

Strukturelle und funktionelle Analyse des mRNA-3'-Prozessierungs-Komplexes

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

(Dr. rer. nat.)

der Naturwissenschaftlichen Fakultät I – Biowissenschaften –

der Martin-Luther-Universität

Halle-Wittenberg,

vorgelegt von

Herrn **Christian Tüting**

geb. am 26.03.1989 in Hannover

Gutachter:

Prof. Dr. Elmar Wahle

Prof. Dr. Andrea Sinz

Prof. Dr. Henning Urlaub

eingereicht am: 23.01.2020

verteidigt am: 14.10.2020

Inhaltsverzeichnis

Inhaltsverzeichnis	iii
Abbildungsverzeichnis	vi
Tabellenverzeichnis	viii
Formelverzeichnis	viii
Abkürzungsverzeichnis	ix
1. Einleitung	1
1.1 Prä-mRNA Reifung	1
1.1.1 Capping	2
1.1.2 Spleißen	2
1.1.3 Spaltung und Polyadenylierung	3
1.2 Struktur der bekannten Faktoren der Polyadenylierungsreaktion	5
1.2.1 CFI	5
1.2.2 CFII	7
1.2.3 CstF	9
1.2.4 CPSF	12
1.2.5 Poly(A)-Polymerase und Poly(A)-bindendes Protein PABPN1	16
1.3 Struktur von Biomolekülen	18
1.3.1 <i>Cross-linking</i> und Massenspektrometrie (XL-MS)	19
1.4 Zielstellung dieser Arbeit	20
2. Material und Methoden	22
2.1 Material	22
2.1.1 Chemikalien	22
2.1.2 Geräte	22
2.1.3 Zelllinien und Kulturmedien	23
2.1.4 DNA-Konstrukte und Baculoviren	23
2.1.5 Enzyme, Proteine und Zellextrakte	25
2.1.6 Software	26
2.2 Methoden	27
2.2.1 Proteinbiochemische Methoden	28

2.2.2 Nukleinsäurebiochemische Methoden	32
2.2.3 Aktivitätsanalyse von Proteinen	32
2.2.4 Chemisches <i>cross-linking</i>	34
2.2.5 Massenspektrometrische Analysen	36
2.2.6 Computergestützte Proteinanalyse und Strukturvorhersage	38
2.2.7 Interaktomanalyse	42
3 Ergebnisse	44
3.1 Strukturelle Analyse der bekannten Polyadenylierungskomplexe	44
3.1.1 Strukturanalyse des Spaltungsfaktor II (CFII)	44
3.1.2 Spaltungs- und Polyadenylierungs-Spezifitätsfaktor (CPSF)	53
3.1.3 Strukturanalyse des Spaltungsstimulations-Faktors (CstF)	61
3.1.4 Strukturanalyse des Spaltungsfaktors I (CFI)	65
3.2 Computergestützte Modellierung der Polyadenylierungsfaktoren	68
3.2.1 Generierung der Entfernungsbegrenzungen	68
3.2.2 Modellierung der Subkomplexe der Polyadenylierung	73
3.3 Identifikation weiterer Faktoren des Polyadenylierungskomplexes	97
3.3.1 Interaktom von CFII	98
3.3.2 Interaktom von CstF	103
3.3.3 Interaktom von CFI	105
3.3.4 Interaktom des Endonuklease-Komplexes	106
3.3.5 Basale Faktoren der Polyadenylierung	108
3.3.6 Weitere identifizierte Proteine	109
4 Diskussion	112
4.1 Proteinexpression und -reinigung und Aktivitätsanalyse	112
4.1.1 Darstellung und Aktivitätsanalyse des CFII Komplexes	113
4.1.2 Darstellung des CPSF-Komplexes	118
4.1.3 Darstellung des CstF-Komplexes und Aktivitätsanalyse	125
4.1.4 Darstellung des CFI-Komplexes und Aktivitätsanalyse	126
4.2 Computergestützte Modellierung der Polyadenylierungsfaktoren	128

4.2.1 Generierung der Entfernungsbeschränkungen	128
4.2.2 Modellierung der Subkomplexe	130
4.2.3 Zusammenfassung der Modellierung	139
4.3 Interaktomanalyse	141
4.3.1 Interaktom des CFII-Komplexes	141
4.3.2 Interaktom des CstF-Komplexes	143
4.3.3 Interaktom des CFI-Komplexes	144
4.3.4 Interaktom des mPSF-Komplexes	144
4.3.5 Interaktom des Endonuklease-Komplexes	145
4.3.6 Basale Faktoren der Polyadenylierung	145
4.3.7 Weitere putative Polyadenylierungsfaktoren	146
4.3.8 Zusammenfassung der Interaktomanalyse	149
5 Zusammenfassung	150
5.1 Summary	153
6 Literaturverzeichnis	156
7. Anhang	175
7.1 Abbildungen	175
7.2 Tabellen	178
7.2.1 Material	178
7.2.2 Massenspektrometrische Methoden	181
7.2.3 Identifizierte <i>cross-links</i>	184
7.2.4 Entfernungsbeschränkungen	202
7.2.5 Sonstige Tabellen	207
7.2 Quellcode	208
7.3 Hinterlegung der Rohdaten	211
Lebenslauf	xiii
Publikationsliste	xiv
Eidesstattliche Erklärung	xv
Danksagung	xvi

Abbildungsverzeichnis

Abbildung 1.1 Schematische Übersicht der Genexpression und mRNA-Reifung.	1
Abbildung 1.2 Schematische Übersicht des 3'-UTR einer mRNA und der daran gebundenen, an der 3'-Prozessierung beteiligten Proteine.	4
Abbildung 1.3 Zusammenfassung der bekannten strukturellen Informationen über den CFI-Komplex.	6
Abbildung 1.4 Zusammenfassung der strukturellen Informationen zum CFII-Komplex.	8
Abbildung 1.5 Zusammenfassung der strukturellen Informationen zum CstF-Komplex.	10
Abbildung 1.6 Zusammenfassung der strukturellen Informationen zum mPSF-Komplex.	13
Abbildung 1.7 Zusammenfassung der bekannten Strukturelemente des Endonuklease-Komplexes.	15
Abbildung 1.8 Struktur der Poly(A)-Polymerase und des Poly(A)-bindenden Proteins PABPN1.	17
Abbildung 1.9 Übersicht der Auflösung strukturgebender Methoden.	18
Abbildung 1.10 Struktur und Reaktionsschema der verwendeten <i>cross-linker</i> .	19
Abbildung 3.1 Reinigung und Aktivitätsanalyse des CFII-Komplexes.	45
Abbildung 3.2 Chemisches <i>cross-linking</i> des CFII-Komplexes.	46
Abbildung 3.3 Tryptischer Verdau und Charakterisierung verkürzter CFII Komplexe.	48
Abbildung 3.4 Analyse der RNA-Bindungen von CFII-Deletionsvarianten.	49
Abbildung 3.5 Reinigung zweier Varianten der Zinkfingerdomäne von Pcf11 und Analyse der Sequenzspezifität.	50
Abbildung 3.6 Sequenzangleichung der humanen FEGP- Wiederholungen.	51
Abbildung 3.7 Rekonstitution der Spaltung der Substrat-RNA durch CFII.	52
Abbildung 3.8 <i>Western-Blot</i> -Analyse der zweistufigen Reinigung von CPSF.	53
Abbildung 3.9 Reinigung und Charakterisierung von mPSF.	54
Abbildung 3.10 Validierung der <i>cross-link</i> -Bedingungen für die Polyadenylierungsreaktion.	55
Abbildung 3.11 Präparation und Analyse des chemischen <i>cross-linking</i> der Polyadenylierungsreaktion.	57
Abbildung 3.12 Expression, Reinigung und Aktivitätsanalyse des Endonuklease-Komplexes.	58
Abbildung 3.13 Chemisches <i>cross-linking</i> des Endonuklease-Komplexes.	59
Abbildung 3.14 Reinigung und Charakterisierung des CstF-Komplexes.	61
Abbildung 3.15 Chemisches <i>cross-linking</i> des CstF-Komplexes.	63
Abbildung 3.16 Reinigung und Aktivitätsanalyse des CFI-Komplexes.	65
Abbildung 3.17 Chemisches <i>cross-linking</i> des CFI-Komplexes.	67
Abbildung 3.18 Empirische Bestimmung der Länge der Aminosäureseitenketten.	68
Abbildung 3.19 Boxplot der gemessenen Entfernungen von Di- bis Decapeptid.	70
Abbildung 3.20 Homologiemodel des humanen Clp1.	74
Abbildung 3.21 Übersicht der generierten Homologiemodelle von Pcf11.	75
Abbildung 3.22 Zusammenfassung der Modellierung des Interaktionspeptides Pcf11 ₁₃₇₇₋₁₄₀₁ .	76
Abbildung 3.23 Strukturmodelle des dimeren <i>Dockings</i> von Clp1* mit Pcf11-Substrukturen.	78
Abbildung 3.24 Zusammenfassung des multimeren <i>dockings</i> mit Clp1* und drei Pcf11-Strukturen.	79
Abbildung 3.25 Finales Strukturmodell von CFII.	80
Abbildung 3.26 Analyse des Konformationsraumes von Pcf11 ₁₄₁₉₋₁₄₈₇ in Relation zum vorher gebildeten CFII-Komplex.	80

Abbildung 3.27 Modellierung der Dimerisierungsdomäne von CstF50.	82
Abbildung 3.28 Darstellung der Modelle für CstF64.	83
Abbildung 3.29 Homologiemodell des humanen CstF77-Dimers.	84
Abbildung 3.30 Darstellung der Komplexstruktur von CstF50 und CstF64.	85
Abbildung 3.31 Strukturmodell des CstF50-CstF77 Dimers.	86
Abbildung 3.32 Modell der trimeren Struktur von CstF.	87
Abbildung 3.33 Homologiemodell der Metallo- β -Lactamase und β -CASP-Domäne von CPSF100.	89
Abbildung 3.34 Homologiemodell des Dimers der C-terminalen Domänen von CPSF73 und CPSF100.	90
Abbildung 3.35 Strukturen der HEAT-Domänen von Symplekin.	91
Abbildung 3.36 Zusammenfassung der Strukturmodellierung der Interaktion von Symplekin und CstF64.	92
Abbildung 3.37 Strukturmodelle des <i>Dockings</i> des Endonuklease-Komplexes.	93
Abbildung 3.38 Strukturmodell des Endonuklease-Komplexes.	94
Abbildung 3.39 Finales Strukturmodell des Endonuklease-Komplexes.	95
Abbildung 3.40 Analyse des Konformationsraumes der N-terminalen HEAT-Domäne von Symplekin im Endonuklease-Komplex.	96
Abbildung 3.41 Analyse des Interaktoms der verkürzten CFII-Komplexe.	99
Abbildung 3.42 Quantitative Analyse der identifizierten Proteine in der spaltungs-aktiven gegenüber der spaltungs-inaktiven CFII-Variante.	100
Abbildung 3.43 Interaktomanalyse des CstF-Komplexes.	103
Abbildung 3.44 Interaktomanalyse des CFI-Komplexes.	105
Abbildung 3.45 Interaktomanalyse des Endonuklease-Komplexes.	107
Abbildung 4.1 Schematische Darstellung des hexameren CstF-Komplexes.	136

Tabellenverzeichnis

Tabelle 1 Salzkonzentrationen der verschiedenen Proteinkomplexe für die Anionenaustauschchromatographie	29
Tabelle 2 Zusammenfassung der Zusammensetzung der Spaltreaktion	33
Tabelle 3 Detaillierte Zusammensetzung des präparativen <i>cross-linkings</i> der Polyadenylierungsreaktion	35
Tabelle 4 Übersicht der verwendeten Materialien zur Affinitätsreinigung der Köderproteine.	43
Tabelle 5 Bindungsenergien der verschiedenen CFII-Varianten.	49
Tabelle 6 Detailanalyse der Verteilung der Entfernungen von Di- bis Pentapeptiden.	71
Tabelle 7 Übersicht der neu definierten Entfernungsbeschränkungen.	72
Tabelle 8 Zusammenfassung des paarweisen Sequenzalignments von Clp1 zu den homologen in Hefe (<i>yClp1</i>) und <i>C. elegans</i> (<i>ceClp1</i>).	73
Tabelle 9 Übersicht der Sequenzbereiche in Pcf11 und die für die Modellierung verwendete Vorlagestruktur.	76
Tabelle 10 Übersicht der in der aktiven CFII-Variante angereicherten Proteine.	101
Tabelle 11 Übersicht der bekannten und identifizierten Polyadenylierungsfaktoren in Kontroll- und gemitteltem CFII- <i>Pulldown</i> .	102
Tabelle 12 Identifizierte Polyadenylierungsfaktoren in dem CstF- <i>Pulldown</i> .	104
Tabelle 13 Identifizierte Interaktionspartner von CF I im <i>Pulldown</i> -Experiment.	106
Tabelle 14 Übersicht der identifizierten Polyadenylierungsfaktoren in dem <i>Pulldown</i> -Experiment des Endonuklease-Komplexes.	107
Tabelle 15 Klassifizierung der in allen Datensätzen angereicherten Proteine, die dem RNA-Metabolismus zugeordnet werden konnten.	109
Tabelle 16 Vereinfachte Darstellung der in allen <i>Pulldowns</i> identifizierte Proteine	109
Tabelle 17 Weitere identifizierte Proteine in den <i>pulldown</i> -Experimenten.	109

Formelverzeichnis

Formel (1): Berechnung der freien Bindungsenergie.	49
Formel (2): Berechnung der Entfernung zweier Punkt im karthesischen Koordinatensystem.	69
Formel (3): Berechnung der iBAQ-Werte.	97
Formel (4): Berechnung der iBAQ-Klassen.	97
Formel (5): Berechnung der Anzahl der Proteine der jeweiligen iBAQ-Klasse.	97

Abkürzungsverzeichnis

Abkürzung	Bedeutung
Å	Ångström (1 Å = 0,1 nm)
(v/v)	Volumen pro Volumen
(w/v)	Masse pro Volumen
(w/w)	Masse pro Masse
A	Adenin
ADP	Adenosindiphosphat
AEBSF	4-(2-Aminoethyl)benzensulfonylfluorid
APS	Ammoniumperoxosulfat
ATP	Adenosintriphosphat
ATP	Adenosintriphosphat-
BLASTp	<i>Basic local alignment search tool for proteins</i>
BSA	Bovines Serumalbumin
BSA ^{met}	Methyliertes bovines Serumalbumin
β-CASP-Domäne	Spezifische Domäne der β-CASP-Familie, ein Akronym der Namen der repräsentativen Proteine dieser Familie (CPSF, Artemis, SNM1, PSO2)
C	Cytosin
<i>C. elegans</i>	<i>Caenorhabditis elegans</i> – Fadenwurm
CFI	Spaltfaktor I (<i>Cleavage Factor I</i>)
CFIA	Spaltfaktor IA (<i>cleavage factor IA</i>) aus Hefe
CFII	Spaltfaktor II (<i>Cleavage Factor II</i>)
CID	CTD (der RNA-Polymerase II)-Interaktions-Domäne
CPSF	Spaltungs- und Polyadenylierungsspezifitäts Faktor (<i>cleavage and polyadenylation specificity factor</i>)
CstF	Spaltung-stimulierender Faktor (<i>cleavage stimulatory factor</i>)
CTD	C-terminale Domäne
CV	Säulenvolumina
Cα	Kohlenstoffatom in Aminosäuren, an dem die Amino- und die Carboxylgruppe sind
Da	Dalton – Atomare Masseneinheit
DEPC	Diethyldicarbonat
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
DSBU	Disuccinimidyl-Dibutyrat-Harnstoff
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamintetraessigsäure
EF-Hand	Spezifisches Proteinstrukturmotiv

eIF4E	Eukaryotischer translations-initiation Faktor 4E
FDR	Falsch-positive-Rate (<i>false discovery rate</i>)
FEGP-Wiederholungen	Aminosäuresequenzwiederholungen
FLAG	Aminosäuresequenz <i>DYKDDDDK</i>
FLAG-Agarose	Agarosepartikel mit FLAG-Peptid
FLAG-Peptid	Peptid mit der Aminosäuresequenz <i>DYKDDDDK</i>
FPOP	<i>Fast photochemical oxidation of proteins</i>
fs	Femtosekunden
G	Guanin
GO terms	<i>Gene Ontology</i> Klassen
GSH-Sepharose	Glutathion gekoppelt an Sepharosepartikel
GST	Glutathion-S-Transferase
GTP	Guanosintriphosphat
HADDOCK	High Ambiguity Driven protein-protein DOCKing
HAT-Domäne	<i>Half a TPR</i> -Domäne
HCD	Kollisionsinduzierte Dissoziation höherer Energien (<i>Higher-energy collisional dissociation</i>)
HCl	Salzsäure
HEAT-Wiederholungen	Proteinmotiv. HEAT ist ein Akronym von Proteinen, in denen dieses Motiv identifiziert wurde (Huntingtin, EF3, PP2A, TOR1)
HeLa-Zellen	Immortalisierte Epithelzellen eines Zervixkarzinoms der Henrietta Lacks
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure
HHPred	
HMM	Verdecktes Markowmodell (<i>hidden markov model</i>)
HPLC	<i>High performance liquid chromatography</i>
iBAQ	<i>Intensity-based absolute quantification</i>
IPTG	Isopropyl- β -D-thiogalactopyranosid
IR	Interaktionsregion
K	Kelvin
K ₅₀	Substratkonzentration, bei der 50 % Sättigung erreicht ist
KCl	Kaliumchlorid
K _D	Dissoziationskonstante
kDa	Kilo-Dalton
KOH	Kaliumhydroxid
kryoEM	Kryo-Elektronenmikroskopie
kryoET	Kryo-Elektronentomographie
LB-Medium	Kulturmedium (LB = <i>lysogeny broth</i>)
LC-SDA	Succinimidyl-6-(4,4'-Azipentan-Amido)-Hexanoat
LFQ	Markierungsfreie Quantifizierung (<i>label free quantification</i>)

<i>m/z</i>	Masse-zu-Ladung-Verhältnis
MD	Molekulardynamik
min	Minuten
MLU	Martin-Luther-Universität
mPSF	Polyadenylierungs-Spezifitäts-Faktor
mRNA	Boten-RNA (<i>messenger RNA</i>)
mRNP	Boten-Ribonukleoproteine
MS	Massenspektrometrie
MS-Spektrum	Massenspektrum
MS/MS	Tandem-Massenspektrometrie
MS/MS-spaltbar	Struktur, die während der MS-Analyse durch Energiezufuhr spezifisch fragmentiert
MW	Molekulargewicht
MβL	Metallo-β-Lactamase
NHS	N-Hydroxysuccinimid
NiNTA-Agarose	Agarose mit kovalent verknüpfter Nickel-Nitriloessigsäure
NMR	Kernspinresonanz (<i>nuclear magnetic resonance</i>)
NP-40	Nonidet™ P-40
nt	Nukleotide
NTA	Nitriloessigsäure
NTC-Komplex	<i>NineTeen complex</i>
NTD	N-terminale Domäne
NTR-Komplex	<i>NTC-related complex</i>
NXT	Kernextrakt
<i>p.a.</i>	<i>por analysi</i>
PABPC1	Zytosolisches Poly(A)-Bindepotein
PABPN1	Nukleäres Poly(A)-Bindepotein
PAGE	Polyacrylamid Gelelektrophorese
PAP1	Poly(A)-Polymerase aus Hefe
PAPα, PAPβ und PAPγ	Kanonische Poly(A)-Polymerasen
PAS	Polyadenylierungssignal
PDB	Protein-Datenbank
PDBID	Identifikationsnummer der Protein-Datenbank
PEG	Polyethylenglycol
PEG6000	Polyethylenglycol mit einer mittleren Molekülmasse von 6000 g/mol
Poly(A)	Polyadenosin
Poly(C)	Polycytosin
Poly(G)	Polyguanin
Poly(U)	Polyuracil

<i>ppm</i>	<i>Parts per million</i>
Prä-mRNA	mRNA-Vorläufer
PSM	<i>Peptide spectrum match</i>
PVDF	Polyvinylidenfluorid
RMSD	<i>root-mean square deviation</i>
RNA	Ribonukleinsäure
RRM	<i>RNA recognition motif</i>
rRNA	Ribosomale-RNA
SAXS	<i>Small-angle X-ray scattering</i>
SDS	Natriumdodecylsulfat
Sf-21 Zellen	Zelllinie des Herbst-Heerwurms (<i>Spodoptera frugiperda</i>)
snRNA	Uridin-reiche nukleare RNAs
SUMO	<i>Small ubiquitin-like modifier</i>
TB-Medium	Kulturmedium (TB = <i>terrific broth</i>)
TBE-Puffer	TRIS-Borat-EDTA-Puffer,
TBST	<i>Tris-buffered saline with Tween20</i>
TEA	Triethanolamin
Temed	N,N,N',N'-Tetramethylethylendiamin
TFA	Trifluoressigsäure
TRIS	Tris(hydroxymethyl)aminoethan
tRNA	Transfer-RNA
U	Uracil
UTR	Untranslatierte Region
UV	Ultraviolett
WT	Wildtyp
x-ray	Röntgenkristallographie
xg	g-Kraft
XL-MS	<i>Cross-linking</i> in Kombination mit Massenspektrometrie
ZK	Zinkknöchel-Domäne

1. Einleitung

1.1 Prä-mRNA Reifung

mRNA wird in 5' – 3' Orientierung als Vorläufermolekül, prä-mRNA, von der RNA-Polymerase II synthetisiert. Die prä-mRNA wird kotranskriptionell *gecappt*, gespleißt sowie am 3'-Ende gespalten und polyadenyliert. Diese Schritte werden im Folgenden näher beschrieben. Anschließend wird die so gereifte mRNA in Partikel, sogenannte mRNPs (*messenger-Ribonucleoprotein*) verpackt und in das Zytoplasma exportiert [1], [2]. Diese kotranskriptionellen Prozesse sind über Interaktionen mit der C-terminale Domäne der großen Untereinheit der RNA-Polymerase eng mit der Transkription der mRNA gekoppelt [3]. Eine Übersicht der Prozesse ist in Abbildung 1.1 schematisch dargestellt.

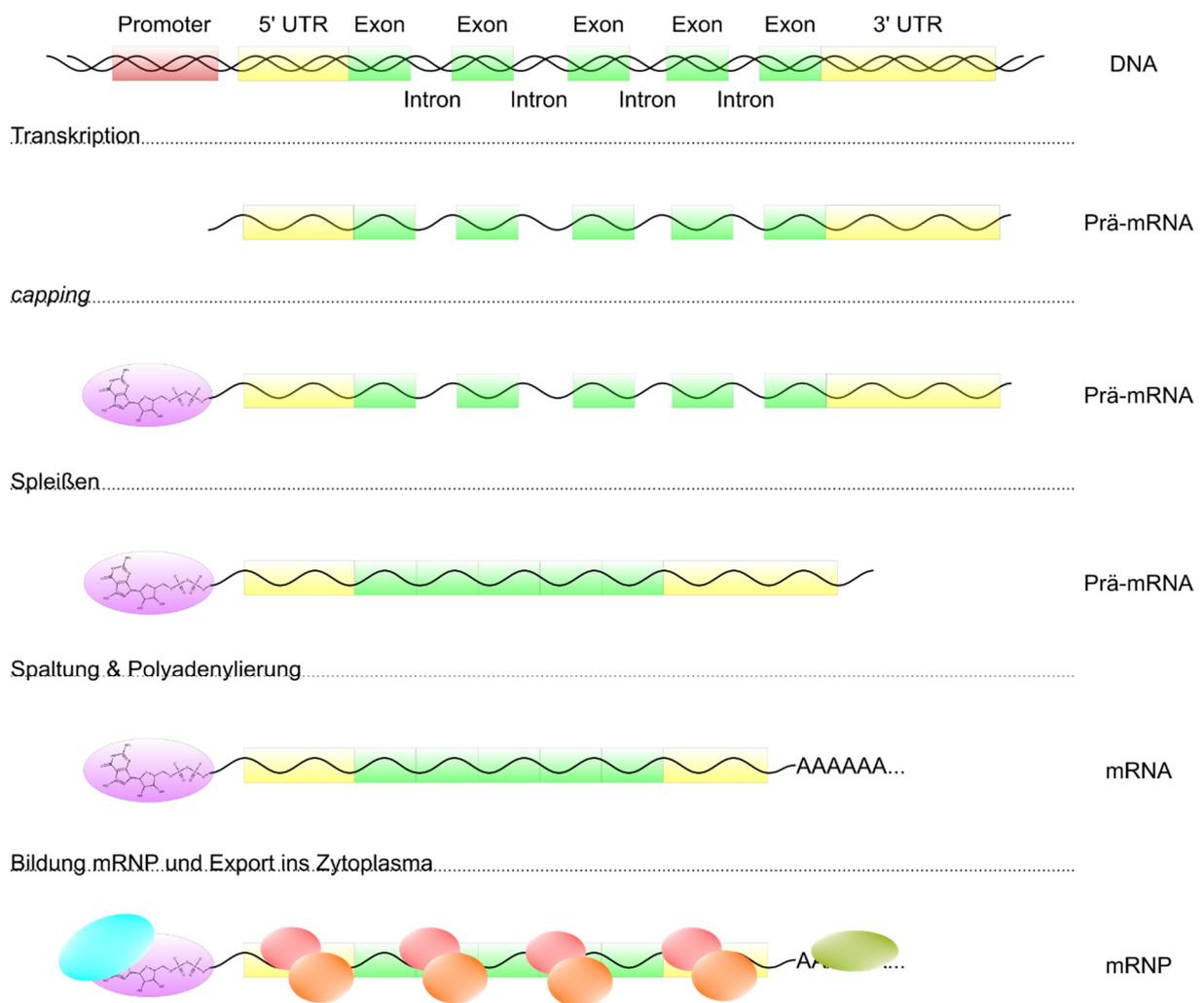


Abbildung 1.1 Schematische Übersicht der Genexpression und mRNA-Reifung. Gezeigt ist ein sehr vereinfachtes Schema des Verlaufes des eukaryotischen Genexpression. Ein Gen setzt sich aus einer Promoterregion (rot), dem 5'-untranslatiertem Bereich (5'-UTR; gelb), den kodierenden Sequenzen, welche sich in Exons (grün) und Introns aufteilen, und dem 3'-untranslatiertem Bereich zusammen. *Transkription* Die RNA-Polymerase II wird in der Promoterregion assembliert und synthetisiert die mRNA in 5'-3' Orientierung. *Capping* Unmittelbar

nach Transkriptionsbeginn und der Bildung des 5'-Ende der mRNA wird dieses durch das Anhängen eines 7-Methylguanosin-*cap* modifiziert (violett). *Spleißen* Die Introns werden durch das Spleißosom entfernt und die Exons zusammengefügt. Hierbei entsteht das intakte offene Leseraster. *Spaltung und Polyadenylierung* Das Polyadenylierungssignal im 3'-UTR wird von den Polyadenylierungsfaktoren erkannt und die prä-mRNA gespalten. Anschließend wird der Poly(A)-Schwanz gebildet, wobei dieser in menschlichen Zellen ca. 250 Adenosine lang ist. *Bildung mRNP und Export ins Zytoplasma* Die gereifte mRNA wird von Proteinen und Proteinkomplexen, die u.a. die *Cap*-Struktur und den Poly(A)-Schwanz erkennen, bedeckt und durch die Kernporen in das Zytoplasma exportiert.

1.1.1 Capping

Die *Cap*-Struktur, die bereits nach ca. 30 polymerisierten Nukleotiden an dem 5'-Ende der prä-mRNA gebildet wird, ist ein elementares Element von mRNAs. Zum einen dient es zur Signalisierung der erfolgreichen Transkriptionsinitiation und dem Start der Elongation [4]. Darüber hinaus dient die *Cap*-Struktur zum Schutz der mRNA vor exonukleolytischem Abbau [5]. Auch für den Export der RNA aus dem Zellkern wird diese Struktur benötigt [6]. Die *Cap*-Struktur ist darüber hinaus von entscheidender Bedeutung bei der Translation von mRNA. Hierbei wird diese von dem Translations-Initiationsfaktor eIF4E gebunden, welcher anschließend die Ribosomen rekrutieren und so eine effiziente Translation ermöglichen [7].

Die *cap*-Struktur wird in drei enzymatischen Schritten gebildet: (1) der Abspaltung des γ -Phosphates am 5'-Ende der prä-mRNA durch eine RNA-Triphosphatase (2) der Addition eines Guanylatrestes, wobei eine 5'-5'-Triphosphatbindung ausgebildet wird (3) der Methylierung dieses Guanins an der N7-Position.

1.1.2 Spleißen

Beim Spleißen werden nicht-kodierende Sequenzbereiche eines Genes, die sogenannten Introns, entfernt und die kodierenden Sequenzbereiche, die Exons, zusammengefügt. Dieses wird durch einen großen Ribonukleoproteinkomplex, das Spleißosom, katalysiert. Kern des Spleißosoms bilden 5 Uridin-reiche nukleare RNAs (snRNA), bekannt unter den Namen U1, U2, U4, U5 und U6 snRNA [8]. Das Spleißen ist Magnesium- und ATP-abhängig. Die beiden Magnesiumionen, welche für die Katalyse benötigt werden, werden durch das Ribophosphatrückgrat der U6 snRNA koordiniert. Das Spleißen von prä-mRNA verläuft über mehrere distinkte Schritte. Häufig erfolgt der Übergang zwischen den Intermediaten hierbei durch Helikasen, wodurch ATP verbraucht wird. Das Spleißosom wird schrittweise an der intronischen Region assembliert. Die U1 snRNA bindet an der 5'-Spleißstelle, welche sich am 3'-Ende des 5'-Exons befindet, gefolgt von der Bindung der U2 snRNA an die 3'-Spleißstelle. Anschließend wird das U4/U6.U5 tri-snRNP gebunden, wobei das präkatalytische Spleißosom gebildet wird. Dieses wird durch die Entfernung der U1 und U4 snRNA und Rekrutierung der NTC- und NTR-Komplexe aktiviert. Anschließend wird die RNA zwischen 5'-Exon und Intron gespalten, wobei sich eine

„Lasso“-Struktur des Introns durch eine 2'-5'-Phosphodiesterbindung bildet. Anschließend erfolgt eine weitere Umesterung, wobei das 5' und 3' Exon verknüpft wird. Nachfolgend werden die so gebildete mRNA und die an der Reaktion beteiligten Komplexe wieder freigegeben [9]. Die Strukturen der verschiedenen Komplexe des menschlichen Spleißosoms konnten mittels Kryoelektronenmikroskopie aufgeklärt werden [10]–[17].

1.1.3 Spaltung und Polyadenylierung

Der letzte Schritt in der Reifung von mRNA ist die Polyadenylierung, welche sich aus zwei Teilschritten zusammensetzt. Zuerst wird die RNA endonukleolytisch gespalten und anschließend polyadenyliert.

Der Poly(A)-Schwanz schützt die mRNA vor 3'-exonukleolytischem Abbau und ist damit maßgeblich für die Stabilität der RNA zuständig. Im Nukleus lagert sich bereits nach wenigen Adenosylresten das nukleäre Poly(A)-bindende Protein PABPN1 an [18]. Dieses bedeckt bis zum Export der mRNA in das Zytoplasma den Poly(A)-Schwanz. Im Zytosol wird das Protein durch das zytosolische Poly(A)-bindende Protein PABPC ersetzt. Über die Menge an gebundenem PABPC an dem Poly(A)-Schwanz wird die Deadenylierung der mRNA reguliert [19]. Fast alle mRNA-Abbauwege werden durch eine Deadenylierung initiiert [20]. Die Deadenylierung wird durch den Ccr4-Not-Komplex katalysiert, welcher neben dem Abbau des Poly(A)-Schwanzes unter anderem auch das *decapping*-Enzym DCP2 rekrutiert, sodass es zu einem effizienterem mRNA-Abbau kommt [21]. Neben dem Schutz der mRNA vor Abbau wird der Poly(A)-Schwanz auch für eine effiziente Translation-Initiation benötigt [22]. Durch eine Interaktion zwischen dem Poly(A)-bindenden Protein PABPC und dem *cap*-bindenden eIF4F-Komplex kommt es zu einer Bildung einer Ringstruktur, der sogenannten *closed-loop* Struktur [23]. Durch diese Struktur wird die Translationseffizienz deutlich gesteigert. Darüber hinaus ist der Prozess der Polyadenylierung, nicht das reine Vorhandensein eines Poly(A)-Schwanzes wichtig, um Exportfaktoren zu rekrutieren und einen effizienten Export in das Zytoplasma zu gewährleisten [24].

Den ersten Schritt der 3'-Prozessierung stellt die endonukleolytische Spaltung da. Diese Reaktion wird durch einen Multiproteinkomplex katalysiert. Die basalen Faktoren sind der CPSF- (*cleavage and polyadenylation specificity factor*), der CstF-Komplex (*cleavage stimulation factor*), die Spaltfaktoren I und II (*cleavage factor I / II*) sowie die Poly(A)-Polymerase. Diese Proteine werden in Kapitel 1.2 genauer charakterisiert. In einer massenspektrometrischen Analyse des Spaltungskomplexes konnten über 80 verschiedene Proteine identifiziert werden, die direkt oder indirekt mit dem Polyadenylierungskomplex interagieren [25]. Inwieweit diese für die Spaltung oder dessen Regulation eine Rolle spielen, muss noch geklärt werden. Dies wird dadurch erschwert, dass es bisher noch nicht gelungen ist, die Spaltung der prä-mRNA *in vitro* aus rekombinanten Proteinen zu rekonstituieren. Zur Analyse der prä-mRNA-Spaltung kann HeLa-

Kernextrakt verwendet werden, welcher *in vitro* spaltungsaktiv ist. Durch eine mehrstufige Fraktionierung dieses Kernextraktes ist es gelungen, essentielle Spaltfaktoren voneinander zu trennen. In einer der Fraktionen kann durch die Zugabe von rekombinantem CstF- und CFII-Komplex sowie der Poly(A)-Polymerase die Spaltung der prä-mRNA rekonstituiert werden [26]. So konnte der Einfluss des CFII-Komplexes auf die Spaltung untersucht werden [27].

Die zweite Teilreaktion, die Polyadenylierung der prä-mRNA, ist hingegen sehr gut verstanden. So sind für die längenspezifische Polyadenylierung der prä-mRNA nur der mPSF-Komplex, ein Subkomplex des CPSF-Komplexes, die Poly(A)-Polymerase sowie das Poly(A)-bindende Protein PABPN1 nötig. Die Reaktion kann durch rekombinant produzierte Proteine rekonstituiert werden [28], [29]. Die Untereinheiten des mPSF-Komplex und PABPN1 stimulieren die Poly(A)-Polymerase auf eine kooperative Weise, wodurch diese prozessiv wird [30], [31]. Der neu gebildete Poly(A)-Schwanz wird direkt von PABPN1 bedeckt, wobei ca. 10 Nukleotide pro Protein bedeckt werden. Nach ca. 250 Adenosylresten wird die Verbindung zwischen Poly(A)-Polymerase und dem 3'-Ende der mRNA getrennt und es findet keine weitere spezifische Polyadenylierung statt [29]. Eine genaue Beschreibung dieses Mechanismus ist jedoch nicht bekannt.

Die Regulierung und Effizienz der 3'-Prozessierung von mRNA erfolgt unter anderem durch besondere Sequenzmerkmale des 3'-untranslatierten Bereichs (3'-UTR), an denen die basalen Faktoren binden. Eine Übersicht der Sequenzelemente und der an der Bindung beteiligten Proteine ist in Abbildung 1.2 dargestellt.

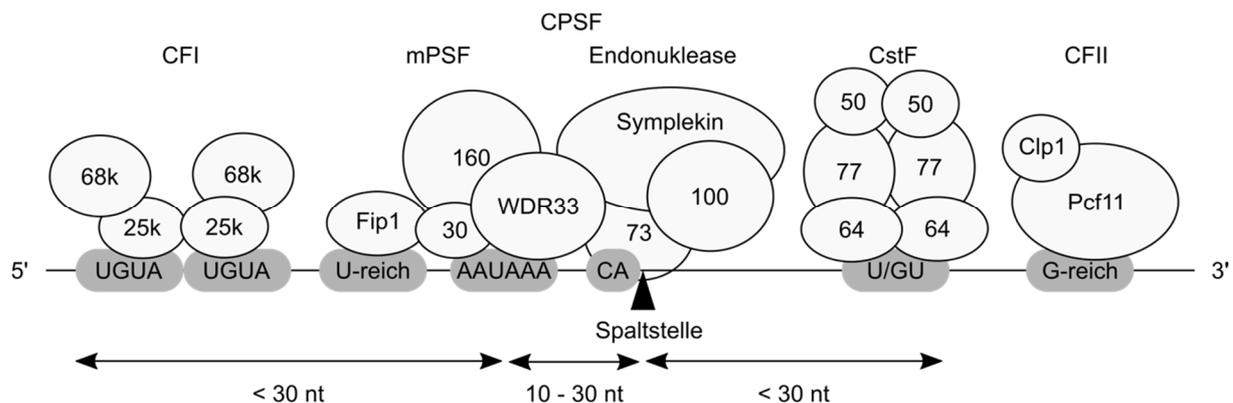


Abbildung 1.2 Schematische Übersicht des 3'-UTR einer mRNA und der daran gebundenen, an der 3'-Prozessierung beteiligten Proteine. Die Erkennungssequenzen sind grau hervorgehoben und die für die Bindung dieser benötigten Proteine überlappen mit diesen. Modifiziert nach *Misra & Green* [32].

Die wichtigste Sequenz stellt hierbei das Polyadenylierungssignal (PAS) mit der Sequenz AAUAAA dar. Diese wird durch die Untereinheiten WDR33 und CPSF30 des CPSF-Komplexes gebunden [28], [33]. Ungefähr 10 bis 30 Nukleotide stromabwärts des PAS liegt die Spaltstelle der mRNA, die sich häufig durch ein CA-Dinukleotid auszeichnet. Die Spaltstelle befindet sich 3'

dieses Dinukleotides, und die Spaltung wird durch CPSF73 katalysiert. Sowohl stromaufwärts als auch stromabwärts liegen weitere Sequenzelemente, die von verschiedenen Polyadenylierungsfaktoren gebunden werden. Stromaufwärts liegt die Bindesequenz des CFI-Komplexes. Dieser bindet zwei Kopien einer UGUA Sequenz [34]. CFI ist maßgeblich an dem Prozess der alternativen Polyadenylierung beteiligt. Hier kann dieser in Abwesenheit eines PAS als primärer Faktor binden und die anderen Faktoren wie den CPSF-Komplex durch direkte Interaktionen rekrutieren [35]. CFI68 besitzt eine RS-Domäne, welche ein Merkmal von Spleißfaktoren darstellt und Protein-Protein-Interaktionen vermittelt. Hierdurch wird das Spleißen direkt an die 3'-Prozessierung gekoppelt [36]. Fip1, Teil des mPSF-Komplexes, welcher für die Bindung des PAS verantwortlich ist, bindet selbst U-reiche RNA-Elemente [37]. Diese zusätzliche Bindestelle des Komplexes stabilisiert die Bindung der RNA und steigert daher die Effizienz der Reaktion. Stromabwärts nach der Spaltstelle liegen zwei weitere Sequenzelemente. Innerhalb eines Rahmens von maximal 30 Nukleotiden liegt eine U- bzw. GU-reiche Region. Dieses wird durch CstF64 des CstF-Komplexes gebunden [38]. Die Bindung von CstF erfolgt kooperativ mit CPSF, was möglicherweise eine Qualitätskontrolle darstellt [39]. Nur bei Vorhandensein sowohl der PAS als auch der CstF-Bindestelle kommt es zur stabilen Bildung eines Prozessierungskomplexes, welches die Spezifität der Polyadenylierung deutlich erhöht [39]. Weiter Stromabwärts der Bindestelle von CstF liegt eine G-reiche Region. Obwohl diese schon länger bekannt ist, konnte dieser bis vor kurzem kein Protein zugeordnet werden, welches dieses spezifisch bindet [40]. Es konnte jedoch kürzlich gezeigt werden, dass diese von dem CFII-Komplexes gebunden wird [27]. Nicht jede Polyadenylierungsstelle besitzt all diese Sequenzelemente und eine Kombination dieser Elemente beeinflusst die Auswahl und die Effizienz der 3'-Prozessierung der jeweiligen mRNA [41].

1.2 Struktur der bekannten Faktoren der Polyadenylierungsreaktion

1.2.1 CFI

Der CFI-Komplex konnte erstmals 1996 identifiziert und charakterisiert werden [42]. Neben der CFI25-Untereinheit besteht der Komplex entweder aus CFI59, CFI68 oder CFI72 [43]. Während CFI59 von einem separaten Gen kodiert wird, handelt es sich bei den beiden anderen um alternative Spleißvarianten eines Gens. CFI59 und CFI68 scheinen redundant zu sein, wobei CFI68 eine stärkere Stimulierung der Polyadenylierung bewirkt [44]. CFI bindet in der Nähe des Polyadenylierungssignales und stabilisiert die Bindung von CPSF, wodurch die Effizienz der Polyadenylierung gesteigert wird [45]. Die RNA-Bindung wird hierbei hauptsächlich durch CFI25 vermittelt, wobei CFI68 diese stabilisiert.

Strukturell ist der CFI-Komplex bereits gut charakterisiert. So konnte gezeigt werden, dass der Komplex als Heterotetramer, mit je zwei Kopien CFI25 sowie zwei Kopien CFI59 oder CFI68, vorliegt [44]. Die Dimerisierung findet hierbei durch CFI25 statt. CFI25 besteht hauptsächlich aus

einer klassischen Nudix-Domäne, einer Hydrolasedomäne mit einer charakteristischen $\alpha\beta$ -Faltung. Der Name dieser Domäne ist ein Akronym für Nukleosid-Diphosphat-gebunden an X, welches das Substrat von Nudix-Domänen beschreibt. Diese werden zu einem Nukleosid-Monophosphat und einem phosphorylierten Rest P-X gespalten [46]. Das aktive Zentrum dieser Domäne ist jedoch in CFI25 durch eine Helix blockiert, sodass das Protein keine Hydrolaseaktivität aufweist [34], [47], [48]. Stattdessen ist die Nudix-Domäne an der spezifischen Bindung von RNA beteiligt. In der Strukturaufklärung von CFI konnte ein RNA-Molekül kokristallisiert werden, welches diese neue Funktion der Nudix-Domäne verdeutlicht.

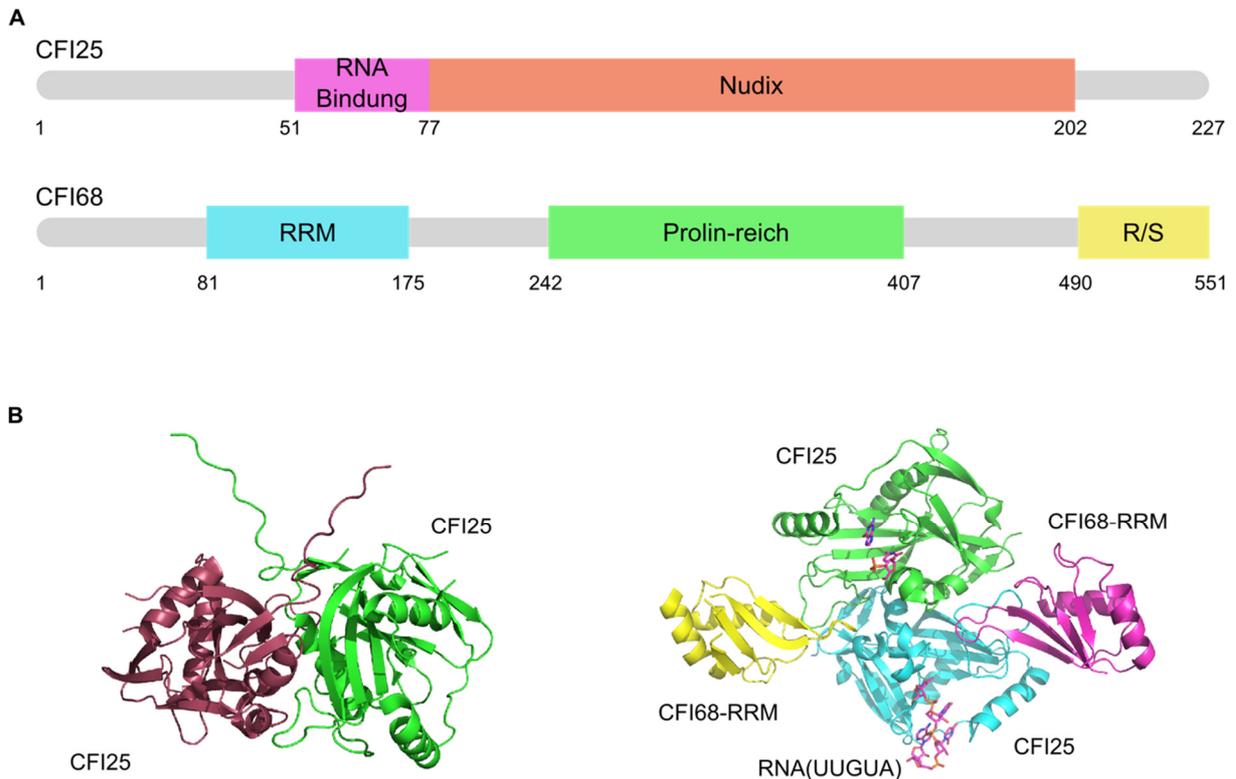


Abbildung 1.3 Zusammenfassung der bekannten strukturellen Informationen über den CFI-Komplex. A). Maßstabgetreue Darstellung der Domänenstruktur der beiden Untereinheiten. Das RNA-Erkennungsmotiv (RRM) sowie die Arginin/Serin-reiche Region (R/S) in CFI68 sind abgekürzt beschriftet. B) Darstellung der bekannten Strukturen von CFI. *Links* Dimere Struktur von CFI25 (PDBID: 3BAP). *Rechts* Tetramere Teilstruktur von CFI. Zentral liegt das CFI25-Dimer (türkis/grün). Je ein Monomer interagiert mit der RRM von CFI68 (gelb/magenta). Die kokristallisierte RNA ist in Stäbchenform dargestellt (PDBID: 3P6Y). Abbildung in Anlehnung an Xiang *et al.* [49].

CFI68 interagiert über seine N-terminale RRM-Domäne (*RNA recognition motif*) mit dem dimeren CFI25 [47], [50]. Die RRM-Domäne von CFI68 ist entgegen ihrem Namen nicht direkt an der RNA-Bindung beteiligt. Darüber hinaus besitzt CFI68 am C-Terminus eine SR-Domäne. Diese zeichnet sich durch häufige Wiederholungen von Serin-Arginin-Dipeptiden aus. Dieses ist ein Merkmal von

Proteinen vielzelliger Organismen, die am Spleißen beteiligt sind [51]. Die beiden Proteine sind in Abbildung 1.3 A schematisch dargestellt.

Die molekulare Struktur von CFI25 sowie der RRM-Domäne von CFI68 konnte durch Röntgenkristallographie aufgeklärt werden [48], [50]. In Rahmen dieser Strukturaufklärung wurde eine synthetische RNA kokristallisiert (Abbildung 1.3 B).

1.2.2 CFII

Der CFII-Komplex wurde erstmals 2000 aus HeLa-Kernextrakt isoliert, wurde jedoch erst 2018 in seiner genauen Zusammensetzung beschrieben [27], [52]. Er besteht aus den Proteinen Clp1 und Pcf11, welche ein Heterodimer bilden. Die homologen Proteine in Hefe, yClp1 und yPcf11, bilden in Hefe mit den Proteinen Rna14p und Rna15p den CFIA-Komplex [53], [54]. Die Rna14p- und Rna15p-Homologen, CstF77 und CstF64, bilden in menschlichen Zellen zusammen mit CstF50 den unabhängigen CstF-Komplex.

Der CFII-Komplex besitzt eine sehr hohe Affinität zu G-reichen Sequenzen, die sich stromabwärts (*downstream*) befinden [27]. CFII scheint eine geringe Affinität zu den anderen Polyadenylierungsfaktoren zu besitzen. So konnte in einem prä-mRNA-*pulldown*-Experiment CFII im Vergleich zu den anderen Polyadenylierungsfaktoren nur substöchiometrisch identifiziert werden [25]. Neben seiner Funktion als essentieller Polyadenylierungsfaktor spielt vor allem Pcf11 eine Rolle bei einer Reihe weiterer Prozesse. So ist es an der Termination der RNA-Polymerase II beteiligt. Es konnte gezeigt werden, dass in Abhängigkeit von einem intakten Polyadenylierungssignales Pcf11 in der Lage ist, die Polymerase II von der DNA zu trennen [55]. Darüber hinaus ist Pcf11 am Abbau des nach der Spaltung entstandenen 3'-RNA Fragmentes durch die Rekrutierung der Exonuklease XRN2 beteiligt [55], [56]. Neben diesen Prozessen hat Pcf11 auch einen Einfluss auf den effizienten Export der gereiften mRNA. Nach Phosphorylierung durch die Kinase WNK1 wird die Interaktion zwischen Pcf11 und der RNA-Polymerase II unterbrochen und die Bildung eines mRNP ermöglicht, welches dann in das Zytoplasma exportiert wird [57]. Es ist anzunehmen, dass diese Prozesse miteinander stark verschränkt sind, da z.B. auch die Rekrutierung von XRN2 die Termination der RNA-Polymerase II fördert [58]. Clp1 hingegen spielt neben seiner Funktion in der mRNA 3'-Prozessierung eine Rolle in dem tRNA-Endonuklease-Komplex während des Spleißens von tRNA [59], [60].

Strukturell ist von dem CFII-Komplex nur Clp1 charakterisiert. Die Struktur der Homologen aus Bäckerhefe und dem Fadenwurm *C. elegans* konnte röntgenkristallographisch aufgeklärt werden [61]–[63] (Abbildung 1.4 B). Im Rahmen dieser Strukturaufklärung konnte zudem die Interaktionsregion von Pcf11 kokristallisiert werden. Clp1 besitzt ein zentrales Walker-A Motiv, eine hochkonservierte Tertiärstruktur [64]. Diese bindet üblicherweise ATP oder GTP und ist daher häufig in Hydrolasen zu finden. In Clp1 aus Säugetieren, aber auch in *C. elegans*, konnte stattdessen eine Polynukleotidkinaseaktivität identifiziert werden [60], [62]. Das Hefehomolog ist,

trotz der hochkonservierten Struktur, katalytisch inaktiv. Diese Aktivität scheint für die mRNA-3'-Prozessierung irrelevant. So konnten in katalytisch inaktiven Mutanten keine Veränderung der prä-mRNA-Spaltung beobachtet werden [27].

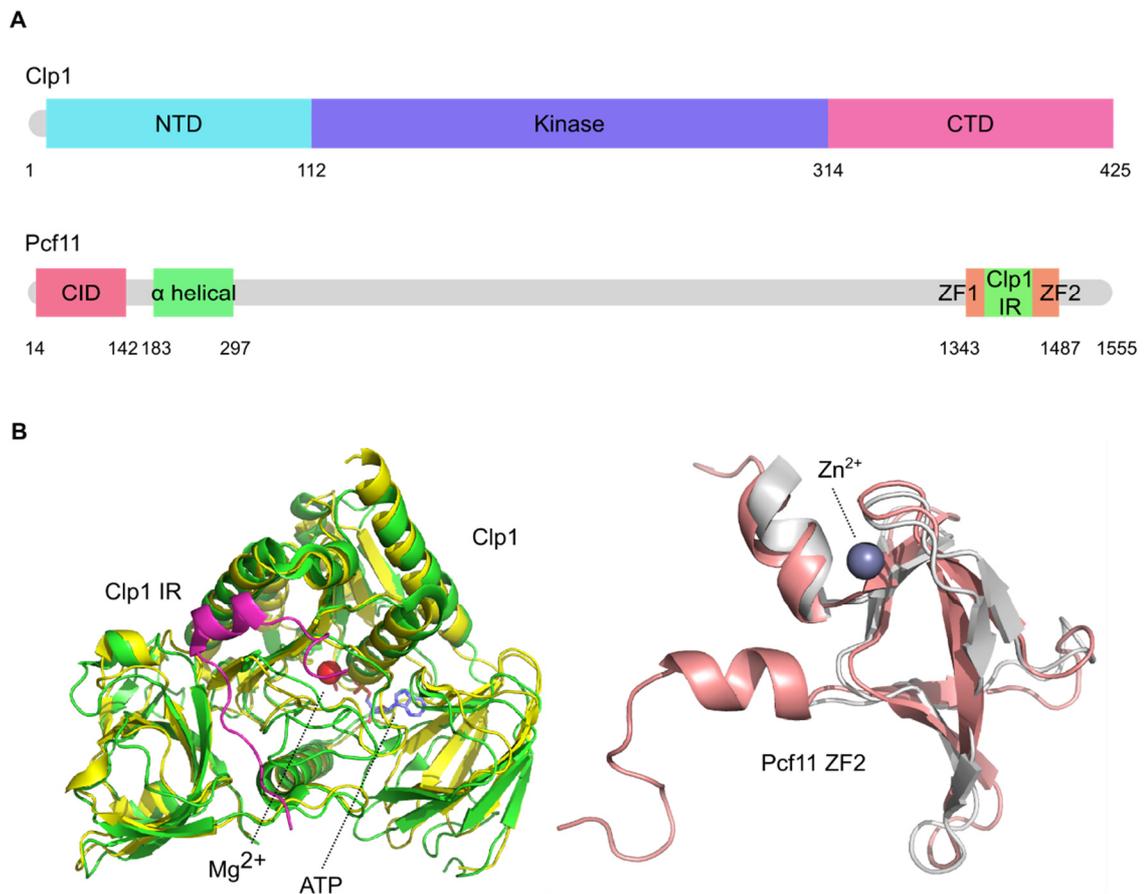


Abbildung 1.4 Zusammenfassung der strukturellen Informationen zum CFII-Komplex. A) Maßstabsgetreues Schema der Domänenorganisation von Clp1 und Pcf11. Die N-terminale (NTD) und C-terminale Domäne (CTD) flankieren die aktive Kinase-Domäne von Clp1. Pcf11 besitzt N-terminal die RNA Polymerase II-CTD-Interaktions-Domäne (CID) und einen α -helikalen Bereich. C-terminal liegt die Zinkfingerdomäne, welche sich aus zwei Zinkfingern (ZF1: 1343-1368; ZF2: 1443-1478) sowie der Clp1-Interaktionsregion (Clp1-IR: 1371-1439) zusammensetzt. B) Bekannte Strukturen des CFII-Komplexes. *Links* Überlagerung der kristallisierten Struktur von Clp1 aus Hefe (grün; PDBID: 2NPI) und *C. elegans* (gelb; PDBID: 4OHV). Überlagert aus der Struktur der Hefe ist das kokristallisierte Pcf11 Peptid (magenta; Clp1 IR), das Magnesiumion sowie das gebundene ATP. *Rechts* Struktur des zweiten Zinkfingers von Pcf11. Überlagert sind die beiden unabhängig publizierten NMR-Strukturen mit gebundenem Zinkion (PDBIDs: 2NAX, 5M9Z). Abbildung in Anlehnung an Xiang *et al.* [49].

Im Gegensatz zu Clp1 ist das humane Pcf11 sehr schlecht strukturell charakterisiert. Das humane Protein ist mit 1555 Aminosäuren deutlich größer als sein Homolog in Hefe, welches nur 626 Aminosäuren besitzt. N-terminal besitzt Pcf11 eine CID (*C-terminal domain (CTD) interaction domain*), welche mit der CTD der RNA-Polymerase II interagiert [65]. Darauf folgt eine helikale

Domäne, welche jedoch nicht weiter charakterisiert ist [66]. Im C-Terminus von Pcf11 liegt die Clp1-Interaktionsregion, welche hoch konserviert ist. Dieser Bereich ist mit Clp1 kokristalliert worden. Flankiert wird diese Region von zwei Zinkfingerdomänen, welche auch hoch konserviert sind. Von dem zweiten Zinkfinger, welcher eine unübliche Sequenz im Vergleich zu anderen Zinkfingerdomänen aufweist, sind zwei unabhängige NMR-Strukturen publiziert worden [67], [68]. Eine Zusammenfassung der bekannten Domänenstruktur ist in Abbildung 1.4 A dargestellt.

1.2.3 CstF

Der humane CstF-Komplex wurde 1990 erstmalig identifiziert [69]. Er besteht aus den drei Untereinheiten CstF50, CstF64 und CstF77. In Bäckerhefe bilden Rna14p und Rna15p, welche homolog zu CstF77 und CstF64 sind, mit yClp1 und yPcf11 den CFIA-Komplex [53], [54]. CstF50 besitzt kein verwandtes Protein in Hefe. Darüber hinaus besitzt CstF64 ein Paralog, CstF64tau, welches von einem anderen Genlokus kodiert wird [70]. Beide Paraloge scheinen redundante Funktionen zu besitzen [71]. CstF bildet einen hexameren Komplex durch die Bildung Dimern von Trimeren [72]. Die genaue Bildung der hexameren Struktur von CstF wird im Rahmen dieser Arbeit näher beleuchtet. Sowohl für CstF77 als auch für CstF50 konnte eine Homodimerisierung nachgewiesen werden. CstF77 bildet die zentrale Struktur des Komplexes und interagiert sowohl mit CstF64 als auch CstF50 [73]. Zwischen den Untereinheiten CstF64 und CstF50 konnte keine Interaktion identifiziert werden [74], [75].

CstF bindet GU-reiche Sequenzen stromabwärts des Polyadenylierungssignales über CstF64 [38], [76]–[78]. CstF ist ein essentieller Faktor der Polyadenylierung und mitverantwortlich für die Effizienz der Reaktion und Selektion der Polyadenylierungsstelle [78], [79].

Der CstF-Komplex ist in seinen Untereinheiten gut untersucht, jedoch ist keine Struktur des gesamten Komplexes bekannt. CstF77 ist am besten charakterisiert. Das Protein besteht aus einer sehr großen HAT-Domäne (*half-a-TPR*) [80]. Diese kann in einen N- und C-terminalen Teil getrennt werden. Dieses ist schematisch in Abbildung 1.5 A dargestellt. Der C-terminale Teil dieser Domäne ist wichtig für die Dimerisierung von CstF77. Dieses konnte biochemisch sowie strukturell gezeigt werden (Abbildung 1.5 B) [72], [74], [81]. Diese Interaktionsfläche ist auch in dem homologen Protein Rna14 aus *Kluyveromyces lactis* zu beobachten [82]. C-terminal besitzt CstF77 eine „Affenschwanz“-Domäne (*monkey tail*), welche für die Interaktion mit CstF64 benötigt wird und auch in den Hefehomologen konserviert ist [72], [83], [84]. Die Struktur dieser Domäne des homologen Proteins Rna14p ist mittels NMR aufgeklärt worden [84]. Diese Region ist sehr Prolin-reich und scheint erst durch die Interaktion mit CstF64 strukturell stabilisiert zu werden.

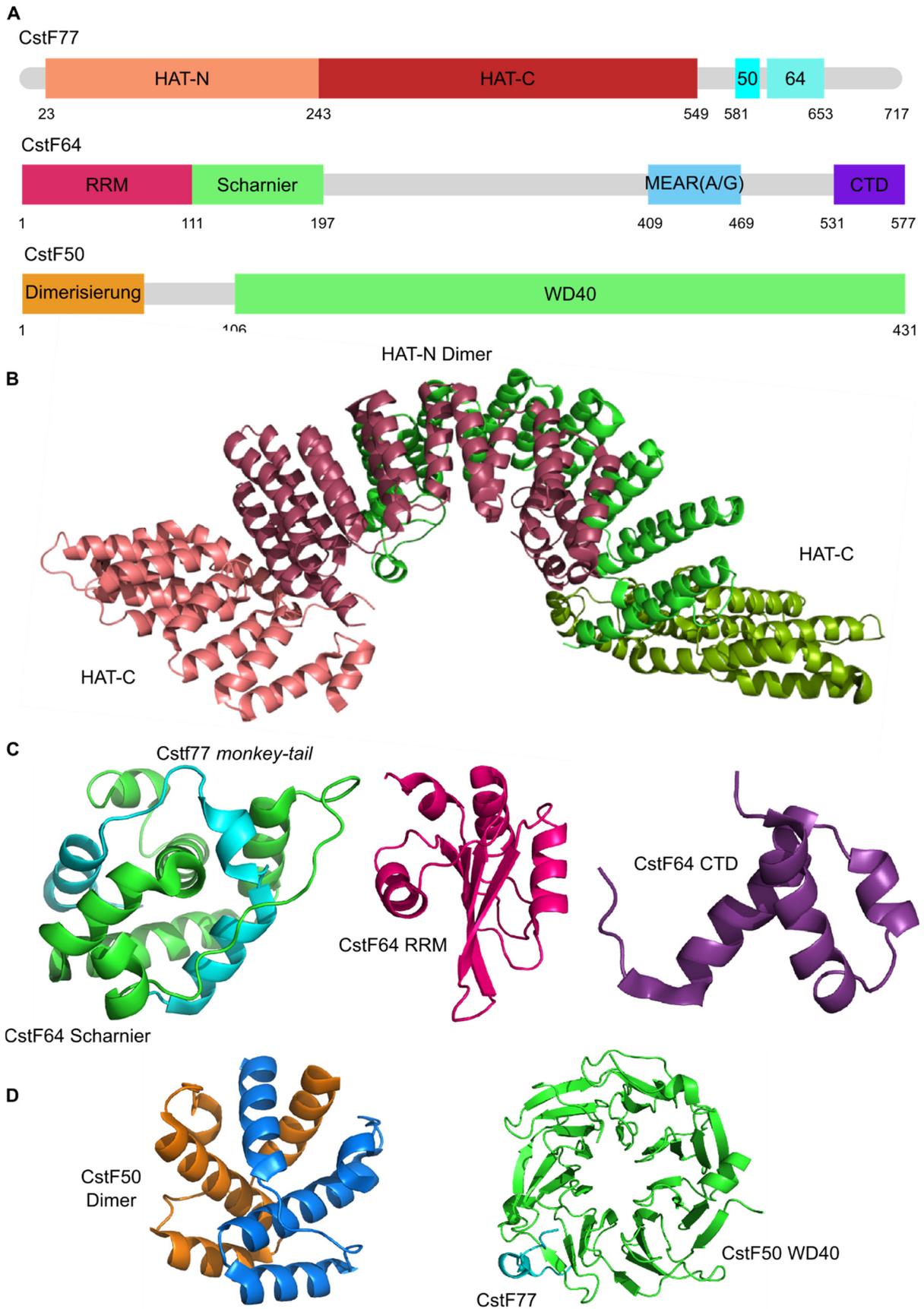


Abbildung 1.5 Zusammenfassung der strukturellen Informationen zum CstF-Komplex. A) Maßstabgerechte Darstellung der Domänenanordnung der CstF-Untereinheiten. Die HAT-Domäne von CstF wurde in N-terminalen und C-terminalen Teil, welcher für die Dimerisierung

verantwortlich ist, getrennt. C-terminal besitzt CstF77 zwei kokristallisierte Interaktionsdomänen für CstF50 (50) und CstF64 (64), welche auch Affenschwanzdomäne genannt wird. CstF64 besitzt N-terminal das RNA-Erkennungsmotiv (RRM), gefolgt von der Scharnierdomäne (*hinge*), welche für die Interaktion mit CstF77 wichtig ist. C-Terminal liegt eine Region mit der repetitiven Aminosäuresequenz MEAR(A/G) gefolgt von der C-terminalen Domäne (CTD). CstF50 besitzt N-terminal die Dimerisierungsdomäne, welche für die Homodimerisierung sorgt. Der Hauptteil des Proteins wird durch eine WD40-Domäne gebildet. B) Homodimere Struktur von CstF77. Die Monomere sind farblich getrennt dargestellt (rot/grün). Der N- und C-terminale Teil der HAT-Domäne ist darüber hinaus farblich differenziert. Die dimere Struktur wird durch die HAT-N Domäne gebildet (PDBID: 2OOE). C) Zusammenfassung der bekannten Strukturen von CstF64. *Links* Die Scharnierdomäne der Hefehomologe Rna15 (CstF64; grün) im Komplex mit der Affenschwanzregion von Rna14 (CstF77; *monkey-tail*; türkis) (PDBID: 2L9B). *Mitte* Die Struktur des RNA-Erkennungsmotiv (RRM) von CstF64, welche für die Bindung der RNA benötigt wird (PDBID: 1P1T). *Rechts* Die C-terminale Domäne von CstF64 (PDBID: 2J8P). D) Darstellung der bekannten Strukturen von CstF50. *Links* Dimere Struktur der N-terminalen Dimerisierungsdomäne von CstF50. Die monomeren Strukturen sind farblich getrennt (PDBID: 2XZ2). *Rechts* Darstellung der WD40-Domäne von CstF50 (grün) im Komplex mit einem Peptid von CstF77 (Aminosäuren 581–600; türkis) (PDBID: 6B3X). Abbildung in Anlehnung an *Xiang et al.* [49].

CstF64 ist die strukturell am schlechtesten charakterisierte Untereinheit des CstF-Komplexes. N-terminal liegt die RRM-Domäne (*RNA recognition motif*), welche für die RNA-Bindung des Komplexes verantwortlich ist [85]. Diese Domäne konnte sowohl aus dem humanen Protein als auch seinem Hefehomolog mittels NMR (human) und Röntgenkristallographie (Hefe) aufgeklärt werden (Abbildung 1.5 C *Mitte*) [38], [86]. Auf die RRM-Domäne folgt eine sogenannte „Scharnier“-Domäne (*hinge*). Über diese Region interagiert CstF64 mit der Affenschwanz-Domäne von CstF77 [74]. Alternativ zu dieser Interaktion kann CstF64 über dieselbe Region mit Symplekin, einer Untereinheit des CPSF-Komplexes, interagieren [87]. Die Struktur des Hefehomologes dieser Region konnte mittels NMR im Komplex mit dem CstF77-Homologen Rna15p gelöst werden (Abbildung 1.5 C *Links*) [84]. Am C-Terminus besitzt CstF64 eine konservierte α -helikale Region, deren Struktur mittels NMR aufgeklärt werden konnte (Abbildung 1.5 C *Rechts*) [88]. Diese Region ist essentiell für die Polyadenylierung und in Hefe wichtig für die Interaktion zwischen Rna15p und Pcf11 [88], [89]. Eine Interaktion zwischen den humanen Varianten von CstF64 und Pcf11 konnte nicht identifiziert werden [27]. Zwischen der Scharnier- und der C-terminalen Domäne liegt eine Glycin/Prolin-reiche Sequenz, gefolgt von einer Region in der die Aminosäuresequenz MEAR(A/G) 12-fach wiederholt vorliegt. Diese Region liegt α -helikal vor, über die Funktion ist jedoch nichts bekannt [90]. Auch ist sie in dem Hefehomologen nicht vorhanden.

CstF50 bildet Homodimere und besitzt kein Homolog in Hefe. Die Dimerisierung erfolgt in einer N-terminalen Domäne, welche mittels Röntgenkristallographie strukturell aufgeklärt ist (Abbildung

1.5 D *Links*) [91]. Darüber hinaus besitzt CstF50 eine WD40-Domäne mit 7 Wiederholungen. Während der Strukturaufklärung dieser Domäne konnte gezeigt werden, dass diese für die Interaktion mit CstF77 verantwortlich ist (Abbildung 1.5 D *Rechts*) [73].

1.2.4 CPSF

Der CPSF-Komplex stellt das Zentrum der Polyadenylierung dar. Er wurde zuerst im Rahmen biochemischer Experimente als Komplex von vier Proteinen, benannt nach ihrem apparentem Molekulargewicht, CPSF30, CPSF73, CPSF100 sowie CPSF160 beschrieben [92]. Erst nachträglich konnten die Proteine Fip1, Symplekin sowie WDR33 dem Komplex zugeordnet werden [25], [37], [74], [93]. Der CPSF-Komplex in seiner bekannten Zusammensetzung scheint ein metastabiler Komplex zu sein, der sich aus den Subkomplexen mPSF (*mammalian polyadenylation specificity factor*) und dem Endonuklease-Komplex zusammensetzt. mPSF besteht aus den Proteinen CPSF30, CPSF160, Fip1 sowie WDR33 [28], [94]. Der Endonuklease-Komplex ist zusammengesetzt aus CPSF73, CPSF100 sowie Symplekin [95], [96]. WDR33 und CPSF30 des mPSF-Komplexes sind für die spezifische Bindung des Polyadenylierungssignales verantwortlich [28], [94]. Auch ist dieser Komplex in der Lage, nach der Spaltung der RNA die Poly(A)-Polymerase zu rekrutieren und die Polyadenylierung zu stimulieren. Der Endonuklease-Komplex, welcher die für die Spaltung nötige Endonuklease CPSF73 enthält, konnte sowohl der mRNA-3'-Prozessierung wie auch der Prozessierung von Histon-RNA, hier jedoch ohne Beteiligung des mPSF-Komplexes, zugeordnet werden [97]. Erst kürzlich konnte gezeigt werden, dass die beiden CPSF-Subkomplexe über WDR33 und CPSF100 interagieren [98]. Eine Interaktion von CstF-Untereinheiten in der Bildung des Endonuklease-Komplexes wird diskutiert. So zeigte sich in fraktioniertem Kernextrakt, dass neben den CPSF-Untereinheiten auch CstF64, nicht jedoch die anderen Untereinheiten von CstF angereichert waren (persönliche Kommunikation mit Florian Kluge). Auch konnte eine direkte Interaktion von Symplekin mit CstF64 nachgewiesen werden [87].

1.2.4.1 mPSF

Die genaue Zusammensetzung des mPSF-Komplexes ist bekannt, da dieser aus rekombinanten Proteinen rekonstituiert werden konnte [28]. Strukturell konnte der mPSF-Komplex durch kryoelektronenmikroskopie, Röntgenkristallographie sowie *cross-linking* gekoppelt mit Massenspektrometrie aufgeklärt werden [99]–[101]. Die Struktur des Komplexes wird durch eine sehr stabile Interaktion zwischen CPSF160 und WDR33 stabilisiert. Sie besitzen eine Interaktionsfläche von ca. 3700 Å² [101]. Darüber hinaus interagiert CPSF30 mit CPSF160 (Interaktionsfläche 1700 Å²), WDR33 (Interaktionsfläche 850 Å²) und Fip1 [101]. Trotz Interaktionen zwischen Fip1 und den beiden anderen Proteinen ist eine stabile Interaktion nur über CPSF30 gegeben [99]. Das Polyadenylierungssignal wird synergistisch von WDR33 und CPSF30 gebunden. Im Folgenden werden die einzelnen Proteine näher dargestellt.

Die Untereinheit CPSF30 besitzt N-terminal fünf aufeinanderfolgende Zinkfingerdomänen und C-terminal eine CCHC-Zinkknöcheldomäne (*zinc knuckle*) [102]. Die Zinkfingerdomänen 2 und 3 sind direkt an der Bindung des Polyadenylierungssubstrates beteiligt, die Domänen 4 und 5 interagieren mit Fip1 [99], [100]. Die Struktur von CPSF30 ist nur bis zum dritten Zinkfinger in atomarer Auflösung geklärt, der C-terminale Bereich konnte aufgrund seiner Flexibilität nicht aufgelöst werden.

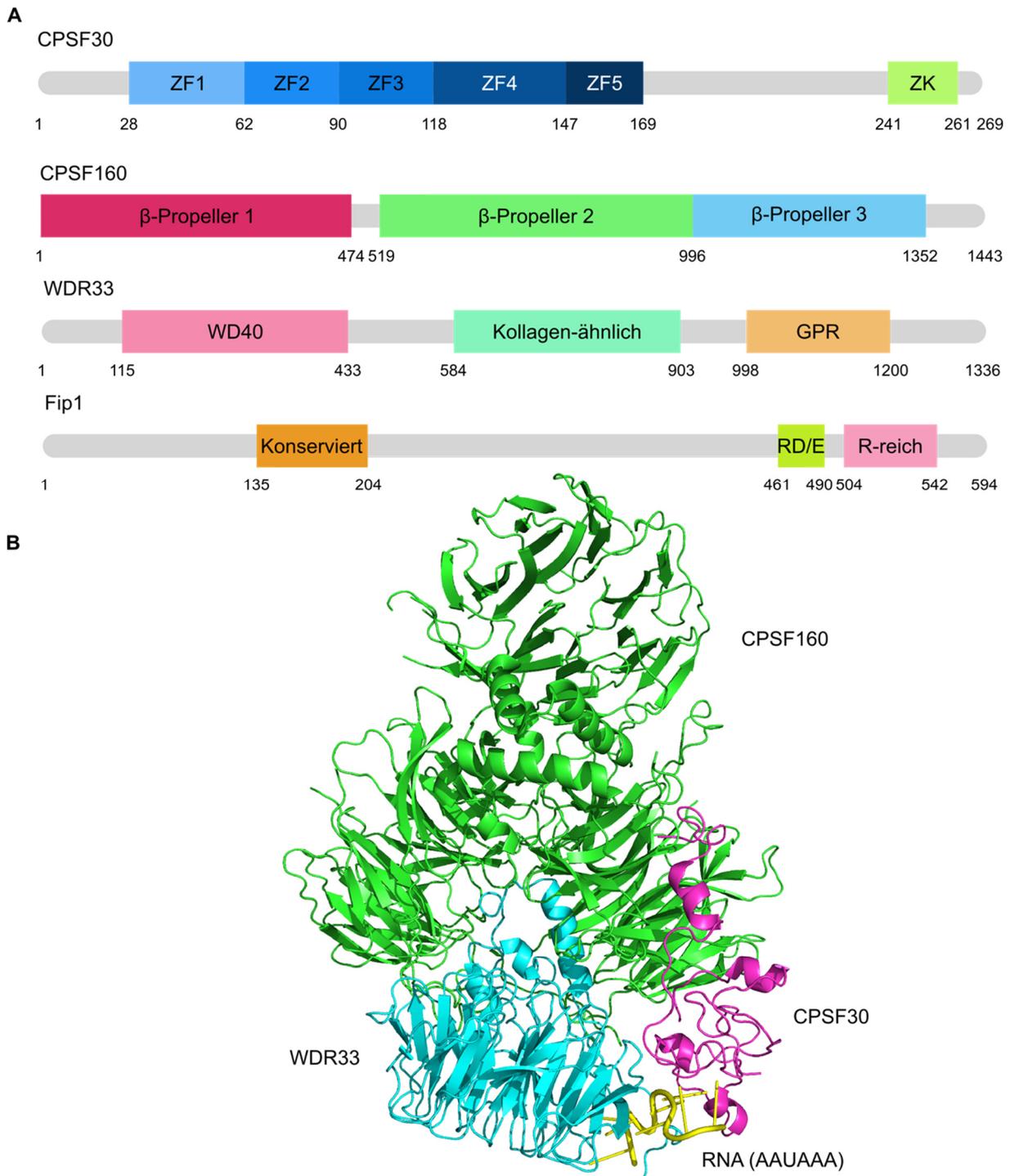


Abbildung 1.6 Zusammenfassung der strukturellen Informationen zum mPSF-Komplex. A) Maßstabgerechte Darstellung der Domänenanordnung der Untereinheiten des mPSF

Komplexes. CPSF30 besitzt fünf aufeinanderfolgende Zinkfingerdomänen (ZF1-5) sowie C-terminal eine Zinkköcheldomäne (ZK). CPSF160 ist aus drei β -Propeller-Domänen zusammengesetzt. WDR33 besitzt N-terminal die namensgebende WD40-Domäne. Zentral liegt ein Kollagen-ähnlicher Sequenzbereich. C-terminal liegt eine GPR (Glycin/Prolin/Arginin)-Domäne. Die konservierte Domäne von Fip1 ist für die Interaktion mit der Poly(A)-Polymerase verantwortlich. C-terminal besitzt Fip1 eine Region, die aus Arginin-Aspartat/Glutamat-Dipeptid-Wiederholungen besteht (R/DE) sowie eine Arginin-reiche Region. Dieses ist für die RNA-Bindung von Fip1 verantwortlich. B) Räumliche Darstellung des mPSF-Komplexes, aufgeklärt durch Kryoelektronenmikroskopie. CPSF160, WDR33 und ein Teilbereich von CPSF-30 konnten aufgelöst werden und sind farblich differenziert. Zusätzlich konnte die gebundene RNA mit dem Polyadenylierungssignal AAUAAA (gelb) mit aufgelöst werden (PDBID: 6BLY). Abbildung in Anlehnung an *Xiang et al.* [49].

1.2.4.2 Endonuklease-Komplex

Der Endonuklease-Komplex besteht aus den Proteinen CPSF73, CPSF100 sowie Symplekin. Der Komplex ist essentiell für die 3'-Prozessierung von mRNA, da er die für die Spaltung benötigte Endonuklease trägt. In der Prozessierung von Histon-mRNA konnte gezeigt werden, dass es zu einer zusätzlichen Assoziation von CstF64 kommt [103].

Die homologen Proteine zu CPSF73 und CPSF100 aus Hefe, Ysh1 und Cft2, bilden zusammen mit Mpe1 den Endonuklease-Komplex [104]. Das humane Homolog zu Mpe1, Rbbp6, ist als Regulator der Polyadenylierung, nicht jedoch als fester Bestandteil von CPSF beschrieben [105]. Das Homolog zu Symplekin, Pta1, ist im Endonuklease-Komplex in Hefe nicht als fester Bestandteil, sondern als Gerüstprotein der 3'-Prozessierung beschrieben [106]. Es scheint daher signifikante Unterschiede zwischen der Zusammensetzung der Subkomplexe zwischen Hefe und Menschen zu geben.

Strukturell am besten aufgeklärt ist CPSF73, die Endonuklease der mRNA-3'-Prozessierung [107]. Die N-terminale Region besteht aus einer Metallo- β -Lactamase-Domäne, welche in der Primärstruktur von einer β -CASP-Domäne unterbrochen wird [108]. Diese Region ist das aktive Zentrum der Endonuklease. Der C-terminale Bereich von CPSF73 ist strukturell wenig klassifiziert, es ist aber bekannt, dass er wichtig für die Interaktion mit CPSF100 und Symplekin ist [109], [110]. Die Struktur der Metallo- β -Lactamase- und β -CASP-Domäne konnte mittels Röntgenkristallographie des humanen CPSF73 sowie seines Hefehomologen Ysh1 geklärt werden [96], [111]. Hierbei zeigte sich, dass das Protein in einer geschlossenen, inaktiven Form sowie in einer geöffneten, aktiven Form vorliegen kann.

CPSF100 ist strukturell dem CPSF73 sehr ähnlich [112]. Es besitzt eine sehr ähnliche Domänenstruktur mit einer N-terminalen Metallo- β -Lactamase- und β -CASP-Domäne. Diese unterscheiden sich durch eine ca. 100 Aminosäure große Insertion in CPSF100, welcher ursprünglich keine Funktion zugeordnet werden konnte [108]. Es konnte jedoch unlängst gezeigt

werden, dass in diesem Bereich die Interaktionsregion zu WDR33 liegt [98]. Darüber hinaus sind die Zink-bindenden Aminosäuren in CPSF100 nicht vorhanden, sodass das Protein auch keine Nukleaseaktivität vorweist [108], [111]. Über die C-terminale Domäne interagiert CPSF100 mit CPSF73. Die Dimerisierung zu Bildung eines aktiven Komplexes scheint eine Eigenschaft der Proteine der β -CASP-Familie zu sein [113]. Im Rahmen der Kristallisation von CPSF73 konnte das Hefehomologe zu CPSF100, Cft2, kristallisiert werden [111]. Auch hier wurde nur der N-terminale Bereich des Proteins kristallisiert, der für die Dimerisierung der Untereinheiten benötigte C-Terminus ist nicht aufgelöst.

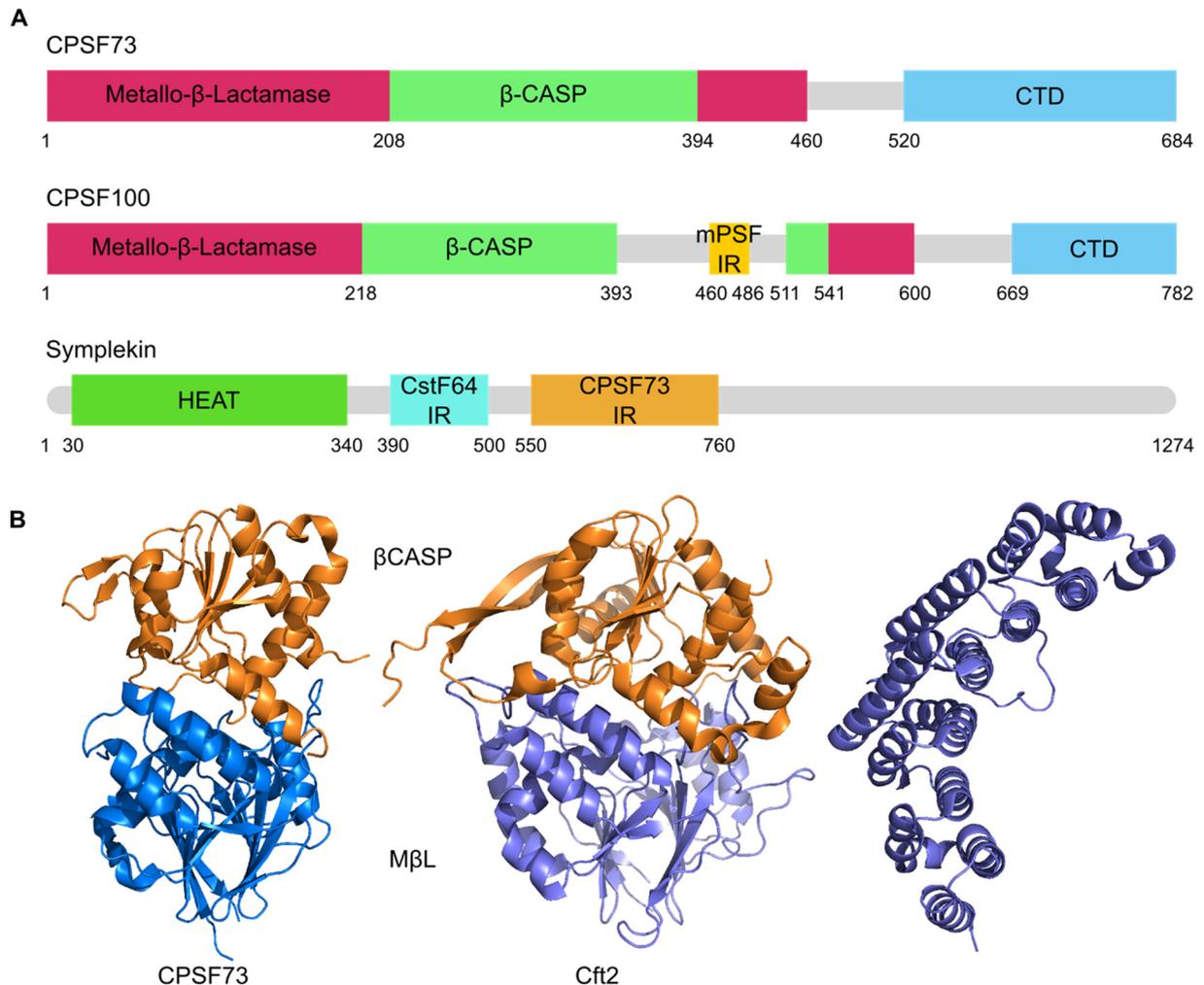


Abbildung 1.7 Zusammenfassung der bekannten Strukturelemente des Endonuklease-Komplexes. A) Maßstabgetreue Darstellung der Domänenanordnung der Untereinheiten des Endonuklease-Komplexes. CPSF73 und CPSF100 sind sich strukturell sehr ähnlich. Beide besitzen eine Metallo- β -Lactamase-Domäne, die in CPSF100 jedoch degeneriert vorliegt und keine Zink-Bindung und damit keine enzymatische Aktivität aufweist. Die Metallo- β -Lactamase-Domäne wird von einer β -CASP-Domäne unterbrochen. Die β -CASP-Domäne in CPSF100 ist wiederum durch eine Insertion (Aminosäuren 393 bis 511) unterbrochen. In dieser sitzt die beschriebene Interaktionsregion zum mPSF-Komplex. Beide Proteine besitzen eine C-terminale Domäne (CTD), welche für die Dimerisierung benötigt wird. Symplekin besitzt eine N-terminale

HEAT-Domäne, welche sich aus sieben antiparallelen α -Helices zusammensetzt. Nachfolgend liegen die Interaktionsregionen für CstF64 (CstF64 IR) sowie CPSF73 (CPSF73 IR). B) Darstellung der bekannten Strukturen der Untereinheiten des Endonuklease-Komplexes. *Links* Kristallstruktur der Endonuklease CPSF73. *Mitte* Kristallstruktur des Hefehomologes zu CPSF100, Cft2. Die Metallo- β -Lactamase- (M β L; blau) sowie die β -CASP-Domäne (β CASP; orange) sind in beiden Proteinen hervorgehoben. *Rechts* Kristallstruktur der HEAT-Domäne von Symplekin (PDBID: 3O2Q). Deutlich werden die sieben antiparallelen Helixpaare, welche diese Domäne bilden. Abbildung in Anlehnung an *Xiang et al.* [49].

Symplekin wurde erst nachträglich der 3'-Prozessierung zugeordnet, erstmalig beschrieben wurde es aber im Zusammenhang mit *tight junctions* [74], [93], [114]. Es konnten Interaktionen mit CPSF73 und CPSF100 sowie zu CstF64 gezeigt werden, jedoch keine spezifische Aktivität, sodass Symplekin als Gerüstprotein (*scaffold*) zur Stabilisierung der anderen Faktoren beschrieben wird [74]. N-Terminal liegen sieben antiparallele α -Helices, welche eine sogenannte HEAT-Domäne bilden. Dieser interagiert direkt mit der Phosphatase Ssu72, welche auch der 3'-Prozessierung zugeschrieben wird, und der C-terminalen-Domäne der RNA-Polymerase II [49], [115]. Darauffolgend liegt eine Region, die mit der Scharnier-Domäne von CstF64 interagieren kann. Diese Interaktion ist kompetitiv zur Bindung von CstF77 [116]. Zentral im Protein sitzt die Interaktionsregion mit CPSF73 [49]. Diese Interaktionsregion konnte aufgrund der Sequenzkonservierung zwischen den Hefeproteinen Ysh1 und Pta1 identifiziert werden. Strukturell ist nur die HEAT-Domäne von Symplekin aus Mensch und Taufliede (*Drosophila melanogaster*) mittels Röntgenkristallographie aufgeklärt [117]–[119]. Hierbei wurde auch die Phosphatase Ssu72 kokristallisiert.

1.2.5 Poly(A)-Polymerase und Poly(A)-bindendes Protein PABPN1

Es sind drei kanonische Poly(A)-Polymerasen (PAP α , PAP β und PAP γ) in Säugetieren bekannt, von denen die Poly(A)-Polymerase PAP α am besten charakterisiert ist. Die Poly(A)-Polymerase ist sowohl essentiell für die Spaltung der prä-mRNA als auch für die Polyadenylierung, wo sie die enzymatische Aktivität vermittelt. Die Poly(A)-Polymerase gehört zur DNA-Polymerase- β -Familie und die ersten 500 Aminosäuren sind in allen Eukaryoten stark konserviert [120]–[122]. In Wirbeltieren besitzt die Polymerase eine C-terminale Verlängerung, die aufgrund von alternativem Spleißen in ihrer Länge variieren kann und die einen hohen Gehalt an Serinen und Threoninen aufweist, jedoch für die Polyadenylierung nicht essentiell ist [121], [123]. Die räumlichen Strukturen des N-Terminus der Poly(A)-Polymerase der homologen Proteinen aus Hefe und Rind konnten aufgeklärt werden [122], [124].

Das nukleäre Poly(A)-bindende Protein PABPN1 wird sowohl für die Stimulation der Poly(A)-Polymerase als auch für die Längenkontrolle des Poly(A)-Schwanzes während der Polyadenylierung benötigt [18], [29]. Nach der Addition von ungefähr 250 Adenosylresten wird die Interaktion zwischen mPSF und der Poly(A)-Polymerase getrennt und die Polymerase verliert

ihre Prozessivität [29]. Die Struktur von PABPN1 unterscheidet sich deutlich von seinen zytosolischen Gegenspielern PABPC. Der N-Terminus ist reich an Glutaminsäure und somit sehr sauer. Darauf folgt eine *coiled-coil*-Domäne, welche für die Stimulation von PAP benötigt wird. Hieran unmittelbar anschließend ist die RNA-erkennende Domäne (RRM), welche in der Kristallstruktur eine homodimere Struktur zeigte [125]. C-terminal liegt eine basische, Arginin-reiche Region. Sowohl die C-terminale Region als auch die RRM sind für die Bindung des Poly(A)-Schwanzes verantwortlich, wobei ungefähr 10 Nukleotide von einer Kopie PABPN1 gebunden werden [126], [127]. Strukturell konnte bisher nur die RRM mittels Röntgenkristallographie aufgeklärt werden [125].

