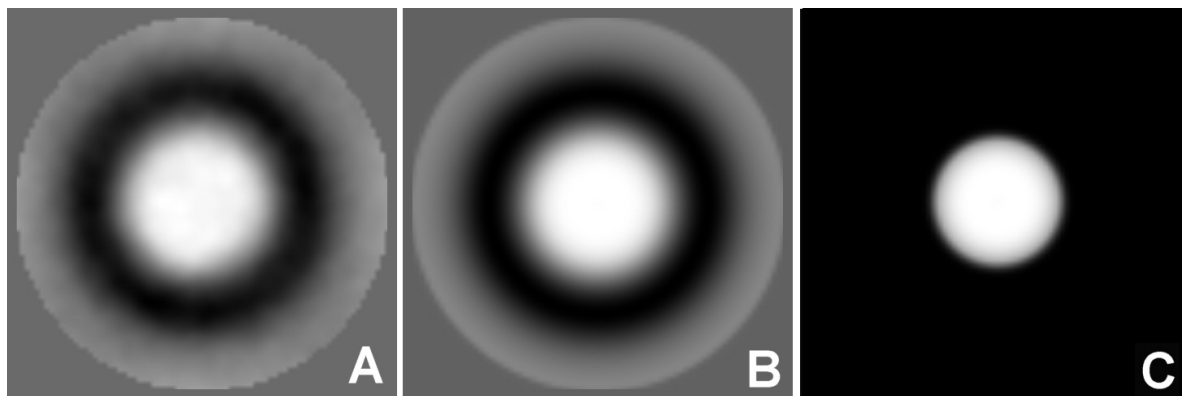


Abbildung 2.2: Referenz für das initiale translationale *Alignment*. Gezeigt ist ein Beispiel für die Generierung einer Referenz für die initiale Massenzentrierung der Einzelpartikelbilder. Der Datensatz bestehend aus 10932 Einzelpartikelbildern wurde aus einer mit Uranylformiat negativ kontrastierten U1 snRNP Probe erhalten. Die totale Summe (A) des Datensatzes wird rotationsgemittelt (B) und anschließend so maskiert (C), dass eine Ausrichtung auf den schwarzen Kontrastmittelsaum nicht möglich ist.

Multi-Referenz-*Alignment*

In folgenden *Alignment*-Schritten werden mehrere Referenzbilder in einem Multi-Referenz-*Alignment*² verwendet [186]. Es werden jene Translations- und Rotationsparameter für die Ausrichtung des Bildes angewendet, die unter Einbeziehung aller Referenzen den höchsten Korrelationskoeffizienten lieferten.

Referenzen aus einer 3D Struktur

Sobald eine 3D Struktur vorliegt, können aus dieser in bestimmten anzugebenden Winkelabständen 2D Projektionen gebildet werden. Diese werden in der Regel für 2D *Alignments* eingesetzt, die der iterativen Verbesserung der 3D Struktur dienen (Abschn. 2.3).

2.2.2 Klassifizierung

Nachdem die 2D Projektionsbilder durch das 2D *Alignment* ausgerichtet worden sind, werden von ähnlichen Projektionsansichten Klassen gebildet [272]. Idealerweise geben die Klassen die Ansichten des Makromoleküls unter verschiedensten Projektionswinkeln wieder. Für die Bildung der Klassen existieren verschiedene Verfahren. Da es aufgrund des Rechenaufwands nicht möglich ist, alle Einzelpartikelbilder Pixel für Pixel zu vergleichen, wird zunächst eine Dimensionsreduktion vorgenommen, wofür leicht unterschiedliche Verfahren [66] [272], wie beispielsweise die Hauptkomponentenanalyse

²MR-*Alignment*: Multi-Referenz-*Alignment*

(PCA³, Hauptkomponentenanalyse) oder Korrespondenzanalyse (CA⁴) [67] eingesetzt werden können. Sowohl im Imagic-5 Softwarepaket, als auch im CowHide Softwarepaket werden Verfahren der Hauptkomponentenanalyse genutzt [270],[156]. Hingegen ist im SPIDER Programmpaket, das für EM Datenverarbeitung ebenfalls weit verbreitet ist, ein Verfahren der Korrespondenzanalyse implementiert.

Hauptkomponentenanalyse

In einem Datensatz, bestehend aus (n) Bildern, kann ein Bild aus (p) Pixeln als p-dimensionalen Vektor dargestellt werden, wobei die Grauwertinformationen der Bildpunkte die Werte liefern. Alle Bilder im Datensatz werden von der Matrix X beschrieben mit $\mathbf{n} \times \mathbf{p} = X$. In diesem p-dimensionalen Raum liegen die Vektoren ähnlicher Bilder nahe beisammen. Um diese Cluster von ähnlichen Bildern in dem Raum zu identifizieren und die Computerrechenkapazität nicht zu sprengen, wird nun die Dimensionalität durch Berechnung eines orthogonalen Systems von Eigenvektoren (in der digitalen Bildverarbeitung Eigenbilder genannt) reduziert. Es wird ein neues Koordinatensystem bestimmt, das trotz seiner reduzierten Komplexität dennoch die relevanten Informationen des Datensatzes wiedergibt. Die erste, neu zu bestimmende Achse liegt in Richtung des höchsten Informationsgehalts des Datensatzes, die zweite wird orthogonal dazu aufgestellt und so fort. Auf diese Weise beschreiben die neuen Achsen die signifikante Varianz des Datensatzes und können benutzt werden, um jedes Bild des Datensatzes als Linearkombination dieser Eigenbilder so gut wie möglich zu beschreiben. Die Anzahl der zu berechnenden Eigenbilder (Abb. 2.3) wird benutzerdefiniert vorgegeben und kann in Imagic-5 maximal 69 betragen für ein 69-dimensionales orthogonales Koordinatensystem. In CowHide ist diese Anzahl nicht derart begrenzt. Die Eigenbilder werden entsprechend ihrer Eigenwerte als Maß für die jeweilige Relevanz des Eigenbildes sortiert, so dass Eigenbilder höherer Ordnung verstärkt Rauschen repräsentieren. Die Bestimmung der Eigenbilder erfolgt iterativ und in Imagic-5 mit der zu Grunde liegenden Formel Gl. 2.1 [23]:

$$X^T N X M U = U \Lambda \quad (2.1)$$

wobei X^T die transformierte Matrix von X , U die Eigenvektormatrix und Λ die Eigenwertmatrix ist. M und N sind eine $p \times p$ und eine $n \times n$ Gewichtungsmatrix, die mit ihrer Dimensionalität die Metrik der Eigenwert-Eigenvektoranalyse vorgeben. Die Matrix U muss die Orthonormalisierungsbedingung [137] erfüllen mit $U^T M U = I_P$ wobei

³PCA: *engl.* Principal Component Analysis

⁴CA: *engl.* Correspondence Analysis, Korrespondenzanalyse

I_P eine Einheitsmatrix ist.

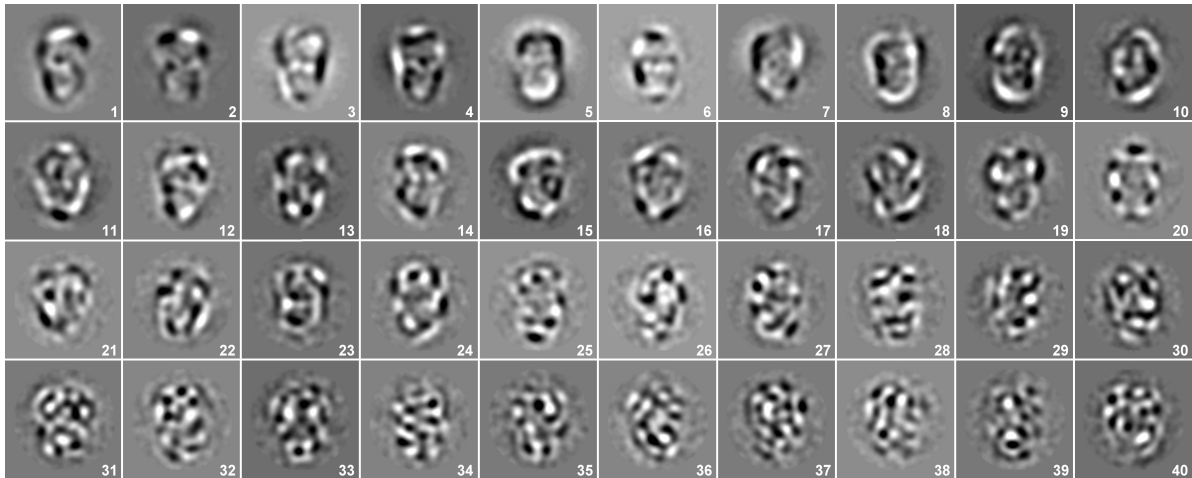


Abbildung 2.3: Eigenbilder. Die ersten 40 Eigenbilder eines Datensatzes aus einer mit Uranylformiat negativ kontrastierten Probe des U1 snRNP Komplexes aus *S. cerevisiae*.

Bildung der Klassen mit Imagic-5

Bei der anschließenden Klassifizierung benutzt das Imagic-5 Programmpaket ein hierarchisches Clusterverfahren (HAC⁵) mit einem anschließenden Verbesserungsschritt, dem sog. 'moving elements refinement'[270]. Dabei werden die Einzelpartikelbilder, die in dem durch die Hauptkomponentenanalyse aufgespannten dimensionsreduzierten Hyperraum in Nachbarschaft liegen, in eine vom Benutzer definierten Anzahl von Klassen sortiert. Für diese Berechnung bleiben nach visueller Begutachtung des Anwenders gegebenenfalls Eigenbilder, die nur Rauschen repräsentieren, unberücksichtigt. Die HAC-Methode in Imagic-5 ist unter den verschiedenen hierarchischen Clustermethoden ein Agglomerationsverfahren. So startet die Methode mit einer Anzahl von Klassen, die der Anzahl von Bildern im Datensatz entspricht, so dass zunächst jedes Bild eine Klasse bildet. Sukzessive werden benachbarte Klassen zusammengefasst bis die gewünschte Anzahl erreicht ist, wobei die Intra-Klassen-Varianz so gering und die Inter-Klassen-Varianz so groß wie möglich sein sollen. Dabei erfolgt die Berechnung des Abstands d zweier Klassen A und B nach der 'Wards'-Methode [23]. Die HAC-Methode ist rigide und lässt keine nachträgliche Änderung der Klassenzuordnung eines Bildes zu, auch wenn dies im fortschreitenden Prozess der Klassenvereinigungen für die fehlerfreie Erfüllung der beiden genannten Kriterien notwendig wäre. Um die Intra-Klassen-Varianz dennoch weiter zu minimieren und die Inter-Klassen-Varianz zu maximieren, wird daher

⁵engl. HAC: Hierachical Ascendant Classification

mit der anschließenden Anwendung der Methode *'moving elements refinement'* jedem Klassenmitglied die Möglichkeit gegeben, die Klassenzugehörigkeit zu wechseln. Dazu wird im Hyperraum der Abstand eines Bildes zu den erhaltenen Klassen berechnet und daraufhin der Klasse mit dem geringsten Abstand zugeordnet. Nach einer jeweiligen Aktualisierung wird dies für jedes Bild fortgesetzt und in einem iterativen Prozess solange wiederholt, bis keine Änderungen in der Klassenzugehörigkeit mehr getroffen werden.

Bildung der Klassen mit CowHide

Die Klassifizierungsmethode im CowHide Programmpaket unterscheidet sich von der in Imagic-5 wesentlich, lässt sich aber ebenfalls gliedern in Klassifizierung und Verbesserungsschritt [156]. Vor Beginn legt auch hier der Anwender die Anzahl der Eigenbilder fest, die in die Berechnung einbezogen werden sollen.

Die CowHide Klassifizierungsmethode ist vergleichbar mit einem K-means-Cluster-Algorithmus via Hauptkomponentenanalyse [50]. Zu Beginn der Klassifizierung werden alle Bilder zu einer einzigen Klasse zusammengefasst, und das Klassenzentrum wird bestimmt. Anschließend werden die Klassenmitglieder mittels PCA in ein orthogonales Koordinatensystem aus Eigenbildern mit Ursprung im Klassenzentrum transformiert. Die Achse der Hauptkomponente dieses Systems aus Eigenvektoren wird zur Festlegung der Teilung der Klasse herangezogen: Alle Einzelbilder, deren Hauptkomponente ≥ 0 ist, werden in einer neuen Klasse zusammengefasst und die verbleibenden Bilder fallen entsprechend in eine zweite. Um die gewünschte Anzahl von Klassen zu erreichen, wird von allen bisher auf diese Weise gebildeten Klassen stets jene mit den meisten Mitgliedern ausgewählt, und mit dieser wird analog wie beschrieben fortgefahren, also zuerst mit der Bestimmung des Klassenzentrums. Sobald 4 entstanden sind oder ein Vielfaches davon, wird zur Verbesserung der Klassifizierung ein zusätzlicher Schritt eingefügt. Das Kriterium für die Ähnlichkeit zweier Bilder ist in diesem Schritt nicht der Abstand im Hyperaum, sondern der Kreuzkorrelationskoeffizient. Jedes Einzelbild wird in die Klasse verschoben, zu deren Klassensumme es den höchsten Kreuzkorrelationskoeffizienten aufweist. Sobald durch dieses gesamte Verfahren die durch den Anwender vorgegebene Anzahl von Klassen erreicht ist, wird für die akkurate Klassenbildung ein abschließender iterativer Schritt hinzugefügt. Dazu werden die Kreuzkorrelationkoeffizienten aller Einzelbilder zu allen Klassensummen gebildet, um dementsprechend neue Zuordnungen zu treffen. Dieser Prozess wird allerdings jetzt solange wiederholt, bis es zu keiner weiteren Änderung der Klassenzugehörigkeit kommt oder bis ein Abbruchkriterium, wie im CowHide Programmpaket die feste Anzahl von Wiederholungen, erfüllt ist.

Bildung der Klassensummen

Die Mitglieder einer Klasse werden aufsummiert und zwecks Normierung wird das Ergebnis dividiert durch die Gesamtzahl der Bilder in dieser Klasse. Indem ähnliche Bilder summiert werden, wird das schwache Signal des Partikels durch die Überlagerung verstärkt, während das Rauschen im Bild nicht korreliert und damit im Vergleich vermindert wird. Die Klassensummen weisen daher ein verbessertes SNR auf.

2.2.3 Zyklische Datenprozessierung

Um stabile Ansichten von 2D Projektionen eines Proteinkomplexes zu erhalten, wird der Zyklus, bestehend aus 2D *Alignment* und Klassifizierung, mehrfach mit dem gegebenen Datensatz wiederholt (Abb. 2.4, oranger Pfeil). Aus den jeweils erhaltenen Klassensummen werden Referenzen für ein erneutes MRA ausgewählt. Der Anwender selektiert dafür charakteristische Projektionsansichten von relativ undefinierten Klassensummen mit uneinheitlichen Klassenmitgliedern. Der Zyklus wird solange wiederholt, bis keine Verbesserung der Klassensummen im Vergleich zu den vorherigen festgestellt werden kann. Durch Variation der Klassenanzahl, die in der Regel so ausgewählt wird, dass 10 – 40 Einzelbilder eine Klasse bilden, kann verhindert werden, dass seltene aber valide Projektionsansichten, beispielsweise durch präferenzielle Bindung des Komplexes an den Kohlefilm, übersehen werden. Es empfiehlt sich daher, vor Auswahl der neuen Referenzen Klassifizierungsergebnisse mit verschiedener Klassenanzahl miteinander zu vergleichen.

Sobald ein erstes 3D Modell der Struktur existiert, das mit *de novo* Rekonstruktionsmethoden (Abschn. 2.3.1) erstellt worden ist, setzt sich der iterative Charakter der Strukturberechnung zur Verbesserung des 3D Modells im sog. *Refinement* fort (Abschn. 2.3 & Abb. 2.4, blauer Pfeil).

2.3 Berechnungen auf 3D Struktur Ebene

Da durch die Aufnahmen am EM 2D Projektionsbilder entstanden sind, transportieren sie neben den 2D Ansichten auch die 3D Informationen des Objekts. Die 3D Informationen sind enthalten in den Grauwerten der Bildpixel. Diese geben die Dichte der Struktur in Projektionsrichtung des Bildes wieder. Aus diesem Grund können elektronenmikroskopische Bilder zur 3D Rekonstruktion herangezogen werden [47]. Sie sind vergleichbar mit CT⁶-Röntgenaufnahmen in der Medizin mit dem Unterschied, dass in der Tomographie die Projektionswinkel vorher bekannt sind. In der Einzelmolekül-TEM müssen die

⁶CT: Computer Tomographie

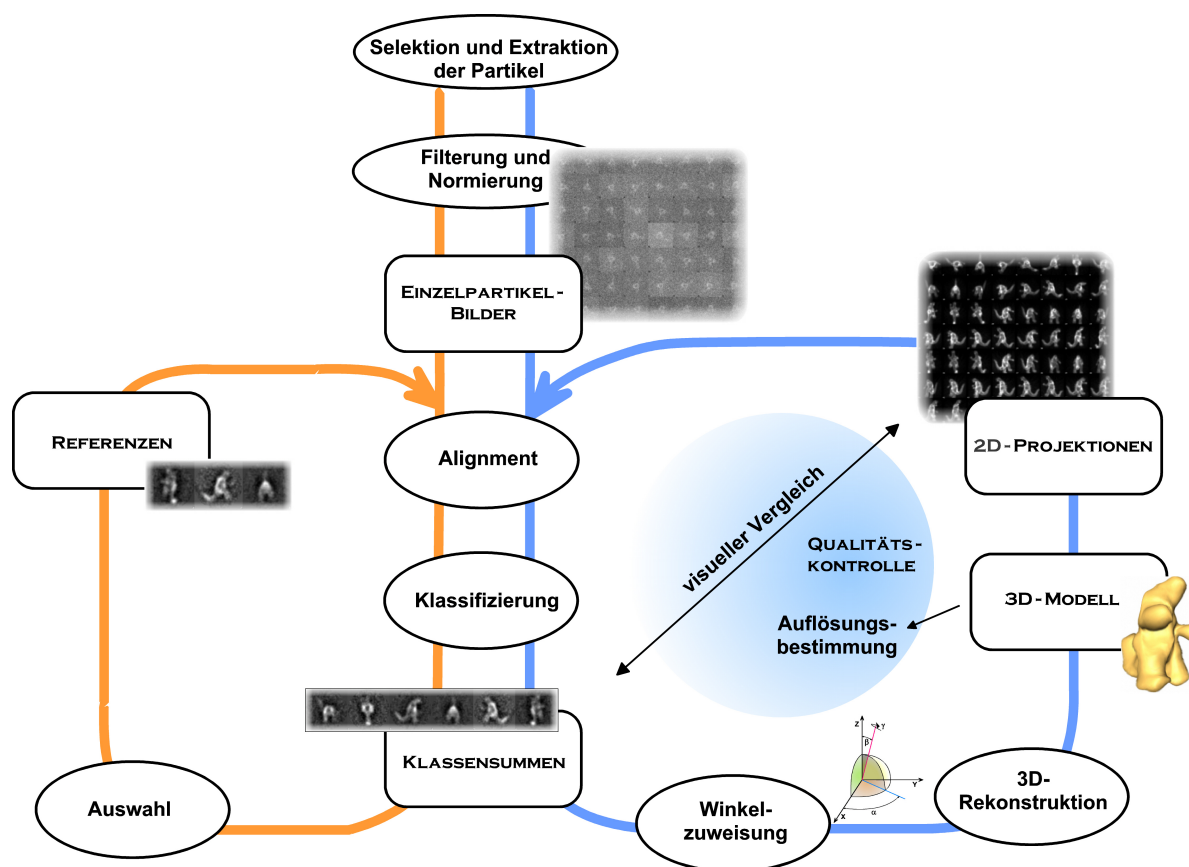


Abbildung 2.4: Übersicht der zyklischen EM-Datenverarbeitung. Nach Vorbereitung eines Datensatzes aus Einzelpartikelbildern durch Selektion und Extraktion der Partikel aus EM-Aufnahmen folgt zunächst ein Zyklus auf 2D Ebene (**oranger Pfeil, links**) bestehend aus Multi-Referenz-Alignment (Abschn. 2.2.1), PCA basierter Klassifizierung (Abschn. 2.2.2) und Auswahl der Referenzen aus den erhaltenen Klassensummen für ein erneutes Alignment. Sobald die Klassensummen durch mehrfache Wiederholung dieses Prozederes gut definiert sind, werden ihnen ihre Eulerwinkel nach einer *de novo* Rekonstruktionsmethode (Abschn. 2.3.1) zugewiesen, und es folgt die Rekonstruktion der 3D Struktur (Abschn. 2.3.4). Für weitere Zyklen zur Verbesserung der 3D Struktur (**blauer Pfeil, rechts, Refinement**) werden stets die 2D Projektionen aus dem aktuell vorhandenen 3D Modell für das MR-Alignment eingesetzt. Diese können auch zur Klassifizierung und Winkelzuweisung verwendet werden (*supervised classification*, Abschn. 2.3.3). Im anderen Fall erfolgt die Klassifizierung des Datensatzes über PCA und die Winkelzuweisung über die *angular reconstitution*-Methode (Abschn. 2.3.3). Zur Kontrolle werden 2D Projektionen mit den erhaltenen Klassensummen visuell verglichen, um gleiche Ansichten zu erkennen. Abschließend wird die Auflösung der 3D Struktur angegeben.

Projektionswinkel *a posteriori* zugewiesen werden. Um eine 3D Struktur aus den über 2D *Alignment* und PCA basierter Klassifizierung erhaltenen Klassensummen zu berechnen, müssen die Projektionsrichtungen der einzelnen Klassensummen ermittelt werden. Sobald dies erreicht ist, kann die Projektion $3D \rightarrow 2D$ bei Entstehung des EM-Bildes mit Hilfe eines Rückprojektionsverfahrens umgekehrt werden (Abschn. 2.3.4), so dass aus den 2D Projektionen die 3D Struktur rekonstruiert werden kann. Um die Projektionsrichtung zu beschreiben, sind in der EM die Angabe der Eulerwinkel α , β und γ gebräuchlich, da mit deren Hilfe eine beliebige Projektionsrichtung eines 3D Objekts beschrieben werden kann. Der eulerschen (z, x, z') Konvention folgend, lässt sich damit der α Winkel in einem initialen karthesischen Koordinatensystem (x, y, z) als Rotation um die z -Achse, der β Winkel als eine anschließende Drehung um die neue x -Achse und der γ Winkel als eine Rotation um die nach den ersten beiden Rotationen erhaltene z' -Achse beschreiben. Diese Konvention wird im CowHide Programmpaket [156] verwendet, mit dem die RCT-3D-Rekonstruktion (Abschn. 2.3.1) des U1 snRNPs durchgeführt worden ist. Im Gegensatz dazu wird im Imagic-5 Programmpaket eine Eulerwinkeldefinition verwendet, die den Kippwinkel β um die y -Achse dreht. In Abb. 2.5 ist eine Eulerwinkeldefinition veranschaulicht dargestellt.

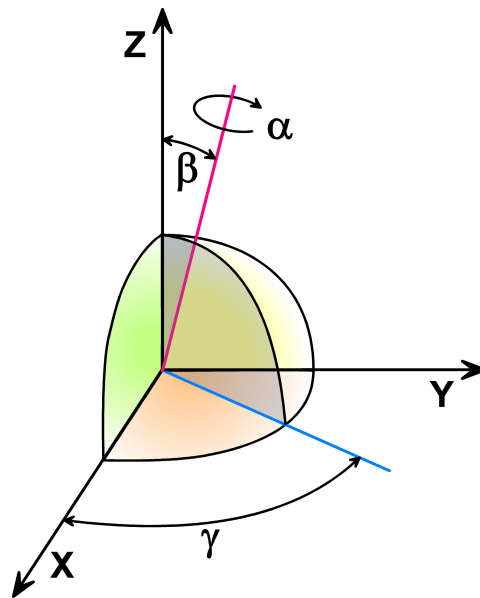


Abbildung 2.5: Die Definition der Eulerwinkel. Die 2D Projektionsrichtungen eines 3D Objektes werden mit den 3 Eulerwinkeln α , β und γ beschrieben, die nacheinander ausgeführt werden. Der β - und γ -Winkel legen den Längen- und Breitengrad fest, womit die neue z -Achse (rot) definiert wird. Der α -Winkel beschreibt die Rotation des Projektionsbildes in der Bildebene um die Hauptachse (z/z' -Achse) des Partikels.

2.3.1 *de novo* Rekonstruktion der 3D Struktur

Um eine neue unbekannte asymmetrische 3D Struktur zu rekonstruieren, ist die *Random Conical Tilt*⁷ - Methode (kurz: RCT) nach [198] eine geeignete *de novo* Rekonstruktionsmethode. Prinzipiell ist es zwar auch möglich, die *angular reconstitution* - Methode nach [271] zu verwenden, wie für die 3D Rekonstruktion des humanen U1 snRNPs geschehen [249]. Dies setzt allerdings eine gleichmäßige Winkelverteilung der Projektionsansichten voraus. Darüber hinaus dürfen keine strukturellen Heterogenitäten im Datensatz vorhanden sein, da zweifelsfrei sein muss, dass die 2D Projektionen von einer 3D Struktur stammen und nicht unterschiedlich konfigurierte 3D Strukturen repräsentieren. Auch unter diesen Voraussetzungen erfordert die *angular reconstitution* - Methode mehr Erfahrung des Anwenders und ist daher für asymmetrische Komplexe fehleranfälliger. Für symmetrische Proteinkomplexe bei denen die Symmetrieart leicht ermittelbar ist, bietet diese Methode hingegen ein vorteilhaftes automatisches Prozedere basierend auf der Ermittlung von Maxima in Sinogramm Korrelationsfunktionen (Abschn. 2.3.3).

Dagegen lässt die RCT-Methode (s. u.) die Rekonstruktion verschiedener 3D Modelle aus einem heterogenen Datensatz zu, so dass mit Hilfe dieser Technik, Proteinkomplexe auf ihre strukturellen Konformationen und auch Konfigurationen untersucht werden können.

Random Conical Tilt - Methode (RCT)

Für diese 3D Rekonstruktionstechnik ist es notwendig, dieselben Proteinkomplexe bei 0° und unter einem festgesetzten Winkel von $\approx 50^\circ$ aufzunehmen. Letzteres ist durch Kippung des Probenhalters im EM möglich, wobei der maximale Kippwinkel technisch allerdings bei $\approx 60^\circ$ begrenzt ist. Der besondere Vorteil dieser Aufnahmetechnik ist, dass 2 von 3 Eulerwinkeln durch die experimentellen Einstellungen während der EM-Aufnahmen festgelegt sind und der fehlende Eulerwinkel aus der Rotation des Einzelpartikelbildes im 2D *Alignment* erhalten wird. Die Händigkeit der 3D Struktur ist ebenfalls bei dieser Methode determiniert.

Die Einzelpartikelbilder aufgenommen bei 0° werden über das beschriebenen Prozedere aus 2D *Alignment* und PCA-basierter Klassifizierung in Klassen identischer Projektionswinkel gruppiert. Die Eulerwinkel werden anschließend für jede einzelne Klasse den entsprechenden Einzelbildern aus der gekippten Aufnahme zugewiesen (Abb. 2.6). Für die Winkelangaben werden die Informationen von (1) Kippachse, (2) Kippwinkel und vom (3) Rotationswinkel des letzten 2D *Alignments* der Einzelpartikelbilder der 0°-Aufnahmen verwendet. Die RCT-Strukturen sind zwar erste 3D Strukturen des Objekts, allerdings in sehr niedriger Qualität aufgrund von 3 Problemen. (1) Da der Kippwinkel

⁷RCT: *Random Conical Tilt*

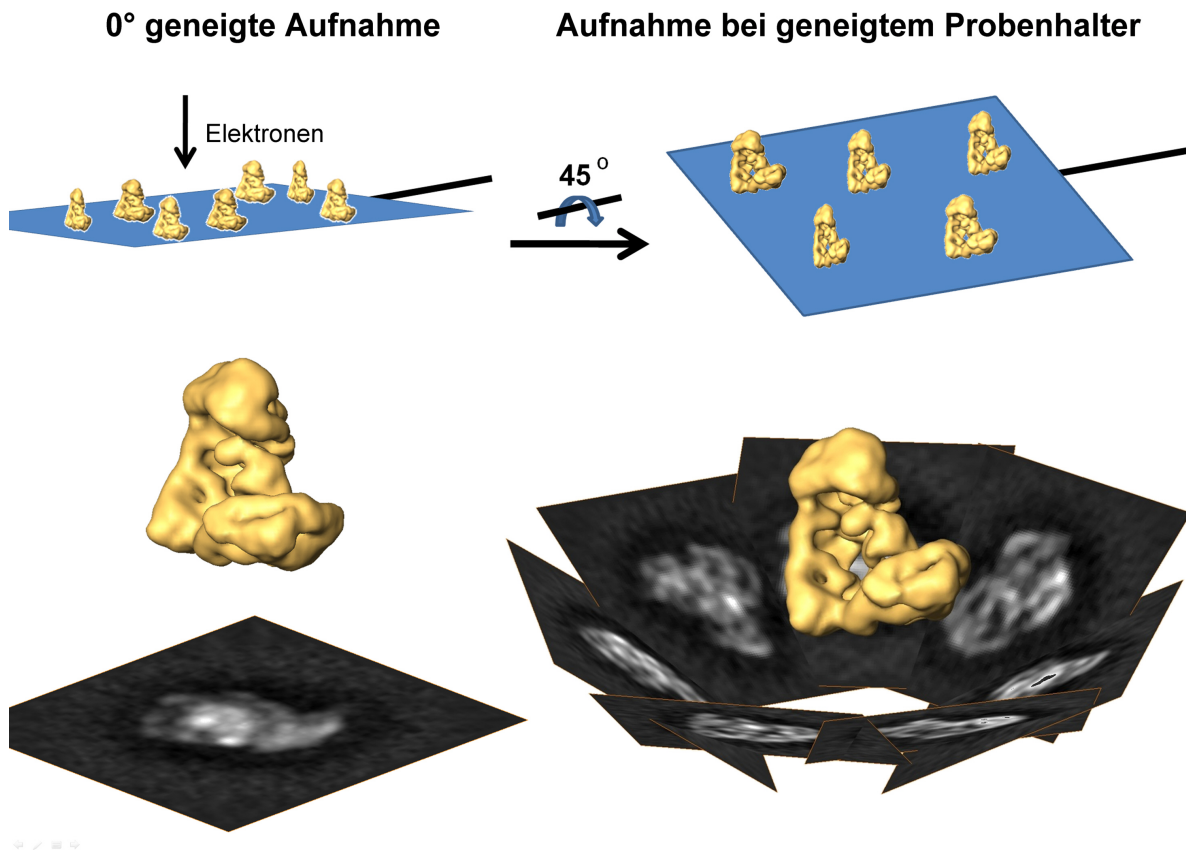


Abbildung 2.6: Die RCT-Methode. Der Proteinkomplex, in diesem schematisierten Beispiel der APC Komplex [56], wird sowohl bei 0°, als auch bei 45° geneigtem Probenhalter aufgenommen. Von den Einzelbildern bei 0° werden in einem zyklischen Verfahren aus 2D *Alignment* und Klassifizierung Klassensummen gebildet. Von den Einzelbildern, die in eine Klasse eingeordnet wurden, werden die korrespondierenden 45° geneigten Einzelbilder benutzt, um die RCT-3D-Struktur zu rekonstruieren. Bild freundlich zur Verfügung gestellt von Prof. H. Stark.

begrenzt ist, sind die RCT-3D-Strukturen in einem Bereich eines Konus um den eulerischen β Winkel von $\approx 90^\circ$ nicht definiert, ein Phänomen, auf das in der Literatur auch mit *missing cone* verwiesen wird. Desweiteren ist (2) das SNR des RCT-3D-Volumens gering, da nur wenige Projektionen für die Rekonstruktion einer RCT-Struktur verwendet wurden, nämlich nur jene, die zuvor in eine Klasse sortiert wurden. Ferner wurden (3) die Proteinkomplexe für die gekippte Aufnahme ein zweites Mal dem Elektronenstrahl exponiert und ein Strahlenschaden aus der ersten Aufnahme wirkt sich mindernd auf die Bildqualität der zweiten Aufnahme aus. Durch ein anschließendes Maximum Likelihood *Alignment* (Abschn. 2.3.2) können diese Einschränkungen ausgeglichen werden.

2.3.2 Maximum Likelihood Methode (ML)

Ziel der Anwendung der Methode des Maximum Likelihood *Alignments*⁸ [237], [218] ist die Summierung von mehreren RCT-Strukturen, die die gleiche 3D Struktur repräsentieren, aber ihren *missing cone* an anderer Stelle tragen. Dadurch lässt sich das geringe SNR einzelner RCT-Strukturen verbessern und das *missing cone* - Phänomen ausgleichen. Ein auf Kreuzkorrelation basiertes *Alignment* ist dafür ungeeignet, da dadurch lediglich die *missing cones* aufeinander ausgerichtet würden. Mit Hilfe des ML-*Alignment* aber lassen sich verlässliche Strukturmodelle auch aus heterogenen Datensätzen erhalten [221]. Eine einzelne RCT-Struktur repräsentiert die 3D Rekonstruktion einer bestimmten Konfiguration und Konformation eines Proteinkomplexes. Angenommen der Proteinkomplex läge in einer einheitlichen Zusammensetzung vor und wäre rigide in Bezug auf seine Konformation, so gehörten alle berechneten RCT-Strukturen zu derselben 3D Struktur und könnten aufsummiert werden, ohne dass es zu Informationsverlust und damit verbundenen Fehlern führte. Datensätze von dynamischen Proteinkomplexen repräsentieren jedoch den Komplex in verschiedenen Assemblierungsstadien und konformationellen Funktionszuständen, so dass die Vielzahl an errechneten RCT-Rekonstruktionen zu einer gewissen Anzahl unterschiedlicher 3D Strukturen gehören. Das ML-*Alignment* wird daher benutzt, um die RCT-Strukturen im 3D Volumen auszurichten, bevor sie ähnlich dem Verfahren auf 2D Ebene in Gruppen klassifiziert und summiert werden können. Zugunsten der Geschwindigkeit wurde das ML-*Alignment* beschränkt auf die Rotationsausrichtung der RCT-Strukturen [214], zumal die 3D Strukturen im Volumen aufgrund des vorherigen Prozederes (2.2) ausreichend massenzentriert sein sollten. Das ML-*Alignment* basiert auf der Bestimmung eines Gewichtungsfaktors [237] für alle einzelnen RCT-Volumina, um damit eine Referenz für das folgende *Alignment* zu bilden. Als initiale Referenz dient ein 3D Volumen aus Rauschen, so dass kein Bias eingebracht

⁸ML-*Alignment*: Maximum-Likelihood-*Alignment*

wird. In den weiteren Alignmentrunden wird dann jeweils eine Referenz verwendet, die aus der gewichteten Summierung aller RCT-Strukturen am Ende eines Alignmentzyklus erhalten wird. Das auf einem Computercluster implementierte ML-Protokoll iteriert standardmäßig über 40 Alignmentzyklen, wobei die Ausrichtung der RCT-Strukturen mit abnehmenden Winkelabständen von anfänglich 15° in der ersten Alignmentrunde bis zu $1,8^\circ$ für die feine Rotationsausrichtung ausgeführt wird. Nach Beendigung des ML-Alignments wird eine Klassifizierung der RCT-Strukturen via Hauptkomponentenanalyse (3D PCA) angefügt. Die Anzahl der RCT-Strukturen pro Klasse beträgt in der Regel 5 – 8. Auf diese Weise werden verschiedene Konformationen und/oder Konfigurationen des Proteinkomplexes sortiert und durch Aufsummierung der zu einer Gruppe gehörenden RCT-Strukturen aufgrund eines verbesserten SNR erkennbar. Die Klassifizierung der RCT-Strukturen kann durch entsprechende Maskierung auch lokal auf eine Region oder Domäne der Struktur beschränkt werden, um auf diese Weise die Heterogenität der Probe zu untersuchen.

2.3.3 Verbesserung der 3D Struktur

Auch auf 3D Ebene setzt sich die zyklisch angelegte EM-Bildverarbeitung im sog. *Refinement* der Struktur fort. Das *Refinement* beruht auf der Verbesserung der Eulerwinkelbestimmung der zur Rekonstruktion zu verwendeten Projektionsbilder, wofür zwei Methoden, die *angular reconstitution*-Methode [271] und die *projection-matching*-Methode [186], zur Verfügung stehen.

Eulerwinkelzuweisung via *angular reconstitution*

Die *angular reconstitution*-Methode [271] ist in das Imagic-5 Programmpaket implementiert und basiert auf dem Theorem der gemeinsamen Linienprojektion im Fourierraum. Grundlegend für diese Projektionstheorie [47], [43] ist die Tatsache, dass die mathematische Operation der Fouriertransformation zu keinem Verlust oder Verzerrung der im 3D Volumen oder Projektionsbild enthaltenen Informationen führt. Daher ist die Beschreibung eines 3D Objekts oder 2D Bildes im Fourierraum äquivalent zur jeweiligen Darstellung im Bildraum (Abb. 2.7). Im Fourierraum entspricht ein beliebiges 2D Projektionsbild eines 3D Objekts einer bestimmten Schnittebene durch das Zentrum. Folglich schneiden sich die Ebenen von 2 verschiedenen Projektionsbildern, erhalten vom selben 3D Objekt, im Fourierraum in einer gemeinsamen Linie, der sog. *common line*. Dieses Erkenntnis lässt sich auch auf den realen Bildraum übertragen: Zwei 2D Projektionsbilder desselben 3D Volumens teilen eine identische 1D Linienprojektion. Die *angular reconstitution*-Methodik ermittelt die gemeinsame Linienprojektion

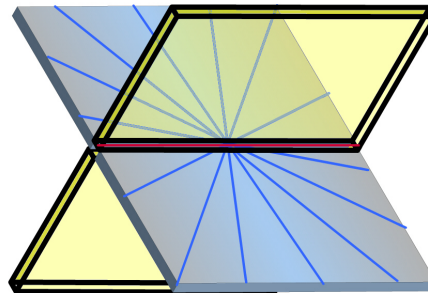
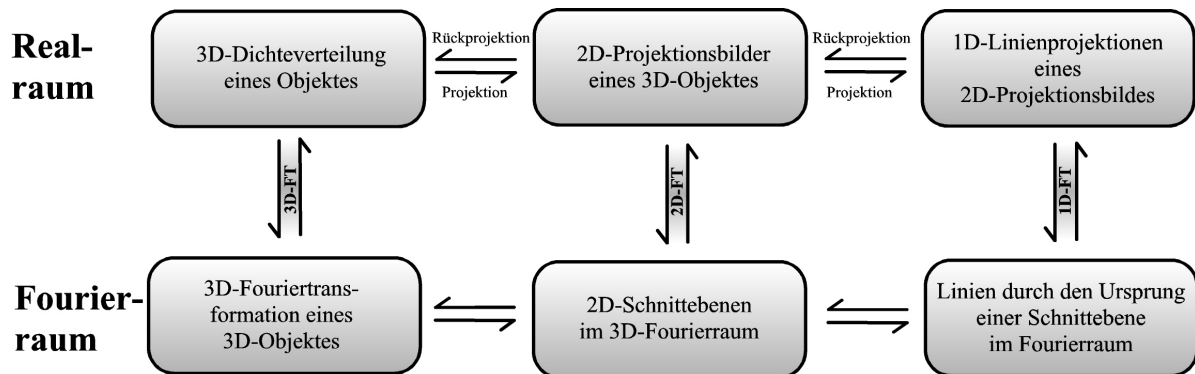


Abbildung 2.7: Äquivalenz von Bild- und Fourierraum. Grundlegend der Projektionstheorie [47], [43], nach der es möglich ist, aus Projektionsbildern desselben 3D Objektes das Objekt selbst zu rekonstruieren, ist die Äquivalenz von Bild- und Fourierraum. Im Fourierraum teilen sich 2 Projektionsansichten, dort sind es Schnittebenen durch den Ursprung, eine gemeinsame Linie, die sog. *common line* (untere Abb. rote Linie).

im Bildraum (Abb. 2.8). Zunächst werden von einem Projektionsbild 360 1D Projektionslinien berechnet. Dazu werden die Grauwerte der Bildpixel (die Dichte im 3D Raum) in einem feststehenden Koordinatensystem in einer definierten Richtung aufsummiert, um die Funktionswerte der ersten Linienprojektion zu erhalten. Anschließend wird das Bild um 1° um eine Achse vertikal zur Bildebene gedreht, wobei das Koordinatensystem und die damit verbundene Projektionsrichtung beibehalten wird. Die Funktionswerte der zweiten Linienprojektion werden somit gleicherweise durch Aufsummierung berechnet. Insgesamt werden auf diese Weise 360 Projektionslinien, einer kompletten Rotation des Bildes entsprechend, erhalten und, aufgelistet in einem sog. Sinogramm, abgebildet (Abb. 2.8, B & D). Das Sinogramm hat somit eine Höhe von 360 Pixeln und die entsprechende Breite des Projektionsbildes (Abb. 2.8, A & C), wobei die Grauwerte der Pixel im Sinogramm den jeweiligen Funktionswert der Linienprojektion angeben. Um nun von 2 verschiedenen Projektionsbildern die gemeinsame Projektionslinie zu ermitteln, werden von beiden Projektionsbildern die Sinogramme gebildet. Anschließend werden die beiden Sinogramme verglichen, indem die jeweiligen Kreuzkorrelationskoeffizienten sämtlicher Kombinationen von 1D Projektionslinien des einen mit denen des anderen Bildes berechnet werden. Als Ergebnis werden die Kreuzkorrelationswerte als Grauwerte in einem Diagramm, dem Sinogramm-Korrelationsbild (Abb. 2.8, E), ausgegeben. Dieses Diagramm ist demnach 360×360 Pixel groß. Die Maxima der Kreuzkorrelationskoeffizienten geben die Lage der gemeinsamen 1D Linienprojektion in den 2D Projektionsbildern an. Während von einem asymmetrischen Objekt 2 Maxima in der Sinogramm-Korrelationsfunktion erhalten werden, richtet sich für symmetrische Objekte die Anzahl der Peaks nach der Art der Symmetrie. Da sich die Linienprojektionen jenseits von 180° gespiegelt wiederholen, ist die Anzahl aber immer ein Vielfaches von 2.

Sobald eine 3D Struktur existiert, werden von dieser gleichmäßig, in diskreten Eulerwinkelabständen, über die gesamte Einheitskugel Projektionen berechnet. Diese 2D Projektionen weisen ein besseres SNR, verglichen mit den Klassensummen, auf und können auch dazu dienen, seltene Projektionsansichten zu erkennen. Diese Projektionen werden in einem erneuten MR-*Alignment* als Referenzen eingesetzt. Der ausgerichtete Datensatz wird anschließend über die PCA-Methode klassifiziert, und es werden so neue Klassensummen erhalten. Um diesen Klassensummen mittels *angular reconstitution* Eulerwinkel zuzuweisen, werden nun Projektionen von einem vorhandenen 3D Modell der Struktur mit ihren bekannten Eulerwinkel als ein Winkelreferenzset, als sog. *anchor set*, eingesetzt. Nachdem von Klassensummen und 2D Projektionen Sinogramme berechnet worden sind, werden diese über Sinogramm-Korrelationsfunktionen miteinander verglichen, um die gemeinsame Linienprojektionen zu finden. Da die Winkelbeziehungen bei den *anchor set* Projektionen bekannt sind, lassen sich mit diesen Informationen die

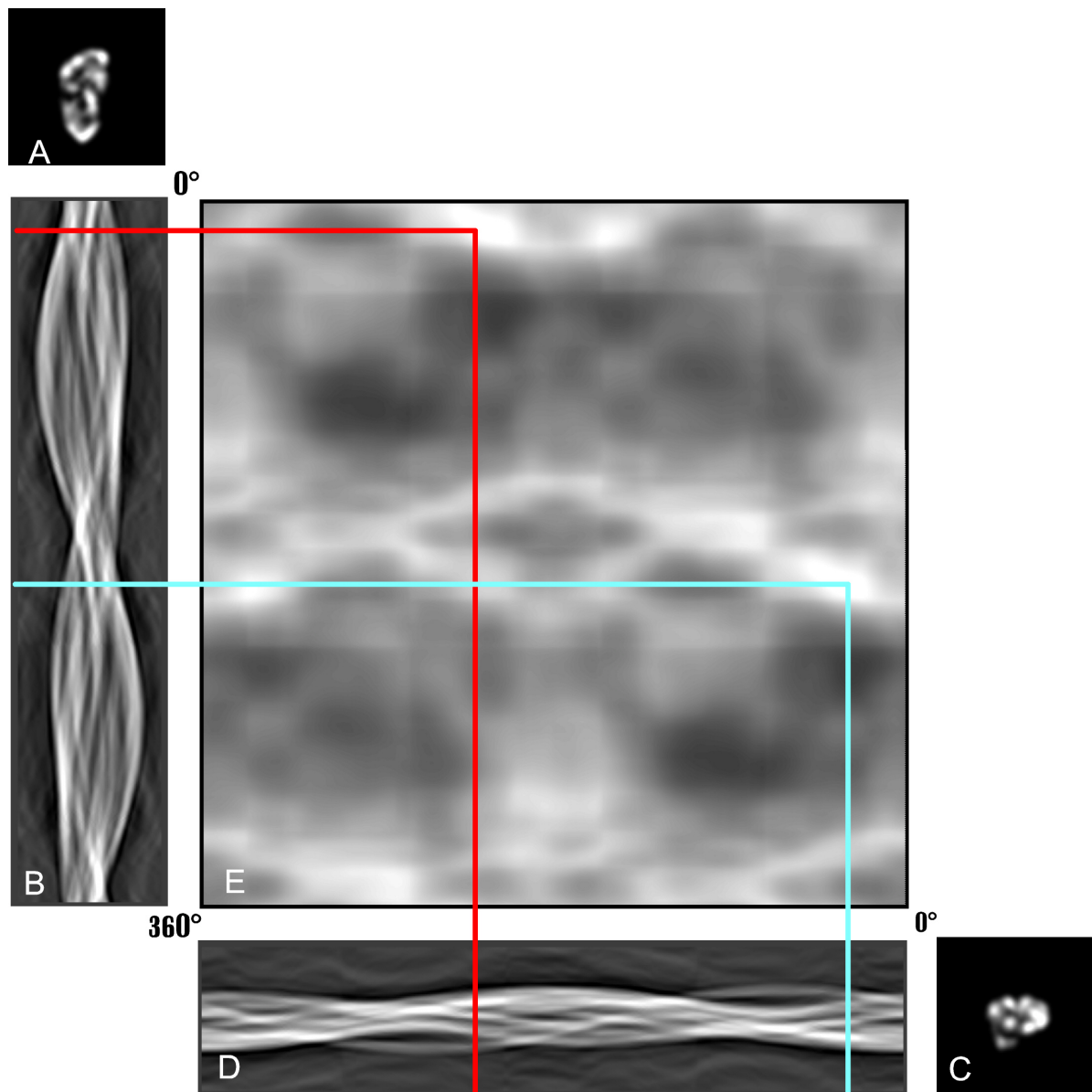


Abbildung 2.8: Ermittlung der gemeinsamen Linienprojektion. In (A) und (C) sind 2 verschiedene Projektionsbilder derselben Struktur vom U1 snRNP dargestellt. Die 360 Linienprojektionen dieser Bilder entstanden durch eine vollständige Rotation des Bildes und der jeweiligen Projektion in 1° Abständen. Untereinander gelegt ergeben die Linienprojektionen das jeweilige Sinogramm (B) und (D) des Projektionsbildes. Der Vergleich der Linienprojektionen von (B) mit (D) erfolgt über die Berechnung der Kreuzkorrelationskoeffizienten sämtlicher Kombinationen, so dass als Ergebnis eine Sinogramm-Korrelationsfunktion erhalten wird. Diese wird in einem Diagramm (E) dargestellt, wobei die Werte der Kreuzkorrelationskoeffizienten in den Grauwerten der Pixel wiedergegeben werden. Helle Pixel sind hohe Werte für den Kreuzkorrelationskoeffizienten. Von einem asymmetrischen Objekt werden 2 Maxima erhalten, da sich die Linienprojektionen jenseits von 180° gespiegelt wiederholen. Für symmetrische Objekte richtet sich die Anzahl der Peaks nach der Art der Symmetrie, ist aber immer ein Vielfaches von 2.

Eulerwinkel der Klassensummen bestimmen. Als Winkelreferenzset können auch Projektionen aus einer 3D Struktur dienen, die direkt nach dem MR-*Alignment* zunächst via *projection matching* rekonstruiert worden ist.

Eulerwinkelzuweisung via *projection-matching*

Im *projection-matching*-Verfahren [186] wird die PCA-basierte Klassifizierung mit anschließender Winkelzuweisung umgangen. In diesem Verfahren werden die Informationen aus dem MR-*Alignment* genutzt, um den Datensatz zu klassifizieren. Es werden jeweils alle jene Einzelpartikelbilder des Datensatzes summiert, welche nach Ermittlung des höchsten Korrelationskoeffizienten entsprechend nach demselben Referenzbild ausgerichtet wurden. Die Anzahl der Klassen/Klassensummen ist damit kleiner oder gleich der Anzahl der in dem MR-*Alignment* eingesetzten Referenzen. Da die Eulerwinkel der verwendeten Referenzbilder mit dem Ausprojizieren aus dem 3D Modell festgelegt worden waren, stehen somit auch die Eulerwinkel für die so kalkulierten Klassensummen fest. Da auf diese Weise in den Klassensummen nur Partikelansichten erhalten werden können, die auch in den Referenzbildern auftreten, wird diese Art der Klassifizierung kontrolliert und dementsprechend als *supervised classification* bezeichnet. Dagegen wird für die Klassifizierung mittels PCA keine Information über das Referenzbild benötigt und die Klassifizierung ist frei von Einflüssen, die sogenannte *unsupervised classification*. Aufgrund dieser Unterschiede eignet sich das Verfahren *projection-matching* für die Verbesserung der Auflösung der 3D Struktur, sobald ein glaubwürdiges 3D Modell existiert. Hingegen ist die oben beschriebene Methode über *angular reconstitution* mit Winkelreferenzset vorzuziehen, wenn die 3D Struktur noch getestet werden sollte, also vornehmlich zu Beginn eines Struktur-*Refinements*. Im letzteren Fall kann die über *projection-matching* kalkulierte Struktur zur Erstellung eines Referenzwinkelsets dienen.

2.3.4 Rekonstruktion der 3D Struktur via Rückprojektion

Um die 3D Struktur zu rekonstruieren, werden alle vorhandenen Projektionsbilder in ihre jeweilige, durch die Eulerwinkelbestimmung vorgegebene, Richtung in das 3D Volumen 'ausgeschmiert'. Dies geschieht vertikal zur Bildebene unter Einbeziehung aller Pixel. Die Grauwerte im Bild geben die Gesamtdichte des Objekts an der jeweiligen Stelle wieder und sie wird gleichmäßig in den Pixelprojektionsraum verteilt (Abb. 2.9). Die Dichte im Gesamtvolumen wird dort hoch, wo sich 2 oder mehr dieser Pixelprojektionsräume schneiden, vornehmlich im Zentrum des Kubus. Da diese Situation äquivalent im Fourierraum ist, kreuzen sich dort die zentralen Schnitte in der Region niedriger Ortsfrequenzen, wodurch die niedrigen Ortsfrequenzen überrepräsentiert werden. Um

diese niedrigen Ortsfrequenzen schwächer zu gewichten, ist im Imagic-5 Softwarepaket ein spezieller *exact filtered back projection*-Algorithmus implementiert. Dieser Algorithmus berechnet individuell für jedes Projektionsbild eine Filterfunktion, und zwar unter Berücksichtigung sowohl der Gesamtzahl der zur 3D Rekonstruktion verwendeten Projektionen, als auch der Anordnung der im Datensatz vorhandenen Eulerwinkel zueinander [273].

