

Biochemical studies on the NEXT complex

Masterarbeit aus dem Fachgebiet

Biochemie

Vorgelegt von

cand. chem. Ksenia Finogenova

geboren am 7. September 1989

in Moskau

Für die Masterprüfung in Biochemie an der
Ludwig-Maximilians-Universität München

Datum der mündlichen Prüfung: 22. Mai 2014

Beginn der Masterarbeit: 14. Juli 2014

Masterarbeit beim Prüfungsausschuss eingereicht am 14. Januar 2015

Erklärung

Ich versichere, dass ich die vorgelegte Masterarbeit am Max-Planck-Institut für Biochemie, Martinsried, unter der Anregung and Anleitung von Prof. Dr. Elena Conti selbstständig durchgeführt und keine anderen als die angegebenen Hilfsmittel und Quellen benützt habe.

München, 14. Januar 2015

1. Berichterstatter: Prof. Dr. Karl-Peter Hopfner

2. Berichterstatter: Prof. Dr. Klaus Förstemann

Abbreviations

CBC = Cap-binding complex

CV = column volume

DNA = Deoxyribonucleic acid

FL = full length

LP = limited proteolysis

NEXT = nuclear exosome targeting

Ni-NTA = Ni²⁺-Nitrilotriacetic acid

Mtr4 = mRNA transport regulator 4

PROMPTs = promoter upstream transcripts

Rbm7 = RNA-binding motif protein 7

RNA = Ribonucleic acid

SEC = Size exclusion chromatography

Ski = Super killer

TRAMP = Trf4/Air2/Mtr4p Polyadenylation

ZCCHC8 = Zinc finger CCHC domain-containing protein 8

Zusammenfassung

Die Mehrheit der eukaryotischen Transkripte kodiert nicht für Proteine, sondern übernimmt regulatorische Aufgaben. Diese werden als nicht kodierende RNAs bezeichnet. Fehlerhafte Prozessierungen oder Transkription kann zu fehlerhaften nicht kodierenden RNAs und folglich zu Schwierigkeiten für das normale Funktionieren der Zelle führen. Daher wurden Kontrollmechanismen entwickelt, um diese RNAs abzubauen. Der wichtigste Spieler hierbei ist das Exosom, eine 3'-5' Exoribonuklease, dessen Kernpartikel in Eukaryoten überraschenderweise katalytisch inaktiv ist (außer in Pflanzen). Aus diesem Grund benötigt das Exosom die Anwesenheit weiterer Untereinheiten mit Nukleaseaktivität und Kofaktoren. Im menschlichen Zellkern wird das Exosom von dem NEXT (nuclear exosome targeting) Komplex unterstützt, welcher aus drei Proteinen besteht: hMtr4 (human mRNA transport regulator 4), eine RNA Helikase, hRbm7 (RNA-binding motif protein 7), welches eine RRM (RNA recognition motif) Domäne besitzt und hZCCHC8 (Zinc finger CCHC domain-containing protein 8).

NEXT rekrutiert das Exosom für den spezifischen Abbau von PROMPTs (promoter upstream transcripts). PROMPTs entstehen bei der Transkription von bidirektionalen Promotoren und müssen abgebaut werden, um ein unidirektionales Transkriptionsprodukt zu gewährleisten. Eine kürzlich durchgeführte Studie zeigte, dass der Rekrutierungsmechanismus für das Exosom neben den NEXT Komplex auch den CBC (cap-binding complex) und andere Proteine involviert, welche zusammen den CBC-NEXT Superkomplex bilden.

Diese Arbeit zeigt erstmalige Ansätze nicht nur zur biochemischen, sondern auch zur strukturellen Erforschung des NEXT Komplexes. hMtr4 kann als Volllänge-Protein gereinigt werden, und zudem hRbm7 als C-terminal verkürztes Konstrukt. Die Reinigung von hZCCHC8 ist erschwert durch die Aggregation des Proteins. Die hier präsentierten Ergebnisse zeigen die ersten Versuche zur Optimierung des Reinigungsprotokolls, wobei ein kurzes Konstrukt von hZCCHC8 im Komplex mit hRbm7 aufgereinigt werden kann.

Das hZCCHC8 Protein kann sowohl hMtr4 als auch hRbm7 binden, während die zwei Proteine untereinander nicht interagieren. Der hauptsächliche Fokus der Arbeit liegt auf der Analyse der Interaktion von hZCCHC8 und hRbm7. Die Ergebnisse zeigen, dass die RRM Domäne von hRbm7 für die hZCCHC8 Bindung genügt. Das Screening von verschiedenen hZCCHC8 Konstrukten weist auf eine bestimmte Region des hZCCHC8 Proteins hin, die für die Interaktion mit hRbm7 wichtig ist. Limitierte Proteolyse des hZCCHC8-hRbm7 Komplexes resultiert in einem minimalen Kern-Komplex, welcher in Kristallisationsexperimenten Spheruliten-ähnliche Strukturen ergab.

Abstract

The majority of eukaryotic transcripts does not code for proteins instead they often have regulatory roles and are referred to as noncoding RNAs. The diversity of noncoding RNAs the cell has to process is astonishing and because aberrant processing events or transcriptional errors leading to defective noncoding RNA would be fatal for the normal functioning, cells have developed surveillance mechanisms to degrade aberrant noncoding RNA. The key player in this mechanism is the exosome, a multimeric protein complex with 3'-5' exoribonuclease activity. Strikingly, the eukaryotic core exosome was found to be catalytically inactive (except in plants) and requires therefore the presence of other subunits with nuclease activity, and cofactors. The human nuclear exosome is assisted by the NEXT (nuclear exosome targeting) complex, which consists of three proteins: hMtr4 (human mRNA transport regulator 4), a RNA Helicase, hRbm7 (RNA-binding motif protein 7), which contains a RRM (RNA recognition motif) domain and hZCCHC8 (Zinc finger CCHC domain-containing protein 8).

NEXT is responsible for targeting PROMPTs (promoter upstream transcripts) for exosome mediated degradation. PROMPTs are the result of transcription at bidirectional promoters and need to be degraded for a unidirectional transcription product. A recent study revealed that the PROMPTs targeting mechanism by the NEXT complex involves the cap-binding complex (CBC). Together with other proteins a CBC-NEXT (CBCN) complex is formed, which recruits the exosome for RNA degradation.

This study presents a first approach towards structural and biochemical studies on the NEXT complex. hMtr4 can be purified as full length protein and hRbm7 as a C-terminal truncated short construct. Purification of hZCCHC8 is complicated by aggregation of the protein. First optimizations of the purification protocol are shown, with which a short construct of hZCCHC8 in complex with hRbm7 can be purified.

hZCCHC8 is the scaffolding protein in NEXT. Studies on interaction of hZCCHC8 and hRbm7 form a major body of the thesis. The RRM of hRbm7 is found to be sufficient to mediate hZCCHC8 interaction. hZCCHC8 construct screening highlights regions of hZCCHC8 important for hRbm7 interaction. Limited proteolysis of the hZCCHC8-hRbm7 complex yields a minimal core complex, which forms spherulite-like structures in crystallization trials.

Contents

List of Figures	VII
1 Introduction	1
1.1 The exosome, a 3'-5' exoribonuclease	1
1.2 Mtr4 and Ski2 containing complexes	3
1.3 The NEXT complex, a nuclear exosome cofactor in human	4
1.4 RNA and protein recognition by the RRM domain	6
1.5 Aims of the thesis	7
2 Material	8
2.1 <i>E. coli</i> strains	8
2.2 Constructs	8
2.3 Vectors	9
2.4 DNA Oligonucleotides	10
2.5 Enzymes	11
2.6 Chemicals and solutions	11
2.7 Buffers and Media	12
2.8 Equipment	15
2.9 Consumables	16
2.10 Software and Webtools	17
3 Methods	18
3.1 Bacterial culture	18
3.1.1 General bacterial culture	18
3.1.2 Transformation	18
3.1.3 Protein production	18
3.2 Cloning	19
3.2.1 PCR	19
3.2.2 Agarose Gel Electrophoresis	19
3.2.3 Ligation Independent Cloning (LIC)	20
3.2.3.1 Vector processing	20
3.2.3.2 Insert processing	20
3.2.3.3 Annealing reaction	21

3.2.4	DNA sequencing	21
3.3	Protein biochemistry	21
3.3.1	Purification of hMtr4 constructs	21
3.3.1.1	Cell lysis	21
3.3.1.2	Ni-NTA purification	22
3.3.1.3	Heparin Ion Exchange Chromatography	22
3.3.1.4	Dialysis & Tag cleavage	23
3.3.1.5	Reverse-Ni-NTA	23
3.3.1.6	Protein concentration	23
3.3.1.7	SEC	23
3.3.2	Purification of hRbm7	24
3.3.2.1	Cell lysis	24
3.3.2.2	Ni-NTA purification	24
3.3.2.3	Dialysis & Tag cleavage	24
3.3.2.4	Reverse-Ni-NTA Purification	24
3.3.2.5	Size exclusion chromatography	24
3.3.3	Purification of hZCCHC8 constructs - initial protocol	25
3.3.3.1	Cell lysis	25
3.3.3.2	GSH-Affinity-purification	25
3.3.3.3	Dialysis	25
3.3.3.4	Reverse-Ni-NTA Purification	26
3.3.3.5	Q Ion Exchange chromatography	26
3.3.3.6	Size exclusion chromatography	26
3.3.4	Purification of hZCCHC8 constructs - optimized protocol	26
3.3.4.1	Cell lysis	27
3.3.4.2	Initial purification steps	27
3.3.4.3	Unfolding of hZCCHC8 by 8 M Urea	27
3.3.4.4	Refolding of hZCCHC8 in presence of hRbm7	27
3.3.4.5	Dialysis with lowering the salt concentration	27
3.3.4.6	Size exclusion chromatography	28
3.3.5	Protein storage	28
3.3.6	Determination of Protein concentration	28
3.3.7	Mass Spectrometry/Peptide mass fingerprint	28
3.3.8	SDS-PAGE	28
3.4	Limited Proteolysis	29
3.4.1	Analytical scale	29
3.4.2	Time course	30
3.4.3	Preparative scale	30
3.5	Crystallization experiments	30

4	Results & Discussion	31
4.1	NEXT domain organisation	31
4.2	NEXT construct design	32
4.3	Expression/Purification of Mtr4 constructs	34
4.3.1	Purification of hMtr4 (Full length)	34
4.3.2	Purification of Mtr4 Δ N	36
4.4	Expression/Purification of hRbm7 constructs	37
4.4.1	Purification of full length hRbm7	38
4.4.2	Purification of the short hRbm7 construct (1-137) and RRM-only construct (1-86)	39
4.5	Purification of hZCCHC8 constructs	41
4.5.1	hZCCHC8 Full length protein is prone to aggregation	41
4.5.2	The short construct (1-317) of the hZCCHC8 protein is also prone to aggregation	42
4.5.3	Optimization of the hZCCHC8 purification protocol	43
4.5.4	Purification of hZCCHC8 full length - optimized protocol	44
4.5.5	Purification of a short construct (41-337) of hZCCHC8 - optimized protocol	46
4.6	hZCCHC8 is the scaffolding protein in the NEXT complex	48
4.7	Production of a hRbm7-hZCCHC8 Interaction core complex by limited proteolysis	49
4.7.1	Analytical-scale limited proteolysis of full length hZCCHC8 in complex with hRbm7 (1-137)	49
4.7.2	Analytical-scale limited proteolysis of hZCCHC8 (1-317) and hRbm7 (1-137)	51
4.7.3	Preparative-Scale limited proteolysis of hZCCHC8 (1-317) and hRbm7 (1-137) complex	53
4.8	Towards the identification of regions important for hRbm7-hZCCHC8 complex formation	55
4.8.1	Critical residues for interaction with Rbm7 lie between residues 264-317 of hZCCHC8	55
4.8.2	The RRM of hRbm7 is sufficient for hZCCHC8 interaction	56
4.9	Crystallization experiments	58
5	Summarized discussion	59
6	Supplementary Material	62
7	Acknowledgments	67
	Bibliography	68

List of Figures

1.1	Cofactors of the exosome in yeast and human.	2
1.2	Structure of Mtr4 bound to RNA and ADP.	3
1.3	Structure of Mtr4 bound to N-terminal regions of Air2 and Trf4 in TRAMP. . .	4
1.4	Mechanism of the CBCN-exosome substrate degradation.	5
1.5	Schematic representations of the canonical RRM fold and the UHM-ligand interaction.	7
4.1	Domain organization of NEXT complex.	31
4.2	Initial working and new NEXT constructs.	32
4.3	The RRM domain in hRbm7 is highly conserved.	33
4.4	Purification of Mtr4 (Full length).	35
4.5	Purification of Mtr4 Δ N.	37
4.6	Expression tests of full-length hRbm7 with different fusion tags.	38
4.7	Purification of full length hRbm7.	38
4.8	Purification of (1-137) construct of hRbm7.	39
4.9	Purification of RRM-only hRbm7 construct (1-86).	40
4.10	Purification of hZCCHC8 (FL) - initial protocol.	42
4.11	Purification of hZCCHC8 short construct (1-317) - initial protocol.	43
4.12	Purification of hZCCHC8 full length - optimized protocol.	45
4.13	Purification of short construct (41-337) hZCCHC8 - optimized protocol.	47
4.14	Pulldown interaction analyses for the NEXT complex.	49
4.15	Analytical-Scale Limited Proteolysis of full length hZCCHC8 with hRbm7 (1-137).	50
4.16	Analytical-Scale Limited Proteolysis of hZCCHC8 (1-317) with hRbm7 (1-137).	51
4.17	Time-course LP of the complex consisting of hZCCHC8 (1-317) and hRbm7 (1-137).	52
4.18	Size-exclusion chromatography of hZCCHC8 (1-317) and hRbm7 (1-137) analytical-scale limited proteolysis.	53
4.19	Analysis of large-scale limited proteolysis by SEC.	54
4.20	Size-exclusion chromatography of hZCCHC8 (1-263) and hRbm7 (1-137).	55
4.21	hZCCHC8 (1-263) fails to form a complex with hRbm7 (1-137).	56
4.22	The RRM of hRbm7 is sufficient for hZCCHC8 interaction.	57
4.23	Spherulite-like structures of the hZCCHC8-hRbm7 minimal core complex.	58
5.1	Schematic representation of interaction between the NEXT proteins.	60

S1	hZCCHC8 is highly conserved around the Zn knuckle and proline rich domain.	62
S2	hZCCHC8-secondary structure prediction.	64
S3	hRbm7-secondary structure prediction.	65
S4	hRbm7 is highly conserved around the RRM motif in metazoa.	66

1 Introduction

In 1986, Walter Gilbert's RNA world hypothesis marked a milestone in the history of origin of life studies (Gilbert, 1986). His hypothesis argued for a world guided on self-replicating RNA molecules predating the DNA based life of today. The RNA is a 'recipe' for proteins and therefore holder of information same as the DNA. The discovery that RNA can possess catalytic activities shaped the idea that the first self-replicating systems consisted of RNA molecules. In contrast, life of today is orchestrated mainly by the ultimate holder of information, the DNA, which appeared only later on the scene. And although it seemed that the importance of RNA declined with this step, new scientific insights on processes such as RNA interference have brought back the interest on RNA.

RNA is formed in a process called transcription and controversially, the majority of the transcripts in eukaryotes does not code for proteins. These transcripts referred to as noncoding RNA include ribosomal RNAs, transfer RNAs, small nuclear and nucleolar RNAs, microRNAs, small interfering RNAs and Piwi-interacting RNAs. Besides participating in important processes such as translation, noncoding RNAs have regulatory roles in gene expression and chromatin structure. Conclusively, transcriptional errors leading to defective RNAs or aberrant processing events could constitute an obstacle for normal functioning of a cell. To avoid this, the cells have evolved surveillance mechanisms to detect and degrade aberrant noncoding RNAs, which involves exoribonucleases such as the exosome (Amaral et al., 2008).

1.1 The exosome, a 3'-5' exoribonuclease

The eukaryotic exosome core consists of 9 subunits forming a barrel-like structure (reviewed in Schmid and Jensen, 2008). Six of these subunits (Rrp41, Rrp42, Rrp43, Rrp45 and Rrp46) form a hexameric ring and are structurally similar to the RNase PH, a 3'-5' exoribonuclease involved in bacterial RNA degradation (reviewed in Lorentzen et al. 2008, Deutscher et al., 1988). On top of this ring a 'cap' is assembled by Csl4, Rrp4, Rrp40 subunits, which contain S1/KH domains typical for RNA-binding proteins. Controversially, despite the presence of six RNase PH proteins the exosome core is catalytically inactive in eukaryotes with exception of plants, a result of accumulated mutations in important residues (Chekanova et al., 2000; Liu et al., 2006). Hence it requires the presence of additional subunits and cofactors. The exonucleolytic activity is provided by the 10th subunit, Rrp44 (in human Dis3 for the nuclear exosome and Dis3L1 for the cytoplasmic exosome), which harbors a C-terminal 3'-5' exoribonuclease- and a N-terminal

endonucleolytic cleavage activity (reviewed in Schmid and Jensen, 2008). Biochemical and structural studies showed, that the 3' end of the RNA passes from the S1/KH proteins cap through the central channel to the exoribonuclease site of Rrp44 (Bonneau et al., 2009; Makino et al., 2013). Another exoribonuclease Rrp6 forms a complex with Rrp47 and interacts with the nuclear exosome in yeast (Burkard and Butler, 2000) (human PM-SCI-100 and C1D). The exosome is involved in several processes, which differ depending on its localization. In the nucleus, the exosome is involved in processing of ribosomal RNAs and surveillance of aberrant nuclear RNAs, while in the cytoplasm the exosome participates in the turnover and surveillance of mRNA. However, for managing these tasks the exosome requires the presence of cofactors (Fig. 1.1), in yeast it is assisted by the TRAMP (Trf4/Air2/Mtr4p Polyadenylation) complex in the nucleus and the Ski (Super killer) complex in the cytoplasm. The human exosome is assisted by the NEXT (nuclear exosome targeting) complex in the nucleus and the hSki complex in the cytoplasm. Interestingly, the TRAMP complex can only be found in the human nucleolus (Lubas et al., 2011). All complexes have a RNA Helicase in common: Mtr4 in TRAMP and NEXT and Ski2 in Ski complex, which are structurally very similar (reviewed in Schmid and Jensen, 2008).

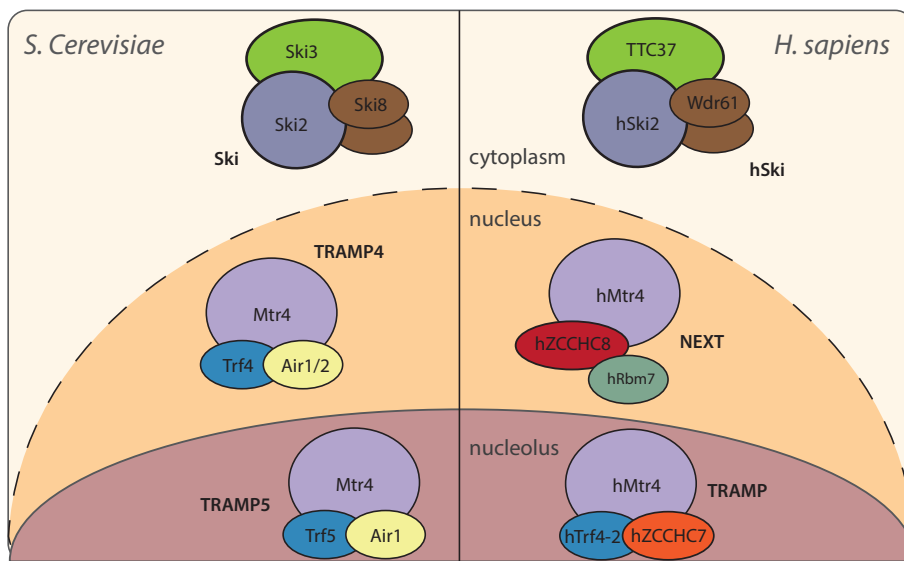


Figure 1.1: Cofactors of the exosome in yeast and human.

The cytoplasmic exosome is assisted by the Ski complex in both yeast and human. The yeast nuclear exosome associates with the TRAMP4 complex for degradation of substrates as aberrant nuclear RNAs, whereas the human requires the presence of the NEXT complex. Here the hMtr4 interacts with a different set of proteins, hZCCHC8 and hRBM7. The cofactor of the nucleolar exosome is the TRAMP5 complex in yeast and the TRAMP complex in human.

1.2 Mtr4 and Ski2 containing complexes

Mtr4 and Ski2 both belong to the DExH-box family of RNA helicases and share over 30% sequence identity (Anderson and Parker, 1998). The structure of yeast Mtr4 in the apo conformation was solved by Jackson et al. (2010) and bound to RNA and ADP by Weir et al. (2010) (Fig. 1.2).

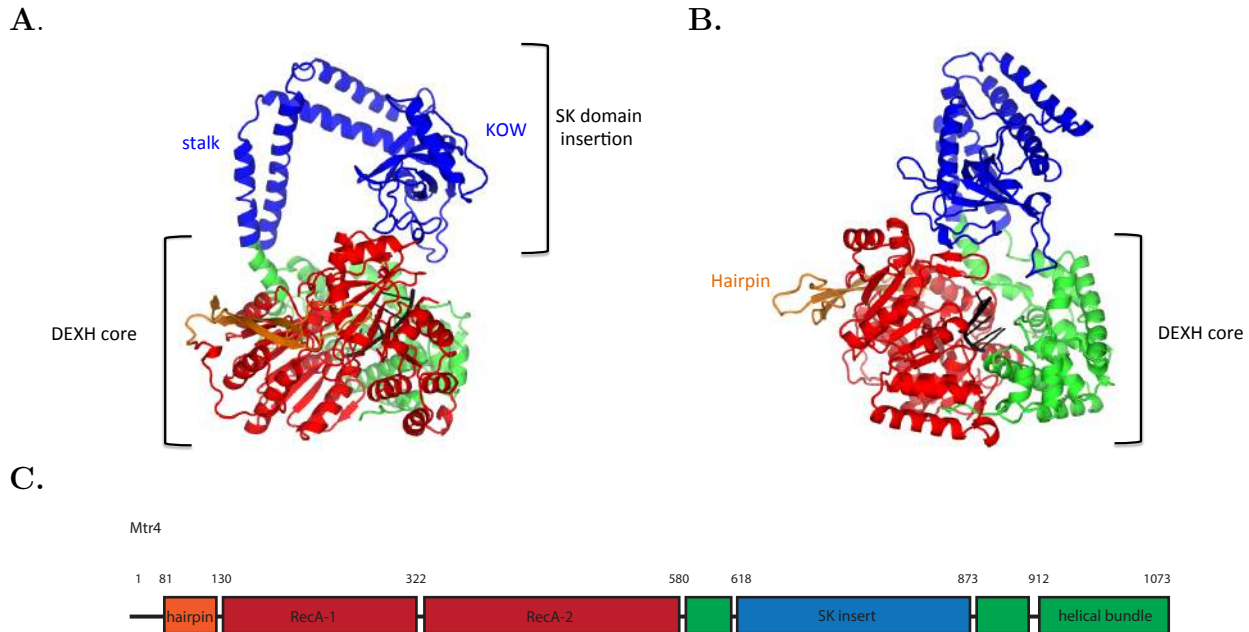


Figure 1.2: Structure of Mtr4 bound to RNA and ADP.

Structure of Mtr4 bound to RNA and ADP at 2.9 Å resolution. **A.** Front view. **B.** Side view. RecA1 and RecA2 domains are shown in red. RNA (black) and ADP are accommodated between RecA1 and RecA2 domains. Together with helical bundle and winged helix (green) they form the helicase-core, which shows structural similarity to other DExH helicases. An arch-like feature bridges the DExH core and the KOW beta-barrel domain (blue). PDB file: 2XGJ (Weir et al., 2010). **C.** Schematic domain organization of Mtr4. Lines represent low-complexity regions. Domains are shown in colors corresponding to the features in the structure.