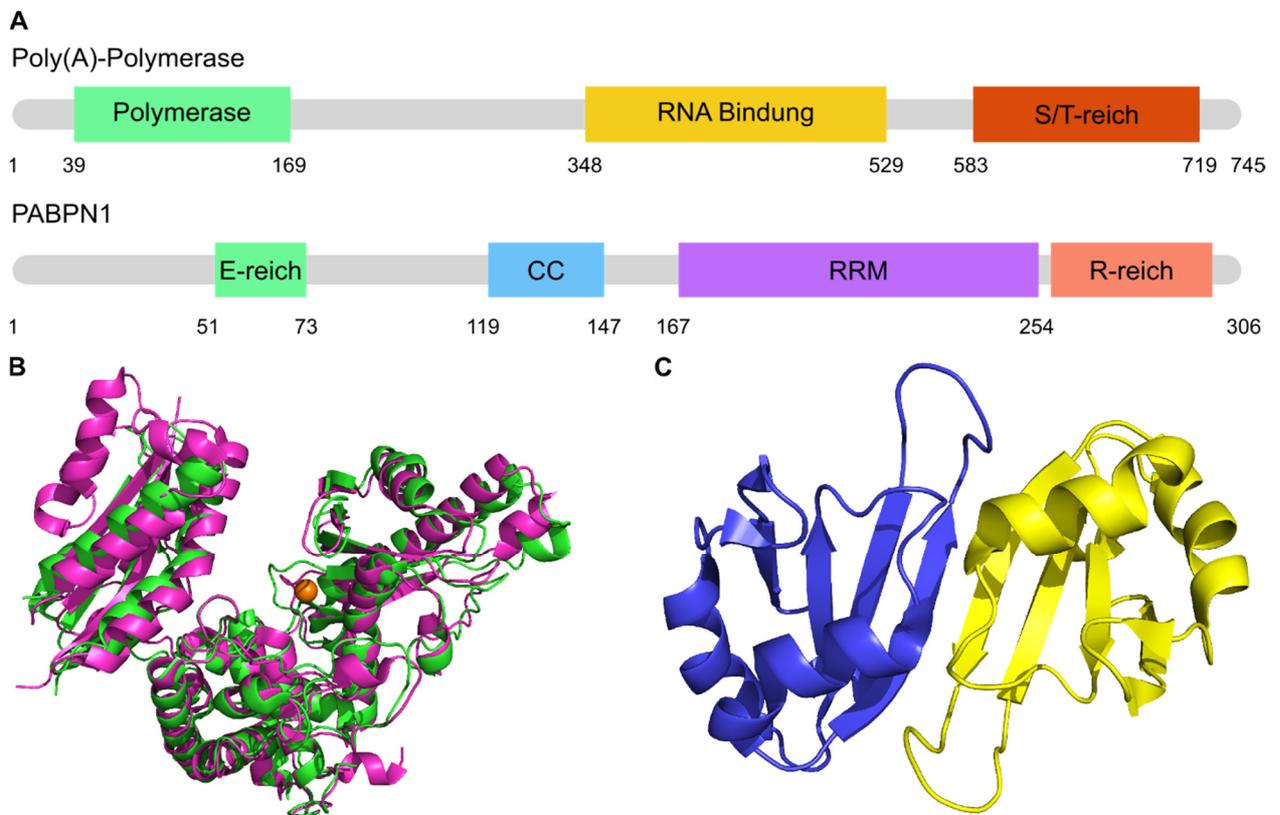


Abbildung 1.8 Struktur der Poly(A)-Polymerase und des Poly(A)-bindenden Proteins PABPN1. A) Maßstabgerechte Darstellung der bekannten Domänen der Poly(A)-Polymerase und von PABPN1. Das katalytische Zentrum (Polymerase) liegt N-terminal. Zentral im Protein ist die RNA-bindende Domäne. C-terminal liegt eine Serin- und Threonin-reiche Region, welche für die Polyadenylierung jedoch nicht benötigt wird. PABPN1 besitzt von N- zu C-Terminus eine saure, Glutaminsäure-reiche Region (E-reich), eine coiled-coil-Region (CC), das für die Dimerisierung und RNA-Bindung verantwortliche RNA-Erkennungsmotiv (RRM) sowie eine Arginin-reiche Region (R-reich; Aminosäuren 259 bis 298). B) Kristallstrukturen der homologen Proteine aus Hefe (magenta; PDBID: 1FA0) und Rind (grün; PDBID 1F5A) mit gebundenem Manganion (orange Sphäre). C) Kristallstruktur der dimerisierten RRM (PDBID: 3B4D). Die monomeren Untereinheiten sind farblich getrennt (blau/gelb). Abbildung in Anlehnung an *Xiang et al.* [49].

1.3 Struktur von Biomolekülen

Die Bestandteile einer Zelle sind stark strukturiert und geordnet. Vor allem RNAs sowie Proteine sind sehr variabel in ihrer Struktur und die Funktion ist von der korrekten, nativen Faltung abhängig. Die Struktur wird hierbei durch eine Vielzahl nicht-kovalenter Wechselwirkungen wie Wasserstoffbrücken, van-der-Waals-Kräfte, hydrophobe Interaktionen und π - π -Wechselwirkungen zwischen aromatischen Bestandteilen fixiert.

Die Aufklärung der Struktur ist elementar für das Verständnis der Funktion eines Biomoleküls. Die Strukturaufklärung kann in verschiedene Bereiche, abhängig von ihrer Auflösung, eingeteilt werden. Eine Übersicht ist in Abbildung 1.9 dargestellt. Unterhalb von 5 Å werden Strukturen als hochauflösend bezeichnet. Hierbei ist es möglich, die Elektronendichte in Röntgenkristallographie bzw. die Coulomb-Potentialdichte in Elektronenmikroskopie einzelnen Atomen zuzuordnen. Die niedrig auflösenden Methoden können weiterhin in direkte und indirekte Methoden differenziert werden. Während SAXS (engl. *small angle X-ray scattering*, Kleinwinkel-Röntgenstreuung) sowie Kryoelektronentomographie als direkte Methoden die Oberfläche des Biomoleküls wiedergeben, werden durch chemische Vernetzung und Massenspektrometrie Aminosäureseitenketten identifiziert, die sich in räumlicher Nähe befinden, und somit indirekt die Oberfläche wiedergeben. Im folgenden Abschnitt wird dieses weiter ausgeführt.

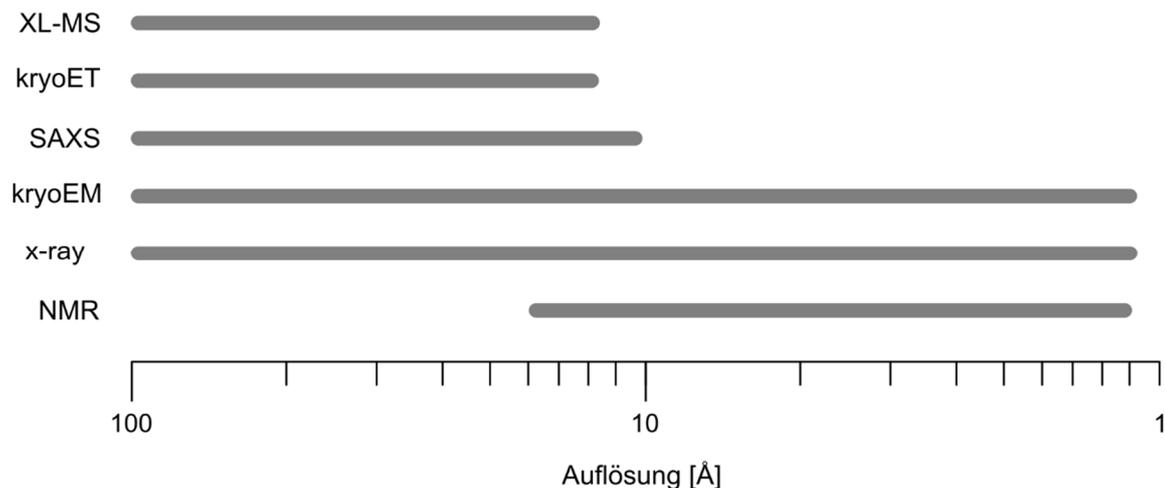


Abbildung 1.9 Übersicht der Auflösung strukturgebender Methoden. Die Abkürzungen der Methoden stehen in absteigender Reihenfolge für: Chemisches *cross-linking* und Massenspektrometrie (XL-MS), Kryoelektronentomographie (kryoET), Röntgen-Kleinwinkelstreuung (SAXS; *small-angle x-ray scattering*), Kryoelektronenmikroskopie (kryoEM), Röntgenkristallographie (x-ray) und Kernspinresonanz (NMR). Der Auflösungsbereich ist logarithmisch in Ångström dargestellt.

1.3.1 Cross-linking und Massenspektrometrie (XL-MS)

Um die Struktur von Proteinen und Proteinkomplexen genauer zu untersuchen, kann man diese mit einem chemischen Reagenz, dem *cross-linker*, kovalent verknüpfen. *Cross-linker* besitzen üblicherweise zwei reaktive Gruppen, getrennt von einem Arm (*linker*), welcher die maximale Entfernung zwischen zwei reagierenden Aminosäuren definiert. Die am häufigsten verwendete reaktive Gruppe ist der *N*-Hydroxysuccinimid-Ester (NHS-Ester), welcher bevorzugt mit freien Aminen, wie an Lysinseitenketten oder N-Termini von Proteinen, aber mit reduzierter Reaktivität auch mit Hydroxylgruppen, wie an Serin-, Threonin- und Tyrosinseitenketten, reagiert [128].

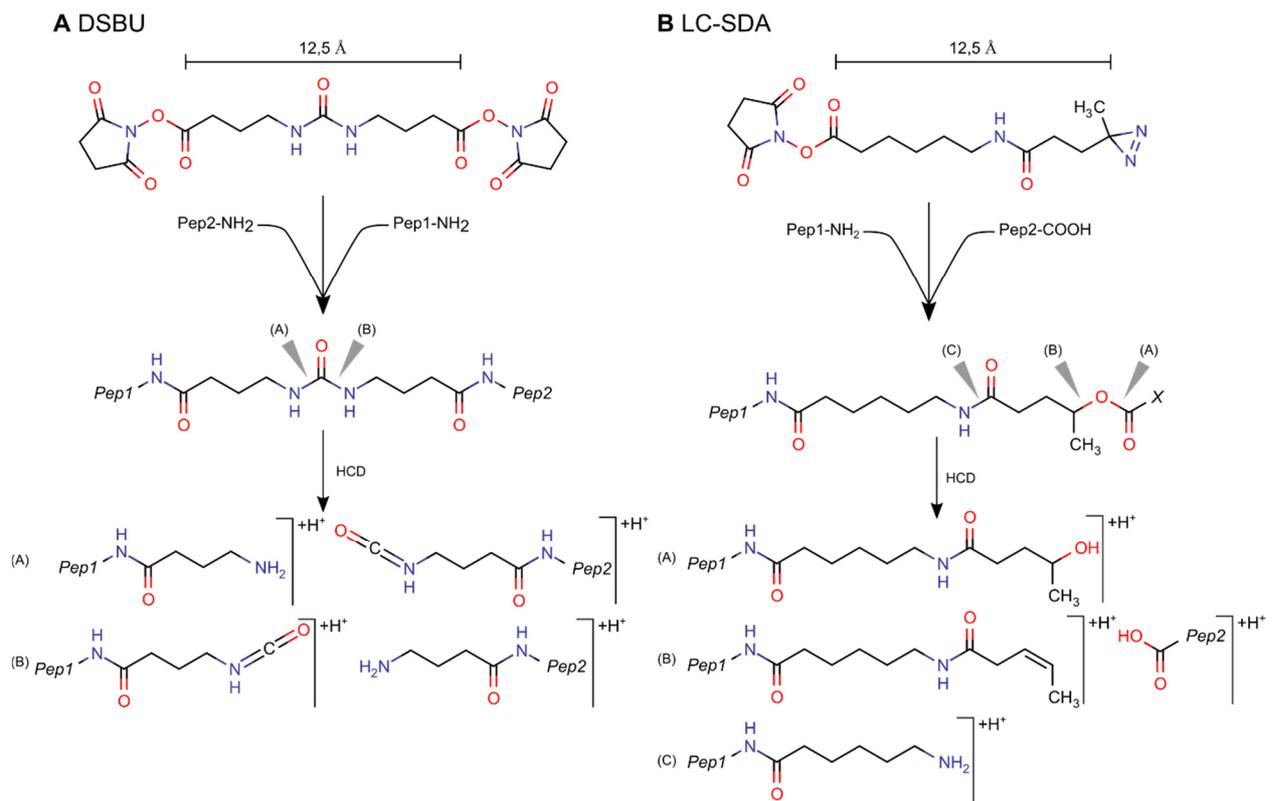


Abbildung 1.10 Struktur und Reaktionsschema der verwendeten *cross-linker*. A) Der *cross-linker* DSBU ist homobifunktional und reagiert vor allem mit freien Aminen. Die zentrale Harnstoffgruppe kann während der massenspektrometrischen Analyse asymmetrisch fragmentieren, angezeigt durch die farbigen Pfeile. Das gebildete Ionenpaar hat einen Masseunterschied von 26 u. B) Der *cross-linker* LC-SDA ist heterobifunktional. Der NHS-Ester reagiert vor allem mit freien Aminen während die Diazirin-Gruppe UV-induziert bevorzugt mit Karbonsäuren reagiert. Der *cross-linker* kann an primär an zwei Stellen fragmentieren: (A) Spaltung im Karbonsäureester. Die Modifizierung an Peptid1 beträgt 213,14 u, während das Keten an Peptid2 instabil ist und spontan zerfällt. (B) Spaltung hinter dem Karbonsäureester. Die Modifizierung an Peptid1 beträgt 195,12 u, während das Peptid2 in seiner Masse unmodifiziert ist. Nachfolgend der ersten Spaltung des *cross-linkers* kann eine weitere Spaltung an Position (C) stattfinden. Die Modifizierung an Peptid1 beträgt in diesem Fall nur noch 113,08 u. Abbildung angelehnt an *Iacobucci und Götze et al.* sowie *Iacobucci et al.* [128], [129].

Anschließend erfolgt eine proteolytische Spaltung der Proteinproben und eine massenspektrometrische Analyse. Hierbei ist eine sehr hohe Massengenauigkeit nötig, um die quervernetzten Peptide eindeutig zu identifizieren. Zur einfacheren Identifizierung können *cross-linker* verwendet werden, die während der Analyse fragmentiert werden können, wobei man ein spezifisches Reporterionenpaar erhält.

Die in dieser Arbeit verwendete *cross-linker* sind Disuccinimidyl-Dibutyrat-Harnstoff (DSBU; Abbildung 1.10A) und Succinimidyl-6-(4,4'-Azipentan-Amido)-Hexanoat (LC-SDA; Abbildung 1.10B). Während DSBU homobifunktional ist und mit primären Aminen und Alkoholen reagiert, ist LC-SDA heterobifunktional und kann über seinen NHS-Ester ebenfalls mit primären Aminen und Alkoholen sowie über die Diazirin-Gruppe mit Karbonsäuren reagieren (Abbildung 1.10B) [129], [130]. Beide *cross-linker* sind mittels Tandem-MS spaltbar und bilden charakteristische Ionenpaare [130]. Die Länge beider beträgt rund 12 Å, sodass unter Berücksichtigung der Flexibilität eines Proteins die maximale Entfernung zwischen den C α -Atomen zweier verknüpfter Lysinen zwischen 25 bis 34 Å beträgt [131].

1.4 Zielstellung dieser Arbeit

Die 3'-Prozessierung ist ein seit langem untersuchter und elementarer Prozess von eukaryotischen Zellen. Es konnten sehr viele Proteine identifiziert werden, die direkt oder indirekt an diesem Prozess beteiligt sind. Während die 3'-Prozessierung in Bäckerhefe biochemisch und strukturell gut verstanden ist, ist dieser Prozess in Säugerzellen aufgrund der höheren Komplexität immer noch unklar. Zwar konnten die Kernkomponenten der Spaltung identifiziert werden, eine detaillierte strukturelle Beschreibung steht aber für viele noch aus. Fortschritte im Bereich der strukturgebenden Methoden, vor allem die „*resolution revolution*“ in der Kryoelektronenmikroskopie, werden zukünftig hilfreich sein, die Polyadenylierungsfaktoren und die Teilschritte der Polyadenylierung strukturell darzustellen. Zur strukturellen Analyse dieser Teilschritte bedarf es jedoch des biochemischen Verständnisses dieses Prozesses. Der erste Teilschritt der Reaktion, die Spaltung der prä-mRNA konnte bisher jedoch nicht *in vitro* rekonstituiert werden. Offen ist die Frage, ob weitere, bisher nicht bekannte Faktoren an der Reaktion beteiligt sind, oder ob die bekannten Faktoren ausreichend, jedoch in ihrer rekombinanten Form inaktiv sind. Beispielhaft ist hier die Struktur der Endonuklease, welche eine geschlossene, inaktive sowie eine geöffnete, aktive Konformation einnehmen kann. Dieses wurde erst kürzlich in Hefe entdeckt und es ist davon auszugehen, dass es in Säugerzellen einer ähnlichen Aktivierung der Endonuklease bedarf.

Im ersten Teil dieser Arbeit sollten die bekannten Polyadenylierungsfaktoren gereinigt, biochemisch analysiert und in Kooperation mit Prof. Andrea Sinz massenspektrometrisch untersucht werden. Hierzu konnte auf die bisherigen Arbeiten der Arbeitsgruppe Wahle zurückgegriffen werden, die die Expression und die Reinigung für die meisten

Hauptkomponenten bereits etabliert hat. Die gereinigten Komplexe sollten chemisch quervernetzt werden und massenspektrometrisch analysiert werden. Die hierbei gewonnenen Informationen wurden genutzt, um die Interaktionsregionen innerhalb der Komplexe qualitativ zu beschreiben. Dieses ermöglicht ein Grundverständnis der Architektur dieser Komplexe, welche für das Verständnis der Aktivität essentiell ist. Da für den mPSF-Komplex, Kern des zweiten Schrittes der 3'-Prozessierung, bereits hochauflösende Strukturen publiziert wurden, nicht aber von der Interaktion zwischen mPSF und Poly(A)-Polymerase, sollten hier mithilfe von *cross-linking* die Dynamik der Polyadenylierung näher beleuchtet werden. Hierzu sollten Bedingungen etabliert werden, in denen die Reaktionsintermediate der Polyadenylierung getrennt und chemisch vernetzt werden können.

Im zweiten Teil der Arbeit sollten die experimentellen Daten des *cross-linkings* für eine computergestützte Modellierung der humanen Komplexe genutzt werden. Hierzu sollten homologe Strukturen identifiziert und für die Generierung von Submodellen genutzt werden. Diese Strukturen sollten dann unter Verwendung der experimentellen *cross-link*-Daten zusammengefügt werden. Neben der direkten Verwendung dieser Daten sollten zusätzlich neue Entfernungsbegrenzungen entwickelt werden, um die Genauigkeit der generierten Modelle weiter zu verbessern.

Im dritten Teil der Arbeit sollten für alle gereinigte Komplexe Interaktionspartner identifiziert werden, wobei die gereinigten Komplexe als Köderproteine in *pull-down*-Experimente eingesetzt werden sollten. Hierbei wurde kommerzieller Kernextrakt verwendet, der in der Lage ist, *in vitro* prä-mRNA zu spalten und somit alle benötigten Proteinfaktoren enthält. Zur Analyse dieser Experimente sollten die Elutionsfraktionen massenspektrometrisch analysiert werden, um eine möglichst breite und genaue Identifizierung von Interaktionspartnern zu ermöglichen.

2. Material und Methoden

2.1 Material

2.1.1 Chemikalien

Chemikalien wurden in p.A. Qualität von der Firma Merck bezogen. Ausnahmen hiervon sind im Anhang in Tabelle A 1 aufgelistet.

2.1.2 Geräte

Für die Anfertigung dieser Arbeit wurde Standard-Laboraüstung verwendet. Besondere Geräte, die darüber hinaus nötig waren, sind folgend aufgelistet:

Gerät	Hersteller
Odyssey Classic (Fluoreszenz-Scanner)	LI-COR Bioscience
SLM AMINCO <i>French Press</i>	SLM Instruments
Sonifier 250 (Ultraschallgerät und -sonde)	Branson
Storm 860 Molecular Imager (Phosphoimager)	Amersham Biosciences
Tri-Carb 2100TR (Szintillationszähler)	Packard
UJ30 (UV-Dioden-Steuergerät)	Panasonic
UV-Diode ANUJ6186 (365 nm; 17200 mW/cm ²)	Panasonic

Für die massenspektrometrischen Analysen wurden folgende Geräte verwendet:

HPLC-System	
Gerät	Hersteller
UltiMate™ 3000 RSLCnano HPLC System	Thermo-Fischer

Chromatographie-Material	
Material	Hersteller
Acclaim™ PepMap™ 100 C18 rp, 300 µm x 5 mm, 5 µm, 100 Å (Vorsäule)	Thermo-Fischer
Acclaim™ PepMap™ 100 C8 rp, 300 µm x 5 mm, 5 µm, 100 Å (Vorsäule)	Thermo-Fischer
PicoFrit™ self-pack, 75 µm x 250 mm, tip ID 15 µm (Leersäule)	New Objective
PicoFrit™ self-pack, 75 µm x 500 mm, tip ID 15 µm (Leersäule)	New Objective
ReproSil-Pur 120 C18-AQ, 1.9 µm, 120 Å (Säulenmaterial)	Dr. Maisch GmbH
Acclaim™ PepMap™ 100 C18 rp, 75 µm x 250 mm, 2 µm, 100 Å	Thermo-Fisher

Massenspektrometer	
Gerät	Hersteller
Orbitrap Fusion™ Tribrid™ (Massenspektrometer)	Thermo-Fisher
Q-Exactive™ Plus (Massenspektrometer)	Thermo-Fisher
Nanospray Flex™ (Elektrospray-Ionisationsquelle)	Thermo-Fisher

2.1.3 Zelllinien und Kulturmedien

2.1.3.1 Bakterien

Für die Kultivierung von *E. coli*-Zellen wurde LB-Medium (Roth) und zur Proteinexpression TB-Medium verwendet.

In dieser Arbeit wurden XL1-Blue, DH10B und Rosetta *E. coli*-stämme verwendet. Der genaue Genotyp der verwendeten Bakterienstämme ist im Anhang in Tabelle A 2 dokumentiert.

TB-Medium

24 g/l Hefeextrakt

20 g/l Trypton

4 ml/l Glycerin

17 mM Kaliumdihydrogenphosphat¹

72 mM Dikaliumhydrogenphosphat¹

¹ Die Phosphatsalze wurden erst unmittelbar vor der Inokulation mit den Bakterien hinzugegeben.

2.1.3.2 Eukaryotische Zelllinien

Für die Kultivierung der Insektenzellen wurde das serumfreie Medium EX-CELL 420 (Sigma-Aldrich) verwendet. Für die Viruspropagation sowie die Proteinexpression wurde folgende Zelllinie verwendet:

Bezeichnung	Vollständiger Name	Herkunft
Sf-21	IPLB-SF-21-AE	Immortalisiert aus Ovarzellen von <i>Spodoptera frugiperda</i> [132]; Invitrogen

2.1.4 DNA-Konstrukte und Baculoviren

2.1.4.1 Primer

Für die DNA-Konstrukte zur Proteinexpression in *E. coli* und Insektenzellen wurden folgende Primer für die Klonierung verwendet. Der zum reversen DNA-Strang komplementäre Primer ist mit „for“ (*forward*), der zum kodieren DNA-Strang komplementäre Primer mit „rev“ (*reverse*) benannt.

Primer	
Name	Sequenz
466 HisSumo-Pcf11zcx for	ATATATGGTCTCTTGGTATTCAGTGTTAC
467 Pcf11zcx rev ¹	TATATGCGCCGCTTATGAAGATGTATTTTG
494 GST-Pcf11zcx for	ATATATGGATCCATTCAGTGTTACTCTTG
481 Pcf11ΔN769 for	ATATAGTCGACATGTTTGAAGGACCCAATAAATTAAGCC
483 Pcf11ΔN1340 for	ATATAGTCGACATGGGTATTCAGTGTTACTCTTGTGG
501 Pcf11ΔN1125 for	ATATAGTCGACATGCTTGAATCAGTATCTTTC
484 Pcf11ΔN rev ²	TATATTCTAGATTAACTGACTCGACTGTGTC

¹ Dieser Primer wurde sowohl für die Klonierung von GST-Pcf11zcx als auch HisSUMO-Pcf11zcx verwendet. ² Dieser Primer wurde für alle N-terminalen Deletionsvarianten von Pcf11 verwendet.

2.1.4.2 Polynukleotid-Substrate

Zur Charakterisierung der Proteine wurden folgende Substrate verwendet. Die kommerziell erworbenen RNA-Substrate wurden hierbei 5' markiert und die DNA-Substrate dienten als Template der *in vitro* Transkription.

Nukleotid-Substrate	
Name	Hersteller / Bereitgestellt von
RNA-Oligo A ₁₄	biomers.net
RNA-Oligo C ₁₄	biomers.net
RNA-Oligo G ₁₄	biomers.net
RNA-Oligo G ₁₄	biomers.net
RNA-Oligo <i>CFI 2sites</i> ¹	biomers.net
pSP64-L3 <i>DraI</i> -verdaut ¹	Lars Schönemann / Uwe Kühn [133]
pSP64-L3 Δ <i>DraI</i> -verdaut ¹	Lars Schönemann / Uwe Kühn [133]
pSP64-L3prä-A15, <i>BbsI</i> -verdaut ¹	Gudrun Scholz / Uwe Kühn [30]
pSP64-L3prä-A15 Δ, <i>BbsI</i> -verdaut ¹	Gudrun Scholz / Uwe Kühn [30]

¹ Die Sequenzen sind in Anhang Tabelle A 3 dargestellt.

2.1.4.3 Plasmide

Für die Proteinexpression in *E. coli* sowie für die Integration der Expressionskonstrukte in die Bacmid-DNA zur Expression in Insektenzellen wurden folgende Plasmide verwendet:

Plasmide für die Expression in <i>E. coli</i>		
Plasmid	Gen	Herkunft
pGEX6p1	-	GE Healthcare
pGEX6p1-Pcf11zcx	GST-Pcf11zcx	Diese Arbeit
pET-SUMOadapt	-	<i>Bosse-Doenecke et al.</i> [134]
pET-SUMOAdapt-Pcf11zcx	HisSUMO-Pcf11zcx	Diese Arbeit

Transferplasmide für die Integration in Bacmid-DNA		
Plasmid	Gene	Herkunft
pFL-HisClp1	His-Clp1	Sandra Grund / Uwe Kühn
pFL-HisClp1-StrepPcf11	His-Clp1, Strep-Pcf11	Peter Schäfer
pFL-HisClp1-Pcf11(770-1555)	His-Clp1, Pcf11ΔN769	Diese Arbeit
pFL-HisClp1-Pcf11(1124-1555)	His-Clp1, Pcf11ΔN1123	Diese Arbeit
pFL-HisClp1-Pcf11(1340-1555)	His-Clp1, Pcf11ΔN1339	Diese Arbeit
pFL-HisFlagClp1-Pcf11(770-1555)	HisFLAG-Clp1, Pcf11ΔN769	Diese Arbeit
pFL-HisFlagClp1-Pcf11(1124-1555)	HisFLAG-Clp1, Pcf11ΔN1123	Diese Arbeit

2.1.4.4 Bacmide und Baculoviren

Übersicht der in dieser Arbeit verwendeten Bacmide und Baculoviren.

Bacmide		
Name	Gene	Herkunft
pBac(pFBDM-CstF) ^{Tn7} -mCherry ^{loxP}	CstF64, StrepII-CstF50, His-CstF77	Peter Schäfer / Matthias Lorbeer
pBac YFP ^{loxP} mPSF ^{Tn7}	CPSF1, CPSF4, Fip1, MycHis-WDR33	Lars Schönemann
pBAC-YFP ^{loxP} -(pFL-HisClp1-Pcf11(770-1555)) ^{Tn7}	His-Clp1, Pcf11ΔN769	Diese Arbeit
pBAC-YFP ^{loxP} -(pFL-HisClp1-Pcf11(1124-1555)) ^{Tn7}	His-Clp1, Pcf11ΔN1123	Diese Arbeit
pBAC-YFP ^{loxP} -(pFL-HisClp1-Pcf11(1340-1555)) ^{Tn7}	His-Clp1, Pcf11ΔN1339	Diese Arbeit
pBAC-YFP ^{loxP} -(pFL-HisFlagClp1-Pcf11(770-1555)) ^{Tn7}	HisFLAG-Clp1, Pcf11ΔN769	Diese Arbeit
pBAC-YFP ^{loxP} -(pFL-HisFlagClp1-Pcf11(1124-1555)) ^{Tn7}	HisFLAG-Clp1, Pcf11ΔN1123	Diese Arbeit
Baculoviren		
Virus	Gene	Herkunft
pBac-YFP- ^{loxP} -CFI ^{Tn7}	His-CFI25, Strep-CFI68	Mathias Lorbeer
pBac-mCherry ^{loxP} -HF-CstF64 ^{Tn7}	HisFLAG-CstF64	Moritz Schmidt
pBac-mPSF-untagged ^{Tn7}	CPSF1, CPSF4, Fip1, WDR33	Anne Hofmeister / Uwe Kühn
pBac-mPSF tagged WDR33 ^{Tn7}	CPSF1, CPSF4, Fip1, MycHis-WDR33	Diese Arbeit
pBac-healed StrepII-CM3 ^{loxP}	CPSF2, CPSF3, Symplekin	Anne Hofmeister / Uwe Kühn
pBac-YFP- ^{loxP} -CFIIm-481 ^{Tn7}	His-Clp1, Pcf11ΔN769	Diese Arbeit
pBac-YFP- ^{loxP} -CFIIm-483 ^{Tn7}	His-Clp1, Pcf11ΔN1339	Diese Arbeit
pBac-YFP- ^{loxP} -CFIIm-TLong ^{Tn7}	His-Clp1, Pcf11ΔN1123	Diese Arbeit
pBac-CstF ^{Tn7}	CstF64, StrepII-CstF50, His-CstF77	Diese Arbeit
pBac-YFP- ^{loxP} -CFIIm-HF-Tlong ^{Tn7}	HisFLAG-Clp1, Pcf11ΔN1123	Diese Arbeit
pBac-YFP- ^{loxP} -CFIIm-HF-FEGP ^{Tn7}	HisFLAG-Clp1, Pcf11ΔN769	Diese Arbeit

2.1.5 Enzyme, Proteine und Zellextrakte

Im Rahmen dieser Arbeit wurden folgende Proteine und Proteinkomplexe der mRNA 3'-Prozessierung detailliert charakterisiert.

Proteine und Proteinkomplexe	Hersteller / Bereitgestellt von
4 ⁺ CstF (FLAG-CstF64tau, CstF64, StrepII-CstF50, His-CstF77)	Moritz Schmidt
Bovine Poly(A)-Polymerase	Uwe Kühn
Boviner CPSF-Komplex	Uwe Kühn
CFI (Strep-CFI68, His-CFI25)	Diese Arbeit
CFII (His-Clp1, Strep-Pcf11)	Diese Arbeit
CFIIΔN1123 (His-Clp1, Pcf11ΔN1123)	Diese Arbeit
CFIIΔN1339 (His-Clp1, Pcf11ΔN1339)	Diese Arbeit
CFIIΔN769 (His-Clp1, Pcf11ΔN769)	Diese Arbeit
CstF (His-CstF77, CstF64, Strep-CstF50)	Diese Arbeit
Endonuklease-Komplex (Symplekin, StrepII-CPSF3, CPSF2, HisFLAG-CstF64)	Diese Arbeit
GST-Pcf11zcz	Diese Arbeit
mPSF (CPSF1, CPSF4, MycHis-WDR33, Fip1)	Diese Arbeit
HF-CFIIΔN1123 (HisFLAG-Clp1, Pcf11ΔN1123)	Diese Arbeit
HF-CFIIΔN769 (HisFLAG-Clp1, Pcf11ΔN769)	Diese Arbeit
Nukleares Poly(A)-bindendes Protein (PABPN1)	Uwe Kühn
SUMO-Pcf11zcz	Diese Arbeit

Darüber hinaus wurden weitere, z.T. kommerziell verfügbare, Proteine, Antikörper und Extrakte verwendet. Diese sind im Anhang in Tabelle A 4 tabelliert.

2.1.6 Software

Die zur Erstellung benötigte sowie im Rahmen dieser Arbeit geschriebene Software ist folgend aufgelistet.

Software	Version	Verwendung	Anbieter/Entwickler/ Bereit gestellt von
Bash	4.4.19	Laufzeitumgebung der in dieser Arbeit entwickelten Software	Free Software Foundation
BLASTp	2.9.0 (Webserver)	Identifizierung von strukturhomologen Proteinen	National Center for Biotechnology Information (NCBI)
Chimera	1.13.1	Visualisierung von Dichtekarten	University of California, San Francisco [135]
DisVis	-(Webserver) ¹	Analyse von Konformationsräumen in Proteinkomplexen	Utrecht Bioinformatics Center [136]
HADDOCK	2.2 (Webserver)	Modellierung von Proteinkomplexen	Utrecht Bioinformatics Center [137], [138]

HHPRED	3.2.0 (Webserver)	Identifizierung von strukturhomologen Proteinen	Max-Planck-Institut für Entwicklungsbiologie [139]
I-TASSER	5.1 (Webserver)	Automatisierte Modellierung von Proteinstrukturen	University of Michigan [140]
ImageJ	1.52	Densitometrische Quantifizierung von Proteinen / Auswertung von Western-Blot-Analysen	Wayne Rasband [141]
JalView	2.11	Generierung und Visualisierung von Protein-Alignments	Biotechnology and Biological Sciences Research Council [142]
MaxQuant	1.6.0.1	Quantifizierung von MS-Daten	Max-Planck-Institut für Biochemie [143], [144]
MeroX	1.6.6	Identifikation von <i>cross-links</i>	Michael Götze [145]
MODELLER	9.21	Homologiemodellierung von Proteinstrukturen	University of California, San Francisco [146]
Proteome Discoverer	2.2	Konversion von massenspektrometrischen Rohdaten	Thermo-Fisher
PyMOL	1.8.4.0	Visualisierung von Proteinstrukturen	DeLano Scientific LLC
R	3.5.2	Statistische Datenanalyse	The R Foundation
SigmaPlot	12.5	Statistische Datenanalyse	Systat

In dieser Arbeit entstanden²

Name	Verwendung	Querverweis
ResLength.sh	Empirische Ermittlung der Länge von Aminosäureseitenketten	3.2.1.1
PepLength.sh	Datenbankweite Analyse von Peptidlängen von bis zu 10 Aminosäuren	3.2.1.2

¹ Keine Versionsnummer angegeben

² Die in dieser Arbeit entstandene Software wurde auf folgendem Betriebssystem getestet: Ubuntu 18.04.2 LTS (64-Bit)

2.2 Methoden

Standardmethoden wurden nach *Sambrook und Russel* durchgeführt [147]. Abweichungen hiervon sowie weitere Methoden werden im Folgenden ausführlicher beschrieben.

Die Präparation der Expressionskonstrukte, die Herstellung der Baculoviren sowie Durchführung der Proteinexpression in Insektenzellen wurden nach dem Benutzerhandbuch des Herstellers durchgeführt [148].

2.2.1 Proteinbiochemische Methoden

2.2.1.1 Proteinexpression

2.2.1.1.1 Proteinexpression in *E. coli*

Für die Proteinexpression wurden ausschließlich *E. coli*-Rosetta™, ein BL21 Stammderivat, verwendet. Die transformierten und selektierten Bakterien wurden in TB-Medium mit 20 µM ZnCl₂ bei 37 °C kultiviert, bis eine Optische Dichte (OD) von 1 erreicht wurde. Anschließend wurde durch Zugabe von 1 mM IPTG die Proteinexpression induziert und 3 Stunden bei 25 °C durchgeführt. Die Zellen wurden durch Zentrifugation (15 Minuten, 4000xg) pelletiert und das Zellpellet in flüssigem Stickstoff eingefroren und bis zur Proteinreinigung bei -80 °C gelagert.

2.2.1.1.2 Proteinexpression in Insektenzellen

Für die Proteinexpression in Insektenzellen wurden Sf21-Zellen verwendet. Nach der Transfektion mit den entsprechenden Baculoviren wurden die Zellen 48 bis 96 Stunden unter Standardbedingungen kultiviert. Die Zellen wurden, sobald ihre Größe auf das 1,5fache angeschwollen und die lebendigen Zellen auf ca. 80 bis 90 % der Gesamtzellzahl gefallen waren, geerntet. Hierzu wurden die Zellen durch Zentrifugation (2 Minuten, 500xg) pelletiert, in flüssigem Stickstoff eingefroren und bis zur Proteinreinigung bei -80 °C gelagert.

2.2.1.2 Zellyse

Zum Zellaufschluss wurden sowohl *E. coli*- als auch die Sf21-Zellen in 1:20 des Kulturvolumens in Lysispuffer resuspendiert. Für Bakterien erfolgte der Zellaufschluss mittels *French Press* (2x 70 bar; SLM AMINCO *French Press*), für Insektenzellen mittels Sonifikation (3x 20 Sekunden; 50 % Intensität; 1 Minute Pause; Branson Sonifier 250). Nach erfolgtem Zellaufschluss wurde das Lysat durch Zentrifugation (30 Minuten, 20000xg) von Zelltrümmern und unlöslichen Bestandteilen geklärt.

Lysispuffer

50 mM Triethanolamin • HCl, pH 8,0
 200 mM Kaliumchlorid
 10 % (w/v) Saccharose
 1 mM DTT
 1 µg/ml Leupeptin
 1 µg/ml Pepstatin
 1 mM PMSF

2.2.1.3 Proteinreinigung

2.2.1.3.1 Affinitätschromatographie

NiNTA-Agarose

Zur Reinigung von Proteinen mit Hexahistidin-*tag* wurden diese pro 15 ml Lysat mit 1 ml, mit Lysispuffer äquilibrierter, NiNTA-Agarose 2 Stunden lang im Kühlraum unter Drehung in einem 50 ml Röhrchen inkubiert. Anschließend wurde das Säulenmaterial in eine Leersäule gegeben. Die Agarose wurde 6x mit 2 Säulenvolumina (CV) Waschpuffer zur Entfernung unspezifischer Proteine gewaschen und das Zielprotein mit 6x 0,5 CV Elutionspuffer eluiert.

Waschpuffer	Elutionspuffer
50 mM Triethanolamin • HCl, pH 8,0	50 mM Triethanolamin • HCl, pH 8,0
200 mM Kaliumchlorid	200 mM Kaliumchlorid
10 % (w/v) Saccharose	10 % (w/v) Saccharose
1 mM DTT	1 mM DTT
20 mM Imidazol	250 mM Imidazol

GSH-Sepharose

Die Reinigung von Proteinen mit GST-tag erfolgte analog zu der Proteinreinigung mit Hexahistidin-tag. Als Matrix wurde hierbei GSH-Sepharose verwendet, und die Proteine wurden mit Glutathion-haltigem Puffer eluiert.

Waschpuffer	Elutionspuffer
50 mM Triethanolamin • HCl, pH 8,0	50 mM Triethanolamin • HCl, pH 8,0
200 mM Kaliumchlorid	200 mM Kaliumchlorid
10 % (w/v) Saccharose	10 % (w/v) Saccharose
1 mM DTT	1 mM DTT
	20 mM Glutathion (reduziert)

2.2.1.3.2 Anionenaustauschchromatographie

Zur Reinigung von Proteinen mittels Anionenaustauschchromatographie wurden die Elutionsfraktionen der Affinitätsreinigung mit salzfreiem Laufpuffer verdünnt, um eine Bindung an das Säulenmaterial zu gewährleisten. Die spezifischen Salzkonzentrationen für die gereinigten Komplexe sind in Tabelle 1 dargestellt.

Tabelle 1 Salzkonzentrationen der verschiedenen Proteinkomplexe für die Anionenaustauschchromatographie

Komplex	Salzkonzentration für die Anionenaustauschchromatographie
CFI	150 mM
CFII (und CFII Deletionsvarianten)	150 mM
CstF	75 mM
mPSF	150 mM
Endonuklease-Komplex	150 mM

Für alle Proteinkomplexe wurde eine 1 ml MonoQ-Säule verwendet. Geladen wurde die Hauptfraktion der Affinitätsreinigung. Proteinmengen und -ausbeuten werden in den Ergebnissen gezeigt. Diese wurde entsprechend dem Proteinkomplex mit Puffer der passenden Salzkonzentration äquilibriert. Die Flussgeschwindigkeit betrug 0,2 ml/min. Nach dem Laden des Proteins auf die Säule wurde diese mit einem Säulenvolumen Puffer gewaschen und

anschließend über 10 Säulenvolumen mit einem Salzgradienten bis 600 mM Kaliumchlorid eluiert. Es wurden 500 µl-Fractionen gesammelt.

MonoQ-Puffer 1	MonoQ-Puffer 2
50 mM Triethanolamin • HCl, pH 8,0	50 mM Triethanolamin • HCl, pH 8,0
10 % (w/v) Saccharose	1 M Kaliumchlorid
1 mM DTT	10 % (w/v) Saccharose
	1 mM DTT

2.2.1.4 Proteinanalyse

2.2.1.4.1 Polyacrylamidgel

SDS-Polyacrylamidgele wurden nach *Laemmli* durchgeführt [149]. Die Detektion von Proteinen erfolgte durch Färbung mit Coomassie-Blau oder Silber.

Coomassie-Blau-Färbung

Nach der Trennung der Proteine wurde das Polyacrylamidgel ca. 10 Minuten fixiert (Fixierlösung) und anschließend mind. 3-mal mit destilliertem Wasser gewaschen. Die Färbung erfolgte durch Erhitzen des Geles in der Färbelösung (ca. 30 Sekunden in einer Mikrowelle) und anschließender 10-minütiger Inkubation. Das Gel wurde über Nacht durch Inkubation in destilliertem Wasser entfärbt.

Fixierlösung
10 % (v/v) Ethanol
10 % (v/v) Essigsäure
Färbelösung
100 mg/l Coomassie G-250
0,3 % (v/v) Salzsäure

Silberfärbung

Zur Silberfärbung von Proteinen nach der Trennung im Polyacrylamidgel wurde das Polyacrylamid-Gel 5 Minuten fixiert (Fixierlösung), 5 Minuten mit destilliertem Wasser gewaschen und anschließend weitere 5 Minuten mit 50 % Aceton vorbehandelt. Es folgte die 1-minütige

Fixierlösung	Lösung 1
50 % (v/v) Aceton	0,167 µg/ml
1,25 % (w/v) Trichloressigsäure	Natriumthiosulfat
0,42 ‰ (v/v) Formaldehyd	
Lösung 2	Lösung 3
2,6 µg/ml Silbernitrat	20 µg/ml Natriumcarbonat
1 % (v/v) Formaldehyd	0,42 ‰ (v/v) Formaldehyd
	42 ng/ml Natriumthiosulfat

Sensibilisierung des Geles (Lösung 1). Danach wurde das Gel 3x mit destilliertem Wasser gewaschen und anschließend 8 Minuten in der Färbelösung inkubiert (Lösung 2). Nach weiterem Waschen des Geles mit Wasser wurde dieses entwickelt (Lösung 3). Waren die Proteinbanden ausreichend sichtbar, wurde die Entwicklung mit 12 % Essigsäure gestoppt, und das Gel konnte fotografiert werden.

2.2.1.4.2 Western-Blot

Zur Identifikation von Proteinen mittels Western-Blot wurden die Proteine nach Trennung im Polyacrylamidgel auf eine Nitrozellulosemembran transferiert. Hierzu wurden das Gel und die Membran zwischen je zwei in Transferpuffer eingeweichte Filterpapiere in die *Blotting*-Apparatur gelegt. Zum Transfer der Proteine wurde eine Stunde eine elektrische Stromdichte von $1,5 \text{ A/cm}^2$ angelegt. Anschließend wurde die Membran mit 5 % Milchpulver in TBST blockiert (mind. 30 Minuten bei Raumtemperatur). Die Bindung des primären Antikörpers, welcher in 0,5 % Milchpulver in TBST verdünnt wurde, erfolgte über Nacht. Anschließend wurde die Membran mit TBST gewaschen (6x je 5 Minuten bei Raumtemperatur) und der sekundäre, fluoreszenzmarkierte Antikörper, 1:15000 verdünnt in TBST, für 2 Stunden mit der Membran inkubiert. Vor der Detektion der Fluoreszenz wurde die Membran nochmals gewaschen.

TBST

50 mM TRIS • HCl, pH 7,4
150 mM Natriumchlorid
0,05 % (v/v) Tween 20

Transferpuffer

25 mM TRIS
192 mM Glycin
20 % (v/v) Methanol

2.2.1.4.3 Edman-Abbau

Zum analytischen Edman-Abbau zur Proteinidentifikation wurde die Probe wie beschrieben mittels SDS-Polyacrylamidgel getrennt und auf eine PVDF-Membran transferiert. Abweichend zu der in 2.2.1.4.2 Western-Blot beschriebenen Methode wurde die PVDF-Membran vor der Verwendung in 100 % Methanol aktiviert. Nach dem Transfer wurde die Membran 5 Minuten gefärbt (0,2 % (w/v) Ponceau S in 1 % Essigsäure) und anschließend mit destilliertem Wasser entfärbt. Die zu analysierende Proteinbande wurde ausgeschnitten, getrocknet und von der Firma Proteome Factory AG (Berlin) analysiert.

2.2.1.4.4 Proteinkonzentrationsbestimmung

Zur Ermittlung der Proteinkonzentration wurden das Zielprotein sowie verschiedene BSA-Mengen mittels Polyacrylamid-Gel aufgetrennt und mit Coomassie eingefärbt und fotografiert. Die Proteinmenge wurde densitometrisch mit ImageJ quantifiziert [141]. Die Konzentrationsbestimmung erfolgte über eine auf demselben Polyacrylamidgel aufgetragene BSA-Eichkurve.

2.2.1.4.5 Chemische Methylierung von BSA

Das in den Aktivitätsanalysen eingesetzte BSA zur Stabilisierung der rekombinanten Proteine wurde zur Inaktivierung von kontaminierenden Nukleasen chemisch nach *Means und Feeney* methyliert [150]. Zur Reinigung des methylierten BSA wurde dieses über Nacht gegen destilliertes Wasser dialysiert und die Konzentration photometrisch bestimmt ($\epsilon = 47790 \text{ M}^{-1}\text{cm}^{-1}$).

2.2.2 Nukleinsäurebiochemische Methoden

Alle nukleinsäurebiochemischen Methoden, einschließlich der Aktivitätsanalysen, wurden mit DEPC-behandeltem Wasser und Puffern durchgeführt, um eine Nukleasekontamination zu vermeiden [151].

2.2.2.1 Denaturierendes Harnstoff-Polyacrylamid

Zur Trennung und Analyse von radioaktiven RNA-Molekülen wurde eine denaturierende Gelelektrophorese nach *Maniatis et al.* verwendet [152]. Abweichend hierzu wurden 8 M Harnstoff, statt wie beschrieben 7 M, verwendet. Die Detektion der RNA erfolgte autoradiographisch.

2.2.2.2 Markierung von RNA

2.2.2.2.1 *In vitro* Transkription

Un- sowie radioaktivmarkierte Substrat-RNAs wurden mittels *in-vitro*-Transkription hergestellt. Die hierbei verwendeten linearisierten DNA-Konstrukte lagen bereits vor. Die Transkription erfolgte mit der SP6 RNA-Polymerase (Roche) nach den Herstellerangaben [153]. Abweichend hierzu wurde das Reaktionsvolumen auf 40 µl und die Reaktionszeit auf 4 Stunden erhöht. Die anschließende Ethanol-Fällung und Gelreinigung wurde nach *Rio, et al.* durchgeführt [154].

2.2.2.2.2 5'-Markierung von RNA

Zur radioaktiven Markierung von RNA wurden die T4 Polynukleotidkinase (NEB) verwendet. Die Reaktion erfolgte nach den Herstellerangaben [155]. Die markierte RNA wurde anschließend durch Phenol-Chloroform-Extraktion gereinigt und in RNase-freiem Wasser aufgenommen.

2.2.3 Aktivitätsanalyse von Proteinen

2.2.3.1 Filterbindungsassay

Die Filterbindungsassays wurden nach *Kühn et al.* durchgeführt [156]. Hierzu wurde die verwendete RNA vor der Reaktion für 3 Minuten bei 95 °C aufgeköcht und anschließend auf Eis abgekühlt. Die Nitrozellulosefilter wurden mit Waschpuffer und 0,5 µg/ml Hefe-RNA äquilibriert.

Das Reaktionsvolumen betrug 50 µl und es wurden, wenn nicht anders angegeben, 50

Reaktionspuffer	Waschpuffer
50 mM Triethanolamin • HCl, pH 8,0	50 mM Triethanolamin • HCl, pH 8,0
100 mM Kaliumchlorid	100 mM Kaliumchlorid
0,2 U/µl RNase Inhibitor	10 % Glycerin
1 mM DTT	
10 % (w/v) Glycerin	
0,5 mM EDTA	
0,2 g/l BSA ^{met}	
0,01 % NP-40	

fmol RNA und steigende Mengen Protein eingesetzt. Die Bindungsreaktion fand bei Raumtemperatur über 20 Minuten statt. Die Salzkonzentration, die sich aus Reaktionspuffer und

dem in der Proteinlösung enthaltenem Salz zusammensetzte, betrug maximal 100 mM. Anschließend wurden die Reaktionen auf die Nitrozellulosefilter gegeben und mit 5 ml kaltem Waschpuffer gewaschen. Die gebundene RNA wurde über Szintillationsmessung der Filter ermittelt.

2.2.3.2 Gel-Shift-Assay

Zur Analyse von RNA-Bindung mittels nativer Gelelektrophorese wurden die Bindungsreaktionen auf einem 5 %igen (60:1) Polyacrylamidgel in 0,5 % TBE-Puffer bei 8 °C aufgetrennt. Die Bindungsreaktionen wurden wie in 2.2.3.1 beschrieben durchgeführt. Das Reaktionsvolumen wurde jedoch auf 20 µl reduziert. Die Polyacrylamidgele wurden anschließend auf Whatman-Papier getrocknet und die radioaktive RNA autoradiographisch detektiert.

2.2.3.3 Spaltungsassay

Der prä-mRNA-Spaltungsassay wurde in 20 µl Reaktionsvolumen modifiziert nach *Schönemann* durchgeführt [26]. Die Reaktion bestand aus Spaltungspuffer (2x konzentriert), 50 fmol Substrat-RNA sowie bis zu 5 µl Proteinlösung. Die finale Salzkonzentration von 100 mM setzte sich aus dem in den Proteinlösungen enthaltenem Salz sowie der entsprechenden Menge Salz im Reaktionspuffer zusammen. Als Kontrolle wurde unfraktionierter HeLa-Kernextrakt verwendet. Die genaue Zusammensetzung der Reaktion ist in Tabelle 2 zusammengefasst. Die Reaktion wurde durch Zugabe der RNA gestartet und für eine Stunde bei 30 °C inkubiert.

Reaktionspuffer
20 mM HEPES • KOH, pH 8,0
100 mM Kaliumchlorid
1 mM Magnesiumchlorid
2 mM DTT
20 mM Kreatinphosphat
0,5 mM Cordycepintriphosphat
3,5 % PEG6000

Tabelle 2 Zusammenfassung der Zusammensetzung der Spaltreaktion

Komponente	Stamm-konzentration	Volumen [µl]	End-konzentration
Reaktionspuffer	2x	12,5	1x
RNase Inhibitor	40 U/µl	0,1	0,2 U/µl
DTT	100 mM	0,2	1 mM
(fraktionierter) Extrakt	-	2	-
Proteinlösung oder 200 mM KCL in 50 mM TEA (pH 8.0)	-	3	-
Substrat-RNA	25 nM	2	2,5 nM

Die Reaktion wurde durch Zugabe von EDTA- und SDS-haltigem Puffer gestoppt und die RNA nach einem Proteinase K Verdau gefällt und mittels denaturierendem Polyacrylamidgel und anschließender Autoradiographie analysiert.

Für die Analyse der Aktivität der CFII-Deletionsvarianten wurde Extrakt verwendet, der über eine Ammoniumsulfat-Präzipitation, anschließende Größenausschluss- sowie folgende Anionenaustauschchromatographie erhalten wurde [26], [27]. Für die Aktivitätsanalyse der CFII- sowie CstF-Reinigungen wurde Kernextrakt, der nach einem modifizierten Protokoll über eine PEG-Fällung, anschließende Ammoniumsulfat-Präzipitation sowie Größenausschlusschromatographie hergestellt wurde, verwendet (Daten von Florian Kluge).

2.2.3.4 *In-vitro*-Polyadenylierungsassay

Zur Analyse der Stimulierung der Poly(A)-Polymerase durch mPSF wurde ein Polyadenylierungsassay nach *Schönemann et al.* sowie *Kühn et al.* durchgeführt [26], [28], [157]. Das für das chemische *cross-linking* modifizierte Protokoll ist in 2.2.4.3 beschrieben.

2.2.4 Chemisches *cross-linking*

Für das chemische *cross-linking* von Proteinen wurden sowohl DSBU als auch LC-SDA verwendet. Da beide *cross-linker* NHS-Ester als reaktive Gruppe besitzen, durften die verwendeten Lösungen und Puffer keine freien Amine besitzen. TRIS, welches üblicherweise als Puffersubstanz verwendet wird, wurde durch Triethanolamin ersetzt. Die *cross-linker* wurden in wasserfreiem DMSO (über Molsieb gelagert) gelöst. Für die Reaktion mit DSBU wurde der *cross-linker* direkt in die Proteinlösung gegeben und die Reaktion für 30 Minuten bei Raumtemperatur durchgeführt. Anschließend wurde die Reaktion durch Zugabe von TRIS gestoppt. Da LC-SDA ein heterobifunktionaler *cross-linker* ist, wurde eine zweistufige Reaktion durchgeführt. Die erste, basierend auf der NHS-Gruppe, wurde analog zu DSBU durchgeführt. Anschließend wurde die Diazirin-Gruppe durch 5-sekündiges Bestrahlen mit einer LED-Diode mit einer Wellenlänge von 365 nm (17200 mW/cm²) auf Eis aktiviert. Diese Reaktion musste nicht gestoppt werden.

2.2.4.1 Empirische Bestimmung der optimalen *cross-linker*-Konzentration

Zur empirischen Bestimmung der optimalen *cross-linker*-Konzentration wurde dieser gegen eine konstante Proteinkonzentration titriert. In Abhängigkeit der maximal erreichbaren Proteinkonzentration wurde der *cross-linker* in einer geometrischen Reihe absteigend von einem 1000-fachem molaren Überschuss titriert. Hierzu wurde 1 µl *cross-linker*, welcher unmittelbar vor der Reaktion in DMSO gelöst und auch in DMSO verdünnt wurde, mit 9 µl gereinigtem Protein gemischt. Die Reaktion wurde 30 Minuten bei Raumtemperatur inkubiert und anschließend mit 10 µl TRIS-haltigem SDS-Ladepuffer (2x konzentriert) gestoppt. Anschließend wurden die Reaktionen mittels SDS-Polyacrylamidgel getrennt und analysiert.

2.2.4.2 Präparatives *cross-linking* gereinigter Proteinkomplexe

Für das präparative *cross-linking* wurde die empirisch bestimmte *cross-linker*-Konzentration verwendet. Hierzu wurden in 100 µl Gesamtvolumen 10 µl *cross-linker* in DMSO mit 90 µl Proteinkomplex für 30 Minuten bei Raumtemperatur inkubiert. Die Reaktion wurde mit 25 µl 1 M

TRIS abgestoppt. Für einen In-Gel-Verdau wurde die Reaktion mit 125 µl SDS-Ladepuffer (2x konzentriert) versetzt und 20 µl pro Spur auf ein SDS-Polyacrylamidgel mit einem Gradienten von 4 bis 12 % aufgetragen. Für einen In-Lösung-Verdau wurde das Reaktionsgemisch mit Aceton gefällt. Die anschließende tryptische Spaltung ist in 2.2.5.1 beschrieben.

2.2.4.3 Präparatives *cross-linking* der Polyadenylierungsreaktion

Zum präparativen *cross-linking* der Intermediate der Polyadenylierung wurde eine zweistufige Reaktion durchgeführt. Zuerst wurde die Bildung eines aktiven Polyadenylierungskomplexes unter Reaktionsbedingungen in einem Gesamtvolumen von 150 µl erlaubt, die Reaktion zu den gegebenen Zeitpunkten zur Isolierung der Reaktionsintermediate gestoppt und anschließend die Reaktion in 200 µl chemisch

Reaktionspuffer
50 mM HEPES • KOH, pH 8,0
4 % (w/v) PEG-6000
0,1 mM EDTA
0,02 % (v/v) NP-40
0,8 mg/ml BSA ^{met}
20 % (v/v) Glycerin

quervernetzt. Beide Reaktionen fanden bei 25 °C statt. Eine detaillierte Übersicht der beschriebenen Reaktion ist in Tabelle 3 dargestellt. Die Reaktion wurde ohne ATP 5 Minuten präinkubiert und die Polyadenylierung anschließend durch die Zugabe von ATP gestartet. Die Kontrollreaktion ohne RNA enthielt stattdessen destilliertes Wasser.

Tabelle 3 Detaillierte Zusammensetzung des präparativen *cross-linkings* der Polyadenylierungsreaktion mit einem Gesamtvolumen von 150 µl.

Komponente	Stammlösung	Volumina [µl]	Endkonzentration
Reaktionspuffer	2x	75	1x
tRNA	1 g/l	7,5	50 µg/ml
RNase Inhibitor	40 U/µl	0,75	0,2 U/µl
DTT	100 mM	1,5	1 mM
MgCl	200 mM	1,5	2 mM
KCl	1 M	7,5	50 mM
Substrat-RNA	1 µM	3	20 nM
Poly(A)-Polymerase	1,6 µM	1,875	20 nM
mPSF	1,6 µM	1,875	20 nM
PABPN1	26,5 µM	1,7	300 nM
Wasser		32,8	
ATP	5 mM	15	0,5 mM
<i>Reaktionsstopp nach 0, 3 oder 10 Minuten:</i>			
EDTA	50 mM	15	4,3 mM
DSBU	22 mM	7,5	0,95 mM
<i>Reaktionsstopp nach 30 Minuten:</i>			
TRIS	1 M	75	ca. 300 mM

Die Reaktion wurde anschließend mit 1,25 ml kaltem Aceton gefällt, in 40 µl SDS-Ladepuffer aufgenommen und je 20 µl pro Spur auf ein SDS-Polyacrylamidgel geladen. Die anschließende tryptische Spaltung ist in 2.2.5.1 beschrieben.

2.2.5 Massenspektrometrische Analysen

2.2.5.1 Probenvorbereitung

2.2.5.1.1 Tryptische Spaltung im Gel

Zur Analyse von spezifischen Proteinen und Reaktionsprodukten in einer heterogenen Lösung wurden diese mittels SDS-Polyacrylamidgel separiert und mit Coomassie gefärbt. Die zu untersuchenden Proteinbanden wurden mit einem scharfen Skalpell ausgeschnitten und in ca. 1 mm³ kleine Würfel zerkleinert. Anschließend wurden die Proteinbanden wie folgt prozessiert. Bei sehr breiten Proteinbanden wurden die angegebenen Volumina verdoppelt, sodass alle Gelstücke immer mit Flüssigkeit bedeckt waren.

<u>Reduzierung und Alkylierung (Überstände verwerfen)</u>	
10 Minuten 37 °C	500 µl Acetonitril
30 Minuten 56 °C	50 µl 10 mM DTT in 100 mM Ammoniumbicarbonat
10 Minuten 37 °C	500 µl Acetonitril
30 Minuten 37 °C	50 µl 55 mM Chloracetamid in 100 mM Ammoniumbicarbonat
10 Minuten 37 °C	500 µl Acetonitril
<u>Tryptischer Verdau (Überstände nicht verwerfen)</u>	
2 Stunden 4 °C	25 µl 12,5 ng/µl Trypsin
über Nacht 37 °C	+ 10 µl 100 mM Ammoniumbicarbonat
<u>Extraktion (Überstände sammeln)</u>	
2x 15 Minuten 37 °C	1:2 (v/v) 5 % Trifluoressigsäure / Acetonitril

Zur Analyse wurden die Proben anschließend auf ca. 50 µl mittels Rotations-Vakuum-Konzentrator eingengt.

2.2.5.1.2 Tryptische Spaltung in Lösung

Zur Verdauung von Proteinen in Lösung wurde die entsprechende Menge mit Aceton präzipitiert. Hierzu wurden zu der Proteinlösung 4 Volumina eiskaltes Aceton gegeben, es wurde gut gemischt und 1 Stunde bei -20 °C inkubiert. Anschließend wurde das ausgefällte Protein durch Zentrifugation pelletiert (10 Minuten, 15000xg). Die Proteolyse in Lösung wurde wie folgt durchgeführt:

<u>Resuspendierung</u>	
25 µl 8 M Harnstoff in 400 mM Ammoniumbicarbonat	
<u>Reduzierung und Alkylierung</u>	
20 Minuten 50 °C	45 mM DTT in dest. Wasser
20 Minuten 37 °C	100 mM Chloracetamid in dest. Wasser
<u>Tryptischer Verdau</u>	
Reaktion auf 200 µl mit destilliertem Wasser auffüllen	
über Nacht 37 °C	1:50 bis 1:100 (w/w) Trypsin/Protein
<u>Reaktionsstopp</u>	
10 µl 10 % Trifluoressigsäure	

Zur Analyse wurden die Proben anschließend auf ca. 50 µl mittels Rotations-Vakuum-Konzentrator eingengt.

2.2.5.2 Massenspektrometrische Analyse

2.2.5.2.1 Geräte- und analysespezifische Parameter

Für die HPLC-Systeme wurden für die Ladepumpe 0,1 % (v/v) TFA in Wasser und für die Nanopumpe 0,1 % (v/v) Ameisensäure in Wasser (Puffer A) bzw. 0,08 % (v/v) Ameisensäure in Acetonitril (Puffer B) verwendet.

Für die Analyse der chemischen *cross-links* des CFII-Komplexes wurde eine C8-Vorsäule sowie die PepMap™ C18-Hauptsäule am Orbitrap Fusion™-Massenspektrometer verwendet. Für alle weiteren massenspektrometrische Analysen wurde eine C18-Vorsäule verwendet. Für das Orbitrap Fusion™-Massenspektrometer wurde die selbstgepackte PicoFrit™ Säule mit einer Länge von 250 mm und für das Q-Exactive™-Massenspektrometer mit einer Länge von 500 mm verwendet.

Die genauen gerätespezifischen Einstellungen der massenspektrometrischen Analysen, einschließlich der Flussgeschwindigkeiten, MS- und MS/MS-Methoden sind im Anhang in 7.2.2 dokumentiert.

2.2.5.2.1 Proteinidentifikation

Zur Verifizierung der Proteinidentität wurden ca. 2 µg Protein in Lösung verdaut. Die Peptide wurden an einem Orbitrap Fusion Tribrid-Massenspektrometer analysiert. Die Auswertung der Rohdaten erfolgte mittels MaxQuant, wobei die Proteinsequenzen der vermuteten Proteine und die MaxQuant-StandardEinstellung verwendet wurden.

2.2.5.2.2 Analyse chemischer *cross-links*

Die massenspektrometrischen Rohdaten zur *cross-link*-Analyse wurden in mgf-Dateien umgewandelt. Für die Identifikation von *cross-links* wurde die MeroX-Software verwendet [145]. Hierzu wurden neben den umgewandelten Rohdaten die Sequenzen der Zielproteine im fasta-Format benötigt. Die Analyse der Daten erfolgte im quadratischen Modus und mit einer FDR-Obergrenze von 5 %. Dieses reizt zwar nicht den Vorteil von spaltbaren *cross-linkern* aus, jedoch war auch eine Identifikation von *cross-link*-Peptiden möglich, die nicht beide Dublett-Paare zeigten.

Für den *cross-linker* DSBU wurde der „Quick Set-Up“ verwendet, wobei abweichend von den Standardangaben HCD als Fragmentierung gewählt wurde.

Da für den *cross-linker* LC-SDA keine Vorgabe in MeroX bestand, wurde der „Quick Set-Up“ für den SDA *cross-linker* verwendet, mit der Abweichung der Verwendung der HCD Fragmentierung. Folgende Änderungen wurden in den *cross-link*-Einstellungen durchgeführt:

Zusammensetzung: C11H17NO2

Modifizierungen von Fragmenten am NHS-reaktiven Peptid			
Name	Zusammensetzung	Masse	Essentiell
Pep§+195u	C ₁₁ H ₁₇ NO ₂	195,123592871	Ja
Pep§+213u	C ₁₁ H ₁₉ NO ₃	213,13649337	Ja
Pep§		0,0	Ja
Pep§+113u	C ₆ H ₁₁ NO	113,08406393	Nein
Modifizierungen von Fragmenten am Diazirin-reaktivem Peptid			
Name	Zusammensetzung	Masse	Essentiell
Pep#		0	Ja

Darüber hinaus wurden keine Änderungen an den Standardeinstellungen durchgeführt. Die identifizierten *cross-links* wurden anschließend zusätzlich manuell beurteilt und ausgewählt.

2.2.6 Computergestützte Proteinanalyse und Strukturvorhersage

2.2.6.1 Bestimmung der Länge der Aminosäureseitenketten

Die Berechnung der Länge der Aminosäureseitenketten erfolgte mit dem in dieser Arbeit geschriebenen und in Quellcode A.1 dargestellten Shellskript. Aminosäurename sowie der Typ der reaktiven Aminosäure wurden entsprechend angepasst. Die bestimmten Längen wurden in eine entsprechende Ergebnisdatei gespeichert, mittels R analysiert und als Boxplot dargestellt.

2.2.6.2 Berechnung der Peptid-Entfernungsbeschränkung

Zur datenbankweiten Analyse von Peptidlängen wurde ein in dieser Arbeit entwickeltes Shellskript genutzt. Dieses ist in Quellcode A.2 dargestellt. Dieses berechnet die Entfernung zwischen zwei C α -Atomen in einem Peptid anhand ihrer Koordinaten im kartesischen Koordinatensystem. Hierbei wurde für jede Strukturdatei eine separate Ergebnisdatei erzeugt, die anschließend auf ihre Konsistenz geprüft wurde. Hierbei wurde ein erlaubter Rahmen für C α -Entfernungen in Dipeptiden von 2 bis 6 Å gewählt. Strukturen, die mindestens eine Dipeptidbindungsgröße außerhalb dieses Rahmens besaßen, wurden von der weiteren Analyse ausgeschlossen.

2.2.6.3 Proteinstrukturvorhersage

2.2.6.3.1 Manuelle Proteinstrukturvorhersage

Template-Identifikation

Zur Identifikation von Strukturhomologen wurden die Online-Services HHPRED (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) und BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) verwendet [139], [158]. Für beide Programme wurde die PDB-Datenbank ausgewählt. Zur Identifikation von homologen Strukturen wurde die Aminosäuresequenz des Zielproteins oder der Zielregion genutzt. HHPRED nutzt zur

Identifikation von Homologen den paarweisen Vergleich von verdeckten Markowmodellen (HMM), während BLASTp den Blast-Algorithmus zur Identifikation nutzt.

Proteinstrukturberechnung

Zur manuellen Proteinstrukturvorhersage wurde das Programm MODELLER genutzt [146]. Dieses basiert auf einer vergleichenden Strukturvorhersage. Auf Grundlage der Vorlagestruktur berechnet das Programm automatisch Beschränkungen wie Entfernungen zwischen zwei Atomen sowie die Torsionswinkel der Peptidbindungen. Zur Optimierung des Strukturmodells kann anschließend zudem eine Molekulardynamik-Simulation (MD) durchgeführt werden, die die chemischen und elektrostatischen Eigenschaften der Aminosäureseitenketten mitberücksichtigt. Die komplette Dokumentation des Programmes und seiner Eigenschaften ist auf der Website des Programmes verfügbar (<https://salilab.org/modeller/documentation.html>).

Dieses Programm ist kostenfrei für den akademischen Gebrauch, bedarf aber einer vorherigen Registrierung (<https://salilab.org/modeller/registration.html>). Zur Strukturvorhersage benötigt werden das *Alignment* des Zielproteins mit den zuvor identifizierten Strukturhomologen im PIR-Format, die Strukturdateien dieser Strukturhomologen im pdb-Format sowie ein Pythonskript, welches die Strukturvorhersage durchführt. Dieses wurde von Jun.-Prof. Panagiotis L. Kastiris zu Verfügung gestellt und modifiziert. Enthielt das Zielprotein im Vergleich zu den Vorlagestrukturen größere Bereiche, die nicht abgedeckt wurden und somit *de novo* modelliert werden mussten, wurde eine zusätzliche Schleifenoptimierung (*loops refinement*) durchgeführt. Der Quellcode für die verwendeten Pythonskripte ohne (Quellcode A.3) und mit Schleifenoptimierung (Quellcode A.4) sind im Anhang zu finden. Für jedes Zielprotein wurden 20 Strukturmodelle generiert, welche über einen internen Bewertungsalgorithmus bewertet wurden.

2.2.6.3.2 Automatisierte Proteinstrukturvorhersage

Zur automatisierten Strukturvorhersage wurde der I-TASSER-Webserver verwendet (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) [140]. Die Nutzung ist kostenfrei, bedarf aber einer vorherigen Registrierung. Zur Proteinstrukturvorhersage wird ausschließlich die Sequenz des Zielproteins benötigt, welche auf dem Server eingereicht wird. Dieser identifiziert automatisch Strukturhomologe und generiert fünf Strukturmodelle des Zielproteins, welche wie auch in MODELLER über einen internen Bewertungsalgorithmus bewertet werden.

2.2.6.4 Docking

2.2.6.4.1 Molekulares Docking

Für die Generierung von oligomeren Proteinkomplexen aus monomeren Strukturen, das *Docking*, wurde der HADDOCK-Webserver (<https://milou.science.uu.nl/services/HADDOCK2.2/haddock.php>) verwendet [138]. HADDOCK, ein Akronym für *High Ambiguity Driven biomolecular DOCKing*, ist ein Programm zur Bildung von Biomolekül-Komplexen basierend auf der 3D-

Struktur der Einzelkomponenten sowie zusätzlichen Interaktionsinformationen. Diese können aus biochemischen Experimenten wie Mutagenesestudien, massenspektrometrischen Methoden wie Wasserstoff/Deuterium-Austausch oder *cross-linking* aber auch NMR-Experimenten stammen. Das Docking von HADDOCK basiert auf drei aufeinanderfolgenden Schritten:

1. *Rigid-body energy minimization*

Hierbei werden die Substrukturen in einer Mindestnähe von 25 Å, unter Berücksichtigung der Interaktionsinformationen, platziert und abwechselnd um ihren Masseschwerpunkt rotiert. Jeder so gebildete Komplex wird anschließend durch Berechnungen wie Interaktionsfläche, Abschätzung der elektrostatischen sowie van-der-Waals-Energien sowie der Verletzung der Interaktionsinformationen bewertet.

2. *Semi-flexible refinement in torsion angle space*

Die besten Komplexstrukturen des ersten Schrittes werden in diesem Schritt weiter optimiert. Dieser besteht aus drei Schritten, die auf dem heuristischen Approximationsverfahren *simulated annealing* (deutsch: simuliertes Ausglühen) basiert. In der ersten Optimierung (1000 Schritte von 2000 bis 50 K in je 8 fs) werden beide Proteine als Festkörper angenommen und ihre Ausrichtung zueinander verbessert. In der zweiten Optimierung (4000 Schritte von 200 bis 50 K in je 4 fs) dürfen sich die Seitenketten in der Interaktionsfläche frei bewegen. Im letzten Optimierungsschritt (1000 Schritte von 500 bis 50 K in 2 fs) dürfen sich Peptidrückgrat und Aminosäureseitenketten in der Interaktionsregion frei bewegen.

3. *Final refinement in explicit solvent refinement*

In diesem Schritt wird die Komplexstruktur in einer expliziten Hydrathülle verfeinert. Hierbei wird die Konformation der Komplexstruktur kaum noch verändert, jedoch entsprechen die hier optimierten Interaktionsregionen aufgrund der Interaktion mit dem Lösungsmittel mehr dem nativen Zustand.

Die Nutzung ist kostenfrei, jedoch wird für das *Docking* von multimeren Proteinkomplexen („*the multi-body interface*“) der sogenannte „Guru level access“ benötigt, welcher vorher unter Angabe der Institution und des Forschungsschwerpunktes beantragt werden muss. Für das molekulare *Docking* werden die Strukturen der Proteine als pdb-Datei benötigt, wobei jeder Struktur beim Einreichen eine Segment-ID (*segid*) manuell zugeordnet werden muss. Mehrkettige Modelle müssen entsprechend den Servervorgaben modifiziert werden, sodass es in jedem Modell zu keiner Überschneidung der Aminosäurenummern kommt. Darüber hinaus müssen nicht aminogene Atomeinträge wie Metallionen entfernt oder richtig umbenannt werden. Die genaue Funktionsweise sowie das Format der Strukturdaten ist in *de Vries et al.* beschrieben [137]. Die Entfernungsbeschränkungen wurden in Abhängigkeit des zu bildenden Komplexen als ein- oder mehrdeutig definiert. Eindeutige Beschränkungen werden beim *Docking* forciert und Verletzungen dieser Beschränkungen mit einer schlechten Gesamtbewertung des gebildeten

Komplexes bestraft. Mehrdeutige Beschränkung hingegen dürfen in dem gebildeten Modell verletzt werden, ohne dass dies einen Einfluss auf die Bewertung hat. Ein Beispiel hierfür ist Hämoglobin, ein Heterotetramer aus je zwei Kopien Hb α (α_1 , α_2) und Hb β (β_1 , β_2). Eine Interaktion zwischen einem α -Molekül und einem β -Molekül kann entweder der Interaktion α_1 - β_1 oder α_1 - β_2 entsprechen. Dementsprechend ist diese Interaktion als mehrdeutig zu definieren.

Die Beschränkungen wurden im folgenden Format definiert:

```
assign (segid X1 and resid X2 and name CA) (segid X3 and resid X4 and name CA) dist1 dist2 dist3
```

wobei X_1 die Segment-ID und X_2 die Aminosäurenummer der ersten Struktur, X_3 die Segment-ID und X_4 die Aminosäurenummer der zweiten Struktur beschreibt. Die Entfernung wird durch die mittlere Entfernung $dist_1$, sowie die maximale untere ($dist_2$) sowie obere Abweichung ($dist_3$) angegeben. Die Entfernung wird in Å angegeben. Die genaue Definition der Entfernungsbeschränkung ist in *Bonvin et al.* zu finden [159].

Für das molekulare *Docking* wurden im Vergleich zu den Standardbedingungen 5000 statt 1000 Strukturen für das *rigid body docking* sowie 400 statt 200 Strukturen für das *semi-flexibel refinement* generiert. Darüber hinaus wurden keine Änderungen angewendet.

Die generierten 200 Strukturen werden in sogenannte Cluster eingeordnet. Hierbei werden neben der Konformation energetische Parameter wie die Interaktionsfläche, elektrostatische und van-der-Waals-Energien sowie das Maß der Beschränkungsverletzung kombiniert. Je größer die Anzahl der Proteine pro Cluster, desto zuverlässiger ist die Struktur des gebildeten Proteinkomplexes.

2.2.6.4.2 Analyse des Konformationsraumes

Ein Konformationsraum beschreibt alle möglichen räumlichen Positionen, die ein Protein, basierend auf bekannten Interaktionsflächen oder -bereichen, im Verhältnis zu einem anderen Protein oder Proteinkomplex einnehmen kann. Zur Berechnung des Konformationsraumes eines Proteins (*mobile*) im Vergleich zu einem Protein oder Proteinkomplex (*target*) wurde der DisVis-Webserver (<https://milou.science.uu.nl/cgi/services/DISVIS/disvis/>) verwendet [136]. Eingereicht werden muss neben den beiden Proteinstrukturen im pdb-Format die Entfernungsbeschränkung als Textdatei. Eine Beschränkung in der Anzahl der Ketten wie in HADDOCK besteht nicht. Die Entfernungen werden wie folgt definiert:

```
Kette1 Nummer1 Atomname1 Kette2 Nummer2 Atomname2 Min Max
```

wobei die erste Atomdefinition (Kette, Nummer, Atomname) die Aminosäure im *target*-Protein und die zweite Definition die Aminosäure im *mobile*-Protein beschreibt. Die Entfernung wird in Å angegeben. Absolut notwendig ist, dass beide Atome in den jeweiligen Strukturen aufgelöst sind.

Das Ergebnis ist eine Beschreibung des Konformationsraumes in Form einer mrc-Datei sowie der statistischen Analyse der möglichen Lösungsstrukturen. Sofern die Entfernungsbeschränkungen mehrdeutig sind, wird die Anzahl der „ungültigen“ Beschränkungen, also die Anzahl der sich widersprechenden Beschränkungen, angegeben.

2.2.6.5 Entfernungsbestimmung in Strukturmodellen

Zur Bestimmung von Entfernungen in den gebildeten Strukturmodellen wurde die Distanzbestimmungsfunktion von PyMOL verwendet [160]. Hierzu wurde der „get_distance“ Befehl mit folgender Syntax verwendet:

```
get_distance /id1//Kette1/Nummer1//CA, /id2//Kette2/Nummer2//CA
```

wobei mit id, der spezifischen Kette und Aminosäurenummer die beiden Ca-Atome der jeweiligen Aminosäuren beschrieben wird. Die Entfernung wird in Å ausgegeben.

2.2.6.6 Strukturkartierung in Kryo-Elektronenmikroskopie-Dichten

Zur Validierung von Strukturmodellen mit publizierten Dichtekarten von kryo-Elektronenmikroskopie-Experimenten wurde das Programm Chimera verwendet [135]. Hierzu wurden das Strukturmodell und die Dichtekarte geladen und im ersten Schritt manuell überlagert. Anschließend erfolgte das automatische Einbetten des Strukturmodelles in die Dichte über die „Fit in Map“-Funktion. Hierbei wurde die zusätzlich die Option „Use map simulated from atoms, resolution:“ mit der publizierten Auflösung der Dichtekarte verwendet.

2.2.7 Interaktomanalyse

2.2.7.1 Pulldown

Zur Identifikation von Interaktionspartnern der rekombinant hergestellten Proteinkomplexe wurden diese mit kommerziellem Kernextrakt gemischt und anschließend affinitätschromatographisch gereinigt. Hierzu wurden 160 µl Kernextrakt sowie das Köderprotein in 400 µl (für die CFII-Varianten, den CstF- sowie den Endonuklease-Komplex) bzw. 640 µl (für den CFI-Komplex) unter nativen Pufferbedingungen (Reaktionspuffer) 20 Minuten bei 25 °C inkubiert. Die

Reaktionspuffer

50 mM HEPES • KOH, pH 7,5
 100 mM Kaliumchlorid
 10 % Glycerin
 1 mM Magnesiumacetat
 0,01 % NP-40
 0,5 mM ATP
 0,5 mM DTT

verwendeten Proteinmengen sind in Tabelle 4 dokumentiert. Zur Reinigung der Köderproteine und deren gebundenen Interaktionspartner wurde das entsprechende Matrixmaterial (siehe Tabelle 4) hinzugegeben und die Reaktion im Kühlraum unter Rotation für 1 Stunde inkubiert.

Nach der Bindung der Köderproteine wurde das Säulenmaterial 2-mal mit 400 µl Reaktionspuffer gewaschen und anschließend 2-mal mit 200 µl eluiert. Der Elutionspuffer für die FLAG-Agarose

0,3 mg/ml FLAG-Peptid. Die Elution der Proteine von der Strep-Tactin-Sepharose wurde mit denaturierendem Puffer durchgeführt (100 mM Ammoniumbicarbonat, 8 M Harnstoff).

Tabelle 4 Übersicht der verwendeten Materialien zur Affinitätsreinigung der Köderproteine.

Köderprotein	Proteinmenge	Matrix-material	Matrix-volumen	Reaktionsvolumen
CFI Δ N769, CFI Δ N1123	34 pmol	ANTI-FLAG M2 Agarose	40 μ l	400 μ l
CstF	25,5 pmol			
Endonuklease-Komplex	34 pmol	Strep-Tactin Sepharose	96 μ l	640 μ l
CFI	0,7 nmol			

2.2.7.2 Identifikation von Interaktionspartnern

Zur Proteinidentifikation wurden 150 μ l der ersten Elutionsfraktion mit Aceton gefällt und anschließend wie in 2.2.5 beschrieben in Lösung verdaut und massenspektrometrisch analysiert.

2.2.7.2.1 MaxQuant

Zur Identifikation von Proteinen wurden die massenspektrometrischen Rohdaten mit MaxQuant (Version 1.6.0.1) analysiert [143]. Hierzu wurde die SwissProt-Datenbank vom 8. Mai 2019 verwendet. Diese enthielt 20427 Proteineinträge. Die MaxQuant-Standardinstellungen wurden für die markierungsfreie Quantifizierung (*label free quantification*) optimiert. Dieses beinhaltet die eindeutige Klassifizierung der verschiedenen Rohdaten in unabhängige Experimente, des Aktivieren der Quantifizierung (LFQ), die Korrelation zwischen den verschiedenen Experimenten zur Verbesserung der Proteinidentifikation (*match between runs*), sowie des Berechnen der iBAQ-Werte. Zur Bewertung der Abundanz der Proteine wurden die iBAQ-Werte logarithmisch zur Basis 2 dargestellt. Zur Berechnung der Anreicherungsfaktoren wurden die ursprünglichen iBAQ-Werte genutzt. Hierbei wurde der Quotient des Experimentes zur Kontrolle berechnet und auf die nächste ganze Zahl gerundet.

2.2.7.2.2 GO term-Anreicherungsanalyse

Zur Klassifizierung der identifizierten Proteine anhand ihrer hinterlegten *GO terms* wurde die PantherDB genutzt [161]. Hierbei wurden alle Proteine ausgewählt, die der 3'-Prozessierung von mRNA zugeordnet wurden. Da die *GO terms* neben manuellen Annotationen auch Funktionen aufgrund von Sequenzähnlichkeiten zu anderen, besser beschriebenen Proteinen aufweisen, kann diese Einordnung zu falsch positiven Ergebnissen führen. Darüber hinaus haben Proteine mit unbekannter Funktion keinen Eintrag, sodass dieses zu falsch negativen Ergebnissen führen kann.

3 Ergebnisse

3.1 Strukturelle Analyse der bekannten Polyadenylierungskomplexe

Zur biochemischen und strukturellen Analyse der in der Einleitung beschriebenen Proteinkomplexe wurden diese in Baculovirus-infizierten Insektenzellen produziert [162]. Die Reinigungsstrategien für CFI, CFII und CstF wurden in vorherigen Arbeiten bereits etabliert [27], [163]. Für die CPSF-Untereinheiten wurde die Reinigung etabliert, welche in 2.1.2 näher beschrieben wird. Die Identität aller gereinigten Proteine wurde massenspektrometrisch verifiziert. Zur Qualitätskontrolle wurden alle Proteinkomplexe auf ihre RNA-Bindung und, sofern möglich, auf ihre Aktivität in der Spaltungs- bzw. Polyadenylierungsreaktion hin untersucht. Für das chemische cross-linking wurde hauptsächlich DSBU, für den CFI-Komplex zusätzlich LC-SDA, verwendet. Alle massenspektrometrischen Analysen wurden mindestens in technischen Duplikaten durchgeführt. Identifizierte *cross-links* wurden manuell validiert und besaßen eine Falscherkennungsrate (*false discovery rate*) von unter 1%.

3.1.1 Strukturanalyse des Spaltungsfaktor II (CFII)

3.1.1.1 Darstellung und Aktivitätsanalyse

CFII wurde affinitätschromatographisch über den Hexahistidin-*tag* an Clp1 und eine anschließende Anionenaustauschchromatographie (Abbildung 3.1 A). Zur Proteinreinigung wurden zwei unabhängige Expressionskulturen verwendet, wobei in den Hauptfraktionen 0,5 bzw. 0,8 mg Komplex enthalten war. Dieses entspricht einer Ausbeute von 1,4 bzw. 2 mg Komplex pro Liter Kulturvolumen. Der gereinigte Proteinkomplex wurde verschiedenen Aktivitätstest unterzogen. So konnte die Spaltung der RNA in fraktioniertem Extrakt rekonstituiert werden (Abbildung 3.1 B). Hierbei ist eine konzentrationsabhängige Spaltung der Substrat-RNA zu beobachten. Keine Spaltung konnte in der Deltavariante, in der das Polyadenylierungssignal von AAUAAA zu AAGAAA mutiert vorliegt und nicht mehr von CPSF erkannt wird, detektiert werden. Darüber hinaus zeigte dieser reproduzierbar eine hochaffine Binding des Polyadenylierungssubstrates L3 (Abbildung 3.1 C). Der gemessene K_{50} -Wert von 0,6 nM entsprach dem Erwartungswert (persönliche Kommunikation mit Peter Schäfer) [27].

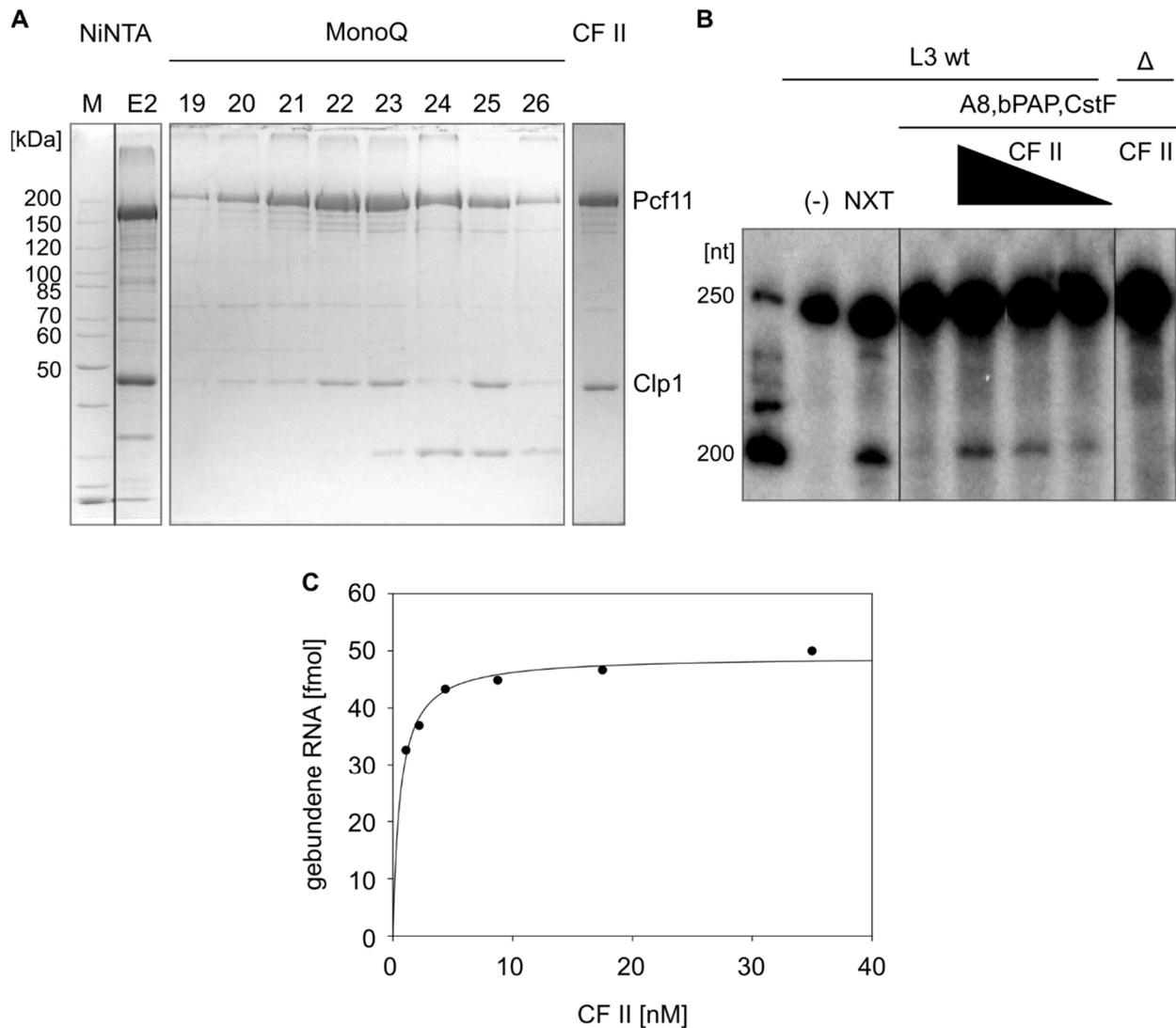


Abbildung 3.1 Reinigung und Aktivitätsanalyse des CFII-Komplexes. A) Reinigung des CFII-Komplexes über die initiale Affinitätschromatographie (NiNTA) und anschließende Anionenaustauschchromatographie (MonoQ). Das für die Analysen quantifizierte Proteine ist gesondert dargestellt (CFII) B) Spaltreaktion von L3-RNA. Die Substrat-RNA wurde ohne Proteine (-), mit kommerziellem Kernextrakt (NXT), sowie in fraktioniertem Extrakt (A8) mit der Komplementation von rekombinanten Proteinen inkubiert. Die Bildung des Produktes konnte in Abhängigkeit der CFII-Konzentration beobachtet werden, nicht jedoch bei Verwendung der mutierten RNA (L3 Δ). C) RNA-Bindung von CFII. Die Substrat-RNA wurde mit steigenden Konzentrationen CFII inkubiert und die Bindung mittels Bindung an Nitrozellulosefilter und anschließender Szintillationsmessung analysiert. Die Messwerte wurden hyperbolisch angepasst.

3.1.1.2 Chemisches *cross-linking* von CFII

Zur Bestimmung der optimalen DSBU-Konzentration wurde der *cross-linker* gegen den gereinigten Proteinkomplex titriert. Bei konstanten 1,1 μM CFII-Komplex wurde der *cross-linker* in einem Bereich von 0,07 bis 2,8 mM eingesetzt, was einem 60- bis 2500-fachem molaren Überschuss entsprach.

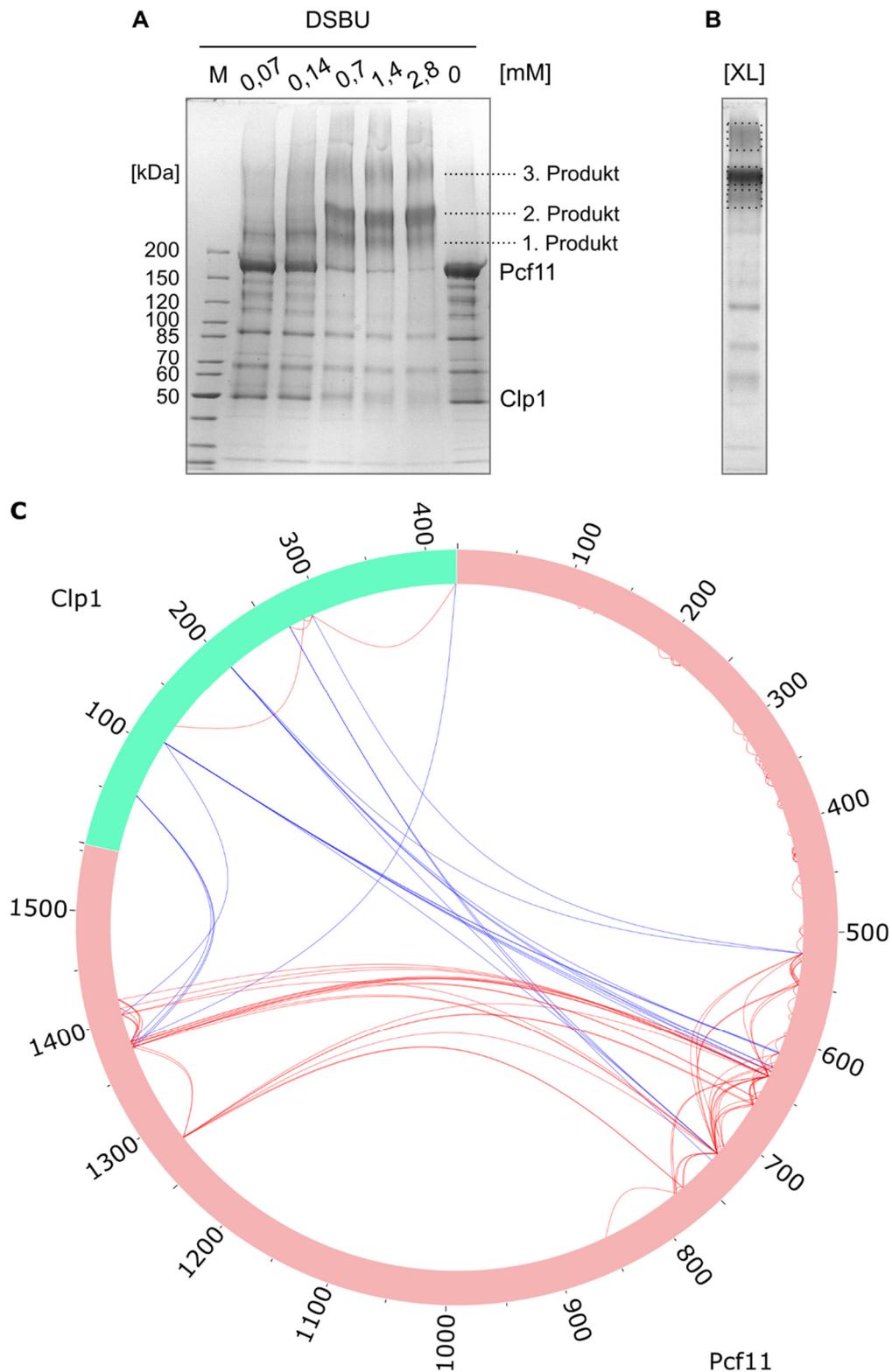


Abbildung 3.2 Chemisches *cross-linking* des CFII-Komplexes. A) Empirische Bestimmung des optimalen Verhältnisses zwischen CFII-Komplex und DSBU. Die Reaktionsprodukte wurden mittels SDS-Polyacrylamidgel aufgetrennt und mit Coomassie gefärbt. Konstante Konzentrationen von CFII (1,1 μM) wurden gegen steigende Konzentrationen des *cross-linkers* eingesetzt. Zur Kontrolle wurde die Trägersubstanz DMSO genutzt („0 mM“). Die monomeren Untereinheiten sowie die gebildeten Reaktionsprodukte sind gekennzeichnet. B) Ausschnitt des präparativen SDS-Polyacrylamidgels. Die drei separat analysierten Produkte sind markiert. C) Visualisierung

der identifizierten *cross-links* in einem Kreisdiagramm. Die Proteine sind im Uhrzeigersinn von N- zu C-Terminus dargestellt. Intra- (rot) und intermolekulare (blau) *cross-links* sind farblich dargestellt.