Von der so rekonstruierten 3D Struktur können 2D Projektionsbilder in gleicher Richtung berechnet werden, wie sie für die Rückprojektion verwendet wurden. Die Qualität und der Fehler der 3D Rekonstruktion kann durch Vergleich der ursprünglich verwendeten mit den erhaltenen 2D Projektionen via Korrelationsberechnungen abgeschätzt werden.

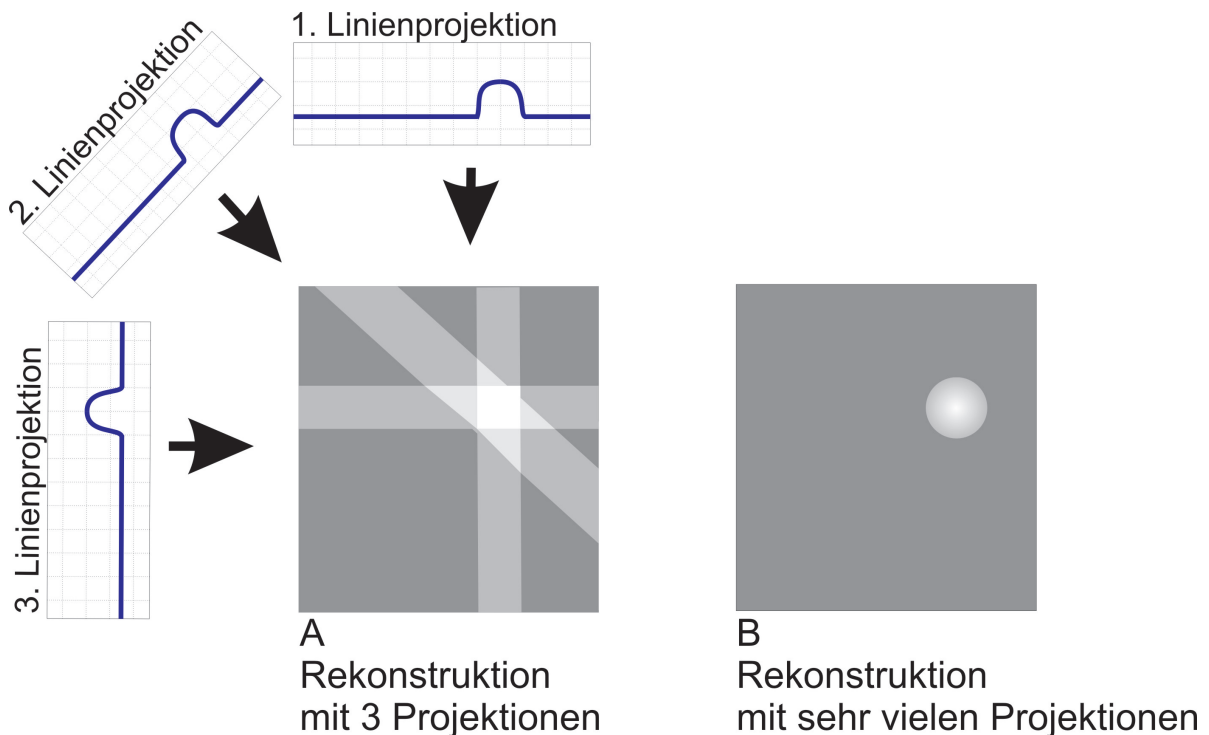


Abbildung 2.9: Illustration des Rückprojektionsverfahrens. (A) Rekonstruktion eines 2D Bildes aus 3 seiner Linienprojektionen. (B) Die Rekonstruktion des Bildes wird detaillierter, je mehr Projektionen zur Verfügung stehen. Gleichzeitig wird in (A) die Limitierung des Verfahrens bei geringer Anzahl an zur Verfügung stehenden Projektionen deutlich und entsprechend die Auswirkung auf die erreichbare Auflösung des zu rekonstruierenden Objekts. Das Rückprojektionsverfahren von 2D Ebene in ein 3D Volumen ist analog und im Text beschrieben. Bild freundlich zur Verfügung gestellt von Dr. M. Lüttich.

2.3.5 Visualisierung der 3D Struktur

Um eine 3D EM Struktur abzubilden, wird ein Schwellenwert für die im 3D Volumen darzustellende Dichte angegeben. Dichten, deren Werte unterhalb dieses Schwellenwertes liegen, gelten als Rauschen und werden nicht visualisiert. Gezeigt wird letztlich eine Oberflächendarstellung der Struktur. Über das Gesamtvolumen der Struktur und unter Angabe der durchschnittlichen Dichte eines Proteins kann die Masse des Proteinkomplexes abgeschätzt werden. Auf diese Weise wird die theoretische Masse des Komplexes bei Festlegung des Dichteschwellenwertes zumindest berücksichtigt.

Für die Visualisierung der 3D EM Struktur wurde das Programm Amira 4.1 (TGS Europe, Merignac Cedex, F) verwendet mit einem Plug-in, der das direkte Einlesen des in Imagic-5 gebräuchlichen Dateiformats erlaubte (Dr. M. Schmeißer, im Hause). Angewendet wurde bei der Strukturdarstellung in der Regel in diesem Programm ein 3D Gaussfilter und gegebenenfalls eine 3D Kontur, um Rauschen zu entfernen.

2.3.6 Bestimmung der Auflösung einer 3D EM Struktur

Um die Auflösung einer 3D EM Struktur zu bestimmen, sind mehrere unterschiedliche Kriterien vorgeschlagen worden, die auf der Bewertung der Auflösung mit Hilfe der Fourier-Schalen-Korrelation (*Fourier-shell-correlation*, FSC) [273] beruhen. Um die Fourier-Schalen-Korrelationsfunktion zu bestimmen, werden die zur 3D Rekonstruktion verwendeten Klassensummen zufällig in zwei gleich große Gruppen geteilt, und sodann werden zwei 3D Strukturen unabhängig voneinander errechnet. Anschließend wird die Korrelation der beiden Strukturen im Fourierraum ermittelt, und zwar als Funktion des Radius (S). Anders formuliert, es werden die beiden 3D Strukturen im Fourierraum Schale für Schale miteinander verglichen (Gl. 2.2). F_1 ist die Fourier transformierte Form der einen 3D Struktur, F_2 entsprechend die der anderen, wobei F_2^* die konjugiert komplexe Form von F_2 kennzeichnet:

$$FSC(S) = \frac{\sum_{R=S} F_1(R)F_2^*(R)}{\sqrt{\sum_{R=S} |F_1(R)|^2 \sum_{R=S} |F_2(R)|^2}} \quad (2.2)$$

mit R für Raumfrequenz. Da nach der Fouriertransformation die niedrigen Raumfrequenzen (R) im Zentrum liegen und die hohen außen, werden feinere Strukturdetails auf den Schalen mit großen Radien beschrieben. Für die Auflösungsbestimmung werden jene äußersten Schalen betrachtet, die sich der Korrelationsberechnung zufolge bei beiden 3D Strukturen noch signifikant ähneln. Dort ist die Auflösungsgrenze erreicht, wobei die Auflösung dem reziproken Wert der Raumfrequenz entspricht. In einer graphischen Darstellung wird der FSC-Wert über ansteigendem Radius (S) (entsprechend

Raumfrequenz) aufgetragen. Nun existieren mehrere Möglichkeiten, einen Wert von der FSC Kurve abzulesen, um einen Auflösungsgrenzwert festzulegen. Für asymmetrische Objekte wird häufig ein Wert nach dem $FSC_{0,5}$ -Kriterium angegeben wie in [12], [24]. Dies ist jene Ortsfrequenz, bei der der Funktionswert der FSC-Kurve 0,5 beträgt. Es handelt sich dabei um einen fixierten Wert, der auf Erfahrung beruht und so konservativ angelegt ist, dass dessen Angabe selten zur Überschätzung der erreichten Auflösung führt. Ein anderer mittlerweile eingeführter fixierter Wert ist das $FSC_{0,143}$ Kriterium [207]. Anspruch war, einen Wert zu haben, der realistischer ist und eine Vergleichbarkeit zu Röntgenkristallstrukturen erlaubt. Da die FSC Kurve allerdings abhängig von der Anzahl der Voxel einer Schale ist, wurden Schwellenwertkurven, beruhend auf dem statistischen σ Faktor, eingeführt und das bereits für das 2D Äquivalent der FSC, der FRC^9 [220], Review [274].

Verbreitet ist danach die Angabe der Auflösungsgrenze nach dem sogenannten 3σ -Kriterium [177]. Hierbei wird zunächst eine Funktion bestimmt, die 3σ -Kurve, die einen Schwellenwert für die Ähnlichkeit zweier 3D Strukturen in Abhängigkeit von der Raumfrequenz angibt. Diese Kurve wurde so gewählt, dass auch statistische Einflüsse von unkorreliertem Rauschen berücksichtigt werden. Bei zufälligem Rauschen in 2 verschiedenen Volumina nämlich werden diese trotzdem eine gewisse Korrelation aufweisen. Erst wenn der stochastisch berechenbare Erwartungswert σ bei einem zufälligen Rauschen in beiden 3D Strukturen in einem bestimmten Frequenzbereich (Schale) um das 3-fache überschritten wird (3σ), liegt eine signifikante Übereinstimmung vor, da in der Regel (in 99% der Fälle) Paare von rauschigen Volumina keine Korrelation in dieser Größenordnung mehr aufweisen. Als Auflösungsgrenze der 3D Struktur wird daraus folgend der Schnittpunkt der 3σ -Kurve mit der $FSC(S)$ -Kurve angegeben. Ebenso wird bisweilen das analog gebildete 5σ -Kriterium aufgeführt. Neben der Ablesung der Auflösungsgrenze kann die $FSC(S)$ -Kurve bei einem untypischen Verlauf auf Probleme in der Bildverarbeitung hindeuten.

Neben diesen Methoden einer Berechnung der Auflösung der Struktur können andere Kriterien hinzugezogen werden, um die Qualität und Richtigkeit einer 3D Struktur einzuschätzen. So ist beispielsweise ein gutes SNR der Klassensummen bei einer durchschnittlichen Mitgliederzahl von 10 – 30 Einzelpartikelbildern in einer Klasse eine gute Voraussetzung für eine nachfolgende möglichst präzise Eulerwinkelzuordnung. Desweiteren sollten die für die Rückprojektion verwendeten Klassensummen den 2D Projektionen aus der entstandenen 3D Struktur möglichst ähnlich sein.

⁹FRC: Fourier-Ring-Correlation

Kapitel 3

Materialien und Methoden

3.1 Materialien

3.1.1 Computer Software und spezielle Hardware

Software	Bezugsquelle
ActivePerl	www.activestate.com/activeperl/ (13.Juli.2009)
AmiraDev 4.1	TGS Europe, Merignac Cedex, France
Boxer from Eman Suite	www.blake.bcm.tmc.edu/eman/ (13.Juli.2009)
CowHide (β -Versionen)	Abt. Stark, im Hause
ImageMagic	www.imagemagick.org (13.Juli.2009)
Imagic-5	Image Science, Berlin, D
Maximum Likelihood für RCT Strukturen	Björn Sander, Göttingen, D
PyMol 0.99	www.desci.com/rel/099/ (14.10.2009)
Python 2.5.1	www.python.org/download/releases/2.5.1/
SIGNATURE	www.brandeis.edu/jzchen/signature (Juli 2009) [37]
TVIPS EM-MENU	Tietz Video Systems, Gauting, D
Xmipp	www.xmipp.cnb.csic.es (13.Juli.2009) [245]

Spezielle Hardware	Beschreibung
Datenbankserver (ML-Alignment)	2 <i>x</i> Dual Core AMD Opteron, 2 <i>GHz</i> , 8 <i>GB</i> RAM, Win2003 Server, 64bit+Fiberchannel RAID
Rechnenode (ML-Alignment)	12 <i>x</i> Dual Core AMD Opteron, 2 <i>GHz</i> , 2 <i>GB</i> RAM, Win XP

Tabelle 3.1: Verwendete Computersoftware und spezielle Hardware

3.1.2 Chemikalien und Pufferlösungen

Alle Puffer wurden mit filtriertem deionisiertem Wasser angesetzt und vor der Verwendung steril filtriert.

Puffer	Zusammensetzung
AGK Puffer	20 mM HEPES/KOH <i>pH</i> 7,9, 200 mM KCl, 1,5 mM MgCl ₂ , 10 % (<i>w/w</i>) Glycerol.
Puffer D	20 mM HEPES/KOH <i>pH</i> 7,9, 50 mM KCl, 0,2 mM EDTA <i>pH</i> 8,0, 8 % (<i>w/w</i>) oder 20 % (<i>w/w</i>) Glycerol, 2 mM Benzamidin, 0,5 mM DTT, 0,5 mM PMSF.
G200 Puffer	20 mM HEPES/KOH <i>pH</i> 7,9, 200 mM KCl, 0,2 mM EDTA <i>pH</i> 8,0, 1,5 mM MgCl ₂ , 0,5 mM PMSF, 0,5 mM DTT.
IPP150 Puffer	10 mM Tris/HCl <i>pH</i> 8,8, 150 mM NaCl, 0,1 % NP-40, 5 % (<i>w/w</i>) Glycerol.
TEV Protease Puffer	10 mM Tris/HCl <i>pH</i> 8,0, 75 mM KCl, 0,1 % NP-40, 0,5 mM EDTA, 5 % (<i>w/w</i>) Glycerol, 1 mM DTT.
Calmodulinbindungs- puffer	10 mM Tris/HCl <i>pH</i> 8,0, 150 mM NaCl, 0,1 % NP-40, 1 mM Mg(OAc) ₂ , 1 mM Imidazol, 2 mM CaCl ₂ , 10 mM β-Mercaptoethanol.
Calmodulinelutions- puffer	10 mM Tris/HCl <i>pH</i> 8,0, 150 mM NaCl, 1 mM Mg(OAc) ₂ , 1 mM Imidazol, 2 mM EGTA, 10 mM β-Mercaptoethanol, 0,02 % NP-40.
Gradientenpuffer (‘GraFix’)	20 mM HEPES/KOH <i>pH</i> 7,9, 200 mM KCl, 0,2 mM EDTA, 1,5 mM MgCl ₂ , 0,5 mM PMSF, 0,5 mM DTT, 10 oder 30 : % (<i>w/w</i>) Glycerol, optional 0,1 % Glutaraldehyd in 30 % (<i>w/w</i>) Glycerollösung.

Tabelle 3.2: Zusammensetzung der verwendeten Pufferlösungen

Chemikalie	Bezugsquelle
Acetobutyrate Zellulose (Triafol)	Sigma-Aldrich, Seelze, D
Aceton	Merck, Darmstadt, D
Alkylbenzoldimethylammonium (0,5 %) (Osvan Lsg.)	Sigma-Aldrich, Seelze, D
Calmodulin Affinitätsmatrix	Statagene, USA
DTT (1,4-Dithiothreitol)	Roth, Karlsruhe, D
EDTA (Dinatrium Salz Dihydrat)	Roth, Karlsruhe, D
EGTA	Roth, Karlsruhe, D
Essigsäureethylester	Merck, Darmstadt, D
Ethanol	Merck, Darmstadt, D
Glutaraldehyd (25 % Lösung)	Electron Microscopy Sciences, Hattfield, PA, USA
Glycerol	Merck, Darmstadt, D
HEPES (N-2-Hydroxyethylpiperazin-N- 2-ethansulfonsäure)	Calbiochem, Darmstadt, D
Hydrogenfluorid	Fluka / Sigma-Aldrich, Seelze, D
IgG-Sepharose	Pharmacia (GE Healthcare), Freiburg, D
Imidazol	Merck, Darmstadt, D
Kaliumhydroxid	Merck, Darmstadt, D
Kaliumchlorid	Merck, Darmstadt, D
Kalziumchlorid	Merck, Darmstadt, D
β -Mercaptoethanol	Roth, Karlsruhe, D
Magnesiumacetat Tetrahydrat	Merck, Darmstadt, D
Magnesiumchlorid	Merck, Darmstadt, D
Natriumchlorid	Merck, Darmstadt, D
Nonidet P-40 (NP-40)	Fluka (Sigma-Aldrich), Seelze, D
Protein Molekulargewichtsmarker	Bio-Rad, USA
Stickstoff (flüssig)	Messer, Sulzbach, D
TMB, 1 Step Ultra TMB-Elisa	Thermo Fisher, Rockford, IL, USA
Tris-(hydroxymethyl)-aminomethan (Tris)	Roth, Karlsruhe, D
Uranylformiat	Polyscience Inc., Warrington, PA, USA
Enzyme, Enzyminhibitoren, Anti- körper	Bezugsquelle
Anti-Kaninchen Ziege polyklonal affinitätsgereinigt, HRP-gekoppelt	Dianova, Hamburg, D
Anti-TAP Antikörper 'anti-CBP'	Biocat, D
Benzamidin	Serva, Heidelberg, D
PMSF (Phenylmethylsulfonylfluorid)	Roche, Mannheim, D
TEV Protease, rekombinant (10 U/ μ l)	Invitrogen, NL

Tabelle 3.3: Verwendete Chemikalien, Enzyme und Antikörper

3.1.3 Labormaterialien und Geräte

Labormaterialien	Bezugsquelle
Dialysemembran 'Spectra/Por'1 (6000-8000 Da MWCO)	Spectrum Laboratories Inc., CA, USA
ELISA-Platten, vorbeschichtet mit Antikörper:	
Ziege-anti Hase	Thermo Fisher, Rockford, IL, USA
Filterpapier 90 mm, Nr. 4 und 5	Whatman, Dassel, D
Glasobjektträger (76 x 26 mm)	Gerhard Menzel GmbH + Co.KG., Braunschweig, D
Glimmer (75 x 25 mm)	Plano, Wetzlar, D
Kohlenstoffstab, höchster Reinheitsgrad	Ringsdorff Werke GmbH, Bonn, D
Kupfernetze (200 Maschen)	Plano, Wetzlar, D
Kupfernetze (200 Maschen) Quantifoil R3.5/1 Cu	Quantifoil Micro Tools GmbH, Jena, D
Parafilm	Roth, Karlsruhe, D
Pipetten	Eppendorf, Hamburg, D
Pipetten 20 ml Kunststoff	Greiner Bio-One GmbH, Essen, D
Poly-Prep Säulen 0,8 x 4 cm	Bio-Rad, USA
Sterilfilter (0,2 μ l oder 0,45 μ l)	Millipore, Billerica, MA, USA
Teflon Präparationsblock (25 und 200 μ l Volumen)	Feinmechanik, im Hause
Zentrifugationsröhrchen	Beckmann, CA, USA

Tabelle 3.4: Verwendete Materialien

Hefestamm	Genotyp
YF-CP1 Snu71p TAP C-Terminus	Mat a, <i>trp1-Δ1</i> , <i>ura3-52</i> , <i>lys2-801</i> , <i>ade2-101</i> , SNU71::TAP Tag- <i>TRP1</i> .
YF-CP1 Snu56p TAP C-Terminus	Mat a, <i>trp1-Δ1</i> , <i>ura3-52</i> , <i>lys2-801</i> , <i>ade2-101</i> , SNU56::TAP Tag- <i>URA3</i> .

Tabelle 3.5: Genotypen der verwendeten Hefestämme

Laborgeräte	Bezugsquelle
CCD-Kamera TemCam F415 (4K x 4K)	Tietz Video Systems, Gauting, D
Elektronenmikroskop CM200 FEG	Philips, Eindhoven, NL
Glow-discharging Apparat	Feinmechanik, im Hause
Gradienten Mischer	Gradient Master, BioComp Instruments, Inc. Fredericton, NB, CAN
Kryo-EM-Probenhalter	Gatan GmbH, München, D
pH-Meter	Mettler Toledo, CH
Raumtemperatur ¹ -EM-Probenhalter	Philips, Eindhoven, NL
Sorvall SLC-6000 Rotor	Kendro, USA
Sorvall T647.5 Rotor	Kendro, USA
Sorvall T865 Rotor	Kendro, USA
Sorvall SS34 Rotor	Kendro, USA
Sorvall TH660 Rotor	Thermo Electron, Langenselbold, D
Sorvall Ultrazentrifuge Discovery 90	Thermo Electron, Langenselbold, D
Tischzentrifuge	Heraeus, D
Vortex	Janke & Kunkel, D
Zellmühle RM100 (elektrisch)	Retsch, D

Tabelle 3.6: Verwendete Laborgeräte

3.2 Biochemische Methoden

3.2.1 Aufreinigung des U1 snRNP aus *S. cerevisiae*

Die Aufreinigung des U1 snRNP aus *S. cerevisiae* wurde in Zusammenarbeit mit Dr. Ramazan Karaduman (Zelluläre Biochemie, im Hause) ausgeführt. Für die Aufreinigung des U1 snRNP aus *S. cerevisiae* wurde eine Tandem-Affinitäts-Aufreinigung², kurz TAP-(tag-)Methode, verwendet ([195], Abb. 3.1 B). Diese Strategie basiert auf der genetischen Einführung eines Fusionsproteins, bestehend aus einem nativen Protein gekoppelt mit dem Affinitätstag, der eine zweistufige Aufreinigung ermöglicht. Der TAP-tag besteht aus zwei IgG Bindungsdomänen, einer Verbindungssequenz und einem Calmodulin-Bindungs-Peptid (CBP). Eine IgG Bindungsdomäne wird gebildet aus Protein A ursprünglich isoliert aus *Staphylococcus aureus*. Die Verbindungssequenz beinhaltet eine Schnittstelle für eine Protease aus dem Tabakatzmosaikvirus (*engl.* tobacco etch virus, TEV Protease). Wahlweise kann der Affinitätstag mit dem C- oder N-Terminus eines Proteins verbunden werden wie in [195] beschrieben worden ist. Für die Strukturbestimmung des U1 snRNP wurde Snu71p am C-Terminus mit dem TAP-tag versehen. Das Prozedere der gesamten Aufreinigung ist in der Abb. 3.1 A schematisch dargestellt.

²TAP: *engl.* Tandem Affinity Purification

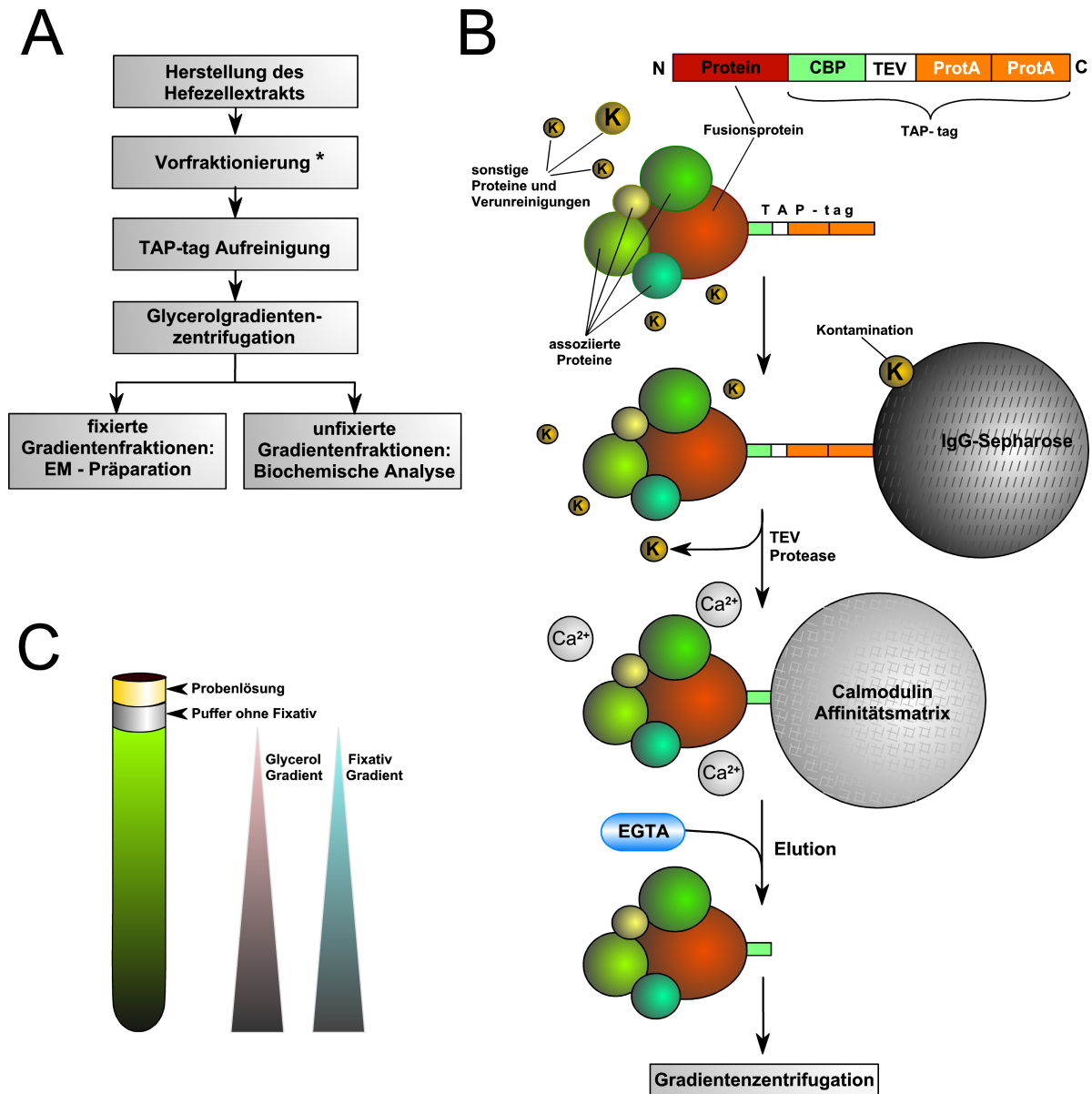


Abbildung 3.1: Aufreinigung des U1 snRNP aus *S. cerevisiae*.

A: Gesamtübersicht des Prozederes der U1 snRNP Aufreinigung. (*) Die Vorfractionierung wurde als zusätzlicher Aufreinigungsschritt zur Verbesserung der Probenqualität für die 3D Strukturbestimmung eingeführt.

B: Schematische Darstellung der Tandem-Affinitätsaufreinigung (TAP) nach [195]. Für die Strukturbestimmung von U1 snRNP war das TAP-tag Fusionsprotein Snu71. CBP: Calmodulin-Bindungs-Peptid, TEV: Schnittstelle der TEV Protease isoliert aus dem Tabakmosaikavirus, ProtA: IgG Bindungsdomäne, Protein A.

C: Bei der Glycerolgradientenzentrifugation sedimentieren die Proteinkomplexe in ansteigende Konzentrationen eines Fixativs [119].

Herstellung des Hefezelleextrakts

Die Präparation des Hefezelleextrakts wurde wie zuvor in [80] beschrieben durchgeführt. Die Hefezellen vom entsprechenden Hefestamm (YF-CP1, Snu71p C-Terminus TAP, s. Tab. 3.5) mit TAP-tag am C-Terminus von Snu71p wurden über Nacht bei 30 °C bis zu einem OD₆₀₀-Wert von 3,0/ml kultiviert. Die Zellen wurden anschließend in einem SLC6000 Rotor 10 min lang bei 5000 rpm und 4 °C pelletiert. Das Zellpellet je Zentrifugenröhrchen wurde mit 500 ml kaltem ddH₂O gewaschen und in AGK Puffer (20 mM HEPES/KOH pH 7,9, 200 mM KCl, 1,5 mM MgCl₂, 10 % (w/w) Glycerol) resuspendiert, und zwar mit 5 ml AGK Puffer je 2 l ursprünglicher Hefezellkultur. Nach der Resuspension wurde die KCl Konzentration auf 200 mM angepasst und 0,5 mM DTT sowie 0,5 mM PMSF hinzugefügt. Die Zellsuspension wurde in flüssigen Stickstoff getropft. Die gefrorenen Hefezelltropfen wurden mit einem elektrischen Mörser (RM100, Retsch, D) unter Zufuhr von Stickstoff 12 min zermahlen bis ein pulverisierter Zustand entstand, der anschließend auf Eis und bei Raumtemperatur³ aufgetaut wurde. Dieses Zellysate wurde nun bei 25000 x g (SS34 Rotor 17000 rpm) und 4 °C 30 min zentrifugiert. Der erhaltene Überstand wurde nochmals 1 h bei 4 °C und 100000 x g (T865 Rotor, 37000 rpm oder T647.5 Rotor, 42000 rpm) zentrifugiert. Nach diesem Zentrifugationsschritt lag das Zellysate in drei Phasen vor. Im Pellet lagen Ribosomen, genomische DNA und feine Zelltrümmer, die oberste Phase beinhaltete Lipide und Liposomen. Die mittlere Phase, das komplette Zellextrakt, wurde zweimal für 1,5 h gegen 5 l Puffer D (20 mM HEPES/KOH pH 7,9, 50 mM KCl, 0,2 mM EDTA pH 8,0, 8 % (w/w) Glycerol, 2 mM Benzamidin, 0,5 mM DTT; 0,5 mM PMSF) dialysiert. Erfolgte die U1 snRNP Aufreinigung ohne Vorfraktionierung enthielt der Puffer D anstelle der 8 % 20 % (w/w) Glycerol.

Vorfraktionierung

Bevor die Aufreinigung nach der TAP-tag Methode erfolgte wurde das dialysierte Zellextrakt mit einer weiteren Gradientenzentrifugation vorfraktioniert. Dazu wurde das Surespin-630 Zentrifugationssystem benutzt. Die Zentrifugationsröhrchen enthielten einen kontinuierlichen 10 – 30 % (w/w) Glycerolgradienten in G200 Puffer (20 mM HEPES/KOH pH 7,9, 200 mM KCl, 0,2 mM EDTA pH 8,0, 1,5 mM MgCl₂, 0,5 mM PMSF, 0,5 mM DTT). Pro Surespin-630 Zentrifugationsröhrchen wurden 7 ml des Hefezelleextrakts geladen. Es wurde 19 h lang bei 28800 rpm und 4 °C zentrifugiert, und anschließend wurden 24 Fraktionen von 1,5 ml Volumen abgenommen. Die Fraktionen 14 – 21 wurden verwendet, um mit der TAP-tag Methode die Aufreinigung fortzusetzen (Abb. 3.1, B). Dazu wurden zunächst die Pufferbedingungen des Extrakts auf 10 mM

³RT: Raumtemperatur

Tris/HCl und 0,1 % NP-40 angepasst.

TAP-tag Aufreinigung

Das vorfraktionierte Zellextrakt wurde 2 h bei 4 °C mit IgG beads (100 µl bead Suspension für 4 l Hefekultur), die zuvor mit 10 ml IPP150 Puffer (10 mM Tris/HCl pH 8,8, 150 mM NaCl, 0,1 % NP-40, 5 % (w/w) Glycerol) gewaschen worden waren, rotierend inkubiert. Nach dieser Inkubationszeit wurde das Extrakt so auf 10 ml BioRad Inkubationssäulen transferiert, dass auf jeder Säule 100 µl beads lagen. Die Säulen wurden nun dreimal mit 10 ml IPP150 Puffer und anschließend einmal mit TEV Protease Puffer (10 mM Tris/HCl pH 8,0, 75 mM KCl, 0,1 % NP-40, 0,5 mM EDTA, 5 % (w/w) Glycerol, 1 mM DTT) gewaschen. Daraufhin wurde der TAP-tag geschnitten, indem die IgG beads in 1 ml TEV Protease Puffer bei 16 °C mit TEV Protease 2 h lang rotierend inkubiert wurden (10 U TEV Protease pro 4 l Hefezellkultur). Aus diesem Prozess wurde die Lösung direkt in neue BioRad Poly-Prep Säulen eluiert, wobei jeweils die Eluate von zwei Säulen vereinigt wurden. Auf den neuen BioRad Inkubationssäulen waren zuvor 100 µl Calmodulin beads (100 µl pro 8 l Hefezellkultur) vorgelegt worden und mit 10 ml Calmodulinbindungspuffer (10 mM Tris/HCl pH 8,0, 150 mM NaCl, 0,1 % NP-40, 1 mM Mg(OAc)₂, 1 mM Imidazol, 2 mM CaCl₂, 10 mM β-Mercaptoethanol) gewaschen worden. Um Verluste zu minimieren, wurden die IgG-Sepharose Säulen nochmals mit jeweils 3 ml Calmodulinbindungspuffer eluiert. Dann wurde die Probenpufferlösung auf den BioRad Säulen mit den Calmodulin beads auf eine Konzentration von 2 mM CaCl₂ justiert und dies wurde 1 h bei 4 °C rotierend inkubiert. Danach wurden die Calmodulinmatrix Säulen entleert und dreimal mit 10 ml Calmodulinbindungspuffer gewaschen. Um den U1 snRNP von der Matrix zu lösen, wurden die Calmodulin beads zweimal mit insgesamt 700 µl Elutionspuffer (= Calmodulinbindungspuffer mit 0,02 % NP-40 anstelle 0,1 % und 2 mM EGTA anstelle 2 mM CaCl₂) eluiert, wobei der Elutionspuffer zunächst jeweils 20 min auf der Calmodulinmatrix inkubiert wurde.

Glycerolgradientenzentrifugation

Als letzten Schritt in der Aufreinigung des U1 snRNP für die EM wurde eine Glycerolgradientenzentrifugation im TH660 Rotor angeschlossen (Abb. 3.1, C). Dazu wurde das Eluat der TAP-tag Aufreinigung auf einen 10 – 30 % (w/w) linearen Glycerolgradienten, der gleichzeitig steigende Konzentration an Glutaraldehyd enthielt, geladen. Die Grundzusammensetzung des Gradientenpuffers lautete 5 mM HEPES/KOH pH 7,9, 200 mM KCl, 0,2 mM EDTA, 1,5 mM MgCl₂, 0,5 mM PMSF, 0,5 mM DTT sowie 10% (w/w) und entsprechend 30% (w/w) Glycerol. Desweiteren wurde in die 30% (w/w) Glycerollösung kurz vor der Vorbereitung der Gradienten Glutaraldehyd in einer finalen

Konzentration von 0,1 % (*w/v*) hinzugefügt. Die Zugabe dieses Fixativs entfiel für die biochemische Analyse des Gradientenprofils. Die Lösungen wurden in einem Gradientenmischer (Gradient Master, BioComp Instruments) mit vorprogrammierten Parametern (75 s, Winkel 83°, 20 rpm) gemischt. Es wurde im Sorvall TH660 Rotor bei 4 °C und 30000 rpm 12 h zentrifugiert. Der Gradient wurde anschließend in 24 Fraktionen von je 175 µl von unten abtropfend gesammelt.

Die biochemischen Proteingelanalysen wurden durchgeführt von Dr. R. Karaduman, Zelluläre Biochemie, im Hause. Es wurde jede der unfixierten Glycerolgradientenfraktionen auf ihre Proteinzusammensetzung getestet. Mittels SDS-Polyacrylamidgelelektrophorese (SDS-PAGE), adaptiert von Laemmli [133], unter Verwendung von 12 % Polyacrylamid Gelen, wurden die Proteine getrennt, und anschließend wurde das Gel mit Silber, basierend auf der Methode nach Blum et al. [19], gefärbt.

3.2.2 Markierung einzelner Proteine im *S. cerevisiae* U1 snRNP

Am U1 snRNP wurden zwecks Lokalisation einzelner Proteine im Protein-RNA-Komplex zwei Strategien verfolgt. Zum einen wurden Markierungen mit einem Antikörper getestet, zum anderen wurden Experimente zur Einführung eines genetischen Labels durchgeführt.

Antikörperbindung

Für die Antikörperbindungsversuche war der polyklonale Anti-CBP-Antikörper (Antikörper abgekürzt 'AK'), der gegen das Calmodulin-Bindungspeptid (CBP) gerichtet ist, ausgewählt worden. Um die Bindung des Anti-CBP-AK zu testen, wurden U1 snRNP Aufreinigungen wie beschrieben aber ohne die Vorfraktionierung und im kleinen Maßstab durchgeführt. Mit je 0,5 l Hefezellkultur konnte eine Bedingung, eine AK-Konzentration, erprobt werden. Das Eluat von der Calmodulinmatix nach dem letzten Schritt der TAP-tag Aufreinigung wurde mit dem Anti-CBP-AK 1 h bei 4 °C rotierend inkubiert und anschließend auf einen 10 – 30 % (*w/w*) Glycerolgradienten geladen. Die erhaltenen Gradientenfraktionen wurden mit der ELISA-Methode auf ihren Anti-CBP-AK-Gehalt untersucht. Dazu wurden mit AK vorbeschichtete (AK:Ziege anti-Hase) ELISA-Platten (Thermo Fisher Scientific, USA) wie in [93], [92] beschrieben und im wesentlichen nach den Instruktionen des Herstellers benutzt. Das Testsystem basierte auf dem HRP⁴-konjugierten anti-Hase Sekundär-AK. Durch die konjugierte Peroxidase wurde die TMB⁵

⁴*Horse-radish-peroxidase*, Meerrettich-Peroxidase

⁵3,3',5,5'-Tetramethylbenzidine

Lösung oxidiert, was durch einen Farbwechsel der Lösung sichtbar wurde. Die Absorptionmessungen wurden bei 450 nm durchgeführt. Um die Menge an Anti-CBP-AK für optimale Bindungseffizienz zu bestimmen, wurde der Anti-CBP-AK in unterschiedlich hohem molarem Überschuss zugegeben. Dabei wurde für die schwer zu bestimmende Menge an U1 snRNP ein Wert von $0,4\text{ pmol}$ aus $0,5\text{ l}$ Hefezellkultur angenommen, was aus Experimenten mit Ethidiumbromid-Färbung der U1 RNA und einem Standard abgeschätzt worden war. Da nur das optimale relative Verhältnis von U1 snRNP und Anti-CBP-AK gesucht war, war dieser Wert ausreichend für die Beurteilung der Gradientenprofile aus den ELISA-Daten, die von Dr. R. Karaduman gesammelt wurden.

Genetischer Label

Die Einführung eines genetischen Labels an Proteine des U1 snRNP wurde ausgeführt von Dr. R. Karaduman. In diesen Experimenten wurde der tDimer2, ein 54 kDa Protein geeignet für Fusionsproteindesign in *S. cerevisiae* [235], an den C-Terminus von verschiedenen U1 snRNP Proteinen fusioniert. Nachdem die korrekte chromosomale Integration des tDimer2 mittels PCR (Polymerase Chain Reaction) geklärt worden war, wurden zwei tDimer2 Fusionsproteine, Nam8 und Snp1, auf ihre korrekte Proteinexpression untersucht. Dazu wurde die U1 snRNP Aufreinigung aus den entsprechenden modifizierten Hefestämmen wie beschrieben (Abschn. 3.2.1) durchgeführt, jeweils einmal ohne und ein wiederholtes Mal mit Vorfraktionierung. Die erhaltenden Fraktionen der Glycerolgradienten wurden anschließend auf ihre Proteinzusammensetzung und das Vorhandensein des entsprechenden Fusionsproteins untersucht.

3.3 Elektronenmikroskopie

3.3.1 Auftragen von Kohlelochfolien auf Kupfernetzträger

Die für die Präparation negativ kontrastierter Proben benötigten Kupfernetzträger mit Kohlelochfilm wurden wie folgt hergestellt: Ein zuvor mit Ethanol (abs.) gereinigter Glasobjektträger wurde für 15 min senkrecht in eine wässrige $0,5\%$ Osvanlösung (Alkylbenzoldimethylammoniumchlorid) gestellt, wodurch die Oberfläche des Objektträgers hydrophob wurde. Anschließend wurde der Objektträger langsam aus der Osvanlösung genommen und für wenige Sekunden in bidestilliertes Wasser getaucht. Der Objektträger wurde daraufhin vorsichtig aus dem Wasser gezogen, ohne dass Wassertropfen auf der hydrophoben Oberfläche haften blieben und circa 10 min an der Luft trocknen gelassen. In einer humiden Atmosphäre wurde der vorbehandelte Objektglaträger auf einen -20°C gekühlten Alublock gelegt, so dass feine Wassertropfen auf der Objektträger-

oberseite kondensierten. Sofort wurde der Objektträger vom Alublock genommen und 500 μl einer 0,5 % Triafolllösung in Essigsäureethylester (Ethylacetat) wurden gleichmäßig über die Oberfläche des schräg gehaltenen Objektträgers pipettiert. Dabei legte sich die hydrophobe Lösung zwischen die feinen Wassertropfen. Nach der Evaporation des Ethylacetats und des Wassers verblieb ein dünner Triafolfilm auf dem Objektträger. Dessen Lochgröße wurde unter dem Lichtmikroskop kontrolliert. Triafolfolien, die eine adäquate Lochgröße (0,5 – 1,0 μl) aufwiesen, wurden weiterverarbeitet und auf Kupfernetzträger (Agar, Essex, UK) aufgebracht.

Um den Triafolfilm vom Glasobjektträger leicht zu lösen, wurde der Triafolfilm zunächst an den Rändern des Objektträgers angeschnitten, und anschließend wurde der beschichtete Objektträger 30 – 60 s in 0,1 % Flußsäure (HF) getaucht. Vorsichtig wurde der beschichtete Objektträger aus dem Flußsäurebad gezogen und langsam in einem flachen Winkel von ungefähr 35 ° in ein Wasserbad getaucht, wobei der Triafolfilm abflotierte. In Ethanol im Ultraschallbad (3 min) vorgereinigte Kupfernetzträger wurden mit ihrer stumpfen Seite nach unten auf den frei auf der Wasseroberfläche schwimmenden Triafolfilm gelegt. Um den vollständig mit Kupfernetzträgern belegten Triafolfilm aus dem Wasserbad zu entnehmen, wurde ein Objektglaträger auf einer Seite mit einer Papieretikette beklebt. Der Triafolfilm mit den Kupfernetzträgern haftete an der Papier beklebten Seite an und wurde über eine 180 ° Drehung im Wasser mit der Triafolseite und den Kupfernetzträgern nach oben zeigend aus dem Wasser genommen. Überschüssiges Wasser wurde mit Filterpapier abgesogen, und es wurde eine Weile trocknen gelassen. Anschließend wurde die Triafolseite der Kupfernetzträger unter Vakuum im Edwards E12E (BOC Edwards, Kirchheim, D) direkt mit Kohle bedampft, wobei die Kupfernetzträger weiterhin auf dem Objektträger lagen. Um den Triafolfilm zwischen Kohlefilm und Kupfernetzträger zu entfernen, wurden die Kupfernetzträger einzeln in eine Petrischale aus Glas gelegt, in der bereits mehrere stark in Ethylacetat getränkte Filterpapiere lagen. Die Petrischale wurde verschlossen und in dieser Ethylacetat angereicherten Atmosphäre wurden die Kupfernetzträger bis zu einer Woche lang belassen. Die Kupfernetzträger wurden nach Evaporation des Ethylacetat aus der Petrischale genommen, auf Glasobjektträger gelegt, und auf ihre Qualität überprüft. Kupfernetzträger, deren Kohlelochfilm zu mehr als 1/5 zerstört waren oder die sonstige Fehlstellen aufwiesen, wurden verworfen oder gegebenenfalls noch für Testzwecke verwendet.

3.3.2 Herstellung kontinuierlicher Kohlefilme

Um einen kontinuierlichen Kohlefilm herzustellen, wurde ein Glimmerplättchen (Plano G250-1, 25 × 75 mm) längs gespalten und mit der frisch gespaltenen Oberfläche nach oben

in einen Metallkorb (Eigenbau, Feinmechanik, im Hause) gelegt. Der Metallkorb war genau auf die Maße eben dieser Sorte Glimmerplättchen zugeschnitten. Sodann wurde der Metallkorb ohne Zeitverzögerung in eine Vakuumanlage (Edwards E12E; BOC Edwards, Kirchheim, D) gestellt, in der durch Anlegen von Hochspannung an mit Kohlestäben (Ringsdorff Werke GmbH, Bonn, D) bestückten Elektroden, Kohle zerstäubt wurde. Der Metallkorb verhinderte weitgehend die direkte Bedampfung des Glimmerplättchens mit Kohlestaub, wodurch Kohleaggregate vermieden wurden. Durch Reflexion des feinkörnigen Kohlestaubs an den mit Glasobjektträgern bestückten Innenseiten des Metallkorbs wurde die in diesem Fall bevorzugte indirekte Kohlebedampfung erreicht. Diese Art der Bedampfung gewährleistete einen gleichmäßigen feinen Kohlefilm, der sich durch hohe Bruchfestigkeit im Gebrauch auszeichnete. Die Schichtdicke des Kohlefilms wurde visuell nach der Kohlefärbung eines beigelegten Filterpapiers beurteilt. Für die Herstellung negativ kontrastierter Proben des U1 snRNP wurden im Vergleich zu anderen Verwendungen dünne Kohlefilme bevorzugt.

3.3.3 Präparation negativ kontrastierter Proben

Um von dem U1 snRNP elektronenmikroskopische Aufnahmen machen zu können, wurden die Partikel auf einen Kupfernetzträger präpariert. Dazu wurde die sogenannte Doppelfolien-Technik nach [264] verwendet, nach der die Proteinkomplexe in Uranylformiat [138] oder Uranylacetat eingebettet werden, wodurch der Kontrast ihrer Umgebung erhöht wird. Die fixierte Probenlösung (Fraktionen 15 – 19 nach der Glycerolgradientenzentrifugation) wurde in eine Teflonschale mit 25 – 30 μl Volumen pipettiert. Auf diese Lösung wurde ein dünner Kohlefilm, der zuvor auf Glimmer als Trägermaterial aufgedampft worden war, abflotiert, ohne aber vollständig vom Träger gelöst zu werden. In diesem Zustand adsorbierten die Partikel an den Kohlefilm (Abb. 3.2, A1). Aufgrund der geringen Probenkonzentration betrug die Inkubationzeit für den U1 snRNP 2,5 – 3 *h*. Anschließend wurde der Kohlefilm samt Träger abgezogen und auf eine 2% (*w/v*) Uranylformiatlösung (200 μl Volumen) transferiert (Abb. 3.2 A2 & B1), wobei der Kohlefilm vollständig vom Träger gelöst wurde und mit der Partikel tragenden Seite auf der Lösung lag (Abb. 3.2, B1). Die Inkubationszeit in dieser Lösung betrug 2 – 3 *min*. Mit einem Kupfernetzträger, dem bereits ein dünner Kohlelochfilm anhaftete, wurde der Kohlefilm mit der adsorbierten Probe so von der Uranylformiatlösung abgenommen (Abb. 3.2, B2 & B3), dass die beiden Kohlefilme direkt einander auflagen. Ein weiterer Kohlefilm war ebenfalls 2 – 3 *min* auf einer Uranylformiatlösung gleicher Konzentration inkubiert worden (Abb. 3.2, C1). Dieser wurde auf die Partikel gelegt, indem er mit dem Probe tragenden Kupfernetzträger aus der Uranylformiatlösung ge-

hoben wurde. Bei dieser Art der Präparation wurde überschüssige Flüssigkeit stets vor Überführung des Trägers in eine andere Lösung vorsichtig mit Filterpapier abgesogen. Ein Teil der so präparierten Proben wurde in flüssigem Stickstoff schockgefroren (kryonegativ-Kontrastierung, [76]). In einer Modifikation der Doppelkohlefolientechnik wurde der zweite Kohlefilm, während dieser auf der Uranylformiatlösung inkubierte, so zerstört, dass auf dem Kupfernetzträger genügend Bereiche ohne Deckkohlefolie erhalten wurden. Über diese Methode wurde eine gute Kontrastmittelverteilung in Einzelkohlefolienbereichen erzielt.

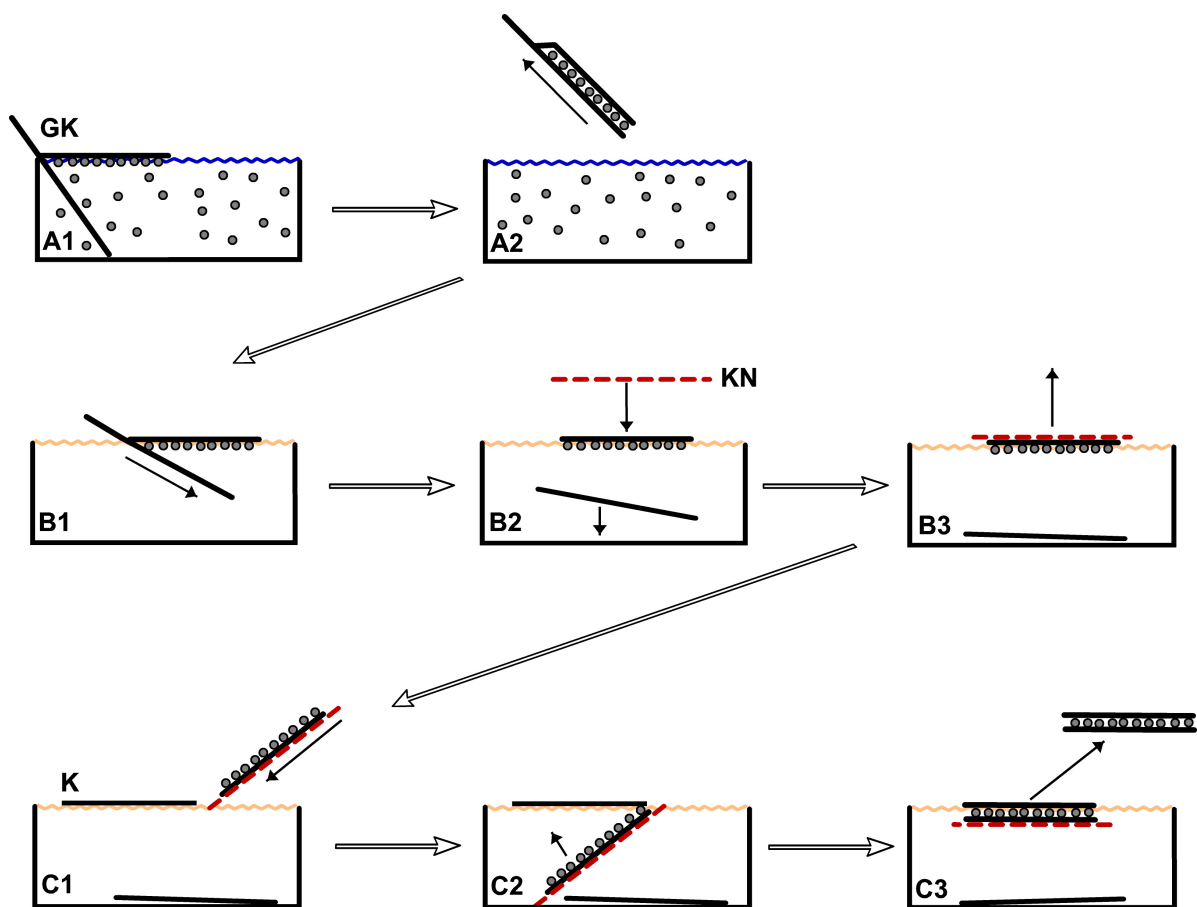


Abbildung 3.2: Illustration der Präparation negativ kontrastierter Proben. Schematisch dargestellt ist die Präparation von negativ kontrastierten Proben mit der Doppelkohlefolientechnik nach [264]. Die Details sind im Text beschrieben unter Bezugnahme der Bezeichnungen A1, A2, B1, B2, B3, C1, C2 und C3. **GK:** Glimmer mit aufgedampftem Kohlefilm, **KN:** Kupfernetzträger mit anhaftendem perforierten Kohlefilm, **K:** Kohlefilm, **blau:** Probenlösung, **gelb:** Uranylformiatlösung.

3.3.4 Aufnahmen am EM

Das U1 snRNP aus *S. cerevisiae* wurde mit einem Philips CM200 FEG Elektronenmikroskop (Philips, Eindhoven, NL) bei einer angelegten Hochspannung von 160 kV und einer 88000-fachen Vergrößerung mikroskopiert. Die Aufnahmen entstanden bei diesen Einstellungen in Spot Scan Serien [52] mit einer $4K \times 4K$ CCD Kamera (TVIPS EM-MENU, Tietz Video Systems, Gauting, D) bei zweifach-Binning. Für die RCT Rekonstruktion wurden so Daten bei RT und 0° und 45° gekipptem Probenhalter aufgenommen. Unter Stickstoffkühlung der Probe entstanden weitere Datensätze unter einem Winkel von 0° . Die Datensätze, die zur 3D Rekonstruktion des U1 snRNP aus *S. cerevisiae* verwendet wurden, sind in der Tabelle 3.7 aufgeführt.