In the structure two RecA domains are packed against each other to accommodate the RNA and ADP molecules. The winged helix and helical bundle domain form together with the RecA domains the helicase core. The DExH core is connected to a region, which elongates in a stalk form and is connected to a beta-barrel domain. This domain is structurally similar to the KOW domain of L24 ribosomal protein and was shown to bind *in vitro* transcribed tRNA^{iMet} (Weir et al., 2010). From yeast studies it is known that Mtr4 unwinds RNA duplexes in 3'-5' direction *in vitro* and participates in the processing of various structured RNAs as snRNAs and snoRNAs (van Hoof et al., 2000). Additionally, as part of the TRAMP complex it is involved in RNA surveillance (reviewed in Schmid and Jensen, 2008). In the TRAMP complex Mtr4 is associated with Air2 and Trf4. A recently published structure showed that the interaction takes place

on the DExH core of Mtr4, which interacts with both proteins of the heterodimer Trf4/Air2 complex. (Fig. 1.3) (Falk et al., 2014). Interestingly, in the Ski complex, that is centered around the helicase Ski2, Ski3 and Ski8 also bind to the helicase core (Halbach et al., 2013). Recently a new Mtr4 containing complex was discovered in human cells, namely the NEXT complex (Lubas et al., 2011). Here hMtr4 associates with a different set of proteins, namely hZCHC8 and hRbm7.

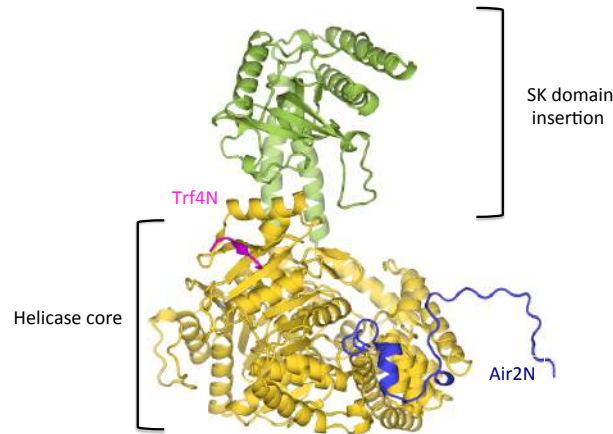


Figure 1.3: Structure of Mtr4 bound to N-terminal regions of Air2 and Trf4 in TRAMP. Structure of Mtr4 bound to the unstructured regions of Air2 and Trf4 in TRAMP at 2.4 Å resolution. The Mtr4 helicase core is shown in yellow-the SK domain is colored green. The N-termini of Air2 and Trf4 are shown in blue and pink, respectively. Both proteins are bound by the helicase core of Mtr4. PDB file: 4U4C (Falk et al., 2014).

1.3 The NEXT complex, a nuclear exosome cofactor in human

The NEXT complex was discovered by searching for new interactions partners of the human exosome (Lubas et al., 2011). The same study showed that the NEXT complex is important for mediating exosome degradation of promoter upstream transcripts (PROMPTs). PROMPTs are the result of transcription at bidirectional promoters and 3' processing at promoter-proximal polyadenylation (pA) sites (Almada et al., 2013; Ntini et al., 2013). Degradation of PROMPTs by the exosome machinery ultimately results in unidirectional transcription products. Depletion of hRbm7 leads to an accumulation of PROMPTs, suggesting that they are bound by hRbm7 (Lubas et al., 2011). Moreover, recent studies revealed the mechanism of PROMPTs degradation. It includes the cap-binding complex (CBC) consisting of the subunits CBP20 and CBP80 (Andersen et al., 2013). RNA containing stem loops, such as small nuclear RNAs (snRNA), is bound co-transcriptionally by the arsenite resistance protein 2 (ARS2). ARS2 forms a complex with the CBC, referred to as CBCA. In the case of functional processed snRNA, the ARS2 protein associates with PHAX (phosphorylated adaptor for RNA export) to enable the RNA export. In contrast, following transcription at bidirectional promoters ARS2 is assumed to bind

to antisense PROMPTs and to stimulate early transcription termination. Additionally, Ars2 associates with the hZC3H18 protein and the NEXT complex, thus forming a CBC-NEXT (CBCN) complex, which finally recruits the exosome for degradation (Fig. 1.4) (Andersen et al., 2013). This mechanism is biochemically similar to CBC-NSS (Nrd1p, Nab3p and Sen1p) complex in *S. cerevisiae* (Vasiljeva and Buratowski, 2006; Andersen et al., 2013). This is also supported by the fact, that until now no sequence homologs of the Nrd1p and Nab3p proteins of the NSS complex have been found in human cells. Other substrates for CBCN complex are misprocessed U1 snRNA and products of genes, encoding replication-dependent histones, which both require a polyadenylation (pA) site-independent mechanism for their 3' end processing (Andersen et al., 2013).

Recent studies showed that the NEXT complex is also participating in DNA damage response (Blasius et al., 2014). In particular the study demonstrated that upon UV-induced DNA damage Rbm7 is phosphorylated on Ser136 and Ser204 by the MK2 Kinase, which is activated in the course of the MAPKAP kinase-2 (MK2)-dependent DNA damage signaling pathway. Consistent with the idea that PROMPTs are mainly bound by Rbm7, UV-induced phosphorylation of Rbm7 lead to an accumulation of PROMPTs. Additionally, immunoprecipitation studies confirmed that FLAG-Rbm7 could bind less efficiently to PROMPTs upon UV treatment compared to FLAG-Rbm7-S136A/S204A, a mutated form of Rbm7, which cannot be phosphorylated by MK2 Kinase (Blasius et al., 2014).

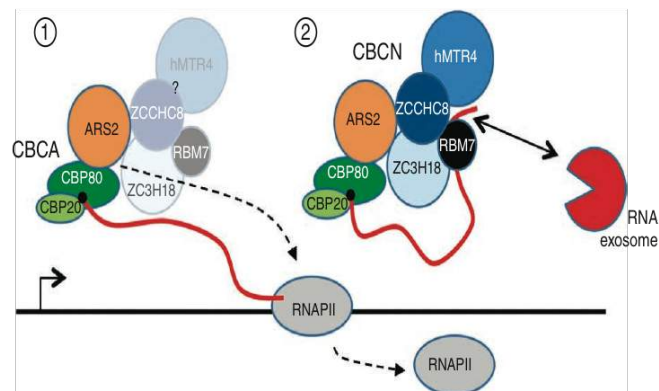


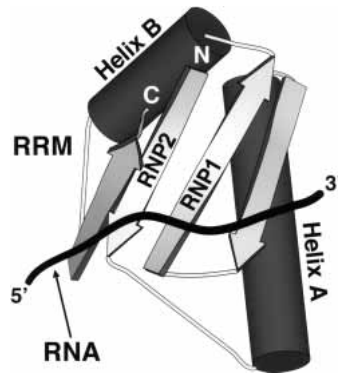
Figure 1.4: Mechanism of the CBCN-exosome substrate degradation.

(1) The CBCA complex is formed during transcription of PROMPTs and 3' extended transcripts and promotes early TSS-proximal transcription termination. The NEXT complex associates with Ars2 either already co-transcriptionally or after transcription termination (2) to form the CBCN complex, which recruits the exosome for degradation. The figure was taken from Andersen et al. (2013).

1.4 RNA and protein recognition by the RRM domain

The RRM domain is one of the most abundant RNA recognition motifs. A lot of research has been focused on the RNA-binding activity of the RRM domain in the last two decades. Important features for RNA binding are two conserved motifs referred to as ribonucleoprotein 1 (RNP1) and ribonucleoprotein 2 (RNP2) with the consensus [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY] and [ILV]-[FY]-[ILV]-X-N-L ($X = \text{any amino acid}$), respectively (reviewed in Maris et al., 2005). The first structure of the RRM domain was published for the U1A RRM, which showed two α -helices stacked against four antiparallel β -strands with topology $\beta\alpha\beta\alpha\beta$, nowadays known as the canonical RRM fold (Fig. 1.5 A) (Nagai et al., 1990). The RNPs are folded as a β sheet, with RNP1 being part of β_3 strand and RNP2 on β_1 strand. RNA binding is facilitated by RNP1 residues 1,3 and 5 and RNP2 residue 2. RRM domains can also participate in protein recognition. Here one can distinguish between three classes of RRM-protein interaction: (1) RRM-RRM interaction (2) RRM, which can bind to RNA and a non-RRM protein and (3) RRM, which interacts with proteins but cannot bind RNA (reviewed in Maris et al., 2005). An example for the first category is the hnRNP A1 (UP1), which contains two RRM domains forming a compact fold. The RRM domains interact with each other via their α -helix 2; two salt bridges between two arginines of the first RRM and two aspartic acids of the second are stabilizing the interaction (Xu et al., 1997; Shamoo et al., 1997). Together with other examples the structure shows that this additional RRM-RRM interaction contributes to a larger RNA-binding surface (Handa et al., 1999; Deo et al., 1999). Examples for the second category are U2B' and CBP20 proteins (Mazza et al., 2001; Price et al., 1998). Both proteins require the presence of cofactors, U2A' and CBP80, respectively, to bind RNA. Both U2B' and CBP20 interact with their cofactors via their α -helices and loop 4, however the binding differs: while the U2B' forms hydrophobic contacts to a leucine-rich repeat domain of U2A', the CBP20 interacts with CBP80 via salt bridges and hydrogen bonds. The β -sheet in both proteins is kept accessible for RNA. Finally some RRM domains participating in protein interaction cannot bind to RNA. This is the case for example when the β -sheet is involved in protein interaction as in the UPF2-UPF3 complex (Kadlec et al., 2004). Another example are unusual RRM motifs referred to as U2AF homology motifs (UHMs) (reviewed in Kielkopf et al., 2004). UHM differ from RRM motifs in the following ways: (1) poor conservation in RNP motifs, (2) Arg-X-Phe motif in the loop region following Helix B and (3) conserved acidic residues in Helix A. The Helix A as well as the Arg-X-Phe motif contribute to formation of a hydrophobic pocket in which the UHM ligand usually inserts a conserved Trp (Fig. 1.5 B). Recently a novel UHM-ligand interaction mode was shown on the example of the pre-mRNA RES (pre-cursor mRNA retention and splicing) complex (Wysoczański et al., 2014).

A.



B.

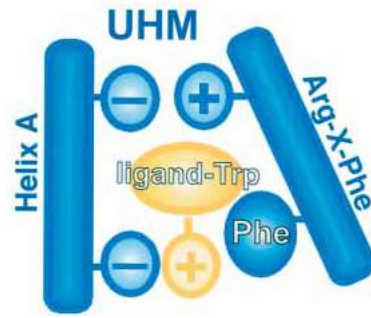


Figure 1.5: Schematic representations of the canonical RRM fold and the UHM-ligand interaction.

A. Schematic architecture of the canonical RRM fold from the structure of U1A protein in complex with RNA (PDB code: 1URN). N- and C-terminus are indicated. RNA is bound by the RNP1/2 motifs. **B.** Schematic model of the UHM-ligand interaction. Helix A together with the loop containing the Arg-X-Phe motif form a hydrophobic pocket. In some cases of UHM-ligand interactions a Trp of the ligand, which inserts into the pocket, seems to play an important role. Figures were taken from Kielkopf et al. (2004).

1.5 Aims of the thesis

The aim of this thesis is to produce, purify and reconstitute the NEXT complex and to elucidate the structural details of the interaction between the NEXT proteins. Interactions between hMtr4, hRBM7 and hZCCHC8 will be tested and characterized by size exclusion chromatography and GST-pulldowns. After successful reconstitution of the NEXT complex the minimal regions necessary for the interaction will be identified by rational construct design and limited proteolysis. With the final goal to obtain structural insights into the organization of the NEXT complex crystallization trials will be performed.

2 Material

2.1 *E. coli* strains

<i>E. coli</i> strains	Description/Genotypes
BL21(DE3) Star pRare	electrocompetent bacteria cells for expression with lower RNase expression and chloramphenicol resistance, F ⁻ ompT hsdSB(rB-, mB-) gal dcm rne131 (DE3) pLysSRARE (Cam ^R)
XL1 Blue	electrocompetent bacteria cells for cloning, endA1 gyrA96(nal ^R) thi-1 recA1 relA1 lac glnV44 F'[:Tn10 (Tet ^r) proAB ⁺ lac/ ^q Δ(lacZ)M15] hsdR17 (r _K -m _K -)

2.2 Constructs

Construct	Description	Clone Nr.
hMtr4 FL	Full length (1-1042) hMtr4 in pEC- K-3C-GST vector	H1
hMtr4ΔN	hMtr4 construct lacking the first 74 N-terminal residues (75-1042) in pEC-K-3C-GST vector	H3
hMtr4ΔN	hMtr4 construct lacking the first 74 N-terminal residues (75-1042) in pEC-K-3C-Z vector	H4
hRbm7 FL	Full length (1-266) hRbm7 in pEC-S-3C-His vector	H5
hRbm7(1-137)	Short construct (1-137) hZCCHC8 in pEC-S-3C-His vector	H7
hRbm7(1-137)	Short construct (1-137) hZCCHC8 in pEC-S-3C-Z vector	H8
hZCCHC8 FL	Full length hZCCHC8 (1-707) in pEC-A-3C-GST vector	H9
hZCCHC8 FL	Full length hZCCHC8 (1-707) in pEC-A-3C-Ztag	H10
hZCCHC8 (1-317)	Short construct hZCCHC8 (1-317) in pEC-A-3C-GST vector	H11
hZCCHC8 (1-317)	Short construct hZCCHC8 (1-317) in pEC-A-3C-Z vector	H12
hRbm7 FL	Full length hRbm7 (1-266) in pEC-K-3C-GB1 vector	H15

Construct	Description	Clone Nr.
hRbm7 (1-137)	Short construct hRbm7 (1-137) in pEC-K-3C-GST vector	H16
hRbm7 FL	Full length hRbm7 (1-266) in pEC-K-3C-TRX vector	H17
hRbm7 FL	Full length hRbm7 (1-266) in pEC-K-3C-GST vector	H18
hRbm7 FL	Full length hRbm7 (1-266) in pEC-K-HI-SUMO vector	H19
hRbm7 FL	Full length hRbm7 (1-266) in pEC-K-HT-MBP vector	H20
hRbm7 FL	Full length hRbm7 (1-266) in pEC-K-CBP-CHis vector	H21
hZCCHC8 (41-263)	Short construct hZCCHC8 (41-263) in pEC-A-3C-GST vector	H22
hZCCHC8 (41-337)	Short construct hZCCHC8 (41-337) in pEC-A-3C-GST vector	H23
hRbm7-RRM	Short construct hRbm7 (RRM only = 1-86) in pEC-S-3C-Z vector	H24
hRbm7 (1-98)	Short construct hRbm7 (1-98) in pEC-S-3C-Z vector	H25
hZCCHC8 (1-263)	Short construct hZCCHC8 (1-263) in pEC-A-3C-GST vector	H26
hZCCHC8 (1-337)	Short construct hZCCHC8 (1-337) in pEC-A-3C-GST vector	H27
hZCCHC8 FL-Strep	Full length hZCCHC8-Strep (1-707) in pEC-A-3C-GST-vector	H28
hZCCHC8 (41-707)-Strep	Short construct hZCCHC8 (41-707)-Strep-tag in pEC-A-3C-GST vector	H29
hZCCHC8 (91-337)-Strep	Short construct hZCCHC8 (91-337)-Strep-tag in pEC-A-3C-GST vector	H30
hZCCHC8 (149-337)-Strep	Short construct hZCCHC8 (149-337)-Strep-tag in pEC-A-3C-GST vector	H31
hZCCHC8 (41-337)-Strep	Short construct hZCCHC8 (41-337)-Strep-tag in pEC-A-3C-GST vector	H32
hZCCHC8 (41-525)-Strep	Short construct hZCCHC8 (41-525)-Strep-tag in pEC-A-3C-GST vector	H33

2.3 Vectors

The vectors used to clone the constructs listed in 2.2 are shown below.

Vector name	Source/Description
pEC-A-3C-GST	house-made; Amp resistance, 3C protease site, N-terminal His-GST tag
pEC-K-3C-GST	house-made; Kan resistance, 3C protease site, N-terminal His-GST tag
pEC-S-3C-GST	house-made; Strep resistance, 3C protease site, N-terminal His-GST tag
pEC-K-HT-GB1	house-made; Kan resistance, TEV protease site, N-terminal His-GB1 tag
pEC-K-3C-TRX	house-made; Kan resistance, 3C protease site, N-terminal His-TrxA tag
pEC-K-HI-Sumo	house-made; Kan resistance, SUMO protease site, N-terminal His-SUMO tag

pEC-K-HT-MBP	house-made; Kan resistance, TEV protease site, N-terminal His-MBP tag
pEC-K-CBP-CHIS	house-made; Kan resistance, TEV protease site, N-terminal CBP tag, C-terminal His-tag
pEC-K-3C-Z	house-made; Kan resistance, 3C protease site, N-terminal His-Z tag
pEC-S-3C-Z	house-made; Strep resistance, 3C protease site, N-terminal His-Z tag

2.4 DNA Oligonucleotides

Oligonucleotides were purchased from Sigma-Aldrich and treated as recommended by the manufacturer.

ORF	5'-3' sequence	Primer ID
hMtr4 2 For 3C	ccagggcccgcactcgatgGCGGACGCATTCGGAGATGAG	SF245
hMtr4 75 For 3C	ccagggcccgcactcgatgATTTTGGAAAGAAGCCCAGGATAGAAGAGTC	SF246
hMtr 1042 Rev	cagaccgccaccgactgcttaCAAGTAGAGGCTGGCAGCAAACACAATATCTCTCTTTG	SF247
hRbm7 137 Rev	cagaccgccaccgactgcttaAGAAGAGAAAGATCTCTGAATTATCTGTGCTGATGAAG	SF 249
hRbm7 266 Rev	cagaccgccaccgactgcttaGTGTCGAGATGAGCGCCATTTTCC	SF250
hZCCHC8 2 For 3C	ccagggcccgcactcgatgGCCGCAGAGGTGTATTTGGCG	SF251
hZCCHC8 263 Rev	cagaccgccaccgactgcttaTCCACAGGCATCCATATACTCTTTTCTCTTTTCCAC	SF252
hZCCHC8 317 Rev	cagaccgccaccgactgcttaCCCTAGCTGGCGCATCCGATAG	SF253
hZCCHC8 707 Rev	cagaccgccaccgactgcttaTTCAGAGGCCTTTTTGTTTTCTGCTGG	SF254
Seq hMtr4 511 For	CAGCGGGA AAAACAGTATGCG	SF256
Seq hMtr4 1000 For	GCCTGCATCTTGTGGTTGATG	SF 257
Seq hMtr4 1522 For	CCCGCAAATTTGATGGGAAGG	SF258
Seq hMtr4 2013 For	GAAGGAGATGACTTTGGCTGG	SF259
Seq hMtr4 2520 For	CTGCAAAGCGAGA ACTGAAG	SF260
Seq hZCCHC8 519 For	CAGTGTGTAGGAAGTGTCTG	SF 261
Seq hZCCHC8 1013 For	CGGGGCTTGCACTCTATGATGG	SF 262
Seq hZCCHC8 1499 For	CCGCTGACTCCCAGTGACTC	SF263
hRbm7 2 For 3C	ccagggcccgcactcgatgGGGGCGGCGGCGGCGGAAG	SF264
hRbm7 2 For Sumo	accaggaacaaccggcgccgctcgatgGGGGCGGCGGCGGCGGAGG	KF01
hRbm7 266 Rev Sumo	gcaaagcaccggcctcgttaGTGTCGAGATGAGCGCCATTTTCC	KF02
hRbm7 2 For TEV	ccaggagcagcctcgatgGGGGCGGCGGCGGCGGAGG	KF03
hRbm7 266 Rev	gcaaagcaccggcctcGTGTCGAGATGAGCGCCATTTTCC	KF04
hZCCHC8 41 For 3C	ccagggcccgcactcgatgGAAAATGGGGTCGGCGACG	SF367

ORF	5'-3' sequence	Primer ID
hZCCHC8 263 Rev	cagaccgccaccgactgcttaTCCACAGGCATCCATATACTCTTTTCTCTTTTCAC	SF268
hZCCHC8 337 Rev	cagaccgccaccgactgcttaATCATAGAGTGCAAGCCCCGAATTCTC	SF269
hRbm7 86 Rev	cagaccgccaccgactgcttaTGATCTAAATTGAATTTTG	SF370
hRbm7 98 Rev	cagaccgccaccgactgcttaTGACAAACTGACATCTTGTGGGGC	SF371
hZCCHC8 707 Rev Strep	cagaccgccaccgactgcttaCTTTTCGAATTGTGGGTGGCTCCACG CGCTTCctcagaggcctttttgtttttctgtggtttcggg	SF372
hZCCHC8 91 For 3C	ccaggggcccgactcgatgGATGGACCTATATTACAGATTCTATT CATGAACAATGCTATTTCAAAGCAATATC	KF05
hZCCHC8 149 For 3C	ccaggggcccgactcgatgGAGGACCACAAAGTGAAGAGTCCTGTGC	KF06
hZCCHC8 337 Rev Strep	cagaccgccaccgactgcttaCTTTTCGAATTGTGGGTGGCTCCACGCGCTTC CATCATAGAGTGCAAGCCCCGAATTCTC	KF07
hZCCHC8 525 Rev Strep	cagaccgccaccgactgcttaCTTTTCGAATTGTGGGTGGCTCCACGCGCTTC CCCTCTGCTGTTCTTCAAGTTCTTCTAGAGTCAGTG	KF08
hZCCHC8 41 For 3C	ccaggggcccgactcgatgGAGAATGGGGTCGGCGACCGGAG	KF09
pGEX_fwd (seq)	GGG CTG GCA AGC CAC GTT TGG TG	
T7 term (seq)	CTA GTT ATT GCT CAG CGG T	

2.5 Enzymes

Enzyme	Source
TEV, Sumo and 3C Proteases	In house expression and purification
DNAseI	Sigma/Fluka
Fusion Flash Polymerase	Finnzymes
PfuUltraII fusion HS Polymerase	Stratagene
T4 DNA Ligase	NEB
T4 DNA Polymerase	NEB
Elastase, Chymotrypsin, Trypsin, Substilisin, Glu C	Roche

2.6 Chemicals and solutions

All chemicals were ordered from Sigma or Fluka, unless otherwise declared.