Es zeigte sich, dass sich konzentrationsabhängig drei hochmolekulare Produkte bilden (Abbildung 3.2 A). Da bei dem höchsten Überschuss des *cross-linkers* noch monomeres Protein zu identifizieren war, wurde die Konzentration von 2,8 mM zur präparativen Analyse genutzt. Pro Reaktion wurden 25 pmol CFII eingesetzt und die drei Reaktionsprodukte mittels SDS-Polyacrylamidgel separiert und getrennt voneinander massenspektrometrisch analysiert (Abbildung 3.2 B). Die Messung fand in technischen Replikaten statt. Eine quantitative Analyse der drei Reaktionsprodukte zeigte keinen Unterschied in ihrer Stöchiometrie (Daten nicht gezeigt). Auch konnte kein Unterschied in den identifizierten *cross-links* beobachtet werden, sodass für die weitere Analyse die Daten zusammengefügt wurden. Insgesamt konnten so nach manueller Validierung 152 einzigartige *cross-link*-Paare identifiziert werden, von denen 18 inter- und 134 intramolekular waren (Abbildung 3.2 C; Tabelle A 5).

3.1.1.3 Identifikation der Interaktionsregion zwischen Clp1 und Pcf11

Die Interaktion zwischen Clp1 und Pcf11 ist in Hefe sehr gut beschrieben. Die Interaktionsregion in Pcf11 (Aminosäuren 475 bis 499) wurde mit Clp1 kokristallisiert (PDBID: 2NPI) [61]. Dieser Bereich ist in humanem Pcf11 (Aminosäuren 1378 bis 1401) konserviert (Abbildung A.3). Neben diesem Bereich konnte durch die *cross-link*-Experimente eine weitere Domäne im Bereich der Aminosäuren 500 bis 750 identifiziert werden, welche sich in räumlicher Nähe zu Clp1 befindet (siehe Abbildung 3.2 C). Die Anzahl und Anordnung der intramolekularen *cross-links* deutet auf eine klar definierte Tertiärstruktur hin und eine räumliche Nähe zum beschriebenen Clp1-Interaktionsbereich ist zu sehen.

Zur näheren Charakterisierung der Interaktion zwischen Clp1 und Pcf11 wurde eine limitierte tryptische Proteolyse durchgeführt. Hierzu wurden je 0,5 µg CFII mit 0,5 ng Trypsin inkubiert, was einem 116-fachem molarem Unterschuss entspricht. Die Reaktion wurde nach definierten Zeitpunkten durch AEBSF gestoppt und mittels SDS-PAGE analysiert. Es zeigte sich, dass Clp1 in dem untersuchten Zeitraum stabil bleibt, während Pcf11 stark abgebaut wird (Abbildung 3.3 A). Darüber hinaus konnte die Bildung eines stabilen Fragmentes von Pcf11 mit einem apparentem Molekulargewicht von ca. 50 kDa beobachtet werden, welches als Pcf11T bezeichnet wurde (Abbildung 3.3 A ★). Eine anschließende affinitätschromatographische Reinigung über den Hexahistidin-*tag* von Clp1 zeigte, dass dieses Fragment weiterhin stabil mit Clp1 assoziiert ist (Abbildung 3.3 B). Eine massenspektrometrische Analyse sowie ein analytischer Edman-Abbau identifizierten das Fragment als Pcf11ΔN1182. Zur weiteren Charakterisierung der Interaktion wurde Pcf11 bis zur ersten Zinkfingerdomäne verkürzt (Pcf11ΔN1339). Dieses wurde mit Clp1 in Insektenzellen analog zum nativen Komplex exprimiert

und gereinigt. Es zeigte sich, dass dieses kürzeste Fragment weiterhin in einem stabilen Komplex mit Clp1 vorliegt (Abbildung 3.3 C). Die Interaktion von Pcf11 zu Clp1 wird daher vor allem durch das beschriebene Peptid im C-terminalen Bereich von Pcf11 vermittelt.

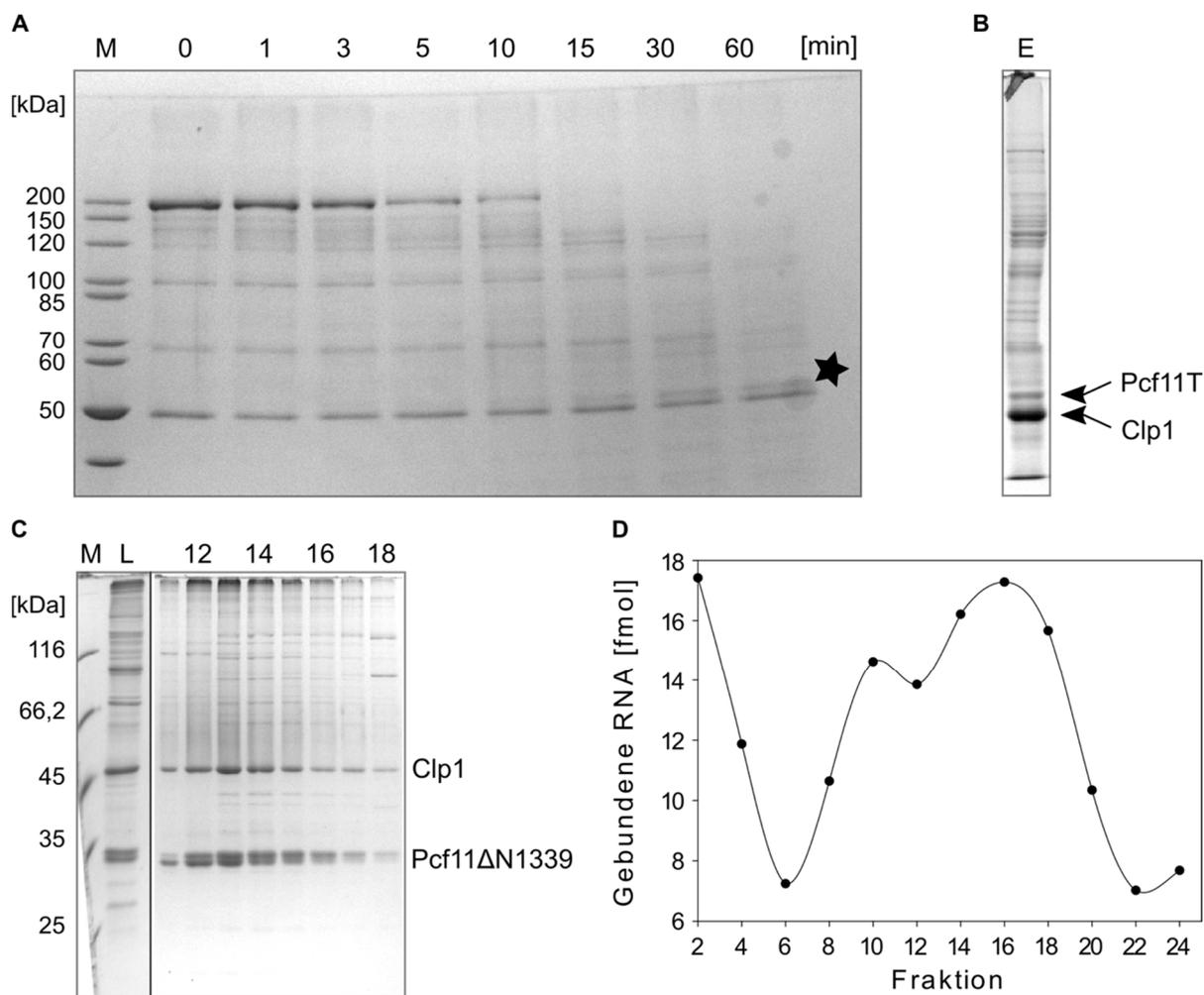


Abbildung 3.3 Tryptischer Verdau und Charakterisierung verkürzter CFII Komplexe. A) Darstellung der tryptischen Proteolyse von CFII über die Zeit. Das stabile Fragment von Pcf11 ist gekennzeichnet (★). B) Elutionsfraktion der affinitätschromatographischen Reinigung nach Inkubation mit Trypsin. Clp1 sowie das tryptische Pcf11-Fragment (Pcf11T) sind gekennzeichnet. C) Reinigung des CFIIΔN1139 Komplexes über eine Anionenaustauschchromatographie. Dargestellt sind das eingesetzte Protein (L) sowie die Peakfraktionen der Reinigung. Gekennzeichnet sind Clp1 sowie Pcf11ΔN1139. D) Verfolgung der RNA-Bindung über die Säulenfraktionen. Dargestellt ist die gebundene RNA in fmol der jeweiligen Fraktionen. Eingesetzt wurden 25 fmol Substrat-RNA und je 1 µl der Proteinfractionen.

3.1.1.4 Untersuchung der RNA Bindung von Pcf11

Natives CFII bindet die untersuchten Polyadenylierungssubstrate mit sub-nanomolarer Affinität (siehe 2.1.1.1). Da für monomeres Clp1 keine RNA-Bindung identifiziert werden konnte (persönliche Kommunikation mit Peter Schäfer), wurden verschiedene verkürzte Pcf11-Varianten im Komplex mit Clp1 auf ihre RNA-Bindung hin untersucht. Neben dem in 2.1.1.3 beschriebenen Fragment wurden analog zwei weitere N-terminal verkürzte Varianten generiert, welche bei

Aminosäure 770 (CFIIΔN769) sowie 1124 (CFIIΔN1123) begannen. Die affinitätschromatographische Reinigung mit anschließender Anionenaustauschchromatographie erfolgte analog zu CFII. Die gereinigten Komplexe sind in Abbildung 3.4 A dargestellt. Zur Charakterisierung dieser Deletionsvarianten wurde ihre Affinität zur Substrat-RNA L3 gemessen (Abbildung 3.4 B). Für CFIIΔN769 konnte ein K_{50} -Wert von 9 nM und für CFIIΔN1123 von 10 nM bestimmt werden. Diese ist gegenüber dem vollständigen CFII, welches einen K_{50} -Wert von 0,6 nM besitzt, ca. um den Faktor 20 erhöht. Darüber hinaus wird in den beiden CFII-Varianten in dem eingesetzten Konzentrationsbereich keine Sättigung erreicht. Die RNA-Affinität des zuvor beschriebenen CFIIΔN1339 konnte nicht bestimmt werden.

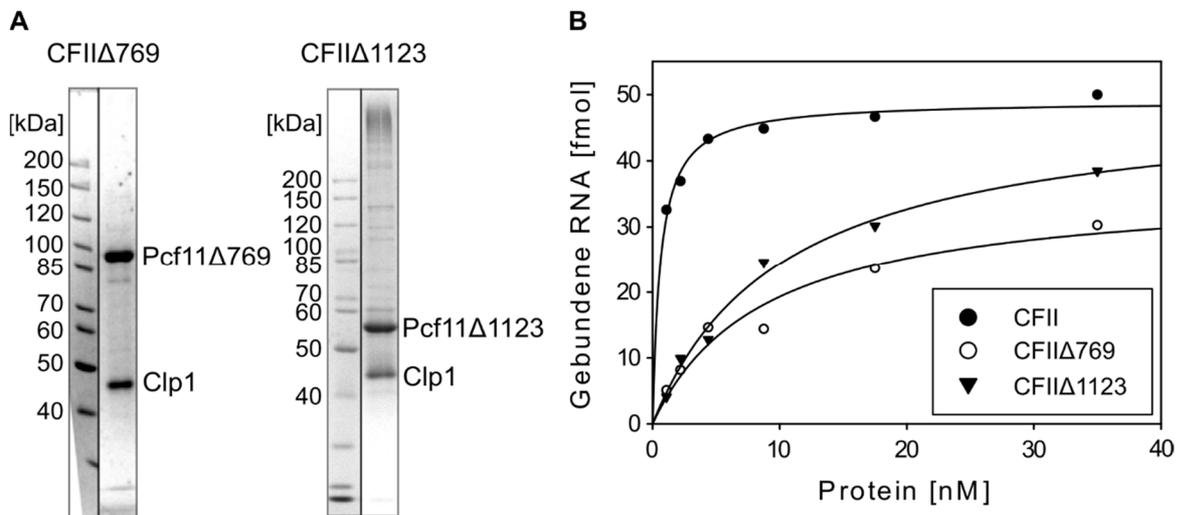


Abbildung 3.4 Analyse der RNA-Bindungen von CFII-Deletionsvarianten. A) Analytisches SDS-Polyacrylamidgel der für die Analyse der RNA-Bindung gereinigten Deletionsvarianten von CFII. B) RNA-Bindung der CFII-Deletionsvarianten. Die Messwerte für das native CFII wurden aus Abbildung 3.1 C übernommen. Die Messwerte wurden zur Bestimmung der K_{50} -Werte hyperbolisch angeglichen.

Unter dem Vorbehalt, dass es sich bei der Messung nicht um eine Gleichgewichtsreaktion handelt, kann aus dem K_{50} -Wert die Bindungsenergie mit folgender Formel berechnet werden:

$$\Delta G = RT * \ln \frac{K_D}{c^\ominus} \quad (1)$$

Wobei R der idealen Gaskonstanten, T der Temperatur von 298,15 K und c^\ominus der Referenzkonzentration von 1 mol/L entspricht.

Tabelle 5 Bindungsenergien der verschiedenen CFII-Varianten. Aus den im Experiment bestimmten K_{50} -Werten wurden die Bindungsenergien nach Formel (1) berechnet.

Komplex	K_{50} [nM]	ΔG [kJ/mol]
CFII	0,6	-52,64
CFIIΔN769	8,8	-45,98
CFIIΔN1123	10,3	-45,59

Die Bindungsenergie einer Wasserstoffbrückenbindung in Proteinen wird mit 2 bis 6 kJ/mol (0,5 bis 1,5 kcal/mol) angegeben [164]. Der Unterschied in der RNA-Affinität zwischen dem Volllänge CFII und den Deletionsvarianten entspricht daher nur der Energie von 1 bis 3 Wasserstoffbrückenbindungen. Es kann also davon ausgegangen sein, dass die RNA-Bindung durch den C-terminalen Bereich in Pcf11 vermittelt wird.

Als potentielle RNA bindende-Region in Pcf11 wurden die beiden beschriebenen Zinkfingerdomänen, die die Clp1 Interaktionsregion flankieren, identifiziert (kurz zcz, für Zinkfinger – Clp1-Interaktionsregion – Zinkfinger). Zur genaueren Untersuchung dieser Zinkfingerdomänen wurden diese, versehen mit einem Hexahistidin-SUMO-*tag* bzw. einem GST-*tag*, in *E. coli* exprimiert. Gereinigt wurden die beiden Fragmente über eine Affinitätschromatographie mittels NiNTA-Agarose bzw. GSH-Sepharose und ein anschließende Anionenaustauschchromatographie (Abbildung 3.5 A).

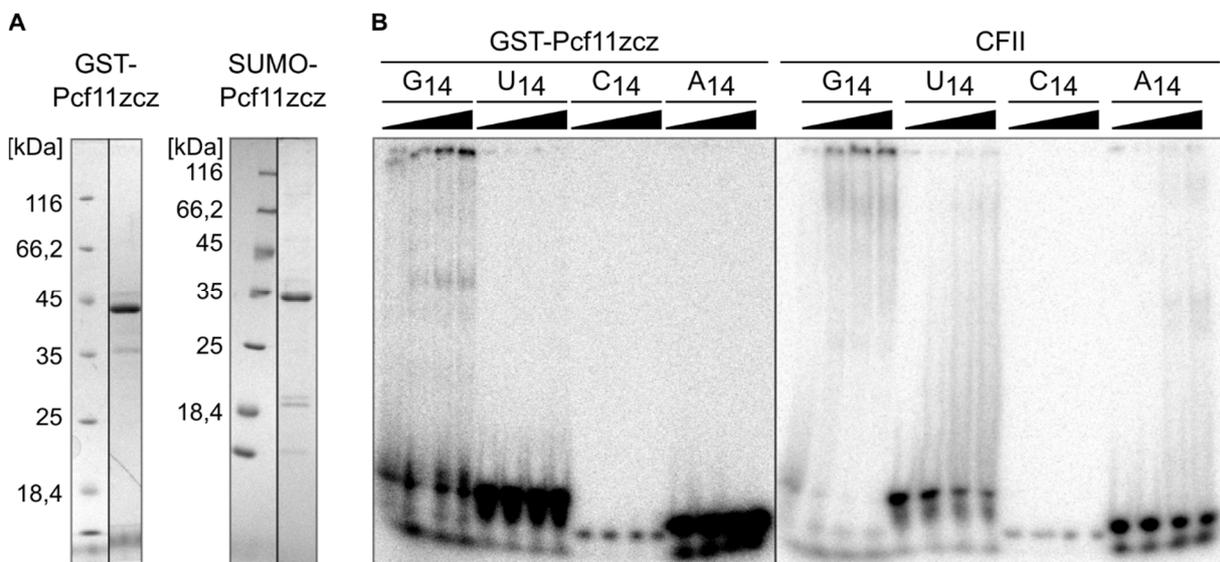


Abbildung 3.5 Reinigung zweier Varianten der Zinkfingerdomäne von Pcf11 und Analyse der Sequenzspezifität. A) Darstellung der beiden Varianten der Zinkfingerdomäne von Pcf11 nach der Anionenaustauschchromatographie. B) *Gel-Shift*-Experimente zur Bestimmung der RNA-Bindung von Homopolymeren von GST-Pcf11 und CFII. Steigende Proteinmengen wurden gegen eine konstante RNA-Menge von 2 fmol in 20 μ l Gesamtvolumen eingesetzt. Verwendet wurden 0, 5, 25 und 100 fmol natives CFII sowie 0, 50, 125, 500 fmol GST-Pcf11zcz.

Trotz einer proteolytischen Spaltstelle in beiden Varianten war eine Entfernung des *tags* nicht möglich. Es konnte für beide verwendeten Proteasen (Sumo-Protease sowie PreScission™) keine Spaltung beobachtet werden, was vermutlich an einer sterischen Blockierung der Spaltstelle durch die Zinkfingerdomäne lag. Beide Fusionsproteine zeigten eine Bindung der L3-RNA im submikromolaren Bereich. Außerdem zeigen die Zinkfinger, wie auch natives CFII, eine Spezifität für Poly-Guanosin, während Poly-Uracil, Poly-Adenosin sowie Poly-Cytosin nicht gebunden werden (Abbildung 3.5 B). Die Abstufung der Spezifität von Poly(G) über Poly(U) zu

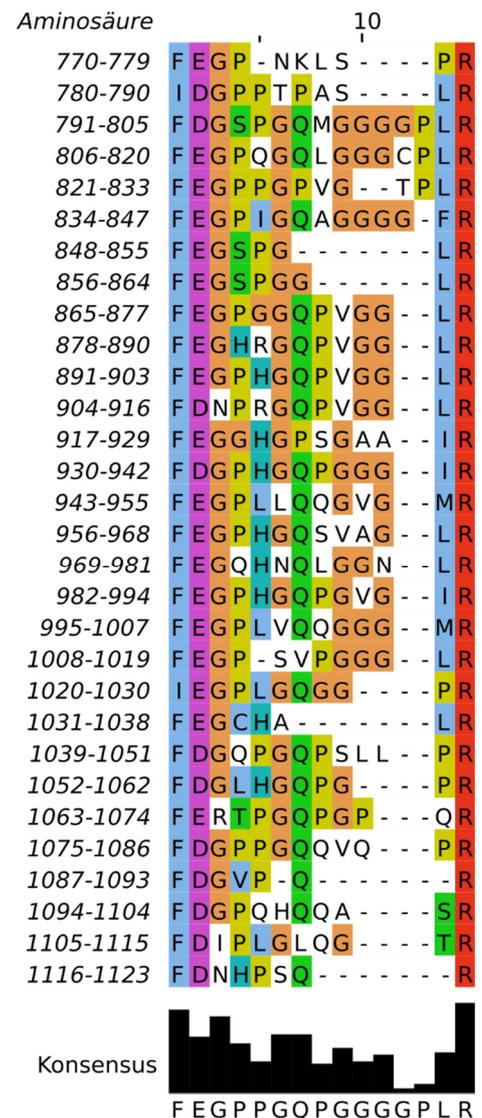
Poly(A) ist in der gereinigten Zinkfingerdomäne nicht zu beobachten. Hier liegt eine exklusive Bindung von Poly(G) vor, während keine Bindung für die anderen Homopolymere zu beobachten ist. Darüber hinaus ist die Affinität im Vergleich zu nativem CFII-Komplex reduziert. Während für CFII bereits 5 fmol reichen, um eine vollständige Bindung von Poly(G) zu erreichen, ist auch bei 0,5 pmol Zinkfingerdomäne noch freies Poly(G) zu beobachten.

Darüber hinaus zeigten beide Fusionsproteine dieselbe Längenabhängigkeit wie natives CFII. So wurde erst Poly(G) ab einer Länge von 12 Nukleotiden messbar gebunden (Daten nicht gezeigt).

3.1.1.5 Analyse der FEGP-Wiederholungen

Pcf11 zeigt zwischen der strukturierten *cross-link*-Domäne Pcf11₅₀₀₋₇₇₀ und dem C-Terminus mit der Zinkfinger- und Interaktionsdomäne einen Sequenzbereich, der sich durch die 30-fache Wiederholung der Konsensussequenz „FEGPPGQPGGGGLR“ auszeichnet (Abbildung 3.6). Dieser Bereich ist in Wirbeltieren konserviert (Anhang Abbildung A. 1). Nicht vorhanden ist dieser Sequenzbereich u.a. in Hefe, der Taufliege (*Drosophila melanogaster*) oder Arabidopsis. Massenspektrometrische Analysen des rekombinanten CFII zeigten, dass die abschließenden Arginine der Wiederholungen methyliert sind. Auf die gesamte Sequenz bezogen sind von den Argininen ca. 14% mono- und 30% dimethyliert. Diese Modifikationen konnten auch in endogenem Pcf11 identifiziert werden, wobei hier die Arginin-Modifizierungen als asymmetrische Methylierung vorliegen [165].

Abbildung 3.6 Sequenzangleichung der humanen FEGP-Wiederholungen. Dargestellt ist die fortlaufende Aminosäuresequenz zwischen den Aminosäuren 770 und 1123. Gleiche oder ähnliche Aminosäuren sind farblich markiert. Der Grad der Konservierung wird durch ein Balkendiagramm sowie die Konsensussequenz unterhalb dargestellt.



Zur weiteren Untersuchung der Funktion dieser Wiederholungen wurden die vorher beschriebene verkürzten CFII-Varianten (siehe 2.1.1.4), welche vor bzw. hinter den Wiederholungen beginnen, auf die Rekonstitution der Spaltaktivität in fraktioniertem Kernextrakt untersucht. Eine konzentrationsabhängige Spaltung der Substrat-RNA L3 konnte nur im CFIIΔN769-Komplex beobachtet werden, welche die FEGP-Wiederholungen enthielten, nicht jedoch im CFIIΔN1123-

Komplex, welcher diese nicht enthielt (Abbildung 3.7). Die FEGP-Wiederholungen sind daher, trotz ihrer unbekannteten Struktur und Funktion, essentiell für die Spaltung Substrat-RNA.

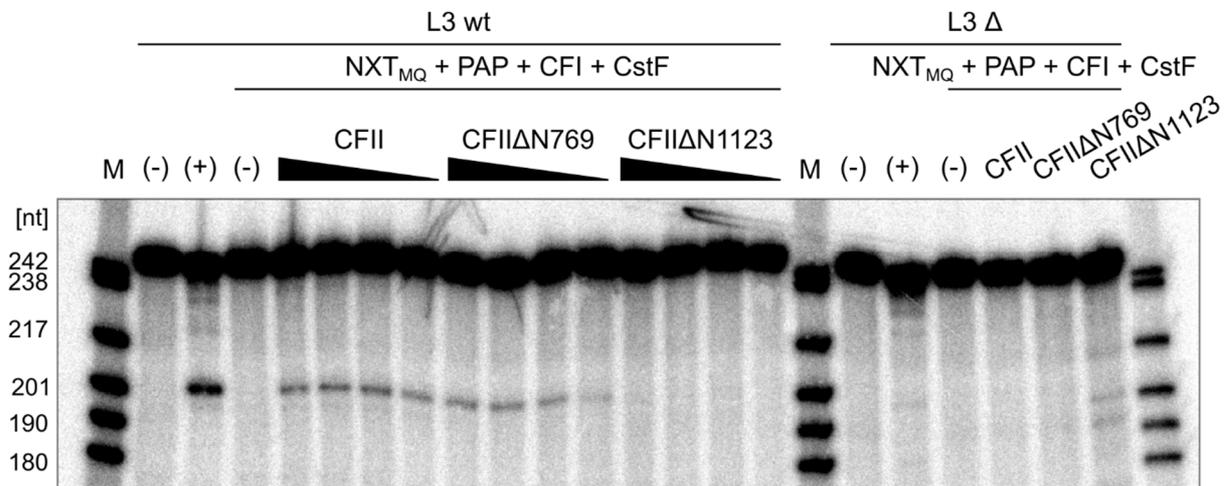


Abbildung 3.7 Rekonstitution der Spaltung der Substrat-RNA durch CFII. Analyse der Spaltung der L3-RNA. Eingesetzt wurden fraktionierter Kernextrakt (NXT_{MQ}), sowie die rekombinanten Proteine Poly(A)-Polymerase (PAP), CFI und CstF. In absteigender Menge (100, 50, 25, 5 fmol) wurden vollständiges CFII, CFII Δ 769 oder CFII Δ 1123 eingesetzt. Als Kontrollen wurde RNA ohne Proteine (-), Kernextrakt (+) sowie rekonstituierter fraktionierter Kernextrakt ohne CFII (-) eingesetzt. Zur Spezifitätskontrolle wurde punktmutierte RNA (L3 Δ) eingesetzt. Die eingesetzten CFII-Mengen entsprachen hier der Maximalmenge von 100 fmol.

Der Versuch, die FEGP-Wiederholungen rekombinant in *E. coli* zu exprimieren und zu reinigen, schlug fehl. Das Protein war unter nativen Standardbedingungen nicht löslich (Daten nicht gezeigt).

3.1.2 Spaltungs- und Polyadenylierungs-Spezifitätsfaktor (CPSF)

Der Spaltungs- und Polyadenylierungs-Spezifitätsfaktor (CPSF) kann in zwei Subkomplexe getrennt werden: den Endonuklease-Komplex, bestehend aus den Untereinheiten CPSF73, CPSF100, Symplekin und CstF64, und den Polyadenylierungskomplex (mPSF), bestehend aus CPSF30, CPSF160, Fip1 und WDR33. Ob CstF64 tatsächlich Teil des CPSF-Komplexes ist, ist bisher noch unklar (siehe Einleitung). Eine stabile Interaktion zwischen Endonuklease- und Polyadenylierungskomplex konnte *in vitro* bisher nicht beobachtet werden. Zur Analyse des CPSF-Komplexes wurde dieser rekombinant in Insektenzellen exprimiert. Hierzu wurden die Zellen mit drei unabhängigen Viren, welche den mPSF-, den Endonuklease-Komplex ohne CstF64 sowie eine Hexahistidin-FLAG-*tag* markierte CstF64-Variante kodierten, koinfiziert. Es erfolgte eine zweistufige Affinitätsreinigung, wobei für beide Reinigungsschritte der Hexahistidin-FLAG-*tag* an CstF64 genutzt wurde. Die Reinigung wurde über eine *Western-Blot*-Analyse mithilfe von Antikörpern gegen CPSF100, stellvertretend für den Endonuklease-Komplex, sowie gegen Fip1, stellvertretend für den mPSF-Komplex, verfolgt. Die initiale Reinigung erfolgte über den Hexahistidin-*tag* an CstF64. Deutlich angereicht sind die Proteine des Endonuklease-Komplexes, jedoch konnten auch Untereinheiten des mPSF-Komplexes identifiziert werden. Eine nachfolgende Reinigung über den FLAG-*tag* von CstF64 trennte diese jedoch voneinander und nur die Proteine des Endonuklease-Komplexes konnten identifiziert werden, welches die stabile Interaktion zwischen CstF64 und den Proteinen des Endonuklease-Komplexes bestätigt (Abbildung 3.8).

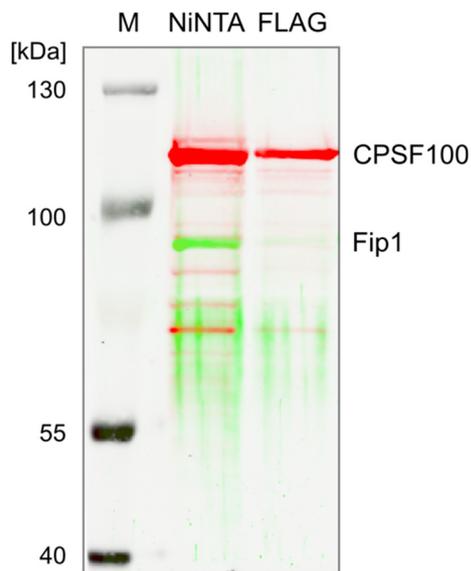


Abbildung 3.8 *Western-Blot*-Analyse der zweistufigen Reinigung von CPSF. Dargestellt sind die Elutionsfraktionen der ersten Reinigung über den Hexahistidin-*tag* (NiNTA) an CstF64 sowie der nachfolgenden Reinigung über den FLAG-*tag* (FLAG), ebenfalls an CstF64. Durch spezifische Antikörper wurden CPSF100 (roter Kanal) sowie Fip1 (grüner Kanal) nachgewiesen.

Zur Strukturanalyse des CPSF-Komplexes werden daher im Folgenden die beiden Untereinheiten unabhängig voneinander exprimiert und gereinigt.

3.1.3.1 Polyadenylierungs-Spezifitäts-Faktor (mPSF)

3.1.3.1.1 Darstellung und Aktivitätsmessung

Der mPSF-Komplex wurde als separierter Komplex gereinigt, wobei dieser im Gegensatz zu dem vorherigen Experiment einen Hexahistidin-*tag* an WDR33 trug. Über diesen wurde der mPSF-Komplex affinitätschromatographisch gereinigt. Es folgte anschließend eine Anionenaustauschchromatographie (Abbildung 3.9 A). Da mPSF die Poly(A)-Polymerase stimuliert, wurde die Reinigung mittels Polyadenylierungsassay verfolgt und verifiziert (Abbildung

3.9 B). Quantifiziert wurde das Protein wie beschrieben über die Konzentration von Fip1 [26]. Insgesamt konnten in den quantifizierten Hauptfraktionen 1,7 mg Protein gereinigt werden, welches einer Proteinausbeute von 2,1 mg pro Liter Kulturvolumen entsprach. Basierend auf dem Coomassie-gefärbtem SDS-Gel der Reinigung liegen die Proteine des mPSF-Komplexes jedoch nicht äquimolar vor.

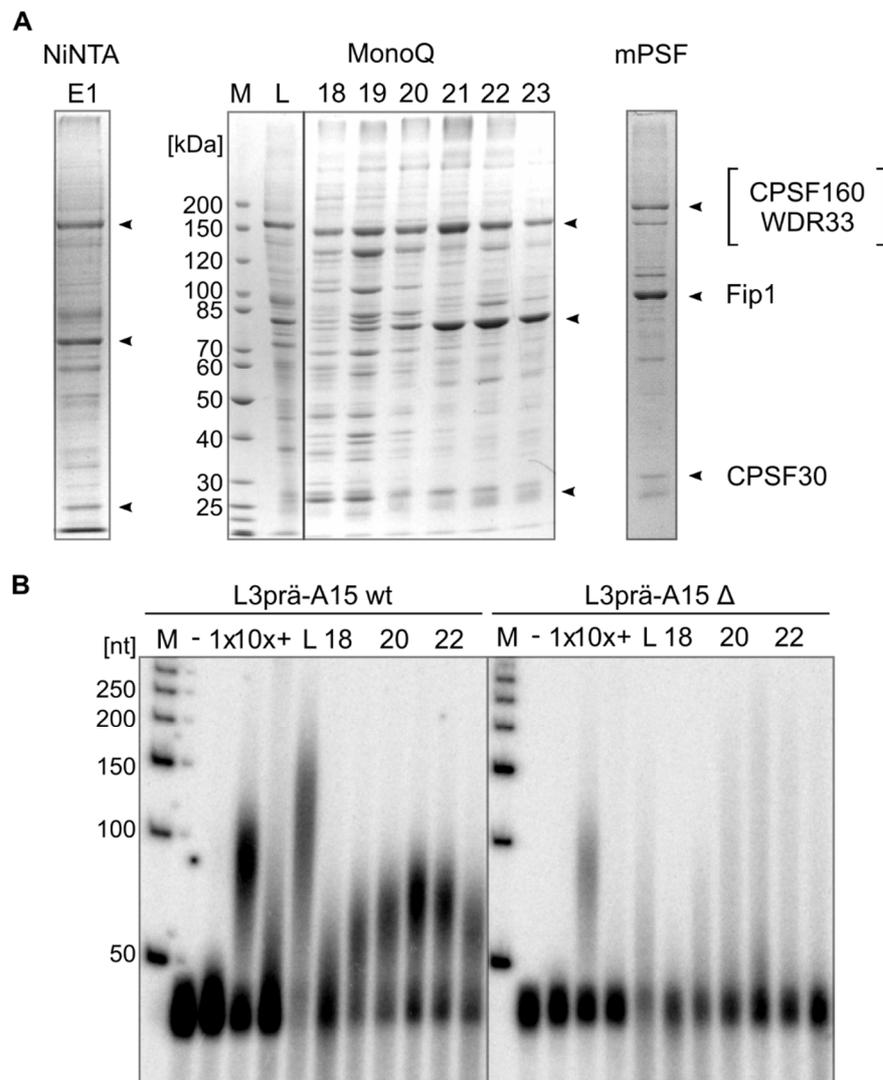


Abbildung 3.9 Reinigung und Charakterisierung von mPSF. A) Darstellung der Reinigung des mPSF-Komplexes. Die Elutionsfraktion der initialen Affinitätschromatographie (NiNTA – E1), die Hauptfraktionen der Anionenaustauschchromatographie (MonoQ) sowie das quantifizierte Protein sind dargestellt (mPSF). Die Untereinheiten CPSF30, Fip1, WDR33 sowie CPSF160 sind mit Pfeilen angezeigt. Da WDR33 und CPSF160 nicht voneinander trennbar waren, sind diese zusammengefasst. B) Stimulierung der Poly(A)-Polymerase durch mPSF. Präadenylierte Substrat-RNA L3prä-A15 in nativer (wt) oder punktmutierter (Δ) Variante wurde in Anwesenheit von 100 fmol Poly(A)-Polymerase mit dem Input der finalen Reinigung (L), sowie den Fraktionen 18 bis 23 (je 1 μ l) und die Reaktionsprodukte mittels Harnstoff-PAGE getrennt. Zur Kontrolle der Reaktion wurde die RNA ohne Protein (-), mit einfacher (1x; 100 fmol) oder zehnfacher Menge

(10x; 1 pmol) Poly(A)-Polymerase ohne Anwesenheit anderer Proteine sowie mit gereinigtem CPSF-Komplex aus Kalbsthymus und 100 fmol Poly(A)-Polymerase (+) inkubiert [166].

3.1.3.1.2 Chemisches *cross-linking* der Polyadenylierungsreaktion

Die *in-vitro*-Polyadenylierung läuft in drei Hauptschritten ab: der Bildung des Komplexes auf der mit 15 Adenosinen präadenylierten Substrat-RNA (L3prä-A15), die Elongation des Poly(A)-Schwanzes, wobei dieser unmittelbar von PABPN1 bedeckt wird, sowie der Reaktionstermination nach ca. 250 Adenosinen, der sogenannten Längenkontrolle [166]. Zur Analyse der Struktur dieser Intermediate der Polyadenylierung mussten Bedingungen definiert werden, unter denen eine Reaktion stattfindet, also ein funktioneller Komplex gebildet wird, die Intermediate getrennt und das chemische *cross-linking* durchgeführt werden können. Unter Standardbedingungen liegt die Poly(A)-Polymerase im Überschuss gegenüber mPSF vor. Da jedoch der aktive Komplex analysiert werden soll, mussten diese äquimolar eingesetzt werden. Die Polyadenylierung ist ein sehr schneller Prozess, der nach weniger als einer Minute bereits abgeschlossen ist. Um ein Zeitrahmen zu schaffen, in denen die Elongationsphase zu isolieren war, wurde die Reaktionstemperatur auf 25 °C reduziert. Um darüber hinaus die Signalintensität in der nachfolgenden massenspektrometrischen Analyse zu erhöhen, wurden die Proteine 10-fach konzentrierter als üblich in die Reaktion eingesetzt. Die Konzentration der Substrat-RNA wurde dementsprechend angepasst. Die Elongation der RNA konnte durch Zugabe von EDTA, welches die Magnesiumionen komplexiert, gestoppt werden. Es konnte keine weitere Verlängerung der RNA in der nachfolgenden Inkubationszeit beobachtet werden (Abbildung 3.10 A). Ein *cross-linking* unter Reaktionsbedingungen sollte somit möglich sein.

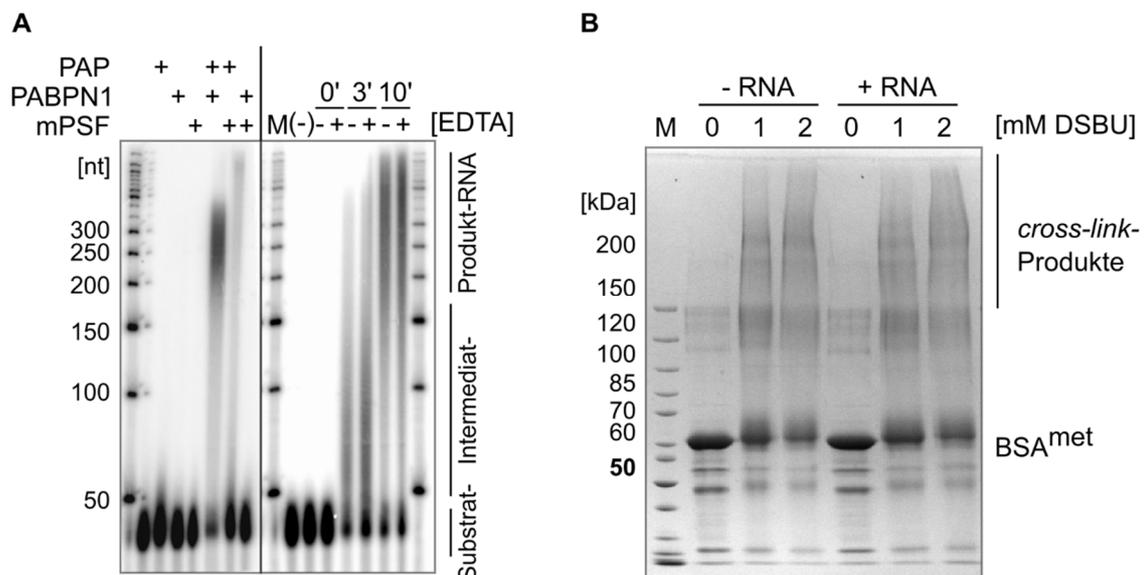


Abbildung 3.10 Validierung der *cross-link*-Bedingungen für die Polyadenylierungsreaktion.

A) Isolierung der Reaktionsintermediate der modifizierten Polyadenylierungsreaktion. Zur Kontrolle wurden die RNA ohne Proteine inkubiert (2. Spur), sowie die binären Kombinationen aus Poly(A)-Polymerase (PAP), dem Poly(A)-bindenden Proteine (PABPN1) sowie mPSF. Die

Reaktionen mit allen Proteinen wurden nach dem angezeigten Zeitpunkt gestoppt. Die Kontrollreaktionen (-) wurden unmittelbar mit Proteinase K verdaut, während die Proben mit EDTA (+) vor dem Verdau über 30 Minuten weiter inkubiert wurden. B) Validierung der optimalen *cross-linker*-Konzentration in der Polyadenylierungsreaktion. Es wurde in Ab- und Anwesenheit von RNA eine Konzentration von 0 bis 2 mM eingesetzt. Gebildete hochmolekulare *cross-link*-Produkte sind gekennzeichnet. Das in der Reaktion vorhandene methylierte BSA (BSA^{met}) ist zusätzlich hervorgehoben.

Zur Bestimmung der optimalen *cross-linker*-Konzentration wurde dieser gegenüber der Polyadenylierungsreaktion titriert. Die Reaktionen wurden sowohl mit als auch ohne RNA durchgeführt, um einen Einfluss der Nukleinsäuren auf die Reaktion zu überprüfen. Eine erste Titration in dem Konzentrationsbereich zwischen 50 bis 500 µM, welches einem bis zu 1000x molarem Überschuss entsprach, zeigte kaum *cross-link*-Produkte (Daten nicht gezeigt). In einer zweiten Reaktion wurde 1 bzw. 2 mM *cross-linker* eingesetzt (Abbildung 3.10 B). Die eingesetzte RNA hatte keinen Einfluss auf die chemische Vernetzung der Proteine. In beiden Reaktionen konnten hochmolekulare *cross-link*-Produkte optisch identifiziert werden. Um artifizielle Zustände durch eine zu hohe Konzentration an *cross-linker* zu vermeiden, wurde für die massenspektrometrische Analyse eine Konzentration von 1 mM gewählt.

Die Polyadenylierungsreaktion enthielt zur Stabilisierung methyliertes BSA. Die Konzentration betrug 0,2 g/l, was einer molaren Konzentration von 2,9 µM entsprach. Dieses ist ungefähr 100-fach stärker konzentriert als die eingesetzten Polyadenylierungsfaktoren. Der Reaktionsansatz wurde zum einen in Lösung verdaut, wobei hier zwar eine bessere Identifikation erwartet wurde, aber auch mit einem hohen BSA-Hintergrund zu rechnen war. Um *cross-link*-Produkte von BSA zu trennen, wurden die Reaktionsprodukte zum anderen mittels eines SDS-Polyacrylamidgeles getrennt und im Gel verdaut. Da BSA als Monomer (ca. 70 kDa) oder Dimer (ca. 140 kDa) vorliegen kann, wurde der Bereich oberhalb von 150 kDa isoliert und verdaut (Abbildung 3.11 A). Bei dem Reaktionsprodukt bei ca. 85 kDa, welches unmittelbar oberhalb des Albumins im SDS-Gel lief, handelte es sich um durch den *cross-linker* modifiziertes Albumin (Daten nicht gezeigt).

Obwohl die Analyse von Peptiden vom Albumin dominiert war, konnten in dem Verdau in Lösung 13 einzigartige *cross-link*-Paare identifiziert werden. Ein intermolekularer *cross-link* war zwischen CPSF160 und dem N-Terminus des Poly(A)-bindendem Protein PABPN1. Der N-Terminus von PABPN1 ist nicht an der RNA-Bindung beteiligt, aber essentiell für die Stimulierung der Poly(A)-Polymerase [127]. Zwei intramolekulare *cross-links* konnten in Fip1 gefunden werden. Die in PABPN1 gefundenen *cross-links* können nicht zweifelsfrei zugeordnet werden, da sie in der RRM liegen, von der eine Dimerisierung gezeigt werden konnte [125]. Diese Dimerisierung ist jedoch nicht notwendig für eine RNA-Bindung, sondern könnte die Folge der Polymerisierung auf dem Poly(A)-Schwanz darstellen. Einer der *cross-links* kann zweifelsfrei einer dimeren Struktur zugeordnet werden. Keine *cross-links* konnten zu den Proteinen CPSF30, WDR33 und der

Poly(A)-Polymerase identifiziert werden. Die Interaktion zwischen CPSF160 und PABPN1 ist RNA-unabhängig identifiziert worden, die eindeutige Dimerbildung von PABPN1 hingegen erst in den Polyadenylierungsendprodukten. Die Ergebnisse der *cross-links* wurden daher in einem Diagramm zusammengefasst (Abbildung 3.11 B; Tabelle A 6).

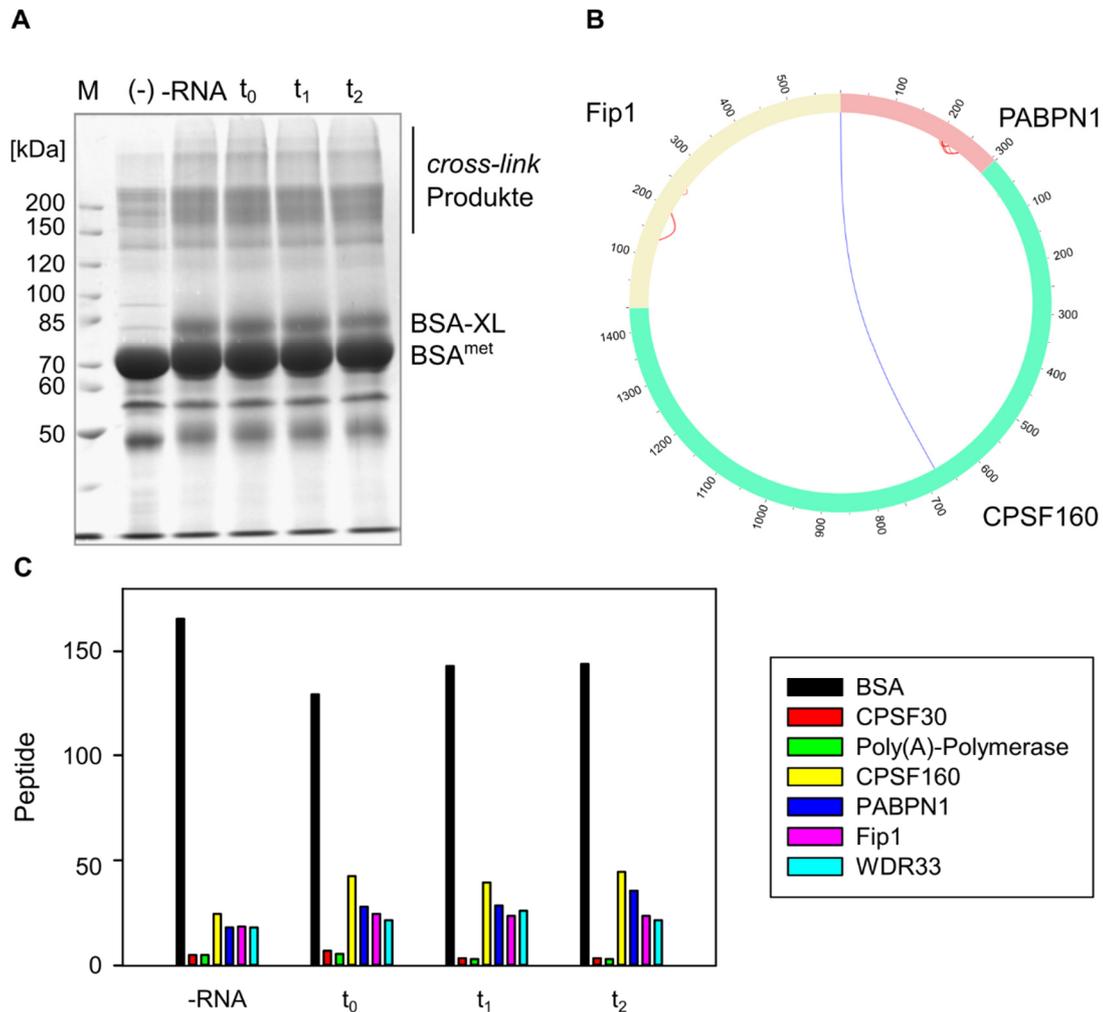


Abbildung 3.11 Präparation und Analyse des chemischen *cross-linking* der Polyadenylierungsreaktion. A) Präparatives *cross-linking* der Polyadenylierungsreaktion. Die Reaktion wurde zur Kontrolle ohne Zugabe von RNA (-RNA), und nach Zugabe der RNA nach 0, 3 und 10 Minuten chemisch (t₀-t₂) vernetzt und die Reaktionsprodukte mittels SDS-Polyacrylamidgel getrennt. Die Reaktionen wurden komplett in Lösung oder die angezeigten *cross-link*-Produkte über 150 kDa sowie die Bande oberhalb von BSA (BSA-XL) ausgeschnitten und anschließend verdaut. B) Kreisdiagramm der identifizierten *cross-links*. Dargestellt sind die 3 Proteine im Uhrzeigersinn sowie die intra- (rot) und intermolekularen (blau) *cross-links*. C) Balkendiagramm der quantitativen Analyse des In-Gel Verdauens. Dargestellt sind die Anzahl der identifizierten Peptide (MS/MS counts) der jeweiligen *cross-link*-Reaktion. BSA bezeichnet das A) mit „BSA-XL“ gekennzeichnete Reaktionsprodukt.

Im zweiten Ansatz wurden die Reaktionsprodukte mittels SDS-Polyacrylamidgel getrennt. Die massenspektrometrische Analyse wurde jedoch weiterhin von BSA-Peptiden dominiert (siehe

Abbildung 3.11 C) und es konnten keine *cross-links* identifiziert werden. Eine detaillierte Analyse der Rohdaten mit MaxQuant zeigte, dass in jedem Gelstück rund 150 Peptide für Albumin, aber nur maximal je 30 bis 50 Peptide der Polyadenylierungsfaktoren, für CPSF30 sowie der Poly(A)-Polymerase nur rund 5 Peptide, identifiziert werden konnten. In dem direkten Vergleich der Reaktionen, welche in An- oder Abwesenheit von RNA vernetzt wurden, zeigt sich eine Steigerung der Anzahl der Peptide (Abbildung 3.11 C – „-RNA“ vs. „t₀“). Dieses deutet auf eine RNA-abhängige Bildung oder Stabilisierung des Komplexes hin. Im Vergleich zwischen den Intermediaten („t₀“ bis „t₂“) ist nur ein minimaler Unterschied in der Menge der Peptide von PABPN1 zu beobachten, was nicht linear zur Länge des Poly(A)-Schwanzes ist. Über die genaue Zusammensetzung des Komplexes zu den verschiedenen Zeitpunkten kann auf Grundlage der vorliegenden Daten keine weitere Aussage getroffen werden.

3.1.3.2 Endonuklease-Komplex

Der Endonuklease-Komplex wurde über den Hexahistidin-*tag* an CstF64 affinitätschromatographisch gereinigt. Anschließend erfolgte eine Anionenaustauschchromatographie, bei der der Komplex in einem sehr klar definierten Peak (Fraktionen 18 und 19) eluiert werden konnte (Abbildung 3.12 A). Die Quantifizierung im Gel zeigte eine stöchiometrische Zusammensetzung des Komplexes. Dies bestätigt die stabile Assoziation von CstF64 zu dem Endonuklease-Komplex. Es konnte aus 800 ml Kulturvolumen 1,9 mg Komplex isoliert werden, was einer Ausbeute von 2,4 mg je Liter Kultur entspricht.

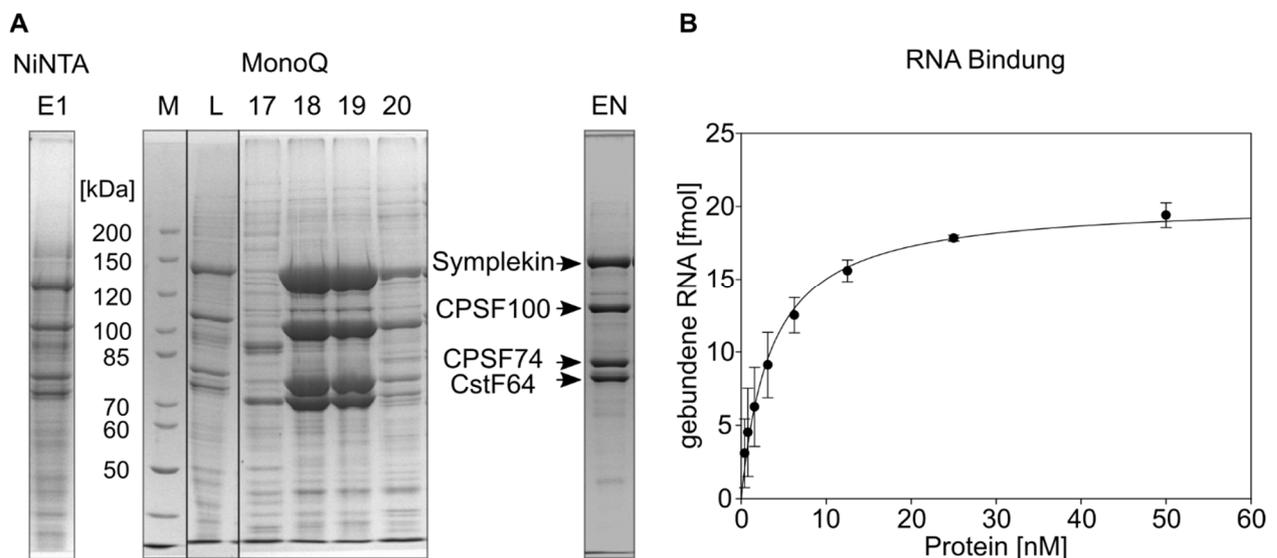


Abbildung 3.12 Expression, Reinigung und Aktivitätsanalyse des Endonuklease-Komplexes. A) Reinigung des Endonuklease-Komplexes über die initiale Affinitätschromatographie (NiNTA E1) und anschließende Anionenaustauschchromatographie (MonoQ) sowie Darstellung des quantifizierten Proteins (EN). B) RNA-Bindung des Endonuklease-Komplexes. Steigende Konzentrationen an Endonuklease-Komplex wurden eingesetzt. Die Bindung der Substrat-RNA L3 wurde mittels Bindung an Nitrozellulosefilter und anschließender Szintillationszählung gemessen und die Messwerte hyperbolisch angepasst.

Da für die Aktivität des Endonuklease-Komplexes, die Spaltung der prä-mRNA, keine beschriebene Nachweismethode vorhanden ist, konnte diese nicht getestet werden. Auch konnte keine Stimulierung der Spaltung von prä-mRNA durch HeLa-Kernextrakt beobachtet werden (Daten von P. Schäfer). Stattdessen wurde die Bindung der Substrat-RNA L3 durch den Endonuklease-Komplex bestimmt (Abbildung 3.12 B). Es konnte ein K_{50} -Wert von 3,5 nM ermittelt werden. Unklar ist jedoch, ob diese RNA-Bindung ausschließlich durch CstF64 oder durch einen oder mehrere weitere Faktoren vermittelt wird.

3.1.3.2.2 Chemisches *cross-linking* des Endonuklease-Komplexes

Zur Identifikation der optimalen Konzentration des *cross-linkers* wurde eine geometrische Titrationsreihe durchgeführt und die Bildung von quervernetzten Produkten mittels SDS-PAGE analysiert. Hierzu wurden konstante 6 μ M an rekombinantem Protein eingesetzt. Ein ca. 100-facher molarer Überschuss des *cross-linkers* zeigt eine deutliche Bildung von Produkten, wobei noch monomeres Protein vorhanden ist (Abbildung 3.13 A).

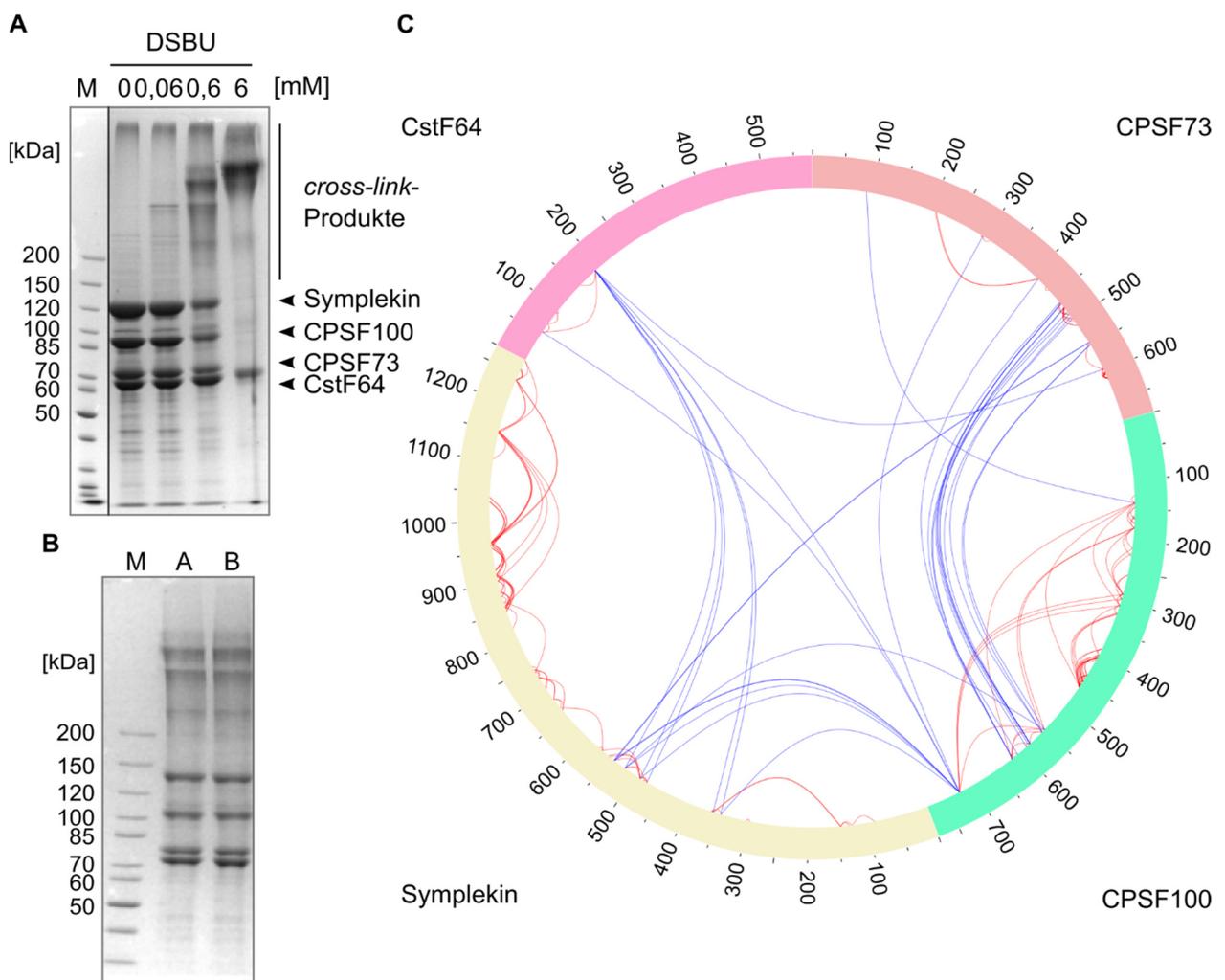


Abbildung 3.13 Chemisches *cross-linking* des Endonuklease-Komplexes. A) Visualisierung der Titration des *cross-linkers* mittels Coomassie-gefärbten SDS-PAGE. Der *cross-linker* wurde bei konstanter Proteinkonzentration von 6 μ M von 10 bis 1000-fachem molarem Überschuss

eingesetzt. Die Untereinheiten sowie die Reaktionsprodukte sind gekennzeichnet. B) Analyse des präparativen *cross-linkings* des Endonuklease-Komplexes. A und B bezeichnet technische Replikate. Zum *cross-linking* wurde 1 mM *cross-linker* verwendet. 10 % der Reaktion wurde für das analytische SDS-Polyacrylamidgel verwendet. C) Visualisierung der identifizierten *cross-links* in einem Kreisdiagramm. Intra- (rot) und intermolekulare (blau) *cross-links* sind eingezeichnet, die Proteine sind im Uhrzeigersinn nach ihrer Größe und von N- zu C-Terminus angeordnet.

Das präparative *cross-linking* wurde bei einer, gegenüber der Titration leicht erhöhten, Konzentration von 1 mM durchgeführt. Es wurden technische Replikate mit je 360 pmol durchgeführt, von denen je 10 % via SDS-Gel analysiert (Abbildung 3.13 B; Tabelle A 7) und 90 % für die massenspektrometrische Analyse in Lösung verdaut wurden. Nach manueller Validierung der Daten konnten 209 einzigartige *cross-link*-Paare identifiziert werden, von denen 176 intra- und 33 intermolekular sind (Abbildung 3.13 C). Da kein intermolekularer *cross-link* zwischen zwei Kopien desselben Proteins identifiziert werden konnte, liegt der Endonuklease-Komplex höchstwahrscheinlich als Heterotetramer mit je einer Kopie der Proteine vor.

3.1.3 Strukturanalyse des Spaltungsstimulations-Faktors (CstF)

3.1.3.1 Darstellung und Aktivitätsmessung

Der CstF-Komplex wurde affinitätschromatographisch über einen Hexahistidin-*tag* an CstF77 und anschließende Anionenaustauschchromatographie gereinigt (Abbildung 3.14 A). Die Proteinexpression wurde in biologischen Replikaten in 400 bzw. 800 ml Kulturvolumen durchgeführt. Die Proteinmengen in den quantifizierten Hauptfraktionen betrugen 1,1 bzw. 6,2 mg, was einer Ausbeute von 2,8 bzw. 7,7 mg Protein pro Liter Kultur entsprach. Zur Überprüfung der Aktivität des gereinigten Komplexes wurden zwei verschiedene Analysen durchgeführt. Zum einen wurde erfolgreich die Komplementation der Spaltung von prä-mRNA in fraktioniertem Extrakt getestet (Abbildung 3.14 B). Zum anderen wurde die Bindung der Substrat-RNA L3 getestet (Abbildung 3.14 C). Hier konnte ein K_{50} -Wert von 11 nM bestimmt werden.

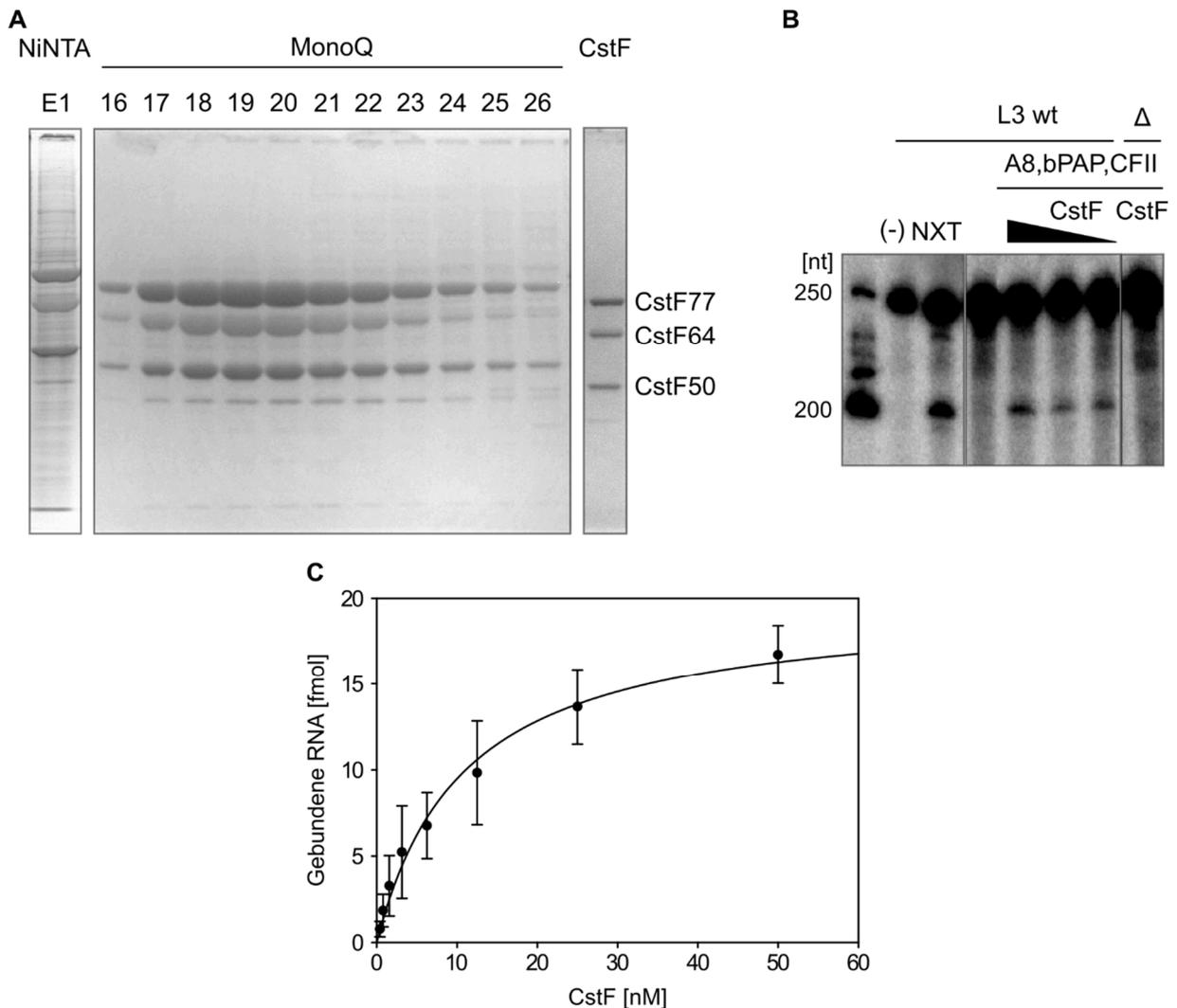


Abbildung 3.14 Reinigung und Charakterisierung des CstF-Komplexes. A) Verfolgung der Reinigung über die Affinitätschromatographie (NiNTA E1) sowie anschließende Anionenaustauschchromatographie (MonoQ) und Quantifizierung (CstF). B) Rekonstituierung der Spaltung von prä-mRNA in fraktioniertem Extrakt. Die Substrat-RNA in nativer Form sowie

punktmutiert (Δ) wurde mit fraktioniertem Extrakt (NXT_{A8}), den rekombinanten Proteinen Poly(A)-Polymerase (PAP), CFII sowie absteigenden Mengen rekombinatem CstF (50; 25; 12,5 fmol) umgesetzt. Als Kontrollen wurde kein Protein (-), kommerzieller Kernextrakt (NXT) sowie eine rekonstituierte Reaktion ohne CstF eingesetzt. Die Kontrollreaktion mit der mutierten RNA wurde mit 50 fmol CstF durchgeführt. C) Bindung der Substrat-RNA L3 durch CstF. Steigende Konzentrationen von CstF wurden eingesetzt und die Bindung der RNA durch Bindung an Nitrozellulose und anschließende Szintillationszählung gemessen. Die Messwerte wurden zur Bestimmung der Dissoziationskonstante hyperbolisch ausgewertet.

3.1.3.2. Chemische *cross-linking*

Zur strukturellen Analyse des CstF-Komplexes wurde der *cross-linker* DSBU verwendet. Zur Bestimmung des optimalen Verhältnisses wurde dieser gegen eine konstante Proteinkonzentration von 6 μM titriert. Es konnten ab einem 100-fachem molaren Überschuss hochmolekulare Produkte und ab einem 1000-fachen Überschuss nur noch ein Hauptprodukt identifiziert werden (Abbildung 3.15 A).

Um Reaktionsartefakte zu vermeiden, wurde für die massenspektrometrische Analyse nur ein 166-facher Überschuss des *cross-linkers* gegenüber dem Komplex verwendet. Es wurden in zwei getrennten Reaktionen je 360 pmol CstF eingesetzt, von denen je zwei massenspektrometrische Analysen durchgeführt wurden, sodass insgesamt 4 Datensätze zur Analyse zur Verfügung standen (Abbildung 3.15 B). Es konnten insgesamt 326 einzigartige *cross-link*-Paare identifiziert werden, von denen 192 intra- und 134 intermolekular waren (Abbildung 3.15 C). Eine Differenzierung von intermolekularen *cross-links* innerhalb der homodimeren Teile des CstF-Komplexes ist nicht ohne weiteres möglich, sodass diese als Teil der intramolekularen *cross-links* betrachtet werden. Von den intermolekularen *cross-links* waren 3 eindeutig homodimeren Substrukturen von CstF₆₄ zuzuordnen (Abbildung 3.15 C; Tabelle A 8). Neben der dargestellten Interaktion Lysin-96 x Lysin-96 konnten Verknüpfungen zwischen den Lysin-55 x Lysin-55 sowie Lysin-189 x Lysin-189 identifiziert werden. Die übrigen 131 intermolekularen *cross-links* sind zwischen verschiedenen Untereinheiten identifiziert worden.

Linien dargestellt. D) MeroX-Detailansicht eines Spektrums eines eindeutig identifizierten *cross-links* eines CstF64 Dimers [145]. Identifiziert werden konnte eine Verknüpfung zwischen den Aminosäuren Lysin-96 des ersten Proteins sowie Lysin-96 des zweiten Proteins. Im α -Peptid wurde eine Trypsin-Spaltstelle übersprungen, sodass dieses eine andere Masse als das β -Peptid aufweist, was eine eindeutige Identifizierung möglich macht. Rechts neben dem Spektrum sind schematisch die identifizierten Ionen dargestellt.

3.1.4 Strukturanalyse des Spaltungsfaktors I (CFI)

3.1.4.1 Darstellung und Aktivitätsmessung

Der CFI-Komplex wurde affinitätschromatographisch zunächst über den Hexahistidin-*tag* an CFI25 und anschließend durch Anionenaustauschchromatographie gereinigt (Abbildung 3.16 A).

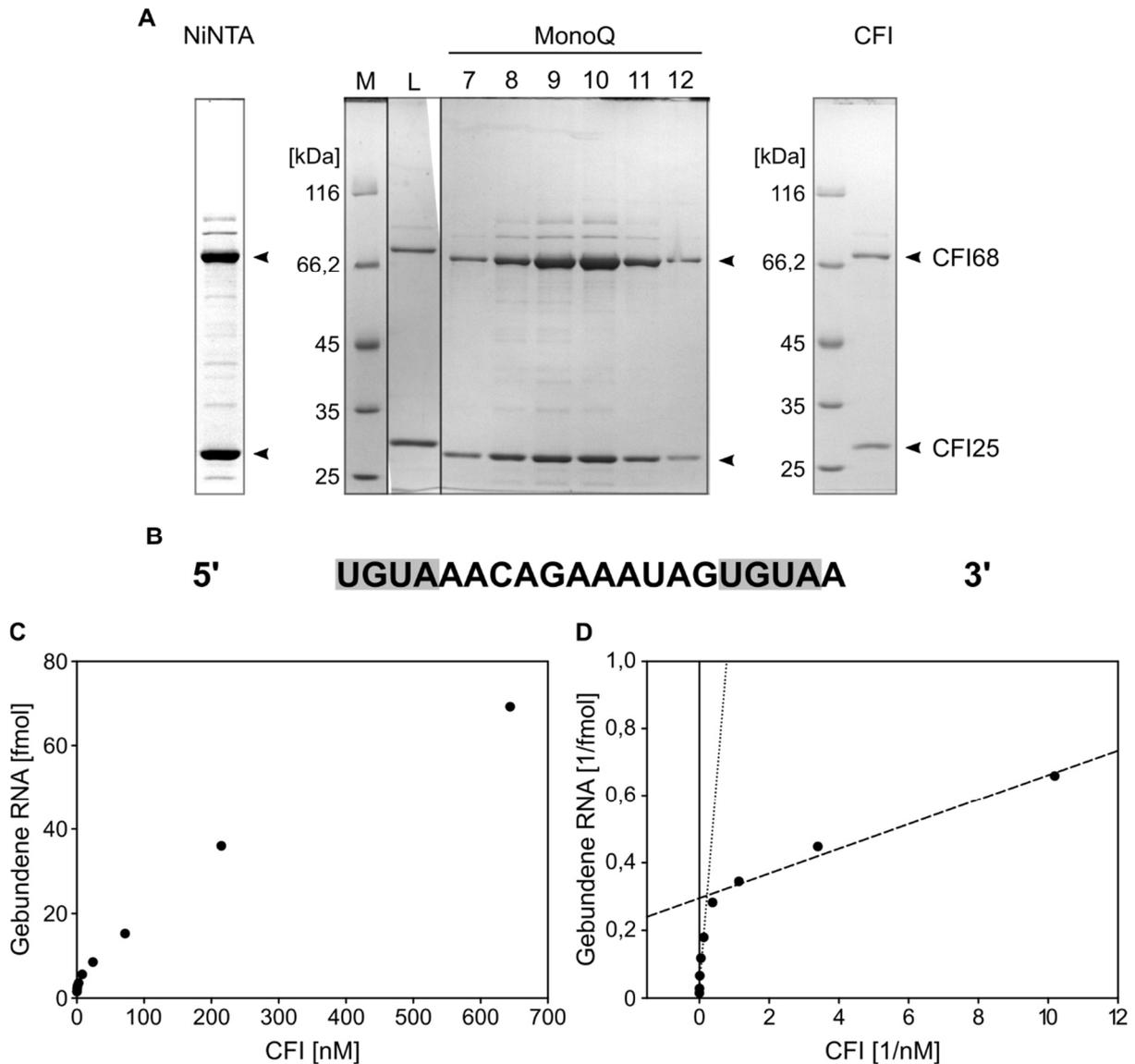


Abbildung 3.16 Reinigung und Aktivitätsanalyse des CFI-Komplexes. A) Darstellung der Reinigung des CFI-Komplexes über die Affinitätschromatographie (NiNTA) und anschließende Anionenaustauschchromatographie (MonoQ) sowie des quantifizierten Proteins (CFI) mittels Coomassie-gefärbten SDS-Polyacrylamidgel. B) Sequenz des in der RNA-Bindung eingesetzten RNA-Oligos. Die CFI-Bindemotive UGUA sind grau hinterlegt. C) Bestimmung der RNA-Bindung durch CFI. Direkte Auftragung der durch Filterbindung und Szintillationsmessung bestimmten RNA-Bindung durch steigende Konzentrationen CFI. D) Doppel-reziproke Auftragung der Messwerte der RNA-Bindung. Linearisiert sind die zwei Phasen der RNA-Bindung.

Die quantifizierten Fraktionen enthielten 0,36 mg CFI-Komplex, was einer Ausbeute von 0,45 mg pro Liter Kulturvolumen entsprach. Zur Aktivitätsanalyse wurde ein synthetisches RNA-Oligo verwendet, welches über zwei CFI-Bindestellen der Sequenz „UGUA“ verfügte (Abbildung 3.16 B). Es konnte eine RNA-Bindung beobachtet werden, wobei jedoch aufgrund der geringen Affinität nur maximal ca. 35 % der RNA gebunden werden konnte (Abbildung 3.16 C). Zur näheren Analyse wurden die Messwerte doppelt-reziprok aufgetragen, welches deutlich die biphasische RNA-Bindung zeigt (Abbildung 3.16 D). Aus der Linearisierung konnten anhand der Abzissenschnittpunkte die apparenten K_{50} -Werte von 0,12 nM für die ersten und 36 nM für die zweite Bindung bestimmt werden. Die Ursache hierfür liegt vermutlich in der heterotetrameren Struktur von CFI und den daraus resultierenden zwei RNA-Bindestellen sowie in der Struktur der RNA, die zwei CFI-Bindestellen aufweist. Hierdurch ist es wahrscheinlich möglich, dass ein CFI-Komplex eine oder zwei RNAs binden kann. Eine genaue Analyse dieser verschiedenen Szenarien fand nicht statt.

3.1.4.2 Chemisches *cross-linking* des CFI-Komplexes

Zur näheren Strukturanalyse des CFI-Komplexes sollte dieser mit zwei verschiedenen *cross-linkern*, DSBU sowie LC-SDA, welcher ein heterobifunktionaler *cross-linker* ist, näher untersucht werden. LC-SDA kann einseitig in wässrigem Milieu über seine NHS-Gruppe mit freien Aminen, auf der anderen Seite über ein Diazirin UV-induziert mit sauren Aminosäuren reagieren [129]. Zur Bestimmung der optimalen *cross-linker*-Konzentrationen wurde gegen konstante 5,8 μ M CFI-Komplex die *cross-linker* von 5 bis 20 mM titriert (Abbildung 3.17 A). Aus Vorversuchen war bekannt, dass ein mindestens 1000-facher Überschuss des *cross-linkers* verwendet werden musste. Es zeigte sich, dass bei 20 mM, welches einem 3500-fachem molarem Überschuss entsprach, sowohl bei DSBU als auch bei LC-SDA noch sehr viel monomeres Protein vorhanden war. Diese Beobachtung deckt sich mit vorherigen Experimenten (Daten nicht gezeigt). Es ist die Bildung von hochmolekularen Produkten oberhalb von 120 kDa sowie eines weiteren Produktes bei 50 kDa zu beobachten. Diese könnten dem nativen tetrameren CFI-Komplex mit einem Molekulargewicht von ca. 140 kDa, sowie einem CFI25-Dimer mit einem Molekulargewicht von 50 kDa entsprechen. Für das präparative *cross-linking* wurde die Konzentration von 20 mM für beide *cross-linker* genutzt. Hierzu wurden in technischen Duplikaten je 307 pmol CFI-Komplex quervernetzt und massenspektrometrisch analysiert. Diese Analyse zeigte insgesamt jedoch nur sehr wenige *cross-links*. Für DSBU konnten 6 *cross-links* identifiziert werden: drei intermolekulare zwischen CFI25 und CFI68, ein intermolekularer zwischen zwei CFI25 Untereinheiten sowie zwei intramolekulare in CFI25. Unter Verwendung von LC-SDA konnten zwei *cross-links* identifiziert werden, wovon je einer intra- und intermolekular war. Die intermolekularen *cross-links* liegen ausnahmslos in der SR-reichen Region von CFI68 (Abbildung 3.17 A; Tabelle A 9).

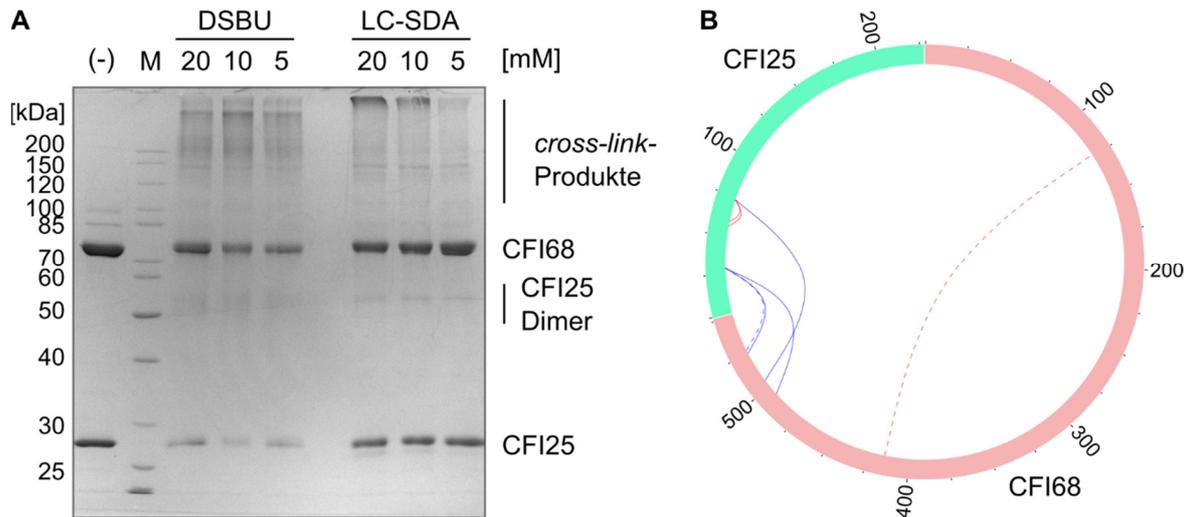


Abbildung 3.17 Chemisches *cross-linking* des CFI-Komplexes. A) Validierung der optimalen *cross-linker* Konzentrationen für CFI. Die beiden *cross-linker* DSBU sowie LC-SDA wurden von 5 bis 20 mM gegen konstante 5,7 μ M CFI-Komplex titriert. Zur Kontrolle wurde der Komplex mit DMSO inkubiert. Die Reaktionsprodukte wurden mittels SDS-Polyacrylamidgel getrennt und mit Coomassie gefärbt. Die monomeren Untereinheiten sowie die Reaktionsprodukte sind gekennzeichnet. B) Kreisdiagramm der identifizierten *cross-links*. Die Untereinheiten sind im Uhrzeigersinn von N- bis C-Terminus dargestellt. Intra- (rot) sowie intermolekulare (blau) Verknüpfungen sind farblich getrennt. *Cross-links*, welche mit DSBU identifiziert wurden sind als durchgezogene Linie dargestellt, welche, die mit LC-SDA identifiziert wurden, mit gestrichelter Linie.

3.2 Computergestützte Modellierung der Polyadenylierungsfaktoren

Um die in 2.1 beschriebenen *cross-link*-Ergebnisse strukturell zu deuten, werden molekulare Modelle der Komplexe benötigt. Für strukturell unbekannte Polyadenylierungsfaktoren wurden im Rahmen dieser Arbeit zuerst Homologiemodelle generiert, die folgend zu Komplexen durch molekulares *Docking* assembliert wurden. Bereits publizierte Strukturen flossen in diese Berechnungen mit ein. Die *cross-link*-Daten dienten hier sowohl als Entfernungsbeschränkung als auch zur Validierung der generierten Strukturen.

Aufgrund der sehr geringen Anzahl der identifizierten *cross-links* für den CFI- und den mPSF-Komplex wurde von einer Strukturmodellierung für diese Komplexe abgesehen.

3.2.1 Generierung der Entfernungsbeschränkungen

3.2.1.1 *Cross-link* Entfernungsbeschränkung

Die durch eine *cross-link*-Analyse identifizierten Aminosäuren müssen in räumlicher Nähe liegen, wobei die Entfernung durch die Länge der Aminosäureseitenkette sowie die Länge des verwendeten *cross-linkers* definiert wird. Darüber hinaus muss die Flexibilität von Proteinen berücksichtigt werden, wobei diese mit zusätzlichen 6 bis 10 Å in die Maximalentfernung berücksichtigt werden kann [131].

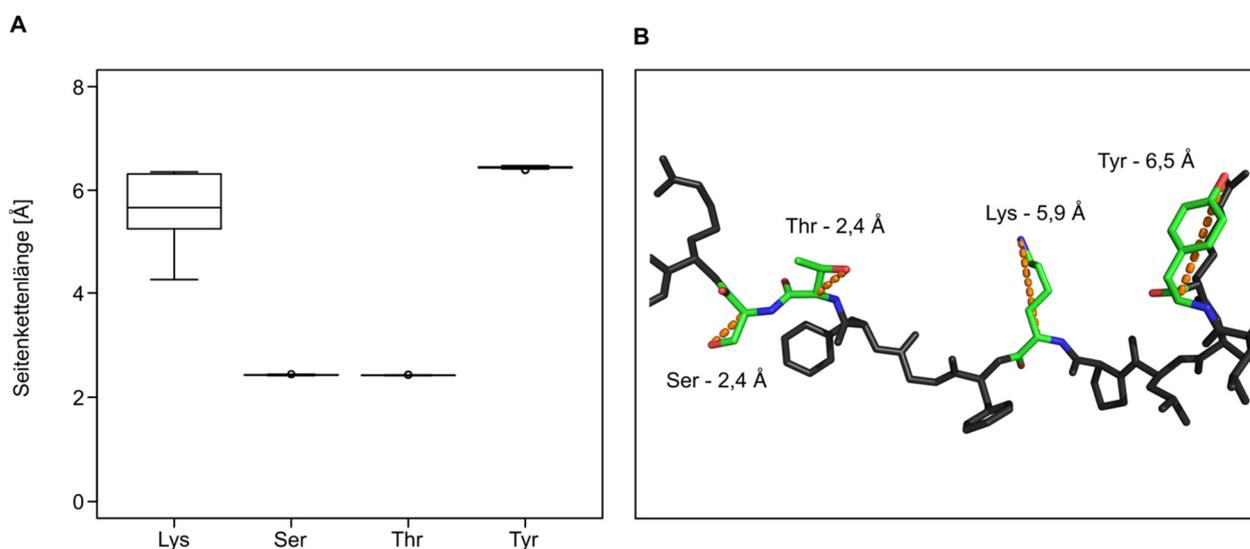


Abbildung 3.18 Empirische Bestimmung der Länge der Aminosäureseitenketten. A) Boxplot der in der 50S ribosomalen Untereinheit von *E. coli* gemessenen Entfernungen zwischen Ca-Atom und reaktiver Seitenkettengruppe. B) Ausschnitt der Struktur der 50S ribosomalen Untereinheit mit hervorgehobenen Serin, Threonin, Lysin und Tyrosin. Die Entfernung zwischen Ca-Atom und reaktiver Seitenkettengruppe ist eingezeichnet.

Der hauptsächlich verwendete *cross-linker* DSBU besitzt zwei NHS-Ester als reaktive Gruppe. Diese reagiert bevorzugt mit freien Aminen, wie in Lysinen oder N-Termini von Proteinen, aber auch mit reduzierter Reaktivität mit Hydroxylgruppen wie in Serinen, Threoninen und Tyrosinen.

Die Länge der Lysinseitenkette wird in der Literatur mit 6,4 Å angegeben [131]. Zur Validierung dieser Länge und zur Definition der Längen von Serin-, Threonin- und Tyrosinseitenketten wurde die 50S Untereinheit des *E. coli* Ribosoms als Strukturmodell zur Grundlage der Analyse genommen (PDBID: 6QUL; 3 Å Auflösung). Insgesamt 326 Lysine, 131 Serine, 169 Threonine und 57 Tyrosine wurden analysiert (Abbildung 3.18 A). Die maximale Länge der Aminosäureseitenkette für Lysin beträgt 6,4 Å, für Serin und Threonin 2,4 Å und für Tyrosin 6,5 Å (Abbildung 3.18 B). Während Serin- und Threoninseitenketten keine Flexibilität aufweisen, ist die Länge von Lysinseitenketten über ca. 2 Å gestreut. Tyrosinseitenketten zeigen aufgrund der Starrheit des Benzolringes keine Varianz in ihrer Länge.

3.2.1.2 Peptid-Entfernungsbeschränkung

Identifizierte *cross-links* beschreiben eine Interaktion sowie eine maximale Entfernung zwischen zwei Aminosäuren. Diese Information kann jedoch nur für weitere Proteinstrukturvorhersagen genutzt werden, wenn auch beide Aminosäuren in den bekannten Strukturen aufgelöst ist. Da dieses, vor allem bei der vorherigen Generierung von Proteinstrukturen durch Homologiemodellierung, jedoch nicht immer gegeben ist, muss, um diese Informationen trotzdem verwenden zu können, eine neue Entfernungsbegrenzung definiert werden. Diese umfasst maximal 9 Aminosäuren, sodass Aminosäuren, welche knapp außerhalb der aufgelösten Struktur liegen, genutzt werden können. Ziel ist es, eine allgemeingültige und sequenzunabhängige Beschreibung zwischen einer euklidischen Entfernung in einer 3D-Struktur und der Entfernung in der Aminosäuresequenz zu erstellen. Diese neue Entfernungsbegrenzung basiert auf einer empirischen Bestimmung der Minimal- und Maximallänge von Peptiden und ist daher sequenzunabhängig.

Hierzu wurden alle Strukturdateien der Proteindatenbank (pdb; Stand 23. März 2019) analysiert, die folgenden Kriterien erfüllten:

- Enthält mindestens ein Protein
- Besitzt mindestens eine Auflösung von 2,8 Å
- Nur eine redundante Struktur bei einer Sequenzidentität von mindestens 30 %
- Die Strukturdaten sind im pdb- und nicht ausschließlich als mmCIF-Format verfügbar

Insgesamt erfüllten 21252 hinterlegte Strukturdaten diese Kriterien. In diesen Dateien wurde die Ca-Ca-Entfernung von Di-, Tri-, Tetra-, Penta-, Hexa-, Hepta-, Octa-, Nona- und Decapeptiden anhand Formel (2) berechnet.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (2)$$

Wobei $P1(x_1|y_1|z_1)$ den räumlichen Koordinaten des ersten Ca-Atoms und $P2(x_2|y_2|z_2)$ den räumlichen Koordinaten des zweiten Ca-Atoms entsprachen.

Analysiert wurden nur die Entfernungen zwischen proteinogenen Aminosäuren. Nicht in die Messung eingeflossen sind artifizielle Aminosäuren wie Selenomethionin ("HETATM"-Index in pdb-Dateien) sowie Aminosäuren, die keine eindeutig definierte Position besaßen. Die gemessenen Entfernungen jeder pdb-Datei wurden in einer separaten Datei gespeichert. Um Fehler, die in den pdb-Dateien vorhanden sein könnten, zu identifizieren wurden anschließend alle Entfernungsdateien auf folgendes Kriterium analysiert:

$$2 \text{ \AA} < \text{Ca-Ca Entfernung im Dipeptid} < 6 \text{ \AA}$$

Die durchschnittliche Ca-Ca Entfernung eines Dipeptides beträgt 3,8 Å, sodass in diesem Kriterium auch stark gestauchte oder gestreckte Bereiche inkludiert sind. Entfernungen außerhalb dieses Bereiches können durch eine fehlerhafte Analyse der pdb-Datei, aber auch durch falsch annotierte Sequenzlücken oder Kettenidentifikationen in den pdb-Dateien entstanden sein. Sollte ein Fehler in einer pdb-Datei gefunden worden sein, wurde diese von der weiteren Analyse ausgeschlossen. Insgesamt wurden so 2052 Entfernungsdateien als fehlerhaft markiert, was ungefähr 10 % aller analysierter Dateien entsprach.