Datensatz	Mikroskopie-Methode	Anzahl	Pixelrahmen	pro Pixel
$0^\circ / 45^\circ$ Partikelpaare	RT, neg. kontrastiert	10168	80×80	4, 4
0° Aufnahmen	RT, neg. kontrastiert	24713	80×80	4, 4
0° Aufnahmen	Stickstoffkühlung, neg. kontrastiert	107147	80×80	4, 4

Tabelle 3.7: Zur 3D Rekonstruktion von U1 snRNP verwendete Datensätze

3.3.5 Bildverarbeitung

Die EM-Bilder, aufgenommen in Spot Scan Serien mit der CCD Kamera, wurden mit Hilfe der ImageMagick Software zu großen Bildern zusammengefügt. Nach einer Reduktion von 16 auf 8 bit in Photoshop 7.1 wurden die einzelnen U1 snRNP unter Benutzung des Programms CowHide (Dr. M. Lüttich, im Hause) in der Regel manuell oder mit Signature [37] automatisch selektiert und anschließend mit dem jeweiligen Programm in einem vorher angegebenen Pixelrahmen in Einzelbilder extrahiert. Die Einzelbilder wurden bei der Extraktion in einem einzigen großen Datensatz zusammengefasst, der in nachfolgenden Schritten mit der Imagic-5 Software [275] und Unterprogrammen des CowHide Softwarepakets wie in Kap. 2 beschrieben prozessiert wurde. Die Bandpassfilterung des Datensatzes sowie das initiale referenzfreie [55] translationale Alignment wurden stets lokal auf dem PC mit der Imagic-5 Software gerechnet. Für MR-Alignments stand eine parallelisierte Grid-Computing Hardware/Software Lösung zur Verfügung [225]. Die Datenpakete wurden dabei automatisch auf alle freien Prozessoreinheiten verteilt, und dies erlaubte eine schnelle Berechnung der Alignment-Parameter. Die Klassifizierung der Daten wurde mit der CowHide Software [156] lokal gerechnet, ebenso wie die *de novo* RCT-Rekonstruktion. Das anschließende ML-Alignment [214] wurde mit Hilfe eines Skripts auf einem Computercluster ausgeführt. Die Refinement-Methoden auf 3D Ebene

(*angular reconstitution, projection matching*) wurden auf dem lokalen PC mit Imagic-5 gerechnet.

Kapitel 4

Ergebnisse

4.1 Aufreinigung des U1 snRNP aus *S. cerevisiae*

Die Aufreinigung des U1 snRNP aus *S. cerevisiae* erfolgte in Zusammenarbeit mit Dr. R. Karaduman (Zelluläre Biochemie, im Hause) und stets wie in Kap. 3, Abschn. 3.2.1 detailliert beschrieben aus dem Hefezelleextrakt über die TAP-Methode [195] mit nachfolgender Gradientenzentrifugation in einem 10 – 30% (*w/w*) Glycerolgradienten, dem im Allgemeinen für die elektronenmikroskopische Probenpräparation Glutaraldehyd als Fixativ zugesetzt wurde (GraFix-Methode) [119].

4.1.1 Aufreinigung ohne Vorfraktionierung

Zunächst war die Aufreinigung des U1 snRNP ohne die eingeführte Vorfraktionierung des Hefezelleextrakts (Kap. 3, Abb. 3.1, A) durchgeführt worden, so dass die TAP-tag Aufreinigung direkt aus dem Hefezelleextrakt erfolgte. Der Calmodulinelutionspuffer enthielt in dieser Aufreinigung 20 *mM* EGTA. Abschließend wurde das isolierte U1 snRNP auf einem 10 – 30% (*w/w*) Glycerolgradienten ohne Fixativzusatz zentrifugiert. Das U1 snRNP sedimentiert mit einem Sedimentationskoeffizienten von 18 *S* [59] und wird während der Zentrifugation im Glycerolgradienten von seinen Zerfallsprodukten und von Proteinaggregaten oder von noch weiteren vorhandenen Verunreinigungen getrennt. Die Proteinanalyse des Glycerolgradienten mittels SDS-PAGE ist in Abb. 4.1 gezeigt. In den Fraktionen 13 – 17 sind die Banden der Proteine des U1 snRNP erkennbar, mit einem Peak in den Fraktionen 14 – 16. Um die Proteinbanden benennen zu können, waren exemplarisch für einen Snu71p-TAP-tag und Gradienten aufgereinigten U1 snRNP die Proteine nach der Trennung auf dem Polyacrylamid Gel massenspektroskopisch nachgewiesen worden (s. Appendix).

Für die elektronenmikroskopische Untersuchung des U1 snRNP wurden die Gradientenfraktionen im Bereich des Peaks, Fraktionen 14 – 20, einzeln präpariert und begut-

achtet. Dazu wurde zunächst das U1 snRNP wie unter Kap. 3 Abschn. 3.3.3 detailliert beschrieben auf einen dünnen Kohlefilm absorbiert. Anschließend wurde die Probe mit Uranylformiat negativ kontrastiert und auf einen Kupfernetzträger aufgebracht. Auf

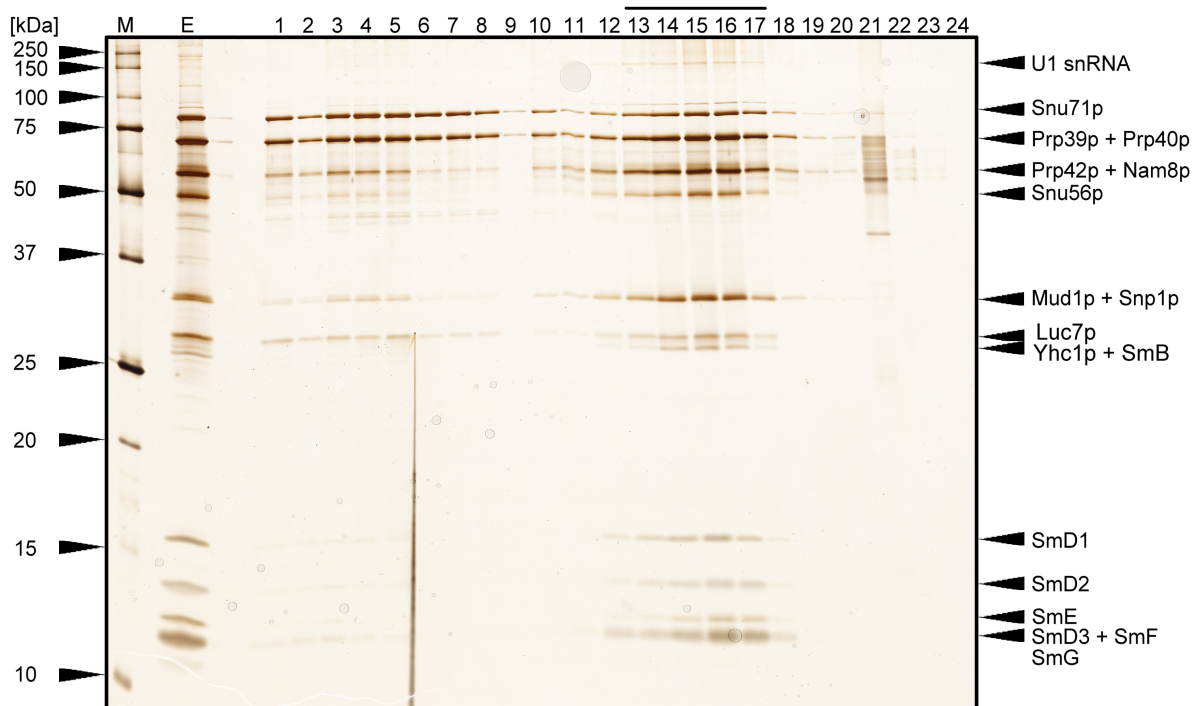


Abbildung 4.1: Proteinanalyse des Gradienten aufgereinigten U1 snRNP. TAP-tag aufgereinigter U1 snRNP wurde auf einem 10 – 30 % (*w/w*) Glycerolgradienten ultrazentrifugiert. Aus den Gradientenfractionen (Gesamtzahl 24, oben 1, unten 24) wurden die Proteine mit Aceton präzipitiert und danach mittels SDS-PAGE auf einem 12% Polyacrylamid Gel analysiert. Das Gel wurde mit Silber gefärbt. In den Fractionen 13 – 17 sind alle Banden der Proteine des U1 snRNP erkennbar mit einem Peak in den Fractionen 14 – 16. M: Proteinstandard zur Angabe der Molekulargewichte; die Massen in [kDa] sind links daneben angegeben, E: Eluat, Probe, die auf den Glycerolgradienten aufgetragen wurde.

diese Weise wurden sowohl U1 snRNP Proben präpariert, denen im Glycerolgradienten kein Fixativ zugesetzt worden war (Abb. 4.2), als auch U1 snRNP Proben, die mit steigenden Konzentrationen an Glutaraldehyd im Gradienten langsam fixiert wurden (Abb. 4.3). Die Randfractionen der Peaks der untersuchten Gradienten wiesen bei gleicher Absorptionszeit eine geringere Partikeldichte auf dem Kupfernetzträger auf, so daß die Hauptfractionen des U1 snRNP nachträglich durch visuelle Betrachtung im Elektronenmikroskop bestimmt werden konnten. Damit konnten auch leichte Verschiebungen des Hauptpeaks, die durch die manuelle Vorbereitung und Ernte der Glycerolgradienten entstehen, erkannt werden. Überdies wurde der Peak bei Glutaraldehydfixierung häufig um eine Fraction in Richtung höherer Sedimentationskoeffizienten verschoben (eigene Beobachtung und persönliche Kommunikation mit Dr. M. Golas), was eine Einzelbe-

trachtung der in Frage kommenden Fraktionen unausweichlich machte.

Um nun sowohl die strukturellen Eigenschaften des U1 snRNP, als auch die Auswirkungen der Fixativzugabe im Glycerolgradienten auf den Komplex besser beurteilen zu können, wurden von beiden U1 snRNP Proben Datensätze aus Einzelpartikelbildern erstellt, wie unter Kap. 2, Abschn. 2.1 beschrieben. Mit diesen Datensätzen wurden in Datenprozessierungszyklen aus MR-*Alignment* und Klassifizierung (Kap. 2, Abschn. 2.2) jeweils Klassensummen erhalten, die die Partikelansichten gegenüber Einzelbildern mit einem besseren SNR zeigen (Abb. 4.2, A & Abb. 4.3, A). Bei der Prozessierung der Datensätze war auffällig, dass unter den Klassensummen der fixierten U1 snRNP Probe kaum Ansichten gefunden wurden, die denen des unfixierten U1 snRNP offensichtlich ähnlich waren (siehe dazu auch Abschn. 4.1.2). Die Klassensummen des unfixierten U1 snRNP zeigten einen elongierten Strukturkörper mit einer breiten Hälfte und einem schmal zulaufenden Teil (Abb. 4.2). Über den gesamten Datensatz ähnelten sich die Ansichten sehr. Dagegen wiesen die Ansichten der ebenfalls elongierten Struktur des fixierten U1 snRNP stärkere Variation auf (Abb. 4.3, A und nicht gezeigt). Dies wurde in mehreren Testdatensätzen so beobachtet. Diese Verschiedenartigkeit der Klassensummen zwischen unfixierter und fixierter Probe war mit anderen Komplexen wie dem Ribosom oder dem B Δ U Komplex nicht beobachtet worden [119]. Von einigen Klassensummen, sowohl des unfixierten, als auch des fixierten U1 snRNP, wurden die Konturen gezeichnet (Abb. 4.2 A, Abb. 4.3 B). Die Umrisse der Projektionsansichten sind zunächst nur eine simplifizierte Darstellung der äußeren Gestalt des betrachteten Partikels. Mit ihrer Hilfe lässt sich allerdings das Ausmaß der strukturellen Heterogenität der unfixierten U1 snRNP Probe verdeutlichen. Die Klassensummen 1 – 6 und 8 in Abb. 4.2, A zeigen den U1 snRNP aus relativ ähnlichen Projektionsrichtungen, da sich ein Muster der Grauwertschwankungen auf dem Hauptkörper der Struktur wiedererkennen lässt. Dennoch weisen die Konturen Unterschiede auf, die sich grundsätzlich durch leichte Konformitätsänderungen der Struktur oder leichte Abweichungen der Projektionsrichtung erklären lassen. Bei Überlagerung der Konturen jedoch, treten 2 Abweichungen systematisch auf (Abb. 4.2, B). Danach gibt es Strukturen des U1 snRNP, denen in der größeren Strukturhälfte im 'Kopfbereich' eine Dichte fehlt (Struktur mit schwarzer Kontur in Abb. 4.2, A). Außerdem weisen die Strukturen in der schmalen Hälfte der Struktur im 'Fußbereich' verstärkt Variabilität auf, die sowohl auf konformationelle Unterschiede, als auch auf eine zusätzlich auftretende Dichte in einigen der Strukturen zurückzuführen ist.

Unter den U1 snRNP Klassensummen der fixierten Probe fällt strukturell vor allem eine Domäne auf, die vom Hauptkörper der Struktur absteht. Diese wurde in den Konturen (Abb. 4.3, B) mit einer gestrichelten Linie angedeutet. Auch wenn diese Dichte nicht als

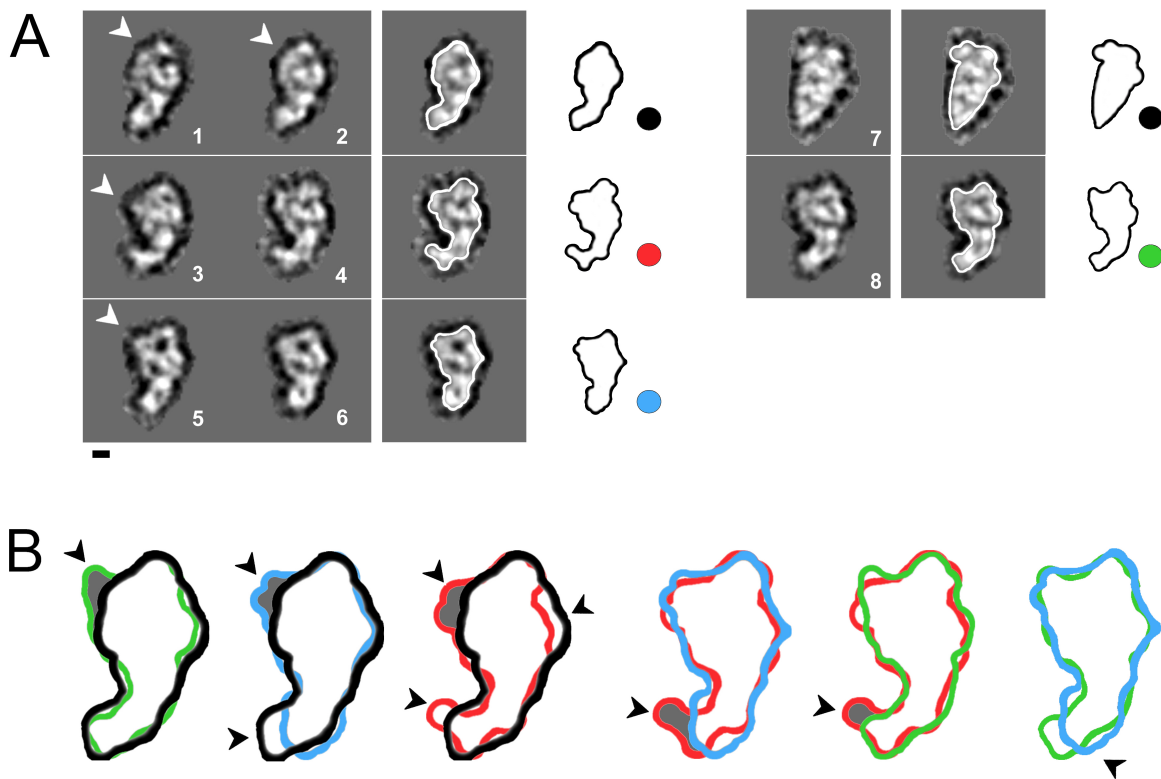


Abbildung 4.2: Charakteristische Klassensummen des U1 snRNP erhalten aus Glycerolgradientenfraktionen ohne Fixativzugabe.

A: Die Klassensummen 1 – 8 sind typische Ansichten des mit Uranylformiat negativ kontrastierten U1 snRNP. Die EM Probenpräparation wurde von TAP-tag aufgereinigtem U1 snRNP Komplex vorgenommen, wobei die Aufreinigung direkt aus dem Hefezellextrakt ohne vorherige Vorfraktionierung erfolgte und mit einer Glycerolgradientenzentrifugation ohne Fixativzugabe abgeschlossen wurde. Die Klassensummen wurden berechnet aus 4325 Einzelbildern, extrahiert aus einem bei RT und 115000-facher Vergrößerung aufgenommenen EM-Bilddatensatz. Der Maßstab entspricht 5 nm . Von den Klassensummen 2, 4, 6, 7 und 8 wurden Masken gebildet, die die Kontur der jeweiligen 2D Ansicht wiedergeben. Zur Unterscheidung sind diese mit einem Farbcode versehen, der in Abb. (B) verwendet wird.

B: Abgesehen von Klassensumme 7 zeigen die verschiedenen Klassensummen den U1 snRNP aus vergleichbaren Projektionsrichtungen. Diese Konturen wurden daher so gut wie möglich übereinander gelegt. Die Pfeile zeigen auf jene Bereiche, die sich nicht zur Deckung bringen lassen. Jene Bereiche, die sich in allen Kombinationen der Konturen nicht überlagern, sind allein mit Konformationsunterschieden oder leicht unterschiedlichen Projektionsrichtungen nicht erklärbar und deuten vielmehr auf eine zusätzliche Dichte hin. Diese Bereiche sind grau unterlegt. Sowohl im Kopf- als auch im Fußbereich besteht demnach Heterogenität in der Konfiguration des Komplexes.

alleiniger Unterschied zwischen der fixierten und der unfixierten Probe geltend gemacht werden kann, so scheint sie bei näherer Betrachtung der Klassensumme 10 (Abb. 4.3, C)

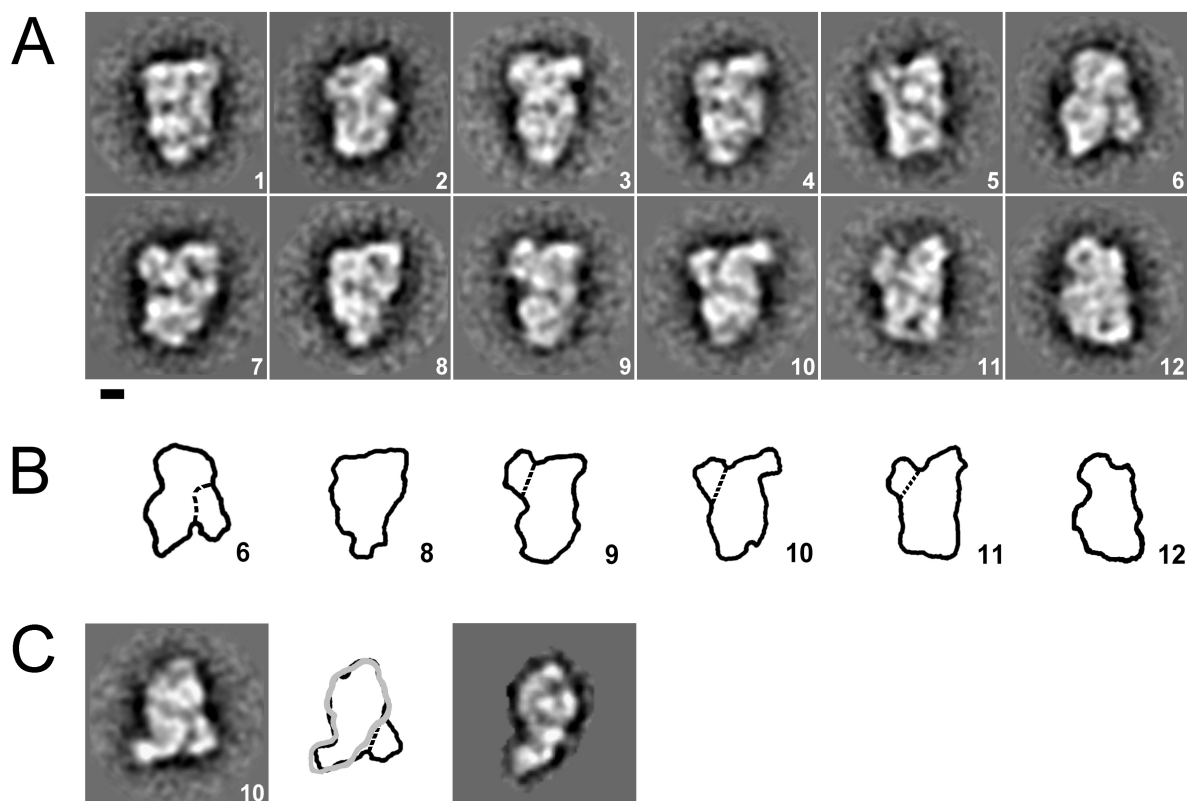


Abbildung 4.3: Charakteristische Klassensummen des U1 snRNP erhalten aus Glycerolgradientenfractionen mit Fixativzugabe.

A: Die Klassensummen 1 – 12 sind einige typische Ansichten des mit Uranylformiat negativ kontrastierten U1 snRNP. Die EM Probenpräparation wurde von TAP-tag aufgereinigtem U1 snRNP Komplex vorgenommen, wobei die Aufreinigung direkt aus dem Hefezellextrakt ohne vorherige Vorfraktionierung erfolgte. Abgeschlossen wurde die TAP-tag Aufreinigung mit einer Glycerolgradientenzentrifugation. Der Glycerolgradient enthielt zur Fixierung des Komplexes Glutaraldehyd in steigenden Konzentrationen (GraFix). Die Klassensummen wurden berechnet aus 10570 Einzelbildern, selektiert aus einem bei RT und 88000-facher Vergrößerung aufgenommenen EM-Bilddatensatz. Der Maßstab entspricht 5 nm. Bei 3 und 9 sowie 4 und 10 handelt es sich um ähnliche Projektionsansichten des U1 snRNP.

B: Von den Klassensummen 6 und 8 – 12 wurden, wie in Abb. 4.2, A veranschaulicht, Masken gebildet, die die Kontur der jeweiligen 2D Ansicht wiedergeben. Mit einer gestrichelten Linie abgegrenzt ist eine Domäne, die vom Hauptkörper der Struktur absteht.

C: Vergleich der Klassensumme 10 mit einer Projektionsansicht des unfixierten U1 snRNP. Die Strukturen der Hauptkörper weisen in beiden Ansichten Ähnlichkeiten auf, auch wenn die Feinstruktur in der unfixierten Klassensumme in diesem Beispiel deutlicher erkennbar ist. Der fixierte U1 snRNP Komplex trägt demnach eine zusätzliche Domäne.

und Vergleich mit den Klassensummen 1 und 2 der unfixierten Probe in Abb. 4.2 ein wesentlicher Unterschied zu sein.

Die Tatsache, dass die U1 snRNP Klassensummen von fixierter und unfixierter Probe so

unterschiedlich sind, kann im Allgemeinen unterschiedliche Gründe haben. Zum einen kann der unfixierte Komplex bei dem Absorptionsprozess auf den Kohlefilm teilweise zerfallen, oder aber es besteht inhärent strukturelle Heterogenität des Komplexes, die bei der biochemischen Aufreinigung noch nicht berücksichtigt werden konnten. Erstere Möglichkeit wird durch die sanfte Fixiermethode im Glycerolgradienten vermieden, wie mit verschiedenen Proben [98], [119] und für den U1 snRNP in Abb. 4.3, C gezeigt werden konnte. Aus diesem Grund sind die fixierten Proben vorzuziehen, insbesondere dann, wenn sie wie in diesem Fall eine offensichtlich andere Struktur aufweisen, da sie sehr wahrscheinlich den intakten Komplex darstellen. Dass durch die Fixiermethode neue intermolekulare Verbindungen geknüpft werden, die keine biologische Relevanz haben und natürlicherweise nicht bestanden haben, ist aufgrund der milden und sukzessiven Fixierung im Gradienten unwahrscheinlich und auch mit anderen Proben nicht beobachtet worden. Darüber hinaus könnte die relativ große Gleichartigkeit der unfixierten Klassensummen untereinander auf präferenzielle Bindung des unfixierten Komplexes an den Kohlefilm hindeuten.

4.1.2 Aufreinigung mit Vorfraktionierung

Da sich ein großer Teil der Klassensummen des U1 snRNP auch von den fixierten Proben in manchen Bereichen ihrer Kontur relativ schlecht vom Kontrastmittelsaum abgrenzen, waren sie somit noch nicht gut definiert. Die Klassensummen wiesen insgesamt betrachtet relativ schlechte SNRs auf. Dies deutete auf noch in nicht zu vernachlässigem Maße vorhandene strukturelle Heterogenität des Komplexes hin. Die Qualität der Probe sollte daher für die 3D Strukturbestimmung durch Modifikation der biochemischen Aufreinigung des U1 snRNP verbessert werden. Da im Hefezellextrakt auch freies und in Subkomplexen vorliegendes TAP-tag Fusionsprotein vermutet wurde, wurde für die TAP-tag Aufreinigung nicht mehr der gesamte Hefezellextrakt verwendet, sondern stattdessen wurde in einem zusätzlichen Aufreinigungsschritt (vgl. Abb. 2.1, Vorfraktionierung) das Hefezellextrakt vorfraktioniert. Dazu wurde das Hefezellextrakt zunächst in einer Surespin-Gradientenzentrifugation, wie in Kap. 3, Abschn. 3.2.1 beschrieben, aufgetrennt. Die erhaltenen 24 Fraktionen wurden mit Standardprotokollen der Northern Blot Methode [213] unter Verwendung entsprechender, radioaktiv markierter, RNA Sonden auf das Vorhandensein der snRNAs U1, U4, U5L, U5S und U6 überprüft (Abb. 4.4, A). Die U1 snRNA konnte in den Fraktionen 7 – 24 nachgewiesen werden.

Um zu ermitteln, welchen ungefähren S-Wert die Komplexe aufweisen, die in eine bestimmte Fraktion des Surespin-Gradienten sedimentierten, wurde aus dem totalen Hefezellextrakt die RNA extrahiert und unter den Gradientenbedingungen der Vorfraktionie-

rung zentrifugiert und fraktioniert (Kap. 3, Abschn. 3.2.1). Von jeder der erhaltenen 24 Fraktionen wurde ein Aliquot abgenommen und auf einem 8 M Urea - 6 % Polyacrylamid Gel mittels PAGE aufgetrennt. Das Gel wurde anschließend mit Silber gefärbt (Abb. 4.4, B). Die ribosomalen RNA (rRNA), die im Extrakt reichlich vorhanden sind, lassen sich den erhaltenen Banden zuweisen. Da von den rRNA der S-Wert bekannt ist, wurde auf diese Art und Weise eine grobe S-Wert-Verteilung im Surespin-Gradienten ersichtlich. Für die weitere U1 snRNP Aufreinigung wurden die Fraktionen 14 – 21 ausgewählt, da 18 S rRNA hauptsächlich in den Fraktionen 16 – 18 nachgewiesen wurde und vom U1 snRNP erwartet wird, dass er ebenfalls mit diesem S-Wert sedimentiert [59]. Mit den Surespin-Fraktionen 14 – 21 wurde die Aufreinigung des U1 snRNP über die TAP-tag Methode dann wie gewöhnlich fortgesetzt. Abschließend wurde der TAP-tag aufgereinigte U1 snRNP auf einem 10 – 30 % (*w/w*) Glycerolgradienten ultrazentrifugiert. Die Proteinanalyse der unfixierten Gradientenfraktionen ist in Abb. 4.5 gezeigt. In den Fraktionen 14 – 18 sind alle Banden der Proteine des U1 snRNP erkennbar einschließlich der U1 snRNA.

Für die EM-Präparationen wurden daraufhin stets die Fraktionen 15 – 19 der Glycerolgradienten, denen das Fixativ Glutaraldehyd zugesetzt worden war, verwendet. Die U1 snRNP wurden dafür mit Uranylformiat negativ kontrastiert (Kap. 3, Abschn. 3.3.3) und anschließend unter RT Bedingungen mikroskopiert. Der Hauptpeak lag in den Fraktionen 16 – 18, wie visuell nach Begutachtung mehrerer Fraktionen im Elektronenmikroskop (Abschn. 4.1.1) beurteilt wurde. Nach Erstellung eines Datensatzes aus Einzelpartikelbildern, wie unter Kap. 2, Abschn. 2.1 beschrieben, wurden die Partikel anschließend in Zyklen aus MR-Alignments und Klassifizierung ausgerichtet und in Gruppen ähnlicher Ansichten sortiert (Kap. 2, Abschn. 2.2). Eine Auswahl von erhaltenen Klassensummen ist in Abb. 4.6, A - C gezeigt. In ihrer Gesamtheit ähnelten sich diese Klassensummen mit den Klassensummen der fixierten Probe des ohne Vorfraktionierung aber ansonsten in gleicher Weise TAP-tag aufgereinigten U1 snRNP. In den Abbildungen 4.3, A können beispielsweise die Klassensummen 4, 9 und 10 mit den Klassensummen 2, 3 und 7 in Abb. 4.6 verglichen werden.

Die äußeren Abmessungen der U1 snRNP Struktur sind in Abb. 4.6, B angegeben. Danach hat der Komplex eine längliche Gestalt mit 233 Å in der Länge und bis zu 163 Å in der Breite.

In Abb. 4.6, C sind von 4 Klassensummen die Umrisse gezeichnet, wobei in Klassensumme 2 eine vom Hauptkörper abstehende Domäne in der Kontur gestrichelt eingezeichnet ist, wie bereits in (Abb. 4.3, B). Zwei dieser Ansichten, Klassensumme 1 und 4, weisen außerdem eine gewisse Ähnlichkeit mit Projektionsansichten aus der unfixierten Probe des U1 snRNP aus der Aufreinigung ohne Vorfraktionierung auf (Abb. 4.6, D).

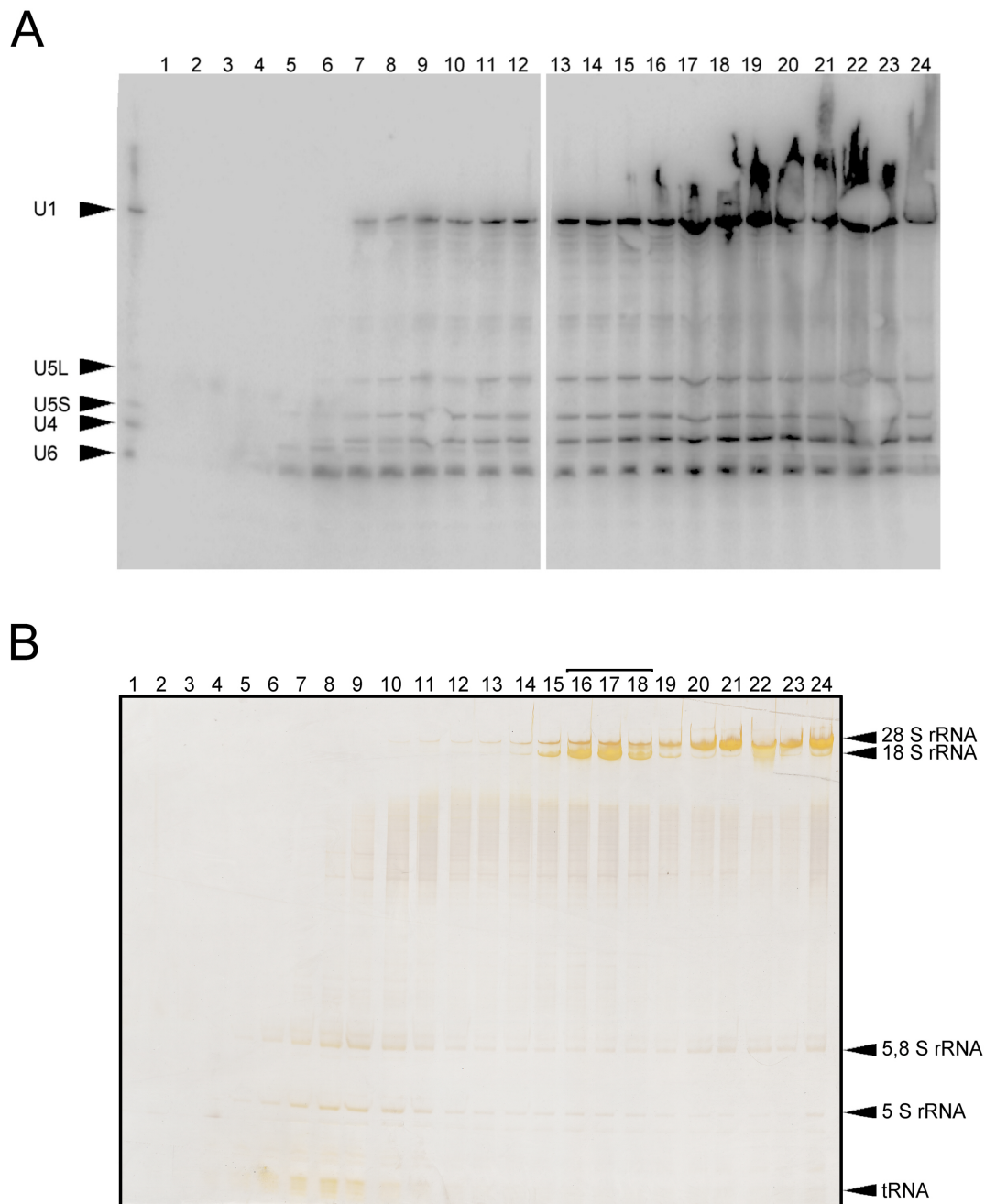


Abbildung 4.4: Analyse der Gradientenfractionen der Vorfraktionierung.

A: Totales Hefeextrakt wurde unter den Gradientenbedingungen der Vorfraktionierung, wie in Kap. 3 im Abschn. 3.2.1 beschrieben, zentrifugiert. Der Gradient wurde anschließend in 24 Fraktionen je 1,5 ml gesammelt. Ein Aliquot jeder Fraktion wurde in einer Northern Blot Analyse mit entsprechenden RNA-Sonden auf die Anwesenheit von U1, U4, U5L, U5S und U6 überprüft. U1 RNA wird in den Fraktionen 7 – 24 mit einem Peak in den Fraktionen 17 – 20 detektiert.

B: Phenolextrahiertes Hefeextrakt wurde unter den Gradientenbedingungen der Vorfraktionierung (vgl. Abb. A) zentrifugiert. Jeweils ein Aliquot der 24 erhaltenen Fraktionen wurde auf ein 8 M Urea - 6 % Polyacrylamid Gel geladen. Das Gel wurde mit Silber gefärbt. Deutlich sichtbar werden die ribosomalen RNAs, von denen der S-Wert bekannt ist. Der Peak der 18 S RNA liegt in den Fraktionen 16 – 18. Mit diesem Wert sedimentiert gleichfalls das U1 snRNP [59].

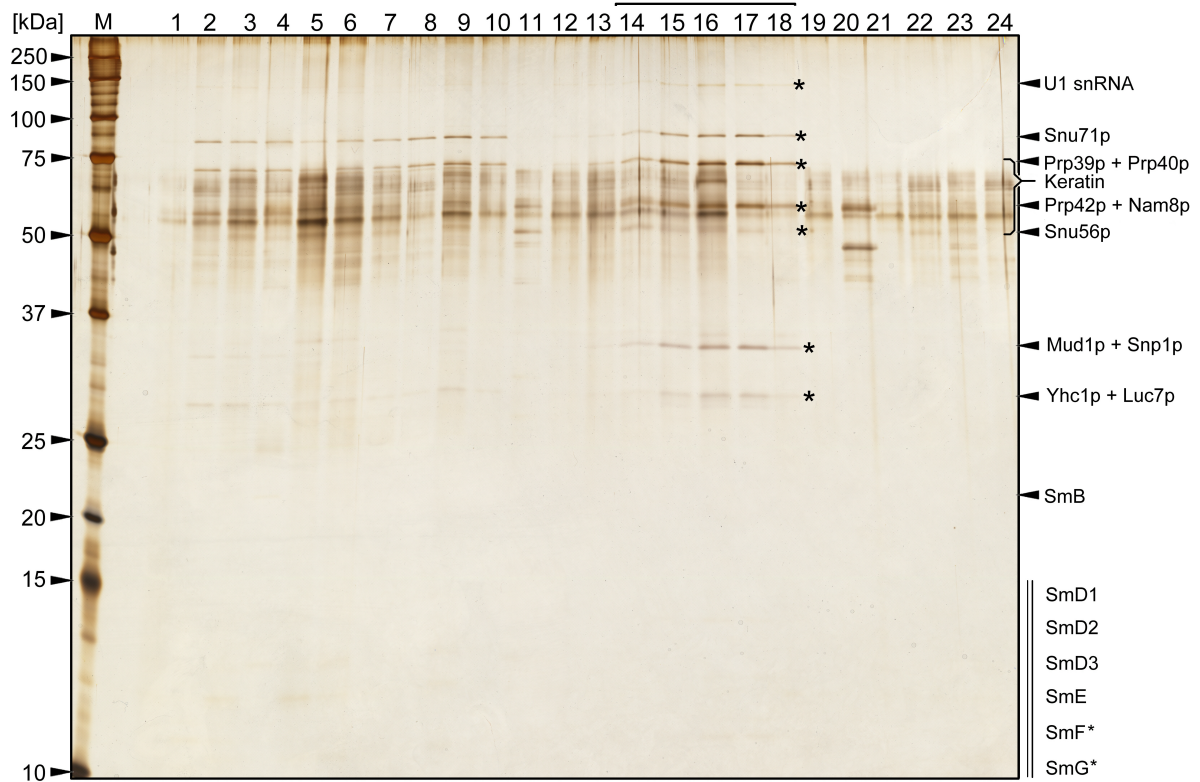


Abbildung 4.5: Proteinanalyse des Gradienten aufgereinigten U1 snRNP. Nach der Vorfraktionierung wurde das U1 snRNP TAP-tag aufgereinigt und anschließend auf einem 10 – 30 % (*w/w*) Glycerolgradienten ultrazentrifugiert (Kap. 3, Abschn. 3.2.1). Aus den Gradientenfraktionen (Gesamtzahl 24, oben 1, unten 24) wurden die Proteine mit Aceton präzipitiert und danach mittels SDS-PAGE auf einem 12% Polyacrylamid Gel analysiert. Das Gel wurde mit Silber gefärbt. In den Fraktionen 14 – 18 sind alle Banden der Proteine des U1 snRNP erkennbar einschließlich der U1 snRNA. M: Proteinstandard zur Angabe der Molekulargewichte, die Massen in [kDa] sind links daneben angegeben. Die zahlreichen weiteren Banden im Bereich von 40 – 60 *kDa* sind eine nachträgliche Keratinverunreinigung der Probe während der Gelanalyse [172] und keine Kontamination im Gradienten, der für die EM verwendet worden war.

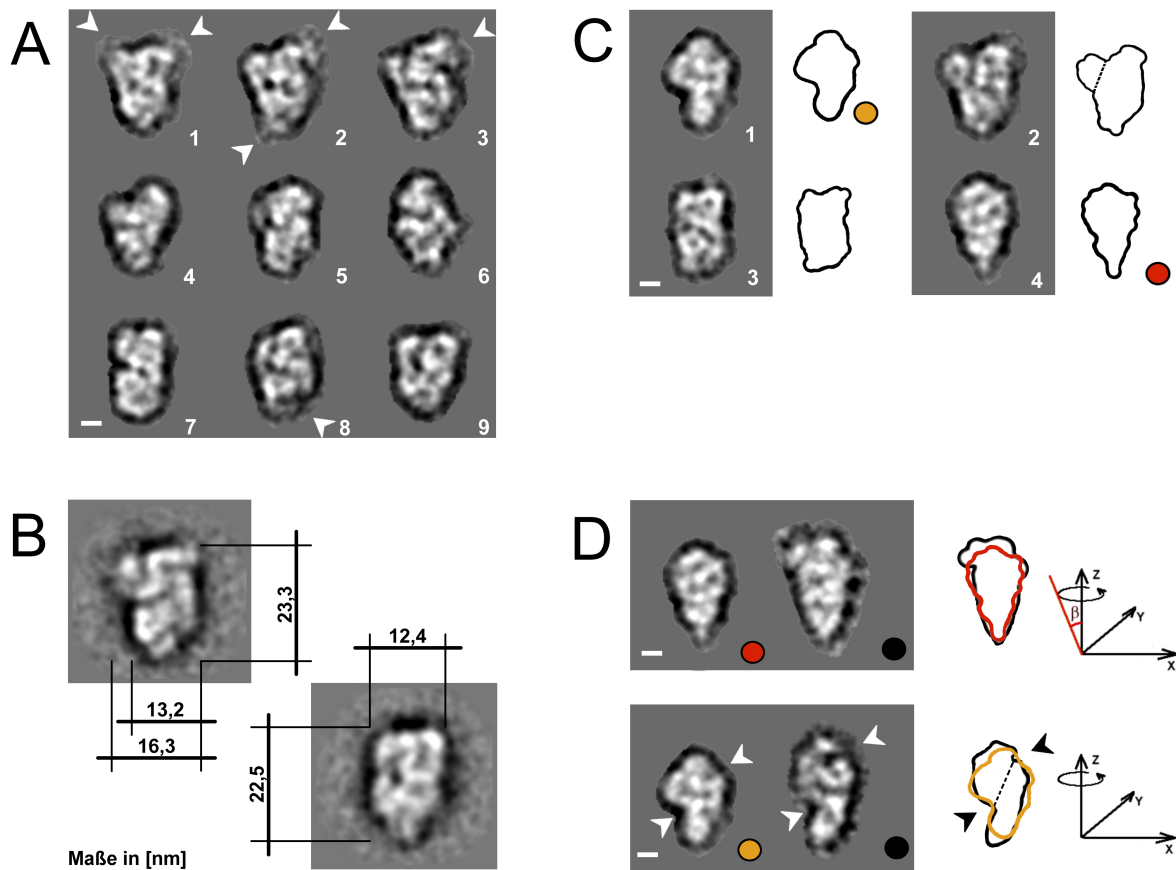


Abbildung 4.6: Charakteristische Klassensummen des über Vorfraktionierung, TAP-tag-Methode und fixierenden Glycerolgradienten aufgereinigten U1 snRNP. Der, wie in Kap. 3, Abschn. 3.2.1 beschrieben, aufgereinigte U1 snRNP wurde bei der EM-Präparation der fixierten Glycerolgradientenfraktionen (16 – 18) mit Uranylformiat negativ kontrastiert (vgl. Kap. 3, Abschn. 3.3.3). Der EM-Bilddatensatz wurde bei RT und 88000-facher Vergrößerung aufgenommen. Die Klassensummen in **A - C** wurden aus 10932 Einzelbildern berechnet. Der Maßstab entspricht 5 nm .

A: Die Klassensummen 1 – 9 zeigen einige 2D Projektionsansichten des U1 snRNP. Die Pfeile deuten auf Bereiche, die sich nicht deutlich vom Kontrastmittelsaum abgrenzen und somit nicht gut definiert sind. Klassensummen 2 und 7 weisen große Ähnlichkeit mit den Projektionsansichten 4 und 9 in Abb. 4.3 auf.

B: Zwei Klassensummen sind mit ihren Abmessungen dargestellt. Die longitudinale Ausdehnung des U1 snRNP liegt demnach bei approx. 230 \AA . Die Ausdehnung in der Breite variiert je nach Projektionsansicht und Meßpunkte von $125 - 165\text{ \AA}$. Die gezeigten Projektionsansichten sind vergleichbar mit den Klassensummen 10 & 8 in Abb. 4.3.

C: Von weiteren 4 Klassensummen wurden Konturen der 2D Ansichten, wie in Abb. 4.2, A vorgeführt, gezeichnet. Die gestrichelte Linie in der Kontur 2 deutet eine vom Hauptkörper der Struktur weitgehend abstehende Domäne an. Die Klassensumme 3 weist große Ähnlichkeit mit der Projektionsansicht 11 in Abb. 4.3, A auf. Die Konturen 1 und 4 wurden in Abb. (D) mit angegebenem Farbcode verwendet.

D: Die Klassensummen 1 und 4 aus Abb. (C) weisen eine leichte Ähnlichkeit mit den Projektionsansichten der unfixierten U1 snRNP Probe aus Abb. 4.2, A, 1 und 7 auf, auch wenn unterschiedliche Projektionsrichtungen, angedeutet mit Hilfe des Koordinatenkreuzes, angenommen werden müssen. In der unteren Abb. deuten die Pfeile auf eine Stauchung der Struktur hin, bedingt durch die stärkere Bindung der oberen Domäne an den unteren Teil der Struktur.

Ein Problem, dass auch mit der TAP-tag Aufreinigung des U1 snRNP mit Vorfraktionierung nicht vollständig behoben werden konnte, waren die in der Datenprozessierung wiederkehrenden undeutlichen Kontrastmittelsäume in den Klassensummen, die auf undefinierte Bereiche hindeuten (Abb. 4.6, A, Pfeile). Eine mögliche Ursache dafür sind flexible Domänen. Da auch auf Ebene der 3D Datenprozessierung Methoden der Trennung verschiedener 3D Substrukturen zur Verfügung standen (*ML-Alignment*, Kap. 2, Abschn. 2.3.2, *supervised classification* Kap. 2, Abschn. 2.3.3), wurde das biochemisch über Vorfraktionierung, TAP-tag-Aufreinigung und abschließender Glycerolgradientenzentrifugation aufgereinigte U1 snRNP (Kap. 3, Abschn. 3.2.1) für die 3D EM-Strukturbestimmung verwendet (Abschn. 4.2.1).

4.1.3 Markierung einzelner Proteine im U1 snRNP

Da es prinzipiell wünschenswert ist, in einer 3D Struktur eines Protein-RNA-Komplexes bestimmte Proteine lokalisieren zu können, waren am U1 snRNP in Zusammenarbeit mit Dr. R. Karaduman (Kap. 3) sowohl Experimente mit spezifischer Antikörper Bindung, als auch mit genetischem Label durchgeführt worden.

mit spezifischen Antikörpern

Immunomarkierung ist in der Einzelmolekül-EM eine gebräuchliche Methode, da Antikörper, wie der Komplex auch, bereits in elektronenmikroskopischen Einzelbildern sichtbar sein können. Ist der Antikörper an den Komplex gebunden, können aus 2D Ansichten des Komplexes gewonnene Informationen über die Lokalisation des Epitops anschließend auf 3D Ebene übertragen werden. Allerdings müssen einige Voraussetzungen gewährleistet sein, um das Zielepitop des Antikörpers zweifelsfrei in der Struktur lokalisieren zu können, von denen die erste die spezifische Bindung des Antikörpers ist.

Für die Antikörperbindungsstudien am U1 snRNP aus *S. cerevisiae* war ein kommerziell erhältlicher polyklonaler Antikörper gegen das Calmodulinbindungspeptid, anti-CBP-AK, ausgewählt worden. Dies hatte zweierlei Gründe. (1) Konventionelle proteinspezifische Antikörper, wenn sie denn für das gewünschte Protein verfügbar sind, weisen häufig in Abhängigkeit von den experimentellen Bedingungen unterschiedliche Sensitivitäten und Spezifitäten auf, was in Folge intensive individuelle Tests und Optimierung der experimentellen Parameter erfordert. (2) Zudem ist nicht gewährleistet, dass die Epitope auf dem ausgewählten Protein im Komplex für den Antikörper zugänglich liegen. Da das Calmodulinbindungspeptid nach der TAP-tag Aufreinigung am Protein verbleibt und mit dem TAP-tag vorher die Aufreinigung möglich war, schienen dies gute Voraussetzungen für die Zugänglichkeit des Epitops und somit für die Bindung des Anti-CBP-AK

an den Komplex zu sein. Desweiteren wurde der anti-CBP-AK bereits in Experimenten mit TAP-tag aufgereinigtem Tri-snRNP (U4/U6.U5 snRNP) aus *S. cerevisiae* erfolgreich als Immunolabel eingesetzt [93].

Für die Bindungsstudien des Anti-CBP-AK an das U1 snRNP wurde das U1 snRNP wie beschrieben (s. Kap. 3, Abschn. 3.2.1), aber ohne Vorfraktionierung, affinitätsaufgereinigt. Das isolierte U1 snRNP wurde mit unterschiedlich hohem molarem Überschuss an anti-CBP-AK (Kap. 3, Abschn. 3.2.2)) inkubiert. Dieses Reaktionsgemisch wurde in einem Glycerolgradienten ultrazentrifugiert, um die Immunkomplexe von freiem Antikörper abzutrennen. Um anschließend den anti-CBP-AK in den Gradientenfraktionen zu detektieren, wurden ELISA-Experimente ausgeführt (Kap. 3, Abschn. 3.2.2), und zwar mit Antikörper vorbeschichteten (Ak:Ziege anti-Hase) ELISA-Platten (Thermo Fisher, USA) und im wesentlichen nach den Instruktionen des Herstellers (siehe auch [92] [93]). Die Absorptionsmessungen wurden bei 450 nm durchgeführt. Nach Normalisierung der Absorptionswerte wurden die Kurven in Abb. 4.7 erhalten, die eine prozentuale Verteilung des anti-CBP-AK in den Gradientenfraktionen bei unterschiedlich hohem molarem Antikörperüberschuss wiedergeben. Ungebundener, freier Antikörper bleibt in den oberen Gradientenfraktionen. Bei den Titrationsexperimenten mit Snu71p-TAP-tag aufgereinigtem U1 snRNP traten keine signifikanten Antikörpermengen in unteren Gradientenfraktionen auf (Abb. 4.7, A und nicht gezeigte Experimente). Bei Wiederholung der Experimente mit auf gleicher Weise aufgereinigtem U1 snRNP mit TAP-tag am Snu56 Protein, sedimentierten bei 20-fachem Überschuss Antikörper in untere Bereiche des Gradienten (Abb. 4.7, B). Die Ko-Migration von Anti-CBP-AK und U1 snRNP war allerdings sehr gering und zudem bei den Titrationsexperimenten mit verschiedenen Antikörperkonzentrationen nicht ausreichend reproduzierbar, was zu zusätzlichen Schwierigkeiten bei der Lokalisation der Antikörper am Komplex in elektronenmikroskopischen Bildern führt (Kap. 5 Diskussion, Abschn. 5.2.1). Eine Lokalisation des anti-CBP-AK Epitops war folglich aufgrund mangelnder Antikörperbindungseffizienz nicht praktikabel.