Kits	source
Qiaquick Spin Miniprep	Qiagen
Qiagen MinElute Gel Extraction	Qiagen
Wizard SV Gel and PCR Clean Up System	Promega

2.7 Buffers and Media

Buffer	Component
Coomassie staining solution (1000 ml)	10% acetic acid 20% EtOH 100 ml Coomassie
(Hepes)-Dialysis Buffer	20 mM Hepes/NaOH pH 7.5 150 mM NaCl 10% Glycerol 5 mM 2-Mercaptoethanol (freshly added)
(Tris)-Dialysis Buffer	20 mM Tris-HCl pH 7.5 200-250 mM NaCl 10 mM Imidazol 10% Glycerol 5 mM 2-Mercaptoethanol (freshly added)
(Hepes)-Dilution Buffer	20 mM Hepes/NaOH pH 7.5 10% Glycerol 5 mM 2-Mercaptoethanol (freshly added)
(Phosphate)-Dilution Buffer	50 mM NaH ₂ PO ₄ /NaOH pH 7.5 10 mM Glycerol
(Hepes)-Elution Buffer	20 mM Hepes/NaOH pH 7.5 250 mM NaCl 250 mM Imidazol 10% Glycerol 5 mM 2-Mercaptoethanol (freshly added)
(Tris)-Elution Buffer	20 mM Tris-HCl pH 7.5 250 mM NaCl 250 mM Imidazol 10% Glycerol 5 mM 2-Mercaptoethanol (freshly added)
(Tris)-GSH-Elution Buffer	20 mM Tris-HCl pH 7.5 500 mM NaCl

Buffer	Component
	10 mM Imidazol 10% Glycerol 30 mM Gluthatione (freshly added)
(Phosphate)-Elution Buffer	50 mM NaH ₂ PO ₄ /NaOH pH 7.5 250 mM NaCl 300 mM Imidazole 10% Glycerol
(Hepes)-SEC Buffer	20 mM Hepes/NaOH pH 7.5 150 mM NaCl 2 mM DTT (freshly added)
(Tris)-SEC Buffer	20 mM Tris-HCl pH 7.5 150 mM NaCl 10% Glycerol 2 mM DTT (freshly added)
Heparin A Buffer	20 mM Hepes/NaOH pH 7.5 150 mM NaCl 10% Glycerol 5 mM 2-Mercaptoethanol (freshly added)
Heparin B Buffer	20 mM Hepes/NaOH pH 7.5 1 M NaCL 10% Glycerol 5 mM 2-Mercaptoethanol (freshly added)
(Hepes)-High Salt Buffer	20 mM Hepes/NaOH pH 7.5 500 mM NaCl 1000 mM NaCl 10 mM Imidazol 10% Glycerol
(Tris)-High Salt Buffer	50 mM Tris-HCl pH 7.5 1000 mM NaCl 15 mM Imidazol 10% Glycerol
(Phosphate)-High Salt Buffer	50 mM NaH ₂ PO ₄ /NaOH pH 7.5 1000 mM NaCl 10 mM Imidazole 10% Glycerol 5 mM 2-Mercaptoethanol (freshly added)
(Hepes)-Lysis Buffer	50 mM Hepes/NaOH pH 7.5 500 mM NaCl 10 mM Imidazol

Buffer	Component
	10% Glycerol 5 mM 2-Mercaptoethanol (freshly added) 1 mM PMSF (freshly added) 5 µg/ml DNaseI
(Tris)-Lysis Buffer	20 mM Tris-HCl pH 7.5 500 mM NaCl 1 mM PMSF (freshly added) 5 µg/ml DNase I (freshly added) 5 mM 2-Mercaptoethanol (freshly added) 10 mM Imidazol 10% Glycerol
Phosphate Buffer for CD spectroscopy	20 mM NaPhosphate 50 mM NaF pH 7.4
Q (Tris)-Buffer A	20 mM Tris-HCl pH 7.5 100 mM NaCl 10% Glycerol 5 mM 2-Mercaptoethanol (freshly added)
Q (Tris)- Buffer B	20 mM Tris-HCl pH 7.5 1000 mM NaCl 10 % Glycerol 5 mM 2-Mercaptoethanol (freshly added)
SDS-PAGE Running Buffer (10x)	0.25 M Trizma Base 1.92 M Glycine 1% SDS
SDS sample buffer (2x)	100 mM Tris pH 6.8 10% 2-Mercaptoethanol 4% SDS 0.2% Bromophenol Blue 20% Glycerol

Media	Component
Luria-Bertani (LB) (Miller, 1972)	1% (w/v) bacto tryptone 0.5% (w/v) bacto yeast extract 170 mM NaCl Adjust pH to 7.6 with NaOH

Media	Component
LB agar plates	1.5% (w/v) bacto agar in LB respective antibiotics at appropriate concentration
SOC Medium	2% (w/v) bacto tryptone 0.5% (w/v) bacto yeast extract 10 mM NaCl 1 mM MgCl ₂ 2.5 mM KCl 10 mM MgSO ₄ 0.4% glucose Adjust pH to 7.2
Terrific Broth (TB) (Sambrook and Russel, 2001)	1.2% (w/v) bacto tryptone 2.4% (w/v) bacto yeast extract 0.4% (w/v) glycerol ddH ₂ O to 900 ml 0.017 M KH ₂ KPO ₄ 0.072 M K ₂ HPO ₄

2.8 Equipment

Equipment	Type	Manufacturer
Bacterial shaker	KS-15/Climo-Shaker ISF1X	Kühner
Cell lysis Sonicator	Sonifyer VS70T	Bandelin Electronics
Centrifuge	Avanti J-20 XP Micro centrifuge 5417C & 5810 RC 5C	Beckman Coulter Eppendorf Sorvall
Chromatography protein purification FPLC	ÄKTA Prime ÄKTA Purifier	GE Healthcare
Chromatography protein purification Columns	HiLoad Superdex 75 HiLoad Superdex 200 Superose 6 HisTrap FF 5 ml HiTrap Heparin 5ml	GE Healthcare
Crystallisation: Pipetting Robot	Phoenix	Art Robbins Instruments
Crystallisation: Visualisation Robot	XtalFocus	ExploraNova La Rochelle

Equipment	Type	Manufacturer
Electroporator	Gene Pulser/Micro Pulser	Bio-Rad
Electro-cuvette	Gene pulser 0.1 cm electrode gap	Bio-Rad
ITC Calorimeter	VP-ITC	MicroCal (GE Healthcare)
Nanodrop Spectrophotometer		PeqLab
pH Meter	Lab860	Schott
Pipettes	eppendorf Research	Eppendorf
Peristaltic pump	Minipuls 2	Limburg- Offheim
SDS PAGE system	Mini-Protean Tetra	Bio-Rad Laboratories
SDS PAGE power supply	Power Pack P25T	Biometra
SDS PAGE Gel Imaging	Gel visualisation Gel scanner V700 photo	Roth Epson
Thermo shaker	Thermomixer comfort	Eppendorf
Transilluminator		Invitrogen
Vortex mixer	Vortex-Genie	Scientific Industries

2.9 Consumables

Product	Manufacturer
Amicon Ultra Centrifugal Filters	Merck Millipore
96 well MRC96T crystallization plate	SwissCl
Crystallization plate PEG Suite	Qiagen
Centrifuge bottle (polypropylene, 250 ml)	Beckman Coulter
Erlenmeyer flask (1 l)	SCHOTT AG
Falcon Tubes (15 ml, 50 ml)	Becton Dickinson
Fraction Tubes	Sarstedt
Laboratory bottle (1 l)	SCHOTT AG
Laboratory vacuum manifold Vac-Ma	Promega GmbH
MF-Millipore Membrane Filters	Merck Millipore
Nitrile laboratory gloves	Semperit AG
PageRuler	Thermo Scientific
Parafilm sealing film	Bemis Company
Pipette tips (10 μ l / 300 μ l / 1000 μ l)	BrandTech Scientific
Reaction tubes (1.5 ml)	Eppendorf AG

Product	Manufacturer
Serological pipettes	Carl Roth GmbH + Co. KG
Syringe	BD

2.10 Software and Webtools

The following software and online servers were used during practical work as well as writing the thesis.

Adobe Illustrator (www.adobe.com/products/illustrator)

BLAST (www.ncbi.nlm.nih.gov/blast)

JabRef (<http://jabref.sourceforge.net>)

Jalview (<http://www.jalview.org/>)

Latex/LyX (<http://www.lyx.org>)

Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>)

PROTPARAM (<http://web.expasy.org/protparam>)

T-COFFEE (<http://www.tcoffee.org/Projects/tcoffee/>)

Unicorn Software (<http://www.gelifesciences.com/webapp/wcs/stores/servlet/catalog/de>)

UNIPROT (<http://www.uniprot.org>)

GELifeSciences-DE/brands/unicorn-control-software)

3 Methods

3.1 Bacterial culture

3.1.1 General bacterial culture

Bacterial cultivation was performed in LB medium or on LB agar plates with the respective antibiotics for positive selection. For liquid culture, 5 ml LB medium was inoculated with bacteria. For standard applications, incubation was performed overnight at 37 °C with constant shaking at 300 rpm (liquid culture) or in the incubator (agar plates).

3.1.2 Transformation

Material

- plasmid DNA
- electrocompetent BL21 Star pRare or XL1Blue cells
- selection plate (LB + antibiotics)
- electroporator
- electro-cuvette
- Shaker (37°C, 300 rpm)
- Incubator 37°C
- SOC Medium

For cloning purposes XL1Blue cells were used. BL21(DE3) Star pRare cells were used for protein expression. The electrocompetent cells were thawed on ice for 1 min. Approximately 0.2 µg of plasmid DNA was added to the bacteria suspension. The cells were then transferred to 0.1 cm electroporation cuvette and an electrical pulse of 1.8 kV (time constant 5.8) was applied. 270 µl of SOC medium was then added to the cell suspension. For outgrowth of bacteria, the suspension was incubated for 60 min with constant shaking and plated on agar containing the respective antibiotic.

3.1.3 Protein production

All proteins were produced in *E. coli* under the control of the bacteriophage T7 transcription using 450 ml of TB medium and 50 ml phosphate buffer in 2 l glass flasks. The cells were

grown to a density of OD₆₀₀ 2 under constant shaking at 220 rpm at 37°C. After reducing the incubation temperature to 18°C, protein expression was induced with 0.4 mM IPTG as the host bacteria contained a chromosomal copy of the T7 RNA polymerase gene (Novagen pET System). Expression was carried out overnight (approximately 16 hours) and the cells were subsequently harvested by centrifugation (6000 rpm, 10 min). Cell pellets were either immediately used for purification of flash frozen in liquid nitrogen and stored at -20°C.

3.2 Cloning

Ligation independent cloning (LIC) method was used for cloning purposes.

3.2.1 PCR

DNA fragments with the open reading frame (ORF) of interest were amplified by polymerase chain reaction (PCR) (first described in Mullis et al. 1986). The PCR reactions were set up as follows:

Reaction mix
10 ng template DNA
5 µl Pfu Buffer (10x Stock)
1 µl Forward Primer (10 µM Stock)
1 µl Reverse Primer (10 µM Stock)
0.5 µl dNTPs (20 mM Stock)
1 µl Pfu Ultra II
ddH ₂ O up to 50 µl

The following thermocycling conditions were used:

Cycle step	Temperature	Duration	Cycles
Initial denaturation	98°C	2 min	1
Denaturation	98°C	30 s	30
Annealing	50-55°C	35 s	
Extension	72°C	1.5 min	
Final extension	72°C	10 min	1

3.2.2 Agarose Gel Electrophoresis

To analyse whether DNA fragments were successfully amplified and to remove the vector DNA used as template, the PCR products were separated by agarose gel electrophoresis. The gels were casted with 1% Agarose in 1xTBE and SYBR Safe stock (Invitrogen) was used as dye for visualisation in a 1:10000 dilution. PCR samples were diluted in 6x Orange Dye (Fermentas) to

a final concentration of 1x. 1 kb plus DNA ladder (Fermentas, 0.5 µg/lane) was used as marker. Electrophoresis was performed in 1x TBE Buffer. DNA fragments were visualized with an UV transilluminator. The correct bands were excised on a blue light transilluminator emitting blue light at a wavelength of 470 nm. DNA was extracted using the Wizard SV Gel and PCR Clean Up System according to manufacturer's instruction.

3.2.3 Ligation Independent Cloning (LIC)

The principle of the LIC system has been described in Aslanidis and de Jong (1990). Briefly, insert and vector DNA are treated with T4 DNA Polymerase (an 3'-5' exonuclease) to create complementary single-stranded overhangs, which mediate DNA ligation. In the absence of all deoxynucleotides, except dATP or dTTP, the T4 polymerase excises the end of a DNA down to the first thymide or adenine, respectively.

3.2.3.1 Vector processing

For processing, 2 µg LIC vector were linearised with 60 U SacII in 100 µl reaction volume. The linearised vector was loaded on a 0.8% Agarose gel and excised from the gel. DNA extraction was performed with Wizard SV Gel and PCR Clean Up System according to manufacturer's instruction.

LIC vector reaction mix

450 ng linearized vector
3 µl T4 DNA Pol. Buffer (Stock 10x)
3 µl dTTP (Stock 25 mM)
1.5 µl DTT (Stock 100 mM)
0.6 µl T4 DNA Pol. LIC qualified (Novagen)
ddH ₂ O up to 30 µl

The reaction was incubated 30 min at room temperature. T4 DNA Polymerase was inactivated 20 min at 75°C.

3.2.3.2 Insert processing

PCR products (obtained as described above) were processed after the following protocol:

LIC vector reaction mix

600 ng gel purified PCR product
2 µl T4 DNA Pol. Buffer (Stock 10x)
2 µl dATP (Stock 25 mM)
1 µl DTT (Stock 100 mM)
0.4 µl T4 DNA Pol. LIC qualified (Novagen)
ddH ₂ O up to 20 µl

The reaction was incubated 30 min at room temperature. T4 DNA Pol. was inactivated 20 min at 75 °C.

3.2.3.3 Annealing reaction

2 µl of insert and 1 µl of vector were incubated for 10 min at room temperature for annealing. 1 µl EDTA (Stock 25 mM) was added and the reaction mix was incubated for another 10 min. The final product was transformed in XL1 Blue cloning cells.

3.2.4 DNA sequencing

All plasmid DNA was sequenced by the core facility of MPI Biochemistry. Vector promoters and terminator regions were typically used as primer sites (see 2.4 for complete list). For larger genes, internal sequencing primers were used, which were interspaced approximately every 500 bp.

3.3 Protein biochemistry

3.3.1 Purification of hMtr4 constructs

All steps were performed on ice or at 4°C.

3.3.1.1 Cell lysis

Material

- (Hepes)-Lysis buffer; 5 mM 2-Mercaptoethanol, DNase I 5 µg/ml, 1 mM PMSF
- Sonicator (40% amplitude, 10-12 min 1s/1s)
- Centrifuge (pre-cooled 4 °C)
- 2x/1x sample buffer
- heating block (95 °C)

The bacterial pellet was resuspended in (2 x weight of pellet x $\frac{\text{ml}}{\text{mg}}$) Lysis buffer and 5 mM of 2-Mercaptoethanol together with 5 µg/ml DNaseI were added. 1 mM PMSF was added shortly before lysis under constant stirring. Cells were lysed using the Sonicator for 10 min. 10 µl of lysate were diluted in 5 µl 2x SDS sample buffer and 15 µl 1x SDS sample buffer, heated at 95 °C for several minutes and stored for analysis by SDS-PAGE (CE = cell extract). The cell lysate was then cleared by centrifugation at 20,000 rpm 1 hour at 4 °C. Sample was taken for SDS PAGE analysis (SUP = supernatant). The centrifuged cell lysate was filtered through a Merck Millipore 5 µm filter.

3.3.1.2 Ni-NTA purification

Material

- (Hepes)-Lysis Buffer; 5 mM 2-Mercaptoethanol
- Ni-NTA HisTrap FF 5ml column
- peristaltic pump
- (Hepes)-Wash Buffer; 5 mM 2-Mercaptoethanol
- 2x/1x sample buffer
- (Hepes)-Elution Buffer; 5 mM 2-Mercaptoethanol
- heating block (95 °C)

A Ni-NTA (Nickel-Nitrilotriacetic acid) 5 ml column was pre-equilibrated with Lysis Buffer. The supernatant from 3.3.1.1 was loaded on the column. A sample was taken for SDS-PAGE analysis (FT=flow-through). Subsequently, the column was washed in 3 steps: 5 column volumes (CV) lysis buffer, 5 CV High Salt buffer. Protein was eluted in total 45 ml volume using Elution buffer containing Imidazole. After each wash and elution steps, samples were taken for analysis by SDS-PAGE (W1 = wash 1, W2 = wash 2, E1 = elution 1, E2 = elution 2, E3 = elution 3).

3.3.1.3 Heparin Ion Exchange Chromatography

Material

- Heparin A and B Buffers; 5 mM 2-Mercaptoethanol
- Heparin Column 5 ml
- ÄKTA Prime and software Unicorn
- Fraction tubes
- (Hepes)-Dilution Buffer; 5 mM 2-Mercaptoethanol
- peristaltic pump

Elution fractions from 3.3.1.2 containing the protein were collected and diluted with (Hepes)-Dilution Buffer to reduce the NaCl concentration to a final concentration of 150 mM. The Heparin column was pre-washed with 5 CV Heparin B Buffer to remove contaminations and equilibrated with 5 CV Heparin A Buffer. Subsequently, the protein was loaded on the Heparin column using the peristaltic pump and the column connected to the ÄKTA Prime system. The column was washed with 5 CV Heparin A Buffer and 5 CV 10% (235 mM NaCl) Heparin B Buffer. The protein was eluted in 50% Heparin B (575 mM NaCl) Buffer. During the run absorption at 280 nm was monitored and fractions corresponding to the absorption peaks analyzed on SDS-PAGE and were collected for later dialysis.

3.3.1.4 Dialysis & Tag cleavage

Material

- (Hepes)-Dialysis Buffer; 5 mM 2-Mercaptoethanol
- His-3C Protease
- Dialysis membrane

To remove the His-GST- or His-Z-tag the His-3C Protease was added to the combined elution fractions containing the purified protein (1 mg of protease per 100 mg protein). A dialysis membrane was pre-equilibrated first in ddH₂O and then in Dialysis Buffer. The protease-containing protein solution was filled in the dialysis membrane and dialysed under stirring at 4°C over night. After dialysis a sample was taken for analysis by SDS-PAGE (DIA=Dialysis).

3.3.1.5 Reverse-Ni-NTA

Following dialysis the untagged protein was purified by Reverse-Ni-NTA to remove the uncleaved protein and the His-tagged 3C protease. Without the tag the protein did not bind to the column matrix and was present in the flow-through fractions. Then the Ni-NTA column was eluted to remove the bound protease and tag. Samples were taken from the flow-through and elution fractions for analysis by SDS-PAGE (FTII= Flow-through II, EII=Elution II).

3.3.1.6 Protein concentration

Before concentrating the sample the Millipore Amicon Ultra Centrifugal Filter (Cut-off 50 kDa) was pre-equilibrated with (Hepes)-Dialysis Buffer. The flow-through fraction from 3.3.1.5 was then concentrated to 2 ml for size exclusion chromatography through centrifugation at 4°C, 4000 rpm.

3.3.1.7 SEC

Material

- Äkta Purifier and software Unicorn
- HiLoad Superdex 200
- Syringe
- correct sample loop
- (Hepes)-SEC Buffer, 2 mM DTT
- 96 deep well plate

The SEC column HiLoad Superdex 200 was equilibrated with SEC buffer. The sample loop was washed three times with the SEC buffer and subsequently loaded with 2 ml of sample. During the run absorption at 280, 260 and 220 nm was monitored and protein containing fractions were analyzed by SDS-PAGE and the fractions containing the purified protein were pooled and concentrated in Millipore Amicon Ultra Centrifugal Filters.

3.3.2 Purification of hRbm7

All steps were performed on ice or at 4°C.

3.3.2.1 Cell lysis

Material same as in 3.3.1.1 with the following change:

- (Phosphate)-Lysis Buffer; 5 mM 2-Mercaptoethanol, DNase I 5 µg/ml, 1 mM PMSF freshly added

Cell lysis of bacteria containing hRbm7 was performed after the same protocol as described in 3.3.1.1 using a (Phosphate)-Lysis Buffer.

3.3.2.2 Ni-NTA purification

Material same as in 3.3.1.2 with the following changes:

- (Phosphate)-Lysis Buffer; 5 mM 2-Mercaptoethanol
- (Phosphate)-High Salt Buffer; 5 mM 2-Mercaptoethanol
- (Phosphate)-Elution Buffer; 5 mM 2-Mercaptoethanol

Ni-NTA purification of hRbm7 constructs was carried out following the same protocol described in 3.3.1.2 with Phosphate Buffers instead of Hepes.

3.3.2.3 Dialysis & Tag cleavage

Material same as in 3.3.1.4 with the following change:

- (Tris)-Dialysis Buffer; 5 mM 2-Mercaptoethanol

Same protocol as in 3.3.1.4 using a (Tris)-Dialysis Buffer.

3.3.2.4 Reverse-Ni-NTA Purification

The untagged hRbm7 constructs were further purified via Reverse-Ni-NTA purification, which followed the same protocol as described in 3.3.1.5 using buffers containing Tris instead of Hepes.

3.3.2.5 Size exclusion chromatography

Material same as in 3.3.1.7 with the following changes:

- HiLoad Superdex 75
- (Tris)-SECBuffer, 2 mM DTT

SEC was used as a final purification step for hRbm7. Hiload Superdex 75 column was pre-equilibrated with (Tris)-SEC Buffer. The protein was separated from contaminants and nucleic acids. Fractions containing the protein were pooled and concentrated.

3.3.3 Purification of hZCCHC8 constructs - initial protocol

All steps were performed on ice or at 4 °C.

3.3.3.1 Cell lysis

Material same as in 3.3.1.1 with the following change:

- (Tris)-Lysis Buffer; 1 mM DTT , DNase I 5 µg/ml, 1 mM PMSF, 2.5 mM EDTA, 5 tablets protease inhibitor cocktail (Roche), 5 mM MgCl₂

Cell lysis of bacteria containing hZCCHC8 constructs was performed after the same protocol as described in 3.3.1.1 using a (Tris)-Lysis Buffer with 1 mM DTT, DNase I 5 µg/ml, 1 mM PMSF, 2.5 mM EDTA, 5 tablets protease inhibitor cocktail (Roche), 5 mM MgCl₂ freshly added.

3.3.3.2 GSH-Affinity-purification

Material

- (Tris)-Lysis Buffer; 1 mM DTT, 2.5 mM EDTA
- Pre-packed GSH column
- peristaltic pump
- (Tris)-High Salt Wash Buffer; 1 mM DTT, 2.5 mM EDTA
- (Tris)-GSH-Elution Buffer

Since all the hZCCH8 constructs were GST-tagged, a GSH affinity column was used as the first purification step. Efficiency of the binding to the column was checked by analyzing the flow through by SDS PAGE (FT= flow through). The column was first washed with 5 CV (Tris)-Lysis Buffer, then with 5 CV (Tris)-High Salt Wash Buffer. An excess of glutathione in the (Tris)-GSH-Elution Buffer was used to elute the protein. Fractions from the wash steps and the elution were analyzed by SDS PAGE.

3.3.3.3 Dialysis

Material same as in 3.3.1.4 with the following change:

- (Tris)-Dialysis Buffer; 5 mM 2-Mercaptoethanol, 1 mM EDTA

To remove the His-GST tag the protein was treated with 3C protease during dialysis. The protocol was identical to 3.3.1.4 using a (Tris)-Dialysis Buffer.

3.3.3.4 Reverse-Ni-NTA Purification

To remove the tag after cleavage by the 3C protease hZCCHC8 was subjected to a Reverse-Ni-NTA purification. The protocol was identical to 3.3.1.5 using buffers containing Tris.

3.3.3.5 Q Ion Exchange chromatography

Material

- Q (Tris)-A and B Buffers, 5 mM 2-Mercaptoethanol, 2.5 mM EDTA
- Q FF Column
- ÄKTA Prime and software Unicorn
- Fraction tubes
- peristaltic pump

Since hZCCHC8 has a low pI, a Q column was reasoned as the most suitable ion exchange column to remove nucleic acid contamination. The Q column was pre-washed with 5 CV Q-Buffer B and pre-equilibrated with 5 CV Q Buffer A. The protein was loaded on the Q column. After washing with 5 CV Q Buffer A and 10% Q Buffer B (190 mM NaCl), a gradient with a target of 50% Buffer B (550 mM NaCl) over a volume of 100 ml was started. The protein eluted at approximately 40% B (460 mM NaCl). Fractions were analyzed on SDS PAGE.

3.3.3.6 Size exclusion chromatography

Material same as in 3.3.1.7 with the following changes:

- Superdex 200
- (Tris)-SEC Buffer, 2 mM DTT

SEC was used as a final purification step for hZCCHC8. Due to low protein amount an analytical S200 was used.

Since purification of hZCCHC8 with the initial protocol was hindered by problems, optimization of the protocol was a major part of the thesis (see Results section). The optimized protocol is presented below.

3.3.4 Purification of hZCCHC8 constructs - optimized protocol

The bacterial expression culture of hZCCHC8 was supplied with 100 μ M ZnSO₄ (added during cooling process to 18°C).

3.3.4.1 Cell lysis

Material same as in 3.3.1.1 with the following change:

- (Tris)-Lysis Buffer; 5 mM 2-Mercaptoethanol, DNaseI
5 µg/ml, 1 mM PMSF, 5 tablets protease inhibitor
cocktail (Roche), 1 µM ZnSO₄

The protocol was identical to 3.3.1.1. Different to initial protocol: EDTA was omitted; 2-Mercaptoethanol was used instead of DTT, 1 µM ZnSO₄ was supplied in every buffer..

3.3.4.2 Initial purification steps

The subsequent purification steps remained the same as described in the initial protocol except for the following changes: EDTA was omitted, 2-Mercaptoethanol was used instead of DTT; 1 µM ZnSO₄ was supplied in every buffer.

3.3.4.3 Unfolding of hZCCHC8 by 8 M Urea

Material:

- 1 M NaCl, 8 M Urea; 5 mM 2-Mercaptoethanol

Following Q-Ion-Exchange column hZCCHC8 was unfolded by dialysis against a buffer containing 8 M Urea, 1 M NaCl over night.

3.3.4.4 Refolding of hZCCHC8 in presence of hRbm7

Material:

- Dialysis Buffer 1: 1 M NaCl, 20 mM Tris/HCl pH 7.5,
10% Glycerol, 10 µM ZnSO₄; 2 mM DTT
- Purified hRbm7
- Dialysis membrane

For refolding the unfolded hZCCHC8 was mixed dropwise with hRbm7 to a final Urea concentration of 2 M Urea. Importantly, an excess of hRbm7 was used (~ 1:1.5). The complex was dialyzed in the Buffer for 2.5 h.