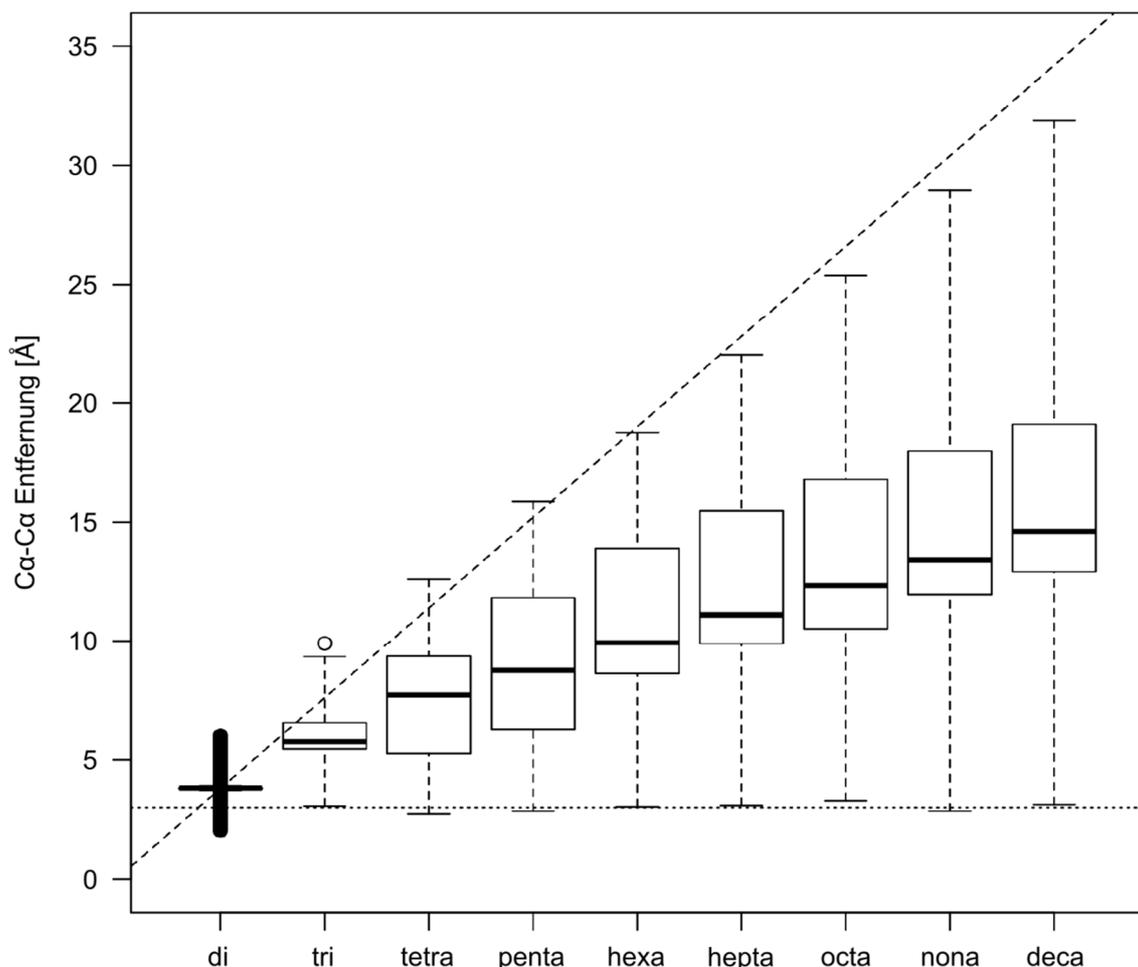


Abbildung 3.19 Boxplot der gemessenen Entfernungen von Di- bis Decapeptid. Die Antennen der Boxen entsprechen dem 3-fachen Interquartilabstand. Datenpunkte ober bzw.

unterhalb dieser Grenzen sind eingezeichnet. Der Querstrich innerhalb der Boxen entspricht dem Median der Datenpunkte. Die horizontale, gepunktete Linie entspricht einer C α -C α -Entfernung von 3 Å, die gestrichelte Linie zeigt die theoretische Maximalentfernung von 3,8 Å pro Peptidbindung.

Insgesamt wurden knapp 70 Millionen Entfernungen analysiert. Diese sind in Abbildung 3.19 als Boxplot dargestellt. Zu beobachten ist, dass die minimale Entfernung zwischen zwei C α -Atomen immer ca. 3 Å beträgt, unabhängig der Länge des Peptides. Außerdem zeigt sich, dass die maximale Länge mit dem Vielfachen von 3,8 Å, der durchschnittlichen Entfernung zwischen zwei C α -Atomen entspricht, korreliert. Eine detaillierte Übersicht der Statistik des Boxplots ist im Anhang in Tabelle A.18 dargestellt.

Da die gemessenen Entfernungen für Di- bis Pentapeptide Ausreißer sowohl unterhalb der 3 Å als auch oberhalb des Vielfachen von 3,8 Å vorwiesen, wurden diese nachfolgenden Kriterien näher klassifiziert:

$$\text{untere Ausreißer} > 3 \text{ \AA} \leq \mu > 3,8 * (n - 1) \text{ \AA} \leq \text{obere Ausreißer}$$

wobei n der Anzahl der beteiligten Aminosäuren und μ dem Erwartungswert der Entfernung entspricht.

Die Detailanalyse zeigte, dass sich mindestens 99% der Entfernungen in dem Erwartungswert aufhalten (Tabelle 6). Untere Ausreißer sind auf *cis*-Peptidbindungen zurückzuführen, die eine Entfernung von 2,94 Å besitzen [167].

Tabelle 6 Detailanalyse der Verteilung der Entfernungen von Di- bis Pentapeptiden.

Entfernungen unterhalb von 3 Å sind als Untere Ausreißer, Entfernungen oberhalb von n-fachen von 3,8 Å, wobei n der Anzahl der im Peptid enthaltenen C α -C α Bindungen entspricht, sind als obere Ausreißer klassifiziert.

	Untere Ausreißer	Erwartungswert	Obere Ausreißer
Dipeptid	19490 (0,2 %)	9845677 (99,78 %)	2303 (0,02 %)
Tripeptid	0	9697041 (100 %)	199
Tetrapeptid	4	9536963 (100 %)	28
Pentapeptid	1	9382830 (100 %)	4

Basierend auf der statistischen Analyse der Entfernungen konnten die Entfernungsbeschränkungen für Peptide, welche aus einer minimalen sowie maximalen Entfernung, definiert werden (Tabelle 7).

Tabelle 7 Übersicht der neu definierten Entfernungsbeschränkungen. Die minimalen und maximalen Entfernungen sind für die angegebenen Peptide darstellt.

	Minimale Entfernung [Å]	Maximale Entfernung [Å]
Dipeptid	3	4
Tripeptid	3	7,6
Tetrapeptid	3	11,4
Pentapeptid	3	15,2
Hexapeptid	3	18,8
Heptapeptid	3	22
Octapeptid	3	25,4
Nonapeptid	3	29
Decapeptid	3	31,9

3.2.2 Modellierung der Subkomplexe der Polyadenylierung

3.2.2.1 Modellierung des CFII-Komplexes

Zur Modellierung des CFII-Komplexes wurden zuerst Strukturhomologe identifiziert und die monomeren Modelle generiert. Anschließend erfolgte das Zusammenführen dieser Proteine zu einem Komplex.

3.2.2.1.1 Generierung der monomeren Subkomplexe

Clp1

Clp1 ist ein hochkonserviertes Protein, welches ein Sequenzalignment mit den Homologen aus Bäckerhefe (*yClp1*) und *C. elegans* (*ceClp1*), von denen beide Kristallstrukturen publiziert sind, deutlich zeigt (Anhang Abbildung A. 2, Tabelle 8). Es ist daher davon auszugehen, dass die humane Struktur den publizierten sehr ähnelt und ohne weiteres modelliert werden kann.

Tabelle 8 Zusammenfassung des paarweisen Sequenzalignments von *Clp1* zu den homologen in Hefe (*yClp1*) und *C. elegans* (*ceClp1*). Das Alignment wurde mit dem Onlineservice EMBOSS und dem Needleman-Wunsch-Algorithmus durchgeführt. Die unterschiedlichen Längen entstehen durch die Summe der ausgerichteten Aminosäuren, Unterbrechungen und Überhänge an den Termini der Proteine.

	<i>yClp1</i>	<i>ceClp1</i>
Identische Aminosäuren	111 / 472 (23,5 %)	203 / 435 (46,7 %)
Ähnliche Aminosäuren	208 / 472 (44,1 %)	286 / 435 (65,6 %)
Unterbrechungen	74 / 472 (15,7 %)	17 / 435 (3,9 %)

Um ein möglichst genaues Ergebnis zu erzielen, wurden die Kristallstrukturen beider Homologe für die Generierung des Homologiemodelles verwendet. In beiden Kristallstrukturen konnte das gebundene Magnesiumion sowie ein Nukleotid, ATP in der Struktur der Hefe, ADP in der von *C. elegans* aufgelöst werden. Da diese einen Einfluss auf die Orientierung von Aminosäureseitenketten durch nicht-kovalente Wechselwirkungen ausüben, wurden diese in die Modellierung inkludiert. Das Modell mit der geringsten Energie besitzt einen RMSD-Wert zu *yClp1* von 1,7 Å und zu *ceClp1* von 0,7 Å. Die überlagerten Strukturen sind in Abbildung 3.20 zu sehen. Zur Validierung der Struktur wurden zudem die intramolekularen *cross-links* genutzt, welche alle in den erlaubten Grenzen lagen (Abbildung 3.20).

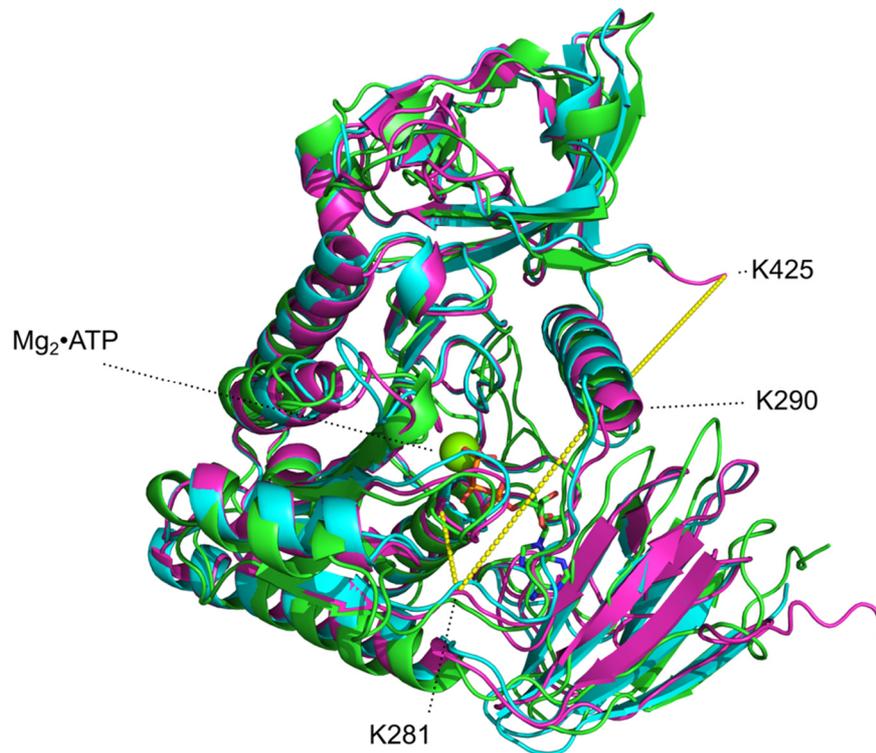


Abbildung 3.20 Homologiemodell des humanen Clp1. Dargestellt sind das generierte Homologiemodell (grün) sowie die beiden Vorlagestrukturen aus Hefe (Türkis; PDBID: 2NPI) sowie *C. elegans* (Magenta; PDBID: 3OHX). Aus der publizierten Struktur überlagert wurden das gebundene Magnesiumion (grüne Sphäre) sowie das ATP-Molekül (orange Stäbchendarstellung). Die experimentellen *cross-link*-Daten sind als gelbe Verbindungen zwischen den α -Atomen der beteiligten Aminosäuren dargestellt.

Pcf11

Von *Pcf11* ist nur einer der beiden Zinkfingerdomänen aus Hefe bekannt, weitere Strukturen liegen bisher nicht vor. Daher mussten für das humane Protein konservierte Homologiestrukturen identifiziert werden. Da für eine Strukturvalidierung weitere experimentelle Daten notwendig sind, wurde sich auf den Sequenzbereich, welcher über intermolekulare *cross-links* zu Clp1 verfügte, beschränkt. Dieses war zum einen der Sequenzbereich von Aminosäure 500 bis 770 (*Pcf11*₅₀₀₋₇₇₀) sowie der Bereich der Zinkfingerdomäne, welche auf den Bereich von Aminosäure 1200 bis zum C-Terminus ausgedehnt wurde (*Pcf11*₁₂₀₀₋₁₅₅₅). Eine Sequenzanalyse mittels HHPred und BLASTp konnte in diesen Sequenzbereichen sieben, nicht redundante homologe Strukturen identifizieren (Tabelle 9) [139], [158].

Davon liegen *Pcf11*₅₇₉₋₆₃₁, *Pcf11*₆₄₆₋₆₅₃ und *Pcf11*₆₅₅₋₆₉₅ im N-terminalen Bereich und *Pcf11*₁₂₆₅₋₁₂₈₁, *Pcf11*₁₃₄₁₋₁₃₆₈, *Pcf11*₁₃₇₇₋₁₄₀₁ und *Pcf11*₁₄₁₉₋₁₄₈₇ in der Zinkfinger-Interaktionsdomäne. Ein Sequenzalignment dieser Bereiche ist im Anhang in Abbildung A.4 dargestellt. Weitere homologe Strukturen konnten nicht identifiziert werden.

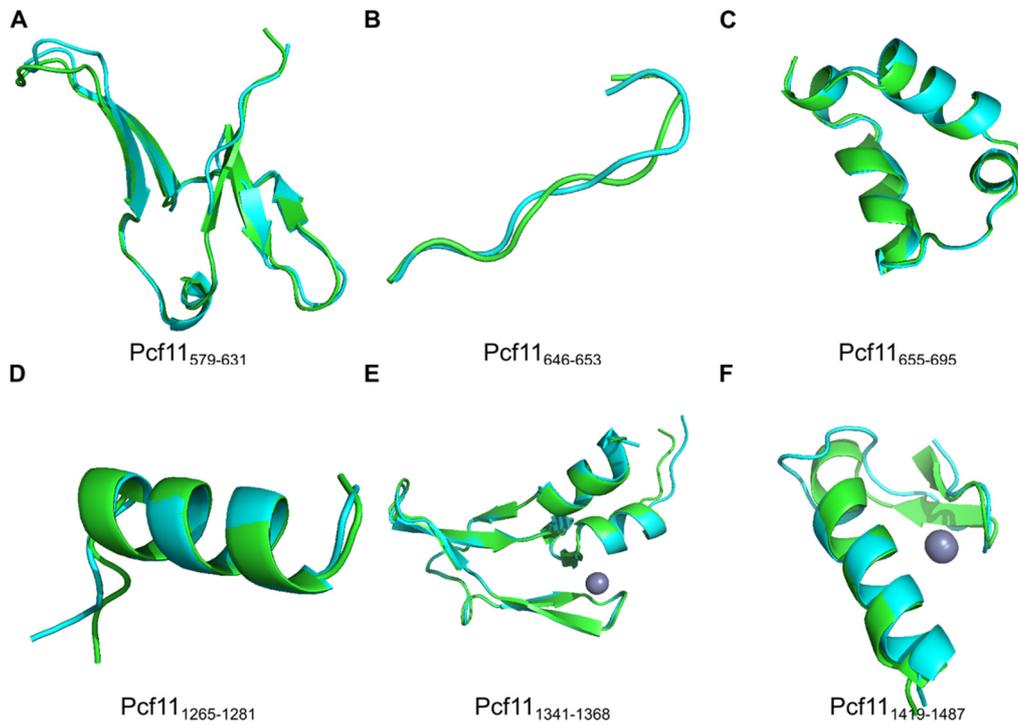


Abbildung 3.21 Übersicht der generierten Homologiemodelle von Pcf11. Das generierte Modell ist immer in Grün, die Vorlagestruktur in Türkis dargestellt. Gebundene Zinkionen sind als graue Sphäre dargestellt. A) Modell von Pcf11₅₇₉₋₆₃₁, welches eine Ekto-Domäne bildet (PDBID: 1XAU). B) Modell von Pcf11₆₄₆₋₆₅₃, einer *coil*-Struktur (PDBID: 3OP8). C) Modell von Pcf11₆₅₅₋₆₉₅, einer EF-Hand-Domäne (PDBID: 2FCD). D) Pcf11₁₂₆₅₋₁₂₈₁, einer einfachen α -helikalen Struktur (PDBID: 5V8W). E) Modell der ersten Zinkfingerdomäne Pcf11₁₃₄₁₋₁₃₆₈ (PDBID: 3WUP). F) Modell der zweiten Zinkfingerdomäne Pcf11₁₄₁₉₋₁₄₈₇ (PDBID: 2NAX).

Es wurden für alle Strukturbereiche wie beschrieben mittels MODELLER Homologiemodelle generiert, wobei aus 20 finalen Modellen das mit der geringsten Energie ausgewählt wurde. Einen Sonderfall stellte Pcf11₁₃₇₇₋₁₄₀₁ dar, da es sich hier um das kokristallisierte Peptid in der Struktur von Clp1 handelt. Dieses wird später ausführlicher beschrieben. Die für die anderen 6 Sequenzen generierten Strukturmodelle sind in Abbildung 3.21 dargestellt. Die Struktur von Pcf11₅₇₉₋₆₃₁ entspricht einer Ekto-Domäne aus BTLA (*B- and T-lymphocyte attenuator*), welches einer extrazellulären Domäne entspricht. Pcf11₆₅₅₋₆₉₅ entspricht einer EF-Hand Domäne aus Myosin, welche üblicherweise Calcium binden. Die beiden Strukturen von Pcf11₁₃₄₁₋₁₃₆₈ und Pcf11₁₄₁₉₋₁₄₈₇ entsprechen den beiden Zinkfingerdomänen. Hierbei wurde in der Modellierung der Strukturen das gebundene Zinkion inkludiert. Die Bereiche Pcf11₆₄₆₋₆₅₃ und Pcf11₁₂₆₅₋₁₂₈₁ bildeten aufgrund ihrer geringen Größe keine distinkte Domäne, sondern stellten nur Strukturelemente dar. Die verwendeten Vorlagestrukturen sind in Tabelle 9 dargestellt.

Tabelle 9 Übersicht der Sequenzbereiche in Pcf11 und die für die Modellierung verwendete Vorlagestruktur.

Sequenzbereich	PDBID	Quelle
Pcf11 ₅₇₉₋₆₃₁	1XAU	[168]
Pcf11 ₆₄₆₋₆₅₃	3OP8	[169]
Pcf11 ₆₅₅₋₆₉₅	2FCD	[170]
Pcf11 ₁₂₆₅₋₁₂₈₁	5V8W	[171]
Pcf11 ₁₃₄₁₋₁₃₆₈	3WUP	[172]
Pcf11 ₁₃₇₇₋₁₄₀₁	4OI4	[62]
Pcf11 ₁₄₁₉₋₁₄₈₇	2NAX	[67]

Für die Generierung der Struktur des Interaktionsbereiches Pcf11₁₃₇₇₋₁₄₀₁ wurden zwei parallele Ansätze verfolgt. Zum einen wurde das Peptid in seiner freien Form, zum anderen in seiner gebundenen Form im Komplex mit Clp1 verwendet. Für die gebundene Form wurde das Homologiemodell von Clp1, welche in 3.2.2.1.1-*Clp1* generiert wurde, mit dem kokristallisiertem Peptid aus Hefe überlagert. Diese hybride Struktur wurde dann im Folgenden für die Modellierung genutzt, wobei neben dem Peptid auch Clp1 remodelliert wurde.

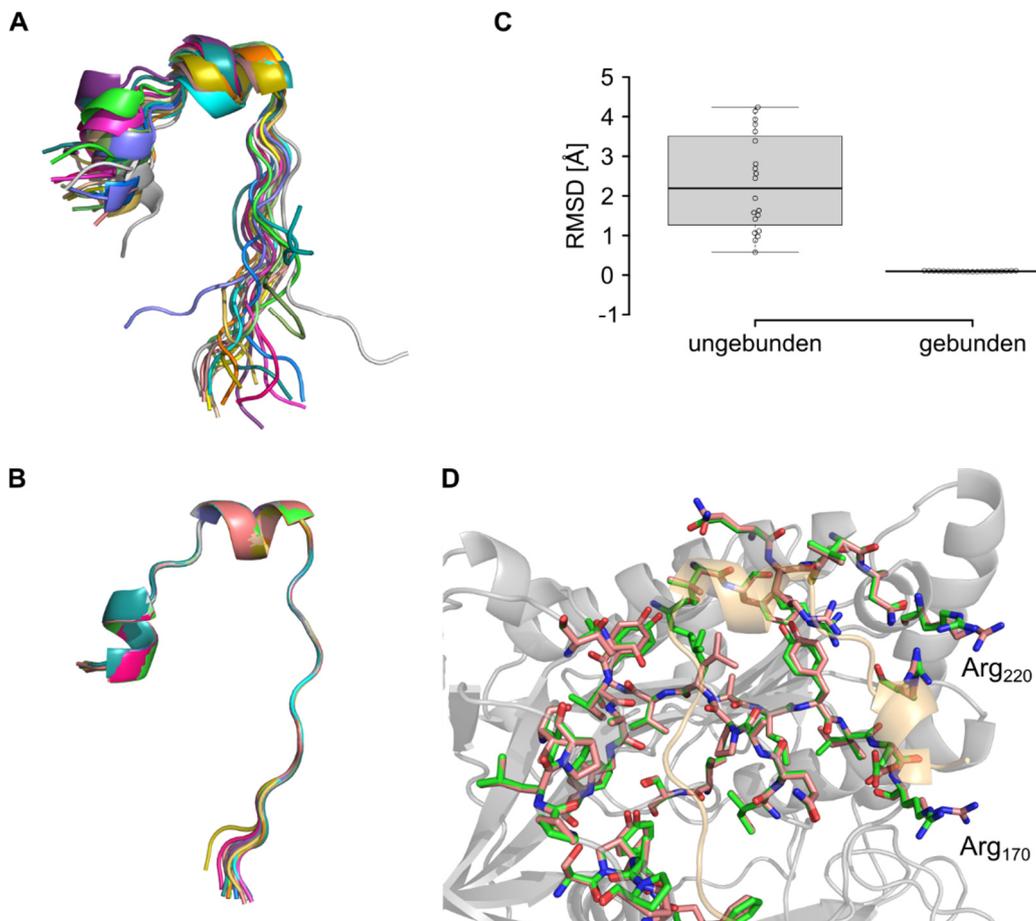


Abbildung 3.22 Zusammenfassung der Modellierung des Interaktionspeptides Pcf11₁₃₇₇₋₁₄₀₁.

A) Modellierung von Pcf11₁₃₇₇₋₁₄₀₁ in seiner ungebundenen Form. Überlagert dargestellt sind alle 20 generierten Strukturmodelle. B) Modellierung von Pcf11₁₃₇₇₋₁₄₀₁ in seiner von Clp1-gebundenen

Form. Überlagert dargestellt sind alle 20 generierten Strukturmodelle. C) Boxplot der strukturellen Abweichung (RMSD in Å) der generierten Modelle von Pcf11₁₃₇₇₋₁₄₀₁ gegenüber der Vorlagestruktur. D) Strukturvergleich der an der Bindung des Interaktionspeptides Pcf11₁₃₇₇₋₁₄₀₁ (Hellgelber Cartoon) beteiligten Aminosäuren in Clp1 (Grauer Cartoon). Die Aminosäuren in Stabdarstellung der humane Vorlagestruktur sind in Rot, die der remodellierten Struktur in Grün dargestellt. Hervorgehoben sind die Aminosäuren Arginin-170 und Arginin-220, welche eine deutliche Konformationsänderung zeigen.

Im direkten Vergleich der Struktur des Peptides zeigte sich, dass im Gegensatz zur gebundenen Form das freie Peptid sehr viele Konformationen annehmen kann (Abbildung 3.22 A/B). Dieses zeigt sich auch in der Berechnung der RMSD-Werte dieser Peptide im Vergleich zur Ausgangsstruktur (Abbildung 3.22 C). Neben der deutlich genaueren Struktur kommt es in der Modellierung in der gebundenen Form zu einer Restrukturierung in der Interaktionsfläche. So kann für die Aminosäuren Arginin-170 und Arginin-220 eine deutliche Konformationsänderung, angepasst an die humane Sequenz des Interaktionspeptides, beobachtet werden (Abbildung 3.22 D).

3.2.2.1.2 Definition der Entfernungsbeschränkungen

Zur Grundlage der Entfernungsbeschränkungen wurden vor allem die experimentellen *cross-link*-Daten verwendet. Hierzu mussten die beiden an der Vernetzung beteiligten Aminosäuren innerhalb der Sequenz der Strukturen liegen. Unter Berücksichtigung der neu definierten Peptidbeschränkungen, konnten alle *cross-links* verwendet werden, die bis zu 10 Aminosäuren außerhalb der aufgelösten Strukturmodelle lagen. Von den 152 einzigartigen *cross-links* lagen 48 in diesem Sequenzbereich, wovon 15 direkt in der Struktur aufgelöst waren und 33 mindestens einseitig strukturell nicht aufgelöst waren. *Cross-links* zu Aminosäure 638 in Pcf11, welches je 8 Aminosäuren zwischen den Strukturen Pcf11₅₇₉₋₆₃₁ und Pcf11₆₄₆₋₆₅₃ liegt, führten zu zwei verschiedenen Beschränkungen. Insgesamt wurden so 44 weitere Entfernungsbeschränkungen durch die Verwendung der Peptidbeschränkung gebildet. Eine zusätzliche Entfernungsbeschränkung basiert komplett auf der Peptidbeschränkung. Das Modell des ersten Zinkfingers endet auf Aminosäure 1371 und das von Clp1 gebundene Peptid von Pcf11 beginnt mit Aminosäure 1377. Der Abstand zwischen diesen Aminosäuren kann auch durch die Peptidbeschränkung beschrieben werden. Die für das Docking verwendete Entfernungsbeschränkungen sind im Anhang in Tabelle A 10 tabelliert.

3.2.2.1.3 Dimeres Docking

Clp1 bildete aufgrund seiner komplett aufgelösten Struktur und der kokristallisierten Interaktion mit Pcf11₁₃₇₇₋₁₄₀₁ den Kern der computergestützten Komplexbildung. Im Weiteren wird dieser Minimalkomplex als Clp1* bezeichnet. Als erster Schritt wurde diese mit den vier strukturierten Domänen binär unter Verwendung der definierten Entfernungsbeschränkungen mittels

HADDOCK zusammengefügt (*gedockt*). Hierbei werden 5000 initiale Modelle generiert, die über zwei Verfeinerungsschritte (*refinement*) in 200 finalen Modellen resultierten. Diese wurden anhand ihrer Konformation, aber auch weiterer Kriterien wie Interaktionsfläche und Bindungsenergien klassifiziert und zu einem Cluster zusammengefügt. Ein Cluster mit einer hohen Zahl an Modellen spricht hierfür für ein vertrauenswürdiges Ergebnis. Die genaue Beschreibung des *Dockings* ist in 2.2.6.4 beschrieben.

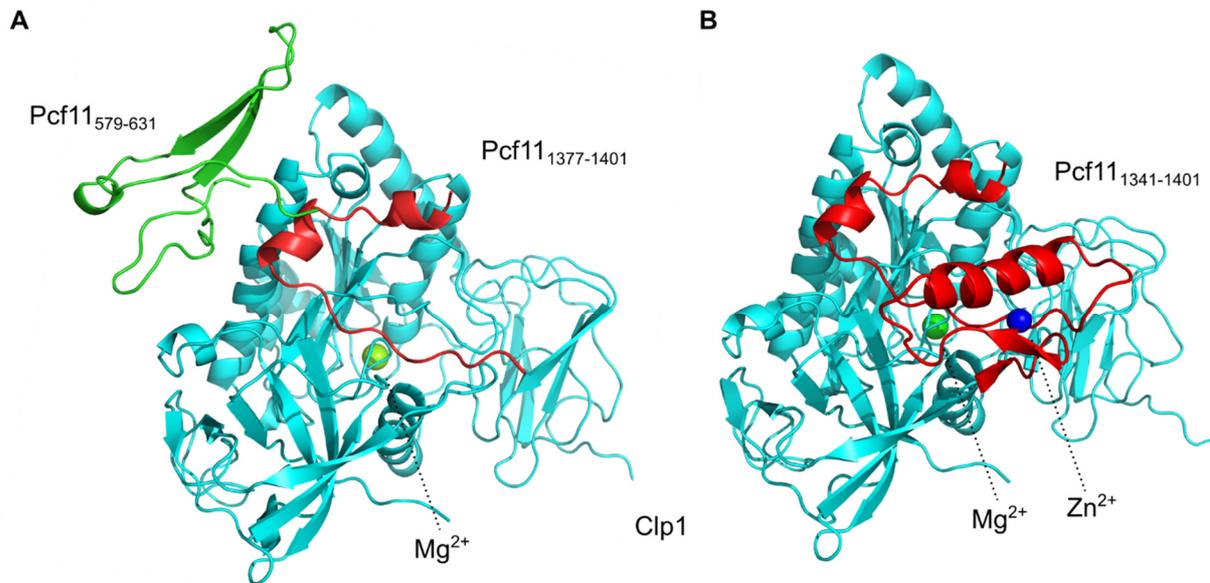


Abbildung 3.23 Strukturmodelle des dimeren Dockings von Clp1* mit Pcf11-Substrukturen.

Clp1 ist in Türkis, das Interaktionspeptid aus Pcf11 in Rot dargestellt und das gebundene Magnesiumion als grüne Sphäre. A) Struktur des Clp1*-Pcf11₅₇₉₋₆₃₁-Dimers. Das hinzugefügte Pcf11-Fragment ist in grün dargestellt. B) Struktur des Clp1-Pcf11₁₃₄₁₋₁₄₀₁-Dimers. Die Verknüpfung zwischen dem Zinkfinger und der Interaktionsregion wurde *de novo* modelliert. Das gebundene Zink-Ion ist als blaue Sphäre dargestellt.

Von den vier analysierten Paaren zeigte sowohl Clp1*-Pcf11₅₇₉₋₆₃₁ mit 195 von 200 Modellen in einem Cluster als auch Clp1*-Pcf11₁₃₄₁₋₁₃₆₈ mit 134 von 200 Modellen in einem Cluster ein sehr gutes Ergebnis. Die beiden anderen Paare generierten keine eindeutige Struktur.

Da zwischen dem ersten Zinkfinger und der kokristallisierten Interaktionsregion nur 9 Aminosäuren lagen, wurde dieser Bereich *de novo* modelliert (Clp1-Pcf11₁₃₄₁₋₁₄₀₁). Die generierten Modelle zeigten keine große Varianz in der Konformation, sodass das beste Modell ausgewählt wurde. Zur weiteren Assemblierung des Komplexes wurden beide generierten Modelle separat mit den weiteren Untereinheiten gedockt, wobei jedoch keine der möglichen Kombinationen einen Komplex von ausreichender Qualität bildeten.

3.2.2.1.4 Multimeres Docking

Da durch das dimere Docking keine weiteren Ergebnisse erzielt werden konnten, wurden ein trimerer Ansatz verfolgt. Hier zeigte einzig die Kombination von Clp1*-Pcf11₅₇₉₋₆₃₁ mit Pcf11₆₅₅₋₆₉₅ und Pcf11₆₄₆₋₆₅₃ die Bildung eines Komplexes (127 Proteine in einem Cluster).

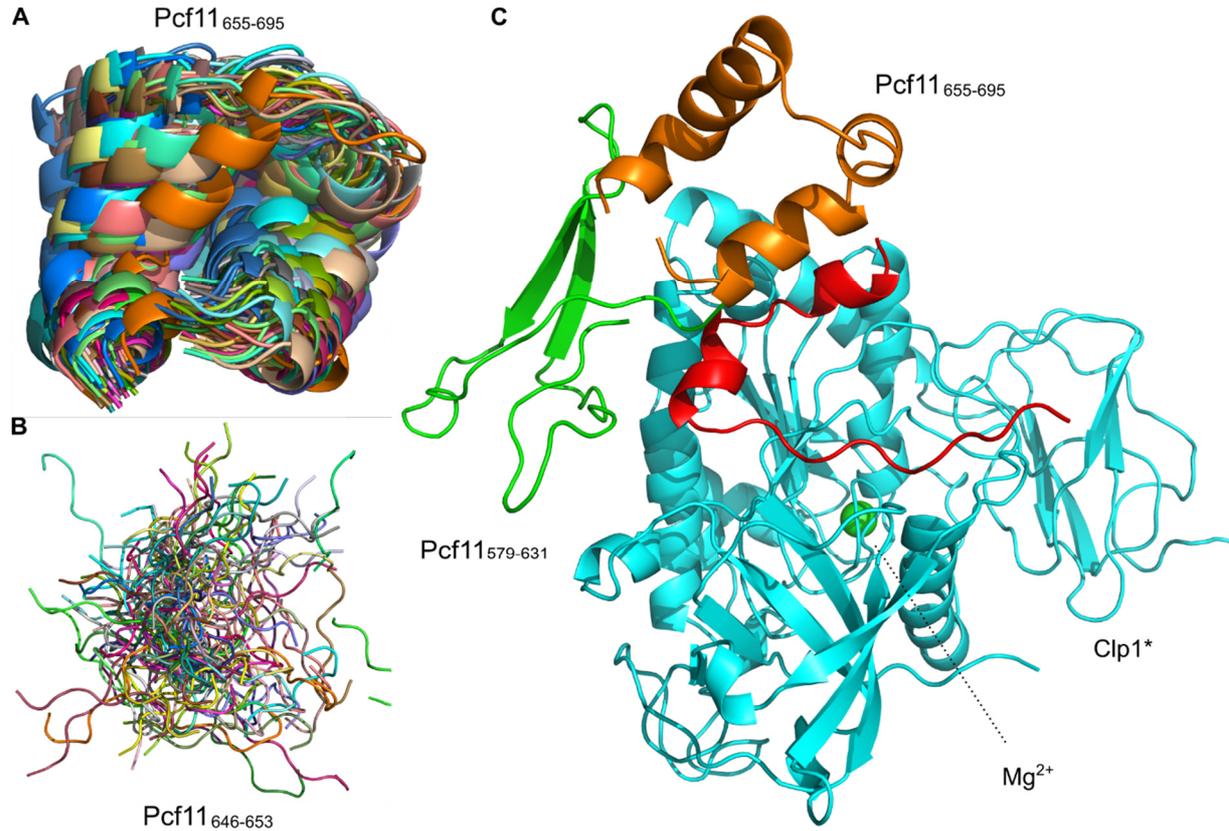


Abbildung 3.24 Zusammenfassung des multimeren dockings mit Clp1* und drei Pcf11-Strukturen. A) Überlagerung aller 127 Strukturen von Pcf11₆₅₅₋₆₉₅, ausgerichtet an Clp1* (nicht dargestellt). B) Überlagerung aller 127 Strukturen von Pcf11₆₄₆₋₆₅₃, ebenfalls an Clp1* ausgerichtet. C) Strukturmodell des Komplexes Clp1*(Türkis/Rot)-Pcf11₅₇₉₋₆₃₁(Grün)-Pcf11₆₅₅₋₆₉₅(Orange), das Magnesium-Ion ist als grüne Sphäre dargestellt.

Eine Detailanalyse dieses Clusters zeigte jedoch, dass Pcf11₆₅₅₋₆₉₅ sehr gut lokalisiert vorlag (Abbildung 3.24 A), das kurze Peptide Pcf11₆₄₆₋₆₅₃ jedoch nicht eindeutig positioniert werden konnte (Abbildung 3.24 B). Dieses wurde daher manuell aus der Struktur entfernt und so das Modell des Clp1*-Pcf11₅₇₉₋₆₃₁-Pcf11₆₅₅₋₆₉₅-Komplexes generiert (Abbildung 3.24 C)

Da alle weiteren Ansätze, die übrigen Strukturen in den Komplex zu integrieren, fehlschlugen, wurden die beiden Komplexe Clp1*-Pcf11₅₇₉₋₆₃₁-Pcf11₆₅₅₋₆₉₅ und Clp1-Pcf11₁₃₄₁₋₁₄₀₁ zusammengefügt (Abbildung 3.25). Zur Validierung dieses hochaufgelösten Modelles wurden die experimentellen *cross-link*-Daten genutzt, um dieses zu validieren. Es zeigte sich, dass 26 der 27 *cross-links* im erlaubten Rahmen liegen.

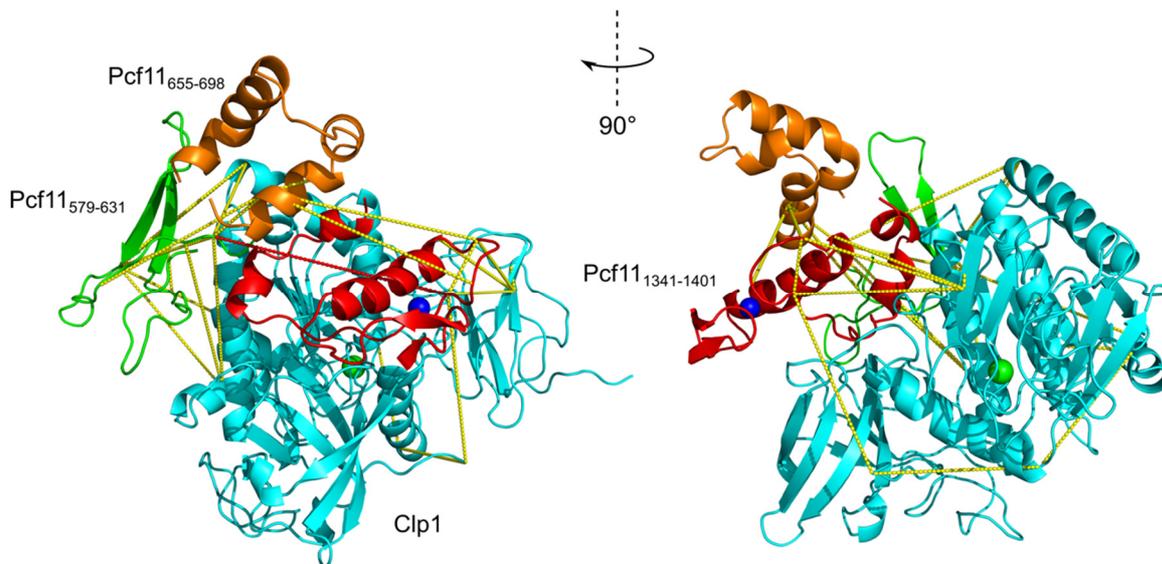


Abbildung 3.25 Finales Strukturmodell von CFII. Dargestellt ist der Clp1-Pcf11₅₇₉₋₆₃₁-Pcf11₆₅₅₋₆₉₅-Pcf11₁₃₄₁₋₁₄₀₁-Komplex in zwei Blickwinkeln. Die Untereinheiten sind farblich getrennt und beschriftet. Die gebundenen Metall-Ionen sind als Sphären dargestellt (Magnesium: Grün; Zink: Blau). Zur Visualisierung der *cross-links* wurden die α -Atome der beteiligten Aminosäuren linear verbunden. *Cross-links*, welche innerhalb des erlaubten Rahmens von 34 Å lagen sind Gelb, solche die oberhalb dieser Grenze liegen Rot dargestellt.

Zur weiteren Analyse der Struktur von CFII wurde für den zweiten Zinkfinger (Pcf11₁₄₁₉₋₁₄₈₇) der mögliche Konformationsraum mittels DisVis analysiert [136]. Hierbei werden mithilfe der experimentellen *cross-link*-Daten die beiden Komplexe zusammengefügt und der Konformationsraum des Zinkfingers berechnet. Zur Analyse standen insgesamt 4 *cross-links* zur Verfügung.

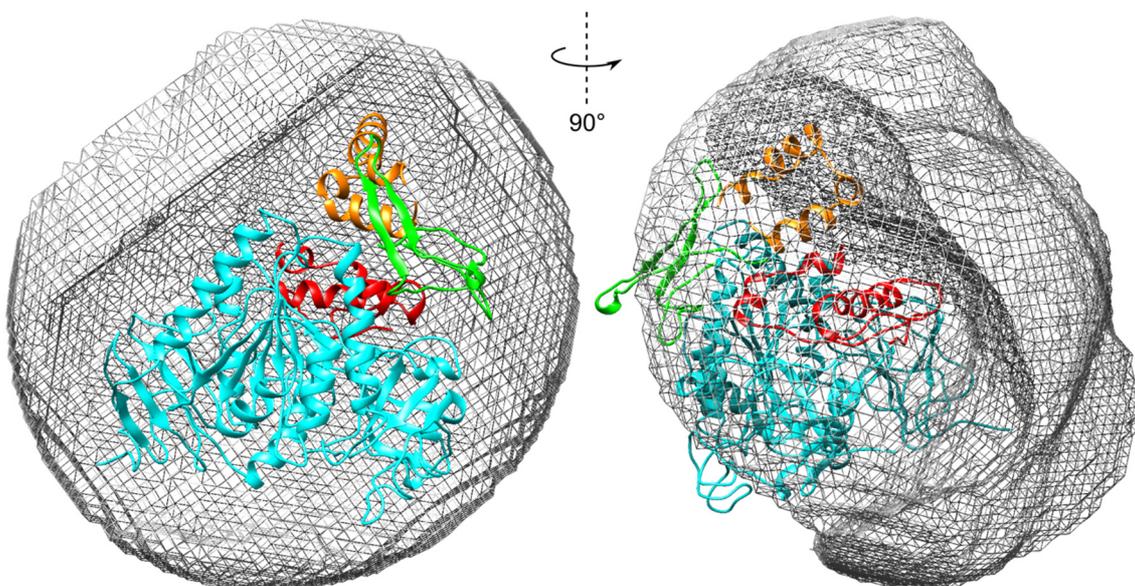


Abbildung 3.26 Analyse des Konformationsraumes von Pcf11₁₄₁₉₋₁₄₈₇ in Relation zum vorher gebildeten CFII-Komplex. Dargestellt ist das Ergebnis der Analyse des Konformationsraumes

des Zinkfingers Pcf11¹⁴¹⁹⁻¹⁴⁸⁷ in Form der Gitteroberfläche. Das Strukturmodell von CFII ist als Cartoon dargestellt, wobei die Submodelle farblich differenziert sind (siehe Abbildung 3.25).

Es zeigte sich, dass durch die Verwendung dieser die Anzahl der möglichen Komplexe von rund 37 Millionen (ohne Entfernungsbeschränkungen) auf 3,6 Millionen (mit 4 Entfernungsbeschränkungen) reduziert. Eine genauere Lokalisation ist aufgrund der limitierenden Daten nicht möglich (Abbildung 3.26).

3.2.2.2 Modellierung des CstF-Komplexes

3.2.2.2.1 Modellierung der separierten Untereinheiten

CstF50

Von CstF50 sind zwei Strukturen publiziert, die nahezu die gesamte Sequenz abdecken. Die N-terminale Homodimerisierungsdomäne konnte von dem homologen Protein aus *Drosophila melanogaster* aufgeklärt werden und deckt die humane Sequenz von Aminosäure 2 bis 61 ab (PDBID: 2XZ2) [91]. Zusätzlich konnte die C-terminale WD40-Domäne des humanen Proteins von Aminosäure 90 bis 427 kristallisiert werden, wobei ein nicht aufgelöster Sequenzbereich von 143 bis 161 vorliegt (PDBID 6B3X) [73]. Zusätzlich ist in dieser Struktur ein Peptid von CstF77 (CstF77₅₈₁₋₅₉₃) vorhanden. Für die Sequenzbereiche, die von den beiden Strukturen nicht abgedeckt werden wurden, konnten keine Strukturhomologe identifiziert werden.

Zur Generierung der humanen Struktur der Homodimerisierungsdomäne (CstF50₂₋₆₁) wurde die dimere Struktur als Vorlage genutzt. Dieses hat den Vorteil, dass die an der Dimerisierung beteiligten Aminosäuren korrekt orientiert vorliegen (vgl. 3.2.2.1.1-*Pcf11*). Aufgrund der hohen Sequenzidentität konnte ein gutes Homologiemodell generiert werden (Abbildung 3.27). Da die publizierte Struktur der WD40-Domäne von CstF50 die humane Struktur darstellt, wurde diese unverändert für die folgenden Experimente verwendet.

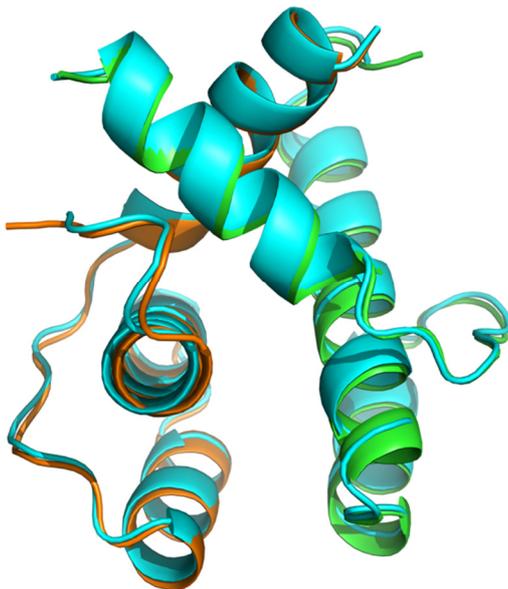


Abbildung 3.27 Modellierung der Dimerisierungsdomäne von CstF50.

Darstellung des generierten Strukturmodelles. Die Vorlagestruktur (Türkis) wurde mit dem generierten humanen Modell überlagert. Die beiden Monomere sind Grün bzw. Orange dargestellt.

CstF64

Von CstF64 sind drei Strukturen publiziert. Der N-Terminus, welcher aus dem RNA-Erkennungsmotiv (RRM) und der Scharnierdomäne besteht, konnte in zwei sich überlappenden Strukturen aufgeklärt werden. Hierbei handelte es sich um die humane Sequenz des RRM (CstF64₈₋₁₁₁; PDBID: 1P1T) sowie die Scharnierdomäne des Hefehomologes Rna15 (entspricht CstF64₁₀₁₋₁₈₉; PDBID: 2L9B), welche zusätzlich die Affenschwanzdomäne von Rna14 (entspricht CstF77₆₁₁₋₆₄₉) enthält [38], [84]. C-terminal konnte von CstF64 eine weitere Struktur aufgelöst werden (CstF64₅₃₁₋₅₇₇).

Da in dieser Region jedoch keine experimentellen *cross-link*-Daten vorhanden waren, wurde diese in der Strukturmodellierung von CstF nicht weiter berücksichtigt.

Zur Generierung der humanen Struktur wurde im ersten Schritt die Struktur der Scharnierdomäne im Komplex mit CstF77 modelliert. Auch hier konnte aufgrund der hohen Sequenzidentität ein sehr gutes Modell generiert werden (Abbildung 3.28 A). Aufgrund der Überlappung dieses Modelles mit der publizierten RRM-Domäne wurden diese im zweiten Schritt mittels MODELLER "zusammengenäht" (Abbildung 3.28 B). Hierbei zeigte sich, dass die generierten Modelle eine sehr geringe Strukturvarianz aufwiesen, was auf eine eindeutige Konformation hindeutet. Das Modell mit der besten Bewertung wurde für die weiteren Experimente ausgewählt. Dieses Modell ist darüber hinaus im Einklang mit den experimentellen *cross-link*-Daten.

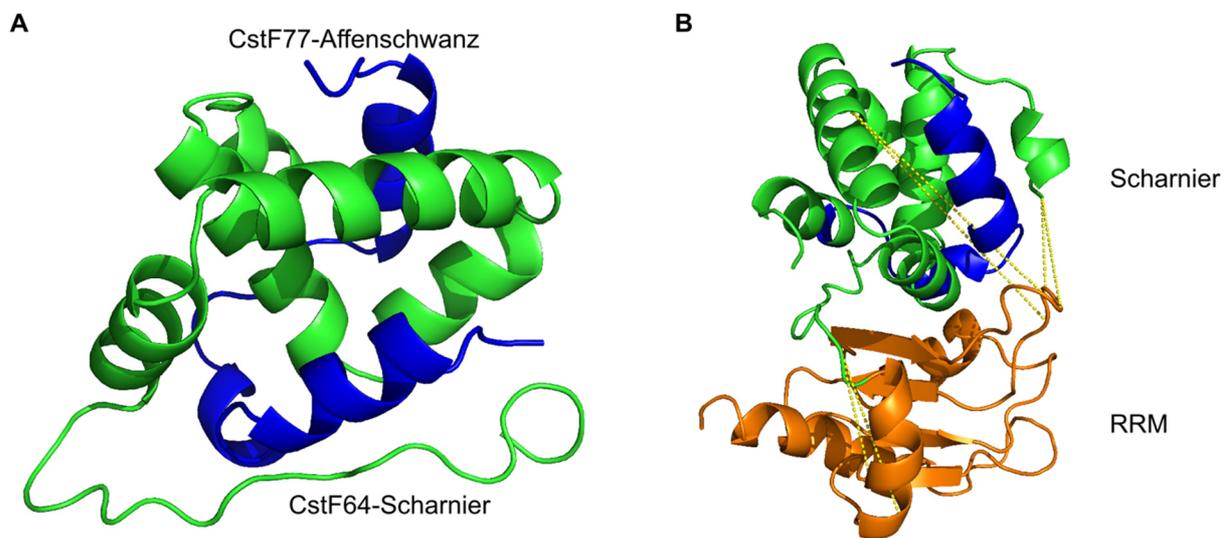


Abbildung 3.28 Darstellung der Modelle für CstF64. A) Struktur des generierten Modelles der Interaktion zwischen CstF64 und CstF77. Die Scharnierdomäne von CstF64 ist in grün, die CstF77-Affenschwanzdomäne in blau dargestellt. B) Verknüpftes Strukturmodell der N-terminalen Region aus CstF64. Die RRM (orange) konnte mit der Scharnierdomäne verbunden werden. Die intramolekularen *cross-links* sind als lineare Verbindungen der Ca -Atome dargestellt (gelb).

CstF77

Das strukturbestimmende Element von CstF77 ist die HAT-Domäne, welche sich aus einem N-terminalen und einem C-terminalen Bereich, welcher homodimerisiert, zusammensetzt. Diese Domäne umfasst die Aminosäuren vom N-Terminus bis 550. Die Struktur dieser Domäne konnte für das murine Homolog gelöst werden (PDBID: 2OOE) [72]. Der Unterschied zur humanen Sequenz zeichnet sich durch eine Punktmutation $\text{K}_{513}\text{-R}_{513}$ sowie das Fehlen des Dipeptids $\text{N}_{241}\text{T}_{242}$ aus. Um diese Unterschiede auszugleichen wurde die humane Struktur modelliert, wobei die dimere Form als Vorlage genutzt wurde. Zur Validierung der Struktur wurden die 116 identifizierten, intramolekularen *cross-links* dieser Domäne auf die Struktur kartiert (Abbildung 3.29). Hierzu wurden alle *cross-links* parallel auf die monomere sowie dimere Struktur kartiert und

die Entfernung bestimmt. Hierbei wurden die *cross-links* jeweils als intramolekular klassifiziert, wenn sie unterhalb der Grenze von 34 Å lagen. So konnten die 118 *cross-links* in 98 intra- und 12 intermolekulare *cross-links* eingeordnet werden. 8 *cross-links* lagen außerhalb der erlaubten Grenzen und konnten weder intra- noch intermolekulare *cross-links* zugeordnet werden (Abbildung 3.29, dargestellt als rote, intramolekulare *cross-links*). Diese sind möglicherweise ein Hinweis darauf, dass die Kristallstruktur nicht in Gänze die native Struktur von CstF77 widerspiegelt.

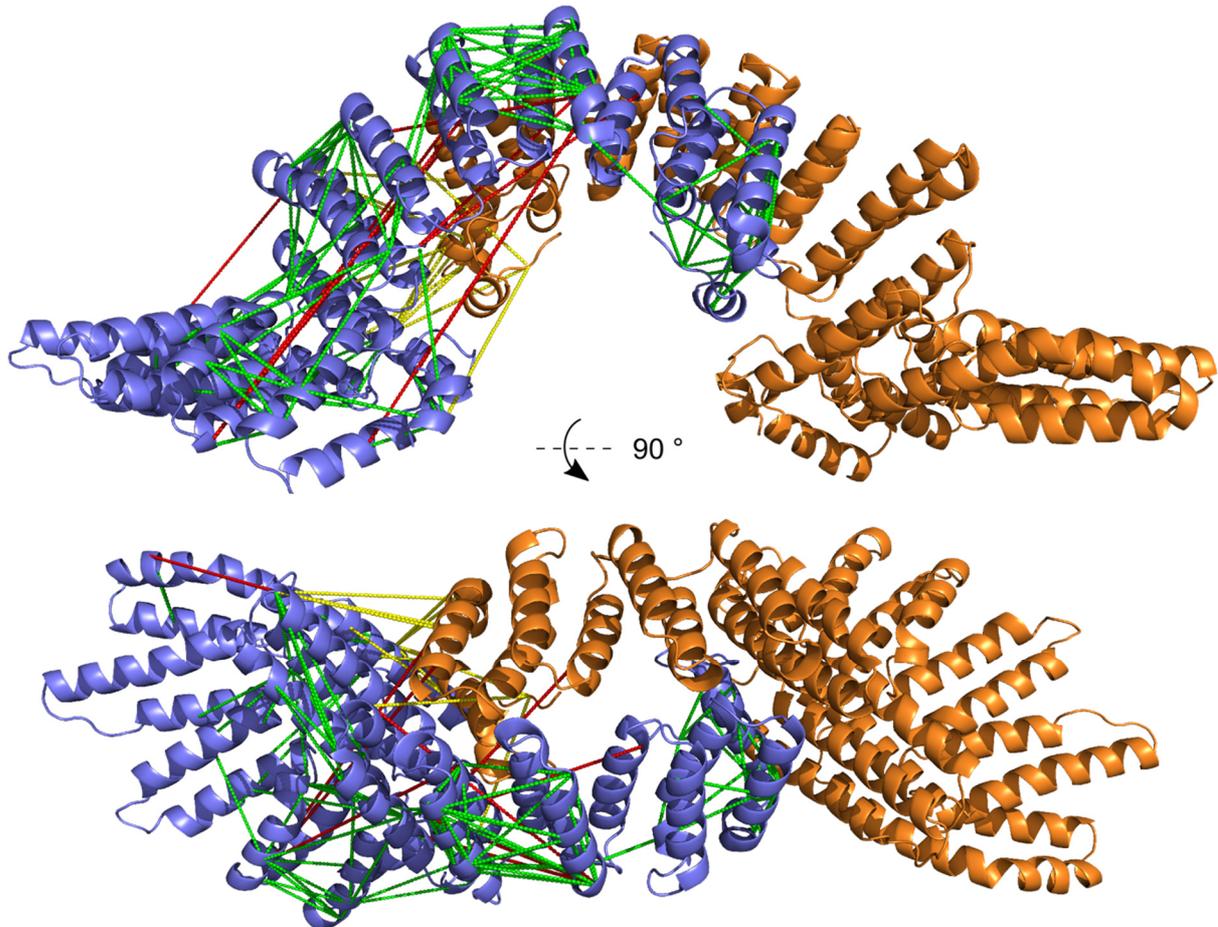


Abbildung 3.29 Homologiemodell des humanen CstF77-Dimers. Dargestellt sind zwei Blickwinkel auf das generierte Strukturmodell des CstF77-Dimers, wobei die beiden monomeren Untereinheiten farblich getrennt sind (Blau / Orange). Auf die Struktur sind die intra-(Grün) und intermolekulare (Gelb) *cross-links* als lineare Verbindung zwischen den Ca-Atomen gezeigt. Zur Übersichtlichkeit wurden diese nur auf einem Monomer und einem Dimer zugeordnet. *Cross-links*, welche aufgrund der Verletzung der Obergrenze weder intra- noch intermolekulare Verknüpfungen zugeordnet werden konnten sind als rote, intramolekulare *cross-links* dargestellt.

3.2.2.2.2 Docking CstF

Zur Assemblierung des CstF-Komplexes wurden zuerst alle dimeren Strukturpaare, die über intermolekulare *cross-links* verfügten, mittels HADDOCK zusammengefügt [138]. Hierbei konnten

die experimentellen *cross-link*-Daten jedoch nur eingeschränkt genutzt werden, da keine eindeutige Zuordnung aufgrund der hexameren Struktur möglich war. Sie wurden daher als mehrdeutige Beschränkung (engl. *ambiguous*) genutzt, was bedeutet, dass das gebildete Strukturmodell nicht alle Beschränkungen erfüllen muss. Die verwendeten Entfernungsbeschränkungen sind im Anhang tabelliert (Tabelle A 11). Erfolglos wurden die Paare CstF50₂₋₆₁-CstF50₉₀₋₄₂₇ und CstF64₈₋₁₈₉-CstF70k₂₁₋₅₄₉ getestet, wobei in CstF50₉₀₋₄₂₇-CstF64₈₋₁₈₉ die gebundenen Peptide von CstF77 mitberücksichtigt wurden (Daten nicht gezeigt).

Für das *Docking* von CstF50₉₀₋₄₂₇-CstF64₈₋₁₈₉ mit gebundenem Peptid (CstF77₅₈₁₋₅₉₃) und der Affenschwanzdomäne (CstF77₆₁₁₋₆₄₉) von CstF77 wurden 37 Entfernungsbeschränkungen verwendet, wobei diese dem Standardbereich (3 bis 34 Å) entsprachen und keine Rücksicht auf kürzere Seitenketten genommen wurde. Alle an den verwendeten *cross-links* beteiligten Aminosäuren waren in den Strukturen aufgelöst.

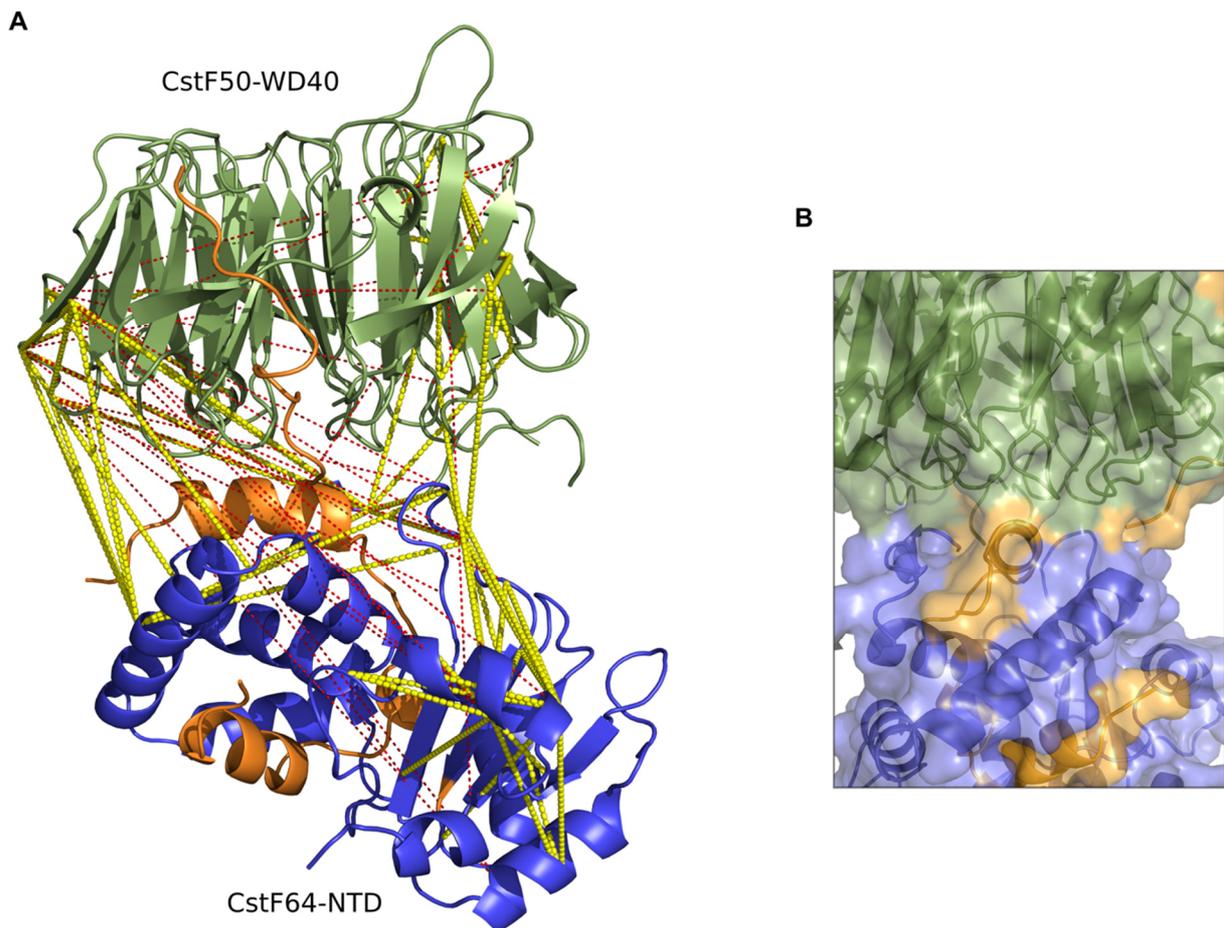


Abbildung 3.30 Darstellung des Komplexstruktur von CstF50 und CstF64. A) Aufgelöst in der Struktur sind die WD40-Domäne von CstF50 (Oliv; CstF50₉₀₋₄₂₇), die N-terminale Domäne von CstF64 (Blau; CstF64₈₋₁₈₉) sowie die beiden gebundenen Regionen von CstF77 (Orange; CstF77₅₈₁₋₅₉₃ sowie CstF77₆₁₁₋₆₄₉). B) Darstellung der Interaktion zwischen CstF50 und CstF64. Die Strukturen sind als Cartoon mit überlagerter Oberfläche dargestellt. Der Fokus ist auf der zentralen Helix von CstF77.

Die Analyse des Dimers zeigte, dass zwei Cluster mit 117 bzw. 81 Proteinen, was 99% der finalen Modelle entspricht, gebildet werden konnten. Diese unterschieden sich durch eine leichte Rotation von CstF64₈₋₁₈₉ in Relation zu der WD40-Domäne von CstF50, wobei der RMSD-Unterschied 2.2 Å beträgt. Aufgrund des geringen Unterschiedes wurde für weitere Analysen das beste Modell des größeren Clusters ausgewählt. Als erste Analyse dieses Komplexes wurden die 82 intra- und intermolekularen *cross-links*, die in dem Sequenzbereich der beiden Strukturen lag, auf die Struktur kartiert (Abbildung 3.30 A). Hierbei zeigte sich, dass 58 innerhalb der erlaubten Entfernung von 34 Å lagen, 24 jedoch außerhalb. Diese entsprechen zudem Aminosäuren auf gegenüberliegenden Seiten des Komplexes, sodass diese *cross-links* vermutlich intermolekulare Vernetzungen zu der zweiten Kopie dieses Dimers entsprechen. Die Interaktionsfläche zwischen CstF50 und CstF64 wird vor allem durch die Helix der Affenschwanzdomäne von CstF77 (CstF77₆₃₄₋₆₄₉) gebildet (Abbildung 3.30 B).

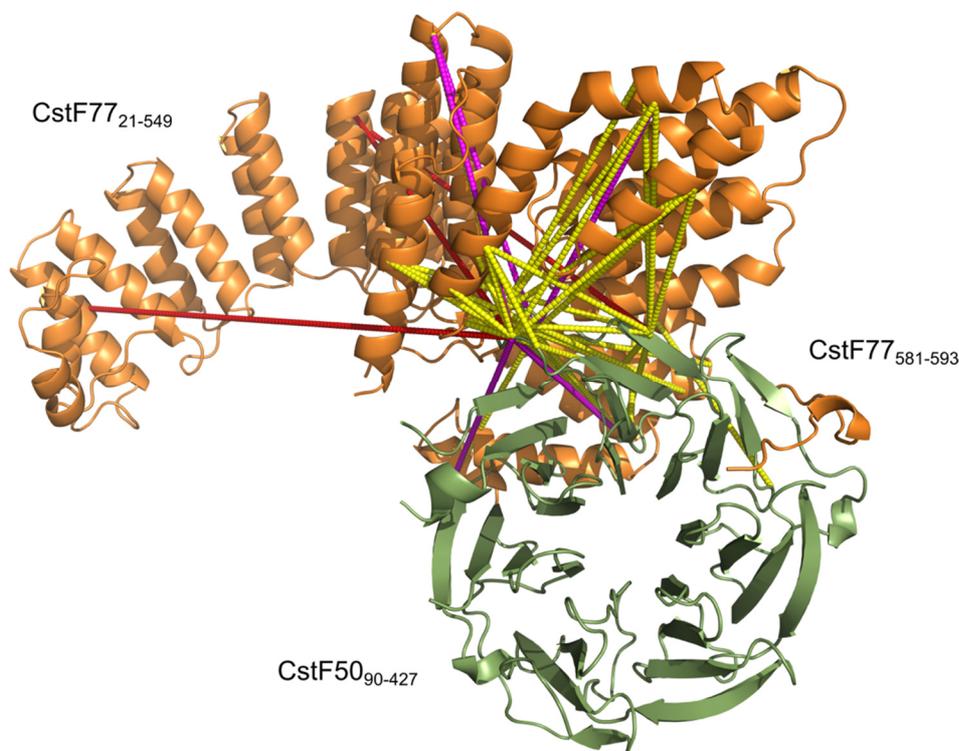


Abbildung 3.31 Strukturmodell des CstF50-CstF77 Dimers. Dargestellt ist die monomere HAT-Domäne von CstF77 (Orange) im Komplex mit der WD40-Domäne von CstF50 (Oliv; CstF50₉₀₋₄₂₇), welche das Peptid CstF₅₈₁₋₅₉₃ gebunden hat. Die *cross-links* sind als lineare Verbindung zwischen den beteiligten Ca-Atomen eingezeichnet. *Cross-links* mit einer Ca-Ca-Entfernung von unterhalb 34 Å sind in Gelb, mit einer Entfernung zwischen 34 und 40 Å in Magenta und solche oberhalb von 40 Å in Rot dargestellt.

In dem *Docking* des CstF50₉₀₋₄₂₇-CstF77₂₁₋₅₄₉ Paares wurden 46 experimentelle *cross-link*-Ergebnisse als direkte Entfernungsbeschränkungen verwendet. Alle generierten 200 Modelle wurden zu einem Cluster zusammengefügt, was einer eindeutigen Konformation entspricht. Zur

Validierung dieses Modelles wurden die experimentellen *cross-link*-Daten auf dieses Modell kartiert (Abbildung 3.31). Es zeigte sich, dass 37 *cross-links* innerhalb des erlaubten Bereiches bis 34 Å lagen. Von den 9 *cross-links*, welche oberhalb dieser Grenze lagen, konnten 6 einem erweiterten Bereich bis 40 Å zugeordnet werden.

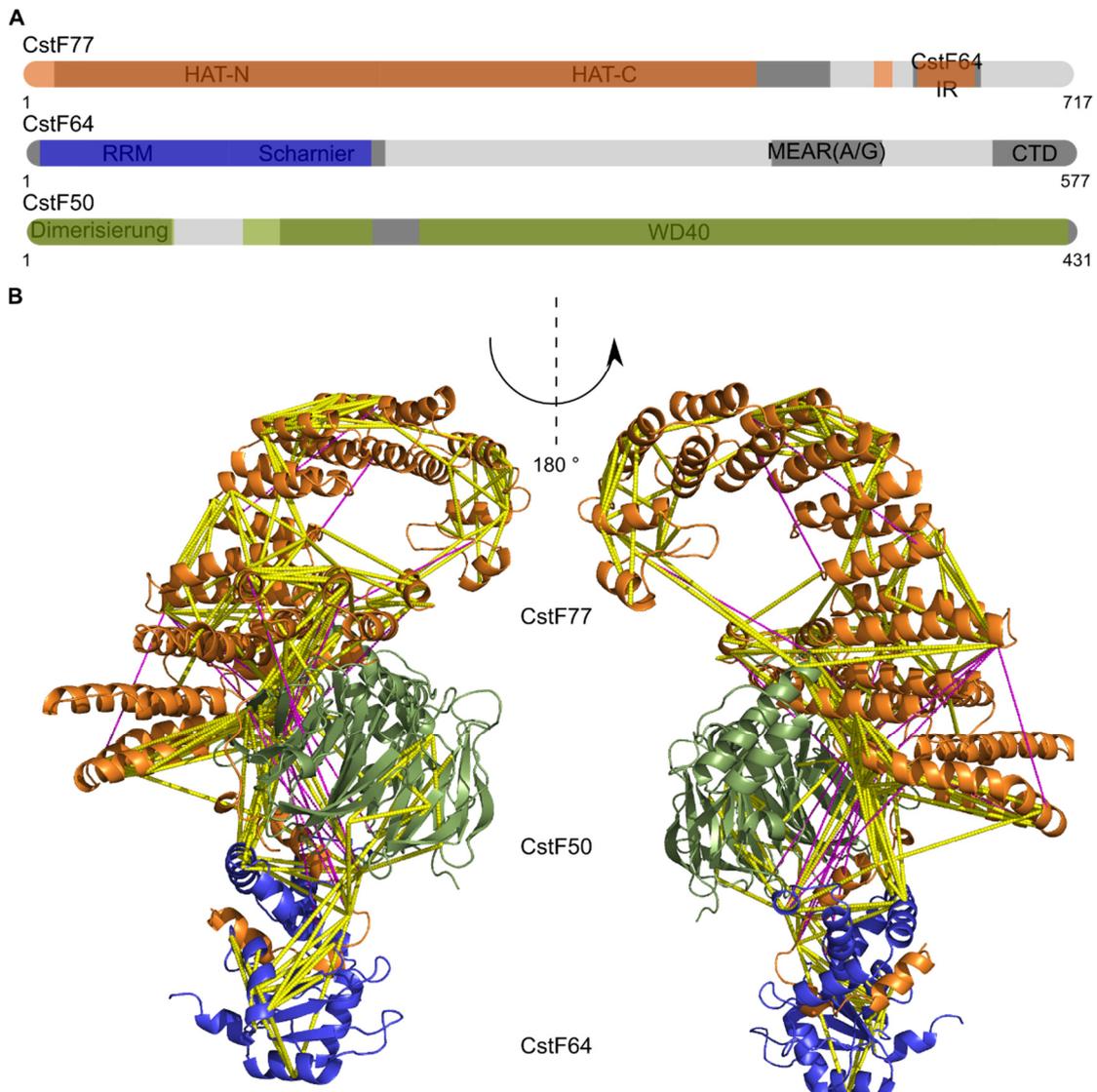


Abbildung 3.32 Modell der trimeren Struktur von CstF. A) Maßstabgerechte Darstellung der Domänenanordnung und der im finalen Modell aufgelösten Bereiche (farblich hervorgehoben) der CstF-Untereinheiten. B) Dargestellt ist die trimere Struktur von CstF, welche durch Überlagerung der WD40-Domäne in CstF50, die in den Dimeren CstF50-CstF64 und CstF50-CstF77 aufgelöst war, generiert wurde. Die Proteine sind farblich getrennt dargestellt. Die experimentellen *cross-links* sind als lineare Verbindung zwischen den beteiligten Ca-Atomen dargestellt. *Cross-links* mit einer Entfernung von bis zu 34 Å sind in Gelb, solche mit einer Entfernung von 34 bis zu 40 Å in Magenta dargestellt.

Zu dem CstF50₉₀₋₄₂₇-CstF77₂₁₋₅₄₉-Komplex wurde im nächsten *Docking* Schritt die Dimerisierungsdomäne CstF50₂₋₆₁ hinzugefügt, wobei 22 *cross-links* verwendet wurden. Es konnte ein Cluster aus 166 Modellen gebildet werden. Da alle weiteren Versuche, den Komplex

weiter zu assemblieren, fehlschlagen, wurden die Komplexe CstF50-CstF77 und CstF50-CstF64 zu einem Trimer zusammengefügt. Da beide Subkomplexe die WD40-Domäne von CstF50 besitzen, wurde diese zur Ausrichtung genutzt. Der so gebildete Komplex stellt die trimere Form von CstF dar (Abbildung 3.32). Zur Validierung dieser Struktur wurden die experimentellen *cross-link*-Daten auf die Struktur kartiert. Insgesamt wurden so 285 einzigartige *cross-link*-Paare verwendet. Hiervon lagen 208 innerhalb des erlaubten Rahmens von bis zu 34 Å (Abbildung 3.32 gelb Linien), 22 in einem erweiterten Rahmen von 34 bis 40 Å (Abbildung 3.32 magenta Linien), und 55 oberhalb davon (nicht dargestellt). Diese widersprechen jedoch nicht dem Modell *per se*, sondern stellen vermutlich *cross-links* in der hexameren Struktur von CstF dar.

Der Versuch, aus der generierten trimeren Struktur durch Symmetrieeoperationen die native hexamere Struktur zu bilden, schlug fehl. Wird die publizierte dimere Struktur von CstF77 als Vorlage verwendet, liegen die CstF50- und CstF64-Untereinheiten zu weit auseinander um im Einklang mit den experimentellen Daten zu sein. Wird die dimere Struktur von CstF50 als Vorlage verwendet, kommt es zu einer Überlagerung der CstF77-Monomeren. Die Ursache hierfür sind die möglicherweise starren Strukturen der Kristallstrukturen. Zur Lösung müsste eine Dynamiksimulation durchgeführt werden, wobei die beiden trimeren Formen zusammengefügt werden.

3.2.2.3 Modellierung des Endonuklease-Komplexes

Der Endonuklease-Komplex besteht aus den vier Untereinheiten CPSF73, CPSF100, Symplekin sowie CstF64. Letzteres wurde strukturell schon analysiert und das generierte Modell CstF64₈₋₁₈₉ (siehe 3.2.2.2.1-CstF64) wurde hier auch verwendet.

3.2.2.3.1 Generierung der monomeren Untereinheiten

CPSF73

Die Struktur Metallo- β -Lactamase- und β -CASP-Domäne des humanen CPSF73 konnte röntgenkristallographisch aufgeklärt werden (Aminosäuren 9 bis 459; PDBID: 2I7T) [111]. Diese Struktur enthält zwei Sequenzlücken (Aminosäuren 113 bis 121 und 289 bis 299). Der in dieser Region lokalisierte intermolekulare *cross-link* kann jedoch durch die Verwendung der Peptidbeschränkung inkludiert werden, sodass diese Lücken nicht durch *de-novo*-Modellierung aufgefüllt wurden. Ein weiterer *cross-link*, welcher in der Kristallstruktur aufgelöst war, entsprach der erlaubten Entfernung (Daten nicht gezeigt).

Der C-terminale Bereich von CPSF73 wird später im Zusammenhang mit CPSF100 näher analysiert.

CPSF100

Die Struktur der Metallo- β -Lactamase- und β -CASP-Domäne des humanen CPSF100 ist nicht bekannt, jedoch wurde das homologe Protein aus Hefe, Ctf2, kristallisiert. Hierbei wurde die unbekannte Region, welche in die β -CASP-Domäne integriert ist, nicht mitberücksichtigt. Aufgrund der hohen Sequenzähnlichkeit konnte die Struktur des humanen Proteins modelliert werden, wobei die Struktur die Sequenzbereiche CPSF100₂₁₋₃₉₂ und CPSF100₅₁₁₋₅₉₈ umschließt (Abbildung 3.33). Zur Validierung des Modelles wurden alle 28 *cross-links*, die im

Sequenzbereich des Modelles lagen, genutzt. Hierbei lagen 27 im erlaubten Bereich, ein *cross-link* lag mit 48 Å außerhalb des erlaubten Bereiches. Der C-terminale Bereich von CPSF100 wird im nächsten Abschnitt näher behandelt.



Abbildung 3.33 Homologiemodell der Metallo- β -Lactamase und β -CASP-Domäne von CPSF100. Die Struktur ist als grüner Cartoon dargestellt. Die intramolekularen *cross-links* sind als Linien zwischen den α -Atomen der beteiligten Aminosäuren dargestellt. In Gelb sind *cross-links* innerhalb und in Rot außerhalb der erlaubten Bereiche gezeigt.

CTD-Interaktion von CPSF73 und CPSF100

In beiden C-terminalen Bereichen von CPSF73 und CPSF100 konnten Strukturhomologe identifiziert werden. Diese homologen Strukturen sind interessanterweise die C-terminalen Regionen der Integratorproteine Int9 und Int11. Der Integrator-Komplex ist in der 3'-Prozessierung von nicht-codierenden RNAs beteiligt. Int9 und Int11, welches ein Paralog von CPSF73 ist, gehören auch zur β -CASP-Familie und die Dimerisierung ist essentiell für die Aktivität [171]. Die Dimerisierungsdomänen der Protein Int9 und Int11 konnte als Komplex kristallisiert werden (PDBID: 5V8W) [171]. Die Struktur der CTD von Int9 ist zu CPSF100₇₁₂₋₇₈₂ und die Struktur der CTD von Int11 zu CPSF73₄₈₈₋₅₈₄ strukturhomolog. Die dimere Struktur der CTDs von CPSF73 und CPSF100 wurden daher komodelliert (Abbildung 3.34). Da keine intermolekularen *cross-links*

in diesem Bereich vorlagen, konnte das Modell hierdurch nicht validiert werden.

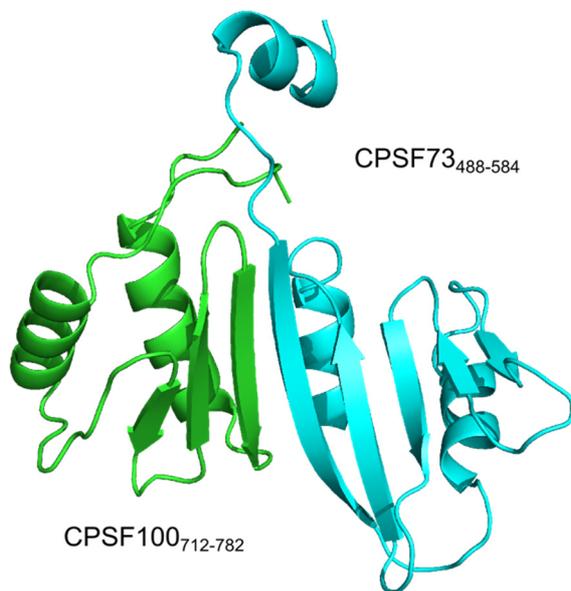


Abbildung 3.34 Homologiemodell des Dimers der C-terminalen Domänen von CPSF73 und CPSF100. Die Dimerisierung der beiden CPSF-Untereinheiten erfolgt über die C-terminalen Domänen (CTD) von CPSF73 (Türkis) und CPSF100 (Grün). Dieses ist homolog zu der Interaktion der Integratorproteine Int9 und Int11.

Symplekin

Die N-terminale HEAT-Domäne des humanen Symplekin konnte kristallisiert und die Struktur aufgeklärt werden (Symplekin₃₀₋₃₄₅; PDBID: 3O2S) [119]. Zur Validierung dieser Struktur wurden die 4 intramolekularen *cross-links* auf die Struktur kartiert, wobei alle im erlaubten Rahmen lagen (Abbildung 3.35 A).

Eine Sequenzanalyse der weiteren Sequenz zeigte nur im Bereich Symplekin₆₆₃₋₁₀₄₂ eine Strukturhomologie zu dem Clathrin-Adaptor Protein AP-1 γ (PDBID: 4P6Z) [173]. Dieser Bereich verfügt keine intermolekularen *cross-links* und kann somit in der Assemblierung des Endonuklease-Komplexes nicht zugeordnet werden. Aufgrund der sehr hohen Anzahl an intramolekularen *cross-links* wurde dieser Bereich trotzdem modelliert und die 40 *cross-links* auf die Struktur kartiert. Es zeigt sich, dass dieser Bereich eine weitere HEAT-Domäne bildet (Abbildung 3.35 B).

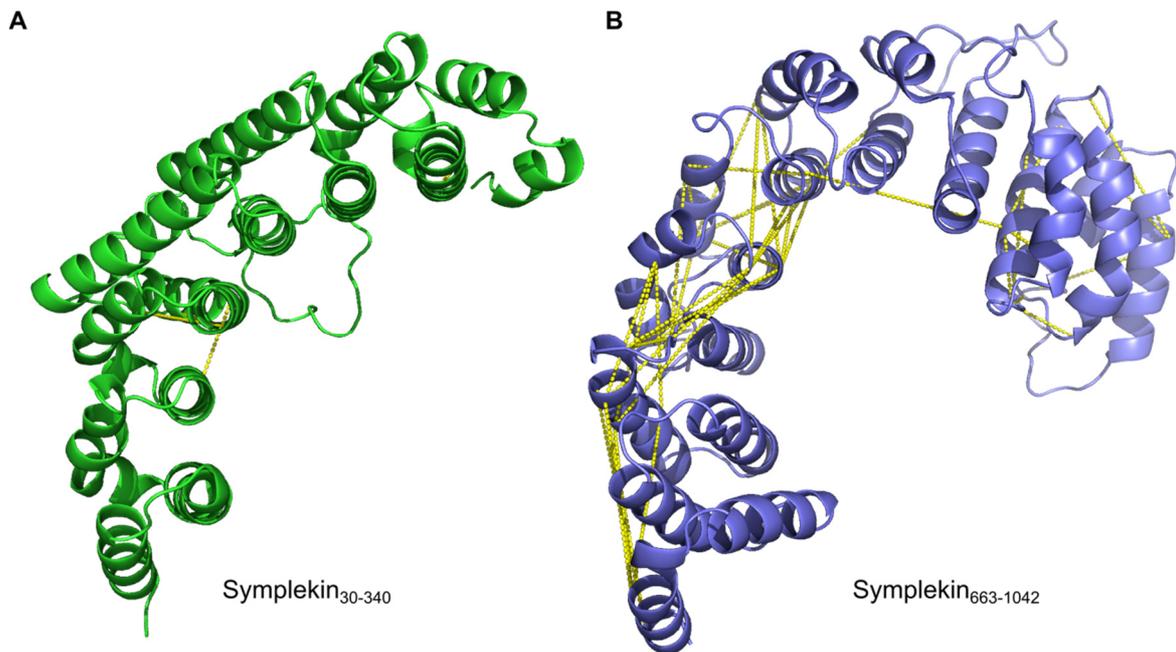


Abbildung 3.35 Strukturen der HEAT-Domänen von Symplekin. A) Publizierte Kristallstruktur der N-terminalen HEAT-Domäne von Symplekin. B) Homologiemodell der C-terminalen HEAT-Domäne von Symplekin. In beiden Modellen sind die intramolekularen *cross-links* als gelbe Linien zwischen den α -Atomen der beteiligten Aminosäuren dargestellt.

Symplekin – CstF64 Interaktion

Symplekin interagiert wie CstF77 mit der Scharnierdomäne von CstF64 [84]. Da die Bindung von CstF64 kompetitiv zwischen Symplekin und CstF77 ist, und die Bindungsoberfläche auf Grund der Größe durch die Struktur von CstF64 vorgegeben wird, ist davon auszugehen, dass die Interaktionsregion von Symplekin eine ähnliche Struktur aufweist wie die Affenschwanzdomäne von CstF77₆₁₁₋₆₄₉ (siehe 2.2.2.2.1-CstF64). Es konnte darüber hinaus gezeigt werden, dass Symplekin₄₄₂₋₄₆₀ an der Bindung beteiligt ist, die genaue Interaktionsregion konnte jedoch aufgrund fehlender Sequenzähnlichkeit zu CstF77 nicht geklärt werden [87]. Zwischen der potentiellen Interaktionsregion von Symplekin und CstF64 konnte ein intramolekularer *cross-link* identifiziert werden, welcher N-terminal der beschriebenen Interaktionsregion liegt: CstF64₁₈₉ x Symplekin₄₇₇. In der Struktur von CstF liegt ebenso in dieser Region ein intermolekularer *cross-link*: CstF64₁₈₉ x CstF77₆₄₇. Da in der Sequenz von Symplekin zwei weitere Lysine, K₄₇₃ und K₄₈₀, vorkommen, hier jedoch kein *cross-link* zu CstF64 identifiziert werden konnte, gibt es nur noch fünf mögliche Sequenzabschnitte von Symplekin für die Interaktionsregion zu CstF64: Symplekin₄₄₀₋₄₇₇ bis Symplekin₄₄₄₋₄₈₁. Es wurden daher die Strukturen aller Varianten im Komplex mit CstF64 modelliert und überlagert (Abbildung 3.36). Hierbei zeigte sich, dass nur in Symplekin₄₄₃₋₄₈₀ das Lysin₄₇₇ exponiert vorliegt (Abbildung 3.36 Magenta). In Symplekin₄₄₄₋₄₈₁, welches das Lysin₄₇₇ an derselben Position wie CstF77 besitzt, ist dieses in der Interaktionsfläche bedeckt. Stattdessen ist das Lysin₄₇₃ exponiert, welches jedoch keinen *cross-link* zu CstF64 aufweist (Abbildung 3.36 Orange).

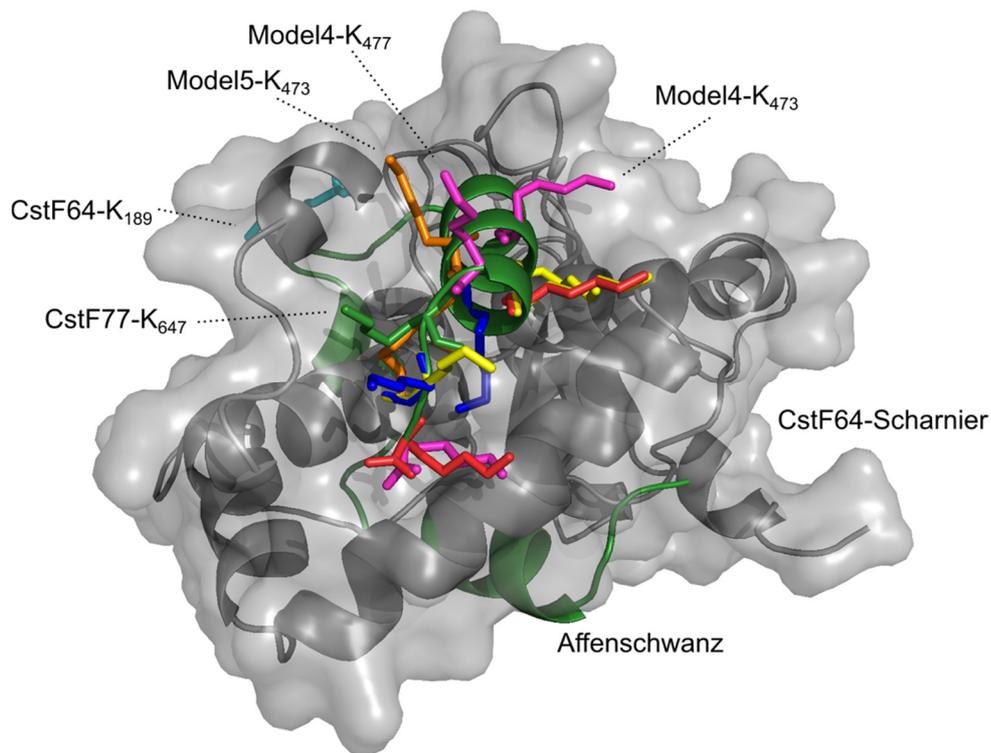


Abbildung 3.36 Zusammenfassung der Strukturmodellierung der Interaktion von Symplekin und CstF64. Die Scharnierdomäne von CstF64 ist sowohl als Cartoon- als auch Oberflächenmodell in grau dargestellt. Die Affenschwanzdomäne von CstF77 ist in Grün dargestellt. In Stäbchenform sind die Lysine der Modelle in der Interaktionsregion hervorgehoben, wobei jedes Modell eine andere Farbe trägt: Model1 (Symplekin₄₄₀₋₄₇₇) – Rot, Model2 (Symplekin₄₄₁₋₄₇₈) – Gelb, Model3 (Symplekin₄₄₂₋₄₇₉) – Blau, Model4 (Symplekin₄₄₃₋₄₈₀) – Magenta; Model5 (Symplekin₄₄₄₋₄₈₁) – Orange.

3.2.2.3.2 Docking Endonuklease-Komplex

Für die molekulare Assemblierung des Endonuklease-Komplexes wurden verschiedene Methoden zur Generierung der Entfernungsbeschränkungen genutzt. Insgesamt wurden 21 Beschränkungen definiert, die auf 19 experimentellen *cross-links* basierten, von denen wiederum 7 unter der Verwendung der Peptidbeschränkung inkludiert werden konnten. Die verwendeten Entfernungsbeschränkungen sind im Anhang aufgeführt (Tabelle A 12). Für zwei Entfernungen, CPSF73₄₅₉ bis CPSF73₄₉₅ sowie Symplekin₃₄₅ bis Symplekin₄₄₃, wurden die unbekannt Sequenzen mittels des I-TASSER-Webserver modelliert [140]. Für jede Sequenz wurden so fünf Modelle generiert, aus deren Durchmesser sich die Entfernungsbeschränkung ableiten ließ. Die fehlenden Strukturbereiche konnten so für CPSF73 mit einem Durchmesser von 5 bis 25 Å und für Symplekin mit 15 bis 45 Å beschrieben werden.

Zur Bildung eines Modelles des Endonuklease-Komplex wurden alle dimeren Kombinationen der Modelle der N-terminalen β-CASP-Domänen von CPSF73 (CPSF73₁₋₄₅₉), CPSF100 (CPSF100₂₂₋

⁵⁹⁶), der dimeren Struktur der CPSF-CTDs (CPSF73₄₉₅₋₅₈₃ – CPSF100₇₂₁₋₇₈₂), dem Symplekin-CstF64-Dimer (Symplekin₄₄₃₋₄₈₀ – CstF64₁₋₁₈₉) sowie der N-terminalen HEAT-Domäne von Symplekin (Symplekin₃₀₋₃₄₅) verwendet. Die neu identifizierte, C-terminale HEAT-Domäne von Symplekin wies keine inter-molekularen *cross-links* auf und konnte daher nicht verwendet werden.