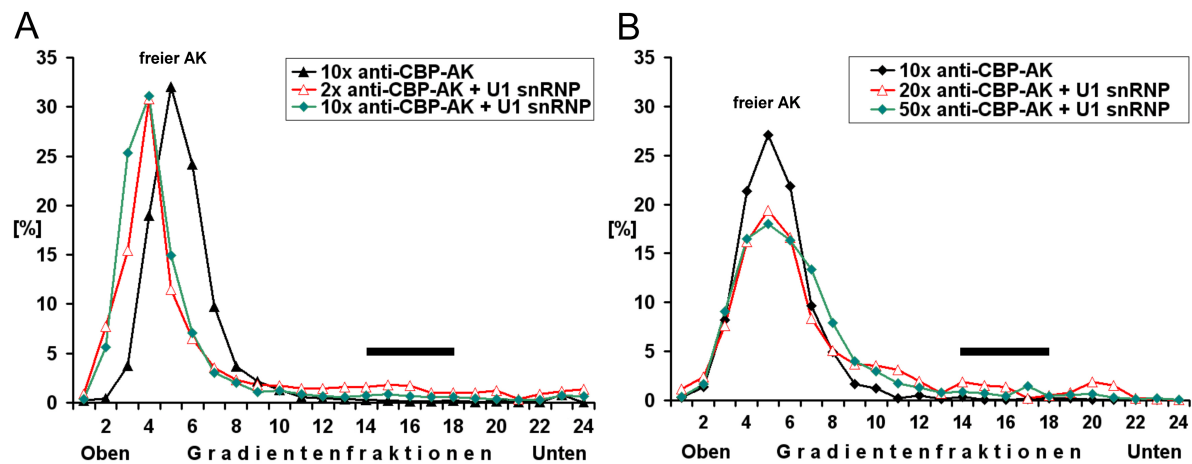


Abbildung 4.7: Titration des anti-CBP Antikörpers gegen TAP-tag aufgereinigten U1 snRNP. TAP-tag aufgereinigter U1 snRNP (ohne Vorfraktionierung, Kap. 3, Abschn. 3.2.1) wurde mit unterschiedlich hohem molarem Überschuss an anti-CBP-AK 1 h bei 4 °C inkubiert und dieses Reaktionsgemisch wurde auf einem 10 – 30 % (*w/w*) Glycerolgradienten ultrazentrifugiert. Die anschließend erhaltenen 24 Gradientenfractionen (x-Achse) wurden mittels ELISA [93] auf den relativen Gehalt an anti-CBP-AK (y-Achse) getestet. Als Kontrollexperiment wurde ein 10-facher molarer Überschuss an anti-CBP-AK ohne U1 snRNP (schwarze Kurve) auf einen Gradienten geladen. Das U1 snRNP wird in den Fraktionen 14 – 18 erwartet, was mit einem schwarzen Balken gekennzeichnet wurde.

A: Snu71p-TAP-tag aufgereinigter U1 snRNP: Es trat keine signifikante Ko-Migration des anti-CBP-AKs mit U1 snRNP auf, so dass es offensichtlich zu keiner ausreichenden Bildung des Immunkomplexes kam. Die leichte Peakverschiebung des freien anti-CBP-AKs ohne U1 snRNP im Kontrollexperiment ist durch die manuelle Ladung und Ernte des Glycerolgradienten bedingt.

B: Snu56p-TAP-tag aufgereinigter U1 snRNP: Es ist eine geringe Ko-Migration des anti-CBP-AKs mit U1 snRNP bei 20-fachen Überschuss des Antikörpers festzustellen.

mit genetischem Label

Die Strategie für die Einführung eines genetischen Labels basiert auf der Möglichkeit ein anhängliches Protein in der EM-Struktur als zusätzliche außerordentliche Dichte erkennen zu können. Je nach Auflösung der Struktur kann zu diesem Zweck bereits ein circa 25 kDa kleines Protein eingesetzt werden, im Falle des U1 snRNP war jedoch das 54 kDa Protein tDimer2 ausgewählt worden. Es handelt sich bei tDimer2 um einen für die Fluoreszenz optimierten Dimer des DsRed Proteins aus *Discosoma* [30]. Im tDimer2 sind die beiden DsRed Untereinheiten über eine kurze Peptidsequenz miteinander verbunden und die konservierten 11-strängigen β -Fass-Strukturen der DsRed Untereinheiten stehen senkrecht zueinander. Die fluoreszierende Eigenschaft des tDimer2 Labels hat keine Bedeutung für die Elektronenmikroskopie, vielmehr waren die Größe und die leichte Verfügbarkeit Gründe für die Wahl. Der tDimer2 liegt auf einem Plasmid in einer optimierten Kasette vor, die speziell für genetisches Fusionsproteindesign in *S. cerevisia* konstruiert worden ist [235]. Analog dazu ist auch das 27 kDa Fluoreszenzprotein yECitrine in einer solchen Kasette verfügbar [235], und mit diesem Label war es bereits möglich gewesen, die Positionen der Lsm Proteine im U6 snRNP und verschiedene Proteine im Tri-snRNP aus *S. cerevisiae* zu identifizieren [93], [115].

Die Expression des U1 snRNP aus genetisch modifizierten Hefestämmen mit chromosomal integriertem tDimer2-Fusionsprotein am C-Terminus von (1) Nam8p und (2) Snp1p wurde getestet. Die Aufreinigung des U1 snRNP erfolgte dafür wie beschrieben (Kap. 3, Abschn. 3.2.1) und ohne Vorfraktionierung über die TAP-methode. Die Isolierung eines intakten 18 S U1 snRNP aus Hefezellen mit Nam8p-tDimer2 war nicht möglich, wie die Proteingelanalyse mit Silberfärbung der Glycerolgradientenfractionen zeigte (ohne Abb.). Eine Aufreinigung des U1 snRNP mit dem tDimer2-Label am Snp1p war möglich wie die Proteinanalyse der Glycerolgradientenfractionen zeigte (Abb. 4.8). Verglichen mit den Proteinanalysen ohne tDimer2 Fusionsprotein in den Abbildungen 4.1, 4.5, 4, 5 wurde eine neue zusätzliche Bande gerade oberhalb der Snu71p Bande erkenntlich, die aufgrund der erwarteten Masse von 88 kDa dem Snp1-tDimer2-Fusionsprotein zugeordnet wurde. Zudem waren die Peakfraktionen des U1 snRNP um zwei Fraktionen nach oben verschoben, was auf ein verändertes Sedimentationsverhalten aufgrund der tDimer2 Integration hindeutete. Elektronenmikroskopische Begutachtung von diesen aufgereinigten U1 snRNP zeigte, dass die Anzahl von Partikeln, die auf den Kohlefilm absorbiert waren, verglichen mit der U1 snRNP Probe ohne Label, erheblich reduziert waren. Dieses war auch in analogen Experimenten in *S. cerevisiae* am APC [56] beobachtet worden (pers. Kommunikation Dr. P. Dube). Aufgrund der strukturellen Heterogenität des U1 snRNP (Abschn. 4.3) war eine Lokalisation des tDimer2 nicht möglich.

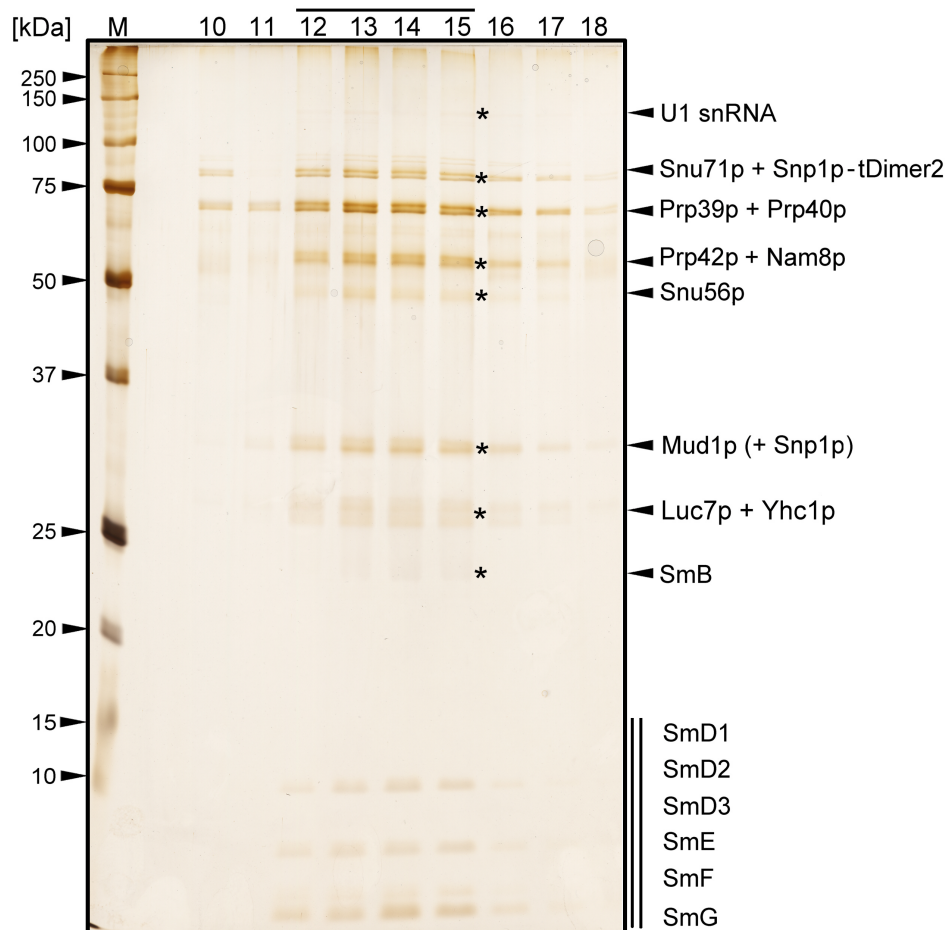


Abbildung 4.8: Untersuchung der Proteinexpression des tDimer2 Fusionsproteins Snp1p im U1 snRNP. Aus einem Hefestamm mit tDimer2 Label am C-Terminus von Snp1p wurde U1 snRNP mittels der TAP-tag-Methode ohne Vorfraktionierung (Kap. 3, Abschn. 3.2.1), aufgereinigt und auf einem 10 – 30 % (*w/w*) Glycerolgradienten zentrifugiert. Aus den Glycerolgradientenfraktionen (Gesamtzahl 24, oben 1, unten 24) wurden die Proteine mit Aceton präzipitiert, danach mittels SDS-PAGE auf einem 12% Polyacrylamid Gel aufgetrennt und schließlich mit Silber gefärbt. Gezeigt sind die Fraktionen, in denen das U1 snRNP erwartet wird. Das Proteinbandenmuster des U1 snRNP ist um zwei Fraktionen nach oben verlagert und ist in den Fraktionen 12 – 15 wiederzuerkennen. Dies deutet nur auf ein verändertes Sedimentationsverhalten des U1 snRNP hin, und kann nicht pauschal auf einen leichteren Partikel zurückgeführt werden. Die Snp1p-tDimer2 Bande liegt gerade oberhalb der Snu71p-Bande.

4.2 3D Strukturen des U1 snRNP

4.2.1 *de novo* 3D Struktur-Rekonstruktion

Um eine initiale 3D Struktur des U1 snRNP unter Berücksichtigung der bereits in 2D Projektionsbildern beobachteten Heterogenität der Proben (Abschn. 4.1) zu erhalten, wurde die sogenannte *Random Conical Tilt*-Methode ([198], Kap. 2, Abschn. 2.3.1) ausgewählt. Dazu wurde das U1 snRNP einschließlich der Vorfraktionierung wie in Kap. 3, Abschn. 3.2.1 beschrieben über den TAP-tag aufgereinigt und während der abschließenden Glycerolgradientenzentrifugation fixiert [119]. Der U1 snRNP der entsprechenden Gradientenfraktionen wurde auf einen dünnen Kohlefilm aufgebracht und mit Uranylformiat negativ kontrastiert (Kap. 3, Abschn.3.3.3). Die Kupfernetzträger dieser Proben wurden bei RT aufbewahrt und mikroskopiert. Zur Aquisition der für die RCT-Rekonstruktion notwendigen Daten wurden elektronenmikroskopische Aufnahmen derselben Probenbereiche bei 0° und 45° Einstellung des EM-Probenhalters direkt nacheinander gemacht. Die 0° -Aufnahme entstand stets zuerst. Die Bilder wurden bei 88000-facher Vergrößerung in euzentrischer Höhe bei einem Defokus von $-0,9 \mu\text{m}$ bis $-1,4 \mu\text{m}$ bezogen auf die 0° -Probenhalterstellung aufgenommen. Ein Ausschnitt eines dieser paarweise entstandenen Aufnahmen ist in Abb. 4.9 gezeigt.

Um einen Datensatz aus Einzelpartikelbildern bei 0° und einen exakt korrespondierenden Datensatz derselben Partikel, aufgenommen unter einem 45° Winkel, zu erhalten, wurden die Partikel mit Hilfe des Programms CowHide manuell selektiert. Beispielhaft sind Einzelpartikelpaare in Abb. 4.9 farblich hervorgehoben. Für die anschließende RCT-Rekonstruktion stand ein Datensatz aus 10713 Einzelpartikelpaaren zur Verfügung. Zunächst wurden die Einzelpartikelbilder der 0° Aufnahmen in Klassen von durchschnittlich 54 Mitgliedern klassifiziert. (Kap. 2, Abschn. 2.2). Von jeder der erhaltenen insgesamt 200 Klassen wurde eine RCT-3D-Struktur berechnet (Abb. 4.10), indem den korrespondierenden 45° -Projektionsansichten aller Mitglieder einer 0° Klassensumme die Winkel zugewiesen wurden (Kap. 2, Abschn. 2.3.1). Da nur wenige Projektionsansichten für eine RCT-Struktur-Berechnung verwendet wurden (durchschnittl. 54) und zudem Projektionsrichtungen fehlen (*missing cone*-Phänomen, Kap. 2, Abschn. 2.3.1), weisen die Volumina der 3D RCT-Rekonstruktionen wie in Abb. 4.10, C deutlich wird, ein geringes SNR auf.

Um das SNR zu verbessern und die Strukturen besser beurteilen zu können, wurden die RCT-3D-Strukturen im *ML-Alignment* rotationsausgerichtet und anschließend mittels der 3D-PCA (Kap. 2, Abschn. 2.3.2) so klassifiziert, dass durchschnittlich 8 RCT-3D-Strukturen eine Klassensumme bildeten. Es wurden auf diese Weise 25 3D Strukturen

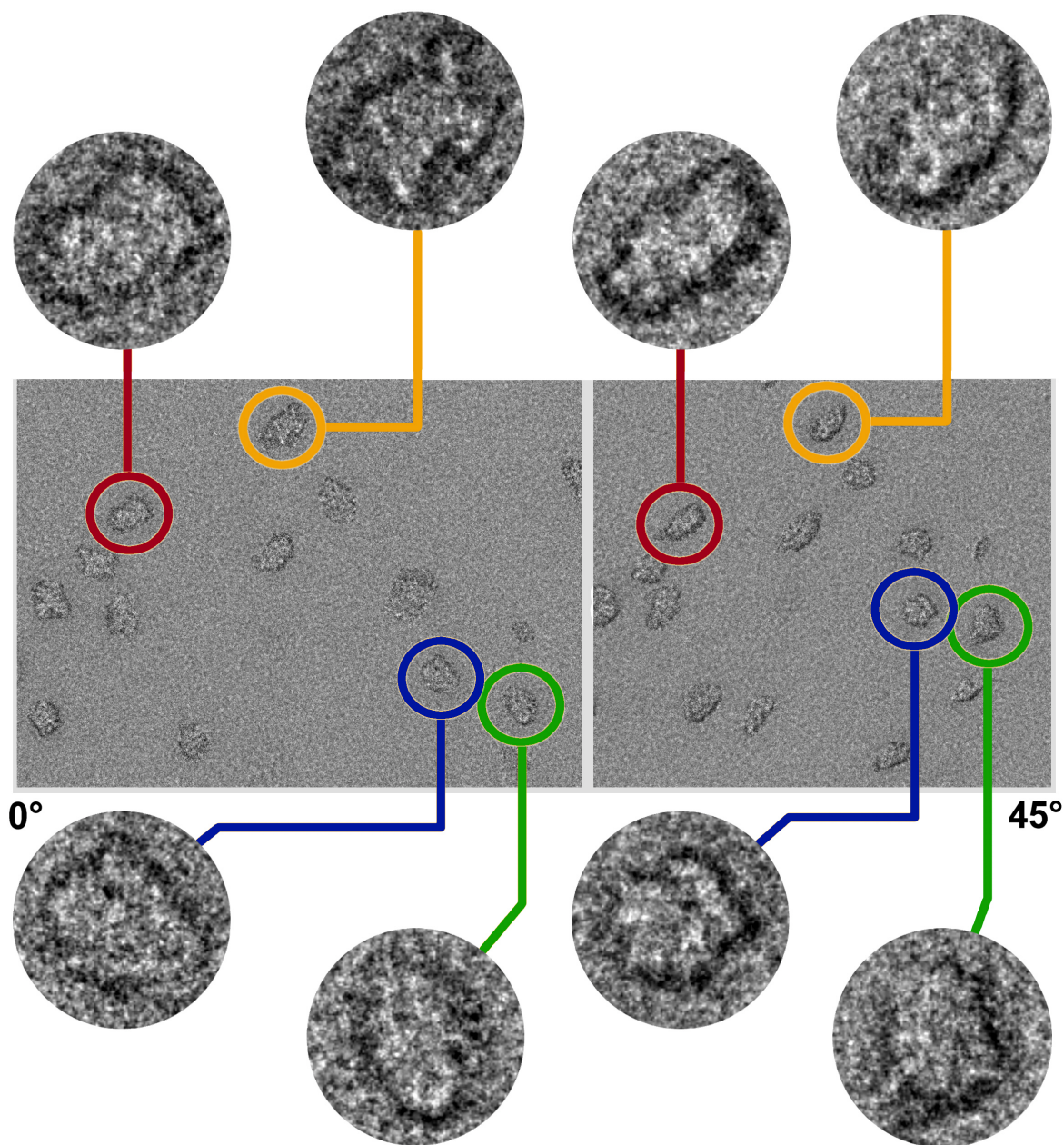


Abbildung 4.9: Elektronenmikroskopische Aufnahmen bei 0° und 45° geneigtem Probenhalter. Gezeigt sind Ausschnitte von elektronenmikroskopischen Aufnahmen des mit Uranylformiat negativ kontrastierten U1 snRNP. Die EM-Bilder entstanden direkt nacheinander bei RT und vom selben Probenbereich auf dem Kupfernetzträger, aber bei unterschiedlicher Neigung des EM-Probenhalters, und zwar unter einem Winkel von 0° (1. Aufnahme, links) und 45° (2. Aufnahme, rechts). Für die RCT-Rekonstruktion wurden unter Verwendung des Programms CowHide die Partikel auf den einander entsprechenden Aufnahmen visuell einander zugeordnet und, wie in dieser Darstellung farblich hervorgehoben, paarweise selektiert und in Einzelbildern extrahiert. Die dargestellten Einzelpartikelbilder sind bereits gefiltert.

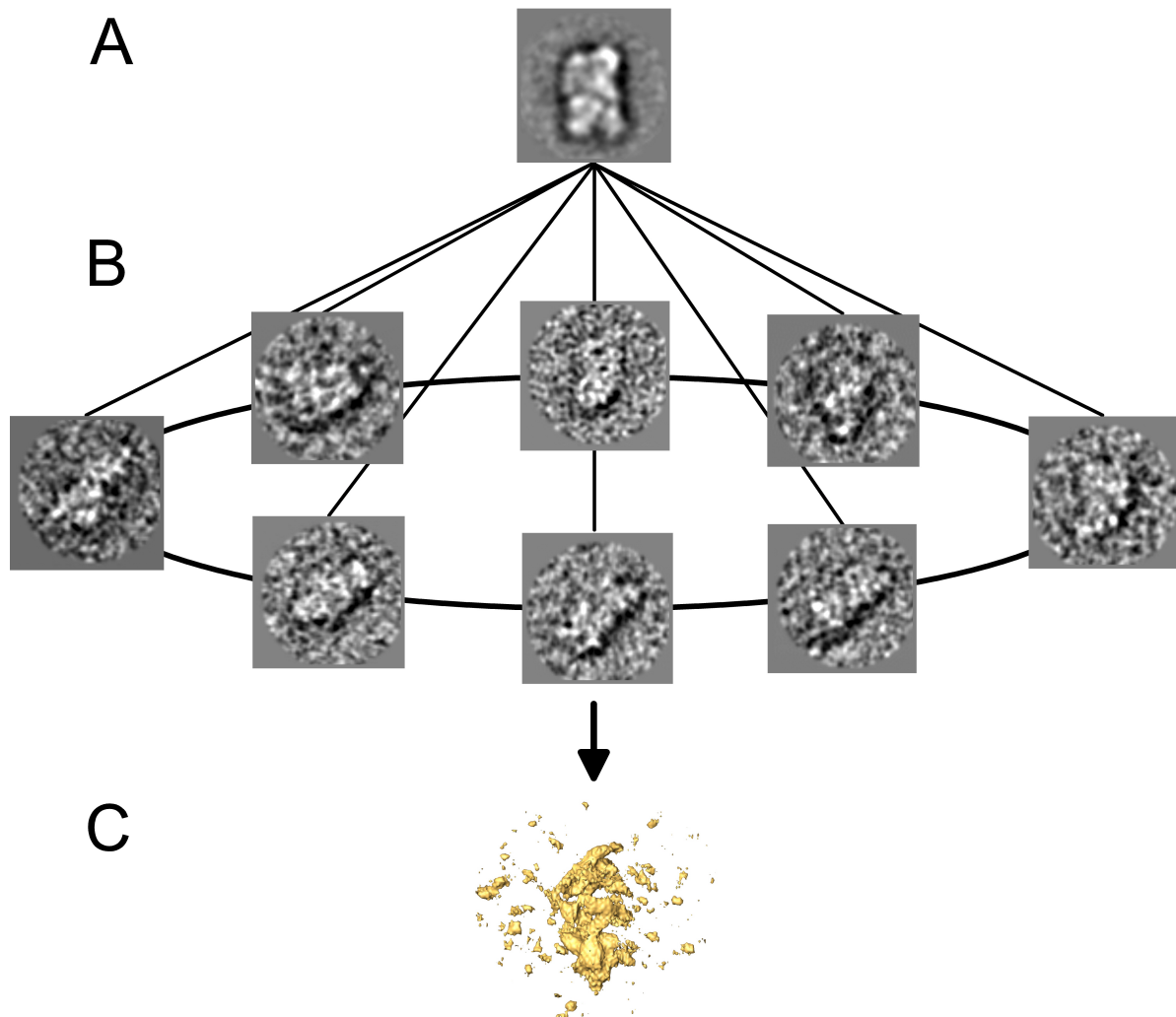


Abbildung 4.10: *de novo* 3D Rekonstruktion des U1 snRNP mittels RCT-Methode.

Aus einer Klassensumme des U1 snRNP (A), berechnet aus einem Datensatz von 10713 Einzelpartikelbildern wie unter Kap. 2, Abschn. 2.2 beschrieben, wurden die korrespondierenden Klassenmitglieder des zugehörigen Datensatzes, der aus EM-Aufnahmen bei 45° geneigtem Probenhalter entstand (8 Mitglieder dieser Klasse sind in B gezeigt), verwendet, um eine 3D RCT-Struktur (C) zu berechnen, wie in Kap. 2, Abschn. 2.3.1 beschrieben. In Abb. C wird das in diesem Stadium der Datenprozessierung generell geringe SNR der ersten RCT-3D-Strukturen ersichtlich.

erhalten. Von diesen wurden nach visueller Begutachtung 5 Strukturen als initiale Referenzen für ein weiteres 3D Struktur-*Refinement* selektiert (Abb. 4.11). Die Wahl fiel auf Strukturen, die miteinander verglichen sehr unterschiedlich waren, um möglichst das gesamte Spektrum an auftretenden Konfigurationen und Konformationen des Komplexes abzudecken. Die Struktur I in Abb. 4.11 war in dem Datensatz einzigartig, während von den übrigen Strukturen II - IV jeweils einige ähnliche vorhanden waren, die vergleichbare Strukturmerkmale aufwiesen. In diesen groben 3D Strukturen wurde ersichtlich, dass in der U1 snRNP Probe Subpopulationen unterschiedlich konfigurierter Komplexe existierten. So ist die Struktur IV sehr viel schmäler, als die Strukturen I und II in Abb. 4.11, was allein mit einer Konformationsumwandlung nicht erklärbar ist.

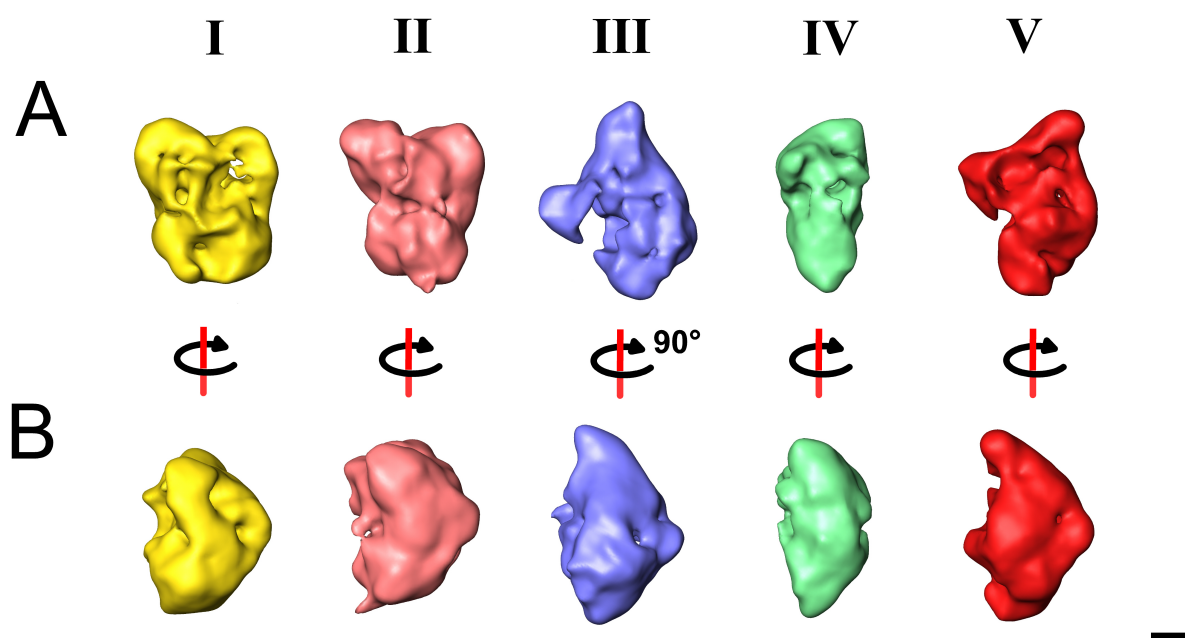


Abbildung 4.11: 3D Strukturen des U1 snRNP nach ML-Alignment und anschließender 3D Klassifizierung. Aus einem Datensatz von 10713 Einzelbildern wurden, wie beispielhaft in Abb. 4.10 für eine Klassensumme gezeigt und wie unter Kap. 2, Abschn. 2.3.1 beschrieben, 200 RCT-3D-Strukturen des U1 snRNP kalkuliert. Nach anschließender Rotationsausrichtung im ML-Alignment, wie unter Kap. 2, Abschn. 2.3.2 beschrieben, wurden die RCT-3D-Strukturen mit einer durchschnittlichen Klassenmitgliederzahl von 8 klassifiziert. In (A) sind 5 charakteristische RCT/ML-Strukturen des U1 snRNP abgebildet und in (B) ihre um 90° im Uhrzeigersinn um die Längsachse gedrehte Ansicht. Diese Strukturen wurden für ein anschließendes Struktur-*Refinement* in der zyklischen Datenprozessierung auf 3D Ebene als initiale Referenzstrukturen verwendet. Der Maßstab in A und B entspricht 5 nm .

4.3 Verbesserung und Validierungen der 3D Strukturen

Um die initialen Strukturen (Abb. 4.11), die aus den RCT-Rekonstruktionen mit anschließendem *ML-Alignment* und 3D Klassifizierung erhalten worden sind (Kap. 2, Abschn. 2.3.1), zu verbessern, wurde ein zyklisches Struktur-*Refinement*-Verfahren abgeschlossen (vgl. Kap. 2, Abschn. 2.3.3 & Abschn. 2.2, Abb. 2.4). Das Verfahren basiert auf der genaueren Eulerwinkelbestimmung für die Projektionsansichten, die zur 3D Rekonstruktion verwendet werden. Um die strukturelle Heterogenität des U1 snRNP, die sich in den 5 ausgewählten Strukturen widerspiegelt, weiterhin zu berücksichtigen, wurde im *Refinement*-Prozedere ein *MR-Alignment*, das in [72] beschrieben ist, verwendet. Nach diesem Verfahren wurden als Referenzen Sets von 2D Projektionsbilder aller 5 3D Modelle gleichzeitig zur Ausrichtung der Bilder des Einzelpartikeldatensatzes benutzt. Anschließend wurden die Einzelpartikelbilder des Datensatzes danach klassifiziert, welches 3D Modell ihrer Ausrichtung zugrunde lag. Auf diese Weise wurden Subdatensätze erhalten, aus denen jeweils das entsprechende 3D Modell neu rekonstruiert werden konnte. Dazu wurde aus den Subdatensätzen zunächst mittels *supervised classification* ein 3D Modell erstellt. Dieses wurde in den Tests zur Validierung von Strukturcharakteristika (Abb. 4.12 & 4.17) jedoch lediglich dazu verwendet, um durch Ausprojizieren ein neues Winkelreferenzset für die *angular reconstitution*-Methode zu erhalten. Um nämlich keinen Bias einzubringen, wurden die Einzelpartikel-Subdatensätze PCA-basiert klassifiziert (CowHide Programm, Kap. 2, Abschn. 2.2.2) und die erforderliche nachfolgende Winkelbestimmung für die Klassensummen erfolgte mittels *angular reconstitution*. Das so rekonstruierte 3D Modell wurde für einen erneuten *Refinement*-Zyklus verwendet. Die für das *Refinement* eingesetzten Einzelpartikeldatensätze basierten auf negativ kontrastierten Proben. Einer der Datensätze entstand durch EM-Aufnahmen bei RT, 2 weitere unter Probenkühlung mit flüssigem Stickstoff.

Validierung eines Strukturfortsatzes

Gleich zu Anbeginn des Struktur-*Refinements* wurde ein Strukturfortsatz in 4 der 3D Modelle beobachtet (3D Modelle II - V), der bei weiteren Zyklen schnell überbetont wurde. Um zu überprüfen, inwieweit dieser Strukturfortsatz Gültigkeit hat, wurde der Einzelpartikeldatensatz aufgetrennt, indem 3D Modelle mit und ohne Fortsatz für ein *MR-Alignment* eingesetzt wurden. Der Datensatz wurde dementsprechend in 2 Subdatensätze, die jeweils die 4 Modelle mit beziehungsweise ohne Strukturfortsatz repräsentierten, gespalten (Abb. A & B 4.12). Für den Datensatz ohne Strukturfortsatz (Test A) wurden

Referenzen mit Strukturfortsatz eingesetzt (Abb. 4.12, A, Pfeile an 3D Referenzen) und entsprechend umgekehrt wurde mit dem Subdatensatz mit Strukturfortsatz (Abb. 4.12, B) verfahren, indem Referenzen ohne Strukturfortsatz eingesetzt wurden (Test B).

Die Testergebnisse zeigten nur für die Struktur II ein einheitliches Ergebnis mit einem gut definierten Fortsatz. Danach gibt es von der Struktur II Substrukturen mit und ohne Fortsatz. Ein angedeuteter Fortsatz trat bereits im RCT/ML 3D Modell der Struktur auf (Abb. 4.11, II). Auch mit einem automatischen *Refinement*-Prozedere, das auf Benutzerinteraktion weitgehend verzichtet und die vorhandenen in Kap. 2 beschriebenen Programme über ein in Python 2.5.1. programmiertes Skript (Dr. Florian Hauer) nacheinander aufruft, wurde der Fortsatz nur in der Struktur II ausgebildet (Daten nicht gezeigt). In den 3D Modellen III und V verschwand der Fortsatz im Test A nicht deutlich genug, um 2 Subpopulationen bestätigen zu können. Bei der Betrachtung der Ergebnisse für das 3D Modell IV fällt auf, dass der Fortsatz im Ergebnismodell von Test B ungewöhnlich spitz zuläuft und der Fortsatz nicht die gleiche Form annimmt wie im Ausgangsmodell. Daher kann nicht abschließend beurteilt werden, bis zu welchem Maß der Fortsatz in der Struktur IV Gültigkeit hat. Gleiches gilt für die Modelle III und V. Überdies wurde nach weiteren *Refinement*-Zyklen auch deutlich, dass der Fortsatz, der in diesen anfänglichen Untersuchungen in 3D Modell II betrachtet wurde, nicht gleichzusetzen ist, mit denen in den übrigen Modellen. In dem 3D Modell II entwickelte sich nämlich auch an anderer Stelle eine schmal zulaufende Strukturausprägung (vgl. Abb. 4.13 II, Pfeile). Es ist möglich, dass an dieser Stelle auch Rauschen korreliert und die zugrunde liegenden Daten somit kein weitergehendes *Refinement* zulassen.

Analyse der im *Refinement* erhaltenen 3D Modelle

In Abb. 4.13 sind die 3D Modelle I - V gezeigt, wie sie nach insgesamt 7 *Refinement*-Zyklen erhalten worden sind. Abgebildet sind die 3D Modelle, die nach dem letzten MR-*Alignment* mit *projection matching* erhalten worden sind. Es lag dem *Alignment* ein Datensatz von 107147 Kryo-negativ-kontrastierten Einzelpartikelbilder zugrunde. In Tabelle 4.1 ist aufgelistet, in welchem Verhältnis sich der Datensatz auftrennte. Desweiteren ist auch angegeben, welche Masse für die Volumina der in Abb. 4.13 gezeigten Strukturen I - V geschätzt wird.

Die Fourierschalenkorrelationsfunktionen (FSC-Kurven) der 3D Strukturen, mit deren Hilfe die Auflösungsgrenze einer 3D EM-Struktur nach verschiedenen Kriterien angegeben werden kann (Kap. 2, Abschn. 2.3.6), ist in Abb. 4.14 gezeigt. Die Strukturen II - V wurden dazu vor der Berechnung im Frequenzbereich von unter 0,1 [$1/\text{\AA}$] gefiltert (Abb. 4.14, grüne Kurven). Ohne diese Filterung wiesen diese 4 FSC-Kurven (Abb. 4.14 rote Kurven von 3D-II bis 3D-V) einen erneuten Anstieg in dem Frequenzbereich auf. Dieser

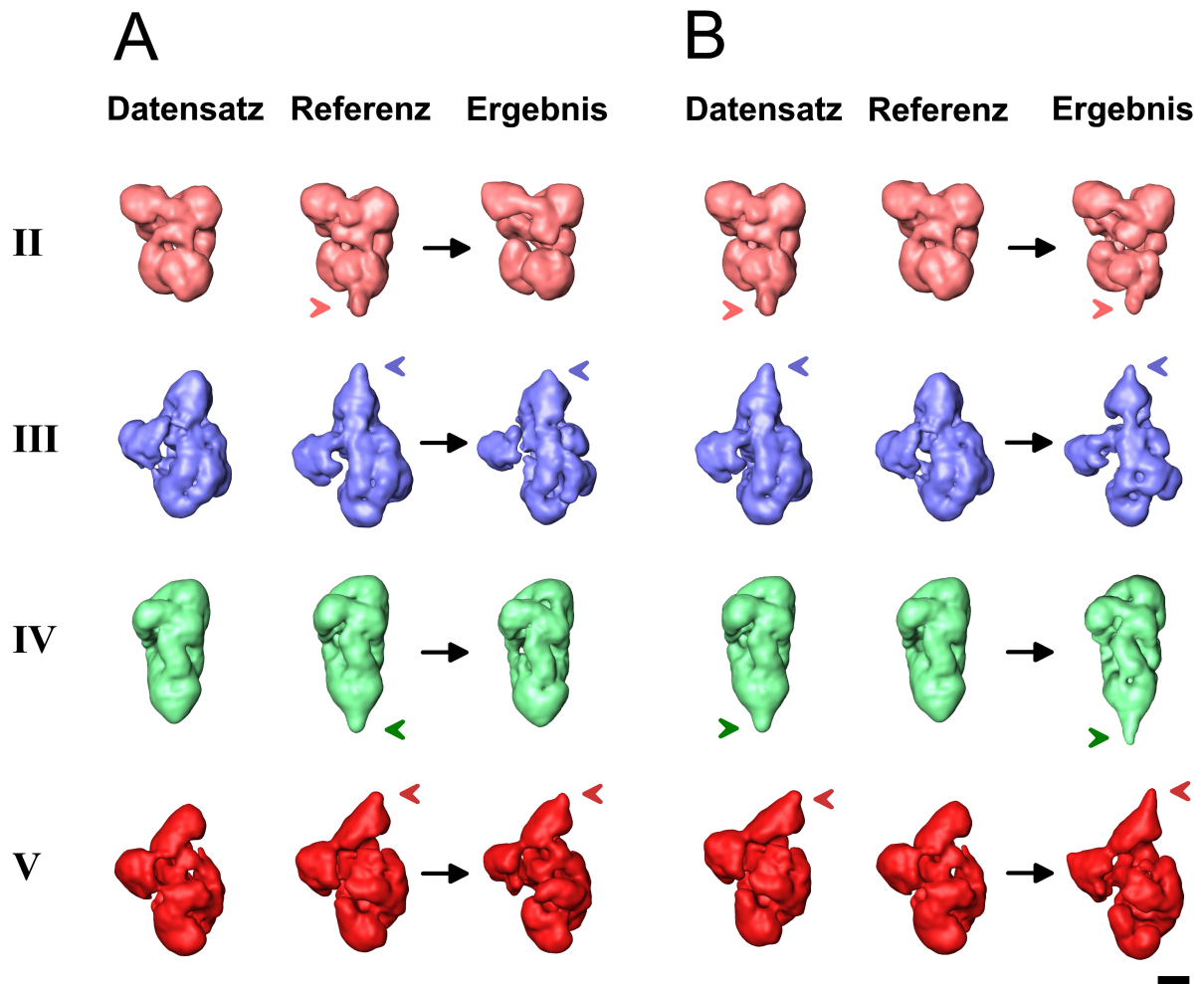


Abbildung 4.12: Refinement-Zyklen zur Validierung einer U1 snRNP Substruktur. Die Einzelbilder eines Datensatzes aus negativ kontrastierten U1 snRNP Partikeln wurden getrennt nach Vorhandensein eines Strukturfortsatzes (Pfeile), wobei die Subdatensätze in (A) mit 11815 Einzelbildern und in (B) mit 11447 Einzelbildern 4 unterschiedliche 3D Strukturen des U1 snRNP repräsentierten (farblich unterschieden). Als initiale Referenzen wurden für den kompletten Datensatz ohne Fortsatz nun die jeweiligen 3D Strukturen mit Fortsatz eingesetzt. Entsprechend umgekehrt wurde mit dem Datensatz mit Fortsatz in B verfahren, indem diesem initiale Referenzen ohne Fortsatz zugewiesen wurden. Die jeweiligen Ergebnisse, die nach 3 Refinement-Zyklen mittels *supervised classification* und *angular reconstitution* erhalten wurden, sind abgebildet. Lediglich Struktur II lässt sich eindeutig in zwei Substrukturen mit und ohne einem vertrauenswürdigen Strukturfortsatz trennen (Pfeilspitze). Der Maßstab entspricht 5 nm.

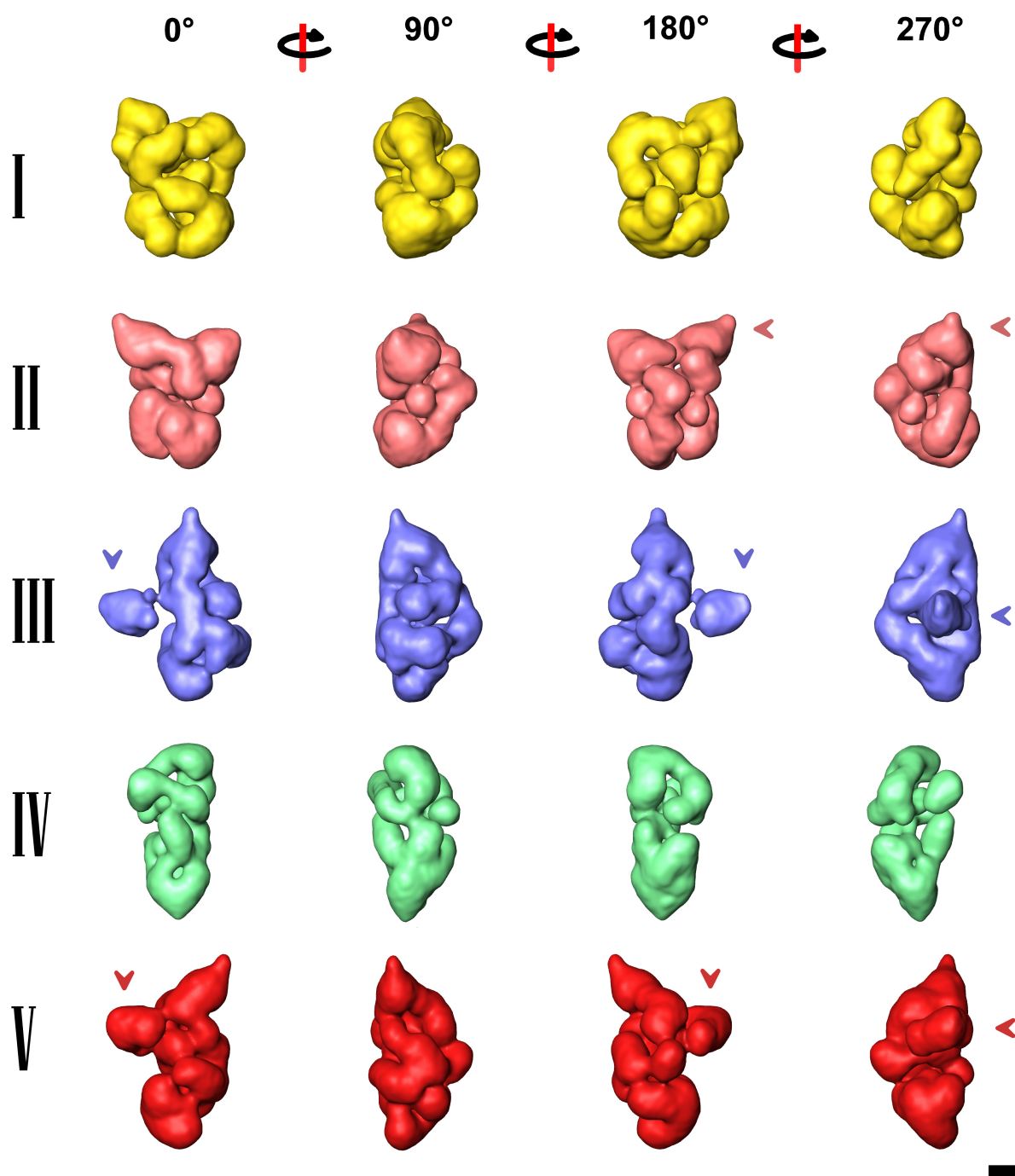


Abbildung 4.13: 3D Strukturen des U1 snRNP aus *S. cerevisiae*. Die abgebildeten Strukturen I - V des U1 snRNP wurden nach einem zyklischen Prozedere zur Verbesserung der 3D Strukturen mittels einer Kombination aus *angular reconstitution* und *projection matching* (s. Kap. 2 Abschn. 2.3.3) erhalten. Als initiale Referenzen dienten die RCT/ML-Strukturen aus Abb. 4.10. Unmittelbar liegt den Strukturen ein gemeinsamer Datensatz bestehend aus 107147 Einzelpartikelbildern zugrunde. Die U1 snRNP Probe des Datensatzes war mit Uranylformiat negativ kontrastiert und anschließend in flüssigem Stickstoff schockgefroren worden (*Cryo-negativ-stain-Methode* nach [76]). Der Datensatz wurde unter Stickstoffkühlung der Probe aufgenommen. Die Strukturen werden von 4 Seiten gezeigt, wobei jeweils um 90° im Uhrzeigersinn um eine Längsachse gedreht wurde. Die Pfeile in Abb. III deuten auf eine Domäne, die nur schwach mit dem Hauptkörper der Struktur assoziiert ist und nachfolgend auf ihre Zugehörigkeit getestet wurde (Abb. 4.17). Der Maßstab entspricht 5 nm .

3D Struktur	Anzahl der Einzelbilder im Subdatensatz	Anteil am Gesamtdatensatz in [%]	ungefähre Masse der Strukturvolumina in [kDA]
Struktur I	6777	6,8	1530
Struktur II	26839	25,0	1250
Struktur III	15991	14,9	1400
Struktur IV	28981	27,0	920
Struktur V	28559	26,7	1420

Tabelle 4.1: 3D Strukturen I - V, kalkuliert aus 107147 Einzelpartikelbildern einer kryonegativ-kontrastierten U1 snRNP Probe.

Anstieg kann grundsätzlich verschiedenen Ursachen geschuldet sein [185]. Der einfachste Grund dafür ist die Anwendung einer Maske auf die 3D Volumina, was in diesem Fall ausgeschlossen werden kann. Sehr wahrscheinlich liegt es hier an einer Korrelation im Rauschen. Aufgrund der beobachteten Heterogenität unter den Klassensummen und den RCT/ML-Strukturen ist es wahrscheinlich, dass sich die zugrunde liegenden Subdatensätze aus mehreren Strukturkonformationen zusammensetzen, die nicht aufgetrennt werden konnten. Die 3D-I Struktur ist nach den FSC-Kurven zu urteilen am besten definiert.

Um zu ermitteln, ob die Winkelverteilung der Projektionsansichten in den Subdatensätzen gleichmäßig ist, wurden die einzelnen Datensätze PCA-basiert mit circa 30 Mitgliedern pro Klasse klassifiziert (Kap. 2, Abschn. 2.2) und die Winkel der erhaltenen Klassensummen wurden mittels *angular reconstitution* bestimmt. Die Projektionsansichten wurden mit Hilfe eines Unterprogramms von CowHide (Dr. M. Schmeißer) auf der Eulerkugel angeordnet und im Adobe Reader 9.0 visualisiert (Abb. 4.15 & 4.16). Danach ist die Winkelverteilung in keinem der Subdatensätze unregelmäßig. Alle Projektionsrichtungen sind gleichmäßig vertreten, auch in dem zugrunde liegenden Datensatz der 3D-I Struktur, für den nur 230 Klassensummen gebildet worden sind. Dort sind die Projektionsansichten der Pole zwar leicht unterrepräsentiert, was in diesem Fall zunächst auf einen statistischen Effekt zurückgeführt werden kann.

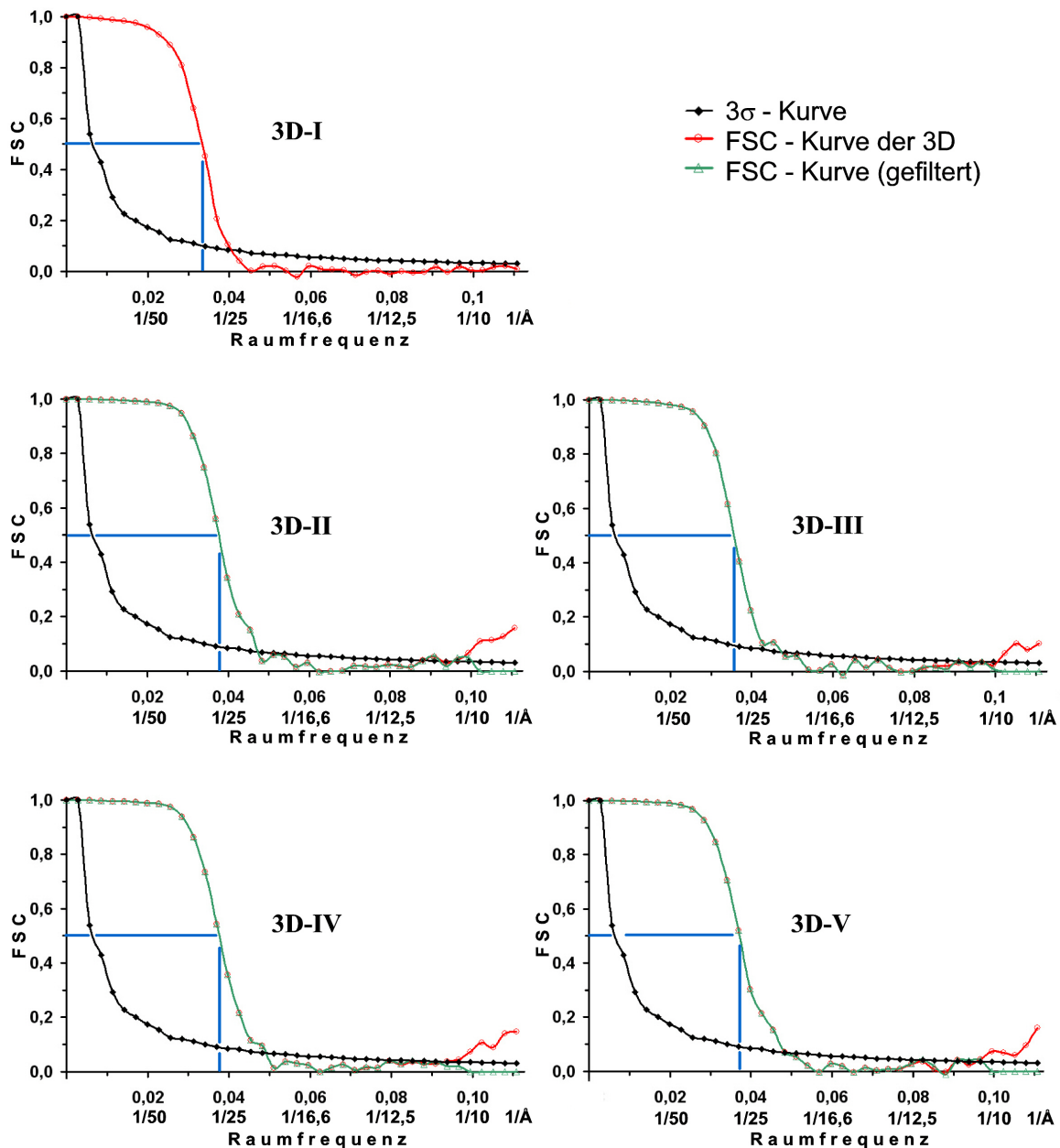


Abbildung 4.14: Auflösung der berechneten 3D Strukturen. Die Diagramme zeigen die Fourier-Schalenkorrelationsfunktionen (FSC) der 5 berechneten U1 snRNP Strukturen **I - V** aus Abb. 4.12. In rot ist jeweils der Verlauf der FSC-Funktion in Abhängigkeit der Raumfrequenz gezeigt. Die Strukturen **II - V** wurden vor der Berechnung im Bereich von unter 10 Å gefiltert und der Verlauf dieser Funktionen ist in grün dargestellt. In den meisten Frequenzbereichen überlagern sich die Verläufe von gefilterter und ungefilterter FSC-Kurve. Die schwarze Kurve gibt den Wert des statistischen Rauschens berechnet nach dem 3σ Kriterium bei der jeweiligen Frequenz an. Am Schnittpunkt dieser Kurve mit der jeweiligen FSC-Kurve wird die maximale Auflösungsgrenze nach dem 3σ Kriterium als reziproker Wert der Raumfrequenz abgelesen, die demnach für die Strukturen **I - V** zwischen 25 und 20 Å liegt. Nach dem $FSC_{0,5}$ -Kriterium beträgt die Auflösungsgrenze den blau eingezeichneten Linien folgend zwischen 25 und 30 Å. Die 3D-I Struktur ist, am Verlauf der FSC-Kurve geurteilt, am besten definiert, während die Strukturen **II - V** wahrscheinlich eine Korrelation von Rauschen aufweisen [185], erkennbar am erneuten Anstieg der FSC-Kurven im hohen Frequenzbereich.