3.3.4.5 Dialysis with lowering the salt concentration

Material:

- Dialysis Buffer 2: 500 mM NaCl, 20 mM Tris/HCl pH
7.5, 10% Glycerol, 10 µM ZnSO₄; 2 mM DTT
- Dialysis Buffer 3: 200 mM NaCl, 20 mM Tris/HCl pH
7.5, 10% Glycerol, 10 µM ZnSO₄; 2 mM DTT

The complex was subsequently dialysed against Dialysis Buffer 2 and 3, each for 2.5 h to lower the NaCl concentration.

3.3.4.6 Size exclusion chromatography

Material same as in 3.3.1.7 with the following changes:

- Superdex 200
- (Tris)-SECBuffer, 2mM DTT, 10 μ M ZnSO₄

SEC was used as a final purification step for hZCCHC8. Due to low protein amount an analytical S200 was used.

3.3.5 Protein storage

For short term storage proteins were either kept on ice or at 4 °C in SEC Buffer. For long term storage 50-100 μ l aliquots were flash frozen in N₂ and stored at -80 °C.

3.3.6 Determination of Protein concentration

Protein concentration was determined using absorbance at 280 nm measured by Nanodrop spectrophotometer. Extinction coefficients were calculated with Protparam.

3.3.7 Mass Spectrometry/Peptide mass fingerprint

The total mass of the proteins or identity of fragments after limited proteolysis were determined by Mass Spectrometry Service of the Core Facility of MPIB, Martinsried.

3.3.8 SDS-PAGE

Material

- 1x SDS Running Buffer
- PageRuler (unstained protein ladder)
- heating block (95 °C)
- SDS gel
- SDS-PAGE system
- power supply unit
- ddH₂O
- Coomassie

Reagent	Stacking	Resolving 15 % [ml]	Resolving 12% [ml]
H ₂ O	2.75	2.29	3.29
Tris pH 8.8	-	2.50	2.50
Tris pH 6.8	1.26	-	-
Acrylamid	0.83	5.00	4.00
SDS	0.05	0.1	0.10
APS	0.1	0.1	0.10
TEMED	0.01	0.01	0.01

The SDS-PAGE system was assembled and a 15% or 12% gel was casted. After assembling the gels and filling up the buffer tank with 1x-SDS Running Buffer, the samples were loaded into the gel pockets. The SDS-PAGE was run at 150 V until the running front reached the resolving gel, then run at 300 V until the running front disappeared. In the next step the SDS-PAGE system was disassembled and the stacking gel was removed. The gel was stained in Coomassie for several hours and subsequently washed in ddH₂O.

3.4 Limited Proteolysis

For production of a hRbm7-hZCCHC8 core complex limited proteolysis experiments were performed in analytical and preparative scale.

3.4.1 Analytical scale

A set of different proteases were used for analytical scale reactions: Trypsin, Elastase, Chymotrypsin, GluC and Subtilisin. The proteases were diluted 1:10, 1:100, 1:1000 (Stock 1 mg/ml) in a buffer containing 20 mM Hepes pH 7.5, 50 mM NaCl, 10 mM MgSO₄. 10 μ l of a 1 mg/ml concentrated hRbm7-hZCCHC8 complex were mix with 3 μ l of the respective protease dilution. The final protease concentration were ranging from 23 ng/ μ l, 2.3 ng/ μ l to 0.23 ng/ μ l. The complex without any protease and the proteases alone at the highest concentration were used as a control. After incubation for 30 min on ice, 2x SDS loading Buffer was added and the samples were heated for 5 min at 95°C to stop the reaction. Fragments were separated and analyzed on SDS PAGE.

The ability of the fragments obtained by proteolysis to form a complex was analyzed by size exclusion chromatography (analytical S200 column). Herefore, the set up was scaled up to 500 μ l and after 30 min incubation with the respective protease (Elastase with a final concentration of 23 ng/ μ l), the reaction was stopped with 1 mM AEBSF. After the chromatography run, peak fractions were analyzed on SDS PAGE.

3.4.2 Time course

For time course limited proteolysis set up as described for analytical scale was incubated for different amount of time (30 min-3h). Hence the proteolysis of the complex by the respective protease was monitored over time.

3.4.3 Preparative scale

For preparative scale the hZCCHC8-hRBM7 complex was purified following the protocol described in 3.3.4. After stepwise lowering of the salt concentration, the complex was diluted to a final concentration of 1 mg/ml and incubated with the respective protease dilution (Elastase with a final concentration of 23 ng/ μ l) over night. Reaction was stopped with 1 mM AEBSF. The trimmed complex was separated from fragments by preparative size exclusion chromatography (Hiload S200 column). Peak fractions were analyzed by SDS PAGE.

3.5 Crystallization experiments

For initial screening in-house screens and commercially available screens were set up. 96 well plates were pipetted using a Phoenix pipetting robot and screened at 4 °C and 18 °C for two different protein concentrations. Each drop contained 0.1 μ l of the protein combined with 0.1 μ l reservoir solution. For visualisation a XtalFocus robot was used. All screenings and visualisations were carried out by the Crystallisation Facility of the MPIB, Martinsried.

4 Results & Discussion

4.1 NEXT domain organisation

The NEXT complex consist of three proteins, namely hMtr4, an ATP-dependent RNA-Helicase and the metazoan-specific proteins hRbm7 and hZCCHC8. Since there is no structural information for the three proteins, the domain organization was predicted using Pfam and Phyre tools (Finn et al., 2014; Kelley and Sternberg, 2009). hZCCHC8 is predicted to contain a zinc knuckle of the type CCHC together with a proline rich domain (Fig. 4.1). hRbm7 is a small protein of 266 residues featuring a RRM (RNA recognition motif) domain, which is the most abundant RNA binding motif (Fig. 4.1). Sequence analysis revealed that hMtr4 is similar to yMtr4, therefore domain organization of hMtr4 was modeled based on yeast Mtr4 structure using Phyre (Kelley and Sternberg, 2009; Weir et al., 2010). It features RecA-1 and RecA-2 domains (shown in red) and a helical domain (green), which as shown in the yeast structure together form the helicase core. The Stalk KOW domain is inserted into the helicase core, therefore here referred to as SK insert (blue) (Weir et al., 2010).

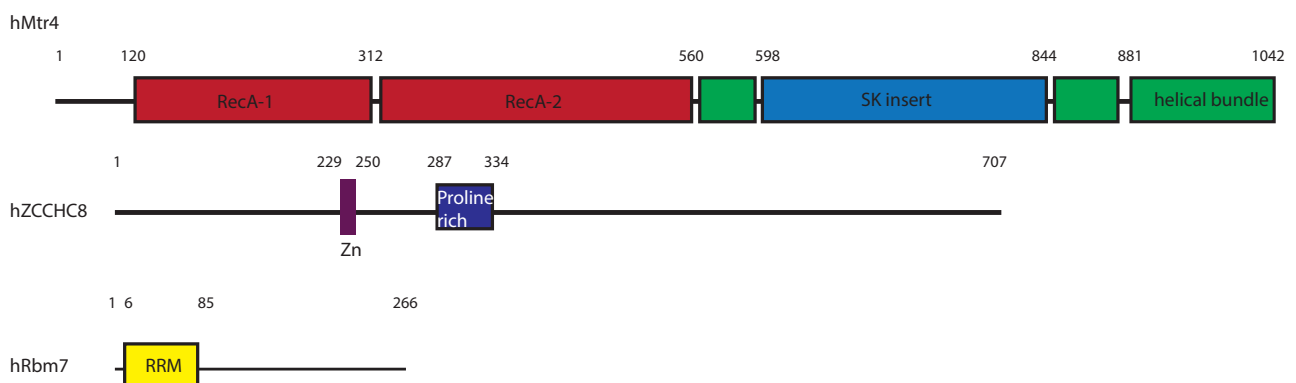


Figure 4.1: Domain organization of NEXT complex.

Schematic representation of the domain architecture in NEXT proteins. Low complexity regions are indicated with lines. hMtr4 features RecA-1, RecA-2 (red) and a helical domain (green). The hZCCHC8 protein contains a Zn knuckle of the ZCCHC8 type and additionally a proline rich domain. The hRbm7 is a small protein, which contains a RRM domain.

4.2 NEXT construct design

Biochemical *in vitro* studies and crystallization experiments rely on the abundance of the purified recombinant protein of interest. A number of NEXT constructs were already available in the beginning of the project and shall be briefly introduced in Fig. 4.2.

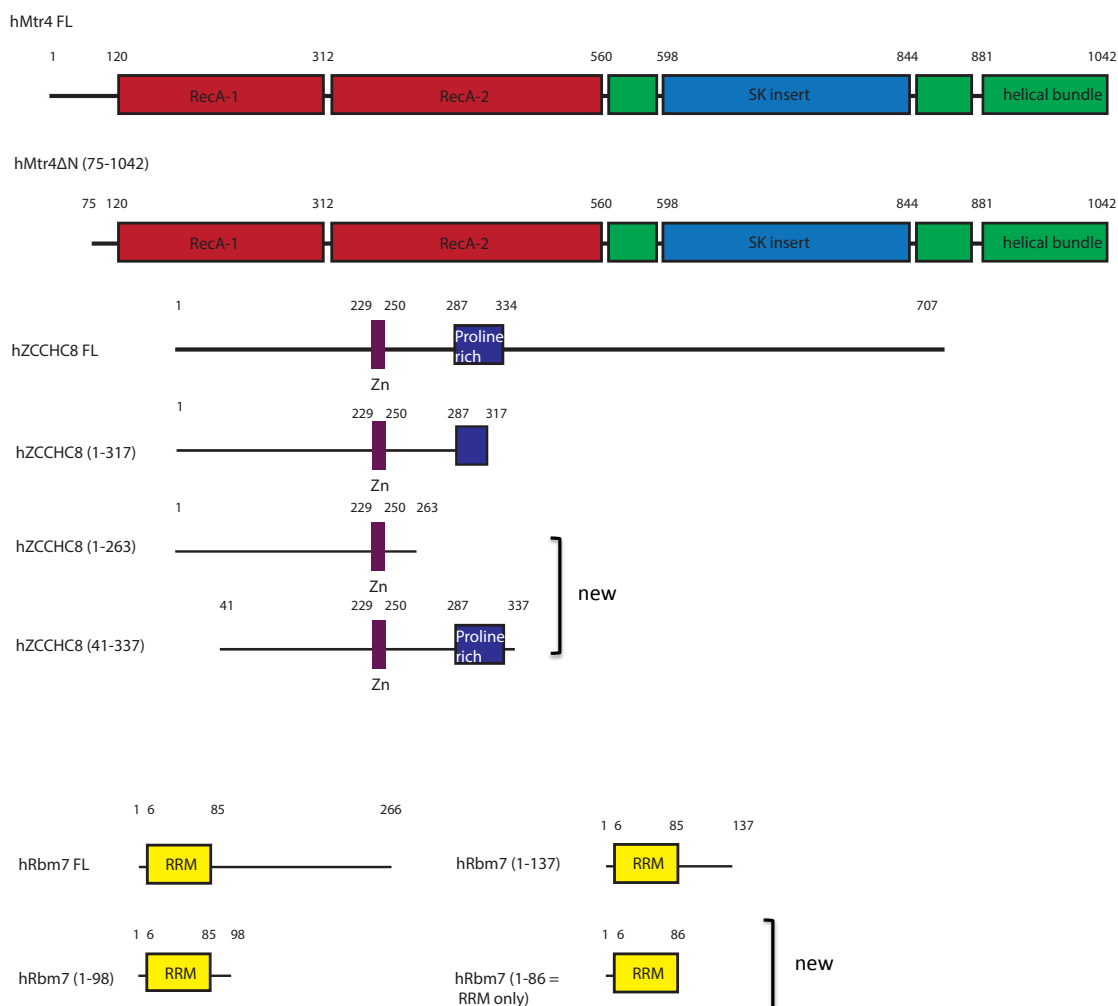


Figure 4.2: Initial working and new NEXT constructs.

Full length constructs of hMtr4, hZCCHC8 and hRbm7 as well as short constructs were available in the beginning of the project and the expression has already been tested in small scale (not shown). New constructs were designed as part of the thesis (new). FL = full length.

Expression tests (by Sebastian Falk and Ariane Fischer, MPI Biochemistry, Martinsried) showed that GST-tagged FL (full length) and short construct of hZCCHC8 as well as GST-tagged FL hMtr4 and Z-tagged hMtr4 Δ N expressed well and are soluble (data not shown). In contrast His-tagged hRbm7 FL showed low expression. Hence, testing new hRbm7 constructs was one of the tasks at the beginning of the project. A short hRbm7 construct encompassing the residues 1-137 showed good expression and purification was established by Sebastian Falk (data not shown).

The design of new NEXT constructs (in particular new hRbm7 and hZCCHC8 constructs) was based on bioinformatics and partially limited proteolysis results shown in Results section 4.7. Highly conserved regions in the respective proteins might represent regions important for interactions between the proteins or for the function and were therefore kept in constructs. hZCCHC8 is highly conserved around the Zn knuckle and the proline rich domain in metazoa (Fig. S1). Additionally, it is predicted to be highly disordered between residues 340-660 (Fig. S2). Reasoning that the last C-terminal helices probably do not have a function, hZCCHC8 constructs were initially designed as C-terminal truncations. Based on limited-proteolysis results described in chapter 4.7, the constructs were additionally N-terminally truncated. Fig. 4.2 shows all new designed constructs, which form the major body of this thesis. A list of all constructs that have been successfully cloned can be found in the Materials section.

Secondary structure prediction of hRbm7 revealed that the C-terminus of hRbm7 is highly disordered (Fig. 4.3 & S3), therefore all short constructs were designed as C-terminal deletions. The protein is highly conserved around the RRM domain in metazoa as shown in Fig. S4, therefore the shortest designed construct was the RRM-only construct, encompassing the residues 1-86. Fig. 4.3 shows a conservation analysis of the hRbm7 RRM domain. Except residue 2 of RNP1 all residues correspond to the conserved motifs of RNP1 and RNP2: [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY] and [ILV]-[FY]-[ILV]-X-N-L (X = any amino acid), respectively. The canonical RRM $\beta\alpha\beta\beta\alpha\beta$ fold was annotated based on secondary structure prediction.

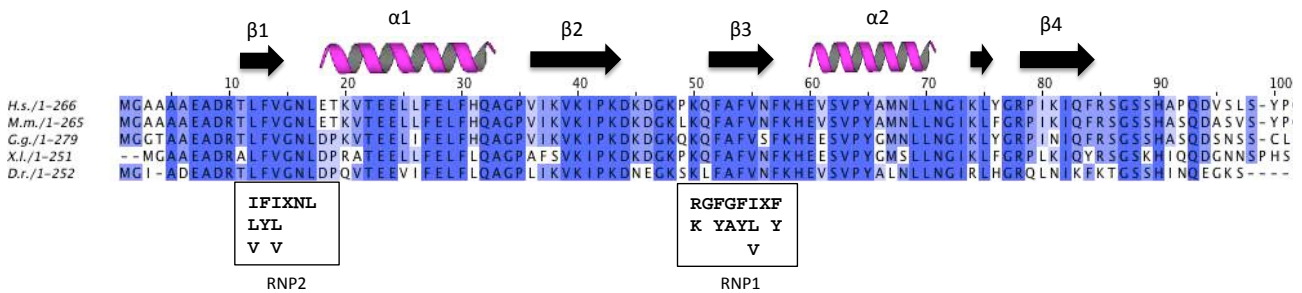


Figure 4.3: The RRM domain in hRbm7 is highly conserved.

Alignment of the RRM domain of Rbm7 from different species. The conserved RNP1 and RNP2 motifs are shown in black boxes (X = any amino acid). β -sheets and α -helices forming the canonical RRM fold were annotated based on secondary structure prediction. All residues except for residue 2 in RNP1 correspond to the conserved RNP1 and RNP2 motifs. H.s. = Homo sapiens. M. m. = Mus musculus. G. g. = Gallus gallus. X.l. = Xenopus laevis. D.r. = Danio rerio.

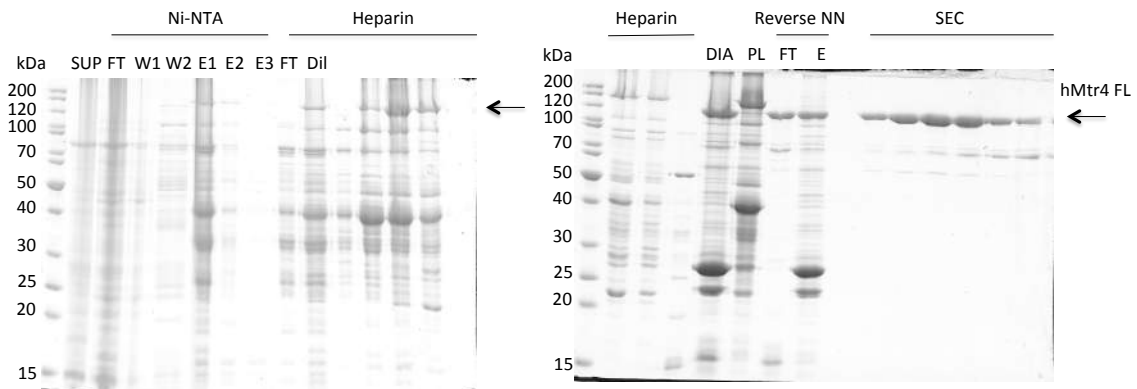
4.3 Expression/Purification of Mtr4 constructs

Expression of hMtr4 constructs was tested by Sebastian Falk (MPI Biochemistry, Martinsried). Good expression and solubility was observed with His-GST- and His-Z-tag. All Mtr4 constructs were expressed in BL21(DE3)Star pRare cells. Purification followed the protocol described in Methods section 3.3.1.

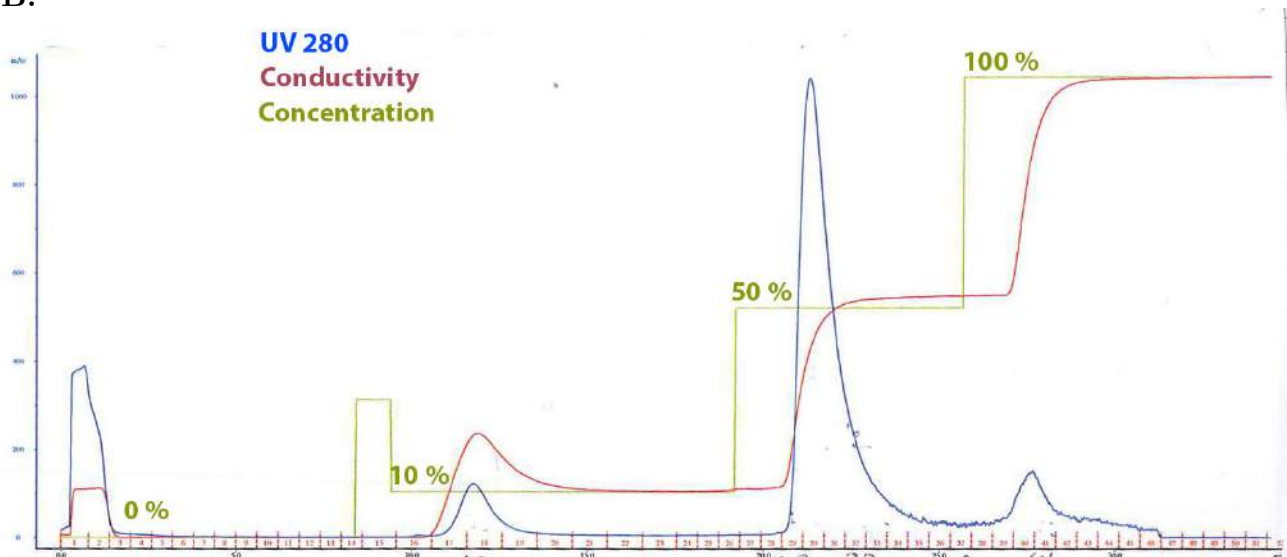
4.3.1 Purification of hMtr4 (Full length)

The FL (full length) hMtr4 is only weakly expressed (not shown). Following cell lysis and centrifugation, the protein was subjected to Ni-Nitrilotriacetate-affinity (Ni-NTA) purification. The protein bound completely to the Nickel column matrix via the His-GST-tag as it was not observed in flow-through fraction (Fig. 4.4 A lane: FT). After several wash steps (Fig. 4.4 A lanes: W1-W2) the 145 kDa protein was eluted in three fractions (Fig. 4.4 A lanes: E1-E3) through the elution buffer component imidazole. An excess of imidazole passed through the column, displaces the His-tag from Ni²⁺ co-ordination, freeing the His-tagged protein. The eluted protein showed strong degradation and was additionally highly contaminated. Despite a 1 M NaCl wash the eluate still contained nucleic acid contamination as indicated by a ratio of $A_{260}/A_{280} > 1.2$ ($\sim 10\%$ nucleic acid). Therefore, it was subsequently purified by Heparin affinity purification. Heparin has a negative charge and a structure, which mimics nucleic acids in their overall binding properties. hMtr4 was loaded on the Heparin column; due to the interaction with Heparin it could not be observed in the flow-through fraction (Fig. 4.4 A lanes: HeparinFT). Absorption at 280 nm, conductivity and concentration were monitored during the run (Fig. 4.4 B). Subsequently, the column was washed with Heparin A Buffer and 10% Heparin B Buffer. Protein that was still contaminated with nucleic acid interacted weakly with the column resin and was eluted at 235 mM NaCl (10% Buffer B). The nucleic-acid free protein eluted at 575 mM NaCl (50% Heparin Buffer B) and the peak in Fig. 4.4 B was analyzed on SDS PAGE (Fig. 4.4 A lanes: Heparin). In the next step the protein was dialysed over night. Since the N-terminal His-GST-tag is followed by a 3C protease site, the tag was cleaved off by 3C proteolysis during dialysis. Subsequently, the protein was subjected to a reverse-Ni-NTA column to remove the tag, the His-tagged protease and the non-cleaved fusion protein. Without the His-GST-tag the 118 kDa protein was partially observed in the flow-through (Fig. 4.4 A lane: Reverse NN-FT) as expected. However, partially the protein interacted non-specifically with the column matrix and was therefore also observed in the elution fraction (Fig. 4.4 A lanes: Reverse NN E). In the last step, the protein was purified by size exclusion chromatography on a Hiload S200 column (Fig. 4.4 C). The size exclusion chromatography profile showed a prominent peak, which was analyzed on SDS PAGE (Fig. 4.4 A lanes: SEC). The peak fractions contained the reasonably pure protein ($A_{260}/A_{280} \sim 0.6$) and were therefore pooled, concentrated and stored at -80°C . Due to the low expression the protein yield after the purification was only $A_{260}/A_{280} \sim 1.5$ mg for a 4 l expression culture.

A.



B.



C.

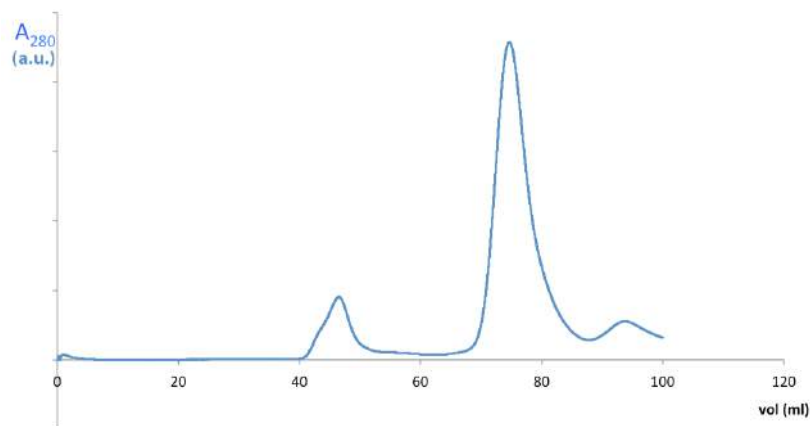


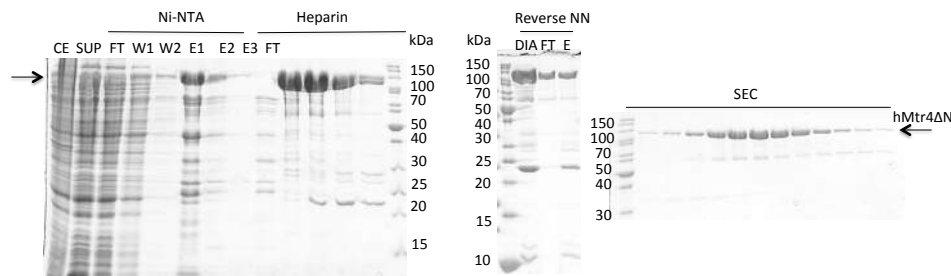
Figure 4.4: Purification of Mtr4 (Full length).