Es konnte ein stabiler Komplex zwischen CPSF100₂₂₋₅₉₆ und dem Dimer der CPSF-CTDs gebildet werden. Hierbei wurden sieben *cross-links* verwendet, die alle in der Struktur aufgelöst werden. Es konnte ein Cluster mit 147 Strukturen gebildet werden. Da dieser Komplex alle Substrukturen von CPSF100 enthielt, entfällt die Nummerierung und der gebildete Komplex wird als CPSF100-CPSF73₄₉₅₋₅₈₃ bezeichnet (Abbildung 3.37 A). Alle sieben intermolekularen *cross-links* in diesem neuen Komplex sind im erlaubten Entfernungsrahmen.

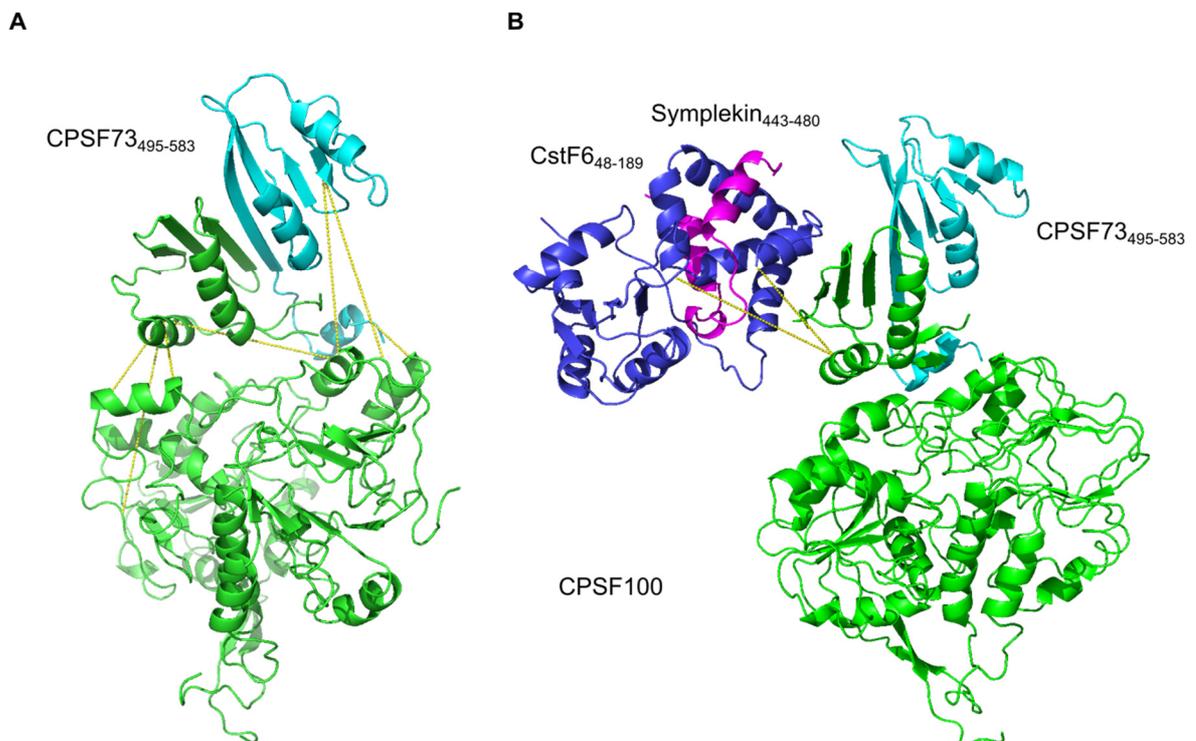


Abbildung 3.37 Strukturmodelle des Dockings des Endonuklease-Komplexes. Die dargestellten Proteine sind farblich differenziert als Cartoon dargestellt (CPSF73 – Cyan; CPSF100 – Grün; Symplekin – Magenta; CstF64 – Blau). *Cross-links* zwischen den neu assemblierten Komplexen sind als gelbe Linien zwischen den α -Atomen der beteiligten Aminosäuren gezeigt. A) Strukturmodell des Dimers zwischen CPSF100₂₂₋₅₉₆ und dem CTD-Dimer von CPSF. B) Strukturmodell des in A) gezeigten Dimers und dem CstF64-Symplekin-Dimer.

Da alle anderen dimeren Versuche, einen Komplex zu bilden, zu keinem Ergebnis führten, wurde der CPSF100-CPSF73₄₉₅₋₅₈₃-Komplex als Grundlage für die weiteren *Docking*-Versuche genutzt. Wiederum wurden alle dimeren Kombinationen *gedockt*. Die Kombination mit dem Symplekin-

CstF64 Dimer führte zu einer Komplexbildung (Abbildung 3.37 B). Unter Verwendung von drei *cross-links*, von denen einer nicht in der Struktur aufgelöst war, sondern unter Verwendung der Peptidbeschränkung inkludiert wurde, konnte ein Cluster mit 191 von 200 Strukturen gebildet werden, welches für eine eindeutige Positionierung spricht. Die beiden intermolekularen *cross-links*, die in dieser Struktur aufgelöst sind, liegen in dem erlaubten Rahmen.

Der so gebildete tetramere Komplex wurde anschließend mit CPSF73₁₋₄₅₉ zusammengefügt. Zwischen diesen Strukturen war nur ein *cross-link* in der Struktur aufgelöst. Unter Verwendung der Peptid-Beschränkung konnten jedoch 5 weitere *cross-links* mit in die Berechnung einbezogen werden. Darüber hinaus wurden die Entfernung zwischen CPSF73₁₋₄₅₉ und CPSF73₄₉₅₋₅₈₃ durch eine Modellierung dieses nicht aufgelösten Bereiches mittels I-TASSER mit bis zu 25 Å berechnet [140].

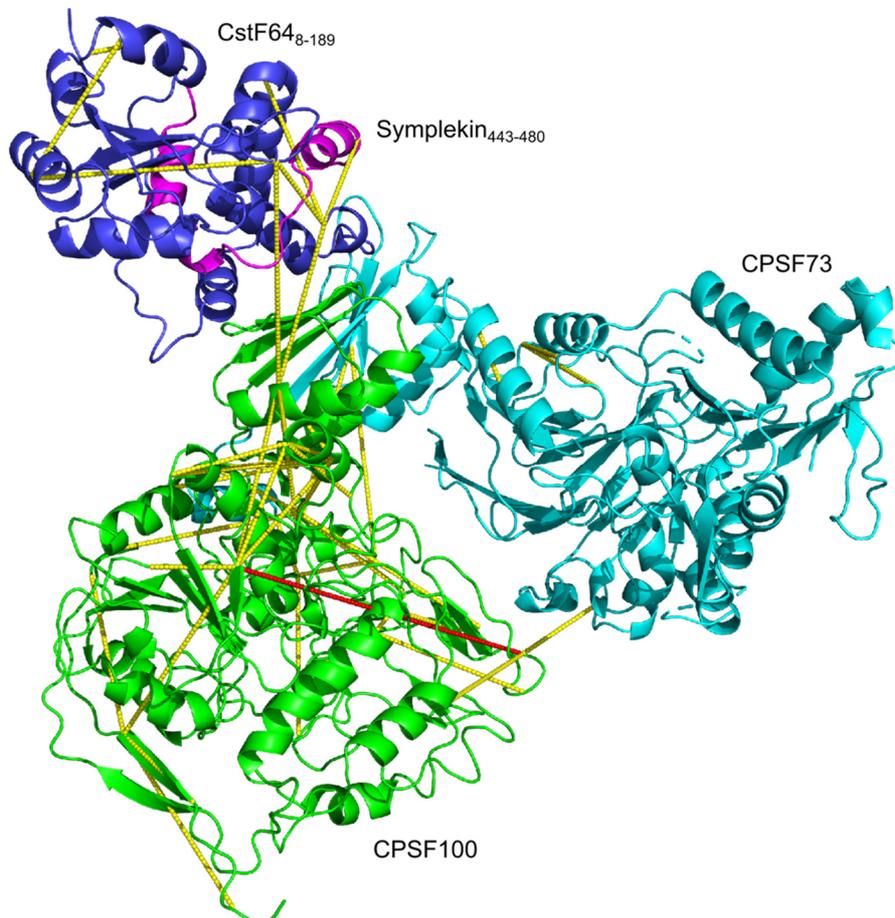


Abbildung 3.38 Strukturmodell des Endonuklease-Komplexes. Das tetramere Strukturmodell des Endonuklease-Komplexes enthält die Strukturen für CPSF73 (Türkis), CPSF100 (Grün) sowie des Dimers der Interaktion zwischen Symplekin (Magenta) und CstF64 (Blau).

Insgesamt konnten also 7 Beschränkungen für das *docking* verwendet werden. Hierbei wurden die 200 generierten Strukturen in 6 Cluster eingeteilt, wobei der größte über 106 Strukturen verfügte. Eine Detailanalyse zeigte jedoch, dass insgesamt 143 der 200 Strukturen eine ähnliche Konformation mit einer RMSD-Varianz von maximal 6 Å aufwiesen. Dieses wurde als natürliche

Varianz in der Flexibilität der Interaktion angesehen und das Modell des größten Clusters mit der besten Bewertung ausgewählt (Abbildung 3.38). Zur Verifizierung wurden alle experimentellen *cross-links* auf diese Struktur kartiert. Von insgesamt 48 *cross-links* lagen 47 im erlaubten Rahmen, wobei der Ausreißer wie beschrieben in der monomeren Struktur von CPSF100 liegt. Zeitgleich zu diesen Experimenten wurde der Endonuklease-Komplex mittels kryo-Elektronenmikroskopie von einer anderen Gruppe untersucht, jedoch konnte aufgrund der Auflösung von 7,4 Å kein Strukturmodell erstellt werden [98]. In diesem waren CPSF73, CPSF100 und ein C-terminaler Bereich von Symplekin (Symplekin₅₃₈₋₁₁₁₀) enthalten, von dem in dieser Arbeit ein Strukturmodell (Symplekin₆₆₃₋₁₀₄₂) generiert wurden. Es zeigte sich, dass die Strukturmodelle und die experimentellen Elektronenmikroskopie-Experimente eine sehr hohe Übereinstimmung besaßen. Das Dimer aus CPSF73 und CPSF100 zeigte eine Übereinstimmung von 68 %, die C-terminale HEAT-Domäne von Symplekin eine Übereinstimmung von 66 %. Dieses ist in Anbetracht der geringen Auflösung und des durch Homologiemodellierung und *Docking* generierte Strukturmodell eine sehr hohe Übereinstimmung und validiert damit dieses Strukturmodell. Es wurde keine weitere Strukturoptimierung anhand dieser Coulombdichte durchgeführt. Das Strukturmodell ist in Abbildung 3.39 gezeigt.

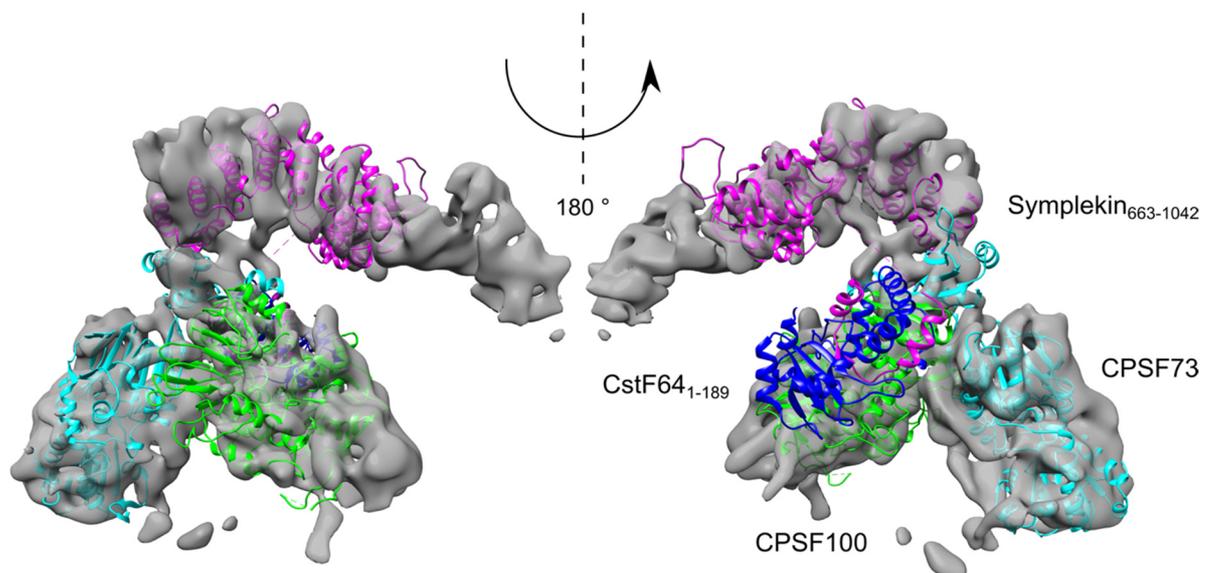


Abbildung 3.39 Finales Strukturmodell des Endonuklease-Komplexes. Das finale Strukturmodell des Endonuklease-Komplexes enthält durch Berücksichtigung der publizierten kryo-EM Dichte (EMDB: 20859; grau) zusätzlich die C-terminale HEAT-Domäne von Symplekin (magenta).

Da ein weiteres molekulares *Docking* mit der N-terminalen HEAT-Domäne von Symplekin fehlschlug, stellt diese die finale Struktur des Endonuklease-Komplexes dar. Auffällig ist, dass in der Struktur von CPSF73 kaum *cross-links* identifiziert wurden. Basierend auf der Analyse der euklidischen Entfernung zwischen zwei C α -Atomen von Lysin mittels Xwalk in der Struktur, sind 160 Interaktion (96 intra- und 64 intermolekular) unterhalb von 30 Å ausgehend von CPSF73

möglich [174]. Eine mögliche Erklärung ist, dass die Oberfläche von CPSF73 nicht zugänglich ist. Diese kann möglicherweise durch die nicht aufgelösten Strukturen von CstF64 oder Symplekin abgeschirmt sein.

Da die Generierung eines molekularen Modells mit der N-terminalen HEAT-Domäne von Symplekin fehlschlug, wurde von diesem der Konformationsraum in Relation zu dem finalen Modell mittels DisVis analysiert (Abbildung 3.40) [136]. Deutlich wird, dass große Teile der beiden CPSF-Untereinheiten nicht oder nur partiell vom N-Terminus bedeckt sein könnten.

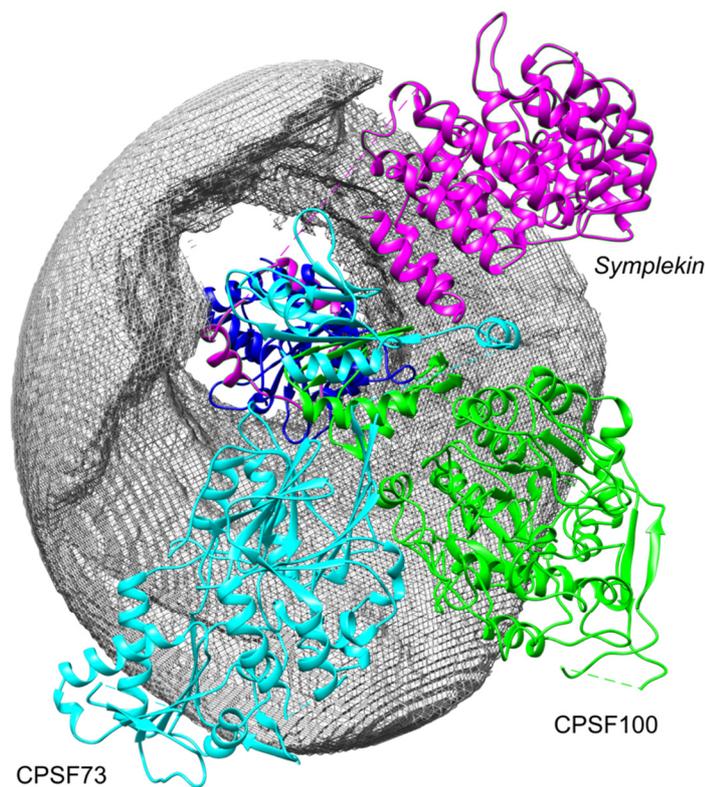


Abbildung 3.40 Analyse des Konformationsraumes der N-terminalen HEAT-Domäne von Symplekin im Endonuklease-Komplex. Das gebildete Strukturmodell des Endonuklease-Komplex wie in Abbildung 3.39 dargestellt. Der mittels DisVis bestimmte mögliche Konformationsraum ist als graues Gitter dargestellt.

3.3 Identifikation weiterer Faktoren des Polyadenylierungskomplexes

Zur Identifikation von weiteren möglicherweise an der Polyadenylierungsreaktion beteiligten Proteinen wurden *Pulldown*-Assays mit den rekombinanten Proteinkomplexen durchgeführt. Hierbei wurden Komplexe mit FLAG- oder Strep-tag verwendet, in molarem Überschuss gegenüber endogenem Komplex mit HeLa-Kernextrakt inkubiert und anschließend affinitätschromatographisch gereinigt. Der Volllängen-CFII- und der mPSF-Komplex, welche beide ausschließlich einen Hexahistidin-tag trugen, zeigten eine sehr unspezifische Anreicherung von Proteinen und wurden daher nicht weiter berücksichtigt. Für alle Proteine konnte eine RNA-Bindung identifiziert werden (siehe 2.1), jedoch wurde keine RNase-Behandlung der Proben durchgeführt. Identifizierte Proteine können daher auch über indirekte Protein-RNA-Protein-Interaktionen mit den Köderproteinen assoziiert sein. Die Elutionsfraktionen wurden massenspektrometrisch analysiert. Zur Datenauswertung wurden MaxQuant und die SwissProt-Datenbank (Stand: 08. Mai 2019) verwendet. Zur Bewertung der Proteinabundanz wurde der iBAQ-Wert genutzt, welcher von MaxQuant wie folgt berechnet wird [144]:

$$iBAQ = \frac{\Sigma \text{Intensität}}{\# \text{theoretische Peptide}} \quad (3)$$

Zur Visualisierung der massenspektrometrischen Daten wurden der binäre Logarithmus der iBAQ-Werte gebildet und die Menge der Proteine in dem Intervall (0,45] berechnet. Hierzu wurde dieser Bereich in 45 Klassen J der Breite $h = 1$ aufgeteilt:

$$h = (b - a)/J \quad (4)$$

wobei a und b das Intervall (0,45] und $J = 45$ die Anzahl der Klassen beschreiben.

Die Anzahl der Proteine in der Klasse $j = 1, \dots, J$ wurde nach Formel (5) berechnet. Die hieraus resultierende Glockenkurve erlaubt eine Einschätzung über die Anzahl der identifizierten Proteine sowie über die Anreicherung. Dieses wird durch eine positive Schiefe (rechtsschief) der Kurve deutlich.

$$n_j = \sum_{i=1}^n \mathbf{1}\{a + (j - 1)h < x_i \leq a + jh\} \quad (5)$$

wobei n_j die Anzahl der Proteine in der Klasse j , a der unteren Grenze des Intervalls von 0 und h der Breite der Klasse von 1 entsprechen.

Zur Klassifizierung wurde zum einen die Liste der bekannten und putativen Polyadenylierungsfaktoren genutzt. Darüber hinaus wurden die gegenüber der Kontrolle angereicherten Proteine durch eine „GO Anreicherungsanalyse“ mittels des PANTHER-Klassifizierungssystems analysiert [161]. Hierbei werden Proteine über ihre GO-Klasse (*Gene Ontology*) eingeordnet. Proteine, die der „mRNA 3'-Prozessierung“-Klasse zugeordnet werden konnten, wurden für jedes Experiment gesondert dargestellt. Die GO-Klassifizierung kann für

schlecht charakterisierte Proteine keine oder nur auf Sequenzvergleichen basierende Einträge enthalten. Auch können Proteine, die zwar *in vivo* an der 3'-Prozessierung beteiligt sind, bisher nicht jedoch als solche identifiziert wurden, falsch oder gar nicht klassifiziert werden.

Proteine anderer RNA-verbundener Prozesse wie Spleiß- oder Transkriptionsfaktoren, Prozessierung von rRNA, tRNA und anderen RNAs wurden, sofern sie nicht auch zusätzlich eine Assoziation zur mRNA-3'-Prozessierung in der Datenbank enthielten, zur Übersichtlichkeit hier nicht dargestellt.

Die Experimente zur Interaktomanalyse von CstF sowie den verkürzten Varianten von CFII wurden parallel durchgeführt, sodass die gleiche Pufferkontrolle hier mehrfach gezeigt wird. Das Experiment zur Interaktomanalyse für den Endonuklease-Komplexes wurde an einem anderen Massenspektrometer gemessen. Daher war eine direkte Korrelation mit den vorherigen Messungen nicht möglich, sodass hier eine separate Pufferkontrolle durchgeführt wurde.

3.3.1 Interaktom von CFII

3.3.1.1 Darstellung und Aktivitätsanalyse der Köderproteine

Für die Identifikation von Interaktionspartnern von CFII wurden die Komplexe CFII Δ N769 und CFII Δ N1123 genutzt (siehe 2.1.1.5). Zur Durchführung des *Pulldowns* wurde zusätzlich zum Hexahistidin-*tag* ein N-terminaler FLAG-*tag* an Clp1 eingeführt. Die Reinigung erfolgte analog zu den in 2.1.1.4 beschriebenen CFII-Komplexen und der FLAG-*tag* wurde nur für das *Pulldown*-Experiment verwendet (Abbildung 3.41 A).

3.3.1.2 Interaktom der verkürzten CFII Varianten

Da sich die beiden Komplexe CFII Δ N769 und CFII Δ N1123 darin unterscheiden, dass CFII Δ N769 aufgrund der FEGP-Wiederholungen die Spaltung in fraktioniertem Kernextrakt rekonstituiert, nicht jedoch in ihrer RNA-Bindung, ist eine Analyse der Interaktome dieser beiden interessant. Es ist anzunehmen, dass die Aktivität durch eine Protein-Protein-Interaktion vermittelt wird und die Identifikation dieses Proteinpartners Aufschluss über die Zusammensetzung des aktiven Spaltungskomplexes der Polyadenylierung geben kann.

Zur Identifikation von Interaktionspartnern wurde je 35 pmol Komplex mit 160 μ l Kernextrakt inkubiert und anschließend über den FLAG-*tag* gereinigt. Im Silber-gefärbtem SDS-Gel sind neben den Köderproteinen weitere angereicherte Proteinbanden zu beobachten. Von den Elutionsfraktion wurde 75% verdaut, massenspektrometrisch analysiert und wie beschrieben ausgewertet. Keine Anreicherung ist in der Pufferkontrolle zu sehen (Abbildung 3.41 B). Es konnten insgesamt 2503 verschiedene Proteine identifiziert werden, davon 1478 in CFII Δ N769, 1789 Proteine CFII Δ N1123 und 2289 in der Pufferkontrolle. Es ist eine deutliche Anreicherung von Proteinen in den CFII-Varianten gegenüber der Kontrolle zu beobachten, welches an der Verschiebung und der Rechtsschiefe der Verteilungsintensität deutlich wird (Abbildung 3.41 C).

Im Vergleich zwischen der aktiven und inaktiven Variante gibt es eine Überlappung von 1289 Proteinen. Darüber hinaus sind 189 Proteine ausschließlich in der CFII Δ N769 und 500 Proteine ausschließlich in CFII Δ N1123 identifiziert worden (Abbildung 3.41 D).

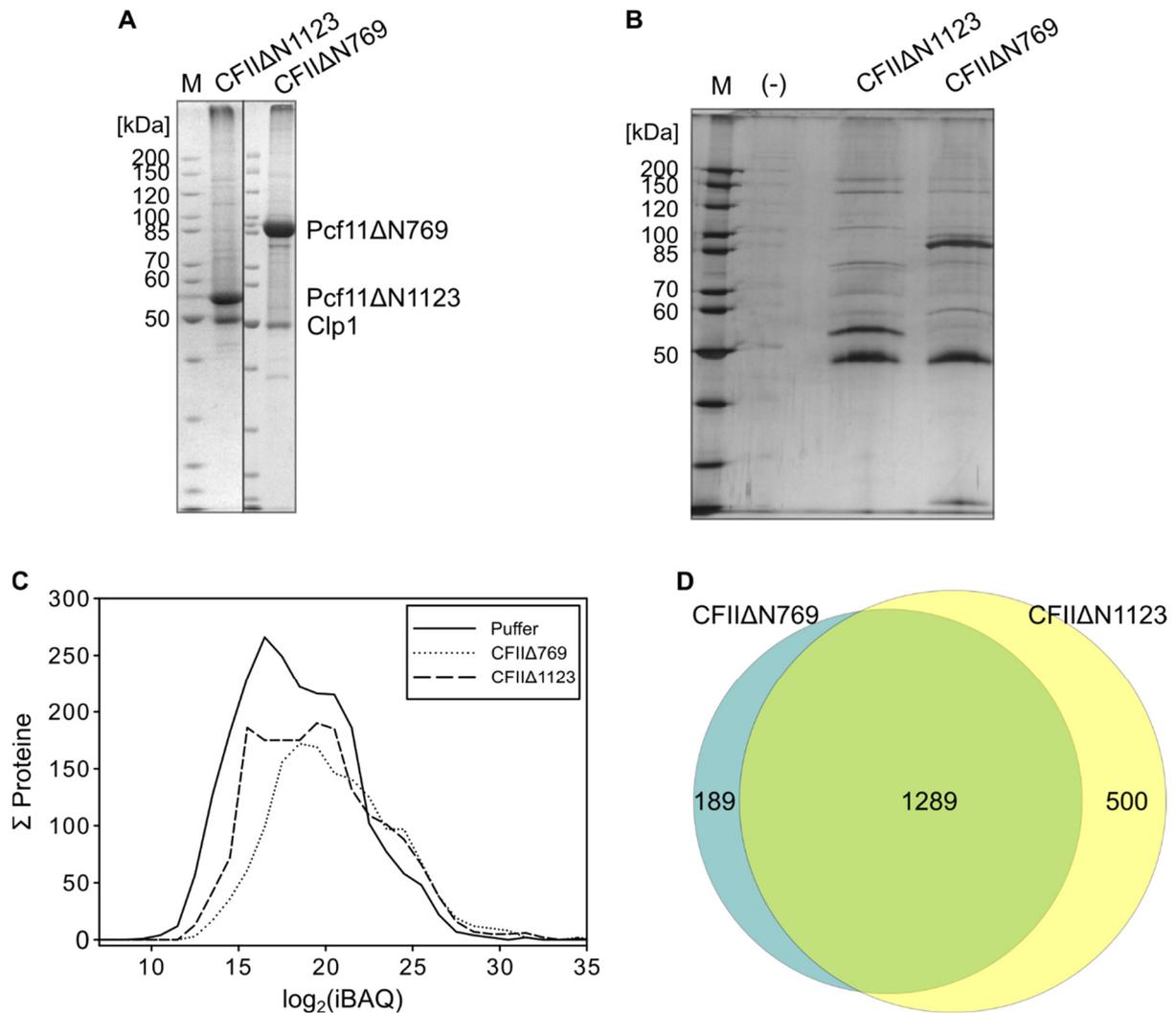


Abbildung 3.41 Analyse des Interaktoms der verkürzten CFII-Komplexe. A) Darstellung der gereinigten verkürzten CFII-Varianten mit kombiniertem Hexahistidin-FLAG-tag. Die Reinigung erfolgte analog zu 2.1.1.4. Die Proteine wurden mit Coomassie eingefärbt und Clp1 sowie die beiden Pcf11-Varianten sind gekennzeichnet. B) Analytisches, silbergefärbtes SDS-Polyacrylamidgel der Elutionsfraktionen des *Pulldown*-Experimentes. Die Pufferkontrolle (-) sowie die beiden CFII-Varianten sind gekennzeichnet. C) Verteilungskurve der identifizierten Proteine. Die Summe der Proteine wurde nach Formel (5) bestimmt. Die Anreicherung von Proteinen in den CFII-*Pulldowns* wird durch die rechtsseitige Schulter der Verteilungskurve gegenüber der Pufferkontrolle deutlich. D) Venn-Diagramm der Summe der identifizierten Proteine in aktiver und inaktiver CFII-Variante.

Zur genaueren Analyse der identifizierten Proteine der *Pulldowns* der CFII-Varianten wurden die binär-logarithmischen iBAQ-Werte gegeneinander aufgetragen (Abbildung 3.42). Die Konfidenzintervalle von 5 % wurden nach Götze *et al.* berechnet [175]. Außerhalb dieser Schwelle

liegen in CFI Δ N769 insgesamt 72, in CFI Δ N1123 67 Proteine. Alle bekannten Subkomplexe des 3'-Prozessierungskomplexes konnten identifiziert werden. Sowohl der CstF- als auch der CPSF-Komplex sind in dem *Pulldown* der spaltungs-inaktiven Variante CFI Δ N1123 angereichert. Die Ursache hierfür ist unklar. Liegen die Interaktionsbereiche jedoch in Clp1 oder im C-terminalen Bereich von Clp1 könnten diese in CFI Δ N1123 zugänglicher sein als im Vergleich zu CFI Δ N769.

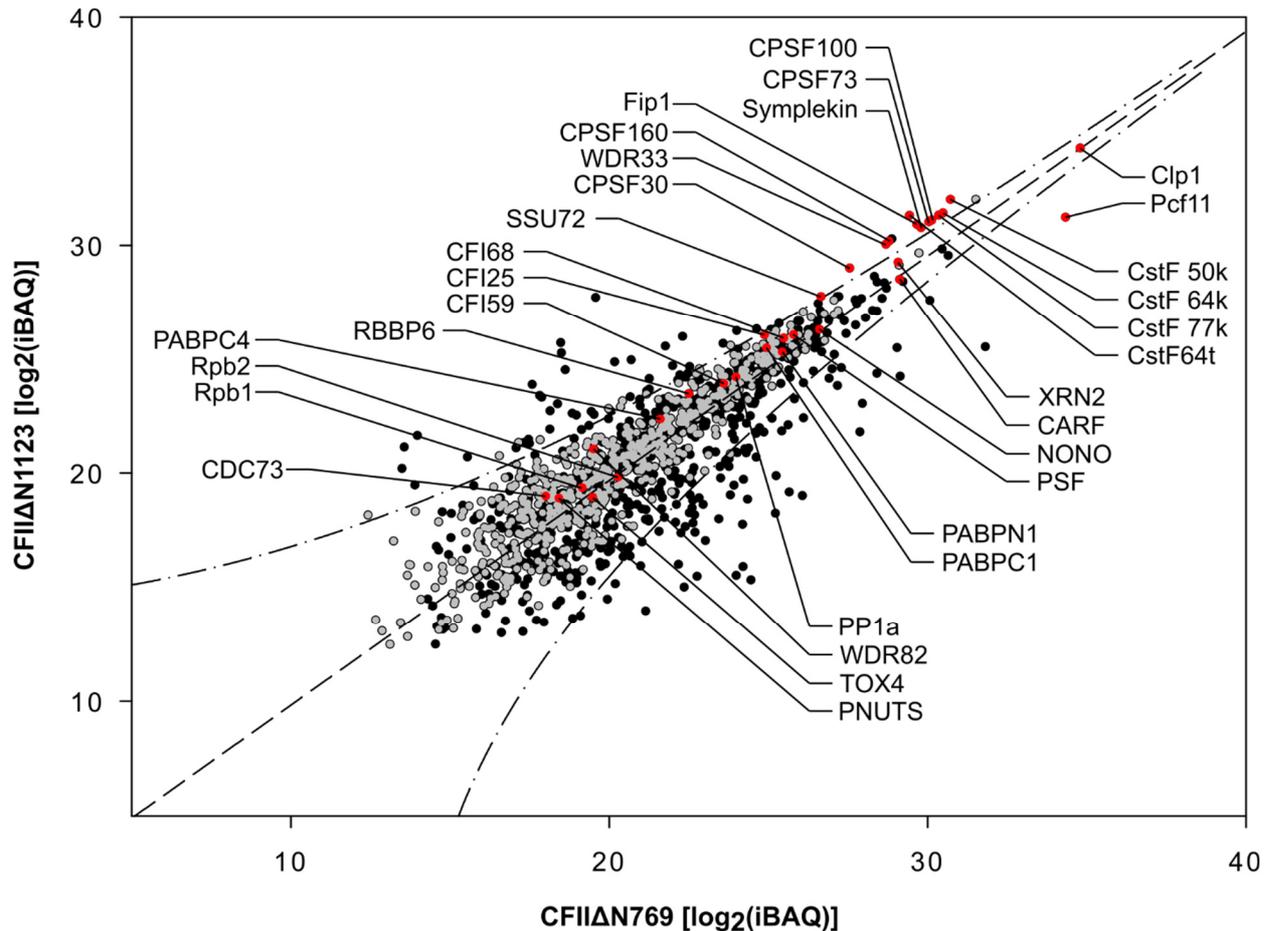


Abbildung 3.42 Quantitative Analyse der identifizierten Proteine in der spaltungs-aktiven gegenüber der spaltungs-inaktiven CFI-Variante. Die Intensitäten aller Proteine, die in beiden CFI-Varianten identifiziert werden konnten, wurden gegeneinander aufgetragen. Proteine, die nur in einer CFI-Variante identifiziert wurden und sich auf der x- bzw. y-Achse befinden, sind nicht dargestellt. Identifizierte Proteine um die Diagonale, welche durch Nullpunkt und Clp1 läuft, sind in keiner der Varianten angereichert. Proteine außerhalb der eingezeichneten Linien sind mit einem p-Wert kleiner 5% angereichert. Polyadenylierungsfaktoren sind rot hervorgehoben und gruppiert beschriftet. Proteine, welche gegenüber der Pufferkontrolle nicht mindestens um den Faktor 2 ($iBAQ_{CFII} / iBAQ_{Puffer}$) angereichert sind, sind grau dargestellt.

Kein Unterschied konnte in der Anreicherung für den CFI-, den XRN2-, sowie den PP1-Komplex beobachtet werden. Der XRN2-Komplex ist für den Abbau des 3'-Fragmentes der gespaltenen prä-mRNA und der Termination der RNA-Polymerase verantwortlich. Hierbei ist eine Interaktion zu Pcf11 beschrieben [56]. Es konnte gezeigt werden, dass der PP1-Komplex auch eine Rolle in der Spaltung und Polyadenylierung spielt [25]. Darüber hinaus konnten drei Poly(A)-bindende

Proteine identifiziert werden, darunter das nukleare Poly(A)-bindende Protein PABPN1, sowie Untereinheiten der RNA Polymerase II (Rpb1 und Rpb2).

Zur Identifikation der in der aktiven gegenüber der inaktiven Variante signifikant angereicherten Proteine, also mit p-Werten unterhalb von 5 % oder solche, die nur in der aktiven Variante identifiziert werden konnten, wurden diese mittels der GO-Begriffe klassifiziert. Auch hier wurde das Kriterium, dass die identifizierten Proteine mindestens Faktor 2 gegenüber der Pufferkontrolle angereichert sein müssen, angewendet. Insgesamt traf dieses Kriterium auf 177 Proteine zu. Die Klassifizierung zeigte nur zwei Proteine der mRNA 3'-Prozessierung, namentlich Pcf11 und die Poly(A)-Polymerase, sowie drei RNA-Helikasen (Tabelle 10). Die größten Klassen waren „Regulatoren der Genexpression“ mit 59 Proteinen und „Regulatoren der RNA Polymerase II Transkription“ mit 43 Proteinen (Daten nicht gezeigt).

Die Identifikation der Poly(A)-Polymerase ist überraschend. Eine Assoziation mit CFII während der prä-mRNA-Spaltung ist bisher nicht beschrieben. Die Poly(A)-Polymerase ist essentiell für die Spaltung. Dass sie nur in CFII Δ N769, der Spaltungs-aktiven Variante, identifiziert werden konnte, erklärt die beobachtete Spaltung der prä-mRNA (siehe 2.1.1.1).

Tabelle 10 Übersicht der in der aktiven CFII-Variante angereicherten Proteine. Dargestellt sind die logarithmischen iBAQ-Werte für Pufferkontrolle, aktive (CFII Δ N769) und inaktive (CFII Δ N1123) Variante. Proteine, die nicht identifiziert werden konnten, sind gekennzeichnet (n.i.).

Name	UniprotID	Puffer [log ₂ (iBAQ)]	CFII Δ N769 [log ₂ (iBAQ)]	CFII Δ N1123 [log ₂ (iBAQ)]
Poly(A)-Faktoren				
PCF11	O94913	18,07	34,33	31,22
PAPOA	P51003	n.i.	19,35	n.i.
RNA-Helikasen				
DDX55	Q8NHQ9	n.i.	22,25	n.i.
DDX19A	Q9NUU7	15,41	17,31	n.i.
AQR	O60306	13	16,58	n.i.

Neben dem Vergleich der beiden Pcf11-Varianten untereinander geben die *Pulldown*-Experimente auch limitierte Informationen zu den allgemeinen Interaktionspartnern von CFII. Da die Daten zum vollständigen CFII nicht auswertbar waren und in den CFII Δ -Varianten der N-Terminus von Pcf11 fehlt, konnte keine Aussage über die über Protein-Protein-Interaktionen getroffen werden, die durch den N-Terminus von Pcf11 vermittelt werden.

Zur Identifizierung der Interaktionspartner von CFII wurden die iBAQ-Werte von CFII Δ N769 und CFII Δ N1123 gemittelt. Wurde ein Protein nur in einer dieser Varianten identifiziert, wurde der iBAQ-Wert unverändert übernommen. Diese allgemeinen CFII Daten wurden dann gegen die Pufferkontrolle abgeglichen. Insgesamt konnten für CFII 1978 Proteine, für die Pufferkontrolle 2289 Proteine mit einer Überschneidung von 1764 Proteinen identifiziert werden. Insgesamt

lagen 903 Proteine gegenüber dem Puffer mit mindestens Faktor 2 ($iBAQ_{CFII} / iBAQ_{Puffer}$) angereichert vor.

Angereicherte bekannte und putative Polyadenylierungsfaktoren sind in Tabelle 11 aufgeführt. Darüber hinaus konnten weitere Proteine und Proteinkomplexe identifiziert werden, welche in 3.3.6 zusammenfassend tabelliert sind.

Tabelle 11 Übersicht der bekannten und identifizierten Polyadenylierungsfaktoren in Kontroll- und gemitteltem CFII-Pulldown. Die Proteine sind in ihre Subkomplexe aufgeteilt. Proteine, die nicht identifiziert werden konnten, sind gekennzeichnet (n.i.). Der Anreicherungsfaktor (AF) berechnete sich aus den iBAQ-Werten (CFII/Puffer). Wurde das Protein in der Pufferkontrolle nicht identifiziert, konnte kein Anreicherungsfaktor berechnet werden (n.b.).

Name	UniprotID	Puffer [log ₂ (iBAQ)]	CFII [log ₂ (iBAQ)]	AF
CPSF-Komplex				
CPSF30	O95639	15,67	28,45	7033
CPSF73	Q9UKF6	16,16	30,62	22537
CPSF100	Q9P2I0	17,9	30,69	7082
CPSF160	Q10570	20,89	29,64	431
Fip1	Q6UN15	19,92	30,42	1448
WDR33	Q9C0J8	21,21	29,52	317
Symplekin	Q92797	16,22	30,36	18054
CFI-Komplex				
CFI25	O43809	21,55	25,57	16
CFI59	Q8N684	16,53	23,76	150
CFI68	Q16630	18,46	25,7	151
CstF-Komplex				
CstF50	Q05048	20,95	31,5	1499
CstF64	P33240	21,48	31,02	744
CstF64tau	Q9H0L4	14,86	30,65	56658
CstF77	Q12996	20,95	30,91	996
CFII-Komplex				
Clp1	Q92989	21,28	34,55	9878
Pcf11	O94913	18,15	33,49	41476
Poly(A)				
PAPα	P51003	n.i.	19,35	n.b
PABPC1	P11940	18,91	25,23	80
PABPN1	Q86U42	21,33	25,37	16
CPSF assoziiert				
Rbbp66	Q7Z6E9	15,29	23,06	218
Ssu72	Q9NP77	n.i.	27,3	n.b

NPM	P06748	23,26	29,11	58
XRN2-Komplex				
XRN2	Q9H0D6	21,74	29,17	172
NONO	Q15233	21,29	26,45	36
CARF	Q9NXV6	n.i.	28,85	n.b
SFPQ	P23246	19,76	25,92	72
tRNA Spleißkomplex				
SEN2	Q8NCE0	15,27	24,97	832
SEN15	Q8WW01	n.i.	25,7	n.b
SEN34	Q9BSV6	18,83	25,51	103
SEN54	Q7Z6J9	18,23	26,07	229

3.3.2 Interaktom von CstF

Für die Analyse des Interaktoms des CstF-Komplexes wurde ein Komplex benutzt, der aus den Untereinheiten CstF50, CstF77, welche beide doppelt vorlagen, sowie je einer Kopie von CstF64 und seinem Paralog CstF64tau, welches ein FLAG-tag besaß, bestand. Der gereinigte Proteinkomplex wurde von Moritz Schmidt zur Verfügung gestellt. Dieses Experiment wurde parallel zu dem *Pulldown* der N-terminal verkürzten Deletionsvarianten von CFII (siehe 2.3.1.2) durchgeführt, sodass dieselben Kontrollen verwendet wurden. Für den *Pulldown* wurde in technischen Duplikaten 34 pmol Protein mit 160 μ L nuklearem Extrakt inkubiert. Das analytische SDS-Polyacrylamidgel zeigte neben den Köderproteinen weitere angereicherte Proteine, während kaum Proteine in der Pufferkontrolle zu beobachten waren (Abbildung 3.43 A).

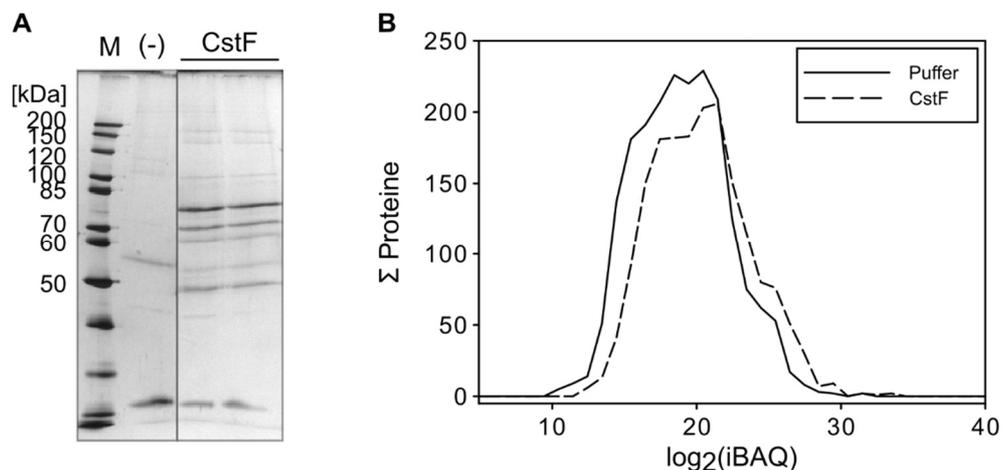


Abbildung 3.43 Interaktomanalyse des CstF-Komplexes. A) Analytisches, silbergefärbtes SDS-Polyacrylamidgel der Elutionsfraktionen der Pufferkontrolle (-) sowie der technischen Duplikate von CstF (CstF). B) Verteilungskurve der identifizierten Proteine. Die Summe der Proteine wurde nach Formel (5) bestimmt. Die Anreicherung von Proteinen ist sowohl durch eine gesamte Rechtsverschiebung der Kurve sowie durch die rechtsseitige Schulter der Intensität zu beobachten.

Von den Elutionsfraktionen wurden 75 % verdaut und massenspektrometrisch analysiert. Es konnten 1781 Proteine identifiziert werden, von denen 1187 gegenüber der Kontrolle um mindestens Faktor 2 ($i\text{BAQ}_{\text{CstF}} / i\text{BAQ}_{\text{Puffer}}$) angereichert waren. Dieses wird auch in der Verteilungskurve deutlich, wo eine klare Verschiebung nach rechts zu beobachten ist (Abbildung 3.43 B). Die Detailanalyse zeigte 53 Proteine, die der mRNA-3'-Prozessierung zugeordnet werden. Neben den bekannten Polyadenylierungsfaktoren, welche in Tabelle 12 dargestellt sind, konnten 33 weitere Proteine identifiziert werden, welche in 3.3.7 aufgeführt sind. Neben CPSF, CFI und CFII konnten unter anderem SR-reiche Spleiß-Faktoren sowie der THO-Komplex identifiziert werden. Nicht identifiziert werden konnte die Poly(A)-Polymerase.

Tabelle 12 Identifizierte Polyadenylierungsfaktoren in dem CstF-Pulldown. Die bekannten Proteine sind nach ihren Subkomplexen gruppiert. Nicht identifizierte Proteine sind gekennzeichnet (n.i.). Der Anreicherungsfaktor (AF) wurde durch Division der iBAQ-Werte berechnet. Für in der Kontrolle nicht identifizierte Proteine konnte kein Anreicherungsfaktor berechnet werden (n.b.)

Name	UniprotID	Puffer [log ₂ (iBAQ)]	CstF [log ₂ (iBAQ)]	AF
CPSF-Komplex				
CPSF30	O95639	15,67	28,75	8659
CPSF73	Q9UKF6	16,16	29,4	9675
CPSF100	Q9P2I0	17,9	29,6	3327
CPSF160	Q10570	20,89	29,23	324
Fip1	Q6UN15	19,92	29,64	843
WDR33	Q9C0J8	21,21	29,46	304
Symplekin	Q92797	16,22	28,9	6562
CPSF assoziiert				
Rbbp6	Q7Z6E9	15,29	22,74	175
Ssu72	Q9NP77	n.i.	26,05	n.b.
NPM	P06748	23,26	29,15	59
CFI-Komplex				
CFI25	O43809	21,55	25,26	13
CFI59	Q8N684	16,53	23,99	176
CFI68	Q16630	18,46	25,36	119
CstF-Komplex				
CstF50	Q05048	20,95	33,37	5480
CstF64	P33240	21,48	31,62	1128
CstF64tau	Q9H0L4	14,86	33,12	313912
CstF77	Q12996	20,95	32,93	4040
CFII-Komplex				
Clp1	Q92989	21,28	29,94	405

Pcf11	O94913	18,15	25,98	228
		Poly(A)		
PAPα	P51003	n.i.	n.i.	n.b.
PABPC1	P11940	18,91	25,82	120
PABPN1	Q86U42	21,33	25,3	16

3.3.3 Interaktom von CFI

Der in 2.1.4 beschriebene CFI-Komplex, welcher an der Untereinheit CFI68 einen Strep-tag besaß, wurde zu Analyse des Interaktoms verwendet. Da in den vorhergehenden Analysen von CstF und CFII der CFI-Komplex im Vergleich zu den anderen bekannten, an der Polyadenylierung beteiligten Komplexen unterrepräsentiert vorlag, wurde eine erhöhte Menge Köderprotein in Relation zum Kernextrakt eingesetzt. Für die technischen Replikate wurden 128 pmol CFI-Komplex mit 160 μ l nuklearem Extrakt inkubiert. Das analytische SDS-Polyacrylamidgel zeigte neben den Köderproteinen noch weitere Proteinbanden, während in der Negativkontrolle eine deutlich geringere Anreicherung von Proteinen detektiert werden konnte (Abbildung 3.44 A). Zur massenspektrometrischen Analyse wurden 75 % der Elutionsfraktion verdaut und analysiert. Es konnten überraschend nur 1255 Proteine in dem *Pulldown* von CFI gegen 2095 Proteinen in der Pufferkontrolle identifiziert werden. Nur 29 Proteine waren gegenüber der Kontrolle angereichert ($\log_2(\text{iBAQ})_{\text{CFI}} > \log_2(\text{iBAQ})_{\text{Puffer}}$). Dieses Ungleichgewicht zeigt sich auch in der Verteilungskurve der Analysen (Abbildung 3.44 B).

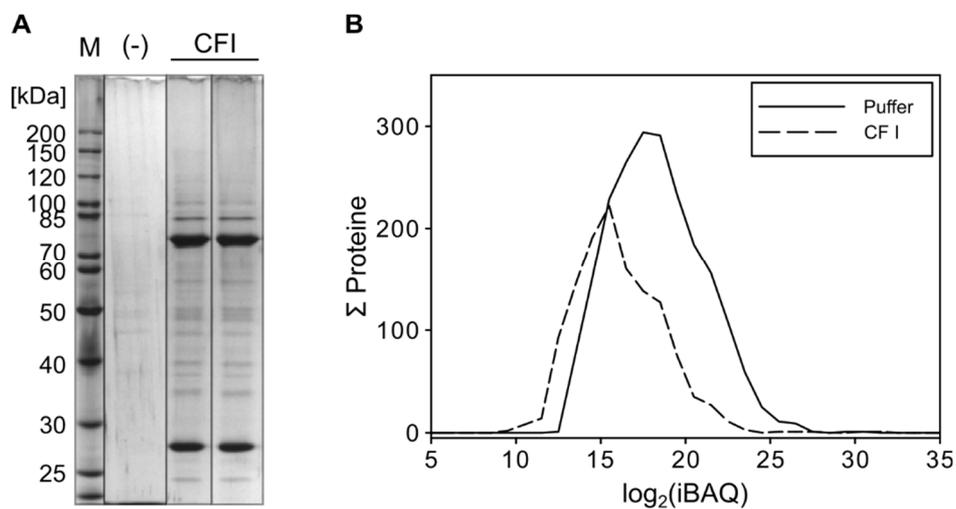


Abbildung 3.44 Interaktomanalyse des CFI-Komplexes. A) Analytisches, silbergefärbtes SDS-Polyacrylamidgel der *Pulldown*-Experimente des CFI-Komplexes sowie der Pufferkontrolle. B) Verteilungskurve der identifizierten Proteine. Die Summe der Proteine wurde nach Formel (5) bestimmt. Deutlich wird, dass in der Pufferkontrolle deutlich mehr Proteine mit höherer Intensität identifiziert wurden.

Von den 29 Proteinen, die ein Signal über dem Hintergrund besaßen, konnten 10 der mRNA-Prozessierung zugeordnet werden, welche in Tabelle 13 tabelliert sind. Von den bekannten

Polyadenylierungsfaktoren konnte neben den Köderproteinen noch CstF50 detektiert werden. Darüber hinaus konnte sieben mit dem Spleißosom-assoziierte Proteine identifiziert werden.

Tabelle 13 Identifizierte Interaktionspartner des CFI-Komplexes im *Pulldown*-Experiment.

Identifizierte Proteine wurden in Polyadenylierungsfaktoren, Spleiß-Faktoren sowie sonstige Proteine klassifiziert.

Name	UniprotID	Puffer [log ₂ (iBAQ)]	CFI [log ₂ (iBAQ)]
Poly(A) Faktoren			
CFI25	O43809	21,84	31,18
CFI68	Q16630	22,27	30,96
CstF50	Q05048	20,59	21,66
Spleiß-Faktoren			
YTDC1	Q96MU7	16,27	16,68
CLASR	Q8N2M8	15,52	16,66
SRSF8	Q9BRL6	15,37	19,01
TRA2A	Q13595	18,3	19,09
TRA2B	P62995	20,32	21,11
SRSF1	Q07955	22,24	22,29
Sonstige			
RBM26	Q5T8P6	19,86	19,93

3.3.4 Interaktom des Endonuklease-Komplexes

Für die Analyse des Interaktoms des Endonuklease-Komplexes wurde der in 2.1.3.2 beschriebene verwendet. Der *Pulldown* erfolgte über den FLAG-*tag* an CstF64. In technischen Duplikaten wurden 34 pmol Köderkomplex mit 160 µl nuklearem Extrakt inkubiert. Im analytischen SDS-Polyacrylamidgel konnten kaum spezifisch angereicherten Proteine identifiziert werden, die Pufferkontrolle hingegen enthielt, bis auf die während der Reinigung eluierten FLAG-Antikörper, keine Proteine (Abbildung 3.45 A). Von den Elutionsfraktionen wurden 75 % verdaut und massenspektrometrisch analysiert. Es konnten 1332 Proteine, von denen 1020 Proteine mindestens Faktor 2 ($iBAQ_{EN}/iBAQ_{Puffer}$) angereichert sind, in den Experimenten gegenüber 1117 Proteinen in der Kontrolle identifiziert werden. Dieses wird auch in der Verteilungskurve deutlich, wo eine klare Anreicherung von Proteinen zu beobachten ist (Abbildung 3.45 B). Hiervon konnten 35 Proteinen den bekannten Polyadenylierungsfaktoren (Tabelle 14) und 30 weitere durch eine Klassifizierung der mRNA 3'-Prozessierung zugeordnet werden (siehe 3.3.7). Unter den Polyadenylierungsfaktoren sind vor allem die bekannten CPSF- und CstF-Untereinheiten, sowie die CPSF-assoziierte Phosphatase SSU72, angereicht. Gegenüber der Pufferkontrolle ist zudem der PP1-Phosphatasekomplex angereicht, jedoch mit deutlich reduzierter Abundanz als die CPSF-Untereinheiten. Nur eine geringe Anreicherung konnte für die Komplexe CFI und CFII

beobachtet werden. Darüber hinaus wurden drei Poly(A)-bindende Proteine, Untereinheiten der RNA Polymerase II sowie der PAF-Komplex angereichert.

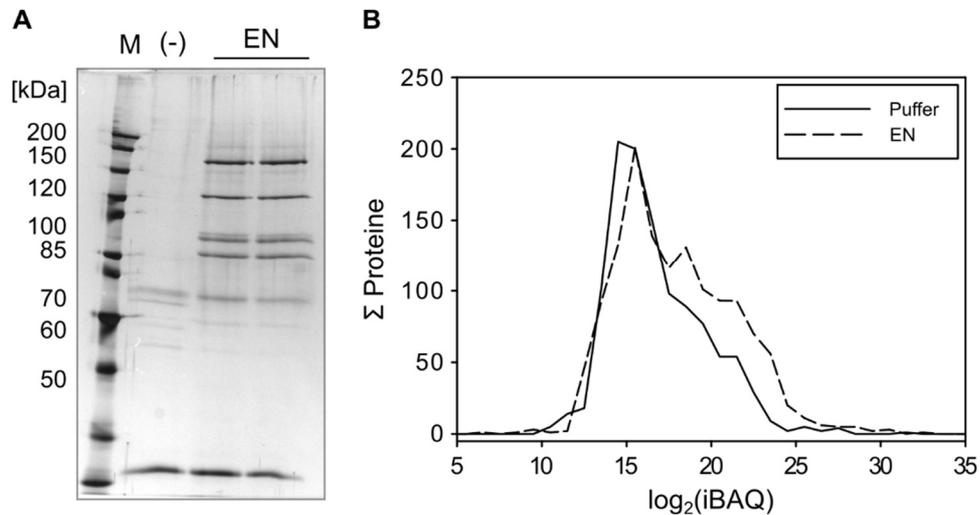


Abbildung 3.45 Interaktomanalyse des Endonuklease-Komplexes. A) Analytisches, silbergefärbte SDS-Polyacrylamidgel der Elutionsfraktionen der Duplikate des Endonuklease-Komplexes (EN) und der Pufferkontrolle (-). B) Verteilungskurve der identifizierten Proteine. Die Summe der Proteine wurde nach Formel (5) bestimmt. Eine Anreicherung von Proteinen im Endonuklease-Komplex (EN) ist deutlich durch die rechtsseitige Schulter der Verteilungskurve zu beobachten.

In den 30 weiteren angereicherten Proteinen, die der mRNA 3'-Prozessierung zugeordnet wurden, sind vor allem SR-reiche Spleißfaktoren sowie der THO-Komplex zu finden.

Tabelle 14 Übersicht der identifizierten Polyadenylierungsfaktoren in dem *Pulldown-Experiment des Endonuklease-Komplexes.* Die Intensitäten der Pufferkontrolle (Puffer) sowie des Endonuklease-Komplexes (EN) sind dargestellt. Nicht identifizierte Proteine sind gekennzeichnet (n.i.). Der Anreicherungsfaktor (AF) wurde durch Division der iBAQ-Werte berechnet. Für in der Kontrolle nicht identifizierte Proteine konnte kein Anreicherungsfaktor berechnet werden (n.b.).

Name	UniprotID	Puffer [$\log_2(\text{iBAQ})$]	EN [$\log_2(\text{iBAQ})$]	AF
CPSF-Komplex				
CPSF30	O95639	15,67	25,13	704
CPSF73	Q9UKF6	16,16	30,61	22381
CPSF100	Q9P2I0	17,9	30,6	6654
CPSF160	Q10570	15,83	25,9	1075
Fip1	Q6UN15	16,55	26,47	969
WDR33	Q9C0J8	17,86	25,98	278
Symplekin	Q92797	16,22	29,88	12944

CPSF assoziiert				
Rbbp66	Q7Z6E9	15,29	18,14	7
Ssu72	Q9NP77	n.i.	28,01	n.b.
NPM	P06748	23,21	24,84	3
CFI-Komplex				
CFI25	O43809	15,64	20,13	22
CFI59	Q8N684	14,03	19,78	54
CFI68	Q16630	15,9	18,75	7
CstF-Komplex				
CstF50	Q05048	18,22	28,18	996
CstF64	P33240	20,23	30,59	1314
CstF64tau	Q9H0L4	n.i.	n.i.	n.b.
CstF77	Q12996	18,32	27,03	419
CFII-Komplex				
Clp1	Q92989	14,41	19,42	32
Pcf11	O94913	14,96	18,04	8
Poly(A)				
PAPα	P51003	n.i.	n.i.	n.b.
PABPC1	P11940	18,7	20,86	4
PABPN1	Q86U42	21,33	22,45	2
PP1α-Komplex				
PP1α	P62136	19,46	23,55	17
WDR82	Q6UXN9	15,15	23,21	267
PNUTS	Q96QC0	16,49	21,72	38
TOX4	O94842	17,42	24,42	128

3.3.5 Basale Faktoren der Polyadenylierung

Zur Identifizierung der basalen Faktoren der Polyadenylierung wurden die Daten der Pulldowns der CFII-Varianten, des CstF- und des Endonuklease-Komplexes verwendet. Die Daten des CFI-, mPSF- sowie vollständigen CFII-Komplexes waren für eine Detailanalyse nicht geeignet.

Hervorgehoben wurden alle Proteine, die in allen drei Datensätzen identifiziert wurden und gegenüber ihrer Kontrollreaktion mindestens Faktor 2 ($i\text{BAQ}_{\text{Komplex}} / i\text{BAQ}_{\text{Puffer}}$) angereichert waren. Dieses traf auf 522 Proteine zu. Zur näheren Klassifizierung dieser Proteine wurden diese, wie beschrieben, in ihre „GO“-Begriffe eingeordnet. 270 dieser Proteine konnten RNA-Metabolismus („*RNA metabolic process*“) zugeordnet werden. Eine weitere Einordnung zeigte, dass 43 Proteine der mRNA-3'-Prozessierung zuzuordnen sind (Tabelle 15).

Tabelle 15 Klassifizierung der in allen Datensätzen angereicherten Proteine, die dem RNA-Metabolismus zugeordnet werden konnten. Dargestellt sind ausgewählte Klassifizierungen. Als weitere Einordnung ist als Referenz das humane Proteom angegeben.

„GO“-Klassifizierung	Referenz (Homo sapiens)	Identifizierte Proteine
mRNA 3'-Prozessierung	82	43
mRNA Polyadenylierung	30	16
mRNA Spleißen (via Spleißosom)	291	127
rRNA Prozessierung	206	49
RNA Polymerase II Termination	34	20

Die Gruppe der mRNA 3'-Prozessierung kann in weitere Komplexe getrennt werden. Eine Übersicht der Komplexe bzw. der Kategorien der Identifizierungen ist vereinfacht in Tabelle 16 dargestellt. Die korrelierenden Intensitäten sind in den jeweiligen *Pulldown*-Experimenten zu finden.

Tabelle 16 Vereinfachte Darstellung der in allen *Pulldowns* identifizierte Proteine

Komplex / Kategorie	Angereicherte Proteine
CPSF-Komplex	CPSF30; CPSF73; CPSF100; CPSF160; Fip1; Symplekin; WDR33; Ssu72
CstF-Komplex	CstF50; CstF64; CstF77
CFI-Komplex	CFI25; CFI59; CFI68
CFII-Komplex	Clp1; Pcf11
Poly(A)-bindend	PABPC1; PABPN1
PAF-Komplex	CDC73; LEO1; PAF1
SR-reiche	SRSF1; SRSF2; SRSF3; SRSF5; SRSF7;
Spleißfaktoren	SRSF9; RNPS1; SRRM1
THO-Komplex	THOC1; THOC2; THOC3; THOC4; THOC5; THOC6; THOC7; DDX39A; DDX39B
Spleißfaktoren	U2AF1; U2AF2
Cap-bindend	NCBP1
Sonstige	RPRD1B; CHTOP

3.3.6 Weitere identifizierte Proteine

In den *pulldown*-Experimenten des CFII-, CstF- und Endonuklease-Komplex konnten neben den beschriebenen basalen Polyadenylierungsfaktoren weitere Proteinkomplexe identifiziert werden, die durch die *GO terms* der mRNA-3'-Prozessierung zugeordnet werden können. Diese sind in Tabelle 17 dargestellt.

Tabelle 17 Weitere identifizierte Proteine in den *Pulldown*-Experimenten. Dargestellt sind die $\log_2(\text{iBAQ})$ -Werte der jeweiligen Proteine, sowie, in Klammern, der Anreicherungsfaktor

(iBAQ_{Pulldown}/iBAQ_{Kontrolle}). Nicht identifizierte Proteine (n.i.) und nicht berechenbare Anreicherungsfaktoren (n.b.) sind gekennzeichnet. Die identifizierten Proteine sind nach Komplexen eingeordnet.

Name	UniprotID	CFII	CstF	EN
XRN2-Komplex				
XRN2	Q9H0D6	29,17 (172)	21,35 (1)	17,87 (3)
NONO	Q15233	26,45 (36)	27,01 (53)	23,58 (6)
CARF	Q9NXV6	28,85 (n.b)	19,33 (n.b)	n.i. (n.b)
SFPQ	P23246	25,92 (72)	26,8 (132)	22,55 (7)
PP1-Komplex				
PP1α	P62136	24,09 (14)	23,95 (13)	23,55 (17)
PP1β	P62140	20,84 (12)	21,99 (27)	20,43 (48)
PP1γ	P36873	20,26 (3)	20,75 (4)	19,99 (7)
WDR82	Q6UXN9	20,46 (2)	21,12 (3)	23,21 (267)
PNUTS	Q96QC0	18,67 (1)	20,14 (3)	21,72 (38)
TOX4	O94842	19,21 (3)	21,16 (13)	24,42 (128)
RNA Polymerase II				
Rpb1	P24928	19,25 (155)	18,42 (87)	17,15 (36)
Rpb2	P30876	20,07 (50)	18,35 (15)	18,04 (12)
PAF-Komplex				
PAF1	Q8N7H5	17,67 (38)	19,08 (101)	19,7 (155)
CDC73	Q6P1J9	18,58 (3)	18,32 (2)	20,61 (11)
CTR9	Q6PD62	16,39 (0)	17,25 (0)	18,81 (n.b)
LEO1	Q8WVC0	15,35 (2)	16,7 (6)	17,01 (7)
WDR61	Q9GZS3	18,94 (1)	19,48 (1)	19,54 (76)
RNA Pol II CTD-modifizierend				
RPRD1B	Q9NQG5	22,21 (99)	19,95 (21)	19,94 (21)
CDK9	P50750	18,24 (n.b)	18,53 (n.b)	13,51 (n.b)
RPR1B	Q9NQG5	22,21 (99)	19,95 (21)	19,94 (21)
RPRD2	Q5VT52	17,95 (0)	20,44 (2)	19,67 (14)
TREX/THO-Komplex				
THOC1	Q96FV9	21,9 (8)	22,3 (11)	20,86 (4)
THOC2	Q8NI27	21,65 (27)	22,06 (36)	18,15 (2)
THOC3	Q96J01	22,3 (142)	22,29 (141)	18,07 (8)
THOC4	Q86V81	26,19 (38)	25,83 (29)	23,2 (5)
THOC5	Q13769	21,27 (29)	22,02 (49)	19,08 (6)
THOC6	Q86W42	22,83 (71)	22,68 (64)	20,19 (11)
THOC7	Q6I9Y2	21,27 (5)	22,67 (14)	20,95 (4)
DX39A	O00148	25,33 (n.b)	25,22 (n.b)	18,14 (10)
DX39B	Q13838	23,13 (33)	24 (61)	21,12 (8)
CHTOP	Q9Y3Y2	23,44 (17)	23,06 (13)	21,87 (6)
SARNP	P82979	22,66 (0)	23,62 (0)	20,4 (4)
TREX assoziiert				
ZC11A	O75152	18,37 (1)	18,76 (2)	16,68 (3)
PDIP3	Q9BY77	23,27 (2)	23,01 (1)	21,85 (3)
NXF1	Q9UBU9	24,24 (362)	19,25 (11)	16,82 (7)
NXT1	Q9UKK6	23,19 (7)	n.i. (n.b)	n.i. (n.b)
Cap-Binde-Komplex				
NCBP1	Q09161	22,42 (13)	22,76 (16)	16,62 (n.b)
NCBP2	P52298	18,29 (2)	22,46 (34)	n.i. (n.b)
Spleißfaktoren				
SRSF1	Q07955	26 (28)	26,56 (41)	22,53 (3)

SRSF2	Q01130	23,37 (10)	25,07 (33)	21,89 (8)
SRSF3	P84103	26,26 (2)	26,92 (3)	21,95 (3)
SRSF5	Q13243	22,17 (7)	23,14 (14)	20,05 (4)
SRSF6	Q13247	22,71 (1)	24,24 (3)	21,74 (5)
SRSF7	Q16629	26,89 (49)	27,01 (53)	23,21 (4)
SRSF8	Q9BRL6	21,2 (0)	21,85 (0)	19,22 (n.b)
SRSF9	Q13242	22,66 (34)	23,51 (61)	17,38 (4)
SRSF10	O75494	24,28 (86)	24,29 (86)	20,64 (7)
SRSF11	Q05519	19,73 (0)	21,07 (1)	18,61 (15)
U2AF1	Q01081	22,06 (46)	23,64 (138)	20,54 (36)
U2AF2	P26368	22,31 (83)	22,06 (70)	19,23 (10)
EJC/ASAP-Komplex				
eIF4A4	P38919	23,71 (3)	22,64 (1)	19,6 (7)
MLN51	O15234	16,61 (n.b)	17,2 (n.b)	n.i. (n.b)
Y14	Q9Y5S9	24,37 (1)	23,29 (0)	21,13 (5)
Magoh 2	Q96A72	19,38 (3)	21,85 (19)	n.i. (n.b)
RNPS1	Q15287	24,86 (27)	25,16 (33)	22,63 (6)
SAP18	O00422	23,52 (16)	22,65 (9)	21,33 (5)
Acinus	Q9UKV3	23,7 (58)	23,28 (43)	20,88 (10)

4 Diskussion

Ziel dieser Arbeit war die strukturelle Charakterisierung der an der mRNA-3'-Prozessierung beteiligten Proteinkomplexe. Hierzu wurden diese Proteinkomplexe rekombinant hergestellt, auf ihre Aktivität hin untersucht und anschließend durch chemisches *cross-linking* und Massenspektrometrie strukturell charakterisiert. Diese Strukturdaten wurden anschließend zur Protein- und Komplexstrukturvorhersage genutzt, wobei verschiedene bioinformatische Methoden zum Einsatz kamen. Darüber hinaus wurden die rekombinant produzierten Proteine verwendet, um in HeLa-Kernextrakt Interaktionspartner zu identifizieren.

4.1 Proteinexpression und -reinigung und Aktivitätsanalyse

Die Expression und -reinigung der hier beschriebenen Proteinkomplexe der mRNA-3'-Prozessierung kann als erfolgreich angesehen werden. Für CFI, CFII und CstF war bereits ein Reinigungsprotokoll etabliert [27], [163]. Für die CPSF-Subkomplexe mPSF- und den Endonuklease-Komplex konnte eine zweistufige Reinigungsstrategie etabliert werden, die sich von der bisherigen Reinigungsstrategie unterschied [26]. Für alle Proteine wurde eine Anionenaustauschchromatographie durchgeführt, um kontaminierende Nukleinsäuren zu entfernen, die unter Umständen die Aktivitätsanalysen hätten beeinflussen können. Alle Puffer, die für die Proteinreinigungen mit anschließendem *cross-linking* verwendet wurden, enthielten Triethanolamin statt TRIS, welches hier üblicherweise verwendet wird. Dieses war aufgrund des freienamins in TRIS zwingend notwendig. Da diese beiden Substanzen chemisch sehr ähnlich sind, konnten alle bisherigen Experimente hiermit reproduziert werden.

Zur Überprüfung der Reinigung wurden von allen Komplexen Aktivitätsanalysen durchgeführt. Hierzu wurde die Bindung von spezifischen RNA-Substraten oder der L3-RNA gemessen. Diese RNA stammt aus dem humanen Adenovirus-2 und besitzt ein sehr starkes Polyadenylierungssignal, sodass sich dieses für die Analyse der Spaltung sehr gut eignet [176]. Diese RNA wurde auch in den Rekonstitutionsexperimenten mit fraktioniertem Extrakt eingesetzt. Hierzu wurde HeLa-Kernextrakt über eine Fällung und anschließende Größenausschlusschromatographie fraktioniert. Es konnte gezeigt werden, dass die Aktivität der resultierenden Fraktion von der Zugabe von CstF, CFII sowie der Poly(A)-Polymerase abhängig ist (Daten von Florian Kluge).

Alle gereinigten Subkomplexe der Polyadenylierungsreaktion wurden erfolgreich chemisch vernetzt und massenspektrometrisch analysiert. Hierzu wurde der in *Iacobucci und Götze et al.* beschriebene Arbeitsablauf angewendet [128]. Die Proteine wurden mit dem MS/MS-spaltbaren *cross-linker* DSBU vernetzt. Zusätzlich wurde für den CFI-Komplex der *cross-linker* LC-SDA verwendet, welcher ebenfalls spaltbar ist. Das richtige Protein zu *cross-linker* Verhältnis musste für jeden Komplex empirisch bestimmt werden, da die Reaktivität nicht theoretisch bestimmt werden kann. Ursache hierfür ist, dass sich freie Amine sowie mit reduzierter Reaktivität freie

Hydroxylgruppen in räumlicher Nähe befinden müssen, welches von der quartären Struktur der Komplexe bestimmt wird. Darüber hinaus können verunreinigende Proteine, welche bei der Reinigung nicht vollständig abgetrennt wurden, mit dem *cross-linker* reagieren und so die benötigte Stoffmenge beeinflussen.

Zur Analyse der Reaktionsprodukte wurden diese, abhängig vom Komplex, entweder mittels SDS-PAGE isoliert oder direkt in der Reaktionslösung tryptisch verdaut. Der Vorteil der Proteolyse in Lösung liegt in einer besseren Identifikation von Peptiden bei gleicher Ausgangsmenge an Zielprotein, jedoch können keine isolierten Produkte analysiert werden.

Im Folgenden werden die Reinigungen der Komplexe sowie die Ergebnisse des chemische *cross-linkings* näher diskutiert.

4.1.1 Darstellung und Aktivitätsanalyse des CFII Komplexes

Die Reinigung des CFII-Komplexes war sehr erfolgreich (Abbildung 3.1). Das Reinigungsprotokoll war bereits etabliert und die Modifizierung der Puffersubstanz von TRIS zu Triethanolamin hatte keinen Einfluss auf die Stabilität, Aktivität und Reinigung des Komplexes. Pcf11 neigt aufgrund seiner Größe zu proteolytischem Abbau, dieses konnte aber durch Verdünnen statt Dialysieren zur Reduzierung des Salzgehaltes vor der Anionenaustauschchromatographie weitgehend verhindert werden. In hoher Konzentration waren Abbaubanden in dem analytischen Polyacrylamidgel zu identifizieren, diese sind jedoch gegenüber dem intakten Protein zu vernachlässigen. Mit dem gereinigtem Proteinkomplex wurden verschiedene Aktivitätsanalysen durchgeführt, die alle den publizierten Daten entsprachen [27]. So konnte ein sub-nanomolarer Affinitätswert gegenüber der Substrat-RNA gemessen werden und das gereinigte war Protein in der Lage, fraktionierten Extrakt zu komplementieren.

4.1.1.1 *Cross-linking* von CFII

Der CFII-Komplex, welcher sich aus Clp1 und Pcf11 zusammensetzt und als Heterodimer vorliegt, war bisher nur partiell strukturell charakterisiert. Homologe der Polynukleotidkinase Clp1 aus Hefe und *C. elegans* konnten röntgenkristallographisch aufgeklärt werden, wobei hier die konservierte Interaktionsregion von Pcf11 kokristallisiert werden konnte. Zur näheren strukturellen Charakterisierung des Komplexes wurde dieser mittels *cross-linking* näher untersucht.

Die Titration des *cross-linkers* zeigte die Bildung dreier hochmolekularer Produkte. Hierbei lag das zweite Produkt in der präparativen Selektion der Produkte am abundantesten vor. Da zum Zeitpunkt des Experimentes noch keine gesicherten Informationen zu der Stöchiometrie des Komplexes und zu möglichen verschiedenen Konformationen vorlagen, wurden diese drei Produkte getrennt analysiert. In der quantitativen Analyse der enthaltenen Peptide in diesen Produkten zeigte sich jedoch, dass alle drei Produkte einem 1:1-Komplex, vermutlich einem

Heterodimer, entsprechen. Parallel durchgeführte Experimente von Peter Schäfer zeigten, dass der CFII-Komplex tatsächlich nur als Heterodimer vorliegt [27]. In der qualitativen Analyse der *cross-link*-Produkte der drei Reaktionsprodukte konnten keine signifikanten Unterschiede in der Verteilung festgestellt werden. Das Vorkommen dieser drei Produkte spiegelt also nicht das Vorkommen verschiedener Konformationen wider, sondern stellt den Grad der chemischen Modifikation durch den *cross-linker* dar. Dies deckt sich auch mit der Beobachtung, dass bei sehr geringen *cross-linker* Konzentrationen nur das erste Produkt zu beobachten ist. Erst mit steigender Konzentration bilden sich auch die beiden anderen Produkte. Sollten verschiedene Konformationen vorliegen, sollten diese auch schon bei geringen Konzentrationen fixiert werden. Da außerdem das zweite Reaktionsprodukt in der präparativen Reaktion am abundantesten vorlag, müsste dieses Produkt unter limitierenden *cross-link*-Konzentrationen als erstes auftreten. Der Unterschied liegt also in dem Grad der Vernetzung und der damit eingehenden Fixierung von Tertiär- und Quartärstrukturen, sodass die Produkte langsamer im Polyacrylamidgel migrieren.

Insgesamt konnten 152 einzigartige *cross-links* identifiziert werden. Wie zu erwarten, konnte eine Interaktion zwischen der beschriebenen Interaktionsregion (Pcf11₁₃₇₇₋₁₄₀₁) und Clp1 beobachtet werden. Am auffälligsten war jedoch der Bereich Pcf11₅₀₀₋₇₇₀. In diesem lagen die meisten intermolekularen *cross-links* und zusätzlich ist dieser Bereich auch sehr stark mit sich selbst und der Interaktionsregion (Pcf11₁₃₇₇₋₁₄₀₁) vernetzt. Hieraus ist abzuleiten, dass es sich hier um eine neue Domäne von Pcf11 handelt, welche eine definierte räumliche Struktur besitzt und an der Interaktion mit Clp1 beteiligt ist. Da diese jedoch nicht in dem durchgeführten limitierten tryptischen Verdau als stabiler Interaktionspartner identifiziert werden konnte, scheint diese Domäne nur eine unterstützende Interaktion mit Clp1 einzugehen, die Hauptinteraktionsfläche stellt jedoch weiterhin die schon beschriebene Interaktionsregion dar.

Auffällig in der Analyse der *cross-links* ist zudem die Region Pcf11₁₋₅₀₀. Hier sind nur wenige und vor allem, auf die Primärsequenz bezogen, sehr kurze *cross-links* identifiziert worden. Das Fehlen von weit entfernten Verknüpfungen kann auf eine ungeordnete Region hindeuten. Die zweite auffällige Region sind die FEGP-Wiederholungen (Pcf11₇₇₀₋₁₁₂₃). Hier konnten keine *cross-links* identifiziert werden, was jedoch auf den geringen Anteil reaktiver Aminosäuren zurückzuführen ist. So sind in den 354 Aminosäuren, welche diesen Bereich bilden, nur ein Lysin, vier Threonine und elf Serine vorhanden. Zur besseren Analyse dieses Bereiches mittels *cross-linking* müssen *cross-linker* mit anderer Reaktivität genutzt werden. Da die Wiederholungen über sehr viele saure Aminosäuren verfügen, liegt die Wahl von Diazirinen nah. Diese photoaktivierbare reaktive Gruppe zeigt eine bevorzugte Reaktion mit Karbonsäuren [129]. Diese reaktive Gruppe ist z.B. in dem in dieser Arbeit verwendeten LC-SDA *cross-linker* vorhanden. Aufgrund seiner Heterobifunktionalität muss jedoch ein freies Amin in räumlicher Nähe sein, was in den FEGP-Wiederholungen nicht zwingend gegeben ist. Vielversprechender ist daher die Verwendung von Photo-Leucin, in der eine Methylgruppe der Seitenkette durch ein Diazirin substituiert ist, da die

FEGP-Wiederholungen über 26 Leucine verfügen. Diese artifizielle Aminosäure kann *in vivo* ohne zusätzliche tRNA statt Leucin in der Proteinbiosynthese verwendet werden [177]. Eine anschließende Proteinreinigung und UV-induziertes *cross-linking* kann dann zur Strukturanalyse verwendet werden [178]. Da die beschriebenen Methoden jedoch für Säugerzellen entwickelt wurden, bedarf es einer Adaption des Protokolls an Insektenzellen. Darüber hinaus gibt es *cross-linker* mit Glyoxal als reaktive Gruppe, die eine Klasse von Arginin-spezifischen *cross-linkern* darstellen [179]. Es ist jedoch davon auszugehen, dass methylierte Arginine, wie sie in diesem Bereich vorkommen, keine oder nur eine reduzierte Reaktivität zeigen.

4.1.1.2 Interaktionsflächen zwischen Pcf11 und Clp1

Pcf11 bildet in Hefe einen stabilen Komplex mit Clp1 sowie den CstF-Homologen Rna14p und Rna15p. Die Interaktion zwischen Pcf11 und Clp1 ist dahingehend sehr gut charakterisiert, dass die betreffende Region von Pcf11 mit Clp1 kokristallisiert ist [61]. Dieser Bereich ist sehr hoch konserviert und kann auch in der humanen Sequenz mit sehr hoher Sequenzidentität (30 % Identität; 55 % Ähnlichkeit) identifiziert werden. Eine weitere Interaktion zwischen humanem Clp1 und Pcf11 konnte durch die *cross-link*-Experimente identifiziert werden. So ist die Region zwischen Aminosäure 500 bis 770 an mehreren Punkten mit Clp1 chemisch quervernetzt.

Zur weiteren Charakterisierung wurde der native CFII-Komplex mit einem Unterschuss an Trypsin verdaut. Hierbei werden vor allem flexible und leicht zugängliche Regionen gespalten. Regionen, die an stabilen Protein-Protein-Interaktionen beteiligt sind, weisen hingegen eine klare Struktur und häufig eine schlechte Lösungsmittelzugänglichkeit auf. Daher ist es durch eine limitierte Proteolyse möglich, diese stabilen Regionen zu identifizieren. Es zeigte sich, dass Clp1 während des Experimentes sehr stabil blieb. Deutlich wird dieses daran, dass es keine optische Veränderung des apparenten Molekulargewichtes sowie der Menge an Clp1 in dem analytischen SDS-Polyacrylamidgel gab. Dieses kann durch die globuläre Struktur von Clp1 erklärt werden, die sehr wenig Angriffsfläche für einen proteolytischen Abbau bietet. Pcf11 hingegen zeigt bereits nach 5 Minuten deutliche Proteolyseprodukte und nach 30 Minuten kann kein Vollängenprotein mehr identifiziert werden. Stattdessen zeigte sich die Bildung eines stabilen Fragmentes. Eine anschließende affinitätschromatographische Reinigung zeigte zudem, dass dieses Fragment weiterhin mit Clp1 assoziiert ist. Aufgrund des hohen Hintergrundes durch längere Pcf11-Varianten, die auch weiterhin mit Clp1 interagieren, war die Identifikation dieses Fragmentes kompliziert. Ein kombinierter Ansatz aus Isolation des Fragmentes, Edman-Abbau sowie massenspektrometrischer Analyse erlaubte schließlich die korrekte Identifikation. Es zeigte sich, dass dieses Fragment die Aminosäuren 1183 bis 1555 umfasst, dem nativen C-Terminus von Pcf11. In diesem liegt die zuvor beschriebene Interaktionsregion. Die in den *cross-link*-Experimenten identifizierte zweite Interaktionsregion Pcf11₅₀₀₋₇₇₀ konnte hier nicht als stabiles Fragment identifiziert werden. Eine mögliche Erklärung liegt in den FEGP-Wiederholungen, die

zwischen den beiden mit Clp1 interagierenden Bereich liegt. Diese enden mit einem Arginin, welches ein Substrat für Trypsin darstellt, und liegen aufgrund des hohen Glycins und Prolins Gehalt vermutlich unstrukturiert vor und sind somit leicht zugänglich für einen proteolytischen Abbau. Die Interaktion zwischen Clp1 und Pcf11₅₀₀₋₇₇₀ wird dann nicht mehr durch die Primärstruktur von Pcf11 stabilisiert. Es ist daher davon auszugehen, dass die Hauptenergie der Bindung zwischen Clp1 und Pcf11 von der hoch konservierten C-terminalen Region in Pcf11 vermittelt wird.

Zur weiteren Analyse der Interaktion zwischen Clp1 und Pcf11 wurde zudem ein N-Terminal verkürztes Pcf11-Fragment (Pcf11 Δ N1339) mit Clp1 in Insektenzellen koexprimiert. Der Beginn dieser Deletionsvariante lag direkt vor der ersten Zinkfingerdomäne, welche unmittelbar vor der Interaktionsregion liegt. Hier zeigte sich, dass die Interaktion zwischen diesem Fragment und Clp1 über die zwei Reinigungsschritte stabil blieb. Die Pcf11 Variante zeigte eine Doppelbande im SDS-Polyacrylamidgel, wobei der Ursprung dieses nicht geklärt werden konnte. Eine C-Terminal verkürzte Variante in Koexpression mit Clp1 wurde in dieser Arbeit nicht untersucht.

4.1.1.3 RNA-Bindung des CFII-Komplexes

Clp1 zeigt eine 5'-Polynukleotidkinaseaktivität, jedoch in monomerer Form keine RNA-Bindung [27]. Die RNA-Bindung in CFII muss also durch Pcf11 vermittelt werden. Beschrieben ist, dass die zweite Zinkfingerdomäne ausschließlich in Abwesenheit von Clp1 RNA binden kann [68]. Darüber hinaus konnte gezeigt werden, dass die C-Terminale-Interaktions-Domäne (CID) RNA binden kann [180]. Im Gegensatz zu den beiden Beobachtungen konnte in dieser Arbeit jedoch die RNA-Bindung einer N-Terminal stark verkürzten Pcf11 Variante im Komplex mit Clp1 (CFII Δ N1339) gemessen werden. Diese Aktivität komigriert mit den Hauptfraktionen, wobei jedoch eine leichte Verschiebung der RNA-Bindung zu beobachten ist (Abbildung 3.3 D). Möglicherweise ist die Proteinpräparation durch ein RNA-bindendes Protein verunreinigt. Eine falsch-positive Identifikation der Aktivität scheint aufgrund der nachfolgenden Experimente jedoch sehr unwahrscheinlich.

Zur näheren Untersuchung der hochaffinen RNA-Bindung wurden zwei weitere Deletionsvarianten charakterisiert. Diese unterscheiden sich vor allem durch das Vorhandensein (CFII Δ N769) bzw. das Fehlen (CFII Δ N1123) der FEGP-Wiederholungen. Für beide Proteine im Komplex mit Clp1 konnte die Bindung der Substrat-RNA nachgewiesen werden. Die IC₅₀-Werte für die beiden Varianten lagen bei ca. 10 nM. Dieses ist um den Faktor 16 schwächer als das native CFII. Aus der Abschätzung der freien Energie aus den K₅₀-Werten kann der energetische Unterschied zwischen den verschiedenen Messungen berechnet werden. Der Unterschied beträgt nur rund 7 kJ/mol. Dieses entspricht der Energie von ungefähr 1 bis 3 Wasserstoffbrückenbindungen, welche mit 2 bis 6 kJ/mol angegeben werden [164]. Der Unterschied zwischen den verkürzten Varianten und dem nativen Protein könnte mit dem Einfluss

der CID auf die RNA-Bindung begründet werden. Da sich zwischen den verkürzten Varianten kein Unterschied in der RNA-Bindung zeigen lassen konnte, wurde darauf geschlossen, dass der C-Terminus von Pcf11 für die Bindung verantwortlich ist. Da sich in dieser Region bis auf die beiden Zinkfingerdomänen keine weiteren bekannten Domänen befinden, wurde vermutet, dass diese Region (Pcf11zcx) die RNA-Bindung vermittelt. Dieses wird im Folgenden näher diskutiert.

Darüber hinaus zeigt nativer CFII-Komplex neben der hochaffinen Bindung von RNA eine starke Sequenzspezifität für guanosinreiche Regionen. Zur Überprüfung der Hypothese, dass die Zinkfinger die RNA-Bindung vermitteln, wurden zwei Fusionsproteine generiert. Diese bestanden aus der Zinkfingerdomäne, bestehend aus den beiden Zinkfingern sowie der Clp1-Interaktionsregion, mit einem GST-*tag* oder Hexahistidin-SUMO-*tag*. Beide *tags* sollten spaltbar sein, jedoch konnte für beide Fusionsproteine keine Spaltung unter Standardbedingungen beobachtet werden. Unmittelbar hinter der Spaltstelle für die Sumo- bzw. die PreScission-Protease lag das an der Bindung des Zink-Ions beteiligte Cystein des ersten Zinkfingers. Aufgrund der globulären Struktur und der Stabilität einer Zinkfingerdomäne war die Spaltstelle höchstwahrscheinlich sterisch blockiert. Zur Analyse der RNA-Bindung wurden daher beide Varianten ohne Spaltung parallel untersucht, um einen Einfluss des artifiziellen *tags* auf die RNA-Bindung auszuschließen. Die Fusionsproteine zeigten eine ähnliche Sequenzspezifität zu Homopolymeren wie das native Protein. Während natives CFII vor allem Poly(G)- und mit deutlich reduzierter Affinität auch Poly(U)- und Poly(A)-, jedoch kein Poly(C)-Oligomer band, konnte für das Fusionsprotein nur eine Bindung von poly(G)-Oligomeren beobachtet werden. Da deutlich mehr Protein eingesetzt werden musste, um Bindung zu identifizieren, kann auf eine geringe Affinität in dem Fusionsprotein geschlossen werden. Während die Zinkfingerdomänen in nativem CFII durch die Interaktion mit Clp1 und der umgebenden Struktur von Pcf11 klar positioniert sind, sind diese möglicherweise in der isolierten Domäne deutlich flexibler.

Darüber hinaus ist eine ähnliche Spezifität für die Länge der RNA-Oligomere zwischen nativem CFII und dem Fusionsprotein zu beobachten. So ist eine effektive Bindung erst ab 12 bis 14 Nukleotiden zu beobachten. Ein isolierter Zinkfinger ist jedoch von der Größe her nicht in der Lage, über 12 Nukleotide zu binden, sodass von einer additiven oder kooperativen Bindung beider Zinkfinger auszugehen ist. Um dieses im Detail zu klären, könnten beide Zinkfingerdomänen isoliert untersucht werden.

4.1.1.4 Analyse der FEGP-Wiederholungen

Die FEGP-Wiederholungen sind das auffälligste Sequenzelement des humanen Pcf11. Diese Region zeichnet sich durch die 30-fache Wiederholung einer Konsensussequenz „FEGPPGQPGGGGPLR“ aus. Eine umfangreiche Sequenzanalyse zeigte, dass diese Wiederholungen sehr spezifisch für Pcf11 und in keinem anderen bekannten Protein zu finden sind. Darüber hinaus sind sie ausschließlich in Wirbeltieren zu finden.

Die FEGP-Wiederholungen haben keinen Einfluss auf die Bindung der Substrat-RNA. In der Rekonstitution von fraktioniertem Extrakt, welcher erst durch Zugabe von CFII, Poly(A)-Polymerase sowie CstF in der Lage ist, die Substrat-RNA zu spalten, konnte aber gezeigt werden, dass die FEGP-Wiederholungen notwendig für die Vermittlung der Spaltaktivität sind. So ist mit CFII Δ N769 eine konzentrationsabhängige Spaltung der RNA mit ähnlicher Effizienz wie mit nativem CFII zu beobachten. In CFII Δ N1123 hingegen findet selbst bei der höchsten Konzentration keine Bildung von Produkt-RNA statt. In der Interaktomanalyse, welche in 4.3.1.2 genauer diskutiert wird, konnte nur in der längeren Variante eine Interaktion mit der Poly(A)-Polymerase identifiziert werden. Diese ist essentiell für die Spaltung der Substrat-RNA. Im Vergleich zur Hefe ist Pcf11 Teil des CFIA-Komplexes mit Clp1, Rna14p und Rna15p. Da hier diese Wiederholungen nicht vorhanden sind, könnte durch die CstF-Homologe die notwendige Interaktion vermittelt werden. Ohne weitere experimentelle Daten ist dieses jedoch reine Spekulation.

Eine massenspektrometrische Analyse der Wiederholungen zeigte zudem eine weitere Besonderheit: Die abschließenden Arginine sind mono- bzw. dimethyliert. Dieses konnte auch in endogenem Pcf11, isoliert aus humanen Zelllinien, gezeigt werden [165]. Diese posttranslationale Modifikation von Pcf11 ist möglicherweise auch daran beteiligt, dass die Expression in *E. coli* fehlschlug. Darüber hinaus könnte durch den hohen Gehalt an Glycinen und Prolinen eine ungefaltete Struktur vorliegen. Nicht auszuschließen ist jedoch auch die Bildung einer Polyprolin-Helix [181]. Zum Nachweis der Sekundärstruktur dieser Wiederholungen kann ein Zirkulardichroismus-Spektrum (CD) aufgenommen werden. Da die isolierten Wiederholungen nicht löslich waren, könnten die in der RNA-Bindung sowie Rekonstitutionsexperimenten eingesetzten Deletionsvarianten hierfür genutzt werden. Hierzu nötig ist eine Optimierung der Pufferbedingungen, sowie eine deutliche Erhöhung der Proteinkonzentration. Dieses wurde im Rahmen dieser Arbeit nicht weiterverfolgt. Alternativ zur Bestimmung der Sekundärstruktur mittels CD-Spektrometrie kann die Struktur des nativen CFII mittels Kryo-Elektronenmikroskopie bestimmt werden. Hierbei müssen die FEGP-Wiederholungen jedoch in einer klar definierten Konformation vorliegen. Mit einem nativen Molekulargewicht von ca. 220 kDa ist CFII ausreichend groß für eine solche Analyse. Hierzu notwendig ist eine initiale Proteinkonzentration von ca. 3 mg/ml, was einer molaren Konzentration von ca. 13,5 μ M beträgt. Eine solche Konzentration sollte durch eine Erhöhung des Kulturvolumens und eine Konzentrierung der gereinigten Fraktionen ohne weitere Optimierung möglich sein.