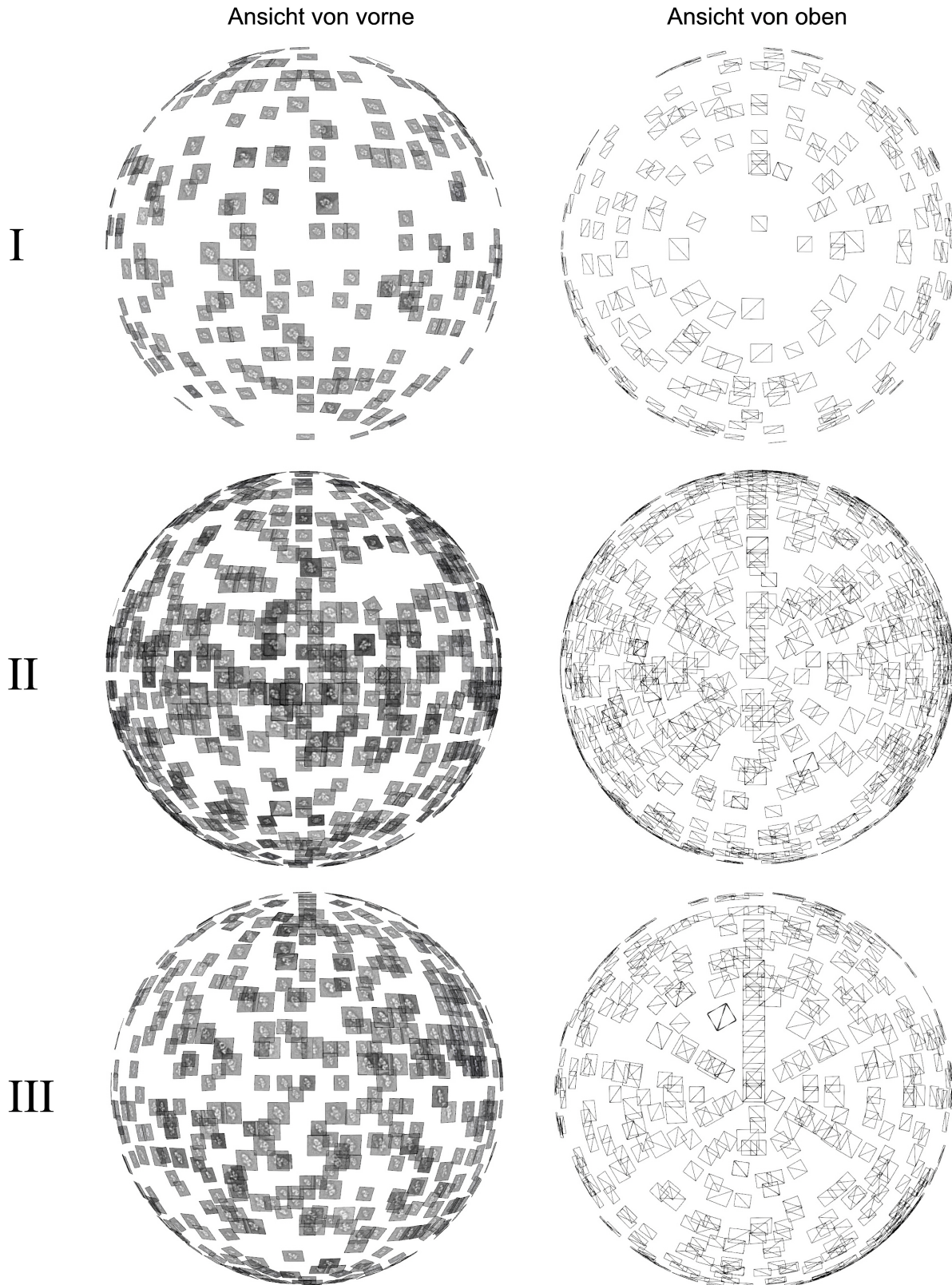


Abbildung 4.15: Projektionswinkelverteilung in den Subdatensätzen I - III. Detaillierte Bildbeschreibung erfolgt unter Abb. 4.16.

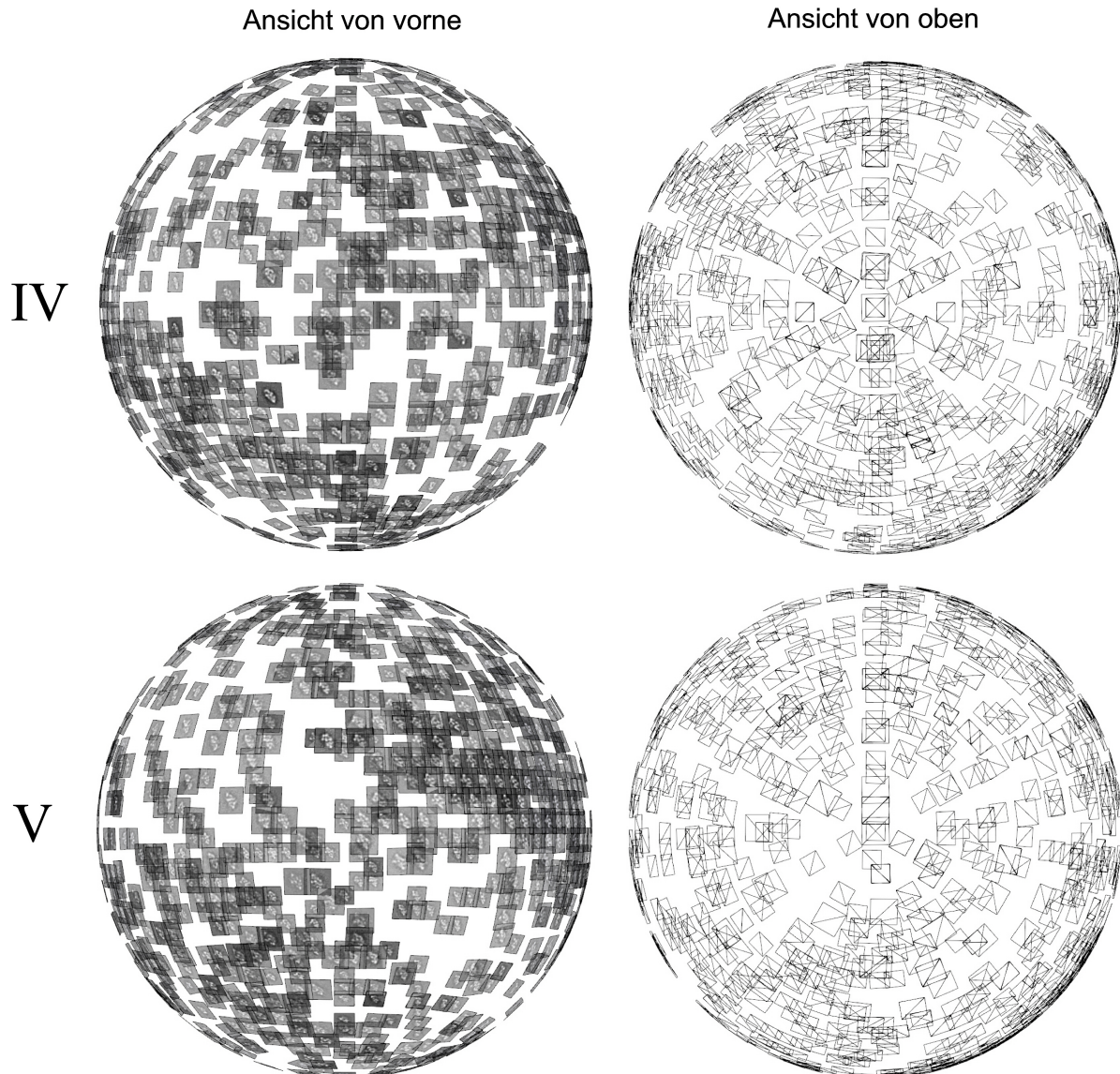


Abbildung 4.16: Projektionswinkelverteilung in den Subdatensätzen IV und V. Bildunterschrift gilt auch für Abb. 4.15.

Die Subdatensätze aus Einzelpartikelbildern, aus denen die 3D Strukturen **I - V** in Abb. 4.12 mittels *projection matching* berechnet worden waren, wurden mit CowHide-Unterprogrammen klassifiziert, wie unter Kap. 2, Abschn. 2.2.2 beschrieben, wobei die durchschnittliche Anzahl an Klassenmitgliedern 30 betrug. Die Anzahl der Klassensummen betrug für die einzelnen Subdatensätze 230 (I), 900 (II), 530 (III) und 950 (IV & V). Den erhaltenen Klassensummen wurden mittels *angular reconstitution* Eulerwinkel zugewiesen, wobei die Projektionen der jeweils zugehörigen 3D Strukturen **I - V** aus Abb. 4.12 als Referenzwinkelset dienten. Die Klassensummen der Subdatensätze **I - V** wurden anschließend mit ihren Eulerwinkeln auf der Eulerkugel unter Benutzung eines Unterprogramms des Softwarepakets von CowHide platziert (Dr. M. Schmeißer, im Hause) und mit dem AdobeReader 9.0 visualisiert. Die Eulerkugeln der Subdatensätze **I - V** sind von jeweils 2 Seiten abgebildet, einer Frontalansicht und der entsprechenden Ansicht von oben mit abstrahierten Projektionsbildern. Projektionsbilder mit großem Quadrat liegen dem Betrachter zugewandt, mit kleinem Quadrat auf der Rückseite der Kugel. Alle Subdatensätze weisen demnach eine gleichmäßige Winkelverteilung auf.

Validierung der anhänglichen großen Domäne in 3D-III

Die 3D Struktur III wies eine große zusätzliche Dichte auf, die nur leicht mit dem Hauptkörper der Struktur verbunden war (Abb. 4.13), aber im bisherigen *Refinement*-Prozess nicht verloren ging. Es sollte daher überprüft werden, ob die Dichte noch der Struktur angerechnet werden darf. Da die 3D Struktur V in derselben Abb. (4.13) eine Ähnlichkeit mit dem Hauptkörper der Struktur III aufwies, und zwar mit der Vorstellung, dass die eine in die andere Form durch einfache Konformationsumwandlung überführt werden kann, wurde ebenso diese 3D-V-Struktur auf ihre anhängliche Domäne untersucht (Abb. 4.17). Dazu wurde zunächst ein Datensatz in Subpopulationen von Einzelparti-

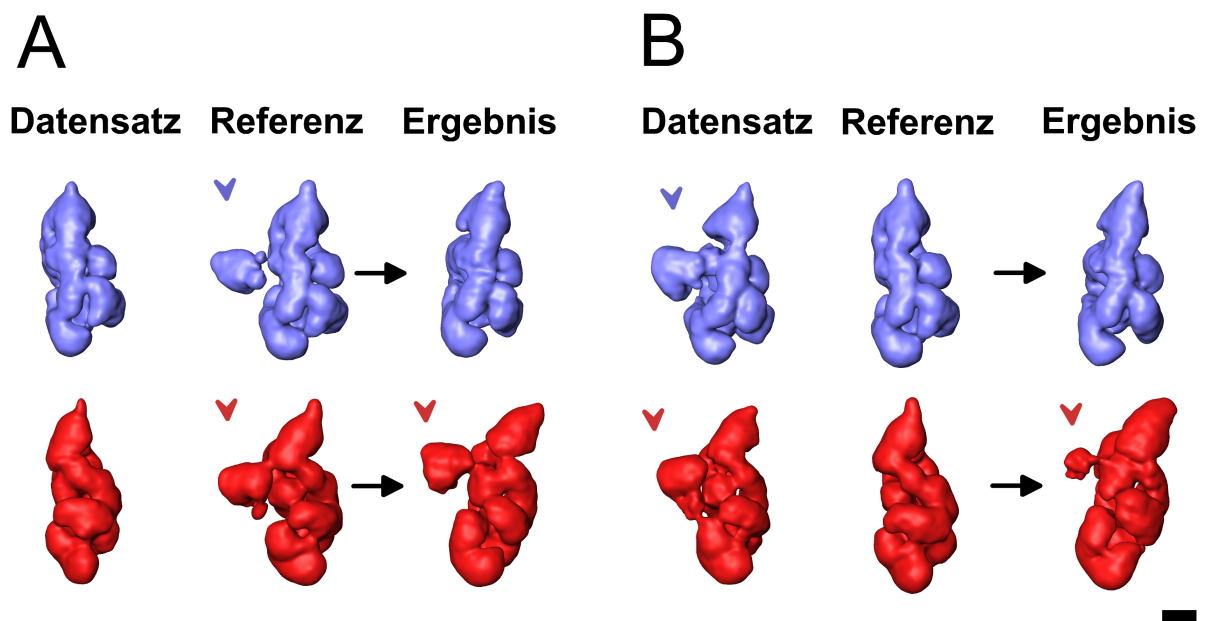


Abbildung 4.17: *Refinement*-Zyklen zur Validierung einer U1 snRNP Domäne. Die Einzelbilder eines Datensatzes aus negativ kontrastierten U1 snRNP Partikeln wurden getrennt nach Vorhandensein der Domäne (Pfeile), wobei die Subdatensätze in (A) mit 14827 Einzelbildern und in (B) mit 10783 Einzelbildern 2 unterschiedliche 3D Strukturen des U1 snRNP repräsentierten (farblich unterschieden). Als initiale Referenzen wurden für den kompletten Datensatz ohne Domäne nun die jeweiligen 3D Strukturen mit Domäne eingesetzt. Entsprechend umgekehrt wurde mit dem Datensatz mit Domäne in B verfahren, indem diesem initiale Referenzen ohne Domäne zugewiesen wurden. Die jeweiligen Ergebnisse, die nach 3 *Refinement*-Zyklen mittels *supervised classification* und *angular reconstitution* erhalten wurden, sind abgebildet. Der Maßstab entspricht 5 nm.

kelbildern aufgetrennt, indem Referenzen mit und ohne Domäne beider Modelle eingesetzt wurden. Anschließend wurde den Subdatensätzen, die die beiden Strukturen ohne Domäne repräsentierten, eine Referenz mit Domäne vorgegeben (Abb. 4.17, A). Der Subdatensatz mit Domänen erhielt Modellvorgaben ohne Domäne. Nach 3 *Refinement*-Zyklen, die wie oben beschrieben ausgeführt wurden, verlor die Domäne in der 3D-III

Struktur deutlich den Kontakt zum Hauptkörper und schwächte sich auch insgesamt in ihrer Ausdehnung ab. Es schien dort aber immer noch verstärkt Rauschen zu korrelieren. Allerdings spricht dieses Ergebnis stark dafür, dass die betrachtete Domäne kein Bestandteil des Komplexes ist. Für die 3D-V Struktur ergab sich ein uneinheitliches Bild. Die Ergebnis-Strukturen waren insgesamt nicht mehr so gut definiert, so dass keine Schlussfolgerungen möglich sind. Da von der Domäne vorher nicht beobachtet worden war, dass der Kontakt zum Hauptkörper vage oder gering ist, sind dort wohl andere Einflüsse der Strukturheterogenität dafür verantwortlich, nicht aber die untersuchte Domäne.

Die fünf 3D Strukturen in Abb. 4.13 sind untereinander verglichen in ihren strukturellen Merkmalen sehr verschiedenartig. Da die ermittelten Massen der gezeigten Strukturvolumina in Relation zueinander ebenfalls sehr unterschiedlich sind, spiegelt dies Heterogenität in der Konfiguration der Komplexe wider. Daneben zeigen die Strukturen III und V, abgesehen von der zusätzlichen Domäne in Struktur V, ähnliche Strukturmerkmale, die mit Konformationsumwandlungen erklärbar sind.

4.4 Fehlerabschätzung der 3D Rekonstruktion bei Heterogenität der Probe

4.4.1 Testdaten und Verfahren

Um die Fehler der 3D Berechnungen bei zugrunde liegenden Datensätzen strukturell heterogener Proteinkomplexe wie dem U1 snRNP aus *S. cerevisiae* besser einschätzen zu können, wurde eine Fehlerberechnung mit einem Testdatensatz durchgeführt. Dazu wurde zunächst ein Testdatensatz, der das Fehlen einer Domäne zu unterschiedlich hohem Prozentsatz simulieren sollte, wie folgt generiert: Ein pdb-File, der kristallographische Daten in Form von Atomkoordinaten eines humanen Protein-RNA-Komplexes beinhaltete, wurde in Imagic-5 eingelesen und in eine 3D Dichteverteilung im Imagic-5 File-Format umgewandelt. Unter den festgelegten Bedingungen beim Umwandeln in die 3D Dichte, wurde eine Masse für die Modellstruktur in Imagic-5 berechnet und damit verbindlich festgesetzt (vgl. Tab. 4.2). Um nun einen unterschiedlich konfigurierten Proteinkomplex mit dem Wegfall einer Domäne zu simulieren, wurde eine Domäne im originalen pdb-File in PyMol 0.99 editiert, indem die Koordinateninformationen dieser Domäne in unterschiedlichem Umfang gelöscht wurden. Diese editierten pdb-Files wurden in gleicher Weise wie das Original in Imagic-5 importiert. Gleichfalls wurden jeweils die Massen der 3D Dichten kalkuliert und prozentual verglichen mit der originalen

3D Struktur (Tab. 4.2). Auf diese Weise entstand eine Reihe von 3D Testdaten bestehend aus insgesamt 10 3D Modellen (Original (Orig.) & A - I; Abb. 4.18) mit prozentual unterschiedlich hohem Verlust einer Domäne.

	3D Modell	Masse in [kDA]	Masseverlust in [%]
1	Modell (original)	320,4	0,0
2	Modell A	308,1	3,8
3	Modell B	297,8	7,1
4	Modell C	288,5	10,0
5	Modell D	277,9	13,3
6	Modell E	272,0	15,1
7	Modell F	264,8	17,4
8	Modell G	258,6	19,3
9	Modell H	249,6	22,1
10	Modell I	244,7	23,6

Tabelle 4.2: 3D Modelle des Testdatensatzes

Testdurchführung

Um den Einfluß der reduzierten Masse auf die Winkelbestimmung mit der Methode der *angular reconstitution* zu prüfen, wurde jedes der Modelle mit einem Winkelabstand von 10° am 'Nordpol' ($\alpha = 0, \beta = 0, \gamma = 0$) startend gleichmäßig über der Eulerkugel ausprojiziert. Der α -Winkel betrug dabei stets 0° . Es wurde so für jedes der 3D Modelle ein in den Projektionsrichtungen/-winkeln gleiches Set von 406 Projektionsbildern erhalten, die sich bis auf die reduzierte Domäne glichen. Die 2D Projektionsbilder des Originalmodells wurden als Referenzset festgelegt. Von allen 2D Projektionsbildersets wurden die Sinogramme erstellt (Kap. 2, Abschn. 2.3.3), um anschließend mit dem festgelegten Referenzset den Projektionsansichten der Modelle A - I die 3 Eulerwinkel mittels der *angular reconstitution*-Methode neu zuzuweisen. Um die Ungenauigkeit der Methode per se festzustellen und um einen Wert bei einer Massenreduktion von 0%, zu erhalten, wurden auch für die Projektionsbilder des Originalmodells, die eigentlich als Referenzset dienten, neue Eulerwinkel bestimmt; in diesem Fall war also Stichprobe und Referenzset identisch. Da diese Ergebnisse bei der statistischen Auswertung für eine Art der Normalisierung benutzt wurden, wird im folgenden dieser Test auch mit 'Nulltest' bezeichnet.

In einem weiteren Test wurden die Projektionsbilder der Modelle A - I vor der Winkelbestimmung mit der *angular reconstitution*-Methode in einem MR-*Alignment* mit allen Projektionsbilder des Originalmodells ausgerichtet. Durch das MR-*Alignment* wird der α -Winkel verändert.

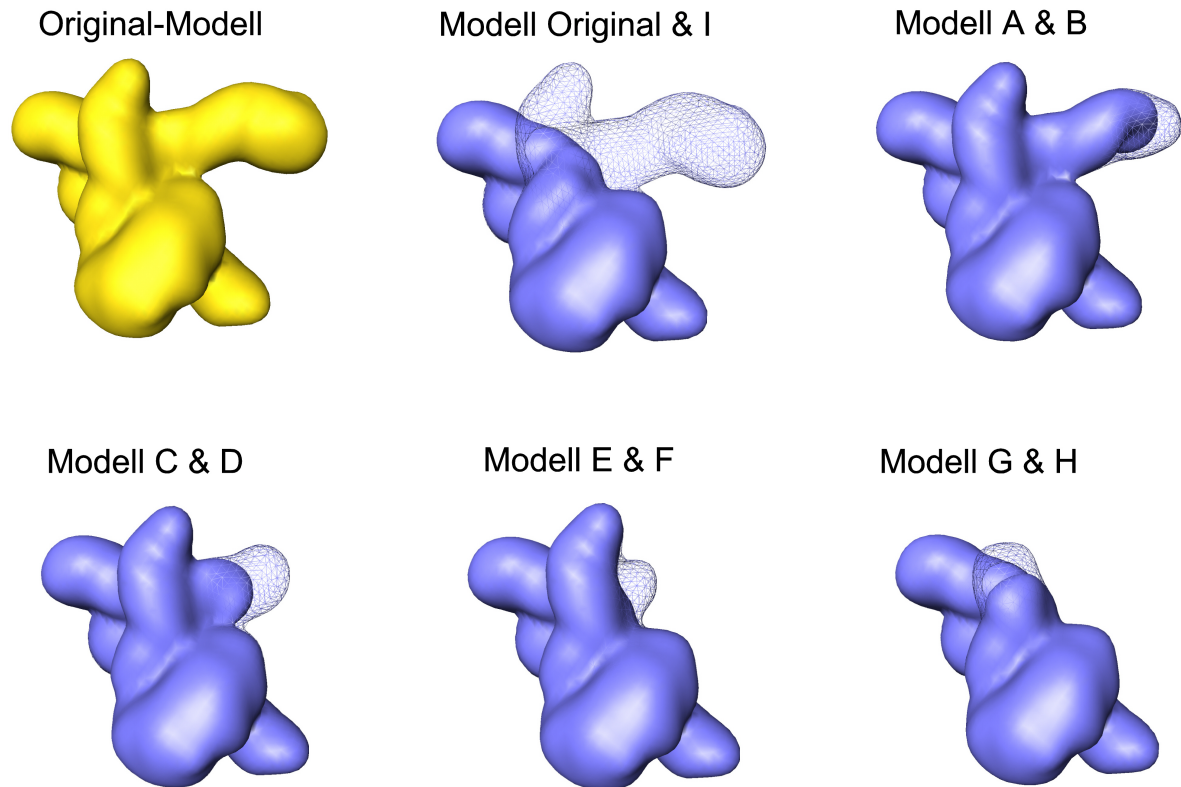


Abbildung 4.18: Visualisierung der 3D Modelle des Testdatensatzes. Die originale 3D Struktur des Testdatensatzes wurde aus einem pdb-File des humanen U1 snRNP durch den Import in Imagic-5 generiert (**Gelbe Struktur**). Die weiteren 3D Strukturen (**Modelle A - I**) des Testdatensatzes entstanden durch Editieren des originalen pdb-Files und wurde unter den gleichen, einmalig festgelegten, Konstanten in Imagic-5 eingelesen. Der kleinsten Struktur im Datensatz, dem Modell I fehlt 23,6% der Masse des Originals. Beide Modelle sind zusammen abgebildet, wobei das Original als blaues Netz gezeigt ist, um so die Größe der editierte Domäne visuell deutlich zu machen. Diese Domäne ist in den weiteren 3D Strukturen des Testdatensatzes (**Modelle A - F**) zu jeweils unterschiedlich hohem Prozentsatz verkleinert worden. Die Verkleinerung wurde im pdb-File im Softwareprogramm PyMol 0.99 vorgenommen. Nach Umwandlung der 3D Koordinateninformationen in eine EM-Dichte im Imagic-5 Dateiformat wurde der prozentuale Masseverlust der Strukturen bezogen auf die Originalstruktur kalkuliert. In den blauen Modellabbildungen zeigen die Umrissse der jeweils einfarbig blauen Struktur die in der Reihe folgend kleinere 3D Modellstruktur, während das blaue Netz dementsprechend die Umrissse der massereichen 3D Modellstruktur zeigt.

3D Modelle Original & I: minus 0% und minus 23,6%

3D Modelle A (Netz) & B (blau): minus 3,8% & minus 7,1%

3D Modelle C (Netz) & D (blau): minus 10,0% & minus 13,3%

3D Modelle E (Netz) & F (blau): minus 15,1% & minus 17,4%

3D Modelle G (Netz) & H (blau): minus 19,3% & minus 22,1%

Testauswertung

Nach Testdurchführung wurden die Abweichungen bestimmt, die zwischen den neu zugewiesenen Eulerwinkeln der Projektionsansichten der Modelle A - I & Orig. gegenüber den ursprünglichen, die auch im Referenz-Projektionsbildersset verwendet wurden, auftraten. Dazu wurde eine Fehlerberechnung durchgeführt, deren Grundlagen in Abb. 4.19 dargestellt sind. Bei dieser Fehlerberechnung wurde der α -Winkel, der eine Drehung des Bildes in der Bildebene bedeutet und ohnehin bei Entstehen der verwendeten Projektionsdatensätze auf Null gesetzt wurde, vernachlässigt, zumal dieser Winkel bei dem ersten Test ohne MR-Alignment auch nicht explizit verändert wurde. Jede Projektionsrichtung ließ sich danach allein mit den Winkeln β und γ im eulerschen Koordinatensystem beschreiben. Unter Annahme eines Radius von 100 [Å], war mit diesen Winkeln für jedes Projektionsbild n ein Punkt P_n auf der Eulerkugel definiert (mit $n = 1$ bis 406). Nach Umrechnung unter Anwendung allgemeiner trigonometrischer Formeln wurden die Punkte mit ihren Koordinaten im kartesischen Koordinatensystem angegeben (Abb. 4.19, A). Da die Projektionsbildersets aus jeweils 406 Projektionsansichten bestanden, wurden entsprechend 406 Punkte berechnet.

Um anschließend die Abweichungen der Winkelzuweisung zu bestimmen, wurde für jedes Projektionsbild n der Abstand Δl_n der Punkte von Referenz $P_{n,Ref}$ und betrachtetem Modell $P_{n,Mod}$ berechnet (Abb: 4.19, B). Diese Strecke wurde benutzt, um den Raumwinkel Ψ zwischen den Vektoren, aufgespannt durch eben diese beiden Punkte ($P_{n,Ref}$ und $P_{n,Mod}$), zu berechnen. In der nachfolgenden statistischen Auswertung der Abweichungen wurde der Raumwinkel Ψ verwendet.

4.4.2 Statistische Auswertung

Mittlere Abweichung

Für die statistische Auswertung wurde zuerst die mittlere Abweichung in Grad, ergo der Mittelwert von Ψ , ermittelt. Dabei war zunächst zur Bestimmung des Winkels Ψ das Referenzwinkelset benutzt worden, das auch als tatsächliches Referenzwinkelset im *angular reconstitution* diente, und mit dem alle Modelle ursprünglich ausprojiziert worden waren. Der Mittelwert von Ψ wurde anschließend für jedes der 10 Modelle (Original & A-I) einzeln unter Einbeziehung aller seiner Projektionsbilder (Anzahl 406) berechnet. Das Ergebnis ist in Abb. 4.20 in einem Diagramm gezeigt (grüne Kurve), in dem der Mittelwert in Grad gegen den prozentualen Masseverlust des Modells aufgetragen wurde. Das Originalmodell weist danach eine durchschnittliche Abweichung von circa 6° auf. Die mittlere Abweichung erhöht sich auf über 10° mit prozentual zunehmendem Masseverlust aufgrund der fehlenden Domäne.

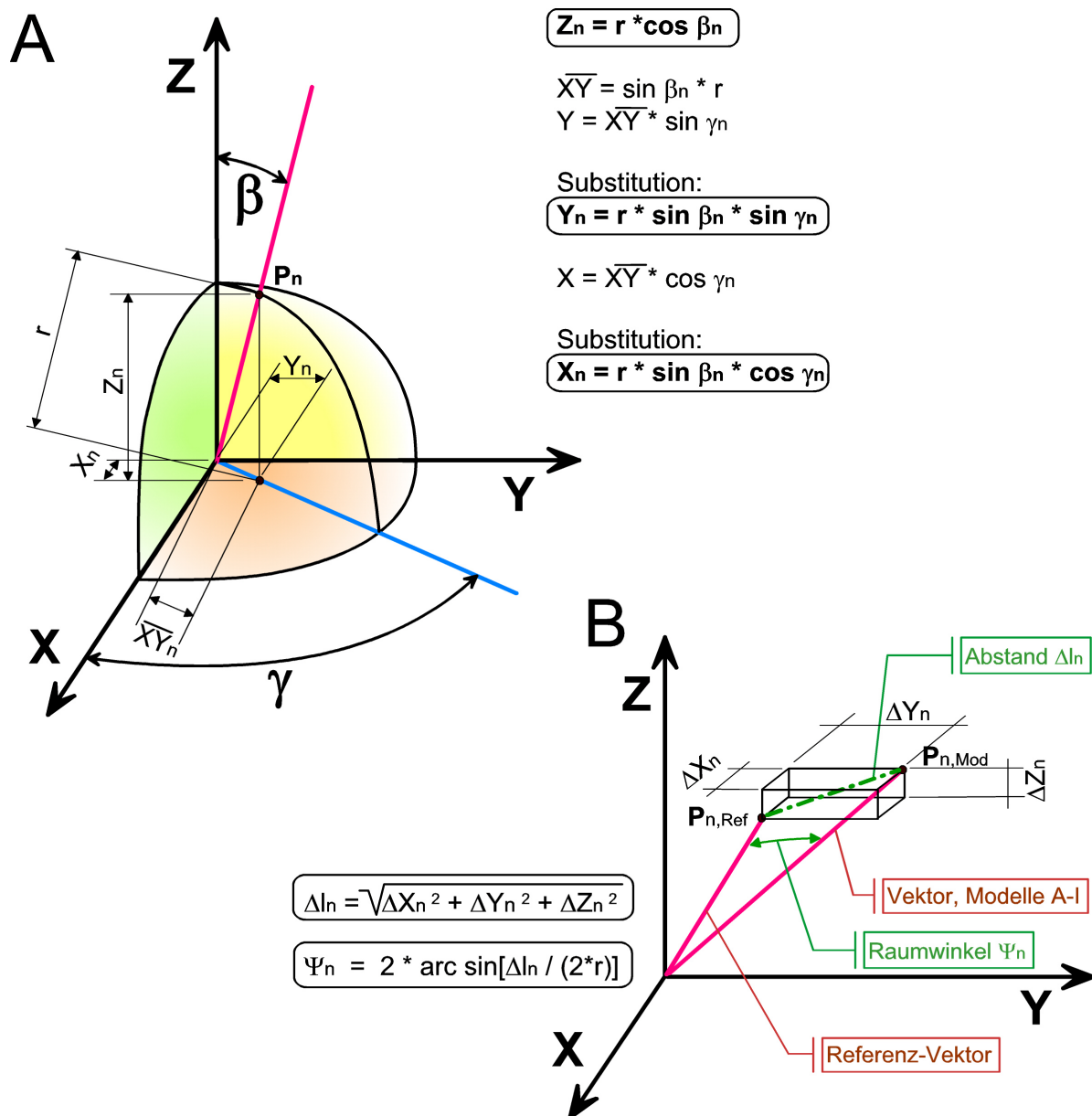


Abbildung 4.19: Trigonometrische Grundlagen der Fehlerberechnung. **A:** Da sich die Fehlerberechnung auf die Eulerwinkelbestimmung mit der Methode *angular reconstitution* im Imagic-5 Programmpaket bezog (Abschn. 2.3.3), liegt dieser die Eulerwinkeldefinition des Imagic-5 Programms zugrunde (vgl. Abschn. 2.3, Abb. 2.5). Die Fehlerberechnung berücksichtigte die Abweichungen der Eulerwinkel β und γ , während der α Winkel vernachlässigbar war (siehe Text). Mit diesen beiden Winkeln sind der Breiten- und Längengrad der Projektionsrichtung des Projektionsbildes definiert. Unter Festlegung eines Radius r von 100 nm lässt sich aus diesen beiden Winkelangaben ein Punkt P_n auf einer Kugel mit einem Durchmesser von 200 nm bestimmen. Die Koordinaten des Punktes P_n wurden mit Hilfe dieser Festlegung des Radius in das kartesische Koordinatensystem unter Verwendung der angegebenen Formel umgerechnet. Da für jedes 3D Modell 406 2D Projektionen gebildet worden waren, denen neue Winkel zugeordnet wurden, bestand jedes Winkelset aus 406 β & γ Winkelpaaren und entsprechend wurden 406 Punkte $P_{n,Mod}$ mit $n = 1$ bis 406 berechnet, für jede Projektion ein Punkt. Analog wurde mit dem Referenzwinkelset verfahren, um die Punkte $P_{n,Ref}$ zu erhalten. **B:** Um die Abweichung zu bestimmen, die bei der neuen Winkelzuweisung mittels *angular reconstitution* entstanden ist, wurde der Abstand berechnet zwischen dem Punkt $P_{n,Ref}$ des Referenzsets und dem Punkt $P_{n,Mod}$ der entsprechenden Projektion des Modells. Letztlich wurde die Abweichung durch den Raumwinkel Ψ , aufgespannt durch die Vektoren der Punkte $P_{n,Mod}$ und $P_{n,Ref}$, angegeben.

Um die Darstellung zu normalisieren, wurden in einem zweiten Schritt die Werte von $P_{n,Mod(orig)}$ als Referenz behandelt und der Winkel Ψ wurde wiederum zwischen Modell

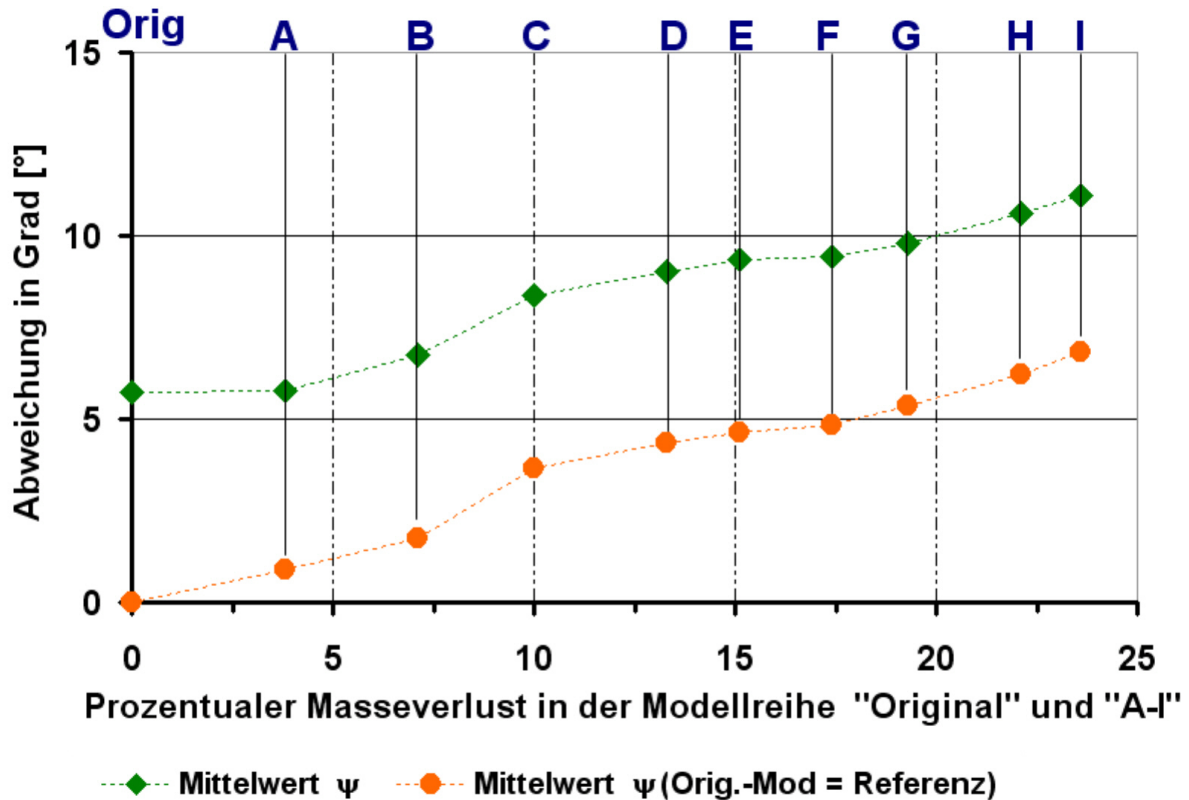


Abbildung 4.20: Mittlere Abweichung der Winkel Ψ in Grad in der Modellreihe. Der Winkel Ψ gibt für jedes Projektionsbild die Abweichung in der Projektionsrichtung an, die durch die Neuzuweisung der Eulerwinkel via *angular reconstitution* entstanden war. Da für jedes Modell 406 Projektionen gebildet worden waren, wurden ebenso 406 Winkel Ψ für jedes Modell erhalten. Zum einen diente zur Bestimmung von Ψ dasjenige Referenzwinkelset, das auch als Referenz im *angular reconstitution* verwendet worden war (Mittelwert Ψ , grüne Kurve). Danach weist das Originalmodell bereits eine mittlere Abweichung, die unter Einbeziehung aller 406 Winkel als Mittelwert von Ψ berechnet worden war, von ungefähr 6° auf. Diese Abweichung vergrößert sich mit Verlust an Masse des Modells, und das kleinste Modell I weist eine mittlere Abweichung von über 10° auf. Da der Mittelwert von Ψ des Originalmodells die Limitierung des Programms zeigt, Stichprobe und Referenz waren hier die gleichen, wurde in einem zweiten Verfahren eine Art der Normalisierung vorgenommen. Dazu wurde die Abweichung Ψ mit dem neu berechneten Winkelset des Originalmodells als Referenz für alle Modelle neu berechnet. Anschließend wurden wiederum die Mittelwerte von Ψ für alle Modelle bestimmt. Für das Originalmodell ergibt sich so eine Abweichung von 0° , da $P_{n,Ref} = P_{n,Orig}$. In dieser Darstellung (Mittelwert Ψ (Orig.-Mod = Referenz, orange Kurve) wird auch deutlicher, dass es sehr wohl einen Unterschied der Abweichung zwischen Originalmodell und Modell A (3,8% Masseverlust) gibt.

und Referenz für alle Modelle ausgerechnet (Abb. 4.20, orange Kurve). So ergab sich für das Originalmodell eine mittlere Abweichung von 0° . Dieses normalisierende Berechnungsverfahren wurde auch in der weiteren Analyse verwendet und ist dann explizit

angegeben.

Häufigkeitsverteilung der Werte

Da bei Betrachtung der Einzelwerte von Ψ auffällig war, dass ein überwiegender Teil in ähnlicher Größenordnung lag und nur ein kleiner Teil erhebliche Abweichung in Grad zeigte, wurde die Häufigkeit des Auftretens einer Winkelabweichung in einer bestimmten Größenordnung näher untersucht. Um die Häufigkeit bestimmter Werte zu ermitteln, wurden zweierlei Klasseneinteilungen vorgenommen. Zum einen wurden die Werte linear in 10° -Schritten klassifiziert, und zum anderen wurde eine logarithmische Klasseneinteilung gewählt. Die Winkelabweichung von 0° bildete in diesem letzteren Fall die erste Klasse und die weiteren Klassen umfassten folgende Winkelgrößen für Ψ :

Klasse	Winkelgrößen für Ψ in Grad [°]
Klasse 0	$\Psi \leq 0^\circ$
Klasse 0, 1	$0^\circ < \Psi \leq 0, 1^\circ$
Klasse 1, 0	$0, 1^\circ < \Psi \leq 1, 0^\circ$
Klasse 10, 0	$1, 0^\circ < \Psi \leq 10, 0^\circ$
Klasse 100	$10, 0^\circ < \Psi \leq 100, 0^\circ$
Klasse 1000	$100, 0^\circ < \Psi \leq 180, 0^\circ$

Tabelle 4.3: Logarithmische Klasseneinteilung für die Häufigkeitsverteilung von Ψ

Diese Analyse der Werte für Ψ wurde sowohl vorgenommen, für die Abweichung von Modell zu Referenz ($P_{n,Mod}$ gegen $P_{n,Ref}$, Abb. 4.21), als auch für die normalisierten Werte, die sich im Falle der Umwidmung von $P_{n,Mod(orig)}$ zum Referenzset ergaben (Abb. 4.22). In Abb. 4.21 ist mit den schwarzen Balken die Häufigkeitsverteilung für das Originalmodell aufgeführt. Damit werden die Abweichungen in Grad, die durch die Methode bedingt sind, widergespiegelt ('Nulltest'). Die Abweichungen liegen demnach zu 96% $\leq 10^\circ$, wobei fast 20% nur eine Abweichung unter $\leq 1^\circ$ aufweisen. Daneben gibt es im Nulltest einige 'Ausreißer', die Abweichungen zwischen 150° und 180° aufweisen. Da nicht gleichzeitig der ansonsten nicht in die Bewertung eingennommene α -Winkel ebensolche Abweichungen zeigte, sind dies echte Fehler und nicht auf die Winkelzuweisungen für das Spiegelbild zurückzuführen.

Im Vergleich dazu ist der Anstieg der Häufigkeit für das Modell I für $10^\circ < \Psi \leq 20^\circ$ auffällig. Scheinbar ist dieser Anstieg und die Reduktion bei Ψ -Werten $\leq 10^\circ$ der erheblich reduzierten Domäne geschuldet. Der Effekt der Reduktion der Domäne wird in der normalisierten Häufigkeitsverteilung in Abb. 4.22 deutlicher. Zunächst ist gut erkennbar, dass ein großer Teil der Winkelzuweisungen identisch ist mit den Winkelzuweisungen, wie sie im 'Nulltest' getroffen wurden. Noch für 76% der Projektionen des

kleinsten Modells wurde die Winkelzuweisung genau so getroffen wie für das Originalmodell (Abb. 4.22), logarithm. Klasseneinteilung). Die Winkelabweichungen nehmen bei

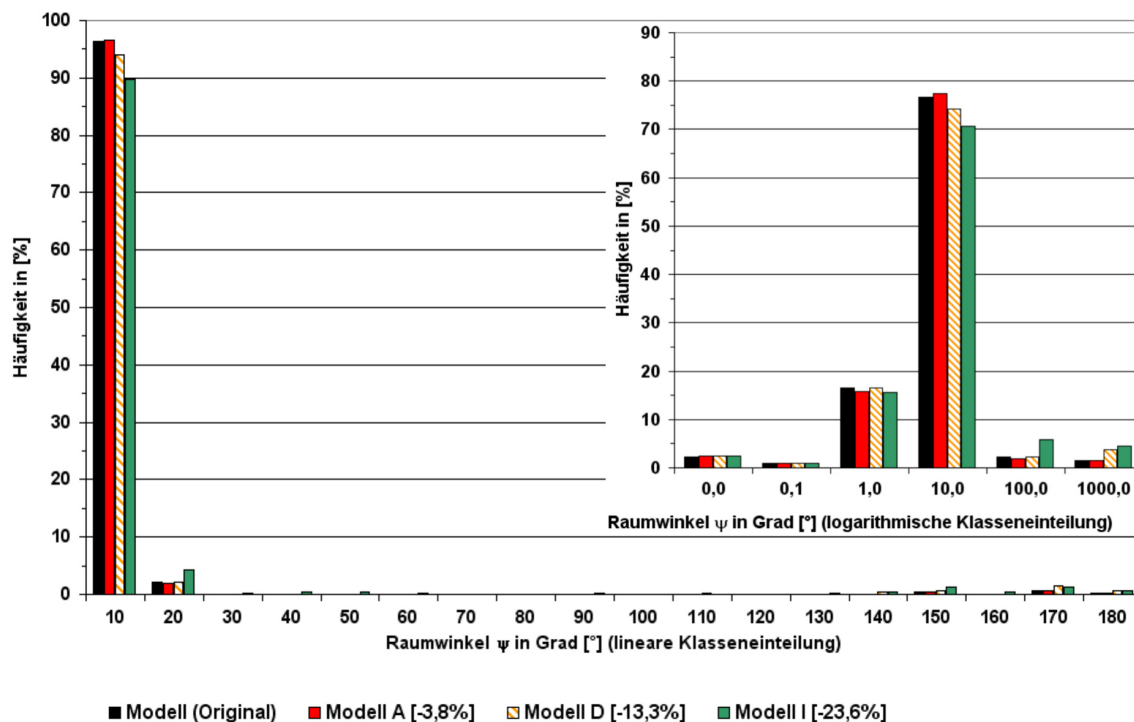


Abbildung 4.21: Häufigkeitsverteilung der Winkelabweichungen nach der *angular reconstitution*-Methodik. Das Ausmaß der Abweichung in der Zuweisung der Projektionsrichtung mittels *angular reconstitution* wurde mit dem Raumwinkel Ψ bestimmt (Abschn. 4.4.1 & Abb. 4.19). Für die Berechnung von Ψ in der hier dargestellten Häufigkeitsanalyse wurden als Referenzwerte ($P_{n,Ref}$) diejenigen verwendet, die sich aus dem tatsächlichen Referenzwinkelset, das im *angular reconstitution* eingesetzt wurde, ergaben. Die erhaltenen Werte für den Winkel Ψ wurden anschließend ihrer Größe entsprechend in Gruppen eingeteilt. Die ermittelte Anzahl an Werten, die in eine bestimmte Klasse fielen, wurde in Prozent umgerechnet und auf der y-Achse aufgetragen. Dies ist für die Modelle Original (schwarz), Modell A (rot), Modell D (gelb) und Modell I (grün) gezeigt. Zur Klassenbildung wurde zum einen eine lineare Klasseneinteilung in 10° -Schritten gewählt, zum anderen eine logarithmische Einteilung (eingefügtes Diagramm).

Lineare Klasseneinteilung: (Klasse **10**) $\Psi \leq 10^\circ$, (Klasse **20**) $10^\circ < \Psi \leq 20^\circ$, (Klasse **30**) $20^\circ < \Psi \leq 30^\circ$, , (Klasse **170**) $160^\circ < \Psi \leq 170^\circ$, (Klasse **180**) $170^\circ < \Psi \leq 180^\circ$.

Logarithmische Klasseneinteilung: (Klasse **0**) $\Psi \leq 0^\circ$, (Klasse **0,1**) $0^\circ < \Psi \leq 0,1^\circ$, (Klasse **1,0**) $0,1^\circ < \Psi \leq 1,0^\circ$, (Klasse **10,0**) $1,0^\circ < \Psi \leq 10,0^\circ$, (Klasse **100**) $10,0^\circ < \Psi \leq 100,0^\circ$, (Klasse **1000**) $100,0^\circ < \Psi \leq 180,0^\circ$.

Zum weitaus überwiegenden Teil ($\geq 90\%$) beträgt demnach die Abweichung in der Richtungsbestimmung für die Projektionsansichten $\leq 10^\circ$ für alle aufgezeigten Modelle, Orig. & A,D,I. Die Ψ -Werte fallen für das Modell I höher aus (vgl. Abb. 4.20, was sich vor allem in einem Anstieg der Häufigkeit bei Werten $10^\circ < \Psi \leq 20^\circ$ und bei groben Winkelabweichungen über 130° bemerkbar macht. Letzterer Effekt, der Anstieg bei groben Winkelabweichungen, ist auch bei Modell D zu beobachten.

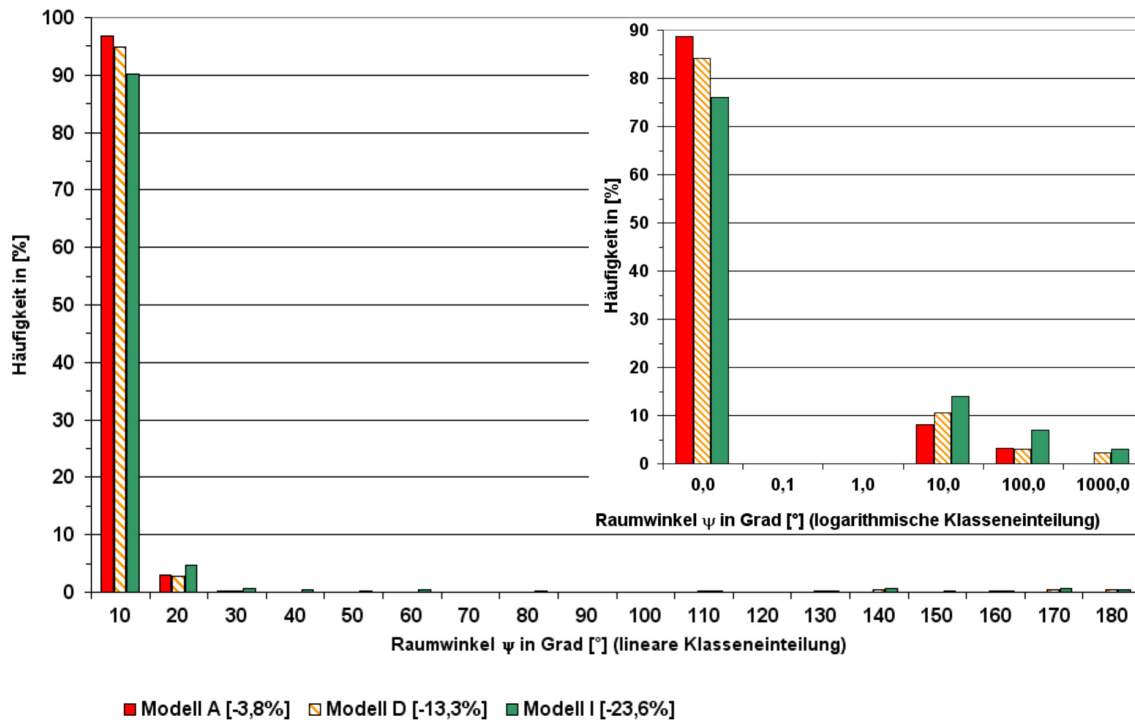


Abbildung 4.22: Häufigkeitsverteilung der Winkelabweichungen nach Normalisierung. Das Ausmaß der Abweichung in der Zuweisung der Projektionsrichtung mittels *angular reconstitution* wurde mit dem Raumwinkel Ψ bestimmt (Abschn. 4.4.1 & Abb. 4.19). Für die Berechnungen von Ψ in der hier dargestellten Häufigkeitsanalyse wurden als Referenzwerte ($P_{n,Ref}$) die Werte gesetzt, die dem Originalmodell im *angular reconstitution* zugewiesen worden waren (siehe Text: 'Nulltest'). Dadurch ergibt sich eine Normalisierung, die die Abweichungen des Originalmodells auf 0 setzt. Die erhaltenen Werte für den Winkel Ψ wurden anschließend ihrer Größe entsprechend in Gruppen eingeteilt. Die ermittelte Anzahl an Werten, die in eine bestimmte Klasse fielen, wurde in Prozent umgerechnet und auf der y-Achse aufgetragen. Dies ist für die Modelle A (rot), D (gelb) und I (grün) gezeigt. Zur Klassenbildung wurde zum einen eine lineare Klasseneinteilung in 10° -Schritten gewählt, zum anderen eine logarithmische Einteilung (eingefügtes Diagramm).

Lineare Klasseneinteilung: (Klasse **10**) $\Psi \leq 10^\circ$, (Klasse **20**) $10^\circ < \Psi \leq 20^\circ$, (Klasse **30**) $20^\circ < \Psi \leq 30^\circ$,, (Klasse **170**) $160^\circ < \Psi \leq 170^\circ$, (Klasse **180**) $170^\circ < \Psi \leq 180^\circ$.

Logarithmische Klasseneinteilung: (Klasse **0**) $\Psi \leq 0^\circ$, (Klasse **0,1**) $0^\circ < \Psi \leq 0,1^\circ$, (Klasse **1,0**) $0,1^\circ < \Psi \leq 1,0^\circ$, (Klasse **10,0**) $1,0^\circ < \Psi \leq 10,0^\circ$, (Klasse **100**) $10,0^\circ < \Psi \leq 100,0^\circ$, (Klasse **1000**) $100,0^\circ < \Psi \leq 180,0^\circ$.

Zum großen Teil werden die Winkelzuweisungen so getroffen, wie für das Originalmodell (Logarith. Einteilung, Klasse **0**, Modell A fast 90%, Modell I ca. 75%). Bei geringer Massereduktion (Modell A) liegen die entstandenen Abweichung überwiegend $\leq 20^\circ$. Bei größeren Masseverlusten nehmen auch die groben Winkelabweichungen und 'Ausreißer' zu (Logarithm. Klasseneinteilung Klasse **100** & **1000**).