A. SDS PAGE. To analyse the progress of the purification samples were taken after every purification step. SUP = supernatant. FT = flow-through. W = wash. E = Elution. Dil = diluted sample. PL = sample for pulldown. NN = Nickel-NTA. SEC = size exclusion chromatography. His-GST-hMtr4 FL and hMtr4 FL are indicated by black arrows. **B.** Heparin Chromatogram: Absorption at 280 nm, conductivity and concentration were monitored during the run. The column was washed with Heparin A and 10% Heparin B and the protein eluted at 50% Heparin B Buffer. **C.** SEC profile: Elution of hMtr4 FL from the SEC column was followed by monitoring absorption at 280 nm. A prominent peak was observed at ~ 75 ml.

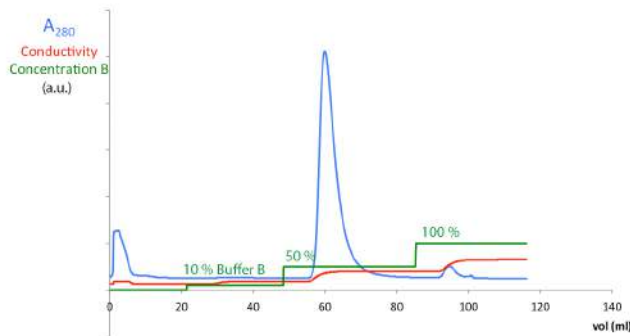
4.3.2 Purification of Mtr4 Δ N

A higher protein expression was observed for the construct of Mtr4 lacking the first 74 N-terminal residues (Fig. 4.5 A lane: CE). This construct has a N-terminal His-Z-tag followed by 3C protease cleavage site. The purification followed the same protocol as described for the full length Mtr4 (Results section 4.3.1). Following cell lysis and centrifugation the soluble 120 kDa protein was purified via a Ni-NTA column, where it was eluted from by an excess of imidazole (Fig. 4.5 A lanes: E1-E3). As can be seen on the SDS PAGE analysis the protein is reasonably pure even after the first purification step and in contrast to the full-length hMtr4 only few degradation bands are present. Subsequently, nucleic acid contaminations were removed with a Heparin column. The protein was eluted with 235 mM NaCl (50% Heparin B Buffer), which resulted in a prominent peak (Fig. 4.5 B). The presence of Mtr4 Δ N in the peak was confirmed by SDS PAGE analysis (Fig. 4.5 A lanes: Heparin). After His-Z-tag removal by 3C protease cleavage during dialysis over night the protein was further purified by reverse-Ni-NTA purification. Without the tag the majority of the protein did not bind to the column matrix (Fig. 4.5 A lane: Reverse NN-FT). Finally, the protein was subjected to size exclusion chromatography, which showed a prominent peak at \sim 80 ml (Fig. 4.5 C). As confirmed by SDS PAGE analysis (Fig. 4.5 A lanes: SEC) the peak contained the reasonably pure 110 kDa protein ($A_{260}/A_{280} \sim 0.6$). The total protein yield amounted to approximately 7 mg per 4 l expression culture.

A.



B.



C.

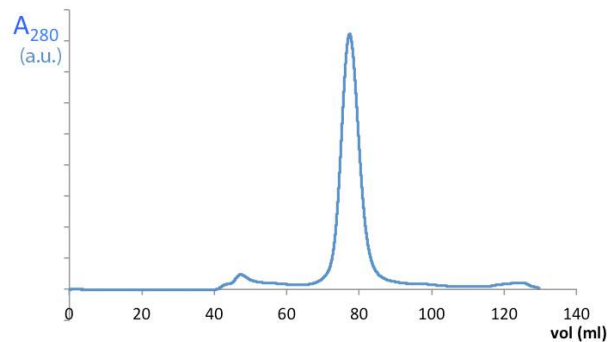


Figure 4.5: Purification of Mtr4 Δ N.

A. SDS PAGE. To analyse the progress of the purification samples were taken after every purification step. CE = cell extract. SUP = supernatant. FT = flow-through. W = wash. E = Elution. Dil = diluted sample. PL = sample for pulldown. NN = Nickel-NTA. SEC = size exclusion chromatography. His-Z-hMtr4 Δ N and hMtr4 Δ N are indicated by black arrows. **B.** Heparin Chromatogram: Absorption at 280 nm, conductivity and concentration were monitored during the run. The column was washed with Heparin A and 10% Heparin B and the protein eluted at 50% Heparin B Buffer. **C.** SEC profile: Elution of hMtr4 Δ N from the SEC column was followed by monitoring absorption at 280 nm. A prominent peak was observed at ~ 80 ml.

4.4 Expression/Purification of hRbm7 constructs

While expression of one short construct of hRbm7 was already tested and established in the beginning of this project, the expression tests of the His-tagged FL (full length) protein failed to give any satisfying results in any of the tested *E. coli* strains. Therefore, different constructs of the full length were designed, differing in the used tag: (1) N-terminal His-TRX-tag, (2) N-terminal His-GST-tag, (3) N-terminal His-Sumo-tag, (4) N-terminal His-MBP-tag, (5) N-terminal CBP- and C-terminal His-tag. Expression was tested in BL21(DE3)Star pRare. FL hRbm7 was found to be well expressed with all the five different fusion proteins (TRX, GST, SUMO, MBP and CBP/C-His) as a distinct band is visible in the crude extract (Fig. 4.6 CE). However, soluble protein is only obtained as a fusion with MBP, as after centrifugation it is observed in the supernatant of the His-MBP lane of Fig. 4.6. Subsequently, the protein was treated with the respective protease for removal of the tag.

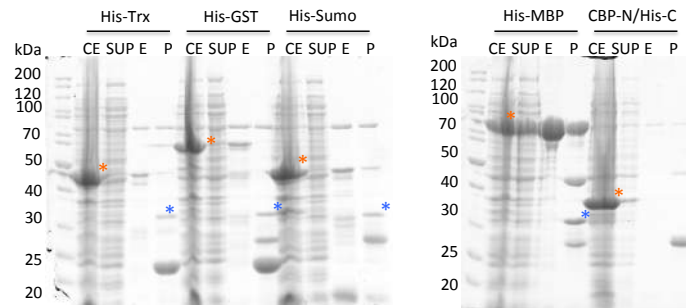


Figure 4.6: Expression tests of full-length hRbm7 with different fusion tags.

Small-scale Ni-purification to assess the expression and solubility of the full length hRbm7 with different fusion proteins. Samples were withdrawn after lysis (CE), centrifugation (SUP), Elution from Ni-NTA (E) and after protease cleavage (P) and analysed by SDS PAGE. hRbm7 FL + tag is indicated by orange stars. hRbm7 FL is highlighted by blue stars.

4.4.1 Purification of full length hRbm7

Based on the results described in the previous section the full length hRbm7 protein was purified with a His-MPB-tag. The 74 kDa protein was well expressed and soluble (Fig. 4.7 lanes CE, SUP). The protein was first purified using a Ni-NTA column. Unexpectedly, it was partially found in flow-through of the column, presumably due to a high amount of protein. The eluted protein of 74 kDa showed a specific degradation/contamination band at 60 kDa but otherwise was reasonable pure (Fig. 4.7 E1-E3). After the Ni-NTA purification the protein was dialysed over night with a simultaneous tag cleavage by 3C protease. However, once the tag was cleaved the protein precipitated. To confirm the presence of the full length protein in the precipitate, a homogenized sample of mixed precipitate and supernatant was analyzed on a SDS PAGE gel and compared to the supernatant fraction (Fig. 4.7 HOM, SUP). The additional band in the homogenized fraction corresponded to the 31 kDa full length hRbm7 protein. Different attempts to improve the solubility were not successful.

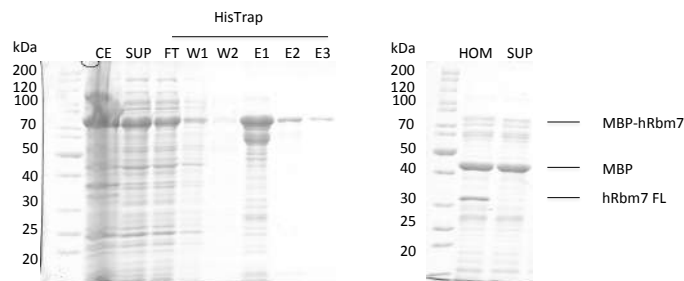


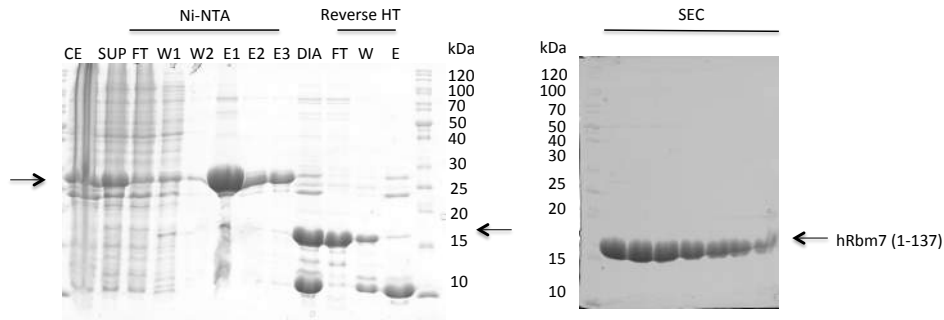
Figure 4.7: Purification of full length hRbm7.

To analyse the progress of the purification samples for SDS PAGE analysis were taken after lysis (CE), centrifugation (SUP) and through-out the Ni-NTA purification. W1-W2 = wash steps 1 and 2. E1-3 = elution fractions. The protein precipitated after tag cleavage. Precipitant was homogenized with the supernatant and analysed on the gel (HOM).

4.4.2 Purification of the short hRbm7 construct (1-137) and RRM-only construct (1-86)

Two short constructs of hRbm7 with a His-Z-tag followed by a 3C site were available and were purified after the same protocol as the full length hRbm7. Purification of the hRbm7 (1-137) construct was already established by Sebastian Falk (MPI Biochemistry, Martinsried). Both constructs were well expressed and were soluble (Fig. 4.8 & 4.9 A lanes: CE, SUP).

A.



B.

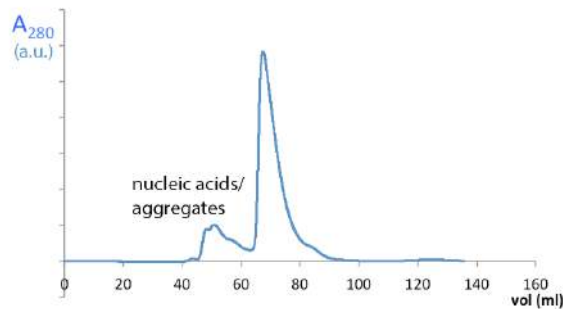


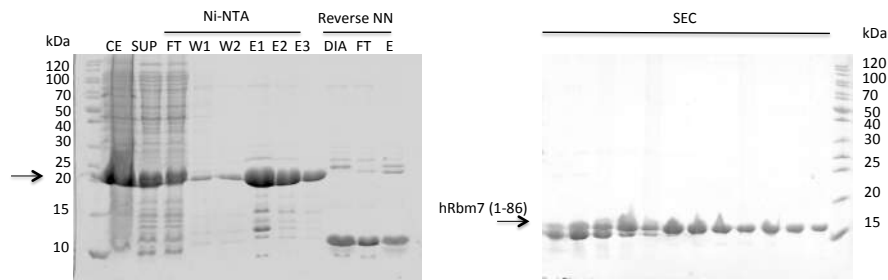
Figure 4.8: Purification of (1-137) construct of hRbm7.

A. SDS PAGE. To analyse the progress of the purification samples were taken after every purification step. CE = cell extract. SUP = supernatant. FT = flow-through. W = wash. E = Elution. PL = sample for pulldown. NN = Nickel-NTA. SEC = size exclusion chromatography. His-Z-hRbm7 (1-137) and hRbm7 (1-137) are indicated by black arrows. **B.** SEC profile: Elution of hRbm7 (1-137) from the SEC column was followed by monitoring absorption at 280 nm. The chromatogram showed a small void peak followed by a prominent peak at ~ 75 ml.

The proteins were first purified via a Ni-NTA column. Both proteins were partially found in the flow-through fraction (Fig. 4.8 & 4.9 A lane: FT), presumably due to high amount of protein. A large amount of protein was eluted in both cases by an excess of imidazole (Fig. 4.8 & 4.9 lanes: E1-E3). In both cases the eluted protein fractions showed slight contaminations or few degradation bands. During dialysis the His-Z-tag was removed by cleavage with a 3C protease (Fig. 4.8 & 4.9 A lane: DIA) and the proteins were subjected to reverse Ni-NTA purification. Without the tag both proteins were found in the flow-through as expected (Fig. 4.8 & 4.9 A lanes: Reverse NN-FT). Finally both proteins were subjected to size exclusion

chromatography on a Hiload S75 column. The hRbm7 (1-137) construct eluted at ~ 75 ml preceded by a small peak, which presumably contained nucleic acids/aggregates (Fig. 4.8 B). The presence of the 17 kDa protein in the prominent peak was confirmed by SDS PAGE (Fig. 4.8 A: SEC) The measured A_{260}/A_{280} ratio of ~ 0.6 indicated a nucleic-acid free pure protein. The peak fractions with the protein were pooled, concentrated and stored at -80°C . The RRM-only construct (1-86) eluted at ~ 90 ml (Fig. 4.9 B second peak), while the first peak presumably contained nucleic acids or aggregates. Fractions not contaminated with the His-Z-tag (Fig. 4.9 A: SEC. lower band) were pooled, concentrated and stored at -80°C . The measured A_{260}/A_{280} ratio amounted to 0.9, which indicates a remaining nucleic acid contamination. Accordingly, the nucleic acid/aggregates void peak was higher for the RRM-only construct compared to the other short construct. The protein yield of a 4 l culture was ~ 40 mg for hRbm7 (1-137) and about 20 mg for hRbm7 (1-86).

A.



B.

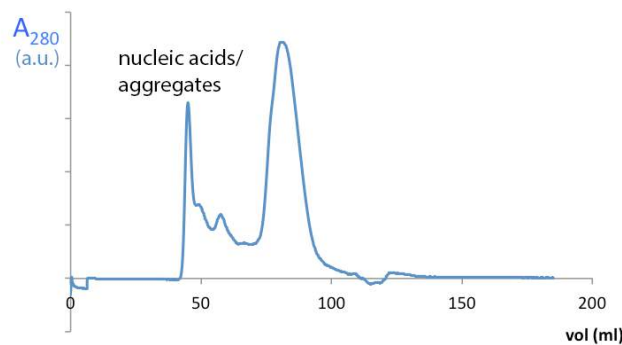


Figure 4.9: Purification of RRM-only hRbm7 construct (1-86).

A. SDS PAGE. To analyse the progress of the purification samples were taken after every purification step. CE = cell extract. SUP = supernatant. FT = flow-through. W = wash. E = Elution. PL = sample for pulldown. NN = Nickel-NTA. SEC = size exclusion chromatography. His-Z-hRbm7 (1-86) and hRbm7 (1-86) are indicated by black arrows. **B.** SEC profile: Elution of hRbm7 (1-86) from the SEC column was followed by monitoring absorption at 280 nm. The chromatogram showed a large void peak followed by a prominent peak at ~ 90 ml.

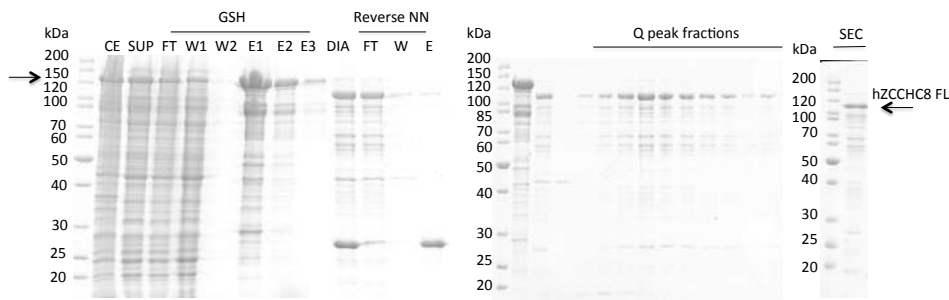
4.5 Purification of hZCCHC8 constructs

Expression of initial hZCCHC8 constructs was tested by Sebastian Falk (MPI Biochemistry, Martinsried.) Satisfying expression and solubility was observed using a His-GST-tag. All hZCCHC8 constructs were expressed in BL21(DE3)Star pRare cells.

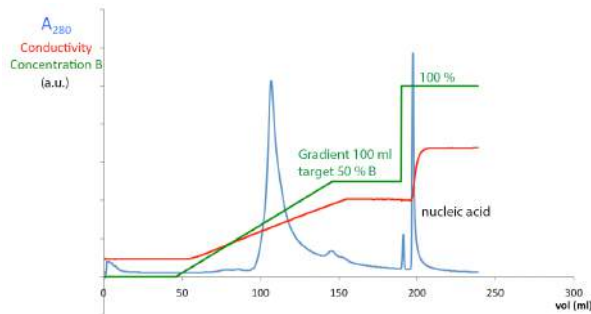
4.5.1 hZCCHC8 Full length protein is prone to aggregation

Purification of the FL (full length) hZCCHC8 fused to His-GST was initially performed following the protocol described in 3.3.3. Following cell lysis and centrifugation, the 107 kDa protein was primarily purified via a GSH-column utilizing the GST-tag (Fig. 4.10 A lanes: GST; the protein migrates slower on the SDS PAGE). Presence of 30 mM Glutathione in the elution buffer displaced the bound protein from the column. The eluted protein was heavily degraded as can be seen in Fig. 4.10 A lane: GSH-E1-E3. Additionally, the eluate was highly contaminated with nucleic acids as indicated by the high A_{260}/A_{280} ratio > 1.2 ($\sim 10\%$ nucleic acid). The protein was dialysed over night with a simultaneous tag cleavage by 3C protease (Fig. 4.10 A lane: DIA). Subsequently, the untagged protein was further purified by a reverse-Ni-NTA purification. As expected without the tag the protein was observed in flow-through fraction (Fig. 4.10 A lane: Reverse NN-FT) whereas the His-GST-tag eluted only in presence of imidazole (Fig. 4.10 A lane: Reverse NN-E). Without the tag the protein has a weight of 79 kDa, however, it migrates slower on the gel (~ 100 kDa). To remove contaminating nucleic acids, that were present after the first purification step, an ion exchange column was chosen for further purification. As the hZCCHC8 protein has a pI of 4.79 a Q column was reasoned to be the most suitable. After loading the protein on the Q column a gradient was started with a target of 50% Buffer B (500 mM NaCl) in 100 ml. Absorption at 280 nm, conductivity and concentration were monitored during the run (Fig. 4.10 B). A prominent peak could be observed at $\sim 34\%$ Buffer B (~ 400 mM NaCl), which contained the hZCCHC8 protein as confirmed by the corresponding SDS PAGE analysis (Fig. 4.10 A lanes: Q peak fractions). Nucleic acids eluted at 100% Buffer B. Finally, the protein was subjected to SEC. An analytical S200 column was used due to a low amount of protein. Absorption at 280 nm showed a peak at approximately 8 ml, which corresponds to the void volume of the respective column. This indicated aggregation of hZCCHC8. The presence of hZCCHC8 in the void peak was confirmed by SDS PAGE (Fig. 4.10 A lane: SEC).

A.



B.



C.

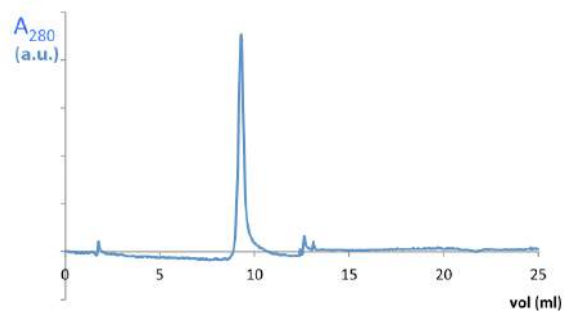


Figure 4.10: Purification of hZCCHC8 (FL) - initial protocol.

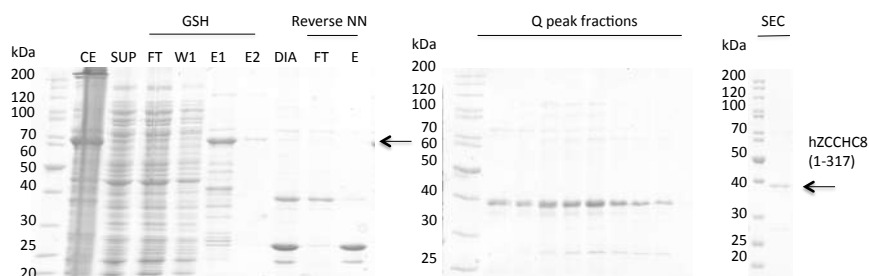
A. SDS PAGE. To analyse the progress of the purification samples were taken after every purification step. CE = cell extract. SUP = supernatant. FT = flow-through. W = wash. E = Elution. Dil = diluted sample. NN = Nickel-NTA. SEC = size exclusion chromatography. His-GST-hZCCHC8 FL and hZCCHC8 FL are indicated by black arrows. **B.** Q-Ion Exchange Chromatogram: Absorption at 280 nm, conductivity and concentration were monitored during the run. The protein was eluted by a linear gradient with a target of 50% Buffer B. **C.** SEC profile: Elution of hZCCHC8 from the SEC column was followed by monitoring absorption at 280 nm. A void peak was observed at ~ 8 ml.

4.5.2 The short construct (1-317) of the hZCCHC8 protein is also prone to aggregation

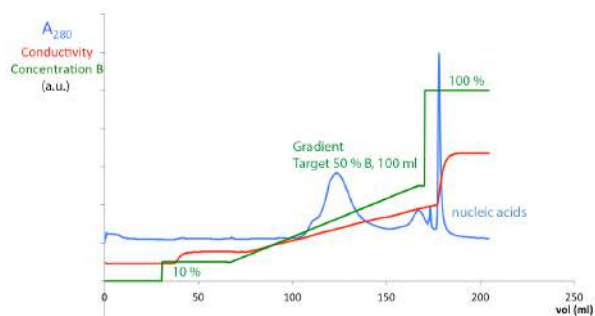
To test whether deletion of the disordered C-terminal residues would improve the behavior of the protein, a short construct of hZCCHC8 (1-317) was purified following the same protocol. Strikingly, the short construct was found to be less soluble as can be seen in Fig. 4.11 A. While there is a prominent band in the cell extract (CE), there is only a faint band in the supernatant (SUP). The protein eluted from the GSH column showed less degradation than the full length protein and after tag cleavage by 3C proteolysis, the 37 kDa-protein was present in the flow-through of the reverse Ni-NTA purification as expected (Fig. 4.11 A lane: Reverse NN-FT). Due to high nucleic acid contamination ($A_{260}/A_{280} > 1.2$) the protein was further purified by a Q-column (Fig. 4.11 B). After a 10% Buffer B wash a gradient was started with a target of 50% B (500 mM NaCl) in 100 ml. Absorption at 280 nm, conductivity and concentration were monitored during the run. The protein eluted at approximately 34% Buffer B (~ 400 mM NaCl) as confirmed by SDS PAGE analysis (Fig. 4.11 A lanes: Q peak fractions). Nucleic acids

eluted at 100% Buffer B. Finally, the protein was subjected to SEC (Fig. 4.11 C). Due to low amounts of protein an analytical S200 column was used. As observed for the full length protein the short construct eluted in void volume of the column indicating an aggregation. Conclusively, the deletion of disordered C-terminal residues did not render the protein soluble.

A.



B.



C.

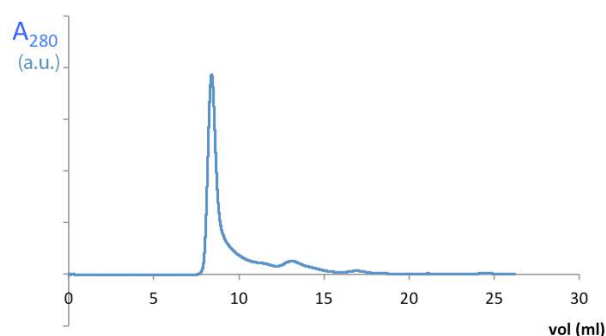


Figure 4.11: Purification of hZCCHC8 short construct (1-317) - initial protocol.