4.1.2 Darstellung des CPSF-Komplexes

Zum CPSF Komplex wurden ursprünglich die vier Untereinheiten CPSF30, CPSF73, CPSF100 sowie CPSF160 gezählt [92]. Erst nachträglich wurden die Untereinheiten WDR33, Fip1 sowie Symplekin diesem Komplex zugeordnet [25], [37], [74], [93]. Eine Interaktion zwischen CPSF-

Untereinheiten und CstF64 wird diskutiert. So konnte für die Interaktion von CstF64 mit Symplekin gezeigt werden, dass diese für die Histon-RNA Prozessierung wichtig ist, nicht jedoch für die Polyadenylierung [87], [182]. Die Bildung eines Histon-Endonuklease-Komplexes aus CPSF73, CPSF100, Symplekin sowie CstF64 konnte darüber hinaus auch gezeigt werden [183]. In dem in dieser Arbeit verwendeten fraktioniertem Extrakt (Daten von Florian Kluge) konnte gezeigt werden, dass in den spaltungsaktiven Fraktionen, welche auf den Zusatz von CstF, CFII und Poly(A)-Polymerase angewiesen sind, CstF64 mit ähnlicher Abundanz wie die CPSF-Untereinheiten angereichert vorliegt, nicht jedoch die anderen CstF-Untereinheiten (persönliche Kommunikation).

Da im Rahmen dieser Arbeit nicht die Zusammensetzung des nativen CPSF-Komplexes, sondern die Aufklärung der Struktur der Untereinheiten im Fokus stand, wurde ein CPSF-Komplex aus den beschriebenen 8 Untereinheiten koexprimiert und gereinigt. Zur Reinigung wurde der kombinierte Hexahistidin-FLAG-*tag* von CstF64 genutzt. In der initialen Reinigung konnte sowohl CPSF100, als Repräsentant des Endonuklease-Komplexes, als auch Fip1, als Repräsentant des mPSF-Komplexes, nachgewiesen werden. In der darauffolgenden FLAG-Immunopräzipitation, wobei der *tag* an CstF64 war, wurde jedoch nur noch CPSF100 nachgewiesen (Abbildung 3.8). CstF64 war also stabil mit dem Endonuklease-Komplex, nicht jedoch mit mPSF assoziiert. Eine Bildung eines stabilen Komplexes aus allen acht Untereinheiten war in dieser Zusammensetzung also nicht zu beobachten. Dieses kann zum einen daran liegen, dass durch die Koexpression von CstF64 ein Histon-Endonuklease-Komplex gebildet wird, der nicht mehr mit dem mPSF-Komplex interagiert. Es ist auch möglich, dass durch die Verwendung von Protein-*tags* die nativen Interaktionsstellen sterisch blockiert werden. Neben dem Hexahistidin-FLAG-*tag* an CstF64 besaß Symplekin einen Strep-*tag*. Eine weitere Erklärung ist, dass diese 8 Proteine nicht ausreichend sind, um einen stabilen Komplex zu assemblieren. Ob eine weitere Untereinheit an der Bildung des Komplexes beteiligt ist, kann hier nicht geklärt werden.

Für die nachfolgenden Untersuchungen wurden daher der mPSF- sowie der Endonuklease-Komplex getrennt betrachtet.

4.1.2.1 Darstellung des mPSF-Komplexes und Aktivitätsanalyse

Der in dieser Arbeit exprimierte und gereinigte mPSF-Komplex bestand aus den vier Untereinheiten CPSF30, CPSF160, Fip1 und WDR33, welches einen Hexahistidin-*tag* trug. Die hierfür genutzten Konstrukte lagen bereits aus vorherigen Arbeiten vor [26]. Gegenüber der bisherigen Reinigungsstrategie wurde der Komplex initial über seinen Hexahistidin-*tag* affinitätschromatographisch gereinigt. Anschließend erfolgte eine Anionenaustauschchromatographie, um kontaminierende Nukleinsäuren zu entfernen. Zur Quantifizierung des Komplexes wurde, wie beschrieben, die Konzentration von Fip1 zugrunde gelegt. In dem analytischen Gel der Reinigung fiel auf, dass dieses optisch gegenüber den anderen Einheiten

angereicherter Vorlag (Abbildung 3.9). Vor allem CPSF160 und WDR33, welche in SDS-Polyacrylamidgelen nicht voneinander separiert werden konnten, waren unterrepräsentiert. Da hier zwei Untereinheiten, die mit Fip1 äquimolar vorliegen sollten, überlagert vorlagen, sollte ein stärkeres Signal in der Coomassie-Färbung vorliegen. Die durch die Quantifizierung von Fip1 abgeschätzte mPSF-Menge überschätzt daher vermutlich die reale Situation und hatte in den folgenden Reaktionen zur Folge, dass deutlich weniger Komplex als berechnet eingesetzt wurden. Darüber hinaus sind im Gegensatz zu allen anderen Komplexen noch sehr viele Kontaminationen nach der Anionenaustauschchromatographie vorhanden. Für hochauflösende Strukturexperimente muss der Komplex noch weiter gereinigt werden. Am nächstliegenden ist hier eine Größenausschlusschromatographie. Neben der weiteren Aufreinigung sollte hierdurch eine Trennung des im Überschuss vorliegenden Fip1 möglich sein, sodass ein stöchiometrischer Komplex erhalten wird.

mPSF vermittelt die Polyadenylierungssignal (PAS)-spezifische Polyadenylierung der Substrat-RNA durch die Poly(A)-Polymerase. Das PAS, welches die Sequenz AAUAAA besitzt, wird durch CPSF30 sowie WDR33 gebunden und rekrutiert die Poly(A)-Polymerase, wodurch diese leicht prozessiv wird [28], [99], [100]. Zur Analyse der Proteinreinigung wurden die Hauptfraktionen auf ihre Fähigkeit die Poly(A)-Polymerase zu stimulieren getestet. In der Kontrolle der punktmutierten Substrat-RNA, welches die Sequenz AAGAAA trägt und damit die Bindung des mPSF-Komplexes verhindert, ist kaum Polyadenylierung zu beobachten. Die beobachtete Aktivität ist hier wohl auf den Überschuss an Fip1, das an U-reiche Sequenzen bindet und somit PAS-unabhängig die Poly(A)-Polymerase rekrutieren kann, zurückzuführen.

4.1.2.1.1 Chemisches *cross-linking* des mPSF-Komplexes

Da der Polyadenylierungsfaktor mPSF bereits zu großen Teilen in seiner Struktur aufgeklärt ist, waren vor allem der Mechanismus und die Dynamik der Polyadenylierungsreaktion von Interesse [99], [100]. Im beschriebenen *in-vitro*-Polyadenylierungsexperiment werden die Proteinfaktoren mPSF sowie die Poly(A)-Polymerase und das Poly(A)-bindende Protein PABPN1 mit einem präadenyliertem Substrat, welches bereits 15 Adenosylreste besitzt, inkubiert [28]. mPSF bindet hierbei durch WDR33 und CPSF30 an das Polyadenylierungssignal AAUAAA. Anschließend muss die Polymerase rekrutiert werden. Das Enzym wird sowohl durch mPSF als auch durch PABPN1, welches sich an den sich bildenden Poly(A)-Schwanz anlagert, stimuliert [29], [30]. Nach der Verlängerung des Poly(A)-Schwanzes auf ca. 250 Nukleotide wird die Interaktion zwischen mPSF und der Poly(A)-Polymerase unterbrochen. Diese Schritte, die Initiation und Bildung des aktiven Komplexes, die Elongation des Poly(A)-Schwanzes sowie die Termination der Polyadenylierung sind biochemisch recht gut verstanden, strukturell jedoch nicht aufgeklärt. Zur Identifikation von Protein-Protein-Interaktionen in den Reaktionsintermediaten sollten diese chemisch vernetzt und analysiert werden. Unter Standardreaktionsbedingungen jedoch ist dieses

aus folgenden Gründen jedoch nur schwer möglich. Zum einen ist die Poly(A)-Polymerase durch die Stimulierung von mPSF und PABPN1 sehr prozessiv, sodass die Reaktion in weniger als einer Minute bereits beendet ist. Zum anderen wird die Poly(A)-Polymerase im Unterschuss gegenüber mPSF und der RNA eingesetzt. Für die *cross-link*-Analyse sollte jedoch der aktive Komplex analysiert werden. Dazu konnten in dieser Arbeit erfolgreich folgende modifizierte Bedingungen definiert werden: Die an der Reaktion beteiligten Faktoren bis auf PABPN1 wurden äquimolar eingesetzt. PABPN1 muss im Überschuss eingesetzt werden, um die vollständige Bedeckung des Poly(A)-Schwanzes zu gewährleisten. Die Reduktion der Reaktionstemperatur konnte von 37 °C auf 25 °C, welches auch der Temperatur der anschließenden *cross-linking* Reaktion entsprach, abgesenkt werden, sodass erst nach ca. 10 Minuten die Reaktion abgeschlossen war. Die Elongationsphase wurde daher nach 3 Minuten isoliert, die Terminationsphase bzw. das Endprodukt nach 10 Minuten. Durch die Zugabe von EDTA konnte die Reaktion sofort abgestoppt werden und die weitere Inkubation über 30 Minuten hatte keinen Einfluss auf die Länge des Poly(A)-Schwanzes. Um bessere *cross-link*-Ergebnisse zu erhalten, wurde darüber hinaus die Protein- und RNA-konzentration 10-fach gegenüber dem beschriebenen Assay erhöht, was aber keinen Einfluss auf die Polyadenylierungsreaktion hatte. Nicht geklärt ist, ob die strukturelle Integrität des Polyadenylierungskomplexes nach Zugabe von EDTA weiterhin gewährleistet ist. Zur Überprüfung könnten die gestoppten Reaktionen mittels Gel-Shift-Experimenten analysiert werden. Durch die Verwendung einer radioaktiv-markierten Substrat-RNA kann so die Größe des gebundenen Proteinkomplexes analysiert werden.

Es konnte gezeigt werden, dass sich hochmolekulare Produkte in Abhängigkeit des chemischen *cross-linkers* bilden. Die Visualisierung der Reaktion mittels SDS-PAGE zeigte darüber hinaus auch, dass das eingesetzte methylierte BSA, welches zur Stabilisierung der Proteine eingesetzt wird, zum großen Teil in seiner Größe unverändert war. Dies bedeutet, dass dieses nicht oder nur kaum mit dem *cross-linker* reagiert hat, was erwartet wurde, da es sich um chemisch methyliertes BSA handelte, und somit keine reaktiven Amine mehr besitzen sollte. Zur massenspektrometrischen Analyse wurden zwei parallele Versuchsdurchführungen verwendet. Zum einen wurde der Reaktionsansatz komplett in Lösung verdaut. Hierbei wurde erwartet, dass neben den Polyadenylierungsfaktoren sehr viele Peptide des Albumins identifiziert werden, da diese nicht abgetrennt wurden. Zum anderen wurde eine präparative SDS-PAGE durchgeführt und die Proteine oberhalb von 150 kDa isoliert und verdaut. Da BSA theoretisch Dimere bilden kann, sollten diese bei ca. 140 kDa vorhanden sein und somit von den hochmolekularen Reaktionsprodukten getrennt werden. Nachteil dieser Methode ist, dass kleinere *cross-link*-Produkte, wie z.B. ein Dimer aus Poly(A)-Polymerase und einer mPSF-Untereinheit, welches jedoch nicht in einem Superkomplex vernetzt sind und somit bei einem apparenten Molekulargewicht von unter 150 kDa migrieren, verloren gehen. Trotz der Trennung von Reaktionsprodukten von BSA konnten keine *cross-links* identifiziert werden. Eine Analyse der

Rohdaten mit MaxQuant zeigte zudem, dass sich zu großen Teilen immer noch BSA in den Proben befand.

Eine Quantifizierung der Polyadenylierungsfaktoren zu den verschiedenen Reaktionszeitpunkten zeigte, dass sich ein RNA-abhängiger Komplex bildet, dieser sich aber über die Reaktionszeit nicht quantitativ nicht verändert (Abbildung 3.11). Eine mögliche Erklärung ist die Bildung eines Kernkomplexes auf der RNA, bestehend aus mPSF, der Poly(A)-Polymerase und mindestens eine Kopie des Poly(A)-bindenden Proteins PABPN1. Alle weiteren Kopien von PABPN1, welche sich auf dem Poly(A)-Schwanz anlagern, sind nicht Teil des quervernetzten Superkomplexes. Dieses ist zu erwarten, da nicht alle Proteine pro Komplex miteinander vernetzt sind und die auf dem Poly(A)-Schwanz angelagerten PABPN1 nur direkten Kontakt zu ihren unmittelbaren Nachbarn haben, nicht zwingend jedoch zu mPSF.

In dem Verdau in Lösung konnten einige *cross-links* identifiziert werden, diese waren jedoch hauptsächlich intramolekular und es zeigten sich keine Unterschiede zu den verschiedenen Reaktionsintermediaten. Nur eine Interaktion zwischen CPSF160 und PABPN1 konnte identifiziert werden. Da auch keine Verbindungen zur Polymerase identifiziert werden konnten, welche die größte räumliche Veränderung während der Reaktion durchläuft, wurden diese Ergebnisse zusammengefasst. Über die Dynamik und den Mechanismus der Polyadenylierungsreaktion konnte aufgrund der Datenlage keine Aussage getroffen werden.

Zur erfolgreicherer Durchführung dieses Experimentes bedarf es noch weiterer Optimierung. Zum einen muss der mPSF-Komplex weiter gereinigt werden. Eine Größenausschlusschromatographie könnte helfen, aktiven Komplex von den verunreinigenden Proteinen zu trennen, aber auch um den Überschuss von Fip1 zu separieren. Dieses hilft, die reale Konzentration des Komplexes besser zu bestimmen. Darüber hinaus können die Reaktionsbedingungen weiter optimiert werden. Da das methylierte BSA eingesetzt wurde, um die in der Standardreaktion sehr geringen Proteinkonzentrationen zu stabilisieren, jedoch bereits 10-fach mehr Protein eingesetzt wurde, kann möglicherweise auf die Verwendung komplett verzichtet werden. Neben BSA befand sich PEG6000 in der Reaktion. Dieses unterstützt die Reaktion, da es als molekularer *crowder* wirkt und somit die lokale Proteinkonzentration durch Lösungsmittelenzug erhöht. PEG ist jedoch nicht Massenspektrometrie-tauglich und muss vor der Analyse entfernt werden. Dieses ist durch Zentrifugen-Filtereinheiten (*centricons*) oder SDS-PAGE möglich, jedoch reduziert man unter Umständen die Peptidausbeute. Es wäre zu testen, ob man auf den Einsatz von PEG verzichten oder die Konzentration reduzieren kann. Darüber hinaus wurden zur Stabilisierung der eingesetzten Substrat-RNA sowohl kommerzieller RNase-Inhibitor sowie tRNA aus *E. coli* eingesetzt. Die tRNA selbst ist nur gering reaktiv gegenüber dem *cross-linker*, jedoch ist diese mit 1 g/l deutlich konzentrierter als die Proteinfaktoren. Auch hier ist

die Optimierung dahingehend, die *cross-link*-Reaktion in ihrer Heterogenität zu reduzieren um letztendlich mehr *cross-link*-Produkte zu erhalten.

4.1.2.2 Darstellung des Endonuklease-Komplexes und Aktivitätsanalyse

Der in dieser Arbeit analysierte Endonuklease-Komplex bestand aus den vier Untereinheiten CPSF73, CPSF100, Symplekin sowie CstF64. Zur Expression wurden zwei getrennte Baculoviren genutzt, ein Virus, der für die drei Untereinheiten CPSF73, CPSF100 sowie Strep-*tag* markiertes Symplekin kodierte (bereitgestellt von Uwe Kühn), sowie ein Virus, welches für das Hexahistidin-FLAG-*tag* markierte CstF64 kodierte [184]. Die Reinigung erfolgte, wie bei allen anderen Komplexen, über eine anfängliche Reinigung unter Verwendung des Hexahistidin-*tags* an CstF64 und anschließende Anionenaustauschchromatographie. Sehr auffällig ist, dass der Komplex in einem scharfen Peak eluierte. Dieses deutet auf eine sehr hohe Homogenität des Komplexes hin. Das analytische SDS-Polyacrylamidgel zeigte darüber hinaus, dass die vier Untereinheiten einen äquimolaren Komplex bilden. Da die Färbung mittels Coomassie jedoch nicht sehr zuverlässig ist, kann dieses noch durch eine analytische Gelfiltration, möglicherweise kombiniert mit dynamischer Lichtstreuung, native Massenspektrometrie oder analytische Ultrazentrifugation verifiziert werden.

Zur Bestimmung der Aktivität wurde die Bindung der Substrat-RNA gemessen. CstF64 besitzt eine RNA-Binde-Domäne [85]. Unklar ist, ob die drei weiteren Untereinheiten auch in der Lage sind, RNA zu binden. Für den Endonuklease-Komplex konnte zur Substrat-RNA L3 eine sehr hohe Affinität von 3,5 nM (K_{50}) bestimmt werden. In der Analyse der RNA-Bindung von CstF, in dem ausschließlich CstF64 die RNA-bindende Untereinheit darstellt, konnte ein K_{50} -Wert von rund 11 nM gemessen werden, wobei hier ausschließlich CstF64 in der Lage ist, RNA zu binden [73]. Obwohl in dem CstF-Komplex CstF64 in zwei Kopien vorlag, konnte für den Endonuklease-Komplex eine dreifach höhere RNA-Affinität bestimmt werden. Daraus folgt, dass entweder CstF64 im Komplex mit den CPSF-Untereinheiten eine höhere RNA-Affinität besitzt, also dass die Komplexbildung einen allosterischen Effekt bewirkt, oder die höhere RNA-Affinität ein Produkt aus mehreren RNA-bindenden Untereinheiten des Endonuklease-Komplexes darstellt. Zur näheren Untersuchung kann ein Komplex ohne CstF64 auf seine RNA-Bindung hin untersucht werden.

Der Endonuklease-Komplex konnte nicht auf seine intrinsische Aktivität der RNA-Spaltung getestet werden, da dieses noch nicht *in vitro* rekonstituiert ist. Der für den CFII- und im Folgenden für den CstF-Komplex genutzte fraktionierte Extrakt zeigte zudem keine Abhängigkeit der Spaltung von zugesetztem Endonuklease-Komplex, da dieser in der Fraktion bereits enthalten war.

Die Darstellung des Endonuklease-Komplexes als Heterotetramer bestätigt die stabile Assoziation von CstF64 zu den drei bekannten CPSF-Untereinheiten. Inwieweit diese Assoziation

für die mRNA-3'-Prozessierung relevant ist, kann hier nicht geklärt werden. Gezeigt werden konnte hingegen, dass die Bindung von CstF64 durch Symplekin wichtig für die Bildung der aktiven Endonuklease in der Histon-RNA-3'-Prozessierung ist [116].

4.1.2.2.1 *Cross-linking* des Endonuklease-Komplexes

Zur strukturellen Analyse wurde der Endonuklease-Komplex chemisch vernetzt und massenspektrometrisch analysiert. Zur Bestimmung der optimalen *cross-linker*-Konzentration wurde dieser gegen den Komplex titriert (Abbildung 3.13). Hierbei kann die Bildung eines hochmolekularen Komplexes beobachtet werden. Auffällig ist, dass die Konzentration von monomeren CstF64 nicht proportional mit den anderen Untereinheiten abnimmt. Entweder liegt CstF64 also in höherer Konzentration und somit überproportional zu dem Endonuklease-Komplex vor oder die drei Untereinheiten CPSF73, CPSF100 und Symplekin bilden einen sehr stabilen Kern des Komplexes und CstF64 ist mit geringerer Energie damit assoziiert. Daher werden statistisch vor allem die ersteren quervernetzt und erst mit steigender Konzentration des *cross-linkers*, und somit mehr als einer Vernetzung pro Komplex, auch CstF64. Zur präparativen Analyse der *cross-links* wurde eine moderate Konzentration des *cross-linkers* gewählt, um artifizielle Zustände zu vermeiden. Deutlich wird dieses in dem großen Anteil an monomeren Protein. Möglicherweise hätte man hier etwas mehr *cross-linker* einsetzen können, um die Ausbeute zu erhöhen. Insgesamt konnten jedoch 209 einzigartige *cross-links* identifiziert werden, was einem sehr guten Ergebnis entspricht. Die meisten inter-molekularen Verknüpfungen konnten zwischen den C-terminalen Domänen (CTD) von CPSF73 und CPSF100 identifiziert werden. Es ist bekannt, dass diese Proteine über diese Bereiche dimerisieren und dieses essentiell für die Aktivität ist [183], [185], [186]. Weitere intermolekulare Verknüpfungen konnten zwischen der beschriebenen Interaktionsregion von Symplekin und CstF64 beobachtet werden [87]. Diese beiden Regionen zeigen auch Verknüpfungen zu den CTDs von CPSF73 und CPSF100. Alle vier Regionen bilden somit den Kern, in dem alle Untereinheiten miteinander interagieren. Auffällig ist zudem, dass trotz ähnlicher Struktur CPSF100, nicht jedoch CPSF73, viele intra-molekulare *cross-links* aufweist. Dieses deutet möglicherweise darauf hin, dass sich CPSF100, aber nicht CPSF73, an der Oberfläche des Komplexes befindet und CPSF73 hierbei von anderen Proteinen bedeckt ist. Auch auffällig ist, dass trotz seiner Größe Symplekin nur sehr wenige Verknüpfungen zu den anderen Proteinen aufweist. Vor allem die C-terminale Region zeigt sehr viele intramolekulare *cross-links*, was bedeutet, dass dieser Bereich klar strukturiert und lösungsmittelzugänglich ist. Da jedoch kein intermolekularer *cross-link* identifiziert wurde, liegt diese Region wahrscheinlich nicht in räumlicher Nähe zu den anderen Proteinen. Da Symplekin als Gerüstprotein für die 3'-Prozessierung beschrieben ist, ist dieser Bereich möglicherweise für Interaktionen mit anderen Proteinen oder Komplexen verantwortlich.

4.1.3 Darstellung des CstF-Komplexes und Aktivitätsanalyse

Der CstF-Komplex in dieser Arbeit bestand aus den drei Untereinheiten CstF50, CstF64 und CstF77. CstF64 *tau*, das Paralog zu CstF64, wurde nicht exprimiert. Das verwendete Expressionskonstrukt war bereits vorhanden [163]. CstF bildet einen sehr stabilen Komplex mit einer sehr guten Expressionsrate, welches an den Ausbeuten von 2,8 bzw. 7,7 mg Komplex pro Liter Expressionskultur deutlich wird (Abbildung 3.14). Die zweistufige Reinigung ist ausreichend, um den Komplex nahezu homogen zu reinigen. In vorherigen Reinigungen konnte beobachtet werden, dass CstF64 in Gegenwart von Imidazol vom Komplex dissoziieren kann. Durch eine kurze Elutionszeit während der anfänglichen Reinigung sowie der anschließenden Anionenaustauschchromatographie, wobei die Probe verdünnt und nicht dialysiert wurde, konnte dieses verhindert werden. In der Konzentrationsbestimmung mittels SDS-PAGE konnte kein signifikanter Unterschied in den Proteinmengen der drei Untereinheiten beobachtet werden. Der gereinigte Komplex wurde sowohl auf seine RNA-Bindung als auch in der Rekonstituierung der Spaltungsreaktion getestet. Es ist bekannt, dass CstF G/U-reiche Sequenzen bindet [76]. Für ein synthetisches RNA-Oligonukleotid (GU)₁₄ konnte eine Dissoziationskonstante von 120 nM durch isotherme Titrationskalorimetrie bestimmt werden [73]. Für die Analyse des hier gereinigten CstF-Komplexes wurde die Bindung an das RNA-Substrat der Spaltung getestet. Hierbei zeigte sich, dass diese hochaffin mit einem K₅₀-Wert von ca. 11 nM gebunden wird. Dieses ist ungefähr Faktor 10 stärker als die Bindung der synthetischen RNA. Neben der RNA-Bindung wurde erfolgreich die Rekonstituierung der RNA-Spaltung in fraktioniertem Extrakt getestet. Hierbei wird deutlich, dass CstF-Komplex zwingend für die Spaltung benötigt wird. Es war jedoch keine Konzentrationsabhängigkeit zu beobachten. Schon geringe Mengen von 12,5 fmol zeigen schon fast vollständige Rekonstitution. CstF muss also möglicherweise nicht äquimolar zur RNA eingesetzt werden.

4.1.3.1 Cross-linking des CstF-Komplexes

Der CstF-Komplex konnte in sehr hoher Konzentration gereinigt werden, was für die Identifikation von *cross-links* von Vorteil ist. In der Titration des *cross-linkers* zeigte sich, dass sich schon bei einem geringen Überschuss hochmolekulare Produkte bilden. Diese resultieren in einem finalen Endprodukt, welches kaum noch in das SDS-Polyacrylamidgel migriert. Dieses deutet auf eine sehr starke Stabilisierung der Quartärstruktur hin. Auch kann beobachtet werden, dass CstF als homogener Komplex vorliegt. Unterschiede würde hier in einem diverserem Bandenmuster sowie einer ungleichen Abnahme der monomeren Proteine zu sehen sein. Insgesamt konnten 326 einzigartige *cross-links* identifiziert werden, was sehr hoch für einen isolierten Komplex ist. Deutlich wird, dass alle drei Untereinheiten miteinander interagieren. Dieses widerspricht der bisherigen Annahme, dass CstF77 als Gerüstprotein die beiden weiteren Untereinheiten zusammenhält, diese sich jedoch nicht "sehen". Vor allem die Region um Aminosäure 200 in CstF50 zeigt sehr viele Verknüpfungen, sowohl mit der N-terminalen Domäne von CstF64,

welche sich aus der RRM- sowie der Scharnierdomäne zusammensetzt, als auch mit der HAT-Domäne von CstF77. Darüber hinaus interagiert die Dimerisierungsdomäne von CstF50 mit CstF77. Auffällig in CstF64 ist das Fehlen von *cross-links* nach Aminosäure 200. Diese Region ist sehr Glycin- und Prolinreich (~30 %) und besitzt nur 3 Lysine. Dieses ist konsistent mit der Verteilung der *cross-links* des Endonuklease-Komplexes, welcher auch CstF64 enthielt. Eine Aussage über mögliche Interaktionen kann hier nicht getroffen werden. CstF77 wird strukturell durch seine HAT-Domäne dominiert. In dieser Region sind die meisten *cross-links* zu beobachten, was zum einen für eine klare Tertiärstruktur, aber auch über eine gute Lösungsmittel-Zugänglichkeit spricht.

Aufgrund der Zusammensetzung des Komplexes, eines Dimers von Trimeren, ist eine eindeutige Differenzierung zwischen "echten" intramolekularen *cross-links* zu solchen, die in einer homodimeren Struktur liegen, nicht möglich. Ausnahmen stellen hier drei *cross-links* in der Struktur von CstF64 dar. In diesen wurde dieselbe Aminosäure in beiden Kopien miteinander verknüpft und eine tryptische Spaltstelle in einem der Peptide ausgelassen. Hierdurch können die Ionenserien der verschiedenen Peptide zum Teil unterschieden werden, sodass es zu einer korrekten Zuordnung kommt. Um eine bessere Differenzierung der *cross-links* zu ermöglichen, muss eine Hälfte des Komplexes mit schweren Isotopen markiert werden. Üblich ist hier die Expression der Proteine in Anwesenheit von ¹⁵N- oder ¹³C-Isotopen [187]. Nach der Expression muss jedoch ein hybrider Komplex aus markiertem und unmarkiertem CstF formiert werden. Ein reines Mischen der Komplexe wird höchstwahrscheinlich nicht zu einem Austausch der Untereinheiten führen, da diese einen sehr stabilen Komplex bilden. Eine theoretische Möglichkeit wäre das Mischen unter denaturierenden Bedingungen und das anschließende Renaturieren des Komplexes. So würde man einen hybriden Komplex aus markierten und unmarkierten Proteinen erhalten. Da dieses jedoch mit sehr viel Aufwand verbunden ist und der Erfolg nicht garantiert ist, steht dieser Aufwand nicht in Relation zum möglichen Ergebnis. Zur genaueren Strukturaufklärung des CstF-Komplexes ist daher vor allem Kryoelektronenmikroskopie in Lage. Mit einem nativen Molekulargewicht von über 300 kDa und der hier beschriebenen Ausbeute von 7,7 mg pro Liter Kulturvolumen kann eine solche Analyse ohne weitere Optimierung mit dem in dieser Arbeit dargestellten Komplex durchgeführt werden.

4.1.4 Darstellung des CFI-Komplexes und Aktivitätsanalyse

Der in dieser Arbeit analysierte CFI-Komplex besteht aus den Untereinheiten CFI25 sowie CFI68 und bildet ein Heterotetramer, mit je zwei Kopien der Untereinheiten. Die hierzu genutzten Expressionskonstrukte lagen bereits vor [163]. Die alternativen Proteine CFI59 und CFI72 wurden nicht berücksichtigt. Die Ausbeute von nur 0,45 mg lag deutlich unterhalb der anderen Komplexe, jedoch wurde das Protein in sehr hoher Reinheit erhalten. Der für die publizierte Kristallisation verwendete Komplex, der eine verkürzte Variante von CFI68 enthielt, wurde in *E. coli* exprimiert

[188]. Hierbei können deutlich höhere Proteinausbeuten erreicht werden. Zur Überprüfung der nativen Faltung wurde die RNA-Bindung des Komplexes analysiert. Dafür wurde ein synthetisches RNA-Oligonukleotid verwendet, welches über zwei Bindemotive für CFI verfügte. Da CFI über zwei RNA-Bindestellen verfügt, sollte so theoretisch eine RNA pro Komplex gebunden werden. Die Analyse der RNA-Bindung zeigt jedoch zwei Bindungsereignisse, eine erste hochaffine Bindung und eine zweite mit deutlich reduzierter Affinität. Eine mögliche Erklärung ist, dass im ersten Schritt ein RNA-Molekül von beiden RNA-bindenden Untereinheiten gebunden wird und die hohe Affinität einen kooperativen Effekt darstellt. Die wenig affine Bindung stellt die Bindung von zwei RNA-Molekülen pro Komplex da. Hierbei tritt keine Kooperativität auf. Eine detaillierte kinetische Untersuchung dieser RNA-Bindung fand im Rahmen dieser Arbeit jedoch nicht statt.

4.1.4.1 *Cross-linking* des CFI-Komplexes

Der Kern des CFI-Komplexes besteht aus einem Homodimer von CFI25 sowie den RRM-Domänen von CFI68, welche je einer Kopie CFI25 zugeordnet sind. Dieser Komplex ist bereits strukturell aufgeklärt worden [188]. Die *cross-link*-Experimente wurden daher vor allem durchgeführt, um Informationen über die Region von CFI68, welche nicht in der Kristallstruktur aufgeklärt wurde, zu erlangen. Es zeigte sich, dass unter der Verwendung des *cross-linkers* DSBU nur sehr wenige Verknüpfungen auftraten. Daher wurde zusätzlich der *cross-linker* LC-SDA verwendet. Dieser ist bifunktional und besitzt eine NHS-Gruppe, welche mit freien Aminen und primären Alkoholen reagiert, sowie eine Diazirin-Gruppe, welche bevorzugt mit Karbonsäuren reagiert. Der *cross-linker* besitzt die gleiche Länge wie DSBU. Unter Verwendung von LC-SDA konnten in den Analysen jedoch nur zwei *cross-links* identifiziert werden. Insgesamt konnten nur 8 *cross-links* identifiziert werden, was eine Aussage über die Struktur sehr limitiert. Problematisch bei der Identifikation von *cross-link*-Produkten ist die hohe Anzahl an Prolinen (122; 15,7 %) in dem Komplex. Zum einen deutet dieses auf eine potentiell flexible oder intrinsisch ungeordnete Region hin, zum anderen neigen Peptide zu einer bevorzugten Fragmentierung N-terminal von Prolinen, was die korrekte Identifikation erschwert [189].

Interessanterweise liegen die identifizierten intermolekularen *cross-links* nicht in dem Bereich, der in der Kristallstruktur aufgelöst ist. Stattdessen sind die Interaktionen ausschließlich zwischen der N-terminalen Region von CFI25, welche den Beginn der RNA-bindenden Region umfasst, sowie der R/S-Region von CFI68. Diese Interaktion ist bei beiden verwendeten *cross-linkern* zu beobachten. Die in der Kristallstruktur identifizierten Interaktionsregionen, die auch die RNA-bindenden Bereiche umfasst, weisen keine *cross-links* auf.

4.2 Computergestützte Modellierung der Polyadenylierungsfaktoren

4.2.1 Generierung der Entfernungsbeschränkungen

Um die Informationen der *cross-links* zu nutzen, müssen diese als Entfernungsverteilung zwischen den beiden verknüpften Aminosäuren beschrieben werden. Üblicherweise wird hier eine maximale Länge zwischen den C α -Atomen definiert, da somit alle möglichen Konformationen berücksichtigt sind. Die Maximalentfernung setzt sich aus der Länge des *cross-linkers*, der Länge der Aminosäureseitenkette sowie der intrinsischen Flexibilität der Proteine zusammen. Die Länge des DSBU-*cross-linkers* beträgt ca. 12 Å, die Länge von Lysin-Seitenketten wird allgemein mit 6,4 Å angegeben. Die theoretische Entfernung zwischen zwei C α -Atomen entspricht daher der Summe von ca. 24,8 Å. Diese berücksichtigt jedoch nicht die Flexibilität von Proteinen. Dieses wurde von *Merkley et al.* in einer breiten Analyse von C α -Atomen in 807 repräsentativen Proteinen näher beleuchtet [131]. Hierbei wird eine C α -Entfernung von bis zu 30 Å empfohlen. Da der hier verwendete DSBU-*cross-linker* etwas länger als der untersuchte BS3-*cross-linker* ist, wurde als Ausgangsentfernung 34 Å gewählt.

4.2.1.1 Bestimmung der Länge von Aminosäureseitenketten

In Abhängigkeit von der Anzahl der identifizierten *cross-links* ist eine möglichst genaue Definition der Entfernungsbeschränkung für die erfolgreiche Modellierung nötig. Je weniger *cross-links* vorhanden sind, desto genauer müssen diese die Struktur beschreiben können, um erfolgreich ein Strukturmodell zu generieren. In der allgemeinen Beschreibung wird von einer Verknüpfung zweier Lysin-Seitenketten ausgegangen. Da DSBU auch mit Hydroxylgruppen in Serinen, Threoninen und Tyrosinen reagieren kann, wurden für diese die Entfernung zwischen C α -Atom und der reaktiven Gruppen empirisch bestimmt. Als Grundlage der Entfernungsbestimmung diente die publizierte Struktur der 50S Untereinheit des *E. coli* Ribosoms (PDBID: 6QUL) [190]. Hierbei zeigte sich zum einen, dass in Lysinen die C α -N ζ -Entfernung meist kürzer als 6,4 Å ist, diese jedoch der Maximalentfernung entspricht. Serin und Threonin besitzen eine C α -O γ - (Ser) bzw. C α -O δ -Entfernung (Thr) von 2,4 Å. Hierbei ist aufgrund der kurzen Seitenkette keine Variation möglich. Für Tyrosin konnte eine C α -O η -Entfernung von 6,5 Å bestimmt werden. Aufgrund der starren Struktur des Benzolringes ist auch hier keine große Varianz zu beobachten.

Für *cross-links* zu Serinen oder Threoninen sollte daher die maximale Entfernung um 4 Å reduziert werden. Dieses kann in der Modellierung von Komplexen, vor allem, wenn nur sehr wenige *cross-links* identifiziert wurden, zu einer Verbesserung des Ergebnisses führen, da der mögliche Interaktionsraum eingeschränkt wird.

4.2.1.2 Peptid-Entfernungsbeschränkungen

Um die *cross-link*-Daten in der Modellierung von Komplex-Strukturen zu verwenden, benötigen die verwendeten Programme die minimale und maximale Entfernung zwischen zwei Atomen, die

in den jeweiligen monomeren Strukturen aufgelöst sind. Liegt die quervernetzte Aminosäure jedoch außerhalb von aufgelösten Bereichen, ist diese für die Strukturmodellierung wertlos. Nicht aufgelöste Bereiche können in publizierten Strukturen z.B. durch ungeordnete Bereiche oder durch Deletionsvarianten entstehen. Wird ein Komplex aus Teilstrukturen, zum Beispiel aus Homologiemodellen von Domänen, assembliert, befinden sich nicht aufgelöste Strukturen außerhalb dieser Teilstrukturen. Um jedoch *cross-link*-Informationen zu verwerten, welche innerhalb eines Rahmens von 10 Aminosäuren außerhalb der vorhandenen Struktur liegen, wurde eine neuartige und allgemeingültige Entfernungsbegrenzung definiert. Diese „Peptid-Entfernungsbegrenzung“ beschreibt die minimale und maximale Länge des nicht aufgelösten Peptides zwischen der letzten aufgelösten und der quervernetzten Aminosäure. Hierzu wurden ca. 21000 pdb-Dateien analysiert, welche repräsentativ für alle publizierten Strukturen standen. Dabei wurden nur Strukturen analysiert, welche in dem alten pdb-, nicht aber ausschließlich in dem neuen mmCIF-Format vorlagen. Grund hierfür ist eine kompliziertere Annotation der Atomkoordinaten, die jedoch aufgrund neuer Anforderungen an die Beschreibung von Komplexstrukturen nötig war [191].

Die Analyse der rund 70 Millionen Entfernungen zeigte, dass unabhängig der Länge des Peptides die minimale Entfernung zwischen den $\text{C}\alpha$ -Atomen der terminalen Aminosäuren immer 3 Å beträgt. Dieses ist grob im Einklang mit dem van-der-Waals-Radius von Kohlenstoff, welcher mit 1,7 Å angegeben ist [192]. Differenziert werden muss diese Minimalentfernung zwischen Di- und Oligopeptiden. In Dipeptiden konnten rund 20000 Entfernungen gemessen werden, die unterhalb von 3 Å lagen. Die Ursache hierfür liegt vermutlich in der *cis*-Konformation von Peptidbindungen. Die $\text{C}\alpha$ - $\text{C}\alpha$ -Entfernung in einer *cis*-Peptidbindung beträgt 2,94 Å [167]. Diese konnten in rund 0,2 % aller Dipeptide identifiziert werden. Dieser Wert deckt sich mit der Literaturangabe zur Häufigkeit von *cis*-Peptidbindungen [193]. In Peptiden, die länger als zwei Aminosäuren sind, stellen $\text{C}\alpha$ - $\text{C}\alpha$ -Entfernungen von 3 Å eine Schleifen-Struktur da. Die gemessenen $\text{C}\alpha$ -Atome sind in räumlicher Nähe, während die Aminosäuren dazwischen die Schleifenstruktur bilden. 3 Å stellt daher die minimale physikalische Entfernung zwischen zwei $\text{C}\alpha$ -Atomen in Lösung dar.

Weiterhin ist zu beobachten, dass die Maximallänge von Peptiden nicht über eine Summe von 3,8 Å pro $\text{C}\alpha$ - $\text{C}\alpha$ -Bindung hinausgeht. In der Analyse der Dipeptide ist dieses die durchschnittliche Entfernung einer $\text{C}\alpha$ - $\text{C}\alpha$ -Bindung. In Polypeptiden stellt diese Maximalentfernung ein maximal gestrecktes Peptid ohne Sekundärstrukturmerkmale dar. Ab einer Länge von ca. 7 Aminosäuren kann beobachtet werden, dass diese theoretische Maximalgrenze nicht mehr erreicht wird. Hier kommt es daher zur Ausbildung lokaler Sekundärstrukturen. Dieses ist auch in der nicht-linearen Abhängigkeit der durchschnittlichen Entfernung von der Anzahl der Aminosäuren zu beobachten.

Da in der Strukturmodellierung ohne weitere Analysen keine Aussage über die Struktur eines Peptides gegeben werden kann, werden für die Definition der Peptidbegrenzung die absoluten

Minimal- bzw. Maximalwerte genutzt. Dieses entspricht daher einer sequenzunabhängigen Beschreibung der Peptidlänge und ist daher allgemeingültig anwendbar.

4.2.2 Modellierung der Subkomplexe

Zur Modellierung der in dieser Arbeit untersuchten Komplexe wurde eine mehrstufige Strukturmodellierung durchgeführt. Im ersten Schritt wurden die Strukturen der monomeren Untereinheiten zusammengetragen, wobei unbekannte Strukturen, wenn möglich, durch Homologiemodellierung generiert wurden. Homologe Strukturen wurden hierbei durch Verwendung von BLASTp und HHPRED identifiziert. Hierbei wurden publizierte, strukturhomologe Proteine identifiziert, die Folgend für die Strukturvorhersage verwendet wurden. *De-novo*-Modellierung von unbekanntem Substrukturen wurde nur für sehr kurze Sequenzbereiche durchgeführt und diese wurden, sofern sie keine klare Konformation zeigten, wieder verworfen. Anschließend erfolgte eine Assemblierung der monomeren Strukturen unter Verwendung der experimentellen *cross-link*-Informationen durch *Docking* unter Verwendung des HADDOCK-Webservers [138]. War die Bildung einer Komplexstruktur nicht weiter möglich, wurde für das kleinere Protein der Konformationsraum mittels DisVis analysiert [136]. Zur Verifizierung aller gebildeter Strukturen wurden diese unter Verwendung der experimentellen *cross-link*-Informationen auf ihre Plausibilität hin untersucht. Ein Modell sollte, wenn möglich, keinen der experimentellen *cross-links* verletzen.

4.2.2.1 Modellierung des CFII Komplexes

4.2.2.1.1 Generierung der monomeren Untereinheiten

Clp1

Da von Clp1 bereits die Strukturen der Proteine aus Hefe und *C. elegans* durch Röntgenkristallographie aufgeklärt werden konnten, und diese eine sehr hohe Sequenzidentität zu dem humanen Protein aufwiesen, konnte ein Homologiemodell generiert werden. In diese Struktur konnten die experimentellen intramolekularen *cross-links* kartiert werden.

Es ist daher davon auszugehen, dass das generierte Modell der nativen Struktur entspricht.

Pcf11

Von Pcf11 ist strukturell sehr wenig bekannt und es sind nur kleine Teilstrukturen dieses Proteins publiziert. Eine *de-novo*-Modellierung dieses Proteins ist aufgrund seiner Größe nicht möglich. Die Analysen beschränkten sich daher auf Bereiche, in denen intermolekulare *cross-links* identifiziert werden konnten, da diese Informationen für die anschließenden *Docking*-Experimente obligatorisch sind. In dem zentralen Teilbereich Pcf11₅₀₀₋₇₇₀ sowie dem C-terminalen Pcf11₁₂₀₀₋₁₅₅₅ konnten insgesamt 7 homologe, distinktive Strukturen identifiziert werden. Eine gesonderte

Rolle hat die Interaktionsregion von Pcf11 zu Clp1. Diese wurde in Clp1 kokristallisiert, was zur Modellierung genutzt wurde und später diskutiert wird.

Die anderen 6 Teilstrukturen wurden separat modelliert. Während die Sequenzen Pcf11₅₇₉₋₆₃₁ und Pcf11₆₅₅₋₆₉₅ im N-terminalen Bereich und die beiden Zinkfingerdomänen im C-terminalen Bereich klar definierte Strukturen vorweisen, sind die beiden weiteren Sequenzen zu kurz, um eine klare Struktur vorzuweisen. Die *coiled-coil*-Struktur von Pcf11₆₄₆₋₆₅₃ sowie die helikale Struktur von Pcf11₁₂₆₅₋₁₂₈₁ sind daher unter Vorbehalt zu betrachten, jedoch dienten sie für das nachfolgende *Docking* als Ankerpunkte.

Für die Modellierung der Interaktionsregion Pcf11₁₃₇₇₋₁₄₀₁ wurden zwei parallele Modellierungsansätze verfolgt. In beiden diente das kristallisierte Peptid von γ Pcf11 als Vorlage. Im ungebundenen Ansatz wurde dieses isoliert genutzt, im gebundenen Ansatz wurde das Peptid in die vorher modellierte humane Clp1-Struktur überlagert. In letzterem Ansatz wurde neben dem Peptid von Pcf11 auch die Struktur von Clp1 remodelliert. Sehr deutlich wird der Qualitätsunterschied der generierten Modelle in der Abweichung von der Vorlagestruktur. In dem ungebundenen Ansatz ist eine durchschnittliche Abweichung von 2 Å mit sehr breiter Streuung zu beobachten, während im gebundenen Ansatz nur eine Konformation gebildet wurde. Darüber hinaus kommt es in der Struktur von Clp1 zu einer Neuausrichtung der an der Bindung beteiligten Aminosäuren. Dieses wird vor allem an den beiden Argininen (Arg170 und Arg220) deutlich, die nun an der Bindung beteiligt sind. In der Modellierung von Strukturen ist es daher sehr sinnvoll, bekannte Interaktionen und Strukturen mit in die Berechnung einfließen zu lassen, da dieses zu deutlich besseren Strukturen führt.

4.2.2.1.2 Definition der Entfernungsbeschränkung

Die für das *Docking* von CFII benötigten Entfernungsbeschränkungen wurden aus den experimentellen *cross-link*-Daten abgeleitet. Hierbei zeigte die in dieser Arbeit definierte „Peptid-Entfernungsbeschränkung“ ihr Potential. Neben 15 *cross-links*, welche innerhalb der Sequenzen der generierten Modelle lagen, konnten so 38 weitere *cross-links* verwendet werden. Darüber hinaus kann diese Beschränkung auch genutzt werden, um Sequenzlücken zu beschreiben. Liegen zwischen zwei aufgelösten Strukturen nur bis zu 10 Aminosäuren, kann diese Entfernung auch beschrieben werden. Dieses konnte im Fall der Interaktionsregion und des ersten Zinkfingers angewendet werden.

4.2.2.1.3 Docking des CFII-Komplexes

Zur Generierung des CFII-Komplexes wurden in sechs parallelen Ansätzen die vorher erstellten Homologiemodelle von Pcf11 mit Clp1, welches die Interaktionsregion von Pcf11 schon gebunden hatte, *gedockt*. Hierbei wurde der HADDOCK-Onlineserver genutzt, der 200 finale

Modelle erzeugt, die in sogenannte Cluster eingeordnet werden. Eine hohe Anzahl von Modellen in einem Cluster ist hierbei ein vertrauenswürdiges Ergebnis.

Die dimeren *Docking*-Versuche waren in zwei Fällen erfolgreich: Clp1*-Pcf11₅₇₉₋₆₃₁ mit einer Clustergröße von 195 von 200 Modellen, was auf eine klar definierte Konformation hinweist und einem sehr guten *docking*-Ergebnis entspricht, und Clp1*-1341-1368, mit einer Clustergröße von 134 von 200 Modellen, was einem akzeptablen Ergebnis entspricht. In dem zweiten Komplex wurde die Sequenzlücke von 9 Aminosäuren zwischen dem Zinkfinger und der Interaktionsregion durch *de-novo*-Modellierung aufgefüllt, da in diesem Bereich Aminosäuren lagen, die direkt an identifizierten *cross-links* beteiligt waren. Hierdurch konnte dann im weiteren Verlauf auf die Verwendung der Peptid-Beschränkung verzichtet werden und die verwendeten Entfernungsbeschränkungen konnten restriktiver definiert werden.

Ausgehend von den gebildeten dimeren Strukturen konnte kein weiterer Komplex gebildet werden, sodass trimeres *Docking*, d.h. mit drei unabhängigen Ausgangsstrukturen, durchgeführt wurde. Hierdurch erhöht sich zwar die Komplexität der Berechnungen, jedoch können mehr Entfernungsbeschränkungen verwendet werden. Aus allen möglichen Kombinationen konnte ein weiterer Komplex gebildet werden: Clp1*-Pcf11₅₇₉₋₆₃₁ in Kombination mit Pcf11₆₄₆₋₆₅₃ und Pcf11₆₅₅₋₆₉₅. Hierbei wurde ein Cluster mit 127 Modellen gebildet. Eine Detailanalyse zeigte jedoch, dass Pcf11₆₅₅₋₆₉₅ sehr gut lokalisiert vorlag, Pcf11₆₄₆₋₆₅₃ jedoch keine klare Konformation besaß. Das letztere Peptid wurde daher manuell aus der Struktur entfernt. Deutlich wird hierbei jedoch der Nutzen von kurzen, wenig strukturierten Sequenzbereichen als „Anker“ im *Docking* der strukturierten Domänen. Pcf11₆₄₆₋₆₅₃ ermöglichte also die korrekte Platzierung von Pcf11₆₅₅₋₆₉₅, ohne jedoch selbst eine klare Position einzunehmen. Möglich wäre hieraus eine weitere Ableitung einer strukturunabhängigen Entfernungsbeschränkung. Sind zwei Strukturen indirekt über einen weiteren Bereich unbekannter Struktur quervernetzt, so kann dieser als molekularer Platzhalter verwendet werden. Die Struktur des unbekanntes Bereiches könnte *de novo* oder besser über eine automatische Template-basierte Modellierung (z.B. mittels I-TASSER) generiert werden und nach dem *Docking* wieder entfernt werden [140].

Da alle weiteren Versuche, einen Komplex zu assemblieren, fehlschlagen, wurden die beiden generierten Komplexe zu einem zusammengefügt. Zur Verifizierung dieses CFII-Komplexes wurden alle in der Struktur enthaltenen *cross-links* verwendet. Es zeigte sich, dass 26 der 27 *cross-links* im erlaubten Rahmen waren. Der eine *cross-link*, welcher außerhalb des erlaubten Rahmens lag, ist zwischen der ersten Zinkfingerdomäne und Pcf11₅₄₉₋₆₃₁. Da beide an der Bindung beteiligten Aminosäuren auf der gleichen Seite der Oberfläche liegen, ist es denkbar, dass die modellierte Struktur eine gestreckte Konformation angenommen hat und der Konflikt durch Optimierung der Struktur durch eine Dynamiksimulation aufgelöst werden kann.

Da der zweite Zinkfinger als einzige klar definierte Struktur nicht dem Komplex zugeordnet werden konnte, wurde der mögliche Konformationsraum mittels DisVis berechnet [136]. Hierbei wird die Strukturkomplementarität unter Berücksichtigung der durch die *cross-links* definierten Entfernungsbeschränkungen berechnet. Alle möglichen Strukturmodelle werden zu einer Dichte, dem Konformationsraum, zusammengefasst. Es wird deutlich, dass der zweite Zinkfinger über einen sehr großen Bereich mit dem gebildeten Komplex interagieren kann, ohne die *cross-links* zu verletzen. Für eine genauere Platzierung sind daher deutlich mehr strukturelle Daten nötig. Dieses kann durch weitere *cross-link*-Experimente unter Verwendung alternativer *cross-linker* geschehen. Da gezeigt werden konnte, dass die Zinkfingerdomänen direkt an der Bindung der RNA beteiligt sind und eine Längenselektivität für 12 Nukleotide aufweisen, kann diese Information möglicherweise als Entfernungsbeschränkung beschrieben werden. Aufgrund des deutlich größeren Freiheitsgrades eines RNA-Rückgrades im Vergleich zum Peptid-Rückgrat ist die mathematische Beschreibung eines RNA-Moleküls deutlich aufwendiger [194].

4.2.2.3 CstF

4.2.2.3.1 Modellierung der monomeren Untereinheiten

Von den drei Untereinheiten lagen bereits einige publizierte Strukturen vor, diese stammen jedoch aus verschiedenen Organismen. Neben den humanen Strukturen der WD40-Domäne von CstF50 und dem RNA-Erkennungsmotiv (RRM) von CstF64 lagen Strukturen der CstF50-Dimerisierungsdomäne aus *Drosophila melanogaster*, die murine Struktur der HAT-Domäne von CstF77 und die dimere Struktur von Teilen von CstF64 und CstF77 aus Hefe vor. Aufgrund des hohen Grades der Konservierung dieser Proteine konnten erfolgreich von allen Strukturen die humanen Modelle generiert werden. Soweit vorhanden, wurden die generierten Modelle mit den experimentellen *cross-link*-Daten abgeglichen. Für die HAT-Domäne von CstF77 konnten von insgesamt 116 intramolekularen *cross-links* nur 8 nicht auf die Struktur kartiert werden. Auffällig ist, dass diese nahezu ausschließlich zwischen dem HAT-N und HAT-C Bereich liegen. Im Verlauf des nachfolgenden *Dockings* zeigte sich, dass die HAT-N-Domäne nicht die für die hexamere Bildung des CstF-Komplexes nötige Konformation besitzt. Da diese Struktur isoliert kristallisiert wurde, ist die Konformation möglicherweise ein Artefakt der Kristallstruktur.

Darüber hinaus konnten die beiden Strukturen von CstF64, die N-terminale RRM Domäne sowie die Scharnierregion, aufgrund ihrer Sequenzüberlagerung zu einer Struktur zusammengeführt werden. Dieser neu modellierte N-Terminus von CstF64 mit gebundenem CstF77 konnte ebenfalls durch die *cross-links* verifiziert werden.

Diese Strukturen deckten einen Großteil der gesamten Sequenz des CstF-Komplexes ab, darunter die wichtigen Domänen, verantwortlich für RNA Bindung und Dimerisierung (siehe Abbildung 3.32 A). Für die fehlenden Bereiche konnten keine Strukturhomologe identifiziert werden.

4.2.2.3.2 Docking

Alle genutzten *cross-links* lagen innerhalb der Teilstrukturen. Aufgrund der großen Anzahl an identifizierten *cross-links* wurden diese unter Standardbedingungen verwendet und keine Einschränkung bzgl. der beteiligten Aminosäuren gemacht.

Zur Generierung der Struktur wurden alle möglichen Paare, welche in den *cross-linking*-Experimenten quervernetzt wurden, mittels HADDOCK zusammengefügt. Dabei zeigten sich nur zwei Kombinationen von Strukturen, die einen vertrauenswürdigen Komplex bildeten: die dimere Struktur der WD40-Domäne von CstF50 zum einen mit der N-terminalen Domäne von CstF64 und zum anderen mit der HAT-Domäne von CstF77. Diese werden im Folgenden näher diskutiert.

CstF50 – CstF64 Dimer

Das CstF50 – CstF64 Dimer konnte in zwei Konformationen, die sich jedoch nur geringfügig unterschieden, zusammengefügt werden. Die Differenz von ca. 2 Å kann durch die Flexibilität der Proteine erklärt werden, sodass die Struktur mit der besten Bewertung ausgewählt wurde. Zur Verifizierung der Struktur wurden alle *cross-links* auf dieser Struktur kartiert. Von den 84 *cross-links* in dieser Struktur lagen 61 innerhalb und 23 außerhalb des erlaubten Rahmens.

Die *cross-links*, die außerhalb des erlaubten Rahmens lagen, widersprechen dem Modell nicht zwingend, sondern könnten auf die multimere Struktur des CstF-Komplexes zurückzuführen sein. Deutlich wird dieses dadurch, dass die an den *cross-links* beteiligten Aminosäuren auf gegenüberliegenden Seiten der Struktur liegen.

In der biochemischen Charakterisierung des CstF-Komplexes konnte gezeigt werden, dass CstF77 das zentrale Strukturelement darstellt und CstF50 und CstF64 nicht miteinander interagieren [74], [75]. Dieses kann durch das Strukturmodell untermauert werden, da in der Interaktionsfläche zwischen CstF50 und CstF64 die Helix CstF77₆₃₁₋₆₄₆ liegt. Diese interagiert über multiple Wasserstoffbrückenbindungen mit beiden anderen Proteinen und stabilisiert die Binding als "molekularer Kleber". Eine stabile Interaktion dieser Domänen ohne die Beteiligung von CstF77 ist daher nicht zu erwarten und deckt sich mit den biochemischen Vorkenntnissen über diesen Komplex.

CstF50 – CstF77 Dimer

Das Modell des Dimers der WD40 Domäne von CstF50 mit der HAT-Domäne von CstF77 zeigt, dass die WD40-Domäne mit der HAT-N-Domäne, welche nicht an der Dimerisierung von CstF77 beteiligt ist, interagiert. Die Interaktionsfläche zwischen diesen steht außerdem nicht in Konflikt mit dem Modell des CstF50-CstF64 Dimers. Zur Validierung des Komplexes wurden alle intermolekularen *cross-links* auf diese Struktur kartiert. Diese wurden, anders als bei den anderen Komplexen, in drei Kategorien eingeteilt. Zum einen alle, die unterhalb von 34 Å und damit in dem erlaubten Rahmen lagen. Darüber hinaus wurde eine erweiterte Kategorie von bis zu 40 Å

gebildet. Es zeigte sich, dass nur 3 von den 46 *cross-links* außerhalb dieses Bereiches liegen. Es ist davon auszugehen, basierend auf ihrer Lage in der HAT-C Domäne, dass diese mit der zweiten Kopie von CstF50 verknüpft sind.

CstF-Trimer

Dem CstF50-CstF77 Dimer konnte anschließend noch der N-terminale Bereich von CstF50, welcher für die Dimerisierung von CstF50 verantwortlich ist, hinzugefügt werden. Dieser liegt neben der HAT-N-Domäne von CstF77. Im hexameren Komplex liegen zentral also die Dimerisierungsdomänen von CstF77 und CstF50, während nach außen gerichtet die HAT-C-Domäne von CstF77 und die WD40-Domäne von CstF50 liegen (Abbildung 3.31, Abbildung 3.32, Abbildung 4.1). Durch Überlagerung der in beiden generierten Modellen, CstF50 – CstF77 sowie CstF50 – CstF64, aufgelösten WD40-Domäne von CstF50 konnte die finale trimere Struktur von CstF gebildet werden (Abbildung 3.32, Abbildung 4.1). Diese wurde durch die *cross-link*-Daten validiert. Insgesamt 208 *cross-links* waren im Einklang mit der Struktur, während 55 dem widersprachen. Diese sind jedoch höchstwahrscheinlich auf die hexamere Struktur zurückzuführen. Durch einfaches Überlagern der aufgelösten Dimerisierungsdomänen konnte kein hexamerer Komplex gebildet werden. Nimmt man die Struktur des HAT-C-Dimers als Vorlage, liegen die beiden dimeren CstF50-CstF64, basierend auf den α -Atomen des Lysin189 von CstF64 zwischen den ein *cross-link* identifiziert wurde, rund 80 Å auseinander (Abbildung A. 5 A). Dieses ist jedoch nicht im Einklang mit den experimentellen *cross-link*-Daten. Nutzt man die Dimerisierungsdomäne von CstF50 als Vorlage für einen hexameren Komplex, so kommt es zu einem Zusammenstoßen der Strukturen, was nicht möglich ist, und zum anderen widerspricht dieses Modell der Struktur der CstF77-HAT-Domäne (Abbildung A. 5 B). Die Ursache hierfür liegt in dem Ursprung der Substrukturen. Da diese häufig isoliert kristallisiert wurden, wurden möglicherweise Konformationen ausgebildet, die nicht der nativen Struktur entsprechen, sondern durch kristallographische Interaktionen stabilisiert wurden. Da HADDOCK während des *Dockings* die Strukturen nur zusammenfügt, nicht jedoch eine längere Dynamiksimulation durchführt, konnten diese Konformationskonflikte nicht aufgelöst werden. Zur Bildung des nativen CstF-Komplexes müssten unter Verwendung der *cross-links* zwischen den trimeren Substrukturen eine Dynamiksimulation durchgeführt werden. Hierbei könnte ein Komplex gebildet werden, der im Einklang mit den experimentellen Daten steht. Optimalerweise wird die hexamere Struktur durch weitere Daten verifiziert. Das in dieser Arbeit gereinigte Protein kann direkt genutzt werden, um mittels *negativ stain* Einzelpartikel-Elektronenmikroskopie oder -tomographie die Oberfläche abzubilden. In die daraus resultierende Struktur könnte das hier generierte Strukturmodell gelegt werden.

Eine schematische Darstellung des nativen CstF-Komplexes, basierend auf den hier vorgestellten Daten, ist in Abbildung 4.1 dargestellt.

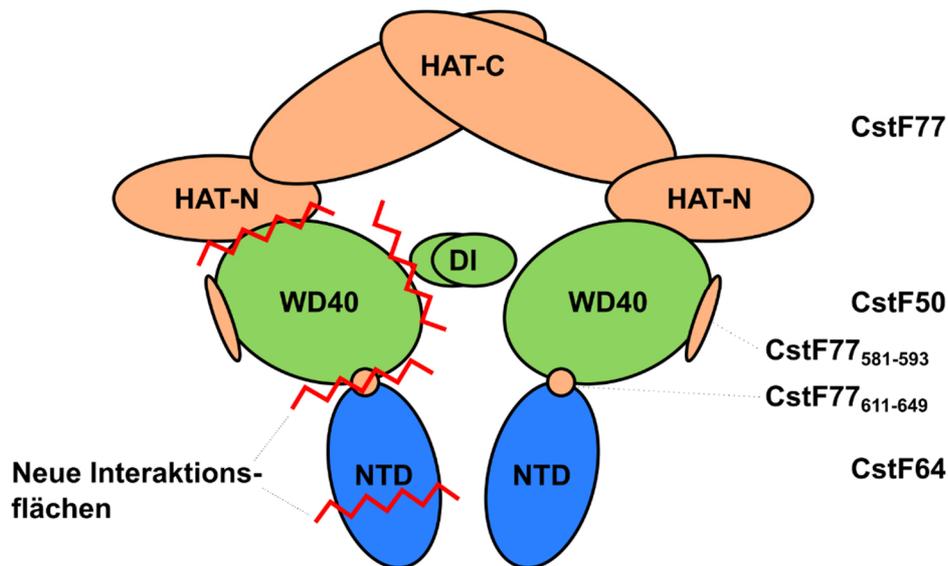


Abbildung 4.1 Schematische Darstellung des hexameren CstF-Komplexes. Das Schema basiert auf den beschriebenen Dimerisierungsdomänen von CstF50 (DI - Dimerisierungsdomäne) und CstF77 (HAT-C) sowie den in dieser Arbeit generierten *Docking*-Ergebnissen. Die Regionen zwischen den von CstF50 und CstF64 gebundenen Bereichen von CstF77 sind nicht aufgelöst, daher kann keine Aussage über die Lokalisation getroffen werden. Vormals unbekannte Interaktionsflächen sind mit roten Linien gekennzeichnet.

4.2.2.4 Modellierung des Endonuklease-Komplexes

4.2.2.4.1 Generierung der monomeren Untereinheiten

CPSF73 – CPSF100

CPSF73 und CPSF100 bilden das aktive Zentrum des Endonuklease-Komplexes. Beide Proteine gehören der β -CASP-Familie an. Gemein haben Proteine dieser Klasse, dass sie eine Metallo- β -Lactamase- und eine β -CASP-Domäne besitzen und im C-terminalen Bereich heterodimerisieren [109]. Dieses kann z.B. auch bei den Integrator-Proteinen Int9 und Int11 beobachtet werden [171].

Strukturell sind von beiden Proteinen die N-terminalen Metallo- β -Lactamase- und β -CASP-Domäne aufgelöst. Während von CPSF73 die humane Struktur kristallisiert wurde, ist für CPSF100 nur die Struktur des homologen Proteins aus Hefe bekannt. Aufgrund der hohen Sequenzidentität konnte hier jedoch ein Homologiemodell erstellt werden, welches die identifizierten intramolekularen *cross-links* bis auf eine Ausnahme erfüllt.

Für die C-terminalen Bereiche, welche für die Dimerisierung der beiden Proteine verantwortlich sind, ist keine Sequenz bekannt. Es zeigte sich jedoch, dass in beiden Proteinen eine Strukturhomologie zu der kristallisierten CTD-Interaktion der Integrator-Proteine Int9 und Int11 zu finden ist [171]. Int11 ist ein Paralog von CPSF73 und für die 3'-Prozessierung von nicht-codierenden RNAs verantwortlich. Da diese Proteine auch zur β -CASP-Familie gehören und

damit strukturell den CPSF-Untereinheiten sehr ähnlich sind, ist anzunehmen, dass auch die CTD-Interaktion konserviert ist. Die dimere Struktur der beiden CTDs von CPSF73 und CPSF100 wurde daher auf Grundlage der Kristallstruktur der CTDs der beiden Integrator-Proteine komodelliert. Da keine *cross-links* in dieser Struktur vorhanden waren, war eine Validierung des Modells nicht möglich.

Symplekin

Symplekin ist das größte Protein des Endonuklease-Komplexes, strukturell aber am schlechtesten charakterisiert. N-terminal liegt eine HEAT-Domäne, die mit der Phosphatase Ssu72 interagiert [119]. Die Struktur dieser Domäne konnte durch Röntgenkristallographie aufgeklärt werden. Darüber hinaus sind die Regionen von Symplekin, welche mit CstF64 und CPSF73 interagieren, grob kartiert, jedoch bisher nicht strukturell identifiziert. Leider konnten für beide Bereiche keine direkten Strukturhomologe identifiziert werden.

Im C-terminalen Bereich konnte eine konservierte Struktur identifiziert werden, in der sehr viele intramolekulare *cross-links* lagen. Auf Basis einer homologen Vorlagenstruktur konnte diese Domäne modelliert werden. Diese bildet eine weitere HEAT-Domäne aus antiparallelen α -Helices. Die Struktur dieser Domäne ist im Einklang mit den experimentellen *cross-link*-Daten. Aufgrund der nicht existierenden *cross-links* zu den anderen Proteinen des Komplexes kann spekuliert werden, dass diese Domäne nicht an der Komplexbildung beteiligt ist und auch nicht in räumlicher Nähe zu den anderen Proteinen liegt. HEAT-Domänen vermitteln vor allem Protein-Protein-Interaktionen, sodass davon auszugehen ist, dass diese Domäne mit anderen Proteinen der mRNA-3'-Prozessierung interagiert [195]. Diese Domäne könnte daher isoliert mit einem FLAG-tag exprimiert werden, um mögliche Interaktionspartner von Symplekin in HeLa-Kernextrakt zu identifizieren.

Für die Interaktionsregion von Symplekin zu CstF64 konnte kein Strukturhomolog identifiziert werden. Aus der Arbeit von *Ruepp et al.* ist jedoch bekannt, dass Symplekin und CstF77 um die Interaktion mit CstF64 kompetieren und dass die Region Symplekin₄₄₂₋₄₆₀ an der Bindung beteiligt ist [87]. Es wurde daher in dieser Arbeit postuliert, dass diese Region von Symplekin in strukturell ähnlicher Art und Weise wie CstF77 an die Scharnierdomäne von CstF64 bindet. Um die Interaktionsregion korrekt zu identifizieren, wurden auf die *cross-links* zwischen CstF64 und CstF77 bzw. Symplekin, welche in dieser Region lagen, zurückgegriffen. Das C-terminale Lysin in CstF64 (K₁₈₉) ist im Komplex mit CstF77 mit dem C-terminalen Lysin der Affenschwanzdomäne (K₆₄₇) quervernetzt. Eine Sequenzanalyse von Symplekin zeigte in der Region C-terminal von der postulierten Interaktionsregion drei Lysine: K₄₇₃, K₄₇₇ und K₄₈₀. Da jedoch nur ein *cross-link* von CstF64₁₈₉ zu Symplekin₄₇₇, nicht aber zu den beiden benachbarten Lysinen, identifiziert werden konnte, muss sich dieses in ähnlicher Position wie das Lysin₆₄₇ in CstF77 befinden. Möglich waren daher nur noch 5 theoretische Strukturen, von denen nur Modell4, Symplekin₄₄₃₋₄₈₀, das Lysin₄₇₇

in korrekter Position vorlag. Dieses liegt im Vergleich zu CstF77 eine Position N-terminaler, ist jedoch lösungsmittelzugänglich und in korrekter euklidischer Entfernung zum Lysin₁₈₉ in CstF64. Zur Verifizierung dieses Modells könnte diese Teilstruktur röntgenkristallographisch untersucht werden. Aufgrund der Größe ist eine erfolgreiche Expression in *E. coli* anzunehmen. Eine Koexpression könnte hierbei zu einer größeren Ausbeute führen.

4.2.2.4.2 Docking des Endonuklease-Komplexes

Für die Definition der Entfernungsbeschränkungen des *Dockings* des Endonuklease-Komplexes wurden im Gegensatz zu den vorherigen Komplexen weitere Methoden angewendet. Insgesamt 12 *cross-links* konnten ohne weiteres verwendet werden, 7 weitere *cross-links* konnten unter Verwendung der Peptid-Beschränkung berücksichtigt werden. Zusätzlich wurden die nicht aufgelösten Strukturbereiche in CPSF73 sowie in Symplekin durch I-TASSER strukturell berechnet. Der Durchmesser des so generierten Modells wurde als Entfernungsbeschränkung genutzt. Dieses stellt eine weitere neue Methode in der Definition von Entfernungsbeschränkungen dar: die automatisierte Strukturvorhersage einer unbekanntenen Region zur Ermittlung der durchschnittlichen Entfernung. Hierbei muss jedoch beachtet werden, dass man unter Umständen die reale Größe der unbekanntenen Struktur unterschätzt und so eine zu restriktive Beschränkung definiert, die zu einem fehlerhaften Modell führt. Eine empirische Validierung dieser Methode steht noch aus.

Im ersten *Docking*-Schritt konnte die β -CASP-Domäne von CPSF100 mit dem CTD-Dimer von CPSF73 und CPSF100 zusammengefügt werden. Die Clustergröße von 147 von 200 Strukturen spricht hier auch für eine eindeutige Positionierung. Das generierte Modell konnte mit sieben *cross-links* validiert werden. Basierend auf diesem Strukturmodell wurde anschließend die dimere Struktur von Symplekin und CstF64 hinzugefügt. Obwohl nur drei *cross-links*, von denen einer durch die Verwendung der Peptid-Beschränkung genutzt werden konnte, zur Verfügung standen, konnte ein Cluster mit 191 von 200 Strukturen generiert werden, was für eine eindeutige Positionierung spricht. Im finalen *Docking*-Experiment wurde dieser Struktur die β -CASP-Domäne von CPSF73 hinzugefügt. Hierbei wurden 6 *cross-links* verwendet, von denen jedoch nur einer direkt in der Struktur aufgelöst vorlag. Zusätzlich wurde der mittels I-TASSER generierte Abstand zwischen der β -CASP-Domäne und der CTD genutzt. Obwohl kein eindeutiger Cluster gebildet werden konnte, zeigte eine Detailanalyse der 200 generierten Modelle jedoch, dass 143 Strukturen eine sehr ähnliche Konformation aufzeigten. Hierbei wird deutlich, dass trotz Einordnung in verschiedene Cluster eine klare Positionierung der Struktur von CPSF73 möglich war. Da parallel zu dieser Arbeit von *Zhang et al.* der CPSF-Komplex mittels kryoelektronenmikroskopie untersucht wurde, konnte zusätzlich auf diese Daten zurückgegriffen werden [98]. Während in dieser Publikation der mPSF-Komplex mit einer Auflösung von 3 Å aufgelöst werden konnte, zeigte der Endonuklease-Komplex nur eine Auflösung von 7,4 Å. Diese

Dichte wurde genutzt, um das in dieser Arbeit gebildete Modell des Endonuklease-Komplexes zu validieren und zu erweitern. Es zeigte sich, dass ohne Strukturoptimierung das Modell mit einer Übereinstimmung von 70 % in die Dichte kartiert werden konnte. Zur weiteren Optimierung der Struktur könnte diese durch Programme wie *Coot* weiter optimiert werden, was jedoch den zeitlichen Rahmen dieser Arbeit überschritt [196]. Da in dieser Publikation zusätzlich der C-terminale Teil von Symplekin enthalten war, konnte das in dieser Arbeit generierte Strukturmodell der HEAT-Domäne trotz fehlender intermolekularer *cross-links* dem Komplex zugeordnet werden. Da ein molekulares *Docking* der N-terminalen HEAT-Domäne von Symplekin nicht möglich war, wurde der Konformationsraum mittels DisVis analysiert. Deutlich wird, dass diese Region von allem mit der C-terminalen Region der CPSF Untereinheiten interagieren kann.

Auffällig in der finalen Struktur ist das Fehlen von *cross-links* in CPSF73. In dem Modell fehlen große Teile der Struktur von Symplekin und CstF64. Diese maskieren möglicherweise in der Quartärstruktur die Oberfläche von CPSF73. Dieses wird durch den Konformationsraum des N-terminalen Bereiches von Symplekin nur zum Teil erklärt. Zur genauen Kartierung der Oberfläche bzw. der Lösungsmittelzugänglichkeit des Endonuklease-Komplexes können massenspektrometrische *foot-printing*-Methoden verwendet werden. Hierzu gehören unter anderem der Wasserstoff-Deuterium-Austausch (HDX). Hierbei wird ein Proteinkomplex mit schwerem Wasser markiert und anschließend massenspektrometrisch analysiert [197]. Diese Methode ist jedoch technisch anspruchsvoll, da die gesamte Analyse unter sauren und möglichst kalten Bedingungen durchgeführt werden muss, um ein *scrambling*, also den Austausch des Deuteriums auf andere, eigentlich nicht oberflächenexponierte Peptide oder Aminosäuren zu verhindern. Alternativ kann die Oberfläche von Proteinen bzw. Proteinkomplexen durch photochemische Oxidation mit Wasserstoffperoxid markiert werden (FPOP) [198]. Hierbei kann die massenspektrometrische Analyse unter Standardbedingungen durchgeführt werden. Beide Methoden erlauben die Beschreibung der Oberflächen und der daran beteiligten Aminosäuren in den Proteinen.

4.2.3 Zusammenfassung der Modellierung

Im Rahmen dieser Arbeit wurden für die Komplexe CFII-, CstF- und den Endonuklease-Komplex molekulare Strukturmodelle auf Grundlage der experimentellen *cross-links* erstellt. Die Möglichkeiten der Strukturbildung sind hierbei limitiert durch die Anzahl der identifizierten *cross-links* aber vor allem auch durch die Qualität der zugrundeliegenden Substrukturen.

Für die allgemeine Verwendung von *cross-link*-Informationen konnte hier eine Peptid-Beschränkung definiert werden, die die Generierung der hier vorgestellten Modelle zum großen Teil erst möglich gemacht hat. Diese basiert auf der datenbankweiten Analyse von C α -Entfernungen von Peptiden mit bis zu 10 Aminosäuren. Darüber hinaus wurde die Länge der an den *cross-links* beteiligten Aminosäuren bestimmt, sodass die Entfernungsbeschränkung sehr

genau definiert werden kann. Die Nutzung dieser Informationen bedarf aktuell aber noch der aufwendigen manuellen Annotation, sodass vor allem die Länge der Aminosäuren nur im Falle des CFII-Komplexes angewendet wurde. Es sollte daher eine Automatisierung angestrebt werden, in der nur noch die Struktur und die Liste der *cross-links* nötig ist, und die Identifikation der Aminosäure sowie die Anwendung der Peptid-Beschränkung selbstständig durchgeführt wird. Darüber hinaus können größere Strukturlücken durch die Verwendung von Programmen wie I-TASSER beschrieben werden, was die Modellierung von Proteinstrukturen weiter verbessern kann. Diese sollte systematisch analysiert und optimiert werden.

Die Strukturmodellierung des CFII-Komplexes war durch die unbekannte Struktur von Pcf11 limitiert. Es konnten nur 4 Sequenzbereiche mit einer klar definierten Struktur und intermolekularen *cross-links* identifiziert werden. Bis auf eine Zinkfingerdomäne konnten diese jedoch erfolgreich an die monomere Struktur von Clp1 *gedockt* werden.

Als äußerst erfolgreich kann die Strukturmodellierung des CstF-Komplexes angesehen werden. Für diesen Komplex waren sehr viele *cross-links* identifiziert worden und große Teile der (z.T. homologen) Strukturen lagen vor. Dass die *cross-links* aufgrund der hexameren Struktur nicht eindeutig zugeordnet werden konnten, hatte keinen nennenswerten Einfluss auf die Komplexbildung. Das generierte trimere Modell ist sowohl im Einklang mit den *cross-link*-Daten als auch mit den biochemischen Vorkenntnissen über diesen Komplex. So kann durch dieses Modell die Beobachtung, dass CstF64 leicht von dem Komplex dissoziieren kann und das CstF77 das Gerüstprotein des Komplexes darstellt, erklärt werden. Die Bildung der nativen hexameren Struktur war aufgrund der starren Strukturen nicht möglich, jedoch bilden diese Ergebnisse gute Grundlagen für eine schnelle Strukturaufklärung in Kombination mit weiteren strukturgebenden Methoden.

Der Endonuklease-Komplex vermittelt die Spaltungsaktivität, welche essentiell für den ersten Schritt der mRNA-3'-Prozessierung ist. Im Fokus dieser Arbeit stand jedoch die Interaktion mit den anderen Proteinen dieses Komplexes und nicht der Mechanismus der prä-mRNA-Spaltung, in der CPSF73 eine Konformationsänderung durchläuft [96], [111]. Für die Interaktion konnte ein Kondensationskern identifiziert werden, an denen die CTDs von CPSF73 und CPSF100 sowie die dimere Interaktionsregion von Symplekin und CstF64 beteiligt sind. Trotz seiner Größe sind bei der Struktur von Symplekin ähnliche Probleme wie bei Pcf11 zu beobachten. Es konnten entweder keine Homologiemodelle oder keine intermolekularen *cross-links* identifiziert werden, sodass außer der bereits genannten Struktur eine weiteren eindeutig positioniert werden konnte. Weitere strukturgebende Experimente könnten hier helfen, die native Struktur des Endonuklease-Komplexes aufzuklären.

4.3 Interaktomanalyse

Die Identifikation von Interaktionspartnern über die Affinitätsreinigung von Köderproteinen aus Zellysate ist eine sehr verbreitete Methode. Als Köderproteine dienten in dieser Arbeit alle an der Polyadenylierung beteiligten Faktoren, wobei unterschiedliche *tags* zur Affinitätsreinigung aus Kernextrakt dienten. Hierbei zeigte sich, dass der FLAG-*tag* aufgrund seiner hochaffinen Anreicherung durch einen Antikörper am besten geeignet ist. Die Proteinidentifizierung erfolgte in allen Fällen mittels Massenspektrometrie, da diese Methode als einzige in der Lage ist, sehr viele verschiedene Proteine in einer semiquantitativen Art und Weise zu detektieren. Da alle analysierten Komplexe RNA-Bindung zeigten, ist davon auszugehen, dass nicht alle identifizierten Interaktionspartner über direkte Protein-Protein-Interaktionen mit dem Köderprotein assoziiert sind. Da sich jedoch der 3'-Prozessierungskomplex auch dadurch auszeichnet, dass die verschiedenen Komplexe an unterschiedliche Sequenzelemente der prä-mRNA binden, und die RNA daher als Gerüst dient, ist von einer RNase-Behandlung des Kernextraktes abgesehen worden. Es ist daher davon auszugehen, dass vor allem RNA-bindende Proteine unspezifisch angereichert wurden.

Zur groben Bewertung der Qualität der identifizierten Proteine wurden diese als Verteilungskurve dargestellt. Angereicherte Proteine resultieren in einer Rechtsschiefe der Glockenkurve. Bei wenig abundanten Proteinen, die substöchiometrisch mit dem Köderprotein interagieren, ist dieser Effekt jedoch nur sehr gering ausgeprägt, und eine Identifikation von angereicherten Proteinen musste daher immer individuell im Vergleich zur Kontrolle stattfinden.

Zur Klassifizierung wurde zum einen eine Liste bekannter und putativer Polyadenylierungsfaktoren genutzt. Diese Liste basiert auf den biochemischen Vorkenntnissen der Proteine sowie auf homologen Proteinen der 3'-Prozessierung in Hefe. Zum anderen wurde eine automatische Einordnung anhand der *GO terms* durchgeführt. Diese basiert zum einen auf der manuellen Annotation basierend auf publizierten Daten, zum anderen auf der automatischen Annotation durch Sequenzvergleiche mit besser charakterisierten Proteinen. Es ist daher davon auszugehen, dass diese Klassifizierung sowohl zu falsch positiven wie falsch negativen Ergebnissen führt. Vor allem gänzlich unbekannte Faktoren der Polyadenylierung können so nicht identifiziert werden.

4.3.1 Interaktom des CFII-Komplexes

Zur Identifizierung von Interaktionspartnern des CFII-Komplexes wurden zwei parallele Ansätze verfolgt. Zum einen wurde das vollständige Protein, welches an Clp1 einen Hexahistidin-*tag* trug, verwendet. Aufgrund des unspezifischen Hintergrundes konnten diese Daten jedoch nicht ausgewertet werden. Zum anderen wurden zwei N-terminale Deletionsvarianten von Pcf11 im Komplex mit Clp1 verwendet. Die längere Variante (CFII Δ N769) konnte in fraktioniertem Extrakt die Spaltung der prä-mRNA rekonstituieren, während die kürzere Variante (CFII Δ N1123) dazu

nicht in der Lage war. Da keine signifikanten Unterschiede in der RNA-Bindung dieser beiden Variante zu detektieren war, wird diese Aktivität vermutlich über Protein-Protein-Interaktionen vermittelt. Es sollte also die vermutete Interaktion der FEGP-Wiederholungen identifiziert werden, welche für die Spaltung der prä-mRNA essentiell ist.

4.3.1.1 Deletionsvarianten

Zur Durchführung des *Pulldown*-Experiments wurden die beiden Deletionsvarianten von Pcf11 im Komplex mit Clp1, CFII Δ N769 und CFII Δ N1123, zusätzlich zudem schon vorhandenen Hexahistidin-*tag* mit einem FLAG-*tag* an Clp1 versehen. Dieses ermöglicht eine deutlich bessere Identifikation von spezifisch gebundenen Proteinen, da die Reinigung durch den FLAG-Antikörper sehr spezifisch ist. Die qualitative Analyse des *Pulldowns* mittels SDS-Polyacrylamidgel unterstreicht dieses. Während in der Kontrolle bis auf den Antikörper keine angereicherten Proteine zu beobachten sind, können in den *Pulldowns* mit Köderprotein weitere klare Banden identifiziert werden. Die Verteilungskurve der massenspektrometrischen Analyse zeigt zudem eine deutliche Rechtsschiefe.

Es konnten insgesamt 177 Proteine identifiziert werden, die in der spaltungsaktiven Variante CFII Δ N769 gegenüber CFII Δ N1123 und der Pufferkontrolle angereichert waren. Dieses schließt auch die Proteine ein, die ausschließlich in CFII Δ N769 identifiziert wurden. Die auffälligste Identifikation ist die der Poly(A)-Polymerase. Diese ist essentiell für beide Teilschritte der 3'-Prozessierung. Während in der Polyadenylierung die Poly(A)-Polymerase mit der Fip1-Untereinheit des mPSF interagiert, ist die Lokalisation der Poly(A)-Polymerase während der Spaltungsreaktion nicht bekannt. Die hier dargestellten Daten deuten darauf hin, dass die Poly(A)-Polymerase durch die FEGP-Wiederholungen von Pcf11 mit dem 3'-Prozessierungskomplex assoziiert sein könnte. In Hefe konnte in einem *Pulldown* von yPcf11, welche sich in dem CFIA Komplex befindet, auch die Poly(A)-Polymerase (PAP1) nachgewiesen werden [67]. Da hier jedoch auch Ysh1, welches einen Teil des Endonuklease-Komplexes darstellt, in gleichem Maße identifiziert wurde, ist eine direkte Interaktion zwischen yPcf11 und PAP1 nicht eindeutig geklärt. yPcf11 verfügt darüber nicht über die FEGP-Wiederholungen.

Über die RNA-Helikasen DDX55 und DDX19A ist keine Beteiligung an der Polyadenylierung berichtet, die RNA-Helikase Aquarius ist Teil des Spleißosoms.

Fasst man die identifizierten Proteine der beiden Deletionsvarianten zusammen, erhält man partielle Informationen über das Interaktom von CFII. Hierbei sind direkte Interaktionen von Clp1, dem C-Terminus von Pcf11 sowie indirekte, über RNA vermittelte Interaktionen zu beobachten. Die Detailanalyse zeigte, dass CFII mit den basalen Polyadenylierungsfaktoren CPSF, inklusive Ssu72 und Rbbp6, welches möglicherweise CPSF-assoziiert ist, CstF und CFI interagiert. Eine sehr starke Interaktion konnte mit XRN2 beobachtet werden. Diese Exonuklease ist für den Abbau des 3'-Fragmentes der gespaltenen mRNA und die Termination der Transkription

verantwortlich [199]. Da auch für Pcf11 ein Einfluss auf die Termination gezeigt werden konnte, ist dieses möglicherweise ein kooperativer Prozess. Auch konnte eine Anreicherung der Proteine Nono und SFPQ (PSF) beobachtet werden. Es wurde berichtet, dass diese einen trimeren Komplex mit XRN2 bilden, welcher die Polyadenylierungsreaktion stimulieren soll [200]. Zusätzlich wurde mit ähnlicher Anreicherung das Protein CARF identifiziert. Es konnte gezeigt werden, dass dieses mit XRN2 interagiert und den Export von XRN2 aus dem Kernplasma inhibiert, welcher relevant für die rRNA-Prozessierung ist [201]. Möglicherweise wird über CARF so auch die Transkriptionstermination durch Kontrolle der XRN2-Konzentration reguliert.

Darüber hinaus konnten Proteine des tRNA-Endonuklease-Komplexes identifiziert werden. Clp1, nicht jedoch Pcf11, ist Teil dieses Komplexes und für die Funktion essentiell [59]. Eine Identifikation von Proteinen dieses Komplexes war daher erwartet.

Da die darüber hinaus identifizierten Komplexe, namentlich der PP1-, der PAF-, der THO-Komplex und der Cap-bindende-Komplex (CBC) sowie SR-reiche Spleißfaktoren auch in den *Pulldowns* des CstF- und des Endonuklease-Komplexes identifiziert wurden, wird ihre Rolle in der 3'-Prozessierung später diskutiert (4.3.7).

4.3.2 Interaktom des CstF-Komplexes

Das *Pulldown*-Experiment mit dem CstF-Komplex, welcher im Gegensatz zu dem in 3.1.3 beschriebenen Komplex je eine Kopie CstF64 und CstF64tau enthielt, wurde mit dem FLAG-*tag* an CstF64tau durchgeführt. Der Komplex wurde von Moritz Schmidt zur Verfügung gestellt. Vergleichbar mit dem *Pulldown* der CFII-Deletionsvarianten ist in dem analytischen SDS-Polyacrylamidgel neben den Köderproteinen eine klare Anreicherung von potentiellen Interaktionspartnern zu beobachten, während in der Pufferkontrolle nur die Proteine des FLAG-Antikörpers zu beobachten sind. In der quantitativen Analyse der identifizierten Proteine konnte eine Rechtsschiefe der Verteilungskurve festgestellt werden. In der qualitativen Analyse der identifizierten Proteine zeigten sich erneut die Kernfaktoren der Polyadenylierung. Hierbei ist der CPSF- und der CFII-Komplex am stärksten angereichert, die Proteine Ssu72 und Rbbp6 lagen substöchiometrisch vor. Der CFI-Komplex konnte nur mit geringerer Intensität identifiziert werden. Keine Identifikation konnte für die Poly(A)-Polymerase erfolgen, dafür jedoch für die beiden Poly(A)-bindenden Proteine PABPC1 und PAPBN1, welches an der Polyadenylierung beteiligt ist.

Interessant ist die Anreicherung von Nucleophosmin (NPM). Es konnte gezeigt werden, dass dieses Protein stromaufwärts der Polyadenylierungsstelle sequenzunabhängig, jedoch PAS-abhängig, bindet. An einer RNA mit mutiertem PAS konnte keine Bindung von NPM identifiziert werden, sodass eine Interaktion mit CPSF, nicht jedoch mit CstF, postuliert ist [202].

Die darüber hinaus identifizierten Komplexe (PP1a-, PAF-, CBC- und THO-Komplex) sowie die SR-reichen Spleißfaktoren werden später diskutiert.

4.3.3 Interaktom des CFI-Komplexes

Zur Identifikation von Interaktionspartnern von CFI wurde das in 3.1.4 gereinigte Protein verwendet. Dieses enthielt neben dem für die Reinigung genutzten Hexahistidin-*tag* an CFI25 zusätzlich einen Strep-*tag* an CFI68. Dieser *tag* basiert auf einer Interaktion zwischen dem Peptid und einem modifiziertem Streptavidin [203]. Die Interaktion ist deutlich spezifischer als die Verwendung eines Hexahistidin-*tags*. Das analytische SDS-Polyacrylamidgel zeigte deutlich die Köderproteine sowie eindeutige Proteinbanden. In der Pufferkontrolle waren diese deutlich schwächer zu beobachten. Die quantitative Analyse der identifizierten Proteine spiegelt das Ergebnis des Acrylamidgeles aber nicht wider. Hier war eine deutlich stärkere Identifikation von Proteinen in der Kontrolle gegenüber den CFI-*Pulldowns* zu beobachten. Die Ursache hierfür ist unklar. Aufgrund des sehr starken Hintergrundes konnten nur sehr wenige Proteine als angereichert identifiziert werden. Neben den beiden Köderproteinen wurde von den bekannten Polyadenylierungsfaktoren nur CstF50 identifiziert. CFI, welche sich aus CFI25/CFI68 zusammensetzte, konnte in gereinigten Spleißosomen identifiziert werden [204], [205]. Neben anderen Spleißfaktoren konnten die Paralogen Proteine Tra2 α und Tra2 β identifiziert werden. Eine Interaktion von CFI mit Tra2 β ist beschrieben und wird durch die R/S-Domäne in CFI68, ein Sequenzmerkmal von Spleißfaktoren, vermittelt [47]. Die Tra2-Proteine werden mit alternativem Spleißen in Verbindung gebracht. Dieses ist dahingehend interessant, dass CFI mit dem Prozess der alternativen Polyadenylierung, einem anderen genregulatorischen Prozess, in Verbindung gebracht wird [206], [207].

Von den bekannten Hauptfaktoren der Polyadenylierung bindet CFI am weitesten stromaufwärts und ist somit theoretisch der erste Polyadenylierungskomplex, welcher an der noch wachsenden prä-mRNA binden kann. Die beobachtete Interaktion mit Spleißfaktoren verstärkt dieses dadurch, dass CFI mutmaßlich schon mit dem Spleißosom assoziiert ist, welches das letzte Intron spaltet. Hierdurch könnte es zu einer sehr effizienten Kopplung der verschiedenen mRNA-Reifungsprozesse kommen.