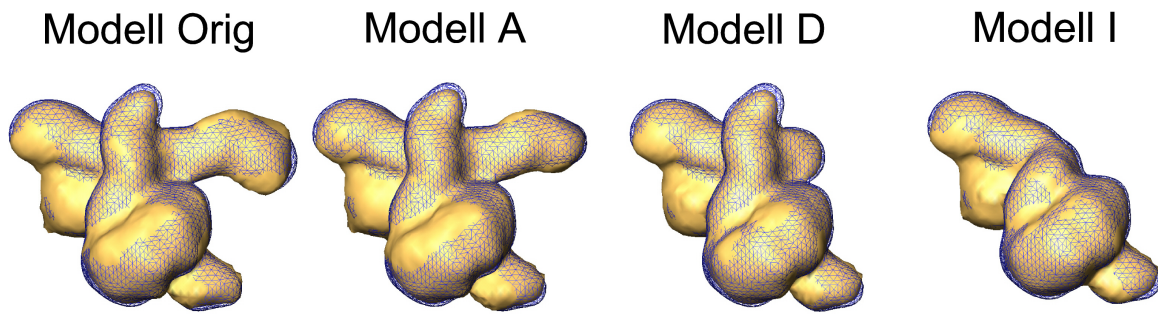


Abbildung 4.23: Rekonstruierte Modelle. Abgebildet sind das Originalmodell (Modell Orig) und die Modelle A, D und I, und zwar in ihren ursprünglichen Formen (blaues Netz) im Testdatensatz und als rekonstruierte Strukturen (gelb). Die Rekonstruktionen wurden mit den 2D Projektionen und ihren im *angular reconstitution* neu zugewiesenen Eulerwinkeln gemacht. Im Originalmodell ist die Domäne, die in den Modellen A-I editiert worden ist, nicht mehr ganz akkurat definiert. Desweiteren fällt die Kopfdomäne in den Modellen Orig, A und D gegenüber den ursprünglichen etwas zu klein aus. Auch die zweite Domäne, in den Abbildungen jeweils links am Hauptkörper, scheint nicht mehr ganz so sauber definiert. Der Test zeigt aber insgesamt, dass sich die Strukturen A-I trotz eines anders konfigurierten Modells als Referenzset mit *angular reconstitution* gut rekonstruieren lassen.

Reduktion der Domäne hauptsächlich im Bereich von $1, 0^\circ < \Psi \leq 20, 0^\circ$ zu. Es kommen allerdings weitere grobe Winkelabweichungen über 130° hinzu neben den 'Ausreißern', die schon für das Originalmodell gefunden wurden. Bei Modell D (13,3% Masseverlust) ist die höhere mittlere Abweichung ebenfalls zu einem guten Teil auf die Zunahme dieser groben Winkelabweichungen zurückzuführen.

Wie sich die gefundenen Fehler bei der Eulerwinkelzuweisung für die einzelnen 2D Projektionsbilder auf die Rekonstruktion der 3D Struktur auswirken, ist in Abb. 4.23 gezeigt. Dazu waren die Modelle Original, A, D und I in Imagic-5 aus den 406 zur Verfügung stehenden Projektionsbildern mit den im Test zugewiesenen Eulerwinkeln rekonstruiert worden. Trotz des Referenzsets mit vollständiger Domäne und den größeren Fehlern bei der Winkelzuweisung ließen sich auch die Modell D und I ohne grobe Fehler gut rekonstruieren.

Fehler mit MR-Alignment

Uneinheitlicher wird der auftretende Fehler bei einer vorherigen Ausrichtung der Projektionsbilder im MR-Alignment. Idealerweise hätte im Alignment der Rotationswinkel α nicht verändert werden dürfen, da im Referenzbildersset des MR-Alignments jeweils ein Projektionsbild des Originalmodells in exakt gleicher Richtung ausprojiziert vorlag wie auch in den Testmodellbildersets. Jedoch gilt dies nur für den Fall gleicher Referenz und Probe, wie festgestellt werden konnte (Originalmodell, Abb. 4.24). Bei den Testmodel-

len mit verkleinerter Domäne, wird der α -Winkel stets verändert. Um die Abweichungen des α -Winkels zu berechnen, wurden die Beträge der α -Winkel auf $\leq 180^\circ$ für alle 406 Projektionsbilder eines Modells umgerechnet und davon wurde der Mittelwert bestimmt (Abb. 4.24, blaue Kurve). Die Abweichungen des α -Winkels sind uneinheitlich in Bezug auf den Masseverlust. So zeigen die Modelle A und B und I die größten Abweichung nach dem *Alignment*, während Modelle C und D die geringsten Abweichungen aufweisen.

Der gleiche Kurvenverlauf ergibt sich für den Winkel Ψ , für dessen Bestimmung die originalen Projektionswinkel als Referenz $P_{n,Ref}$ eingesetzt wurden. Auch wenn der α -Winkel bei dem Test mit *Alignment* im Prinzip nicht unberücksichtigt werden darf, so zeigten die Analysen doch aufgrund des bekannten Ausgangszustands, dass ganz offensichtlich die Projektionsrichtungsbestimmung für einen zunehmenden Teil an Projektionsbildern nach einem MR-*Alignment* fehlerbehafteter und uneinheitlicher wird.

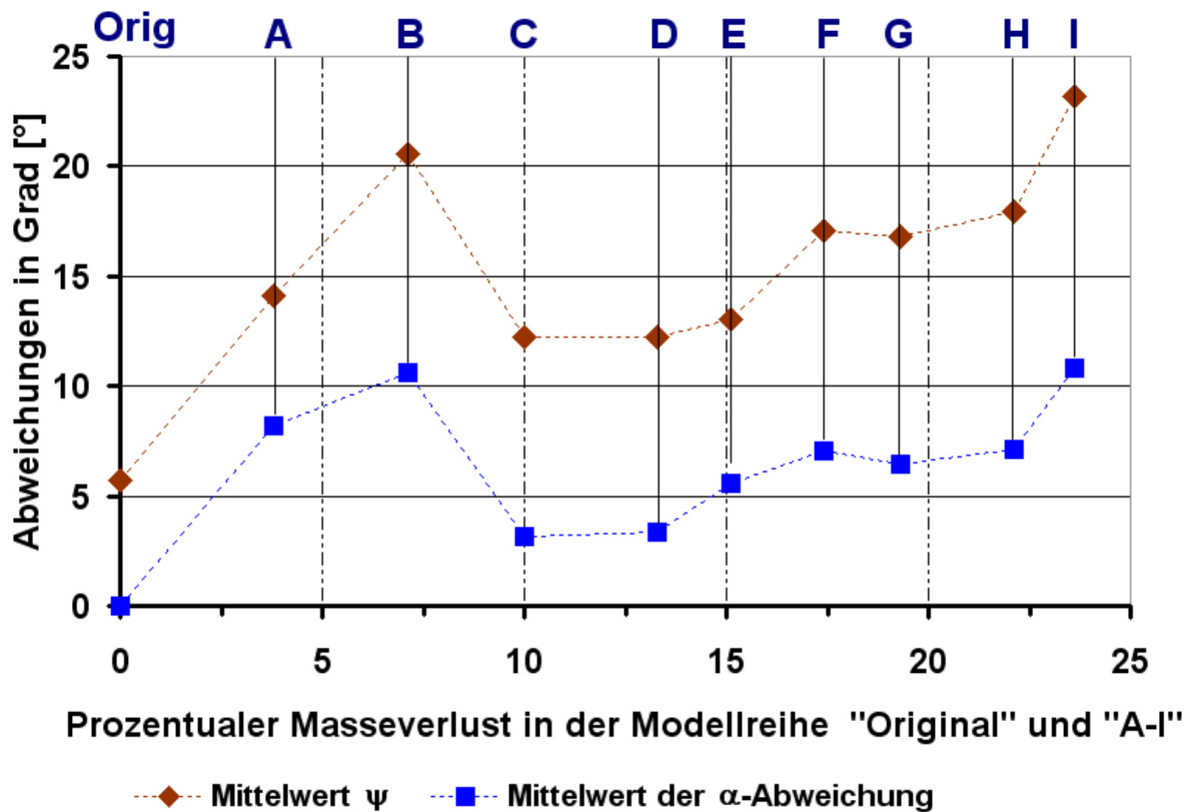


Abbildung 4.24: Mittlere Abweichungen des α -Winkels und des Ψ -Winkels nach MR-Alignment. Um die Auswirkungen eines MR-Alignments bei heterogenen Proben zu testen, wurden die Projektionsbilder der Modelle Orig. & A-I in einem MR-Alignment mit den Projektionsbildern des Originalmodells als Referenzen ausgerichtet. Um nachfolgend die Abweichungen des α -Winkels zu berechnen, wurden die α -Winkel in Bezug auf die Ausgangsposition so umgerechnet, dass Abweichungen $\leq 180^\circ$ für alle 406 Projektionsbilder eines Modells erhalten wurden. Davon wurde der Mittelwert bestimmt (blaue Kurve). Interessanterweise sind die resultierenden Abweichungen von α für die Modelle A und B höher, als für die Modelle C bis H. Der Einfluss der reduzierten Domäne auf die α -Abweichung ist somit uneinheitlich und nicht linear mit abnehmender Masse.

Den im MR-Alignment ausgerichteten Projektionsbildern wurden anschließend im *angular reconstitution* neue Eulerwinkel zugewiesen wiederum mit dem Originalmodell als Referenzwinkelset. Anschließend wurde Ψ bestimmt. Der berechnete Mittelwert von Ψ ändert sich in der Modellreihe in gleicher Weise wie die α -Werte.

FSC-Kurven der rekonstruierten 3D Modelle

Um die Bedeutung dieser gefundenen Eulerwinkelabweichungen für die 3D Rekonstruktionen abzuschätzen, wurden die FSC-Kurven aller berechneten 3D Strukturen bestimmt. Die FSC-Kurven von jeweils 3 Modellen sind für die Datensätze ohne und mit vorherigem MR-*Alignment* in Abb. 4.25 gezeigt. Es konnte keine Wirkung auf die Auflösungsgrenze nach dem 3σ oder $FSC_{0,5}$ Kriterium festgestellt werden. Die FSC-Kurven nehmen alle einen sehr ähnlichen Verlauf. Interessanterweise tritt bei 3 der FSC-Kurven

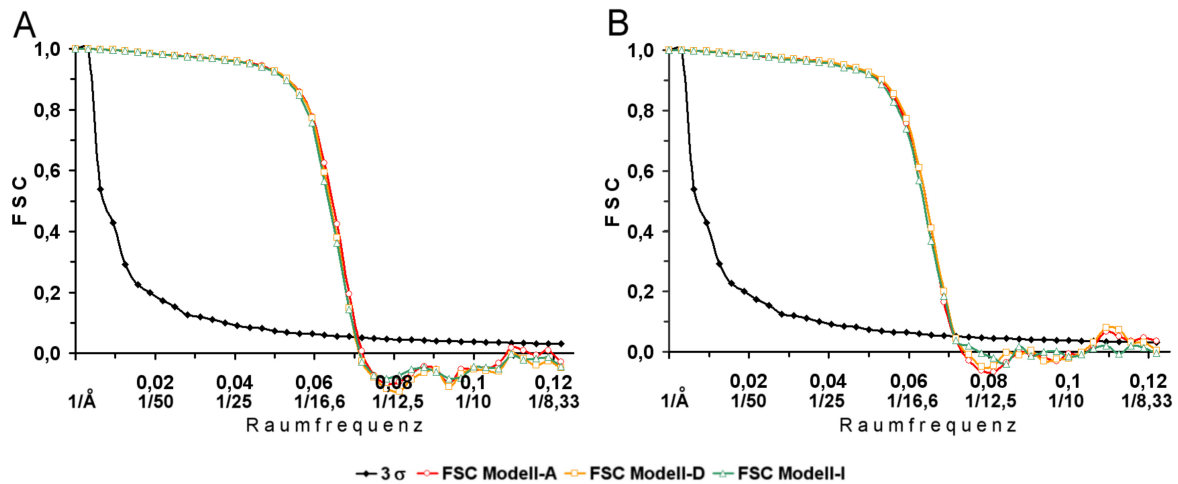


Abbildung 4.25: FSC-Funktionen der berechneten 3D Strukturen der Testdatenreihe. Nach der Berechnung der Eulerwinkel für die 2D Projektionsbilder der Modelle A - I mittels *angular reconstitution* wurden mit diesen Projektionsbildern neue 3D Strukturen rekonstruiert und ihre jeweilige Fourierschalenkorrelationsfunktionen bestimmt.

(A) FSC-Funktionen der Modelle A ($-3,8\%$, rot), D ($-13,3\%$, gelb) und I (grün, $-23,6\%$). Die Modelle

(B) FSC-Funktionen der Modelle A ($-3,8\%$, rot), D ($-13,3\%$, gelb) und I ($-23,6\%$, grün). Bei diesem Testverfahren war vor der Winkelbestimmung der 2D Projektionsbilder ein MR-*Alignment* erfolgt.

Die Reduktion der Domäne nimmt keinen Einfluß auf die Bestimmung der Auflösung nach dem 3σ - oder $FSC_{0,5}$ -Kriterium. Der Verlauf der Kurven ist insgesamt sehr ähnlich. (B) Auffallend ist im hohen Frequenzbereich der Anstieg der FSC-Kurven von Modell A und D über die 3σ -Kurve hinaus und auch der verzögerte Abfall der FSC-Kurve des Modells I im Bereich des Schnittpunkts mit der 3σ -Kurve.

aus den Datensätzen mit MR-*Alignment* ein erneuter leichter Anstieg im hohen Frequenzbereich über die 3σ -Kurve hinaus auf. Dies ist ein Effekt, der in verstärkter Form bei den 3D Rekonstruktion II-V des U1 snRNP beobachtet worden ist (Abschn. 4.3, Abb. 4.14).

Zusammenfassung der Testdatenauswertung

Die Genauigkeit der 3D Struktur Berechnungen bei zugrunde liegender Heterogenität

der Probe ist, wie die Ergebnisse mit den Testdaten zeigen, tatsächlich limitiert. Die zwei Datenprozessierungsschritte, deren Limitierungen mit den Testdaten einzeln überprüft wurden, waren zum einen die zur Rekonstruktion verwendete *angular reconstitution*-Methodik und zum anderen das *MR-Alignment*. Die Auswertung zeigte, dass der Fehler im *angular reconstitution* umso größer wird, je größer die Massendifferenz zwischen zu rekonstruierender Struktur und Referenzstruktur ist. Insgesamt reagiert diese Methode aber noch ziemlich robust trotz der Konfigurationsunterschiede.

Im Gegensatz dazu zeigte die Fehlerauswertung, dass offensichtlich das Nadelöhr der Strukturbestimmung das *MR-Alignment* ist. Bei der Rotationsausrichtung ähnlicher Einzelbilder, konnten diese nicht mehr mit ihrem Strukturhauptkörper akkurat zur Deckung gebracht werden. Dabei gab es keinen einfachen Zusammenhang zwischen Höhe der Abweichung und dem Massenverlust des Testdatensatzes gegenüber dem Referenzset. Letztlich bedeutet dies, dass Konfigurationsunterschiede nicht angemessen im *ML-Alignment* berücksichtigt werden können. Dies wirkt sich auf die Qualität der rekonstruierten 3D Strukturen aus, wie die FSC-Funktionen zeigten [185]. Es führt zu einem erneuten Anstieg der FSC-Kurven der berechneten 3D Strukturen im hohen Frequenzbereich, ein Effekt, der auch bei den U1 snRNP Strukturen II-IV beobachtet wurde.

Kapitel 5

Diskussion

5.1 Heterogenität des Hefe U1 snRNP Komplexes

Das U1 snRNP aus *S. cerevisiae* war für die elektronenmikroskopische 3D Strukturbestimmung aus totalem Hefeextrakt über die Tandem-Affinitätsaufreinigung [195] mit dem TAP-tag am Snu71 Protein aufgereinigt worden (Kap. 3, Abschn. 3.2.1). Abschließend wurde der so isolierte U1 snRNP auf einem 10 – 30% Glycerolgradienten zentrifugiert, wobei der Partikel für die EM-Präparation gleichzeitig mit der GraFix-Methode mit Glutaraldehyd fixiert wurde [119].

Die Methode, das U1 snRNP über den TAP-tag aufzureinigen, wurde in [205] beschrieben. In der Arbeit und in [64] konnte gezeigt werden, dass auch das Protein Luc7 eine Komponente des U1 snRNP aus *S. cerevisiae* ist. Mit den zuvor zur Isolierung des U1 snRNP aus *S. cerevisiae* verwendeten Aufreinigungstechniken mittels Immunoaffinitätschromatographie mit Antikörpern, gerichtet gegen die *m₃G* Kappe und in Kombination entweder mit einer direkt anschließenden Glycerolgradientenzentrifugation [59], oder mit einer weiteren Affinitätschromatographie unter Verwendung eines His-getaggten U1 snRNP Proteins [169], [79], war dieses Protein, Luc7, noch nicht entdeckt worden. Aufgrund dessen war erwartet worden, mit der beschriebenen TAP-Methode ein vollständig assembliertes U1 snRNP für die Strukturbestimmung zu erhalten.

5.1.1 Modifikationen der Aufreinigung

Bei elektronenmikroskopischer Begutachtung des U1 snRNP fielen schon während der Datenprozessierung auf 2D Ebene mit den Methoden, wie sie unter Kap. 2, Abschn. 2.2 beschrieben sind, Merkmale struktureller Heterogenität auf. Indizien dieser Heterogenität waren unter anderem die unsauberen Kontrastmittelsäume und Bereiche, die nur mit zusätzlicher Masse erklärbar waren (Abb. 4.2 und Abb. 4.3). Daher wurden ver-

schiedene Modifikationen im Prozedere der U1 snRNP Aufreinigung eingeführt, um die strukturelle Heterogenität einzuschränken. Zu bedenken ist, dass nicht von jeder dieser Modifikationen in dieser Arbeit 2D EM-Bilder des entsprechend isolierten U1 snRNP gezeigt wurden, sondern die Auswahl der EM-Bilder wurde auf die signifikanten Fälle beschränkt.

Herstellung des totalen Hefezelleextrakts

In Aufreinigungen des *S. cerevisiae* Tri-snRNP aus totalem Hefezelleextrakt war aufgefallen, dass die Snu66p Bande in der Proteingelanalyse nach erfolgter Aufreinigung verloren geht, wenn die Zellen unter Druck mit der French-Press aufgebrochen wurden. Wurden die Hefezellen dagegen mit der Zelmühle (RM100, Retsch, D) zermahlen, wurde dieser Effekt nicht beobachtet [92]. Für die TAP-tag Aufreinigung des U1 snRNP war ebenfalls für den Zellaufschluß stets die French-Press, wie in [205] auch, verwendet worden. Zwar konnte in der genannten Publikation gezeigt werden, dass sich alle 10 spezifischen *S. cerevisiae* Proteine nach erfolgter U1 snRNP Isolierung nachweisen ließen, doch um auszuschließen, dass die Probleme der beobachteten strukturellen Heterogenität des U1 snRNP aufgrund eines unsanften Zellaufschlusses und einer teilweisen Zerrüttung des Komplexes zustande kamen, wurde nachfolgend auch für den Zellaufschluß in der U1 snRNP Aufreinigung ausnahmslos die Zelmühle verwendet. In [258] wird sogar vorgeschlagen, die Zellen für den Zellaufschluss nicht in flüssigem Stickstoff zu frieren, bevor sie zermahlen werden, sondern anstelle die (Glas-)Perlenmühle zu verwenden und ferner auf die Dialyse zu verzichten, um durch den Pufferaustausch möglicherweise induzierte Rearrangements oder Zerfall der spleißosomalen Komplexe zu verhindern.

Glycerolgradientenzentrifugation 'GraFix'

Die Glycerolgradientenzentrifugation ist mit Fixativzugabe (GraFix) eine Methode, um die Qualität der nachfolgenden EM-Probenpräparation zu verbessern [119]. Während Zerfallsprodukte des Komplexes und noch vorhandene niedermolekulare Verunreinigungen oben im Gradienten verbleiben, sedimentieren möglicherweise während des Aufreinigungsprozederes gebildete große Aggregate an den Boden des Zentrifugenröhrchens. Die steigende Konzentration an Glutaraldehyd im Glycerolgradienten fixiert den Komplex während des Zentrifugenlaufes langsam in seiner vorliegenden molekularen Zusammensetzung und schränkt dessen Konformationsänderungen zunehmend ein. Die Fixierung gewährleistet vor allem eine bessere Stabilität und Integrität des Komplexes während der Absorption an den Kohlefilm bei der notwendigen EM-Probenpräparation auf einen Kupfernetzträger (Kap. 3, Abschn. 3.3.3). So konnte gezeigt werden, dass der fixierte eIF3 Komplex aus *S. cerevisiae* im Gegensatz zum unfixierten nicht während der

Präparation zerfiel [98]. In [119] konnte darüber hinaus nachgewiesen werden, dass die EM-Bilder von fixierten Proben ein besseres SNR aufweisen, was sich positiv in der Datenprozessierung auswirkt.

Im Falle des U1 snRNP Komplexes war jedoch auffällig, dass sich die Projektionsansichten von fixierter und unfixierter Probe stark unterschieden (Kap. 4, Abb. 4.2, 4.3). Für die Quervernetzungsreaktion von Glutaraldehyd mit Proteinen sind zwar mehrere Mechanismen vorgeschlagen worden und möglich [153], die wichtigste Reaktion unter ihnen ist aber die Ausbildung von Schiff'schen Basen zwischen den Aldehydgruppen des Glutaraldehyds und den primären Amidgruppen des Proteins. Da Glutaraldehyd zwei Aldehydgruppen besitzt, kommt es zu Quervernetzungen. Durch die langsame Fixiermethode im Gradienten während der Probenseparatisierung im Zentrifugationslauf werden aber sehr wahrscheinlich keine intermolekularen Verbindungen neu geknüpft, da die Konzentrationen sowohl der Probe, als auch des Fixativs zu gering sind. Aufgrund dieser Überlegungen ist es wahrscheinlich, dass das U1 snRNP bei der zuerst verwendeten Präparation der unfixierten Probe zwar nicht wie im Falle des eIF3 Komplexes [98] vollständig, aber doch teilweise zerfiel (Abb. 4.3). Inwieweit bestehende intermolekulare Kontakte, die durch die Fixierung mit Glutaraldehyd konserviert werden, allerdings im Falle des U1 snRNP funktionell relevant sind, oder auf unspezifischen Wechselwirkungen beruhen, kann nicht eindeutig beantwortet werden.

Im übrigen sahen die 2D Ansichten der unfixierten Proben (Abb. 4.2) den veröffentlichten 2D Einzelbildern des über Immunoaffinitätsaufreinigung isolierten U1 snRNP aus *S. cerevisiae* sehr ähnlich [59]. Die EM-Präparation der Partikel erfolgte in [59] ebenfalls ohne vorherige Fixierung des Komplexes. Möglicherweise wäre also auch dort ein Unterschied der 2D Ansichten von fixierter zu unfixierter Probe aufgefallen.

Nochmals angemerkt sei, dass sich die Heterogenität der Probe nicht nur in dem Unterschied von fixiertem zu unfixiertem Komplex widerspiegelte, sondern auch innerhalb eines Datensatzes beobachtet wurde. Die Klassensummen negativ kontrastierter fixierter Proben zeigten nicht selten an einigen Stellen Unschärfe am Kontrastmittelsaum. Dies sind dann schlecht definierte Bereiche, deren Ursache flexible Domänen sein können (Kap. 4, Abb. 4.6).

Variation der Salzkonzentration im Glycerolgradienten

Auch die Salzkonzentration während einer Proteinaufreinigung hat Einfluß auf das Resultat. So konnte für den U5 snRNP gezeigt werden, dass die Salzkonzentration die Konformation des Komplexes veränderte [218]. Der unter Hochsalzbedingungen aufgereinigte U5 snRNP wies eine elongiertere Struktur auf, aufgrund einer aufgerichteten Domäne, die ansonsten unter herkömmlicher Salzbedingung geknickt war.

Ebenso war das U1 snRNP unter verschiedenen Salzbedingungen bei Verwendung der His-tag Affinitätschromatographie aufgereinigt worden [79]. Bei 700 *mM* KCl im Glycerolgradienten konnte Prp42p (Snu65p) nicht mehr nachgewiesen werden, ebenso wie das Protein Npl3p, das in der dort verwendeten U1 snRNP Aufreinigung auftrat, und das mittlerweile als SR-ähnliches Protein in Hefe charakterisiert worden ist [128]. Aufgrund dieser Hinweise war die Salzkonzentration im finalen Schritt der Aufreinigung, im Glycerolgradienten, auch bei Verwendung der TAP-tag Aufreinigungsmethode variiert worden (Abb. 5.1). Nach der Trennung der Proteine mittels SDS-PAGE, wurden die Proteine mit Silber gefärbt. Danach scheint eine Konzentration von 500 *mM* geeignet für die Aufreinigung des U1 snRNP zu sein. Die Nachteile der hohen Salzkonzentration im Glycerolgradienten wogen allerdings für die elektronenmikroskopische Probenpräparation schwer. Für eine gute negativ Kontrastierung der Komplexe mussten Waschschritte eingeführt werden, die die Konzentration des Partikels auf dem Kohlefilm wiederum stark reduzierten. Desweiteren zeigte sich, dass eine Kombination mit der GraFix-Methode keine Qualitätsverbesserung der Probe brachte. Die Fixierung der Probe kann dadurch, dass die Glycerolgradientenzentrifugation noch als Aufreinigungsschritt benötigt wird, nur sinnvoll nach der Fraktionierung erfolgen. Unter Anbetracht dieser weitreichenden Auswirkungen, wäre eine elegantere Methode, den Komplex beispielsweise mit Hochsalzwaschschritten schon während der Affinitätsaufreinigung zu behandeln. Dies erfordert eine erneute Testung der Parameter für die Isolierung des U1 snRNP Komplexes. Dann jedoch könnte die Glycerolgradientenzentrifugation als Schritt der Probenvorbereitung für die EM und nicht mehr als benötigter Aufreinigungsschritt der Probe dienen.

Vorfraktionierung des Hefezelleextrakts

Da im totalen Hefezelleextrakt auch TAP-tag versehenes Fusionsprotein (Snu71) vermutet wurde, dass nicht im U1 snRNP vorliegt, sondern entweder frei oder in Subkomplexen auftritt, wurde eine Vorfraktionierung des totalen Hefezelleextraktes vorgenommen. Bei den Vorexperimenten fiel in der Northern Blot Analyse (Abb. 4.4) auf, dass sich U1 RNA quasi über den gesamten Surespin-Gradienten streute, d.h. mit verschiedenen S-Werten, auch höheren S-Werten, als den erwarteten 18 *S*, sedimentierte. Dadurch, dass nur ein Teil der Fraktionen eingesetzt wurden, und zwar jene in denen 18 *S* U1 erwartet wurde, sollte die Möglichkeit via des TAP-tags auch andere Snu71p-haltige Komplexe aufzureinigen, eingeschränkt werden. Mit diesem zusätzlichen Aufreinigungsschritt wurden U1 snRNP Proben erhalten, die schließlich für die 3D Rekonstruktion der Struktur verwendet wurden. Das verbliebene Ausmaß an Heterogenität sollte mit den Möglichkeiten der EM-Datenverarbeitung ermittelt werden.

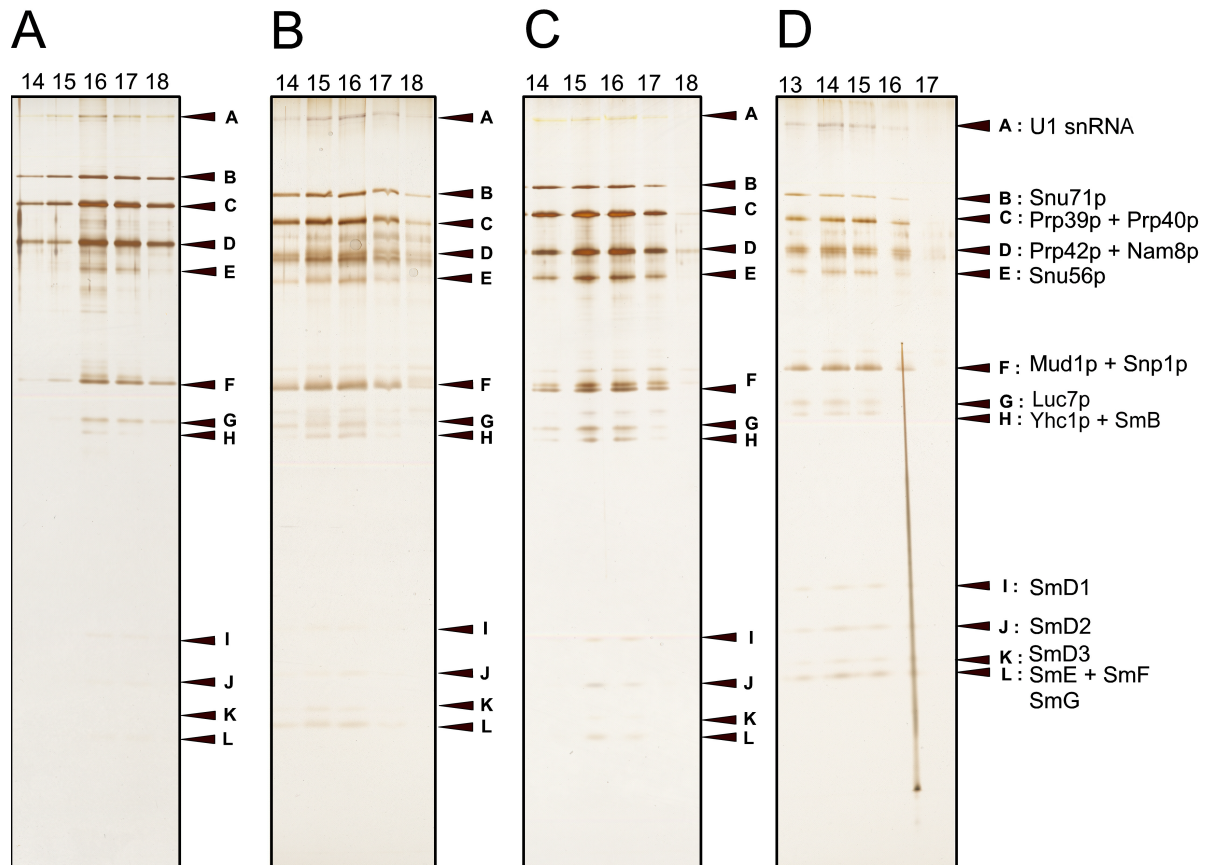


Abbildung 5.1: Abhängigkeit der Salzkonzentrationen in der Aufreinigung des U1 snRNP. Ohne Vorfraktionierung TAP-tag aufgereinigter U1 snRNP wurde auf 4 verschiedenen 10 – 30 % (*w/w*) Glycerolgradienten, die sich nur in ihrer Salzkonzentration unterschieden, ultrazentrifugiert. In (A) betrug die KCl Konzentration 100 *mM*, in (B) 300 *mM*, in (C) 500 *mM* und in (D) 700 *mM*. Aus den Gradientenfraktionen (Gesamtzahl 24, oben 1, unten 24) wurden die Proteine mit Aceton präzipitiert und danach mittels SDS-PAGE auf einem 12 % Polyacrylamid Gel analysiert. Die Gele wurden mit Silber gefärbt. Gezeigt sind jeweils die Ausschnitte mit den Fraktionen, in die das U1 snRNP sedimentiert ist. Mit ansteigender Salzkonzentration ist der U1 snRNP zunehmend in höheren Fraktionen erkennbar, was trotz manueller Ernte der Fraktionen auf eine geringer werdende Masse hindeutet.

5.2 3D Strukturen des Hefe U1 snRNP

Die Einzelmolekül-TEM ist eine geeignete Methode, um die Strukturen von makromolekularen spleißosomalen Komplexen in der Größenordnung von 240 *kDa* bis ungefähr 5 *MDa* zu bestimmen [248]. Von derart großen Komplexen sind röntgenkristallographische Strukturbestimmungen aufgrund der Notwendigkeit, zuvor Kristalle zu züchten, kaum möglich. Jüngst gelang es allerdings, eine röntgenkristallographische Struktur des humanen U1 snRNP zu erhalten [192].

Da nun für die Einzelmolekül-EM das Erfordernis, starre Kristalle zu bilden, entfällt, ist diese Methode geeignet, Strukturen unterschiedlicher Konformationen eines Komplexes zu bestimmen [56], [218], [141]. Aufgrund der gestiegenen Computer-Prozessorleistungen konnte dieser Möglichkeit in der Vergangenheit auch vermehrt Rechnung getragen werden, und zwar mit der Entwicklung neuer Datenverarbeitungsstrategien und Programme, die die nachträgliche Separatisierung verschiedener 3D Strukturen erlaubt [141]. Zunächst ist die Kombination von RCT-3D-Rekonstruktion mit dem *ML-Alignment* [237], [214], [218], (Kap. 2, Abschn. 2.3.2) zu nennen, wodurch bereits in der *de novo* Struktur-determination die Erkennung unterschiedlicher Konformationen oder Konfigurationen eines Komplexes möglich ist (Abb. 4.11). Dadurch, dass im anschließenden *Refinement* mehrere Strukturen im *supervised classification* [72] angeboten werden können, lassen sich prinzipiell aus einem heterogenen Datensatz mehrere gut aufgelöste Strukturen eines Komplexes erhalten, die den Komplex beispielsweise in verschiedenen Konformationen oder mit unterschiedlicher Ligandenbesetzungen zeigen.

Da bekanntermaßen häufig Flexibilitäten in spleißosomalen Komplexen auftreten, die zu nicht ausreichender Homogenität der Proben für hochaufgelöste 3D Strukturen führen [283], sollten für das U1 snRNP aus *S. cerevisiae* mit den einleitend beschriebenen Möglichkeiten der 3D EM-Strukturbestimmung die unterschiedlichen Strukturen aufgezeigt werden. Zur Verbesserung der Homogenität der Probe waren mit der Einführung der Vorfraktionierung die biochemischen Modifikationen des bestehenden Aufreinigungsprotokolls ausgereizt. Desweiteren schien die Fixiermethode im Gradienten nach dem 'Gra-Fix' Protokoll [119], die Integrität des Komplexes in Bezug auf eine Domäne, wie in Abb. 4.3, C gezeigt, zu gewährleisten. Die beobachteten verschwommenen Kontrastmittelsäume (Abb. 4.6, A) deuteten bewegliche Domänen an wie sie beispielsweise auch im humanen spleißosomalen BΔU-Komplex beobachtet wurden [22], [21]. Sie waren somit für spleißosomale Komplexe nicht ungewöhnlich, sondern vielmehr von biologischem Interesse.

Von dem U1 snRNP aus *S. cerevisiae* lagen im Gegensatz zum humanen U1 snRNP [249], [192] noch keine veröffentlichten Strukturdaten vor, abgesehen von den 2D Ansich-

ten in der Vergleichsstudie von Tri-snRNP und U1 snRNP in *S. cerevisiae* und Mensch [59]. Die letztgenannte Vergleichsstudie deutete bereits auf erhebliche strukturelle Unterschiede des U1 snRNP in den beiden Organismen hin.

***de novo* Rekonstruktion des U1 snRNP**

Die *de novo* Rekonstruktion der U1 snRNP 3D-Strukturen wurde mit der *Random Conical Tilt*-Methode [198] von negativ kontrastierten Proben ausgeführt. Dazu wurden Aufnahmen derselben Partikel unter einem Aufnahmewinkel von 0° und 45° benötigt (Abb. 4.9). Die Partikel in der zweiten Aufnahme sind aufgrund der vorherigen Strahl exposition schon ein wenig beschädigt. Mit diesem Wissen wurde die 0° Aufnahme stets zuerst gemacht, um einen qualitätvollen Einzelpartikel-Datensatz ohne unnötige Artefakte für die Einteilungen in Klassen (Kap. 2, Abschn. 2.2) zur Verfügung zu haben. Um für jede Klasse eine RCT-3D-Struktur zu berechnen, wurden dann die 45° Aufnahmen verwendet (Abb. 4.10). Zwar sind das Kontrastmittel und die Aufnahmen bei RT eine gewisse Qualitätseinschränkung, diese sind aber aus folgenden Gründen unumgänglich oder vertretbar: (1) Um das U1 snRNP während der Aufreinigung intakt zu halten oder auch Aggregationen zu vermeiden, wurde im gesamten Aufreinigungsprotokoll Glycerol verwendet. Dieses Glycerol verringert den Kontrast im Elektronenmikroskop bei nativer Probenbereitung für die EM. Kurz beschrieben, bei nativer Probenbereitung werden die Strukturen durch schnelles Einfrieren mit flüssigem Ethan in einem dünnen amorphen Eisfilm konserviert [140]. Das Glycerol muss dafür aus dem oben genannten Grund vorher entfernt werden, was einen zusätzlichen Schritt der Umpufferung erfordert, bei dem es zu Materialverlust und Gefährdung der Integrität der Struktur kommt. (2) Die Aufnahmen bei gekipptem Probenhalter bei der dann erforderlichen Stickstoffkühlung der Probe sind aufgrund der Instabilität des Probenhalters extrem störanfällig. (3) Im Falle des U1 snRNP war eine reine native Präparationsmethode, bei der der Eisfilm direkt über der Kohlelochfolie aufliegt, aufgrund der geringen Konzentration sowieso nicht möglich, so dass eine Absorptionszeit für das U1 snRNP von 3 h an einen kontinuierlichen Kohlefilm unumgänglich blieb.

Da die RCT-Rekonstruktionen lediglich der Erstellung initialer 3D Strukturen dienen, die im folgenden *Refinement* verbessert wurden, waren die Artefakte der negativ Kontrastierung und der Aufnahmen bei RT tolerabel. Tatsächlich wurden bisher aufgrund dieser Limitierungen nur wenige spleißosomale Komplexe überhaupt mit der nativen Präparationstechnik untersucht [249], [218], [10]. Bei dem Tri-snRNP in [218] wurde dazu die Probe über Nacht mit Glutaraldehyd fixiert, bevor der Tri-snRNP in Glycerol-freien Puffer überführt wurde. Eine Fixierung über Nacht hatte im Falle des U1 snRNP nicht funktioniert, da möglicherweise die Probe an den Wänden des Gefäßes absorbierte oder

der Komplex aggregierte. Da in [76], [217] gezeigt werden konnte, dass auch mit negativ kontrastierten Proben unter ansonsten Kryobedingungen gute bis hohe Auflösungen erreicht werden können, wurde für das Struktur-*Refinement* ebenfalls ein Datensatz aus kryo-negativ-kontrastierten Proben verwendet. Die Aufnahmen entstanden dann unter Stickstoffkühlung.

Um initiale 3D Strukturen für ein *Refinement*-Prozedere zu erhalten, wurden die 200 RCT-Strukturen in einem Maximum-Likelihood-Alignment [218] ausgerichtet und anschließend PCA-basiert klassifiziert mit durchschnittlich 8 Mitgliedern pro Klasse. Auf diese Weise wurden 25 3D-Strukturen erhalten. Bei Begutachtung dieser groben initialen Strukturen war bereits deutlich, dass der Datensatz aus mehreren sehr unterschiedlichen Strukturen bestand. Es wurden daher 5 unterschiedliche Strukturen danach ausgewählt, dass sie das Spektrum an den im Datensatz auftretenden Strukturen möglichst repräsentativ erfassten (Abb. 4.11).

***Refinement* des U1 snRNP**

Im '*Refinement*' wurden die Strukturen auf verschiedene Merkmale getestet. Zum einen fiel ein Strukturfortsatz auf (Abb.: 4.12), der in den Strukturen II-V untersucht wurde. Danach existierte die 3D-II in der U1 snRNP Probe sowohl mit, als auch ohne Fortsatz. Im Gegensatz dazu konnte für die übrigen Strukturen keine Aussage über den Strukturfortsatz getroffen werden.

Um die erhaltenen Strukturen besser beurteilen zu können, wurden zunächst die ungefähren Massen in Imagic-5 bestimmt. Die Methode ist zwar ungeeignet für eine absolute Massenbestimmung, sie lässt aber Aussagen über die Relationen der Strukturen untereinander zu. Die Massen variieren demnach stark und fallen tendenziell höher aus, als die in der Literatur beschriebene [79]. Damit wird zum einen die angenommene 1:1 Stöchiometrie aller Proteine im Komplex zumindest in Frage gestellt. Es kann aber auch nicht ausgeschlossen werden, dass noch weitere Proteine am U1 snRNP in substöchiometrischen Maßstab haften, die ebenfalls am Prozess der 5' Ss-Erkennung mitwirken oder an der Regulation der mRNA-Prozessierung insgesamt beteiligt sind.

In [258] beispielsweise wurde berichtet, dass mit der Snu71-TAP-tag Aufreinigung auch Sto1p, das Homolog des humanen CBP80 Proteins, das die m^7G -Kappe der mRNA bindet und Bestandteil des CBC ist, bei der Aufreinigung des U1 snRNP aus *S. cerevisiae* gefunden wurde. Ähnliche Beobachtungen wurden auch im Hause gemacht, indem die kleine Untereinheit Cbc2/Mud13p des CBC gefunden wurde (s. Appendix, Tab. 1). Diese Ergebnisse waren allerdings scheinbar abhängig von der in der Aufreinigung verwendeten Salzkonzentration. Auch abhängig von der Salzkonzentration war die Bestimmung des Proteins, Gcr3, in der massenspektroskopischen Analyse (s. Appendix, Tab. 1). Gcr3

wurde als Protein beschrieben, das in *S. cerevisiae* an der Expression von glykolytischen Genen beteiligt ist [268].

In den früheren Aufreinigungen über einen His-tag wurde auch Npl3p mit den Proteinen des U1 snRNP isoliert [79]. Es wurde in der Literatur als ein SR-ähnliches Protein beschrieben, das neben anderen Funktionen in *S. cerevisiae* den Spleißprozess von einem Teil der prä-mRNAs erleichtert und in den frühen Erkennungs- und Assemblierungsprozessen mitwirkt [128], [202]. Auch wenn es keine direkten Hinweise auf Npl3p in diesen Proben gab, kann es nach der beobachteten Heterogenität auch nicht ausgeschlossen werden, dass es substöchiometrisch an das U1 snRNP gebunden hat. Gleiches gilt für Sto1, Cbc2 und Gcr3. Eine Massenspektroskopische Analyse genau dieser Aufreinigung könnte Aufschluss darüber geben, ob diese an einem Teil der U1 snRNP gebunden haben könnten.

Für die an 3D Struktur III muss nach den hier vorliegenden Daten angenommen werden, dass die nur über einen schmalen Steg mit dem Hauptkörper der Struktur verbundene Domäne ein Artefakt des *Refinements* ist (Abb. 4.17).

Um auszuschließen, dass ein Teil der Probleme im Struktur-*Refinement* präferentieller Bindung der Partikel an den Kohlefilm durch die verwendete Präparationsmethode mit Uranylformiat, wie unter Kap. 3, Abschn. 3.3.3 beschrieben, geschuldet ist, wurden die Projektionsansichten jeder 3D Struktur ihren Eulerwinkeln entsprechend auf der Eulerkugel angeordnet. Dazu waren die Subdatensätze, die aus dem *supervised classification* erhalten worden waren, PCA-basiert klassifiziert worden. Den so gebildeten Klassensummen waren mittels *angular reconstitution* die Eulerwinkel zugewiesen worden. Das Ergebnis zeigte eine gleichmäßige Projektionswinkelverteilung für alle 3D Strukturen.

Die Fourier-Schalen-Korrelations-Kurven der 3D Strukturen I-V wurden errechnet, um Aussagen über die Qualität der erhaltenen 3D Strukturen treffen zu können. Auch kann die Auflösungsgrenze nach verschiedenen Kriterien mit Hilfe der Fourier-Schalen-Korrelationsfunktionen angegeben werden (Kap. 2, Abschn. 2.3.6). Die 3D-I Struktur zeigt einen eindeutigen Verlauf der FSC-Kurve. Nach dem 3σ -Kriterium hat sie eine Auflösung von 25 Å, nach dem $FSC_{0,5}$ -Kriterium eine Auflösung von 32 Å. Da die Kurve ansonsten keine Auffälligkeiten aufweist, und die Datenverarbeitung zeigte, dass stets eine kleine, aber in allen verwendeten Datensätzen konstante Anzahl von Projektionsbildern die 3D Struktur repräsentierte (6,8 % nach Tab. 4.1), scheint dies eine valide stabile Struktur im Datensatz zu sein. Die Masse ist allerdings circa doppelt so hoch wie erwartet [79]. Danach wäre abzuklären, ob es sich um einen Dimer oder eine Teildimerisierung der Proteine im U1 snRNP handeln könnte. Dagegen sind die Strukturen II-V anders zu beurteilen, deren FSC-Kurven im hohen Frequenzbereich einen erneuten Anstieg verzeichnen, und zwar über die 3σ -Kurve hinaus. Dies kann auf ver-

schiedene Fehlerquellen hindeuten. Eine scharfe Maskierung der beiden 3D Volumina, die jeweils für die Berechnung der FSC-Kurve verwendet wurden, kann ausgeschlossen werden (eigene Beobachtung). Wahrscheinlich handelt es sich um einen Effekt, der sich im *Alignment* von Rauschen begründet [185].

Dies sei anhand der Struktur II veranschaulicht: Mit dem Test in Abb. 4.12 konnte gezeigt werden, dass die Struktur II sowohl mit, als auch ohne Strukturfortsatz im Datensatz vorhanden sein muss. Es ist folglich anzunehmen, dass sich im Subdatensatz, der für die Rekonstruktion der Struktur II in Abb. 4.13 verwendet wurde, auch eine Anzahl an Einzelbildern mit dem Strukturfortsatz befindet, wodurch eine gewisse Korrelation im Rauschen zustande kommen kann. Desweiteren blieb auch fraglich, wie stabil eine Dichte, sichtbar in der 180° und 270° gedrehten Ansicht in Abb. 4.13, II an den Hauptkörper der Struktur gebunden ist. Möglicherweise trägt diese auch zur Heterogenität in dem Subdatensatz bei. So könnte der Subdatensatz, bestehend aus 26839 Einzelbildern durchaus 4 verschieden konfigurierte 3D Strukturen repräsentieren, mit durchschnittlich ungefähr 6700 Partikeln. In der 3D-III Struktur könnte der gleiche Effekt in der anhänglichen Domäne begründet sein. Bei einem Test zu ihrer Validierung, verlor die Domäne zwar eindeutig den Kontakt zum Hauptkörper (Abb. 4.17), es blieb aber eine Dichte des Rauschens ungewöhnlich kompakt in dem Bereich zurück. Auch der spitz zulaufende Fortsatz zählt zu diesen Quellen, in denen wohl hauptsächlich Rauschen kumuliert. Welche Bereiche in den Strukturen IV und V neben dem spitz zulaufenden Fortsatz problematisch sind, konnte nicht festgestellt werden. Nicht ausgeschlossen werden kann, dass sich weitere Strukturkonfigurationen im Datensatz verbergen, die mit den 5 Modellvorgaben noch nicht ausreichend berücksichtigt wurden und somit ihre Projektionsansichten in die vorhandenen Subdatensätze während des *Alignments* verteilt wurden. Aus diesem Grund waren in die Berechnung der in Abb. 4.13 dargestellten 3D Strukturen Projektionsbilder, die mit niedrigem CCC-Wert auf ein Referenzbild ausgerichtet wurden, nicht in die Rekonstruktion einbezogen worden. Allerdings waren auf diese Weise nur 1 % der Bilder nicht berücksichtigt worden.

Die für die Strukturen II - V ablesbaren Werte für die Auflösung sind vor diesem Hintergrund betrachtet, und der genannten Artefakte wegen, überschätzt. Ungewöhnlich ist auch der Verlauf der FSC-Kurven im Bereich der Schnittpunkte mit der 3σ -Kurve bei diesen Kurven, da diese dort nicht gleichmäßig auf Null abfallen.

Beschreibung der Rekonstruktionsungenauigkeiten

Eine zu große strukturelle Heterogenität verursacht Ungenauigkeiten in den rekonstruierten 3D Strukturen und limitiert dadurch deren Qualität, wie bei den Beschreibungen der 5 gefundenen Strukturen in der U1 snRNP Probe aus *S. cerevisiae* ersichtlich wurde.

Um diese Ungenauigkeiten quantifizieren zu können, wurde ein Testdatensatz erstellt. Die Modelle des Testdatensatzes waren zu unterschiedlich hohem Prozentsatz in einer Domäne beschnitten worden, um strukturelle Heterogenität der Konfiguration eines Komplexes zu simulieren. Mit den Modelldaten wurde die 3D Rekonstruktion mit der Methode *angular reconstruction* wie sie im Softwarepaket Imagic-5 implementiert ist, getestet. Die Winkelzuweisungen für die Projektionsansichten werden danach mit zunehmendem Verlust der Domäne tatsächlich ungenauer. Die Mittelwertabweichung der Projektionsrichtungen in Grad nahm von ungefähr 5° auf über 10° zu. Bei der Häufigkeitsverteilung der Winkelabweichungen fiel auf, dass auch mit zunehmendem Verlust der Domäne die überwiegende Anzahl an Projektionen die identische Projektionsrichtung zugewiesen bekamen wie das Originalmodell (noch 76 % bei 23 % Masseverlust). Umgekehrt heißt dies aber auch, dass dem kleinsten Modell in circa 25 % aller seiner Projektionen davon abweichende Projektionsrichtungen zugewiesen wurden. 60 % dieser Abweichungen lagen zwar unter $\leq 10^\circ$, es wurde aber deutlich, dass der Fehler der Rekonstruktion einem Teil an Projektionsrichtungen geschuldet ist, bei dem die Zuweisung fehlerhaft verläuft. Dass dieser Fehler nur bei einem Teil der Projektionen auftritt, mag der Grund dafür sein, dass die Methode dennoch, auch bei Konfigurationsunterschieden von Test-Struktur zu Referenzstruktur, relativ robust ist wie die Rekonstruktion der Modelle in 4.23 zeigte. Bei den Extremabweichungen von $170 - 180^\circ$ konnte durch Begutachtung des α Winkels ausgeschlossen werden, dass es sich um die spiegelbildliche Zuweisung handelte.

Desweiteren wurden die Fehler untersucht, die im MR-*Alignment* dadurch entstehen, dass eine unterschiedliche Konfiguration des Komplexes nicht vollständig berücksichtigt werden kann. Auch dieser Fall konnte mit Hilfe der Testdaten simuliert werden. Es wurde dabei nur die Rotationsabweichung im *Alignment* betrachtet. Die Auswertung der α -Winkel-Abweichungen Abb. 4.24 ergab keinen einheitlichen Zusammenhang zur Höhe des Massenverlustes. Auch geringe Massenunterschiede um 5 % können danach schon ähnlich hohe Fehler im Rotationsalignment hervorrufen wie die bei 20 % Massenunterschied von Datensatz zu Referenzset. Da das *Alignment* nicht eindeutig den mit der Referenz deckungsgleichen Teil der Struktur erkennt, kommt es zu Verschiebungen während der Ausrichtung. Es folgen daraus Klassensummen mit verringertem SNR, weil das Signal der Struktur nicht gut genug zur Deckung gebracht wurde. Im Rauschen liegen damit noch Teile des Signals, die akkumulieren. Sichtbar wird dies in den FSC-Kurven der daraus rekonstruierten 3D Strukturen in Form eines neuerlichen Anstiegs über die 3σ -Kurve hinaus, wie im Falle der Modelle A und D in Abb. 4.25 gezeigt werden konnte. Auch der Effekt des verzögerten Abfalls der FSC-Kurve im Bereich des Schnittpunktes mit der 3σ -Kurve konnte im Falle des Modells I beobachtet

werden. Es ist leicht vorstellbar, dass sich diese Effekte mit realen Daten, die zudem durch die Aufnahmetechnik per se veräussert sind, verstärken. So wurden beide Effekte in den FSC-Kurven für die 3D Strukturen II bis V (Abb. 4.14) in ausgeprägterer Form beobachtet, sollten aber bei einer einwandfreien 3D Rekonstruktion nicht auftreten (3D Struktur I). Zusammenfassend demonstrierte dieser Test, die direkte Abhängigkeit der Fourier-Schalen-Korrelation mit dem *Alignment*-Fehler [185] und die Limitierungen des *Alignments* bei stark heterogenen Proben wie dem Hefe U1 snRNP, die mit Konfigurationsunterschieden des Komplexes verbunden sind.