A. SDS PAGE. To analyse the progress of the purification samples were taken after every purification step. CE = cell extract. SUP = supernatant. FT = flow-through. W = wash. E = Elution. Dil = diluted sample. NN = Nickel-NTA. SEC = size exclusion chromatography. His-GST-hZCCHC8 (1-317) and hZCCHC8 (1-317) are indicated by black arrows. **B.** Q-Ion Exchange Chromatogram: Absorption at 280 nm, conductivity and concentration were monitored during the run. The protein was eluted by a linear gradient with a target of 50% Buffer B. **C.** SEC profile: Elution of hZCCHC8 from the SEC column was followed by monitoring absorption at 280 nm. A void peak was observed at ~ 8 ml.

4.5.3 Optimization of the hZCCHC8 purification protocol

Different protocol changes were assayed to tackle the aggregation problem of the hZCCHC8 protein. This included e.g. testing of several additives as CHAPS, Arg or Urea, co-lysis with an interaction partner or the addition of Zn^{2+} since hZCCHC8 contains one Zn knuckle. Optimization attempts are summarized in a table below.

Table 4.1: Protocol optimization attempts for the purification of hZCCHC8.

S6 Size exclusion column (instead of S200)
Varying the salt concentration in the SEC-Buffer
No EDTA in buffers; supplying every buffer and the bacterial expression culture with ZnSO ₄
2-Mercaptoethanol as reducing agent instead of DTT
Co-lysis with full-length Mtr4
Co-lysis with short construct (1-137) of hRbm7
Co-lysis (short construct (1-317) of hZCCHC8) with full length hRbm7
Co-lysis (short construct (1-317) of hZCCHC8) with short construct (1-137) hRbm7
Adding 0.5 % CHAPS in lysis buffer, 10 mM MgSO ₄ in every Buffer
Using a different concentrator (different membrane for the concentration)
Concentrating the protein with 0.5% CHAPS/ 200 mM Glycine/ 200 mM Arginine/ 0.5 M Urea in buffer

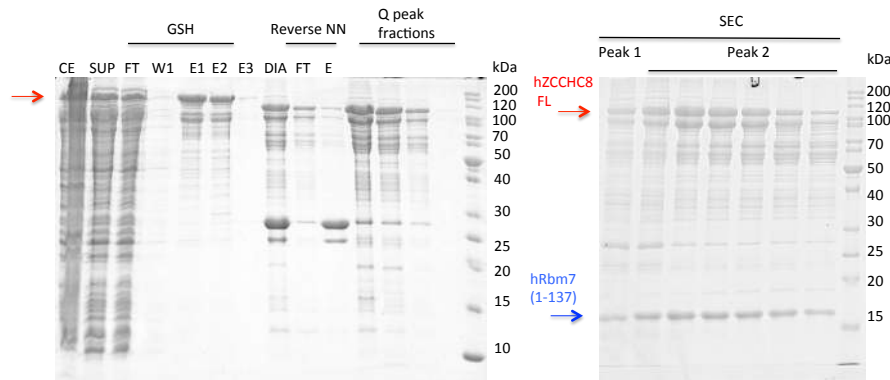
Despite none of the above listed attempts helped to solve the aggregation problems, some of them were found useful and therefore were established in the optimized protocol. Adding 0.5% CHAPS in the lysis buffer slightly improved the solubility of the hZCCHC8 (1-317) short construct (not shown). Supplying buffers and bacterial expression culture with ZnSO₄ was rationalized as prerequisite for purification of a Zn knuckle protein.

The first successful optimization attempt was achieved by unfolding and refolding the hZCCHC8 protein in presence of its binding partner hRbm7 (see Results section 4.6 for interaction studies). After the Q-column the hZCCHC8 protein was dialyzed against 8 M Urea-1 M NaCl over night. In the next step the unfolded or partially unfolded protein was diluted in a 1:1 ratio with hRbm7 and subjected to several dialysis steps, in which the salt concentration was decreased gradually. The final optimized protocol is described in Methods section 3.3.4. The following subsection shows purification of different hZCCHC8 constructs with the applied protocol changes.

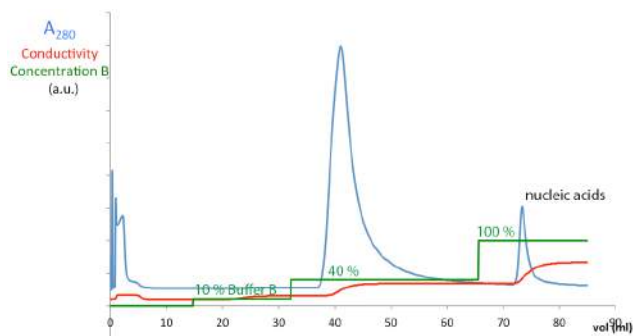
4.5.4 Purification of hZCCHC8 full length - optimized protocol

The first steps of the hZCCHC8 purification remained the same in both protocols. After cell lysis and centrifugation, the protein was first purified over a GSH column (Fig. 4.12 A GSH).

A.



B.



C.

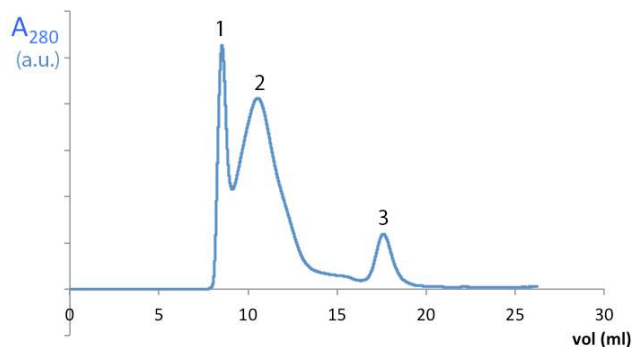


Figure 4.12: Purification of hZCCHC8 full length - optimized protocol.

A. SDS PAGE. To analyse the progress of the purification samples were taken after every purification step. CE = cell extract. SUP = supernatant. FT = flow-through. W = wash. E = Elution. NN = Nickel-NTA. SEC = size exclusion chromatography. His-GST-hZCCHC8 FL and hZCCHC8 FL are indicated by red arrows. hRbm7 (1-137) is indicated by the blue arrow. **B.** Q-Anion Exchange Chromatogram: Absorption at 280 nm, conductivity and concentration were monitored during the run. The protein was eluted at 40% Buffer B. **C.** SEC profile: Elution of hZCCHC8 from the SEC column was followed by monitoring absorption at 280 nm. Two peaks were observed: the first at ~ 8 ml, the second at ~ 11 ml.

During dialysis over night the GST tag was removed via 3C protease cleavage and the protein was subsequently further purified by a reverse-Ni-NTA column. The Ni-NTA column was connected to a Q-column. The untagged protein did not bind to the Ni-NTA column but bound to the following Q-column (Fig. 4.12 A lane: Reverse HT-NN-FT: only a weak band is visible corresponding to the small portion of protein, which did not bind to the Q column). An excess of imidazole eluted the GST-tag from the Ni-NTA column (Fig. 4.12 A Reverse NN-E). Since previous Q-Anion exchange runs with linear gradients showed that the hZCCHC8 protein eluted at approximately 34% Buffer B, the subsequent purification were done with a step-wise increase of Buffer B concentration to 40% (460 mM NaCl), at which the protein elutes (Fig. 4.12 B). The eluted peak contained the hZCCHC8 full length protein as shown on the SDS PAGE gel (Fig. 4.12 A lanes: Q peak fractions). After the Q-column the protein was dialysed over night against a buffer containing 8 M Urea and 1 M NaCl inducing unfolding of the protein. Subsequently, the protein was carefully and drop-wise mixed with its binding partner hRbm7 (short construct

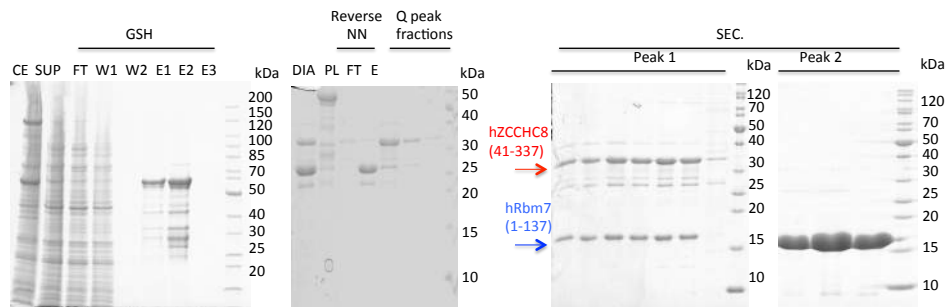
1-137) to a final Urea dilution of 2 M. In this step a refolding of hZCCHC8 is expected to occur. Finally, the complex was subjected to several dialysis series where the NaCl concentration was lowered stepwise to a final concentration of 200 mM NaCl. After concentration, the protein complex was subjected to size exclusion chromatography. Due to low protein amounts an analytical S200 column was used. The chromatogram showed three peaks (Fig. 4.12), the first in void volume (1) and the second between 10-12 ml corresponding to a presumably oligomeric stage of the complex (2). The third peak (3) contains hRbm7 (1-137) as known from previous studies. SDS PAGE analysis confirmed the presence of hZCCHC8 and hRbm7 in peaks 1-2 (Fig. 4.12 A lanes: SEC.). The void peak is larger implying that the majority of the protein is still aggregating, however, a significant part of the complex is not present in the void volume. Better results with the optimized protocol were achieved for a short construct of hZCCH8 protein as shown in the next subchapter.

4.5.5 Purification of a short construct (41-337) of hZCCHC8 - optimized protocol

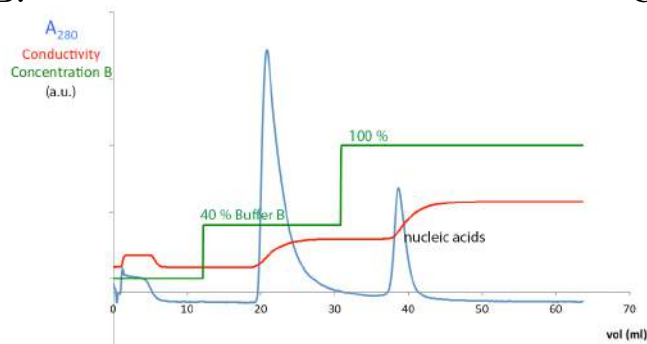
The shorter hZCCHC8 construct encompassing residues 41-337 was purified using the same optimized protocol as for the full length protein. While there is a prominent hZCCHC8 band in cell extract fraction on the SDS PAGE analysis (Fig. 4.13 A lane: CE), there is only a very weak band in the supernatant fraction after centrifugation (Fig. 4.13 A lane: SUP). This indicates that the short construct is highly insoluble even in presence of 0.5% CHAPS. After elution from the GSH column the 63 kDa protein showed some degradation bands (Fig. 4.13 A lanes: E1-E2). After tag cleavage the sample was subjected to a Ni-NTA-column connected to a Q-column. The untagged protein should not bind to the Ni-column but to the Q-column. As expected the hZCCHC8 was not found in the flow-through of the Ni- and Q-column (Fig. 4.13 A lane: Reverse NN-FT). The His-GST-tag bound to the Ni-column was eluted by an excess of imidazole (Fig. 4.13 A lane: Reverse NN-E). The protein eluted from the Q-column at 40% Buffer B (460 mM NaCl) (Fig. 4.13 B) while nucleic acids eluted at 100% Buffer B. To unfold the protein it was dialysed against a buffer containing 8 M Urea and 1 M NaCl over night. Refolding occurred by a drop-wise mixing of hZCCHC8 protein to its binding partner hRbm7 (1-137) (final urea concentration of 2 M). Subsequently, the complex was dialysed against a series of buffers, where the NaCl concentration was step-wise lowered. After concentration the complex was further purified on size exclusion chromatography (Fig. 4.13 C). Due to low amounts of protein an analytical S200 was used. The chromatography showed two prominent peaks, which were analyzed on SDS PAGE (Fig. 4.13 A lanes: SEC peak 1-2). While the second peak contained an excess of hRbm7, the first peak contained both proteins in the complex. No void peak could be observed. The complex appeared to be reasonably pure except for faint degradation bands and a slight His-GST-tag contamination (Fig. 4.13 A: SEC band at ~ 28 kDa). Complex containing fractions were pooled and concentrated. The A_{260}/A_{280} ratio was

~ 0.75 indicating a remaining slight nucleic acid contamination. Nevertheless the complex was subjected to crystallization.

A.



B.



C.

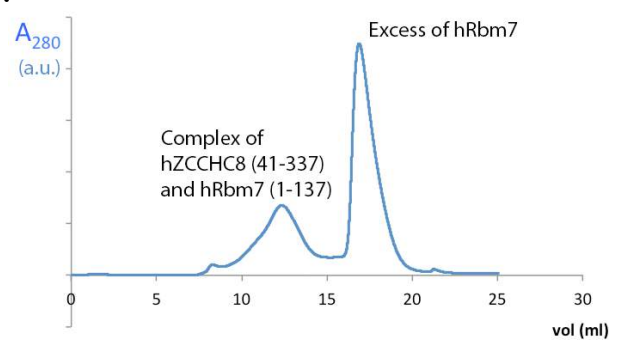


Figure 4.13: Purification of short construct (41-337) hZCCHC8 - optimized protocol.

A. SDS PAGE. To analyse the progress of the purification samples were taken after every purification step. CE = cell extract. SUP = supernatant. FT = flow-through. W = wash. E = Elution. Dil = diluted sample. NN = Nickel-NTA. SEC = size exclusion chromatography. hZCCHC8 (41-337) is indicated by a red arrow. hRbm7 (1-137) is highlighted by a blue arrow. **B.** Q-Ion Exchange Chromatogram: Absorption at 280 nm, conductivity and concentration were monitored during the run. The protein was eluted by at 40% Buffer B. **C.** SEC profile: Elution of hZCCHC8 from the SEC column was followed by monitoring absorption at 280 nm. Three peaks were observed, the first at ~ 12 ml, the second at ~ 18 ml . The third peak represents an excess of hRbm7 (1-137).

4.6 hZCCHC8 is the scaffolding protein in the NEXT complex

In order to analyze the interaction of the three proteins within the NEXT complex pull-downs with full length proteins were performed. Herefore GST-tagged versions of the NEXT proteins were mixed with untagged and a pull-down using GSH-agarose was done. Fig. 4.14 shows SDS PAGE analyses of the pull-downs. Upper gels: GST-tagged hMtr4 and GST-tagged hZCCHC8 FL proteins were tested for interaction with FL MBP-tagged hRbm7. The left gel shows the input- right gel shows the elution. Presence of individual proteins is represented by a plus sign on top of the SDS PAGE lane whereas absence is indicated by a minus sign. Red star indicates the GST tagged hZCCHC8 107 kDa protein (protein migrates higher on SDS PAGE). MBP-tagged hRBM7 protein is highlighted by a blue star. The FL protein is 74 kDa large, however, Ni-NTA purification of the full length protein yielded also an additional band, which is assumed to be a specific degradation band (Fig. 4.7). Both bands are visible on the pull-downs shown below, however as the specific degradation band is more easily detectable and for simplicity reasons it is the one which is marked with the blue stars. GST-tagged hMtr4 is indicated by a green star. While the GST-tagged hZCCHC8 pulled down hRbm7 (Fig. 4.14 lane 7), GST-tagged hMtr4 did not show any interaction (Fig. 4.14 lane 9). MBP-tagged hRbm7 did not bind non-specifically to the glutathione beads (Fig. 4.14 lane 10). The showed hRbm7-hZCCHC8 interaction might appear unreliable as only a degradation band of hRBM7 is seen in the pull-down. However, the interaction was confirmed by size exclusion chromatography of short hRbm7 and hZCCHC8 constructs as for example shown in Fig 4.13.

The lower SDS PAGE shows pull-downs of GST-tagged-hMtr4 (145 kDa) with untagged hZCCHC8 (79 kDa, migrates at ~ 100 kDa). As can be seen in Fig. 4.7 lane 14, the GST-tagged hMtr4 was able to pull-down hZCCHC8. hZCCHC8 did not bind non-specifically to the glutathione beads (Fig. 4.7 lane 16).

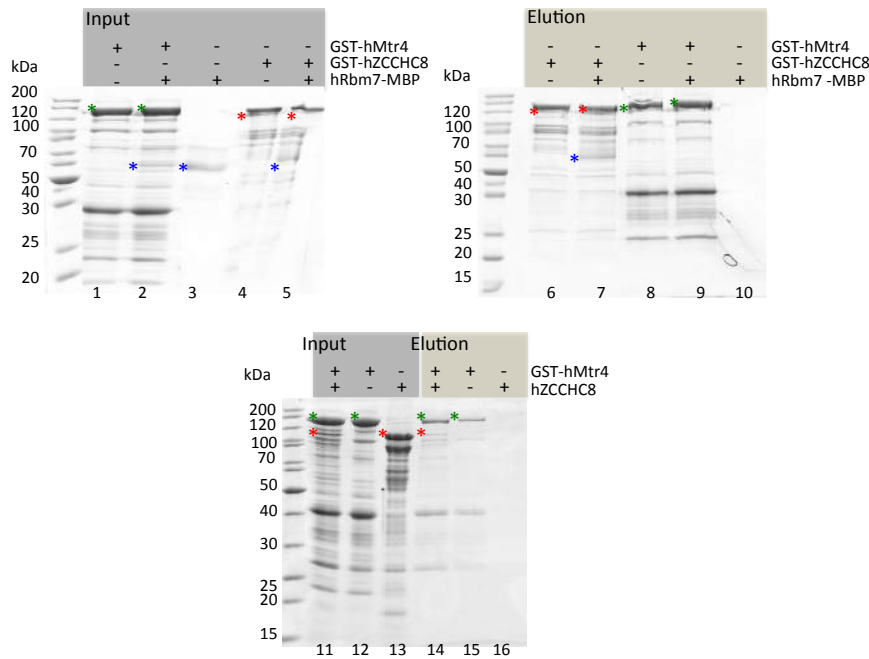


Figure 4.14: Pulldown interaction analyses for the NEXT complex.

Interaction between the NEXT proteins was analyzed with glutathione pulldowns. Upper gel left: Input. Upper gel right: Elution. Lower: Input and Elution. Presence of the respective proteins is indicated with a plus sign; absence with a minus sign. hZCCHC8 is highlighted by a red star; hRbm7 - blue; hMtr4 - green. Input is shown as a control that all the respective proteins were present before pulldown analyses. GST-tagged hMtr4 and hZCCHC8 were tested for interaction with MBP-tagged hRbm7 (upper gels). Lower gel shows interaction analysis of GST-tagged hMtr4 with untagged hZCCHC8.

4.7 Production of a hRbm7-hZCCHC8 Interaction core complex by limited proteolysis

4.7.1 Analytical-scale limited proteolysis of full length hZCCHC8 in complex with hRbm7 (1-137)

To identify the minimal interaction core of the hRbm7-hZCCHC8 complex limited proteolysis was performed in analytical scale. Flexible regions of proteins, which are not participating in the complex formation are accessible for proteases and can therefore be trimmed down. The stability of the hZCCHC8 (1-317) and hRbm7 (1-137) complex towards several different proteases (Trypsin - cuts at Arg, Lys; Elastase - cuts at Ala, Val; Chymotrypsin - cuts at large hydrophobics; GluC - cuts at Glu and Subtilisin-cuts at large uncharged residues) was tested. After incubation of the hZCCHC8-hRbm7 complex with the respective proteases, degradation fragments were visualized by SDS-PAGE analysis (Fig. 4.15).

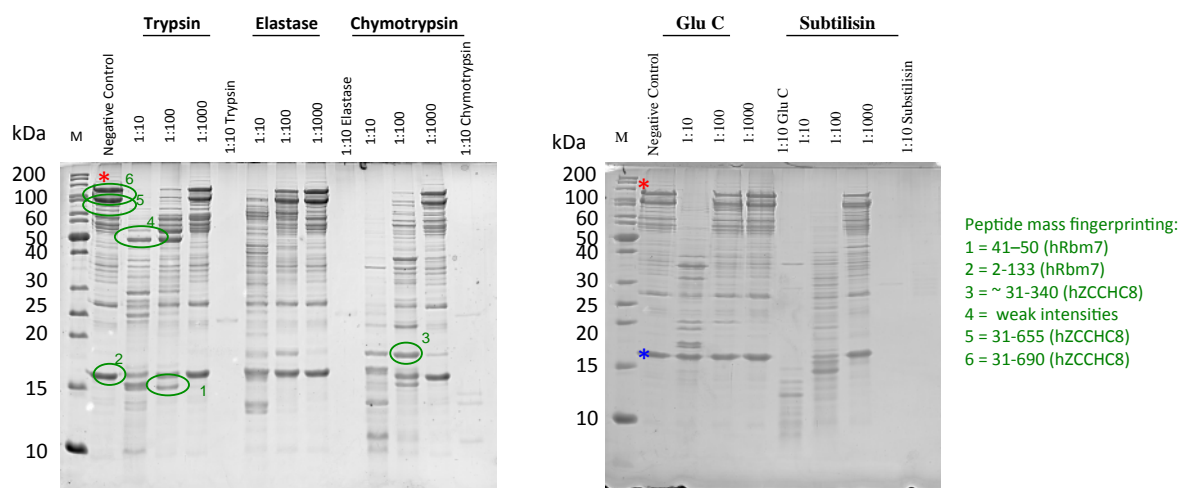


Figure 4.15: Analytical-Scale Limited Proteolysis of full length hZCCHC8 with hRbm7 (1-137).

Trypsin, Elastase, Chymotrypsin, GluC and Subtilisin were used to proteolyze parts of hRbm7 (1-137) and hZCCHC8 FL, which do not participate in complex formation. Herefore the complex was incubated with the respective proteases in decreasing concentration (ranging from 23 ng/ μ l, 2.3 ng/ μ l and 0.23 ng/ μ l final concentration). Bands submitted to peptide mass fingerprinting are labeled by a green circle. The results are shown on the right. hZCCHC8 FL is indicated by a red star. hRbm7 (1-137) is highlighted by a blue star.

Unproteolyzed complex is shown as a control. In the case of GluC, proteolytic cleavage was only observed at high protease concentration. However, digestion with Subtilisin resulted in almost complete degradation. Proteolysis with Chymotrypsin yielded a stable hZCCHC8 fragment (green circle nr. 3), which can be found in Elastase digest as well. Also Trypsin trimmed the complex down to a prominent stable hZCCHC8 fragment (green circle nr. 4), stable even at high protease concentration. hRbm7 (blue star) seems to be protected in the complex against proteolysis as only slight degradation fragments are visible for Trypsin and Elastase digests (for trypsin green circle nr. 1). Fragments possibly representing the minimal regions of the respective proteins for complex formation were subjected to peptide mass fingerprinting. As controls unproteolyzed bands (green circle 5-6 for hZCCHC8 and 2 for hRbm7) were analyzed as well. Due to technical problems the peptide mass fingerprint did not cover the sequences of the short fragments very well. Residues 31-690 were found for the full length (707 aa) hZCCHC8 and 2-133 for hRbm7 (1-137). For hRbm7 fragment only one peptide, which covers 10 amino acids between residues 41-50, was found. The stable hZCCHC8 fragment of the Chymotrypsin and Elastase proteolysis is a C-terminal truncation (31-340) of the full length protein. This is in agreement with already shown interaction of the hZCCHC8 construct (41-337) with hRbm7 (Fig. 4.13). Peptide mass fingerprinting of Trypsin digest band (green circle nr. 4) yielded fragments with very weak intensities that could not reliably be interpreted.

4.7.2 Analytical-scale limited proteolysis of hZCCHC8 (1-317) and hRbm7 (1-137)

As the analytical-scale limited proteolysis of hZCCHC8 FL and hRbm7 (1-137) complex did not give any new results, proteolysis was repeated again with a shorter construct of hZCCHC8 (1-317) in complex with hRbm7. This hZCCHC8 construct is also able to interact with hRbm7 (not shown). Having established Trypsin, Elastase and Chymotrypsin as the proteases best suitable for the purpose of producing a hRbm7-hZCCHC8 core complex, the same proteases were used for the following analysis (Fig. 4.16).