4.3.4 Interaktom des mPSF-Komplexes

Der zur Interaktomanalyse verwendete Komplex, welcher wie in 3.1.3.1 beschrieben, gereinigt und charakterisiert wurde, verfügte nur über einen Hexahistidin-*tag* an WDR33. Aufgrund des hohen Hintergrundes, wie es auch schon beim Hexahistidin-*tag*-tragendem CFII-Komplex zu beobachten war, kann keine Aussage über Interaktionspartner getroffen werden. Zur Identifikation von Interaktionspartnern von mPSF sollte stattdessen eine Variante mit FLAG-*tag* benutzt werden, bei der das Verhältnis zum Hintergrund deutlich besser ist.

4.3.5 Interaktom des Endonuklease-Komplexes

Zur Interaktomanalyse wurde der in 3.1.3.2 beschriebene Endonuklease-Komplex, welcher an der putativen Untereinheit CstF64 einen FLAG-*tag* trug, genutzt. Das analytische SDS-Polyacrylamidgel war vergleichbar mit den anderen FLAG-*Pulldown* Experimenten. In der Pufferkontrolle waren neben der leichten und schweren Kette des verwendeten Antikörpers keine deutlich angereicherten Proteine zu sehen. Dem gegenüber konnten in dem Experiment neben den Köderproteinen weitere Proteine beobachtet werden. Dieses ist auch durch die quantitative Analyse der massenspektrometrischen Daten sichtbar, in denen man eine deutliche rechtsschiefe Verteilung der Intensitäten beobachten konnte. In der qualitativen Analyse wurde deutlich, dass neben den Köderproteinen der CstF-Komplex mit ähnlicher Intensität angereichert werden konnte. Dieses ist möglicherweise auf die Doppelrolle von CstF64 in beiden Komplexen zurückzuführen. Auch war die Phosphatase Ssu72 ähnlich stark angereichert. Diese interagiert mit Symplekin, welches u.a. durch die Aufklärung der dimeren Struktur von Symplekin und Ssu72 gezeigt werden konnte [119]. Mit reduzierter Intensität gegenüber den Köderproteinen, jedoch deutlich angereichert gegenüber der Kontrolle, liegen die anderen basalen Polyadenylierungsfaktoren mPSF, CFI und CFII vor.

Die Rolle der weiteren identifizierten Komplexe wird später diskutiert.

4.3.6 Basale Faktoren der Polyadenylierung

Die bekannten basalen Polyadenylierungsfaktoren konnten durch die *Pulldown*-Experimente bestätigt werden. So konnten in allen Experimenten der CFI-, CFII, CstF- sowie der CPSF-Komplex identifiziert werden. Diese bilden den bekannten Kern der Polyadenylierungsmaschinerie. Basierend auf der Intensität lagen CFII, CstF und CPSF äquimolar vor, während CFI unterrepräsentiert war. Ob dieses auf einem tatsächlich substöchiometrischen Komplex oder auf geringerer Affinität beruht, kann nicht geklärt werden. Neben diesen bekannten Komplexen konnte PABPN1, nicht jedoch die Poly(A)-Polymerase, in allen Analysen identifiziert werden. Während PABPN1 für die korrekte Polyadenylierung benötigt wird, ist die Poly(A)-Polymerase essentiell für sowohl die Spaltung als auch die Polyadenylierung. Dass diese nur in dem Interaktom von CFII Δ N769, nicht jedoch in anderen Analysen identifiziert wurde, deutet auf eine transiente oder sehr schwache Interaktion zwischen der Polymerase und den anderen Faktoren hin. Möglicherweise stellt dieses einen regulatorischen Mechanismus der Polyadenylierung dar. Als CPSF-assoziiert und als Teil des Polyadenylierungskomplexes gelten darüber hinaus Ssu72 und Rbbp6, wobei unklar ist, ob dieses tatsächlich CPSF-assoziiert ist. Während Ssu72 konsistent in allen Experimenten deutlich angereichert vorlag, war für Rbbp6 nur eine substöchiometrische Identifikation möglich. Dieses überrascht, da das Hefehomolog Mpe1 essentiell für die Polyadenylierung und die Aktivierung der Endonuklease ist [96], [208]. Das jedoch in dem Interaktom des Endonuklease-Komplexes Rbbp6 im Vergleich zu CFII und CstF

am geringsten angereichert vorlag, deutet auf eine unterschiedliche Aktivierung der Endonuklease in humanen Zellen im Vergleich zu Hefe hin.

4.3.7 Weitere putative Polyadenylierungsfaktoren

In der zusammenfassenden Analyse der FLAG-*Pulldowns* von CFII, CstF und des Endonuklease-Komplexes konnten weitere Proteine und Komplexe identifiziert werden, die im Folgenden detailliert diskutiert werden.

4.3.7.1 Poly(A)-bindende Proteine

Neben dem nuklearen PABPN1, welches für die Längenkontrolle in der Synthese des Poly(A)-Schwanzes benötigt wurde, konnte in allen *Pulldowns* das zytoplasmatische PABPC1 identifiziert werden. Es ist bekannt, dass PABPC1 unter Stresssituationen auch im Nukleus lokalisiert ist und zu einer Hyperadenylierung von mRNA führt, welche den Export inhibiert [209], [210]. Darüber hinaus konnte gezeigt werden, dass in HeLa-Zellen, welche für die Herstellung des in den *Pulldown*-Experimenten genutzten Kernextraktes verwendet wurden, PABPC1 auch nuklear lokalisiert ist [211], [212]. Da eine Interaktion zwischen den verwendeten Köderproteinen und PABPC1 nicht bekannt ist, ist die Identifikation vermutlich auf eine indirekte, RNA-vermittelte Interaktion zurückzuführen.

4.3.7.2 PP1-Phosphatase-Komplex

Der PP1-Phosphatase-Komplex besteht aus den Proteinen PNuts, Tox4, WDR82 sowie einer von drei Isoformen der katalytischen Untereinheit (PP1 α , PP1 β oder PP1 γ) [213]. Die Zusammensetzung dieses Komplexes unterscheidet sich deutlich von der des Phosphatase-Komplexes in Hefe, dem sogenannten APT-Komplex (*associated with Pta1*) [214]. Der APT-Komplex besteht aus Pta1, dem Homolog zu Symplekin, Ssu72, Glc7, dem Homolog zu PP1 α , Swd2/Cps35, dem Homolog zu WDR82, sowie den beiden Proteinen Ref2 und Pti1, zu denen kein Homolog in Säugern bekannt ist [215]. Von den Proteine PNuts und Tox4 im humanen PP1-Phosphatase-Komplex sind wiederum keine Homologe in Hefe bekannt [213]. Der PP1-Komplex ist sowohl dem Spleißen als auch der Polyadenylierung zugeordnet [25], [216], [217]. Auch konnte gezeigt werden, dass durch die Dephosphorylierung von CFI oder CFII die Spaltung der prä-mRNA reversibel inhibiert werden kann [217].

Die Untereinheiten des PP1-Phosphatase-Komplexes konnten in allen *Pulldowns* identifiziert werden. Die stärkste Anreicherung wurde im Interaktom des Endonuklease-Komplexes beobachtet. Hier waren neben den katalytischen Untereinheiten auch die regulatorischen Proteine WDR82 und Tox4 sehr stark angereicht. Diese waren in den Interaktomen vom CFII und CstF nicht zu beobachten. Es ist daher anzunehmen, dass der PP1-Phosphatase-Komplex über den Endonuklease-Komplex mit der 3'-Prozessierung assoziiert ist. Ferner kann man aufgrund

der sehr hohen Anreicherung von WDR82 und Tox4 spekulieren, dass die direkte Interaktion über diese beiden Proteine stattfindet.

In der Verteilung der verschiedenen katalytischen Untereinheiten ist für PP1 β die stärkste Anreicherung zu beobachten, während PP1 γ am geringsten angereichert ist. Die unterschiedlichen katalytischen Untereinheiten besitzen sowohl überlappende als auch spezifische Funktionen und unterscheiden sich bezüglich ihrer Assoziation mit unterschiedlichen regulatorischen Untereinheiten [218]–[220]. Die Anreicherung von PP1 β bedeutet nicht zwangsläufig, dass dieses die bevorzugte katalytische Untereinheit der Polyadenylierungsmaschinerie darstellt. Hierzu muss die Anreicherung in Verhältnis zu der Abundanz der Untereinheiten in dem verwendeten HeLa-Kernextrakt gesetzt werden, um eine spezifische Anreicherung zu identifizieren.

4.3.7.3 PAF-Komplex

Der humane PAF-Komplex besteht aus den fünf Untereinheiten Paf1, Ctr9, Leo1, Cdc73 und WDR61, welches ursprünglich als Ski8 beschrieben wurde [221]. Der homologe Komplex in Hefe interagiert direkt über yCdc73 und yRtf1, von dem kein humanes Homolog identifiziert werden konnte, mit der aktiven RNA-Polymerase II, aber hat jedoch keinen Einfluss auf die Elongation [222]. Der humane PAF-Komplex interagiert zudem mit dem Ski-Komplex, welcher für den Abbau von RNA durch das Exosom verantwortlich ist [221], [223]. Eine Deletion des PAF-Komplexes hat nur Einfluss auf eine geringe Anzahl von Zielgenen, sodass auszuschließen ist, dass der PAF-Komplex einen essentiellen Bestandteil der mRNA-Prozessierung darstellt [221], [222]. Daraus wurde postuliert, dass der PAF-Komplex eine Rolle in der Qualitätskontrolle von mRNA spielt und dieser Prozess ko- und nicht posttranskriptionell stattfindet.

Identifiziert wurde der PAF-Komplex mit allen Untereinheiten in dem *Pulldown* des Endonuklease-Komplexes, wobei hier Paf1 und WDR61 deutlich angereichert vorlagen. In den Interaktomen von CFII und CstF konnte hingegen nur Paf1 als signifikant angereichert identifiziert werden, wohingegen die anderen Faktoren nur mit sehr geringen Intensitäten vorlagen. Hieraus lässt sich vermuten, dass der PAF-Komplex über Paf1 und eventuell WDR61 mit dem Endonuklease-Komplex interagiert, nicht jedoch mit CFII oder CstF, wo die Identifikation auf eine indirekte, über den Endonuklease-Komplex-vermittelte Interaktion zurückzuführen sein könnte. Dieses ließe sich durch eine analytische Gelfiltration des hier beschriebenen Endonuklease-Komplexes in Kombination mit rekombinantem Paf1, welches sich in *E. coli* exprimieren lässt, verifizieren [224].

4.3.7.4 TREX/THO-Komplex

Der TREX/THO-Komplex besteht mindestens aus den sieben THO-Untereinheiten (Thoc1 bis Thoc7), den RNA-Helikasen DDX39A und DDX39B, sowie den Proteinen SARNP und CHTOP [225], [226]. Darüber hinaus sind die Proteine PDIP3 und ZC11A sowie die Exportfaktoren NXF1

und NXT1, welche ein Heterodimer bilden, mit dem TREX/THO-Komplex assoziiert [225], [227]. Der TREX/THO-Komplex ist wichtig für den Export von mRNA, die genaue Funktion ist jedoch nicht bekannt [228]. Der Komplex interagiert spezifisch mit gespleißter RNA, mutmaßlich über eine direkte Interaktion mit dem EJC-Spleißkomplex, welcher nach Entfernung der intronischen Region weiterhin mit der mRNA assoziiert vorliegt [226], [229]. Zusätzlich interagiert er mit verschiedenen Komplexen sowohl am 5'-Ende als auch am 3'-Ende von mRNA. So konnte eine direkte Interaktion zwischen dem Cap-bindenden Komplex CBC und Untereinheiten des THO-Komplexes identifiziert werden [230]. Auch interagiert Thoc4 (auch bekannt als Aly/REF) sowohl in Hefe als auch im Menschen direkt mit der C-terminalen Domäne von Pcf11 [231].

In allen analysierten Interaktomen konnten der TREX/THO-Komplex, beide Isoformen der Helikase DDX39, sowie alle assoziierten Proteine mit Ausnahme von NXT1 identifiziert werden. Einen Sonderfall stellt SARNP dar, da dieses sowohl in dem Interaktom von CFII als auch in CstF im Vergleich zur Kontrolle abgereichert vorliegt. Basierend auf der Anreicherung und der Intensität interagiert der TREX/THO-Komplex sowohl mit CFII als auch mit CstF. Die geringe Anreicherung in dem *Pulldown* des Endonuklease-Komplexes ist daher vielleicht auf eine indirekte, über CFII und CstF vermittelte Interaktion zurückzuführen. Der heterodimere Exportfaktor NXF1-NXT1 konnte nur in dem Interaktom von CFII identifiziert werden. Aufgrund der hohen Anreicherung von NXF1 kann man von einer direkten Interaktion mit CFII ausgehen. In einer *high-throughput*-mRNA-Interaktomanalyse konnte eine Interaktion zwischen NXF1 und Pcf11 identifiziert werden [232].

4.3.7.5 Cap-bindende Proteine

Die bekanntesten Cap-bindenden Proteine sind eIF4E und der nukleare Cap-bindende Komplex CBC. Der CBC-Komplex besteht aus den beiden Untereinheiten NCBP1 (Cbp80) und NCBP2 (Cpb20) [233]. Der CBC-Komplex lagert sich unmittelbar nach der Synthese an die Cap-Struktur der prä-mRNA an. Die Bindung des CBC-Komplexes stimuliert sowohl das Spleißen als auch die Spaltung der RNA in der 3'-Prozessierung [6], [234]. Keinen Einfluss hat der Komplex jedoch auf die anschließende Polyadenylierung der prä-mRNA. Der CBC-Komplex wurde vor allem in dem *Pulldown*-Experiment von CstF identifiziert. Hier liegen die beiden Untereinheiten, basierend auf der Intensität, äquimolar vor.

Es konnte kürzlich gezeigt werden, dass auch eIF4E die Spaltung und Polyadenylierung stimulieren kann [235]. Diese soll über eine direkte Interaktion mit Faktoren der Polyadenylierung vermittelt werden. eIF4E konnte jedoch nur in dem *Pulldown*-Experiment von CstF mit sehr geringer Intensität ($\log_2(\text{iBAQ}) = 16,5$) identifiziert werden, sodass davon auszugehen ist, dass die Interaktion zwischen der Cap-Struktur und der 3'-Prozessierung vor allem durch den CBC-Komplex vermittelt wird.

4.3.7.6 SR-reiche und andere Spleißfaktoren

Neben fast allen bekannten SRSF-Proteinen (*serine and arginine rich splicing factors*) wurden zwei Helikasen und zwei Proteine des U2-Hilfskomplexes (U2AF) in den Interaktomanalysen gefunden. Es ist anzunehmen, dass diese keinen direkten Einfluss auf die Polyadenylierung haben. Über den CFI-Komplex ist eine Interaktion zwischen Polyadenylierung und Spleißen beschrieben [36].

4.3.7.7 EJC/ASAP-Komplex

Der EJC-Komplex besteht aus den Proteinen eIF4A3, Magoh, Y14, MLN51 sowie dem ihm assoziiertem ASAP Komplex, bestehend aus RNPS1, SAP18 und Acinus [236]–[238]. Dieser Komplex wird während des Spleißens an die prä-mRNA angelagert und ist vor allem für den *nonsense-mediated*-Abbau von mRNA verantwortlich [239], [240]. Der EJC-Komplex bleibt bis zur Translation mit der mRNA assoziiert und stimuliert diese [195].

Eine Interaktion zwischen dem EJC/ASAP-Komplex und den Proteinen der 3'-Prozessierung ist nicht beschrieben. Die Identifikation dieser Proteine ist daher vermutlich auf die stabile Interaktion mit mRNA und somit auf eine indirekte Interaktion zurückzuführen.

4.3.8 Zusammenfassung der Interaktomanalyse

Im Rahmen der Interaktomanalysen der beiden Deletionsvarianten des CFII-Komplexes sowie des CstF- und des Endonuklease-Komplexes konnte ein umfangreiches Bild der mRNA-Prozessierung dargestellt werden. Neben der im Vordergrund stehenden 3'-Prozessierung konnten Faktoren des Spleißens, wie die SR-reichen Faktoren und der EJC-Komplex, der 5'-Prozessierung in Form des CBC-Komplexes, sowie des mRNA-Exports durch den TREX/THO-Komplex identifiziert werden. Darüber hinaus sind durch den PAF- sowie den ASAP-Komplex Faktoren der mRNA-Qualitätskontrolle identifiziert worden. Da es für die Zelle essentiell ist, dass die transkribierte mRNA korrekt prozessiert wurde, laufen diese Prozesse *in vivo* sehr stark verschränkt ab.

Eine Identifikation eines neuen Faktors der Polyadenylierung ist im Rahmen dieser Arbeit nicht möglich gewesen. Die spezifische Rolle der hier gezeigten Komplexe in der 3'-Prozessierung muss im Rahmen weiterer biochemischer Experimente im Detail geklärt werden. Auch ist es denkbar, dass aufgrund von fehlenden Informationen die Identifikation weiterer Proteine, die an der Prozessierung beteiligt sind, nicht möglich war.

5 Zusammenfassung

In dieser Arbeit wurden die bisher bekannten und an der mRNA-3'-Prozessierung beteiligten Proteinkomplexe gereinigt, biochemisch charakterisiert und mittels *cross-linking* und Massenspektrometrie strukturell untersucht. Darüber hinaus wurden diese Strukturinformationen genutzt, um durch computergestützte Proteinstrukturvorhersage Strukturmodelle der Proteinkomplexe zu erhalten.

Im ersten Teil der Arbeit wurden die bekannten Proteinkomplexe CFI, CFII, CstF sowie die beiden CPSF-Subkomplexe mPSF und der Endonuklease-Komplex auf der Basis bereits vorhandener Expressionsklone und unter Nutzung etablierter Verfahren gereinigt. Darüber hinaus wurden verschiedene Deletionsvarianten von CFII sowie die Zinkfingerdomäne von Pcf11 exprimiert und gereinigt. Bis auf den mPSF-Komplex, bei dem Fip1 im Überschuss vorlag, konnten diese Proteinkomplexe in sehr hoher Reinheit dargestellt werden.

Der CFII-Komplex war aktiv in der Spaltung der Substrat-RNA. Darüber hinaus wurde gezeigt, dass die Zinkfingerdomäne in Pcf11, welche direkt mit Clp1 interagiert, auch für die Bindung der RNA verantwortlich ist. Es konnte eine weitere Region in Pcf11, die FEGP-Wiederholungen, identifiziert werden. Diese zeichnet sich durch die 30-fache Wiederholung einer Konsensussequenz aus, die ausschließlich in Wirbeltieren existiert. Es wurde identifiziert, dass die Argininreste in dieser Region nach der Expression im Baculovirus-System dimethyliert vorliegen. Auch konnte gezeigt werden, dass die Region essentiell für die Aktivität von CFII in der RNA-Spaltung ist. Durch das chemische *cross-linking* konnte darüber hinaus eine dritte Region, Pcf11₅₀₀₋₇₇₀, identifiziert werden, die zum einen vermutlich eine klare Tertiärfaltung aufzeigt und zum anderen direkt an der Interaktion mit Clp1 beteiligt ist.

Auch der gereinigte CstF-Komplex war aktiv in der Spaltung der Substrat-RNA. Dieser Komplex wurde vor allem strukturell untersucht. Es zeigte sich, dass, entgegen der bisherigen Annahme, CstF50 und CstF64 direkt miteinander interagieren.

Der CPSF-Komplex, bestehend aus mPSF und Endonuklease-Komplex, zeigte keine stabile Assoziation. Es wurden daher beide Komplexe getrennt analysiert. Der mPSF-Komplex, aktiv in der Polyadenylierung, konnte hierbei nur partiell gereinigt werden. Es zeigten sich zum einen, dass noch sehr viele Fremdproteine die Proteinpräparation kontaminierten und zum anderen, dass die Untereinheiten des mPSF-Komplexes nicht äquimolar vorlagen. Da die Struktur des isolierten mPSF-Komplexes bereits aufgeklärt wurde, sollte der Mechanismus der Polyadenylierung analysiert werden. Hierbei konnte der Grundstein für weitere Experimente gelegt werden, der Mechanismus selbst konnte jedoch nicht geklärt werden.

Der Endonuklease-Komplex konnte in sehr hoher Reinheit dargestellt werden und bildet ein stabiles Heterotetramer. Der isolierte Komplex band die Substrat-RNA mit hoher Affinität. Es gab aber leider keine technische Möglichkeit, zu testen ob der Komplex aktiv in der Spaltung der Substrat-

RNA ist. Die Gründe hierfür wurden diskutiert. In der Strukturanalyse zeigte sich wie angenommen, dass die C-terminalen Bereiche von CPSF73 und CPSF100 miteinander interagieren. Symplekin, das größte Protein des Endonuklease-Komplexes, zeigte nur im zentralen Bereich Interaktionen mit den anderen Proteinen. Vor allem im C-terminalen Bereich konnten viele intramolekulare *cross-links* identifiziert werden.

Auch der CFI-Komplex wurde in sehr hoher Reinheit gewonnen. Wie erwartet, wurde eine RNA, welche zwei spezifische Bindestellen besaß, spezifisch gebunden. Das chemische *cross-linking* zeigte nur sehr wenige *cross-links*, so dass über die Struktur keine Aussagen getroffen werden konnten.

Im zweiten Teil der Arbeit wurden die identifizierten *cross-links* zur Vorhersage der Struktur der Proteinkomplexe verwendet. Hierzu wurden zusätzlich neue Methoden entwickelt, die dazu dienen, die *cross-link*-Daten möglichst umfangreich und präzise nutzen zu können. So wurde die Länge der reaktiven Seitenketten bestimmt, um die maximale Länge der Entfernungsbeschränkungen zu reduzieren. Darüber hinaus wurde durch die empirische Bestimmung der Entfernung zwischen zwei C α -Atomen in Peptiden eine sequenzunabhängige und allgemeingültige Methode entwickelt, um die Länge einer unbekannt Proteinsequenz zu bestimmen. Diese Methoden wurden genutzt, um die Struktur des CFII, CstF- sowie Endonuklease-Komplexes zu beschreiben.

Der CFII-Komplex, welcher sich aus Pcf11 und Clp1 zusammensetzt, war strukturell bisher nur in Form der homologen Strukturen von Clp1 aus Hefe und *C. elegans* sowie einer Zinkfingerdomäne von Pcf11 aus Hefe beschrieben. Durch die Kombination verschiedener bioinformatischer Werkzeuge konnten insgesamt 7 homologe Teilstrukturen von Pcf11 identifiziert werden und anschließend zusammen mit der Struktur von Clp1 zu einem Komplexmodell zusammengefügt werden.

Die Untereinheiten des CstF-Komplexes, CstF50, CstF64 und CstF77, wurden in verschiedenen Arbeiten bereits strukturell untersucht, so dass von allen Proteinen schon Teilstrukturen vorhanden waren. Diese wurden jedoch jeweils isoliert betrachtet, so dass die Struktur des nativen CstF-Komplexes unbekannt war. Unter Verwendung der *cross-link*-Daten konnte sequentiell die Struktur der trimeren Form des CstF-Komplexes erzeugt werden. Hervorzuheben ist hier die Interaktion zwischen CstF50 und CstF64, welche durch eine α -Helix von CstF77 stabilisiert wird. Die Analyse der nativen hexameren Struktur war nicht erfolgreich, jedoch kann das hier vorgestellte Modell genutzt werden, um zukünftige strukturelle Experimente zu unterstützen.

Wie im CstF-Komplex waren Strukturen der ungebundenen Untereinheiten bekannt, die Struktur des nativen Endonuklease-Komplexes war jedoch unbekannt. Diese z.T. homologen Strukturen

konnten in dieser Arbeit zu einem Komplex zusammengefügt werden, der im Einklang mit den experimentellen *cross-link*-Daten war. Zeitgleich mit dieser Arbeit wurde eine kryo-EM Dichte für den Endonuklease-Komplex mit einer Auflösung von 7.4 Å publiziert, die genutzt werden konnte um die C-terminale HEAT-Domäne von Symplekin, welche keine intermolekularen *cross-links* aufwies, korrekt zu platzieren. Da jedoch weiterhin große Teile von CstF64 und Symplekin sowohl keine homologen Strukturen als auch keine intermolekulare *cross-links*, die für eine Strukturbeschreibung benötigt werden, aufzeigten, ist die native Struktur des Endonuklease-Komplexes weiterhin nicht in Gänze bekannt. Auch beschreibt dieses statische Modell nicht die bekannte Konformationsänderung von CPSF73, da diese Dynamik in dieser Arbeit nicht adressiert wurde.

Der dritte Teil der Arbeit fokussierte sich auf die Identifikation weiterer Interaktionspartner der zuvor beschriebenen Proteinkomplexe. Hierzu wurden zwei Deletionsvarianten des CFII-Komplexes, sowie der CstF-, CFI- und Endonuklease-Komplex verwendet, um in *Pulldown*-Experimenten in HeLa-Kernextrakt mögliche Interaktionspartner zu identifizieren. Es zeigte sich, dass ausschließlich die Verwendung eines FLAG-tags zu auswertbaren Ergebnissen führte. Die beiden CFII-Varianten unterschieden sich in dem Vorhandensein der FEGP-Wiederholungen und der damit vermittelten Aktivität in der Spaltung der Substrat-RNA. In den Interaktionsstudien zeigte sich, dass die FEGP-Wiederholungen vermutlich eine Assoziation mit der Poly(A)-Polymerase erlauben. Diese ist für beide Teilschritte der 3'-Prozessierung notwendig. Eine biochemische Verifizierung dieser Identifikation steht noch aus. Darüber hinaus wurde der XRN2-Komplex spezifisch als Interaktionspartner für den CFII-Komplex identifiziert.

Außerdem wurden die Interaktome des CstF- sowie des Endonuklease-Komplexes analysiert. Für CstF konnte hier überraschend eine Anreicherung des Protein Nucleophosmin beobachtet werden, vom dem eine Interaktion zu CPSF postuliert wurde. Im Endonuklease-Komplex wurde vor allem die Phosphatase Ssu72 angereichert, eine direkte Interaktion mit Symplekin ist aber bereits beschrieben und strukturell aufgeklärt.

In allen drei analysierten Interaktomen wurden darüber hinaus weitere Proteinkomplexe identifiziert, die einen umfangreichen Überblick über die mRNA-Prozessierung, einschließlich des *cappings*, Speißen, 3'-Prozessierung, des Exportes sowie der Qualitätskontrolle, aufzeigen. Hervorzuheben ist hier der TREX/THO-Komplex der in allen Analysen angereichert wurde. Es ist bekannt, dass dieser Komplex mit der RNA-Polymerase II, der Cap-Struktur und mit Proteinen der 3'-Prozessierung interagieren kann und wichtig für den Export der RNA ist. Inwieweit dieser Komplex auch wichtig für die 3'-Prozessierung ist, muss noch untersucht werden.

5.1 Summary

In this thesis all known essential protein complexes involved in mRNA 3'-processing were purified, biochemically characterized and structurally analyzed by chemical cross-linking and mass spectrometry. This structural information was further used to generate model of these complexes, using various bioinformatic and computational structural biology tools.

In the first part of this thesis the known protein complexes CFI, CFII, CstF and both CPSF subcomplexes, mPSF and the endonuclease-complex, were purified using already existing expression constructs and established protocols. In addition, different deletion variants of CFII and the zinc finger domain of Pcf11 were expressed. All protein complexes could be purified to high purity with the exception of the mPSF complex, where Fip1 was in excess.

The CFII complex was active in cleavage of the substrate RNA. It could also be shown for the first time that the zinc finger domain which directly interacts with Clp1 is responsible for RNA binding. In addition, another region in Pcf11, the so called FEGP repeats, could be characterized in this context: This region consists of 30 repeats of a consensus sequence and only exists in Pcf11 from vertebrae. After expression in the baculoviral expression system it could be shown that the terminal arginine residue in the repeats is frequently methylated. This region is also essential for the cleavage activity of CFII. A third region (Pcf11₅₀₀₋₇₇₀) was identified by cross-linking of the entire complex, being proximal to the repeat region. This additional domain most likely has a well-defined tertiary structure and is also involved in the interaction with Clp1.

The purified CstF complex, essential for the cleavage reaction, was active in cleavage of the substrate RNA after purification. This complex was structurally analyzed by cross-linking and results show that contrary to the current hypothesized structural model of CstF50 and CstF64, both proteins have been identified to directly interact.

The CPSF complex, consisting of mPSF and the endonuclease complex, doesn't show a stable association. Therefore, both complexes were analyzed separately. The mPSF complex, active in polyadenylation, could only partially purified. It was observed, that a lot of contaminating proteins were in the purification and that the subunits were not equimolar. Although the structure of isolated mPSF is known, it cannot provide insights into the mechanism of polyadenylation (initiation, elongation and termination). Therefore, preliminary experiments were performed to define optimal conditions for the cross-linking of the purified complexes and results provide a framework for further optimization of the biochemistry to polyadenylation.

The endonuclease complex could be purified to very high homogeneity and forms a stable heterotetramer. The isolated complex is binding the substrate RNA with high affinity. Due to technical limitations, it was not possible to test the cleavage activity of this complex. The reasons were discussed. The structural analysis showed that the C-terminal regions of CPSF73 and CPSF100 are interacting. Symplekin, the largest protein of the endonuclease complex only

showed interaction in the central region with the other proteins. Also, the CFI complex could be purified to high purity. As expected, an RNA with two binding motifs was bound, therefore indicating an active state. Chemical cross-linking showed few cross-links, so no final model about the structure of CFI could be derived.

In the second part of this thesis, the identified cross-links were used as experimental input for optimizing protein structure prediction. Additional methods to use this cross-link information most extensive and precisely as possible were developed: (1) Residue-specific distances for the cross-linking reaction were applied to reduce the maximum length of distance restraints; (2) by measuring 70,000,000 C α distances in peptides, a sequence independent and universal method was developed to describe the average length of an unknown protein sequence – a restraint that can be used in structure prediction. These methods were used to predict the molecular architecture of the CFII, CstF and endonuclease complexes.

Of the CFII complex, consisting of Pcf11 and Clp1, only homologous structure of Clp1 from baker's yeast and *C. elegans* and one out of two zinc finger domains from Pcf11 from baker's yeast were known. By combining different bioinformatic tools, 7 homologous substructures of Pcf11 were identified and docked with Clp1, to form a novel complex structure which matches the experimental data.

The subunits of CstF, CstF50, CstF64 and CstF77, were already known, so that from all proteins monomeric structures were available. However, they as they were known in their unbound states, therefore, the native structure of the CstF complex was unknown. By using cross-linking data as distance restraints, although formation of the native hexameric complex was not possible, the trimeric form of CstF could be build. Particular highlight is the interaction between CstF50 and CstF64, which is stabilized by a α -helix of CstF77; This model can be used to assist future experiments and possibly as a starting structure for understanding the full architecture of the native CstF complex.

The structure of the endonuclease complex, comparable to the CstF complex, was unknown, but structures of the monomeric subunits were available. In this thesis, these, partly homologous models could be docked to a final complex, which is in good agreement with the derived experimental data. During the writing of the thesis, a cryo-EM density map with a resolution of 7,4 Å was published by another group, and was used to validate and further extend the endonuclease complex. In particular, a large domain of Symplekin, which lacks intermolecular cross-links could be assigned in the formed complex. But as large parts of CstF64 and Symplekin neither have homologous structures nor intermolecular cross-links, which are essential for optimal docking results, the native structure of the endonuclease complex remains unknown. In addition, extensive conformational changes of the endonuclease complex, shown in yeast, have not been addressed yet, and the structural elucidation of those states remains a goal for future studies.

The third part of this thesis focuses on the identification of interaction partners of the above described protein complexes. Therefore, two deletion variants of CFII, the CstF, CFI and endonuclease complex were used to identify interaction partners in HeLa nuclear extracts by pulldown experiments. By benchmarking different tags, it was evident that only the usage of a FLAG-tag may deliver usable data. The CFII deletion variants differ in the existence of the FEGP repeats and, thereby, in the activity in the cleavage reaction. In the interactome studies of these variants performed in this thesis, it was shown that the FEGP repeats presumably interact with the poly(A) polymerase, which is essential for both the cleavage and polyadenylation steps for mRNA maturation. Additionally, the XRN2 complex was identified as a specific interactor of the CFII complex. The interactome of CstF and the endonuclease complex were also analyzed. Surprisingly, in the CstF pulldown an enrichment for Nucleophosmin, previously described as interaction partner of CPSF not CstF, was identified. In the endonuclease complex pulldown, the phosphatase Ssu72 was enriched, but a direct interaction with Symplekin is already known and structurally characterized.

In all three analyzed interactome studies additional protein complexes were identified, which give a broader view about mRNA processing, including capping, splicing, 3'-processing, export and quality control. In particular, the TREX/THO complex was enriched in all studies. It is known that this complex interacts with the RNA polymerase II, with the cap structure and also proteins of the 3'-processing machinery and that it is important for the export of mRNA. In what extend this complex is involved in the actual 3'-processing has to be validated.

Overall, in this thesis a combination of molecular biology, biochemistry, cross-linking mass spectrometry and computational structural biology methods was applied to provide structural insights into the mRNA 3'-processing machinery. In particular, the abovementioned protocols and derived structural models can provide the basis for subsequent experiments and structure-based modelling in combination with novel techniques, e.g. cryo-EM, to provide further insights into the atomic structures involved in this essential cellular mechanism.

6 Literaturverzeichnis

- [1] N. J. Proudfoot, A. Furger, und M. J. Dye, „Integrating mRNA Processing with Transcription“, *Cell*, Bd. 108, Nr. 4, S. 501–512, Feb. 2002.
- [2] S. R. Carmody und S. R. Wentz, „mRNA nuclear export at a glance“, *J Cell Sci*, Bd. 122, Nr. 12, S. 1933–1937, Juni 2009.
- [3] T. Saldi, M. A. Cortazar, R. M. Sheridan, und D. L. Bentley, „Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing“, *J Mol Biol*, Bd. 428, Nr. 12, S. 2623–2635, Juni 2016.
- [4] M. Lidschreiber, K. Leike, und P. Cramer, „Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II“, *Mol. Cell. Biol.*, Bd. 33, Nr. 19, S. 3805–3816, Okt. 2013.
- [5] N. E. Shirokikh und T. Preiss, „Translation initiation by cap-dependent ribosome recruitment: Recent insights and open questions“, *Wiley Interdiscip Rev RNA*, Bd. 9, Nr. 4, S. e1473, 2018.
- [6] E. Izaurralde, J. Stepinski, E. Darzynkiewicz, und I. W. Mattaj, „A cap binding protein that may mediate nuclear export of RNA polymerase II-transcribed RNAs.“, *The Journal of Cell Biology*, Bd. 118, Nr. 6, S. 1287–1295, Sep. 1992.
- [7] N. Sonenberg und A. C. Gingras, „The mRNA 5' cap-binding protein eIF4E and control of cell growth“, *Curr. Opin. Cell Biol.*, Bd. 10, Nr. 2, S. 268–275, Apr. 1998.
- [8] M. R. Lerner und J. A. Steitz, „Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus“, *Proc. Natl. Acad. Sci. U.S.A.*, Bd. 76, Nr. 11, S. 5495–5499, Nov. 1979.
- [9] Y. Shi, „Mechanistic insights into precursor messenger RNA splicing by the spliceosome“, *Nat. Rev. Mol. Cell Biol.*, Bd. 18, Nr. 11, S. 655–670, Nov. 2017.
- [10] K. Bertram, D. E. Agafonov, O. Dybkov, D. Haselbach, M. N. Leelaram, C. L. Will, H. Urlaub, B. Kastner, R. Lührmann, und H. Stark, „Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation“, *Cell*, Bd. 170, Nr. 4, S. 701-713.e11, Aug. 2017.
- [11] D. Haselbach, I. Komarov, D. E. Agafonov, K. Hartmuth, B. Graf, O. Dybkov, H. Urlaub, B. Kastner, R. Lührmann, und H. Stark, „Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex“, *Cell*, Bd. 172, Nr. 3, S. 454-464.e11, 25 2018.
- [12] K. Bertram, D. E. Agafonov, W.-T. Liu, O. Dybkov, C. L. Will, K. Hartmuth, H. Urlaub, B. Kastner, H. Stark, und R. Lührmann, „Cryo-EM structure of a human spliceosome activated for step 2 of splicing“, *Nature*, Bd. 542, Nr. 7641, S. 318–323, 16 2017.
- [13] X. Zhang, C. Yan, J. Hang, L. I. Finci, J. Lei, und Y. Shi, „An Atomic Structure of the Human Spliceosome“, *Cell*, Bd. 169, Nr. 5, S. 918-929.e14, Mai 2017.
- [14] X. Zhan, C. Yan, X. Zhang, J. Lei, und Y. Shi, „Structure of a human catalytic step I spliceosome“, *Science*, Bd. 359, Nr. 6375, S. 537–545, 02 2018.

- [15] X. Zhan, C. Yan, X. Zhang, J. Lei, und Y. Shi, „Structures of the human pre-catalytic spliceosome and its precursor spliceosome“, *Cell Res.*, Bd. 28, Nr. 12, S. 1129–1140, Dez. 2018.
- [16] X. Zhang, C. Yan, X. Zhan, L. Li, J. Lei, und Y. Shi, „Structure of the human activated spliceosome in three conformational states“, *Cell Res.*, Bd. 28, Nr. 3, S. 307–322, März 2018.
- [17] X. Zhang, X. Zhan, C. Yan, W. Zhang, D. Liu, J. Lei, und Y. Shi, „Structures of the human spliceosomes before and after release of the ligated exon“, *Cell Res.*, Bd. 29, Nr. 4, S. 274–285, Apr. 2019.
- [18] E. Wahle, „A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation“, *Cell*, Bd. 66, Nr. 4, S. 759–768, Aug. 1991.
- [19] M. W. Webster, Y.-H. Chen, J. A. W. Stowell, N. Alhusaini, T. Sweet, B. R. Graveley, J. Collier, und L. A. Passmore, „mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases“, *Mol Cell*, Bd. 70, Nr. 6, S. 1089-1100.e8, Juni 2018.
- [20] C.-Y. A. Chen und A.-B. Shyu, „Mechanisms of deadenylation-dependent decay“, *Wiley Interdiscip Rev RNA*, Bd. 2, Nr. 2, S. 167–183, Apr. 2011.
- [21] A. Yamashita, T.-C. Chang, Y. Yamashita, W. Zhu, Z. Zhong, C.-Y. A. Chen, und A.-B. Shyu, „Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover“, *Nat. Struct. Mol. Biol.*, Bd. 12, Nr. 12, S. 1054–1063, Dez. 2005.
- [22] D. Munroe und A. Jacobson, „mRNA poly(A) tail, a 3' enhancer of translational initiation.“, *Mol Cell Biol*, Bd. 10, Nr. 7, S. 3441–3455, Juli 1990.
- [23] Q. Vicens, J. S. Kieft, und O. S. Rissland, „Revisiting the Closed-Loop Model and the Nature of mRNA 5'–3' Communication“, *Molecular Cell*, Bd. 72, Nr. 5, S. 805–812, Dez. 2018.
- [24] Y. Huang und G. G. Carmichael, „Role of polyadenylation in nucleocytoplasmic transport of mRNA.“, *Molecular and Cellular Biology*, Bd. 16, Nr. 4, S. 1534–1542, Apr. 1996.
- [25] Y. Shi, D. C. Di Giammartino, D. Taylor, A. Sarkeshik, W. J. Rice, J. R. Yates, J. Frank, und J. L. Manley, „Molecular architecture of the human pre-mRNA 3' processing complex“, *Mol Cell*, Bd. 33, Nr. 3, S. 365–376, Feb. 2009.
- [26] L. Schönemann, *In-Vitro-Rekonstitution des 3'-Prozessierungsfaktors CPSF (cleavage and polyadenylation specificity factor)*. Dissertation an der Martin-Luther Universität Halle-Wittenberg. Halle (Saale), 2017.
- [27] P. Schäfer, C. Tüting, L. Schönemann, U. Kühn, T. Treiber, N. Treiber, C. Ihling, A. Graber, W. Keller, G. Meister, A. Sinz, und E. Wahle, „Reconstitution of mammalian cleavage factor II involved in 3' processing of mRNA precursors“, *RNA*, Bd. 24, Nr. 12, S. 1721–1737, 2018.

- [28] L. Schönemann, U. Kühn, G. Martin, P. Schäfer, A. R. Gruber, W. Keller, M. Zavolan, und E. Wahle, „Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33“, *Genes Dev.*, Bd. 28, Nr. 21, S. 2381–2393, Nov. 2014.
- [29] U. Kühn, M. Gündel, A. Knoth, Y. Kerwitz, S. Rüdell, und E. Wahle, „Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor“, *J. Biol. Chem.*, Bd. 284, Nr. 34, S. 22803–22814, Aug. 2009.
- [30] Y. Kerwitz, U. Kühn, H. Lilie, A. Knoth, T. Scheuermann, H. Friedrich, E. Schwarz, und E. Wahle, „Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA“, *EMBO J.*, Bd. 22, Nr. 14, S. 3705–3714, Juli 2003.
- [31] K. G. Murthy und J. L. Manley, „The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation“, *Genes Dev.*, Bd. 9, Nr. 21, S. 2672–2683, Nov. 1995.
- [32] A. Misra und M. R. Green, „From polyadenylation to splicing: Dual role for mRNA 3' end formation factors“, *RNA Biol.*, Bd. 13, Nr. 3, S. 259–264, 2016.
- [33] W. Keller, S. Bienroth, K. M. Lang, und G. Christofori, „Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA.“, *EMBO J.*, Bd. 10, Nr. 13, S. 4241–4249, Dez. 1991.
- [34] Q. Yang, G. M. Gilmartin, und S. Doublé, „Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing“, *Proc. Natl. Acad. Sci. U.S.A.*, Bd. 107, Nr. 22, S. 10062–10067, Juni 2010.
- [35] K. Venkataraman, K. M. Brown, und G. M. Gilmartin, „Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition“, *Genes Dev.*, Bd. 19, Nr. 11, S. 1315–1327, Juni 2005.
- [36] S. Millevoi, C. Loulergue, S. Dettwiler, S. Z. Karaa, W. Keller, M. Antoniou, und S. Vagner, „An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries“, *EMBO J.*, Bd. 25, Nr. 20, S. 4854–4864, Okt. 2006.
- [37] I. Kaufmann, G. Martin, A. Friedlein, H. Langen, und W. Keller, „Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase“, *EMBO J.*, Bd. 23, Nr. 3, S. 616–626, Feb. 2004.
- [38] J. M. Pérez Cañadillas und G. Varani, „Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein“, *EMBO J.*, Bd. 22, Nr. 11, S. 2821–2830, Juni 2003.
- [39] K. G. Murthy und J. L. Manley, „Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus“, *J. Biol. Chem.*, Bd. 267, Nr. 21, S. 14804–14811, Juli 1992.

- [40] J. Hu, C. S. Lutz, J. Wilusz, und B. Tian, „Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation“, *RNA*, Bd. 11, Nr. 10, S. 1485–1493, Okt. 2005.
- [41] A. R. Gruber, G. Martin, W. Keller, und M. Zavolan, „Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors“, *Wiley Interdisciplinary Reviews: RNA*, Bd. 5, Nr. 2, S. 183–196, 2014.
- [42] U. Rügsegger, K. Beyer, und W. Keller, „Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors“, *J. Biol. Chem.*, Bd. 271, Nr. 11, S. 6107–6113, März 1996.
- [43] U. Rügsegger, D. Blank, und W. Keller, „Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits“, *Mol. Cell*, Bd. 1, Nr. 2, S. 243–253, Jan. 1998.
- [44] S. Kim, J. Yamamoto, Y. Chen, M. Aida, T. Wada, H. Handa, und Y. Yamaguchi, „Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation“, *Genes Cells*, Bd. 15, Nr. 9, S. 1003–1013, Sep. 2010.
- [45] K. M. Brown und G. M. Gilmartin, „A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im“, *Mol. Cell*, Bd. 12, Nr. 6, S. 1467–1476, Dez. 2003.
- [46] A. G. McLennan, „The Nudix hydrolase superfamily“, *Cell. Mol. Life Sci.*, Bd. 63, Nr. 2, S. 123–143, Jan. 2006.
- [47] S. Dettwiler, C. Aringhieri, S. Cardinale, W. Keller, und S. M. L. Barabino, „Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization“, *J. Biol. Chem.*, Bd. 279, Nr. 34, S. 35788–35797, Aug. 2004.
- [48] M. Coseno, G. Martin, C. Berger, G. Gilmartin, W. Keller, und S. Doublé, „Crystal structure of the 25 kDa subunit of human cleavage factor Im“, *Nucleic Acids Res.*, Bd. 36, Nr. 10, S. 3474–3483, Juni 2008.
- [49] K. Xiang, L. Tong, und J. L. Manley, „Delineating the Structural Blueprint of the Pre-mRNA 3'-End Processing Machinery“, *Mol Cell Biol*, Bd. 34, Nr. 11, S. 1894–1910, Juni 2014.
- [50] H. Li, S. Tong, X. Li, H. Shi, Z. Ying, Y. Gao, H. Ge, L. Niu, und M. Teng, „Structural basis of pre-mRNA recognition by the human cleavage factor Im complex“, *Cell Res.*, Bd. 21, Nr. 7, S. 1039–1051, Juli 2011.
- [51] H. Shen und M. R. Green, „RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans“, *Genes Dev*, Bd. 20, Nr. 13, S. 1755–1765, Juli 2006.
- [52] H. de Vries, U. Rügsegger, W. Hübner, A. Friedlein, H. Langen, und W. Keller, „Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors“, *EMBO J.*, Bd. 19, Nr. 21, S. 5895–5904, Nov. 2000.

- [53] N. Amrani, M. Minet, M. Le Gouar, F. Lacroute, und F. Wyers, „Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro“, *Mol. Cell. Biol.*, Bd. 17, Nr. 7, S. 3694–3701, Juli 1997.
- [54] L. Minvielle-Sebastia, P. J. Preker, T. Wiederkehr, Y. Strahm, und W. Keller, „The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation“, *Proc. Natl. Acad. Sci. U.S.A.*, Bd. 94, Nr. 15, S. 7897–7902, Juli 1997.
- [55] S. West und N. J. Proudfoot, „Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination“, *Nucleic Acids Res.*, Bd. 36, Nr. 3, S. 905–914, Feb. 2008.
- [56] W. Luo, A. W. Johnson, und D. L. Bentley, „The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model“, *Genes Dev.*, Bd. 20, Nr. 8, S. 954–965, Apr. 2006.
- [57] A. Volanakis, K. Kamieniarz-Gdula, M. Schlackow, und N. J. Proudfoot, „WNK1 kinase and the termination factor PCF11 connect nuclear mRNA export with transcription“, *Genes Dev.*, Bd. 31, Nr. 21, S. 2175–2185, 01 2017.
- [58] J. D. Eaton, L. Davidson, D. L. V. Bauer, T. Natsume, M. T. Kanemaki, und S. West, „Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity“, *Genes Dev.*, Bd. 32, Nr. 2, S. 127–139, 15 2018.
- [59] S. V. Paushkin, M. Patel, B. S. Furia, S. W. Peltz, und C. R. Trotta, „Identification of a Human Endonuclease Complex Reveals a Link between tRNA Splicing and Pre-mRNA 3' End Formation“, *Cell*, Bd. 117, Nr. 3, S. 311–321, Apr. 2004.
- [60] S. Weitzer und J. Martinez, „The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs“, *Nature*, Bd. 447, Nr. 7141, S. 222–226, Mai 2007.
- [61] C. G. Noble, B. Beuth, und I. A. Taylor, „Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor“, *Nucleic Acids Res.*, Bd. 35, Nr. 1, S. 87–99, 2007.
- [62] A. Dikfidan, B. Loll, C. Zeymer, I. Magler, T. Clausen, und A. Meinhart, „RNA specificity and regulation of catalysis in the eukaryotic polynucleotide kinase Clp1“, *Mol. Cell*, Bd. 54, Nr. 6, S. 975–986, Juni 2014.
- [63] A. F. Dupin und S. Fribourg, „Structural basis for ATP loss by Clp1p in a G135R mutant protein“, *Biochimie*, Bd. 101, S. 203–207, Juni 2014.
- [64] J. E. Walker, M. Saraste, M. J. Runswick, und N. J. Gay, „Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold“, *EMBO J.*, Bd. 1, Nr. 8, S. 945–951, 1982.
- [65] A. Meinhart und P. Cramer, „Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors“, *Nature*, Bd. 430, Nr. 6996, S. 223–226, Juli 2004.

- [66] X. Xu, N. Pérébasquine, L. Minvielle-Sébastien, S. Fribourg, und C. D. Mackereth, „Chemical shift assignments of a new folded domain from yeast Pcf11“, *Biomol NMR Assign*, Bd. 9, Nr. 2, S. 421–425, Okt. 2015.
- [67] F. Yang, P. Hsu, S. D. Lee, W. Yang, D. Hoskinson, W. Xu, C. Moore, und G. Varani, „The C terminus of Pcf11 forms a novel zinc-finger structure that plays an essential role in mRNA 3'-end processing“, *RNA*, Bd. 23, Nr. 1, S. 98–107, 2017.
- [68] J. Guéguéniat, A. F. Dupin, J. Stojko, L. Beaurepaire, S. Cianférani, C. D. Mackereth, L. Minvielle-Sébastien, und S. Fribourg, „Distinct roles of Pcf11 zinc-binding domains in pre-mRNA 3'-end processing“, *Nucleic Acids Res.*, Bd. 45, Nr. 17, S. 10115–10131, Sep. 2017.
- [69] Y. Takagaki, J. L. Manley, C. C. MacDonald, J. Wilusz, und T. Shenk, „A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs“, *Genes Dev.*, Bd. 4, Nr. 12A, S. 2112–2120, Dez. 1990.
- [70] B. Dass, K. W. McMahon, N. A. Jenkins, D. J. Gilbert, N. G. Copeland, und C. C. MacDonald, „The Gene for a Variant Form of the Polyadenylation Protein CstF-64 Is on Chromosome 19 and Is Expressed in Pachytene Spermatocytes in Mice“, *J. Biol. Chem.*, Bd. 276, Nr. 11, S. 8044–8050, März 2001.
- [71] C. Yao, E.-A. Choi, L. Weng, X. Xie, J. Wan, Y. Xing, J. J. Moresco, P. G. Tu, J. R. Yates, und Y. Shi, „Overlapping and distinct functions of CstF64 and CstF64 τ in mammalian mRNA 3' processing“, *RNA*, Bd. 19, Nr. 12, S. 1781–1790, Dez. 2013.
- [72] Y. Bai, T. C. Auperin, C.-Y. Chou, G.-G. Chang, J. L. Manley, und L. Tong, „Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors“, *Mol. Cell*, Bd. 25, Nr. 6, S. 863–875, März 2007.
- [73] W. Yang, P. L. Hsu, F. Yang, J.-E. Song, und G. Varani, „Reconstitution of the CstF complex unveils a regulatory role for CstF-50 in recognition of 3'-end processing signals“, *Nucleic Acids Res*, Bd. 46, Nr. 2, S. 493–503, Jan. 2018.
- [74] Y. Takagaki und J. L. Manley, „Complex protein interactions within the human polyadenylation machinery identify a novel component“, *Mol. Cell. Biol.*, Bd. 20, Nr. 5, S. 1515–1525, März 2000.
- [75] Y. Takagaki und J. L. Manley, „A polyadenylation factor subunit is the human homologue of the *Drosophila* suppressor of forked protein“, *Nature*, Bd. 372, Nr. 6505, S. 471–474, Dez. 1994.
- [76] Y. Takagaki und J. L. Manley, „RNA recognition by the human polyadenylation factor CstF“, *Mol. Cell. Biol.*, Bd. 17, Nr. 7, S. 3907–3914, Juli 1997.
- [77] C. C. MacDonald, J. Wilusz, und T. Shenk, „The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location.“, *Molecular and Cellular Biology*, Bd. 14, Nr. 10, S. 6647–6654, Okt. 1994.

- [78] G. M. Gilmartin und J. R. Nevins, „Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA“, *Mol. Cell. Biol.*, Bd. 11, Nr. 5, S. 2432–2438, Mai 1991.
- [79] Y. Takagaki, L. C. Ryner, und J. L. Manley, „Four factors are required for 3'-end cleavage of pre-mRNAs.“, *Genes Dev.*, Bd. 3, Nr. 11, S. 1711–1724, Jan. 1989.
- [80] P. J. Preker und W. Keller, „The HAT helix, a repetitive motif implicated in RNA processing“, *Trends Biochem. Sci.*, Bd. 23, Nr. 1, S. 15–16, Jan. 1998.
- [81] P. Legrand, N. Pinaud, L. Minvielle-Sébastien, und S. Fribourg, „The structure of the CstF-77 homodimer provides insights into CstF assembly“, *Nucleic Acids Res.*, Bd. 35, Nr. 13, S. 4515–4522, 2007.
- [82] A. R. Paulson und L. Tong, „Crystal structure of the Rna14-Rna15 complex“, *RNA*, Bd. 18, Nr. 6, S. 1154–1162, Juni 2012.
- [83] J. A. Hockert, H.-J. Yeh, und C. C. MacDonald, „The hinge domain of the cleavage stimulation factor protein CstF-64 is essential for CstF-77 interaction, nuclear localization, and polyadenylation“, *J. Biol. Chem.*, Bd. 285, Nr. 1, S. 695–704, Jan. 2010.
- [84] M. Moreno-Morcillo, L. Minvielle-Sébastien, S. Fribourg, und C. D. Mackereth, „Locked tether formation by cooperative folding of Rna14p monkeytail and Rna15p hinge domains in the yeast CF IA complex“, *Structure*, Bd. 19, Nr. 4, S. 534–545, Apr. 2011.
- [85] Y. Takagaki, C. C. MacDonald, T. Shenk, und J. L. Manley, „The human 64-kDa polyadenylation factor contains a ribonucleoprotein-type RNA binding domain and unusual auxiliary motifs“, *Proc. Natl. Acad. Sci. U.S.A.*, Bd. 89, Nr. 4, S. 1403–1407, Feb. 1992.
- [86] C. Pancevac, D. C. Goldstone, A. Ramos, und I. A. Taylor, „Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors“, *Nucleic Acids Res.*, Bd. 38, Nr. 9, S. 3119–3132, Mai 2010.
- [87] M.-D. Ruepp, C. Schweingruber, N. Kleinschmidt, und D. Schümperli, „Interactions of CstF-64, CstF-77, and symplekin: Implications on localisation and function“, *MBoC*, Bd. 22, Nr. 1, S. 91–104, Nov. 2010.
- [88] X. Qu, J.-M. Perez-Canadillas, S. Agrawal, J. De Baecke, H. Cheng, G. Varani, und C. Moore, „The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing“, *J. Biol. Chem.*, Bd. 282, Nr. 3, S. 2101–2115, Jan. 2007.
- [89] S. Gross und C. Moore, „Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I“, *Proc. Natl. Acad. Sci. U.S.A.*, Bd. 98, Nr. 11, S. 6080–6085, Mai 2001.
- [90] J. M. Richardson, K. W. McMahon, C. C. MacDonald, und G. I. Makhatadze, „MEARA sequence repeat of human CstF-64 polyadenylation factor is helical in solution. A

- spectroscopic and calorimetric study“, *Biochemistry*, Bd. 38, Nr. 39, S. 12869–12875, Sep. 1999.
- [91] M. Moreno-Morcillo, L. Minvielle-Sébastien, C. Mackereth, und S. Fribourg, „Hexameric architecture of CstF supported by CstF-50 homodimerization domain structure“, *RNA*, Bd. 17, Nr. 3, S. 412–418, März 2011.
- [92] S. Bienroth, E. Wahle, C. Suter-Crazzolara, und W. Keller, „Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors“, *J. Biol. Chem.*, Bd. 266, Nr. 29, S. 19768–19776, Okt. 1991.
- [93] I. Hofmann, M. Schnölzer, I. Kaufmann, und W. W. Franke, „Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in *Xenopus laevis* oocytes“, *Mol. Biol. Cell*, Bd. 13, Nr. 5, S. 1665–1676, Mai 2002.
- [94] S. L. Chan, I. Huppertz, C. Yao, L. Weng, J. J. Moresco, J. R. Yates, J. Ule, J. L. Manley, und Y. Shi, „CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing“, *Genes Dev.*, Bd. 28, Nr. 21, S. 2370–2380, Jan. 2014.
- [95] Y. Shi und J. L. Manley, „The end of the message: multiple protein–RNA interactions define the mRNA polyadenylation site“, *Genes Dev.*, Bd. 29, Nr. 9, S. 889–897, Jan. 2015.
- [96] C. H. Hill, V. Boreikaitė, A. Kumar, A. Casañal, P. Kubík, G. Degliesposti, S. Maslen, A. Mariani, O. von Loeffelholz, M. Girbig, M. Skehel, und L. A. Passmore, „Activation of the Endonuclease that Defines mRNA 3' Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex“, *Mol Cell*, Bd. 73, Nr. 6, S. 1217–1231.e11, März 2019.
- [97] K. D. Sullivan, M. Steiniger, und W. F. Marzluff, „A Core Complex of CPSF73, CPSF100, and Symplekin May Form Two Different Cleavage Factors for Processing of Poly(A) and Histone mRNAs“, *Molecular Cell*, Bd. 34, Nr. 3, S. 322–332, Mai 2009.
- [98] Y. Zhang, Y. Sun, Y. Shi, T. Walz, und L. Tong, „Structural Insights into the Human Pre-mRNA 3'-End Processing Machinery“, *Mol. Cell*, Nov. 2019.
- [99] M. Clerici, M. Faini, R. Aebersold, und M. Jinek, „Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex“, *Elife*, Bd. 6, 23 2017.
- [100] M. Clerici, M. Faini, L. M. Muckenfuss, R. Aebersold, und M. Jinek, „Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex“, *Nat. Struct. Mol. Biol.*, Bd. 25, Nr. 2, S. 135–138, Feb. 2018.
- [101] Y. Sun, Y. Zhang, K. Hamilton, J. L. Manley, Y. Shi, T. Walz, und L. Tong, „Molecular basis for the recognition of the human AAUAAA polyadenylation signal“, *PNAS*, Bd. 115, Nr. 7, S. E1419–E1428, Feb. 2018.
- [102] S. M. Barabino, W. Hübner, A. Jenny, L. Minvielle-Sebastia, und W. Keller, „The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins“, *Genes Dev.*, Bd. 11, Nr. 13, S. 1703–1716, Juli 1997.

- [103] W. F. Marzluff und K. P. Koreski, „Birth and Death of Histone mRNAs“, *Trends Genet.*, Bd. 33, Nr. 10, S. 745–759, 2017.
- [104] A. Casañal, A. Kumar, C. H. Hill, A. D. Easter, P. Emsley, G. Degliesposti, Y. Gordiyenko, B. Santhanam, J. Wolf, K. Wiederhold, G. L. Dornan, M. Skehel, C. V. Robinson, und L. A. Passmore, „Architecture of eukaryotic mRNA 3'-end processing machinery“, *Science*, Bd. 358, Nr. 6366, S. 1056–1059, 24 2017.
- [105] D. C. Di Giammartino, W. Li, K. Ogami, J. J. Yashinski, M. Hoque, B. Tian, und J. L. Manley, „RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs“, *Genes Dev*, Bd. 28, Nr. 20, S. 2248–2260, Okt. 2014.
- [106] J. Zhao, M. Kessler, S. Helmling, J. P. O'Connor, und C. Moore, „Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor“, *Mol. Cell. Biol.*, Bd. 19, Nr. 11, S. 7733–7740, Nov. 1999.
- [107] Z. Dominski, „The hunt for the 3' endonuclease“, *Wiley Interdiscip Rev RNA*, Bd. 1, Nr. 2, S. 325–340, Okt. 2010.
- [108] I. Callebaut, D. Moshous, J.-P. Mornon, und J.-P. de Villartay, „Metallo- β -lactamase fold within nucleic acids processing enzymes: the β -CASP family“, *Nucleic Acids Res*, Bd. 30, Nr. 16, S. 3592–3601, Aug. 2002.
- [109] Z. Dominski, X. Yang, M. Purdy, E. J. Wagner, und W. F. Marzluff, „A CPSF-73 Homologue Is Required for Cell Cycle Progression but Not Cell Growth and Interacts with a Protein Having Features of CPSF-100“, *Molecular and Cellular Biology*, Bd. 25, Nr. 4, S. 1489–1500, Feb. 2005.
- [110] A. Zhelkovsky, Y. Tacahashi, T. Nasser, X. He, U. Sterzer, T. H. Jensen, H. Domdey, und C. Moore, „The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in *Saccharomyces cerevisiae*“, *RNA*, Bd. 12, Nr. 3, S. 435–445, Jan. 2006.
- [111] C. R. Mandel, S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J. L. Manley, und L. Tong, „Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease“, *Nature*, Bd. 444, Nr. 7121, S. 953–956, Dez. 2006.
- [112] A. Jenny, L. Minvielle-Sebastia, P. J. Preker, und W. Keller, „Sequence Similarity Between the 73-Kilodalton Protein of Mammalian CPSF and a Subunit of Yeast Polyadenylation Factor I“, *Science*, Bd. 274, Nr. 5292, S. 1514–1517, Nov. 1996.
- [113] Z. Dominski, „Nucleases of the Metallo- β -lactamase Family and Their Role in DNA and RNA Metabolism“, *Critical Reviews in Biochemistry and Molecular Biology*, Bd. 42, Nr. 2, S. 67–93, Jan. 2007.
- [114] B. H. Keon, S. Schäfer, C. Kuhn, C. Grund, und W. W. Franke, „Symplekin, a novel type of tight junction plaque protein“, *J. Cell Biol.*, Bd. 134, Nr. 4, S. 1003–1018, Aug. 1996.
- [115] M. A. Ghazy, X. He, B. N. Singh, M. Hampsey, und C. Moore, „The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72

- function but is dispensable for pre-mRNA 3'-end processing", *Mol. Cell. Biol.*, Bd. 29, Nr. 8, S. 2296–2307, Apr. 2009.
- [116] M.-D. Ruepp, C. Schweingruber, N. Kleinschmidt, und D. Schümperli, „Interactions of CstF-64, CstF-77, and symplekin: implications on localisation and function“, *Mol. Biol. Cell*, Bd. 22, Nr. 1, S. 91–104, Jan. 2011.
- [117] J. E. Mayfield, S. Fan, S. Wei, M. Zhang, B. Li, A. D. Ellington, F. A. Etzkorn, und Y. J. Zhang, „Chemical Tools To Decipher Regulation of Phosphatases by Proline Isomerization on Eukaryotic RNA Polymerase II“, *ACS Chem. Biol.*, Bd. 10, Nr. 10, S. 2405–2414, Okt. 2015.
- [118] S. A. Kennedy, M. L. Frazier, M. Steiniger, A. M. Mast, W. F. Marzluff, und M. R. Redinbo, „Crystal structure of the HEAT domain from the Pre-mRNA processing factor Symplekin“, *J. Mol. Biol.*, Bd. 392, Nr. 1, S. 115–128, Sep. 2009.
- [119] K. Xiang, T. Nagaike, S. Xiang, T. Kilic, M. M. Beh, J. L. Manley, und L. Tong, „Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex“, *Nature*, Bd. 467, Nr. 7316, S. 729–733, Okt. 2010.
- [120] M. Edmonds, „Polyadenylate polymerases“, *Meth. Enzymol.*, Bd. 181, S. 161–170, 1990.
- [121] G. Martin und W. Keller, „Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases“, *EMBO J.*, Bd. 15, Nr. 10, S. 2593–2603, Mai 1996.
- [122] G. Martin, W. Keller, und S. Doublé, „Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP“, *EMBO J.*, Bd. 19, Nr. 16, S. 4193–4203, Aug. 2000.
- [123] W. Zhao und J. L. Manley, „Complex alternative RNA processing generates an unexpected diversity of poly(A) polymerase isoforms“, *Mol. Cell. Biol.*, Bd. 16, Nr. 5, S. 2378–2386, Mai 1996.
- [124] J. Bard, A. M. Zhelkovsky, S. Helmling, T. N. Earnest, C. L. Moore, und A. Böhm, „Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP“, *Science*, Bd. 289, Nr. 5483, S. 1346–1349, Aug. 2000.
- [125] H. Ge, D. Zhou, S. Tong, Y. Gao, M. Teng, und L. Niu, „Crystal structure and possible dimerization of the single RRM of human PABPN1“, *Proteins*, Bd. 71, Nr. 3, S. 1539–1545, Mai 2008.
- [126] A. Nemeth, S. Krause, D. Blank, A. Jenny, P. Jenö, A. Lustig, und E. Wahle, „Isolation of genomic and cDNA clones encoding bovine poly(A) binding protein II“, *Nucleic Acids Res.*, Bd. 23, Nr. 20, S. 4034–4041, Okt. 1995.
- [127] U. Kühn, A. Nemeth, S. Meyer, und E. Wahle, „The RNA binding domains of the nuclear poly(A)-binding protein“, *J. Biol. Chem.*, Bd. 278, Nr. 19, S. 16916–16925, Mai 2003.
- [128] C. Iacobucci, M. Götze, C. H. Ihling, C. Piotrowski, C. Arlt, M. Schäfer, C. Hage, R. Schmidt, und A. Sinz, „A cross-linking/mass spectrometry workflow based on MS-cleavable cross-

- linkers and the MeroX software for studying protein structures and protein-protein interactions“, *Nat Protoc*, Bd. 13, Nr. 12, S. 2864–2889, 2018.
- [129] C. Iacobucci, M. Götze, C. Piotrowski, C. Arlt, A. Rehkamp, C. Ihling, C. Hage, und A. Sinz, „Carboxyl-Photo-Reactive MS-Cleavable Cross-Linkers: Unveiling a Hidden Aspect of Diazirine-Based Reagents“, *Anal. Chem.*, Bd. 90, Nr. 4, S. 2805–2809, 20 2018.
- [130] M. Q. Müller, F. Dreiocker, C. H. Ihling, M. Schäfer, und A. Sinz, „Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS“, *Anal. Chem.*, Bd. 82, Nr. 16, S. 6958–6968, Aug. 2010.
- [131] E. D. Merkley, S. Rysavy, A. Kahraman, R. P. Hafen, V. Daggett, und J. N. Adkins, „Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine–lysine distances“, *Protein Sci*, Bd. 23, Nr. 6, S. 747–759, Juni 2014.
- [132] J. L. Vaughn, R. H. Goodwin, G. J. Tompkins, und P. McCawley, „The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae)“, *In Vitro*, Bd. 13, Nr. 4, S. 213–217, Apr. 1977.
- [133] T. Humphrey, G. Christofori, V. Lucijanic, und W. Keller, „Cleavage and polyadenylation of messenger RNA precursors in vitro occurs within large and specific 3' processing complexes“, *EMBO J.*, Bd. 6, Nr. 13, S. 4159–4168, Dez. 1987.
- [134] E. Bosse-Doenecke, U. Weininger, M. Gopalswamy, J. Balbach, S. M. Knudsen, und R. Rudolph, „High yield production of recombinant native and modified peptides exemplified by ligands for G-protein coupled receptors“, *Protein Expr. Purif.*, Bd. 58, Nr. 1, S. 114–121, März 2008.
- [135] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, und T. E. Ferrin, „UCSF Chimera--a visualization system for exploratory research and analysis“, *J Comput Chem*, Bd. 25, Nr. 13, S. 1605–1612, Okt. 2004.
- [136] G. C. P. van Zundert und A. M. J. J. Bonvin, „DisVis: quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes“, *Bioinformatics*, Bd. 31, Nr. 19, S. 3222–3224, Okt. 2015.
- [137] S. J. de Vries, M. van Dijk, und A. M. J. J. Bonvin, „The HADDOCK web server for data-driven biomolecular docking“, *Nat Protoc*, Bd. 5, Nr. 5, S. 883–897, Mai 2010.
- [138] G. C. P. van Zundert, J. P. G. L. M. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, M. van Dijk, S. J. de Vries, und A. M. J. J. Bonvin, „The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes“, *J. Mol. Biol.*, Bd. 428, Nr. 4, S. 720–725, Feb. 2016.
- [139] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, und V. Alva, „A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core“, *J. Mol. Biol.*, Bd. 430, Nr. 15, S. 2237–2243, 20 2018.

- [140] A. Roy, A. Kucukural, und Y. Zhang, „I-TASSER: a unified platform for automated protein structure and function prediction“, *Nat Protoc*, Bd. 5, Nr. 4, S. 725–738, Apr. 2010.
- [141] C. A. Schneider, W. S. Rasband, und K. W. Eliceiri, „NIH Image to ImageJ: 25 years of image analysis“, *Nat. Methods*, Bd. 9, Nr. 7, S. 671–675, Juli 2012.
- [142] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, und G. J. Barton, „Jalview Version 2--a multiple sequence alignment editor and analysis workbench“, *Bioinformatics*, Bd. 25, Nr. 9, S. 1189–1191, Mai 2009.
- [143] J. Cox und M. Mann, „MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification“, *Nat. Biotechnol.*, Bd. 26, Nr. 12, S. 1367–1372, Dez. 2008.
- [144] S. Tyanova, T. Temu, und J. Cox, „The MaxQuant computational platform for mass spectrometry-based shotgun proteomics“, *Nature Protocols*, Bd. 11, Nr. 12, S. 2301–2319, Dez. 2016.
- [145] M. Götze, J. Pettelkau, R. Fritzsche, C. H. Ihling, M. Schäfer, und A. Sinz, „Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis“, *J. Am. Soc. Mass Spectrom.*, Bd. 26, Nr. 1, S. 83–97, Jan. 2015.
- [146] B. Webb und A. Sali, „Comparative Protein Structure Modeling Using MODELLER“, *Curr Protoc Bioinformatics*, Bd. 54, S. 5.6.1-5.6.37, 20 2016.
- [147] J. Sambrook und D. W. Russell, *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press, 2001.
- [148] Life Technologies Corporation, *Bac-to-Bac® Baculovirus Expression System*. Carlsbad, CA, 2015.
- [149] U. K. Laemmli, „Cleavage of structural proteins during the assembly of the head of bacteriophage T4“, *Nature*, Bd. 227, Nr. 5259, S. 680–685, Aug. 1970.
- [150] G. E. Means und R. E. Feeney, „Reductive alkylation of amino groups in proteins“, *Biochemistry*, Bd. 7, Nr. 6, S. 2192–2201, Juni 1968.
- [151] W. C. Summers, „A simple method for extraction of RNA from E. coli utilizing diethyl pyrocarbonate“, *Anal. Biochem.*, Bd. 33, Nr. 2, S. 459–463, Feb. 1970.
- [152] T. Maniatis, A. Jeffrey, und H. Van deSande, „Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis“, *Biochemistry*, Bd. 14, Nr. 17, S. 3787–3794, Aug. 1975.
- [153] Roche Diagnostics GmbH, „SP6 RNA Polymerase“. 2016.
- [154] D. C. Rio, Hrsg., *RNA: a laboratory manual*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press, 2011.
- [155] New England Biolabs, „Radioactive Labeling with T4 PNK or T4 PNK (3' phosphatase minus)“.
- [156] U. Kühn, A. Nemeth, S. Meyer, und E. Wahle, „The RNA binding domains of the nuclear poly(A)-binding protein“, *J. Biol. Chem.*, Bd. 278, Nr. 19, S. 16916–16925, Mai 2003.

- [157] U. Kühn, M. Gündel, A. Knoth, Y. Kerwitz, S. Rüdell, und E. Wahle, „Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor“, *J. Biol. Chem.*, Bd. 284, Nr. 34, S. 22803–22814, Aug. 2009.
- [158] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis, und T. L. Madden, „NCBI BLAST: a better web interface“, *Nucleic Acids Res.*, Bd. 36, Nr. Web Server issue, S. W5-9, Juli 2008.
- [159] A. M. J. J. Bonvin, E. Karaca, P. L. Kastiris, und J. P. G. L. M. Rodrigues, „Defining distance restraints in HADDOCK“, *Nat Protoc*, Bd. 13, Nr. 7, S. 1503, 2018.
- [160] W. L. DeLano, „Pymol: An open-source molecular graphics tool“, *CCP4 Newsletter on protein crystallography*, Bd. 40, Nr. 1, S. 82–92, 2002.
- [161] H. Mi, A. Muruganujan, D. Ebert, X. Huang, und P. D. Thomas, „PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools“, *Nucleic Acids Res.*, Bd. 47, Nr. D1, S. D419–D426, Jan. 2019.
- [162] I. Berger, D. J. Fitzgerald, und T. J. Richmond, „Baculovirus expression system for heterologous multiprotein complexes“, *Nat. Biotechnol.*, Bd. 22, Nr. 12, S. 1583–1587, Dez. 2004.
- [163] M. Lorbeer, *Darstellung und Charakterisierung der 3'-Prozessierungsfaktoren CF Im und CstF*. Diplomarbeit an der Martin-Luther Universität Halle. Halle (Saale), 2012.
- [164] S.-Y. Sheu, D.-Y. Yang, H. L. Selzle, und E. W. Schlag, „Energetics of hydrogen bonds in peptides“, *Proc. Natl. Acad. Sci. U.S.A.*, Bd. 100, Nr. 22, S. 12683–12687, Okt. 2003.
- [165] A. Guo, H. Gu, J. Zhou, D. Mulhern, Y. Wang, K. A. Lee, V. Yang, M. Aguiar, J. Kornhauser, X. Jia, J. Ren, S. A. Beausoleil, J. C. Silva, V. Vemulapalli, M. T. Bedford, und M. J. Comb, „Immunoaffinity enrichment and mass spectrometry analysis of protein methylation“, *Mol. Cell Proteomics*, Bd. 13, Nr. 1, S. 372–387, Jan. 2014.
- [166] E. Wahle, „Poly(A) tail length control is caused by termination of processive synthesis“, *J. Biol. Chem.*, Bd. 270, Nr. 6, S. 2800–2808, 10 1995.
- [167] W. G. Touw, R. P. Joosten, und G. Vriend, „Detection of trans–cis flips and peptide-plane flips in protein structures“, *Acta Crystallogr D Biol Crystallogr*, Bd. 71, Nr. Pt 8, S. 1604–1614, Juli 2015.
- [168] C. A. Nelson, M. D. Fremont, J. R. Sedy, P. S. Norris, C. F. Ware, K. M. Murphy, und D. H. Fremont, „Structural Determinants of Herpesvirus Entry Mediator Recognition by Murine B and T Lymphocyte Attenuator“, *The Journal of Immunology*, Bd. 180, Nr. 2, S. 940–947, Jan. 2008.
- [169] A. Kolyada, C.-J. Lee, A. De Biasio, und N. Beglova, „A Novel Dimeric Inhibitor Targeting Beta2GPI in Beta2GPI/Antibody Complexes Implicated in Antiphospholipid Syndrome“, *PLoS ONE*, Bd. 5, Nr. 12, S. e15345, Dez. 2010.

- [170] M. Pennestri, S. Melino, G. M. Contessa, E. C. Casavola, M. Paci, A. Ragnini-Wilson, und D. O. Cicero, „Structural Basis for the Interaction of the Myosin Light Chain Mlc1p with the Myosin V Myo2p IQ Motifs“, *Journal of Biological Chemistry*, Bd. 282, Nr. 1, S. 667–679, Jan. 2007.
- [171] Y. Wu, T. R. Albrecht, D. Baillat, E. J. Wagner, und L. Tong, „Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance“, *Proc. Natl. Acad. Sci. U.S.A.*, Bd. 114, Nr. 17, S. 4394–4399, 25 2017.
- [172] N. Suzuki, A. Rohaim, R. Kato, I. Dikic, S. Wakatsuki, und M. Kawasaki, „A novel mode of ubiquitin recognition by the ubiquitin-binding zinc finger domain of WRNIP1“, *The FEBS Journal*, Bd. 283, Nr. 11, S. 2004–2017, Juni 2016.
- [173] X. Jia, E. Weber, A. Tokarev, M. Lewinski, M. Rizk, M. Suarez, J. Guatelli, und Y. Xiong, „Structural basis of HIV-1 Vpu-mediated BST2 antagonism via hijacking of the clathrin adaptor protein complex 1“, *Elife*, Bd. 3, S. e02362, Apr. 2014.
- [174] A. Kahraman, L. Malmström, und R. Aebersold, „Xwalk: computing and visualizing distances in cross-linking experiments“, *Bioinformatics*, Bd. 27, Nr. 15, S. 2163–2164, Aug. 2011.
- [175] M. Götze, J. Dufourt, C. Ihling, C. Rammelt, S. Pierson, N. Sambrani, C. Temme, A. Sinz, M. Simonelig, und E. Wahle, „Translational repression of the *Drosophila* nanos mRNA involves the RNA helicase Belle and RNA coating by Me31B and Trailer hitch“, *RNA*, Bd. 23, Nr. 10, S. 1552–1568, 2017.
- [176] C. L. Moore, H. Skolnik-David, und P. A. Sharp, „Analysis of RNA cleavage at the adenovirus-2 L3 polyadenylation site.“, *EMBO J*, Bd. 5, Nr. 8, S. 1929–1938, Aug. 1986.
- [177] M. Suchanek, A. Radzikowska, und C. Thiele, „Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells“, *Nat. Methods*, Bd. 2, Nr. 4, S. 261–267, Apr. 2005.
- [178] B. Häupl, C. H. Ihling, und A. Sinz, „Combining affinity enrichment, cross-linking with photo amino acids, and mass spectrometry for probing protein kinase D2 interactions“, *Proteomics*, Bd. 17, Nr. 10, S. e1600459, Mai 2017.
- [179] Q. Zhang, E. Crosland, und D. Fabris, „Nested Arg-specific bifunctional crosslinkers for MS-based structural analysis of proteins and protein assemblies“, *Anal Chim Acta*, Bd. 627, Nr. 1, S. 117–128, Okt. 2008.
- [180] Z. Zhang, J. Fu, und D. S. Gilmour, „CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11“, *Genes Dev.*, Bd. 19, Nr. 13, S. 1572–1580, Juli 2005.
- [181] A. A. Adzhubei, M. J. E. Sternberg, und A. A. Makarov, „Polyproline-II Helix in Proteins: Structure and Function“, *Journal of Molecular Biology*, Bd. 425, Nr. 12, S. 2100–2132, Juni 2013.

- [182] A. Skrajna, X.-C. Yang, M. Dadlez, W. F. Marzluff, und Z. Dominski, „Protein composition of catalytically active U7-dependent processing complexes assembled on histone pre-mRNA containing biotin and a photo-cleavable linker“, *Nucleic Acids Res.*, Bd. 46, Nr. 9, S. 4752–4770, 18 2018.
- [183] N. G. Kolev, T. A. Yario, E. Benson, und J. A. Steitz, „Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation“, *EMBO Rep*, Bd. 9, Nr. 10, S. 1013–1018, Okt. 2008.
- [184] M. Schmidt, *Rekombinante Darstellung und Charakterisierung von CstF-Komplexen*. Masterarbeit an der Martin-Luther Universität Halle. Halle (Saale), 2018.
- [185] I. L. de la Sierra-Gallay, L. Zig, A. Jamalli, und H. Putzer, „Structural insights into the dual activity of RNase J“, *Nature Structural & Molecular Biology*, Bd. 15, Nr. 2, S. 206–212, Feb. 2008.
- [186] Y. Nishida, H. Ishikawa, S. Baba, N. Nakagawa, S. Kuramitsu, und R. Masui, „Crystal structure of an archaeal cleavage and polyadenylation specificity factor subunit from *Pyrococcus horikoshii*“, *Proteins: Structure, Function, and Bioinformatics*, Bd. 78, Nr. 10, S. 2395–2398, Mai 2010.
- [187] A. Strauss, F. Bitsch, G. Fendrich, P. Graff, R. Knecht, B. Meyhack, und W. Jahnke, „Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells“, *J. Biomol. NMR*, Bd. 31, Nr. 4, S. 343–349, Apr. 2005.
- [188] Q. Yang, M. Coseno, G. M. Gilmartin, und S. Doublé, „Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping“, *Structure*, Bd. 19, Nr. 3, S. 368–377, März 2011.
- [189] C. Bleiholder, S. Suhai, A. G. Harrison, und B. Paizs, „Towards Understanding the Tandem Mass Spectra of Protonated Oligopeptides. 2: The Proline Effect in Collision-Induced Dissociation of Protonated Ala-Ala-Xxx-Pro-Ala (Xxx = Ala, Ser, Leu, Val, Phe, and Trp)“, *J. Am. Soc. Mass Spectrom.*, Bd. 22, Nr. 6, S. 1032–1039, Juni 2011.
- [190] A. Scaiola, M. Leibundgut, D. Boehringer, P. Caspers, D. Bur, H. H. Locher, G. Rueedi, und D. Ritz, „Structural basis of translation inhibition by cadazolid, a novel quinoxolidinone antibiotic“, *Sci Rep*, Bd. 9, Nr. 1, S. 1–9, Apr. 2019.
- [191] P. D. Adams *u. a.*, „Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB)“, *Acta Crystallographica Section D Structural Biology*, Bd. 75, Nr. 4, S. 451–454, Apr. 2019.
- [192] „van der Waals Volumes and Radii | The Journal of Physical Chemistry“. [Online]. Verfügbar unter: <https://pubs.acs.org/doi/abs/10.1021/j100785a001>. [Zugegriffen: 15-Juli-2019].
- [193] D. Pal und P. Chakrabarti, „Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations“, *J. Mol. Biol.*, Bd. 294, Nr. 1, S. 271–288, Nov. 1999.

- [194] K. S. Keating, E. L. Humphris, und A. M. Pyle, „A new way to see RNA“, *Quarterly reviews of biophysics*, Bd. 44, Nr. 4, S. 433, Nov. 2011.
- [195] S. H. Yoshimura und T. Hirano, „HEAT repeats - versatile arrays of amphiphilic helices working in crowded environments?“, *J. Cell. Sci.*, Bd. 129, Nr. 21, S. 3963–3970, 01 2016.
- [196] P. Emsley, B. Lohkamp, W. G. Scott, und K. Cowtan, „Features and development of Coot“, *Acta Crystallogr. D Biol. Crystallogr.*, Bd. 66, Nr. Pt 4, S. 486–501, Apr. 2010.
- [197] T. E. Wales und J. R. Engen, „Hydrogen exchange mass spectrometry for the analysis of protein dynamics“, *Mass Spectrom Rev*, Bd. 25, Nr. 1, S. 158–170, Feb. 2006.
- [198] D. T. Johnson, L. H. Di Stefano, und L. M. Jones, „Fast photochemical oxidation of proteins (FPOP): A powerful mass spectrometry-based structural proteomics tool“, *J. Biol. Chem.*, Bd. 294, Nr. 32, S. 11969–11979, Aug. 2019.
- [199] J. D. Eaton und S. West, „An end in sight? Xrn2 and transcriptional termination by RNA polymerase II“, *Transcription*, Bd. 9, Nr. 5, S. 321–326, 2018.
- [200] S. Kaneko, O. Rozenblatt-Rosen, M. Meyerson, und J. L. Manley, „The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination“, *Genes Dev*, Bd. 21, Nr. 14, S. 1779–1789, Juli 2007.
- [201] S. Sato, H. Ishikawa, H. Yoshikawa, K. Izumikawa, R. J. Simpson, und N. Takahashi, „Collaborator of alternative reading frame protein (CARF) regulates early processing of pre-ribosomal RNA by retaining XRN2 (5'-3' exoribonuclease) in the nucleoplasm“, *Nucleic Acids Res*, Bd. 43, Nr. 21, S. 10397–10410, Dez. 2015.
- [202] V. Palaniswamy, K. C. M. Moraes, C. J. Wilusz, und J. Wilusz, „Nucleophosmin is selectively deposited on mRNA during polyadenylation“, *Nat Struct Mol Biol*, Bd. 13, Nr. 5, S. 429–435, Mai 2006.
- [203] A. Skerra und T. G. M. Schmidt, „Use of the Strep- tag and streptavidin for detection and purification of recombinant proteins“, in *Methods in Enzymology*, Bd. 326, Elsevier, 2000, S. 271–304.
- [204] S. Awasthi und J. C. Alwine, „Association of polyadenylation cleavage factor I with U1 snRNP“, *RNA*, Bd. 9, Nr. 11, S. 1400–1409, Nov. 2003.
- [205] J. Rappsilber, U. Ryder, A. I. Lamond, und M. Mann, „Large-scale proteomic analysis of the human spliceosome“, *Genome Res.*, Bd. 12, Nr. 8, S. 1231–1245, Aug. 2002.
- [206] R. Vallejos Baier, J. Picao-Osorio, und C. R. Alonso, „Molecular Regulation of Alternative Polyadenylation (APA) within the Drosophila Nervous System“, *Journal of Molecular Biology*, Bd. 429, Nr. 21, S. 3290–3300, Okt. 2017.
- [207] B. R. Graveley, „Getting in the Loop: New Insights into the Mechanism of Poly(A) Site Recognition“, *Structure*, Bd. 19, Nr. 3, S. 279–281, März 2011.
- [208] S. D. Lee und C. L. Moore, „Efficient mRNA Polyadenylation Requires a Ubiquitin-Like Domain, a Zinc Knuckle, and a RING Finger Domain, All Contained in the Mpe1 Protein“, *Mol Cell Biol*, Bd. 34, Nr. 21, S. 3955–3967, Nov. 2014.

- [209] E. Afonina, R. Stauber, und G. N. Pavlakis, „The human poly(A)-binding protein 1 shuttles between the nucleus and the cytoplasm“, *J. Biol. Chem.*, Bd. 273, Nr. 21, S. 13015–13021, Mai 1998.
- [210] G. R. Kumar, L. Shum, und B. A. Glaunsinger, „Importin α -Mediated Nuclear Import of Cytoplasmic Poly(A) Binding Protein Occurs as a Direct Consequence of Cytoplasmic mRNA Depletion“, *Mol Cell Biol*, Bd. 31, Nr. 15, S. 3113–3125, Aug. 2011.
- [211] M. Görlach, C. G. Burd, und G. Dreyfuss, „The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity“, *Exp. Cell Res.*, Bd. 211, Nr. 2, S. 400–407, Apr. 1994.
- [212] E. Hampden-Smith, „The Localization of PABPC1 in HeLa Cells“, *Eukaryon*, Bd. 10, 2014.
- [213] J.-H. Lee, J. You, E. Dobrota, und D. G. Skalnik, „Identification and characterization of a novel human PP1 phosphatase complex“, *J. Biol. Chem.*, Bd. 285, Nr. 32, S. 24466–24476, Aug. 2010.
- [214] M. Lidschreiber, A. D. Easter, S. Battaglia, J. B. Rodríguez-Molina, A. Casañal, M. Carminati, C. Baejen, P. Grzechnik, K. C. Maier, P. Cramer, und L. A. Passmore, „The APT complex is involved in non-coding RNA transcription and is distinct from CPF“, *Nucleic Acids Res*, Bd. 46, Nr. 21, S. 11528–11538, Nov. 2018.
- [215] A. Kumar, M. Clerici, L. M. Muckenfuss, L. A. Passmore, und M. Jinek, „Mechanistic insights into mRNA 3'-end processing“, *Curr. Opin. Struct. Biol.*, Bd. 59, S. 143–150, Dez. 2019.
- [216] Y. Shi, B. Reddy, und J. L. Manley, „PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins“, *Mol. Cell*, Bd. 23, Nr. 6, S. 819–829, Sep. 2006.
- [217] K. Ryan, „Pre-mRNA 3' Cleavage is Reversibly Inhibited In Vitro by Cleavage Factor Dephosphorylation“, *RNA Biol*, Bd. 4, Nr. 1, S. 26–33, 2007.
- [218] H. Ceulemans, W. Stalmans, und M. Bollen, „Regulator-driven functional diversification of protein phosphatase-1 in eukaryotic evolution“, *BioEssays*, Bd. 24, Nr. 4, S. 371–381, Apr. 2002.
- [219] J. Kirchner, S. Gross, D. Bennett, und L. Alphey, „Essential, Overlapping and Redundant Roles of the Drosophila Protein Phosphatase 1 α and 1 β Genes“, *Genetics*, Bd. 176, Nr. 1, S. 273–281, Mai 2007.
- [220] J. A. Gibbons, L. Kozubowski, K. Tatchell, und S. Shenolikar, „Expression of Human Protein Phosphatase-1 in *Saccharomyces cerevisiae* Highlights the Role of Phosphatase Isoforms in Regulating Eukaryotic Functions“, *Journal of Biological Chemistry*, Bd. 282, Nr. 30, S. 21838–21847, Juli 2007.
- [221] B. Zhu, S. S. Mandal, A.-D. Pham, Y. Zheng, H. Erdjument-Bromage, S. K. Batra, P. Tempst, und D. Reinberg, „The human PAF complex coordinates transcription with events downstream of RNA synthesis“, *Genes Dev.*, Bd. 19, Nr. 14, S. 1668–1673, Juli 2005.

- [222] C. L. Mueller, S. E. Porter, M. G. Hoffman, und J. A. Jaehning, „The Paf1 complex has functions independent of actively transcribing RNA polymerase II“, *Mol. Cell*, Bd. 14, Nr. 4, S. 447–456, Mai 2004.
- [223] Y. Araki, S. Takahashi, T. Kobayashi, H. Kajihō, S. Hoshino, und T. Katada, „Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast“, *EMBO J*, Bd. 20, Nr. 17, S. 4684–4693, Sep. 2001.
- [224] X. Chu, X. Qin, H. Xu, L. Li, Z. Wang, F. Li, X. Xie, H. Zhou, Y. Shen, und J. Long, „Structural insights into Paf1 complex assembly and histone binding“, *Nucleic Acids Research*, Bd. 41, Nr. 22, S. 10619–10629, Dez. 2013.
- [225] N. Viphakone, G. M. Hautbergue, M. Walsh, C.-T. Chang, A. Holland, E. G. Folco, R. Reed, und S. A. Wilson, „TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export“, *Nat Commun*, Bd. 3, S. 1006, 2012.
- [226] S. Masuda, R. Das, H. Cheng, E. Hurt, N. Dorman, und R. Reed, „Recruitment of the human TREX complex to mRNA during splicing“, *Genes Dev.*, Bd. 19, Nr. 13, S. 1512–1517, Juli 2005.
- [227] E. G. Folco, C.-S. Lee, K. Dufu, T. Yamazaki, und R. Reed, „The proteins PDIP3 and ZC11A associate with the human TREX complex in an ATP-dependent manner and function in mRNA export“, *PLoS ONE*, Bd. 7, Nr. 8, S. e43804, 2012.
- [228] K. Strässer, S. Masuda, P. Mason, J. Pfannstiel, M. Oppizzi, S. Rodriguez-Navarro, A. G. Rondón, A. Aguilera, K. Struhl, R. Reed, und E. Hurt, „TREX is a conserved complex coupling transcription with messenger RNA export“, *Nature*, Bd. 417, Nr. 6886, S. 304–308, Mai 2002.
- [229] G. Singh, A. Kucukural, C. Cenik, J. D. Leszyk, S. A. Shaffer, Z. Weng, und M. J. Moore, „The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus“, *Cell*, Bd. 151, Nr. 4, S. 750–764, Nov. 2012.
- [230] H. Cheng, K. Dufu, C.-S. Lee, J. L. Hsu, A. Dias, und R. Reed, „Human mRNA export machinery recruited to the 5' end of mRNA“, *Cell*, Bd. 127, Nr. 7, S. 1389–1400, Dez. 2006.
- [231] S. A. Johnson, G. Cubberley, und D. L. Bentley, „Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11“, *Mol. Cell*, Bd. 33, Nr. 2, S. 215–226, Jan. 2009.
- [232] A. Castello, B. Fischer, K. Eichelbaum, R. Horos, B. M. Beckmann, C. Strein, N. E. Davey, D. T. Humphreys, T. Preiss, L. M. Steinmetz, J. Krijgsveld, und M. W. Hentze, „Insights into RNA biology from an atlas of mammalian mRNA-binding proteins“, *Cell*, Bd. 149, Nr. 6, S. 1393–1406, Juni 2012.
- [233] T. Gonatopoulos-Pournatzis und V. H. Cowling, „Cap-binding complex (CBC)“, *Biochem J*, Bd. 457, Nr. Pt 2, S. 231–242, Jan. 2014.