5.2.1 Auswirkung der Heterogenität auf die Lokalisation markierter Proteine im U1 snRNP

Grundsätzlich ist es mit Hilfe der Einzelmolekül-EM möglich, Untereinheiten eines Komplexes über Antikörperbindung und/oder genetischen Fusionsprotein-Tags zu lokalisieren, da das Label im EM-Bild des Komplexes als zusätzliche Dichte erscheint [155]. Um diese zusätzliche Dichte zweifelsfrei wahrnehmen zu können, sind allerdings einige Voraussetzungen notwendig. Zunächst ist im Falle eines Immunolabels eine spezifische Bindung des Antikörpers erforderlich, und für das genetische Label muss die korrekte Expression des Fusionsproteins sichergestellt sein. Neben diesen unerlässlichen Bedingungen muss für die verlässliche Bestimmung der Position des Labels die strukturelle Stabilität des Komplexes mit und ohne Label gewährleistet sein. Die Genauigkeit, mit der dann die Position des Labels angegeben werden kann, ist von der Auflösung der Struktur und der Größe des Labels abhängig. Ist die Auflösung der Struktur hoch, kann ein kleineres Label verwendet werden, und die Position kann dementsprechend präziser bestimmt werden.

Ein Label lässt sich prinzipiell sowohl auf 2D Ebene, als auch in der 3D Struktur des Komplexes nachweisen. Letzteres stellt allerdings höhere Anforderungen an die Rigidität der Struktur des Komplexes und des Labels. Ist das Label flexibel, und das gilt allgemein für Antikörper, aber auch für größere genetische Tags, ist der Nachweis auf 2D Ebene die bevorzugte Methode [93] [118] [115]. In diesen Fällen werden Einzelbilder des Komplexes und/oder die Klassensummen herangezogen, um das Label zu zeigen. Für einen besseren Kontrast in den EM-Bildern, werden die Komplexe in Kontrastmittel (Uranylformiat oder -acetat) eingebettet (vgl. Kap 3, Abschn. 3.3.3). Trotzdem ist das Label in der 2D Ansicht des Komplexes nur in geeigneter Projektionsrichtung gut erkennbar, und zwar dann, wenn sich seine Position am Rand der 2D Projektionsansicht befindet (Abb. 5.2, A, Position 1). Dann nämlich wird die Kontur der Ansicht an der Position des Labels gegenüber unbesetztem Partikel verändert. In diesem Fall ist seine

Position in der x-y-Ebene lokalisiert, es bleibt jedoch eine Ungenauigkeit in Richtung der z-Achse (der Projektionsrichtung). Liegt hingegen das Label in Projektionsrichtung auf dem Strukturkörper (Abb. 5.2, A, Position 2) führt dies zwar zu einer Veränderung der Grauwerte im EM-Projektionsbild aufgrund der erhöhten Elektronendichte; diese sind jedoch visuell im 2D Bild kaum erkennbar, da die Grauwertschwankungen im Vergleich zum unbesetzten Komplex relativ gering ausfallen. Somit ist in diesen Positionen eine Lokalisation in den 2D Projektionsansichten nur selten möglich.

Etwas schwieriger wird die präzise Lokalisation des Labels bei Flexibilitäten des Labels oder der Struktur im markierten Bereich (Abb. 5.2, B, I & II). Da Antikörper eine biegsame Struktur aufweisen, ist dies ein allgemeines Problem der Immunomarkierungen. Im Gegensatz zu einer statischen Dichte wird die bewegliche Dichte des Antikörpers bei der Berechnung der Klassensummen durch statistische Effekte leicht unterdrückt, so daß eine Erkennung des Antikörpers in diesen Fällen häufig nur im EM-Einzelbild möglich ist. Gleiches gilt, wenn sowohl Struktur, als auch Label flexibel sind. Dann ist der Unterschied bei vorhandenem Label zur unbesetzten Struktur noch erheblich schwieriger zu erkennen, da die entsprechenden Klassensummen in beiden Fällen in diesem Bereich unscharf definiert sind.

Ist die Struktur des Komplexes wie im Falle des U1 snRNP aus *S. cerevisiae* nicht nur in seiner Konformation dynamisch, sondern auch in seiner Konfiguration unterschiedlich, wird die zweifelsfreie Erkennung eines Labels, unabhängig von seiner Art, de facto unmöglich. Es kann nämlich in diesen Fällen nicht mehr unterschieden werden, ob die zusätzliche Dichte vom Label oder von der Struktur herrührt (Abb. 5.2, B, III). Diese Verwechselbarkeit gilt nicht nur in der speziellen Situation des Auftretens des Labels in der Nähe einer optional vorhandenen Domäne, sondern es handelt sich vielmehr um ein generelles Problem. Da a priori nicht bekannt ist, wo das Label positioniert sein wird, bleibt nämlich stets die Frage, ob die Variation der Dichte an einer bestimmten Stelle der Dynamik des Partikels oder dem Label geschuldet ist, zumal gewöhnlich die Besetzung des Partikels mit dem Label nicht sicher zu 100 % erfolgt. So wurden beispielsweise bei Experimenten in *S. cerevisiae* mit dem tDimer2-Label an Proteinen des Tri-snRNP Subpopulationen der betreffenden Proteine ohne tDimer2 beobachtet (in [92] Abb.3.8, Größe der Subpopulation approx. 20 %, pers. Kommunikation Dr. I. Hacker, Dr. M. Golas), was bereits aus der Proteingelanalyse ersichtlich wurde. Desweiteren besteht die Möglichkeit, dass das Label an der optional auftretenden Domäne lokalisiert ist, was die Besetzung des Partikels mit dem Label zusätzlich reduziert und die Interpretation erschwert.

Summa summarum ist eine Auswertung von Labelpositionen von derart heterogenen Komplexen wie dem U1 snRNP aus *S. cerevisiae* mit dieser Art Labeln nicht zuverlässig.

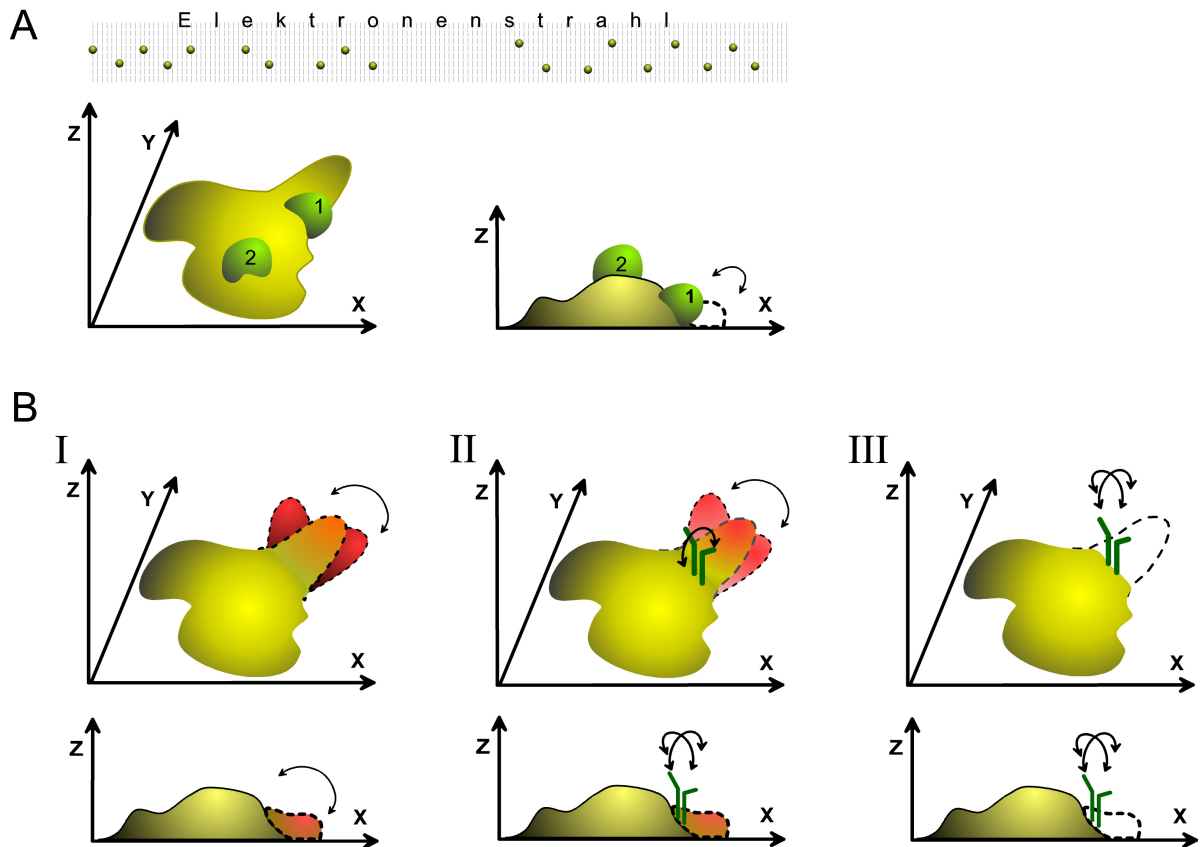


Abbildung 5.2: Schwierigkeiten bei der Auswertung von Positionen eines Labels.

A: Das Label ist auf 2D Ebene nur in geeigneter Projektionsrichtung gut erkennbar, und zwar wenn er am Rand der Struktur liegt und damit die Kontur der Ansicht verändert (Position 1). Nicht ermittelt werden kann die Lage des Labels in Richtung der z -Achse, der Richtung des Elektronenstrahls.

B: Bei flexiblen Strukturdomänen des Komplexes und/oder des Labels (I & II) wird die Interpretation dadurch erschwert, dass variierende Dichte aufgrund statistischer Effekte bei der Datenverarbeitung minimiert wird. Der Besetzungsunterschied kann in diesen Fällen meist nicht in Klassensummen, sondern nur in EM-Einzelbildern nachgewiesen werden. (III) Unmöglich wird die Analyse von Markierungsexperimenten bei unterschiedlichen Konfigurationen des Komplexes, da die jeweils vorhandene Dichte nicht eindeutig auf die Struktur oder auf den Label zurückgeführt werden kann.

sig möglich.

Bei den Vorexperimenten für die Markierungen konnten allerdings zwei interessante Beobachtungen gemacht werden.

- (1) Die Fusion des tDimer2-tags an das Nam8 Protein erzeugte zwar, wie erwartet [57], einen lebensfähigen Hefestamm, aber es war nicht möglich, den assemblierten 17 – 18 S U1 snRNP zu isolieren. Möglicherweise hat hier der große genetische Tag den Aufbau des Komplexes zu stark beeinträchtigt. Es konnte schon vorher gezeigt werden, dass aus einem Δ Nam8p Hefestamm ein U1 snRNP, der aus Hefeextrakt mit 450 mM KCl in den Puffern aufgereinigt wurde, als 13 – 14 S Partikel im Glycerolgradienten sedimentierte [79]. Von diesem U1 snRNP waren alle Hefe-spezifischen Proteine mit Ausnahme von Prp42p dissoziiert. Die elektronenmikroskopischen 2D Ansichten dieses Partikels sahen denen des humanen U1 snRNP ziemlich ähnlich [78]. Aus dem Wildtyp-Hefestamm dagegen konnte das U1 snRNP noch bei unphysiologisch hoher Salzkonzentration (700 mM KCl) als 17 S Partikel aufgereinigt werden. Diese und die in dieser Arbeit gemachten Beobachtung lassen die Vermutung zu, dass Nam8p, trotzdem es deletiert werden kann, eine kritische Funktion für die Integrität des U1 snRNP zukommt. Da eine Deletion von Nam8p nicht letal wirkt, wurde bereits in [78] spekuliert, dass in diesem Fall andere U1 snRNP Proteine die Position von Nam8p übernehmen könnten.
- (2) Bei den Bindungsstudien mit dem U1 snRNP und einem Antikörper, der das Calmodulin-Bindungs-Peptid des TAP-tags erkennt, fiel folgendes auf: Es war nicht möglich, signifikante Antikörpermengen in den Glycerolgradientenfraktionen, in denen das U1 snRNP mit TAP-tag am Snu71p erwartet wird, zu detektieren. Für die analogen Experimente mit dem TAP-tag an Snu56p deuteten die Vorexperimente zumindest an, dass Antikörper mit dem U1 snRNP sedimentierten, wenn auch nur in sehr geringem Maße. Dies ist insofern interessant, als dass mit [58] Daten vorliegen, die einen Subkomplex von Prp40p, Luc7p und Snu71p vermuten lassen. Beruhend darauf wurde auch spekuliert, dass Snu71p, das nicht essentiell ist, auch nicht absolut notwendig für die U1 snRNP Funktion ist. Möglicherweise lässt daher die Bindung des Antikörpers an das Snu71p des isolierten U1 snRNP die Integrität dieses Subkomplexes zerfallen, so dass Snu71p dissoziiert. Dieser Fall erklärte die oben beschriebene Beobachtung.

5.3 Schlussfolgerungen

5.3.1 U1 snRNP in *S. cerevisiae*

In dieser Arbeit wurden erstmals dreidimensionale EM-Strukturen des U1 snRNP aus *S. cerevisiae* gezeigt. Der starke Polymorphismus der Strukturmodelle des isolierten

S. cerevisiae 18 S U1 snRNP, die mit neusten EM-Datenverarbeitungsverfahren, die die Trennung verschiedener 3D Strukturen in einem Datensatz erlauben, ermittelt worden waren, deutet auf zellbiologische Prozesse hin, die in der Aufreinigung des Komplexes nicht berücksichtigt werden konnten. So wurden sehr wahrscheinlich mehrere Assemblierungsstadien des U1 snRNP gleichzeitig aufgereinigt, was bei der biochemischen Proteingelanalyse nicht bemerkt werden kann, aber durch die in dieser Arbeit gewonnenen strukturbiologischen Einblicke ersichtlich wurde.

In der Vergangenheit wurden mit verschiedenen Aufreinigungsprotokollen stets weitere Proteine gefunden, die mit der U1 snRNA in Hefe isolierbar waren [149] [162] [79] [169] [64], und die aus diesem Grund dem U1 snRNP als weitere Komponente zugerechnet wurden. Umgekehrt wurden in einem Teil dieser Aufreinigungen zusätzliche Proteine isoliert, die nicht als Bestandteil des U1 snRNP Komplexes gelten konnten, wie Npl3p [79] und Sto1 [258], der großen Untereinheit des CBC.

Der aufgrund dieser Ergebnisse definierte 18 S U1 snRNP besteht abgesehen vom CBC und dem Substrat, der prä-mRNA, aus allen Proteinen des CC1 in *S. cerevisiae*. Beide Proteine des CBC wurden allerdings auch schon nach der TAP-tag Aufreinigung als U1 snRNP-assoziiert gefundenen, und zwar die große Untereinheit, Sto1p, in [258] und die kleine Untereinheit, Cbc2/Mud13 Protein, im eigenen Hause (Anhang Tab. 1), Der CC1 scheint somit schon weitgehend vororganisiert vorzuliegen und die Tatsache, dass diese Aufreinigungen so möglich waren, spiegelt insofern nur die Konsequenz wieder, mit der sich das U1 snRNP in *S. cerevisiae* zu CC1 formiert. In Anbetracht eines effektiven Spleißprozesses, in dem die Einweisung einer prä-mRNA in den Spleißzyklus kotranskriptionell geschieht [5], bedarf es einer entsprechenden Reaktionskinetik bei Eintreffen des Substrats, und so scheint diese Vororganisation durchaus sinnvoll, zumal im speziellen Fall des Hefe-Organismus. Dort nämlich darf davon ausgegangen werden, dass sich die CC, und auch die folgenden Spleißkomplexe, viel gleichartiger assemblieren, als im humanen System, da auf die Fülle von Spleißregulatoren und -faktoren, die im humanen System benötigt werden, um die geringer konservierten Sequenzen an den Spleißstellen zu erkennen, verzichtet werden kann. Auch entfällt in *S. cerevisiae* die komplexe Regulation des alternativen Spleißprozesses. So war es aus Hefe auch möglich, das insgesamt kompaktere, vororganisierte Spleißsystem unter bestimmten Bedingungen als Penta-snRNP, bestehend aus den snRNP U1,U2,U4/U6 und U5, zu isolieren [253]. Ebenfalls war die Isolierung eines Tetra-snRNP ohne U1 möglich, was allerdings auch schon mal in ähnlicher Weise aus humanem Zellextrakt gelang [125]. Hiermit wird möglicherweise die Vororganisation der makromolekularen Komplexe widergespiegelt, was wiederum auch die Schwierigkeit zeigt, Bedingungen zu finden, um definierte Subkomplexe aufzureinigen.

Das System der Assemblierung der CC ist in Hefe offensichtlich so kompakt, dass bisher kein auf das humane Äquivalent reduzierter U1 snRNP aus *S. cerevisiae* isoliert worden ist. Allerdings konnte interessanterweise aus einem Δ NAM8 Hefestamm ein stabiler 14 S U1 snRNP isoliert werden, der sich homolog dem humanen U1 snRNP zusammensetzte und als einziges weiteres Protein Prp42 trug [79]. Die 2D Ansichten dieses Komplexes ähnelten denen des humanen U1 snRNP sehr [78].

Bisher ist über die Biogenese des 18 S U1 snRNP und der CC1 im Hefe-Organismus wenig bekannt. Die hier gefundenen strukturellen Daten fügen sich jedoch gut ein in funktionelle Daten, die über das U1 snRNP vorliegen, und die Hinweise geben, dass sich das 18 S U1 snRNP sehr viel modulartiger als bisher gemeinhin geglaubt, assembliert. Eine Quelle der strukturellen Heterogenität könnte mit dem Subkomplex, bestehend aus Snu71p, Luc7p und Prp40p, zusammenhängen. Dieser wird in [58] überhaupt erst als Subkomplex im U1 snRNP bezeichnet, wobei Prp40p scheinbar die Proteinkontakte zu Luc7p und Snu71p vermittelt. Der Subkomplex scheint für die kotranskriptionelle Regulation des Spleißprozesses zu sorgen, da Prp40p den Kontakt zur phosphorylierten CDT Domäne der RNA-Polymerase II herzustellen vermag [165]. Für Luc7p konnte bereits gezeigt werden, dass in seiner Anwesenheit die CBP-proximale 5' Ss bevorzugt gespleißt wird [64], und es liegt eine Interaktion mit dem CBC auf genetischer Ebene vor [65]. Die Funktion von Snu71p bleibt unklar. Interessanterweise wurde Luc7p erst mit der Tap-tag Aufreinigung über Snu71p als U1 snRNP assoziiert entdeckt. In das Bild eines Subkomplexes bestehend aus diesen drei Proteinen passt auch die massenspektroskopische Analyse eines Glycerolgradienten nach erfolgter U1 snRNP Aufreinigung (Appendix, Abb. 4 Tab. 1), wonach diese drei Proteine auch in den oberen Glycerolgradientenfraktionen einen Peak haben.

Die in dieser Arbeit unterschiedlichen 3D Strukturen III und V könnten den Prozess des Andockens dieses Subkomplexes widerspiegeln. Während die extra Domäne in Struktur III nicht verifiziert werden konnte, scheint die 3D Struktur V nicht ohne zu existieren (Abb. 4.17). Das Andocken des Subkomplexes könnte die offensichtlichen Konformationsunterschiede des Rückgrats des Hauptstrukturkörpers induzieren. Die ungefähre Masse der zusätzlichen Domäne in 3D Struktur V liegt mit 195 *kDa* sehr gut im Rahmen der erwarteten Masse von 170 *kDa* für diesen Subkomplex. Auch haben die Hauptstrukturkörper von III und V jeweils ohne die Domäne eine ziemlich ähnlich veranschlagte Masse, tendenziell ist die von Struktur III etwas größer. Diese Differenz muss keine Bedeutung haben, aber vielleicht verlässt zeitgleich ein anderes Protein schräg gegenüber den Komplex, um die Brechung des Strukturrückgrats (Abb. 5.3) zuzulassen. Dies ist zwar spekulativ, Labelexperimente könnten hier in Zukunft aber Aufschluss geben, sobald eine Aufreinigung eines definierten Assemblierungsstadiums des U1 snRNP möglich

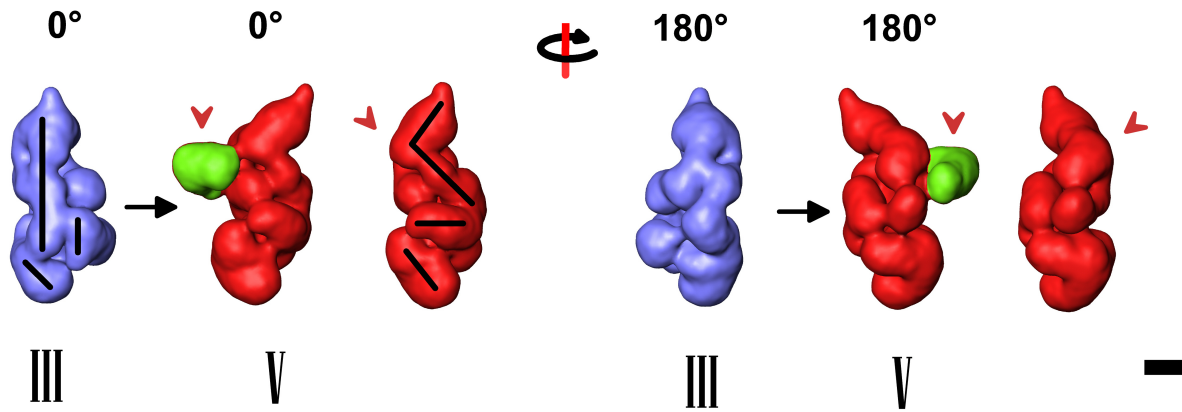


Abbildung 5.3: Modell der Biogenese des U1 snRNP. Der Subkomplex (grün), bestehend aus Prp40p, Snu71p und Luc7p, wie er in [58] beschrieben worden ist, könnte in einem Modell der Biogenese des U1 snRNP leicht verzögert vollständig an übrigen prä-assemblierten Hauptkörper des 18 S U1 snRNP (blau) andocken wodurch Konformationsumwandlungen des Rückgrats der Hauptstruktur von einer linearen zu einer doppel-S-förmigen Anordnung (rot in Struktur V) induziert werden (schwarz eingezeichnet). Der Maßstab beträgt 5 nm.

ist (s. Abschn. 5.3.2).

Eine weitere Ursache für die Heterogenität können Subpopulationen des U1 snRNP ohne Prp39p sein. Diese wurden von Lockart & Rymond in [149] beschrieben, wobei jedoch nicht abschließend geklärt werden konnte, ob diese auch funktionelle Bedeutung haben, zumal Prp39p essentiell für die Bildung der CC ist. Daher könnte es sich um ein natürliches Zwischenprodukt bei der Biogenese des U1 snRNP gehandelt haben, was bereits in der genannten Publikation diskutiert worden ist. Da es nun dieses Indiz auf An- und Abwesenheit von Prp39p gibt, kommt dies als eine mögliche Ursache der beobachteten strukturellen Heterogenität des 18 S U1 snRNP in Frage. Prp42p ist ebenfalls essentiell für die Funktion von U1 snRNP, da es bei Deletion zu keiner Bildung der CC1/CC2 kam [162]. Dennoch wurde es in [169] noch nicht mit dem U1 snRNP isoliert. Beide Proteine, Prp39 und Prp42, für die bisher keine humanen homologen Proteine gefunden worden sind, tragen mit ihren TPR-Motiven Merkmale von Gerüstproteinen. Da in der Hefe nicht die Fülle von Spleißfaktoren existiert, beispielsweise kanonische SR Proteine, die assistieren könnten bei der Überwindung der Distanz von 5' Ss zu Verzweigungspunkt und 3' Ss in den frühen Präspliceosom-Komplexen [202], erklärt dies vielleicht die Funktion für Prp39p und Prp42p. Sie könnten als Adaptorproteine helfen, ein Gerüst auszubauen, um 5' Ss und 3' Ss für die Spleißreaktion einander näher zu bringen. Dies erklärte auch die elongierte Struktur des 18 S U1 snRNP, der in dieser Arbeit beobachtet wurde und bereits in 2D Ansichten in [59] gefunden wurde, im Vergleich zum 14 S

U1 snRNP [78] und zum humanen U1 snRNP [249] [192].

5.3.2 Ausblick

Für eine umfassende Strukturbestimmung aller beobachteten Zustände des U1 snRNP sollte der Prozess der Assemblierung idealerweise angehalten werden können, um einen in der Zusammensetzung definierten Komplex zu erhalten. Da keine natürlichen Antibiotika wie im Falle des Ribosoms zur Verfügung stehen, sind mehrere Optionen zu prüfen. Eine Möglichkeit ist, stringenteren Aufreinigungsbedingungen zu wählen, auch wenn diese potentiell die native Struktur des Komplexes gefährden. Unphysiologische Salzkonzentration um 500 *mM* schon frühzeitig während der Aufreinigung, könnten 'Kontaminationen' ausschließen und dennoch eine sinnvolle Anwendung der GraFix-Methode gewährleisten. Da Snu71p in einem Subkomplex mit Luc7p und Prp40p vermutet wird, ist eine Aufreinigung über ein anderes TAP-tag-Fusionsprotein, das im Hauptkörper der Struktur vermutet wird, ebenfalls eine einfache Modifikation, um die Experimente durchzuführen und die Ergebnisse zu vergleichen.

Da das TAP-tag Protokoll mit der Elution des Komplexes von der Calmodulinmatrix mit dem Entzug von divalenten Ionen im Puffer verbunden ist, birgt dieser Schritt grundsätzlich eine Gefahr für die strukturelle Integrität des Komplexes. Da aber bisher unabhängig vom Protokoll stets zwei Affinitätsaufreinigungsschritte für die Isolierung der snRNP benötigt wurden [252], könnte eine andere Kombination im TAP-tag eingeführt werden, mit der der Calmodulinmatrix-Aufreinigungsschritt ersetzt wird und von der erwartet wird, dass sie die Struktur weniger beeinflusst.

Im Sinne einer Strukturbestimmung unter möglichst nativen Pufferbedingungen, sind in der Hefe jedoch aufgrund der Anwendung vielfältiger genetischen Techniken andere Wege denkbar. Es besteht die Möglichkeit, durch Deletion einzelner Proteine arretierte Komplexe zu isolieren, wie möglicherweise aus dem Δ Nam8p Hefestamm bereits geschehen [78]. Da viele der 18S U1 snRNP in *S. cerevisiae* essentiell sind, ist die Auswahl für Gendeletionen allerdings begrenzt. Vielleicht könnte mit Hefestämmen gearbeitet werden, die eine konditionelle Proteindeletion tragen [256].

Da ein großer Teil der Hefe U1 snRNA deletiert werden kann, ohne dass es zu schweren Einschränkungen des Wachstums der Hefezellen kommt [239], könnte eine U1 snRNP Aufreinigung beispielsweise aus einer Δ 192 – 507 U1 RNA Deletionsmutante ausgeführt werden, um festzustellen, inwieweit sich das U1 snRNP assembliert und ob es zu geringerer Heterogenität führt, wenn ein Großteil der RNA fehlt.

Eine fundamental andere Methode wäre, gezielt U1 snRNP als Teil des CC1 zu isolieren.

Hierzu könnte während der Aufreinigung eine kurze RNA-Sequenz mit 5' Ss angeboten werden. Das Substrat sollte keine 3' Ss oder Verzweigungspunktabschnitte enthalten, um die Bildung von CC2 und nachfolgenden E-Komplexen zu vermeiden. Letzteres wird auch weitgehend durch ATP-freie Bedingungen erreicht. Bisher galt die Arretierung der Assemblierung im CC-Zustand als schwierig, da *in vitro* die CC nur beobachtet worden waren, wenn U2 snRNP inhibiert wurde oder kein ATP zur Verfügung stand [229] [143]. Es wurde daher angenommen, dass die Anlagerung von U1 snRNP an den Komplex im Vergleich zur nachfolgenden Anlagerung des U2 snRNP langsam verläuft. Tardiff & Rosbash hingegen schlagen aufgrund ihrer Ergebnisse eine frühe, schnelle Anlagerung des U1 snRNP an die prä-mRNA *in vivo* vor und eine verzögerte Assemblierung von U2 snRNP und Tri-snRNP [258]. Die hier gefundenen strukturellen Ergebnisse über den 18S U1 snRNP zeigen, dass es notwendig sein wird, in der weiteren Strukturaufklärung gezielt einen Zustand zu begünstigen. Da ein Unterschied zwischen *in vitro* und kotranskriptionellen *in vivo* Assemblierungsstadien diskutiert wird, wäre demnach eine Strukturbestimmung von CC1 auch unter diesem Aspekt wünschenswert.

Literaturverzeichnis

- [1] N. Abovich and M. Rosbash. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*, 89(3):403–12, 1997.
- [2] T. Achsel, H. Brahm, B. Kastner, A. Bachi, M. Wilm, and R. Luhrmann. A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J*, 18(20):5789–802, 1999.
- [3] P. Ajuh, B. Kuster, K. Panov, J. C. Zomerdijk, M. Mann, and A. I. Lamond. Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO J*, 19(23):6569–81, 2000.
- [4] M. Allen, A. Friedler, O. Schon, and M. Bycroft. The structure of an FF domain from human HYPA/FBP11. *J Mol Biol*, 323(3):411–6, 2002.
- [5] J. Ares, M. and N. J. Proudfoot. The spanish connection: transcription and mRNA processing get even closer. *Cell*, 120(2):163–6, 2005.
- [6] S. Arning, P. Gruter, G. Bilbe, and A. Kramer. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA*, 2(8):794–810, 1996.
- [7] G. Ast. How did alternative splicing evolve? *Nat Rev Genet*, 5(10):773–82, 2004.
- [8] O. Autorenangabe. Cryomicroscopy and radiation damage. International Study Group for Cryo Electron Microscopy Workshop. Schloss Ringberg am Tegernsee, 9-14 November 1981. *Ultramicroscopy*, 10(1-2):1–177, 1982.
- [9] S. Awasthi, R. Palmer, M. Castro, C. D. Mobarak, and S. W. Ruby. New roles for the Snp1 and Exo84 proteins in yeast pre-mRNA splicing. *J Biol Chem*, 276(33):31004–15, 2001.
- [10] M. Azubel, S. G. Wolf, J. Sperling, and R. Sperling. Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol Cell*, 15(5):833–9, 2004.
- [11] N. L. Barbosa-Morais, M. Carmo-Fonseca, and S. Aparicio. Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. *Genome Res*, 16(1):66–77, 2006.
- [12] R. Beckmann, D. Bubeck, R. Grassucci, P. Penczek, A. Verschoor, G. Blobel, and J. Frank. Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science*, 278(5346):2123–6, 1997.

- [13] E. Ben Asher, O. Groudinsky, G. Dujardin, N. Altamura, M. Kermorgant, and P. Slonimski. Novel class of nuclear genes involved in both mRNA splicing and protein synthesis in *Saccharomyces cerevisiae* mitochondria. *Mol Gen Genet*, 215: 517–528, 1989.
- [14] S. M. Berget. Exon recognition in vertebrate splicing. *J Biol Chem*, 270(6):2411–4, 1995.
- [15] J. A. Berglund, K. Chua, N. Abovich, R. Reed, and M. Rosbash. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UA-CUAAC. *Cell*, 89(5):781–7, 1997.
- [16] J. A. Berglund, N. Abovich, and M. Rosbash. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev*, 12(6):858–67, 1998.
- [17] J. A. Berglund, M. L. Fleming, and M. Rosbash. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA*, 4(8):998–1006, 1998.
- [18] D. L. Black. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, 72:291–336, 2003.
- [19] H. Blum, H. Beier, and H. Gross. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*, 8:93–99, 1987.
- [20] P. Bochnig, R. Reuter, P. Bringmann, and R. Luhrmann. A monoclonal antibody against 2,2,7-trimethylguanosine that reacts with intact, class U, small nuclear ribonucleoproteins as well as with 7-methylguanosine-capped RNAs. *Eur J Biochem*, 168(2):461–7, 1987.
- [21] D. Boehringer. Elektronenmikroskopische 3D Strukturbestimmung des Spleissosoms. *Dissertation, Georg-August-Universitaet Goettingen*, 2005.
- [22] D. Boehringer, E. M. Makarov, B. Sander, O. V. Makarova, B. Kastner, R. Luhrmann, and H. Stark. Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat Struct Mol Biol*, 11(5):463–8, 2004.
- [23] L. Borland and M. van Heel. Classification of image data in conjugate representation spaces. *J. Opt. Soc. Am. A*, 7(4):601–610, 1990.
- [24] B. Bottcher, S. A. Wynne, and R. A. Crowther. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, 386(6620): 88–91, 1997.
- [25] P. Bringmann, B. Appel, J. Rinke, R. Reuter, H. Theissen, and R. Luhrmann. Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J*, 3(6): 1357–63, 1984.
- [26] E. Brody and J. Abelson. The spliceosome: yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science*, 228(4702):963–7, 1985.

- [27] D. A. Brow. Allosteric cascade of spliceosome activation. *Annu Rev Genet*, 36: 333–60, 2002.
- [28] C. B. Burge, T. Tuschl, and P. A. Sharp. Splicing of Precursors to mRNAs by the Spliceosomes, 1999.
- [29] H. Busch, R. Reddy, L. Rothblum, and Y. C. Choi. SnRNAs, SnRNPs, and RNA processing. *Annu Rev Biochem*, 51:617–54, 1982.
- [30] R. E. Campbell, O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias, and R. Y. Tsien. A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A*, 99(12):7877–82, 2002.
- [31] S. M. Carty, A. C. Goldstrohm, C. Sune, M. A. Garcia-Blanco, and A. L. Greenleaf. Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proc Natl Acad Sci U S A*, 97(16): 9015–20, 2000.
- [32] S. P. Chan, D. I. Kao, W. Y. Tsai, and S. C. Cheng. The Prp19p-associated complex in spliceosome activation. *Science*, 302(5643):279–82, 2003.
- [33] C. H. Chen, W. Y. Tsai, H. R. Chen, C. H. Wang, and S. C. Cheng. Identification and characterization of two novel components of the Prp19p-associated complex, Ntc30p and Ntc20p. *J Biol Chem*, 276(1):488–94, 2001.
- [34] C. H. Chen, W. C. Yu, T. Y. Tsao, L. Y. Wang, H. R. Chen, J. Y. Lin, W. Y. Tsai, and S. C. Cheng. Functional and physical interactions between components of the Prp19p-associated complex. *Nucleic Acids Res*, 30(4):1029–37, 2002.
- [35] H. R. Chen, S. P. Jan, T. Y. Tsao, Y. J. Sheu, J. Banroques, and S. C. Cheng. Snt309p, a component of the Prp19p-associated complex that interacts with Prp19p and associates with the spliceosome simultaneously with or immediately after dissociation of U4 in the same manner as Prp19p. *Mol Cell Biol*, 18(4): 2196–204, 1998.
- [36] J. Y. Chen, L. Stands, J. P. Staley, J. Jackups, R. R., L. J. Latus, and T. H. Chang. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol Cell*, 7(1): 227–32, 2001.
- [37] J. Z. Chen and N. Grigorieff. SIGNATURE: a single-particle selection system for molecular electron microscopy. *J Struct Biol*, 157(1):168–73, 2007.
- [38] H. Cheng, K. Dufu, C. S. Lee, J. L. Hsu, A. Dias, and R. Reed. Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, 127(7):1389–400, 2006.
- [39] S. C. Cheng and J. Abelson. Spliceosome assembly in yeast. *Genes Dev*, 1(9): 1014–27, 1987.
- [40] S. C. Cheng, W. Y. Tarn, T. Y. Tsao, and J. Abelson. PRP19: a novel spliceosomal component. *Mol Cell Biol*, 13(3):1876–82, 1993.

- [41] D. F. Colgan and J. L. Manley. Mechanism and regulation of mRNA polyadenylation. *Genes Dev*, 11(21):2755–66, 1997.
- [42] H. V. Colot, F. Stutz, and M. Rosbash. The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes Dev*, 10(13):1699–708, 1996.
- [43] R. Crowther, D. De Rosier, and A. Klug. The reconstruction of a three dimensional structure from projections and its application to electron microscopy. *Proc. Roy. Soc. London*, A317:319–340, 1970.
- [44] L. D. D’Andrea and L. Regan. TPR proteins: the versatile helix. *Trends Biochem Sci*, 28(12):655–62, 2003.
- [45] A. K. Das, P. W. Cohen, and D. Barford. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J*, 17(5):1192–9, 1998.
- [46] R. Das, Z. Zhou, and R. Reed. Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell*, 5(5):779–87, 2000.
- [47] D. De Rosier and A. Klug. Reconstruction of threedimensional structures from electron microscopy. *Nature*, 217:130–134, 1968.
- [48] J. Deckert, K. Hartmuth, D. Boehringer, N. Behzadnia, C. L. Will, B. Kastner, H. Stark, H. Urlaub, and R. Luhrmann. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol*, 26(14):5528–43, 2006.
- [49] F. Del Gatto-Konczak, C. F. Bourgeois, C. Le Guiner, L. Kister, M. C. Gesnel, J. Stevenin, and R. Breathnach. The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5’ splice site. *Mol Cell Biol*, 20(17):6287–99, 2000.
- [50] C. Ding and X. F. He. Cluster Structure of k-means clustering via principal component analysis. *Advances in Knowledge Discovery and Data Mining, Proceedings*, 3056:414–418, 2004.
- [51] I. Dix, C. Russell, S. B. Yehuda, M. Kupiec, and J. D. Beggs. The identification and characterization of a novel splicing protein, Isy1p, of *Saccharomyces cerevisiae*. *RNA*, 5(3):360–8, 1999.
- [52] K. H. Downing. Spot-scan imaging in transmission electron microscopy. *Science*, 251(4989):53–9, 1991.
- [53] H. Du and M. Rosbash. Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing. *RNA*, 7(1):133–42, 2001.
- [54] H. Du and M. Rosbash. The U1 snRNP protein U1C recognizes the 5’ splice site in the absence of base pairing. *Nature*, 419(6902):86–90, 2002.
- [55] P. Dube, P. Tavares, R. Lurz, and M. van Heel. The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J*, 12(4):1303–9, 1993.

- [56] P. Dube, F. Herzog, C. Gieffers, B. Sander, D. Riedel, S. A. Muller, A. Engel, J. M. Peters, and H. Stark. Localization of the coactivator Cdh1 and the cullin subunit Apc2 in a cryo-electron microscopy model of vertebrate APC/C. *Mol Cell*, 20(6):867–79, 2005.
- [57] K. Ekwall, M. Kermorgant, G. Dujardin, O. Groudinsky, and P. P. Slonimski. The NAM8 gene in *Saccharomyces cerevisiae* encodes a protein with putative RNA binding motifs and acts as a suppressor of mitochondrial splicing deficiencies when overexpressed. *Mol Gen Genet*, 233(1-2):136–44, 1992.
- [58] C. Ester and P. Uetz. The FF domains of yeast U1 snRNP protein Prp40 mediate interactions with Luc7 and Snu71. *BMC Biochem*, 9:29, 2008.
- [59] P. Fabrizio, S. Esser, B. Kastner, and R. Luhrmann. Isolation of *S. cerevisiae* snRNPs: comparison of U1 and U4/U6.U5 to their human counterparts. *Science*, 264(5156):261–5, 1994.
- [60] P. Fabrizio, J. Dannenberg, P. Dube, B. Kastner, H. Stark, H. Urlaub, and R. Luhrmann. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell*, 36(4):593–608, 2009.
- [61] J. Fleckner, M. Zhang, J. Valcarcel, and M. R. Green. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev*, 11(14):1864–72, 1997.
- [62] P. Forch, O. Puig, N. Kedersha, C. Martinez, S. Granneman, B. Seraphin, P. Anderson, and J. Valcarcel. The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Mol Cell*, 6(5):1089–98, 2000.
- [63] P. Forch, O. Puig, C. Martinez, B. Seraphin, and J. Valcarcel. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *EMBO J*, 21(24):6882–92, 2002.
- [64] P. Fortes, D. Bilbao-Cortes, M. Fornerod, G. Rigaut, W. Raymond, B. Seraphin, and I. W. Mattaj. Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site recognition. *Genes Dev*, 13(18):2425–38, 1999.
- [65] P. Fortes, J. Kufel, M. Fornerod, M. Polycarpou-Schwarz, D. Lafontaine, D. Tollervey, and I. W. Mattaj. Genetic and physical interactions involving the yeast nuclear cap-binding complex. *Mol Cell Biol*, 19(10):6543–53, 1999.
- [66] J. Frank. Multivariate Data Analysis and Classification of Images in 'Three-Dimensional Electron Microscopy of Macromolecular Assemblies'. *Oxford University Press, Inc*, pages 145–192, 2006.
- [67] J. Frank and M. van Heel. Correspondence analysis of aligned images of biological particles. *J Mol Biol*, 161(1):134–7, 1982.
- [68] D. Frendewey and W. Keller. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell*, 42(1):355–67, 1985.

- [69] M. Fromont-Racine, J. C. Rain, and P. Legrain. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet*, 16(3):277–82, 1997.
- [70] X. D. Fu. The superfamily of arginine/serine-rich splicing factors. *RNA*, 1(7):663–80, 1995.
- [71] Y. Furuichi and A. J. Shatkin. Caps on Eukaryotic mRNAs. *Encyclopedia of Life Sciences, John Wiley and Sons, Ltd. www.els.net*, 2007.
- [72] H. Gao, M. Valle, M. Ehrenberg, and J. Frank. Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *J Struct Biol*, 147(3):283–90, 2004.
- [73] W. Gilbert, C. W. Siebel, and C. Guthrie. Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. *RNA*, 7(2):302–13, 2001.
- [74] A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. Life with 6000 genes. *Science*, 274(5287):546, 563–7, 1996.
- [75] V. Goguel, X. L. Liao, B. C. Rymond, and M. Rosbash. U1 snRNP can influence 3'-splice site selection as well as 5'-splice site selection. *Genes Dev*, 5(8):1430–8, 1991.
- [76] M. M. Golas, B. Sander, C. L. Will, R. Luhrmann, and H. Stark. Molecular architecture of the multiprotein splicing factor SF3b. *Science*, 300(5621):980–4, 2003.
- [77] M. M. Golas, B. Sander, C. L. Will, R. Luhrmann, and H. Stark. Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol Cell*, 17(6):869–83, 2005.
- [78] A. Gottschalk. Umfassende Analyse der Proteinkomponenten spleissosomaler Ribonukleoprotein-Komplexe aus *Saccharomyces Cerevisiae*. *Dissertation, Philipps-Universitaet Marburg*, 1999.
- [79] A. Gottschalk, J. Tang, O. Puig, J. Salgado, G. Neubauer, H. V. Colot, M. Mann, B. Seraphin, M. Rosbash, R. Luhrmann, and P. Fabrizio. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA*, 4(4):374–93, 1998.
- [80] A. Gottschalk, G. Neubauer, J. Banroques, M. Mann, R. Luhrmann, and P. Fabrizio. Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J*, 18(16):4535–48, 1999.
- [81] A. Gottschalk, C. Bartels, G. Neubauer, R. Luhrmann, and P. Fabrizio. A novel yeast U2 snRNP protein, Snu17p, is required for the first catalytic step of splicing and for progression of spliceosome assembly. *Mol Cell Biol*, 21(9):3037–46, 2001.

- [82] A. Gottschalk, B. Kastner, R. Luhrmann, and P. Fabrizio. The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p. *RNA*, 7(11):1554–65, 2001.
- [83] O. Gozani, R. Feld, and R. Reed. 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(2):233–43, 1996.
- [84] O. Gozani, J. Potashkin, and R. Reed. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol*, 18(8):4752–60, 1998.
- [85] P. J. Grabowski and P. A. Sharp. Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonucleoprotein particles in the spliceosome. *Science*, 233(4770):1294–9, 1986.
- [86] P. J. Grabowski, S. R. Seiler, and P. A. Sharp. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell*, 42(1):345–53, 1985.
- [87] R. J. Grainger and J. D. Beggs. Prp8 protein: at the heart of the spliceosome. *RNA*, 11(5):533–57, 2005.
- [88] R. J. Grainger and J. D. Beggs. Spliceosomal Machinery. *Encyclopedia of Life Sciences, John Wiley and Sons, Ltd. www.els.net*, 2007.
- [89] B. R. Graveley. Sorting out the complexity of SR protein functions. *RNA*, 6(9):1197–211, 2000.
- [90] C. Guthrie. Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science*, 253(5016):157–63, 1991.
- [91] C. Guthrie and B. Patterson. Spliceosomal snRNAs. *Annu Rev Genet*, 22:387–419, 1988.
- [92] I. Hacker. Electron microscopic localization of tagged proteins in yeast *S. cerevisiae* spliceosomal U4/U6.U5 tri-snRNP. *Dissertation, Georg-August-Universitaet Goettingen*, 2008.
- [93] I. Hacker, B. Sander, M. M. Golas, E. Wolf, E. Karagoz, B. Kastner, H. Stark, P. Fabrizio, and R. Luhrmann. Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nat Struct Mol Biol*, 15(11):1206–12, 2008.
- [94] S. Hacker and H. Krebber. Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. *J Biol Chem*, 279(7):5049–52, 2004.
- [95] J. Hamm, M. Kazmaier, and I. W. Mattaj. In vitro assembly of U1 snRNPs. *EMBO J*, 6(11):3479–85, 1987.
- [96] K. Hartmuth, H. Urlaub, H. P. Vornlocher, C. L. Will, M. Gentzel, M. Wilm, and R. Luhrmann. Protein composition of human pre-spliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A*, 99(26):16719–24, 2002.

- [97] C. Hashimoto and J. A. Steitz. U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res*, 12(7):3283–93, 1984.
- [98] F. Hauer. Three-dimensional electron microscopy of structurally heterogenous biological macromolecules. *Dissertation, Georg-August-Universitaet Goettingen*, 2009.
- [99] V. Heinrichs, M. Bach, G. Winkelmann, and R. Luhrmann. U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science*, 247(4938):69–72, 1990.
- [100] R. Henderson. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q Rev Biophys*, 28(2):171–93, 1995.
- [101] H. Hermann, P. Fabrizio, V. A. Raker, K. Foulaki, H. Hornig, H. Brahms, and R. Luhrmann. snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J*, 14(9):2076–88, 1995.
- [102] P. J. Hilleren, H. Y. Kao, and P. G. Siliciano. The amino-terminal domain of yeast U1-70K is necessary and sufficient for function. *Mol Cell Biol*, 15(11):6341–50, 1995.
- [103] T. Hirano, N. Kinoshita, K. Morikawa, and M. Yanagida. Snap helix with knob and hole: essential repeats in *S. pombe* nuclear protein nuc2+. *Cell*, 60(2):319–28, 1990.
- [104] E. Jankowsky and H. Bowers. Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res*, 34(15):4181–8, 2006.
- [105] A. Jarmolowski and I. W. Mattaj. The determinants for Sm protein binding to *Xenopus* U1 and U5 snRNAs are complex and non-identical. *EMBO J*, 12(1):223–32, 1993.
- [106] T. H. Jessen, C. Oubridge, C. H. Teo, C. Pritchard, and K. Nagai. Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J*, 10(11):3447–56, 1991.
- [107] M. H. Jones and C. Guthrie. Unexpected flexibility in an evolutionarily conserved protein-RNA interaction: genetic analysis of the Sm binding site. *EMBO J*, 9(8):2555–61, 1990.
- [108] L. Joyeux and P. A. Penczek. Efficiency of 2D alignment methods. *Ultramicroscopy*, 92(2):33–46, 2002.
- [109] M. S. Jurica and M. J. Moore. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*, 12(1):5–14, 2003.
- [110] M. S. Jurica, L. J. Licklider, S. R. Gygi, N. Grigorieff, and M. J. Moore. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA*, 8(4):426–39, 2002.

- [111] M. S. Jurica, D. Sousa, M. J. Moore, and N. Grigorieff. Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat Struct Mol Biol*, 11(3):265–9, 2004.
- [112] C. Kambach, S. Walke, R. Young, J. M. Avis, E. de la Fortelle, V. A. Raker, R. Luhrmann, J. Li, and K. Nagai. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, 96(3):375–87, 1999.
- [113] H. Y. Kao and P. G. Siliciano. The yeast homolog of the U1 snRNP protein 70K is encoded by the SNP1 gene. *Nucleic Acids Res*, 20(15):4009–13, 1992.
- [114] H. Y. Kao and P. G. Siliciano. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol Cell Biol*, 16(3):960–7, 1996.
- [115] R. Karaduman, P. Dube, H. Stark, P. Fabrizio, B. Kastner, and R. Luhrmann. Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSm complex as revealed by electron microscopy. *RNA*, 14(12):2528–37, 2008.
- [116] B. Kastner and R. Luhrmann. Electron microscopy of U1 small nuclear ribonucleoprotein particles: shape of the particle and position of the 5' RNA terminus. *EMBO J*, 8(1):277–86, 1989.
- [117] B. Kastner and R. Luhrmann. Purification of U small nuclear ribonucleoprotein particles. *Methods Mol Biol*, 118:289–98, 1999.
- [118] B. Kastner, U. Kornstadt, M. Bach, and R. Luhrmann. Structure of the small nuclear RNP particle U1: identification of the two structural protuberances with RNP-antigens A and 70K. *J Cell Biol*, 116(4):839–49, 1992.
- [119] B. Kastner, N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F. Herzog, J. M. Peters, D. Poerschke, R. Luhrmann, and H. Stark. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods*, 5(1):53–5, 2008.
- [120] D. H. Kim and J. J. Rossi. The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex. *RNA*, 5(7):959–71, 1999.
- [121] R. W. King, J. M. Peters, S. Tugendreich, M. Rolfe, P. Hieter, and M. W. Kirschner. A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. *Cell*, 81(2):279–88, 1995.
- [122] A. L. Kistler and C. Guthrie. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes Dev*, 15(1):42–9, 2001.
- [123] M. M. Konarska and P. A. Sharp. Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell*, 46(6):845–55, 1986.
- [124] M. M. Konarska and P. A. Sharp. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell*, 49(6):763–74, 1987.

- [125] M. M. Konarska and P. A. Sharp. Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. *Proc Natl Acad Sci U S A*, 85(15):5459–62, 1988.
- [126] K. M. Kotovic, D. Lockshon, L. Boric, and K. M. Neugebauer. Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol Cell Biol*, 23(16):5768–79, 2003.
- [127] A. Kramer and U. Utans. Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. *EMBO J*, 10(6):1503–9, 1991.
- [128] T. L. Kress, N. J. Krogan, and C. Guthrie. A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. *Mol Cell*, 32(5):727–34, 2008.
- [129] L. Kretzner, B. C. Rymond, and M. Rosbash. *S. cerevisiae* U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA. *Cell*, 50(4):593–602, 1987.
- [130] L. Kretzner, A. Krol, and M. Rosbash. *Saccharomyces cerevisiae* U1 small nuclear RNA secondary structure contains both universal and yeast-specific domains. *Proc Natl Acad Sci U S A*, 87(2):851–5, 1990.
- [131] A. Krol, E. Westhof, M. Bach, R. Luhrmann, J. P. Ebel, and P. Carbon. Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res*, 18(13):3803–11, 1990.
- [132] S. A. Lacadie and M. Rosbash. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. *Mol Cell*, 19(1):65–75, 2005.
- [133] U. K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259):680–5, 1970.
- [134] B. Lagerbauer, T. Achsel, and R. Luhrmann. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proc Natl Acad Sci U S A*, 95(8):4188–92, 1998.
- [135] A. I. Lamond. The spliceosome. *Bioessays*, 15(9):595–603, 1993.
- [136] C. Le Guiner, F. Lejeune, D. Galiana, L. Kister, R. Breathnach, J. Stevenin, and F. Del Gatto-Konczak. TIA-1 and TIAR activate splicing of alternative exons with weak 5' splice sites followed by a U-rich stretch on their own pre-mRNAs. *J Biol Chem*, 276(44):40638–46, 2001.
- [137] L. Lebart, A. Morineau, and K. M. Warwick. Multivariate descriptive statistical analysis (correspondence analysis and related techniques for large matrices). *Wiley and Sons, Ltd., New York*, 1984.
- [138] R. Leberman. Use of uranyl formate as a negative stain. *J Mol Biol*, 13(2):606, 1965.
- [139] P. Legrain, B. Seraphin, and M. Rosbash. Early commitment of yeast pre-mRNA to the spliceosome pathway. *Mol Cell Biol*, 8(9):3755–60, 1988.