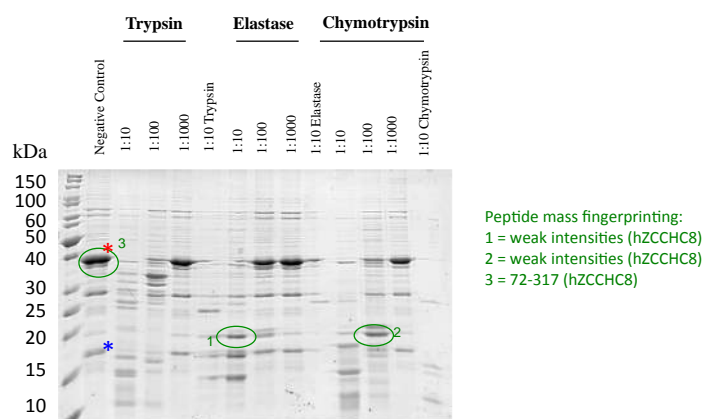


Figure 4.16: Analytical-Scale Limited Proteolysis of hZCCHC8 (1-317) with hRbm7 (1-137).

Short construct of hZCCHC8 (1-317) (red star) and hRbm7 (1-137) (blue star) form a stable complex (neg. control). The complex was subjected to proteolysis by Trypsin, Elastase and Chymotrypsin in decreasing concentration ranging from 23 ng/ μ l, 2.3 ng/ μ l and 0.23 ng/ μ l final concentration) for removal of protein parts not participating in complex formation. Bands submitted to peptide mass fingerprinting are labeled by a green circle. The results are shown on the right.

Proteolysis by Elastase and Chymotrypsin yielded both a degradation band of \sim 20 kDa, which possibly represent minimal regions of hZCCHC8 required for interaction with hRbm7 and were hence subjected to peptide mass fingerprinting. However, due to technical problems, no peptides could be identified.

To test if fingerprinting results would improve in case we would subject fragment bands with stronger intensities, a time-course limited proteolysis with a higher protein concentration (2 mg/ml instead of 1 mg/ml) was performed. Herefore, the complex was incubated with 23 ng/ μ l (final concentration) Elastase over different amount of time (30 min - 3 h). The protease cleavage process was hence monitored over time. Fig. 4.17 shows the obtained results. The proteolyzed fragments resemble the proteolysis of the complex with a 1:100 dilution of Elastase (2.3 ng/ μ l final concentration) (Fig. 4.16). The hZCCHC8 (1-317) band disappeared gradually over time. At the same time two degradation bands at \sim 20 kDa appeared and their signal became stronger with the longer incubation. Both fragments were subjected to peptide mass

fingerprint, which together with some technical optimizations, yielded interpretable results. For the lower band (green circle nr. 1) the analysis found peptides encompassing residues 26-314 of hZCCHC8, the upper band (nr. 2) were peptides containing residues 74-314 of hZCCHC8. Strikingly, the analysis indicated that the lower band in the SDS PAGE (nr. 1) contained a larger fragment of the protein. This could be explained by the longer construct migrating faster due many negatively charged residues. The band of the full 1-317 construct was analysed as a control, here the analyses found peptides between 6-314 of hZCCHC8. Taken together the analysis shows that while the C-terminus the construct of hZCCHC8 (1-317) seems to be stable, the construct is proteolyzed at the N-terminus. Hence, N-terminal residues are most likely not involved in hZCCHC8-hRbm7 complex formation.

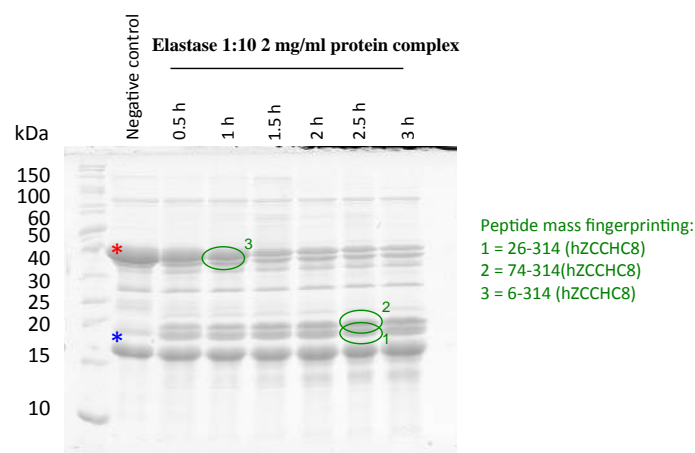


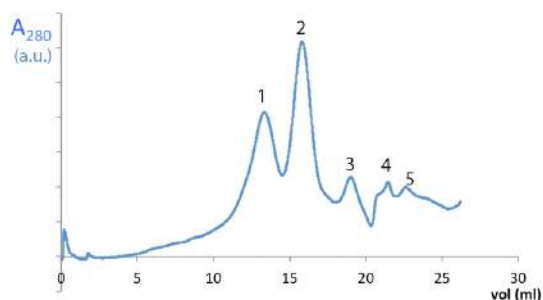
Figure 4.17: Time-course LP of the complex consisting of hZCCHC8 (1-317) and hRbm7 (1-137).

Short construct of hZCCHC8 (1-317) (red star) and hRbm7 (1-137) (blue star) were subjected to a time-course limited proteolysis with Elastase (23 ng/ μ l final concentration) over 3 hours. While over the time the band representing the hZCCHC8 (1-317) construct gradually disappeared, two degradation bands at 20 kDa became more visible (green circle nr.1-2). To analyze whether the two fragments represented minimal regions of hZCCHC8 for interaction with hRbm7, the bands were submitted for peptide mass fingerprinting. Results are shown in green.

To test whether the fragments obtained through the proteolytic process of hZCCHC8 (1-317) and hRbm7 (1-137) are still able to form a complex, the limited proteolysis set-up was scaled up to 500 μ l. After 30 min of proteolysis with the respective concentration of 23 ng/ μ l Elastase the sample was subjected to size exclusion chromatography. Fragments of hZCCHC8 and hRbm7, which are still able to form a complex, were expected to co-elute in the same peak. The chromatogram showed two prominent peaks (Fig. 4.18 A - labeled 1 and 2), which were analyzed on SDS PAGE (Fig. 4.18 B). Unproteolyzed complex is shown as a control; hZCCHC8 is marked with a red and hRbm7 with a blue star. The proteolyzed complex before size exclusion chromatography is labeled as "1:10 Elastase". The proteolytic cleavage resembles strongly the 1:10 Elastase proteolysis shown in Fig. 4.16. The first peak in Fig. 4.18 contains the upper band at \sim 20 kDa (presumably hZCCHC8 fragment) and hRbm7 (1-137) at 17 kDa. The lower

fragment at ~ 13 kDa coeluted with hRbm7 in the second peak. Assuming both fragments are hZCCHC8 would indicate that both fragments form a stable complex with hRbm7.

A.



B.

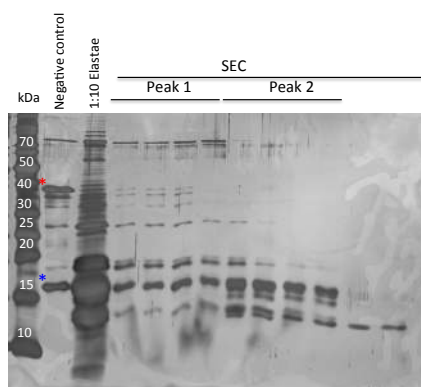


Figure 4.18: Size-exclusion chromatography of hZCCHC8 (1-317) and hRbm7 (1-137) analytical-scale limited proteolysis.

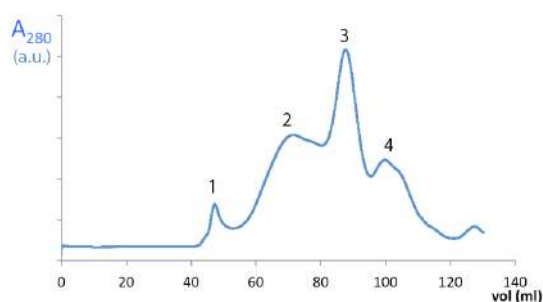
Stability of the trimmed hRbm7-hZCCHC8 complex was tested by size exclusion chromatography. The chromatogram showed several small and two prominent peaks (A 1-2), which were analyzed on SDS PAGE visualised by silver staining (B). hZCCHC8 (1-317) is indicated by a red star. hRbm7 (1-137) is highlighted by a blue star. The first peak contained hRbm7 (17 kDa) and presumably a hZCCHC8 fragment of 20 kDa. A smaller (presumably hZCCHC8) fragment of ~ 13 kDa co-eluted with hRbm7 in the second peak together with an additional fragment of ~ 15 kDa.

4.7.3 Preparative-Scale limited proteolysis of hZCCHC8 (1-317) and hRbm7 (1-137) complex

Purification of the hZCCHC8 (1-317) construct followed the same protocol as described for the hZCCHC8 (41-337) construct in 4.5.5. After the dialysis steps to lower the salt concentration, the complex hZCCHC8 (1-317) with hRbm7 (1-137) was diluted to a concentration of 1 mg/ml and subjected to Elastase proteolysis over night. The optimal Elastase concentration for an over night proteolysis of the 1 mg/ml complex was determined before in analytical scale (data not shown). A 1:200 dilution of Elastase (1.5 ng/ μ l final concentration) was found optimal and was therefore also used in preparative scale. After an over night incubation with the

Elastase, the reaction was stopped with AEBSF and the proteolyzed complex was subjected to SEC. The chromatogram (Fig. 4.19 A) showed one prominent peak (3) and several smaller peaks. The peaks were analyzed by SDS PAGE (Fig. 4.19 B). The smaller peaks contained individual degradation by-products and fragments not forming a complex. The prominent peak (3) contained the same fragments already observed in the second peak of the analytical-scale size exclusion chromatography (Fig. 4.18 A, peak 2.). Strikingly, fragments, which co-eluted in the first peak of analytical-scale chromatography were not observed in preparative-scale. Hence, the complex was more drastically proteolyzed in the preparative-scale than in analytical. Based on previous results, it is very likely that the band in Fig. 4.19 B at ~ 17 kDa is hRbm7. One could therefore speculate, that the other fragments originate from hZCCHC8. One of those bands, therefore, possibly represents the regions of hZCCHC8, which are still able to bind to hRbm7. The corresponding fractions were pooled, concentrated and subjected for crystallization.

A.



B.

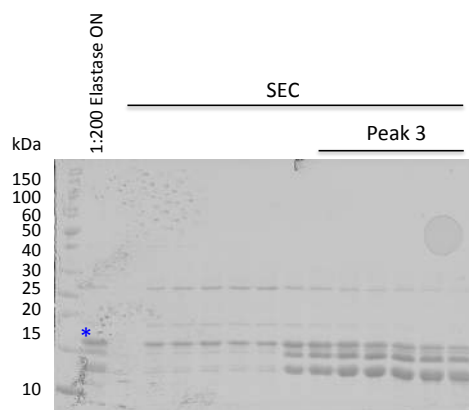


Figure 4.19: Analysis of large-scale limited proteolysis by SEC.

For preparative-scale limited proteolysis, the purified complex of hZCCHC8 (1-317) and hRbm7 (1-137) was proteolyzed with Elastase (1.5 ng/ μ l final conc) over night. After concentration the trimmed complex was separated from fragments via size exclusion chromatography (**A**), which exhibited one prominent (3) and several small peaks. Analysis of the prominent peak (3) by SDS PAGE (**B**) showed the presence of hRbm7 (1-137) (blue star) together with two fragments of ~ 15 and 12 kDa.

4.8 Towards the identification of regions important for hRbm7-hZCCHC8 complex formation

4.8.1 Critical residues for interaction with Rbm7 lie between residues 264-317 of hZCCHC8

The hZCCHC8 (1-263) construct was purified following the same protocol as described for hZCCHC8 (41-337) construct in 4.5.5. Following unfolding in 8 M urea, refolding in presence of hRbm7 and several dialysis steps to gradually lower the salt concentration, both proteins were subjected to size exclusion chromatography (Fig. 4.20 A).

A.

B.

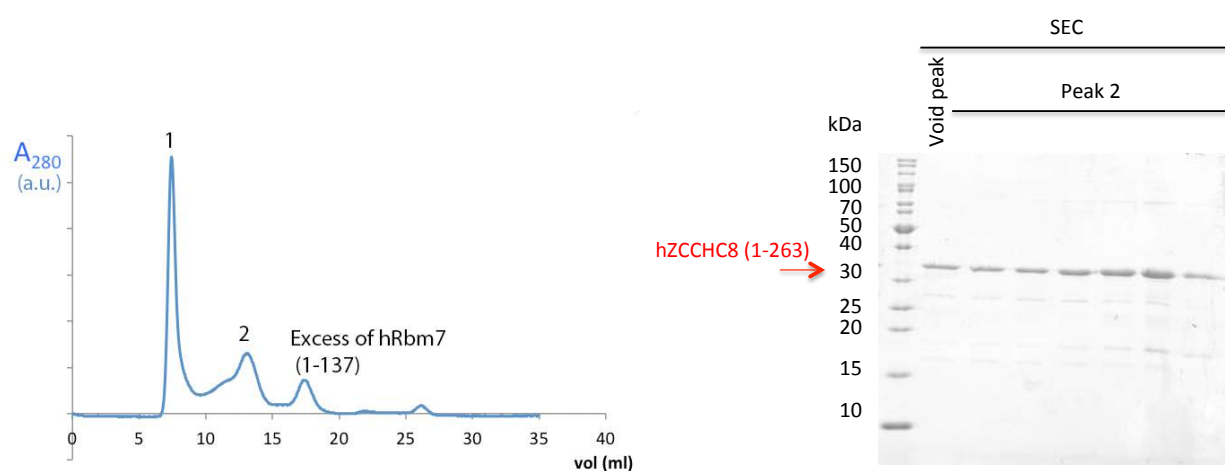


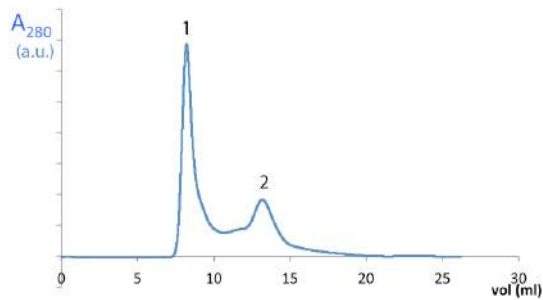
Figure 4.20: Size-exclusion chromatography of hZCCHC8 (1-263) and hRbm7 (1-137).

Following unfolding via urea, the hZCCHC8 (1-263) construct was refolded in presence of hRbm7 (1-137). Complex formation was tested via size exclusion chromatography (A). The chromatogram exhibited one prominent peak in the void volume (1) and two non-void peaks. SDS PAGE analysis (B) of the peaks showed, that peak 1 and 2 contained only the hZCCHC8 (1-263) construct indicating that the two proteins failed to form a complex.

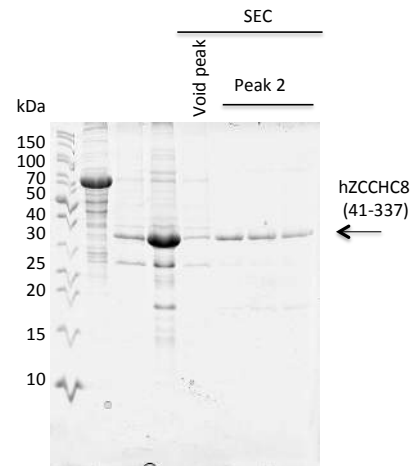
The chromatogram showed a prominent peak in the void volume (1) and two additional peaks. From previous studies it is known that the peak at 17 ml originates from hRbm7. The other two peaks were analyzed on SDS-PAGE (Fig. 4.20 B). Notably both peaks contained only hZCCHC8 (1-263) (30 kDa) indicating that this construct failed to form a complex with hRbm7. To support this the same hZCCHC8 construct was purified under identical conditions with the exception that hRbm7 was omitted during the refolding step. Subsequently, the protein was subjected to size exclusion chromatography as done before with the mixture of hZCCHC8 and hRbm7 (Fig. 4.21 A). Comparing the two size exclusion chromatograms in Fig. 4.20 A and Fig. 4.21 A, one can see that the two profiles look very similar except that the peak corresponding to

hRbm7 is missing. Non-void peak fractions containing the hZCCHC8 (1-263) construct were pooled and concentrated. To verify that the construct is not able to bind hRbm7, it was mixed with hRbm7 in excess and incubated for 30 min to allow complex formation. Subsequently, complex formation was analyzed via size exclusion chromatography (Fig. 4.21 C). The excess of hRbm7 is clearly visible as a large peak at 17 ml. hZCCHC8 (1-263) eluted alone in the first peak as confirmed by SDS PAGE analysis (Fig. 4.21 D).

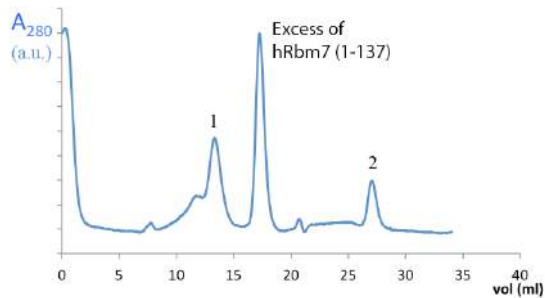
A.



B.



C.



D.

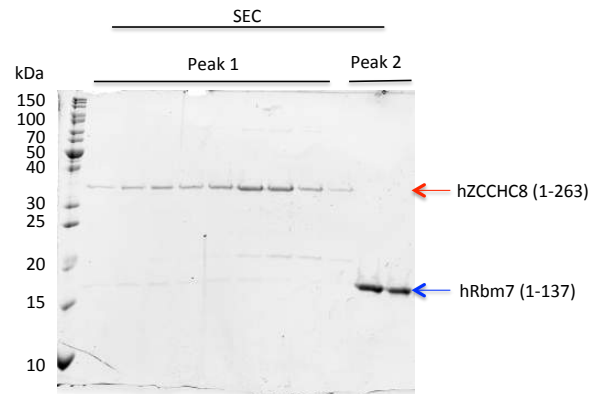


Figure 4.21: hZCCHC8 (1-263) fails to form a complex with hRbm7 (1-137).

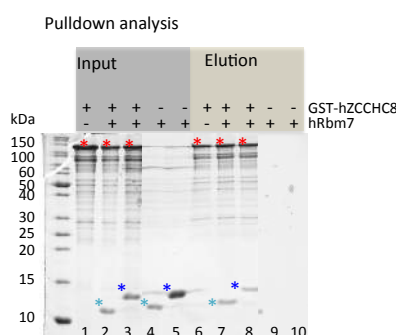
To support the results shown in Fig. 4.20, the hZCCHC8 construct was purified as previously (A and B) and the protein from peak 2 in A was mixed with an excess of hRbm7 and subjected to SEC (C). The peaks were analyzed by SDS-PAGE (D).

4.8.2 The RRM of hRbm7 is sufficient for hZCCHC8 interaction

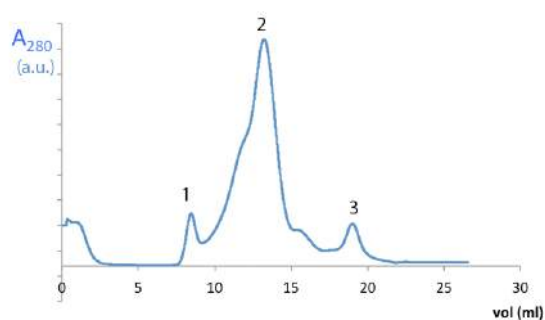
Two short hRbm7 constructs (1-98 and (1-86) were tested for interaction with hZCCHC8 FL via GSH-pulldown (Fig. 4.22 A). The input is shown on the left, the elution on the right. GST-tagged hZCCHC8 FL (red star) was used to test the interaction with the hRbm7 constructs

(blue stars). As can be seen in lanes 7 and 8, hZCCHC8 was able to pull down both hRbm7 constructs, indicating that even the RRM of hRbm7 is sufficient to bind hZCCHC8. To support these results, the complex of hRbm7 (1-86 = RRM) and hZCCHC8 (31-337) was purified in preparative scale (Fig. 4.22 B & C). Herefore, hZCCHC8 was purified after the same protocol as described in 4.5.5. After unfolding with 8 M urea, the protein was refolded in presence of hRbm7 (1-86 = RRM) allowing complex formation. The complex was then dialysed several times to lower the salt concentration. Finally, the complex was subjected to size exclusion chromatography (Fig. 4.22 B). The chromatogram showed one small void peak (1) (the peak contained nucleic acids as no protein band was observed on the corresponding SDS PAGE analysis), one prominent peak with a shoulder (2) and a small hRbm7 excess peak at ~ 17 ml (3). The prominent peak was analyzed on SDS PAGE (Fig. 4.22 C) and the presence of both proteins in the peak was confirmed. hZCCHC8 (41-337) shows a slight degradation and a GST tag contamination (28 kDa). The peak fractions containing the complex were nevertheless pooled, concentrated and submitted for crystallization.

A.



B.



C.

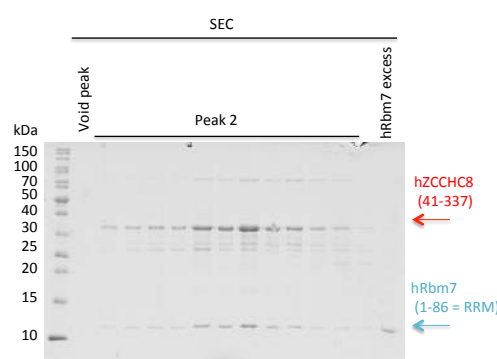


Figure 4.22: The RRM of hRbm7 is sufficient for hZCCHC8 interaction.

Interaction of the hRbm7 RRM (1-86) and hZCCHC8 FL was analyzed by a GSH pull-down analysis (A) and in preparative scale (B and C). A. GST-hZCCHC8 FL is marked by a red star, hRbm7 (1-98) by a dark blue and hRbm7 (1-86) by a light blue star. The input is shown on the left - elution on the right. B. The complex of hZCCHC8 (41-337) with the RRM of hRbm7 was purified in preparative scale. C. The presence of the complex in the prominent peak was confirmed by SDS PAGE analysis.

4.9 Crystallization experiments

Crystallization experiments were performed on complexes containing different hZCCHC8 and hRbm7 constructs, with different concentrations and at different temperatures. The attempts are summarized in the following table:

Table 4.2: Crystallization of different hZCCHC8-hRbm7 complexes.

Protein complex	Conc. (mg/ml)	Screens	Temp.	Hits
Z8 (41-337) with R7 (1-137)	3/6	Qiagen JSCG/Morpheus/ Complex screen 1-2	RT	-
Z8 (41-337) with R7 (1-86 = RRM)	4	Qiagen JSCG/Morpheus/Complex screen 1-2	RT	-
Z8 (1-317) with R7 (1-137) after prep. LP	3	Qiagen JSCG/Morpheus/Qiagen PACT/Wizard/Hampton	4°C	Spherulites

Z8 = hZCCHC8. R7 = hRBM7. Prep. LP = preparative-scale limited proteolysis.

In many screens the minimal core complex of hZCCHC8-hRbm7, which was obtained after limited proteolysis (see Results section 4.7.3), yielded spherulite-like structures. The conditions had a basic pH in common. Examples of the spherulite-like structures are shown in Fig. 4.23. Several of those structures were fished for analysis on SDS PAGE, however, even with silver staining no bands were visible.

A.



B.



Figure 4.23: Spherulite-like structures of the hZCCHC8-hRbm7 minimal core complex. Examples of spherulite-like structures, which were obtained for the minimal hZCCHC8-hRbm7 core complex. **A.** 0.3 M Magnesium Formate. 0.1 M Tris pH 8.5. **B.** 0.1 M Bicine pH 9.0 10% PEG 6000.

5 Summarized discussion

The NEXT complex is a cofactor of the human nuclear exosome and consists of three proteins. hMtr4 is a Ski2-like helicase present in lower and higher eukaryotes, whereas hRbm7 and hZCCHC8 are metazoan-specific proteins. While the structure of Mtr4 is known from yeast studies (Weir et al., 2010; Jackson et al., 2010), there is no structural information for the other proteins comprising the NEXT complex. This thesis presents a first approach towards the structural and biochemical studies on the interaction between the NEXT components.