- [234] S. M. Flaherty, P. Fortes, E. Izaurralde, I. W. Mattaj, und G. M. Gilmartin, „Participation of the nuclear cap binding complex in pre-mRNA 3' processing“, *Proc. Natl. Acad. Sci. U.S.A.*, Bd. 94, Nr. 22, S. 11893–11898, Okt. 1997.
- [235] M. R. Davis, M. Delaleau, und K. L. B. Borden, „Nuclear eIF4E Stimulates 3'-End Cleavage of Target RNAs“, *Cell Rep*, Bd. 27, Nr. 5, S. 1397-1408.e4, Apr. 2019.
- [236] J. Lykke-Andersen, M. D. Shu, und J. A. Steitz, „Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1“, *Science*, Bd. 293, Nr. 5536, S. 1836–1839, Sep. 2001.
- [237] T. Ø. Tange, T. Shibuya, M. S. Jurica, und M. J. Moore, „Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core“, *RNA*, Bd. 11, Nr. 12, S. 1869–1883, Dez. 2005.
- [238] Z. Wang, L. Ballut, I. Barbosa, und H. Le Hir, „Exon Junction Complexes can have distinct functional flavours to regulate specific splicing events“, *Sci Rep*, Bd. 8, Nr. 1, S. 9509, Juni 2018.
- [239] T. Shibuya, T. Ø. Tange, N. Sonenberg, und M. J. Moore, „eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay“, *Nat. Struct. Mol. Biol.*, Bd. 11, Nr. 4, S. 346–351, Apr. 2004.
- [240] M. H. Viegas, N. H. Gehring, S. Breit, M. W. Hentze, und A. E. Kulozik, „The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway“, *Nucleic Acids Res.*, Bd. 35, Nr. 13, S. 4542–4551, 2007.
- [241] P.-E. Chazal, E. Dagueneau, C. Wendling, N. Ulryck, C. Tomasetto, B. Sargueil, und H. Le Hir, „EJC core component MLN51 interacts with eIF3 and activates translation“, *Proc. Natl. Acad. Sci. U.S.A.*, Bd. 110, Nr. 15, S. 5903–5908, Apr. 2013.

7. Anhang

7.1 Abbildungen

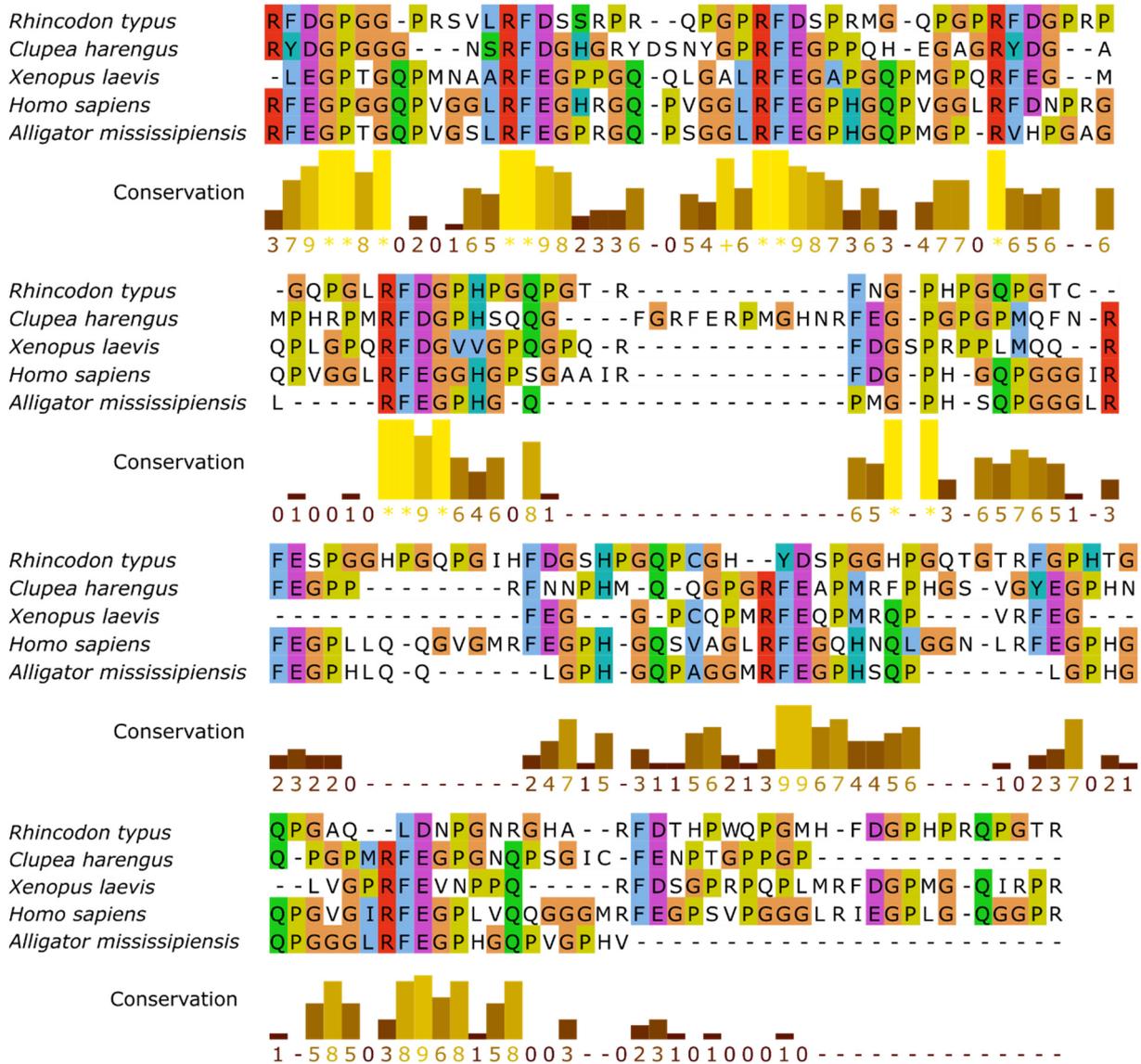


Abbildung A. 1 Alignment FEGR-Wiederholungen verschiedener Organismen. Ausgewählt wurden Organismen verschiedener Großgruppen: Knorpelfische: Walhai (*Rhincodon typus*; Aminosäuren 1160 bis 1317); Knochenfische: Atlantischer Hering (*Clupea harengus*; Aminosäuren 957 bis 1103); Amphibien: Glatter Krallenfrosch (*Xenopus laevis*; Aminosäuren 957 bis 1090); Reptilien: Mississippi-Alligator (*Alligator mississippiensis*; Aminosäuren 989 bis 1108); Säugetiere: Mensch (*Homo sapiens*; Aminosäuren 864 bis 1108). Die Sequenzen wurde händisch anhand des Auftretens der charakteristischen Sequenzmerkmale, dem Starttriplet FEG und dessen Derivate sowie dem schließenden Arginin, ausgerichtet.

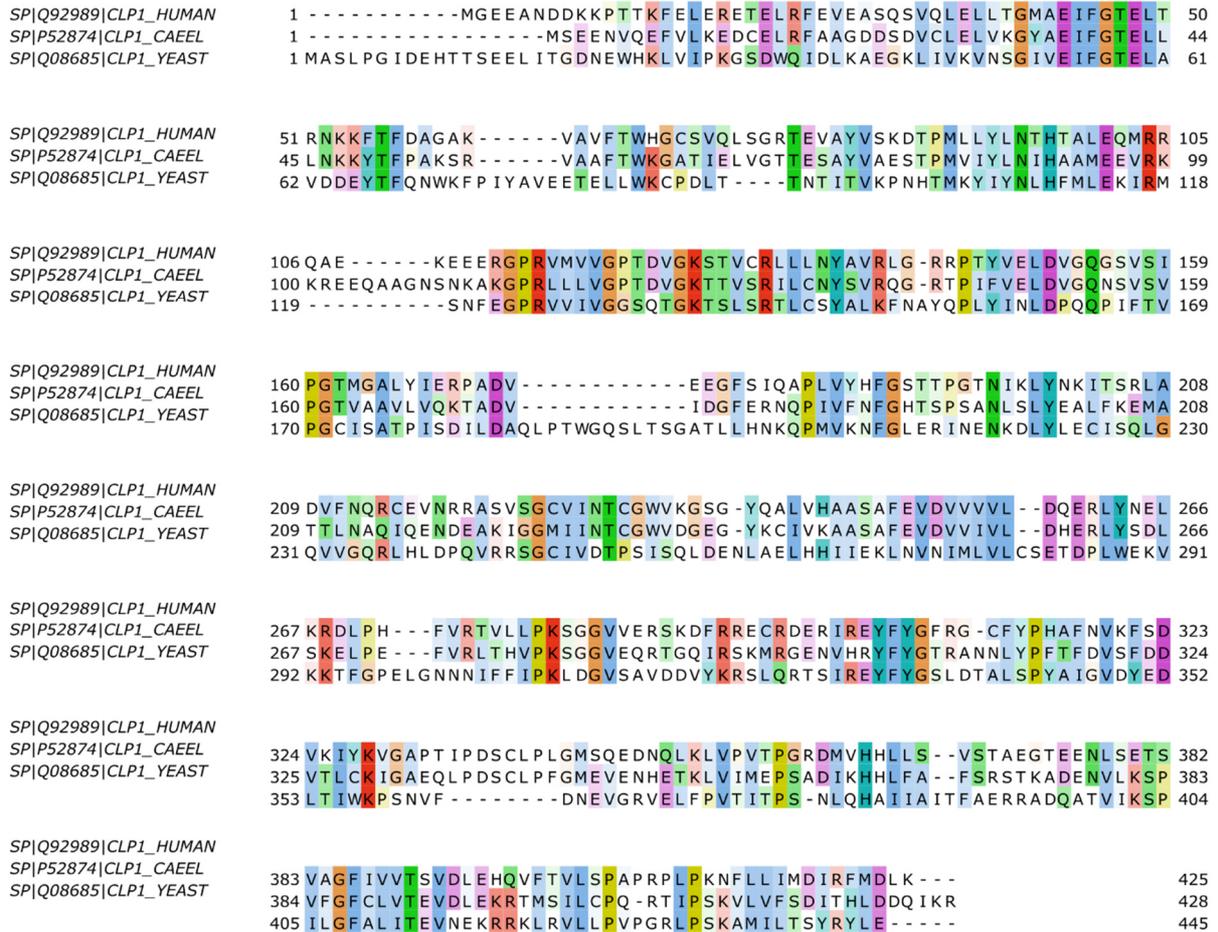


Abbildung A. 2 Clp1 Alignment. Dargestellt ist das Sequenzalignment der Clp1-Sequenzen aus Menschen (CLP1_HUMAN), *C. elegans* (CLP1_CAEEL) sowie Bäckerhefe (CLP1_YEAST). Das Alignment wurde von UniProt generiert. Ähnliche oder identische Aminosäuren sind farblich hervorgehoben.

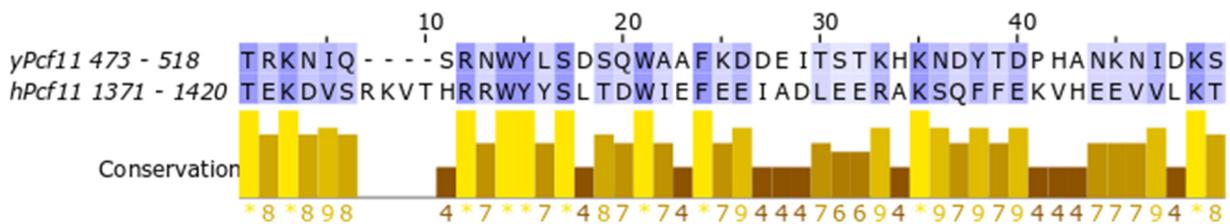


Abbildung A.3 Sequenzalignment der Interaktionsregion von Pcf11 aus Hefe und Mensch. Die Aminosäuresequenz der Hefe (*Saccharomyces cerevisiae*; yPcf11) von den Aminosäuren 473 bis 518 ist gegen die humane Sequenz (*Homo sapiens*; hPcf11) von den Aminosäuren 1371 bis 1420 dargestellt. Identische oder ähnliche Aminosäuren sind blau hinterlegt. Unterhalb des Alignment ist die Sequenzkonservierung als Balkendiagramm dargestellt.

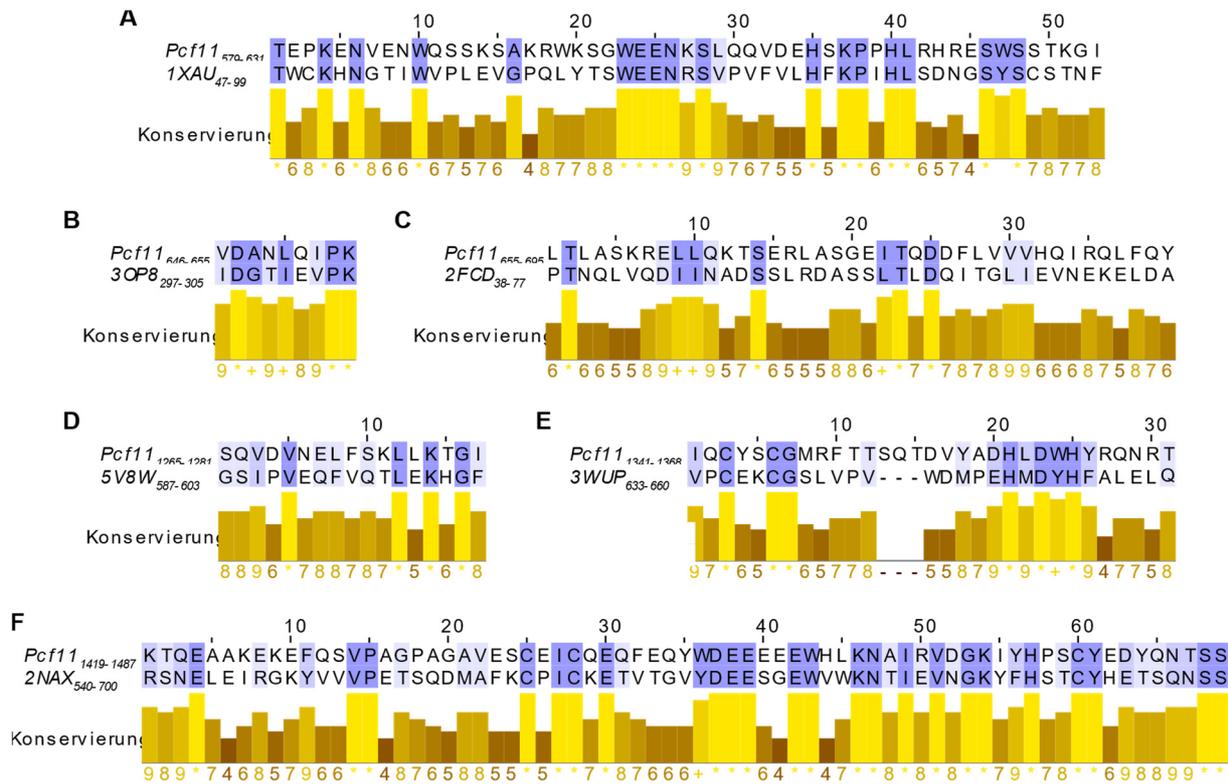


Abbildung A.4 Sequenzalignement von *Pcf11* mit den identifizierten homologen Strukturen.

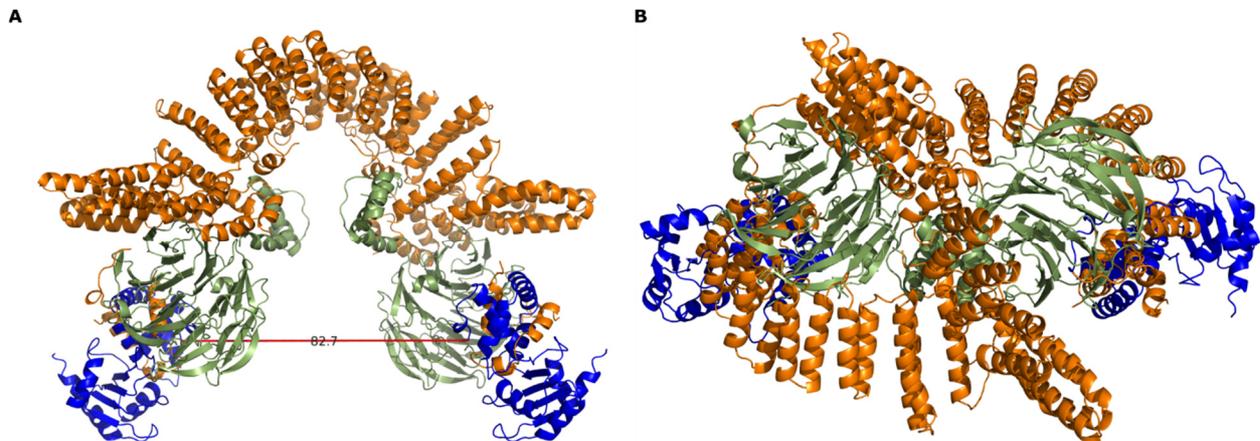


Abbildung A. 5 Hexamere Strukturen von *CstF*. A) Die Struktur wurde durch Überlagerung der HAT-Domänen des trimeren Modells mit der publizierten Struktur des dimeren *CstF77* generiert. Die eingezeichnete Linie symbolisiert den identifizierten *cross-link* zwischen den Lysinen 189 von *CstF64*. B) Diese Struktur ist durch Überlagerung der *CstF50*-Dimerisierungsdomänen des trimeren Modells mit der publizierten Struktur des dimeren *CstF50* generiert worden. Es kommt hierbei zum Zusammenstoß der HAT-Domäne der einen mit *CstF64* der anderen Struktur. Darüber hinaus widerspricht dieses Modell der Struktur der dimeren *CstF77*-HAT-Domäne.

7.2 Tabellen

7.2.1 Material

Tabelle A 1 Chemikalienliste. Übersicht der in dieser Arbeit verwendeten Chemikalien einschließlich ihres Herstellers.

Chemikalien	Hersteller
β-Mercaptoethanol	Sigma-Aldrich
4-(2-Aminoethyl)benzonsulfonylfluorid (AEBSF)	AppliChem
Aceton	Roth
Acrylamid 30 % (37,5:1)	Serva
Acrylamid 40 % (19:1)	Serva
Agar-Agar	Carl Roth
Agarose, Ultrapure	Invitrogen
Ammoniumbicarbonat	Sigma-Aldrich
Ammoniumhydrogencarbonat	Sigma-Aldrich
Coomassie G250	Serva
Diethylpyrocarbonat (DEPC)	Sigma-Aldrich
Dithiothreitol (DTT)	Gerbu
DMSO (auf Molsieb)	Sigma-Aldrich
DSBU	CF Plus Chemicals
Essigsäure	Roth
Ethanol	Honeywell
FLAG-Peptid (Sequenz: DYKDDDDK)	Sigma-Aldrich
Formaldehyd (37 %)	Sigma-Aldrich
Glycerin	Roth
Glycin	Serva
Glycogen	Serva
Harnstoff	Roth
Hefe-RNA (total)	Roche
Irgasafe Plus	PerkinElmer
Isopropyl-β-D-thiogalactopyranosid	Sigma-Aldrich
Kreatinphosphat	Sigma-Aldrich
LC-SDA	Thermo-Fisher
Leupeptin	Roche
Magermilchpulver	Sucofin
Methanol	Roth
Natriumdodecylsulfat (SDS)	Serva
Nonident™ P-40 (NP-40)	Sigma-Aldrich
PageRuler <i>Prestained Protein Ladder</i> (10 - 180 kDa)	Thermo Fisher
PageRuler <i>Unstained Protein Ladder</i> (10 - 200 kDa)	Thermo Fisher
PageRuler <i>Unstained Protein Ladder</i> (14,4 - 116 kDa)	Thermo Fisher

Pepstatin	Roche
Phenylmethylsulfonylfluorid (PMSF)	Sigma-Aldrich
Ponceau S	Sigma-Aldrich
tRNA (<i>E. coli</i>)	Roche
Trypanblau-Lösung	BioRad
X-Gal	Peqlab

Nukleotide

Chemikalie	Hersteller
α -[³² P]-UTP	PerkinElmer / Hartmann Analytic ¹
γ -[³² P]-ATP	PerkinElmer / Hartmann Analytic ¹
3'-O-Me-m7G(5')ppp(5')G RNA <i>Cap Structure Analog</i>	NEB
Cordycepin-Triphosphat	Sigma-Aldrich
Nukleotid-Triphosphate	Sigma-Aldrich

Chromatographie-Material

Material	Hersteller
ANTI-FLAG M2 Affinity Gel	Merck
GSH-Sepharose	GE Healthcare
MonoQ (1 ml)	Pharmacia
NiNTA Agarose	Qiagen
NiNTA Magnetic Agarose Beads	Qiagen
Strep-Tactin Sepharose	IBA

Sonstige Materialien

Material	Hersteller
Dialyseschläuche	Serva
Plasmid Mini Kit	Qiagen
Plasmid Midi Kit	Qiagen
Nitrozellulosefilter NC20 (für Filterbindungsassay)	Whatman
Nitrozellulosemembran Protran 0.1 (für <i>Western-Blot</i>)	GE Healthcare

¹ Während der Erstellung dieser Arbeit wurde der Hersteller der radioaktiv-markierten Nukleotide gewechselt.

Tabelle A 2 Bakterienstämme. Übersicht der in dieser Arbeit verwendeten Bakterienstämme, einschließlich ihres Genotypes und ihrer Herkunft.

Name	Genotyp	Herkunft
XL1-Blue	endA1 gyrA96(nal ^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB ⁺ lacI ^q Δ(lacZ)M15] hsdR17(r _K ⁻ m _K ⁺)	Stratagene
DH10B	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara leu)7697 galU galK λ-rpsL nupG/pMON14272/ pMON7124	Invitrogen
DH10B MultiBac	Enthält das MultiBac-Bacmid. F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara leu)7697 galU galK λ-rpsL nupG/pMON14272/ pMON7124	T. J. Richmond, ETH Zürich [162]
Rosetta (DE3)	Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL (DE3) F'[lac ⁺ lacI ^q pro] gor522::Tn10 trxB pRARE (Cam ^R)	Merck

Tabelle A 3 RNA-Sequenzen. Sequenzen der in dieser Arbeit als Substrate verwendeten RNA. Das Polyadenylierungssignal (wt und Δ) ist fett markiert.

Name	Sequenz
RNA-Oligo <i>CFI 2sites</i> ¹	UGUAAACAGAAUAGUGUAA
pSP64-L3 <i>DraI</i> -verdaut ¹	GAAUACAAGCUUGGGCUGCAGGUCGACUCUAGAGGAUCCCCG GGCGAGCUCGAAUUCGGCUUCCUGGAGCGCCACUCGCCCUAC UUCGCGAGCCACAGUGCGCAGAUUAGGAGCGCCACUUCUUUU UGUCACUUGAAAAACAUGUAAAAUAAUGUACUAGGAGACAC UUUC AAUAAA GGCAAUGUUUUUAAUUUGUACACUCUCGGGUG AUUAUUUACCCCCACCCUUGCCGUCUCGCGCCGUUU
pSP64-L3 Δ <i>DraI</i> -verdaut ¹	GAAUACAAGCUUGGGCUGCAGGUCGACUCUAGAGGAUCCCCG GGCGAGCUCGAAUUCGGCUUCCUGGAGCGCCACUCGCCCUAC UUCGCGAGCCACAGUGCGCAGAUUAGGAGCGCCACUUCUUUU UGUCACUUGAAAAACAUGUAAAAUAAUGUACUAGGAGACAC UUUC AAGAAA GGCAAUGUUUUUAAUUUGUACACUCUCGGGUG AUUAUUUACCCCCACCCUUGCCGUCUCGCGCCGUUU
pSP64-L3prä-A15, <i>BbsI</i> -verdaut ¹	GAAUACAAGCUUGGGCUGCAGGUCACUAGGAGACACUUUC AA UAAA GGCAAUGUUUUUAAUUUGU-A ₁₅
pSP64-L3prä-A15 Δ, <i>BbsI</i> -verdaut ¹	GAAUACAAGCUUGGGCUGCAGGUCACUAGGAGACACUUUC AA GAAA GGCAAUGUUUUUAAUUUGU-A ₁₅

Tabelle A 4 Protein, Enzyme und Antikörper. Übersicht der in dieser Arbeit verwendeten Proteine, Enzyme, Antikörper und Extrakte sowie ihre Herkunft.

Proteine	Hersteller / Bereitgestellt von
BSA	Merck
Methyliertes BSA	Diese Arbeit
Nukleares Poly(A)-bindendes Protein (PABPN1)	Uwe Kühn
Phusion DNA-Polymerase	Bodo Moritz
PreScission™	GE Healthcare
Proteinase K	Merck
Restriktionsendonukleasen	NEB
RNase-Inhibitor	NEB
SP6 RNA-Polymerase	Roche
Sumo-Protease (Ulp1)	Bodo Moritz
T4 DNA-Ligase	NEB
T4 RNA-Polynukleotidkinase	NEB
Taq DNA-Polymerase	NEB
Trypsin	Promega
Antikörper	
α -CPSF100, SA1366 (Antikörper)	Gudrun Scholz, Uwe Kühn
α -hFip1, SZ1005 (Antikörper)	AG Keller, Universität Basel
α -Kaninchen, IRDye (Antikörper)	LI-COR Bioscience
Extrakte	
HeLa-Kernextrakt	Ipracell
HeLa-Kernextrakt (Fraktion A8)	Florian Kluge
HeLa-Kernextrakt (Fraktion MQ)	Lars Schönemann

7.2.2 Massenspektrometrische Methoden

Proteinidentifikation

Zur Verifizierung der Proteinidentität wurde das Q-Exactive™ Plus-Massenspektrometer mit folgendem Gradienten und Analyseinstellungen verwendet:

Gradient	
Laden	15 min auf Vorsäule (Flussrate: 30 μ l/min)
Elution	30 min linear von 3 auf 50 % B, in 2 min auf 99 % B, 5 min bei 99 %, in 2 min auf 3 % B (Flussrate: 300 μ l/min; Säulentemperatur 45 °C)
MS-Methode	

MS	299 – 1799 m/z, Auflösung 140.000 bei m/z 200, AGC-target (<i>automated gain control</i>) 3.000.000, <i>maximum injection time</i> 30 ms
MS/MS	10 <i>most intense</i> aus MS, <i>charge states</i> 2-6, <i>isolation window</i> 1,6 m/z, Auflösung 35.000 bei m/z 200, AGC target 100.000, <i>maximum injection time</i> 100 ms, HCD (28 % NCE (<i>normalized collision energy</i>)), <i>dynamic exclusion</i> für 30 s

Analyse chemischer cross-links

Zur massenspektrometrischen Analyse von chemischen *cross-links* wurden beide beschriebenen Massenspektrometer verwendet.

Für das Orbitrap Fusion™-Massenspektrometer wurden zur Identifikation von CFII *cross-links* folgende Einstellungen verwendet:

Gradient	
Laden	15 min auf Vorsäule (Flussrate: 30 µl/min)
Elution	90 min linear von 1 auf 35 % B, in 5 min auf 85 % B, 5 min bei 85 %, in 1 min auf 1 % B (Flussrate: 300 µl/min; Säulentemperatur 40 °C)
MS-Methode	
MS	300 – 1500 m/z, Auflösung 120.000 bei m/z 200, AGC-target (<i>automated gain control</i>) 400.000, <i>maximum injection time</i> 50 ms, <i>internal calibration</i>
MS/MS	<i>most intense</i> aus MS, sovieler wie innerhalb 5 s mgl., <i>charge states</i> 3-8, <i>isolation window</i> 2 m/z, Auflösung 15.000 bei m/z 200, AGC target 50.000, <i>maximum injection time</i> 250 ms, HCD (30 % oder 35% NCE (<i>normalized collision energy</i>)) oder <i>stepped</i> HCD (24, 29, 34 % NCE), <i>dynamic exclusion</i> für 60 s

Für das Q-Exactive™-Massenspektrometer wurden folgende Einstellungen verwendet:

Gradient	
Laden	15 min auf Vorsäule (Flussrate: 75 µl/min)
Elution	90 min linear von 3 auf 40 % B, in 2 min auf 99 % B, 5 min bei 99 %, in 2 min auf 3 % B (Flussrate: 230 µl/min; Säulentemperatur 45 °C)
MS-Methode	
MS	375 – 1799 m/z, Auflösung 140.000 bei m/z 200, AGC-target (<i>automated gain control</i>) 3.000.000, <i>maximum injection time</i> 100 ms
MS/MS	10 <i>most intense</i> aus MS, <i>charge states</i> 3-7, <i>isolation window</i> 2 m/z, Auflösung 17.500 bei m/z 200, AGC target 200.000, <i>max. inj. time</i> 250 ms, <i>stepped</i> HCD (27, 30, 33 % NCE (<i>normalized collision energy</i>)), <i>dynamic exclusion</i> für 60 s

Interaktomanalyse

Für die Interaktomanalyse wurden beide beschriebenen Massenspektrometer verwendet. Hierbei wurden für diese folgende Einstellungen verwendet.

Für das Orbitrap Fusion™-Massenspektrometer:

Gradient	
Laden	15 min auf Vorsäule (Flussrate: 30 µl/min)
Elution	150 min linear von 3 auf 35 % B, in 5 min auf 85 % B, 5 min bei 85 %, in 1 min auf 3 % B (Flussrate: 300 µl/min; Säulentemperatur 45 °C)
MS-Methode	
MS	350 – 1500 m/z, Auflösung 120.000 bei m/z 200, <i>AGC-target (automated gain control) 400.000, maximum injection time 50 ms, internal calibration</i>
MS/MS	Orbitrap: <i>most intense</i> aus MS, so viele wie innerhalb 5 s mgl., <i>charge states 2-6, isolation window 1,5 m/z</i> , Auflösung 15.000 bei m/z 200, <i>AGC target 50.000, maximum injection time 120 ms</i> , HCD (28 % NCE (<i>normalized collision energy</i>)), <i>dynamic exclusion</i> für 60 s Linear Ion Trap: <i>most intense</i> aus MS, so viele wie neben OTHCD-scans mgl., <i>charge states 2-6, isolation window 1,5 m/z, AGC target 10.000, maximum injection time 35 ms</i> , CID (35 % NCE (<i>normalized collision energy</i>)), <i>dynamic exclusion</i> für 60 s

Für das Q-Exactive™-Massenspektrometer:

Gradient	
Laden	15 min auf Vorsäule (Flussrate: 75 µl/min)
Elution	180 min linear von 3 auf 40 % B, in 2 min auf 99 % B, 5 min bei 99 %, in 2min auf 3 % B (Flussrate: 230 µl/min; Säulentemperatur 45 °C)
MS-Methode	
MS	375 – 1799 m/z, Auflösung 140.000 bei m/z 200, <i>AGC-target (automated gain control) 3.000.000, maximum injection time 100 ms</i>
MS/MS	10 <i>most intense</i> aus MS, <i>charge states 2-7, isolation window 2 m/z</i> , Auflösung 17.500 bei m/z 200, <i>AGC target 200.000 maximum injection time 150 ms</i> , stepped HCD (27, 30, 33 % NCE (<i>normalized collision energy</i>)), <i>dynamic exclusion</i> für 60 s

7.2.3 Identifizierte *cross-links*

Übersicht aller einzigartiger, in dieser Arbeit identifizierter *cross-links*.

Tabelle A 5 Cross-links CFII-Komplex.

Protein1	Protein2	Aminosäure1	Aminosäure2
Clp1	Clp1	53	54
Clp1	Clp1	127	281
Clp1	Clp1	281	290
Clp1	Clp1	290	425
Clp1	Pcf11	53	54
Clp1	Pcf11	54	1371
Clp1	Pcf11	54	1373
Clp1	Pcf11	54	1378
Clp1	Pcf11	109	614
Clp1	Pcf11	109	615
Clp1	Pcf11	109	629
Clp1	Pcf11	109	723
Clp1	Pcf11	109	638
Clp1	Pcf11	109	1405
Clp1	Pcf11	202	615
Clp1	Pcf11	202	628
Clp1	Pcf11	202	629
Clp1	Pcf11	202	633
Clp1	Pcf11	202	723
Clp1	Pcf11	267	615
Clp1	Pcf11	290	520
Clp1	Pcf11	425	1378
Pcf11	Pcf11	113	117
Pcf11	Pcf11	129	130
Pcf11	Pcf11	196	211
Pcf11	Pcf11	211	213
Pcf11	Pcf11	211	222
Pcf11	Pcf11	213	221
Pcf11	Pcf11	221	222
Pcf11	Pcf11	293	309
Pcf11	Pcf11	309	315
Pcf11	Pcf11	315	331
Pcf11	Pcf11	331	337
Pcf11	Pcf11	331	342
Pcf11	Pcf11	337	345
Pcf11	Pcf11	342	345

Pcf11	Pcf11	342	347
Pcf11	Pcf11	345	351
Pcf11	Pcf11	347	351
Pcf11	Pcf11	348	355
Pcf11	Pcf11	351	354
Pcf11	Pcf11	351	355
Pcf11	Pcf11	351	363
Pcf11	Pcf11	355	363
Pcf11	Pcf11	377	382
Pcf11	Pcf11	377	385
Pcf11	Pcf11	382	385
Pcf11	Pcf11	405	420
Pcf11	Pcf11	406	411
Pcf11	Pcf11	406	420
Pcf11	Pcf11	420	424
Pcf11	Pcf11	422	424
Pcf11	Pcf11	422	428
Pcf11	Pcf11	424	428
Pcf11	Pcf11	424	431
Pcf11	Pcf11	428	429
Pcf11	Pcf11	428	431
Pcf11	Pcf11	428	436
Pcf11	Pcf11	429	436
Pcf11	Pcf11	431	436
Pcf11	Pcf11	448	456
Pcf11	Pcf11	448	470
Pcf11	Pcf11	494	504
Pcf11	Pcf11	496	504
Pcf11	Pcf11	504	520
Pcf11	Pcf11	520	527
Pcf11	Pcf11	520	545
Pcf11	Pcf11	520	547
Pcf11	Pcf11	520	549
Pcf11	Pcf11	520	638
Pcf11	Pcf11	520	714
Pcf11	Pcf11	520	723
Pcf11	Pcf11	524	549
Pcf11	Pcf11	527	545
Pcf11	Pcf11	527	547
Pcf11	Pcf11	527	549
Pcf11	Pcf11	545	549

Pcf11	Pcf11	545	638
Pcf11	Pcf11	549	559
Pcf11	Pcf11	549	723
Pcf11	Pcf11	549	638
Pcf11	Pcf11	559	560
Pcf11	Pcf11	560	576
Pcf11	Pcf11	576	582
Pcf11	Pcf11	595	598
Pcf11	Pcf11	598	615
Pcf11	Pcf11	598	638
Pcf11	Pcf11	615	629
Pcf11	Pcf11	615	633
Pcf11	Pcf11	615	638
Pcf11	Pcf11	615	723
Pcf11	Pcf11	624	661
Pcf11	Pcf11	627	723
Pcf11	Pcf11	629	723
Pcf11	Pcf11	629	638
Pcf11	Pcf11	629	661
Pcf11	Pcf11	629	1378
Pcf11	Pcf11	633	661
Pcf11	Pcf11	633	723
Pcf11	Pcf11	633	1373
Pcf11	Pcf11	633	1378
Pcf11	Pcf11	638	1373
Pcf11	Pcf11	638	1411
Pcf11	Pcf11	638	1419
Pcf11	Pcf11	638	661
Pcf11	Pcf11	638	667
Pcf11	Pcf11	638	668
Pcf11	Pcf11	638	669
Pcf11	Pcf11	638	714
Pcf11	Pcf11	638	718
Pcf11	Pcf11	638	721
Pcf11	Pcf11	638	723
Pcf11	Pcf11	638	775
Pcf11	Pcf11	638	777
Pcf11	Pcf11	638	1278
Pcf11	Pcf11	638	1376
Pcf11	Pcf11	638	1378
Pcf11	Pcf11	638	1405

Pcf11	Pcf11	654	667
Pcf11	Pcf11	660	723
Pcf11	Pcf11	661	667
Pcf11	Pcf11	661	723
Pcf11	Pcf11	661	1373
Pcf11	Pcf11	661	1376
Pcf11	Pcf11	667	714
Pcf11	Pcf11	667	723
Pcf11	Pcf11	667	1278
Pcf11	Pcf11	668	1278
Pcf11	Pcf11	669	723
Pcf11	Pcf11	714	723
Pcf11	Pcf11	723	767
Pcf11	Pcf11	723	766
Pcf11	Pcf11	723	775
Pcf11	Pcf11	723	777
Pcf11	Pcf11	723	1278
Pcf11	Pcf11	723	1373
Pcf11	Pcf11	723	1376
Pcf11	Pcf11	723	1405
Pcf11	Pcf11	767	775
Pcf11	Pcf11	767	1278
Pcf11	Pcf11	775	851
Pcf11	Pcf11	1278	1373
Pcf11	Pcf11	1278	1376
Pcf11	Pcf11	1373	1405
Pcf11	Pcf11	1373	1419
Pcf11	Pcf11	1373	1378
Pcf11	Pcf11	1376	1405
Pcf11	Pcf11	1378	1405
Pcf11	Pcf11	1378	1411
Pcf11	Pcf11	1378	1419
Pcf11	Pcf11	1378	1420
Pcf11	Pcf11	1405	1411

Tabelle A 6 *Cross-links* mPSF-Komplex.

Protein1	Protein2	Aminosäure1	Aminosäure2
bPABP2	bPABP2	202	223
bPABP2	bPABP2	207	209
bPABP2	bPABP2	207	225
bPABP2	bPABP2	207	247
bPABP2	bPABP2	209	223
bPABP2	bPABP2	209	247
bPABP2	bPABP2	213	247
bPABP2	bPABP2	225	247
bPABP2	CPSF1	1	375
CPSF1	CPSF1	52	60
FIP1	FIP1	135	191
FIP1	FIP1	135	193
FIP1	FIP1	235	248

Tabelle A 7 *Cross-links* Endonuklease-Komplex.

Protein1	Protein2	Aminosäure1	Aminosäure2
CPSF73	CPSF73	207	410
CPSF73	CPSF73	209	410
CPSF73	CPSF73	292	295
CPSF73	CPSF73	295	309
CPSF73	CPSF73	428	435
CPSF73	CPSF73	462	487
CPSF73	CPSF73	465	487
CPSF73	CPSF73	473	496
CPSF73	CPSF73	473	501
CPSF73	CPSF73	473	510
CPSF73	CPSF73	474	496
CPSF73	CPSF73	474	510
CPSF73	CPSF73	474	487
CPSF73	CPSF73	545	595
CPSF73	CPSF73	592	595
CPSF73	CPSF73	595	604
CPSF73	CPSF73	595	611
CPSF73	CPSF73	595	612
CPSF73	CPSF73	600	604
CPSF73	CPSF73	600	610
CPSF73	CPSF73	600	611

CPSF73	CPSF73	600	612
CPSF73	CPSF73	602	604
CPSF73	CPSF73	602	612
CPSF73	CPSF73	603	604
CPSF73	CPSF73	603	611
CPSF73	CPSF73	603	612
CPSF73	CPSF100	90	138
CPSF73	CPSF100	295	725
CPSF73	CPSF100	410	611
CPSF73	CPSF100	462	622
CPSF73	CPSF100	465	592
CPSF73	CPSF100	465	1
CPSF73	CPSF100	465	584
CPSF73	CPSF100	465	587
CPSF73	CPSF100	473	584
CPSF73	CPSF100	474	1
CPSF73	CPSF100	474	584
CPSF73	CPSF100	482	622
CPSF73	CPSF100	487	622
CPSF73	CPSF100	492	557
CPSF73	CPSF100	496	584
CPSF73	CPSF100	549	551
CPSF73	CPSF100	549	557
CPSF73	CstF64	595	189
CPSF73	Symplekin	545	541
CPSF73	Symplekin	545	542
CPSF100	CPSF100	125	138
CPSF100	CPSF100	138	309
CPSF100	CPSF100	138	460
CPSF100	CPSF100	138	624
CPSF100	CPSF100	140	172
CPSF100	CPSF100	145	179
CPSF100	CPSF100	145	164
CPSF100	CPSF100	158	164
CPSF100	CPSF100	161	172
CPSF100	CPSF100	161	179
CPSF100	CPSF100	179	213
CPSF100	CPSF100	179	226
CPSF100	CPSF100	179	290
CPSF100	CPSF100	179	551
CPSF100	CPSF100	179	557

CPSF100	CPSF100	213	290
CPSF100	CPSF100	221	290
CPSF100	CPSF100	257	290
CPSF100	CPSF100	257	298
CPSF100	CPSF100	288	298
CPSF100	CPSF100	290	298
CPSF100	CPSF100	290	309
CPSF100	CPSF100	290	725
CPSF100	CPSF100	298	309
CPSF100	CPSF100	298	725
CPSF100	CPSF100	309	324
CPSF100	CPSF100	309	398
CPSF100	CPSF100	309	725
CPSF100	CPSF100	322	459
CPSF100	CPSF100	324	450
CPSF100	CPSF100	324	455
CPSF100	CPSF100	386	398
CPSF100	CPSF100	386	450
CPSF100	CPSF100	386	460
CPSF100	CPSF100	398	403
CPSF100	CPSF100	398	408
CPSF100	CPSF100	398	450
CPSF100	CPSF100	398	460
CPSF100	CPSF100	402	444
CPSF100	CPSF100	403	437
CPSF100	CPSF100	403	438
CPSF100	CPSF100	403	444
CPSF100	CPSF100	403	448
CPSF100	CPSF100	403	455
CPSF100	CPSF100	403	459
CPSF100	CPSF100	403	450
CPSF100	CPSF100	407	461
CPSF100	CPSF100	437	459
CPSF100	CPSF100	437	450
CPSF100	CPSF100	438	459
CPSF100	CPSF100	438	450
CPSF100	CPSF100	444	450
CPSF100	CPSF100	444	455
CPSF100	CPSF100	448	450
CPSF100	CPSF100	448	455
CPSF100	CPSF100	450	459

CPSF100	CPSF100	455	459
CPSF100	CPSF100	455	460
CPSF100	CPSF100	455	461
CPSF100	CPSF100	551	624
CPSF100	CPSF100	551	725
CPSF100	CPSF100	557	624
CPSF100	CPSF100	557	584
CPSF100	CPSF100	584	592
CPSF100	CPSF100	584	624
CPSF100	CPSF100	611	624
CPSF100	CPSF100	617	624
CPSF100	CstF64	725	189
CPSF100	CstF64	725	55
CPSF100	Symplekin	551	520
CPSF100	Symplekin	725	490
CPSF100	Symplekin	725	520
CPSF100	Symplekin	725	345
CPSF100	Symplekin	725	541
CPSF100	Symplekin	725	542
CstF64	CstF64	55	189
CstF64	CstF64	55	72
CstF64	CstF64	72	96
CstF64	CstF64	94	98
CstF64	CstF64	154	189
CstF64	Symplekin	189	520
CstF64	Symplekin	189	345
CstF64	Symplekin	189	358
CstF64	Symplekin	189	477
CstF64	Symplekin	189	490
Symplekin	Symplekin	90	127
Symplekin	Symplekin	128	147
Symplekin	Symplekin	132	146
Symplekin	Symplekin	146	358
Symplekin	Symplekin	147	358
Symplekin	Symplekin	296	302
Symplekin	Symplekin	343	361
Symplekin	Symplekin	358	361
Symplekin	Symplekin	477	483
Symplekin	Symplekin	477	490
Symplekin	Symplekin	480	490
Symplekin	Symplekin	483	520

Symplekin	Symplekin	490	520
Symplekin	Symplekin	490	551
Symplekin	Symplekin	490	554
Symplekin	Symplekin	490	542
Symplekin	Symplekin	490	567
Symplekin	Symplekin	567	574
Symplekin	Symplekin	567	665
Symplekin	Symplekin	660	727
Symplekin	Symplekin	660	684
Symplekin	Symplekin	665	684
Symplekin	Symplekin	684	725
Symplekin	Symplekin	684	727
Symplekin	Symplekin	705	742
Symplekin	Symplekin	708	742
Symplekin	Symplekin	727	737
Symplekin	Symplekin	730	742
Symplekin	Symplekin	737	742
Symplekin	Symplekin	737	748
Symplekin	Symplekin	737	866
Symplekin	Symplekin	820	857
Symplekin	Symplekin	850	883
Symplekin	Symplekin	850	909
Symplekin	Symplekin	850	912
Symplekin	Symplekin	854	883
Symplekin	Symplekin	854	909
Symplekin	Symplekin	854	912
Symplekin	Symplekin	857	883
Symplekin	Symplekin	857	909
Symplekin	Symplekin	857	912
Symplekin	Symplekin	857	1150
Symplekin	Symplekin	862	875
Symplekin	Symplekin	875	883
Symplekin	Symplekin	875	901
Symplekin	Symplekin	883	901
Symplekin	Symplekin	883	909
Symplekin	Symplekin	900	909
Symplekin	Symplekin	901	1150
Symplekin	Symplekin	901	956
Symplekin	Symplekin	909	956
Symplekin	Symplekin	912	956
Symplekin	Symplekin	912	964

Symplekin	Symplekin	912	960
Symplekin	Symplekin	919	964
Symplekin	Symplekin	919	966
Symplekin	Symplekin	919	960
Symplekin	Symplekin	919	1147
Symplekin	Symplekin	960	1033
Symplekin	Symplekin	960	1040
Symplekin	Symplekin	961	1040
Symplekin	Symplekin	964	1033
Symplekin	Symplekin	964	1147
Symplekin	Symplekin	964	1150
Symplekin	Symplekin	964	1239
Symplekin	Symplekin	966	1033
Symplekin	Symplekin	966	1147
Symplekin	Symplekin	966	1150
Symplekin	Symplekin	966	1239
Symplekin	Symplekin	1103	1150
Symplekin	Symplekin	1147	1239
Symplekin	Symplekin	1150	1239
Symplekin	Symplekin	1150	1274
Symplekin	Symplekin	1239	1256
Symplekin	Symplekin	1239	1257
Symplekin	Symplekin	1239	1259
Symplekin	Symplekin	1239	1274

Tabelle A 8 *Cross-links* CstF-Komplex.

Protein1	Protein2	Aminosäure1	Aminosäure2
CstF50	CstF50	4	212
CstF50	CstF50	4	208
CstF50	CstF50	5	212
CstF50	CstF50	5	208
CstF50	CstF50	5	204
CstF50	CstF50	9	208
CstF50	CstF50	9	204
CstF50	CstF50	9	143
CstF50	CstF50	133	206
CstF50	CstF50	137	208
CstF50	CstF50	204	212
CstF50	CstF50	206	331
CstF50	CstF50	206	302

CstF50	CstF50	206	212
CstF50	CstF50	208	315
CstF50	CstF50	208	302
CstF50	CstF50	208	298
CstF50	CstF50	208	212
CstF50	CstF50	298	326
CstF50	CstF50	298	331
CstF50	CstF50	299	326
CstF50	CstF50	311	326
CstF50	CstF50	311	331
CstF50	CstF50	315	326
CstF50	CstF50	315	331
CstF50	CstF50	320	331
CstF50	CstF50	326	383
CstF50	CstF64	203	145
CstF50	CstF64	204	189
CstF50	CstF64	204	364
CstF50	CstF64	204	145
CstF50	CstF64	204	69
CstF50	CstF64	204	55
CstF50	CstF64	204	98
CstF50	CstF64	204	53
CstF50	CstF64	204	49
CstF50	CstF64	206	189
CstF50	CstF64	206	148
CstF50	CstF64	206	145
CstF50	CstF64	206	57
CstF50	CstF64	206	55
CstF50	CstF64	206	98
CstF50	CstF64	208	189
CstF50	CstF64	208	154
CstF50	CstF64	208	98
CstF50	CstF64	208	96
CstF50	CstF64	208	69
CstF50	CstF64	208	44
CstF50	CstF64	208	53
CstF50	CstF64	212	189
CstF50	CstF64	212	145
CstF50	CstF64	212	98
CstF50	CstF64	212	55
CstF50	CstF64	298	55

CstF50	CstF64	302	154
CstF50	CstF64	315	98
CstF50	CstF64	326	17
CstF50	CstF64	328	189
CstF50	CstF64	383	189
CstF50	CstF77	4	221
CstF50	CstF77	4	49
CstF50	CstF77	5	550
CstF50	CstF77	5	546
CstF50	CstF77	5	311
CstF50	CstF77	5	192
CstF50	CstF77	5	77
CstF50	CstF77	5	49
CstF50	CstF77	5	24
CstF50	CstF77	9	547
CstF50	CstF77	9	546
CstF50	CstF77	9	311
CstF50	CstF77	9	228
CstF50	CstF77	9	77
CstF50	CstF77	143	24
CstF50	CstF77	144	221
CstF50	CstF77	144	228
CstF50	CstF77	144	647
CstF50	CstF77	192	86
CstF50	CstF77	203	221
CstF50	CstF77	204	647
CstF50	CstF77	204	503
CstF50	CstF77	204	228
CstF50	CstF77	204	225
CstF50	CstF77	204	223
CstF50	CstF77	204	221
CstF50	CstF77	204	203
CstF50	CstF77	204	192
CstF50	CstF77	204	682
CstF50	CstF77	204	311
CstF50	CstF77	204	77
CstF50	CstF77	204	49
CstF50	CstF77	204	24
CstF50	CstF77	204	52
CstF50	CstF77	204	254
CstF50	CstF77	204	392

CstF50	CstF77	206	647
CstF50	CstF77	206	311
CstF50	CstF77	206	228
CstF50	CstF77	206	221
CstF50	CstF77	206	203
CstF50	CstF77	206	192
CstF50	CstF77	206	682
CstF50	CstF77	206	77
CstF50	CstF77	206	52
CstF50	CstF77	206	204
CstF50	CstF77	206	254
CstF50	CstF77	208	647
CstF50	CstF77	208	228
CstF50	CstF77	208	203
CstF50	CstF77	208	682
CstF50	CstF77	208	681
CstF50	CstF77	208	221
CstF50	CstF77	208	192
CstF50	CstF77	208	24
CstF50	CstF77	208	49
CstF50	CstF77	208	52
CstF50	CstF77	208	77
CstF50	CstF77	208	79
CstF50	CstF77	208	254
CstF50	CstF77	208	311
CstF50	CstF77	212	228
CstF50	CstF77	212	203
CstF50	CstF77	212	192
CstF50	CstF77	212	49
CstF50	CstF77	212	52
CstF50	CstF77	212	204
CstF50	CstF77	212	221
CstF50	CstF77	212	254
CstF50	CstF77	212	393
CstF50	CstF77	212	647
CstF50	CstF77	213	647
CstF50	CstF77	240	86
CstF50	CstF77	298	682
CstF50	CstF77	302	682
CstF50	CstF77	302	681
CstF64	CstF64	33	364

CstF64	CstF64	44	98
CstF64	CstF64	53	189
CstF64	CstF64	53	145
CstF64	CstF64	53	55
CstF64	CstF64	53	96
CstF64	CstF64	55	189
CstF64	CstF64	55	145
CstF64	CstF64	55	98
CstF64	CstF64	55	96
CstF64	CstF64	55	72
CstF64	CstF64	55	55
CstF64	CstF64	57	96
CstF64	CstF64	57	98
CstF64	CstF64	59	96
CstF64	CstF64	59	98
CstF64	CstF64	64	96
CstF64	CstF64	72	96
CstF64	CstF64	72	98
CstF64	CstF64	94	98
CstF64	CstF64	96	106
CstF64	CstF64	96	102
CstF64	CstF64	96	96
CstF64	CstF64	98	106
CstF64	CstF64	98	103
CstF64	CstF64	145	189
CstF64	CstF64	154	189
CstF64	CstF64	189	364
CstF64	CstF64	189	189
CstF64	CstF64	364	567
CstF64	CstF77	44	204
CstF64	CstF77	44	399
CstF64	CstF77	49	204
CstF64	CstF77	53	228
CstF64	CstF77	145	682
CstF64	CstF77	145	647
CstF64	CstF77	154	647
CstF64	CstF77	189	682
CstF64	CstF77	189	647
CstF64	CstF77	189	228
CstF64	CstF77	189	203
CstF64	CstF77	189	24

CstF64	CstF77	189	49
CstF64	CstF77	189	204
CstF64	CstF77	189	221
CstF64	CstF77	189	254
CstF64	CstF77	189	311
CstF64	CstF77	189	337
CstF64	CstF77	189	392
CstF64	CstF77	189	399
CstF64	CstF77	189	714
CstF64	CstF77	562	503
CstF64	CstF77	567	204
CstF77	CstF77	19	24
CstF77	CstF77	20	30
CstF77	CstF77	23	30
CstF77	CstF77	23	52
CstF77	CstF77	23	338
CstF77	CstF77	24	550
CstF77	CstF77	24	547
CstF77	CstF77	24	36
CstF77	CstF77	24	52
CstF77	CstF77	30	438
CstF77	CstF77	36	438
CstF77	CstF77	49	647
CstF77	CstF77	49	86
CstF77	CstF77	49	83
CstF77	CstF77	49	52
CstF77	CstF77	49	77
CstF77	CstF77	52	647
CstF77	CstF77	52	337
CstF77	CstF77	52	86
CstF77	CstF77	52	83
CstF77	CstF77	52	64
CstF77	CstF77	52	63
CstF77	CstF77	54	86
CstF77	CstF77	69	102
CstF77	CstF77	77	503
CstF77	CstF77	77	500
CstF77	CstF77	77	383
CstF77	CstF77	86	126
CstF77	CstF77	86	94
CstF77	CstF77	86	337

CstF77	CstF77	94	363
CstF77	CstF77	94	358
CstF77	CstF77	94	274
CstF77	CstF77	94	337
CstF77	CstF77	102	228
CstF77	CstF77	104	337
CstF77	CstF77	107	228
CstF77	CstF77	126	274
CstF77	CstF77	132	274
CstF77	CstF77	146	274
CstF77	CstF77	192	682
CstF77	CstF77	192	203
CstF77	CstF77	192	204
CstF77	CstF77	204	647
CstF77	CstF77	204	311
CstF77	CstF77	221	254
CstF77	CstF77	223	254
CstF77	CstF77	225	254
CstF77	CstF77	228	682
CstF77	CstF77	228	647
CstF77	CstF77	228	254
CstF77	CstF77	254	572
CstF77	CstF77	254	550
CstF77	CstF77	254	547
CstF77	CstF77	254	521
CstF77	CstF77	254	519
CstF77	CstF77	254	513
CstF77	CstF77	254	510
CstF77	CstF77	255	519
CstF77	CstF77	260	521
CstF77	CstF77	274	328
CstF77	CstF77	274	318
CstF77	CstF77	274	311
CstF77	CstF77	311	682
CstF77	CstF77	311	647
CstF77	CstF77	311	516
CstF77	CstF77	311	513
CstF77	CstF77	311	506
CstF77	CstF77	311	328
CstF77	CstF77	311	355
CstF77	CstF77	311	358

CstF77	CstF77	318	355
CstF77	CstF77	318	358
CstF77	CstF77	321	355
CstF77	CstF77	328	355
CstF77	CstF77	328	358
CstF77	CstF77	337	392
CstF77	CstF77	337	363
CstF77	CstF77	337	358
CstF77	CstF77	338	455
CstF77	CstF77	352	547
CstF77	CstF77	352	438
CstF77	CstF77	363	392
CstF77	CstF77	381	392
CstF77	CstF77	383	392
CstF77	CstF77	383	393
CstF77	CstF77	383	399
CstF77	CstF77	383	426
CstF77	CstF77	392	438
CstF77	CstF77	392	431
CstF77	CstF77	392	426
CstF77	CstF77	392	399
CstF77	CstF77	392	400
CstF77	CstF77	393	438
CstF77	CstF77	393	431
CstF77	CstF77	393	399
CstF77	CstF77	393	400
CstF77	CstF77	399	440
CstF77	CstF77	399	438
CstF77	CstF77	399	431
CstF77	CstF77	399	426
CstF77	CstF77	400	440
CstF77	CstF77	400	438
CstF77	CstF77	400	431
CstF77	CstF77	426	438
CstF77	CstF77	427	438
CstF77	CstF77	431	440
CstF77	CstF77	440	480
CstF77	CstF77	463	506
CstF77	CstF77	480	553
CstF77	CstF77	480	519
CstF77	CstF77	480	503

CstF77	CstF77	480	547
CstF77	CstF77	480	550
CstF77	CstF77	481	553
CstF77	CstF77	503	521
CstF77	CstF77	503	519
CstF77	CstF77	503	513
CstF77	CstF77	503	516
CstF77	CstF77	503	539
CstF77	CstF77	506	521
CstF77	CstF77	506	516
CstF77	CstF77	506	513
CstF77	CstF77	510	519
CstF77	CstF77	513	550
CstF77	CstF77	513	521
CstF77	CstF77	513	528
CstF77	CstF77	519	550
CstF77	CstF77	519	542
CstF77	CstF77	519	546
CstF77	CstF77	521	542
CstF77	CstF77	521	547
CstF77	CstF77	528	547
CstF77	CstF77	529	547
CstF77	CstF77	546	553
CstF77	CstF77	550	553
CstF77	CstF77	647	682
CstF77	CstF77	647	714

Tabelle A 9 *Cross-links* CFI-Komplex.

Protein1	Protein2	Aminosäure1	Aminosäure2	Cross-linker
CFI25	CFI25	56	73	DSBU
CFI25	CFI25	59	73	DSBU
CFI25	CFI25	73	73	DSBU
CFI25	CFI68	29	494	DSBU
CFI25	CFI68	29	525	DSBU
CFI25	CFI68	29	527	LC-SDA
CFI25	CFI68	73	505	DSBU
CFI68	CFI68	124	415	LC-SDA

7.2.4 Entfernungsbeschränkungen

Übersicht der für das *Docking* verwendeten Entfernungsbeschränkungen für den CFII-, CstF- und Endonuklease-Komplex.

Tabelle A 10 *Docking*-Beschränkungen des CFII-Komplexes.

Protein 1	Aminosäure 1	Protein 2	Aminosäure 2	Min [in Å]	Max [in Å]
Clp1	54	Pcf11 ₁₃₄₁₋₁₃₆₈	1371	3	31
Clp1	109	Pcf11 ₅₇₉₋₆₃₁	614	3	27
Clp1	109	Pcf11 ₅₇₉₋₆₃₁	615	3	31
Clp1	109	Pcf11 ₅₇₉₋₆₃₁	629	3	31
Clp1	202	Pcf11 ₅₇₉₋₆₃₁	615	3	31
Clp1	202	Pcf11 ₅₇₉₋₆₃₁	628	3	27
Clp1	202	Pcf11 ₅₇₉₋₆₃₁	629	3	31
Clp1	267	Pcf11 ₅₇₉₋₆₃₁	615	3	31
Clp1	54	Pcf11 ₁₃₄₁₋₁₃₆₈	1371	6	53,4
Clp1	54	Pcf11 ₁₃₄₁₋₁₃₆₈	1371	6	36
Clp1	202	Pcf11 ₅₇₉₋₆₃₁	631	6	28
Clp1	109	Pcf11 ₅₇₉₋₆₃₁	631	6	28
Clp1	109	Pcf11 ₅₇₉₋₆₃₁	631	6	53,4
Pcf11 ₅₇₉₋₆₃₁	624	Pcf11 ₆₅₅₋₆₉₅	661	3	27
Pcf11 ₅₇₉₋₆₃₁	629	Pcf11 ₆₅₅₋₆₉₅	661	3	31
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₆₅₅₋₆₉₅	661	6	28
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₁₃₄₁₋₁₃₆₈	1371	6	32
Pcf11 ₅₇₉₋₆₃₁	629	Pcf11 ₁₃₇₇₋₁₄₀₁	1378	3	31
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₁₃₇₇₋₁₄₀₁	1378	6	28
Pcf11 ₆₄₆₋₆₅₃	653	Pcf11 ₆₅₅₋₆₉₅	667	3	31
Pcf11 ₆₅₅₋₆₉₅	667	Pcf11 ₁₂₆₅₋₁₂₈₁	1278	3	31
Pcf11 ₆₅₅₋₆₉₅	668	Pcf11 ₁₂₆₅₋₁₂₈₁	1278	3	27
Pcf11 ₆₅₅₋₆₉₅	661	Pcf11 ₁₃₄₁₋₁₃₆₈	1371	6	36
Pcf11 ₆₅₅₋₆₉₅	661	Pcf11 ₁₃₇₇₋₁₄₀₁	1377	6	28
Pcf11 ₁₃₄₁₋₁₃₆₈	1371	Pcf11 ₁₃₇₇₋₁₄₀₁	1401	9	49
Pcf11 ₁₃₄₁₋₁₃₆₈	1371	Pcf11 ₁₄₁₉₋₁₄₈₇	1419	6	36
Pcf11 ₁₃₄₁₋₁₃₆₈	1371	Pcf11 ₁₃₇₇₋₁₄₀₁	1401	9	55,8
Pcf11 ₁₃₄₁₋₁₃₆₈	1371	Pcf11 ₁₃₇₇₋₁₄₀₁	1378	6	36
Pcf11 ₁₂₆₅₋₁₂₈₁	1278	Pcf11 ₁₃₄₁₋₁₃₆₈	1371	6	36
Pcf11 ₁₂₆₅₋₁₂₈₁	1278	Pcf11 ₁₃₇₇₋₁₄₀₁	1377	6	28
Pcf11 ₁₃₇₇₋₁₄₀₁	1378	Pcf11 ₁₄₁₉₋₁₄₈₇	1419	6	57
Pcf11 ₁₃₇₇₋₁₄₀₁	1378	Pcf11 ₁₄₁₉₋₁₄₈₇	1419	3	31
Pcf11 ₁₃₇₇₋₁₄₀₁	1378	Pcf11 ₁₄₁₉₋₁₄₈₇	1420	3	27
Pcf11 ₁₃₇₇₋₁₄₀₁	1401	Pcf11 ₁₄₁₉₋₁₄₈₇	1419	6	57

Pcf11 ₅₇₉₋₆₃₁	598	Pcf11 ₆₄₆₋₆₅₃	646	6	53,4
Pcf11 ₅₇₉₋₆₃₁	615	Pcf11 ₆₄₆₋₆₅₃	646	6	53,4
Pcf11 ₅₇₉₋₆₃₁	629	Pcf11 ₆₄₆₋₆₅₃	646	6	53,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₆₅₅₋₆₉₅	661	6	53,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₆₅₅₋₆₉₅	667	6	53,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₆₅₅₋₆₉₅	668	6	49,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₆₅₅₋₆₉₅	669	6	49,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₁₃₄₁₋₁₃₆₈	1371	9	58,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₁₄₁₉₋₁₄₈₇	1419	9	79,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₁₄₁₉₋₁₄₈₇	1419	6	53,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₁₂₆₅₋₁₂₈₁	1278	6	53,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₁₃₇₇₋₁₄₀₁	1377	9	50,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₁₃₇₇₋₁₄₀₁	1378	6	53,4
Pcf11 ₅₇₉₋₆₃₁	631	Pcf11 ₁₃₇₇₋₁₄₀₁	1401	6	53,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₆₅₅₋₆₉₅	661	6	53,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₆₅₅₋₆₉₅	667	6	53,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₆₅₅₋₆₉₅	668	6	49,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₆₅₅₋₆₉₅	669	6	49,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₁₃₄₁₋₁₃₆₈	1371	9	58,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₁₄₁₉₋₁₄₈₇	1419	9	79,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₁₄₁₉₋₁₄₈₇	1419	6	53,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₁₂₆₅₋₁₂₈₁	1278	6	53,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₁₃₇₇₋₁₄₀₁	1377	9	50,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₁₃₇₇₋₁₄₀₁	1378	6	53,4
Pcf11 ₆₄₆₋₆₅₃	646	Pcf11 ₁₃₇₇₋₁₄₀₁	1401	6	53,4
Pcf11 ₁₃₄₁₋₁₃₆₈	1371	Pcf11 ₁₃₇₇₋₁₄₀₁	1377	3	19

Tabelle A 11 Docking-Beschränkungen des CstF-Komplexes.

Protein 1	Aminosäure 1	Protein 2	Aminosäure 2	Min [in Å]	Max [in Å]
CstF64 ₈₋₁₈₉	145	CstF50 ₉₀₋₄₂₇	203	3	34
CstF64 ₈₋₁₈₉	145	CstF50 ₉₀₋₄₂₇	204	3	34
CstF64 ₈₋₁₈₉	145	CstF50 ₉₀₋₄₂₇	206	3	34
CstF64 ₈₋₁₈₉	145	CstF50 ₉₀₋₄₂₇	212	3	34
CstF64 ₈₋₁₈₉	148	CstF50 ₉₀₋₄₂₇	206	3	34
CstF64 ₈₋₁₈₉	154	CstF50 ₉₀₋₄₂₇	208	3	34
CstF64 ₈₋₁₈₉	154	CstF50 ₉₀₋₄₂₇	302	3	34
CstF64 ₈₋₁₈₉	117	CstF50 ₉₀₋₄₂₇	326	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	204	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	206	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	208	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	212	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	328	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	383	3	34
CstF64 ₈₋₁₈₉	144	CstF50 ₉₀₋₄₂₇	208	3	34
CstF64 ₈₋₁₈₉	149	CstF50 ₉₀₋₄₂₇	204	3	34
CstF64 ₈₋₁₈₉	153	CstF50 ₉₀₋₄₂₇	204	3	34
CstF64 ₈₋₁₈₉	153	CstF50 ₉₀₋₄₂₇	208	3	34
CstF64 ₈₋₁₈₉	155	CstF50 ₉₀₋₄₂₇	204	3	34
CstF64 ₈₋₁₈₉	155	CstF50 ₉₀₋₄₂₇	206	3	34
CstF64 ₈₋₁₈₉	155	CstF50 ₉₀₋₄₂₇	212	3	34
CstF64 ₈₋₁₈₉	155	CstF50 ₉₀₋₄₂₇	298	3	34
CstF64 ₈₋₁₈₉	157	CstF50 ₉₀₋₄₂₇	206	3	34
CstF64 ₈₋₁₈₉	169	CstF50 ₉₀₋₄₂₇	204	3	34
CstF64 ₈₋₁₈₉	169	CstF50 ₉₀₋₄₂₇	208	3	34
CstF64 ₈₋₁₈₉	196	CstF50 ₉₀₋₄₂₇	208	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	204	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	206	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	208	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	212	3	34
CstF64 ₈₋₁₈₉	189	CstF50 ₉₀₋₄₂₇	315	3	34
CstF50 ₉₀₋₄₂₇	204	CstF77 ₆₁₁₋₆₄₉	647	3	34
CstF50 ₉₀₋₄₂₇	206	CstF77 ₆₁₁₋₆₄₉	647	3	34
CstF50 ₉₀₋₄₂₇	208	CstF77 ₆₁₁₋₆₄₉	647	3	34
CstF50 ₉₀₋₄₂₇	212	CstF77 ₆₁₁₋₆₄₉	647	3	34
CstF50 ₉₀₋₄₂₇	213	CstF77 ₆₁₁₋₆₄₉	647	3	34

CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	24	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	24	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	49	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	49	3	31
CstF50 ₉₀₋₄₂₇	212	CstF77 ₂₁₋₅₄₉	49	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	52	3	31
CstF50 ₉₀₋₄₂₇	206	CstF77 ₂₁₋₅₄₉	52	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	52	3	31
CstF50 ₉₀₋₄₂₇	212	CstF77 ₂₁₋₅₄₉	52	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	77	3	31
CstF50 ₉₀₋₄₂₇	206	CstF77 ₂₁₋₅₄₉	77	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	77	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	79	3	31
CstF50 ₉₀₋₄₂₇	192	CstF77 ₂₁₋₅₄₉	86	3	31
CstF50 ₉₀₋₄₂₇	240	CstF77 ₂₁₋₅₄₉	86	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	192	3	31
CstF50 ₉₀₋₄₂₇	206	CstF77 ₂₁₋₅₄₉	192	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	192	3	31
CstF50 ₉₀₋₄₂₇	212	CstF77 ₂₁₋₅₄₉	192	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	203	3	31
CstF50 ₉₀₋₄₂₇	206	CstF77 ₂₁₋₅₄₉	203	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	203	3	31
CstF50 ₉₀₋₄₂₇	212	CstF77 ₂₁₋₅₄₉	203	3	31
CstF50 ₉₀₋₄₂₇	206	CstF77 ₂₁₋₅₄₉	204	3	31
CstF50 ₉₀₋₄₂₇	212	CstF77 ₂₁₋₅₄₉	204	3	31
CstF50 ₉₀₋₄₂₇	203	CstF77 ₂₁₋₅₄₉	221	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	221	3	31
CstF50 ₉₀₋₄₂₇	206	CstF77 ₂₁₋₅₄₉	221	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	221	3	31
CstF50 ₉₀₋₄₂₇	212	CstF77 ₂₁₋₅₄₉	221	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	223	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	225	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	228	3	31
CstF50 ₉₀₋₄₂₇	206	CstF77 ₂₁₋₅₄₉	228	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	228	3	31
CstF50 ₉₀₋₄₂₇	212	CstF77 ₂₁₋₅₄₉	228	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	254	3	31
CstF50 ₉₀₋₄₂₇	206	CstF77 ₂₁₋₅₄₉	254	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	254	3	31

CstF50 ₉₀₋₄₂₇	212	CstF77 ₂₁₋₅₄₉	254	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	311	3	31
CstF50 ₉₀₋₄₂₇	206	CstF77 ₂₁₋₅₄₉	311	3	31
CstF50 ₉₀₋₄₂₇	208	CstF77 ₂₁₋₅₄₉	311	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	392	3	31
CstF50 ₉₀₋₄₂₇	212	CstF77 ₂₁₋₅₄₉	393	3	31
CstF50 ₉₀₋₄₂₇	204	CstF77 ₂₁₋₅₄₉	503	3	31
CstF50 ₉₀₋₄₂₇	212	CstF50 ₂₋₆₁	4	3	31
CstF50 ₉₀₋₄₂₇	208	CstF50 ₂₋₆₁	4	3	31
CstF50 ₉₀₋₄₂₇	212	CstF50 ₂₋₆₁	5	3	31
CstF50 ₉₀₋₄₂₇	208	CstF50 ₂₋₆₁	5	3	31
CstF50 ₉₀₋₄₂₇	204	CstF50 ₂₋₆₁	5	3	31
CstF50 ₉₀₋₄₂₇	208	CstF50 ₂₋₆₁	9	3	31
CstF50 ₉₀₋₄₂₇	204	CstF50 ₂₋₆₁	9	3	31
CstF50 ₉₀₋₄₂₇	143	CstF50 ₂₋₆₁	9	3	31
CstF77 ₂₁₋₅₄₉	221	CstF50 ₂₋₆₁	4	3	31
CstF77 ₂₁₋₅₄₉	49	CstF50 ₂₋₆₁	4	3	31
CstF77 ₂₁₋₅₄₉	550	CstF50 ₂₋₆₁	5	3	31
CstF77 ₂₁₋₅₄₉	546	CstF50 ₂₋₆₁	5	3	31
CstF77 ₂₁₋₅₄₉	311	CstF50 ₂₋₆₁	5	3	31
CstF77 ₂₁₋₅₄₉	192	CstF50 ₂₋₆₁	5	3	31
CstF77 ₂₁₋₅₄₉	77	CstF50 ₂₋₆₁	5	3	31
CstF77 ₂₁₋₅₄₉	49	CstF50 ₂₋₆₁	5	3	31
CstF77 ₂₁₋₅₄₉	24	CstF50 ₂₋₆₁	5	3	31
CstF77 ₂₁₋₅₄₉	547	CstF50 ₂₋₆₁	9	3	31
CstF77 ₂₁₋₅₄₉	546	CstF50 ₂₋₆₁	9	3	31
CstF77 ₂₁₋₅₄₉	311	CstF50 ₂₋₆₁	9	3	31
CstF77 ₂₁₋₅₄₉	228	CstF50 ₂₋₆₁	9	3	31
CstF77 ₂₁₋₅₄₉	77	CstF50 ₂₋₆₁	9	3	31

Tabelle A 12 Docking-Beschränkungen des Endonuklease-Komplexes.

Protein 1	Aminosäure 1	Protein 2	Aminosäure 2	Min [in Å]	Max [in Å]
CPSF100 ₂₂₋₅₉₆	551	CPSF100 ₇₂₁₋₇₈₂	725	3	31
CPSF100 ₂₂₋₅₉₆	290	CPSF100 ₇₂₁₋₇₈₂	725	3	31
CPSF100 ₂₂₋₅₉₆	309	CPSF100 ₇₂₁₋₇₈₂	725	3	31
CPSF100 ₂₂₋₅₉₆	551	CPSF73 ₄₉₅₋₅₈₃	549	3	31
CPSF100 ₂₂₋₅₉₆	557	CPSF73 ₄₉₅₋₅₈₃	492	3	31
CPSF100 ₂₂₋₅₉₆	557	CPSF73 ₄₉₅₋₅₈₃	549	3	31
CPSF100 ₂₂₋₅₉₆	584	CPSF73 ₄₉₅₋₅₈₃	496	3	31
CPSF100 ₇₂₁₋₇₈₂	725	CstF64 ₁₋₁₈₉	189	3	31
CPSF100 ₇₂₁₋₇₈₂	725	CstF64 ₁₋₁₈₉	55	3	31
Symplekin ₄₄₃₋₄₈₀	480	CPSF100 ₇₂₁₋₇₈₂	725	6	59,9
CPSF100 ₂₂₋₅₉₆	584	CPSF73 ₁₋₄₅₉	459	6	49
CPSF100 ₂₂₋₅₉₆	587	CPSF73 ₁₋₄₅₉	459	6	49
CPSF100 ₂₂₋₅₉₆	596	CPSF73 ₁₋₄₅₉	410	6	85
CPSF100 ₇₂₁₋₇₈₂	725	CPSF73 ₁₋₄₅₉	305	6	69,8
CPSF100 ₂₂₋₅₉₆	592	CPSF73 ₁₋₄₅₉	459	6	49
CPSF100 ₂₂₋₅₉₆	138	CPSF73 ₁₋₄₅₉	90	3	31
CPSF73 ₄₉₅₋₅₈₃	495	CPSF73 ₁₋₄₅₉	459	5	25
CstF64 ₁₋₁₈₉	189	Symplekin ₃₀₋₃₄₅	345	3	31
CPSF100 ₇₂₁₋₇₈₂	725	Symplekin ₃₀₋₃₄₅	345	3	31
Symplekin ₄₄₃₋₄₈₀	443	Symplekin ₃₀₋₃₄₅	345	15	45

7.2.5 Sonstige Tabellen

Tabelle A.18 Statistische Übersicht des Boxplots der Analyse der gemessenen Entfernungen.

Peptid	Min.	1st Qu	Median	Durchschnitt	3rd Qu	Max.	n
di	2,04	3,79	3,81	3,81	3,82	6	9867470
tri	3,06	5,46	5,76	6	6,55	9,9	9697240
tetra	2,7	5,3	7,7	7,4	9,4	12,6	9366765
penta	2,9	6,3	8,8	9,1	11,8	15,9	8882130
hexa	3	8,6	9,9	11	13,9	18,8	8248834
hepta	3,1	9,9	11,1	12,5	15,5	22	7472178
octa	3,3	10,5	12,3	13,7	16,8	25,4	6556509

nona	2,9	12	13,4	14,9	18	29	5505454
deca	3,1	12,9	14,6	16	19,1	31,9	4322608

7.2 Quellcode

```

1  #!/bin/bash
2  pdbid=6QUL
3
4  #generate output file
5
6  touch #RESN.csv
7
8  echo "#RESN" > #RESN.csv
9
10 # load pdb file from folder pdb/ and
11 # readout ca entries and save to array
12 # array organisation: ["ca", "#RESN", "x", "y", "z", "chain"]
13
14 cat $pdbid.pdb | grep "ATOM" | grep "#RESN" | grep "#RESI " | awk '{ print
15 substr($0,14,2) "\t" substr($0,23,4) "\t" substr($0,31,8) "\t" substr($0,39,8) "\t" substr($0,47,8
16 ) "\t" substr($0,22,1)}' > #RESN.tt
17
18 readarray ca < #RESN.tt
19
20 length=${#ca[@]}
21
22 for (( j=0;j<$length;j=j+2 ))
23
24 do
25 # starting coordinate:
26
27 x1=$( echo ${ca[$j]} | awk '{print $3}' )
28
29 y1=$( echo ${ca[$j]} | awk '{print $4}' )
30
31 z1=$( echo ${ca[$j]} | awk '{print $5}' )
32
33 # ending coordinate:
34
35 x2=$( echo ${ca[$j+1]} | awk '{print $3}' )
36
37 y2=$( echo ${ca[$j+1]} | awk '{print $4}' )
38
39 z2=$( echo ${ca[$j+1]} | awk '{print $5}' )
40
41 # calculate ca-ca distance
42
43 # formula: d = sqrt( (x1-x2)^2 + (y1-y2)^2 + (z1-z2)^2 )
44
45 dist=$(bc <<< "scale=4;sqrt( ($x1 - $x2)^2 + ($y1 - $y2)^2 + ($z1 - $z2)^2 )")
46
47 # generate output
48
49 echo $dist >> #RESN.csv
50
51 done

```

Quellcode A.1 Shellscript zur Bestimmung der Aminosäureseitenketten. #RESN ist Platzhalter für den 3-Buchstabencode der zu analysierenden Aminosäure (LYS, SER, THR, TYR). #RESI beschreibt den Typ des terminalen Atoms der jeweiligen Aminosäure: CA|NZ für Lysine, CA|OG für Serine und Threonine und CA|OH für Tyrosine.

```

1  #!/bin/bash
2  pdbid=$1
3  #generate output files

```

```
4 touch pdb/$pdbid.csv
5 # load pdb file from folder pdb/ and
6 # readout ca entries and save to array
7 # array organisation: ["ca","resn","x","y","z","chain"]
8 cat pdb/$pdbid.pdb | grep "ATOM" | grep "CA " | awk '{ print
9 substr($0,14,2)"\t"substr($0,23,4)"\t"substr($0,31,8)"\t"substr($0,39,8)"\t"substr($0,47,8)
10 "\t"substr($0,22,1)}' > $pdbid.tt2
11 mv $pdbid.tt2 $pdbid.tt
12 readarray ca < $pdbid.tt
13 ca_length=${#ca[@]}
14 for (( j=0;j<$ca_length;j++ ))
15 do
16 # starting coordinate:
17 x1=$(awk '{print $3}' <<< ${ca[$j]})
18 y1=$(awk '{print $4}' <<< ${ca[$j]})
19 z1=$(awk '{print $5}' <<< ${ca[$j]})
20 # loop till maxlength, to get all distances from starting point
21 maxlength=9
22 output=""
23 for (( n=1;n<=$maxlength;n++ ))
24 do
25 # get resn and chains
26 resn1=$(awk '{print $2}' <<< ${ca[$j]})
27 resn2=$(awk '{print $2}' <<< ${ca[$j+$n]})
28 chain1=$(awk '{print $6}' <<< ${ca[$j]})
29 chain2=$(awk '{print $6}' <<< ${ca[$j+$n]})
30 # check, if there is a gap or end of chain
31 if [ $resn1 -eq $((resn2-n)) ] && [ $chain1 = $chain2 ]
32 then
33 # ending coordinate
34 x2=$(awk '{print $3}' <<< ${ca[$j+$n]})
35 y2=$(awk '{print $4}' <<< ${ca[$j+$n]})
36 z2=$(awk '{print $5}' <<< ${ca[$j+$n]})
37 # calculate ca-ca distance
38 dist=$(bc <<< "scale=4;sqrt( ($x1 - $x2)^2 + ($y1 - $y2)^2 + ($z1 - $z2)^2 )")
39 # generate output
40 output=$output"$dist";"
41 fi
42 done
43 echo $output >> pdb/$pdbid.csv
44 done
45 rm $pdbid.tt
```

Quellcode A.2 Quellcode des Shellskriptes zur Berechnung der Peptidlängen in Di- bis Decapeptiden.

```
1 # Comparative modeling with ligand transfer from the template
2 from modeller import * # Load standard Modeller classes
3 from modeller.automodel import * # Load the automodel class
4 log.verbose() # request verbose output
5 env = environ() # create a new MODELLER environment to build this model in
6 # directories for input atom files
7 env.io.atom_files_directory = ['.', '../atom_files']
8 # Read in HETATM records from template PDBs
9 env.io.hetatm = True
10 a = automodel(env,
11     alnfile = 'alignment.ali', # alignment filename
12     knowns = 'template', # codes of the templates
13     sequence = 'target') # code of the target
14 a.starting_model= 1
15 a.ending_model = 20
16 #Very thorough VTFM optimization:
17 a.library_schedule = autosched.slow
18 a.max_var_iterations = 300
19 # Thorough MD optimization:
20 a.md_level = refine.slow
21 # Repeat the whole cycle 2 times and do not stop unless obj.func. > 1E6
22 a.repeat_optimization = 2
23 a.max_molpdf = 1e6
24 # do the actual comparative modeling
25 a.make()
```

Quellcode A.3 automodel.py. Pythonskript zur Proteinstrukturvorhersage mittels MODELLER ohne Verfeinerung von Schleifenstrukturen.

```
1 # Comparative modeling with ligand transfer from the template
2 from modeller import * # Load standard Modeller classes
3 from modeller.automodel import * # Load the automodel class
4 log.verbose() # request verbose output
5 env = environ() # create a new MODELLER environment to build this model in
6 # directories for input atom files
7 env.io.atom_files_directory = ['.', '../atom_files']
8 # Read in HETATM records from template PDBs
9 env.io.hetatm = True
10 a = loopmodel(env,
11     alnfile = 'alignment.ali', # alignment filename
```

```
12     knowns = 'template',           # codes of the templates
13     sequence = 'target')          # code of the target
14 a.starting_model= 1
15 a.ending_model = 20
16 #Loops
17 a.loop.starting_model = 1         # First loop model
18 a.loop.ending_model  = 20        # Last loop model
19 a.loop.md_level      = refine.slow # Loop model refinement level
20 #Very thorough VTfM optimization:
21 a.library_schedule = autosched.slow
22 a.max_var_iterations = 300
23 # Thorough MD optimization
24 a.md_level = refine.slow
25 # Repeat the whole cycle 2 times and do not stop unless obj.func. > 1E6
26 a.repeat_optimization = 2
27 a.max_molpdf = 1e6
28 # do the actual comparative modeling
29 a.make()
```

Quellcode A.4 loops.py. Pythonskript zur Proteinstrukturvorhersage mittels MODELLER mit Verfeinerung von Schleifenstrukturen.

7.3 Hinterlegung der Rohdaten

Die Rohdaten der *cross-link* Experimente zu CFII sind in der ProteomeXchange-Datenbank mit dem Identifier PXD010177 hinterlegt. Details zu den identifizierten *cross-links* sind darüber hinaus im Ergänzungsmaterial (*Supplemental material*) von Schäfer *et al.* [27] zu finden.

Alle weiteren Rohdaten der weiteren *cross-link* Experimente und den Pulldown-Experimenten, sowie die MeroX-Ergebnisdateien, welche genaue Details zu den identifizierten *cross-links* enthalten, sind öffentlich zugänglich unter <http://dx.doi.org/10.25673/34748> hinterlegt.

Lebenslauf**Persönliche Daten**

Name: Christian Tüting
Geburtstag und -ort: 26. März 1989 in Hannover
Staatsangehörigkeit: deutsch
Familienstand: verheiratet
Adresse: Steinweg 1, 06110 Halle (Saale)

Ausbildung

2008: **Abitur**, *Hölty Gymnasium Wunstorf*

2009-2013 **Bachelorstudium – Biochemie**, *MLU Halle-Wittenberg*

2013 **Bachelor of Science – Biochemie**, *MLU Halle-Wittenberg, Institut für Biochemie und Biotechnologie, Enzymologie (Prof. Dr. Mike Schutkowski)*, zum Thema: „Entwicklung und Etablierung eines Trypsin-gekoppelten Sirtuin 5 Assays“

2013-2015 **Masterstudium – Biochemie**, *MLU Halle-Wittenberg*

2015 **Master of Science – Biochemie**, *Fraunhofer-Institut für Zelltherapie und Immunologie IZI, Außenstelle Molekulare Wirkstoffbiochemie und Therapieentwicklung, Außenstelle Molekulare Wirkstoffbiochemie und Therapieentwicklung (Dr. Stephan Schilling)*, zum Thema: „Expression und Reinigung von humanem Meprin alpha“

2015-2018 **Wissenschaftlicher Mitarbeiter, Doktorand**, *MLU Halle-Wittenberg, Institut für Biochemie und Biotechnologie, Allgemeine Biochemie (Prof. Dr. Elmar Wahle)*

Seit 2010 **Wissenschaftlicher Mitarbeiter**, *MLU Halle-Wittenberg, Institut für Biochemie und Biotechnologie, WE ZIK HALOmem NWG III, Kryo-Elektronenmikroskopie (Jun.-Prof. Dr. Panagiotis L. Kastiris)*

Publikationsliste**Publikationen**

Roessler, C., Tüting, C., Meleshin, M., Steegborn, C., & Schutkowski, M. (2015). A novel continuous assay for the deacylase sirtuin 5 and other deacetylases. *Journal of medicinal chemistry*, 58(18), 7217-7223.

Schäfer, P., Tüting, C., Schönemann, L., Kühn, U., Treiber, T., Treiber, N., Ihling, C., Graber, A., Keller, W., Meister, G., Sinz, A., & Wahle, E. (2018). Reconstitution of mammalian cleavage factor II involved in 3' processing of mRNA precursors. *RNA*, 24(12), 1721-1737.

Konferenzbeiträge

Poster Tüting, C., Schäfer, P., Schönemann, L., Ihling, C., Götze, M., Kühn, U., Sinz, A., & Wahle, E. Structural Analysis of Protein-Protein Interactions in the Polyadenylation Complex by Chemical Cross-linking and Mass Spectrometry. 49. Jahrestagung der DGMS (28. Februar – 2. März 2016) in Hamburg (DEU).

Poster & Vortrag Tüting, C., Schäfer, P., Schönemann, L., Ihling, C., Götze, M., Kühn, U., Sinz, A., & Wahle, E. Structural Analysis of Protein-Protein Interactions in the Polyadenylation Complex by Chemical Crosslinking and Mass Spectrometry. 6th Symposium on Structural Proteomics (17. November – 18. November) in Dortmund (DEU).

Poster Tüting, C., Schäfer, P., Schönemann, L., Ihling, C., Götze, M., Kühn, U., Sinz, A., & Wahle, E. Structural Analysis of Protein-Protein Interactions in the Polyadenylation Complex by Chemical Cross-linking and Mass Spectrometry. *Structural Mass Spectrometry Workshop* (3. Mai – 8. Mai 2017) in Rehovot (ISR).

Poster & Vortrag Tüting, C., Schäfer, P., Schönemann, L., Ihling, C., Götze, M., Kühn, U., Sinz, A., & Wahle, E. Structural Analysis of Protein-Protein Interactions in the Polyadenylation Complex by Chemical Cross-linking and Mass Spectrometry. 31st ASMS Sanibel Conference on Mass Spectrometry (24. Januar – 27. Januar 2019) in St. Petersburg (USA).

Poster & Vortrag Tüting, C., Iacobucci, C., Ihling, C., Sinz, A., & Kastritis, P. L. Structural variation in protein complexes revealed by database-wide mapping of chemical cross-links. 1st HALOmem International Meeting (14. November – 15. November 2019) in Halle (Saale) (DEU).

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die eingereichte schriftliche Fassung der Arbeit entspricht der auf dem elektronischen Speichermedium. Weiterhin versichere ich, dass die vorliegende Arbeit noch nicht als Abschlussarbeit an anderer Stelle eingereicht wurde.

Halle (Saale), 23.01.2020

Christian Tüting

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mir diese Arbeit erst ermöglicht haben und mich auf dem Weg begleitet haben.

Ganz besonders möchte ich meinem Doktorvater, **Elmar Wahle**, für die Betreuung meiner Arbeit danken. Dies gab mir die Möglichkeit, viele neue Methoden, von Klonierung, Proteinreinigung und -charakterisierung, bis Massenspektrometrie zu lernen.

Auch möchte ich mich bei **Andrea Sinz** zum einen für die Übernahme des Zweitgutachtens, für die erfolgreiche Kooperation und die gemeinsamen Konferenzen bedanken. Ohne die unzähligen massenspektrometrischen Analysen hätte ich meine Arbeit nicht in dieser Art und Weise verfassen können.

Ich möchte mich sehr herzlich bei **Prof. Hennig Urlaub** für die Übernahme des Drittgutachtens bedanken.

Auch **Jun.-Prof. Panagiotis L. Kastritis** möchte ich danken. Er hat mir gezeigt, wie ich die *cross-link*-Daten nutzen kann, um Strukturmodelle zu generieren, was am Ende ein eigenes Kapitel meiner Arbeit geworden ist.

Darüber hinaus möchte ich natürlich allen Kollegen in den Arbeitsgruppen AG Wahle und AG Sinz, mit denen ich während meiner Arbeit zusammenarbeiten durfte, meinen Dank ausdrücken. Hervorzuheben sind hier **Dr. Uwe Kühn**, **Dr. Lars Schönemann**, **Peter Schäfer**, **Moritz Schmidt** und **Florian Kluge**, die mit mir an der mRNA-3'-Prozessierung geforscht haben, **Dr. Christian Ihling**, **Dr. Michael Götz** und **Dirk Tänzler** für die Massenspektrometrie, das *cross-linking* und die Datenauswertung und **Lisa Schmidt** für das Korrekturlesen meiner Arbeit.

Auch danken möchte ich dem **Graduiertenkolleg 1591** für die Finanzierung meiner Promotionsstelle und der Konferenzen, auf denen ich meine Daten vorstellen durfte, bedanken.

Ein großer Dank gilt auch meiner Familie und meinen Freunden, die mich diesen Weg begleitet haben: meiner Frau und besten Freundin Nicole, meinen Eltern und meinen Brüdern, ganz besonders Michael für das Korrekturlesen und der Hilfe bei der Statistik. Auch danken möchte ich meinen Freunden in Halle wie in Hannover, die stets für mich da waren.

In stillem Gedenken an Manni († 2018).