- [140] J. Lepault and J. Dubochet. Electron microscopy of frozen hydrated specimens: preparation and characteristics. *Methods Enzymol*, 127:719–30, 1986.
- [141] A. E. Leschziner and E. Nogales. Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. *Annu Rev Biophys Biomol Struct*, 36:43–62, 2007.
- [142] J. D. Lewis, D. Gorlich, and I. W. Mattaj. A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. *Nucleic Acids Res*, 24(17):3332–6, 1996.
- [143] X. C. Liao, H. V. Colot, Y. Wang, and M. Rosbash. Requirements for U2 snRNP addition to yeast pre-mRNA. *Nucleic Acids Res*, 20(16):4237–45, 1992.
- [144] X. C. Liao, J. Tang, and M. Rosbash. An enhancer screen identifies a gene that encodes the yeast U1 snRNP A protein: implications for snRNP protein function in pre-mRNA splicing. *Genes Dev*, 7(3):419–28, 1993.
- [145] X. L. Liao, L. Kretzner, B. Seraphin, and M. Rosbash. Universally conserved and yeast-specific U1 snRNA sequences are important but not essential for U1 snRNP function. *Genes Dev*, 4(10):1766–74, 1990.
- [146] D. Libri, N. Graziani, C. Saguez, and J. Boulay. Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes Dev*, 15(1):36–41, 2001.
- [147] P. Linder, N. K. Tanner, and J. Banroques. From RNA helicases to RNPases. *Trends Biochem Sci*, 26(6):339–41, 2001.
- [148] Z. Liu, I. Luyten, M. J. Bottomley, A. C. Messias, S. Houngninou-Molango, R. Sprangers, K. Zanier, A. Kramer, and M. Sattler. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science*, 294(5544):1098–102, 2001.
- [149] S. R. Lockhart and B. C. Rymond. Commitment of yeast pre-mRNA to the splicing pathway requires a novel U1 small nuclear ribonucleoprotein polypeptide, Prp39p. *Mol Cell Biol*, 14(6):3623–33, 1994.
- [150] M. C. Lopez, M. Sanchez, E. Ferminan, and A. Dominguez. Disruption of six *Saccharomyces cerevisiae* genes from chromosome IV and basic phenotypic analysis of deletion mutants. *Yeast*, 14(13):1199–208, 1998.
- [151] P. J. Lopez and B. Seraphin. Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition. *RNA*, 5(9):1135–7, 1999.
- [152] J. R. Lorsch and D. Herschlag. The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. *Biochemistry*, 37(8):2194–206, 1998.
- [153] R. Lubig, P. Kusch, K. Roper, and H. Zahn. Zum Reaktionsmechanismus von Glutaraldehyd mit Proteinen. *Monatshefte fur Chemie*, 112:1313–1323, 1981.
- [154] S. J. Ludtke, P. R. Baldwin, and W. Chiu. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol*, 128(1):82–97, 1999.

- [155] R. Luhrmann and H. Stark. Structural mapping of spliceosomes by electron microscopy. *Curr Opin Struct Biol*, 19(1):96–102, 2009.
- [156] M. Luttich. Analytische Methoden zur hochauflösenden Strukturbestimmung in der Kryo-Elektronen-Mikroskopie. *Dissertation, Georg-August-Universität Göttingen*, 2007.
- [157] S. Lybarger, K. Beickman, V. Brown, N. Dembla-Rajpal, K. Morey, R. Seipelt, and B. C. Rymond. Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. *Mol Cell Biol*, 19(1):577–84, 1999.
- [158] E. M. Makarov, O. V. Makarova, H. Urlaub, M. Gentzel, C. L. Will, M. Wilm, and R. Luhrmann. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, 298(5601):2205–8, 2002.
- [159] O. V. Makarova, E. M. Makarov, H. Urlaub, C. L. Will, M. Gentzel, M. Wilm, and R. Luhrmann. A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J*, 23(12):2381–91, 2004.
- [160] A. E. Mayes, L. Verdone, P. Legrain, and J. D. Beggs. Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J*, 18(15):4321–31, 1999.
- [161] S. McCracken, N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, S. Shuman, and D. L. Bentley. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev*, 11(24):3306–18, 1997.
- [162] M. R. McLean and B. C. Rymond. Yeast pre-mRNA splicing requires a pair of U1 snRNP-associated tetratricopeptide repeat proteins. *Mol Cell Biol*, 18(1):353–60, 1998.
- [163] L. Merendino, S. Guth, D. Bilbao, C. Martinez, and J. Valcarcel. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature*, 402(6763):838–41, 1999.
- [164] M. J. Moore and P. A. Sharp. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature*, 365(6444):364–8, 1993.
- [165] D. P. Morris and A. L. Greenleaf. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*, 275(51):39935–43, 2000.
- [166] C. Mura, D. Cascio, M. R. Sawaya, and D. S. Eisenberg. The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. *Proc Natl Acad Sci U S A*, 98(10):5532–7, 2001.
- [167] T. Nakagawa and H. Ogawa. Involvement of the MRE2 gene of yeast in formation of meiosis-specific double-strand breaks and crossover recombination through RNA splicing. *Genes Cells*, 2(1):65–79, 1997.

- [168] R. L. Nelissen, C. L. Will, W. J. van Venrooij, and R. Luhrmann. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J*, 13(17):4113–25, 1994.
- [169] G. Neubauer, A. Gottschalk, P. Fabrizio, B. Seraphin, R. Luhrmann, and M. Mann. Identification of the proteins of the yeast U1 small nuclear ribonucleo-protein complex by mass spectrometry. *Proc Natl Acad Sci U S A*, 94(2):385–90, 1997.
- [170] K. M. Neugebauer, J. A. Stolk, and M. B. Roth. A conserved epitope on a subset of SR proteins defines a larger family of Pre-mRNA splicing factors. *J Cell Biol*, 129(4):899–908, 1995.
- [171] A. Newman. Splicing of pre-mRNA. *Encyclopedia of Life Sciences, John Wiley and Sons, Ltd. www.els.net*, 2005.
- [172] D. Ochs. Protein contaminants of sodium dodecyl sulfate-polyacrylamide gels. *Anal Biochem*, 135(2):470–4, 1983.
- [173] C. L. O’Day, G. Dalbadie-McFarland, and J. Abelson. The *Saccharomyces cerevisiae* Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J Biol Chem*, 271(52):33261–7, 1996.
- [174] H. Ogawa, K. Johzuka, T. Nakagawa, S. H. Leem, and A. H. Hagihara. Functions of the yeast meiotic recombination genes, MRE11 and MRE2. *Adv Biophys*, 31: 67–76, 1995.
- [175] M. D. Ohi, A. J. Link, L. Ren, J. L. Jennings, W. H. McDonald, and K. L. Gould. Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol Cell Biol*, 22(7):2011–24, 2002.
- [176] M. D. Ohi, C. W. Vander Kooi, J. A. Rosenberg, L. Ren, J. P. Hirsch, W. J. Chazin, T. Walz, and K. L. Gould. Structural and functional analysis of essential pre-mRNA splicing factor Prp19p. *Mol Cell Biol*, 25(1):451–60, 2005.
- [177] E. V. Orlova, P. Dube, J. R. Harris, E. Beckman, F. Zemlin, J. Markl, and M. van Heel. Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *J Mol Biol*, 271(3):417–37, 1997.
- [178] C. Oubridge, N. Ito, P. R. Evans, C. H. Teo, and K. Nagai. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, 372(6505):432–8, 1994.
- [179] R. A. Padgett. mRNA Splicing: Role of snRNAs. *Encyclopedia of Life Sciences, Wiley and Sons, Ltd. www.els.net*, 2005.
- [180] R. A. Padgett and C. B. Burge. Splice Sites. *Encyclopedia of Life Sciences, John Wiley and Sons, Ltd. www.els.net*, 2005.

- [181] E. Pascolo and B. Seraphin. The branchpoint residue is recognized during commitment complex formation before being bulged out of the U2 snRNA-pre-mRNA duplex. *Mol Cell Biol*, 17(7):3469–76, 1997.
- [182] A. A. Patel and J. A. Steitz. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*, 4(12):960–70, 2003.
- [183] J. R. Patton and T. Pederson. The Mr 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 5' stem-loop of U1 RNA and interacts with Sm domain proteins. *Proc Natl Acad Sci U S A*, 85(3):747–51, 1988.
- [184] H. Peled-Zehavi, J. A. Berglund, M. Rosbash, and A. D. Frankel. Recognition of RNA branch point sequences by the KH domain of splicing factor 1 (mammalian branch point binding protein) in a splicing factor complex. *Mol Cell Biol*, 21(15):5232–41, 2001.
- [185] P. A. Penczek. Resolution Measures in Single Particle Analysis. Single Particle Reconstructions and Visualization. *EMAN Tutorial, Houston, TX*, 2005.
- [186] P. A. Penczek, R. A. Grassucci, and J. Frank. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy*, 53(3):251–70, 1994.
- [187] D. N. Perkins, D. J. Pappin, D. M. Creasy, and J. S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–67, 1999.
- [188] R. Perriman and J. Ares, M. ATP can be dispensable for prespliceosome formation in yeast. *Genes Dev*, 14(1):97–107, 2000.
- [189] C. W. Pikielny and M. Rosbash. mRNA splicing efficiency in yeast and the contribution of nonconserved sequences. *Cell*, 41(1):119–26, 1985.
- [190] C. W. Pikielny, B. C. Rymond, and M. Rosbash. Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. *Nature*, 324(6095):341–5, 1986.
- [191] D. A. Pomeranz Krummel and K. Nagai. Spliceosome. *Encyclopedia of Life Sciences, John Wiley and Sons, Ltd. www.els.net*, 2005.
- [192] D. A. Pomeranz Krummel, C. Oubridge, A. K. Leung, J. Li, and K. Nagai. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature*, 458(7237):475–80, 2009.
- [193] C. C. Ponting and C. Phillips. Rapsyn's knobs and holes: eight tetratricopeptide repeats. *Biochem J*, 314 (Pt 3):1053–4, 1996.
- [194] O. Puig, A. Gottschalk, P. Fabrizio, and B. Seraphin. Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes Dev*, 13(5):569–80, 1999.
- [195] O. Puig, F. Caspar, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Seraphin. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, 24(3):218–29, 2001.

- [196] O. Puig, E. Bragado-Nilsson, T. Koski, and B. Seraphin. The U1 snRNP-associated factor Luc7p affects 5' splice site selection in yeast and human. *Nucleic Acids Res*, 35(17):5874–85, 2007.
- [197] C. C. Query, R. C. Bentley, and J. D. Keene. A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. *Mol Cell Biol*, 9(11):4872–81, 1989.
- [198] M. Radermacher, T. Wagenknecht, A. Verschoor, and J. Frank. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. *J Microsc*, 146(Pt 2):113–36, 1987.
- [199] P. L. Raghunathan and C. Guthrie. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol*, 8(15):847–55, 1998.
- [200] J. C. Rain, Z. Rafi, Z. Rhani, P. Legrain, and A. Kramer. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA*, 4(5):551–65, 1998.
- [201] V. A. Raker, K. Hartmuth, B. Kastner, and R. Luhrmann. Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol Cell Biol*, 19(10):6554–65, 1999.
- [202] O. Ram and G. Ast. SR proteins: a foot on the exon before the transition from intron to exon definition. *Trends Genet*, 23(1):5–7, 2007.
- [203] J. Rappsilber, U. Ryder, A. I. Lamond, and M. Mann. Large-scale proteomic analysis of the human spliceosome. *Genome Res*, 12(8):1231–45, 2002.
- [204] R. Reed. Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev*, 6(2):215–20, 1996.
- [205] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*, 17(10):1030–2, 1999.
- [206] M. Rosbash and B. Seraphin. Who's on first? The U1 snRNP-5' splice site interaction and splicing. *Trends Biochem Sci*, 16(5):187–90, 1991.
- [207] P. B. Rosenthal and R. Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol*, 333(4):721–45, 2003.
- [208] S. W. Ruby. Dynamics of the U1 small nuclear ribonucleoprotein during yeast spliceosome assembly. *J Biol Chem*, 272(28):17333–41, 1997.
- [209] S. W. Ruby and J. Abelson. An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science*, 242(4881):1028–35, 1988.

- [210] S. W. Ruby and J. Abelson. Pre-mRNA splicing in yeast. *Trends Genet*, 7(3): 79–85, 1991.
- [211] B. Rutz and B. Seraphin. Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA*, 5(6): 819–31, 1999.
- [212] B. C. Rymond and M. Rosbash. Cleavage of 5' splice site and lariat formation are independent of 3' splice site in yeast mRNA splicing. *Nature*, 317(6039):735–7, 1985.
- [213] J. Sambrook, E. F. Fritsch, and T. Maniatis. Molecular cloning - A Laboratory Manual. *Cold Spring Harbor Laboratory Press*, 1989.
- [214] B. Sander. Three-dimensional electron cryo-microscopy of the human small nuclear ribonucleoprotein complexes [U4/U6], U5 and [U4/U5.U5] based upon novel image-processing techniques. *Dissertation*, 2005.
- [215] B. Sander, M. M. Golas, and H. Stark. Corrim-based alignment for improved speed in single-particle image processing. *J Struct Biol*, 143(3):219–28, 2003.
- [216] B. Sander, M. M. Golas, and H. Stark. Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra. *J Struct Biol*, 142(3):392–401, 2003.
- [217] B. Sander, M. M. Golas, and H. Stark. Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy. *J Struct Biol*, 151(1):92–105, 2005.
- [218] B. Sander, M. M. Golas, E. M. Makarov, H. Brahm, B. Kastner, R. Luhrmann, and H. Stark. Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol Cell*, 24(2):267–78, 2006.
- [219] D. G. Sashital, G. Cornilescu, C. J. McManus, D. A. Brow, and S. E. Butcher. U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat Struct Mol Biol*, 11(12):1237–42, 2004.
- [220] W. O. Saxton and W. Baumeister. The correlation averaging of a regularly arranged bacterial cell envelope protein. *J Microsc*, 127(Pt 2):127–38, 1982.
- [221] S. H. Scheres, H. Gao, M. Valle, G. T. Herman, P. P. Eggermont, J. Frank, and J. M. Carazo. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat Methods*, 4(1):27–9, 2007.
- [222] D. Scherly, W. Boelens, W. J. van Venrooij, N. A. Dathan, J. Hamm, and I. W. Mattaj. Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J*, 8(13):4163–70, 1989.
- [223] D. Scherly, W. Boelens, N. A. Dathan, W. J. van Venrooij, and I. W. Mattaj. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B'' and their cognate RNAs. *Nature*, 345(6275):502–6, 1990.

- [224] D. Scherly, C. Kambach, W. Boelens, W. J. van Venrooij, and I. W. Mattaj. Conserved amino acid residues within and outside of the N-terminal ribonucleoprotein motif of U1A small nuclear ribonucleoprotein involved in U1 RNA binding. *J Mol Biol*, 219(4):577–84, 1991.
- [225] M. Schmeisser, B. C. Heisen, M. Luettich, B. Busche, F. Hauer, T. Koske, K. H. Knauber, and H. Stark. Parallel, distributed and GPU computing technologies in single-particle electron microscopy. *Acta Crystallogr D Biol Crystallogr*, 65(Pt 7): 659–71, 2009.
- [226] J. C. Schuette, F. V. t. Murphy, A. C. Kelley, J. R. Weir, J. Giesebrecht, S. R. Connell, J. Loerke, T. Mielke, W. Zhang, P. A. Penczek, V. Ramakrishnan, and C. M. Spahn. GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J*, 28(6):755–65, 2009.
- [227] B. Schwer. A new twist on RNA helicases: DExH/D box proteins as RNPsases. *Nat Struct Biol*, 8(2):113–6, 2001.
- [228] B. Seraphin. Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J*, 14(9): 2089–98, 1995.
- [229] B. Seraphin and M. Rosbash. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell*, 59(2):349–58, 1989.
- [230] B. Seraphin and M. Rosbash. The yeast branchpoint sequence is not required for the formation of a stable U1 snRNA-pre-mRNA complex and is recognized in the absence of U2 snRNA. *EMBO J*, 10(5):1209–16, 1991.
- [231] B. Seraphin, L. Kretzner, and M. Rosbash. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *EMBO J*, 7(8):2533–8, 1988.
- [232] K. W. Shannon and C. Guthrie. Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. *Genes Dev*, 5(5):773–85, 1991.
- [233] A. J. Shatkin. Capping of eucaryotic mRNAs. *Cell*, 9(4 PT 2):645–53, 1976.
- [234] A. J. Shatkin and J. L. Manley. The ends of the affair: capping and polyadenylation. *Nat Struct Biol*, 7(10):838–42, 2000.
- [235] M. A. Sheff and K. S. Thorn. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast*, 21(8):661–70, 2004.
- [236] A. Shevchenko, M. Wilm, O. Vorm, and M. Mann. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem*, 68(5):850–8, 1996.
- [237] F. J. Sigworth. A maximum-likelihood approach to single-particle image refinement. *J Struct Biol*, 122(3):328–39, 1998.
- [238] R. S. Sikorski, M. S. Boguski, M. Goebel, and P. Hieter. A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell*, 60(2):307–17, 1990.

- [239] P. G. Siliciano, M. H. Jones, and C. Guthrie. *Saccharomyces cerevisiae* has a U1-like small nuclear RNA with unexpected properties. *Science*, 237(4821):1484–7, 1987.
- [240] P. G. Siliciano, W. J. Kivens, and C. Guthrie. More than half of yeast U1 snRNA is dispensable for growth. *Nucleic Acids Res*, 19(23):6367–72, 1991.
- [241] R. Singh and R. Reddy. Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc Natl Acad Sci U S A*, 86(21):8280–3, 1989.
- [242] V. Smith and B. G. Barrell. Cloning of a yeast U1 snRNP 70K protein homologue: functional conservation of an RNA-binding domain between humans and yeast. *EMBO J*, 10(9):2627–34, 1991.
- [243] N. Sonenberg. eIF4E, the mRNA cap-binding protein: from basic discovery to translational research. *Biochem Cell Biol*, 86(2):178–83, 2008.
- [244] E. J. Sontheimer, S. Sun, and J. A. Piccirilli. Metal ion catalysis during splicing of premessenger RNA. *Nature*, 388(6644):801–5, 1997.
- [245] C. O. Sorzano, R. Marabini, J. Velazquez-Muriel, J. R. Bilbao-Castro, S. H. Scheres, J. M. Carazo, and A. Pascual-Montano. XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol*, 148(2):194–204, 2004.
- [246] M. Spingola, L. Grate, D. Haussler, and J. Ares, M. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA*, 5(2):221–34, 1999.
- [247] J. P. Staley and C. Guthrie. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, 92(3):315–26, 1998.
- [248] H. Stark and R. Luhrmann. Cryo-electron microscopy of spliceosomal components. *Annu Rev Biophys Biomol Struct*, 35:435–57, 2006.
- [249] H. Stark, P. Dube, R. Luhrmann, and B. Kastner. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, 409(6819):539–42, 2001.
- [250] M. Steinkilberg and H. Schramm. Eine Verbesserte Drehkorrelationsmethode für die Strukturbestimmung biologischer Makromoleküle durch Mittelung elektronenmikroskopischer Bilder. *Z Physiol Chem*, 361:1363–1369, 1980.
- [251] T. A. Steitz and J. A. Steitz. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A*, 90(14):6498–502, 1993.
- [252] S. W. Stevens and J. Abelson. Yeast pre-mRNA splicing: methods, mechanisms, and machinery. *Methods Enzymol*, 351:200–20, 2002.
- [253] S. W. Stevens, D. E. Ryan, H. Y. Ge, R. E. Moore, M. K. Young, T. D. Lee, and J. Abelson. Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell*, 9(1):31–44, 2002.

- [254] E. Tan and H. Kunkel. Characteristics of soluble nuclear antigen precipitating with sera of patients with systemic lupus erythematosus. *Journal of Immunology*, 96(3):464–71, 1966.
- [255] J. Tang and M. Rosbash. Characterization of yeast U1 snRNP A protein: identification of the N-terminal RNA binding domain (RBD) binding site and evidence that the C-terminal RBD functions in splicing. *RNA*, 2(10):1058–70, 1996.
- [256] J. Tang, N. Abovich, M. L. Fleming, B. Seraphin, and M. Rosbash. Identification and characterization of a yeast homolog of U1 snRNP-specific protein C. *EMBO J*, 16(13):4082–91, 1997.
- [257] N. K. Tanner and P. Linder. DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell*, 8(2):251–62, 2001.
- [258] D. F. Tardiff and M. Rosbash. Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. *RNA*, 12(6):968–79, 2006.
- [259] W. Y. Tarn and J. A. Steitz. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci*, 22(4):132–7, 1997.
- [260] W. Y. Tarn, K. R. Lee, and S. C. Cheng. The yeast PRP19 protein is not tightly associated with small nuclear RNAs, but appears to associate with the spliceosome after binding of U2 to the pre-mRNA and prior to formation of the functional spliceosome. *Mol Cell Biol*, 13(3):1883–91, 1993.
- [261] W. Y. Tarn, K. R. Lee, and S. C. Cheng. Yeast precursor mRNA processing protein PRP19 associates with the spliceosome concomitant with or just after dissociation of U4 small nuclear RNA. *Proc Natl Acad Sci U S A*, 90(22):10821–5, 1993.
- [262] W. Y. Tarn, C. H. Hsu, K. T. Huang, H. R. Chen, H. Y. Kao, K. R. Lee, and S. C. Cheng. Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO J*, 13(10):2421–31, 1994.
- [263] P. A. Thuman-Commike. Single particle macromolecular structure determination via electron microscopy. *FEBS Lett*, 505(2):199–205, 2001.
- [264] G. W. Tischendorf, H. Zeichhardt, and G. Stoffler. Determination of the location of proteins L14, L17, L18, L19, L22, L23 on the surface of the 50S ribosomal subunit of *Escherichia coli* by immune electron microscopy. *Mol Gen Genet*, 134(3):187–208, 1974.
- [265] K. D. Toennies. Grundlagen der bildverarbeitung. *Pearson Studium*, 2005.
- [266] W. Y. Tsai, Y. T. Chow, H. R. Chen, K. T. Huang, R. I. Hong, S. P. Jan, N. Y. Kuo, T. Y. Tsao, C. H. Chen, and S. C. Cheng. Cef1p is a component of the Prp19p-associated complex and essential for pre-mRNA splicing. *J Biol Chem*, 274(14):9455–62, 1999.
- [267] K. Tycowski, N. Kolev, N. Conrad, V. Fok, and J. A. Steitz. The ever-growing world of small nuclear ribonucleoproteins, 2006.

- [268] H. Uemura and Y. Jigami. GCR3 encodes an acidic protein that is required for expression of glycolytic genes in *Saccharomyces cerevisiae*. *J Bacteriol*, 174(17):5526–32, 1992.
- [269] A. van Dam, I. Winkel, J. Zijlstra-Baalbergen, R. Smeenk, and H. Cuypers. Cloned human snRNP proteins B and B' differ only in their carboxyl-terminal part. *EMBO J*, 8:3853–3860, 1989.
- [270] M. van Heel. Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy*, 13(1-2):165–83, 1984.
- [271] M. van Heel. Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, 21(2):111–23, 1987.
- [272] M. van Heel and J. Frank. Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, 6(2):187–94, 1981.
- [273] M. van Heel and G. Harauz. Exact filters for general geometry three dimensional reconstruction. *Optik*, 4:17–24, 1986.
- [274] M. van Heel and M. Schatz. Fourier shell correlation threshold criteria. *J Struct Biol*, 151(3):250–62, 2005.
- [275] M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, and M. Schatz. A new generation of the IMAGIC image processing system. *J Struct Biol*, 116(1):17–24, 1996.
- [276] M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, and A. Patwardhan. Single-particle electron cryo-microscopy: towards atomic resolution. *Q Rev Biophys*, 33(4):307–69, 2000.
- [277] R. W. van Nues and J. D. Beggs. Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of *Saccharomyces cerevisiae*. *Genetics*, 157(4):1451–67, 2001.
- [278] V. P. Vidal, L. Verdone, A. E. Mayes, and J. D. Beggs. Characterization of U6 snRNA-protein interactions. *RNA*, 5(11):1470–81, 1999.
- [279] R. M. Vidaver, D. M. Fortner, L. S. Loos-Austin, and D. A. Brow. Multiple functions of *Saccharomyces cerevisiae* splicing protein Prp24 in U6 RNA structural rearrangements. *Genetics*, 153(3):1205–18, 1999.
- [280] E. Villa, J. Sengupta, L. G. Trabuco, J. LeBarron, W. T. Baxter, T. R. Shaikh, R. A. Grassucci, P. Nissen, M. Ehrenberg, K. Schulten, and J. Frank. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc Natl Acad Sci U S A*, 106(4):1063–8, 2009.
- [281] T. Villa, J. A. Pleiss, and C. Guthrie. Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell*, 109(2):149–52, 2002.

- [282] J. D. Wagner, E. Jankowsky, M. Company, A. M. Pyle, and J. N. Abelson. The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO J*, 17(10):2926–37, 1998.
- [283] M. C. Wahl, C. L. Will, and R. Luhrmann. The spliceosome: design principles of a dynamic RNP machine. *Cell*, 136(4):701–18, 2009.
- [284] E. Wahle. mRNA Formation: 3' End. *Encyclopedia of Life Sciences, John Wiley and Sons, Ltd. www.els.net*, 2001.
- [285] H. Y. Wang, W. Lin, J. A. Dyck, J. M. Yeakley, Z. Songyang, L. C. Cantley, and X. D. Fu. SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol*, 140(4):737–50, 1998.
- [286] Y. Wang, J. D. Wagner, and C. Guthrie. The DEAH-box splicing factor Prp16 unwinds RNA duplexes in vitro. *Curr Biol*, 8(8):441–51, 1998.
- [287] D. K. Wiest, C. L. O'Day, and J. Abelson. In vitro studies of the Prp9.Prp11.Prp21 complex indicate a pathway for U2 small nuclear ribonucleoprotein activation. *J Biol Chem*, 271(52):33268–76, 1996.
- [288] C. L. Will and R. Luhrmann. Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol*, 13(3):290–301, 2001.
- [289] C. L. Will and R. Luhrmann. Spliceosome Structure and Function, 2006.
- [290] C. L. Will, C. Schneider, A. M. MacMillan, N. F. Katopodis, G. Neubauer, M. Wilm, R. Luhrmann, and C. C. Query. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J*, 20(16):4536–46, 2001.
- [291] S. Wu, C. M. Romfo, T. W. Nilsen, and M. R. Green. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, 402(6763):832–5, 1999.
- [292] S. H. Xiao and J. L. Manley. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev*, 11(3):334–44, 1997.
- [293] J. Xie, K. Beickman, E. Otte, and B. C. Rymond. Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. *EMBO J*, 17(10):2938–46, 1998.
- [294] S. L. Yean, G. Wuenschell, J. Termini, and R. J. Lin. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature*, 408(6814):881–4, 2000.
- [295] P. D. Zamore and M. R. Green. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A*, 86(23):9243–7, 1989.

- [296] D. Zhang and M. Rosbash. Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev*, 13(5):581–92, 1999.
- [297] D. Zhang, N. Abovich, and M. Rosbash. A biochemical function for the Sm complex. *Mol Cell*, 7(2):319–29, 2001.
- [298] M. Zhang and M. R. Green. Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes Dev*, 15(1):30–5, 2001.
- [299] Z. Zhou, L. J. Licklider, S. P. Gygi, and R. Reed. Comprehensive proteomic analysis of the human spliceosome. *Nature*, 419(6903):182–5, 2002.

Teil I

Appendix

Massenspektroskopische Analyse

Die massenspektroskopische Analyse der Proteinzusammensetzung des isolierten U1 snRNP wurde in Zusammenarbeit mit der Abteilung 'Bioanalytische Massenspektrometrie' unter Leitung von Dr. Henning Urlaub (Max-Planck-Institut für biophysikalische Chemie, Göttingen) durchgeführt.

Das U1 snRNP war für diese Analysen über die Tandem-Affinitätsaufreinigung und Glycerolgradientenzentrifugation wie in Kap. 3, Abschn. 3.2.1 beschrieben, ohne Vorfraktionierung, aufgereinigt worden. Desweiteren ist in der Aufreinigung ein Calmodulinelutionspuffer mit 20 mM EGTA verwendet worden. Für die massenspektroskopischen Analysen wurde der U1 snRNP zum einen unter normalen Salzbedingungen (200 mM KCl im Glycerolgradienten, (Abb. 4) und zum anderen unter Hochsalzbedingungen (400 mM KCl im Glycerolgradienten, Abb. 5) aufgereinigt worden.

Die Proteine aller Fraktionen wurden mittels SDS-PAGE in einem Polyacrylamidgel getrennt und die Proteine wurden anschließend mit Silber gefärbt. Von den Peakfraktionen mit U1 snRNP wurden für die massenspektroskopische Analyse die Proteinbanden einer Spur wie in den Abbildungen 4 und 5 gezeigt, ausgeschnitten, und die Proteine wurden mit der Endoprotease Trypsin verdaut und Gel-extrahiert wie in [236] beschrieben. Die erhaltenen Peptide wurden in einem Flüssig-Chromatographie-gekoppelten Elektrospray-Ionisator, Q-ToF (*quadrupole time of flight*) (Q-ToF Ultima, Waters), analysiert. Die ausgegebenen Daten über die Peptide wurden mit der nichtredundanten Peptidatendbank des NCBI mithilfe des Mascot-Suchalgorithmus (Matrix Science Inc., Boston) [187] verglichen, um schließlich die in der Probe erhaltenen Proteine zu identifizieren. In den Tabellen sind die Proteine, die mit einem Score von ≥ 25 identifiziert wurden und aus *S. cerevisiae* stammen, aufgenommen. Es entfielen demnach Proteine, die unterhalb des methodisch bedingten Schwellenwertes von 25 für eine vertrauenswürdige Erkennung lagen, und Proteine, die in der Datenbanksuche in anderen Spezies gefunden wurden.

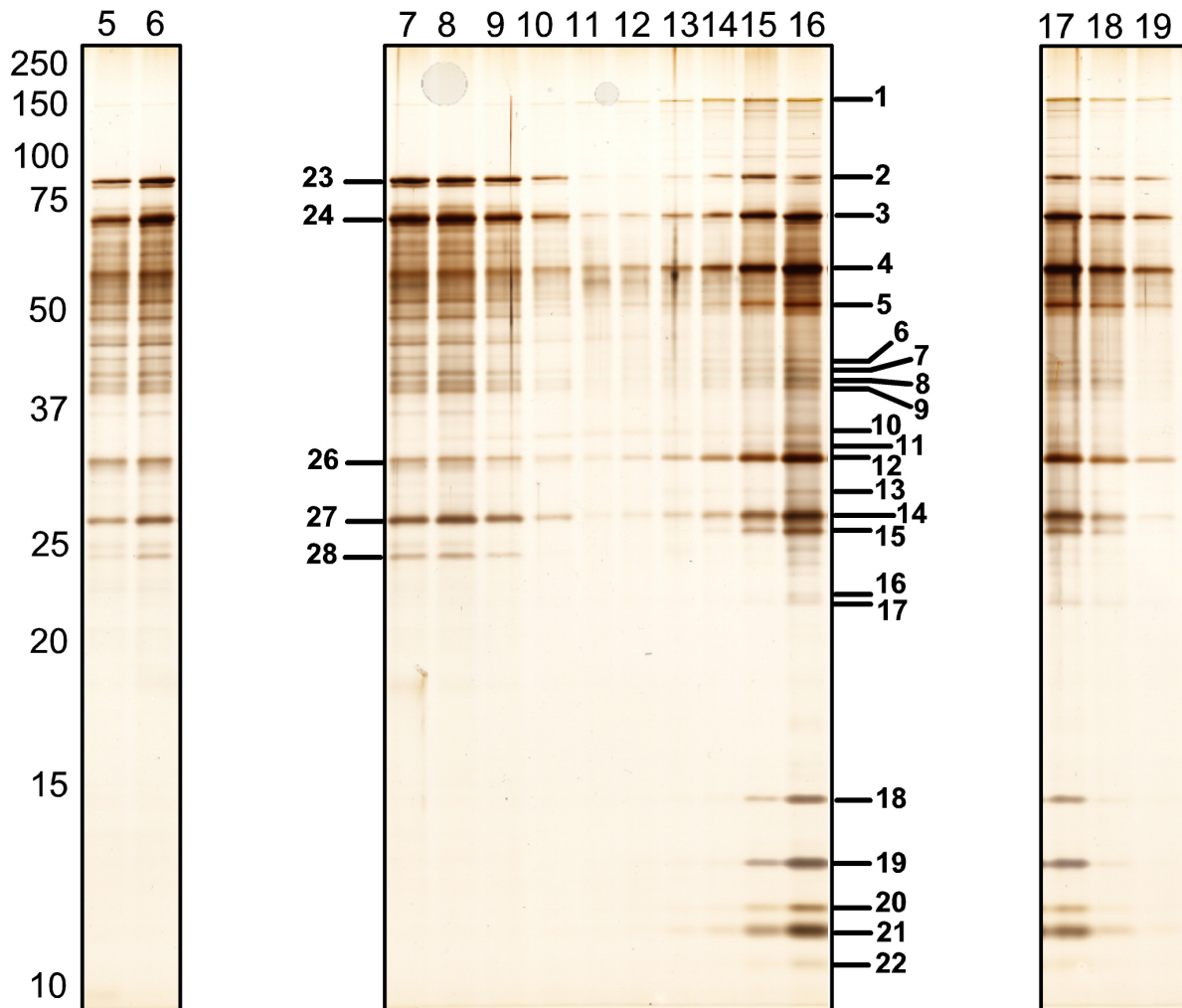


Abbildung 4: Verwendete Proteingelbanden für die massenspektroskopische Analyse der Proteine des U1 snRNP, aufgereinigt bei 200 mM KCl. U1 snRNP war affinitätsaufgereinigt worden, und zwar im wesentlichen wie in Kap. 3, Abschn. 3.2.1 beschrieben, aber ohne Vorfraktionierung. Nach der abschließenden Glycerolgradientenzentrifugation mit 200 mM KCl im Gradienten wurden die Proteine des U1 snRNPs mittels SDS-PAGE in einem 12% Polyacrylamidgel getrennt. Gezeigt ist ein Ausschnitt mit den Fraktionen 5 – 19. Die Proteinbanden der Fraktionen 7 und 16 wurden für die massenspektroskopische Analyse der Proteine durchnummeriert und ausgeschnitten. Die Nummern geben den Bezug für die Proteingelbande in den Tabellen 1 und 2.

Proteingel- bande	Protein	Protein- masse in [kDA]	Genbank- Akzessions- nummer	Anzahl der Peptide	Score
Bande 1	Prp42p	65,1	gi 6320441	2	59
Bande 2	Snu71p	71,3	gi 6321450	11	364
	Snp1p	34,4	gi 6322128	1	50
Bande 3	Prp40p	69,0	gi 6322840	20	746
	Prp39p	74,4	gi 6323596	18	563
	Snu71p	71,3	gi 6321450	8	238
Bande 4	Prp42p	65,1	gi 6320441	27	716
	Nam8p	56,9	gi 4026	5	268
	Snu71p	71,3	gi 6321450	7	235
	Prp40p	69,0	gi 6322840	7	218
Bande 5	Snu56p	56,5	gi 6320447	5	183
	Nam8p	56,9	gi 4026	2	179
	Gcr3	99,5	gi 464222	4	86
	Prp40p	69,0	gi 6322840	3	37
Bande 6	Nam8p	56,9	gi 4026	1	68
	Snp1p	34,4	gi 6322128	1	67
	Prp39p	74,4	gi 6323596	2	63
	Prp40p	69,0	gi 6322840	2	52
	Snu71p	71,3	gi 6321450	1	41
Bande 7	Snu71p	71,3	gi 6321450	4	191
	Nam8p	56,9	gi 4026	2	115
	Prp40p	69,0	gi 6322840	3	63
Bande 8	Prp40p	69,0	gi 6322840	6	199
	Snu71p	71,3	gi 6321450	5	93
	Snp1p	34,4	gi 6322128	1	69
	Gcr3	99,5	gi 464222	2	58
Bande 9	Prp40p	69,0	gi 6322840	9	287
	Snp1p	34,4	gi 6322128	1	48
Bande 10	Snu56p	56,5	gi 6320447	4	173
	Snp1p	34,4	gi 6322128	2	61
Bande 11	Snp1p	34,4	gi 6322128	4	149
	Snu56p	56,5	gi 6320447	4	106
	Prp40p	69,0	gi 6322840	1	31
Bande 12	Mud1p	34,8	gi 476073	8	419
	Snp1p	34,4	gi 6322128	12	338
	Snu71p	71,3	gi 6321450	1	40
	Snu56p	56,5	gi 6320447	1	32
Bande 13	Smb1b	22,4	gi 6320867	2	75
	Prp40p	69,0	gi 6322840	1	45
Bande 14	Luc7p	30,2	gi 6320116	15	621
	Prp40p	69,0	gi 6322840	5	154
	Smb1p	22,4	gi 6320867	4	135
	Yhc1p	27,0	gi 6323329	1	28

Tabelle 1: Teil 1, Massenspektroskopische Analyse der Proteinbanden 1 – 14 aus Abb. 4. In die Liste aufgenommen wurden nur Proteine, die mit einem Score von ≥ 25 identifiziert wurden und aus *S. cerevisiae* stammen.

Proteingel- bande	Protein	Protein- masse in [kDA]	Genbank- Akzessions- nummer	Anzahl der Peptide	Score
Bande 15	Yhc1p	27,0	gi 6323329	6	236
	Smb1p	22,4	gi 6320867	3	109
	Snp1p	34,4	gi 6322128	1	48
Bande 16	Keratin	-	-	-	-
Bande 17	Cbc2p	23,8	gi 6325079	4	104
Bande 18	Smd1p	16,3	gi 6321510	3	75
Bande 19	Smd2p	12,8	gi 6323305	3	66
Bande 20	-	-	-	-	-
Bande 21	Smx3p/ SmF	9,6	gi 6325440	2	104
	Smd3p	11,2	gi 6323176	2	52
Bande 22	Smx2p/ SmG	8,5	gi 14318502	1	58
Bande 23	Snu71p	71,3	gi 6321450	22	661
Bande 24	Prp40p	69,0	gi 6322840	29	1016
	Snu71p	71,3	gi 6321450	9	341
Bande 25	Prp40p	69,0	gi 6322840	14	463
	Snu71p	71,3	gi 6321450	8	252
	Pyruvat- Decarboxylase	60,0	gi 4109	1	29
Bande 26	Snu71p	71,3	gi 6321450	3	80
Bande 27	Luc7p	30,2	gi 6320116	9	337
	Prp40p	69,0	gi 6322840	6	177
Bande 28	Luc7p	30,2	gi 6320116	5	232
	Snu71p	71,3	gi 6321450	2	22

Tabelle 2: Teil 2, Fortsetzung der Tabelle 1: Massenspektroskopische Analyse der Proteinbanden 15 – 28 aus Abb. 4.

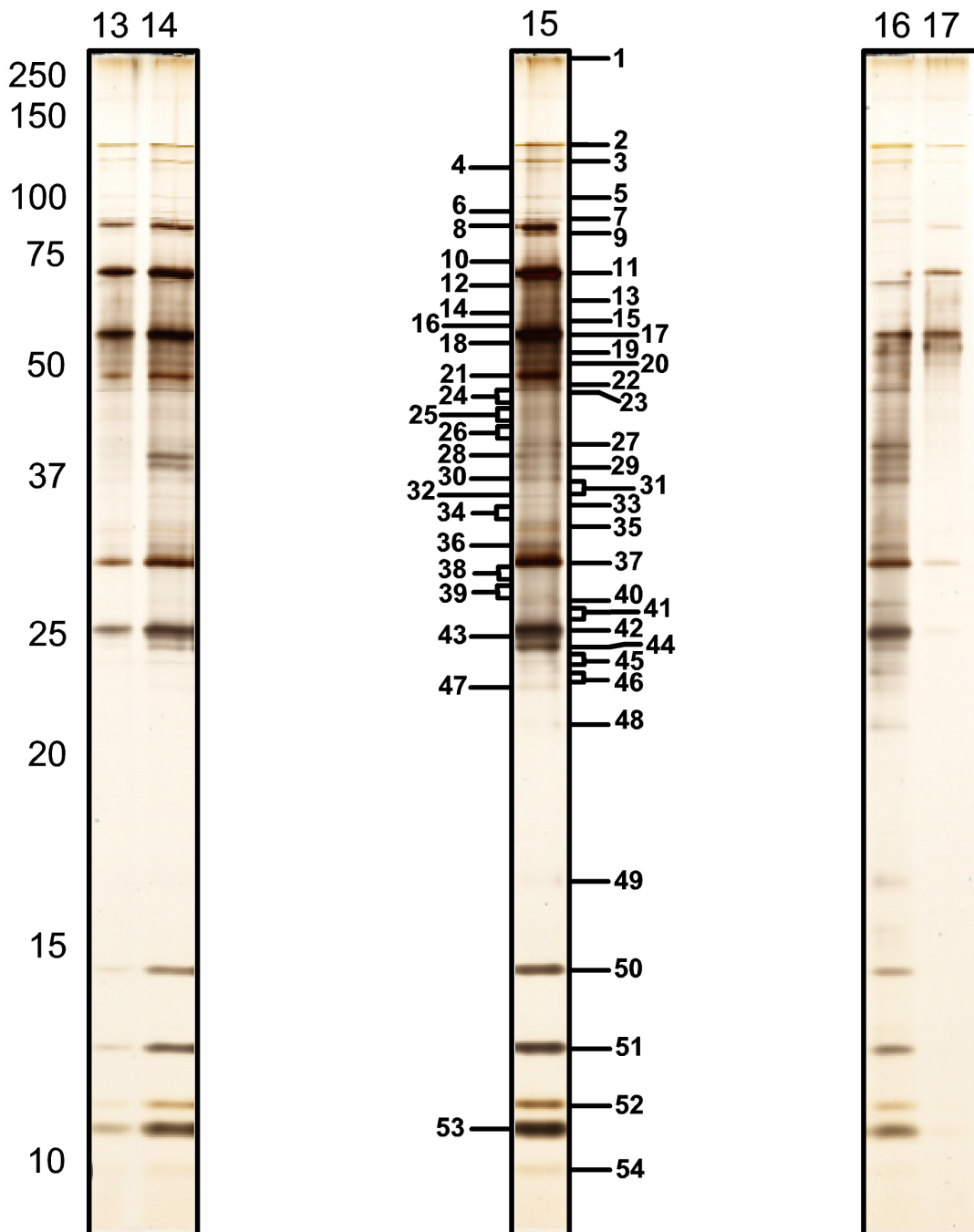


Abbildung 5: Verwendete Proteingelbanden für die massenspektroskopische Analyse der Proteine des U1 snRNP, aufgereinigt bei 400 *mM* KCl. U1 snRNP war affinitätsaufgereinigt worden, und zwar im wesentlichen wie in Kap. 3, Abschn. 3.2.1 beschrieben, aber ohne Vorfraktionierung. Nach der abschließenden Glycerolgradientenzentrifugation mit 400 *mM* KCl im Gradienten wurden die Proteine des U1 snRNP's mittels SDS-PAGE in einem 12% Polyacrylamidgel getrennt. Gezeigt ist ein Ausschnitt mit den Fraktionen 13 – 17. Die Proteinbanden der Fraktion 15 wurden für die folgende massenspektroskopische Analyse der Proteine durchnummeriert und ausgeschnitten. Die Nummern geben den Bezug für die Proteingelbande in den Tabellen 3, 4 und 5.

Proteingel- bande	Protein	Protein- masse in [kDA]	Genbank- Akzessions- nummer	Anzahl der Peptide	Score
Bande 1	-	-	-	-	-
Bande 2	-	-	-	-	-
Bande 3	Snplp	34,4	gi 6322128	1	37
Bande 4	-	-	-	-	-
Bande 5	-	-	-	-	-
Bande 6	-	-	-	-	-
Bande 7	Snu71p	71,3	gi 6321450	3	88
Bande 8	Snu71p	71,3	gi 6321450	32	428
Bande 9	Snu71p	71,3	gi 6321450	8	146
Bande 10	-	-	-	-	-
Bande 11	Prp39p	74,7	gi 6323596	24	728
	Prp40p	69,0	gi 6322840	17	598
	Snu71p	71,3	gi 6321450	4	161
Bande 12	Prp39p	74,7	gi 6323596	7	174
	Prp40p	69,0	gi 6322840	9	135
	Snu71p	71,3	gi 6321450	6	95
Bande 13	Prp39p	74,7	gi 6323596	6	179
	Snu71p	71,3	gi 6321450	4	83
	Nam8p	56,9	gi 4026	1	49
	Prp40p	69,0	gi 6322840	2	31
Bande 14	Snu71p	71,3	gi 6321450	5	73
Bande 15	Snu71p	71,3	gi 6321450	5	73
	Prp42p	65,1	gi 6320441	2	61
Bande 16	Nam8p	56,9	gi 4026	2	130
	Prp42p	65,1	gi 6320441	3	73
Bande 17	Prp42p	65,1	gi 6320441	27	457
	Nam8p	56,9	gi 6321878	7	230
	Prp40p	69,0	gi 6322840	7	187
	Snu71p	71,3	gi 6321450	3	57
	Prp39p	74,7	gi 6323596	4	47
Bande 18	Nam8p	56,9	gi 6321878	2	128
	Prp42p	65,1	gi 6320441	12	124
	Prp40p	69,0	gi 6322840	2	60
	Prp39p	74,7	gi 6323596	1	34
Bande 19	Nam8p	56,9	gi 6321878	3	142
	Prp39p	74,7	gi 6323596	4	62
	Prp42p	65,1	gi 6320441	3	62
	Prp40p	69,0	gi 6322840	3	53
Bande 20	Nam8p	56,9	gi 6321878	2	78
	Prp40p	69,0	gi 6322840	4	72
	Snplp	30,0	gi 6322128	1	30

Tabelle 3: Teil 1, Massenspektroskopische Analyse der Proteinbanden 1 – 20 aus Abb. 5. In die Liste aufgenommen wurden nur Proteine, die mit einem Score von ≥ 25 identifiziert wurden und aus *S. cerevisiae* stammen.

Proteingel- bande	Protein	Protein- masse in [kDA]	Genbank- Akzessions- nummer	Anzahl der Peptide	Score
Bande 21	Snu56p	56,5	gi 6320447	11	288
	Nam8p	56,9	gi 6321878	1	100
Bande 22	Snu56p	gi 6320447	8	196	
	Prp40p	69,0	gi 6322840	2	53
Bande 23	Prp40p	69,0	gi 6322840	3	56
Bande 24	Keratin	-	-	-	
Bande 25	Nam8p	56,9	gi 6321878	1	79
	Snplp	34,4	gi 6322128	2	68
Bande 26	Nam8p	56,9	gi 6321878	1	70
	Snplp	34,4	gi 6322128	1	59
Bande 27	Snplp	34,4	gi 6322128	1	65
	Nam8p	56,9	gi 4026	1	45
Bande 28	Nam8p	56,9	gi 4026	1	68
	Snplp	34,4	gi 6322128	1	35
Bande 29	Snplp	34,4	gi 6322128	2	74
	Snu71p	71,3	gi 6321450	2	43
	Prp40p	69,0	gi 6322840	2	42
Bande 30	Snplp	34,4	gi 6322128	1	31
Bande 31	Snplp	34,4	gi 6322128	2	85
Bande 32	Snplp	34,4	gi 6322128	2	79
Bande 33	Snplp	34,4	gi 6322128	2	100
Bande 34	Snplp	34,4	gi 6322128	2	84
Bande 35	-	-	-	-	-
Bande 36	Snplp	34,4	gi 6322128	2	88
Bande 37	Mud1p	34,4	gi 6319595	15	610
	Snplp	34,4	gi 6322128	11	278
	Smb1p	22,4	gi 6320867	4	81
	Yhc1p	27,0	gi 6323329	2	51
	Axl1p	13,8	gi 6325379	1	27
Bande 38	Mud1p	34,4	gi 6319595	2	57
	Smb1p	22,4	gi 6320867	2	55
Bande 39	Snplp	34,4	gi 6322128	2	84
	Smb1p	22,4	gi 6320867	3	54
	Mud1p	34,4	gi 6319595	1	46
Bande 40	Snplp	34,4	gi 6322128	3	90
	Smb1p	22,4	gi 6320867	2	51
	Yhc1p	27,0	gi 6323329	3	29
Bande 41	Luc7p	30,1	gi 6320116	2	91
	Snplp	34,4	gi 6322128	1	55
	Smb1p	22,4	gi 6320867	3	41
	Yhc1p	27,0	gi 6323329	3	36

Tabelle 4: Teil 2, Fortsetzung der Tabelle 3: Massenspektroskopische Analyse der Proteinbanden 21 – 41 aus Abb. 5.

Proteingel- bande	Protein	Protein- masse in [kDA]	Genbank- Akzessions- nummer	Anzahl der Peptide	Score
Bande 42	Luc7p	30,1	gi 6320116	13	433
	Yhc1p	27,0	gi 6323329	3	94
	Smb1p	22,4	gi 6320867	3	91
	Snplp	34,4	gi 6322128	1	44
Bande 43	Luc7p	30,1	gi 6320116	13	433
	Yhc1p	27,0	gi 6323329	3	72
	Snplp	34,4	gi 6322128	1	69
Bande 44	Yhc1p	27,0	gi 6323329	3	72
	Luc7p	30,1	gi 6320116	13	433
	Snplp	34,4	gi 6322128	2	54
	Smb1p	22,4	gi 6320867	2	34
Bande 45	-	-	-	-	-
Bande 46	-	-	-	-	-
Bande 47	-	-	-	-	-
Bande 48	-	-	-	-	-
Bande 49	-	-	-	-	-
Bande 50	Smd1p	16,3	gi 6321510	6	127
Bande 51	Smd2p	12,8	gi 6323305	10	109
Bande 52	Sme1p	10,4	gi 6324733	1	43
Bande 53	Smd3p	11,2	gi 6324733	4	119
	Smx3p/ SmF	9,7	gi 6325440	2	81
	Smd2p	12,8	gi 6323305	10	109
Bande 54	Smx2p/ SmG	8,5	gi 14318502	1	35

Tabelle 5: Teil 3, Fortsetzung der Tabelle 3: Massenspektroskopische Analyse der Proteinbanden 42 – 54 aus Abb. 5.

Erklärung

Hiermit erkläre ich, dass ich die Dissertation mit dem Titel 'Dreidimensionale elektronenmikroskopische Strukturanalyse des Hefe U1 snRNP' selbständig verfasst und die benutzten Hilfsmittel und Quellen sowie die zu Hilfeleistungen herangezogenen Institutionen vollständig angegeben habe. Diese Dissertation wurde in der Vergangenheit nicht, auch nicht in Teilen, als Master-, Diplom- oder sonstige Prüfungsarbeit verwendet.

Göttingen, den 18.11.2009

Ilonka Bartoszek

Persönliche Daten

Persönliche Daten

Name: Bartoszek
Vorname: Ilonka
Geburtstag: 30. März 1978
Geburtsort: Lübeck
Staatsangehörigkeit: deutsch

Publikationen

2008 Hassenklöver,T., Kurtanska,S., Bartoszek,I., Junek,S., Schild,D.,
Manzini,I., (2008) Nucleotide-induced Ca^{2+} signaling in sustentacu-
lar supporting cells of the olfactory epithelium. *Glia* 56(15):1614-24.