Biochemical *in vitro* studies and crystallization experiments rely on the abundance of the purified recombinant protein of interest. Therefore, in addition to already existing NEXT protein constructs, new hZCCHC8 and hRbm7 constructs were cloned based on bioinformatics and limited proteolysis results (see Results section 4.2). The thesis showed the purification of all three NEXT proteins. hMtr4 full length and a short construct lacking the N-terminus could both be purified as shown in Results section 4.3. The short construct represents the yeast Mtr4 Δ N construct, which was crystallized in Weir et al. (2010). Full length hRbm7 precipitates once the MBP tag is cleaved. However, purification of three short hRbm7 constructs yielded high amounts of pure protein (see Results section 4.4). Purification of hZCCHC8 was complicated by aggregation of the protein as indicated by elution in the void volume during size exclusion chromatography. Different approaches to solve the problem were used. The problem could be solved by unfolding the protein in 8 M urea and refolding in presence of the binding partner, hRbm7. When this unfolding/refolding protocol is applied in the purification of the hZCCHC8 full length protein, already a small portion of protein does not elute in the void volume anymore (see Results section 4.5.4). Even more, a N- and C-terminal truncated hZCCHC8 construct, comprising residues 41-337, behaves much better and does not elute in the void volume at all (see Results section 4.5.5). However, the protocol remains to be optimized as following problems remained even after inclusion of hZCCHC8 unfolding/refolding in the purification: (1) The peak corresponding to the hRbm7 and hZCCHC8 complex still contains a shoulder (2) GST tag contamination (3) low protein yield of the short hZCCHC8 constructs due to low solubility. To circumvent this one could purify the hZCCHC8 protein under denaturing conditions from inclusion bodies isolated from *E. coli*.

Pulldown analysis performed in this thesis showed that hZCCHC8 can independently bind hMtr4 and hRbm7, whereas no interaction could be observed between hMtr4 and hRbm7 (see Results section 4.6) (summarized in Fig. 5.1) This suggests that hZCCHC8 is the scaffolding protein in NEXT.

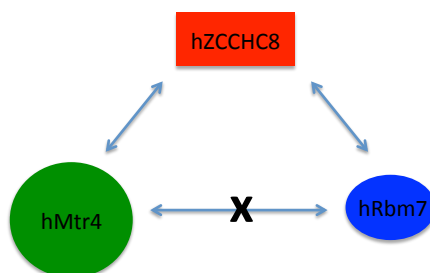


Figure 5.1: Schematic representation of interaction between the NEXT proteins.

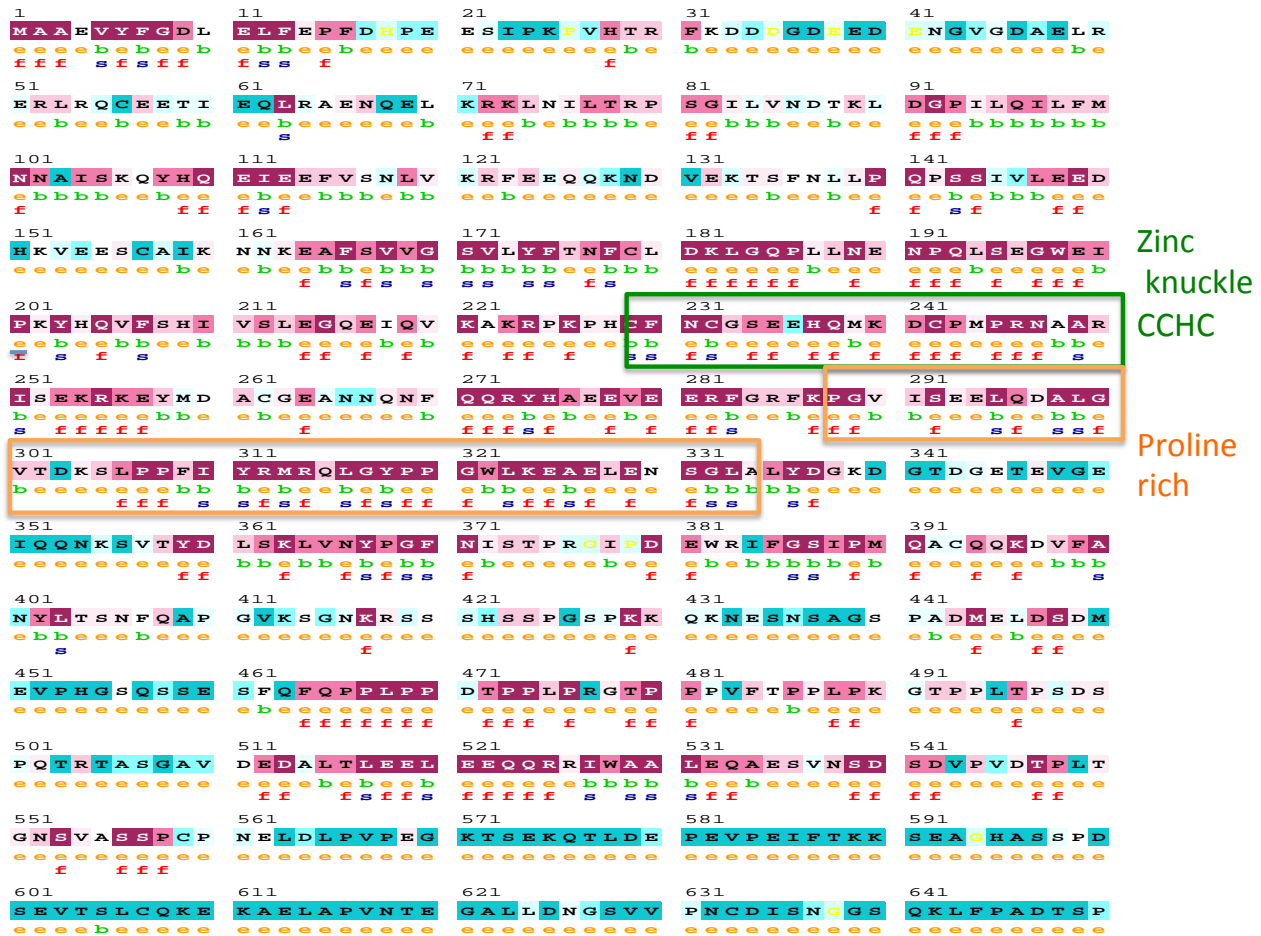
So far there are two Ski2-like helicase containing complexes known: (1) TRAMP and (2) the Ski complex. In both complexes the RNA helicases associate with a different set of proteins: in TRAMP, Mtr4 interacts with Trf4 and Air2. The Ski complex is formed by Ski2 with Ski3 and two copies of Ski8. Interestingly, in both complexes the interaction takes place on the helicase core of Mtr4/Ski2, whereas the SK insert domain is not required (Falk et al., 2014; Halbach et al., 2013). It would be therefore interesting to test in the future, if the SK domain of hMtr4 is also dispensable for the NEXT complex formation.

Insights into minimal regions of the hZCCHC8-hRbm7 interaction were gained by constructs screening. The short hZCCHC8 constructs encompassing residues 1-263 failed to form a complex with hRbm7 (see Results section 4.8.1), whereas hZCCHC8 (1-317) was able to bind hRbm7. This indicates that important residues for hRbm7 interaction lie between the residues 264-317. The RRM of hRbm7 was found to be sufficient for mediating interaction with hZCCHC8 (see Results section 4.8.2) implicating a different function in addition to RNA binding as suggested by Lubas et al. (2011). RRM domains have been previously reported to mediate protein-protein interactions. Hereby, one can distinguish between RRMs, which can interact with other RRMs (1), RRMs which can bind non-RRM proteins and RNA (2) and RRMs which can only bind to proteins but not to RNA (3) (reviewed in Maris et al. 2005). The mechanisms of the RRM-protein interactions differs in the three classes. Binding of RNA is enabled by an accessible β -sheet in the respective RRM motifs. In RRMs, in which RNA binding is abolished, the β -sheet is either participating in the protein interaction or the aromatic residues important for establishing interaction with RNA bases are missing in the RNP1/2. The later case is a novel form of RRMs, referred to as UHMs, which in addition to divergence in RNP motifs, exhibit an Arg-X-Phe motif in the loop region following Helix B and conserved acidic residues in Helix A (reviewed in Kielkopf et al., 2004). The RNP-like motifs as well as the Arg-X-Phe motif contribute to formation of a hydrophobic pocket, in which the UHM ligand usually inserts a characteristic, conserved Trp (exceptions are found e.g. in the RES complex as shown in Wysoczański et al., 2014). Upstream of the Trp the ligand exhibits basic residues. hRbm7 seems to belong to the second class of protein-interacting RRMs. First of all, work by Lubas et al. (2011) suggested that hRbm7 binds RNA/PROMPTs. Secondly the RNP motifs in hRbm7

are highly conserved and show no divergence. Finally it has no Arg-X-Phe motif in the loop connecting the last helix (Helix B) and the β -strand. Regions 264-317 of hZCCHC8, which have been identified as important regions for hRBM7 interaction, do not contain a conserved Trp and are rather acidic. While the β -sheet of hRbm7 is probably participating in RNA recognition, the mechanisms of hRbm7-hZCCHC8 interaction remains unknown. A structure of hRbm7-hZCCHC8 complex would gain insights into the mechanism.

Finally, reconstitution of the hRbm7-hZCCHC8 core complex was probed via limited proteolysis in analytical and preparative scale (see Results section 4.7). Analytical scale limited proteolysis of the short construct hZCCHC8 (1-317) and hRbm7 (1-137) suggested that the N-terminus of hZCCHC8 is not participating in the complex formation (see Results section 4.7.2). The minimal core complex could be produced by preparative-scale limited proteolysis and yielded spherulite-like structures in crystallization experiments (see Results sections 4.7.3 & 4.9). It remains to be tested whether seeding would induce crystallization.

6 Supplementary Material



The conservation scale:

1 2 3 4 5 6 7 8 9
 Variable Average Conserved

- e - An exposed residue according to the neural-network algorithm.
- b - A buried residue according to the neural-network algorithm.
- f - A predicted functional residue (highly conserved and exposed).
- s - A predicted structural residue (highly conserved and buried).
- x - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Figure S1: hZCCHC8 is highly conserved around the Zn knuckle and proline rich domain.

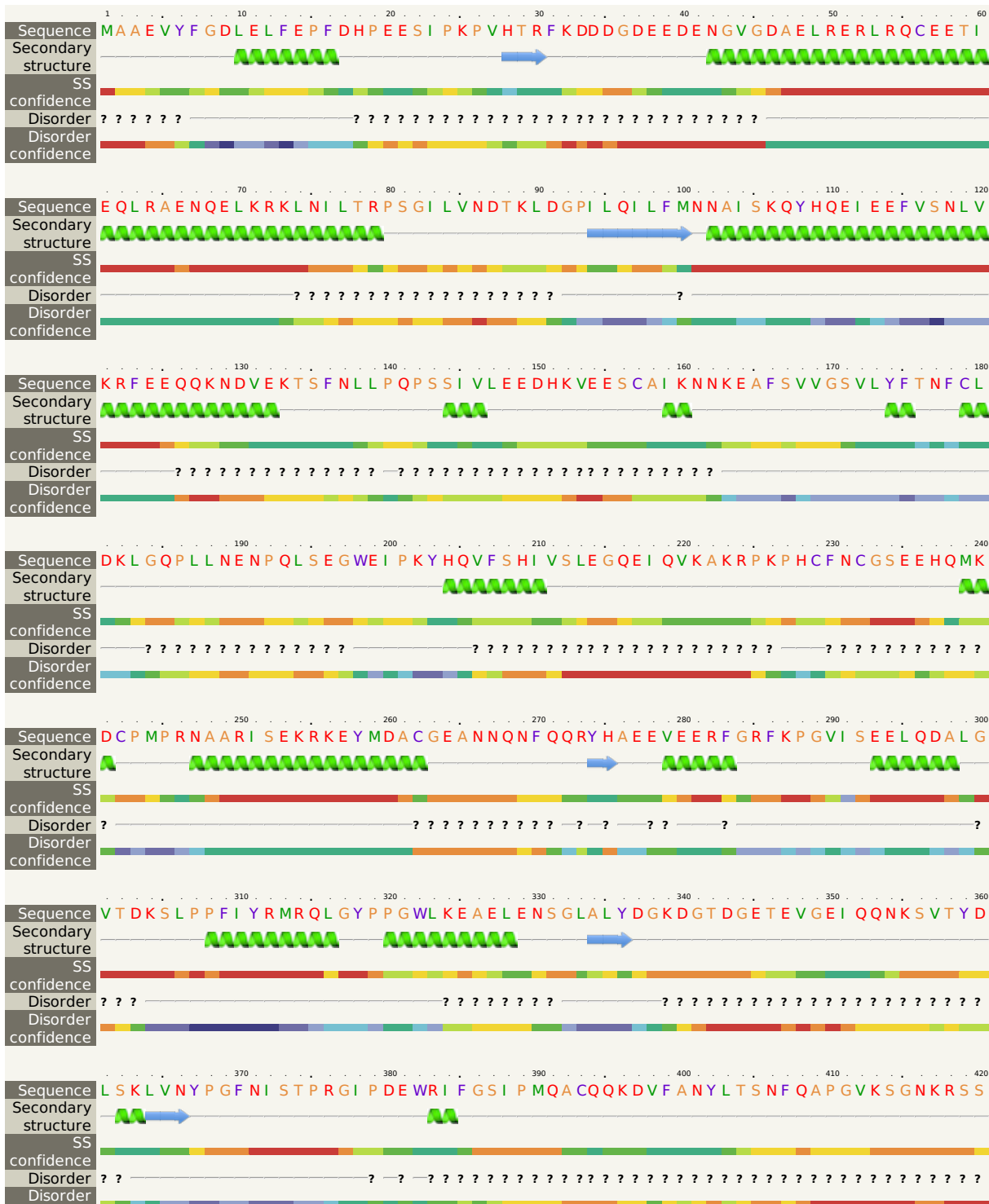
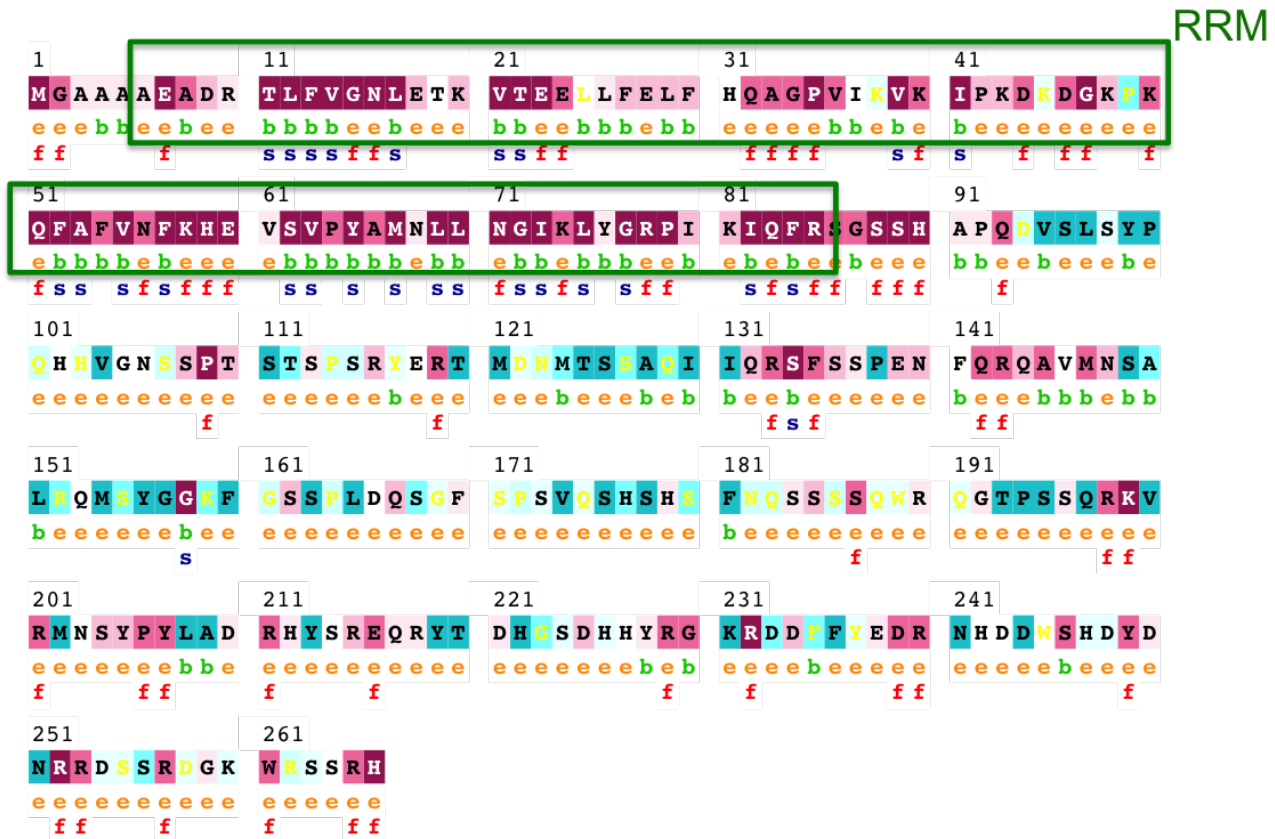




Figure S3: hRbm7-secondary structure prediction.



The conservation scale:



Variable Average Conserved

- e - An exposed residue according to the neural-network algorithm.
- b - A buried residue according to the neural-network algorithm.
- f - A predicted functional residue (highly conserved and exposed).
- s - A predicted structural residue (highly conserved and buried).
- X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Figure S4: hRbm7 is highly conserved around the RRM motif in metazoa.

7 Acknowledgments

Finally, I want to use the opportunity to express my gratitude to several people for making this last half a year pleasant and productive: First I want to thank Prof. Dr. Elena Conti for giving me the possibility to be part of her group during the practical work of the master thesis. This gave me the chance of working in a well-equipped lab, which improved and broadened my practical skills.

A very special thanks goes to my supervisor Dr. Sebastian Falk, who always encouraged me and was open for questions. With his sheer unlimited patience he extensively introduced me into protein biochemistry. While he supported my independent scientific working, he was also always helpful when I still needed advice. In the last weeks of thesis writing I appreciated his experience in scientific notation and I am deeply grateful for his editing assistance. I also have to acknowledge Dr. Christian Benda for helpful discussions and sharing his knowledge.

Finally, thanks goes also to the facility of MPI for Biochemistry, Martinsried, for all the help with the laboratory work.

Bibliography

- Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. and Sharp, P. A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* *499*, 360–363.
- Amaral, P. P., Dinger, M. E., Mercer, T. R. and Mattick, J. S. (2008). The Eukaryotic Genome as an RNA Machine. *Science* *319*, 1787–1789.
- Andersen, P. R., Domanski, M., Kristiansen, M. S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M., Sandberg, R., Hyman, A., LaCava, J., Rout, M. P., Andersen, J. S., Bertrand, E. and Jensen, T. H. (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat Struct Mol Biol* *20*, 1367–1376.
- Anderson, J. S. J. and Parker, R. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *The EMBO Journal* *17*, 1497–1506.
- Aslanidis, C. and de Jong, P. J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research* *18*, 6069–6074.
- Blasius, M., Wagner, S. A., Choudhary, C., Bartek, J. and Jackson, S. P. (2014). A quantitative 14-3-3 interaction screen connects the nuclear exosome targeting complex to the DNA damage response. *Genes & Development* *28*.
- Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E. and Conti, E. (2009). The Yeast Exosome Functions as a Macromolecular Cage to Channel RNA Substrates for Degradation. *Cell* *139*, 547–559.
- Burkard, K. T. D. and Butler, J. S. (2000). A Nuclear 3-5 Exonuclease Involved in mRNA Degradation Interacts with Poly(A) Polymerase and the hnRNA Protein Npl3p. *Molecular and Cellular Biology* *20*, 604–616.
- Chekanova, J. A., Shaw, R. J., Wills, M. A. and Belostotsky, D. A. (2000). Poly(A) Tail-dependent Exonuclease AtRrp41p from *Arabidopsis thaliana* Rescues 5.8S rRNA Processing and mRNA Decay Defects of the Yeast *ski6* Mutant and Is Found in an Exosome-sized Complex in Plant and Yeast Cells. *Journal of Biological Chemistry* *275*, 33158–33166.
- Deo, R. C., Bonanno, J. B., Sonenberg, N. and Burley, S. K. (1999). Recognition of Polyadenylate RNA by the Poly(A)-Binding Protein. *Cell* *98*, 835–845.

- Deutscher, M. P., Marshall, G. T. and Cudny, H. (1988). RNase PH: an *Escherichia coli* phosphate-dependent nuclease distinct from polynucleotide phosphorylase. *Proceedings of the National Academy of Sciences* *85*, 4710–4714.
- Falk, S., Weir, J., Hentschel, J., Reichelt, P., Bonneau, F. and Conti, E. (2014). The Molecular Architecture of the TRAMP Complex Reveals the Organization and Interplay of Its Two Catalytic Activities. *Molecular Cell* *55*, 856–867.
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J. and Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research* *42*, D222–D230.
- Gilbert, W. (1986). Origin of life: The RNA world. *Nature* *319*, 618–618.
- Halbach, F., Reichelt, P., Rode, M. and Conti, E. (2013). The Yeast Ski Complex: Crystal Structure and RNA Channeling to the Exosome Complex. *Cell* *154*, 814–826.
- Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. (1999). Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature* *398*, 579–585.
- Jackson, R. N., Klauer, A. A., Hintze, B. J., Robinson, H., van Hoof, A. and Johnson, S. J. (2010). The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J* *29*, 2205–2216.
- Kadlec, J., Izaurralde, E. and Cusack, S. (2004). The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat Struct Mol Biol* *11*, 330–337.
- Kelley, L. A. and Sternberg, M. J. E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protocols* *4*, 363–371.
- Kielkopf, C. L., Luecke, S. and Green, M. R. (2004). U2AF homology motifs: protein recognition in the RRM world. *Genes & Development* *18*, 1513–1526.
- Liu, Q., Greimann, J. C. and Lima, C. D. (2006). Reconstitution, Activities, and Structure of the Eukaryotic {RNA} Exosome. *Cell* *127*, 1223 – 1237.
- Lorentzen, E., Basquin, J. and Conti, E. (2008). Structural organization of the RNA-degrading exosome. *Current Opinion in Structural Biology* *18*, 709 – 713.
- Lubas, M., Christensen, M., Kristiansen, M., Domanski, M., Falkenby, L., Lykke-Andersen, S., Andersen, J., Dziembowski, A. and Jensen, T. (2011). Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex. *Molecular Cell* *43*, 624–637.
- Makino, D. L., Baumgartner, M. and Conti, E. (2013). Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* *495*, 70–75.
- Maris, C., Dominguez, C. and Allain, F. H.-T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal* *272*, 2118–2131.

- Mazza, C., Ohno, M., Segref, A., Mattaj, I. W. and Cusack, S. (2001). Crystal Structure of the Human Nuclear Cap Binding Complex. *Molecular Cell* *8*, 383–396.
- Miller, J. H. (1972). *Experiments in molecular genetics*. Cold Spring Harbor.
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* *1*, 263–73.
- Nagai, K., Oubridge, C., Jessen, T. H., Li, J. and Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* *348*, 515–520.
- Ntini, E., Jarvelin, A. I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P. R., Andersen, P. K., Preker, P., Valen, E., Zhao, X., Pelechano, V., Steinmetz, L. M., Sandelin, A. and Jensen, T. H. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* *20*, 923–928.
- Price, S. R., Evans, P. R. and Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* *394*, 645–650.
- Sambrook, J. and Russel, D. W. (2001). *Molecular cloning: a laboratory manual*. New York, Cold Spring Harbour Laboratory Press.
- Schmid, M. and Jensen, T. H. (2008). The exosome: a multipurpose RNA-decay machine. *Trends in Biochemical Sciences* *33*, 501 – 510.
- Shamoo, Y., Krueger, U., Rice, L., Williams, K. and Steitz, T. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat. Struct. Biol.* *4*, 215–22.
- van Hoof, A., Lennertz, P. and Parker, R. (2000). Yeast Exosome Mutants Accumulate 3'-Extended Polyadenylated Forms of U4 Small Nuclear RNA and Small Nucleolar RNAs. *Molecular and Cellular Biology* *20*, 441–452.
- Vasiljeva, L. and Buratowski, S. (2006). Nrd1 Interacts with the Nuclear Exosome for 3'-Processing of RNA Polymerase II Transcripts. *Molecular Cell* *21*, 239–248.
- Weir, J. R., Bonneau, F., Hentschel, J. and Conti, E. (2010). Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proceedings of the National Academy of Sciences* *107*, 12139–12144.
- Wysoczański, P., Schneider, C., Xiang, S., Munari, F., Trowitzsch, S., Wahl, M. C., Luhrmann, R., Becker, S. and Zweckstetter, M. (2014). Cooperative structure of the heterotrimeric pre-mRNA retention and splicing complex. *Nat Struct Mol Biol* *21*, 911–918.
- Xu, R., Jokhan, L., Cheng, X., Mayeda, A. and Krainer, A. (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure* *5*, 559–70.